Kastendiek *et al. BioData Mining* (2024) 17:56
https://doi.org/10.1186/s13040-024-00411-y

BioData Mining

RESEARCH Open Access

Exploring glioma heterogeneity through omics networks: from gene network discovery to causal insights and patient stratification

Nina Kastendiek[1], Roberta Coletti[2], Thilo Gross[1,3,4] and Marta B. Lopes[2,5*]

*Correspondence: marta.lopes@fct.unl.pt

[1] Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg 26129, Germany
[2] Center for Mathematics and Applications (NOVA Math), NOVA School of Science and Technology (NOVA FCT), Caparica 2829-516, Portugal
[3] Helmholtz Institute for Functional Marine Biodiversity (HIFMB), Oldenburg 26129, Germany
[4] Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven 27570, Germany
[5] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology (NOVA FCT), Caparica 2829-516, Portugal

Abstract

Gliomas are primary malignant brain tumors with a typically poor prognosis, exhibiting significant heterogeneity across different cancer types. Each glioma type possesses distinct molecular characteristics determining patient prognosis and therapeutic options. This study aims to explore the molecular complexity of gliomas at the transcriptome level, employing a comprehensive approach grounded in network discovery. The graphical lasso method was used to estimate a gene co-expression network for each glioma type from a transcriptomics dataset. Causality was subsequently inferred from correlation networks by estimating the Jacobian matrix. The networks were then analyzed for gene importance using centrality measures and modularity detection, leading to the selection of genes that might play an important role in the disease. To explore the pathways and biological functions these genes are involved in, KEGG and Gene Ontology (GO) enrichment analyses on the disclosed gene sets were performed, highlighting the significance of the genes selected across several relevent pathways and GO terms. Spectral clustering based on patient similarity networks was applied to stratify patients into groups with similar molecular characteristics and to assess whether the resulting clusters align with the diagnosed glioma type. The results presented highlight the ability of the proposed methodology to uncover relevant genes associated with glioma intertumoral heterogeneity. Further investigation might encompass biological validation of the putative biomarkers disclosed.

Keywords: Cancer omics data, Graphical lasso, Network centrality, Causal networks, Modularity detection, Spectral clustering

Background

Gliomas are primary malignant brain tumors with a typically bad prognosis, accounting for 80% of malignancies in the brain [1]. Due to their large heterogeneity, gliomas encompass different cancer types, where each type possesses distinct characteristics that influence patient prognosis and therapeutic options. The World Health Organization (WHO) Classification of Tumors of the Central Nervous System (CNS) has been

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

changing throughout the years. In the most recent version of the WHO CNS tumor classification [2], the glioma types are classified mainly based on the sample's molecular profiles instead of histological features. Gliomas in adults are characterized by three types: astrocytoma, oligodendroglioma, and glioblastoma [2]. The latter exhibits a wildtype status for the IDH genes, and it is the most common and aggressive type, with a median survival of only 14 months [3]. Conversely, oligodendrogliomas and astrocytomas present mutations in at least one IDH gene, being molecularly distinguished by the presence/absence of the combined loss of the short arm chromosome 1 and the long arm of chromosome 19 (1p/19q codeletion), respectively [2].

Studying gliomas on a molecular level is essential for understanding the biological mechanisms behind these brain tumors, ultimately advancing cancer medicine. Omics data comprise various layers of biological information, e.g. genomics (DNA sequences), transcriptomics (RNA transcripts), proteomics (protein abundances), methylomics (DNA methylation level), and metabolomics (metabolite profiles). This type of data is increasingly available with technological advances, e.g., through high-throughput RNA sequencing (RNA-seq) and protein mass spectrometry. Typically, omics data is a matrix of variables (e.g. genes) and samples. In gene expression data, for example, each entry represents the expression level of a particular gene in a particular sample. These datasets are often high-dimensional, making it challenging to derive meaningful insights.

Network science is highly recognized in the field of cancer research to study the complexity of diseases and extract relevant biological information [4–6]. Biological networks help to explain the disease by studying the interactions between biological entities such as genes. For example, genomics mutations data is analyzed by leveraging biological interaction networks to discover genetic drivers in cancer [7–9]. To identify altered pathways, differential network analysis examines differences between group-specific molecular interaction networks, e.g. between cancer and healthy groups [10, 11]. From transcriptomics data, gene co-expression networks can be constructed, where the nodes represent genes that are connected with an edge if the corresponding genes are significantly co-expressed in the RNA-seq dataset across different samples. Inferring and analyzing these networks for specific cancer types can uncover key network genes as potential biomarkers for diagnostic and therapeutic purposes [12–14].

Network inference methods attempt to identify dependencies between variables. The simplest way to construct a network from omics data is to use pairwise association measures, such as Pearson's correlation coefficient, which then represent the edge weights in the graph. However, this approach cannot distinguish between direct and indirect effects, resulting in dense and poorly interpretable networks. To overcome this, conditional dependencies, or partial correlations, can be employed that capture only direct interactions between two variables by taking into account the influence of all other variables so that that indirect effects do not create a link in the network. The most commonly used are Gaussian Graphical Models (GGMs), which assume normally distributed variables. The network structure is inferred from the inverse covariance matrix, representing direct dependencies [15]. In omics data, the number of samples is often much smaller than the number of variables, making statistical modelling difficult due to the risk of overfitting. The graphical lasso method estimates the structure of an undirected GGM by applying a lasso penalty to the inverse covariance matrix [15, 16].

Another approach to GGM estimation is based on Bayesian methods, which can also be used to construct directed causal networks [15]. Learning Bayesian networks typically involves algorithms that explore possible network structures and identify the most probable structure from the data [17]. To get reliable results, this method is often combined with constraints from prior biological knowledge [15]. A key property of Bayesian Networks is that they are acyclic, which can be problematic for modeling gene networks, as these may contain feedback loops [18].

This study aims to investigate the molecular complexity of gliomas at the transcriptome level. We developed a comprehensive network-based framework that integrates network discovery and analysis to reveal causal relationships. The goal is to identify potential biomarkers or pathways that could improve our understanding of glioma heterogeneity. For this, we utilize the graphical lasso method to discover gene co-expression networks as it introduces sparsity to the network structure and does not rely on any additional information. To infer causal networks, we estimate the Jacobian matrix of the underlying biological system by a method that allows to infer causality by using the covariance from sample data and information about the network structure [19, 20]. The Jacobian matrix of a dynamical system describes how the system responds to small perturbations, here showing how small changes in the expression of one gene affect the expression of other genes. The estimated Jacobian matrix provides insights into the directional influences among genes, from which we construct a directed network. In this work, we explore the promising combination of the graphical lasso method, which provides insights on the underlying network structures, with Jacobian estimation to infer potential causal relationships in gene expression data.

Identifying important groups of genes in these networks can help uncover biomarkers and explore new treatment options aimed at these genes. In network analysis, centrality measures rank nodes based on their connections and position in the network. These metrics can identify genes of biological importance in the networks [21, 22], with each centrality measure reflecting a different notion of a gene's role in the network. For example, the degree, or strength in weighted networks, is the sum of edges connected to a node [23]. In gene networks, well-connected genes may play a central role in biological processes. Eigenvector centrality extends this concept by considering not only the direct connections a node has but also the importance of the nodes it is connected to [24], highlighting genes that exert a broader influence over the network rather than just locally within their immediate neighbors. Closeness centrality calculates how close a node is to all other nodes in the network, by measuring the shortest path lengths between all pairs of nodes [25]. Genes with high closeness centrality can interact quilckly with other genes across the network, suggesting potential roles in regulatory or signaling pathways. The betweenness centrality, on the other hand, quantifies the extent to which a node lies on the shortest paths between other nodes [25], where a high betweenness centrality indicates that a gene acts as a bridge between different groups of related genes in the network. We focus on strength, eigenvector, and closeness centrality because these measures emphasize different aspects of a gene's connectivity and influence within the network, such as the number and quality of connections and proximity to other nodes. Betweenness centrality captures a gene's role as an intermediary between different parts of the network, which refers to a property that is not central to our analysis. Community

or modularity detection in biological networks are crucial for uncovering functional relationships between genes, by grouping together nodes that reflect shared roles or behaviors within the network. Stochastic Block Models (SBMs) offer a probabilistic approach by fitting a model to the existing network structure to recover the community structure in the graph [26, 27].

Exploring modularity, hubs and shortest paths in biological networks is crucial for understanding their structure and function [28]. We use strength, eigenvector and closeness centrality as well as modularity detection by SBMs to identify groups of important genes in the networks for glioblastoma, astrocytoma and oligodendroglioma. To explore the pathways and functions these genes are involved in, we perform KEGG (Kyoto Encyclopedia of Genes and Genomes) and Gene Ontology (GO) enrichment analysis on the selected genes in each glioma type. Enrichment analysis helps understanding the biological significance of gene sets. While KEGG provides a resource for linking genes to metabolic and signaling pathways, GO categorizes genes based on their molecular functions and processes.

The use of these tools allows us to determine whether the genes identified in our study are already known as potential drivers of other cancers or diseases, and if they can be considered therapeutic targets. Beside this, we also explore whether they can capture the heterogeneity between glioma types as potential diagnostic biomarkers. As patients may respond differently to treatments, patient stratification is critical for precision medicine approaches, ensuring that patients with similar characteristics receive the most effective treatments, adapted to their specific type of tumor. For grouping the patients, we build on another type of network that can be constructed from omics data. In patient similarity networks, patients are represented as nodes and the similarities between them as edges connecting the nodes. These networks are used for clustering patients into groups with similar molecular characteristics, which may be indicative of common underlying biological mechanisms in a disease. We employ spectral clustering of patient similarity networks based on the gene expression data of the selected genes, and evaluate how much the resulting clusters are in agreement with the diagnosed glioma types.

In a previous study, Martins et al. [29] explored the capability of applying network discovery and clustering techniques to glioma data. Glioma patients were grouped according to 2016-WHO classification guidelines, and undirected networks were used to perform variable selection before applying the K-means clustering. The results pointed out some inconsistencies between clusters and 2016-WHO classes, fostering further studies based on the new 2021-WHO diagnostic label assignments, which we use here.

This study presents a comprehensive network-based approach to glioma patient data to investigate the complex gene expression profiles in glioma patient data, leading to the identification of potentially important genes and patient stratification. For this, RNA-seq glioma data from The Cancer Genomics Atlas (TCGA) were used, with patients grouped according to the updated 2021-WHO glioma types [30]. From these vast datasets, the variables (i.e., genes) were preselected using the corresponding TCGA proteomics dataset containing a set of proteins in the major biochemical pathways in cancer, i.e., the genes encoding for the proteins present in the proteomics dataset were retained for further analysis. A gene co-expression network for each tumor type was constructed using the graphical lasso method to reveal the interplay

of genes in each glioma type. Next, the obtained network structure was used to infer causality among genes through the Jacobian matrix estimation, a crucial step providing insights into the direction of gene interactions. Centrality measures and modularity detection were then applied to the constructed networks to identify central genes as potential biomarkers of each glioma type. As a final step, spectral clustering was applied to patient similarity networks based on the selected variables from the gene networks to stratify patients into distinct groups and evaluate whether these patient groups matched the clinically diagnosed glioma types.

Materials and methods

Gene co-expression networks via graphical lasso

The graphical lasso is a statistical method for estimating the graphical structure in a Gaussian graphical model. It is particularly useful in high-dimensional scenarios where the number of variables is higher than the number of samples, as it induces sparsity in the model. The objective is to find a sparse graph that represents the conditional independence structure among the variables by applying a lasso penalty to the inverse covariance matrix. Suppose n multivariate normal observations of dimension p, and let $\mathbf{\Sigma}$ and $\mathbf{S}$ be the empirical and theoretical covariance matrix, respectively. The graphical lasso, first implemented by Friedman et al. [16], estimates the inverse covariance matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$ by solving the optimization problem

$$\max_{\mathbf{\Theta}} \left(\log\det \mathbf{\Theta} - \mathrm{tr}(S\mathbf{\Theta}) - \lambda||\mathbf{\Theta}||_1\right),$$

where λ is the regularization parameter and $||\mathbf{\Theta}||_1$ is the L1 norm of $\mathbf{\Sigma}^{-1}$ (sum of the absolute values of all its elements).

If element ij in the estimated inverse covariance matrix $\mathbf{\Sigma}^{-1}$ is zero, then variables i and j are conditionally independent [16]. The non-zero elements quantify the sign and strength of the direct relationship between the corresponding pair of variables, while controlling for the influence of all other variables. The inverse covariance matrix thus provides information about the partial correlation structure of the nodes [15].

To avoid overfitting when the number of variables is higher than the number of samples, the regularization term induces sparsity in the model by pushing entries in the inverse covariance matrix to zero. A larger value of λ enforces more sparsity. Here, the Stability Approach to Regularization Selection (StARS) is used for selecting the regularization parameter λ. StARS measures the stability of network topology in the estimated graphical model across different subsamples of the data. The idea is that the graph structure should be stable if the chosen regularization parameter is appropriate. The method repeatedly draws subsamples from the dataset, fits a graphical lasso model for each subsample, and tracks how often each edge appears across the subsamples. For each λ, StARS calculates the variability in edge selection across the subsamples and selects the smallest value of λ such that the instability remains below a threshold. In this way λ is optimized to use the minimum necessary regularization to ensure the network's reproducibility under random sampling [31].

