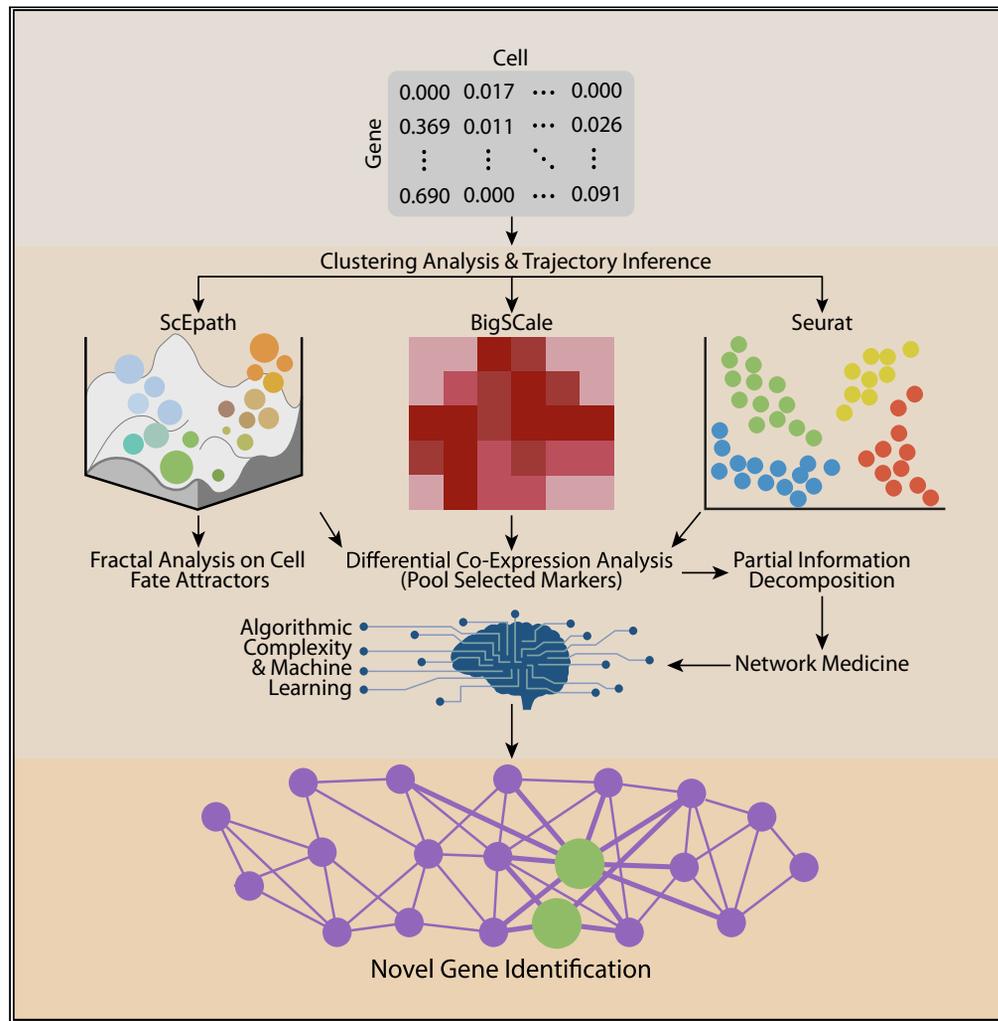


Article

Algorithmic reconstruction of glioblastoma network complexity



Abicumarán
Uthamacumarán,
Morgan Craig

morgan.craig@umontreal.ca

Highlights

Complex cell fate attractors capture glioblastoma differentiation dynamics

Graph theoretic approaches decode master regulators of GBM glioblastoma cell fate decisions

Network dynamics of pediatric glioblastoma resemble adult GSCs

Transcriptional networks may help reprogram glioblastoma behavioral patterns

Uthamacumarán & Craig,
iScience 25, 104179
May 20, 2022 © 2022 The Authors.
<https://doi.org/10.1016/j.isci.2022.104179>



Article

Algorithmic reconstruction of glioblastoma network complexity

Abicumaran Uthamacumaran¹ and Morgan Craig^{2,3,4,*}

SUMMARY

Glioblastoma is a complex disease that is difficult to treat. Network and data science offer alternative approaches to classical bioinformatics pipelines to study gene expression patterns from single-cell datasets, helping to distinguish genes associated with the control of differentiation and aggression. To identify the key molecular regulators of the networks driving glioblastoma/GSC and predict their cell fate dynamics, we applied a host of data theoretic techniques to gene expression patterns from pediatric and adult glioblastoma, and adult glioma-derived stem cells (GSCs). We identified eight transcription factors (OLIG1/2, TAZ, GATA2, FOXG1, SOX6, SATB2, and YY1) and four signaling genes (ATL3, MTSS1, EMP1, and TPT1) as coordinators of cell state transitions and, thus, clinically targetable putative factors differentiating pediatric and adult glioblastomas from adult GSCs. Our study provides strong evidence of complex systems approaches for inferring complex dynamics from reverse-engineering gene networks, bolstering the search for new clinically relevant targets in glioblastoma.

INTRODUCTION

Glioblastoma is the most lethal pediatric and adult brain tumor. Despite advances in treatment, recurrence will occur in all glioblastoma patients, and mean survival in adults is only 15 months (Alifieris and Trafalis, 2015). Glioblastoma 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 glioblastoma 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 glioblastoma 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 glioblastoma exhibits molecular patterns and collective behaviors which are fundamentally different from those of adult glioblastoma (Paugh et al., 2010; Jones et al., 2017; Schwartzentruber et al., 2012; Sturm et al., 2012). There is a greater epigenetic burden in pediatric glioblastoma marked by specific histone H3.3 modifications and aberrant DNA methylation profiles (Schwartzentruber et al., 2012; Sturm et al., 2012; Lulla et al., 2016; Harutyunyan et al., 2019). However, the complex signaling dynamics distinguishing pediatric and adult glioblastoma 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 glioblastoma groups steers their cell fate dynamics would allow for the prediction of causal patterns in disease progression and therapeutic responses.

Glioma-derived stem cells (GSCs) are believed to be a small subset of glioblastoma cancer cells that largely contribute to emergent glioblastoma 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 glioblastoma 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

¹Department of Physics, Concordia University, Montréal, Québec H3G 1M8, Canada

²Sainte-Justine University Hospital Research Centre, Montreal, Québec H3T 1C5, Canada

³Département de mathématiques et de statistique, Université de Montréal, Montréal, Québec H3C 3J7, Canada

⁴Lead contact

*Correspondence: morgan.craig@umontreal.ca
<https://doi.org/10.1016/j.isci.2022.104179>



expression patterns) underlying GSC cell fate choices and transitions to their mature phenotypes hinders successful clinical interventions in the treatment of glioblastoma (Jung et al., 2019; Xiong et al., 2019; Yabo et al., 2021).

Statistical approaches are traditionally 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. Furthermore, fundamental limitations associated with the raw counts of the scRNA-Seq complicate the inference of networks in complex diseases like glioblastoma. 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 glioblastoma networks while retaining the essential information representative of their complex dynamics requires tools from the interdisciplinary paradigm of complex systems theory. Thereby, the main objective of our study comprises complex systems reconstruction of cell fate decisions and behavioral patterns in glioblastoma dynamics. Specifically, we employed complex systems tools such as network medicine, attractor reconstruction, and statistical measures of complexity to decode cellular dynamics in glioblastoma ecosystems and gain quantitative insights into the molecular drivers of glioblastoma differentiation dynamics and their adaptive behaviors such as glioblastoma plasticity and multiscale heterogeneity.

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). 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 temporal 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 (Thurner et al., 2018; Huang et al., 2009; Barabási and Oltvai, 2004).

The molecular networks coordinating the emergence of GSC and glioblastoma phenotypes are such complex networks. To reveal the mechanisms underlying GSC cell fate decisions and transitions to their mature glioblastoma phenotypes, we deployed several approaches from complex systems theory on data from single-cell RNA Sequencing (scRNA-Seq) count matrices. We compared pediatric glioblastoma to adult glioblastoma to identify the signaling network patterns distinguishing pediatric and adult glioblastoma from GSCs. For this, we relied on 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 glioblastoma/GSC cell fate control and regulation. We found that distinct gene expression signatures regulate the cell fate decisions in the glioblastoma 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/glioblastoma 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 glioblastoma, identify potential drug targets for precision oncology, and provide a roadmap for data theoretic approaches to other such complex systems.

RESULTS

Key driver genes mediating the cell fate transition dynamics in glioblastoma/GSC epigenetic landscapes are identified using the scEpath algorithm

We applied our complex theory-based analysis to pediatric and adult IDH-wt glioblastoma single-cell RNA-Seq (scRNA-Seq) datasets (Table 1). Waddington landscape reconstruction identified causal patterns (attractors) to which the distinct transcriptional states within each patient group cluster (Figures 1A–1C). Distinct patient group clusters were determined by the scEpath algorithm (colored by similarity in gene expression (i.e., phenotypes) in Figure 1). Three and four meta-clusters were identified in the pediatric

Table 1. Summary of single-cell datasets

Patient Group	Single-Cell Dataset	# Patient Samples (n) and Single-Cells (N) for Seurat/BigScale	# Patient Samples (n) and single-cells (N) for scEpath Analysis	# of Cell Fate Trajectories in scEpath Waddington Landscape
Pediatric glioblastoma	Neftel et al. (18)	n = 7 N = 1850	n = 7 N = 1850	2
Adult glioblastoma	Neftel et al. (18)	n = 18 N ~21,500	n = 7 N = 2221	4
Adult GSC	Richards et al. (19)	n = 28 N~69,000	n = 13 N = 1504	2

The total number of patient samples (n) and number of single-cells within each patient group (N) used for each step of the clustering and single-cell trajectory inference process are shown.

glioblastoma (Figure 1A) and adult glioblastoma (Figure 1B), respectively, whereas sub-populations are observed within each meta-cluster indicating the presence of phenotypic heterogeneity and epigenetic plasticity. 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 (Figures 1D–1F) and found similarities in the gene expression signatures and similar oscillatory patterns in EMP1, MTSS1, PHGDH and OLIG1/2 (Figure 2). These markers were selected in the clustering and trajectory inference process as explained above. Their similarity was assessed by their expression variation along the cell fate trajectories in pseudotime (Figure 2). We also identified OLIG1/2 as critical transcription factors in the adult GSC phenotypic transitions (Figure 1F). Other signatures such as EGFR and PDGFRA were observed in the top 2 PCA loadings of glioblastoma samples in both Seurat and BigScale clustering. However, they were not expressed as highly in all clusters and hence, only a few genes were found to be relevant during filtering when imposing the condition that the gene marker must be expressed in all patient groups and all cell clusters in TSNE/UMAP pattern space (Figure S1).

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 amongst the identified critical gene markers. Using clustering algorithms (see STAR Methods), we found that PTPRZ1 and S100B showed nearly identical expression dynamics in pediatric glioblastoma 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 glioblastoma and adult GSC (Figures 2A and 2C). The expression metric used to compare the dynamics of the different pseudotime-dependent genes correspond to the cubic spline smoothed average normalized expression along the pseudotime interval of [0,1].

