scientific reports

Check for updates

OPEN

Medulloblastoma's master regulators and their association with patients' risk

Gustavo Lovatto Michaelsen[1,2,3,9], Tayrone de Sousa Monteiro[1,9], Danilo Oliveira Imparato[1], João Vitor Almeida da Costa[1], Daniel Rocha Silva[2], Iara Dantas de Souza[1], Marcel Câmara da Ribeiro-Dantas[4], Otávio Cabral-Marques[5,6], Marialva Sinigaglia[2,3] & Rodrigo Juliani Siqueira Dalmolin[1,7,8]✉

Medulloblastoma (MB) is the most common malignant pediatric brain tumor, accounting for approximately 20% of all childhood brain tumors. Despite recent advances, current treatments like surgery, radiation, and chemotherapy still lead to severe side effects and high morbidity. Limited knowledge exists regarding the regulatory mechanisms behind the MB transcriptional alterations in high-aggressive subgroups like Group 3 and Group 4, hindering the development of targeted therapies. Identifying key transcriptional regulators, known as master regulators (MRs), can elucidate the dysregulated pathways underlying MB progression and uncover potential treatment targets. In this study, we utilize primary MB gene expression samples to infer its regulatory network. Subsequently, we applied the Master Regulator Analysis identifying the transcription factors *BHLHE41*, *RFX4*, and *NPAS3* as its main transcriptional regulators, showing tumor suppressor features. We also identified eight risk MRs highly associated with patient outcome: four regulators (*MYC*, *REL*, *ZSCAN5 A*, and *ZFAT*) with activities associated with poor prognosis, and four (*PAX6*, *ARNT2*, *ZNF157*, and *HIVEP3*) acting antagonistically, being associated with good outcome. Our results offer key insights into the molecular mechanisms driving these tumors and identify novel potential therapeutic targets, addressing the urgent need for more effective and less toxic treatments.

Keywords Medulloblastoma, Transcription factor, Regulome, Systems biology

Medulloblastoma (MB) is an embryonal neuroepithelial cancer of the cerebellum, ranked as the most frequent malignant pediatric brain tumor[1]. It is characterized by four distinct molecular subgroups: wingless/int1 (WNT), Sonic Hedgehog (SHH), and the numerically designated Group 3 (G3) and Group 4 (G4), each with unique molecular and clinical features[2–4]. Approximately 90% of WNT tumors present a mutation in the *CTNNB1*gene, which activates the WNT signaling pathway and prompts WNT-responsive genes that drive tumor growth[5]. Patients of this subgroup have the best prognosis of all four groups, with survival rates around 95–100%[6]. In SHH, tumorigenesis typically results from copy number variations or mutations in the SHH signaling pathway, most commonly in the *PTCH1* and *SUFU*genes[7]. Despite recent advances, significant knowledge gaps remain in understanding the dysregulated pathways and master regulators (MRs) driving the highly aggressive G3 and G4 subgroups, which share similarities in both molecular and clinical domains[8].

Over the past decade, molecular targeted therapies have been tested for SHH, G3, and G4 subgroups, aiming to reduce side effects without compromising treatment efficacy and optimize overall survival[9]. SHH pathway inhibitors have shown promising results in clinical trials for SHH-driven MBs. However, one major obstacle is the development of drug resistance to these pathway inhibitors and novel side effects, such as the impairment of normal bone development in SHH patients. Specific tumor drivers in G3 and G4 MB have not been well

[1]Bioinformatics Multidisciplinary Environment-BioME, Digital Metropole Institute, Federal University of Rio Grande do Norte, Natal 59076-550, RN, Brazil. [2]Children's Cancer Institute, Porto Alegre 90620-110, RS, Brazil. [3]National Science and Technology Institute for Children's Cancer Biology and Pediatric Oncology – INCT BioOncoPed, Porto Alegre 90035-003, RS, Brazil. [4]Graduate Program of Biotechnology, Universidade Potiguar, Natal 59056-000, RN, Brazil. [5]Department of Immunology, Institute of Biomedical Sciences, University of São Paulo Brazil, São Paulo 05508-000, Brazil. [6]DO'R Institute for Research, São Paulo 01401-002, Brazil. [7]Department of Biochemistry, Federal University of Rio Grande do Norte, Natal 59064-741, RN, Brazil. [8]UFRN Central University, Campus, R. do Horto, Lagoa Nova, Natal 59078-900, RN, Brazil. [9]Gustavo Lovatto Michaelsen and Tayrone de Sousa Monteiro contributed equally. ✉email: rodrigo.dalmolin@imd.ufrn.br

identified, and fewer specific approaches for these subgroups are available. Despite the promising results, significant limitations persist, with various clinical trials not achieving the expected outcomes[10–12].

Alterations in transcription factors (TFs) activity are a common regulatory mechanism in various types of cancer[13]. Dysregulated TFs, through mechanisms such as altered expression, chromosomal translocations, gene amplification or deletion, point mutations, and indirectly through non-coding DNA mutations, represent a unique class of drug targets that mediate aberrant gene expression in biological processes key for cancer development[14]. A well-known example of this kind of dysregulation is the amplification of *MYC*in MB, particularly in the G3 subgroup, which promotes cell proliferation and increases tumor aggressiveness[15]. Several approaches to target TF activity in cancer have been demonstrated, both clinically and preclinically. Recent innovations include modulation of auto-inhibition, proteolysis targeting chimeras, and combinations of TF inhibitors with kinase inhibitors to block treatment resistance[16–19]. These advances in drug development hold great promise for yielding TF-targeted therapies that could significantly impact future cancer treatments. These innovations underscore the importance of continuing to improve MB molecular characterization, with the aim of discovering new treatment targets and refine existing molecular targeted therapies.

Identifying key transcriptional regulators, known as MRs, can elucidate the dysregulated pathways underlying MB progression and uncover potential treatment targets. This approach has been previously employed in various cancers, including Ewing sarcoma, neuroblastoma, and breast cancer, among others[20–25]. In this study, we inferred the MB's regulatory network and applied the Master Regulator Analysis (MRA) integrated with patient survival data. Our goal was to gain insights into the molecular mechanisms driving these tumors and identify novel TFs linked to the disease, addressing the urgent need for more effective and less toxic treatments. By uncovering the MRs associated with MB progression, this work offers a valuable resource for developing novel therapeutic strategies tailored to each subgroup's distinct biology.

Methods

Dataset selection and processing

The gene expression data used in this study were obtained from the Gene Expression Omnibus (GEO)[26] under the accession identifier GSE85217 and clinical data retrieved from its corresponding publication[27]. This dataset includes tumor expression profiles from 763 MB patients, spanning various age groups and representing all four molecular subgroups. As controls, we used a second dataset containing expression data from eight healthy fetal cerebellum samples and five healthy adult cerebellum samples. This control dataset was also retrieved from GEO, under accession identifier GSE167447. Both datasets were generated using the Affymetrix Human Gene 1.1 ST Array platform. All expression datasets were preprocessed and normalized by the oligo R/Bioconductor package[28]. For genes detected by multiple probe sets, the probe set with the highest standard deviation was selected for further analyses. Quality control was performed on both datasets using the *arrayQualityMetrics* R/Bioconductor package, and outlier samples were removed (Table S1)[29]. The remaining GSE85217 samples were subsequently divided into two sets: one for network inference, comprising 732 samples, and another for signature construction, containing 24 samples from the three most aggressive MB subgroups (SHH, G3, and G4)[30]. For the signature subset, eight from each subgroup were randomly selected, with selection adjusted to reflect the molecular subtype prevalence in the dataset (Table S1)[31].

