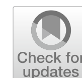
Acta Neurochirurgica (2024) 166:72
https://doi.org/10.1007/s00701-024-05922-5

ORIGINAL ARTICLE



Development and validation of a 23-gene expression signature for molecular subtyping of medulloblastoma in a long-term Chinese cohort

Yuyuan Wang[1,2,3,4] · Jianhan Huang[1,2,3,4] · Xian Yin[1] · Qinghua Xu[5,6] · Yifeng Sun[5] · Yu Yao[1,2,3,4] · Ji Xiong[7]

Received: 3 September 2023 / Accepted: 16 November 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2024

Abstract

Purpose Medulloblastoma is the most common childhood malignant brain tumor and is a leading cause of cancer-related death in children. Recent transcriptional studies have shown that medulloblastomas comprise at least four molecular subgroups, each with distinct demographics, genetics, and clinical outcomes. Medulloblastoma subtyping has become critical for subgroup-specific therapies. The use of gene expression assays to determine the molecular subgroup of clinical specimens is a long-awaited application of molecular biology for this pediatric cancer.

Methods In the current study, we established a medulloblastoma transcriptome database of 460 samples retrieved from three published datasets (GSE21140, GSE37382, and GSE37418). With this database, we identified a 23-gene signature that is significantly associated with the medulloblastoma subgroups and achieved a classification accuracy of 95.2%.

Results The 23-gene signature was further validated in a long-term cohort of 142 Chinese medulloblastoma patients. The 23-gene signature classified 21 patients as WNT (15%), 41 as SHH (29%), 16 as Group 3 (11%), and 64 as Group 4 (45%). For patients of WNT, SHH, Group 3, and Group 4, 5-year overall-survival rate reached 80%, 62%, 27%, and 47%, respectively ($p<0.0001$), meanwhile 5-year progression-free survival reached 80%, 52%, 27%, and 45%, respectively ($p<0.0001$). Besides, SHH/TP53-mutant tumors were associated with worse prognosis compared with SHH/TP53 wild-type tumors and other subgroups. We demonstrated that subgroup assignments by the 23-gene signature and Northcott's NanoString assay were highly comparable with a concordance rate of 96.4%.

Conclusions In conclusion, we present a novel gene signature that is capable of accurately and reliably assigning FFPE medulloblastoma samples to their molecular subgroup, which may serve as an auxiliary tool for medulloblastoma subtyping in the clinic. Future incorporation of this gene signature into prospective clinical trials is warranted to further evaluate its clinical.

Keywords Medulloblastoma · Molecular subgroup · Gene expression signature · Real-time qPCR assay · 23-gene signature · Transcriptome database

Yuyuan Wang and Jianhan Huang are co-first authors and have equal contribution to this study.

✉ Ji Xiong
dabenx@163.com

[1] Department of Neurosurgery, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai 200040, China

[2] Neurosurgical Institute of Fudan University, Shanghai 200040, China

[3] Shanghai Clinical Medical Center of Neurosurgery, Shanghai 200040, China

[4] Shanghai Key Laboratory of Brain Function and Restoration and Neural Regeneration, Shanghai 200040, China

[5] Canhelp Genomics Research Center, Canhelp Genomics Co., Ltd, Hangzhou 31100, Zhejiang Province, China

[6] Institute of Machine Learning and Systems Biology, College of Electronics and Information Engineering, Tongji University, Shanghai 200092, China

[7] Department of Pathology, Huashan Hospital, Fudan University, No. 12 Wulumuqi Zhong Road, Shanghai 200040, China

Introduction

Medulloblastoma (MB) is the most common malignant pediatric brain tumor, accounting for ~ 12% of childhood cancer deaths worldwide [22]. Approximately 80% of medulloblastomas occur in children under the age of 15. In adults, medulloblastomas are rare (1 ~ 2%). The standard therapies include surgery followed by radiotherapy and chemotherapy, based on clinical factors and histopathological subgroups. With an overall 5-year survival rate of 50–80% [2], survivors suffer severe long-term side effects including neurocognitive impairment, physical/endocrine deficiencies caused by indiscriminate postoperative chemoradiotherapy, and they are at risk for recurrence and secondary malignancies [8].

The WHO 2021 Classification of Tumors of the Central Nervous System has classified medulloblastoma into four molecular subgroups: WNT, SHH-P53 wildtype, SHH-P53 mutant, and non-WNT/non-SHH (including Group 3 and Group 4), with disparate demographics, clinical characteristics, genetic features [13, 21]. The WNT and SHH subgroup are characterized by mutations leading to constitutive activation of the Wingless and Sonic Hedgehog Signaling pathways, respectively. By contrast, the genetics and biology underlying the Group 3 and Group 4 remain less clear. The recognition of consensus subgroups has deeply changed our understanding of how medulloblastoma should be diagnosed and treated in the clinic [13]. Subgroup-driven clinical trials are currently underway to evaluate the efficacy of SHH pathway inhibitors such as Vismodegib and Sonidegib at diagnosis or in recurrent or refractory SHH subgroup tumors [12].

Although the retrospective classification of various medulloblastoma cohorts into molecular subgroups has been scientifically insightful, medulloblastoma subtyping has not yet been applied in the clinical setting for either patient stratification or patient selection for targeted therapy. There is currently no well-accepted gold standard test for medulloblastoma subgroup assignment. Although an excellent tool for biomarkers discovery, gene expression microarray is likely inappropriate for routine clinical use due to the need for high-quality RNA extracted from flash-frozen samples, technical variability of microarrays, and relatively high cost. Recent studies have shown that DNA methylation profiling and specific gene expression platforms, such as Illumina microarray and NanoString nCounter assay, could be applied to formalin-fixed paraffin-embedded (FFPE) samples with robust performance; however, these approaches are still not widely available in the majority of clinical centers due to high cost and sophisticated workflow. In contrast, medulloblastoma subgroup assignment using immunohistochemistry (IHC) performed on FFPE cases has shown certain promise, although challenges remain due to variability in antibody batches, sample preparation methods, staining procedures, and inter-observer reliability.

