The Rachel and Selim Benin
School of Computer Science and Engineering
Department of Developmental Biology & Cancer research
The Hebrew University – Hadassah Medical School

Computational method in Molecular Biology Lab - 76554

Charting the direct channels of communication between enteroendocrine cells in the human gut

Liron Gershuny
308350503

Supervised by Dr. Danny Ben-Zvi

October 2021

1. Abstract

Endocrine cells are scattered throughout the gastrointestinal mucosa from the stomach to the colon and constitute one of the largest endocrine systems in the body. Enteroendocrine cells labeled by chromogranin-A or chromogranin-B comprise about 1% of all epithelial cells in the gastrointestinal tract. The enteroendocrine system consists of at least 15 different cell types that can be classified based on their main hormonal products. Although enteroendocrine cells are very scarce, they are essential regulators of digestion, gut motility, appetite, and metabolism.

In this project, I studied how endocrine cells communicate with each other with an emphasis on the human intestinal tract. In order to do this, I collected data extracted from intestinal organoids and used new computational biological tools in order to reach interesting biological conclusions on the subject. Subsequently, I applied the methodology I developed to data collected in the laboratory about endocrine cells in the human stomach and I found reinforcement for results obtained from the gut in addition to other aspects.

2. Introduction

2.1. Literature Overview

The human digestive system is a set of structures and organs that help the digestion process in the body. These organs and structures play a vital role in the digestion of food and liquids. The intestine is the main organ of the digestive tract. It carries out various tasks such as mixing and transporting intraluminal contents, producing enzymes and absorbing nutrients. The intestine is the longest part of a person's digestive tract. It begins at its pylorus and ends at its ileocecal valve. The basic roles of the small digestive system are to transport, digest, and absorb food.

Intestinal epithelial cells are consistently regenerate each 4–5 days through a course of cell division, development, and movement. Regenerating depends on proliferative cells (undifferentiated organisms) that dwell at the base of the gastrointestinal organs. After being framed at the base, the new cells move upwards and out of the grave, developing en route. In the end, they go through apoptosis and shed off into the digestive lumen. Thusly, the coating of the digestive system continually regenerated while the quantity of cells making up the epithelial layer stays consistent.

The intestinal epithelium consists of six distinct cell types. Each type develops as per its particular separation program as it moves up and out of the base. A large number of the qualities fundamental for separation into the distinctive epithelial cell types have been recognized and described. The cell types delivered are enterocytes, Goblet cells, enteroendocrine cells, Paneth cells, transit-amplifying cells, and stem cells.

With more than thirty unique chemicals distinguished as being delivered in the gastrointestinal (GI) plot, the gut has been depicted as the biggest endocrine organ in the body. The order of these chemicals and the cells that produce them, the enteroendocrine cells (EECs), has given the establishment to stomach-related physiology. EECs assume a vital part in the control of GI discharge and motility, the guideline of food consumption, postprandial glucose levels, and digestion. EECs sense luminal content and deliver flagging particles that can enter the flow to go about as exemplary chemicals on far-off targets, act locally on adjoining cells, and particular neuronal pathways including intestinal and extraneous neurons.

Enteroendocrine cells (EECs) have been displayed to create from the equivalent pluripotent immature microorganisms as the other three cell heredities of the gastrointestinal epithelium: absorptive enterocytes, goblet cells, and Paneth cells. The course of separation of EECs starts as multiplying pluripotent foundational microorganisms at the foundation of the digestive base and advances as daughter cells relocate, in a vertical straight style, towards the epithelial sleeve at the luminal surface. In the mean of single-cell analysis, EECs can be filtered using high levels expression of specific genes like NGN3.

Gastrointestinal EECs are limited to the mucosa, predominately situated inside its more profound half, and involve just a little minority (<1%) of the general epithelial cell populace, regularly lying detached from each other blended by non-endocrine epithelial cells suggesting their operation via endocrine signaling over paracrine

signaling. At this point information investigating the role of EECs in obsessive conditions influencing the GI plot are preliminary, yet recommend that EEC might play a part in the pathology of conditions including peevish entrail disorder and colorectal adenocarcinoma.

Distinct types of EECs communicating more than 20 peptides/chemicals have been recognized along the GI mucosa and it is currently evident that numerous EECs contain more than one flagging atom. EEC secretory substances are delivered because of assorted kinds of boosts and impact an assortment of physiological capacities. For example, gastrin, emitted by open-type G cells of the gastric antrum and pylorus in light of luminal amino acids and calcium, controls gastric corrosive emission by following up on shut sort enterochromaffin-like cells of the gastric corpus, which discharge histamine. Thusly, histamine enacts parietal cells to secret gastric corrosive. Somatostatin, which is delivered by shut kind D cells of the gastric corpus in light of digestive chemicals like CCK or transmitters, hinders gastric corrosive discharge by direct restraint on parietal cells and through the hindrance of histamine discharge. Paradoxically, somatostatin delivered from open kind D cells in the antrum represses gastric corrosive discharge through restraint of the creation and arrival of gastrin.

