Abstract

Background: Single-cell RNA sequencing, (scRNA-seq), has been used to quantify the epigenetic landscape of cells undergoing developmental processes. Dimensionality reduction algorithms allow for cells to be placed into a two-component space where a progression through a smooth developmental path can be quantified as pseudotime. Trends in autocorrelation from time series data have been shown to predict critical transitions. Single-cell gene expression levels from murine embryonic liver progenitor cells were dimensionally reduced and modeled as time series data and autocorrelation was measured along certain intervals of pseudotime.

Results: The scRNA-seq data was dimensionally reduced to a smooth trajectory and was found to concur with previously published results. The gene expression levels and their corresponding pseudotime values were fit to ARMA models and autocorrelation was measured according to their respective lag. A correlation of varying strengths between critical transition of gene expression levels occurring and increase of autocorrelation between a pseudotime value of 30 and 38. The clustering of gene expression levels show that more genes are activated/repressed during cholangiocyte development in comparison to hepatocyte development.

Conclusion: The measurements in autocorrelation show a potential method that finds statistical characteristics and uses them to predict if cells are about to undergo a critical transition. The development of an algorithm that mines all gene expression levels for the increase in autocorrelation and correlates these genes with a critical transition occurring. This tool would allow for predictive measurement of cells undergoing a developmental or health-to-disease transition. predicted. Difficulties in the development of the algorithm include the need for visual inspection of autocorrelation functions to generate an ARMA model, and the definition of a step change in autocorrelation and expression levels.

Theory/Background

Epigenetic Landscape and Liver Development

Biological cells have been shown to undergo epigenetic changes during lineage specification. These changes are referred to as the "epigenetic landscape", first introduced by Conrad Hal Waddington [1]. Waddington presented that the epigenetic changes of a cell is like a marble that travels down an inclined surface that contains different hills which control where the marble travels. The marble is representative of a cell that has the potential of specifying into different types. The hills represent different gene regulatory networks that influence the lineage of the cell. The position of the marble represents how differentiated, or how plastic, the cell is.

The development of the rigorous quantification of this landscape has been a debated topic amongst mathematicians [2] [3] [4] [5]. Probabilistic models have been generated to relate the state the cells position in their epigenetic landscape to the gene regulatory networks [2] [6] [7] [8]. A more recent answer the question of defining cellular states employs the measurement and analysis of single-cell gene expression levels [9] [10] [11] [12] [13] [14]. The expression levels of single cells are being measured by a technique called single cell RNA-sequencing, (scRNA-seq). The population of gene expression levels with a large, ($n>100$), sample size has sequestered the development of algorithms that allow for a massively-parallel analysis. An abundance of different algorithms have been created to characterize cells based on their cellular gene expression levels [15] [16] [17] [18] [19] [20] [21] [22]. The goal of these algorithms is to identify which cells are most and least like each other by combinations and variations of dimensionality reductions

techniques such as principal component analysis, (PCA), and t-distributed stochastic neighbor embedding, (t-SNE). The reduction allows for cells to be represented by two components which are responsible for the variation amongst cells.

scRNA-seq has been used to address various human and murine cellular processes including immune cell differentiation, embryonic development, and liver cell differentiation [23] [24] [25] [26] [27] [28] [29]. The topic that this paper will specifically cover is hepatic cell development and reprogramming in murine models. The development of liver cells has been shown to have distinct gene regulatory networks that vary in metabolic function [30]. The development of a hepatoblast-hepatocyte differentiation process has been shown throughout murine embryogenesis, liver lobule location, and dioxin exposure [29] [28] [31] [32]. It has been found that the expression of highly regulated enzymes in oxygen demanding metabolic pathways are correlated to differentiation. Metabolic pathways include by not limited to glycolysis, gluconeogenesis, urea cycle, cholesterol transport and metabolism, and heme synthesis/degradation. Factors such as the hemodynamics and bile flow have also been analyzed to see their contribution to hepatic differentiation [31] [33].

This paper will analyze scRNA-seq data generated from murine liver progenitor cells, (LPCs), from different embryonic days [29]. Monocle 2, a dimensionality reduction algorithm, will be employed to map cells along their developmental pathways based on their gene expression levels [34]. The algorithm uses an adaptive kernel to align cells on a smooth trajectory. This was determined to be favorable because it allows for the assignment of a value which is a measurement of how far along cells are in their development trajectory. The assigned value is called pseudotime, and it allows for gene expression levels to be modeled as time series data.

Numerical Modeling

Time series data are commonly fit by using auto-regressive moving-average, (ARMA), models [35] [36]. ARMA models are a way of representing data based on preceding values in a time series. ARMA models consist of an auto-regressive, (AR), and a moving-average, (MA), component. The AR component is a linear regressive model that fits a representation of leading values to the lagging ones. **Equation 1** shows the representation of an AR model using stationary gene expression levels, (E).

$$E_t = \varphi_1 E_{t-1} + \varphi_2 E_{t-2} + .. + \varphi_p E_{t-p} + \varepsilon_t, \quad \{1\}$$

Where p is the number of lags chosen to represent the data and ε_t is the error term associated with the regression. If p is greater than one, the fitting of the expression levels become a multiple regression problem where the minimization of residuals needs to be taken into consideration [37]. The MA component of the models takes the form shown in **Equation 2** below:

$$E_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + .. + \theta_q \varepsilon_{t-q}, \quad \{2\}$$

Where q is the number of residual terms to be explicitly represented, and θ is a term that represents the weight of its associated error term. ARMA models are fitted to stationary data points which are typically made that way by differencing. ARMA models are fitted to stationary values by visual inspection of the autocorrelation function, (ACF), and partial autocorrelation function, (PACF), at different lags to observe how autocorrelation changes between lags and if autocorrelation exceeds

statistical significance. Criteria used to choose a model for each expression level analyzed is shown in **Table 1** [35] [36].