In adult glioblastoma, NACA and PABPC1, and TPT1 and PSAP had similar expression patterns across all four differentiation paths (Figure 2B). S100B, OLIG1, and PHGDH all had a broad expression profile in path 4 (Figure 2B). Furthermore, the presence of four cell clusters in adult glioblastoma landscape (Figure 1B) is in good agreement with previous classifications of four molecular subtypes of adult glioblastoma (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 (Figure 2C). For instance, PTPRZ1, NACA and PABPC1, were all found to have similar expression dynamics in both transition paths (Figure 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 (Figures 2A–2C).

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 glioblastoma's Waddington landscape (Figures 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

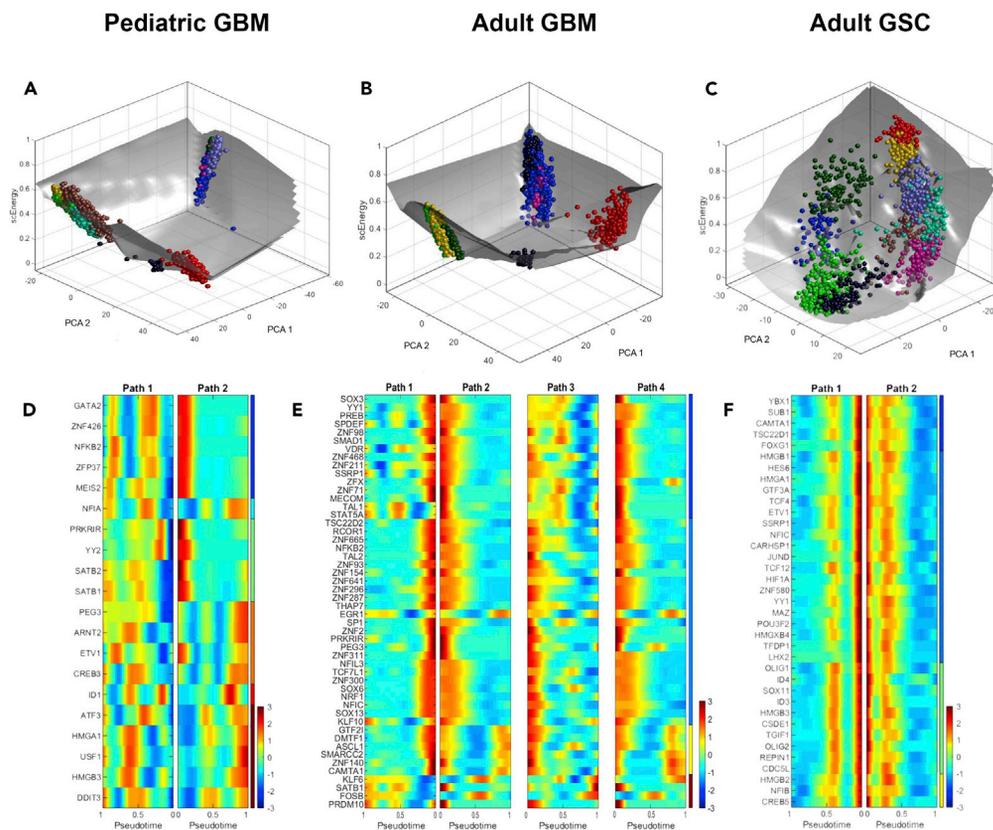


Figure 1. Reconstructing pseudotime dynamics in glioblastoma/GSC cell fate decisions of the Waddington landscape

Average normalized 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 glioblastoma, (B) adult glioblastoma, (C) adult GSC. (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 glioblastoma/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 for D) pediatric glioblastoma, E) adult glioblastoma, and F) adult GSCs. Additional results are reported in [Figure S3](#) in the [Supplementary Information](#).

(see [Figure 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 glioblastoma/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 glioblastoma patient's tumor.

PIDC network inference algorithm reconstructs the regulatory network configurations driving glioblastoma/GSC cell fate transitions

We next reverse engineered the signaling networks coordinating the information flow in glioblastoma and GSC using Partial Information Decomposition and Context (PIDC). In graph theory, each link (edge) of a network has a weight capacity, a threshold, above or below which a state-transition can occur in the

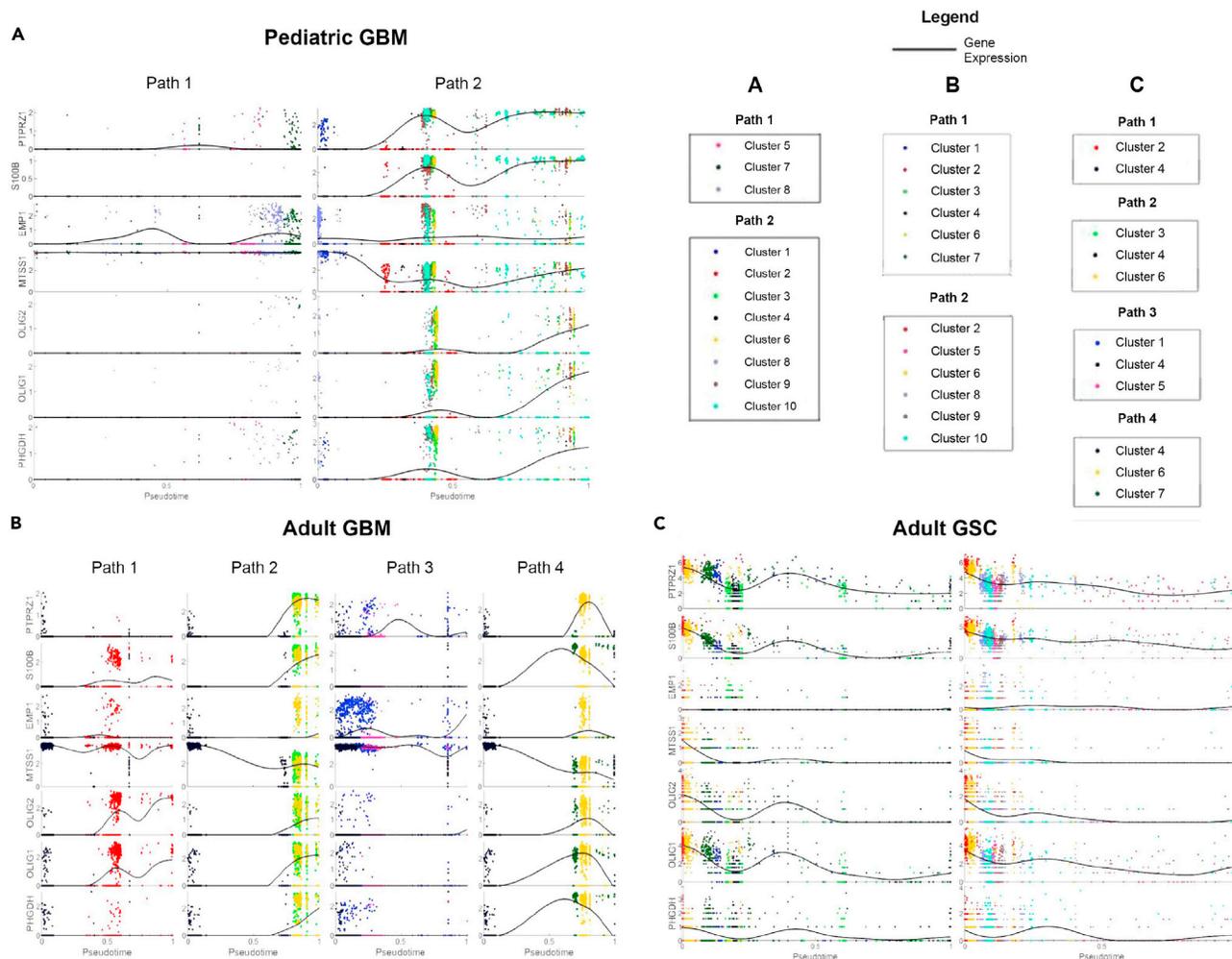


Figure 2. Reconstructing pseudotime dynamics in glioblastoma/GSC cell fate decisions of the Waddington landscape

Average normalized 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 glioblastoma, (B) adult glioblastoma, (C) adult GSCs.

network's topology resulting in changes in its behaviors (dynamics) (Latora et al., 2017; Rodrigues, 2019). The connectivity/topology of the graph and the transitions/updating of the PID scores (weights) creates a flow pattern characterizing the network dynamics. The transfer of information across the network is defined as information flow (Latora et al., 2017; Rodrigues, 2019). The information flow can cause topological transitions (rearrangements) of the network configuration, and hence, result in adaptive cell fate behaviors or correspond to the emergence of distinct cancer phenotypes.

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 glioblastoma, suggesting a strong relationship between these two genes in the complex network steering their cell fate decisions (Figure 3A). We found that S100A10 and EMP1 have the highest interaction in adult glioblastoma with a PID score of 1.9517 (Figure 3B), whereas NACA and TPT1 had the highest interaction in adult GSC with a PID score of 1.9628 (Figure 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 (Figures 3E–3G) had

