Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR

Claudia L Kleinman1,2,17, Noha Gerges2,17, Simon Papillon-Cavanagh1, Patrick Sin-Chan3, Albena Pramatarova1, Dong-Anh Khuong Quang2, Véronique Adoue1, Stephan Busche1, Maxime Caron1, Haig Djambazian1, Amandine Bemmo1, Adam M Fontebasso1, Tara Spence3, Jeremy Schwartzentuber1, Steffen Albrecht5, Peter Hauser6, Miklos Garami6, Almos Klekner7, Laszlo Bognar7, Jose-Luis Montes8, Alfredo Staffa1, Alexandre Montpetit1, Pierre Berube1, Magdalena Zakrzewska9, Krzysztof Zakrzewski10, Pawel P Liberski9, Zhifeng Dong11, Peter M Siegel11, Thomas Duchaine12, Christian Perrotti13, Adam Fleming14, Damien Faury14, Marc Remke15, Marco Gallo15, Peter Dirks15, Michael D Taylor15, Robert Sladek12, Tomi Pastinen1, Jennifer A Chan13, Annie Huang3,16,18, Jacek Majewski1,2,18 & Nada Jabado2,4,18

Embryonal tumors with multilayered rosettes (ETMRs) are rare, deadly pediatric brain tumors characterized by high-level amplification of the microRNA cluster C19MC1,2. We performed integrated genetic and epigenetic analyses of 12 ETMR samples and identified, in all cases, C19MC fusions to TTYH1 driving expression of the microRNAs. ETMR tumors, cell lines and xenografts showed a specific DNA methylation pattern distinct from those of other tumors and normal tissues. We detected extreme overexpression of a previously uncharacterized isoform of DNMT3B originating at an alternative promoter that is active only in the first weeks of neural tube development. Transcriptional and immunohistochemical analyses suggest that C19MC-dependent DNMT3B deregulation is mediated by RBL2, a known repressor of DNMT3B4,5. Transfection with individual C19MC microRNAs resulted in DNMT3B upregulation and RBL2 downregulation in cultured cells. Our data suggest a potential oncogenic re-engagement of an early developmental program in ETMR via epigenetic alteration mediated by an embryonic, brain-specific DNMT3B isoform.

Brain tumors are currently the leading cause of cancer-related mortality and morbidity in children, with neoplasms of embryonal origin representing the most common group. The World Health Organization has divided these embryonal tumors into three broad entities comprising medulloblastomas, atypical teratoid rhabdoid tumors and central nervous system (CNS) primitive neuroectodermal tumors (PNETs). ETMRs (named for their histology)6 are a rare, newly suggested entity within PNETs6,7 that mainly affect infants, with the vast majority of affected individuals dying within 2 years of diagnosis, despite intensive therapeutic intervention8,9. Most of these tumors are characterized by high-level amplification of the polycistronic microRNA (miRNA) cluster C19MC on chromosome 19q13.41 (refs. 1,2), one of the largest miRNA clusters that spans ~100 kb and contains 46 miRNA genes10. Transcriptional signatures indicate that ETMRs may originate from an early neural stem cell precursor, and high LIN28A and low OLIG2 levels have recently been proposed as surrogate diagnostic markers8,9. However, there is insufficient information to improve disease management and a crucial need to identify relevant targets for the design of novel therapeutic agents.

C19MC amplification is accompanied by a striking overexpression of oncogenic miRNAs that cannot be explained solely by the copy number aberration1. To decipher the molecular pathogenesis of ETMR, we performed an integrated genetic and epigenetic analysis

1McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. 2Department of Human Genetics, McGill University, Montreal, Quebec, Canada. 3Division of Hematology-Oncology, Arthur & Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. 4Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada. 5Department of Pathology, McGill University Health Centre, Montreal, Quebec, Canada. 6Second Department of Paediatrics, Semmelweis University, Budapest, Hungary. 7Department of Neurosurgery, Medical and Health Science Center, University of Debrecen, Debrecen, Hungary. 8Division of Neurosurgery, Department of Surgery, Montreal Children’s Hospital, McGill University Health Centre, Montreal, Quebec, Canada. 9Department of Molecular Pathology and Neuropathology, Medical University of Lodz, Lodz, Poland. 10Department of Neurosurgery, Polish Mother’s Memorial Hospital Research Institute, Lodz, Poland. 11Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, Quebec, Canada. 12Department of Biochemistry, McGill University, Montreal, Quebec, Canada. 13Department of Pathology & Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada. 14Division of Pediatric Hematology-Oncology, Department of Pediatrics, McGill University and the McGill University Health Centre Research Institute, Montreal, Quebec, Canada. 15Division of Neurosurgery, Arthur & Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. 16Program in Cell Biology, Arthur & Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada. 17These authors contributed equally to this work. 18These authors jointly directed this work. Correspondence should be addressed to N.J. (nada.jabado@mcgill.ca), J.M. (jacek.majewski@mcgill.ca) or A.H. (annie.huang@sickkids.ca).

