Supplemental Methods

Contact for reagents and resources
Further information and reasonable 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.

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) (1).

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 Genomics LongRanger underwent tumor-normal variant calling using samtools mpileup (v1.7), to produce a single tumor-normal mpileup file (2). The mpileup files were processed to yield variants using the varscan2 somatic function and high confidence variants were isolated using the varscan2 processSomatic 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 fpfilter 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 copynumber function, and further adjusted for GC content using the copycaller function. Adjusted calls underwent segmentation using a recursive algorithm, circular binary segmentation as implemented in the package DNAcopy (3). Segmented CNA was analyzed and readjusted using copycaller with recenter-up or recenter-down, 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 (4). 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
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 (5):

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

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 (6) 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 was created using the as.phylo function from the ape package (7).

Copy number spectrum plots
CNA data from varscan2 copy number analysis were used to visualize copy number gains and losses. The copy number spectrum function for the GenVisR package provides a method of visualizing primary and recurrence CNA changes (8). 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 differences 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 (9). Mutational signatures were generated for each subclonal population group, all groups were above 50 variants, with the exception of 3f (patient 3), 1ef (patient 1) and 2b (patient 2). DeconstructSigs was run using 30 signatures from the

Catalogue Of Somatic Mutations In Cancer public repository as the reference signatures (10). 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 (11,12). 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 (11,13):

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

Here, the leading *1/f* dependence, highlighted by Williams et al. 2016 (12), 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 (11,13).

Due to the sequencing resolution limit (~150x physical coverage) we introduced a lower cut-off on the allele frequencies at $f_{min} = 0.12$ (14). 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. (11).

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**, **Fig. 6B-E** and **Fig. 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. S7E-I**. For a full description of the model, please refer to Supplementary Theory in reference (13).

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. S7J-N**), while the distribution of shared mutations is typically peaked around larger values, as expected (**Fig. S7O-R**).

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 manufacturer's instructions. Viability and number 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, generating ~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 (15).

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 (16) v2.4.2a to hg38 human reference genome (from iGenome). R Bioconductor DESEq2 package (17) 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 (18,19). 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 (20). Assessment of allelic fractions for *ATRX* mutations was

done using an independent dataset (21) available at www.pedpancan.com. GTEx data were assessed at https://www.gtexportal.org/home (22). The Database of Genomic Variants was accessed at http://dgv.tcag.ca/dgv/app/home (23). Developmental human brain gene expression data was extracted from BrainSpan (http://www.brainspan.org).

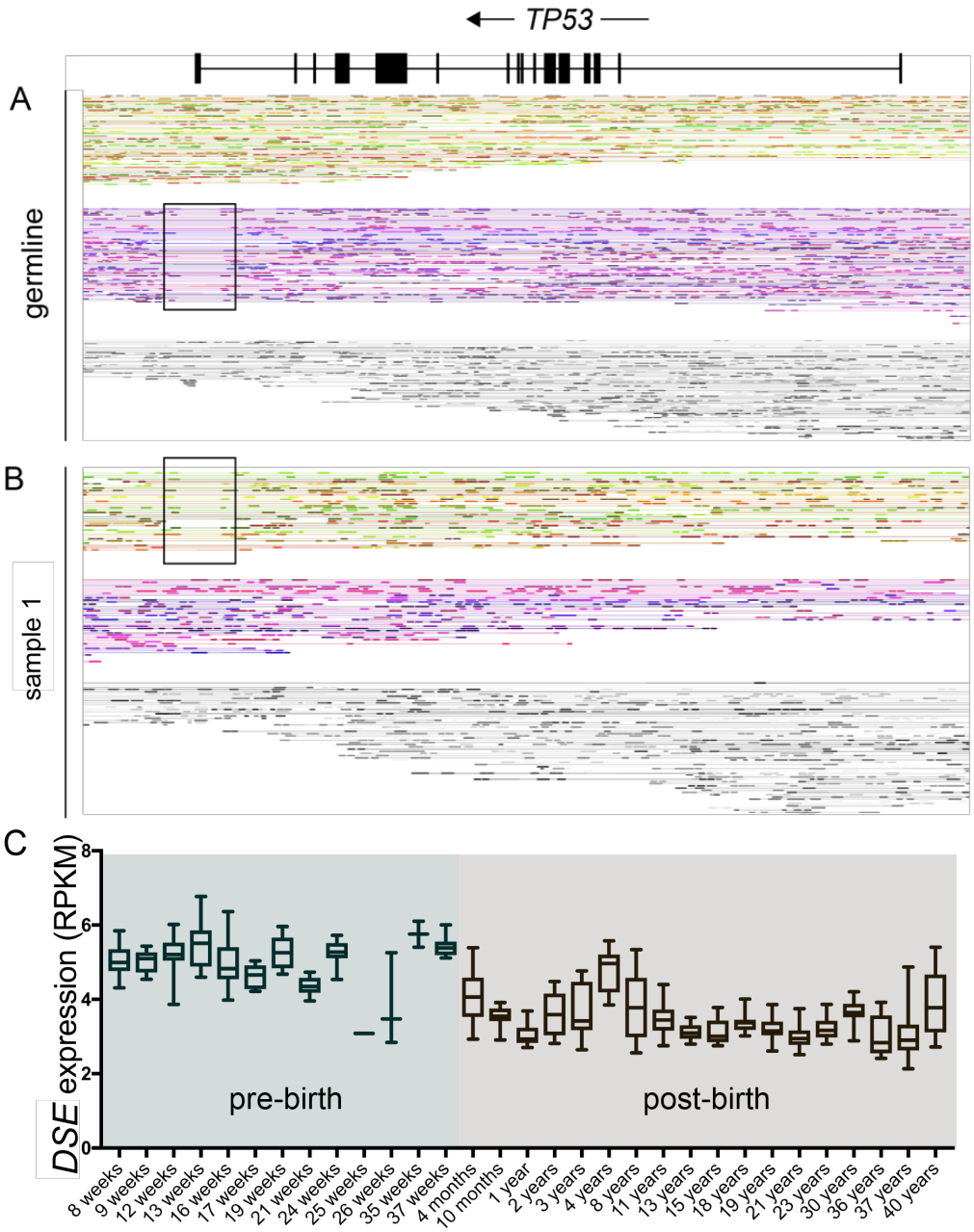
Statistical analysis of enrichment of OMIM genes in the pGBM mutational dataset

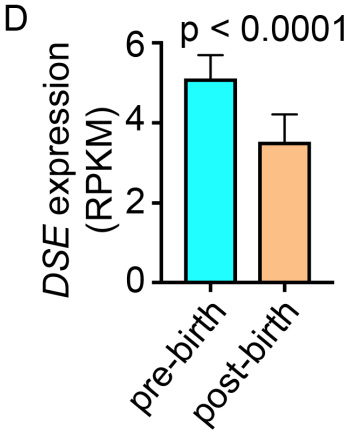
The number *X* of genes harboring non-synonymous, stop gain/stop loss, frame shift, non-frame shift SNVs and indels was identified for primary and recurrent tumors. From these, genes with a phenotype in OMIM were selected (*Y*). The OMIM genes (n = 13,520) with a clinical phenotype were identified using 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.

