Modeling Cell Fate Differentiation as a State-Based Process

Shankara Anand
Department of Chemical Engineering
Stanford University

Varun Gupta
Department of Computer Science
Stanford University

CS 221 Fall 2016

ABSTRACT: Stem cell fate differentiation is a complex, biological process with vast medical benefits in predictive modeling. Advancements in cellular assays over the past decade have provided researchers with comprehensive data-sets for single-cell gene expression analysis, where one can isolate specific cells in a heterogeneous set of stem cells and elucidate genetic factors responsible for varying states of differentiation. Using state-based models, we attempt to show shortest/optimal fate differentiation paths for stem cells by modeling each cell as a state to predict fate progression solely through genetic markers. We successfully implemented uniform cost search and a Markov decision process, thus finding sophisticated state-based models for mapping cell differentiation.

1 INTRODUCTION

1.1 *Background*

One of biology's most complex and ubiquitous phenomena is early stem cell fate differentiation. Cell fate is mediated by an interplay of both environmental factors and intracellular components, particularly gene transcription factors.[1] In particular, Hematopoietic Stem Cells (HSCs) are multipotent cells capable of differentiation into both lymphoid and myeloid blood cells.[2] Remarkably, these cells are capable of self-renewal, meaning during cell division, they are able to generate at least one daughter cell with the potential to become an HSC. The robust nature of HSCs allows them to maintain stability across vast, circulatory networks and has made them the most successful candidates for clinical stem cell therapy to date.[2] Thus, there exists a strong interest in better understanding HSC differentiation given a set of existing transcription factor data for better targeted medical applications.

Previously, scientists have approached cell differentiation through bulk assays, which provide insight on average cell population characteristics. This data is then extrapolated to describe individual cell behavior. However, cells are often at varying stages of differentiation within a given population, meaning a homogeneous approach is not entirely accurate.[3] Specifically, information for rare populations, such as cells early on in the differentiation process, are often masked by population averages studied in assays. Therefore, relevant data informing cell fate decisions is better modeled at an individual cell level.[3]

In the past decade, microfluidic high-throughput qRT-PCR has made it possible to isolate individual cells and collect single-cell gene expression data.[4] This data essentially maps a given cell in a population to transcription factor levels for a set of target genes. Statistical analysis of such data has provided insights into interactions among individual transcription factors and how they may determine cellular fate. For example, unsupervised analysis of mouse embryonic stem cells suggests regulatory networks in early stage blood-cell development.[5,6] Our aim in this project is to account for the temporal nature of cell division by modeling each single cell in our data set as a state and observing path propagation to a final, differentiated state.

1.2 *Task Definition*

We seek an optimal path to a given cell fate for mouse HSCs using a single-cell gene expression dataset collected and studied by Moignard et al.[6] While

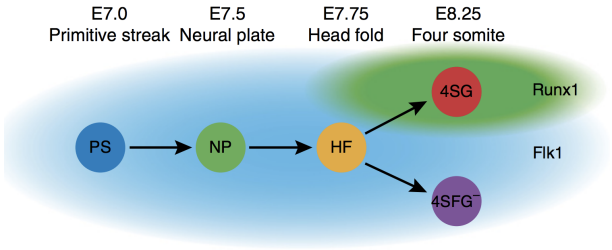

Figure 1: **Cell Differentiation Classifications:** The above pathway, described by Moignard et al, demonstrates three ordinal stages of development for mouse embryonic cells: primitive streak (PS), neural plate (NP), and head fold (HF). From HF, cells may differentiate into 4SG, erythroid fate, or 4SFG-, endothelial fate.

Moignard et al used both biological assays and statistical single-cell gene expression analyses to cluster HSCs and identify gene transcription factors that control differentiation, we use the state-based models - particularly a search problem and a Markov decision process - to model cell differentiation over time. Because there does not exist a biological assay to track gene expression changes through cell differentiation, our task is to model this development as a path an undifferentiated, or progenitor, cell takes to its final, determined fate.