Regulatory network inference and master regulator analysis

MRA and regulatory network inference were conducted using *RTN*(Reconstruction of Transcriptional Networks), an R/Bioconductor package[24]. This package offers various techniques for the inference and exploration of transcriptional regulatory networks, such as the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) and the MRA algorithms[32,33]. First, the Transcriptional Network Inference was applied to calculate the interdependence between each TF and all its potential regulated genes, using a list of human TFs comprising 1,639 genes and the 732 samples from GSE85217[34]. TF gene symbols were converted to official gene symbols using the *alias2Symbol* function from the R/Bioconductor package *limma*[35]. This conversion was also applied to all genes present in the Affymetrix Human Gene 1.1 ST Array platform, resulting in 1,635 TFs and 19,615 potential gene targets. Following the *RTN's* package vignette, the network construction pass through a permutation step, a bootstrap analysis, which creates a consensus bootstrap network, and finally the ARACNe algorithm aiming to remove the weakest interaction in any triplet formed by two TFs and a common target gene, preserving the dominant TF-target pair. This process ultimately led to the inference of the MB transcriptional network.

Parallel to the network inference, gene signatures for the three MB subgroups were defined. These signatures were obtained by performing three independent differential gene expression analyses comparing the three MB subgroup (GSE85217: $N = 8$ SHH, $N = 8$ G3, $N = 8$ G4) against the control fetal cerebellum samples (GSE167447: $N = 8$). The differential expression analysis was performed using the *limma* R/Bioconductor package (BH adjusted p-value < 0.05, $-1.5 >$ logFC > 1.5)[36]. Signature for the WNT subgroup was not made. Another three differential gene expression analyses using all samples for each subgroup ($N = 222$ SHH, $N = 142$ G3, $N = 323$ G4) was also made to ensure that the eight samples' signatures comprehend the heterogeneity of the dataset. Subsequently, the MRA searched for regulons in the regulatory network that were enriched for each of the corresponding subgroup-specific gene signatures (BH adjusted p-value < 0.01), aiming to identify important TFs for the disease, known as MR. Only regulons with more than 15 targets, 1,156 of the 1,581 present in the network, were considered in this analysis.

Selecting master regulators by significance

A subset of MRs shared across SHH, G3, and G4 subgroups was selected for further analysis. This subset was identified by first ranking the shared MRs according to their associated p-value from the MRAs from each

subgroup. Later, calculate the mean ranking for each of them across the three subgroups. Finally, the subset was selected based on the differences between their average ranking in an ordered list referred to as gaps. All consecutive gaps (differences between sorted values) are computed, and the target gap (gap with the highest value) is isolated. To avoid bias, the mean and standard deviation of the remaining gaps are calculated (excluding the target gap itself), ensuring the reference distribution reflects typical variability. A z-score is then derived to quantify how many standard deviations the target gap lies above the mean of the other gaps. This z-score provides a parametric measure of the gap's extremity, assuming approximate normality in the distribution of the remaining gaps.

Enrichment analysis

Gene Ontology (GO) biological process, KEGG pathways, and Reactome pathways enrichment analysis was performed to explore the biological links for the investigated regulons[37–39]. This analysis considered all genes regulated by the regulons of interest in the regulatory network. The *clusterProfiler*R/Bioconductor package was used to perform the enrichment analysis[40]. Terms enriched had the *p*-value adjusted with the Bonferroni correction method.

Network visualization and over-representation analysis

The *RedeR* R/Bioconductor package was employed to visualize the networks inferred by the *RTN*package[41]. There are three kinds of network representations in this study:

*i*The first method is the tree-and-leaf representation[42]. It aims to represent the regulatory network in two dimensions depicting regulons as nodes, labeled according to the TFs that regulate the respective group of genes. Node size represents the amount of genes in the regulon. The proximity between nodes represents the extent of overlap of regulated genes among regulons; the closer two regulons are portrayed on the same branch, the greater the number of regulated genes they share. This arrangement ultimately forms clusters of regulons based on the similarity of their contents.

ii The second one is the association map. In this visualization, regulons are also depicted as nodes, but the edges between them vary in width to represent the number of shared regulated genes between two regulons. Solid edges indicate regulatory agonism (positive regulation), while dotted edges denote antagonism (negative regulation).

iii The third method is the regulatory map. In this visualization, TFs are represented as diamonds, while target genes are shown as circles. Target genes are color-coded: red for up-regulated and blue for down-regulated genes, as determined by differential expression analysis of MB samples against controls. Edges between nodes are colored to reflect the type of regulation: red for activation and blue for inhibition, based on the MB regulatory network. If the color of the node and the edge conflict (e.g., there is an up-regulated target being inhibited), the edge is shown as dotted; otherwise, it is continuous.

Over-representation analysis and pairwise comparison

Regions in the network with an over-representation of MRs were calculated using the *phyper* function from the *stats* R package. Pairwise comparisons between MB's subgroups, cerebellum fetal control and cerebellum adult control were done using Dunn's nonparametric test with the *ggstatsplot*R package[43]. Adjusted *p*-values were obtained through Bonferroni correction.

Master regulators activity and survival analysis

The *RTN* package also infers the regulatory activities of a regulon for each sample, determined by an enrichment score. The *RTN* package calculates the activity of a particular regulon for each sample by first ranking a list of all genes by the differential gene expression observed when comparing that sample's expression with the average expression for all GSE85217 samples. Subsequently, it applies the Two-Tailed Gene Set Enrichment Score Analysis. If the up-regulated genes in the ranked list are primarily associated with positive targets, and the down-regulated genes are largely linked to negative targets, the regulon will receive a high positive enrichment score, indicating that the regulon is likely activated in the sample. This suggests a direct correlation between the TF and its regulated genes. Conversely, if up-regulated genes are predominantly linked to negative targets and down-regulated genes are linked to positive targets, the regulon will receive a high negative enrichment score. This negative score implies inhibition of the regulon in the sample, indicating an inverse relationship between the TF and its target genes. The outcomes of these analyses were portrayed as a heatmap with dendrograms, illustrating the regulon activity for all patients and how regulons are grouped based on the similarity of their enrichment scores across all samples. MRs' regulatory activities were also analyzed in conjunction with the overall survival of GSE85217 patients using the *RTNsurvival*R/Bioconductor package[44]. Risk MRs were identified with Cox multivariate regression analyses (BH adjusted p-value < 0.01) using the regulatory activities of all MRs as covariates in the model[36]. Kaplan-Meier curves followed by log-rank tests (p-value < 0.01) were used to visualise and confirm the prognostic value of the Risk MRs identified by the Cox model. Samples from the WNT subgroup were excluded from survival analysis, focusing on the most aggressive MB subgroups (SHH, G3 and G4)[30].

Results

Medulloblastoma master regulators

The MRA aims to identify TFs that play a critical role in driving specific gene expression patterns within a network. It's done by assessing the enrichment of a gene expression signature in each refined regulon in the

network. Therefore, we infer the MB regulatory network and its subgroup-specific gene signatures. The MB regulatory network was constructed with gene expression information of 732 MB samples and the list of human TFs. A total of 1,635 TFs and 19,615 potential gene targets formed 1,581 regulons (Fig. S1). The MB subgroup-specific (SHH, G3, and G4) gene signatures were defined by differential expression analyses of cancer against healthy fetal cerebellum tissue samples (Fig. S2, Table S2). Venn diagrams show that the signates made with the sets of eight cancer samples produces a similar list of differential expressed genes compared to using all samples available (Fig. S3). By applying the signatures to the MB network, the MRA inferred 51, 82, and 77 MRs for SHH, G3, and G4 subgroups, respectively, totalyzing 131 MRs, 20 of them identified as MRs across all three subgroups (Fig. S4).

