



A novel tool for multi-omics network integration and visualization: A study of glioma heterogeneity

Roberta Coletti ^{a,*,}, João F. Carrilho ^{a,b,}, Eduarda P. Martins ^{c,d,}, Céline S. Gonçalves ^{c,d,}, Bruno M. Costa ^{c,d,}, Marta B. Lopes ^{a,b,e,}

^a Center for Mathematics and Applications (NOVA Math), NOVA School of Science and Technology, 2829-516, Caparica, Portugal

^b NOVA School of Science and Technology, NOVA University of Lisbon, 2829-516, Caparica, Portugal

^c Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus Gualtar, 4710-057, Braga, Portugal

^d ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal

^e UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, 2829-516, Caparica, Portugal

ARTICLE INFO

Dataset link: <http://firebrowse.org/>, <https://github.com/sysbiomed/MONET>, <https://github.com/RobertaColetti/MINGLE>

Keywords:

MINGLE
Network
Multi-omics
Glioma
GBM
Graph
Visualization
Variable selection
Biomarkers

ABSTRACT

Gliomas are highly heterogeneous tumors with generally poor prognoses. Leveraging multi-omics data and network analysis holds great promise in uncovering crucial signatures and molecular relationships that elucidate glioma heterogeneity. However, the complexity of the problem and the high dimensionality of the data increase the challenges of integrating information across various biological levels.

This study develops a comprehensive framework aimed at identifying potential glioma-type-specific biomarkers through innovative variable selection and integrated network visualization. We designed a two-step framework for variable selection using sparse network estimation across various omics datasets. This framework incorporates MINGLE (Multi-omics Integrated Network for Graphical Exploration), a novel methodology designed to merge distinct multi-omics information into a single network, enabling the identification of underlying relations through an innovative integrated visualization. The analysis was conducted using glioma omics datasets, with patients grouped based on the latest glioma classification guidelines.

Our investigation of the glioma data led to the identification of variables potentially serving as glioma-type-specific biomarkers. The integration of multi-omics data into a single network through MINGLE facilitated the discovery of molecular relationships that reflect glioma heterogeneity, supporting the biological interpretation. Scripts and files for reproducing the analysis or adapting it to other applications, are available in R software.

1. Introduction

Gliomas are tumors originating from glial or stem cells in the brain, which exhibit a remarkable level of heterogeneity, both at the molecular and histo-pathological levels [1]. This inherent complexity has presented significant challenges in glioma classification and understanding, ultimately impacting treatment decisions and patient outcomes [2]. The advent of modern techniques for data acquisition has allowed the generation of huge amounts of data describing different molecular entities within a biological system, *i.e.*, omics data [3]. The analysis of this type of data has substantially increased the opportunities to disclose important molecular insights for the different glioma types, contributing to improve patient stratification [4,5]. As a result, the World Health Organization (WHO) recently updated the standards for the classification of Tumors of the Central Nervous System (CNS), releasing the last edition in 2021 (2021-WHO) [6], ultimately

distinguishing between three main adult-type diffuse gliomas: astrocytoma, oligodendroglioma, and glioblastoma (GBM). These advances have also contributed to the development of modern therapies (such as immunotherapy), which currently complement classical approaches (chemotherapy, surgery, and radiation) [7]. Despite this, glioma patients often show poor overall survival [8], indicating that further efforts are needed to improve the life expectancy and quality of glioma patients.

In this context, mathematical and statistical methods may be extremely useful, being specifically developed to solve various biological problems [9–12]. In particular, graphical models might deal with the estimation of networks of molecular features at different biological levels, allowing the disclosure of potentially important relations between multiple molecular entities and clinical outcomes. Nevertheless, traditional approaches fail when dealing with high-dimensional data

* Corresponding author.

E-mail address: roberta.coletti.91@gmail.com (R. Coletti).

<https://doi.org/10.1016/j.combiomed.2025.109811>

Received 19 March 2024; Received in revised form 29 January 2025; Accepted 4 February 2025

Available online 17 February 2025

0010-4825/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

when the number of variables (p) vastly outnumbers the number of samples (n), as occurs with omics data [13]. Dimensionality reduction techniques can address this issue by reducing the number of features while maintaining overall information. In recent years, several methods have been proposed to infer sparse networks from biological data, with a particular focus on gene regulatory networks based on transcriptomics data [14]. However, the incorporation of multiple molecular levels through multi-omics integration leads to a more comprehensive perspective of the complex molecular landscape regulating cell behavior [15]. The multi-omics approach enables researchers and clinicians to identify subtle patterns and associations unfeasible to discern by considering a single omics view [16–18]. On the other hand, this integration might be challenging owing to the high dimensionality and heterogeneity of these data [19]. Here, we propose a novel methodology specifically designed to explore glioma heterogeneity based on multi-omics data. After a network estimation step, a key component of the methodology includes the design of a *Multi-omics Integrated Network for Graphical Exploration (MINGLE)*, a novel graphical method integrating prior information to enhance the interpretability of multi-omics networks. Despite the designed methodology being applicable to different omics, here we consider RNA-sequencing data (*transcriptomics*) and DNA-methylation data (*methyloomics*) of patients from The Cancer Genome Atlas (TCGA) glioma projects (namely, TCGA-LGG and TCGA-GBM), previously updated according to the 2021-WHO diagnostic annotations [20]. The selection of these two omics was motivated by the close relationship existing between them. Indeed, whereas RNA sequencing drives all cellular processes, methylation of small regions of DNA might affect transcript reading [21], significantly impacting cancer regulation [22]. In particular, in the context of glioma, methylation might explain the pathogenesis and clinical behavior of the different cancer types [23,24], therefore being promising information to include in a multi-omics study aimed at understanding glioma heterogeneity. The designed pipeline consists of two sequential steps of network-based variable selection, which are achieved by using graphical models. Specifically, we applied the Graphical Lasso [25] and Joint Graphical Lasso [26] to estimate sparse networks from each omics layer, by focusing on the analysis of the connected nodes. The estimated undirected relations were then used to merge the information obtained from the two omics into a single integrated graph using MINGLE. Various methodologies have been proposed in the literature for analyzing multilayer networks [27–29]. These approaches primarily aim to create a unified network that reveals the overall influence of different factors within a complex system. Typically, each factor corresponds to a distinct layer with interactions represented by networks that share the same nodes. While this approach is effective for integrating the effects of multiple layers on the same variables, it does not align with the objectives of this study, which focuses on networks from omics referring to different variables. The novelty of MINGLE lies in defining a new network, where edges represent the underlying mechanisms that refer to variables distinct from the network's nodes. Given the established influence of DNA methylation on gene expression, the MINGLE networks defined in this study have genes as nodes, whereas the edges are computed from methyloomics networks. This simple but innovative graphical representation enables visualization of the genes' relations based on DNA-methylation, advancing towards multi-omics interpretation.

The analysis here reported is focused on GBM, as it represents the most aggressive and frequent malignant glioma with very poor survival. Disclosing the underlying molecular relationships potentially associated with GBM evolution is extremely important for identifying novel therapeutic targets, designing innovative and more effective therapies, and ultimately improving patient survival and quality of life. Nevertheless, equivalent analyses can be easily performed on different glioma types using the provided R codes and tables, which allow the independent exploration of our results. To validate the provided

outcomes, mathematical validation of the inferred networks was performed. Classification was used to formally evaluate the ability of the selected features to distinguish among the glioma types. Importantly, survival analysis of a subset of variables was also applied to assess the prognostic value of these features.

2. Materials and methods

2.1. Network inference and variable selection

Let $\mathbb{G}$ be a graph with p nodes $\mathbf{X} = (X_1, \dots, X_p)$ having a multivariate normal distribution, and with theoretical and empirical covariance matrix Σ and S , respectively. The edges of the graph $\mathbb{G}$ can be estimated based on the non-zero entrances of $\Theta = \Sigma^{-1}$, also known as precision matrix [30]. In particular, each entrance $\theta_{i,j}$ describes the conditional dependence between the variables X_i and X_j , given the others X_k , $k = 1, \dots, p$, $k \neq i, j$.

Given a dataset with n observation of $X_1, \dots, X_p$ variables, if $p \gg n$ estimating the fully-connected precision matrix is unfeasible, and methods leading to sparse networks should be employed. The graphical lasso (*glasso*) [25] finds sparse Σ and Θ matrices, by solving the following Gaussian log-likelihood maximization problem:

$$\max_{\Theta} \{ \log(\det \Theta) - \text{tr}(S\Theta) - \rho \|\Theta\|_1 \}. \quad (1)$$

The presence of the regularization term $\rho \|\Theta\|_1$ makes the solution sparse, which degree is regulated by the parameter ρ .

Let $\mathbb{G}_1, \dots, \mathbb{G}_k$ be k graphs with the same nodes $\mathbf{X} = (X_1, \dots, X_p)$, and different sets of edges, represented by the precision matrices $\Theta = (\Theta^{(1)}, \dots, \Theta^{(k)})$. The Joint Graphical Lasso (JGL) methods [26] allows the joint estimation of k precision matrices, by a generalization the function (1):

$$\max_{\Theta} \left\{ \sum_{l=1}^k n_l (\log(\det \Theta^{(l)}) - \text{tr}(S^{(l)}\Theta^{(l)}) - P(\Theta)) \right\}, \quad (2)$$

where n_l and $S^{(l)}$, for $l = 1, \dots, k$ are the number of observations of the l -th dataset, and the corresponding empirical covariance matrices. The penalization term $P(\Theta)$ affects the sparsity of the estimated Θ ; in this work, we employed the Fused Graphical Lasso penalty:

$$P(\Theta) = \lambda_1 \sum_{l=1}^k \sum_{\substack{i,j=1 \\ i \neq j}}^p |\theta_{i,j}^{(l)}| + \lambda_2 \sum_{\substack{l < l' \\ i,j=1 \\ i \neq j}}^k |\theta_{i,j}^{(l)} - \theta_{i,j}^{(l')}|, \quad (3)$$

where the parameter λ_1 regulates the amount of sparsity (as ρ in *glasso* Eq. (1)), while λ_2 encourages similarity across edges.

The sparsity of the precision matrix determines a network-based variable selection. Indeed, $\theta_{i,j} = 0$, for a certain i and j ($i \neq j$), means that the variables X_i and X_j are conditionally independent given the others X_k , $k = 1, \dots, p$, $k \neq i, j$. In highly-sparse networks, a certain number of variables will appear as isolated nodes, implying that the corresponding features could be removed, as they do not impact on the overall network.

2.2. Sparse multinomial logistic regression with the elastic net penalty

Let $\mathbf{X} = (X_1, \dots, X_p)$ be a set of p variables among which n observations are collected. Let we assume that each observation is attributed with a categorical variable C , with L possible levels. Zhu and Hastie [31] proposed a symmetrical approach for the estimation of the probabilities of a given sample being attributed to level $\ell = 1, \dots, L$:

$$\Pr(C = \ell | \mathbf{x}) = \frac{e^{\beta_{0\ell} + \beta_{\ell}^T \mathbf{x}}}{\sum_{k=1}^L e^{\beta_{0k} + \beta_k^T \mathbf{x}}}, \quad \ell = 1, \dots, L, \quad (4)$$

where $\beta_\ell = (\beta_{1\ell}, \dots, \beta_{p\ell})$. The coefficients $\beta_{0\ell}$ and β_ℓ are then estimated by minimizing the following function:

$$l(\beta_{01}, \dots, \beta_{0L}, \beta_1, \dots, \beta_L) = -\frac{1}{n} \sum_{i=1}^n \log \Pr(C = c_i | x_i) + \lambda \sum_{k=1}^L F_\alpha(\beta_k), \quad (5)$$

where n represents the number of observations, x_i the i th observation, c_i the class which x_i is attributed to and λ the strength of the penalty. F_α represents the elastic net penalty function, defined as

$$F_\alpha(\beta) = \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right), \quad \beta = (\beta_1, \dots, \beta_p), \quad (6)$$

with α controlling the balance between the Lasso and Ridge penalties and λ controlling the strength of the penalty. The regularization term induces sparsity in the estimated coefficients β_ℓ , allowing variable selection.

2.3. Dataset preprocessing

The glioma datasets are publicly available from the TCGA portal, under the projects TCGA-LGG and TCGA-GBM [32–34]. To update the glioma types according to the latest 2021 WHO classification of brain tumors [6], we followed the methodology described in our previous work [20], integrating the molecular profiles of the samples [35]. We considered only patients included in both omics datasets (transcriptomics and methylomics), and having survival time information, to perform survival analysis on a subset of selected features. After this preprocessing, the samples were grouped into three datasets consisting of 249 astrocytomas, 166 oligodendrogliomas and 112 GBM samples.

