



OPEN

A method for in silico exploration of potential glioblastoma multiforme attractors using single-cell RNA sequencing

Marcos Guilherme Vieira Junior¹✉, Adriano Maurício de Almeida Côrtes^{2,3}, Flávia Raquel Gonçalves Carneiro^{4,5,6}, Nicolas Carels⁷ & Fabrício Alves Barbosa da Silva⁸✉

We presented a method to find potential cancer attractors using single-cell RNA sequencing (scRNA-seq) data. We tested our method in a Glioblastoma Multiforme (GBM) dataset, an aggressive brain tumor presenting high heterogeneity. Using the cancer attractor concept, we argued that the GBM's underlying dynamics could partially explain the observed heterogeneity, with the dataset covering a representative region around the attractor. Exploratory data analysis revealed promising GBM's cellular clusters within a 3-dimensional marker space. We approximated the clusters' centroid as stable states and each cluster covariance matrix as defining confidence regions. To investigate the presence of attractors inside the confidence regions, we constructed a GBM gene regulatory network, defined a model for the dynamics, and prepared a framework for parameter estimation. An exploration of hyperparameter space allowed us to sample time series intending to simulate myriad variations of the tumor microenvironment. We obtained different densities of stable states across gene expression space and parameters displaying multistability across different clusters. Although we used our methodological approach in studying GBM, we would like to highlight its generality to other types of cancer. Therefore, this report contributes to an advance in the simulation of cancer dynamics and opens avenues to investigate potential therapeutic targets.

Keywords Single-cell RNA sequencing, Glioblastoma multiforme, Biomarkers, Cancer attractors, Confidence regions, Multistability

Despite substantial progress in comprehension and therapeutic approaches, cancer remains a predominant global cause of mortality. For instance, Glioblastoma multiforme (GBM), the most common and aggressive brain tumor, presents 15 months of average overall survival (OS) with roughly 10% probability of achieving a 5-year OS^{1,2}. Additionally, single-cell RNA sequencing (scRNA-seq) has emphasized the notable heterogeneity in GBM and many types of cancer^{3–6}. The better knowledge of tumor heterogeneity has shown that it might be driving the aggressiveness of these malignancies^{7,8}, emphasizing the need to investigate its underlying dynamics. Particularly, extensive research has examined the influence of mutations and epigenetics on the complex carcinogenesis process^{9–11}, which connects to the malignant state's development according to the gene regulatory networks (GRN) dynamics. In this direction, pivotal studies have identified the correspondence between cell types or subtypes with stable states from system dynamics theory, often termed 'attractors'^{12–14}. These insights into the

¹Graduate Program in Computational and Systems Biology, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro 21040-900, Brazil. ²Department of Applied Mathematics, Institute of Mathematics, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro 21941-909, Brazil. ³Systems Engineering and Computer Science Program, Coordination of Postgraduate Programs in Engineering (COPPE), Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro 21941-972, Brazil. ⁴Center of Technological Development in Health (CDTS), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro 21040-361, Brazil. ⁵Laboratório Interdisciplinar de Pesquisas Médicas, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro 21040-900, Brazil. ⁶Program of Immunology and Tumor Biology, Brazilian National Cancer Institute (INCA), Rio de Janeiro 20231-050, Brazil. ⁷Laboratory of Biological System Modeling, Center of Technological Development in Health (CDTS), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro 21040-361, Brazil. ⁸Scientific Computing Program, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro 21041-222, Brazil. ✉email: marcosvieira.research@gmail.com; fabricio.silva@fiocruz.br

tumor's molecular complexity set the stage for developing frameworks integrating complex systems approaches to cancer research. One important application of systems dynamics theory to cancer is the cancer attractor concept. According to this concept, cancer is a pathological cellular development that creates or increases propensity towards such states^{14–16}. The cancer attractor concept gives a theoretical background to interpret the patterns of gene expression distributions observed in scRNA-seq datasets of tumors, offering insights into cancer's underlying dynamics. It implies that clusters of gene expression observed in scRNA-seq of malignant cells represent cellular populations orbiting within specific attractor states¹⁵, with the clusters' distributions reflecting the regulatory mechanisms, here called the constraints governing the cellular dynamics¹⁷ (see Fig. 1). This framework helps to overcome the lack of time series measurements and opens avenues for investigating the dynamics underlying scRNA-seq snapshot-like data. However, standard scRNA-seq downstream data analysis concentrates on machine learning dimensionality reduction algorithms to perform clustering exploration^{18–20}, focusing on a static characterization to the detriment of a model-building approach. In this direction, developing methods integrating the available data into theoretical models is fundamental to further advancements.

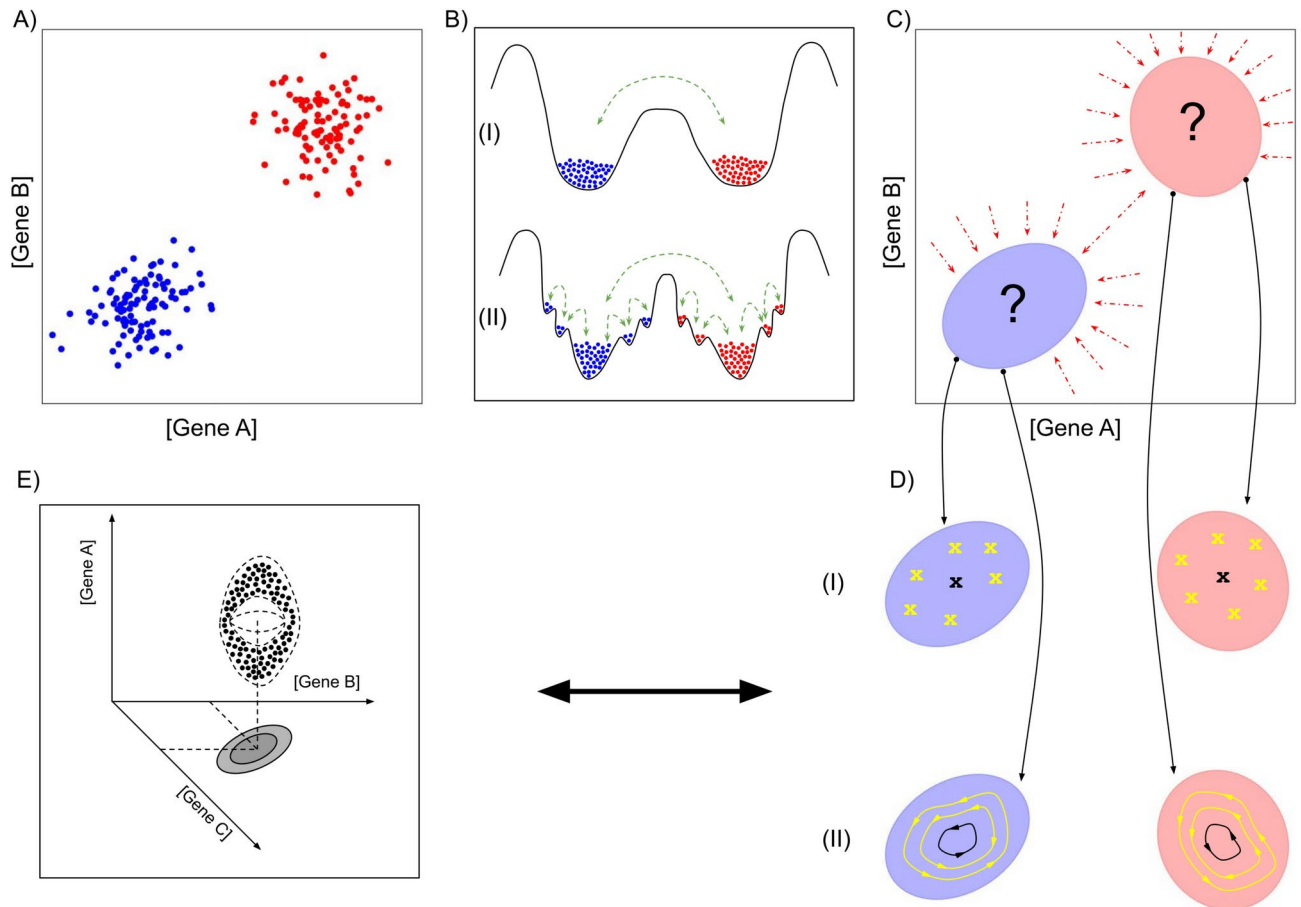


Figure 1. Schematic illustration regarding investigating cancer attractors using scRNA-seq data. **(A)** Depicts an example of the dispersion plot regarding the expression level of genes A and B in an illustrative group of cells. The colors differentiate each cluster with similar expression levels, which are supposed to reflect similar biological regulation constraints. **(B)** Represents a possible interpretation of each cluster's stability. **(I)** Illustrate the clusters as two broader stabilities, called basins of attraction. **(II)** Represents each cluster as a group of smaller basins, each with its attractors. The dotted green arrows point to the transitions. When the attractors are accessible within the same genetic and/or environmental conditions, we call multistability. When they can't traverse to each other and depend on genetic and/or environmental changes, we call alternative stable states. In this case, these changes lead to a shift in the equilibrium states. **(C)** Depicts the uncertainty regarding the underlying dynamics that lead the cells' gene expression to stay mainly limited to the cluster's region, here called confidence region. The point dot red arrows indicate the biological constraints pushing the system toward the regions. **(D)** Illustrate different attractor possibilities. **(I)** The crosses represent the fixed points of stochastic dynamics. The black color illustrates stability in the cluster centroid coordinate and the yellow color other stable states, either alternative stable states or multistability. **(II)** Shows the possibility of underlying multiple limit cycles (closed orbits). The black color illustrates the presence of a single attractor, while the yellow represents the possibility of multiple attractors (alternative stable states or multistability). **(E)** Illustrates the dimensional information loss when projecting from 3D to 2D marker space, which justifies the investigation using confidence regions to study the stability and underlying dynamics.

Recently, multiple frameworks have been developed to integrate complex systems approaches to cancer investigation^{14,21}. For example, significant advancements propose the presence of chaotic cancer attractors²². Additionally, investigations showed parallels between the malignant state development and ecological systems^{23–25}. These parallels allow the integration of knowledge used to answer pivotal questions in ecology, for instance, the investigation of alternative stable states and multistability^{26–29}. According to these concepts, the dynamic interactions of different species and the environment can lead to different equilibrium states. In the cancer attractors' context, different equilibrium states resulting from genetic and/or environmental regulation changes are called alternative stable states. In contrast, potentially accessible stable states under the same genetic and/or environmental conditions are called multistability (see Fig. 1B, D). Combining ecological characteristics of the cancer niche with the cancer attractor concept provides a robust framework to investigate scRNA-seq data. For example, analyzing data distributions can help to understand how a tumor's genetic alterations and phenotypic variability can affect intratumor heterogeneity. One possible path in this direction is the characterization of steady states, focusing the investigation on the data distributions instead of investigating the detailed attractors' trajectories^{30,31}. Vieira et al.³⁰ demonstrated a viable framework for an *in silico* investigation of the stability regarding scRNA-seq data clusters centroids (see Fig. 1D). Nevertheless, the inherent complexity of biological systems yet imposes theoretical and computational limits. In this direction, advancements are still necessary for developing clinical applications.

