의학박사 학위논문

Methylation and Molecular Analysis of Ependymoma in association with Location and Age

위치와 나이를 고려한 뇌실막종의 메틸화 및 분자적 분석

2017년 7월

서울대학교 대학원

의학과 병리학 전공

조 화 진

ABSTRACT

Methylation and Molecular Analysis of Ependymoma in association with Location and Age

Hwa Jin Cho
Department of Medicine
The Graduate School
Seoul National University

Ependymomas are tumors of the central nervous system (CNS) that occur in patients of all ages. Important parameters of the World Health Organization (WHO) guidelines to improve the prognosis of CNS-related tumors such as ependymomas include patient age, tumor resection range, tumor location, and histopathology grade. However, recent studies suggest that more focus on tumor location and patient age in relation to transcriptomics, genetics, and epigenetics can provide a more accurate assessment of clinical prognosis compared to adoption of the WHO-delegated grading scheme alone. In this study, we collected primary intracranial ependymoma tumor tissues from 13 Korean patients (nine adults and four children), and performed whole-exome sequencing, analysis of an ion-proton comprehensive cancer panel, RNA sequencing, and methylation analysis with the Infinium HumanMethylation450 BeadChip array to identify somatic mutations, copy number variations, and fusion genes, and determine the methylation status

and associated genes exhibiting significant differential expression according to variations in tumor location and patient age. We found some somatic mutations with a *YAP1* fusion in a pediatric patient with a good prognosis, in agreement with a previous study, as well as a few novel gene fusions. However, methylation microarray revealed a novel finding: genes associated with neurogenesis and neuron differentiation were significantly hypermethylated in the adult group and genes activated in the *HOX* genes related to the hindbrain development pathway were significantly hypermethylated in the supratentorial location group. Our collective results from these analyses confirmed the existence of significant differential expression in tumor-specific genes according to tumor location and patient age. These results provide valuable insight into the epigenetic and genetic profiles of intracranial ependymomas and suggest potential novel treatment strategies based on location- and age-based ependymoma-related prognostic factors.

Keywords: Ependymoma, DNA Methylation, Molecular Analysis, Classification

Student Number: 2011-21905

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Ependymomas can develop in all age groups; however, the incidence is greatly affected by location (1), with 90% of tumors in pediatric cases occurring intracranially and 70% arising in the posterior fossa (PF) (2). Supratentorial (ST) ependymomas affect both pediatric and adult patients (1), in whom PF and spinal ependymomas arise with almost equal frequency (3).

The World Health Organization (WHO) classification of tumors of the central nervous system (CNS) delineates four important prognostic factors: age, extent of resection, location, and histopathological grading (4). Pediatric cases tend to be more severe because ependymomas commonly occur in the PF in children, whereas those in adults more frequently occur in the spinal cord. A multi-institutional retrospective analysis of pediatric ependymoma patients revealed that anaplastic histopathological features and incomplete tumor resection are indications of a poor outcome in patients younger than 3 years of age (5).

Molecular genetics studies have also revealed distinct features associated with ependymoma subgroups. Intracranial and spinal ependymomas showed mutually exclusive gene mutations (6), and differential gene expression patterns among different CNS regions were recently identified (7). In addition, Taylor et al. (7) found distinct patterns of gene expression and chromosome alterations among distinct ependymoma subsets, which correlated with the anatomic location of the tumor but not with clinical parameters or with tumor histological grade. Furthermore, ependymomas recapitulate the gene expression profiles of regionally specified radial glial cells, representing a mechanism that explains the clinical

heterogeneity among histologically similar ependymomas (7). This suggested that ependymoma treatment should target the appropriate cell-signaling pathways based on tumor location.

Palm et al. (8) performed class discovery through a hierarchical clustering approach using 1000 RNA transcripts, and found that both location and histological grade exhibited different gene expression patterns, suggesting the involvement of different pathways between low-grade (LG) and high-grade (HG) ependymomas. Korshunov et al. (9) identified recurrent genetic aberrations associated with ependymomas using array-based comprehensive genomic hybridization (aCGH), with correlations between copy number aberrations and clinical, histopathological, and survival data. They further identified three distinct subgroups showing characteristic patterns of genomic aberrations. Moreover, some studies have focused on the methylation phenotype associated with ependymoma using methylation probes such as that described by Rogers et al. (2), or through integration of methylation and mRNA expression data to identify changes in methylation-induced expression. Notable genes identified were those involved in the control of cell growth, cell death, and the immune system, including members of the c-Jun N-terminal kinase pathway and peroxisome proliferator-activated receptor-gamma binding. Interestingly, epigenetic silencing of tumor-suppressor genes by hypermethylation was only found to be effective in the pathogenesis of ST and spinal ependymomas, but not in PF ependymomas (2).

Mack et al. (10) studied epigenomic alterations in PF ependymomas, which occur more often in infants and children and are related to somatic single-nucleotide variants (SNVs); these tumors are relatively rare in comparison to other organ

malignancies. They divided PF ependymomas into two groups according to age: PF group A (PFA) and PF group B (PFB), with PFA predominantly found in infants and associated with a poor prognosis despite maximally aggressive therapy, and PFB occurring in older children and adults exhibiting favorable prognoses (10). These two groups represented two very distinct molecular subgroups based on unsupervised clustering of CpG-methylation sites; certain genes, including *CRIP1*, *CYP26C1*, and *PKP1*, exhibited increased CpG methylation in most of the PFA tumors. Furthermore, Witt et al. (11) delineated two clinically and molecularly distinct subgroups of PF ependymomas based on gene sets. Group A ependymomas were characterized by numerous cancer-related networks, including platelet-derived growth factor signaling and mitogen-activated protein kinase signaling, whereas group B ependymomas included genes involved in ciliogenesis/microtubule assembly and mitochondria/oxidative metabolism.

Therefore, subgroups according to tumor location and patient age are associated with distinct transcriptomic, genetic, epigenetic, and clinical features, indicating that these classifications could be more informative for prognostic prediction than WHO grading alone (10). Through the assessment of correct grouping and selection of corresponding genes in each group, clinicians might be able to identify risk groups for recurrence and identify patients most suitable for adjuvant therapy. Indeed, Pajtler et al. (14) assessed 500 ependymal tumors, including cases of subependymoma and myxopapillary ependymoma, which were classified across all CNS compartments, age groups, and histological grades. Through DNA-methylation profiling, they identified nine molecular subgroups, with the PFA and RELA-positive ependymoma groups exhibiting a dismal prognosis, indicating that

risk stratification by molecular subgrouping is superior to histologic grading.

Here, we analyzed the genomic and epigenomic alterations in Korean patients with intracranial ependymomas using three different methods of molecular analysis and Illumina 450K methylation arrays. Our results revealed differences and commonalities in ependymoma characteristics according to each method. Our findings will help to achieve a more accurate classification of molecular subgroups to facilitate prognosis and personalized medicine.

Materials and methods

Study subjects and experimental design for molecular analysis

We collected fresh-frozen samples from 13 patients who underwent tumor resection for ependymoma at Seoul National University Hospital (Seoul, South Korea). Nine adults and four children younger than 7 years old were included and the female-to-male ratio was 8:5. Nine tumors were located in the PF, whereas the others were ST ependymomas with no cases of spinal ependymoma in the cohort. In a pathological review, six tumors were diagnosed as grade II and seven were diagnosed as grade III anaplastic ependymoma in accordance with the WHO classification of tumors of the CNS (11). We further used the criteria of identification of increased cellularity, frequent mitosis, abundant necrosis, and microvascular proliferation. High mitotic activity was defined when at least five mitotic figures were detected based on the results from a European clinical trial (12). We also performed Ki-67 immunohistochemical staining in all samples to detect the proliferation index of the tumor growth fraction using an image analysis system (ScanScope XT; Aperio).

Ependymomas in two adults and three children were recurrent, with one resulting in death. The five surviving patients harbored grade-III tumors, four of which were located in the PF area.

