Published in final edited form as:

Acta Neuropathol. 2012 April ; 123(4): 485–499. doi:10.1007/s00401-012-0959-7.

Subgroup Specific Alternative Splicing in Medulloblastoma

Adrian M Dubuc^{#1,2,3}, A. Sorana Morrissy^{#1,2}, Nanne K Kloosterhof^{4,5}, Paul A Northcott^{1,2,3}, Emily PY Yu⁶, David Shih^{1,2,3}, John Peacock^{1,2,3}, Wieslawa Grajkowska⁷, Timothy van Meter⁸, Charles G Eberhart⁹, Stefan Pfister¹⁰, Marco A Marra¹¹, William A Weiss¹², Stephen W Scherer^{13,14}, James T Rutka^{1,3}, Pim J French⁴, and Michael D Taylor^{1,2,3}

¹ Division of Neurosurgery, Arthur & Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. ² Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. ³ Department of Laboratory Medicine & Pathobiology, University of Toronto, Toronto, Ontario, Canada. ⁴ Department of Neurology, Erasmus MC, Rotterdam, The Netherlands ⁵ Department of Paediatric Oncology and Hematology, Erasmus MC—Sophia Children's Hospital, Rotterdam, Netherlands ⁶ Program in Biology and Pharmacology, University of Western Ontario, London, Ontario, Canada ⁷ Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland. ⁸ Department of Neurosurgery, Medical College of Virginia, Richmond, Virginia, USA. ⁹ Department of Pathology, Johns Hopkins University, Baltimore, Maryland, USA ¹⁰ German Cancer Research Centre, University of Heidelberg, Heidelberg, Germany. ¹¹ British Columbia Cancer Agency, Genome Science Centre, Vancouver, British Columbia, Canada ¹² Helen Diller Family Comprehensive Cancer Centre, University of California, San Francisco, California, United States of America. ¹³ The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. ¹⁴ Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

These authors contributed equally to this work.

Abstract

Medulloblastoma is comprised of four distinct molecular variants: WNT, SHH, Group 3, and Group 4. We analyzed alternative splicing usage in 14 normal cerebellar samples and 103 medulloblastomas of known subgroup. Medulloblastoma samples have a statistically significant increase in alternative splicing as compared to normal fetal cerebella (2.3-times; $P < 6.47 \times 10^{-8}$). Splicing patterns are distinct and specific between molecular subgroups. Unsupervised hierarchical clustering of alternative splicing events accurately assigns medulloblastomas to their correct subgroup. Subgroup-specific splicing and alternative promoter usage was most prevalent in Group 3 (19.4%) and SHH (16.2%) medulloblastomas, while observed less frequently in WNT (3.2%), and Group 4 (9.3%) tumors. Functional annotation of alternatively spliced genes reveals over-representation of genes important for neuronal development. Alternative splicing events in medulloblastoma may be regulated in part by the correlative expression of antisense transcripts, suggesting a possible mechanism affecting subgroup specific alternative splicing. Our results identify additional candidate markers for medulloblastoma subgroup affiliation, further support the existence of distinct subgroups of the disease, and demonstrate an additional level of transcriptional heterogeneity between medulloblastoma subgroups.

Keywords

medulloblastoma; alternative splicing; neuronal development; molecular subgroup; pediatric cancer

Introduction

Medulloblastoma (MB) is the most common malignant brain tumor in children [12, 41, 45] and has recently been demonstrated to exhibit considerable inter-tumoral heterogeneity [7, 14]. Recent publications have dissected medulloblastoma at a molecular level into four distinct variants – namely WNT, SHH, Group 3 and Group 4 [9, 27, 41, 43, 53]. These subgroups differ in their epidemiology, copy number profiles, transcriptional networks, mutational spectra, and clinical characteristics [42, 45, 48]. The study of subgroup specific gene expression has assisted in the identification of cells of origin for WNT [20] and SHH medulloblastomas [63]. Subgroup specific targeted therapy is imminent, with promising preliminary responses to SHH-pathway inhibitors in humans and mice [8, 49]. However, nearly half of all MBs are represented by Group 3 or 4 tumors [13, 43] with dismal overall survival and in which the molecular mechanisms driving tumorigenesis remain largely unknown [21]. The optimal mechanism for ‘real time’ assignment of subgroup affiliation in the setting of a clinical trial is not currently settled. To further understand the transcriptional dissimilarity between subgroups we undertook an analysis of alternative splicing and promoter use in medulloblastoma.

Alternative splicing of pre-mRNA is a dynamic mechanism that adds complexity to the human transcriptome, thereby significantly increasing the diversity of expressed proteins [28]. Transcriptional selection of splice sites usage occurs through the processes of: exon skipping, alternative transcriptional start site usage, intron retention, and alternative polyadenylation sites, collectively referred to in the current manuscript as alternative splicing [28, 35, 40, 62]. One or more of these alternative splicing mechanisms are estimated to affect 75% [19] to 92% [62] of all genes in the human genome. Tight regulation of normal, tissue-specific, and developmental splicing is mediated by a complex network of RNA-binding proteins that recognize exonic or intronic cis-regulatory elements, enhancing or repressing inclusion of an exon in a transcript [19]. Recent evidence has also shown that transcription of an overlapping gene, encoded on the opposite DNA strand (antisense transcription) can affect splicing outcomes [57, 58].

Alternative splicing has been reported to be “cancer-specific”, producing protein isoforms that favor cellular growth, or metastasis [4, 11, 36, 59, 60]. Research efforts to target cancer specific isoforms are ongoing [2, 23], and in notable cases have led to clinical trials addressing the efficacy of isoform-specific monoclonal antibodies [54]. A limited number of studies detailing medulloblastoma-restricted isoform-expression have been conducted. Notable findings include alternative splicing of *ERRB4* [17], *PTC* [56], and *GLII* [65], which impact critical signaling and developmental pathways relevant to the pathogenesis of medulloblastoma. We undertook a comprehensive investigation of alternative splicing across medulloblastoma subgroups in a large cohort of primary tumours (n=103). Using data from the Affymetrix exon array platform, we identified multiple, recurrent, subgroup-specific alternative start site, and exon dropping events. Furthermore, we identified sense-antisense (S-AS) transcription, with subgroup specific expression of antisense transcripts correlating with alternative splicing in medulloblastoma, which may represent a putative mechanism contributing to isoform variability. Our data further highlights the transcriptional dissimilarity between subgroups, suggests additional markers for assignment of subgroup

affiliation, provide additional tools for cell of origin studies, and provides a hypothesis based on SAS transcription that may explain patterns of subgroup specific alternative splicing.