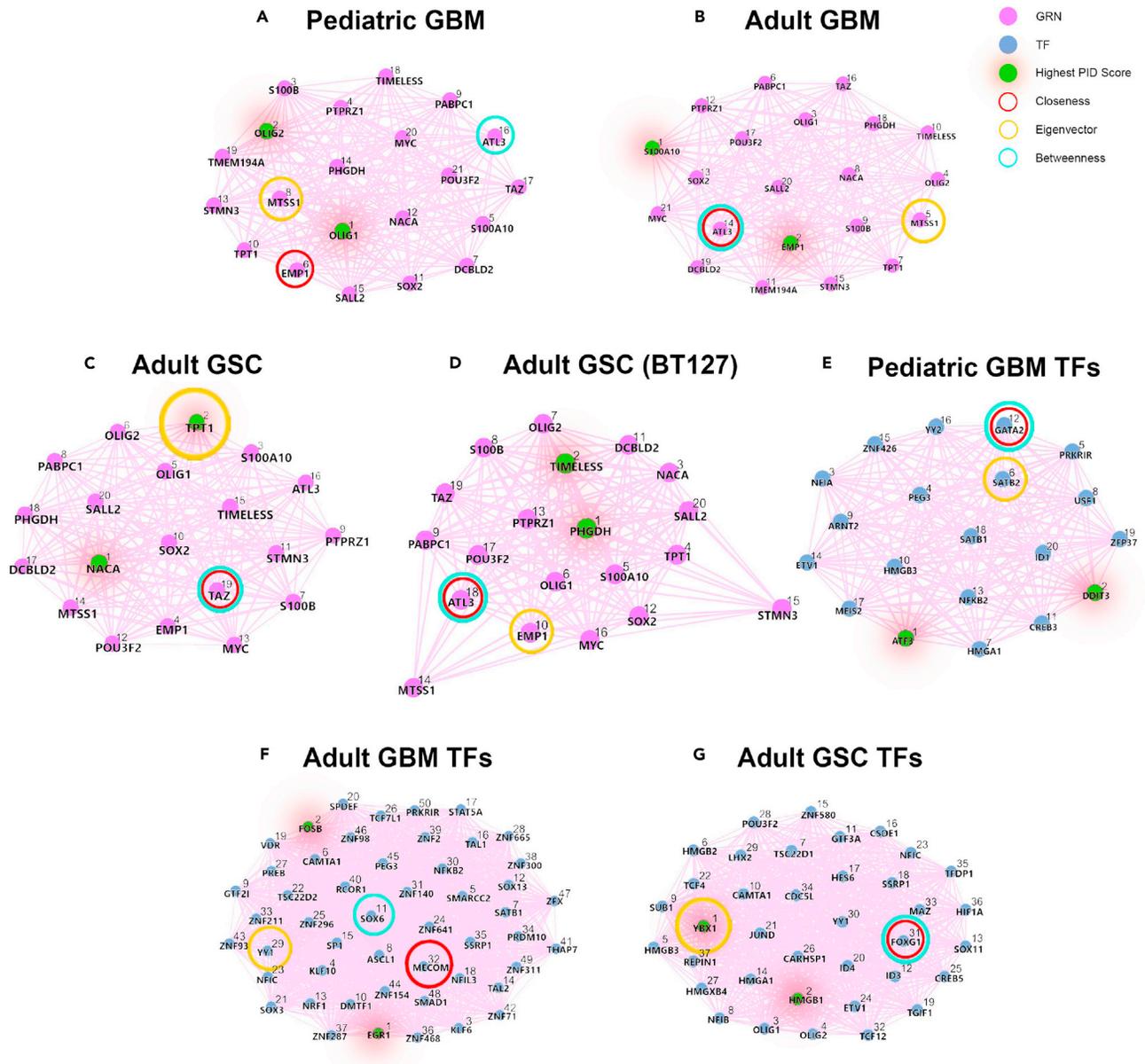


Figure 3. Mathematical modelling identifies key regulatory genes driving glioblastoma networks

Gene regulatory networks of (A) pediatric glioblastoma, (B) adult glioblastoma, (C) adult GSC, (D) adult GSC sample BT127, (E) pediatric glioblastoma transcription factors, (F) adult glioblastoma transcription factors, and (G) 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 (A–D) whereas the scEpath TF networks are colored in teal (E–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. In addition, three different colored rings are used to identify the nodes of the networks with the highest network centrality measures as identified in Figure 4. See Figure S5 in the Supplementary Information for additional results.

similar pseudotemporal expression dynamics (Figures S4A–S4C in the Supplementary Information). ATF3 and DDIT3 were the top interaction markers from the critical TFs identified for pediatric glioblastoma with a PID score of 1.971 (Figure 3E). EGR1 and FOSB in the adult glioblastoma group (Figure 3F), and YBX1 and HMGB1 were identified as the top interaction TF markers, with PID score of 1.992 (Figure 3G). These results suggest the reconfiguration of the nodes within the same complex signaling network may characterize GSC cells from glioblastoma cells and distinguish pediatric glioblastoma from adult glioblastoma cell fate dynamics.

Table 2. General properties of inferred complex networks

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

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.

Network centrality measures identify master regulators of information flow across the regulatory networks underlying glioblastoma/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 glioblastoma, and adult GSC networks (Table 2). 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 (italic columns, Table 2), the GSC TF network had the highest diameter whereas the glioblastoma 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 (bold columns, Table 2) 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 et al., 2019). The nodes corresponding to the highest closeness centrality for each GRN network were found to be Node 6 (EMP1) for pediatric glioblastoma, Node 14 (ATL3) for adult glioblastoma, 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 (Figure 4A). Nodes corresponding to the maximal closeness in the pediatric glioblastoma, adult glioblastoma, 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 (Figure 4B).

Betweenness centrality indicates the presence of regulatory bottlenecks (Iacono et al., 2019; Latora et al., 2017; Rodrigues, 2019). In our analyses, the highest betweenness measures for the pediatric glioblastoma, adult glioblastoma, 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 (Figure 4C). The trends in maximal betweenness values (Figure 4C) were in good agreement with the nodes contributing to the maximal closeness values discussed in Figure 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 node 12 (GATA2) for pediatric glioblastoma, node 11 (SOX6) for adult glioblastoma, and node 31 (FOXG1) for adult GSC, with values of 0.3801, 0.2279, and 0.1539, respectively (Figure 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 glioblastoma, node 5 (MTSS1) for adult glioblastoma, node 10 (EMP1) for BT127, and node 2 (TPT1) for GSC, with measures of 0.2796, 0.2827, 0.2909, and 0.2805, respectively. The eigenvector centrality, also known as the prestige score or authority score, measures the influence of a node on all other nodes in the network in directing the information dynamics (i.e., the PID scores) (Rodrigues, 2019). In our case, a high eigenvector centrality node implies the node is connected to many other nodes with high PID scores.

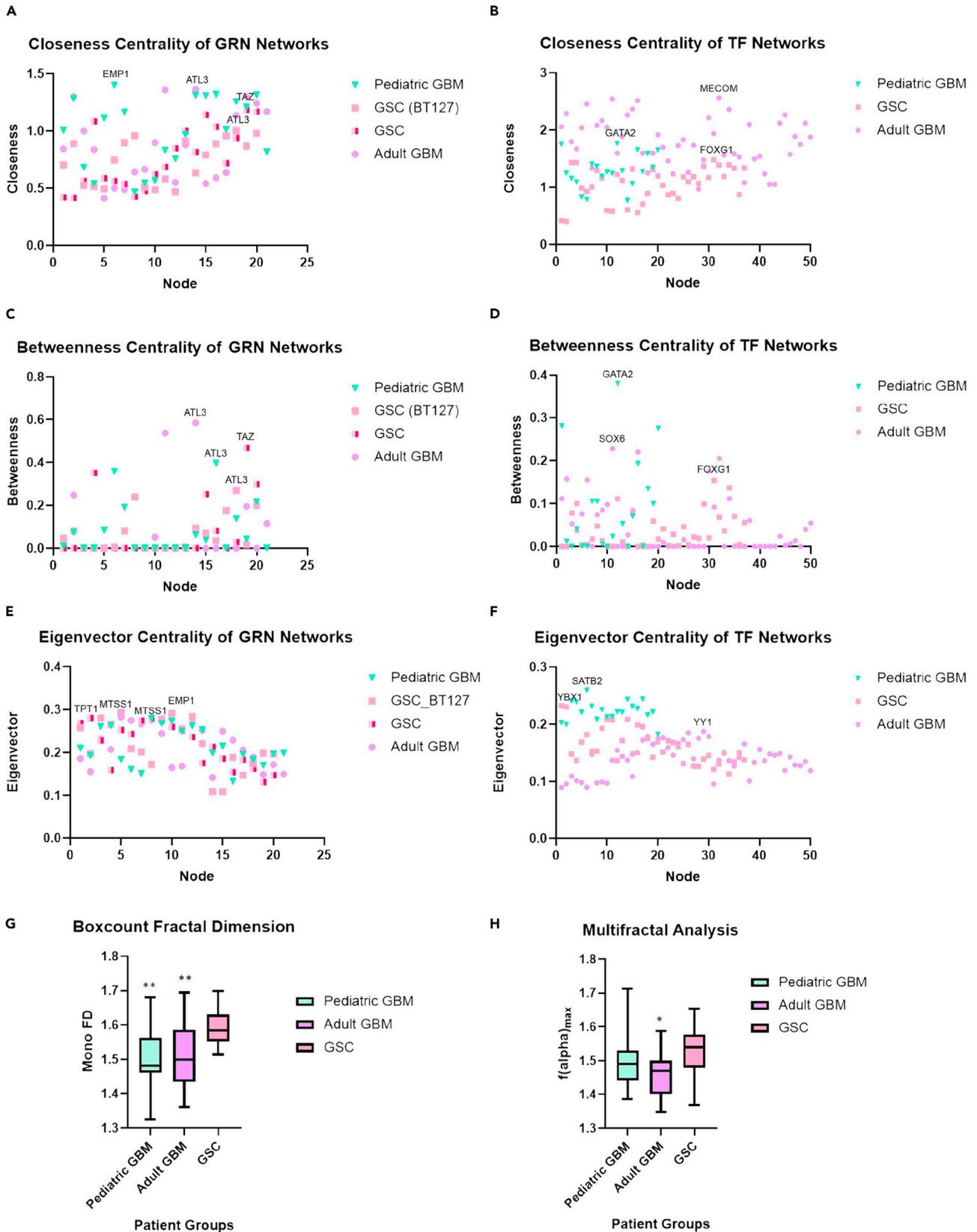


Figure 4. Centrality measures distinguish master regulators of information flow in glioblastoma networks

Three network centrality measures are assessed on the reconstructed glioblastoma/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 glioblastoma/GSC cell fate transition dynamics.

(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.

(H) Multifractal analysis of cell fate attractors on scEpath Waddington landscapes.

A p value of 0.0031 between the adult GSC and adult glioblastoma, and $p = 0.0011$ between adult GSC and pediatric glioblastoma 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.

The maximal eigenvector is a measure of the hub-score, i.e., the highest authority node of hub networks (Latora et al., 2017; Rodrigues, 2019). The maximal eigenvector centrality of the TF networks was found to be node 6 (SATB2) for pediatric glioblastoma, and node 29 (YY1) for adult glioblastoma, 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 remodeling and transcription regulation during neuronal differentiation (Gyorgy et al., 2008). Interestingly, all transition genes with high centrality measures identified in our network analyses, including EMP1, MTSS1, ATL3, and TPT1 have a TF-binding site for YY1 (Stelzer et al., 2016; GeneCards, 2021) (see Table 3).

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 glioblastoma ($n = 7$), adult glioblastoma ($n = 18$), and GSC ($n = 28$)). The mean fractal dimension scores of the pediatric glioblastoma, adult glioblastoma, 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 glioblastoma groups were nearly identical whereas 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 ($n = 54$) (Figure 4H). Only the difference between GSC versus adult glioblastoma was found to be statistically significant ($p = 0.0201$) by a Kolmogorov-Smirnov test. The pediatric glioblastoma, and adult glioblastoma 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.

DISCUSSION

Here we applied a collection of data theoretic and complexity science approaches to single cell RNA-seq data from pediatric and adult glioblastoma, and adult GSCs to distinguish genes regulating communication within these cellular populations. Our findings demonstrate the application of these tools for deciphering glioblastoma/GSC signaling networks to understand how network configuration orchestrates information flow and determines cell fate dynamics.

Multiple clustering algorithms were deployed to cross-validate their findings and ensure that the differential markers extracted for network analysis were robust, complementary, and of high importance in cell fate transition/differentiation mapping. There is a high degree of heterogeneity displayed by glioblastoma stem cells. The complementarity of our results in our independent and orthogonal approaches are outlined in Table 3 by the associations identified between the transition genes and the scEpath TFs. Our approach using distinct clustering techniques and verifying their matching or complementary results was deployed to minimize the effects of expression heterogeneity and validate our findings (Krieger et al., 2020).

