Reverse Engineering of Medulloblastoma Regulatory Network and Inference of Master Regulons

Gustavo Lovatto Michaelsen[1,2,3,†], Tayrone de Sousa Monteiro[1,†], Danilo Oliveira Imparato[1], João Vitor Almeida da Costa[1], Daniel Rocha Silva[2], Marialva Sinigaglia[1,2,3], Rodrigo Juliani Siqueira Dalmolin[1,4],

[1] Bioinformatics Multidisciplinary Environment-BioME, Digital Metropole Institute, Federal University of Rio Grande do Norte, Natal, RN, 59076-550, Brazil; | [2] Children's Cancer Institute, Porto Alegre, RS, 90620-110, Brazil; | [3] National Science and Technology Institute for Children's Cancer Biology and Pediatric Oncology – INCT BioOncoPed, Porto Alegre, RS, 90035-003, Brazil; | [4] Department of Biochemistry, Federal University of Rio Grande do Norte, Natal, RN, 59064-741, Brazil; | [†] These authors contributed equally to this work; | bioinfo_pesquisa2@ici.ong.

Introduction

Among malignant pediatric brain tumors, medulloblastoma (MB) is the most common and aggressive, comprising almost 10% of all childhood and juvenile brain neoplasms. Currently, the molecular classification recognizes four MB subgroups that are genetically distinct: wingless (WNT), sonic hedgehog (SHH), Group 3 and Group 4. High-aggressive subgroups such as Group 3 and Group 4 lack well-defined tumor drivers, hindering the development of targeted therapies. Considering that the lack of accurate markers is a major problem for improving clinical outcome in patients with the disease, there is a growing search for new high precision and robust molecular biomarkers with viable clinical application. Identifying key transcriptional regulators, known as master regulators (MRs), can elucidate the dysregulated pathways underlying MB progression and uncover potential treatment targets.

Objective

In this study we construct the MB regulatory network and identify key transcription factor for its development known as Master Regulators (MRs). Our goal was to gain insights into the molecular mechanisms driving these tumors and identify novel therapeutic opportunities, addressing the urgent need for more effective and less toxic treatments.

Methods

Flowchart of Master Regulator Analysis (MRAs). To obtain the gene signatures for each MB subgroup, small subsets of tumor gene expression samples, obtained from the GEO database (GSE 85212), were compared against healthy cerebellum control samples (GSE1167447). On the other hand, the MB regulatory network was inferred using all the other MB samples from GSE85212 and a list of human transcription factors (TFs). The network infers regulatory units composed of a TF and its target genes in the disease. Afterward, MRAs identify the regulatory units enriched in target genes from each MB subgroup's signatures, named as Master Regulators.
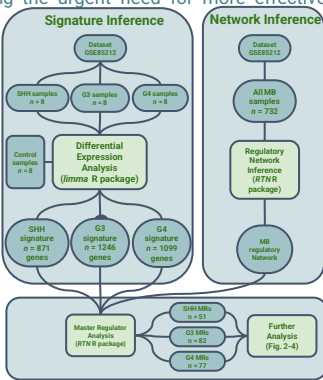
Results

The MB regulatory network was constructed based on the GSE85217 dataset (N = 732). A total of 1635 TFs and 19615 potential gene targets formed 1581 regulons. Alongside, subgroup-specific gene signatures were defined by contrasting transcriptional information of malignant against healthy tissue. Finally, with the regulatory network defined and the signatures built, Master Regulator Analysis (MRA) inferred 51, 82, and 77 MRs for SHH, G3, and G4 subgroups respectively (**Fig. 1**).
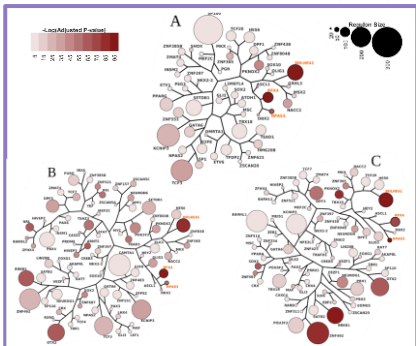
Figura 1. Three-and-leaf representation of the master regulators identified in the A SHH subgroup, B Group 3 subgroup, and C Group 4 subgroup. Nodes symbolize regulons, labeled according to the transcription factors that regulate the respective group of genes. Master regulators are colored according to their significance level. The proximity between nodes 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 ones shared across all subgroups.