An organoid is a 3D multicellular in vitro tissue construct that mimics its corresponding in vivo organ, such that it can be used to study aspects of that organ in the tissue culture dish. Such constructs are derived from stem cells that could be either pluripotent (embryonic or induced) or adult stem cells from various

organs. In general, the processes that form these tissues in vitro approximate natural development or tissue maintenance. Challenges arise from the fact that these cultures are both more complex than the in vitro models researchers are using, and yet they are not quite complex enough. Organoids typically lack vasculature and immune cells which means that they are limited in how big they can grow without cell death. One of the organoids that researchers already modeled is the intestine which I used in this project.

Bone Morphogenetic Proteins (BMPs) are a group of signaling molecules that belongs to the Transforming Growth Factor-β (TGF-β) superfamily of proteins. Initially discovered for their ability to induce bone formation, BMPs are now known to play crucial roles in all organ systems. BMPs are important in embryogenesis and development, and also in the maintenance of adult tissue homeostasis. Moreover, BMP signaling is involved in the formation of intestinal villus patterning. Villi contribute to increasing the effective absorption of nutrients by extending the surface area in the intestine. Gain or lose function of BMP signaling altered the patterning of clusters and emergence of villi in mouse intestinal model.

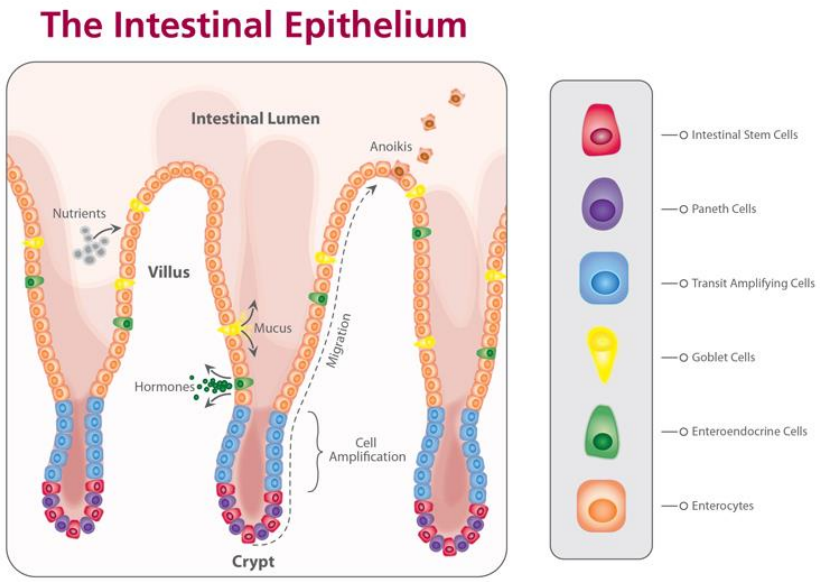

Figure 1: Diagram of the intestinal epithelium highlighting the identity and spatial arrangement of key epithelial cell types. The adult intestinal epithelium is primarily composed of six cell types arranged in a crypt-villus structure. Emphasizing EECs do not touch one another working via endocrine signaling.

2.2. Research Question

The main research question is if endocrine cells work together as a distributed machine, or does each do its own thing, completely independent of the other. My hypothesis is that EECs are working together as a group meaning that each EEC type communicates with any other EEC type in order to accomplish the same purpose.

The project aims to find the means of communication between the enteroendocrine cells in the human gut, and in particular, I will quantify all the means of communication, and later I will characterize significant means of communication. Finally, I will apply the results to data extracted from the human stomach.

The idea behind those goals is to find out more options to externally affect the hormonal balance inside the digestive system as well as to open up additional horizons for the production of drugs that will affect a particular target cell.

2.3. Working Premises

Human EECs are rare and have been largely inaccessible for in vitro studies. Therefore, to generate the full spectrum of human EECs, the data I worked with was established from organoids of healthy adult patients from different regions of the intestinal tract. The human EEC organoid cultures mimic almost perfectly in structure, function and generate all cell types of the epithelial lining. Furthermore, these organoids uniquely allow a proteomic analysis of hormones secretion and are found to be essentially identical.

3. Methods

To produce the desired results I used computational biological tools and packages available in R and Python.

3.1. Bulk Single-Cell mRNA Sequencing

RNA sequencing (RNA-seq) is a genomic approach for the detection and quantitative analysis of messenger RNA molecules in a biological sample and is useful for studying cellular responses. scRNA-seq permits the comparison of the transcriptomes of individual cells.

While scRNA-seq workflows are conceptually carefully related to population-stage transcriptomics protocols, statistics from scRNA-seq experiments have several functions that require specific bioinformatics tactics. First, the records are incredibly sparse thanks to an excessive frequency of dropout events. Moreover, due to the digital nature of gene expression on the single-cellular stage and the related phenomenon of transcriptional bursting, transcript tiers are subject to temporal fluctuation, similarly contributing to the high frequency of zero observations in scRNA-seq facts. Therefore, the numbers of expressed genes detected from unmarried cells usually decrease as compared with populace-degree ensemble measurements. Because of this imperfect insurance, I determined to work with normalized transcript levels used for bulk RNA-seq.