Received 15 June, accepted 13 November; published online 8 December 2013; doi:10.1038/ng.2849
using deep sequencing and global DNA methylation arrays of primary tumor samples (Supplementary Table 1), which allowed us to detect the downstream effects of this amplification. Exome sequencing and copy number variant (CNV) analysis of four ETMR samples, including a recurrence (Supplementary Table 1), confirmed the presence of the amplicon and allowed us to map its boundaries to the TTYH1 locus in all cases (Fig. 1a), suggesting a fusion between this gene and the C19MC cluster. We next sequenced the transcriptome of 12 ETMR samples (including 2 recurrences) and 5 PNET control samples and validated the presence of a fusion between TTYH1 and the miRNA cluster exclusively in all ETMR samples. We determined the genomic locations of the fusion breakpoints at single-base resolution, with consistent results using three different alignment strategies (Supplementary Fig. 1). Each breakpoint was unique to a given sample (Supplementary Fig. 2), falling 5.5–16 kb upstream of the C19MC cluster and occurring either within introns (7/8 cases) or exons (1/8 cases) of TTYH1, downstream of its promoter. Sanger sequencing confirmed the presence of the TTYH1-C19MC fusion in all tested ETMR samples. An analysis of the genomic regions surrounding the fusion breakpoints suggested that a shared mechanism, conferred by sequences with the potential to hybridize and form secondary structures, could perturb replication and lead to genomic instability (Supplementary Fig. 3).

Sequence reads displayed an unusual distribution across the chimeric gene (Fig. 1b–d) that shed light on the mechanism of expression and processing of the fused transcript. We found an anomalous proportion of intronic reads in this region (50% in ETMRs compared to 21% in PNETs; t-test P = 0.017) that did not correspond to the proportion observed across the rest of the genome (for example, there was no difference in the frequency of intronic reads for GAPDH or ACTB when comparing the tumor types; P > 0.5), suggesting that the processing of miRNAs from their primary transcripts interferes with proper splicing and the usually rapid degradation of introns that follows. Indeed, the substantial amount of intronic reads allowed us to visually confirm the locations of the breakpoints (Supplementary Fig. 2). Small RNA sequencing showed expression of the C19MC miRNAs at levels 150–1,000 times higher than in control PNET samples (Supplementary Fig. 4 and Supplementary Table 2). Sharp decreases in RNA sequencing (RNA-seq) coverage were observed where the miRNAs are processed out of their primary transcripts, overlapping the exact locations where the mature miRNAs are detected by small RNA sequencing (top rows) suggesting that both regions are driven by the same promoter. Next, we confirmed the presence of transcripts extending from the first exon of TTYH1 up to the first miRNAs in C19MC by targeted long-read

---

**Figure 1** Genomic regions involved in the rearrangement leading to ETMR. (a) Amplified region of chromosome 19 detected by CNV analysis of exome sequencing samples. (b) TTYH1 gene. Top rows show RNA-seq reads mapping to TTYH1 for one ETMR sample (pink) and one control PNET sample (green); the black arrow marks the breakpoint location in the ETMR sample. The ETMR sample shows unusual accumulation of intronic reads up to the breakpoint. The bottom row shows the TTYH1 gene model. (c) C19MC locus. Purple triangles mark the locations of miRNA genes. Accumulation of reads corresponding to the fused transcript is shown for two ETMR samples. The top row corresponds to sample ETMR1, which has a 14-kb deletion in the C19MC locus. The bottom row depicts a typical ETMR sample. RNA-seq data for all other samples are provided in Supplementary Figure 2. (d) Alignment to an *in silico* reconstruction of the fusion, with reads spanning the inferred breakpoint. Chimeric reads are colored to illustrate the regions mapping to TTYH1 (orange) and CM19C (red). A deletion of two nucleotides is observed at the breakpoint for this sample, which was confirmed by Sanger sequencing. (e) Mature miRNAs are detected by small RNA sequencing (top rows) where RNA-seq coverage decreases sharply (bottom rows), indicative of proper miRNA processing. Gene models for the mature and precursor miRNAs are shown. (f) Long-range PCR targeted amplification followed by PacBio sequencing was performed on total RNA extracted from the xenograft X-ETMR8. Reads were aligned to a reference genome that combines human genome hg19 and an *in silico* reconstruction of the TTYH1-C19MC fusion. Top, schematic of the TTYH1 fusion; the first two miRNAs of the C19MC cluster are indicated. Bottom, long reads (>1 kb) mapping to the fusion are shown in blue; each line represents a single read. Solid boxes represent mapping regions; gaps are represented by lines joining the boxes.
sequencing using the PacBio platform (Fig. 1f and Supplementary Fig. 6). Taken together, these results argue that the TTYH1 promoter drives expression of the C19MC cluster, explaining the strikingly high expression levels of C19MC miRNAs in ETMRs.

