

RESEARCH

Open Access



Unraveling the glycosphingolipid metabolism by leveraging transcriptome-weighted network analysis on neuroblastic tumors

Arsenij Ustjanzew^{1*}, Annekathrin Silvia Nedwed¹, Roger Sandhoff², Jörg Faber^{3,4}, Federico Marini^{1,5} and Claudia Paret^{3,4,5}

Abstract

Background Glycosphingolipids (GSLs) are membrane lipids composed of a ceramide backbone linked to a glycan moiety. Ganglioside biosynthesis is a part of the GSL metabolism, which involves sequential reactions catalyzed by specific enzymes that in part have a poor substrate specificity. GSLs are deregulated in cancer, thus playing a role as potential biomarkers for personalized therapy or subtype classification. However, the analysis of GSL profiles is complex and requires dedicated technologies, that are currently not included in the commonly utilized high-throughput assays adopted in contexts such as molecular tumor boards.

Methods In this study, we developed a method to discriminate the enzyme activity among the four series of the ganglioside metabolism pathway by incorporating transcriptome data and topological information of the metabolic network. We introduced three adjustment options for reaction activity scores (RAS) and demonstrated their application in both exploratory and comparative analyses by applying the method on neuroblastic tumors (NTs), encompassing neuroblastoma (NB), ganglioneuroblastoma (GNB), and ganglioneuroma (GN). Furthermore, we interpreted the results in the context of earlier published GSL measurements in the same tumors.

Results By adjusting RAS values using a weighting scheme based on network topology and transition probabilities (TPs), the individual series of ganglioside metabolism can be differentiated, enabling a refined analysis of the GSL profile in NT entities. Notably, the adjustment method we propose reveals the differential engagement of the ganglioside series between NB and GNB. Moreover, *MYCN* gene expression, a well-known prognostic marker in NTs, appears to correlate with the expression of therapeutically relevant gangliosides, such as GD2. Using unsupervised learning, we identified subclusters within NB based on altered GSL metabolism.

Conclusion Our study demonstrates the utility of adjusting RAS values in discriminating ganglioside metabolism subtypes, highlighting the potential for identifying novel cancer subgroups based on sphingolipid profiles. These findings contribute to a better understanding of ganglioside dysregulation in NT and may have implications for stratification and targeted therapeutic strategies in these tumors and other tumor entities with a deregulated GSL metabolism.

*Correspondence:

Arsenij Ustjanzew

arsenij.ustjanzew@uni-mainz.de

Full list of author information is available at the end of the article



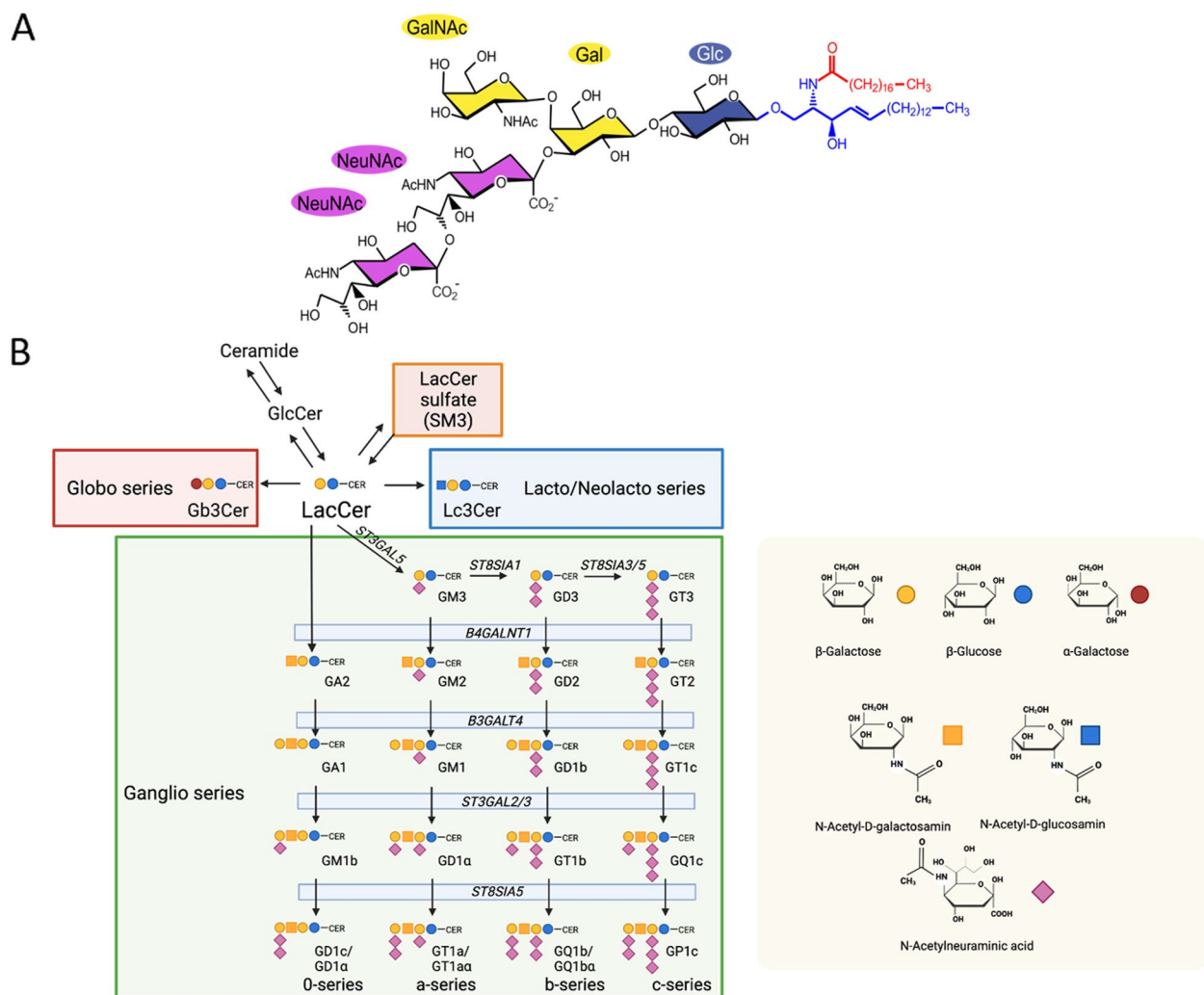
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords Neuroblastoma, Ganglioneuroma, Ganglioneuroblastoma, Ganglioside, GD2, Glycosphingolipids, Reaction activity score, Metabolic graph

Background

Gangliosides are sialylated glycosphingolipids (GSLs) found in high concentration in the nervous system [1, 2]. Gangliosides are composed of a lipid anchor, consisting of sphingosine, a fatty acid of different lengths, and a polar head group composed of sialylated oligosaccharides (Fig. 1A)

[1, 2]. The GSL biosynthesis pathway, shown in Fig. 1B, is a complex series of coordinated reactions occurring in the Golgi apparatus [3, 4]. Early steps in this pathway, which guide into a-, b-, and c-series are performed with enzymes of relative high substrate specificity, whereas downstream enzymes are promiscuitive and elongate all four series [5]. This circumstance results in the fact that the same enzymes



are involved in reactions of all four series, for example, the protein encoded by *B3GALT4* converts GA2, GM2, GD2, and GT2 into the more complex gangliosides GA1, GM1, GD1b, and GT1c, respectively. Further, the pattern of ganglioside expression changes during the development in normal brain tissues. Simple gangliosides such as GD2 are expressed in the embryo, while mature neurons express complex gangliosides of the a-, and b-series, namely GM1a, GD1a, GD1b, and GT1b [6–9].

GSL profiles are often found to be deregulated in tumors [10, 11]. These changes have been particularly studied in neuroblastic tumors (NT) [12–15]. NTs are the most common extra-cranial solid tumors in children and include neuroblastoma (NB), ganglioneuroblastoma (GNB), and ganglioneuroma (GN) [16–18]. They arise from primitive neuroectodermal cells of the neural crest, which give rise to the sympathetic nervous system during the embryonic development [18]. One prominent feature of NTs is their biological and clinical heterogeneity. NT range from immature, undifferentiated to mature, differentiated tumors, where GNB and GN represent the mature end of this range [18, 19]. NBs often express simple gangliosides, particularly GD2, lacking or having reduced concentrations of the more complex a-, and b-series gangliosides. More mature NT such as GNB express very low amounts of GD2 and show instead an increased level of complex gangliosides of the a- and b-series [15, 20].

Ganglioside profiles in NB have been linked to prognosis and are relevant for therapy. Patients with tumors expressing higher concentrations of the b-series gangliosides, GD1b, and GT1b have a better prognosis [21]. The ganglioside GD2 is used as a target for the treatment of high-risk NB with monoclonal antibodies (Dinutuximab and Naxitamab, both already used in clinical protocols) and CAR-T cells (in clinical studies for several tumor entities) [22–24].

The simple gangliosides GD3 and GD2 regulate several receptors and signaling cascades and contribute to the undifferentiated state of neural stem cells and NB [25]. Discordant hypotheses exist on the mechanism of GD2 accumulation, with some reports claiming a low expression of *ST8SIA1* (required for the synthesis of GD3) and others a high expression of the same gene as the reason for GD2 expression [26, 27].

To understand and exploit GSLs, a better characterization of their molecular profiles is necessary. However, the measurement of the different GSLs is complex and requires dedicated protocols and instruments such as thin-layer chromatography and mass spectrometry [15]. RNA sequencing (RNA-seq) data obtained from patient-derived samples, on the other hand, provides

a powerful opportunity for studying the expression of the genes required for ganglioside synthesis, as it is frequently collected during high-dimensional molecular cancer profiling, such as in the context of a Molecular Tumor Board [28]. However, the translation of transcriptome data into ganglioside profiles is challenging, and few models have been proposed so far to predict the expression of selected gangliosides from RNA-seq data [29–31].

In the following text, the terms network and graph are used interchangeably. The GSL biosynthesis pathway can be studied as a metabolite-centric network, where nodes are metabolites and edges represent the metabolic reactions. The edges can be weighted by the reaction activity scores (RAS) (first introduced by Graudenzi A. et al. (2018)), which can be considered as measurement of the activity of each reaction in the metabolic graph and is calculated using the normalized gene expression levels of a patient sample and the gene-protein-reaction (GPR) association rules of a genome-scale metabolic model [32–34]. As a result of the low specificity of the involved enzymes, the reaction activity cannot be distinguished across the four series by simply weighting edges of the network by the respective RAS values. This prevents the analysis of a cell type-specific GSL profile based on transcriptional regulation and hinders a more detailed understanding of the underlying biological processes and functions associated with the ganglioside metabolism pathway.

In this study, our primary objectives were 1) to develop a method based on RNA-seq data that can discriminate between the four series of the ganglioside metabolism pathway, 2) to test the method on NT entities, and finally 3) to interpret the results in the context of published ganglioside measurements. We hereby introduced three adjustment methods for the RAS by incorporating the network topology and its transition probabilities (TPs) across the four series. We demonstrated how these weighting schemes can be used in 1) an exploratory, unsupervised manner, and 2) in the direct comparison between several defined groups by applying our approach on two NT datasets. On the one hand, our approach enhances our comprehension of GSL deregulation in NT, and its transferability suggests potential application to other tumor entities, prospectively leading to the development of new pre-screening tools for assessing GSL metabolism in patients. On the other hand, it is conceivable that the method applied here may be transferable to other metabolic pathways involving non-specific enzymes, broadening its utility beyond the scope of ganglioside metabolism.

Methods

Workflow

We followed the workflow schematically illustrated in Fig. 2. Two publicly accessible RNA-seq datasets containing NT samples were acquired. The analysis workflow was conducted separately for each dataset. Both datasets were preprocessed by data normalization and log10 transformation with one pseudocount. The resulting normalized count matrices were used for further processing steps. A GSL metabolic graph was constructed (see further details in “Graph construction” section below), representing the reactions

between metabolites and genes in the dataset. The RAS for each reaction of the metabolic graph were computed resulting in a weighted, directed graph per sample and a RAS matrix over all samples.

A transition probability (TP) matrix was computed for each weighted graph, describing the probability of moving from one node to another in the metabolic pathway based on the RAS values of the outgoing edges. Two alternative versions of the TP matrix were also computed to address the discrimination of the ganglioside 0-, a-, b-, and c-series (see subsection “Alternative transition probability matrices”). Finally, the

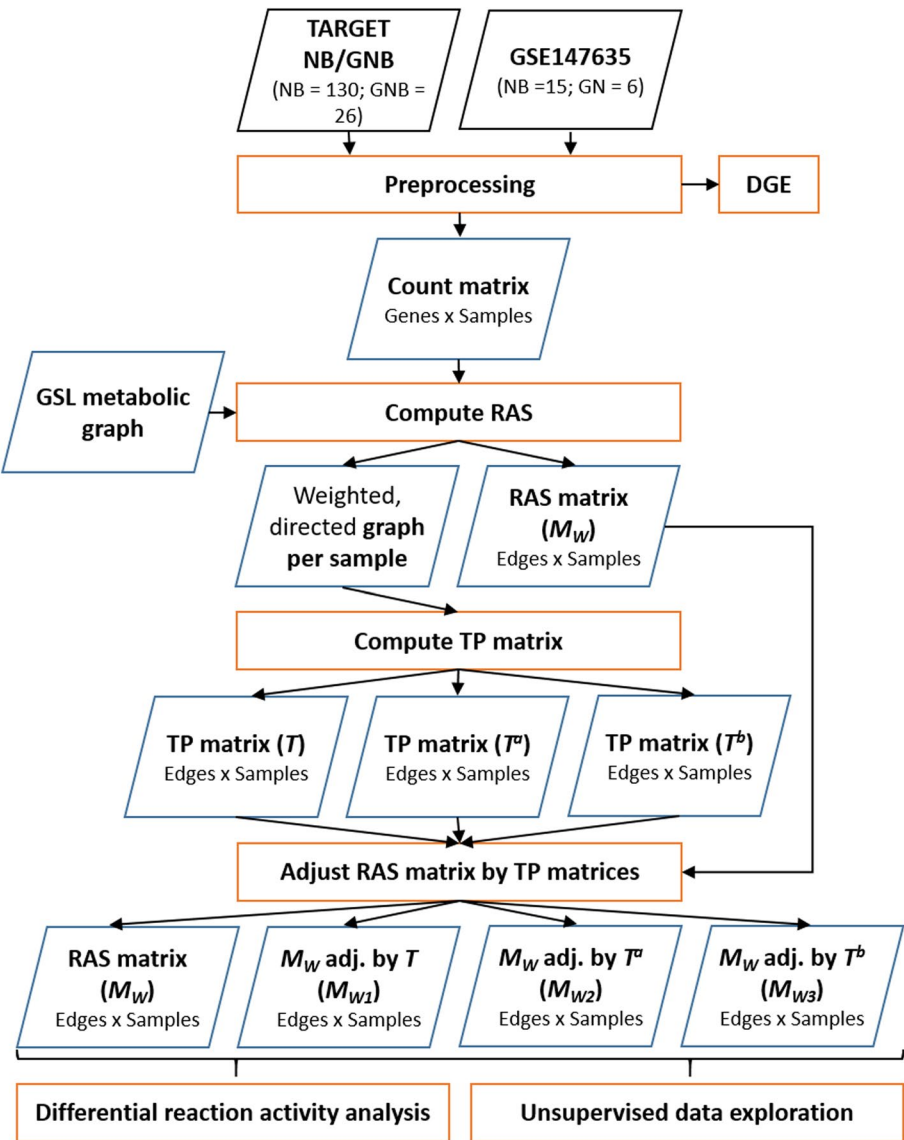


Fig. 2 Workflow diagram representing the data processing steps (orange rectangles), the input data (black parallelograms), and the intermediate outputs (blue parallelograms). NB=Neuroblastoma; GNB=Ganglioneuroblastoma; GN=Ganglioneuroma; DGE=Differential gene expression; GSL=Glycosphingolipid; RAS=Reaction activity score; TP=Transition probability

RAS matrix was adjusted using the transition matrices, resulting in four distinct matrices.