Causal discovery by Jacobian reconstruction

Inferring causality from observed correlations, i.e., moving from undirected to directed networks, is a key challenge. In omics studies, it holds the promise to unveil relevant links driving tumorigenesis. When a system is subject to some noise, it is possible to employ the theory of stochastic processes showing that correlations emerging from samples can be interpreted as a 'signature' of the underlying deterministic system [19]. This leads to a relationship between the covariance of the samples and the system's Jacobian matrix, which is derived according to Steuer et al. [19] and Barter et al. [20] as follows.

The Jacobian matrix of a dynamical system describes how the system responds to perturbations. Given a system of N variables defined by $\boldsymbol{X} = (X_1, ..., X_N)^{\mathrm{T}}$, their response to small fluctuations around an equilibrium can be approximated as

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X} = \mathbf{J}\boldsymbol{X},$$

with the Jacobian $\mathbf{J}$ of dimension N. The system with noise can be modeled by the Langevin-type equation

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \sum_j J_{ij}X_j + \sqrt{2D_i}\xi_i(t),$$

where $\xi_i(t)$ is Gaussian white noise, with zero mean and unit variance and D_i is the mean amplitude of the fluctuations of X_i. By using the corresponding stationary Fokker-Planck equation for the probability distribution, the Lyapunov equation is obtained [32]

$$\mathbf{J}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\mathbf{J}^T = -2\mathbf{D},$$

where $\boldsymbol{\Gamma}$ is the covariance matrix with entries $\Gamma_{ij} = \langle X_i X_j \rangle$ and $\mathbf{D}$ is the fluctuation matrix, describing the internal noise of the system.

Based on this systematic relationship, it is possible to recover the deterministic system by inferring the system's Jacobian matrix from the sample data, if additional knowledge about the system is available [19].

Note that the Jacobian matrix is generally non-symmetric, whereas the covariance matrix is always symmetric. This leads to a number of constraints in the Lyapunov equation that is lower than the degrees of freedom of the Jacobian, making it impossible to determine causality from correlation alone. However, if additional information, such as the topology of interactions, is available, the full Jacobian can be reconstructed. Barter et al. [20] proposed an algorithm to solve the Lyapunov equation for $\mathbf{J}$, by considering additional constraints. Specifically, it is needed to set $J_{ij} = 0$ for $N(N-1)/2$ pairs of nodes i and j, forcing the absence of the corresponding variable relations [20].

Biomarker selection through node centrality

Centrality measures in network analysis are used to identify the most important nodes within a network. There are various measures that each offer different perspectives on

a node's importance based on its connections and position within the network structure. Here, the strength, eigenvector and closeness centralities are used, which are both calculated from a non-negative adjacency matrix $\mathbf{A}$ describing relations among a set of N nodes with centralities $c_1, ..., c_N$.

The strength of a node is quantified by summing the weights of the edges the node i is connected to [23]

$$c_i = \sum_j A_{ij}.$$

With strength centrality, it is presumed that a node with significant or several interactions has a higher level of importance in the network.

In the eigenvector centrality, the sum of a node's connections is calculated, whereby each connection is weighted by the other node's centrality

$$\lambda c_i = \sum_j A_{ij} c_j,$$

where λ is a constant required so that the equations have a non-zero solution. This equation can be expressed in matrix notation

$$\lambda \boldsymbol{c} = \mathbf{A}\boldsymbol{c},$$

where λ and $\boldsymbol{c}$ represent, an eigenvalue and the associated eigenvector of the matrix $\mathbf{A}$, respectively. Among the possible eigenvalues of $\mathbf{A}$, the leading one (i.e., the eigenvalue with the largest real part) is considered, since its eigenvector is guaranteed to have positive entries [33]. Then, the N components of the leading eigenvector represent the centralities of each node. In this concept of centrality, the importance of a node is proportional to the importance of its neighboring nodes, creating a recursive relationship where connections to highly central nodes enhance a node's own centrality [24].

Closeness centrality measures the proximity of a node to all other nodes in a network. Let $\mathrm{d}(j, i)$ be the shortest path distance between node i and j, then the closeness centrality of node i is defined by

$$c_i = \frac{1}{\sum_{j \neq i} \mathrm{d}(j, i)}.$$

Nodes with high closeness centrality can quickly interact with all others due to shorter path lengths [25]. In networks where the edges have weights corresponding to the strength of relationships, the distances can be represented as the reciprocal of the weights $d_{ij} = 1/A_{ij}$ so that higher weights (stronger relationships) translate into shorter distances.

In directed networks, it must be distinguished between outward and inward centrality. Outward centrality quantifies the influence of a given node over the network, while inward centrality reflects how much a node is influenced by the others [34]. With an adjacency matrix where rows reflect the out-degrees and columns the in-degrees of nodes, outward strength centrality is calculated by summing over the rows and inward strength centrality is computed by summing over columns. For eigenvector centrality,

outward centrality calculates the right eigenvector of $\mathbf{A}$, satisfying $\lambda \boldsymbol{c} = \mathbf{A}\boldsymbol{c}$ while inward centrality calculates the left eigenvector, satisfying $\lambda \boldsymbol{c}^t = \boldsymbol{c}^t \mathbf{A}$ [34]. Consequently, outward centrality can be calculated similarly to undirected networks and inward centrality is calculated based on the transpose of $\mathbf{A}$. For closeness centrality in directed networks, out-closeness is calculated by measuring the paths from node i to all other nodes and in-closeness is calculated by considering paths from all other nodes to node i. For this study, outward centrality was calculated, following the notion that genes that exert a high influence over the network could be potentially important drivers in the disease process.

Modularity detection by Stochastic Block Models

In modularity detection, the task is to organize a network into modules, or communities, by grouping nodes that have a similar role in the network. This can be regarded as a way to select groups of functionally related genes instead of focusing only on single nodes. Stochastic Block Models (SBMs) provide a probabilistic approach for this purpose [26].

An SBM is a generative model that is based on a structure of N nodes divided into B groups. The partition is given by a vector $\boldsymbol{b}$, where each entry $b_i \in \{1, ..., B\}$ defines the group membership of node i. Given the partition of the nodes $\boldsymbol{b}$, the idea is to generate a network in which the nodes are grouped accordingly, by maximizing the probability $P(\mathbf{A}|\boldsymbol{b})$, where $\mathbf{A}$ is the adjacency matrix. The probability $P(\mathbf{A}|\boldsymbol{b})$ determines the likelihood of edges existing between nodes based on their group memberships. With a Bayesian inference approach, we can revert the problem, and infer the most probable modularity structure that could have generated an observed network, by computing $P(\boldsymbol{b}|\mathbf{A})$ through the Bayesian rule [27]. The number of groups B and the most likely partition $\boldsymbol{b}$ can be computed in combination by optimizing the Integrated Classification Likelihood [35].

SBMs lead to relatively homogeneous degree distributions within modules, which might not always align with networks estimated from real-world data. To address this limitation, Degree-corrected SBMs (DcSBMs) account for variations in node degrees within the same module [27].

Spectral clustering of patient similarity networks

Spectral clustering is a powerful method for grouping complex data into distinct clusters by partitioning a similarity graph. It is particularly useful for clustering samples based on high-dimensional omics data. In a patient similarity network, the samples to be clustered are regarded as nodes of a graph. The edges between the nodes represent the similarities between the samples, which are typically evaluated using the Euclidean distance across all variables. Taking the nearest neighbor graph to capture the local structure of the data is especially important for detecting non-linear structures in the data.

Spectral clustering captures the manifold structure of complex data structures through the eigendecomposition of the graph's Laplacian matrix. The eigenvectors corresponding to the smallest non-zero eigenvalues form a new, lower-dimensional, feature space where the data points are separated into distinct clusters [36]. John et al. [37] proposed the *Spectrum* method for spectral clustering of complex omics data, implemented as an R package. It constructs the patient similarity matrix from the omics data, computes the

nearest neighbor graph, reduces noise to better reveal underlying structures, and finally performs the spectral clustering. For more details of the method, refer to [37].

Data and analysis workflow

The transcriptomics and proteomics data were retrieved from GDAC Broad Firehose (https://gdac.broadinstitute.org), a portal collecting data generated by The Cancer Genome Atlas (TCGA) [38]. In particular, we considered normalized RNA-seq and protein abundance data from the TCGA-GBM and TCGA-LGG projects [39–41]. These datasets group the glioma patients into glioblastoma and lower-grade glioma (LGG) in line with the 2007 WHO classification [42]. LGGs aggregate astrocytoma, oligodendroglioma, and oligoastrocytoma samples. The dataset was updated to align with the 2021 WHO classification, thereby reallocating oligoastrocytoma cases to one of the three defined glioma types [30].

The transcriptomics data, derived from RNA sequencing, provides the gene expression levels of over 20,000 genes. The proteomics data, obtained through reverse phase protein arrays (RPPA), is a functional proteomics dataset accounting for nearly 200 proteins involved in major biochemical signaling pathways in cancer [43]. For the computational feasibility of the methods employed in this study, a dimensionality reduction of the transcriptomics dataset was necessary. Therefore, the analysis was specifically focused on the expression levels of those genes from the transcriptomics dataset that encode the proteins contained in the proteomics dataset, according to the map provided by the GDAC Broad Firehose portal. This resulted in a significant variable reduction while still ensuring that important genes involved in regulating cancer pathways were included in the analysis. This subset of the transcriptomics dataset (transcriptomics$_S$) comprises 145 RNA-seq variables measured over 206 samples for glioblastoma, 255 samples for astrocytoma, and 166 samples for oligodendroglioma. The reduced dataset was normalized by a high-dimensional Gaussian copula with non-parametric marginals (nonparanormal), which transforms the variable distribution to achieve normality [44]. The nonparanormal normalization was performed by the *huge.npn* R function. For the network inference process, we used only the transformed variables that were normally distributed in all glioma types according to the Jarque-Bera test (*jarque.test* function from the *moments* R package).

A preliminary spectral clustering using the *Spectrum* method was performed to get insights into the biological information contained in the reduced transcriptomics$_S$ dataset compared to the full transcriptomics dataset and the proteomics dataset. The *Spectrum* R tool was employed with a fixed cluster number set to $c = 3$ to match the three glioma types, i.e., glioblastoma, astrocytoma, and oligodendroglioma. The data is handled automatically and no additional parameters are necessary.

In the methodological workflow, we started with network discovery to infer association and causal gene networks for the three glioma types, using the graphical lasso method and the Jacobian matrix estimation, respectively. Figure 1 describes the analysis workflow. The undirected networks were inferred from the transcriptomics data for each glioma type separately by the graphical lasso method using the *huge* R function. The StARS method was employed using the *huge.select* R function to determine the optimal regularization parameter λ for each of the three networks. The resulting values for

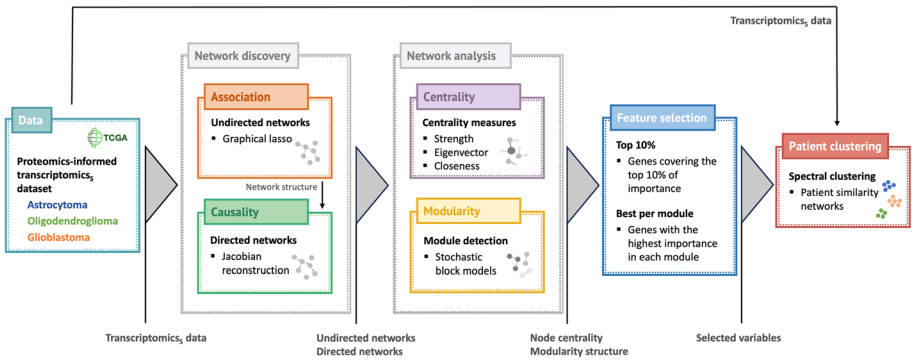


Fig. 1 Schematic representation of the workflow encompassing network discovery, network analysis, feature selection and patient clustering applied to the TCGA glioma transcriptomics$_S$ dataset (transcriptomics$_S$ stands for the proteomics-informed transcriptomics dataset)

λ were closely aligned across all types, with $\lambda \simeq 0.3$, leading to a similar level of sparsity in the graphs. Based on the inverse covariance matrices, undirected gene co-expression networks were constructed for each glioma type, providing information about the partial correlation structure of the gene relations at the transcriptomics level. Owing to the induced sparsity, some genes had no connections, being excluded for further analysis.

To infer causality among gene relations, the Jacobian matrix was computed for each glioma type, following the algorithm of [20]. The fluctuation matrix describing the internal noise of the system is assumed diagonal with $D_i = \frac{\sigma_i^2}{2}$, for $i = 1, \ldots, N$, where σ_i^2 are the variances of the individual variables [32]. The structural information needed as additional constraints was retrieved from the sparsity of the gene co-expression networks previously estimated. The zero entries in the inverse covariance matrix $\Sigma_{ij}^{-1} = 0$ defined the zero entries in the Jacobian matrix $J_{ij} = 0$. For each glioma type, a directed network was deduced based on the respective Jacobian matrix, describing how one gene influences another. Self-loops and the sign of the edge weights were ignored for further analysis since the identification of important genes requires focusing only on the strength of the relationship between the genes.

