Profiling *in vitro* Neurogenesis Trajectories Resolved at the Single-cell Level to Compare Retinoic Acid Titrations with a novel Nearest Neighbor URD Trajectory Tool (NNUTT)

By

Emily A. Puleo, Department of Biomedical Engineering
Kristen I. Fread, Ph.D. Candidate, Department of Biomedical Engineering
Eli R. Zunder, Ph.D., Department of Biomedical Engineering
Austin B. Keeler, Ph.D., Department of Biomedical Engineering
Corey M. Williams, Ph.D. Candidate, Department of Biomedical Engineering
Amy L. Van Deusen, Ph.D. Candidate, Department of Biomedical Engineering
Sarah M. Goggin, Ph.D. Candidate, Department of Biomedical Engineering

Word Count: 4187
Figure Count: 6
Equation Count: 0
Supplement Count: 0
Reference Count: 16

Approved: Eli Zunder Date 5/6/2021
Eli R. Zunder, Ph.D., U.Va. Biomedical Engineering

Profiling *in vitro* Neurogenesis Trajectories Resolved at the Single-cell Level to Compare Retinoic Acid Titrations with a novel Nearest Neighbor URD Trajectory Tool (NNUTT)

Emily A. Puleo[a,1], Kristen I. Fread[b], Eli R. Zunder[c], Austin B. Keeler[d], Corey M. Williams[b], Amy L. Van Deusen[b], Sarah M. Goggin[b]

[a] Fourth Year Biomedical Engineering Student, University of Virginia
[b] Ph.D Candidate, Zunder Lab, Department of Biomedical Engineering, University of Virginia
[c]Assistant Professor, Zunder Lab, Department of Biomedical Engineering, University of Virginia
[d]Ph.D. , Deppmann Lab, Department of Biology, University of Virginia
[1] Correspondence: eap3zr@virginia.edu

Abstract

An estimated 25 million Americans are affected by neurological diseases[1]. It is critical to understand abnormal neurogenesis pathways to better understand how and why these diseases develop. This project establishes a method to characterize and compare cell populations over time using high dimensional mass cytometry data and show the relationship between retinoic acid (RA) concentration and timing of neural development. RA-treated cells are expected to differentiate into neural progenitors[2]. Cell cultures with varying concentrations of RA (0 M, 1 nM, and 1 uM) were collected at Days 6 and 14, stained for antibodies, and run on a mass-cytometer to yield 41 protein measurements per cell. The UMAP force-directed algorithm was applied to the dataset creating graphs to visualize the high dimensional dataset. Leiden clustering identified cell populations based on marker similarity. UMAP analysis revealed unique cell populations specific to the 0 M RA condition and indicated a need to optimize the mass cytometry panel to include antibodies to measure markers of cardiomyocytes that were present in the cultures. The Nearest Neighbor URD Trajectory Tool (NNUTT) was developed to compare cell trajectories between cultures. NNUTT expands URD to map cells from different cultures to a common dendrogram to compare timing of population development along the trajectory. NNUTT calculates Euclidean distances between cell types and maps the closest cell to an underlying graph structure of neural development. Initial NNUTT analysis of RA titration data showed that the titrations have no shifts in the glial developmental pathway. To ensure the conclusion's validity, NNUTT will be optimized to include distance thresholding. Overall, NNUTT and mass cytometry analysis of this RA titration on neural differentiation of mouse embryonic stem cells demonstrated that glial developmental pathways are relatively conserved between RA titrations, but the resulting abundance and diversity of cell types generated changes dynamically between conditions.

Keywords: retinoic acid, neural differentiation, mass cytometry, UMAP, URD

Introduction

Defects in neurogenesis contribute to the development of neurological disorders. By defining the mechanisms and signaling pathways that control neurogenesis we can understand how specific neural cell types develop and how this process is impaired in diseases, like autism, epilepsy, and Alzheimer's[1,3]. Retinoic acid (RA) mediates Vitamin A functions necessary for growth and development, and cells treated with RA are expected to differentiate into neural and mesoderm progenitors[3,4]. It's known that people with autism have decreased levels of RA in their serum and lower levels are correlated with

disorder severity[3]. Therefore, it is critical to study neurogenesis in depth and identify the cell populations that arise as a result of the presence and concentration of RA during neural development.

Previous studies of RA titrated differentiations have only analyzed neural development in fewer than 10 antibody markers of cellular identity and use methods that do not allow for measurements at the single-cell level, such as polymerase chain reaction and immunofluorescence[4–7]. These methods limit cell characterization because they are not optimized to use high numbers of markers. Using only a few markers to define an entire cell population can be misleading because many markers are present in multiple cell types. Low marker volumes do not allow for the most accurate representation of the cells studied in neural development.

Using current methods, only a few markers can be studied at a time. Thus, when researchers study a disease caused by impaired neural development, they choose markers most relevant to that disease, resulting in the characterization of one disease state[8]. Since the relevant markers vary between diseases, the methods used to study those markers will also vary between diseases. It's clear that there is not a standard platform to study the development of neural cell populations in depth. However, recent advances in the field have tried to capture cell type diversity and development with techniques like scRNAseq, flow cytometry, and mass cytometry of *in vivo* neural development in mice, but not in *in vitro* models of neural development[9].

Single-cell analysis techniques produce high dimensional datasets; mass cytometry alone producing ~40-50 markers per single cell on samples exceeding 1 million individual cells[10]. To visualize such complex datasets, researchers identify cell populations by using clustering techniques and graph-based visualizations. Cluster analysis allows researchers to analyze datasets with many dimensions effectively. UMAP (uniform manifold approximation projection) analysis groups cells into clusters based on cell marker similarity and the clusters are positions closer to or farther from other clusters that are similar or dissimilar to them. Another useful tool for studying cell populations and cell developmental pathways that exist in single cell datasets is URD. URD is a simulated diffusion-based computational method named after Urd, the Norse mythological figure that decides all fates. URD dendrograms take "root" and "tip" inputs in the form of specific cells from beginning and end time points of developmental datasets. Originally established in mapping zebrafish embryogenesis, the URD algorithm uses random walk, simulated-diffusion approaches, and time information for all of the inputted samples to link all of the other cells between the root and the tips.[11] This results in a dendrogram of the development of the cells from the first time point to the terminal populations and each cell is given a pseudotime associated with their positioning on the dendrogram. However, once this base trajectory and analysis map is created, there is not a way to visualize how perturbations of the developmental system shift the baseline developmental pathway.