Before the application of the network inference methods, the variables of each glioma-type dataset were normalized through the *huge.npn* function from the *huge* R package [36], which applies a non-parametric normal (nonparanormal) transformation aiming to obtain variables with multivariate Gaussian distribution. For our analysis, we considered only variables normally distributed according to the Jarque–Bera test [37].

The TCGA transcriptomics datasets are constituted by a matrix containing the expression of 20,501 genes, already normalized by Transcripts Per Million (TPM) and upper quartile normalizations. After the nonparanormal normalization, a dataset of 16,654 normally distributed genes was obtained. The complete TCGA methylomics datasets contain 485,577 variables, expressing the ratio of methylation of each gene's methylation site. To reduce dimensionality, filters based on previous knowledge were applied, after removing variables with zero variance and invalid entries. In particular, we only considered C-phosphate-G (CpG) sites, as more than 98% of DNA methylation occurs in a CpG dinucleotide [38], and we excluded from the analysis: probes related to sexual chromosomes [39,40], cross-reactive and polymorphic CpGs, as suggested in [41]. At the end of this procedure, methylation datasets contain 340,427 variables. Given that the employed network inference methods are unable to deal with this high dimensionality, we developed a framework to integrate the two omics, while reducing the dimensionality of this dataset, as explained in the next section.

2.4. Pipeline and implementation

The general framework and the integrated multi-omics strategy developed for this study are schematized in Fig. 1. As a first step after data preprocessing (Fig. 1, box 1), JGL was applied to the RNA-sequencing dataset to perform a joint estimation of networks relying on distinct glioma subtypes (Fig. 1, box 2.a), which was desirable for interpretation purposes. We considered three classes corresponding to the main glioma types (astrocytoma, oligodendroglioma, and GBM), which were grouped into a list that we used as the input for the JGL function from the *JGL* package [42]. We tested several combinations of parameters (Table S1), yet only the results obtained by setting $\lambda_1 = 0.90$ and $\lambda_2 = 0.001$ were explored. JGL determined a selection of 853 variables, specifically 517, 611, and 300 for astrocytoma,

oligodendroglioma, and GBM, respectively (we refer to Section 2.1 for the explanation of the variable selection strategy). To explore the differences among the three glioma types, we focused the analysis on the genes involved in exclusive edges (i.e., edges estimated for a single glioma type), which we assumed to be more representative of the corresponding disease condition. Therefore, in the second step, the subsets of these exclusive pairs of genes were considered and associated with the corresponding CpG sites in the DNA-methylation dataset by the function *data* from the *methylGSA* R package [43]. This association led to three datasets comprising 6650, 6811, and 2917 variables for astrocytoma, oligodendroglioma, and GBM, respectively. Given the high number of variables, we performed another variable selection using the *glasso* method (Fig. 1, box 2.b), which follows the same methodology as JGL by considering one glioma type at a time. This approach was necessary as, in the methylomics layer, we focused on relationships exclusive to individual glioma types. The *glasso* method was implemented using the *huge.glasso* function from the *huge* R package [36] with $\rho = 0.85$, which led to an interpretable number of features that was sufficiently large for a comprehensive analysis (Table S2). An extensive discussion of the parameter selection process for both network inference methods is available in Supplementary Materials S2.

The results provided by the two omics were integrated through MINGLE (Fig. 1, box 2.c), which defines new gene networks based on the estimated methylomics networks (see Section 2.6 for details).

The reliability of the JGL and *glasso* results was tested by performing a mathematical validation, as explained in Supplementary Material S2. Also, sparse multinomial logistic regression and survival analysis models were employed to investigate the relevance of the biological information carried by the sets of network-based selected variables. Classification was executed by the *glmnet* R function [44], performing sparse multinomial logistic regression with the elastic net penalty. Different values of parameter $\alpha \in [0, 1]$ were tested, while the optimal value for λ was estimated through 10-fold cross-validation. For each parameter choice, 100 runs were computed, where the dataset was split into 75%-train and 25%-test sets. A variable is considered to be selected by the regularized classification method if it is identified by the algorithm for a given glioma type in at least 75% of the runs. Classification performances were evaluated by computing the Area under the ROC Curve (AUC). The complete results are presented in Tables S7 and S8. Additionally, the classification outcomes were supported by graphical visualization of the 2D-sample projections performed by Uniform Manifold Approximation and Projection (UMAP) and t-distributed stochastic neighbor embedding (t-SNE), two state-of-the-art nonlinear methods for dimensionality reduction. These projections were computed by applying the *umap* and *Rtsne* R functions, with the default parameters (details in Supplementary Materials Section S5).

The final subset of variables detected by our integrated multi-omics study was further investigated by survival analysis to determine their impact on patient prognosis. To this aim, the univariate Kaplan–Meier method was used, allowing for the estimation of the time of death in a setting of incomplete observations [45]. In particular, for each variable, patients were first stratified into high- and low-risk groups by applying two functions from the *survminer* R package [46]: *surv_cutpoint*, which determines the best cut-off point to divide the two curves, and *surv_categorize*, which separates the samples in the two groups. Then, the *survfit* function from the *survminer* R package was employed to estimate the survival probability. This output was used to plot the Kaplan–Meier curves of the statistically significant variables according to the log-rank p -value by the *ggsurvplot* function from the same package.

The presented results are obtained by codes implemented in software R (version 4.3.2). The code was implemented on a MacBook Air Apple M1 with an 8-core CPU. The scripts for reproducing the analysis with personalized choices of regularization parameters, and the developed MINGLE function, are available at the link: <https://github.com/RobertaColetti/MINGLE>.

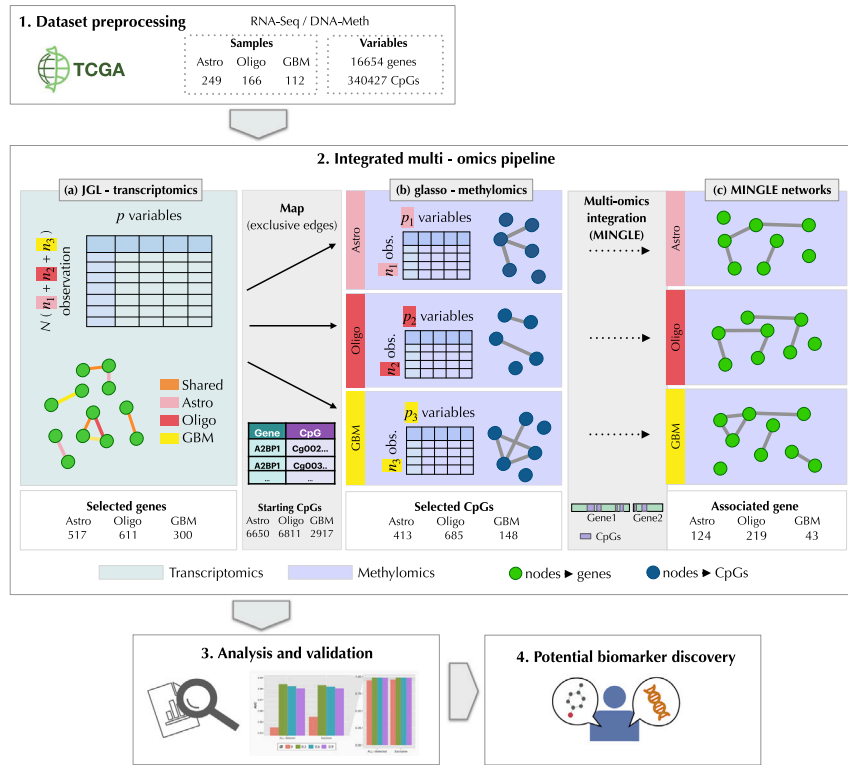


Fig. 1. Workflow of the analysis and multi-omics integration scheme. 1. Dataset preprocessing: data were downloaded from TCGA portal, and samples were grouped according to the updated glioma diagnoses. The three matrices obtained were then preprocessed to ensure considering only variables with normal distribution, as described in the corresponding “Dataset preprocessing” section. 2. Integrated multi-omics pipeline: Green and purple panels refer to networks obtained from transcriptomics and methylomics data, respectively. The nodes are colored differently to highlight the molecular entities they represent, *i.e.*, green nodes are genes, while blue nodes are CpG sites. The glioma-type transcriptomics datasets were used as input for JGL and glasso algorithm (a), which estimated joint gene networks for the three glioma types. Only the genes linked by exclusive edges were further analyzed, mapping each one with the respective CpG sites. Methylomics data were divided for the three glioma types, and the network estimation was performed separately by glasso (b), obtaining three distinct graphs. These final outcomes were used to define the MINGLE networks, *i.e.*, networks of genes whose edges are computed based on the estimated methylomics relations (c). 3. Analysis and validation: The obtained results were then analyzed and validated through sparse multinomial logistic regression, uniform manifold approximation and projection, and survival analysis models. 4. Potential biomarker discovery: Network analysis and visualization led to a list of potential biomarkers. Abbreviations: RNA-Seq, RNA sequencing; DNA-Meth, DNA methylation; Astro, astrocytoma; Oligo, oligodendroglioma; GBM, glioblastoma; CpG, CpG site.

2.5. Analysis of the results

The first two steps of the proposed integrated multi-omics pipeline lead to a subset of features (genes or CpGs) that are selected based on the estimated networks. The main goal of our analysis was to identify, among these features, those with a key role in the disease network, which could be potential biomarkers specific for a glioma type. We focused our analysis on GBM, but the same study for other glioma types can be independently explored and reproduced by the provided files and codes described in Supplementary Materials S1. Given the biological differences between transcriptomics and methylomics layers, we developed different strategies for each step of our pipeline.

Concerning the transcriptomics results, we first evaluated the joint networks to detect subnetworks of interest for deeper analysis in light of shared or exclusive relations among the glioma types. We focused on each subnetwork to disclose relationships that could potentially be relevant to GBM.

From the methylomics layer, the glasso algorithm returns a sparse precision matrix for each glioma type. Being that the starting DNA methylation datasets were built to further explore the genes involved in the exclusive relations, each resulting methylomics network is related to a single glioma type. The analysis of these networks cannot be performed as it was done for transcriptomics, given the complexity of the interpretation of such relations. Indeed, genes usually have many regions in which methylation might occur, which could result in a network in which several links are related to the same gene, affecting overall understanding and interpretability. To avoid this bias towards the identification of gene-specific methylation subnetworks and to

integrate the results from the two omics, we defined new networks of genes based on methylomics outcomes by MINGLE.

2.6. Multi-omics integrated network for graphical exploration (MINGLE)

MINGLE networks were designed to integrate the multi-omics information obtained from the defined two-step pipeline. In particular, the nodes were defined by the genes associated with the selected CpG sites, whereas the edges were computed based on the precision matrix estimated by glasso on DNA-Methylation data. This novel representation allowed us to observe the same network from a different perspective, *i.e.*, changing the nodes from CpGs to genes, while edges express a relation based on the underlying DNA-methylation processes, with a great advantage for biological interpretation.

We defined two measures for computing the network edges, leading to different MINGLE networks: one based on the strength of the connections (Strength MINGLE network, S_M), and another based on the number of connections (Degree MINGLE network, D_M), as shown in Fig. 2. Let g_i and g_j be two genes associated with certain sets of selected CpG sites M_i and M_j , respectively. To develop the S_M , we computed the edge connecting the genes i and j as:

$$a_{i,j}^S = \text{median}_{\theta_{k,l} \neq 0} (|\theta_{k,l}|)^{\frac{k \in M_i}{l \in M_j}}, \quad (7)$$

where k and l are the indices corresponding to all the CpG sites associated with the genes i and j , respectively.

