Supplementary

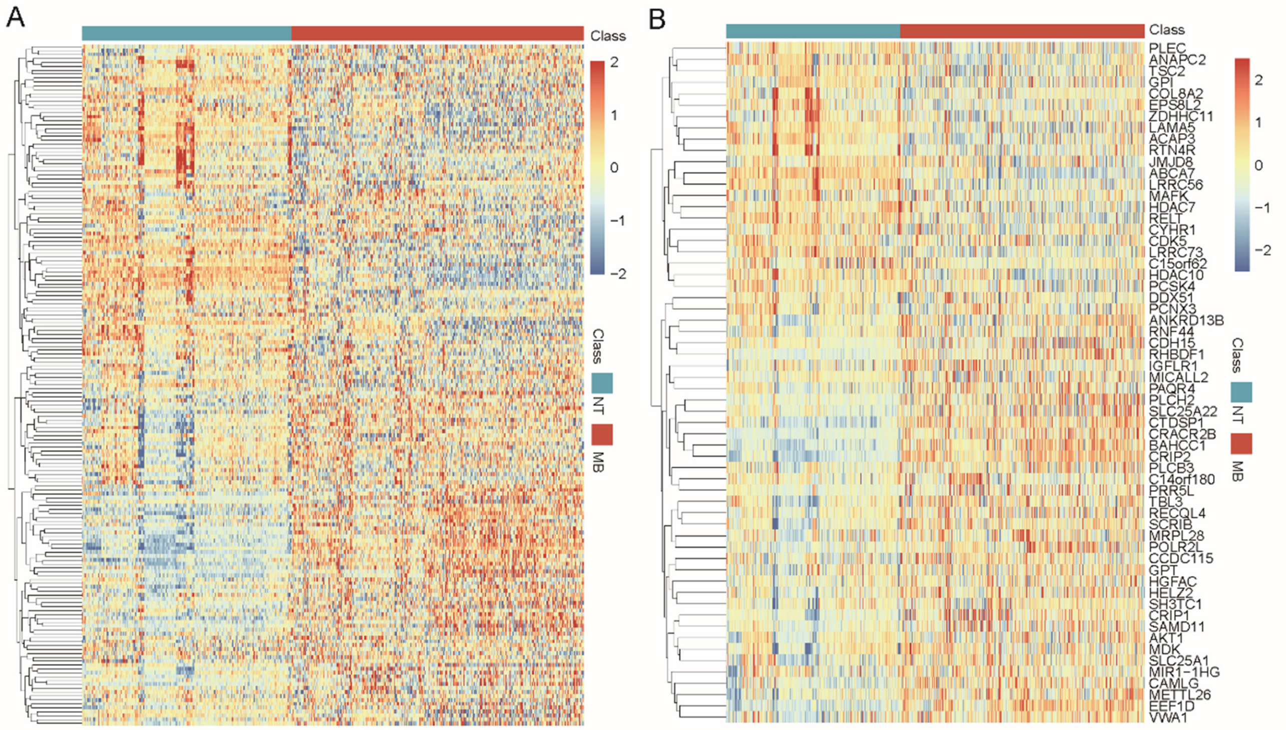
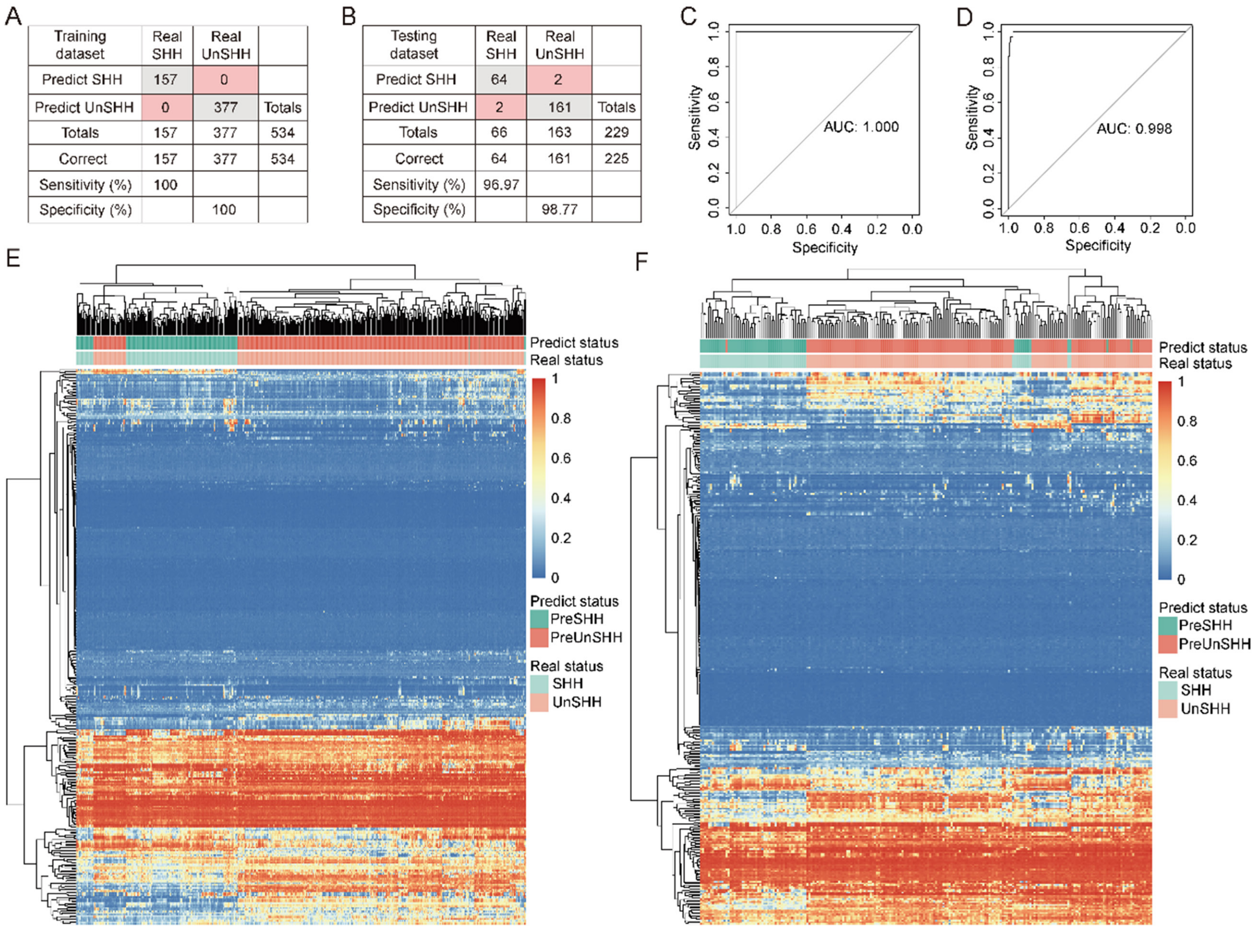


