Supplemental Material

Molecular heterogeneity alongside high expression and mutations of *CXorf67* in posterior fossa type A (PFA) ependymoma

Kristian W. Pajtler, Ji Wen, Martin Sill, Tong Lin, Wilda Orisme, Bo Tang, Jens-Martin Hübner, Vijay Ramaswamy, Sujuan Jia, James D. Dalton, Kelly Haupfear, Hazel A. Rogers, Chandanamali Punchihewa, Ryan Lee, John Easton, Gang Wu, Timothy A. Ritzmann, Rebecca Chapman, Lukas Chavez, Fredrick A. Boop, Paul Klimo Jr., Noah D. Sabin, Robert Ogg, Stephen C. Mack, Brian D. Freibaum, Hong Joo Kim, Hendrik Witt, David T.W. Jones, Baohan Vo, Amar Gajjar, Arzu Onar-Thomas, Martine F. Roussel, Jinghui Zhang, J. Paul Taylor, Thomas E. Merchant, Richard Grundy, Ruth G. Tatevossian, Michael D. Taylor, Stefan M. Pfister, Andrey Korshunov, Marcel Kool, David W. Ellison

Supplemental material contents

Experimental Procedures

Tumor samples and patient characteristics

Clinical samples and data were assembled in accordance with local institutional review board approval at all principal institutions involved in this study: St. Jude Children's Research Hospital, Heidelberg University Hospital (German Cancer Research Center; DKFZ), including approved samples from the NN Burdenko Neurosurgical Institute (Moscow), The Hospital for Sick Children (Toronto), and the Children's Brain Tumour Research Centre (University of Nottingham). Inclusion criteria were an institutional histopathologic diagnosis of WHO grade II or grade III ependymoma, primary tumor located within the posterior fossa, and PFA molecular group, following confirmatory analysis (see below).

Of the 675 samples in the study series, frozen tissue that had been stored at -80^{o}C was available for DNA extraction in 92 cases, while DNA was extracted from formalin-fixed and paraffin embedded (FFPE) tissue for remaining samples (n=583). DNA was extracted according to standard procedures using proteinase K digestion, phenol:chloroform:isoamyl extraction and ethanol precipitation for frozen samples, and using the Qiagen DNeasy FFPE extraction kit (Qiagen, Hilden, Germany), or Maxwell® 16 Plus LEV DNA purification kit (Promega, Madison WI) for FFPE samples. All DNA samples were quantified using Picogreen (Invitrogen, Carlsbad, CA). Patient cohorts from different centers were comparable (Table S1); age at diagnosis and gender ratio were very similar, but aspects of the therapeutic approaches differed.

DNA methylation array processing

Illumina Infinium HumanMethylation450 (450K) arrays were used to obtain genome-wide DNA methylation profiles for tumor samples, according to manufacturer's instructions (Illumina, San Diego, CA). Data were generated at the Genomics and Proteomics Core Facility of the DKFZ, the Hartwell Center at St. Jude Children's Research Hospital, and the Princess Margaret Genomics Centre in Toronto, Canada, as previously described [15,18,23]. For fresh-frozen tumor samples, >500ng of DNA was used as input material, while 250-500ng of DNA was used for FFPE samples. On-chip quality metrics of all samples were carefully controlled. Samples were also checked for unexpected genotype matches by pairwise comparison of the 65 genotyping probes included on the 450K array. Copy number variation (CNV) analysis from 450K methylation array data was performed using the Conumee package, version 1.9.0 [8]. Two sets of 50 control samples displaying a balanced copy-number profile from both male and female donors were used.

All computational analyses were performed in R (version 3.3.1). Raw signal intensities were obtained from IDAT-files using the minfi Bioconductor package version 1.21.4 [2]. Each sample was individually normalized by performing a background correction (shifting of the 5% percentile of negative control probe intensities to 0) and a dye-bias correction (scaling of the mean of normalization control probe intensities to 10,000) for both color channels.

Subsequently, a correction for the type of material tissue (FFPE / frozen) was performed by fitting univariate, linear models to the log_2-transformed intensity values (removeBatchEffect function, limma package version 3.30.11 [21]). The methylated and unmethylated signals were corrected individually. Beta-values were calculated from the retransformed intensities using an offset of 100 (as recommended by Illumina).

Before unsupervised clustering analysis, the following filtering criteria were applied: removal of probes targeting the X and Y chromosomes (n=11,551), removal of probes containing a single-nucleotide polymorphism (dbSNP132 Common) within five base pairs of and including the targeted CpG-site (n=7,998), and probes not mapping uniquely to the human reference genome (hg19) allowing for one mismatch (n=3,965). The complete methylation values for the series of tumors have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), under accession number GSE104210.

Methylation data (IDAT files) for all 675 tumors were uploaded into the DKFZ brain tumor DNA methylation-based molecular classifier (www.molecularneuropathology.org), to confirm that all ependymomas were assigned to the PFA molecular group (Capper et al., 2018). This website enables tumor data to be compared with data from a reference cohort comprising 2682 tumors, including almost all known neuropathological entities. The site generates an automated report, assigning each tumor to a defined 'DNA Methylation Class' where the calibrated classifier score is greater or equal to 0.9, or 'no match to a defined DNA Methylation Class' where the score is less than 0.9. All tumors included in this study were assigned to the "ependymoma, posterior fossa group A" Methylation Class.

Unsupervised clustering analysis

The 5000 most variable methylation probes were selected by standard deviation. Based on these probes a Consensus Clustering was performed using the ConsensusClusterPlus package version 1.38 [27]. The following non-default parameters were used: maxK=20, rep=1000, distance="euclidean", innerLinkage="ward.D2". Elbow criteria were used to estimate the optimal number of clusters. A cluster validation index is calculated for varying number of clusters and how this index changes is observed. The optimal number of clusters is found at a point where the index does not change significantly, i.e. the so called 'elbow' in a plot that shows the index for increasing numbers of clusters. To explore how well clustering results correlate with OS and PFS, Cox proportional hazard models were fitted for all numbers of clusters resulting from the consensus clustering and the risk prediction performance of these models were measured by the Concordance-statistic. The results based on these criteria in Figure S3B indicate that nine clusters are most stable and clinically meaningful and increasing the number of clusters further will not significantly improve cluster stability and risk predication. Principal component scores were calculated by applying singular value decomposition (svds function of the RSpectra package version 0.12-0) to the mean centered beta values of the 5000 most variable probes and multiplying the resulting left singular vectors with diagonal matrix that includes the singular values. The first 50 principal component scores were used for t-SNE (t-Distributed Stochastic Neighbor Embedding,

Rtsne package version 0.11). The following non-default parameters were used: theta=0, pca=F, max_iter=2500 perplexity=20.

Gene expression profiling

Tumor samples from the DKFZ and St. Jude (SJ), for which RNA of sufficient quantity and quality was available, were analyzed on the Affymetrix GeneChip human Genome U133 Plus 2.0 (U133v2) array (Affymetrix, Santa Clara, CA) at the Microarray Department of the University of Amsterdam, the Netherlands, and the Hartwell Center at St. Jude Children's Research Hospital. The DKFZ data have been previously reported and are deposited within the NCBI's Gene Expression Omnibus (GEO) under dataset accession number GSE64415. SJ data have been deposited at the NCBI GEO with data accession number GSE100240.

U133v2 data were available for 79 samples covering all methylation clusters (PFA-1a n=7, PFA-1b n=12, PFA-1c n=9, PFA-1d n=7, PFA-1e n=12, PFA-1f n=2, PFA-2a n=9, PFA-2b n=16, PFA-2c n=5). Gene expression levels were processed and normalized by Robust Multi-Chip Averaging (RMA) using the Affymetrix Expression Console software. The combined dataset of SJ and DKFZ samples were processed using a batch effect adjustment (removeBatchEffect in limma). Probe sets from sex chromosomes were removed. For unsupervised clustering, hierarchical clustering was performed using Euclidean distance and Ward's method. ConsensusClusterPlus, with 100 resamplings, was used to estimate the consensus matrix and to evaluate cluster stability [27]. For supervised clustering, probe sets mapping to the same gene symbol were aggregated by taking the mean. In total, 22,056 genes for which gene symbol annotations were available were kept for the analysis. Differential expression analysis was performed by a linear model and empirical bayes approach (limma, [21]), testing each methylation group against all others. The 20 genes with smallest p-values in each methylation cluster may be visualized in a supervised heatmap representation (Figure S3A). Gene expression values have been mean centered and rescaled by dividing by the standard deviation before visualization.

Polyadenylated (poly-A) RNA transcriptome sequencing (RNA-seq) was undertaken for 28 tumor samples at the Hartwell Center, St. Jude Children's Research Hospital, following previously described methods [32,16]. These specific data have been reported previously [16], and they are available at the European Bioinformatics Institute (EBI) under dataset accession number EGAS00001000254. The EBI ID specific to RNA-seq data is EGAD00001000854. RNA-seq gene expression levels were profiled using HTSeq, and scale normalization was performed using the TMM method in edgeR [1]. Clustering was performed using Euclidean distance and Ward's method. ConsensusClusterPlus was then used to evaluate cluster stability [27]. Principal component analysis (PCA) was also applied to visualize differences between the two subgroups. Statistical analysis of gene expression profile differences between the PFA-1 and PFA-2 subgroups was undertaken using the 'limma voom' method [9]. The top 58 most differentially expressed genes are listed in Table S2.

Gene Set Enrichment Analysis (GSEA)

GSEA analysis was used to identify systematic differences in gene expression between different subgroups [24]. The combined u133v2 dataset of SJ cohort and DKFZ cohort were first processed using a batch effect adjustment (removeBatchEffect in limma). Each subgroup was then compared to all other subgroups. Gene ontology and MSigDB canonical pathway collections were included in the analysis using gene-set permutations (1000), with an FDR cutoff of 0.05. To reduce pathway redundancy, enriched gene sets were further grouped and visualized in the Cytoscape software with the Enrichment Map plugin [14]. Distinct gene sets between the molecular subgroups were illustrated in the inner pie chart using different color codes, and the enrichment or depletion was illustrated in the outer circle. The network structure was simplified by manual curation to remove edges with weak similarity between clustered gene sets.

Mutation analyses

Amplicon sequencing

Amplicon sequencing across the *CXorf67* coding region was undertaken for 234 tumor samples. Primer pairs for PCR amplification tiling across the *CXorf67* coding region were designed using Primer 3 (see tables below). The primers were divided into two non-overlapping sets prior to pooling for multiplexed PCR. The PCR reaction was set up using GoTaq (Promega, Madison WI), 4.6ul of primer pool (A or B) and 50ng of DNA. Thermocycling was carried out using the following conditions: 95C for 5min; 30 cycles of [95C for 20sec, 55C for 20sec, 72C for 20sec]; 72C for 2min. Libraries of the multiplexed amplicons were prepared using the Kapa Hyper DNA libarary prep kit (Kapa Biosystems, Wilmington, MA) and Nextflex 96 adapters (Bioo Scientific, Austin, TX), with each sample pool pair ligated using an adapter with the sample index sequence. Sequencing was performed using the Illumina sequencing by synthesis method, which employs a paired-end 150-cycle protocol with an 8bp index read (Illumina, San Diego, CA). The bcl files were converted to fastq format using bcl2fastq. Illumina standard adapters were first removed using cutadapt v1.8.1. Primer sequences were then trimmed off using cutadapt anchored mode. The trimmed reads were mapped using BWA v0.7.12 [10]. SNVs were called and reviewed using Bambino [6]. For tumor samples without matched germline, dbSNP147 and ExAC databases were used to filter out human polymorphic variants.

Primer Pool A					
Primer Name	**Forward Primer Sequence**	**Pool Conc. (µM)**	**Primer Name**	**Reverse Primer Sequence**	**Pool Conc. (µM)**
CXORF67F12	CACCTTCTTGCTCTACCAGTTC	2.5	CXORF67R12	AAGGGCGGTTTCGTTGTT	2.5
CXORF67F14	ATCAAGATCCTGCTGCTTCC	2.5	CXORF67R14	GGCCTCATCGGTGATGAAA	2.5
CXORF67F16	TCATAGCTGCTGTGCTGAC	2.5	CXORF67R16	CTTCCTACGACGGCTGTTC	2.5
CXORF67F18	CGTCTGTTTCCTGAGCCTTT	5	CXORF67R18	CAGAAGCTACCAGGGTGATG	5
CXORF67F20	CATCACCCTGGTAGCTTCTG	2.5	CXORF67R20	CAGATTCAGGGCTTAGACGAG	2.5
CXORF67F22	GATTCTGCGCCAGTCTCTG	2.5	CXORF67R22	CTGATCTTTGCTGGGTGGAA	2.5
CXORF67F24	GTGGGACTGGCTCAGAAAG	7.5	CXORF67R24	GAGGAGGAGTTGGAAGCATAG	7.5
CXORF67F26	AGTAGGTCTCCTGGCCTAAG	7.5	CXORF67R26	GGTGAGCTGTGCTCTCTATTT	7.5

Primer Pool B					
Primer Name	**Forward Primer Sequence**	**Pool Conc. (µM)**	**Primer Name**	**Reverse Primer Sequence**	**Pool Conc. (µM)**
CXORF67F13	GACATGGAGAAGGAGCAGAAG	2.5	CXORF67R13	GAGATGCTTGGCTGGAGAC	2.5
CXORF67F15	AGCCAAGCATCTCCCTCGG	7.5	CXORF67R15	AACTTCGTGAGGACTGCGG	7.5
CXORF67F17	CTGGCCCAGACCAAGAG	2.5	CXORF67R17	CGGAACCCTGGAGAAGATG	2.5
CXORF67F19	CTCCTAAGCCACGCATCTG	2.5	CXORF67R19	ACGAGCAGGGTCTAGATGA	2.5
CXORF67F21	CACCGCGCCACTCATCTA	2.5	CXORF67R21	AGAGCCTGGCGCAGAAT	2.5
CXORF67F23	GTCAGACGCCGGTCATCG	7.5	CXORF67R23	CTTGGGACCTCAGGATCAGG	7.5
CXORF67F25	TCTTCCTTCCCATCCCTCA	2.5	CXORF67R25	GGGCTAGGAGTGGAGATAGAT	2.5
CXORF67F27	TTTATGCTCTGAGCCCTGTC	2.5	CXORF67R27	AGGCAGCCCATTCACAAA	2.5

Sanger sequencing

Samples were screened for mutations in the following histone genes: *H3F3A*, *HIST1H3A*, *HIST1H3B*, and *HIST1H3C*. PCR primers and conditions for the screen were as previously described [28].

CXorf67 analysis

The mutation rate for *CXorf67* was compared between different tumor types and cohorts. These included the entire dataset from the St. Jude Children's Research Hospital – Washington University Pediatric Cancer Genome Project (whole genome sequencing [WGS] data from 788 tumors and 25 childhood cancer cohorts) and an ependymoma dataset of whole exome sequencing (WES)/ WGS from the German Cancer Research Center (DKFZ), including 21 PFA and 18 PFB tumors. WES data for 28 intracranial germ cell tumors were downloaded from the NCBI dbGaP database. Sequencing reads mapped to *CXorf67* were extracted and downloaded from the GDC data portal for 10726 tumors and 37 cohorts in the TCGA (The Cancer Genome Atlas) and TARGET (Therapeutically Applicable Research To Generate Effective Treatments) projects. All SNV/Indel detections were performed using SNPdetector pipeline and manually reviewed as previously described [31,32]. Samples from the same individual were considered as a single case for mutation rate calculation.

Immunohistochemistry

Immunohistochemistry on human tumors used heat-mediated antigen retrieval in Dako pH 9 Antigen Retrieval Solution (Agilent, Santa Clara, CA) and employed antibodies to OTX2 (1H12G8B2; 1: 20,000 dilution, GeneTex, Inc. Irvine, CA), CXorf67 (SAB2107156; 1: 200 dilution, Sigma-Aldrich, St. Louis, MO), mutant histone H3 K27M (ABE419; 1: 600 dilution, EMD Millipore, Millerica, MA) and Tri-Methyl-Histone H3 (Lys27) (C36B11; 1: 200 dilution, Cell Signaling Technology, Danvers, MA).

