

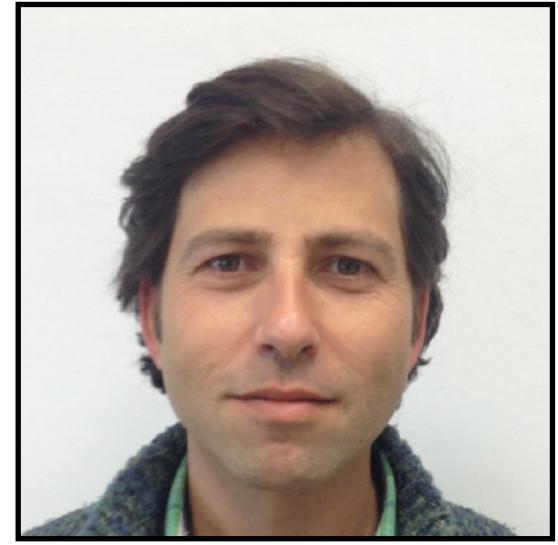
Single-cell RNA-Seq analysis

Chris Smillie

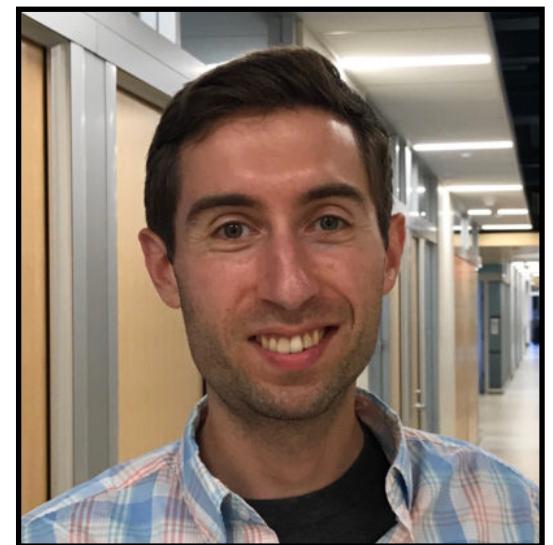
2018 Computational Genomics Workshop



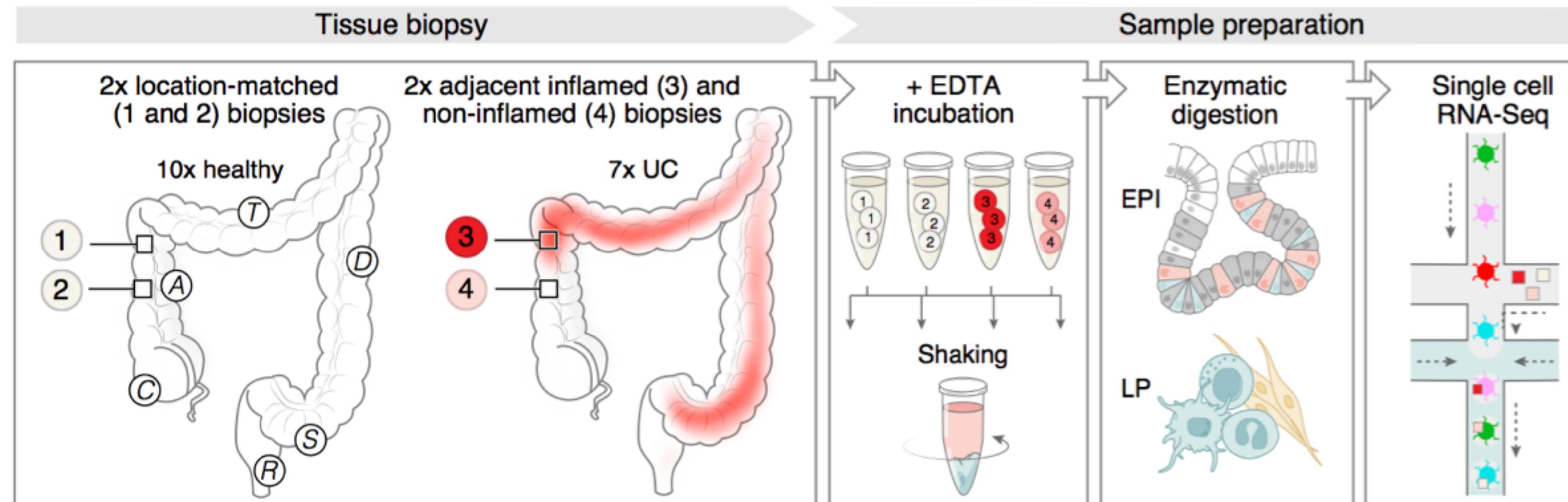
Single-cell transcriptomes of biopsy specimens from 10 healthy individuals and 7 UC patients



Moshe Biton
(Regev/Xavier Labs)



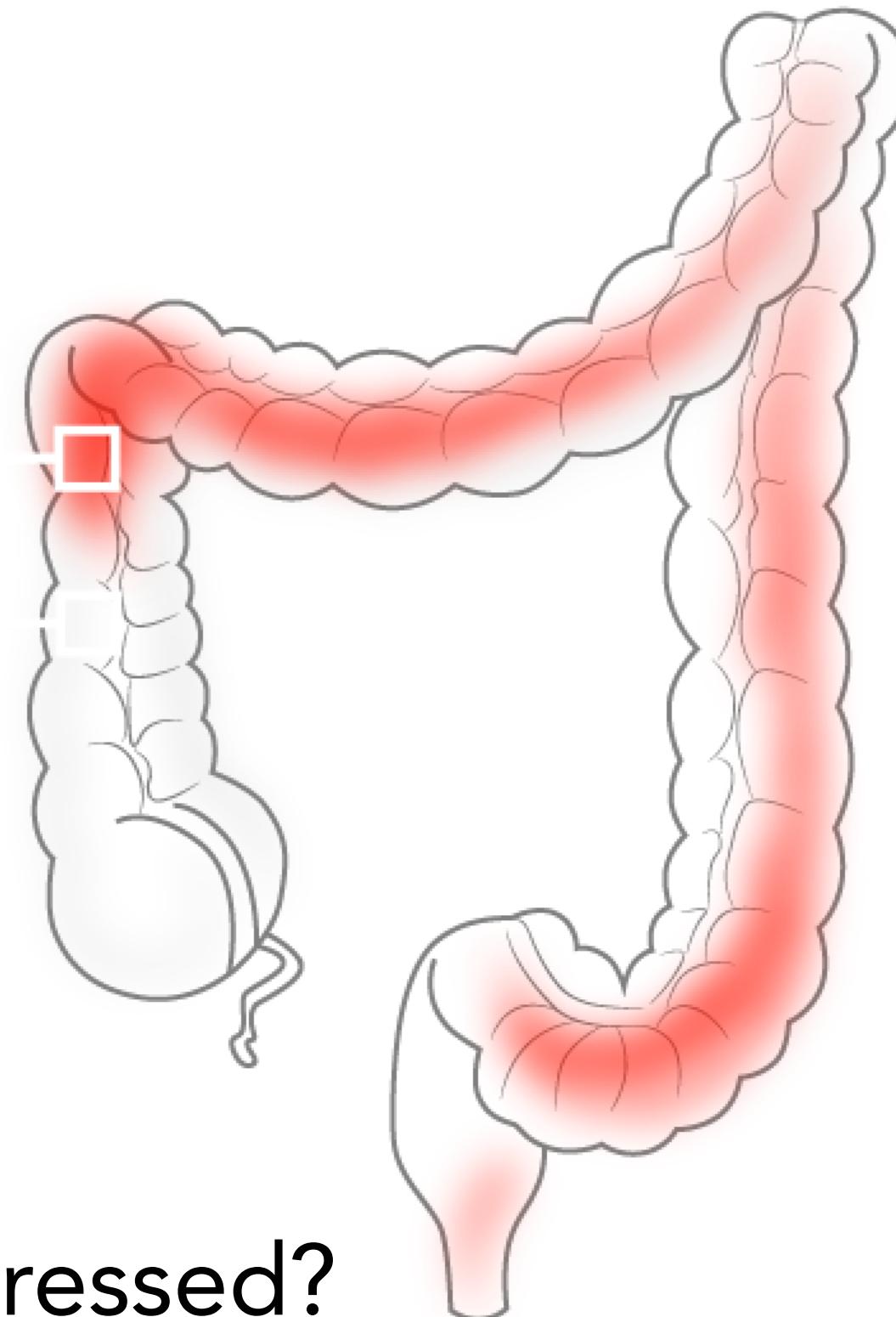
Jose Ordovas
(Shalek Lab)



115,517 cells, 17 people = **34** colon biopsies = **68** 10X channels

What can we learn from single cell data?

1) What **cell types** are present?



2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

6) Can we use single cell data to understand IBD risk genes?

Noise in single cell data

Technical sources:

- 10-20% of transcripts sequenced per cell
- PCR amplification
- Library complexity

Biological sources:

- Transcriptional stochasticity
- Cell unhappiness

Noise in single cell data

Technical sources:

- 10-20% of transcripts sequenced per cell
- PCR amplification
- Library complexity

Biological sources:

- Transcriptional stochasticity
- Cell unhappiness

Solutions:

- Integrate over cells (cell clusters)
- Integrate over genes (gene signatures)
- Statistical methods that account for sparsity and complexity

What can we learn from single cell data?

1) What **cell types** are present?