Supplemental Methods References

1. 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.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
3. Seshan VE, Olshen AB. DNAcopy : A Package for Analyzing DNA Copy Data. Bioconductor Vignette. 2014;1–7.
4. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. 2016;8:289–317.
5. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486:400–4.
6. 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.
7. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90.
8. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: Genomic Visualizations in R. Bioinformatics. 2016;32:3012–4.
9. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
10. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: Somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45:D777–83.
11. Simons BD. Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis. Proc Natl Acad Sci. 2016;113:128–33.
12. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. Nat Genet. 2016;48:238–44.
13. 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.
14. Shin HT, Choi Y La, Yun JW, Kim NKD, Kim SY, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. Nat Commun. 2017;8:1377.
15. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.

16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
18. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
19. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
20. Paugh BS, Qu C, Jones C, Liu Z, Adamowicz-Brice M, Zhang J, et al. Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. J Clin Oncol. 2010;28:3061–8.
21. 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.
22. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.
23. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42:D986-992.


TP53
A
germline
B
sample 1
C
DSE expression (RPKM)
pre-birth
post-birth
8 weeks
9 weeks
12 weeks
13 weeks
16 weeks
17 weeks
19 weeks
21 weeks
24 weeks
25 weeks
26 weeks
35 weeks
37 weeks
4 months
10 months
1 year
2 years
3 years
4 years
8 years
11 years
13 years
15 years
18 years
19 years
21 years
23 years
30 years
36 years
37 years
40 years


D
p < 0.0001
DSE expression (RPKM)
pre-birth
post-birth

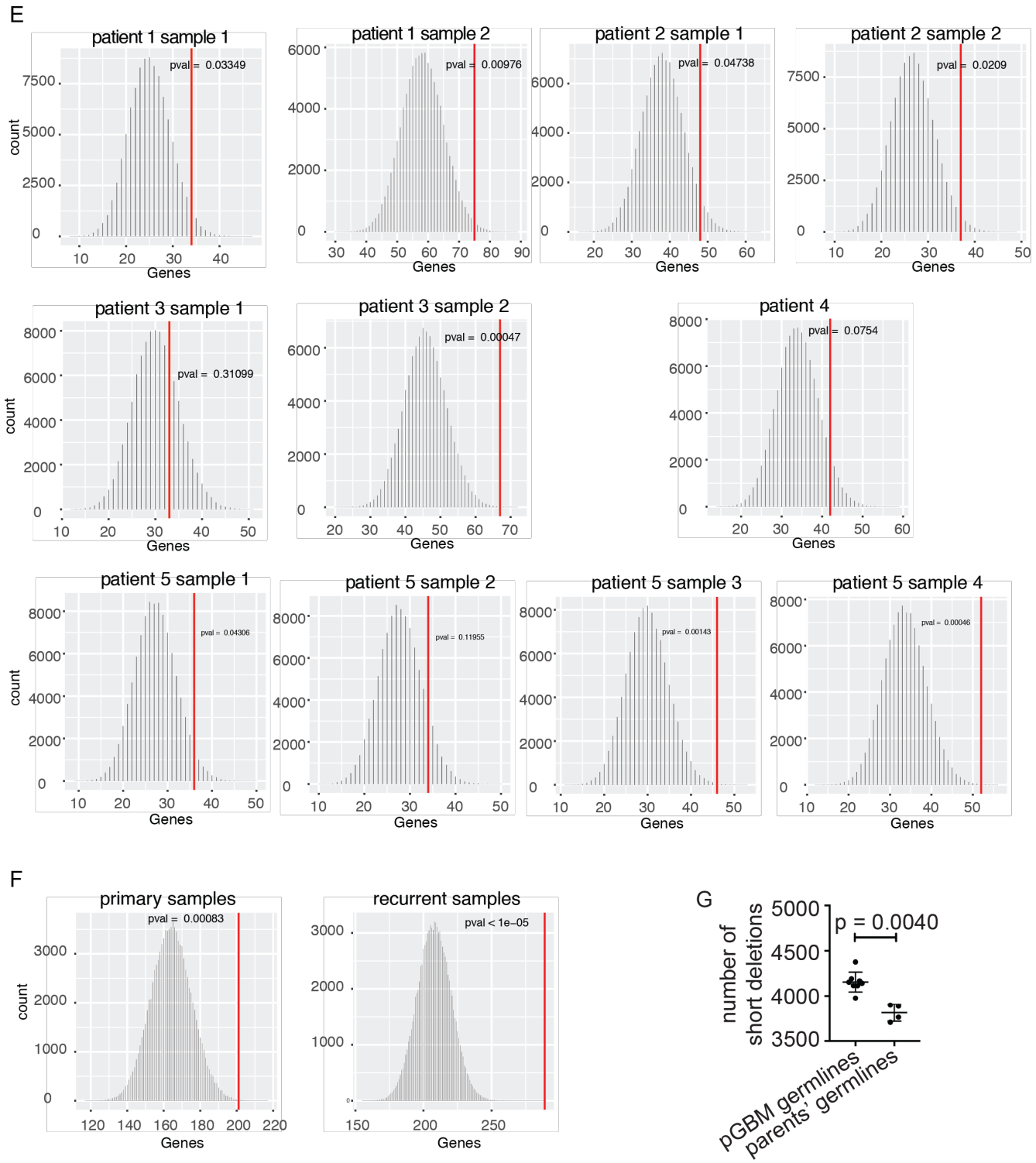


Fig. S1. Heterozygous *TP53* deletion in the germline and tumor of patient 1 and expression timelines of *DSE*.

(A) Whole-genome sequencing with linked-reads shows a heterozygous deletion of the last exon of *TP53* in the germline of patient 1.

(B) Linked-read whole-genome sequencing data for the primary tumor of patient 1.

Linked-reads are clustered based on parental origin (green-orange color scheme for one parental chromosome, and blue-purple color scheme for the chromosome derived from the other parent). Black boxes highlight the site of the deletion. Haplotypes are arbitrarily assigned, hence the attribution of the deletion to haplotypes of different colors in germline and tumor.