3.2. Seurat

After reading the data, I created a counting matrix (QC) whose values represent quantification of expression for each gene expressed in a particular cell. I have now used a package called Seurat designed for single-cell data analysis and created an object of it that is a container containing the data itself as well as analyzes that I perform on the data. Now, in the preprocessing phase, I normalized the Log-Normalize data that normalizes the amount of gene expression for each cell relative to the total expression, multiplied the result by a fixed factor of 10,000, and performs a log-to-result transformation. Already at this point, I could see which genes are overexpressed in my data and thus focus my look at them as well as perform a kind of self-examination that these are indeed genes that I expected to be overexpressed in this type of data coming from the gut.

To perform PCA I had to perform scaling of the data for each expression of each gene so that the mean across all cells would be 0 and the variance across all cells would be 1. At this point, I performed a PCA procedure in which the various features found earlier are used as input. Finally, I used the K-nearest neighbor algorithm in which edges were drawn between cells with similar feature expression patterns, and then attempt to partition this graph into highly interconnected ‘communities’.

3.3. CellPhoneDB

To decode intercellular communication networks that will ultimately explain tissue function in homeostasis and their alterations in disease it is necessary to measure the expression of ligands and receptors in multiple cell types.

Identifying ligand-receptor interactions from scRNA-seq requires both the annotation of complex ligand-receptor relationships from the literature and a statistical method that integrates the resource with scRNA-seq data and selects relevant interactions from the dataset.

To enable a comprehensive, systematic analysis of cell-cell communication molecules I used CellPhoneDB, a public repository of ligands, receptors, and their interactions. This repository relies on the use of public resources to annotate receptors and ligands, as well as manual curation of specific families of proteins involved in cell-cell communication. Furthermore, CellPhoneDB includes subunit architecture for both ligands and receptors to represent heteromeric complexes accurately, which is critical because cell-cell communication relies on multi-subunit protein complexes that go beyond the binary representation used in most databases and studies.

As input, I took the clusterization I performed in the clusterization step and I added the single-cell sequencing data. The tool used the platform of Google Collab and required many hours of running to produce the desired results.

As an output, I received lists of ligand-receptor pairs that known to have a biological relationship between them to which I could match the cells in my clusterization, quantify the total number of connections between the cells in a specific group versus the total number of relationships

with other groups. Moreover, CellPhoneDB indicates pairs of bonds that the genes that characterize them are overexpressed and therefore these pairs are marked as significant.

3.4. Cytoscape

To characterize the types of communication and connections found with the help of CellPhoneDB, I chose to build a network in which each node constitutes a ligand or receptor, and I will connect in a line those that are found to be related. I can symbolize these lines in different colors, thicknesses, and transparency depending on how significant the connection is according to the results obtained from CellPhoneDB. To do this I used the Cytoscape tool, which allows me to use tables of connections as input and produces a visualization of a network as required.

4. Results

4.1. Cell Filtering

Out of a sampling of 8,448 cells collected from the intestinal epithelium, preliminary filtration biologically by FACS performed in order to be left with only enteroendocrine cells. To do more filtrations I turned to literary sources that point to markers unique to these kinds of cells. The main markers by which I filtered were FABP1-positive enterocytes, OLFM4, positive stem cells, MUC2-positive goblet cells, LYZ-positive Paneth cells. The filtering was performed one after the other according to markers expression level, respectively to the unstudied expression levels in the data. After filter, I left with 2,993 cells when I now know that the vast majority of them are enteroendocrine cells.

4.2. Clusterization

Clustering is a type of unsupervised learning comprising many different methods. In this study, I focused on k-means clustering algorithm that uses correlation distance. My main goals were to divide the filtered cells in the data into similar groups in terms of their gene expression but not for further filtering clearly non-endocrine. Later, I wished to be able to characterize each group according to the gene expression in it so that I can give it concrete names. In general, I used default values as recommended by Seurat (see Figure 2), changing some parameters in running the program produces slightly different results.

I wished to learn the underlying manifold of the data to place similar cells together in low-dimensional space. To do so I used non-linear

dimensional reduction such as PCA. In order to visualize the results, I used UMAP in which cells within the graph-based clusters determined above should co-localize on these dimension reduction plots. As input to the UMAP, I used the same PCs from the PCA output process.

To complete the clusterization process I found the biomarkers that characterize each cluster according to the different expressions in it. I could do this by visualizing the expression of certain genes from the UMAP I produced. After reading similar studies, I found several including the one from which I took the data for the project, that point to EECs groups that characterize the human gut. For example, a group of cells called M-X cells characterized by overexpression of the MLN, GHRL, ENPP1, and ACSL1 genes. When I direct my gaze at these genes and the clusters characterized by their overexpression, I can say with confidence that this cluster is probably of the M-X cell type.

