UNDERGRADUATE RESEARCH THESIS

Single Cell Analysis of Patients with Perianal Chron's Disease

Ferguson, Katherine E

First Reader:

DocuSigned by:
Dr. Greg Gibson
D55B0EF58EF44F8...
Dr. Greg Gibson

Second Reader:

DocuSigned by:
Dr. Peng Qiu
A4A8655E0903420...
Dr. Peng Qui

Table of Contents

Abstract

Perianal Chron's Disease is a debilitating form of Chron's Disease that often involves rectal tissue in its manifestation in the body. There is a greater need for understanding of the role of the epithelial tissue in the rectum in the pathology of Chron's Disease. This study aims to determine cellular make-up of rectal-derived organoids grown from Chron's Disease patients. Through the use of single-cell transcriptomics, this study will determine the cell types present in four rectal-derived organoids and use differential expression to analyze differences in epithelial tissue of Chron's Disease patients in comparison to a healthy individual.

Background

Perianal Fistulizing Crohn's disease (perianal CD) is a crippling manifestation of Crohn's disease (CD) and is associated with a poor quality of life. A fistula, which is an abnormal connection between two hollow tissue formations, is a marker of perianal CD when it forms, and can lead to long term damage of the sphincter and perianal tissue.[1] Fistulas frequently invade the anal sphincter and can involve other structures of the pelvis. These fistulas can lead to serious complications such as infection of the bone, abscesses, and fecal incontinence.[2] While the incidence of perianal disease in children is unknown, it is estimated to be around 8%-13% of patients[1,6] and usually requires a multidisciplinary approach for treatment and therapy. Previous studies have implicated growth factors, including TNF, involvement of inflammation, and confirmed the role of the immune system in perianal CD.[2] As the pathogenesis of Crohn's disease itself remains poorly understood, most studies are more generalized in terms of study of the overall disease. Gettler et al (2019) is one of the most recent studies establishing how multiple loci combine to cause Crohn's disease.[8] They were able to utilize multiple datasets and new computational analysis methods to narrow the search for causal links in the genome, focusing on monocytes from patients with Crohn's disease. Zheng et al complemented this research by performing single-cell analysis on T-cells in patients with pediatric Crohn's disease.[7] Their use of single-cell techniques allowed them to not only look at gene expression in each cell type, but also to gain a better understanding of the variety of expression interactions among T-cell subsets in terms of what percent of each cell population is experiencing certain gene expression profiles.

The rectum has also been shown to be involved in the pathogenesis of perianal CD in 92% of patients.[3] This suggests that the rectal epithelium, along with disruptive contributions of

the immune system, contributes to the pathogenesis of perianal CD. There is a need for a greater understanding of the pathology of perianal CD and what role the epithelium plays in the expression of this type of CD. This research will contribute to this investigation by exploring cell-type specific gene expression in human-derived organoids from patients with perianal CD through the use of organoid single cell profiling. Current single cell profiling experiments have focused on the immune cells - mostly T-cell involvement. We will be able to extend on that current knowledge and provide a deeper understanding by looking at multiple cell lines of tissue involved in perianal CD.[5]

The objective of our study was to observe gene expression in single cells of rectal-derived organoids in an attempt to determine cell types found in the rectal organoid and then determine the amount of variability between the individual organoids. We utilized four organoids, both fresh and frozen, consisting of one control patient and three patients with different stages of Chron's disease. These patient derived organoids should closely resemble the donor profiles at the cell-type level of gene expression. We hypothesized that the cell make-up of these organoids should include LGR5+ stem cells, cyclic transient amplifying epithelial cells (also known as transitional cells), mucin producing goblet cells and mature epithelial cells. Using single-cell sequencing and the single cell transcriptome analysis software Seurat, I identify cell-type clusters in the four organoids. Initial steps toward incorporating single nucleotide variant (SNV) detection will enable assessment of which individuals are represented in each of those clusters.