Of the 13 samples, some specimens did not pass the DNA or RNA quality test and showed an insufficient DNA or RNA concentration. We conducted experiments

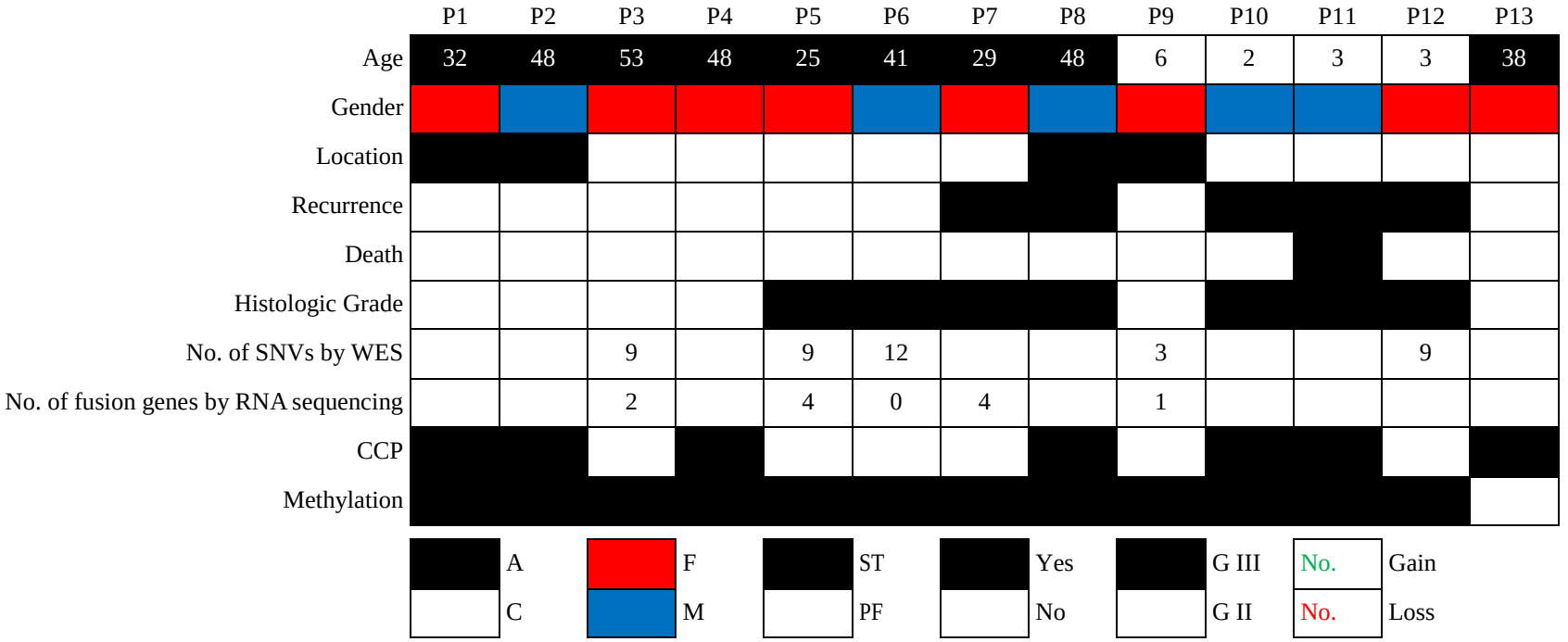
with only specimens that passed the quality check: 12 were used for methylation profiling, five were used for whole-exome sequencing (WES), seven were applied for targeted sequencing using an ion-proton comprehensive cancer panel (CCP), and five were subjected to RNA sequencing. Clinicopathological information of the patients and an overview of the experimental design are provided in Table 1, and the genomic and clinical characteristics of the patients are summarized in Table 2.

Table 1. Clinicopathological characteristics and overview of the experimental design.

										Experimetal method			
ID	Gender	Age	Histologic grade	Location	OP	Adjuvant Tx	Recurrence (n)	Death	F/U (m)	Methylation	WES	RNA sequencing	CCP
P1	F	32	LG	BG & thalamus	GTR	RT	N	N	21	Y	F	F	Y
P2	M	48	LG	Frontoparietal lobe	STR	GKRS	N	N	17	Y	F	F	Y
P3	F	53	LG	4^{th} ventricle	NTR	ND	N	N	30	Y	Y	Y	F
P4	F	48	LG	4^{th} ventricle	NTR	GKRS, RT	N	N	13	Y	F	F	Y
P5	F	25	HG	4^{th} ventricle	GTR	RT	N	N	16	Y	Y	Y	F
P6	M	41	HG	4^{th} ventricle	NA	RT	N	N	187	Y	Y	Y	F
P7	F	29	HG	4^{th} ventricle	GTR	RT	Y (1)	N	64	Y	F	Y	F
P8	M	48	HG	Temporal lobe, BG & thalamus	GTR	GKRS, RT	Y (1)	N	11	Y	F	F	Y
P9	F	6	LG	Frontal lobe	GTR	RT	N	N	49	Y	Y	Y	F
P10	M	2	HG	4^{th} ventricle & medulla	NTR	RT	Y (1)	N	72	Y	F	F	Y
P11	M	3	HG	4^{th} & 3^{rd} ventricle	NTR	GKRS, RT	Y (2)	Y	36	Y	F	F	Y
P12	F	3	HG	4^{th} ventricle & medulla	NTR	GKRS, RT	Y (5)	N	66	Y	Y	F	F
P13	F	38	LG	4^{th} ventricle	NA	ND	N	N	41	F	F	F	Y
Total													
13	F 8 M 5	Adult 9 Child 4	HG 7 LG 6	ST: 4 PF: 9			5	1		12	5	5	7

OP operation method, Tx treatment, F/U follow up, m month, WES whole exome sequencing, CCP comprehensive cancer panel, F female, M male, LG low grade HG high grade, GTR gross total resetion, STR subtotal resection, NTR,near total resection, NA not applicable, RT radiation therapy, GKRS gamma knife radiation surgery, ND not done, N no, Y yes, F failed

Table 2. Summary of clinicopathologic and genomic characteristics.



A adult, C child, F female, M male, ST supratentorial, PF posterior fossa, G grade

DNA methylation

DNA extraction and preparation

Genomic DNA (at least 500 ng) was isolated from all tissues. Bisulfite conversion of genomic DNA was achieved using the EZ DNA methylation gold kit (Zymo Research, Irvine, CA, USA). Bisulfite treatment changes the unmethylated cytosine nucleotides to thymidines, whereas methylated cytosines remain unchanged. This difference allows for the detection of C/T nucleotide polymorphisms at each CpG site. Genomic DNA was sent to Macrogen (Seoul, South Korea) for hybridization on Illumina 450K methylation arrays (Illumina, San Diego, CA, USA). We used the Illumina Infinium human methylation 450 BeadChip kit (Illumina) to provide coverage of CpG sites throughout the genetic regions according to the manufacturer instructions. BeadChips were scanned with an Illumina iScan scanner (Illumina).

Data processing and analysis

Following image analysis, data processing was performed using Illumina GenomeStudio software version 2011.1 (Methylation Module v1.9.0; Illumina) and R version 3.0.2 (http://www.r-progect.org), and data preprocessing and background correction were performed using the R library and dye-bias equalization. Detected CpGs were filtered for each sample according to the *p*-value. On average, ~485,139.3 and 485,262.6 CpGs were detected with a $p < 0.05$. We then used the beta mixture quantile in the R package to correct for probe-design bias. To perform data transformation, the β-value, M-value, and δ-mean were calculated. The β-value represents the ratio of methylated probe intensity to the sum of methylated

and unmethylated probe intensities. The M-value is calculated as the $\log_2$-ratio of the intensities of the methylated probes versus those of the unmethylated probes (13), and δ-mean represents the difference between the average β-value of the experimental results and the controls. The odds ratio was calculated by transforming the δ-mean into the M-value and measuring the ratio between the unmethylated intensity and methylated intensity for the test sample against that of a control. Fold change was calculated to represent the ratio of methylation rates between test results and controls in the expression analysis.

Hierarchical clustering analyses

Using the M-values for the significant data, we plotted a heat map by hierarchical clustering based on the distance similarity for samples and CpGs according to the Euclidean distance and complete linkage.

Assessment of differential methylation patterns

To determine the difference in DNA methylation profiles among samples, especially according to patient and clinical attributes such as age and tumor location, we adopted two basic approaches. First, we classified the methylation status of CpG sites in the regulatory regions of target genes as hyper- or hypomethylated depending on their methylation level. We adopted a β-value >0.7 and <0.3 as the cut-offs for hypermethylation and hypomethylation, respectively. For the second approach, we tried to determine the difference in the absolute DNA methylation status between samples, regardless of their classification as hypermethylated or hypomethylated groups. For instance, if the β-value from a

pediatric patient's sample is 0 and that of an adult patient's sample is 0.6, the first approach would not detect this difference as statistically significant, despite the 0.6 β-value difference. In some cases, the difference obtained with the second method was greater than that in the first approach. Therefore, to capture the absolute DNA methylation difference, we subtracted the β-values between samples for the same CpG sites from both the age and tumor location groups. In this approach, we applied a β-value cut-off of 0.4 to identify significant methylation differences between groups.

Gene-enrichment and functional-annotation analyses

Through these two methylation difference measurement approaches, we could successfully extract target genes showing a significant difference in DNA methylation of the regulatory regions, especially within the promoter and gene body regions. To determine their biological functions, we performed a gene set analysis using ToppGene, a gene list enrichment analysis and candidate gene prioritization system based on functional annotations and protein interaction networks (14).

Merging with public microarray data

We compared our methylation data regarding age and tumor location with publicly available datasets based on Agilent microarray data from The German Cancer Research Center (DKFZ, Heidelberg; GSE27287) to validate and reduce the number of candidate genes. The public data are associated with information on age, gender, and location. We collected genes showing a β-value difference > 0.4, which

were merged with the microarray data by matching the identified hypermethylated genes and concurrent downregulated genes of the microarray data. Differentially expressed genes between groups were selected, and those showing a Benjamini-Hochberg q-value < 0.05 were selected after performing a *t*-test.

WES

DNA sequencing

DNA from five resected fresh-frozen ependymoma samples was extracted using the EZ DNA methylation gold kit (Zymo Research). DNA from blood samples of the same patients were used as germline controls. To enrich the coding regions, the SureSelect target-enrichment system capture process (Agilent, Santa Clara, CA, USA) was used. Sequence reads were produced by Illumina HiSeq 2500 (15).