The (adjusted) RAS matrices were further used for (1) unsupervised data exploration, and (2) differential reaction analysis between groups. For the first approach, unsupervised machine learning methods were used for exploratory analysis of the adjusted RAS values. This includes dimensionality reduction with UMAP, identification of clusters of samples based on their RAS profiles, and the identification of statistically differential expressed reactions between clusters. In the differential reaction analysis, the dataset-specific groups were compared by computing a log₂ fold-change of the RAS values. *P*-values per reaction were determined by using the Kolmogorov-Smirnov (KS) test.

The workflow was executed entirely in the R environment (version 4.1.3). Box plots, dot plots and scatter plots were generated with ggplot2 (version 3.4.2) [35] and plotly (version 4.10.1). Heatmaps were drawn with the pheatmap library (version 1.0.12). The networks were visualized with the R package igraph (version 1.4.2) [36].

Data acquisition and preprocessing

Two datasets were used:

(1) Expression data and corresponding clinical information were obtained from GSE147635 [37] (*n*=21). This dataset contains 6 GN and 15 NB samples. The RASflow pipeline was used as the framework for the quantification of gene expression [38]. HISAT2 tool [39] in combination with HTSeq [40] was used on the human genome GRCh38 (GCA_000001405.15) and the GENCODE annotation (version 31) [41] to retrieve gene counts. In the further course of the text this dataset is called GN/NB dataset.

(2) The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) NB gene-expression profile (“RSEM expected_count”) and clinical data (*n* = 162) were acquired from the UCSC Xena database (<http://xena.ucsc.edu/>) [39] as $\log_2(\text{expected_count} + 1)$. The gene counts were back-transformed to integer counts. This dataset contains 26 GNB and 130 NB. Six samples of unknown subtypes were excluded. Further, this dataset contains the “MYCN.status” variable, resulting in 33 MYCN-amplified (+MYCN), 128 without a MYCN amplification (-MYCN), and one sample with unknown MYCN amplification status. This one sample of unknown status was removed. A GNB sample has the annotation stating that it is MYCN-amplified. A MYCN amplification is rare in GNB, so we also removed it from the dataset due to lack of power as it was detected in only one sample. The final dataset remained 154 samples in total. In the following, we refer to this dataset as the TARGET GNB/NB dataset.

All datasets were separately normalized by the median of ratios method using the DESeq2 R package (version 1.34.0) [42]. A pseudocount was added to the counts and then log₁₀ transformed. For the distance heatmaps of both datasets, the raw counts were normalized and a variance stabilizing transformation was applied using the vst function of DESeq2.

Graph construction

A metabolite-centric graph, where nodes are metabolites and edges connect the metabolites involved in the same reactions, was built with the R packages NetPathMiner (version 1.30.0) [43] and igraph (version 1.4.2) [36]. The metadata of the edges contains the involved enzymes. The graph consists of the four pathways (hsa00600, hsa00601, hsa00603, hsa00604) from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [44] (the KGML files were retrieved on 07/07/2022) and represents the sum of sphingolipid & glycosphingolipid metabolism including the lacto-, neolacto-, globo-, and ganglio series. The reactions R06010 and R06004 representing the degradation of GM1 to GM2 and GM2 to GM3 were removed from the graph. This is justified because these reactions represent the degradation process and are not competing enzymatic processes of the biosynthesis. The resulting metabolic graph is used in later steps as a template graph. Additionally, most enzymes in the ganglioside biosynthesis pathway are exclusively involved in the GSL pathways [45]. Out of the 90 considered enzymes, 56 are also involved in non-GSL pathways, which are often related to lipids or glycans, such as the Sphingolipid signaling pathway, Lysosome, and Ether lipid metabolism. The final graph annotation can be seen in Fig. S1 and Additional file 2 and an igraph object can be found at https://github.com/arsenij-ust/NT_GSL_analysis.

Weighting of the metabolic graph

Computing RAS values

The normalized gene expression values $C_{i,j}$ of a patient sample in the form of an $n \times m$ matrix C , where n is the number of genes and m is the number of samples, are used to compute the RAS. RAS defines the amount of activity in a certain condition, for each reaction of the metabolic graph (Fig. 3A). RAS was computed as described in Graudenzi A. et al. [32] based on the GPR association rules of the genome-scale metabolic model of the Homo sapiens (Human1) [46]. GPRs are logical formulas explaining the association between gene products in the process of catalyzation of a given reaction. These formulas involve logical operators, such as AND and OR. Thus, for each sample $s = 1, \dots, n$, and each reaction r the RAS is computed based on the following distinctions:

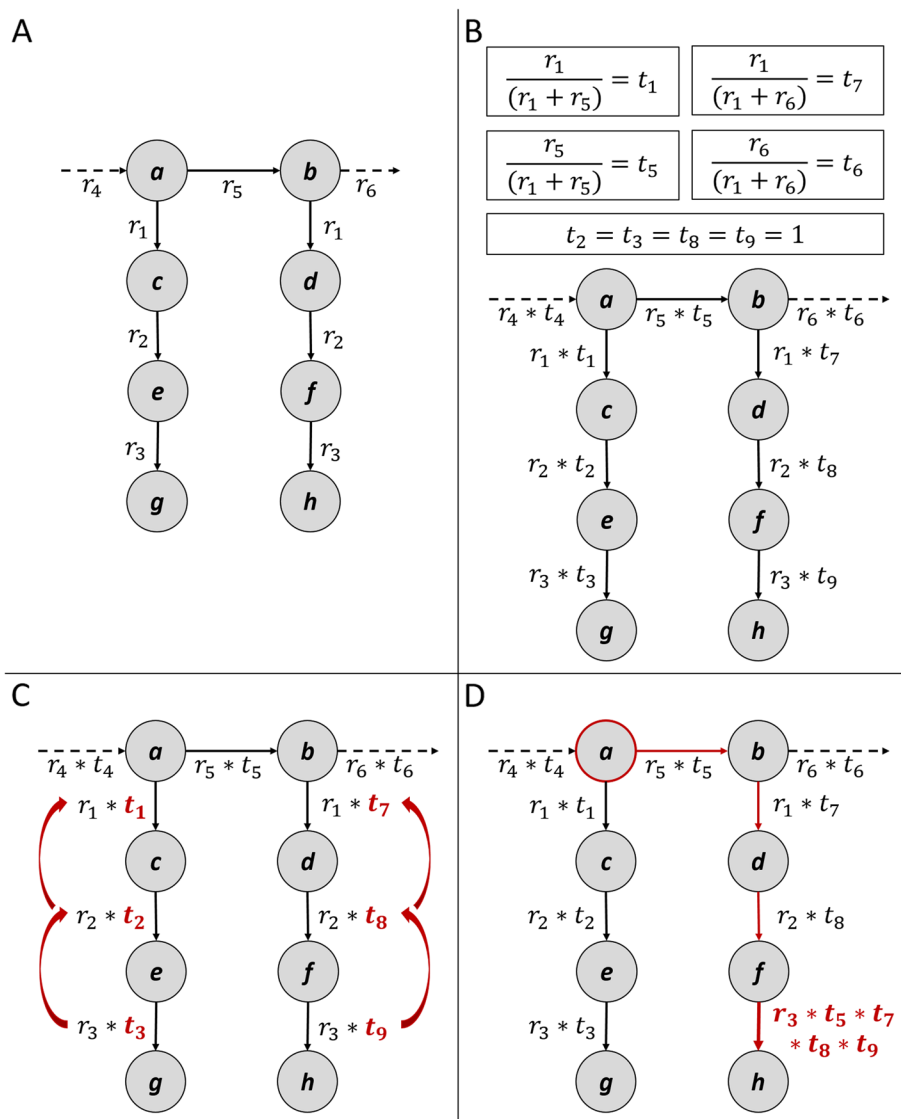


Fig. 3 Visual examples of introduced adjustment methods based on transition probability. **A** An exemplary graph, where nodes represent metabolites (*a* - *h*) and edges are RAS values of the respective reactions (*r*). Although formally the reactions from nodes *a* to *c* and *b* to *d* are separate reactions with distinct reaction IDs, the RAS values of these reactions are identical because of the involvement of the exact same genes. Therefore, we denote these reactions with identical *r*-numbers (*r*₁) in this figure. The same also applies to edges *r*₂ and *r*₃. Ingoing and outgoing edges are dashed. **B** Edge weights are adjusted by multiplying the TP *t* with the RAS value of the respective reaction *r*. Exemplary calculations of the TPs are given for *t*₁, *t*₅, *t*₆, and *t*₇. The TP of *t*₂, *t*₃, *t*₈, and *t*₉ are equal to 1 because nodes *c*, *d*, *e*, and *f* have only one outgoing edge. **C** First alternative adjustment method, in which the TP is equal to 1, recursively takes the TP of the previously incoming edges. **D** Second alternative adjustment method, where exemplarily the RAS value *r*₃ is adjusted by the product of TPs along the path starting from the defined node *a* to node *h*

1) Reactions with AND operator (i.e., enzyme subunits):

$$RAS_{s,r} = \min(C_{i,j} : U_r) \quad (1)$$

where U_r is the set of genes that encode the subunits of the enzyme catalyzing reaction *r*.

2) Reactions with OR operator (i.e., enzyme isoforms):

$$RAS_{s,r} = \sum_{i \in I_r} \quad (2)$$

where I_r is the set of genes that encode isoforms of the enzyme that catalyzes reaction *r*.

RAS values of a sample are used to weight the associated edges, resulting in a weighted directed graph $G = (V, E, W)$, where *V* and *E* are the set of all nodes

(metabolites) and edges (reactions) in G , and W is the weighted adjacency matrix containing the RAS values. Our resulting metabolic graph contains 116 reactions, of which two reactions have AND logic, 51 reactions have OR logic, and 63 reactions have only one assigned gene.

Computing the transition probability matrix

Let $G_1, G_2, \dots, G_s$ be a list of s graphs (one graph per sample), where each graph $G_i = (V, E, W)$ has an associated transition matrix T_i . T_i is computed as follows:

Let W_i be the $n_i \times n_i$ weighted adjacency matrix for graph G_i , where the element $w_{i,j}$ represents the RAS value of the reaction catalyzed from node i to node j in graph G_i .

The row sums of W_i can be computed as:

$$rs_{i,j} = \sum_{k=1}^{n_i} w_{i,j,k} \quad (3)$$

where $rs_{i,j}$ represents the sum of the j th row of W_i .

The transition matrix T_i can be computed by dividing each element of W_i by the corresponding row sum (Fig. 3B):

$$t_{i,j} = \frac{w_{i,j}}{rs_{i,j}} \quad (4)$$

where $t_{i,j}$ represents the element of the transition matrix T corresponding to the TP from node i to node j in graph G_i . In other words, $t_{i,j}$ is the probability of node j being reached in one step by a random walker located in node i .

Alternative transition probability matrices

Further, two alternative transition matrices T^a & T^b are computed to address the problem of identical RAS values of the 0-, a-, b-, and c-series of the ganglioside metabolism pathway.

In case of T^a , the transition matrix is computed as described above. If $i = 1$ in $t_{i,j}$, we set $t_{i,j}$ to the TP of the incoming edge of node i with the largest probability. If the TP of the incoming edge used to set $t_{i,j}$ is also 1, we recursively repeat the process to find the TP of the previous edge until a value $\neq 1$ is found. If no incoming edges exist, the TP stays 1 (Fig. 3C).

In case of T^b , let x be a node in G_i . We can compute the transition matrix T^b for G_i by computing the sum of TPs along all simple paths from x to all other nodes in G_i . Specifically, for each simple path $P_{x,j}$ from x to j in G_i , we compute the sum of TPs along the path as follows:

$$t_{P_{x,j}} = \prod_{(i,k) \in P_{x,j}} t_{i,k} \quad (5)$$

where (i, k) represents the edge in G_i from node i to node k . This is similar to the product rule for probability along a branch in a tree diagram. This method is illustrated in Fig. 3D.

We set $t_{P_{x,j}}$ as the new TP value of the last edge of $P_{x,j}$. As G is not strongly connected, we account for nodes that are not reachable from x by setting the TP of the respective edges to 0.

In this study, we selected lactosylceramide (C01290) as the x -node, since the sphingolipid metabolism pathway splits into the lacto-/neolacto-, globo-, and ganglio series at this point.