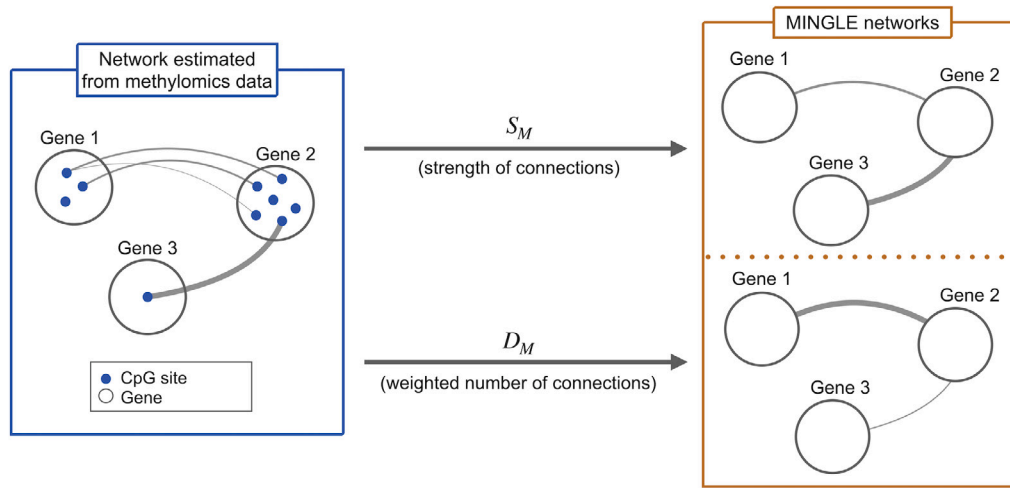


Fig. 2. MINGLE conceptual diagram. Scheme representing the transition from the glasso methylomics network (blue box) to MINGLE networks (orange box). The sparse graph estimated from methylomics data is a network of CpG sites (blue nodes). Since each site represents a specific region of a gene (white circles), MINGLE employs this known associations to define networks of genes, whose edges are computed based on the estimated methylomics relations. MINGLE networks can be designed based on the strength of the connections (S_M , from Eq. (7)), and the number of connections (D_M , from Eq. (8)). These two definitions lead to networks with different edge weights, while the overall structure of the graphs is the same. Abbreviations: S_M , Strength MINGLE network; D_M , Degree MINGLE network.

To define the D_M , we considered a weighted degree measure, computing the edge connecting two genes g_i and g_j as:

$$a_{i,j}^D = w_{i,j} \sum_{\substack{k \in M_i \\ l \in M_j}} q_{k,l}, \quad (8)$$

where $q_{k,l} = 1$ if $\theta_{k,l} \neq 0$, and $q_{k,l} = 0$ otherwise. The weights $w_{i,j} = \frac{P_i + P_j}{2}$, are calculated based on the percentages of selected CpG sites for each gene (P_i and P_j). The presence of the weights avoids the edges from being dependent on the different availability of CpG sites per gene, as genes with many CpG sites would naturally have more probability to appear highly connected, introducing a bias in the computed edges.

This representation helps in the interpretation of results and the identification of important genes at the methylomics level. Once the nodes of interest are identified, the original methylomics relations can be recovered to perform a more detailed analysis of the estimated methylomics networks.

The proposed MINGLE networks were developed to address the complex problem of integrating multi-omics information in the context of gliomas. However, independently of the type of data considered, MINGLE can be employed for a wide range of applications, not only limited to biological issues. Specifically, MINGLE allows an integrative graph visualization, expressing relations among variables of one layer with respect to the *underlying network*, i.e., previously known relations characterizing another information layer, being very useful in all cases that would benefit from an integrative graph visualization. The designed MINGLE function (available at <https://github.com/RobertaColetti/MINGLE>) only requires (1) a single-layer network structure (underlying network), (2) a functional map between the nodes of the two layers, and (3) the desired edge type, returning the adjacency matrix of MINGLE network. The simple implementation and high computational efficiency of the MINGLE function (Supplementary Materials S3), allow the application of the designed tool to a wide range of different problems and domains, effortlessly tackling large-scale datasets. The supplementary file “MINGLE_Manual.pdf” provides a detailed guide for the correct usage of the MINGLE R function, along with suggestions to harness its versatility. The manual includes two examples focused on glioblastoma, although they can be easily generalized to other diseases. These examples are based on omics networks derived from the current study and previous research [47].

3. Results and discussion

In this section, we show and discuss the main results of our integrated multi-omics analysis, with a focus on GBM. We provide the R script to reproduce the analysis and additional tables that could be independently explored. The list of the provided files and their contents, along with a usage guide, is reported in Supplementary Materials S1.

3.1. Transcriptomics networks

Fig. 3(a) shows the jointly estimated network for the three glioma types, obtained from RNA-Sequencing data with $\lambda_1 = 0.9$ and $\lambda_2 = 0.001$. Many subnetworks can be distinguished clearly, most of which are small, involving two to seven nodes. Six large subnetworks are also present (highlighted with green circles in Fig. 3(a)), which include both exclusive and shared relations among the glioma types. Despite the difficulty of interpretation of such big and dense networks, we observe a prevalence of exclusive edges in some of them. For example, the green circle A is mainly related to GBM, while B1 and B2 include many oligodendroglioma edges. The small subnetworks are easier to discuss, since they are often related to one glioma type. In particular, when these exclusive relations constitute a subnetwork with more than two nodes, it could represent an important regulation of the associated disease. For example, the gray circle C in Fig. 3(a) is mainly exclusive to GBM, while the two gray circles D1 and D2 include relations of both astrocytoma and oligodendroglioma (which are known to exhibit similarities compared to GBM).

With the purpose of deeply investigating the GBM glioma type, we focused on the estimated GBM network, which is constituted of 39 subnetworks, divided into 3 big and 36 small subnetworks (less than 6 nodes each). To reduce the complexity of the analysis, we focused on the relations we assume to be more representative of this glioma type, removing the small subnetworks involving only shared genes, or genes of the same family (i.e., trivial links). Our filter removed 23 small subnetworks related either to shared genes or trivial relations, maintaining the relations shown in Fig. 3(b). By analyzing these subnetworks, we detected some potentially interesting relations, such as:

- (1) the interaction between *CXCR6*, *CD2*, *CD96*, *CD3E*, *GZMA* and *LCK* (subnetwork C in Fig. 3(a)). Noteworthy, all these genes are known to be involved in immune-related mechanisms, in some

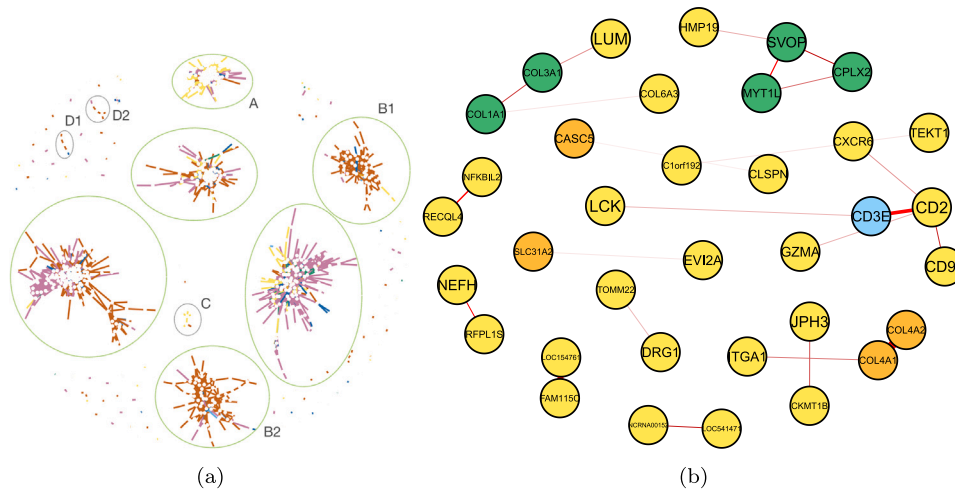


Fig. 3. Graphical representations of the estimated networks. (a) Complete JGL networks. This representation is edge-based, *i.e.*, exclusive or shared links are highlighted by different colors. (b) GBM small subnetworks after our filtering process. This representation is node-based, *i.e.*, exclusive or shared genes are highlighted by different colors. Full-page version of this figure is available at supplementary materials Section S7

■ Glioblastoma (GBM)
 ■ Oligodendroglioma (O)
 ■ Astrocytoma (A)
 ■ Shared all
 ■ Shared O-GBM
 ■ Shared A-GBM.

cases explicitly linked to GBM [48–53].

For instance, a recent study on GBM discovered that *CXCR6*, *CD2*, and *CD96* are differentially expressed in infiltrating CD4+ T cells compared with normal-appearing brain tissue [53]. In particular, *CXCR6* was also studied as a therapy target in a mouse model [50], where it has been demonstrated that blocking *CXCR6* reduces the resistance to anti-PD-1 therapy in GBM.

- (2) the strong exclusive relations, such as the one between *FAM115C* and *LOC154761*. Both these genes are differentially regulated in glioma [54,55], yet *FAM115C* also shows prognostic relevance, as it influences the overall glioma patient's survival [55,56].
- (3) the relations differing with the other glioma types. For instance, in the subnetwork with *COL4A1*, *COL4A2*, and *ITGA1*, the relation between *COL4A1* and *COL4A2* was common to all the glioma types, but only in GBM they are linked with *ITGA1*, which is associated with temozolomide chemotherapy resistance [57]. The biological validation of this estimated link could clarify if the relation among these genes could somehow play a role in GBM treatment resistance.

The interpretation of larger subnetworks might be more challenging, but it could be essential to discover potentially important relations, due to the presence of both mixed and exclusive relations. We report an example of how to accomplish this in the Supplementary Materials S4.

3.2. MINGLE networks

To further investigate the GBM-exclusive relations by multi-omics integration, we built the corresponding MINGLE networks, *i.e.*, the network of genes based on the GBM-methylomics graph estimated by glasso, defined as described in Section 2.6. For GBM, our results led to the selection of 148 CpG sites, associated with 43 genes. Among them, 23 genes were involved in inter-gene relations, depicted in Fig. 4. Interestingly, it might be worth noting that, differently from the estimated transcriptomics graphs, most of the genes here represented are part of the same network, indicating the presence of a complex structure of gene relations at the methylomics level. It is also evident that the two omics data revealed different network structures, as we can spot some connections among CpG sites of genes that were not related in RNA Sequencing-based networks. Isolated genes are not visualized in these networks, because of their inherent characteristic of having relations solely among their own CpG sites. Despite this, these genes hold relevance as their intra-gene relations might potentially influence GBM processes. The two representations in Fig. 4 show the

same network with different edge weights, which is represented by edge thickness, allowing the disclosure of distinct but complementary properties of the same relations between CpG sites of pair genes. For instance, D_{MG} (Fig. 4(a)), which is based on the number of connected probes, shows that the two genes *PLEK* and *PTPN6* are connected by multiple associations at methylomics level. However, these relations are weak, as revealed by the S_{MG} network, whose edges are computed based on the strength of the connections between the gene-related CpG sites. Conversely, *COL4A3*, *DOCK2*, *MAL*, *FA2H*, and *CPLX2* genes appear strongly connected in Fig. 4(b), while involving few methylomics relations.

To support the reliability of the MINGLE-based findings, research literature on the role of these genes in cancer was performed. In particular, as the presented study is focused on identifying potentially important regulatory mechanisms, we investigated the functionalities of the identified groups of genes. To the best of our knowledge, there are no specific studies linking *COL4A3*, *DOCK2*, *MAL*, *FA2H*, and *CPLX2*. However, each of them is known to affect intracellular signaling and cellular functions, which could further impact the tumor landscape. For example, a previous study has shown that collagen family genes are positively correlated with several immune cell types in LGG and GBM samples [58]. Although *COL4A3* (a member of the collagen family gene) was not specifically mentioned in this study, we may hypothesize its potential influence on GBM heterogeneity, including its tumor microenvironment landscape. This influence may occur through its effect on different pathways, potentially resulting in different phenotypical profiles and contributing to tumor differentiation.

If the knowledge about this group of genes in the literature is limited, our research on *PLEK* and *PTPN6* led to significant findings. Previous reports on other cancer types evidence that both *PLEK* and *PTPN6* play a critical role in cellular signaling, suggesting these genes as potentially relevant also in the context of GBM. In particular, *PLEK*, known for its involvement in cytoskeletal organization, may be implicated in various cellular signaling cascades crucial for GBM aggressiveness and heterogeneity, such as PI3K/AKT pathway [59–61]. On the other hand, *PTPN6* has been described as able to affect the activation status of various intracellular signaling pathways, including JAK/STAT, PI3K/AKT, and MAPK pathways [62]. Interestingly, in glioma, *PTPN6* is known to be affected by epigenetic regulations influencing prognosis and treatment responses [63]. Since methylation is one of the main epigenetic mechanisms, the fact that this gene has been pointed out by our analysis based on this omic layer further reinforces the trustworthiness of our results, supporting the need for additional biological validation studies on this pair of genes.

