

Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling

Stephen C. Mack^{1,2,3,4*}, Kristian W. Pajtker^{5,6,7*}, Lukas Chavez^{5,6,8*}, Konstantin Okonechnikov^{5,6}, Kelsey C. Bertrand^{1,2,9}, Xiuxing Wang^{3,4,10}, Serap Erkek^{5,6,11}, Alexander Federation¹², Anne Song^{3,4}, Christine Lee^{3,4}, Xin Wang¹³, Laura McDonald¹³, James J. Morrow¹⁴, Alina Saiakhova¹⁴, Patrick Sin-Chan¹³, Qiulian Wu^{3,4,10}, Kulandaimanuvél Antony Michaelraj¹³, Tyler E. Miller^{3,4,15}, Christopher G. Hubert^{3,4}, Marina Ryzhova¹⁶, Livia Garzia¹³, Laura Donovan¹³, Stephen Dombrowski^{3,4,17}, Daniel C. Factor¹⁴, Betty Luu¹³, Claudia L. L. Valentim^{3,4}, Ryan C. Gimple^{3,4,10,15}, Andrew Morton^{3,4,14}, Leo Kim^{3,4,10}, Briana C. Prager^{3,4,10}, John J. Y. Lee¹³, Xiaochong Wu¹³, Jennifer Zuccaro¹³, Yuan Thompson¹³, Borja L. Holgado¹³, Jüri Reimand^{18,19}, Susan Q. Ke^{3,4}, Adam Tropper^{3,4}, Sisi Lai^{3,4}, Senthuran Vijayarajah^{9,20}, Sylvia Doan²¹, Vaidehi Mahadev^{3,4}, Ana Fernandez Miñan²², Susanne N. Gröbner^{5,6}, Matthias Lienhard²³, Marc Zapatka²⁴, Zhiqin Huang²⁴, Kenneth D. Aldape²⁵, Angel M. Carcaboso²⁶, Peter J. Houghton²⁷, Stephen T. Keir²⁸, Till Milde^{5,7,29}, Hendrik Witt^{5,6,7}, Yan Li¹⁴, Chao-Jun Li³⁰, Xiu-Wu Bian³¹, David T. W. Jones^{5,6}, Ian Scott¹³, Sheila K. Singh³², Annie Huang^{13,33}, Peter B. Dirks¹³, Eric Bouffet^{13,33}, James E. Bradner³³, Vijay Ramaswamy^{13,34}, Nada Jabado³⁵, James T. Rutka¹³, Paul A. Northcott³⁶, Mathieu Lupien¹⁹, Peter Lichter²⁴, Andy Korshunov^{37,38}, Peter C. Scacheri¹⁴, Stefan M. Pfister^{5,6,7}, Marcel Kool^{5,6,7}, Michael D. Taylor^{13,§} & Jeremy N. Rich^{3,4,10,17,§}

Genomic sequencing has driven precision-based oncology therapy; however, the genetic drivers of many malignancies remain unknown or non-targetable, so alternative approaches to the identification of therapeutic leads are necessary. Ependymomas are chemotherapy-resistant brain tumours, which, despite genomic sequencing, lack effective molecular targets. Intracranial ependymomas are segregated on the basis of anatomical location (supratentorial region or posterior fossa) and further divided into distinct molecular subgroups that reflect differences in the age of onset, gender predominance and response to therapy^{1–3}. The most common and aggressive subgroup, posterior fossa ependymoma group A (PF-EPN-A), occurs in young children and appears to lack recurrent somatic mutations². Conversely, posterior fossa ependymoma group B (PF-EPN-B) tumours display frequent large-scale copy number gains and losses but have favourable clinical outcomes^{1,3}. More than 70% of supratentorial ependymomas are defined by highly recurrent gene fusions in the NF- κ B subunit gene *RELA* (ST-EPN-RELA), and a smaller number involve fusion of the gene encoding the transcriptional activator *YAP1* (ST-EPN-YAP1)^{1,3,4}.

Subependymomas, a distinct histologic variant, can also be found within the supratentorial and posterior fossa compartments, and account for the majority of tumours in the molecular subgroups ST-EPN-SE and PF-EPN-SE. Here we describe mapping of active chromatin landscapes in 42 primary ependymomas in two non-overlapping primary ependymoma cohorts, with the goal of identifying essential super-enhancer-associated genes on which tumour cells depend. Enhancer regions revealed putative oncogenes, molecular targets and pathways; inhibition of these targets with small molecule inhibitors or short hairpin RNA diminished the proliferation of patient-derived neurospheres and increased survival in mouse models of ependymomas. Through profiling of transcriptional enhancers, our study provides a framework for target and drug discovery in other cancers that lack known genetic drivers and are therefore difficult to treat.

To pinpoint genes that depend on enhancers for their role in tumour formation, we characterized regions of actively transcribed chromatin in 42 primary intracranial ependymomas using histone 3 lysine 27 acetylation chromatin immunoprecipitation and sequencing (H3K27ac

¹Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA. ²Department of Pediatric Hematology and Oncology, Texas Children's Cancer and Hematology Centers, Houston, Texas, USA. ³Department of Stem Cell Biology and Regenerative Medicine, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. ⁴Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio 44195, USA. ⁵Hopp Children's Cancer Center at the NCT Heidelberg (KITZ), 69120 Heidelberg, Germany. ⁶Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany and German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. ⁷Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University Hospital, 69120 Heidelberg, Germany. ⁸Department of Medicine, Division of Medical Genetics, University of California – San Diego School of Medicine, La Jolla, California 92093, USA. ⁹Department of Pediatrics, Cleveland Clinic, Cleveland, Ohio 44195, USA. ¹⁰Department of Medicine, Division of Regenerative Medicine, University of California – San Diego School of Medicine, La Jolla, California, USA. ¹¹European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany. ¹²Department of Genomic Sciences, University of Washington, Seattle, Washington 98195, USA. ¹³Division of Neurosurgery, Program in Developmental and Stem Cell Biology, Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁴Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA. ¹⁵Department of Pathology, Case Western Reserve University, Cleveland, Ohio 44106, USA. ¹⁶Department of Neuropathology, NN Burdenko Neurosurgical Institute, 4th Tverskaya-Yamskaya 16, Moscow 125047, Russia. ¹⁷Rose Ella Burkhardt Brain Tumor & Neuro-Oncology Center, Cleveland Clinic Neurological Institute, Department of Neurosurgery, Cleveland Clinic, Cleveland, Ohio 44195, USA. ¹⁸Computational Biology Program, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ¹⁹Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ²⁰Department of Pediatrics, Division of Critical Care, University of Utah School of Medicine, Salt Lake City, Utah, USA. ²¹Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA. ²²Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas and Universidad Pablo de Olavide, Sevilla, Spain. ²³Department of Computational Molecular Biology, Max-Planck-Institute for Molecular Genetics, 14195 Berlin, Germany. ²⁴Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69121 Heidelberg, Germany. ²⁵Department of Pathology, University Health Network, Toronto, Ontario M5G 1L7, Canada. ²⁶Preclinical Therapeutics and Drug Delivery Research Program, Fundació Sant Joan de Deu, 08950 Barcelona, Spain. ²⁷Nationwide Children's Hospital, Center for Childhood Cancer and Blood Diseases, Columbus, Ohio. ²⁸Duke University Medical Center, Department of Surgery, Durham, North Carolina, USA. ²⁹Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), INF 280, D-69120 Heidelberg, Germany. ³⁰MOE Key Laboratory of Model Animal for Disease Study, Model Animal Research Center, Nanjing University, National Resource Centre for Mutant Mice, Nanjing, China. ³¹Institute of Pathology and Southwest Cancer Center, Southwest Hospital, The Third Military Medical University, and The Key Laboratory of Tumor Immunopathology, The Ministry of Education of China, Chongqing, China. ³²Department of Surgery, Division of Neurosurgery, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada. ³³Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. ³⁴Division of Hematology and Oncology, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ³⁵Departments of Paediatrics and Human Genetics, McGill University and the McGill University Health Centre Research Institute, Montreal, Quebec H3Z 2Z3, Canada. ³⁶Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. ³⁷Department of Neuropathology, University of Heidelberg, 69120 Heidelberg, Germany. ³⁸Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), 69121 Heidelberg, Germany.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

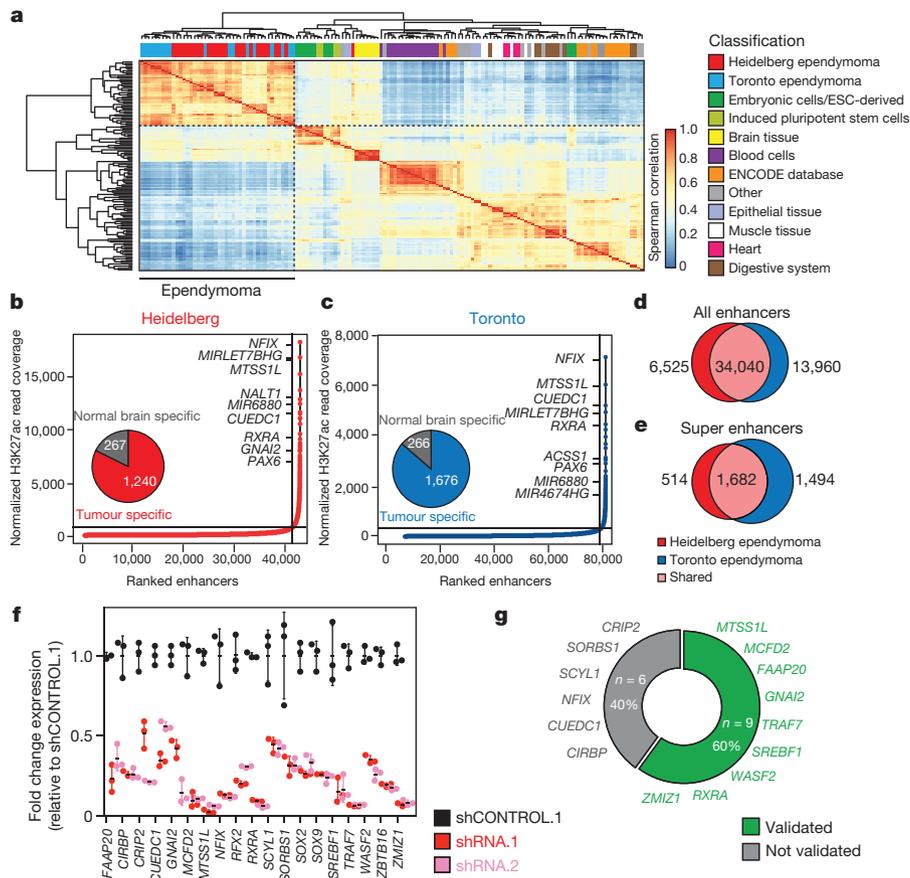


Figure 1 | H3K27ac profiles define active regulatory elements of ependymoma. **a**, Unsupervised hierarchical clustering of the top 10,000 variant enhancer loci detected in ependymomas compared to the Roadmap Epigenomics Consortium samples; $n = 143$ independent samples. **b**, **c**, Inflection plot indicating identified ependymoma super enhancers. **d**, **e**, Venn diagrams depicting the number of shared enhancers (**d**) and super enhancers (**e**) between the Heidelberg ($n = 24$) and Toronto

ChIP-seq), a histone mark of active chromatin, on two independent cohorts of fresh-frozen primary ependymoma specimens in two different facilities ('Heidelberg' and 'Toronto'), each with a different H3K27 acetylation-specific antibody. Our analysis focused on the intersection of shared enhancers between these two datasets, integrated with whole-exome sequencing (WES), whole-genome sequencing (WGS), RNA sequencing (RNA-seq), DNA copy-number analysis, and DNA methylation profiling (Extended Data Figs 1, 2; Supplementary Tables 1–7). 'Active' typical enhancers were defined as significant H3K27ac peaks more than 2.5 kb from the nearest transcriptional start site. To perform unsupervised hierarchical clustering, the top 10,000 variant enhancer loci from both cohorts were compared to the Roadmap Epigenomics and ENCODE databases⁵ (Fig. 1a, Extended Data Figs 3, 4). Ependymoma enhancer profiles were distinct from those of other tissue types, marked by acquisition and loss of hundreds of enhancer loci (Extended Data Fig. 4). Consistent with prior literature, super enhancer domains were substantially associated with greater transcriptional load^{6–9} (Extended Data Fig. 4). We identified 2,196 and 3,176 super enhancers in the Heidelberg and Toronto cohorts, respectively, and both cohorts shared a large proportion of super enhancer regions (Fig. 1b–e, Supplementary Tables 8–10, Extended Data Fig. 4). The vast majority of super enhancers were tumour-specific and enriched with cancer-associated genes reported in other solid cancers, including *PAX6*, *SKI*, *FGFRL1*, *FGFR1*, and *BOC* (Fig. 1b, c, Supplementary Table 10, Extended Data Fig. 4). Several of these genes, such as *EPHB2* and *CCND1*, have been previously validated as ependymoma oncogenes^{10–12} (Extended Data Fig. 5).

($n = 18$) independent ependymoma sample cohorts. **f**, Quantitative reverse transcription PCR showing knockdown efficiency of 15 ependymoma super-enhancer-associated genes ($n = 3$ technical replicates, error bars show s.d. Results were reproduced in independent biological duplicates). **g**, Percentage of top ependymoma super enhancer genes that demonstrate greater than 50% decrease in viability over seven days. Cell survival from knockdown of each gene was assayed and independently replicated as biological triplicates.