This project aims to fill the gap in our knowledge of neural development by characterizing *in vitro* mouse neural development with mass cytometry and creating a tool to map perturbations of neural differentiation back to an URD reconstruction of neural development. This project creates a data generation and analysis method which will allow for up to 41 cellular measurements to be made on the single-cell level. This method allows for cell populations to be characterized using high dimensional analysis to ensure accurate characterization. Furthermore, since each cell has measurements in multiple dimensions, it is possible to group the cells together into different populations based on similarities in marker expression levels[12]. Additionally, the use of mass cytometry allows for many markers to be used simultaneously which is useful for creating a standard platform to model disease states[12]. Since the platform allows for the analysis of multiple markers, markers relevant to any particular disease state can be substituted into the mass cytometry panel. Modifications can be made to the *in vitro* model and any disease concerning neural development can be studied in depth. This project also added to the current analysis method and developed the Nearest Neighbor URD Trajectory Tool (NNUTT) to compare cell trajectories across differentiations and

Original Panel

Stem	Early Neural	Mature Neural	Glial	Mesoderm	Endoderm
Thy1	TuJ1	NeuN	Olig2	Islet1	VCAM
c-Myc	CD54	DCX	A2B5	SMA	Sox17
Ki67	Nestin	MAP2	Sox2	GAD65	FoxA2
Sall4	Sox1	IB4	BLBP	PDGFRa	
Nanog	Neurogenin2		ALDH1L1	GFAP	
SSEA-1	Pax6			PECAM1	
c-Kit	N-cadherin				
Lin28	Pax3				
ALDH1A1		Schwann	Gap Junctions	Cell-Cell Adhesion	
Klf4		Sox10	Connexin43	B-catenin	
p75NTR				CD44	

Optimized Panel

Replace		with	
	Glial		Mesoderm
	ALDH1L1		TroponinT
	Mesoderm		GATA4
	PDGFRa		

Fig. 1. Optimized Panel Design The 41 antibody markers are each specific to a cell population important to neural development. The presence or absence of each of these markers within each cell collected in the differentiation yields information about the types of populations in the neural differentiation. The past (original) antibody panel and the panel used in this research is shown and categorized by marker type. The edits to the original panel result in the optimized version which is better equipped to target mesoderm populations.

titrations of RA by building on the current URD analysis method.

The novel data generation and NNUTT will be implemented in the study of three neural differentiations at the 0 M RA, 1 nM RA, and 1 uM RA concentrations. One goal of this project is to study the biology of neurogenesis. It is expected that the 1 nM RA neural differentiation will experience delayed development of cell populations as compared to the 1 uM RA neural differentiation[6]. It is also expected that the 0 M RA differentiation will produce unique populations not found in either the 1 nM or 1 uM differentiations and as a result have less neural cells than the other titrations. Another component of the project involves developing a computational tool and the goal of the new computational tool is to compare cell trajectories within different differentiations and compare the timing of the development of the cell populations within those trajectories. To investigate these hypotheses, the mass cytometry antibody panel was optimized (Aim 1), neural differentiations were performed (Aim 2), and a novel trajectory comparison tool, NNUTT, was developed (Aim 3).

Results

Panel Design

Based on previous differentiations, a new antibody panel was designed to more accurately capture the similarities and differences between cells populations across RA titrations (Figure 1). Visual inspection of past differentiations showed that the 0 M RA culture had drastically different cell phenotypes than the 1 nM RA and 1 uM RA cultures due to the presence of beating cardiomyocyte like cells. Research suggests that these beating cells in the 0 M RA are mesodermal cell types, however the antibody panel at the time did not have adequate mesoderm markers to thoroughly confirm this. A new panel was created that encompassed more mesoderm markers and omitted extraneous markers that did not add any information about the cell populations in the cultures.

The previous antibody panel and associated dataset was observed to identify the cellular markers that defined certain populations. The dataset consisted of the full 1 uM RA time course as well as Day 6 and Day 14 1 nM RA and Day 6 and Day 14 0 M RA. First, the range of intensity of each marker was analyzed to decide if the markers were changing intensity between

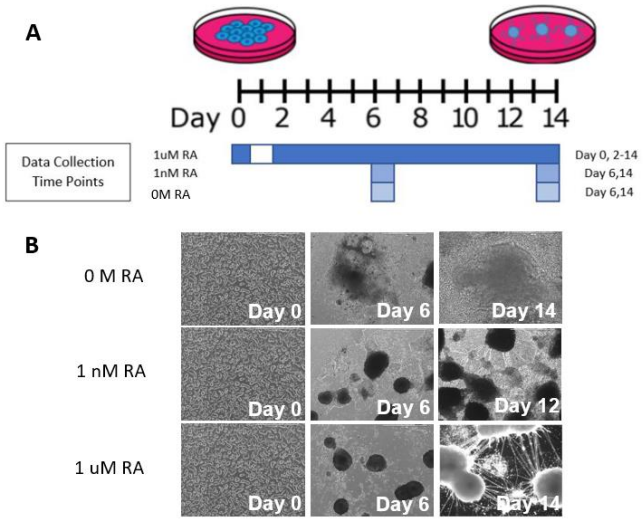

Fig. 2. Tissue Culture (A) This graphic shows the full time course of the experiment and the data collection points throughout the experiment. **(B)** Representative brightfield images of the 1 uM RA and 0 M RA neural differentiation showcasing the drastic differences in cell populations based on visual inspection only.

populations. This initial screen of the markers yielded a preliminary list of markers that may be strong indicators of cell populations. A literature search for mesoderm markers was conducted to find markers often used to identify mesoderm cell types. The list of potential markers was cross-referenced with markers available in the Zunder lab. The optimized mass cytometry panel includes the top identifying markers for each cluster identified after UMAP analyses and Leiden clustering, markers specific to the mesoderm population, and to introduce more dimensionality to the panel more identifying markers from the statistical tests were included.