Methods

Our study continues investigation into the pathogenesis and appearance of perianal disease by analyzing cell-type specific gene expression in human-derived organoids from patients with perianal CD through the use of organoid single cell profiling. We will examine four organoids, three from Crohn's Disease patients and one control patient. We will aim to profile 2,000 cells per organoid and have a read depth of 50,000 reads from those cells. From those reads, we will first impute cell types of epithelial tissue based on a previous study that was able to identify five cell types from a similar organoid culture system.[4] We will accomplish this by using single-cell RNA sequencing methods and the bioinformatics tool *Seurat*, which identifies significant genes that define clusters of cells. Those marker genes are then used to determine the cell types corresponding to each cluster of cells based on previous literature of marker genes and what is known about the expression of those genes in epithetical tissue types.

Since the sequencing was based on pooling of organoids from four individuals, after determining the cell types of each of the clusters, we will next develop a pipeline for assessing which cells came from which individual. Since some samples were from inflamed tissue, this would allow us to assess whether different clusters are likely to reflect the impact of inflammation.

To determine the individual that each cell belongs to, we would preferably need the genotypes of each patient, which we currently do not have access to. Instead, while we wait for the genotype data, I have attempted to identify individuals through a relatively new technique called RNAseq short variant calling, which extracts SNPs and indels in the cells that can be used to determine and individuals signature and help determine which cells belong to the same person. This will at least enable us to identify that some of the populations of types of cells have an uneven representation

of the pooled population of cells from individuals. By gaining a better understanding of which cells are involved in the formation and expression of proteins that cause fistulas, we may be able to contribute to development of novel treatments, and potentially help prevent the formation of perianal fistulas.

Results

A prior study on single cell sequencing of human intestinal organoids, Fuji et. al.,[4] suggested that we would be able to detect LGR5+ stem cells, transient amplifying cells, goblet cells, early and mature enterocytes in our data set. We used the marker genes identified in that study as a basis for detecting similar cells in our organoids. This aim was accomplished through the use of transcriptome profiling. We utilized both t-distributed stochastic neighbor embedding (t-SNE) (Figure 1B) and uniform manifold approximation and projection (umap) (Figure 1A) dimensional reduction to visualize nine discrete cell subpopulations. We decided to utilize the umap dimensional reduction method for the rest of the visualization of our differential expression analysis because the distance between clusters was more clearly defined. We then performed differential expression to determine marker genes of each of the nine discrete sub populations (Figure 1C). We found that cluster 7 contained gene expression patterns consistent with intestinal stem cells, including LGR5 and EPHB3 (Figure 2A). Clusters 1 and 2 showed up regulation in PCNA, TOP2A, MKI67, PTTG1, and CKS2 (Figure 2A). These genes all serve as markers for transient amplifying cells. This was suggestive that the umap_1 axis of our single cell analysis was reflective of the cell timeline, from stem cells on the right side of the axis and adult enterocytes to the left side of the axis, and warranted further investigate into differential expression between clusters separated along the umap_2 axis. We found SOX4+ tuft cells in

cluster 9 (Figure 2C) and used MUC2 as a marker gene for goblet cells. Goblet cells were found in clusters 4, 5, 2 and 2 as well as in cluster 9 (Figure 2B). KRT19 and FABP1 served as early

