ETH zürich

Trajectory inference in glioblastoma

Lab rotation report

Wiona Sophie Glänzer

09.01.-17.02.2023

Advisors:
Valentina Boeva
Florian Barkmann

D-INFK
ETH Zürich

Abstract

Glioblastoma is a hard-to-treat malignant brain tumor characterized by high cellular diversity. Previous research has identified expression meta-modules that define cellular states within Glioblastoma and cells that appear to be in transition between these states. In this project, we tested whether trajectory inference methods, developed to analyze cell differentiation processes, could be utilized to identify trajectories connecting these cell states. While some reservations remain about the validity of this approach, the tested methods successfully identified the expected trajectories and revealed gene expression changes along them. These findings suggest that trajectory inference techniques might offer valuable insights into Glioblastoma's cellular heterogeneity.

Contents

1 Introduction

Single-cell RNA sequencing allows us to measure gene expression in individual cells and thus investigate the diversity in cell types and states. Since the publication of the first scRNA-seq method in 2009 [1], its successors have been applied to a wide range of research fields within the life sciences, among them the study of cancer. Many computational tools have been developed to fully utilize the information obtained through the new experimental methods: Clustering of sequenced cell populations into different cell types or other subgroups enables studying tumor composition. Differential expression analysis makes it possible to identify which genes characterize certain cell types and show differences between healthy and diseased individuals or entire species [2]. Trajectory inference (TI) aims to sort cells along paths leading from less developed to more developed states. By investigating cells' development processes in a continuous manner instead of only considering averages for different stages, more precise differential expression results can be achieved [3], and a greater understanding of the studied process obtained.

While single-cell analysis methods have generally found great application in cancer research, leading to a detailed exploration of the heterogeneity of tumors, trajectory inference has only recently begun to be applied in this context [4],[5]. Thus we wanted to explore how it can be employed and whether it can help us understand the nature of changes in tumor composition, which occur both in natural tumor development and in response to therapies.

The first method developed to generate pseudo-temporal orderings was Monocle, published in 2014. Monocle was first applied to the differentiation of human skeletal myoblasts [3]. Another common validation example for TI methods is hematopoiesis [6]. Nearly all trajectory inference methods developed thus far predominantly concentrate on studying differentiation processes, which differ significantly from cancer progression or cancer cell development. Therefore it is unclear whether assumptions made by these methods also hold in the context of cancer, and we wanted to investigate this.

We chose Glioblastoma (also called Glioblastoma Multiform or GBM), a malignant brain tumor, as an example to test trajectory inference in cancer because GBMs display significant heterogeneity, and detailed studies have identified subgroups of cells within them between which transitions could exist [7],[8],[9]. Therefore they are an ideal example to test trajectory inference methods in a malignant context. The inter- and intra-tumoral heterogeneity of GBM is a critical obstacle to their treatment, and research in this field has the potential to lead to improvements in patient outcomes in the future [10].

In this report, we will first give a more detailed introduction to current knowledge of Glioblastoma and further explore the problem of applying trajectory inference to GBM gene expression data. Then we will present the methods we have chosen to test and their respective results. Finally, we will discuss the outcomes and suggest further ways to test trajectory inference in cancer.

2 Opportunities and challenges for trajectory inference in GBMs

Glioma is a type of brain tumor originating from glial cells. It is divided into subtypes depending on which type of glial cells it develops from. Gliomas developing from astrocytes are called Astrocytomas. Glioblastomas are high-grade Astrocytomas [11]. Based on gene expression and mutations, GBMs can further be subclassified into proneural (PDGFRA amplification, point mutations in IDH1), neural (expression of neuron markers NEFL, GABRA1, SYT1, and SLC12A5), classical (EGFR amplification) or mesenchymal (NF1 deletion) GBMs. There are associations between tumor subtype and age, patient outcome, and treatment efficacy [12].