The next step of the workflow was the analysis of the networks inferred, based on centrality measures and modularity detection. Strength, eigenvector and closeness centrality were calculated to determine the importance of the genes in the glioma networks, while the modularity structure was computed to reveal functionally related groups of genes that may be important. For modularity detection via DcSBMs, the *greed* R function was used. If the adjacency matrix of the network is supplied, the function computes the most likely modularity structure obtained by DcSBM [35].

In the proposed methodology, relevant genes were selected from the undirected and directed networks, by identifying the genes covering the top 10% of importance in the networks and the most important gene from each module, based on strength, eigenvector and closeness centrality. This leads to eight sets of potential biomarkers for each glioma type.

We used the *enrichKEGG* and *enrichGO* functions from the *clusterProfiler* package in R to perform KEGG and GO enrichment analysis on the selected genes in each glioma type. These functions take a list of gene IDs and return pathways or GO terms that are

significantly enriched. KEGG and GO enrichment can be specified for Homo sapiens through the arguments *organism* =*"hs"* and *OrgDb* = *org.Hs.eg.db*, respectively. Both *enrichKEGG* and *enrichGO* functions allow for False Discovery Rate (FDR) control via the Benjamini-Hochberg procedure (*pAdjustMethod* = *"BH"*), which adjusts p-values for multiple testing. To retrieve Entrez IDs from gene symbols, the *mget* function from the *org.Hs.eg.db* package is employed. For each glioma type, we combine the eight sets of selected genes to form one gene set for each glioma type, that is passed to the enrichment analysis.

In a final step of the analysis, spectral clustering using the *Spectrum* method was applied to the resulting glioma transcriptomics datasets accounting for the different sets of genes selected by the network analysis step, to test whether these reduced datasets allow the distinction of known glioma types. The Adjusted Rand Index (ARI) [45] was computed to quantify how much the identified clusters were in agreement with the actual updated labels. The quality of the clustering in terms of tightness and separation was evaluated by the Average Silhouette Width (ASW) [46]. In the *Spectrum* method the first $c = 3$ eigenvectors were used for clustering the data; thus, the feature space based on those was used to calculate the ASW.

Results

To assess if the biological information of the considered omics data is capable of capturing glioma intertumoral heterogeneity, a preliminary unsupervised study was performed. In particular, the *Spectrum* clustering method was applied to the full transcriptomics dataset (16,217 features), the proteomics dataset (174 features), and the transcriptomics$_S$ dataset (143 features), i.e., covering only genes coding for the proteins collected in the proteomics dataset.

The confusion matrices of the clustering results together with the two clustering performance measures ASW and ARI are provided in Table 1. The full RNA transcriptomics dataset produced the best clustering outcome, with the highest scores in both measures, suggesting these data reflect glioma heterogeneity. In contrast, the proteomics dataset lead to poor-quality clusters, not matching the known glioma types. This outcome indicates that the proteomics dataset used was not informative for glioma patient stratification. Interestingly, the clustering obtained for the proteomics-informed dataset, i.e., transcriptomics$_S$, achieved performances comparable to those obtained from the full transcriptomics dataset. The lower input dimension massively reduced the

Table 1 Results from spectral clustering based on the transcriptomics and proteomics datasets (transcriptomics$_S$ stands for the proteomics-informed transcriptomics dataset)

	Transcriptomics			**Proteomics**			**Transcriptomics$_S$**		
	16217 features			**174 features**			**143 features**		
Astrocytoma	4	239	12	57	42	107	4	243	8
Glioblastoma	190	15	1	33	14	69	181	24	1
Oligodendroglioma	0	6	160	37	49	57	0	17	149
ARI	0.82			0.01			0.75		
ASW	0.71			0.45			0.65		

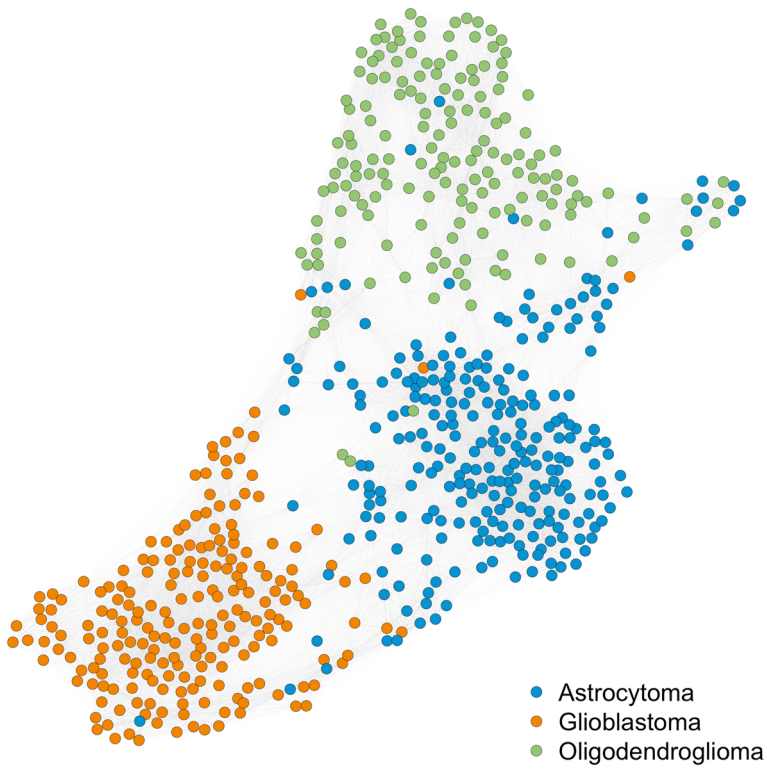


Fig. 2 Patient similarity network computed from the proteomics-informed transcriptomics dataset transcriptomics$_S$. Each node represents a patient, and colors are assigned based on the associated glioma type. The nodes are placed using a force-directed layout algorithm

computational complexity, making this subset of features a valuable input dataset for our analysis.

Figure 2 shows the patient similarity network derived from the transcriptomics$_S$ dataset by the *Spectrum* algorithm. The graph is visualized using the force-directed layout algorithm by Fruchterman and Reingold, which positions nodes that are directly connected closer to each other and increases the distance of isolated nodes [47]. The similarity network shows a remarkable distinction between the three glioma types, despite some samples being closer to others from a different class. For instance, few glioblastoma cases are allocated in the area of astrocytoma and oligodendroglioma, suggesting some molecular affinities that might deserve further investigation. Overall, astrocytoma seems to share more similarities with the other two classes, while oligodendroglioma and glioblastoma can be better distinguished. This prior result is in line with biological knowledge since astrocytoma is characterized by the same cell type as glioblastoma [48], while having mutations in the same gene family as oligodendroglioma [2].

Network discovery and analysis

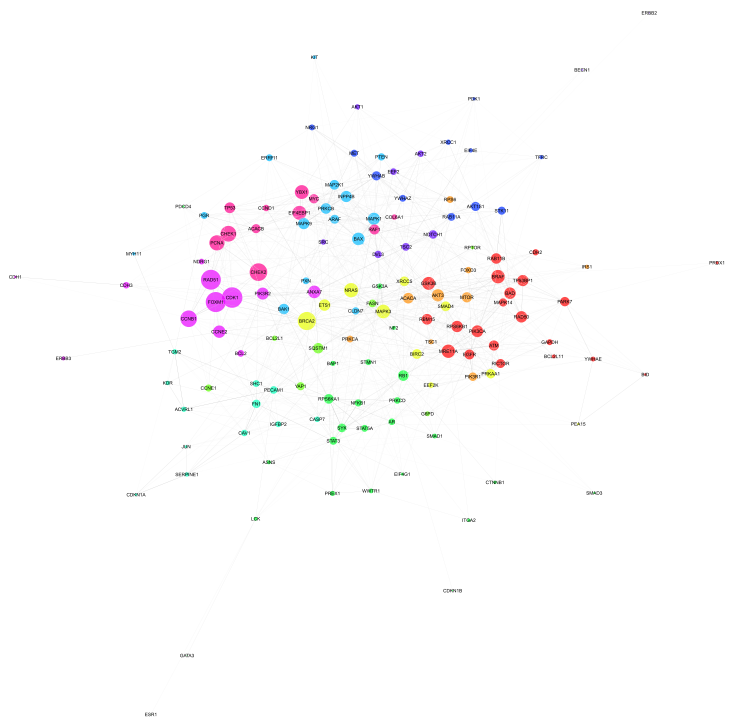
The estimation of the undirected glioma networks by the graphical lasso method was performed separately for glioblastoma, astrocytoma, and oligodendroglioma. The resulting gene co-expression networks comprised a set of genes that have at least one connection, i.e., 141 genes for glioblastoma, 143 genes for astrocytoma, and 141 genes for

oligodendroglioma. Based on these network structures, the Jacobian matrices were computed for each type. We recall that the Jacobian matrix is not symmetric, as two genes can influence each other in both directions and in different ways ($J_{mn} \neq J_{nm}$). Moreover, the entries can be positive or negative, where a positive entry J_{mn} means that gene m has an activating influence on gene n, while a negative entry indicates an inhibiting influence. However, the sign of the influence was not taken into account in our analysis; instead, our focus was on the direction and strength of the relationships between the genes. While in the undirected networks, the edges reflect conditional dependencies between the genes, the edges in the directed networks reflect the extent to which the expression of one gene affects the expression of another gene.

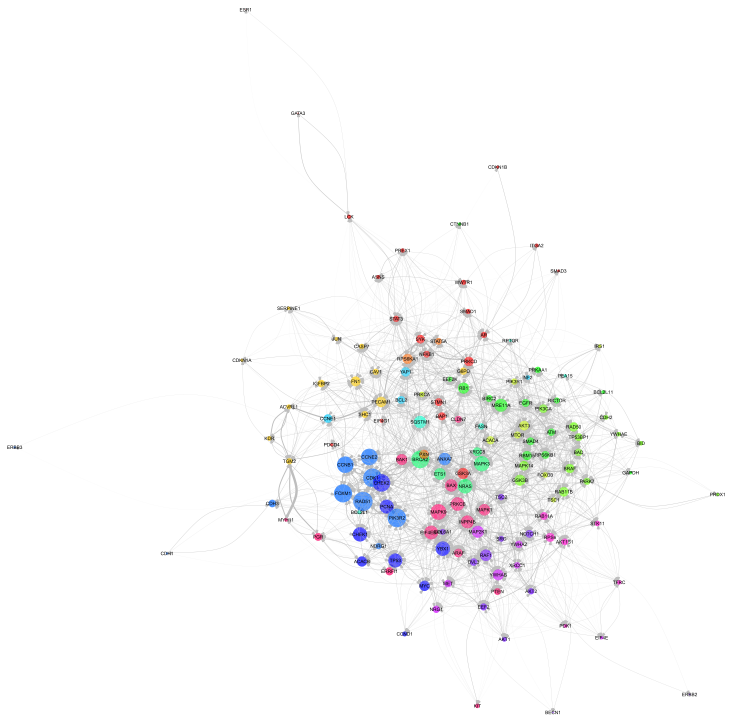
Given the difficulty of a visual analysis of these dense networks, modularity detection by SBMs was employed to uncover groups of related genes and centrality measures were used to detect key genes as potential biomarkers. To this aim, for each glioma type, both undirected and directed networks were considered. The modularity detection algorithm applied to the undirected graphs identified 11 modules in both the astrocytoma and glioblastoma network and 9 modules in the oligodendroglioma network. In the directed networks, 15 modules were detected for both astrocytoma and oligodendroglioma, and 14 modules for glioblastoma. The modules group together genes that have a similar role in the network, indicating that genes within the individual modules are likely to be functionally related and participate in related biological processes.

The detection of such modules and the computation of gene importance improves the interpretability of the overall networks. Strength centrality is a local measure that highlights well-connected genes with significant or numerous direct interactions. Such genes may play central roles in biological processes, acting as direct participants in potentially important interactions [21, 28]. Eigenvector centrality extends beyond the local neighborhood and recursively considers the influence of a node's neighbors. This allows for the identification of genes with a broader influence over the network, highlighting those that may play roles in the regulation or coordination of multiple biological pathways [21]. By capturing both direct and indirect influence, eigenvector centrality points to genes that may serve as central coordinators within complex systems. While eigenvector centrality provides a balanced average over all paths in the network, closeness centrality focuses specifically on the shortest, most direct paths to all other nodes. Genes with high closeness centrality can interact quickly with other genes, suggesting they may be important in signaling and regulatory roles [21, 28].

From the computation of the centrality measures emerged that eigenvector centrality assigns a more distinct ranking among the genes, with some genes having significantly higher importance values than others, whereas strength and especially closeness centrality return a more uniform value to each gene. Eigenvector centrality proves to be particularly valuable in identifying hub genes, offering the most interpretable outcome from a visual point of view. Figure 3 shows, as an example, the astrocytoma undirected and directed network in which the nodes are labelled with the respective gene names and the node size is proportional to the gene importance based on eigenvector centrality. The modularity structure is depicted by the node colors, with genes in the same module sharing a color. The nodes are placed using the force-directed layout algorithm by Fruchterman and Reingold [47].