The original data-set used by Moignard et al[6] contains 3935 mouse embryonic stem cells belonging to five different cell classifications: 624 cells of type PS, 552 NP, 1005 HF, 983 4SG, and 770 4SFG$^-$ **(figure 1)**, which represent broad cell types corresponding to different stages of stem cell maturation.[6] (*Note: 4G and 4GF are also used to denote erythroid and endothelial cell fates throughout this paper*) The normalized expression levels of 46 genes are reported for each stem cell in the data-set. Using this as an outline, we determined a state-based model, with individual cells as states, would be appropriate in modeling our system.

Note that Moignard et al identify and reproduce known regulatory networks of particular transcription factors that control endothelial fates (Cdh5, Erg, HoxB4, Sox7, Sox17) and erythroid fates (Gata1, Gfi1b, Hbbbh1, Ikaros, Myb, Nfe2). To illustrate these known regulatory networks, we use the Wilcoxon signed-rank test to estimate the quantitative difference in expression of specific genes across time periods (and thus cell types) **(figure 2)**. For example, for gene Sox17, we use the Wilcoxon signed-rank test to compare the expression levels in PS and NP, NP and HF, HF and 4SGA, HF and 4SFGA, and 4SGA and 4SFGA. Thus, for each gene, we used the Wilcoxon signed-rank test 5 times. With 5 tests for each of 46 genes, we ran 230 tests overall.

Clearly, the Wilcoxon analysis confirms model predictions of which cell types exhibit increased expression of which genes. For instance, the transition from HF to 4G, the endothelial progenitor, is marked by increases in the Ets family (Ets, Etv, and Erg) and

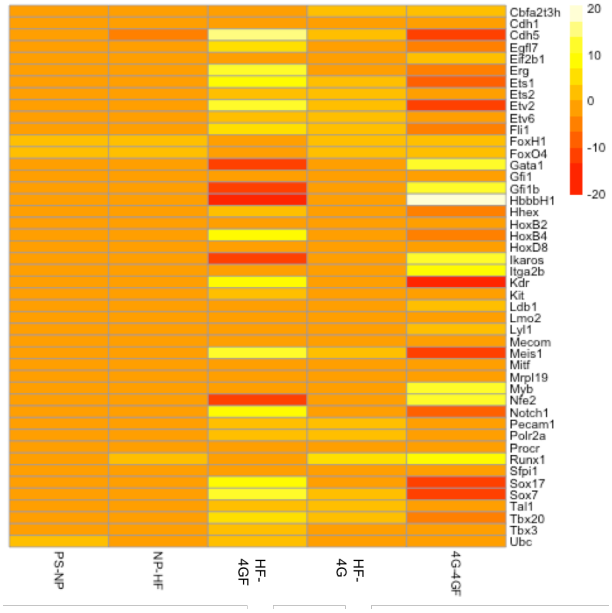

Figure 2: **Wilcoxon Signed-rank Test:** this heat map compares expression levels over mean cell populations.

the cadherins (responsible for proteins that bind endothelial tissue together), as well as HoxB4, Notch1, and the Sox factors. Not only that, but erythroid genes are suppressed, including the zinc finger erythroid proteins of Gata1, Nfe2, and Gfi1b, as well as HbbbH1, which is responsible for hemoglobin, and Ikaros, which is critical to proper erythroid development in mice. We evaluate state-based models by testing whether computed optimal paths and policies reproduce, at the very least, these known regulatory networks and transcription dynamics.

2 APPROACH TO PROBLEM

We approach this problem with two state-based models: a search problem and a Markov decision process (MDP). For both models, we use the same framework for state definition. Because no temporal assay exists for tracking an individual cell through its fate differentiation, we model each cell in the data-set as a state into which any cell can enter at a point in time during its differentiation. Therefore, each state will hold information about its gene expression profile and original cell type, with the assumption that changes in gene expression are indicative of cell determination towards a terminal state. States have the following form:

$$S_i = ([g_{1i}, g_{2i}, ..., g_{ji}, ..., g_{46i}], X_i)$$

$$for\ i = 1,2...3935$$

g_{ij} is the expression level of the j^{th} gene in the i^{th} cell, $X_i \in \{PS, NP, HF, 4SG, 4SFG^-\}$. We will now demonstrate how these states are used in our two implementations.