In the current study, we aimed to develop an accurate, reliable, and economical gene expression assay for medulloblastoma subtyping in the routine clinical setting. We identified and developed a 23-gene signature showing highly accurate for medulloblastoma subgroups assignment. The findings were further validated in a long-term cohort of FFPE samples to explore the applicability of the gene signature in the clinic and to evaluate its utility for guiding diagnostic and prognostic options in Chinese medulloblastoma patients.

Materials and methods

Gene expression database curation

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the Helsinki Declaration. This study was approved by the Ethics Committee of Huashan Hospital, Fudan University in 2018.10 (Shanghai, China) (No. KY2018-424). Gene expression dataset of 460 medulloblastoma samples with confirmed molecular subgroups were collected from a public data repository—NCBI Gene Expression Omnibus, and curated to form a comprehensive medulloblastoma transcriptome database. Gene expression profiling of medulloblastoma samples obtained from three public data sets (GSE21140, GSE37382, and GSE37418) was conducted on different Affymetrix oligonucleotide microarray platforms, GeneChip Human Genome U133Plus 2.0 Array, Human Exon 1.0 ST Array, and Gene 1.1 ST Array. Detailed descriptions of the specimen characteristics and clinical features are retrieved from the original studies (Table 1) [15, 16, 21].

Microarray data processing and normalization

Gene expression data analysis was performed using R software and packages from the Bioconductor project [3, 9, 20]. We used the Single Channel Array Normalization (SCAN) approach from the SCAN-UPC package to process Affymetrix microarray data [18, 19]. Upon normalizing each raw CEL file, SCAN outputs probe-level expression values. We further used the custom mapping files from the BrainArray resource to summarize probe-level intensities directly to gene-level expression values [5]. Thus, probes mapping to multiple genes and other problems associated

Table 1 Clinical, pathological, and molecular features of medulloblastoma cohorts

Demographic	Training set ($n = 230$)	Testing set ($n = 230$)
Tissue type		
Frozen	231	230
Age		
≥ 18	14	14
< 18	213	215
NA	3	1
Gender		
Male	155	162
Female	74	67
NA	1	1
Histological subgroup		
Classic	170	158
Desmoplastic/nodular	22	20
Large cell/anaplastic	23	32
MB with extensive nodularity	3	4
NA	13	16
Molecular subgroup		
WNT	8	8
SHH	47	47
Group 3	45	44
Group 4	131	131

with older generations of Affymetrix probe designs were avoided. After normalization, we applied the ComBat approach to adjust for batch effects [10].

Gene expression signature identification and performance assessment

To identify a gene expression signature, we used the support vector machine-recursive feature elimination (SVM-RFE) algorithm for feature selection and classification modelling. For multi-class classification, a one-versus-all approach was used by which multiple binary classifiers were first derived for each subtype. The results are reported as the subtype classifying the test sample with the highest confidence. For each specimen, the predicted subtype was compared with the reference diagnosis, and a true positive result was indicated when the predicted subgroup matched the reference diagnosis. When the predicted subtype and reference diagnosis did not match, the specimen was considered a false positive. For each subtype on the panel, sensitivity was defined as the ratio of true positive results to the total positive samples analyzed, while specificity was defined as the ratio (1 − false positive) / (total tested − total positive).

Tumor material and patient characteristics

This study was approved by the Ethics Committee of Huashan Hospital, Fudan University (Shanghai, China) (No. KY2018-424). All patients provided written informed consent. All tumor samples were obtained from children ($6 \leq \text{age} < 18$ years) and adult ($\text{age} \geq 18$ years) patients during initial surgery before any adjuvant therapy. Histopathological assessment by a senior neuropathologists for all cases confirmed medulloblastoma, grade IV, using 2021 WHO criteria. The tumors were classified by histology as classic (CMB), desmoplastic/nodular (DNMB), anaplastic or large cell subtype (LC/A MB). A cohort of 142 FFPE samples collected from January 2005 to December 2015 were used for gene expression analysis. In this cohort, 84 patients were children, and 58 patients were adults with the sex ratio of 1.18:1 (male: female). Radiotherapy was administrated to 78 (55%) patients and 97 (68%) were treated with chemotherapy. The median follow-up time for all patients was 34.1 months with a range of 0.1 to 110.2 months from date of diagnosis to last contact or death. There had been 88 deaths by the end of the study.

Immunohistochemistry

For cases classified as SHH group by the gene expression signature, IHC evaluation of p53 was performed to determine the TP53 protein expression as previously described [24]. The *p53* antibody (SC-126, Santa Cruz Biotechnologies, USA; working dilution 1:1000) was performed on a BenchMark XT automated immunostainer (Ventana Medical Systems, Tucson, AZ, USA) with Cell Conditioning 1 heat retrieval solution (Ventana Medical Systems, Tucson, AZ, USA). The cutoff value for tumor positivity was set at 30% of tumor cell staining. Colon carcinoma was used as a positive control, and parallel slides omitting the primary antibody were run as a negative control.

Sample processing and subgroup analysis

Total RNA was isolated from FFPE samples using a FFPE Total RNA Isolation Kit (Canhelp Genomics Co., Ltd, Hangzhou, China) as described previously [26]. The concentration of total RNA was determined by NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) at 260-nm absorbance, and the RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Foster City, CA, USA). For each specimen, reverse transcription was performed on isolated total RNA using a High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems). The real-time PCR (RT-PCR) reaction was carried out in 7500 Real-Time PCR instrument (Applied Biosystems).

The PCR program consisted of an initiation step at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. In addition, NanoString nCounter Assay was used for molecular subgrouping based on 22 medulloblastoma subgroup-specific genes and three housekeeping genes (*GAPDH, ACTB and LDHA*) as previously described [17].