Neural Differentiations

Three neural differentiations were conducted at the 0 M RA, 1 nM RA, and 1 uM RA concentrations. Cell samples were collected at Day 0 and Days 2-14 for the 1 uM RA time course (Figure 2A). For the 1 nM RA and 0 M RA titrations samples were collected at Day 6 and 14. Visual inspection of the cultures revealed the presence of axons in the 1 uM RA on Day 7 and the presence of axons in the 1 nM RA on Day 8 (Figure 2B). Beating cells were observed in the 0 M RA culture on Day 11. Immunofluorescent staining of the cultures (not shown) revealed an established neuronal antibody marker, TuJ1, staining in 0 M, 1 nM, and 1 uM cultures, but TroponinT, a cardiomyocyte marker, largely present in the 0nM RA condition but not 1 nM or 1 uM conditions, confirming what was seen by visual inspection of beating cardiomyocyte patches in the cultures.

Clustering and UMAP Application

UMAP (uniform manifold approximation projection) analysis techniques were applied to the datasets gathered from the neural differentiations to visualize high dimensional datasets and identify cell types. Clustering is based on marker expression similarity and clusters are identified based on their marker profile. The markers used are from the original panel (Figure 1). Samples from each titration collected on the same day were clustered together to determine the similarities and differences between the populations at the same point in the differentiation. UMAP analysis for Days 6 and 14 were completed (Figure 3). Notably, at Day 6 there is one cluster (2) that only the 0 M RA condition contributes to. This means that, as predicted, the 0 M RA condition has populations unique to its condition. Also, the neural clusters (top left clusters) are predominantly made up of the 1 nM RA and 1 uM RA conditions as seen by how the UMAPs colored by condition show cells in those clusters (Figure 3A). At Day 14, interestingly there is another 0 M RA only cluster (14) expressing markers Islet1, SMA, and GAD65. Again, at Day 14 the neural

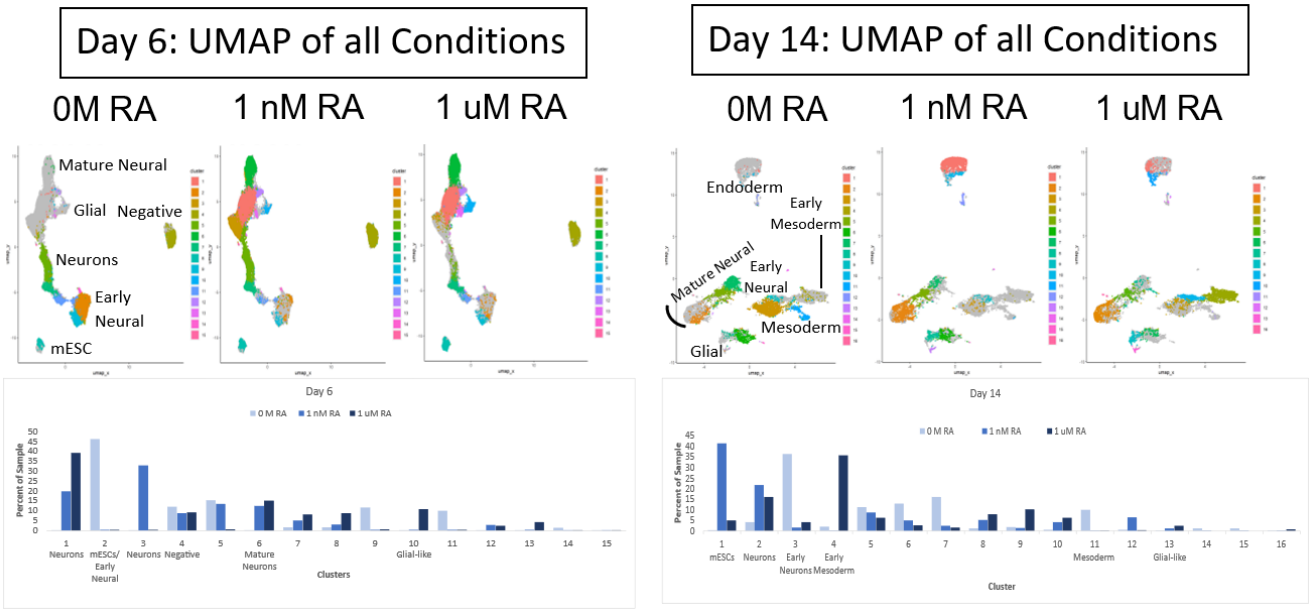

Fig. 3. Clustering Analysis of all RA titrations (A) This is a UMAP of all conditions at Day 6. The three plots to the right show the UMAP colored by each condition to show what cells of each condition are in each cluster. The bar plots quantitatively show the percentage of the condition that makes up each cluster. **(B)** This panel is the same as (A), but the conditions at Day 14 are shown.

clusters are made up mostly on the 1 nM RA and 1 uM RA conditions. (Figure 3B).

