Characterization of the Differentiation Patterns of Histone H3 K27M Neural Stem Cells

Liam Tran

Bachelor of Science in Biomolecular Engineering

Haussler/Salama Lab

Jack Baskin School of Engineering

University of California, Santa Cruz

Abstract

This project seeks to characterize the nature of a hypothesized differentiation stall induced by the Histone H3 K27M mutation in neural stem cells. This mutation occurs in 30% of pediatric high-grade gliomas and 80% of Diffuse Intrinsic Pontine Gliomas, a more aggressive form of pediatric high-grade gliomas. Investigation of the cell-type specific timing of the differentiation stall in the neural development pathway was conducted with immunofluorescence staining of cerebral organoids for the following protein markers: SOX2 (undifferentiated pluripotent stem cells, neural epithelium and radial glia neural stem cells), SATB2 (upper layer neurons), VIM (radial glia neural stem cells), TBR2 (intermediate progenitor cells), PAX6 (radial glia cells), CTIP2 (deep layer neurons), and RELN (layer 1 Cajal-Retzius neurons). To supplement this visual data, single-cell RNA sequencing analysis was used to further examine the cell types present in cerebral organoids harboring the H3 K27M Mutation. Analysis of both sets of data supports the hypothesis that the differentiation stall occurs between neuroepithelium and radial glia cells, but further optimization of the single-cell RNA sequencing pipeline is ongoing to provide a more accurate location of the differentiation stall.

Acknowledgments

I would like to thank my advisors David Haussler and Sofie Salama for their time and support since I have been a part of the Haussler Wet Lab. I would also like to thank Kristof Tigyi for answering my questions and taking the time to show me new laboratory techniques. I would especially like to thank Lauren Sanders for her incredible support and mentorship throughout this project. Additionally, I would like to thank Ryan Hoffman, Alexander Hoefler, and Jennifer Yang for their immunofluorescence tips and tricks as well as the rest of the Haussler Wet Lab for being supportive and answering any questions I had. Finally, I would like to thank my parents and my sister, Lily Tran, for their support throughout my undergraduate career.

Abbreviations

pHGG - Pediatric High-Grade Glioma

DIPG - Diffuse Intrinsic Pontine Glioma

GFP - Green Fluorescent Protein

GTR - Gross Total Resection

CNS - Central Nervous System

IF - Immunofluorescence

Co-IP - Co-Immunoprecipitation

TFM - Tissue Freezing Medium

PBS - Phosphate-Buffered Saline

hESC - Human Embryonic Stem Cell

EDTA - Ethylenediaminetetraacetic acid

BSA - Bovine Serum Albumin

DAPI - 4,6-diamidino-2-phenylindole

Contents

List of Figures

1 Problem Statement

The purpose of this project is to determine whether neural stem cells harboring the Histone H3 K27M mutation are capable of differentiating into the same types of mature neural cells as wild-type Histone H3 neural stem cells. The Histone H3 K27M mutation is of interest because it occurs in 30% of pediatric high-grade gliomas (pHGGs), a class of children's cancer occurring in the brain, and 80% of Diffuse Intrinsic Pontine Gliomas (DIPGs), a more aggressive form of the tumor occurring in the brainstem [9]. In fact, 80% and 90% of patients do not survive two years after diagnosis of pHGGs and DIPGs respectively, due to the lack of treatment options [9].

pHGGs, and especially DIPGs, are extremely difficult to treat surgically due to the location of these tumors in the brain. pHGGs occur in the cortex while DIPGs occur primarily in the brainstem and pons. Complications arise during surgical resection of pHGGs because it is difficult to differentiate between normal brain tissue and the tumor tissue [2]. It is nearly impossible to treat DIPGs by surgical resection. There is also a lack of therapeutic drug treatments to combat these tumors. In order to uncover therapeutic targets and alternative treatment approaches, a thorough understanding of the mechanism by which this mutation drives tumorigenesis is needed. Previous studies suggest this mutation causes differentiation to stall during the pathway from neural epithelial cells to mature brain cells such as astrocytes or oligodendrocytes [5]. The precise nature and location of this stall is unclear and more research needs to be done to pinpoint where it occurs. I propose to characterize the differentiation patterns of mutant and wild-type neural stem cells by: (1) performing immunofluorescence staining on cerebral organoids harboring the mutation, and (2) utilizing a bioinformatic tool developed in the Haussler/Salama Lab at the University of California, Santa Cruz to analyze single-cell RNA sequencing data to further identify cell states in cerebral organoids at relevant time points during neural development.

Cerebral organoids have several advantages over conventional two-dimensional cell culture techniques. Organoids' most important advantage is that they are able to replicate living

tissue far more accurately than two-dimensional techniques. *In vivo*, cells naturally organize themselves into complex structures as they form living tissue. These structures are referred to as the cytoarchitecture of the tissue. Organoids are far better equipped to mimic the cytoarchitecture of tissue in the lab. Because organoids are capable of modeling early neural development, they also provide excellent models for for the progression of diseases which arise over time during neural development such as DIPG [10].

The engineered cell lines used in this project to develop cerebral organoids are human embryonic stem cells (H9) [18] transduced with either a lentivirus harboring the Histone H3 K27M mutation and a green fluorescent protein (GFP) sequence, or a control lentivirus with wild-type histone H3 and GFP. After culturing and maintenance, the individual transduced clones were isolated and tested for robust, inducible expression of GFP and H3. The successful clones were cultured until they were ready for aggregation and seeded into cerebral organoids at 0.5% and 10% along with H9 cells in a differentiation assay developed by Andrew Field et al. in the Haussler/Salama Lab [4]. GFP expression allows identification of the control and mutant cell lines within the organoids under a fluorescence microscope.

At specific time points physiologically relevant to neural development, I will harvest mutant and control cerebral organoids, and perform immunofluorescence staining on them. In this project, antibodies with affinity for protein markers specific to cell states involved in the neuron differentiation pathway are used to identify which cell states are present in the organoid. Analysis of which cell states are present will aid in pinpointing where the differentiation stall occurs. For example, if the differentiation stall occurs between the transition from neural epithelial stem cells to radial glia cells, there will be fewer GFP-positive radial glia cells in the H3 K27M mutant organoids when compared to the H3 wild-type organoids.

The planned immunofluorescence stains for this project include neural epithelial, radial glia, intermediate progenitor, deep layer neuron, upper layer neuron, nuclear, and neural stem cells. I have attempted radial glia and intermediate progenitor stains, but further optimization of the protocol is required to remove noisy signals that are obscuring the relevant

visual data. I will optimize the protocol by performing serial dilutions of each antibody on the organoids as well as experiment with different concentrations of blocking serum. The visual data generated by the immunofluorescence staining will be supplemented by single-cell RNA sequencing data analysis of the organoids.

2 Background

Pediatric high-grade gliomas (pHGGs) are the primary cause of cancer related deaths in children [2]. They comprise 8-12% of all pediatric central nervous system (CNS) tumors. While rare (incidence rate of 0.85 per 100,000 people), pHGGs are characterized by their extremely aggressive nature. The survival rate of pHGGs 5 years after diagnosis ranges from 15-35% [3]. The bleak mortality rate of pHGGs stems from the location of these tumors. They occur in the brain, making surgical treatment very difficult or impossible. A more aggressive subclass of pHGGs, Diffuse Pontine Intrinsic Glioma (DIPGs), occurs primarily in the brainstem, complicating surgical removal in a similar manner. Unfortunately, there is no standardized method of treatment for CNS tumors but most experts tend to recommend a gross total resection (GTR) followed by focal radiation therapy [3]. A GTR refers to a surgical procedure to excise the tumor completely from the brain. However, DIPGs are impossible to treat surgically since the risk of fatal brain damage is so great [12]. The issue with this method is that it is difficult to distinguish the tumor from normal brain tissue [2]. In the case of DIPGs, these tumors actually show a resistance to conventional chemotherapy treatments [11].

The lack of effective treatment options means more research needs to be done to understand the mechanism of tumorigenesis in pHGGs. A firm grasp of the molecular mechanism may uncover therapeutic targets to treat this class of tumors. Unfortunately, there is currently no consensus model for the timing of pHGG development or the tumor cell of origin.

2.1 The Histone H3 K27M Mutation

Histones are proteins that play a fundamental role in DNA compaction and gene regulation. Histones are composed of smaller protein subunits called: H2A, H2B, H3, and H4. Together, these subunits form one half of the histone dimer, so the whole protein is actually an octamer made up of eight subunits. The whole complex of histone and its coiled DNA is

referred to as a nucleosome. Fully stretched out, each strand of human DNA is about 2 meters long. This means that DNA needs to be condensed into structures known as chromatin so that it can fit into the nucleus of a cell, roughly 10 μM in diameter. Chromatin is further condensed into chromosomes. Additionally, DNA has a strong negative charge because of

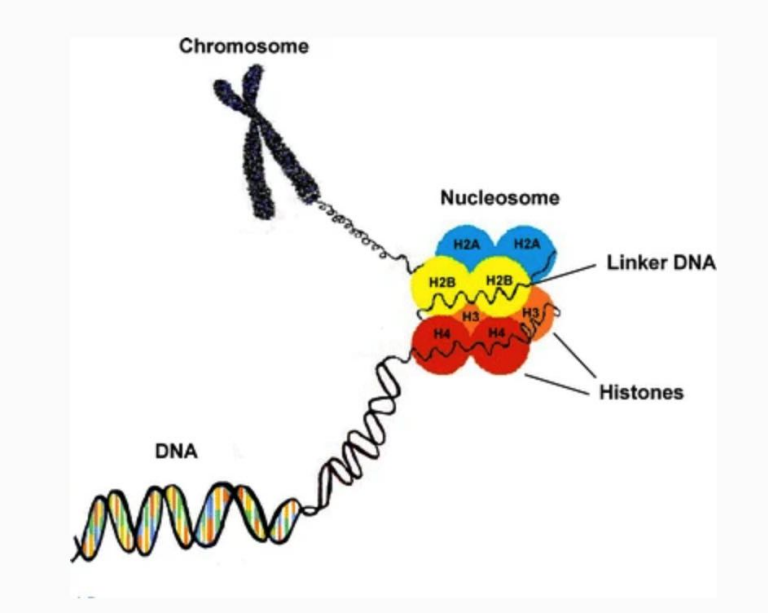


Figure 1: DNA is compacted into chromatin which is then organized into chromosomes. Histones are the protein responsible for making this level of compaction possible. Anier, Kaili, and Anti Kalda. "Epigenetics in the Central Nervous System." Current Geriatrics Reports 1.4 (2012): 190-198.

the phosphate backbone that links nucleotides together. The histone protein, which is a structural protein responsible for organizing DNA, is positively charged and compacts DNA into the nucleus of a cell by using the electrostatic attraction between the protein and DNA to wrap DNA around itself. Each histone protein can coil 146 base pairs, the equivalent of 1.7 turns. Because DNA is coiled so tightly around histones, they also play a role in transcription and gene regulation. For transcription to occur, RNA polymerase needs to get between the strands of DNA so that it can read one strand as a template to begin making the RNA transcript. This is impossible if DNA is wrapped around a histone. Therefore, histones have a significant impact on gene regulation. Consequently, mutations occurring in histones or other proteins that control chromatin dynamics often lead to cancer [8].

The mutation K27M (lysine to methionine at position 27) in histone H3 is found in 30% of pediatric high grade gliomas (pHGGs) and 80% of Diffuse Pontine Intrinsic Gliomas (DIPGs) [9]. In fact, a new molecular class of tumors is named after the H3 K27M mutation according to the World Health Organization 2016 brain tumor classification meeting [17]. This class of tumors is extremely difficult to treat since 80% occur in the brainstem, making surgical treatment impossible. Additionally, those diagnosed with H3 K27M tumors have an extremely low survival rate of two years after diagnosis (20%) [9]. The mutation from lysine to methionine removes a lysine-specific methyltransferase site and trimethylation of K27 represses transcription of genes that are thought to drive tumorigenesis [9]. Interestingly, this mutation has a global effect within the cell. There are many genes that encode for histone H3 and the H3 K27M mutation generally only occurs in one copy. So the mutant histone only comprises a small fraction of the total histone H3 protein population within the cell. Even if the H3 K27M population within the cell is as low as 3-17% of the total H3 population, trimethylation of K27 is decreased throughout the whole H3 population in the cell [9]. Additionally the H3 K27M mutation seems to stall differentiation, keeping cells in a stem-like state [5]. It is unknown which stage this differentiation stall occurs or in what cell type [9]. Based on a study by Pathania et al. [14], and previous unpublished H3 K27M experiments carried out in the Haussler Lab, it is thought that the H3 K27M mutation occurs during early brain development even though the tumor does not become apparent until around 6-9 years of age.

