Clinics in Oncology

Research Article
Published: 05 Oct, 2023

Extensive Coding and Antisense-Transcript Based Characterization of Indian Medulloblastoma a Précised Approach

Neetu S[1#], Dinesh KS[1,2#], Ratnesh KT[1#], Nawazish A[1], Pratap S[1], Abhishek M[1], Ajay KS[3], Uma Shankar S[3], Chhitij S[4], Bal Krishna O[4*] and Devendra KG[1,2]*

[1]Molecular Biology Unit, Centre for Advance Research, King George's Medical University, India

[2]Department of Pediatric Surgery, Super Specialty Pediatric Hospital & Post Graduate Teaching Institute, India

[3]Department of Pathology, King George's Medical University, India

[4]Department of Neurosurgery, King George's Medical University, India

[#]These authors contributed equally to this work

Abstract

Medulloblastoma (MB) is a malignant pediatric brain-tumor with a high death rate. A comprehensive protein-coding and regulatory-transcriptomics are needed for personalized medicine. Molecularly sub-grouped 22 histopathologically-characterized Medulloblastoma (MB) into Group 4-MB (G4), Group 3-MB (G3), Wnt-MB, and SHH-MB by two accepted methods molecular inversion probe-based array and transcript-probe-based NanoString assay. Out of 22, 12 samples (10 MB and 2 controls) including all subgroups of MB were considered for long and short-read sequencing. Assembly was done through PacBio-ToFU pipeline and assessed through SQANTI using a pre-treating pipeline with 47 distinctive-features. Based on canonical/highly-conserved splice-junctions induced transcriptomics-complexity the MB-patients were categorized into low- or high-risk tumors which may be considered for precise treatment. Significant upregulation of Antisense-transcripts (AS), EZH2 in SHH, CENPX in SHH/G-3, COL18A1 in G-4 MB and transcripts TTR in SHH/G3/G4-MB; PDLIM3 and SFRP1 in SHH; NRL in G-3; and NNAT, an imprinted gene in G-4 MB were observed. The identified ASs may be used as a target for developing anti-antisense synthetic peptides or transcriptional-inhibition through classical CRISPR/Cas9 in combination with SHH/NRL/NNAT inhibitors may be used as a precise approach to target MB. Further, TTR identified in cerebrospinal fluid may be used as a marker to understand the prognosis of the disease.

Keywords: Medulloblastoma; SQANTI; Antisense transcripts; TTR; PDLIM3; SFRP1; NRL; NNAT; EZH2-AS1; CENPX; COL18A1-AS1

OPEN ACCESS

Correspondence:

Neetu Singh, Molecular Biology Unit, Centre for Advance Research, King George's Medical University, Lucknow 226 003, UP, India, Tel: +91-9984-801-444;

Bal Krishna Ojha, Department of Neurosurgery, King George's Medical University, Lucknow 226 003, UP, India, Tel: +91-9415-108-077;

Received Date: *21 Aug 2023*

Accepted Date: *14 Sep 2023*

Published Date: *05 Oct 2023*

Citation:

Neetu S, Dinesh KS, Ratnesh KT, Nawazish A, Pratap S, Abhishek M, et al. Extensive Coding and Antisense-Transcript Based Characterization of Indian Medulloblastoma a Précised Approach. Clin Oncol. 2023; 8: 2020.

ISSN: 2474-1663

Introduction

Present knowledge regarding transcriptome complexity including Alternative Splicing (AlS) and Alternative Polyadenylation (APA) aspects in molecularly grouped Medulloblastoma (MB) etiology is presently insufficient. MB counts for 16% to 20% of all brain tumors. Of these, 5.91% are Group-4 (G4), 4.05% are Group-3 (G3), 4.05% are Sonic Hedgehog (SHH), and 1.59% are WNT, and 0.380% are ill-defined [1]. Regardless of innumerable high-throughput gene-expression studies, the differential expression of coding RNAs has been thoroughly assessed. Additionally, some lncRNA-based MB subclass-precise chromosomal remodeling have been reported independently [2]. These post-transcriptional AlS and APA mechanisms which subsequently enhance proteome diversity have expressed to establish vital roles in cancer development, progression and mortality [3-5] and other diseases like diabetes [6,7], as well as in neurological disorders [8-12] and accordingly may engage in a cardinal role in the setting up MB complexity.

Technically, AlS has been accomplished chiefly using exon-microarrays and short-read RNA-sequencing. The two procedures are worthwhile for the recognition of AlS events such as exon skipping or intron retention and have entrenched the participation of AlS in MB. Nevertheless, both exon-microarrays and short-read RNA-sequencing have restrictions for the alignment of the actually expressed transcripts. As short-reads defy the continuation of the transcript sequences [13-15]. Hence, fails to resolve actual ambiguities of MB subgroups. Limited MB studies are available

which have catalogued specific transcriptomes to understand functional aspects of isoform diversity, identification of antisense, and long noncoding transcripts. However, a sudden increase in single-molecule sequencing technique to acquire full-length transcripts has helped to eliminate the need for hybrid assembly created by the alignment of short-read sequences against long-read sequences for transcript reconstruction.

Currently, different long-read transcriptome sequencing platforms are available: Pacific Biosciences (PacBio) [15-18], and Nanopore [19]. Here, we have used the popular PacBio Iso-Seq protocol and short-read sequencing through Illumina for enhancing coverage of different samples categorized in MB sub-groups through molecular inversion probe constructed assay (OncoScan Array) and transcript-probe-build NanoString assay [20,21].

As PacBio sequencing undergo, comparatively increased raw error rate (~15%) [22] and a lower yield compared to Illumina. Hence for PacBio error correction, short-read sequencing through Illumina was also conducted and hybrid assembly was used for transcript identification to remove limitations of short-read breaks which disrupt the actual open reading frames of the transcripts [13,15]. In lieu, we utilized SQANTI (Structural and Quality Annotation of Novel Transcript Isoforms) pre-processing pipeline for the examination of long- and short-read transcriptomics and have identified MB subgroup-specific as well as common differentially expressed full-length transcripts, full-length novel isoforms, alternative lengths of poly A transcripts and number of poly A motifs and, antisense transcripts [23]. These can be applicable in addressing development of small molecules. Additionally, a panel that can extensively and uniquely portray MB for time ahead clinical trials and precision medicine.

Methodology

Molecularly to sub-classify histopathologically characterized 25 MB into G4-MB, G3-MB, WNT-MB, and SHH-MB were processed through two endorsed methods Molecular Inversion Probe constructed assay (OncoScan Array) and transcript-probe-build NanoString assay (Supplementary Table 1). We collected raw tumor tissue in liquid nitrogen immediately after surgery for molecularly characterization and simultaneous immune-histopathological and examinations. A total of immunohistopathologically confirmed 22 posterior fossa-MB was treated for aCGH assays for evaluating Copy Number Variations (CNVs) on OncoScan array [24]. For control, posterior fossa tissue from the conventional trauma adult patient was utilized and labeled as MC. The other three controls were matched MB controls (M14C, M97C, M76C). All the specimens were acquired after informed consent.

Molecular inversion probe assay

22 MB samples (M1, M2, M7, M9, M12, M13, M22, M24, M28, M30, M32, M36, M39, M42, M48, M50, M55, M60, M62, M73, M76, M80) and two controls were molecularly characterized through OncoScan-Array (OA). Genomic DNA extraction, processing, scanning and analysis was done as reported previously [24] (Supplementary Table 2).

Transcript-probe-based NanoString nCounter assay

Twelve samples were processed through NanoString (NS) based method. Six samples (M-39, M48, M62, M73, M80, M83) along with 4 known samples of each group (G3-M1, G4-M2, SHH-M28, and WNT-M48) were processed on commercially available transcript-based code-set on NS against two controls (Supplementary Table 3-codeset). Probes for nCounter were designed for 22 code sets for different sub-types of Medulloblastoma from mRNA of *Homo sapiens*. Experiments have been performed according to the instruction manual (NanoString Technologies). nSolver™ Analysis Software 3.0 (NanoString Technologies) was used to perform background subtraction, spike-in-control normalization, and reference gene normalization. Using 5 housekeeping genes (ACTB, GAPDH, RPL19, TBP, TUBB), transcript counts were normalized to the and 2 MC1 and MC4 controls in the nCounter panel and the 5 housekeeping genes (ACTB, GAPDH, RPL19, TBP, TUBB) and compared to the control samples (Table 1).