Statistical analysis

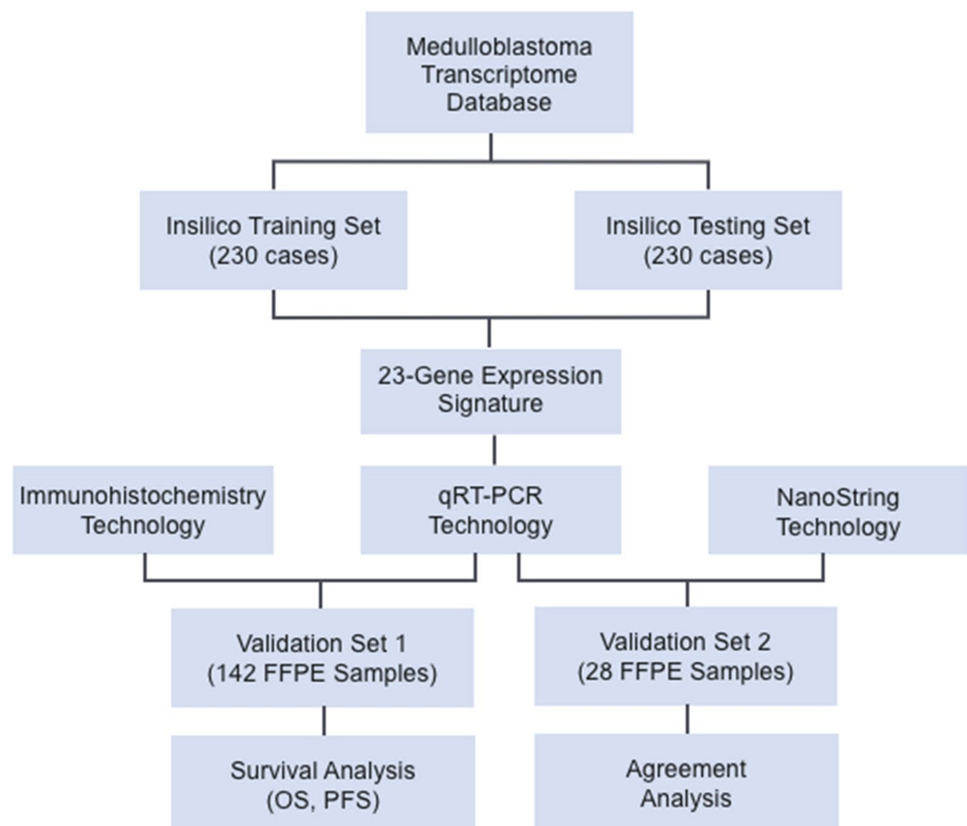
Statistical analysis was performed in the R statistical environment (v3.6.3). The distribution of survival times was determined using Kaplan–Meier estimates. The log-rank test was used to compare survival curves between groups. Overall survival (OS) for all analyses was defined as the time from diagnosis until death. Progression-free survival (PFS) was defined as the time from date of surgical resection until date of tumor progression confirmed by imaging (or by death if no progression had been confirmed previously). A p value of 0.05 was premeditated as the level of statistical significance.

Results

Establishment of medulloblastoma transcriptome database

To create a medulloblastoma transcriptome database for subgroup classification, we performed a systematic search of major biological data repositories to collect gene expression data sets of WNT, SHH, Group 3, and Group 4 tumor samples. Overall, we accumulated the gene expression profiles of 460 samples to form a comprehensive medulloblastoma transcriptome database. To identify a reliable gene expression signature, we adopted a training–testing-validation approach in this study. The medulloblastoma transcriptome database was split into a training set and a testing set. These two datasets shared equivalent distributions of four molecular subgroups. The training set was used to select candidate biomarkers and optimal classification algorithms, while the testing set was used to evaluate gene signature performance. Then, a fully specified gene signature was transferred from Affymetrix microarray to RT-PCR platform and further validated in a clinical cohort of FFPE samples. Figure 1 depicts the whole study design.

Fig. 1 Study design



Identification and validation of a 23-gene signature for medulloblastoma subgroup classification

The training set comprised 230 medulloblastoma tumors, including eight WNT, 47 SHH, 44 Group 3, and 131 Group 4 samples. After data normalization and annotation steps, a matrix of 12,263 unique genes in 230 samples (nearly 3 million data points) was prepared for downstream bioinformatics analyses. The SVM-RFE algorithm was used to select a candidate list of informative genes from high-dimension genomic data. For each subtype, we used the SVM-RFE approach to (1) evaluate and rank the contributions of each gene to the optimal separation of a specific subtype from other subtypes; (2) select the Top-10 ranked genes as the most differentially expressed for that specific subtype; (3) repeat the process for each subtype, and obtain four lists of the Top-10 gene set. After removing redundant features, a list of 23 unique genes was identified (Table 2). The 23-gene list consisted of five to seven genes for each subgroup: WNT (*DKK2*, *WIF1*, *EPHA7*, *LGR5*, *GAD*, *NKD1*, *TNC*), SHH (*PDLIM3*, *ROBO1*, *OTX2*, *TRAC*, *TBR1*), Group 3 (*GABRA5*, *IMPG2*, *IGF2BP3*, *MAB21L2*, *TTR*, *ZFPM2*), and Group 4 (*EOMES*, *KCNA*, *KHDRBS2*, *PEX5L*, *RALYL*). Functional enrichment analysis revealed that five genes (*EPHA7*, *ROBO1*, *OTX2*, *EOMES*, and *TBR1*) were over-represented in the "Cell morphogenesis involved in neuron differentiation" gene set ($p = 3.4e-3$). Besides, *DKK2*, *NKD1*, and *WIF1* were specifically over-represented in the "WNT signaling pathway" ($p = 1.6e-2$). Unsupervised hierarchical cluster analysis was further performed based on the 23-gene expression profiles of 230 samples. As shown in Fig. 2A, the samples were clustered into four groups that closely followed the molecular subtypes of medulloblastoma.