Data analysis

Raw sequence reads were mapped against the hg19 reference genome using a Burrows-Wheeler aligner (16). After creating a BAM file, polymerase chain reaction (PCR) duplicates were removed using Picard. (https://broadinstitute.github.io/picard/index.html). Insertion and deletion (indel) realignment and base quality score recalibration were performed using Genome Analysis Tool Kit (17, 18). VarScan2 software was utilized to detect somatic SNVs and small indels (19). ANNOVAR was used to annotate variants. To identify confident and rare somatic variants in the coding regions and splice sites, we applied the following filtering criteria: 1) tumor total allele count $\geq$ 10, tumor altered allele count $\geq$ 3, and normal altered allele count = 0; and 2) rare variants based on a minor allele frequency (MAF) < 0.5% in Exome Aggregation Consortium Version 0.3 (ExAC03, http://exac.broadinstitute.org) (20).

CCP

To identify the mutation status in 409 cancer-related genes, we conducted targeted sequencing of seven tumor samples using the Ion AmpliSeq Comprehensive Cancer Panel (Life Technologies; Thermo Fisher Scientific, Waltham, MA, USA). Mutatations on the list showing non-synonymous SNVs and indels in coding regions with an altered allele frequency ≥ 0.1 and rare variants based on a MAF < 0.5% in Exome Aggregation Consortium Version 0.3 were used for further analysis.

RNA sequencing

Using the Illumina TruSeq RNA sample preparation kit (Illumina), mRNA was converted into a library of template molecules for subsequent cluster generation. Purified mRNA was fragmented into small fragments copied into cDNA using reverse transcriptase. DNA polymerase I and RNase H were used to synthesize second-strand cDNA. An additional adenine was added to the 3′ end of the cDNA fragments and ligated to the adapters. Products were enriched by PCR using HiSeq 2000 (Illumina) for bridged-amplification reactions and imaging, and images of single-base extensions at a specific cluster were generated.

Sequence quality checks were performed using FastQC (v0.10.0; www.bioinformatics.babraham.ac.uk/projects/fastqc/). Alignment was performed by TopHat (v1.3.3; https://ccb.jhu.edu/software/tophat/index.shtml), a fast splice-junction mapper for RNA-Seq reads, with hg19 as reference (21). Cufflinks (v2.0.2; http://cole-trapnell-lab.github.io/cufflinks/) was used for transcript assembly and testing for the differential expression and regulation of RNA-sequencing samples (22). Gene expression levels were expressed in units of fragments per kilobase of transcript per million mapped reads (FPKM).

To detect SNVs and establish annotations, SAMtools was used to manipulate alignments in the SAM format (http://samtools.sourceforge.net/cns0.shtml). We used ANNOVAR (http://www.openbioinformatics.org/annovar/) to annotate functional genetic variants and DeFuse (v0.4.3; http://sourceforge.net) was used to discover fusion genes.

Genes exhibiting differences in expression between groups were detected by comparing the mean of the original value for each group against the mean of the Z-

score for each group. Genes exhibiting the largest differences in expression were analyzed and validated according to a literature search for their relevance to ependymoma and inclusion in the human disease database (MalaCards; http://malacards.org/card/ependymoma?limit[MaladiesGenes]=128#related_genes) A *t*-test was applied to analyze expression differences in different groups.

Results

Histopathologic features

Of the 13 samples, six were diagnosed as grade II ependymoma and seven were diagnosed as grade III anaplastic ependymoma. Increased cellularity was observed in five of the grade II ependymomas and in all of the grade III ependymomas. The mitotic count ranged from 0 to 4 (mean: 2.5) in cases of grade II ependymoma and from 11 to 26 (mean 19.0) in cases of grade III ependymoma. Microvascular proliferation was observed in only one case of grade III ependymoma. Abundant necrosis was observed in four cases of grade III ependymoma and focal necrosis was observed in three cases of grade II ependymoma. The Ki-67 staining index ranged from less than 1% to 11.64% (mean 5.0%) in grade II ependymomas, and ranged from 4.48% to 32.97% (mean 16.2%) in grade III ependymomas.

Methylation

Hierarchical clustering

Significant differences in the methylation rates associated with patient age and tumor grade and location were selected and analyzed by hierarchical clustering. Hierarchical clustering according to age was performed by satisfying the significance criterion ($p < 0.05$) with an absolute δ-mean value > 0.2 (Figure 1). The ST and PF groups (location) were clustered in each age group. Hierarchical clustering by grade showed clustering of children and adults in the PF group (Figure 2), although no further clustering according to ST and PF areas could be achieved (data not shown); therefore, clustering was divided into the PF-adult and PF-children/ST groups for further comparisons. Overall, methylation patterns showed meaningful differences with respect to both patient age and location.

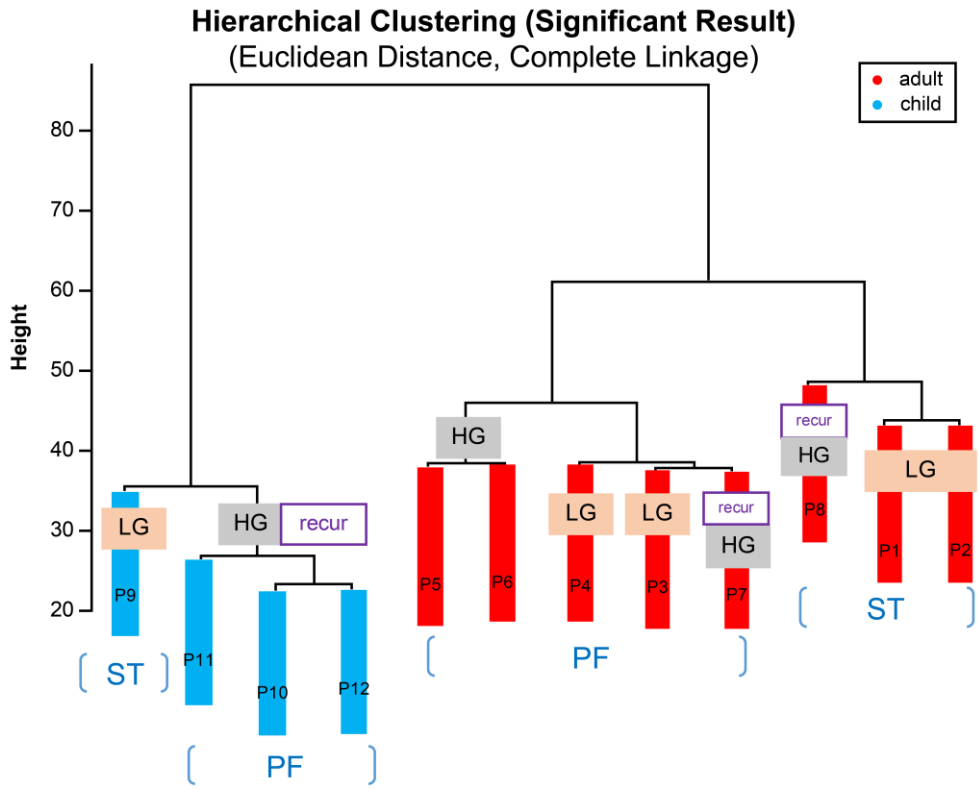


Figure 1. Two-way hierarchical clustering dendrogram of methylation patterns between adults and children. Tumor locations in the brain are clustered according to the supratentorial (ST) and posterior fossa (PT).

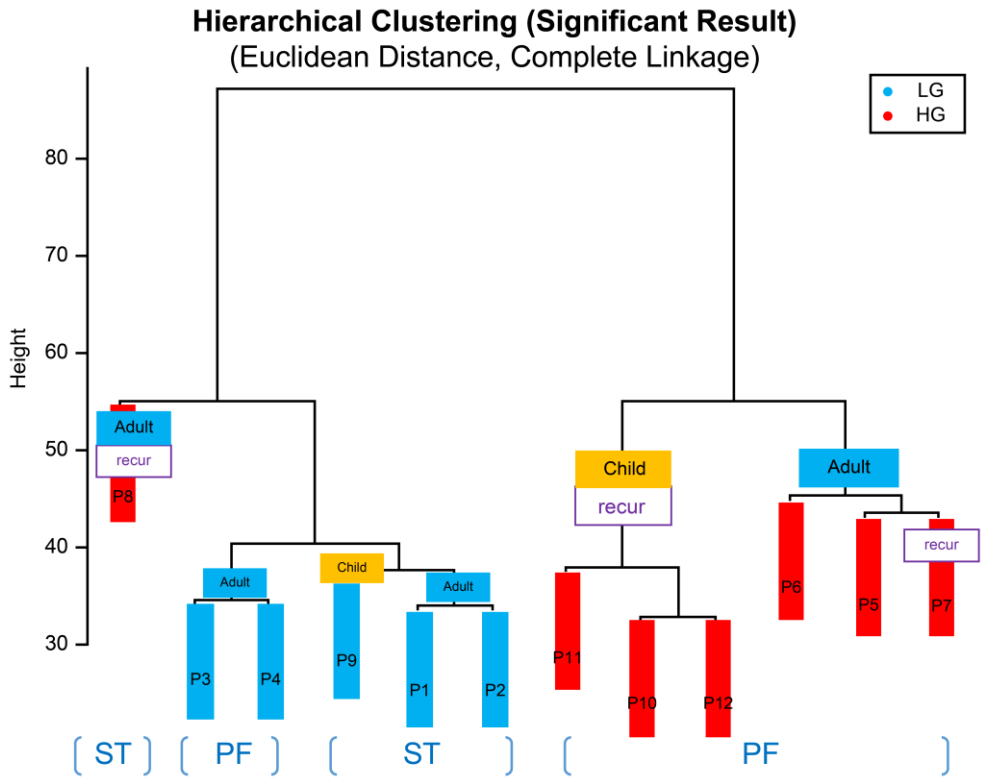


Figure 2. Two-way hierarchical clustering dendrogram of methylation patterns between low-grade and high-grade tumors. High-grade tumors located in the posterior fossa (PF) clustered in children and adults.

