Biotechnology Research and Innovation (2019) xxx, xxx–xxx

SBBiotec
SOCIEDADE BRASILEIRA DE BIOTECNOLOGIA

Biotechnology
Research & Innovation

http://www.journals.elsevier.com/biotechnology-research-and-innovation/

RESEARCH PAPER

Transcriptome analysis and gene networks in a rare pediatric tumor

Michel L. Leite[a], **Elio F. Vanin**[b,1], **Stephen Iannaccone**[b,c], **Nicolau B. da Cunha**[a], **Sérgio de Alencar**[a,*], **Fabricio F. Costa**[a,b,c,d,e,*]

[a] *Laboratory of Data Science, Genomic Sciences and Biotechnology Program, UCB, Brasilia, SGAN 916 Modulo B, Bloco C, 70.790-160, Brasília, DF, Brazil*
[b] *Cancer Biology and Epigenomics Program, USA*
[c] *Developmental Biology, Ann & Robert H Lurie Children's Hospital of Chicago Research Center and Northwestern University's Feinberg School of Medicine, 2430 N. Halsted St., Box 220, Chicago, IL, USA*
[d] *The Polsky Center for Entrepreneurship and Innovation, University of Chicago, 1452 E 53rd St, Chicago, IL, USA*
[e] *MATTER Chicago, 222 W. Merchandise Mart Plaza, Suite 12th Floor, Chicago, IL, 60.654, USA*

Received 11 April 2019; accepted 15 August 2019

KEYWORDS
Ependymomas;
Transcriptome analysis;
Expression;
Big data analytics;
Epigenetic regulation;
Potential biomarkers.

Abstract Ependymoma, a rare pediatric tumor originating from ependymal cells located in the lining of ventricular surfaces in the brain, presents great challenges in treatment despite advances in neurosurgical techniques. In order to identify new molecular biomarkers that could improve clinical management and outcomes, we have used RNA-seq to profile the Whole Transcriptome of three ependymoma samples and one normal control brain tissue, producing a total of 2.5 Gigabases of sequencing data. Different protein-coding gene databases were interrogated for known annotated genes to calculate RPKM and Fold Changes (FGs) for each clustered gene. Using this approach, we were able to identify Differentially Expressed Genes (DEGs) in ependymomas. KEGG pathways and Gene Ontology (GO) categories enriched in the DEGs were analyzed using Enrichr, and protein interaction networks were then built using MetaCore™. Thirty differentially expressed protein-coding genes associated with neurogenesis were identified by TaqMan Real-Time PCR, 17 of these showing statistically significant differential expression. Based on these results, we have identified *IGF-2* as highly over-expressed in ependymomas and that this is due to loss of DNA methylation in its promoter region. In conclusion, we believe that some of these genes, specially *IGF-2*, may be of clinical importance, opening new avenues for disease management and new therapies.

* Corresponding authors at: Laboratory of Data Science, Genomic Sciences and Biotechnology Program, UCB - SGAN 916 Modulo B, Bloco C, 70.790-160, Brasília, DF, Brazil.
E-mails: sergiodealencar@gmail.com (S. Alencar), fcosta@genomicenterprise.com (F.F. Costa).
[1] In Memoriam.

https://doi.org/10.1016/j.biori.2019.08.002

Introduction

Ependymomas (EPN) are the third most prevalent tumor of the central nervous system (CNS) in childhood, accounting for 6–12% of all pediatric brain tumors (Khatua, Ramaswamy, & Bouffet, 2017). According to the National Institutes of Health (NIH), EPN affects less than 200,000 people in the United States and is therefore classified as a rare tumor (Reni, Gatta, Mazza, & Vecht, 2007). These tumors can occur throughout neuroaxis, arise from alterations that occur during normal process of proliferation and differentiation of radial glial stem cells (Poppleton & Gilbertson, 2007; Taylor et al., 2005) (Supplementary Fig. 1).

Despite histopathological resemblances, EPN at different regions and locations in the brain, such as infratentorial (IT), supratentorial (ST), posterior fossa (PF), and the spinal cord (SC), have a disparate prognoses (Cage et al., 2013), most likely due to EPN vary in their clinicopathologic features, molecular characteristics, and lethality (Merchant et al., 2019). Such a discrepancy could be caused either by different cellular origins or by distinct genetic mutations of these tumors. Thus, although there are therapies available to treat EPN such as maximal cirrhotic resection, radiotherapy and chemotherapy (Lourdusamy et al., 2017), treatment for EPN remains challenging.

Recent studies have contributed to a better understanding of the genetic abnormalities in EPN. These investigations have shed some light on the molecular mechanisms underlying the development of EPN while proposing molecular markers of potential value for tumor stratification. For example, Comparative Genomic Hybridization (CGH) analysis has revealed genetic aberrations in EPN, such as gains in chromosome 1q that are associated with a poor prognosis (Rand et al., 2008). On the other hand, loss of the 6q25.3 region has been associated with improved prognosis (Monoranu et al., 2008).

Gene expression analysis has identified molecular signatures in EPN that comprise members of both the Notch and Hedgehog signaling pathways (de Bont, Packer, Michiels, Boer, & Pieters, 2008; Modena et al., 2006). Some of these studies have also shown that the over-expression of the enzymatic subunit of the human telomerase (hTERT) can be associated with a poor prognostic for this type of tumor (Mendrzyk et al., 2006; Modena et al., 2012; Shago et al., 2008). In addition, EPN have shown frequent gain and amplification of the epidermal growth factor receptor (EGFR) gene located at 7p11.2 as an independent prognostic factor for poor survival (Mendrzyk et al., 2006).

In 2011, research from our group has identified EPN-specific microRNA and epigenetic signatures that enabled stratification of EPN according to the region of occurrence (IT or ST), and to certain clinical behavior (Costa et al., 2011; Xie et al., 2009). Using a method for methylation analysis in Alu repetitive elements, we were able to classify EPN according to their behavior (Woodfine, Huddleston, & Murrell, 2011). We were also able to identify a microRNA cluster in an imprinted region of human chromosome 14q32 that is associated with a poor prognosis in EPN (Liu et al., 2010). However, there is an increasing need for studies aimed at identifying novel and better molecular biomarkers that will improve clinical decisions and therefore improve outcomes for EPN.

In this study, we sought to identify possible molecular biomarkers that could improve clinical management and better treatment for EPN, enabling a more favorable prognosis for the patients. In this direction, we performed the transcriptome analysis of 33 EPN (IT and ST) and 5 normal brain samples. We have been able to identify top the 5 genes differentially expressed in EPN, some of which, such as *IGF-2*, that are of potential clinical importance. Methylation analysis showed that *IGF-2* has loss methylation in the DMR2 region of its promoter in tumors in which it is highly expressed. This is the first report that describes the possible association of over-expression of *IGF-2* together with epigenetic changes in EPNs.

Material and methods

Patients and tumor samples

The samples used in all analysis were collected from the Falk Brain Tumor Tissue bank at the Ann & Robert H. Lurie Children's Hospital of Chicago (Chicago, IL) under an IRB-approved protocol. Initially, we have used one ''normal'' brain sample of a patient who died from non-brain related illnesses, which were microdissected and enriched for ependymal cells and 3 ependymomas to generate RNA-Seq data as our initial dataset. After that, we have used thirty-tree EPN samples (9 ST and 24 IT), that were surgically collected (before any treatment) from pediatric patients between 1997 and 2010. Five normal brain specimens (control) represent the core of the lateral ventricular lining of the basal ganglia and thalamus of patients who died from non-brain related illnesses, which were microdissected and enriched for ependymal cells. Clinical pathological information such as age, gender, tumor location, and WHO Grade for each EPN specimen was also collected by retrospective medical record review and is shown in Table 1.

RNA extraction

Approximately 0.5 cm^3 of frozen tissue of each samples was used for RNA extraction. Total RNA was extracted using a standard Trizol protocol (Invitrogen). Total RNA quantity and quality were evaluated using a spectrophotometer (Nanodrop ND-1000, Thermo Scientific) and agarose gel electrophoresis.

Library construction and sequence generation

Libraries were subjected to emulsion PCR and SOLiD sequencing following the manufacturer's instructions (Life Technologies/Applied Biosystems). An average of 20–30 million 35-base pair reads passed quality controls for each sample that was analyzed.

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

Table 1 List of thirty-three ependymoma samples and five normal controls used in this study.