The MB regulatory network presented regions with a high representation of MRs, one in its bottom region, and another in its upper part referred to as (**Fig. 2**). The cluster A had 29 MRs from G4 of a total of 77 present in the network, while the expected number by change would be 5. Overrepresentation analysis revealed a hypergeometric p-value of 1.17e-16. Similar results were encountered in cluster B but for G3 MRs and cluster C for MRs shared in the three MB subgroups.
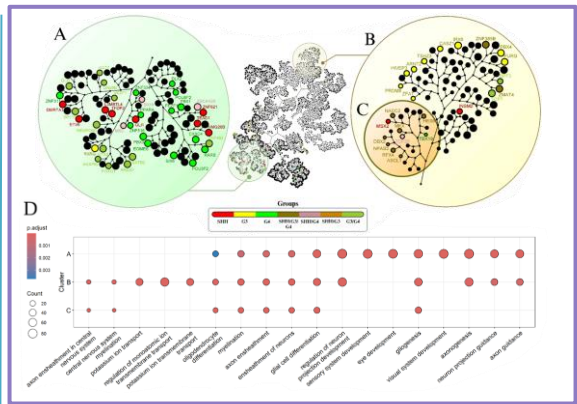
Figure 2. Tree-and-leaf representation of all regulons inferred for the medulloblastoma regulatory network. Nodes symbolize regulons, labeled according to the transcription factors that regulate each group of genes. The proximity between nodes corresponds to the amount of regulated genes they share. The highlighted clusters are overrepresented with master regulators, specifically cluster A with Group 4 regulators, cluster B with Group 3 regulators, and cluster C with regulators shared for subgroups SHH, Group 3, and Group 4. D Main Gene Ontology biological processes enriched for the regulons of clusters A, B, and C, as determined by clusterProfileR's enrichment analysis.

The impact of the MRs activities on patients' outcome were accessed through multivariable Cox regression using the 131 MRs identified across the three subgroups. Risk Master Regulators (RMRs) activities were represented across all samples as a heatmap (**Fig. 3**).
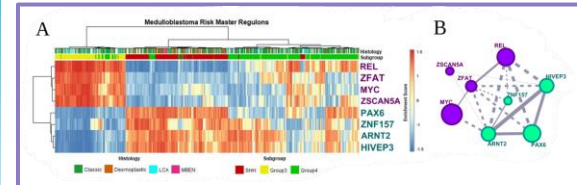
Figure 3. A Heatmap of regulatory activity for medulloblastoma RMRs with dataset GSE85217. Subgroups and histology are shown according to the classification provided by the work that published the data (Cavalli et al. 2017). **B** Association maps of medulloblastoma RMRs. Node size expresses the amount of genes in the regulon and edge width reflects the quantity of genes mutually regulated by a pair of regulons. Continuous edges symbolize regulatory agreement and dotted edges indicate regulatory antagonism. Node colors represent the two major clusters of the regulon activity dendrogram.

To assess how the RMRs were regulating their target genes, we constructed a regulatory map comprising the eight RMRs and the 159 targets they regulate (**Fig. 4**). Very distinct patterns were observed between the high and the low-risk regulators.
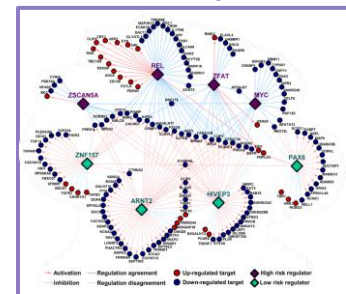
Figure 4. Regulatory map of the medulloblastoma RMRs. Regulators associated with worse outcome (purple diamonds), have their target gene regulation majorly agreeing with their designated function (activation/inhibition) defined by the regulatory network.

Very distinct patterns were observed between the high and the low-risk regulators. For the high RMRs, the regulation role they were assigned by the regulatory network, whether it was the activation or the inhibition of a particular target gene, agreed if that gene was up or down-regulated in the signature.

Conclusion

Our study brought a new understanding of the transcription regulators involved in MB development and agressivement. Transcription factors such as *BHLHE41*, *RFX4*, and *NPAS3* still have mostly unknown characterization in the disease and here they were shown to regulate a great portion of the genes involved in the MB tumorigenesis.