Integrative DNA methylation analysis of pediatric brain tumors reveals tumor type-specific developmental trajectories and epigenetic signatures of malignancy

Meeri Pekkarinen*[1], Kristiina Nordfors*[2,3], Joonas Uusi-Mäkelä*[1], Ville Kytölä*[1], Minna Rauhala[3,4], Henna Urhonen[1], Sergei Häyrynen[1], Ebrahim Afyounian[1], Olli Yli-Harja[5,6], Wei Zhang[7], Pauli Helen[4], Olli Lohi[2,3,5], Hannu Haapasalo[5,8], Joonas Haapasalo[3,4,8], Matti Nykter[1], Juha Kesseli[1], Kirsi J. Granberg+[1,9]

Affiliations:
1. Prostate Cancer Research Center, Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Center, Tampere University Hospital, Tampere, Finland
2. Tampere Center for Child Health Research, Tays Cancer Center, Tampere University and Tampere University Hospital, Tampere, Finland
3. Tays Cancer Center, Tampere University Hospital, Tampere, Finland
4. Department of Neurosurgery, Tays Cancer Centre, Tampere University Hospital and Tampere University, Tampere, Finland
5. Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Center, Tampere University Hospital, Tampere, Finland
6. Institute for Systems Biology, Seattle, WA, United States
7. Cancer Genomics and Precision Oncology, Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC, United States
8. Fimlab Laboratories ltd., Tampere University Hospital, Tampere, Finland
9. Tampere Institute for Advanced Study, Tampere University, Tampere, Finland

* Equal contribution
+ Corresponding author: kirsi.granberg@tuni.fi, +358 50 318 5819

Abstract

Understanding oncogenic epigenetic mechanisms in brain tumors is crucial for improved diagnosis and treatment. To evaluate tumor type-specific features, we compared atypical teratoid/rhabdoid tumors (AT/RTs), medulloblastomas, and choroid plexus tumors with each other by integrating DNA methylation (507 samples), gene expression (120 samples), and transcription factor (TF) -binding data. Our results suggest that aberrant DNA methylation plays a vital role in the oncogenesis of AT/RTs, which are known for their aggressiveness and exceptionally low mutation rates. In AT/RT, elevated DNA methylation masks the binding sites of TFs driving neural development and is associated with reduced gene expression for specific neural regulators, whereas DNA methylation patterns predict a more advanced differentiation state for MB. By analyzing the data together with pluripotent stem cells, we found two regulatory programs in AT/RT: pluripotent stem cell -like DNA methylation patterns and AT/RT-specific DNA methylation uniquely associated with polycomb repressive complex 2 (PRC2) members. Both methylation patterns cover neural TF binding sites in a TF-specific manner, suggesting their relevance for oncogenic regulation.

Introduction

Central nervous system (CNS) tumors can arise at any age and have the highest cancer-associated mortality rate in pediatric patients [1,2]. Atypical teratoid rhabdoid tumors (AT/RTs), medulloblastomas (MBs), and choroid plexus tumors (PLEXs) are all detected in infants [3]. AT/RTs and MBs are aggressive, grade IV tumors according to the World Health Organization (WHO) classification [4], and high-grade (III) PLEXs are typically detected in children [4–6]. In particular, AT/RTs, SHH and group 3 subtypes of MBs, and grade III PLEXs have a poor prognosis [7–9]; thus, improving patient outcomes is an urgent task. Because aberrant epigenetic regulation plays a major role in many aggressive pediatric tumors [10], including AT/RTs [11], we wanted to specify the role of DNA methylation in AT/RT, MB, and PLEX. For AT/RT, the significance of epigenetics is also highlighted by their deserted DNA alteration landscape: the sole recurrent genetic alteration is the inactivation of *SMARCB1* or *SMARCA4*, both of which are subunits of SWItch/Sucrose-Non-Fermentable (SWI/SNF) chromatin remodeling complex [12]. Surprisingly, many histone modifications are also depleted in AT/RT [11], suggesting an alternate epigenetic driver of AT/RT development.

DNA methylation is one of the major mechanisms underlying epigenetic regulation. It contributes to cell differentiation [13]; is altered in different malignancies [14–16]; and is, together with histone methylation, a clinically relevant therapeutic target [17,18]. In CNS tumors, DNA methylation is used for accurate tumor classification and early cancer detection [19–21]. DNA methylation-based classification of CNS tumors also reveals tumor type- or subtype-specific DNA methylation patterns, linking DNA methylation to tumor type-specific oncogenesis [21].

Traditionally, researchers have compared and analyzed subtypes of AT/RT, MB, and PLEX separately [22–25]. For example, molecular AT/RT subtypes have been reported to have distinct genomic profiles, *SMARCB1* genotypes, and chromatin landscapes, including DNA methylation differences among the subtypes [22,26]. In this study, we aimed to define DNA methylation patterns and related gene regulation, which are characteristic of each tumor type irrespective of their internal heterogeneity, to gain insight into these tumors and their oncogenic regulation. AT/RT, MB, and PLEX were used for the comparison because they have a partly shared history and characteristics [27–29].

Our analysis revealed elevated DNA methylation patterns in AT/RT, suggesting that the binding of TFs that drive neural differentiation is often inhibited in AT/RTs. In addition, polycomb repressive complex 2 (PRC2) sites were hypermethylated in AT/RTs. We combined gene expression data with DNA methylation and observed upregulation of especially neural genes in MBs when compared with AT/RTs and/or PLEXs. Because the SWI/SNF complex is critical for the targeted opening of chromatin at certain genomic loci during neural development by recruiting EP300, leading to H3K27 acetylation [12,30], we analyzed our results with respect to DNA methylation data from pluripotent stem cells (PSCs) and fetal brain (FB). We found that some sites shared similar methylation patterns in AT/RTs and PSCs. However, these sites are different from PRC2 target sites, suggesting that there are two different DNA methylation programs in AT/RTs.

Results

DNA hypermethylated in AT/RT when compared with MB or PLEX

To study oncogenic epigenetic regulation in AT/RTs, MBs, and PLEXs, we collected genome-wide DNA methylation Illumina microarray data (i450K) from 497 tumors and expression data from 110 tumors. In addition, we generated data from 10 matched tumors with reduced representation bisulfite sequencing (RRBS) and RNA sequencing (**Figure 1A**) (**Supplementary Table 1**), and 89 normal brain DNA methylation samples were used as controls in i450k-based DNA methylation analysis. DNA methylation data separated the tumor types and their subtypes in t-distributed stochastic neighbor embedding (tSNE) analysis (**Figure 1B**). RRBS samples were positioned adjacent to the tumor subgroups that matched their clinical diagnosis [21]. To clarify the role of aberrant DNA methylation in AT/RTs, MBs and, PLEXs, we identified differentially methylated genomic regions between them, transcriptional regulators enriched in these regions, and how differential methylation was associated with gene expression.

Differential methylation analyses resulted in 11,016 and 64,983 differentially methylated regions (DMRs) in the i450k and RRBS data, respectively (**Figure 1C**). Higher DMR counts in RRBS data analysis were predictable because the method has a wider genomic coverage and i450k DMRs were, on average, longer than those from RRBS (always 1kb) [31,32]. Furthermore, tumor location information and normal samples were used to filter out tumor location-related differences in i450k DNA methylation data, which decreased the DMR counts (**Figure 1C** vs. **Supplementary Figure 1A).** There was a higher proportion of intergenic DMRs in RRBS than in i450k data (36%, 30%, and 36% vs. 14%, 10% and, 11% in AT/RT-MB, ATRT-PLEX, and PLEX-MB, respectively) (**Supplementary Figure 1C**, **Supplementary Table 1**). DMRs were distributed across all chromosomes (**Supplementary Figure 1D**).

The majority of DMRs were more methylated in AT/RT than in other tumor types (84% and 72% in i450k and RRBS data, respectively) and less methylated in MB than in other tumor types (77% and 79% in i450k and RRBS data, respectively) (**Figure 1D**). Consistently, the highest DNA methylation levels were detected in AT/RTs and lowest in MBs (**Figure 1E**).

As higher DNA levels were reported in AT/RT-TYR and AT/RT-SHH subgroups [26], we asked whether higher DNA methylation was subgroup-specific. No AT/RT subgroup specificity was observed in our setting; the highest DNA methylation levels were consistently detected in each AT/RT subgroup when compared with the subgroups of other tumor types (Wilcoxon p values < 0.005, effect size Cliff's delta $|\delta| > 0.2$) (**Supplementary Figure 1B**, **Supplementary Table 1**).

Regions inside topologically associating domains (TADs) potentially share similar patterns of epigenetic regulation [33]. Our analysis also found several large-scale DNA methylation differences within TADs (**Figure 1F**). Especially in the AT/RT-MB comparison, there were clusters of large-scale differences in chromosomes 1, 11, 13, 17, and 19. Chromosome 17, which carries many DMRs (**Supplementary Figure 1D**) and harbors recurrent gains and losses in MB [34], is notable, because it harbors 14 large-scale methylation-related TADs, 13 of which were differentially methylated in the AT/RT-MB comparison.