Table 1: Model Fitting Criteria

	AR(p)	MA(q)	ARMA(p, q)
ACF Behavior	Tails off in an exponentially damped sinusoidal fashion	Cuts off after q lag(s)	Tails off in an exponentially damped sinusoidal fashion
PACF Behavior	Cuts off after the p lag(s)	Tails off in an exponentially damped sinusoidal fashion	Tails off in an exponentially damped sinusoidal fashion

The above criteria were applied when fitting a model to each of the stationary expression levels. An ARMA model was fit to each of the expression levels to determine how autocorrelation can be measured by evaluation of the ACF at the p lag of the model.

A system is said to display a critical slowing down prior to a critical transition. This critical slowing down describes a systems recovery rate after exposure to a perturbation, which causes its intrinsic rates of change to decrease. In other words, the increase in correlation of each measured value at sequential points in time is an indicator that a critical transition will occur. A measure of how similar measured values are to each other at equidistant intervals is defined as autocorrelation [35] [36]. In other words, autocorrelation is a measure of how associated (or correlated) preceding values are to the leading value. Incipient bifurcations have been detected by an increase in autocorrelation of different data representations [38] [39] [40] [41] [42] [43] [44] [45]. The ACF is a representative measure of how autocorrelated leading and lagging sets of data are. The ACF is measured at p lags and is calculated using **Equation 3** below:

$$\mathrm{ACF} = \frac{\sum_{t=1}^{T-p}(E_{t+p}-\bar{E})(E_t-\bar{E})}{\sum_{t=1}^{T}(E_t-\bar{E})^2}, \quad \{3\}$$

Where E represents the leading, (t+p), or lagging, (t), stationary expression level vector, and $\bar{E}$ represents an average stationary expression level. An increase in autocorrelation is measured in by calculating a regression coefficient of the leading and lagging stationary expression levels preceding the bifurcation point [45]. Autocorrelation is also measured by sliding a window of 10% length that moves in increments of p across interpolated expression levels [44] [38] [39]. Here measurements in both the static and dynamic manners are presented. See **Methods** for further explanation of specific application.

It is important to keep in mind is that for p>1, the autocorrelation function evaluated at lag p is not the same as the regression coefficient for the AR(p) model. The autocorrelation function is a representation of how correlated the leading and lagging data values are, and not always the regression coefficient for the AR(p) model.

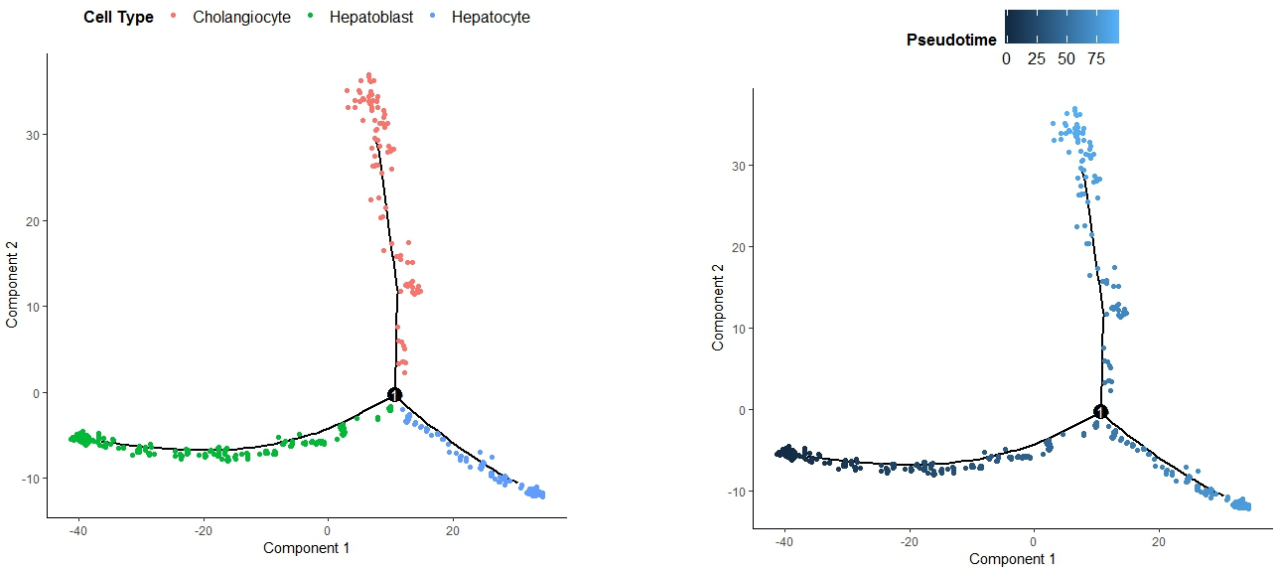
scRNA-seq Application

Gene expression levels generated from the previously mentioned scRNA-seq data and their respective pseudotime values were fit to ARMA models. Increases in autocorrelation were found by measuring ACF values along certain intervals of pseudotime with a sliding window and were measured qualitatively by plotting the leading and lagging stationary expression levels prior to the bifurcation point determined by the Monocle 2 algorithm. The autocorrelation was measured over a 10% sliding window as mentioned previously. It is expected in this case that the cells in an undifferentiated state will show an increase in autocorrelation preceding their transition into a differentiated state.

Results

Dimensionality Reduction

Cells were organized in pseudochronological order and were assigned a pseudotime value that measures how far along the cell is along it's trajectory. The value assigned is a direct measurement of how differentiated the cell is; the higher the pseudotime value, the more differentiated the cell is. **Figure 1 a** and **b** depict the cellular trajectory with assigned cell type and pseudotime values, respectively.



a.) Cell Type Depiction

b.) Pseudotime Depiction

Figure 1: Developmental Trajectory of Cells[1]

The figure depicts more of a gradual transition in the hepatoblast-hepatocyte branch and shows discontinuity in the hepatoblast-cholangiocyte branch. A heatmap depicting gene expression levels of 9539 genes amongst both branches is shown in **Figure 2**. Cluster 9 appears to be genes that are heavily up-regulated in hepatocyte development and Clusters 2, 4, 5, 7, and 10 are genes that are up-regulated specific to cholangiocyte development. Cluster 8 appears to be genes that are specific

[1] Data accessed from GSE90047.

to down-regulation during cholangiocyte development. Overall, more genes are activated or repressed during cholangiocyte development than in hepatocyte development.

