Università degli Studi di Roma "Sapienza"

DOTTORATO DI RICERCA IN MEDICINA MOLECOLARE

XXV CICLO

PhD Thesis

"TRANSCRIPTIONAL LANDSCAPE OF NEURONAL and CANCER STEM CELLS"

PhD Student: Dr. Evelina Miele

Tutor: Prof. Elisabetta Ferretti

Coordinator: Prof. Alberto Gulino

Anno Accademico 2011-2012

INDEX

Introduction

Tumor mass is composed by a heterogeneous cell population including a subset of “cancer stem cells” (CSC), arisen by either transformed somatic stem cells or progenitor/differentiated cells reprogrammed towards cancer stemness (1,2). Thus, CSC share features with both cancer and stem cells such as self-renewal, epithelial/mesenchimal transition and constitutive activation of stemness-promoting signals (1).

CSC are believed to trigger tumor formation, drive resistance to conventional therapeutics and underlie patients’ relapse (2). Indeed, stem cell signatures have been associated with poor prognosis in various tumors (3, 4, 5, 6). This background makes the identification of CSC molecular features mandatory to understand tumor progression and to design novel CSC specific therapeutic strategies.

Medulloblastoma (Mb) is the most common childhood malignant brain tumor and a leading cause of cancer-related morbidity and mortality. Current multimodal therapies are effective in about 50% of patients but often cause long-term side effects, i.e. developmental, neurological, neuroendocrine and psychosocial deficits (7).

For many years, Mb was treated as a single tumor entity despite the divergent histology, patients’outcome and drug sensitivity, and also the diversity of the cell of origin. Very recently the scenario of human Mb has dramatically changed since its heterogeneous biology has been addressed by gene expression high-throughput micro-arrays (8, 9, 10) and exome deep sequencing techniques (11,12). These led to the identification of four Mb subgroups (WNT, SHH, Group 3 and Group 4) uncovering the existence of a highly diverse mutational spectra and gene expression (13, 14, 7).

However a quantitative approach has not yet been applied to the transcriptional landscape of Medulloblastoma stem cells (MbSC) through RNA Next Generation Sequencing (RNA-Seq) technology. This is a relevant issue, since RNA-Seq is able to interrogate the genome wide global transcriptome including new transcripts, alternative spliced isoforms and non-coding RNAs.

The cellular origin of 3 of the 4 MB subgroups has been identified. Lower rhombic lip progenitors of the dorsal brainstem are considered the triggering cells in WNT tumors. In SHH subgroup initiation cells are Prominin1^{+} CD15^{+} stem cells from the subventricular zone requiring the commitment to Math1^{+} granule cell progenitors [GCP] of the external granule cell layer [EGL]. Math1^{+} or Math1^{-} EGL-GCP or Prominin1^{+}/lineage-negative stem cells sustain, instead, the MYC-driven Group 3 tumors (7, 2).

MbSC derived from SHH tumors and postnatal normal cerebellar stem cells (NcSC) have been reported to share several features. A key signal for both of them is Hedgehog (Hh) (15). Furthermore, both NcSC and MbSC display up-regulation of stemness genes (e.g *Sox2*, *Nestin*, *Nanog, Prom1*) (15). Finally, constitutive activation of the Hh pathway by conditional deletion of *Ptch1* inhibitory receptor in NcSC, promote Mb in vivo (16, 17,18), producing a mouse model of the human SHH tumor. Acquisition of stemness features may therefore represent the first step of oncogenic conversion. Cooperation with additional oncogenic signals is however needed to enhance MbSC tumorigenicity (19). Therefore, MbSCs may be characterized by both stemness and additional cancer-specific features. However, the identification of the genes responsible for these characteristics of MbSC by large-scale expression analysis is poorly investigated.

In order to understand the MbSCs transcriptional programs we analyze MbSC derived from *Ptch1*$^{+/-}$ tumors (*Ptch1*$^{+/-}$ MbSC) by RNA-Seq. The choice of a genetically determined model of Mb has allowed us to work with cancer stem cells together with appropriate normal counterpart, and to analyze biological replicates for statistical analysis.

We identify transcripts characterizing MbSC, including novel isoforms, and long non-coding RNAs. Pathway analysis of genes significantly deregulated in the MbSC setting includes also epithelial/mesenchimal transition features (Focal Adhesion and ECM-receptor interaction) and Hedgehog signaling with high expression and biological relevance of Gas1 and Hdac2. Some of these stemness- and cancer-related genes are conserved and deregulated in human Mbs. Interestingly, a subset of them belongs to the “cell stress response” and has prognostic relevance in human disease.

We have provided a high-resolution analysis of previously unidentified MbSC transcriptomes and identified prognosis-related transcripts of human tumors as potential candidate therapeutic targets.

Results

MbSC, NcSC and Diff-NcSC deep-sequencing transcriptomes.

To identify MbSC- and NcSC-specific transcriptomes we used Mbs developed in *Ptch1*$^{+/-}$ mice, a genetically defined model of the SHH tumor. This murine model allowed us to analyze the paired normal stem cell counterpart (NcSC isolated from cerebella of wild type animals) from which MbSC are believed to derive (Wang J Exp Neurol. 2012). NcSC and *Ptch1*$^{+/-}$ MbSC were isolated and cultured as previously described (15). Both cells showed positive staining for the neural stem cell markers Sox2 and Nestin, (NcSC 70%, 85% and MbSC 90%, 70% respectively), with very low expression of Neurod1 or Math1 neuronal differentiation markers (Figure 1A and ref. 15). The Hh target gene Gli1 was detectable in 40% of NcSC (Figure 1B) and in 80% of MbSC, as expected (15, 16). NcSC gave rise to differentiated cells (Diff-NcSC) including astrocytes (S100^{+} cells), neurons (Tuj1^{+} cells), purkinje (Pvalb^{+} cells) and oligodendrocytes (Cspg4^{+}*)* when cultured in the presence of platelet-derived growth factor (PDGF) (Figure 1C).