Immunoblotting

Total protein from uninfected and infected cell lines (HEK293, Daoy, and human neural stem cells) was extracted using M-PER mammalian extraction buffer or a histone extraction kit buffer (Abcam). Western blot analysis was performed using standard techniques. Antibodies employed included: CXorf67 (1:1000, HPA004003, Sigma-Aldrich), ß-Actin (1:1000, 8H10D10, Cell Signaling Technology), SUZ12 (1:1000, 3737, Cell Signaling), EZH2 (1:1000, 612667, BD Biosciences), H3K27me3 (1:1000, 9733S, Cell Signaling Technology) and Acetyl-Histone H3 (Lys27) (1:1000, 8173S, Cell Signaling Technology).

Immunofluorescence

Cultured cells (HEK293 and Daoy) were fixed with 4% PFA and treated using standard immunofluorescence techniques. The cells were incubated with primary antibodies at 4°C overnight (CXorf67, 1:500 and H3K27me3, 1:1000). The following day, cells were washed and incubated with secondary antibodies (anti-rabbit AlexaFluor-647, and anti-mouse AlexaFluor-568) for two hours at room temperature. All cells were subsequently imaged using the Zeiss LSM 780 NLO point-scanning confocal/multiphoton microscope.

Cloning and Lentiviral Production

Human cDNA of *CXorf67* and three *CXorf67* mutants (T73S, I88F and Y184C) were cloned into the pCDH-MSCV-MCS-EF1a-copGFP vector using Clontech In-Fusion HD Cloning system to generate expression constructs. All constructs were verified by sequencing and used to make lentiviruses. To produce lentivirus, 293TN cells were co-transfected with expression constructs and packaging plasmid mix (Lenti-vpak Lentiviral Packaging kit, OriGene), Lentivirus-containing media were collected, centrifuged and filtered. Resulting lentiviruses were concentrated using PEG-It virus precipitation solution (System Biosciences). The viral titer was determined by flow cytometric analysis of neural stem cells transduced with different dilutions of cDNA encoding lentiviruses.

CXorf67 infection studies

NIH approved H9-derived (WA09) Human Neural Stem Cells (NSC) were purchased from Fisher Scientific, and cultured following the manufacturer's instructions. NSC and HEK293 cells were infected with lentivirus expressing wild type and three mutant *CXorf67* constructs.

CXorf67 CRISPR-Cas-9 Gene Editing

CXorf67 CRISPR sgRNA gene editing guides were cloned into the Lenti-virus pCas9-mCherry vector.

CXorf67 Guide 1:	Forward Oligo 5' CACCGTAGGCCAGGAGACCTACTCG 3'
	Reverse Oligo 5' AAACCGAGTAGGTCTCCTGGCCTAC 3'
CXorf67 Guide 2:	Forward Oligo 5'CACCGCTGATGAGAATCCTTCCTGT3'
	Reverse Oligo 5' AAACACAGGAAGGATTCTCATCAGC 3'

The vector was digested with BsmBI and the two RNA guides were inserted into the vector. The ligation reactions were transformed into Stbl3 cells. Positive clones were confirmed using Sanger Sequencing. Large scale DNA preps for sequenced constructs were used to generate lentivirus, as described above. The lentivirus was used to infect Daoy and HEK293 cells. Flow sorted cells or single clone infected cells were cultured for immunofluorescence, immunoblotting and growth rate analysis.

In vitro cell growth rate analysis

Daoy single clone cells with or without CXorf67 knocked-out (CRISPR-Cas-9 gene editing) were analyzed using a cell viability assay. Samples were collected at day 0, day 1, day 3, day 4 and day 6. Cell numbers were evaluated using the luminescence cell viability assay method (CellTiter-Glo kit, Promega) using four replicates for each group. For each replicate, the luminescence (RLU) values were first normalized by day 0 to estimate the growth rate for each day the cells were collected. The relative growth rates were calculated by normalizing each group of samples to the un-infected Daoy control cells.

Co-immunoprecipitation and Mass Spectrometry

Whole cell lysates from Daoy and U2OS cells were prepared for immunoprecipitation (IP) using a PBS-based buffer (1x PBS, 5mM EDTA, 10% glycerol, 0.2% NP-40 (ThermoFisher Scientific Cat #. 28324)). Nuclear extracts of Daoy cells were prepared for co-IP with NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific, Cat. #78835), and protein was quantitated using the BCA protein assay kit (ThermoFisher Scientific, Cat. #2322). Whole cell lysates were prepared for IP using the CXorf67 antibody. Nuclear extracts were divided into four equal parts for immunoprecipitation (IP) with different antibodies. Each IP was incubated overnight with antibody (CXorf67, HPA004003, Sigma Aldrich; SUZ12, 3737, Cell Signaling Technology; EZH2, 5246, Cell Signaling Technology; Normal Rabbit IgG, 2729, Cell Signaling Technology) at 4°C, and then incubated for three hours with protein G Dynabeads (ThermoFisher Scientific, Cat. #10007D). The beads were washed three times with washing buffer, and then resuspended in elution buffer with 4X gel loading buffer and NuPAGE™ sample reducing agent (ThermoFisher Scientific, Cat#NP0004).

Each IP was performed in duplicate for mass spectrometry (MS). Samples were run on 4-12% Bris-Tris protein gel (ThermoFisher Scientific), stained with SimplyBlue™ SafeStain (ThermoFisher Scientific, Cat#LC6060) and submitted to the St. Jude Proteomics Shared Resource. MS analysis was performed using an optimized platform as previously reported [29]. Proteins in the gel band were in-gel digested by trypsin. Resulting peptides were loaded on a nanoscale reverse-phase column, and eluted by a gradient (~30 min). Eluted peptides were detected by an inline LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). Acquired data were searched against the protein database using the Sequest algorithm and filtered to reduce the protein false discovery rate to below 1%. For binary comparison, statistical analysis was also performed based on the G-test. FDR was derived according to biological replicates and null hypothesis [33]. MS results were analyzed by the Significance Analysis of INTeractome (SAINT) approach with CRAPome online tools using default settings [3,13]. The interacting proteins showing SAINT scores >0.9 in all three affinity purification assays with antibodies to CXorf67, SUZ12 and EZH2 were considered to be interacting with the PRC2 complex with high confidence.

In silico analysis of CXorf67

A number of factors were considered when assessing whether the somatic mutation 'hotspot' within *CXorf67* might have a functional effect on the resulting protein. The degree of order was determined on the DisMeta server (http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder/) [7], which shows consensus results of the following protein disorder predictors (DisEMBL [11], DISOPRED2 [26], DISpro (http://www.ics.uci.edu/~baldig/dispro.html), FoldIndex, GlobPlot2 [12], IUPred [4], RONN [30], and VSL2 [17]) and PONDR [Predictors of Natural Disordered Regions (http://pondr.com/)]. The potential protein binding regions in disordered domain were predicted by ANCHOR

[5] (http://anchor.enzim.hu/). The linear net charge per residue (NCPR) of the protein was calculated using the CIDER algorithm (http://pappulab.wustl.edu/CIDER/).

Statistical analysis of clinical data

Progression-free survival (PFS) was defined as the interval between date of diagnosis and date of progression or death. Overall survival (OS) was defined as the interval between date of diagnosis and death. Patients without an event in PFS or OS were censored at the time of last contact. PFS and OS distributions were estimated using the Kaplan-Meier method and compared between two or more groups of patients using the log-rank test. To investigate associations between multiple covariates and progression-free survival or overall survival, Cox proportional hazards regression models were employed. Descriptive statistics were provided for a set of demographic and clinical variables. Pearson chi-squared tests were used to examine the association between methylation subgroups and categorical clinical variables. The Pearson chi-square test was replaced by Fisher's exact test where one or more cells had small counts. The Wilcoxon rank sum test was used for age at diagnosis, which was treated as continuous.

Radiologic and neurosurgical evaluation of diagnostic MRIs and intraoperative macroscopic pathology

Pre-operative MR imaging was evaluated by a pediatric neuroradiologist (N.S.), who was unaware of the molecular and neurosurgical findings. The content of the pre-operative imaging examination was variable, according to protocols at each patient's treatment center. Evaluated patients (n=40) were selected because the scope and quality of available images were adequate to determine tumor location and the severity of hydrocephalus according to established criteria [19,20,22,25]. Tumor location was classified as central (*CEN*) or lateral (*LAT*) based on whether the mass was centered in the fourth ventricle or outside the fourth ventricle, usually in the cerebellopontine angle cistern. Hydrocephalus was graded on a four-point scale: grade 0 – no evidence of hydrocephalus, grade 1 – mild ventricular enlargement, grade 2 – substantial ventricular enlargement, grade 3 – substantial ventricular enlargement with clear transependymal flow of cerebrospinal fluid.

Operative records were reviewed by two pediatric neurosurgeons (F.A.B & P.K.), who were unaware of the molecular and neuroradiological findings. The putative tumor origin was classified as floor of the fourth ventricle (*F*), roof of the fourth ventricle (*R*), or lateral recess/cerebellopontine angle (*CPA*), based on the operating surgeon's description in the neurosurgical record and the reviewing neurosurgeon's judgment of imaging and clinical data. Analysis was performed with R statistical software (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ version 3.0.2).

Supplemental References

1. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-169. doi:10.1093/bioinformatics/btu638

2. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363-1369. doi:10.1093/bioinformatics/btu049

3. Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods 8:70-73. doi:10.1038/nmeth.1541

4. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433-3434. doi:10.1093/bioinformatics/bti541

5. Dosztanyi Z, Meszaros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25:2745-2746. doi:10.1093/bioinformatics/btp518

6. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH (2011) Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics 27:865-866. doi:10.1093/bioinformatics/btr032

7. Huang YJ, Acton TB, Montelione GT (2014) DisMeta: a meta server for construct design and optimization. Methods Mol Biol 1091:3-16. doi:10.1007/978-1-62703-691-7_1

8. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles AK, Pages H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115-121. doi:10.1038/nmeth.3252

9. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15:R29. doi:10.1186/gb-2014-15-2-r29

10. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760. doi:10.1093/bioinformatics/btp324

11. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. Structure 11:1453-1459

12. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31:3701-3708

13. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY, Halim VA, Bagshaw RD, Hubner NC, Al-Hakim A, Bouchard A, Faubert D, Fermin D, Dunham WH, Goudreault M, Lin ZY, Badillo BG, Pawson T, Durocher D, Coulombe B, Aebersold R, Superti-Furga G, Colinge J, Heck AJ, Choi H, Gstaiger M, Mohammed S, Cristea IM, Bennett KL, Washburn MP, Raught B, Ewing RM, Gingras AC, Nesvizhskii AI (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods 10:730-736. doi:10.1038/nmeth.2557

14. Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5:e13984. doi:10.1371/journal.pone.0013984

15. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, Kratochwil F, Wani K, Tatevossian R, Punchihewa C, Johann P, Reimand J, Warnatz HJ, Ryzhova M, Mack S, Ramaswamy V, Capper D, Schweizer L, Sieber L, Wittmann A, Huang Z, van Sluis P, Volckmann R, Koster J, Versteeg R, Fults D, Toledano H, Avigad S, Hoffman LM, Donson AM, Foreman N, Hewer E, Zitterbart K, Gilbert M, Armstrong TS, Gupta N, Allen JC, Karajannis MA, Zagzag D, Hasselblatt M, Kulozik AE, Witt O, Collins VP, von Hoff K, Rutkowski S, Pietsch T, Bader G, Yaspo ML, von Deimling A, Lichter P, Taylor MD, Gilbertson R, Ellison DW, Aldape K, Korshunov A, Kool M, Pfister SM (2015) Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell 27:728-743. doi:10.1016/j.ccell.2015.04.002

16. Parker M, Mohankumar KM, Punchihewa C, Weinlich R, Dalton JD, Li Y, Lee R, Tatevossian RG, Phoenix TN, Thiruvenkatam R, White E, Tang B, Orisme W, Gupta K, Rusch M, Chen X, Li Y, Nagahawhatte P, Hedlund E, Finkelstein D, Wu G, Shurtleff S, Easton J, Boggs K, Yergeau D, Vadodaria B, Mulder HL, Becksfort J, Gupta P, Huether R, Ma J, Song G, Gajjar A, Merchant T, Boop F, Smith AA, Ding L, Lu C, Ochoa K, Zhao D, Fulton RS, Fulton LL, Mardis ER, Wilson RK, Downing JR, Green DR, Zhang J, Ellison DW, Gilbertson RJ (2014) C11orf95-RELA fusions drive oncogenic NF-kappaB signalling in ependymoma. Nature 506:451-455. doi:10.1038/nature13109

17. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 7:208. doi:10.1186/1471-2105-7-208

18. Ramaswamy V, Hielscher T, Mack SC, Lassaletta A, Lin T, Pajtler KW, Jones DT, Luu B, Cavalli FM, Aldape K, Remke M, Mynarek M, Rutkowski S, Gururangan S, McLendon RE, Lipp ES, Dunham C, Hukin J, Eisenstat DD, Fulton D, van Landeghem FK, Santi M, van Veelen MC, Van Meir EG, Osuka S, Fan X, Muraszko KM, Tirapelli DP, Oba-Shinjo SM, Marie SK, Carlotti CG, Lee JY, Nageswara Rao AA, Giannini C, Faria CC, Nunes S, Mora J, Hamilton RL, Hauser P, Jabado N, Petrecca K, Jung S, Massimi L, Zollo M, Cinalli G, Bognar L, Klekner A, Hortobagyi T, Leary S, Ermoian RP, Olson JM, Leonard JR, Gardner C, Grajkowska WA, Chambless LB, Cain J, Eberhart CG, Ahsan S, Massimino M, Giangaspero F, Buttarelli FR, Packer RJ, Emery L, Yong WH, Soto H, Liau LM, Everson R, Grossbach A, Shalaby T, Grotzer M, Karajannis MA, Zagzag D, Wheeler H, von Hoff K, Alonso MM, Tunon T, Schuller U, Zitterbart K, Sterba J, Chan JA, Guzman M, Elbabaa SK, Colman H, Dhall G, Fisher PG, Fouladi M, Gajjar A, Goldman S, Hwang E, Kool M, Ladha H, Vera-Bolanos E, Wani K, Lieberman F, Mikkelsen T, Omuro AM, Pollack IF, Prados M, Robins HI, Soffietti R, Wu J, Metellus P, Tabori U, Bartels U, Bouffet E, Hawkins CE, Rutka JT, Dirks P, Pfister SM, Merchant TE, Gilbert MR, Armstrong TS, Korshunov A, Ellison DW, Taylor MD (2016) Therapeutic Impact of Cytoreductive Surgery and Irradiation of Posterior Fossa Ependymoma in the Molecular Era: A Retrospective Multicohort Analysis. J Clin Oncol 34:2468-2477. doi:10.1200/JCO.2015.65.7825

19. Raybaud C (2016) MR assessment of pediatric hydrocephalus: a road map. Childs Nerv Syst 32:19-41. doi:10.1007/s00381-015-2888-y

20. Raybaud C, Ramaswamy V, Taylor MD, Laughlin S (2015) Posterior fossa tumors in children: developmental anatomy and diagnostic imaging. Childs Nerv Syst 31:1661-1676. doi:10.1007/s00381-015-2834-z

21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. doi:10.1093/nar/gkv007

22. Sabin ND, Merchant TE, Li X, Li Y, Klimo P, Jr., Boop FA, Ellison DW, Ogg RJ (2016) Quantitative imaging analysis of posterior fossa ependymoma location in children. Childs Nerv Syst 32:1441-1447. doi:10.1007/s00381-016-3092-4

23. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DT, Capper D, Sill M, Buchhalter I, Northcott PA, Leis I, Ryzhova M, Koelsche C, Pfaff E, Allen SJ, Balasubramanian G, Worst BC, Pajtler KW, Brabetz S, Johann PD, Sahm F, Reimand J, Mackay A, Carvalho DM, Remke M, Phillips JJ, Perry A, Cowdrey C, Drissi R, Fouladi M, Giangaspero F, Lastowska M, Grajkowska W, Scheurlen W, Pietsch T, Hagel C, Gojo J, Lotsch D, Berger W, Slavc I, Haberler C, Jouvet A, Holm S, Hofer S, Prinz M, Keohane C, Fried I, Mawrin C, Scheie D, Mobley BC, Schniederjan MJ, Santi M, Buccoliero AM, Dahiya S, Kramm CM, von Bueren AO, von Hoff K, Rutkowski S, Herold-Mende C, Fruhwald MC, Milde T, Hasselblatt M, Wesseling P, Rossler J, Schuller U, Ebinger M, Schittenhelm J, Frank S, Grobholz R, Vajtai I, Hans V, Schneppenheim R, Zitterbart K, Collins VP, Aronica E, Varlet P, Puget S, Dufour C, Grill J, Figarella-Branger D, Wolter M, Schuhmann MU, Shalaby T, Grotzer M, van Meter T, Monoranu CM, Felsberg J, Reifenberger G, Snuderl M, Forrester LA, Koster J, Versteeg R, Volckmann R, van Sluis P, Wolf S, Mikkelsen T, Gajjar A, Aldape K, Moore AS, Taylor MD, Jones C, Jabado N, Karajannis MA, Eils R, Schlesner M, Lichter P, von Deimling A, Pfister SM, Ellison DW, Korshunov A, Kool M (2016) New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. Cell 164:1060-1072. doi:10.1016/j.cell.2016.01.015

24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-15550. doi:10.1073/pnas.0506580102

25. U-King-Im J, Taylor MD, Raybaud C (2010) Posterior fossa ependymomas: new radiological classification with surgical correlation. Childs Nerv Syst 26:1765-1772. doi:10.1007/s00381-010-1251-6

26. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337:635-645. doi:10.1016/j.jmb.2004.02.002

27. Wilkerson MD, Hayes DN (2010) ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26:1572-1573. doi:10.1093/bioinformatics/btq170

28. Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, Zhu X, Qu C, Chen X, Zhang J, Easton J, Edmonson M, Ma X, Lu C, Nagahawatte P, Hedlund E, Rusch M, Pounds S, Lin T, Onar-Thomas A, Huether R, Kriwacki R, Parker M, Gupta P, Becksfort J, Wei L, Mulder HL, Boggs K, Vadodaria B, Yergeau D, Russell JC, Ochoa K, Fulton RS, Fulton LL, Jones C, Boop FA, Broniscer A, Wetmore C, Gajjar A, Ding L, Mardis ER, Wilson RK, Taylor MR, Downing JR, Ellison DW, Zhang J, Baker SJ, St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome P (2014) The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat Genet 46:444-450. doi:10.1038/ng.2938

29. Xu P, Duong DM, Peng J (2009) Systematical optimization of reverse-phase chromatography for shotgun proteomics. J Proteome Res 8:3944-3950. doi:10.1021/pr900251d

30. Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21:3369-3376. doi:10.1093/bioinformatics/bti534

31. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, Ulyanov A, Wu G, Wilson M, Wang J, Brennan R, Rusch M, Manning AL, Ma J, Easton J, Shurtleff S, Mullighan C, Pounds S, Mukatira S, Gupta P, Neale G, Zhao D, Lu C, Fulton RS, Fulton LL, Hong X, Dooling DJ, Ochoa K, Naeve C, Dyson NJ, Mardis ER, Bahrami A, Ellison D, Wilson RK, Downing JR, Dyer MA (2012) A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature 481:329-334. doi:10.1038/nature10733

32. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK, Downing JR, Mullighan CG (2012) The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481:157-163. doi:10.1038/nature10725

33. Zhou JY, Afjehi-Sadat L, Asress S, Duong DM, Cudkowicz M, Glass JD, Peng J (2010) Galectin-3 is a candidate biomarker for amyotrophic lateral sclerosis: discovery by a proteomics approach. J Proteome Res 9:5133-5141. doi:10.1021/pr100409r

Supplemental figure legends

Figure S1. Separation of PFA subgroups and subtypes from PFB ependymomas

(A) TSNE plot of DNA methylation array data from 675 PFA and 50 PFB ependymomas. Samples are colored according to their respective consensus cluster affiliation throughout the manuscript and in the key (lower right).

Figure S2. Gene expression profiling and PFA subtype heterogeneity

(A) Analysing the cumulative distribution function (CDF) of consensus matrices related to DNA methylation data in order to determine an optimal number of subtype clusters. The optimal number of clusters is found at a point where the relative change in area under the CDF curve for a specific number of clusters does not change significantly. Increasing the number of clusters beyond nine does not improve the stability of the clustering. (B) Complementary analysis on PFA ependymoma outcome data confirming nine subtypes as optimal. To explore how well subtype clustering results correlate with OS and PFS, Cox proportional hazard models were fitted for all numbers of clusters resulting from the consensus clustering and the risk prediction performance of these models was measured by the Concordance-statistic. The results based on these criteria indicate that nine clusters are most stable and clinically meaningful and that increasing the number of clusters further will not significantly improve cluster stability and risk predication. (C) Heatmap from a supervised clustering analysis demonstrating nine subtypes in data from a subset of PFA ependymomas (n=79) analyzed by Affymetrix u133v2 arrays.

Figure S3. Gene expression profiling and PFA subgroup heterogeneity

(A) Unsupervised clustering analysis of gene expression profiles (Affymetrix u133v2 arrays; n=79) using the top 100 most differentially expressed genes across the dataset. Compared to transcriptome sequencing data from only 28 tumors (Figure 2), expression array data from a larger subset of tumors begin to display the heterogeneity shown by the DNA methylation profiling of 675 tumors. (B) Gene set enrichment analysis (GSEA) demonstrating distinct pathways and biological processes for PFA-1 (red) and PFA-2 (blue) ependymomas (FDR corrected $p < 0.05$). Data for this analysis were generated on Affymetrix u133v2 arrays using RNA from the subset of SJ and DKFZ tumors (n=79) used for the clustering analyses in Figure S2A. Cytoscape and Enrichment Map were used for the visualization of GSEA results by manual curation. Each node represents an enriched gene set; these are grouped and annotated by their similarity. Node size is proportional to the number of genes within each gene set. Green line thickness represents the proportion of shared genes between two nodes.

Figure S4. Radiologic characteristics of PFA-1 and PFA-2 ependymomas

(A) Factor plot displaying distribution of radiological and clinical features evaluated for a subset of 40 St. Jude PFA ependymoma patients. The rows of the upper part of the plot are ordered and sorted according to decreasing difference between PFA-1 (n=23) and PFA-2 subgroups (n=17). (B) Patient age at diagnosis, which did not differ between subgroups. (C-G) Conditional density plots summarizing features of PFA-1 and PFA-2 ependymomas (color scheme – upper left): (C) location – lateral (*LAT*) or central (*CEN*); (D) origin – cerebellopontine angle (*CPA*), floor of fourth

ventricle (*F*), or roof of fourth ventricle (*R*); (E) relapse – local (*LOC*) or distant (*DIST*); (F) hydrocephalus – grade 0/1 or grade 2/3; (G) resection – gross total resection (*GTR*) or subtotal resection (*STR*). (H-M) Magnetic resonance imaging scans at diagnosis. The color bars alongside each image identify radiological and clinical features according to the color scheme in upper left of figure.

Figure S5. Clinical and genetic characteristics of PFA-1 and PFA-2 ependymomas

(A) Clinicopathologic characteristics of PFA-1 and PFA-2 ependymomas; only site of relapse is significantly different. (B) Progression-free survival (PFS) and (C) overall survival (OS) for the two subgroups are practically identical. Differences in the frequencies of chromosome arm copy number gain (D) or loss (E) are evident when PFA-1 (n=460) and PFA-2 (n=209) ependymomas are compared. Significant differences are found for: 1q (P=0.005), 9p (P<0.001), 9q (P<0.001), 10q (P=0.005), 17p (P=0.029), chromosome 2 (P<0.001), chromosome 8 (P=0.014), chromosome 9 (P<0.001), chromosome 11 (P=0.007), chromosome 19 (P<0.001).

Figure S6. Copy number alterations (CNAs) across nine subtypes of PFA ependymoma

Genome-wide CNAs (gain – green, loss – red) detected by DNA methylation profiling across nine subtypes of PFA ependymoma (n=669). Note high-frequency CNAs on chromosomes 1q, 6q, and 10q in PFA-1c tumors, and 22q loss in PFA-1f tumors.

Figure S7. Chromosome 1q gain in PFA ependymomas

(A) Ideogram of chromosome 1 showing relative up-regulation of q arm genes in PFA-1c tumors. The average $\log_2$-fold change of gene expression in PFA-1c ependymomas compared to other PFA tumors across each 1Mb genomic bin is depicted. (B) This GSEA analysis compares PFA-1c and other PFA ependymoma subtypes using MSigDB C1 collection and listing up-regulated cytogenetic band gene sets with FDR P value<0.05. (C) Enrichment plots for the top 3 most significant 1q gene sets are shown. Three gene sets located on 1p are included for comparison. (D) Progression-free survival (PFS) for PFA-1c tumors (red) is significantly worse (P=0.0002) than that for other PFA tumors (black). (E) PFS is significantly worse (P=0.0001) for patients with tumors that harbor 1q gain (orange) than for those with tumors that lack this CNA (green). (F) PFA-1e tumors (pink), with 1q gain in only 4.8% of tumors, are also associated with poor PFS (P=0.0107). (G) Among PFA-1c tumors (orange – 1q gain; green – no 1q gain), 1q status did not influence outcome (P=0.9394).

Figure S8. DNA methylation profiling of diffuse midline gliomas, H3 K27M-mutant and PFA-1f ependymomas

PFA-1f ependymomas enriched for H3 K27M-mutant tumors show no overlap with H3 K27M-mutant diffuse midline gliomas in a clustering analysis.

Figure S1

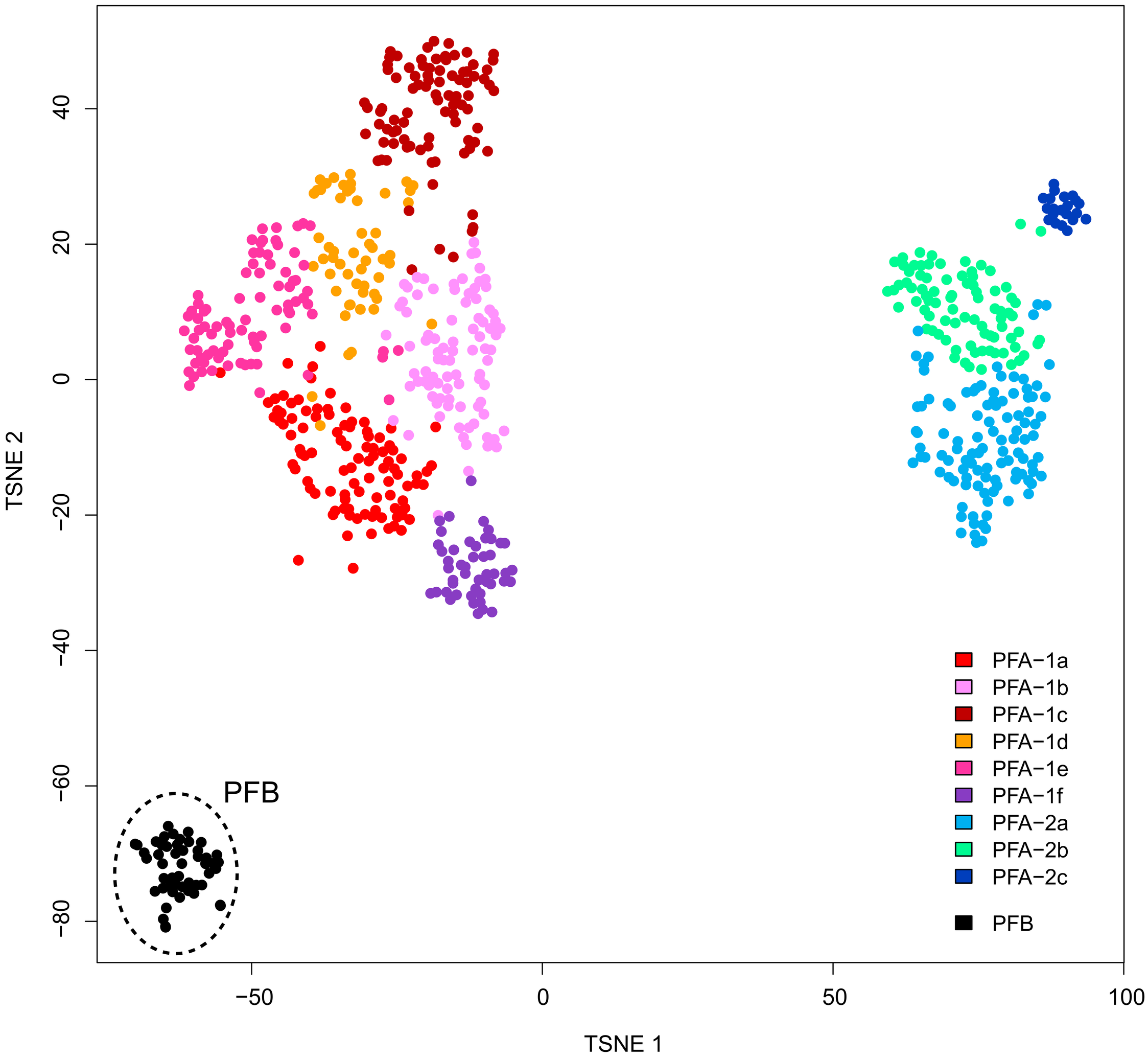
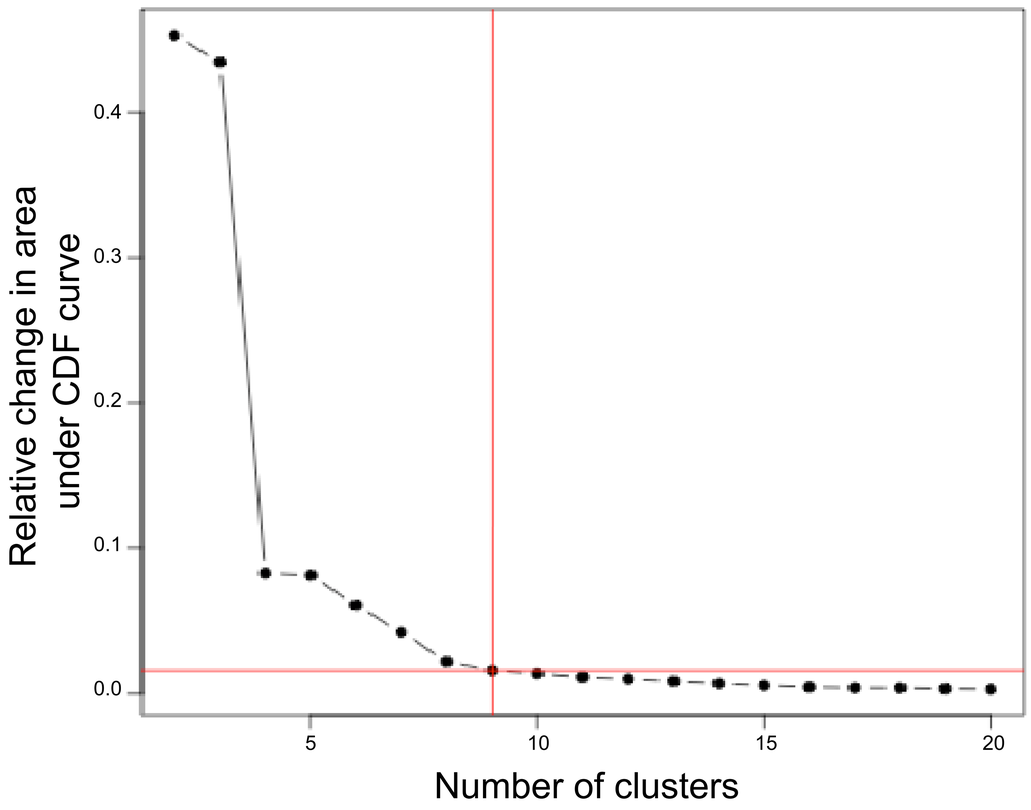
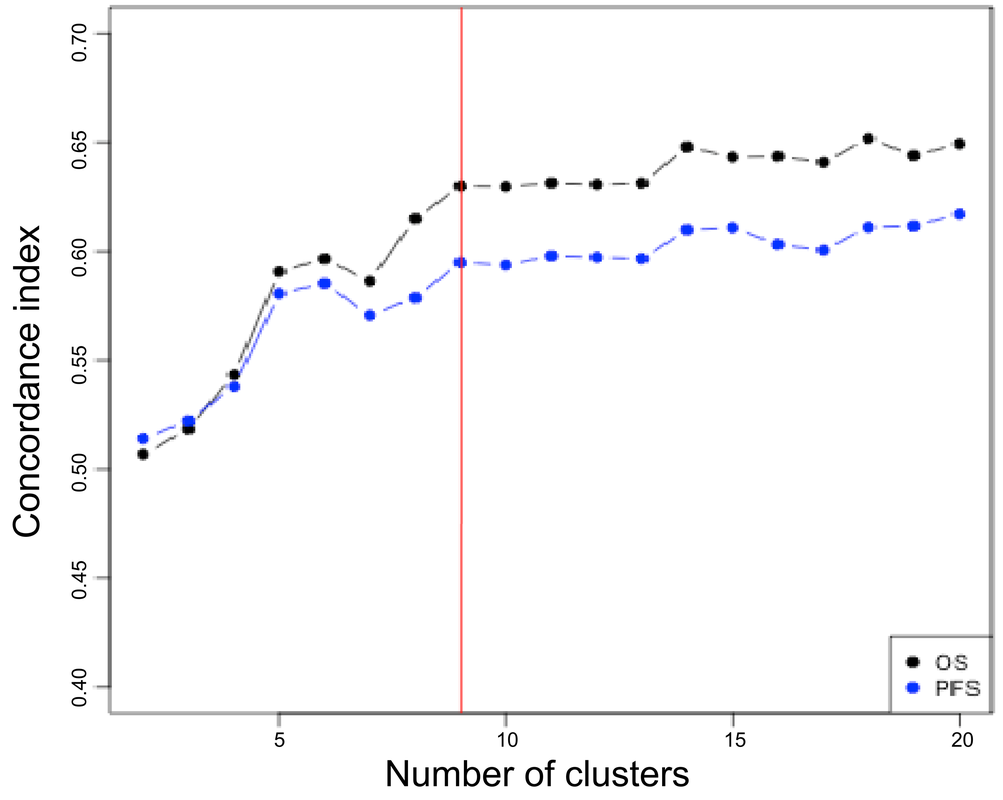