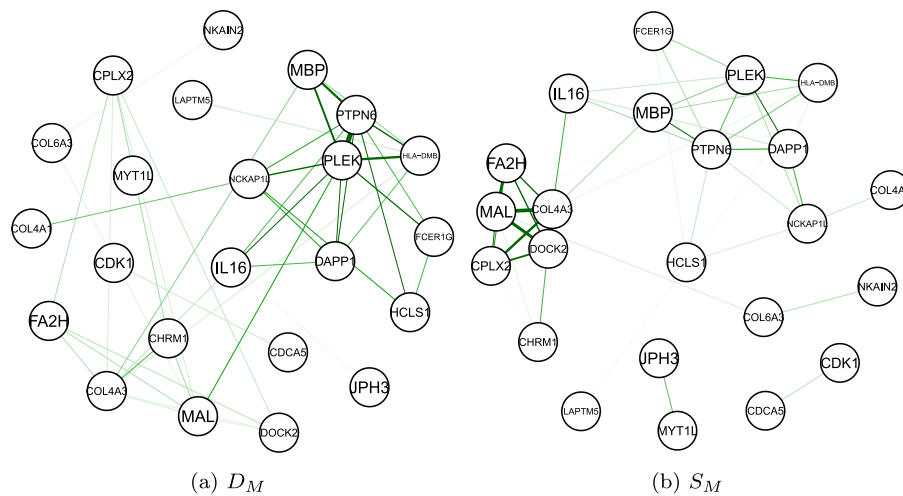


Fig. 4. GBM MINGLE networks. The structure of the two graphs is the same (i.e., genes are connected by the same link), while the edge weights are computed by taking into account (a) the number, or (b) the strength of the relations in the methylomics networks. Full-page version of this figure is available at supplementary materials Section S7. Abbreviations: S_M , Strength MINGLE network; D_M , Degree MINGLE network.

The elucidation of the combined role of these two genes in glioma may also hold significant clinical and therapeutic implications. Evidence suggests that the pleckstrin signaling pathway, in which *PLEK* is involved, associates with immune responses and regulates the RTK/PI3K signaling pathway, widely known to be of critical relevance in GBM [64,65], and for which targeted therapeutic agents have been investigated [66]. Particularly interesting is the role of *PLEKHG5*, a pleckstrin-family member, which affects the viability of GBM cells, and *MGMT* expression [67], a clinical biomarker predictive of sensitivity to temozolomide-based chemotherapy in GBM patients. Therefore, exploring how *PLEK* expression may impact GBM sensitivity to RTK/PI3K-targeting therapies, alone or in combination with pleckstrin-targeting agents, is warranted. Noteworthy, novel inhibitors of *PLEK* proteins have recently been developed and tested for various cancers [68], offering promising prospects for novel GBM therapies.

As shown, the proposed MINGLE network representations are very useful to detect genes characterized by potentially relevant methylomics relations, yet losing the original meaning of the edges as conditional dependence. However, once nodes of interest are identified, it is always possible to go back to the original network to further explore the CpG site relations. For instance, Figs. 5 and 6 show the estimated networks among the probes of the discussed genes. The two figures provide two different representations, one with nodes distributed based on their distance in the network (Figs. 5(a) and 6(a)), which simplifies the observation of the overall network structure, and another with CpGs sorted with respect to their location in the genome (Figs. 5(b) and 6(b)). The latter can be particularly useful to detect if there is a specific region of the gene that appears to be more involved in methylomics relations, which could suggest a specific area particularly accessible to methylation.

All the information necessary to reproduce similar networks for other variables of interest is collected in the provided Supplementary files.

3.3. Classification

To ensure that the estimated sparse networks retain biologically relevant variables, we performed patients' classification through sparse multinomial logistic regression. Good classification performances will indicate that the glioma differences across types have been preserved.

For transcriptomics, the classification algorithm was applied to four datasets, comprising all the glioma samples and: (i) the starting 16,655 variables set, (ii) the set of selected variables (union of the variables

selected for each glioma type), (iii) variables linked by glioma-type exclusive edges, and (iv) glioma-type exclusive variables (the nodes appearing only in one glioma-type network). By performing classification on these four datasets, we aim to investigate not only if the three glioma types are well characterized by the set of identified genes, but also if exclusive nodes (or exclusive pairs) can provide comparable performances, yet further reducing the number of variables. Moreover, by employing regularized classification, we are able to identify single genes that allow the separation of patients into the different glioma types, thus being promising candidates as diagnostic markers. It is important to note that the diagnostic variables selected via model regularization from the original starting dataset (case (i)) are different from the ones selected from the reduced set (cases (ii)-(iv)), indicating the existence of different sets of genes suitable for classification purposes. Nevertheless, the proposed approach also estimates a network structure, which represents a remarkable advantage in terms of biological interpretability, being extremely important for a better understanding of the glioma molecular mechanisms.

Regularization can be modulated by the parameter α , which was set between 0 (Ridge; no feature selection) and 1 (LASSO; feature selection). Table S7 resumes the classification outcomes for different values of α . For $\alpha = 0$, better performances are obtained by considering the complete set of variables. However, despite the considerable variable reduction (< 20% of the initial dataset), our identified subsets are also providing excellent performances, with AUC > 0.9. Notably, reducing the variables to the exclusive nodes (case (iv)) provides better results, supporting that the genes that are exclusive of one glioma-type network are more biologically representative of the disease. Classification performances enhance with a small regularization.

Regarding methylomics results, we reduced the cases of study to the first two, i.e., starting variables versus selected ones. The other two cases involving exclusive variables were not taken into account, as the methylomics analysis was specifically defined to target CpG sites associated with nodes linked by exclusive relations. The results of classification in the two cases show comparable AUC values for all the values of the regularization parameter (Table S8), proving that the subset of selected CpGs is not losing the original diagnostic information.

To further validate our sets of selected variables, and to provide graphical support to the classification outcomes, unsupervised 2D-sample projection was performed by UMAP and t-SNE methods. In particular, we considered the full starting dataset and the reduced one, constituted by the subset of variables detected by the network-based approach, for both transcriptomics and methylomics data (cases (i) and

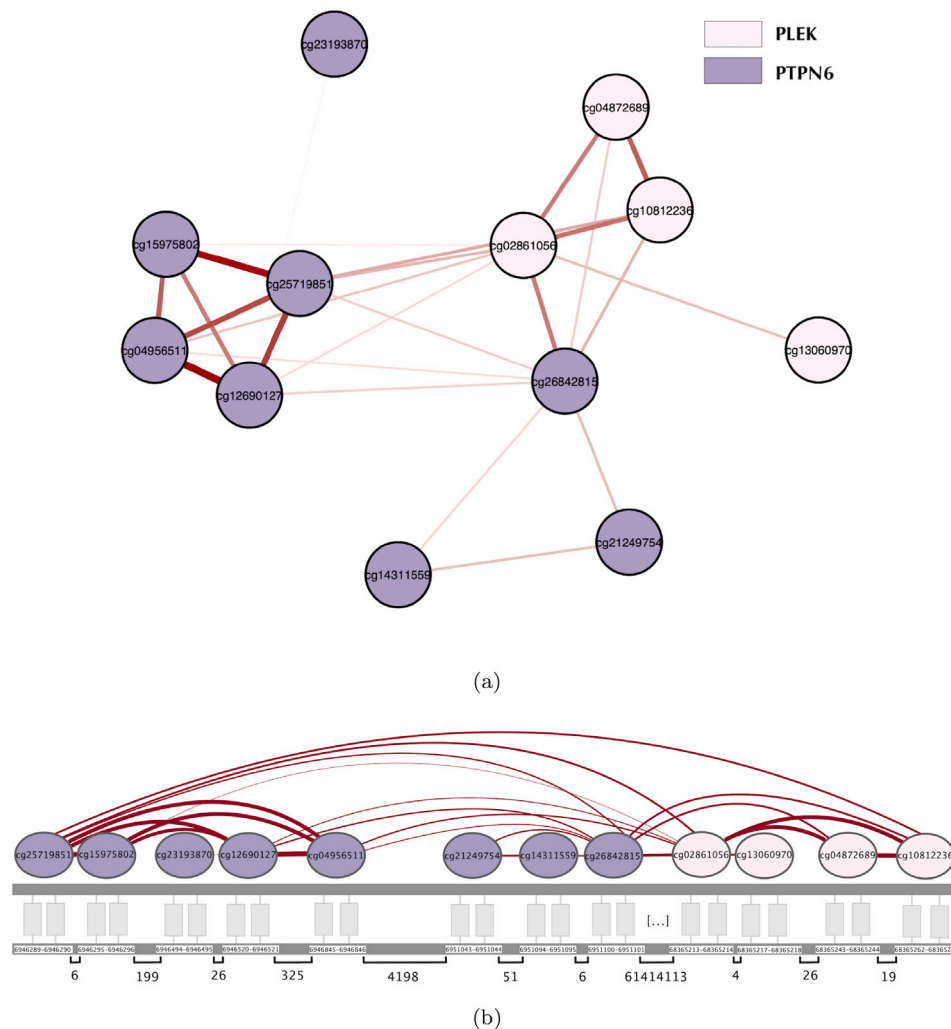


Fig. 5. Original estimated relations between CpG sites of the two genes *PLEK* (pink) and *PTPN6* (purple). (a) The network layout is based on the Fruchterman-Reingold force-directed algorithm, which distributes nodes based on their distance in the network (neighbors are closer). (b) The nodes are represented according to their location in the genome. For each CpG, the corresponding nucleotides are reported, along with the computed distance between two subsequent CpGs. Full-page version of the figure (b) is available at supplementary materials Section S7.

(ii) defined above). These four datasets (two per omics) were used as input of UMAP and t-SNE methods to obtain a 2D projection of the samples colored according to their diagnostic labels. These graphical results (Supplementary Figure S7) align with the classification outcomes, showing well-separated groups despite the considerable variable reduction.

A comprehensive analysis of the classification results and of the 2D-sample projections, for both transcriptomics and methylomics data, is reported in Supplementary Materials S5.

3.4. Survival analysis

Survival analysis was performed to detect features having an impact on GBM's overall survival. To this end, we focused on the final outcome of our multi-omics integration procedure, which selected 148 CpGs associated with 43 genes. We investigated both omics, including in the survival analysis all the 43 genes, and the top 20 nodes of the methylomics networks, taking into account both, the number of connections, and the median sum of weights (in case of multiple variables with the same ranking, this selection has been expanded to include all the nodes with equivalent importance). The Kaplan–Meier curves for the statistically significant variables are provided as supplementary material (see files *SA-DNAMeth-count*, *SA-DNAMeth-weight*, and *SA-RNASeq*). Based

on transcriptomics data, we detected 27 over 43 genes significantly affecting GBM survival (p -value < 0.05). Interestingly, many of them are isolated in the defined MINGLE networks. Literature research on the genes with lower p -values (Fig. 7(a)) revealed that most of them act in cancer-related processes, also in the context of glioma [69–72]. The only exception regards *PLEK*, about which there are no exhaustive studies investigating its role in cancer. However, based on our estimated networks, this gene is directly linked with *FCER1G* and *PTPN6*, both known to influence glioma patients' prognosis and treatment responses [63,73].

In methylomics, a total of 42 variables were investigated, namely 22 highly connected, and 20 strongly connected nodes. Among them, 23 CpGs had an impact on GBM prognosis, corresponding to a total of 13 genes.

Fig. 7(b) shows the Kaplan–Meier curves for the 5 CpGs with lower p -values. Again, we encounter genes already pointed out by previous analyses, such as *MAL*, *BIRC5*, and *PLEK*. The CpG with the strongest impact on GBM survival is associated with gene *NCKAP1L*, whose probes bind to *PLEK* and *PTPN6* CpG sites (Fig. 4), recurrent genes in our study. Although *NCKAP1L* is not yet studied in glioma, it is known to be a tumor suppressor gene [74], and barely expressed in several cancers [74–76]. Further survival analysis results are available in Supplementary Materials S6.

