

Project presentation: A Single Cell data analysis for Chron's Disease with COTAN

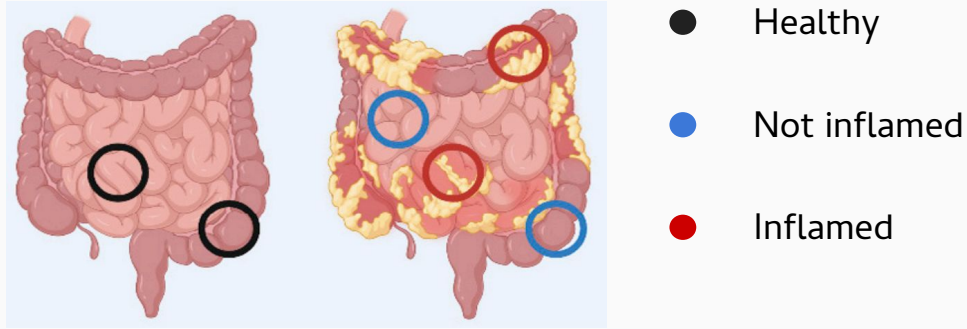
Ninniri Matteo (student ID: 543873)

Piras Andrea (student ID: 619640)

Outline

- Introduction to Chron's Disease
- Datasets summary
- Data preprocessing with COTAN
- Clustering & making sense of it
- Gene set enrichment analysis
- Finding new markers
- Single patient analysis

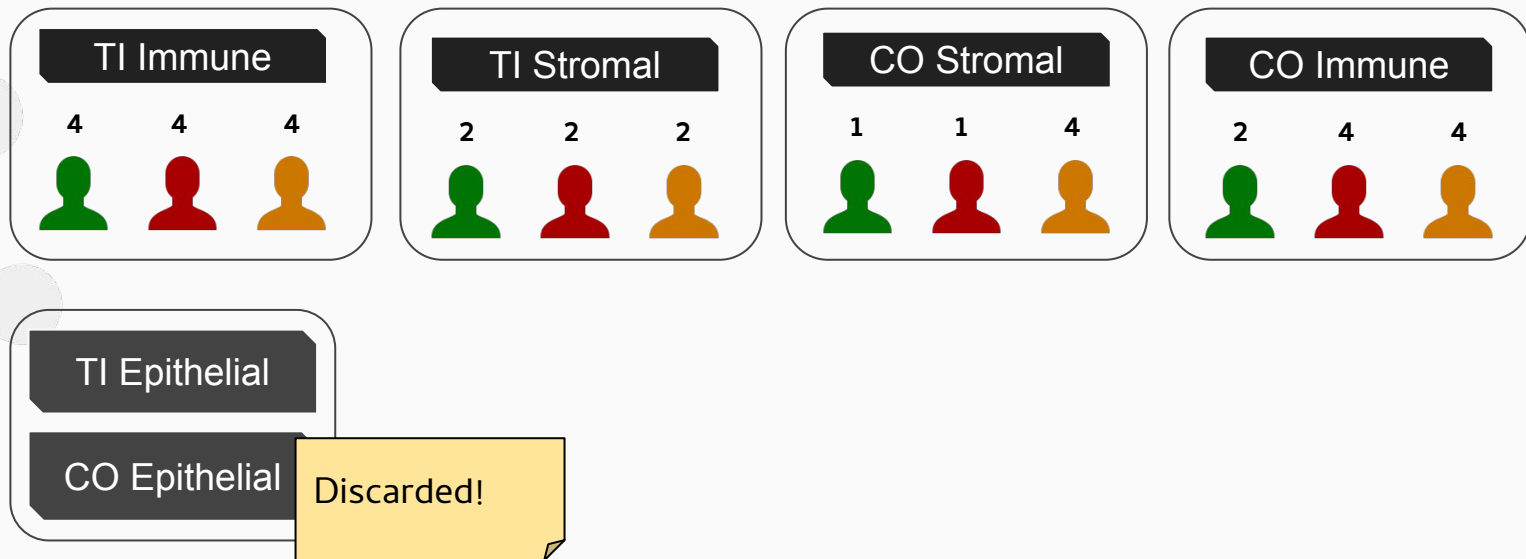
Introduction to Chron's disease



- Chronic gastrointestinal autoimmune disease
- Understand the causes is very complex
- Single-cell RNA sequencing