Long read sequencing using PacBIO RS-II

The samples were categorized as following G3/4: M-1, M-2, M-39, M-73; G3/4 with SHH: M80, M83; SHH with G3/4: M13, M22, M28, M76; WNT-M48, M32, M60; WNT with SHH -M62. Out of which a total of 11 samples were processed for long Read Sequencing. G3/4: M-1 (G3/4 through OA and NS), M-39 (G3/4 through NS), M-73(G3/4 through NS); M80 (G3/4 with SHH through NS); SHH with G3/4: M13 (OA-SHH, NS-SHH with G3), M28 (OA-SHH, NS-SHH with G3 and G4 genes); M76 (NS-SHH with G3 and G4 genes); WNT: M32 (WNT alpha through 6-, 8+, 10q-OA) and M60 (WNT beta with no loss of Ch6) ; G4 M-2 (G4 through OA; G4 through NS) (Table 1).

Control: Two control samples including one pediatric matched control and one adult control from trauma patients were also processed for long read sequencing (Table 1).

Short Read Sequencing (SRS) through MiSeq

A total of 12 samples were processed for SRS G3/4: M-1, M-39, M-73, M-83; SHH with G3/4: M13, M28, M76; WNT: M32; G4: M97T, M-104; Control: M97C and M14C (Table 1).

(M-2, M60, and M80 samples were inadequate hence were not processed for short-read sequencing).

Overview of bioinformatics

To procure consensus transcripts, assembly was orchestrated through PacBio and ToFU pre-processing pipeline. Calculable (indel) errors were removed by correlating consensus transcripts to the reference genome and an amended transcriptome was generated. Further, false transcripts were removed using a SQANTI filter which provided us with a curated transcriptome. Illumina short reads were mapped against the RefSeq transcriptome annotation, the corrected, and the curated PacBio transcriptomes. In individual samples (M1, M39, M73, M83, M13, M28, M76, M22, M97, M104). SQANTI was used for the classification of long-read transcripts and with support of short-read transcripts were able to identify full-length-transcriptome, known and novel-transcript-isoforms, antisense transcripts, fusion transcripts. All the transcripts were further characterized as coding if they showed the presence of putative ORF. SQANTI also assessed the quality of data and the pre-processing pipeline using 47 unique descriptors. Canonical and Non-canonical splice junctions were also identified. All the above samples were aligned to the human reference genome and a differential expression level matrix was created between tumor and control samples. Using standard settings (minimum fold change 2.0 and p-value correction was through Bonferroni and cut off p-value was 0.05) heat map was generated between control and experimental samples. Using standard t-test each subtype was compared with control as well as subtypes were also compared

Table 1: Detailed Characterization of Medulloblastoma based on previous reports.

		Categorization of transcript profiling through Long Read Sequencing – G Fold Approach	Categorization of transcript profiling based on NanoString [21]	Categorization of transcript profiling based on through short read Sequencing	Broad and Focal Amplifications [20]	OncoScan Array
1	M1	OTX1: 3.7875; DSCAM: 2.5474; PROM2: 2.00402; SNCAIP: 4.31335; OTX2: 2.43606	GP-3- 4 genes; 2 genes of GP-4; 1 gene of WNT	GP-3- 5 genes; 2 genes of GP-4; 1 gene of WNT	SNCAIP duplication – GP-4-Beta and OTX2 gain in GP-3 Beta	
2	M2	MYCN: 2.89776; CDK6: 2.03726; DACH1: 2.61703; TIGAR: 2.60825	GP-3- 3 genes; 3 genes of GP-4	Sample inadequate	MYCN and CDK6 amplification –GP-4 alpha	1q+, 7+, 8-, i17q, 11-
						MYCN (HCN); CDK6
						GP-4 alpha
3	M7				CDK6 amplification –GP-4 gamma	1q+, 7q+, 8p-, i17q, 11-
						CDK6
						GP-4 gamma
4	M9	Sample inadequate	Sample inadequate			6-, 8+
						WNT Alpha
5	M12					2+, 4+, 7+, 9+, 11+, 17+, 19+, 20+
						CDK6 and MYCN-Anaplastic Ependymoma (Histopathologically reported Medulloblastoma
6	M13	SOX11: 2.28819; DACH1: 2.14536	SHH-4 genes and 2 genes of GP-3	SHH-5 genes; 2 genes of GP-3; 1 gene of WNT	9-, 10q-, 17p- and MYCN amplification, GLI2 amplification, YAP1 amplification	9-LOH, 3+, MYCN amplification-SHH alpha
				YAP1; MYCN		
7	M22	-	SHH-2 genes; 2 WNT genes; 2 gene of GP-3 and 1 gene of GP-4			9-, 3q+ -SHH alpha
8	M24				MYCN and CDK6 amplification –GP-4 alpha	MYCN
						CDK6
						1q+, 7+, 8p-, i17q, 11-
						GP-4 alpha
9	M28	CTNNB1 -2.68278; SOX2-OT -2.49813; SOX9 -2.29142; SFRP1 -2.74285; PPM1H -2.82398	SHH-4 genes and 1 gene of GP-3 and 1 gene of GP-4	SHH-5 genes and 2 genes of GP-3 and 1 gene of GP-4;	9-, 10q-, 17p- and MYCN amplification, GLI2 amplification, YAP1 amplification SHH alpha	6- (HCL), 10q-, MYCN (HCN)
		PTCHD1; -2.76498; PTCH1: -3.01062		YAP1; MYCN	6- WNT alpha	WNT alpha and SHH alpha
10	M30					1+, 2+, 4+, 5+, 6+, 7+, 9+, 12+, 14+, i16, 17+, 19+, 21+
						CDK6 and MYCN (Histopathologically Meduloblastoma-Expired)
11	M32	CNTN1: -2.08981; SOX2-OT: -3.15584; SOX9: -3.52228; CTNNB1: -2.42989; SFRP1 -3.11352; **WIF1: 4.5439; DKK4**: 2.61646; PTCHD1: -2.74897	Sample inadequate	Sample inadequate	6- , 8+, CTNNB1 mutation-WNT alpha	6-, 8+ (MYC amplification), 10q-
					10q- SHH alpha	WNT
12	M36				CDK6 amplification –GP-4 gamma	7+, 8-, i17q, X-
						CDK6
						GP-4 gamma
13	M39	OTX2: 2.69446; MEIS3 2.19684	GP-3 - 4 genes and 2 genes of GP-4	GP-3 - 4 genes and 2 genes of GP-4	OTX2 Gain: GP-3 beta	-
14	M42					1+, 2+, 4+, 5+, 6+, 7+, 9+, 10+,11+, 12+, 13+, 14+, 17+, 18+, 19+, 20+ 21+
						CDK6 and MYCN (Histopathologically Meduloblastoma-Expired)
15	M48		No up-regulation of any subclass		6- , 8+, CTNNB1 mutation-WNT alpha	6- WNT alpha
16	M50	-	-	-	CDK6 amplification –GP-4 gamma	7+, 8-, i17q, X-
						CDK6 GP-4 gamma

17	M55	-	-	-	CDK6 amplification –GP-4 gamma	1q+, 7q+, 8p-, i17q CDK6 GP-4 gamma
18	M60	SOX4: -2.55303			WNT Beta	No alterations
19	M62		SHH-1gene and 1 gene of WNT			
20	M73	TCF4: -2.10229; SOX2-OT: -2.21847; GABBR1 -2.25801; CTNNB1: -2.38753; MYC: 2.47426; OTX2: 2.60189	GP-3 - 5 genes and 1 gene of GP-4	GP-3 - 5 genes; 1 gene of GP-4 and 1 gene of WNT	MYC Amplification- GP-3 gamma; OTX2 gain in GP-3 Beta	-
21	M76	BMP5:2.943; BCYRN1: -3.5674	SHH-5 genes; 2 genes of GP-3; 1 gene of WNT	SHH-5 genes; 2 genes of GP-3; 1 gene of WNT; YAP1; MYCN	9-, 10q-, 17p- and MYCN amplification, GLI2 amplification, YAP1 amplification	-
22	M80	No. of reads were less to be compared	SHH -1 gene; GP-3 - 3 genes and 3 genes of GP-4	-		-
23	M83	-	SHH -1 gene; GP-3 - 3 genes and 4 genes of GP-4	GP-3 - 3 genes; 2 genes of GP-4 and WNT -1 gene		
24	M97	-		To be analysed		
25	M104	-		4 genes of GP-4; GP-3 - 1 genes		

among themselves.