Aside from inter-tumoral heterogeneity, there is also intra-tumoral heterogeneity in GBMs. In our project, we mainly built on results obtained by Neftel et al. [8] that describe this cellular diversity. In their 2019 study, they performed single-cell RNA-sequencing with SMART-Seq2 of 28 tumors. Within single tumors, they found expression programs that were exhibited by many cells and identified those programs that could

be found in many of the studied tumors. Through a process of hierarchical clustering, they found four meta-modules of which two are further subdivided into two related modules. The expression of these modules then defines cellular states in GBMs. They name these states after the cell types whose developmental trajectories they recapitulate:

1. neural-progenitor-like (NPC1 and NPC2)
2. oligodendrocyte-progenitor-like (OPC)
3. astrocyte-like (AC)
4. mesenchymal-like (MES1 and MES2)

The proportions of cells of these different types differed between tumors and were found to be consistent with the subclassification of GBMs into proneural, neural, classical, and mesenchymal GBMs described above. Neftel et al. went on to identify genetic mutations that determine the cell state proportions of a tumor. To find such mutations, they looked at genes that were not part of the meta-modules but still correlated with a high frequency of a particular state within the tumor. They found that *EGFR* aberrations are relevant for the occurrence of AC-like cells, amplifications of *PDGFRA* are related to OPC-like abundance, amplifications of *CDK4* influence the proportion of NPC-like cells, and *NF1* point mutations correlated with MES-like abundance. They hypothesize that these genetic changes could be facilitating or inhibiting transitions between cell states. They also directly show that plasticity between the states exists and can be induced via the overexpression of the above genes in mouse neural progenitor cells [8].

We aimed to determine whether existing trajectory inference methods could identify trajectories that represent these transitions and thus, whether they are applicable in this context. If so, this could, for example, enable the investigation of gene expression changes along sub-module transitions and uncover factors that influence why certain cell states are more prevalent in some tumors than in others.

As described in the introduction, TI has been developed in the context of cell differentiation. We identified several assumptions made by many TI methods that could be violated by scRNA-seq data from tumors:

1. *Unidirectionality:* It is assumed that the differentiation process only occurs in one direction, from less differentiated states to more differentiated states [13]. In the case of cell states in GBMs, this unidirectionality does not seem to hold. Neftel et al. show that xenografts of cells of only one state into mice brains initiated the development of tumors containing several cell states. With cellular barcoding, they demonstrated that these cells in other states are either originally implanted cells or their progenitors [8]. Thus it is possible that cells can transition between the four defined cell states in many directions, and unidirectionality cannot without limitations be assumed in our context.

2. *Markov assumption:* This assumption posits that the development of a cell is only dependent on its current state and not its past states. The authors of Palantir note that this assumption is violated by individual cells, however, it is a sensible assumption for abstracted cell states [13]. In the case of GBM cell states it is unclear whether this assumption holds or not.

3. *Time progression:* Pseudotime is inferred based on distances from the starting cell. In the case of cell differentiation, it can thus correspond to the actual time course of the differentiation process. When applying trajectory inference to cancer, however, it is not clear whether a process of cells developing into a certain state over time is captured or whether other factors cause the gradient that is inferred. For example, the spatial structure of tumors could also create a gradient in expression patterns that could be inferred to be a trajectory.
   A possible way to separate these factors would be to use spatial data to check whether the trajectory matches a spatial pattern. To confirm the temporal nature, RNA velocity could be used. RNA velocity as the time derivative of gene expression indicates in which direction a change in expression is occurring. However, as it is calculated based on the ratio of unspliced and spliced RNA, it is most suitable to processes occurring on time scales of a few hours and probably not applicable in this case.

4. *Expression changes as drivers:* In cell differentiation the underlying DNA sequence is typically constant apart from very few mutations, while changes in the expression of genes are driven by epigenetic changes. In cancer, on the other hand, large-scale genetic mutations are present and may also accumulate over time. While this assumption is not explicitly built into TI methods, it does distinguish these two processes from one another and may make the validation of TI methods in the differentiation context less applicable to the cancer development process.

Given that these assumptions don't necessarily hold for GBM data, it is important to properly verify whether the results of TI are correct. Studies that verified the correctness of inferred trajectories in other contexts worked on well-studied differentiation processes, while we are looking at transitions that are not well understood yet. This makes it challenging to assess the accuracy of the inferred trajectories. Additionally, some studies of TI methods use time series data to test their methods. While there is no time series data of GBMs in humans, using data from mouse models would be an option to more rigorously test whether the trajectories found with TI are correct in the future.

3 Trajectory inference methods and respective results

Over 80 methods for trajectory inference have been developed. A 2019 review by Saelens et al. [14] compared available methods systematically. To do so, the authors have developed an R package called `dynbenchmark` to represent the different outputs given by the methods in a comparable way. They point out that the methods have different degrees of generalization. They can either infer only specific topologies as trajectories (such as linear or tree-shaped) or general trajectories (including circular trajectories). In our case, we wanted to use general methods, as results from Neftel et al. [8] suggest that non-linear and circular trajectories could be present in our data. This condition ruled out many of the existing methods.