To further investigate the disease MRs, we constructed a tree-and-leaf representation showing the MRs of SHH, G3, and G4 subgroups and their significance level (Fig. 1A). In a MRA, the significance level or *p*-values indicate the extent to which a TF regulates signature genes, adjusted for the size of its regulon. A smaller *p*-value suggests that a TF regulates a greater proportion of signature genes relative to its regulon size. Consequently, a highly significant MR is likely to have a regulatory function more directly associated with the biological role represented by the gene signature. Aiming at the most important regulators shared across the MB's subgroups, three MRs stood out being the most significant regulators in all three subgroups (Table S3, Fig. 1A). Ranking the 20 MRs shared across the MB's subgroups by significance level, *BHLHE41*, *RFX4* and *NPAS3* average ranking stands out compared to the others (Z-score *p*-value = 0.003, Fig. S5). The MR *BHLHE41* exhibited the highest statistical significance within the SHH subgroup, the second highest within the G3, and the third highest within the G4. The regulon RFX4 was also highly significant between groups, being ranked first for G3, second for SHH, and sixth for G4. The MR *NPAS3* was the third most statistically significant in the SHH subgroup, the eighth most in the G3, and the second most in the G4. GO biological processes enrichment analyses for the genes belonging to each regulon revealed shared processes between *BHLHE41* and *RRFX4* linked to myelination and

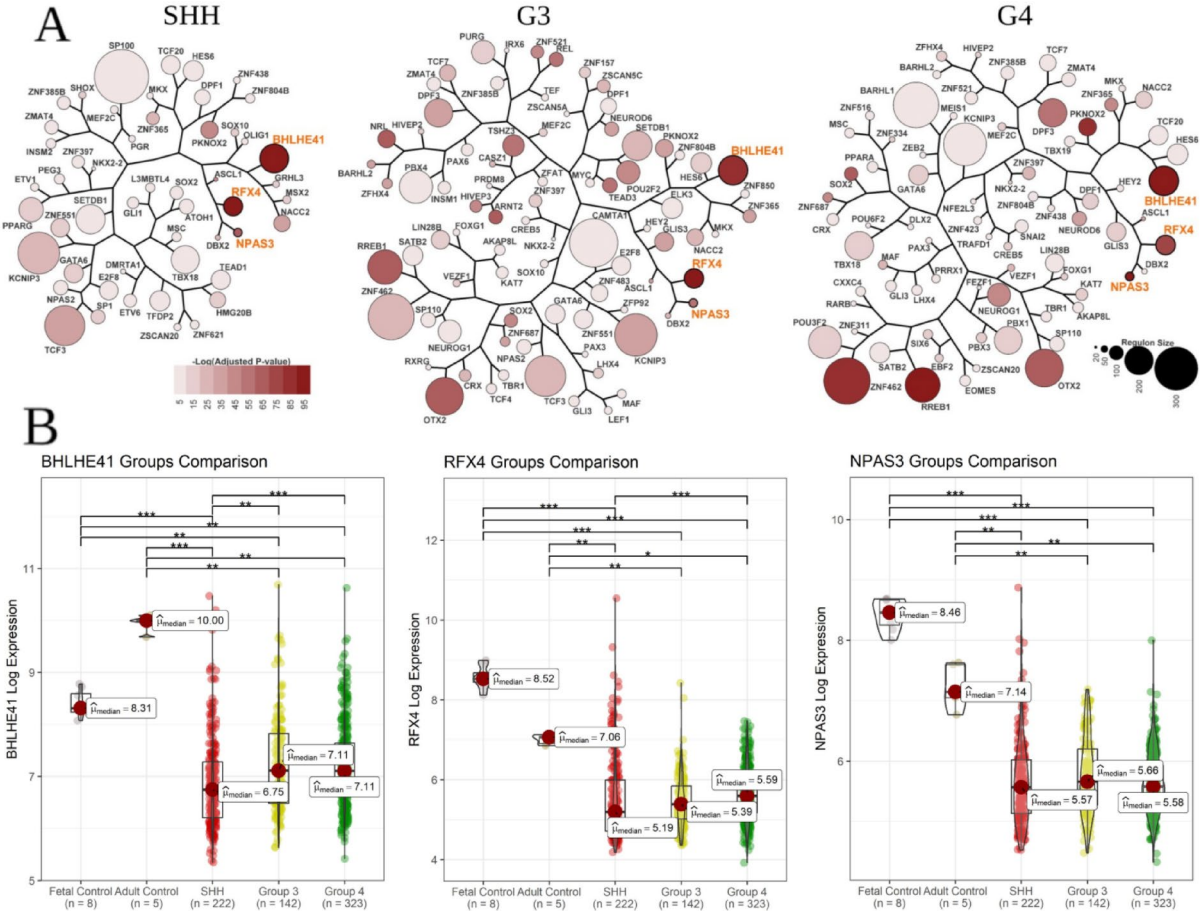

Fig. 1. **(A)** Three-and-leaf representation of the MRs identified for the SHH subgroup (left), G3 subgroup (center), and G4 subgroup (right). Nodes represent regulons, labeled according to the TFs that regulate the respective group of genes. MRs are colored according to their significance level. The proximity between nodes in the same branch represents the extent of overlap of regulated genes among regulons, i.e., the closer two regulons are portrayed, the greater the number of regulated genes they share. Regulons labeled in orange are the most significant shared across all subgroups. **(B)** Pairwise comparisons with Dunn's test between the MB subgroups, human fetal and adult healthy cerebellum controls. Bars show only significant results. Adjusted *p*-values ≤ 0.05 are denoted by one asterisk (*), ≤ 0.01 by two asterisks (**), and ≤ 0.001 by three asterisks (***).

glial cells development, while *NPAS3* processes are connected to synapse regulation and brain development (Fig. S6). To assess how these three MRs were regulating their target genes, we constructed their regulatory map in all evaluated subgroups. According to the network, these three MRs majorly regulate their target genes positively (Fig. S7), despite these targets being found down-regulated in tumor samples compared to control. It suggests that those ubiquitous regulons are suppressed in MB. In this line, we compare the gene expression of *BHLHE41*, *RFX4* and *NPAS3* across the MB's subgroups, human fetal, and adult healthy cerebellum as controls (Fig. 1B). Consistently, the expression of the three TFs is lower in the MBs subgroups compared to both adult and fetal controls.

Risk master regulator analysis

The impact of the MRs activities on patient outcome were accessed through multivariate Cox regression using the 131 MRs identified across the three subgroups. In this analysis, all GSE85217 samples, excluding WNT patients, were used. Eight Risk Master Regulators (RMRs) were identified with an adjusted p-value < 0.01 (Table S4). Interestingly, all of them were MRs exclusive from G3. The prognostic value of RMRs were confirmed with Kaplan-Maier analysis. For each RMR, samples were divided into two groups based on their enrichment score, one with positive activity and the other with negative activity. Kaplan-Meier curves were then constructed for these patient groups and compared using the Log-rank test. Two groups of RMRs can be identified: one comprising *REL*, *ZFAT*, *MYC*, and *ZSCAN5 A* (Fig. 2A), and another with *PAX6*, *ZNF157*, *ARNT2*, and *HIVEP3* (Fig. 2B). For all RMRs in Fig. 2A, patients with high activity had a worse outcome compared with patients with lower activity. Conversely, for the RMRs in Fig. 2B, lower activity was associated with a worse outcome.