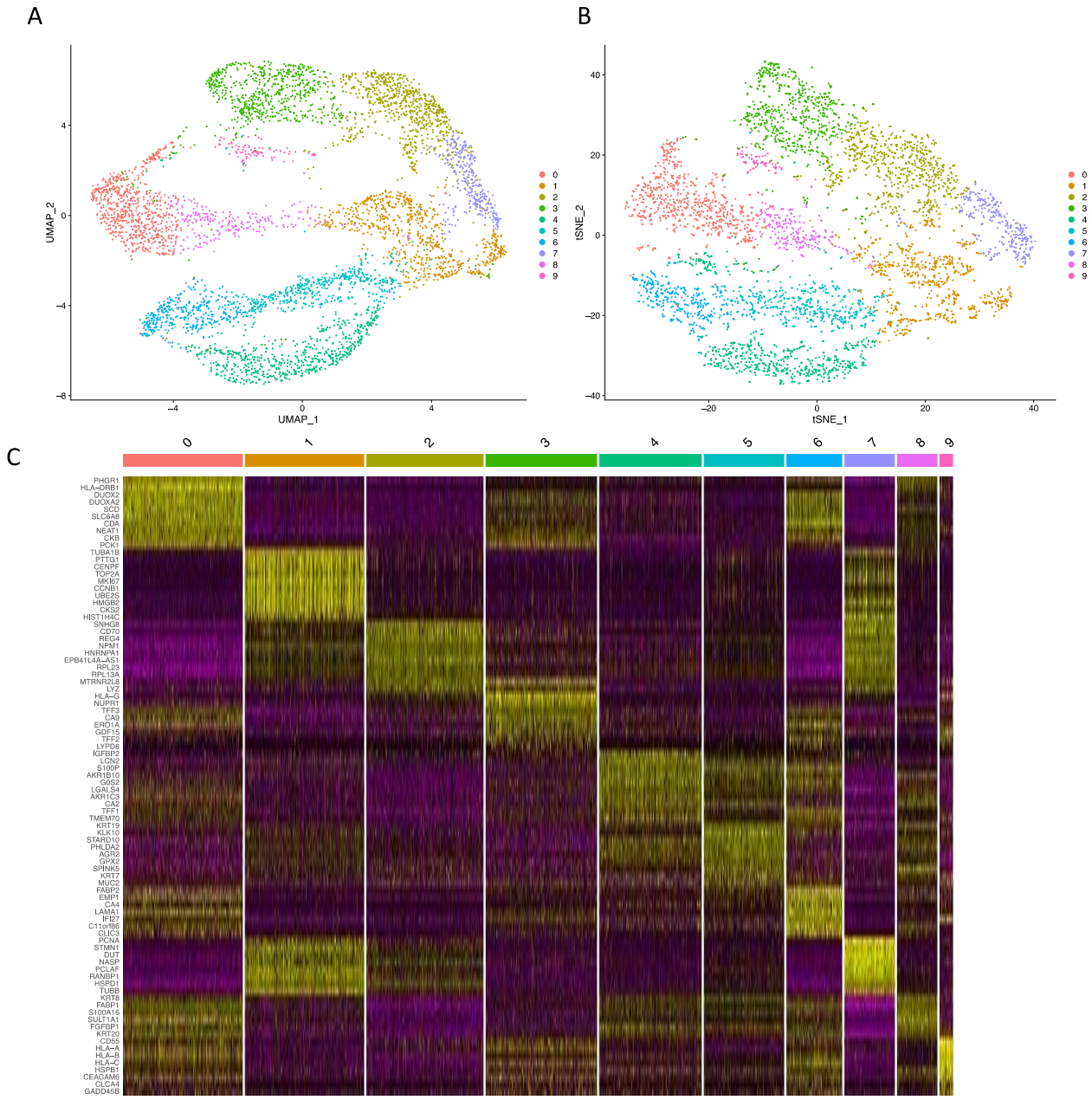

Figure 1. Differential Expression among Nine Discrete Cell Sub-populations

(A) scRNA-seq analysis of four rectal-derived organoids (2 fresh and 2 frozen). Cell clusters are visualized by umap. Genes were selected as variable features using the *FindVariableFeatures* function in the *Seurat* package and used for dimensional reduction. To find clusters we

performed shared nearest neighbor clustering modules with the *FindNeighbors* function in *Seurat* and set the resolution for finding clusters to 0.8

(B) scRNA-seq analysis of four rectal-derived organoids (2 fresh and 2 frozen). Cell clusters are visualized by t-SNE and clustering was performed using the same parameters for umap visualization in (A).

(C) Heat map of the top ten markers of all nine sub-populations, found using the *FindAllMarkers* function in the *Seurat* package

enterocytes (Figure 2C). We found this to be consistent with our hypothesis that the umap_1 axis reflected the differences along the cell timeline. We then began to investigate the differences along the umap_2 axis. We found HLA-G to be a in clusters two and eight, on the positive side of the umap_2 axis. We also found significant upregulation of REG4 in clusters 8 and 2. To further investigate the hypothesis that we had observerable differentiation between enterocytes and goblet cells between individuals we attempted to run a trajectory analysis and SNV variant detection. We ran initial *Slingshot* trajectory using our umap object (Figure 2D) and found two distinct lineages stemming from our LGR5+ cluster. One lineage starts at the LGR5+ cluster and goes up the umap_2 axis and to the left of the umap_1 axis through clusters 2, 8 and then 9. The second lineage stems from the LGR5+ stem cells and then proceeds to our identified transient amplifying cell cluster (cluster 1). At cluster 1, the lineage then splits in two, with one branch going through clusters 8 and then zero and the second branch going through 5, 4 and then 6. We also attempted to run *GATK*'s single nucleotide variant (snv) detection pipeline and found initial success that indicated this could be useful in our analysis but were unable to make the pipeline successful in our research timeframe.