TTYH1, a member of the Tweety family, encodes a chloride channel restricted to neural tissue with an ill-defined role in neuron physiology. We analyzed TTYH1 expression in more than 600 samples covering 62 cell types, compiling expression data from the BrainSpan, Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics projects. TTYH1 was expressed in embryonic stem cells and in most neural structures analyzed, both during development and in adult brain, whereas limited expression was detected in other cell types and tissues (Supplementary Fig. 7). Thus, a gene fusion between the TTYH1 promoter and the C19MC cluster would account not only for the high expression levels of C19MC miRNAs in ETMRs but also for the brain specificity of tumors harboring the C19MC amplification.

miRNAs could help modulate the epigenome and regulate global DNA methylation patterns. We therefore analyzed ETMR samples using the Illumina 450K DNA methylation array and compared the methylation patterns obtained to those of 223 samples comprising normal brain and 16 other tumor types, including embryonal tumors (Fig. 2 and Supplementary Table 3). Hierarchical clustering by methylation levels consistently identified ETMR samples as a distinct group (Fig. 2a and Supplementary Fig. 8). Variance-ranked clustering was extremely robust, with as few as 100 individual sites sufficient to properly group ETMR samples (Supplementary Fig. 8c). This specific methylation pattern for ETMRs was further validated by randomly resampling sites (P < 0.01), indicating a distinctive genome-wide alteration in DNA methylation in ETMRs that is not restricted to a few discrete regions (Supplementary Fig. 9). Moreover, this methylation profile was retained in cell lines and mouse xenografts derived from tumors included in this study (Fig. 2a), arguing for a strong association of the miRNA cluster not only with tumor formation, as previously suggested, but also with global changes in DNA methylation.

Given the unique DNA methylation profile, we focused our exome and transcriptome analyses on genes directly involved in establishing and maintaining chromatin structure. Very few of these genes showed deregulated expression in ETMRs (Supplementary Table 4).
Exome sequencing identified few nonrecurrent mutations in individual samples (Supplementary Table 5), which occurred in genes not known to affect DNA methylation (including mutations in VANGL1, involved in neural tube development, and genes involved in the WNT pathway, Fzd6 and Cttnnb1, each occurring in one sample). Transcriptome analysis identified altered expression of only one gene known to affect DNA methylation—DNMT3B, a de novo DNA methyltransferase gene involved in early neural crest specification and progenitor cell development\(^1\),\(^6\)—which was significantly overexpressed in ETMRs compared to PNETs (\(P = 5 \times 10^{-5}\)) and other pediatric brain tumors. Notably, in ETMRs, transcription of Dnmt3B was found to start at an internal alternative promoter, resulting in the inclusion of exon 1B, with a predicted addition of 11 amino acids to the protein (Fig. 2g and Supplementary Figs. 10 and 11). This exon and the protein-coding sequence contained within it are evolutionarily conserved across placental mammals, suggesting that it has a conserved function.

We assessed Dnmt3b expression in The Cancer Genome Atlas (TCGA) data set of 5,445 tumor samples and normal tissues of fetal and adult origin (Fig. 2b,c and Supplementary Table 6a). ETMRs had significantly higher Dnmt3b mRNA expression levels than any other tumor type, except for a subgroup of acute myeloid leukemia, where high expression of this gene was recently associated with poor prognosis\(^7\). Strikingly, we did not detect inclusion of Dnmt3b exon 1B in any tumor type except ETMR (Fig. 2c) nor in any single normal tissue from the ENCODE or Roadmap Epigenomics projects (Supplementary Fig. 12). Analysis of the chromatin state at the Dnmt3b locus showed that promoter 1A, which is used in the transcription of the vast majority of Dnmt3b isoforms, was demethylated (Supplementary Fig. 13) and surrounded by open chromatin in almost all cell types studied, both fetal and adult (Supplementary Fig. 14). In contrast, the alternate promoter 1B, which is used in exon 1B transcription, was silenced by methylation (Fig. 2d) and located in a region of closed chromatin that was open almost exclusively in fetal brain (Supplementary Fig. 14). We observed DNA demethylation at this promoter locus only in ETMRs and fetal brain (Fig. 2d), with methylation increasing with the gestational age of the samples and paralleling decreased Dnmt3b exon 1B production (Supplementary Fig. 15). These data suggest that stable silencing of the alternate promoter, which normally occurs in fetal and adult tissues, is absent in ETMRs. As the only tissue type where promoter 1B was unmethylated and surrounded by open chromatin was fetal brain, we next profiled Dnmt3b expression in 580 samples from the BrainSpan Project covering 16 CNS structures across the full course of human brain development, (Supplementary Table 6b) finding exon 1B expression to be restricted to a narrow developmental window during early neurogenesis, around 8 weeks after conception (Fig. 2e,f).