TSNE 2
40
20
0
−20
−40
−60
−80
PFB
PFA−1a
PFA−1b
PFA−1c
PFA−1d
PFA−1e
PFA−1f
PFA−2a
PFA−2b
PFA−2c
PFB
−50
0
50
100
TSNE 1

Figure S2

A



B



C

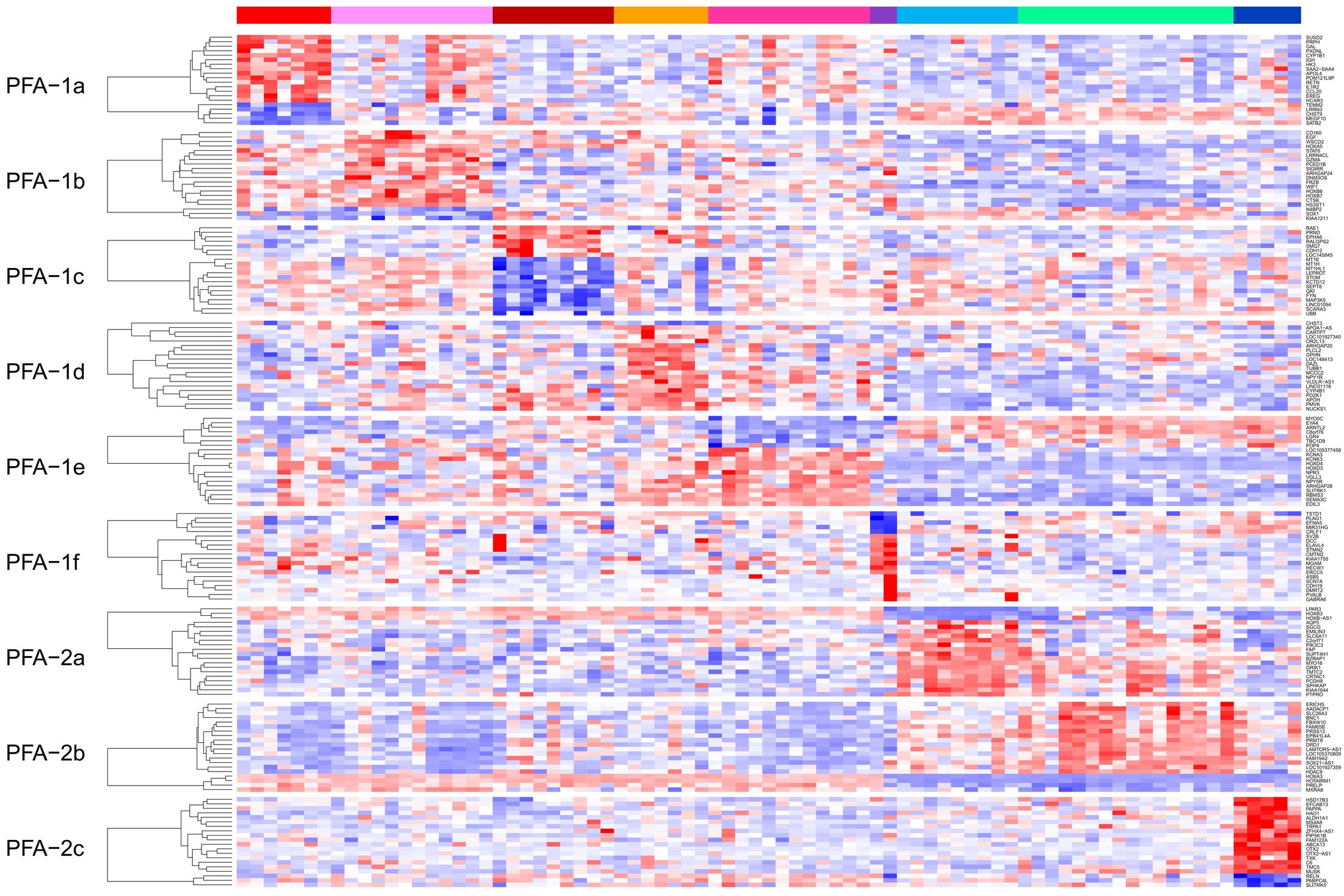
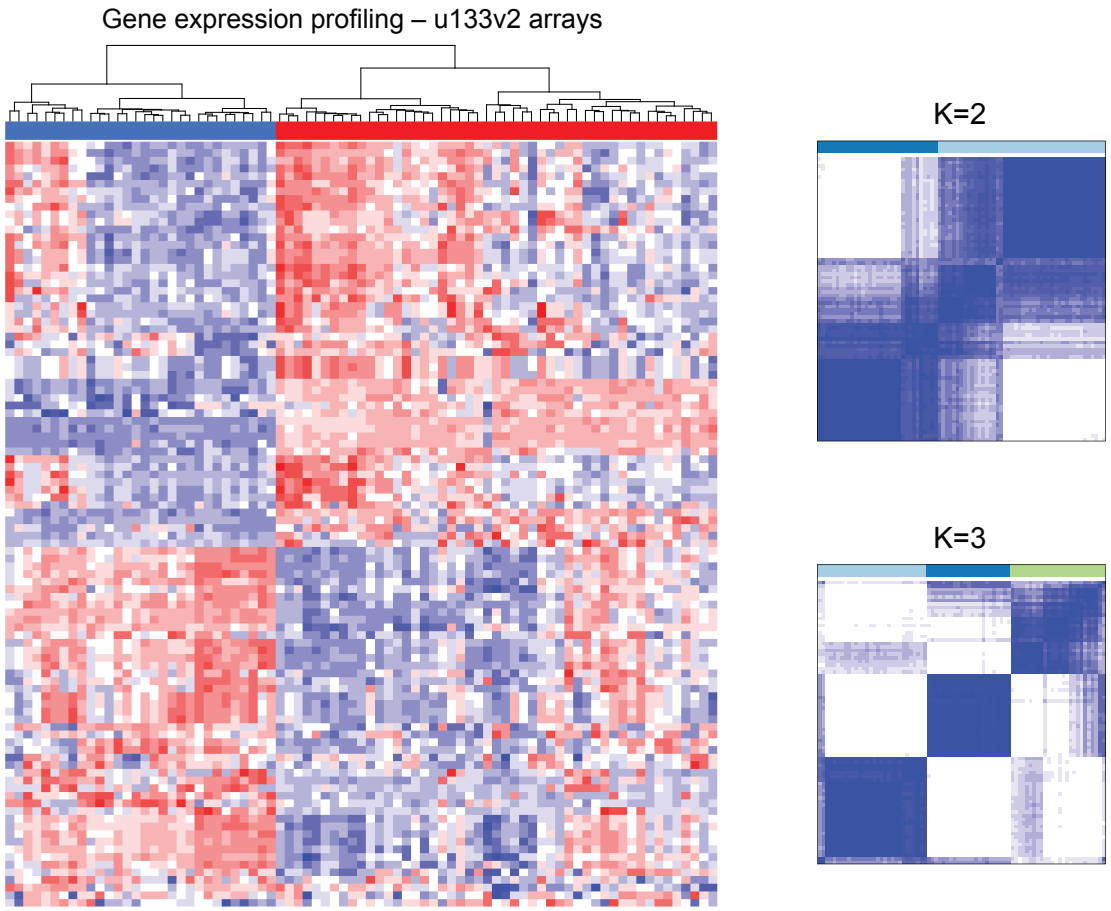


Figure S3

A


Gene expression profiling – u133v2 arrays
K=2
K=3

B

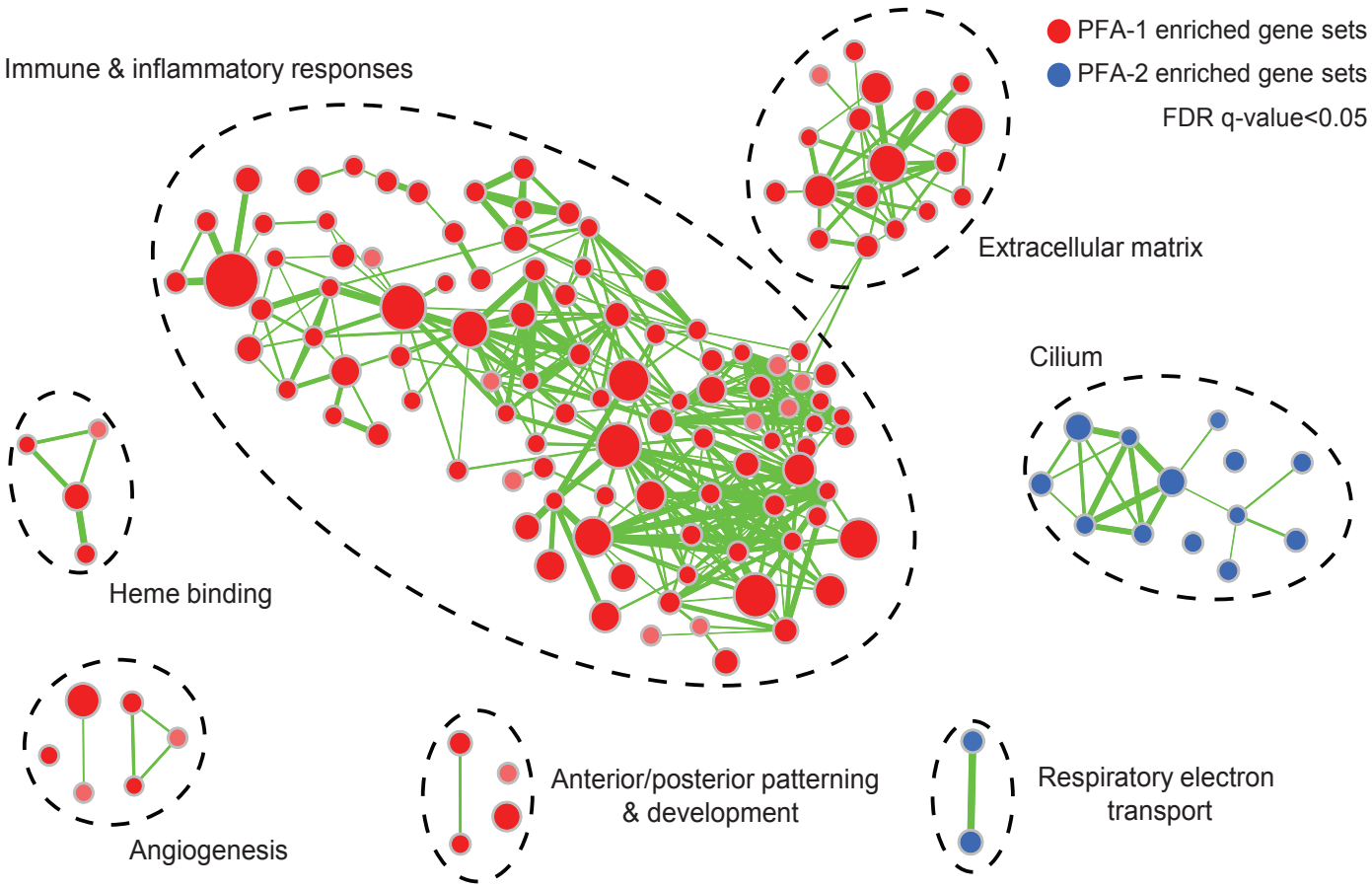
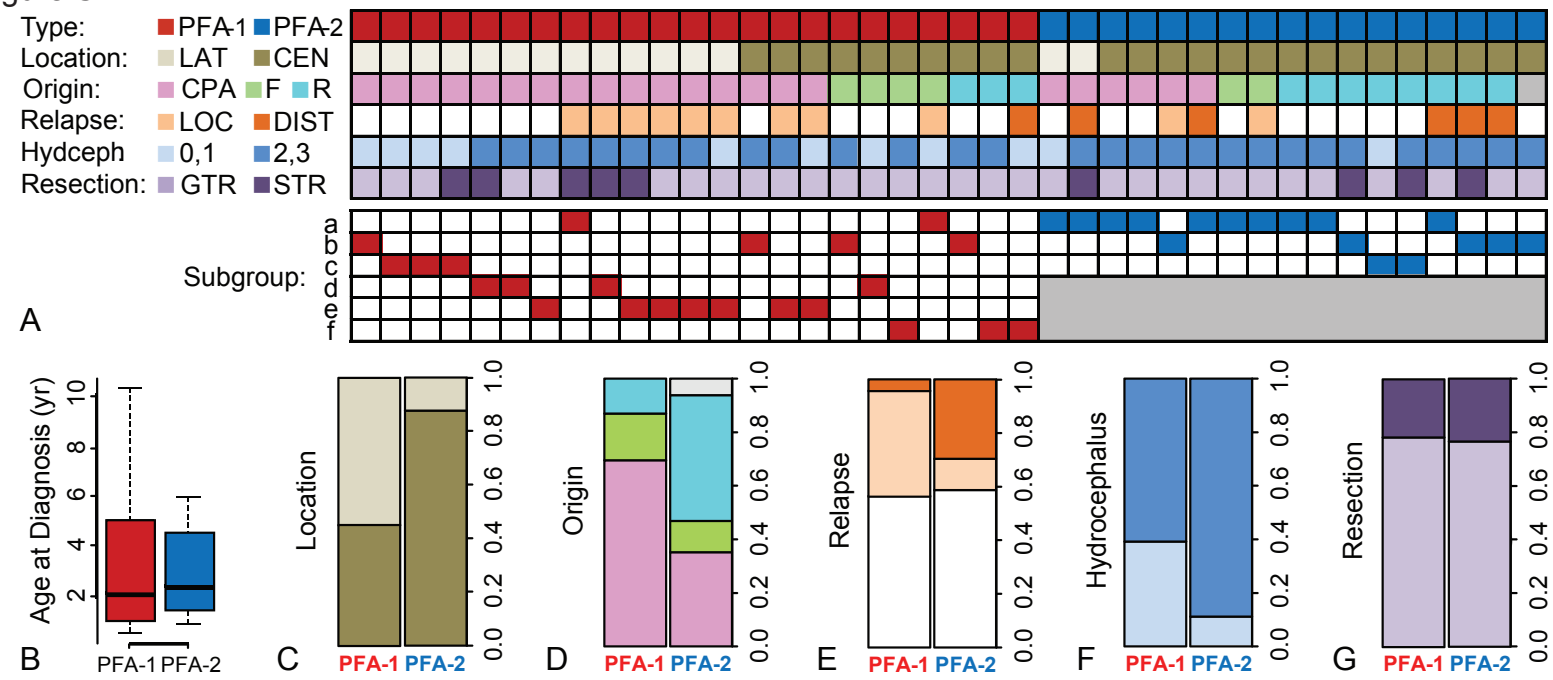
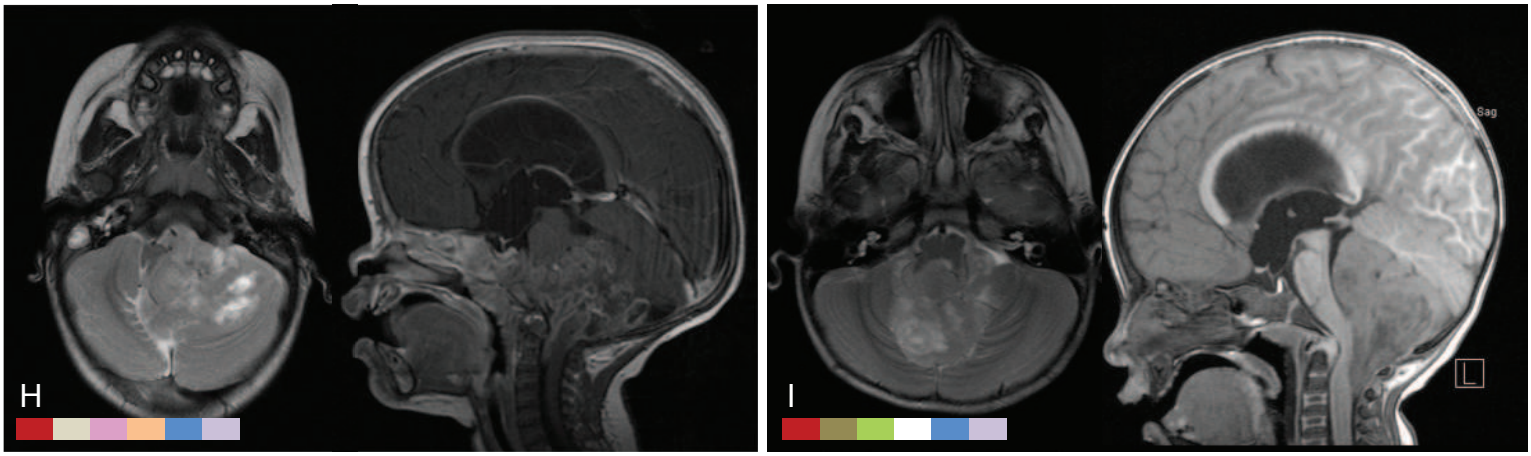
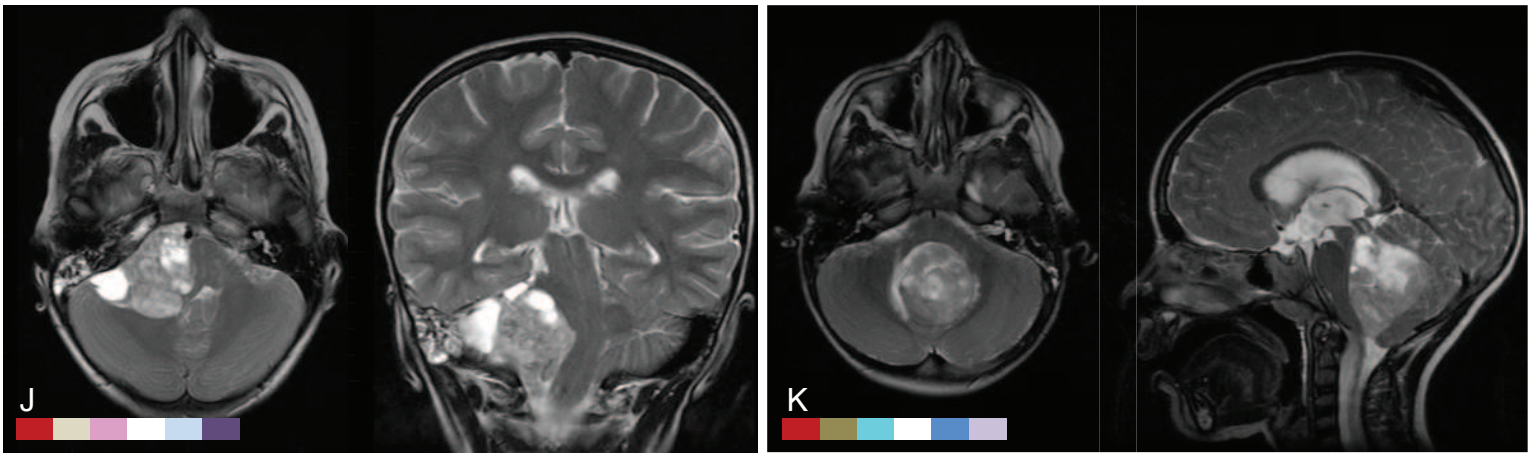