Methylation profiles based on age and location

We counted the number of genes exhibiting significant differences in methylation based on patient age and location. Among these genes expressed in the child and adult groups, 114 (1.9%) genes were significantly enriched in the neuroactive ligand-receptor interaction Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway category, with a false discovery rate < 0.05 (Table 3). Among the genes showing differential methylation patterns between the ST and PF groups, 124 (3.1%) exhibited enrichment in the KEGG pathway category related to olfactory transduction (Table 3).

Table 3. Genes exhibiting significant differences in methylation based on patient age and location according to KEGG pathway enrichment categories.

	ID	Pathway	Hit count	List total	Benjamini
Age	hsa4080	Neuroactive ligand-receptor interaction	114	1582	8.66E-04
	hsa4510	Focal adhesion	93	1582	5.71E-04
	hsa5414	Dilated cardiomyopathy	49	1582	1.13E-03
	hsa5412	Arrhythmogenic right ventricular cardiomyopathy	42	1582	1.24E-03
Location	hsa4740	Olfactory transduction	124	1119	1.19E-04

Many genes were identified to have a β-value difference > 0.4 between adults and children (data not shown): 66 genes were included in the adult-hypomethylated group and 425 genes were included in the adult-hypermethylated group. Three of these genes (*PDE4C*, *HOXA9*, and *TBX5*) in the adult-hypermethylated and child-hypomethylated group showed the most significant age-related differences, as reported by Koch et al. (23) in an epigenetic study.

We next classified the genes showing hypomethylation and hypermethylation in each age group (Table 4). Overall, 10 genes showed hypomethylation in adults, including two genes located in the promoter regions and eight genes located in the body regions. By contrast, 88 genes showed hypermethylation in adults, including 25 genes located in the promoter regions and 65 genes located in the body regions.

Table 4. Genes displaying hypermethylation and hypomethylation in adults and children according to the CpG site.

	CpG location	
Age group (β value)	Promoter	Body
Adults (<0.3) Children (>0.7)	LOC641518 NRN1L	BTBD19, CLDN6, FNDC1, IGF2BP1, LEF1, LTBP4, NRXN2, WNK4
Adults (>0.7) Children (<0.3)	C10orf108, CMYA5, COMMD3, DEFB133, FGF22, GFI1, HIST1H4I, HOXC8, IPO11, LOC285375, MAGEC3, MAP9, MDFIC, MIR503, NKX2-5, NTNG1, RPA3, SNX31, SOSTDC1, SYT5, TCF21, TFAP2A, THRB, ZBTB39, ZC3H12D	ADARB2, AGAP1, ASPH, B3GALT2, C17orf50, C19orf38, C20orf56, CABP1, CD82, CDC73, CDKN2C, CNPY1, CNTNAP5, COL21A1, DIP2C, EIF2C2, EIF4G3, EPB41L3, EXOC2, EXOC4, F2RL1, GABRB3, GALNT9, GIMAP5, GSTCD, HIST1H2BK, HLA-DRB6, HMHA1, HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, HOXD1, IPO11, KCNMA1, LOC404266, LUM, MGC16121, MSLN, NEK9, NELL1, NEUROG1, NR3C1, PRDM13, PRLHR, PTPRH, PTPRN2, RBM20, SAMSN1, SLC6A20, SPATS2L, SPN, STXBP5, TFAP2A, TFAP2B, TMEM132D, TMEM173, TNN, UNC5A, UPP1, WNT3, ZNF683, ZNF827

Gene Ontology (GO) analysis revealed that three (*FNDC1*, *CLDN6*, and *WNK4*) of the 420 genes in the cell-cell junction GO category showed hypomethylation in the adult group (Table 5). A much higher proportion of genes displaying hypermethylation in adults was included in GO sets, including 19 of the 1638 genes associated with neurogenesis and 16 of the 1405 genes associated with neuron differentiation (Table 5). Among the extracted genes, only *HOXA9* has been previously reported as a significant age-associated gene (23).

Table 5. Ontology of genes displaying hypermethylation and hypomethylation in adults and children.

Age group (β value)	GO ID	Function	Hit count	List total	Hit in query list
Adults (<0.3) & Children (>0.7)	5911	cell-cell junction	3	420	FNDC1, CLDN6, WNK4
Adults (>0.7) & Children (<0.3)	22008	neurogenesis	19	1638	GABRB3, CDKN2C, UNC5A, HOXC8, NEUROG1, HOXD1, STXBP5, WNT3, TNN, KCNMA1, NTNG1, PRDM13, NKX2-5, LUM, NR3C1, EPB41L3, TFAP2A, GFI1, EXOC4
	30182	neuron differentiation	16	1405	GABRB3, UNC5A, HOXC8, NEUROG1, HOXD1, STXBP5, WNT3, TNN, KCNMA1, NTNG1, NKX2-5, LUM, NR3C1, EPB41L3, TFAP2A, GFI1
	9790	embryo development	15	1146	HOXA7, HOXB6, HOXC4, HOXC8, NEUROG1, THRB, TCF21, WNT3, SOSTDC1, NKX2-5, NR3C1, TFAP2A, TFAP2B, GFI1, EXOC4
	9887	animal organ morphogenesis	15	1130	HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, NEUROG1, HOXD1, TCF21, WNT3, NKX2-5, CDC73, TFAP2A, GFI1, EXOC4
	48598	embryonic morphogenesis	12	652	HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, HOXC8, WNT3, ASPH, SOSTDC1 NKX2-5,TFAP2A
	7389	pattern specification process	12	495	HOXA7, HOXA9, HOXB6, HOXC4, NEUROG1, TCF21, WNT3, NKX2-5, CDC73, TFAP2A, GFI1, EXOC4
	1501	skeletal system development	11	534	HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, HOXC8, THRB, HOXD1, LUM, TFAP2A
	3002	regionalization	9	370	HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, HOXC8, WNT3, NKX2-5
	48562	embryonic organ morphogenesis	8	327	HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6, HOXC8, WNT3
	9952	anterior/posterior pattern specification	8	225	HOXA7, HOXB6, HOXC4, NEUROG1, TCF21, NKX2-5, TFAP2A, GFI1
	48706	embryonic skeletal system development	7	137	HOXA7, HOXA9, HOXB6, HOXC5, HOXC6, HOXD1, TFAP2A
	43525	positive regulation of neuron apoptotic process	4	69	HOXA9, WNT3, TFAP2A, TFAP2B
	35136	forelimb morphogenesis	4	45	KCNMA1, NR3C1, TFAP2A, TFAP2B

Moreover, many genes showed a β-value difference > 0.4 between the PF and ST regions (data not shown). Sixty-one genes were included in the PF-hypomethylated group and 14 genes were included in the PF-hypermethylated group. No significant genes were detected in the PF-hypermethylated group by GO analysis.

We classified genes with hypomethylated and hypermethylated phenotypes in each location group (Table 6). Overall, 18 genes showed a hypomethylated phenotype in the PF, including hypomethylation in the promoter regions for 10 genes and in the body region for 9 genes. Only two genes showed a hypermethylated phenotype in the PF, located in the body region in both cases.

Table 6. Genes displaying hypermethylation and hypomethylation in the supratentorial and posterior fossa regional groups according to CpG sites.

Location group (β value)	CpG location	
	Promoter	Body
PF (<0.3) & ST (>0.7)	DYRK2, FHIT, HLA-DRB1, HOXD4, MIR10B, MRI1, SNORD22, SNORD30, SNORD31, ZNF709	ABR, GREB1, GSC, HLA, DRB1, HLA-DRB6, HOXC4, PAX6, SHANK2, SNHG1
PF (>0.7) & ST (<0.3)		MIR548N, SPTB

PF posterior fossa, ST supratentorial

GO and pathway analysis revealed that five genes (*HOXD4*, *GSC*, *HOXC4*, *PAX6*, and *ABR*) overlapped in a biological process category (Table 7). Among them, *HOXD4*, *HOXC4*, and *PAX6* identified in the PF-hypomethylated/ST-hypermethylated group are associated with the pathway related to activation of *HOX* genes in hindbrain development during early embryogenesis. There were no significantly enriched genes detected in the PF-hypermethylated/ST-hypomethylated group.

Table 7. Ontology of genes displaying hypermethylation and hypomethylation in the supratentorial and posterior fossa regional groups.