Previous studies on the mutation seem to agree on a few key ideas. A review by Brandon Lowe et al. summarizes the current research on lysine-to-methionine mutations. It is well publicized that the H3 K27M mutation occurs through missense mutations in the H3F3A and HIST3H1B genes which encode for H3.3 and H3.1 variants of the H3 subunit, respectively [9]. Many other studies conducted on the mutation were also able to show that the mutation decreases H3 K27me3 (lysine trimethylation at position 27). However, the full mechanism has yet to be uncovered.

Researchers have attempted many different methods in an effort to uncover this mechanism. The lysine residue at position 27 in histone H3 provides a target for the protein polycomb repressive complex 2 (PRC2). PRC2 belongs to a class of proteins that regulate genomic H3 K27 methylation. Methylation of histone H3 at position 27 is a repressive mark that leads to less transcription of genes with methylated histone 3 proteins in the region around their promoters. It is thought that removal of transcriptional repression in the presence of this mutation aberrantly activates oncogenes (genes that can cause cells to become tumor cells under certain conditions). Removal of the lysine residue prevents methylation and causes cells to become oncogenic. Lowe et al. discusses several models for how this

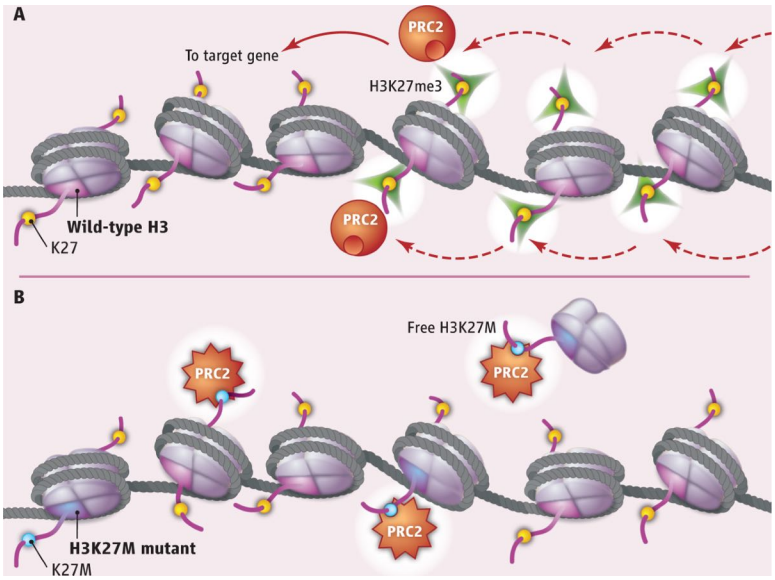


Figure 2: Tethering and Sequestration model. The protein PRC2 binds with higher affinity to methionine and does not dissociate. This decreases the repressive effect of the trimethyl mark on a global scale within the cell. Morgan, Marc Alard, and Ali Shilatifard. "(Poly) combing the pediatric cancer genome for answers." Science 340.6134 (2013): 823-824.

occurs. Lowe's main model is the "Tethering and Sequestration" model depicted in Figure 2. This model proposes that the mutation from lysine to methionine stabilizes the binding of PCR2 to the mutant histone, removing the protein's ability to methylate other histones within the cell [9]. This would explain the global reduction in trimethyl marks within mutant cells. Support for this model stems from co-immunoprecipitation (Co-IP) data that showed

enhanced co-purification of H3 K27M with PCR2 compared to Wild-type H3. However, there is also evidence against this model; overexpression of H3 K27M did not show an increased affinity for PCR2 in Co-IP experiments performed by Herz et al. [9]. In fact, *in vitro* nucleosome binding assays have determined that the affinity for PCR2 is roughly equivalent between the mutant and wild-type histones. Clearly, there is still more research to be done before the mechanism of pathogenesis is fully understood. One study was conducted by Lewis et al. to further characterize H3 K27M's involvement in cancer development through sequencing methods and immunohistochemistry [8]. Sequencing traced the mutation back to the H3F3A and HIST3H1B genes while immunohistochemistry further validated the idea that H3 K27M decreases trimethylation globally.

The same study by Pathania et al. [14] used mice to build a pHGG model with the H3 K27M mutation as well as knockout of Trp53, a gene important in suppressing tumor growth. They did this by introducing the mutation into mouse embryonic stem cells when they discovered early on in their study that introducing the mutation to mice postzygotic cells was fatal (a postzygotic mutation is one acquired by the organism during its lifespan). This suggests that tumorigenesis occurs during development *in utero*. However, this study was not able to determine the precise cell state in which tumorigenesis begins. Additionally, previous studies have shown the H3 K27M mutation alone does not drive tumorigenesis. Pathania et al. knocked out Trp53 in combination with inducing the H3 K27M mutation at specific cell states and found that the mutation plus the knockout was able to drive tumorigenesis.

While the research by Pathania introduced the idea that the H3 K27M mutation required only one more additional hit for tumorigenesis to occur, it was conducted in mice embryonic stem cells rather than human ones. Kosuke Funato's study advanced Pathania's model by developing a DIPG model using human embryonic stem cells (hESCs). With this improved model, Funato et al. demonstrated that H3 K27M expression along with Trp53 loss and PDGFRA (platelet-derived growth factor receptor alpha) activation, a gene encod-

ing a protein involved in many cell signaling pathways, transformed the neural progenitor cells to a more primitive stem cell state. In hESCs, the H3 K27M mutation alone was able to increase cell proliferation with knockdown of Trp53 and activation of PDGFRA further increasing cell proliferation rates [6]. Additionally, this study further validates the idea that tumorigenesis only occurs in a highly-specific cell state because tumorigenesis did not occur in fully undifferentiated cells or differentiated cells [6]. This supports the hypothesis that tumorigenesis occurs sometime during neural development.

In a study by Mariella Filbin, single cell analysis of these histone mutant gliomas has revealed that the mutation causes maintenance of a population of cycling, undifferentiated malignant cells [5]. To further characterize H3 K27M neural cells, Filbin performed single cell RNA sequencing. Single cell RNA sequencing is used in biology and genetics research to uncover what is being transcribed within a cell. Transcription levels are an indicator of the proteins present in the cell as well as the level of gene expression. This allows researchers to categorize the H3 K27M cells into specific cell states. Single cell RNA sequencing also allows researchers to compare the K27M cells to normal cell types as well as other malignant cells within the same tumor. This approach also relates single cell states to genetics through inferred copy number variations as well as mutation detection in expressed transcripts. Filbin

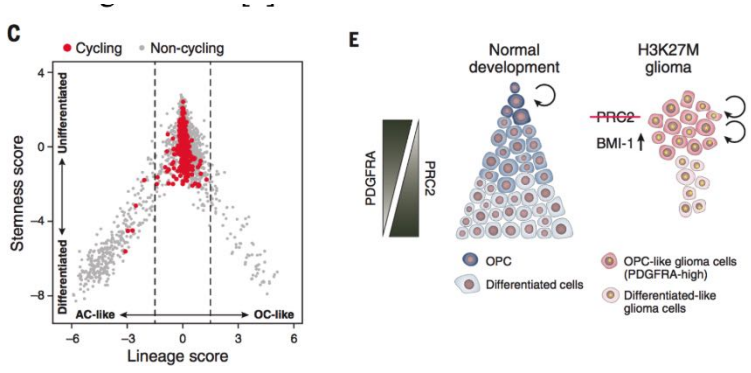


Figure 3: Filbin was able to score malignant cells based on "stemness". Red indicates a positive stemness score. Figure was obtained from [5].

took six samples obtained via biopsy of H3 K27M gliomas and performed single cell RNA seq on them. They then compared the transcriptomes of these samples to known cell types.

What Filbin and her collaborators uncovered was that H3 K27M gliomas contained increased amounts of oligodendrocyte precursor cells (OPC-like) [5]. OPC-like cells occur early in the differentiation pathway to oligodendrocytes. This means these cells are more stem-like. Filbin et al. was able to score these OPC-like cells based on their stemness as shown in Figure 3. These cells have greater tumor-propagation and proliferation properties than their more differentiated counterparts [5]. These cells propagate the disease in patients but now that they have been identified as the predominant population in H3 K27M gliomas, it may be possible to search for therapeutic targets specific to this cell type. Their analysis was in line with the "tethering and sequestration" model mentioned in the Lowe Review. They found that few genes were down-regulated in H3 K27M gliomas, which supports the idea that PRC2 transcriptional repressive activity via trimethylation is repressed by the mutation. Additionally, PRC2 target genes were upregulated [5]. In normal brain cells, PRC2 targets were highly expressed in OPC-like cells and expressed significantly lower in more differentiated cell types. This would suggest that differentiation into oligodendrocytes requires PRC2 repressive activity [5]. This indicates that the H3 K27M mutation stalls differentiation into mature oligodendrocyte cells. Lastly, Filbin concludes that the dominant malignant population in H3 K27M gliomas are OPC-like cells.

2.2 Neural Stem Cells and Differentiation

Since pHGGs and DIPGs seem to occur during early brain development, stem cells are relevant models. Stem cells are defined by their ability to self-renew and differentiate into other specialized cell types [16]. More specialized cells are adept at performing a few key functions. It is important to note that stem cells themselves are not specialized, that is to say that they cannot carry out the same functions differentiated cells can. Every tissue in the human body began as a mass of pluripotent stem cells that received cues to differentiate from the environment. These signals activate the expression of genes that help the cell to adapt to environmental changes.

Because pluripotent stem cells are the precursor to every cell type in the human body including neurons, they provide an excellent model for studying early neural development. Similarly, they provide an excellent model for studying disease progression during development. The ability of a stem cell to differentiate in a controlled manner *in vitro* is key to their usefulness as models for neural development [13].

2.3 Cerebral Organoids

In the past, two dimensional cell culture has been invaluable to expanding knowledge of cell biology. Simple cellular mechanisms can be modelled using two dimensional cell culture methods but when it comes to complex three dimensional cellular structures such as the brain, two dimensional culture methods are incapable of fully recapitulating the properties of these types of tissue. Additionally, obtaining human tissue can be difficult, so an alternative method of modelling the brain is needed [1].

Organoids are an exciting new technology that better mimic living tissue in the lab than any two dimensional cell culture. This is because organoids are essentially three dimensional cell cultures that can better mimic the cytoarchitecture of tissue *in vivo* [10]. They can also be grown into a wide range of tissue types.

Cerebral organoids, or "mini brains", are currently the best way to model brain tissue. They are amazingly able to mimic different regions of the brain as well as recapitulate aspects of human cortical development [7]. Because cerebral organoids are so adept at mimicking human brain development, they expand on stem cell culturing techniques to even better model neural development and disease progression. Most importantly, cerebral organoids are able to model early, human neural stem cells that are the most likely cells of origin of H3 K27M gliomas. This capability is not something that can be found in animal models or conventional cell culture models.

2.4 The EMT Hypothesis

A recently submitted paper in the Haussler Lab showed that single-cell and bulk RNA sequencing analysis revealed that H3 K27M tumors and pre-EMT brain cells share gene expression similarities. The epithelial-mesenchymal transition (EMT) pathway refers to the process that neural epithelial cells undergo to transition into radial glia neural stem cells. These radial glia stem cells are capable of converting into more mature cell types.