Alternative transition probability matrices

Finally, the original RAS matrix W is adjusted by T , T^a , and T^b , as follows, resulting in three additional weight matrices per graph G_i .

$$W_1 = W * T \quad (6)$$

$$W_2 = W * T^a \quad (7)$$

$$W_3 = W * T^b \quad (8)$$

Consequently, we obtain one matrix M_W , that represents unadjusted RAS values; three matrices M_{W_1} (Eq. 6), M_{W_2} (Eq. 7), M_{W_3} (Eq. 8), which combine the RAS matrix with the TPs lowering the RAS values proportionally. These matrices represent *reactions* $\times$ *samples*, where the value $M_{i,j}$ represents the edge weight of reaction i in sample j .

UMAP plot generation from the adjusted RAS matrices

We generated multiple Uniform Manifold Approximation and Projection (UMAP) [47] plots using the adjusted RAS matrices as input and varied the number of neighbors. The UMAP algorithm is a nonlinear dimensionality reduction method that maps high-dimensional data onto a low-dimensional space while maintaining the local structure of the data. We visually inspected the UMAP plots to identify patterns in the data. The UMAP R package (version 0.2.10) was used. The choice of parameters in UMAP is crucial for achieving a meaningful and balanced representation of the data. UMAP has two primary parameters: *n_neighbors* and *min_dist*, which are instrumental in controlling the trade-off between local and global structure in the resulting projection. We adopted an iterative approach, investigating different *n_neighbors* and *min_dist* values for each RAS adjustment method within our study. Through this process, we aimed to find a balanced representation between the local and global structures of the data. For the GN/NB dataset the final *n_neighbors* and *min_dist* values were 3 and

0.1 across all RAS matrices. For the TARGET NB/GNB dataset we selected the value 20 for $n_neighbors$ and 0.1 for min_dist in case of the RAS matrix and the value 12 for $n_neighbors$ and 0.1 for min_dist for the adjusted RAS matrices. Additionally, to ensure the reproducibility of our research given UMAP's stochastic nature, we used the *random_state* parameter to set a seed for random number generation.

Clustering the UMAP results

The HDBSCAN clustering method of the *dbscan* R package (version 1.1-11) [48] was used to process the UMAP manifolds after creating the UMAP graphs. To find clusters, HDBSCAN examines local density peaks in the data. High-dimensional data with a variety of densities and shapes may be processed with this scalable technology. We adjusted the minimum size of clusters (*minPts*) by inspecting the sample number per cluster for different values of *minPts* (Fig. S8) and selected a *minPts* value, where the number of outlier samples (cluster 0) is minimal and the number of samples per cluster is moderate. For the TARGET NB/GNB dataset the *minPts*-value was set to 10 and for the GN/NB dataset the *minPts*-value was set to 5. Note, that an optimal *minPts* value strongly depends on the UMAP layout.

Estimating the stability of UMAP results

As UMAP representations are the result of a heuristic algorithm, each time the UMAP plot is generated the position of the samples in the UMAP plot differs, although the local structure of the data is still preserved. To estimate the stability of the UMAPs and identify the number of clusters for more than one UMAP without cherry-picking one UMAP plot, we iteratively computed 1000 UMAPs on $M_W, \dots, M_{W_3}$, summed up the x- and y-coordinates of each sample, and plotted 50 random selected iterations as a line plot. Additionally, we computed the C-index, the Calinski-Harabasz index, and the silhouette score across the iterations to assess the clustering performance by using the R package *clusterCrit* (version 1.3.0).

Marker identification

Marker reactions were identified for each cluster based on the results of the clustering analysis. In our study, we used the R function *findMarkers* from the *scrn* package (version 1.22.1) [49] to identify markers for each cluster identified by HDBSCAN clustering. We set the *pval.type* to “all” and the *test.type* to “t” to perform a two-sample t-test between each cluster and the rest of the samples. The function returns a list of significant markers for each cluster along with their log-fold change, false-discovery rate, and *p*-values. The resulting list of markers allowed

us to gain insights into the biological processes and reactions that are enriched in each cluster.

Differential reaction activity analysis between two groups

Using the respective groups of the datasets, namely (1) NB MYCN+ against NB MYCN-, (2) GNB against NB, and (3) GN against NB, the log2 fold-change of the average adjusted RAS per reaction was calculated between the groups. The nonparametric statistical KS test can be used to compare a sample distribution to a reference or theoretical distribution. The null hypothesis is that the sample comes from the reference distribution. For each reaction the KS test was performed with a default *p*-value threshold equal to 0.05, to test if there is a significant difference in the adjusted RAS distributions across the samples in the two groups. The *p*-values were adjusted for multiple testing with the Benjamini-Hochberg (BH) method.

Differential gene expression analysis & gene set enrichment analysis

Differential gene expression (DGE) analysis of the RNA-seq data between 19 TARGET GNB/NB samples related to the identified “cluster 1” and 136 samples of the other clusters was performed using the DESeq2 package (version 1.34.0) by fitting the negative binomial generalized linear model for each gene and using the Wald test for significance testing. Genes of the count matrix with less than 10 counts in the sum of all samples were excluded. The False Discovery Rate (FDR) was set to 0.05. Benjamini-Hochberg correction was used to obtain adjusted *p*-values. The “apeglm” log2 fold shrinkage method was used [50]. Additional file 3 provides the result of DGE.

Differentially expressed genes were used to perform enrichment analysis by testing the over-representation of Gene Ontology (GO) terms (only terms assigned to the category Biological Process were analyzed). The analysis was conducted by the R packages *topGO* (version 2.46.0) [51] and *pcaExplorer* (version 2.20.2) [52]. The *elim* algorithm [53] was used for the enrichment testing as well as Fisher's exact test. This step was performed independently for up- and down-regulated genes. All expressed genes of the RNA-seq data were used as background genes. An enrichment map was computed by the R package *GeneTonic* [54].

Results

Comparison of RAS Adjustment Methods: Investigating Differential Reaction Activities of the GSL pathway using the TARGET GNB/NB dataset

In this study, we utilized the two datasets, TARGET GNB/NB and the GN/NB dataset, to investigate the transcriptional regulation of the ganglioside metabolism

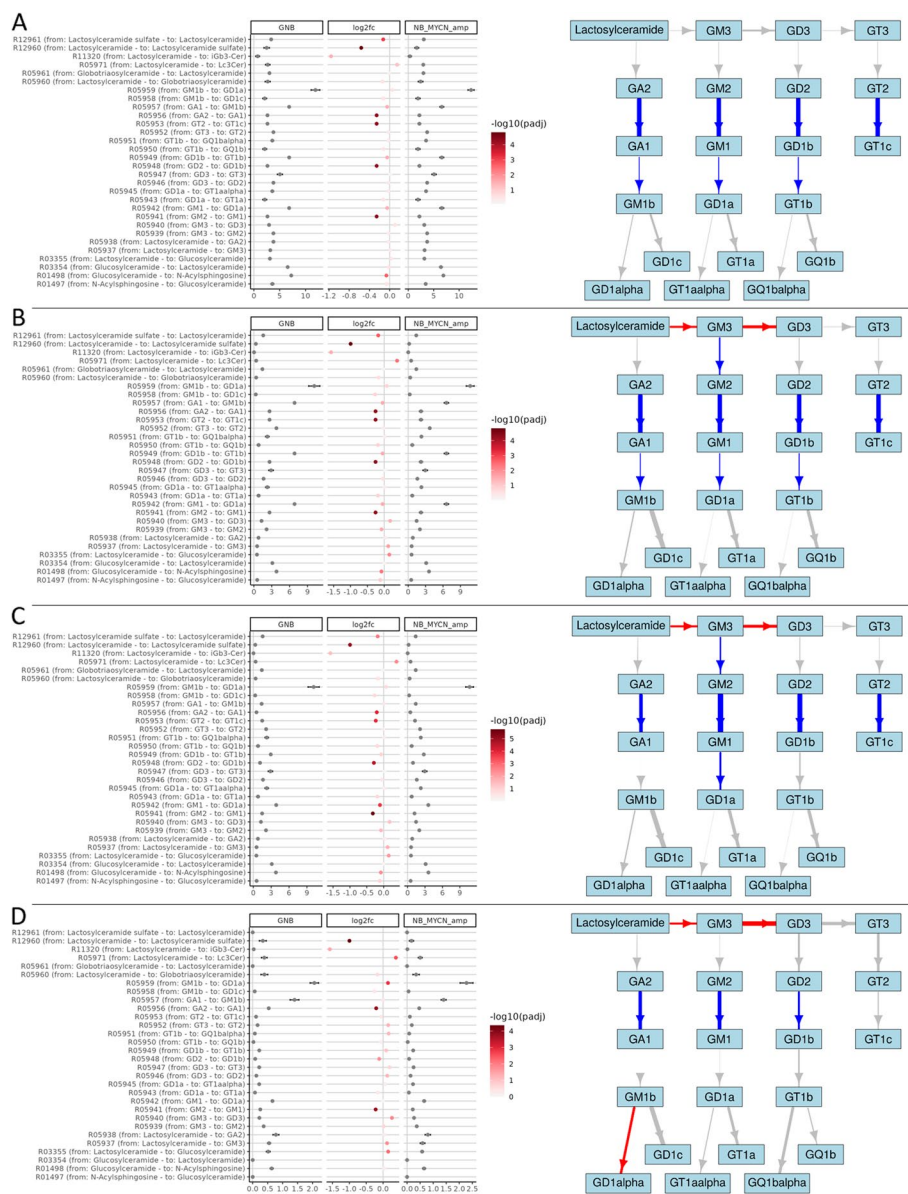


Fig. 4 Comparison of RAS-adjusted methods on the GSL pathway between NB +MYCN (“NB_MYCN_amp” in figure) and GNB, as well as the log2 fold-change between the two groups. The whiskers indicate the standard deviation. The color intensity of log2 fold-change points indicates the negative decimal logarithm of the adjusted p -values. On the right, the subgraph containing the GSL pathway is visualized. Nodes represent the gangliosides and edges are the metabolic reactions (arrows). The arrow color describes the log2 fold-change direction, where red means that the reaction is more active in NB +MYCN compared to GNB and blue means the reaction is stronger in GNB compared to NB +MYCN. Gray arrows represent reactions that are not significant (adjusted p -value > 0.05). The thickness of the arrow represents the relative log2 fold-change. The subplots show A) values from M_W (RAS values), B) M_{W1} (RAS values adjusted by TP), C) M_{W2} (RAS adjusted by the recursive TP), and D) M_{W3} (RAS adjusted by TP of paths)

pathway in NB and GNB or NB and GN, with a particular focus on distinguishing the individual series involved in ganglioside metabolism. In the following, the RAS values of the TARGET NB/GNB dataset were calculated based on the final metabolic graph and the respective gene expression values of the samples. The RAS matrix W

was adjusted with one of the three transition probability matrices described in the “Methods” section, resulting in the four matrices M_W , M_{W1} , M_{W2} , M_{W3} , each of which consists of reactions $\times$ samples, and where the value $M_{i,j}$ denotes the edge weight of reaction i in sample j . We analyzed three comparisons based on the TARGET GNB/NB

dataset, 1) NB +MYCN vs. GNB (Fig. 4), 2) NB +MYCN vs. NB -MYCN (Fig. 5), and 3) NB -MYCN vs. GNB (Fig. S2). A similar comparison was performed on the GN/NB dataset and can be found in Fig. S3. Log2 fold-changes and RAS values for all comparisons are provided in Additional file 4.

Figure 4 illustrates the first comparison (NB +MYCN vs. GNB). A, B, C, and D represent the four matrices. The figure is reduced to the reactions relevant to ganglioside biosynthesis. The absolute values of the respective comparison group and the log2 fold-change are shown on the left. The coloring of the log2 fold-change dots reflects the negative decimal-logarithmic p -value.

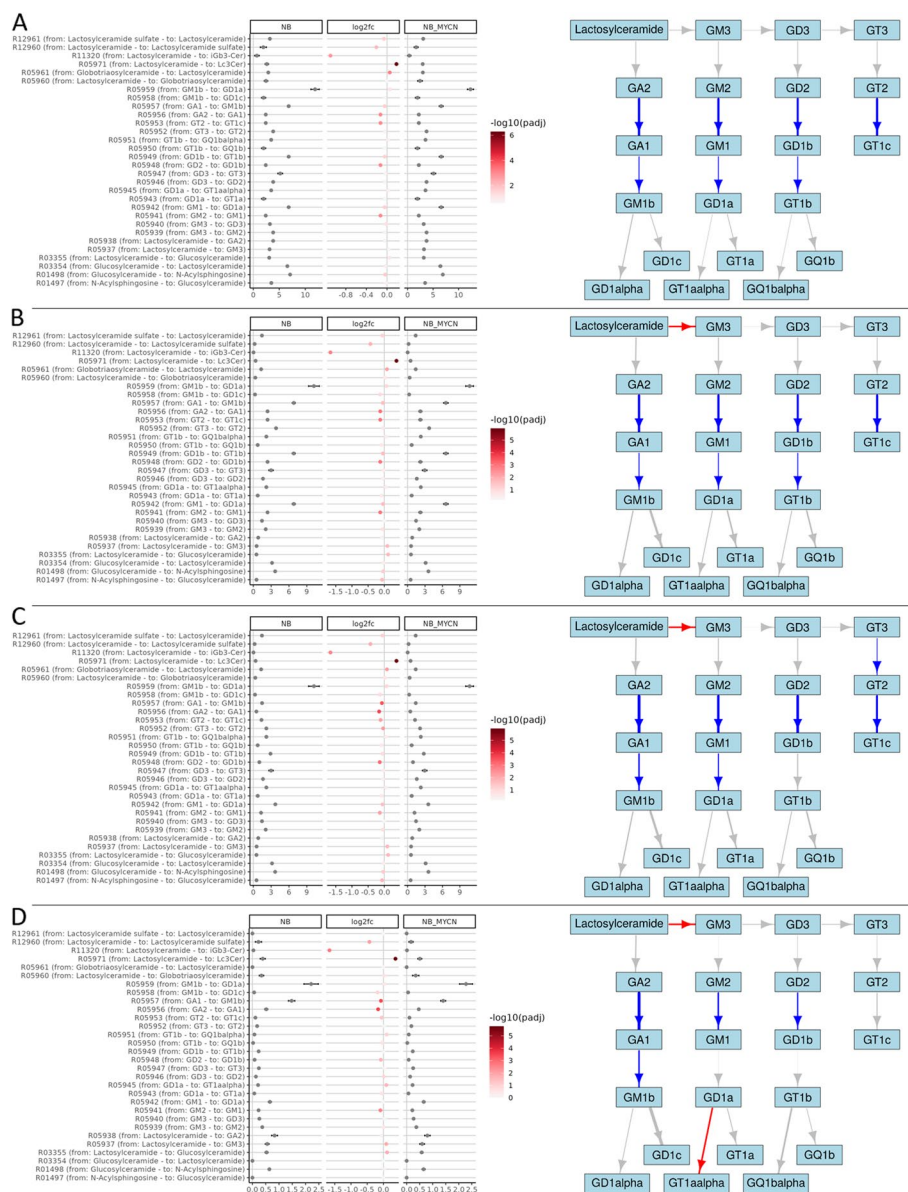


Fig. 5 Comparison of RAS-adjusted methods on the GSL pathway between NB -MYCN and NB +MYCN. Left dot plots illustrate the adjusted RAS values of NB -MYCN (NB) and NB +MYCN (NB_MYCN), as well as the log2 fold-change between the two groups. The whiskers indicate the standard deviation. The color intensity of log2 fold-change points indicates the negative decimal logarithm of the adjusted p -values. On the right, the subgraph containing the GSL pathway is visualized. Nodes represent the gangliosides and edges are the metabolic reactions (arrows). The arrow color describes the log2 fold-change direction, where red means that the reaction is more active in NB +MYCN compared to NB -MYCN and blue means the reaction is stronger in NB -MYCN compared to NB +MYCN. Gray arrows represent reactions that are not significant (adjusted p -value > 0.05). The thickness of the arrow represents the relative log2 fold-change. The subplots show A) values from M_W (RAS values), B) M_{W1} (RAS values adjusted by TP), C) M_{W2} (RAS adjusted by the recursive TP), and D) M_{W3} (RAS adjusted by TP of paths)

