ALGORITHMIC RECONSTRUCTION OF GBM NETWORK COMPLEXITY

Abicumaran Uthamacumaran[1] and Morgan Craig[2,3]

[1]Department of Physics, Concordia University, Montreal, Canada
[2]Sainte-Justine University Hospital Research Centre, Montreal, Canada
[3]Département de mathématiques et de statistique, Université de Montréal, Montreal, Canada

ABSTRACT

Glioblastoma (GBM) is a complex disease that is difficult to treat. Establishing the complex genetic interactions regulating cell fate decisions in GBM can help to shed light on disease aggressivity and improved treatments. Networks and data science offer novel approaches to study gene expression patterns from single-cell datasets, helping to distinguish genes associated with control of differentiation and thus aggressivity. Here, we applied a host of data theoretic techniques, including clustering algorithms, Waddington landscape reconstruction, trajectory inference algorithms, and network approaches, to compare gene expression patterns between pediatric and adult GBM, and those of adult GSCs (glioma-derived stem cells) to identify the key molecular regulators of the complex networks driving GBM/GSC and predict their cell fate dynamics. Using these tools, we identified critical genes and transcription factors coordinating cell state transitions from stem-like to mature GBM phenotypes, including eight transcription factors (OLIG1/2, TAZ, GATA2, FOXG1, SOX6, SATB2, YY1) and four signaling genes (ATL3, MTSS1, EMP1, and TPT1) as clinically targetable novel putative function interactions differentiating pediatric and adult GBMs from adult GSCs. Our study is among the first to provide strong evidence of the applicability of complex systems approaches for reverse-engineering gene networks from patient-derived single-cell datasets and inferring their complex dynamics, bolstering the search for new clinically- relevant targets in GBM.

Keywords: Glioblastoma, Complex Systems, Networks, Data Science, Waddington landscape Reconstruction, Computational Oncology.

1. INTRODUCTION

Glioblastoma (GBM) is the most lethal pediatric and adult brain tumour. Despite advances in treatment, recurrence will occur in all GBM patients, and mean survival is only 15 months (Alifieris and Trafalis, 2015). GBM is a morbid disease that is driven by a high degree of heterogeneity and phenotypic plasticity in response to the interactions with their tumor microenvironment (Jung et al., 2019). The cell fate transitions and cellular decision-making in GBM cell populations are regulated by the dynamics of complex signaling networks (Suvà et al., 2014; Jia et al., 2017). Recent advances linking single-cell datasets and computational algorithms have improved our understanding of these complex networks and their orchestration of cell fate decisions of GBM transcriptional states (phenotypes) (Jin et al., 2018; Iacono et al., 2019). Despite this progress, quantitative approaches that reconstruct the information flow and dynamics of these complex networks remain under-applied. Pediatric GBM exhibits molecular patterns and behaviors which are fundamentally different from those of adult GBM (Paugh et al., 2010; Jones et al., 2017). There is a greater epigenetic burden in pediatric GBM marked by specific histone modifications and DNA methylation profiles (Lulla et al., 2016). However, the complex signaling dynamics distinguishing pediatric and adult GBM subgroups, and the similarities within the molecular networks driving their cancer stemness, remain poorly investigated (Paugh et al., 2010; Jones et al., 2017). Answering the question of whether the reconfiguration of these underlying signaling networks in

both GBM groups steers their cell fate dynamics would allow for the prediction of causal patterns in the disease progression and therapeutic responses.

Glioma-derived stem cells (GSCs) are believed to be a small subset of GBM cancer cells that largely contribute to emergent GBM adaptive behaviors such as phenotypic plasticity, clonal heterogeneity, self-renewal, aggressiveness (resilience), relapse/recurrence, and therapy resistance (Jung et al., 2019; Xiong et al., 2019). However, many different phenotypes in the tumor microenvironment, including immune cells, healthy cells, extracellular matrices, and blood vessels, form complex feedback loops with malignant GBM cells (Jung et al., 2019; Xiong et al., 2019). GSCs form complex networks with their tumor microenvironment. Signaling dynamics within this microenvironment and its reconfiguration govern the fitness and stemness of GSCs. A lack of quantitative understanding of the causal mechanisms (gene expression patterns) underlying GSC cell fate choices and transitions to their mature phenotypes hinders successful clinical interventions in the treatment of GBM (Jung et al., 2019; Xiong et al., 2019).

Traditionally, statistical approaches are used to study cell fate dynamics and infer complex networks from large-scale single cell transcriptomics by differential expression analysis through a combination of single cell data processing and clustering algorithms (Iacono et al., 2019). However, these algorithmic pipelines are inadequate for capturing the complex patterns and emergent behaviors of cancer network dynamics. Further, fundamental limitations associated with the raw counts of the scRNA-Seq complicate the inference of networks in complex diseases like GBM. These limitations include drop out events (zero counts) and the inherent noise and sparsity of single cell data. To extract quantitatively meaningful differences between GSC and GBM networks, while retaining the essential information representative of their complex dynamics, requires tools from the interdisciplinary paradigm of *complex systems theory*.

Complex systems theory, or complexity science, is the study of irreducible systems composed of many interacting parts in which the systems exhibit emergent behaviors. Emergence denotes systems in which the nonlinear interactions between the system and its environment give rise to complex patterns and unpredicted collective dynamics (Wolfram, 1988; Shalizi, 2006). Some general properties of complex systems include nonlinear dynamics, adaptive processes, self-organized structures, interconnectedness, collective behaviors, pattern formation, fractality, sudden phase-transitions, computational irreducibility, non-locality, long-term unpredictability, undecidability and multi-scaled, multi-nested feedback loops (Wolfram, 1988; Shalizi, 2006). The presence of multi-scaled feedback loops, in particular, is the defining feature of complex networks (Thurner et al., 2018). Traditional reductionist approaches are inadequate to quantify the properties and behaviors of complex networks (Wolfram, 1988; Shalizi, 2006). Complex systems theory advocates the use of computational algorithms and tools from network science to dissect these complex networks (Huang et al., 2009; Barabási and Oltvai, 2014; Thurner et al., 2018).

The molecular networks coordinating the emergence of GSC and GBM phenotypes are such complex networks. To reveal the mechanisms underlying GSC cell fate decisions and transitions to their mature GBM phenotypes, we deployed a number of approaches from complex systems theory on data from single-cell RNA Sequencing (scRNA-Seq) count matrices. We compared pediatric GBM to adult GBM to identify the signaling network patterns distinguishing pediatric and adult GBM from GSCs. For this, we relied upon clustering algorithms, Waddington landscape reconstruction, multivariate information theory, network science (graph theory), and machine learning algorithms to map possible *cell fate dynamics* and identify robust expression markers (critical TFs and genes) driving the complex networks underlying GBM/GSC cell fate dynamics. We found that distinct gene expression signatures regulate the cell fate decisions in the GBM and GSC patient groups we studied. In particular, we identified a set of key

gene targets as master regulators of cell fate decision dynamics in all patient groups, and the critical drivers of GSC stemness networks. Mapping their energy landscape dynamics and cell fate trajectories in pseudotime (cellular transition dimension), we represented the GSC/GBM cell fate decisions as dynamical systems which allowed us to identify genes such as GATA2, FOXG1, SATB2, YY1, and SOX6, amidst others, as master regulators of information flow in their signaling networks. Our results help to understand how cellular fate decisions in GBM, identify potential drug targets for precision oncology, and provide a roadmap for data theoretic approaches to other such complex systems.

2. METHODS

2.1. Single-cell datasets: Gene expression matrices for pediatric GBM, adult GBM, and adult GSC were obtained from the SingleCell Portal repositories from Neftel et al., 2019 and Richards et al., 2021. Briefly, GBM patient samples from Neftel et al. contained the single cell RNA-Seq counts of four phenotypes (or cellular states): macrophages, malignant GBM cells, oligodendrocytes, and T-cells. Adult GSC consisted only of stem cells. Overall, our dataset included 28 adult GSC datasets, 7 pediatric GBM, and 18 adult GBM scRNA-Seq expression count matrices. A description of the experimental approaches used to derive these datasets from their original studies is provided in the Supplementary Information.

2.2. Clustering techniques

2.2.1. Seurat algorithm: scRNA-Seq count matrices were pre-processed to obtain normalized and binarized count expressions. Seurat initially performs a cluster analysis by principal component analysis (PCA) dimensionality reduction followed by a graph-based clustering (sNN graph-shared nearest neighbor) and Louvain community detection algorithm (modularity optimization) to cluster cells by their expression similarity onto a Uniform Manifold Approximation and Projection (UMAP) or t-Distributed Stochastic Neighbor Embedding (TSNE) space (i.e., unsupervised nonlinear dimensionality reduction techniques) (Stuart et al., 2019). Positive markers were visualized in UMAP space. We clustered similarly expressed cells together in low dimensional space by finding differentially expressed features/markers corresponding to the highest PCA loadings in the graph-based clusters. We imposed the condition that selected markers for the network reconstruction must be expressed in all clusters of the three patient groups (pediatric GBM, adult GBM, and adult GSC).

2.2.2. BigSCale algorithm: BigSCale is a framework for clustering, phenotyping, pseudotiming, and inferring gene regulatory and protein-protein interaction networks from single-cell data (Iacono et al., 2019). A SingleCellExperiment class was created from the scRNA-Seq raw count matrices for BigSCale processing, and counts were replaced by z-scores from an unsupervised recursive clustering-based iterative differential expression analysis between clusters of cells. The markers specific to a cluster were sorted from the highest (most significant) to the lowest (least significant) z-score for the selection of cluster-specific differential and co-expressed gene markers within the top 25 PCA components. A z-score threshold of 5.0 was used for as a cut off threshold. This imposed cut-off acts as a filtering mechanism to retain only the markers with significant expression changes per cluster. As in the Seurat analysis, we imposed the condition that selected markers for the network reconstruction must be expressed in all clusters of the three patient groups.