(a)



(b)

Fig. 3 Astrocytoma **a** undirected network and **b** directed gene network. The nodes are labelled with the corresponding gene names. The modularity structure is depicted by the node colors, with genes in the same module sharing a color. The node area is sized proportional to the node's eigenvector centrality, making influential genes appear larger. The nodes are placed using a force-directed layout algorithm

For ease of network visualization and interpretation, an interactive online visualization tool is provided in https://netzwerk-ninja.github.io/glioma/gene-networks/. The tool enables the exploration of the undirected and directed networks estimated for all three glioma types based on the different centrality measures. The rankings of the network genes according to the centrality measures can be found under https://netzwerk-ninja.github.io/glioma/gene-rankings/, where the genes are colored according to the modularity structure.

Coupling module detection with eigenvector centrality allows the detection of entire groups of genes with high importance within the network. From Fig. 3, it is possible to detect the most relevant module based on eigenvector centrality in both undirected and directed astrocytoma networks, i.e., the purple module in Fig. 3a and the blue module in Fig. 3b. Figure 4 focuses on those subnetworks, allowing a deeper discussion of the included gene relations. Interestingly, despite a generally different module structure of directed and undirected networks, the most important module in each of these two cases is mostly constituted by the same genes, with the only exception of *BCL2*, which is exclusively present in the purple module of the undirected astrocytoma network. This high node overlap among the two graph representations allows the disclosure of how the estimated gene relationships change from the undirected to the directed network. We observe that the edge weights of the undirected module appear relatively uniform, while varying significantly in the directed network. In this case, many pairs of nodes have links in both directions, yet their weights highly differ, suggesting a dominant influence of one gene over the other. For instance, from Fig. 4a, *PIK3R2* is predicted to have a considerable impact on *RAD51*, influencing it by a direct strong positive link, and by a path mediated by *FOXM1*, which might indicate an important gene regulatory process.

A literature review on these genes revealed findings that align with our results, highlighting a potential common ground in DNA repair mechanisms. Specifically, *PIK3R2* regulates the activity of the enzyme Phosphoinositide 3-kinase (PI3K), which

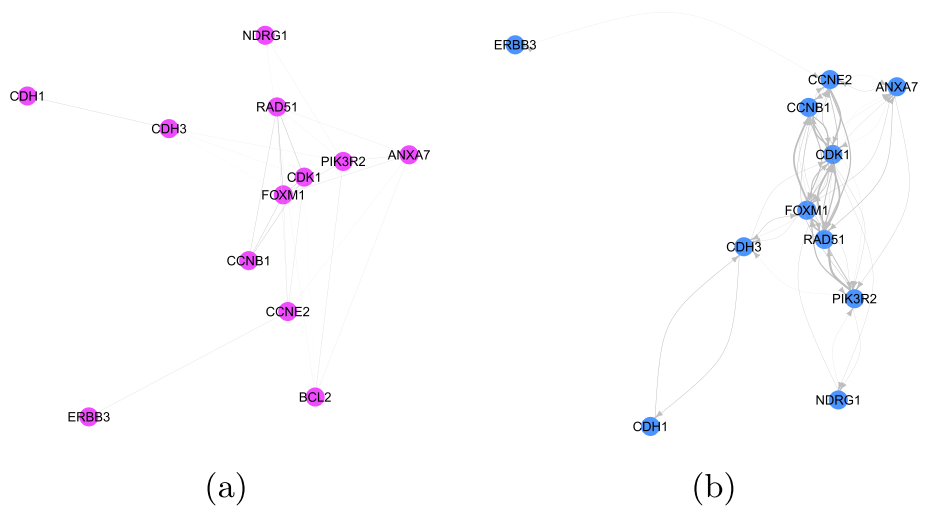


Fig. 4 Most relevant module in astrocytoma **a** undirected network and **b** directed network. These modules include the same genes, with the only exception of *BCL2*, which is exclusively present in the module of the undirected astrocytoma network. Node colors and placements correspond to Fig. 3

is linked to many cell functions, such as growth, proliferation, and cell motility. This gene is widely known in cancer, where it is frequently mutated, and associated with tumor proliferation and increased invasion [49, 50]. Due to these characteristics, recent studies proposed *PIK3R2* as a cancer prognostic marker [51, 52].

On the other hand, *FOXM1* is a transcription factor gene involved in cell proliferation, self-renewal, and tumorigenesis [53]. Commonly upregulated in cancer [54], it is considered a prognostic marker and a therapeutic target [55].

The existence of a link going from *PIK3R2* to *FOXM1* is supported by literature, as *FOXM1* activation is controlled by PI3K/ATK pathway [56], which is crucial in glioma progression [57].

Instead, *RAD51* has its primary function in DNA-damage repairs [58], a process extremely important in cancer, and broadly investigated in glioma due to its implication in treatment failure. In particular, it has been shown that higher levels of *RAD51* are associated with poor glioma survival, due to its role in repairing DNA damages caused by radiation and chemotherapy, contributing to treatment resistance and tumor recurrence [59, 60].

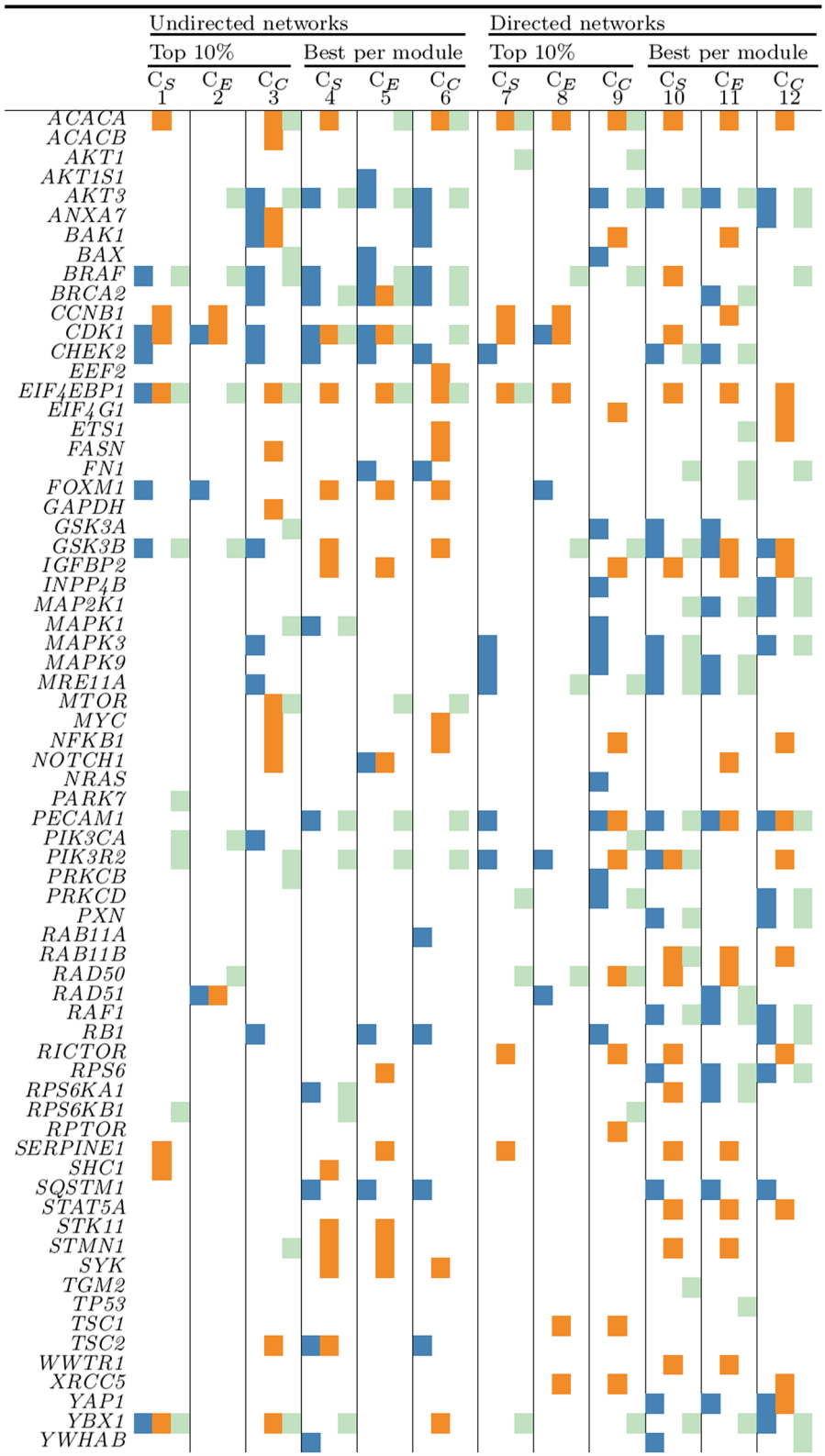
Many studies pointed out the *FOXM1* role in regulating DNA damage response [61–63]. In pediatric glioma, it has been shown that inhibition of PI3K reduces the DNA repair functions by a pathway involving *FOXM1*. In particular, *FOXM1* is responsible for activating promoters of genes having DNA damage repair functions, such as *RAD51* [64]. In ovarian cancer, the depletion of *PIK3R2* results in increased DNA damage and a reduction of the proteins encoded by the *RAD51* gene [65].

All these findings suggest that the genes identified in our approach deserve further exploration to understand their biological functions in glioma. If such regulations are confirmed by biological validations, we may postulate that the *PIK3R2* gene might impact glioma therapy resistance by alternating DNA repair processes, with important clinical implications.

For further biological interpretation and potential biomarker detection, the modularity structures and node importances in the networks were used to select the following sets of genes from the networks: genes covering the top 10% of importance in undirected networks based on strength centrality (set 1), eigenvector centrality (set 2), and closeness centrality (set 3); the most important gene per module in undirected networks based on strength centrality (set 4), eigenvector centrality (set 5), and closeness centrality (set 6); genes covering the top 10% of importance in the directed networks based on strength centrality (set 7), eigenvector centrality (set 8), and closeness centrality (set 9); the most important gene per module in directed networks based on strength centrality (set 10), eigenvector centrality (set 11), and closeness centrality (set 12). The selected genes for each glioma type from all twelve sets are listed in Table 2.

The overview provided in Table 2 allows the identification of promising genes that may play significant roles in the glioma disease due to their importance in the networks. For instance, there are genes identified as important in the different networks, indicating their potential relevance in processes shared across different types. Among them, *ACACA* plays a key role in glioblastoma and oligodendroglioma networks, *AKT3* in astrocytoma and oligodendroglioma, and *CDK1* in the three glioma types.

Table 2 Selected genes in the 8 sets: undirected networks vs. directed networks, genes covering the top 10% of importance vs. most important gene per module, strength centrality (C_S) vs. eigenvector centrality (C_E) vs. closeness centrality (C_C) (Astrocytoma, Glioblastoma, Oligodendroglioma)



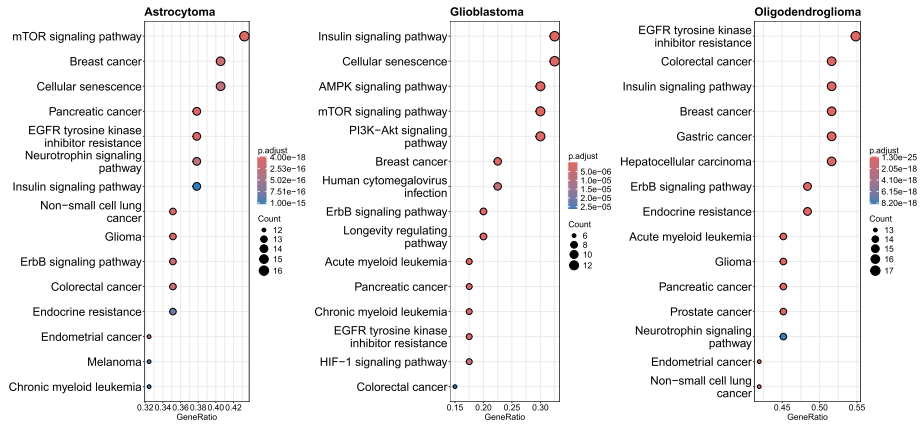


Fig. 5 KEGG enrichment analysis showing the top 15 KEGG pathways significantly enriched in the gene sets. The x-axis displays the ratio of genes associated with each KEGG pathway relative to the total gene set. The dot size reflects the number of genes linked to each pathway and the dot color represents the adjusted *p*-values from the enrichment analysis

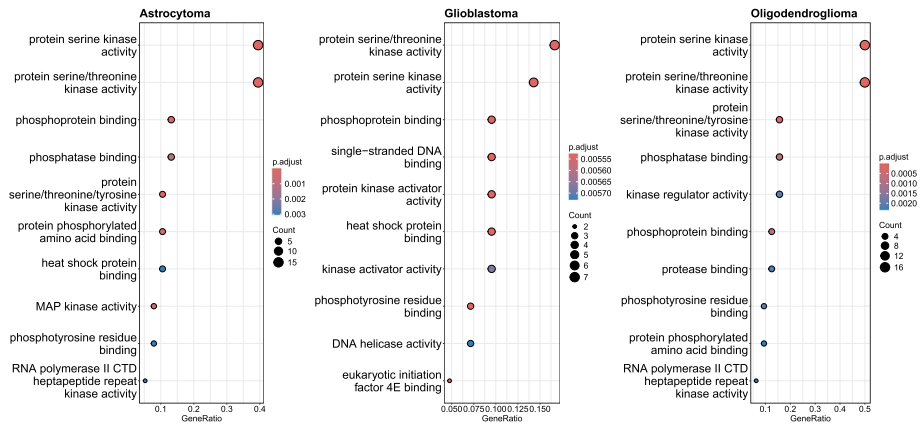


Fig. 6 GO enrichment analysis showing the top 10 GO terms significantly enriched in the gene sets. The x-axis displays the ratio of genes associated with each GO term relative to the total gene set. The dot size reflects the number of genes linked to each term and the dot color represents the adjusted *p*-values from the enrichment analysis

On the other hand, there are genes that were detected as key variables for a single glioma type, making them potential markers for glioma intertumoral heterogeneity. Indeed, many genes were selected exclusively for glioblastoma, such as *CCNB1*, *NFKB1* and *SERPINE1*, potentially characterizing the most aggressive type of glioma.