Sample name	Sex	Tumor location	Specimen	WHO grade	Prior treat-ment	Patient status	Age at diagnosis (years)	Relapse	Specimen time to relapse (years)	Follow-up (years)
Ep01	F	IT	Primary	III	None	DOD	1.1	Yes	1.8	4.0
Ep03	M	IT	Primary	III	None	DOD	0.9	Yes	1.0	7.1
Ep04	F	IT	Primary	III	None	DOD	0.2	Yes	0.6	0.6
Ep05	M	IT	Primary	II	C/RT	NED	5.2	Yes	2.2	7.1+
Ep06	F	IT	Primary	II	None	DOD	7.6	Yes	1.2	2.4
Ep08	M	IT	Recurrent	III	C/RT	AWD	4.1	Yes	0.2	10.3+
Ep09	M	IT	Recurrent	II	C/RT	DOD	0.8	Yes	0.9	6.8
Ep11	M	IT	Primary	III	None	NED	5.2	No	NR	4.9+
Ep12	M	IT	Recurrent	III	RT	DOD	7.0	Yes	0.6	10.0
Ep16	M	IT	Recurrent	II	RT	DOD	12.8	Yes	0.8	3.5
Ep17	M	IT	Primary	II	C/RT	DOD	2.5	Yes	1.0	11.0
Ep19	F	ST	Primary	III	None	NED	16.4	Yes	0.4	8.8+
Ep20	F	ST	Recurrent	III	RT	NED	16.4	No	NR	8.8+
Ep21	F	IT	Recurrent	II	C/RT	NED	9.7	No	NR	13.1+
Ep22	M	IT	Recurrent	III	C	DOD	1.2	Yes	1.7	11.2
Ep24	F	IT	Recurrent	III	C/RT	DOD	9.2	Yes	1.1	6.8
Ep25	F	ST	Recurrent	II	C/RT	DOD	9.2	Yes	0.8	6.8
Ep26	F	IT	Recurrent	III	None	NED	11.1	No	NR	8.4+
Ep28	M	IT	Primary	III	None	DOD	0.8	Yes	1.0	6.8
Ep30	F	ST	Recurrent	III	C/RT	DOD	8.2	Yes	0.2	10.6
Ep31	M	IT	Primary	III	None	DOD	5.8	Yes	2.5	3.4
Ep32	F	ST	Recurrent	III	None	NED	16.4	Yes	1.5	8.8+
Ep33	F	IT	Primary	II	None	NED	1.4	No	NR	8.4+
Ep35	M	IT	Recurrent	II	C	DOD	1.2	Yes	1.2	11.2
Ep36	M	IT	Primary	III	None	DOD	1.3	Yes	1.2	1.4
Ep39	F	ST	Recurrent	III	None	NED	6.3	Yes	1.2	10.7+
Ep40	F	ST	Recurrent	III	RT	DOD	8.2	Yes	0.9	10.6
Ep42	F	IT	Primary	II	None	DOD	1.9	Yes	0.4	0.5
Ep44	F	IT	Primary	III	C/RT	DOD	9.2	Yes	0.5	6.8
Ep45	F	IT	Recurrent	III	C/RT	DOD	1.1	Yes	0.2	4.0
Ep46	F	ST	Primary	III	None	NED	11.1	No	NR	6.4+
Ep47	M	ST	Recurrent	III	None	NED	2.2	No	NR	13.4+
Ep48	M	IT	Primary	II	None	DOD	15.4	Yes	NR	2.5
N1	F	Normal Control	NA	NA	NA	NA	–	–	–	–
N2	F	Normal Control	NA	NA	NA	NA	–	–	–	–
N3	M	Normal Control	NA	NA	NA	NA	–	–	–	–
N4	M	Normal Control	NA	NA	NA	NA	–	–	–	–
N5	F	Normal Control	NA	NA	NA	NA	–	–	–	–

DOD = Died of Disease; AWD = Alive with Disease; NED = No Evidence of Disease; NA = Not Applicable.
NR = no relapse to date; C = Chemotherapy; RT = Radiotherapy.
For patients who are DOD, follow up represents time from diagnosis to death. For patients who are AWD or NED, follow up represents time from diagnosis to end of the study period April 1, 2010.

Data pre-processing and transcriptome mapping to human genome

Data pre-processing was carried out by removing low-quality reads using the Trimmomatic tool (Bolger, Lohse, & Usadel, 2014). Leading and trailing nucleotides with quality below eight were excluded, as well as 'N' bases. A 4-base wide sliding window was set to scan the reads and cut when the average quality per base dropped below 20. All pre-processed reads below 36 bases long were removed.

The high-quality reads generated were mapped to the March 2006 human reference sequence (NCBI Build

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

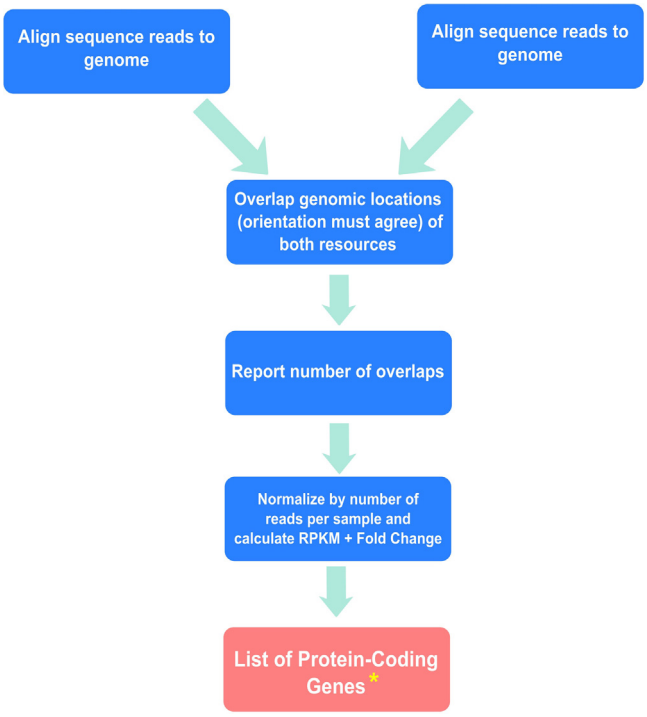

Fig. 1 The pipeline of big data analyses. In this figure, we demonstrate the analyzes made to evaluate protein-coding genes that were differentially expressed in EPN tumors when compared to the control (normal brain tissue).

36.1) with the Novoalign software version 2.0 (Novocraft Technologies, http://www.novocraft.com). The number of mapped reads from each run is shown in Supplementary Table 1. After obtaining these mapped locations, different custom made Perl scripts were used to overlap these intervals with other annotations, including human mapped positions of the Reference Sequence (RefSeq) dataset (O'Leary et al., 2016), genomic locations provided by UCSC (Kent, Sugnet, Furey, & Roskin, 2002), the miRBase version 18 dataset (Griffiths-Jones, 2004), and the fRNAdb dataset at ncrna.org (Kin et al., 2007). Another Perl script was then used to calculate the number of times that each annotated transcript was overlapped by a mapped read in each of the samples.

Identification of differentially expressed genes

Protein-coding gene databases in NCBI were searched for known genes and to calculate the Reads Per Kilobase per Million mapped reads (RPKM) for each clustered gene. Approximately 20,000 protein-coding genes of the transcriptome data were ascertained and annotated to the human genome using all available databases. Multiple gene lists were then generated corresponding to different RPKM-based comparisons between three EPN samples (tumors 1, 2, and 3) and one control tissue, ultimately enabling the identification of protein-coding genes that are differentially expressed between EPN and normal samples (Fig. 1) Supplementary Table 2). RPKM and FCs of expression between normal control tissue and tumor samples for each clustered gene were also calculated. Differential expression analysis of the data generated by RNA-Seq was conducted with the Bioconductor Package, specifically edgeR (Robinson, McCarthy, & Smyth, 2009). EdgeR normalizes the expression values based on the total number of reads and calculates the statistical significance of changes in expression between EPN samples and the normal control specimen. Validation calculations were done using the clValid package for the R statistical programming environment. Finally, the list of differentially expressed genes was filtered with the following criteria: p-value $\leq$0.01 (control vs. all tumor samples), at least a 10-fold change (control vs. all tumor samples), and a minimum expression of at least 500 sequence reads/gene in one tissue sample.

Gene ontology and KEGG enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed for the list of Differentially Expressed Genes (DEGs) in order to identify their prevalence in Biological Processes, Molecular Functions, Cellular Components and Pathways. This analysis was carried out using Enrichr (Kuleshov et al., 2016) functional annotation tool, with the Fisher Exact *P*-value set to <0.01. The smallest *P*-value indicates the highest degree of enrichment.

Protein-Coding Gene Networks

The list of differentially expressed protein-coding genes was also analyzed using the GeneGO (Metacore) pathway analysis software in order to obtain the biological roles of DEGs in EPN. Based on GeneSpring software (Agilent), GeneGO (Metacore), and published EPN data, a subset of 30 genes was selected for Real-Time TaqMan RT-PCR confirmation in a different cohort of 33 EPN samples.

Gene expression confirmation

Customized TaqMan Low Density Arrays (TLDA) (Applied Biosystems) representing a subset of 30 differentially expressed Protein-Coding genes (*ANXA1*, *C-FOS*, *CALM3*, *CAMK2A*, *COL6A2*, *DLG2*, *EGR1*, *MECOM*, *GABRA4*, *GLIS3*, *ID4*, *IGF-2*, *IGFBP5*, *ITGB4*, *LGR5*, *MMP14*, *MSX1*, *NES*, *NFATC4*, *NOTCH2*, *PAX6*, *ROR2*, *SLC1A2*, *TNC*, *TNFRSF11B*, *UNC13A*, *VIM*, *WNT5A*, *YWHAH* and *ZIC2*) and two endogenous control genes (*GAPDH* and 18S) were used to validate the differential expression of proteins observed in the initial transcriptome data in a new cohort of samples containing 33 EPNs compared to 5 normal samples.