Results

Stratification of 24 MB samples and two control samples

Twenty-four Medulloblastoma samples (M1, M2, M7, M9, M12, M13, M22, M24, M28, M30, M32, M36, M39, M42, M48, M50, M55, M60, M62, M73, M76, M80, M83, M97, M104) either processed through MIP assay, Transcript-probe-based NanoString nCounter assay or through long- and short-read sequencing have been stratified in various groups as well as sub-groups as per recent broad and focal amplifications [20] and transcript based markers as depicted in Table 1 [21] and supplementary data (Supplementary Table 3a and Supplementary Figure 1a, Supplementary Table 3b and Supplementary Figure 1b, Supplementary Table 3c and Supplementary Figure 1c, Supplementary Table 3d and Supplementary Figure 1d; Supplementary Table 4 transcript-probe based NanoString nCounter characterization).

Structural classification of Transcripts through SQUANTI including both long and short read Sequencing

Through SQANTI analyses, we were able to categorize unique transcript isoforms, structurally into Full Splice Match (FSM), Incomplete Splice Match (ISM), Antisense (AS), intergenic, Novel in Catalog (NIC), Novel not in Catalog (NNC). For each structural category, we filtered between coding and non-coding transcripts. Coding transcripts were filtered between canonical and non-canonical junctions based on splice junctions (GT-AG as well as GCAG and AT-AC as canonical splicing and all the other possible combinations as non-canonical splicing). Further Canonical junctions were categorized as true and false based on Reverse Transcriptase template Switching (RTS) which was approximately 2% for known splice junctions [23].

The result of 11 samples including two controls was processed for sequencing raw data analysis sequenced by the PacBio RSII (Long-Read Sequencing) sequencer, further long-reads were also supported by short-read sequence through Illumina. Importantly most of the known canonical junctions were supported by short-read data depicted through individual SQUANTI reports of the individual samples (Supplementary Data 5 PDF of 11 samples processed through SQUANTI). The results found as SQANTI reports categorized unique transcript isoforms of all Medulloblastoma samples in FSM; ISM; NIC; NNC; Antisense-transcripts (AS); Intergenic (IG)-transcripts (Supplementary Table 6a). Further, the transcript isoforms were classified into coding and non-coding sequences through SQANTI-GeneMarkS-T-algorithm (Supplementary Table 6b). Coding transcripts of FSM, ISM, and NIC represented known-canonical with mean value of 92.00 (SD ± 2.48), 85.89 (SD ± 2.04) and 42.62 (SD ± 20.69) respectively (Supplementary Table 6c). However, ~0 to 1.61% were non-canonical in nature in FSM, ISM, and NIC (Supplementary Table 6c). Canonical junctions were categorized as RTS true i.e., percent of transcripts in which reverse transcription switching occurred were 4.43% (SD ± 0.84) in FSM, 3.48% (SD ± 0.76) in ISM, 5.35% (SD ± 0.89) in NIC, while no reverse transcription i.e., 0% was observed in canonical transcripts of NNC, AS and IG transcripts (Supplementary Table 6d).

The transcripts were also compared with annotated Transcription Start Sites (TSSs) and Polyadenylation sites (PolyA)

Distance to Annotated-Polyadenylation PolyA/Transcription Termination Site (TTS)/annotated Transcription Start Site (TSS) Site in FSM, ISM, NIC, NNC. AS and IG transcripts showed no applicability of PolyA/TTS/TSS. The distance of Poly A transcript/TTS/TSS from annotated Poly A /TTS/TSS site is (equal to 1). More than 1 was depicted as positive and less than 1 negative value. Negative values indicated upstream of annotated PolyA site/TTS, while downstream of the annotated-TSS site respectively, (Supplementary Table 6e).

Detection of poly A motif from 3'end both for FSM and ISM: Poly A Motif were identified showing negative value to annotated and specific annotated PolyA site peaking between -10 to -20 in all structural categories of transcript did not show the significant change between control and all MB samples (Supplementary Table 6f). The most common motif identified was AATAAA in all control and tumor samples ~60%.

Novel Antisense (AS) transcripts: Canonical, RTS-false, coding (as per ORF length polyA transcripts), and predicted Non-sense Mediated Decay (NMD False) were considered as filters for identifying AS transcripts. Based on aforesaid criteria sub-group-

Table 2: Based on following criteria, Canonical, RTS-false, coding (as per ORF length polyA transcripts), and predicted non-sense mediated decay (NMD False) following AS transcripts were identified. For example, CROCC, TOM1, STON2, and AK5 were WNT-specific; CENPX and EZR were common between SHH and Group-3; while ZNF391 and ZNF 865 were specific to Group-3 and COL18A1 and ZC3HAV1L to Group-4.

MB Subgroups	Chromosome	Gene ID	Entrez Gene Summary	GeneCards Summary for Gene	GeneCards Summary for AS Gene
Dominant SHH	17	CENPX		CENPX (Centromere Protein X) is a Protein Coding gene. Among its related pathways are Homology Directed Repair and Chromosome Maintenance.	
Dominant SHH	6	EZR	The cytoplasmic peripheral membrane protein encoded by this gene functions as a protein-tyrosine kinase substrate in microvilli. As a member of the ERM protein family, this protein serves as an intermediate between the plasma membrane and the actin cytoskeleton. This protein plays a key role in cell surface structure adhesion, migration and organization, and it has been implicated in various human cancers. A pseudogene located on chromosome 3 has been identified for this gene. Alternatively spliced variants have also been described for this gene. [provided by RefSeq, Jul 2008]	EZR (Ezrin) is a Protein Coding gene. Diseases associated with EZR include Autosomal Recessive Non-Syndromic Intellectual Disability and Neurofibromatosis, Type Ii. Among its related pathways are Blood-Brain Barrier and Immune Cell Transmigration: VCAM-1/CD106 Signaling and wtCFTR and delta508-CFTR traffic / Generic schema (norm and CF). Gene Ontology (GO) annotations related to this gene include RNA binding and protein domain specific binding. An important paralog of this gene is RDX.	EZR-AS1 (EZR Antisense RNA 1) is an RNA Gene, and is affiliated with the lncRNA class.
Dominant GP3	17	CENPX		CENPX (Centromere Protein X) is a Protein Coding gene. Among its related pathways are Homology Directed Repair and Chromosome Maintenance.	
Dominant GP3	6	ZNF391		ZNF391 (Zinc Finger Protein 391) is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include nucleic acid binding and DNA-binding transcription factor activity. An important paralog of this gene is ZNF300.	
	19	ZNF865		ZNF865 (Zinc Finger Protein 865) is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include transcription coregulator activity. An important paralog of this gene is ZNF667.	
Dominant GP4	7	ZC3HAV1L		ZC3HAV1L (Zinc Finger CCCH-Type Containing, Antiviral 1 Like) is a Protein Coding gene. An important paralog of this gene is ZC3HAV1.	ZC3HAV1L (Zinc Finger CCCH-Type Containing, Antiviral 1 Like) is a Protein Coding gene. An important paralog of this gene is ZC3HAV1.
			This gene encodes a CCCH-type zinc finger protein. This antiviral protein inhibits viral replication by recruiting cellular RNA degradation machineries to degrade viral mRNAs. The encoded protein plays an important role in the innate immune response against multiple DNA and RNA viruses, including Ebola virus, HIV and SARS-CoV-2 (which causes COVID-19). [provided by RefSeq, Sep 2021]	ZC3HAV1 (Zinc Finger CCCH-Type Containing, Antiviral 1) is a Protein Coding gene. Diseases associated with ZC3HAV1 include Eastern Equine Encephalitis. Among its related pathways are HIV Life Cycle and Oncogenic MAPK signaling. Gene Ontology (GO) annotations related to this gene include *RNA binding* and *NAD+ ADP-ribosyl transferase activity*. An important paralog of this gene is PARP12.	
		COL18A1	This gene encodes the alpha chain of type XVIII collagen. This collagen is one of the multiplexins, extracellular matrix proteins that contain multiple triple-helix domains (collagenous domains) interrupted by non-collagenous domains. A long isoform of the protein has an N-terminal domain that is homologous to the extracellular part of frizzled receptors. Proteolytic processing at several endogenous cleavage sites in the C-terminal domain results in production of endostatin, a potent antiangiogenic protein that is able to inhibit angiogenesis and tumor growth. Mutations in this gene are associated with Knobloch syndrome. The main features of this syndrome involve retinal abnormalities, so type XVIII collagen may play an important role in retinal structure and in neural tube closure. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2014]	COL18A1 (Collagen Type XVIII Alpha 1 Chain) is a Protein Coding gene. Diseases associated with COL18A1 include Knobloch Syndrome 1 and Glaucoma, Primary Closed-Angle. Among its related pathways are Integrin Pathway and ERK Signaling. Gene Ontology (GO) annotations related to this gene include *identical protein binding* and *structural molecule activity*. An important paralog of this gene is COL4A1.	COL18A1-AS1 (COL18A1 Antisense RNA 1) is an RNA Gene, and is affiliated with the lncRNA class. Diseases associated with COL18A1-AS1 include Glioma Susceptibility 1.