Figure S1 Expression levels of differentially distributed cfDNA region-related genes derived from gene express array data. (A) Gene expression level of significant differential cfDNA distribution regions defined by gene expression arrays. Totally 175 of 262 genes were contained in gene expression array datasets. (B) Gene expression levels inferred as regulated by promoter regions with differential cfDNA read counts are presented. MB: medulloblastoma patient sample. NT: non-tumor, hydrocephalus patient sample. cfDNA, cell free DNA.

https://dx.doi.org/10.21037/tcr-2025-aw-2286



A

Training dataset	Real SHH	Real UnSHH	
Predict SHH	157	0	
Predict UnSHH	0	377	Totals
Totals	157	377	534
Correct	157	377	534
Sensitivity (%)	100		
Specificity (%)		100	

B

Testing dataset	Real SHH	Real UnSHH	
Predict SHH	64	2	
Predict UnSHH	2	161	Totals
Totals	66	163	229
Correct	64	161	225
Sensitivity (%)	96.97		
Specificity (%)		98.77	

Figure S2 Classification of SHH MB based on the methylation state derived from 450K methylation array. (A,B) Confusion tables of classify results of the diagnostic perdition model in the training (A) and testing (B) data sets. Totally 763 samples were divided into 2 groups, SHH MB samples were defined as SHH; Group 3, Group 4 and WNT MB samples were defined as UnSHH. (C,D) ROC of the diagnostic prediction model based on methylation state in the training (C) and testing (D) data sets. (E,F) Unsupervised hierarchical clustering of markers used in the diagnostic prediction model in the training (E) and testing (F) data sets. AUC, area under the curve; ROC, receiver operating characteristic.