Cell	Cbfa2t3h	Cdh1	Cdh5	Egfl7
HFA1_001	18.03714015	25	13.62318312	15.53445242
HFA1_002	25	25	25	19.69280303
HFA1_003	25	25	25	21.03942464
HFA1_004	25	25	25	25
HFA1_005	19.32622431	25	15.39946959	16.13481182
HFA1_007	25	25	25	20.26317394
HFA1_008	18.95254452	25	25	18.10729015
HFA1_009	25	19.83847404	25	25
HFA1_010	17.05357252	25	12.01550254	15.63221166

Figure 3: **Single Cell Gene Expression Data:** The above is a portion of our raw data used for our cell fate differentiation modeling. Each cell is mapped to a set of K genes with normalized expression levels.

2.1 *Search Problem*

We first model cell differentiation as a search problem and use the Uniform Cost Search algorithm (UCS) to find shortest paths to differentiated fates for each progenitor cell. Through search, we deterministically explore every single path from a progenitor cell to a terminal fate, which involves either an erythroid or endothelial cell fate. Although this approach is simpler from a modeling perspective, it is computationally intensive because UCS samples the entire state space for every progenitor cell to find its shortest path. Additionally, the search problem model assumes differentiation is deterministic.

Start State: Every possible progenitor cell (PS, NP, HF).

Successors/Actions: All possible cell states accessible by one evolution from a given cell. For the given cell types, the following actions are possible regardless of its gene expression profile. Below are the mappings of current states to accessible successors.

PS $\rightarrow$ PS, NP
NP $\rightarrow$ NP, HF
HF $\rightarrow$ HF, 4SG, 4SFG$^-$

Cost: We model the cost for entering a successor state in the search problem using a distance metric representing genetic expression variation between the current state and the successor state. Intuitively, the greater variation in the gene expression profiles of a cell A and a cell B, the higher cost this model assigns for exploring that successor state.

Distance Metric
We use a high-dimensional Euclidean distance for costs in the search problem. The data used by Moignard et al[6] provides gene expression levels of 46 genes for each of the 3935 single cells **(figure 3)**. By computing the Euclidean distance across 46 dimensions (genes) for a pair of cells based on normalized expression values, we obtain a primitive metric for cell similarity that we use to compute transition cost

$$Cost(cell_1), cell_2) = \sqrt{\frac{(g_{11}-g_{12})^2}{\alpha_1} + \frac{(g_{21}-g_{22})^2}{\alpha_2} + ... + \frac{(g_{K1}-g_{K2})^2}{\alpha_K}}$$

A

$$\alpha = \frac{(\sigma_x)^2}{\sum_{j=1}^{K}(\sigma_j)^2}$$

B

Figure 4: **Cost Function and Weight Spread:** (**A**) This assigns a cost to an action based on genetic similarity. In this equation, we account for spread weights, α = 1,...K. (**B**) This is the mathematical equation we are using to compute our weight spread for a given gene, g_x, over all i=1,...K genes.

(figure 4A). One downside to this approach is that it homogeneously weighs each of the differences across 46 genes. For instance, if gene g_j were crucial to cell function and controls differentiation from cell type PS to NP, a change δ of this individual gene expression would be weighted equally with the same change δ of an arbitrary gene in our model that has little to no effect on this transition. Therefore, we weight distance contributions of individual genes through the following method using a "Weight Spread."

Weight Spread
To prevent weighting all of our genes equally in our system, we weight our distances using the standard deviation of expression levels for a gene in each cell in our data set **(figure 4B)**. The intuition for this is as follows: if the expression profile of a gene X has very little variance amongst all single cells in our data set, whenever we do have a change in expression for gene X during a transition, we can assume that it has a large effect on cell fate determination. To compute this weight spread value, we compute the standard deviation of our gene expression levels per gene, and then normalize it over all values. We proceed to incorporate this into our equation as a weight spread, α.

End State: We use an arbitrary termination state (END) that an erythroid or endothelial fate state ($4SG$, $4SFG^-$) will deterministically enter. This allows each progenitor to find the shortest distance to a given final state once entering our known terminal states.