Dominant WNT	22	TOM1	This gene was identified as a target of the v-myb oncogene. The encoded protein shares its N-terminal domain in common with proteins associated with vesicular trafficking at the endosome. It is recruited to the endosomes by its interaction with endofin. Several alternatively spliced transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Oct 2008]	TOM1 (Target of Myb1 Membrane Trafficking Protein) is a Protein Coding gene. Diseases associated with TOM1 include Immunodeficiency 85 And Autoimmunity and Congenital Muscular Dystrophy-Dystroglycanopathy Type A6. Among its related pathways are Innate Immune System and C-MYB transcription factor network. Gene Ontology (GO) annotations related to this gene include *clathrin binding*. An important paralog of this gene is TOM1L2.	ENSG00000273176 (Novel Transcript, Antisense To TOM1) is an RNA Gene, and is affiliated with the lncRNA class.
	14	STON2	This gene encodes a protein which is a membrane protein involved in regulating endocytotic complexes. The protein product is described as one of the clathrin-associated sorting proteins, adaptor molecules which ensure specific proteins are internalized. The encoded protein has also been shown to participate in synaptic vesicle recycling through interaction with synaptotagmin 1 required for neurotransmission. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jan 2012]	STON2 (Stonin 2) is a Protein Coding gene. Among its related pathways are Clathrin-mediated endocytosis and Vesicle-mediated transport. An important paralog of this gene is STON1-GTF2A1L.	
	1	AK5	This gene encodes a member of the adenylate kinase family, which is involved in regulating the adenine nucleotide composition within a cell by catalyzing the reversible transfer of phosphate groups among adenine nucleotides. This member is related to the UMP/CMP kinase of several species. It is located in the cytosol and expressed exclusively in brain. Alternatively spliced transcript variants encoding distinct isoforms have been identified for this gene. [provided by RefSeq, Jul 2008]	AK5 (Adenylate Kinase 5) is a Protein Coding gene. Diseases associated with AK5 include Limbic Encephalitis and Anterograde Amnesia. Among its related pathways are AMPK Enzyme Complex Pathway and purine nucleotides de novo biosynthesis. Gene Ontology (GO) annotations related to this gene include *nucleoside diphosphate kinase activity* and *nucleoside kinase activity*. An important paralog of this gene is AK9.	ENSG00000287647 (Novel Transcript, Antisense To AK5) is an RNA Gene, and is affiliated with the lncRNA class.
		CROCC		CROCC (Ciliary Rootlet Coiled-Coil, Rootletin) is a Protein Coding gene. Diseases associated with CROCC include Alstrom Syndrome and Joubert Syndrome 1. Gene Ontology (GO) annotations related to this gene include *structural molecule activity* and *kinesin binding*. An important paralog of this gene is CROCC2.	

specific novel AS transcripts were identified. For example, CROCC, TOM1, STON2, and AK5 were WNT-specific; CENPX and EZR were common between SHH and Group-3; while ZNF391 and ZNF 865 were specific to Group-3 and COL18A1 and ZC3HAV1L to Group-4 (Table 2).

Differential analysis among different subtypes of MB and control: For this the combined long and short read for each sample was considered and differential analysis of all subtypes of MB *vs.* control was analyzed, considering Bonferroni less than <0.05, we identified 35 transcripts (Supplementary Table 7-Bonferroni). Further, categorized and characterized 9 MB in different subgroups compared to two control samples through RPKM values as transcript-based markers suggested by Northcott et al. Supplementary Table 8. On further comparison among different subtypes, i.e., comparison of SHH *vs.* G3/G4 samples through our long and short-read sequencing (Table 3a-3c) as well as transcripts which have been previously reported to categorize different subtypes of medulloblastoma (Table 4a-4c). We identified TTR, PDLIM3, and SFRP1 as highly significantly up-regulated genes, which were upregulated in SHH compared to both G3 and G4 (Table 3a, 3b, 4a, 4b). NRL was specifically increased in G3 compared to SHH and G4 (Table 4a, 4c). While NNAT was up-regulated in G4 compared to SHH and G3 (Table 3c).

Customized nCounter based Codeset for comprehensive characterization

Most common and sub-group specific 24-AS transcripts, 35 differentially regulated transcripts and 22 previously reported transcripts have been incorporated as probe-based direct transcript-based characterization has been prepared to run on NCounter NanoString (Supplementary Table 9). Importantly, all the transcripts were reported as false predicted Nonsense-Mediated Decay (NMD) transcripts.

Discussion

Initially, we followed a personalized approach to characterize each Medulloblastoma (MB) sample based on the Molecular Inversion Probe (MIP)- OncoScan array and transcript-based NanoString array. Individual samples did not show single subgroup-specific genes, they showed genes specific to two or more than two subgroups G3/4: M-1, M-2, M-39, M-73; G3/4 with SHH: M80, M83; SHH with G3/4: M13, M22, M28, M76; WNT-M48, M32, M60; WNT with SHH -M62. This categorization of samples increased the heterogeneity among various subgroups of MB and categorized as high-risk tumors as well as favors more to comprehensive treatment instead of only surgical and radiological treatments. The mixed subgroup was annotated

Table 3a: Only RPKM (SHH *vs.* Group 3).

Gene	SHH Mean	SHH SD	Group 3 Mean	Group 3 SD	P value	Significance
SLC1A7	2.2	3.5	228.2	310.3	P>0.05	ns
RD3	2	2.8	61.5	99.6	P>0.05	ns
CHRM3	3.9	5.5	30.8	16.1	P>0.05	ns
TBR1	0.3	0.1	204.2	207.1	P>0.05	ns
FEZF2	0.1	0.1	25.2	40.4	P>0.05	ns
IMPG2	0.7	0.3	202.2	370.6	P>0.05	ns
SST	0.6	0.8	256	539.6	P>0.05	ns
SLC6A3	5.3	7.9	113	145.3	P>0.05	ns
NEUROG1	0.4	0.8	133	104	P>0.05	ns
TULP1	0.3	0.5	72.6	100.9	P>0.05	ns
PTP4A1	37	30.8	46.1	24	P>0.05	ns
GNGT1	0.3	0.5	28.8	45	P>0.05	ns
FEZF1	0.2	0	12.6	13.6	P>0.05	ns
SLC13A4	318.7	635.8	2.7	2.1	P>0.05	ns
WDR38	116.3	232.5	0.1	0.1	P>0.05	ns
SPAG6	45.7	91	2.6	2.8	P>0.05	ns
ARMC3	32.1	63.7	0.5	0.4	P>0.05	ns
DYDC2	22	43.4	1.3	1.4	P>0.05	ns
PRAP1	70.9	141.5	0.5	0.6	P>0.05	ns
FOLR1	447.3	894	1.3	1.6	P>0.05	ns
C11orf88	43.2	86.3	0.4	0.4	P>0.05	ns
MFRP	121.7	243.4	0.1	0.1	P>0.05	ns
KCNA5	71.6	142.9	96.3	103.5	P>0.05	ns
CFAP73	68.9	136.6	0.5	0.4	P>0.05	ns
TMEM233	4.7	5.9	71.2	56.5	P>0.05	ns
FOXG1	0.2	0.1	49.3	42.1	P>0.05	ns
SIX6	0	0	39	35.3	P>0.05	ns
GABRA5	1.1	1.5	92.4	175.7	P>0.05	ns
CPLX3	9.1	13.5	152.4	134.9	P>0.05	ns
ST6GALNAC2	36.9	73.8	4.5	9.6	P>0.05	ns
TTR	**8630.4**	**17260**	**51.1**	**92.9**	**P<0.001**	***
RAX2	0.2	0.3	114.8	242.6	P>0.05	ns
NNAT	857.3	1010.7	769.9	477.9	P>0.05	ns
CLIC6	140.7	280.1	2.7	2.1	P>0.05	ns
PRAME	27.2	54.3	102.9	171.2	P>0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

as follows, the group whose most of the genes were identified in the sample was considered as dominant/major sub-group and next was the one or two genes of another subgroup. Hence mixed sub-group-specific 11 samples were processed for long-read sequencing and a total of 12 samples were processed for short-read sequencing with an aim to capture the extent of transcriptome complexity and to obtain full-length transcripts in different sub-groups of MB based on molecular classification. For this, we used long- and -short-read transcriptome sequencing and analyzed it through SQANTI software. Since, it calculates transcript length distribution, reference transcript length, number of supporting FL reads, transcript expression, reference coverage at both 3' and 5' ends, number of exons, and number of transcripts per gene (Supplementary Data 5 (PDF)).