2.2.3. ScEpath algorithm: The single cell Energy path (scEpath) is a Waddington Landscape reconstruction algorithm with an unsupervised clustering framework for cell lineage hierarchy mapping and studying the pseudotemporal transcriptional dynamics in cell fate decisions. scEpath is a trajectory inference algorithm for reconstructing the 3D-energy landscape of cells and infer regulatory relationships from their transcriptional dynamics (Jin et al., 2018). The information flow and reconfiguration of the networks underlying the cellular decision-making steers the topography of the cell population's energy landscape (also referred to as a fitness landscape, attractor landscape, or Waddington's epigenetic landscape (Waddington, 1957)). A cell state or rather a cell fate corresponds to a specific transcriptional (gene expression) program and phenotype of a given cellular population. Cell clusters higher on the energy landscape correspond to stem-cell like states (unstable attractors) with higher differentiation potency, while cell states stuck in lower energies (valleys, or stable attractors) correspond to differentiated (mature) phenotypes with lower potency/plasticity (Figure 1). scEpath allows for the visualization of cell fate transition probabilities in the population, mapping of cell lineage trajectories in pseudotemporal ordering, and inference of cell fate decisions from patient-derived scRNA-Seq datasets using the following steps: (i) preprocessing of scRNA-seq count matrix, (ii) gene regulatory network (GRN) inference, (iii) single cell energy (scEnergy) calculation, (iv) 3D energy landscape reconstruction via principal component analysis and structural clustering; (v) Transition probabilities calculation, (vi) Inference of cell lineage hierarchy via a probabilistic directed graph, (vii) pseudotime trajectory inference and, (viii) downstream analyses of identifying critical transcription factors (TFs) governing the cell-fate commitments (Jin et al., 2018). A detailed description of the scEpath algorithm is provided in the Supplementary Information.

To perform scEpath analysis on our data, we first pre-processed teh log-normalized count matrices with respect to their gene expression values by filtering out zero counts. We then ran the scEpath MATLAB code from Jin et al. (2018) on these processed datasets. GSC patient samples BT127, BT48, and BT84 were used for all scEpath analyses on GSC. Seven pediatric GBM samples from Neftel et al. (2019) and seven adult GBM samples, selected to match the cell count of the pediatric patient group, from Neftel et al. (2019), were analyzed. We confirmed that the number of patients did not influence the results and analysis by selecting different random sets of adult GBM samples. We then ran energy (Waddington) landscapes reconstruction on the following population sizes: pediatric GBM: n= 7, N= 1850 cells; adult GBM: and n = 7, N= 2221 cells; adult GSC: n=3, N=1504 cells.

scEpath smoothens the average expression of each gene using cubic regression splines, leading to smoothed gene expression along a lineage path, to map the pseudotemporal gene expression dynamics along the inferred trajectories of the cell fates on the landscape (Jin et al., 2018). We inferred key regulatory TFs for the cell fate differentiation by considering all PDG genes with a standard deviation > 0.5 and a Bonferroni-corrected p-value below a significance level $\alpha = 0.01$ for the expression greater than a threshold (e.g., log2(fold-change) > 1).

2.2.4. Fractal and multifractal analysis: Fractals are signatures of complex systems (Mandelbrot, 1982), and the fractal dimension is a non-integer, fractional dimension characterizing the statistical self-similarity and roughness of a pattern (Mandelbrot, 1982). A higher fractality in

tumor structures may imply that the tumor is more complex, resilient (i.e., withstands environmental perturbations), aggressive, and difficult to treat (Coffey, 1998; Jain and Baish, 1998). We used ImageJ plugin FracLac (v2.5) to compute the fractal dimension (FD) of analyzed samples using the BoxCount algorithm on the cell state attractors (patterns of cellular distributions on the scEpath energy landscapes). To calculate the fractal dimension, landscape images were converted to black and white. Attractor fractal dimensions reconstructed from the cell fate landscapes found to be non-integer were considered to exhibit a fractional dimension in phase-space. Higher fractal indices indicate more complex dynamics that are irregular and asymptotically unpredictable, since in dynamical systems theory, patterns of systems exhibiting deterministic chaos have a fractal dimension (i.e. strange attractors) (Strogatz, 2015).

2.2.5. Partial Information Decomposition and Context network inference: Using the differential expression markers identified by the various approaches discussed above, we reconstructed the underlying complex networks driving the GBM/GSC cell state dynamics on the Waddington energy landscapes. Network inference tools study the statistical dependencies between genes amidst distributions of expression levels in populations of sampled cells (Chan et al., 2017) by inferring a graph-theoretic representation of the functional relationships between the drivers of complex behaviors such as cell fate transitions. Partial Information Decomposition and Context (PIDC) networks have been suggested to outperform traditional gene regulatory network inference approaches using mutual information, Boolean networks, or Bayesian inference methods for network reconstruction (Chan et al., 2017). We used this PIDC network inference algorithm to obtain a network structure of GBM and GSC samples.

To reconstruct the GRN networks, we used the InformationMeasures.jl and NetworkInference.jl packages in Julia. PIDC network inference uses partial information decomposition (PID) to infer regulatory interaction networks from gene expression datasets. We used the NetworkInference package to establish the (undirected) networks from the multivariate information measure (PID) calculated from the gene expression matrices. Gene expression counts were first discretized via Bayesian blocks discretization and the maximum likelihood estimator (Chan et al., 2017).
The PIDC network *pattern* is the simplest network the algorithm can construct such that the distance between the nodes (genes or TFs) are minimized given their weights (PID score). Network measures characterizing the structure, properties, and information flow of these complex networks were extracted using the NetworkInference package on Julia LightGraphs. We then quantified the most differentially expressed genes identified by the clustering algorithms using PID scores.

2.2.6. Block Decomposition Method Calculations: The *algorithmic complexity* of key nodes (genes) of the inferred signaling networks were evaluated to further identify robust markers distinguishing GBM and GSC. The K-complexity of a string s, $K(s)$, also known as Kolmogorov or algorithmic complexity, is the shortest computer program length needed to output that string. This can also alternatively be interpreted as the length of the shortest description of a system (Zenil et al., 2016). Since $K(s)$ does not depend on a choice of probability distribution like Shannon entropy, it is more robust for the assessment of system complexity (Zenil et al., 2016). Formally, the Kolmogorov complexity of a discrete dynamical system is given by

$$K(s|e) = \min\{|p|: U(p,e) = s\},$$

for a string or array s, where p is the program that produces s and halts running on a universal Turing machine U with input e. Then, K(s) is a function that takes a string or matrix s to be the length of the shortest program p that generates s. However, $K(s)$ *is* in principle incomputable and must be approximated. We therefore used the block decomposition method (BDM) to approximate the $K(s)$ of a dataset, which provides local estimates of the algorithmic complexity [Zenil et al., 2016]. BDM is available in the online algorithmic complexity calculator [OACC] and its R-implementation (see Data and Code Availability). The BDM is defined as

$$BDM = \sum_{i=1}^{n} K(block_i) + log_2(|block_i|),$$

where the block size must be specified for the n-number of blocks. When the block sizes are higher, better approximations of the K-complexity are obtained (Zenil et al., 2016; Zenil et al., 2019).

To calculate the BDM, we selected scRNA-Seq counts of seven randomly chosen patient samples from each of the three patient groups. String length was kept the same for all gene candidates from each sample. Accordingly, we chose the cell count expressions of 46 cells from each patient sample for this analysis. The R-implementation of the Online Algorithmic Complexity Calculator was used to compute the BDM estimates of K-complexity for each expression string. scNA-Seq counts of the top gene interactions with highest PID scores were selected from each network and binarized. We then performed BDM on these binarized strings using a block size of 12 and alphabet size of 2 bits to estimate the K-complexity (i.e., BDM score).

2.3. Machine learning: Binary classification was performed using Google Colab using Scikit-learn on the BDM scores of three classification groups (pediatric GBM, adult GBM, and GSC). The linear support vector machine (SVM) and AdaBoost random forest (RF) classifier modules from Scikit-learn were trained using both an 80:20 and a 50:50 training: testing size split with the BDM scores of the gene markers with the top PID scores identified in the PIDC networks for both the GRNs and TF networks. Seven patient samples were selected for each of the three groups for the classification training and validation, as defined above. GraphPad Prism 8.4.3 was used for additional statistical analyses.

3. RESULTS

3.1. *Key driver genes mediating the cell fate transition dynamics in GBM/GSC epigenetic landscapes are identified using the scEpath algorithm.*

The Waddington landscape reconstruction identified causal patterns (attractors) to which the distinct transcriptional states within each patient group cluster (Fig 1A-C). Many genes encoding transcription factors (TFs) were identified as the transition genes required for cells to transition from one attractor to another. We mapped the expressions of these transition genes across the inferred cell fate trajectories (Fig 1D-F) and found similarities in the gene expression signatures and similar oscillatory patterns in EMP1, MTSS1, PHGDH and OLIG1/2 (Fig 2). Previous studies have shown that EMP1 facilitates the aggressivity and metastatic potential of GBM cells potentially by activation of the PI3K/AKT/mTOR signaling pathway (Miao et al., 2019). MTSS1 is a metastasis suppressor gene, the methylation profile of which has been shown to steer GBM migration (Luxen et al., 2017). Further, MTSS1 is involved in regulating EMT transitions and hence, regulating GSC stemness in GBM populations (Yu et al., 2020). PHGDH encodes the enzyme phosphoglycerate dehydrogenase and is involved in serine biosynthesis,

providing a glia-neuron cross- talk (Ehmsen et al., 2013). Given that PHGDH requires NAD^+ as a cofactor, there may be a link between PHGDH expression and the promotion of glycolytic GBM phenotypes and hypoxic tumor microenvironments (Engel et al., 2020). Moreover, the PHGDH-FOXM1 axis has been shown to promote glioma invasion (Liu et al., 2013). Lastly, OLIG1/2 are master neurodevelopmental transcription factors essential for GBM propagation and GSC stemness (Suvà et al., 2014). Together, these findings suggest that reconstructing the underlying regulatory network of these TFs and assessing their statistical properties via network science measures may provide key insights into the information flow across these target genes.