On the right side, the ganglioside pathway as part of the GSL metabolic graph is shown. Non-significant reactions (with an adjusted p -value > 0.05) are grayed. Reactions that are more active in NB +MYCN are red and those that are more active in GNB are blue. The width of the arrows indicates the relative log2 fold-change. As expected, in Fig. 4A the log2 fold-changes as well as the absolute values are the same for reactions with the same enzymes. One can only see the tendency that GNBs tend to form complex gangliosides. While this tendency changes and becomes more fine-grained when looking at the adjusted RAS values. Over the plots of Fig. 4 that show adjusted RAS (Fig. 4B, C, and D), the reaction R05940 (from GM3 to GD2, involving *ST8SIA1*) remains significant and indicates a higher activity in NB +MYCN. The reaction R05937 (from lactosylceramide to GM3, involving *ST3GAL5*) is also significant and higher in NB +MYCN but indicates a smaller log2 fold change and a larger p -value. Another notable reaction that is significantly higher in NB +MYCN and shows a big fold-change, is R05971 (from lactosylceramide to Lc3Cer, involving *B3GNT5*) suggesting a tendency of NB +MYCN to enter the lacto- and neolacto series. Also as expected, in Fig. 4B only the first reaction values of the four series are changed. Figure 4D indicates that the reaction from GM1b to GD1alpha of the 0-series is significantly more active in NB +MYCN.

Overall, the scores between NB +MYCN and GNB are not highly differentially expressed and relatively similar. Only few reactions differ more than 10% between the two entities (not shown in the figure), indicating fine-grained genomic changes that result in relatively small log2 fold-changes. In summary, depending on the adjusted method, different ganglioside series are predicted to be active in NB +MYCN and GNB. Particularly, the adjustment methods that discriminate all reactions between the four series (Fig. 4C and D) suggest the a-series to be more active in GNB, while the higher activity of the reactions R05937 & R05940 in NB +MYCN hints to an accumulation of gangliosides of the b-, or c- series.

Similar results can be observed in the comparison between NB and GN (Fig. S3). Although more reactions are significant due to the low sample number, there seems to be a similar tendency of GN to metabolize the simpler gangliosides to more complex ones, because the reactions R05941 & R05948 show a larger log2 fold-change compared to NB (e.g. Fig. S3B). Reaction R05940 is also similar to the above comparison of GNB and NB hinting towards a higher activity of the b-series. However, there are also differences, so the large fold-change values of reactions R05938 and R05947 can be interpreted as an increased activity of the 0- and c-series in NB.

In contrast to the previous comparison, the differential reaction activity between NB -MYCN and NB +MYCN is overall smaller, since it is the same tumor entity. Notably, the small differences are also reflected in the fact that the changes between the adjustments are more subtle. In all four subfigures of Fig. 5, there is a trend for the upstream reaction responsible for the more complex gangliosides to have slightly higher median activity in NB -MYCN compared to samples with MYCN amplification. This would indicate that MYCN amplified samples tend to have a higher GD2 concentration, as GD2 is further metabolized in -MYCN samples. The patterns in Fig. 5B, C, and D indicate that reaction R05971, in which lactosylceramide is metabolized to GM3, is significantly more active in MYCN-amplified NBs. The three reactions R05957, R05949, and R05942, which show particularly high absolute adjusted RAS values in Fig. 5A and B, are reduced by the two alternative adjustment methods and brought to a similar activity level of other reactions. Whereas the absolute RAS values of reaction R05959 remain constantly high across the four subfigures. Similarly to the previous comparison, the reaction R05971 indicates a large positive fold-change and is highly significant. Again, this could mean a stronger tendency of NB +MYCN to enter the lacto- and neolacto series. Overall, the scores between NB +MYCN and NB -MYCN are not highly differentially expressed and are very similar, resulting in a relatively small log2 fold-change range. These results suggest that simple gangliosides may accumulate in samples with a MYCN amplification.

The metabolic graph used in our analysis results from the integration of four KEGG pathways. We examined whether there is an NT type-specific difference in the activity of the individual pathways of the graph. The average activity of the RAS was calculated as the sum of all reactions involved in the respective pathway. As shown in Fig. S4, there are no major differences in the overall activity of each pathway between NB & GN in the respective dataset, and NB -MYCN, NB +MYCN, and GNB of the TARGET GNB/NB dataset. In both comparisons, the sphingolipid metabolism pathway in NB & NB +MYCN tends to be less active than in GN and GNB. This can be explained by strong activity differences of sulfate and sulfatide metabolic reactions.

Correlation analysis of GD2 relevant genes and MYCN

The previous results suggest that the expression of gangliosides in NT changes depending on the subtype (NB, GNB) and the MYCN amplification status. This may affect the expression of clinically relevant gangliosides such as GD2. Therefore, we further analyzed the expression of genes required for GD2 synthesis and MYCN in the GN/NB and TARGET NB/GNB datasets.

Our analysis reveals that the biological samples of the GN/NB dataset exhibited a clustering pattern that was highly dependent on the tumor entity, as determined by hierarchical clustering (Fig. 6A). In GN, *B3GALT4* exhibited notably high expression levels, while *ST8SIA1* displayed relatively lower expression levels when compared to NB supporting our previous finding suggesting the expression of the a-series and of complex gangliosides in more differentiated tumors (Fig. S3). Moreover, we observed a strong negative correlation between the gene

expression of *MYCN* and *B3GALT4* ($r = -0.92$) (Fig. 6B), as well as a strong positive correlation between *MYCN* and *ST8SIA1* ($r = 0.89$) (Fig. 6C). Interestingly, the gene *B4GALNT1*, which is associated with the ganglioside pathway and is involved in the biosynthesis of GD2, did not exhibit a high correlation with *MYCN* ($r = 0.26$) (Fig. 6D). GNs show the lowest expression of *MYCN* and the highest expression of *B3GALT4* suggesting an involvement of *MYCN* in the further metabolism of GD2 to GD1b, as expected in more differentiated tumors.

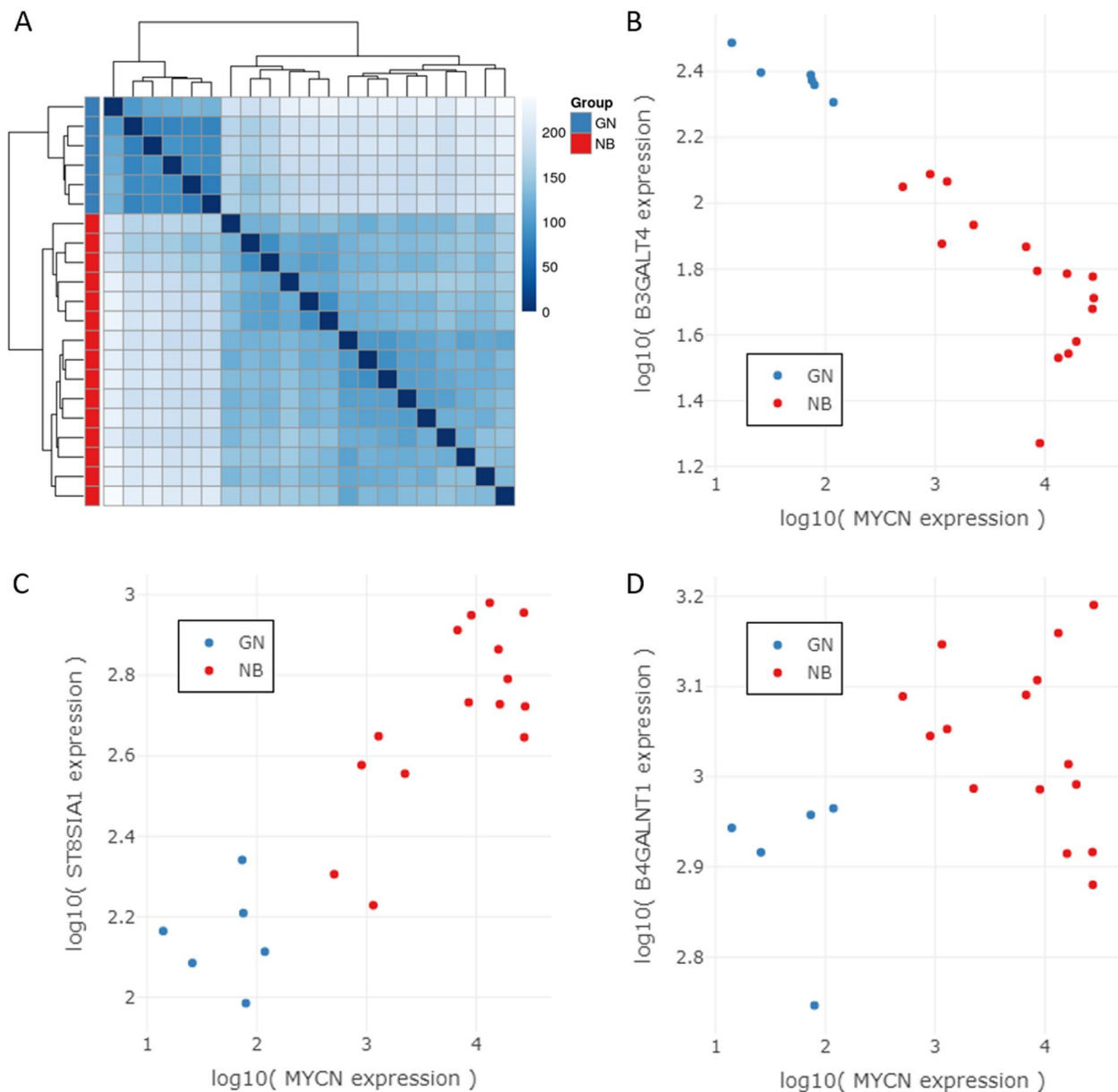


Fig. 6 **A** Heatmap illustrating the sample-to-sample distances of variance stabilized RNA-seq data (GN/NB dataset); GN ($n = 6$), and NB ($n = 15$); Intensity of the blue color indicates high and low similarity between samples. **B** Gene expression values of *MYCN* vs. *B3GALT4*. **C** Gene expression values of *MYCN* vs. *ST8SIA1*. **D** Gene expression values of *MYCN* vs. *B4GALNT1*. Red and blue dots indicate NB and GN samples, respectively

In contrast, the TARGET GNB/NB dataset does not show a clear clustering of samples based on the tumor entity or the MYCN amplification status in the overall distance matrix (Fig. 7A). Nevertheless, the MYCN status seems to cluster better than the tumor entity.