Moreover, through SQANTI analyses we were able to categorize transcripts structurally into Full Splice Match (FSM), Incomplete Splice Match (ISM), Antisense (AS), intergenic, Novel in Catalog (NIC), Novel Not in Catalog (NNC). For each structural category, we filtered between coding and non-coding transcripts. Coding transcripts were filtered between canonical and non-canonical junctions. Further, canonical junctions (originally described in transcripts and most commonly found (like in ~99% of introns) and have GT at the donor site (just after the 5' end of the cut) and AG at the acceptor site (just before the 3' end of the cut)) were categorized as true and false RTS stages. Additionally, false predicted Nonsense-Mediated Decay (NMD) transcripts were considered for analysis. As a canonical function, true NMD may prevent translation of

Table 3b: Only RPKM (SHH *vs.* Group 4).

Gene	SHH Mean	SHH SD	Group 4 Mean	Group 4 SD	P value	Significance
SLC1A7	2.2	3.5	0.2	0.2	P>0.05	ns
RD3	2	2.8	0.1	0	P>0.05	ns
CHRM3	3.9	5.5	65.9	110.7	P>0.05	ns
TBR1	0.3	0.1	400.1	669.8	P>0.05	ns
FEZF2	0.1	0.1	3.9	6.5	P>0.05	ns
IMPG2	0.7	0.3	0.9	0.8	P>0.05	ns
SST	0.6	0.8	0	0.1	P>0.05	ns
SLC6A3	5.3	7.9	32.4	56.1	P>0.05	ns
NEUROG1	0.4	0.8	88.7	110.5	P>0.05	ns
TULP1	0.3	0.5	0.5	0.7	P>0.05	ns
PTP4A1	37	30.8	12.3	11	P>0.05	ns
GNGT1	0.3	0.5	0.1	0	P>0.05	ns
FEZF1	0.2	0	40.7	70.4	P>0.05	ns
SLC13A4	318.7	635.8	1.1	1.8	P>0.05	ns
WDR38	116.3	232.5	31.3	54.2	P>0.05	ns
SPAG6	45.7	91	119.3	199.4	P>0.05	ns
ARMC3	32.1	63.7	55.1	94.1	P>0.05	ns
DYDC2	22	43.4	85.2	146.6	P>0.05	ns
PRAP1	70.9	141.5	1.5	2.6	P>0.05	ns
FOLR1	447.3	894	33.6	29.2	P>0.05	ns
C11orf88	43.2	86.3	76.9	133.2	P>0.05	ns
MFRP	121.7	243.4	1.1	1.7	P>0.05	ns
KCNA5	71.6	142.9	82.3	108.9	P>0.05	ns
CFAP73	68.9	136.6	109.1	188.7	P>0.05	ns
TMEM233	4.7	5.9	40.5	39.9	P>0.05	ns
FOXG1	0.2	0.1	15.6	26.9	P>0.05	ns
SIX6	0	0	39.1	40.2	P>0.05	ns
GABRA5	1.1	1.5	7.3	8.4	P>0.05	ns
CPLX3	9.1	13.5	62	54.2	P>0.05	ns
ST6GALNAC2	36.9	73.8	23.7	40.9	P>0.05	ns
TTR	**8630.4**	**17260**	**123.1**	**197.2**	**P<0.001**	***
RAX2	0.2	0.3	0.4	0.5	P>0.05	ns
NNAT	857.3	1010.7	4126.6	3424.2	P>0.05	ns
CLIC6	140.7	280.1	15.4	13.9	P>0.05	ns
PRAME	27.2	54.3	2.2	2.3	P>0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

Table 3c: Only RPKM (Group 3 *vs.* Group 4).

Gene	Group 3 Mean	Group 3 SD	Group 4 Mean	Group 4	P value	Significance
SLC1A7	228.2	310.3	0.2	0.2	P>0.05	ns
RD3	61.5	99.6	0.1	0	P>0.05	ns
CHRM3	30.8	16.1	65.9	110.7	P>0.05	ns
TBR1	204.2	207.1	400.1	669.8	P>0.05	ns
FEZF2	25.2	40.4	3.9	6.5	P>0.05	ns
IMPG2	202.2	370.6	0.9	0.8	P>0.05	ns
SST	256	539.6	0	0.1	P>0.05	ns
SLC6A3	113	145.3	32.4	56.1	P>0.05	ns
NEUROG1	133	104	88.7	110.5	P>0.05	ns

TULP1	72.6	100.9	0.5	0.7	P>0.05	ns
PTP4A1	46.1	24	12.3	11	P>0.05	ns
GNGT1	28.8	45	0.1	0	P>0.05	ns
FEZF1	12.6	13.6	40.7	70.4	P>0.05	ns
SLC13A4	2.7	2.1	1.1	1.8	P>0.05	ns
WDR38	0.1	0.1	31.3	54.2	P>0.05	ns
SPAG6	2.6	2.8	119.3	199.4	P>0.05	ns
ARMC3	0.5	0.4	55.1	94.1	P>0.05	ns
DYDC2	1.3	1.4	85.2	146.6	P>0.05	ns
PRAP1	0.5	0.6	1.5	2.6	P>0.05	ns
FOLR1	1.3	1.6	33.6	29.2	P>0.05	ns
C11orf88	0.4	0.4	76.9	133.2	P>0.05	ns
MFRP	0.1	0.1	1.1	1.7	P>0.05	ns
KCNA5	96.3	103.5	82.3	108.9	P>0.05	ns
CFAP73	0.5	0.4	109.1	188.7	P>0.05	ns
TMEM233	71.2	56.5	40.5	39.9	P>0.05	ns
FOXG1	49.3	42.1	15.6	26.9	P>0.05	ns
SIX6	39	35.3	39.1	40.2	P>0.05	ns
GABRA5	92.4	175.7	7.3	8.4	P>0.05	ns
CPLX3	152.4	134.9	62	54.2	P>0.05	ns
ST6GALNAC2	4.5	9.6	23.7	40.9	P>0.05	ns
TTR	51.1	92.9	123.1	197.2	P>0.05	ns
RAX2	114.8	242.6	0.4	0.5	P>0.05	ns
NNAT	**769.9**	**477.9**	**4126.6**	**3424.2**	**P<0.001**	***
CLIC6	2.7	2.1	15.4	13.9	P>0.05	ns
PRAME	102.9	171.2	2.2	2.3	P>0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

Table 4a: Reported transcript markers in our data in RPKM (SHH *vs.* Group 3).

Gene	SHH Mean	SHH SD	Group 3 Mean	Group 3 SD	P value	Significance
DKK2	1.8	2	0.6	0.5	P > 0.05	ns
EMX2	0.7	0.5	4	4.9	P > 0.05	ns
GAD1	0.9	1.2	5.5	5.3	P > 0.05	ns
TNC	34.4	41.9	12.6	9.4	P > 0.05	ns
WIF1	8.2	16.1	0.9	1	P > 0.05	ns
ATOH1	45.8	44.6	0	0	P > 0.05	ns
EYA1	42.2	23.7	0.2	0.3	P > 0.05	ns
HHIP	41.4	31.3	1.9	1.6	P > 0.05	ns
PDLIM3	295.8	188.7	11.4	12	P > 0.05	ns
SFRP1	**424.7**	**233.3**	**13.9**	**14.6**	**P<0.01**	**
GABRA5	1.2	1.5	105.3	200.2	P > 0.05	ns
IMPG2	0.7	0.3	251	408.9	P > 0.05	ns
MAB21L2	42.1	76	59.6	50.5	P > 0.05	ns
NPR3	6.7	11.2	16.2	17.7	P > 0.05	ns
NRL	**1.8**	**0.9**	**533.6**	**868.5**	**P<0.001**	***
EOMES	3.3	4.9	179.7	162	P > 0.05	ns
KCNA1	3.3	6.4	16.8	16.5	P > 0.05	ns
KHDRBS2	1.4	2.5	14.7	12	P > 0.05	ns
OAS1	8	1.6	11.5	9.6	P > 0.05	ns
RBM24	1.7	3.1	3.9	3.6	P > 0.05	ns
UNC5D	3.9	6.3	10.1	15.2	P > 0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

Table 4b: Reported transcript markers in our data in RPKM (SHH *vs.* Group 4).