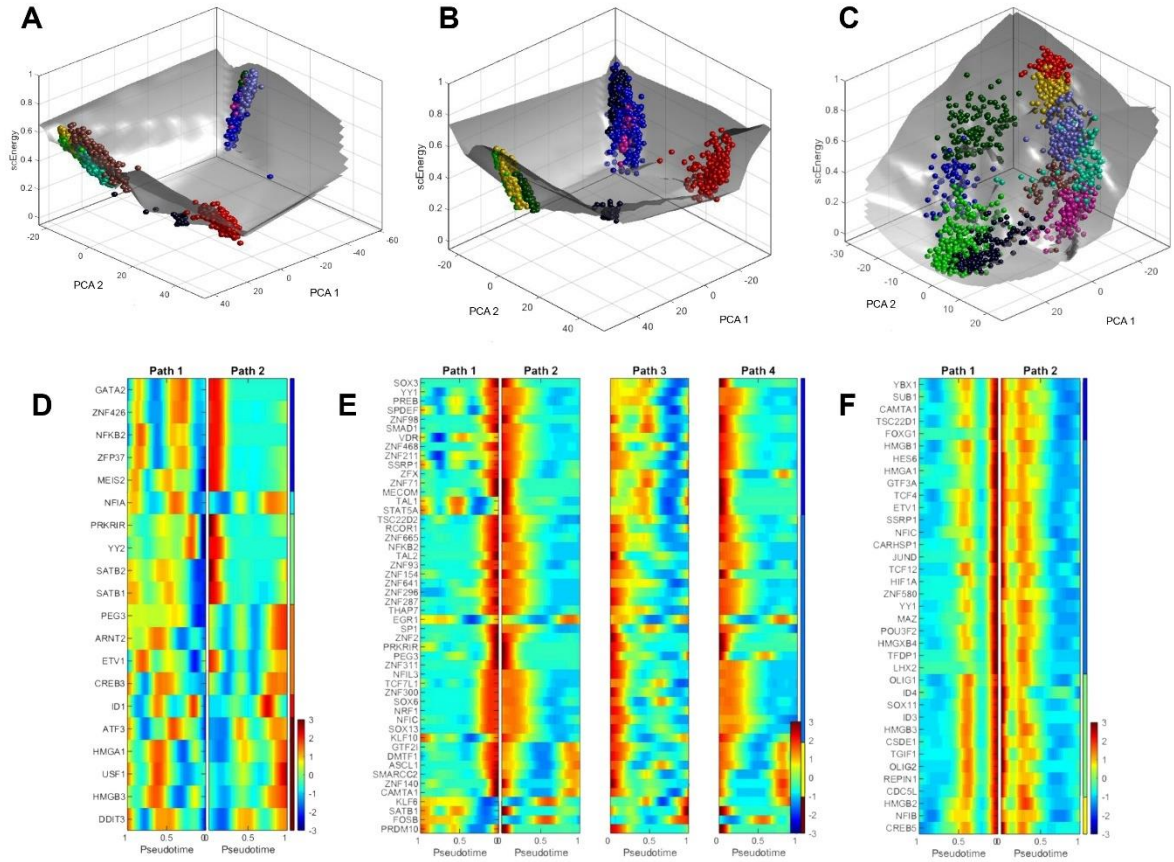

FIGURE 1. Waddington landscape reconstruction differentiates adult and pediatric GBM, and adult GSC critical genes and transcription factors for cell fate transitions. A-C) The distinct phenotypes of each patient group were clustered on the Waddington energy landscape by their similarity in gene expression. The cell clustering patterns on the landscape are referred to as *attractors*. Balls represent the transcriptional states (cell fates) and paths correspond to the cell fate differentiation trajectories on the Waddington landscape. D-F) Heat maps of the critical transcription factors involved in the differentiation and cell fate transitions between the distinct attractors (phenotype clusters) of the GBM/GSC Waddington landscape. The color gradient represents the intensity of the gene expression in pseudotime trajectory, where blue implies low expression and red implies high expression of the gene

(TF) during the cell fate choices along the cell differentiation trajectories. The path corresponds to the inferred trajectories in between the cell state attractors on the Waddington landscape.

3.2. *Pseudotime expression dynamics identifies oscillatory patterns in critical gene targets.*

Given the key driver genes and transcription factors identified by scEpath trajectory inference, we next sought to infer similarities in gene expression dynamics during cell fate transitions within each patient group amidst the identified critical gene markers. Using the above-discussed clustering algorithms (see Methods), we found that PTPRZ1 and S100B showed nearly identical expression dynamics in pediatric GBM along both cell fate trajectories on the Waddington landscape, whereas genes such as EMP1, MTSS1, and PHGDH had more complex dynamics and exhibited oscillations during cell fate dynamics in pediatric GBM and adult GSC (Fig 2A and Fig 2C). PTPRZ1 encodes a member of the receptor protein tyrosine phosphatase involved in developmental processes of the central nervous system (Harroch et al., 2002). Studies have shown PTPRZ1 promotes GBM invasion, coordinates GBM cellular homeostasis, and regulates the cross- talk between tumor-associated macrophages (TAM) and GSC cells (Shi et al., 2017; Bhaduri et al., 2020). S100B encodes a glial-specific calcium-binding protein involved in complex processes such as cell cycle progression and astrocyte differentiation (Brozzi et al., 2009). Interestingly, S100B also has been shown to regulate glioma growth via TAM-associated chemoattraction (Wang et al., 2013). Our findings thus suggest a possible regulatory feedback loop between PTPRZ1 and S100B in GBM cell fate dynamics.

In adult GBM, NACA and PABPC1, and TPT1 and PSAP had similarity in expression patterns across all four differentiation paths (Fig 2B). S100B, OLIG1, and PHGDH all had a broad expression profile in path 4 (Fig 2B). Furthermore, the presence of four cell clusters in adult GBM landscape (Fig 1B) is in good agreement with previous classifications of four molecular subtypes of adult GBM (Verhaak et al., 2010). The expression of EGFR and PDGFRA were distinctly higher in one of the four cell fate clusters/attractors (Figure S4B). However, the expression of IDH1 exhibited oscillatory dynamics in all four paths/attractors (data not shown). In adult GSC, many of the identified markers had similar gene expression profiles in pseudotemporal ordering (Fig 2C). For instance, PTPRZ1, NACA and PABPC1, were all found to have similar expression dynamics in both transition paths (Fig 2C). Notably, OLIG 1 and OLIG2 were found to have similar expression patterns in all three patient groups across all cell fate transition trajectories of the landscape (Fig 2A-C).

Notably, we identified that genes such as STMN3, MTSS1 and TAZ are critical regulators in one transition pathway, while PSAP, TPT1, and PTPRZ1 are relevant for the other transition trajectory on the pediatric GBM's Waddington landscape (Fig 2A and S4A). The same trends in pseudotemporal gene expression patterns in STMN3 and PTPRZ1 have also been found in the adult GSC cell fate trajectories (see Fig S4C in the Supplementary Information). In all three patient groups, OLIG1, OLIG2, PHGDH, and TIMELESS had similar expression profiles within the distinct cell fate transition paths indicating potentially some network coordination or collective oscillations. Some signals (e.g., BCAN and CLU) were found to exhibit oscillations that may be indicative of complex dynamics with time-series expression analysis (Supplementary Information). These findings suggest that the identified markers involved in GBM/GSC cell fate decisions exhibit similar patterns in their expression dynamics, and that the identified critical genes are functionally putative master orchestrators of cell fate transitions/differentiation of the heterogeneous phenotypes within a GBM patient's tumor.