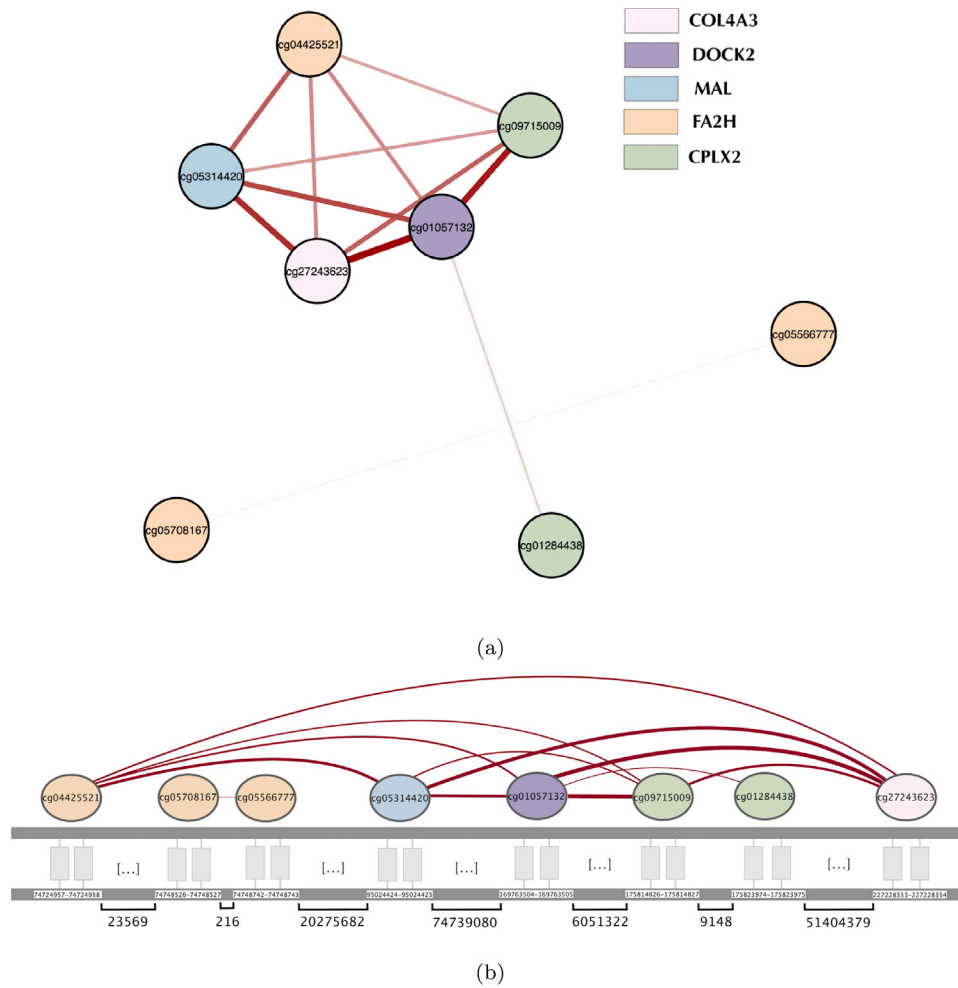


Fig. 6. Original estimated relations between CpG sites of the genes *COL4A3* (pink), *DOCK2* (purple), *MAL* (light blue), *FA2H* (orange) and *CPLX2* (green). (a) Network layout is based on the Fruchterman-Reingold force-directed algorithm, which distributes nodes based on their distance in the network (neighbors are closer). (b) The nodes are represented according to their location in the genome. For each CpG, the corresponding nucleotides are reported, along with the computed distance between two subsequent CpGs. Full-page version of the figure (b) is available at supplementary materials Section S7.

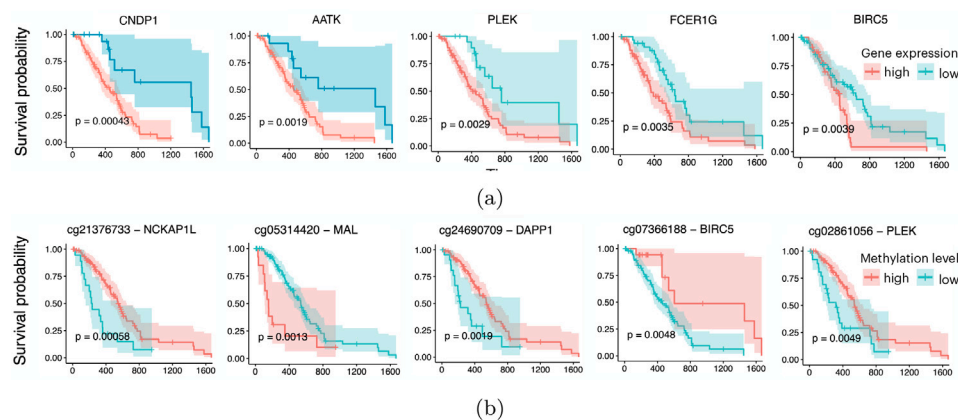


Fig. 7. Kaplan-Meier curves of the top 5 (a) genes and (b) CpG sites with lower p -values. The complete results are available at the supplementary files *SA-DNAMeth-count*, *SA-DNAMeth-weight*, and *SA-RNASeq*.

4. Conclusion

The presented work represents a pioneering study on multi-omics integration from glioma patients grouped following the latest glioma classification guidelines release. The graphical integration through MINGLE led to the identification of gene subnetworks that could be

associated with the specific glioma type. The different network structures obtained from transcriptomics and methylomics further stress the importance of a multi-omics study, which allows the disclosure of unknown underlying mechanisms that would be impossible to detect by a single-omics analysis. The genes identified from both omics layers were discussed in terms of their promising role in glioma regulation,

especially in GBM, with a focus on their potential clinical applications. Future studies are required to confirm direct interactions between these groups of genes, which could unveil new therapeutic GBM targets. Ground-breaking techniques such as single-cell sequencing and spatial transcriptomics would be promising for elucidating these gene-to-gene interactions, identifying regulatory networks, and understanding their effects on tumor landscape and heterogeneity.

Patient classification and 2D-samples' projection confirmed the potentiality of the subset of identified variables as biomarkers of glioma heterogeneity, being able to provide a good distinction among the defined glioma types, especially if considering exclusive variables.

Survival analysis focused on GBM patient stratification revealed that around 50% of the investigated variables have prognostic value, pointing out potential interesting biomarkers from a multi-omics perspective. In particular, most of these genes, which are involved in GBM-exclusive relations from the transcriptomics layer, appear as isolated nodes in the MINGLE network, suggesting that they are solely characterized by links among their own CpG sites. If these findings are biologically verified, these genes might hold significant promise as therapeutic targets for GBM. Indeed, despite their relevance to the disease, the predicted limited impact of such genes on the overall network suggests that targeted modifications may increase the specificity and decrease effects in other crucial regulatory processes.

The proposed framework could be generalized to the cases in which merging network information coming from different omics could be useful for biological interpretation purposes. For instance, MINGLE networks could be designed from other omics with previously known associations, such as transcriptomics and proteomics, or genomics. Also, MINGLE could be applied to other data types and contexts, as a general tool for integrating multi-layer networks when the functional link between the variables of two layers is available. The main goal of MINGLE is to support the understanding of a network based on underlying relations. To achieve this, the resulting MINGLE network should remain interpretable, avoiding excessive dimensionality. Conversely, MINGLE is capable of handling large-scale underlying networks, with a high number of nodes. There are some limitations of MINGLE due to its mathematical definition. Firstly, the MINGLE networks are undirected, which only identifies the presence of relationships between variables. When considering causal networks, the proposed tool can be extended to take into account directional edges, enabling the discovery of causal underlying relationships. Secondly, MINGLE was specifically designed to merge information from two omics. If the intention is to integrate more than two layers, MINGLE can be applied recursively. A future extension could include the simultaneous integration of multiple layers.

Furthermore, this work provides customizable tables that serve as a tool to independently explore the obtained results. These tables collect genes and CpG-site relations, specific for each glioma type, resulting from our network inference, classification, and survival outcomes. The exploration of such files might lead to the detection of other promising groups of features as novel potential biomarkers or worth of future laboratory validation. Groups of genes of interest can also be identified by applying module detection strategies to the estimated networks. Different methodologies can be used for this scope, which have proven effective in many applications [77–81], offering the potential to uncover additional insights into glioma. These results might motivate further investigation into the role of the genes we identified as potential biomarkers, with the aim of improving the clinical management of glioma patients.

As the next step, we plan to conduct biological validation to confirm the importance of the detected variables. The genes will be compiled into a list and ranked based on their importance within the network, as well as their influence on classification and survival analysis. The top-ranking genes will undergo a thorough literature review and clinical validation using publicly available databases to assess their prognostic value across various datasets and identify the most promising targets. Following this, we will engineer glioma cell lines by overexpressing or silencing the candidate genes to assess their relevance through *in vitro* and *in vivo* experiments.

CRediT authorship contribution statement

Roberta Coletti: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **João F. Carrilho:** Software, Methodology, Formal analysis. **Eduarda P. Martins:** Writing – review & editing, Conceptualization. **Céline S. Gonçalves:** Writing – review & editing, Conceptualization. **Bruno M. Costa:** Writing – review & editing, Funding acquisition, Conceptualization. **Marta B. Lopes:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references CEECINST/00042/2021, 2021.02600.CEECIND (doi: [10.54499/2021.02600.CEECIND/CP1664/CT0015](https://doi.org/10.54499/2021.02600.CEECIND/CP1664/CT0015)), UMINHO/BIM-CNCG/2023/102, UIDB/00297/2020 and UIDP/00297/2020 (NOVA Math, Center for Mathematics and Applications), UIDB/00667/2020 and UIDP/00667/2020 (UNIDEMI), CEECIND/00072/2018/CP1581/CT0011 (doi: [10.54499/CEECIND/00072/2018/CP1581/CT0011](https://doi.org/10.54499/CEECIND/00072/2018/CP1581/CT0011)), CEECINST/00077/2018/CP1640/CT0002 (doi: [10.54499/CEECINST/00077/2018/CP1640/CT0002](https://doi.org/10.54499/CEECINST/00077/2018/CP1640/CT0002)), 2022.04859.PTDC (doi: [10.54499/2022.04859.PTDC](https://doi.org/10.54499/2022.04859.PTDC)), UIDB/50026/2020 (doi: [10.54499/UIDB/50026/2020](https://doi.org/10.54499/UIDB/50026/2020)), UIDP/50026/2020 (doi: [10.54499/UIDP/50026/2020](https://doi.org/10.54499/UIDP/50026/2020)), LA/P/0050/2020 (doi: [10.54499/LA/P/0050/2020](https://doi.org/10.54499/LA/P/0050/2020)), and the research project PTDC/CCI-BIO/4180/2020 “MONET – Multi-omic networks in gliomas” (doi: [10.54499/PTDC/CCI-BIO/4180/2020](https://doi.org/10.54499/PTDC/CCI-BIO/4180/2020)). This work was also supported by the project NORTE-01-0145-FEDER-000055, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). The results presented here are based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.combiomed.2025.109811>.

Data availability

The glioma data used for this study are publicly available at <http://firebrowse.org/>, and can be accessed by selecting LGG and GBM projects. The updated diagnostic labels considered to group patients according to the latest guidelines for glioma classification are at <https://github.com/sysbiomed/MONET>. The data prepared for this study, together with the scripts to reproduce the analysis, are available at <https://github.com/RobertaColetti/MINGLE>.

References

- [1] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S.M. Pfister, R. Nishikawa, M. Rosenthal, P.Y. Wen, R. Stupp, G. Reifenberger, Glioma, *Nat. Rev. Dis. Prim.* 1 (2015) 15017, <http://dx.doi.org/10.1038/nrdp.2015.17>.
- [2] J.G. Nicholson, H.A. Fine, Diffuse glioma heterogeneity and its therapeutic implications, *Cancer Discov.* 11 (2021) 575–590, <http://dx.doi.org/10.1158/2159-8290.CD-20-1474>.

