METHODS

Patients and tumour samples. Tumour samples, clinical information, and animal studies were approved by local ethics institutional review boards (IRBs) from both the Heidelberg and Toronto institutions. Informed consent was obtained from all patients. No subject underwent chemotherapy or radiotherapy before the surgical removal of the primary tumour. In the sequencing cohort of tumour samples, at least 80% of tumour cell content was estimated by staining cryosections (~5 μm thick) of each sample with haematoxylin and eosin as described previously[2]. Diagnoses were confirmed by histopathologic assessment by at least two neuropathologists, including a central pathology review that used the 2007 World Health Organization classification for Central Nervous System tumours.

WES and WGS DNA library preparation and Illumina sequencing. Tumour and control samples were individually processed; in every case, thorough histological examination proved that each tumour consisted of over 80% tumour cells (in most cases >95%). DNA from tumour and control samples (blood) was prepared and sequenced individually. The Agilent SureSelect Human All Exon 50-Mb target enrichment kit (v3 initially, switched to v4 subsequently) was used to capture all human exons for deep sequencing, using the vendor's protocol v2.0.1. The SureSelect Human All Exon Kit targets regions of 50 Mb in total size, which is approximately 1.7% of the human genome. In brief, 3 μg genomic DNA was sheared with a Covaris S2 to a mean size of 150 bp. Five hundred nanograms of library DNA was hybridized for 24 h at 65 °C with the SureSelect baits. The captured fragments from the tumour samples and controls were sequenced in 105-bp single-end mode on an Illumina HiSeq2000 deep sequencing instrument (based on Illumina, Inc., v3 sequencing chemistry). The median coverage of whole-exome sequenced tumour samples was 157-fold (range 43–469-fold) and for control samples (blood DNA) 146-fold (range 80–222-fold). In addition, whole-genome libraries (before the exome hybridization step) were sequenced (three lanes each in paired-end 105-bp mode) on the HiSeq2000, as described[15].

To increase the coverage of the samples for whole-exome sequencing, we used the following strategy. Exome capture was initially carried out with Agilent SureSelect (Human All Exon 50 Mb) in-solution reagents using the default Illumina adapters (without barcode). To introduce Illumina Multiplex barcodes into the existing libraries at a later stage, 15 ng final exome-enriched library (without barcode) was used as a template in a 50-μl PCR reaction. The Herculase II Fusion enzyme (Agilent) was used together with the NEBNext Universal PCR primer for Illumina and NEBNext Index primer (NEB #E7335S) under the following conditions. The initial denaturation step for 2 min at 98 °C was followed by four cycles of 30 s 98 °C, 30 s 57 °C, 1 min 72 °C, and a final step of 10 min at 72 °C. Six or seven barcoded samples were then sequenced on the Hiseq2000 in 2 × 100-bp paired-end mode.

WGS and WES data processing. Fastq files were processed by the standardized alignment and variant-calling pipeline developed and applied in the context of the Pan-cancer Analysis of Whole Genomes (PCAWG) project (https://github.com/ICGC-TCGA-PanCancer)[18]. Here, we used the human genome assembly hs37d5 (https://ncbi.nlm.nih.gov/assembly/2758) as a reference genome and GENCODE19 (http://gencodegenes.org/releases/19.html) as gene annotations. Germline or somatic origin of the variants and indels was determined on the basis of their presence or absence in the matched control tissue.

RNA-seq data processing. Sequencing reads were aligned to the GRCh37 1000G reference using STAR 2.3.0[19] by reporting only reads with one best alignment (–outFilterMultimapNmax 1). Uniquely aligned reads were counted at gene regions using the package Subread v1.4.6 based on Gencode v19 annotations. Differential gene expression analysis between subgroups was performed using the R/Bioconductor package DESeq2 with contrast adjustment for multiple groups comparison. Fusion gene discovery was performed by the InFusion toolkit v.0.6.3[20].

Chromatin immunoprecipitation. ChIP of 5–10 mg flash-frozen primary ependymoma tumour was performed using 5 mg H3K27ac antibody per ChIP experiment (Abcam-AB4729 (Toronto) or Active Motif-39133 (Heidelberg)). Enriched DNA was quantified using Picogreen (Invitrogen) and ChIP libraries were amplified and barcoded using the Thruplex DNA-seq library preparation kit (Rubicon Genomics) according to the manufacturer's recommendations. Following library amplification, DNA fragments were agarose gel (1.0%) size-selected (<1 kb), assessed using Bioanalyzer (Agilent Technologies) and sequenced at The Centre for Applied Genomics (The Hospital for Sick Children) using Illumina Hi-Seq 2000 100-bp (Toronto cohort) and 50-bp (Heidelberg) single-end sequencing