(C) RNA-seq data for *DSE* at different developmental time points in the human brain. Data were extracted from BrainSpan.
(D) Summary of expression profiles for *DSE* in pre-birth and post-birth brain tissue shown in (C). Bars represent mean ± SD. P value was calculated with the non-parametric, two-tailed Mann Whitney test.
(E) Permutation tests to assess enrichment of OMIM genes among mutated genes in pGBM samples. The analysis was done for each individual sample.
(F) Permutation tests to assess enrichment of OMIM genes among mutated genes in pGBM samples, grouped as "primary" or "recurrent" samples.
(G) Comparison of short deletions (range: 50-30,000 bp) in the germlines of pGBM patients (n = 8) or control germlines (n = 4). Controls were the germlines of two sets of parents of two pGBM patients. Each column shows the distribution of values in each category, together with mean ± SD. P-value was obtained with non-parametric, two-tailed Mann Whitney test.

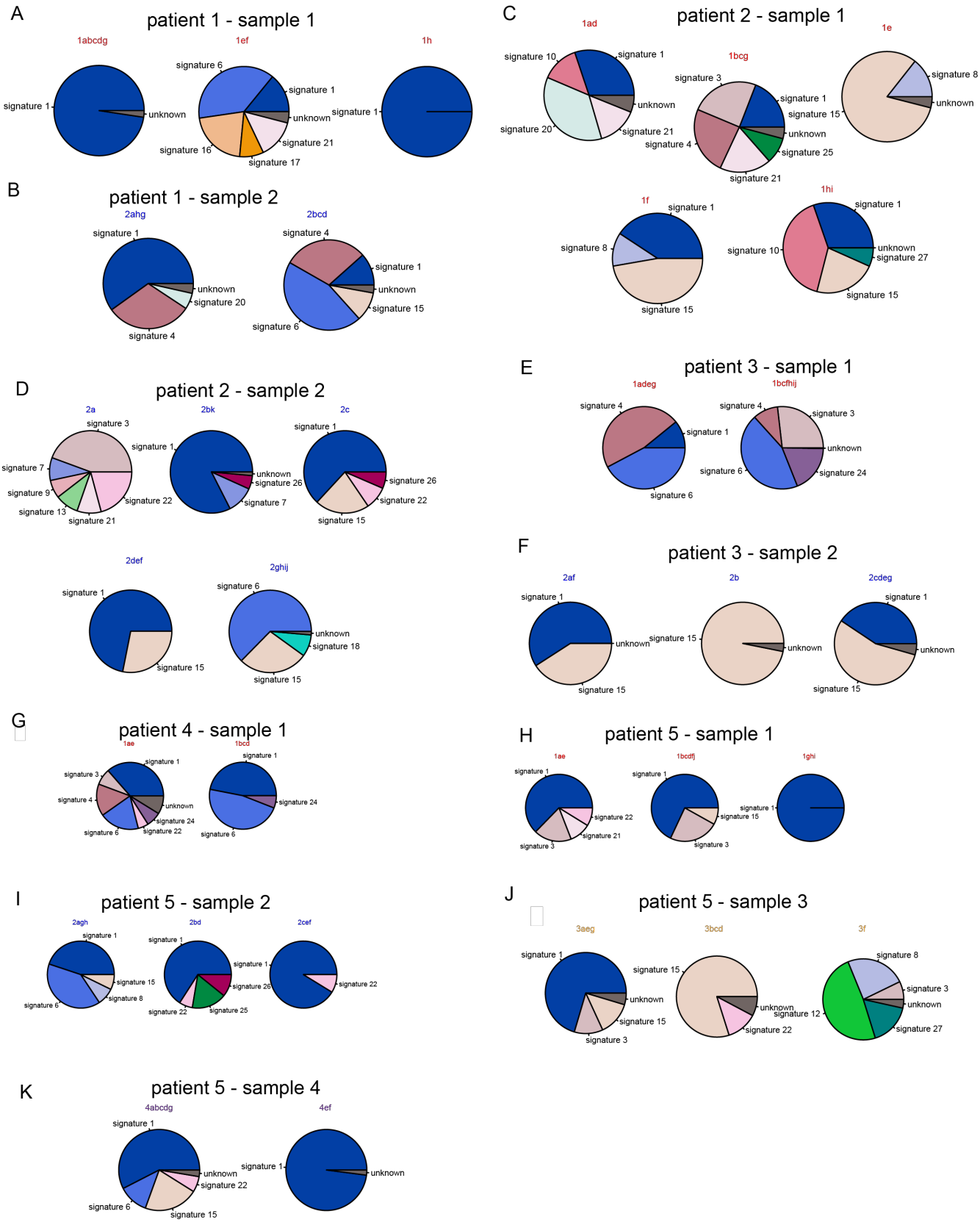


Fig. S2. Mutational signatures in pGBM samples.
(A) Mutational signatures in the diagnostic sample of patient 1.
(B) Mutational signatures in the second resection of patient 1.
(C) Mutational signatures in the primary tumor of patient 2.
(D) Mutational signatures in the recurrence of patient 2.

(E) Mutational signatures in the primary tumor of patient 3.
(F) Mutational signatures in the recurrence of patient 3.
(G) Mutational signatures in the primary tumor of patient 4.
(H) Mutational signatures in the primary tumor of patient 5.
(I) Mutational signatures in the first recurrence of patient 5.
(J) Mutational signatures in the second recurrence of patient 5.
(K) Mutational signatures in the third recurrence of patient 5.
All signatures were derived from SNVs for each tumor. Each pie chart applies to a specific genetic subclone or a group of sub clones, as indicated by the colored labels on top of each diagram.