PFA-1 enriched gene sets
PFA-2 enriched gene sets
FDR q-value<0.05
Immune & inflammatory responses
Extracellular matrix
Cilium
Heme binding
Angiogenesis
Anterior/posterior patterning & development
Respiratory electron transport

Figure S4


Type: PFA-1 PFA-2
Location: LAT CEN
Origin: CPA F R
Relapse: LOC DIST
Hydceph 0,1 2,3
Resection: GTR STR
Subgroup: a b c d e f
A

Age at Diagnosis (yr)
PFA-1 PFA-2
B
Location
C
Origin
D
Relapse
E
Hydrocephalus
F
Resection
G


H
I


J
K

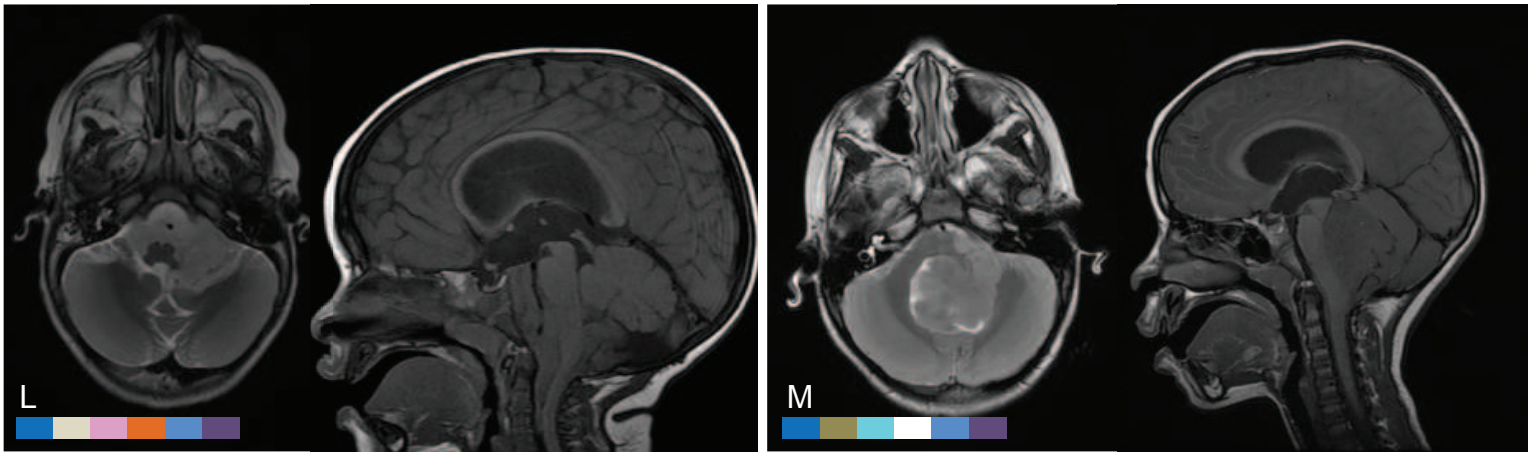
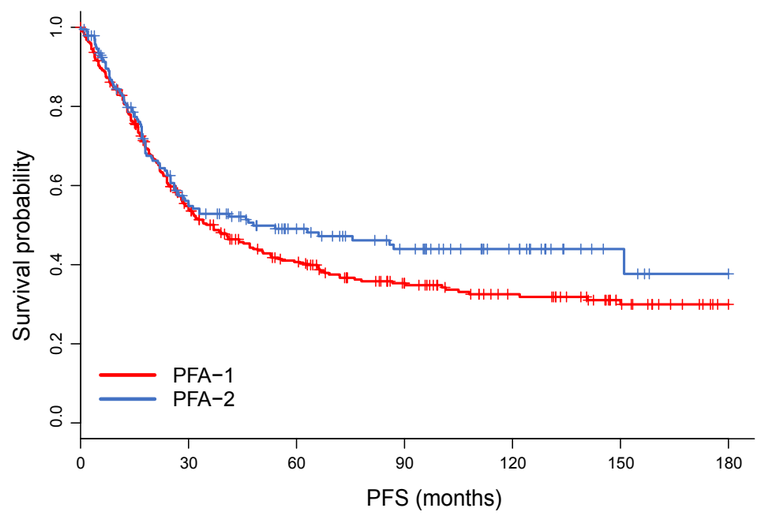

L
M

Figure S5

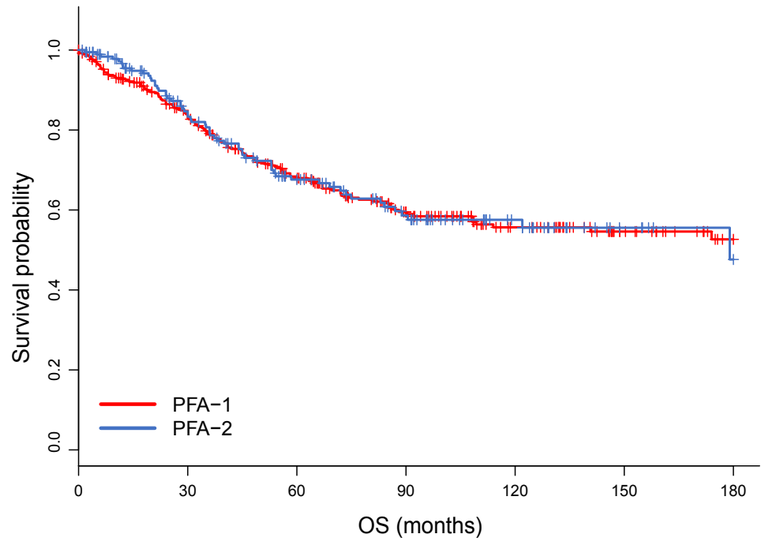
A

	PFA-1 (N=464)	PFA-2 (N=211)	ALL (N=675)	p value
Age at Dx (years)				0.8587
n	382	188	570	
Median (Range)	2.8 (0.0 - 54.2)	2.8 (0.0 - 32.0)	2.8 (0.0 - 54.2)	
Gender				0.3127
n	464	211	675	
Female	201 (43.3%)	82 (38.9%)	283 (41.9%)	
Male	263 (56.7%)	129 (61.1%)	392 (58.1%)	
Level of Resection				0.4589
n	381	188	569	
GTR	238 (62.5%)	124 (66.0%)	362 (63.6%)	
STR	143 (37.5%)	64 (34.0%)	207 (36.4%)	
Radiotherapy				0.7368
n	381	188	569	
Yes	304 (79.8%)	153 (81.4%)	457 (80.3%)	
No	77 (20.2%)	35 (18.6%)	112 (19.7%)	
Chemotherapy				0.0845
n	374	187	561	
Yes	229 (61.2%)	100 (53.5%)	329 (58.6%)	
No	145 (38.8%)	87 (46.5%)	232 (41.4%)	
WHO grade				0.1377
n	211	104	315	
2	71 (33.6%)	44 (42.3%)	115 (36.5%)	
3	140 (66.4%)	60 (57.7%)	200 (63.5%)	
Relapse				0.0204
n	26	10	36	
Distant	6 (23.1%)	7 (70.0%)	13 (36.1%)	
Local	18 (69.2%)	2 (20.0%)	20 (55.6%)	
Local & Distant	2 (7.7%)	1 (10.0%)	3 (8.3%)	
Institute				0.8588
n	464	211	675	
DKFZ-1	79 (17.0%)	36 (17.1%)	115 (17.0%)	
DKFZ-2	49 (10.6%)	18 (8.5%)	67 (9.9%)	
St. Jude	87 (18.8%)	45 (21.3%)	132 (19.6%)	
Toronto	203 (43.8%)	89 (42.2%)	292 (43.3%)	
UN-UK	46 (9.9%)	23 (10.9%)	69 (10.2%)	