Materials and Methods

Tissue Samples and RNA preparation

Primary medulloblastoma (n=103) and normal cerebella (fetal - n=9, adult - n=5) samples were profiled on Affymetrix Genechip Human Exon 1.0ST Arrays. Samples, obtained in accordance with Hospital for Sick Children (Toronto, Canada) Research Ethic Board, were snap frozen with liquid nitrogen at local host institutions and stored at -80°C . RNA was extracted using standard TRIzol (Invitrogen) protocol and quantification was performed using a Nanodrop ND-1000 Spectrophotometer. The quality of RNA was assessed on an Agilent 2100 Bioanalyzer by The Toronto Centre for Applied Genomics (TCAG, Toronto, Canada).

Expression Profiling and Molecular Subgrouping

As previously described in Northcott *et al.*, 2009 [41], Affymetrix Genechip Human Exon 1.0ST Array were processed at the TCAG (Toronto, Canada). Molecular subgroups were established as previously described [43].

Subgroup Specific Splice Variant Detection

Pattern-Based Correlation

Pattern-based Correlation (PAC) splice variant values, which represent the theoretical expression of each probe in relation to gene expression levels, were generated for each probe set and used to identify putative alternative-splicing. PAC values were calculated for each probe set in each sample except for samples where its meta-probe set level was < 8.5 . In this manner, PAC values are calculated only in samples where the meta-probe set (a measure for gene-level expression) is expressed. We further focused only on probe sets whose expression is correlated with its meta- probe set ($R^2 > 0.64$) as a measure to filter out poor performing and cross-hybridizing probe sets. ANOVA analysis was performed to identify differentially expressed splice variants between molecular subgroups of MBs. Differentially expressed splice variants were then further selected based on the degree of alternative-splicing (PAC values > 2 criteria, corresponding to 4-fold difference in relative expression), see also French *et al*, 2007 [18]. Our analysis detected distinct changes in intra-transcript levels, revealing 1986 probe sets mapping to 1286 genes demonstrating 1 probe with a statistically significant PAC-score, suggestive of alternative-splicing.

Splice Index (SI)

As an alternative bioinformatic approach, we re-processed the data to generate Splice Index values for each probe set. First, to filter out probe sets with poor performance or low signal, we calculated the median intensity across samples for each of the 287,189 core probe sets on the array. 161,720 probe sets with an average intensity above the median of these values (6.58) were retained for further analysis. Next, probe sets were mapped to Ensembl genes (hg18). A total of 12,209 genes that (1) were represented by a minimum of 6 core probe sets on the array, and (2) had a minimum of 20% of probe sets above the filtering threshold, were retained for further analysis. A splice index (SI) value was calculated for each probe set in these 12,209 genes as previously described [37]. Briefly, in each sample, probe set intensity values were normalized by the corresponding gene expression value. The resulting SI value indicates whether the exon is included in the transcript (higher SI) or excluded (lower SI). SI values were first filtered to retain probe sets with a high dynamic range across samples, as described next. For each probe set, we calculated the difference between the 5th percentile

and 95th percentile of the SI values across the 117 samples. The top 5% of probe sets (7464) with the largest 95th-5th percentile differences were selected as a probable target of alternative splicing. To determine the number of alternative splicing events across each sample, a z-score was calculated for each probe set. Samples with probe sets whose z-score fell two standard deviations away from the mean ($-2 \leq \text{z-score} \leq 2$) were identified as alternative splicing events.

Comparing alternatively spliced probe sets identified by PAC or SI predictions a collective splice series was generated, whereas probe sets/genes identified by both algorithms were defined as the consensus splice series. The collective splice series was used to identify subgroup-specific alternative splicing events and hypersplice medulloblastomas whereas the consensus splice series permitted the identification of hallmark alternative splicing events prevalent in each molecular subgroup.

Unsupervised Clustering of Splice Index values

Splice Index (SI) values for medulloblastoma (n=103) and normal cerebella (n=14) samples were used for clustering analysis. We performed unsupervised Hierarchical Clustering (HCL) using Pearson's Correlation as a distance metric with bootstrapping analysis (100 iterations) with all 7464 probe sets using TM4 Microarray SoftwareSuite (MeV v4.6, Dana-Farber Cancer Institute, Boston). We repeated this analysis with the top 50% of probe sets (3732) with the highest standard deviation, as well as the top 25% of probe sets (1866), 13.4% (1000 probe sets) and 6.25% of probe sets (467). We identified the strongest support for clustering using 1000 probe sets which identified 6 core clusters, composed of 2 normal subgroups (fetal and adult cerebella) and 4 medulloblastoma subgroups. Non-negative matrix factorization (NMF) (Dana-Farber Cancer Institute, Boston) was used as second, unsupervised clustering algorithm. Using both the top 1000 probe sets with the highest standard deviation, as well as all 7464 probe sets, we determined the cophenetic correlation for k=2 to k=8 subgroups. We identified the strongest support for k=7 for both the filtered (1000 probe sets 0.9629) and unfiltered (7464 probe sets 0.8801), producing 2 normal clusters and 5 medulloblastoma subgroups.

qRT-PCR Validation of Alternative Start Sites and Exon Dropping Events

Validation of splice isoforms was performed using qRT-PCR. In brief, cDNA was synthesized from RNA using Superscript III First-Strand Synthesis supermix (Invitrogen). Five-hundred (500ng) nanograms of RNA was incubated with 2×First Strand Reaction mix (Invitrogen) and random hexamers (50ng/μL) for 10-minutes at 25°C and then 1-hour at 50°C prior to heat-inactivation of the enzyme mixture at 85°C for 5-minutes. Primers designed using Primer Express software were generated targeting regions at the 5' and 3' end of the transcript. Primer sequences can be found in supplemental data (Table S5). Fifty-nanograms (50ng) of cDNA was profiled on ABI Step One qRT-PCR instrumentation using SYBR green. A transcript ratio was calculated as the fold-change difference between the 5' versus 3' end. All transcript ratios were normalized to pooled fetal cerebellar cDNA.

Bioinformatic Analysis of Over-represented Genes and Pathways

Ingenuity Pathway Analysis (IPA) (Ingenuity Systems) was used to annotate predominant themes and pathways. Specifically, the top statistically significant canonical pathways and molecular functions were used to classify genes. Over representation of Gene Ontology (GO) groups targeted by alternative splicing were assessed using BINGO v2.3 (A Biological Network Gene Ontology Tool) [31] a Cytoscape plug-in [10]. In brief, a hypergeometric test was used to assess over represented GO Biological Processes. Benjamini & Hochberg False Discovery Rate (FDR) correction was applied and only themes with a statistical significance of $P < 0.05$ were included in the analysis.

Sense-Antisense Transcription

Filtered Affymetrix exon array probe sets were mapped to ~1,765 sense-antisense gene pairs (defined as overlapping by a minimum of 1bp, and encoded on opposing strands, as in Morrissy et al., 2011 [37]). A total of 376 genes with at least 20% of probe sets expressed above the filtering threshold were further considered. These genes had a total of 4,344 filtered probe sets. SI values for these probe sets were calculated as described above. Spearman's rank correlation coefficients were calculated between the SI values of each probe set in a sense gene, and the expression values of the antisense gene (across all samples). P-values for correlations were calculated using the `cor.test` function in R (R Development Core Team 2008), and were multiple-test corrected using the stringent Bonferroni method. For each S-AS gene pair, each gene partner was, in turn, analyzed as the sense gene and as the antisense gene (in order to identify cases where both genes had antisense-correlated splicing events).