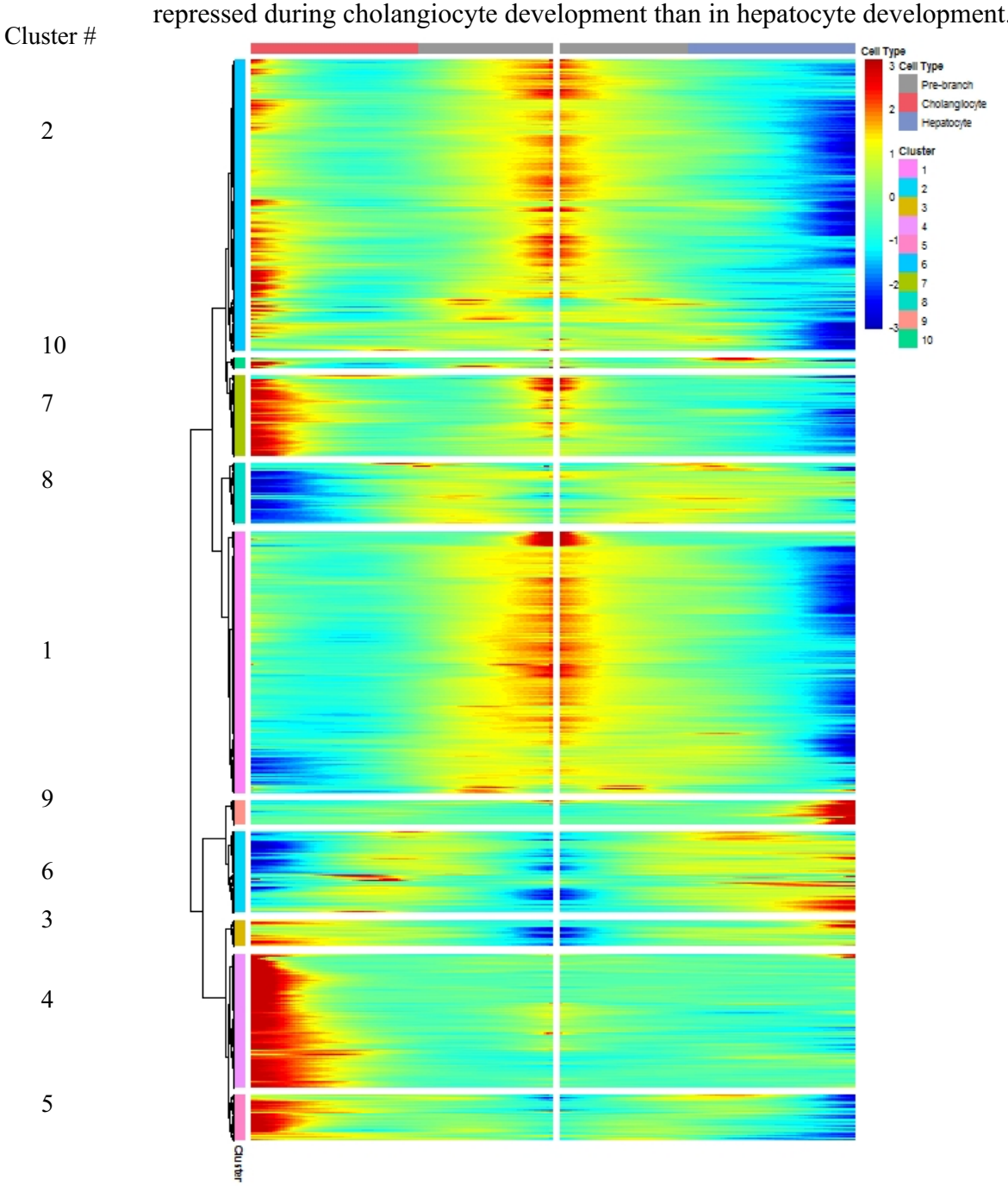


Figure 2: Heatmap of Genes Active in Hepatic Development[2]

Further details on expression level filtering and transformation are shown in **Methods**. A heatmap that depicts the 20 genes identified to be active in hepatic development is shown in **Figure S1** [29].

[2] Asymmetries in expression levels exist due to interpolation. Heatmap generated using the Monocle and pheatmap package available in *R*.

Incipient Bifurcation Detection via Measurement of Autocorrelation

Known gene markers in the hepatic differentiation were picked and a select few were chosen as markers that undergo a critical transition [29] [30] [31] [46]. ARMA models were fitted to 27 stationary gene expression levels, and static and dynamic autocorrelation were qualitatively and quantitatively measured, respectively (see **Methods** for exact details).

The leading and lagging stationary expression levels that show the strongest static correlations are presented here in **Figure 3**.