For Day 6, the bar plots confirm that most of the 0 M RA condition is in the mESC/neural progenitor cluster and not at all in the later neural clusters. The 1 nM RA and 1 uM RA are both found in the later clusters, however, only the 1 nM condition is found in the early neurons indicating that the 1 uM condition has already matured past this population. For Day 14, the bar plots show that the 0 M RA condition is mostly in an early neural subtype. Surprisingly, most of the 1 nM condition is found in an endoderm-like cluster indicating that now all these cells have differentiated in neurons, but most of the remaining 1 nM RA cells are found in a neural cluster. Interestingly, the 1 uM RA condition at Day 14 is found primarily in the early mesoderm subtype. Since the 1 uM subtype and not the 0 M subtype is shown in the early mesoderm cluster despite visual inspections of the cultures that indicate otherwise, it is clear that the markers of the original panel do not accurately capture the cell types in the 0 M condition.
This analysis allows us to compare the types and abundance of clusters or populations of cells across the differentiations, but it does not reveal any information about the timing of these populations.

URD Application and Nearest Neighbor URD Trajectory Tool (NNUTT) Design

An URD dendrogram was built using an established 1 uM RA dataset of a 12 day neural differentiation of mouse embryonic stem cells. This URD analysis reconstructed neural development as the baseline map of neural development *in vitro*. The Day 0 mESCs were used at the “root” of the dendrogram and the terminal cell populations at Day 12 were used as the “tips” of the dendrogram. Day 12 rather than Day 14, the final day of the differentiation, was used because of excessive cell death at Days 13 and 14 and similarity of the terminal cell populations at Day 14 to the terminal cell populations at Day 12. The original URD reconstruction enables identification of cell trajectories within a dataset, but does not allow for comparison of trajectories between our titration samples of RA[11].

NNUTT was built to compare the timing of similar trajectories across different neural conditions. The tool takes all the cells from one condition (Either 0 M RA, 1 nM RA, or 1 uM RA in this case) and maps them to the most similar cells on the URD that the condition is being compared to (Figure 4). The most similar cells are based on minimum Euclidean distance between the cells the construct the URD and the cells that are being mapped. This operates under the assumption that the minimum distance indicates that the cells have the most similarity of all the other cells. Additionally, NNUTT analysis assumes that all cell types are represented in the base map, which is not necessarily true for that dataset in this paper and thus, is a

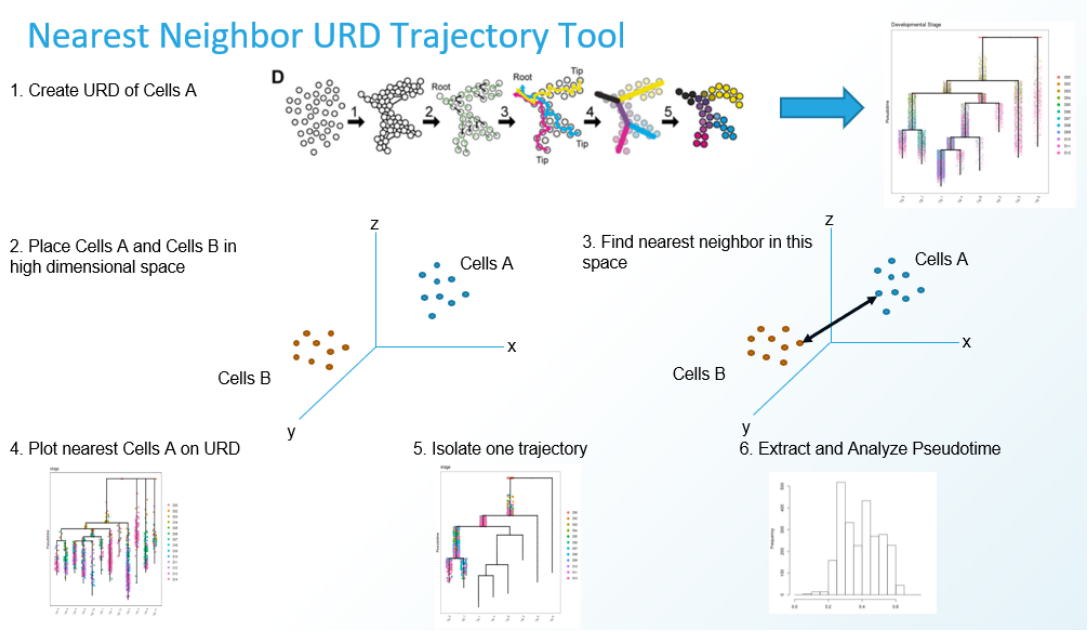

Fig. 4. Nearest Neighbor URD Trajectory Tool (NNUTT) For the URD-Trajectory Mapping tool, the required inputs are an URD of a full time course (1 uM RA time course in this case) and a UMAP of another condition (1 nM RA or 0 M RA). The Euclidean distances between the URD cells and the UMAP cells are calculated based on the marker similarities of the cells. The UMAP cells are mapped to the URD cell that is the shortest Euclidean distance away and the mapped and the cells that fall along certain segments are plotted on the URD resulting in the mapping of a single trajectory.

limitation of the tool. Each cell in the condition being mapped has its distance calculated to every cell that makes up the URD and the cell on the URD with the smallest distance is selected as that cell's most similar cell. The URD constructed from the 1 uM time course is used as the URD that the cells of each condition will be mapped to. Then, only the similar cells along certain segments of the URD are plotted so identify the most similar cells along a specific cell trajectory.

Subsequently, the pseudotime information for each of the mapped conditions can be extracted to quantitatively define differences in the timing of cell development along a certain trajectory (towards a glial fate, neural fate, etc.).

NNUTT Validation

The novel URD-Trajectory Mapping tool NNUTT was validated by mapping certain clusters with known marker identities onto an URD of the 1 uM RA time course. Since it was observed that when the mature neural cluster from the Day 14 UMAP (cluster 11 in Figure 3) mapped to the neural branch of the URD, it was confirmed that indeed the cells from the UMAP were being accurately mapped to similar cells on the URD.