Results

Alternative Splicing in Medulloblastoma is Subgroup Specific

To further highlight the transcriptional differences between medulloblastoma subgroups we analyzed alternative splicing consisting of the differential use of exons, promoters and polyadenylation sites in a large cohort of medulloblastomas (n=103) and normal cerebella (n=14). Using two independent bioinformatics algorithms – Splice Index (SI) [37] and Pattern Based Correlation (PAC) [18] – we created a ‘collective splicing series’ of 9096 putatively spliced probe sets that map to well annotated exons in 4622 genes (Figure S1a; Table S3). The majority of these alternatively spliced probe sets (64%) mapped to non-terminal exons, whereas 15% and 21% of our collective splice series affected probe sets which could be mapped to the first or last exon, respectively (Figure S2). Most of the identified alternative splicing occurred in medulloblastoma samples (79%), while only a minority (15%) was specifically enriched in the normal cerebella (Figure 1a). Subgroup-specific splicing events were most prevalent in Group 3 and SHH tumors (19.4% and 16.2% respectively) and less abundant in Group 4 (9.3%) and WNT (3.2%) medulloblastomas. Half (51.9%) of all medulloblastoma-enriched splicing events occurred across subgroups in a mixed population of medulloblastomas (Figure 1b). We identified genes with known roles in medulloblastoma and cerebellar development including: *AXIN2* (WNT), *GLII* [47], *TSC1* [5] and (SHH) [24](Table S5; Table S6). We also observed the previously reported medulloblastoma-specific affecting *ERBB4*[17].

During cerebella development, a significant increase in alternative splicing is observed as the normal cerebellum develops from the fetus to adulthood ($P < 4.34 \times 10^{-8}$) (Figure 1c). The adult cerebella demonstrates 3.93-times higher median levels of alternative splicing (Figure 1c; Figure S3a) relative to the fetal cerebella, however within fetal or adult samples there exists no direct correlation between age and the observed frequency of alternative splicing (Figure S3b; Figure S3c). Medulloblastomas display on average 2.3-times the median levels present in the developing fetal cerebella ($P < 6.47 \times 10^{-8}$), which nonetheless remain 0.59-times lower than those observed within the developed, adult cerebella ($P < 1.89 \times 10^{-2}$). A subgroup-specific analysis of medulloblastoma alternative splicing reveals no statistically significant differences in the observed number of spliced probe sets across WNT, SHH and Group 3 tumors whereas Group 4 medulloblastomas possess a reduced frequency of alternative splicing ($P < 2.31 \times 10^{-2}$). Although medulloblastoma is largely a pediatric disease, adult tumors (age > 16) represent 13.6% (14/102) of our tumor cohort. Pediatric versus adult medulloblastomas do not display any statistically significant differences ($P < 4.92 \times 10^{-1}$) in the observed frequency of alternative splicing events (Figure S4a). Furthermore, there is no correlation between the age of the patient and the frequency of alternative splicing when

medulloblastoma is analyzed as a single disease, (Figure S4b) however a weak, positive trend towards increasing alternative splicing with age was observed in non- Group 4 tumors (Figure S4c).

The extensive intra-subgroup variance in abundance of alternative spliced probesets permits stratification of medulloblastomas into two broader groups, distinguished by the frequency of alternative splicing. The first group, referred to as “hyperspliced”, is composed of 26 samples with splicing frequencies above the 75th percentile across all medulloblastomas. The second group, with splicing frequencies comparable to those present in normal cerebella, is referred to as “non-hyperspliced”. Notably, we observed relative differences in the distributions of Group 3 and 4 subgroups across both hyperspliced and non-hyperspliced medulloblastomas. An increase (+16%) in the distribution of Group 3 tumors was observed in the hyperspliced group with the inverse relationship (−20%) for Group 4 medulloblastomas (Figure 1d). Hyperspliced medulloblastomas demonstrate 2.18 (WNT) to 4.97 (Group 3) times greater frequency of median splicing events relative to non-hyperspliced tumors in the same sub-group (Figure 1e). Strikingly, there is a significant decrease in the overall survival of patients with hyperspliced tumors ($P < 3.08E-2$) (Figure 1f) with a trend towards increased mortality across all molecular subgroups of the disease (Figure 1g). There exists no significant change in the frequency of alternative splicing that occurs in the presence of metastasis (Figure S5a), nor is there any change in the incidence of metastasis which correlates with the presence of the hyperspliced phenotype (Figure S5b). Whether this hyperspliced phenomena is a true biological event with clinical significance, or an artifact associated with the current sample cohort, the algorithms used for analysis, or the platform used, remains to be proven through identification and validation of the hypersplice phenotype on a separate cohort of medulloblastomas with exon level expression data derived from another hybridization or sequencing-based platform.

Unsupervised Clustering of Splice Indices Identifies Four Medulloblastoma Subgroups

Through unsupervised hierarchical clustering (HCL) of Splice Index (SI) values, we were able to recapitulate the clustering pattern produced by gene-level transcriptional data [41, 43], generating six major clusters with four clear medulloblastoma subgroups, in addition to normal fetal and adult cerebella clusters (Figure 2a). Ninety-two percent (92%, 95/103) of samples clustered according to their predicted molecular subgroup, while six samples (8%, 8/103) were misclassified. Clustering discrepancies occurred largely (87.5%, 7/8) between Group 3 and 4 medulloblastomas – two molecular subgroups previously shown to display a higher concordance in copy number and transcriptional profiles. The clustering pattern observed was highly robust (Figure S6a) with >98% confidence associated with the clustering patterns of WNT, SHH and normal cerebellar samples (Figure S6b). Fetal and adult normal cerebella clustered together with confidence scores >81% irrespective of the number of probe sets used to generate the clusters, suggesting they display a distinct alternative splicing pattern from the medulloblastoma samples profiled (Figure S6a). There is clear sub-structure identified within Group 3 medulloblastomas, with half of all Group 3 hyperspliced tumors clustering with a high confidence (77%) (Figure S6b), further supporting the necessity of characterizing intra-subgroup heterogeneity. Using an independent and unsupervised learning algorithm, Non-negative Matrix Factorization (NMF), we were able to reproduce our HCL clustering patterns. NMF provided the highest support (Cophenetic correlation 0.9629) for 7 molecular subgroups consisting of the 6 major groups identified by HCL (Fetal cerebella, Adult cerebella, WNT, SHH, Group 3, Group 4) and one additional subgroup (Figure 2b). The additional subgroup consisted of a minority ($n=3$) of SHH cases clustering separately from other SHH tumors (Figure 2c). NMF produced an accuracy similar to that of HCL, with 93% (96/103) of medulloblastomas clustering as expected. Importantly, the clustering pattern produced by alternative splicing is

not driven by gene expression, as there is only 47.2% (244/516) overlap in the genes sets used to generate stable alternative slicing clustering and gene-level transcriptional clustering (Table S10). Using information generated from both HCL and NMF clustering, we identified highly recurrent hallmark alternative splicing events enriched in each of the molecular subgroups (Figure 2d).