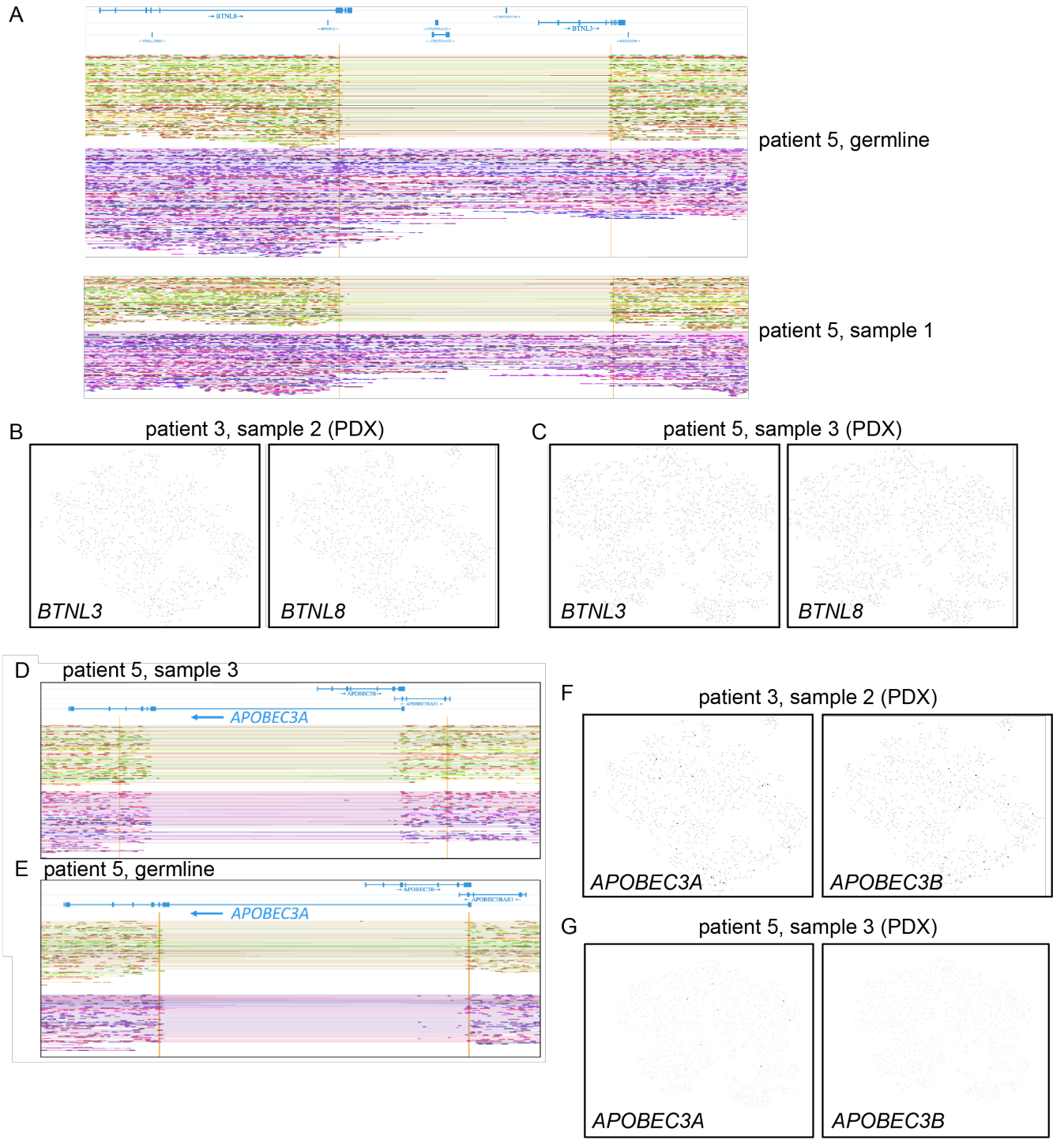


Fig. S3. Deletions at immune-modulatory loci and their effect on gene expression.
(A) Linked read WGS data displaying a deletion at the *BTNL8/3* loci in the germline and in the primary tumor of patient 5.
(B) scRNA-seq for a PDX derived from the recurrence of patient 3. Expression levels for *BTNL3* and *BTNL8* are shown. Grey indicates lack of expression.
(C) scRNA-seq for a PDX derived from the third recurrence of patient 5. Expression levels for *BTNL3* and *BTNL8* are shown. Grey indicates lack of expression.
(D) Linked read WGS data displaying a deletion at the *APOBEC3A/B* locus in the second recurrence of patient 5.
(E) Linked read WGS data displaying a deletion at the *APOBEC3A/B* locus in the germline of patient 5.

(F) scRNA-seq for a PDX derived from the recurrence of patient 3. Expression levels for *APOBEC3A* and *APOBEC3B* are shown. Grey indicates lack of expression.
(G) scRNA-seq for a PDX derived from the second recurrence of patient 5. Expression levels for *APOBEC3A* and *APOBEC3B* are shown. Grey indicates lack of expression.