Further, the neural progenitor cluster from the Day 14 UMAP (cluster 2 in Figure 3) was also mapped and these cells also mapped to the neural branches. Visual inspection confirmed that the neural progenitor cluster was mapped to the same URD segment as the mature neural cluster; however, the progenitor cluster was shown higher up on the URD. This indicates that the progenitor cluster precedes the mature cluster on the pseudotimeline and confirms that the NNUTT is satisfactorily mapping similar cells.

Discussion

Findings

Implementation of NNUTT resulted in defining the pseudotime distributions of mappings of each condition to the glial trajectory of the base URD dendrogram and discovering unexpected similarities in the distributions.

Pseudotime analysis was completed by NNUTT-Trajectory Mapping of each of the 0 M RA, 1 nM RA, and 1 uM RA conditions to the glial trajectories of the 1 uM RA URD. URD analysis of the 1 uM RA time course (Day 0 through Day 12) resulted in eight terminal populations (Figure 5A). By overlaying each of the 39 markers onto the URD, it was discovered that

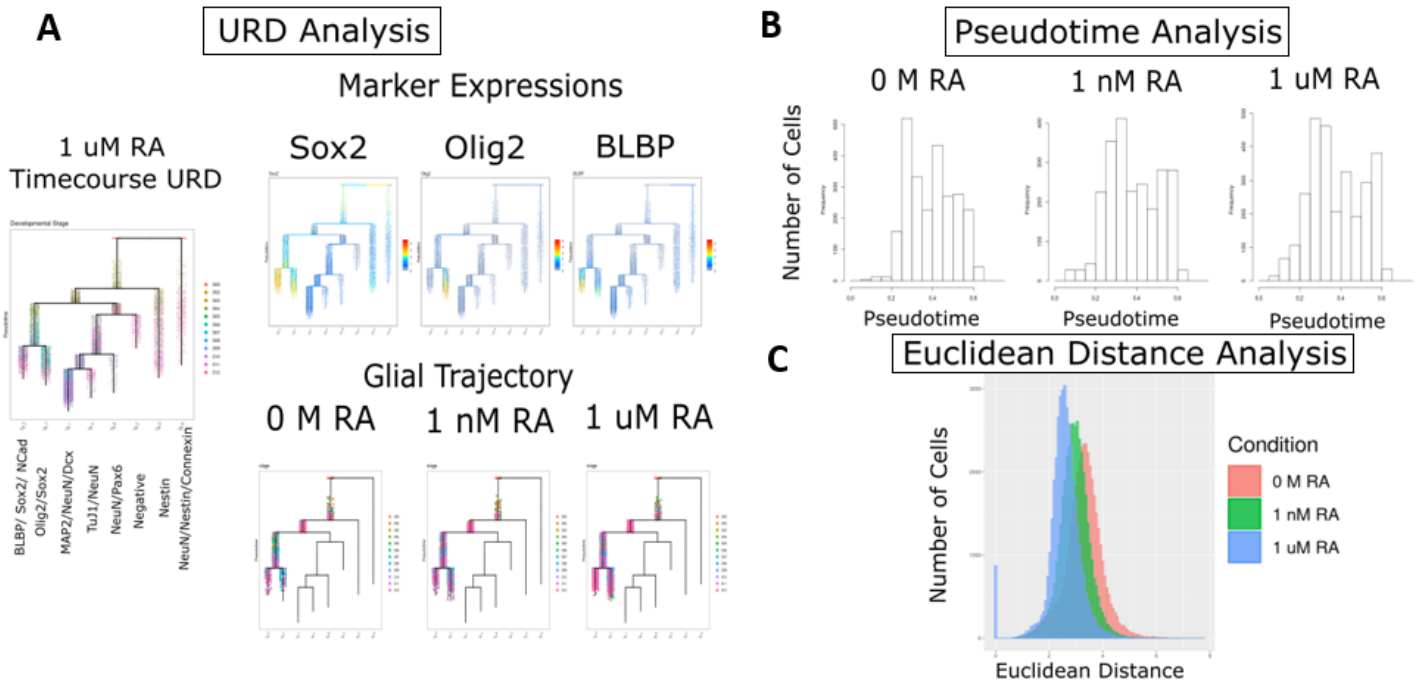

Fig. 5. Pseudotime Analysis of Glial Trajectory and URD-Trajectory Mapping Tool Analysis (A) URD from 1 uM RA time course beginning at Day 0 and ending at Day 12 with classifications of terminal cell populations. The URD-Trajectory Mapping tool was used to plot the cells from each RA condition along the glial trajectory. The overlay of Sox2, Olig2, and BLBP indicates that the plotted trajectory is in fact a glial trajectory. **(B)** The cell counts of each condition from the mapped trajectory is plotted over pseudotime. **(C)** Histograms of the Euclidean distance between the actual cells from each condition and the plotted most similar cells.

two of the terminal segments are glial progenitors. Sox2, Olig2, and BLBP confirm that the two leftmost populations on the dendrogram are glial progenitors (Figure 5B). The most similar cells on the dendrogram to the 0 M RA, 1 nM RA, and 1 uM RA conditions were found and plotted using NNUTT (Figure 5B). Analysis of the most similar cells to each condition was completed and histograms of the cell counts and cell pseudotime distributions along the trajectories were plotted (Figure 5C). Cells that mapped to timepoint 0 (on a scale from 0 to 1) were excluded from each of the histograms because the amount of cells at that pseudotime point did not allow for the other cells to be visualized. It was also determined that the 0 timepoint was not relevant for pseudotime distribution analysis because it does not indicate the timing of cell populations along the trajectory, it only indicates how many cells in the trajectory are similar to the starting point.

This analysis revealed that the cell counts and pseudotime distributions were unexpectedly similar between the populations. It was hypothesized that the 0 M RA condition would have less neural progenitor cells than the 1 nM RA and 1 uM RA conditions due to unique cell populations existing in the 0 M culture resulting in less neural cells. However, the histograms indicate that there are similar amounts of neural cells in all of the different RA conditions. Additionally, the pseudotime distributions were also unexpectedly similar. It was hypothesized that the 1 nM RA culture condition would be delayed when compared to the 1 uM RA condition. However, the pseudotime distributions do not show this.