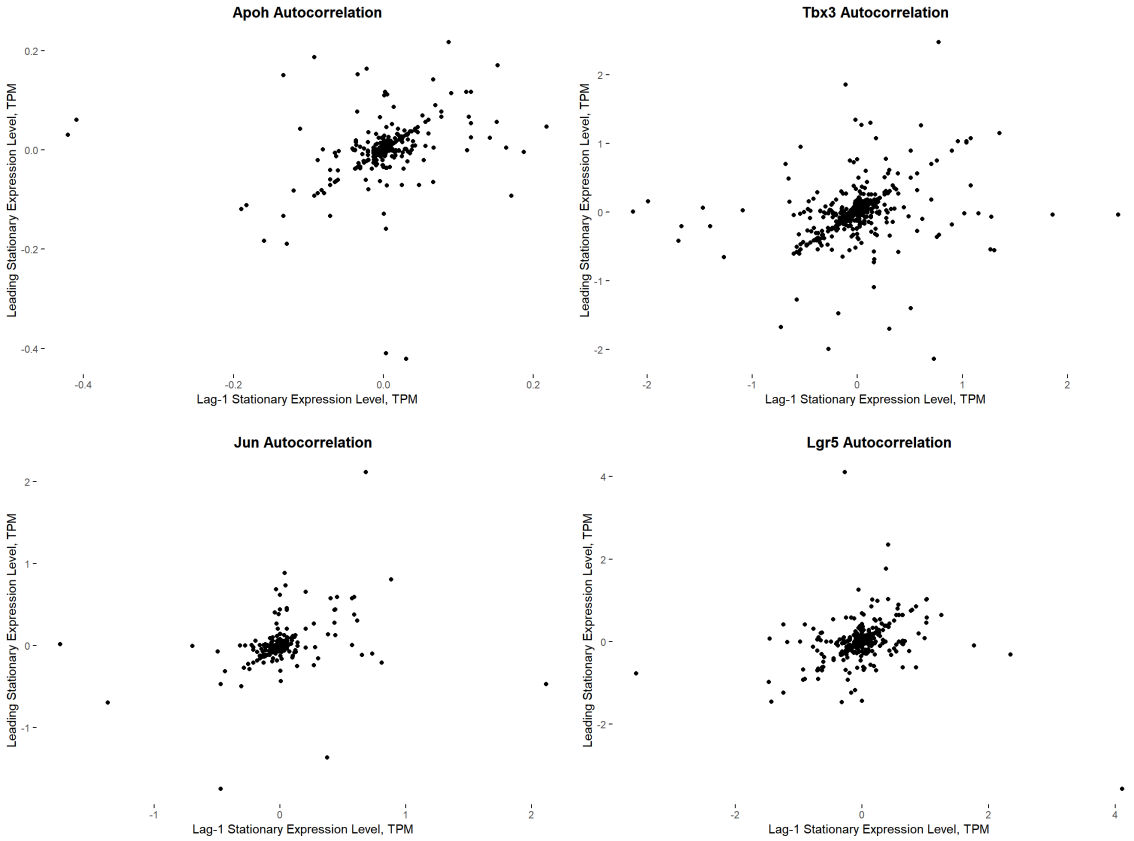
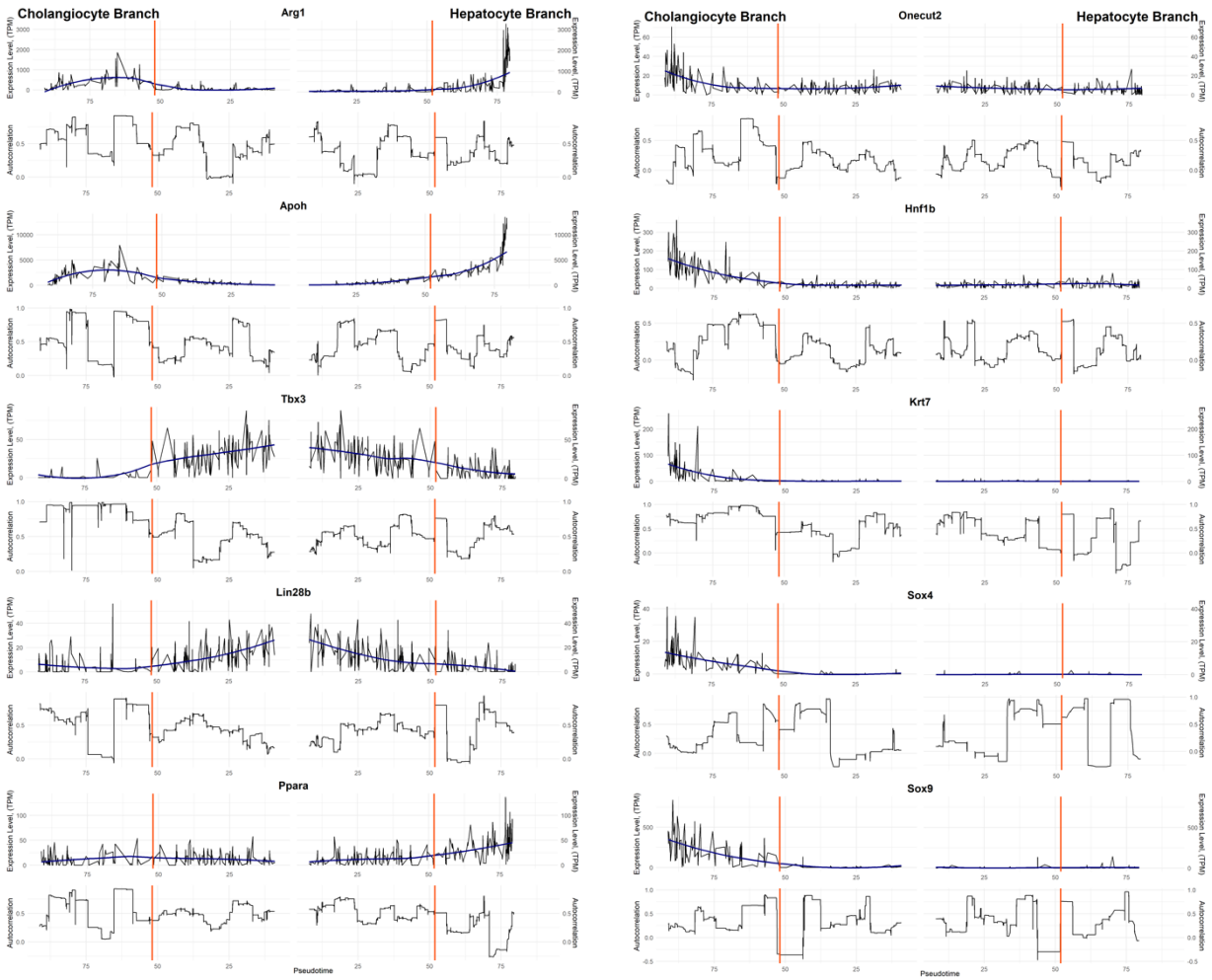


Figure 3: Static Autocorrelation Levels

Each of the plots contain a correlation between the leading and lagging expression levels. This is indicative of a trend being present prior to the bifurcation point in these expression levels. The specific span of points in pseudotime where the expression levels follow a linear trend could be the same for each of the stationary expression levels. Further addition of another dimension to **Figure 3** can be done to reveal similarities; further filtering of each of the 27 stationary gene expression levels could be done to find commonalities of strong linear correlations between lags and a specific span of pseudotime.