We performed a paired-end, high-throughput mRNA-Seq on three biological replicates for each cell type (NcSC, Diff-NcSC and MbSC). On average, 124±39 millions reads per biological replicate (~ 85% of the sequenced ones) were aligned to the mouse genome (Figure 1 D; Table S1). Aligned reads were then assembled using the reference annotation based transcript assembly (RABT) from the Cufflinks suite, using the UCSC transcripts as reference (20,21,22). The abundance of each transcript was estimated in fragments per kilobase of exon per million fragments mapped (FPKM). The similarity of the FPKM distributions among different samples (Figure 1E) and the balance in the MA plots among biological replicates and cell types (Figure 2) confirm the absence of biases in the sequencing and/or normalization phases. The overall expression data of the biological replicates cluster together and are separated from the expression values of each cell types (Figure 1F), demonstrating that our sample preparation and sequencing approach steps reflect

the biology and expression profile of the cells under examination. The transcriptomes of all three analyzed samples show similar features. In all of them expressed genes are between 14 and 15 thousand, 86% of which are genes already annotated in the mouse reference transcriptome. They also show a similar proportion of alternative splicing (see Figure 1G). We also identified transcribed isoforms not annotated in the reference version of UCSC and in the following refer to them as putative novel isoforms. The transcribed regions not annotated in UCSC were first filtered to remove those overlapping with repeated regions or annotated in other databases (Ensembl, RefSeq, NONCODE) and defined as Novel Transcriptionally Active Regions (nTAR). Their classification according to their splicing patterns (single or poly exons structure) and their position with respect to known expressed loci (intronic or intergenic) is shown in Figure 1H. We found 157 intronic single exon nTARs with an average length of 489 bases; 614 intergenic nTARs with a single exon structure (average length of 492 bases) and 351 poly-exons nTARs with an average length of 2118 bases.

Long non-coding RNAs (lincRNAs) are also known to play a role in gene regulation (23). In our dataset we found 283 lincRNA annotated in UCSC, 51 of them subject to alternative splicing and 39 containing putative novel isoforms (24). As expected, the expression analysis of lincRNAs showed an overall lower expression level compared to protein-coding genes. The lincRNA and protein coding gene FPKM distributions have a different median (0.33 vs. 4.89) and upper quartile values (1.08 vs. 16.21) (Figure 3).

Overall, our RNA-seq deep sequencing approach yielded an accurate analysis of global stem and differentiated cell transcriptomes.

Classification of differentially expressed transcripts.

In order to characterize MbSC and NcSC transcriptomes, we performed a differential expression analysis using the Cuffdiff program. We identified 4898 significantly differentially expressed transcripts (DETs) among NcSC, MbSC and Diff-NcSC. Among DETs we observed 431 upregulated and 934 downregulated transcripts in MbSC with respect to NcSC. Diff-NSC displayed 1207 downregulated and 1734 upregulated transcripts with respect

to NcSC. When the transcriptomes of MbSC versus Diff-NcSC were analyzed, we detected 1206 upregulated in and 2066 downregulated DETs. We performed an unsupervised clustering of all annotated DETs in at least one out of the three comparisons (Figure 4A). The analysis shows that the transcripts can be grouped in two main clusters, one composed of differentiated cells, the other by stem-like cells (NcSC, MbSC) with a much higher internal diversity in the latter with respect to the former. This MbSC/NcSC group further separated into two consistent subgroups, one composed of NcSC, the other by MbSC (Figure 4A).

A similar cell-type specific pattern of expression was also observed by unsupervised clustering of putative novel isoforms (j) (Figure 5).

Since Hh pathway sustains both MbSC and NcSC, we searched for its transcripts by a hypergeometric test and found that MbSC and NcSC are enriched in transcripts belonging to Hh (Figure 3B). Notably, among Hh pathway transcripts, Gas1 and Hdac2 were dramatically upregulated in stem cells context. Gas1 is an Hh co-receptor required to enhance the pathway activity in multiple developmental contexts (25, 26) with an early role in cell fate specification of multiple neuronal progenitors (27). Hdac2 de-acetylates the Hh effectors Gli1 and Gli2, enhancing Hh signaling (28).

To verify the biological significance of Gas1 and Hdac2 in stemness, we abrogated their expression by specific siRNA. Our results show that Gas1 and Hdac2 silencing significantly reduced NcSC and MbSC self-renewal ability suggesting a role of these genes in neuronal and cancer stem cell maintenance (Figure 3C and D).

Searching for molecular networks linking DETs, we used Gene Ontology (GO) enrichment analysis. Figure 6 schematically summarizes the heatmaps of DETs belonging to KEGG pathways that characterize the cells. The first heatmap (a) “neuroactive ligand-receptor interaction” include the “gamma-aminobutyric acid (GABA)/glutamate/adrenergic receptor family” (*Gabra3, 4, Gabrb1, 3, Gria1, 2, 4, Grin1, 2a, 3a, Adra1a, Chrna4*), down-regulated in MbSCs vs NcSC. Among dowregulated genes in MbSCs other functional categories are enriched, such as “potential regulation”, “synaptic organization/transmission” and

"glutamate/ligand-gated ion channel activity" as well as the KEGG pathways "calcium signaling" and "axon guidance". Conversely, upregulated genes in MbSCs significantly enrich the category of "amino-acid biosynthesis".