- [3] V. Ehrlich, C. Gerner, C. Hoelzl, S. Knasmüller, W. Mikulits, M. Mišák, A. Nersisyan, A. Szakmary, K.-H. Wagner, Use of conventional and -omics based methods for health claims of dietary antioxidants: a critical overview, 2008/05/01, Br. J. Nutr. 99 (2008) ES3–ES52, <http://dx.doi.org/10.1017/S0007114508965752>.
- [4] Z. Zhao, K. Zhang, Z. Wang, K. Wang, X. Liu, F.A. Wu, J. Chen, A comprehensive review of available omics data resources and molecular profiling for precision glioma studies (Review), Biomed. Rep. 10 (2019) 3–9, <http://dx.doi.org/10.3892/br.2018.1168>.
- [5] B. Delgado-Martín, M.Á. Medina, Advances in the knowledge of the molecular biology of glioblastoma and its impact in patient diagnosis, stratification, and treatment, Adv. Sci. 7 (2020) 1902971, <http://dx.doi.org/10.1002/adv.201902971>.
- [6] D.N. Louis, A. Perry, P. Wesseling, D.J. Brat, I.A. Cree, D. Figarella-Branger, C. Hawkins, H.K. Ng, S.M. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling, D.W. Ellison, The 2021 WHO classification of tumors of the central nervous system: a summary, Neuro-Oncol. 23 (2021) 1231–1251, <http://dx.doi.org/10.1093/neuonc/noab106>.
- [7] J. Wu, X. Wang, X. Yuan, Q. Shan, Z. Wang, Y. Wu, J. Xie, Kinesin family member C1 increases temozolomide resistance of glioblastoma through promoting DNA damage repair, Cell Transplant. 30 (2021) 0963689721991466, <http://dx.doi.org/10.1177/0963689721991466>.
- [8] Q.T. Ostrom, M. Prince, C. Neff, G. Cioffi, K. Waite, C. Kruchko, J.S. Barnholtz-Sloan, CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019, Neuro-Oncol. 24 (2022) v1–v95, <http://dx.doi.org/10.1093/neuonc/noac202>.
- [9] C. Ma, M. Wu, S. Ma, Analysis of cancer omics data: a selective review of statistical techniques, Brief. Bioinform. 23 (2022) bbab585, <http://dx.doi.org/10.1093/bib/bbab585>.
- [10] Q. Yang, B. Li, J. Tang, X. Cui, Y. Wang, X. Li, J. Hu, Y. Chen, W. Xue, Y. Lou, Y. Qiu, F. Zhu, Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data, Brief. Bioinform. 21 (2020) 1058–1068, <http://dx.doi.org/10.1093/bib/bbz049>.
- [11] F. Li, J. Yin, M. Lu, Q. Yang, Z. Zeng, B. Zhang, Z. Li, Y. Qiu, H. Dai, Y. Chen, F. Zhu, Consig: consistent discovery of molecular signature from OMIC data, Brief. Bioinform. 23 (2022) bbac253, <http://dx.doi.org/10.1093/bib/bbac253>.
- [12] S. Martins, R. Coletti, M.B. Lopes, Disclosing transcriptomics network-based signatures of glioma heterogeneity using sparse methods, BioData Min. 16 (2023) 26, <http://dx.doi.org/10.1186/s13040-023-00341-1>.
- [13] E.-I. Delatola, M. Dakna, Statistical Inference in High-Dimensional Omics Data, 2018 John Wiley & Sons, Inc., 2018, pp. 196–206, <http://dx.doi.org/10.1002/9781119183952.ch11>.
- [14] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, R. Guthke, Gene regulatory network inference: Data integration in dynamic models—A review, Biosystems 96 (1) (2009) 86–103, <http://dx.doi.org/10.1016/j.biosystems.2008.12.004>.
- [15] Q. Shi, C. Zhang, M. Peng, X. Yu, T. Zeng, J. Liu, L. Chen, Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data, Bioinformatics 33 (2017) 2706–2714, <http://dx.doi.org/10.1093/bioinformatics/btx176>.
- [16] S. Ren, Y. Shao, X. Zhao, C.S. Hong, F. Wang, X. Lu, J. Li, G. Ye, M. Yan, Z. Zhuang, C. Xu, G. Xu, Y. Sun, Integration of metabolomics and transcriptomics reveals major metabolic pathways and potential biomarker involved in prostate cancer*, Mol. Cell. Proteom. 15 (2016) 154–163, <http://dx.doi.org/10.1074/mcp.M115.052381>.
- [17] Y. Hasin, M. Seldin, A. Lusi, Multi-omics approaches to disease, Genome Biol. 18 (2017) 83, <http://dx.doi.org/10.1186/s13059-017-1215-1>.
- [18] I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika, Multi-omics data integration, interpretation, and its application, Bioinform. Biol. Insights 14 (2020) <http://dx.doi.org/10.1177/1177932219899051>.
- [19] N. Wani, K. Raza, Integrative approaches to reconstruct regulatory networks from multi-omics data: A review of state-of-the-art methods, Comput. Biol. Chem. 83 (2019) 107120, <http://dx.doi.org/10.1016/j.compbiolchem.2019.107120>.
- [20] M.L. Mendonça, R. Coletti, C.S. Gonçalves, E.P. Martins, B.M. Costa, S. Vinga, M.B. Lopes, Updating TCGA glioma classification through integration of molecular profiling data following the 2016 and 2021 WHO guidelines, 2023, <http://dx.doi.org/10.1101/2023.02.19.529134>, BioRxiv.
- [21] H. Cedar, DNA methylation and gene activity, Cell 53 (1) (1988) 3–4, [http://dx.doi.org/10.1016/0092-8674\(88\)90479-5](http://dx.doi.org/10.1016/0092-8674(88)90479-5).
- [22] C. Liu, H. Tang, N. Hu, T. Li, Methyloinformatics and cancer: the current state of methylation profiling and marker development for clinical care, Cancer Cell Int. 23 (2023) 242, <http://dx.doi.org/10.1186/s12935-023-03074-7>.
- [23] Q. Wu, A.E. Berglund, A.B. Etame, The impact of epigenetic modifications on adaptive resistance evolution in glioblastoma, Int. J. Mol. Sci. 22 (2021) <http://dx.doi.org/10.3390/ijms22158324>.
- [24] A. Ozair, V. Bhat, R.S. Alish, A.A. Khosla, R.R. Kotecha, Y. Odia, M.W. McDermott, M.S. Ahluwalia, DNA methylation and histone modification in low-grade gliomas: Current understanding and potential clinical targets, Cancers 15 (2023) <http://dx.doi.org/10.3390/cancers15041342>.
- [25] J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, Biostatistics 9 (2008) 432–441, <http://dx.doi.org/10.1093/biostatistics/kxm045>.
- [26] P. Danaher, P. Wang, D.M. Witten, The joint graphical lasso for inverse covariance estimation across multiple classes, J. R. Stat. Soc. Ser. B Stat. Methodol. 76 (2014) 373–397, <http://dx.doi.org/10.1111/rssb.12033>.
- [27] A. Santra, S. Bhowmick, S. Chakravarty, Efficient community re-creation in multilayer networks using boolean operations, Procedia Comput. Sci. 108 (2017) 58–67, <http://dx.doi.org/10.1016/j.procs.2017.05.246>.
- [28] X. Sun, H. Zhuge, Merging complex networks, in: 2011 Seventh International Conference on Semantics, Knowledge and Grids, 2011, pp. 233–236, <http://dx.doi.org/10.1109/SKG.2011.41>.
- [29] M.D. Domenico, V. Nicosia, A. Arenas, V. Latora, Structural reducibility of multilayer networks, Nat. Commun. 6 (2015) 6864, <http://dx.doi.org/10.1038/ncomms7864>.
- [30] S.L. Lauritzen, Graphical Models, Clarendon Press, 1996.
- [31] J. Zhu, T. Hastie, Classification of gene microarrays by penalized logistic regression, Biostatistics 5 (3) (2004) 427–443.
- [32] T.C.G.A.R. Network, Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas, N. Engl. J. Med. 372 (2015) 2481–2498, <http://dx.doi.org/10.1056/NEJMoa1402121>.
- [33] C.W. Brennan, R.G. Verhaak, A. McKenna, B. Campos, H. Nounshmehr, S.R. Salama, S. Zheng, D. Chakravarty, J.Z. Sanborn, S.H. Berman, R. Beroukheim, B. Bernard, C.-J. Wu, G. Genovesi, I. Shmulevich, J. Barnholtz-Sloan, L. Zou, R. Vegesna, S.A. Shukla, G. Ciriello, W.K. Yung, W. Zhang, C. Sougnez, T. Mikkelsen, K. Aldape, D.D. Bigner, E.G.V. Meir, M. Prados, A. Sloan, K.L. Black, J. Eschbacher, G. Finocchiaro, W. Friedman, D.W. Andrews, A. Guha, M. Iacocca, B.P. O'Neill, G. Foltz, J. Myers, D.J. Weisenberger, R. Penny, R. Kucherlapati, C.M. Perou, D.N. Hayes, R. Gibbs, M. Marra, G.B. Mills, E. Lander, P. Spellman, R. Wilson, C. Sander, J. Weinstein, M. Meyerson, S. Gabriel, P.W. Laird, D. Haussler, G. Getz, L. Chin, C. Benz, J. Barnholtz-Sloan, W. Barrett, Q. Ostrom, Y. Wolinsky, K.L. Black, B. Bose, P.T. Boulos, M. Boulos, J. Brown, C. Czerinski, M. Eppley, M. Iacocca, T. Kempista, T. Kitko, Y. Koymann, B. Rabeno, P. Rastogi, M. Sugarman, P. Swanson, K. Yalamanchi, I.P. Otey, Y.S. Liu, Y. Xiao, J. Auman, P.-C. Chen, A. Hadjipanayis, E. Lee, S. Lee, P.J. Park, J. Seidman, L. Yang, R. Kucherlapati, S. Kalkanis, T. Mikkelsen, L.M. Poisson, A. Raghunathan, L. Scarpace, B. Bernard, R. Bressler, A. Eakin, L. Iype, R.B. Kreisberg, K. Leinonen, S. Reynolds, H. Rovira, V. Thorsson, I. Shmulevich, M.J. Annala, R. Penny, J. Paulauski, E. Curley, M. Hatfield, D. Mallory, S. Morris, T. Shelton, C. Shelton, M. Sherman, P. Yena, L. Cuppini, F. DiMeo, M. Eoli, G. Finocchiaro, E. Maderna, B. Pollo, M. Saini, S. Balu, K.A. Hadley, L. Li, C.R. Miller, Y. Shi, M.D. Topal, J. Wu, G. Dunn, C. Giannini, B.P. O'Neill, B.A. Aksoy, Y. Antipin, L. Borsu, S.H. Berman, C.W. Brennan, E. Cerami, D. Chakravarty, G. Ciriello, J. Gao, B. Gross, A. Jacobsen, M. Ladanyi, A. Lash, Y. Liang, B. Reva, C. Sander, N. Schultz, R. Shen, N.D. Socci, A. Viale, M.L. Ferguson, Q.-R. Chen, J.A. Demchok, L.A. Dillon, K.R.M. Shaw, M. Sheth, R. Tarnuzzer, Z. Wang, L. Yang, T. Daviden, M.S. Guyer, B.A. Ozenberger, H.J. Sofia, J. Bergsten, J. Eckman, J. Harr, J. Myers, C. Smith, K. Tucker, C. Winemiller, L.A. Zach, J.Y. Ljubimova, G. Eley, B. Ayala, M.A. Jensen, A. Kahn, T.D. Pihl, D.A. Pot, Y. Wan, J. Eschbacher, G. Foltz, N. Hansen, P. Hothi, B. Lin, N. Shah, J. geun Yoon, C. Lau, M. Berens, K. Ardlie, R. Beroukheim, S.L. Carter, A.D. Cherniack, M. Noble, J. Cho, K. Cibulskis, D. DiCara, S. Frazer, S.B. Gabriel, N. Gehlenborg, J. Gentry, D. Heiman, J. Kim, R. Jing, E.S. Lander, M. Lawrence, P. Lin, W. Mallard, M. Meyerson, R.C. Onofrio, G. Saksena, S. Schumacher, C. Sougnez, P. Stojanov, B. Tabak, D. Voet, H. Zhang, L. Zou, G. Getz, N.N. Dees, L. Ding, L.L. Fulton, R.S. Fulton, K.-L. Kanchi, E.R. Mardis, R.K. Wilson, S.B. Baylin, D.W. Andrews, L. Harshyne, M.L. Cohen, K. Devine, A.E. Sloan, S.R. VandenBerg, M.S. Berger, M. Prados, D. Carlin, B. Craft, K. Ellrott, M. Goldman, T. Goldstein, M. Grifford, D. Haussler, S. Ma, S. Ng, S.R. Salama, J.Z. Sanborn, J. Stuart, T. Swatloski, P. Waltman, J. Zhu, R. Foss, B. Frentzen, W. Friedman, R. McTiernan, A. Yachnis, D.N. Hayes, C.M. Perou, S. Zheng, R. Vegesna, Y. Mao, R. Akbani, K. Aldape, O. Bogler, G.N. Fuller, W. Liu, Y. Liu, Y. Lu, G. Mills, A. Protopopov, X. Ren, Y. Sun, C.-J. Wu, W.A. Yung, W. Zhang, J. Zhang, K. Chen, J.N. Weinstein, L. Chin, R.G. Verhaak, H. Nounshmehr, D.J. Weisenberger, M.S. Bootwalla, P.H. Lai, T.J.T. Jr., D.J.V.D. Berg, P.W. Laird, D.H. Gutmann, N.L. Lehman, E.G. VanMeir, D. Brat, J.J. Olson, G.M. Mastrogiannis, N.S. Devi, Z. Zhang, D. Bigner, E. Lipp, R. McLendon, The somatic genomic landscape of glioblastoma, Cell 155 (2013) 462–477, <http://dx.doi.org/10.1016/j.cell.2013.09.034>.
- [34] R. McLendon, A. Friedman, D. Bigner, E.G.V. Meir, D.J. Brat, G.M. Mastrogiannis, J.J. Olson, T. Mikkelsen, N. Lehman, K. Aldape, W.K.A. Yung, O. Bogler, S. VandenBerg, M. Berger, M. Prados, D. Muzny, M. Morgan, S. Scherer, A. Sabo, L. Nazareth, L. Lewis, O. Hall, Y. Zhu, Y. Ren, O. Alvi, J. Yao, A. Hawes, S. Jhangiani, G. Fowler, A.S. Lucas, C. Kovar, A. Cree, H. Dinh, J. Santibanez, V. Joshi, M.L. Gonzalez-Garay, C.A. Miller, A. Milosavljevic, L. Donehower, D.A. Wheeler, R.A. Gibbs, K. Cibulskis, C. Sougnez, T. Fennell, S. Mahan, J. Wilkinson, L. Ziaugra, R. Onofrio, T. Bloom, R. Nicol, K. Ardlie, J. Baldwin, S. Gabriel, E.S. Lander, L. Ding, R.S. Fulton, M.D. McLellan, J. Wallis, D.E. Larson, X. Shi, R. Abbott, L. Fulton, K. Chen, D.C. Koboldt, M.C. Wendl, R. Meyer, Y. Tang, L. Lin, J.R. Osborne, B.H. Dunford-Shore, T.L. Miner, K. Delehaunty, C. Markovic, G. Swift, W. Courtney, C. Pohl, S. Abbott, A. Hawkins, S. Leong, C. Haipek, H. Schmidt, M. Wiechert, T. Vickery, S. Scott, D.J. Dooling, A. Chinwalla, G.M. Weinstock, E.R. Mardis, R.K. Wilson, G. Getz, W. Winckler, R.G.W. Verhaak, M.S. Lawrence, M. O'Kelly, J. Robinson, G. Alexe,