RNA was purified as previously described and each RT reaction contained total RNA and reagents from the TaqMan Reverse Transcription Kit, according to the manufacturer's protocols. Reactions were run in a 384-well TLDA block with a 10 min initial denaturing step at 94.5 °C, followed by 40 cycles of 30 s at 97 °C and 1 min at 60 °C. The relative expression (delta C_T) for each of the genes was then determined using both the 18S RNA and the GAPDH mRNA as the controls.

Gene clustering

PCR results for these selected genes were scrutinized to see if samples could be appropriately categorized into the 3 distinct sample sets of normal, ST EPNs, and IT EPNs based

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

solely on the Delta C_T values of the 30 genes represented on the TLDA plate. To determine the best method of clustering, internal validation coefficients were calculated for several methods of unsupervised clustering including hierarchical, K-means, self-organizing maps (SOM), partitioning around medioids (PAM), model-based, self-organizing tree algorithm (SOTA), and divisive analysis (diana). The best method of clustering would minimize the connectivity, maximize the silhouette width, and maximize the Dunn index for all 3 clusters. The best method was determined to be hierarchical clustering based on Pearson correlation with average (for hierarchical clustering) linkage correlation. Results of the clustering dendrogram are shown as a heat map of expression values in Fig. 3.

DNA extraction and bisulfite conversion

DNA was extracted by the phenol-chloroform method. Approximately 1 μg of extracted DNA was treated with sodium bisulfite using the EZ DNA Lightning Methylation Kit (Zymo Research) following the manufacturer's instructions. The bisulfite converted DNA was then subjected to PCR to amplify the DMR2 region of *IGF-2*. The following primers (IDT) were used: Forward Primer (5′-GGT TAG GAG GAG GTT GTA GG-3′) and Reverse Biotinylated Primer (5′-CCA AAA CAA CTT CCC CAA ATA-3′). The following program was used to amplify the DNA: 94 °C for 10 min followed by 45 cycles of the following: a) 4 °C for 30 s; b) 54 °C for 30 s; c) 72 °C for 45 s and 72 °C for 5 min for the final extension. PCR products were visualized by agarose gel electrophoresis.

Pyrosequencing and quantitative analysis of *IGF-2* DNA methylation

We used the PyroMark Kit (Qiagen) for pyrosequencing reactions following the manufacturer's instructions. After PCR Amplification, triplicates of each sample were added to a 96 well plate with 30 μL of water and 40 μL of a PyroMark Binding Buffer/Streptavidin-Sepharose bead mix (GE Healthcare). Following this, DNA was prepared and pyrosequencing was carried out using the PyroMark Annealing for the DMR2 of the *IGF-2* gene: S primer (5′-GGG TGG GTA GAG TAA-3′). The PyroMark MD machine was then used to create a run using an assay to sequence by synthesis the DMR2 region of *IGF-2*. The relative methylation levels at 7 CpG sites in each replicate tested were analyzed using the PyroMark Q CpG Software (Biotage) in order to generate a mean methylation for each replicate. These replicate methylation percentages were then averaged for each sample to generate the final mean methylation percentage for each tumor and control sample as shown in the Results section.

For the methylation analysis and comparisons, figures were plotted using GraphPad Prism 6 Software (GraphPad Software, Inc) and underwent statistical analyses using the same software (Fig. 5). Significance for each figure was calculated using a standard unpaired, parametric, and two-tailed *t*-test with Welch's correction for unequal standard deviations between two groups compared (33 EPN samples and 5 normal brain specimens).

Results

Sequence generation and mapping to the human genome

Whole Transcriptome sequencing produced approximately 20–30 million reads per run that included one run from a primary EPN (Tumor 1), two runs of the recurrence of that EPN (Tumor 2), three runs of a second primary EPN (Tumor 3) and one run of a Normal Control brain sample enriched for ependymal cells (Control 1). Multiple gene lists were generated corresponding to different RPKM-based comparisons between tumor and control tissues, ultimately enabling the identification of protein-coding genes that are differentially expressed in EPNs (Supplementary Tables 3 to 5).

Approximately 10 million reads of good quality were then used for further analysis for each sample. A total of 70 million reads, representing a total of approximately 2.5 Gigabases of good quality reads with an average of 40% mapped to the human genome reference sequence, were retained for all samples. Then, a filtering process was carried in order to identify statistically significant differentially expressed genes. This analysis generated lists of 459 differentially expressed protein-coding genes.

GO enrichment analysis of protein-coding DEGs

After filtering statistically significant DEGs between normal and ependymal cancer cells identified, a list of 459 entries were obtained. This list was then submitted to Enrichr fGene Ontology (GO) Enrichment analysis, and significant GO terms including biological process, cellular component, and molecular function were collected (Fig. 2). For upregulated DEGs, pattern specification process was the most significant enrichment of biological process; sarcolemma was the highest enrichment for cellular component; and motor activity was the highest enrichment of molecular function, as shown in Supplementary Table 3. For downregulated DEGs, synaptic transmission was the most significant enrichment of biological process; synapse part was the highest enrichment of cellular component and SNARE binding was the highest enrichment of molecular function, as shown in Supplementary Table 4.

KEGG enrichment analysis of protein-coding DEGs

Following KEGG enrichment analysis of the 459 entries, significant terms were collected (Fig. 2). The pathways by 22 upregulated DEGs were related to ECM-receptor interaction and focal adhesion (Supplementary Table 5). A total of 214 downregulated DEGs were significantly enriched in GABAergic synapse, retrograde endocannabinoid signaling, nicotine addiction and synaptic vesicle cycle, as shown in Supplementary Table 5.

Protein-coding gene network analyses

We then evaluated the correlation of all protein-coding RNAs identified as statistically differentially expressed in different pathways and networks and found approximately 25

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

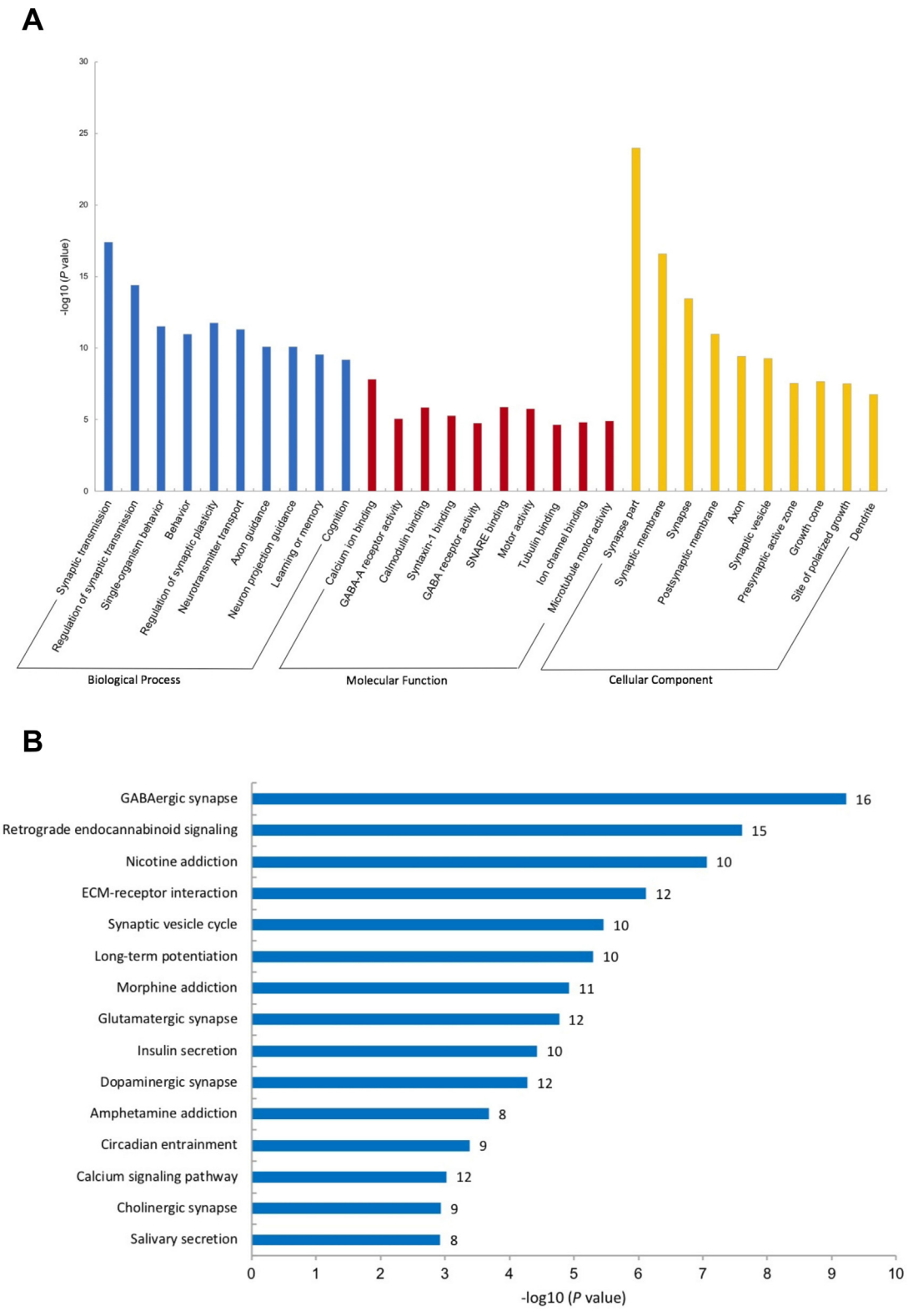

Fig. 2 GO and KEGG Enrichment Analysis of protein-coding DEGs in ependymomas. Using a list of 459 statistically significant Differentially Expressed Genes between normal and ependymal cancer cells identified, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment analysis were carried out. Fig. 2A represents GO enrichment information on Biological Process, Molecular Function and Cellular Component and Fig. 2B represents KEGG enrichment information.