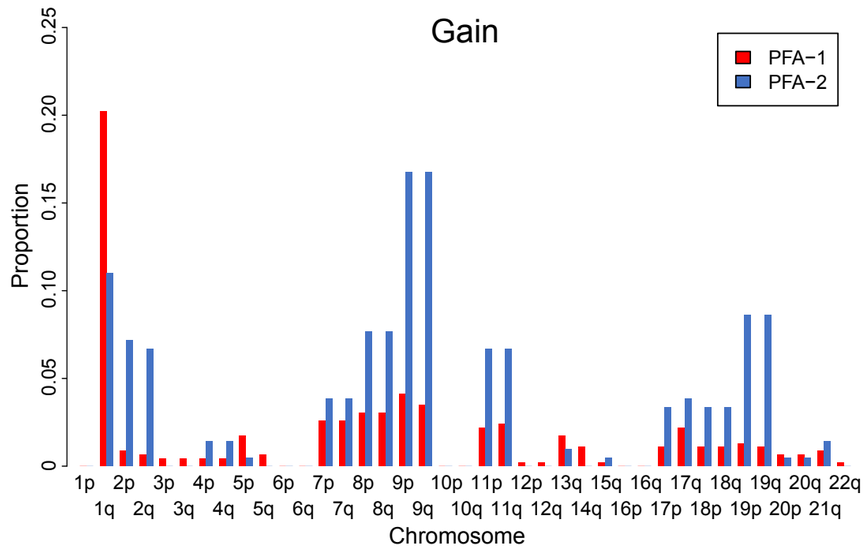
B



C



D



E

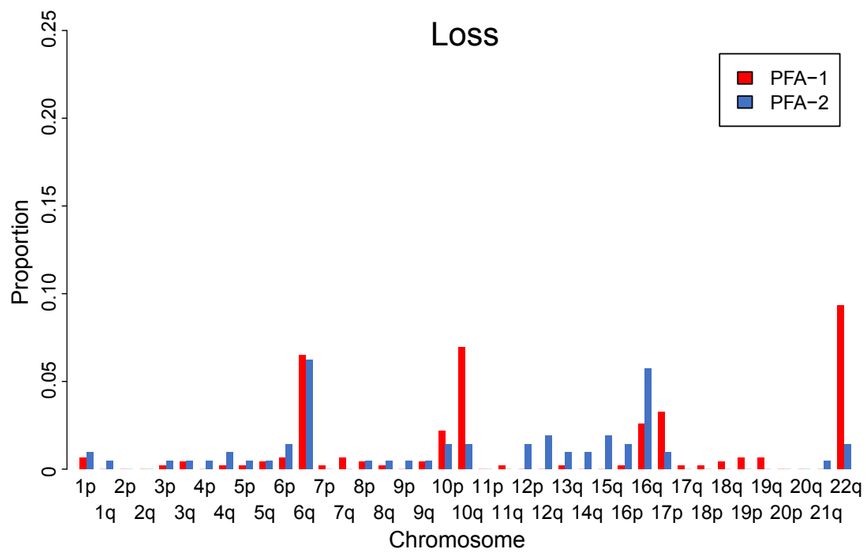


Figure S6

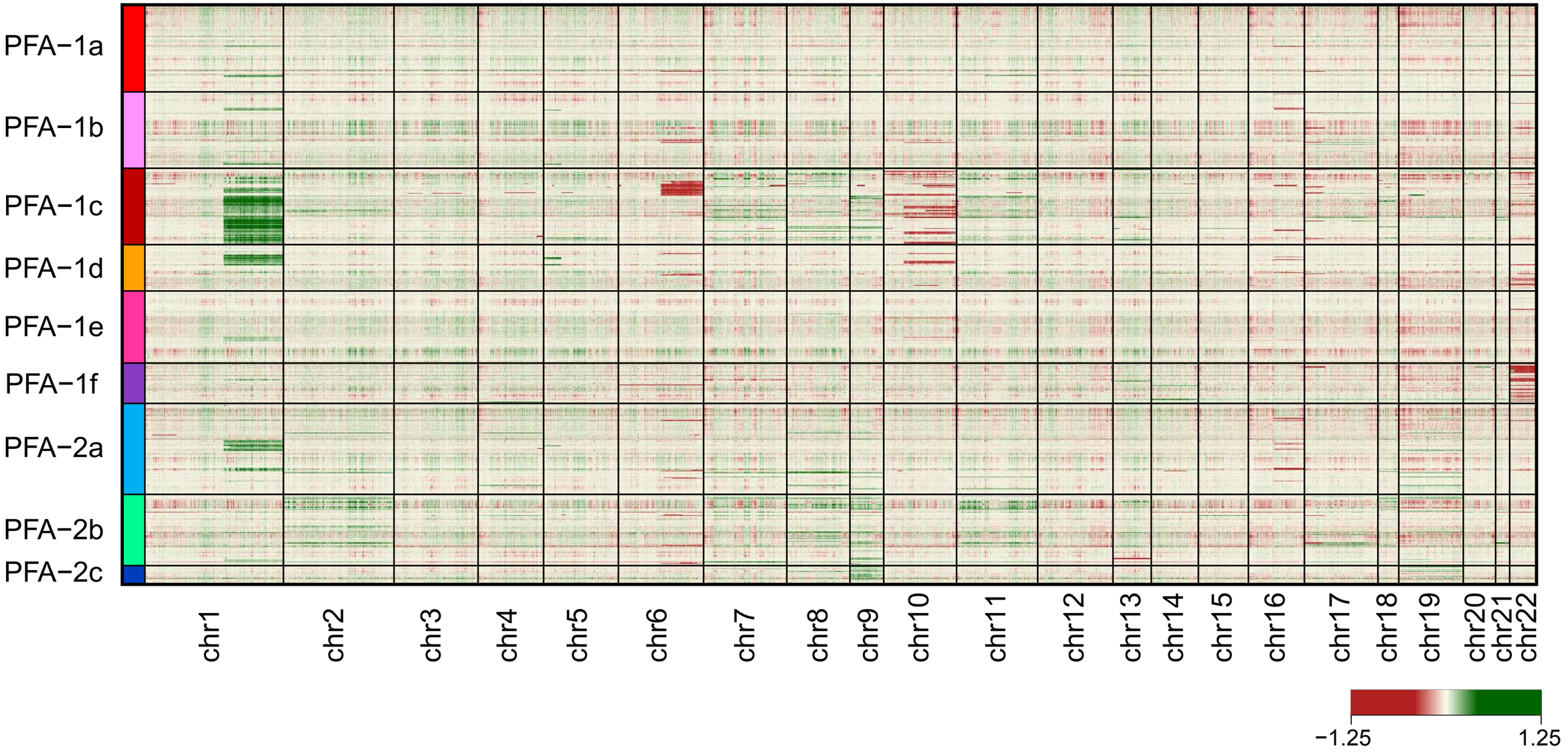
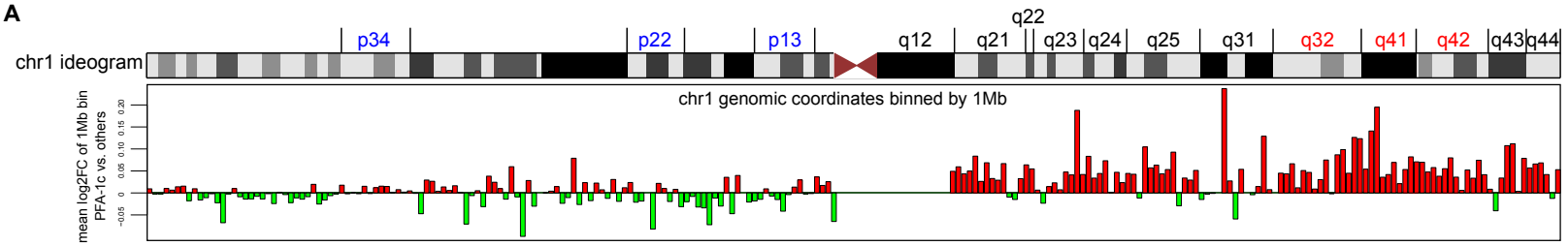

PFA−1a
PFA−1b
PFA−1c
PFA−1d
PFA−1e
PFA−1f
PFA−2a
PFA−2b
PFA−2c
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
−1.25
1.25

Figure S7

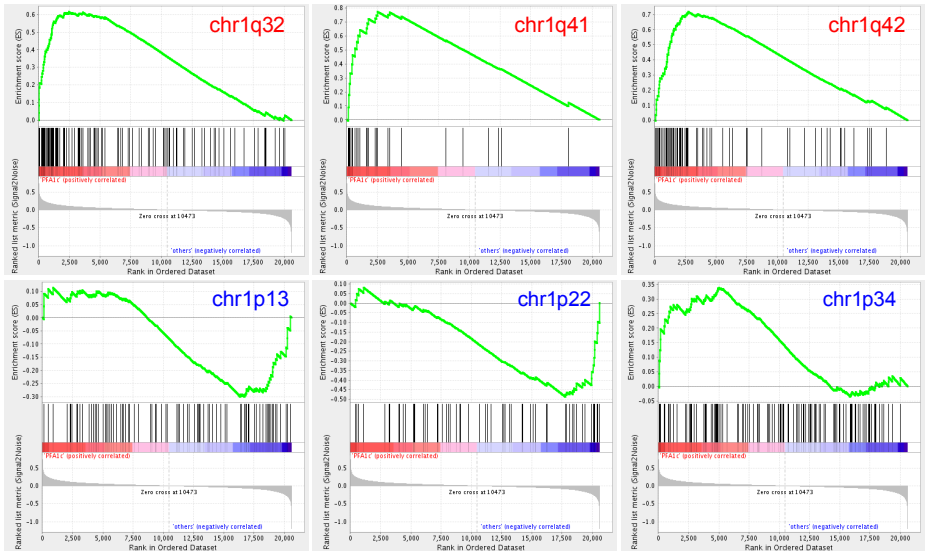
A



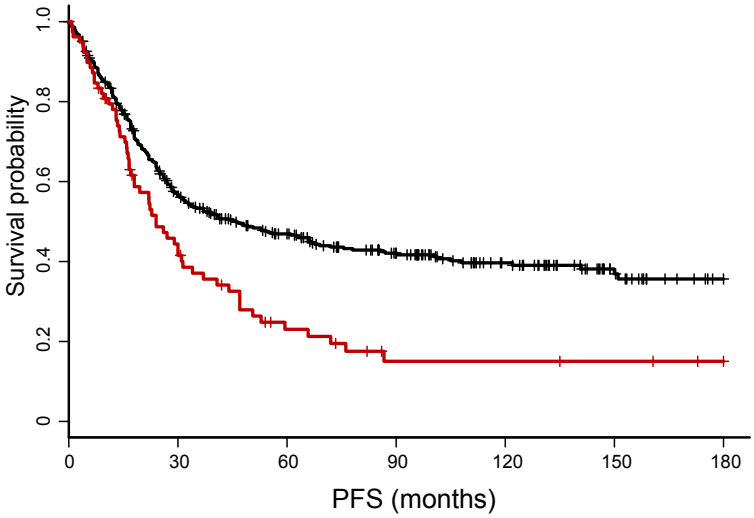
B

Gene sets upregulated in PFA-1c (FDR pvalue<0.05)					
NAME	SIZE	ES	NES	NOM p	FDR q-val
chr1q42	79	0.72	2.58	0	0
chr1q32	120	0.61	2.35	0	0
chr1q41	25	0.77	2.22	0	0
chr1q44	32	0.70	2.09	0	8.93E-04
chr1q21	195	0.50	2.09	0	7.14E-04
chr1q22	56	0.60	1.98	0	1.70E-03
chr15q14	36	0.59	1.87	2.14E-03	1.08E-02
chr7p15	53	0.56	1.85	0	1.29E-02
chr13q13	22	0.65	1.77	0	2.69E-02
chr1q24	34	0.58	1.77	4.26E-03	2.47E-02
chr1q43	17	0.66	1.73	6.19E-03	3.92E-02
chr1q25	58	0.49	1.71	0	4.27E-02
chr5p12	11	0.74	1.71	1.05E-02	4.27E-02
chr11p14	16	0.66	1.70	5.95E-03	4.20E-02