Panels b, c and d show transcripts of "cell cycle" (b), "DNA replication" (c) and "p53 pathway" (d) networks. These genes are overrepresented in stem cell context and are involved in "mitotic and cell cycle activities" (e.g. cyclins Ccne2, Ccne1, Ccnd1, Ccnb2, Ccnd2, Ccna2), "nuclear division", "DNA replication" (e.g. mini-chromosome complex and helicases Mcm2-7, Pola1, Pole, Pole2, Dna2), "DNA repair", "mitotic cell cycle checkpoints" (Cdk1, Ccnd1, Mad2l1, Cep192) and "metabolic processes" (Cbs, Nob1, Rps6, Dis3, Ebna1bp2, Ddx56, Pa2g4, Shmt1, 2).

Panel e and f, "ECM-receptor interaction" and "focal adhesion" respectively, highlighted transcripts upregulated in Diff-NcSC. Among these, others enrich functional categories belonging to "extracellular matrix", "cell adhesion", "cell cycle arrest", "synaptic organization", "membrane organization" and "extracellular region part" "calcium ion binding" (S100a1, 3, 4, 6, 11, 13, S100b) and "integrins" (Itga2, 4, 7, Itgb4, 8) (Figure 6).

Gene patterns characterizing MbSC.

To define sub-groups of transcripts characterizing cell subtypes, we performed subset analysis by Venn diagrams to pick up DETs specifically upregulated in the cell contexts (Figure 7). The first subgroup is composed of 161 transcripts, which are downregulated in both MbSC and Diff-NcSC compared to NcSC. We hypothesized that this group includes genes specifically expressed in normal stem cells (Figure 7A). The second sub-group includes 111 transcripts specifically upregulated only in Diff-NcSC compared to NcSC and reciprocally downregulated in MbSC, assumed to be the genes expressed in normal differentiated cells (Figure 7B). Finally, we selected two additional subsets of genes (MbSC1 and MbSC2) highly expressed in MbSC compared to NcSC: the MbSC1 subset is characterized by 11 transcripts specifically upregulated in MbSC, compared to NcSC, and reciprocally downregulated in Diff-NcSC, thus assumed to be the genes highly

expressed only in MbSC (Figure 7C and D, left box). This latter subset includes 7 known transcripts and 4 novel isoforms. The MbSC2 is assumed to feature cancer properties of MbSC and includes 30 DETs (Figure 7D, right box) selected according to the following criteria: i) highly expressed (FPKM > 15), ii) significantly up-regulated in MbCSC versus NcSC (adjusted p value <0.05, log FC >1.5), iii) not significantly upregulated in Diff-NSC compared to NcSC, iv) published to be involved in various cancers. MbSC1 and MbSC2 subsets are thus suggested to play a role in both stemness and cancer behavior.

Among them, we noticed specific genes involved in mouse Mb formation (i.e. the Notch ligand Jagged-1, the chemokine receptor CXCR4 and the angiogenic VEGFA) (29,30,31) or Hh-dependent development (i.e. neuropilin-1) (32).

MbSC transcriptome correlates with clinical outcome of human medulloblastomas.

To understand whether the above described gene networks identified in Mb mouse model could be conserved in human tumors, we queried GeneMANIA (33) and identified putative human networks involved in pathophysiology. Notably, ATF4, TRIB3, DDIT3 and ASNS are involved in **"response to stress**" (Figure 7E), while CXCR4, JAG1 and NRP2 are linked together and with the CXCR4 ligand CXCL12, FLT4 and VEGFA and C in controlling cell locomotion (Figure 8A). These observations suggest that both MbSC1 and MbSC2 signatures might be involved in human Mb.

Therefore, we evaluated the selected MbSC specific transcripts (listed in figure 7D) in human Mb samples (n=36). Details of patients included in the study are reported in Table 1. Mb patients were stratified according to their risk subgroups (based on patients' age, metastasis and post-surgical residual) (34). Twenty-three patients were high risk (HR) (< 3 year age, metastatic and residual tumor mass) and 13 were average risk (AR). Patients were also stratified with respect to molecular subgroups (15 SHH + 21 non-SHH) (13,14).

We first applied an unsupervised clustering algorithm based on non-negative matrix factorization (NMF) (35) that identified aggregate pattern of gene expression analyzed in human samples. We obtained 2 or

8 compact clusters as judged by the cophenetic value. In both cases one of the clusters is significantly enriched in Shh-patients (Figure 8B,C and D).

Twenty-three out of 37 genes (62%) are significantly higher in human Mbs (Fig. 9A), mirroring similar features observed in mouse model.

To get insights into the biological relevance of the differentially expressed gene signature, we investigated whether any of the selected genes could have a prognostic value. Patients were divided into two subgroups (containing at least five patients) according to whether the expression of the selected gene was more than 3 standard deviation above the median of the controls or not. Interestingly, Kaplan-Meier (KM) analysis of the two groups indicated a worse prognosis with a shorter survival rate of ATF4 high expressing patients (regardless of the Mb subgroup classification) compared to patients with lower transcript levels (Fig. 9B). To investigate whether such a gene signature may also represent a further prognostic indicator besides the currently known ones (i.e. HR versus AR), we performed KM analyses in HR patients. High levels of ATF4, GADD45a and ERO1L also correlated with a worse prognosis in HR patients (Fig. 9 C,D and E). Overall, ATF4, ERO1L and GADD45a belonging to the "**cell stress response**", correlate with clinical outcome of Mb patients.

Interestingly, this gene signature extends to human MbSC, as well. Indeed the levels of these genes are significantly overexpressed in human MbSC with respect to the bulk pre-neurosphere cell population (Figure 9F).

Discussion

We describe here the first report of the use of a large scale RNA-seq technology of mouse MbSC revealing their previously unknown transcriptional landscape. We have obtained high-throughput trascriptome that meets best requirements (36) for the measurement of genes and identification of alternative splicing as well as analysis for detection of noncoding transcripts.