Dnmt3b is essential for key events in embryonic differentiation\(^8\), and its depletion leads to early arrest of development\(^9\). In embryonic stem cells, it is an integral component of a regulatory feedback loop involving Oct4, Sall4, Nanog and the miR-290–295 cluster, and its transcription is repressed by retinoblastoma-like 2 (Rbl2)\(^10\),\(^11\). The mRNA for this repressor is predicted by five different algorithms (Supplementary Table 7) to be targeted by miRNAs in C19MC that bear an identical sequence seed to members of the miR-290–295 cluster. We thus investigated Rbl2 and Dnmt3b expression levels in ETMR samples and PNET control samples (Fig. 3), both at the mRNA (RNA-seq and quantitative PCR) and protein (immunohistochemical staining) levels. Consistent with targeting by C19MC miRNAs, Rbl2 transcript levels were lower in ETMRs compared to PNETs (by twofold; Fig. 3b and Supplementary Fig. 16), and Rbl2 protein levels were undetectable in ETMRs (Fig. 3a) (protein expression was absent in tumor cells in all six ETMRs and present in all four PNETs examined; \(P = 0.005\), Fisher’s exact test). Combined transcription with three members of the C19MC cluster, selected on the basis of their prediction to target Rbl2 by four distinct algorithms (miR-519a) or their previous description as oncogenic in ETMR\(^1\) (miR-520g), plus an additional miRNA overexpressed as part of the C19MC cluster (miR-517c), of human neural stem cells (hNSCs) decreased Rbl2 expression (Fig. 3d) and was inversely correlated with increased Dnmt3b levels (Fig. 3a,b and Supplementary Fig. 16). Stable individual overexpression of each oncogenic miRNA in PFSK-1, a human PNET cell line, showed that the miRNAs predicted to target Rbl2 cause a decrease in Rbl2 levels.

---

**Figure 3** Expression analysis of Rbl2 and Dnmt3b in patient samples, an ETMR-derived xenograft and neural stem cells transfected with C19MC miRNAs. (a) Immunohistochemistry performed on ETMR and PNET tumor samples, confirming Rbl2 downregulation and Dnmt3b induction in ETMRs. Insets show magnified views of the boxed regions. Scale bars, 100 µm (insets, 20 µm). (b) Quantitative PCR performed on an ETMR tumor sample, an ETMR-derived xenograft and a PNET tumor sample; expression is normalized to the combined expression levels for Gapdh and Actb. Error bars, s.d. from three replicates, calculated from the sum of the weighted variances of the components (\(C_v\) value and PCR efficiency). (c, d) Changes in expression of Rbl2, Dnmt3b, Dnmt3a and Dnmt3d after transfection with empty vector control (pcDNA or pcDH) or a combination of members of the C19MC cluster in hNSCs (c) or with each of these constructs individually or two additional oncogenic miRNAs in the PNET cell line PSFK-1 (d). Both cell lines lack endogenous expression of C19MC\(^1\). Expression levels are normalized to those for Actb. Error bars, s.d. from three replicates. Significance was determined by unpaired two-tailed Student’s t test. *\(P < 0.01\) (c); *\(P < 0.05\) (d). (e) Proposed model of ETMR tumorigenesis.

---

NOTES

---

4 ADVANCE ONLINE PUBLICATION NATURE GENETICS
and concurrent substantial increase in DNMT3B levels (Fig. 3c). These findings were further validated by stable overexpression in PFSK-1 cells of another miRNA predicted to target RBL2 (miR-512-3p) and one not predicted to target it (miR-517-a) (Fig. 3d). Overexpression of miR520g, the oncogenic miRNA previously investigated in ETMR, did not lead to changes in RBL2 or DNMT3B levels in cells, suggesting that it is not likely to be involved in ETMR tumorigenesis, as further inferred by its deletion in an ETMR sample in our data set (Fig. 1d and Supplementary Fig. 4). Notably, levels of DNMT1, a maintenance DNA methylationase, and of DNMT3A, a second de novo methyltransferase also shown to be highly expressed during brain development, remained unchanged following overexpression of the three combined miRNAs or of the five individual ones tested. Exon 1B was not present in overexpressed DNMT3B, a fact we attribute to the cell of origin, that is, postnatal brain-derived neural stem cells and the PNET cell line. RBL2 targeting by specific miRNAs in C19MC is thus one plausible mechanism leading to alternate DNMT3B promoter use.