Using scEpath, we identified three and four meta-clusters in the pediatric glioblastoma (Figure 1A) and adult glioblastoma (Figure 1B), respectively, while sub-clusters within each meta-cluster indicated the presence of phenotypic heterogeneity and plasticity. However, the number of meta-clusters was ambiguous in the adult GSC landscape (Figure 1C), as shown by the continuous progression from the higher energy state clusters (stem-like fates) to the lower energy states indicating the potential presence of a complex attractor. An alternative measure to assess the significance of the scEpath clustering is the transition paths (cell fate

Table 3. Interactions between transition genes and transcription factors identified in network analysis

Transition Genes	Transcription Factors
ATL3	YY1, FOSB, SOX6, GATA2, ATF3, EGR1, MYC
MTSS1	YY1, ATF3, MYC
EMP1	YY1, FOSB, GATA2, ATF3, MYC
TPT1	YY1, ATF3, FOSB, SOX6, EGR1, OLIG1/2
PTPRZ1	YY1, YY2, EGR1, NANOG, POUF51
S100B	YY1, GATA2, EGR1, SOX6, MYC

Amidst the critical transition genes listed, the first four were identified as the central regulators of information flow across the glioblastoma/GSC regulatory networks, while PTPRZ1 and S100B were other differential markers identified in our analyses. The list is not inclusive of all possible gene-TF interactions but restricted to the analysis of only the high importance (i.e., highest network centrality measures) scEpath TFs identified in our findings. The TF-gene interactions were identified using the GeneCards human gene database (GeneCards, 2021).

trajectories). We predicted that the number of clusters identified in the pediatric glioblastoma group corresponds to the neuronal, astrocytic-mesenchymal, and oligodendrocytic lineages, mirroring the healthy brain's neurodevelopmental hierarchy (Jessa et al., 2019; Couturier et al., 2020). Similarly, the four clusters identified in the adult glioblastoma group correspond to the four groups identified by Nefitel et al. (2019), namely the OPC-like (oligodendrocytic progenitor cell), NPC-like (neuronal progenitor cell-like), AC-like (astrocytic cell-like), and MES-like (mesenchymal cell) lineages. Furthermore, the infiltrated immune cells (i.e., T-cells and macrophages) grouped into the MES-like state (Nefitel et al., 2019). Pediatric glioblastoma cells showed less differentiation than the adult glioblastoma samples, as indicated by the higher energy cell-states, suggesting a closer resemblance to the GSC sample. The two cell fate trajectories observed in the adult GSC sample may correspond to the transcriptional gradient of two cellular states observed in the original study by Richards et al. (Figure 2C), which were shown to mirror normal neurodevelopment and inflammatory wound responses (Richards et al., 2021).

The cell fate trajectories along the scEpath Waddington landscape (Figures 1A–1C) were determined by the transition probabilities of the probabilistic directed graph reconstructed from the cell fate clusters, where the weighted edges of the networks correspond to the average normalized gene expression (see STAR Methods for additional details). scEpath used the minimum directed spanning tree to find the maximum probability flow and minimal number of edges along the network, since cell fates transition to lower energy states during differentiation. The resulting tree approximates the cell state transition network and infers the observed developmental trajectories/lineage structures. The weighted edges of the cell state transition network were found to be proportional to the gene expression values seen in Figure 3, where the number of developmental trajectories inferred are indicated by the path numbers in Figure 1. Thus, two cell fate trajectories were detected in the pediatric glioblastoma and adult GSC samples whereas four developmental trajectories were observed in adult glioblastoma.

In pediatric glioblastoma, the expression of transcription factors in pseudotime was shown to be highly nonlinear. Certain genes, including GATA2, were even found to be oscillatory in one trajectory while demonstrating 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 glioblastoma and adult GSC cells. Furthermore, we found that genes such as EMP1, MTSS1, PTPRZ1 and S100B exhibited 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 glioblastoma/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 glioblastoma/GSC Waddington landscape. The fractal dimension measure was used to quantify the complexity/roughness of the cell fate patterns (attractor) on the Waddington landscape reconstruction. Our findings revealed that the cell fate differentiation patterns of adult GSCs had a higher fractal index, followed by pediatric glioblastoma, whereas adult glioblastoma exhibited the lowest fractality. We predict a higher fractal index

may be a signature of phenotypic plasticity or aggressivity, and lower control predictability in cancer systems.

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 glioblastoma and GSC cells (Figures 4A–4F). The graph-theoretic measures allowed for the identification of key driver genes/markers controlling glioblastoma differentiation dynamics. Many of these signals are neurodevelopmental transcription factors involved in healthy brain development, essential for conferring and maintaining cancer stem cells (GSCs). Maximal centrality scores indicated that they are key regulators of the network information flow in both glioblastoma groups and GSCs. The functional significance of these transcription factors (see [Supplementary Information](#)) suggests their critical role in stem cell decision-making and differentiation dynamics. Our findings indicate that these genes may be strong candidates for therapeutic interventions points for the treatment of glioblastoma. Other signaling interactions such as PTPRZ1 and S100B were identified in our analyses as potent clinically druggable targets in the treatment of glioblastoma. Furthermore, we predicted that GATA2 and MTSS1 may provide a common ground for interlinking leukemogenesis, the complex signaling dynamics of leukemia/lymphoma affecting children, and pediatric glioma/glioblastoma (Menendez-Gonzalez et al., 2019; Schemionek et al., 2015).

Although the graph theoretic network centrality measures identified the critical genes governing glioblastoma differentiation dynamics, BDM was used to distinguish which of the differential network markers can accurately classify/differentiate the three patient group samples (see [STAR Methods](#) and [Supplementary Information](#)). We identified FOSB, HMGB1 and EGR1 as differential signatures which can accurately predict the patient groups in our single-cell analyses (see [Supplemental Information](#)). The algorithmic complexity measured by the BDM allowed for the identification of critical network genes differentiating glioblastoma and GSC phenotypes with the minimal information. The rationale for using gene/TF markers' BDM as a phenotypic discriminant is that the algorithmic complexity denotes the shortest algorithm or minimal set of information within the complex networks inferred required to classify the distinct patient groups. As such, the identified genes/TFs may be useful biomarkers for prognostic screening and disease phenotyping in clinical medicine.

From the transcription factor (TF) networks identified by scEpath (Table 3), we distinguished some TFs to form interactions with some of the differential gene markers, suggesting cellular reprogramming targets for controlling glioblastoma cell fate dynamics. Our study therefore quantifies how these markers' expressions vary in the cell fate transitions from stem-like to mature phenotypes. For a discussion on the biological significance of key genes and transcription factors identified in our analyses, see the [Supplementary Results](#) in the [Supplementary Information](#).

The cell fate transition markers identified in our study, including PTPRZ1, EMP1, S100B, and MTSS1, are in good agreement with the findings from the original studies (SCP393 and SCP503). Although some of the signatures we identified overlap with the differential expression patterns of the original studies, they did not compare the co-expression of these markers between GSC and glioblastoma. Markers differentiating distinct cellular states have been previously investigated (for instance, the original study by Neftel et al. identified copy number amplifications of the CDK4, EGFR, and PDGFRA loci and mutations of the NF1 locus, each favoring one of the four glioblastoma phenotypes (Neftel et al., 2019)). Our study instead analyzed the expression patterns which fluctuate or form a differentiation gradient across the distinct cell states. Furthermore, whereas previous studies have associated the differentiation markers of glioblastoma progression identified here, our study demonstrates their novel integrated application to elucidate the roles of these network biomarkers in glioblastoma cell fate decisions and differentiation dynamics. Indeed, whereas many of the identified genes or TFs have been previously studied in the context of neurodevelopmental regulation and glioma cell fate dynamics, most of those selected in our analyses are not yet documented in glioblastoma cell fate control. As such, we propose the identified interactions in Table 3 may provide clinically relevant glioblastoma-specific precision therapeutics, and that our network analyses provide a quantitative tool to characterize which of the markers were of high importance (i.e., high centrality measures) in cell fate control, plasticity regulation, and transition dynamics. Future studies should exploit tools from algorithmic complexity theory including algorithmic network perturbation analysis (i.e., quantify the BDM changes across a network by node or link deletion) to better elucidate the inferred network dynamics in cancer cell fate control and regulation.

Although previous glioblastoma gene regulatory network inference methods vary from our approaches, our findings are consistent with their results. For example, Sun et al. found 15 hub genes in glioblastoma-specific miRNA-TF networks, including PDGFRA and SOX11, and 6 hub TFs (including GATA1) as key regulators of glioblastoma dynamics (Sun et al., 2012). In our study, we also identified PDGFRA and SOX11 as hub genes of the inferred glioblastoma networks, and found that GATA2, an alternate isoform, overlapped with these findings. However, Sun et al. (2012) did not compare glioblastoma of different age groups nor consider glioblastoma-derived stem cells for reconstruct their differentiation networks. Similarly, a network inference study by Ping et al. (2015) revealed 17 hub genes in glioblastoma networks, including EGFR and PDGFRA, as gene signatures of the proneural glioblastoma subtype, both of which were identified in our analyses. In another study, GSEA and IPA-based gene enrichment pathway analysis discovered TAZ as a key regulator of glioblastoma networks (Bozdag et al., 2014), which was also identified as a master regulator of glioblastoma differentiation dynamics in our analyses.

Using multi-omic analyses, Suva et al. distinguished OLIG2, POU3F2 SALL2, and SOX2 as hub genes of glioblastoma stemness networks critical for their tumor-propagation potential (Suvà et al., 2014). Our findings identified OLIG2 as a master control gene of glioblastoma differentiation dynamics and established a connection between SOX2 expression and the critical hub gene FOXG1. Furthermore, some epigenetic profiling studies have shown that aberrant histone modifications and methylation profiles are molecular signatures driving pediatric glioblastoma and distinguishing them from their adult counterparts (Jones et al., 2017; Lulla et al., 2016; Sturm et al., 2012). Sturm et al. (2012) revealed that the TFs OLIG1, OLIG2 and FOXG1 are the master regulators of the hub gene networks driving these oncohistone pediatric glioblastoma variants (i.e., K27M and G34 V/R). Similar findings were recently reported by Wang et al. (2021), who identified the same set of TFs as critical drivers of pediatric high-grade gliomas' epigenetic landscapes. We identified all three TFs reported by Sturm et al. and Wang et al. in our network approaches as critical regulators of glioblastoma cell fate dynamics and behavioral patterns. Thus, our findings recapitulate the complex network dynamics driving the oncohistone variants of pediatric glioblastoma and validate and extend previous findings.