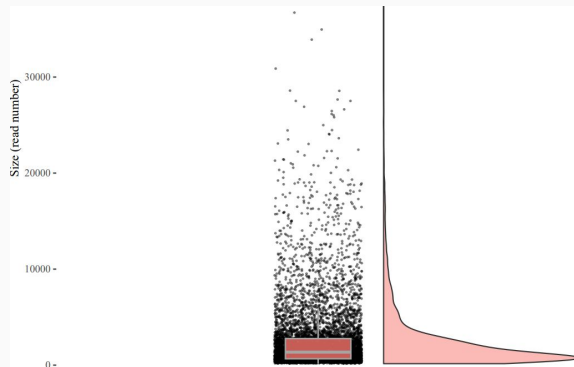
Datasets summary

- 71 donors
- 25 Healthy
- 46 With the disease
- ~720.000 cells total
- Only samples with > 1000 cells

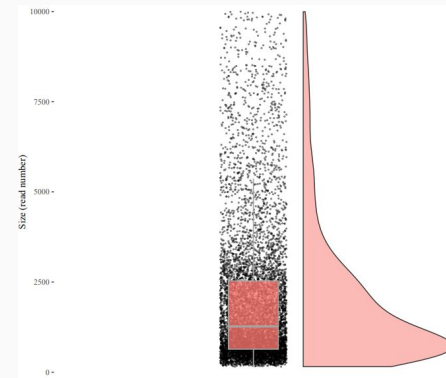


Data preprocessing with COTAN- Cells and genes

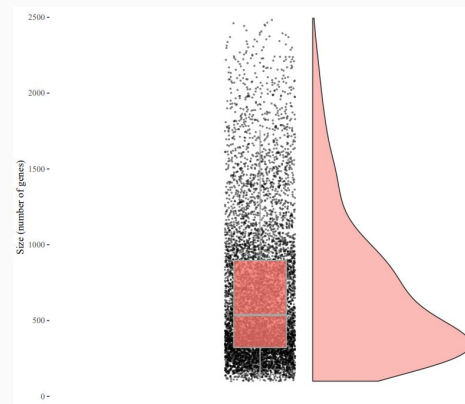
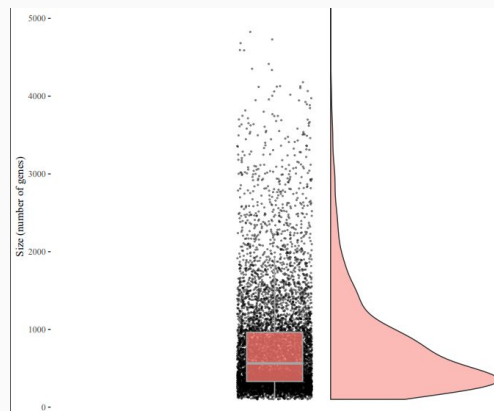
Cell count