To determine whether super enhancers reveal pathways and genes on which ependymoma cells depend, and which could be actionable by targeted therapy, the 15 top-ranking ependymoma super enhancer genes were validated in a series of 60 RNA interference short hairpin RNA (shRNA) knockdown time-course studies to demonstrate the feasibility of our approach to uncover novel cancer targets (Extended Data Fig. 6). Following transduction of ST-EPN-RELA patient-derived (EPI-NS) cells with shRNA constructs, the two most effective and specific shRNA constructs per gene were functionally validated (Fig. 1f). Globally, depletion of the top-ranking tumour-specific super enhancer genes impaired cell growth to varying degrees over seven days, compared to non-targeting shRNA controls (Extended Data Fig. 7). Using a stringent cut-off of shRNA-mediated growth inhibition by two independent shRNA constructs (shRNA.1 and shRNA.2) of at least 50% decrease in cell viability over seven days, a majority (60%) of ependymoma super enhancer genes were required for cellular maintenance, supporting super enhancer mapping as a viable approach for therapeutic target identification (Fig. 1g).

We next investigated whether the differences in enhancer landscapes between molecular subgroups of ependymoma reflect transcriptional differences. In both cohorts, unsupervised hierarchical clustering of all enhancers demonstrated an unbiased segregation of ependymoma molecular subgroups (Fig. 2a–d, Extended Data Fig. 5). Molecular differences between ependymoma subgroups were supported by robust segregation at the DNA methylation level (Fig. 2c). Subgroup-specific typical enhancers were enriched within large H3K27 acetylated domains (that is, super enhancers), and confirmed

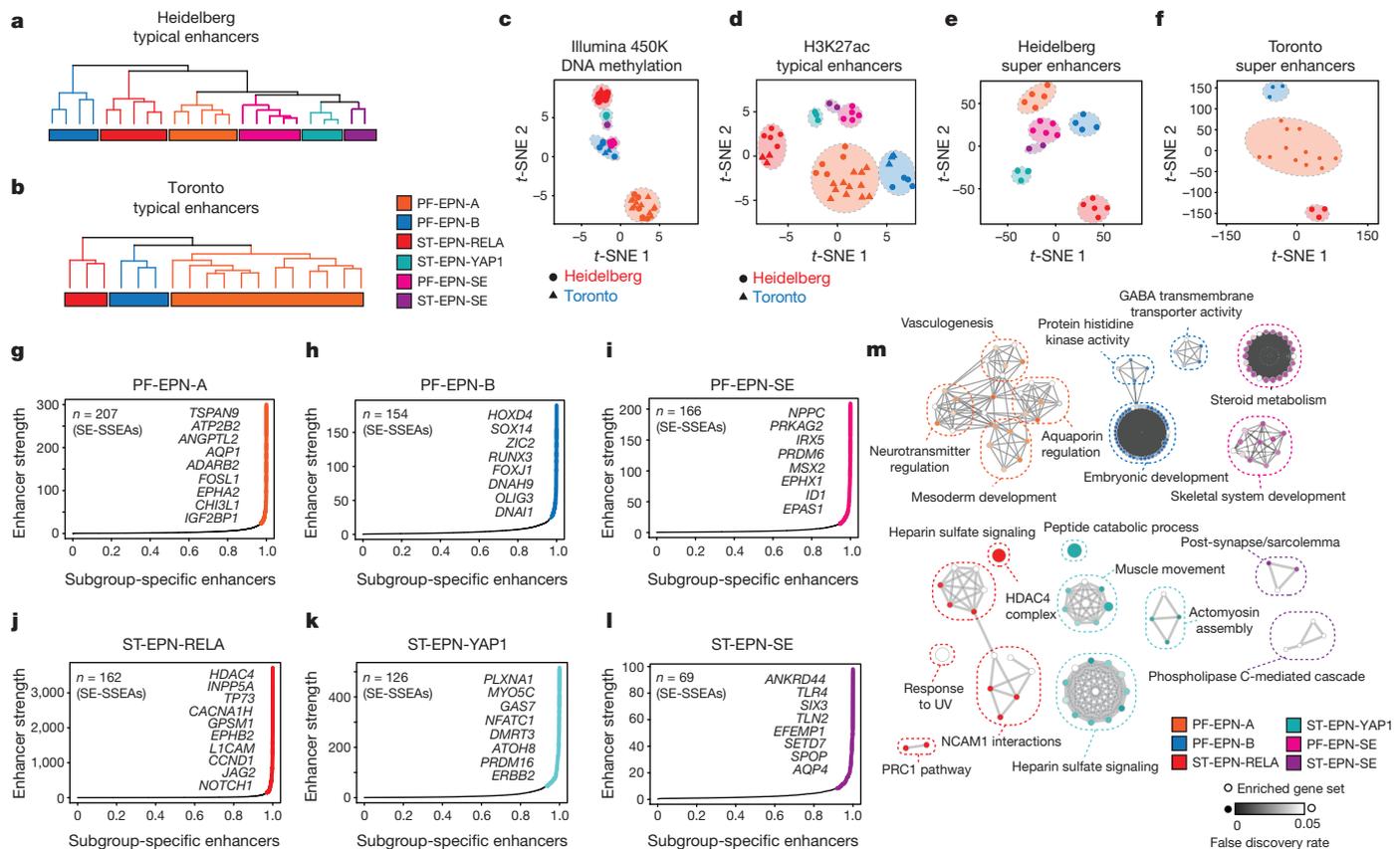


Figure 2 | Active enhancers delineate subgroups of ependymoma.

a, b, Unsupervised hierarchical clustering of all H3K27ac enhancer loci in Heidelberg ($n = 24$) and Toronto ($n = 18$) independent sample cohorts. **c**, Combined t -distributed stochastic neighbour embedding (t -SNE) analysis of the top 10,000 variably methylated Illumina 450K CpG probes. **d**, Combined t -SNE analysis of all enhancer loci. $n = 43$ independent samples. **e, f**, t -SNE analysis of the H3K27ac marked super enhancer

regions in ependymoma. $n = 42$ independent samples. **g, i**, Inflection plot indicating super enhancers with subgroup-specific enhancer activity (SE-SSEA) in ependymomas. $n = 24$ independent samples. **m**, G-Profiler pathway analysis of ependymoma subgroup super-enhancer-associated genes with significant enrichment indicated as the false discovery rate (FDR)-corrected P value. $n = 24$ independent samples.

by unsupervised segregation of ependymoma subgroups using super enhancer regions (Fig. 2e, f, Extended Data Fig. 5). We termed this distinct class of super enhancers with subgroup-specific enhancer activity SE-SSEAs, and similarly typical enhancers with subgroup-specific activity TE-SSEAs. Over 86% of SE-SSEAs observed in the Heidelberg cohort were confirmed by the Toronto cohort as active super enhancers in the respective subgroup (Extended Data Fig. 5), thus uncovering a distinct subset of super enhancers that were most common in the PF-EPN-A, PF-EPN-B and ST-EPN-RELA subgroups of ependymoma (Fig. 2g–l, Extended Data Fig. 5, Supplementary Tables 11–16). Owing to the low prevalence of ST-EPN-YAP1, ST-EPN-SE, and PF-EPN-SE tumours, these tumours were not represented in the Toronto cohort, and further downstream analysis was based on the Heidelberg cohort alone (Fig. 2g–l, Extended Data Fig. 5, Supplementary Tables 11–16). SE-SSEA genes were associated with subgroup-specific gene expression, further supporting the role of super enhancers as important contributors to transcriptional output (Extended Data Fig. 5, Supplementary Tables 11–16). SE-SSEA genes also converged on a subset of signalling pathways that distinguished the molecular subgroups of ependymoma, such as the polycomb repressive complex 1 (PRC1) and histone deacetylase (HDAC4) pathways in ST-EPN-RELA tumours, both of which can be inhibited by small molecules (Fig. 2m, Extended Data Fig. 5, Supplementary Table 17).

To translate identified SE-SSEA genes in subgroups of ependymoma into novel therapeutic leads, we first focused on ST-EPN-RELA tumours, where we observed an SE-SSEA proximal to *CACNA1H* and associated with its subgroup-restricted gene expression (Extended Data Fig. 8). CRISPR–dCas9–KRAB mediated repression of active

constituent enhancers within the *CACNA1H* super enhancer resulted in downregulation of *CACNA1H* gene expression (Extended Data Fig. 8). Compared to a PF-EPN-A primary culture (S15-NS), cell proliferation of an ST-EPN-RELA patient-derived primary culture model (EPI-NS) was specifically impaired by shRNA-mediated knockdown of *CACNA1H* or pharmacologic blockade of its activity using the calcium channel inhibitor mibefradil (Extended Data Fig. 8). In a similar fashion, we found the super-enhancer-regulated gene *IGF2BP1* preferentially in a subset of PF-EPN-A tumours. shRNA-mediated targeting of *IGF2BP1* in PF-EPN-A ependymoma cultures, but not ST-EPN-RELA primary cultures, impaired cell proliferation, implicating *IGF2BP1* as a potential cancer dependency gene in PF-EPN-A ependymomas (Extended Data Fig. 8). Our findings thus identify candidate oncogenes that are associated with super enhancers as well as novel pathways specific to subgroups of ependymoma.

The regulation of cell-type-specific gene expression is often dominated by only a small number of core transcription factors out of the hundreds expressed within a given cell type¹³. As many important transcription factor motifs, such as FOSL1, FOSL2, SOX9, RFX2, and SOX2, were enriched across shared enhancers of ependymoma (Fig. 3a, Supplementary Table 18), we sought to identify the principal transcription factors of ependymoma that govern ependymoma cell identity across subgroups using core regulatory circuitry analysis^{8,14} (Fig. 3b, Extended Data Fig. 9, Supplementary Table 19). A small set of highly active transcription factors was identified, including SOX9, RFX2, SOX2, ZBTB16, HES1, NFIA, and NFIB, which were highly expressed in ependymoma compared to a large collection of normal brain tissues (Fig. 3b, Extended Data Fig. 9). By contrast, transcription

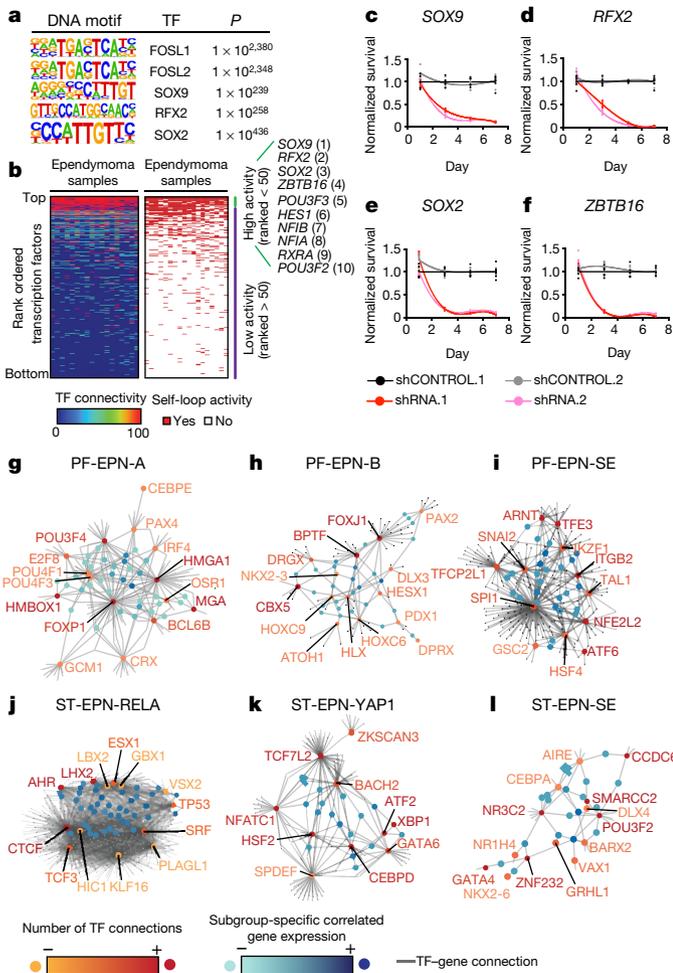


Figure 3 | Transcription factor circuitries of ependymoma. **a**, DNA motifs enriched within shared ependymoma-typical enhancers that overlay with ATAC-seq peaks derived from the EP1-NS cell culture model as determined by HOMER motif analysis (see Methods and Supplementary Table 18). TF, transcription factor. **b**, Heatmap of transcription factors ranked by predicted activity using core circuitry analysis (left) and presence or absence of self-loop activity (right). $n = 18$ independent samples from Toronto cohort. **c–f**, shRNA constructs targeting super-enhancer-associated genes ordered by normalized cell survival. Highlighted in red are shRNAs targeting super-enhancer-associated core transcription factors. Each gene assayed with six technical replicates and replicated in three independent biological experiments. **g–l**, Connections between subgroup-specific transcription factors integrated with gene expression in subgroups of ependymoma. $n = 24$ independent samples.

factors that exhibited lower relative activity showed no significant difference in gene expression compared to normal brain (Extended Data Fig. 9). RNA interference (RNAi) was used to functionally demonstrate that the ependymoma core transcription factors *SOX9*, *RFX2*, *SOX2* and *ZBTB16* were essential for ependymoma cell maintenance (Fig. 3c–f, Extended Data Fig. 7). We hypothesized that this core model would be further specified by additional transcription factors that delineate the transcriptional differences between molecular subgroups of ependymoma. An integrative analysis was performed to assess subgroup-specific enhancers, the expression of their target genes within local topological associated domains¹⁵, and the enrichment of subgroup-specific transcription factor-binding motifs at these subgroup-specific enhancer loci. Using this approach, we modelled regulatory circuitry maps of each molecular subgroup of ependymoma, as defined by distinct sets of transcription factors, which might be used to establish and/or maintain ependymoma subgroup identity (Fig. 3g–l, Supplementary Table 20).