Dynamic trends of autocorrelation throughout the stationary expression levels are shown in **Figure 4** to visualize behavior of successive observations throughout different points in pseudotime. The



a.) Hepatoblast-Hepatocyte Genes

b.) Hepatoblast-Cholangiocyte Genes

Figure 4: Gene Expression Levels and Dynamic Measurement of Autocorrelation[3]

purpose of **Figure 4** is to reveal specific ranges in pseudotime where stationary expression levels contain a rise in autocorrelation. **Figure 4 a** and **b** depict expression levels and measured autocorrelation values for genes active in the hepatoblast-hepatocyte and hepatoblast-cholangiocyte differentiation process, respectively. The 10 out of 27 genes were picked based on which had characteristic fluctuations in autocorrelation; **Figure S2** displays all 27 gene expression levels and autocorrelation values.

Autocorrelation values fluctuate before and after the bifurcation point, (shown by the orange line). Autocorrelation is represented by the respective lag that is representative of the autoregressive model fit to the levels, (see markdown documentations for ARMA model fitting of each of the 27 genes). A general increase in autocorrelation values is shown between pseudotime values of 30

[3] Asymmetries in graph are due to linear interpolation. The smoothened purple line generated for the expression levels were generated by using Locally Weighted Scatterplot Smoothing, (LOESS), with a formula of y~x.

and 38. The amount increased varies from gene to gene, with some having more of a dramatic increase rather than others.

Conclusion & Future Work

Gene expression values were dimensionally reduced and isolated to identify up and down regulated trends in gene expression levels throughout the differentiation process. The dimensionality reduction uses reverse graph embedding showed similar differentiation patterns obtained by PCA [29]. Further work integrating two similar experiments using scRNA-seq data of embryonic mouse liver cells could be done by using data found in [28]. Integration of liver data from different zones in an adult mouse liver could also be done so show parallels in development and zonation [31].

Autocorrelation was shown to be prevalent in stationary gene expression levels prior to the transition point statically and dynamically. Further filtering of static gene expression levels between pseudotime values of 30 and 38 should be done to identify a minimum r^2 value and slope that represents an indicator of an incipient bifurcation. Statistical tests should be ran to test a dimensionally-reduced cluster of genes, (i.e. genes clustered from heatmap that are up-regulated in hepatoblast-hepatocyte branch), have a minimum slope and r^2.

A more automated algorithm that fits ARMA models and determines dynamic autocorrelation values could be employed to mine gene expression levels that contain characteristics of incipient bifurcation. The potential algorithm would have the potential to identify predictive markers that would indicate a developmental, (or health-to-disease), transition is about to occur based on scRNA-seq data obtained. The algorithm would be able to detect a sudden increase in autocorrelation and run statistical tests that are able to identify a correlation between the increase in autocorrelation to the observed critical transition. The algorithm would be able to be used on other datasets based on the parameters used in the scRNA-seq data used from the embryonic mouse hepatic cells.

Further simulations that quantify the steady-states of gene expression levels throughout hepatic development should be done to simulate the epigenetic landscape [2]. A model that represents changes in clustered genes that significantly change throughout development can be made to clearly identify what is considered a critical transition. Characteristics of the models could be used to identify what is considered a critical transition in gene expression in different scRNA-seq datasets.

Methods

Autoregressive Model Fitting & Autocorrelation Measurement

An autoregressive model was fit to each of the stationary gene expression levels based on visual inspection of their autocorrelation functions. The gene expression levels were made stationary by taking the 1st difference of the TPM values. The coefficients for the autoregressive model in the dynamic measurement of autocorrelation was found by using the acf() function in *R* which is a part of the "stats" package. The scedasticity for each fitted model was found by fitting stationary expression values using the arma() function which is a part of the "tseries" package in *R*. The determination of ARMA models via visual inspection is shown in the ARMA files in the Markdown Documentation.

The dynamic measurement of autocorrelation was taken by sliding a window of 10% of the grid size which moved in increments of the number of lag coefficients in the autoregressive model. For example, if an AR(2) model was fit to the stationary expression levels, the window would move in increments of 2 data points. The number of iterations was found by the line of pseudocode below

$$\#\ \text{of Iterations} = (\text{length}(\text{ExpressionLevels}) - \text{d} - \text{WindowLength})\%/\%\text{p}, \quad \{4\}$$

Where d represents the number of differences (which is 1 in all cases), %/% represents integer division, and p represents the number of autoregressive coefficients fit to the stationary expression levels.

The static plots of the stationary expression levels were constructed by trimming data points by how many autoregressive coefficients are used to fit the stationary expression levels. The leading and lagging expression levels were found by executing the lines of pseudocode below.

$$\text{Leading} = \text{ExpressionLevels}[(1 + \text{p}): \text{length}(\text{ExpressionLevels})], \quad \{5\}$$

$$\text{Lagging} = \text{ExpressionLevels}[1: (\text{length}(\text{ExpressionLevels}) - \text{p})], \quad \{6\}$$

The expression levels represented above span from the beginning of pseudotime until the bifurcation point at a pseudotime value of 51.9. Each of the 27 genes were inspected and the top 4 best correlations were presented.