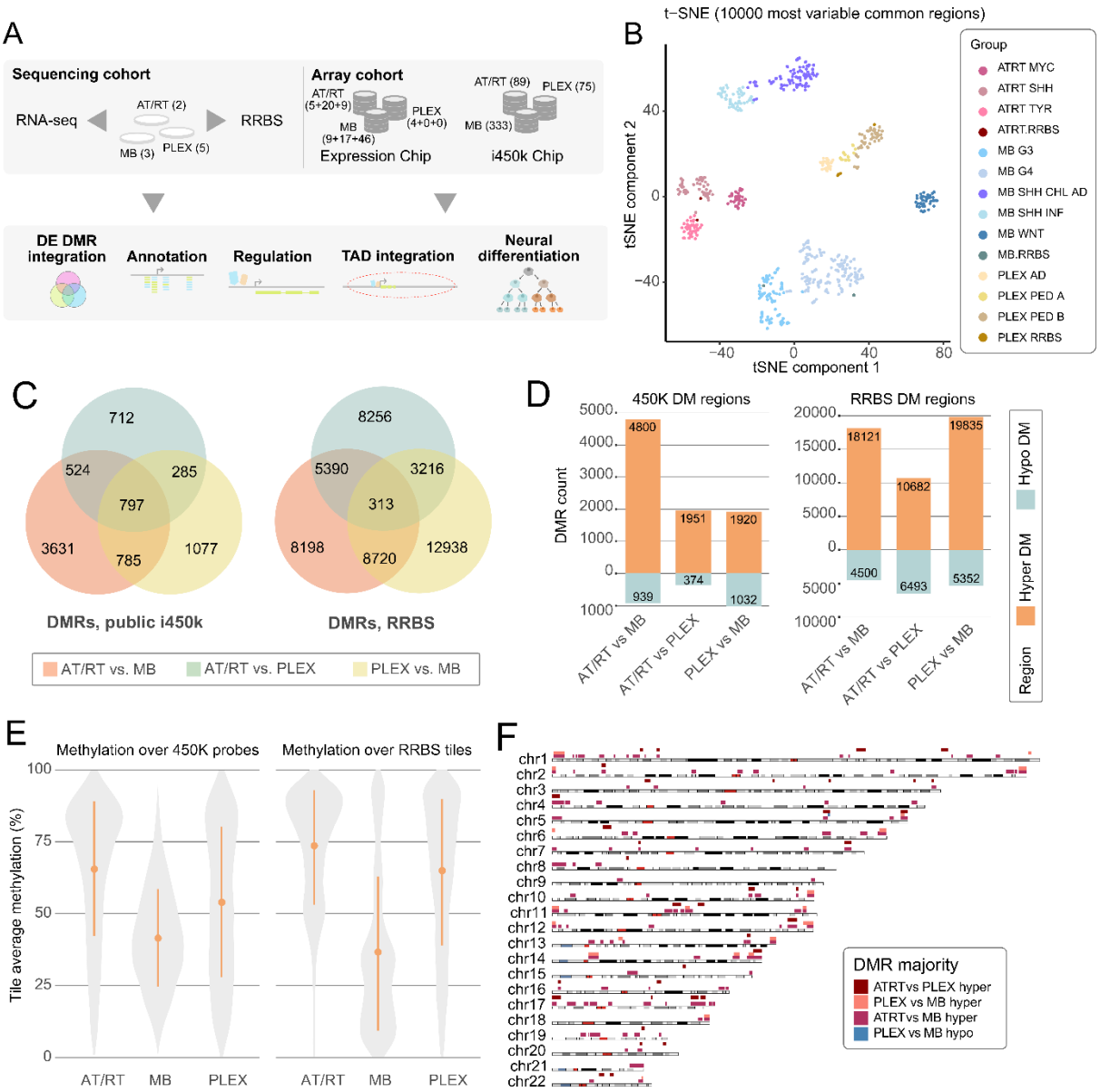

Figure 1: **Characterization of DNA methylation differences among AT/RT, MB, and PLEX.**

a) Illustration of our data analysis and integration approach. The number of samples in each cohort is shown, and also for the three separate expression cohorts analyzed using expression microarrays.
b) Tumor types are separated based on DNA methylation in a tSNE visualization, when the 10 000 most variable regions shared between i450k and RRBS data were used for the visualization. Methylation subgroups of each tumor type (omitted from [21]) in i450k were also visualized.
c) Venn diagram showing the number of differentially methylated regions (DMRs) in each comparison. Tumor type-specific DMRs are marked into the intersecting areas. A higher number of DMRs were detected in the RRBS than in the i450k data. For i450k results, DMRs were filtered using methylation data from normal brain samples.
d) Numbers of hypermethylated and hypomethylated regions for each comparison and for i450k and RRBS data.

e) AT/RT had the highest overall DNA methylation level; MB was the least methylated. Violin plots visualized the distribution of DNA methylation in the 10 000 most variable sites in the RRBS and i450k microarray datasets.
f) Several TADs were influenced by large-scale methylation differences, especially in AT/RT-MB comparison. Karyoplot visualizes the topologically associating domains (TADs) that harbor large-scale DNA methylation differences. Color indicates a comparison in which a difference is observed.

AT/RT hypermethylated sites are enriched with neural differentiation TFs

To gather further information on the function and possible upstream regulation of cancer-specific DMRs, we analyzed which TFs bind to these sites by using the Gene Transcription Regulation Database (GTRD) [35]. Because most TFs do not bind to methylated DNA or closed chromatin [36,37], we presumed that methylation in the TF binding site typically means decreased DNA binding and thus decreased activity for the enriched TF. DMRs, which are concordantly hypermethylated or hypomethylated in one tumor type when compared with the others, were used for the analysis (**Figure 2A**).

Of the 183 enriched TFs, 136 (74%) were unique to one tumor type, 30 (16%) to two tumor types, and 17 (9.3%) to all three (**Figure 2B**). Altogether, 77 TFs were enriched in AT/RT hypermethylated DMRs (adjusted p value < 0.05, one-sided Fisher's exact test) and none in AT/RT hypomethylated sites (**Figure 2B**). In PLEXs, all 47 TFs were enriched in hypomethylated DMRs. For MBs, there were only 10 TFs enriched in hypomethylated DMRs and 114 TFs in hypermethylated DMRs, which is somewhat surprising as a moderate number of DMRs were hypermethylated in MB. We hypothesized that these MB hypermethylated DMRs represent regulatory hubs that comprise several TF binding sites. Indeed, a higher proportion of MB hypermethylated DMRs was located in CpG islands than that for MB hypomethylated, AT/RT hypermethylated, or PLEX hypomethylated DMRs (**Supplementary Figure 2A, Supplementary Table 1**). No tumor subgroup specificity was observed in the DNA methylation patterns of differentially methylated TF binding sites (selected examples in **Figure 2C**).

We used a literature search to functionally annotate the enriched TFs and obtained the annotation for 105 (57%) of them (**Supplementary Table 2**). Themes were chosen based on the knowledge of these tumors, reported TF functions, and recurrent detection of certain gene family members [11,24,25,38]. Notably, 15 of 39 (38%) TFs, which were uniquely enriched in AT/RT hypermethylated DMRs, were linked to neural differentiation and neural cells (**Figure 2B**). These included *e.g.* NEUROD1, ASCL1, and MYCN. Furthermore, four neural differentiation-related TFs (NEUROG2, HAND2, PHOX2B, and PAX7) were enriched in both AT/RT hypermethylated and MB hypomethylated DMRs (**Figure 2B, Supplementary Table 2**). We also noticed that 16 neural differentiation-related TFs (19% of the total 85) were uniquely enriched in sites hypermethylated in MBs. However, five of these (REST, SIN3A, ZEB1, TBX1, and ZHX2) inhibit neural differentiation [39–42]. In addition, six of the remaining 11 TFs were active in the later phases of neural differentiation, whereas seven TFs hypermethylated in AT/RTs were active in earlier stages [43–55]. Furthermore, among the TFs uniquely enriched in MB hypermethylated regions, 9 TFs (11%) were related to the pluripotent stem cell state and

early differentiation (**Supplementary Table 2**) [39,56–58]. Together these findings suggest that DNA methylation probably reduces the DNA binding of neural TFs, and thus neural differentiation, in AT/RT, whereas a more advanced neural differentiation is allowed in MB together with methylation-related silencing of pluripotency-related TF binding sites.

Next, we studied the expression of the enriched TFs in our datasets (**Supplementary Figures 2 and 3, Supplementary Table 3**). ERF was the only TF that had a significantly higher expression in AT/RTs than other tumors (adjusted $p < 0.05$, absolute logFC > 1), whereas 13 TFs, including EBF3 and NEUROD1, had significantly higher expression in MBs than in other tumors (**Figure 2D**, **Supplementary Figures 2, 3, and 4**). The expression of 164 (90%) TFs was not significantly associated with any tumor type (**Supplementary Figures 2 and 3**). We observed that neural differentiation TFs enriched in AT/RT hypermethylated regions were expressed in AT/RTs and that some, such as NEUROG2 and HAND2, had even the highest expression in AT/RTs (**Figure 2D**). Thus, we conclude that DNA methylation changes in TF binding sites affected the function of these TFs independently of their gene expression.

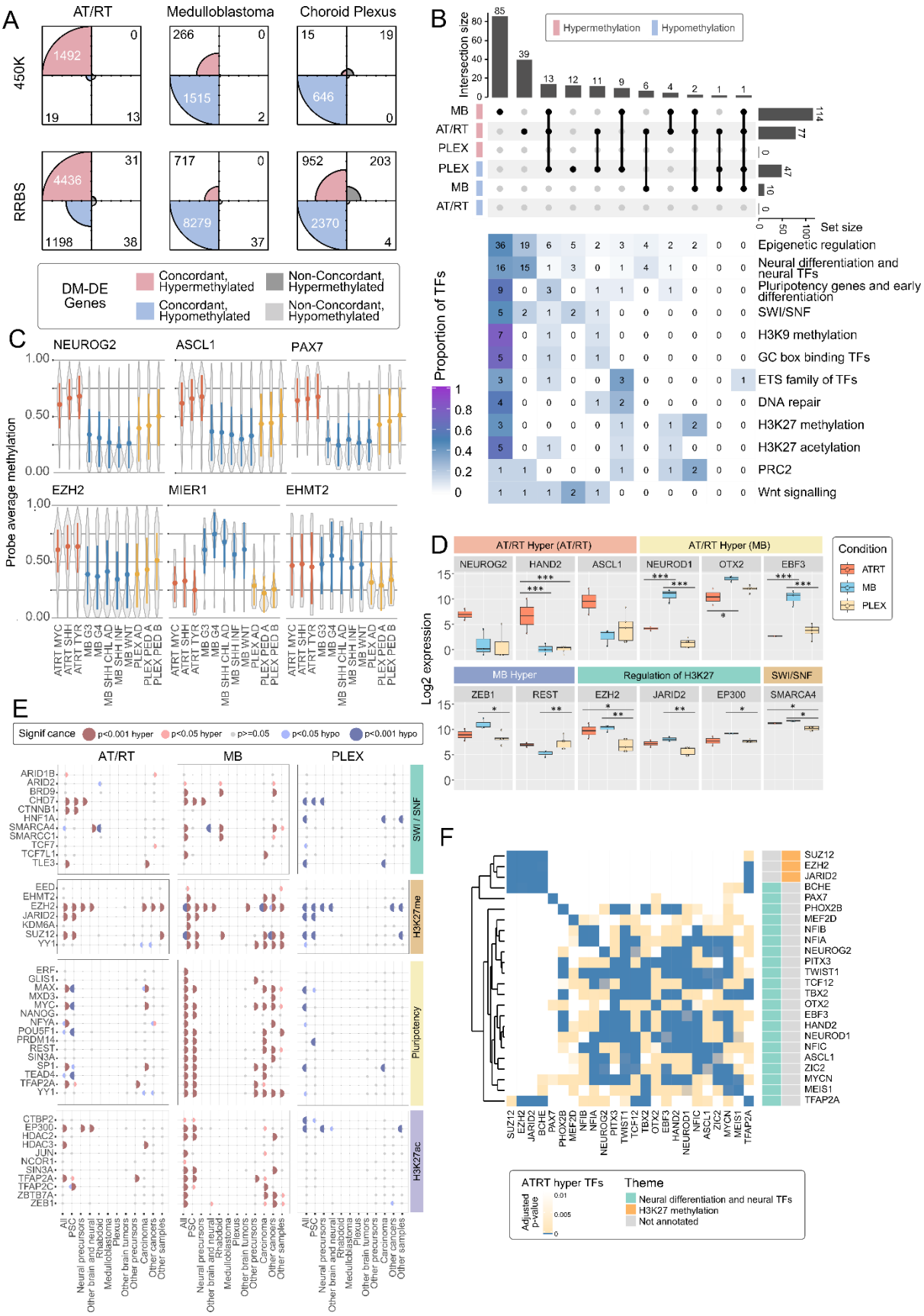

Figure 2: Differential DNA methylation is linked to upstream regulation and to altered chromatin accessibility for differentiation-related TFs.