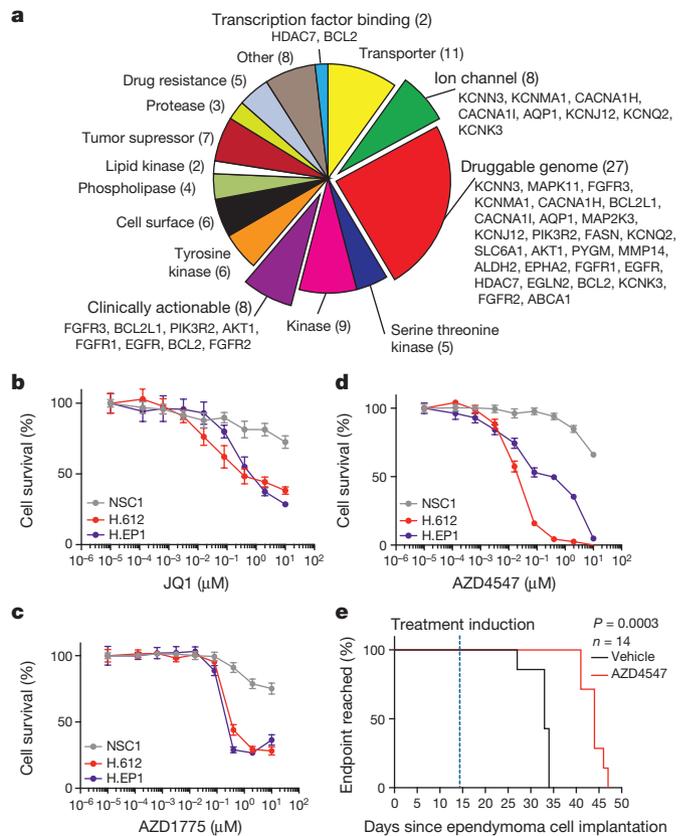


Figure 4 | Active regulatory maps identify candidate drugs against ependymoma. **a**, Pie chart of candidate drug compounds detected by integrating shared super enhancers with the Washington University Drug Gene Interaction Database. **b–d**, Ependymoma cells and neural stem cell line 1 (NSC1) controls treated with JQ1 (**b**), AZD1775 (**c**) or AZD4547 (**d**) for 72 h and assessed using an Alamar Blue stain. Error bars show s.d. Experiment performed as six technical replicates and replicated in biological triplicates. **e**, Kaplan–Meier curve for immunodeficient mice bearing H.612 ependymomas, treated with vehicle or AZD4547 (25 mg kg⁻¹ d⁻¹). Significance of endpoint difference was assessed using a log-rank test. Median survival ratio of treatment (AZD4547):control (vehicle) is 44 days:33 days, and reported as a ratio of 1.333 with a 95% confidence interval of 0.4677–3.801.

We leveraged subgroup-specific super-enhancer-regulated transcription factors to provide further insight into the lineage programs of ependymoma (Extended Data Fig. 10). The rationale for these experiments stemmed from our observation that in zebrafish embryos, several subgroup-specific super enhancers were active in specific regions within the developing central nervous system (Extended Data Fig. 9). We identified a *FOXJ1* transcription factor network that was enriched in PF-EPN-B ependymoma (Extended Data Fig. 10). *FOXJ1* is expressed during mouse embryonic development at E13.5 (during the expansion of radial glial cells (RGCs), which are candidate cells-of-origin of ependymoma) and its expression is restricted in the regions surrounding the choroid plexus in the mouse forebrain and hindbrain (Extended Data Fig. 10). Compared to other brain tumour types, *FOXJ1* expression was increased in ependymomas, with the highest levels in PF-EPN-B tumours¹⁶ (Extended Data Fig. 10). Furthermore, the ependymal differentiation program in RGC-derived *FOXJ1*-expressing cells versus *FOXJ1*-knockout cells was significantly and specifically enriched in PF-EPN-B ependymomas (Extended Data Fig. 10). From these data, we hypothesized that the transcriptional program of PF-EPN-B tumours closely resembles a more differentiated cell type along the ependymal lineage compared to ependymomas previously shown to match more primitive RGC precursor populations¹¹.

To inform the clinical translation of ependymoma dependencies, we prioritized targets for which small molecules were available by integrating our analysis of tumour-specific super-enhancer-regulated genes with the Washington University Drug Gene interaction database¹⁷ (Fig. 4a, Supplementary Table 21). *HDAC7*, *EPHA2*, *FGFR1* and *CACNA1H* were identified as candidate genes on which ependymomas depend that could be responsive to small-molecule inhibitors (Fig. 4a). Numerous subtype-restricted lead compounds were also identified (Supplementary Table 22). Active super enhancers marking molecular dependencies for ependymomas suggested that ependymoma cells would be responsive to inhibition of the BET bromodomain family of proteins by JQ1, which blocks protein 'readers' of H3K27 acetylation. JQ1 inhibited the proliferation of ependymoma cells at clinically achievable nanomolar concentrations and showed limited efficacy against normal brain cell proliferation (Fig. 4b). Our super enhancer analysis identified *FGFR1* small-molecule inhibitors as possible pan-ependymoma therapies, whereas inhibitors of another super-enhancer-associated gene product, *WEE1*, are likely to be active for subsets of ependymoma. *AZD4547* (*FGFR1* inhibitor) and *AZD1775* (*WEE1* inhibitor) exhibited potent and clinically achievable antitumour activity (Fig. 4c, d). Treatment of immunodeficient mice bearing posterior fossa ependymoma intracranial xenografts (H.612) with *AZD4547* extended survival (Fig. 4e), suggesting that chromatin landscapes can inform therapeutic paradigms.

Our study of active chromatin landscapes within ependymomas identified tumour- and subgroup-specific super-enhancer-driven genes in ependymoma as potential leads for further testing. By integrating our data with drug interaction databases, we identified and validated novel cancer dependencies of ependymoma that are responsive to pharmacologic inhibition. Our study further demonstrates that knowledge of enhancer landscapes can be used to dissect the molecular differences between histologically similar tumour entities and to provide unique information that may inform precision therapies. These differences are captured by the characterization of variant enhancer and super enhancer loci, in addition to the reverse engineering of core transcriptional regulatory circuitries in tumours. Finally, as shown in ependymomas and other tumours, knowledge of core and subgroup-specific transcription factors reveals a molecular basis for the oncogenic transcriptional programs of cancer, and provides insight into lineage programs that persist in the neoplastic state⁸.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 December 2016; accepted 22 November 2017.

Published online 20 December 2017.

- Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27**, 728–743 (2015).
- Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
- Witt, H. *et al.* Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer Cell* **20**, 143–157 (2011).
- Parker, M. *et al.* *C11orf95-RELA* fusions drive oncogenic NF- κ B signalling in ependymoma. *Nature* **506**, 451–455 (2014).
- Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
- Lin, C. Y. *et al.* Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
- Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Taylor, M. D. *et al.* Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* **8**, 323–335 (2005).

- Johnson, R. A. *et al.* Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature* **466**, 632–636 (2010).
- Mohankumar, K. M. *et al.* An *in vivo* screen identifies ependymoma oncogenes and tumor-suppressor genes. *Nat. Genet.* **47**, 878–887 (2015).
- Ramsey, S. A. *et al.* Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. *Bioinformatics* **26**, 2071–2075 (2010).
- Saint-André, V. *et al.* Models of human core transcriptional regulatory circuitries. *Genome Res.* **26**, 385–396 (2016).
- Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
- Abedalthagafi, M. S. *et al.* Decreased *FOXJ1* expression and its ciliogenesis programme in aggressive ependymoma and choroid plexus tumours. *J. Pathol.* **238**, 584–597 (2016).
- Griffith, M. *et al.* DGIdb: mining the druggable genome. *Nat. Methods* **10**, 1209–1210 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by an Alex's Lemonade Stand Young Investigator Award (S.C.M.), The CIHR Banting Fellowship (S.C.M.), The Cancer Prevention Research Institute of Texas (S.C.M., RR170023), Sibylle Assmus Award for Neurooncology (K.W.P.), the DKFZ-MOST (Ministry of Science, Technology & Space, Israel) program in cancer research (H.W.), James S. McDonnell Foundation (J.N.R.) and NIH grants: CA154130 (J.N.R.), R01 CA169117 (J.N.R.), R01 CA171652 (J.N.R.), R01 NS087913 (J.N.R.) and R01 NS089272 (J.N.R.). R.C.G. is supported by NIH grants T32GM00725 and F30CA217065. M.D.T. is supported by The Garron Family Chair in Childhood Cancer Research, and grants from the Pediatric Brain Tumour Foundation, Grand Challenge Award from CureSearch for Children's Cancer, the National Institutes of Health (R01CA148699, R01CA159859), The Terry Fox Research Institute and Brainchild. M.D.T. is also supported by a Stand Up To Cancer St. Baldrick's Pediatric Dream Team Translational Research Grant (SU2C-AACR-DT1113). Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. We thank S. Archer for technical writing and editing expertise. In addition, we thank the High-Throughput Sequencing Unit of the DKFZ Genomics and Proteomics Core Facility for technical support and acknowledge technical assistance by M. Mauerermann, T. Wedig, A. Wittmann and L. Siebert. Additional support came from the ICGC DE-Mining grant (#01KU1505). We thank The Children's Hospital at Westmead (CHW) Tumour Bank for support of tumour samples (H.W.). We thank D. Schumick (Cleveland Clinic Art Department) and G. Hsu (<http://www.hsubiomedicalvisual.com>) for their assistance with creative artwork.

Author Contributions S.C.M., K.W.P. and L.C. designed, performed and analysed the majority of the experiments in this study. Q.W. performed genetic knockdown experiments along with *in vivo* drug studies. K.C.B. performed all of the ChIP QC including library preparations and pre- and post-qPCR for the entire cohort. A.F., K.O. and S.E. performed the transcription factor network mapping of the super enhancer data. J.J.M. and T.E.M. assisted with super enhancer analysis and overall interpretation of data and analysis. Xin W., L.M., A.F.M. and I.S. led all of the zebrafish experiments in terms of establishment, interpretation and analysis. L.G., A.M., Y.T. and B.L.H. performed timed mating and tissue isolation in developing mouse embryos. J.R. assisted with pathway analysis of super enhancers. J.J.Y.L. assisted with ChIP experiments and library preparations. A.S. guided analysis of super-enhancer-subgroup stratification. D.C.F. performed RNA-seq pre-processing and analysis. B.L. helped with tissue isolation, preparation and submission for ChIP sequencing and DNA methylation analysis. Xia.W. and L.G. directed breeding and establishment of *meis1*-GFP mice. C.L.L.V., R.C.G. K.A.M. and A.T. performed data integration and mining of drug databases and identification of lead therapeutic compounds. A.M. performed super-enhancer-saturation analysis. P.C.S. assisted with study design, data analysis interpretation and manuscript review. S.Q.K., J.Z., V.M. and S.L., assisted with qPCR of numerous targets in genetic knockdown and differentiation experiments. P.J.H., T.M., A.M.C., S.K.S. and S.T.K. provided ependymoma models, controls and helped design the study. Xiu.W., L.D., S.D., L.K. and B.C.P. assisted with normal NSC drug treatments with drug inhibitors used in this study. C.L., C.-J.L., X.-W.B., C.G.H., M.R., S.D., S.V., S.N.G., H.W., D.T.W.J., P.A.N., P.L., A.K., N.J., J.T.R., E.B., A.H., K.D.A., P.B.D., Y.L., M.L., Z.H., M.Z., V.R., J.E.B., S.M.P., P.S.-C. and P.C.S., assisted with data interpretation, manuscript preparation and review. M.D.T., J.N.R. and M.K. conceived, designed, interpreted and funded the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.N.R. (drjeremrych@gmail.com), M.D.T. (mdtaylor@sickkids.ca) or M.K. (m.kool@dkfz-heidelberg.de).

Reviewer Information *Nature* thanks S. Pomeroy, W. Weiss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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 previously². 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 described¹⁵.

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://genencode.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 2.3.0¹⁹ 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 v0.6.3²⁰.

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 described²¹. 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 v0.0.11. 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 classifications².

Identification of super-enhancer-associated pathways and drug-gene interactions. Differential super-enhancer-associated genes in ependymomas or ependymoma subgroups were imported into G-Profiler²² 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 gene sets 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 targets¹⁷.

Analysis of super enhancers with subgroup-specific enhancer activity (SSEA).

To identify subgroup-specific enhancer activity, we employed the R/Bioconductor package QSEA v.0.0.11²³. Previously calculated enhancer regions (see above) were provided as regions of interest and tiled into 500-bp windows. For each sample, H3K27ac ChIP-seq enrichments were calculated at these tiled enhancers and were library size-normalized by TMM. In addition, matched blood and tumour WGS data were imported and copy number variations were calculated for all endymoma samples using the findCNV() function of the QSEA package. CNV-aware subgroup-specific enhancer activity was then calculated by comparing H3K27ac ChIP-seq enrichments in one subgroup against the other subgroups by fitting general linear models with respect to the presence of CNVs (non-default parameters are norm_method = "nrpk", minRowSum = 10, fdr_th = 10⁻⁵, direction = "gain"). We excluded 500-bp windows that were significant in more than one subgroup. For each subgroup, we stitched all significant 500-bp windows within a distance of 12.5 kb together, summed their normalized H3K27ac ChIP-seq enrichment values (nRPKM), and ranked them accordingly. Analogous to the definition of super enhancers, we define the first occurrence of a slope > 1 (from high to low enrichment) as a threshold for distinguishing between extended stretches of significant SE-SSEAs and TE-SSEAs.

Calculating core regulatory networks for super-enhancer-associated transcription factors.

To quantify the interaction network of transcription factor regulation, we calculated the inward and outward binding degree of all super-enhancer-associated transcription factors¹⁴. For all promoters within 100 kb, the most acetylated promoter was assigned as the target of the super enhancer (excluding promoters that overlap super enhancers). If there were no active promoters within 100 kb, the super enhancer was assigned to the nearest active promoter. All super-enhancer-associated promoters annotated to regulate a transcription factor were considered as the node-list for network construction. For any given transcription factor (TFi), the IN degree was defined as the number of transcription factors with an enriched binding motif at the proximal super enhancer or promoter of TFi. The OUT degree was defined as the number of transcription factor-associated super enhancers containing an enriched binding site for TFi. Within any given super enhancer, enriched transcription factor binding sites were determined at putative nucleosome-free regions (valleys) flanked by high levels of H3K27ac. Valleys were calculated using an adapted algorithm¹³. In these regions, we searched for enriched transcription factor binding sites using the FIMO59 algorithm with transcription factor position weight matrices defined in the TRANSFAC database²⁴. An FDR cut-off of 0.01 was used to identify enriched transcription factor-binding sites.