Differential expression analysis shows that MbSC and NcSC share “stem-like” gene patterns that make them a cluster distinct from a second cluster corresponding to Diff-NcSC. According to the $Ptc1^{+/-}$ Mb model studied, we found that the Hh pathway is enhanced in stem context, confirming previous findings (15). More specifically, we reveal a link of NcSC/MbSC with Gas1, a target of Hh signaling and also a Hh co-receptor required to enhance the pathway activity (25,26,27), that we show here to promote self-renewal. We also report the upregulation of Hdac2, which regulates the acetylation of the Hh Gli effectors enhancing their activity and confirming its functional role in Mb (28). Similar considerations can be made for MbSC-associated upregulation of the transmembrane protein Nrp2, an important enhancer of Hh signaling forming a positive feedback circuit (32).

Gene signatures defining SHH subgroup human primary Mbs were previously identified to include transcripts that we found to be upregulated in MbSC and NcSC. These transcripts include *HHIP* (8,14), *SFRP1* (9,14), *MICAL1* (37), *BCHE* (36), *NTRK3* (8) and *SPHK1* (8). Insm1, a target of Hh (38), is also enriched in stem context.

Unsupervised clustering allowed us to identify DETs specifically upregulated in mouse MbSC, composing a subset of genes which correlates with stemness, cancer or both features. Notably, these transcripts were conserved to a large extent in a cohort of human Mb. Most importantly, some of them correlate with tumor prognosis, thus attributing a clinical validation to our data.

More specifically, a prognostic value can be attributed to ATF4, GADD45a and ERO1L transcripts that distinguish mouse MbSC from normal NcSC and are not downregulated in differentiated cells, which are therefore likely to mark MbSC with some cancer-associated differentiation-related features.

Remarkably, ATF4, GADD45A and ERO1L are also upregulated in MbSC isolated from human primary Mbs, with respect to bulk pre-neurosphere cell population and encode proteins involved in the response to oncogenic and metabolic stress. This is in line with the observations that tumor progression and metastatic activity need to overcome a series of microenvironmental obstacles, in order to adapt escaping mechanisms and promote disease recurrence. Among these obstacles, the disordered microvasculature impairs oxygen and nutrient supply while genotoxic stress is caused by chemo/radiotherapy. To this regard, ATF4 overexpression is observed in response to a diverse array of micro-environmental stresses including amino acid depletion, oxidative stress and endoplasmic reticulum (ER) stress. It also characterizes tumors under hypoxic- and nutrient-deprived conditions, where it promotes metabolic homeostasis and cancer cell survival by transcriptional regulation of amino acid uptake and biosynthesis, autophagy, redox balance, angiogenesis and resistance to current chemotherapeutic drugs (39). Interestingly, the analysis of transcripts expressed in mouse MbSC and conserved in human primary medulloblastomas, reveals a number of genes that functionally interact with ATF4 either as activated negative loops (i.e. TRIB3 and DDIT3) (40) or survival response to anoxia (i.e. VEGF) (41). An imbalance of these signals is therefore suggested to be involved in the ATF4-controlled clinical outcome that we have described.

ER is very sensitive to tumor-associated hypoxia, resulting in the unfolded protein response-linked ER-stress, critically regulated by ATF4 (42). Notably, we observed that the ER oxidoreductin 1 (ERO1L) also plays a prognostic role. ERO1L is the primary disulfide bond-generating catalyst in the ER and plays an essential role in protein folding, whose subversion is involved in diseases such as cancer. Indeed, by oxidizing Protein Disulfide Isomerase to keep its function, ERO1L prevents the suppression of disulfide bond formation caused by hypoxic stress. ERO1L overexpression has been reported as an adaptive response to the prevalent hypoxic

conditions of growing tumor cells, whereby hypoxia enhances HIF-1-mediated transcriptional upregulation of ERO1L that in turn promotes appropriate VEGF protein folding and secretion, thus favoring tumor angiogenesis and progression (43).

Finally, the member of the stress sensor GADD45 family, GADD45a, is implicated in the response to physiological and environmental stress. These include oncogenic stress, which interact in a complex manner with other cellular proteins involved in cell cycle regulation and stress response (such as PCNA, p21, cdc2/cyclinB1, and the p38 and JNK stress response kinases) (44). The result can be cell cycle arrest, DNA repair, cell survival, senescence, and apoptosis. Altered expression of GADD45 has been observed in multiple types of solid tumors and hematopoietic malignancies, where it can function to either promote or suppress tumor development, depending on the activating oncogene, the cell type and the engagement of different signaling pathways (44). For instance, GADD45a suppresses RAS-driven tumorigenesis via Jnk-mediated apoptosis and p38-mediated senescence. In contrast GADD45a promotes Myc-driven tumorigenesis via GSK3b/b-catenin signaling which suppresses MMP10 expression, resulting in increased tumor vascularization, decreased apoptosis and senescence, ultimately accelerating tumor growth (44). Notably, c-MYC is known to promote Mb formation in vivo and to confer stemness and mitogenic properties (45, 46). Indeed, we observed that c-myc is is highly expressed in MbSC compared to NcSC (Figure 8). Accordingly, several c-myc target genes (i.e. ASNS, ATF4, EMP1, FBLN5, GADD45A, HK2, MTHFD2, NRP2, PRPH, SLC1A4, DDIT3, ERO1L, TRIB3) belong to the signature we have observed to be upregulated in MbSC (http://www.myccancergene.org/site/myctargetdb.asp).

Overall, our data suggest that the oncogenic activation of mechanisms driving the adaptation to metabolic and micro environmental stress plays a major role in enhancing the aggressiveness of MbSC.

In summary, we have shown that our model system is effective in prioritizing potential molecular mechanisms of human Mb. Moreover we provided evidence for the existence of a new subset of genes useful for patients' stratification to predict clinical outcome.