Location group (β value)	ID	Function	Hit count	List total	Hit in Query List
	GO ID				
PF (<0.3) & ST (>0.7)	9790	embryo development	5	1146	HOXD4,GSC,HOXC4,PAX6,ABR
	9887	animal organ morphogenesis	5	1130	HOXD4,GSC,HOXC4,PAX6,ABR
	48598	embryonic morphogenesis	5	652	HOXD4,GSC,HOXC4,PAX6,ABR
	48568	embryonic organ development	5	483	HOXD4,GSC,HOXC4,PAX6,ABR
	48562	embryonic organ morphogenesis	5	327	HOXD4,GSC,HOXC4,PAX6,ABR
	7389	pattern specification process	4	495	HOXD4,GSC,HOXC4,PAX6
	3002	regionalization	4	370	HOXD4,GSC,HOXC4,PAX6
	48863	stem cell differentiation	3	303	HOXD4,GSC,PAX6
	90596	sensory organ morphogenesis	3	297	GSC,PAX6,ABR
	9952	anterior/posterior pattern specification	3	225	HOXD4,HOXC4,PAX6
	Reactome ID				
	1339140	Activation of anterior HOX genes in hindbrain development during early embryogenesis	3	127	HOXD4,HOXC4,PAX6
	1339139	Activation of HOX genes during differentiation	3	127	HOXD4,HOXC4,PAX6
PF (>0.7) & ST (<0.3)	No significant genes				

We also performed validation of the differentially expressed genes identified using publicly available datasets based on Agilent microarray data from the DKFZ, Heidelberg (GSE27287).

No genes overlapped between the DKFZ data and our differentially expressed genes selected with the criteria of a β-value >0.7 for hypermethylation and <0.3 for hypomethylation for either the age or location group comparisons. However, among the genes that showed a β-value difference > 0.4, a few differentially expressed and hypermethylated genes located in the gene body regions were reported to be downregulated in the microarray data (Table 8). No genes overlapped with the microarray dataset when the ST area was more hypermethylated than the PF area.

Table 8. Genes showing a significant correlation between methylation and genomic microarray data in different groups

	CpG location	
hypermethylated group	Promoter	Body
Children		IGF2BP1
Adults		ZFR2 ELTD1
PF		LHX2 ECM2 C15orf27

PF posterior fossa

WES

We performed WES for five paired tumor-normal tissue samples, including two LG tumors and three HG tumors. Three of the patients were adults and two were children. Four cases were located in the PF and one case was in the ST.

A total of 58 non-synonymous SNVs (nsSNVs) were identified in five paired samples, ranging from 7 to 14 per sample. The median number of nsSNVs per megabase was 0.26 in adults and 0.20 in children. After filtering for rare mutations, each sample had 3 to 12 nsSNVs. The mean number of rare nsSNVs in children was smaller than that in adults, although the difference was not statistically significant. A full list of the rare somatic nsSNVs identified is presented in Table 9.

Table 9. Summary of the full list of rare somatic non-synonymous single nucleotide variants identified in whole-exome sequencing.

Sample ID	Gene symbol	Transcript ID	Exonic function	CDS change	AA change	dbSNP	COSMIC ID	Normal #-of Ref. allele	Tumor #-of Ref. allele	# Alt. allele	freq.
P3	PRAMEF25	NM_001306072	nonsynonymous SNV	c.C55G	p.L19V			8	8	7	0.47
P3	IGFN1	NM_001164586	nonsynonymous SNV	c.C5231A	p.A1744E			8	9	9	0.50
P3	KCNH7	NM_173162	nonsynonymous SNV	c.G959A	p.S320N			67	19	7	0.27
P3	HARS2	NM_001278731	nonsynonymous SNV	c.A53G	p.Q18R			154	139	44	0.24
P3	TRPV6	NM_018646	nonsynonymous SNV	c.G298A	p.A100T			114	95	33	0.26
P3	OR4C13	NM_001001955	nonsynonymous SNV	c.T310A	p.F104I		COSM927746	22	8	4	0.33
P3	UBASH3B	NM_032873	nonsynonymous SNV	c.T1584A	p.D528E			36	17	10	0.37
P3	SPSB3	NM_080861	nonsynonymous SNV	c.G685A	p.G229S			16	9	10	0.53
P3	ACSM5	NM_017888	nonsynonymous SNV	c.C1139T	p.A380V			29	12	3	0.20
P6	CDKN2C	NM_078626	nonsynonymous SNV	c.C230T	p.A77V		COSM191010	137	88	47	0.35
P6	FMN2	NM_020066	nonsynonymous SNV	c.C3029T	p.P1010L		COSM1285015	18	17	5	0.23
P6	BAZ2B	NM_001289975	nonsynonymous SNV	c.A1142G	p.N381S	rs368374349		127	96	29	0.23
P6	TRIO	NM_007118	nonsynonymous SNV	c.G8613A	p.M2871I			57	47	17	0.27
P6	SNX13	NM_015132	nonsynonymous SNV	c.A2101G	p.R701G			150	104	43	0.29
P6	ELN	NM_001278913	nonsynonymous SNV	c.C1801T	p.L601F			100	99	32	0.24
P6	PLXNA4	NM_020911	nonsynonymous SNV	c.C3176T	p.A1059V			56	46	19	0.29
P6	PARP12	NM_022750	nonsynonymous SNV	c.G1969A	p.D657N	rs201580615		35	24	10	0.29
P6	LPAR5	NM_001142961	nonsynonymous SNV	c.G832C	p.V278L			54	43	40	0.48
P6	CDH19	NM_001271028	nonsynonymous SNV	c.A431G	p.N144S			48	25	14	0.36
P6	NFATC1	NM_001278669	nonsynonymous SNV	c.C1040T	p.A347V			65	25	14	0.36
P6	BCOR	NM_001123383	stopgain	c.C2539T	p.Q847X			31	17	6	0.26

P5	POTEF	NM_001099771	nonsynonymous SNV	c.A1330G	p.N444D			13	12	5	0.29
P5	STAC	NM_001292049	nonsynonymous SNV	c.A104C	p.E35A			218	69	74	0.52
P5	MYBPC3	NM_000256	nonsynonymous SNV	c.G1294A	p.A432T	rs371167525		128	81	29	0.26
P5	GOLGA6L2	NM_001304388	nonsynonymous SNV	c.A2345C	p.D782A			22	15	5	0.25
P5	CASC5	NM_144508	nonsynonymous SNV	c.G6327T	p.W2109C			44	39	12	0.24
P5	LOC10106038	NM_001001418	nonsynonymous SNV	c.G429C	p.Q143H			24	15	4	0.21
P5	L̄OC10106038	NM_001001418	nonsynonymous SNV	c.G431A	p.R144H			24	16	4	0.20
P5	Z̄NF585B	NM_152279	nonsynonymous SNV	c.G1879A	p.V627I			18	16	5	0.24
P5	MAGEC1	NM_005462	nonsynonymous SNV	c.C820T	p.P274S	rs146798989		16	9	3	0.25
P9	NBPF15	NM_173638	nonsynonymous SNV	c.A715G	p.T239A	rs202133347		20	25	7	0.22
P9	MUC3A	NM_005960	nonsynonymous SNV	c.A1325C	p.N442T	rs76608602		31	14	5	0.26
P9	LURAP1L	NM_203403	nonsynonymous SNV	c.T166G	p.C56G	rs201963967		59	36	13	0.28
P12	TMEM200B	NM_001003682	nonsynonymous SNV	c.C853G	p.R285G			75	39	20	0.34
P12	OR2M5	NM_001004690	nonsynonymous SNV	c.G365A	p.R122H			11	8	10	0.56
P12	RIF1	NM_001177665	nonsynonymous SNV	c.A3545C	p.K1182T			128	64	17	0.21
P12	WDFY1	NM_020830	nonsynonymous SNV	c.C955A	p.Q319K			17	8	3	0.27
P12	AGAP5	NM_001144000	nonsynonymous SNV	c.A742G	p.M248V			20	7	3	0.30
P12	OR8B8	NM_012378	nonsynonymous SNV	c.T542A	p.I181N			232	127	59	0.32
P12	NBR1	NM_001291572	nonsynonymous SNV	c.A2300C	p.N767T			21	24	7	0.23
P12	ASXL1	NM_015338	stopgain	c.C2278T	p.Q760X		COSM97035	140	79	47	0.37
P12	MAP7D2	NM_001168466	stopgain	c.G1159T	p.E387X			48	28	14	0.33

CDS coding DNA sequence, AA amino acid, # Number, Ref. reference, Alt., alternate, freq frequency

In four samples with both WES and RNA sequencing data available, the SNVs were checked to determine whether they were transcribed or edited in RNA sequencing. SNVs identified in seven genes (*HARS2, SPSB3, BAZ2B, TRIO, SNX13, PLXNA4,* and *PARP12*) in WES were transcribed in the RNA sequencing analysis (Table 10). No case of editing was identified.

Table 10. List of transcribed genes identified by whole-exome and RNA sequencing.