Extensive Alternative Splicing of Cerebellar Development Genes in Non-WNT Medulloblastoma.

To identify genes and pathways disproportionately affected by alternative splicing we performed Ingenuity Pathway Analysis (IPA) in a subgroup specific manner (Figure 3a). We identified pathways with known roles in the pathogenesis of medulloblastoma, including p53 signaling (WNT tumors, $P < 1.09E-2$) [44] and CREB signaling (SHH tumors; $P < 1.70E-4$) [46]. Among medulloblastomas, *TP53* mutations are most common in the WNT subgroup [44]. In non-WNT medulloblastomas, we identified a high incidence of neuronal development pathways affected by alternative splicing. Of the top ten statistically significant pathways, 60% (6/10) in both SHH and Group 3 medulloblastomas, and 40% (4/10) of Group 4 tumors, affected neuronal functions (Figure S7). Normal cerebella exhibited some overlap with these findings, however neuronal functions are less frequently targeted (30%, 3/10). Instead, cell cycle pathways (30%, 3/10) are enriched in the normal cerebella (Table S11).

Using Cytoscape BINGO [10, 31], an independent algorithm for the visualization of Gene Ontology (GO) functions, we performed a subtractive analysis, removing gene ontologies present in the normal cerebella and identifying biological processes enriched exclusively in medulloblastoma. The results complemented our pathway analysis demonstrating a strong enrichment of neuronal networks, including nervous system development ($P < 1.30E-2$), axonal guidance ($P < 3.36E-3$) and glutamatergic synaptic transmission ($P < 2.19E-2$) in Group 3 medulloblastomas (Figure 3b). Additionally, this analysis identified signaling pathways previously implicated in medulloblastoma pathogenesis including the Roundabout (*ROBO-SLIT*, Group 3, $P < 1.13E-2$) [61] and *PDGF* pathways (Group 3, $P < 2.53E-2$) [1, 30]. Similarly, alternative splicing events in SHH and Group 4 tumors comprised a high percentage of neuronal pathways, and networks such as the regulation of cell migration ($P < 8.06E-3$, SHH) and extracellular structure organization ($P < 1.65E-2$, Group 4) (Figure S7, Table S12-S15).

Given the high incidence of mortality associated with hyperspliced medulloblastomas, we analyzed differential use of exons, alternative polyadenylation sites and alternative start sites between hyperspliced and non-hyperspliced tumors in an effort to identify possible molecular changes contributing to this phenotype. Using supervised clustering, we first identified the top 5% of probesets with the greatest differential splice indices between the two groups (Figure S9a; Figure S9b). We then examined the molecular function of the most differential probesets using Ingenuity Pathway Analysis (IPA). The majority of the top canonical pathways differentiating hyperspliced from non-hyperspliced tumours affected known cancer signaling pathways (60%, 6/10) (Figure S9c).

Validation of Subgroup-Specific Alternative Splicing Events

We selected high confidence subgroup specific splicing candidates from our consensus splicing series to validate. These events were predominantly found in a single molecular subgroup (>75%), and displayed gross changes in transcript structure of isoforms. To assess subgroup-specific isoform expression, a transcript ratio was calculated as change in 3' versus 5' expression levels, and normalized to fetal cerebellar levels. Exon-specific primers were then designed to distinguish between 3' and 5' exon cassettes. We validated alternative

splicing events for *INADL* (WNT), *CHN2* (Group 3), *NBEA* (Group 4) and *SNAP25* (Mixed MB). *INADL* is a cell polarity and tight junction protein [52] with a 3' alternative promoter isoform that is present in >80% of WNT tumors and a minority of Group 4 tumors (8%) (Figure 4a). Upon validation, WNT medulloblastomas displayed a 2-10 times greater transcript ratio (i.e. higher levels of the shorter isoform) relative to non-WNT tumors and normal cerebella (Figure 4b). *CHN2*, a Rho-GTPase Activating Protein [6, 25, 29] and *NBEA*, a neuronal differentiation protein [33] also demonstrated 3' alternative promoters enriched in Group 3 and Group 4 subgroups, respectively (Figure 4a). The *NBEA* truncated isoform was predicted to occur in 91% of Group 4 tumors and a minority of Group 3 medulloblastomas (37%), whereas the 3' *CHN2* isoform predominated in Group 3 (74%) and WNT (37%) medulloblastomas, evident in our validation (Figure 4b). Finally, we validated alternative splicing targeting *SNAP25*, a synaptosomal protein necessary for neurotransmitter exocytosis [3], with a known exon-5 cassette (*SNAP25a* versus *SNAP25b*). We observed a greater 5a to 5b ratio in normal fetal cerebella and the majority non-WNT medulloblastomas. In contrast, adult cerebella and WNT tumors demonstrate higher 5b levels. Each of these splicing events causes the disruption of one or more protein domains, likely resulting in a significant change in gene function (Figure S10).

Sense-Antisense (S-AS) Transcription Correlates with Alternative Splicing

Recent reports have demonstrated changes in alternative splicing patterns that correlate with the expression of an antisense gene [37]. Sense-antisense (S-AS) transcription occurs when overlapping genes on opposing DNA strands are co-expressed in the same cell [26, 51]. Antisense transcription can regulate splicing decisions and alter the balance of isoforms expressed from the sense strand through a variety of mechanisms, such as direct transcriptional interference based on the physical consequences of convergent polymerase complexes (PolII) transcribing both strands of the sense-antisense gene locus [50]. In normal human cells, antisense transcription is significantly correlated to splicing at hundreds of loci, showing that this is likely a common mechanism of transcriptional regulation (Figure 5a), occurring in >75% of all genes and often altered in a cancer-specific context [32].

To determine whether antisense transcription contributes to alternative splicing in medulloblastoma, we analyzed 188 overlapping gene pairs (i.e. 376 genes) that are encoded in opposing orientations and simultaneously expressed in a given tumor. Alternative splicing was predicted to occur in either the sense or antisense gene in 88 (46.8%) sense-antisense partners. We measured the correlation between exon inclusion in the sense genes (splice index (SI) values; see Methods), and the expression of the antisense gene partner, across all 117 samples, identifying significant correlations between splicing and antisense gene expression in all (100%, 88/88) gene pairs ($p < 0.05$, after Bonferroni correction) (Figure S11a). Our results suggest that S-AS transcription may play a role in the regulation of alternative splicing of these genes in medulloblastoma.