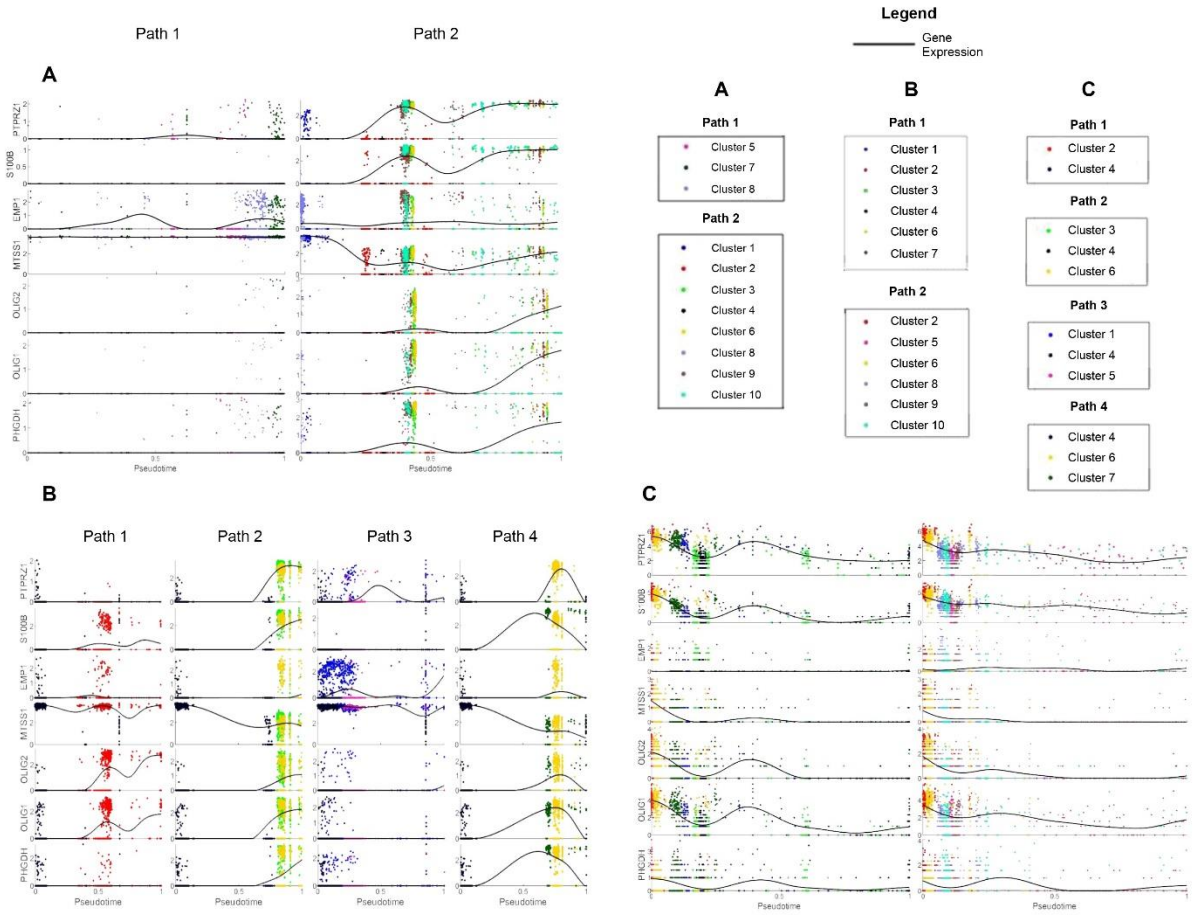

FIGURE 2. Reconstructing pseudotime dynamics in GBM/GSC cell fate decisions of the Waddington landscape. Gene expression in cells plotted along pseudotime after fitting with a cubic smoothing spline (black line). Cells are colored according to cell clusters defined by scEpath. The expression patterns of the top genes identified by scEpath and BigSCale algorithms (via correlation metrics) showed significant changes along the pseudotime trajectory inferred by scEpath algorithm. Selected gene markers in A) pediatric GBM, B) adult GBM, C) adult GSC.

3.3. *PIDC Network Inference algorithm reconstructs the regulatory network configurations driving GBM/GSC cell fate transitions.*

We next reverse engineered the signaling networks coordinating the information flow in GBM and GSC using Partial Information Decomposition and Context (PIDC). Though the network topography may seem similar, the arrangement of the interactions from highest influence on the information flow (i.e., top PID scores) to those of the weakest interactions (lowest PID scores) vary for each patient group. As seen in Figure 3A, OLIG1 and OLIG2 have the highest PID score of 1.9508, followed by S100B and PTPRZ1 interaction with a PID score of 1.9303 in pediatric GBM, suggesting a strong relationship between these two genes in the complex network steering their cell fate decisions (Fig 3A). We found that S100A10 and EMP1 have the highest interaction in adult GBM with a PID score of 1.9517 (Fig 3B), whereas NACA and TPT1 had the highest interaction in adult GSC with a PID score of 1.9628 (Fig 3C). A distinct pattern was observed in the PIDC regulatory network of adult GSC sample BT127 (highest quality GSC cells). The

highest interaction was observed between PHGDH and TIMELESS at a PID score of 2.762. Other top interactions identified for the TF networks (Fig 3E-G) had similar pseudotemporal expression dynamics (Figure S4 A-C in the Supplementary Information). ATF3 and DDIT3 were the top interaction markers from the critical TFs identified for pediatric GBM with a PID score of 1.971 (Fig 3E). EGR1 and FOSB in the adult GBM group (Fig 3F), and YBX1 and HMGB1 were identified as the top interaction TF markers, with PID score of 1.992 (Fig 3G). These results suggest the reconfiguration of the nodes within the same complex signaling network may characterize GSC cells from GBM cells and distinguish pediatric GBM from adult GBM cell fate dynamics.

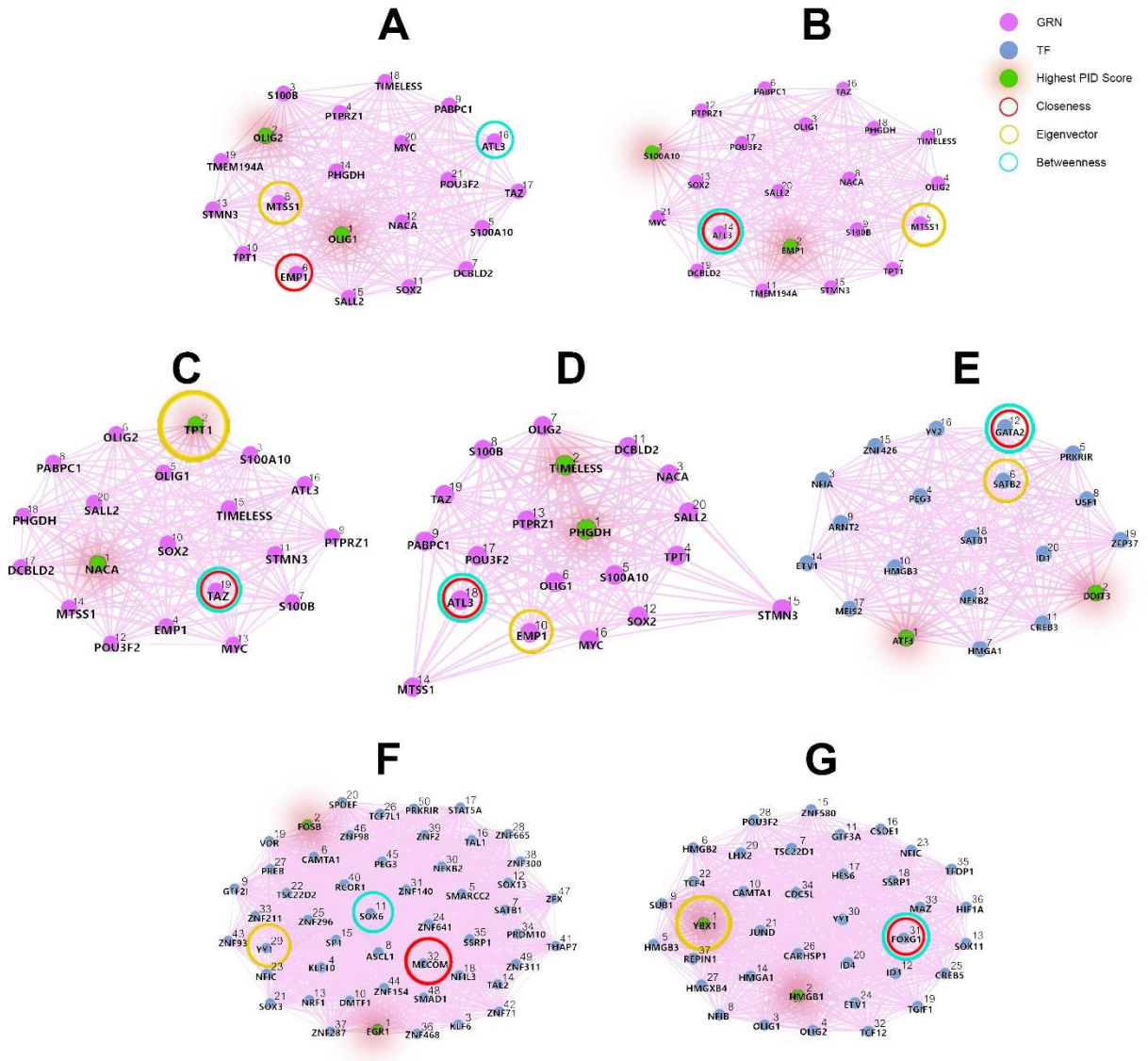

FIGURE 3. Mathematical modelling identifies key regulatory genes driving GBM networks. Gene regulatory networks of A) pediatric GBM, B) adult GBM, C) adult GSC, D) adult GSC sample BT127, E) pediatric GBM transcription factors, F) adult GBM transcription factors, and E) adult GSC transcription factors. In each, the signaling networks show the information flow between critical signals required for the complex cell fate dynamics. The GRN networks identified by Seurat and BigSCale are colored in violet

nodes (Fig 3A-D) while the scEpath TF networks are colored in teal (Fig 3E-G). The ranks were assigned a priority index by the PID content as indicated by the numbers on the nodes. A high PID content implies a high mutual information (dependence) of those gene interactions in the information flow network. The number index on the nodes of the network correspond to the PID score in a decreasing order, where rank 1 denotes the top (highest) value. As shown in the legend, the nodes with the highest PID score are colored in green with a red shadow. Additionally, three different colored rings are used to identify the nodes of the networks with the highest network centrality measures as identified in Figure 4.

3.4. *Network centrality measures identify master regulators of information flow across the regulatory networks underlying GBM/GSC cell fate decision-making.*

Centrality is a key property of complex networks that influences the network dynamics and information flow (Iacono et al., 2019). The nodes (genes or TFs) with the highest centrality in the regulatory networks are the most biologically important signals. By measuring network centrality, we identified the primary genes regulating communication flow across each of the pediatric and adult GBM, and adult GSC networks (Table 1). In particular, we calculated the global clustering coefficient that measures the total number of closed triangles (link density) in a network. A clustering coefficient at its maximal value of 1 indicates that the neighbors of the gene (node) i form a complete graph (i.e., they all connect to each other) versus the converse for a clustering coefficient of 0 (Barabási and Posfai, 2016). We observed a lower clustering coefficient of 0.94 for the BT127 network in Figure 3D. In the transcription factor networks reconstructed from the scEpath heatmaps (violet columns, Table 1), the GSC TF network had the highest diameter, while the GBM networks (both pediatric and adult) had smaller diameters. The diameter is relatively in the same order of magnitude for the PIDC networks reconstructed from the Seurat-BigSCale markers (pink columns, Table 1) as they correspond essentially to the same set of genes interactions. The degree of centrality of all networks in Figure 3 was 1.0 at all nodes, except for the BT127 PIDC network which had a degree centrality of value of 1.0 only at nodes 1, 5, 10, 12, 13, and 16, and a clustering coefficient of value 0.96. The degree centrality of nodes 2, 7, and 8 were 0.89, the degree centrality of nodes 14 and 15 were roughly 0.5, and the remaining nodes had a degree centrality of 0.95.