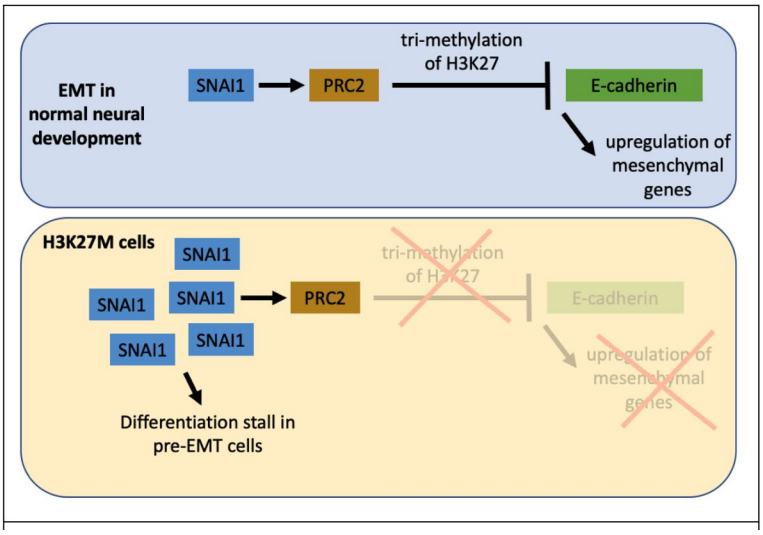


Figure 4: Depiction of the proposed mechanism of the differentiation stall induced by H3 K27M. In normal brain development, PRC2 is not inhibited by the H3 K27M mutation and the H3 K27 residue is trimethylated. SNAI1 then uses this as a signal to continue the EMT pathway. In cells harboring the mutation, trimethylation does not occur and the EMT pathway is stopped leading to the differentiation stall. This figure was obtained from a recently submitted paper in the Haussler Lab.

One of the genes highly expressed in mutant pHGGs is SNAI1, a pro-EMT transcription factor. SNAI1 relies on the protein complex PRC2 for H3K27Me3 to occur and regulate EMT gene expression but in H3K27M pHGGs trimethylation does not occur. Therefore, the abundance of SNAI1 and decrease in other mesenchymal gene expression in H3K27M gliomas indicates that the EMT process is halted and keeps neural stem cells in their current state.

This process also occurs regularly during normal embryonic development [15]. Addition-

ally, it has been found that this pathway is typically upregulated in tumor cells but it is incomplete. This has the consequence of keeping the cells in a stem-like state [15]. Figure 5 shows single cell RNA sequencing data from the Haussler lab organoid assay which was scored on each cell's EMT completeness. It was found that neural epithelium and early radial glia cells did not express a complete EMT pathway while intermediate progenitors, late radial glia, and mature neurons did. It seems that H3 K27M cells exist in a pre-EMT state corresponding to neural stem cells and early radial glia. A comprehensive literature review done by the authors of this EMT paper revealed that H3 K27M gliomas occurred exclusively in pre-EMT cells. This data gives rise to the idea that H3 K27M halts the EMT process and consequently prevents neural stem cells harboring the mutation from differentiating further. Because the cell of origin of H3 K27M gliomas is at least a pre-EMT cell type, this would provide further evidence that the H3 K27M mutation causes a differentiation stall before radial glia but also introduces the idea that the differentiation stall occurs because the mutation inhibits the EMT pathway by preventing SNAI1 deposition of H3K27Me3. This trimethylation is necessary for repressing epithelial genes so that EMT can occur. The Haussler lab has developed a human embryonic stem cell based cerebral organoid model of the development and differentiation of the early cerebrum [4]. This analysis shows that organoids provide an excellent model for early neural development since they are capable of recapitulating the EMT pathway and therefore can model both pre-EMT and post-EMT cell types. Therefore, the cerebral organoid model developed in the Haussler Lab is ideal for investigation of the nature of the differentiation stall. Characterization of Histone H3 K27M mutation in pediatric glioma and brain development using standardized cerebral organoids could drastically improve survivorship by early detection. Understanding the mechanism of pathogenesis may also uncover alternative routes of treatment such as identifying biomarkers specific to mutant cells.

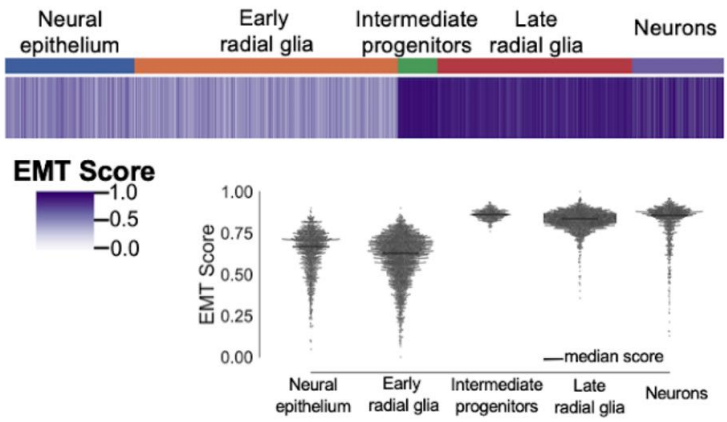


Figure 5: Single cell RNA sequencing data from the Haussler lab organoid assay was scored on each cell's EMT completeness. Neural epithelial and early radial glia cells did not score highly while intermediate progenitors, late radial glia, and neurons expressed more complete EMT characteristics. Neural epithelial and early radial glia cells are more stem-like, giving rise to the idea that stem-like cancer cells amay also be pre-EMT. Data from [4].

2.5 Immunofluorescence Staining

Immunofluorescence staining is a technique used to identify different cell types within a tissue sample with antibodies. The first step when staining is to freeze and preserve the tissue in tissue freezing medium. Then the sample is sliced into 16-20 μM sections using a cryostat, a machine that allows the user to slice with precision by keeping the sample at -20 °C. These sections are then placed on a microscope slide and are ready for staining. The sample is washed with a primary antibody that has an affinity for a protein marker specific to a certain cell type and washed again with a secondary antibody with affinity for the primary antibody. The secondary antibody has a fluorescent tag that allows visualization under a fluorescent microscope. In this project, the tissue samples that I will be staining are from cerebral organoids.

2.6 Single-Cell RNA Sequencing Pipeline

Single cell RNA sequencing is a method that provides you with quantification of which genes are being transcribed within individual cells. In this project, cerebral organoids are

cultured to the physiologically relevant timepoints during neural development and then harvested. Then, the organoids are dissociated to single cells and prepped for single cell RNA sequencing.

After receiving the raw data, I will use Scanpy, a toolkit in Python that enables quick preprocessing of the data. I will also be using an unpublished tool developed by Lucas Seninge, a graduate student in the Haussler Lab, called ScoreCT to categorize the cells in the sample by cell type based on gene expression. This computational approach to identifying the cell types with my organoids will be used to validate the visual data I produce through immunofluorescence staining.

3 Modelling the Effects of the Histone H3 K27M Mutation on Neural Stem Cell Differentiation

3.1 Development of Human Embryonic Stem Cell Lines Harboring the H3 K27M Mutation

Prior to my involvement in this project, human embryonic stem cell lines were developed in the Haussler Lab to model the development of pHGGs and DIPGs. The first cell line

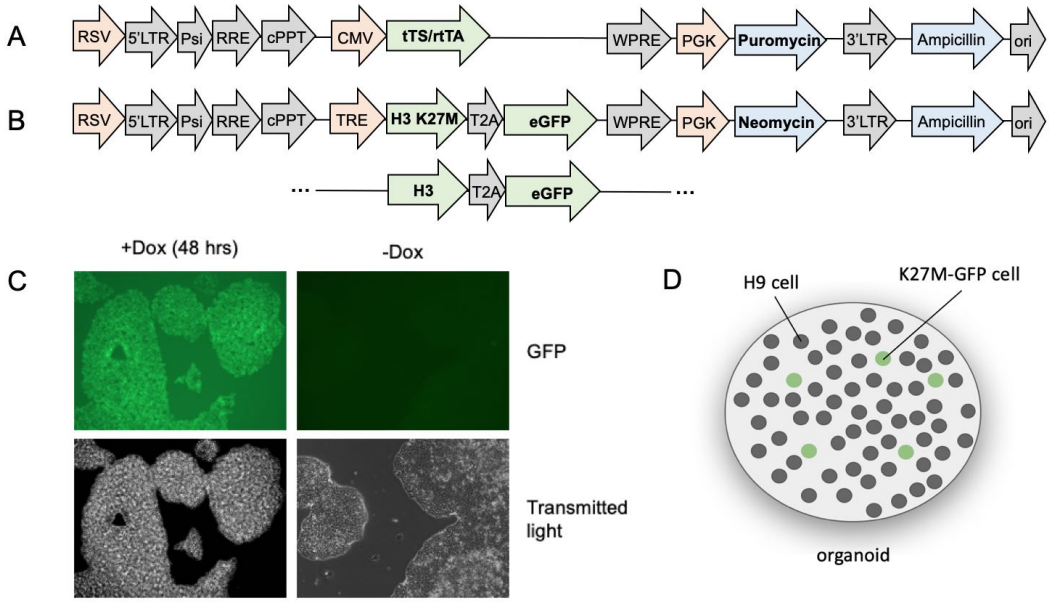


Figure 6: Schematic of single integrant cell lines used in this experiment. In the absence of doxycycline, tTS binds to TRE promoter to actively suppress gene transcription whereas rtTA is unable to bind to TRE promoter. In the presence of the inducing drug, tTS dissociates from TRE promoter whereas rtTA binds to TRE promoter to activate gene transcription. Figure created by Lauren Sanders.

harbors the histone H3 K27M mutation and is referred to as 5K30. The second cell line is

my wild-type histone H3 control and is referred to as 5H1. The last cell line I am working with is a clone of a commonly used hESC cell line, H9. This last cell line is named C5, for clone 5 and was transduced with the tTS/RTTA construct shown in Figure 6 A. The 5K30 and 5H1 cell lines were engineered via viral transduction to express either the wild-type histone or mutant histone in the presence of doxycycline. This was done by transducing the C5 cell line either with a viral plasmid encoding the H3 K27M mutation or the wild-type H3 sequence shown in Figure 6 B. In this assay, doxycycline is an antibiotic that binds to the tTS construct of the viral plasmid and causes it to dissociate from the TRE promoter. This has the effect of allowing the rtTA sequence to bind to the TRE promoter and activating expression of the mutant or wild-type histone. In the absence of doxycycline, the tTS sequence remains bound to the TRE promoter to actively suppress transcription as depicted in Figure 5A and 5B. The 5K30 and 5H1 cell lines were also designed to express GFP along with the mutant or wild-type histone so that they are identifiable under a fluorescent microscope.

These cell lines were generated through lentiviral transduction by Lauren Sanders and stored until more organoids are needed for this project. Lentiviral transduction is a method to insert the plasmids depicted in Figure 6A and 6B into human embryonic stem cells using a virus that integrates the plasmid DNA into the host cells' genome.

3.2 Culturing Human Embryonic Stem Cells (hESCs)

In the Haussler Lab, I learned proper stem cell culture techniques that have assisted me in this project. hESCs are cultured in a sterile cell culture dish in a biosafety cabinet to prevent any outside contaminants from reaching the cells. Stem cells only proliferate in very specific conditions so it is necessary that they are cultivated in a very controlled environment. Factors such as pH, carbon dioxide levels, and temperature all play a role in the health of the cells. Therefore, the cell plates are stored at 37°C in incubators located in the lab. These incubators are capable of monitoring the temperature and carbon dioxide levels of the stem cells. When culturing hESCs, it is essential to maintain their pluripotent status. The hESCs

are fed with media containing amino acids, carbohydrates, vitamins, and other proteins needed for the cells to proliferate. Growth factors are added to the media to further regulate proliferation and differentiation. The media used in this project was StemFlex. Stemflex contains growth factors, particularly a stabilized form of fibroblast growth factor 2 (FGF2), that facilitate the maintainence of the hESC pluripotent state. This media is replaced every other day and double fed on the weekend to provide a stable supply of nutrients to the cell culture while removing toxic waste byproducts from the dish. Earlier culturing protocols required daily feeding, but the addition of FGF2 enables the ability to feed hESCs every other day.

When the cell colonies become confluent (when the cells cover most of the floor of the plate), it is necessary to passage or divide the cells into separate plates. This prevents overcrowding and gives the cells enough space to continue proliferating. This was done by coating new plates with vitronectin, a protein that allows the cells to adhere to the new plate. Without vitronectin, the cells would not adhere and die. The old plate is treated with 0.5 mM Ethylenediaminetetraacetic acid (EDTA) in PBS for 3-6 minutes. EDTA is a chemical commonly used in stem cell culture to detach cells from the plastic dish. After EDTA detachment, the EDTA is removed and new media is added to the plate. This media is used to spray the cells off the plate so that they are floating in the media. At this point, the cells are triturated (broken down into smaller colonies) and split into the vitronectin coated plates. This process is repeated as often as necessary.