a) The majority of AT/RT DMRs were hypermethylated (99% and 79% in i450K and RRBS data, respectively), whereas hypomethylated DMRs were more commonly observed in MB (85% and 88% in i450K and RRBS data, respectively) and PLEX (98% and 71% in i450K and RRBS data, respectively). Four-field plots for each tumor type show the number of concordant and non-concordant hypermethylated and hypomethylated regions. Concordant means that DMRs are either hypermethylated or hypomethylated in the specified tumor type when compared with two other tumor types, whereas an opposite difference was observed for non-concordant DMRs. Concordant DMRs were defined as tumor-specific, differentiating each tumor type from the others.
b) The majority of TFs are specifically enriched in AT/RT hyper, MB hyper, or PLEX hypo DMRs and are linked to relevant cellular functions. Upper part: Upset plot of the number of enriched TFs for tumor-specific hypermethylated and hypomethylated regions. Some TFs were associated with several DMR groups. Lower part: TFs were annotated into themes based on their reported functions. The number of TFs in a given theme for each upset plot column is shown. The color of the heatmap shows the fraction of TFs in each theme (row).
c) Binding sites of neural differentiation factors NEUROG2, ASCL1, and PAX7 are more methylated in AT/RT, irrespective of the tumor subtype. Out of TFs linked to histone lysine methylation, EZH2 (involved in histone H3 lysine 27 trimethylation) behaved similarly, but a distinct DNA methylation pattern was observed for MIER1 and EHMT2 (involved in histone H3 lysine 9 methylation) binding sites. Violin plots visualizing the distribution of DNA methylation with selected TF binding sites in all i450k DMRs. The tumor subgroups are presented separately.
d) Some TFs enriched in DMRs hypermethylated in AT/RT (AT/RT hyper) showed differential expression between tumor types. The expression of selected TFs enriched in tumor-specific DMRs are visualized with boxplots (*,**, and *** refer to $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively).
e) The cell-type-specific binding patterns provide information on the epigenetic state of TF regulomes in the analyzed tumors. GTRD data were split based on the sample of origin in the listed categories (at the bottom), and TF binding site enrichment was calculated for each of them separately. TFs were organized into the themes listed in Figure 2B. The dot is not marked when a given TF is not measured in a given GTRD category.
f) PRC2 subunits rarely co-localize with neural TFs in AT/RT hypermethylated sites. Heatmap visualization of the enrichment p-value (one-sided Fisher's exact test) for colocalization. All the adjusted p values of 0.05 or higher are marked in white. TF themes for each TF were annotated on the right-hand side of the heatmap.

Analysis of split GTRD data underlines the epigenetic states of neural cell differentiation in AT/RT and MB

GTRD information is from various cells of origin [35]; thus, we analyzed TF binding data in detail by considering the sample source. We divided GTRD into 11 categories based on the cell of origin (**Figure 2E, Supplementary Figure 5, Supplementary Table 2**). Using this split GTRD

data, we found 15 neural differentiation and neural cells-related TFs measured in brain tumors or neural cells. Ten of these (67%) were enriched in AT/RT hypermethylated sites, MYCN binding sites in brain tumors were enriched in AT/RT hypomethylated sites, and only ZEB1 (6.7%) was enriched in MB hypermethylated sites (**Supplementary Table 2, Supplementary Figure 5**). Of the 16 neural TFs enriched in MB hypermethylated sites, 12 (75%) were measured in pluripotent cells or other cell types, but not in brain tumors or neural cells (**Supplementary Table 2, Supplementary Figure 5**).

TFs related to pluripotency and early cell differentiation were enriched especially in the MB hypermethylated regions. Altogether, 14 (93%) of these 15 TFs were enriched in MB hypermethylated DMRs in both the i450k and RRBS datasets, and the remaining TF, NFYA, was enriched only in the i450k data (**Figure 2E, Supplementary Table 2**). All 11 pluripotent TFs, measured from pluripotent cells, were enriched in regions hypermethylated in MB.

DNA methylation is associated with PRC2 and EP300 binding in a tumor-type-specific manner

Because dysfunctional PRC2 and EZH2 were linked to AT/RT development [11], we analyzed the association of PRC2 members with tumor-type-specific DMRs. Five of six (83%) TFs involved in PRC2 function and all seven (100%) TFs involved in H3K27 methylation [59–61] were enriched in sites hypermethylated in both MB and AT/RT or in MB (**Figure 2B**). EZH2 was enriched in DMRs hypermethylated in AT/RT, hypermethylated in MB, or hypomethylated in MB, suggesting different modes of regulation (**Figure 2E, Supplementary Table 2**). The enriched binding of EZH2 to hypermethylated regions could mean that the PRC2 complex guides DNA methylation instead of advancing H3K27 methylation, as reported in the literature previously [62].

We continued our analysis with split GTRD-data. When examining PRC2 subunits, all the EZH2 binding sites, but those originating from other brain tumors, were enriched in regions hypermethylated in AT/RT (**Figure 2E**). EZH2 was also enriched in regions hypermethylated in MB irrespective of the sample source, but only the EZH2 binding sites in carcinomas and other cancers significantly overlapped with MB hypomethylated regions. The results were somewhat different for another PRC2 subunit SUZ12: only its binding sites in PSCs and other samples were enriched in regions hypermethylated in AT/RT or hypomethylated in PLEX, whereas in MB hypermethylated regions, the SUZ12 binding sites in rhabdoid tumors, carcinomas, and other cancers were also enriched. Overall, the results suggest that DNA methylation is directed to some of the sites EZH2 occupies in pluripotent cells and during neural differentiation in a cancer-type-specific manner.

Next, we examined the SWI/SNF complex because it is altered in AT/RTs due to *SMARCB1* or *SMARCA4* inactivation and acts with EP300 to open chromatin during neural differentiation [12,29]. We found that the SMARCA4 binding sites in rhabdoid tumors were enriched in sites hypermethylated in MB and in sites hypomethylated in AT/RT. The opposite is true in other brain and neural cells: the SMARCA4 binding sites are enriched in sites hypermethylated in AT/RT and sites hypomethylated in MB (**Supplementary Table 2, Figure 2E**). This finding was supported by EP300 binding sites in neural precursors or other brain and neural cells, which were enriched in regions hypermethylated in AT/RT. By contrast, the EP300 binding sites in pluripotent cells were enriched in regions hypermethylated in MB. The results suggest

that the neural SWI/SNF-EP300 target sites have been methylated in AT/RT. Furthermore, SWI/SNF binding sites are partly unique in MB and AT/RT, but the complex remains able to promote open chromatin in the regions to which it binds.

PRC2 subunits and neural differentiation -related TFs rarely co-localize

We performed TF binding co-localization analysis for regions hypermethylated in AT/RT or MB, or hypomethylated in MB or PLEX (**Figure 2F**, **Supplementary Figures 6 and 7**). Most TFs related to neural differentiation share their binding sites and are clustered, but significant co-localization (adjusted $p < 0.01$, one-sided Fisher's exact test) was rarely observed between PRC2 members and TFs related to neural differentiation (**Figure 2F**, **Supplementary Figures 6 and 7**). In regions hypermethylated in AT/RT, the only TFs co-localizing with PRC2 subunits were the neural TFs BCHE and TFAP2A, which were co-localized with all the PRC2 subunits. Only co-localization with one PRC2 subunit was detected for MBD1 and PHOX2B in MB hypermethylated and hypomethylated regions, respectively (**Figure 2F**, **Supplementary Figure 6**). This result suggests that PRC2-related DNA methylation largely affects parts of the genome other than the sites that recruit neural TFs during differentiation, BCHE, and TFAP2A, representing possible TFs that act at their intersection.

DNA methylation -related regulation of genes involved in neural differentiation

We next evaluated the association between DNA methylation and gene expression. Principal component analysis (PCA) visualization showed that samples clustered to their respective groups based on gene expression (**Supplementary Figure 8**), although the separation was less clear than with DNA methylation (**Figure 1B**).

We used Venn diagrams to summarize genes that were similarly regulated in both microarray and sequencing experiments (**Figure 3A**). Because gene expression is generally inhibited by DNA methylation [63], only the genes having an opposite direction in DNA methylation and gene expression change (DM-DE genes) were considered [63]. There were 554, 247, and 879 differentially expressed (DE) genes (adjusted p value < 0.05, minimum two-fold change) in the AT/RT-MB, AT/RT-PLEX, and MB-PLEX comparisons, and 69, 53, and 100 of these were DM-DE genes with methylation differences in the genomic neighborhood, respectively. Seventy-two DM-DE genes harbored DNA methylation differences in the gene promoter and 108 DM-DE genes in the gene-annotated enhancer region, respectively. The majority of the DM-DE genes (207 of 237 genes, 87%) were detected in only one comparison.

DM-DE genes were summarized using oncoprint-like visualization (**Figure 3B**, full version in **Supplementary Figure 8**). Nearly all AT/RT-specific DE genes (134 of 144, 93%) were downregulated, and only 10 (7%) were upregulated. DNA methylation differences were more linked to decreased expression in AT/RT and increased expression in MB and PLEX: of all the 237 DM-DE genes, 134 (93% of AT/RT-related), 85 (42% of MB-related), and 44 (28% of PLEX-related) were downregulated in AT/RT, MB, and PLEX, respectively, when compared with at least one tumor type (**Supplementary Figure 8**). By contrast, 10 (6.9% of AT/RT-related), 116 (58% of MB-related), and 114 (72% of PLEX-related) were upregulated in AT/RT, MB, and PLEX, respectively.