The closeness centrality identified genes/TFs occupying a central position in a network (Iacono, 2019). The nodes corresponding to the highest closeness centrality for each GRN network were found to be: Node 6 (EMP1) for pediatric GBM, Node 14 (ATL3) for adult GBM, Node 18 (ATL3) for GSC BT127, and Node 19 (TAZ) for GSC with closeness values of 1.398, 1.361, 1.006, and 1.184, respectively (Fig 4A). It has been established that TAZ via its interaction with EGFR is critical for GBM therapy resistance due to its ability to promote cancer stemness and phenotypic plasticity (Gao et al., 2021). TAZ is a role player in promoting phenotypic plasticity and differentiation to mesenchymal phenotypes via the Epithelial-Mesenchymal Transition (EMT) pathway (Bhat et al., 2011). Cancer cells undergoing EMT can acquire cancer stemness, and hence GBM cells expressing TAZ may acquire GSC properties and behaviors (Mani et al., 2008). TAZ may further promote hybrid OXPHOS/glycolytic phenotypes providing a metabolic advantage to adapt to the fluctuant microenvironment (Jia et al., 2017).

Amongst all identified critical network dynamics regulators, ATL3 may be the most influential as it was found to have the highest score for most of the centralities across all patient groups. ATL3 encodes integral membrane GTPases involved in the formation of endoplasmic reticulum networks and calcium signaling homeostasis (Polisetty et al., 2012). ATL3 has also been identified as a differentially expressed membrane marker in GBM (Polisetty et al., 2012).

Nodes corresponding to the maximal closeness in the pediatric GBM, adult GBM, and adult GSC TF networks were found to be node 12 (GATA2), node 32 (MECOM), and node 31 (FOXG1), respectively with closeness measures of 1.761, 2.563, and 1.478 respectively (Fig 4B). GATA2 is a hematopoietic transcription factor essential to the maintenance and differentiation of hematopoietic stem cells that has been found to associated to hematopoietic malignancies (Rodrigues et al., 2012). GATA2 has been shown to promote glioma progression through the EGFR/ERK pathway (Wang et al., 2015), and MECOM is an oncogenic TF known for its downstream signaling with TGF-β and cell cycle regulators that is associated with poor GBM prognosis (Hou et al., 2016). Lastly, FOXG1 is a neurodevelopmental TF highly expressed in GBM cells. Elevated expressions of FOXG1 and SOX2 promote GSC phenotypes in GBM (Bulstrode et al., 2017).

Betweenness centrality indicates the presence of regulatory bottlenecks (Iacono et al., 2019). In our analyses, the highest betweenness measures for the pediatric GBM, adult GBM, BT127 adult GSC, and adult GSC GRN networks were node 16 (ATL3), node 14 (ATL3), node 18 (ATL3), and node 19 (TAZ), respectively with betweenness values of 0.3947, 0.5842, 0.2690, and 0.4678, respectively (Fig 4C). The trends in maximal betweenness values (Fig 4C) were in good agreement with the nodes contributing to the maximal closeness values discussed in Fig 4A, indicating that identified nodes are critical targets governing the information flow in these complex networks. The highest betweenness values for the TF networks were found to be to node 12 (GATA2) for pediatric GBM, node 11 (SOX6) for adult GBM, and node 31 (FOXG1) for adult GSC, with values of 0.3801, 0.2279, and 0.1539, respectively (Fig 4D). The highest values of eigenvector centrality, a measure of information flow across the network, for the GRNs were found to be node 8 (MTSS1) for pediatric GBM, node 5 (MTSS1) for adult GBM, node 10 (EMP1) for BT127, and node 2 (TPT1) for GSC, with measures of 0.2796, 0.2827, 0.2909, and 0.2805, respectively.

SOX6 has functional relevance in the normal development of the central nervous system (Azim et al., 2009), and the role of SOX family members of TFs in cancers and maintenance/control of pluripotent stem cell fates is well-established (Grimm et al., 2020). S100B, one of the markers identified in our differential expression analysis of GBM cells, is a transcriptional target of SOX6 (Saito et al., 2007). TPT1 is a cancer-associated gene that encodes a protein known to promote glioma growth and progression, primarily via calcium binding and microtubule-cytoskeletal dynamics (Gu et al., 2014). Further, TPT1 was established as essential for activating pluripotency (stemness) genes such as Oct-4 and Nanog upon nuclear transfer (Koziol and Gurdon, 2007). Previous findings have shown TPT1 is key to phenotypic reprogramming in cancers and hence, regulates complex adaptive processes such as tumor reversion (Amson et al., 2013). Hence, our findings suggest TPT1 promotes GSC stemness.

The maximal eigenvector centrality of the TF networks were found to be node 6 (SATB2) for pediatric GBM, and node 29 (YY1) for adult GBM, and node 1 (YBX1) for GSC, with values of 0.2594, 0.1874, and 0.2322, respectively. SATB2 is a nuclear matrix-associated protein involved in chromatin remodelling and transcription regulation during neuronal differentiation (Gyorgy et al., 2008). Recently, SATB2 was shown as a driver of GBM progression preferentially expressed by GSCs (Tao et al., 2020). SATB2 may be a critical gatekeeper of cancer stemness in GBM cell populations (Roy et al., 2020). Hence, it should be further investigated as a clinical drug target. The overexpression of the transcription factor YY1 has been reported in GBM and shown to promote chemoresistance (Baritaki et al., 2009), and it has been further established that YY1 interacts with NF-kB dependent pro-inflammatory signaling pathways in GBM-

immune interactions (Sarvagalla et al., 2019). Further, EMP1 and MTSS1 both genes have a TF-binding site for YY1 (Stelzer et al., 2016; GeneCard Database, GCID: GC12P013196; GC08M124550).

We also performed fractal analysis on the attractors (cell clustering patterns) in the scEpath Waddington landscapes. The fractal dimension scores obtained on the cell state attractors on the energy landscape were compared across all groups (pediatric GBM (n=7), adult GBM (n = 18), and GSC (n=28)). The mean fractal dimension scores of the pediatric GBM, adult GBM, and adult GSC groups were 1.502 ± 0.099, 1.509 ± 0.091, and 1.588 ± 0.051, respectively (Figure 4G). The FD scores of the two GBM groups were nearly identical, while a statistically significant difference was observed from the GSC group. The multifractal spectrum $f(\alpha)$ was extracted from the multifractal spectra of the individual cancer samples energy landscape (n=54) (Fig 4H). Only the difference between GSC versus adult GBM was found to be statistically significant (p=0.0201) by a Kolmogorov-Smirnov test. The pediatric GBM, and adult GBM and GSC groups had a maximal multifractal spectrum $f(\alpha)$ value of 1.499 ± 0.092, 1.462 ± 0.066, and 1.521 ± 0.075, respectively.

Network Properties	Pediatric GBM	Adult GBM	Adult GSC	BT127 (Adult GSC)	Pediatric GBM (TF Network)	Adult GBM (TF Network)	Adult GSC (TF Network)
$G(V,E)$	(21,210)	(21,210)	(20,190)	(20,174)	(20,190)	(50, 1225)	(37,666)
Center	16	14	19	18	8	15	31
Diameter	3.037	3.283	3.490	3.106	1.869	1.474	3.936
Global Clustering Coefficient	1.00	1.00	1.00	0.94	1.00	1.00	1.00

TABLE 1. General Properties of Inferred Complex Networks. $G(V,E)$ denotes the graph with the number of vertices V (the genes) and number of edges E for each inferred GRN network. The center values designate to the node index (gene) acting as the center of the simple weighted network. The clustering coefficient captures the degree to which the neighbors of a given node link to each other.