Aiming for such advancements, this report enhances the framework of Vieira et al.³⁰ by investigating the viability of constraining the stability analysis to a restricted number of marker genes' dimensions. Specifically, we propose a biological-informed clustering that probes known markers' dimensions and correlates the data density to the presence of stable states. This approach improves the biological interpretation of clusters and reduces the dimensional complexity of the problem. To this end, we investigated the efficacy of a density-based clustering algorithm, that aligns with attractor interpretation. However, using markers' projected dimensions led to the issue of information loss (see Fig. 1E). To overcome this problem, we investigated the data density by using confidence regions defined from clustered data. Particularly, we employed ellipsoidal statistics, as described in³². To confirm this approach's feasibility, we compared the density of experimental clusters to clusters obtained by Gaussian sampling. This methodology allowed us to verify the possibility of finding markers reflecting the density of a higher dimensional system. Additionally, we tested the theoretical presence of alternative stable states and multistability by simulating the stochastic dynamics of a GRN containing the tumor markers. We investigated whether clusters might represent regions containing multiple attractors and the clusters' interchangeability. Biologically, the attractors' multiplicity could be due to genetic mutations, epigenetic modifications, or tumor microenvironment conditions. These diverse regulations would affect the GRN dynamics and determine different cell fates.

To evaluate this enhanced methodology, we utilize annotated GBM scRNA-seq data provided by Darmanis et al.³³. The dataset encompasses four patients with a different number of cells classified as each of the four GBM subtypes according to the Verhaak et al. classification³⁴ (Classical, Mesenchymal, Proneural, and Neural). However, this classification is still being developed, with studies pointing to different directions regarding the subtypes and their underlying dynamics^{6,35,36}. This way, we assembled a list of marker genes corresponding to the GBM subtypes to compose our investigation. To simulate the stochastic dynamics, we employed a GRN investigated in Vieira et al.³⁰, which provided us with prior parameters' ranges to analyze. The different gene regulations were modeled by varying the Hill function coefficients and the Vieira et al.³⁰ methodology was enhanced to include confidence regions to select the estimated activation and inhibition strengths. This enhancement allowed the selection and analysis of the parameters' configurations leading to stability and multistability within the confidence regions defined in the markers dimension. The final parameter configuration was assumed to represent possible GRN rewirings and microenvironmental conditions informed by the scRNA-seq data constraints, allowing us to make a parallel with the underlying biological system.

This investigation provided a feasible way to analyze the presence of cancer attractors. Using ellipsoidal statistics within known marker genes' dimensions effectively reduces the problem's complexity, advancing in the direction of practical applications. Additionally, defining confidence regions allows straightforward criteria to automate the selection of multiple parameter configurations. For instance, criteria to select parameters achieving stability within the physiological ranges informed by the constraints specified by the scRNA-seq data clusters (Fig. 1C, D). The combined results allow a data-driven quantification of attractors and multistability. Although we used our methodological approach to study a GBM dataset, we would like to highlight its generality to other types of cancer by testing the corresponding known marker genes. This methodology can be a complementary verification of biomarkers, probing their potential to define cancer attractors. Further, advancing the multistability analysis might be an important way to identify states presenting a higher potential for cancer recurrence. In this way, our investigation opens new avenues for applying single-cell omics technologies to cancer diagnostic and investigating potential therapeutic targets (theranostics).

Methods

Method overview

We present a method to investigate the presence of cancer attractors on annotated scRNA-seq data using confidence regions and their integration into a GRN stochastic dynamic model. Based on our working hypothesis (illustrated in Fig. 1), we seek to probe the data density in the markers' projected dimension compared to a higher dimension and use stochastic simulations to corroborate and quantify the presence of stability. In the cancer context, this framework advances in the direction of practical application for the concept of cancer attractor, while quantifying the presence of stable states and multistability. Figure 2 outlines the steps involved in the proposed method.

The analysis protocol developed for this investigation consisted of two main phases (Fig. 2), one denoted by (I), representing the aspects of scRNA-seq data analysis, and the other by (II), associated with the models' simulation steps. Below we highlight the conceptual implications of the complete steps of each phase.

Phase (I): We started from a chosen cancer scRNA-seq dataset, in our case a GBM dataset, and selected the markers genes (I-A—“[GBM scRNA-seq dataset and selected markers](#)”). This step requires a known list of gene markers according to the cancer data under investigation. The idea is to find the minimum markers' dimension that displays well-defined clusters. To this end, the marker genes and the scRNA-seq data were processed to get the datasets and to construct the GRN (II-A—“[Data preparation and GRN construction](#)”). The GRN already reduces the problem dimensionality and establishes the context within which the combination of the markers will be defined. After verifying the markers' combination, we clustered the data using a density-based clustering method, which aligns with our attractor hypothesis. Then, we uncovered the clusters' centroids and their covariance matrices to define the confidence regions (I-B to I-D—“[Clustering scRNA-seq data: centroids and confidence region](#)”). These confidence regions are the core of our investigation, shifting the focus from analyzing each attractor to probing the space containing the attractors. A positive result in this step allows moving to the stochastic dynamics investigation.

Phase (II): After confirming the feasibility of defining the clusters in the markers' dimension, the corresponding confidence regions can be probed in silico for the presence of stable states. In this direction, we started by specifying a GRN dynamic model (II-B—“[GRN dynamics and implementation](#)”). This step selects the regulation functions and models the nature of the GRN interactions. Following, it is necessary to specify the scaling parameters corresponding to the regulation strengths. To execute this step, we used one model investigated in our previous work³⁰. One important characteristic of this implementation is the possibility of using linear programming for parameter estimation while ensuring parameter biological interpretability. The parameter estimation integrates the scRNA-seq information by using the clusters' centroids as steady states (I-C and II-C—“[Fixed points and parameter estimation](#)”). Finally, we integrate the confidence regions defined in the markers' dimension into the GRN stochastic simulations. This integration aims to select the parameter combinations that achieve stability according to the scRNA-seq data constraints. Biologically, the confidence region aims to ensure we get parameters whose dynamics stay constrained to physiological ranges informed by the experimental data. Additionally, it enables the identification of alternative stable states and multistability (I-D and II-D—“[Stochastic simulations: identifying attractors and multistability](#)”). After discovering the parameters, it is possible to check each region's stability by quantifying the number of parameter configurations, and the more likely clusters to present multistability by identifying the parameters' configuration leading to stable states in multiple regions. The following sections detail the steps of each phase.

GBM scRNA-seq dataset and selected markers

We used the data curated and analyzed by Darmanis et al.³³. This dataset contains single-cell resolution RNA sequencing outputs from patients diagnosed with four GBM subtypes. The authors investigated tumor heterogeneity by contrasting the tumor core with its periphery. This dataset aggregates samples from four patients, all diagnosed with primary GBM and characterized by a negative *IDH1* signature (indicating an absence of mutations in the *IDH* gene). After quality control, the dataset retained information from 3589 cells, including various cell types from the central nervous system (CNS), such as vascular, immune, neuronal, and glial cells.

Darmanis et al.³³ identified cellular clusters from the dimensionality reduction with tSNE and subsequently clustered the dataset via the k-means algorithm. To determine cellular identities, they cross-referenced the clustering results against a previous scRNA-seq dataset from healthy human brain samples. The cross-reference step led to unidentified clusters categorized as neoplastic, with the remaining data related to the major cell types of the CNS and considered non-neoplastic cells (labeled as Regular, as we will refer to it). Subsequent analysis

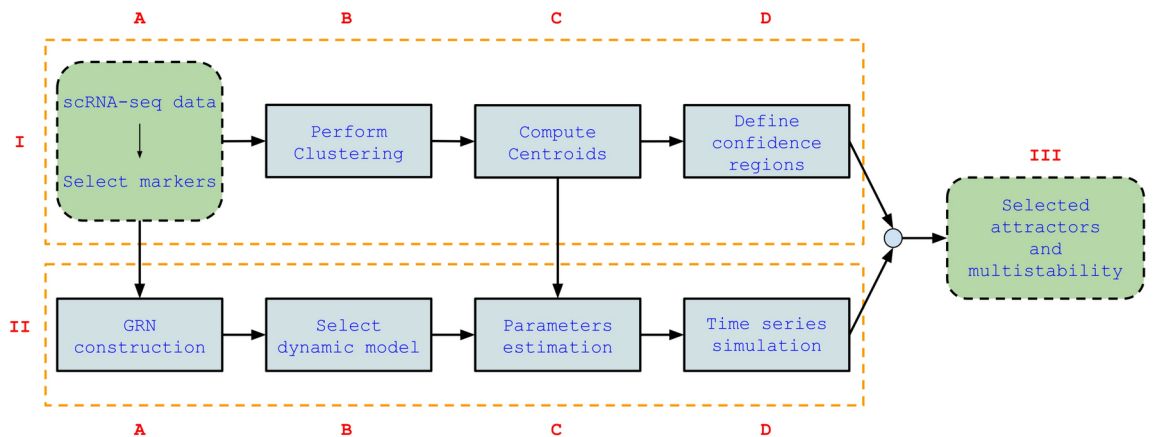


Figure 2. Diagram depicting each stage of our method. We considered two major phases, one denoted by (I), representing the aspects of scRNA-seq data analysis, and the other denoted by (II), associated with the simulation stages. Each phase will be further detailed in the respective methods sections, concluding with the parameters selection and their corresponding attractors.

revealed that 94% of neoplastic clusters originated from tumor core and presented high expression of genes like *EGFR* and *SOX9*. To further improve the confidence in the clusters' identification, the authors conducted an additional comparison with datasets of single-cell and bulk RNA-seq data from healthy human brains and GBM samples, which corroborated the results.

In addition to the *EGFR* gene, observed by Darmanis et al.³³ as presenting high expression values, we focused on *IDH1*, and *CD44* due to their significant roles in GBM pathology. *CD44*, identified as a stem cell marker, has been linked to increased tumor severity³⁷. Notably, the coexpression of *CD44* and *EGFR* has been associated with shorter OS in GBM patients, underscoring the clinical relevance of the *CD44-EGFR* axis in GBM aggressiveness³⁸. Additionally, there is evidence of overexpression of wild-type *IDH1* in Glioblastoma, and several studies have proposed that upregulation of *IDH1* may represent a common metabolic adaptation in GBMs, contributing to enhanced macromolecular synthesis, aggressive tumor growth, and increased resistance to therapy³⁹.

Data preparation and GRN construction

Besides the *EGFR*, *IDH1*, and *CD44* marker genes, we selected a list of genes related to the GBM subtypes or associated with GBM's aggressiveness^{33,34,36–39}. We utilized the 'transcription regulation network construction' tool of the MetaCore⁴⁰ platform to construct the GRN, completing its connectivity. We chose to compose our GRN with regulatory interactions (edges) and genes (vertices) characterized by the binding of transcription factors (TF) to their target gene promoters. As these interactions directly affect the amount of mRNA, we modeled them as direct connections between the transcription factor vertex and the vertex representing the targeted gene.

We used R⁴¹ for the initial data processing and GRN preparation⁴². The complete steps are shown in Fig. 3. Column 'A' shows the phases concerned with the scRNA-seq data processing, and column 'B' shows the processing of the MetaCore output. We started processing the data using the Seurat package⁴³ and applying a *sctransform* normalization to reduce technical bias (A1), recovering biologically significant distributions^{44,45}. We did not remove cell cycle effects because we wanted to preserve as much information as possible and avoid incorporating low-accuracy information of tumor cells⁴⁶. We selected the interactions classified as *Transcription Regulation* (B1) and intersected the GRN genes with the scRNA-seq data (A2 and B2). After reducing the genes of investigation, we filtered the scRNA-seq data into smaller datasets, as described below.