The activities of the RMRs were represented across all samples as a heatmap (Fig. 2C). The group of four genes in purple showed high activity mainly in the G3 samples, and its activity was associated with a worse prognosis. The other group in green presented high activity associated with a better prognosis, with SHH samples predominantly showing high activities for this group. The opposite roles in survival between the two groups might be due to their shared target genes regulation. An association map was constructed showing that shared target genes are regulated in the same direction within each cluster and in opposite directions between the two groups (Fig. 2D).

Risk master regulators regulatory map

To better understand how the opposite roles in survival between the two groups of RMRs are influenced by the regulation of their target genes from the G3 signature, we created a regulatory map. This map includes the eight RMRs and the 159 target genes they regulate, with 39 of these targets regulated by more than one RMR (Fig. 3).

We observed two distinct regulatory patterns between the high-risk and low-risk RMRs. For high-risk RMRs, the regulatory effects—whether activating or inhibiting a specific target gene—matched the gene's expression in G3 tumor samples. If a high-risk RMR activated a target gene, that gene was upregulated in the tumor samples, and if it inhibited a gene, the gene was downregulated. In contrast, the low-risk RMRs showed an opposite pattern. Their regulatory roles did not align with the expression of the genes they targeted. Even when a target gene was regulated by both high- and low-risk RMRs, there was a clear difference in how each group of RMRs regulated the gene.

As an example for a high-risk regulator, the *REL* TF was assigned to positively regulate 13 target genes, and all of them were up-regulated in the tumor compared with normal samples. Similarly, for the 32 target genes that REL negatively regulates, all were down-regulated in tumor samples. For the low-risk regulator ZNF157, all the 19 target genes it was assigned to positively regulate were down-regulated in the tumor. Similarly, the four target genes it was assigned to negatively regulate were up-regulated in the tumor. The only exception is the gene NELL1, which represents less than 1% of the total regulatory interactions in the map.

Master regulators regions and its biological functions

The MB regulatory network, represented as a tree-and-leaf diagram (see Methods), highlights two regions with a significant concentration of MRs: Cluster A in the lower region and Cluster B in the upper region (Fig. 4). From the 77 G4 MRs identified in the regulatory network, 29 were located in the cluster A, representing a 5.73-fold enrichment compared to what would be expected by chance (hypergeometric p-value = 1.17e-16). Similar results were found in cluster B for G3 MRs, with 20 MRs out of a total of 82 (hypergeometric p-value = 3.06e-09). A third cluster, cluster C, was identified within cluster B, and showed an overrepresentation of MRs shared across all three MB subgroups (highlighted in gold in Fig. 4). Out of the 21 shared 'gold' MRs, 7 were found in Cluster C, indicating a 26-fold enrichment over the expected number (hypergeometric p-value = 1.68e-09).

Biological processes enrichment analysis was performed for each set of genes regulated by MRs: G4 for Cluster A, G3 for Cluster B, and MRs shared among the three MB subgroups for Cluster C (Fig. 4D, Table S5). Cluster C is enriched for processes related to myelination and glial cells, such as ensheathment of neurons, axon ensheathment, myelination, central nervous system (CNS) myelination, axon ensheathment in the CNS, gliogenesis, and glial cell differentiation. Cluster B is also enriched for all such processes, while cluster A shows enrichment for all except two. Exclusive processes of cluster A are sensory system related, while cluster B is ion transport related. Processes shared between these two clusters are related to neuronal projection, like axon guidance, neuron projection guidance, axogenesis, and regulation of neuron projection development. KEGG and Reactome pathways enrichment analysis were also done with the set genes (Fig. S8).

Discussion

In this work, we constructed the MB regulatory network and detected the TFs most associated with the disease, focusing on the most aggressive MB subgroups[30]. We did not investigate the WNT subgroup. It is the rarest and least aggressive subgroup of all four, representing 10% of cases and manifesting a 5-year survival rate exceeding