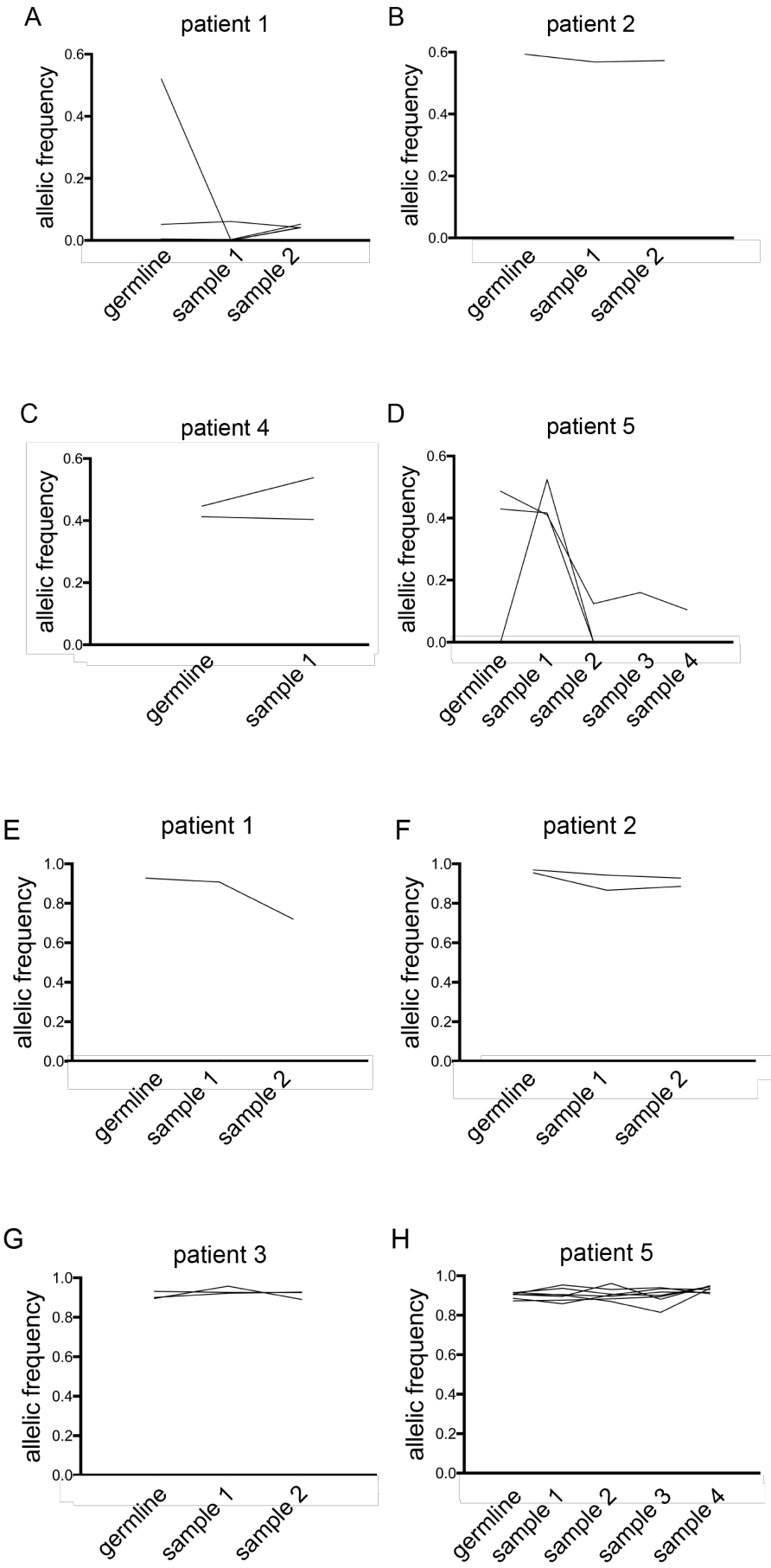

A
patient 1
allelic frequency
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
B
patient 2
allelic frequency
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
C
patient 4
allelic frequency
0.6
0.4
0.2
0.0
germline
sample 1
D
patient 5
allellic frequency
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
sample 3
sample 4
E
patient 1
allelic frequency
1.0
0.8
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
F
patient 2
allelic frequency
1.0
0.8
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
G
patient 3
allelic frequency
1.0
0.8
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
H
patient 5
allelic frequency
1.0
0.8
0.6
0.4
0.2
0.0
germline
sample 1
sample 2
sample 3
sample 4

Fig. S4. Dynamics of germlines SVs during tumor evolution.
(A) Allelic frequencies for large SVs that have a germline allelic frequency > 0.4 and < 0.7 in the germline of patient 1.
(B) Allelic frequencies for large SVs that have a germline allelic frequency > 0.4 and < 0.7 in the germline of patient 2
(C) Allelic frequencies for large SVs that have a germline allelic frequency > 0.4 and < 0.7 in the germline of patient 4.
(D) Allelic frequencies for large SVs that have a germline allelic frequency > 0.4 and < 0.7 in the germline of patient 5.
(E) Allelic frequencies for large SVs that have a germline allelic frequency > 0.7 in the germline of patient 1.
(F) Allelic frequencies for large SVs that have a germline allelic frequency > 0.7 in the germline of patient 2
(G) Allelic frequencies for large SVs that have a germline allelic frequency > 0.7 in the germline of patient 3.
(H) Allelic frequencies for large SVs that have a germline allelic frequency > 0.7 in the germline of patient 5.
The germline of patient 3 had no large SVs with germline allelic frequencies between 0.4 and 0.7.
The germline of patient 4 had no large SVs with germline allelic frequencies > 0.7.
All SVs in these figures were called with the longranger package.

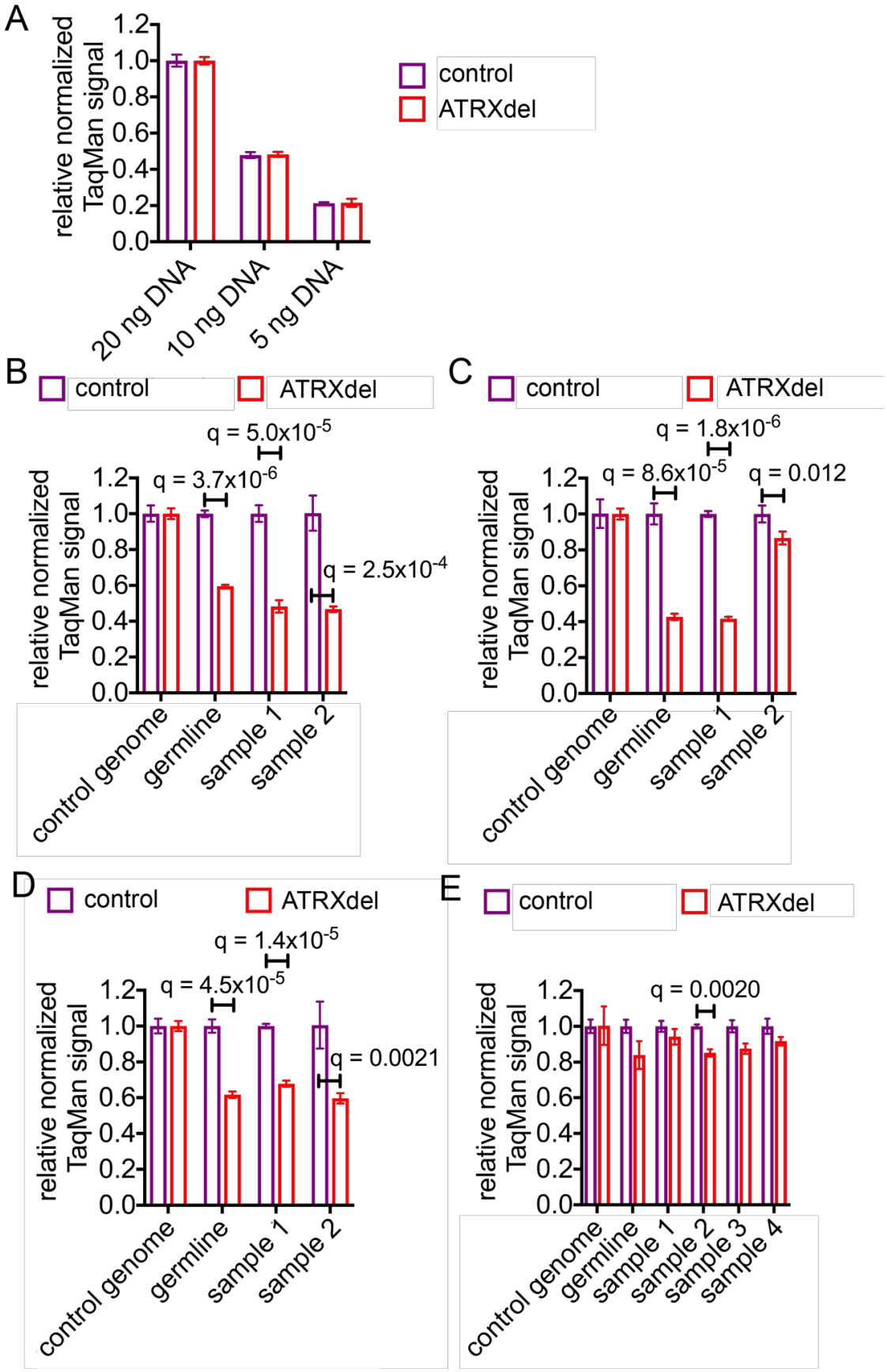


Fig. S5. Copy number assays using TaqMan probes confirm the deletion at the *ATRX* locus.