We considered two major dataset groups (A3). The first group consisted of only Neoplastic cells in the tumor core to avoid incorporating the different features specific to Neoplastic cells in the periphery. The second group included the Regular (non-neoplastic) data in the tumor core and periphery. We removed the genes from scRNA-seq data that presented only null values and divided the data into six different datasets (A4). Five datasets related to Neoplastic cells located in the tumor core: one for all Neoplastic data from the tumor core (we will refer to it as BT_All), and one for each one of the four patients, referred by Darmanis et al.³³ as BT_S1, BT_S2, BT_S4, and BT_S6. The last dataset was for all patients' cells labeled as Regular, located both in the tumor core and periphery, which we will refer to as BT_Regular. The number of cells' data in each dataset was 265 for BT_S1, 502 for BT_S2, 134 for BT_S4, 126 for BT_S6, 1027 for BT_All (the sum of each patient), and 2489 for BT_Regular.

Concerning the GRN construction, the intersected genes list comprised 40 genes and their interconnections, which generated a new network in a table format. We employed the new table as an input to a code developed to convert them into two adjacency matrices, one for activation interactions and the other for inhibition interactions⁴⁷. These matrices will be used to automatically construct the dynamic model ("GRN dynamics and implementation").

Clustering scRNA-seq data: centroids and confidence region

Each point of the scRNA-seq data can be expressed as a vector $\mathbf{X} = (X_1, X_2, \dots, X_N)$, with $N = 40$ being the total number of genes or transcription factors present in the GRN and each value X_i corresponding to the scRNA-seq data mRNA molecule quantification. For each dataset mentioned in "Data preparation and GRN construction", namely BT_All, BT_S1, BT_S2, BT_S4, and BT_S6, BT_Regular, the data points are distributed in a 40-dimensional space and agglomerated according to biological processes. Instead of analyzing the whole dimension or utilizing a machine learning method for dimensional reduction, we leveraged biological insights provided by the cancer gene markers (*EGFR*, *IDH1*, and *CD44*) and conducted the cluster analysis of the BT_All dataset in the projected 3-dimensional space, each axis being one of the three markers. We employed Mathematica⁴⁸ to analyze the datasets described in "Data preparation and GRN construction". We used the Neighborhood Contraction (NbC⁴⁹) clustering method, a density-based method that identifies clusters of varying shapes and densities without a prior cluster number definition. We configured the built-in Mathematica function with the 'PerformanceGoal' set to quality, the 'CriterionFunction' set to standard deviation, and the 'DistanceFunction' set to Euclidean distance.

Density-based clustering methods do not present an intrinsic representative point interpretation like centroid-based methods. Nevertheless, we considered each cluster's average a representative point and defined it as the centroids. We visually verified the clusters' symmetry in the 3-marker gene space and used each cluster covariance matrix to construct confidence regions around the centroid coordinates. These confidence regions formed the basis of our cancer attractor investigation and were characterized as the region constrained by an ellipsoid defined as³²:

$$\{\Xi : (\Xi - \mu_{\text{ref}})^T \mathbf{C}_{\text{ref}}^{-1} (\Xi - \mu_{\text{ref}}) \leq \chi_{p,\alpha}^2\} \quad (1)$$

where Ξ represents the data points coordinates in the 3-marker gene spaces, μ_{ref} is a cluster's centroid from the dataset chosen as reference, \mathbf{C} is the cluster covariance matrix, and $\chi_{p,\alpha}^2$ is the critical value of the chi-squared

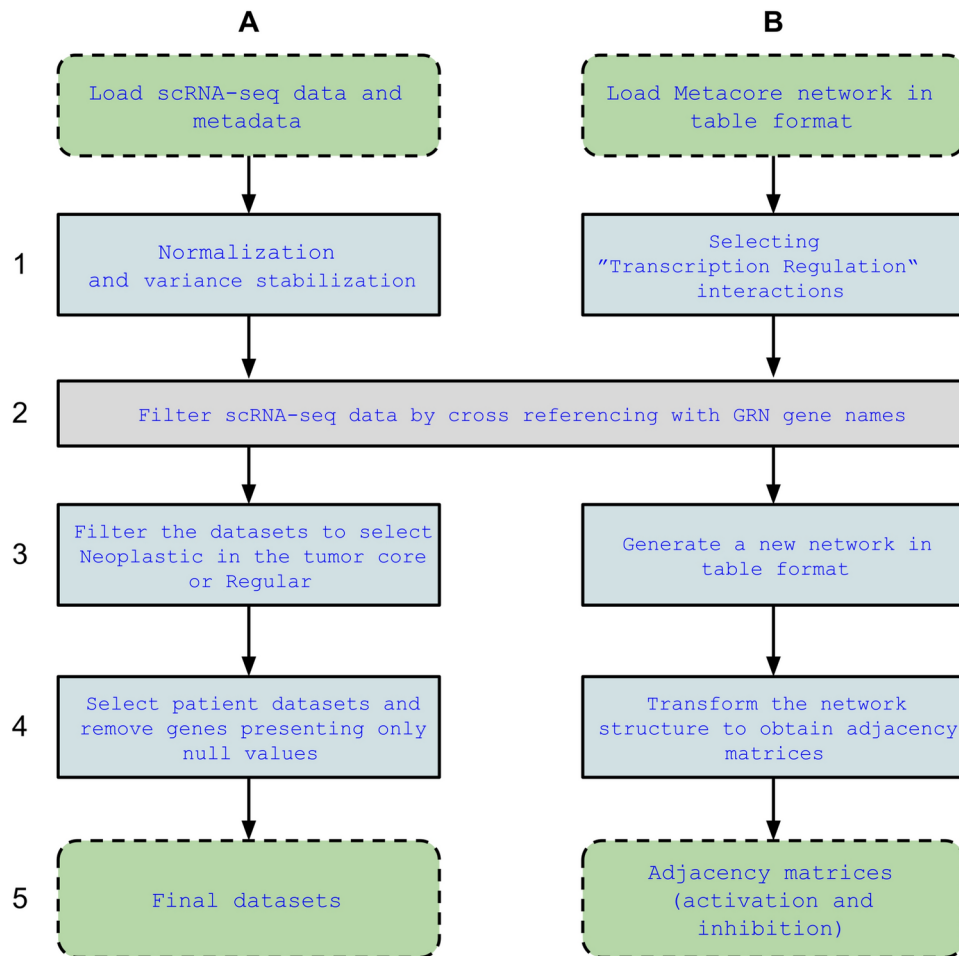


Figure 3. Diagram depicting each stage of the GRN construction and data preparation. Column A displays the steps regarding the scRNA-seq processing, which assembled the datasets for the investigation. Column B displays the steps regarding the GRN and its adjacency matrices construction. The datasets' preparation and the GRN processing were implemented in the same code, with the resulting genes after selecting the "Transcription Regulation" mechanism being intersected with the scRNA-seq data so the new datasets presented the same genes as the GRN.

distribution with p degrees of freedom at significance level α . We selected two significance levels, one leading to a 95% (two standard deviations) and another to a 68% (one standard deviation) confidence region. The 95% confidence region reflected a high uncertainty about the boundary limits and a small type I error of rejecting a centroid when it indeed belonged to the experimental cancer attractor confidence region. The 68% region investigated a narrower region corresponding to high type I error.

About the concentration of datapoints

Before proceeding with the dynamics analysis, we highlight the rationale behind our hypothesis of taking the clusters' mean as centroids, that is, as representative points of each agglomerate. Besides the visual inspection in the 3-dimensional space, as already mentioned, we investigated multiple confidence regions (95%, 68%, and 20%) obtained from the BT_All clusters. We sampled data from uncorrelated multi-variate Gaussian distribution with parameters coming from the empirical data. For each cluster \mathcal{C}_i of the BT_All data, we computed the empirical mean vector μ_i and the empirical full covariance matrix C_i . We sampled from the distribution $\mathcal{N}(\mu_i, \text{diag}(C_i))$ a sample 10 times the number of points in the respective BT_All clusters, and obtained the correspondent Gaussian ellipsoids. First, we checked the proportion of Gaussian distributed points in the Gaussian confidence regions, that is, checking if Eq. (1) is satisfied for the determined values. We verified this by considering all genes and the reduced marker genes' dimensions. This statistical experiment ascertained that the confidence regions contained the expected proportion of uncorrelated data. Next, we made the same verification using de scRNA-seq data concerning the Gaussian ellipsoids to investigate the point concentration around the defined centroid. Finally, we obtained the proportions considering the scRNA-seq data within the confidence regions generated by the three marker genes dimension's full covariance matrix C_i . These steps ensured (i) the approximation for the centroids using mean and (ii) the compatibility between analyzing the 40 dimensions and the three marker genes

dimensions. In other words, the centroid using the mean value informed a densely populated region for the complete and the reduced dimension, strengthening our initial hypothesis for subsequent using the coordinates in the parameter estimation.

At this point, the reader may question why a centroid-based cluster analysis should not be used directly. The first reason is to have an automatic and visually unbiased definition of the number of clusters. The second and most important one is verifying the clusters' biological meaning concerning patients' gene signatures and their GBM subtypes, as will be shown in the "Results". Furthermore, this establishes the starting point for our dynamic analysis of verifying the high-density clusters as highly probable regions for finding cancer attractors.

GRN dynamics and implementation

To investigate in silico the presence of cancer attractors, we constructed a GRN dynamic model (Fig. 2 II-B). Due to the inherently stochasticity of biological systems⁵⁰, we modeled the dynamics using Langevin dynamics equation^{51,52}:

$$\frac{dx(t)}{dt} = F(x) + \xi(t), \quad (2)$$

where $x(t)$ is the gene expression level as a function of time (implicit dependence) relative to random variables of X , $F(x)$ is the deterministic term representing regulation due to network interactions, and $\xi(t)$ is the stochastic term accounting for the presence of intrinsic (intracellular contributions) and extrinsic noise (microenvironment contributions)^{53,54}.

We used the Hill function to model regulation interactions of the GRN⁵⁵, with the driving force F described by:

$$F_i = -k_i X_i + a_i \sum_{j \in \mathcal{A}_i} \frac{X_j^n}{S^n + X_j^n} + b_i \sum_{j \in \mathcal{I}_i} \frac{S^n}{S^n + X_j^n}, \quad (3)$$

where, for each gene i , represented by the component X_i , the index sets \mathcal{A}_i and \mathcal{I}_i represent the genes that interact with gene i through activation and inhibition, respectively. The value j represents the edge that bridges the regulation of transcription factors interacting with their target gene promoters. Note that in the case of self-activation or self-inhibition, one has $i \in \mathcal{A}_i$ or $i \in \mathcal{I}_i$, respectively. The parameter S denotes the value where the Hill function reaches its maximum inclination, n represents the intensity of the transition, a_i are the activation coefficients, b_i are the inhibition coefficients, and k_i are the self-degradation constants.