It should be noted that there is a good deal of heterogeneity within the single-cell datasets across and with the patient groups. The original datasets contained 8 pediatric glioblastoma samples, 20 adult glioblastoma samples, and 28 adult GSC samples. For the initial clustering (i.e., differential discovery using Seurat and BigSCale), samples—two adult glioblastoma and one pediatric glioblastoma—with the highest drop-out rate (i.e., zero counts) were removed as a data filtering and quality control step before normalization. Subsequently, the number of adult glioblastoma samples in the scEpath analysis was randomly selected to closely match the cell count numbers of the adult GSC patient groups. The down-sampling of GSC samples was necessary since scEpath analysis has a computational limitation on the number of samples which can be processed (roughly 2500 cells). As noted in the Methods, selecting a different combination of GSC samples did not change the results and including the removed samples did not change the differential marker discovery or expression analyses. Indeed, the global clustering patterns remained the same although there was greater dispersion in the local sub-clusters in the Seurat and BigSCale pattern space. However, including all $n = 8$ pediatric glioblastoma patient samples generated a shorter list of transition genes with abrupt transitions between the distinct phenotypes.

This proof-of-concept study provides a comprehensive method to dissect the cybernetics of cancer cellular ecosystems and their cell fate dynamics. Current bioinformatic pipelines in cancer data science largely fail to reconcile the complex dynamics and temporal features of glioblastoma transcriptional states, as they either take a reductionist approach to inferring gene expression patterns or rely on statistical correlation methods. In contrast, our framework provides a pipeline for causal pattern discovery and thereby allows the prediction/forecasting of how the differentially expressed transition genes control and regulate cell fate decision-making. Furthermore, our approach allows for the mapping of these cancer cell fate behaviors to information flow across the inferred complex networks. Thus, these causal inference tools shed light on emergent behaviors in cell fate decisions such as transcriptional heterogeneity from a dynamical systems perspective. As such, we propose our methodological framework may provide a complementary and potentially more useful means to assess how the heterogeneous cancer phenotypes exhibit adaptive (emergent) behaviors and help forecast their dynamic response to drug/therapeutic perturbations at the level of molecular interactions.

CONCLUSION

This study demonstrates the use of complex systems approaches in deciphering the cybernetics of glioblastoma/GSC networks, and shows how signaling dynamics differ between pediatric glioblastoma, adult glioblastoma, 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 glioblastoma, suggesting both precision therapeutic targets and glioblastoma reprogramming factors.

Prospective studies should explore the use of artificial neural networks, including deep learning algorithms, for single-cell transcriptomic analyses. Further, causal inference-based network inference methods such as Bayesian networks and algorithmic information dynamics should be investigated for glioblastoma regulatory networks reconstruction. The epigenetic regulation of our identified transcriptional networks must be explored using high-throughput multi-omics datasets. Our network approaches should be extended to protein-protein interaction networks, epigenetic networks, and metabolic networks to investigate multi-omic levels of glioblastoma heterogeneity, including oncohistone variants (i.e., K27M, K36M, G34 V/R) and IDH1/2-mutants observed in pediatric gliomas (glioblastoma).

LIMITATIONS OF THE STUDY

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). There may be other hidden causal interactions interconnecting the nodes of the complex networks we inferred that were not identified because of lack of data. Furthermore, the lack of time-series scRNA-Seq counts is a barrier to understanding the complex dynamics of glioblastoma/GSC networks. The pseudotemporal dynamics consist of inferred cell fate trajectories in a dimensionality-reduced data space (i.e., PCA space) by transcriptional similarity of cell fates. Ribosomal proteins and certain cytoskeletal markers (housekeeping genes) were also not pooled with the differential expression signatures for network inference (Figure S1).

Furthermore, although shown to be insightful in identifying robust network patterns controlling glioblastoma cell fate dynamics, PID is considered a weak statistical method which raises a key issue with respect to the problem of causal inference in network medicine (Zenil et al., 2016, 2019). Generally, the choice of a network inference metric is limited to statistical measures such as PID, which are considered to outperform other weaker methods such as mutual information and correlation metrics. Prospective studies should exploit BDM as a network metric followed by BDM-based network perturbation analysis should be considered (Zenil et al., 2019). However, BDM methods are not optimized for more complex networks in fuzzy systems and are currently limited to binary information networks, wherein the scRNA-Seq counts (and thus the network weights/edges) must be binarized at a chosen threshold. Prospective studies should also consider exploiting soft computing methods in cancer network inference.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Material availability
 - Data and code availability
- METHOD DETAILS
 - General methodological framework
 - Single-cell datasets
 - Clustering techniques
 - scEpath Waddington Landscape reconstruction algorithm
 - Mapping pseudotemporal ordering and cell lineage bifurcations in glioblastoma/GSC cell fates
 - Fractal and multifractal analysis
 - Partial Information Decomposition and Context network inference
 - Block Decomposition Method calculations
 - Machine learning and block decomposition analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104179>.

ACKNOWLEDGMENTS

MC was funded by Natural Science and Engineering Research Council of Canada Discovery Grant RGPIN-2018-04546 and an Fonds de recherche du Québec-Santé Research Scholar grant (J1).

AUTHOR CONTRIBUTIONS

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

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

Received: February 7, 2022

Revised: March 16, 2022

Accepted: March 24, 2022

Published: May 20, 2022

SUPPORTING CITATIONS

The following reference appears in the Supplemental Information: [Amson et al., 2013](#); [Azim et al., 2009](#); [Baritaki et al., 2009](#); [Bhaduri et al., 2020](#); [Bhat et al., 2011](#); [Brozzi et al., 2009](#); [Bulstrode et al., 2017](#); [de la Rocha et al., 2014](#); [Ehmsen et al., 2013](#); [Engel et al., 2020](#); [Gao et al., 2021](#); [Grimm et al., 2020](#); [Gu et al., 2014](#); [Hagiwara, 2011](#); [Harroch et al., 2002](#); [Hou et al., 2016](#); [Hoxha et al., 2020](#); [Kozioł et al., 2007](#); [Liu et al., 2013](#); [Lü et al., 2020](#); [Luxen et al., 2017](#); [Mani et al., 2008](#); [Miao et al., 2019](#); [Polisetty et al., 2012](#); [Roy et al., 2020](#); [Saito et al., 2007](#); [Shi et al., 2017](#); [Tao et al., 2020](#); [Thompson et al., 2009](#); [Wang et al., 2013, 2015](#); [Yamasaki et al., 2001](#); [Yu et al., 2020](#).

REFERENCES

- Aliferis, C., and Trafalis, D.T. (2015). Glioblastoma multiforme: pathogenesis and treatment. *Pharmacol. Ther.* 152, 63–82. <https://doi.org/10.1016/j.pharmthera.2015.05.005>.
- Amson, R., Pece, S., Marine, J.C., Di Fiore, P.P., and Teclerman, A. (2013). TPT1/TCTP-regulated pathways in phenotypic reprogramming. *Trends Cell Biol.* 23, 37–46. <https://doi.org/10.1016/j.tcb.2012.10.002>.
- Azim, E., Jabaudon, D., Fame, R.M., and Macklis, J.D. (2009). SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nat. Neurosci.* 12, 1238–1247. <https://doi.org/10.1038/nn.2387>.
- Baish, J.W., and Jain, R.K. (1998). Cancer, angiogenesis and fractals. *Nat. Med.* 4, 984. <https://doi.org/10.1038/1952>.
- Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. <https://doi.org/10.1038/nrg1272>.
- Barabási, A.L., and Posfai, M. (2016). *Network Science* (Cambridge University Press).
- Baritaki, S., Chatzinikola, A.M., Vakis, A.F., Soultzlis, N., Karabetos, D.A., Neonakis, I., Bonavida, B., and 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 Invest.* 27, 184–192. <https://doi.org/10.1080/07357900802210760>.
- Bhaduri, A., Di Lullo, E., Jung, D., Müller, S., Crouch, E.E., Espinosa, C.S., Ozawa, T., Alvarado, B., Spatazza, J., Cadwell, C.R., et al. (2020). Outer radial glia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell* 26, 48–63.e6. <https://doi.org/10.1016/j.stem.2019.11.015>.
- Bhat, K.P., Salazar, K.L., Balasubramanian, V., Wani, K., Heathcock, L., Hollingsworth, F., James, J.D., Gumin, J., Diefes, K.L., Kim, S.H., et al. (2011). The transcriptional coactivator TAZ regulates mesenchymal differentiation in malignant glioma. *Genes Dev.* 25, 2594–2609. <https://doi.org/10.1101/gad.176800.111>.
- Bozdag, S., Li, A., Baysan, M., and Fine, H.A. (2014). Master regulators, regulatory networks, and pathways of glioblastoma subtypes. *Cancer Inform.* 13, 33–44. <https://doi.org/10.4137/CIN.S14027>.
- Brozzi, F., Arcuri, C., Giambanco, I., and Donato, R. (2009). S100B protein regulates astrocyte shape and migration via interaction with src kinase: implications for astrocyte development, activation, and tumor growth. *J. Biol. Chem.* 284, 8797–8811. <https://doi.org/10.1074/jbc.M805897200>.
- Bulstrode, H., Johnstone, E., Marques-Torres, M.A., Ferguson, K.M., Bressan, R.B., Blin, C., Grant, V., Gogolok, S., Gangoso, E., Gagrira, S., et al. (2017). Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes Dev.* 31, 757–773. <https://doi.org/10.1101/gad.293027.116>.
- Chan, T.E., Stumpf, M., and Babbie, A.C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5, 251–267.e3. <https://doi.org/10.1016/j.cels.2017.08.014>.
- 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>.
- Couturier, C.P., Ayyadury, S., Le, P.U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., et al. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Comm.* 11, 3406. <https://doi.org/10.1038/s41467-020-17186-5>.
- de la Rocha, A.M., Sampron, N., Alonso, M.M., and Matheu, A. (2014). Role of SOX family of transcription factors in central nervous system tumors. *Am. J. Cancer Res.* 4, 312–324.
- Ehmsen, J.T., Ma, T.M., Sason, H., Rosenberg, D., Ogo, T., Furuya, S., Snyder, S.H., and Wolosker, H. (2013). D-serine in glia and neurons derives from 3-phosphoglycerate dehydrogenase.