After cut

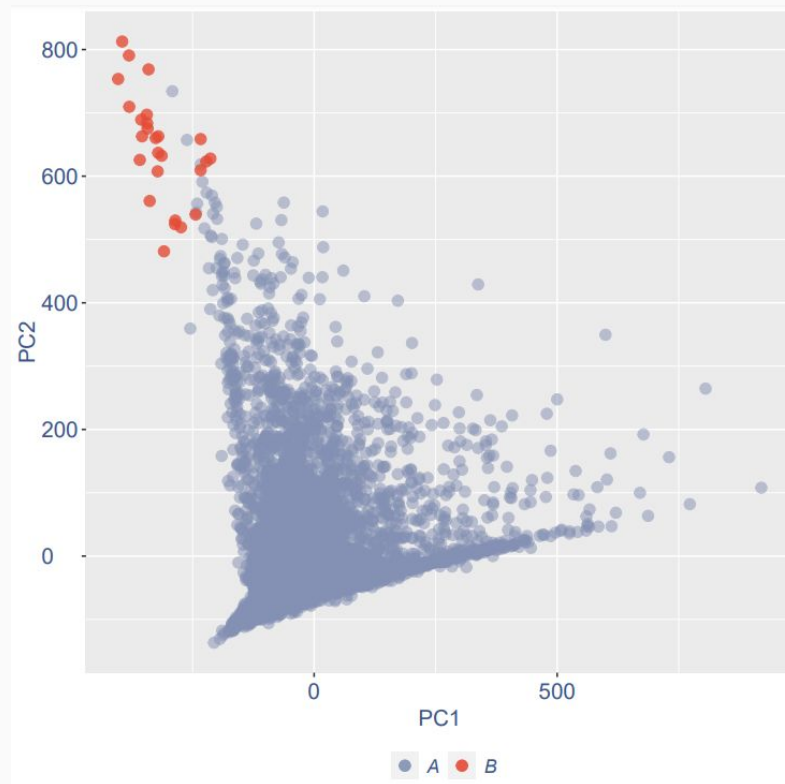


Gene count



Data preprocessing - Cluster B

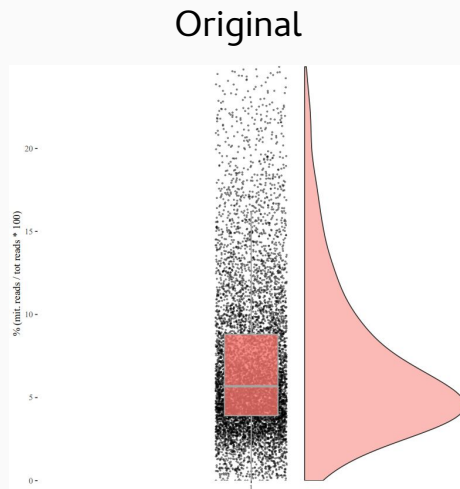
Original



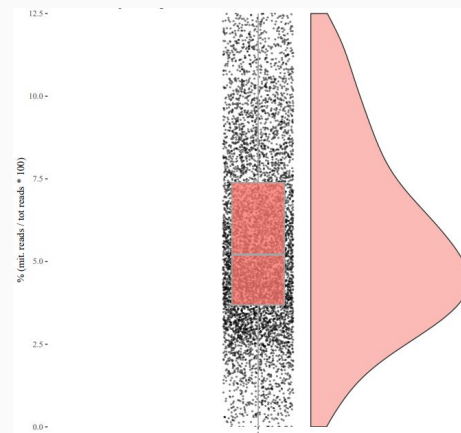
Cluster B removal

Data preprocessing - Mitochondrial and UDE

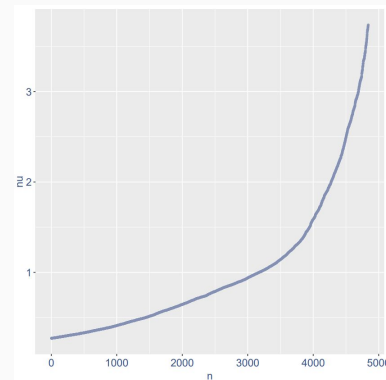
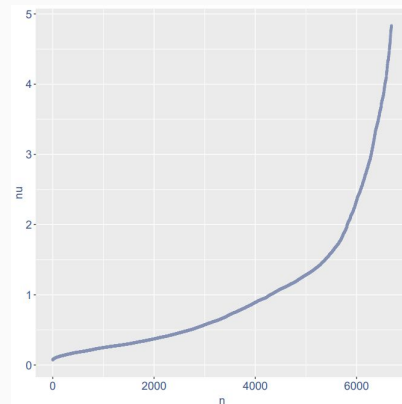
Mitochondrial percentage



After cut



UDE (UMI efficiency
detection)



Making sense of the cell clusters

Two approaches:

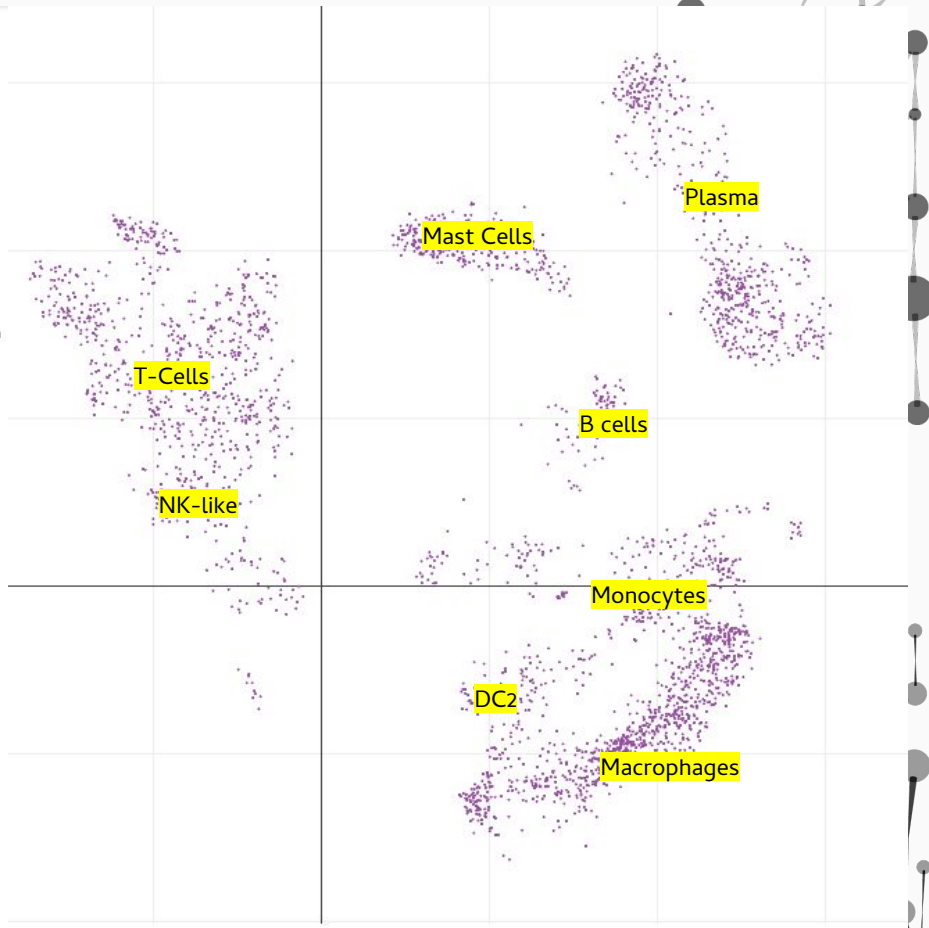
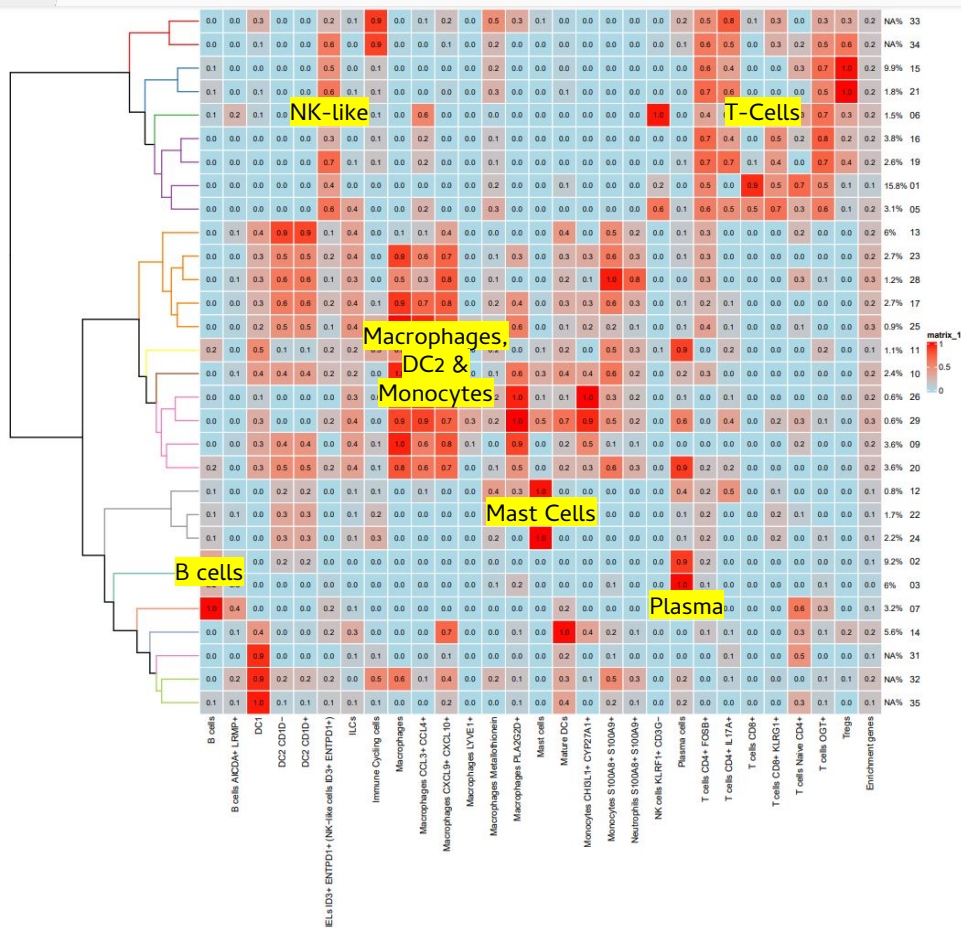
- Literature search
- Use well known cell types