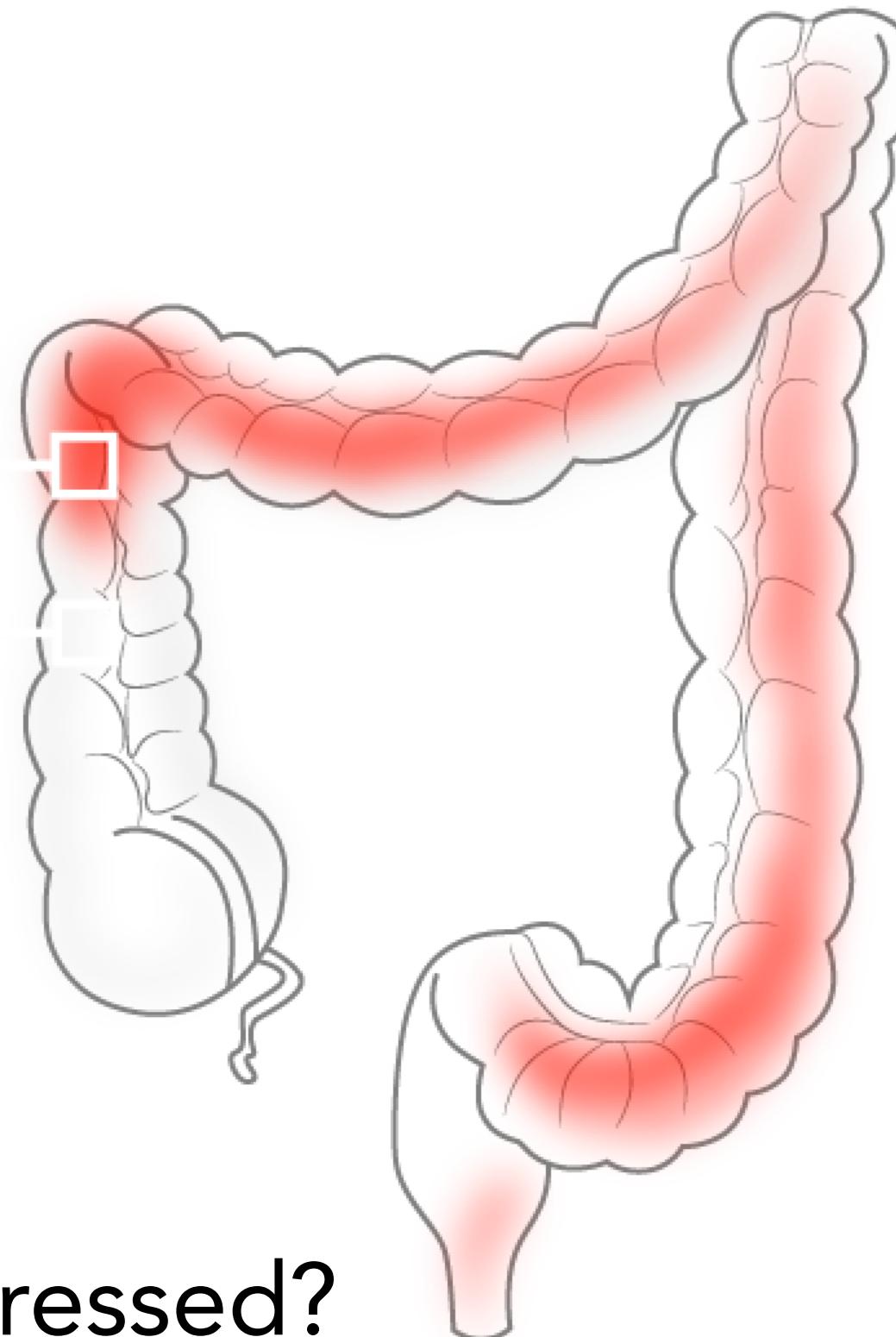
2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

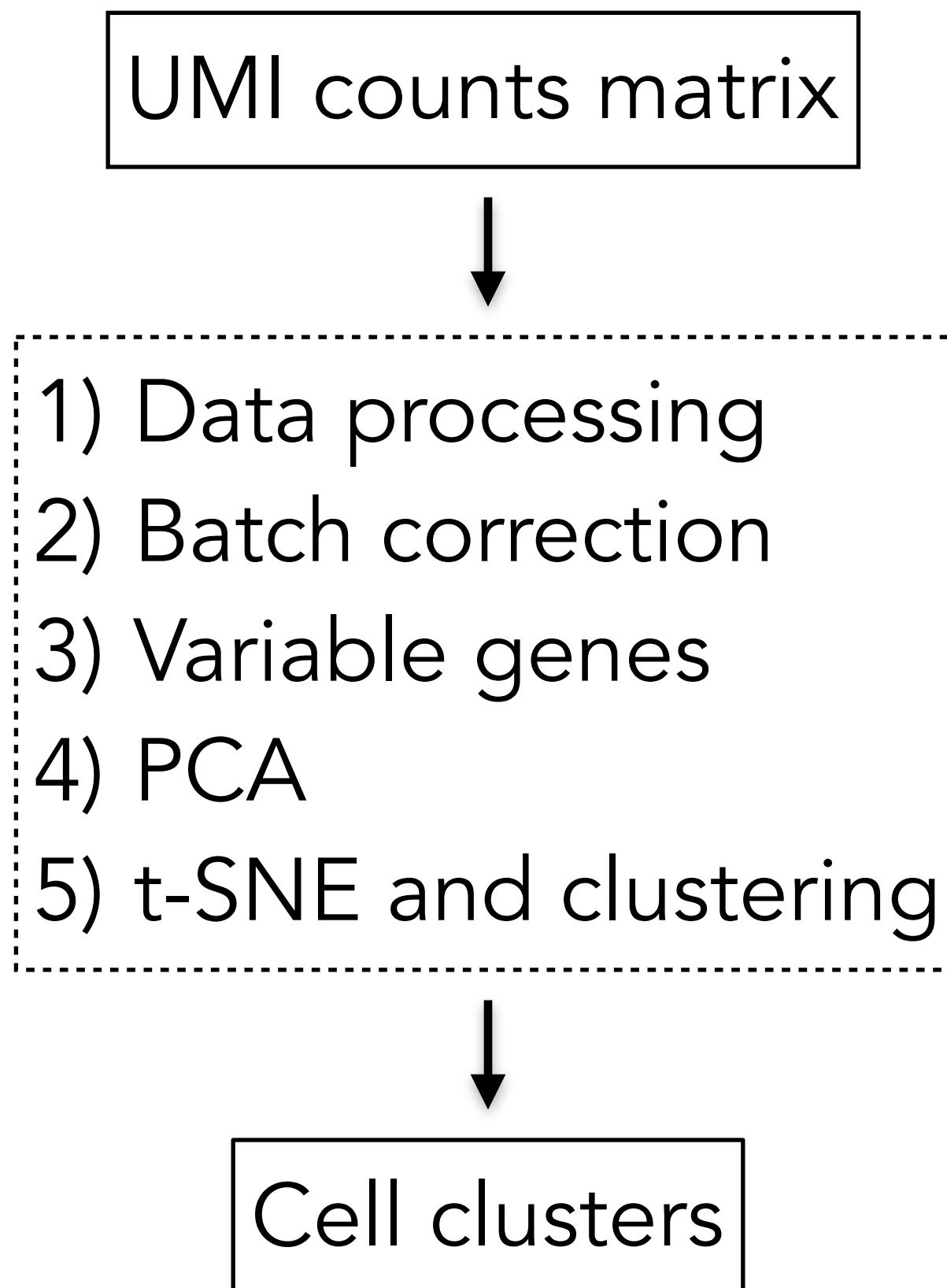
4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

6) Can we use single cell data to understand IBD risk genes?