ChIP–seq data pre-processing, enhancer and super enhancer analysis. Mapping of ChIP–seq data was performed as described[21]. Analogous to ref. 8, H3K27ac peak finding was performed using MACS1.4 with default parameter settings except with a *P*-value threshold of 1×10^{-9}. Peak finding for each ependymoma was performed separately, and as a control background for each H3K27ac ChIP–seq sample, its matched genomic DNA was used where available. Peaks that could not be identified in at least two primary ependymomas and peaks contained completely within the region surrounding ±2.5 kb of transcriptional start sites were excluded from any further analysis. Afterwards, the H3K27ac peaks of the individual samples were merged into a single set of (non-overlapping) peaks. When comparing against the Roadmap Epigenomics Dataset, reads from ependymoma samples were trimmed to 36 bp to be consistent with processed Roadmap Epigenomics Data, and then pre-processed as described above. To reduce potential batch effects, enhancer H3K27 acetylation profiles were quantile-normalized using the preprocessCore package in R. Super enhancers were identified using the rank ordering of super enhancers (ROSE) algorithm, which classified as a super enhancer any set of two or more H3K27ac peaks (detected by MACS1.4, $P < 1^{-9}$) within a 12.5-kb distance, and further than 2.5 kb from a transcriptional start site. Super enhancers were further defined by those demonstrating the greatest levels of H3K27 acetylation as detected by graphing an inflection plot and selecting values for which the slope of a fitted curve exceeded a value of 1. In the case of tumour-specific super enhancers, all regions were removed that contained any overlap with a super enhancer detected in at least one normal brain region consisting of: anterior caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, mid frontal lobe, and substantia nigra.

***t*-SNE analysis of Illumina DNA methylation and enhancer data.** All DNA methylation analyses were performed in R v3.3.0 (R Development Core Team, 2015). Raw signal intensities were obtained from IDAT-files using minfi Bioconductor v1.18.2. Each sample was individually normalized by performing a background correction (shifting the 5th percentile of negative control probe intensities to 0) and a dye-bias correction (scaling the mean of normalization control probe intensities to 10,000) for both colour channels. No further normalization or transformation steps were performed, and standard beta-values were used for downstream methylation analyses. The following criteria were applied to filter out probes prone to yield inaccurate methylation levels: removal of probes targeting the X and Y chromosomes ($n = 11{,}551$), removal of probes that overlap common SNPs (dbSNP132 Common) within the CpG or the following base ($n = 7{,}998$), and removal of probes not mapping uniquely to the human reference genome (hg19) ($n = 3{,}965$). To enable comparability with the Illumina Infinium HumanMethylationEPIC array, we also removed probes not represented on this array ($n = 32{,}260$). In total, 428,799 probes were kept for analysis. For unsupervised hierarchical clustering, we selected the 10,000 most variably methylated probes across the dataset (s.d. > 0.264). Distance between samples was calculated by using 1-Pearson correlation coefficient as the distance measure. The resulting distance matrix was used to perform *t*-SNE analysis with Rtsne package v0.11. The following non-default parameters were used: theta = 0, is_distance = T, pca = F, max_iter = 10000.

For clustering of H3K27ac ChIP–seq data from the Heidelberg and Toronto cohorts together, we processed both cohorts in single-end mode without background using the R/Bioconductor package QSEA v.0.0.11. For each sample, we quantified sequencing reads as reads per kilobase per million (RPKM) at previously derived enhancers, neglecting enhancers at mitochondrial and sex chromosomes. Distance between samples was calculated by using 1-Spearman correlation coefficient as the distance measure. The resulting distance matrix was used to perform the *t*-SNE analysis (Rtsne package v0.11). The following non-default parameters were used: theta = 0, is_distance = T, pca = F, max_iter = 5000.

Unsupervised hierarchical clustering analysis of variant enhancer loci. A matrix of the normalized H3K27ac density was generated in HOMER (v3.12) based on the identified consensus typical enhancers. Variant enhancer loci (VELs) were defined as enhancers, which exhibited the greatest median absolute deviation (MAD) across all samples used for clustering. In the case of unsupervised hierarchical clustering between ependymoma, Roadmap Epigenomics, and ENCODE samples, the top 10,000 VELs were retained. These enhancers were used for unsupervised hierarchical clustering using a Pearson correlation as a distance metric. In the case of super enhancers, a matrix was generated in HOMER using the consensus super enhancer BED files of normalized H3K27ac densities across all samples. Non-negative matrix factorization was performed using all super enhancer regions, using the methodology described previously, with 20 iterations, across 10 rank classifications[2].

Identification of super-enhancer-associated pathways and drug–gene interactions. Differential super-enhancer-associated genes in ependymomas or ependymoma subgroups were imported into G-Profiler[22] for pathway analysis, restricted to GO, KEGG and REACTOME gene sets. Cytoscape (v3.2.1) and the EnrichmentMap plug-in was used to generate networks for genesets enriched with an FDR cut-off of <0.05. super-enhancer-associated genes were also used to query the Washington University Drug Gene Interaction database, restricted to expert-curated drug–target interactions to identify novel and druggable gene targets[17].