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

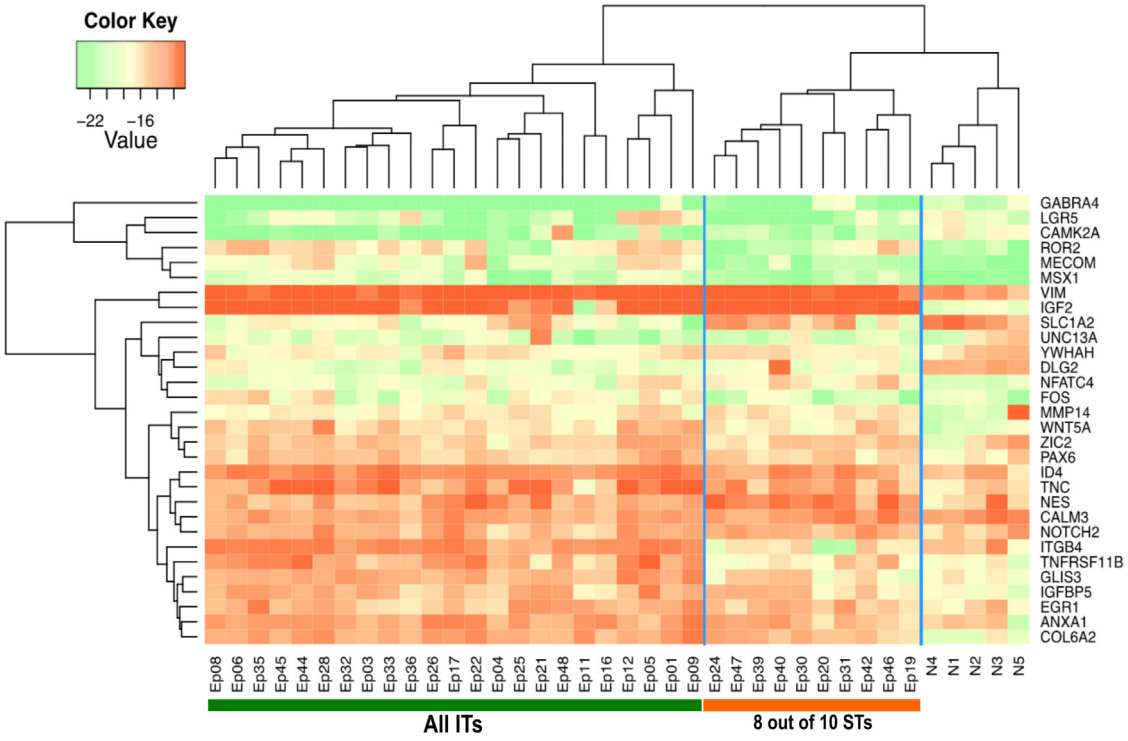

Fig. 3 Expression confirmation by Real-Time PCR. Heat map showing 17 genes that were differentially expressed. Amongst these genes, *IGF-2*, *COL6A2*, *WNT5A*, *Vimentin*, and *IGFBP5* were identified as the top differentially expressed in all comparisons between EPNs and controls.

protein-coding genes associated to neurogenesis and development (Supplementary Fig. 2). Five more genes were added based on an extensive literature analysis totaling 30 protein-coding genes (Supplementary Table 6). These 30 genes were then selected for confirmation by Customized TaqMan Real-Time PCR as described in the Material and Methods Section.

Expression confirmation by real-time PCR

We were able to confirm 17 genes that are differentially expressed in the new cohort. From these, five (*IGF-2*, *COL6A2*, *WNT5A*, *Vimentin* and *IGFBP5*) were identified as the top 5 differentially expressed in all comparisons between 33 EPNs and 5 control samples (Fig. 3). Hierarchical clustering of these 30 genes as well as the controls showed that the controls formed their own cluster, that is no other samples were in the same cluster. It also showed that the majority of the ST samples (8 of 10) were in the same cluster (Fig. 3). Therefore, these 30 genes of interest on the TLDA plate are not only effective at distinguishing between normal and disease states, but also aid in further diagnosis between disease states. A Venn Diagram comparing normal brain control samples with IT samples or just ST, showed a highly overexpression of *IGF-2*, *COL6A2*, *WNT5A*, *Vimentin* and *IGFBP5* genes in the EPN samples (Fig. 4). However, when IT samples were compared to ST samples, all five genes, except for *IGF-2*, were upregulated. *IGF-2*, the most over-expressed gene in our analysis of EPNs against normal controls, was further studied for epigenetic mechanisms of gene expression regulation.

DNA methylation analysis of the DMR2 of IGF-2 by pyrosequencing

Methylation results of *IGF-2* pyrosequencing are shown in Fig. 5. In each subset of the figure (A–F, from left to right) the overall percent methylation of 7 CpG sites in DMR2 in *IGF-2* of each sample in the subset is plotted along with a long horizontal line demarcating the arithmetic mean of percent DNA methylation in the subset with the whiskers denoting the 95% confidence interval. In Fig. 5A, the control samples were compared versus all the EPN samples, regardless of location or nature of the tumor. In this particular figure, although a slight decrease in overall methylation is seen across the board in the EPN samples, the difference was not statistically significant according to a *t*-test as indicated by the p-value considerably larger than the standard of 0.05. This was probably due to the large variation present in the tumor samples, and particular subsets of the tumor samples were then tested versus the controls to reveal some trends in the data. In Fig. 5B, the infratentorial tumors were tested against the controls. Although several tumors in this subset are below the arithmetic mean in the control group, the variance of methylation in this subset was not statistically significant. However, an appreciable difference between controls and ST tumors was found, as shown in Fig. 5C.

Further evaluation of the tumors under and above the line denoting the arithmetic mean revealed that the lone tumor with a percent methylation nearing 40% was a primary tumor, while the rest with methylation levels below 20% were all relapses. This trend of relapsed samples falling

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

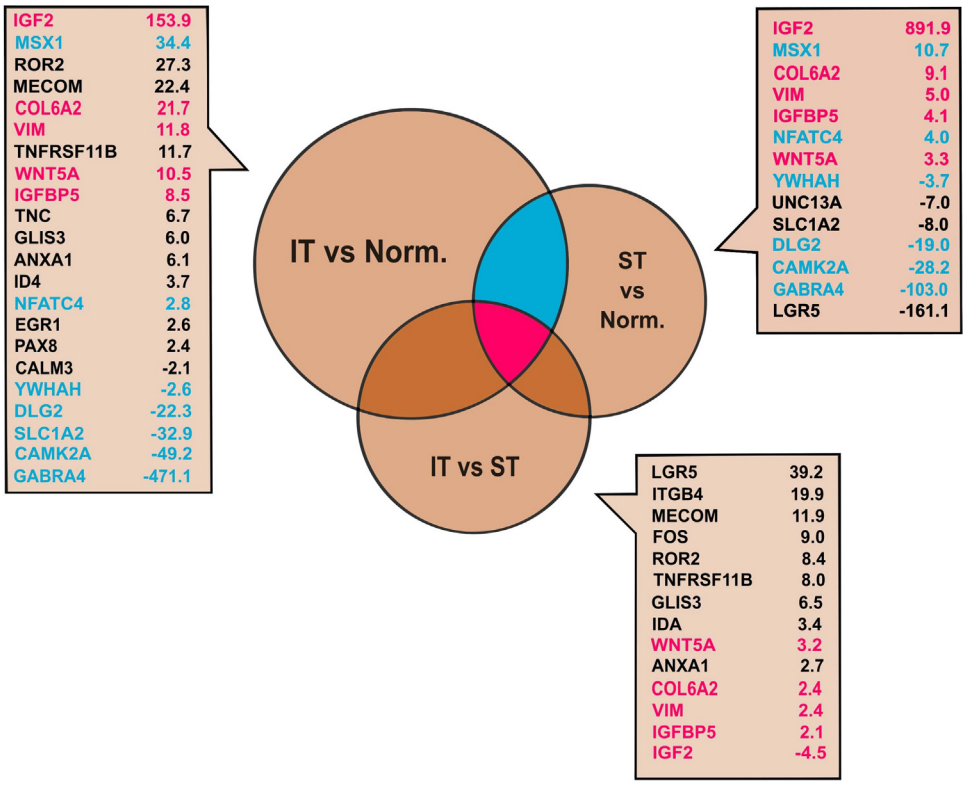

Fig. 4 Venn Diagram. When infratentorial and supratentorial EPN samples was compared with normal control samples, the genes *IGF-2*, *COL6A2*, *WNT5A*, *Vimentin* and *IGFBP5* were highly overexpressed in the tumors, although IGF-2 was downregulated when infratentorial samples were compared.

below the group arithmetic mean is also seen in IT tumors as well (Fig. 5B and D). Although the p-value was low ($p = 0.07$), the relapsed tumors did not show statistically significant hypomethylation of the DMR2 region of *IGF-2* versus the controls. Despite this, the quantitative degree of hypomethylation is indeed considerable, with the ST relapse tumor samples having extremely low methylation compared to the other relapses, ranging from 5.3 to 17.0%.