Stromal

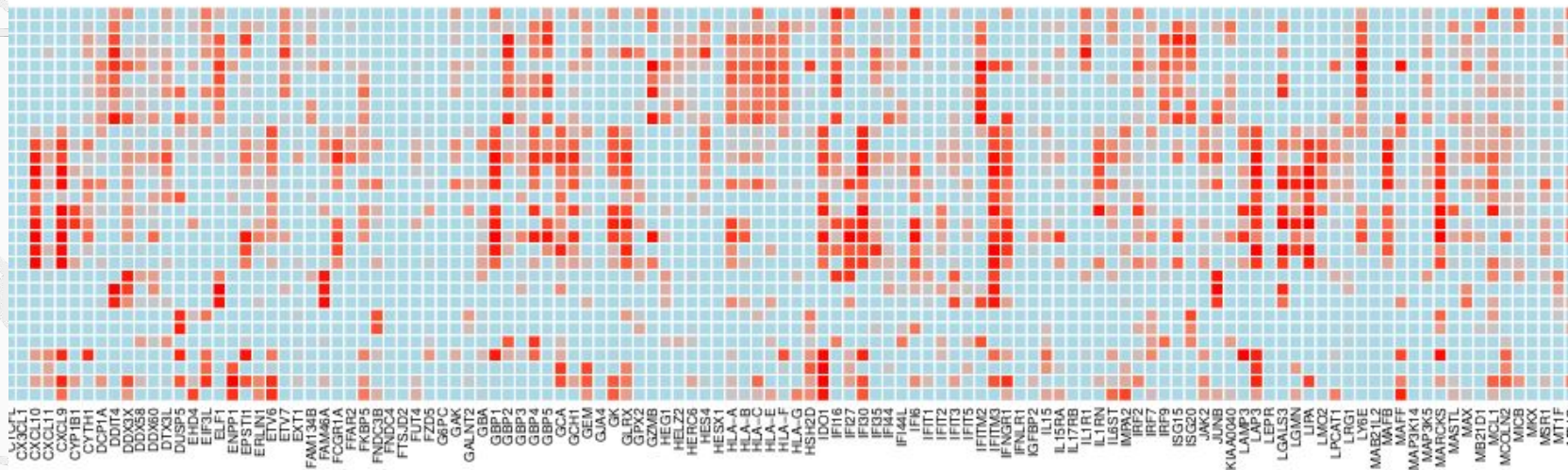
- Activated fibroblast.CCL19+.ADAMDEC1+
- Endothelial cells CA4+ CD36+
- Endothelial cells CD36+
- Endothelial cells DARC+
- Endothelial cells LTC4S+ SEMA3G+
- Fibroblast ADAMDEC1++
- Fibroblast KCNN3+ LY6H+
- Fibroblast NPY+ SLITRK6+
- Fibroblast SFRP2+ SLP1+
- Fibroblast SMOC2+ PTGIS+
- Glial cell
- Inflammatory fibroblasts IL11+ CHI3L1+
- Lymphatics
- Myofibroblast GREM1+ GREM2+
- Myofibroblast HHIP+ NPNT+
- Pericyte HIGD1B+ STEAP4+
- Pericytes RERGL+ NTRK2+
- Stromal Cycling Cells