After performing this procedure, I was able to perform a full annotation of 13 different groups and subgroups divided according to the expression of specific markers in the UMAP I produced. It must be said that even after all the filters I performed at the beginning of the work I could still find cell groups that do not belong to the EECs family, such as more stem cells, and these will constitute for me later a good sanity check for the results. Moreover, a merge of different clusters was done manually by me which led to a decrease from 18 different clusters as the output of Seurat PCA algorithm to 13 different annotations by my analysis.

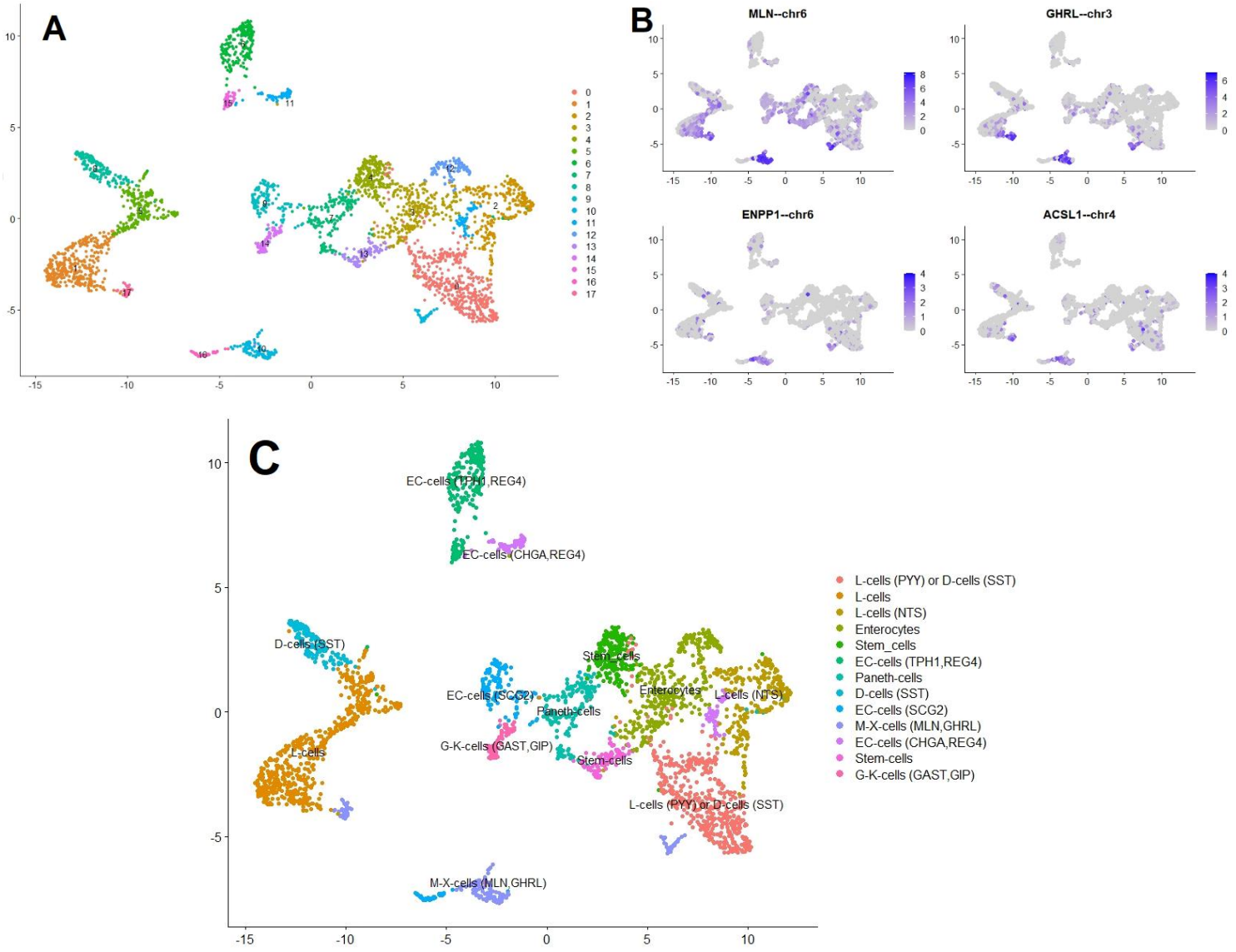

Figure 2: Single-Cell Transcriptome Atlas of Human Gut Enteroendocrine Cells

(A)UMAP displaying the human EEC atlas (n=2,993 cells) before annotations. Different colors represent the 18 separate clusters (k=10).

(B)UMAP displaying the expression levels of M-X cell type marker genes in the different human EEC subtypes from intestinal organoids. Bars display color-coded unique transcript expression (logarithmic scale).

(C)UMAP displaying the human EEC atlas (n=2,993 cells) after annotations. Different colors represent the 13 separate clusters (k=10).