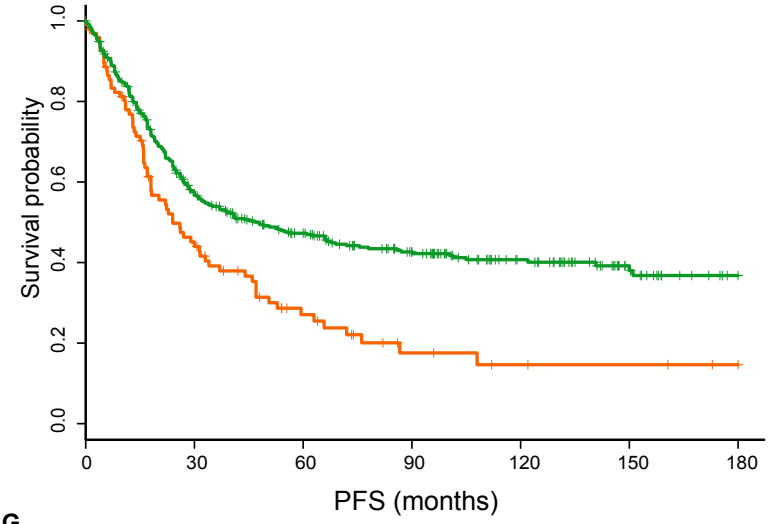
C



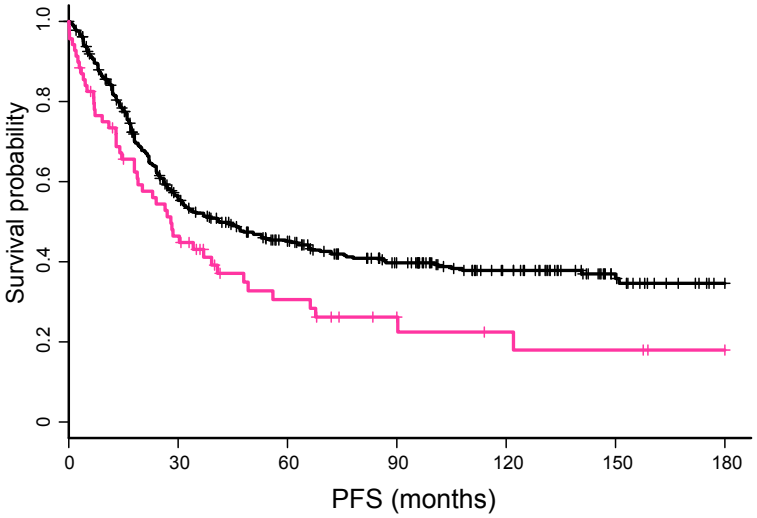
D



E



F



G

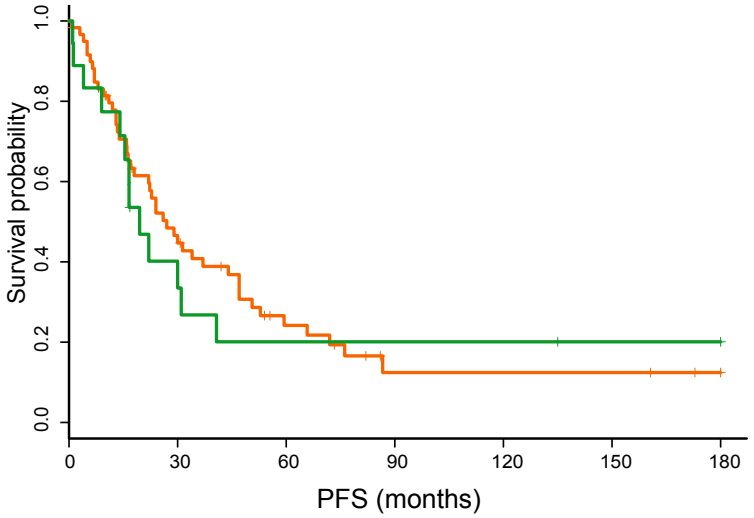


Figure S8

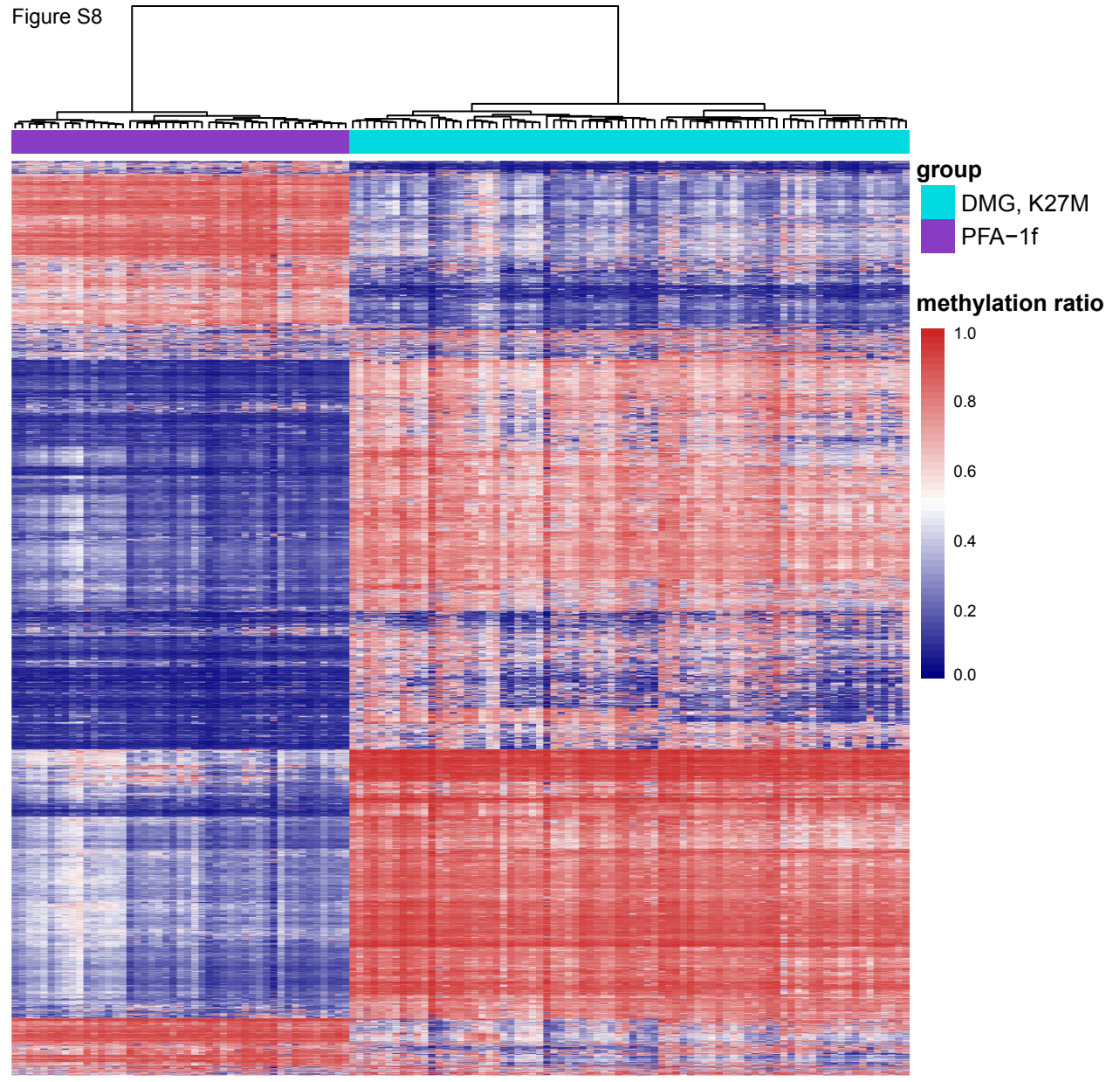


Table S1 Clinical characteristics of PFA ependymomas from different centers

	St. Jude (N=132)	DKFZ1 (N=115)	DKFZ2 (N=67)	Sick Kids (N=292)	UN-UK (N=69)	All (N=675)
Age @ diagnosis (yrs)						
n	132	85	64	220	69	570
Median (range)	2.1 (0.0 - 15.9)	3.1 (0.6 - 12.0)	3.0 (0.0 - 15.0)	3.0 (0.0 - 54.2)	2.8 (0.8 - 16.3)	2.8 (0.0 - 54.2)
Gender						
n	132	85	64	220	69	570
Female	59 (44.7%)	25 (29.4%)	24 (37.5%)	100 (45.5%)	25 (36.2%)	233 (40.9%)
Male	73 (55.3%)	60 (70.6%)	40 (62.5%)	120 (54.5%)	44 (63.8%)	337 (59.1%)
Level of Resection						
n	131	85	64	220	69	569
GTR	97 (74.0%)	40 (47.1%)	40 (62.5%)	150 (68.2%)	35 (50.7%)	362 (63.6%)
STR	34 (26.0%)	45 (52.9%)	24 (37.5%)	70 (31.8%)	34 (49.3%)	207 (36.4%)
Radiotherapy						
n	131	85	64	220	69	569
Yes	126 (96.2%)	54 (63.5%)	55 (85.9%)	170 (77.3%)	52 (75.4%)	457 (80.3%)
No	5 (3.8%)	31 (36.5%)	9 (14.1%)	50 (22.7%)	17 (24.6%)	112 (19.7%)
Chemotherapy						
n	132	85	64	211	69	561
Yes	46 (34.8%)	69 (81.2%)	58 (90.6%)	102 (48.3%)	54 (78.3%)	329 (58.6%)
No	86 (65.2%)	16 (18.8%)	6 (9.4%)	109 (51.7%)	15 (21.7%)	232 (41.4%)
WHO grade						
n	99	84	64	NA	68	315
2	49 (49.5%)	21 (25.0%)	10 (15.6%)	NA	35 (51.5%)	115 (36.5%)
3	50 (50.5%)	63 (75.0%)	54 (84.4%)	NA	33 (48.5%)	200 (63.5%)

N = total number of tumors from each center used for subgroup analysis by DNA methylation profiling.
n = number of patients for which clinical information was available.
DKFZ; Deutsches Krebsforschungszentrum – German Cancer Research Center
UN-UK; University of Nottingham, United Kingdom
GTR; gross total resection – STR; subtotal resection
NA; not available

Table S2 Relatively upregulated genes in a comparison of PFA-1 and PFA-2 ependymomas using both transcriptome sequencing and Affymetrix arrays

PFA-1 ependymomas						PFA-2 ependymomas					
Transcriptome sequencing			Affymetrix			Transcriptome sequencing			Affymetrix		
gene	adj.P.Val	B	gene	adj.P.Val	B	gene	adj.P.Val	B	gene	adj.P.Val	B
HOTAIRM1	0.000000	*27.42*	*HOTAIRM1*	3.33E-44	94.38	*EN2*	0.000000	14.65	*EN2*	1.03E-29	64.61
HOXB3	0.000000	*26.65*	*HOXA3*	1.54E-39	85.02	*MPPED2*	0.000001	12.07	*CNPY1*	1.71E-21	46.05
HOXB2	0.000000	*25.07*	*HOTAIRM1*	5.00E-35	75.77	*CNPY1*	0.000002	11.92	*MPPED2*	7.39E-16	33.00
HOXB4	0.000000	*22.26*	*HOXA2*	6.94E-29	62.67	*PTPN3*	0.000007	10.25	*LRRC8D*	1.51E-15	32.21
HOXA2	0.000000	*20.51*	*HOXB3*	1.71E-27	59.57	*ST6GAL2*	0.000016	9.46	*ARNTL2*	9.39E-15	30.39
HOXA3	0.000000	*19.58*	*SKAP2*	9.64E-26	55.70	*AC008060.5*	0.000019	8.58	*PTPN3*	1.40E-14	29.98
HOXA1	0.000000	*18.79*	*SKAP2*	1.74E-24	52.89	*B3GALTL*	0.000040	8.18	*ARNTL2*	2.42E-14	29.42
HOXA-AS2	0.000000	*18.59*	*HOXA4*	1.47E-23	50.79	*AC005235.1*	0.000034	8.11	*TCF7L2*	1.98E-13	27.36
MIR10A	0.000000	*16.18*	*HOXB2*	1.74E-22	48.37	*LRRC8D*	0.000055	7.79	*SH3GL3*	1.99E-13	27.33
HOXB-AS1	0.000000	*15.75*	*HOXB4*	1.03E-21	46.61	*SYT10*	0.000058	7.63	*RASSF10*	1.99E-13	27.30
HOXB-AS2	0.000000	*14.07*	*HOXA1*	2.74E-21	45.53	*WBSCR17*	0.000084	7.43	*SLC12A4*	4.23E-13	26.50
AC106786.1	0.000000	*13.60*	*SKAP2*	1.38E-19	41.75	*PCBP3*	0.000095	7.16	*TCF7L2*	4.97E-13	26.32
ZNF503	0.000000	*13.28*	*HOXA-AS2*	6.57E-19	40.14	*EFCC1*	0.000123	6.97	*ARNTL2*	5.84E-13	26.14
LPAR3	0.000004	*10.98*	*SKAP2*	1.75E-17	36.96	*LPO*	0.000123	6.92	*PTPN3*	7.11E-13	25.92
PRELP	0.000006	*10.34*	*ZNF503*	2.21E-17	36.67	*SMCO2*	0.000223	6.31	*NTS*	1.23E-12	25.35
SEMA3C	0.000009	*10.02*	*PRDM6*	3.10E-17	36.30	*NEDD4L*	0.000223	6.13	*FAM65B*	1.56E-12	25.10
PRDM6	0.000022	*8.95*	*PRDM6*	6.29E-17	35.58	*EYA4*	0.000271	6.02	*EXOC5*	1.94E-12	24.87
ZNF503-AS2	0.000052	*8.06*	*WIF1*	6.45E-17	35.51	*LOXL4*	0.000339	5.82	*GPR37*	2.32E-12	24.67
SKAP2	0.000080	*7.50*	*SEMA3C*	3.21E-16	33.88	*RNF43*	0.000281	5.82	*TCF7L2*	3.45E-12	24.25
SLC35F2	0.000082	*7.44*	*HOXA5*	9.83E-16	32.69	*SERPINI1*	0.000368	5.61	*ITIH5*	3.94E-12	24.10
HOXC4	0.000146	*6.81*	*CCDC85A*	1.51E-15	32.21	*ITIH5*	0.000411	5.31	*LINC01114*	4.16E-12	24.00
HOXD4	0.000123	*6.17*	*PRELP*	1.82E-15	32.00	*GPR39*	0.000740	4.85	*BZRAP1*	4.16E-12	23.97
PPP4R4	0.000268	*5.97*	*SLC35F2*	2.21E-13	27.18	*PROB1*	0.000602	4.82	*GRIK1*	4.16E-12	23.97
AJAP1	0.000339	*5.69*	*PRELP*	3.17E-13	26.80	*TOX3*	0.000761	4.77	*EYA4*	5.91E-12	23.60
HOXA4	0.000241	*5.60*	*SEMA3C*	9.02E-13	25.67	*ARNTL2*	0.000740	4.76	*GREB1L*	6.02E-12	23.56
FGFR2	0.000368	*5.42*	*HOXD3*	1.25E-11	22.82	*BZRAP1*	0.000730	4.69	*TCF7L2*	1.07E-11	22.99
FGFRL1	0.000403	*5.28*	*HOXD4*	2.84E-11	21.96	*RFX3*	0.000744	4.55	*RNF43*	1.44E-11	22.67
TICAM1	0.000513	*5.12*	*SEMA3B*	2.84E-11	21.95	*NTS*	0.000898	4.50	*UNC79*	1.51E-11	22.61
LRRTM4	0.000689	*4.85*	*C3orf70*	6.63E-11	21.04	*ADAMTS3*	0.000761	4.43	*FAM216B*	2.83E-11	21.99
GPR133	0.000676	*4.77*	*PRELP*	2.39E-10	19.63	*TCTEX1D1*	0.001023	4.15	*EYA4*	3.15E-11	21.84
CXXC5	0.000689	*4.69*	*PIK3R1*	2.86E-10	19.41	*NR1H4*	0.001269	4.13	*EPB41L4A*	3.24E-11	21.79
CCDC85A	0.000874	*4.56*	*RBMS3*	2.99E-10	19.35	*KRT81*	0.001204	4.12	*COL21A1*	3.42E-11	21.72
SEMA3B	0.000754	*4.54*	*VLDLR*	3.26E-10	19.26	*TEX15*	0.001241	4.10	*WBSCR17*	3.42E-11	21.71
CNKSR3	0.000874	*4.39*	*RBMS3*	4.09E-10	19.01	*ASB18*	0.001217	3.99	*CATSPERD*	5.04E-11	21.32
DPT	0.001234	*4.18*	*HOXA-AS2*	4.18E-10	18.98	*SH3GL3*	0.001349	3.84	*C8orf76*	7.45E-11	20.92
ZNF703	0.001190	*3.88*	*FGFR2*	5.36E-10	18.69	*FAM65C*	0.001349	3.81	*FAM65B*	8.35E-11	20.78
HCN1	0.001349	*3.74*	*RBMS3*	5.36E-10	18.69	*LINC00907*	0.001275	3.79	*TCERG1L*	8.35E-11	20.78
LRP2	0.001533	*3.73*	*KCNE4*	6.16E-10	18.51	*LMX1B*	0.001349	3.78	*GPR37*	8.35E-11	20.77
LINC00626	0.001557	*3.56*	*HOXB-AS1*	9.43E-10	18.07	*TRPC6*	0.001349	3.76	*MPPED2*	9.07E-11	20.66
HSPB1	0.001483	*3.51*	*ADGRD1*	1.18E-09	17.81	*LECT1*	0.001557	3.69	*TOX3*	9.07E-11	20.66
STAC2	0.001523	*3.39*	*MXRA8*	1.35E-09	17.67	*AL592528.1*	0.001349	3.68	*LYPD1*	1.03E-10	20.52
VLDLR	0.001670	*3.32*	*LRP2*	1.90E-09	17.23	*SEMA3D*	0.001349	3.65	*KCNE5*	1.34E-10	20.26
AC007743.1	0.001672	*3.32*	*RBMS3*	2.08E-09	17.12	*TP73*	0.001349	3.63	*SVOPL*	1.66E-10	20.04
COTL1	0.001582	*3.29*	*HCN1*	3.86E-09	16.40	*ELMOD1*	0.001483	3.61	*EPB41L4A*	1.73E-10	19.99
RBMS3	0.001872	*3.20*	*COTL1*	4.04E-09	16.35	*LGR4*	0.001349	3.61	*PLEKHA5*	1.90E-10	19.88
WIF1	0.002180	*3.15*	*BGN*	4.25E-09	16.29	*TMEM254*	0.001369	3.50	*GLUD1*	2.09E-10	19.78
INPP4B	0.001872	*3.03*	*FGFR2*	4.52E-09	16.21	*CHL1*	0.001483	3.47	*LINC00907*	2.30E-10	19.68
RHOBTB3	0.001976	*2.95*	*TNFRSF11A*	4.89E-09	16.09	*GREB1L*	0.001423	3.47	*ITIH5*	2.49E-10	19.57
TNFRSF11A	0.002407	*2.89*	*SUSD4*	6.94E-09	15.67	*SPATA42*	0.001872	3.41	*LOC101928370*	2.86E-10	19.42
EGF	0.003270	*2.44*	*RHOBTB3*	7.31E-09	15.59	*TDRD1*	0.001882	3.40	*ZFAND5*	2.86E-10	19.41
LAMA2	0.002694	*2.44*	*NETO1*	7.55E-09	15.56	*ITGA8*	0.001731	3.39	*SOX21-AS1*	3.75E-10	19.10
PAX8	0.003270	*2.43*	*GUCY1A3*	9.65E-09	15.25	*KCNG2*	0.001904	3.35	*LOC101927008*	3.75E-10	19.10
HOXD-AS1	0.003270	*2.39*	*BGN*	9.81E-09	15.23	*SVOPL*	0.001872	3.28	*TOX3*	4.62E-10	18.86
SERPINA12	0.003270	*2.38*	*DPT*	1.04E-08	15.17	*NRSN1*	0.001976	3.24	*GRHL1*	5.68E-10	18.62
PLSCR4	0.002763	*2.35*	*HOXB-AS1*	1.21E-08	14.98	*ERC2*	0.001904	3.20	*F2RL1*	5.71E-10	18.60
STXBP5L	0.002872	*2.24*	*COTL1*	1.21E-08	14.96	*CATSPERD*	0.001976	3.17	*CCDC88C*	5.71E-10	18.59
PIK3R1	0.003009	*2.17*	*FJX1*	1.25E-08	14.93	*SPATA6*	0.001672	3.16	*SDK1*	6.55E-10	18.44
ARHGEF40	0.002990	*2.16*	*HOXB3*	1.49E-08	14.73	*CSPP1*	0.001872	3.00	*WDR38*	8.13E-10	18.23

Table S3 Univariate survival data

All PFA ependymomas				
Variable	**PFS Hazard Ratio**	**PFS p-value**	**OS Hazard Ratio**	**OS p-value**
Age Years	0.98 (0.95, 1.01)	0.15	0.97 (0.93, 1.01)	0.099
Gender Male:Female	1.62 (1.28, 2.04)	<0.0001	1.57 (1.16, 2.13)	0.0031
Resection GTR:STR	0.52 (0.42, 0.65)	< 0.0001	0.48 (0.36, 0.64)	< 0.0001
Radiotherapy Yes:No	0.61 (0.46, 0.80)	0.0003	0.45 (0.33, 0.63)	< 0.0001
Grade II:III	0.76 (0.56, 1.04)	0.0826	0.73 (0.48, 1.09)	0.1246
1q gain Yes:No	1.67 (1.28, 2.18)	0.0001	2.08 (1.49, 2.90)	< 0.0001
6q loss Yes:No	1.72 (1.06, 2.80)	<0.0001	1.72 (1.06, 2.80)	0.0264
CXorf67 mutation Yes:No	0.69 (0.30, 1.59)	0.4205	0.69 (0.30, 1.59)	0.3796