						RNA sequencing	
Sample ID	Gene symbol	Transcript ID	Exonic function	CDS change	AA change	total depth	alt depth
P3	HARS2	NM_001278731	nonsynonymous SNV	c.A53G	p.Q18R	30	8
P3	SPSB3	NM_080861	nonsynonymous SNV	c.G685A	p.G229S	167	81
P6	BAZ2B	NM_001289975	nonsynonymous SNV	c.A1142G	p.N381S	41	20
P6	TRIO	NM_007118	nonsynonymous SNV	c.G8613A	p.M2871I	98	21
P6	SNX13	NM_015132	nonsynonymous SNV	c.A2101G	p.R701G	90	29
P6	PLXNA4	NM_020911	nonsynonymous SNV	c.C3176T	p.A1059V	20	10
P6	PARP12	NM_022750	nonsynonymous SNV	c.G1969A	p.D657N	71	20

In CNV analysis (Figure 3), patient 3 (53-year-old female; LG and PF) showed gains in chromosomes 1, 4, 5, 7, 8, 9, 12, 14, 15, 18, and 19, and showed no chromosome deletions. Patient 5 (25-year-old female; HG and PF) showed gains in chromosomes 12, 19, and 20 and losses in chromosomes 3, 6, 13, 16, 17, 21, and 22. Patient 6 (41-year-old male; HG and PF) showed an increased copy number in chromosome 1q and losses in chromosomes 3, 5, 17, 21, and 22. Patient 9 (6-year-old female; LG and ST) showed copy loss in chromosome 11, and patient 12 (3-year-old female; HG and PF) showed copy gain in chromosome 19.

In summary, three of four PF area cases showed a copy gain in chromosome 19, which is a rare finding. Moreover, children showed fewer CNVs than adults, as expected.

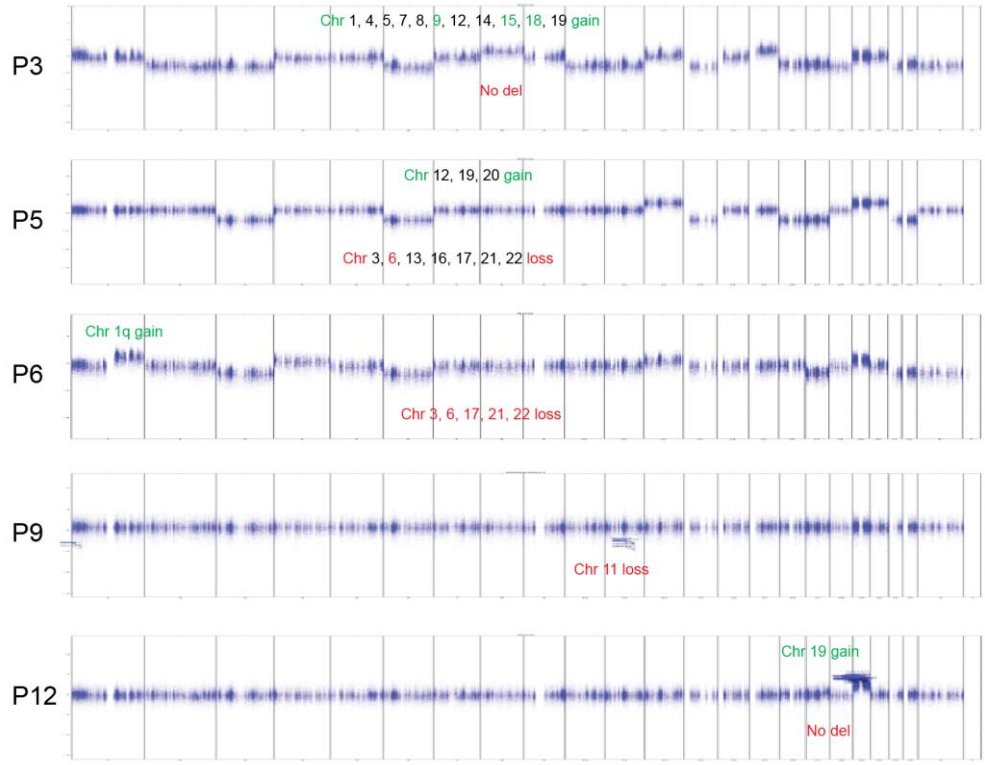


Figure 3. Copy number variations of five patients (adults: P3, P5, P6; children: P9, P12) using whole-exome sequencing.

CCP

To identify the mutation status of cancer-related genes, we used the Ion AmpliSeq CCP. Among the 409 oncogenes included in the panel, 249 SNVs and indels in 156 genes were identified (data not shown). Three SNVs (in *FGFR3*, *KIT*, and *TAF1L*) and one indel (in *KMT2C*) were identified in all seven patients (Table 11). We excluded two indels that occurred in a repeat area, which is a common error of the Ion Ampliseq platform.

Mutations in the *CASC5* gene were identified in both the WES and CCP analyses, although the positions of these mutations were different. Despite filtering common variants according to the population MAF database, probable germline variants were difficult to filter out without corresponding information on paired normal samples.

Table 11. List of mutations identified in all seven patients using the Ion AmpliSeq Comprehensive Cancer Panel.

Gene symbol	Transcript ID	Exonic function	CDS change	AA change
FGFR3	NM_022965	nonsynonymous SNV	c.G1708A	p.V570I
KIT	NM_000222	nonsynonymous SNV	c.T789G	p.H263Q
TAF1L	NM_153809	nonsynonymous SNV	c.G3004C	p.A1002P
KMT2C	NM_170606	stopgain	c.2447dupA	p.Y816_I817delinsX

RNA sequencing

The RNA sequencing results for five patients revealed the presence of multiple fusion genes (Table 12), including a *YAP1* fusion in patient 9 (child; ST and LG), which is well-recognized in cases of pediatric ST ependymoma and is associated with a outcome. Other fusion genes identified in three patients were novel fusion genes.

Table 12. Fusion genes identified using RNA sequencing.

Patient ID	Fusion genes
P3	PDE4DIP - AL592284.1
	PTMS - PCBD2
P5	HNRNPA2B1 - RGPD2
	ZFP36L1 - RGPD2
	EEF1G - PCBD2
	MT-ND5 - MT-ND2
P7	RAB6C - OR4Q3
	GFAP - PCBD2
	IRAK2 - NDFIP1
	GFAP - PCBD2
P9	YAP1 - MAMLD1

We compared the results according to groups with respect to severity (LG, n = 2 vs. HG, n = 3), age/location (adult or PF area, n = 4 vs. children or ST, n = 1), and sex (female, n = 4 vs. male, n = 1). Genes exhibiting differences in expression between groups in each comparison were detected by comparing the mean of the original value for each group against the mean of the corresponding Z-score. Genes exhibiting the largest differences in expression were analyzed and validated according to a literature search for their relevance to ependymoma and inclusion in the human disease database (MalaCards;http://malacards.org/card/ependymoma?limit[MaladiesGenes]=128#related_genes) (Table 13).

The difference in the original values associated with the RNA sequencing results indicated that genes exhibiting enhanced levels of expression might be preferentially ranked. In this case, genes such as *GFAP* and *S100B* exhibited high levels of expr

ession related to the presence of ependymoma. Therefore, they were unable to be used as genes defining the presence of group differences. Because differences in Z-scores reveal relative differences in expression, an originally small value shows no significance. Therefore, the possibility of finding genes that can distinguish groups is high when using intersecting genes (Table 13).

The differences in expression levels were largest between females and males in all group comparisons based on correlations. The *t*-test showed that 595 genes exhibited significant differential expression between the LG and HG groups ($p < 0.05$). Of these genes, four (*EZR*, *PDPN*, *SHC3*, and *PRG2*) are directly related to ependymoma according to MalaCard analysis.

Table 13. Genes displaying significant differences in expression according to different grades, age/location, and sex for both the original genetic value and Z-score value.

Group (N)	Gene	Gene in Malacards
Adult/PF (4) vs. Children/ST (1)		
Adult/PF < Children/ST	WLS, ARL4D, DUSP1, CYR61, TIMP3, FAM84B	TIMP3
LG (2) vs. HG (3)		
LG > HG	CETN2, HSPB8, TMEM66, ID4	
Female (4) vs. Male (1)		
Female > Male	RPL38	
Female < Male	MIR3936, METRN, ATP5B	

N number of cases in each group, PF posterior fossa, ST supratentorial, LG low grade, HG high grade

Discussion

Our data support the recent WHO guidelines describing ependymoma prognosis parameters in three aspects. First, the concept of “cell of origin” suggests that distinct groups of stem cells are specific to anatomical sites, and that different origins can explain the predominant locations of ependymomas detected in different age groups (7). Our methylation analysis and RNA sequencing results revealed differences in gene expression profiles between different age groups and anatomical locations. In addition, hierarchical clustering of 500 randomly selected CpG sites satisfying cut-offs for their association with genes exhibiting significant differential expression ($p < 0.05$ and Δmean) demonstrated prominent differences between the ST and PF groups in both children and adults. Although ependymoma histology cannot reveal clinical differences corresponding to patients with ST or PF localization, the genetic profiles implied the presence of distinct phenotypes related to tumor location.

In the epigenetics analysis using a methylation array, large differences in methylation levels between age groups were detected for many genes. In adults, several more genes were associated with hypermethylation than hypomethylation, which were largely related to neurogenesis and neuron differentiation based on GO analysis, representing a meaningful association with respect to tumor properties. Genes associated with hypomethylation in the PF group mainly included those in the *HOX* family, which Taylor et al. (7) suggested as a representative gene that maintains the cancer stem cell phenotype in spinal ependymomas.

Molecular alterations are common in ependymoma, and accurate molecular

subgrouping is necessary to predict patient outcome and determine the best therapeutic approach. In the revised fourth edition of the WHO guidelines for CNS diseases, nine molecular groups were suggested based on a single study conducted by Pajtler et al. (16). This analysis focused on the use of DNA-methylaton profiling to determine ependymona subgroups, and identified two ST subgroups characterized by the fusion genes *RELA* and *YAP1* (14). Their results also indicated molecular subgroups showing distinct copy number profiles, which agreed with our findings of CNVs based on WES analysis.

