Supplemental Data

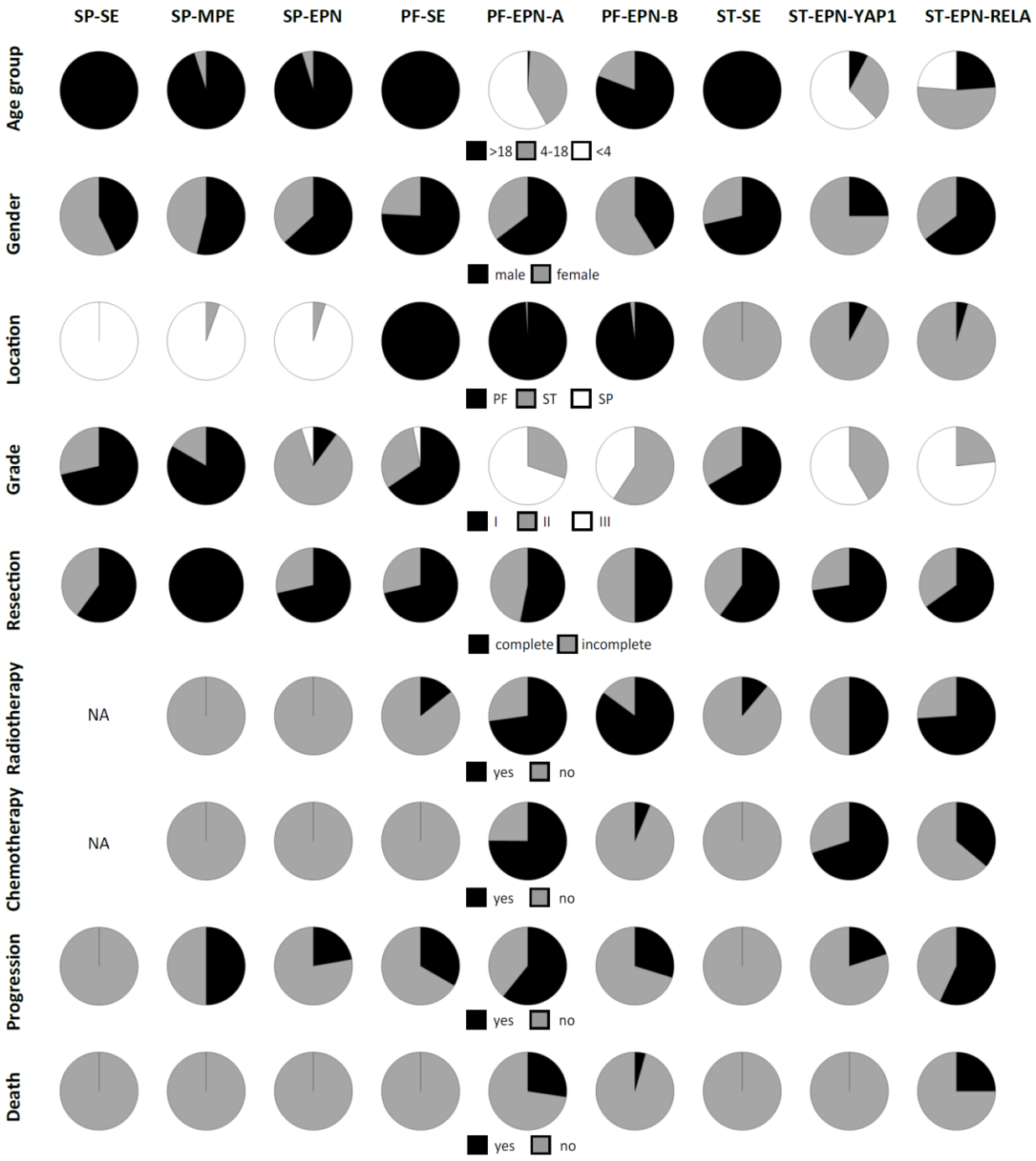
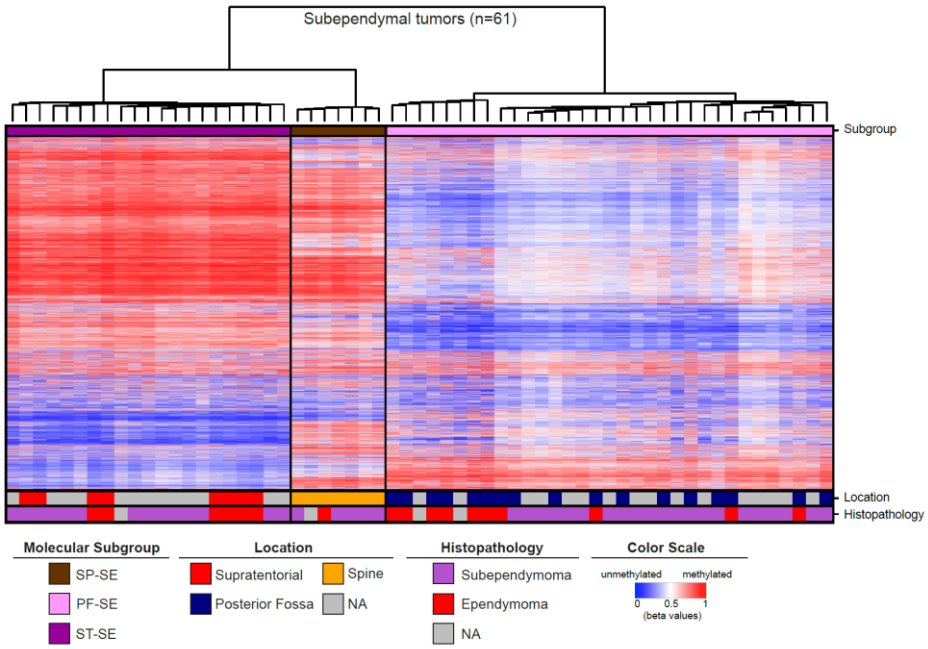