3.3 Aggregation of Human Embryonic Stem Cells into Cerebral Organoids and Cryopreservation

Aggregation refers to the technique of taking a 2-dimensional stem cell culture and forming the cells into organoids, or 3-dimensional culture. Aggregation of the cerebral organoids was done by Lauren Sanders with stem cells that were passaged only a few times. Each chimeric organoid was comprised of either 90% or 99.5% C5 cells with 5K30 or 5H1 cells

making up the rest of the organoid. This was done so that the 5H1 and 5K30 cell populations were easily distinguishable from the rest of the organoid. It also allows for easy comparison of the development of the mutant GFP positive cells versus the normal C5 cells which should be developing normally. The 5H1 cell line was created as a control for having an extra copy of histone H3. Doxycycline was added a week after the aggregation so that the stem cells could form organoids without the pressure of the histone H3 K27M mutation.

After aggregation and culturing the cerebral organoids until the desired time point, it is time to harvest the organoids from their plate in the incubator. This was done with wide-bore P1000 pipette tips. I placed each organoid into its own Eppendorf tube and removed the extra media that naturally comes from the organoid with a normal P1000 pipette tip. I then added 200 μL of 4% PFA into each tube and incubated the organoids for 15-30 minutes. This step helps preserve the structural integrity of the sample by causing covalent crosslinking between protein complexes within the sample. I then rinsed the organoids three times with PBS buffer to remove as much PFA as possible. I then incubated each organoid in 30% sucrose overnight. The sucrose draws out moisture so that the organoids are easier to section with the cryostat. The organoids can then better retain their structural integrity during sectioning. I then used wide-bore pipette tips to place the organoids along with 100 μL of 30% sucrose and placed each one in its own plastic tray with tissue freezing medium (TFM). I then covered the organoid with more TFM and placed them in the -80 °C fridge in the lab. These organoids are now ready to section whenever it is convenient to do so.

4 Visual Analysis of Cell States in Cerebral Organoids with Immunofluorescence Staining

4.1 Cryostat Sectioning

After harvesting the organoids, I freeze them in TFM and store them at -80 °C. Then, I slice the organoids into 16-18 μM sections and apply them to microscope slides using a machine called a cryostat. A cryostat allows the user to section the sample with high precision using a specifically manufactured razor and control both the inside containment temperature and the temperature of the block the user is slicing on. This technique is commonly used in histology, the study of microscopic anatomy, to visualize biological structures within tissue samples.

Cryostat sectioning is a highly precise method that requires a lot of practice to figure out what works for you. There is a general technique for acquiring high quality sections from a tissue sample but there is a lot of variation between users. Due to the nature of sectioning, it is common to lose significant chunks of your sample while slicing. This issue is exaggerated in the case of small organoids. I spent the first couple months on this project practicing my cryostat sectioning technique on organoids from previous experiments in the Haussler Lab until I felt that I could obtain quality sections.

4.2 Immunofluorescence Staining of Cerebral Organoids

In order to investigate the effects of the H3 K27M mutation on neural development, I am using a cerebral organoid assay developed in the Haussler Lab. I am working with Lauren Sanders, a PhD student in Professor David Haussler's wet lab, who has developed the previously mentioned brain organoid assay in which a certain percentage of neural stem cells in the organoid harbor the H3K27M mutation along with a green fluorescent protein (GFP) sequence (0.5-10%).

I am currently optimizing the conditions for cryopreservation and immunofluorescent (IF) staining of cell markers in these organoids at weeks 2, 5 and 10. I will be staining for cell type specific protein markers including (but not limited to) PAX6 (radial glia cells), RELN (layer 1 Cajal-Retzius neurons), VIM (radial glia neural stem cells), TBR2 (intermediate progenitor cells), CTIP2 (deep layer neurons), SOX2 (undifferentiated pluripotent stem cells, neural epithelium, and radial glia neural stem cells), and SATB2 (upper layer neurons). By staining for these protein markers, I will be able to identify which cell states are present in the mutant and control organoids at different time points relevant to neural development. In this way, I hope to characterize the nature of the differentiation stall induced by the H3 K27M mutation.

I am performing these experiments based on Field et al., 2019 [4]. I will use the immunofluorescence data I produce to visually examine the patterns of differentiation in H3 K27M cells as compared to the surrounding wild-type neural stem cells. Changes in cryopreservation include performing a titration of sucrose levels (10% sucrose, 20% sucrose, 30% sucrose) and varying fixation processes (extending the amount of time allotted to have the samples sitting in 4% paraformaldehyde (PFA)). I will section the organoids into 16-18 μM slices using a cryostat and I will adhere the slices to glass microscope slides. Finally, for addressing the differences in IF staining, I will be conducting different numbers of washes between antibody additions, incubation times for both primary and secondary antibodies, and titrating the concentrations of the antibodies added.

As mentioned before, immunofluorescence staining allows visual analysis of cell states present within a tissue sample. This is done using two separate antibody washes with a series of 0.05% tween phosphate-buffered saline (PBST) buffer washes in between. The first wash uses primary antibodies with affinity for a protein marker corresponding to a specific cell state and the second wash involves secondary antibodies with affinity for the primary antibodies. Additionally, the secondary antibodies fluoresce at certain wavelengths of light so that they are visible under a fluorescent microscope.

For my first attempt at performing immunofluorescence staining, I performed serial dilutions on the primary and secondary antibodies in blocking solution (3% bovine serum albumin in PBS) to determine which concentrations provide the best results. The primary antibodies used were anti-GFP (Abcam product number ab13970) and anti-PAX6 (Abcam product number ab195045). The secondary antibodies used were AlexaFluor 488 (ab150169) and AlexaFluor (ab150075). I also chose to stain for radial glia cells (PAX6) and GFP because these seemed to be the best characterized antibodies the Haussler Lab had in stock. The first step in the protocol is to obtain your sections from -80°C and letting them warm to room temperature. I chose to stain week five 5K30 and 5H1 samples in parallel so that I could compare afterwards. These organoids were aggregated on September 16, 2019 and harvested on October 29, 2019. To remove the tissue freezing medium from the microscope slide and the sample, I washed the samples by placing 200 μL of PBST over each sample for 5 minutes. I repeated this process three times while letting the sample dry on a paper towel but not letting the section turn white. At this point, I incubated the samples in 4% PFA in PBS buffer for thirty minutes. Without this second fixation step, the sample would diffuse and fall apart during the washing steps in this protocol. After this step, I performed another set of three PBST washes and incubated the samples in 50 μL of permeabilizing solution (0.15 mL TritonX in 100 mL of PBS). The purpose of permeabilizing solution is to allow antibodies to access the interior of the sample. This is important if the protein you are staining for is intracellular as the fixation step leaves the cell membrane intact. After another series of three PBST washes, I incubated the samples in 50 μL of blocking solution for two hours. Blocking solution is used to reduce nonspecific binding of the antibodies. After this incubation, I added primary antibodies to each sample at dilutions of 1:250, 1:500, and 1:1000. The samples were then incubated overnight at 4 °C.

The next day, I performed another series of three PBST washes and applied the secondary antibodies at a dilution of 1:250 for two hours. I then performed another set of PBST washes and applied 50 μL of mounting medium and a glass cover slip. This was then allowed to dry

overnight.

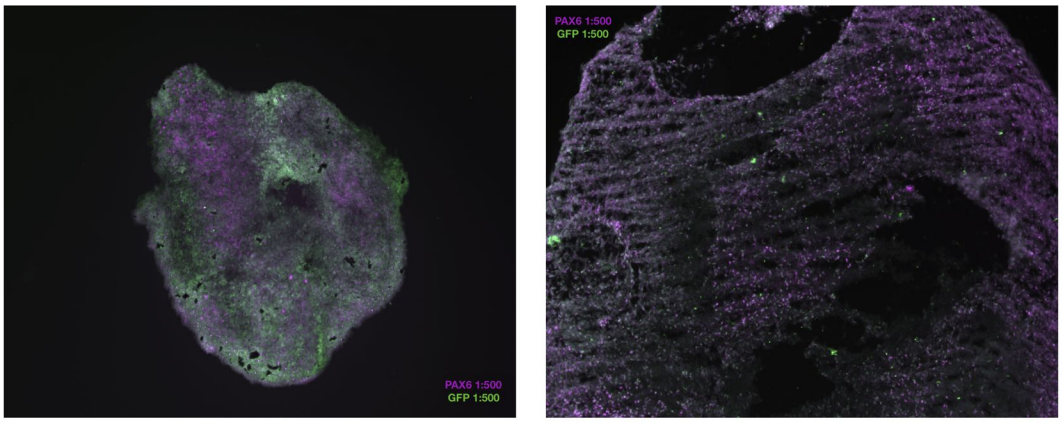


Figure 7: A.) Mutant GFP Organoid, stained for Radial Glia/GFP, 10X magnification B.) Control/"normal" cerebral organoid, stained for Radial Glia/GFP, 10X magnification

My preliminary results of my first stain are depicted in Figure 7. These early results were promising as the mutant 5K30 organoid seemed to have clear separation between the radial glia (purple) and mutant GFP positive (green) cell populations suggesting that the GFP positive cells did not differentiate into radial glia. The wild-type organoid did not have this characteristic and seemed to be predominantly radial glia suggesting that the GFP positive cells were indeed differentiating as expected. This was in line with my hypothesis that the differentiation stall occurs between neuroepithelial cells and radial glia. Encouraged by these results, I proceeded to stain for GFP and TBR2 using the same conditions as before.

Upon closer examination of each single channel (AlexaFluor 647 or 488 alone) with Sofie Salama and Ben Abrams (Director of the Microscopy Center at UCSC) and higher magnification, it became apparent that there was too much nonspecific staining masking the relevant signals in the image. To reduce the amount of nonspecific staining and tease out the real signals, I increased the concentration of BSA in the blocking solution to 10% and I drastically decreased the secondary antibody concentration to 1:1000 or 1:2000.

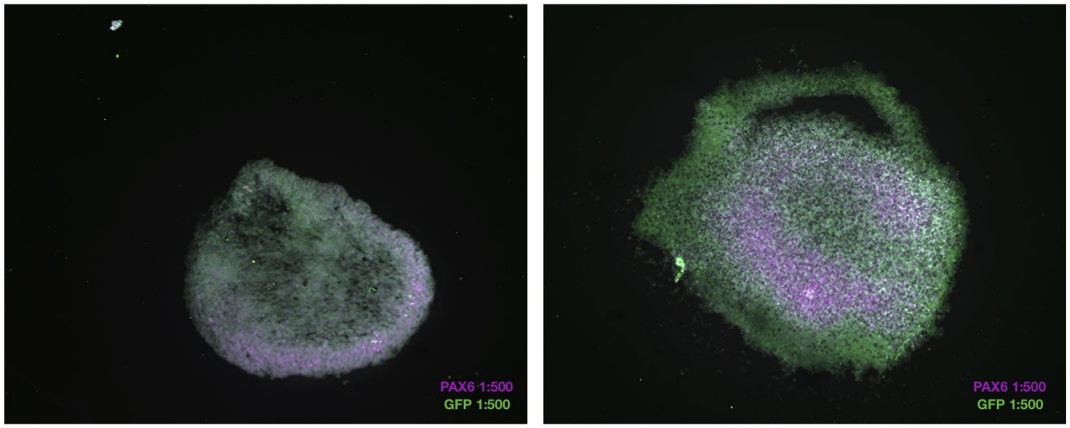


Figure 8: A.) Mutant GFP Organoid, stained for Radial Glia/GFP, 10X magnification B.) Control/ "normal" cerebral organoid, stained for Radial Glia/GFP, 10X magnification.

I chose to stain for PAX6 again since I did not fully trust the signals I was getting from my first stains. At this point I did not have more 5H1 samples to stain so I only stained 5K30 samples. I had slightly better results but there was still a lot of nonspecific staining, especially in the case of GFP. I believe that this is because GFP is already fluorescent so you can see it under the microscope even without IF staining.