Saelens et al. score each method according to its accuracy, scalability, stability, and usability. Among the general methods, PAGA, implemented in Python, scored highest in each one of these categories [14]. While many newer methods claim to have superior performance [15], no systematic review of TI methods has been published since 2019, and thus we decided to include PAGA in our project.

Among the newer methods, Palantir is widely used, can find general trajectories, has been tested extensively by the authors, and is also available in Python [13]. Thus we chose this as another method to experiment with. Due to the project's limited scope, we primarily focused on these two trajectory inference methods, although numerous other options are available. As more recent methods tend to come out as R packages, we would suggest using R when working with trajectory inference in the future.

3.1 Preprocessing

The data was downloaded from the *CELLxGENE* glioblastoma atlas [16]. The counts are already normalized and log-transformed. Sequences obtained through 10x v2 sequencing and from the study by Neftel et al. [8] were used. Only cells annotated as malignant were kept. Selecting patients with enough remaining cells, five patients were included in our analysis.

Cells were assigned cell cycle scores with the function `scanpy.tl.score_genes_cell_cycle()`. It takes the average expression of a list of genes for each cell cycle stage from [17] and compares it to the average expression of a randomly sampled list of genes. Afterward, cells were clustered with Leiden clustering. Clusters with more than 75 percent cycling cells (S or G2M phase) were removed.

To annotate meta-modules, the genes identified by Neftel et al. were passed into the function `sc.tl.score_genes()`. Each cell was assigned to the meta-module with the maximum score.

3.2 PAGA

PAGA or Partition-based graph abstraction is an unsupervised method for trajectory inference. To apply it, we followed the steps outlined in [18]. A graphical representation of the process together with the results for one patient with a proneural tumor can be found in figure 1. As a preprocessing step, we applied PCA to our data. The PCA positions of cells are then used for all other visualizations. An example result can be seen in panel *a* of figure 1. Each plot is colored by the cell type score for one of the cell states AC, OPC, NPC, and MES. The light yellow areas with the highest scores are differently positioned in each plot, indicating that the meta-modules are separated in PCA space, but many intermediary cells can also be seen.

PAGA then represents the data as a neighborhood graph of cells. This graph is weighted with a kernel that can be chosen by the user, the default being the Gaussian kernel. From this neighborhood graph, the method constructs a so-called PAGA graph. This is done by partitioning the graph through Louvain clustering. Louvain clustering maximizes the modularity and the choice of the resolution parameter affects the number of clusters. The authors of PAGA argue that this allows inferring trajectories on different scales, i.e. on a cell-to-cell or group-to-group level [6]. We chose a resolution of 1 such that there were several clusters within each visually identified submodule, but not too many clusters as this would require more user input that is potentially inaccurate (more details on this can be found below).

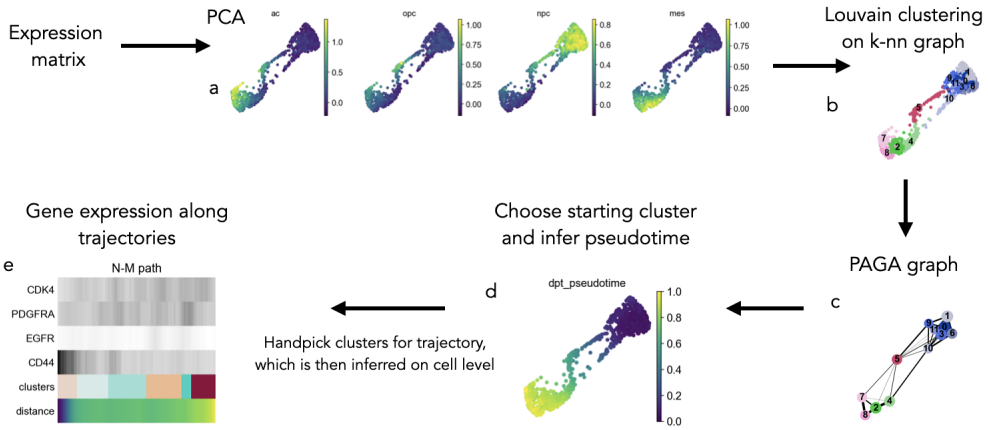

Figure 1: Process of using PAGA for TI inference, patient 124, the last step is an example image with data in a different format

The user chooses a starting cluster and PAGA then uses diffusion pseudotime to temporally order the cells. Diffusion pseudotime applies random walks on diffusion maps to calculate distances between cells based on their gene expression. This distance measure assumes a unidirectional progression, which is typical for cell differentiation, but might not be an accurate assumption in the malignant case. As can be seen in 1, by comparing the plots colored by cell type scores in panel *a* with that colored by pseudotime in panel *d*, the inferred pseudotime matches a potential transition between submodules.