(A) TaqMan probes were assessed for their sensitivity with decreasing amount of genomic DNA as a template. All probes responded comparably to changes in amounts of DNA.
(B) Copy number TaqMan assay for patient 1.
(C) Copy number TaqMan assay for patient 2.
(D) Copy number TaqMan assay for patient 3.
(E) Copy number TaqMan assay for patient 5.
Bars represent average ± SD. For all assays, multiple t test statistics was used to derive q values.

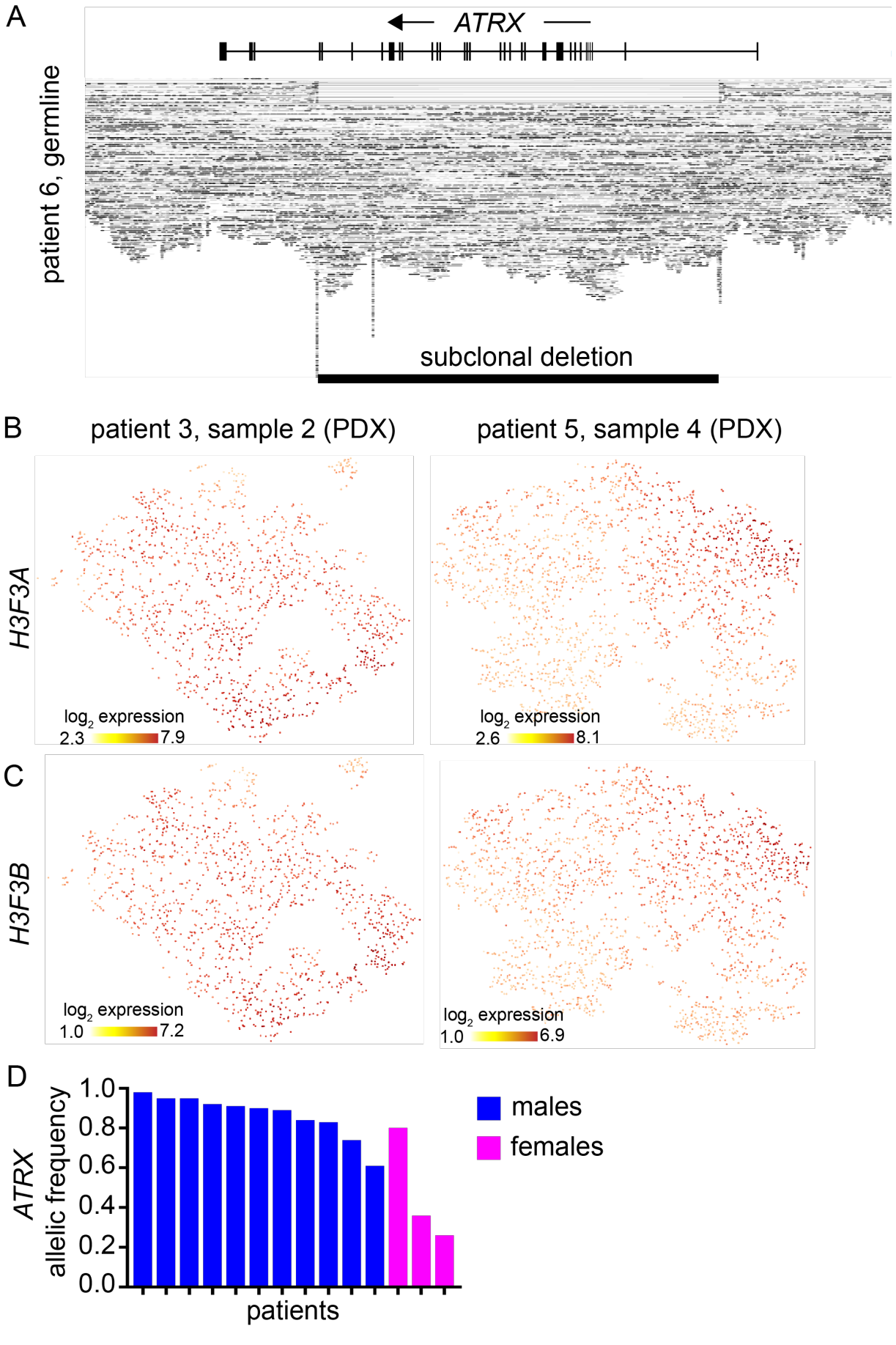
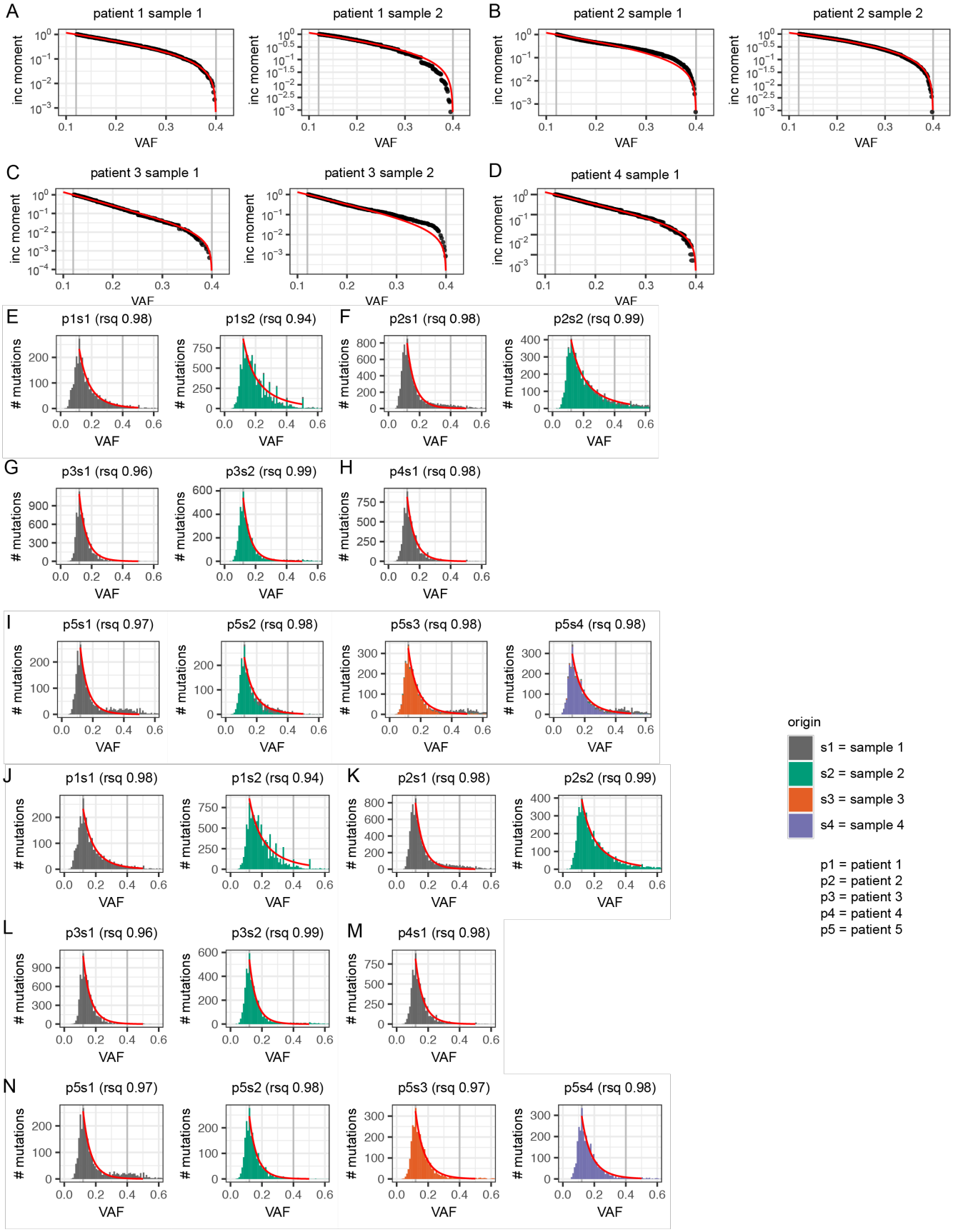