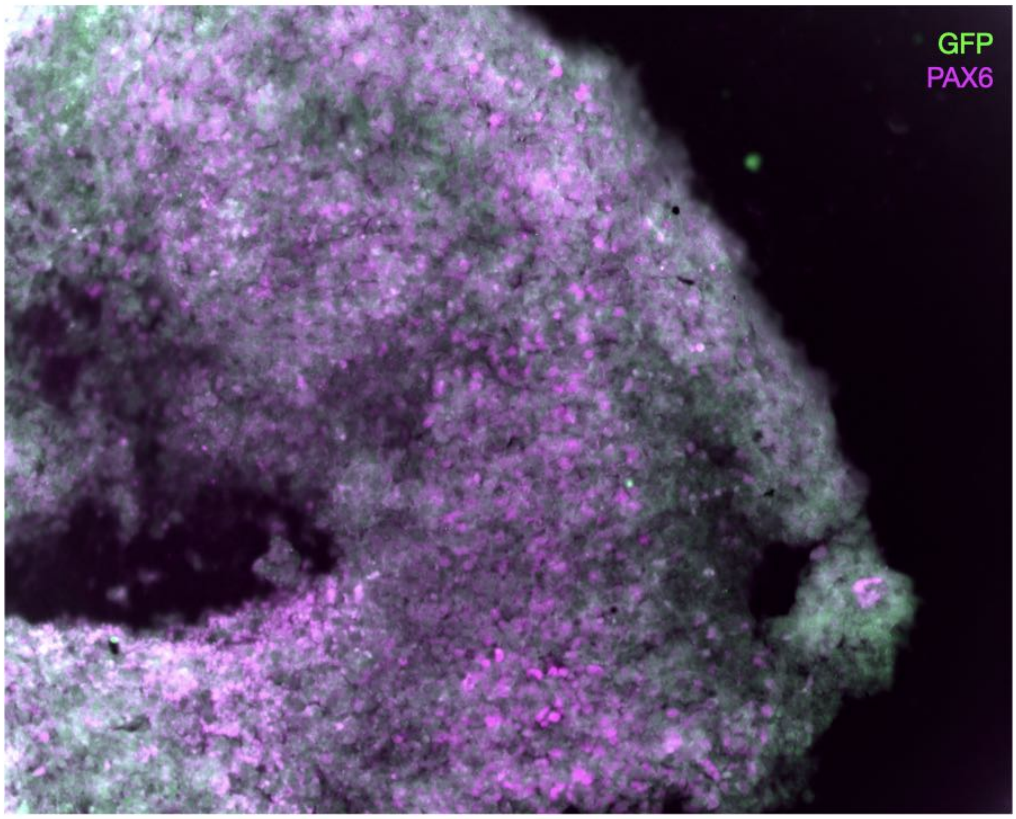


Figure 9: 5K30 organoid stained for PAX6 and GFP, 20x magnification. This sample shows improved removal of nonspecific staining over figures 6 and 7. There is higher specificity for PAX6 shown by distinct clusters of purple.

I have been staining for GFP anyways because the 4% PFA fixation step can kill off most of the GFP fluorescence. Due to the large amount of nonspecific signal, I started to experiment with staining samples with and without GFP to see if it was even necessary since a lot of things (dust, small particles, etc.) autofluoresce at 488 nm, the wavelength of light used to visualize GFP under the microscope. I also started using a rocker during the PBST wash steps in an effort to remove all antibodies that are binding to things other than the protein markers of interest. I then trimmed the number of initial wash steps before PFA fixation to one wash because I believe that three washes was compromising the structural integrity of the samples and even removing a lot of the cells from the sample. Following the recommendation of Sofie Salama, I incorporated DAPI into my stains as well. DAPI

(4',6-diamidino-2-phenylindole) is a very simple stain that targets nuclear DNA. It is used as a quality control to ensure that proper cell morphology is present in the sample.

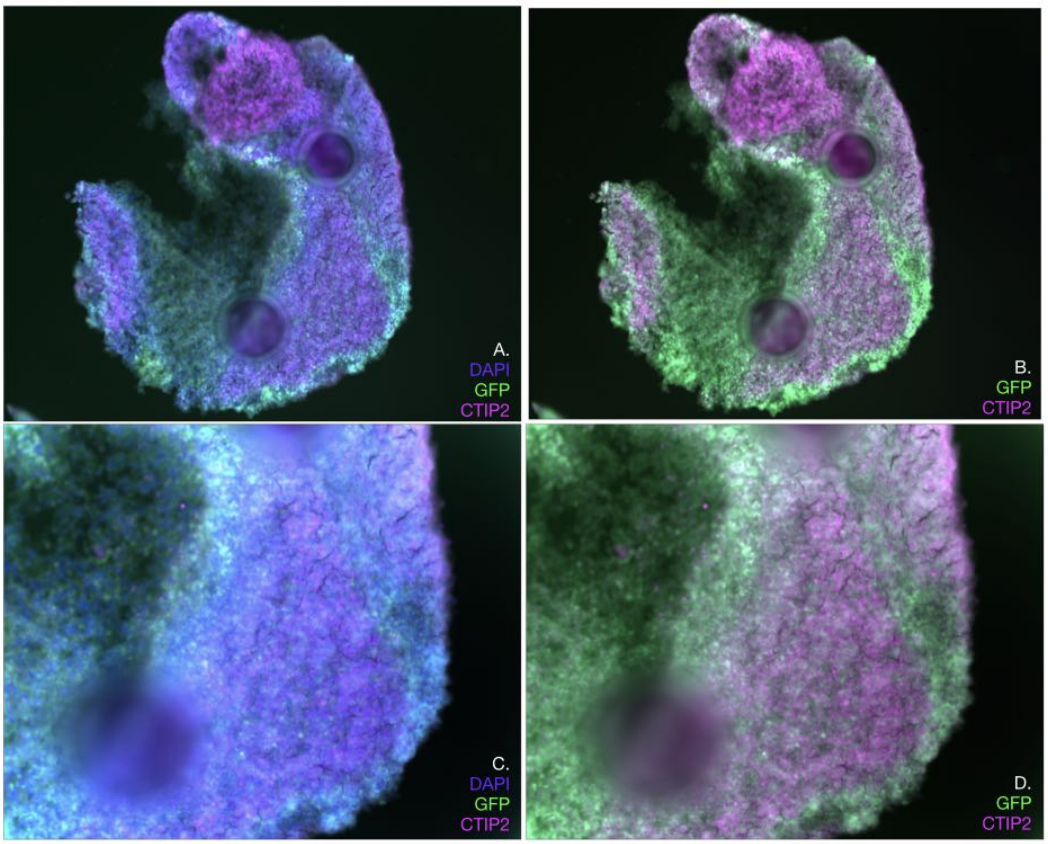


Figure 10: Week 5 0.5% 5K30 organic stained for CTIP2, 1:1000 dilution. A and B were taken at 10x magnification and C and D were taken at 20x magnification. A and C show the nuclear stain, DAPI. B and D show the section without DAPI.

The results of all these adjustments to the protocol are depicted in figure 10. Figure 10 was especially encouraging as it depicts clear separation between AF647 (purple) and AF488 (green). This would indicate that the mutant 5K30 cell populations are not differentiating into deep layer neurons (the cell type the CTIP2 protein marker corresponds to), supporting the hypothesis that the H3 K27M mutation stalls differentiation. However, my hypothesis is I expect the differentiation stall to occur on the pathway to radial glia so I aimed to stain for PAX6 and TBR2 next now that it seemed I had optimized the IF staining protocol. Additionally, Figure 10 depicts a week 5 5K30 organoid so I needed to stain a 5H1 organoid

that was at a similar stage of development.

Confident that I was now capable of producing high quality IF images, I sought to stain a week 5 5H1 organoid for CTIP2 in order to compare with the 5K30 organoid shown in Figure 10. This 5H1 organoid is shown in Figure 11.

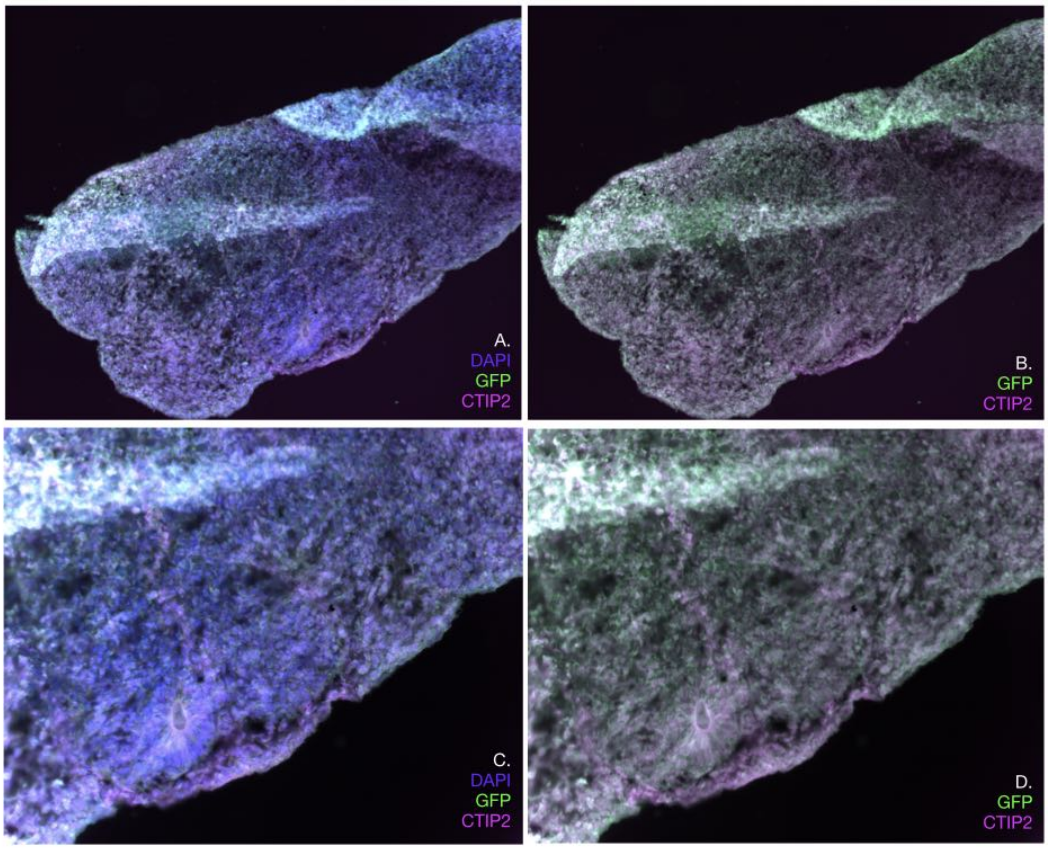


Figure 11: Week 5 0.5% 5H1 organic stained for CTIP2, 1:1000 dilution. A and B were taken at 10x magnification and C and D were taken at 20x magnification. A and C show the nuclear stain, DAPI. B and D show the section without DAPI.

For this most recent stain, I followed the exact protocol I used to obtain the images in Figure 10 but with differing results. There is a lot less separation of cell populations in this 5H1 organoid relative to the 5K30 one. It is also apparent that the fluorescent signal is weaker in Figure 11 compared to Figure 10. My hypothesis is that the signal is weaker because there are less cells in Figure 11, thus resulting in a thinner section. While the results of the stain depicted in Figure 11 are not as satisfactory, there is still information to obtain

from this sample. Upon examination of the samples shown without DAPI, it seems that there is a much higher amount of AF647 signal relative to AF488. This would indicate that the control 5H1 organoid in Figure 11 is differentiating into deep layer neurons as expected. Additionally, the AF647 signal overlaps significantly with the AF488 signal meaning the individual 5H1 cells are differentiating as expected. In contrast with Figure 10, there is very clear separation of AF647 and AF488 signals. This shows that the individual 5K30 cells are not differentiating into deep layer neurons to the same extent the 5H1 cells are.