With the help of the provided pseudotime, the user then picks clusters that belong to the studied trajectory and determines their order. In our example, this could for example be a path through clusters 0, 3, 9, 4, and 2 to represent a transition from the NPC to the MES state. This step also allows one to incorporate any other information one might have into the inferred trajectories by picking the clusters and their order accordingly. On the other hand, it can introduce errors since the 2D visual representation does not allow for accurately identifying the most likely trajectory. From this rough trajectory, PAGA then calculates the cell-level trajectory and it can visualize pseudotime distance and gene expression changes along the trajectory. The function provided by the PAGA package to display this graph did not apply to our input data format. Thus the final step in figure 1 is represented by an image of this output for example data

from the same publication in a different format that had been used in previous versions of the analysis. It is desirable to have a trajectory with a monotonous progression of pseudotime along the trajectory and if this is not the case it is possible to identify which clusters cause this and to reorder them. A more detailed description of the PAGA algorithm can be found in [6].

Overall, we found that mostly monotonous trajectories could be constructed and that there were some gene expression changes for the four genes identified by Neftel et al. along most trajectories. It would be possible to further investigate the validity of the found trajectories, however, it is relatively labor-intensive to construct them and it is to be expected that methods that replace the manual steps with algorithms give more reproducible and reliable results. We thus continued our analysis with a newer method called Palantir.

3.3 Palantir

3.3.1 The method

Palantir is also a graph-based method that models the differentiation process as a Markov process. It was developed with the motivation of better understanding the continuity in cell development processes. Crucially, Palantir makes two assumptions about the process that is investigated, namely, it assumes unidirectionality and the Markov property [13].

To analyze our data with Palantir, we followed a tutorial provided by the authors [19] and afterward created custom plots to visualize the results. Figure 2 shows the steps followed in applying Palantir.

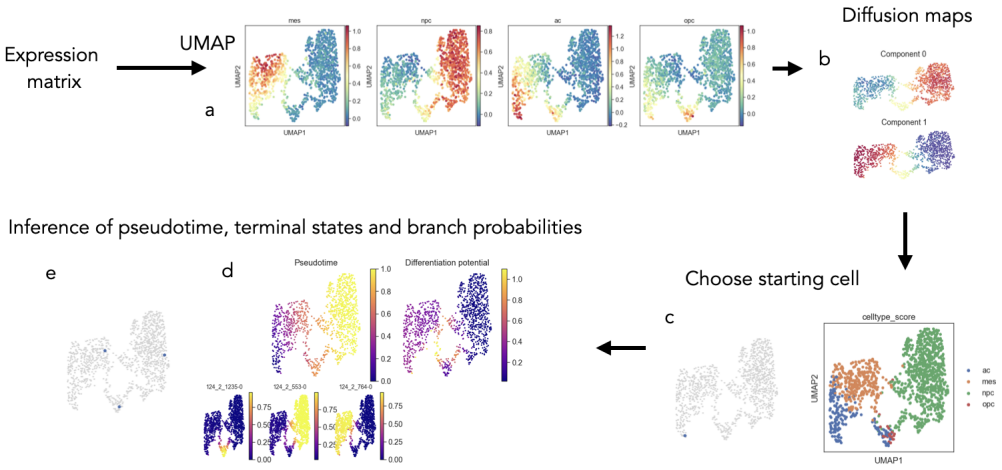

Figure 2: Palantir Process and results visualizations (patient 124)

For visualization, UMAP is first applied to reduce the dimensionality of the data. A separation of cells in the different states can be seen in Figure 2 panel *a*. For the pseudotime calculations, Palantir uses diffusion maps and the authors state that it can capture more trajectories than other methods because more than one diffusion component is used. The starting cell of the trajectories has to be picked by the user. In a usual trajectory inference scenario studying cell differentiation, this would be a stem cell. In the case of GBM data, the meta-modules are more informative for cell state than stemness, as certain meta-modules have more stem cell markers. Therefore we picked the cells with the maximum meta module score for one meta module as starting cells. In our example in Figure 2 this is an AC cell (see panel *c*). The observed dependence of results on the choice of starting cell is discussed in subsection 3.3.2.

