Mathematical modelling of interacting subpopulations in glioblastoma using pseudotime

Nicholas Harbour[1], Lee Curtin[3], Michael Chappell[2], Matthew Hubbard[1], Kristin Swanson[3], Markus Owen[1]
[1]School of Mathematical Sciences, University Nottingham, UK, [2]Sir Peter Mansfield Imaging Centre, University Nottingham, UK,
[3]Mathematical NeuroOncology Lab, Mayo Clinic, AZ, USA

INTRODUCTION

A major challenge in treating glioblastoma (GBM) is the heterogeneity in cellular composition within and between patients. Single cell transcriptomics suggests as many as eighteen distinct cell types in GBM[1]. Cellular deconvolution techniques, such as CIBERSORTx, can reveal the cellular composition from more commonly available bulk RNAseq[2]. Understanding how interactions between these cellular populations impacts GBM progression may lead to novel treatments that exploit the unique cellular composition of individual tumours.

This poster shows preliminary results from the first six months of my PhD. We consider how pseudotime techniques can be used to infer mathematical models from RNAseq data. In the first case study we assume all patients follow the same progression trajectory. Then we use pseudotime techniques to correctly reorder samples. From this we can learn the ODE system that was used to generate the data. In the second case study we consider a more complex scenario in which both the sample time and initial condition is unknown. We show that we are still able to learn information about the parameters that generate this data.

PSEUDOTIME

Improvements in transcriptomics have produced a wealth of biological data. However, RNAseq only provides a snapshot at a particular time point. Biological systems are inherently dynamic. There are many difficulties in capturing accurate time series data, particularly in GBM progression:

- Typically, A GBM patient will have biopsies taken at most twice due to the difficulty of the operation.
- There is significant heterogeneity in the rate of progression dynamics within and between patients.

We use a minimum spanning tree (MST) approach to calculate pseudotime, based on techniques such as Monocle[3]. The key assumptions of our approach are:

- States which are transcriptionally similar are also temporally close together.
- All patients' progression follows a limited number of possible trajectories.

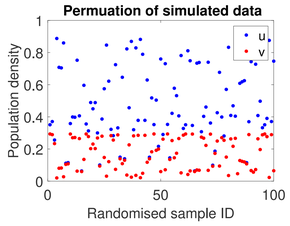
LOTKA-VOLTERRA STUDY

Mathematical models can be used to predict the temporal evolution of interacting populations. A prototypical model for studying this is the Lotka-Volterra (LV) competition model. For two species this takes the form:

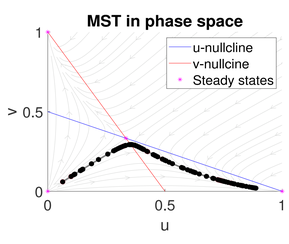
$$\frac{du}{dt} = u(1 - u - av)$$
$$\frac{dv}{dt} = \rho v(1 - v - bu)$$

where a, b represent the strength of competition, and ρ the relative proliferation rate. First, assume that all patients are on the same trajectory. However, the time at which samples are taken is random and unknown.

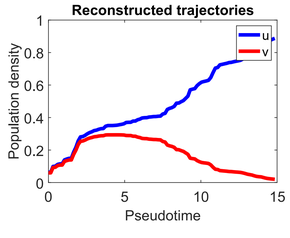
Simulate LV system with random sample times



Calculate MST in phase space and find longest path



Reconstruct trajectory



Learn ODE system[4]
True parameters:
$a = 2, b = 2, \rho = 1$

Inferred ODE system

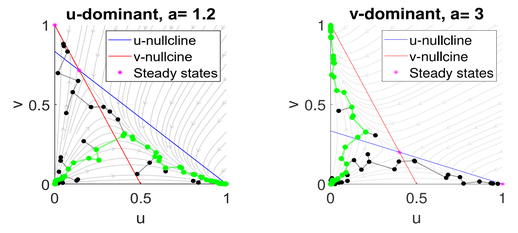
$$\frac{du}{dt} = u(1.12 - 1.15u - 2.23v)$$
$$\frac{dv}{dt} = v(1.05 - 1.05v - 2.06u)$$

POPULATION STUDY

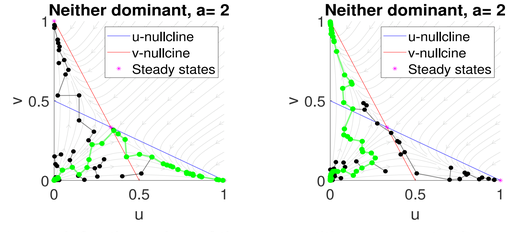
- Simulate 200 samples each with a different initial condition (IC) of cancer cell populations, for a random amount of time. The random IC's will ensure that all samples are on different trajectories, despite the fact that all samples follow the same dynamics.
- Carry out repeated subsampling, each time taking a subsample of 100.
- For each subsample calculate the MST, starting from the node closest to the origin.
- Repeat these steps for different values of the competition parameter a, keeping $b = 2$ fixed.

Possible trajectories in phase space

When the competition parameters are significantly different, the majority of points lie on qualitatively similar trajectories.

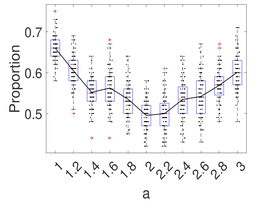


With equal competition parameters, samples are evenly distributed between two qualitatively different trajectories.

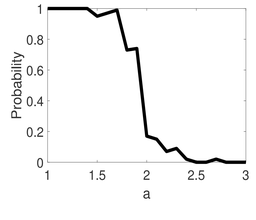


We can infer the value of the competition parameter by calculating the the number of nodes in the longest path through the MST, as well as where the final node is located.

Proportion of nodes in longest path



Proportion of trajectories that end at $u = 1$



FUTURE WORK

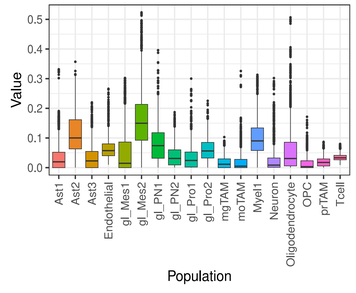
The ultimate goal is to be able to use pseudotime ordering to accurately infer dynamical models from static RNAseq data. Some of the current challenges are:

- Not all patients will follow the same dynamics.
- Real biopsy data contains significant amounts of noise.
- There are many more than two cellular populations and a variety of complex interactions.

The future steps in this work include:

- Introduce multiple trajectories/branching into MST.
- Test the robustness of our pseudotime inference with the addition of noise.
- More complicated dynamical systems.

Example CIBERSORTx data for GBM



REFERENCES

1. O. Al-Dalahmah, et al., Re-convolving the compositional landscape of primary and recurrent glioblastoma using single nulcues RNA sequencing. bioRxiv 2021
2. C. Steen, et al., Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. Methods Mol. Biol. 2117, 135-157, 2020.
3. C.Trapnell, et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol., 32, 381-386, 2014.
4. C. Rudy, et al., Data-driven discovery of partial differential equations. Sci. Adv., 3, e1602614, 2017