2.2 *Markov Decision Process*

Our next implementation for modeling cell differentiation was a Markov decision process. The motivation behind using an MDP is that cell fate determination is, by nature, a stochastic process[6]. Therefore, for each possible transition from one cell to another, there is a chance that our cell will not evolve to a cell in the next differentiation stage and, instead, stay

within its given defined stage.

Start State: We will sample our full space by creating an arbitrary start state, START, that will deterministically transition to a PS cell. For this first state, there will be a uniformly random chance that it will choose a cell in PS, ideally allowing our MDP to sample all possible PS cells. Thus, we will have 1 start state that will then randomly enter 1 of 624 possible PS cell types.

Actions: In this MDP, actions represent the capabilities of a cell at a given time point from an evolutionary perspective. We model two such capabilities germane to cellular differentiation: stay and evolve.

Stay
A cell of type X_i may continue to be of type X_i. However, due to the stochastic and regulatory natures of gene expression, the expression profile of the cell must change, even if the type of the cell remains the same. We use transition probabilities within this state space to chose which precise change in genetic expression to undergo. Because in our raw data we only have gene expression data of K genes, we make the assumption that computing a transition probability based on this (see below) Thus, for a state S_i:

$$Succ(S_i) = \{S_j | X_j = X_i \; and \; j \; \neq i\}$$

If X_i = PS: 623 successors of type PS
If X_i = NP: 552 successors of type NP

Evolve
A cell of type X_i evolves to be of type X_{i+1}. Thus, our model defines evolution as moving from one cell type to the next. This means that when in a given state, if a single cell chooses evolve as an action, it commits to entering another cell type requiring the use of our transition probability metric **(figure 5)**. Thus, for a state S_i:

$$Succ(S_i) = \{S_j | X_j = X_{i+1}\}$$

A special case exists for states with the cell type that immediately precedes the final cell fates - in this case, HF. The action Evolve is replaced with Evolve to $4SG$ and Evolve to $4SFG^-$.

If X_i = PS: 553 successors of type NP
If X_i = NP: 1004 successors of type HF
If X_i = HF $\rightarrow 4SG$: 983 successors of type $4SG$
If X_i = HF $\rightarrow 4SFG^-$: 770 successors of type $4SFG^-$

Transition Probabilities: Transition probabilities in the MDP capture the stochasticity involved with changes in the gene expression profiles of HSCs.

$$T(cell_1, cell_2) = \frac{1}{\sqrt{\frac{(g_{11}-g_{12})^2}{\alpha_1} + \frac{(g_{21}-g_{22})^2}{\alpha_2} + ... + \frac{(g_{K1}-g_{K2})^2}{\alpha_K}}}$$

Figure 5: **Transition Probability:** This computes transition probability using genetic similarity as a likelihood of transitioning to a given cell state. In this equation, we account for spread weights, α = 1,...K.

For an MDP that models cell differentiation, the probability that one state transitions to another, or one cell modifies its gene expression profile to match that of another cell, will be based on a distance metric representing genetic expression variation between the two. The intuition here is that it is more likely for a cell A to match the gene expression profile of a cell B in a single time-step if its current gene expression profile is close to that of B.

Distance Metric
The distance metric we used for the preliminary implementation of the MDP is Euclidean distance as done for UCS **(figure 4A)**. By computing the Euclidean distance across 46 dimensions (genes) for a pair of cells based on normalized expression values, we obtain a primitive metric for cell similarity that we use to compute transition probabilities **(figure 5)**. Additionally, we also weight these probabilities as done in UCS **(figure 4B)**.

Rewards Function
Rewards for our MDP are based on a set of known genes outlined by Moignard et al[6] that were shown through both empirical and statistical analysis to be crucial to erythroid ($4SG$) or endothelial ($4SFG^-$) cell fates, denoted by *:

$4SG^* \rightarrow$ Gata1, Gfi1b, HbbbH1, Ikaros, Myb, Nfe2
$4SFG^{-*} \rightarrow$ Cdh5, Erg, HoxB4, Sox7, Sox17

To incorporate these known genes into our reward, we computed a metric of "Decisiveness" (**figure 6**).