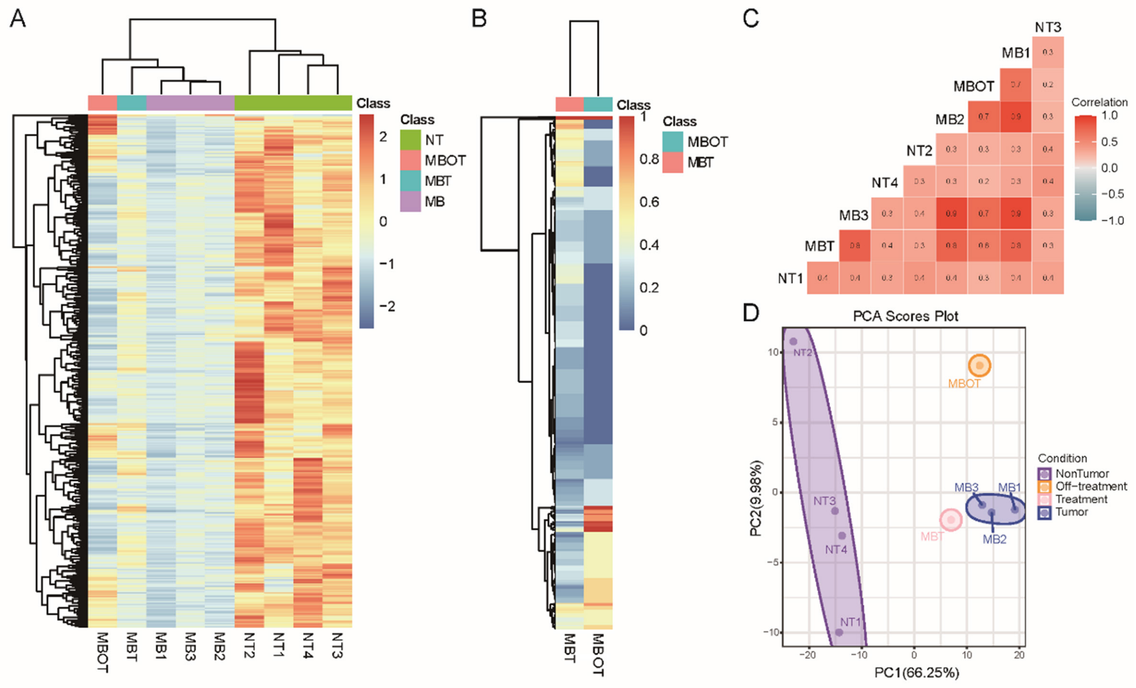


Figure S3 cfDNA distribution feature of a patient with MB through the treatment process. (A) cfDNA read distributions of a patient with MB on (MBT) and off treatment (MBOT). cfDNA read counts were calculated in significantly differential regions of patients with MB and hydrocephalus, and a patient with MB on and off treatment. Clustering was performed based on read counts. (B) Comparison of cfDNA reads distribution of one MB patients in different treatment states. (C) Similarities among between cfDNA samples from multiple sources. (D) Classification of cfDNA samples based on differentially distributed genome regions. cfDNA, cell free DNA; PC, principal components.

https://dx.doi.org/10.21037/tcr-2025-aw-2286

Table S1 The information of datasets used in this study collected from the GEO database