Based on the gene expression and DNA methylation patterns of the 237 DM-DE genes (**Supplementary Figure 9**), these genes were largely hypermethylated in AT/RT (226 had a high methylation in AT/RTs) and overexpressed in PLEX or MB or in both in respect to AT/RT (179 and 150 genes in array and RNA-seq data, respectively) (**Figure 3B, Supplementary Figure 8**). The number of AT/RT-specific genes was partly reduced because DNA methylation was decreased in MB or PLEX alone or because decreased DNA methylation in MB and PLEX led to increased expression in only one of them.

Thirty-two DM-DE genes were specific to a given tumor type, that is, differential methylation in the same genomic context (genomic neighborhood, promoter or enhancer) was observed in two tumor comparisons (**Figure 3B-C**, **Supplementary Figure 8**). Of the seven tumor type-specific genes that are hypermethylated and underexpressed, *CXXC5* and *TCEA3* are specific to AT/RT, *RSPH9*, *CLDN23*, *C5AR1,* and *IGDCC3* specific to MB, and *CHTF18* specific to PLEX. Of the 25 tumor type-specific genes that were hypomethylated and overexpressed, nine genes, including neural genes NEUROG1, EBF3, NBEA, BARHL1, and AMER2, are specific to MB. Of these, EBF3, BARHL1, and AMER2 have been reported to be overexpressed in MB [64–66].

We separately visualized the genomic neighborhoods of *CXXC5*, *TCEA3*, *NEUROG1,* and *NEUROD2*. Several DMRs were present in the *CXXC5* locus but only one was present in the other gene loci (**Figure 3D**, **Supplementary Figures 10 and 11, Supplementary Table 3**). The *CXXC5* locus was also located in a large-scale methylation area (in Chr5q31.2, **Figure 1F**), which is hypermethylated in ATRT when compared with MB or PLEX. In addition, *TCEA3* had a DMR in the promoter region, indicating classical regulation through promoter methylation. The binding sites of the SWI/SNF subunits SMARCA4 or BRD9 measured from AT/RT cells were detected in the *CXXC5*, *NEUROG1,* and *NEUROD2* DMRs. Notably, EP300 binding sites measured from neural cells were also found around *CXXC5*, representing a possible neural SWI/SNF-EP300 target site described earlier (**Figure 3D**).

We identified *NEUROG1* and *NEUROD2* as DM-DE genes that were less expressed in AT/RTs (**Figure 3B**), and *NEUROG2* and *NEUROD1* as TFs whose binding sites were enriched specifically in AT/RT hypermethylated regions (**Figure 2D, Supplementary Figure 5**). Because these transcription factors (TFs) are key regulators of neural differentiation [67,68], we visualized their expression together with the expression of their target genes *TBR1, RCOR2, BLHE22, NHLH1, NHLH2, HES6, DAB1,* and *CDK5R1* (**Figure 3E, Supplementary Figure 10**). Except for NEUROG2, which had the highest expression in AT/RT, the highest expression of all these genes was observed in MB. This result suggests that their induction, which should occur during neural differentiation, has been successful mainly in MB.

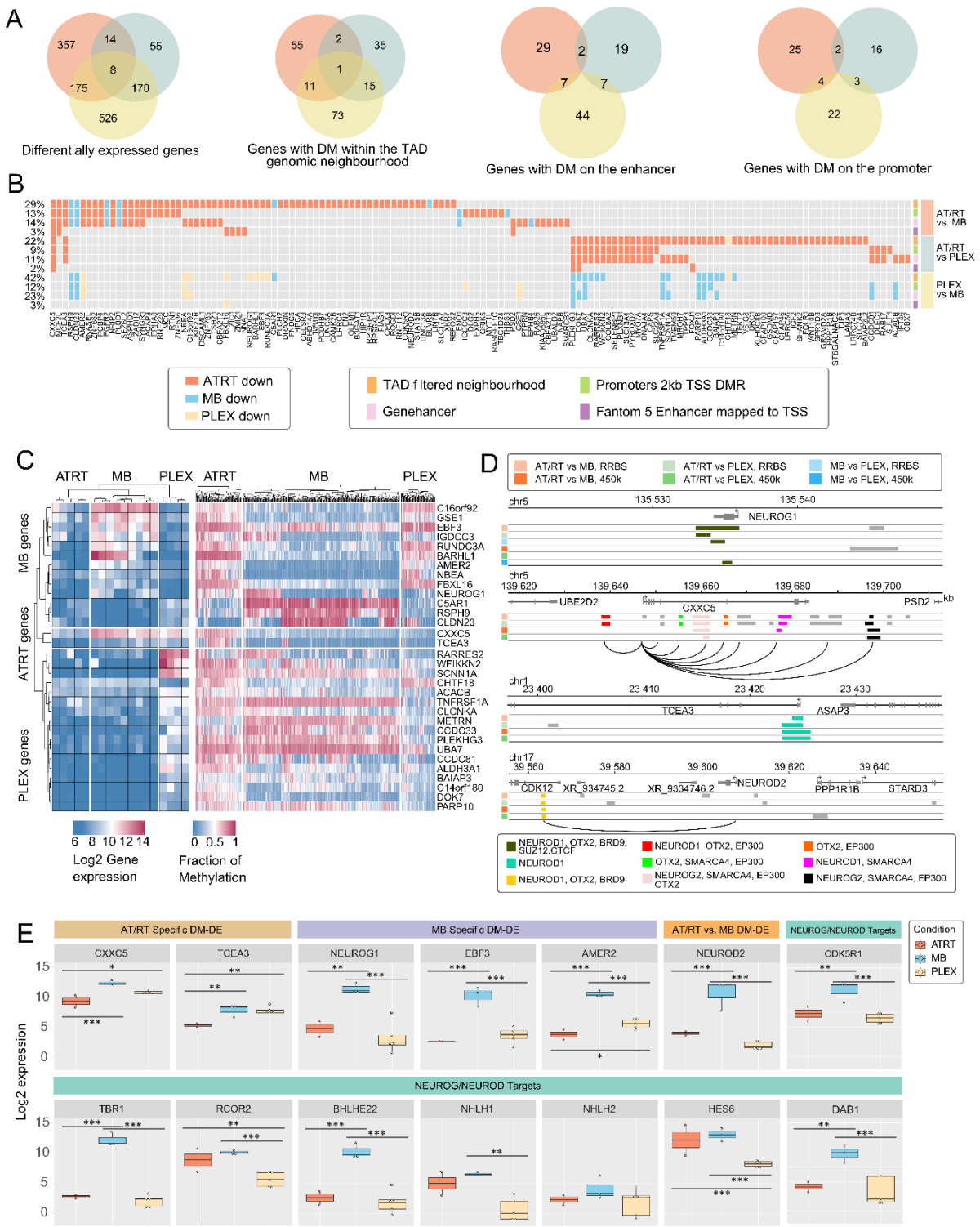

Figure 3: Methylation differences are associated with differential expression.

a) Several differentially expressed (DE) genes are associated with differential methylation (DM). Gene expression and methylation patterns were studied in four contexts: 1) differential gene expression, 2) DE coupled with DM in the genomic neighborhood (+-200kb from the transcription start site [TSS] within the same topologically associating

domain [TAD]), 3) DE coupled with DM in gene-linked enhancer, and 4) DE coupled with DM in the gene promoter (2kb upstream and 500bp downstream from the TSS). Venn diagrams show the genes that behaved similarly to both sequencing and array data. Differentially expressed genes associated with differential methylation (cases 2-4) are called DM-DE genes, and only cases where the sign of DM change was opposite to DE were included.

b) The DM-DE genes were hypermethylated and underexpressed in AT/RT. Oncoprint showing the DM-DE genes, which are differentially expressed in AT/RT-PLEX and/or AT/RT-MB comparisons, and the direction of differential gene expression. For each comparison, different possible DMR locations (gene neighborhood within TAD, promoter (2kb), enhancer (Fantom 5), enhancer (Genehancer)) were visualized separately. Enhancer regions have been linked to DE genes.

c) Sample-wise levels of gene expression (on the left) and DNA methylation (average methylation of variable sites, right) for tumor-specific DM-DE genes. Microarray data were used for both heatmaps.

d) Different DMR patterns were present in the DM-DE genes. Visualization of DMRs around *NEUROG1*, *CXXC5*, *TCEA3*, and *NEUROD2*. *CXXC5* and *TCEA3* were AT/RT-specific downregulated DM-DE genes, *NEUROG1* was an MB-specific upregulated gene, and *NEUROD2* was a DM-DE gene in AT/RT-MB comparison. Distal DMRs are connected to the TSS via an arch. Colors indicate which relevant TFs have binding sites in these regions in the GTRD AT/RT, MB, and neural samples. SMARCA4, SUZ12, CTCF, and BRD9 were measured from rhabdoid samples, EP300 from neural samples, NEUROG2 from other precursors, and NEUROD1 and OTX2 from MB-samples.

e) Expression patterns of selected DM-DE genes and targets of NEUROG and NEUROD TFs. The highest expression is typically detected in MB (*,**, and *** refer to $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively).

AT/RTs harbor unique and PSC-like DNA methylation patterns

On the basis of TF binding patterns, genes regulated by DNA methylation, and recent results in the field [69], we wanted to analyze how the observed DNA methylation differences relate to normal neural cell differentiation. To achieve this objective, we obtained i450k DNA methylation data from pluripotent cells (PSCs) and primary fetal brain (FB) [70,71] and defined their methylation differences with respect to tumors and to each other. PSC and FB samples clustered to separate clusters in the tSNE visualization (**Figure 4A**), as expected. We then divided the DMRs between the tumors to 1) those that are tumor type-specific, PSC-like, or FB-like and 2) whether they were differentially methylated between PSC and FB (**Supplementary Table 4**).

DMR counts in different categories were visualized to estimate how AT/RT hypermethylated, MB hypomethylated, and MB hypermethylated DMRs are related to DNA methylation in normal cells (**Figure 4B,** left side). Among the 898 DMRs hypermethylated in AT/RT, 470 (52%) were similarly methylated in PSCs as AT/RTs and hypermethylated in PSCs when compared with FBs (group 5, **Figure 4B**); thus, they were linked to neural differentiation. Only 31 (3.5 %) of AT/RT hypermethylated DMRs were FB-like (group 8), highlighting the higher

similarity of AT/RTs to PSCs than FB. However, there were also 300 (33%) AT/RT-specific DMRs, 124 (14%) of which were not related to differentiation (group 4), and the remainder (176, 20%) were also hypermethylated in PSCs when compared with FB (group 1).