Decisiveness
We define *decisiveness* as a state's preference to choose an erythroid or an endothelial fate. The intuition here is that our model will reward our cells for choosing an action that pushes them closer to a given cell fate. Scientifically, cells optimize for fate determination because it indicates an evolutionary genetic optimum. There are a number of factors such as environment, cell mass, and metabolic function that support cell growth which our model assumes push our progenitor cells towards final fate determination.[7,8] Therefore, for a given action, we take the weighted euclidean difference for each known erythroid and each known endothelial gene listed above (*) and sum these values for each category. Then, the difference in these differences is the reward assigned, meaning our new state is

$$|\sum_{i \in 4SG^*} dist(s(g_i) - s'(g_i)) - \sum_{j \in 4SFG^{-*}} dist(s(g_j) - s'(g_j))|$$

Figure 6: **Decisiveness:** The following computes a metric for rewarding a cell transitioning to a state towards differentiation. These metrics were developed using known transcription factors that were shown to correlate to erythroid or endothelial development, denoted by *: $4SG^*$ = {Gata1, Gfi1b, HbbbH1, Ikaros, Myb, Nfe2} and $4SFG^{-*}$ = {Cdh5, Erg, HoxB4, Sox7, Sox17}. In the above equation, s is our current state, $s^{'}$ is the successor, and g_x is a given gene for that state.

preferentially moving towards fate determination of either path.

End State

Cells reach an end state in our MDP when reaching a fully differentiated erythroid or endothelial cell fate ($4SG$ or $4SFG^{-}$).

3 RESULTS

To visualize the states and cell fate given 46 dimensions (genes), we use principle component analysis to project the data into the two most important dimensions of variability. We show in **figure 7A** the original HSC data categorized into stages of cell differentiation done by Moignard et al.[6] In **figure 7B**, we show a baseline implementation of clustering that characterizes all progenitor cells into erythroid and endothelial fates. We assign each progenitor cell (of types PS, NP, and HF) to an endothelial or erythroid fate based on distances to the centroid of known endothelial (4GF) and known erythroid (4G) cells. We make two observations: first, the baseline assigns cells based on a sharp decision boundary, leaving no room for more granular differentiation behavior; second, the baseline assigns a group of cells in the top left portion of the differentiation manifold to erythroid fates, despite the longer distance across the manifold between the two clusters. Clearly, the baseline classification scheme is problematic.

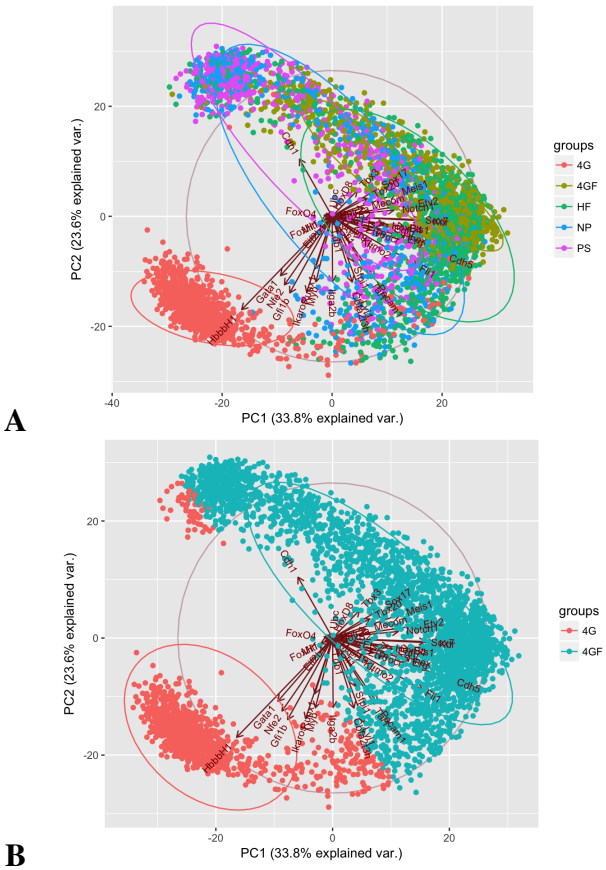

Figure 7: **Biplots of Data with Baseline Implementation:** (**A**) Biplot of original dataset provided by Moignard et al[6]. Centroids of each cell stage visualized by clusters shown. (**B**) Biplot of a baseline implementation for progenitor cell assignment. Shown are all cells assigned into either erythroid (4G) or endothelial (4GF) by euclidean distance to original erythroid/endothelial cluster centroids. Note the group of cells in the top of the manifold that are erroneously assigned erythroid fates.