Clustering single cells into cell types



Data processing and batch correction

Data processing

UMI counts matrix



$\log_2(\text{TPM} + 1)$



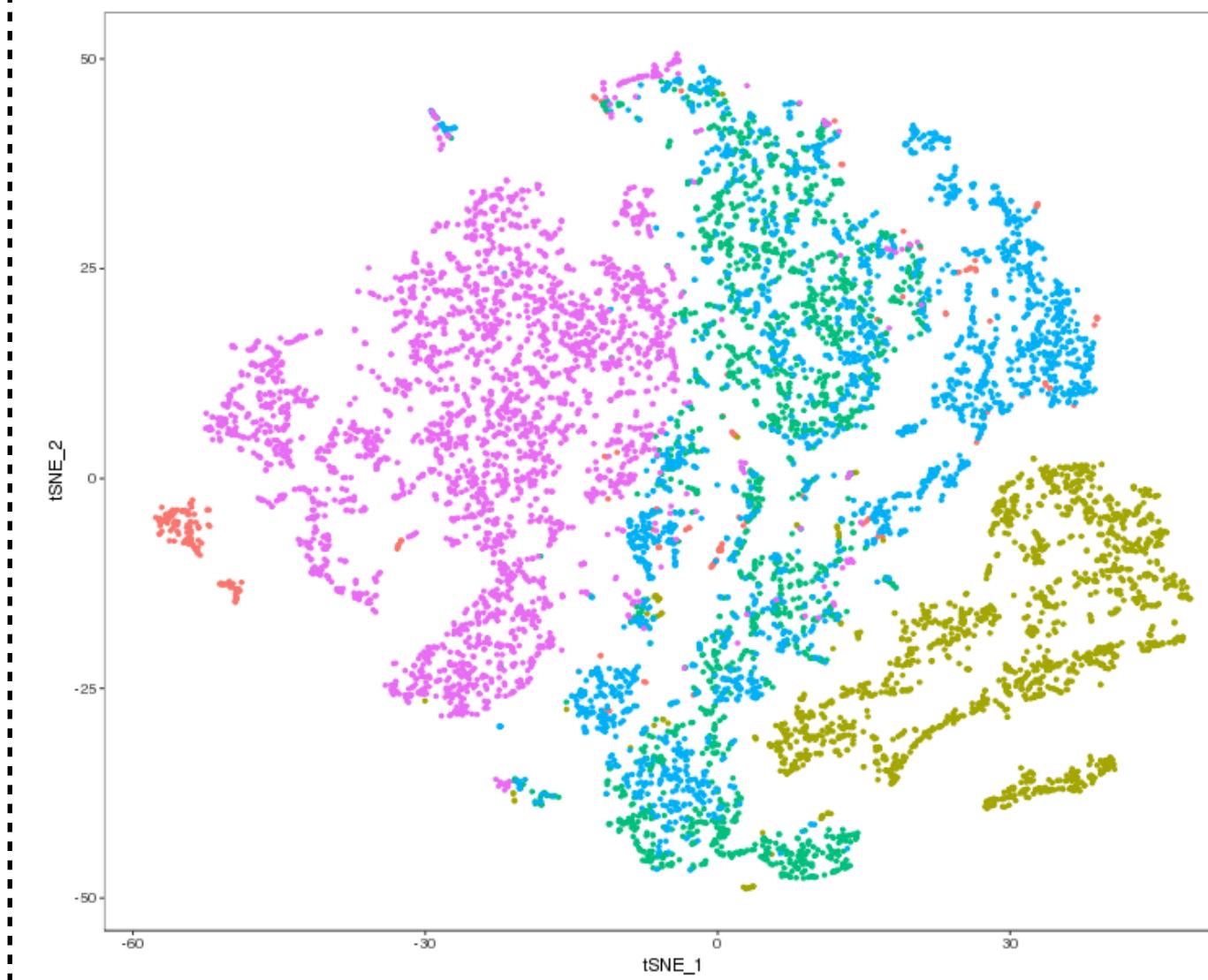
Filter low complexity cells



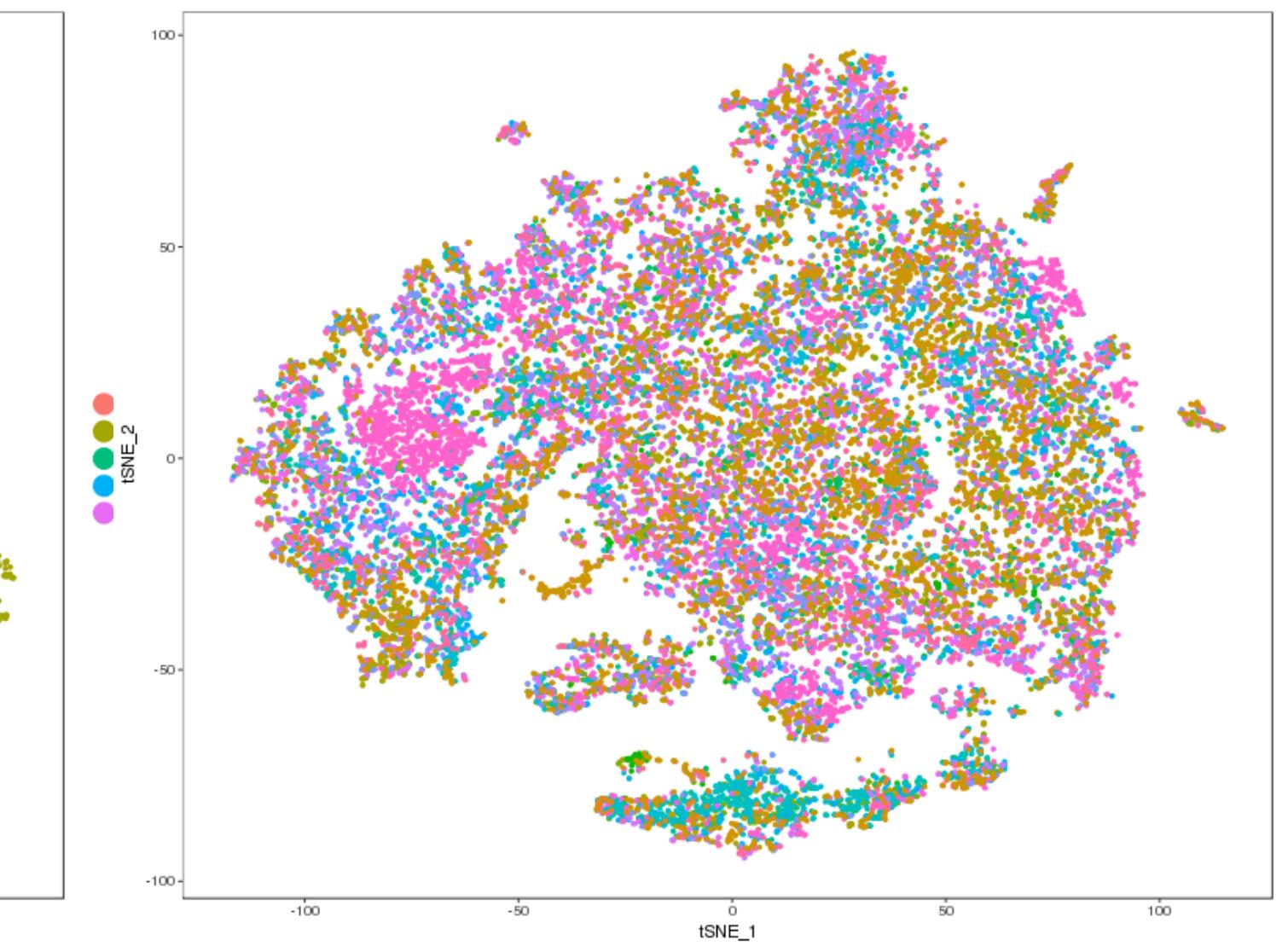
Filter lowly expressed genes

Batch correction (ComBat)

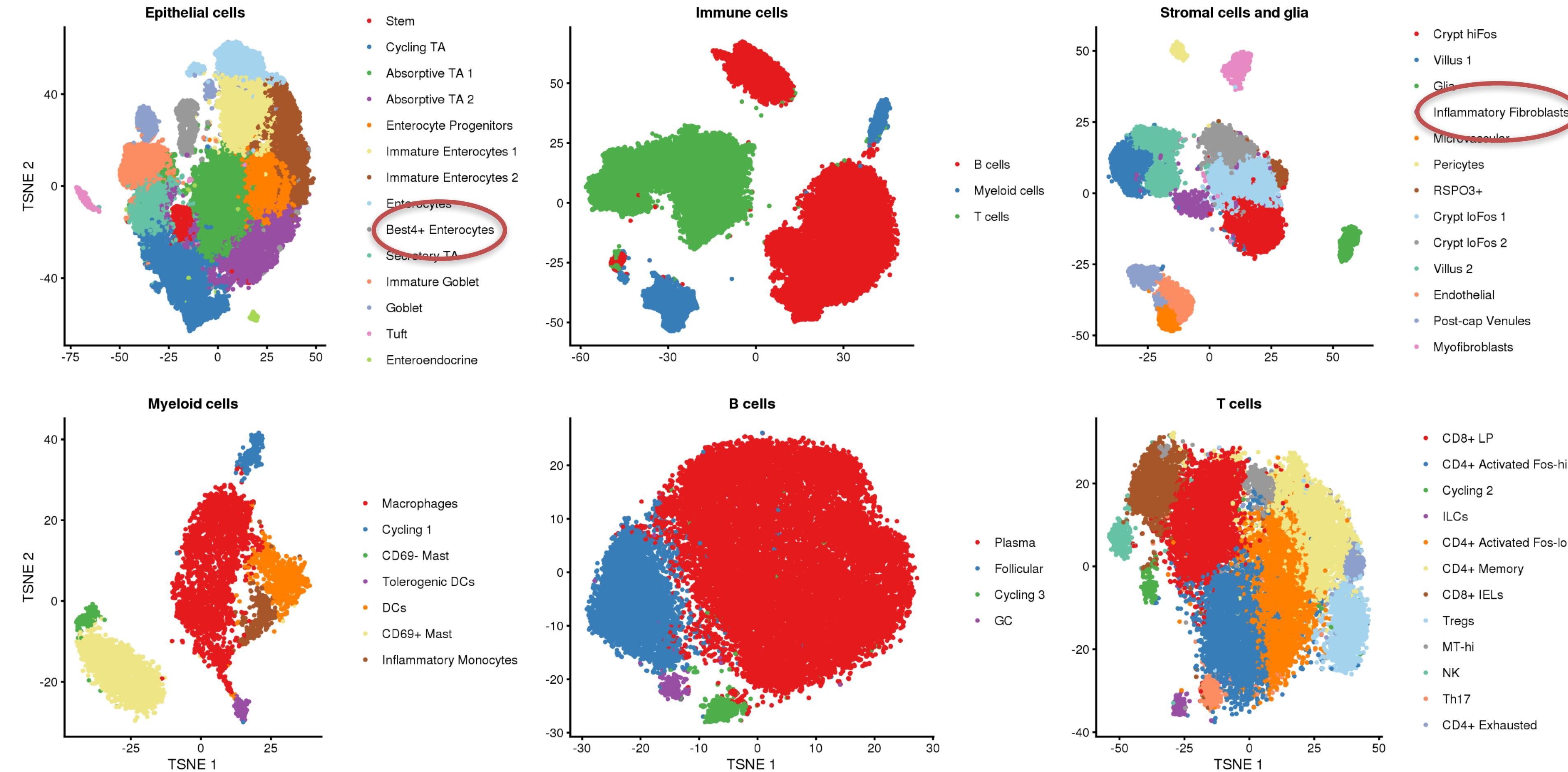
Pre-correction



Post-correction



Single cell atlas reveals nearly all cell types in colon



What can we learn from single cell data?