Gene	SHH Mean	SHH SD	Group 4 Mean	Group 4 SD	P value	Significance
DKK2	1.8	2	0.7	1	P > 0.05	ns
EMX2	0.7	0.5	0.5	0.6	P > 0.05	ns
GAD1	0.9	1.2	0.2	0.2	P > 0.05	ns
TNC	34.4	41.9	14.2	22.3	P > 0.05	ns
WIF1	8.2	16.1	2.3	3.8	P > 0.05	ns
ATOH1	45.8	44.6	0.1	0.2	P > 0.05	ns
EYA1	42.2	23.7	2	2.7	P > 0.05	ns
HHIP	41.4	31.3	2	1.9	P > 0.05	ns
PDLIM3	**295.8**	**188.7**	**22.2**	**37.5**	**P<0.001**	***
SFRP1	**424.7**	**233.3**	**6.3**	**5.5**	**P<0.001**	***
GABRA5	1.2	1.5	7.3	8.3	P > 0.05	ns
IMPG2	0.7	0.3	0.9	0.8	P > 0.05	ns
MAB21L2	42.1	76	56.2	69.6	P > 0.05	ns
NPR3	6.7	11.2	0.9	1.4	P > 0.05	ns
NRL	1.8	0.9	10.9	13.8	P > 0.05	ns
EOMES	3.3	4.9	45.3	76.8	P > 0.05	ns
KCNA1	3.3	6.4	85.7	105.2	P > 0.05	ns
KHDRBS2	1.4	2.5	32.7	56.6	P > 0.05	ns
OAS1	8	1.6	13.9	19.2	P > 0.05	ns
RBM24	1.7	3.1	7.5	6.8	P > 0.05	ns
UNC5D	3.9	6.3	73.7	127.4	P > 0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

Table 4c: Reported transcript markers in our data in RPKM (Group 3 *vs.* Group 4).

Gene	Group 3 Mean	Group 3 SD	Group 4 Mean	Group 4 SD	P value	Significance
DKK2	0.6	0.5	0.7	1	P > 0.05	ns
EMX2	4	4.9	0.5	0.6	P > 0.05	ns
GAD1	5.5	5.3	0.2	0.2	P > 0.05	ns
TNC	12.6	9.4	14.2	22.3	P > 0.05	ns
WIF1	0.9	1	2.3	3.8	P > 0.05	ns
ATOH1	0	0	0.1	0.2	P > 0.05	ns
EYA1	0.2	0.3	2	2.7	P > 0.05	ns
HHIP	1.9	1.6	2	1.9	P > 0.05	ns
PDLIM3	11.4	12	22.2	37.5	P > 0.05	ns
SFRP1	13.9	14.6	6.3	5.5	P > 0.05	ns
GABRA5	105.3	200.2	7.3	8.3	P > 0.05	ns
IMPG2	251	408.9	0.9	0.8	P > 0.05	ns
MAB21L2	59.6	50.5	56.2	69.6	P > 0.05	ns
NPR3	16.2	17.7	0.9	1.4	P > 0.05	ns
NRL	**533.6**	**868.5**	**10.9**	**13.8**	**P<0.01**	**
EOMES	179.7	162	45.3	76.8	P > 0.05	ns
KCNA1	16.8	16.5	85.7	105.2	P > 0.05	ns
KHDRBS2	14.7	12	32.7	56.6	P > 0.05	ns
OAS1	11.5	9.6	13.9	19.2	P > 0.05	ns
RBM24	3.9	3.6	7.5	6.8	P > 0.05	ns
UNC5D	10.1	15.2	73.7	127.4	P > 0.05	ns

*Bonferroni posttests test; P value <0.05 (significant); P<0.01 (moderately significant); P<0.001 (highly significant); ns (non-significance)

mutant mRNAs harboring Premature Termination Codons (PTCs) by targeting them for degradation [25]. Results suggested that known canonical splice junctions were common in FSM and ISM. Additionally, in NIC, NNC, AS and IG categories the identified novel transcripts possessed known canonical junctions, the percentage of non-canonical junctions was negligible suggestive of less involvement of immunoglobulins. As most of the non-canonical splice sites have been observed during immunoglobulin expression [26].

Furthermore, negative values indicated upstream of annotated PolyA site/TTS, while downstream of the annotated-TSS site described alternative polyadenylation and transcription initiation in FSM and ISM may lead to the formation of intergenic and antisense transcripts. Alternative polyadenylation and transcription initiation sizeable contribution to cellular phenotypes in normal human physiology has been reported by Reyes and Huber in 2018. Alternative start and termination sites of transcription driven most transcript isoform differences across human tissues [27].

Most importantly, a significant change was observed between AS-coding ($p \leq 0.037$)/noncoding ($p \leq 0.039$) transcripts in control *vs.* all MB samples. The coding sense/AS-transcript summary has been presented in Table 3. Importantly, SHH-MB specific EZR antisense (EZR-AS1; lncRNA) has also been identified in both human Esophageal Squamous Cell Carcinoma (ESCC) tissues and cell lines and showed its correlation with EZR expression. Both *in vivo* and *in vitro* studies revealed that EZR-AS1 promoted ESCC cell migration through up-regulation of EZR expression. Mechanistically, the antisense lncRNA EZR-AS1 forms a complex with RNA polymerase II to activate the transcription of EZR. Moreover, EZR-AS1 recruit SET and MYN-Domain containing 3 (SMYD3) to GC-rich binding region downstream of the EZR promoter, causing the binding of SMYD3 and local enrichment of H3K4me3 with subsequent enhanced EZR transcription and expression driving to tumor aggressiveness, drug resistance and poor prognosis [28].

Miele et al. [29] suggested EZH2, histone methyltransferase as druggable target in SHH-MB and suggested promotable MC3629 as EZH2 inhibitor (EZH2i) and SYMD3 inhibitors [30]. SHH/G3 specific CENPX/MHF2 are involved in the Fanconi anemia pathway in DNA repair and maintenance of genome integrity. Recently, the importance of Fanconi anemia genes including CENPX/MHF2 in colorectal cancer act as a promising prognostic and predictive biomarkers for disease management and treatment [31]. G4 specific COL18A1-sense transcript has been reported as a direct target of miR-9. However, COL18A1-AS1 (COL18A1 Antisense RNA 1) transcript is an RNA gene, and is affiliated with the lncRNA class. It has been reported that lncRNA sequester miRNA, hence COL18A1 antisense RNA 1 may sequester MYC and OCT4 directed miR-9 and may dysregulate the expression of COL18A1-sense transcript [32]. Hence, antisense oligonucleotide-based [33] or transcriptional inhibition through classical CRISPR/Cas9 to delete region of interest in lncRNA loci may be used as a lncRNA for SHH/G3/G4 MB based therapeutics can be used for treatment of MB [34].