Notable examples of such events include the well-annotated S-AS pairs *NBEA-MAB21L1*, *NNAT-BLCAP* and *BCL2L12-IRF3*. All three examples have previously been identified, and validated [15, 39, 55], using independent molecular techniques, demonstrating the validity and strength of our approach. Of specific interest to us was *NBEA*, previously identified as alternatively spliced in Group 4 medulloblastomas where it is enriched in isoforms distinguished by an alternative transcriptional start site located mid-gene (Figure 4). Analysis of the antisense-correlated splicing events identified the presence of two predominant *NBEA* isoforms (Figure 5b). Expression of the longer *NBEA* isoform was negatively correlated ($r = -0.61$) to the expression of the antisense gene (*MAB21L1*). In contrast, expression of the shorter *NBEA* isoform, previously identified as up-regulated in Group C and Group D medulloblastomas (Figure 4b), was positively correlated ($r = 0.74$) to

MAB21L1 expression (also up-regulated in Group D tumors (Figure S11b)). These results suggest that subgroup specific expression of *MAB21L1* contributes to the regulation of subgroup specific alternative promoter usage of *NBEA*.

In the set of 88 S-AS gene pairs with significant correlations between alternative splicing and concomitant antisense transcription, there exist subgroup-restricted (Figure 5c, middle and bottom panels) and subgroup-independent events (Figure 5c, top panel). To understand the putative biological relevance of S-AS genes with antisense-correlated splicing events, we performed pathway analysis, and identified an enrichment of critical cellular functions such as cell death, cell cycle regulation and cellular development (Figure S11c). Further validation of the relationship between antisense transcription and the regulation of alternative splicing in larger datasets with more comprehensive AS transcriptional data will allow further testing of our model. Antisense genes that are expressed in a highly subgroup specific manner should be considered as candidate marker genes for subgroup assignment.

Discussion

We present the first subgroup specific analysis of alternative splicing in medulloblastoma using two independent bioinformatic approaches. We identified differential splicing across each medulloblastoma subgroup as well as normal fetal and adult cerebella. While age-matched normal tissue was unavailable, normal fetal (20 to 40 weeks of age) and adult (22 to 82 years of age) cerebella were used as normal controls, representing the developing and developed cerebella. Based on the distribution of alternative splicing events across all tumors, we were able to identify a ‘hypersliced’ phenotype evident in one-quarter of all medulloblastomas. Although hyperspliced medulloblastomas are evident across all molecular subgroups, they are significantly under-represented in Group 4 tumors. Notably, they display decreased overall survival and a strong trend towards increased mortality in non-Group 4 medulloblastomas, indicating that survival trends are not influenced by the over-representation of a single aggressive molecular variant. Hyperspliced tumors do not display any change in the observed frequency of metastatic disease (M+), nor do we observe higher levels of alternative splicing in M+ tumors. Bioinformatic analysis of probesets differentiating hyperspliced from non-hyperspliced medulloblastomas identified numerous cancer-signaling pathways whose instability may, in part, explain the aggressive nature of hyperspliced tumors. Clinical and biological relevance of the observed hyperspliced phenotype awaits its validation on a separate set of tumors studied using an independent technology.

Exon arrays were designed to allow complementary analyses to be performed, both at the level of gene expression as well as alternative splicing [16]. For the latter purpose, the relative difference in exon-level probeset expression can be effectively interpreted as alternative splicing, alternative promoter usage, and alternative polyadenylation events [11, 19, 34, 38]. Although a comprehensive survey of transcript structures cannot be conducted using this platform, particularly in genes with low expression values or in genes with numerous co-expressed isoforms, obvious and consistent events corresponding to the above categories can be measured reliably and reproducibly [16, 64]. Although sequencing technologies can explicitly profile exonexon junctions, their comparatively prohibitive cost ensures that array-based approaches are an efficient and informative method for rapidly assaying large numbers of tumors, in an effort to address the known heterogeneity of this disease. However, a complete catalog of alternative splicing events remains to be identified through the use of alternative technologies.

Our results are supported by previous research demonstrating a role for alternative splicing in medulloblastoma. Most recently, Menghi et al, 2011 [34] examined alternative splicing in

a modest cohort of 14 medulloblastomas using the Splice Index algorithm which differentiated SHH from non-SHH medulloblastomas. The authors identified 174 high confidence splicing events, with an additional 285 probable events. Of these, they validated 11/14 alternatively spliced exons present in a SHH-specific (3/11) or in a medulloblastoma-enriched (8/11) manner. Our analysis identified 100% of the validated events presented by Menghi et al. [34], largely following the subgroup associations reported in their investigation. There are some discrepancies, such as the presence of a SHH-restricted isoform of *TRRAP* reported by the Menghi study, which was observed in all molecular subgroups of medulloblastoma in our dataset.

We examined whether the variation in alternative splicing patterns reflected the transcriptional heterogeneity that defines the four molecular subgroups of medulloblastoma using two independent methods of unsupervised clustering: Hierarchical Clustering (HCL) and Non-negative Matrix Factorization (NMF). HCL produced the most robust clustering with four molecular subgroups of medulloblastoma, findings largely recapitulated by NMF, which identified the four core subgroups and one additional minor medulloblastoma cluster. The extra NMF cluster composed of three SHH medulloblastomas, two of which are large cell anaplastic and hyperspliced, while the remaining SHH tumor is classic and non-hyperspliced. These tumors cluster apart from other SHH tumors in our HCL analysis, suggesting they may represent outliers. As the clustering patterns produced were largely driven by exon from genes independent from those used to produce transcriptional clustering, our results suggest that subgroup-specific alternative splicing events are an independent and equally informative measure of the heterogeneity that exists within the medulloblastoma transcriptome.

By examining alternative splicing events that predominate within each medulloblastoma subgroup, we identified hallmark events, many of which have neuronal functions. These prevalent splicing events, found in >50-70% of tumors in each molecular variant, may identify genes important in tumorigenesis, and may aid in the identification of the cell type of origin. WNT and non-WNT tumors display significantly different pathways affected by alternative splicing. Pathway analysis of WNT medulloblastomas revealed genes important to cerebellar development and medulloblastomas pathogenesis targeting tight junction signaling ($P < 1.49E-2$) and p53 signaling ($P < 1.08E-2$) [44]. In non-WNT medulloblastomas there was a high incidence of pathways affecting nervous system development and differentiation. Important neuronal signaling pathways include CREB signaling in neurons (SHH tumors, $P < 1.70E-4$) and RAR activation (Group 4 tumors, $P < 2.77E-3$), both of which are core or peripheral elements of retinoic acid (RA) signaling. Retinoid treatment has been previously used to induce differentiation in medulloblastoma cells [22]. Furthermore, our analysis has identified subgroup-specific pathways targeting Roundabout (ROBO/SLIT) in Group 3 tumors. Deregulation of ROBO/SLIT genes may contribute to the high incidence of metastasis and brain tumor invasion observed in aggressive Group 3 medulloblastomas [61].