Discussion

Determining the cell types of the cells present in our rectal-derived organoids proved to be difficult due to the similarity in gene expression across the cell types. Differential expression among individuals in our sample set may also have contributed to cluster detection. We were able to accurately identify many of the key cell types you would expect in a

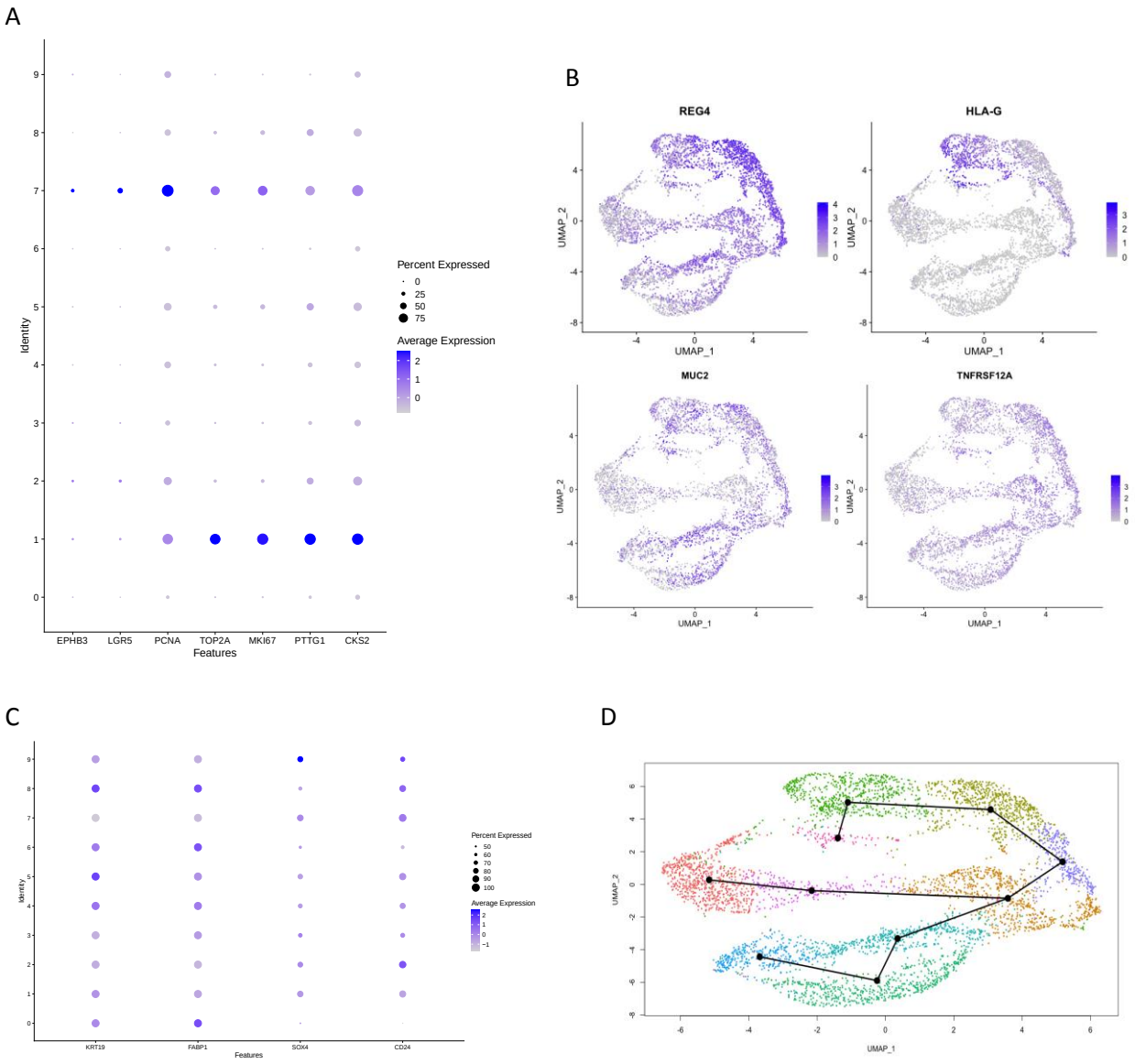

Figure 2. Differential Expression Analysis of Sub-populations of Cells

(A) Dot plot of stem cell markers EPHB3, LGR5 and PCNA and transient amplifying cell markers TOP2A, MKI67, PTTG1 and CKS2