- R. Beroukhi, S. Carter, D. Chiang, J. Gould, S. Gupta, J. Korn, C. Mermel, J. Mesirov, S. Monti, H. Nguyen, M. Parkin, M. Reich, N. Stransky, B.A. Weir, L. Garraway, T. Golub, M. Meyerson, L. Chin, A. Protopopov, J. Zhang, I. Perna, S. Aronson, N. Sathiamoorthy, G. Ren, J. Yao, W.R. Wiedemeyer, H. Kim, S.W. Kong, Y. Xiao, I.S. Kohane, J. Seidman, P.J. Park, R. Kucherlapati, P.W. Laird, L. Cope, J.G. Herman, D.J. Weisenberger, F. Pan, D.V.D. Berg, L.V. Neste, J.M. Yi, K.E. Schuebel, S.B. Baylin, D.M. Absher, J.Z. Li, A. Southwick, S. Brady, A. Aggarwal, T. Chung, G. Sherlock, J.D. Brooks, R.M. Myers, P.T. Spellman, E. Purdom, L.R. Jakkula, A.V. Lapuk, H. Marr, S. Dorton, Y.G. Choi, J. Han, A. Ray, V. Wang, S. Durinck, M. Robinson, N.J. Wang, K. Vranizan, V. Peng, E.V. Name, G.V. Fontenay, J. Ngai, J.G. Conboy, B. Parvin, H.S. Feiler, T.P. Speed, J.W. Gray, C. Brennan, N.D. Socci, A. Olshen, B.S. Taylor, A. Lash, N. Schultz, B. Reva, Y. Antipin, A. Stukalov, B. Gross, E. Cerami, W.Q. Wang, L.-X. Qin, V.E. Seshan, L. Villafania, M. Cavatore, L. Borsu, A. Viale, W. Gerald, C. Sander, M. Ladanyi, C.M. Perou, D.N. Hayes, M.D. Topal, K.A. Hoadley, Y. Qi, S. Balu, Y. Shi, J. Wu, R. Penny, M. Bittner, T. Shelton, E. Lenkiewicz, S. Morris, D. Beasley, S. Sanders, A. Kahn, R. Sfeir, J. Chen, D. Nassau, L. Feng, E. Hickey, J. Zhang, J.N. Weinstein, A. Barker, D.S. Gerhard, J. Vockley, C. Compton, J. Vaught, P. Fielding, M.L. Ferguson, C. Schaefer, S. Madhavan, K.H. Buetow, F. Collins, P. Good, M. Guyer, B. Ozenberger, J. Peterson, E. Thomson, T.C.G.A.R. Network, T. source sites: Duke University Medical School, E. University, H.F. Hospital, M.D.A.C. Center, U. of California San Francisco, G. sequencing centres: Baylor College of Medicine, B.I. of M I T, Harvard, W.U. in St Louis, C. genome characterization centres: Broad Institute/Dana-Farber Cancer Institute, H.M.S.-F.C. Institute, J.H. of Southern California, H.I. University, L.B.N. Laboratory, M.S.-K.C. Center, C.H.U. of North Carolina, B.C. Resource, D.C. Center, P. teams: National Cancer Institute, N.H.G.R. Institute, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, *Nature* 455 (2008) 1061–1068, <http://dx.doi.org/10.1038/nature07385>.
- [35] M. Ceccarelli, F.P. Barthel, T.M. Malta, T.S. Sabedot, S.R. Salama, B.A. Murray, O. Morozova, Y. Newton, A. Radenbaugh, S.M. Pagnotta, S. Anjum, J. Wang, G. Manyam, P. Zoppoli, S. Ling, A.A. Rao, M. Grifford, A.D. Cherniack, H. Zhang, L. Poisson, C.G. Carloti, D.P. da Cunha Tirapelli, A. Rao, T. Mikkelsen, C.C. Lau, W.A. Yung, R. Rabadan, J. Huse, D.J. Brat, N.L. Lehman, J.S. Barnholtz-Sloan, S. Zheng, K. Hess, G. Rao, M. Meyerson, R. Beroukhi, L. Cooper, R. Akbani, M. Wrensch, D. Haussler, K.D. Aldape, P.W. Laird, D.H. Gutmann, S. Anjum, H. Arachchi, J.T. Auman, M. Balasundaram, S. Balu, G. Barnett, S. Baylin, S. Bell, C. Benz, N. Bir, K.L. Black, T. Bodenheimer, L. Boice, M.S. Bootwalla, J. Bowen, C.A. Bristow, Y.S. Butterfield, Q.-R. Chen, L. Chin, J. Cho, E. Chuah, S. Chudamani, S.G. Coetzee, M.L. Cohen, H. Colman, M. Couce, F. D'Angelo, T. Davidsen, A. Davis, J.A. Demchok, K. Devine, L. Ding, R. Duell, J.B. Elder, J.M. Eschbacher, A. Fehrenbach, M. Ferguson, S. Frazer, G. Fuller, J. Fulop, S.B. Gabriel, L. Garofano, J.M. Gastier-Foster, N. Gehlenborg, M. Gerken, G. Getz, C. Giannini, W.J. Gibson, A. Hadjipanayis, D.N. Hayes, D.I. Heiman, B. Hermes, J. Hilty, K.A. Hoadley, A.P. Hoyle, M. Huang, S.R. Jefferys, C.D. Jones, S.J. Jones, Z. Ju, A. Kastl, A. Kendler, J. Kim, R. Kucherlapati, P.H. Lai, M.S. Lawrence, S. Lee, K.M. Leraas, T.M. Lichtenberg, P. Lin, Y. Liu, J. Liu, J.Y. Ljubimova, Y. Lu, Y. Ma, D.T. Maglinte, H.S. Mahadeshwar, M.A. Marra, M. McGraw, C. McPherson, S. Meng, P.A. Mieczkowski, C.R. Miller, G.B. Mills, R.A. Moore, L.E. Mose, A.J. Mungall, R. Naresh, T. Naska, L. Neder, M.S. Noble, A. Noss, B.P. O'Neill, Q.T. Ostrom, C. Palmer, A. Pantazi, M. Parfenov, P.J. Park, J.S. Parker, C.M. Perou, C.R. Pierson, T. Pihl, A. Protopopov, A. Radenbaugh, N.C. Ramirez, W.K. Rathmell, X. Ren, J. Roach, A.G. Robertson, G. Saksena, J.E. Schein, S.E. Schumacher, J. Seidman, K. Senecal, S. Seth, H. Shen, Y. Shi, J. Shih, K. Shimmel, H. Siotte, S. Sifri, T. Silva, J.V. Simons, R. Singh, T. Skelly, A.E. Sloan, H.J. Sofia, M.G. Soloway, X. Song, C. Sougne, C. Souza, S.M. Staughtais, H. Sun, C. Sun, D. Tan, J. Tang, Y. Tang, L. Thorne, F.A. Trevisan, T. Triche, D.J.V.D. Berg, U. Veluvolu, D. Voet, Y. Wan, Z. Wang, R. Warnick, J.N. Weinstein, D.J. Weisenberger, M.D. Wilkerson, F. Williams, L. Wise, Y. Wolinsky, J. Wu, A.W. Xu, L. Yang, L. Yang, T.I. Zack, J.C. Zenklusen, J. Zhang, W. Zhang, J. Zhang, E. Zmuda, H. Noshmeh, A. Iavarone, R.G.W. Verhaak, Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma, *Cell* 164 (2016) 550–563, <http://dx.doi.org/10.1016/j.cell.2015.12.028>.
- [36] H. Jiang, X. Fei, H. Liu, K. Roeder, J. Lafferty, L. Wasserman, X. Li, T. Zhao, *huge: High-dimensional undirected graph estimation*, 2021, R package version 1.3.5.
- [37] C.M. Jarque, *Jarque-Bera test*, *Int. Encycl. Stat. Sci.* (2011).
- [38] B. Jin, Y. Li, K.D. Robertson, DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2 (2011) 607–617, <http://dx.doi.org/10.1177/1947601910393957>.
- [39] M.L. Johansen, L.C. Stetson, V. Vadmal, K. Waite, M.E. Berens, J.R. Connor, J. Lathia, J.B. Rubin, J.S. Barnholtz-Sloan, Gliomas display distinct sex-based differential methylation patterns based on molecular subtype, *Neuro- Oncol. Adv.* 2 (2020) vdaa002, <http://dx.doi.org/10.1093/oaajnl/vdaa002>.
- [40] O.A. Grant, Y. Wang, M. Kumari, N.R. Zabet, L. Schalkwyk, Characterising sex differences of autosomal DNA methylation in whole blood using the illumina EPIC array, *Clin. Epigenetics* 14 (2022) 62, <http://dx.doi.org/10.1186/s13148-022-01279-7>.
- [41] Y. an Chen, M. Lemire, S. Choufani, D.T. Butcher, D. Grafodatskaya, B.W. Zanke, S. Gallinger, T.J. Hudson, R. Weksberg, Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium HumanMethylation450 microarray, *Epigenetics* 8 (2013) 203–209, <http://dx.doi.org/10.4161/epi.23470>.
- [42] P. Danaher, JGL: Performs the joint graphical lasso for sparse inverse covariance estimation on multiple classes, 2018, R package version 2.3.1.
- [43] X. Ren, P.F. Kuan, methylGSA: a bioconductor package and shiny app for DNA methylation data length bias adjustment in gene set testing, *Bioinformatics* 35 (11) (2019) 1958–1959, <http://dx.doi.org/10.1093/bioinformatics/bty892>.
- [44] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Softw.* 33 (2010) 1–22, <http://dx.doi.org/10.18637/jss.v033.i01>.
- [45] E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* 53 (282) (1958) 457–481.
- [46] A. Kassambara, M. Kosinski, P. Biecek, *Survminer: Drawing survival curves using 'ggplot2'*, 2021, R package version 0.4.9.
- [47] R. Coletti, M.B. Lopes, Multi-omics data integration and network inference for biomarker discovery in glioma, in: N. Moniz, Z. Vale, J. Cascalho, C. Silva, R.S. ao (Eds.), *Progress in Artificial Intelligence*, Springer Nature, Switzerland, Cham, 2023, pp. 247–259.
- [48] M.S. Rooney, S.A. Shukla, C.J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity, *Cell* 160 (2015) 48–61, <http://dx.doi.org/10.1016/j.cell.2014.12.033>.
- [49] U. Bommhardt, B. Schraven, L. Simeoni, Beyond TCR signaling: Emerging functions of lck in cancer and immunotherapy, *Int. J. Mol. Sci.* 20 (2019) <http://dx.doi.org/10.3390/ijms20143500>.
- [50] A. Pant, L. Saleh, K. Patel, R. Medikonda, M. Lim, IMM-13. Efficacy of CXCR6 Blockade as a potentiator of anti-PD-1 therapy for the treatment of glioblastoma, *Neuro-Oncol.* 22 (2020) <http://dx.doi.org/10.1093/neuonc/noaa215.443>, ii107–ii107.
- [51] W. Wu, Q. Zhou, T. Masubuchi, X. Shi, H. Li, X. Xu, M. Huang, L. Meng, X. He, H. Zhu, S. Gao, N. Zhang, R. Jing, J. Sun, H. Wang, E. Hui, C.C. Wong, C. Xu, Multiple signaling roles of CD3e and its application in CAR-T cell therapy, *Cell* 182 (2020) 855–871.e23, <http://dx.doi.org/10.1016/j.cell.2020.07.018>.
- [52] X. Zheng, X. Liao, L. Nie, T. Lin, H. Xu, L. Yang, B. Shen, S. Qiu, J. Ai, Q. Wei, LCK and CD3E orchestrate the tumor microenvironment and promote immunotherapy response and survival of muscle-invasive bladder cancer patients, *Front. Cell Dev. Biol.* 9 (2021).
- [53] M. Mitsdoerffer, L. Aly, M. Barz, T. Engleitner, C. Sie, C. Delbridge, G. Lepennetier, R. Öllinger, M. Pfaller, B. Wiestler, R. Rad, B. Meyer, B. Knier, F. Schmidt-Graf, J. Gempt, T. Korn, The glioblastoma multiforme tumor site promotes the commitment of tumor-infiltrating lymphocytes to the TH17 lineage in humans, *Proc. Natl. Acad. Sci.* 119 (2022) e2206208119, <http://dx.doi.org/10.1073/pnas.2206208119>, doi: 10.1073/pnas.2206208119.
- [54] L. Li, X. Liu, X. Ma, X. Deng, T. Ji, P. Hu, R. Wan, H. Qiu, D. Cui, L. Gao, Identification of key candidate genes and pathways in glioblastoma by integrated bioinformatical analysis, *Exp. Ther. Med.* (2019) <http://dx.doi.org/10.3892/etm.2019.7975>.
- [55] Y. Li, J. Li, H. Chen, B. Lu, F. Lu, H. Chen, H. Liu, C. Qian, TCAF2 is associated with the immune microenvironment, promotes pathogenesis, and impairs prognosis in glioma, *Gene* 883 (2023) 147667, <http://dx.doi.org/10.1016/j.gene.2023.147667>.
- [56] R. Coletti, M.L. de Mendonça, S. Vinga, M.B. Lopes, Inferring diagnostic and prognostic gene expression signatures across WHO glioma classifications: A network-based approach, *Bioinform. Biol. Insights* 18 (2024) 11779322241271535, <http://dx.doi.org/10.1177/11779322241271535>.
- [57] H. Li, Q. Liu, Z. Chen, M. Wu, C. Zhang, J. Su, Y. Li, C. Zhang, Hsa_circ_0110757 upregulates ITGA1 to facilitate temozolomide resistance in glioma by suppressing hsa-miR-1298-5p, *Cell Death Dis.* 12 (2021) 252, <http://dx.doi.org/10.1038/s41419-021-03533-x>.
- [58] W. Yin, H. Zhu, J. Tan, Z. Xin, Q. Zhou, Y. Cao, Z. Wu, L. Wang, M. Zhao, X. Jiang, C. Ren, G. Tang, Identification of collagen genes related to immune infiltration and epithelial-mesenchymal transition in glioma, *Cancer Cell Int.* 21 (2021) 276, <http://dx.doi.org/10.1186/s12935-021-01982-0>.
- [59] C. Edlich, G. Stier, B. Simon, M. Sattler, C. Muhle-Goll, Structure and phosphatidylinositol-(3,4)- bisphosphate binding of the C-terminal PH domain of human pleckstrin, *Structure* 13 (2005) 277–286, <http://dx.doi.org/10.1016/j.str.2004.11.012>, doi: 10.1016/j.str.2004.11.012.
- [60] X. Wang, L.B. Hills, Y.H. Huang, Lipid and protein co-regulation of PI3K effectors Akt and Itk in lymphocytes, *Front. Immunol.* 6 (2015) <http://dx.doi.org/10.3389/fimmu.2015.00117>.
- [61] M. Lenoir, I. Kufareva, R. Abagyan, M. Overduin, Membrane and protein interactions of the pleckstrin homology domain superfamily, *Membranes* 5 (2015) 646–663, <http://dx.doi.org/10.3390/membranes5040646>.
- [62] A. Varone, D. Spano, D. Corda, Shp1 in solid cancers and their therapy, *Front. Oncol.* 10 (2020).
- [63] L. Sooman, S. Ekman, G. Tsakonas, A. Jaiswal, S. Navani, P.-H. Edqvist, F. Pontén, S. Bergström, M. Johansson, X. Wu, E. Blomquist, M. Bergqvist, J. Gullbo, J. Lennartsson, PTPN6 expression is epigenetically regulated and influences survival and response to chemotherapy in high-grade gliomas, *Tumor Biol.* 35 (2014) 4479–4488, <http://dx.doi.org/10.1007/s13277-013-1590-5>.