In order to verify if the trends in the DNA methylation data are inversely correlated to gene expression of *IGF-2* we compared the percent methylation of samples with the delta Ct values generated using RT-PCR for samples in which expression data was available. Overall, there was a good correlation between loss of DNA methylation and gene expression in the samples; however, in some cases we did not see that trend (Fig. 6).

Discussion

Our analyzes of differentially expressed transcripts in infant tumor samples provided possible diagnostic molecular markers. 20–30 million good quality reads for each sample, with an average of 30–40% mapping to the human reference genome, were obtained from RNA-Seq and WT sequencing. Analysis of protein-coding genes identified 459 genes differentially expressed in EPNs. Among them, 137 belong to a specific network of protein-coding genes that are affected in these neoplasms. From these, at least 25 have direct connections (Fig. 2). Two of the genes (*IGF-2* and *SLC1A2*) present in the same network were confirmed by preliminary Real-Time PCR reactions (Supplementary Fig. 3).

SLC1A2 was already implicated in tumor induced epilepsy and gliomas (Choi, Stradmann-Bellinghausen, Savaskan, & Regnier-Vigouroux, 2015; Niesen et al., 2013). *SLC1A2* is a glutamate transporter that allows uphill transport of glutamate into cells against a concentration gradient (Jiménez-Jiménez et al., 2014; Kanai et al., 2013; Zhou & Danbolt, 2013). The function of these transporters plays a crucial role in protecting neurons against glutamate excitotoxicity in the central nervous system (Kanai & Hediger, 2004; Kanai et al., 2013).

Our Real-Time PCR analysis indicated that 17 out of 30 genes were statistically significant in 33 EPNs samples (Fig. 3). Pair-wise calculation of p-values between 3 samples classification (normal, IT, and ST) identified 5 genes (*IGF-2*, *COL6A2*, *WNT5A*, *Vimentin* and *IGFBP5*) that were significant. These genes likely play a critical role in distinguishing sub-states in EPNs. It is already known that the location of the tumor in the brain is directly correlated to outcomes in EPNs and these genes may help quantitatively assess tumor origin in the near future (Lee, Chung, & Kim, 2016; Sayegh et al., 2014).

COL6A2 have been associated to different tumor types (Chen et al., 2014; Cheon et al., 2014; Lee et al., 2016) and directly to malignant brain tumors. Its over-expression is implicated together with other genes to vascular gene expression patterns in malignant brain tumors (Liu et al.,

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

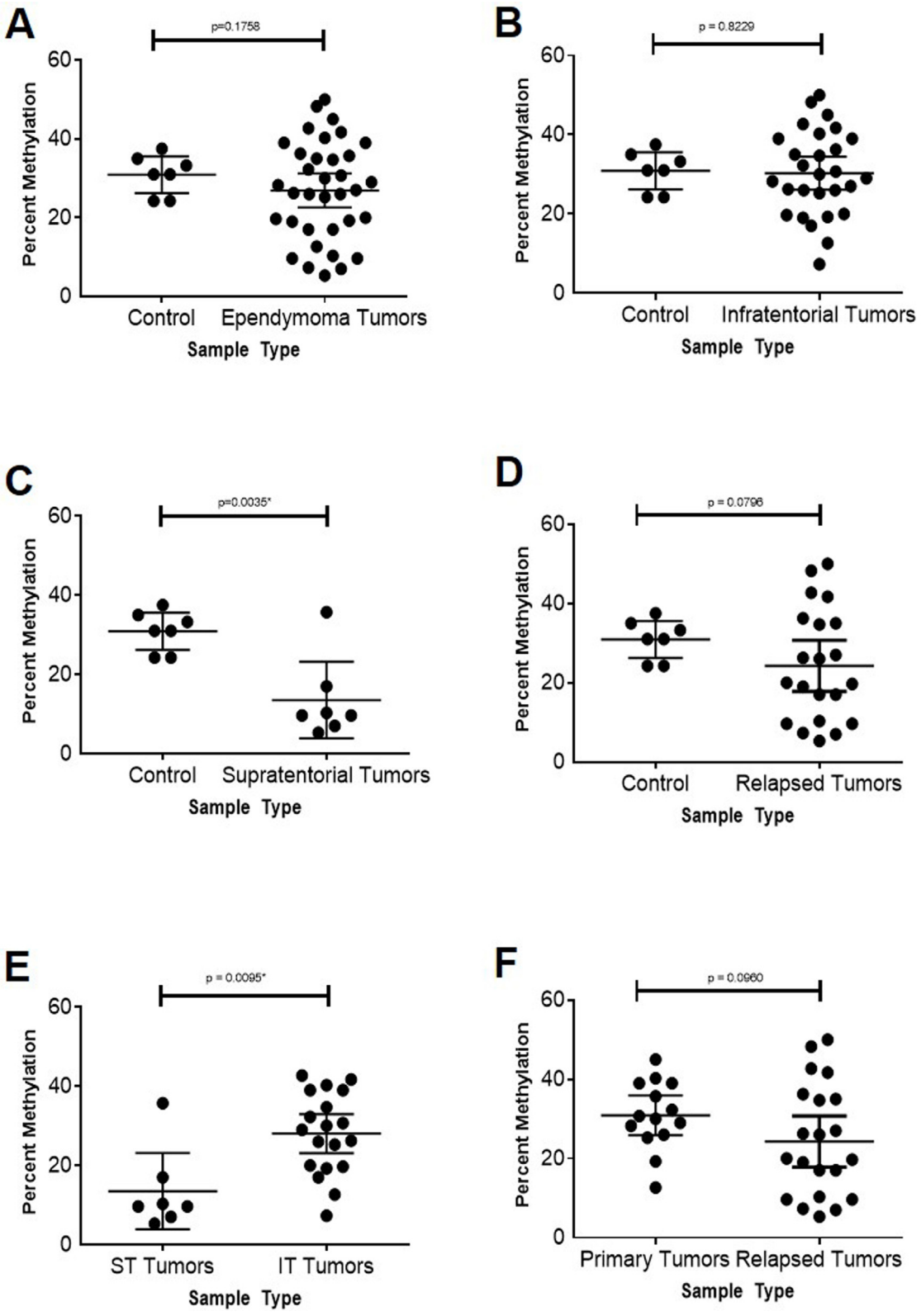

Fig. 5 DNA methylation analysis. Analysis of the overall percent methylation of 7 CpG sites in DMR2 in *IGF-2*. Several subsets to evaluated the differentially methylation pattern were performed. When compared with the control, supratentorial tumor samples showed a statistically significant in percentage of methylation (C).

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

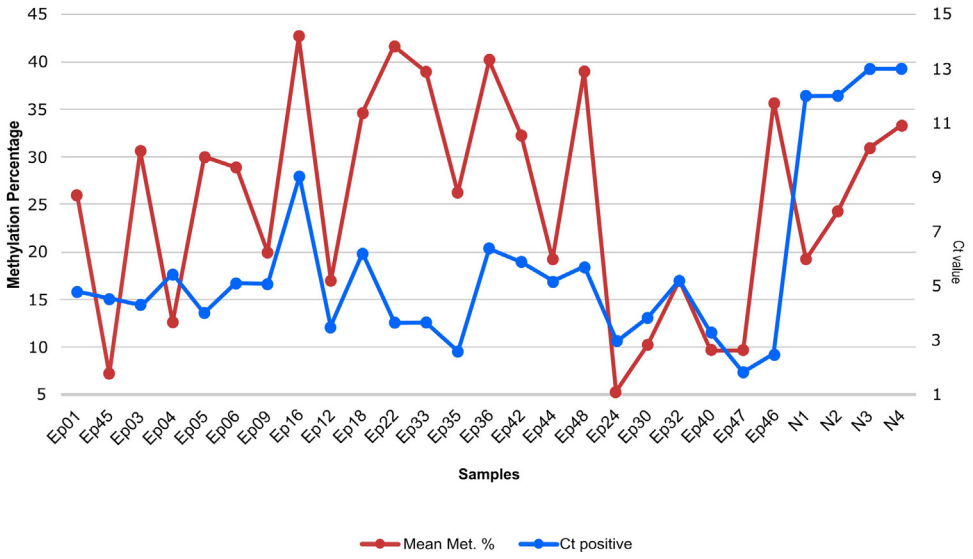

Fig. 6 Correlation between loss of DNA methylation and gene expression. Using RT-PCR for samples, comparison the percent methylation of samples with the delta Ct values were performed.