- J. Neurosci. 33, 12464–12469. <https://doi.org/10.1523/JNEUROSCI.4914-12.2013>.
- Engel, A.L., Lorenz, N.I., Klann, K., Munch, C., Steinbach, C., Ronellenfitch, J.P., and Luger, A.L. (2020). Serine-dependent redox homeostasis regulates glioblastoma cell survival. *Br. J. Cancer* 122, 1391–1398. <https://doi.org/10.1038/s41416-020-0794-x>.
- Gao, M., Fu, Y., Zhou, W., Gui, G., Lal, B., Li, Y., Xia, S., Ji, H., Eberhart, C.G., Lateral, J., and Ying, M. (2021). EGFR activates a TAZ-driven oncogenic program in glioblastoma. *Cancer Res.* 81, 3580–3592. <https://doi.org/10.1158/0008-5472.CAN-20-2773>.
- GeneCards: the human gene database (Weizmann Institute of Science). <https://www.genecards.org/> (GCID:GC12P013196;GC08M124550).
- Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., and Corydon, T.J. (2020). The role of SOX family members in solid tumours and metastasis. *Sem. Cancer Biol.* 67, 122–153. <https://doi.org/10.1016/j.semcancer.2019.03.004>.
- Gu, X., Yao, L., Ma, G., Cui, L., Li, Y., Liang, W., Zhao, B., and Li, K. (2014). TCTP promotes glioma cell proliferation in vitro and in vivo via enhanced β -catenin/TCF-4 transcription. *Neurooncology* 16, 217–227. <https://doi.org/10.1093/neuonc/not194>.
- Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. *Eur. J. Neurosci.* 27, 865–873. <https://doi.org/10.1111/j.1460-9568.2008.06061.x>.
- Hagiwara, N. (2011). Sox6, jack of all trades: a versatile regulatory protein in vertebrate development. *Dev. Dynam.* 240, 1311–1321. <https://doi.org/10.1002/dvdy.22639>.
- Harroch, S., Furtado, G., Brueck, W., Rosenbluth, J., Lafaille, J., Chao, M., Buxbaum, J.D., and Schlessinger, J. (2002). A critical role for the protein tyrosine phosphatase receptor type Z in functional recovery from demyelinating lesions. *Nat. Genet.* 32, 411–414. <https://doi.org/10.1038/ng1004>.
- Harutyunyan, A.S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C.L.C., Belle, J., Mikael, L.J., et al. (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Comm.* 10, 1262. <https://doi.org/10.1038/s41467-019-09140-x>.
- Hou, A., Zhao, L., Zhao, F., Wang, W., Niu, J., Li, B., Zhou, Z., and Zhu, D. (2016). Expression of MECOM is associated with unfavorable prognosis in glioblastoma multiforme. *Oncotargets Ther.* 9, 315–320. <https://doi.org/10.2147/OTT.S95831>.
- Hoxha, S., Shepard, A., Troutman, S., Diao, H., Doherty, J.R., Janiszewska, M., Witwicki, R.M., Pipkin, M.E., Ja, W.W., Karetka, M.S., and Kissil, J.L. (2020). YAP-mediated recruitment of YY1 and EZH2 represses transcription of key cell-cycle regulators. *Cancer Res.* 80, 2512–2522. <https://doi.org/10.1158/0008-5472.CAN-19-2415>.
- Huang, S., Ernberg, I., and Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* 20, 869–876. <https://doi.org/10.1016/j.semcdb.2009.07.003>.
- Iacono, G., Massoni-Badosa, R., and Heyn, H. (2019). Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* 20, 110. <https://doi.org/10.1186/s13059-019-1713-4>.
- Jessa, S., Blanchet-Cohen, A., Krug, B., Vladouiu, M., Coutelier, M., Faury, D., Poreau, B., De Jay, N., Hébert, S., Monlong, J., et al. (2019). Stalled developmental programs at the root of pediatric brain tumors. *Nat. Genet.* 51, 1702–1713. <https://doi.org/10.1038/s41588-019-0531-7>.
- Jia, D., Jolly, M.K., Kulkarni, P., and Levine, H. (2017). Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers* 9, 70. <https://doi.org/10.3390/cancers9070070>.
- Jin, S., MacLean, A.L., Peng, T., and Nie, Q. (2018). scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 34, 2077–2086. <https://doi.org/10.1093/bioinformatics/bty058>.
- Jones, C., Karajannis, M.A., Jones, D., Kieran, M.W., Monje, M., Baker, S.J., Becher, O.J., Cho, Y.J., Gupta, N., Hawkins, C., et al. (2017). Pediatric high-grade glioma: biologically and clinically in need of new thinking. *Neurooncology* 19, 153–161. <https://doi.org/10.1093/neuonc/now101>.
- 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>.
- Koziol, M.J., Garrett, N., and Gurdon, J.B. (2007). Tpt1 activates transcription of oct4 and nanog in transplanted somatic nuclei. *Curr. Biol.* 17, 801–807. <https://doi.org/10.1016/j.cub.2007.03.062>.
- Krieger, M.S., Moreau, J.M., Zhang, H., Chien, M., Zehnder, J.L., and Craig, M. (2020). A blueprint for identifying phenotypes and drug targets in complex disorders with empirical dynamics. *Patterns* 1, 100138. <https://doi.org/10.1016/j.patter.2020.100138>.
- Latora, V., Nicosia, V., and Russo, G. (2017). Centrality Measures. *Complex Networks: Principles, Methods and Applications* (Cambridge University Press).
- Liu, J., Guo, S., Li, Q., Yang, L., Xia, Z., Zhang, L., Huang, Z., and Zhang, N. (2013). Phosphoglycerate dehydrogenase induces glioma cells proliferation and invasion by stabilizing forkhead box M1. *J. Neurooncol.* 111, 245–255. <https://doi.org/10.1007/s11060-012-1018-x>.
- Lü, L., Niu, L., and Hu, J. (2020). At last in" the physiological roles of the tubular ER network. *Biophys. Rep.* 6, 105–114. <https://doi.org/10.1007/s41048-020-00113-y>.
- Lulla, R.R., Saratsis, A.M., and Hashizume, R. (2016). Mutations in chromatin machinery and pediatric high-grade glioma. *Sci. Adv.* 2, e1501354. <https://doi.org/10.1126/sciadv.1501354>.
- Luxen, D., Gielen, G.H., Waha, A., Isselstein, L., Müller, T., Koch, P., Hammes, J., Becker, A., Simon, M., Wurst, P., et al. (2017). MTSS1 is epigenetically regulated in glioma cells and inhibits glioma cell motility. *Trans. Oncol.* 10, 70–79. <https://doi.org/10.1016/j.tranon.2016.11.006>.
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature* (W.H. Freeman and Company).
- Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shiptsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133, 704–715. <https://doi.org/10.1016/j.cell.2008.03.027>.
- 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., et al. (2019). Gata2 as a crucial regulator of stem cells in adult hematopoiesis and acute myeloid Leukemia. *Stem Cell Rep.* 13, 291–306. <https://doi.org/10.1016/j.stemcr.2019.07.005>.
- Miao, L., Jiang, Z., Wang, J., Yang, N., Qi, Q., Zhou, W., Feng, Z., Li, W., Zhang, Q., Huang, B., et al. (2019). Epithelial membrane protein 1 promotes glioblastoma progression through the PI3K/AKT/mTOR signaling pathway. *Oncol. Rep.* 42, 605–614. <https://doi.org/10.3892/or.2019.7204>.
- 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, 835–849.e21. <https://doi.org/10.1016/j.cell.2019.06.024>.
- Paugh, B.S., Qu, C., Jones, C., Liu, Z., Adamowicz-Brice, M., Zhang, J., Bax, D.A., Coyle, B., Barrow, J., Hargrave, D., et al. (2010). Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J. Clin. Oncol.* 28, 3061–3068. <https://doi.org/10.1200/JCO.2009.26.7252>.
- Ping, Y., Deng, Y., Wang, L., Zhang, H., Zhang, Y., Xu, C., Zhao, H., Fan, H., Yu, F., Xiao, Y., and Li, X. (2015). Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data. *Nucleic Acids Res.* 43, 1997–2007. <https://doi.org/10.1093/nar/gkv074>.
- 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., et al. (2012). LC-MS/MS analysis of differentially expressed glioblastoma membrane proteome reveals altered calcium signaling and other protein groups of regulatory functions. *Mol. Cell Proteom.* 11, M111.013565. <https://doi.org/10.1074/mcp.M111.013565>.
- 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. <https://doi.org/10.1038/s43018-020-00154-9>.

Rodrigues, F.A. (2019). Network centrality: an introduction. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1901.07901>.

Roy, S.K., Shrivastava, A., Srivastava, S., Shankar, S., and Srivastava, R.K. (2020). SATB2 is a novel biomarker and therapeutic target for cancer. *J. Cell Mol. Med.* 24, 11064–11069. <https://doi.org/10.1111/jcmm.15755>.

Saito, T., Ikeda, T., Nakamura, K., Chung, U.I., and Kawaguchi, H. (2007). S100A1 and S100B, transcriptional targets of SOX trio, inhibit terminal differentiation of chondrocytes. *EMBO Rep.* 8, 504–509. <https://doi.org/10.1038/sj.embor.7400934>.

Schemionek, M., Kharabi Masouleh, B., Klaile, Y., Krug, U., Hebestreit, K., Schubert, C., Dugas, M., Büchner, T., Wörmann, B., Hiddemann, W., et al. (2015). Identification of the adapter molecule MTSS1 as a potential oncogene-specific tumor suppressor in acute myeloid leukemia. *PLoS One* 10, e0125783. <https://doi.org/10.1371/journal.pone.0125783>.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Khuong Quang, D.A., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 482, 226–231. <https://doi.org/10.1038/nature10833>.

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

Shi, Y., Ping, Y.F., Zhou, W., He, Z.C., Chen, C., Bian, B.S., Zhang, L., Chen, L., Lan, X., Zhang, X.C., et al. (2017). Tumour-associated macrophages secrete pleiotrophin to promote PTPRZ1 signalling in glioblastoma stem cells for tumour growth. *Nat. Comm.* 8, 15080. <https://doi.org/10.1038/ncomms15080>.

Stelzer, G., Rosen, R., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards suite: from gene data mining to disease genome sequence analysis. *Curr. Prot. Bioinform.* 54, 1.30.1–1.30.33. <https://doi.org/10.1002/cpbi.5>.

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

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y.,

Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., et al. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22, 425–437. <https://doi.org/10.1016/j.ccr.2012.08.024>.

Sun, J., Gong, X., Purow, B., and Zhao, Z. (2012). Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma. *PLoS Comp. Biol.* 8, e1002488. <https://doi.org/10.1371/journal.pcbi.1002488>.

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., et al. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* 157, 580–594. <https://doi.org/10.1016/j.cell.2014.02.030>.

Tao, W., Zhang, A., Zhai, K., Huang, Z., Huang, H., Zhou, W., Huang, Q., Fang, X., Prager, B.C., Wang, X., et al. (2020). SATB2 drives glioblastoma growth by recruiting CBP to promote FOXM1 expression in glioma stem cells. *EMBO Mol. Med.* 12, e12291. <https://doi.org/10.15252/emmm.202012291>.

Thompson, M.R., Xu, D., and Williams, B.R. (2009). ATF3 transcription factor and its emerging roles in immunity and cancer. *J. Mol. Med.* 87, 1053–1060. <https://doi.org/10.1007/s00109-009-0520-x>.

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

Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110. <https://doi.org/10.1016/j.ccr.2009.12.020>.

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

Wang, H., Zhang, L., Zhang, I.Y., Chen, X., Da Fonseca, A., Wu, S., Ren, H., Badie, S., Sadeghi, S., Ouyang, M., et al. (2013). S100B promotes glioma growth through chemoattraction of myeloid-derived macrophages. *Clin. Cancer Res.*

19, 3764–3775. <https://doi.org/10.1016/j.canlet.2018.07.034>.

Wang, Z., Yuan, H., Sun, C., Xu, L., Chen, Y., Zhu, Q., Zhao, H., Huang, Q., Dong, J., and Lan, Q. (2015). GATA2 promotes glioma progression through EGFR/ERK/Elk-1 pathway. *Med. Oncol.* 32, 87. <https://doi.org/10.1007/s12032-015-0522-1>.