We modeled the regulations using the two-directional graphs (digraphs) outputs of the GRN processing step of Fig. 3, and rewritten Eq. (3) as:

$$\mathbf{F} = -\mathbf{kX} + \text{rowsum}(\mathbf{M}^a \odot \mathbf{V}^a) + \text{rowsum}(\mathbf{M}^b \odot \mathbf{V}^b), \quad (4)$$

with $\mathbf{k} = \text{diag}(k_1, \dots, k_N)$ a diagonal matrix, \mathbf{M}^a the activation matrix with entries $(\mathbf{M}^a)_{ij} = a_{ij}$, \mathbf{M}^b the inhibition matrix with entries $(\mathbf{M}^b)_{ij} = b_{ij}$, \mathbf{V}^a the activation Hill functions matrix with entries

$$(\mathbf{V}^a)_{ij} = \frac{X_j^n}{S^n + X_j^n}, \quad \text{with } j \in \mathcal{A}_i, \quad (5)$$

\mathbf{V}^b the inhibition Hill functions matrix with entries

$$(\mathbf{V}^b)_{ij} = \frac{S^n}{S^n + X_j^n}, \quad \text{with } j \in \mathcal{I}_i. \quad (6)$$

The \odot denotes the Hadamard product (element-wise matrix product), and $\text{rowsum}(\cdot)$ returns the vector with the row-wise sums of the matrix.

Fixed points and parameter estimation

After uncovering the clusters in the 3-dimensional gene markers space, we carried (lifted) the labels to the complete 40-dimensional space. We verified the symmetry of data clusters and proposed investigating the underlying dynamics by estimating the model parameters (Fig. 2 II-C) using the centroid coordinates as approximations for the fixed points coordinates. This assumption allowed us to consider the following:

$$F = \frac{dX}{dt} \approx 0, \quad (7)$$

which sought to be the first investigation of the presence of stability (cancer attractors).

This choice allowed us to estimate the parameters of equation (4) computing 2 parameters per equation (one for activation and one for inhibition). A possible biological interpretation was of an activation and inhibition intensity proportional to the target gene, for example, due to epigenetic regulations.

We assumed uniform and constant degradation coefficients for all mRNA molecules and used $k_i = k$ for all gene i . After that, we wrote equation (4) as follow:

$$k\mathbf{X} = \mathbf{V}\mathbf{c}, \quad (8)$$

with $\mathbf{V} = (\mathbf{V}^a \mid \mathbf{V}^b)$, $\mathbf{c} = (\mathbf{c}^a \mid \mathbf{c}^b)$, for $(\mathbf{c}^a)_i = a_i$ and $(\mathbf{c}^b)_i = b_i$.

As in our previous investigation³⁰, we proposed a parameter estimation including multiple centroids simultaneously. This choice aimed to capture the contributions of different equilibrium states and avoid overfitting individual clusters. Mathematically, for each centroid vector \mathbf{X}_α , we build the matrices \mathbf{V}_α and the vectors $\gamma_\alpha = k\mathbf{X}_\alpha$, and stack them as

$$\mathbf{M} = [\mathbf{V}_1 \mid \cdots \mid \mathbf{V}_{N_{clusters}}]^T, \quad (9)$$

$$\boldsymbol{\gamma} = [\gamma_1 \mid \cdots \mid \gamma_{N_{clusters}}]^T. \quad (10)$$

We estimated the parameters using a L_1 -norm robust regression, implemented as a linear programming problem⁵⁶ using the Simplex algorithm in the Mathematica environment⁴⁸. By doing so, we solved the following L_1 -norm minimization problem:

$$\min \|\mathbf{M}\mathbf{c} - \boldsymbol{\gamma}\|_1, \quad (11)$$

$$0.01 \leq \mathbf{c} \leq 10. \quad (12)$$

We computed the solutions by choosing $k = 1$ and defining a lower and upper limit for the parameter estimation. After a coarse search verification for different values, we defined n ranging from 1 to 4 in increments of 0.5, S from 0.5 to 4 in increments of 0.5, and the lower and upper limits for the linear programming algorithm as 0.01 and 10, respectively.

Each hyperparameter (n and S) combination was intended to characterize possible dynamic deviations related to malignant states and the corresponding regulation parameters (activation and inhibition) to represent distinct GRN rewiring. To test and quantify multistability in these regions, we used all clusters' combinations to estimate the parameters (Eqs. (9) and (10)).

Stochastic simulations: identifying attractors and multistability

We sought to investigate the dynamics stability achieved for each set of parameters estimated (Fig. 2 II-D). This step was the core of our investigation of clusters, pointing to regions with a higher probability of finding stable states (cancer attractors) and multistability across clusters. The confidence regions were our choice to instrumentalize the verifications (Fig. 4).

To quantify the presence of one or more stable states inside a confidence region, we wrote the dynamics as a system of stochastic differential equation (SDE):

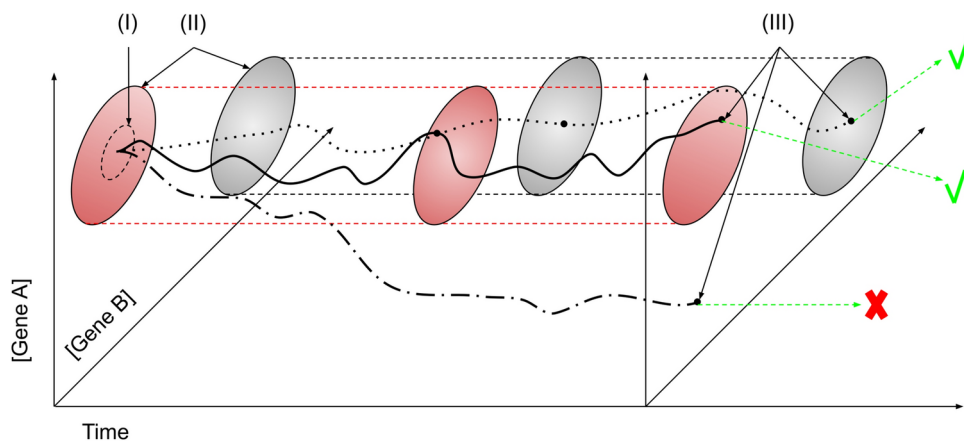


Figure 4. Schematic illustrating the trajectories within two markers dimension and using the confidence region to select parameters' configurations. (I) Display the constrained region for the perturbation of initial conditions. Without the perturbation, the initial conditions are the center of the ellipsoids. (II) Show the ellipsoids used as confidence regions to select the parameters. The red and gray ellipsoids define a tube through the time dimension allowing us to verify if each parameter configuration achieves trajectories within the desired constraints. (III) Depicts the final time steps. Parameters' configurations are saved if the final steps are within the red or gray regions (green 'V' mark), otherwise, they are not considered (red 'X' mark).

$$dX = \nu(X, t) dt + \sigma(X, t) dW, \quad (13)$$

with the drift $\nu(X, t)$ as the driven force $F(X)$ including the estimated parameters obtained from the multiple centroid combinations, the noise proportional to each state to avoid negative values for near zero gene expressions and computed as $\sigma(X, t) = \eta X$ (with η a proportionality constant), and a Wiener standard process dW .

We chose a low noise so that the trajectories would not be trapped in unstable states and tested the method considering different simulation times (t_{sim}). We decided to test 20, 100, 200, and 400 arbitrary units (a.u.) using time steps (Δt) of 0.1, 0.05 and 0.01. We observed that a simulation time of 200 (a.u.) using time steps of 0.05 was enough to obtain the equilibrium states, as increasing the time or reducing the steps gave the same results. To simplify the definition of stable states, we used the low noise choice and approximate:

$$\mathbf{X}_{sim}(200) \approx \bar{\mathbf{X}}_{sim} \approx \boldsymbol{\mu}_{sim}, \quad (14)$$

where the final step of time of 200 a.u. (after 4.000 simulation steps) is approximated as the centroid coordinates. We highlight that, due to the GRN constraints, $\boldsymbol{\mu}_{sim}$ is not necessarily the same as $\boldsymbol{\mu}_{ref}$ used in the parameter estimation and justify the definition of confidence regions. This approximation allowed employing equation (1) for each sampling to verify if the final equilibrium state lies in some of each cluster confidence region defined by the BT_All data.

We proposed to test the following null hypothesis H_0^0 : “There is no parameter configuration that leads to attractors in the confidence region” to verify the existence of an attractor. This hypothesis implies that the observed experimental data points are oscillations or random observations within the state space. Additionally, we proposed to test H_0^1 : “There is only a single attractor in the confidence region” for the existence of multiple cancer attractors inside the same region. This hypothesis implies that the experimental data distribution regarding each cluster contains only a single attractor. The first hypothesis could be rejected by showing that at least one of the parameters’ combinations could lead to an attractor inside one or more regions, and the second by demonstrating the existence of parameters leading to more than one attractor inside a cluster’s confidence region.

We tested the previous hypotheses by solving Eq. (13) numerically using the Euler-Maruyama and Stochastic Runge Kutta method, both with an Itô interpretation and fixing $\eta = 0.001$. As we obtained the same results, we proceeded with Euler-Maruyama, which was shown to be more time-efficient. Instead of exploring the 40-dimensional space searching for attractors, we leveraged the biological relevance of the scRNA-seq data clusters and chose the centroids as initial conditions. Additionally, we proposed testing the sensibility to the initial conditions by adding Gaussian noise and exploring a limited number of totally random initial conditions sampled across the space (Fig. 4).

In this way, we defined the initial conditions as:

$$\mathbf{X}_{sim}(0) = \mathbf{X}_\alpha + \beta \cdot \varepsilon^0, \quad (15)$$

$$\mathbf{X}_{sim}(0) = \varepsilon^1, \quad (16)$$

where $\mathbf{X}_\alpha = \{X_{\alpha,1}, X_{\alpha,2}, \dots, X_{\alpha,n}\}$ is the centroids coordinates of an α cluster, ε^0 a noise such that each $\varepsilon_1^0 \sim \mathcal{N}(0.5, 0.1)$ with β a proportionality constant so we could remove or amplify the perturbation, and ε^1 a noise such that each $\varepsilon_1^1 \sim U(0, 10)$. We used $\beta = 0$ to test H_0^0 , H_0^1 and investigate the presence of multistability. In sequence, $\beta = 1$ and ε^1 were applied to analyze the effect of perturbations around the centroids and explore the state space.

Results

Glioblastoma GRN

Starting with the data preparation (“[Data preparation and GRN construction](#)”), we assembled the datasets and constructed the GRN interactions table and the adjacency matrices used in the implementation of the stochastic dynamic (Fig. 2 I-A and II-A). Figure 5 displays the GRN presenting the interactions of our GBM dynamics model. The resulting structure comprised 40 vertices and 242 edges: 187 activations, 11 self-activations, 41 inhibitions, and three self-inhibitions. The complete list of interactions is available in the ‘GRN_info’ folder of the repository provided in the “Data availability” section.