Identification of regulatory networks at enhancers with subgroup-specific enhancer activity.

Subgroup-specific transcription factor-regulatory networks were constructed as previously described with only a few amendments^{8,25}. H3K27ac data of the samples within the same subgroup were combined. For each subgroup, nucleosome-free regions (NFRs) were identified using the findPeaks function of HOMER²⁶ (<http://homer.salk.edu/homer/ngs/index.html>) with option -nfr. ENCODE transcription factor motifs and their mapped positions in the genome were downloaded from <http://compbio.mit.edu/encode-motifs/>. For each transcription factor, contingency tables containing the number of NFRs overlapping and non-overlapping with the respective transcription factor were constructed. The significance of enrichment of transcription factors in NFRs of enhancers with subgroup-specific activity was determined using the χ^2 test. The resulting *P* values were corrected for multiple testing (FDR < 0.01). Transcription factor enrichments were calculated as the ratio between observed counts over expected counts. To identify enhancer target genes, we accessed publicly available topology-associated domains (TADs) previously obtained in IMR90 cells. Each SSEA was assigned to its enclosing TAD and protein-coding genes within the same TAD were identified. Correlation tests (Spearman's rank correlation coefficient) for SSEA H3K27ac enrichment and gene expression level within the same TAD were performed. After repeating this procedure for each enhancer, all *P* values obtained were combined and corrected for multiple testing using the Bioconductor package qvalue. Correlations with an FDR less than 1% were preserved. To derive subgroup-specific transcription factor regulatory networks, we selected the top 50% enriched transcription factors in each subgroup, which also have the highest expression in the respective subgroup compared to the other subgroups. The resulting networks highlight transcription factors (red or orange nodes) whose binding sites are significantly enriched at enhancers with SSEA. By gene-enhancer correlation analysis restricted by TAD domains (see above), these transcription factors were assigned to their likely target genes (blue nodes). Networks were visualized using Gephi (<http://gephi.github.io/>).

ATAC-seq chromatin preparation and sequencing. Freshly cultured ependymoma cells were prepared for ATAC-seq as described²⁷. In brief, nuclei were prepared from ~50,000 cells by spinning at 600g for 10 min at 4 °C, followed by

a PBS wash and centrifugation at 600g for 5 min. Cells were lysed using ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1%), and centrifuged for 10 min at 600g at 4 °C. The supernatant was removed and pellet re-suspended in 50 μ l transposase mix (25 μ l 2 × TD buffer, 2.5 μ l transposase, 22.5 μ l water) (FC-121-1030 Illumina) for 30 min at 37 °C. Library amplification was performed using the NEBnext High Fidelity 2 × PCR Master Mix (#M0541S New England Biolabs) according to previously published PCR conditions²⁷. PCR reactions were purified using a QIAGEN miniElute kit, and a following size selection step using standard gel extraction protocol to isolate ~240–360 bp. ATAC-seq library preparations were sequenced using single-end 50-bp reads on the Illumina HiSeq 2000 platform. Raw reads were adaptor-trimmed using Trim Galore (v0.2.5) and aligned to the genome with Bowtie (v1.0.1) with the m1 option enabled to allow only uniquely aligned high-quality reads. Peaks were called using the MACS2 software (v2.1.0.20140616) with the options -q 0.05 to retain significant peaks, -shiftsize 50 to account for the transposase fingerprint, and otherwise default parameters were used. Tag count libraries and bedgraph files were constructed using HOMER software (v4.7).

Ependymoma culture experiments. Ependymoma cell cultures were isolated from patients and cultured on laminin (Sigma) and in neurobasal medium (Invitrogen) consisting of: sodium pyruvate (Invitrogen), B27 (Invitrogen), glutamine (Cleveland Clinic Media Core), human EGF (Invitrogen), human basic FGF (Invitrogen), and penicillin/streptomycin (Cleveland Clinic Media Core). Medium was replenished every other day while leaving ~50% conditioned medium to encourage continued cell proliferation. Cell viability assays were performed in 96 wells using an Alamar Blue stain (Invitrogen) according to the manufacturer's instructions. Drug-response assays were performed by seeding cells overnight, treating the following day with increasing drug concentrations, and reading by Alamar Blue Absorption following 72 h of treatment. AZD4547 and MK1775 were obtained from Selleck Chemicals. JQ1 was provided by the laboratory of J. E. Bradner (Harvard). All cell lines were STR profiled for authenticity and confirmed to be mycoplasma free using a PCR-based detection strategy with positive and negative controls.

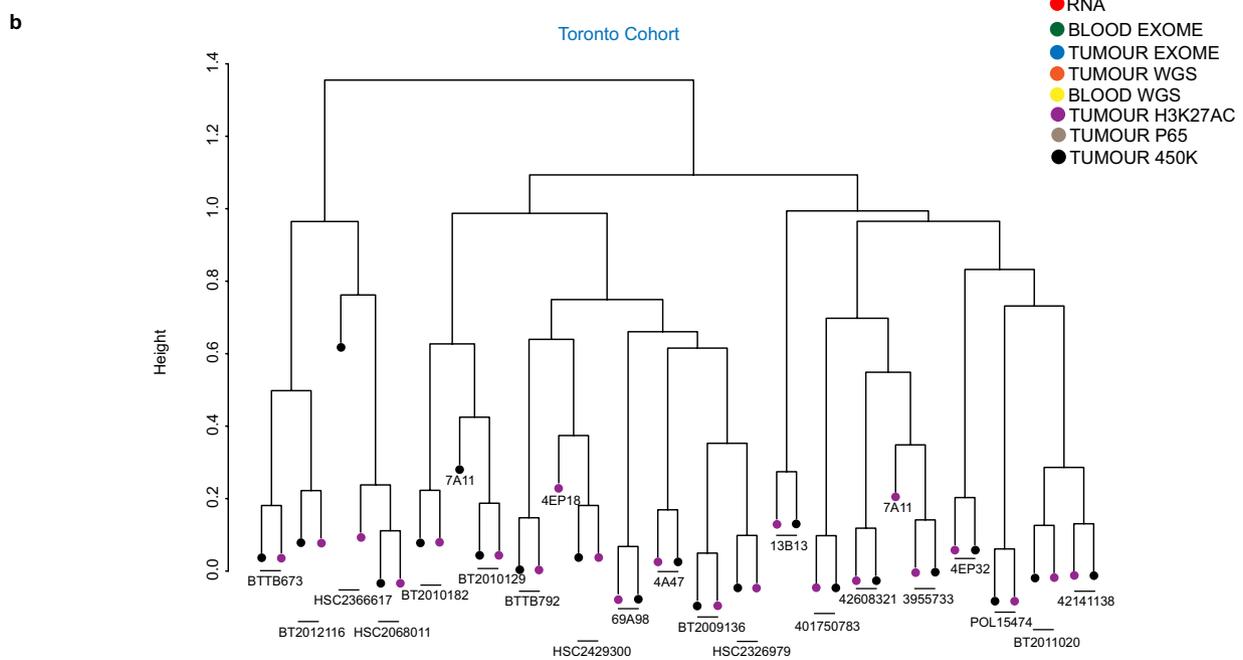
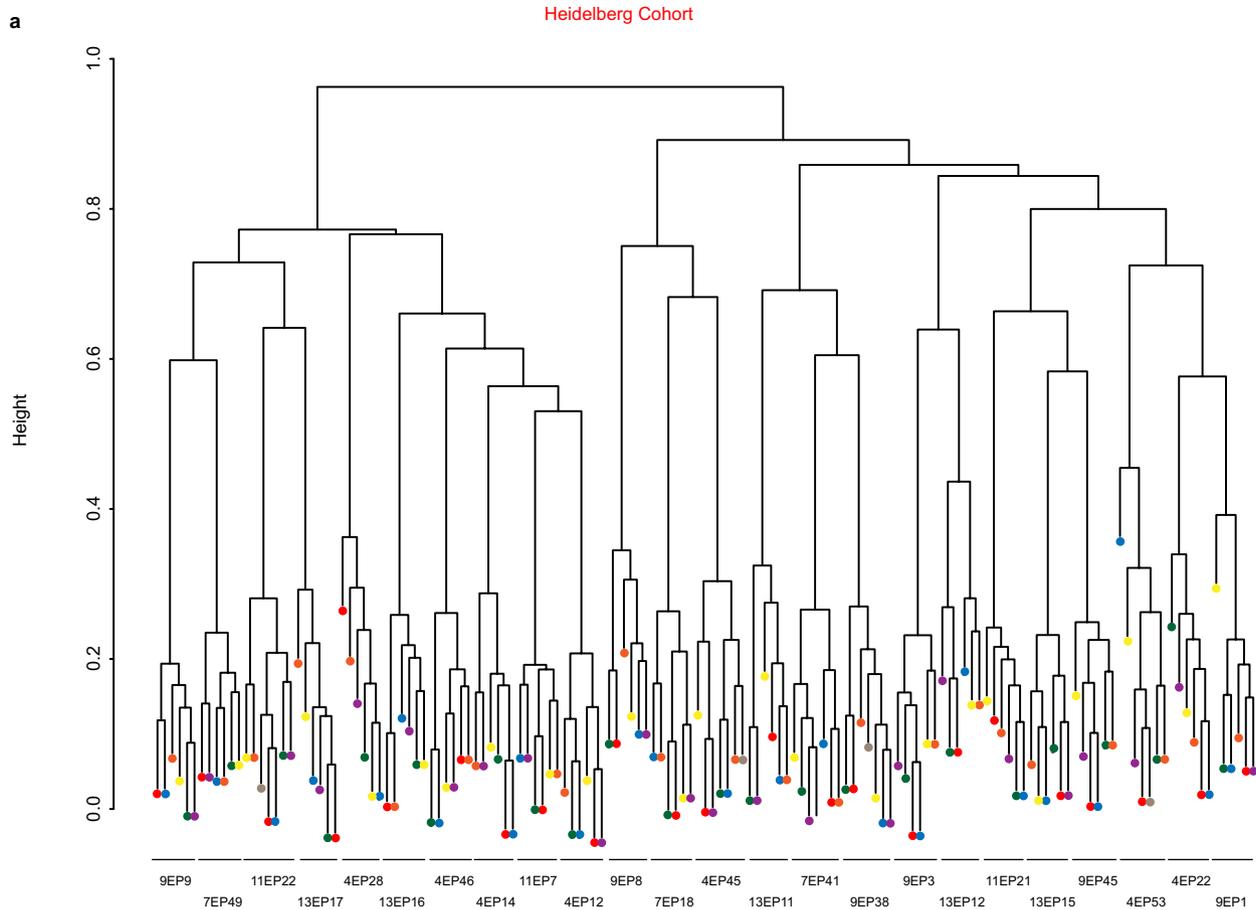
RNA interference of enhancer-associated genes. Lentiviral shRNA clones (Sigma Mission RNAi) targeting super-enhancer-associated genes, and two non-targeting controls (SHC002, SHC007) were purchased from Sigma. (Supplementary Table 23). These vectors were co-transfected into HEK 293FT cells with the packaging vectors psPAX2 (Addgene) and pCI-VSVG (Addgene) using a calcium phosphate method to produce viable lentivirus. Knockdown efficiency of different lentiviral shRNA clones in cells was determined by quantitative reverse transcription PCR. Cells infected with lentivirus expressing the indicated shRNAs were plated in 96-well plates at 1,000 cells per well. Cell viability was determined after the indicated number of days after plating using Alamar Blue Assay (Life Technologies) or CellTitreGlo (Promega).

CRISPR-Cas9-mediated repression of enhancer regions. CRISPR-Cas9 sgRNAs were identified and designed using the MIT CRISPR design tool, and control (pLenti-Guide-Puro D103) non-targeting sgRNAs were selected from the GeCKOv2 library. All sgRNA sequences may be found in Supplementary Table 23. sgRNAs were cloned into plenti-Guide-Puro (Addgene, 52963). Lentivirus expressing dCAS9-KRAB (gift from M. Meyerson laboratory)²⁸ were used to infect EP1-NS, following which cells were selected for 48 h with 10 μ g/ml blasticidin. These cells were then infected with selected lentiGuide-Puro sgRNA constructs and selected for 48 h with 1 μ g/ml puromycin. These cells were plated for 48 h following selection in 96-well plates and cell viability was assessed using an Alamar Blue Stain (Life Technologies).

In vivo animal experiments. We followed the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioural Research from the National Research Council to estimate the minimal number of animals necessary to assess statistical significance. The number of animals per arm was based upon the following calculation: $N = 1 + 2C(s/d)^2$ where *n* is the number of animals per arm, $C = 7.85$ when $\alpha = 0.05$ and $1 - \beta = 0.8$ (significance level of 5% with a power of 80%), *s* is standard deviation, and *d* is the difference to be detected. All animal experiments were performed in accordance with local IACUC regulations and protocols. Animal experiments were conducted in a single-blinded fashion, and endpoints were assessed by an independent animal technician in the laboratory. 250,000 H612 cells were xenografted intracranially into NOD/SCID/ γ female mice. Tumours were allowed to develop for 14 days then independently randomized into a treatment or vehicle group. AZD4547 (25 mg/kg/d) or vehicle (Sigma: 1% Tween-80) were administered daily by oral gavage. Survival of mice was plotted using a Kaplan–Meier curve and quantified using a log-rank test. Our study did not measure tumour size or volume directly. We monitored neurological signs and behaviours associated with brain tumour development in accordance with our IACUC protocols and regulations.