Figure S1, related to Table 1. Pie charts of clinical variables across nine molecular subgroups of ependymal tumors

A



B

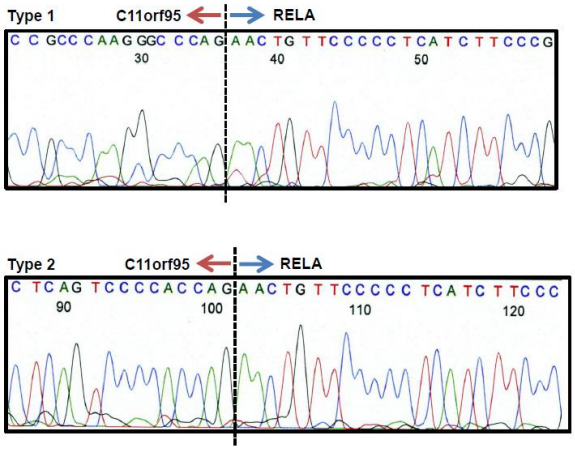


Figure S2, related to Figure 1. Molecular subgroups of ependymal tumors

(A) Separate hierarchical clustering of tumors belonging to the molecular subgroups annotated as SP-SE, PF-SE and ST-SE revealed stable and distinct clusters for all three subgroups. Histopathological diagnosis is indicated, demonstrating that all three molecular subgroups also comprise grade II ependymomas.

(B) Representative electropherograms of *RELA* fusion types 1 and 2 in ST-EPN-RELA cases.

Table S1, related to Figure 1. Validation of ST-EPN-RELA cases for the major *RELA* fusion types

Sample ID	Molecular subgroup	Fusion status[a]	RELA fusion type
1EP12	ST-EPN-RELA	positive	type1
1EP5	ST-EPN-RELA	positive	type 1, 2
2EP20	ST-EPN-RELA	positive[b]	type 1
3EP11	ST-EPN-RELA	positive	type 1
3EP13	ST-EPN-RELA	positive	type 1, 2
3EP19	ST-EPN-RELA	positive	type 2
3EP28	ST-EPN-RELA	positive	type 1
3EP30	ST-EPN-RELA	positive	RELA intron 2-3 at break point
3EP50	ST-EPN-RELA	positive	type 1
3EP54	ST-EPN-RELA	positive	type 1, 2
3EP58	ST-EPN-RELA	positive	type 1, 2
3EP61	ST-EPN-RELA	positive	type 1
3EP67	ST-EPN-RELA	positive	type 1, 2
3EP72	ST-EPN-RELA	positive	type 1
3EP8	ST-EPN-RELA	positive	type 1
4EP44	ST-EPN-RELA	negative	
4EP45	ST-EPN-RELA	positive	type 1
4EP47	ST-EPN-RELA	positive	type 1, 2
4EP49	ST-EPN-RELA	positive	type 1
4EP51	ST-EPN-RELA	positive	type 1
4EP53	ST-EPN-RELA	positive	type 1, 2
7EP1	ST-EPN-RELA	positive	type 1
7EP17	ST-EPN-RELA	negative	
7EP35	ST-EPN-RELA	positive	type 1
3EP36	ST-EPN-RELA	negative[b,c]	

[a]RT-PCR for most frequent fusion types; [b]additional RNA sequencing; [c]*PTEN* fused to *TAS2R1* leading to frame-shift and *PTEN* disruption