A
ATRX
patient 6, germline
subclonal deletion
B
patient 3, sample 2 (PDX)
patient 5, sample 4 (PDX)
H3F3A
log2 expression
2.3
7.9
log2 expression
2.6
8.1
C
H3F3B
log2 expression
1.0
7.2
log2 expression
1.0
6.9
D
ATRX
allelic frequency
1.0
0.8
0.6
0.4
0.2
0.0
patients
males
females

Fig. S6. Genetic heterogeneity at the *ATRX* locus.
(A) An example of putative deletion at the *ATRX* locus detected in the germline of patient 6 (Toronto cohort).
(B) Expression levels for *H3F3A* in a xenograft derived from the recurrence of patient 3 (left panel) and in a xenograft derived from recurrence 3 of patient 5 (right panel).
(C) Expression levels for *H3F3B* in a xenograft derived from the recurrence of patient 3 (left panel) and in a xenograft derived from recurrence 3 of patient 5 (right panel).
(D) Allelic frequency for *ATRX* mutations in pediatric GBM patients in the DKFZ pediatric pan-cancer study (Gröbner et al., 2018).


A
patient 1 sample 1
patient 1 sample 2
B
patient 2 sample 1
patient 2 sample 2
C
patient 3 sample 1
patient 3 sample 2
D
patient 4 sample 1
inc moment
VAF
E
p1s1 (rsq 0.98)
p1s2 (rsq 0.94)
F
p2s1 (rsq 0.98)
p2s2 (rsq 0.99)
G
p3s1 (rsq 0.96)
p3s2 (rsq 0.99)
H
p4s1 (rsq 0.98)
mutations
I
p5s1 (rsq 0.97)
p5s2 (rsq 0.98)
p5s3 (rsq 0.98)
p5s4 (rsq 0.98)
J
K
L
M
N
p5s3 (rsq 0.97)
origin
s1 = sample 1
s2 = sample 2
s3 = sample 3
s4 = sample 4
p1 = patient 1
p2 = patient 2
p3 = patient 3
p4 = patient 4
p5 = patient 5

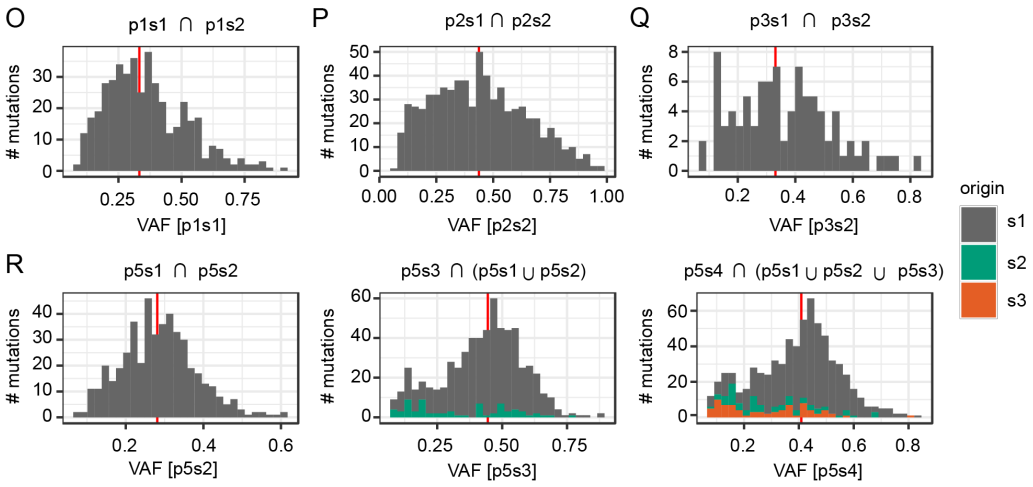


Fig. S7. Distribution of allelic frequencies and clonal sizes in pGBM samples.
(A-D) Fit of the first incomplete moment of the VAF distribution based on a neutral hierarchical model of tumor cell dynamics (red line) to data obtained from patients 1-4 (black points). The grey lines indicate allele frequency cut-offs due to sequencing resolution (lower) and ploidy (upper).
(E-H) Fit of the raw VAF distribution based on a neutral hierarchical model of tumor cell dynamics (red line) to data obtained from patients 1-5 (colored bars). The grey lines indicate allele frequency cut-offs due to sequencing resolution (lower) and ploidy (upper). The filling of the histogram indicates the origin of the mutations (i.e. first detection). Y axis represents the number of mutations that occurred at a specified VAF.
(I-N) Fit of the raw VAF distribution based on a neutral hierarchical model of tumor cell dynamics (red line) to data obtained from patients 1-5 (colored bars). In this case, the data for recurrences has been filtered to include only *de novo* mutations defined as those that reach detection threshold only during recurrence. The grey lines indicate allele frequency cut-offs due to sequencing resolution (lower) and ploidy (upper).
(O-R) Histograms showing VAF distributions for mutations that are shared between recurrent samples as indicated. The red line indicates the median of the distribution.
For details on the model and fits, see main text and Methods.