Sliding Window

The expression levels in each branch of pseudotime were interpolated and analyzed separately. Each branch of pseudotime was separated by the corresponding hepatocyte and cholangiocyte state designation assigned to each cell. The hepatocyte branch of pseudotime was isolated by filtering in the cells that only contained the hepatoblast and hepatocyte state assignment, and the cholangiocyte branch was done similarly. The bifurcation point was defined as the maximum pseudotime value that a hepatoblast existed which was 51.90215.

Each branch was interpolated to 10,000 evenly spaced data points using the user defined function createlineargrid() which calculates the interpolated expression by the following formula:

$$E_i = E_{r1} + \frac{(E_{r2} - E_{r1}) * (t_i - t_{r1})}{t_{r2} - t_{r1}}, \quad \{7\}$$

Where E_i is the interpolated expression value, E_{r2} is the real expression value that is associated with the larger real pseudotime value t_{r2}, E_{r1} is the real expression level that is associated with the smaller real pseudotime value t_{r1}, and t_i is the interpolated pseudotime value. The function makes use of the linspace() function a part of the pracma package in *R*.

The expression levels of 27 expressed genes were interpolated to be evenly spaced. A sliding window 1/10 of the interpolated grid size was used to measure autocorrelation of the expression values that were differenced twice. **Figure 5** below depicts the sliding window.

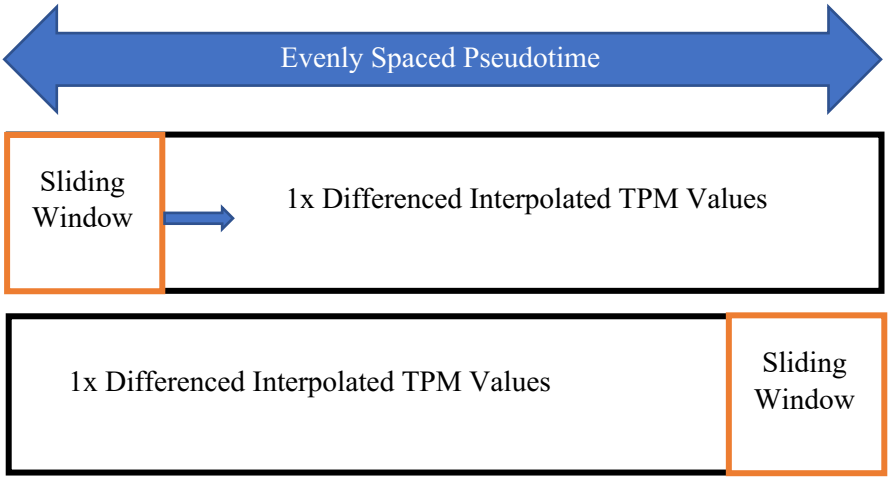


Figure 5: Sliding Window Depiction

The sliding window was iterated in p increments throughout each branch of pseudotime and autocorrelation was measured the autocorrelation value at the p lag. See **ARMA** files in the markdown documentation.

Dimensionality Reduction

Monocle 2 was used to reduce the dimension of single-cell RNA sequence data set available on the NCBI's Gene Expression Omnibus, (GEO), accession number GSE90047 [6] [7]. The TPM values of the 447 *in vivo* cells and 16738 expressed genes were used. A gene was determined to be expressed by having at least 1 TPM in 10 of the cells, and the ERCC spike-ins were removed. The gene list of expressed genes is shown in table S1. After implementing DDRTree, each of the cell states were deduced by looking at jitter plots that displayed genes known to be active in hepatoblasts, hepatocytes, and cholangiocytes [7]. Pseudotime values were assigned to each cell, and the matrix of cells was organized in ascending chronological order. All expressed genes were clustered using the methods like the plot_genes_branched_heatmap() function in the Monocle package, which implements the hierarchical clustering algorithm ward.D2 [9]. The supplementary methods reveal the data processing steps prior to the dimensionality reduction and gene clustering.

Monocle's DDRTree was used to produce the two-component cellular manifold. The algorithm starts off by log transforming the data,

$$E_{log} = \log_{10}(TPM + 1), \quad \{8\}$$

Where E_{log} is the log-transformed expression and TPM is the expression level. 1 is added to prevent infinite values. The log-transformed data is then z-score scaled by the following expression

$$E_{scale} = \frac{E_{log} - mean(E_{log})}{\text{standard deviation}(E_{log})}, \quad \{9\}$$

The scaled expression matrix is multiplied by its transpose and then is fed into the "irlba" package in *R* which computes the singular value decomposition of the product. The right singular vectors of the matrix are returned. The right singular vectors are transposed and multiplied by the expression matrix, which yields the cell embeddings values for principal component 1 and 2. The expression matrix, cell embeddings which act as initial values for both sets of latent points, and the parameter values of K and lambda, and other parameter values are passed on to the DDRTree_reduce_dim() exported C function [47]. Code that implements the reduceDimension() function is show in **Dim_Red** in the markdown documentation.

Gene Clustering

Genes were clustered using the plot_genes_branched_heatmap() function in the Monocle package [48]. The function utilizes the pheatmap package, and the expression data is altered prior to its implementation. The following steps are done to the expression data prior to the implementation of the ward.D2 clustering algorithm [49]:

1. Interpolation of expression data in each branch
2. Spline fit to data using VGAM package with three degrees of freedom
3. Data is log transformed by Log_{10}(Fitted Expression+1)
4. A new matrix is arranged so that the ends of pseudotime are on the outside of the matrix and the beginning is on the inside. The expression values on the leftmost side are the ending values for the hepatocyte branch of pseudotime and the expression values on the rightmost side are ending values for the cholangiocyte branch.
5. Data is z-scored scaled for each expression trajectory along both branched by (Log Expression – mean(Log Expression)/standard deviation(Log Expression)
6. The distance between each gene is found by 1-r/2, where r is the correlation coefficient between the scaled expression levels of each gene.