By contrast, 116 of 123 (94%) MB hypermethylated DMRs and 699 of 957 (73%) MB hypomethylated DMRs were MB specific (**Figure 4B**). Furthermore, 241 (25%) MB hypomethylated DMRs were FB-like and differentiation-related (group 18). Within MB specific regions, 445 (51%) of the MB hypomethylated regions were differentiation-related and hypermethylated in PSC (group 15), 321 (34%) of which were also PSC-like when compared with AT/RT (group 16). A similar trend was observed for 123 MB-specific hypermethylated regions, 31 (25%) of which were also more methylated in FB than in PSC (group 13). Thus, a high proportion of even tumor-type-specific DMRs change their DNA methylation during differentiation and are differentiation-related in a manner that underlines the similarity of MBs to FB and AT/RTs to PSCs.

These results further support our prior findings that a majority of DMRs are hypermethylated in AT/RT when compared with MB or PLEX. The majority of these sites are demethylated upon neural differentiation, and a large proportion of them are PSC-like. However, we also observed AT/RT-specific DNA methylation, part of which is not differentiation-related and represents unique targeting of DNA methylation in these malignant cells.

We calculated the proportion of large-scale DMRs (**Figure 1F**) in each DMR category (**Figure 4B,** right side) to determine whether the observed DNA methylation differences were more linked to focal or large-scale regulation of DNA methylation. Notably only 14 (11%) of MB hyper DMRs represented large-scale DNA methylation differences, whereas 384 (49%) and 505 (55%) of AT/RT hyper and MB hypo DMRs, respectively, were involved in large-scale methylation differences (**Figure 4B**). Thus, MB hyper DMRs appear to target specific cytosines or focal areas, such as specific promoters and enhancers, whereas focal and large-scale regulation of DNA methylation is observed for regions hypermethylated in AT/RT or hypomethylated in MB.

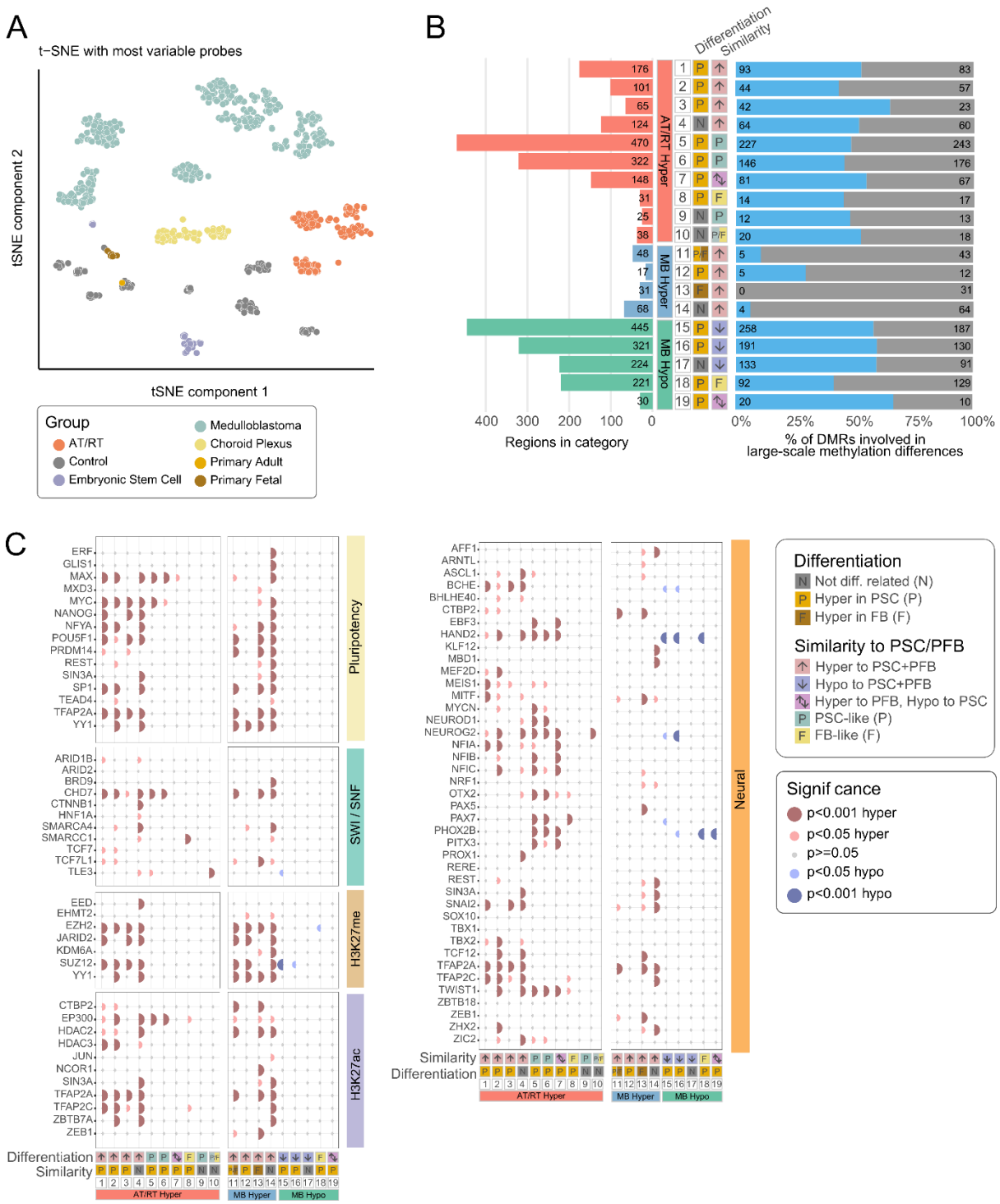

Figure 4: AT/RTs harbor more pluripotent stem-cell-like properties than MBs or PLEXs

a) Pluripotent stem cells (PSCs), primary adult brain, and primary fetal brain (FB) are separated from tumor samples based on DNA methylation in a tSNE visualization, when the 10 000 most variable regions in i450k data were used for visualization.

b) A large proportion of AT/RT hyper DMRs were PSC-like, whereas the majority of MB DMRs were specific to MB. Very few MB hypermethylated DMRs were associated with large-scale differences in DNA methylation. Tumor-type-specific DMRs (**Figure 2a**) were categorized based on DNA methylation levels in PSC and FB samples. The bar

plot on the left shows the number of DMRs in different categories. Annotations show whether DMRs are PSC-like (C), FB-like (B), or specific (S) (different from PSC or FB) and whether they were differentially methylated between PSC (C) or FB (B) or if they were not related to differentiation (N). Groups 2 and 3 are subgroups of 1. Group 2 included only DMRs which were also hypermethylated in PSCs compared with PLEX or MB. Group 3 included only DMRs that were PSC-like in PLEX. Group 5 comprised groups 6 and 7, and group 11 comprised groups 12 and 13. Group 16, comprising only DMRs that are PSC-like in AT/RT, is a subgroup of 15. The proportion of DMRs in large-scale methylation differences within annotated DMR categories is shown on the right. The number of DMRs is also marked in the figure.

c) The DMR category-related binding patterns revealed TFs involved in tumor-specific, normal cell-like, and differentiation-related regulation of DNA methylation. TF binding site enrichment was calculated separately for each DMR category (bottom). TFs were organized into the themes listed in Figure 2B. The dot is not marked when a given TF is not measured in a given GTRD category.

AT/RT-specific methylation is associated with PRC2

We analyzed TF binding separately in the DMR categories presented in **Figure 4B** to define TFs relevant to different regulatory programs. This step was performed as we previously observed that PRC2 subunits are only rarely co-localized with neural TFs in AT/RT hypermethylated sites (**Figure 2F**) and with other TFs in MB hypermethylated sites (**Supplementary Figure 6 and 7**). The binding of EZH2 and other PRC2 members was significantly enriched in AT/RT-specific, but not in PSC-like or FB-like, AT/RT hyper DMRs (**Figure 4C**). Therefore PRC2 might target DNA methylation in regions specific to AT/RTs, which are not linked to DNA methylation in sites shared with PSCs. In MB, PRC2 members were enriched in hypermethylated regions, but EZH2 was also enriched in hypomethylated regions that were FB-like.

The SWI/SNF subunits SMARCA4 and SMARCC1 are especially associated with DMRs that are AT/RT-specific and not differentiation-related. SMARCA4, whose binding sites in other brain and neural cells are enriched in AT/RT hypermethylated DMRs (**Figure 2E**), is enriched in sites that are AT/RT-specific and demethylated during neural differentiation. These sites might represent hypermethylation of the neural SWI/SNF-EP300 target sites discussed earlier.

Contrary to prior findings, a large proportion of neural TFs was enriched in PSC-like AT/RT hyper DMRs. This finding suggests that these PSC-like sites remain methylated, and thus inaccessible, in AT/RT, which has the potential to partly maintain AT/RT in a low differentiation state and promote malignancy. PSC-like methylation was not linked to PRC2. Notably, neural TFs BCHE and TFAP2A, the only neural TFs co-localized with PRC2 subunits in AT/RT hypermethylated sites (**Figure 2F**), were enriched only in AT/RT-specific AT/RT hyper-DMR categories (**Figure 4C**), linking them closely to PRC2-related DNA methylation.

In conclusion, our results suggest that AT/RT-specific methylation is linked to malfunctional SWI/SNF and PRC2 complexes. This phenomenon has the potential to drive oncogenic epigenetic regulation, which is unique to AT/RT. In addition, we detected regions that harbor

binding sites for several neural regulators and remained similar to PSCs that were originally methylated via a PRC2-independent mechanism.

Discussion

Our study aimed to identify DNA methylation patterns, epigenetically regulated TFs, and abnormally expressed genes specific to AT/RTs, PLEXs, or MBs by using both array and sequencing data. Although the characteristics of the materials, e.g., the genomic coverage, sample amounts, and suitable algorithms differed, the key findings align.

Our results showed that many TFs that would bind to the sites hypermethylated in AT/RT are the drivers of neural differentiation. NEUROG2 and NEUROD1 are especially notable because they regulate a wide range of factors that affect neural differentiation [72]. Both belong to the basic helix–loop–helix (bHLH) gene family, members of which bind to unmethylated DNA and are thus sensitive to DNA methylation [37]. Furthermore, *e.g.* NEUROD1 target genes [72,73] OTX2 and EBF3, both with the highest expression in MB, have their binding sites also hypermethylated in AT/RTs, indicating the silencing of the whole differentiation cascade of genes, which has been described [52,74–77]. In our downstream analysis, *NEUROG1* and *NEUROD2* were also defined as DM-DE genes. Thus, NEUROG-NEUROD family TFs appear to be targets of DNA methylation-driven regulation at different levels, via decreased gene expression or altered DNA binding ability, both of which contribute to suppressed neural differentiation. Several other neural genes, such as *NBEA, BARHL1,* and AMER2 [78–80], were more methylated and less expressed in AT/RT than in MB. Our results are supported by prior notions that neuronal differentiation factors have lowered functionality in AT/RTs [11,81,82].