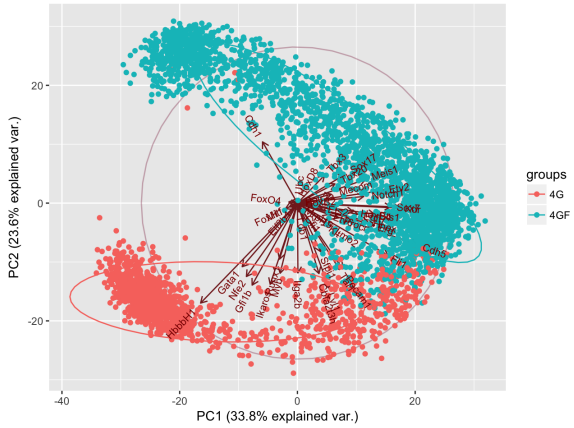

Figure 8: **Uniform Cost Search Implementation:** The above biplot shows cell fate assignments based on the shortest paths calculated for each progenitor to either an erythroid (4G) or an endothelial (4GF) fate by UCS.

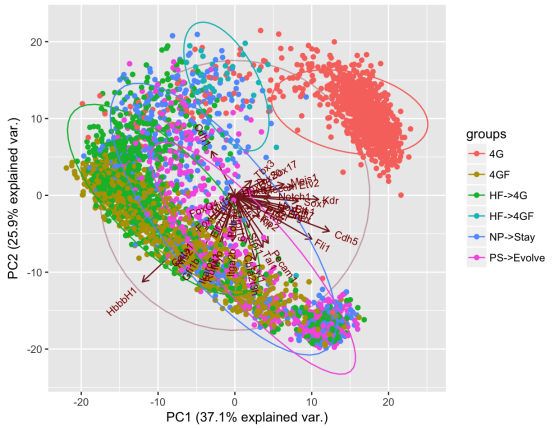

Figure 9: **Markov Decision Process Implementation:** The above biplot shows cell fate assignments based on the optimal policies found for the MDP model of cell differentiation. With optimal policies computed by value iteration, the above shows which cells will optimally stay in a given stage or transition.

3.1 *Uniform Cost Search Results*

With UCS, we compute the shortest path from each progenitor cell to a given cell fate (**figure 8**).

3.2 *Markov Decision Process Results*

With value iteration, we compute the optimal policy after roughly 20 iterations with a discount factor $\gamma = 0.9$ and tolerance $\epsilon = 1.0$ to ensure reasonable convergence. We display the optimal decision for each state. (**figure 9**).

4 DISCUSSION AND ANALYSIS

4.1 *Uniform Cost Search Analysis*

To compare the shortest paths computed by UCS to known regulatory networks and transcription dynamics, we identify the average transitions in gene expression levels for all cell states. We separate results into two categories: PS progenitor cells assigned to erythroid fates, and PS progenitor cells assigned to endothelial fates. For each category, we compute the average change in gene expression profiles between adjacent states (ex. PS $\rightarrow$ NP, NP $\rightarrow$ HF, etc.) and report these transitions in heat maps (**figure 10**).

We track genes of interest in our model by comparing how UCS chose its shortest path to known regulatory networks elucidated by Moignard et al. and earlier authors.

By looking at genes that have been shown to indicate erythroid fate determination (Gata1, Gfi1b, HbbH1, Ikaros, Myb, Nfe2)[6], we see that Gata1, Gfi1b, HbbbH1, Ikaros, Myb, Nfe2 all have lower average transition values in the NP to HF for cells that were eventually assigned to endothelial fates. For these same genes, we see higher average transition values for cells that were assigned to erythroid fates, confirming that our model implementation was successful in reproducing what was shown by Moignard et al.