1) What **cell types** are present?

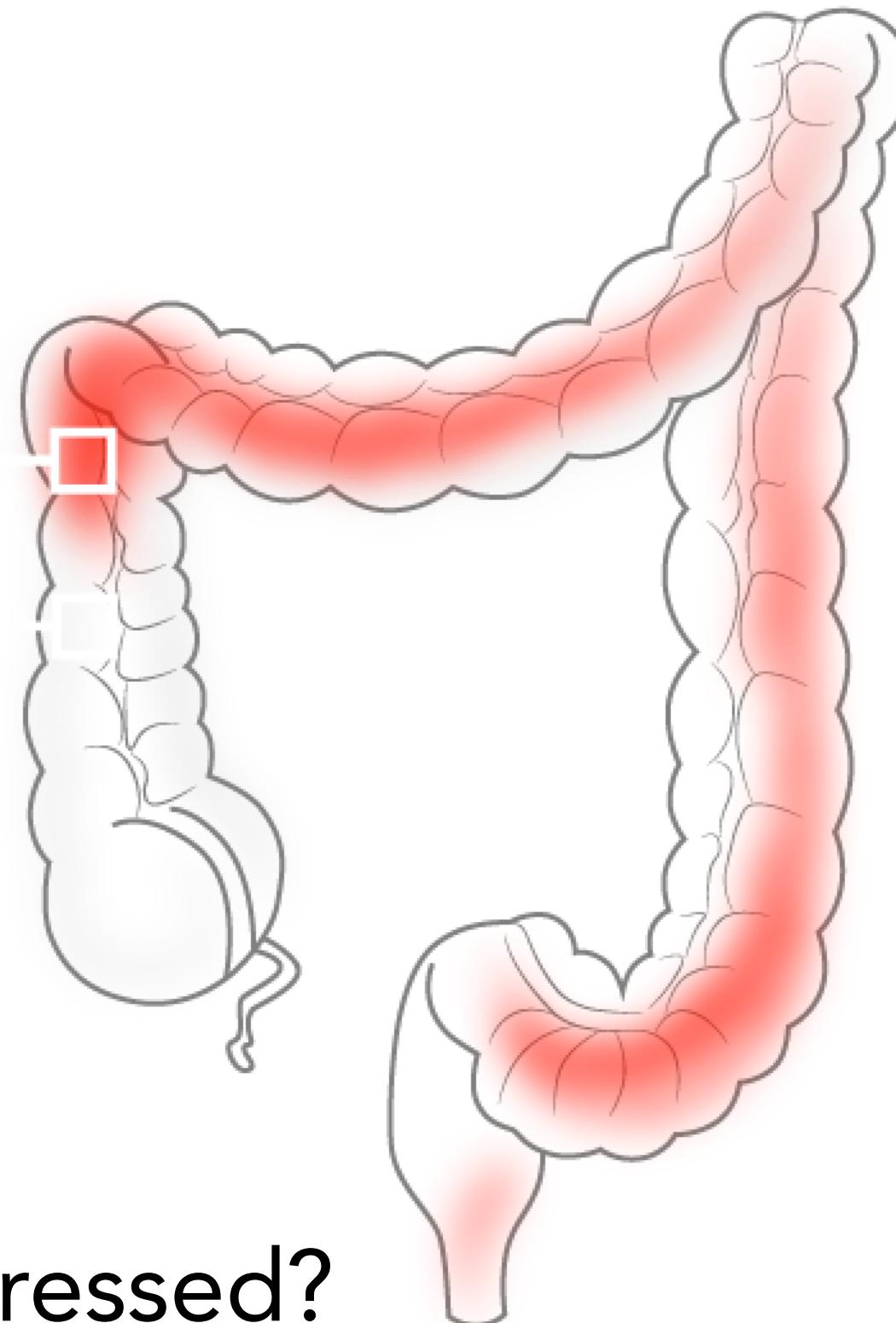
2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

6) Can we use single cell data to understand IBD risk genes?



Identifying changes in cell proportions with single cells

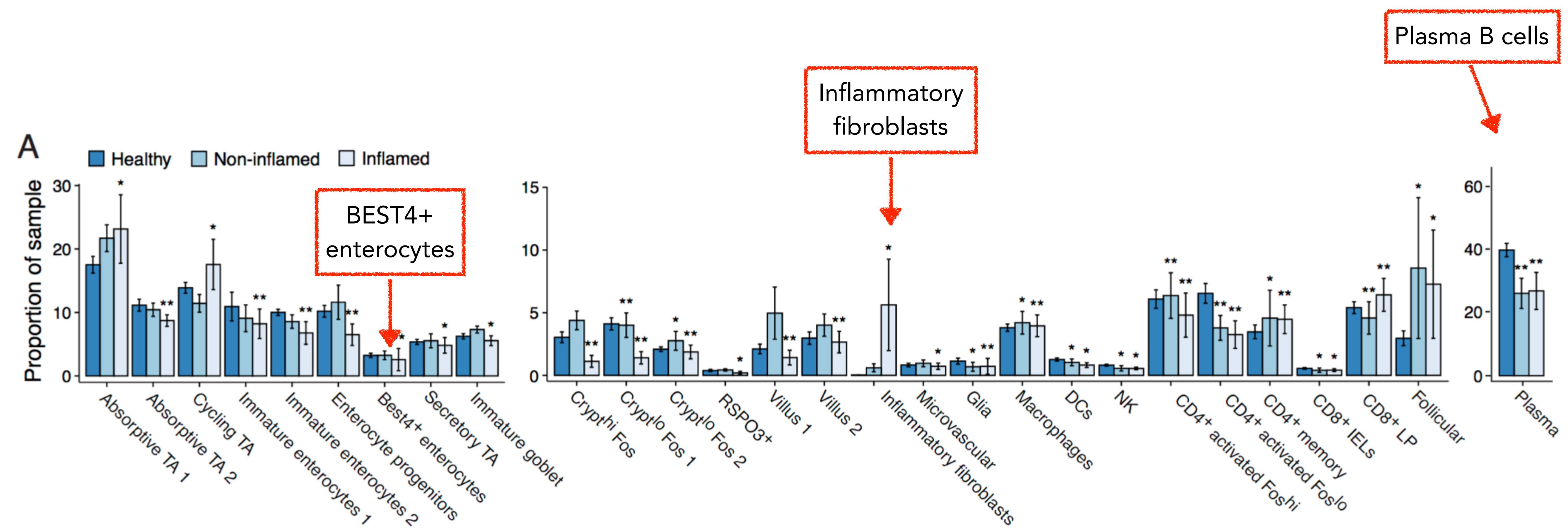
Problems:

- Cell filtering: can differentially impact cell types
- Proportionality: when one cell increases, all others must decrease

Solutions:

- Calculate cell proportions using minimally filtered data
- Use statistical model that accounts for proportionality
(e.g. *Dirichlet-multinomial regression*)