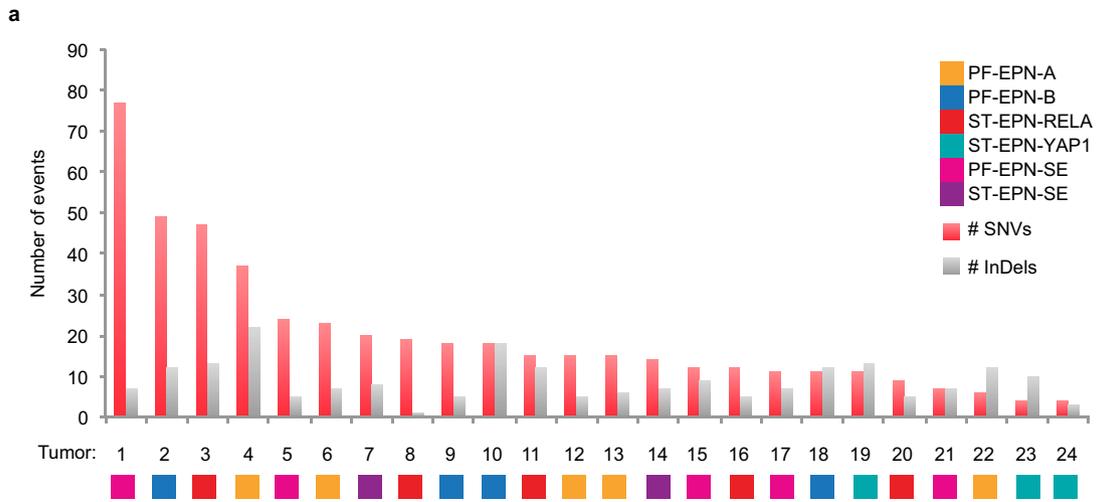
Data availability. All raw data files were deposited in the European Genome-phenome archive (<https://www.ebi.ac.uk/ega/home>) under the accession number: EGAS00001002696.

18. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korb, J. O. Data analysis: create a cloud commons. *Nature* **523**, 149–151 (2015).
19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
20. Okonechnikov, K. *et al.* InFusion. Advancing discovery of fusion genes and chimeric transcripts from deep RNA-sequencing data. *PLoS One* **11**, e0167417 (2016).
21. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
22. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
23. Lienhard, M. *et al.* QSEA-modelling of genome-wide DNA methylation from sequencing enrichment experiments. *Nucleic Acids Res.* **45**, e44 (2016).
24. Matys, V. *et al.* TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
25. Johann, P. D. *et al.* Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29**, 379–393 (2016).
26. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
27. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
28. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
29. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
30. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).



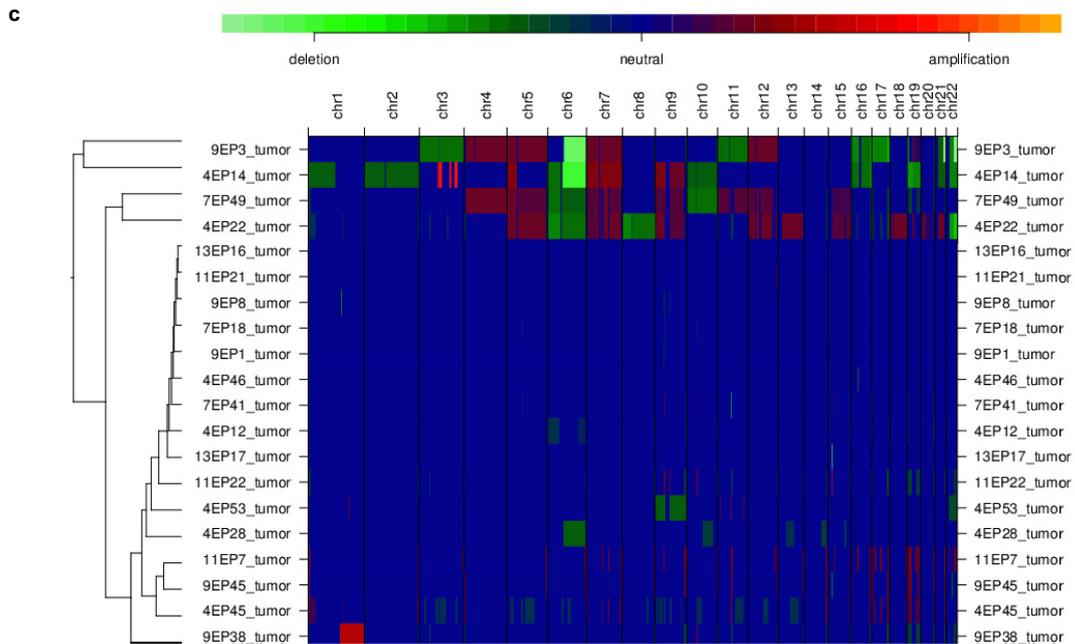
Extended Data Figure 1 | DNA fingerprint analysis of ependymoma sequence data. a, b, Unsupervised clustering of ChIP-seq, RNA-seq, WES, WGS, and Illumina DNA methylation profiles with genotypes that have an

average heterozygosity score greater than 0.25 in the Heidelberg ($n = 25$ independent samples) (a) and Toronto cohorts ($n = 18$ independent samples) (b).



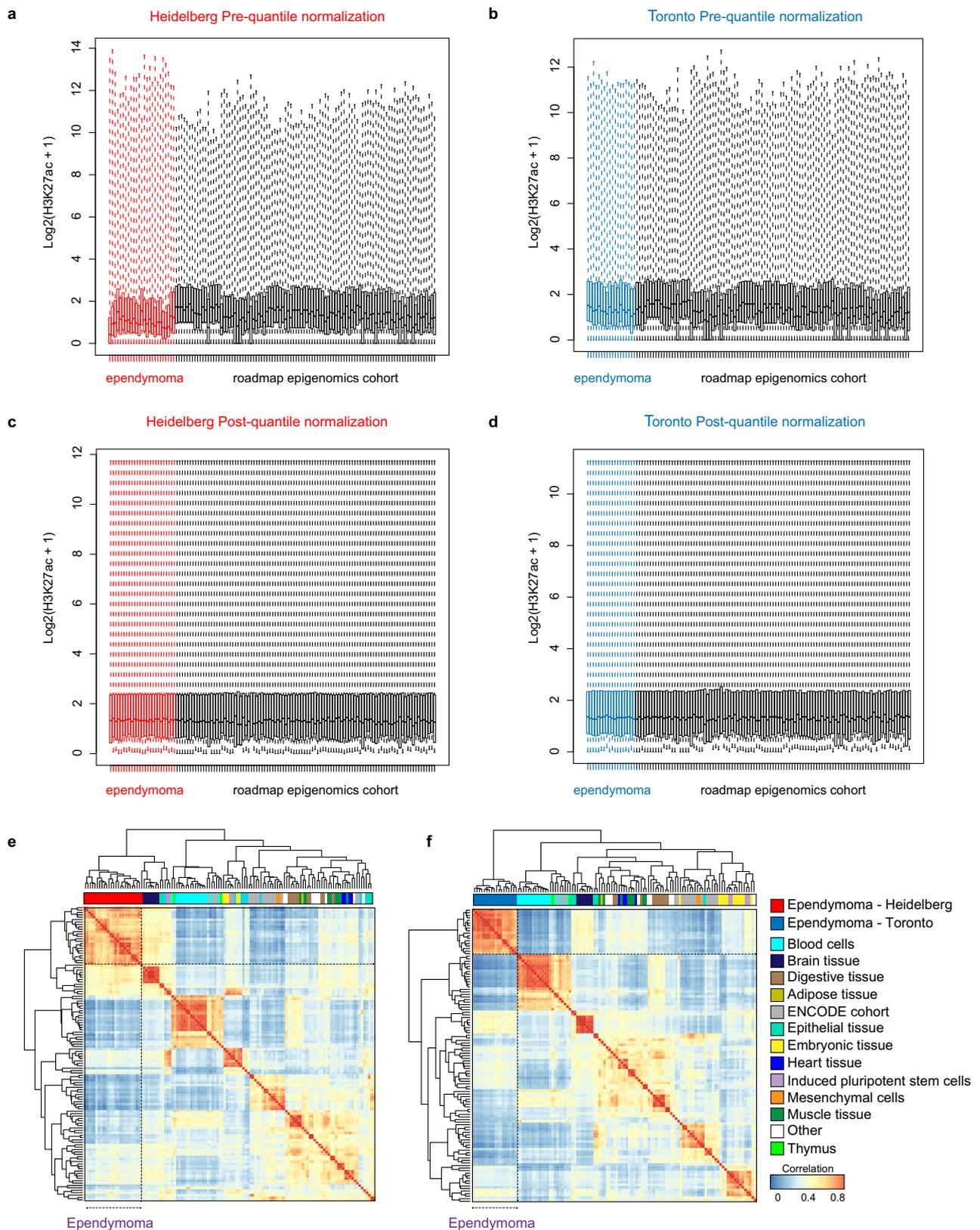
b

Recurrently affected genes (SNVs)		Recurrently affected genes (indels)	
2/24 (8.33333%)	SHOX	2/24 (8.33333%)	WISP1
2/24 (8.33333%)	PDE3B	2/24 (8.33333%)	SORCS1
2/24 (8.33333%)	PAK2	2/24 (8.33333%)	snoU13
2/24 (8.33333%)	MYO15A	2/24 (8.33333%)	SHCBP1
2/24 (8.33333%)	ICOSLG	2/24 (8.33333%)	RP3-470B24.5
2/24 (8.33333%)	HT T	2/24 (8.33333%)	PTPRK
2/24 (8.33333%)	EPHX4	2/24 (8.33333%)	NDRG1
2/24 (8.33333%)	AGXT2	2/24 (8.33333%)	FAM169A
2/24 (8.33333%)	AC079354.1	2/24 (8.33333%)	DLEU2
		2/24 (8.33333%)	CROCCP2



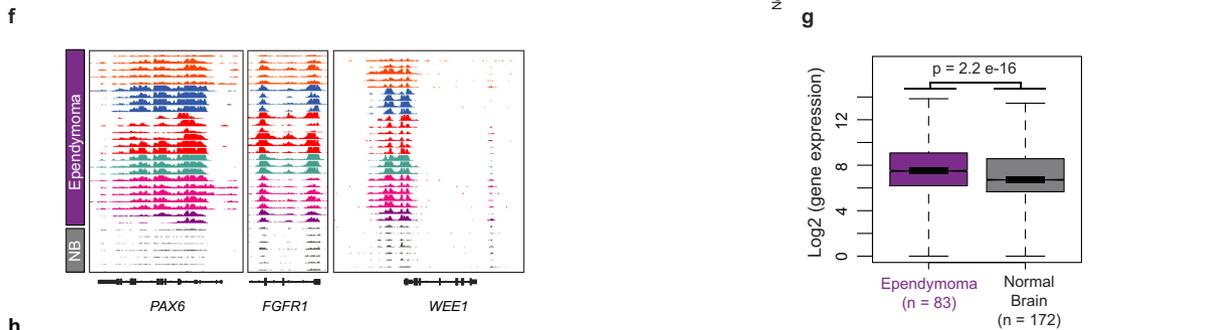
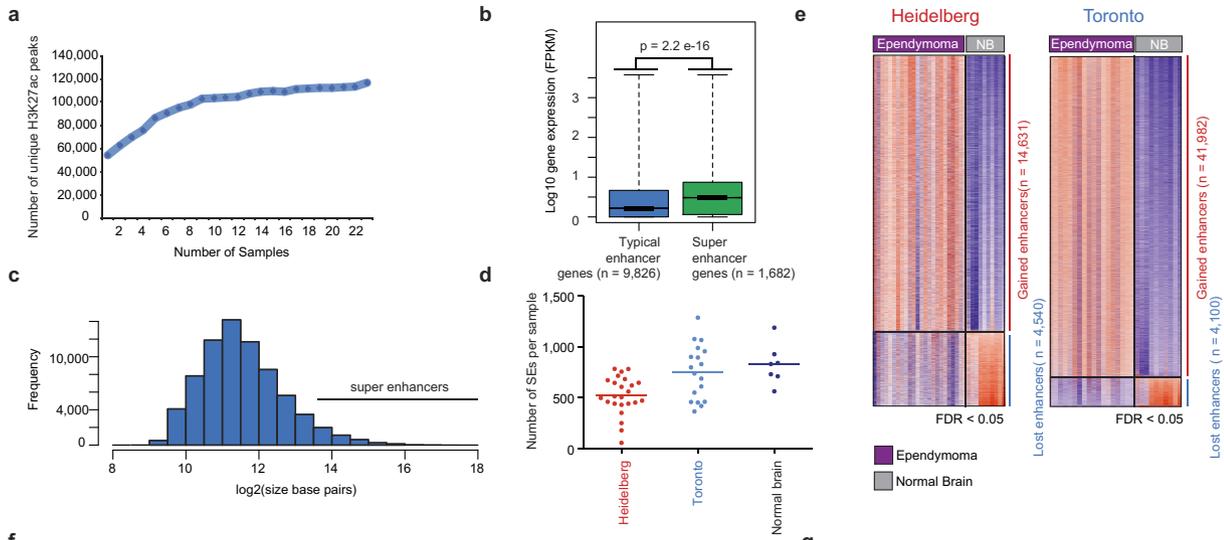
Extended Data Figure 2 | Summary of genome sequencing and copy number data. **a**, Number of somatic single nucleotide variants (SNVs) detected per endependymoma sample. **b**, Frequency of somatic mutations detected across the Heidelberg endependymoma cohort ($n = 24$ independent

samples). **c**, Unsupervised hierarchical clustering of copy number alterations detected by WGS in primary endependymoma samples ($n = 24$ independent samples).