A classification model comprising 23 subgroup-specific genes was trained using the entire training set and then applied to the independent test set. The test set was composed of eight WNT, 47 SHH, 44 Group 3, and 131 Group 4 specimens. The unsupervised hierarchical clustering of 23 genes in 230 testing samples also revealed distinct patterns between WNT, SHH, Group 3, and Group 4 (Fig. 2B). With the 23-gene signature, eight samples were classified as WNT, 47 as SHH, 50 as Group 3, and 125 as Group 4. The gene expression-based assignments reached a 95.2% overall agreement with the reference diagnoses (219 of 230; 95% CI: 0.92 to 0.98). Sensitivities ranged from 94 (Group 4) to 100% (WNT, SHH), while specificities ranged from 96 (Group 3) to 100% (WNT). The 23-gene signature allowed robust discrimination of medulloblastoma subgroups in 219 of 230 test samples with an accuracy over 95%, which was

Table 2 Twenty-three candidate genes

Gene symbol	Description	Cytoband	Subtype-related	Regulation
IGF2BP3	Insulin like growth factor 2 MRNA binding protein 3	7p15.3	Group 3	UP
GABRA5	Gamma-aminobutyric acid type A receptor subunit alpha5	15q12	Group 3	UP
IMPG2	Interphotoreceptor matrix proteoglycan 2	3q12.3	Group 3	UP
MAB21L2	Mab-21 Like 2	4q31.3	Group 3	UP
TTR	Transthyretin	18q12.1	Group 3	UP
ZFPM2	Zinc finger protein, FOG family member 2	8q23.1	Group 3	UP
EOMES	Eomesodermin	3p24.1	Group 4	UP
KCNA1	Potassium voltage-gated channel subfamily A member 1	12p13.32	Group 4	UP
KHDRBS2	KH RNA binding domain containing, signal transduction associated 2	6q11.1	Group 4	UP
PEX5L	Peroxisomal biogenesis factor 5 like	3q26.33	Group 4	UP
RALYL	RALY RNA binding protein like	8q21.2	Group 4	UP
OTX2	Orthodenticle homeobox 2	14q22.3	SHH	Down
TBR1	T-Box brain transcription factor 1	2q24.2	SHH	Down
TRAC	T cell receptor alpha constant	14q11.2	SHH	Down
ROBO1	Roundabout guidance receptor 1	3p12.3	SHH	UP
PDLIM3	PDZ and LIM domain 3	4q35.1	SHH	UP
LGR5	Leucine rich repeat containing G protein-coupled receptor 5	12q21.1	WNT	UP
DKK2	Dickkopf WNT signaling pathway inhibitor 2	4q25	WNT	UP
EPHA7	EPH receptor A7	6q16.1	WNT	UP
GAD1	Glutamate decarboxylase 1	2q31.1	WNT	UP
NKD1	NKD inhibitor of WNT signaling pathway 1	16q12.1	WNT	UP
TNC	Tenascin C	9q33.1	WNT	UP
WIF1	WNT inhibitory factor 1	12q14.3	WNT	UP