Sample information	Description	Sample type	Sampling method	Wet-lab strategy	Platform	SRA No.	Data type	Accession No.
MB1	Medulloblastoma patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina Nextseq 500	SRR10725743	Fastq	GSE142241
MB2	Medulloblastoma patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina Nextseq 500	SRR10725745	Fastq	
MB3	Medulloblastoma patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina Nextseq 500	SRR10725747	Fastq	
NT1	Hydrocephalus patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina NovaSeq 6000	SRR10725749	Fastq	
NT2	Hydrocephalus patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina NovaSeq 6000	SRR10725750	Fastq	
NT3	Hydrocephalus patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina NovaSeq 6000	SRR10725751	Fastq	
NT4	Hydrocephalus patient cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina NovaSeq 6000	SRR10725752	Fastq	
Treatment	Medulloblastoma patient with treatment cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina Nextseq 500	SRR10725753	Fastq	
Off treatment	Medulloblastoma patient after treatment cfDNA WGBS data	Cerebrospnial fluid	Lumbar puncture	Bisulfite-Seq According with the manufacture of Pico Methyl-Seq Library Prep Kit (Zymo Research).	Illumina Nextseq 500	SRR10725754	Fastq	
ATAC-seq	Medulloblastoma tissues ATAC-seq data	Tissue	Surgical resection	ATAC-seq Cells were incubated in lysis buffer, centrifuged, and resuspended in Transposase reaction mix. Tagmentated DNA were purified and amplified to add sequencing primers to generate sequencing libraries.	Illumina HiSeq 2000	SRR25635612 SRR25635613 SRR25635614 SRR25635615 SRR25635620 SRR25635621 SRR25635627 SRR25635629 SRR25635630 SRR25635631	Fastq	GSE240985
RNA-seq	Medulloblastoma or para-tumor tissues RNA-seq data	Tissue	Surgical resection	RNA-seq Medulloblastoma tumor or para-tumor tissues were flash frozen on dry ice, and RNA was harvested using Trizol reagent. Illumina TruSeq RNA Sample Prep Kit (Cat#FC-122-1001) was used with 1 μg of total RNA for the construction of sequencing libraries.	Illumina HiSeq 2000	–	Reads count matrix	GSE164677
Gene expression profile	Medulloblastoma and normal brain gene expression level were qualified using gene expression array (expression matrix file)	Tissue	Surgical resection	As directed by affimetrix.	Affimetrix array	–	Gene expression matrix	GSE124814
DNA methylation profile	Medulloblastoma and normal brain DNA methylation states were qualified using 450K arrays	Tissue	Surgical resection	Illumina HumanMethylation450 BeadChip	Illumina HumanMethylation450 BeadChip	–	IDAT files	GSE85218

GEO, gene expression omnibus.

Table S2 The statistic of mappable reads counts and genome coverages

Sample	Mappable reads count	Average reads length (bp)	Genome coverage*	10^6 level reads count	10^6 reads genome coverage
MB1	330,178,552	83	8.7×	991,910	0.03×
MB2	50,110,085	82	1.3×	1,002,712	0.03×
MB3	60,818,825	79	1.5×	972,974	0.02×
NT1	8,488,362	300	0.8×	1,000,000	0.1×
NT2	7,509,358	300	0.7×	1,000,000	0.1×
NT3	28,305,339	300	2.7×	1,000,000	0.1×
NT4	10,801,670	300	1.0×	1,000,000	0.1×

*, the genome coverage was calculated following the formula: genome coverage = mappable reads count × average reads length/ reference genome length (hg19).

Table S3 The comparison between our study and other similar study

Items	Present study	Peneder's study
Cancer type	Medulloblastoma (brain tumor)	Ewing sarcoma (bone neoplasms)
Origin of cfDNA	Cerebrospinal fluid	Blood
Analysis method	Shallow whole genome sequencing	Shallow whole genome sequencing
Genome bin length	1 kb	100 kb
Split of cfDNA fragments	No	Shor: 100–150 bp; long: 151–200 bp
Classify indicator	Reads count in each bin	Short/long ratio
Linked biologic features	Chromatin accessibility Gene expression DNA methylation Super-enhancer DNase high sensitivity H3K27ac state	Chromatin accessibility Chromosomal translocations Copy number alternations DNA methylation Super-enhancer DNase high sensitivity

cfDNA, cell free DNA.

https://dx.doi.org/10.21037/tcr-2025-aw-2286