We assessed the Pearson correlation between *MYCN* expression and the three genes of the ganglioside metabolism related to GD2 biosynthesis within each of the NT groups (NB +MYCN, NB -MYCN, and GNB) and across different combinations of these groups to

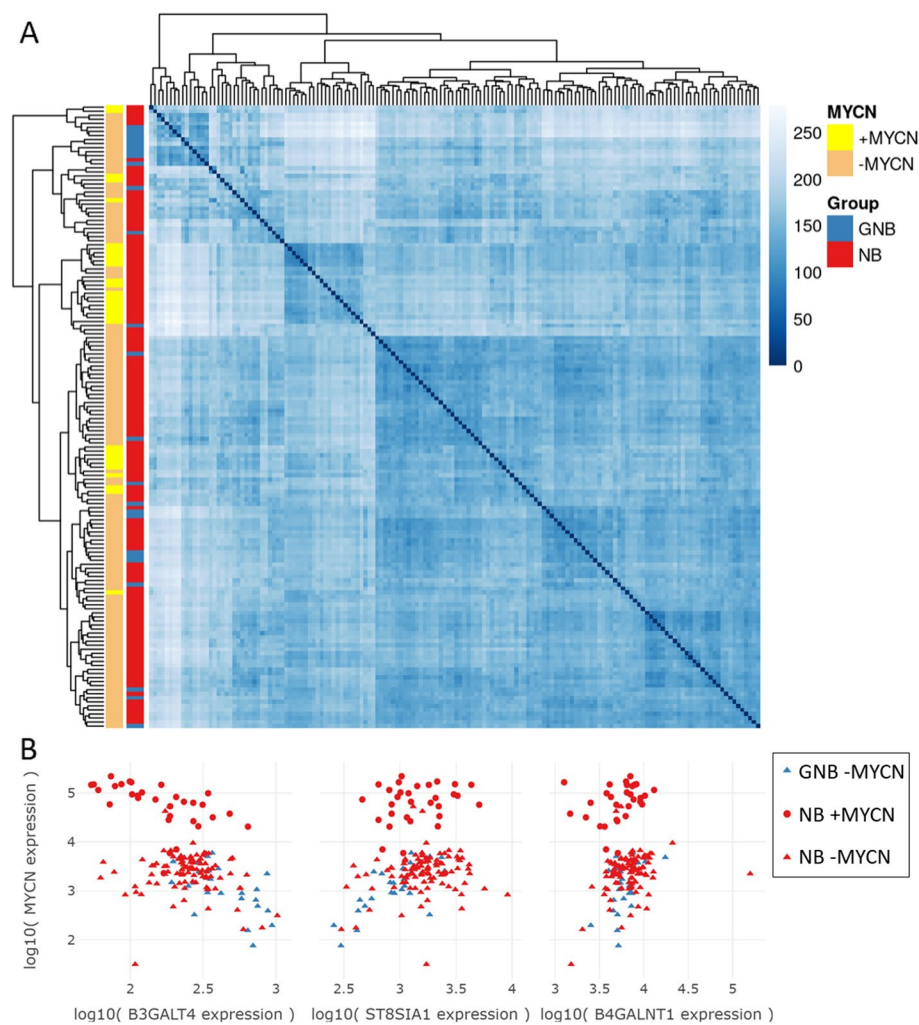


Fig. 7 **A** Heatmap illustrating the sample-to-sample distances of variance stabilized RNA-seq data (TARGET NB/GNB) ($n = 154$); blue annotation = GNB ($n = 25$), and red annotation = NB ($n = 129$); yellow annotation = +MYCN ($n = 31$), orange annotation = -MYCN ($n = 123$); Intensity of the blue color indicate high and low similarity between samples. **B** Gene expression values of *MYCN* vs. *B3GALT4*, *ST8SIA1*, and *B4GALNT1*. Red and blue data points indicate NB and GNB samples respectively. Round-shaped data points symbolize +MYCN samples and triangle-shaped ones show -MYCN samples

Table 1 Pearson correlation coefficient between expression values of GD2-related genes and MYCN for different sample subsets

Comparison	<i>B3GALT4</i> vs. <i>MYCN</i> [r]	<i>ST8SIA1</i> vs. <i>MYCN</i> [r]	<i>B4GALNT1</i> vs. <i>MYCN</i> [r]
only NB +MYCN	-0.52	0.09	0.26
only NB -MYCN	-0.14	0.18	0.3
only GNB	-0.59	0.73	0.51
NB +MYCN & GNB	-0.77	0.51	0.02
NB -MYCN & GNB	-0.33	0.38	0.34
all samples	-0.51	0.21	0.04

identify co-expression or co-regulation patterns, as well as to compare gene relationships between the subtypes (Table 1). The correlation coefficient r shows strong negative values between *MYCN* and *B3GALT4* within *MYCN* amplified NB samples, GNB, their combination, as well as across all samples. Between *MYCN* and *ST8SIA1* a strong positive correlation exists only within GNB and in combination with NB + *MYCN* & GNB. Although GNB samples have a strong positive correlation between *MYCN* and *B4GALNT1*, the other NT subtypes and combinations of subtypes do not indicate a co-expression pattern. The visualization of the three GD2-related genes vs. *MYCN* can be seen in Fig. 7B.

Taken together, these data support the role of *MYCN* in the differentiation of NT tumors [55] by regulating the expression of the *B3GALT4* gene, which is required for the synthesis of complex gangliosides.

Unsupervised data exploration of GN/NB dataset indicates distinct subgroups in NB based on adjusted RAS values

Using an unsupervised learning approach, our goal was to show that by adjusting the RAS values of the GSL metabolism and differentiating the ganglioside series, the identification of novel subgroups would potentially be possible. Therefore, we first examined the GN/NB dataset with UMAP and performed HDBSCAN clustering on the UMAP coordinates to determine visually appropriate parameters. To estimate the stability of the identified clusters in dependence of the adjustment method, we calculated UMAPs and HDBSCAN clustering on 1000 iterations and summed the x- and y-coordinates per sample. The results were visualized as line plots per adjustment method in Fig. 8.

In Fig. 8A, it can be seen that the NT types based on RAS form two clusters over all iterations. By adjusting the RAS values with the TP, these two NT type based clusters can be further observed (Fig. 8B). By applying the third adjustment method based on TP along defined paths, two further clusters become apparent within the NB type (Fig. 8C). In this case, the number of variables for the UMAP calculation are reduced because reactions outside of possible paths are removed. As the TP differs significantly in dependence of the path length, the TP matrix was scaled to make the scores homoscedastic for the UMAP. The adjustment method based on the recursive TP identifies most of the time three clusters but shows a weaker cluster stability, as more outliers are found (HDBSCAN annotates outlier samples as a pseudo-cluster 0) and more samples are switching between the clusters (Fig. 8D).

As additional diagnostic metrics for the clustering performance, the C-index, the Calinski-Harabasz index, and the silhouette score were computed across the iterations

(Fig. S9). The C-index in both datasets based on the adjusted/non-adjusted RAS values is in the low range, indicating coherent and well-separated clusters. Due to the high homogeneity of the NB/GNB dataset, the C-index is slightly higher compared to the GN/NB dataset. The clear separation in the GN/NB dataset is also evident in the high values of the silhouette score. Interestingly, the RAS adjusted by the recursive paths in the NB/GNB dataset leads to a clear formation of two widely separated clusters. The higher values of the Calinski-Harabasz index and the silhouette score support this observation. In summary, although the scatter between the iterations is higher in the more homogeneous NB/GNB dataset than in the NB/GN dataset, it can be said that the results of the iterations are consistent, except for a few outliers.

Using a UMAP result calculated on the path-based TP-adjusted RAS matrix as an example (Fig. 9A), we demonstrate the further procedure to identify characteristic reactions for the respective clusters applying t-tests. For each cluster, we have chosen one exemplary reaction with the highest p -value for visualization. In Fig. 9C the reaction R11321, which takes part in the globo series of the GSL biosynthesis pathway and is represented by the gene *B3GALNT1*, shows a significant p -value of 0.0001 and FDR of 0.01 for cluster 1. The reaction R06038, which participates in the lacto- and neolacto series of the GSL biosynthesis pathway and is represented by the fucosyltransferase (FUT) genes, shows a significant p -value of 0.0001 and FDR of 0.003 for cluster 2 (Fig. 9D). Lastly, reaction R12960 characterizes the third cluster (GN samples), which is responsible for the metabolism of lactosylceramide sulfate by galactosylceramide sulfotransferase with a p -value smaller 0.0001 and FDR < 0.0001 (Fig. 9E). In accord to these results, the presence of sulfatides is expected to be high in Schwannian stroma, which is the main component of GN [19].

In addition, we checked the correlation between the identified clusters and the *MYCN* expression. As shown in Fig. 9B, there is indeed a difference between the *MYCN* expression of NBs from clusters 1 and 2. As already shown in Fig. 6, one can see the clear difference in *MYCN* expression between NB (clusters 1 and 2) and GN (cluster 3).

In summary, these results suggest that unsupervised data exploration techniques can be applied to the adjusted weighting schemes and yield additional findings on subtypes and their specific characteristics. Our analysis suggests the presence of two NB subclusters with an altered sphingolipid composition. One of the subclusters is characterized by, among other factors, increased activity of the FUT enzymes and a correlation with the *MYCN* gene expression.

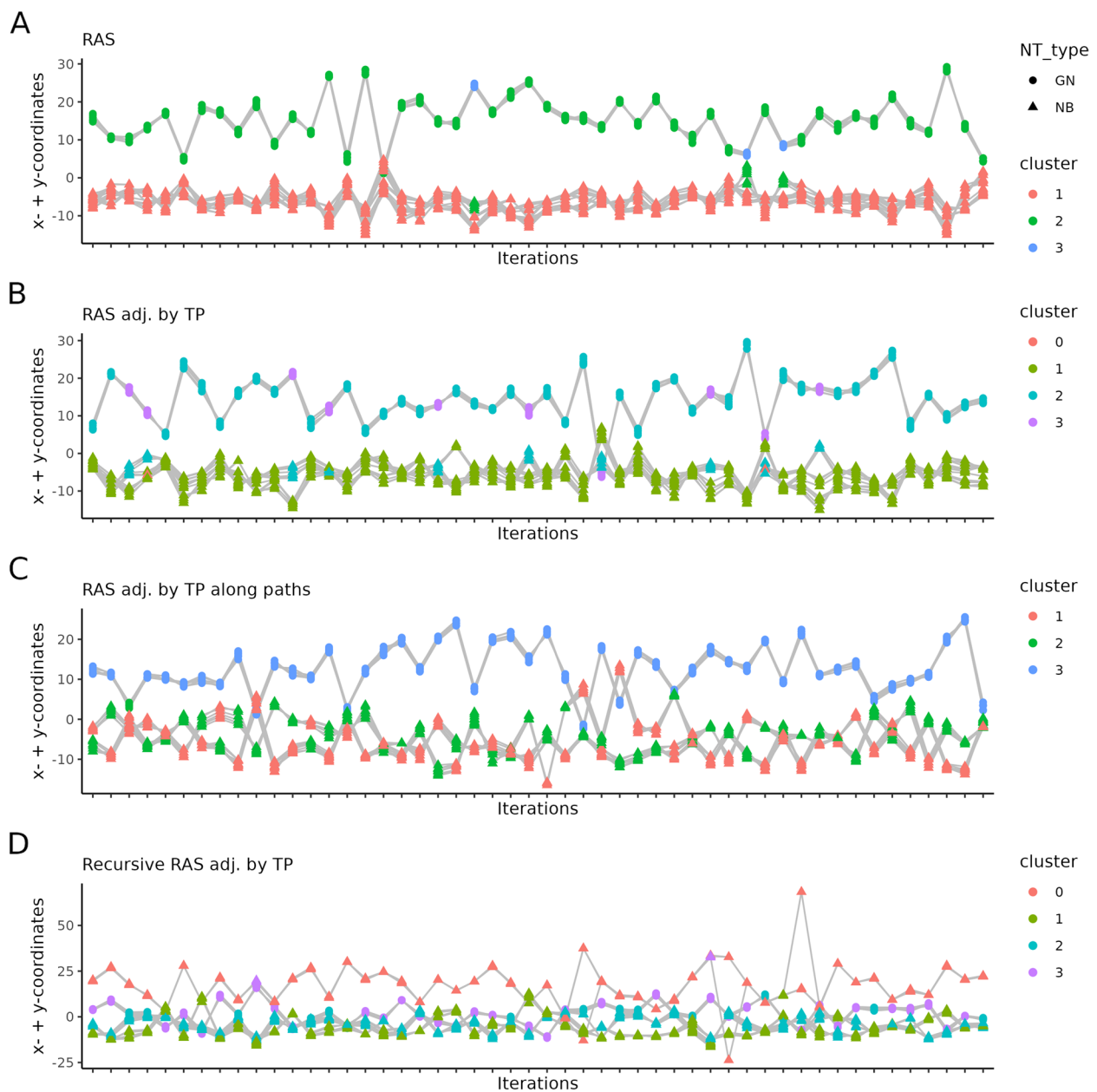


Fig. 8 Stability assessment of the unsupervised clustering of the adjusted RAS matrices. UMAP dimensionality reduction with HDBSCAN clustering was performed on 21 samples of the GN/NB dataset over 1000 iterations. UMAP's parameter $n_neighbor$ attribute was set to 3. The subplots show 50 random selected iterations of UMAP followed by HDBSCAN clustering (x-axis) for A) M_W (RAS values), B) M_{W1} (RAS values adjusted by TP), C) M_{W2} (RAS adjusted by the path-specific TP), and D) M_{W3} (RAS adjusted by the recursive TP). The color indicates the cluster of the respective sample. GN is represented through a circle, while NBs are rectangular. The same samples are connected with lines between the iterations, indicating cluster stability. The y-axis represents the sum of x- and y-coordinates of the UMAP per sample

Unsupervised learning of TARGET GNB/NB samples indicates a Fucosyltransferase driven subcluster

We have also applied the unsupervised procedure described above to the TARGET GNB/NB dataset. We computed the UMAP on the unadjusted RAS values for the following analysis (Fig. 10A). Cluster 1 shows a

high similarity of samples and was selected for further investigation.