Materials & Methods

Stem Cell cultures and treatments

Neural Stem Cells (NcSC) and Medulloblastoma Cancer Stem Cells (MbSC) were isolated and cultivated as already described (15). More in detail NcSC were isolated from murine WT postnatal cerebella (P4). MbSC were isolated from fresh tumor specimens obtained from Ptch1+/- mice, which develop Sonic Hedgehog-dependent medulloblastoma (16). Both NcSC and MbSC were obtained after mechanical and enzymatic dissociation and cultured in serum-free DMEM-F12 plus glucose 0.6%, insulin 25 mg/ml, N-acetyl-L-cystein 60 mg/ml, heparin 2 mg/ml, B27 1x, EGF 20 ng/ml and bFGF 20 ng/ml, to maintain them under proliferative conditions. NSC were also treated to differentiate in vitro after withdrawal of EGF/bFGF and addition of differentiating factors (Platelet Derived Growth Factor, PDGF) for 4 days (15). For the neurosphere -forming assay, cells were plated at clonal density (1–2 cells/mm2) into 96-well plates and cultured in selective medium as described above.

Knockdown of mouse Gas1 and Hdac2 was performed with specific smart pool siRNAs (20 nM) (Dharmacon, Lafayette, CO) using interferin siRNA Transfection Reagent (Polyplus Transfection, NY).

Immunofluorescence

Neurospheres were dissociated mechanically in cell dissociation solution non-enzymatic buffer (Sigma) to obtain aggregates of around 5-10 cells and plated on poly-lysine-coated Lab-Tek chamber slides (cover slips) and allowed to adhere for 3 h or differentiated for 3 days as described above to detect differentiation markers. Cells were fixed with 4% paraformaldehyde for 20 min at room temperature, incubated in blocking solution (5% normal goat serum (NGS), 1% BSA, 0.1% Triton X-100) and stained overnight with primary antibodies diluted in blocking solution and for 2 h with secondary antibodies. Primary antibodies were mouse anti-Sox2

(MAB4343 Millipore Billerica, MA), mouse anti-Nestin (Abcam ab 11306); mouse anti-Gli1, #2643 (Cell Signaling Technology Inc); mouse anti-Parvalbumin P3088 (Sigma); rabbit anti- S100 S2644 (Sigma); mouse antib-III tubulin (MAB1637) (Millipore). In all, 488-conjugated anti-mouse and antirabbit secondary antibodies were purchased from Molecular Probes (Invitrogen, Eugene, OR). Nuclei were counterstained with HoecHst reagent. Cover slips were mounted with fluorescence mounting medium (S3023) (Dako, Carpinteria, CA). Images were acquired with Carl Zeiss microscope (Axio Observer Z1) using Apotome technology and AxioVision Digital Image Processing Software.

***Western blot assay*.**

Cells were lysed in Tris–HCl pH 7.6, 50mM, deoxycholic acid sodium salt 0.5%, NaCl 140 mM, NP40 1%, EDTA 5mM, NaF 100mM, Na pyrophosphate 2mM and protease inhibitors. Lysates were separated on 8% acrylamide gel and immunoblotted using standard procedures. Rabbit anti-Gas1 (Santa Cruz Biotechnology, sc-33175), rabbit anti- Hdac2 (Santa Cruz Biotechnology, sc-7899), rabbit anti- Jagged1 (Cell Signaling # 2620), rabbit anti-HSP70, Santa Cruz Biotechnology, sc-33575), rabbit anti-c-myc (Cell Signalling #9402) and HRP-conjugated secondary antisera (Santa Cruz Biotechnology) were used followed by enhanced chemiluminescence (ECL Amersham, Amersham, UK).

Libraries preparation and RNA sequencing

RNA extraction. Total RNA from biological replicates of mNSC, differentiated NSCs and mMbSCs was extracted using Trizol reagent (Life technologies, USA) according to the manufactures' protocol. The quality of total RNA was checked on BioAnalyzer (Agilent) using RNA nano kit according to the manufactures' instructions. Only RNA with RIN value > 9 were used for libraries preparation.

mRNAseq libraries were constructed according to the standard Illumina protocol by using TruSeq RNA Sample Prep Kit. mRNA was converted from total RNA into a library of template molecules suitable for subsequent

deep sequencing analysis. Briefly, in a first step, poly-dA containing mRNA molecules were purified using poly-dT oligo-attached magnetic beads. Following purification, mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments went through an end-repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were purified and enriched by PCR to create the final cDNA library.

Sequencing data analysis

Mapping of RNA-Seq reads using TopHat. All sequenced reads (70 bases in paired-ends) were aligned to the *Mus musculus* genome (NCBI37/mm9) using the spliced read aligner TopHat version V1.3.1. We ran two iterations of TopHat alignments to maximize the use of splice site information derived across all samples (47). To this purpose, the reads from each sample were first aligned using the paired-end alignment option when possible (default parameters and 'min-anchor=5'). Next, we generated a pooled splice-sites (or "junctions") file by combining all predicted splice sites across all alignments. We then re-aligned each sample using the pooled splice-sites file without the option to search for new splice junctions. This strategy led to the reconstruction of about 80% of the junctions annotated in UCSC (Genes knownGene) (173,464).

RNA-Seq transcriptome assembly. The transcriptome of each sample was assembled separately from the mapped reads using Cufflinks V1.2.1 with default parameters and paired-end information with UCSC reference gtf annotation. If more than one lane per sample was available the corresponding aligned reads were pooled prior to running the assemblers. Similarly to other studies we established an FPKM threshold to define whether a transcript is expressed using the lower bound of the expression distributions (Figure 1E) (48).

Differential expression test. Once all reads were assembled with Cufflinks, the output was carefully cleaned by non-expressed transcripts and transcripts not present in at least two out of the three biological replicas to avoid

assembly artifacts and we used Cuffcompare to produce a combined annotation file subsequently used as input for Cuffdiff along with the original alignments files produced by TopHat. Cuffdiff was used to re-estimates transcript abundance and differentially expressed transcripts (DETs) and to perform pairwise comparisons of expression, splicing and promoter usage between NSC/CSC and NSC/DIF.