Pseudotime for each cell is initiated as the shortest distance from the starting cell. It is then recalculated from sampled points along pseudotime. If unidirectionality is violated this could lead to cells being the same distance from a starting cell but in different temporal directions and thus not close to each other, which

could lead to unwanted effects. Using the directionality on the neighborhood graph imposed by Pseudotime, a Markov chain is constructed and random walks are simulated on the graph to infer terminal states as outliers of the stationary distribution, and branch probabilities. All edges going out from terminal states are deleted. Each terminal state defines a branch. In Figure 2, the terminal states are plotted in panel e, and one terminal state was found in each cell state, apart from the starting state.

Definition 3.1. *The* Branch probability $p_C(t)$ *for a cell* C *and terminal state* t *is the probability that a random walk starting in* C *would end in* t*. It is a measure of the commitment of a cell to a particular fate.*

Apart from pseudotime and branch probabilities for each terminal state, Palantir outputs a differentiation potential for each cell.

Definition 3.2. *The* Differentiation potential (DP) *of a cell is the entropy over the branch probabilities. Let* T *be the set of terminal states.*

$$DP(C) = -\sum_{t \in T} p_C(t) log(p_C(t))$$

A high differentiation potential signifies that a cell is not yet committed to a branch and still has high plasticity, while the minimum DP of 0 means that a cell is fully committed to a branch. Panel d of 2 shows plots of Pseudotime and Differentiation potential for the example tumor. The three plots in the bottom row display the branch probabilities for each of the three branches. More details of the Palantir algorithm can be found in [13].

3.3.2 Evaluating trajectories

To test whether the inferred terminal states and trajectories correspond to the state transitions we wanted to observe, we looked at the location of starting cells and terminal states in the UMAP visualization of the data points, the relationship between branch probabilities and pseudotime and the correlation between cell-type score and branch probabilities, as well as between differentiation potential and branch probabilities.

Across all patients and starting cell types, we found terminal states representing other cell types than the starting cell or rarely cells of the same type as the starting cell but projected near other cell types. Most often terminal states of all other cell types were identified (for example see Figure 2 panel e). When changing the starting cell to a similar cell, usually the same terminal states were inferred. When changing the starting cell to a cell of a different type, usually a terminal state close to the original starting cell was inferred. Based on these results we could clearly assign all found trajectories to a type of transition that it most likely represents.

Visualizing the branch probabilities over pseudotime allows us to compare all existing trajectories for one patient and a particular starting cell. In Figure 3, the cells are colored by their highest branch probability. The y-axis of each plot is the branch probability of one of the branches. The starting cell was an OPC cell and there are one MES terminal state and two NPC terminal states. A clear separation between the trajectories can be observed. This separation does not appear for all trajectories at the same pseudotime point. Instead, the blue NPC trajectory splits off earlier. This suggests a topology in which there is first a separation into two trajectories and later in pseudotime there is another separation occurring on one of the branches.