The scaled expression levels as well as the their distance is used in the pheatmap() function in R which passes along arguments to the hclust() function in the *R* package stats. The ward.D2 is a hierarchical clustering algorithm that arranges each gene that is the most like each based on their scaled branch expression level. It was found that similar results can be obtained using kmeans clustering if the data was processed as stated above. The code for the heatmap implementation is shown in **Dim_Red** in the markdown documentation.

References

[1] C. Waddington, An introduction to modern genetics, London: George Allen & Unwin, 1939.

[2] S. Bhattacharya, Q. Zhang and M. E. Andersen, "A deterministic map of Waddington's epigenetic landscape for cell fate specification," *BMC Systems Biology,* vol. 5, no. 1, p. 85, 27 May 2011.

[3] J. Slack, "Conrad Hal Waddington: the last renaissance biologist?," *Nat Rev Genet,* vol. 3, no. 11, pp. 889-895, 2002.

[4] R. Thom, Structural Stability and Morphogenesis, Reading: W.A. Benjamin, Inc, 1976.

[5] R. Thom, "An Inventory of Waddingtonian Concepts," in *Therical BIology: Epigenetic and Evolutionary Order from Complex Systems*, Edinburgh University Press, 1989, pp. 1-7.

[6] S. Huang, "Reporogramming cell fates: reconciling rirty with robustness," *Nat Rev Mol Cell Biol,* vol. 31, no. 5, pp. 546-560, 2009.

[7] J. Wang, L. Xu, E. Wang and S. Huang, "The Potential Landscape of Genetic CIrcuits Imposes the Arrow of Time in Stem Cell Differentiation," *Biophys J,* vol. 99, no. 1, pp. 29-39, 2010.

[8] J. H. S. Zhou, "Understanding gene circuits at cell-fate branch points for rational cell reprogramming," *Trends Genet,* vol. 27, no. 2, pp. 55-62, 2011.

[9] P. S. L. Angerer, S. Tritschler, F. A. Wolf, D. Fischer and F. J. Theis, "Single cells make big data: New challenges and opportunities in transcriptomics," *Current Opinion in Systems Biology,* vol. 4, pp. 85-91, 2017.

[10] E. Callaway, "The Trickiest Family Tree in Biology," *Nature,* vol. 547, pp. 20-22, 2017.

[11] X. Zhang and N. Yosef, "A new way to build cell lineages," *eLife,* vol. e25654, no. DOI: 10.7554/eLife.25654, pp. 1-4, 2017.

[12] P. Kumar, Y. Tan and P. Cahan, "Understanding development and stem cells using single cell-based analyses of gene expression," *The Company of Biologists,* vol. 144, no. doi:10.1242/dev.133058, pp. 17-32, 2017.

[13] A. Nowogrodzki, "The Cell Seeker," *Nature,* vol. 547, pp. 24-26, 2017.

[14] A. Tanay and A. Regev, "Scaling single-cell genomics from phenomenology to mechanism," *Nature,* vol. 541, pp. 331-338, 2017.

[15] O. B. Poirion, X. Zhu, T. Ching and L. Garmire, "Single-Cell Transcriptomics Bioinformatics and Computational Challenges," *Frontiers in Genetics,* vol. 7, no. 163, pp. 1-11, 2016.

[16] L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner and F. J. Theis, "Diffusion pseudotime robustly reconstructs lineage branching," *Nature Methods,* vol. 13, no. 10, pp. 845-847, 2016.

[17] Z. Ji and H. Ji, "TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis," *Nucleic Acids Research,* vol. 44, no. 13, p. e117, 2016.

[18] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner and C. Trapnell, "Reversed graph embedding resolves complex single-cell trajectories," *Nature Methods,* vol. ADVANCE ONLINE PUBLICATION, no. doi:10.1038/nmeth.4402, pp. 1-4, 2017.

[19] M. T. M. D. Setty, S. Reich-Zeliger, O. Angel, T. M. Salame, P. Kathail, K. Choi, S. Bendall, N. Friedman and D. Pe'er, "Wishbone identifies bifurcating developmental trajectories from single-cell data," *Nature Biotechnology,* vol. 34, no. 6, pp. 637-645, 2016.

[20] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen and J. L. Rinn, "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells," *Nature Biotechnology,* vol. 32, no. 4, pp. 381-386, 2014.

[21] Y. Deng, F. Bao, Q. Dai, L. F. Wu and S. J. Altschuler, "Scalable analysis of cell-type composition from single-cell transcriptomics using deep recurrent learning," *Nature Methods,* no. https://doi.org/10.1038/s41592-019-0353-7, 2019.

[22] A. Frishberg, N. Peshes-Yaloz, O. Cohn, D. Rosentul, Y. Steuerman, L. Valadarsky, G. Yankovitz, M. Mandelboim, F. A. Iraqi, I. Amit, L. Mayo, E. Bacharach and I. Gat-Viks, "Cell composition analysis of bulk genomics using," *Nature Methods,* no. https://doi.org/10.1038/s41592-019-0355-5, 2019.

[23] A. Giladi and I. Amit, "Immunology, one cell at a time," *Nature,* vol. 547, pp. 27-29, 2017.

[24] M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. Chattopadhyay, H. Swerdlow, R. Stija and P. Smibert, "Simultaneous epitope and transcriptome measurement in single cells," *Nature,* vol. 14, no. 9, pp. 865-868, 2017.

[25] S. Nestorowa, F. Hamey, B. P. Sala, E. Diamanti, M. Shepherd, E. Laurenti, N. Wilson, D. G. Kent and B. Gottgens, "A single-cell ressolution map of mouse hematopoietic stem and progenitor cell differentiation," *Blood,* vol. 128, no. 8, pp. e20-e31, 2016.