We suggest a model in which antisense transcription may represent one mechanism able to mediate alternative splicing outcomes. We found that approximately 47% of S-AS gene pairs expressed in medulloblastoma had a significant relationship between sense gene splicing outcomes, and expression of the antisense gene partner. An example of this relationship is the alternative splicing of *NBEA*, a neuronal development protein. Our data show that a majority of Group 4 (91%) tumors express a truncated *NBEA* isoform containing only 27% of the full length coding sequence, lacking important functional domains. Expression of this short *NBEA* isoform is significantly correlated to expression of *MAB21L1*, a gene encoded on the opposite strand of *NBEA*, and believed to function in embryonal development [55]. Many of the identified S-AS events occurred in a subgroup

specific manner, indicating that antisense transcription is likely an important component in the regulation of subgroup-specific alternative splicing and medulloblastoma tumorigenesis.

Our data reveals important clinical and biological trends associated with alternative splicing in the medulloblastoma transcriptome, suggests a putative mechanism for subgroup specific alternative splicing, and further highlights the transcriptional heterogeneity present across, and within, subgroups of medulloblastoma.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Abouantoun TJ, MacDonald TJ. Imatinib blocks migration and invasion of medulloblastoma cells by concurrently inhibiting activation of platelet-derived growth factor receptor and transactivation of epidermal growth factor receptor. *Mol Cancer Ther.* 2009; 8(5):1137–1147. [PubMed: 19417143]
2. Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, et al. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol.* 2009; 16(7):717–724. [PubMed: 19543290]
3. Bark C, Bellinger FP, Kaushal A, Mathews JR, Partridge LD, et al. Developmentally regulated switch in alternatively spliced SNAP-25 isoforms alters facilitation of synaptic transmission. *J Neurosci.* 2004; 24(40):8796–8805. [PubMed: 15470145]
4. Bartel F, Taubert H, Harris LC. Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer cell.* 2002; 2(1):9–15. [PubMed: 12150820]
5. Bhatia B, Northcott PA, Hambardzumyan D, Govindarajan B, Brat DJ, et al. Tuberous sclerosis complex suppression in cerebellar development and medulloblastoma: separate regulation of mammalian target of rapamycin activity and p27 Kip1 localization. *Cancer Res.* 2009; 69(18):7224–7234. [PubMed: 19738049]
6. Bruinsma SP, Baranski TJ. Beta2-chimaerin in cancer signaling: connecting cell adhesion and MAP kinase activation. *Cell Cycle.* 2007; 6(20):2440–2444. [PubMed: 17726379]
7. Bueren AO, von Hoff K, Pietsch T, Gerber NU, Warmuth-Metz M, et al. Treatment of young children with localized medulloblastoma by chemotherapy alone: results of the prospective, multicenter trial HIT 2000 confirming the prognostic impact of histology. *Neuro Oncol.* 2011; 13(6):669–679. [PubMed: 21636711]
8. Buonamici S, Williams J, Morrissey M, Wang A, Guo R, et al. Interfering with resistance to smoothened antagonists by inhibition of the PI3K pathway in medulloblastoma. *Sci Transl Med.* 2010; 2(51):51ra70.
9. Cho Y-J, Tsherniak A, Tamayo P, Santagata S, Ligon A, et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol.* 2011; 29(11):1424–1430. [PubMed: 21098324]
10. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat protoc.* 2007; 2(10):2366–2382. [PubMed: 17947979]
11. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 2010; 24(21):2343–2364. [PubMed: 21041405]
12. Dubuc AM, Northcott PA, Mack S, Witt H, Pfister S, et al. The genetics of pediatric brain tumors. *Curr Neurol Neurosci Rep.* 2010; 10(3):215–223. [PubMed: 20425037]
13. Eberhart CG. Molecular diagnostics in embryonal brain tumors. *Brain Pathol.* 2011; 21(1):96–104. [PubMed: 21129063]
14. Ellison DW. Childhood medulloblastoma: novel approaches to the classification of a heterogeneous disease. *Acta Neuropathol.* 2010; 120(3):305–316. [PubMed: 20652577]

15. Evans HK, Weidman JR, Cowley DO, Jirtle RL. Comparative phylogenetic analysis of blcap/nnat reveals eutherian-specific imprinted gene. *Mol Biol Evol.* 2005; 22(8):1740–1748. [PubMed: 15901842]
16. Fan W, Khalid N, Hallahan AR, Olson JM, Zhao LP. A statistical method for predicting splice variants between two groups of samples using GeneChip expression array data. *Theor Biol Med Model.* 2006; 3:19. [PubMed: 16603076]
17. Ferretti E, Di Marcotullio L, Gessi M, Mattei T, Greco a, et al. Alternative splicing of the ErbB-4 cytoplasmic domain and its regulation by hedgehog signaling identify distinct medulloblastoma subsets. *Oncogene.* 2006; 25(55):7267–7273. [PubMed: 16878160]
18. French PJ, Peeters J, Horsman S, Duijm E, Siccama I, et al. Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Res.* 2007; 67(12):5635–5642. [PubMed: 17575129]
19. Gardina PJ, Clark T a, Shimada B, Staples MK, Yang Q, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics.* 2006; 7:325. [PubMed: 17192196]
20. Gibson P, Tong Y, Robinson G, Thompson MC, Currie DS, et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature.* 2010; 468(7327):1095–1099. [PubMed: 21150899]
21. Gilbertson RJ. Finding the perfect partner for medulloblastoma prognostication. *J Clin Oncol.* 2011; 29(29):3841–3842. [PubMed: 21911713]
22. Hallahan AR, Pritchard JJ, Chandraratna R a S, Ellenbogen RG, Geyer JR, et al. BMP-2 mediates retinoid-induced apoptosis in medulloblastoma cells through a paracrine effect. *Nat Med.* 2003; 9(8):1033–1038. [PubMed: 12872164]
23. Hilmi C, Guyot M, Pagès G. VEGF spliced variants: possible role of antiangiogenesis therapy. *J Nucleic Acids.* 2012; 2012:162692. [PubMed: 22013509]
24. Ingham PW. The patched gene in development and cancer. *Curr Opin Genet Dev.* 1998; 8(1):88–94. [PubMed: 9529611]
25. Kai M, Yasuda S, Imai S- I, Kanoh H, Sakane F. Tyrosine phosphorylation of beta2-chimaerin by Src-family kinase negatively regulates its Rac-specific GAP activity. *Biochim Biophys Acta.* 2007; 1773(9):1407–1415. [PubMed: 17560670]
26. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, et al. Antisense transcription in the mammalian transcriptome. *Science.* 2005; 309(5740):1564–1566. [PubMed: 16141073]
27. Kool M, Koster J, Bunt J, Hasselt NE, Lakeman A, et al. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PloS one.* 2008; 3(8):e3088. [PubMed: 18769486]
28. Kwan T, Benovoy D, Dias C, Gurd S, Serre D, et al. Heritability of alternative splicing in the human genome. *Genome Res.* 2007; 17(8):1210–1218. [PubMed: 17671095]
29. Leung T, How BE, Manser E, Lim L. Cerebellar beta 2-chimaerin, a GTPase-activating protein for p21 ras-related rac is specifically expressed in granule cells and has a unique N-terminal SH2 domain. *J Biol Chem.* 1994; 269(17):12888–12892. [PubMed: 8175705]
30. Li X- N, Shu Q, Su JM, Adesina a M, Wong KK, et al. Differential expression of survivin splice isoforms in medulloblastomas. *Neuropathol Appl Neurobiol.* 2007; 33(1):67–76.
31. MacDonald TJ, Brown KM, LaFleur B, Peterson K, Lawlor C, et al. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet.* 2001; 29(2):143–152. [PubMed: 11544480]
32. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics.* 2005; 21(16):3448–3449. [PubMed: 15972284]
33. Mata MD, Alonso CR, Fededa JP, Pelisch F, Cramer P, et al. A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Mol Cell.* 2003; 12(2):525–532. [PubMed: 14536091]
34. Medrihan L, Rohlmann A, Fairless R, Andrae J, Döring M, et al. Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. *J Physiol.* 2009; 587(Pt 21):5095–5106. [PubMed: 19723784]