PFA-1 ependymomas				
Variable	**PFS Hazard Ratio**	**PFS p-value**	**OS Hazard Ratio**	**OS p-value**
Age Years	0.98 (0.96, 1.01)	0.34	0.97 (0.93, 1.01)	0.19
Gender Male:Female	1.82 (1.37, 2.41)	<0.0001	1.74 (1.20, 2.53)	0.0032
Resection GTR:STR	0.48 (0.36, 0.62)	< 0.0001	0.45 (0.32, 0.64)	< 0.0001
Radiotherapy Yes:No	0.53 (0.39, 0.73)	<0.0001	0.38 (0.26, 0.56)	< 0.0001
Grade II:III	0.68 (0.47, 1.00)	0.0501	0.61 (0.37, 1.02)	0.0582
1q gain Yes:No	1.71 (1.26, 2.31)	0.0004	2.33 (1.60, 3.41)	< 0.0001
6q loss Yes:No	2.48 (1.58, 3.89)	<0.0001	2.16 (1.24, 3.76)	0.0055
CXorf67 mutation Yes:No	0.56 (0.24, 1.28)	0.1634	0.43 (0.13, 1.37)	0.1427

PFA-2 ependymomas				
Variable	**PFS Hazard Ratio**	**PFS p-value**	**OS Hazard Ratio**	**OS p-value**
Age Years	0.92 (0.84, 1.01)	0.079	0.94 (0.84, 1.05)	0.28
Gender Male:Female	1.26 (0.82, 1.93)	0.2984	1.26 (0.75, 2.13)	0.379
Resection GTR:STR	0.63 (0.41, 0.96)	0.0287	0.56 (0.33, 0.92)	0.0209
Radiotherapy Yes:No	0.83 (0.49, 1.42)	0.4956	0.66 (0.36, 1.20)	0.1684
Grade II:III	0.99 (0.57, 1.72)	0.9756	1.08 (0.53, 2.22)	0.8343
1q gain Yes:No	1.45 (0.79, 2.67)	0.2293	1.47 (0.70, 3.10)	0.3067
6q loss Yes:No	1.66 (0.83, 3.32)	0.1468	0.98 (0.36, 2.72)	0.971
CXorf67 mutation Yes:No	1.37 (0.49, 3.87)	0.5544	1.5 (0.45, 5.07)	0.5125

Table S4 Clinical and genetic characteristics of PFA ependymoma subtypes

	PFA-1a (N=101)	PFA-1b (N=89)	PFA-1c (N=89)	PFA-1d (N=54)	PFA-1e (N=84)	PFA-1f (N=47)	PFA-2a (N=106)	PFA-2b (N=83)	PFA-2c (N=22)	ALL (N=675)	p value
Age at Diagnosis (yrs)											<0.0001
n	85	77	79	34	69	38	96	72	20	570	
Median (range)	1.8 (0.4 - 7.0)	3.2 (0.6 - 54.2)	7.7 (0.6 - 52.3)	3.1 (1.9 - 9.6)	1.0 (0.0 - 6.9)	5.0 (1.8 - 19.0)	1.9 (0.0 - 32.0)	3.9 (0.0 - 14.7)	4.3 (2.0 - 9.2)	2.8 (0.0 - 54.2)	
Gender											0.0236
n	101	89	89	54	84	47	106	83	22	675	
Female	42 (41.6%)	50 (56.2%)	36 (40.4%)	21 (38.9%)	27 (32.1%)	25 (53.2%)	36 (34.0%)	34 (41.0%)	12 (54.5%)	283 (41.9%)	
Male	59 (58.4%)	39 (43.8%)	53 (59.6%)	33 (61.1%)	57 (67.9%)	22 (46.8%)	70 (66.0%)	49 (59.0%)	10 (45.5%)	392 (58.1%)	
Level of Resection											0.0013
n	85	77	79	34	69	37	96	72	20	569	
GTR	56 (65.9%)	62 (80.5%)	48 (60.8%)	13 (38.2%)	36 (52.2%)	23 (62.2%)	65 (67.7%)	44 (61.1%)	15 (75.0%)	362 (63.6%)	
STR	29 (34.1%)	15 (19.5%)	31 (39.2%)	21 (61.8%)	33 (47.8%)	14 (37.8%)	31 (32.3%)	28 (38.9%)	5 (25.0%)	207 (36.4%)	
Radiotherapy											<0.0001
n	85	77	79	34	69	37	96	72	20	569	
Yes	59 (69.4%)	66 (85.7%)	72 (91.1%)	30 (88.2%)	41 (59.4%)	36 (97.3%)	74 (77.1%)	61 (84.7%)	18 (90.0%)	457 (80.3%)	
No	26 (30.6%)	11 (14.3%)	7 (8.9%)	4 (11.8%)	28 (40.6%)	1 (2.7%)	22 (22.9%)	11 (15.3%)	2 (10.0%)	112 (19.7%)	
Chemotherapy											0.0601
n	83	73	78	34	69	37	95	72	20	561	
Yes	56 (67.5%)	36 (49.3%)	44 (56.4%)	23 (67.6%)	45 (65.2%)	25 (67.6%)	57 (60.0%)	35 (48.6%)	8 (40.0%)	329 (58.6%)	
No	27 (32.5%)	37 (50.7%)	34 (43.6%)	11 (32.4%)	24 (34.8%)	12 (32.4%)	38 (40.0%)	37 (51.4%)	12 (60.0%)	232 (41.4%)	
Institute											0.0885
n	101	89	89	54	84	47	106	83	22	675	
St. Jude	25 (24.8%)	17 (19.1%)	10 (11.2%)	12 (22.2%)	18 (21.4%)	5 (10.6%)	28 (26.4%)	13 (15.7%)	4 (18.2%)	132 (19.6%)	
DKFZ-1	10 (9.9%)	17 (19.1%)	15 (16.9%)	12 (22.2%)	17 (20.2%)	8 (17.0%)	9 (8.5%)	23 (27.7%)	4 (18.2%)	115 (17.0%)	
DKFZ-2	9 (8.9%)	6 (6.7%)	11 (12.4%)	4 (7.4%)	8 (9.5%)	11 (23.4%)	10 (9.4%)	7 (8.4%)	1 (4.5%)	67 (9.9%)	
Toronto	46 (45.5%)	41 (46.1%)	41 (46.1%)	22 (40.7%)	36 (42.9%)	17 (36.2%)	44 (41.5%)	34 (41.0%)	11 (50.0%)	292 (43.3%)	
UN-UK	11 (10.9%)	8 (9.0%)	12 (13.5%)	4 (7.4%)	5 (6.0%)	6 (12.8%)	15 (14.2%)	6 (7.2%)	2 (9.1%)	69 (10.2%)	

	PFA-1a (N=101)	PFA-1b (N=89)	PFA-1c (N=89)	PFA-1d (N=54)	PFA-1e (N=84)	PFA-1f (N=47)	PFA-2a (N=106)	PFA-2b (N=83)	PFA-2c (N=22)	ALL (N=675)	p value
Pathology grade											<0.0001
n	46	43	40	23	37	22	50	43	11	315	
WHO II	11 (23.9%)	25 (58.1%)	8 (20.0%)	15 (65.2%)	8 (21.6%)	4 (18.2%)	17 (34.0%)	19 (44.2%)	8 (72.7%)	115 (36.5%)	
WHO III	35 (76.1%)	18 (41.9%)	32 (80.0%)	8 (34.8%)	29 (78.4%)	18 (81.8%)	33 (66.0%)	24 (55.8%)	3 (27.3%)	200 (63.5%)	
Gain chromosome 1q											<0.0001
n	101	88	86	54	84	47	105	82	22	669	
Yes	4 (4.0%)	7 (8.0%)	63 (73.3%)	14 (25.9%)	4 (4.8%)	1 (2.1%)	18 (17.1%)	5 (6.1%)	0	116 (17.3%)	
No	97 (96.0%)	81 (92.0%)	23 (26.7%)	40 (74.1%)	80 (95.2%)	46 (97.9%)	87 (82.9%)	77 (93.9%)	22 (100%)	553 (82.7%)	
Loss chromosome 6q											<0.0001
n	101	88	86	54	84	47	105	82	22	669	
Yes	2 (2.0%)	4 (4.5%)	20 (23.3%)	3 (5.6%)	0	1 (2.1%)	5 (4.8%)	8 (9.8%)	0	43 (6.4%)	
No	99 (98.0%)	84 (95.5%)	66 (76.7%)	51 (94.4%)	84 (100%)	46 (97.9%)	100 (95.2%)	74 (90.2%)	22 (100%)	626 (93.6%)	
Loss chromosome 22q											<0.0001
n	101	88	86	54	84	47	105	82	22	669	
Yes	2 (2.0%)	3 (3.4%)	7 (8.1%)	5 (9.3%)	2 (2.4%)	24 (51.1%)	2 (1.9%)	1 (1.2%)	0	46 (6.9%)	
No	99 (98.0%)	85 (96.6%)	79 (91.9%)	49 (90.7%)	82 (97.6%)	23 (48.9%)	103 (98.1%)	81 (98.8%)	22 (100%)	623 (93.1%)	
***CXorf67* mutation**											0.5755
n	39	32	24	16	31	9	44	32	7	234	
Yes	4 (10.3%)	6 (18.8%)	1 (4.2%)	1 (6.3%)	3 (9.7%)	0	3 (6.8%)	4 (12.5%)	0	22 (9.4%)	
No	35 (89.7%)	26 (81.2%)	23 (95.8%)	15 (93.7%)	28 (90.3%)	9 (100%)	41 (93.2%)	28 (87.5%)	7 (100%)	212 (90.6%)	

N = total number of tumors from each center used for subgroup analysis by DNA methylation profiling.
n = number of patients for which clinical information was available.
DKFZ; Deutsches Krebsforschungszentrum – German Cancer Research Center
UN-UK; University of Nottingham, UK
GTR; gross total resection – STR; subtotal resection

Table S5 Immunoprecipitation / Mass Spectrometry data

Bait	PreyGene	Spec	SpecSum	iProb	ctrlCounts	AvgP/SAINT	MaxP	FDR
CXorf67	CXorf67	69|24|25	118	1.00|1.00|1.00	0|0|0	1	1	0
	EED	30|28|27	85	1.00|1.00|1.00	0|0|0	1	1	0
	NUMA1	130|155|121	406	1.00|1.00|1.00	0|0|0	1	1	0
	EZH2	49|52|57	158	1.00|1.00|1.00	0|0|0	1	1	0
	SUZ12	76|76|81	233	1.00|1.00|1.00	0|0|0	1	1	0
	RBMX	16|8|8	32	1.00|1.00|1.00	0|0|0	1	1	0
	RBBP4	20|13|11	44	1.00|1.00|1.00	0|0|0	0.9992	1	0.0001
	MTF2	27|8|17	52	1.00|1.00|1.00	0|0|0	0.9975	0.9975	0.0004
	YLPM1	15|21|18	54	1.00|1.00|1.00	0|0|0	0.9967	0.9975	0.0007
	EZH1	9|6|10	25	0.98|0.99|1.00	0|0|0	0.99	0.995	0.0017
	RBBP7	11|3|3	17	0.99|0.97|0.96	0|0|0	0.9717	0.9875	0.0041
	SAFB	7|3|2	12	0.97|0.97|0.93	0|0|0	0.9575	0.97	0.0073
	SAFB2	6|3|2	11	0.95|0.94|0.90	0|0|0	0.9325	0.9525	0.0119
	GLYR1	13|55|53	121	0.71|0.99|0.99	0|0|0	0.8992	0.9925	0.0183
	PPP1CA	2|2|4	8	0.83|0.89|0.94	0|0|0	0.885	0.935	0.0247
EZH2	EZH2	40|49	89	1.00|1.00	0|0	1	1	0
	KIF5B	29|19	48	1.00|1.00	0|0	1	1	0
	CIT	6|10	16	1.00|1.00	0|0	1	1	0
	EED	18|28	46	1.00|1.00	0|0	1	1	0
	SUZ12	42|62	104	1.00|1.00	0|0	1	1	0
	RBBP4	11|11	22	1.00|1.00	0|0	1	1	0
	MTF2	9|8	17	1.00|1.00	0|0	0.9975	0.9975	0.0004
	CXorf67	6|5	11	1.00|1.00	0|0	0.9963	0.9975	0.0008
	MYH9	9|6	15	1.00|1.00	0|0	0.995	0.995	0.0012
	NPLOC4	9|9	18	0.99|0.99	0|0	0.9925	0.9925	0.0019
	UFD1L	3|3	6	0.99|0.99	0|0	0.9912	0.9925	0.0025
	SAFB	4|5	9	0.99|0.99	0|0	0.9887	0.99	0.0038
	RBBP7	4|6	10	0.98|1.00	0|0	0.9887	0.995	0.0038
	RBMX	4|7	11	0.98|0.99	0|0	0.985	0.9925	0.0046
	HSP90AB1	2|3	5	0.97|0.98	0|0	0.9737	0.98	0.0061
SUZ12	DHX15	14|16|16	46	1.00|1.00|1.00	0|0|0	1	1	0
	SNRNP200	35|38|28	101	1.00|1.00|1.00	0|0|0	1	1	0
	BCLAF1	190|186|167	543	1.00|1.00|1.00	0|0|0	1	1	0
	EED	26|48|41	115	1.00|1.00|1.00	0|0|0	1	1	0
	EIF4A3	36|83|63	182	1.00|1.00|1.00	0|0|0	1	1	0
	THRAP3	152|155|152	459	1.00|1.00|1.00	0|0|0	1	1	0
	HNRNPUL2	47|62|65	174	1.00|1.00|1.00	0|0|0	1	1	0
	SF3B1	39|59|32	130	1.00|1.00|1.00	0|0|0	1	1	0
	CPSF1	15|47|36	98	1.00|1.00|1.00	0|0|0	1	1	0
	SUZ12	60|92|71	223	1.00|1.00|1.00	0|0|0	1	1	0
	ACIN1	62|16|12	90	1.00|1.00|1.00	0|0|0	0.9992	1	0.0003
	SF3B3	26|63|48	137	1.00|1.00|1.00	0|0|0	0.9992	1	0.0003
	EZH2	27|67|62	156	1.00|1.00|1.00	0|0|0	0.9992	1	0.0003
	TRA2A	14|17|8	39	1.00|1.00|1.00	0|0|0	0.9992	1	0.0003
	PRMT1	16|32|35	83	1.00|1.00|1.00	0|0|0	0.9992	1	0.0003
	PRPF8	42|26|17	85	1.00|1.00|1.00	0|0|0	0.9983	1	0.0006
	MTF2	27|20|9	56	1.00|1.00|1.00	0|0|0	0.9983	1	0.0006
	RBMX	88|29|22	139	1.00|1.00|1.00	0|0|0	0.9983	1	0.0006
	PNN	34|12|9	55	1.00|1.00|1.00	0|0|0	0.9983	1	0.0006
	PABPC1	30|21|22	73	1.00|1.00|1.00	0|0|0	0.9975	0.9975	0.0009
	PRPF19	17|28|26	71	1.00|1.00|1.00	0|0|0	0.9975	0.9975	0.0009
	DDX21	6|10|10	26	1.00|1.00|1.00	0|0|0	0.9975	0.9975	0.0009
	HNRNPA3	23|19|16	58	1.00|1.00|1.00	0|0|0	0.9975	0.9975	0.0009
	PABPC4	15|18|15	48	0.99|1.00|1.00	0|0|0	0.9958	0.9975	0.0012
	SF3B2	23|22|19	64	1.00|1.00|1.00	0|0|0	0.9958	0.9975	0.0012
	MOV10	11|8|7	26	1.00|1.00|1.00	0|0|0	0.995	0.995	0.0016
	CXorf67	30|25|15	70	1.00|1.00|1.00	0|0|0	0.995	0.995	0.0016
	HNRNPH3	11|15|14	40	0.99|1.00|1.00	0|0|0	0.995	0.9975	0.0016
	PABPN1	22|20|14	56	1.00|1.00|0.99	0|0|0	0.9942	0.995	0.0019
	SF3A1	11|17|11	39	0.98|1.00|1.00	0|0|0	0.9942	1	0.002

Bait: Antibody used for immunoprecipitation
PreyGene: Immunoprecipitated proteins
Spec: Spectral count in each replicate
SpecSum: Summation of Spec
iProb: SAINT probabilities computed for each replicate
ctrlCounts: Spectral counts in controls
AvgP/SAINT: SAINT score, defined as average of individual probabilities (iProb) over replicates
MaxP: Maximum of individual probabilities (iProb) over replicates
FDR: Bayesian false discovery rate

CXorf67 – highlighted in green
Components of PRC2 – highlighted in orange