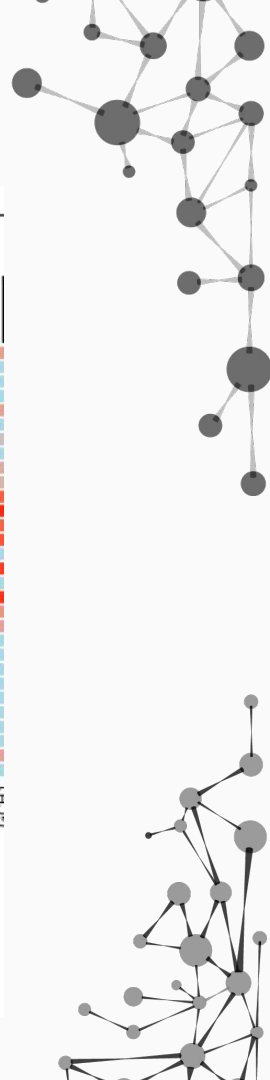
Immune

- B cell
- B cells AICDA+ LRMP+
- DC1
- DC2 CDID+
- DC2 CDID-
- IELs ID3+ ENTPD1+
- ILCs
- Immune Cycling cells
- Macrophages
- Macrophages CCL3+ CCL4+
- Macrophages CXCL9+ CXCL10+
- Macrophages LYVE1+
- Macrophages Metallothionein
- Macrophages PLA2G2D+
- Mast cells
- Mature DCs
- Monocytes CHI3L1+ CYP27A1+
- Monocytes S100A8+ S100A9+
- NK cells KLRF1+ CD3G-
- Neutrophils S100A8+ S100A9+
- NK-like cells ID3+ ENTPD1+
- Plasma cells
- T cells CD4+ FOXP3+
- T cells CD4+ IL17A+
- T cells CD8+
- T cells CD8+ KLRG1+
- T cells Naive CD4+
- T cells OGT+
- Tregs