patient #	sample 1 (diagnostic sample)	sample 2	sample 3	sample 4	blood (germline)
1	✓*	✓*			✓
2	✓	✓			✓
3	✓	✓			✓
4	✓				✓
5	✓	✓	✓	✓	✓

Supplemental Table S1. Calgary pediatric GBM cohort. Red asterisk indicates that for patient 1, two resections were performed with no treatment in between. The sample obtained at the first resection is referred to as "diagnostic sample," and the sample obtained at the second surgery is referred to as "second resection."

patient #	sex	age (years)	radiation	chemotherapy	OS (days)
1	M	12	none	none	> 1434
2	F	9	local	temozolomide	281
3	F	< 1	none	carboplatin, etoposide	325
4	M	< 1	none	none	16
5	M	15	cerebrospinal, boost	cisplatin, vincristine, cyclophosphamide	192

Supplemental Table S2. Therapeutic approach taken for each pediatric GBM patient in the Calgary cohort. We have treatment and outcome information for each patient. OS: Overall survival.

patient #	sex	age (years)	used for
6	M	14	WGS
7	F	14	WGS
8	F	11	WGS
G477	F	15	RNA-seq
G626r	F	14	RNA-seq
G752r	M	1	RNA-seq

Supplemental Table S3. Toronto pediatric GBM cohort used for WGS with linked reads and bulk RNA-seq. For patient 6, 7 and 8, germline and primary tumor tissue were available. For G477, G626r and G752r, RNA-seq was performed on bulk tumor and on their primary cultures.

sample	longest phase block (Mbp)	large SVs (> 30 kbp)	short deletions (50-30,000 bp)	Average sequencing coverage (X)	Average physical coverage (X)
patient 1 (Calgary) – sample 1	9	49	4239	36.3	181.5
patient 1 (Calgary) - sample 2	1	26	6838	30.5	152.5
patient 1 (Calgary) - germline	11	20	4153	32.9	164.5
patient 2 (Calgary) - sample 1	12	23	4596	35.9	179.5
patient 2 (Calgary) - sample 2	12	25	4849	37.9	189.5
patient 2 (Calgary) - germline	21	19	4165	33.8	169.0
patient 3 (Calgary) - sample 1	4	15	4481	32.0	160.0
patient 3 (Calgary) - sample 2	9	20	4367	34.1	170.5
patient 3 (Calgary) - germline	12	20	4141	33.7	168.5
patient 4 (Calgary) - sample 1	12	35	4086	33.8	169.0
patient 4 (Calgary) - germline	24	31	3977	34.4	172.0
patient 5 (Calgary) - sample 1	9	30	3953	36.8	184.0
patient 5 (Calgary) - sample 2	9	23	3945	35.3	176.5
patient 5 (Calgary) - sample 3	11	25	3876	34.8	174.0
patient 5 (Calgary) - sample 4	10	20	3877	36.9	184.5
patient 5 (Calgary) - germline	18	27	4116	33.7	168.5
patient 6 (Toronto) - primary	71	20	3755	38.5	192.5
patient 6 (Toronto) - germline	9	23	4114	37.1	185.5
patient 7 (Toronto) - primary	15	58	4055	38.9	194.5
patient 7 (Toronto) - germline	3	28	4376	36.3	181.5
patient 8 (Toronto) - primary	15	83	4566	39.1	195.5
patient 8 (Toronto) - germline	10	22	4191	36.4	182
patient 1 - parent 1	24	14	3884	35.1	175.5
patient 1 - parent 2	26	14	3762	36.2	181
patient 3 - parent 1	38	15	3897	36.6	183
patient 3 - parent 2	37	25	3705	38.4	192

Supplemental Table S4. Features of each sample analyzed by whole-genome sequencing with linked reads.

genes	number of pGBMs with mutations
EBF4	2
FAM160A2	2
RNF169	2
VWF	2

Supplemental Table S5. Genes mutated in diagnostic pGBM samples in more than one patient.

genes	number of pGBMs with mutations
ASXL1	2
HTR1E	2
NBEAL1	2
POC5	2
PRUNE2	2
SZT2	2

Supplemental Table S6. Genes mutated in follow-up resections in more than one patient.

Sample	f_0 (inc moment)	R-squared (inc moment)	f_0 (neg bin dist)	N_0 (neg bin dist)	R-squared (neg bin dist)
patient 1, sample 1 (p1s1)	0.1597	0.9997	0.1363	0.0148	0.9786
patient 1, sample 2 (p1s2)	0.1963	0.9977	0.2808	0.0063	0.9376
patient 2, sample 1 (p2s1)	0.1316	0.9926	0.0776	0.0022	0.9785
patient 2, sample 2 (p2s2)	0.2709	0.9997	0.2834	0.0137	0.9869
patient 3, sample 1 (p3s1)	0.0608	0.9978	0.0748	0.0015	0.9630
patient 3, sample 2 (p3s2)	0.0731	0.9983	0.0701	0.0028	0.9856
patient 4, sample 1 (p4s1)	0.0737	0.9993	0.0776	0.0022	0.9796
patient 5, sample 1 (p5s1)	0.1472	0.9883	0.0779	0.0070	0.9710
patient 5, sample 2 (p5s2)	0.1529	0.9990	0.1183	0.0130	0.9767
patient 5, sample 3 (p5s3)	0.1132	0.9982	0.1077	0.0084	0.9764
patient 5, sample 4 (p5s4)	0.1401	0.9994	0.1383	0.0117	0.9823
p1s1 (de novo)	0.1597	0.9997	0.1363	0.0148	0.9786
p1s2 (de novo)	0.1815	0.9974	0.2622	0.0061	0.9360
p2s1 (de novo)	0.1316	0.9925	0.0776	0.0022	0.9785
p2s2 (de novo)	0.2223	0.9995	0.2432	0.0129	0.9863
p3s1 (de novo)	0.0608	0.9978	0.0748	0.0015	0.9630
p3s2 (de novo)	0.0696	0.9986	0.0691	0.0027	0.9859
p4s1 (de novo)	0.0737	0.9993	0.0776	0.0022	0.9796
p5s1 (de novo)	0.1472	0.9883	0.0779	0.0070	0.9710
p5s2 (de novo)	0.0738	0.9994	0.0768	0.0071	0.9790
p5s3 (de novo)	0.0811	0.9984	0.0966	0.0075	0.9731
p5s4 (de novo)	0.0965	0.9979	0.1156	0.0099	0.9802

Supplemental Table S7. **Results of the fits of the first incomplete moment and the negative binomial like distribution to the samples.** For details on the models and fits, see main text and Supplemental Materials and Methods.