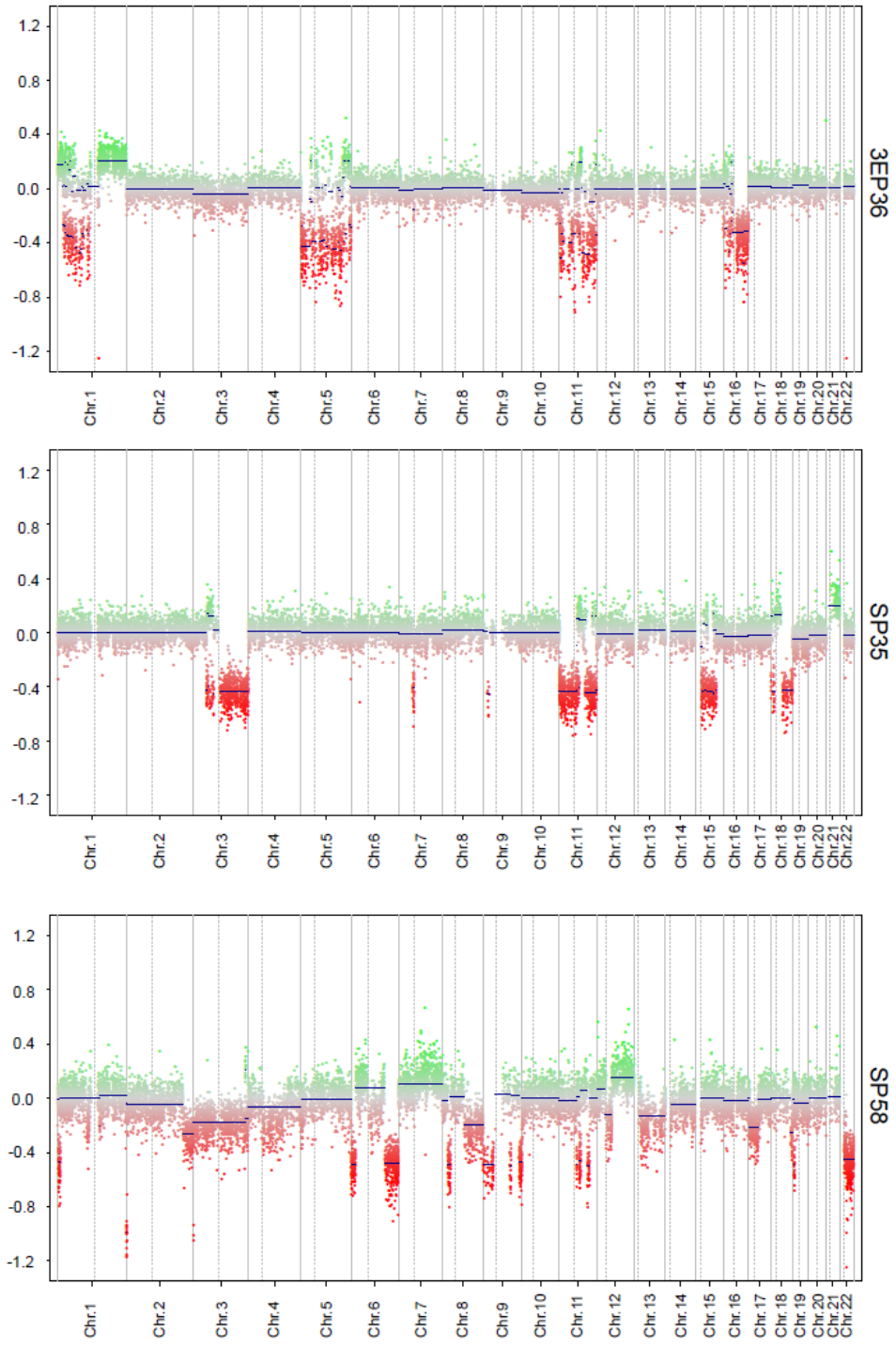


Figure S3, related to Figure 3. The ST-EPN-RELA subgroup reveals distinct copy-number aberrations

Representative DNA methylation array-based copy-number variation plots of ST-EPN-RELA cases revealing signs of chromosomal fragmentations (chromothripsis), in particular of chromosome 11.

Table S2, related to Figure 3. Provided as a separate Excel file. Copy-number aberrations of all nine molecular subgroups of ependymal tumors

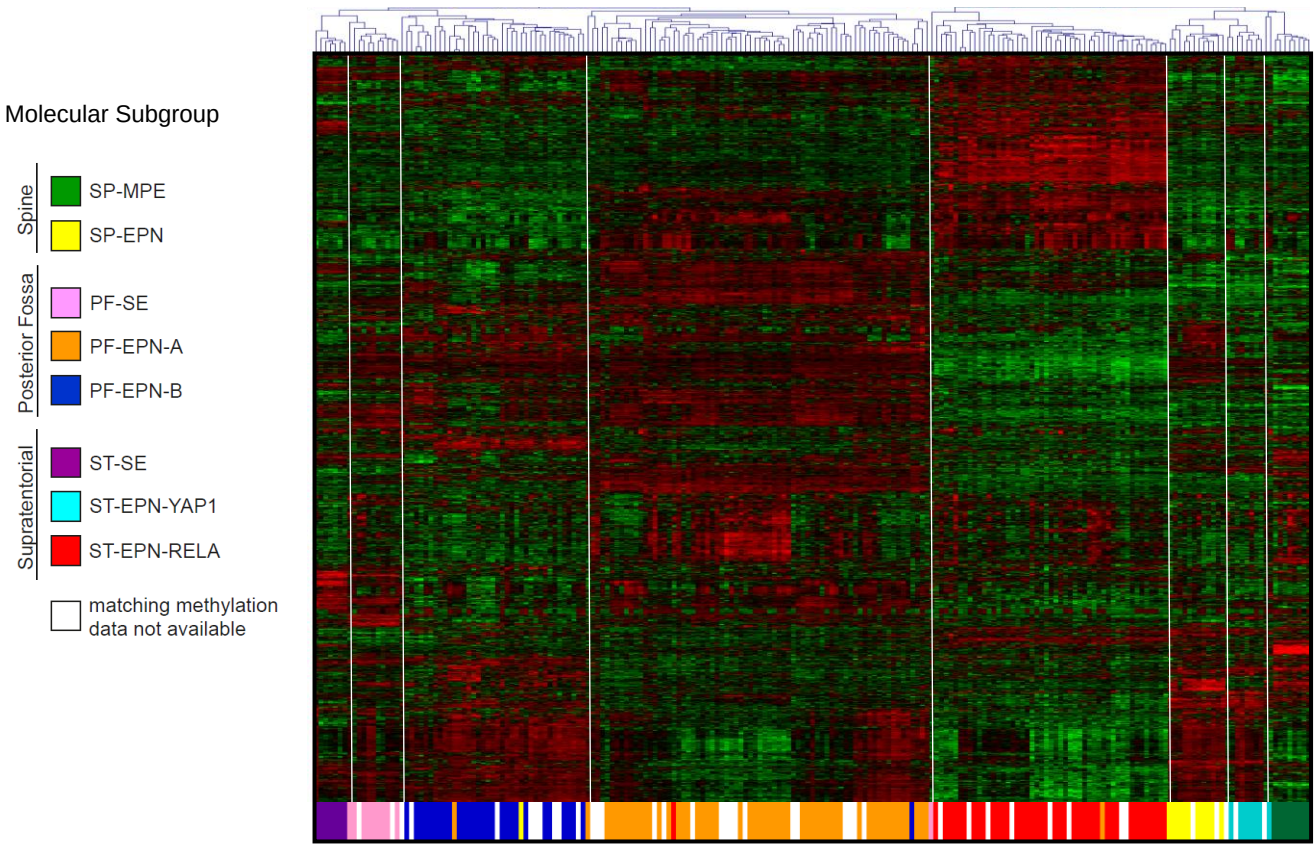
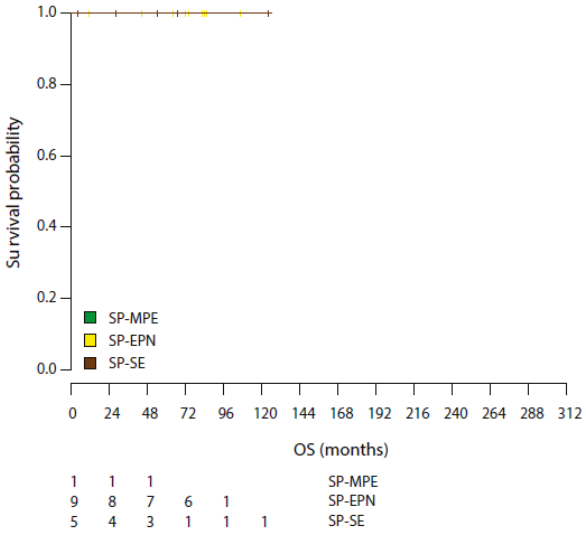