We analyzed the genes selected for each glioma type using the KEGG and GO enrichment analysis, which allowed us to identify key pathways and functional categories that are significantly overrepresented. Figure 5 highlights the top 15 KEGG pathways significantly enriched in the gene set for each glioma type, while Fig. 6 illustrates the top 10 enriched GO terms. The x-axis shows the ratio of genes associated with each KEGG pathway or GO term relative to the total gene set. The dot size reflects the number of genes linked to each pathway or term, and the dot color represents the adjusted p-values from the enrichment analysis, accounting for multiple testing corrections. Smaller *p*-values indicate greater statistical significance.

Besides the links to glioma and other cancer types, the *EGFR*, *PI3K*/*AKT* and *mTOR* pathways appear among the top significant in the KEGG enrichment analysis. Interestingly, our network analysis had already highlighted the *PI3K*/*AKT* pathway, and its importance in glioma has been discussed previously. The *mTOR* signaling pathway acts both as a downstream effector and upstream regulator of *PI3K* [66], and has become an important therapeutic target for glioblastoma [67, 68]. The activity of this pathway is also associated with disease progression in astrocytoma and oligodendroglima [69]. Glioblastoma frequently exhibits overexpression or mutation of *EGFR*, triggering the activation of many downstream signaling pathways, such as the *PI3K*/*Akt*/*mTOR* pathway [57]. The *AMPK* pathway, here significantly enriched in the glioblastoma gene set, is known for its role in inflammation and cancer, particularly in regulating immune responses within the tumor microenvironment [70]. Its involvement in glioma, specifically in glioblastoma, has also been studied, where it appears to promote tumor formation [71–73].

In the GO enrichment analysis, protein serine kinase activity and protein serine/threonine kinase activity are significantly enriched in all three glioma types. Protein kinases regulate various signalling pathways and cellular processes in cell life such as metabolism, transcription and apoptosis [74]. While protein serine kinase is a borader term, protein serine/threonine kinase activity is directly involved in the *PI3K*/*Akt*/*mTOR* and *AMPK* pathways, as *AKT* and *mTOR* as well as *AMPK* are serine/threonine kinases [57, 73]. Kinase inhibitors have become important in cancer therapy [75, 76] and present a promising therapeutic strategy in glioma.

Patient clustering

Since the dataset exploration performed as a first stage of our analysis assessed that the initial dataset was suitable for patient stratification by spectral clustering, we would like to preserve this ability after variable selection. Therefore, as a final step, clustering was performed by considering only data from the selected features. The confusion matrices of the clustering results together with ASW and ARI values are given in Table 3.

For most subsets, the clustering performance decreased compared to the one provided by the initial dataset. However, despite the considerable variable reduction, most of the final subsets lead to high-quality clusters, with ASW score above 0.5. The ability of class recovery was better for the subsets selected from the undirected networks, where ARI values are around 0.5 or higher. This suggests that the undirected networks based on conditional dependencies better capture the known glioma classes compared to the directed networks derived from the estimated Jacobian matrix. However, the biomarker selection process may also play a role since the centrality measures applied to the directed networks focus solely on the out-direction of links, overlooking bidirectional interactions and feedback loops. This may limit their ability to capture the complexity of gene interactions.

The best clustering results were obtained by considering the most important genes from the modules in the undirected graphs according to closeness centrality (set 6). The ARI value of 0.72 value is only slightly below that of the initial dataset and the ASW value of 0.68 is even higher, indicating that these genes show great potential as diagnostic biomarkers capturing glioma heterogeneity. Figure 7 visualizes the clustering based on set 6 in the patient similarity graph. Each node represents a patient, colors are

Table 3 Results from spectral clustering based on the variables selected from undirected vs. directed networks; genes covering the top 10% of importance vs. most important gene per module; strength centrality (C_S) vs. eigenvector centrality (C_E) vs. closeness centrality (C_C))

	Undirected networks - Top 10%									**Undirected networks - Best per module**									**Directed networks - Top 10%**									**Directed networks - Best per module**								
	C_S			**C_E**			**C_C**			**C_S**			**C_E**			**C_C**			**C_S**			**C_E**			**C_C**			**C_S**			**C_E**			**C_C**		
	15 features			**10 features**			**29 features**			**23 features**			**23 features**			**26 features**			**16 features**			**13 features**			**30 features**			**32 features**			**32 features**			**29 features**		
	(set 1)			**(set 2)**			**(set 3)**			**(set 4)**			**(set 5)**			**(set 6)**			**(set 7)**			**(set 8)**			**(set 9)**			**(set 10)**			**(set 11)**			**(set 12)**		
Astrocytoma	181	24	50	171	21	63	232	21	2	204	26	25	207	31	17	225	26	4	235	18	2	85	46	124	186	41	28	224	12	19	232	21	2	219	8	28
Glioblastoma	32	170	4	45	140	21	14	192	0	16	186	4	12	194	0	10	196	0	66	139	1	15	186	5	13	192	1	27	178	1	28	178	0	32	173	1
Oligodendroglioma	17	4	145	135	4	27	68	1	97	46	4	116	12	9	145	19	3	144	70	2	94	36	11	119	120	10	36	155	2	9	40	2	124	29	0	137
ARI	0.47			0.20			0.57			0.51			0.64			0.72			0.36			0.30			0.35			0.38			0.59			0.58		
ASW	0.55			0.42			0.56			0.51			0.63			0.68			0.59			0.43			0.60			0.65			0.49			0.54		

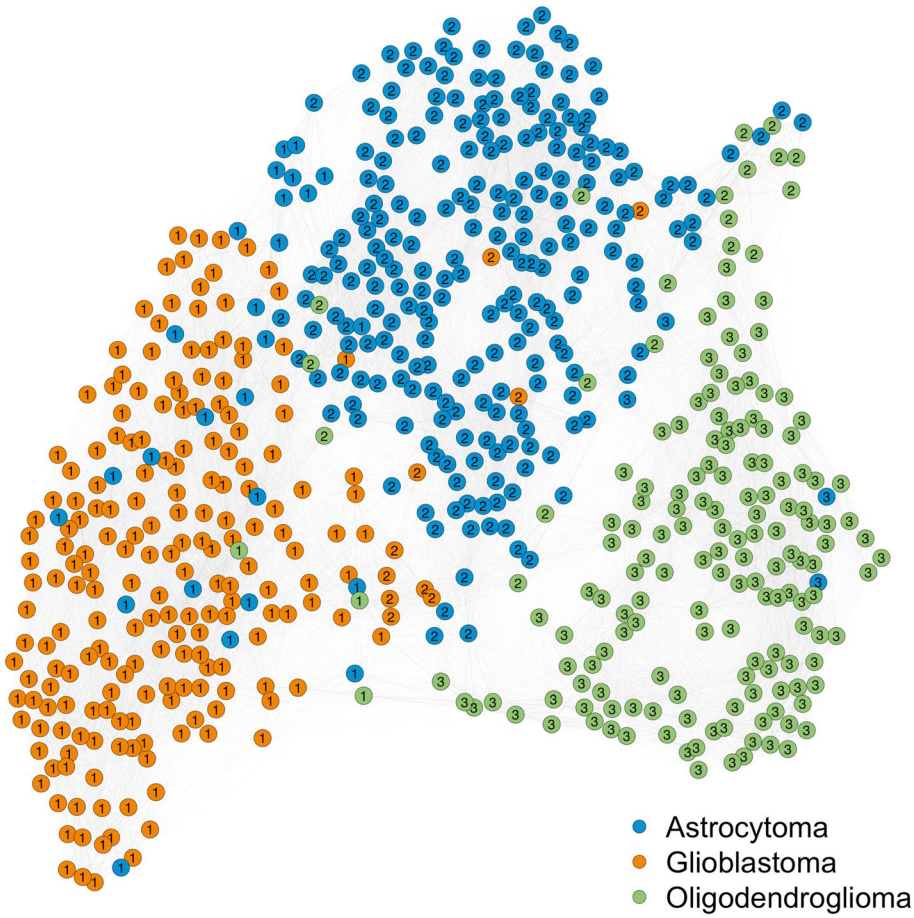


Fig. 7 Patient similarity network computed based on the genes included in set 6 (the most important genes from the modules in the undirected graphs according to closeness centrality). Each node represents a patient, colors are assigned based on the associated glioma type and the node labels represent the assigned cluster. The nodes are placed using a force-directed layout algorithm

assigned based on the associated glioma type, and the node labels represent the assigned cluster. This representation highlights the definition of well-defined and distinctly separated clusters, as indicated by the high average silhouette score obtained for set 6. The few misaligned observations (i.e., patients grouped in a cluster that does not match their assigned glioma type) tend to be located on the border of their respective glioma type. This suggests a similarity across some glioma types, as expected. Particularly, we can observe similarity between oligodendroglioma (green) and astrocytoma (blue), and between astrocytoma and glioblastoma (orange), which aligns with biological knowledge and prior studies [77–79]. On the other hand, a small number of these misaligned observations appears to be more closely aligned with patients in their assigned cluster than those of their actual glioma type. This could indicate that these patients have a distinct transcriptomic molecular profile, potentially identifying outliers or hinting at a need for a different stratification of patient groups.

Generally, the clustering performance improved when considering the modularity structure of the networks, though this result might also be affected by the number of genes included in the analysis, which is generally larger than the one derived from the

top 10% approach. However, when looking at the sets of genes selected by closeness centrality, the modularity approach led to better ARI values with fewer genes. Closeness centrality provided a relatively uniform distribution of importance between the genes, which explains why the top 10% includes a higher number of genes compared to the other centrality measures. The worst performances in terms of ARI and ASW values were provided by considering genes from the top 10% based on eigenvector centrality. Nevertheless, if the number of genes included in the analysis might affect clustering results, it is worth to note that both these cases are characterized by the smallest number of features.

The confusion matrices reveal that oligodendroglioma and astrocytoma are often assigned to the same cluster. This outcome aligns well with the known similarities in the genetic profiles of these glioma types. Astrocytomas and oligodendrogliomas are both low-grade gliomas exhibiting mutations in the same gene family, i.e., the IDH genes. Glioblastomas, on the other hand, display a wild-type IDH status, and are characterized by genetic or histological alterations, uncommon to astrocytomas and oligodendrogliomas. In diagnosis, this IDH mutation is a crucial marker that distinguishes astrocytomas and oligodendrogliomas, from the more aggressive glioblastomas [2]. This outcome is also confirmed by previous studies based on the same datasets, investigating glioma heterogeneity by network inference, which also highlights the similarity between astrocytoma and oligodendroglioma [29, 80]. Furthermore, there is more overlap between glioblastoma and astrocytoma than between glioblastoma and oligodendroglioma, which is reasonable since both glioblastoma and astrocytoma arise from the same cell type, namely astrocytes.

In this study, clustering was used as an unsupervised approach to assess the effectiveness of the selected genes in differentiating the glioma classes. An additional literature review on the subset of genes leading to the best clustering results (set 6) was conducted to explore their specific functions in gliomas and cancer in general. The findings can be found in Table 4 in Appendix. In the literature review, all selected features are mentioned in relation to human cancers, many in the context of gliomas. Among studies about gliomas, many studies are focused on glioblastoma, widely studied due to being the most common and aggressive type of glioma. The *ACACA* gene plays a significant role in the glioblastoma and oligodendroglioma networks but has not yet been linked to glioma in the literature. *CHEK2* deservers further exploration in astrocytoma, and *EIF-4EBP1* should be considered for investigation in all types of glioma, since they have not yet been studied experimentally in glioma.