Erkek et al. reported that active enhancers and PRC2-repressed chromatin regions are more depleted in AT/RT than in other brain tumors or the pediatric normal brain [11]. Instead, a higher proportion of the genome was in a quiescent state and thus did not have the measured chromatin marks [11]. It is likely that some of the quiescent regions acquired DNA methylation. PRC2 marks many genes that are methylated during cell differentiation into neural precursor cells [62]. EZH2 has been reported to interact with DNA methyltransferases [62,83] and to directly control DNA methylation [83]. In rhabdoid tumor cells, DNA methylation is decreased, and PRC2 members concomitantly released from the p16 locus upon SMARCB1 reexpression [84], also suggesting an interplay between PRC2 and DNA methylation. Our results suggest that EZH2 contributes to AT/RT-specific hypermethylation, whereas the majority of neural TFs were not associated with PRC2. BCHE and TFAP2A were exceptions: they co-localized with PRC2 members and were solely enriched in AT/RT-specific regions. TFAP2A is known to regulate neural crest induction and cranial placode specification [85,86]; thus, it might mark regions relevant for developmental branches being silenced in AT/RT, MB, and even in the brain in general. BCHE is expressed early in nervous system development and plays a role in neural stem cell development [49]. This could possibly contribute to the observed differentiation blockage. Further studies are necessary to clarify the roles of these genes.

PSC-like DNA methylation and gene expression patterns were shown to exist in human AT/RT samples [69]. We showed that areas similarly methylated in both AT/RTs and PSCs are differentiation-related, giving them the function and strengthening the idea that these are important regulatory regions controlling the fate of these cells. We also showed that many of these regions involve large-scale methylation differences, implying larger changes in the 3D-structure of chromatin. What is the relationship between these PSC-like sites and AT/RT-specific, PRC2-related sites that do not co-localize with the binding sites of most neural TFs? One possibility is that altered PRC2-mediated regulation is halting AT/RT cells to a specific point in the developmental trajectory, prohibiting the demethylation steps necessary for further neural development. This notion is supported by a report showing that EZH2 inhibition suppresses an ESC-like expression pattern in SMARCB1 deficient tumors [69]. EZH2 inactivation or deletion prevents AT/RT development and growth [69,81,87,88], making it a relevant regulator for these tumors. The PSC-like hypermethylation in AT/RT possibly reflects an unaccomplished differentiation-related DNA demethylation and could thus be a downstream effect, which does not need to overlap with PRC2 binding.

In summary, our analysis revealed DNA methylation patterns specific to AT/RT, MB, and PLEX and associated them with gene regulation, cell differentiation, and malignancy. AT/RTs carry very few genetic alterations, which target the SWI/SNF epigenetic regulatory complex; therefore, aberrant epigenetic regulation is relevant for these tumors. Because other studied epigenetic marks are mainly depleted [11], whereas DNA methylation is increased in AT/RT, DNA methylation is likely to play a role in AT/RT malignancy, at least partly by halting normal neural development. Further research will show whether treatment options affecting differentiation blockage could generate positive outcomes for patients suffering from this devastating disease.

Materials and Methods

Reduced representation bisulfite sequencing

DNA was isolated from the frozen samples using a QIAamp DNA Mini kit (QIAGEN) with RNAse treatment and from the FFPE sample with a turXTRAC FFPE DNA kit (Covaris). Library construction and sequencing were performed at the Beijing Genomics Institute (BGI), Hong Kong. Samples were sequenced using Illumina HiSeq2000 technology. Paired-end sequencing of 50 bp was used. The data goal was 5G.

RNA sequencing

RNA was isolated using a mirVana isolation kit (Invitrogen). Library construction and sequencing were performed at Novogene, Hong Kong. Samples were sequenced using Illumina HiSeq2000 technology. Paired-end sequencing of 150 bp was used. The data goal was 20 million raw reads per sample.

Preprocessing and analysis of DNA methylation datasets

i450k data. The full CNS tumor reference cohort established by Capper et al. was filtered, batch-effect corrected, and normalized as described in their study [21], resulting in an initial set of 428799 probes and 2801 samples. Supplementary clinical annotations were also mapped onto the samples. AT/RT, MB, PLEX, and CONTR samples, excluding the control subtype "INFLAM" and all samples without location information, were extracted (N = 584).

To detect DMRs between tumor types, AT/RT-MB, AT/RT-PLEX, and PLEX-MB comparisons were conducted, and each cancer was also tested against the CONTR set. Additional filtering (sex chromosomes, cross-hybridizing chromosomes) followed by DMR calling for the remaining 396700 probes was performed with DMRcate v1.18.0, switching the hg19 coordinates into hg38 with rtracklayer v1.42.2 [89–91]. Location information was included in the linear model to adjust for biological confounders. To reduce normal cell effects, the DMRs of each main comparison (*e.g.*, AT/RT-MB) was overlapped with the control DMRs (*e.g.*, AT/RT-CONTR, MB-CONTR). The resulting regions were reduced by taking only the intersecting regions between the main comparison and at least one of the control comparisons [92].

DMRs were annotated to annotatr v1.8.0 pre-built hg38 genic regions, CpG regions, and FANTOM5 enhancers [93,94]. In addition, the pre-built promoter annotations (1 kb upstream from the transcription start site [TSS]) were used to generate broader promoters (2 kb upstream and 500 bp downstream from the TSS), and genomic neighborhoods spanning 200 kb in both directions from the TSS.

RRBS data. Quality control, adapter trimming, and MspI restriction site trimming were performed using the Babraham Bioinformatics Group preprocessing tools FastQC v0.11.7 and Trim Galore v0.5.0. Reads were queried against bisulfite-converted reference genomes of potential contaminants with FastQ Screen v0.13.0 to check the origin of the libraries [95]. Mapping against hg38 (UCSC) reference and mapping quality control without duplicate removal was performed using Bismark v0.19.1 [96]. The quality information of all these steps was summarized over the dataset using MultiQC [97].

Methylation percentage calling and differential methylation analysis were performed using methylKit v1.8.1 [91,98]. CpGs with coverage of less than 10 reads or hitting into scaffolds, sex chromosomes, or mitochondrial DNA were removed. The CpGs were tiled into 1000 bp non-overlapping regions, i.e., tiles, which were used for DMR calling with overdispersion correction. Regions with methylation percentage difference greater than 25% and q value (FDR) smaller than 0.05, were extracted and then annotated similarly as i450k data.

Dimensionality Reduction. The β values for all AT/RT, MB, and PLEXUS samples were aggregated on RRBS tiles based on lift-over hg38 probe coordinates. This step was performed by taking the complete overlaps and then calculating the mean beta value for probes hitting each tile [92]. The scales in the dataset were unified, and the 10 000 most varying regions were used as inputs for various dimensionality reduction algorithms.

TF binding analyses

Preliminary Statistical Testing. TF binding site (TFBS) enrichment analysis was performed using one-sided Fisher's exact test (FET). TFBS data were obtained from GTRD v.19.04 metaclusters [35]. Before testing, DMRs with beta fold change >= 0.25 were filtered by overlap operations between the comparisons to obtain the regions specific to each cancer with a concordant direction of methylation change. Hypermethylated and hypomethylated regions were tested separately. For the i450k data, the background set was generated from previously filtered 396700 probes, and for the RRBS data, the methylKit tiles were used as a background.

When performing the tests with i450k data, the hg38 coordinates of each 50 bp probe overlapping with DMRs were first obtained. These were extended to 500 bp in both directions, and the methylation fold-change of the original DMR was used to represent the methylation difference in each extended probe. Regions were considered cancer-specific if they overlapped with at least 600 bp between comparisons, and with RRBS, because of the non-overlapping nature of the DMRs, the exact overlaps could be used.

FET p values were adjusted by Benjamini-Hochberg (BH) correction [99]. TFs with adjusted p-values smaller than 0.05, were considered significant. Sets of validated TFs with significant enrichment in the analogous DMR-sets, that is, in hypermethylated or hypomethylated cancer regions of the i450k and RRBS experiments, were constructed. Each validated TF had been studied in the literature, and because common themes existed, they were manually annotated into 11 categories when relevant.

Customized TFBS enrichment analysis with cell-type splitting. GTRD annotations for cell types linked to a given TFBS were further examined and manually divided into 11 groups (**Supplementary Table 2**). The GTRD was then split based on the terms of the groups, and these groups were used as input material for the FET to replace the full GTRD.

Studying average methylation levels at binding sites. Methylation of the relevant i450k probes near the binding sites of the validated TFs (maximum gap 500 bp) was visualized as violin plots. The probes were selected by pooling the cancer regions and then finding complete overlaps with the hg38 probe coordinates. The beta values over each tumor subclass were aggregated probewise using the mean.

TFBS co-localization. The co-localization of validated TFs was studied with TF binding sites hitting target areas (hypermethylated AT/RT regions, hypermethylated MB regions, hypomethylated MB regions, or hypomethylated PLEX regions). TFs were tested pairwise against each other to construct four-field matrices for each case. Regions with 1) no binding site from either TF, 2) binding sites from both TFs, and 3-4) binding sites from one TF but not the other were counted and the significance was tested using one-sided FET and stored as symmetric TF matrix of P values. These P value matrices were corrected using the BH method. Adjusted P values were subsetted by the TF names of the validation group (AT/RThyper-TFs, MBhyper-TFs, MB-hypoTFs, PLEXhypoTFs) and visualized using the ComplexHeatmap package [100].

Differential expression analysis for expression datasets and integration with methylation data

Array datasets. Normalized expression arrays of three public datasets were downloaded from GEO: Illumina HumanHT-12 V3.0 [101] (accession GSE42658) and two Affymetrix Human Genome U133 Plus 2.0 sets [102] (accession GSE35493 and [103], accession GSE73038). The Illumina set contains measurements of all three cancers, and Affymetrix sets provide additional support for AT/RT-MB comparison. Before DE analysis with limma, the expression matrices were log2-transformed if that had not been performed by the normalization method [104]. Probes were annotated into symbolic gene names, and the mean of all the probes mapped to the same gene was chosen to represent the expression values. Genes with BH adjusted p values smaller than 0.05 and log2-transformed fold changes (absolute change) of at least 1 were considered significant.