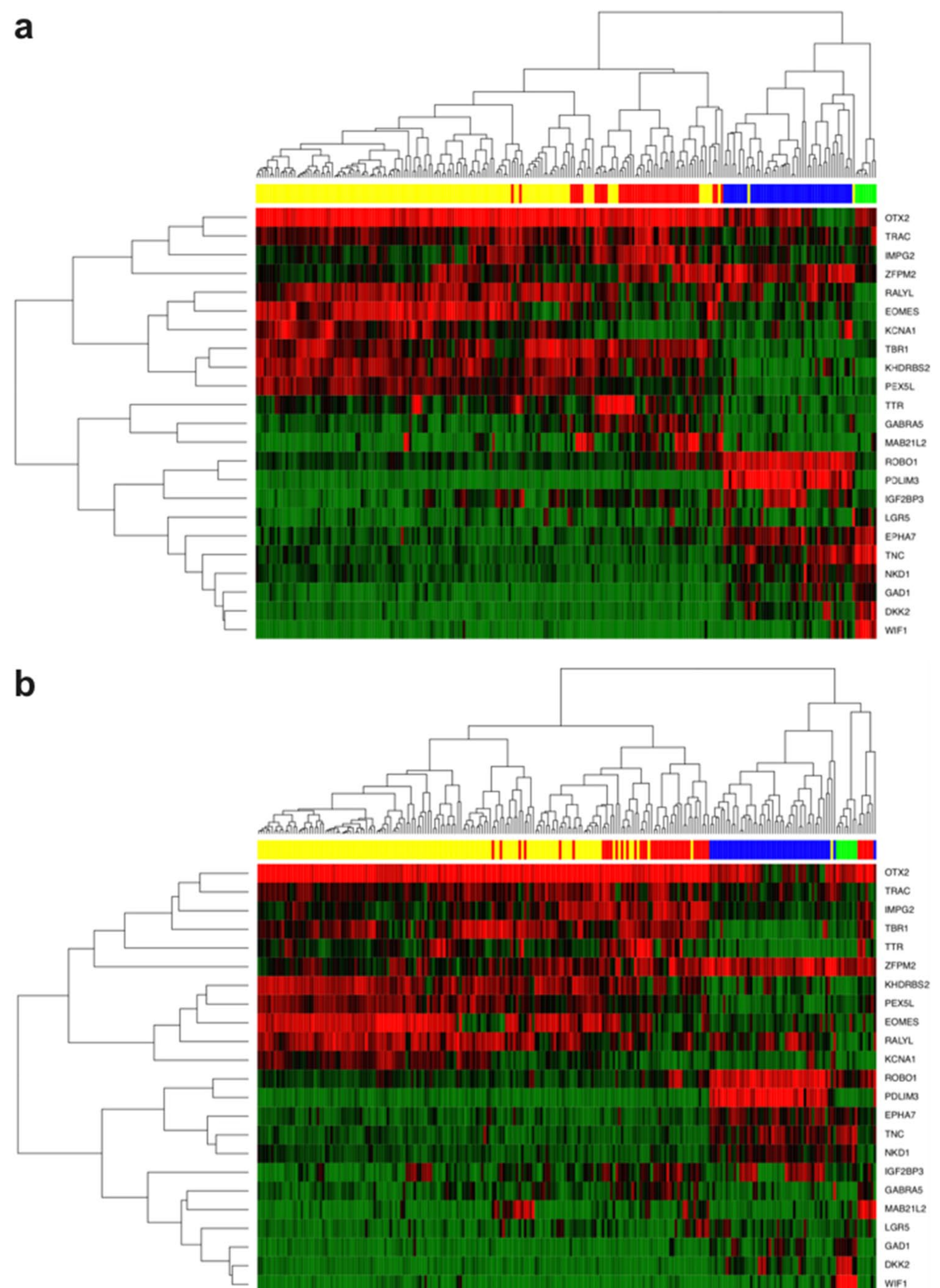


Fig. 2 Hierarchical clustering analysis of 23-gene expression data in the training set and test set. **a** Hierarchical clustering of 230 samples from the training set. Normalized gene expression intensities were shifted to mean = 0 and rescaled to STD = 1 to enhance the expression differences. The average linkage hierarchical clustering method was performed where the metric of similarity was Pearson's correlation between every pair of samples. The left panel shows a dendrogram of hierarchical clustering of 23 genes. Colored pixels capture the magnitude of the expression for any gene, where shades of red and green represent over-expression and under-expression, respectively, relative to the mean for each gene. The upper panel shows a dendrogram of hierarchical clustering of samples. The histological type of each sample is indicated in the bottom panel, with WNT shown in green, SHH shown in blue, Group 3 indicated in yellow, and Group 4 in red. The samples clustered into four groups that closely follow the histological types. **b** Hierarchical clustering of 230 samples from the testing set

comparable to the reported performance of transcriptomic and DNA methylation methods.

Clinical validation of the 23-gene signature in Chinese medulloblastoma patients

Microarray and RNA-sequencing technology provide a global assessment of transcriptomic variations; however, their sensitivity and robustness are limited in individual gene analyses, and they remain difficult to use in routine clinical practice. RT-PCR technology is generally considered the "gold standard" for measuring individual gene expression profiles and is often used to confirm the findings of microarray and RNA-sequencing analyses. More importantly, RT-PCR can be reliably adapted to clinical FFPE samples. Together with its simplicity, flexibility, and low cost, RT-PCR gene expression assay is made widely applicable in clinical practice. Hence, we further transformed the 23-gene signature from the Affymetrix microarray to the RT-PCR platform. Our RT-PCR assay successfully measured the 23-gene expression patterns in all of 142 FFPE samples that were archived for 5–15 years. With the 23-gene signature,

Table 3 Characteristics of 142 MB patients

Characteristic, %	WNT ($n=21$)	SHH ($n=41$)	Group3 ($n=16$)	Group4 ($n=64$)	*p* value
Gender					
Male	7 (33)	25 (61)	10 (63)	35 (55)	0.18
Female	14 (67)	16 (39)	6 (37)	29 (45)	
Tumor location					
Midline	9 (43)	19 (46)	12 (75)	55 (86)	**0.00**
Hemisphere	12 (57)	22 (54)	4 (25)	9 (14)	
Tumor size					
≥4 cm	5 (24)	16 (39)	11 (69)	37 (58)	**0.01**
<4 cm	16 (76)	25 (61)	5 (31)	27 (42)	
V4 floor involvement					
Yes	8 (38)	21 (51)	13 (81)	55 (86)	**0.00**
No	13 (62)	20 (49)	3 (19)	9 (14)	
Metastasis					
Yes	1 (5)	3 (7)	8 (50)	29 (45)	**0.00**
No	20 (95)	38 (93)	8 (50)	35 (55)	
Surgical resection					
GTR	19 (90)	33 (80)	10 (63)	37 (58)	**0.01**
STR	2 (10)	8 (20)	6 (37)	27 (42)	
Histology					
CMB	10 (48)	9 (22)	10 (63)	55 (86)	**0.00**
DNMB	6 (28)	18 (44)	2 (12)	1 (2)	
LC/A MB	5 (24)	14 (34)	4 (25)	8 (12)	
Chemotherapy					
Yes	10 (48)	35 (85)	12 (75)	40 (63)	**0.01**
No	11 (52)	6 (15)	4 (25)	24 (37)	
Radiotherapy					
Yes	12 (57)	21 (51)	11 (69)	34 (53)	0.66
No	9 (43)	20 (49)	5 (31)	30 (47)	

V4 fourth ventricle, *GTR* gross total resection, *STR* subtotal resection, *CMB* classic medulloblastoma, *DNMB* desmoplastic/nodular medulloblastoma, *LC/A MB* large cell/anaplastic medulloblastoma

21 (15%) cases were classified as WNT, 41 (29%) as SHH, 16 (11%) as Group 3, and 64 (45%) as Group 4 (Table 3).

We further investigated the correlations between clinicopathological features and the molecular subgroups. In Group 3 and group 4 subgroups, tumors were more overrepresented in midline location ($p<0.001$) with larger tumor size ($p=0.010$). In addition, WNT and SHH subgroups presented rarer metastases ($p<0.001$) and fewer fourth ventricle involvement ($p<0.001$) than group 3 and group 4. Furthermore, we also examined the extent of tumor resection based on post-operative imaging and found that most patients who underwent GTR were prevalent in the WNT and SHH subgroups ($p=0.011$). Most DNMB tumors belonged to the SHH subgroup, and the proportion of DNMB tumors is the highest in the SHH subgroup. As for another three subgroups, tumors with classic histology accounted for the largest proportion, especially in the Group 4 subgroup, in which 86% of the tumors had classic histology ($p<0.001$). In terms of treatment, Group 3 and SHH subgroups were more administrated to chemotherapy ($p=0.012$), but there is no significant difference in radiation therapy between the four subgroups.