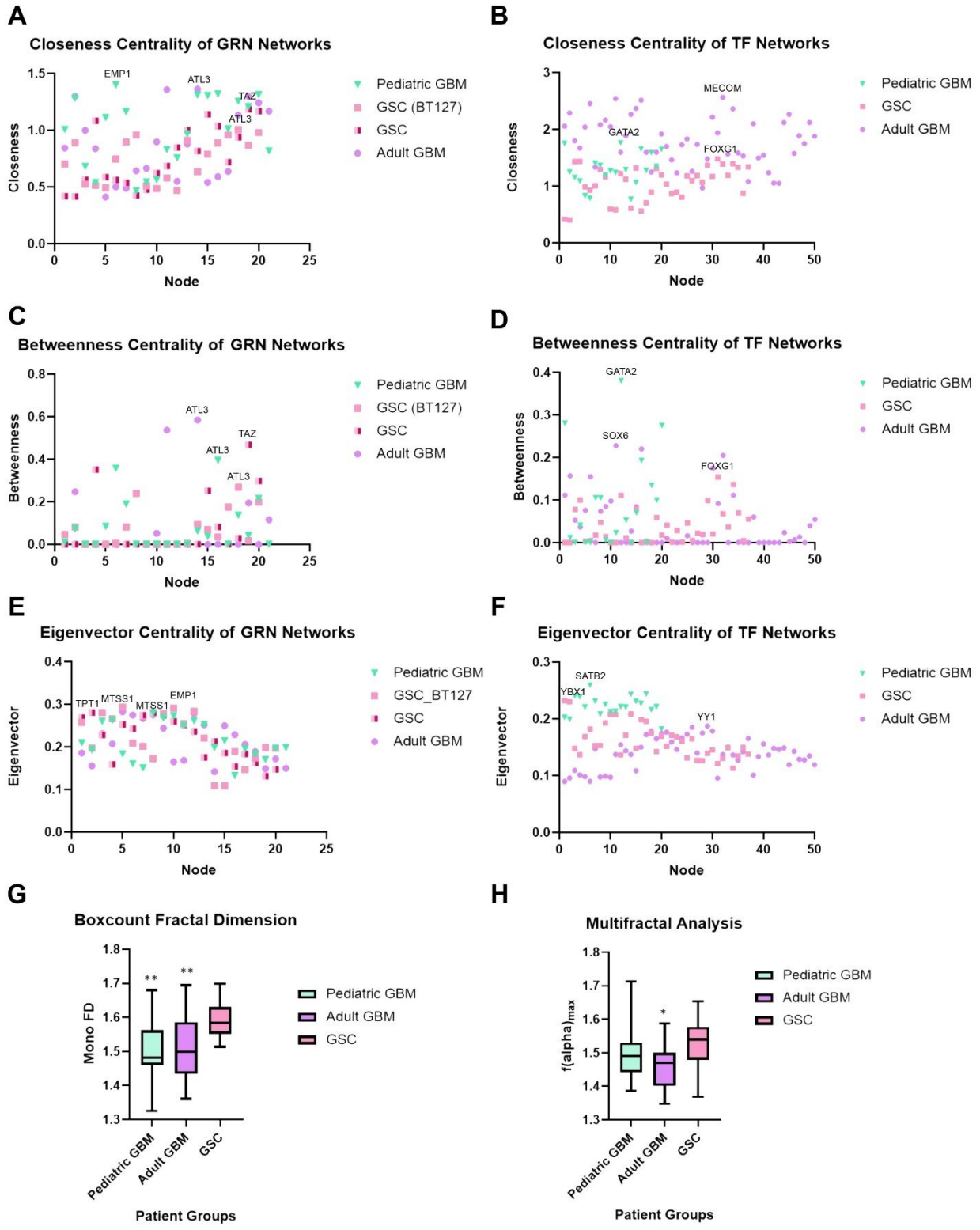
A
Closeness Centrality of GRN Networks
Closeness
Node
EMP1
ATL3
TAZ
ATL3
Pediatric GBM
GSC (BT127)
GSC
Adult GBM
B
Closeness Centrality of TF Networks
Closeness
Node
MECOM
GATA2
FOXG1
Pediatric GBM
GSC
Adult GBM
C
Betweenness Centrality of GRN Networks
Betweenness
Node
ATL3
TAZ
ATL3
ATL3
Pediatric GBM
GSC (BT127)
GSC
Adult GBM
D
Betweenness Centrality of TF Networks
Betweenness
Node
GATA2
SOX6
FOXG1
Pediatric GBM
GSC
Adult GBM
E
Eigenvector Centrality of GRN Networks
Eigenvector
Node
TPT1
MTSS1
MTSS1
EMP1
Pediatric GBM
GSC_BT127
GSC
Adult GBM
F
Eigenvector Centrality of TF Networks
Eigenvector
Node
SATB2
YBX1
YY1
Pediatric GBM
GSC
Adult GBM
G
Boxcount Fractal Dimension
Mono FD
Patient Groups
Pediatric GBM
Adult GBM
GSC
H
Multifractal Analysis
f(alpha)max
Patient Groups
Pediatric GBM
Adult GBM
GSC

FIGURE 4. Centrality measures distinguish master regulators of information flow in GBM networks. *Three network centrality measures are assessed on the reconstructed GBM/GSC networks.* Three different network centralities were computed on the reconstructed networks: closeness, betweenness, and eigenvector centrality. The genes (nodes) occupying the highest of these centrality measures correspond to critical nodes steering the information flow in the complex signaling networks governing GBM/GSC cell fate transition dynamics (Latora, 2017; Rodrigues, 2019). A) Closeness centrality of inferred GRNs. B) Closeness centrality of TF networks. C) Betweenness centrality of gene regulatory networks. D) Betweenness centrality of transcription factor networks. E) Eigenvector centrality of gene regulatory networks. F) Eigenvector centrality of transcription factor networks. G) Fractal dimension of cell state attractors on scEpath energy landscapes. A p-value of 0.0031 between the adult GSC and adult GBM, and p= 0.0011 between adult GSC and pediatric GBM was calculated for the box-count algorithm's fractal dimension scores using the Kolmogorov-Smirnov test. Multifractal analysis of cell fate attractors on scEpath Waddington landscapes.

DISCUSSION

Here we applied a collection of data theoretic and complexity science approaches to pediatric and adult GBM, and adult GSC single cell RNA-seq data to distinguish genes regulating communication within these cellular populations. Our findings are among the first to demonstrate the application of these tools for deciphering GBM/GSC signaling networks, which is critical for understanding how their configuration orchestrates information flow and determines cell fate dynamics. Thus, this proof-of-concept study suggests that complex systems approaches can provide insights into the quantitative differences between GBM and GSC gene expression dynamics during cell fate transitions/differentiation.

In pediatric GBM, the expression of transcription factors in pseudotime were shown to be highly nonlinear and even oscillatory for certain genes, including GATA2 in one trajectory while in an increasing or decreasing gradient of expression along the other cell fate trajectory. Likewise, patterns of other critical transition genes (TFs) were identified along the attractor dynamics between the distinct transcriptional states of adult GBM and adult GSC cells. Further, we found that genes such as EMP1, MTSS1, PTPRZ1 and S100B exhibit distinct gene expression oscillations in one differentiation trajectory (path) over the other(s) (Figure 2). These genes were also found to have TF-binding sites for the scEpath identified TFs with the highest network centrality measures in our downstream analysis (Figure 3). Together, these findings are indicative of a highly interconnected network of gene-TF interactions governing GBM/GSC cell fate decisions, and further suggest that the information flow across the inferred networks may steer cell fate decisions towards complex attractors on the GBM/GSC Waddington landscape.

The abnormally high PIDC scores for the TIMELESS and PHGDH interaction observed in the BT127 adult GSC sample suggests a possible mechanism distinguishing GSC from GBM. PHGDH is a key enzyme in metabolic energy transduction that is also involved in chemoresistance and promoting glycolytic phenotypes (tumor metabolism regulation) (Engel et al., 2020). Given that PHGDH is expressed highly in developing glial and astrocytic cells (Yamasaki et al., 2001), this interaction may be the primary source of energy used by GSCs, namely, oxyglutarate or pyruvate derivatives, as predicted by the Warburg effect (Liu et al., 2013).

Using network centrality measures, we identified OLIG1/2, TAZ, GATA2, FOXG1, SOX6, SATB2, YY1, and gene targets ATL3, MTSS1, EMP1, and TPT1 as critical genes governing the cell fate dynamics of GBM

and GSC cells (Fig 4A-F). Many of these signals are neurodevelopmental transcription factors essential in conferring and maintaining cancer stem cells (GSCs), and maximal centrality scores indicated that they are key regulators of the network information flow in both GBM groups and GSCs. The functional significance of these transcription factors (see Supplementary Information) suggests their critical role in stem cell decision-making and differentiation. Our findings suggest that these genes may be strong candidates for therapeutic interventions points for the treatment of GBM. Other signaling interactions such as PTPRZ1 and S100B were identified in our analyses as potent clinically druggable targets in the treatment of GBM. Furthermore, we predict GATA2 may provide a common ground for interlinking leukemogenesis, the complex signaling dynamics of leukemia/lymphoma affecting children, and pediatric glioma/GBM (Menendez-Gonzalez et al., 2019).

Interestingly, ATL3, whose function remains poorly investigated, occupied the highest number of maximal centrality measures across all patient groups. Further, given many of the identified differentially expressed signals in our GBM networks are involved in calcium signaling, we speculate that ATL3 may serve as a hub for regulating ER stress and extracellular vesicles transport mediating GBM cell-cell communication networks. Moreover, identified gene candidates play functional roles in the epigenetic regulation of GBM via signature histone modifications and DNA methylation profiles. For example, recurrent H3F3A mutations of two critical amino acids (K27 and G34) of histone H3.3 in one-third of pediatric GBM are regulated by OLIG1/2 and FOXG1 (Sturm et al., 2012), and YY1 interacts with many epigenetic modifiers including those involved in histone acetylation and chromatin remodelling (Sarvagalla et al., 2019).

A limitation of our study is that we did not have access to pediatric GSC cells, given that adult GSC data have only recently been described (Richards et al., 2021). Additionally, there may be other hidden causal interactions interconnecting the nodes of the complex networks we inferred that were not identified due to lack of data. Further, the lack of time-series scRNA-Seq counts is a barrier to understanding the complex dynamics of GBM/GSC networks. Ribosomal proteins and certain cytoskeletal markers (housekeeping genes) were also not pooled with the differential expression signatures for network inference (Figure S1).

This study demonstrates the use of complex systems approaches in deciphering the cybernetics of GBM/GSC networks, and shows how signaling dynamics differ between pediatric GBM, adult GBM, and adult GSC populations. By identifying transcription factors and genes, our combined approach serves as one part of the precision medicine toolbox for the treatment of GBM, suggesting both therapeutic targets and GBM reprogramming factors.

***DATA AND CODE AVAILABILITY**

Algorithms:

https://github.com/iaconogi/BigSCale2
http://complexitycalculator.com/
https://github.com/algorithmicnaturelab/OACC
https://github.com/Tchanders/NetworkInference.jl
https://juliagraphs.org/LightGraphs.jl/v1.3/linalg/

Pedregosa *et al.,* Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

http://www.unit-conversion.info/texttools/convert-text-to-binary/

https://imagej.nih.gov/ij/plugins/fraclac/FLHelp/Multifractals.htm

https://bioconductor.org/packages/release/bioc/vignettes/SingleCellExperiment/inst/doc/intro.html (to create a SingleCellExperiment)

https://www.uniprot.org/uniprot/Q6DD88

Single-cell Datasets:

https://singlecell.broadinstitute.org/single_cell/study/SCP393/single-cell-rna-seq-of-adult-and-pediatric-glioblastoma#study-summary

https://singlecell.broadinstitute.org/single_cell/study/SCP503/gradient-of-developmental-and-injury-reponse-transcriptional-states-define-functional-vulnerabilities-underpinning-glioblastoma-heterogeneity#study-download

AUTHOR CONTRIBUTIONS: AU performed the algorithms, wrote, and edited the manuscript. MC supervised, wrote, and edited the manuscript.