Datasets, variables, and simulation configuration

The following table summarizes the information regarding the datasets, variables, and simulation configuration. Table 1 presents three blocks ‘GBM scRNA-seq Dataset and Description’, ‘Model Parameters’, and ‘Simulation Settings’. The ‘GBM scRNA-seq Dataset and Description’ block summarises the information of “[Data preparation and GRN construction](#)”, presenting a succinct description of the datasets analyzed in this investigation and the number of cells of each one. The ‘Model Parameters’ block summarises the information of “[Fixed points and parameter estimation](#)”, displaying the specified values for the parameter estimation. Specifically, the fixed k value, the tested Hill coefficients (n and S), and the estimated parameters a and b . The parameter ranges were based on our previous investigation³⁰. The block ‘Simulation Settings’ summarises the values of the variables corresponding to the stochastic dynamics simulation and the corresponding numerical configurations. The

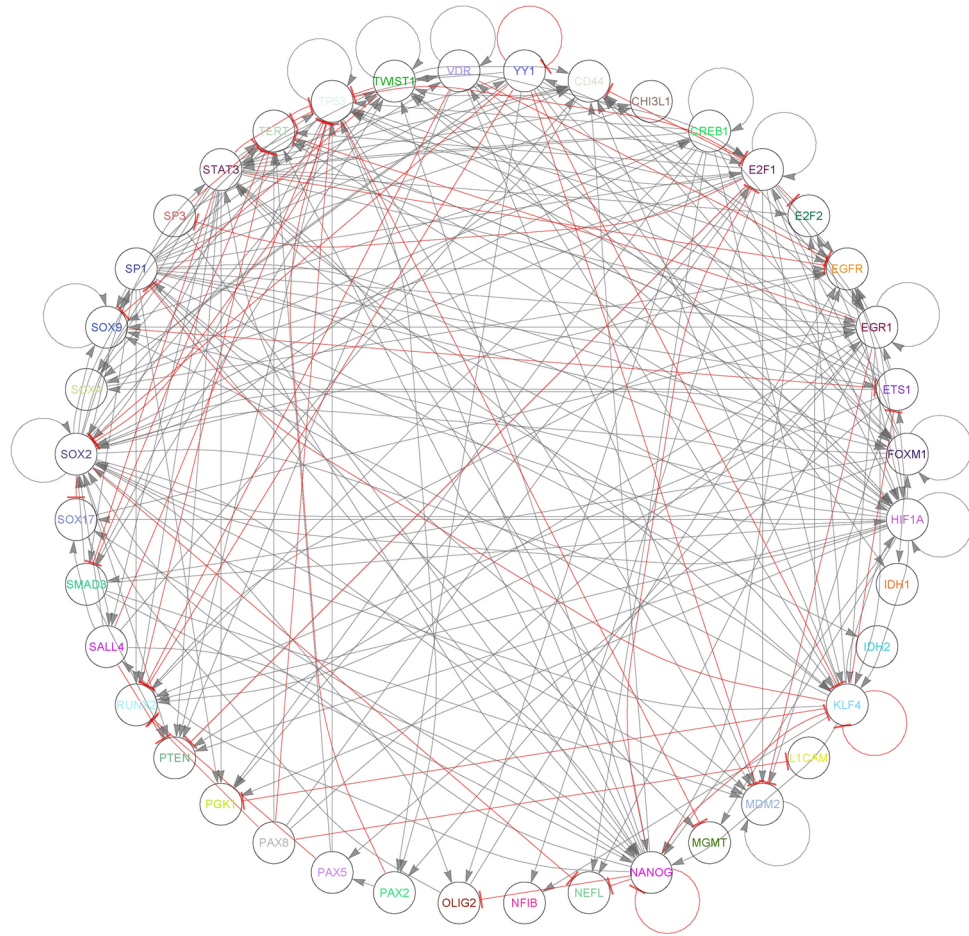


Figure 5. Gene regulatory network used in implementing the GBM dynamics model. Black lines with flat arrows represent activations, and red lines with arrowheads represent inhibitions. It contains 40 vertices and 242 edges, with 187 activations, 11 self-activations, 41 inhibitions, and 3 self-inhibitions. (With permission from ref.³⁰).

noise values were used to investigate the centroids' stability regarding the confidence region constraints. We display the chosen values concerning the simulation time, time step, and numerical method. The complete tested values list is described in “[Stochastic simulations: identifying attractors and multistability](#)”.

Clustering scRNA-seq data markers dimensions

We executed an initial analysis in the R environment that revealed the genes *EGFR*, *IDH1*, and *CD44* with apparently multimodal distributions. Figure S1 (Neoplastic dataset) and Fig. S2 (Regular dataset) show the pairwise scatter plots to investigate the gene correlations and inter-patient variability. It also shows the density histogram and boxplots of gene expression distribution for each patient dataset. We moved to the clustering phase (Fig. 2 I-B to I-D; “[Clustering scRNA-seq data: centroids and confidence region](#)”), confirming the clearer observation of data agglomerates in these markers' dimension when visually comparing to other combinations. The density-based clustering of the BT_All dataset obtained 7 clusters (labeled from A to G), with Table S1 showing the means and standard deviations of the corresponding markers genes. We additionally tested the clustering using Manhattan distance, corroborating the number of clusters. By grouping high/low expression levels in the CD44-EGFR dimension, we got four groups (A, B-C, D-E, and F-G). Concerning the IDH1 gene, cluster A presented low values, and the remaining groups alternated low and high. We computed the corresponding centroids and defined the confidence regions. We also clustered the remaining datasets to compare with the BT_All dataset. We obtained the datasets BT_S1, BT_S2, BT_S4, BT_S6, and BT_Regular presenting 5, 8, 9, 7, and 6 clusters, respectively. It is important to note that clustering individual patients with fewer data densities might lead to different classifications.

Evaluating the BT_All clusters

We visually inspected the data distribution on the three marker gene dimensions and confirmed data agglomerating around centroid coordinates. We proceed to the quantification of data to compare the concentration of data points within multiple confidence regions (95%, 68%, and 20%) defined for the 40 genes dimension and the three marker genes dimensions (section [About the concentration of datapoints](#)). The results are presented in

GBM scRNA-seq dataset and description		
Dataset	Description	Number of cells
BT_S1	Neoplastic cells from tumor core of patient BT_S1	265
BT_S2	Neoplastic cells from tumor core of patient BT_S2	502
BT_S4	Neoplastic cells from tumor core of patient BT_S4	134
BT_S6	Neoplastic cells from tumor core of patient BT_S6	126
BT_All	All patients neoplastic cells from tumor core	1027
BT_Regular	All patients regular cells from tumor core and periphery	2489
Model parameters		
Symbol	Description	Value/range
k	Self-degradation constant	$k = 1$
S	Hill function inflection point value	$0.5 \leq S \leq 4$, in steps of 0.5
n	Hill function transition intensity	$1 \leq n \leq 4$, in steps of 0.5
a_i	Activation coefficients for gene i	$0.01 \leq a_i \leq 10$
b_i	Inhibition coefficients for gene i	$0.01 \leq b_i \leq 10$
Simulation settings		
Symbol	Description	Value/range
η	Proportionality constant in multiplicative noise	$\eta = 0.001$
t_{sim}	Simulation time	$t_{sim} = 200$ a.u.
Δt	Time steps used in simulations	$\Delta t = 0.05$ a.u.
N_{method}	Numerical method used for solving SDEs	Euler-Maruyama
\mathbf{X}_α	Centroid vector for cluster α	$\mathbf{X}_\alpha = \{X_{\alpha,1}, X_{\alpha,2}, \dots, X_{\alpha,n}\}$
$\mathbf{X}_{sim}(0)$	Initial condition for simulations, using centroids and/or noise	$\mathbf{X}_{sim}(0) = \mathbf{X}_\alpha + \beta \cdot \varepsilon^0$ OR $\mathbf{X}_{sim}(0) = \varepsilon^1$
ε^0	Gaussian noise applied to initial conditions	$\varepsilon_i^0 \sim \mathcal{N}(0.5, 0.1)$
ε^1	Uniformly distributed random noise for initial conditions	$\varepsilon_i^1 \sim U(0, 10)$
β	Proportionality constant for perturbation in initial conditions	0 and 1

Table 1. Combined description of variables, symbols, and GBM scRNA-seq datasets used in the study.

Table 2. First, we checked the proportions of Gaussian data inside the confidence regions generated by its data clusters. We confirmed the expected proportions defined by the respective confidence values, disregarding sampling fluctuations of up to 3 percentage points.

Next, we evaluated the proportions of the BT_All data points concerning the ellipsoids defined by the Gaussian clusters. For the 95% confidence regions, we observed percentages below the expectation to a minimum of around 84%. For the 68% and 20% confidence regions, all values were over the expected independently of the degrees of freedom considered. Confirming our expectations, we observed an increasing percentage of data points for the 20% confidence region. The values increase to 3.5 times the expected percentage of 20% of the cluster size when compared to Gaussian distribution. This result confirmed the points agglomerating around the centroid, which might be evidence of an increasing probability of the presence of attractors.

The last verification was to evaluate the percentage of BT_All data points with BT_All clustered data and restrict the analysis to the three marker space dimensions. The confidence region for the complete genes dimensions is problematic due to the typical clusters' singular covariance matrices. Interestingly, our results show that the percentages were practically the same as for the Gaussian clusters confidence region. Only clusters B and G of the 20% confidence region showed a 7% difference. These results enabled our investigation to proceed using the BT_All clusters defined within the three marker genes dimension as a criterion for the parameter selection.

To investigate the biological meaning of the clusters concerning each patient, we quantified the proportion of points of each dataset (BT_S1, BT_S2, BT_S4, BT_S6, and BT_Regular) within the 68% and 95% confidence regions defined by the BT_All dataset. Figure S3 illustrates the case for the 95% confidence regions. We correlated the proportions within each confidence region to the results provided in the supplementary material of Darmanis et al.³³ by comparing the number of cells identified as Classical, Mesenchymal, Neural, and Proneural

Confidence (%)	df	Data	Ellipsoid	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)	G (%)
95	40	Gaussian	Gaussian	96.15	94.60	95.66	95.27	95.44	95.21	97.50
95	3	Gaussian	Gaussian	94.53	94.83	95.32	95.54	94.84	94.71	95.31
95	40	BT_All	Gaussian	88.89	86.21	87.32	88.04	84.81	87.40	96.88
95	3	BT_All	Gaussian	84.62	91.95	91.71	93.48	93.29	94.96	93.75
95	3	BT_All	BT_All	84.62	91.95	91.71	94.02	92.58	94.12	93.75
68	40	Gaussian	Gaussian	67.44	67.93	67.85	68.21	67.67	68.66	66.25
68	3	Gaussian	Gaussian	66.84	66.44	68.44	67.34	67.99	68.07	68.12
68	40	BT_All	Gaussian	85.47	80.46	77.56	76.09	74.91	78.15	81.25
68	3	BT_All	Gaussian	77.78	81.61	84.39	67.93	75.62	77.31	71.88
68	3	BT_All	BT_All	77.78	78.16	83.41	68.48	75.97	75.63	71.88
20	40	Gaussian	Gaussian	18.80	20.69	19.51	20.54	19.65	20.50	16.88
20	3	Gaussian	Gaussian	22.31	20.11	21.41	19.84	20.46	19.83	20.31
20	40	BT_All	Gaussian	69.23	70.11	62.44	57.61	61.13	64.71	50.00
20	3	BT_All	Gaussian	70.09	41.38	55.12	23.91	35.69	42.86	25.00
20	3	BT_All	BT_All	70.09	34.48	55.12	23.37	34.98	40.34	18.75

Table 2. Percentage of points within confidence regions when comparing BT_All with uncorrelated Gaussian samples.