2010). *WNT5A* is a gene that was extensively associated to cancer (Endo, Nishita, Fujii, & Minami, 2015; Lin, Liu, Zhao, Sun, & Chen, 2014; Lu et al., 2015; Shojima et al., 2015). Vimentin was also implicated in brain tumors, mainly because this protein is associated to brain regeneration (Liu et al., 2014; Yung, Luna, & Borit, 1985). Recent studies show that Vimentin is directly implicated in the epithelial-to-mesenchymal transition process (EMT), and this process is usually defective in cancers (Niknami, Eslamifar, Emamirazavi, Ebrahimi, & Shirkoohi, 2017; Polioudaki et al., 2015; Satelli et al., 2017).

IGFBP5 protein was described as overexpressed in different tumor types (Güllü et al., 2015; Lee et al., 2014; Neuzillet et al., 2017), especially high-grade gliomas (Kološa, Motaln, Herold-Mende, Koršič, & Lah, 2015; Wang, Wang, Zhang, & Fuller, 2006), and it is involved in invasiveness in meningiomas (Sandberg Nordqvist & Mathiesen, 2002). IGFBP5 protein has an important role in the process of neurogenesis. Interestingly, *IGFBP5* and the other gene that was identified as highly expressed in EPNs in our study, *IGF-2*, are in the same network and both correlate to more aggressive and invasive tumors (Rosner, Pham, Moriggl, & Hengstschläger, 2017; Russell et al., 2017).

IGF-2 is an insulin growth factor associated to loss of imprinting in cancer (Damaschke et al., 2017; Küffer et al., 2018). *IGF-2* is already known as an oncogene in different types of cancer (Kessler, Haybaeck, & Kiemer, 2016; Martinez-Quetglas et al., 2016; Unger et al., 2017). This gene was identified as the most over-expressed gene in all our comparisons for EPNs for the first time. Recently, it has been shown that increased proliferation of chemoresistant glioma cells occurs by the coordinated activation of IGF-1R and AKT due to the presence of IGF-2 not sequestered by IGFBP6 (Oliva et al., 2018). In 2008, Corcoran and colleagues demonstrated that the *IGF-2* expression is not required for tumor initiation but it is important to proliferation and maintenance of medulloblastoma cells *in vitro* tests (Corcoran, Raveh, Barakat, Lee, & Scott, 2008).

Here, we were also able to identify subgroups of tumors based on location. Five genes that were identified, including *IGF-2*, are over-expressed in normals compared to IT as well as normal samples to STs. Interestingly, when comparing IT tumors to STs, *IGF-2* is down-regulated in STs. This might be explained by the difference in metabolism of tumors at different locations in the brain while *IGF-2* over-expression was reported in different brain tumors (Korshunov et al., 2003), however this is the first time that a group reports these results for EPN.

We selected the DMR2 of *IGF-2* for the DNA methylation analysis by pyrosequencing based on the study by Ye, Kohtz, Pollonini, Riccio, and Alberini (2015). Baseline levels of methylation of DMR2 in *IGF-2* in normal brain tissue were reported in studies to be between 30 and 50% (Guo et al., 2013; Murrell et al., 2008; Pidsley, Dempster, Troakes, Al-Sarraj, & Mill, 2012; Tabano et al., 2010). Our data for the controls tested had an arithmetic mean of ~30%, rendering the results found in the aforementioned paper validated and allowing tests for statistically significant hypomethylation of EPN samples versus these controls.

In our experiments, the DMR2 of the *IGF-2* gene was focused on to determine whether or not there was appreciable hypomethylation of the DMR2 in the *IGF-2* gene versus controls in EPN samples (Fig. 6). Despite of overall the tumors set did not show any statistically significant loss of methylation, subsets, especially ST, of tumors did show significant hypomethylation, with relapsed tumors also showing considerable hypomethylation relative to the controls. These results indicate that hypomethylation of DMR2 of *IGF-2* and supratentorial location of a tumor could be biomarker for tumors that are more likely to relapse.

DNA methylation plays an important role in the tumorigenesis process by regulating oncogenes activation, tumor-suppressor silencing, and chromosomal stability (He et al., 2008). It was already demonstrated that hypomethylation of *IGF-2* DMR0 is associated with a poor prognosis in colorectal tumor (Baba et al., 2010). In 2010, Miroglio

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

and colleagues have demonstrated that hypomethylation in lymphocytes DNA could be a specific biomarker for familial adenomatous polyposis patients (Miroglio et al., 2010).

In conclusion, using different approaches, we have been able to identify genes differentially expressed in EPNs, some of which, such as *IGF-2*, are of potential clinical importance. DNA methylation analysis showed that *IGF-2* has loss of methylation in the DMR2 region of its promoter in tumors in which it is highly expressed. This is the first report that describes the possible association of over-expression of *IGF-2* together with epigenetic changes in EPNs. More studies are warranted to further evaluate the impact of *IGF-2* over-expression in EPNs and how this pathway could be used as a biomarker and to develop new types of drugs.

Conflicts of interest

The authors declare that they do not have any conflicts of interest in this study.

Acknowledgements

We acknowledge Cofactor Genomics (www.cofactorgenomics.com) that helped generate the whole Transcriptome data in this study. We greatly appreciate the infrastructure provided by Dr Marcelo B. Soares' Laboratory at the Ann & Robert Lurie Children's Research Center in Chicago, United States. We also thank the insightful intellectual and technical help from Dr Maria de Fatima Bonaldo and Jared M. Bischof, respectively. We finally thank Dr Tadanori Tomita, Dr Stewart Goldman and Dr Chris Hamm for providing the samples and helping in the initial phase of this study. We are also indebt for the initial help with the raw data from Dr Fabio Passetti from FIOCRUZ in Rio de Janeiro, Brazil. Finally, we acknowledge Kevin G. Lindsay-Rivera for helping with the DNA methylation analysis. This article is dedicated to Dr Elio F. Vanin who has passed away in 2016, acknowledging him for his tremendous help in analyzing the data and for his invaluable intellectual insights in this study. We also thank the Financial Support given by the Maeve McNicholas Memorial Foundation that supported FFC for more than 7 years.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.biori.2019.08.002.

References

Baba, Y., Nosho, K., Shima, K., Huttenhower, C., Tanaka, N., Hazra, A., ... & Ogino, S. (2010). Hypomethylation of the IGF2 DMR in colorectal tumors, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastroenterology*, *139*(6), 1855–1864. http://dx.doi.org/10.1053/j.gastro.2010.07.050

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170

Cage, T. A., Clark, A. J., Aranda, D., Gupta, N., Sun, P. P., Parsa, A. T., ... & Auguste, K. I. (2013). A systematic review of treatment outcomes in pediatric patients with intracranial ependymomas. *Journal of Neurosurgery Pediatrics*, *11*(6), 673–681. http://dx.doi.org/10.3171/2013.2.PEDS12345

Chen, Y. C., Huang, R. L., Huang, Y. K., Liao, Y. P., Su, P. H., Wang, H. C., ... & Lai, H. C. (2014). Methylomics analysis identifies epigenetically silenced genes and implies an activation of β-catenin signaling in cervical cancer. *International Journal of Cancer*, *135*(1), 117–127. http://dx.doi.org/10.1002/ijc.28658

Cheon, D. J., Tong, Y., Sim, M. S., Dering, J., Berel, D., Cui, X., ... & Orsulic, S. (2014). A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clinical Cancer Research*, *20*(3), 711–723. http://dx.doi.org/10.1158/1078-0432.CCR-13-1256

Choi, J., Stradmann-Bellinghausen, B., Savaskan, N., & Regnier-Vigouroux, A. (2015). Glioblastoma cells induce differential glutamatergic gene expressions in human tumor-associated microglia/macrophages and monocyte-derived macrophages glutamatergic gene expressions. *Cancer Biology & Therapy*, *63*(August), E466. http://dx.doi.org/10.1080/15384047.2015.1056406

Corcoran, R. B., Raveh, T. B., Barakat, M. T., Lee, E. Y., & Scott, M. P. (2008). Insulin-like growth factor 2 is required for progression to advanced medulloblastoma in patched1 heterozygous mice. *Cancer Research*, *68*(21), 8788–8795. http://dx.doi.org/10.1158/0008-5472.CAN-08-2135

Costa, F. F., Bischof, J. M., Vanin, E. F., Lulla, R. R., Wang, M., Sredni, S. T., ... & Soares, M. B. (2011). Identification of microRNAs as potential prognostic markers in ependymoma. *PloS One*, *6*(10) http://dx.doi.org/10.1371/journal.pone.0025114

Damaschke, N. A., Yang, B., Bhusari, S., Avilla, M., Zhong, W., Blute, M. L., ... & Jarrard, D. F. (2017). Loss of Igf2 gene imprinting in murine prostate promotes widespread neoplastic growth. *Cancer Research*, *77*(19), 5236–5247. http://dx.doi.org/10.1158/0008-5472.CAN-16-3089

de Bont, J. M., Packer, R. J., Michiels, E. M., Boer, M. L. d., & Pieters, R. (2008). Biological background of pediatric medulloblastoma and ependymoma: A review from a translational research perspective. *Neuro-Oncology*, *10*(6), 1040–1060. http://dx.doi.org/10.1215/15228517-2008-059

Endo, M., Nishita, M., Fujii, M., & Minami, Y. (2015). *Insight into the role of Wnt5a-induced signaling in Normal and cancer cells. International review of cell and molecular biology* (Vol. 314) Elsevier Ltd. http://dx.doi.org/10.1016/bs.ircmb.2014.10.003

Griffiths-Jones, S. (2004). The microRNA registry. *Nucleic Acids Research*, *32*(90001), 109D–111D. http://dx.doi.org/10.1093/nar/gkh023

Güllü, G., Peker, I., Haholu, A., Eren, F., Küçükodaci, Z., Güleç, B., ... & Akkiprik, M. (2015). Clinical significance of miR-140-5p and miR-193b expression in patients with breast cancer and relationship to IGFBP5. *Sexual Development : Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, *29*, 21–29.