4.3. Communication Quantification

After initial basic runs of CellPhoneDB, I found that the number of pairs is extremely high so I decided to emphasize pairs that the literature points to as more relevant to endocrine cells located in the human gut assuming that they don't touch each other in real life. To do this CellPhoneDB allows me to upload an independent database according to which the calculations will be performed. At this point, I downloaded all the existing data in the tool, made a comparison with the relationships that found to be significant in the basic runs, and filtered out the ones I would not want to focus on and thus significantly increased the expression of the couples more relevant to me. This change in the running of the program resulted in a decrease in the number of intercellular connections of more than 60% thus providing a more accurate picture of the nature of the connections between the endocrine cells in the gut.

The results yield some key conclusions that I could have considered: One, the amount of intercellular connections for cells from the same cluster is lower than the number of intercellular connections between different clusters suggests that autocrine signaling or endocrine between self is not relevant- each cell or cell type works relatively alone. In addition, clusters that are subgroups of a single cell type, such as L-cells, communicate differently depending on the expression of the overexpressed genes suggests that it's actually two different cell types- these cell types secrete the same hormone but have other functions or they are not endocrine at all. Second, it can be said that a large number of intercellular connections between different clusters indicate genes responsible for intestinal motility

suggests that these cell types (see G-K and L-NTS clusters in Figure 3) work together.

Referring to non-endocrine clusters such as Stem cells, Enterocytes- Probably those types weren't filtered out because of a lack of filtering or not the proper filtering parameters. Thanks to it, I can show that no communication between EECs and these cell types are found while there is communication with Paneth cell type, which is non-endocrine too. This field can be for future studies on the interaction between endocrine cells and other cell types.

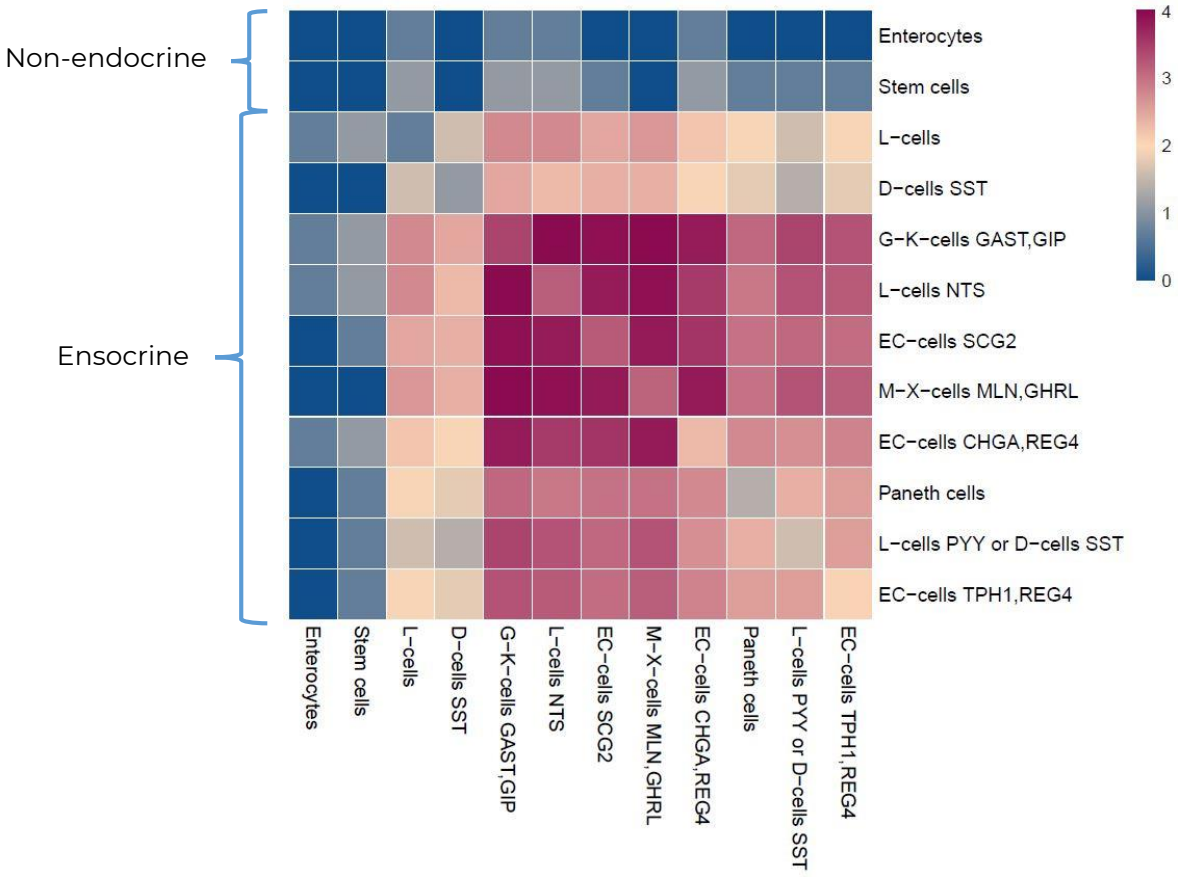

Figure 3: Heatmap showing the total number of interactions between EEC types and non-EEC types such as Stem cells, Enterocytes and Paneth cells in the intestinal dataset obtained with CellPhoneDB. Bar display color-coded amount of interactions (logarithmic scale).