Figure S4, related to Figure 4. Subgrouping of ependymal tumors based on gene expression data recapitulates subgroups defined by DNA methylation profiling

Unsupervised hierarchical cluster analysis of gene expression data from 209 ependymal tumors. DNA methylation profiling based subgroup affiliation indicated by color coding was available for 168 cases.

Table S3, related to Figure 4. Provided as a separate Excel file. Subgroup associated signature genes and full list of subgroup-specific differentially expressed genes

A



B

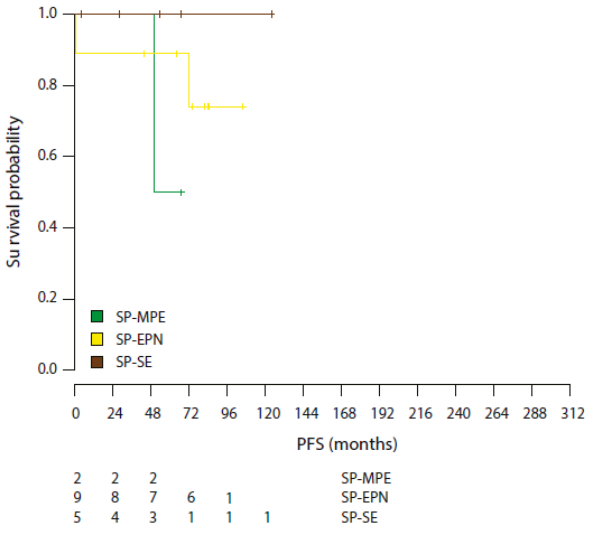
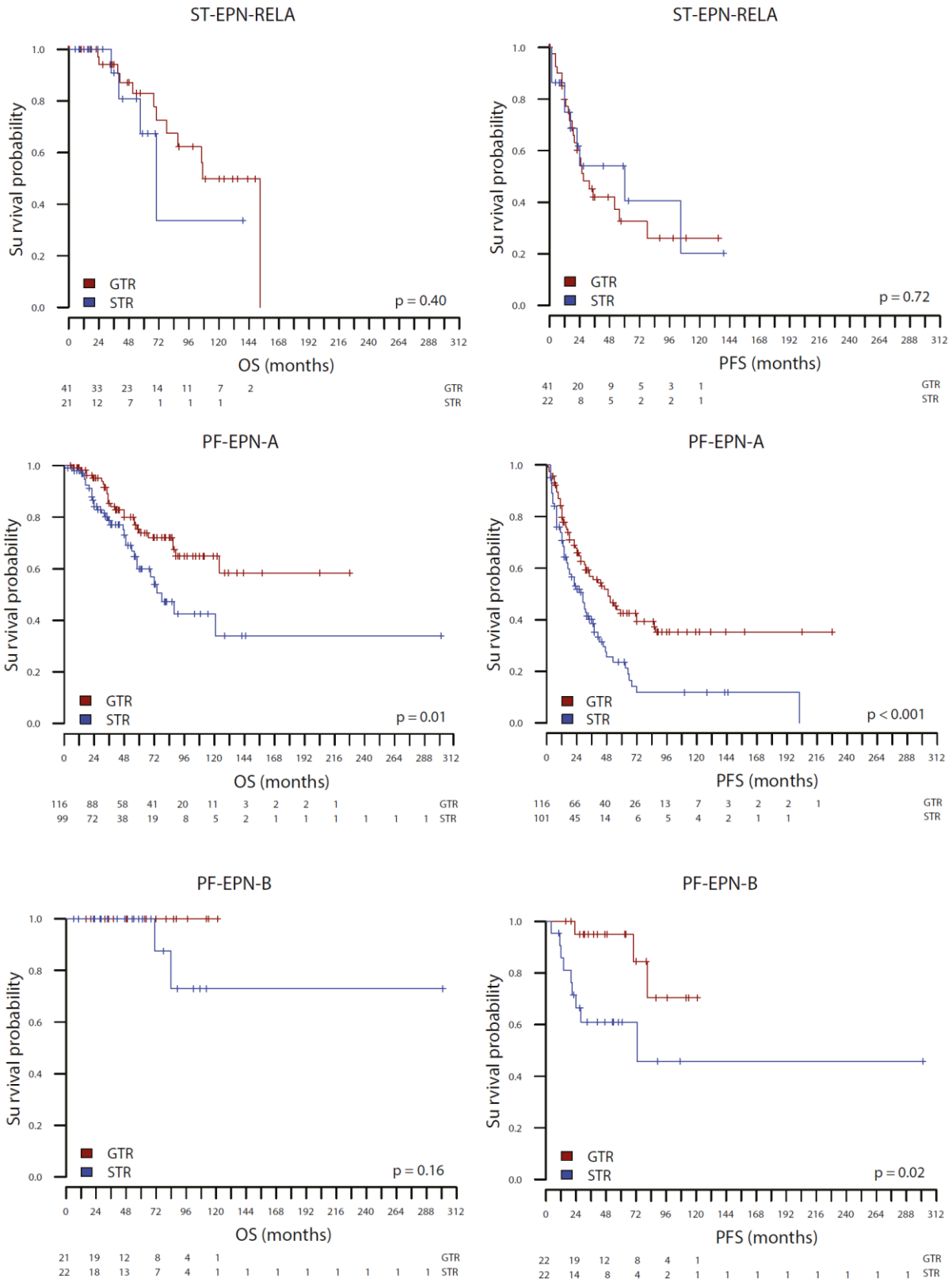