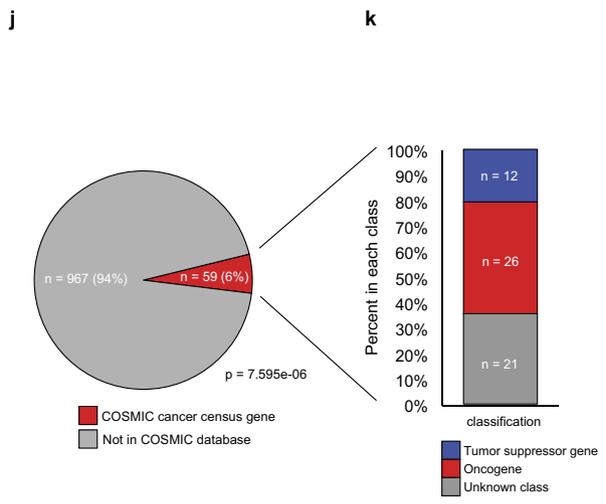
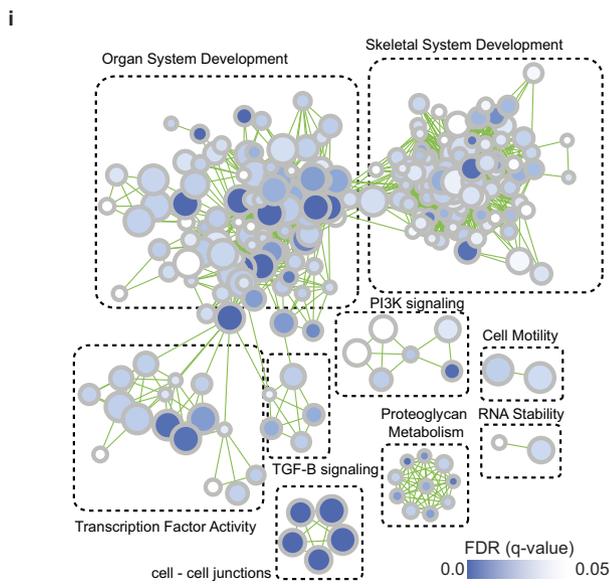


Extended Data Figure 3 | Preprocessing and clustering of ependymoma H3K27ac profiles. **a, b,** Box plots of H3K27ac enhancer profiles ($n = 556,676$ enhancer loci evaluated per sample) before quantile normalization for both Heidelberg ($n = 24$ independent samples) (**a**) and Toronto ($n = 18$ independent samples) (**b**) cohorts compared to Roadmap Epigenomics and ENCODE cohorts ($n = 98$ independent samples). Box plots are shown with the centre (median), upper and lower quartile range, and dotted line indicating minima and maxima per sample. **c, d,** Box plots

of H3K27ac enhancers after quantile normalization for both Heidelberg ($n = 24$ independent samples) (**c**) and Toronto ($n = 18$ independent samples) (**d**) cohorts compared to the Roadmap Epigenomics cohort ($n = 98$ independent samples). **e, f,** Unsupervised hierarchical clustering of enhancer profiles as measured using the top 10,000 variant enhancer loci identified in the Roadmap Epigenomics cohort with the Heidelberg ($n = 122$ independent samples) (**e**) and Toronto cohorts ($n = 116$ samples) (**f**) and compared in a pair-wise fashion using a Spearman correlation.



Comparison	Heidelberg cohort	Number confirmed by Toronto cohort	Percent confirmed by Toronto cohort
Overall typical enhancers (PFA, PFB, RELA)	40,565	34,040	83.9
Overall super enhancers (PFA, PFB, RELA)	2,196	1,682	76.6
PF-EPN-A super enhancers with subgroup specific activity	229	218	95.2
PF-EPN-B super enhancers with subgroup specific activity	202	178	88.1
ST-EPN-RELA super enhancers with subgroup specific activity	182	172	94.5

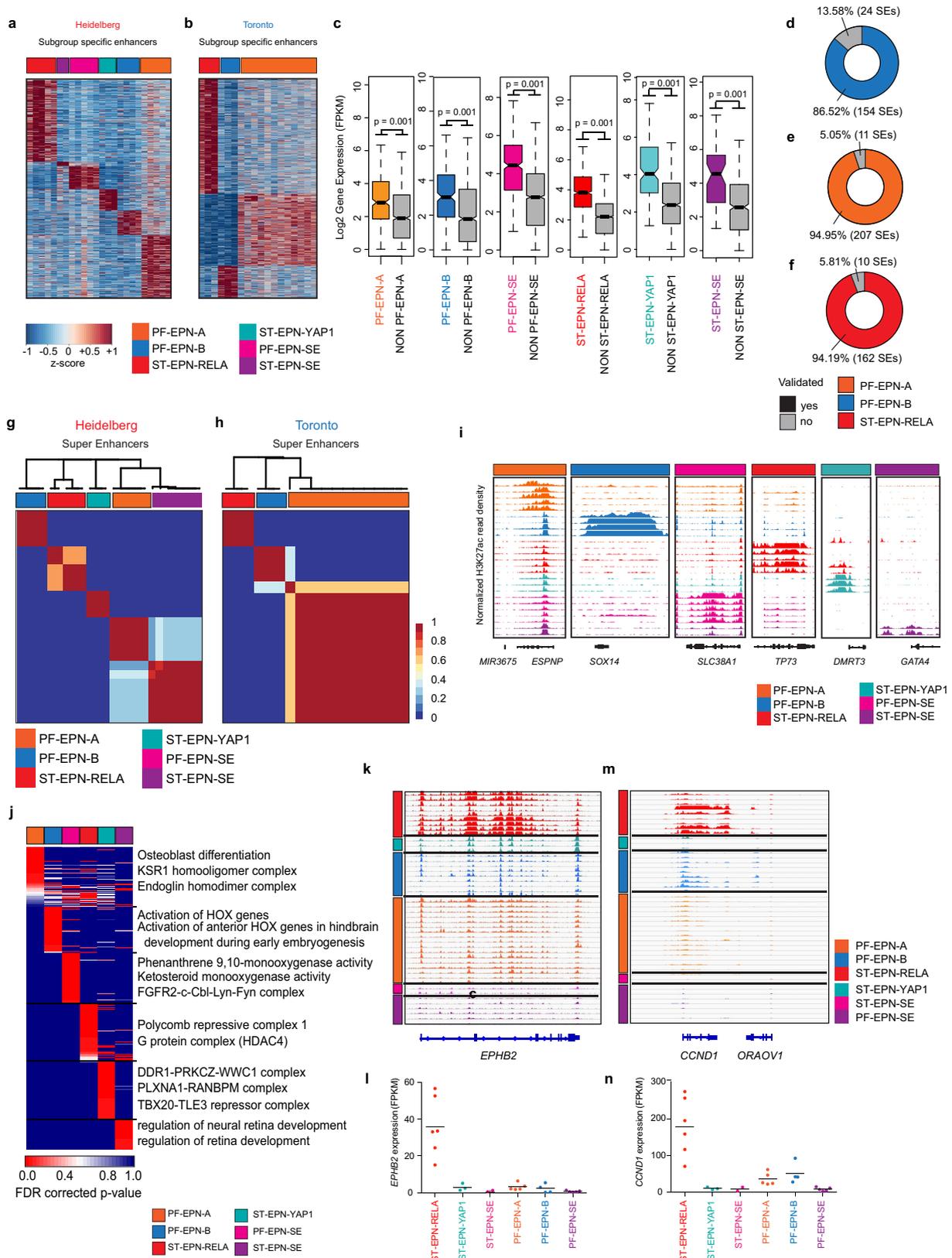


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Ependymoma enhancer supporting data.

a, Number of unique H3K27ac peaks detected by MACS1.4 ($P < 1 \times 10^{-9}$ cut-off) with increasing sample number in the Heidelberg cohort ($n = 24$ independent samples). **b**, Box plot of gene expression values comparing typical enhancer ($n = 9,826$ genes) versus super enhancer ($n = 1,682$ genes) associated genes. Statistical analysis was assessed using a two-sided Wilcoxon rank-sum test. Box plots show the centre (median), upper and lower quartile range, and dotted line indicating minima and maxima. **c**, Frequency of enhancer and super enhancer regions as a function of size in base pairs. **d**, Dot plots illustrating the numbers of super enhancers detected in the Heidelberg ($n = 24$ independent samples), Toronto ($n = 18$ independent samples) and normal brain ($n = 7$ independent samples) cohorts. The horizontal bar indicates the mean. **e**, Heatmap illustrating significant gained and lost enhancer loci in both ependymoma cohorts compared to normal brain samples. Comparisons were evaluated using a two-sided Wilcoxon rank-sum test with FDR correction and a cut-off of $FDR < 0.05$. **f**, Example plots of normalized and scaled H3K27ac RPKM profiles at example ependymoma candidate genes in Heidelberg

ependymomas and normal brain (NB) ($n = 32$ independent samples). **g**, Comparison of gene expression of ependymoma super-enhancer-associated genes derived from ref. 11 ($n = 83$ independent samples) with normal brain ($n = 172$ independent samples). Statistical analysis was assessed using a two-sided Wilcoxon rank-sum test. **h**, Table comparing the number and per cent confirmation between the Heidelberg ($n = 24$ independent samples) and Toronto ependymoma cohorts ($n = 18$ independent samples). **i**, G-Profiler pathway-enrichment analysis of ependymoma-specific super-enhancer-associated genes in the Toronto cohort ($n = 18$ independent samples), with statistical significance determined using a hypergeometric test. **j**, Overlap analysis measured by a two-sided binomial test between tumour-specific ependymoma super enhancers and cancer census genes from the Catalogue of Somatic Mutations in Cancer (COSMIC) database. **k**, Classification of tumour-specific ependymoma super enhancer genes also found in the COSMIC database²⁹ as tumour suppressor genes ($n = 12$), oncogenes ($n = 26$), or unknown ($n = 21$).

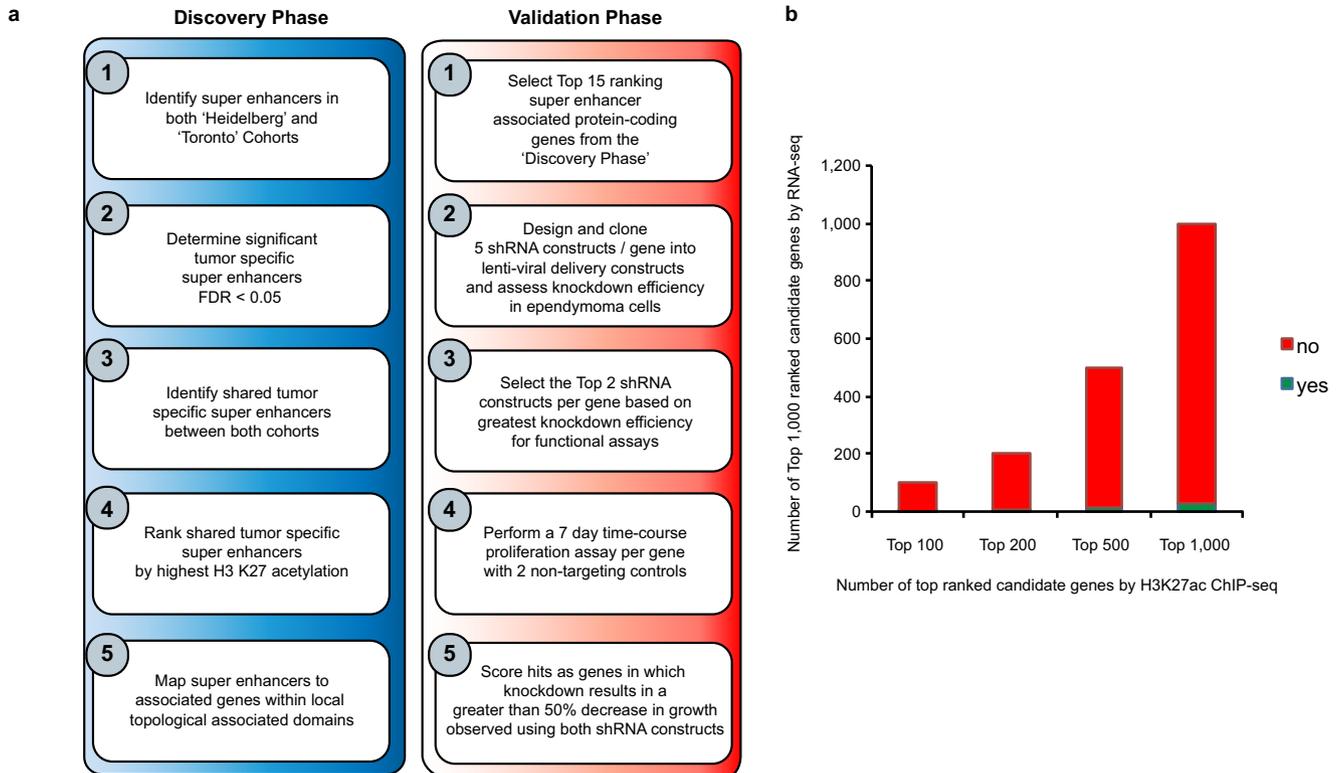


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Subgroup-specific enhancers of ependymoma.