If we look at our genes of interest that indicate endothelial fate determination (Cdh5, Erg, HoxB4, Sox7, Sox17), we see that Cdh5, Erg, and HoxB4 have a higher transition value from NP to HF, and Sox17 for HF to 4GF for cells that were eventually assigned to endothelial fates. This data corroborates network patterns elucidated in literature[6]. For cells eventually assigned to erythroid fates, we see that Sox7, HoxB4, and Cdh5 all have strong negative transition values at some point along the process, indicating they likely are being suppressed for the formation of erythroids.

Out of the above genes, Erg is a transcription factor of interest in that its activity is nearly identical in our model for erythroids and endothelial cells (**figure 10**). As mentioned previously, increases in the Ets family (Ets, Etv, and Erg) have been linked to endothelial fate determination. However, in our data for UCS, we see no tangible, qualitative difference between our PS cells that were eventually assigned erythroid fates or endothelial fates (**figure 10**). By inspection of the biplots in **figures 7A and 8**, we note that the progenitor PS, NP, and HF cells that are assigned an erythroid fate by UCS but also display increases in Erg transcription in the computed average transitions most likely do not enter into terminal erythroid states in the isolated cluster in the bottom left portion of the manifold (direction of decreasing Erg transcription), but rather into terminal erythroid states in the bottom right portion of the manifold (direction of increasing Erg transcription) obscured by other cell types.

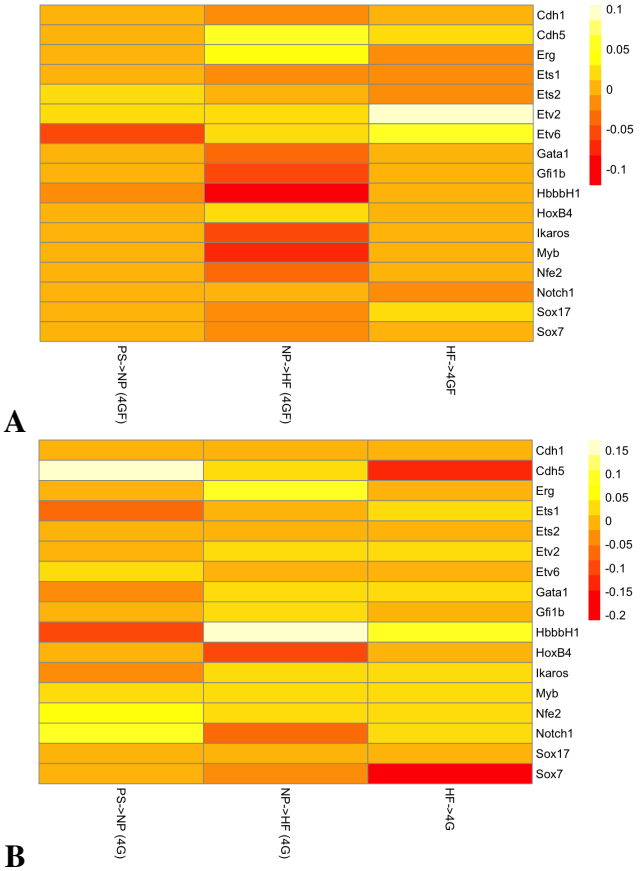

Figure 10: **Analysis of UCS Implementation:** (**A**) This heat map for our cells that have been assigned to endothelial fates (4GF) outlines the average transition change between cell stages per gene of interest. (**B**) Heap map for cells assigned to erythroid fates (4G).

4.2 *Markov Decision Process Analysis*

According to the optimal policy found through value iteration, all PS cells should evolve, all NP cells should stay, 973 HF cells should evolve into 4G, and only 32 HF cells should evolve into 4GF. These results are intuitively consistent with the decisiveness paradigm that we use for rewards. An inspection of the standard ellipsoids of each type of cell and type of decision (**figure 9**) indicates that NP cells exhibit the greatest variation in gene expression, and thus from a probabilistic standpoint, there is a greater possibility that an NP cell at one end of the manifold can transition to an NP cell at the other end of the manifold - a transition correlated with moving from one state that leans endothelial to another state that leans erythroid (or vice versa), which would manifest high decisiveness rewards. On the other hand, PS cells represent a smaller band of cells, and thus would almost always

benefit from evolving into any NP cell and then taking advantage of the aforementioned rewards.