Through differential analysis of all subtypes of MB *vs.* control, considering Bonferroni less than <0.05, we identified 35 transcripts (Supplementary Table 7), which categorized samples as mixed subgroups and more genes were identified which were common between both G3 and G4/SHH and G3/G4. Importantly, we identified highly significantly up-regulated marker genes to categorize three subgroups of MB. First and foremost was TTR up-regulated in SHH/G3/G4 except for WNT. Transthyretin (TTR) helps in the transport of both thyroxine and retinol (vitamin A) from the blood to the choroid plexus, the cerebrospinal fluid, and the brain (GeneCard). As TTR was overexpressed in all the three subtypes SHH/G3/G4 of MB, hence can be used as a cerebrospinal fluid marker for prognosis. TTR is also involved in the canonical retinoid cycle in rods (twilight vision); retinoid cycle disease events; diseases associated with visual transduction; non-integrin membrane-ECM interactions-retinoid metabolism and transport; metabolism of fat-soluble vitamins may corroborate with the eye disorders in MB; previously reported SHH specific PDLIM3 and SFRP1 were as marker gene in our study also. PDLIM3 Actin-associated LIM Protein (ALP), also known as PDZ and LIM domain protein 3 is a protein that in humans is encoded by the PDLIM3 gene. ALP is highly expressed in cardiac and skeletal muscle, where it localizes to Z-discs and intercalated discs) and plays a role in muscle differentiation (GeneCard). PDLIM3 splicing was regulated during development and in a tissue-specific manner, and aberrant PDLIM3 splicing was associated with splicing of SERCA1 in Myotonic dystrophy type 1 patient [35], hence suggestive of myotonic involvement in MB. Further, SFRP1 (Secreted-Frizzled-Related-Protein) family of WNT inhibitors function and act as novel tumor suppressor genes epigenetically silenced in MB [36] and is significantly increased in the brain and cerebrospinal fluid of patients with Alzheimer's disease [37]. However, increase in SHH tumors with respect to G3 and G4 very well differentiates it from SHH MB and epigenetic silencing in G3 and 4 suggests pro-tumorigenesis in G3/G4 [38]. Additionally, PDLIM3 and SFRP1 were also validated in reported signatures including (Schwalbe et al. [39] 8-gene SHH signature-BCHE, GLI1, ITIH2, MICAL1, PDLIM3, PTCH2, RAB33A, and SFRP1), Northcott et al. [21] a 5-gene SHH signature (PDLIM3, EYA1, HHIP, ATOH1, and SFRP1) and Amakye et al. [40] unique 5-gene SHH signature including GLI1, SPHK1, SHROOM2, PDLIM3, and OTX2 was used for targeted treatment for SHH-dependent medulloblastoma [40,41].

Neural Retina-specific Leucine zipper protein (NRL) increased in G-3 compared to SHH and G-4, hence highly G-3 specific. It acts as a transcriptional activator by binding to promoter regions of several rod-specific genes, including Rhodopsin (RHO), PDE6B [42,43], and BCL-XL [44]. It also functions as a transcriptional coactivator by stimulating transcription factors like CRX and NR2E3 [42,44]. Hence, targeting NRL pathway as a therapeutic strategy may be preferred as suggested by Moore et al. [45], to react Retinitis pigmentosa.

Neuronatin (NNAT) was up-regulated in G-4 compared to SHH and G-3. It is a proteolipid involved in regulating the control of ion channels in mammalian brain development located on Chromosome 20 [46]. It also increases the differentiation of pluripotent stem cells into cells with a neural fate by increasing their calcium levels- miRNA-708 mediated [47,48]. UniProtKB/Swiss-Prot summary for NNAT gene suggests NNAT participates in the maintenance of segment identity in the hindbrain and pituitary development, and maturation or maintenance of the overall structure of the nervous system (UniProtKB/Swiss-Prot Summary for NNAT Gene). Diseases associated with NNAT include megalencephalic leukoencephalopathy with subcortical cysts 1 and myopathy, centronuclear, 6, with fiber-type disproportion. NNAT is maternally imprinted and only expressed on the paternal allele in normal adults i.e., exclusively expressed only in the father and results in the expression of only one copy of NNAT in the offspring. However, defect in inheriting NNAT may be more vulnerable to the negative

effects of mutations as it is haploid in nature due to the absence of no substitute allele [49] or dysregulation in methylation pattern in maternal allele as shown in ovarian cancer may dysregulate functions in the control of ion channels and pluripotent stem cells and alter the development of the nervous system [50]. The imprint control region loses methylation resulting in aberrant cell growth and a worse prognosis has been reported in embryonic neoplasms such as Wilms' tumor [51], aggressive variants of human glioblastoma multiforme medulloblastoma [52-54], and neuro-blastoma [55]. Neuronatin is also highly expressed in large cell lung cancer [56,57]. Contrarily, unexpected hypermethylation of the Neuronatin locus has also been associated with pediatric leukemia [58] and is also involved in whole-animal glucose homeostasis [59,60].

Conclusion

Descriptive analysis of transcriptome complexity and full-lengthiness was made easy by SQANTI software and molecularly categorized samples; the identified antisense transcripts can be used as a target for developing anti-antisense synthetic peptides/ transcriptional inhibition through classical CRISPR/Cas9 to delete region of interest in lncRNA loci which in combination with SHH/ NRL/NNAT inhibitors may be used as a targeted approach to control MB. Further, for precise medicine, the following approach can be used obtaining the tumor material just after surgery, molecular characterization of the tumor material within three weeks and after clinical correlation, stratify the patients into low- standard- and high-risk and relates to different biology through our transcript based customized panel. Through this approach will be able to understand the prognosis of disease following prognostic markers like up-regulated TTR in SHH/G3/G4 MB; increased levels of PDLIM3 and SFRP1 in SHH specific MB; Increased NRL-G-3 specific marker regulates the expression of several rod-specific genes; Enhanced expression of NNAT suggests G-4 MB as NNAT imprinting disease. Importantly, TTR identified in cerebrospinal fluid may be used as a marker to understand the prognosis of the disease.

Funding

This project has been funded by Department of Science and Technology as Early Career Research Award, Department of Science and Technology- ECR/2016/001616*:* IsoSeq analysis and functional annotation of molecularly classified Medulloblastoma on PacBio RSII platform.

References

1. Gröbner S, Worst B, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321-7.
2. Chen YD, Zhang N, Qiu XG, Yuan J, Yang M. LncRNA CDKN2BAS rs2157719 genetic variant contributes to medulloblastoma predisposition. J Gene Med. 2018;20(1).
3. Ladomery M. Aberrant alternative splicing is another hallmark of cancer. Int J Cell Biol. 2013;2013:463786.
4. Liu S, Cheng C. Alternative RNA splicing and cancer. Wiley Interdiscip Rev RNA. 2013;4(5):547-66.
5. Chen J, Weiss WA. Alternative splicing in cancer: Implications for biology and therapy. Oncogene. 2015;34(1):1-14.
6. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, et al. The human pancreatic islet transcriptome: Expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet. 2012;8(3):e1002552.
7. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43(12):e78.
8. Yang YY, Yin GL, Darnell RB. The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. Proc Natl Acad Sci U S A. 1998;95(22):13254-9.
9. D'Souza I, Poorkaj P, Hong M, Nochlin D, Lee VM, Bird TD, et al. Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc Natl Acad Sci USA. 1999;96(10):5598-603.
10. Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, et al. A muscle blind knockout model for myotonic dystrophy. Science. 2003;302(5652):1978-80.
11. Ladd AN. CUG-BP, Elav-like family (CELF)-mediated alternative splicing regulation in the brain during health and disease. Mol Cell Neurosci. 2013;56:456-64.
12. Lee JA, Damianov A, Lin CH, Fontes M, Parikshak NN, Anderson ES, et al Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. Neuron. 2016;89(1):113-28.
13. Steijger T, Abril JF, Engström PG, Kokocinski F, Hubbard TJ; RGASP Consortium, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013;10(12):1177-84.
14. Tilgner H, Raha D, Habegger L, Mohiuddin M, Gerstein M, Snyder M. Accurate identification and analysis of human mRNA isoforms using deep long read sequencing. G3 (Bethesda). 2013;3(3):387-97.
15. Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single-molecule long-read transcriptome. Proc Natl Acad Sci U S A. 2014;111(27):9869-74.
16. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol. 2013;31(11):1009-14.
17. Li S, Yamada M, Han X, Ohler U, Benfey PN. High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. Dev Cell. 2016;39(4):508-22.
18. Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, Chen F, et al. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. Nat Biotechnol. 2015;33(7):736-42.
19. Oikonomopoulos S, Wang YC, Djambazian H, Badescu D, Ragoussis J. Benchmarking of the oxford nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. Sci Rep. 2016;6:31602.
20. Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B, et al. Intertumoral heterogeneity within medulloblastoma subgroups. Cancer Cell. 2017;31(6):737-54.e6.
21. Northcott PA, Shih DJ, Remke M, Cho YJ, Kool M, Hawkins C, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol. 2012;123(4):615-26.
22. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, De Pristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012;13:375.
23. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, et al. SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28(3):396-411.
24. Singh N, Sahu DK, Mishra A, Agarwal P, Goel MM, Chandra A, et al. Multiomics approach showing genome-wide copy number alterations and differential gene expression in different types of North-Indian pediatric

brain tumors. Gene. 2016;576(2 Pt 2):734-42.