Conclusions

In this study, we proposed a network-based methodology to discover relevant gene interaction networks associated with different glioma types. We offer a novel perspective on gene expression in glioma, revealing valuable molecular insights and potential biomarkers that may contribute to diagnosis and therapy. The methodology encompasses the estimation gene co-expression networks from transcriptomics data using the graphical lasso method, the estimation of causal relationships by the Jacobian matrix, modularity detection by SBMs and the analysis of gene importance based on centrality measures.

Further clustering of patient similarity networks evaluates the suitability of selected network genes to stratify the patients into the diagnosed glioma types.

While the sparse gene co-expression networks identified by the graphical lasso method stand as valuable information regarding the molecular associations involved in different glioma types, the causal links disclosed through the Jacobian matrix further increase the biological understanding of the directions of the corresponding interactions. Although these cannot be considered as direct gene regulatory networks due to the absence of transcription factors that usually mediate gene-gene interplay [81], the estimated causal networks provide an important indication of gene interactions, a crucial starting point to foster further research. In particular, by analyzing the directed gene relations in a astrocytoma module, we were able to detect very promising regulatory mechanisms, which aligns with prior studies, yet not comprehensively investigated in glioma. The confirmation of these findings might increase the overall glioma understanding, with potentially important clinical implications. In the astrocytoma directed network as well as in the glioblastoma and oligondendroglioma networks, several more causal interactions between genes can be identified. With increased computational resources, it would be interesting to explore network inference using a larger dataset. For the Jacobian reconstruction, further research might consider other prior information about the network structure available from curated databases regarding known molecular interactions involved in glioma or cancer in general.

The analysis of the resulting networks through node centrality and modularity detection provided valuable insights into gene importance and the structure of gene interactions in the glioma networks. This led to a selection of genes as potential biomarkers. For each glioma type, we performed KEGG and GO enrichment analysis based on the selected genes, showing biological pathways and processes that are enriched in the set of genes. While the selected genes are already valuable as disease drivers and therapeutic targets, we further applied spectral clustering to assess their ability to group patients into the known glioma types.

By calculating closeness centrality and selecting the most central gene from each module in the undirected networks, a set of promising genes that captures intertumoral heterogeneity in gliomas was identified from a mathematical and computational standpoint. Spectral clustering based on these genes yielded clusters that closely aligned with the established glioma subtypes. Glioma is a highly heterogeneous disease, exhibiting greater variability than is captured by the three established glioma types. The structure we identified enables the discovery of new patient subgroups, sharing molecular characteristics.

We conducted a literature review to evaluate previous reports on the role of these selected genes in glioma or cancer in general, which is essential to assess their potential for diagnosis and therapy. While several genes have already been associated with glioma, many have not yet been comprehensively investigated. To promote further studies, we provided a ranked list that helps researchers prioritizing genes for further experimental validation, involving biological testing of the most promising candidates to examine their role in the development, progression, and therapy of glioma.

Appendix

Table 4 Selected genes and their function in gliomas and cancer in general

Gene	Function
ACACA	Aberrant expression increases cancer risk [82]; down-regulation suppresses malignant progression of prostate cancer [83]
AKT3	Amplification promotes glioma progression [84]; it delays tumor progression, therefore strategies that inhibit *AKT3* may be unhelpful in some glioblastoma patients [85]
ANXA7	De-regulated in many cancers, appears to have a tumor-suppression role in glioblastoma [86]; tumor suppressor, augments EGFR signaling in glioblastomas [87]
BAK1	Targeting BAK1 enhances chemosensitivity of glioblastoma stem cells to Temozolomide [88]
BRAF	Lower-grade gliomas of childhood are characterized by *BRAF* mutations [89]; *BRAF* mutation associated with improved survival in glioma [90]
BRCA2	Downregulation can sensitize glioma cells to killing by anti-cancer drugs [91]; factor for poor prognosis in glioma patients [92]
CDK1	Inhibiting CDK1 activity promotes apoptosis in glioblastoma [93]; significantly enriched in glioblastoma cell cycle [94]
CHEK2	*CHEK2* gene polymorphism might correlate with prognosis of glioblastoma patients [95], activated upon DNA damage; associated with breast cancer [96]
EEF2	Regulates autophagy in human glioblastoma cells [97]; targeting *EEF2* kinase can enhance the antitumor activity of Temozolomide against glioma [98], tumor-associated antigen overexpressed in various types of cancers [99]
EIF4EBP1	Increased expression in malignant gliomas [100]; overexpression linked to poor survival and disease progression in hepatocellular carcinoma [101]
ETS1	Key role on vascular abnormality in glioblastoma [102]
FASN	Up-regulation of *FASN* correlates strongly with glioma grade [103, 104]; inhibition of *FASN* decreases expression of stemness markers in glioma stem cells [105]
FN1	Upregulated in glioblastoma; promotes glioblastoma cell proliferation by altering *PTPRM* methylation [106]
FOXM1	Promotes glioma cells progression [107]; overexpressed in malignant glioma [108]
GSK3B	*GSK3B* activity is a critical regulator of glioblastoma stem-like cell survival and apoptosis [109]; potential anticancer targets for astrocytoma therapy [110]; *AKT/GSK3* signaling pathway plays a significant role in the pathogenesis of glioblastoma [111]
MTOR	Important role in glioblastoma [66]; therapeutic target for glioblastoma [67, 68]; the activity of the *PI3K/Akt/mTOR* is also associated with astrocytoma and oligodendroglioma [69]
MYC	*MYC* inhibition effective against glioma [112], c-Myc protein required for maintenance of glioma cancer stem cells [113]
NFKB1	Significantly up-regulated in tumor tissue of glioblastoma, astrocytoma and oligodendroglioma but not in other cancer types [114], implicated in carcinogenesis, in some cases as tumor-driver, in others as tumour-suppressor [115]
PECAM1	Downregulated in glioblastoma; correlated to the overexpression of HIF-1α, a transcriptional regulator increasing tumor aggressiveness, invasiveness and resistance to radiotherapy and chemotherapy [116]
PIK3R2	Regulates the activity of PI3K, which is linked to many cell functions, and well known in glioma [117–119]; frequently mutated, and associated with tumor proliferation and increased invasion [49, 50]; proposed as a cancer prognostic marker [51, 52]
RAB11A	The *RAB11* family protein family is the master regulator of vesicular trafficking and is prone to be altered in human cancers [120]
RB1	Downregulated in human brain tumors [121]; alteration of the *RB1*-pathway related genes are associated with shorter survival in low-grade gliomas [122]; in glioblastoma, loss of RB1 expression correlates with hypermethylation, which is not reported in low-grade astrocytoma [123]
SQSTM1	Acts as a regulator in autophagy in glioblastoma cells [124]; and cancer in general [125]
SYK	*SYK* inhibition blocks proliferation and migration of glioma cells [126]; associated with malignant phenotype and immune checkpoints in diffuse glioma [127]
TSC2	*AKT* activation in human glioblastomas enhances proliferation via *TSC2* and *S6* kinase signaling [128], *TSC2* protects glioblastoma from cell death induced by Photodynamic therapy [129]
YBX1	Potential regulator of tumor invasion in glioblastoma [130], highly overexpressed in multiple cancer types

Acknowledgements
The results presented here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Authors' contributions
NK, RC, and MBL conceptualized and designed the study. NK performed research, formal analysis, and wrote the original draft. NK and RC implemented the codes and curated the data. NK and MBL created the visual contents. RC, TG, and MBL supervised the work. MBL administrated the project and acquired funding. All authors defined the methodologies and reviewed the final manuscript.

Funding
This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references CEECINST/00042/2021, UIDB/00297/2020 and UIDP/00297/2020 (NOVA Math, Center for Mathematics and Applications), UIDB/00667/2020 and UIDP/00667/2020 (UNIDEMI), and the research project PTDC/CCI-BIO/4180/2020 "MONET - Multi-omic networks in gliomas" (doi: 10.54499/PTDC/CCI-BIO/4180/2020). This work was also supported by Erasmus + under the project "MIX IT".

Data availability
The updated diagnostic labels of the samples of the TCGA glioma dataset used in this study are available at: https://github.com/sysbiomed/MONET. The interactive online visualization tool for the exploration of the networks estimated for the three glioma types based on the different centrality measures is provided at https://netzwerk-ninja.github.io/glioma/gene-networks/. The rankings of the network genes according to the centrality measures can be found under https://netzwerk-ninja.github.io/glioma/gene-rankings/. R codes and scripts used to perform this study will be made available upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 July 2024 Accepted: 25 November 2024
Published online: 18 December 2024

References

1. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, et al. Glioma. Nat Rev Dis Prim. 2015;1(1):15017. https://doi.org/10.1038/nrdp.2015.17.
2. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-Oncol. 2021;23:1231–51. https://doi.org/10.1007/s00401-016-1545-1.
3. Mohammed S, Dinesan M, Ajayakumar T. Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: a retrospective study. Rep Pract Oncol Radiother. 2022;27(6):1026–36. https://doi.org/10.5603/RPOR.a2022.0113.
4. Lopes MB, Martins EP, Vinga S, Costa BM. The Role of Network Science in Glioblastoma. Cancers. 2021;13(5):1045. https://doi.org/10.3390/cancers13051045.
5. Axenie C, Bauer R, Martínez MR. The Multiple Dimensions of Networks in Cancer: A Perspective. Symmetry. 2021;13(9):1559. https://doi.org/10.3390/sym13091559.
6. Kosvyra A, Ntzioni E, Chouvarda I. Network analysis with biological data of cancer patients: A scoping review. J Biomed Inform. 2021;120:103873. https://doi.org/10.1016/j.jbi.2021.103873.
7. Reyna MA, Leiserson MDM, Raphael BJ. Hierarchical HotNet: identifying hierarchies of altered subnetworks. Bioinformatics. 2018;34(17):i972–80. https://doi.org/10.1093/bioinformatics/bty613.
8. Cho A, Shim JE, Kim E, Supek F, Lehner B, Lee I. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. Genome Biol. 2016;17(1):129. https://doi.org/10.1186/s13059-016-0989-x.
9. Xi J, Wang M, Li A. Discovering mutated driver genes through a robust and sparse co-regularized matrix factorization framework with prior information from mRNA expression patterns and interaction network. BMC Bioinformatics. 2018;19(1):214. https://doi.org/10.1186/s12859-018-2218-y.
10. Gill R, Datta S, Datta S. Differential Network Analysis in Human Cancer Research. Curr Pharm Des. 2014;20(1):4–10. https://doi.org/10.2174/1381612820011401131223 16.
11. Ha MJ, Baladandayuthapani V, Do KA. DINGO: differential network analysis in genomics. Bioinformatics. 2015;31(21):3413–20. https://doi.org/10.1093/bioinformatics/btv406.
12. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun. 2014;5(1):3231. https://doi.org/10.1038/ncomms4231.