Functional Analysis. We used the DAVID gene functional classification tool (ref). In our analysis we considered only GO categories with a false discovery rate < 0.05 and pathways with Benjamini correction < 0.05. (49,50).

Clustering analysis. Unsupervised clustering and heatmap were generated in R by using differentially expressed transcripts as input. We used the R package Pvclust to generate sample clustering with 10000 bootstrap replicas and heatmap.2 to generate the heatmaps. (51)

Human samples.

Thirty-six human primary medulloblastoma patients' for whom a 5-years clinical follow-up was available were used in this study. Tumor specimens were collected during surgical resection with the approval of Institutional Review Board, as previously described (51). RNA of normal human cerebellum (four adult samples from 25- to 70-year-old subjects were purchased from Biocat (Heidelberg, Germany) and Ambion (LifeTech, Foster City, CA). Human MBs were further stratified according the risk subgroups (based on patients age, metastasis and post-surgical residual) (34). Twenty-three patients were high risk (HR) (< 3 year age, metastatic and residual tumor mass) and 13 were average risk (AR). Finally, patients were stratified with respect to tumors "molecular subgroups" (15 SHH + 21 non-SHH) by realt-time PCR evaluating the expression levels (with respect to normal controls) of the following genes: GLI1, SFRP1, PTCH2, MICAL1, RAB33A, PDLIH3, BCHE, ITIH2 (ITIH2-Hs00158297_m1; BCHE-Hs00163746_m1; PTCH2-Hs00184804_m1, MICAL1-Hs00917696_m1; RAB33A-Hs-00191243_m1; PDLIM3-Hs01062534_m1; GLI1-Hs00171790_m1; HHIP-Hs01011009_m1; SFRP1-Hs00610060_ m1, performed in triplicates. Patients were considered to belong to SHH subgroup if all the analyzed genes were expressed at least three standard deviations over the average of normal cerebellar

controls. (14, 37). All the Clinical and molecular features of patients included in the study are schematized in Table 1.

Quantitative RT-PCR

cDNA synthesis was performed using High Capacity cDNA reverse transcription kit from Applied Biosystems-Life Technologies (Foster City, CA) as previoly described (53). Quantitative reverse transcription (RT–PCR) analysis was performed using a custom 384-TaqMan Low Density Array (TLDA) containing triplicates for the following assay: ADAM15-Hs00984794_m1, ADAM19-Hs00224960_m1, CTH-Hs00542276_m1, FBLN5-Hs00197064_m1, GABRR2-Hs00266703_m1, MTHFD2-Hs00759197_s1, P2RX3-Hs01125554_m1, SLC1A5-Hs00194540_m1, ADAM15-Hs00984794_m1, ALDH1L2-Hs00402876_m1, ARHGEF2-Hs00190884_m1, ASNS-Hs00370265_m1, ATF4-Hs00909569_g1, ATF5-Hs01119208_m1, CA14-Hs00201626_m1, CXCR4-Hs00607978_s1, DDIT3-Hs00358796_g, DDR2-Hs01025953_m1, DUSP4-Hs01027785_m1,EMP1-Hs00608052_m1, ERO1L-Hs00205880_m1, GADD45A-Hs00169255_m1, GTF3A-Hs00157851_m1, GTPBP2-Hs00368050_m1, HK2-Hs00606086_m1, JAG1-Hs01070032_m1, LONP1-Hs00998404_m1, NRP2-Hs00187290_m1, NUPR1-Hs01044304_g1, PODXL-Hs01574644_m1, PRPH-Hs00196608_m1, RCOR2-Hs00293511_m1, SLC3A2-Hs00374243_m1,STC2-Hs00175027_m1,TEAD2-Hs00366217_m1, TRIB3-Hs01082394_m1, VEGFA-Hs00900055_m1, VLDLR-Hs01045922_m1. mRNA expression was analyzed on cDNAs using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems) using TaqMan gene expression assay according to the manufacturer's instructions (Applied Biosystems). Each amplification reaction was performed in triplicate, and the average of the three threshold cycles was used to calculate the amount of transcripts in the sample (SDS software, AB). mRNA quantification was expressed, in arbitrary units, as the ratio of the sample quantity to the calibrator or to the mean values of control samples. All values were normalized to 4 endogenous controls, PGK1, GAPDH, ß-ACTIN and HPRT.

Shh-type MBs were assessed by real-time PCR of the following assay ITIH2-Hs00158297_m1; BCHE-

Hs00163746_m1; PTCH2-Hs00184804_m1, MICAL1-Hs00917696_m1; RAB33A-Hs-00191243_m1; PDLIM3-Hs01062534_m1; GLI1-Hs00171790_m1; HHIP-Hs01011009_m1; SFRP1-Hs00610060_ m1, performed in triplicates (36).

***Statistical analysis*.**

Where not differently indicated, statistical analysis was performed using StatView 4.1 software (Abacus Concepts, Berkeley, CA). The Mann–Whitney U test for unpaired data was used to analyze differences in gene expression of each gene between cancer patients and controls. The results are expressed as mean±S.D. from at least three experiments as indicated in the figure legends.

Survival analysis

Kaplan–Meier analysis was used to determine the survival effect based on gene expression data. Patients were divided into two subgroups (containing at least five patients) according to whether the expression of the selected gene signature was more than 3 standard deviation above the median of the controls or not.

The statistical significance of Kaplan–Meier curves was evaluated with the log-rank test.

MNF methods

Non-negative matrix factorization (NMF) is an efficient and powerful method for the identification of distinct molecular patterns and for class discovery (35). It is an algorithm based on decomposition by parts that can reduce the dimension of expression data from thousands of genes to a handful of metagenes. We could obtain 2 or 8 compact clusters as judged by the cophenetic value provided by the NMF method. In both cases one of the clusters is significantly enriched in Shh-patients.