miRNAs regulate gene expression and diverse biological processes and are increasingly associated with human disease, including cancer. C19MC,21,22, the largest miRNA cluster found in the human genome, is specific to primates, imprinted and expressed mainly during the first trimester of gestation.10,20 Abnormal expression of some members of this cluster has been associated with parathyroid, breast and liver cancers and with thyroid adenomas.21–23 DNMT3B overexpression, on the other hand, has been associated with a number of cancers.24–27 and suppresses the WNT/β-catenin pathway, previously shown to be abnormally regulated in ETMR, whereas hypomorph germline DNMT3B mutations are responsible for ICF syndrome, a genetic disorder associated with defects in neurogenesis.15,28 The expression of TTYH1, C19MC and exon 1B–containing DNMT3B peaks 8–10 weeks after conception, which may suggest an early event leading to ETMR development. We propose a model of tumorigenesis (Fig. 3e) in which a fusion between TTYH1 and C19MC leads to extreme overexpression of the miRNA cluster; the effect, potentially mediated by RBL2, would be an abnormal maintenance of chromatin state at promoter 1B of DNMT3B, leading to the incorporation of exon 1B, an alternate exon 1 specific to early brain embryogenesis. ETMR tumorigenesis in this model would thus be promoted by the ‘reawakening’ or aberrant maintenance of an early neurogenesis developmental pathway throughout brain development and after birth, as TTYH1 continues to be expressed at later stages in the brain. If this model is confirmed, DNMT3B would represent a clear candidate for future therapies targeting these deadly tumors, which have been described by pathologists as resembling undifferentiated neural tubes.

ACKNOWLEDGMENTS

This work was supported by the Cancer Research Society and was funded in part by Hungarian Scientific Research Fund (OTKA) contract T-04639, by National Development and Research Fund (NKFP) contracts 1A/002/2004 (P.H., M. Garami and L.B.) and TAMOP-4.2.2A/11-1/KONV-2012-0025 (A.K. and L.B.) and by Canadian Institutes of Health Research grant 102684 (A.H.). N.J. is a member of the Penny Cole laboratory and is the recipient of a Chercheur Clinicien Senior Award. S.B. is supported by an RMGRA (Rezseau Médical de Génétique Appliquée; Network of Applied Genetic Medicine) fellowship. T.P. and J.M. hold Canada Research Chairs (tier 2). C.L.K. is the recipient of a fellowship award from the Fonds de Recherche du Qu’bec – Santé, N.G. is the recipient of a studentship award from Cedars, and A.M.F. is the recipient of a studentship from the Canadian Institutes of Health Research and the McGill–Canadian Institutes of Health Research Training Program in Systems Biology. R.S. is supported by the Canadian Institutes of Health Research.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.


METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Exome sequencing and small RNA sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession SRP032767. RNA-seq data have been deposited in SRA under accession SRP032476. Methylation data have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession GSE52556.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ONLINE METHODS
Characteristics of the samples and pathological review. All samples were obtained with informed consent. After approval was obtained from the institutional review boards of the respective hospitals, samples were treated and independently reviewed by senior pediatric neuropathologists (S.A. and M.Z.) according to World Health Organization guidelines. The clinical characteristics of the patients are summarized in Supplementary Table 1a. Samples were taken at the time of the first surgery before further treatment, except for two samples corresponding to relapses.

Exome sequencing and CNV analysis. Standard instructions from the manufacturer were used for target capture with the Illumina TruSeq exome enrichment kit and 100-bp paired-end sequencing on the Illumina HiSeq 2000 platform, with bioinformatic processing and variant annotation as previously described29. To look for CNVs, we used a CNV detection algorithm that prioritizes rare variants by using a large set of exome samples as controls and comparing the exonic read depth of the test sample against the distribution of read depths for the control set30. We first obtained coverage information for 200 controls and the test sample in RPMK and eliminated batch-to-batch variation in the control set by applying principal-component analysis (PCA) and removing the top contributing principal components (~2–10) from the data. Then we segmented the test sample into regions of similar RPMK ratios using circular binary segmentation (CBS)31. Finally, we calculated the statistical deviation of the test sample from the distribution of the controls and generated a P value for each segment, controlling the false discovery rate with the Benjamini-Hochberg procedure. Although the C19MC cluster was not included in the exome capture kit, all neighboring genes were, and, for each ETMR case, the breakpoint was inferred to fall within the TTYH1 gene.