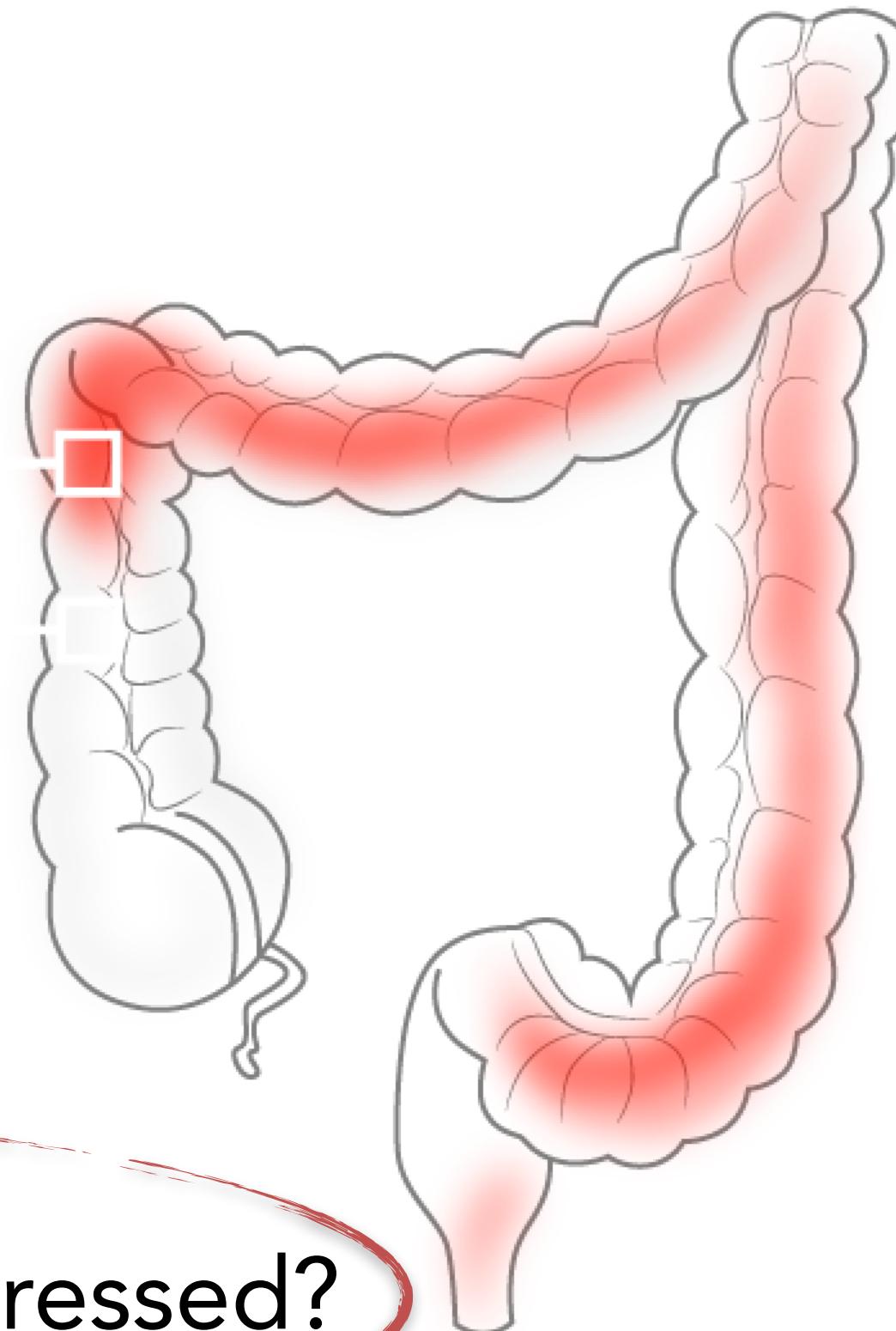
Remodeling of the colon during disease



Credit: Adam Haber original implementation of method

What can we learn from single cell data?

1) What **cell types** are present?



2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

6) Can we use single cell data to understand IBD risk genes?

Finding differentially expressed genes in specific cell types

Problems:

- Single cell data are sparse
- Many factors can impact gene expression
- Most genes are correlated with cell complexity

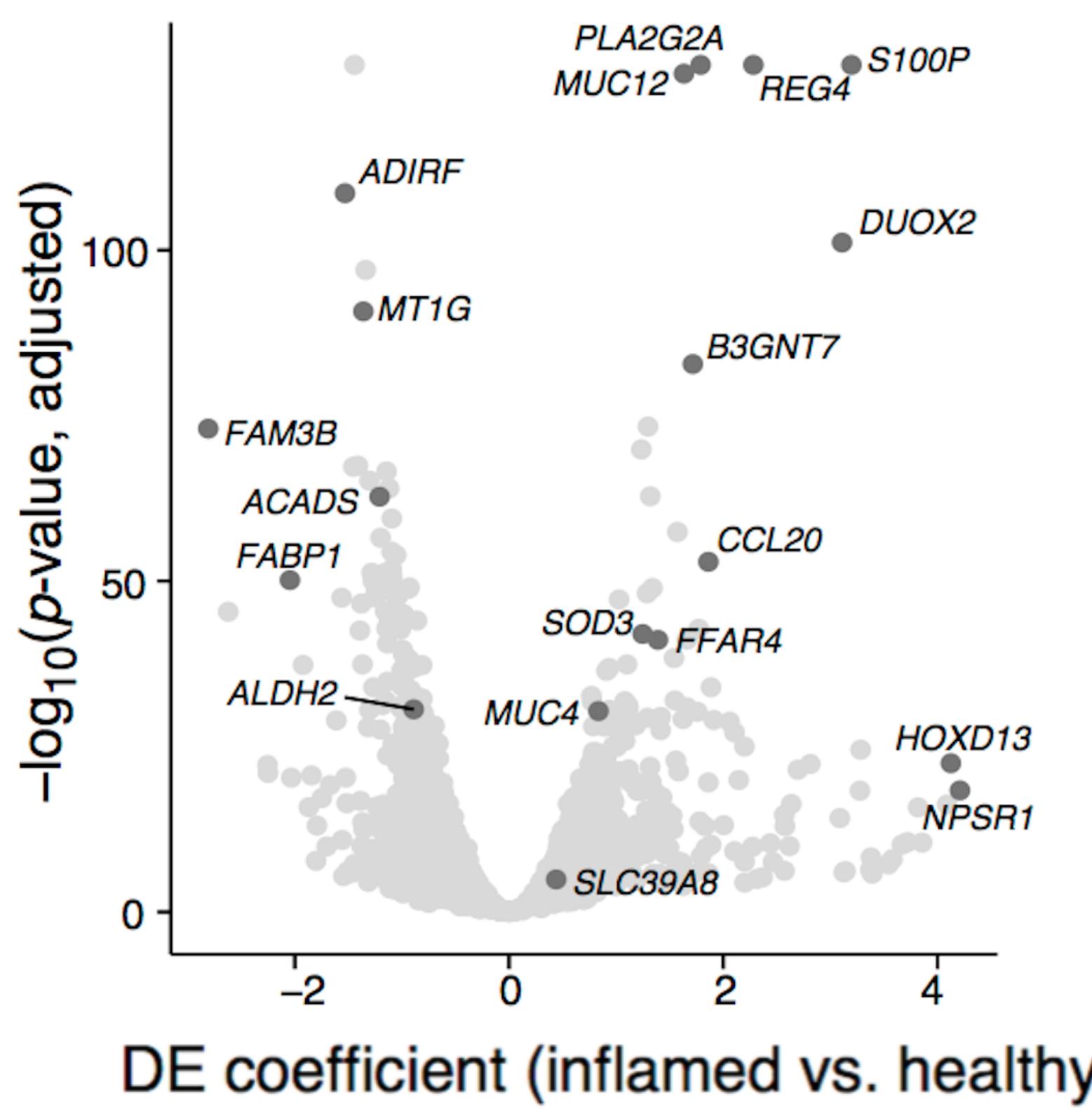
Solutions:

- Use a statistical model that accounts for sparse data (e.g. MAST)
- Regression framework to control for many sources of variation:

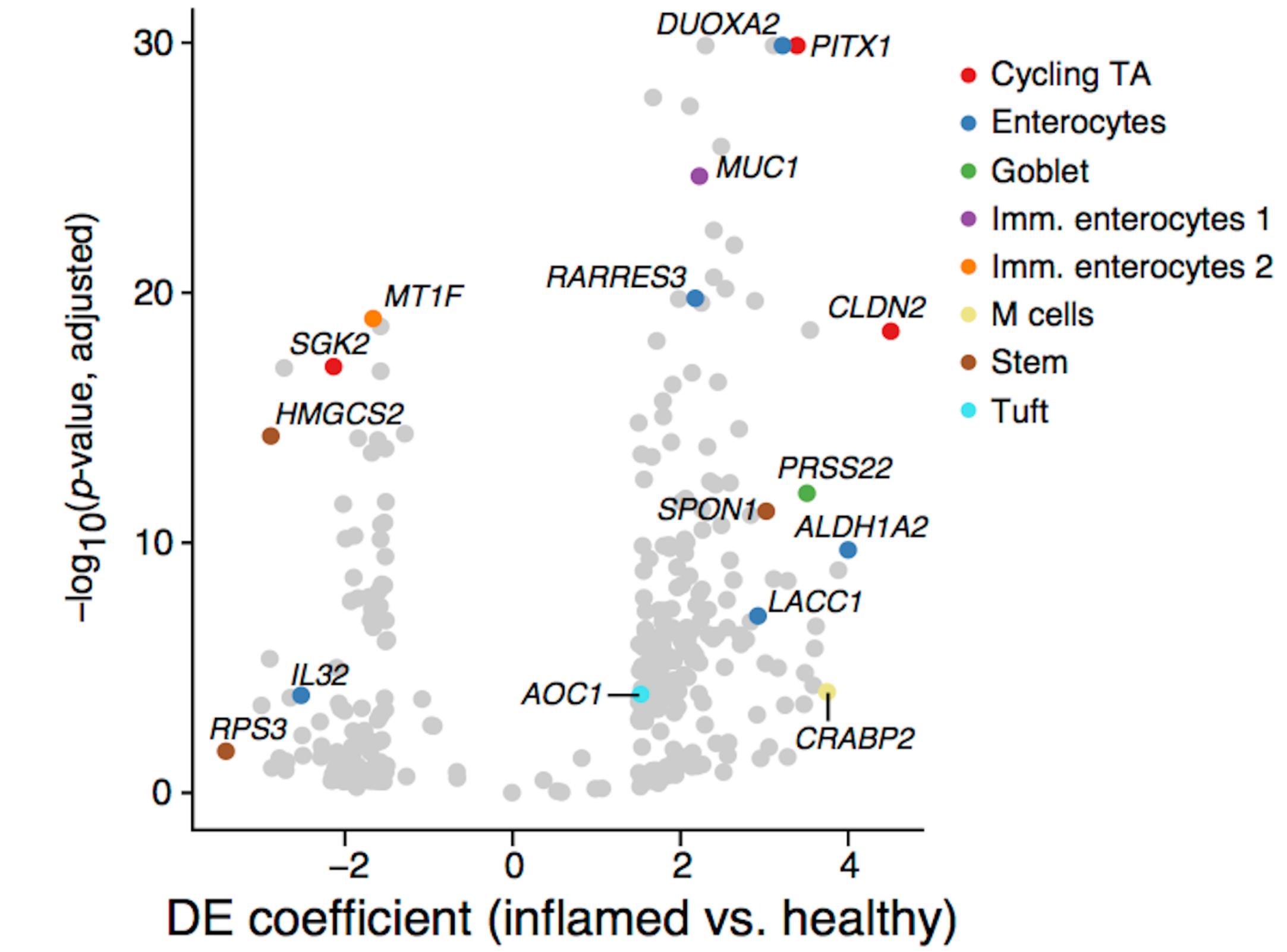
$$\text{Expression} \sim (\text{Cell quality}) + (\text{Cell type}) + (\text{Disease}) + (\text{Cell type} * \text{Disease})$$