RNA-seq data. After sequencing data QC with FastQC v0.11.7 (https://www.bioinformatics.babraham.ac.uk)
Reads were mapped and quantified against hg38 by kallisto v0.44.0 [105]. The kallisto transcript abundances were transformed to gene level counts and scaled to library size using tximport v.1.10.1. DE analysis was then performed using DESeq2 v1.22.2, with default settings [106]. Genes with BH adjusted p values smaller than 0.05, and log2-transformed fold changes (absolute change) of at least 1 were considered significant.

Integration. The DE results were merged with the annotated DMRs by symbolic gene names to find cases where the direction of the methylation change in a regulatory region (i.e., promoter, enhancer, genomic neighboorhood) was opposite to the gene expression change.
GEO expression array results were integrated with i450k DMRs, and RNA-seq based results were integrated with RRBS DMRs from matched cases to provide data for validation. Validation was performed by extracting the genes that showed similar behavior in both approaches. Since annotatr uses transcript-level annotations, the validated hits, *e.g.*, for promoters, represent DE genes that show regulatory methylation in at least one of its transcript specific promoters in both datasets. Additional filtering and validation steps were performed for the enhancers and genomic neighborhoods, described later. The results of the experiments were summarized using Venn diagrams and an oncoprint-like visualization using the ComplexHeatmap package [100].

A recent regulatory genomics workflow was adapted to integrate DMRs and DE genes with precalculated promoter-enhancer annotations provided by the FANTOM5 database [94,107]. DE-DMR pairs in genomic neighborhoods were queried between i450k and RRBS experiments with a maximum gap of 5 kb, and the regions inside this “validation distance” with similar methylation direction were extracted. Moreover, to ensure that a given DMR can spatially interact with the TSS of a given gene, the results were studied with TAD information from Hi-C experiments [108,109] (http://3dgenome.fsm.northwestern.edu/). The region between the DMR and TSS of each record overlapped with TADs originating from five cell types: SKNDZ, Cortex DLPFC, Hippocampus, H1-ESC, and H1-NPC [108]. Genes without any TAD boundaries in any of their transcript TSS-DMR pairs were extracted.

Linking DMRs to TADs for large-scale DNA methylation analysis

DMRs were overlapped with *Cortex DLPFC* and H1-NPC TADs by using Bedtools GroupBy v2.25.0 and in-house scripts to calculate hyper- and hypomethylated records in each TAD [108,110]. The input material was either a broadened (+-500bp) probe extracted from the i450k DMRs or the methylKit DMRs of the RRBS analysis. The resulting TADs for each comparison were filtered such that at least five similarly directed probes/tiles with both the i450k and RRBS approach should hit the TAD and, in addition, less than 1/10 of the hits can undergo methylation changes in the opposite direction. Moreover, the distance between the first and last probe or tile in the TAD had to be over 50 kb. The filtered TADs from different cell lines were pooled and reduced to produce single TAD coordinates for each comparison and then visualized with karyoploteR [111].

Comparing brain tumors to ESCs and primary fetal brain

Two ESC samples, nine primary FB samples (GSE116754) [71], and 25 ESC samples (GSE31848) [70], were preprocessed using the Capper et al. data [21]. Beta-value density plots and tSNE-clustering for the 10 000 most varying probes were used to check that samples were separated based on tumor and normal sample type. The DMRs were called with DMRcate by comparing AT/RT, MB, and PLEX samples against ESC and FB cells. In addition, ESC samples were compared with FBs.

From these seven comparisons, hypermethylated and hypometylated DMRs were extracted ($q < 0.05$, abs. average beta fold change >= 0.25 and length between 100 and 5000 bp). The region lists were overlapped with the same cancer-specific regions extracted for TFBS analysis (min 1 bp overlap). As a result, the experiment was summarized in a “developmental DMR matrix” of 2301 non-overlapping regions showing which of the comparisons or cancers each region overlapped. The resulting region sets were categorized based on the differences detected in tumor-tumor, tumor-normal, and ESC-FB comparisons. TFBS enrichment analysis was performed for these categories by extracting and extending the probes hitting the regions and testing them against GTRD using one-sided FET, as described.

Acknowledgements

We would like to acknowledge Mrs. Paula Kosonen, Mrs. Päivi Martikainen, Mrs. Marika Vähä-Jaakkola, Mrs. Marja Pirinen, Mrs. Sari Toivola, Mrs. Hanna Selin, Mrs Marita Nieminen, Mrs Leena Jalonen and Mrs Satu Ranta for sample handling and logistics, Mrs. Aliisa Tiihonen for scientific input, Mrs. Laura Huhtala for scientific input and for commenting on the manuscript, Mrs. Suvi Lehtipuro for preprocessing of i450k data, and Mr. Tomi Häkkinen for technical support. Personnel at Tampere University Hospital and Fimlab laboratories Ltd. are acknowledged for their contribution to sample collection. We are grateful for them and the patients for permitting the analysis of patient material. The study was financially supported by the Academy of Finland (#312043 (MN), #310829 (MN), #333545 (KG)), Cancer Foundation Finland (MN, KG), Sigrid Jusélius Foundation (MN, KG), Emil Aaltonen Foundation (KG), Finnish Cancer Institute (MN), Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (MN, KG), Väre Research foundation (KN), The Finnish Pediatric Research Foundation (KN) and Finnish Cultural Foundation (JU). We

acknowledge the CSC - IT Centre for Science, Finland, for providing computational resources and the Biocenter Finland (BF) and Tampere Genomics Facility for the service.

References

1. Bhakta, N. *et al.* Childhood cancer burden: a review of global estimates. *Lancet Oncol.* **20**, e42–e53 (2019).
2. Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro. Oncol.* **21**, v1–v100 (2019).
3. Burger, P. C. *et al.* Atypical Teratoid/Rhabdoid Tumor of the Central Nervous System: A Highly Malignant Tumor of Infancy and Childhood Frequently Mistaken for Medulloblastoma: A Pediatric Oncology Group Study. *Am. J. Surg. Pathol.* **22**, 1083 (1998).
4. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
5. Udaka, Y. T. & Packer, R. J. Pediatric Brain Tumors. *Neurol. Clin.* **36**, 533–556 (2018).
6. Ostrom, Q. T. *et al.* The descriptive epidemiology of atypical teratoid/rhabdoid tumors in the United States, 2001-2010. *Neuro. Oncol.* **16**, 1392–1399 (2014).
7. von Hoff, K. *et al.* Frequency, risk-factors and survival of children with atypical teratoid rhabdoid tumors (AT/RT) of the CNS diagnosed between 1988 and 2004, and registered to the German HIT database. *Pediatr. Blood Cancer* **57**, 978–985 (2011).
8. Sun, M. Z. *et al.* Current management of choroid plexus carcinomas. *Neurosurg. Rev.* **37**, 179–92; discussion 192 (2014).
9. Sidaway, P. Medulloblastoma: prognostic subtypes revealed. *Nature reviews. Clinical oncology* vol. 18 131 (2021).
10. Faria, C. M. C., Rutka, J. T., Smith, C. & Kongkham, P. Epigenetic mechanisms regulating neural development and pediatric brain tumor formation. *J. Neurosurg. Pediatr.* **8**, 119–132 (2011).

11. Erkek, S. *et al.* Comprehensive Analysis of Chromatin States in Atypical Teratoid/Rhabdoid Tumor Identifies Diverging Roles for SWI/SNF and Polycomb in Gene Regulation. *Cancer Cell* **35**, 95–110.e8 (2019).
12. Sokpor, G., Xie, Y., Rosenbusch, J. & Tuoc, T. Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front. Mol. Neurosci.* **10**, 243 (2017).
13. Stricker, S. H. & Götz, M. DNA-Methylation: Master or Slave of Neural Fate Decisions? *Front. Neurosci.* **12**, 5 (2018).
14. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
15. Feinberg, A. P., Koldobskiy, M. A. & Göndör, A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* **17**, 284–299 (2016).
16. Saghafinia, S., Mina, M., Riggi, N., Hanahan, D. & Ciriello, G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* **25**, 1066–1080.e8 (2018).
17. Patnaik, S. & Anupriya. Drugs Targeting Epigenetic Modifications and Plausible Therapeutic Strategies Against Colorectal Cancer. *Front. Pharmacol.* **10**, 588 (2019).
18. Sato, T., Issa, J.-P. J. & Kropf, P. DNA Hypomethylating Drugs in Cancer Therapy. *Cold Spring Harb. Perspect. Med.* **7**, (2017).
19. van der Pol, Y. & Mouliere, F. Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA. *Cancer Cell* **36**, 350–368 (2019).
20. Pan, Y., Liu, G., Zhou, F., Su, B. & Li, Y. DNA methylation profiles in cancer diagnosis and therapeutics. *Clin. Exp. Med.* **18**, 1–14 (2018).
21. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
22. Torchia, J. *et al.* Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors. *Cancer Cell* **30**, 891–908 (2016).
23. Ho, B. *et al.* Molecular subgrouping of Atypical Teratoid / Rhabdoid Tumors (ATRT) – a reinvestigation and current consensus. *Neuro-Oncology* (2019)

doi:10.1093/neuonc/noz235.

24. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
25. Merino, D. M. *et al.* Molecular characterization of choroid plexus tumors reveals novel clinically relevant subgroups. *Clin. Cancer Res.* **21**, 184–192 (2015).
26. 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).
27. Biegel, J. A. *et al.* Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Res.* **59**, 74–79 (1999).
28. Gessi, M., Giangaspero, F. & Pietsch, T. Atypical teratoid/rhabdoid tumors and choroid plexus tumors: when genetics ‘surprise’ pathology. *Brain Pathol.* **13**, 409–414 (2003).
29. Judkins, A. R. *et al.* INI1 protein expression distinguishes atypical teratoid/rhabdoid tumor from choroid plexus carcinoma. *J. Neuropathol. Exp. Neurol.* **64**, 391–397 (2005).
30. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* **8**, 14648 (2017).
31. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
32. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
33. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Science Advances* **5**, eaaw1668 (2019).
34. McCabe, M. G., Bäcklund, L. M., Leong, H. S., Ichimura, K. & Collins, V. P. Chromosome 17 alterations identify good-risk and poor-risk tumors independently of clinical factors in medulloblastoma. *Neuro. Oncol.* **13**, 376–383 (2011).
35. Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y. & Kolpakov, F. GTRD: a database on gene transcription regulation—2019 update. *Nucleic Acids Res.* **47**, D100–D105 (2019).