13. Li YK, Hsu HM, Lin MC, Chang CW, Chu CM, Chang YJ, et al. Genetic co-expression networks contribute to creating predictive model and exploring novel biomarkers for the prognosis of breast cancer. Sci Rep. 2021;11(1):7268. https://doi.org/10.1038/s41598-021-84995-z.
14. Redekar SS, Varma SL, Bhattacharjee A. Gene co-expression network construction and analysis for identification of genetic biomarkers associated with glioblastoma multiforme using topological findings. J Egypt Natl Cancer Inst. 2023;35(1):22. https://doi.org/10.1186/s43046-023-00181-4.
15. Hawe JS, Theis FJ, Heinig M. Inferring Interaction Networks From Multi-Omics Data. Front Genet. 2019;10:535. https://doi.org/10.3389/fgene.2019.00535.
16. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9(3):432–41. https://doi.org/10.1093/biostatistics/kxm045.
17. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian Networks to Analyze Expression Data. J Comput Biol. 2000;7(3–4):601–20. https://doi.org/10.1089/106652700750050961.
18. Liu E, Li L, Cheng L. Gene Regulatory Network Review. In: Encyclopedia of Bioinformatics and Computational Biology. Amsterdam: Elsevier; 2019. pp. 155–64. https://doi.org/10.1016/B978-0-12-809633-8.20218-5.
19. Steuer R, Kurths J, Fiehn O, Weckwerth W. Observing and interpreting correlations in metabolomicnetworks. Bioinformatics. 2003;19(8):1019–26. https://doi.org/10.1093/bioinformatics/btg120.
20. Barter E, Brechtel A, Drossel B, Gross T. A closed form for Jacobian reconstruction from time series and its application as an early warning signal in network dynamics. Proc R Soc A Math Phys Eng Sci. 2021;477(2247):2020–074220200742. https://doi.org/10.1098/rspa.2020.0742.
21. Koschützki D, Schreiber F. Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. Gene Regul Syst Biol. 2008;2:GRSB.S702. https://doi.org/10.4137/GRSB.S702.
22. Azuaje FJ. Selecting biologically informative genes in co-expression networks with a centrality score. Biol Direct. 2014;9(1):12. https://doi.org/10.1186/1745-6150-9-12.
23. Newman MEJ. Analysis of weighted networks. Phys Rev E. 2004;70(5):056131. https://doi.org/10.1103/PhysRevE.70.056131.
24. Bonacich P. Power and Centrality: A Family of Measures. Am J Sociol. 1987;92(5):1170–82. https://doi.org/10.1086/228631.
25. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw. 1978;1(3):215–39. https://doi.org/10.1016/0378-8733(78)90021-7.
26. Abbe E. Community Detection and Stochastic Block Models: Recent Developments. J Mach Learn Res. 2018;18(177):1–86.
27. Peixoto TP. 11. In: Bayesian Stochastic Blockmodeling. Hoboken: Wiley; 2019. pp. 289–332. https://doi.org/10.1002/9781119483298.ch11.
28. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005;118(21):4947–57. https://doi.org/10.1242/jcs.02714.
29. Martins S, Coletti R, Lopes MB. Disclosing transcriptomics network-based signatures of glioma heterogeneity using sparse methods. BioData Min. 2023;16(1):26. https://doi.org/10.1186/s13040-023-00341-1.
30. Mendonça ML, Coletti R, Gonçalves CS, Martins EP, Costa BM, Vinga S, et al. Updating TCGA glioma classification through integration of molecular profiling data following the 2016 and 2021 WHO guidelines. bioRxiv. 2023. https://doi.org/10.1101/2023.02.19.529134.
31. Liu H, Roeder K, Wasserman L. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. 2010. https://doi.org/10.48550/ARXIV.1006.3316.
32. Van Kampen NG. Stochastic Processes in Physics and Chemistry. Amsterdam: North-Holland Personal Library; 2007.
33. Horn RA, Johnson CR. Matrix Analysis. New York: Cambridge University Press; 2012.
34. Newman M. Networks. Oxford: Oxford University Press; 2018.
35. Côme E, Jouvin N, Latouche P, Bouveyron C. Hierarchical clustering with discrete latent variable models and the integrated classification likelihood. ADAC. 2021;15:957–86. https://doi.org/10.1007/s11634-021-00440-z.
36. Von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007;17(4):395–416. https://doi.org/10.1007/s11222-007-9033-z.
37. John CR, Watson D, Barnes MR, Pitzalis C, Lewis MJ. Spectrum: fast density-aware spectral clustering for single and multi-omic data. Bioinformatics. 2020;36(4):1159–66. https://doi.org/10.1093/bioinformatics/btz704.
38. TCGA. The Cancer Genome Atlas. 2023. https://www.cancer.gov/tcga. Accessed 15 Sept 2023.
39. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. Cell. 2013;155(2):462–77. https://doi.org/10.1007/s00401-016-1545-1.
40. TCGA. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med. 2015;372:2481–2498. https://doi.org/10.1056/NEJMoa1402121.
41. TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(23):1061–8. https://doi.org/10.1007/s00401-016-1545-1.
42. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO Classification of Tumours of the Central Nervous System. Acta Neuropathol. 2007;114(2):97–109. https://doi.org/10.1007/s00401-007-0243-4.
43. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods. 2013;10(11):1046–7. https://doi.org/10.1038/nmeth.2650.
44. Liu H, Lafferty J, Wasserman L. The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. J Mach Learn Res. 2009;10(80):2295–328.
45. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2(1):193–218. https://doi.org/10.1007/BF01908075.
46. Kaufman L, Rousseeuw PJ. Finding Groups in Data. An Introduction to Cluster Analysis. New York: Wiley InterScience; 1990.
47. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Experience. 1991;21(11):1129–64. https://doi.org/10.1002/spe.4380211102.
48. Reifenberger G, Collins VP. Pathology and molecular genetics of astrocytic gliomas. J Mol Med. 2004;82:656–70. https://doi.org/10.1007/s00109-004-0564-x.

49. Andrade VP, Morrogh M, Qin LX, Olvera N, Giri D, Muhsen S, et al. Gene expression profiling of lobular carcinoma in situ reveals candidate precursor genes for invasion. Mol Oncol. 2015;9:772–82. https://doi.org/10.1016/j.molonc.2014.12.005.
50. Vallejo-Díaz J, Chagoyen M, Olazabal-Morán M, González-García A, Carrera AC. The Opposing Roles of PIK3R1/p85α and PIK3R2/p85β in Cancer. Trends Cancer. 2019;5:233–44. https://doi.org/10.1016/j.trecan.2019.02.009.
51. Liu Y, Wang D, Li Z, Li X, Jin M, Jia N, et al. Pan-cancer analysis on the role of PIK3R1 and PIK3R2 in human tumors. Sci Rep. 2022;12:5924. https://doi.org/10.1038/s41598-022-09889-0.
52. Chicco D, Alameer A, Rahmati S, Jurman G. Towards a potential pan-cancer prognostic signature for gene expression based on probesets and ensemble machine learning. BioData Min. 2022;15:28. https://doi.org/10.1186/s13040-022-00312-y.
53. Wierstra I, Alves J. FOXM1, a typical proliferation-associated transcription factor. Biol Chem. 2007;388:1257–74. https://doi.org/10.1515/BC.2007.159.
54. Liao GB, Li XZ, Zeng S, Liu C, Yang SM, Yang L, et al. Regulation of the master regulator FOXM1 in cancer. Cell Commun Signal. 2018;16:57. https://doi.org/10.1186/s12964-018-0266-6.
55. Borhani S, Gartel AL. FOXM1: a potential therapeutic target in human solid cancers. Expert Opin Ther Targets. 2020;24:205–17. https://doi.org/10.1080/14728222.2020.1727888.
56. Chesnokov MS, Borhani S, Halasi M, Arbieva Z, Khan I, Gartel AL. FOXM1-AKT Positive Regulation Loop Provides Venetoclax Resistance in AML. Front Oncol. 2021;11. https://doi.org/10.3389/fonc.2021.696532.
57. Li X, Wu C, Chen N, Gu H, Yen A, Cao L, et al. PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma. Oncotarget. 2016;7(22):33440–50. https://doi.org/10.18632/oncotarget.7961.
58. Laurini E, Marson D, Fermeglia A, Aulic S, Fermeglia M, Pricl S. Role of Rad51 and DNA repair in cancer: A molecular perspective. Pharmacol Ther. 2020;208:107492. https://doi.org/10.1016/j.pharmthera.2020.107492.
59. King HO, Brend T, Payne HL, Wright A, Ward TA, Patel K, et al. RAD51 Is a Selective DNA Repair Target to Radiosensitize Glioma Stem Cells. Stem Cell Rep. 2017;8:125–39. https://doi.org/10.1016/j.stemcr.2016.12.005.
60. Morrison C, Weterings E, Mahadevan D, Sanan A, Weinand M, Stea B. Expression Levels of RAD51 Inversely Correlate with Survival of Glioblastoma Patients. Cancers. 2021;13(21). https://doi.org/10.3390/cancers13215358.
61. Zona S, Bella L, Burton MJ, de Moraes GN, Lam EWF. FOXM1: An emerging master regulator of DNA damage response and genotoxic agent resistance. Biochim Biophys Acta (BBA) - Gene Regul Mech. 2014;1839:1316–22.
62. Zhang N, Wu X, Yang L, Xiao F, Zhang H, Zhou A, et al. FoxM1 Inhibition Sensitizes Resistant Glioblastoma Cells to Temozolomide by Downregulating the Expression of DNA-Repair Gene Rad51. Clin Cancer Res. 2012;18:5961–71. https://doi.org/10.1158/1078-0432.CCR-12-0039.
63. Tabnak P, Bashkandi AH, Ebrahimnezhad M, Soleimani M. Forkhead box transcription factors (FOXOs and FOXM1) in glioma: from molecular mechanisms to therapeutics. Cancer Cell Int. 2023;23:238. https://doi.org/10.1186/s12935-023-03090-7.
64. Pal S, Kozono D, Yang X, Fendler W, Fitts W, Ni J, et al. Dual HDAC and PI3K Inhibition Abrogates NFKB- and FOXM1-Mediated DNA Damage Response to Radiosensitize Pediatric High-Grade Gliomas. Cancer Res. 2018;78:4007–21. https://doi.org/10.1158/0008-5472.CAN-17-3691.
65. Mak VCY, Li X, Rao L, Zhou Y, Tsao SW, Cheung LWT. p85β alters response to EGFR inhibitor in ovarian cancer through p38 MAPK-mediated regulation of DNA repair. Neoplasia. 2021;23:718–30. https://doi.org/10.1016/j.neo.2021.05.009.
66. Akhavan D, Cloughesy TF, Mischel PS. mTOR signaling in glioblastoma: lessons learned from bench to bedside. Neuro-Oncol. 2010;12(8):882–9. https://doi.org/10.1093/neuonc/noq052.
67. Duzgun Z, Eroglu Z, Biray Avci C. Role of mTOR in glioblastoma. Gene. 2016;575(2):187–90. https://doi.org/10.1016/j.gene.2015.08.060.
68. Divé I, Klann K, Michaelis JB, Heinzen D, Steinbach JP, Münch C, et al. Inhibition of mTOR signaling protects human glioma cells from hypoxia-induced cell death in an autophagy-independent manner. Cell Death Discov. 2022;8(1):409. https://doi.org/10.1038/s41420-022-01195-y.
69. Mohamed E, Kumar A, Zhang Y, Wang AS, Chen K, Lim Y, et al. PI3K/AKT/mTOR signaling pathway activity in IDH-mutant diffuse glioma and clinical implications. Neuro-Oncol. 2022;24(9):1471–81. https://doi.org/10.1093/neuonc/noac064.
70. Keerthana CK, Rayginia TP, Shifana SC, Anto NP, Kalimuthu K, Isakov N, et al. The role of AMPK in cancer metabolism and its impact on the immunomodulation of the tumor microenvironment. Front Immunol. 2023;14:1114582. https://doi.org/10.3389/fimmu.2023.1114582.
71. Leli NM, Koumenis C. Pro-tumorigenic AMPK in glioblastoma. Nat Cell Biol. 2018;20(7):736–7. https://doi.org/10.1038/s41556-018-0129-9.
72. Lorenz NI, Sauer B, Zeiner PS, Strecker MI, Luger AL, Schulte D, et al. AMP-kinase mediates adaptation of glioblastoma cells to conditions of the tumour microenvironment. 2022. https://doi.org/10.1101/2022.03.25.485624.
73. Chhipa RR, Fan Q, Anderson J, Muraleedharan R, Huang Y, Ciraolo G, et al. AMP kinase promotes glioblastoma bioenergetics and tumour growth. Nat Cell Biol. 2018;20(7):823–35. https://doi.org/10.1038/s41556-018-0126-z.
74. Johnson LN. The regulation of protein phosphorylation. Biochem Soc Trans. 2009;37(4):627–41. https://doi.org/10.1042/BST0370627.
75. Wu P, Hu YZ. PI3K/Akt/mTOR Pathway Inhibitors in Cancer: A Perspective on Clinical Progress. Curr Med Chem. 2010;17(35):4326–41. https://doi.org/10.2174/092986710793361234.
76. Glaviano A, Foo ASC, Lam HY, Yap KCH, Jacot W, Jones RH, et al. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. Mol Cancer. 2023;22(1):138. https://doi.org/10.1186/s12943-023-01827-6.
77. Coletti R, Leiria De Mendonça M, Vinga S, Lopes MB. Inferring Diagnostic and Prognostic Gene Expression Signatures Across WHO Glioma Classifications: A Network-Based Approach. Bioinforma Biol Insights. 2024;18:11779322241271536. https://doi.org/10.1177/11779322241271535.
78. Ohgaki H, Kleihues P. Genetic profile of astrocytic and oligodendroglial gliomas. Brain Tumor Pathol. 2011;28(3):177–83. https://doi.org/10.1007/s10014-011-0029-1.