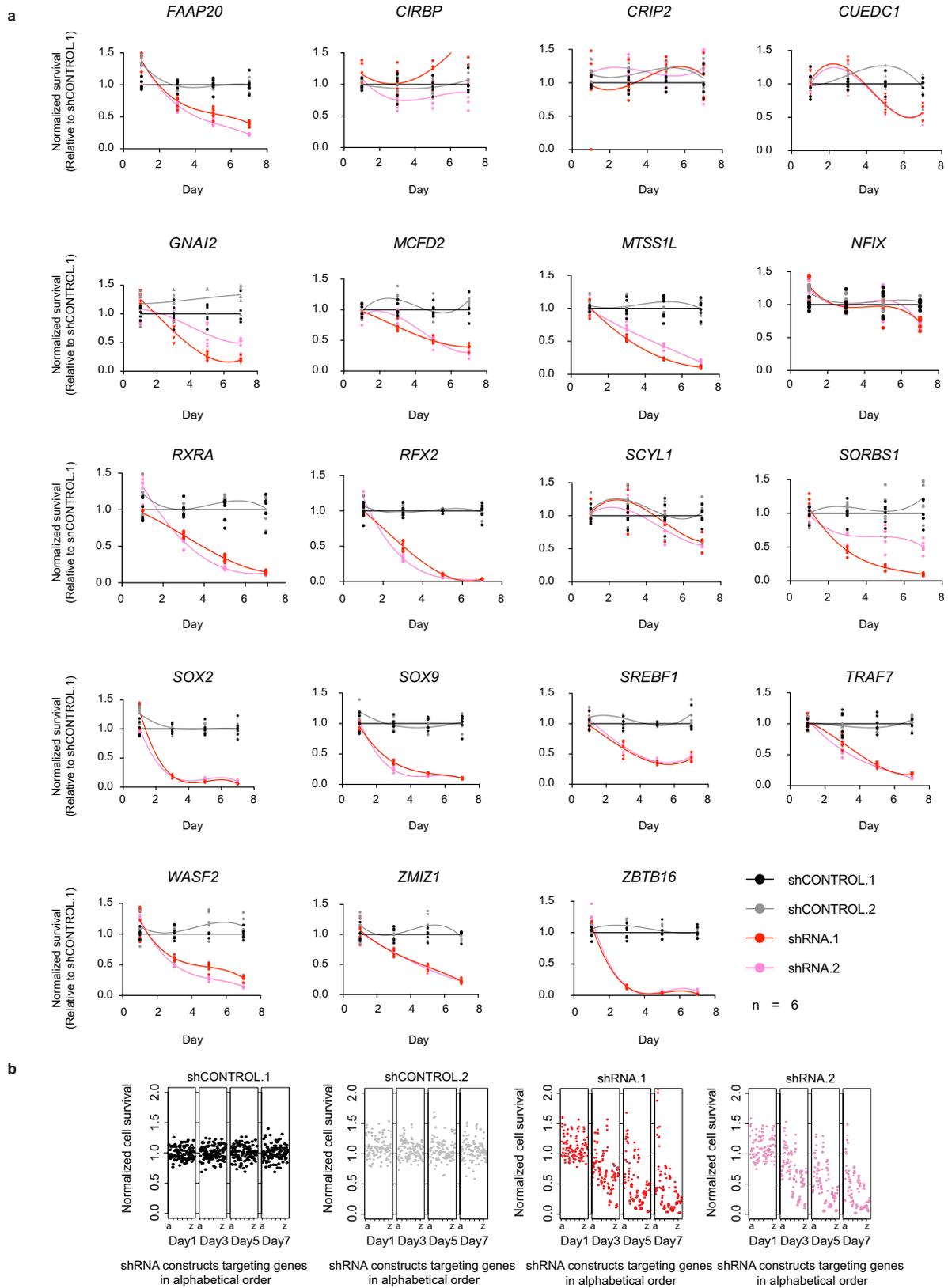
a, b, Heatmap of all subgroup-specific active enhancers detected in ependymomas in independent samples in the Heidelberg ($n = 24$ independent samples) (**a**) and Toronto ($n = 18$ independent samples) (**b**) cohorts. **c**, Box plot of gene expression for ependymoma SE-SSEA-associated genes in the Heidelberg cohort ($n = 24$ independent samples). Comparisons were made using a two-sided Wilcoxon rank-sum test. Box plots show the centre (median), upper and lower quartile range, and dotted lines indicate minima and maxima. **d–f**, Venn diagrams of the number and percentage of subgroup-specific super-enhancer-associated loci validated between the Heidelberg and Toronto cohorts. **g, h**, Non-negative factorization of ependymoma super enhancer profiles in the Heidelberg ($n = 24$ independent samples) and Toronto ($n = 18$ independent samples) cohorts. **i**, Normalized H3K27ac profiles for

subgroup-specific genomic example loci in the Heidelberg cohort with at least three biological replicates per subgroup, with the exception of ST-EPN-SE, shown as a biological duplicate. **j**, G-Profiler pathway-enrichment analysis of ependymoma subgroup-specific super-enhancer-associated genes in the Heidelberg cohort ($n = 24$ independent samples) with statistical significance determined using a hypergeometric test. **k–n**, H3K27ac profiles surrounding the *EPHB2* (**k**) and *CCND1* (**m**) loci in the Heidelberg cohort with at least three biological replicates per subgroup, with the exception of ST-EPN-SE, shown as a biological duplicate. *EPHB2* (**l**) and *CCND1* (**n**) expression by RNA-seq across ependymoma subgroups in the Heidelberg cohort with horizontal bars indicating the median value and each dot representing an independent ependymoma sample ($n = 24$ independent samples).



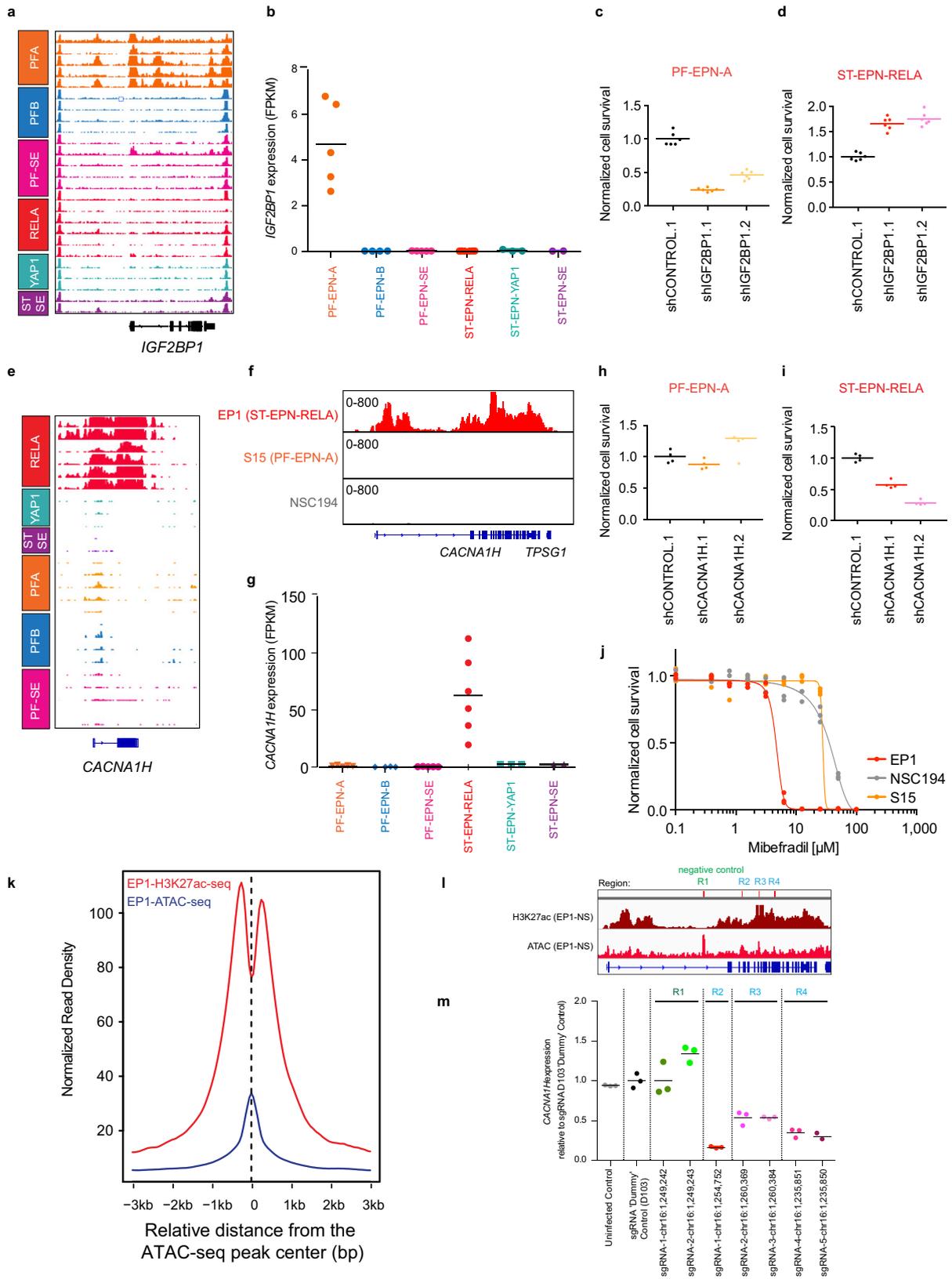
Extended Data Figure 6 | Workflow describing the functional validation of ependymoma super enhancer genes. **a**, Workflow of super-enhancer target-gene prioritization for functional evaluation. **b**, Bar chart comparing the top-ranked super-enhancer-associated genes against top-ranked genes detected by RNA-seq defined as significantly increased or overexpressed

compared to normal brain controls across all ependymoma samples ($n = 42$ independent samples). Significant genes were identified by a two-sided Wilcoxon rank-sum test with FDR correction and ranked by FDR corrected P value with a cut-off of less than 0.05.



Extended Data Figure 7 | RNA interference of ependymoma super enhancer genes. **a**, Individual shRNA time-course knockdown experiments in EP1-NS (ST-EPN-RELA) cells, using two shRNA constructs (shRNA.1 and shRNA.2) compared to two controls (shCONTROL.1 and shCONTROL.2). Shown are time-course experiments for 19 genes performed in six technical replicates.

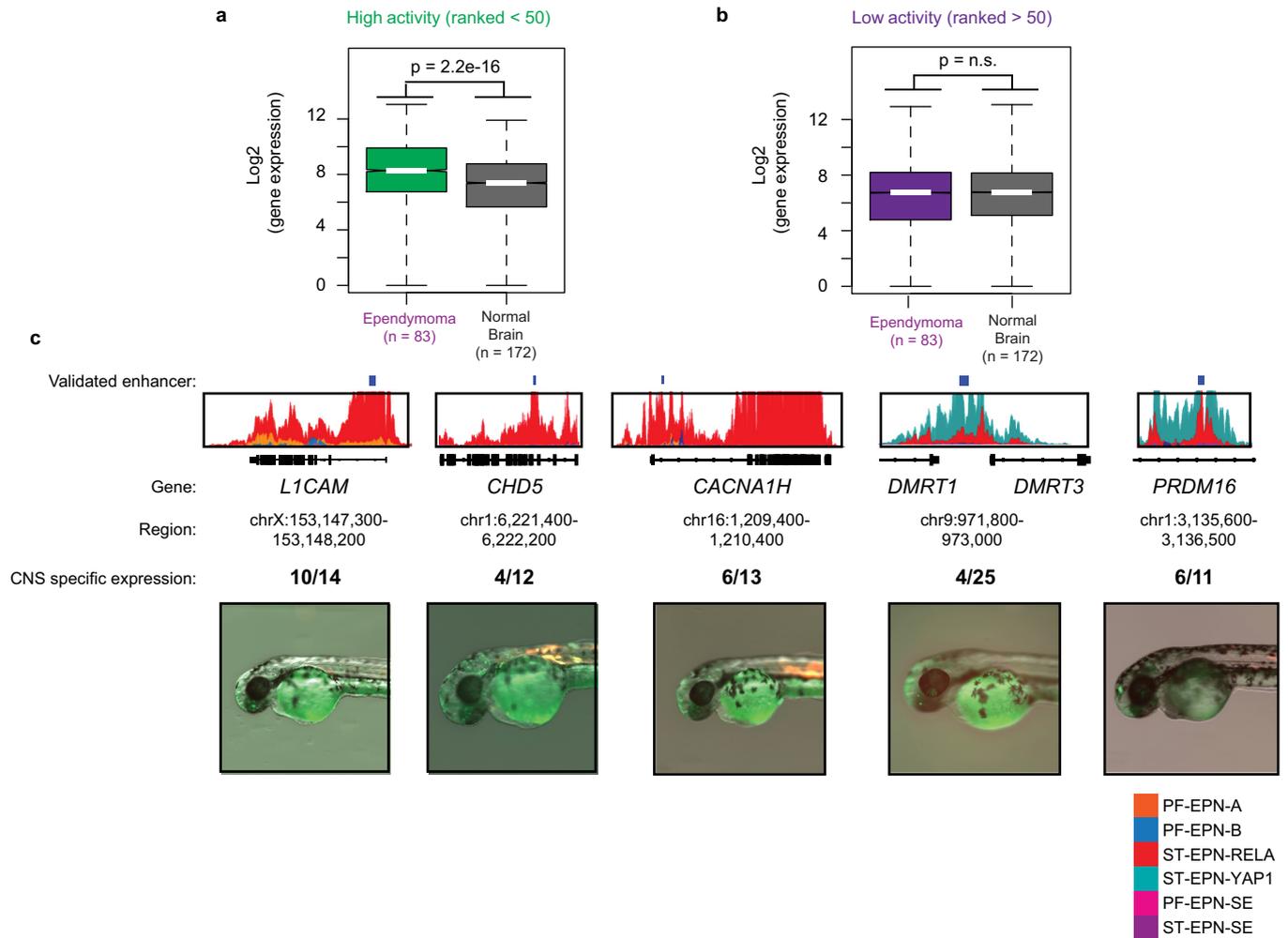
b, Ependymoma cell viability (EP1-NS) following treatment with shRNAs targeting super-enhancer-associated genes over a seven-day time course (in alphabetical order). Cell viability data for treatment with non-targeting controls: shCONTROL.1 (black), shCONTROL.2 (grey), and for two gene-specific shRNA constructs: shRNA.1 (red) and shRNA.2 (pink).



Extended Data Figure 8 | See next page for caption.