DE genes in epithelial cells during ulcerative colitis

Shared across all epithelial cells

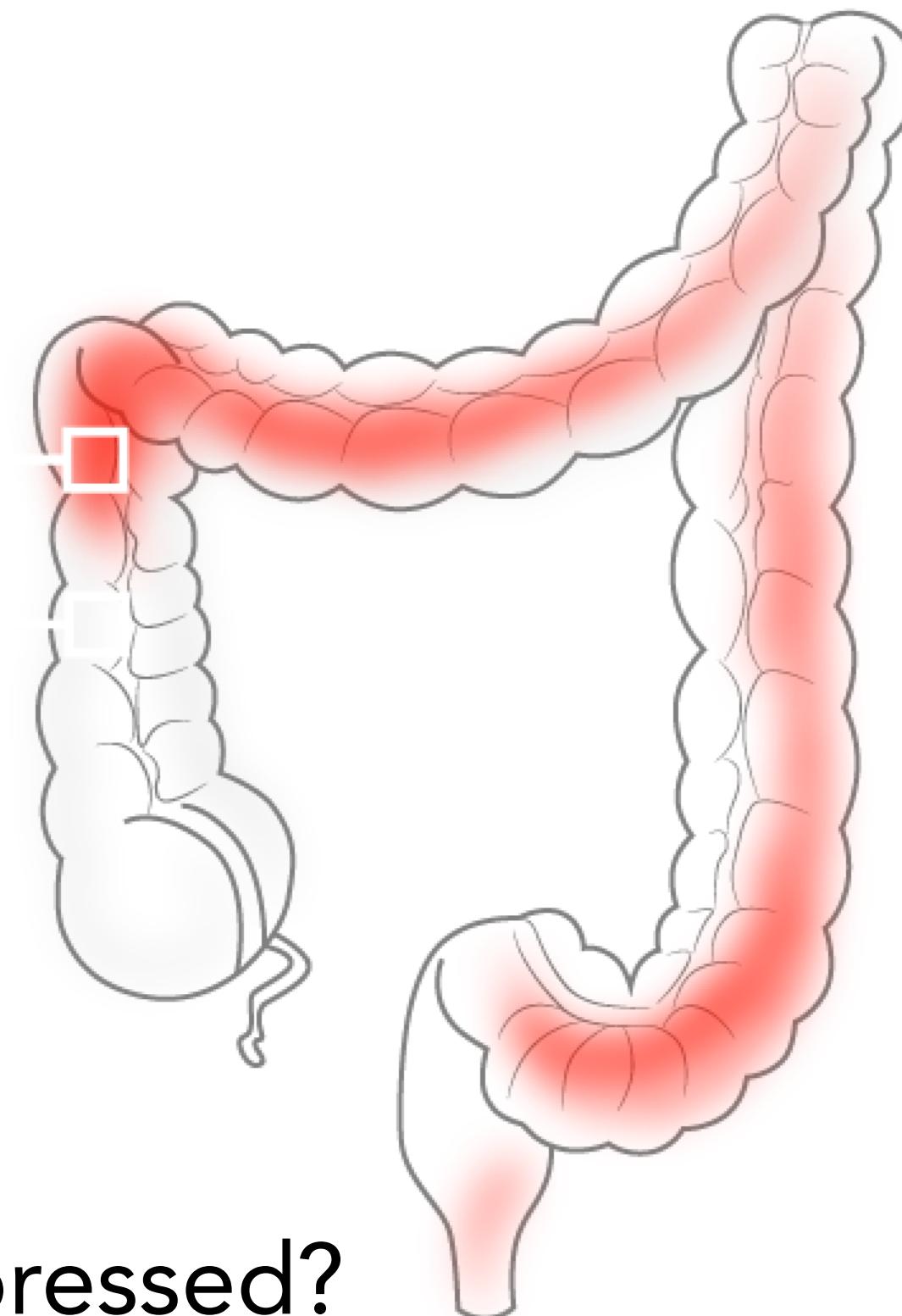


Unique to specific cell types



What can we learn from single cell data?

1) What **cell types** are present?



2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

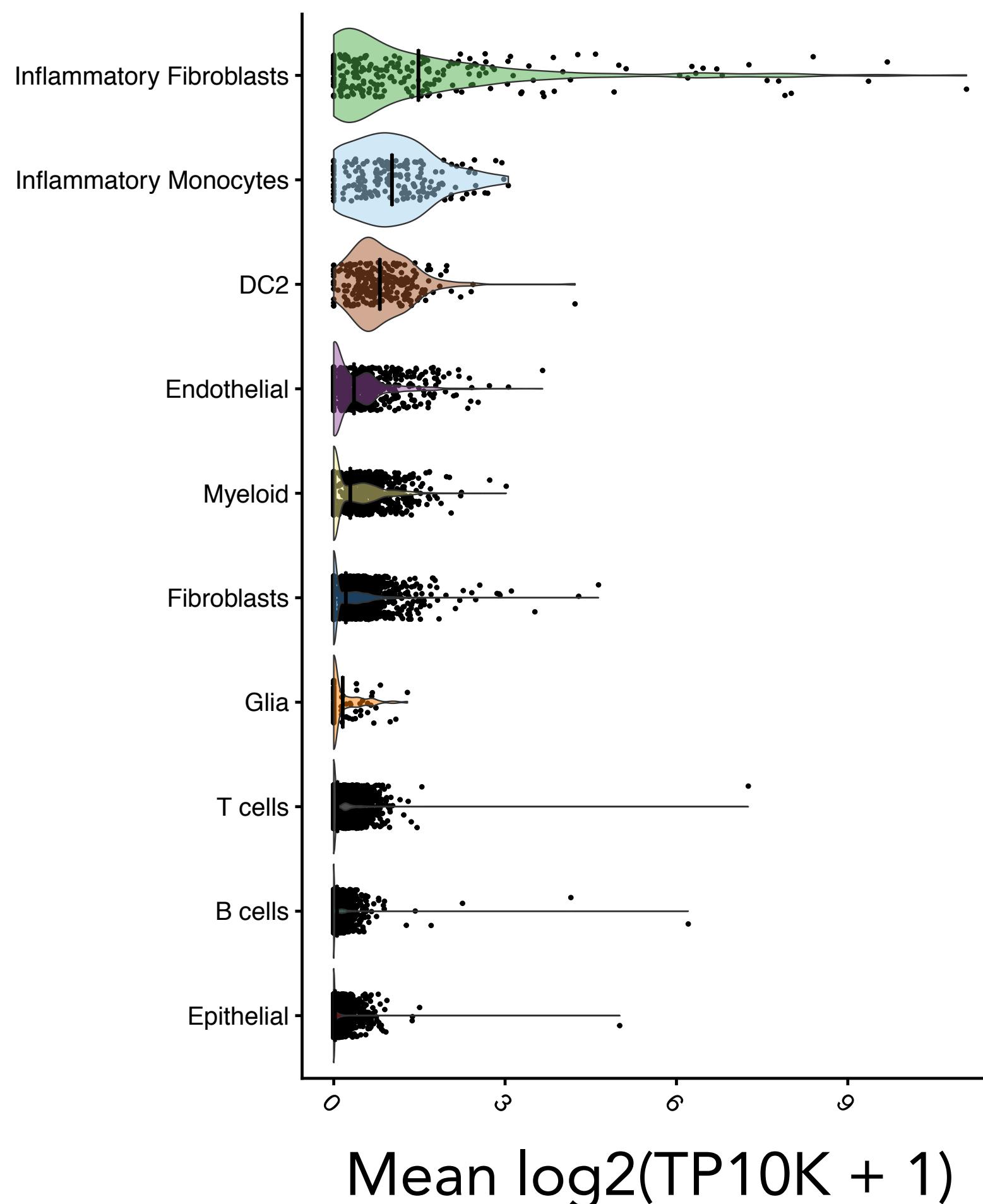
6) Can we use single cell data to understand IBD risk genes?