Confidence (%)	Dataset	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)	G (%)	\emptyset (%)
95	BT_All	9.64	7.89	18.79	17.04	26.48	11.30	2.92	5.94
95	BT_S1	23.40	20.00	48.68	0.75	0.38	1.51	0.00	5.28
95	BT_S2	2.59	1.59	3.98	29.08	37.85	14.94	4.18	5.78
95	BT_S4	12.69	4.48	4.48	13.43	32.09	20.15	3.73	8.96
95	BT_S6	5.56	11.11	30.16	7.14	30.16	7.94	3.17	4.76
95	BT_Regular	43.47	2.45	3.58	0.60	0.28	22.38	10.16	17.08
68	BT_All	8.86	6.62	16.65	12.27	20.93	8.76	2.24	23.66
68	BT_S1	21.89	17.36	44.15	0.00	0.00	1.51	0.00	15.09
68	BT_S2	2.39	1.20	2.99	21.71	30.68	11.75	3.19	26.10
68	BT_S4	11.94	4.48	2.99	6.72	23.13	13.43	2.99	34.33
68	BT_S6	3.97	7.94	27.78	6.35	23.81	7.14	2.38	20.63
68	BT_Regular	39.53	1.81	3.05	0.08	0.24	17.24	6.27	31.78

Table 3. Percentage of points inside the ellipsoids defined by BT_All data clusters normalized by the total number of points of each dataset.

with the four groups of the CD44-EGFR axis. Table S2 synthesizes the supplementary material of Darmanis et al., displaying the percentage of cells of each GBM subtype concerning each patient.

We observed distinct signatures for each patient (Tables S3 and 3, where the \emptyset symbol represents the number of data points located outside the defined regions). Additionally, by correlating the order of the number of cells in these markers' dimensions confidence regions, we observed that the Classical and Mesenchymal subtypes seem to be divided into smaller groups. The Classical subtype appears to correlate with B-C and D-E clusters. All these clusters present high EGFR, with B-C presenting low CD44 and D-E high CD44. The clusters F-G seem to correlate with the number of cells of patients BT_S2 and BT_S4, classified as presenting Mesenchymal subtype by Darmanis et al.³³. By this comparison, the Classical subtype presents an expression of the CD44 stemness marker. For patients BT_S1, BT_S2, and BT_S4 the Mesenchymal subtype only presented low expression of EGFR. For patient BT_S6, the Mesenchymal subtype could also include clusters D-E. The Proneural subtype might be distributed within these clusters, requiring deeper investigations such as analyzing additional markers. We highlight that some works suggest the Neural subtype as non-tumor-specific and point to different directions regarding the subtypes' characterization^{6,35,36,57}.

Finally, we compared the BT_Regular data points and BT_All confidence regions to ascertain the regions less likely related to BT_Regular data (more likely BT_All related). The results revealed differences in confidence region occupation within each dataset, with cluster A containing more BT_Regular cells (Fig. 6a, b). To check the existence of different cancer attractors and multistability, we proceeded with the in silico simulations.

Ascertaining cancer attractors and multistability

We specified the GRN dynamic model (Fig. 2 II-B—“GRN dynamics and implementation”), proceeded with the parameters' estimation (Fig. 2 I-C and II-C—“Fixed points and parameter estimation”), and computed the

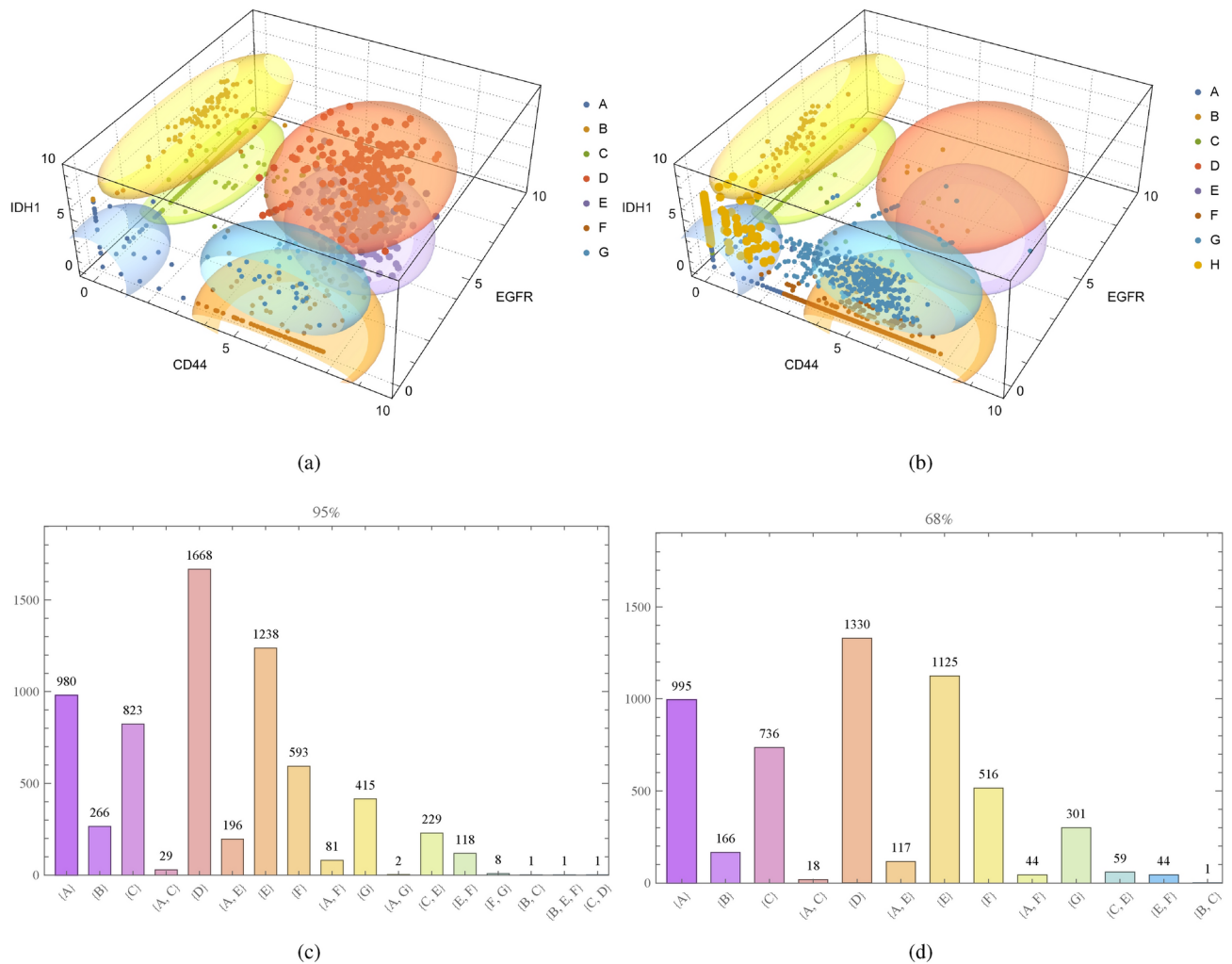


Figure 6. Ellipsoids representing the 95% confidence regions for each cluster of BT_All data and the total number of parameters' configurations leading to stability within each confidence region. **(a)** BT_All data and **(b)** BT-Regular data. The letters and colors represent each cluster. **(a)** The number of parameters achieved for the 95% confidence region. **(b)** The number of parameters achieved for the 65% confidence regions.

stochastic simulations (Fig. 2 I-D and II-D—“Stochastic simulations: identifying attractors and multistability”). We attempted to get stability across different clusters' confidence regions by constructing a list of all 127 cluster combinations to use in Eqs. (9) and (10). We ran the parameter estimation depending on Eq. (14) to find the activations and inhibitions. For each combination, we generated one trajectory departing from each of the seven cluster centroids using the 56 Hill function parameters combinations (n and S). Figure 6c, d summarises the outcomes using two values for the confidence region (95% and 68%) in the parameters selection. The x-axis shows the achieved stability, and the y-axis indicates the number of parameters leading to each one, considering all of the $127 \times 56 \times 7$ trials. The results show the clusters with the most parameters leading to one stable state and a few displaying multistability. For instance, it revealed a predisposition for multistability, including clusters A, C, E, and F. Additionally, we observed tristability only for the 95% confidence region (clusters B, E, and F). As discussed later, the x-axis represents the achieved stability, not the combinations used in the parameter estimation. All parameters and clusters relations are available in the 'outputs_xlsx' folder in code repository (see “Data availability”).

The results show the presence of multiple parameters' configurations leading to stable states inside all clusters' confidence regions, enabling the rejection of H_0^0 . Additionally, we achieved multistability for various clusters' combinations. To investigate if the attractors inside each region are the same, we quantified what parameters led to each multistability within each confidence region. Figure S4 illustrates the results, with the titles displaying the achieved stable states, the y-axis showing the Hill function parameter combination number (from the total of 56 combinations), and the x-axis showing the parameter frequency. The results show that the 68% confidence regions mainly presented fewer parameter combinations than the 95% region. However, the reduction was not necessarily proportional to the decreasing volume. For instance, parameter 53 of Fig. S4a was reduced to zero counts, S4b did not change, and parameter 53 of Fig. S4d was only reduced from 27 to 25 cases. The parameters

absent in the 68% regions must be stable states within the boundaries of the 68% and 95% regions, demonstrating the existence of parameters combination leading to different attractors inside the region and rejecting H_0^1 .

Next, we investigated what clusters were used in Eqs. (9) and (10) to reach each stability from Fig. S4. Each plot title of Fig. S5 displays the achieved stable states, the y-axis shows the clusters' combination used in the parameter estimation, and the x-axis displays the number of parameters for each case. These results highlighted that our method explored the multistability according to the constraints of our GRN's model, not arbitrarily achieving any desired multistability.

In the final verification, we investigated the sensibility to initial conditions. We sampled five initial conditions for each one of the seven centroids using $\beta = 1$ (Eq. (15)) and five from ε^1 (Eq. (16)). We limited this investigation to the 95% confidence regions and observed the same results of Fig. 6c. This result showed the robustness of the found stable states and pointed out that sampling from unperturbed centroids was a method to identify stable states, avoiding sampling through the entire 40-dimensional space. All results can be reproduced with the code present in the repository (see "Data availability").

Discussion

Typical scRNA-seq downstream data analysis uses machine learning algorithms to reduce the dimensionality and perform clustering analysis to identify cell types or subtypes^{18–20}. This approach allows the integration of numerous biological information within the reduced dimensions to aggregate into the clustering. However, the snapshot nature of the data neglects the underlying dynamics leading to and characterizing each cell type or subtype. To advance this understanding, we departed from a curated and annotated GBM dataset from the study of Darmanis et al.³³ and proposed a biological-informed clustering to investigate the presence of cancer attractors dynamics¹⁵. Our choice of reducing the analysis to marker genes dimension is justified due to their use in specifying the state of a system. To this end, they must show constrained expression levels instead of oscillating from low to high levels. The latter behavior would make them useless, as it would be a transitory classification since the expression levels would vary substantially for each snapshot the data is captured. Concentrating on marker gene dimensions also allowed us to enhance subsequent biological interpretation of the clusters. This choice is supported by previous investigations demonstrating the potential of using a small number of biomarkers to describe complex systems⁵⁸. We proposed that the constrained regions within the marker genes dimension space would be the clusters, as highly probable regions of finding stable states. Additionally, we suggested that the clusters could contain multiple stable states or even represent multiple interchangeable stable states. This investigation was divided into two significant steps: exploring the clusters and an in-silico simulation to search for stable states.