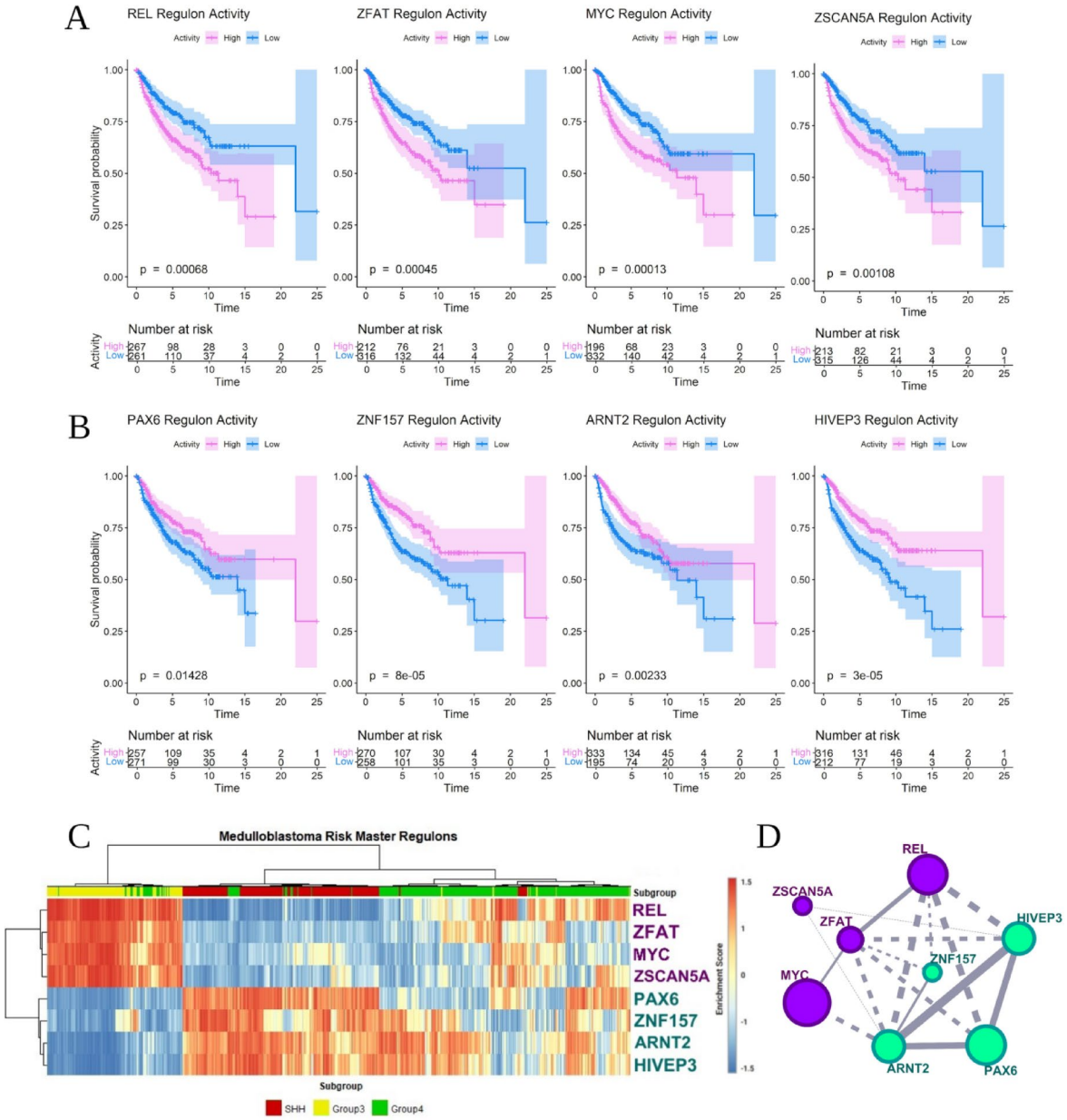

Fig. 2. RMRs activities and their impact on patients' outcome. Kaplan-Meier curves in years for each RMRs between high and low-risk patients classified by their respectively enrichment scores for worse outcome RMRs (**A**)*REL*, *ZFAT*, *MYC*, and *ZSCAN5 A*, and (**B**) better outcome RMRs *PAX6*, *ZNF157*, *ARNT2*, and *HIVEP3*. The shaded regions at the Kaplan-Meier curve denote its confidence intervals. (**C**) Heatmap of regulatory activity for MB's RMRs with dataset GSE85217. (**D**) Association maps of MB's RMRs. Node size expresses the amount of genes in the regulon and edge width reflects the quantity of genes mutually regulated by a pair of regulons. Continuous edges symbolize regulatory agonism and dotted edges indicate regulatory antagonism. Node colors represent the two major clusters of the regulon activity dendrogram.

90%[5], with secondary neoplasms and therapy complications exerting a heavy influence on the 10% portion of lethal victims[45]. The development of targeted therapies to WNT MBs have not been proved to be fruitful, since targeting the WNT pathway would meddle with critical developmental processes and could extinguish the convenient chemosensitivity of WNT-activated MBs. Current clinical trials do not apply targeted treatments on this subgroup, instead, treatment de-intensification is being tested for patients of low and intermediate risk[11].

To refine our analysis, we selected the three most statistically significant MRs that were shared across the three subgroups. Among the 20 MRs shared between SHH, G3, and G4, three MRs–*BHLHE41*, *RFX4*, and *NPAS3*–stood out as the most significant across all subgroups, meaning they regulate, proportionally to their regulon size, the highest number of target genes within each subgroup's gene signatures. This approach ensures that our results emphasize regulators deeply involved in the specific biological processes under investigation, rather than more general regulators affecting diverse pathways. Interestingly, all positively regulated targets of

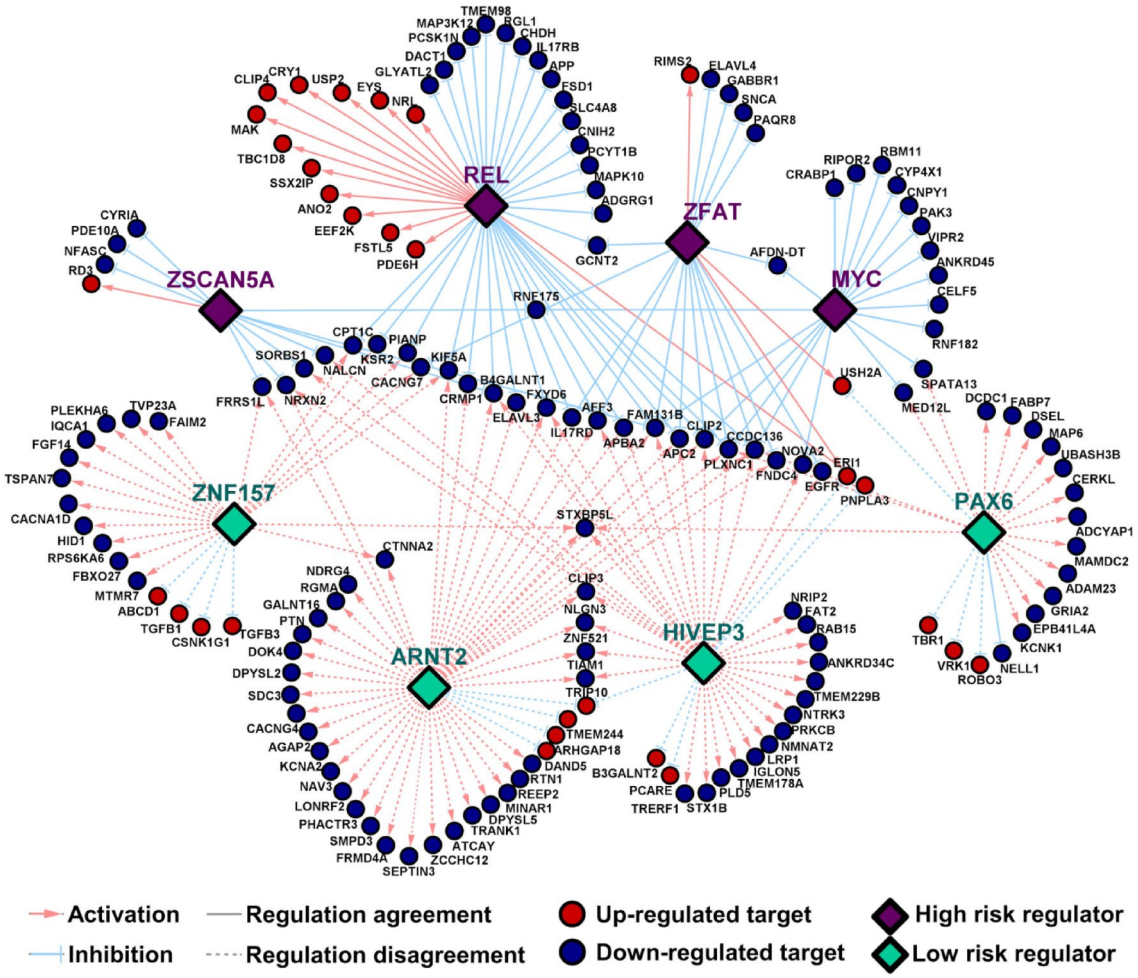

Fig. 3. Regulatory map of the MB RMRs. High-risk regulators (purple diamonds), have their target gene regulation predominately agreeing with their designated function (activation/inhibition) defined by the regulatory network. Low-risk regulators (green diamonds) have their target gene regulation predominately disagreeing with their designated function (activation/inhibition) defined by the regulatory network.

those MRs were downregulated in tumor samples (Fig. S7), suggesting an impairment in the regulatory function of *BHLHE41*, *RFX4*, and *NPAS3* in MB samples, particularly SHH, G3, and G4 subgroups. Additionally, their expression was also lower in tumor samples compared to the cerebellum controls.

BHLHE41(also known as DEC2/Sharp1) regulates 187 targets in the MB network, including 75 from the SHH gene signature, 77 to the G3 signature, and 65 to the G4 signature. This TF is expressed in the suprachiasmatic nucleus of the hypothalamus, peaking during the subjective day, and it is known to play important roles in normal cells, particularly as an essential regulator of the circadian rhythm[46]. Disruptions in this biological circuit are often associated with an increased risk of cancer. Research indicates that individuals who work night shifts face a higher likelihood of developing prostate, lung, and colon cancers compared to those who have never worked at night[47]. In the context of pediatric cancer, excessive daytime sleepiness in children is linked to a greater incidence of CNS tumors, with sleep disorders reported in approximately 40% of all childhood cancer cases[48]. *BHLHE41*appears to have a complex, context-dependent role in tumor progression. It promotes initial tumor growth in renal cancer but suppresses later progression. Conversely, it seems to act as a tumor suppressor in lung, breast, and colon cancers by inhibiting invasion, metastasis, and inducing cell death[49–52]. In MB, according to our results, *BHLHE41* tends to have a tumor suppressor role given its regulatory function impairment as a highly significant MR. Furthermore, its down-regulation in tumor samples compared to both controls tested: healthy fetal cerebellum and healthy adult cerebellum emphasize its inhibition in the tumor.