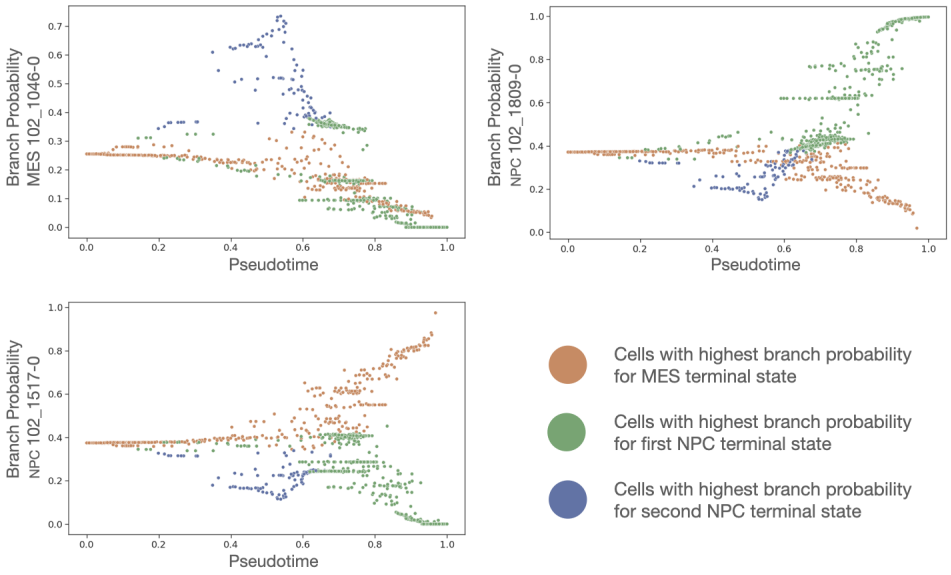

Figure 3: Branch probabilities over pseudotime, colored by the branch with maximum probability (patient 102)

To confirm that not only start and end points of the trajectory correspond to the expected cell types but that there is also a continuous transition toward the terminal cell type along the trajectory, we visualized cell-type score vs. branch probability. Cells with a higher maximum branch probability are more committed to one branch and should thus also have a cell type score closer to that of the respective terminal cell type. Figure 4 shows such plots for an NPC to MES trajectory. While there appears to be a roughly linear correlation between the cell type scores corresponding to starting and terminal cells, we observe that the other scores first rise and then go down again. This could be because cells in transition states are more similar to each other than at the extremes of the cell types.

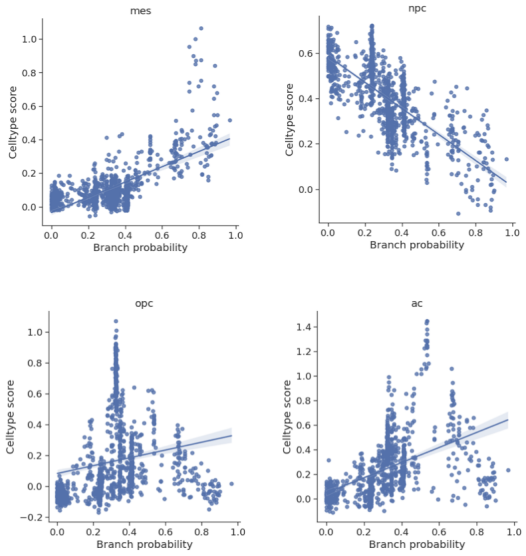

Figure 4: Correlation of cell type score with branch probability for a branch representing an NPC to Mes trajectory (patient 124)

The results of all these visualizations suggest that the found trajectories indeed correspond to the transitions we expected to find.

3.3.3 Genetic changes along the trajectories

Finally, we wanted to investigate whether the genes that Neftel et al. hypothesize to down- or upregulate transitions between states were connected to the trajectories we identified. Figure 5 shows Gene expression vs. Pseudotime plots for these genes and a trajectory from the AC to the NPC state. As in the Palantir tutorial, MAGIC imputation was applied before plotting.

For the *EGFR* gene, which is related to AC abundance, a clear negative correlation can be observed. However, looking at the plot, there appear to be two relatively straight lines, suggesting that there is a more complicated relation between this gene and the trajectories. One possibility could be that this represents two possible paths from the AC cluster to the NPC cluster, one above the blank space which can be seen in the UMAP plot, and one below (see Figure 2 panel c). This could represent two possibilities for transitioning between the states that are mediated by different genes. Alternatively, this could also represent backward and forward transitions between the states, that are not correctly ordered in pseudotime, as unidirectionality is assumed. In the *PDGFRA* plot (gene related to OPC state), a similar triangular pattern appears. The phenomenon of several lines with the same sign but different magnitude slopes appeared in several transitions and patients, just as the UMAP representations with circular patterns that could represent several different trajectories grouped into one trajectory.

Looking at the plot for *CDK4*, which is related to the NPC state, we can see a different pattern. Here the expression of the gene is roughly the same at the beginning and end of the trajectory but higher in the center. This pattern also occurred several times and could mean that the gene is upregulated during the transition but not in the well-defined beginning and end states.

Finally, the *NF1* plot shows an increase in *NF1*. While *NF1* was not named to play a role in either transition to the AC or to the NPC state, this trend could result from other correlations, such as a higher similarity of some states to others. In the UMAP plot 2, we can see that the MES state (which is related to NF1) appears between the AC and the NPC state. Also in other patients such trends in unrelated genes occurred, though they were most often more similar to the *CDK4* plot shown here with an expression increase in the center.