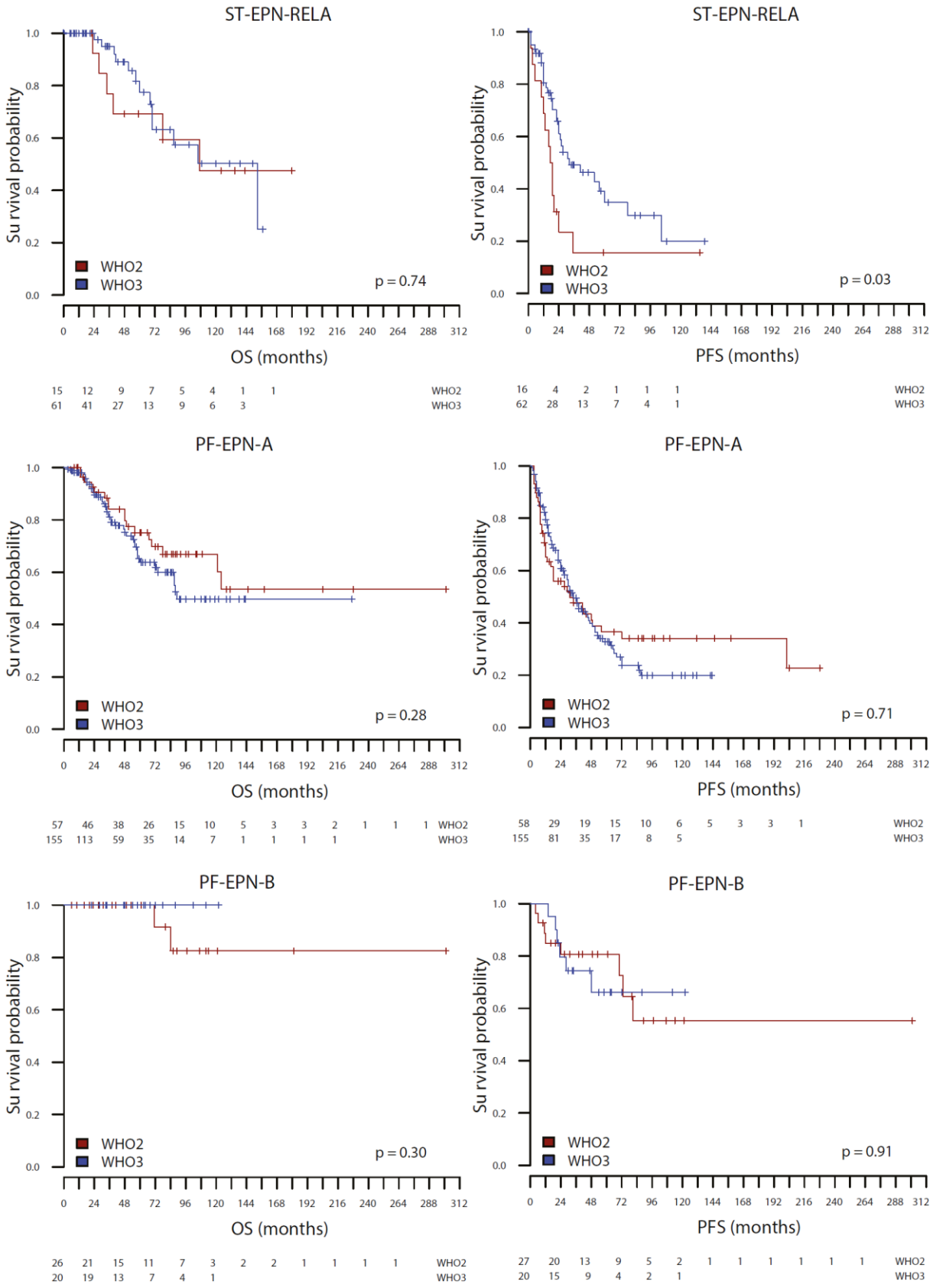
Figure S5, related to Figure 5. Molecular subgroups of ependymal tumors correlate with distinct clinical outcome

Kaplan-Meier curves for overall (A) and progression-free (B) survival for spinal molecular ependymal tumor subgroups defined by methylation profiling. Only significant p values were included.