RFX4 TF was the second most significant MR in the SHH subgroup. *RFX4* plays a critical role in the Sonic Hedgehog signaling pathway by modulating ciliogenesis, essential for the proper functioning of this pathway in a region-specific manner during the CNS development. *RFX4*modulates SHH pathway activity and Gli protein processing through its control over ciliogenesis and cilia-related gene transcription[53]. This regulation is crucial for proper dorsoventral CNS patterning and is executed in a region-specific manner, highlighting *RFX4*'s importance in developmental processes. Additionally, *RFX4*is involved in neural tube morphogenesis and is required for dorsal and ventral midline formation in the caudal neural tube in zebrafish[54]. *RFX4* is also an

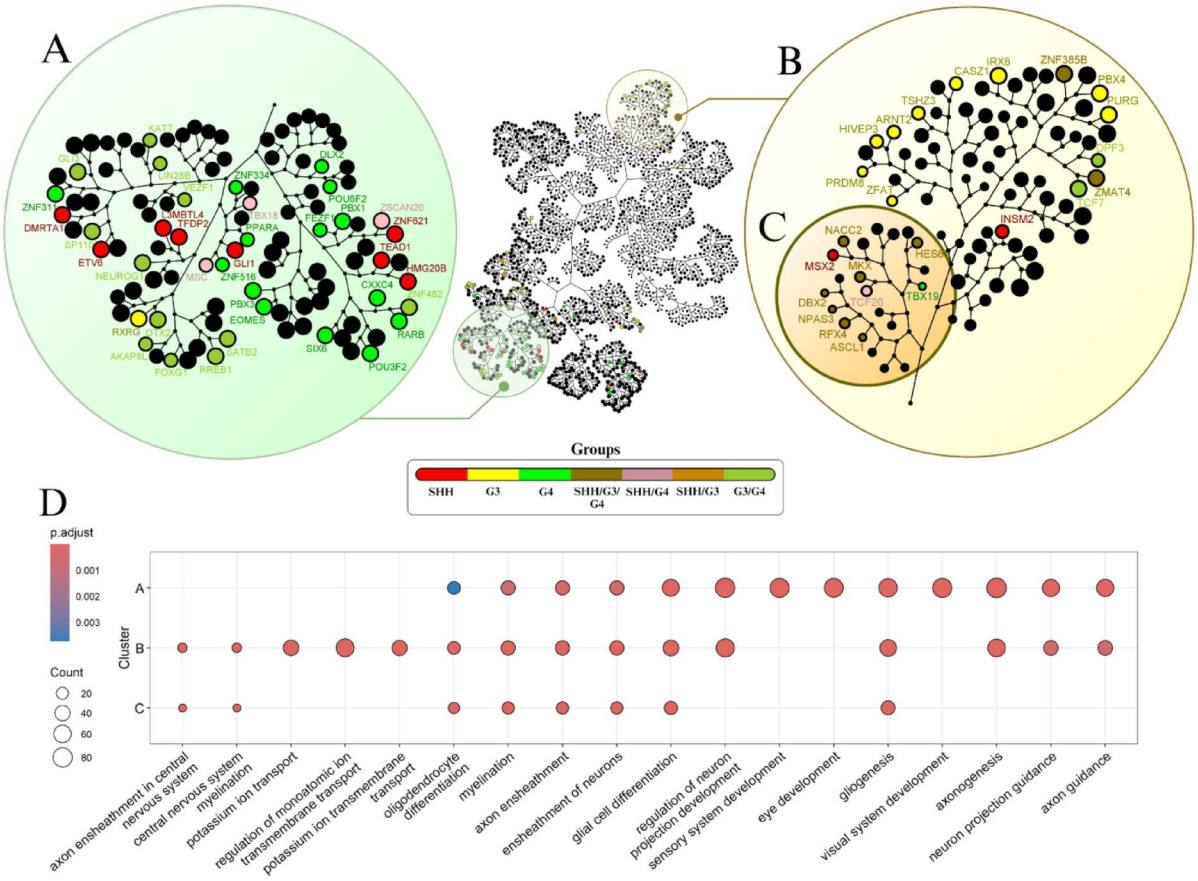

Fig. 4. Tree-and-leaf representation of all regulons inferred for the MB regulatory network. Nodes represent regulons, labeled according to the TFs that regulate each group of genes. The proximity between nodes corresponds to the amount of regulated genes they share. The highlighted clusters are overrepresented with MRs, specifically cluster **(A)** with G4 regulators, cluster **(B)** with G3 regulators, and cluster **(C)** with regulators shared for subgroups SHH, G3, and G4. **(D)** Top Gene Ontology biological processes enriched for the regulons of clusters A, B, and C, as determined by clusterProfileR's enrichment analysis.

intrinsic factor for neuronal differentiation, inducing proneural genes such as *POU3 F2* and *NEUROD1*, crucial for neural cell differentiation[55]. Our results emphasize *RFX4*'s importance in CNS development and suggest its regulatory inactivity as a critical factor in transitioning from normal neuronal tissue to cancerous cells in SHH, G3, and G4 MB subgroups.

NPAS3(Neuronal PAS Domain Protein 3) is primarily known for its role in mental health and neurodevelopment as shown in the GO enrichment, but research suggests it may also act as tumor suppressor, although this area is less well-studied compared to its neurological functions[56–58]. With a tumor suppressor role particularity in astrocytomas, *NPAS3* shows those features roles driving its progression by modulating cell cycle, proliferation, apoptosis and cell migration. Additionally, the overexpression of *NPAS3*in malignant glioma cell lines suppressed its transformation, while an opposite expression induced more aggressive growth[56–58]. Notably, *NPAS3*was already categorized as a MRs, but in its involvement in neurodevelopment and mental health, specifically by regulating neuropsychiatric risk genes[59]. In our study, interestingly *NPAS3* and *RFX4* MRs are side-by-side in the regulatory network, as shown in Fig. 4C, suggesting that they may share some biological functions regulating many of the same target genes and potentially shared tumor suppressing features.

We identified eight MRs associated with patient survival. Some, such as *PAX6* and *MYC*, are well-known genes linked to MB, while the roles of the others are still not described in the disease[15,60,61]. The identification of *PAX6* and *MYC*, both of which are established genes of MB biology, suggests that our method is capable of detecting TFs with strong prognostic and functional relevance, surviving as a supportive indicator of the method's ability to prioritize biologically relevant regulators. *MYC*is a critical molecular biomarker for MB prognosis, especially in G3, where its amplification is strongly associated with a worse prognosis[62]. Our analysis revealed that *MYC* regulatory activity is highest in G3 patients, lower in SHH patients, and more broadly spread in G4 patients (Fig. 2A). *MYC* regulatory activity follows the same pattern as other high-risk RMRs, while an opposite activity pattern is observed in the low-risk RMRs. These activity patterns suggest that RMRs with divergent patterns have antagonistic regulatory roles, regulating target genes in opposite directions, while RMRs with the same pattern have synergistic roles, regulating target genes in the same direction (Fig. 2B). Furthermore, RMRs with opposite activation patterns show opposite patterns in their target gene regulation outcomes. High-risk RMRs'

regulatory roles in the MB network corresponded with their target gene expression outcomes, while low-risk RMRs exhibited regulatory disagreements. This observation suggests that high-risk RMRs actively promote tumor progression and aggressiveness during tumor development, whereas low-risk RMRs had their regulatory capabilities attenuated or inhibited during tumorigenesis, preventing them from acting as suppressors. It is important to note that the regulatory network construction and the identification of differentially expressed genes from the G3 signature are independent analyses. Although the samples used in both analyses were from the same dataset, they were different MB samples.