For the initial step of cluster exploration, our initial goal was to define a cluster representative point. This point should exhibit the properties expected by the presence of attractors, that is, an increasing density around it. To execute this verification, we selected a density-based clustering algorithm aligned with our search for cancer attractors. We used an algorithm with automated identification of the number of clusters, which ensured that our analysis remained independent of visualization biases⁴⁹. By evaluating the gene expression of 3 GBM marker genes (*EGFR*, *IDH1*, and *CD44*) in 4 patients, we found seven possible cellular clusters (Table S1). Concerning the proportion and spreading of points for each patient dataset alone, we observed that the low number of data points led to erratic clustering results, highlighting the relevance of defining the clusters using multiple patient data. Next, we considered each cluster's average as representative points, from that moment on called centroids. We described the confidence regions using the centroids' coordinates and each cluster covariance matrix³², confirming our defined centroids representing increasingly dense regions when investigating the concentration of points within smaller confidence regions and comparing them with the expected concentration of uncorrelated Gaussian distributions. The results presented in Table 2 show that we could get information about the density across the 40-dimensional transcription factor space by clustering in the marker space, validating our centroids definition. Besides, the specified regions presented a powerful way to investigate the datasets and simulations. We analyzed individual patient data within each confidence region to Darmanis et al.³³ subtypes classifications, observing distinct signatures for each patient. Besides, we observed that in the EGFR-CD44 dimension, the Classical and Mesenchymal subtypes split into smaller groups. Considering the role of these markers in GBM aggressiveness, this subdivision might reveal essential features related to GBM dynamics. We compared the neoplastic dataset to the regular one and observed regions more likely to be associated with malignant states. Finally, we employed the confidence regions to select parameters in the in-silico simulations.

The final step was to investigate if the density of points could imply a significant concentration of stable states for the simulations. Positive results would strengthen the hypothesis, associating higher density with the probability of the presence of cancer attractors. Our strategy for this verification was to use the centroids as a first approximation of stable states. Upon this first approximation, we applied a GBM GRN used in our previous investigation³⁰. The GRN was expanded using the MetaCore platform⁴⁰, ensuring the objective of increasing the network connectivity. Next, we used Hill functions dynamics, enabling our investigation to extend previous contributions^{52,55,59,60}. Concerning the parameter estimation, we considered one parameter for activation and one for inhibition per gene. We stacked multiple combinations of clusters during the estimation to explore the presence of multistability, ensuring a data-driven parameter estimation. By achieving the parameter estimation, we addressed the limitations of previous investigations using arbitrary parameters and dealt with the dependence on time series data⁶¹. We implemented the stochastic dynamics with enough noise to repel unstable states and investigated the time necessary to reach stable states. After that, we used the previously investigated confidence regions to filter the parameters that led to stable states inside them. This framework successfully found numerous parameters presenting stable states and multistability. We show the dependency of stability with GRN constraints and different stable states' likelihood. These results strengthened our hypothesis that the density near the clusters' centroids indicates higher probability regions of finding stable states.

Our findings are aligned with the ecological perspective of cancer. Investigations of alternative stable states have been a pivotal question in ecology. For instance,²⁶ has shown how alternative stable states might coexist under the same parameters, representing interchangeable states, or appear and disappear due to parameter changes. Depending on the nature of alternative states, the dimension of a basin of attraction could even be related to the observed rate of changes^{27,28}. Recently, some results have shown microbiome shifts between alternative stable states of the dynamics around complex attractors²⁹. Other authors have investigated the presence of multistability in complex ecological communities⁶². These findings align with the distinct cell populations coexisting in the tumors^{63–66}. In our results, the multiple stable states could be interpreted as alternative stable states resulting from the dynamics of a complex GBM GRN. As in ecological studies, the gene expression states are also coupled to the environment, known as the tumor microenvironment⁶⁷. However, the tumor cells in the microenvironment are usually heterogeneous in their mutations and epigenetic regulation⁶⁷. Our investigation suggested that mutations or epigenetic regulations might characterize various parameter perturbations with a low probability of returning to previous configurations. This low probability for reversibility might characterize the malignant state of genome attractors resulting from distinct subpopulations⁶⁸. In this way, our results could represent stable states that are not interchangeable but represent different molecular phenotypes coexisting in the same region of markers' gene expression space.

Biologically, genetic mutations and epigenetic changes affect the parameter values and consequently the cellular fates¹⁵. Assuming that the environment correlates with the values of parameters, knowing the more conceivable parameters would indicate the cellular states more likely to emerge. Additionally, selecting parameters presenting multistability implies selecting more than one stable state, which could represent subtypes likely to coexist, as observed in *IDH*-wild-type GBMs³⁶. All these features together might be underlying the observed plasticity of the malignant state, as an entire cluster would be the outcome of multiple attractors and parameter combinations. A deeper understanding of each cluster's characteristics and the parameters leading to them could greatly assist our understanding of tumor heterogeneity and drug resistance mechanisms. For instance, these alternative trajectories could represent different biological circumstances, such as patient reactions to therapies, the tumor's various levels of hypoxia and nutrient access^{69,70}, genetic and epigenetic alterations⁷¹, and the immune system response⁷². All these characteristics impact the tumor heterogeneity and the disease outcome⁷³. The success in finding parameters leading to multistability indicated that the proposed methodology is robust and adequate for complex GRNs. Also, it might present a scalable and straightforward alternative to previous proposals^{74,75}.

Despite our simplified model, we propose that further advances seeking to correlate the parameters with biological observation could help quantify malignant states. With biologically meaningful parameters, the analysis presented in Figs. S4 and S5 would describe the conditions and probabilities of observing each cluster and the changes needed to obtain desired outcomes. In this way, our method is a basis for an algorithm to define therapeutic targets for individual patients and other types of cancer.

Conclusion

Single-cell data still presents multiple challenges to overcome⁷⁶. With the increasing availability, many cluster algorithms to explore single-cell cancer datasets have been developed⁷⁷. However, incorporating dynamic information is a typically disregarded aspect. In previous work, we have extensively explored different dynamic models and multistability³⁰. The present investigation delved into a selected model, proposing a data-driven stable state quantification. While the studied parameters still do not represent specific biological processes, they characterize the system behavior and illustrate trends observed in experiments.

We proposed a framework for a biomarker-guided uncovering of potential cancer attractors given scRNA-seq data. The pipeline executed biomarker-oriented clustering and ellipsoidal statistics to identify high-density regions indicative of cancer attractors. The clusters' centroids were used as a first stability approximation, leading to the parameters' estimation using linear programming. Further, exploring GRN stochastic dynamics allowed the verification of cancer attractor candidates. The results revealed the biomarkers' potential to identify cancer attractors and the corresponding probable regions. Also, it disclosed candidates for multistability, exposing states likely to transit to each other, which presents a high potential for cancer recurrence in case any cells remain within those regions after treatment.

This methodology may complement the investigation of biomarkers and their potential to define cancer attractors, giving essential insights concerning the underlying dynamics driving cancer progression and therapy. For example, in identifying attractors and stability within confidence regions, we can advance in investigating the genes implicated in cancer attractors, paving the way to propose inhibitions leading to destabilizing the attractors within the framework of personalized oncology.

Data availability

The code and data analyzed/generated to produce the results of the current study are available in the [Biomarker-Guided-scRNA-Seq-Cancer-Attractor-Analysis](#) repository.

Received: 17 May 2024; Accepted: 30 September 2024

Published online: 29 October 2024

References

- Gallego, O. Nonsurgical treatment of recurrent glioblastoma. *Curr. Oncol.* **22**, 273–281. <https://doi.org/10.3747/co.22.2436> (2015).
- Duhamel, M. *et al.* Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. *Nat. Commun.* **13**. <https://doi.org/10.1038/s41467-022-34208-6> (2022).

3. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401. <https://doi.org/10.1126/science.1254257> (2014).
4. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196. <https://doi.org/10.1126/science.aad0501> (2016).
5. Kim, C. et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* **173**, 879–893. e13. <https://doi.org/10.1016/j.cell.2018.03.041> (2018).
6. Nefel, C. et al. 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> (2019).
7. Marusyk, A. et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58. <https://doi.org/10.1038/nature13556> (2014).
8. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell* **168**, 613–628. <https://doi.org/10.1016/j.cell.2017.01.018> (2017).
9. Esteller, M. Epigenetics in cancer. *N. Engl. J. Med.* **358**, 1148–1159. <https://doi.org/10.1056/nejmra072067> (2008).
10. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558. <https://doi.org/10.1126/science.1235122> (2013).
11. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38–55. <https://doi.org/10.1016/j.cell.2013.03.008> (2013).
12. Huang, S. Systems biology of stem cells: Three useful perspectives to help overcome the paradigm of linear pathways. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2247–2259. <https://doi.org/10.1098/rstb.2011.0008> (2011).
13. Moris, N., Pina, C. & Arias, A. M. Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* **17**, 693–703. <https://doi.org/10.1038/nrg.2016.98> (2016).
14. Strauss, B., Bertolaso, M., Ernberg, I. & Bissell, M. Rethinking cancer: A new paradigm for the postgenomics era. In *Vienna Series in Theoretical Biology* (MIT Press, 2021).
15. Huang, S., Ernberg, I. & Kauffman, S. 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> (2009).
16. Li, Q. et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. *Proc. Natl. Acad. Sci.* **113**, 2672–2677. <https://doi.org/10.1073/pnas.1519210113> (2016).
17. Covert, M. W., Famili, I. & Palsson, B. O. Identifying constraints that govern cell behavior: A key to converting conceptual to computational models in biology?. *Biotechnol. Bioeng.* **84**, 763–772. <https://doi.org/10.1002/bit.10849> (2003).
18. Peyvandipour, A., Shafi, A., Saberian, N. & Draghici, S. Identification of cell types from single cell data using stable clustering. *Sci. Rep.* **10**. <https://doi.org/10.1038/s41598-020-66848-3> (2020).
19. Miao, Z. et al. Putative cell type discovery from single-cell gene expression data. *Nat. Methods* **17**, 621–628. <https://doi.org/10.1038/s41592-020-0825-9> (2020).
20. Zhang, S., Li, X., Lin, J., Lin, Q. & Wong, K.-C. Review of single-cell RNA-seq data clustering for cell-type identification and characterization. *RNA* **29**, 517–530. <https://doi.org/10.1261/rna.078965.121> (2023).
21. Uthamacumaran, A. A review of complex systems approaches to cancer networks. *Complex Syst.* **29**, 779–835. <https://doi.org/10.25088/complexsystems.29.4.779> (2020).
22. Uthamacumaran, A. A review of dynamical systems approaches for the detection of chaotic attractors in cancer networks. *Patterns* **2**, 100226. <https://doi.org/10.1016/j.patter.2021.100226> (2021).
23. Álvarez-Arenas, A., Podolski-Renic, A., Belmonte-Beitia, J., Pesic, M. & Calvo, G. F. Interplay of Darwinian selection, Lamarckian induction and microvesicle transfer on drug resistance in cancer. *Sci. Rep.* **9**. <https://doi.org/10.1038/s41598-019-45863-z> (2019).
24. Pienta, K. J., Hammarlund, E. U., Axelrod, R., Amend, S. R. & Brown, J. S. Convergent evolution, evolving evolvability, and the origins of lethal cancer. *Mol. Cancer Res.* **18**, 801–810. <https://doi.org/10.1158/1541-7786.mcr-19-1158> (2020).
25. Scarborough, J. A., Eschrich, S. A., Torres-Roca, J., Dhawan, A. & Scott, J. G. Exploiting convergent phenotypes to derive a pan-cancer cisplatin response gene expression signature. *npj Precis. Oncol.* **7**. <https://doi.org/10.1038/s41698-023-00375-y> (2023).
26. Beisner, B., Haydon, D. & Cuddington, K. Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382. [https://doi.org/10.1890/1540-9295\(2003\)001\[0376:assie\]2.0.co;2](https://doi.org/10.1890/1540-9295(2003)001[0376:assie]2.0.co;2) (2003).
27. Petraitis, P. S. & Dudgeon, S. R. Detection of alternative stable states in marine communities. *J. Exp. Mar. Biol. Ecol.* **300**, 343–371. <https://doi.org/10.1016/j.jembe.2003.12.026> (2004).
28. Petraitis, P. & Hoffman, C. Multiple stable states and relationship between thresholds in processes and states. *Mar. Ecol. Prog. Ser.* **413**, 189–200. <https://doi.org/10.3354/meps08691> (2010).
29. Fujita, H. et al. Alternative stable states, nonlinear behavior, and predictability of microbiome dynamics. *Microbiome* **11**. <https://doi.org/10.1186/s40168-023-01474-5> (2023).
30. Junior, M. G. V., Cortés, A. M. d. A., Carneiro, F. R. G., Carels, N. & Silva, F. A. B. d. Unveiling the dynamics behind glioblastoma multiforme single-cell data heterogeneity. *Int. J. Mol. Sci.* **25**. <https://doi.org/10.3390/ijms25094894> (2024).
31. Ding, Y., Gao, J. & Magdon-Ismael, M. Efficient parameter inference in networked dynamical systems via steady states: A surrogate objective function approach integrating mean-field and nonlinear least squares. *Phys. Rev. E* **109**, 034301. <https://doi.org/10.1103/physreve.109.034301> (2024).
32. Friendly, M., Monette, G. & Fox, J. Elliptical insights: Understanding statistical methods through elliptical geometry. *Stat. Sci.* **28**. <https://doi.org/10.1214/12-sts402> (2013).
33. Darmanis, S. et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* **21**, 1399–1410. <https://doi.org/10.1016/j.celrep.2017.10.030> (2017).
34. Verhaak, R. G. et al. 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> (2010).
35. Sidaway, P. Glioblastoma subtypes revisited. *Nat. Rev. Clin. Oncol.* **14**, 587–587. <https://doi.org/10.1038/nrclinonc.2017.122> (2017).
36. Fine, H. A. Malignant gliomas: Simplifying the complexity. *Cancer Discov.* **9**, 1650–1652. <https://doi.org/10.1158/2159-8290.cd-19-1081> (2019).
37. Mooney, K. L. et al. The role of cd44 in glioblastoma multiforme. *J. Clin. Neurosci.* **34**, 1–5. <https://doi.org/10.1016/j.jocn.2016.05.012> (2016).
38. Wang, W. et al. Internalized cd44s splice isoform attenuates egfr degradation by targeting rab7a. *Proc. Natl. Acad. Sci.* **114**, 8366–8371. <https://doi.org/10.1073/pnas.1701289114> (2017).
39. Calvert, A. E. et al. Cancer-associated idh1 promotes growth and resistance to targeted therapies in the absence of mutation. *Cell Rep.* **19**, 1858–1873. <https://doi.org/10.1016/j.celrep.2017.05.014> (2017).
40. Clarivate Analytics. MetaCore, 2019. Available online: <https://portal.genego.com>. (accessed on 16 April 2022).
41. R Development Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: <https://www.R-project.org/>. (accessed on 16 April 2022).
42. Vieira, M. Gene Expression Network Analysis, 2023; GitHub Repository. Available online: <https://github.com/marcosgvjunior/gene-expression-network-analysis>. (accessed on 16 April 2022).
43. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502. <https://doi.org/10.1038/nbt.3192> (2015).
44. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**. <https://doi.org/10.1186/s13059-019-1874-1> (2019).

45. Lab, S. Using Sctransform in Seurat, 2022. GitHub Repository. Available online: https://satijalab.org/seurat/articles/sctransform_vignette.html. (accessed on 17 July 2022).
46. Witkiewicz, A. K., Kumarasamy, V., Sanidas, I. & Knudsen, E. S. Cancer cell cycle dystopia: Heterogeneity, plasticity, and therapy. *Trends Cancer* **8**, 711–725. <https://doi.org/10.1016/j.trecan.2022.04.006> (2022).
47. Vieira, M. Graph Matrix and Combinatorics, 2023; GitHub Repository. <https://github.com/marcosgvjunior/graph-matrix-andcombinatorics>. (accessed on 17 July 2022).
48. Wolfram Research, Inc. Mathematica, Version 13.1; Mathematica: Champaign, IL, USA, 2022.
49. Wolfram Research, Inc. Neighborhood Contraction, 2023. Available online: <https://reference.wolfram.com/language/ref/method/NeighborhoodContraction.html>. (accessed on 9 April 2023).
50. Meister, A., Du, C., Li, Y. H. & Wong, W. H. Modeling stochastic noise in gene regulatory systems. *Quant. Biol.* **2**, 1–29. <https://doi.org/10.1007/s40484-014-0025-7> (2014).
51. Gillespie, D. T. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306. <https://doi.org/10.1063/1.481811> (2000).
52. Li, C. & Wang, J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLoS Comput. Biol.* **9**, e1003165. <https://doi.org/10.1371/journal.pcbi.1003165> (2013).
53. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186. <https://doi.org/10.1126/science.1070919> (2002).
54. Volfson, D. et al. Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864. <https://doi.org/10.1038/nature04281> (2005).
55. Santillán, M. On the use of the hill functions in mathematical models of gene regulatory networks. *Math. Model. Nat. Phenomena* **3**, 85–97. <https://doi.org/10.1051/mmnp:2008056> (2008).
56. Wolfram Research, Inc. Constrained Optimization, 2023. Available online: <https://library.wolfram.com/infocenter/Books/8506/ConstrainedOptimization.pdf>. Accessed 12th July 2022.
57. Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56.e6. <https://doi.org/10.1016/j.ccell.2017.06.003> (2017).
58. Cohen, A. A. Complex systems dynamics in aging: New evidence, continuing questions. *Biogerontology* **17**, 205–220. <https://doi.org/10.1007/s10522-015-9584-x> (2015).
59. Wang, J., Xu, L., Wang, E. & Huang, S. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.* **99**, 29–39. <https://doi.org/10.1016/j.bpj.2010.03.058> (2010).
60. Ferrell, J. E. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* **22**, R458–R466. <https://doi.org/10.1016/j.cub.2012.03.045> (2012).
61. Liu, Y.-Y. & Barabási, A.-L. Control principles of complex systems. *Rev. Mod. Phys.* **88**, 035006. <https://doi.org/10.1103/revmodphys.88.035006> (2016).
62. Aguadé-Gorgorió, G., Arnoldi, J.-F., Barbier, M. & Kéfi, S. A taxonomy of multiple stable states in complex ecological communities. *Ecol. Lett.* **27**, e14413. <https://doi.org/10.1111/ele.14413> (2024). E14413 ELE-01065-2023.R2. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.14413>.
63. Fassoni, A. C. & Yang, H. M. An ecological resilience perspective on cancer: Insights from a toy model. *Ecol. Complex.* **30**, 34–46. <https://doi.org/10.1016/j.ecocom.2016.10.003> (2017) (**Dynamical Systems In Biomathematics**).
64. Kemwoue, F. F. et al. Bifurcation, multistability in the dynamics of tumor growth and electronic simulations by the use of pspice. *Chaos Solitons Fractals* **134**, 109689. <https://doi.org/10.1016/j.chaos.2020.109689> (2020).
65. Lauko, A., Lo, A., Ahluwalia, M. S. & Lathia, J. D. Cancer cell heterogeneity & plasticity in glioblastoma and brain tumors. *Semin. Cancer Biol.* **82**, 162–175. <https://doi.org/10.1016/j.semcancer.2021.02.014> (2022) (**Cancer Cell Heterogeneity and Plasticity: From Molecular Understanding to Therapeutic Targeting**).
66. Januškevičienė, I. & Petrikaitė, V. Heterogeneity of breast cancer: The importance of interaction between different tumor cell populations. *Life Sci.* **239**, 117009 <https://doi.org/10.1016/j.lfs.2019.117009> (2019).
67. Hanahan, D. Hallmarks of cancer: New dimensions. *Cancer Discov.* **12**, 31–46. <https://doi.org/10.1158/2159-8290.cd-21-1059> (2022).
68. Kasperski, A. & Kasperska, R. Study on attractors during organism evolution. *Sci. Rep.* **11**. <https://doi.org/10.1038/s41598-021-89001-0> (2021).
69. Chen, Z., Han, F., Du, Y., Shi, H. & Zhou, W. Hypoxic microenvironment in cancer: Molecular mechanisms and therapeutic interventions. *Signal Transduct. Target. Ther.* **8**[SPACE] <https://doi.org/10.1038/s41392-023-01332-8> (2023).
70. Sullivan, M. R. & Vander Heiden, M. G. Determinants of nutrient limitation in cancer. *Crit. Rev. Biochem. Mol. Biol.* **54**, 193–207 <https://doi.org/10.1080/10409238.2019.1611733> (2019).
71. Bell, C. C. & Gilan, O. Principles and mechanisms of non-genetic resistance in cancer. *Br. J. Cancer* **122**, 465–472. <https://doi.org/10.1038/s41416-019-0648-6> (2019).
72. Gonzalez, H., Hagerling, C. & Werb, Z. Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev.* **32**, 1267–1284. <https://doi.org/10.1101/gad.314617.118> (2018).
73. Zhu, L. et al. A narrative review of tumor heterogeneity and challenges to tumor drug therapy. *Ann. Transl. Med.* **9**, 1351–1351. <https://doi.org/10.21037/atm-21-1948> (2021).
74. Angeli, D., Ferrell, J. E. & Sontag, E. D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci.* **101**, 1822–1827. <https://doi.org/10.1073/pnas.0308265100> (2004).
75. Wu, S., Zhou, T. & Tian, T. A robust method for designing multistable systems by embedding bistable subsystems. *npj Syst. Biol. Appl.* **8**. <https://doi.org/10.1038/s41540-022-00220-1> (2022).
76. Lähnemann, D. et al. Eleven grand challenges in single-cell data science. *Genome Biol.* **21**. <https://doi.org/10.1186/s13059-020-1926-6> (2020).
77. Mahalanabis, A. et al. Evaluation of single-cell RNA-seq clustering algorithms on cancer tumor datasets. *Comput. Struct. Biotechnol. J.* **20**, 6375–6387. <https://doi.org/10.1016/j.csbj.2022.10.029> (2022).

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brazil (CAPES) through the Social Demand Program (Programa de Demanda Social, DS) under File Number 88887.597339/2021-00-Finance Code 001. We would also like to mention INOVA Fiocruz program for their support of this research

Author contributions

MGVJ designed the analysis, developed the codes, analyzed/interpreted the data, and wrote the manuscript. AMAC revised the mathematical model and its implementation. NC and FRGC ensured biological accuracy. FABS provided structural critiques and improvements. All authors participated in text revision and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-74985-2>.

Correspondence and requests for materials should be addressed to M.G.V.J. or F.A.B.d.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024