A


ST-EPN-RELA
Su rvival probability
GTR
STR
p = 0.40
OS (months)
41 33 23 14 11 7 2 GTR
21 12 7 1 1 1 STR
ST-EPN-RELA
Su rvival probability
GTR
STR
p = 0.72
PFS (months)
41 20 9 5 3 1 GTR
22 8 5 2 2 1 STR
PF-EPN-A
Su rvival probability
GTR
STR
p = 0.01
OS (months)
116 88 58 41 20 11 3 2 2 1 GTR
99 72 38 19 8 5 2 1 1 1 1 1 1 STR
PF-EPN-A
Su rvival probability
GTR
STR
p < 0.001
PFS (months)
116 66 40 26 13 7 3 2 2 1 GTR
101 45 14 6 5 4 2 1 1 STR
PF-EPN-B
Su rvival probability
GTR
STR
p = 0.16
OS (months)
21 19 12 8 4 1 GTR
22 18 13 7 4 1 1 1 1 1 1 1 1 STR
PF-EPN-B
Su rvival probability
GTR
STR
p = 0.02
PFS (months)
22 19 12 8 4 1 GTR
22 14 8 4 2 1 1 1 1 1 1 1 1 STR

B


ST-EPN-RELA
Su rvival probability
WHO2
WHO3
p = 0.74
OS (months)
15 12 9 7 5 4 1 1 WHO2
61 41 27 13 9 6 3 WHO3
ST-EPN-RELA
Su rvival probability
WHO2
WHO3
p = 0.03
PFS (months)
16 4 2 1 1 1 WHO2
62 28 13 7 4 1 WHO3
PF-EPN-A
Su rvival probability
WHO2
WHO3
p = 0.28
OS (months)
57 46 38 26 15 10 5 3 3 2 1 1 1 WHO2
155 113 59 35 14 7 1 1 1 1 WHO3
PF-EPN-A
Su rvival probability
WHO2
WHO3
p = 0.71
PFS (months)
58 29 19 15 10 6 5 3 3 1 WHO2
155 81 35 17 8 5 WHO3
PF-EPN-B
Su rvival probability
WHO2
WHO3
p = 0.30
OS (months)
26 21 15 11 7 3 2 2 1 1 1 1 WHO2
20 19 13 7 4 1 WHO3
PF-EPN-B
Su rvival probability
WHO2
WHO3
p = 0.91
PFS (months)
27 20 13 9 5 2 1 1 1 1 1 1 WHO2
20 15 9 4 2 1 WHO3