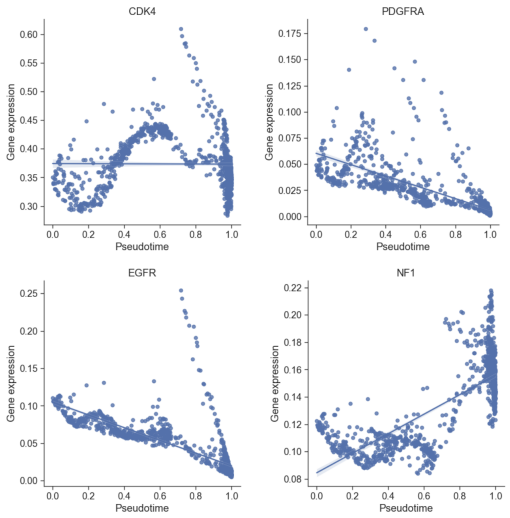

Figure 5: Genetic changes through pseudotime for a trajectory from AC to NPC state (patient 124)

Overall, the analysis of genetic changes along the trajectories suggests that the genes named by Neftel et al. play a role in the transitions between cell states. However, the plots also point to some hidden effects in the

trajectories that could, for example, result from the violation of unidirectionality or other reasons that prevent mechanistically different trajectories from being separated. Since we only looked at a few patients (for some types of GBMs just one patient), we could not test whether the relationships we found were consistent and statistically significant. To get a clear picture, our analysis would have to be expanded to a larger dataset.

4 Discussion

We tested two established trajectory inference methods on scRNA-seq data from Glioblastoma samples. The output trajectories matched the trajectories we expected to find, and there were strong correlations between branch probability and strength of submodule identity. The genes responsible for the submodule proportions in tumors also appear connected to these transitions, largely not by simple linear relationships. We conclude that at least some of the existing trajectory inference methods can be applied to expression data from tumors and lead to useful results.

Some questions about the theoretical validity of the assumptions that trajectory inference makes in the case of malignant data nevertheless remain. Without access to ground truth trajectories and no established understanding of how the investigated transition processes work, it is hard to test these assumptions. Sometimes it is possible to simulate data with known ground truth for cases like this, however, specifically regarding the state transitions, it is unclear how such data could be generated to accurately match the biological system. Therefore the best option, for now, is to apply methods to measured data and determine the validity of the results for example by looking at the correlations we discussed. Comparing the results of more different methods could also validate inferred trajectories. Further, time series data from mice models could also serve as another factor to verify the results of TI in Glioblastoma.

For this project, we only used gene expression data. It would also be interesting to see how our findings generalize, for example, to epigenetic data and if multimodal models improve the quality of the trajectories. Epigenetic analysis could help in elucidating the mechanisms behind the trajectories. Understanding how cells transition between the states may aid in comprehending and preventing the therapeutic resistance of Glioblastoma.

References

[1] Fuchou Tang et al. "mRNA-Seq whole-transcriptome analysis of a single cell". In: *Nature Methods 2009 6:5* 6 (5 Apr. 2009), pp. 377–382. ISSN: 1548-7105. DOI: 10.1038/nmeth.1315. URL: https://www.nature.com/articles/nmeth.1315.

[2] Malte D Luecken and Fabian J Theis. "Current best practices in single-cell RNA-seq analysis: a tutorial". In: *Molecular systems biology* 15 (6 June 2019). ISSN: 1744-4292. DOI: 10.15252/MSB.20188746. URL: https://pubmed.ncbi.nlm.nih.gov/31217225/.

[3] Cole Trapnell et al. "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells". In: *Nature Biotechnology 2014 32:4* 32 (4 Mar. 2014), pp. 381–386. ISSN: 1546-1696. DOI: 10.1038/nbt.2859. URL: https://www.nature.com/articles/nbt.2859.

[4] Bo Pang et al. "Single-cell RNA-seq reveals the invasive trajectory and molecular cascades underlying glioblastoma progression". In: *Molecular Oncology* 13 (12 Dec. 2019), p. 2588. ISSN: 18780261. DOI: 10.1002/1878-0261.12569. URL: /pmc/articles/PMC6887585/%20/pmc/articles/PMC6887585/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6887585/.