25. Karousis ED, Mühlemann O. Nonsense-mediated mRNA decay begins where translation ends. Cold Spring Harb Perspect Biol. 2019;11(2):a032862.

26. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. Nucleic Acids Res. 1987;15(17):7155-74.

27. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. 2018;46(2):582-92.

28. Zhang XD, Huang GW, Xie YH, He JZ, Guo JC, Xu XE, et al. The interaction of lncRNA EZR-AS1 with SMYD3 maintains overexpression of EZR in ESCC cells. Nucleic Acids Res. 2018;46(4):1793-809.

29. Miele E, Valente S, Alfano V, Silvano M, Mellini P, Borovika D, et al. The histone methyltransferase EZH2 as a druggable target in SHH medulloblastoma cancer stem cells. Oncotarget. 2017;8(40):68557-70.

30. Bernard BJ, Nigam N, Burkitt K, Saloura V. SMYD3: A regulator of epigenetic and signaling pathways in cancer. Clin Epigenetics. 2021;13(1):45.

31. Parsa FG, Nobili S, Karimpour M, Aghdaei HA, Nazemalhosseini-Mojarad E, Mini E. Fanconi anemia pathway in colorectal cancer: A novel opportunity for diagnosis, prognosis and therapy. J Pers Med. 2022;12(3):396.

32. Chen X, Yang F, Zhang T, Wang W, Xi W, Li Y, et al. MiR-9 promotes tumorigenesis and angiogenesis and is activated by MYC and OCT4 in human glioma. J Exp Clin Cancer Res. 2019;38(1):99.

33. Dean NM, Bennett CF. Antisense oligonucleotide-based therapeutics for cancer. Oncogene. 2003;22(56):9087-96.

34. Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med. 2018;24(3):257-77.

35. Ohsawa N, Koebis M, Suo S, Nishino I, Ishiura S. Alternative splicing of PDLIM3/ALP, for α-actinin-associated LIM protein 3, is aberrant in persons with myotonic dystrophy. Biochem Biophys Res Commun. 2011;409(1):64-9.

36. Kongkham PN, Northcott PA, Croul SE, Smith CA, Taylor MD, Rutka JT. The SFRP family of WNT inhibitors function as novel tumor suppressor genes epigenetically silenced in medulloblastoma. Oncogene. 2010;29(20):3017-24.

37. Esteve P, Rueda-Carrasco J, Inés Mateo M, Martin-Bermejo MJ, Draffin J, Pereyra G, et al. Elevated levels of secreted-frizzled-related-protein 1 contribute to Alzheimer's disease pathogenesis. Nat Neurosci. 2019;22(8):1258-68.

38. Regel I, Eichenmüller M, Mahajan UM, Hagl B, Benitz S, Häberle B, et al. Downregulation of SFRP1 is a protumorigenic event in hepatoblastoma and correlates with beta-catenin mutations. J Cancer Res Clin Oncol. 2020;146(5):1153-67.

39. Schwalbe EC, Lindsey JC, Nakjang S, Crosier S, Smith AJ, Hicks D, et al. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: A cohort study. Lancet Oncol. 2017;18(7):958-971.

40. Amakye D, Robinson D, Rose K, Cho J, Ligon KL, Sharp T, et al. The predictive value of a 5-gene signature as a patient pre-selection tool in medulloblastoma for hedgehog pathway inhibitor therapy. Cancer Res. 2012;72(8suppl 1):4818.

41. Kieran MW, Hargrave D, Wen PY. A phase 3, multicenter, open-label, randomized, controlled study of the efficacy and safety of oral Sonidegib (LDE225) versus temozolomide in patients with hedgehog pathway-activated relapsed medulloblastoma [abstract NO-069]. Neuro Oncol. 2013;15(suppl 3).

42. Kanda A, Friedman JS, Nishiguchi KM, Swaroop A. Retinopathy mutations in the bZIP protein NRL alter phosphorylation and transcriptional activity. Hum Mutat. 2007;28(6):589-98.

43. Hernan I, Gamundi MJ, Borràs E, Maseras M, García-Sandoval B, Blanco-Kelly F, et al. M96T variant of NRL and shRNA-based suppression and replacement of NRL mutants associated with autosomal dominant retinitis pigmentosa. Clin Genet. 2012;82(5):446-52.

44. Garancher A, Lin CY, Morabito M, Richer W, Rocques N, Larcher M, et al. NRL and CRX Define photoreceptor identity and reveal subgroup-specific dependencies in medulloblastoma. Cancer Cell. 2018;33(3):435-49.e6.

45. Moore SM, Skowronska-Krawczyk D, Chao DL. Targeting of the NRL pathway as a therapeutic strategy to treat retinitis pigmentosa. J Clin Med. 2020;9(7):2224.

46. Shusuke N, Tianzhang Y, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, et al. "DNA Methylation signatures in development and aging of the human prefrontal cortex". Am J Hum Genet. 2012;90(2):260-72.

47. Joseph RM. Neuronatin gene: Imprinted and misfolded: Studies in Lafora disease, diabetes and cancer may implicate NNAT-aggregates as a common downstream participant in neuronal loss. Genomics. 2014;103(2-3):183-8.

48. Ryu S, Mcdonnell, KA, Choi H, Gao D, Hahn M, Joshi N, et al. Suppression of miRNA-708 by polycomb group promotes metastases by calcium-induced cell migration. Cancer Cell. 2013;23(1):63-76.

49. Jirtle RL, Weidman JR. Imprinted and more equal American Scientist. 2007;95:143-9.

50. Reid BM, Fridley BL. DNA methylation in ovarian cancer susceptibility. Cancers (Basel). 2021;13(1):108.

51. Hubertus J, Zitzmann F, Trippel F, Müller-Höcker J, Stehr M, von Schweinitz D, et al. Selective methylation of CpGs at regulatory binding sites controls NNAT expression in Wilms tumors. PLoS One. 2013;8(6):e67605.

52. Xu DS, Yang C, Proescholdt M, Bründl E, Brawanski A, Fang X, et al. Neuronatin in a subset of glioblastoma multiforme tumor progenitor cells is associated with increased cell proliferation and shorter patient survival. PLoS One. 2012;7(5):e37811.

53. Yokota N, Mainprize TG, Taylor MD, Kohata T, Loreto M, Ueda S, et al. Identification of differentially expressed and developmentally regulated genes in medulloblastoma using suppression subtraction hybridization. Oncogene. 2004;23(19):3444-53.

54. Siu IM, Bai R, Gallia GL, Edwards JB, Tyler BM, Eberhart CG, et al. Coexpression of neuronatin splice forms promotes medulloblastoma growth. Neuro Oncol. 2008;10(5):716-24.

55. Higashi M, Tajiri T, Kinoshita Y, Tatsuta K, Souzaki R, Maehara Y, et al. High expressions of neuronatin isoforms in favorable neuroblastoma. J Pediatr Hematol Oncol. 2007;29(8):551-6.

56. Zhong S, Fields CR, Su N, Pan YX, Robertson KD. Pharmacologic inhibition of epigenetic modifications, coupled with gene expression profiling, reveals novel targets of aberrant DNA methylation and histone deacetylation in lung cancer. Oncogene. 2007;26(18):2621-34.

57. Okubo C, Minami Y, Tanaka R, Uchihara T, Anami Y, Furuya S, et al. Analysis of differentially expressed genes in neuroendocrine carcinomas of the lung. J Thorac Oncol. 2006;1(8):780-6.

58. Kuerbitz SJ, Pahys J, Wilson A, Compitello N, Gray TA. Hypermethylation of the imprinted NNAT locus occurs frequently in pediatric acute leukemia. Carcinogenesis. 2002;23(4):559-64.

59. Millership SJ, Da Silva Xavier G, Choudhury AI, Bertazzo S, Chabosseau P, Pedroni SM, et al. Neuronatin regulates pancreatic β cell insulin content and secretion. J Clin Invest. 2018;128(8):3369-81.

60. Zang L, Shimada Y, Nakayama H, Chen W, Okamoto A, Koide H, et al. Therapeutic silencing of centromere protein X ameliorates hyperglycemia in zebrafish and mouse models of type 2 diabetes mellitus. Front Genet. 2019;10:693.