Wang, J., Huang, T.Y., Hou, Y., Bartom, E., Lu, X., Shilatifard, A., Yue, F., and Saratsis, A. (2021). Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Sci. Adv.* 7, eabg4126. <https://doi.org/10.1126/sciadv.abg4126>.

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. (Addison-Wesley), pp. 183–189.*

Xiong, S., Feng, Y., and Cheng, L. (2019). Cellular reprogramming as a therapeutic target in cancer. *Trends Cell Biol.* 29, 623–634. <https://doi.org/10.1016/j.tcb.2019.05.001>.

Yabo, Y.A., Niclou, S.P., and Golebiewska, A. (2021). Cancer cell heterogeneity and plasticity: a paradigm shift in glioblastoma. *Neurooncology*, noab269. <https://doi.org/10.1093/neuonc/noab269>.

Yamasaki, M., Yamada, K., Furuya, S., Mitoma, J., Hirabayashi, Y., and Watanabe, M. (2001). 3-phosphoglycerate dehydrogenase, a key enzyme for serine biosynthesis, is preferentially expressed in the radial glia/astrocyte lineage and olfactory ensheathing glia in the mouse brain. *J. Neurosci.* 21, 7691–7704. <https://doi.org/10.1523/jneurosci.21-19-07691.2001>.

Yu, J., Shen, W., Gao, B., Xu, J., and Gong, (2020). B. Metastasis suppressor 1 acts as a tumor suppressor by inhibiting epithelial-to-mesenchymal transition in triple-negative breast cancer. *Int. J. Biol. Markers* 35, 74–81. <https://doi.org/10.1177/1724600820905114>.

Zenil, H., Kiani, N.A., and Tegnér, J. (2016). Methods of information theory and algorithmic complexity for network biology. *Semin. Cells Dev. Biol.* 51, 32–43. <https://doi.org/10.1016/j.semcdb.2016.01.011>.

Zenil, H., Kiani, N.A., Marabita, F., Deng, Y., Elias, S., Schmidt, A., Ball, G., and Tegnér, J. (2019). An algorithmic information calculus for causal discovery and reprogramming systems. *iScience* 19, 1160–1172. <https://doi.org/10.1016/j.isci.2019.07.043>.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Glioblastoma Single Cell Datasets	https://singlecell.broadinstitute.org/single_cell/study/SCP393/single-cell-rna-seq-of-adult-and-pediatric-glioblastoma#study-summary	SCP393
GSC Single Cell Datasets	https://singlecell.broadinstitute.org/single_cell/study/SCP503/gradient-of-developmental-and-injury-reponse-transcriptional-states-define-functional-vulnerabilities-underpinning-glioblastoma-heterogeneity#study-download	SCP503
Software and algorithms		
https://github.com/Abicumaran/GBM_Complexity_I	In this paper	https://doi.org/10.5281/zenodo.6371481
Seurat v3	https://github.com/satijalab/seurat/	
BigScale V2	https://github.com/iaconogi/BigScale2	
scEpath	https://github.com/sqjin/scEpath	
Online Algorithmic Complexity Calculator (OACC)	https://github.com/algorithmicnaturelab/OACC	
Network Inference	https://github.com/Tchanders/NetworkInference.jl	
Julia LightGraphs v1.3	https://github.com/JuliaGraphs/SimpleWeightedGraphs.jl	
SciKit-learn	https://github.com/scikit-learn/scikit-learn	
FracLac v2.5	https://imagej.nih.gov/ij/plugins/fractalac	

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the [lead contact](#), Morgan Craig: morgan.craig@umontreal.ca

Material availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the [key resources table](#).
- All original code has been deposited at https://github.com/Abicumaran/GBM_Complexity_I and is publicly available as of the date of publication. A DOI is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request. Further details of the software/algorithm packages are provided below:

Seurat

Project name: Seurat V3

Project home page: <https://github.com/satijalab/seurat/>

Archived version: 10.1016/j.cell.2019.05.031

Programming language: R

License: GNU Public License (GPL 3.0)

BigScale

Project name: BigScale V2

Project home page: <https://github.com/iaconogi/BigScale2>

Archived version: 10.1186/s13059-019-1713-4

Programming language: R

Other requirements: C++

scEpath

Project name: single-cell Energy path (scEpath)

Project home page: <https://github.com/sqjin/scEpath>

Archived version: 10.1093/bioinformatics/bty058

Programming language: MATLAB

Other requirements: C++

OACC

Project name: Online Algorithmic Complexity Calculator V3

Project home page: <https://github.com/algorithmicnaturelab/OACC>

Archived version: 10.1016/j.isci.2019.07.043

Programming language: R

License: GNU Public License (GPL 3.0)

Network inference

Project name: NetworkInference.jl and Partial Information Decomposition (PID)

Project home page: <https://github.com/Tchanders/NetworkInference.jl>

Archived version: 10.1016/j.cels.2017.08.014

Programming language: Julia

License: MIT "Expat" License

Julia LightGraphs

Project name: LightGraphs.jl V1.3

Project home page: <https://github.com/JuliaGraphs/SimpleWeightedGraphs.jl>

Programming language: Julia

Other requirements: Jupyter Notebook and HTML

License: MIT "Expat" License

SciKit-learn

Project name: Scikit-learn

Project home page: <https://scikit-learn.org/> or <https://github.com/scikit-learn/scikit-learn>

Archived version: <http://jmlr.org/papers/v12/pedregosa11a.html>

Operating system(s): Platform independent

Programming language: Python (\geq V3.7)

Other requirements: NumPy (\geq 1.14.6), SciPy (\geq 1.1.0), joblib (\geq 0.11), threadpoolctl (\geq 2.0.0),

Google Colab or Jupyter Notebook

License: 3-Clause BSD license

FracLac

Project name: FracLac V2.5

Project home page: <https://imagej.nih.gov/ij/plugins/fraclac>

Programming language: Java

License: National Institute of Health (NIH) Public License

METHOD DETAILS

General methodological framework

To understand glioblastoma network complexity, we integrated several pediatric and adult IDH-wt glioblastoma single-cell RNA-Seq (scRNA-Seq) datasets in an analytical pipeline that combines several network reconstruction and analysis tools (see subsections below). Details of the datasets used are provided in [Table 1](#). Single-cell datasets were first filtered and normalized in a quality control step, and patient samples were removed from the scRNA-Seq counts expression matrix due to low unique molecular identifier (UMI)/high drop-out rates.

Next, gene expression matrices were analyzed independently using the various clustering and trajectory inference algorithms discussed below. Here we provide a short summary. For the Seurat algorithm, the top 10 principal component analysis (PCA) loadings were used for the differential marker discovery; the top 25 PC loadings were used for the BigScale analysis. To identify the differential markers expressed in all clusters, the top 10 markers within these PC loadings were pooled and analyzed on the UMAP/tSNE patterning space of the cell fate clusters for each patient group. Similarly, the top 2 PCA loadings were used by the scEpath pseudotime analysis. The normalized scRNA-Seq counts of the discovered markers from the Seurat and BigScale algorithms were pooled together, and separately analyzed for each patient group. The expression counts of these markers were then run through the PIDC Network Inference algorithm to obtain gene receptor networks. The differential transcription factors identified in the pseudotemporal progression heatmaps were selected for scEpath analysis. Only the markers specific to each patient group were selected for the PIDC network inference. Lastly, complex networks analysis was performed on the reconstructed networks using transitivity and centrality scores to assess the network structure and dynamics (information flow) to identify key regulators of glioblastoma/GSC cell fate decisions. Further, algorithmic complexity measures (see below) were used to identify gene markers which could accurately discriminate the patient groups by machine learning classifiers. Within the established gene networks, algorithmic complexity was used to identify robust discriminants that could accurately distinguish the three patient groups (i.e., pediatric glioblastoma, adult glioblastoma, and adult GSC), based on the performance of machine learning classifiers on their algorithmic complexity scores (see [Supplementary Information](#)).

Single-cell datasets

Gene expression matrices for pediatric glioblastoma, adult glioblastoma, and adult GSC were obtained from the SingleCell Portal repositories from [Nefitel et al. \(2019\)](#) and [Richards et al. \(2021\)](#) ([Table 1](#)). Briefly, glioblastoma patient samples from [Nefitel et al. \(2019\)](#) contained the single cell RNA-Seq counts of four phenotypes (or cellular states): macrophages, malignant glioblastoma cells, oligodendrocytes, and T-cells. Adult GSC consisted only of stem cells. Overall, our dataset included 28 adult GSC datasets, 7 pediatric glioblastoma, and 18 adult glioblastoma scRNA-Seq expression count matrices.

As a quality control measure for the Seurat and BigScale clustering, two adult glioblastoma samples and one pediatric glioblastoma sample were dropped in the filtering process (prior to clustering) due to high zero-counts (i.e., low UMI). Importantly, we confirmed that our findings were insensitive to the number of patient samples within each patient group: including these removed samples did not change the differential expression analysis. To further validate this finding, one sample was randomly chosen and dropped from the total number of samples from each patient group to verify whether the clustering analysis changed (i.e., leave-out-one cross-validation) and we confirmed the clustering results were identical. Beyond 2500 cells, the computational time complexity of the scEpath algorithm increased. Thus, the total cell counts of all three patient groups were kept at the maximum computational threshold for the scEpath analysis (see GitHub link in [Data and code availability](#) in the [key resource table](#)). Further, to visualize the cell fate

attractor dynamics at the same fine-scale resolution for all patient groups, cell counts were kept roughly the same for each glioblastoma type. Selecting a different combination of adult GSC samples did not change the scEpath landscape or results, as the trial of multiple random selections (>6 distinct combinations) reproduced identical results.

Smart-seq2 whole transcriptome amplification, library construction, and sequencing were taken from (Nefitel et al., 2019). For a subset of samples in (Nefitel et al., 2019), single cells were processed via the 10× Chromium 30 Single Cell Plat-form using the Chromium Single Cell 30 Library, Gel Bead and Chip Kits (10× Genomics, Pleasanton, CA). 7,000 cells were added to each channel of a chip partitioned into Gel Beads in Emulsion (GEMs), followed by cell lysis and barcoded reverse transcription of RNA in droplets-Seq. De-emulsion was followed by amplification, fragmentation, and addition of adaptor and sample index (Nefitel et al., 2019). Similar treatment conditions were applied for the GSC count matrices with >69,000 adult GSC cells extracted from 26 patients (Richards et al., 2021).

Among filtered cells, an average of 5,730 genes per cell were found as a quality measure. Expression levels were quantified as $E_{i,j} = \log_2\left(\frac{TPM_{i,j}}{10} + 1\right)$, where $TPM_{i,j}$ refers to transcript-per-million for gene i in sample j , as calculated by RSEM (Nefitel et al., 2019). TPM values were divided by 10 given that the complexity of single cell libraries was estimated to be on the order of 100,000 transcripts. For the remaining cells, the aggregate expression of each gene was calculated as $E_{a_i} = \log_2(\text{average}(TPM_{i,j}) + 1)$ for $i = 1 \dots n$ (Nefitel et al., 2019), then defined the relative expression over the remaining cells by centering the expression levels per gene, i.e., $Er_{i,j} = E_{i,j} - \text{average}[E_i]$

Clustering techniques

Clustering algorithms were used to identify differential markers co-expressed within all patient groups and distinguish a robust network regulating the cell fate dynamics across all phenotypes.