36. Medvedeva, Y. A. *et al.* Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics* **15**, 119 (2014).

37. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).

38. Swartling, F. J., Čančer, M., Frantz, A., Weishaupt, H. & Persson, A. I. Deregulated proliferation and differentiation in brain tumors. *Cell Tissue Res.* **359**, 225–254 (2015).

39. Su, X., Kameoka, S., Lentz, S. & Majumder, S. Activation of REST/NRSF target genes in neural stem cells is sufficient to cause neuronal differentiation. *Mol. Cell. Biol.* **24**, 8018–8025 (2004).

40. Halder, D. *et al.* Suppression of Sin3A activity promotes differentiation of pluripotent cells into functional neurons. *Sci. Rep.* **7**, 44818 (2017).

41. Wang, H. *et al.* ZEB1 Represses Neural Differentiation and Cooperates with CTBP2 to Dynamically Regulate Cell Migration during Neocortex Development. *Cell Rep.* **27**, 2335–2353.e6 (2019).

42. Singh, S. *et al.* Zeb1 controls neuron differentiation and germinal zone exit by a mesenchymal-epithelial-like transition. *Elife* **5**, (2016).

43. Kimiwada, T. *et al.* Clock genes regulate neurogenic transcription factors, including NeuroD1, and the neuronal differentiation of adult neural stem/progenitor cells. *Neurochem. Int.* **54**, 277–285 (2009).

44. Hamilton, K. A. *et al.* Mice lacking the transcriptional regulator Bhlhe40 have enhanced neuronal excitability and impaired synaptic plasticity in the hippocampus. *PLoS One* **13**, e0196223 (2018).

45. Li, X. *et al.* Epigenetic Regulation of the Stem Cell Mitogen Fgf-2 by Mbd1 in Adult Neural Stem/Progenitor Cells *. *J. Biol. Chem.* **283**, 27644–27652 (2008).

46. Augustyniak, J. *et al.* Bezafibrate Upregulates Mitochondrial Biogenesis and Influence Neural Differentiation of Human-Induced Pluripotent Stem Cells. *Mol. Neurobiol.* **56**, 4346–4363 (2019).

47. Wu, Q. *et al.* DISC1 Regulates the Proliferation and Migration of Mouse Neural

Stem/Progenitor Cells through Pax5, Sox2, Dll1 and Neurog2. *Front. Cell. Neurosci.* **11**, 261 (2017).

48. Wang, H., Gui, H., Rallo, M. S., Xu, Z. & Matise, M. P. Atrophin protein RERE positively regulates Notch targets in the developing vertebrate spinal cord. *J. Neurochem.* **141**, 347–357 (2017).
49. Tiethof, A. K., Richardson, J. R. & Hart, R. P. Knockdown of Butyrylcholinesterase but Not Inhibition by Chlorpyrifos Alters Early Differentiation Mechanisms in Human Neural Stem Cells. *Toxics* **6**, (2018).
50. Barber, B. A. *et al.* Dynamic expression of MEIS1 homeoprotein in E14.5 forebrain and differentiated forebrain-derived neural stem cells. *Ann. Anat.* **195**, 431–440 (2013).
51. Katoh, M. Network of WNT and other regulatory signaling cascades in pluripotent stem cells and cancer stem cells. *Curr. Pharm. Biotechnol.* **12**, 160–170 (2011).
52. Pataskar, A. *et al.* NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **35**, 24–45 (2016).
53. Yang, S.-H. *et al.* Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency. *Cell Rep.* **7**, 1968–1981 (2014).
54. Hong, S. *et al.* Functional roles of Nurr1, Pitx3, and Lmx1a in neurogenesis and phenotype specification of dopamine neurons during in vitro differentiation of embryonic stem cells. *Stem Cells Dev.* **23**, 477–487 (2014).
55. Luo, Z. *et al.* Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol. Cell* **57**, 685–694 (2015).
56. Ding, N. *et al.* MED19 and MED26 are synergistic functional targets of the RE1 silencing transcription factor in epigenetic silencing of neuronal gene expression. *J. Biol. Chem.* **284**, 2648–2656 (2009).
57. Ding, Z., Gillespie, L. L. & Paterno, G. D. Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol. Cell. Biol.* **23**, 250–258 (2003).
58. Senft, A. D. *et al.* Combinatorial Smad2/3 Activities Downstream of Nodal Signaling

Maintain Embryonic/Extra-Embryonic Cell Identities during Lineage Priming. *Cell Rep.* **24**, 1977–1985.e7 (2018).

59. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
60. Curry, E. *et al.* Dual EZH2 and EHMT2 histone methyltransferase inhibition increases biological efficacy in breast cancer cells. *Clin. Epigenetics* **7**, 84 (2015).
61. Jiang, W., Wang, J. & Zhang, Y. Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. *Cell Res.* **23**, 122–130 (2013).
62. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304 (2009).
63. Newell-Price, J., Clark, A. J. L. & King, P. DNA Methylation and Silencing of Gene Expression. *Trends Endocrinol. Metab.* **11**, 142–148 (2000).
64. Pöschl, J. *et al.* Expression of BARHL1 in medulloblastoma is associated with prolonged survival in mice and humans. *Oncogene* **30**, 4721–4730 (2011).
65. Chatterjee, A. *et al.* Genome-wide methylation sequencing of paired primary and metastatic cell lines identifies common DNA methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis. *Oncotarget* **8**, 6085–6101 (2017).
66. Kumar, V. *et al.* Impact of miRNA-mRNA Profiling and Their Correlation on Medulloblastoma Tumorigenesis. *Mol. Ther. Nucleic Acids* **12**, 490–503 (2018).
67. Dixit, R. *et al.* Neurog1 and Neurog2 control two waves of neuronal differentiation in the piriform cortex. *J. Neurosci.* **34**, 539–553 (2014).
68. Massari, M. E. & Murre, C. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology* vol. 20 429–440 (2000).
69. Terada, Y. *et al.* Human Pluripotent Stem Cell-Derived Tumor Model Uncovers the Embryonic Stem Cell Signature as a Key Driver in Atypical Teratoid/Rhabdoid Tumor. *Cell Rep.* **26**, 2608–2621.e6 (2019).

70. Nazor, K. L. *et al.* Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012).

71. Colunga, T. *et al.* Human Pluripotent Stem Cell-Derived Multipotent Vascular Progenitors of the Mesothelium Lineage Have Utility in Tissue Engineering and Repair. *Cell Rep.* **26**, 2566–2579.e10 (2019).

72. Seo, S., Lim, J.-W., Yellajoshyula, D., Chang, L.-W. & Kroll, K. L. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *EMBO J.* **26**, 5093–5108 (2007).

73. Seko, Y. *et al.* Derivation of human differential photoreceptor cells from adult human dermal fibroblasts by defined combinations of CRX, RAX, OTX2 and NEUROD. *Genes Cells* **19**, 198–208 (2014).

74. Monsoro-Burq, A. H. PAX transcription factors in neural crest development. *Semin. Cell Dev. Biol.* **44**, 87–96 (2015).

75. Hufnagel, R. B., Le, T. T., Riesenberg, A. L. & Brown, N. L. Neurog2 controls the leading edge of neurogenesis in the mammalian retina. *Dev. Biol.* **340**, 490–503 (2010).

76. Henke, R. M., Meredith, D. M., Borromeo, M. D., Savage, T. K. & Johnson, J. E. Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. *Dev. Biol.* **328**, 529–540 (2009).

77. Kageyama, R., Shimojo, H. & Ohtsuka, T. Dynamic control of neural stem cells by bHLH factors. *Neurosci. Res.* **138**, 12–18 (2019).

78. Pfister, A. S., Tanneberger, K., Schambony, A. & Behrens, J. Amer2 protein is a novel negative regulator of Wnt/β-catenin signaling involved in neuroectodermal patterning. *J. Biol. Chem.* **287**, 1734–1741 (2012).

79. Medrihan, L. *et al.* Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. *J. Physiol.* **587**, 5095–5106 (2009).

80. Li, S., Qiu, F., Xu, A., Price, S. M. & Xiang, M. Barhl1 Regulates Migration and Survival of Cerebellar Granule Cells by Controlling Expression of the Neurotrophin-3 Gene. *J.*

Neurosci. **24**, 3104–3114 (2004).

81. Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* **18**, 316–328 (2010).
82. Albanese, P., Belin, M.-F. & Delattre, O. The tumour suppressor hSNF5/INI1 controls the differentiation potential of malignant rhabdoid cells. *Eur. J. Cancer* **42**, 2326–2334 (2006).
83. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
84. Kia, S. K., Gorski, M. M., Giannakopoulos, S. & Verrijzer, C. P. SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. *Mol. Cell. Biol.* **28**, 3457–3464 (2008).
85. Wang, W.-D., Melville, D. B., Montero-Balaguer, M., Hatzopoulos, A. K. & Knapik, E. W. Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Dev. Biol.* **360**, 173–185 (2011).
86. Tchieu, J. *et al.* A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell Stem Cell* **21**, 399–410.e7 (2017).
87. Alimova, I. *et al.* Inhibition of EZH2 suppresses self-renewal and induces radiation sensitivity in atypical rhabdoid teratoid tumor cells. *Neuro. Oncol.* **15**, 149–160 (2013).
88. Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7922–7927 (2013).
89. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).
90. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
91. Team, R. C. RA language and environment for statistical computing. R Foundation for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. (2019).
92. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS*

Comput. Biol. **9**, e1003118 (2013).

93. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).
94. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).
95. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* **7**, 1338 (2018).
96. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
97. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
98. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
99. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* vol. 57 289–300 (1995).
100. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
101. Henriquez, N. V. *et al.* Comparative expression analysis reveals lineage relationships between human and murine gliomas and a dominance of glial signatures during tumor propagation in vitro. *Cancer Res.* **73**, 5834–5844 (2013).
102. Birks, D. K. *et al.* Pediatric rhabdoid tumors of kidney and brain show many differences in gene expression but share dysregulation of cell cycle and epigenetic effector genes. *Pediatr. Blood Cancer* **60**, 1095–1102 (2013).
103. Sturm, D. *et al.* New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell* vol. 164 1060–1072 (2016).
104. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

105. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Erratum: Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 888 (2016).

106. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

107. Ferrero, E. Using regulatory genomics data to interpret the function of disease variants and prioritise genes from expression studies. *F1000Res.* **7**, 121 (2018).

108. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology* vol. 19 (2018).

109. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* vol. 485 376–380 (2012).

110. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

111. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* vol. 33 3088–3090 (2017).