We further investigated the prognostic utility of the molecular assignments in our cohort. As shown in Fig. 3A, B, we observed a significant difference in OS and PFS rates amongst the molecular subgroups from the Kaplan Meier analysis ($p<0.0001$). The 5-year OS rate for the WNT group was 0.802 (95% CI: 0.566–0.933) compared to 0.622 (95% CI: 0.457–0.764) for SHH, 0.467 for Group 4 (95% CI: 0.343–0.595), and 0.274 for Group 3 (95% CI: 0.097–0.549). The 5-year PFS rate for the WNT group was 0.799 (95% CI: 0.563–0.931) compared to 0.524 (95% CI: 0.364–0.679) for SHH Group, 0.448 for Group 4 (95% CI: 0.534–0.827), and 0.274 for Group 3 (95% CI: 0.097–0.549). Furthermore, 33 out of 41 cases classified as SHH group were further evaluated and subdivided into SHH/*TP53* mutation group ($n=8$) and SHH/*TP53* wild-type group ($n=25$) based on *p53* IHC

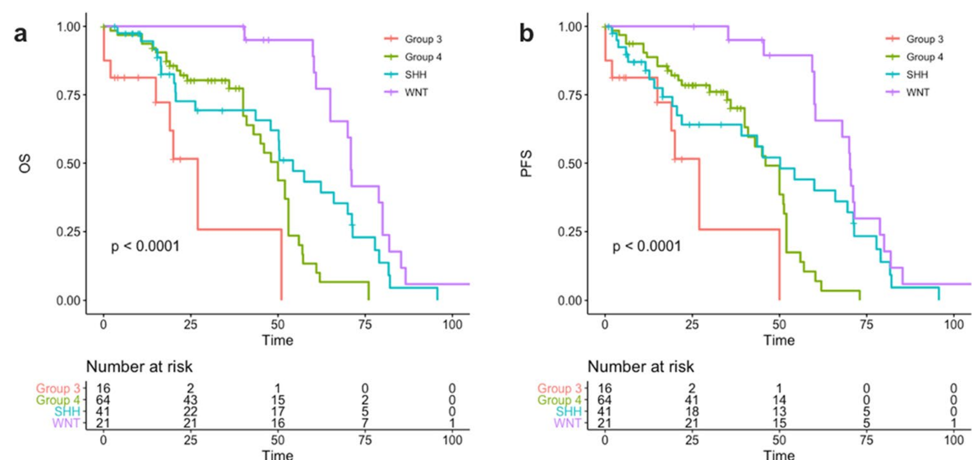


Fig. 3 Kaplan–Meier plots of A overall survival and B progression-free survival analysis of four molecular subgroups. *p* value was obtained using the log-rank test

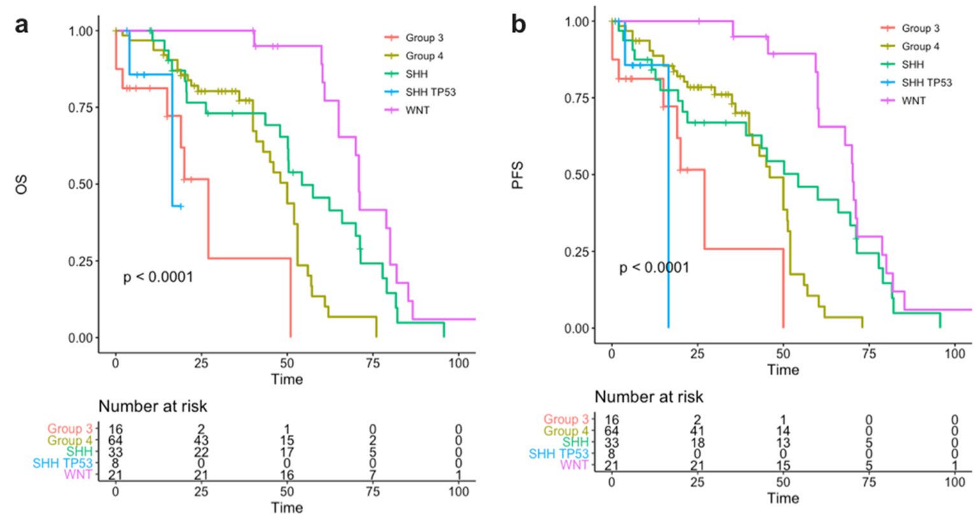


Fig. 4 Kaplan–Meier plots of A overall survival and B progression-free survival analysis of five molecular subgroup. *p* value was obtained using the log-rank test

Name	Total	Gene
Common genes	11	DKK2 EOMES GABRA5 GAD1 IMPG2 KCNA1 KHDRBS2 MAB21L2 PDLIM3 TNC WIF1
The 23-gene assay specific genes	12	EPHA7 IGF2BP3 LGR5 NKD1 OTX2 PEX5L RALYL ROBO1 TBR1 TRAC TTR ZFPM2
NanoString nCounter assay specific genes	11	ATOH1 EGFL11 EMX2 EYA1 HHIP NPR3 NRL OAS1 RBM24 SFRP1 UNC5D

Fig. 5 Venn diagram analysis of the 23-gene set with NanoString nCounter assay

results. As shown in Fig. 4A, B, the OS and PFS between five subgroups revealed significant differences, and SHH TP53-mutant tumors were associated with extremely poor prognosis ($p < 0.0001$).

Comparison of the 23-gene expression assay and NanoString nCounter assay

Northcott et al. had previously reported a Nanostring-based 22-gene expression assay for medulloblastoma molecular classification. As shown in Fig. 5, 11 genes were overlapped between the 23-gene expression assay and NanoString nCounter assay, including several well-studied genes like *DKK2*, *WIF1*, *EOMES*, *GABRA5*, and so on. The remaining 12 genes out of the 23-gene expression assay were distinct from NanoString nCounter assay. To compare the concordance between the 23-gene expression assay and NanoString nCounter assay, a group of 29 medulloblastoma specimens collect between January 2012 and December 2015 were analyzed by both methods. Of 29 specimens, one case was

Table 4 Confusion matrix by molecular subgroups of the 28 medulloblastoma cases

	NanoString nCounter assay				
		WNT	SHH	Group 3	Group 4
23-gene expression assay	WNT	1			
	SHH		15		1
	Group 3			4	
	Group 4				7

NanoString nCounter assay results are shown across the top row, and the 23-gene expression assay results are shown in the left column

eliminated due to insufficient tumor tissue, and the remaining 28 specimens were successfully analyzed by the 23-gene expression assay and NanoString nCounter assay. As shown in Table 4, the 23-gene expression assay classified one case as WNT, 16 as SHH, four as Group 3, and seven as Group 4. Meanwhile, according to NanoString nCounter assay results, one specimen was classified as WNT, 15 as SHH, four as Group 3, and eight as Group 4. There was one case classified as Group 4 by NanoString nCounter assay assigned to SHH Group by the 23-gene expression assay. Overall, the concordance rate reached 96.4% (27/28, 95%CI: 0.80–1.00) between the NanoString nCounter assay and 23-gene expression assay, highlighting that the results of these two transcriptomic methods were highly comparable.