4.4. Communication Characterization

Building this Cytoscape network has helped me reach biological conclusions that are difficult to discern without going beyond easy visualization like this. At this point, I have reached important conclusions. First, endocrine cells communicate with each other via TGFbeta and not via the hormones they secrete to target cells. I concluded this after seeing significant associations between a BMP receptor and some of its legends like BMP2, BMP4, and BMP7. Second, endocrine cells communicate within a single villus or crypt, and not across long distances, the literature indicates that BMP acts on a short range of less than a millimeter also communicates in the form of diffusion and not through the bloodstream.

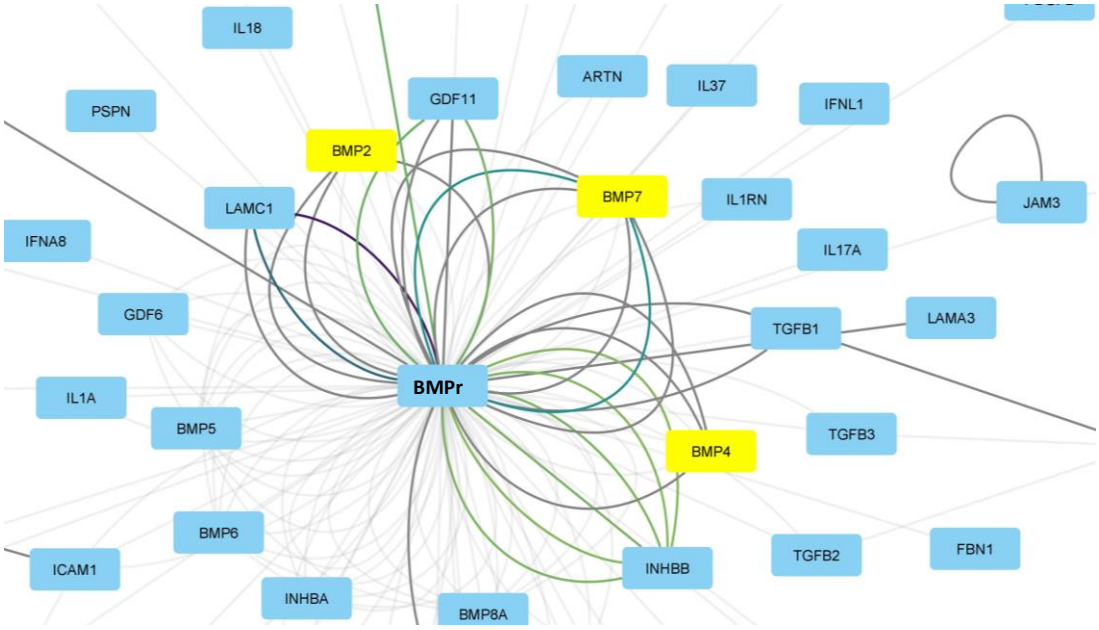

Figure 4: A subnetwork of ligand-receptor connections in the intestinal EEC dataset obtained with CellPhoneDB significant connections output and Cytoscape. This part of the network shows the significant connections with the TGFbeta ligands family.

4.5. Stomach

In order to get a broader picture of the nature of communication between endocrine cells, to check whether the procedure I have determined produces logical results for other databases as well as to connect with existing laboratory data about endocrine cells from the human stomach I chose to perform the same procedure for endocrine cells from the fundus and body regions of the human stomach.

To do this, I filtered the total cells to only 1,068 endocrine cells on which I performed a clustering procedure. That resulted in division into only three groups - two of which known to me from the data collected for the gut and are M-X cells and EC-cells to which a new group of cells called ECL-cells that characterized by the HDC and MAOB biomarkers.

The results obtained from CellPhoneDB show a lesser amount of connections of about 25% of the number of connections found between endocrine cells in the human gut. Similar results were obtained in the context of communication with the help of BMP, to which significant connections were also added as part of the Notch signaling pathway.