To better understand the biological roles of the MRs, we analyzed the MB regulatory network with a topological view. This way, it revealed two regions with overrepresentation of MRs: one cluster overrepresented with G4 regulators (Fig. 4A) and another with G3 regulators (Fig. 4B). Despite being in opposite regions of the network, these two regions share many enriched biological processes, such as axogenesis, myelination, and gliogenesis, emphasizing their importance for G3 and G4 subgroups. Within cluster B, we identified a smaller region called cluster C (Fig. 4C), which is 26-fold overrepresented with MRs shared among the three MB subgroups. The biological processes enriched by these MRs are strongly linked to myelination, gliogenesis, and glial cell differentiation. Since many MRs in this region are master for all three subgroups, these processes might act as foundational mechanisms for MB development, with the MRs in this region being important regulators. Astrocytes, a glial cell subtype, play an indispensable role in MB tumorigenesis and progression by secreting hedgehog proteins[63]. Astrocytic trans-differentiation, where tumor neural progenitors differentiate into astrocytes in the tumor microenvironment, stimulates microglia to produce growth factors that promote tumor progression[64]. Additionally, in mouse models, astrocyte progenitors transformed by MYC, a classic biomarker for G3, lead to G3-like tumors based on histology and gene expression profiles[31,62]. Therefore, the MRs in cluster C may regulate these mechanisms or other unknown processes.

Conclusion

Our study provides new insights into the transcription regulators involved in MB development. TFs such as *BHLHE41*, *RFX4*, and *NPAS3*, which had mostly unknown roles in the disease, were shown to regulate a significant portion of the genes involved in the MB tumorigenesis. Their presumably inability to properly regulate their targets suggests that their inhibition may be necessary during disease progression, indicating their potential as tumor suppressors for the SHH, G3, and G4 MB subgroups. Another set of MRs were identified as associated with patients' outcome. The eight RMRs described exhibited distinct activities and regulatory patterns, depending on whether they were classified as low or high-risk for patient outcomes. Altogether, the TFs identified in this study stand out as promising biomarkers with potential application for future target treatments for MB. Integrating experimental validation with our computational framework is a necessary next step to advance these MRs from candidate targets to actionable drivers. Such efforts will not only solidify their biological roles but also accelerate the development of targeted therapies, addressing the urgent unmet need for reducing treatment toxicity in pediatric MB. Further studies are needed to verify the exact biological mechanisms behind the roles of *BHLHE41*, *RFX4* and *NPAS3* MRs potentially acting as tumor suppressors in MB and the mechanisms behind the impact in survival of the RMRs identified in our study.

Data availability

The data on which this study is based is open and available under the accession identifiers GSE85217 and GSE167447 in the Gene Expression Omnibus (GEO) data repository (https://www.ncbi.nlm.nih.gov/geo/).

Received: 15 December 2024; Accepted: 30 April 2025
Published online: 10 May 2025

References

1. Northcott, P. A. Medulloblastoma. *Nat. Rev. Dis. Primers* **5**, 11 (2019).
2. Thompson, M. C. et al. Genomics identifies Medulloblastoma subgroups that are enriched for specific genetic alterations. *J. Clin. Oncol.* **24**, 1924–1931 (2006).
3. Northcott, P. A. et al. Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29**, 1408–1414 (2011).
4. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* **23**, 1231–1251 (2021).
5. Taylor, M. D. et al. Molecular subgroups of Medulloblastoma: the current consensus. *Acta Neuropathol.* **123**, 465–472 (2012).
6. Massimino, M. et al. Childhood Medulloblastoma. *Crit. Rev. Oncol. Hematol.* **105**, 35–51 (2016).
7. Northcott, P. A. et al. The whole-genome landscape of Medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
8. Juraschka, K. & Taylor, M. D. Medulloblastoma in the age of molecular subgroups: a review. *J. Neurosurg. Pediatr.* **24**, 353–363 (2019).
9. Wang, J., Garancher, A., Ramaswamy, V. & Wechsler-Reya, R. J. Medulloblastoma: from molecular subgroups to molecular targeted therapies. *Annu. Rev. Neurosci.* **41**, 207–232 (2018).
10. Robinson, G. W. et al. Irreversible growth plate fusions in children with Medulloblastoma treated with a targeted Hedgehog pathway inhibitor. *Oncotarget* **8**, 69295–69302 (2017).
11. Menyhárt, O. & Győrffy, B. Molecular stratifications, biomarker candidates and new therapeutic options in current Medulloblastoma treatment approaches. *Cancer Metastasis Rev.* **39**, 211–233 (2020).
12. Kool, M. et al. Genome sequencing of SHH Medulloblastoma predicts genotype-related response to smoothened Inhibition. *Cancer Cell.* **25**, 393–405 (2014).
13. Bushweller, J. H. Targeting transcription factors in cancer - from undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624 (2019).
14. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
15. Roussel, M. F. & Robinson, G. W. Role of MYC in Medulloblastoma. *Cold Spring Harb. Perspect. Med.* **3**, (2013).
16. Bondeson, D. et al. Catalytic in vivo protein knockdown by small-molecule protacs. *Nat. Chem. Biol.* **11**, 611–617 (2015).
17. Chen, Y. N. P. et al. Allosteric Inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148–152 (2016).