The identification of the marker reactions showed that in cluster 1, among others, the reactions involving genes of fucosyltransferases (*FUT3*, *FUT4*, *FUT5*, *FUT6*, and *FUT9*), e.g. reaction R06025 of the lacto- and neolacto

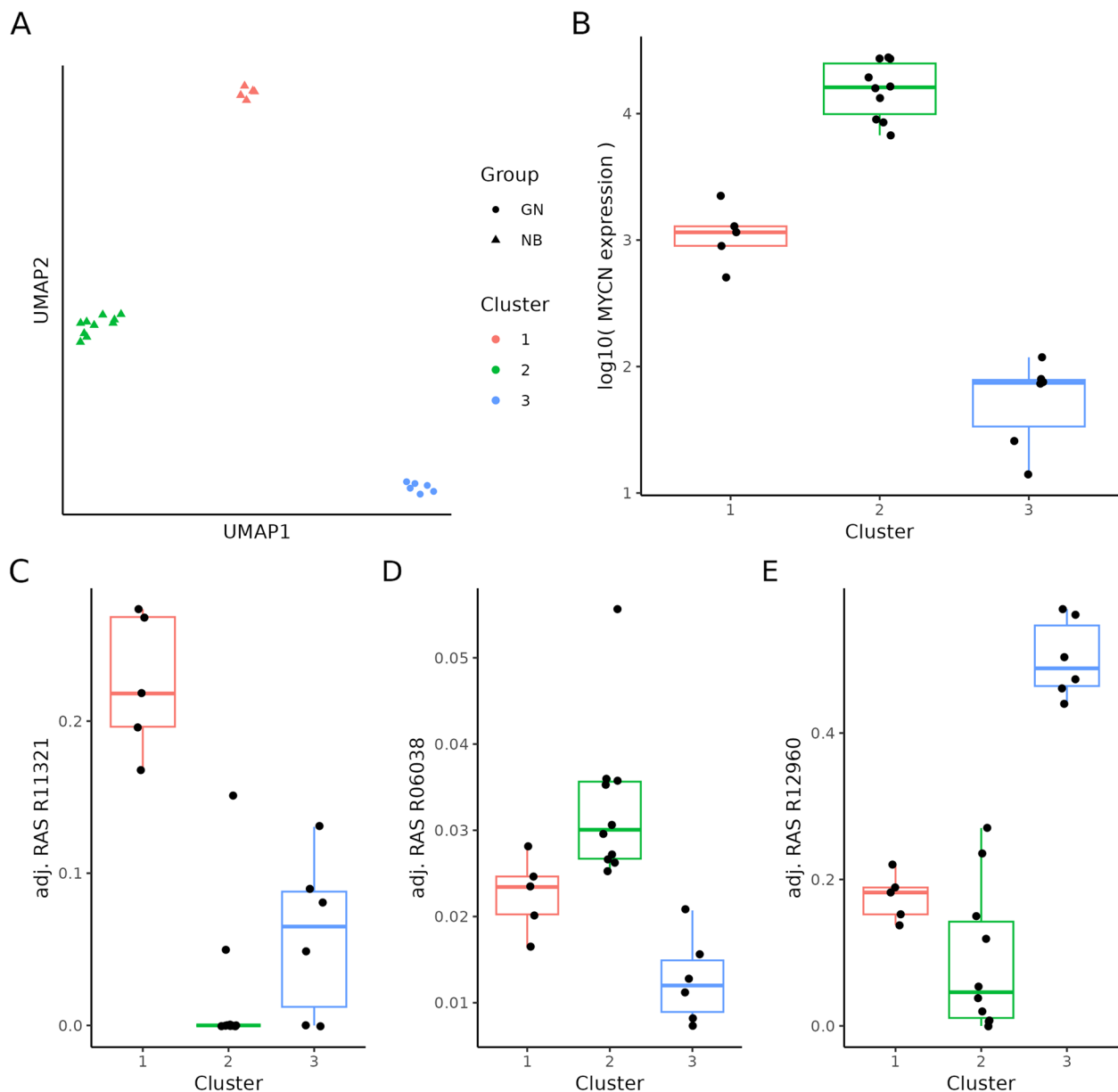


Fig. 9 Unsupervised clustering and identification of marker reactions. **A** The HDBSCAN clustering of the UMAP coordinates based on the M_{W_3} (RAS values adjusted by TP along paths) shows a clear separation between GN and NB, as well as a separation of NB samples into two distinct groups. Round symbols represent GN and triangles are NB samples. **B** Box plot of normalized MYCN expression grouped by identified clusters. With a t-test statistic we demonstrate characteristic reactions per cluster. As an example, the box plot in **C** shows the adjustment RAS values of reaction R11321 characteristically for cluster 1, **D** the reaction R06038 characteristically for cluster 2, and **E** the reaction R12960 characteristically for cluster 3. Boxes range from first to third quantile, the middle line indicates the median, the whiskers show the highest and lowest values no further than $1.5 \cdot IQR$ from the hinge. Black dots represent the samples

series, are significantly more active with $p < 0.0001$ and $FDR < 0.001$ than in the samples of the remaining clusters (Fig. 10B). Since the analysis was performed on unadjusted RAS values, all reactions involving all named FUTs have the same p -values and FDRs. To further investigate the transcriptional deregulation of

cluster 1, differential gene expression analysis was performed between the cluster 1 samples against all remaining samples (Fig. 10C). This identified 91 downregulated and 498 upregulated genes that had an adjusted p -value < 0.05 and a $\log_2$ fold-change < -0.6 or > 0.6 , which corresponds to a fold-change of 1.5 in gene expression

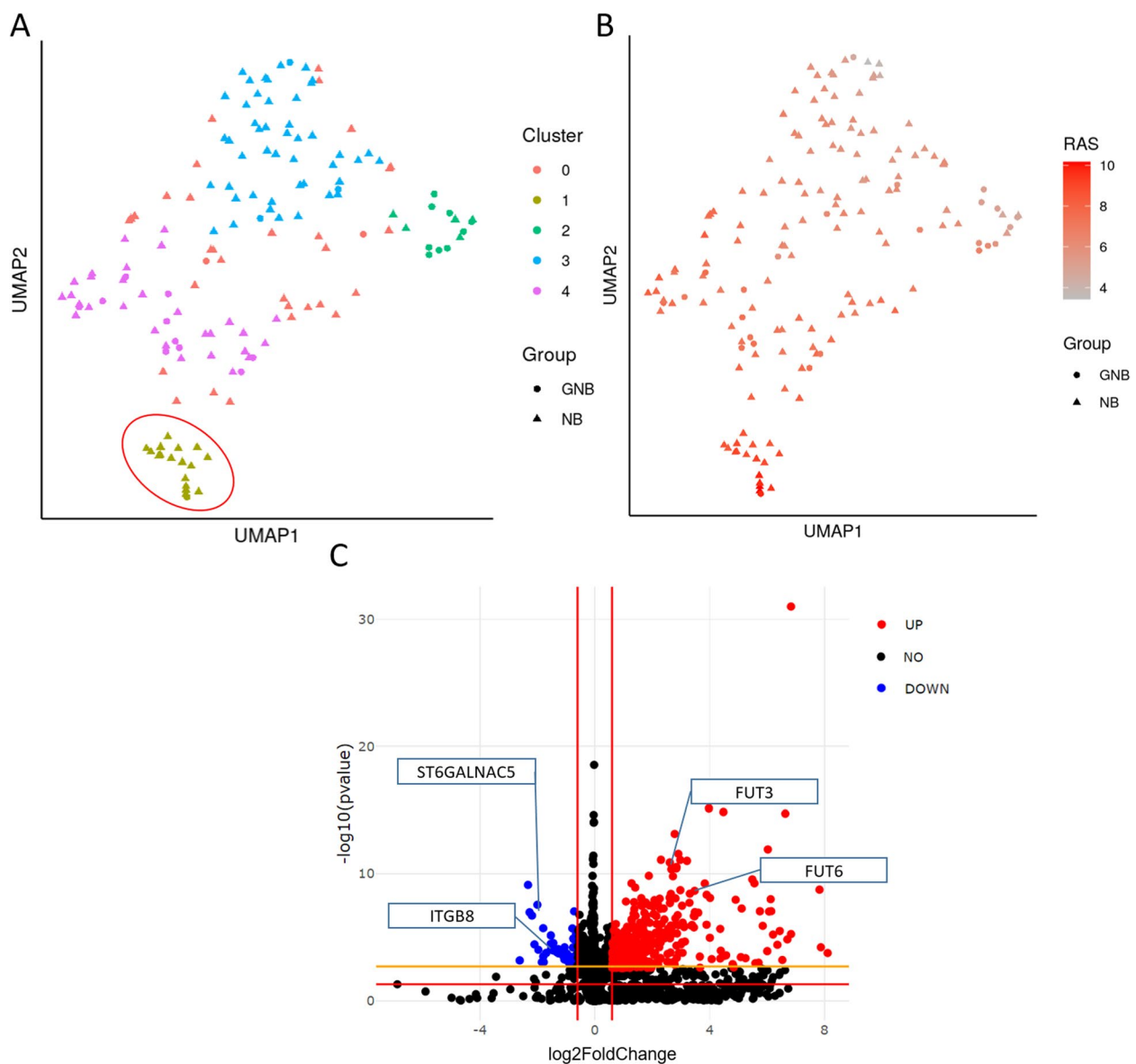
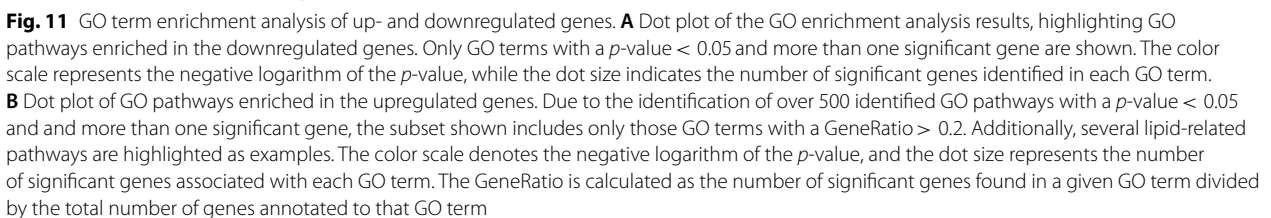


Fig. 10 Unsupervised clustering and identification of marker reactions. **A** One representative UMAP representation out of the 1000 computed iterations, colored by the HDBSCAN clustering (in accord with the HDBSCAN method samples of cluster 0 are identified as outliers). Circles represent GNB and triangles are NB samples. The red selection indicates samples of cluster 1. **B** The same UMAP representation as in **B**) colored by the computed values for the R06025 reaction, which was identified as highly characteristically for cluster 1. **C** Volcano plot of differentially expressed genes between samples of cluster 1 and all other samples. Red dots denote upregulated genes in samples of cluster 1 compared to samples of all other clusters. Blue dots show downregulated genes in cluster 1. Not significant expressed genes are black. The red lines indicate the p -value threshold of 0.05, and log2 fold change thresholds of 0.6 and -0.6. The orange line indicates the adjusted p -value threshold of 0.05. The genes *FUT3/6*, *ITGB8*, and *ST6GALNAC5* were additionally labeled

(Additional file 3). Additionally, a heatmap of the down-regulated genes can be found in Fig. S6. Among the upregulated genes, the *FUT3* and *FUT6* genes are found, as expected. Among the downregulated genes, genes of the GSL pathway are also present, namely *ITGB8* and *ST6GALNAC5*. In Fig. S5 we further investigated the association of *MYCN* expression across the UMAP

representation. The specific cluster under investigation comprises only a single sample exhibiting a *MYCN* expression exceeding 4. We could not detect any discernible correlation in *MYCN* expression patterns within the UMAP. This observation suggests that the impact of *MYCN* may not be sufficiently robust to manifest discernibly in the GSL profile.

deacetylation, myelin assembly, lower regulation of synaptic transmission, and glial cell development (Fig. 11A). Furthermore, the GO terms generation of neurons, neuron differentiation, brain development, positive regulation of lipid metabolic process, and lipid homeostasis are significantly enriched, yet the gene ratio between the differential genes and the term annotated to the GO term



is lower (Additional file 3). In Fig. S7 we further investigated the enrichment map of the GO enriched terms from downregulated genes in cluster 1 and identified six communities. Two major communities indicate an association with 1) neuron and nervous system related terms (green), and 2) lipid-related and regulatory terms. Among GO terms enriched within upregulated genes, we detected the following pathways standing out in particular with a high gene ratio (> 0.2) and a particularly low p -value: histidine metabolic process, regulation of lipoprotein lipase activity, negative regulation of cholesterol transport, reverse cholesterol transport, epoxigenase P450 pathway, intestinal lipid absorption, and triglyceride homeostasis (Fig. 11B).

Taken together, we investigated the GSL metabolism patterns with a particular focus on the ganglioside metabolism pathway in NB, GNB, and GN using the introduced adjusted RAS values. Our analysis revealed subtle but significant differences in gene expression-driven reaction activities. Notably, comparisons between MYCN-amplified NB and GNB suggested fine-grained transcriptomic changes, indicating distinct ganglioside series activities. Correlation analyses underscored the influence of tumor entity and MYCN amplification on ganglioside expression and suggested an involvement of MYCN in the accumulation of GD2. Unsupervised techniques, such as UMAP and HDBSCAN clustering, showed an association between the clusters with the tumor type in case of the NB/GN dataset and revealed distinct subgroups within NB in both datasets. In particular, a FUT-driven subcluster within NB showed potential associations with GO terms related to neuron and nervous system, lipids, and regulatory mechanisms. The comparison between the RAS adjustment methods led to slightly different results when investigating the reaction activities of the ganglioside series and different UMAP embeddings during the unsupervised analysis, making it difficult to assess which method is more reliable. Nevertheless, the approach enabled us to better quantitatively discriminate the ganglioside series and in this context showed improvements compared to unchanged RAS values.

Discussion

In this study, we focused on the specific issue of low substrate specificity among ganglioside metabolism enzymes and the resulting difficulty to obtain conclusions about the activity within the four series of the ganglioside metabolism when utilizing only transcriptome data. To overcome this challenge, we propose a straightforward approach that integrates network topology information and transcriptome data. We constructed a directed graph based on four GSL related metabolic pathways from the KEGG database. Transcriptome datasets of three NT

entities were aggregated to RAS values for each sample and used to weighting the graph edges. Three different adjustment methods based on the TP between nodes were applied to the RAS weighing and compared among each other. This integrated methodology provides a solution for unraveling the complexities of ganglioside pathways.

Even though the study of the effect of complex metabolic processes plays an important role in many fields, such as signal transduction, biomarkers, and cancer physiology, it is still difficult to infer metabolic activity from indirect measurements, such as transcriptome data [56]. Many of the methods developed for this purpose can be divided into at least two fields, constraint-based methods (e.g., flux-based analysis) [56, 57] and graph analysis, also known as network science techniques [58–60]. In the latter case, graph construction can take different forms, such as metabolite-centric, bipartite, or reaction-centric graphs [61]. Both fields mostly use genome-scale metabolic models (GEMs), which are mathematical representations of the current state of knowledge on organ-, species-, or condition-specific metabolic properties. Constraint-based approaches, while being powerful tools, they require reaction rates, assume a steady-state condition and one or more suitable optimization objectives, which are often elusive in cancer types [32, 62, 63]. Therefore, we have intentionally refrained from using whole GEM and constraint-based methods in order to reduce the complexity of the analysis and focus only on the GSL metabolism. Acknowledging the multifaceted nature of GSL metabolism regulation, encompassing epigenetic, transcriptional, and post-translational control [64–67], we intentionally limit our model to gene expression data, as this simplification enables a focused exploration within the defined scope of our study and allows the usage of a widely available data type.