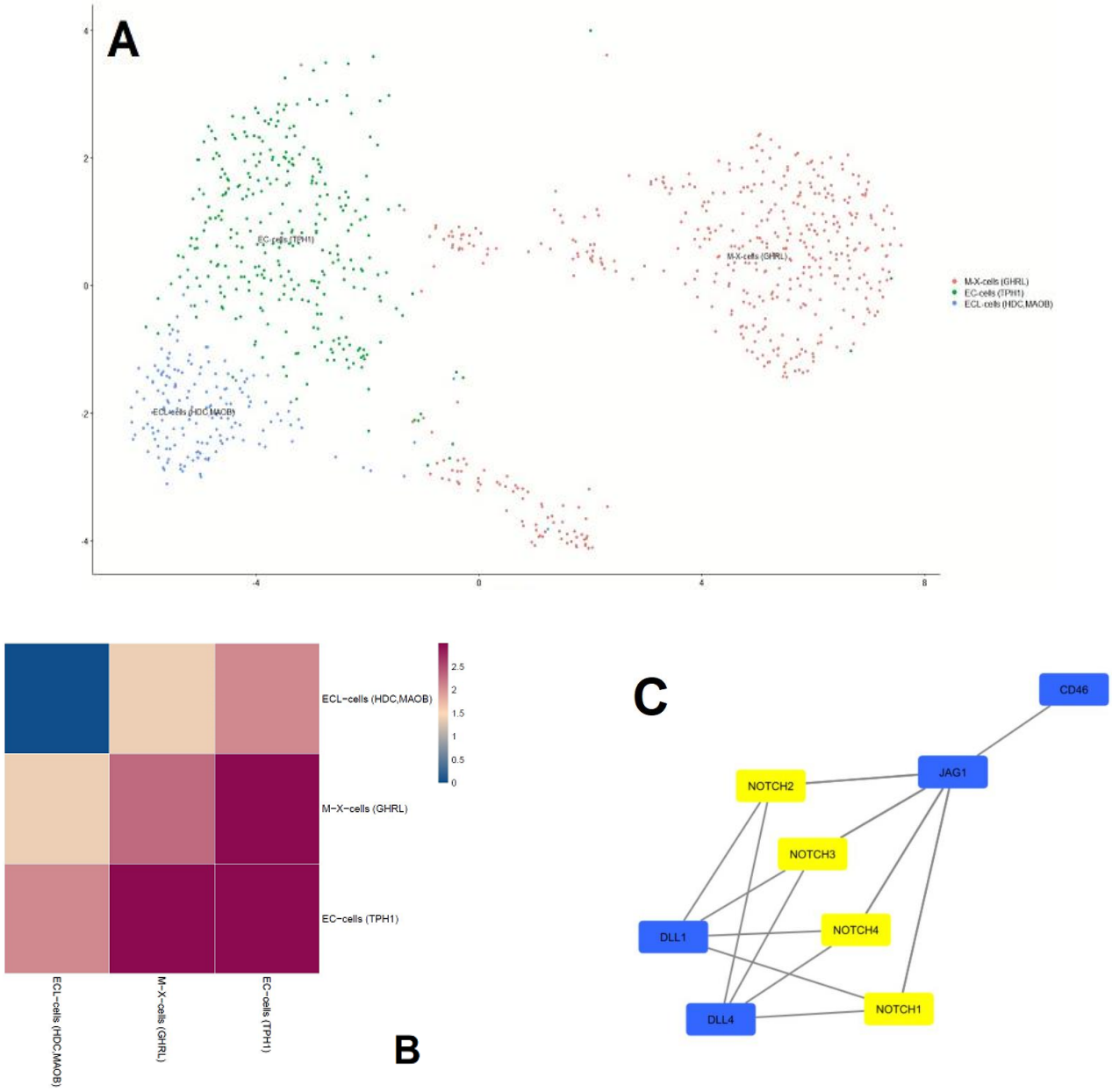

Figure 5: Single-Cell Analysis of the Human Stomach

(A)UMAP displaying the fundus and body regions of the human stomach EEC atlas (n=1,068 cells) after annotations. Different colors represent 3 separate clusters (k=10).

(B)Heatmap showing the total number of interactions between EEC types in the fundus and body regions of the human stomach dataset obtained with CellPhoneDB. Bar display color-coded amount of interactions (logarithmic scale).

(C)A subnetwork of ligand-receptor connections in the fundus and body regions of the human stomach EEC dataset obtained with CellPhoneDB significant connections output and Cytoscape. This part of the network shows the significant connections with the NOTCH ligands family.

5. Discussion

My main goal in working on the project was to shed light on how endocrine cells communicate with each other because there is not much literature that addresses this and there seems to be a lot more to discover in this area. In my opinion, and the opinion of many other researchers, if we know more about this mode of communication we can contribute to the effort for personal treatment and medication while taking advantage of the fact that these endocrine cells affect most parts of the body in one way or another. For example, GLP1 hormone secreted by L-cells known as a treatment for diabetes and obesity, the Modulators of Serotonin expressed by EC-cells effect Diarrhea and constipation, and HDC on reflux.

I decided to focus on the human intestinal area because I knew in advance that this area considered rich in endocrine cells and has a high probability of finding different types of these cells in which will lead to a wealth of results to which I can refer. In addition, to expand the canvas, I chose to produce results from the fundus area in the human stomach in the same way so that I can compare two different parts of the human digestive system and thus deepen the results obtained or even bring new results that I can address.