18. Tisato, V., Voltan, R., Gonelli, A., Secchiero, P. & Zauli, G. MDM2/X inhibitors under clinical evaluation: perspectives for the management of hematological malignancies and pediatric cancer. *J. Hematol. Oncol.* **10**, 133 (2017).
19. Zecena, H. et al. Systems biology analysis of mitogen activated protein kinase inhibitor resistance in malignant melanoma. *BMC Syst. Biol.* **12**, 33 (2018).
20. da Ribeiro-Dantas, M. C. et al. Reverse engineering of ewing sarcoma regulatory network uncovers PAX7 and RUNX3 as master regulators associated with good prognosis. *Cancers (Basel)* **13**, (2021).
21. Castro, M. A. A. et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* **48**, 12–21 (2016).
22. Lim, W. K., Lyashenko, E. & Califano, A. Master regulators used as breast cancer metastasis classifier. *Pac. Symp. Biocomput.* 504–515. https://doi.org/10.1142/9789812836939_0048 (2009).
23. Albanus, R. D. et al. Reverse engineering the neuroblastoma regulatory network uncovers MAX as one of the master regulators of tumor progression. *PLoS ONE* **8**, e82457 (2013).
24. Fletcher, M. N. C. et al. Master regulators of FGFR2 signalling and breast cancer risk. *Nat. Commun.* **4**, 2464 (2013).
25. Sartor, I. T. S., Zeidán-Chuliá, F., Albanus, R. D., Dalmolin, R. J. S. & Moreira, J. C. F. Computational analyses reveal a prognostic impact of TULP3 as a transcriptional master regulator in pancreatic ductal adenocarcinoma. *Mol. Biosyst.* **10**, 1461–1468 (2014).
26. Clough, E. et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* **52**, D138–D144 (2024).
27. Cavalli, F. M. G. et al. Intertumoral heterogeneity within Medulloblastoma subgroups. *Cancer Cell* **31**, 737–754e6 (2017).
28. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).
29. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics–a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009).
30. Liu, Y., Xiao, B., Li, S. & Liu, J. Risk factors for survival in patients with Medulloblastoma: A systematic review and Meta-Analysis. *Front. Oncol.* **12**, 827054 (2022).
31. Tao, R. et al. MYC drives group 3 Medulloblastoma through transformation of Sox2 + Astrocyte progenitor cells. *Cancer Res.* **79**, 1967–1980 (2019).
32. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7** (Suppl 1), S7 (2006).
33. Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
34. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
35. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statistic. Soc.: Series B (Methodological)* **57**, 289–300 (1995).
37. Gene Ontology Consortium, et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
38. Kanehisa, M. & Goto, S. K. E. G. G. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
39. Milacic, M. et al. The reactome pathway knowledgebase 2024. *Nucleic Acids Res.* **52**, D672–D678 (2024).
40. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
41. Castro, M. A. A., Wang, X., Fletcher, M. N. C., Meyer, K. B. & Markowetz, F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biol.* **13**, R29 (2012).
42. Cardoso, M. A. et al. TreeAndLeaf: an R/Bioconductor package for graphs and trees with focus on the leaves. *Bioinformatics* **38**, 1463–1464 (2022).
43. Patil, I. Visualizations with statistical details: The ggstatsplot approach. *JOSS* 6, 3167 (2021).
44. Groeneveld, C. S. et al. RTNsurvival: an R/Bioconductor package for regulatory network survival analysis. *Bioinformatics* **35**, 4488–4489 (2019).
45. Ellison, D. W. et al. Definition of disease-risk stratification groups in childhood Medulloblastoma using combined clinical, pathologic, and molecular variables. *J. Clin. Oncol.* **29**, 1400–1407 (2011).
46. Honma, S. et al. Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* **419**, 841–844 (2002).
47. Parent, M. É., El-Zein, M., Rousseau, M. C., Pintos, J. & Siemiatycki, J. Night work and the risk of cancer among men. *Am. J. Epidemiol.* **176**, 751–759 (2012).
48. Rosen, G. & Brand, S. R. Sleep in children with cancer: case review of 70 children evaluated in a comprehensive pediatric sleep center. *Support. Care Cancer* **19**, 985–994 (2011).
49. Chen, S. et al. BHLHE41 Overexpression Alleviates the Malignant Behavior of Colon Cancer Cells Induced by Hypoxia via Modulating HIF-1α/EMT Pathway. *Gastroenterol. Res. Pract.* 6972331 (2022). (2022).
50. Montagner, M. et al. SHARP1 suppresses breast cancer metastasis by promoting degradation of hypoxia-inducible factors. *Nature* **487**, 380–384 (2012).
51. Nagata, T. et al. BHLHE41/DEC2 expression induces autophagic cell death in lung Cancer cells and is associated with favorable prognosis for patients with lung adenocarcinoma. *Int. J. Mol. Sci.* **22**, (2021).
52. Bigot, P. et al. Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. *Nat. Commun.* **7**, 12098 (2016).
53. Ashique, A. M. et al. The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci. Signal.* **2**, ra70 (2009).
54. Sedykh, I. et al. Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Dev. Dyn.* **247**, 650–659 (2018).
55. Choi, W. et al. RFX4 is an intrinsic factor for neuronal differentiation through induction of proneural genes POU3F2 and NEUROD1. *Cell. Mol. Life Sci.* **81**, 99 (2024).
56. Moreira, F. et al. NPAS3 demonstrates features of a tumor suppressive role in driving the progression of Astrocytomas. *Am. J. Pathol.* **179**, 462–476 (2011).
57. Erbel-Sieler, C. et al. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc. Natl. Acad. Sci. USA* **101**, 13648–13653 (2004).
58. Jastrzębski, M. K., Wójcik, P., Stępnicki, P. & Kaczor, A. A. Effects of small molecules on neurogenesis: neuronal proliferation and differentiation. *Acta Pharm. Sin. B* **14**, 20–37 (2024).
59. Michaelson, J. J. et al. Neuronal PAS domain proteins 1 and 3 are master regulators of neuropsychiatric risk genes. *Biol. Psychiatry* **82**, 213–223 (2017).
60. Zagozewski, J. et al. An OTX2-PAX3 signaling axis regulates group 3 Medulloblastoma cell fate. *Nat. Commun.* **11**, 3627 (2020).
61. Shahi, M. H. et al. Regulation of Sonic hedgehog-GLI1 downstream target genes PTCH1, Cyclin D2, Plakoglobin, PAX6 and NKX2.2 and their epigenetic status in Medulloblastoma and Astrocytoma. *BMC Cancer* **10**, 614 (2010).
62. Ramaswamy, V. et al. Risk stratification of childhood Medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol.* **131**, 821–831 (2016).
63. Liu, Y. et al. Astrocytes promote Medulloblastoma progression through Hedgehog secretion. *Cancer Res.* **77**, 6692–6703 (2017).
64. Yao, M. et al. Astrocytic trans-Differentiation completes a multicellular paracrine feedback loop required for Medulloblastoma tumor growth. *Cell* **180**, 502–520e19 (2020).

Acknowledgements
This study was financed by the governmental Brazilian Agency Coordination for the Improvement of Higher Education Personnel (CAPES – Portuguese: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, project number 88887.692895/2022-00 and 88887.752882/2022-00), National Council of Technological and Scientific Development (CNPq – Portuguese: Conselho Nacional de Desenvolvimento Científico e Tecnológico, project number 312305/2021-4), and PROPESQ-UFRN. We thank the STRING, SRA database, and GWAS catalog for freely providing their data. We also would like to thank NPAD/UFRN for computational resources.

Author contributions
Original manuscript writing, data acquisition, processing, analyses, study design were performed by GLM and TSM. The authors DOI, JVAC and DRS contributed to data acquisition, processing, and analyses. RJSD supervised the project and together with IDS, MCRD, OCM and MS contributed to data interpretation, study conception and design. All authors contributed to article reviewing and approved the final manuscript.

Funding
This study was financed by the Brazilian Agency Coordination for the Improvement of Higher Education Personnel (CAPES – Portuguese: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - project number 88887.692895/2022-00 and 88887.752882/2022-00), National Council of Technological and Scientific Development (CNPq – Portuguese: Conselho Nacional de Desenvolvimento Científico e Tecnológico, project number 312305/2021-4 and grant number 406484/2022-8 (INCT BioOncoPed)), PROPESQ-UFRN, Ronald McDonald Institute through the McDia Feliz 2022 Campaign - Financial Code CO 2022062 and Children's Cancer Institute from Brazil. We also would like to thank the Instituto Metrópole Digital (IMD, UFRN) and NPAD for providing the computing infrastructure required to conduct the analyses done in this work, as well as the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) for making available all datasets used in this study.

Declarations

Competing interests
The authors declare no competing interests.

Conflict of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-00763-3.

Correspondence and requests for materials should be addressed to R.J.S.D.