The vast majority of the HF cells transition into endothelial fates (4GF). This result is actually consistent with the Wilcoxon analysis, which indicates that there are larger differences in gene expression for both erythroid genes (significant decreases) and endothelial genes (significant increases) between HF cells and 4GF cells than between HF cells and 4G cells (**figure 2**). In other words, as the Wilcoxon analysis indicates, transitioning into an endothelial state from the HF state requires a large "course correction" of sorts, whereas transitioning into an erythroid state from the HF state is a result of "staying the course" with respect to gene expression. By rewarding decisiveness, the MDP actually rewards course corrections, thus pushing cells to make evolutionary decisions that reward large transitions. Therefore, while decisiveness mirrors some aspects of the cellular differentiation in mouse embryonic HSCs, it over-rewards decisions that probabilistically result in aggressive changes to gene profile. This paradigm ascribes extreme variability over short periods of time to mouse embryonic development and thus does not reflect more measured cellular differentiation policies.

5 CONCLUSION

In conclusion, we demonstrate two possible state-based models for modeling cell fate differentiation. The search-based method with Uniform Cost Search produces average transition metrics that compare closely to to known regulatory networks of transcription factors described by Moignard et al. Although finding the shortest paths for each progenitor cell through UCS is computationally intensive, we find promising results for the genes of interest and visually demonstrate more reasonable differentiation behavior in the assigned biplot (**figure 8**) than that in the baseline (**figure 7B**). On the other hand, the Markov decision process is more difficult to adjust to the dynamics of cellular differentiation. Although we intuitively justify the rewards paradigm to favor decisiveness, the MDP and the resulting optimal policy highly favors aggressive changes to gene expression, which is less accurate in reproducing cell differentiation dynamics.

For future directions, our search-based approach may easily be equipped in future studies of single-cell gene expression analysis. It is capable of outlining expression level changes for genes of known interest that reflect experimentally determined transcription factors crucial for determination and thus may elucidate new target genes for testing. The MDP model requires more work in determining a suitable rewards function to promote granularity in the optimal policies. Currently, the model successfully models aggressive differentiation behavior; however, a more ideal MDP should promote a wider sampling of the state space.

6 REFERENCES

[1] Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell, 126(4), 663676. $http://doi.org/10.1016/j.cell.2006.07.024$

[2] Schroeder, T. (2010). Hematopoietic Stem Cell Heterogeneity: Subtypes, Not Unpredictable Behavior. Cell: Stem Cell, 6(3), 203207. $http://doi.org/10.1016/j.stem.2010.02.006$

[3] Moignard, V., & Gttgens, B. (2014). Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. BioEssays, 36(4), 419426. $http://doi.org/10.1002/bies.201300102$

[4] Citri, A., Pang, Z. P., Sdhof, T. C., Wernig, M., & Malenka, R. C. (2011). Comprehensive qPCR profiling of gene expression in single neuronal cells. Nature Protocols, 7(1), 118127. $http://doi.org/10.1038/nprot.2011.430$

[5] Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schtte, J., Calero-Nieto, F. J., Gttgens, B. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nature Cell Biology, 15(4), 363372. $http://doi.org/10.1038/ncb2709$

[6] Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., Gttgens, B. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nature Biotechnology, 33(3), 26976. $http://doi.org/10.1038/nbt.3154$

[7] Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. Current Opinion in Microbiology, 13(3), 344349. $https://doi.org/10.1016/j.mib.2010.03.003$

[8] Feist, A. M., Palsson, B. O., Edwards, J., Palsson, B., Dobzhansky, T., Edwards, J., Papin, J. (2016). What do cells actually want? Genome Biology, 17(1), 110. $https://doi.org/10.1186/s13059-016-0983-3$

7 GITHUB

`https://github.com/varunabgupta/fate`