RNA sequencing. Brain samples were homogenized in nitrogen using a mortar and pestle. Brain powder was immersed in TRIzol (Invitrogen) and subjected to a second round of homogenization using a POLYTRON instrument for 30 s. Total RNA was extracted using the miRNeasy kit (including a step of DNase I treatment), and RNA quality was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries for RNA-seq and small RNA sequencing were prepared according to Illumina TruSeq protocols (with the Gold option and strand-specific preparation for RNA-seq samples). Size selection (130–170 bp) of libraries for small RNA sequencing was performed using 3% gel cassettes (Sage Science). The quality of the libraries was assessed using an Agilent 2100 Bioanalyzer. Samples were indexed (three or four per lane for RNA-seq and ten per lane for small RNA sequencing) and sequenced on an Illumina HiSeq 2000 (100-bp paired-end and 50-bp single-end reads for RNA-seq and small RNA sequencing, respectively).

Sequencing runs were processed with Illumina CASAVA 1.8 software. Reads were trimmed using Trimmomatic v.0.22 (ref. 32), removing low-quality bases at the ends of reads (phred33 < 30) and clipping Illumina adaptor sequences using the palindrome mode in Trimmomatic. Reads shorter than 32 bp after trimming (10 bp for small RNA sequencing) were discarded. Resulting strand-specific, high-quality RNA-seq reads were aligned to human reference genome build hg19 using TopHat v1.4.1 (ref. 33) coupled with Bowtie v0.12.8 (ref. 34), with UCSC gene model annotations to guide spliced alignment. Picard tools were used to mark duplicated reads. Reads mapping to multiple (more than ten) locations were discarded for downstream analysis. In the case of small RNA sequencing data, alignment to the human reference genome hg19 was performed with Bowtie v2.0.2 (ref. 34), reporting up to ten alternative locations for each read, using a seed length of 15 nucleotides and allowing 1 mismatch within the seed.

Integrative Genomics Viewer35 was used for visualization. Multiple quality control metrics were obtained using FastQC, SAMtools36, BEDtools37 and custom scripts (Supplementary Table 1b). Bigwig tracks for visualization were generated with custom scripts, using BEDtools and UCSC tools.

Gene fusion detection from RNA-seq data. To determine at single-base resolution the breakpoints of the genomic rearrangement leading to the fusion of C19MC and TTYH1, three alignment strategies were used (Supplementary Fig. 1). First, focusing on the amplified regions detected by exome sequencing data, approximate breakpoints were obtained from the paired-end alignment to the reference genome described above by identifying mate pairs where each mate mapped to a different gene (Supplementary Fig. 1a). In ETMRs, these pairs had a an unusually long insert size and inverse orientation (negative insert size found in the 99.5th percentile of the inferred insert size distribution, with mates facing outward). Second, a different alignment algorithm, STAR38, was used to map the sequencing reads to the reference genome. In the seed-finding phase of STAR, a sequential search for a maximal mappable prefix is performed, in which the longest substring of the read that exactly matches a genomic location is found. This approach represents a natural way of finding the precise locations of splice junctions or fusion breakpoints in a read sequence and is more accurate than split-read methods that use arbitrary splitting (Supplementary Fig. 1b). Thus, a large number of reads spanning the breakpoint were correctly mapped and reported as soft clipped. Finally, we performed an alignment using an in silico reconstruction of the chimeric gene as the reference (Supplementary Fig. 1c).

Validation of gene fusion by PacBio sequencing. We reverse transcribed 100 ng of DNase I–treated RNA from the xenograft X-ETMR8 using ThermoScript reverse transcriptase (Invitrogen) and either gene-specific primers or a random hexamer primer. TTYH1-C19MC fusion transcripts were then amplified using a combination of forward primers annealing in exons 1, 3 or 5 of the TTYH1 gene and reverse primers annealing downstream of the fusion breakpoint, either close to the breakpoint or close to the first miRNA gene in the C19MC cluster (Supplementary Fig. 15). PCR runs comprised 40 cycles using either Takara PrimeSTAR GXL DNA polymerase (Clontech) or Q5 High-Fidelity DNA polymerase (New England BioLabs).

Each PCR product was prepared for sequencing using 200 ng of input material. First, a polyA tail was added to the 3′ end using a terminal deoxynucleotidyl transferase (TdT). A short polyT sequencing primer was then annealed to this tail, where the P4 PacBio (Pacific Biosciences) sequencing polymerase was subsequently bound, forming the DNA polymerase complex. For each sample, 30 pM of the complex was loaded onto SMRTcells using the PacBio MagBead kit. Libraries were sequenced with 120-min movies for each sample on a PacBio RSII instrument with stage start enabled. Raw sequencing reads were trimmed for quality using SMRT Pipe software from PacBio, with a quality cutoff of 0.75 and removing reads shorter than 50 bp. High-quality reads were then aligned to a reference genome consisting of a combination of human genome hg19 and an in silico reconstruction of the TTYH1-C19MC chimeric sequence. Alignments were performed with BLAT39, requiring at least 85% sequence identity and considering only the best scoring hit for each read, with the score defined as the number of matches minus the number of mismatches.