79. Seifert M, Garbe M, Friedrich B, Mittelbronn M, Klink B. Comparative transcriptomics reveals similarities and differences between astrocytoma grades. BMC Cancer. 2015;15(1):952. https://doi.org/10.1186/s12885-015-1939-9.
80. Coletti R, Mendonça ML, Vinga S, Lopes MB. Inferring diagnostic and prognostic gene expression signatures across WHO glioma classifications: A network-based approach. 2023. https://doi.org/10.48550/arXiv.2305.12207.
81. Schlitt T, Brazma A. Current approaches to gene regulatory network modelling. BMC Bioinformatics. 2007;8(S6):S9. https://doi.org/10.1186/1471-2105-8-S6-S9.
82. Bhattacharjee K, Nath M, Choudhury Y. Fatty acid synthesis and cancer: Aberrant expression of the ACACA and ACACB genes increases the risk for cancer. Meta Gene. 2020;26:100798. https://doi.org/10.1016/j.mgene.2020.100798.
83. Zhang H, Liu S, Cai Z, Dong W, Ye J, Cai Z, et al. Down-regulation of ACACA suppresses the malignant progression of Prostate Cancer through inhibiting mitochondrial potential. J Cancer. 2021;12(1):232–43. https://doi.org/10.7150/jca.49560.
84. Turner KM, Sun Y, Ji P, Granberg KJ, Bernard B, Hu L, et al. Genomically amplified Akt3 activates DNA repair pathway and promotes glioma progression. Proc Natl Acad Sci. 2015;112(11):3421–6. https://doi.org/10.1073/pnas.1414573112.
85. Joy A, Kapoor M, Georges J, Butler L, Chang Y, Li C, et al. The role of AKT isoforms in glioblastoma: AKT3 delays tumor progression. J Neuro-Oncol. 2016;130(1):43–52. https://doi.org/10.1007/s11060-016-2220-z.
86. Guo C, Liu S, Greenaway F, Sun MZ. Potential role of annexin A7 in cancers. Clin Chim Acta. 2013;423:83–9. https://doi.org/10.1016/j.cca.2013.04.018.
87. Yadav AK. Monosomy of Chromosome 10 Associated With Dysregulation of Epidermal Growth Factor Signaling in Glioblastomas. JAMA. 2009;302(3):276. https://doi.org/10.1001/jama.2009.1022.
88. Chen J, Fu X, Wan Y, Wang Z, Jiang D, Shi L. miR-125b inhibitor enhance the chemosensitivity of glioblastoma stem cells to temozolomide by targeting Bak1. Tumor Biol. 2014;35(7):6293–302. https://doi.org/10.1007/s13277-014-1821-4.
89. Appin CL, Brat DJ. Molecular Genetics of Gliomas. Cancer J. 2014;20(1):66–72. https://doi.org/10.1097/PPO.0000000000000020.
90. Vuong HG, Altibi AMA, Duong UNP, Ngo HTT, Pham TQ, Fung KM, et al. BRAF Mutation is Associated with an Improved Survival in Glioma-a Systematic Review and Meta-analysis. Mol Neurobiol. 2017. https://doi.org/10.1007/s12035-017-0599-y.
91. Quiros S, Roos WP, Kaina B. Rad51 and BRCA2 - New Molecular Targets for Sensitizing Glioma Cells to Alkylating Anticancer Drugs. PLoS ONE. 2011;6(11):e27183. https://doi.org/10.1371/journal.pone.0027183.
92. Meimand SE, Pour-Rashidi A, Shahrbabak MM, Mohammadi E, Meimand FE, Rezaei N. The Prognostication Potential of BRCA Genes Expression in Gliomas: A Genetic Survival Analysis Study. World Neurosurg. 2022;157:e123–8. https://doi.org/10.1016/j.wneu.2021.09.107.
93. Huang S, Xiao J, Wu J, Liu J, Feng X, Yang C, et al. Tizoxanide Promotes Apoptosis in Glioblastoma by Inhibiting CDK1 Activity. Front Pharmacol. 2022;13:895573. https://doi.org/10.3389/fphar.2022.895573.
94. Bo L, Wei B, Li C, Wang Z, Gao Z, Miao Z. Identification of potential key genes associated with glioblastoma based on the gene expression profile. Oncol Lett. 2017;14(2):2045–52. https://doi.org/10.3892/ol.2017.6460.
95. Simon M, Ludwig M, Fimmers R, Mahlberg R, Müller-Erkwoh A, Köster G, et al. Variant of the CHEK2 gene as a prognostic marker in glioblastoma multiforme. Neurosurgery. 2006;59(5):1078–85. https://doi.org/10.1227/01.NEU.0000245590.08463.5B.
96. Apostolou P, Papasotiriou I. Current perspectives on CHEK2 mutations in breast cancer. Breast Cancer Targets Ther. 2017;9:331–5. https://doi.org/10.2147/BCTT.S111394.
97. Wu H, Yang JM, Jin S, Zhang H, Hait WN. Elongation Factor-2 Kinase Regulates Autophagy in Human Glioblastoma Cells. Cancer Res. 2006;66(6):3015–23. https://doi.org/10.1158/0008-5472.CAN-05-1554.
98. Liu XY, Zhang L, Wu J, Zhou L, Ren YJ, Yang WQ, et al. Inhibition of Elongation Factor-2 Kinase Augments the Antitumor Activity of Temozolomide against Glioma. PLoS ONE. 2013;8(11):e81345. https://doi.org/10.1371/journal.pone.0081345.
99. Oji Y, Tatsumi N, Fukuda M, Nakatsuka SI, Aoyagi S, Hirata E, et al. The translation elongation factor eEF2 is a novel tumor-associated antigen overexpressed in various types of cancers. Int J Oncol. 2014;44(5):1461–9. https://doi.org/10.3892/ijo.2014.2318.
100. Hauffe L, Picard D, Musa J, Remke M, Grünewald TGP, Rotblat B, et al. Eukaryotic translation initiation factor 4E binding protein 1 (EIF4EBP1) expression in glioblastoma is driven by ETS1- and MYBL2-dependent transcriptional activation. Cell Death Discov. 2022;8(1):91. https://doi.org/10.1038/s41420-022-00883-z.
101. Cha YL, Li PD, Yuan LJ, Zhang MY, Zhang YJ, Rao HL, et al. EIF4EBP1 Overexpression Is Associated with Poor Survival and Disease Progression in Patients with Hepatocellular Carcinoma. PLOS ONE. 2015;10(2):e0117493. https://doi.org/10.1371/journal.pone.0117493.
102. Tang J, Li Y, Liu B, Liang W, Hu S, Shi M, et al. Uncovering a Key Role of ETS1 on Vascular Abnormality in Glioblastoma. Pathol Oncol Res. 2021;27:1609997. https://doi.org/10.3389/pore.2021.1609997.
103. Tao BB, He H, Shi XH, Wang CL, Li WQ, Li B, et al. Up-regulation of USP2a and FASN in gliomas correlates strongly with glioma grade. J Clin Neurosci. 2013;20(5):717–20. https://doi.org/10.1016/j.jocn.2012.03.050.
104. Grube S, Dünisch P, Freitag D, Klausnitzer M, Sakr Y, Walter J, et al. Overexpression of fatty acid synthase in human gliomas correlates with the WHO tumor grade and inhibition with Orlistat reduces cell viability and triggers apoptosis. J Neuro-Oncol. 2014;118(2):277–87. https://doi.org/10.1007/s11060-014-1452-z.
105. Yasumoto Y, Miyazaki H, Vaidyan LK, Kagawa Y, Ebrahimi M, Yamamoto Y, et al. Inhibition of Fatty Acid Synthase Decreases Expression of Stemness Markers in Glioma Stem Cells. PLoS ONE. 2016;11(1):e0147717. https://doi.org/10.1371/journal.pone.0147717.
106. Song J, Zhao D, Sun G, Yang J, Lv Z, Jiao B. PTPRM methylation induced by FN1 promotes the development of glioblastoma by activating STAT3 signalling. Pharm Biol. 2021;59(1):902–9. https://doi.org/10.1080/13880209.2021.1944220.

107. Cheng SX, Tu Y, Zhang S. FoxM1 Promotes Glioma Cells Progression by Up-Regulating Anxa1 Expression. PLoS ONE. 2013;8(8):e72376. https://doi.org/10.1371/journal.pone.0072376.
108. Liu M, Dai B, Kang SH, Ban K, Huang FJ, Lang FF, et al. FoxM1B Is Overexpressed in Human Glioblastomas and Critically Regulates the Tumorigenicity of Glioma Cells. Cancer Res. 2006;66(7):3593–602. https://doi.org/10.1158/0008-5472.CAN-05-2912.
109. Gürsel DB, Banu MA, Berry N, Marongiu R, Burkhardt JK, Kobylarz K, et al. Tight regulation between cell survival and programmed cell death in GBM stem-like cells by EGFR/GSK3b/PP2A signaling. J Neuro-Oncol. 2015;121(1):19–29. https://doi.org/10.1007/s11060-014-1602-3.
110. Tsai HP, Lin CJ, Lieu AS, Chen YT, Tseng TT, Kwan AL, et al. Galectin-3 Mediates Tumor Progression in Astrocytoma by Regulating Glycogen Synthase Kinase-3β Activity. Curr Issues Mol Biol. 2023;45(4):3591–602. https://doi.org/10.3390/cimb45040234.
111. Majewska E, Szeliga M. AKT/GSK3β Signaling in Glioblastoma. Neurochem Res. 2017;42(3):918–24. https://doi.org/10.1007/s11064-016-2044-4.
112. Annibali D, Whitfield JR, Favuzzi E, Jauset T, Serrano E, Cuartas I, et al. Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis. Nat Commun. 2014;5(1):4632. https://doi.org/10.1038/ncomms5632.
113. Wang J, Wang H, Li Z, Wu Q, Lathia JD, McLendon RE, et al. c-Myc Is Required for Maintenance of Glioma Cancer Stem Cells. PLoS ONE. 2008;3(11):e3769. https://doi.org/10.1371/journal.pone.0003769.
114. Yan C, Su L, Qi X, Yang Y, Wu J. Oncogenic involvement of NF-kB signaling pathway in glioma. Cancer Res. 2020;80(16_Supplement):4735. https://doi.org/10.1158/1538-7445.AM2020-4735.
115. Concetti J, Wilson CL. NFKB1 and Cancer: Friend or Foe? Cells. 2018;7(9):133. https://doi.org/10.3390/cells7090133.
116. Musumeci G, Castorina A, Magro G, Cardile V, Castorina S, Ribatti D. Enhanced expression of CD31/platelet endothelial cell adhesion molecule 1 (PECAM1) correlates with hypoxia inducible factor-1 alpha (HIF-1α) in human glioblastoma multiforme. Exp Cell Res. 2015;339(2):407–16. https://doi.org/10.1016/j.yexcr.2015.09.007.
117. Cheng CK, Fan Q, Weiss WA. PI3K Signaling in Glioma-Animal Models and Therapeutic Challenges. Brain Pathol. 2009;19(1):112–20. https://doi.org/10.1111/j.1750-3639.2008.00233.x.
118. Zając A, et al. Involvement of PI3K Pathway in Glioma Cell Resistance to Temozolomide Treatment. Int J Mol Sci. 2021;22(10):5155. https://doi.org/10.3390/ijms22105155.
119. Langhans J, Schneele L, Trenkler N, Von Bandemer H, Nonnenmacher L, Karpel-Massler G, et al. The effects of PI3K-mediated signalling on glioblastoma cell behaviour. Oncogenesis. 2017;6(11):398. https://doi.org/10.1038/s41389-017-0004-8.
120. Ferro E, Bosia C, Campa CC. RAB11-Mediated Trafficking and Human Cancers: An Updated Review. Biology. 2021;10(1):26. https://doi.org/10.3390/biology10010026.
121. Mathivanan J, Rohini K, Gope ML, Anandh B, Gope R. Altered structure and deregulated expression of the tumor suppressor gene retinoblastoma (RB1) in human brain tumors. Mol Cell Biochem. 2007;302(1–2):67–77. https://doi.org/10.1007/s11010-007-9428-3.
122. Kim Y, Lachuer J, Mittelbronn M, Paulus W, Brokinkel B, Keyvani K, et al. Alterations in the RB1 Pathway in Low-grade Diffuse Gliomas Lacking Common Genetic Alterations. Brain Pathol. 2011;21(6):645–51. https://doi.org/10.1111/j.1750-3639.2011.00492.x.
123. Nakamura M, Yonekawa Y, Kleihues P, Ohgaki H. Promoter Hypermethylation of the RB1 Gene in Glioblastomas. Lab Investig. 2001;81(1):77–82. https://doi.org/10.1038/labinvest.3780213.
124. Tan X, Zou L, Qin J, Xia D, Zhou Y, Jin G, et al. SQSTM1/p62 is involved in docosahexaenoic acid-induced cellular autophagy in glioblastoma cell lines. In Vitro Cell Dev Biol Anim. 2019;55(9):703–12. https://doi.org/10.1007/s11626-019-00387-8.
125. Puissant A, Fenouille N, Auberger P. When autophagy meets cancer through p62/SQSTM1. Am J Cancer Res. 2012;2(4):397–413.
126. Moncayo G, Grzmil M, Smirnova T, Zmarz P, Huber RM, Hynx D, et al. SYK inhibition blocks proliferation and migration of glioma cells and modifies the tumor microenvironment. Neuro-Oncol. 2018;20(5):621–31. https://doi.org/10.1093/neuonc/noy008.
127. Zhou Q, Wei M, Shen W, Huang S, Fan J, Huang H. SYK Is Associated With Malignant Phenotype and Immune Checkpoints in Diffuse Glioma. Front Genet. 2022;13:899883. https://doi.org/10.3389/fgene.2022.899883.
128. Riemenschneider MJ, Betensky RA, Pasedag SM, Louis DN. AKT Activation in Human Glioblastomas Enhances Proliferation via TSC2 and S6 Kinase Signaling. Cancer Res. 2006;66(11):5618–23. https://doi.org/10.1158/0008-5472.CAN-06-0364.
129. Fettweis G, Di Valentin E, L'homme L, Lassence C, Dequiedt F, Fillet M, et al. RIP3 antagonizes a TSC2-mediated pro-survival pathway in glioblastoma cell death. Biochim Biophys Acta (BBA) - Mol Cell Res. 2017;1864(1):113–24. https://doi.org/10.1016/j.bbamcr.2016.10.014.
130. Gupta MK, Polisetty RV, Sharma R, Ganesh RA, Gowda H, Purohit AK, et al. Altered transcriptional regulatory proteins in glioblastoma and YBX1 as a potential regulator of tumor invasion. Sci Rep. 2019;9(1):10986. https://doi.org/10.1038/s41598-019-47360-9.

Publisher's Note

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