FUNDING: MC was funded by Natural Science and Engineering Research Council of Canada Discovery Grant RGPIN-2018-04546.

REFERENCES

1. Alifieris, C., & Trafalis, D. T. (2015). Glioblastoma multiforme: Pathogenesis and treatment. *Pharmacology & therapeutics*, *152*, 63–82. https://doi.org/10.1016/j.pharmthera.2015.05.005

2. Jung, E., Alfonso, J., Osswald, M., Monyer, H., Wick, W., and Winkler, F. (2019). Emerging intersections between neuroscience and glioma biology. Nat. Neurosci. 22, 1951–1960, https://doi.org/10.1038/ s41593-019-0540-y.

3. Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., Shalek, A. K., … Bernstein, B. E. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell*, *157*(3), 580–594. https://doi.org/10.1016/j.cell.2014.02.030

4. Jia, D., Jolly, M. K., Kulkarni, P., & Levine, H. (2017). Phenotypic Plasticity and Cell Fate Decisions in Cancer: Insights from Dynamical Systems Theory. *Cancers*, *9*(7), 70. https://doi.org/10.3390/cancers9070070

5. Jin, S., MacLean, A.L., Peng, T., Nie, Q. (2018). scEpath: Energy Landscape-based Inference of Transition Probabilities and Cellular Trajectories from Single-cell Transcriptomic Data. Bioinformatics 34(12):2077-2086. 10.1093/bioinformatics/bty058.

6. Iacono, G., Massoni-Badosa, R. & Heyn, H. (2019). Single-cell transcriptomics unveils gene regulatory network plasticity. Genome Biol 20**:** 110. 10.1186/s13059-019-1713-4.

7. Paugh, B. S., Qu, C., Jones, C., Liu, Z., Adamowicz-Brice, M., Zhang, J., et al. Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. J Clin Oncol. *28*(18), 3061–3068 (2010)

8. Jones, C., Karajannis, M. A., Jones, D., Kieran, M. W., Monje, M., Baker, S. J. et al. Pediatric high-grade glioma: biologically and clinically in need of new thinking. *Neuro-oncology*, *19*(2), 153–161 (2017)

9. Lulla RR, Saratsis AM, Hashizume R. Mutations in chromatin machinery and pediatric high-grade glioma. Sci Adv. 2016 Mar 18;2(3): e1501354.

10. Xiong, S., Feng, Y., and Cheng, L. (2019). Cellular reprogramming as a therapeutic target in cancer. Trends Cell Biol. 29 (8): 623–34. 10.1016/j.tcb.2019.05.001.

11. Wolfram, S. (1988). "Complex Systems Theory," in Emerging Syntheses in Science: Proceedings of the Founding Workshops of the Santa Fe Institute, Santa Fe, New Mexico (D. Pines, ed.), Redwood City, CA: Addison-Wesley, pp. 183–189.

12. Shalizi, C.R. (2006). Methods and Techniques of Complex Systems Science: An Overview, Complex Systems Science in Biomedicine (T. S. Deisboeck and J. Y. Kresh, eds.), New York: Springer pp. 33–114.

13. Thurner, S., Klimek, P., and Hanel, R. (2018). Introduction to the Theory of Complex Systems (USA: Oxford University Press).

14. Huang, S., Ernberg, I., & Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in cell & developmental biology*, *20*(7), 869–876. https://doi.org/10.1016/j.semcdb.2009.07.003

15. Barabási, A.-L., and Oltvai, Z.N. (2014). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113, https://doi.org/10.1038/nrg1272.

16. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M. et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell 178(4). pp. 835–849.e21. 10.1016/j.cell.2019.06.024.

17. Richards, L.M., Whitley, O.K.N., MacLeod, G., Cavalli, F.M.G., Coutinho, F.J., Jaramillo, J.E., Svergun, N., Riverin, M., Croucher, D.C., Kushida, M. et al. (2021). Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. Nat Cancer. 2:157-173. 10.1038/s43018-020-00154-9.

18. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell, 177(7): 1888–1902.e21. 10.1016/j.cell.2019.05.031

19. Waddington, C.H. The strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology (Allen and Unwin, London. 1957)

20. Mandelbrot, B. (1982). The Fractal Geometry of Nature (W.H. Freeman and Company).

21. Coffey, S.D. (1998). Self organization, complexity, and chaos: the new biology for medicine. Nat. Med 4, 882–885, https://doi.org/10.3109/07357909709047603.

22. Baish, J.W., and Jain, R.K. (1998). Cancer, angiogenesis and fractals. Nat. Med. 4, 984, https://doi.org/10.1038/1952.

23. Strogatz, S.H. (2015). Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering (Boulder, CO: Westview Press).

24. Chan, T. E et al. "Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures." *Cell systems* vol. 5,3 (2017): 251-267.e3. doi:10.1016/j.cels.2017.08.014

25. Zenil, H., Kiani, N.A., and Tegnér, J. Methods of Information Theory and Algorithmic Complexity for Network Biology, Seminars in Cell & Developmental Biology, 51: 32–43 [2016]

26. Zenil, H., Kiani, N. A., Marabita, F., Deng, Y., Elias, S., Schmidt, A., Ball, G., & Tegnér, J. (2019). An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems. iScience, 19: 1160–1172. 10.1016/j.isci.2019.07.043

27. Miao, L., Jiang, Z., Wang, J., Yang, N., Qi, Q., Zhou, W., Feng, Z., Li, W., Zhang, Q., Huang, B., Chen, A., Zhang, D., Zhao, P., & Li, X. (2019). Epithelial membrane protein 1 promotes glioblastoma progression through the PI3K/AKT/mTOR signaling pathway. *Oncology reports*, *42*(2), 605–614. https://doi.org/10.3892/or.2019.7204

28. Luxen, D., Gielen, G. H., Waha, A., Isselstein, L., Müller, T., Koch, P., Hammes, J., Becker, A., Simon, M., Wurst, P., Endl, E., Pietsch, T., Gessi, M., & Waha, A. (2017). MTSS1 is epigenetically regulated in glioma cells and inhibits glioma cell motility. *Translational oncology*, *10*(1), 70–79. https://doi.org/10.1016/j.tranon.2016.11.006

29. Yu J, Shen W, Gao B, Xu J, Gong B. Metastasis suppressor 1 acts as a tumor suppressor by inhibiting epithelial-to-mesenchymal transition in triple-negative breast cancer. Int J Biol Markers. 2020 Mar;35(1):74-81.

30. Ehmsen, J. T., Ma, T. M., Sason, H., Rosenberg, D., Ogo, T., Furuya, S., Snyder, S. H., & Wolosker, H. (2013). D-serine in glia and neurons derives from 3-phosphoglycerate dehydrogenase. *The*

Journal of neuroscience : the official journal of the Society for Neuroscience, *33*(30), 12464–12469. https://doi.org/10.1523/JNEUROSCI.4914-12.2013

31. Harroch, S., Furtado, G., Brueck, W. *et al.* A critical role for the protein tyrosine phosphatase receptor type Z in functional recovery from demyelinating lesions. *Nat Genet* **32,** 411–414 (2002). https://doi.org/10.1038/ng1004

32. Shi, Y., Ping, Y. F., Zhou, W., He, Z. C., Chen, C., Bian, B. S., Zhang, L., Chen, L., Lan, X., Zhang, X. C., Zhou, K., Liu, Q., Long, H., Fu, T. W., Zhang, X. N., Cao, M. F., Huang, Z., Fang, X., Wang, X., Feng, H., … Bian, X. W. (2017). Tumour-associated macrophages secrete pleiotrophin to promote PTPRZ1 signalling in glioblastoma stem cells for tumour growth. *Nature communications*, *8*, 15080. https://doi.org/10.1038/ncomms15080

33. Bhaduri, A., Di Lullo, E., Jung, D., Müller, S., Crouch, E. E., Espinosa, C. S., Ozawa, T., Alvarado, B., Spatazza, J., Cadwell, C. R., Wilkins, G., Velmeshev, D., Liu, S. J., Malatesta, M., Andrews, M. G., Mostajo-Radji, M. A., Huang, E. J., Nowakowski, T. J., Lim, D. A., Diaz, A., … Kriegstein, A. R. (2020). Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell stem cell*, *26*(1), 48–63.e6. https://doi.org/10.1016/j.stem.2019.11.015

34. Brozzi, F., Arcuri, C., Giambanco, I., & Donato, R. (2009). S100B Protein Regulates Astrocyte Shape and Migration via Interaction with Src Kinase: IMPLICATIONS FOR ASTROCYTE DEVELOPMENT, ACTIVATION, AND TUMOR GROWTH. *The Journal of biological chemistry*, *284*(13), 8797–8811. https://doi.org/10.1074/jbc.M805897200

35. Wang, H., Zhang, L., Zhang, I. Y., Chen, X., Da Fonseca, A., Wu, S., Ren, H., Badie, S., Sadeghi, S., Ouyang, M., Warden, C. D., & Badie, B. (2013). S100B promotes glioma growth through chemoattraction of myeloid-derived macrophages. *Clinical cancer research*, *19*(14), 3764–3775.

36. Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 17(1):98-110 (2010)

37. Barabási, A-L. and Oltvai, Z.N. (2014). Network Biology: Understanding the Cell's Functional Organization, Nature Review Genetics 5(2): 101-113. 10.1038/nrg1272.