Validation of gene fusion by Sanger sequencing. Total RNA (500 ng to 1 μg) was used to synthesize first-strand cDNA with 500 ng of random hexamers (Invitrogen) and SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer’s instructions. Genomic DNA was isolated from tissue using the DNaseasy Blood and Tissue kit (Qiagen) according to the manufacturer’s protocols.

 Primer pairs for PCR amplification and sequencing of each breakpoint were generated using Primer3 (ref. 40). Primers were designed to obtain a product of approximately 300 bp, with the fusion breakpoint near the center of the fragment. To minimize the amplification of nonspecific genomic sequences, primer pairs were filtered using UCSC In Silico PCR. PCR products were purified before sequencing by filtration using Millipore MultiScreen filter plates (EMD Millipore Corporation). Products were sequenced by conventional Sanger methods with the BigDye Terminator v3.1 Cycle Sequencing kit (Life Technologies) and purified by ethanol precipitation before obtaining DNA chromatograms on a 3730xl DNA Analyzer (Applied Biosystems).

Analysis of gene expression from RNA-seq data. To estimate gene expression levels, we used all exonic reads mapping uniquely within the maximal genomic locus containing each gene and its known isoforms, normalizing by library size using DESeq41. For clustering and correlation analysis, variance-stabilized expression values were derived using DESeq. Hierarchical clustering was performed using Pearson’s correlation as the distance metric and average linkage as the agglomeration method. Differential expression analysis was...
performed using DESeq, where statistical significance was calculated using the negative binomial distribution as a null distribution for gene expression values, with the variance and mean estimated from the data and linked by local regression. The expression levels obtained for all coding genes can be found in Supplementary Table 8. Genes were subsequently annotated with the DAVID functional annotation tool\textsuperscript{32}, v6.7.

**Analysis of gene expression by quantitative PCR.** RNA was extracted from tissue (around 20 mg) or cell lines (1–2 × 10\textsuperscript{6} cells total) using the RNaseasy Lipid Tissue kit (Qiagen). RNA integrity was assessed with the Experion system (Bio-Rad). We then reverse transcribed 100 ng of total RNA using iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad), following the manufacturer’s instructions. SsoFast EvaGreen Supermix (Bio-Rad) was used for quantitative PCR, starting with 1 μl of cDNA. Each sample was run in triplicate on a LightCycler 96 instrument (Roche). Cycling conditions comprised 95 °C for 30 s, 95 °C for 5 s and 60 °C for 20 s (n = 40 cycles). Product specificity was examined by melting curve analysis. Quantitative PCR analysis was performed using LC96 Application Software 2.0. Results were normalized to expression levels for reference genes (ACTB and GAPDH).

**DNA methylation analysis.** We subjected 500 ng of genomic DNA to bisulfite conversion using the EZ DNA Methylation kit (Zymo Research), according to the manufacturer’s protocols. Modified genomic DNA was processed as described in the Infinitum Assay Methylation Protocol Guide Rev C (November 2010) and analyzed on Infinium HumanMethylation450 BeadChips (Illumina), measuring DNA methylation at single-CpG resolution on the basis of genotyping of C\textsuperscript{G}U polymorphisms. On such a chip, fluorescence intensities are detected in two separate color channels for the methylated (I\textsubscript{M}) and unmethylated (I\textsubscript{U}) alleles. For genomic annotation of probes, RefSeq gene and CGI annotations were downloaded from UCSC hg19 (NCBI Reference Sequence Database Release 37), and miRNA annotations were obtained from miRBase version 17. We removed probes that had missing values, bound multiple genomic regions, bound targets with known sequence variants (1000 Genomes Project, SNPs with a minor allele frequency of ≥2/120) or were located on the X or Y chromosome. The final set contained 309,015 probes.

Intensities from both channels (I\textsubscript{M} and I\textsubscript{U}) were combined into a single β value per locus ranging from 0 (no methylation) to 1 (complete methylation on both alleles), where

\[
\beta = \frac{I_M}{I_M + I_U}
\]

Finally, we used a subset quantile within-array normalization (SWAN) algorithm\textsuperscript{43} to match the intensity peaks of type I and type II probes using the R package minfi\textsuperscript{43}. We detected artifactual samples with the help of quality control graphs and individual density distribution. In total, 16 of 215 samples were removed from further analysis.