4.3 Digital Filtering of Immunofluorescence Data

At this point, I was unable to continue performing IF stains due to the shelter-in-place order from the COVID-19 pandemic. Fortunately, I received digital filtering tips to reduce the amount of nonspecific staining/autofluorescence prior to the order from Ryan Hoffman, a PhD student in the Haussler Lab. Armed with this new knowledge, I began taking a second look at the data I already have. It is apparent that when taking a look at my first stain

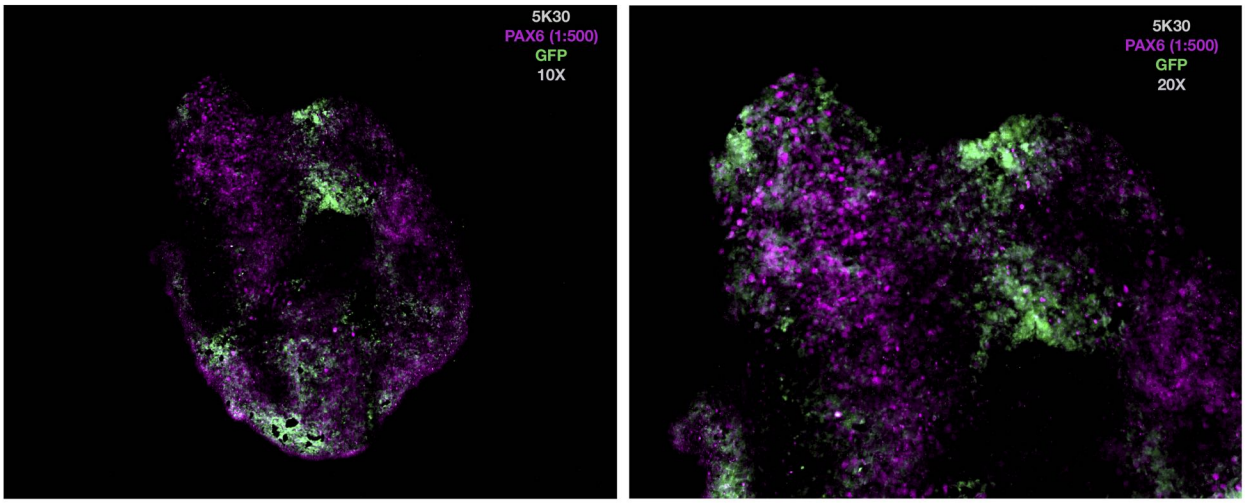


Figure 12: This figure depicts the same week 5 0.5% mutant organoid depicted in 7. With the use of digital filtering, the signal is much cleaner. It is apparent that the mutant GFP cells (green) are not differentiating into radial glia (purple).

(Figure 7) with the digital filtering shown in Figure 12, there is clear separation between

the radial glia and mutant GFP cells harboring the H3 K27M mutation. This suggests that the mutant GFP cells are not differentiating into radial glia, supporting my hypothesis that the differentiation stall occurs along this pathway. In comparison, the wild-type organoid

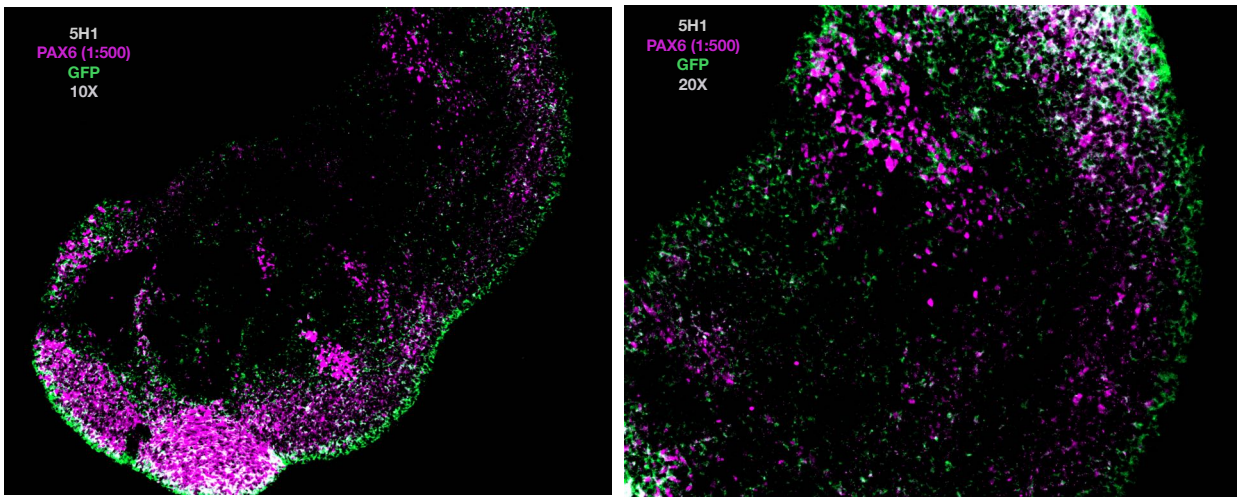


Figure 13: This figure depicts the same week 5 0.5% H3 wild-type organoid stained for radial glia (purple) and GFP (green). These images were filtered digitally. There is a higher level of costaining (white), meaning the purple and green overlap. This means the GFP positive cells are differentiating into radial glia cells. Although there is more overlap, it is difficult to determine if this is because this organoid harbors the wild-type histone rather than the mutant one because there is still some separation of cell populations.

shown in Figure 13 has more overlap between cell populations. This would suggest that the wild-type cells are indeed differentiating into radial glia. However, there is still some separation of cell populations visible so it is unwise to draw conclusions from these two images. Staining of mutant and control organoids at a later timepoint is needed to further validate the hypothesis that the differentiation stall occurs before radial glia.

As I seeked to optimize the staining protocol, I obtained quite a few images of different stains not depicted in section 4.2. The reason I did not show these images earlier is that I believed them to be of insufficient quality for visual analysis. When I learned how to digitally filter these IF images, I had hope that there was information to be gleaned from my old stains. After filtering the radial glia stains shown in Figures 12 and 13, I did the same filtering on week 5 organoids stained for intermediate progenitor cells (TBR2). Shown

in Figure 14 is a 0.5% H3 wild-type organoid. The GFP positive cells seem to outline the organoid and there is also some costaining. This is as expected since the wild-type organoid should develop normally into intermediate progenitor cells.

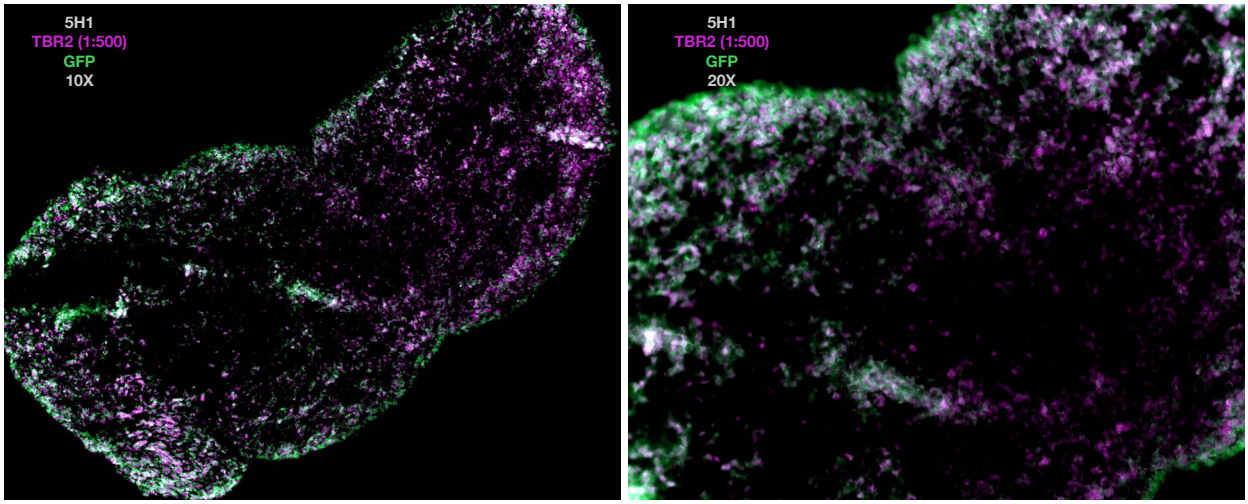


Figure 14: This figure depicts the a week 5 0.5% H3 wild-type organoid stained for intermediate progenitor cells (purple) and GFP (green). These images were filtered digitally. There is some costaining (white), meaning the purple and green overlap. This would suggest that the GFP positive cells are differentiating into intermediate progenitor cells.

I also filtered an image of a comparable week 5 H3 K27M organoid for comparison. This is shown in Figure 15. Unfortunately, large pieces of the organoid are missing so it is difficult to make any meaningful comparison between the wild type and mutant TBR2 organoids. The sections missing are likely due to sectioning issues which I wanted to address prior to the shelter-in-place order. My hypothesis is that the freezing step after harvesting the organoids from their plates in the shaking incubator is an extremely important step. Up until this point, I had been harvesting the organoids and incubating them in 30% sucrose overnight to dehydrate them so that the cells would not rupture when exposed to low temperatures. I would take up 100 μL of 30% sucrose along with the organoid and deposit all of this into TFM. When I was sectioning, I would notice a significant difference in texture between the sucrose and TFM which I believed to be warping my sections. I planned to start mixing the sucrose and TFM together to create a more uniform solution to address this issue. Sadly, I

will not be able to test this until the shelter-in-place order ends.

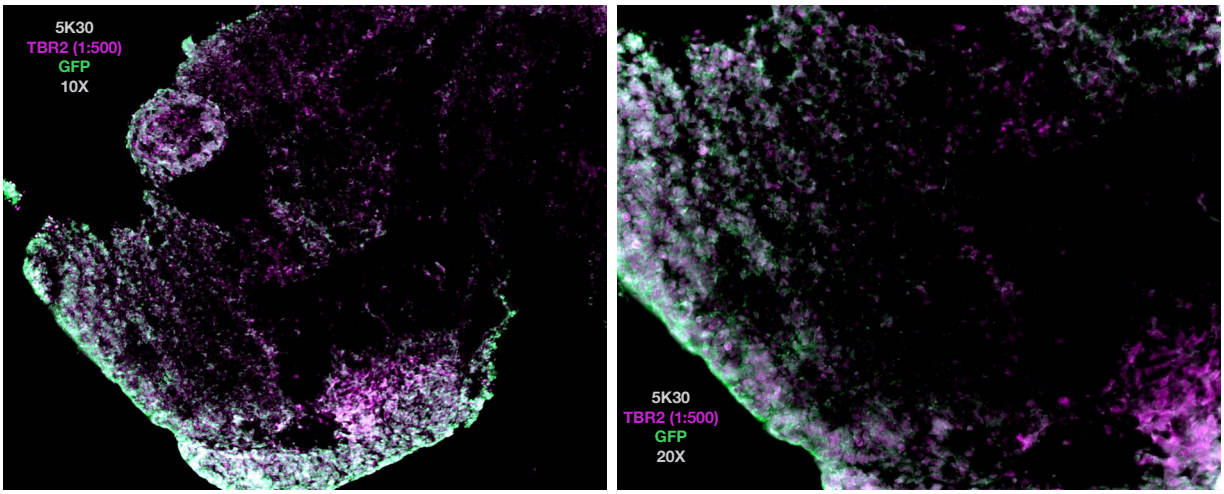


Figure 15: This figure depicts a week 5 0.5% mutant H3 K27M organoid stained for intermediate progenitor cells (purple) and GFP (green). These images were filtered digitally. While there is overlap between the GFP and TBR2 cell populations, it is difficult to draw any conclusions about the differentiation patterns of these cells because large sections of the organoid are missing. This is visible in the 10X magnification image on the left.

I continued to go back and digitally filter my existing data and next took a look at the mutant organoid shown in Figure 10. At the time, I believed this to be my most successful stain but after consulting with Ryan Hoffman, I became aware that there was still far too much nonspecific staining possibly as a result of cryostat sectioning issues. I believe that I was losing the structural integrity of my cells during the freezing step.

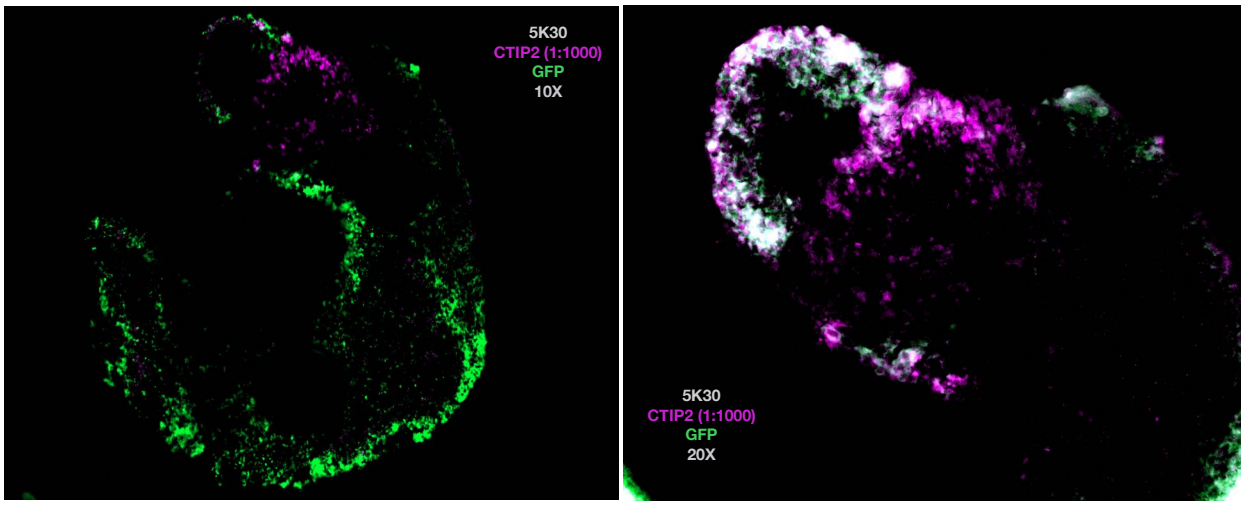


Figure 16: This figure depicts the same week 5 0.5% H3 K27M organoid stained for deep layer neurons (purple) and GFP (green) shown in Figure 10. The deep layer neurons are localized in one region of the organoid which is as expected since this cell type only just begins to form at week 5. When looking at the organoid under 20X magnification, costaining between the mutant GFP cells and the deep layer neurons in clearly visible. This suggests that the mutant cells are differentiating into deep layer neurons.