38. Barabási, A.L., and Posfai, M. Network Science (Cambridge University Press, UK. 2016)

39. Gao, M., Fu, Y., Zhou, W., Gui, G., Lal, B., Li, Y., Xia, S., Ji, H., Eberhart, C. G., Laterra, J., & Ying, M. (2021). EGFR Activates a TAZ-driven Oncogenic Program in Glioblastoma. *Cancer research*, canres.2773.2020. Advance online publication. https://doi.org/10.1158/0008-5472.CAN-20-2773

40. Bhat, K. P., Salazar, K. L., Balasubramaniyan, V., Wani, K., Heathcock, L., Hollingsworth, F., James, J. D., Gumin, J., Diefes, K. L., Kim, S. H., Turski, A., Azodi, Y., Yang, Y., Doucette, T., Colman, H., Sulman, E. P., Lang, F. F., Rao, G., Copray, S., Vaillant, B. D., … Aldape, K. D. (2011). The transcriptional coactivator TAZ regulates mesenchymal differentiation in malignant glioma. *Genes & development*, *25*(24), 2594–2609.

41. Mani, SA et al. The epithelial-mesenchymal transition generates cells with properties of stem-cells. Cell, 2008, 133:704-715

42. Polisetty, R. V., Gautam, P., Sharma, R., Harsha, H. C., Nair, S. C., Gupta, M. K., Uppin, M. S., Challa, S., Puligopu, A. K., Ankathi, P., Purohit, A. K., Chandak, G. R., Pandey, A., & Sirdeshmukh, R. (2012). LC-MS/MS analysis of differentially expressed glioblastoma membrane proteome reveals altered calcium signaling and other protein groups of regulatory functions. *Molecular & cellular proteomics : MCP*, *11*(6), M111.013565. https://doi.org/10.1074/mcp.M111.013565

43. Rodrigues, N. P., Tipping, A. J., Wang, Z., & Enver, T. (2012). GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *The international journal of biochemistry & cell biology*, *44*(3), 457–460. https://doi.org/10.1016/j.biocel.2011.12.004

44. Menendez-Gonzalez, J. B., Vukovic, M., Abdelfattah, A., Saleh, L., Almotiri, A., Thomas, L. A., Agirre-Lizaso, A., Azevedo, A., Menezes, A. C., Tornillo, G., Edkins, S., Kong, K., Giles, P., Anjos-Afonso, F., Tonks, A., Boyd, A. S., Kranc, K. R., & Rodrigues, N. P. (2019). Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. *Stem cell reports*, *13*(2), 291–306. https://doi.org/10.1016/j.stemcr.2019.07.005

45. Hou, A., Zhao, L., Zhao, F., Wang, W., Niu, J., Li, B., Zhou, Z., & Zhu, D. (2016). Expression of MECOM is associated with unfavorable prognosis in glioblastoma multiforme. *OncoTargets and therapy*, *9*, 315–320. https://doi.org/10.2147/OTT.S95831

46. Bulstrode, H., Johnstone, E., Marques-Torrejon, M. A., Ferguson, K. M., Bressan, R. B., Blin, C., Grant, V., Gogolok, S., Gangoso, E., Gagrica, S., Ender, C., Fotaki, V., Sproul, D., Bertone, P., & Pollard, S. M. (2017). Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes & development*, *31*(8), 757–773.

47. Wang, Z., Yuan, H., Sun, C., Xu, L., Chen, Y., Zhu, Q., Zhao, H., Huang, Q., Dong, J., & Lan, Q. (2015). GATA2 promotes glioma progression through EGFR/ERK/Elk-1 pathway. *Medical oncology (Northwood, London, England)*, *32*(4), 87. https://doi.org/10.1007/s12032-015-0522-1

48. Azim, E., Jabaudon, D., Fame, R. M., & Macklis, J. D. (2009). SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nature neuroscience*, *12*(10), 1238–1247. https://doi.org/10.1038/nn.2387

49. Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., & Corydon, T. J. (2020). The role of SOX family members in solid tumours and metastasis. *Seminars in cancer biology*, *67*(Pt 1), 122–153. https://doi.org/10.1016/j.semcancer.2019.03.004

50. Gu, X., Yao, L., Ma, G., Cui, L., Li, Y., Liang, W., Zhao, B., & Li, K. (2014). TCTP promotes glioma cell proliferation in vitro and in vivo via enhanced β-catenin/TCF-4 transcription. *Neuro-oncology*, *16*(2), 217–227. https://doi.org/10.1093/neuonc/not194

51. Saito T, Ikeda T, Nakamura K, Chung UI, Kawaguchi H. S100A1 and S100B, transcriptional targets of SOX trio, inhibit terminal differentiation of chondrocytes. EMBO Rep. 2007 May;8(5):504-9. doi: 10.1038/sj.embor.7400934 (2007)

52. Hagiwara N. (2011). Sox6, jack of all trades: a versatile regulatory protein in vertebrate development. Developmental dynamics: an official publication of the American Association of Anatomists, 240(6), 1311–1321. https://doi.org/10.1002/dvdy.22639

53. de la Rocha, A. M., Sampron, N., Alonso, M. M., & Matheu, A. (2014). Role of SOX family of transcription factors in central nervous system tumors. *American journal of cancer research*, *4*(4), 312–324.

54. Koziol, M. J., Garrett, N., & Gurdon, J. B. (2007). Tpt1 activates transcription of oct4 and nanog in transplanted somatic nuclei. *Current biology: CB*, *17*(9), 801–807.

55. Amson, R., Pece, S., Marine, J. C., Di Fiore, P. P., & Telerman, A. (2013). TPT1/ TCTP-regulated pathways in phenotypic reprogramming. *Trends in cell biology*, *23*(1), 37–46.

56. Gyorgy, A. B., Szemes, M., de Juan Romero, C., Tarabykin, V., & Agoston, D. V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. *The European journal of neuroscience*, *27*(4), 865–873. https://doi.org/10.1111/j.1460-9568.2008.06061.x

57. Tao, W., Zhang, A., Zhai, K., Huang, Z., Huang, H., Zhou, W., Huang, Q., Fang, X., Prager, B. C., Wang, X., Wu, Q., Sloan, A. E., Ahluwalia, M. S., Lathia, J. D., Yu, J. S., Rich, J. N., & Bao, S. (2020). SATB2 drives glioblastoma growth by recruiting CBP to promote FOXM1 expression in glioma stem cells. *EMBO molecular medicine*, *12*(12), e12291. https://doi.org/10.15252/emmm.202012291

58. Roy, S. K., Shrivastava, A., Srivastav, S., Shankar, S., & Srivastava, R. K. (2020). SATB2 is a novel biomarker and therapeutic target for cancer. *Journal of cellular and molecular medicine*, *24*(19), 11064–11069. https://doi.org/10.1111/jcmm.15755

59. Baritaki, S., Chatzinikola, A. M., Vakis, A. F., Soulitzis, N., Karabetsos, D. A., Neonakis, I., Bonavida, B., & Spandidos, D. A. (2009). YY1 Over-expression in human brain gliomas and meningiomas correlates with TGF-beta1, IGF-1 and FGF-2 mRNA levels. *Cancer investigation*, *27*(2), 184–192.

60. Sarvagalla, S., Kolapalli, S. P., & Vallabhapurapu, S. (2019). The Two Sides of YY1 in Cancer: A Friend and a Foe. *Frontiers in oncology*, *9*, 1230. https://doi.org/10.3389/fonc.2019.01230

61. Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan - Golan Y, Kohn A, Rappaport N, Safran M, and Lancet D. *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis* , Current Protocols in Bioinformatics(2016), 54:1.30.1 - 1.30.33.doi: 10.1002 / cpbi.5.

62. GeneCards: The Human Gene Database. https://www.genecards.org/ (GCID: GC12P013196; GC08M124550) (Weizmann Institute of Science; accessed on July 3, 2021)

63. Latora, V. et al., Centrality Measures. Complex Networks: Principles, Methods and Applications (Cambridge University Press, Cambridge, UK. 2017**)**

64. Rodrigues, F.A. Network centrality: an introduction. arXiv:1901.07901 [physics.soc-ph] (2019)

65. Engel, A.L., Lorenz, N.I., Klann, K. *et al.* Serine-dependent redox homeostasis regulates glioblastoma cell survival. *Br J Cancer* **122,** 1391–1398 (2020). https://doi.org/10.1038/s41416-020-0794-x

66. Yamasaki, M., Yamada, K., Furuya, S., Mitoma, J., Hirabayashi, Y., & Watanabe, M. (2001). *3-Phosphoglycerate Dehydrogenase, a Key Enzyme forl-Serine Biosynthesis, Is Preferentially Expressed in the Radial Glia/Astrocyte Lineage and Olfactory Ensheathing Glia in the Mouse Brain. The Journal of Neuroscience, 21(19), 7691–7704.* doi:10.1523/jneurosci.21-19-07691.2001

67. Liu, J., Guo, S., Li, Q., Yang, L., Xia, Z., Zhang, L., Huang, Z., & Zhang, N. (2013). Phosphoglycerate dehydrogenase induces glioma cells proliferation and invasion by stabilizing forkhead box M1. *Journal of neuro-oncology*, *111*(3), 245–255. https://doi.org/10.1007/s11060-012-1018-x

68. Schemionek, M., Kharabi Masouleh, B., Klaile, Y., Krug, U., Hebestreit, K., Schubert, C., … Koschmieder, S. (2015). *Identification of the Adapter Molecule MTSS1 as a Potential Oncogene-Specific Tumor Suppressor in Acute Myeloid Leukemia. PLOS ONE, 10(5), e0125783.*

69. Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D. A., Jones, D. T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., Kool, M., Zapatka, M., Becker, N., Zucknick, M., Hielscher, T., Liu,

X. Y., Fontebasso, A. M., Ryzhova, M., Albrecht, S., Jacob, K., … Pfister, S. M. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer cell*, *22*(4), 425–437. https://doi.org/10.1016/j.ccr.2012.08.024