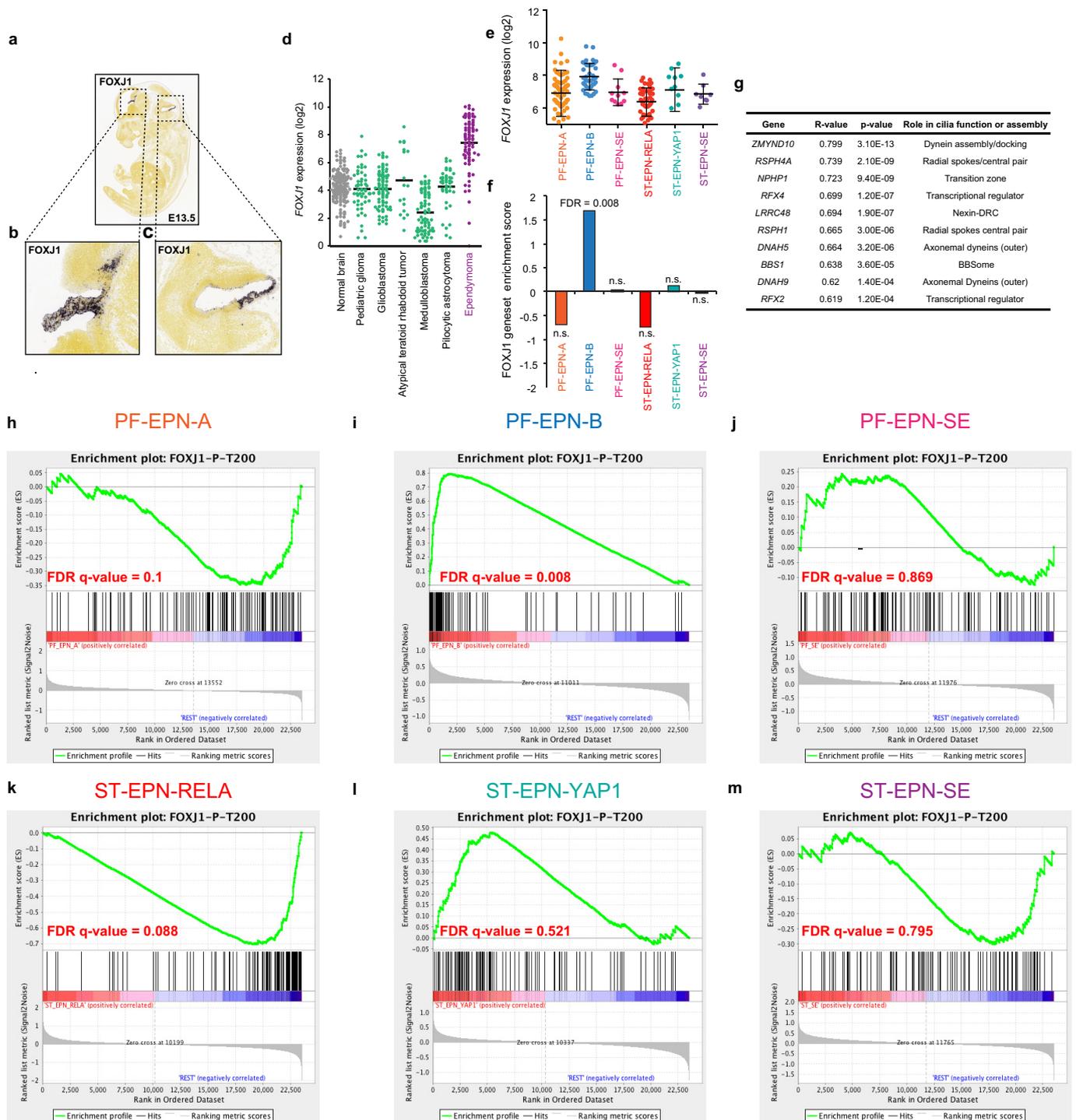
Extended Data Figure 8 | Validation of ependymoma subgroup-specific super enhancer genes. **a**, H3K27ac profiles at the ependymoma-specific super enhancer locus *IGF2BP1* in the Heidelberg cohort ($n = 24$ independent samples) with at least three biological replicates per subgroup, with the exception of ST-EPN-SE, which is shown as a biological duplicate. **b**, *IGF2BP1* gene expression derived from RNA-seq data for the Heidelberg cohort ($n = 24$ independent samples) with a horizontal bar for each subgroup indicating the mean. **c, d**, Normalized survival of PF-EPN-A (S15) primary cultures (**c**) and EP1-NS cell cultures (**d**) following shRNA knockdown of *IGF2BP1* with two independent non-overlapping shRNA constructs compared to shCONTROL.1. Experiments performed as six technical replicates and independently validated in three biological replicates. Horizontal bars indicates mean values. **e**, H3K27ac profiles at the ependymoma-specific super enhancer locus *CACNA1H* in the Heidelberg cohort with at least three biological replicates per subgroup, with the exception of ST-EPN-SE, which is shown as a biological duplicate. **f**, H3K27ac profiles surrounding the *CACNA1H* locus in a ST-EPN-RELA model (EP1-NS), a PF-EPN-A model (S15) and a normal neural stem cell control performed in biological duplicates. **g**, *CACNA1H* gene expression derived from RNA-seq data for the Heidelberg cohort ($n = 24$ independent samples) with a horizontal bar for each subgroup

indicating the mean. **h, i**, Normalized survival of PF-EPN-A (S15) primary cultures (**h**) and EP1-NS (**i**) cell cultures following shRNA knockdown of *CACNA1H* with two shRNA constructs compared to shCONTROL.1. Experiments performed as four technical replicates and independently validated in three biological replicates. Horizontal bars indicate mean values. **j**, Normalized cell survival of EP1-NS, S15, and NSC194 cells treated with increasing concentrations of mibefradil. Shown are technical triplicates, results replicated in biological triplicates. **k**, Overlay of ATAC-seq and H3K27ac-seq data centred upon ATAC-seq peak regions identified in the ST-EPN-RELA cell culture EP1-NS. **l**, CRISPR-dCAS9 targeting of *CACNA1H* active enhancers impairs *CACNA1H* expression. H3K27ac-seq (top) and ATAC-seq (bottom) surrounding the *CACNA1H* locus, indicating regions targeted by CRISPR-dCAS9 sgRNA complexes. Region 1 (R1) indicates a negative control region devoid of H3K27ac (green), while regions 2–4 (R2–R4) indicate experimental regions under evaluation. Experiments replicated in biological duplicates. **m**, Gene expression for various sgRNA constructs relative to a ‘dummy’ targeting control (D103), negative control (green), and uninfected control. All group comparisons were made using a two-sided Wilcoxon rank-sum test; error bars show s.d. and horizontal bars indicate mean value. Experiments were replicated in biological triplicates.



Extended Data Figure 9 | Validation of ependymoma transcription factors. **a, b**, Gene expression of ‘high activity’ transcription factors (ranked <50) (**a**) and ‘low activity’ transcription factors (ranked >50) (**b**) in ependymoma ($n = 83$ independent samples) versus normal brain tissue ($n = 172$ independent samples). Box plots showing median value

(horizontal bar), interquartile range and dotted line representing the data range. Comparison between groups was assessed using a two-sided Wilcoxon rank-sum test. **c**, Constituent enhancer activity in the central nervous system (CNS) of developing zebrafish embryos derived from subgroup-specific super enhancers identified in ependymomas.



Extended Data Figure 10 | Putative cell lineage programs of origin uncovered by transcription factor mapping. **a–c**, Immunohistochemical staining of Foxj1 at day 13.5 of mouse embryonic development (E13.5). Staining in discrete regions encompassing the choroid plexus and ependymal layer are shown in the forebrain (**b**) and hindbrain (**c**). **d**, \log_2 normalized gene expression of FOXJ1 in ependymoma ($n = 83$ independent samples) compared to independent sample cohorts of the following tissue types: normal brain ($n = 172$), paediatric glioma ($n = 53$), glioblastoma ($n = 84$), atypical rhabdoid teratoid tumours ($n = 18$), medulloblastoma ($n = 62$) and pilocytic astrocytoma ($n = 41$). Horizontal bar indicates the mean value. **e**, Subgroup-specific gene expression of FOXJ1 derived from ref. 1 ($n = 209$ independent samples). Error bars

indicate s.d. and interquartile range; horizontal bar indicates median. **f**, Gene set enrichment analysis³⁰ demonstrating significant enrichment of the FOXJ1 transcriptional program derived from E14.5 mouse embryos specifically in PF-EPN-B tumours ($n = 209$ independent samples). FDR corrected significance evaluated by gene set enrichment analysis. **g**, Significant FOXJ1 gene-expression correlations with proteins known to regulate cilia assembly and function. *P* values for significant positive or negative correlations have been corrected for multiple testing using the Bonferroni method. **h–m**, FOXJ1 gene set enrichment plots of PF-EPN-A (**h**), PF-EPN-B (**i**), PF-EPN-SE (**j**), ST-EPN-RELA (**k**), ST-EPN-YAP1 (**l**) and ST-EPN-SE (**m**) ependymomas. FDR-corrected significance evaluated by gene set enrichment analysis, $n = 209$ independent samples.

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

We determined our sample size based on the inclusion of at least 2 samples per tumor subgroup. We utilized two independent cohorts and studied the intersection of these cohorts to further support the reproducibility and reliability of our datasets.

2. Data exclusions

Describe any data exclusions.

No data was excluded from our study.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

This is most relevant to the animal component of our study in which case animals were randomly assigned to control versus treatment arms in a balanced and blinded fashion with the assistance of an independent lab technician.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Animal experiments were conducted in a single-blinded fashion, and endpoints assessed by an independent animal technician in the lab.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

We used standard MACS1.4 for peak calling, and also the DEseq2 package in R/ bioconductor which are available to the public.

For clustering and differential peak analysis we also used HOMER: <http://homer.ucsd.edu/homer/ngs/peaks.html>.

For super enhancer calling we used ROSE: http://younglab.wi.mit.edu/super_enhancer_code.html.

For core regulatory circuitry analysis we used CRCmapper: https://bitbucket.org/young_computation/crcmapper.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All material in our study is publicly and freely available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

For H3K27ac ChIP-seq we utilized two widely used antibodies for chip-seq in human samples:

1) Abcam (AB4729, Lot#: GR184557). Validation as specific, ChIP grade, and recognition of human tissue can be found here: <http://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729-references.html>

2) Active Motif (39133, Lot#:3184008). Validation as specific, ChIP grade, and recognition of human tissue can be found here: https://www.citeab.com/antibodies/82304-39133-histone-h3k27ac-antibody-pab?utm_campaign=Widget+All+Citations&utm_medium=Widget&utm_source=Active+Motif

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Models used were patient-derived primary ependymoma cultures

b. Describe the method of cell line authentication used.

STR authentication was used for cell line authenticity.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cell lines were routinely tested and confirmed to be free of Mycoplasma contamination using a Mycoplasma detection PCR strategy with positive and negative controls.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly mis-identified cell lines were used.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Female immunodeficient NOD-SCID gamma (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) mice at 5-8 weeks old were used in our study

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Our study involved human research specimens for which we have provided all information made accessible by our human institutional review board in the Supplementary Tables of our manuscript.

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

► Data deposition

1. For all ChIP-seq data:

- a. Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

We have made all of the raw and processed data publicly available for our manuscript such that interested readers may download and reprocess datasets, as well as obtain files that can be visualized easily. We have deposited all raw and processed data pertaining to our manuscript in a public repository available for complete access. Our accession number and link to data can be found here: European Genome Phenome archive under the accession number: EGAS00001002696.

3. Provide a list of all files available in the database submission.

A list of files with detailed sequencing statistics is provided in Supplementary Table 1 - 4.

4. If available, provide a link to an anonymized genome browser session (e.g. [UCSC](#)).

Our raw and processed data is provided within the European Genome Phenome archive, and can be readily viewed with a standard genome viewer such as IGV. Unfortunately, due to data restrictions, we are unable to host a live public session on UCSC.

► Methodological details

5. Describe the experimental replicates.

We have performed H3K27ac ChIP-seq experiments in two independent non-overlapping cohorts using two different antibodies, and assessed the overlap and consistency of data on a cohort-level.

6. Describe the sequencing depth for each experiment.

PCR during library preparation were performed according to recommendations by Rubicon Genomics DNA-seq kit. A minimum of 30 million mapped reads was obtained for each sample (Supplementary Table 3-4). Sequencing in the Heidelberg cohort was performed in a 50bp single-end format. Sequencing in the Toronto-Cleveland cohort was performed in a 100bp single end format. Additional sequencing details can be found in Supplementary Tables 1-4 and Methods section of the manuscript.

7. Describe the antibodies used for the ChIP-seq experiments.

For H3K27ac ChIP-seq we utilized two widely used antibodies for chip-seq in human samples:

1) Abcam (AB4729, Lot#: GR184557). Validation as specific, ChIP grade, and recognition of human tissue can be found here: <http://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729-references.html>

2) Active Motif (39133, Lot#:3184008). Validation as specific, ChIP grade, and recognition of human tissue can be found here: https://www.citeab.com/antibodies/82304-39133-histone-h3k27ac-antibody-pab?utm_campaign=Widget+All+Citations&utm_medium=Widget&utm_source=Active+Motif

8. Describe the peak calling parameters.

Peak calling was performed using default parameters of MACS (version 1.4), setting a p-value threshold of $1e-9$, and utilizing matched WCE control sequences where available.

9. Describe the methods used to ensure data quality.

We utilized a p-value threshold of $1e-9$, and MACS1.4 default Mfold parameters for high confidence peaks used for model building. We obtained greater than 40,000 high-confidence peaks in all samples using these threshold parameters (Extended Data Figure 3).

10. Describe the software used to collect and analyze the ChIP-seq data.

We used standard MACS1.4 for peak calling, and also the DEseq2 package in R/bioconductor which are available to the public.

For clustering and differential peak analysis we also used HOMER: <http://homer.ucsd.edu/homer/ngs/peaks.html>.

For super enhancer calling we used ROSE: http://younglab.wi.mit.edu/super_enhancer_code.html.

For core regulatory circuitry analysis we used CRCmapper: https://bitbucket.org/young_computation/crcmapper.