Discussion

Medulloblastoma is the most common childhood malignant brain tumor and is a leading cause of cancer-related death in children. Over the last decade, medulloblastoma has been widely recognized as a heterogeneous disease with disparate genomics, biology, and clinical outcomes. WNT, SHH-P53 wildtype, SHH-P53 mutant, and non-WNT/non-SHH (including Group 3 and Group 4), with disparate demographics, clinical characteristics, genetic features [13, 21]. Precise molecular classification of medulloblastoma in routine clinical setting has become critical for subgroup-specific therapies and subgroup-driven clinical trials design [1, 6]. Given the most favorable outcome, WNT group patients are currently controlled by reduction intensive treatment to reduce long-term sequelae and other adverse side-effects. In contrast, patients with Group 3 represent terrible prognosis that still lacks new therapeutic options. Treatment of Group 3 patients with *MYC* amplification could be suppressed by using BET bromodomain inhibition such as JQ1 and Milciclib [27]. Beyond *MYC*, Gholamin et al. identify a humanized anti-CD47 antibody that could be a potentially safe and effective therapeutic agent for managing Group 3 tumors [7].

Several studies have reported the promise of gene expression- and methylation- based signatures in the assignment of medulloblastoma subgroups. However, molecular subtyping using gene expression and/or methylation microarray is not economical or flexible for surgical pathology laboratories and hardly compatible with long-term archived FFPE samples [25]. IHC panels may be applied as surrogates and are widely used; however, intraobserver and interobserver variations of IHC evaluation undermine their accuracy. Even more importantly, IHC approach is lack of specific antibodies for distinguishing Group 3 and Group 4 tumors confirmed in multiple studies to be both genetically and clinically distinct. A previous study revealed no significant differences of OS and PFS between the three molecular groups (WNT, SHH, non-WNT/SHH) assigned by IHC method [29]. Northcott et al. had developed a gene expression assay based on NanoString nCounter Technology and achieved a high accuracy of 98% in fresh-frozen tissue [17]. However, the assay performance decreased to 68% accuracy when applied to FFPE tissue samples. In addition, the integrated and enclosed platform, high cost, and sophisticated workflow may limit this NanoString-based gene expression assay for routine clinical use.

In the current study, we performed gene expression analyses and identified a 23-gene signature for molecular subgroups of medulloblastoma. We demonstrated that the gene signature could accurately identify the subgroups in an independent test set of medulloblastoma tumors (219/230, 95.2% accuracy), which is comparable to that of other signatures established by mRNA or methylation markers (ranging 95.7 from 98%) [4, 11, 17]. We compared the signature genes with the 22 genes identified by Northcott et al. The two mRNA signature shared 11 common genes including several well-studied markers associated with WNT group (*WIF1* and *DKK2*), SHH group (*PDLIM3*), Group 3 (*GABRA5*, *IMPG2*, and *MAB21L2*), and Group 4 (*EOMES* and *KCNA1*). Furthermore, 12 out of 23 genes were distinct from the Northcott et al. gene panel. Among these genes, *EPHA7* (WNT), *TBR1* (Group 4), and *OTX2* (Group 3 and Group 4) were found to be associated with subgroup-specific somatic variants [14]. Whittier et al. also reported that *LGR5* is significantly and uniquely over-expressed in the WNT group tumors [28]. Future research focus on the rest of our signature genes may provide additional opportunities to improve our understanding of the biological and prognostic significance of medulloblastoma subgroup.

To determine whether the 23-gene signature could really support patients' risk stratification in routine practice, we transferred it into a clinical applicable PCR-based gene expression assay and further validated the assay in a large independent cohort of 142 Chinese younger and adult medulloblastoma patients. The robust performance of our gene expression assay in FFPE materials was very satisfactory, even for specimens archived 15 years ago. This may be critical for the circumstance of relapsed medulloblastoma patients as their biopsied samples may archived for several years.

In this cohort, our data showed a preponderance of Group 4 tumors (45%), followed by SHH group tumors (29%), WNT group tumors (15%), and Group 3 tumors (11%). This is consistent with previous studies which showed Group 4 and SHH group are always the most common subgroup. It is also reported that Group 3 is predominantly amongst infants and children, whereas WNT group was uncommon in pediatric patients. Given that our cohort does not include infant patients, thus the proportion of WNT group is slightly higher and the proportion of Group 3 is lower than previous studies which showed 10% and 15% prevalence of WNT and Group 3 tumors, respectively [23].

Our results demonstrated that the molecular subgroups by the 23-gene signature could provide valuable insights in to the patients' risk stratification. The survival analysis of OS and PFS revealed substantial differences between molecular subgroups. Patients classified as WNT group had the most favorable prognosis, Group 4 patients had an intermediate prognosis, while SHH and Group 3 patients had the worst outcome. Noteworthily, Nataliya et al. reported that *TP53* mutation was enriched in 21% of SHH group tumors and associated with extremely poor survival groups [30]. In our cohort, we found the presence of SHH/*TP53* mutant tumors in 24% (8/33) of SHH group, which was indeed associated with markedly poorer prognosis.

In conclusion, we have identified, developed, and validated a novel 23-gene expression assay for medulloblastoma molecular classification. We show that this gene expression assay can be accurately and reliably applied to long-term archived FFPE specimens, hence may serve as an auxiliary diagnostic tool to aid pathologists with the growing unmet clinical needs for medulloblastoma subgroup determination. Future incorporation of this gene expression assay into prospective clinical trials is warranted in order to further evaluate its clinical utility for either patient prognostication and patient selection for targeted therapy.