References

1. Vermeulen L, de Sousa e Melo F, Richel DJ, Medema JP. The developing cancer stem-cell model: clinical challenges and opportunities. Lancet Oncol. 2012 Feb;13(2):e83-9. doi: 10.1016/S1470-2045(11)70257-1. Review.

2. Wang J, WecHsler-Reya RJ. The role of stem cells and progenitors in the genesis of medulloblastoma. Exp Neurol. 2012 Nov 22. doi:pii: S0014-4886(12)00431-1.

3. Shats I, Gatza ML, Chang JT, Mori S, Wang J, et al. (2011) Using a stem cell based signature to guide therapeutic selection in cancer. Cancer Res 71: 1772–1780.

4. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, et al. (2008) An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 40: 499–507.

5. Wang T, Ong CW, Shi J, Srivastava S, Yan B, et al. (2011) Sequential expression of putative stem cell markers in gastric carcinogenesis. British Journal of Cancer 105: 658–665.

6. Bauer L, Langer R, Becker K, Hapfelmeier A, Ott K, et al. (2012) Expression Profiling of Stem Cell-Related Genes in Neoadjuvant-Treated Gastric Cancer: A NOTCH2, GSK3B and b-catenin Gene Signature Predicts Survival. PLoS ONE 7(9): e44566.

7. Northcott PA, Jones DT, Kool M, Robinson GW, Gilbertson RJ, Cho YJ, Pomeroy SL, Korshunov A, Lichter P, Taylor MD, Pfister SM. Medulloblastomics: the end of the beginning. Nat Rev Cancer. 2012 Dec;12(12):818-34.

8. Northcott P a et al. (2011) Medulloblastoma comprises four distinct molecular variants. *Journal of clinical oncology:*29:1408-14.

9. Cho Y-J et al. (2011) Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *Journal of clinical oncology :* 29:1424-30.

10. Northcott PA, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature. 2012 Aug 2;488(7409):49-56. doi: 10.1038/nature11327.

11. Jones DT, et al. Dissecting the genomic complexity underlying medulloblastoma. Nature. 2012 Aug 2;488(7409):100-5. doi: 10.1038/nature11284.

12. Pugh TJ, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature. 2012 Aug 2;488(7409):106-10. doi: 10.1038/nature11329.

13. Taylor MD, Northcott PA, Korshunov A, Remke M, Cho YJ, Clifford SC, Eberhart CG, Parsons DW, Rutkowski S, Gajjar A, Ellison DW, Lichter P, Gilbertson RJ, Pomeroy SL, Kool M, Pfister SM. Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol. 2012 Apr;123(4):465-72.

14. Northcott, PA. et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol. 123, 615–626 (2012).

15. Po A et al. (2010) Hedgehog controls neural stem cells through p53-independent regulation of Nanog. *the The European Molecular Biology Organization Journal* 29:2646-2658.

16. Goodrich LV, Milenković L, Higgins KM, Scott MP. Altered neural cell fates and medulloblastoma in mouse patched mutants. Science. 1997 Aug 22;277 (5329):1109-13.

17. Yang ZJ, Ellis T, Markant SL, Read TA, Kessler JD, Bourboulas M, Schüller U, Machold R, Fishell G, Rowitch DH, Wainwright BJ, Wechsler-Reya RJ. Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. Cancer Cell. 2008 Aug 12;14(2):135-45.

18. Schüller U, Heine VM, Mao J, Kho AT, Dillon AK, Han YG, Huillard E, Sun T, Ligon AH, Qian Y, Ma Q, Alvarez-Buylla A, McMahon AP, Rowitch DH, Ligon KL. Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. Cancer Cell. 2008 Aug 12;14(2):123-34

19. Takebe N, Harris PJ, Warren RQ, Ivy SP. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. Nat Rev Clin Oncol. 2011 Feb;8(2):97-106.

20. Trapnell C et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28:511-515.

21. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25.

22. Li H et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19505943

23. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat Rev Neurosci. 2012 Jul 20;13(8):528-41. doi: 10.1038/nrn3234

24. Cabili MN et al. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*. 15.

25. Izzi L, Lévesque M, Morin S, Laniel D, Wilkes BC, Mille F, Krauss RS, McMahon AP, Allen BL, Charron F. Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation. Dev Cell. 2011 Jun 14;20(6):788-801. doi: 10.1016/j.devcel.2011.04.017.

26. Martinelli DC, Fan CM. Gas1 extends the range of Hedgehog action by facilitating its signaling. Genes Dev. 2007 May 15;21(10):1231-43.

27. Allen BL, Song JY, Izzi L, Althaus IW, Kang JS, Charron F, Krauss RS, McMahon AP. Overlapping roles and collective requirement for the coreceptors GAS1, CDO, and BOC in SHH pathway function.Dev Cell. 2011 Jun 14;20(6):775-87. doi: 10.1016/j.devcel.2011.04.018.

28. Canettieri G (2010),Di Marcotullio L, Greco A et al. Histone Deacetylase And Cullin3-Ren(Kctd11) Ubiquitin Ligase Interplay Regulates Hedgehog Signalling Through Gli Acetylation. Nature Cell Biology, 12; P. 132-142.

29. Hallahan AR, Pritchard JI, Hansen S, Benson M, Stoeck J, Hatton BA, Russell TL, Ellenbogen RG, Bernstein ID, Beachy PA, Olson JM. The SmoA1 mouse model reveals that notch signaling is critical for the growth and survival of sonic hedgehog-induced medulloblastomas. Cancer Res. 2004 Nov 1;64(21):7794-800.