35. Menghi F, Jacques TS, Barenco M, Schwalbe EC, Clifford SC, et al. Genomewide analysis of alternative splicing in medulloblastoma identifies splicing patterns characteristic of normal cerebellar development. *Cancer Res.* 2011; 71(6):2045–2055. [PubMed: 21248070]
36. Modrek B, Lee C. A genomic view of alternative splicing. *Nat Genet.* 2002; 30(1):13–19. [PubMed: 11753382]
37. Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, et al. Nextgeneration tag sequencing for cancer gene expression profiling. *Genome Res.* 2009; 19(10):1825–1835. [PubMed: 19541910]
38. Morrissy AS, Griffith M, Marra MA. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res.* 2011; 21(8):1203–1212. [PubMed: 21719572]
39. Nagao K, Togawa N, Fujii K, Uchikawa H, Kohno Y, et al. Detecting tissuespecific alternative splicing and disease-associated aberrant splicing of the PTCH gene with exon junction microarrays. *Hum Mol Genet.* 2005; 14(22):3379–3388. [PubMed: 16203740]
40. Nakajima A, Nishimura K, Nakaima Y, Oh T, Noguchi S, et al. Cell typedependent proapoptotic role of Bcl2L12 revealed by a mutation concomitant with the disruption of the juxtaposed Irf3 gene. *Proc Natl Acad Sci USA.* 2009; 106(30):12448–12452. [PubMed: 19617565]
41. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature.* 2010; 463(7280):457–463. [PubMed: 20110989]
42. Northcott PA, Fernandez-L A, Hagan JP, Ellison DW, Grajkowska W, et al. The miR-17/92 polycistron is up-regulated in sonic hedgehog-driven medulloblastomas and induced by N-myc in sonic hedgehog-treated cerebellar neural precursors. *Cancer Res.* 2009; 69(8):3249–3255. [PubMed: 19351822]
43. Northcott, P a; Hielscher, T.; Dubuc, A.; Mack, S.; Shih, D., et al. Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct. *Acta Neuropathol.* 2011; 122(2):231–240. [PubMed: 21681522]
44. Northcott, P a; Korshunov, A.; Witt, H.; Hielscher, T.; Eberhart, CG., et al. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol.* 2011; 29(11):1408–1414. [PubMed: 20823417]
45. Pfaff E, Remke M, Sturm D, Benner A, Witt H, et al. TP53 mutation is frequently associated with CTNNB1 mutation or MYCN amplification and is compatible with long-term survival in medulloblastoma. *J Clin Oncol.* 2010; 28(35):5188–5196. [PubMed: 21060032]
46. Pfister SM, Korshunov A, Kool M, Hasselblatt M, Eberhart C, et al. Molecular diagnostics of CNS embryonal tumors. *Acta Neuropathol.* 2010; 120(5):553–566. [PubMed: 20882288]
47. Pons S, Trejo JL, Martínez-Morales JR, Martí E. Vitronectin regulates Sonic hedgehog activity during cerebellum development through CREB phosphorylation. *Development.* 2001; 128(9):1481–1492. [PubMed: 11290288]
48. Raffel C. Medulloblastoma: molecular genetics and animal models. *Neoplasia.* 2004; 6(4):310–322. [PubMed: 15256053]
49. Remke M, Hielscher T, Korshunov A, Northcott P a, Bender S, et al. FSTL5 Is a Marker of Poor Prognosis in Non-WNT/Non-SHH Medulloblastoma. *J Clin Oncol.* 2011; 29(29):3852–3861. [PubMed: 21911727]
50. Rudin CM, Hann CL, Lattera J, Yauch RL, Callahan C a, et al. Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. *N Engl J Med.* 2009; 361(12):1173–1178. [PubMed: 19726761]
51. Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. *Trends Genet.* 2005; 21(6):339–345. [PubMed: 15922833]
52. Shendure J, Church GM. Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol.* 2002; 3(99):RESEARCH0044. [PubMed: 12225583]
53. Storrs CH, Silverstein SJ. PATJ, a tight junction-associated PDZ protein, is a novel degradation target of high-risk human papillomavirus E6 and the alternatively spliced isoform 18 E6. *J Virol.* 2007; 81(8):4080–4090. [PubMed: 17287269]
54. Thompson MC, Fuller C, Hogg TL, Dalton J, Finkelstein D, et al. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol.* 2006; 24(12):1924–1931. [PubMed: 16567768]

55. Tjink BM, Buter J, de Bree R, Giaccone G, Lang MS, et al. A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus. *Clin Cancer Res*. 2006; 12(20 Pt 1):6064–6072. [PubMed: 17062682]
56. Tsang WH, Shek KF, Lee TY, Chow KL. An evolutionarily conserved nested gene pair - Mab21 and Lrba/Nbea in metazoan. *Genomics*. 2009; 94(3):177–187. [PubMed: 19482073]
57. Uchikawa H, Toyoda M, Nagao K, Miyauchi H, Nishikawa R, et al. Brain and heart-specific Patched-1 containing exon 12b is a dominant negative isoform and is expressed in medulloblastomas. *Biochem Biophys Res Commun*. 2006; 349(1):277–283. [PubMed: 16934747]
58. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456(7221):470–476. [PubMed: 18978772]
59. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008; 14(5):802–813. [PubMed: 18369186]
60. Wei J, Zaika E, Zaika A. p53 Family: Role of Protein Isoforms in Human Cancer. *J Nucleic Acids*. 2012; 2012:687359. [PubMed: 22007292]
61. Werbowetski-Ogilvie TE, Seyed Sadr M, Jabado N, Angers-Loustau a, Agar NYR, et al. Inhibition of medulloblastoma cell invasion by Slit. *Oncogene*. 2006; 25(37):5103–5112. [PubMed: 16636676]
62. Xin D, Hu L, Kong X. Alternative promoters influence alternative splicing at the genomic level. *PloS one*. 2008; 3(6):e2377. [PubMed: 18560582]
63. Yang Z- J, Ellis T, Markant SL, Read T-A, Kessler JD, et al. Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer cell*. 2008; 14(2):135–145. [PubMed: 18691548]
64. Zhang C, Li H- R, Fan J- B, Wang-Rodriguez J, Downs T, et al. Profiling alternatively spliced mRNA isoforms for prostate cancer classification. *BMC Bioinformatics*. 2006; 7:202. [PubMed: 16608523]
65. Zhu H, Lo H-W. The Human Glioma-Associated Oncogene Homolog 1 (GLI1) Family of Transcription Factors in Gene Regulation and Diseases. *Curr Genomics*. 2010; 11(4):238–245. [PubMed: 21119888]