From data of about 8,500 cells, I filtered using biomarkers that distinguish endocrine cells, which brought me to a total amount of almost 3,000 cells, the vast majority of which are endocrine. After analyzing the clustering results of the remaining data with the genes expressions in each cluster, I was able to annotate 13 different groups in the cluster. This annotation includes endocrine cell groups known to science, division into several subgroups according to a specific gene, as

well as some groups, like Stem cells and Enterocytes, that are not endocrine cells but remained in the data for sanity check of the results and lack filtering.

CellPhoneDB indicates the amount of many ligand-receptor relationships between the M-X cell group characterized by the MLN and GHRL biomarkers and the EC-cell group characterized by the CHGA and REG4 biomarkers. This result explains the many movements of the gut, as there is a direct link between these biomarkers and bowel movements. In addition, I found that the amount of ligand-receptor bonds between cells from the same cluster is significantly less than those that bind different cell types and thus can better explain how the signals transmitted between cells in the gut.

By visualizing the results, I could discern that the most significant ligand-receptor bonds belong to the BMP family, which binds to the BMP receptor, and indicate that the endocrine cells communicate in this way and not necessarily through the hormones they secrete into target cells. An important biological conclusion that emerges from this result is that endocrine cells communicate within a single villus and less over long distances since communication via BMP occurs at short distances characterized by diffusion rather than through the bloodstream.

The results obtained for the stomach indicate a lesser amount of ligand-receptor bonds between the endocrine cells compared to the amount in the gut. This result indicates the importance of the endocrine cells in the gut and I assume that as I examine closer areas of

the stomach with the gut I will see an increasing amount of endocrine cell types and respectively other types of communication.
Furthermore, the results from the stomach reinforce my hypothesis that the communication between endocrine cells using BMP signaling versus specific hormones. The results show a significant relationship as part of the Notch communication pathway that influences binary fate decisions of cells that must choose between the secretory and absorptive lineages in the gut.

To conclude, the main research question was if endocrine cells work together as a distributed machine, or does each do its own thing, completely independent of each other. My hypothesis was partially right because yes, the EECs do work together and communicate with each other but not via hormones, i.e. not functionally, but in terms of development. That raises the question- what will be the right number of EEC types in one villus. If there are so few endocrine cells in each villus, how can we make sure we don't have villi with just EC cells, or another missing the L-cell that works with EC cells? My project suggests the answer to it connected with TGFbeta, further research needs to be done.

While working on the project I considered two weak points.
The data I worked with from the human gut was collected from organoids, which raises doubts about the reliability of the results as it is not clear how similar the data are to those existing in vivo. I hypothesize that because I focus on endocrine cells the importance of this reliability decreases and therefore the results do yield a good explanation for the form of communication between the endocrine cells. In addition, in

vivo data extraction is complicated to an impossible procedure in the case of the gut.

Another weakness is the reliability of CellPhoneDB results, which gives equal statistical weight between different signal types such as assembly, disassembly, and disconnection. This may mean that the results I have obtained do not properly indicate the communication pathways between the cells. In my opinion, because I chose to run the tool with data that I rearranged so that emphasizing specific markers helps me to trust the results I obtained. In addition, CellPhoneDB has non-validated interactions as it has been curated from many different resources. Moreover, CellPhoneDB count many interactions together, e.g. all BMP ligands interact with all BMP receptors, which leads to an exponential count of interactions.

As a future thought, I would like to examine the role of the BMP pathway in endocrine cells communication and the impact of this pathway on health and illness. In addition, I would like to perform the same process for endocrine cells in the antrum area of the stomach to complete the overall picture between the entrance to the stomach and the intestine as part of the digestive chain.

In conclusion, much remains to be researched on the communication pathways between endocrine cells in general and the gastrointestinal tract in particular. I hope that the results I have obtained in this project will form a good basis for further in-depth research on the subject

6. References

1) Beumer, J., Puschhof, J., Bauzá-Martinez, J., Martínez-Silgado, A., Elmentaite, R., James, K. R., ... & Clevers, H. (2020). High-resolution mRNA and secretome atlas of human enteroendocrine cells. Cell, 181(6), 1291-1306.
2) Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W. M., ... & Satija, R. (2019). Comprehensive integration of single-cell data. Cell, 177(7), 1888-1902.
3) Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nature protocols, 15(4), 1484-1506.
4) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), 2498-2504.
5) Buffa, R., Capella, C., Fontana, P., Usellini, L., & Solcia, E. (1978). Types of endocrine cells in the human colon and rectum. Cell and tissue research, 192(2), 227-240.
6) Gunawardene, A. R., Corfe, B. M., & Staton, C. A. (2011). Classification and functions of enteroendocrine cells of the lower gastrointestinal tract. International journal of experimental pathology, 92(4), 219-231.
7) Gehart, H., van Es, J. H., Hamer, K., Beumer, J., Kretzschmar, K., Dekkers, J. F., ... & Clevers, H. (2019). Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. Cell, 176(5), 1158-1173.
8) Sander, G. R., & Powell, B. C. (2004). Expression of notch receptors and ligands in the adult gut. Journal of Histochemistry & Cytochemistry, 52(4), 509-516.