Mapping published datasets onto single cells

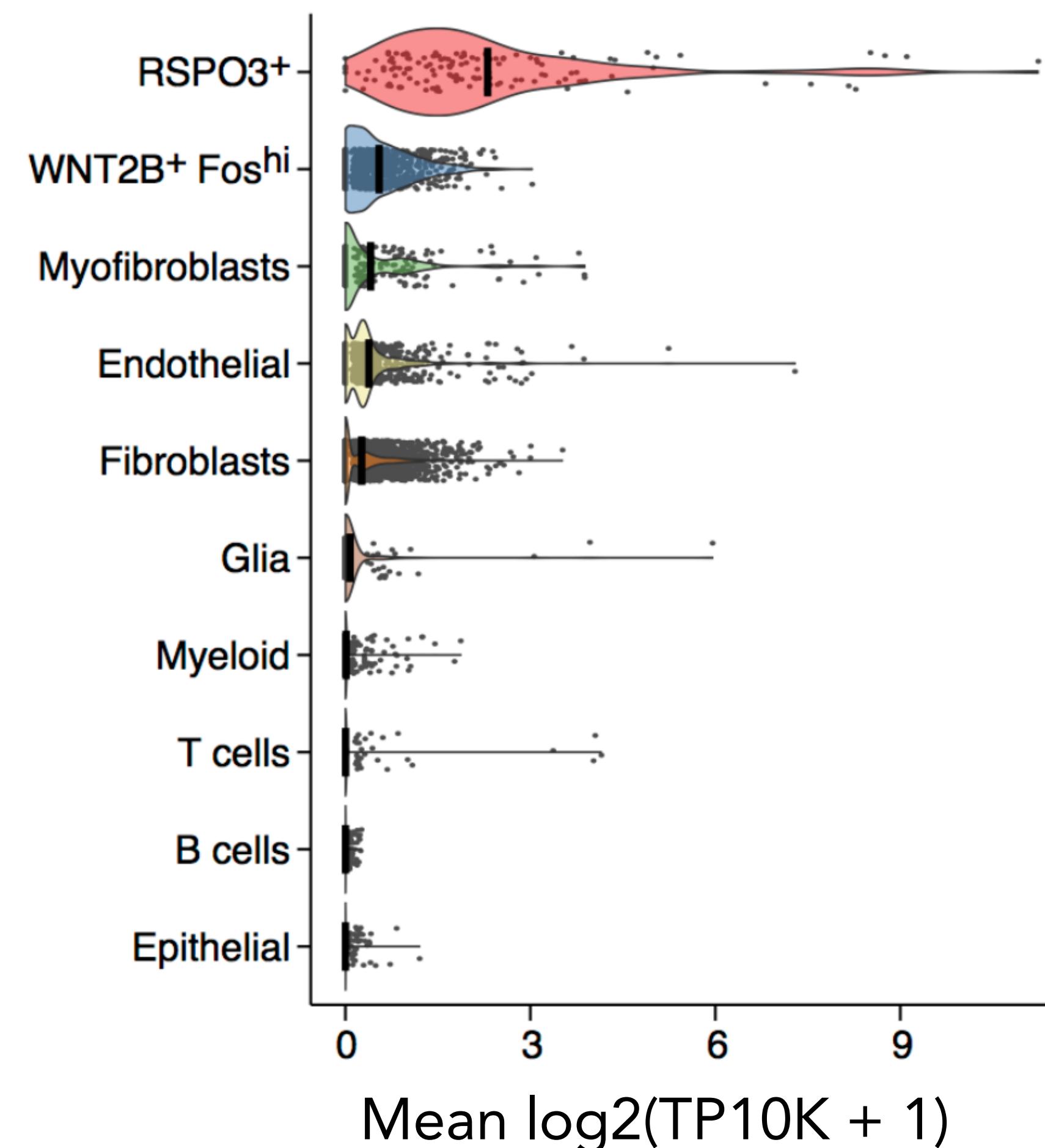
- 1) Find gene signature (Google/Scholar/PubMed are your friends)
 - example: google "IL13RA2 TNFRSF11B IL11"
- 2) For each cell, calculate the mean expression of all genes in the signature
- 3) For each cell type, calculate the mean signature score across cells

Identifying cell types underlying published gene signatures

Inflammatory fibroblasts enriched for
“Anti-TNF resistance” signature

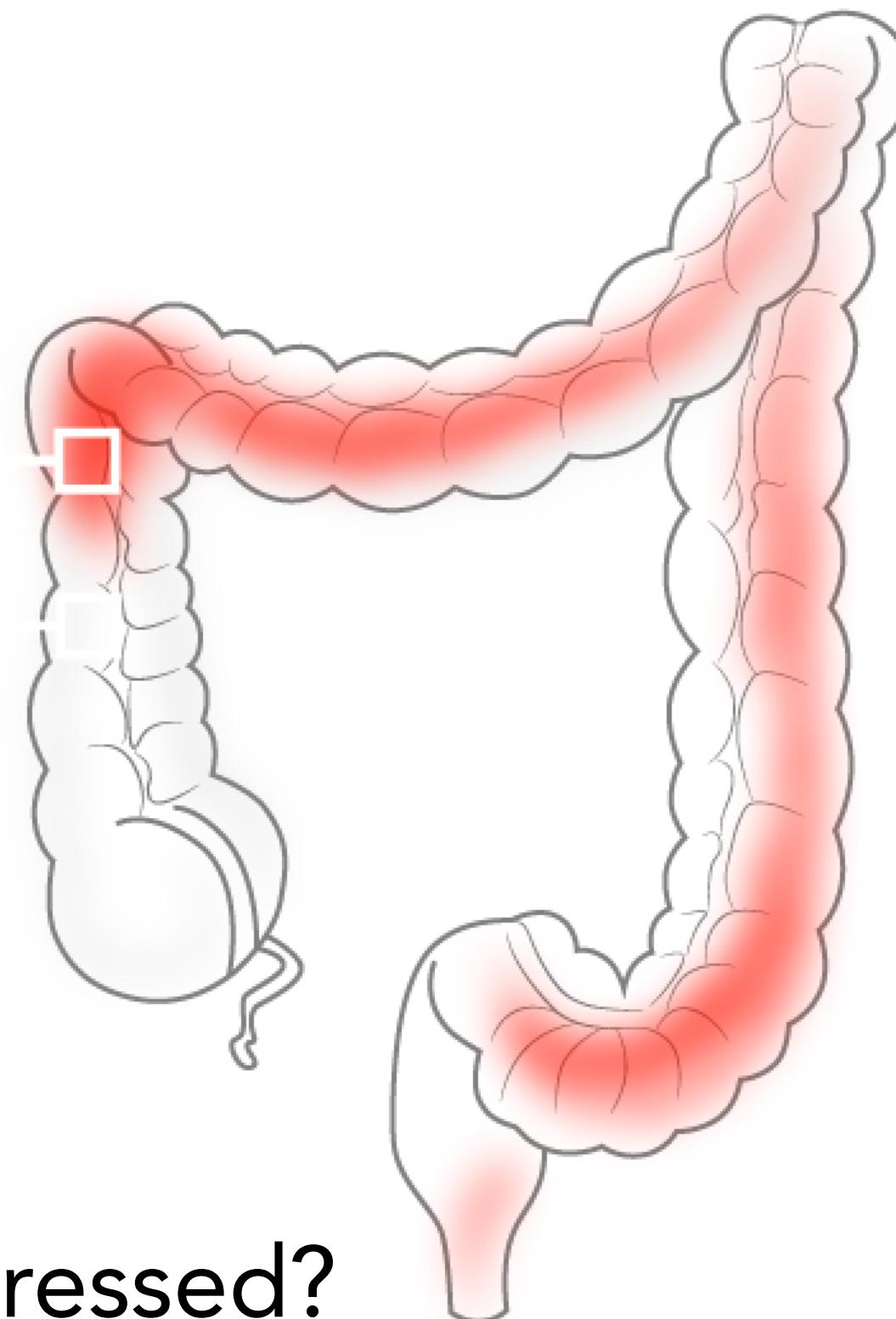


RSPO3+ fibroblasts enriched for
“CRC poor prognosis” signature



What can we learn from single cell data?

1) What **cell types** are present?



2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

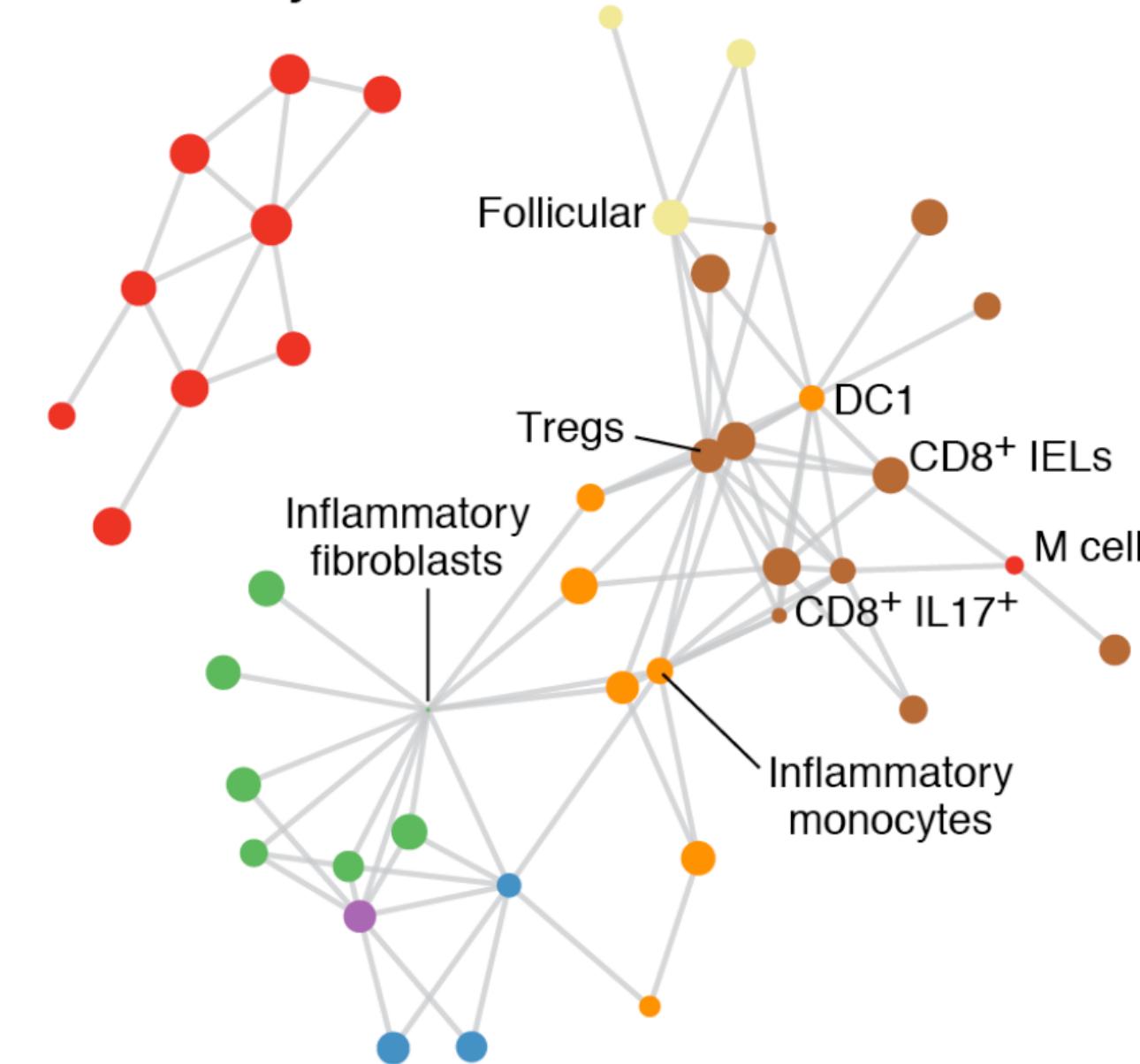
6) Can we use single cell data to understand IBD risk genes?