(B) Feature plot of potential Chron's Disease markers HLA-G and REG4 and of goblet cell markers MUC2 and TNFRSF12A

(C) Dot plot of early enterocyte markers KRT19, FABP1 and CD24 in addition to tuft cell marker SOX4

(D) Slingshot trajectory analysis lineage plot, with black lines determining path of cell timeline starting at cluster 7.

conventionally grown rectal-derived human organoid, including LGR5+ stem cells, MUC2+ goblet cells, transient amplifying cells and early enterocytes. We also found HLA-G and REG4 to be significant markers that separated some of the goblet and enterocyte cells from other clusters of goblet and enterocyte cells. HLA-G has been cited in previous research as a potential marker of both ulcerative colitis and Chron's Disease when found in intestinal samples.[8] Increased expression of REG4 has also been linked to inflammation.[9] This suggests that there may be a distinct profile of gene expression in rectal cells of patients with Crohn's Disease. Further investigation is required in order to confirm this observation. We are looking to complete application of the GATK pipeline to our data set and potentially to complete the genotyping of our individuals which will enable unambiguous determination the identity of the individuals to whom each cell belongs.

Work Cited

1. Kamm, M. A., Ng, S. C. (2008). Perianal fistulizing Crohn's disease: a call to action. *Clinical Gastroenterology and Hepatology*, *6*(1), 7-10.
2. Panés J, Rimola J. Perianal fistulizing Crohn's disease: pathogenesis, diagnosis and therapy. Nat. Rev. Gastroenterol. Hepatol. 2017; 14: 652–64
3. Hellers, G., Bergstrand, O., Ewerth, S., Holmstrom, B. Occurrence and outcome after primary treatment of anal fistulae in Crohn's disease. *Gut* **21**, 525-7 (1980).
4. Fujii, Masayuki, Mami Matano, Kohta Toshimitsu, Ai Takano, Yohei Mikami, Shingo Nishikori, Shinya Sugimoto, and Toshiro Sato. "Human Intestinal Organoids Maintain Self-Renewal Capacity and Cellular Diversity in Niche-Inspired Culture Condition." *Cell Stem Cell* 23.6 (2018): 787-93.e6. Web.

5. Zheng H., Ordovas-Montanes J., et. al, P154 Single-Cell Analysis Of T Cell Pathogenesis In Pediatric Crohn's Disease, *Inflammatory Bowel Diseases*, Volume 25, Issue Supplement_1, February 2019, Page S69, https://doi.org/10.1093/ibd/izy393.171
6. Safar B, Sands D. Perianal Crohn's disease. Clin Colon Rectal Surg. 2007;20(4):282–293. doi:10.1055/s-2007-991027
7. Gettler K., Giri M., Kenigsberg E., Martin J. el al. "Prioritizing Crohn's disease genes by integrating association signals with gene expression implicates monocyte subsets." Genes and Immunity (2019).
8. Gomes RG, Brito CAA, Martinelli VF, et al. HLA-G is expressed in intestinal samples of ulcerative colitis and Crohn's disease patients and HLA-G5 expression is differentially correlated with TNF and IL-10 cytokine expression. *Hum Immunol*. 2018;79(6):477–484. doi:10.1016/j.humimm.2018.03.006
9. van Beelen Granlund, A., Østvik, A. E., Brenna, Ø., Torp, S. H., Gustafsson, B. I., & Sandvik, A. K. (2013). REG gene expression in inflamed and healthy colon mucosa explored by in situ hybridisation. *Cell and tissue research*, *352*(3), 639–646. https://doi.org/10.1007/s00441-013-1592-z