[26] A. Olsson, M. Venkatasubramanian, V. K. Chaudhri, B. J. Aronow, N. Salomonis, H. Singh and H. L. Grimes, "Single-cell analysis of mixed-lineage states leading to a binary cell fate choice," *Nature,* vol. 537, pp. 698-702, 2016.

[27] B. Sala-Pijuan, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, T. R. C. V., S. Srinivas, B. D. N. J. Simons, J. Marioni and B. Göttgens, "A single-cell molecular map of mouse gastrulation and early organogenesis," *Nature,* vol. 566, pp. 490-495, 2019.

[28] X. Su, Y. Shi, X. Zou, Z. Lu, G. Xie, J. Y. H. Yang, C. Wu, X. H. K. Cui, Q. Luo, Y. Qu, N. Wang, L. Wang and Z. Han, "Single-cell RNA-Seq analysis reveals dynamic trajectories during mouse liver development," *BMC Genomics,* vol. 18, no. 946, pp. 1-14, 2017.

[29] L. Yang, W.-H. Wang, W.-L. Qiu, Z. Guo, E. Bi and C.-R. Xu, "A Single-Cell Transcriptomic Analysis Reveals Precise Pathways and Regulatory Mechanisms Underlying Hepatoblast Differentiation," *Hepatology,* vol. 66, no. 5, pp. 1387-1401, 2017.

[30] C. Gerard, J. Tys and F. Lemaigre, "Gene regulatory networks in differentiation and direct reprogramming of hepatic cells," *Seminars in Cell & Developmental Biology,* vol. 66, no. June, pp. 43-50, 2017.

[31] K. Halpern, R. Shenhav, O. Matcovitch-Natan, T. Beata, D. Lemze, M. Golan, E. Massasa, S. Baydatch, S. Landen, A. Moor, A. Brandis, A. Giladi, A. Stokar-Avilhail, E. Dvaid, I. Amit and S. Itzkovitz, "Single-cell spatial reconstruction reveals global division of labour in the mammalian liver," *Nature,* vol. 542, pp. 352-356, 2017.

[32] R. Nault, K. A. Fader, J. R. Harkema and T. Zacharewski, "Loss of liver-specific and sexually dimorphic gene expression by aryl hydrocarbon receptor activation in C57BL/6 mice," *PLoS ONE,* vol. 12, no. 9, 2017.

[33] G. Gentric, C. Desdouets and S. Celton-Morizur, "Hepatocytes Polyploidization and Cell Cycle Control in Liver Physiopathology," *International Journal of Hepatology,* vol. 2012, no. doi:10.1155/2012/282430, pp. 1-8, 2012.

[34] Q. Mao, L. Wang, I. Tsang and Y. Sun, "Principal Graph and Structure Learning Based on Reversed Graph Embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 11, pp. 2227-2241, 2017.

[35] S. Bisgaard and M. Kulahci, Time Series Analysis and Forecasting By Example, Hoboken: John Wiley & Sons, Inc., 2011.

[36] R. Shumway and D. Stoffer, Time Series Analysis and Its Applications, Gewerbestrasse: Springer Nature, 2017.

[37] W. Navidi, Principles of Statistics for Engineers and Scientists, New York: McGraw Hill, 2010.

[38] V. Dakos, M. Scheffer, E. van Nes, V. Brovkin, V. Petoukhov and H. Held, "Slowing down as an early warning signal for abrupt," *PNAS,* vol. 105, no. 38, pp. 14308-14312, 2008.

[39] V. N. Livinia and T. Lenton, "A modified method for detecting incipient bifurcations in a dynamical," *Geophysical Research Letters,* vol. 34, no. L03712, pp. 1-2, 2007.

[40] T. J. and S. Jan, "Climate tipping as a noisy bifurcation: a predictive technique," *IMA Journal of Applied Mathematics,* no. 76, pp. 27-46, 2010.

[41] D. Henk, Nonlinear Climate Dynamics, New York: Cambridge University Press, 2013.

[42] E. van Nes and M. Scheffer, "Slow Recovery from Perturbations as a Generic Indicator," *The American Naturalist,* vol. 169, no. 6, pp. 738-747, 2007.

[43] J. Thompson and J. Sieber, "Predicting Climate Tipping as a Noisy Bifurcation: A Review," *International Journal of Bifurcation and Chaos,* vol. 21, no. 2, pp. 399-423, 2011.

[44] H. Held and T. Kleinen, "Detection of climate system bifurcations by degenerate fingerprinting," *Geophysical Research Letters,* vol. 31, no. L23207, 2004.

[45] M. Scheffer, J. Bascompte, W. Brock, V. Brovkin, S. Carpenter, V. Dakos, H. Held, E. van Nes, M. Rietkerk and G. Sugihara, "Early-warning signals for critical transitions," *Nature,* vol. 461, pp. 53-59, 2009.

[46] J. Berg, J. Tymoczko, G. Gatto and L. Stryer, Biochemistry, New York: W.H. Freeman and Company, 2015.

[47] X. Qiu, "cole-trapnell-lab/ DDRTree," March 2018. [Online]. Available: https://github.com/cole-trapnell-lab/DDRTree/blob/master/R/DDRTree.R. [Accessed 09 August 2018].

[48] X. Q. C. Trapnell, "Monocle," 2018. [Online]. Available: http://cole-trapnell-lab.github.io/monocle-release/docs/#acknowledgements.. [Accessed 09 August 2018].

[49] F. Murtagh and P. Legendre, "Ward's Hierarchical Clustering Method:," 13 December 2011. [Online]. Available: https://arxiv.org/pdf/1111.6285.pdf. [Accessed 09 August 2018].

[50] Q. M. Y. T. L. W. R. C. H. A. P. C. T. Xiaojie Qiu, "Reversed graph embedding resolves complex single-cell trajectories," *Nature Methods,* vol. 14, no. 10, pp. 979-982, 2017.

[51] E. Zivot, "Unit Root Tests," 05 April 2005. [Online]. Available: https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf. [Accessed 09 August 2018].