After filtering, I was able to tease out some real signals from the image shown in Figure 16. The deep layer neurons (CTIP2) are localized at the top of the organoid which is expected since this type of neuron is usually not present or just beginning to appear at week 5. I then took a look at a wild-type week 5 organoid stained for deep layer neurons.

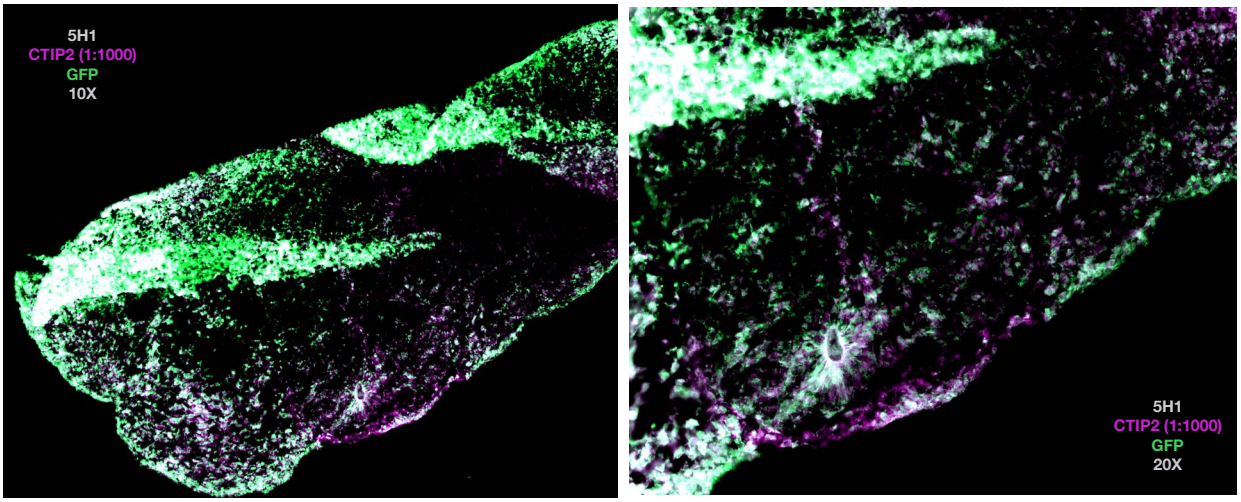


Figure 17: This figure depicts the same week 5 0.5% H3 wild-type organoid stained for deep layer neurons (purple) and GFP (green) shown in Figure 11. Again, there are clearly problems with sectioning as the organoid is folded over itself in a couple areas. Also, there are some deep layer neurons present but in low quantities which is expected for this timepoint.

Figure 17 shows this control organoid but the organoid is folded over itself due to the sectioning process. It shows some deep layer neurons as expected. Because of the week 5 timepoint, I cannot make any conclusions about the cTIP2 differentiation pattern differences between H3 K27M cells and wild-type H3 cells since deep layer neurons only begin to show themselves at week 5. Analysis of organoids stained for CTIP2 at week 10 would provide a better picture of the differentation patterns of these cells.

With the use of digital filtering, I have been able to tease out the real signal out of my previous IF staining data and uncovered that the source of the nonspecific staining could be an artifact of cryostat sectioning issues, specifically dehydration with 30% sucrose and freezing in TFM. I aim to test this hypothesis as soon as I am able to.

5 Computational Analysis of Single-cell RNA Sequencing of Cerebral Organoids

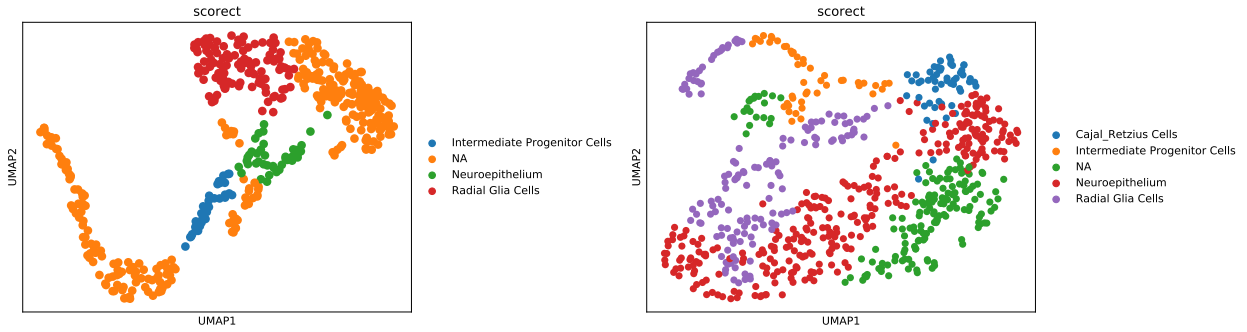
In conjunction with IF staining, I also analyzed single-cell RNA sequencing data of week 2 5H1 and 5K30 cerebral organoids. I used an unpublished cell type annotation tool developed in the Haussler Lab by Lucas Seninge, a PhD student. To run, the tool requires the user to do preprocessing on the single-cell RNA sequencing data using Scanpy. Scanpy is a package developed by the Theis Lab at the Institute of Computational Biology in Munich. The preprocessing performs louvain clustering on the cell data and then ranks the clusters by gene.

I wrote a preprocessing script based on the instructions from Scanpy that can be found in the appendix. The output of this script is the ".h5ad" file needed as the input for the annotation tool. The other input required is gene marker file provided by the user. I used a marker file containing a list of cell types and their associated gene markers. This is depicted in Figure 18.

Neural Stem	Intermediate	Radial Glia C	Neurons	Cajal_Retzius	Neuroepithelium	
ABCG2	EOMES	SLC1A3	CHAT	RELN	HES3	
CXCR4	HES1	FABP7	ISL1	EOMES	NR2F1	
FGFR4	SOX2	NES	ISL2	NHLH1	LIN28	
FZD1		PAX6	NEUROG2	LHX9	OCLN	
GLUT1		SOX2	OLIG2		CDH1	
NOTCH1		VIM	HB9		SOX10	
NOTCH2		GLI3	MAP2			
CD15		PDGFD	NEFL			
CD133		HES1	NEFM			
BMI1		PTN	DLG4			
SMARCA4		EMX2	NR1A2			
FABP7		NNAT	PPP1R1B			
MSI1		TLE4	GAD2			
MSI2NES			GAD1			
NEUROD1			VAMP2			
GNL3			VMAT2			
SOX1			TH			
SOX2			NR4A2			
SOX9			PAX2			
VIM			PAX6			
NEUROD6						
SOX9						
VIM						
NEUROD6						

Figure 18: This is .csv file used as an input for the Cell Type Annotation Tool developed in the Haussler Lab. The cell types specified in each column are neural stem cells, intermediate progenitor cells, radial glia, neurons, Cajal-Retzius neurons, and neuroepithelium cells. Under each cell type is a list of genes corresponding to that cell type. The tool takes this input and classifies each group in the single-cell RNA sequencing data as a particular cell type based on the gene expression in the data.

The cell types I expected to find the sequencing data were neural stem cells, radial glia, Cajal-Retzius neurons, and neuroepithelium cells. After generating the marker file and running the data through the preprocessing script, I ran the annotation tool on the Scanpy-processed sequencing data. This is shown in Figure 19.



(a) Week 2 Wild-type 5H1 organoid sequencing data.

(b) Week 2 H3 K27M 5K30 organoid sequencing data.

Figure 19: (a) depicts the cell types present in a week two wild-type organoid. Intermediate progenitor cells, neuroepithelium, and radial glia cells are the classified cell types present. The predominant population of cells is unclassified. (b) shows the cell types present in the 5K30 organoid. These are Cajal-Retzius neurons, intermediate progenitor cells, neuroepithelium, and radial glia cells. A much smaller proportion of cell are unclassified in this data set.

Unfortunately, the sequencing data used in this computational pipeline was shallow, meaning that there were few numbers of reads per base. I believe that this affected the output of this pipeline so the next step would be to perform this on a deeper data set. The reason that the data set was not higher quality could potentially be attributed to the library preparation and cells that were sent to be sequenced.

6 Conclusion and Future Directions

In this project, I have developed and optimized an IF staining protocol to examine the differentiation patterns of neural stem cells harboring the histone H3 K27M mutation within cerebral organoids. I have stained five week old cerebral organoids for the following cell states in the neural differentiation pathway with the hypothesis that the mutation stalls differentiation between neural stem cell and radial glia: PAX6 (radial glia), TBR2 (intermediate progenitor), and CTIP2 (deep layer neuron). After much trial and error, I have been able to provide evidence to support my hypothesis that the differentiation stall does occur before radial glia cells but I have been unable to characterize the exact location and nature of the differentiation stall. This is because I need to grow and harvest more cerebral organoids for staining since I used many of them optimizing the IF protocol.

6.1 Future Immunofluorescence Staining

To further pinpoint the location of the differentiation stall, I will need to complete additional IF stains for SOX2 (undifferentiated pluripotent stem cells, neural epithelium, and radial glia neural stem cells), SATB2 (upper layer neurons), VIM (radial glia neural stem cells), and RELN (layer 1 Cajal-Retzius neurons). I expect that the differentiation stall occurs near the intermediate progenitor step in the neural differentiation pathway based on my preliminary results. However, I would like to perform more TBR2 stains with a TBR2 antibody from a different manufacturer since the TBR2 antibody in stock in the Haussler Lab has performed poorly and does not bind well.

I also have ideas to further remove nonspecific staining and autofluorescence. These include a 100 mM glycine quenching step after the first 4% PFA fixation or using an ice-cold methanol fixation step rather than 4% PFA. This is because PFA has properties that could be contributing to the observed autofluorescence.

6.2 Single-Cell RNA Sequencing

To supplement the visual data produced from IF staining, I have also modified a preprocessing script from Scanpy and combined it with a cell type annotation tool developed in the Haussler Lab on week 2 mutant and wild-type organoids. The pipeline requires the user to provide a cell type annotation file containing genes corresponding to different cell types. The pipeline runs smoothly but the quality of the data for the week 2 organoids is poor so I would like to run the tool on a better data set as well as edit the annotation file so that it can annotate the single-cell RNA sequencing data more accurately. I will also be running the tool on data from week 5 organoids so that the data is more comparable with the IF data I already have.

Due to the novel COVID-19 pandemic, the single-cell RNA sequencing of my cerebral organoids has been delayed. In lieu of data from my organoids, I am analyzing single-cell RNA sequencing datasets available to the public. The first dataset (accession: GSE106245) is from the study by Andrew Field et al. [4] and it contains cells from week 2 and week 5 organoids. The cells in this dataset have been assigned neural epithelial, radial glia, intermediate progenitor, Cajal-Retzius, and neuron. This dataset will represent wild-type histone H3 cells. The second dataset is from Mariella Filbin et al. [4] (accession: GSE102130) and contains cells from six H3 K27M gliomas and two H3 wild type gliomas. The Filbin dataset is not annotated by cell type so I will use ScoreCT to assign these values. I am currently performing differential gene expression analysis on these datasets with Scanpy to see if any differentially expressed genes overlap between datasets. If my hypothesis that the H3 K27M mutation induces a differentiation stall is correct, a possible result could be that differential gene expression analysis reveals that H3 K27M glioma cells closely resemble the neural epithelium cells from Field's organoid data. I am also going to use ScoreCT to assign cell types to Filbin's H3 K27M gliomas and H3 WT gliomas to investigate whether the cell state in H3 K27M gliomas are at a different stage in development relative to H3 WT gliomas. If my hypothesis is correct, I would expect to see more pre-EMT cell types present in the

H3 K27M gliomas compared to the H3 WT gliomas.