Guo, J., Wu, L., Wang, L., Shangguan, S., Chang, S., Wang, Z., ... & Zhang, T. (2013). Altered methylation of IGF2 DMR0 is associated with neural tube defects. *Molecular and Cellular Biochemistry*, *380*(1–2), 33–42. http://dx.doi.org/10.1007/s11010-013-1655-1

He, X., Chang, S., Zhang, J., Zhao, Q., Xiang, H., Kusonmano, K., ... & Wang, J. (2008). MethyCancer: The database of human DNA methylation and cancer. *Nucleic Acids Research*, *36*(SUPPL. 1), 836–841. http://dx.doi.org/10.1093/nar/gkm730

Jiménez-Jiménez, F. J., Alonso-Navarro, H., Martínez, C., Zurdo, M., Turpín-Fenoll, L., Millán-Pascual, J., ... & Agúndez, J. A. G. (2014). The solute carrier family 1 (glial high affinity glutamate transporter), member 2 gene, SLC1A2, rs3794087 variant and assessment risk for

restless legs syndrome. *Sleep Medicine*, *15*(2), 266–268. http://dx.doi.org/10.1016/j.sleep.2013.08.800

Kanai, Y., Clémençon, B., Simonin, A., Leuenberger, M., Lochner, M., Weisstanner, M., ... & Hediger, M. A. (2013). The SLC1 high-affinity glutamate and neutral amino acid transporter family. *Molecular Aspects of Medicine*, *34*(2–3), 108–120. http://dx.doi.org/10.1016/j.mam.2013.01.001

Kanai, Y., & Hediger, M. A. (2004). The glutamate/neutral amino acid transporter family SLC1: Molecular, physiological and pharmacological aspects. *Pflugers Archiv European Journal of Physiology*, *447*(5), 469–479. http://dx.doi.org/10.1007/s00424-003-1146-4

Kent, W. J., Sugnet, C. W., Furey, T. S., & Roskin, K. M. (2002). The human genome browser at UCSC. *Journal of Medicinal Chemistry*, *19*(10), 1228–1231. http://dx.doi.org/10.1101/gr.229102

Kessler, S., Haybaeck, J., & Kiemer, A. (2016). Insulin-Like growth factor 2 - the Oncogene and its accomplices. *Current Pharmaceutical Design*, *22*(39), 5948–5961. http://dx.doi.org/10.2174/1381612822666160713100235

Khatua, S., Ramaswamy, V., & Bouffet, E. (2017). Current therapy and the evolving molecular landscape of paediatric ependymoma. *European Journal of Cancer*, *70*, 34–41. http://dx.doi.org/10.1016/j.ejca.2016.10.013

Kin, T., Yamada, K., Terai, G., Okida, H., Yoshinari, Y., Ono, Y., ... & Asai, K. (2007). fRNAdb: A platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Research*, *35*(SUPPL. 1), 145–148. http://dx.doi.org/10.1093/nar/gkl837

Kološa, K., Motaln, H., Herold-Mende, C., Koršič, M., & Lah, T. T. (2015). Paracrine effects of mesenchymal stem cells induce senescence and differentiation of glioblastoma stem-like cells. *Cell Transplantation*, *24*(4), 631–644. http://dx.doi.org/10.3727/096368915X687787

Korshunov, A., Neben, K., Wrobel, G., Tews, B., Benner, A., Hahn, M., ... & Lichter, P. (2003). Gene expression patterns in Ependymomas Correlate with tumor location, grade, and patient age. *The American Journal of Pathology*, *163*(5), 1721–1727. http://dx.doi.org/10.1016/S0002-9440(10)63530-4

Küffer, S., Gutting, T., Belharazem, D., Sauer, C., Michel, M. S., Marx, A., ... & Ströbel, P. (2018). Insulin-like growth factor 2 expression in prostate cancer is regulated by promoter-specific methylation. *Molecular Oncology*, *12*(2), 256–266. http://dx.doi.org/10.1002/1878-0261.12164

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. http://dx.doi.org/10.1093/nar/gkw377

Lee, C. H., Chung, C. K., & Kim, C. H. (2016). Genetic differences on intracranial versus spinal cord ependymal tumors: A meta-analysis of genetic researches. *European Spine Journal*, *25*(12), 3942–3951. http://dx.doi.org/10.1007/s00586-016-4745-4

Lee, S., Yoon, D. S., Paik, S., Lee, K.-M., Jang, Y., & Lee, J. W. (2014). microRNA-495 Inhibits Chondrogenic Differentiation in Human Mesenchymal Stem Cells by Targeting*Sox9*. *Stem Cells and Development*, *23*(15), 1798–1808. http://dx.doi.org/10.1089/scd.2013.0609

Lin, L., Liu, Y., Zhao, W., Sun, B., & Chen, Q. (2014). Wnt5A expression is associated with the tumor metastasis and clinical survival in cervical cancer. *International Journal of Clinical and Experimental Pathology*, *7*(9), 6072–6078.

Liu, Y., Carson-Walter, E. B., Cooper, A., Winans, B. N., Johnson, M. D., & Walter, K. A. (2010). Vascular gene expression patterns are conserved in primary and metastatic brain tumors. *Journal of Neuro-oncology*, *99*(1), 13–24. http://dx.doi.org/10.1007/s11060-009-0105-0

Liu, Z., Li, Y., Cui, Y., Roberts, C., Lu, M., Wilhelmsson, U., ... & Chopp, M. (2014). Beneficial effects of gfap/vimentin reactive astrocytes for axonal remodeling and motor behavioral recovery in mice after stroke. *Glia*, *62*(12), 2022–2033. http://dx.doi.org/10.1002/glia.22723

Lourdusamy, A., Luo, L. Z., Storer, L. C., Cohen, K. J., Resar, L., & Grundy, R. G. (2017). Transcriptomic analysis in pediatric spinal ependymoma reveals distinct molecular signatures. *Oncotarget*, *8*(70), 115570–115581. http://dx.doi.org/10.18632/oncotarget.23311

Lu, C., Wang, X., Zhu, H., Feng, J., Ni, S., & Huang, J. (2015). Over-expression of ROR2 and Wnt5a cooperatively correlates with unfavorable prognosis in patients with non-small cell lung cancer. *Oncotarget*, *6*(28), 24912–24921. http://dx.doi.org/10.18632/oncotarget.4701

Martinez-Quetglas, I., Pinyol, R., Dauch, D., Torrecilla, S., Tovar, V., Moeini, A., ... & Llovet, J. M. (2016). IGF2 is up-regulated by epigenetic mechanisms in hepatocellular carcinomas and is an actionable oncogene product in experimental models. *Gastroenterology*, *151*(6), 1192–1205. http://dx.doi.org/10.1053/j.gastro.2016.09.001

Mendrzyk, F., Korshunov, A., Benner, A., Toedt, G., Pfister, S., Radlwimmer, B., ... & Lichter, P. (2006). Identification of gains on 1q and epidermal growth factor receptor overexpression as independent prognostic markers in intracranial ependymoma. *Clinical Cancer Research*, *12*(7 I), 2070–2079. http://dx.doi.org/10.1158/1078-0432.CCR-05-2363

Merchant, T. E., Bendel, A. E., Sabin, N. D., Burger, P. C., Shaw, D. W., Chang, E., ... & Ramaswamy, V. (2019). Conformal radiation therapy for pediatric ependymoma, chemotherapy for incompletely resected ependymoma, and observation for completely resected, supratentorial ependymoma. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *37*(12), 974–983. http://dx.doi.org/10.1200/JCO.18.01765

Miroglio, A., Jammes, H., Tost, J., Ponger, L., Gut, I. G., Abdalaoui, H. E., ... & Dandolo, L. (2010). Specific hypomethylated CpGs at the IGF2 locus act as an epigenetic biomarker for familial adenomatous polyposis colorectal cancer. *Epigenomics*, *2*(3), 365–375. http://dx.doi.org/10.2217/epi.10.24

Modena, P., Buttarelli, F. R., Miceli, R., Piccinin, E., Baldi, C., Antonelli, M., ... & Massimino, M. (2012). Predictors of outcome in an AIEOP series of childhood ependymomas: A multifactorial analysis. *Neuro-Oncology*, *14*(11), 1346–1356. http://dx.doi.org/10.1093/neuonc/nos245

Modena, P., Lualdi, E., Facchinetti, F., Veltman, J., Reid, J. F., Minardi, S., ... & Sozzi, G. (2006). Identification of tumor-specific molecular signatures in intracranial ependymoma and association with clinical characteristics. *Journal of Clinical Oncology*, *24*(33), 5223–5233. http://dx.doi.org/10.1200/JCO.2006.06.3701