Hierarchical clustering was performed using the 10,000 most variant sites from all samples. Distance was computed using 1 − product-moment correlation coefficient between features, and clustering was performed using DESeq, where statistical significance was calculated using the negative binomial distribution as a null distribution for gene expression values, with the variance and mean estimated from the data and linked by local regression. The expression levels obtained for all coding genes can be found in Supplementary Table 8. Genes were subsequently annotated with the DAVID functional annotation tool\textsuperscript{32}, v6.7.

**Overexpression of C19MC members in human neural stem cells and the PFSK-1 cell line.** The 3-miR expression plasmid was generated by designing a cluster of three miRNA precursor stem loop structures in a pCDH-CMV-MCS-4EF1-copGFP vector under a constitutive CMV promoter and was subcloned into the pcDNA3.1 vector (Invitrogen) for expression in hNSCs. Generation of hNSCs with stable expression of 3-miR and propagation of the stable cell line were performed as previously described\textsuperscript{31}. Expression of miRNAs was validated by quantitative RT-PCR using TaqMan probes (Applied Biosystems). Constructs for individual miRNAs were generated similarly and were stably overexpressed individually in PFSK-1 cells. hNSCs were obtained by P. Dirks, and the PFSK-1 cell line (a PNET from a supratentorial tumor from a 22-month-old) was purchased from the American Type Culture Collection (ATCC). Both have been reported previously\textsuperscript{30}. All cell lines used in this study were tested for mycoplasma and were found to be free of this pathogen.

**Immunohistochemistry.** Immunohistochemical staining for DNMT3B and RBL2 was performed on paraffin-embedded sections that were subjected to antigen retrieval in 10 mM citrate buffer (pH 6.0) for 15 min at temperatures below the boiling point. Slides were incubated overnight at 4 °C with a mouse monoclonal antibody to Dnmt3b (Abcam, ab13604) used at a 1:200 dilution or a polyclonal rabbit antibody to RBL2 (Abcam, ab76234) used at a 1:100 dilution. After incubation with primary antibody, secondary biotin-conjugated antibodies were applied to tissue sections for 30 min. After washing with PBS, slides were developed using diaminobenzidine (Dako) as the chromogen. All slides were counterstained with Harris hematoxylin. Slides were independently scored for RBL2 and DNMT3B positivity by three individuals, and the results were merged after consensus scoring was obtained. Briefly, samples were considered to be negative for RBL2 or DNMT3B staining if tumor cells showed no nuclear positivity, and the core included a control brain or reactive glia with positive nuclear staining\textsuperscript{1}.

**ETMR cell lines and xenografts.** Tumor tissue was collected at the time of autopsy from a 2-year-old male with recurrent ETMR and immediately placed in chilled transport medium consisting of 1x PBS with glucose (4.5 g/l), penicillin (50 U/ml) and streptomycin (50 U/ml). After rinsing in transport medium, tissue was transferred to NeuroCult human NS-5 proliferation medium (Stem Cell Technologies) supplemented with heparin (0.0002%); Stem Cell Technologies), epidermal growth factor (EGF; 20 ng/ml; Peprotech), fibroblast growth factor (bFGF; 20 ng/ml; R&D) and sonic hedgehog (recombinant human SHH N terminus, 0.1 μg/ml; R&D), and tissue was mechanically disaggregated into single cells and small clusters of cells by gentle trituration. The cell suspension was passed through a 70-μm filter, and cells were plated in stem cell medium at a density of 20,000 cells/ml in low-adhesion flasks for suspension cells (Sarstedt) and maintained at 37 °C in a 5% CO\textsubscript{2} incubator. Tumor spheres formed within 7–10 d. Cells were passaged immediately after spheres reached 200–400 μm in diameter in xenografts. Spheres were disaggregated to single cells with Accumax (Innovative Cell Technologies) and resuspended in sterile PBS. We injected 100,000 cells in a volume of 2–3 μl into the right striatum of 6– to 8-week-old male NOD-SCID mice (Jackson Laboratory) using the following stereotactic coordinates: −1.0 anteroposterior, 2.0 medial-lateral and 3.0 dorsoventral. Mice were euthanized, and brains were collected for histology and extraction of DNA and RNA 8 weeks after injection. The McGill animal care committee approved the protocols.

**Expression of RBL2.** RBL2 expression was assessed at the RNA level from histologically defined tumor regions. Tumor RNA was purified from formalin-fixed, paraffin-embedded samples using the RNeasy Lipid Tissue kit (Qiagen). mRNA expression levels were normalized to expression levels for ACTB using the ΔΔC\textsubscript{t} method.