To further investigate the mapped cells, the Euclidian distances between each of the actual cells from each condition and the plotted most similar cells were plotted against the counts of cells at that distance. Statistical analysis (a series of Welch Two Sample t-tests of all possible sample pairs with $p<2.2e-16$) of these populations indicated that the 0 M RA conditions had the largest Euclidean distance, indicating that these cells are the most dissimilar of the mapped cells of each condition. The 1 nM RA mapped cells had the next largest Euclidean distance and the 1 uM RA had the smallest distance showing that its cells were the most similar which is to be expected as the dendrogram was built from cells with the same culture conditions. This discrepancy in distances explains that the cell counts and pseudotime distributions look similar because the cells that are being mapped to on the URD are likely the same for each condition; however, the similarity varies between the conditions. Thus, the conditions are not being represented accurately by the data. To address this limitation and potential mapping of non-similar cell types, NNUTT will be expanded to include a thresholding parameter for mapping of cell types to the baseline URD.

Significance

This newly developed method, NNUTT, of comparing entire trajectories to one another will be useful in a variety of applications, such as the study of other cell fates, small molecule perturbations to neural differentiations, and CRISPR/Cas9 mediated studies of neural differentiation. This tool can also be applied to fields that look at the development of populations over time and have an interest in comparing either the rates of development or comparing the similarities of trajectories.

Design Constraints

Panel

When optimizing the mass cytometry antibody panel, it was necessary to include a maximum of 41 metal-conjugated antibodies. This required prioritization of antibodies based on: relevance to neurogenesis, marker reliability, and pertinence to suspected uncaptured populations in the dataset.

NNUTT

This tool operates under the assumption that the smallest Euclidean distance between two cells indicates that those cells are the most similar of all other cell pairings. This tool is currently limited in that there is not a specified similarity value. For example, two cells may be considered similar by the algorithm because of a small distance; however, the cell still may be very different to each other when examining marker expressions because of a lack of a similar cell population in both datasets.

Limitations

This project uses the mass cytometry antibody panel in Figure 1A and therefore was unable to capture the mesoderm subpopulations in enough detail to distinguish them as different from some of the neural populations, such as the early neural populations. COVID-19 inhibited running more differentiations, so

no new data was collected using the optimized panel in Figure 1. Furthermore, because COVID-19 interrupted wet lab activities, the RA titrations were not able to be replicated, thus these results represent experiments with n=1. This low sample size does not allow for adequate significance testing.

Future Directions

Future plans for this project include running the experiment with replicates and using the optimized mass cytometry antibody panel. This project only investigated the glial trajectory in detail, but analyzing the other cell trajectories in detail would provide useful information regarding the similarities and differences across the RA differentiation conditions. With regards to NNUTT, there are several recommendations that may improve the algorithm. First, setting a distance threshold for the maximum allowable distance between mapped similar cells will further ensure that the cells mapped to the trajectory are in fact similar. Also, further analysis of the cells that were mapped to the URD by applying clustering and UMAP analyses should be done. That analysis will illuminate the true marker expressions and intensities to determine how similar the mapped cells are to the URD. Since these recommendations are concerned with ensuring that the cells that map to the URD are indeed similar, a metric for similarity must be decided upon, whether that be a Euclidean distance threshold or a measure of marker similarity.

Materials and Methods

Tissue Culture

E14tg2a mouse embryonic stem cells were thawed onto mouse embryonic fibroblast cells (MEFs) treated with mitomycin-C. The cells were cultured in mESC media with factors 2i and LIF to maintain pluripotency. The mESC media consisted of 1000 parts DMEM F12 with L-glutamine and sodium bicarbonate, 100 parts ES Cell FBS, 10 parts sodium pyruvate, 10 parts non-essential AAs, 10 parts Pen Strep, and 1 part 100 mM beta-mercaptoethanol. The media is filtered with a 0.2 um filter and stored at 4^{o} C. TrypLE was used to dissociate cells for passaging and cells were passage 1:16 onto gelatin coated tissue culture treated plates. Cells were passaged at least three times after thawing in order to dilute mitotically inactivated MEFs. For the 14-day neural differentiation, at Day 0 mESCs cultured in 2i and LIF were dissociated and cultured as embryoid bodies (EBs) at 5000 cells/drop in a DMEM/FBS medium. At Day 2.5, the EBs were plated onto gelatin coated 6-well plates in basal N2B27 medium. The 0 M RA condition's media contained the B27 supplement without Vitamin A and the 1 nM RA and 1 uM RA conditions' media contained the B27 supplement with Vitamin A. The N2B27 medium consisted of 50% DMEM F12, 50% Neurobasal Media, 1X N2, 1X B27 with(out) vitamin A, 100mM b-mercaptoethanol, and 1X Pen Strep. The EBs in the N2B27 media were additionally given either 0 M, 1 nM, or 1 uM of retinoic acid in order to induce neural differentiation. This process if shown in Figure 6. The media was replaced every 2 days until the end at of differentiation at 14 days. Media changes were on Day 2, 4, 6, 8, 10, and 12.