We acknowledge that a limitation of our analysis method is the involvement of many considered enzymes in reactions outside of GSL metabolism. This could obscure significant differences between the various subpopulations in our dataset. Currently, the state of the art in finding subgroups are simple RNA-Seq workflows, often in combination with other molecular data, that do not account for single reactions at all. Our primary aim was to explore the GSL metabolism and identify metabolic differences between several groups, which is not efficiently possible with the aforementioned workflows. In this context, we focused on neuroblastic tumor entities due to the existing literature on the deregulation of GSL metabolism. Given that the primary enzymes involved in ganglioside synthesis are specific to their substrates [45], we opted to simplify our approach by avoiding the use of a GEM. Regarding the technical application of the three

RAS weighting methods on a GEM, we conclude that the simple TP adjustment can be applied to a whole-genome network, as it depends solely on local network topology and is unaffected by enzyme involvement in different pathways. Notably, our results from the GSL-subnetwork would remain consistent even if applied to the entire network. The recursive weighting method is also applicable to the whole metabolic graph and would not alter the results. In contrast, the path-based adjustment method relies heavily on the central node from which simple paths are calculated. Applying it to the whole network is challenging due to the absence of a central metabolic node. A potential solution could be to select a central node per pathway and prevent overlapping paths, but this approach would require further testing.

It can be observed that both adjustment methods (path-based and recursive weighting) have introduced differences in the weighting of the ganglioside series compared to the unadjusted RAS weighting. Even the simple addition of the TP with the RAS values does not provide complete differentiation. The interpretation of the results in relation to the unsupervised analysis is difficult, as the different weightings lead to different clusters. In the case of recursive adjustment, even the association of the clusters with the tumor entities dissolves and shows more outlier samples. However, this may also be due to the clustering parameters. In the comparison between two specific conditions, the differences introduced between the ganglioside series are of a subtle nature. In any case, we understand that further work is needed to evaluate the adjustment methods on other tissue and tumor entities, that may indicate stronger differences in the ganglioside patterns and to validate the results in the context of laboratory measurements, e.g. immunohistochemistry or mass spectrometry. In this first approach, we wanted to focus on well-studied neuroblastic tumor entities that have clinical relevance in terms of ganglioside metabolism.

Using this approach, we were able to predict an activation of the a-series in GNB vs. NB. This is in accord with a previous analysis showing that GNB expresses GM1a and GD1a [15]. NB samples were shown to express complex gangliosides of the b-series (GD1b and GT1b), while the presence of both the a- and b-series was observed particularly in GNB. This could be regulated by the *ST8SIA1* gene product (R05940) which is indeed more expressed in NB according to our result and underscores the hypothesis of N. W. Mabe et al. that *ST8SIA1* acts as a bottleneck on the GD2-production [26]. A lower expression of *ST8SIA1* could be a signal to the cells to enter the a-series. The co-expression of complex gangliosides of the a- and b-series is typical of mature neuronal tissues, which express very low amount of GD2, and the

significantly higher expression of *B3GALT4* and lower expression of *ST8SIA1* in GN support our hypothesis that the combination of different genes is required to predict the final composition of NT. Importantly, reduced expression of *ST8SIA1* could change the composition of gangliosides towards the a-series with reduction in GD2 expression and indeed downregulation of *ST8SIA1* in NBL confers resistance to the anti-GD2 antibody [26, 68].

Our data suggest also that MYCN amplified tumors, known for their unfavorable prognosis, accumulate less complex gangliosides, driven by the activities of reactions R05956, R05941, R05948, and R05953 (all attributed to the gene *B3GALT4*). This aligns with previous data showing the absence of complex gangliosides in NB cases featuring MYCN amplification [15]. Notably, the simple ganglioside GD2 is particularly abundant in these tumors [12]. Loss of complex gangliosides has been linked to a poor prognosis in NB [11, 69]. As MYCN amplification is also associated with the differentiation of NT tumors, more complex gangliosides can be observed in GN and GNB, which are more differentiated tumor entities. This highlights the need for further investigations to unravel the relationship between MYCN amplification, regulation of differentiation, the depletion of complex gangliosides, and the concurrent accumulation of GD2.

GSL profiles are important not only in NT. GSL deregulation has been described in different tumor entities, including hard to treat tumors such as diffuse midline glioma, ependymoma [70], and triple negative breast cancer [71]. Deregulation of GSL affects the biology of the tumors. For example, expression of GD1a facilitates the adhesion of tumor cells to the endothelium and high GD2 expression has been described in metastasis of different tumor types [72, 73]. GSL can be exploited for therapy and classification. Tumors with deregulated gangliosides synthesis can be responsive to treatment with UGCG inhibitors, such as miglustat and eliglustat, which are already used in the clinic for the treatment of children with Gaucher disease [15, 70, 74]. Tumors with expression of GD2 can be targeted with monoclonal antibodies [75] or CAR-T cells [76]. Finally, GSL profiles allow identifying subgroups in heterogeneous tumor entities such as medulloblastoma [15]. Therefore, the approach we propose may facilitate in the future the identification of therapeutic or diagnostic relevant GSL profiles across tumor entities.

Current prognostic markers are based on genomic alterations, particularly MYCN amplification and ALK alterations in NB. Ganglioside profiles have not been integrated into a clinical routine so far, also because of the technical requirements necessary for their quantification. The use of RNA-seq data to predict ganglioside profiles may help in the future to improve the prognosis

and possibly the therapy of NT patients. Current biomarkers are indeed not sufficient to identify all patients at risk. For example, MYCN amplification is associated with a poor prognosis but children with tumors without MYCN amplification can also have a poor prognosis. GN is generally considered a benign tumor and therefore can be treated by surgery alone. However, ganglioneuroblastoma intermixed can have aggressive behavior and can need multimodal therapy. Thus, ganglioside profiling with our method may help in the future for better risk stratification of NT tumors and could help to identify subgroups with specific biological features. For instance, we were able to define a subcluster involving genes of fucosyltransferases. FUT genes are required for the synthesis of Sialyl Lewis x (SSEA1; CD15) which belongs to the neolacto series. SSEA1 is expressed in neuronal stem cells and could define a subgroup with higher potential for proliferation and self-renewal as already shown in other tumor entities [77]. Further, Cuello et al. identified an overexpression of FUT genes in MYCN amplified cell lines and patient tumors with aberrant glycosylation and aggressive behaviour [78].

Conclusions

In this paper, we described a methodological approach introducing three TP based adjustments for a RAS weighted metabolic graph with nodes as metabolites and edges representing metabolic reactions. We illustrated how these weighting schemes can be applied to a conditional comparison and an unsupervised data exploration. Furthermore, we tested our approach on three NT entities with the focus on the GSL metabolism and particularly the ganglioside biosynthesis pathway.

Our results suggest that indeed the weighting schemes make the individual series of ganglioside metabolism distinguishable, allowing a differentiated analysis of the GSL profile in NT entities. We showed that the results have a biological rationale and are largely consistent with published data on the ganglioside composition of NT. Our method may be helpful in the comparison of GSL deregulation between two specific sample groups, as well as for the identification of previously unknown cancer subgroups that exhibit distinguishability based on their sphingolipid profile. Our approach can also be used for other metabolic pathways involving enzymes that have low specificity to the substrate and participate in several metabolic reactions unlocking the potential to leverage widely available transcriptome datasets to obtain deeper insights in such clinically relevant scenarios.

Abbreviations

BH	Benjamini-Hochberg
DGE	Differential gene expression
FDR	False Discovery Rate

FUT	Fucosyltransferase
GEM	Genome scale metabolic model
GlcCer	Glucosylceramide
GN	Ganglioneuroma
GNB	Ganglioneuroblastoma
GO	Gene Ontology
GPR	Gene-protein-reaction
GSL	Glycosphingolipids
KEGG	Kyoto Encyclopedia of Genes and Genomes
KS	Kolmogorov-Smirnov
LacCer	Lactosylceramide
NB	Neuroblastoma
NT	Neuroblastic tumor
RAS	Reaction activity score
RNA-seq	RNA sequencing
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TP	Transition probability
UMAP	Uniform Manifold Approximation and Projection
+MYCN	MYCN-amplified samples
-MYCN	Samples without MYCN amplification

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40170-024-00358-y>.

Additional file 1

Additional file 2. Annotation of reactions: 1) Table with edge reaction ID, incoming edges, outgoing edges, involved genes as gene symbol and ensemble; 2) Table of considered genes and their associated KEGG pathways

Additional file 3. DGE & GO enrichment results

Additional file 4. Table containing the log2 fold-changes, p-values, adjusted p-values, and absolute values for the following comparisons across all reactions of the whole graph

Acknowledgements

This work was supported by the computing infrastructure provided by the Core Facility Bioinformatics at the IMBEL.

Authors' contributions

All authors read and approved the final manuscript. AU: Visualization, Writing - Original Draft, Data Curation, Conceptualization, Methodology, Software, Formal analysis, Investigation. ASN: Visualization, Writing - Review & Editing. RS: Writing - Review & Editing, Investigation. JF: Writing - Review & Editing. FM: Writing - Review & Editing, Methodology. CP: Supervision, Writing - Original Draft, Conceptualization, Investigation.

Funding

Open Access funding enabled and organized by Projekt DEAL. The work of FM was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektnummer 318346496 - SFB1292/2 TP19N.

Data availability

All analyses included in this manuscript are documented in the github repository https://github.com/arsenij-ust/NT_GSL_analysis.

The GN/NB dataset analyzed during the current study are available in the Gene Expression Omnibus repository (<https://www.ncbi.nlm.nih.gov/geo/>) under the GEO accession GSM4437037.

The TARGET GNB/NB dataset analyzed during the current study are available in the UCSC Xena database (<http://xena.ucsc.edu/>) under the combined cohort of TCGA, TARGET and GTEx (<https://xenabrowser.net/datapages/?cohort=TCGA%20TARGET%20GTEx>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz 55131, Germany. ²Lipid Pathobiochemistry, German Cancer Research Center, Heidelberg 69120, Germany. ³Department of Pediatric Hematology/Oncology, Center for Pediatric and Adolescent Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz 55131, Germany. ⁴University Cancer Center (UCT), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz 55131, Germany. ⁵Research Center for Immunotherapy (FZI), Mainz 55131, Germany.