We compared our WES data with those of Korshunov et al. (9), who proposed three distinct molecular stages for intracranial ependymoma through analysis of patterns of cytogenetic alterations in 122 primary ependymoma samples from adults and children. Group 1 ependymoma is characterized by numerous aberrations affecting the whole chromosomes or chromosome arms, including gains in chromosome 9, 15q, and/or 18, and/or loss of chromosome 6 without a 1q gain and/or *CDKN2A* deletion. These alterations were observed in 34% of all patients and were associated with a 100% 5-year overall survival rate. Patients 3 and 5 in our study exhibited characteristics that would place them in this group, and these patients were disease-free for 2 years and 1 year, respectively.

Group 2 ependymoma is characterized by a largely balanced genomic profile, and was observed in 42% of patients with 5-year overall survival rates of 78% (9). In our study, patient 9 and patient 12 could be associated with this group, as patient 9 with ST/LG ependymoma showed no recurrence for 49 months. However, patient 12, a child with PF/HG ependymoma, experienced recurrence five times, despite repeated gamma knife radiosurgery and radiotherapy.

Group 3 ependymoma involves tumors exhibiting copy number gains in chromosome 1q and/or homozygous *CDKN2A* deletion, with 25% of patients classified in this group exhibiting 5-year survival rates of 32%, which constituted the worst prognosis overall. Patient 6 in our study was included in this group, and experienced no recurrence or death over a 16-month follow-up period. Overall, these results indicated that the CNV data alone could not be applied for accurate molecular classification.

In addition, copy gain of chromosome 19 was observed in three of the four patients with PF ependymoma, which is a rare finding. Trisomy 19 ependymomas were classified as ST WHO grade-III tumors in children (24), and are associated with a clear cell morphology, which was not evident in our cases. Thus, classification by CNV seems to have poor reproducibility.

The published revision to the WHO CNS disease guidelines included ependymoma and anaplastic ependymoma, but a grading system was not applied because various studies did not establish an association between tumor grade and biological behavior or survival. In our study, hirerarchical clustering of the methylation array according to grade showed a weak tendency for grouping. RNA sequencing revealed several genes with significant differential expression between LG and HG tumors. However, the difference between females and males was greatest overall among the groups compared. Nevertheless, this result should be interpreted with caution given the limitations of the sample; namely, the ST group only included one child, and all patients in the PF group were adults. Considering the fact that posterior ependymomas are most common among children, our patient cohort might not represent the major features present in ependymoma groups.

Compared to methylation microarray analysis, several molecular analysis tools fail to find sufficient evidence to support genetic features associated with a disease or subtype. Before filtering the WES data, the absolute number of somatic mutations per megabase was larger in adults, which is likely attributed to the presence of age-related mutations. After filtering, few nsSNVs were detected by WES, especially in children, compared to reports for other malignancies, which is in line with previous findings of PF ependymoma in children (25). In addition, few SNVs and indel mutations were identified in the CCP analysis, and no mutation was common between the WES dataset and CCP.

We detected a few fusion genes by RNA sequencing, including a *YAP1* fusion in a pediatric patient who had a tumor in the ST area and showed a good prognosis. This result is in line with Pajtler's molecular classification. Other novel fusion genes identified are not listed in the COSMIC database, possibly indicating that they are not functional or actionable. Another well-known fusion gene other than *YAP1* is *RELA*, which is identified in about 70% of childhood ST ependymoma cases (26). The C11orf95-*RELA* fusion is the most common structural variant found in ependymomas (16). In our RNA fusion study, only one available sample was from the ST and a *YAP1* fusion was detected in that sample. However, we did not find a *RELA* fusion in the five ST samples, which is likely due to the small sample size.

Among the genes displaying a significant difference between the adult/PF and children/ST groups with respect to both the original genetic value and Z-scores in RNA sequencing, only *TIMP3* was listed in the human disease database Malacards. *TIMP3*, which plays a role in CNS development and the proteolysis pathway, was

identified as one of the candidate methylated genes that is re-expressed in short-term cell cultures of ependymoma samples from pediatric patients following treatment with the demethylating agent 5-Aza-2′-deoxycytidine (27). This suggests a high possibility that methylation of *TIMP3* plays a major role in ependymoma. However, *TIMP3* was not observed to be differentially methylated in our analysis according to age and location.

Drugs that target CpG methylation or histone demethylase inhibitors are considered as potential therapeutic agents for treating pediatric PFA ependymomas (28). Based on our results, targeting certain genes of the *HOX* family (*HOXA7, HOXA9, HOXB6, HOXC4, HOXC5, HOXC6,* and *HOXC8*) in adults and targeting others (*HOXC4, HOXD4*) in ST tumors may be a useful treatment strategy. A few *HOX* family genes have been implicated in spinal cord ependymomas (*HOXA9* (7), *HOXB5* (29)) and were shown to be highly expressed in spinal tumors based on gene expression data. Some *HOX* family members, especially *HOXA9*, are selectively expressed in spinal ependymoma (30). However, in our study, *HOXA9* was hypermethylated in the adult group of intracranial ependymoma. Koch et al. (23) reported *HOXA9* as an age-related gene in different tissues such as the dermis or cervical smear cells. Further study may be necessary to discover the function of *HOXA9* in intracranial ependymoma in association with aging.

FHIT was among the genes showing large differences in methylation patterns between the ST and PF groups, which was ST-hypermethylated and PF-hypermethylated in the promoter area. *FHIT* is a tumor suppressor gene involved in cell-cycle regulation, and is one of the most commonly altered genes in human cancer, showing inactivation in 60% of human tumors (31). Although a previous

study showed that *FHIT* was methylated in 22% of ependymoma samples, the exact molecular mechanism or functional pathway is still unknown (32).

Furthermore, among the genes showing a methylation rate difference > 0.4, three *(PDE4C, HOXA9,* and *TBX5*) were also identified as age-related genes in different tissues by Koch et al. (23) in an epigenetic study. These genes were included in the adult-hypermethylated and child-hypomethylated group. Thus, using the absolute β-value is preferred over using the difference in methylation rate to determine target genes in each age or location group because of the potential interference of age-related methylation. We used a β-value of 0.7 and 0.3 as the hypermethylation and hypomethylation cut-off, respectively, because we could not obtain significant data using the cutoff points of 0.8 or 0.2 proposed by Rogers et al. (2).

LHX2 was more highly methylated in the PF than in the ST area in our data and also showed decreased expression in the PF area in the Agilent microarray data of the DKFZ (GSE27287). In a previous study, *LHX2* was found to be markedly overexpressed in ST ependymoma compared with infratentorial ependymoma by quantitative PCR (33). In addition, *LHX2* was identified as a differentially expressed and overexpressed gene in ST pilocytic astrocytoma (34). LHX2, LIM homeobox protein 2, is a transcription factor involved in brain development and neural stem cell differentiation, and inactivation in $LHX2^{-/-}$ mice was shown to result in impaired cortial hem formation (35). In this regard, *LHX2* might show region-specific expression in the ST area.

We also found differences in gene expression patterns according to different tumor grades using RNA sequencing. The distinction between grade II and grade III ependymoma is often difficult, with poor interobserver reproducibility, and its

molecular biology and clinical outcome are heterogeneous (36). In this study, applying the diagnostic criterion of increased cellularity could not distinguish between these grades, whereas the addition of other criteria, including mitosis and microvacular proliferation, did allow for discriminating between grade II and III ependymoma. Although the revised WHO classification states that grading lacks a direct association with tumor biology (36), a few large cohort studies support the value of grading (12, 37). Palm et al. (8) also found that both location and histological grade exhibited differential gene expression patterns in RNA transcripts, and WHO grade II ependymomas differed from anaplastic ependymoma according to the expression of genes implicated in several signaling pathways, including the Wnt/β-catenin pathway.

Our experimental design may have some limitations related to the shortage of tumor samples because of the overall rarity of CNS tumors. In addition, there was no germline control used for RNA sequencing and CCP because of the shortage of samples. However, this limitation was overcome to a certain degree by comparing the results from three independent types of molecular analysis tools.

In summary, we analyzed ependymomas from patients according to different tumor grades and locations, as well as patient age, using next-generation sequencing methods, including methylation array, WES, CCP, and RNA-sequencing analyses. Although the number of specimens was limited, the combinations of the various conditions established sufficient diversity within the cohort to obtain reliable data. This constitutes the first report of a next-generation sequencing analysis of ependymomas in Korean patients, and thus our results are expected to provide valuable insight into the genetic and epigenetic profiles of intracranial

ependymoma.