- [64] Q.-W. Fan, W.A. Weiss, Targeting the RTK-PI3K-mTOR Axis in Malignant Glioma: Overcoming Resistance, Springer, Berlin, Heidelberg, 2011, pp. 279–296, http://dx.doi.org/10.1007/82_2010_67.
- [65] S. Venkatesan, M.L.M. Lamfers, C.M.F. Dirven, S. Leenstra, Genetic biomarkers of drug response for small-molecule therapeutics targeting the RTK/Ras/PI3K, p53 or Rb pathway in glioblastoma, *CNS Oncol.* 5 (2016) 77–90, <http://dx.doi.org/10.2217/cns-2015-0005>.
- [66] B. Huang, W. Pan, W. Wang, Y. Wang, P. Liu, W. Geng, Overexpression of pleckstrin homology domain-containing family a member 4 is correlated with poor prognostic outcomes and immune infiltration in lower-grade glioma, in: M. Malaguarnera (Ed.), *Dis. Markers* 2022 (2022) 1292648, <http://dx.doi.org/10.1155/2022/1292648>.
- [67] K.E. Witte, C. Slotta, M. Lütkemeyer, A. Kitke, R. Coras, M. Simon, C. Kaltschmidt, B. Kaltschmidt, PLEKHG5 regulates autophagy, survival and MGMT expression in U251-MG glioblastoma cells, *Sci. Rep.* 10 (2020) 21858, <http://dx.doi.org/10.1038/s41598-020-77958-3>.
- [68] X. Han, Y. Mei, R.K. Mishra, H. Bi, A.D. Jain, G.E. Schiltz, B. Zhao, M. Sukhanova, P. Wang, A.A. Grigorescu, P.C. Weber, J.J. Piwinski, M.A. Prado, J.A. Paulo, L. Stephens, K.E. Anderson, C.S. Abrams, J. Yang, P. Ji, Targeting pleckstrin-2/Akt signaling reduces proliferation in myeloproliferative neoplasm models, *J. Clin. Investig.* 133 (2023) <http://dx.doi.org/10.1172/JCI159638>.
- [69] M.L. Woods, A. Weiss, A.M. Sokol, J. Graumann, T. Boettger, A.M. Richter, R.T. Schermuly, R.H. Dammann, Epigenetically silenced apoptosis-associated tyrosine kinase (AATK) facilitates a decreased expression of cyclin D1 and WEE1, phosphorylates TP53 and reduces cell proliferation in a kinase-dependent manner, *Cancer Gene Ther.* (2022) <http://dx.doi.org/10.1038/s41417-022-00513-x>.
- [70] M.J. Duffy, N. O'Donovan, D.J. Brennan, W.M. Gallagher, B.M. Ryan, Survivin: A promising tumor biomarker, *Cancer Lett.* 249 (2007) 49–60, <http://dx.doi.org/10.1016/j.canlet.2006.12.020>.
- [71] L. Xu, W. Yu, H. Xiao, K. Lin, BIRC5 is a prognostic biomarker associated with tumor immune cell infiltration, *Sci. Rep.* 11 (2021) 390, <http://dx.doi.org/10.1038/s41598-020-79736-7>.
- [72] P. Gautam, S.C. Nair, M.K. Gupta, R. Sharma, R.V. Polisetty, M.S. Uppin, C. Sundaram, A.K. Puligopu, P. Ankathi, A.K. Purohit, G.R. Chandak, H.C. Harsha, R. Sirdeshmukh, Proteins with altered levels in plasma from glioblastoma patients as revealed by iTRAQ-based quantitative proteomic analysis, *PLoS One* 7 (2012) e46153, <http://dx.doi.org/10.1371/journal.pone.0046153>.
- [73] H. Xu, Q. Zhu, L. Tang, J. Jiang, H. Yuan, A. Zhang, M. Lou, Prognostic and predictive value of FCER1G in glioma outcomes and response to immunotherapy, *Cancer Cell Int.* 21 (2021) 103, <http://dx.doi.org/10.1186/s12935-021-01804-3>.
- [74] X. ping Zhong, A. Kan, Y. hong Ling, L. he Lu, J. Mei, W. Wei, S. hua Li, R. ping Guo, NCKAP1 improves patient outcome and inhibits cell growth by enhancing Rb1/p53 activation in hepatocellular carcinoma, *Cell Death Dis.* 10 (2019) 369, <http://dx.doi.org/10.1038/s41419-019-1603-4>.
- [75] J.K. Cowell, Y. Teng, N.G. Bendzunas, R. Ara, A.S. Arbab, E.J. Kennedy, Suppression of breast cancer metastasis using stapled peptides targeting the WASF regulatory complex, *Cancer Growth Metastasis* 10 (2017) 1179064417713197, <http://dx.doi.org/10.1177/1179064417713197>.
- [76] M.E. Lomakina, F. Lallemand, S. Vacher, N. Molinie, I. Dang, W. Cacheux, T.A. Chipysheva, V.D. Ermilova, L. de Koning, T. Dubois, I. Bièche, A.Y. Alexandrova, A. Gautreau, Arpin downregulation in breast cancer is associated with poor prognosis, *Br. J. Cancer* 114 (2016) 545–553, <http://dx.doi.org/10.1038/bjc.2016.18>.
- [77] N. Kastendiek, R. Coletti, T. Gross, M.B. Lopes, Exploring glioma heterogeneity through omics networks: from gene network discovery to causal insights and patient stratification, *BioData Min.* 17 (2024) 56, <http://dx.doi.org/10.1186/s13040-024-00411-y>.
- [78] W. Saelens, R. Cannoodt, Y. Saeys, A comprehensive evaluation of module detection methods for gene expression data, *Nat. Commun.* 9 (2018) 1090, <http://dx.doi.org/10.1038/s41467-018-03424-4>.
- [79] P. Lecca, A. Re, Detecting modules in biological networks by edge weight clustering and entropy significance, *Front. Genet.* 6 (2015) <http://dx.doi.org/10.3389/fgene.2015.00265>.
- [80] W. Hwang, Y.-R. Cho, A. Zhang, M. Ramanathan, A novel functional module detection algorithm for protein-protein interaction networks, *Algorithms Mol. Biol.* 1 (2006) 24, <http://dx.doi.org/10.1186/1748-7188-1-24>.
- [81] S. Tripathi, S. Moutari, M. Dehmer, F. Emmert-Streib, Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules, *BMC Bioinformatics* 17 (2016) 129, <http://dx.doi.org/10.1186/s12859-016-0979-8>.