Figure 1. Subgroup-Specific Alternative Splicing in Medulloblastoma

(a) Distribution of 9096 alternatively spliced probe sets, identified by Splice Index (SI) and Pattern-Based Correlation (PAC) algorithms, across 103 primary medulloblastoma and 14 normal cerebella samples demonstrates strong enrichment patterns in medulloblastoma (79%) with a minority of alternative splicing events restricted to the normal cerebella (15%). (b) Subgroup association of 7509 probe sets with medulloblastoma-enriched splicing patterns identifies elevated levels of Group 3 (19.4%, 1454 probe sets) and SHH (16.2%, 1216 probe sets) enriched alternative splicing with lower levels present in WNT (3.2%, 241 probe sets) and Group 4 (9.3%, 697 probe sets) tumors. Half of all medulloblastoma-

enriched splicing events (51.9%, 3901 probe sets) were identified in medulloblastomas from multiple subgroups.

(c) Using Splice Index (SI) and Pattern Based Correlation (PAC) algorithms the number of alternative splicing events per sample was identified, producing similar trends for both algorithms. Subgroup-specific splicing patterns revealed a significant developmental increase in alternative splicing from the fetal to adult normal cerebella with a further increase in splicing observed in medulloblastomas.

(d) Distribution of the molecular subgroups of medulloblastoma in hyperspliced (n=26) and non-hyperspliced tumors (n=77) reveals an increased frequency of WNT (+6%) and Group 3 tumors (+16%) in hyperpliced medulloblastomas, and a decreased frequency in Group 4 tumors (-20%).

(e) Subgroup-specific distribution of alternative splicing per sample for hyperspliced versus non-hyperspliced tumors. Hyperspliced tumors demonstrate a significant increased number of alternatively spliced exons across each subgroup – ranging from 2.18 (WNT) to 4.97 (Group 3) times higher levels of splicing.

(f) Hyperspliced medulloblastomas display a significantly decreased overall survival ($P < 3.08E-2$) relative to non-hyperspliced medulloblastomas.

(g) Each molecular subgroup of medulloblastomas demonstrates a trend towards increased mortality for hyperspliced tumors, with an 80% or greater increase in mortality for WNT, SHH and Group 3 medulloblastomas.

Figure 2. Unsupervised Clustering of Splice Indices Identifies Four Subgroups of Medulloblastomas

- (a) Unsupervised Hierarchical Clustering (HCL) of the top 1000 probe sets with the highest standard deviation across Splice Index (SI) values generates robust clustering of four core medulloblastoma subgroups with two normal cerebella clusters.
- (b) Non-Negative Matrix Factorization using the same 1000 probe sets used for HCL clustering produces the highest support (cophenetic correlation) for 7 subgroups.
- (c) Non-Negative Matrix Factorization demonstrates 7 core subgroups composed of the adult and fetal normal samples, the corresponding four subgroups identified by HCL, and one additional subgroup composed of a small set of SHH medulloblastomas (n=3).

(d) Medulloblastoma subgroup-enriched alternative splicing events identified by SI and PAC analysis revealing events present in >50-70% of each subgroup.

Figure 3. Pathway and Gene Ontology analysis of Subgroup-Specific Splicing Events Identifies Recurrent Targeting of Cerebellar Development Pathways in non-WNT medulloblastomas

(a) Ingenuity Pathway Analysis (IPA) of the top ten pathways affected by alternative splicing across each molecular subgroup of medulloblastoma. Known signaling pathways: such as Tight junction signaling (WNT, $P<1.49E-2$) and CREB signaling (SHH, $P<1.70E-4$) were identified in our analysis as well as an abundance of neuronal pathways in non-WNT medulloblastomas.

(b) Cytoscape BINGO analysis of the significant Gene Ontologies (GO) targeted by alternative splicing in Group 3 tumors, after subtracting events present in the normal

cerebella, identifies neuronal pathways targeting axonogenesis and glutamatergic synaptic transmission.

Figure 4. Validation of Subgroup-Specific Hallmark Alternative Splicing Events in Medulloblastoma

(a) Exon array RMA signal intensity plots of highly frequent and subgroup-specific alternative transcripts. Alternative promoter usage generates a WNT-specific 3' isoform of *INADL*, while alternative promoter usages in Group 3 and Group 4 tumors produce known isoforms of *CHN2* and *NBEA*, respectively. *SNAP25* demonstrates subgroup-specific expression of a known exon cassette with Fetal and non-WNT medulloblastomas expressing elevated levels of *SNAP25a* and Adult and WNT tumors expressing higher levels of *SNAP25b*.

(b) Primers generated against the 5' and 3' gene regions of the full transcript were used to assess the intra-transcript variability. A transcript ratio, based on the 3':5' expression, was calculated and normalized to normal fetal cerebella permitting the identification of subgroup specific isoform expression. For each of the reported genes we observe the expected isoform expression restricted to the predicted subgroups.

Figure 5. Sense-Antisense Transcription Correlates with Alternative Splicing in Medulloblastoma

(a) Schematic of sense-antisense (S-AS) transcription depicting overlapping genes on opposing strands with concomitant expression. A switch in the predominant sense-strand isoform occurs in the context of antisense transcription.

(b) Analysis of *MAB21L1* transcription (antisense gene) and correlated NBEA (sense gene) alternative splicing events. The splice index values of 38 *NBEA* 5' exons are inversely correlated with expression of *MAB21L1* (i.e. these exons are excluded from the *NBEA* isoform when *MAB21L1* is expressed). Conversely, the 22 3' exons have splice index values

that are positively correlated to *MAB21L1* expression (i.e. these exons are included in the NBEA isoform when MAB21L1 is expressed).

(c) Examples of S-AS gene pairs demonstrating inverse relationships between sense strand exon inclusion and antisense strand transcription. These correlations demonstrate mixed medulloblastomas (*C13ORF3-MRP63*) or subgroup-enriched (*DDX31-GTF3C4*, Group 3 & Group 4) patterns suggesting S-AS events may play a critical role in normal development (*SLC26A10-B4GALNT*, Normal CB) and the pathogenesis of medulloblastoma.