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 (k-nearest neighbor (kNN) graph) based on the Euclidean distance of the 10 PCA loadings using the FindNeighbors function and Louvain community detection algorithm (modularity optimization) using the FindClusters function (parameter can be tuned between 0.4 and 1.2 for optimal results), to cluster cells by their Jaccard index-expression similarity (see Seurat Clustering tutorial in GitHub code). All clustering parameters were kept to their default settings. Next, the cells within the graph-based clusters were visualized on 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). Differential markers from the top 10 PCA loadings were visualized in UMAP space (analysis does not vary for TSNE space) using the FindAllMarkers function with parameters: min.pct = 0.25 and logfc.threshold = 0.25. We clustered similarly expressed cells together in the low dimensional space by finding differentially expressed features/markers corresponding to the highest ten PCA loadings in the graph-based clusters. To identify markers that govern disease progression and transcriptional dynamics, we imposed the condition that selected markers for the network reconstruction must be expressed in all clusters of the three patient groups (pediatric glioblastoma, adult glioblastoma, and adult GSC).

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. Cellular clustering was established by first computing all pairwise cell distances using the Pearson correlation to generate a distance matrix. Following, cells were assigned to cluster groups via the Ward's linkage/method (an agglomerative hierarchical clustering algorithm). Iterative differential expression analysis was performed between the clusters of cells and the differential markers within the identified clusters were assessed using the getMarkers function (see BigScale 2 tutorial in Github code). 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 as a cut-off threshold while the `min_ODscore` parameter was kept default at 2.33. 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.

ScEpath algorithm

We applied single cell Energy path (scEpath) to reconstruct the 3D-energy landscape of cells and infer regulatory relationships from their transcriptional dynamics (Jin et al., 2018). 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. In this trajectory inference algorithm, information flow and network reconfiguration underlying the cellular decision-making steer the topography of cell populations' energy landscapes (also referred to as a cell fate landscape, attractor landscape, or Waddington's epigenetic landscape (Waddington, 1957)). A cell state (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-celllike 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 below.

To perform the scEpath analysis on our data, we first pre-processed the log-normalized (within patient-groups) count matrices with respect to their gene expression values by filtering out zero counts. The differential markers were selected from the first two significant PCA components. We then ran the scEpath MATLAB code from Jin et al., 2018 on these processed datasets. GSC patient samples BT127, BT48, and BT84 from Richards et al. (2021) were used for all scEpath analyses on GSC. Seven pediatric glioblastoma samples from Neftel et al., 2019 and seven adult glioblastoma 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 glioblastoma samples. We then ran energy (Waddington) landscapes reconstruction on the following population sizes: pediatric glioblastoma: $n = 7$, $N = 1850$ cells; adult glioblastoma: $n = 7$, $N = 2221$ cells; adult GSC: $n = 3$, $N = 1504$ cells.

scEpath smooths the average normalized expression of each gene using cubic regression splines to map the pseudotemporal gene expression dynamics along the inferred trajectories of the cell fates on the landscape, leading to smoothed gene expression along a lineage path (Jin et al., 2018). Leveraging this, 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., $\log_2(\text{fold-change}) > 1$). The probabilistic-directed graph network and the cell lineage hierarchy inference parameters were kept at default settings (`quick_construct = 1`; `tau = 0.4`; `alpha = 0.01`; `theta1 = 0.8`). The pseudotime-dependent genes were identified using parameters `sd_thresh = 0.5`; `sig_thresh = 0.01`; `nboot` (see Hyperparameter-optimized code in GitHub link).

scEpath Waddington Landscape reconstruction algorithm

The scEpath algorithm performs principal component analysis (PCA) analysis on the energy matrix $E = (E_{ij})$ and fits a potential energy surface using piecewise linear interpolation over the first two PCA components and the single cell energy (scEnergy) of each cell. Cells are then colored according to unsupervised

clustering which groups cells with similar gene expression patterns (transcriptional states). The energy of each cell state (scEnergy), E_i , on the Waddington landscape is computed according to:

$$E_i(\mathbf{y}) = \sum_{j=1}^n E_{ij}(\mathbf{y}) = - \sum_{j=1}^n y_{ij} \ln \frac{y_{ij}}{\sum_{k \in N_i} y_{kj}}$$

where y_{ij} represents the normalized gene expression level (between 0 and 1) of gene i and cell j , and $N(i)$ is the neighborhood of node- i in the network. Each gene is assigned a local energy state E_{ij} (Jin et al., 2018). The scEnergy is combined with a distance-based measure and structural clustering to reconstruct the 3D energy landscapes.

The cell-state on the scEpath energy landscape corresponds to which discrete bin its mRNA levels fluctuate within (Jin et al., 2018). The cell states distribution on the scEnergy landscape can thus be defined as attractors, a term from dynamical systems theory used to describe a causal pattern to which the cell state dynamics are bound. Assessing the fractal dimension of this attractor (cell state patterns) for each patient group's energy landscape provides key insights into the complexity of the cell states and the transition gene dynamics governing their cellular decision-making.

Mapping pseudotemporal ordering and cell lineage bifurcations in glioblastoma/GSC cell fates

scEpath performs PCA on the energy matrix and uses the first two PCA components as the reaction coordinates followed by structural clustering of cells via an unsupervised cell-cell similarity metric called single-cell interpretation via multikernel learning (SIMLR) (Jin et al., 2018). To infer cell lineages, scEpath constructs a probabilistic directed graph in which nodes represent phenotypic clusters (attractors), with edges weighted by cell state transition probabilities (Figure S2). By default, scEpath defines the cell clustering patterns as the set of cells occupying 80% percent of the total energy in each cluster. The cell state transition probability such that a cell state will be in cell cluster k with a particular scEnergy was calculated using the Boltzmann–Gibbs distribution by the scEpath algorithm (Jin et al., 2018). The directions of the probability flow are determined by the energy flow with significant changes from high to low. scEpath learns the maximum probability flow in the probabilistic directed graph defined by a weighted matrix W determined by the gene expression counts. This problem is equivalent to finding the minimum directed spanning tree by setting the edge weights to be $1-W$. Thus, scEpath implements Edmonds' algorithm to determine the minimum directed spanning tree (MDST) connecting the cell clusters (attractors) and hence, the candidate cell lineage bifurcations of cell fate decisions (Jin et al., 2018).

Fractal and multifractal analysis

We applied fractal analysis to quantify the complexity of the phenotypic patterns on the scEpath cell fate attractor landscape. 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. 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; Baish and Jain, 1998). As such, the fractal index provides a quantitative measure of the cell fates' phenotypic plasticity (i.e., higher for stem cell-like fates) and disease progression.

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).

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 glioblastoma/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, thus allowing for the quantification of the relationships between identified differential transition markers and tracking how these relationships change across distinct phenotypes. Partial Information Decomposition and Context (PIDC) networks have been suggested to outperform traditional gene regulatory network inference approaches using correlation metrics, 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 glioblastoma and GSC samples.

The Julia packages InformationMeasures.jl and NetworkInference.jl were used to reconstruct the GRN networks. PIDC network inference uses partial information decomposition (PID) to infer regulatory interaction networks from gene expression datasets. We used the NetworkInference.jl 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 then computed and the most differentially expressed genes were identified by the clustering algorithms using PID scores.

We used Julia LightGraphs to infer the PIDC network. This algorithm works by constructing undirected simple weighted graphs that optimize the shortest path of vertices/nodes based on the weights of the network (i.e., the PID scores). The network nodes are ordered by the top PID scores in decreasing order. Given three variables (genes) x , y , and z on the network, partial information decomposition (PID) maps the information obtained from a source set of genes $S = \{x, y\}$ about the target gene z . The information is redundant, synergistic, and unique. PID is an information-theoretic similar to pairwise mutual information, taking into consideration the information dynamics between three-variables at a time (instead of pairwise-correlations) (Chan et al., 2017). Thus, the PID score between the source gene set S and the target gene Z is given by:

$$I(X; X, Y) = Synergy(Z; X, Y) + UniqueY(Z; X) + UniqueX(Z; y) + Redundancy(Z; X, Y)$$

where $(Z; X)$ is the unique information between the source gene X and target gene Z when the other source gene is Y . The PIDC inference algorithm calculates the PID scores.

The PIDC inference algorithm may be simply defined as follows. PID values are estimated for every gene triplet (with each gene treated as the target gene in the others), and from these the proportional unique contribution (PUC), represented as $U_{x, y}$ (as defined below), is estimated for every pair of genes. For each gene X , an empirical distribution $f_X(u)$ is estimated. The confidence of an edge between a pair of genes depends on the corresponding cumulative distribution functions $f_X(u)$ (assumed as either a Gamma or Gaussian empirical probability distribution, for each gene within the pair). These confidence scores are then used to rank all possible network edges. The Julia LightGraphs algorithm uses the cumulative probability distributions for each gene to obtain a final confidence score for network edges (Chan et al., 2017). The corresponding PID scores are obtained as output of the algorithm.

We define the PUC between two genes X and Y as the sum of this ratio calculated using every other gene Z in a network:

$$U_{X, Y} = \sum_{Z \in S(x, y)} \frac{Unique_z(X; Y)}{I(X; Y)} + \sum_{Z \in S(x, y)} Unique_z(X; Y)$$

This measure captures the mean proportion of mutual information (MI) between two genes X and Y . Using an unsupervised Louvain community detection algorithm, the PID network is then inferred from the PID scores of the genes. Such network analyses coupled to the scEpath algorithm helped us identify gene modules which may exhibit oscillatory dynamics in gene expression as cells under-go state transition, and putative gene interactions which may be involved in regulating glioblastoma/GSC cell fate choices.

Block Decomposition Method calculations

We evaluated the algorithmic complexity of key nodes (genes) of the inferred signaling networks to further identify robust markers distinguishing glioblastoma and GSC. Algorithmic complexity is a complementary

measure that identifies the minimal amount or set of information in our inferred complex networks which regulate the phenotypic plasticity dynamics across the patient groups, and as such the genes/TFs with highest algorithmic complexity could be robust disease screening tools in precision oncology. 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, 2019). 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 to be the length of the shortest program p that generates s . However, $K(s)$ is in principle incomputable and must be approximated using the coding theorem method (Zenil et al., 2019). 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 Availability of Data and Material). 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, 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. scRNA-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) (see [Supplementary Information for BDM Results](#)).

Machine learning and block decomposition analysis

Binary classification was performed using Google Colab using Scikit-learn on the BDM scores of three classification groups (pediatric glioblastoma, adult glioblastoma, and GSC). The linear support vector machine (SVM) and AdaBoost random forest (RF) classifier modules (with hyperparameter tuning to optimal learning rates) 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.