30. Sengupta R, Dubuc A, Ward S, Yang L, Northcott P, Woerner BM, Kroll K, Luo J, Taylor MD, Wechsler-Reya RJ, Rubin JB . CXCR4 activation defines a new subgroup of Sonic hedgehog-driven medulloblastoma Cancer Res. 2012 Jan 1;72(1):122-32. doi: 10.1158/0008-5472.CAN-11-1701. Epub 2011 Nov 3.

31. Slongo ML, Molena B, Brunati AM, Frasson M, Gardiman M, Carli M, Perilongo G, Rosolen A, Onisto M. Functional VEGF and VEGF receptors are expressed in human medulloblastomas Neuro Oncol. 2007 Oct;9(4):384-92. Epub 2007 Aug 17.

32. Hillman RT, Feng BY, Ni J, Woo WM, Milenkovic L, Hayden Gephart MG, Teruel MN, Oro AE, Chen JK, Scott MP. Neuropilins are positive regulators of Hedgehog signal transduction. Genes Dev. 2011 Nov 15;25(22):2333-46. doi: 10.1101/gad.173054.111. Epub 2011 Nov 3.

33. Warde-Farley D et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* 38:W214-W220.

34. Gajjar A, Hernan R, Kocak M, Fuller C, Lee Y, McKinnon PJ, Wallace D, Lau C, Chintagumpala M, Ashley DM, Kellie SJ, Kun L, Gilbertson RJ. Clinical, histopathologic, and molecular markers of prognosis: toward a new disease risk stratification system for medulloblastoma. J Clin Oncol. 2004 Mar 15;22(6):984-93. Epub 2004 Feb 17.

35. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A. 2004 Mar 23;101(12):4164-9. Epub 2004 Mar 11.

36. .Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet. 2012 Sep;13(9):667-72.

37. Schwalbe EC, Lindsey JC, Straughton D, Hogg TL, Cole M, Megahed H, Ryan SL, Lusher ME, Taylor MD, Gilbertson RJ, Ellison DW, Bailey S, Clifford SC. Rapid diagnosis of medulloblastoma molecular subgroups. Clin Cancer Res (2011) 17:1883–1894.

38. De Smaele E, Fragomeli C, Ferretti E, Pelloni M, Po A, Canettieri G, Coni S, Di Marcotullio L, Greco A, Moretti M, Di Rocco C, Pazzaglia S, Maroder M, Screpanti I, Giannini G, Gulino A. An integrated approach identifies Nhlh1 and Insm1 as Sonic Hedgehog-regulated genes in developing cerebellum and medulloblastoma. Neoplasia. 2008 Jan;10(1):89-98..

39. Singleton DC, Harris AL. Targeting the ATF4 pathway in cancer therapy. Expert Opin Ther Targets. 2012;16(12):1189-202.

40. Rzymski T, Paantjens A, Bod J, Harris AL. Multiple pathways are involved in the anoxia response of SKIP3 including HuR-regulated RNA stability, NF-kappaB and ATF4. Oncogene. 2008 Jul 31;27(33):4532-43

41. Oskolkova OV, Afonyushkin T, Leitner A, von Schlieffen E, Gargalovic PS, Lusis AJ, Binder BR, Bochkov VN. ATF4-dependent transcription is a key mechanism in VEGF up-regulation by oxidized phospholipids: critical role of oxidized sn-2 residues in activation of unfolded protein response. Blood. 2008 Jul 15;112(2):330-9. doi: 10.1182/blood-2007-09-112870.

42. Rzymski T, Milani M, Pike L, Buffa F, Mellor HR, Winchester L, Pires I, Hammond E, Ragoussis I, Harris AL Regulation of autophagy by ATF4 in response to severe hypoxia. Oncogene. 2010 Aug 5;29(31):4424-35. doi: 10.1038/onc.2010.191. Epub 2010 May 31.

43. May D, Itin A, Gal O, Kalinski H, Feinstein E, Keshet E. Ero1-L alpha plays a key role in a HIF-1-mediated pathway to improve disulfide bond formation and VEGF secretion under hypoxia: implication for cancer. Oncogene. 2005 Feb 3;24(6):1011-20.

44. Liebermann DA, Tront JS, Sha X, Mukherjee K, Mohamed-Hadley A, Hoffman B. Gadd45 stress sensors in malignancy and leukemia. Crit Rev Oncog. 2011;16(1-2):129-40.

45. .Pei Y, Moore CE, Wang J, Tewari AK, Eroshkin A, Cho YJ, Witt H, Korshunov A, Read TA, Sun JL, Schmitt EM, Miller CR, Buckley AF, McLendon RE, Westbrook TF, Northcott PA, Taylor MD, Pfister SM, Febbo PG, Wechsler-Reya RJ. An animal model of MYC-driven medulloblastoma. Cancer Cell. 2012 Feb 14;21(2):155-67.)

46. Park AK, Lee SJ, Phi JH, Wang KC, Kim DG, Cho BK, Haberler C, Fattet S, Dufour C, Puget S, Sainte-Rose C, Bourdeaut F, Grill J, Delattre O, Kim SK, Park WY. Prognostic classification of pediatric

medulloblastoma based on chromosome 17p loss, expression of MYCC and MYCN, and Wnt pathway activation. Neuro Oncol. 2012 Feb;14(2):203-14.

47. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* (Oxford, England) 25:1105-11.
48. Bolker B et al. (2009) The gplots Package. *Text*.
49. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4:44-57.
50. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A (2011) Differential expression in RNA-seq: A matter of depth. *Genome Research* 21:2213-2223.
51. Bootstrap M et al. (2009) Package ' pvclust. *Bioconductor* 96:94-102.
52. Ferretti E (2008), De Smaele, E., Miele, E et al. Concerted microRNA control of Hedgehog signalling in cerebellar neuronal progenitor and tumour cells. Embo J.
53. Ferretti E (2009), De Smaele E, et al.. MicroRNA profiling in human medulloblastoma. Int J Cancer.124, 568-77.