References

[1] *Brain Organoids or Cerebral Organoids Derived from hPSCs*. URL: https://www.stemcell.com/technical-resources/area-of-interest/organoid-research/neural-organoids/overview.html (visited on 03/26/2020).

[2] Bo Chen et al. "Gross Total Resection of Glioma with the Intraoperative Fluorescence-guidance of Fluorescein Sodium". In: *International Journal of Medical Sciences* 9.8 (Oct. 6, 2012), pp. 708–714. ISSN: 1449-1907. DOI: 10.7150/ijms.4843. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3477680/ (visited on 03/04/2020).

[3] Jason Fangusaro. "Pediatric High Grade Glioma: a Review and Update on Tumor Clinical Characteristics and Biology". In: *Frontiers in Oncology* 2 (2012). Publisher: Frontiers. ISSN: 2234-943X. DOI: 10.3389/fonc.2012.00105. URL: https://www.frontiersin.org/articles/10.3389/fonc.2012.00105/full (visited on 03/04/2020).

[4] Andrew R. Field et al. "Structurally Conserved Primate LncRNAs Are Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-Specific Genes". In: *Stem Cell Reports* 12.2 (Feb. 12, 2019), pp. 245–257. ISSN: 2213-6711. DOI: 10.1016/j.stemcr.2018.12.006. URL: http://www.sciencedirect.com/science/article/pii/S2213671118305253 (visited on 03/04/2020).

[5] Mariella G. Filbin et al. "Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq". In: *Science* 360.6386 (Apr. 20, 2018). Publisher: American Association for the Advancement of Science Section: Report, pp. 331–335. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aao4750. URL: https://science.sciencemag.org/content/360/6386/331 (visited on 03/04/2020).

[6] Kosuke Funato et al. "Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation". In: *Science* 346.6216 (Dec. 19, 2014). Publisher: American Association for the Advancement of Science Section: Report, pp. 1529–1533.

ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1253799. URL: https://science.sciencemag.org/content/346/6216/1529 (visited on 03/04/2020).

[7] Madeline A. Lancaster et al. "Cerebral organoids model human brain development and microcephaly". In: *Nature* 501.7467 (Sept. 19, 2013). ISSN: 0028-0836. DOI: 10.1038/nature12517. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3817409/ (visited on 03/26/2020).

[8] Peter W. Lewis et al. "Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma". In: *Science* 340.6134 (May 17, 2013). Publisher: American Association for the Advancement of Science Section: Report, pp. 857–861. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1232245. URL: https://science.sciencemag.org/content/340/6134/857 (visited on 03/04/2020).

[9] Brandon R. Lowe et al. "Histone H3 Mutations: An Updated View of Their Role in Chromatin Deregulation and Cancer". In: *Cancers* 11.5 (May 2019). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 660. DOI: 10.3390/cancers11050660. URL: https://www.mdpi.com/2072-6694/11/5/660 (visited on 03/04/2020).

[10] "Method of the Year 2017: Organoids". In: *Nature Methods* 15.1 (Jan. 2018). Number: 1 Publisher: Nature Publishing Group, pp. 1–1. ISSN: 1548-7105. DOI: 10.1038/nmeth.4575. URL: https://www.nature.com/articles/nmeth.4575 (visited on 03/04/2020).

[11] Surya Nagaraja et al. "Histone Variant and Cell Context Determine H3K27M Reprogramming of the Enhancer Landscape and Oncogenic State". In: *Molecular Cell* 76.6 (Dec. 19, 2019), 965–980.e12. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2019.08.030. URL: http://www.sciencedirect.com/science/article/pii/S1097276519306859 (visited on 03/04/2020).

[12] *New hope for treating childhood brain cancer*. ScienceDaily. Library Catalog: www.sciencedaily.com. URL: https://www.sciencedaily.com/releases/2019/04/190404165102.htm (visited on 03/26/2020).

[13] Anca M Paşca et al. "Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture". In: *Nature methods* 12.7 (July 2015), pp. 671–678. ISSN: 1548-7091. DOI: 10.1038/nmeth.3415. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489980/ (visited on 03/26/2020).

[14] Manav Pathania et al. "H3.3K27M Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas". In: *Cancer Cell* 32.5 (Nov. 13, 2017), 684–700.e9. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2017.09.014. URL: http://www.sciencedirect.com/science/article/pii/S153561081730418X (visited on 03/04/2020).

[15] Joëlle Roche. "The Epithelial-to-Mesenchymal Transition in Cancer". In: *Cancers* 10.2 (Feb. 2018). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 52. DOI: 10.3390/cancers10020052. URL: https://www.mdpi.com/2072-6694/10/2/52 (visited on 03/04/2020).

[16] *Stem Cell Basics VII. — stemcells.nih.gov*. URL: https://stemcells.nih.gov/info/basics/7.htm (visited on 03/26/2020).

[17] *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary — SpringerLink*. URL: https://link.springer.com/article/10.1007/s00401-016-1545-1 (visited on 03/04/2020).

[18] James A. Thomson et al. "Embryonic Stem Cell Lines Derived from Human Blastocysts". In: *Science* 282.5391 (Nov. 6, 1998). Publisher: American Association for the Advancement of Science Section: Report, pp. 1145–1147. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.282.5391.1145. URL: https://science.sciencemag.org/content/282/5391/1145 (visited on 04/01/2020).

7 Appendix: Scripts

7.1 Scanpy Preprocessing Script

```
import numpy as np
import pandas as pd
import scanpy as sc
from numba import jit
from numba.types import string
class H3K27M:


    def __init__(self, file):
        self.file = file
        sc.settings.verbosity = 3  # verbosity: errors (0), warnings (1),
        ↪  info (2), hints (3)
        sc.logging.print_versions()
        sc.settings.set_figure_params(dpi=80)


        adata = sc.read_10x_mtx(self.file, var_names='gene_symbols',
        ↪  cache=True)
        adata.var_names_make_unique()
        #self.highestExp(adata)
        #self.mitoInfo(adata)
        self.pcAnalysis(adata)
    def highestExp(self, data):
        """
        Show those genes that yield the highest fraction of counts in each
↪  single cells, across all cells.
```

```
        Filters genes that are present in less than 3 cells.
        :param data:
        :return:
        """
        sc.pl.highest_expr_genes(data, n_top=20, )
        sc.pp.filter_cells(data, min_genes=200)
        sc.pp.filter_genes(data, min_cells=3)
    def mitoInfo(self, data):
        mito_genes = data.var_names.str.startswith('MT-')
        data.obs['percent_mito'] = np.sum(data[:, mito_genes].X, axis=1).A1
        ↪  / np.sum(data.X, axis=1).A1
        data.obs['n_counts'] = data.X.sum(axis=1).A1
        """
        A violin plot of the computed quality measures.
        """
        sc.pl.violin(data, ['n_genes', 'n_counts', 'percent_mito'],
        ↪  jitter=0.4, multi_panel=True)
        """
        Remove cells that have too many mitochondrial genes expressed or
↪   too many total counts.
        """
        sc.pl.scatter(data, x='n_counts', y='percent_mito')
        sc.pl.scatter(data, x='n_counts', y='n_genes')
        """
        Filtering
        """
        data = data[data.obs.n_genes < 2500, :]
```

```
    data = data[data.obs.percent_mito < 0.05, :]
    """
    Total-count normalize (library-size correct) the data matrix    to
↪  10,000 reads per cell, so that counts become comparable among cells.
    """
    sc.pp.normalize_total(data, target_sum=1e4)
    """
    Logarithmize the data.
    """
    sc.pp.log1p(data)


    data.raw = data
    """
    Identify highly-variable genes.
    """
    sc.pp.highly_variable_genes(data, min_mean=0.0125, max_mean=3,
    ↪  min_disp=0.5)
    sc.pl.highly_variable_genes(data)
    """
    Filtering
    """
    data = data[:, data.var.highly_variable]
    """
    Regress out effects of total counts per cell and the percentage of
↪  mitochondrial genes expressed. Scale the data to unit variance.
    """
    sc.pp.regress_out(data, ['n_counts', 'percent_mito'])
```

```
        """
        Scale each gene to unit variance. Clip values exceeding standard
↪   deviation 10.
        """
        sc.pp.scale(data, max_value=10)
    def pcAnalysis(self, data):
        genes = []
        genes = input("Enter genes of interest: ")
        genes = genes.split()
        print(genes)
        results_file =
        ↪  '/Users/LiamT/Desktop/H3_K27M/hu_ls_h3k27_wk02_aggr_results.h5ad'
        """
        Reduce the dimensionality of the data by running principal
↪   component analysis (PCA), which reveals the main axes of variation and
↪   denoises the data.
        """
        sc.tl.pca(data, svd_solver='arpack')
        #sc.pl.pca(data, color='CST3')
        """
        Let us inspect the contribution of single PCs to the total
↪   variance in the data.
        This gives us information about how many PCs we should consider in
↪   order to compute the neighborhood relations of cells,
        e.g. used in the clustering function sc.tl.louvain() or tSNE
↪   sc.tl.tsne().
```

```
        In our experience, often, a rough estimate of the number of PCs
↪   does fine.
        """
        sc.pl.pca_variance_ratio(data, log=True)
        data.write(results_file)
        """
        Computing the neighborhood graph
        """
        sc.pp.neighbors(data, n_neighbors=10, n_pcs=40)
        """
        In some ocassions, you might still observe disconnected clusters
↪   and similar connectivity violations. They can usually be remedied by
↪   running:
        """
        #tl.paga(adata)
        #pl.paga(adata, plot=False)  # remove `plot=False` if you want to
        ↪  see the coarse-grained graph
        #tl.umap(adata, init_pos='paga')
        sc.tl.umap(data)
        sc.pl.umap(data, color=genes)
        sc.pl.umap(data, color=genes, use_raw=False)
        """
        Clustering the neighborhood graph
        """
        sc.tl.leiden(data)
        genes.append('leiden')
        sc.pl.umap(data, color=genes)
```

```
        data.write(results_file)
        """
        Finding marker genes
        """
        sc.tl.rank_genes_groups(data, 'leiden', method='t-test')
        sc.pl.rank_genes_groups(data, n_genes=25, sharey=False)
        sc.settings.verbosity = 2  # reduce the verbosity
def main():
    inString = input("input file path: ")
    #results = input("input results file name: ")
    h5Object = H3K27M(inString)
if __name__ == "__main__":
    main()
```

7.2 ScoreCT

```
#!/usr/bin/env python
# coding: utf-8

# In[8]:



# Import modules
import os
import sys
import pandas as pd
import scanpy.api as sc
```

```
sys.path.append('../src/')
import scorect_api as ct

# Load example scRNA-seq data (already analyzed)
adata = sc.read('hu_ls_wtype_wk02_aggr_results.h5ad')

# Load reference table - scoreCT api provides a wrapper for gmt, tsv and
↪  csv formats.
ref_marker = ct.read_markers_from_file('Ref_Marker_LHT.csv')

sc.tl.rank_genes_groups(adata, groupby='leiden', n_genes=len(adata.var),
↪  use_raw=False)

marker_df = ct.wrangle_ranks_from_anndata(adata)
print(marker_df.head())

# Score cell types for each cluster
# Let's set parameters first - K represents the number of genes included
↪  in the ranking
# m represents the number of bins used to divide the top K genes.
K = 300
m = 5
# Get the background genes - here, all the genes used to run the
↪  differential gene expression test
background = adata.var.index.tolist()
# Now run the function
ct_pval, ct_score = ct.celltype_scores(nb_bins=m,
```

```
                                        ranked_genes=marker_df,
                                        K_top = K,
                                        marker_ref=ref_marker,
                                        background_genes=background)

# Let's have a look at the pvalue for cluster 0
ct.plot_pvalue(ct_pval, clusters=0)

# Now assign clusters to cell types
cluster_assign = adata.obs['leiden']
celltype_assign = ct.assign_celltypes(cluster_assignment=cluster_assign,
↪  ct_pval_df=ct_pval, ct_score_df=ct_score, cutoff=0.1)
# Add to anndata object
adata.obs['scorect'] = celltype_assign
# Let's compare with the true assignment now!
sc.pl.umap(adata, color='scorect', save='hu_ls_wtype_wk02_aggr_results.pdf')


# In[ ]:
```