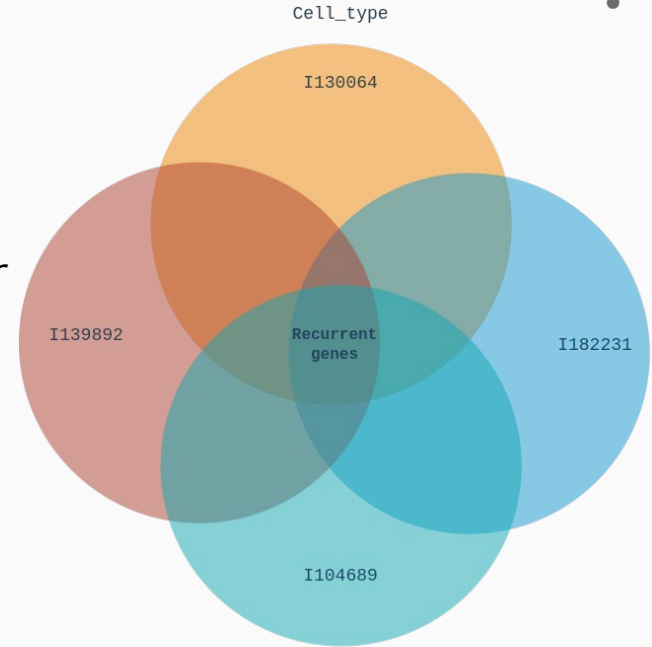
Gene Enrichment





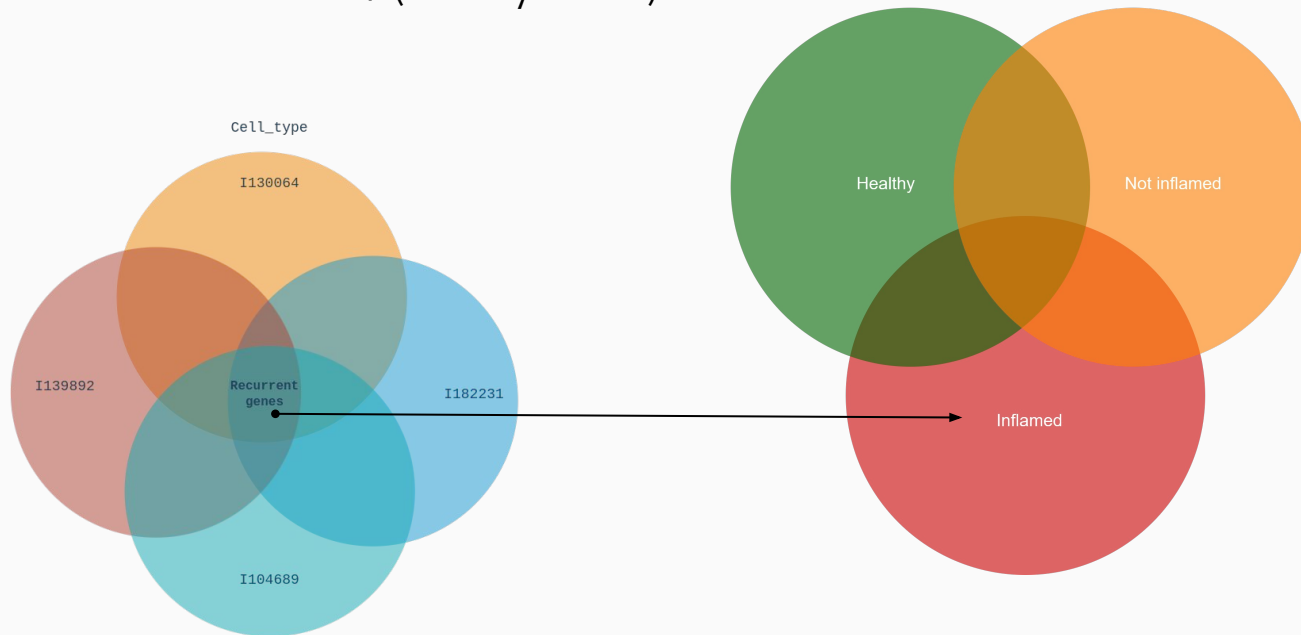
How to spot new disease markers?

1. Get multiple patients
2. Clusterize their enrichment
3. For each cell type:
 - a. Intersect the cluster where they appear



To obtain the genes exclusive to a condition, just subtract the union of the other sets from it

- Healthy markers = Healthy / (Not infl. \cup Infl.)
- Not infl. markers = Not infl. / (Healthy. \cup Infl.)
- Infl. markers = Infl. / (Healthy \cup Infl.)



Results

Terminal Ileum Immune dataset: Inflamed samples

- IELs ID3+ ENTPD1+
 - ANXA2R
- Monocytes S100A8+ S100A9+
 - THBD
- Neutrophils S100A8+ S100A9+
 - TCF7L2
- Plasma cells
 - CD38, COMMD3
- T cells CD4+ FOXP3+ & T cells OGT+
 - SOCS1;SLFN5;RARRES3



Results

Terminal Ileum Immune dataset: Non-inflamed samples

- B cells
 - MX1, HSH2D, AIM2, STAP1
- B cells AICDA+ LRMP+
 - MX1, HSH2D, AIM2
- DC1
 - LAMP3
- DC2 CD1D- & DC2 CD1D+
 - PLSCR1