Monoranu, C. M., Huang, B., Zangen, I. L., Rutkowski, S., Vince, G. H., Gerber, N. U., ... & Roggendorf, W. (2008). Correlation between 6q25.3 deletion status and survival in pediatric intracranial ependymomas. *Cancer Genetics and Cytogenetics*, *182*(1), 18–26. http://dx.doi.org/10.1016/j.cancergencyto.2007.12.008

Murrell, A., Ito, Y., Verde, G., Huddleston, J., Woodfine, K., Silengo, M. C., ... & Riccio, A. (2008). Distinct methylation changes at the IGF2-H19 locus in congenital growth disorders and cancer. *PloS One*, *3*(3), 1–7. http://dx.doi.org/10.1371/journal.pone.0001849

Neuzillet, Y., Chapeaublanc, E., Krucker, C., De Koning, L., Lebret, T., Radvanyi, F., ... & Bernard-Pierrot, I. (2017). IGF1R activation and the in vitro antiproliferative efficacy of IGF1R inhibitor are inversely correlated with IGFBP5 expression in bladder cancer. *BMC Cancer*, *17*(1), 1–12. http://dx.doi.org/10.1186/s12885-017-3618-5

Niesen, C. E., Xu, J., Fan, X., Li, X., Wheeler, C. J., Mamelak, A. N., ... & Wang, C. (2013). Transcriptomic Profiling of

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002

Human Peritumoral Neocortex Tissues Revealed Genes Possibly Involved in Tumor-Induced Epilepsy. *PloS One*, *8*(2), 1–9. http://dx.doi.org/10.1371/journal.pone.0056077

Niknami, Z., Eslamifar, A., Emamirazavi, A., Ebrahimi, A., & Shirkoohi, R. (2017). The association of vimentin and fibronectin gene expression with epithelial-mesenchymal transition and tumor malignancy in colorectal carcinoma. *EXCLI Journal*, *16*, 1009–1017. http://dx.doi.org/10.17179/excli2017-481

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., ... & Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. http://dx.doi.org/10.1093/nar/gkv1189

Oliva, C. R., Halloran, B., Hjelmeland, A. B., Vazquez, A., Bailey, S. M., Sarkaria, J. N., ... & Griguer, C. E. (2018). IGFBP6 controls the expansion of chemoresistant glioblastoma through paracrine IGF2/IGF-1R signaling. *Cell Communication and Signaling*, *16*(1), 1–14. http://dx.doi.org/10.1186/s12964-018-0273-7

Pidsley, R., Dempster, E., Troakes, C., Al-Sarraj, S., & Mill, J. (2012). *Imprinting control region on 11P15. 5 is associated with cerebellum weight, (February)*. pp. 155–163.

Polioudaki, H., Agelaki, S., Chiotaki, R., Politaki, E., Mavroudis, D., Matikas, A., ... & Theodoropoulos, P. A. (2015). Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. *BMC Cancer*, *15*, 399. http://dx.doi.org/10.1186/s12885-015-1386-7

Poppleton, H., & Gilbertson, R. J. (2007). Stem cells of ependymoma. *British Journal of Cancer*, *96*(1), 6–10. http://dx.doi.org/10.1038/sj.bjc.6603519

Rand, V., Prebble, E., Ridley, L., Howard, M., Wei, W., Brundler, M. A., ... & Grundy, R. G. (2008). Investigation of chromosome 1q reveals differential expression of members of the S100 family in clinical subgroups of intracranial paediatric ependymoma. *British Journal of Cancer*, *99*(7), 1136–1143. http://dx.doi.org/10.1038/sj.bjc.6604651

Reni, M., Gatta, G., Mazza, E., & Vecht, C. (2007). Ependymoma. *Critical Reviews in Oncology/hematology*, *63*(1), 81–89. http://dx.doi.org/10.1016/j.critrevonc.2007.03.004

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://dx.doi.org/10.1093/bioinformatics/btp616

Rosner, M., Pham, H. T. T., Moriggl, R., & Hengstschläger, M. (2017). Human stem cells alter the invasive properties of somatic cells via paracrine activation of mTORC1. *Nature Communications*, *8*(1) http://dx.doi.org/10.1038/s41467-017-00661-x

Russell, M. R., Graham, C., D'Amato, A., Gentry-Maharaj, A., Ryan, A., Kalsi, J. K., ... & Graham, R. L. J. (2017). A combined biomarker panel shows improved sensitivity for the early detection of ovarian cancer allowing the identification of the most aggressive type II tumours. *British Journal of Cancer*, *117*(5), 666–674. http://dx.doi.org/10.1038/bjc.2017.199

Sandberg Nordqvist, A. C., & Mathiesen, T. (2002). Expression of IGF-II, IGFBP-2, -5, and -6 in meningiomas with different brain invasiveness. *Journal of Neuro-oncology*, *57*(1), 19–26. http://dx.doi.org/10.1023/A:1015765613544

Satelli, A., Batth, I., Brownlee, Z., Mitra, A., Zhou, S., Noh, H., ... & Li, S. (2017). EMT circulating tumor cells detected by cell-surface vimentin are associated with prostate cancer progression. *Oncotarget*, *8*(30), 49329–49337. http://dx.doi.org/10.18632/oncotarget.17632

Sayegh, E. T., Aranda, D., Kim, J. M., Oh, T., Parsa, A. T., & Oh, M. C. (2014). Prognosis by tumor location in adults with intracranial ependymomas. *Journal of Clinical Neuroscience*, *21*(12), 2096–2101. http://dx.doi.org/10.1016/j.jocn.2014.05.011

Shago, M., Bouffet, E., Wong, V., Rutka, J., Alon, N., Hawkins, C., ... & Malkin, D. (2008). Telomere maintenance and dysfunction predict recurrence in paediatric ependymoma. *British Journal of Cancer*, *99*(7), 1129–1135. http://dx.doi.org/10.1038/sj.bjc.6604652

Shojima, K., Sato, A., Hanaki, H., Tsujimoto, I., Nakamura, M., Hattori, K., ... & Kikuchi, A. (2015). Wnt5a promotes cancer cell invasion and proliferation by receptor-mediated endocytosis-dependent and -independent mechanisms, respectively. *Scientific Reports*, *5*, 8042. http://dx.doi.org/10.1038/srep08042

Tabano, S., Colapietro, P., Cetin, I., Grati, F. R., Zanutto, S., Mandò, C., ... & Miozzo, M. (2010). Epigenetic modulation of the IGF2/H19 imprinted domain in human embryonic and extra-embryonic compartments and its possible role in fetal growth restriction. *Epigenetics*, *5*(4), 313–324. http://dx.doi.org/10.4161/epi.5.4.11637

Taylor, M. D., Poppleton, H., Fuller, C., Su, X., Liu, Y., Jensen, P., ... & Gilbertson, R. J. (2005). Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell*, *8*(4), 323–335. http://dx.doi.org/10.1016/j.ccr.2005.09.001

Unger, C., Kramer, N., Unterleuthner, D., Scherzer, M., Burian, A., Rudisch, A., ... & Dolznig, H. (2017). Stromal-derived IGF2 promotes colon cancer progression via paracrine and autocrine mechanisms. *Oncogene*, (March), 1–15. http://dx.doi.org/10.1038/onc.2017.116

Wang, H., Wang, H., Zhang, W., & Fuller, G. N. (2006). Overexpression of IGFBP5, but not IGFBP3, correlates with the histologic grade of human diffuse glioma: A tissue microarray and immunohistochemical study. *Technology in Cancer Research & Treatment*, *5*(3), 195–199. http://dx.doi.org/10.1177/153303460600500303

Woodfine, K., Huddleston, J. E., & Murrell, A. (2011). Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics & Chromatin*, *4*(1), 1. http://dx.doi.org/10.1186/1756-8935-4-1

Xie, H., Wang, M., Bonaldo, M., de, F., Smith, C., Rajaram, V., Goldman, S., ... & Soares, M. B. (2009). Epigemic analysis of Alu repeats in human ependymomas. *Nucleic Acids Research*, *37*(13), 4331–4340. http://dx.doi.org/10.1093/nar/gkp393

Ye, X., Kohtz, A., Pollonini, G., Riccio, A., & Alberini, C. M. (2015). Insulin like growth factor 2 expression in the rat brain both in basal condition and following learning predominantly derives from the maternal allele. *PloS One*, *10*(10), 1–15. http://dx.doi.org/10.1371/journal.pone.0141078

Yung, W. K., Luna, M., & Borit, A. (1985). Vimentin and glial fibrillary acidic protein in human brain tumors. *Journal of Neuro-oncology*, *3*(0167-594X (Print)), 35–38.

Zhou, Y., & Danbolt, N. C. (2013). GABA and glutamate transporters in brain. *Frontiers in Endocrinology*, *4*(November), 1–14. http://dx.doi.org/10.3389/fendo.2013.00165

Please cite this article in press as: Leite, M. L., et al. Transcriptome analysis and gene networks in a rare pediatric tumor. *Biotechnology Research and Innovation* (2019), https://doi.org/10.1016/j.biori.2019.08.002