[5] Abicumaran Uthamacumaran and Morgan Craig. "Algorithmic reconstruction of glioblastoma network complexity". In: *iScience* 25 (5 May 2022), p. 104179. ISSN: 2589-0042. DOI: 10.1016/J.ISCI.2022.104179.

[6] F. Alexander Wolf et al. "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells". In: *Genome Biology* 20 (1 Mar. 2019), pp. 1–9. ISSN: 1474760X. DOI: 10.1186/S13059-019-1663-X/FIGURES/4. URL: https://link.springer.com/articles/10.1186/s13059-019-1663-x%20https://link.springer.com/article/10.1186/s13059-019-1663-x.

[7] Spyros Darmanis et al. "Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma". In: *Cell reports* 21 (5 Oct. 2017). One of 2 datasets used in ¡i¿Pang, B., Xu, J., Hu, J., Guo, F., Wan, L., Cheng, M., Pang, L. (2019). Single-cell RNA-seq reveals the invasive trajectory and molecular cascades underlying glioblastoma progression. Molecular Oncology, 13(12), 2588. https://doi.org/10.1002/1878-0261.12569¡/i¿, pp. 1399–1410. ISSN: 2211-1247. DOI: 10.1016/J.CELREP.2017.10.030. URL: https://pubmed.ncbi.nlm.nih.gov/29091775/.

[8] C. Neftel et al. "An integrative model of cellular states, plasticity and genetics for glioblastoma". In: *Cell* 178 (4 Aug. 2019), p. 835. ISSN: 10974172. DOI: 10.1016/J.CELL.2019.06.024. URL: /pmc/articles/PMC6703186/%20/pmc/articles/PMC6703186/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6703186/.

[9] Charles P. Couturier et al. "Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy". In: *Nature Communications 2020 11:1* 11 (1 July 2020), pp. 1–19. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17186-5. URL: https://www.nature.com/articles/s41467-020-17186-5.

[10] Christopher M. Jackson, John Choi, and Michael Lim. "Mechanisms of immunotherapy resistance: lessons from glioblastoma". In: *Nature Immunology 2019 20:9* 20 (9 July 2019), pp. 1100–1109. ISSN: 1529-2916. DOI: 10.1038/s41590-019-0433-y. URL: https://www.nature.com/articles/s41590-019-0433-y.

[11] *Gliomas by Hopkins Medicine*. https://www.hopkinsmedicine.org/health/conditions-and-diseases/gliomas. Retrieved: 2023-02-12.

[12] Roel G.W. Verhaak et al. "Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1". In: *Cancer Cell* 17 (1 Jan. 2010), pp. 98–110. ISSN: 1535-6108. DOI: 10.1016/J.CCR.2009.12.020.

[13] Manu Setty et al. "Characterization of cell fate probabilities in single-cell data with Palantir". In: *Nature Biotechnology 2019 37:4* 37 (4 Mar. 2019), pp. 451–460. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0068-4. URL: https://www.nature.com/articles/s41587-019-0068-4.

[14] Wouter Saelens et al. "A comparison of single-cell trajectory inference methods". In: *Nature Biotechnology 2019 37:5* 37 (5 Apr. 2019), pp. 547–554. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0071-9. URL: https://www.nature.com/articles/s41587-019-0071-9.

[15] Louise Deconinck et al. "Recent advances in trajectory inference from single-cell omics data". In: *Current Opinion in Systems Biology* 27 (Sept. 2021), p. 100344. ISSN: 2452-3100. DOI: 10.1016/J.COISB.2021.05.005.

[16] *Chanzuckerberg Initiative. (n.d.). CZ CELLxGENE Discover.* https://cellxgene.cziscience.com/. Retrieved: 2023-02-01.

[17] Rahul Satija et al. "Spatial reconstruction of single-cell gene expression data". In: *Nature Biotechnology 2015 33:5* 33 (5 Apr. 2015), pp. 495–502. ISSN: 1546-1696. DOI: 10.1038/nbt.3192. URL: https://www.nature.com/articles/nbt.3192.

[18] *Trajectory inference for hematopoiesis in mouse.* https://scanpy-tutorials.readthedocs.io/en/latest/paga-paul15.html. Retrieved: 2023-01-25.

[19] *Palantir analysis notebook.* https://nbviewer.org/github/dpeerlab/Palantir/blob/master/notebooks/Palantir_sample_notebook.ipynb. Retrieved: 2023-01-25.