C

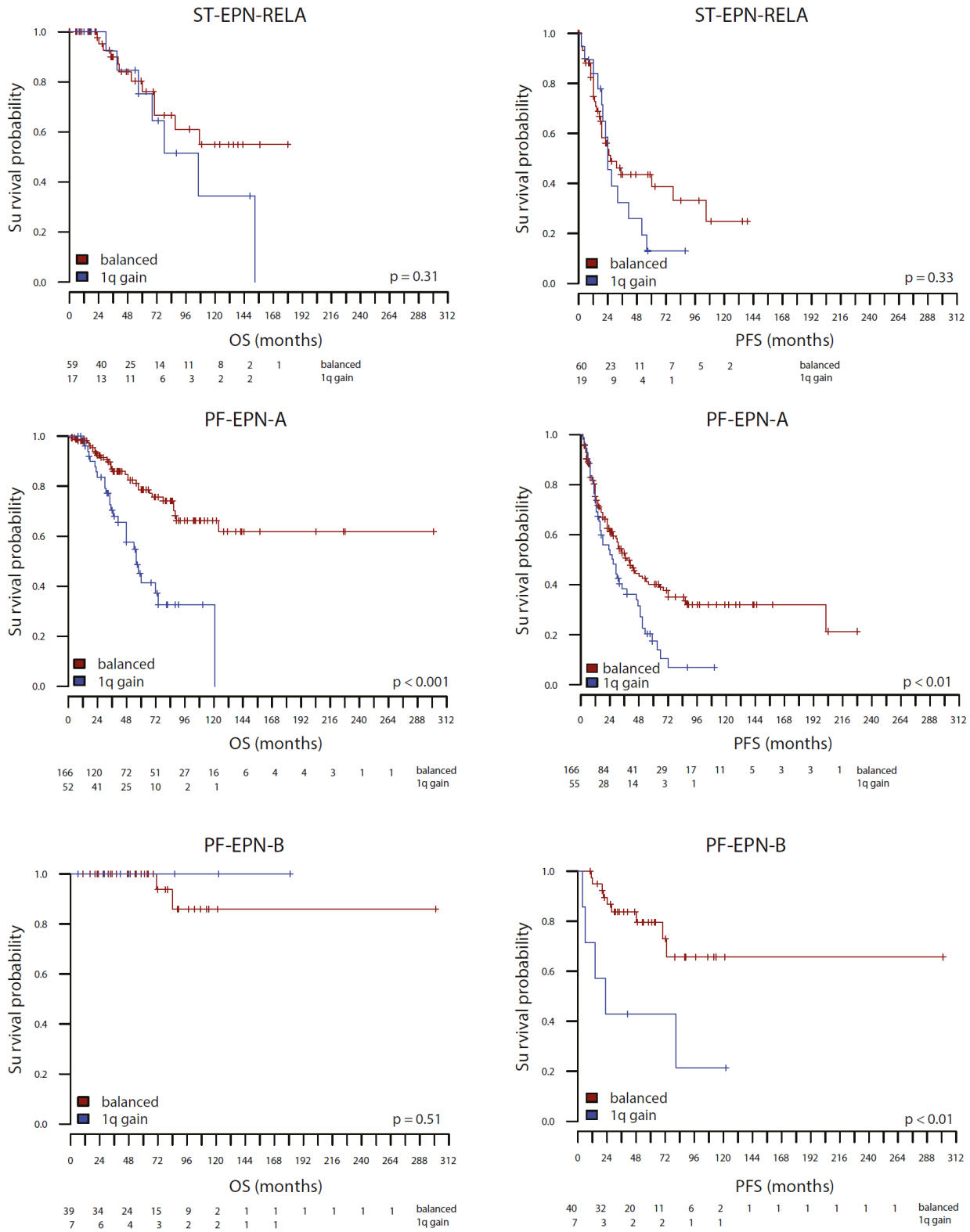

ST-EPN-RELA
Su rvival probability
balanced
1q gain
p = 0.31
OS (months)
59 40 25 14 11 8 2 1 balanced
17 13 11 6 3 2 2 1q gain
ST-EPN-RELA
balanced
1q gain
p = 0.33
PFS (months)
60 23 11 7 5 2 balanced
19 9 4 1 1q gain
PF-EPN-A
balanced
1q gain
p < 0.001
OS (months)
166 120 72 51 27 16 6 4 4 3 1 1 balanced
52 41 25 10 2 1 1q gain
PF-EPN-A
balanced
1q gain
p < 0.01
PFS (months)
166 84 41 29 17 11 5 3 3 1 balanced
55 28 14 3 1 1q gain
PF-EPN-B
balanced
1q gain
p = 0.51
OS (months)
39 34 24 15 9 2 1 1 1 1 1 1 balanced
7 6 4 3 2 2 1 1 1q gain
PF-EPN-B
balanced
1q gain
p < 0.01
PFS (months)
40 32 20 11 6 2 1 1 1 1 1 1 balanced
7 3 2 2 1 1 1q gain

Figure S6, related to Table 2. Subgroup-specific prognostic values in PF-EPN-A, PF-EPN-B and ST-EPN-RELA

Kaplan-Meier curves for overall and progression-free survival for gross-total and subtotal resection (A), for WHO grades II and III (B) and for status of chromosome 1q (C) with and without gain (= balanced). The p values were computed by log rank tests.

Table S4, related to Table 2. Univariate Cox proportional hazard models for overall and progression-free survival of molecular subgroups of ependymal tumors

Overall survival			
Variable	**Hazard ratio**	**95% CI**	**P value[b]**
1q gain (yes vs no)	3.24	2.06-5.12	<0.0001
WHO III vs WHO II	1.61	0.99-2.63	0.056
Age, years (<4 vs >18)	2.93	1.14-7.53	0.026
Age, years (4-18 vs >18)	3.98	1.57-10.09	0.004
Resection (STR vs GTR)	1.72	1.10-2.69	0.018
Chemotherapy (yes vs no)	2.31	1.42-3.77	<0.001
Radiotherapy (yes vs no)	1.32	0.77-2.26	0.317
PF-SE vs PF-EPN-B[a]	1.40	0.06-31.10	0.830
PF-EPN-A vs PF-EPN-B[a]	5.84	1.60-21.23	0.007
ST-SE vs PF-EPN-B[a]	1.74	0.08-38.75	0.725
ST-EPN-YAP vs PF-EPN-B[a]	1.41	0.06-31.42	0.829
ST-EPN-RELA vs PF-EPN-B[a]	5.36	1.38-20.83	0.015
Likelihood ratio test			
Full model vs model without methylation subgroups (OS)			0.04
Full model vs model without WHO (OS)			0.79
Progression-free survival			
Variable	**Hazard ratio**	**95% CI**	**P value[b]**
1q gain (yes vs no)	2.09	1.51-2.88	<0.0001
WHO III vs WHO II	1.32	0.95-1.82	0.096
Age, years (<4 vs >18)	3.02	1.79-5.12	<0.0001
Age, years (4-18 vs >18)	2.64	1.56-4.48	<0.001
Resection (STR vs GTR)	1.71	1.27-2.30	<0.001
Chemotherapy (yes vs no)	1.59	1.17-2.16	0.003
Radiotherapy (yes vs no)	1.06	0.75-0.14	0.737
PF-SE vs PF-EPN-B[a]	1.76	0.30-4.64	0.818
PF-EPN-A vs PF-EPN-B[a]	2.96	1.64-5.32	<0.0001
ST-SE vs PF-EPN-B[a]	0.30	0.18-5.17	0.406
ST-EPN-YAP vs PF-EPN-B[a]	0.75	0.14-4.17	0.747
ST-EPN-RELA vs PF-EPN-B[a]	2.90	1.51-5.60	<0.001
Likelihood ratio test			
Full model vs model without methylation subgroups (PFS)			0.01
Full model vs model without WHO (PFS)			0.56

CI, confidence interval; [a]with Firth's correction; [b]Wald test
OS, overall surviva; PFS, progression-free survival

Supplemental Experimental Procedures

DNA methylation analyses

All samples were analyzed on the Illumina Infinium HumanMethylation450 BeadChip arrays at the German Cancer Research Center (DKFZ) Genomics and Proteomics Core Facility according to the manufacturer's instructions and as previously described (Sturm et al., 2012). All samples were checked for expected and unexpected genotype matches by pairwise correlation of the 65 genotyping probes on the 450k array. Data analysis was performed using the open source statistical programming language R (R Core Team, 2014). Raw data files generated by the iScan array scanner were read and preprocessed using *minfi* Bioconductor package (Aryee et al., 2014). With the *minfi* package the same preprocessing steps as in Illumina's Genomestudio software were performed. In addition, to take possible batch effects due to different DNA extraction protocols of fresh frozen versus paraffin embedded material into account, a batch adjustment was performed. The batch effects were estimated by fitting a linear model to the $\log_2$ transformed intensity values of the methylated and unmethylated channel. After removing the component due to the batch effect the residuals were transformed back to intensity scale and methylation beta-values were calculated as described in Illumina's protocols. In addition, 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 = 24,536), and probes not mapping uniquely to the human reference genome (hg19) allowing for one mismatch (n = 9,993). In total, 438,370 probes were kept for analysis. For unsupervised hierarchical clusterings the 5,000, 10,000, 20,000 or 50,000 most variable methylated CpG probes measured by standard deviation were selected. Tumor samples were clustered by applying Ward's method and dissimilarity based on Euclidean distance. The reproducibility of the hierarchical clustering was validated by calculating p values by bootstrap hierarchical clustering using 1000 bootstrap steps as implemented in the R-package *pvclust* (Suzuki and Shimodaira, 2006).