Received: 15 January 2024 Accepted: 11 October 2024

Published online: 24 October 2024

References

- Schnaar RL. The biology of gangliosides. *Adv Carbohydr Chem Biochem*. 2019;76:113–48. <https://doi.org/10.1016/bs.accb.2018.09.002>.
- Sipione S, Monyor J, Galleguillos D, Steinberg N, Kadam V. Gangliosides in the Brain: Physiology, Pathophysiology and Therapeutic Applications. *Front Neurosci*. 2020;14:572965. <https://doi.org/10.3389/fnins.2020.572965>.
- Gault CR, Obeid LM, Hannun YA. An overview of sphingolipid metabolism: from synthesis to breakdown. *Adv Exp Med Biol*. 2010;688:1–23. https://doi.org/10.1007/978-1-4419-6741-1_1.
- Schnaar RL, Sandhoff R, Tiemeyer M, Kinoshita T. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. *Glycosphingolipids*. 4th ed. Cold Spring Harbor (NY); 2022. pp. 129–40. <https://doi.org/10.1101/glycobiology.4e.11>.
- Sandhoff R, Sandhoff K. Emerging concepts of ganglioside metabolism. *FEBS Lett*. 2018;592(23):3835–64. <https://doi.org/10.1002/1873-3468.13114>.
- Vasques JF, de Jesus Goncalves RG, da Silva-Junior AJ, Martins RS, Gubert F, Mendez-Otero R. Gangliosides in nervous system development, regeneration, and pathologies. *Neural Regen Res*. 2023;18(1):81–6. <https://doi.org/10.4103/1673-5374.343890>.
- Yu RK, Nakatani Y, Yanagisawa M. The role of glycosphingolipid metabolism in the developing brain. *J Lipid Res*. 2009;50 Suppl(Suppl):S440–5. <https://doi.org/10.1194/jlr.R800028-JLR200>.
- Svennerholm L, Bostrom K, Fredman P, Mansson JE, Rosengren B, Rynmark BM. Human brain gangliosides: developmental changes from early fetal stage to advanced age. *Biochim Biophys Acta*. 1989;1005(2):109–17. [https://doi.org/10.1016/0005-2760\(89\)90175-6](https://doi.org/10.1016/0005-2760(89)90175-6).
- Vajn K, Viljetic B, Degmecic IV, Schnaar RL, Heffer M. Differential distribution of major brain gangliosides in the adult mouse central nervous system. *PLoS ONE*. 2013;8(9):e75720. <https://doi.org/10.1371/journal.pone.0075720>.
- Jin X, Yang GY. Pathophysiological roles and applications of glycosphingolipids in the diagnosis and treatment of cancer diseases. *Prog Lipid Res*. 2023;91:101241. <https://doi.org/10.1016/j.plipres.2023.101241>.
- Cao S, Hu X, Ren S, Wang Y, Shao Y, Wu K, et al. The biological role and immunotherapy of gangliosides and GD3 synthase in cancers. *Front Cell Dev Biol*. 2023;11:1076862. <https://doi.org/10.3389/fcell.2023.1076862>.
- Balis FM, Busch CM, Desai AV, Hibbitts E, Naranjo A, Bagatell R, et al. The ganglioside G(D2) as a circulating tumor biomarker for neuroblastoma. *Pediatr Blood Cancer*. 2020;67(1):e28031. <https://doi.org/10.1002/pbc.28031>.
- Terzic T, Cordeau M, Herblot S, Teira P, Cournoyer S, Beaunoyer M, et al. Expression of Disialoganglioside (GD2) in Neuroblastic Tumors: A Prognostic Value for Patients Treated With Anti-GD2 Immunotherapy. *Pediatr Dev Pathol*. 2018;21(4):355–62. <https://doi.org/10.1177/1093526617723972>.
- Lee MC, Kim BW, Kim JS, Kim KS, Lee JH, et al. Neuronal differentiation of human neuroblastoma SH-SY5Y cells by gangliosides. *Brain Tumor Pathol*. 1997;14(1):5–11. <https://doi.org/10.1007/BF02478862>.
- Paret C, Ustjanzew A, Ersali S, Seidmann L, Jennemann R, Ziegler N, et al. GD2 Expression in Medulloblastoma and Neuroblastoma for Personalized Immunotherapy: A Matter of Subtype. *Cancers (Basel)*. 2022;14(24). <https://doi.org/10.3390/cancers14246051>.
- Shawraba F, Hammoud H, Mrad Y, Saker Z, Fares Y, Harati H, et al. Biomarkers in Neuroblastoma: An Insight into Their Potential Diagnostic and Prognostic Utilities. *Curr Treat Options Oncol*. 2021;22(11):102. <https://doi.org/10.1007/s11864-021-00898-1>.
- Rodriguez EF, Jones R, Miller D, Rodriguez FJ. Neurogenic Tumors of the Mediastinum. *Semin Diagn Pathol*. 2020;37(4):179–86. <https://doi.org/10.1053/j.semdp.2020.04.004>.
- Shimada H, Ikegaki N. Genetic and Histopathological Heterogeneity of Neuroblastoma and Precision Therapeutic Approaches for Extremely Unfavorable Histology Subgroups. *Biomolecules*. 2022;12(1). <https://doi.org/10.3390/biom12010079>.
- Shimada H, Ambros IM, Dehner LP, Hata J, Joshi VV, Roald B. Terminology and morphologic criteria of neuroblastic tumors: recommendations by the International Neuroblastoma Pathology Committee. *Cancer*. 1999;86(2):349–63.
- Wu ZL, Schwartz E, Seeger R, Ladisch S. Expression of GD2 ganglioside by untreated primary human neuroblastomas. *Cancer Res*. 1986;46(1):440–3.
- Schengrund CL. Gangliosides and Neuroblastomas. *Int J Mol Sci*. 2020;21(15). <https://doi.org/10.3390/ijms21155313>.
- Mastrangelo S, Rivetti S, Triarico S, Romano A, Attina G, Maurizi P, et al. Mechanisms, Characteristics, and Treatment of Neuropathic Pain and Peripheral Neuropathy Associated with Dinutuximab in Neuroblastoma Patients. *Int J Mol Sci*. 2021;22(23). <https://doi.org/10.3390/ijms222312648>.
- Slatnick LR, Jimeno A, Gore L, Macy ME. Naxitamab: a humanized anti-glycolipid disialoganglioside (anti-GD2) monoclonal antibody for treatment of neuroblastoma. *Drugs Today (Barc)*. 2021;57(11):677–88. <https://doi.org/10.1358/dot.2021.57.11.3343691>.
- Mount CW, Majzner RG, Sundaresh S, Arnold EP, Kadapakkam M, Haile S, et al. Potent antitumor efficacy of anti-GD2 CAR T cells in H3–K27M(+) diffuse midline gliomas. *Nat Med*. 2018;24(5):572–9. <https://doi.org/10.1038/s41591-018-0006-x>.
- Yanagisawa M, Yoshimura S, Yu RK. Expression of GD2 and GD3 gangliosides in human embryonic neural stem cells. *ASN Neuro*. 2011;3(2). <https://doi.org/10.1042/AN20110006>.
- Mabe NW, Huang M, Dalton GN, Alexe G, Schaefer DA, Geraghty AC, et al. Transition to a mesenchymal state in neuroblastoma confers resistance to anti-GD2 antibody via reduced expression of ST8SIA1. *Nat Cancer*. 2022;3(8):976–93. <https://doi.org/10.1038/s43018-022-00405-x>.
- Ruan S, Raj BK, Lloyd KO. Relationship of glycosyltransferases and mRNA levels to ganglioside expression in neuroblastoma and melanoma cells. *J Neurochem*. 1999;72(2):514–21. <https://doi.org/10.1046/j.1471-4159.1999.0720514.x>.
- Rieke DT, de Bortoli T, Horak P, Lamping M, Benary M, Jelas I, et al. Feasibility and outcome of reproducible clinical interpretation of high-dimensional molecular data: a comparison of two molecular tumor boards. *BMC Med*. 2022;20(1):367. <https://doi.org/10.1186/s12916-022-02560-5>.
- Sha Y, Han L, Sun B, Zhao Q. Identification of a Glycosyltransferase Signature for Predicting Prognosis and Immune Microenvironment in Neuroblastoma. *Front Cell Dev Biol*. 2021;9:769580. <https://doi.org/10.3389/fcell.2021.769580>.
- Sorokin M, Kholodenko I, Kalinovskiy D, Shamanskaya T, Doronin I, Konovalov D, et al. RNA Sequencing-Based Identification of Ganglioside GD2-Positive Cancer Phenotype. *Biomedicines*. 2020;8(6). <https://doi.org/10.3390/biomedicines8060142>.
- Yang J, Han L, Sha Y, Jin Y, Li Z, Gong B, et al. A novel ganglioside-related risk signature can reveal the distinct immune landscape of neuroblastoma and predict the immunotherapeutic response. *Front Immunol*. 2022;13:1061814. <https://doi.org/10.3389/fimmu.2022.1061814>.
- Graudenzi A, Maspero D, Di Filippo M, Gnugnoli M, Isella C, Mauri G, et al. Integration of transcriptomic data and metabolic networks in cancer

- samples reveals highly significant prognostic power. *J Biomed Inform.* 2018;87:37–49. <https://doi.org/10.1016/j.jbi.2018.09.010>.
33. Galuzzi BG, Vanoni M, Damiani C. Combining denoising of RNA-seq data and flux balance analysis for cluster analysis of single cells. *BMC Bioinformatics.* 2022;23(Suppl 6):445. <https://doi.org/10.1186/s12859-022-04967-6>.
 34. Weglarz-Tomczak E, Rijlaarsdam DJ, Tomczak JM, Brul S. GEM-Based Metabolic Profiling for Human Bone Osteosarcoma under Different Glucose and Glutamine Availability. *Int J Mol Sci.* 2021;22(3). <https://doi.org/10.3390/ijms22031470>.
 35. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org>.
 36. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2006;Complex Systems:1695. <https://igraph.org>.
 37. Weiss T, Taschner-Mandl S, Janker L, Billek A, Rifatbegovic F, Kromp F, et al. Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal growth factor-like protein 8. *Nat Commun.* 2021;12(1):1624. <https://doi.org/10.1038/s41467-021-21859-0>.
 38. Zhang X, Jonassen I. RASflow: an RNA-Seq analysis workflow with Snakemake. *BMC Bioinformatics.* 2020;21(1):110. <https://doi.org/10.1186/s12859-020-3433-x>.
 39. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
 40. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
 41. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. *Nucleic Acids Res.* 2021;49(D1):D916–23. <https://doi.org/10.1093/nar/gkaa1087>.
 42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
 43. Mohamed A, Hancock T, Nguyen CH, Mamitsuka H. NetPathMiner: R/Bioconductor package for network path mining through gene expression. *Bioinformatics.* 2014;30(21):3139–41. <https://doi.org/10.1093/bioinformatics/btu501>.
 44. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
 45. Furukawa K, Tokuda N, Okuda T, Tajima O, Furukawa K. Glycosphingolipids in engineered mice: insights into function. *Semin Cell Dev Biol.* 2004;15(4):389–96. <https://doi.org/10.1016/j.semcdb.2004.03.006>.
 46. Robinson JL, Kocabas P, Wang H, Cholley PE, Cook D, Nilsson A, et al. An atlas of human metabolism. *Sci Signal.* 2020;13(624). <https://doi.org/10.1126/scisignal.aaz1482>.
 47. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2018. <https://www.ncbi.nlm.nih.gov/pubmed/30531897>. <https://doi.org/10.1038/nbt.4314>.
 48. Hahsler M, Piekenbrock M, Doran D. dbSCAN: Fast Density-Based Clustering with R. *J Stat Softw.* 2019;91:1–30. <https://doi.org/10.18637/jss.v091.i01>.
 49. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* 2016;5:2122. <https://doi.org/10.12688/f1000research.9501.2>.
 50. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics.* 2019;35(12):2084–92. <https://doi.org/10.1093/bioinformatics/bty895>.
 51. Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for Gene Ontology. R package version 2.28.0. Cranio. 2016.
 52. Marini F, Binder H. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics.* 2019;20(1):331. <https://doi.org/10.1186/s12859-019-2879-1>.
 53. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics.* 2006;22(13):1600–7. <https://doi.org/10.1093/bioinformatics/btl140>.
 54. Marini F, Ludt A, Linke J, Strauch K. GeneTonic: an R/Bioconductor package for streamlining the interpretation of RNA-seq data. *BMC Bioinformatics.* 2021;22(1):610. <https://doi.org/10.1186/s12859-021-04461-5>.
 55. Kang JH, Rychahou PG, Ishola TA, Qiao J, Evers BM, Chung DH. MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. *Biochem Biophys Res Commun.* 2006;351(1):192–7. <https://doi.org/10.1016/j.bbrc.2006.10.020>.
 56. Ng RH, Lee JW, Baloni P, Diener C, Heath JR, Su Y. Constraint-Based Reconstruction and Analyses of Metabolic Models: Open-Source Python Tools and Applications to Cancer. *Front Oncol.* 2022;12:914594. <https://doi.org/10.3389/fonc.2022.914594>.
 57. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol.* 2012;10(4):291–305. <https://doi.org/10.1038/nrmicr.2012.737>.
 58. Manipuri I, Granata I, Maddalena L, Guarracino MR. Clustering analysis of tumor metabolic networks. *BMC Bioinformatics.* 2020;21(Suppl 10):349. <https://doi.org/10.1186/s12859-020-03564-9>.
 59. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics.* 2003;19(11):1423–30. <https://doi.org/10.1093/bioinformatics/btg177>.
 60. Holme P, Huss M, Jeong H. Subnetwork hierarchies of biochemical pathways. *Bioinformatics.* 2003;19(4):532–8. <https://doi.org/10.1093/bioinformatics/btg033>.
 61. Dusad V, Thiel D, Barahona M, Keun HC, Oyarzun DA. Opportunities at the Interface of Network Science and Metabolic Modeling. *Front Bioeng Biotechnol.* 2020;8:591049. <https://doi.org/10.3389/fbioe.2020.591049>.
 62. Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol.* 2010;13(3):344–9. <https://doi.org/10.1016/j.mib.2010.03.003>.
 63. Zhang Y, Boley D. Nonlinear multi-objective flux balance analysis of the Warburg Effect. *J Theor Biol.* 2022;550:111223. <https://doi.org/10.1016/j.jtbi.2022.111223>.
 64. Memon RA, Holleran WM, Uchida Y, Moser AH, Ichikawa S, Hirabayashi Y, et al. Regulation of glycosphingolipid metabolism in liver during the acute phase response. *J Biol Chem.* 1999;274(28):19707–13. <https://doi.org/10.1074/jbc.274.28.19707>.
 65. Russo D, Capolupo L, Loomba JS, Sticco L, D'Angelo G. Glycosphingolipid metabolism in cell fate specification. *J Cell Sci.* 2018;131(24). <https://doi.org/10.1242/jcs.219204>.
 66. Hardardottir I, Grunfeld C, Feingold KR. Effects of endotoxin and cytokines on lipid metabolism. *Curr Opin Lipidol.* 1994;5(3):207–15. <https://doi.org/10.1097/00041433-199405030-00008>.
 67. Metallo CM, Vander Heiden MG. Understanding metabolic regulation and its influence on cell physiology. *Mol Cell.* 2013;49(3):388–98. <https://doi.org/10.1016/j.molcel.2013.01.018>.
 68. Ruan S, Lloyd KO. Glycosylation pathways in the biosynthesis of gangliosides in melanoma and neuroblastoma cells: relative glycosyltransferase levels determine ganglioside patterns. *Cancer Res.* 1992;52(20):5725–31.
 69. Berois N, Osinaga E. Glycobiology of neuroblastoma: impact on tumor behavior, prognosis, and therapeutic strategies. *Front Oncol.* 2014;4:114. <https://doi.org/10.3389/fonc.2014.00114>.
 70. El Malki K, Wehling P, Alt F, Sandhoff R, Zahnreich S, Ustjanzew A, et al. Glucosylceramide Synthase Inhibitors Induce Ceramide Accumulation and Sensitize H3K27 Mutant Diffuse Midline Glioma to Irradiation. *Int J Mol Sci.* 2023;24(12). <https://doi.org/10.3390/ijms24129905>.
 71. Ly S, Anand V, El-Dana F, Nguyen K, Cai Y, Cai S, et al. Anti-GD2 antibody dinutuximab inhibits triple-negative breast tumor growth by targeting GD2(+) breast cancer stem-like cells. *J Immunother Cancer.* 2021;9(3). <https://doi.org/10.1136/jitc-2020-001197>.
 72. Taki T, Ishikawa D, Ogura M, Nakajima M, Handa S. Ganglioside GD1alpha functions in the adhesion of metastatic tumor cells to endothelial cells of the target tissue. *Cancer Res.* 1997;57(10):1882–8.
 73. Hatano K, Miyamoto Y, Nonomura N, Kaneda Y. Expression of gangliosides, GD1a, and sialyl paragloboside is regulated by NF-kappaB-dependent transcriptional control of alpha2,3-sialyltransferase I, II, and VI in human castration-resistant prostate cancer cells. *Int J Cancer.* 2011;129(8):1838–47. <https://doi.org/10.1002/ijc.25860>.
 74. Jennemann R, Federico G, Mathow D, Rabionet M, Rampoldi F, Popovic ZV, et al. Inhibition of hepatocellular carcinoma growth by blockade of

- glycosphingolipid synthesis. *Oncotarget*. 2017;8(65):109201–109216. <https://doi.org/10.18632/oncotarget.22648>.
75. Wingerter A, El Malki K, Sandhoff R, Seidmann L, Wagner DC, Lehmann N, et al. Exploiting Gangliosides for the Therapy of Ewing's Sarcoma and H3K27M-Mutant Diffuse Midline Glioma. *Cancers (Basel)*. 2021;13(3). <https://doi.org/10.3390/cancers13030520>.
76. Miguel Llordes G, Medina Perez VM, Curto Simon B, Castells-Yus I, Vazquez Sufuentes S, Schuhmacher AJ. Epidemiology, diagnostic strategies, and therapeutic advances in diffuse midline glioma. *J Clin Med*. 2023;12(16). <https://doi.org/10.3390/jcm12165261>.
77. Son MJ, Woolard K, Nam DH, Lee J, Fine HA. SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. *Cell Stem Cell*. 2009;4(5):440–52. <https://doi.org/10.1016/j.stem.2009.03.003>.
78. Cuello HA, Segatori VI, Alberto M, Gulino CA, Aschero R, Camarero S, et al. Aberrant O-glycosylation modulates aggressiveness in neuroblastoma. *Oncotarget*. 2018;9(75):34176–34188. <https://doi.org/10.18632/oncotarget.26169>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.