Cell Dissociation and Fixation

At Day 0, Day 2.5, and every 24 hours from Day 3 to Day 14, cell samples from the 1 uM RA in vitro neural differentiation were collected. Cell samples from the 0 M RA and 1 nM RA in vitro neural differentiations were collected only on Day 6 and 14. Cell collection means the cell samples were washed with 1X DPBS then incubated with 1X TrypLE for 5 minutes at 37C. After incubation, samples were dissociated by rapid mechanical dissociation with a P1000 pipette and filtered through a 70-µM filter. The filtered cells were immediately fixed at room temperature for 10 minutes

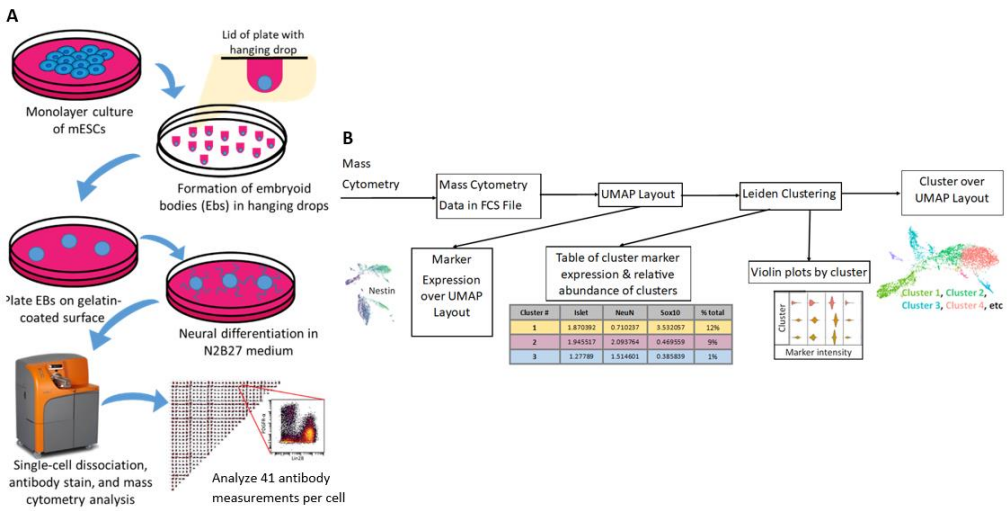

Fig. 6. Tissue Culture Methods and Analysis Approach (A) On Day 0, mESCs were cultured in vitro with 2i/LIF to maintain pluripotency. Differentiation of mESCs was induced by embryoid body (EB) formation using the hanging drop method. On Day 2.5, the EBs were collected and plated onto gelatin in basal N2B27 medium with RA to promote neural differentiation. Every 24 hours after Day 2.5 the cells were collected, fixed with paraformaldehyde, and stored at -80C. Samples were thawed, barcode multiplexed, and stained with metal-conjugated antibodies to measure 41 markers by single-cell mass cytometry. **(B)** The UMAP uniform manifold approximation and projection algorithm was applied to the dataset to create a map which subsamples and clusters cell populations based on marker similarity. Cell populations are then identified based on the marker profile of a cluster.

in 1.6% paraformaldehyde and then stored at -80°C in cell staining media (0.5% BSA, 0.02% NaN_3 in PBS).

The final sample suspensions were mixed with 100 µL of 2× cisplatin solution (10 µM in PBS; Sigma Aldrich, P4394). The samples were incubated for 30 seconds before quenching with 1.3 mL of PBS containing 0.5% BSA. After 3 minutes of centrifugation at 300 × g at 4ºC, resulting cell pellets were washed once with PBS containing 0. 5% BSA before fixation in 1 mL of 1.6% paraformaldehyde solution (Electron Microscopy Services, CAS 30525-89-4) in PBS for 10 minutes at room temperature. After 3 minutes of centrifugation at 600 × g at 4ºC, the cell samples were washed once with PBS before final resuspension in 1 mL of cell staining medium (CSM; 0.5% BSA, 0.02% NaN_3 in PBS) and subsequently stored at -80ºC.

Mass Cytometry Antibody Preparation

Mass cytometry antibodies used in this experiment are summarized in Figure 1 (original panel). Antibodies were conjugated to isotopically pure lanthanide metals with the MaxPar antibody conjugation kit according to manufacturer's protocol. Antibodies are stored at 4ºC post-conjugation in a candor stabilization buffer. Some antibodies were first validated by flow cytometry before mass cytometry antibody conjugation. Antibody cocktails for cell staining are prepared fresh each day of antibody staining.

Mass-Tag Cell Barcoding (MCB)

After all the samples were collected, the cell samples were thawed and individually Mass-tag Cell Barcoded (MCB) as previously described[13,14]. Cells are permeabilized with saponin and then incubated with unique mixtures of Palladium containing MCB reagents in DMSO at 1:100 DMSO to cell solution with saponin. After a 10-minute incubation at room temperature. Each sample is washed 3 times with cell staining media (CSM) and then all samples are pooled together. MCB barcoding allows for up to 20 barcoded samples to be pooled together. For this experiment, each unique cell sample (considering timepoint and condition) had a unique barcode. The 1 uM RA timecourse was pooled together and the 1 nM RA and 0 M RA samples were pooled together.

Mass Cytometry Cell Staining and Measurement

After barcoding and pooling of samples, the pooled cells are stained with metal-conjugated antibodies to measure up to 41 proteomic-level markers of neurogenesis by single-cell mass cytometry. First, the samples are blocked with 10% donkey serum, stained with surface antibody metal-conjugated antibodies, then methanol permeabilized for 10 minutes at 4C before being stained for intracellular markers. Then, the samples are incubated with Iridium interculator for DNA staining. This incubation at 4C ranges from overnight to up to 2 weeks before analysis on CyTOF mass cytometer (DVS Sciences). Normalization beads (DVS Sciences beads containing Lanthanum-139, Praseodymium-141, Terbium-159, Thulium-169, and Lutetium-175) are added to the samples for normalization as described previously[15]. The sample and bead mixtures are filtered through a 40 uM filter and analyzed at ~250 cells per second on CyTOF. Next, the samples are normalized and de-barcoded to individual FCS files for each sample.

Mass Cytometry Data Processing

The raw FCS files containing mass cytometry data were processed using the Cytobank software. First, cell doublets, debris, and beads were gated out and discarded from the dataset. FCS files from different CyTOF runs are batch corrected.[16] Next, the ranges for each marker's positive and negative ranges were adjusted so that markers with smaller ranges could be read as effectively as marker's with larger ranges.