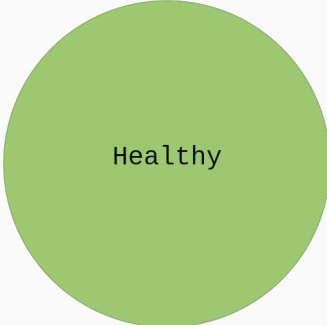


Not-inflamed

Results

Terminal Ileum Immune dataset: healthy controls

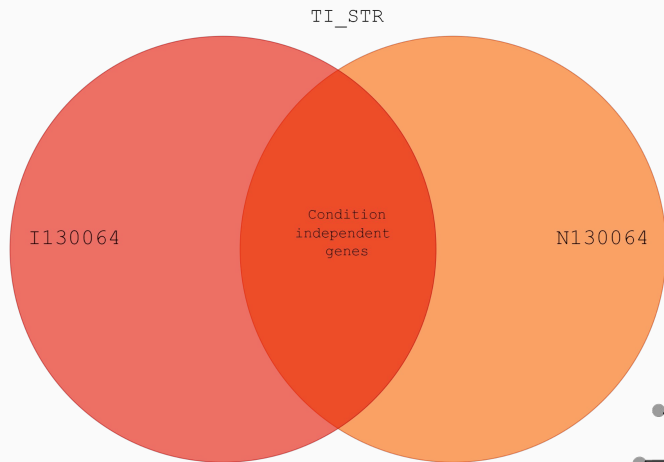
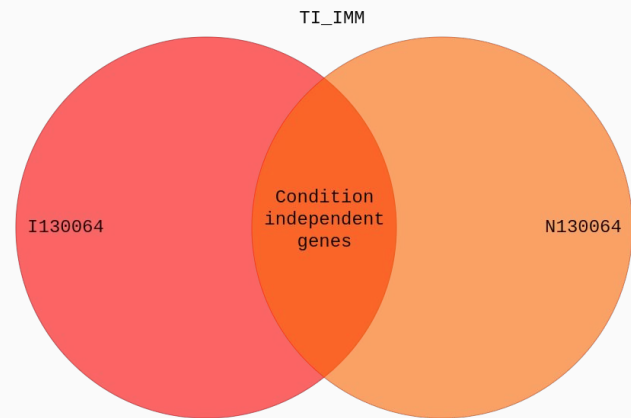
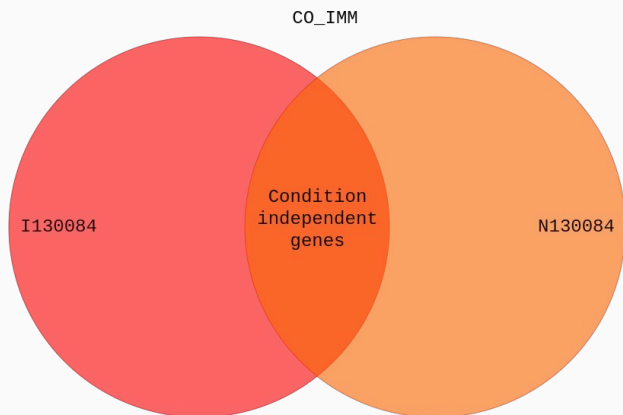
- B cells & B cells AICDA+ LRMP+
 - NCF1, P XK
- Macrophages CCL3+ CCL4+
 - LIPA
- Macrophages CXCL9+ CXCL10+
 - CASP7
- Macrophages LYVE1+
 - FCGR1A
- Monocytes CHI3L1+ CYP27A1+
 - NDC80, HELZ2, IL15RA
STARD5, SLC16A1, PI4K2B



Healthy

- T cells CD8+ & T cells CD8+ KLRG1+:
 - CCL5, GZMBCCL5, GZMB
- T cells Naive CD4+
 - SUN2, SOCS1, SLFN5, FAM134B, ANXA2R
- Tregs
 - SUN2, SOCS1, SLFN5, FAM134B
- T cells CD4+ IL17A+
 - DDX58

Patient-specific markers



Future developments

- Better metrics for clustering
- More K-means iterations for the enrichment (column_km_repeats)

The slide features a light gray background with decorative network-like graphics in the corners. On the left, there are several light gray clusters of circles connected by lines. On the right, there are two clusters of dark gray circles connected by lines, one near the top and one near the bottom.

Thank you for the attention!

How to determine the optimal number of clusters?

1. Enumerate all the cluster combinations
2. For each combination:
 - 2.1. For each cell type:
 - 2.1.1. Get the clusters where the cell appears (one for each patient)
 - 2.1.2. Jaccard
 - 2.2. Average the scores
3. Optimal combination: the one with the highest average Jaccard