References

1. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK. WHO Classification of Tumours of the Central Nervous System. International Agency for Research on Cancer: Lyon, 2007. 2007(4th Edition):74-80.
2. Rogers HA, Kilday JP, Mayne C, Ward J, Adamowicz-Brice M, Schwalbe EC, et al. Supratentorial and spinal pediatric ependymomas display a hyper methylated phenotype which includes the loss of tumor suppressor genes involved in the control of cell growth and death. Acta neuropathologica. 2012; 123(5):711-25.
3. Kudo H, Oi S, Tamaki N, Nishida Y, Matsumoto S. Ependymoma diagnosed in the first year of life in Japan in collaboration with the International Society for Pediatric Neurosurgery. Child's nervous system : ChNS : official journal of the International Society for Pediatric Neurosurgery. 1990;6(7):375-8.
4. Schiffer D, Giordana MT. Prognosis of ependymoma. Child's nervous system : ChNS : official journal of the International Society for Pediatric Neurosurgery. 1998;14(8):357-61.
5. Horn B, Heideman R, Geyer R, Pollack I, Packer R, Goldwein J, et al. A multi-institutional retrospective study of intracranial ependymoma in children: identification of risk factors. Journal of pediatric hematology/oncology. 1999;21(3):203-11.
6. Rubio MP, Correa KM, Ramesh V, MacCollin MM, Jacoby LB, von Deimling A, et al. Analysis of the neurofibromatosis 2 gene in human ependymomas and astrocytomas. Cancer research. 1994;54(1):45-7.
7. Taylor MD, Poppleton H, Fuller C, Su X, Liu Y, Jensen P, et al. Radial glia cells are candidate stem cells of ependymoma. Cancer Cell. 2005;8(4):323-35.
8. Palm T, Figarella-Branger D, Chapon F, Lacroix C, Gray F, Scaravilli F, et al. Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis. Cancer. 2009;115(17):3955-68.
9. Korshunov A, Witt H, Hielscher T, Benner A, Remke M, Ryzhova M, et al. Molecular staging of intracranial ependymoma in children and adults. Journal of clinical oncology : official journal of the American Society of Clinic

al Oncology. 2010;28(19):3182-90.
10. Witt H, Mack SC, Ryzhova M, Bender S, Sill M, Isserlin R, et al. Deli neation of two clinically and molecularly distinct subgroups of posterior foss a ependymoma. Cancer Cell. 2011;20(2):143-57.
11. McLendon RE, Wiestler OD, Kros JM, Korshunov A, Ng HK. WHO Cla ssification of Tumours of the Central Nervous System. Lyon. 2007;IARC:74-8.
12. Ellison DW, Kocak M, Figarella-Branger D, Felice G, Catherine G, Piet sch T, et al. Histopathological grading of pediatric ependymoma: reproducibili ty and clinical relevance in European trial cohorts. Journal of negative results in biomedicine. 2011;10:7.
13. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Compar ison of Beta-value and M-value methods for quantifying methylation levels b y microarray analysis. BMC bioinformatics. 2010;11:587.
14. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene li st enrichment analysis and candidate gene prioritization. Nucleic acids resear ch. 2009;37(Web Server issue):W305-11.
15. SureSelect - How it Works. Agilent Techonologies http://wwwgenomi csagilentcom/articlejsp?pageId=3083.
16. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, Kratochwil F, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell. 2015;27(5):728-43.
17. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, e t al. A framework for variation discovery and genotyping using next-generati on DNA sequencing data. Nat Genet. 2011;43(5):491-8.
18. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, L evy-Moonshine A, et al. From FastQ data to high confidence variant calls: th e Genome Analysis Toolkit best practices pipeline. Current protocols in bioinf ormatics. 2013;43:11.0.1-33.
19. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cance r by exome sequencing. Genome research. 2012;22(3):568-76.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of g enetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.
21. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junction

s with RNA-Seq. Bioinformatics (Oxford, England). 2009;25(9):1105-11.
22. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannot ated transcripts and isoform switching during cell differentiation. Nature biot echnology. 2010;28(5):511-5.
23. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging. 2011;3(10):1018-27.
24. Rousseau E, Palm T, Scaravilli F, Ruchoux MM, Figarella-Branger D, S almon I, et al. Trisomy 19 ependymoma, a newly recognized genetico-histolo gical association, including clear cell ependymoma. Molecular cancer. 2007;6:4 7.
25. Mack SC, Witt H, Piro RM, Gu L, Zuyderduyn S, Stutz AM, et al. Epi genomic alterations define lethal CIMP-positive ependymomas of infancy. Nat ure. 2014;506(7489):445-50.
26. Parker M, Mohankumar KM, Punchihewa C, Weinlich R, Dalton JD, Li Y, et al. C11orf95-RELA fusions drive oncogenic NF-kappaB signalling in epe ndymoma. Nature. 2014;506(7489):451-5.
27. Karakoula K, Jacques TS, Phipps KP, Harkness W, Thompson D, Hardi ng BN, et al. Epigenetic genome-wide analysis identifies BEX1 as a candidate tumour suppressor gene in paediatric intracranial ependymoma. Cancer lette rs. 2014;346(1):34-44.
28. Wu J, Armstrong TS, Gilbert MR. Biology and management of epend ymomas. Neuro-oncology. 2016;18(7):902-13.
29. Korshunov A, Neben K, Wrobel G, Tews B, Benner A, Hahn M, et al. Gene expression patterns in ependymomas correlate with tumor location, gr ade, and patient age. The American journal of pathology. 2003;163(5):1721-7.
30. Gu S, Gu W, Shou J, Xiong J, Liu X, Sun B, et al. The Molecular Fea ture of HOX Gene Family in the Intramedullary Spinal Tumors. Spine. 2017;42 (5):291-7.
31. Pekarsky Y, Zanesi N, Palamarchuk A, Huebner K, Croce CM. FHIT: fr om gene discovery to cancer treatment and prevention. Lancet Oncol. 2002;3 (12):748-54.
32. Michalowski MB, de Fraipont F, Michelland S, Entz-Werle N, Grill J, P asquier B, et al. Methylation of RASSF1A and TRAIL pathway-related genes is frequent in childhood intracranial ependymomas and benign choroid plexus

papilloma. Cancer genetics and cytogenetics. 2006;166(1):74-81.
33. Andreiuolo F, Puget S, Peyre M, Dantas-Barbosa C, Boddaert N, Philippe C, et al. Neuronal differentiation distinguishes supratentorial and infratentorial childhood ependymomas. Neuro-oncology. 2010;12(11):1126-34.
34. Sharma MK, Mansur DB, Reifenberger G, Perry A, Leonard JR, Aldape KD, et al. Distinct genetic signatures among pilocytic astrocytomas relate to their brain region origin. Cancer research. 2007;67(3):890-900.
35. Bulchand S, Grove EA, Porter FD, Tole S. LIM-homeodomain gene Lhx2 regulates the formation of the cortical hem. Mechanisms of development. 2001;100(2):165-75.
36. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Ellison DW, Figarella-Branger D, et al. WHO Classification of Tumours of the Central Nervous System. International Agency for Research on Cancer: Lyon, 2016. 2016;Revised 4th Edition:101-14.
37. Godfraind C. Classification and controversies in pathology of ependymomas. Child's nervous system : ChNS : official journal of the International Society for Pediatric Neurosurgery. 2009;25(10):1185-93.

국문초록

위치와 나이를 고려한 뇌실막종의 메틸화 및 분자적 분석

서울대학교 대학원
의학과 병리학 전공
조 화 진

뇌실막종은 중추신경에 발생하는 종양으로, 소아에서는 주로 두개내에, 성인에서는 척추와 두개내에 발생한다. 세계보건기구 지침에 따르면, 뇌실막종의 예후에는 환자의 나이, 종양의 절제 범위, 종양의 위치 및 조직병리학적 등급이 고려인자로 작용한다고 하였다. 하지만 최근 발표된 연구에서는 종양의 위치와 환자의 나이에 중점을 두어 유전자전사적 (transcriptomic), 유전적, 그리고 후생유전학적 분석을 시행하여 WHO 등급보다 더 정확한 임상적 예후를 예측할 수 있게 되었다. 본 연구에서는 성인환자 9 례, 소아환자 4 례로 구성된 한국인 환자의 원발성 두개내 뇌실막종을 대상으로 Infinium HumanMethylation450 BeadChip array 를 이용하여 종양의 위치와 환자의 나이에 따른 메틸화 정도와 관련 유전자를 밝히려 하였다. 신경발생과 신경 분화에 관련된 유전자들이 성인에서 과메틸화 표현형을

보였고, 후뇌 발달 경로의 HOX 유전자 활성화 유전자들이 천막상 그룹에서 과메틸화 표현형을 보였다. 또한 전장엑솜시퀀싱 (whole exome sequencing), 포괄적 암패널과 RNA 시퀀싱 분석을 하여 체세포 돌연변이, 복제수변이 (copy number variation) 및 융합유전자들을 찾았다. 이를 통해 소수의 체세포 돌연변이들과 새로운 융합유전자들을 찾아내었고, 이전 연구에서와 같이 예후가 좋은 한 소아 환자에서 YAP1 융합을 관찰하였다. 이러한 포괄적인 분석들을 이용하여 종양의 위치와 환자 나이에 따라 유의하게 차이를 보이는 종양특이 유전자 발현을 확인할 수 있었다. 따라서 본 연구를 통하여 두개내 뇌실막종의 후생학적 및 유전학적 특성에 대한 값진 통찰을 제공하였고, 위치와 나이에 입각한 정확한 뇌실막종 관련 예후 인자를 찾아 잠재적 치료전략을 모색하는데 큰 역할을 할 수 있을 것으로 기대된다.

주요어 : 뇌실막종, DNA 메틸화, 분자분석, 분류

학 번 : 2011-21905