Author contribution Collecting clinic data of patients, writing this paper: YW, JH. Collecting samples and performing NanoString assay: YY, XY. Real-time PCR experiment: QX. Data analysis and language modification: YS. Experimental design: JX.

Funding The present study was funded by a grant from the Shanghai Municipal Science and Technology Commission (Grant No.:16410712000).

Data availability The data that support the findings of this study are openly available on NCBI (URL: https://pubmed.ncbi.nlm.nih.gov/), including three published datasets (GSE21140, GSE37382, and GSE37418).

Declarations

Conflict of interest Authors Qinghua Xu and Yifeng Sun are employees of Canhelp Genomics Co., Ltd. No other potential conflicts of interest were disclosed by the authors.

References

1. Archer TC, Mahoney EL, Pomeroy SL (2018) Medulloblastoma: molecular classification-based personal therapeutics. Neurotherapeutics 14(2):265–273
2. Castillo-Rodríguez RA, Dávila-Borja VM, Juárez-Méndez S (2018) Data mining of pediatric medulloblastoma microarray expression reveals a novel potential subdivision of the Group 4 molecular subgroup. Oncol Lett 15(5):6241–6250
3. Chang CC, Lin CJ (2007) LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2(3, 27):1–27
4. Cruzeiro GAV, Salomão KB, de Biagi CAO (2019) A simplified approach using Taqman low-density array for medulloblastoma subgrouping. Acta Neuropathol Commun 7(1):33–10
5. Dai M, Wang P, Boyd AD (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33(20):e175–e175. https://doi.org/10.1093/nar/gni179
6. Gajjar AJ, Robinson GW (2014) Medulloblastoma-translating discoveries from the bench to the bedside. Nat Publ Group 11(12):714–722
7. Gholamin S, Mitra SS, Feroze AH (2017) Disrupting the CD47-SIRPα anti-phagocytic axis by a humanized anti-CD47 antibody is an efficacious treatment for malignant pediatric brain tumors. Sci Transl Med 9(381):eaaf2968
8. Gopalakrishnan V, Tao R-H, Dobson T, Brugmann W, Khatua S (2015) Medulloblastoma development: tumor biology informs treatment decisions. CNS Oncol 4(2):79–89
9. Ihaka R, Arnqvist HJ, Ballermann BJ, Gentleman R, King GL (1988) Receptors for and effects of insulin and IGF-I in rat glomerular mesangial cells. Am J Physiol 254(3 Pt 1):C411–C416
10. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8(1):118–127
11. Kunder R, Jalali R, Sridhar E (2013) Real-time PCR assay based on the differential expression of microRNAs and protein-coding genes for molecular classification of formalin-fixed paraffin embedded medulloblastomas. Neuonc 15(12):1644–1651
12. Li Y, Song Q, Day BW (2019) Phase I and phase II sonidegib and vismodegib clinical trials for the treatment of paediatric and adult MB patients: a systemic review and meta-analysis. Acta Neuropathol Commun 7(1):123–128
13. Louis DN, Perry A, Wesseling P (2021) The 2021 World Health Organization classification of tumors of the central nervous system: a summary. Neuro Oncol 23(8):1231–1251
14. Northcott PA, Buchhalter I, Morrissy AS (2017) The whole-genome landscape of medulloblastoma subtypes. Nature 547(7663):311–317
15. Northcott PA, Korshunov A, Witt H (2011) Medulloblastoma comprises four distinct molecular variants. Am Soc Clin Oncol Educ Book 29(11):1408–1414
16. Northcott PA, Shih DJH, Peacock J (2012) Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature 488(7409):49–56
17. Northcott PA, Shih DJH, Remke M (2012) Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol 123(4):615–626
18. Piccolo SR, Sun Y, Campbell JD, Lenburg ME, Bild AH, Johnson WE (2012) A single-sample microarray normalization method to facilitate personalized-medicine workflows. Genomics 100(6):337–344
19. Piccolo SR, Withers MR, Francis OE, Bild AH, Johnson WE (2013) Multiplatform single-sample estimates of transcriptional activation. Proc Natl Acad Sci USA 110(44):17778–17783

20. Reimers M, Carey VJ (2006) Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 411:119–134
21. Robinson G, Parker M, Kranenburg TA (2012) Novel mutations target distinct subgroups of medulloblastoma. Nature 488(7409):43–48
22. Schwalbe EC, Williamson D, Lindsey JC (2013) DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. Acta Neuropathol 125(3):359–371
23. Sengupta S, PomeranzKrummel D, Pomeroy S (2017) The evolution of medulloblastoma therapy to personalized medicine. F1000Res. 6(490):490
24. Tabori U, Baskin B, Shago M (2010) Universal poor survival in children with medulloblastoma harboring somatic TP53 mutations. Am Soc Clin Oncol Educ Book 28(8):1345–1350
25. Taylor MD, Northcott PA, Korshunov A (2012) Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol 123(4):465–472
26. Wang Q, Gan H, Chen C (2017) Identification and validation of a 44-gene expression signature for the classification of renal cell carcinomas. J Exp Clin Cancer Res 36(1):176
27. Wang J, Garancher A, Ramaswamy V, Wechsler-Reya RJ (2018) Medulloblastoma: from molecular subgroups to molecular targeted therapies. Annu Rev Neurosci 41:207–232
28. Whittier KL, Boese EA, Gibson-Corley KN (2013) G-protein coupled receptor expression patterns delineate medulloblastoma subgroups. Acta Neuropathol Commun 1(1):66–15
29. Zhang Z-Y, Xu J, Ren Y (2014) Medulloblastoma in China: clinicopathologic analyses of SHH, WNT, and non-SHH/WNT molecular subgroups reveal different therapeutic responses to adjuvant chemotherapy. PLoS One 9(6):e99490
30. Zhukova N, Ramaswamy V, Remke M (2013) Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. Am Soc Clin Oncol Educ Book 31(23):2927–2935

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.