ZunderLabPipeline

The finalized FCS files entered the ZunderLabPipeline for high dimensional data analysis. The cell events in the FCS files are used in UMAP and URD analysis.

<u>UMAP Analysis: Clustering</u>

UMAP analysis is used to reduce the dimensionality of the mass cytometry data. To project the cell events in two dimensions, the UMAP algorithm was used in Python and R (v3.6.3). Leiden clustering was completed in Python and the umap function in R was used to create the plots. UMAP settings used are as follows: n_neighbors = 15, min_dist = 0.000001. n_components =2, metric = "Euclidean", n_epochs = 1000, input = "data", init = "spectral", set_op_max_ratio = 1, local_connectivity = 1, bandwidth = 1, alpha =1, gamma = 1, negative_sample_rate = 5, A = NA, B = NA, spread = 1, random_state = 1, transform_state = NA, knn_repeats = 1, verbose = FALSE, learn_args = NA.

<u>URD Analysis</u>

URD analysis was used to discover cell population trajectories within the 1 uM RA dataset. The URD was constructed using the cell marker information for Days 0 and 2-12. At each day, the cells were clustered and negative clusters were removed. The "root" of the tree was defined as all of the clusters found at Day 0 and the "tips" were defined as each of the clusters as Day 12 that were terminal. Clusters indicative of residual stem cells or earlier neurons etc. were not assigned at tips.

The URD library in R was used to construct the trajectories. After the roots were specified, pseudotime information was determined using the floodPseudotime function (settings: n=50, minimum.cells.flooded = 2, verbose=F). Next, the tips were specified and then biased random walks from each tip were simulated and processed using the simulateRandomWalksFromTips (settings: URD_Object, tip.group.id="tip.clusters", root.cells=root.cells, transition.matrix = URD_Object.biased.tm, n.per.tip = 25000, root.visits = 1, max.steps = 5000, verbose = F) and processRandomWalksFromTips (settings: verbose =F) functions. The URD tree was built using the buildTree function with the following parameters: pseudotime = "pseudotime", tips.use=1:num.tips, divergence.method = "preference", cells.per.pseudotime.bin = 25, bins.per.pseudotime.window = 8, save.all.breakpoint.info = T, p.thresh=0.001.

End Matter

Author Contributions and Notes

E.A.P. and K.I.F. designed research; E.A.P. and K.I.F. performed research; E.R.Z, C.M.W., S.G, A.V., A.K., K.I.F, and E.A.P contributed to data analysis pipeline; E.A.P. wrote new software; E.A.P. and K.I.F. analyzed data; and E.A.P. wrote the paper.
The authors declare no conflict of interest.

Acknowledgments

I would like to thank Eli Zunder, my advisor in UVA's Department of Biomedical Engineering, for providing expertise and advice throughout the years I have worked on this project. I would also like to Kristen Fread, Ph.D. candidate in the Zunder lab, who has spent countless hours offering support and advice. Furthermore, I would like to thank all members of the Zunder lab for giving input and coding expertise.

References

1. Gooch, C. L., Pracht, E. & Borenstein, A. R. The burden of neurological disease in the United States: A summary report and call to action. *Ann. Neurol.* **81**, 479–484 (2017).
2. Engberg, N., Kahn, M., Petersen, D. R., Hansson, M. & Serup, P. Retinoic Acid Synthesis Promotes Development of Neural Progenitors from Mouse Embryonic Stem Cells by Suppressing Endogenous, Wnt-Dependent Nodal Signaling. *STEM CELLS* **28**, 1498–1509 (2010).
3. Zhou, W. & Li, S. Decreased levels of serum retinoic acid in chinese children with autism spectrum disorder. *Psychiatry Res.* **269**, 469–473 (2018).
4. Kam, R. K. T., Deng, Y., Chen, Y. & Zhao, H. Retinoic acid synthesis and functions in early embryonic development. *Cell Biosci.* **2**, 11 (2012).
5. Korecka, J. A. *et al.* Phenotypic Characterization of Retinoic Acid Differentiated SH-SY5Y Cells by Transcriptional Profiling. *PLOS ONE* **8**, e63862 (2013).
6. Jacobs, S. *et al.* Retinoic acid is required early during adult neurogenesis in the dentate gyrus. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3902–3907 (2006).
7. Mishra, S., Kelly, K. K., Rumian, N. L. & Siegenthaler, J. A. Retinoic Acid Is Required for Neural Stem and Progenitor Cell Proliferation in the Adult Hippocampus. *Stem Cell Rep.* **10**, 1705–1720 (2018).
8. Gouti, M. *et al.* A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* (2017) doi:10.1016/j.devcel.2017.04.002.
9. Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* **11**, 3406 (2020).
10. Fluidigm | Mass Cytometry. https://www.fluidigm.com/singlearticles/mass-cytometry.
11. Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, (2018).
12. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A Continuous Molecular Roadmap to iPSC Reprogramming through Progression Analysis of Single-Cell Mass Cytometry. *Cell Stem Cell* **16**, 323–337 (2015).
13. Fread, K. I., Strickland, W. D., Nolan, G. P. & Zunder, E. R. AN UPDATED DEBARCODING TOOL FOR MASS CYTOMETRY WITH CELL TYPE-SPECIFIC AND CELL SAMPLE-SPECIFIC

STRINGENCY ADJUSTMENT. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **22**, 588–598 (2017).

14. Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm - PubMed. https://pubmed.ncbi.nlm.nih.gov/25612231/.
15. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytom. Part J. Int. Soc. Anal. Cytol.* **83**, 483–494 (2013).
16. Schuyler, R. P. *et al.* Minimizing Batch Effects in Mass Cytometry Data. *Front. Immunol.* **10**, (2019).