Copy number variation analysis based on 450k Illumina arrays was performed as previously described (Sturm et al., 2012), with some improvements (Hovestadt *et al*., manuscript in preparation). In brief, the combined intensities of all available CpG probes were normalized by a unique combination of 87 control samples without known copy number alterations that most closely resembled the query sample using a linear regression approach (R function: lm). Probes were then combined into genomic bins of 50kb, and bins comprising of less than 15 probes were iteratively merged with the adjacent bin comprising of fewer probes. Subsequent steps were performed as previously described. Chromothripsis was called when massive fragmentations of a distinct chromosomal region resulting in simultaneous gains and losses rather than mutual exclusive events were observed (Figure S3) (Korbel and Campbell, 2013). DNA methylation data for the entire cohort has been deposited at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) and are accessible under accession number GSE65362.

Gene expression analyses

Tumor samples for which RNA of sufficient quantity and quality was available (n = 225) were analyzed on the Affymetrix GeneChip Human Genome U133 Plus 2.0 array at the Microarray Department of the University of Amsterdam, the Netherlands. Sample library preparation, hybridization, and quality control were performed according to protocols recommended by the manufacturer. The MAS5.0 algorithm of the GCOS program (Affymetrix Inc) was used for normalization of the expression data. Detection p values were assigned to each probe set using the MAS5.0 algorithm. Quality of the arrays was ensured by inspection of the beta-actin and GAPDH 5'-3' ratios as well as the percentage of present calls generated by MAS5.0. Data was analyzed using R2 (http://R2.amc.nl) or TMEV software (Saeed et al., 2003). Gene expression data have been deposited at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) and are accessible under accession number GSE64415.

Pathway enrichment analysis

Sub-group specific lists of differentially expressed genes were estimated with linear models using the Limma package in R (Smyth, 2004) after normalizing gene expression data with the VSN method (Huber et al., 2002). Each subgroup was compared to all other subgroups and a collection of normal brain control samples (Griesinger et al., 2013; Roth et al., 2006) using anatomical locations as cofactors (spinal cord, posterior fossa, supratentorial, control brain). Differential expression was computed with F-test and p values were corrected for multiple testing with the Bonferroni method ($q < 0.01$). Only up-regulated genes from each subgroup were selected for pathway analysis to reduce redundancy in gene lists. Significantly enriched pathways were computed with the g:Profiler software (Reimand et al., 2011), using ordered enrichment analysis on significance-ranked genes and custom filtering (5 - 1,000 genes in pathway, at least 3 differentially expressed genes in pathway, FDR corrected $q < 0.01$). Biological processes from Gene Ontology, pathways from the KEGG and Reactome databases, and protein complexes from the CORUM database were included in the analysis and other functional annotations were filtered. Pathways were visualized in the Cytoscape software with the Enrichment Map plugin (Merico et al., 2010).

RNA sequencing

Non-strand-specific sequencing libraries were prepared from a total of 55 samples using the mRNA-Seq Sample Prep Kit (Illumina) according to the manufacturer's instructions**.** Sequencing was carried out on the HiSeq 2000 platform using 2x51 cycles according to the manufacturer's instructions. Paired-end sequencing read data from transcriptome sequencing (fastq format) was used for *de novo* annotation of fusion transcripts using the SOAPfuse (Jia et al., 2013) and TopHat-Fusion (Kim and Salzberg, 2011) algorithms with gene annotations from the Ensembl database (v70).

Validation of gene fusions

C11orf95-RELA, *YAP1-MAMLD1* and *YAP1-FAM118B* fusions in tumor samples were confirmed by reverse-transcription PCR. RNA was extracted from snap-frozen tumor samples, and reverse-transcribed using SuperScript VILO (Life Technologies). PCR was undertaken with GoTaq® PCR Master Mix (Promega, Madison, WI) using primers 5'- GGGGGCTGAGGAGGAGGAG -3' (*C11orf95-RELA* type 1; forward), 5'- CCTGCACCTGGACGACAT - 3' (*C11orf95-RELA* type 2; forward) and 5'- TGTGGAGATCATTGAGCAGC - 3' (*C11orf95-RELA* type 1 and 2; reverse) to identify samples with the most common variants of *C11orf95-RELA* fusions (Parker et al., 2014). *YAP1* fusions were detected using primers 5'- CCAACCAGCAGCAACAGATG -3' (*YAP1-MAMLD1*; forward), 5'- GTGAGCCCATGATGTTGCC -3' (*YAP1-MAMLD1*; reverse), 5'- GAGTGCTCCAGTGAAACAGC -3' (*YAP1-FAM118B*; forward), 5'- GCCTGTTCCAATCACTAGCA -3' (*YAP1-FAM118B*; reverse). The PCR product was purified using the Qiagen MinElute PCR purification Kit (Qiagen, CA, USA). Direct sequencing of the purified PCR product was performed using an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA). Obtained sequence data was analyzed using the BLASTN software (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Supplemental References

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., and Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363-1369.

Griesinger, A. M., Birks, D. K., Donson, A. M., Amani, V., Hoffman, L. M., Waziri, A., Wang, M., Handler, M. H., and Foreman, N. K. (2013). Characterization of distinct immunophenotypes across pediatric brain tumor types. J Immunol *191*, 4880-4888.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics *18 Suppl 1*, S96-104.

Jia, W., Qiu, K., He, M., Song, P., Zhou, Q., Zhou, F., Yu, Y., Zhu, D., Nickerson, M. L., Wan, S.*, et al.* (2013). SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol *14*, R12.

Kim, D., and Salzberg, S. L. (2011). TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol *12*, R72.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G. D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One *5*, e13984.

Roth, R. B., Hevezi, P., Lee, J., Willhite, D., Lechner, S. M., Foster, A. C., and Zlotnik, A. (2006). Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics *7*, 67-80.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M.*, et al.* (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques *34*, 374-378.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540-1542.