Inferring cell-cell interactions from single cell data

- 1) Map receptor-ligand pairs onto the cell types
 - example: *LGR5+ intestinal stem cell, RSPO3+ fibroblasts*
- 2) Search for pairs of cell types with more interactions than expected by chance
- 3) Look for changes in this network between health and disease

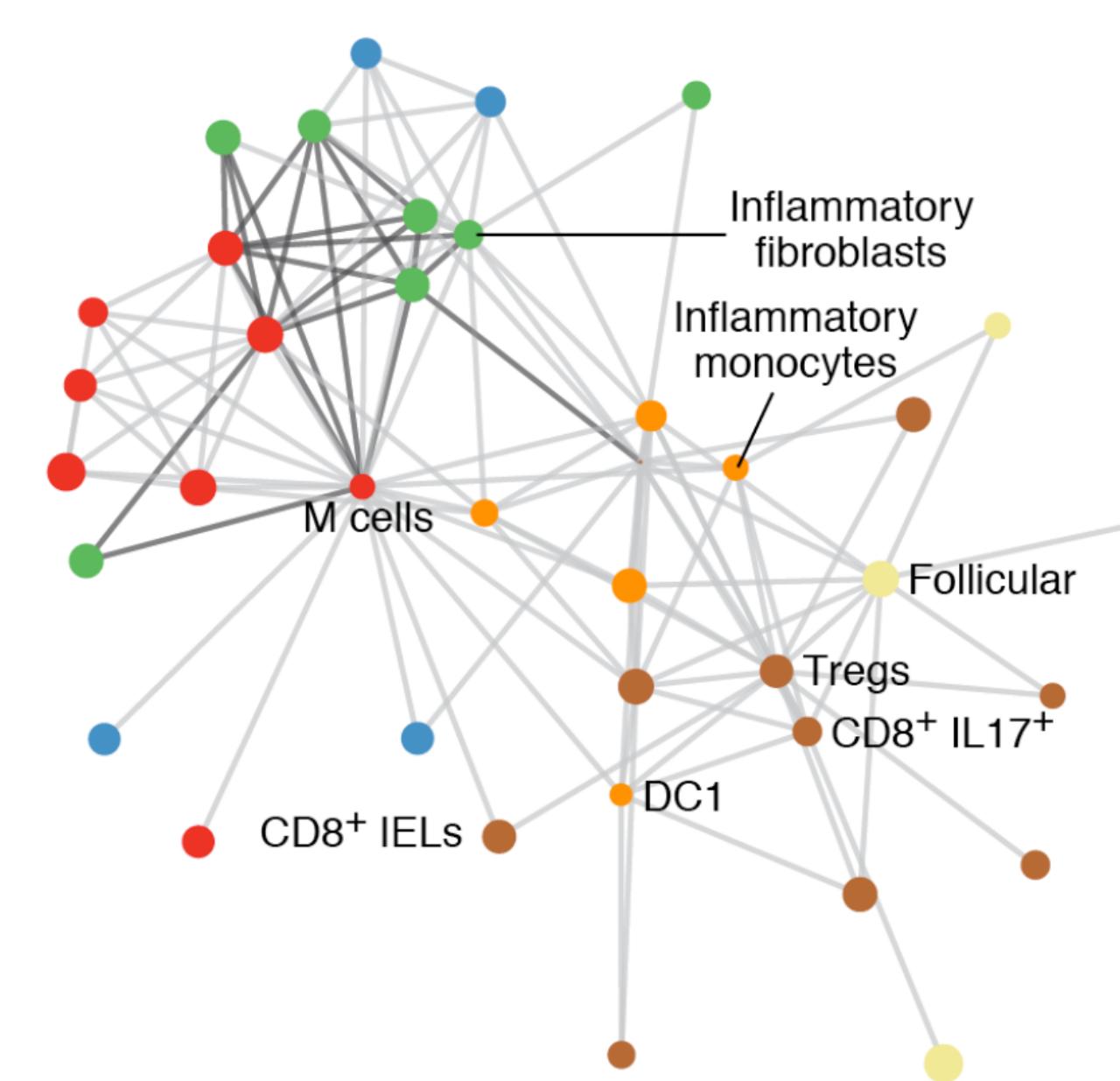
Changes in cell-cell interactions during ulcerative colitis

Healthy



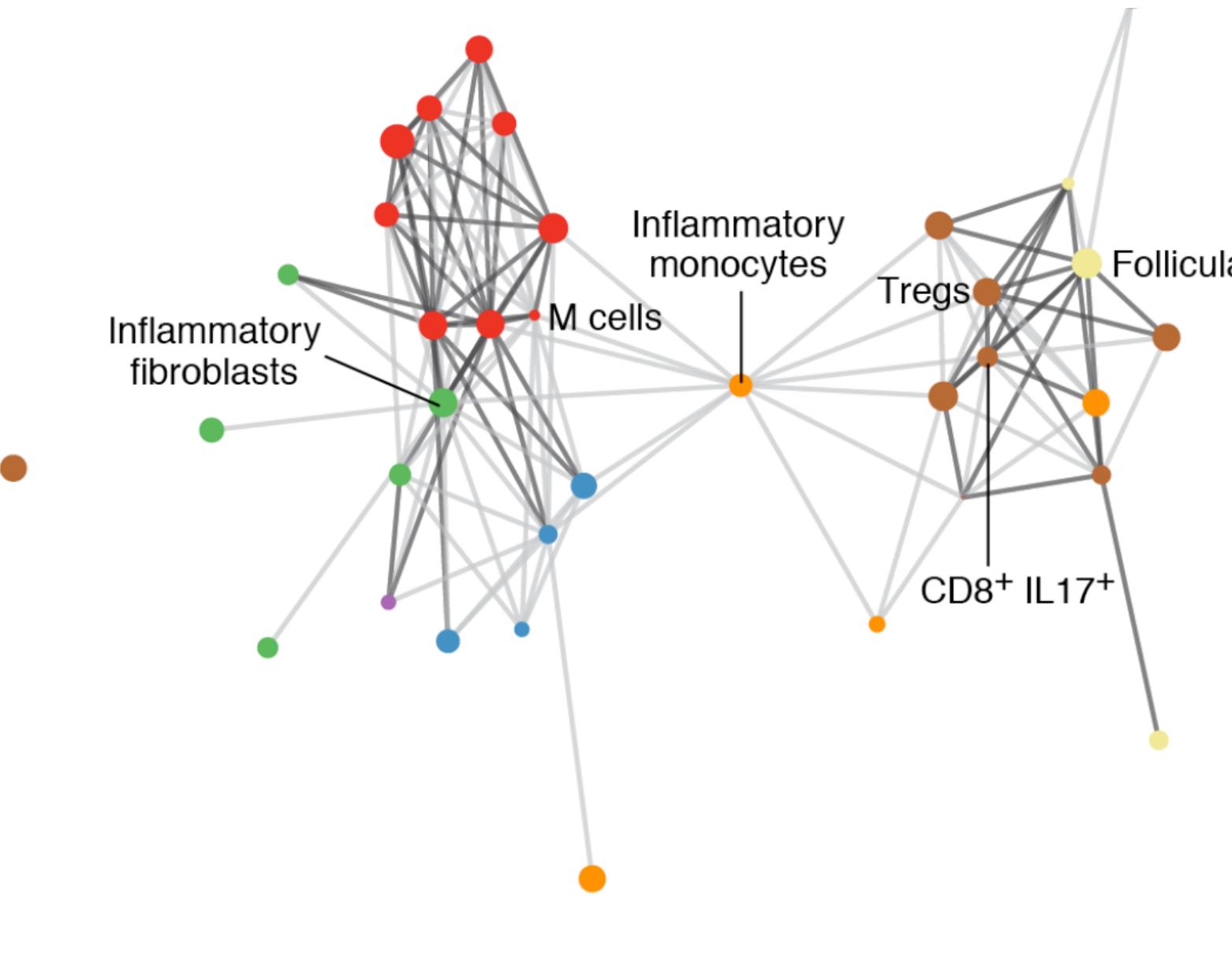
*Most interactions within
cellular compartments*

Non-inflamed



*Disease targets
DC - T cell interactions*

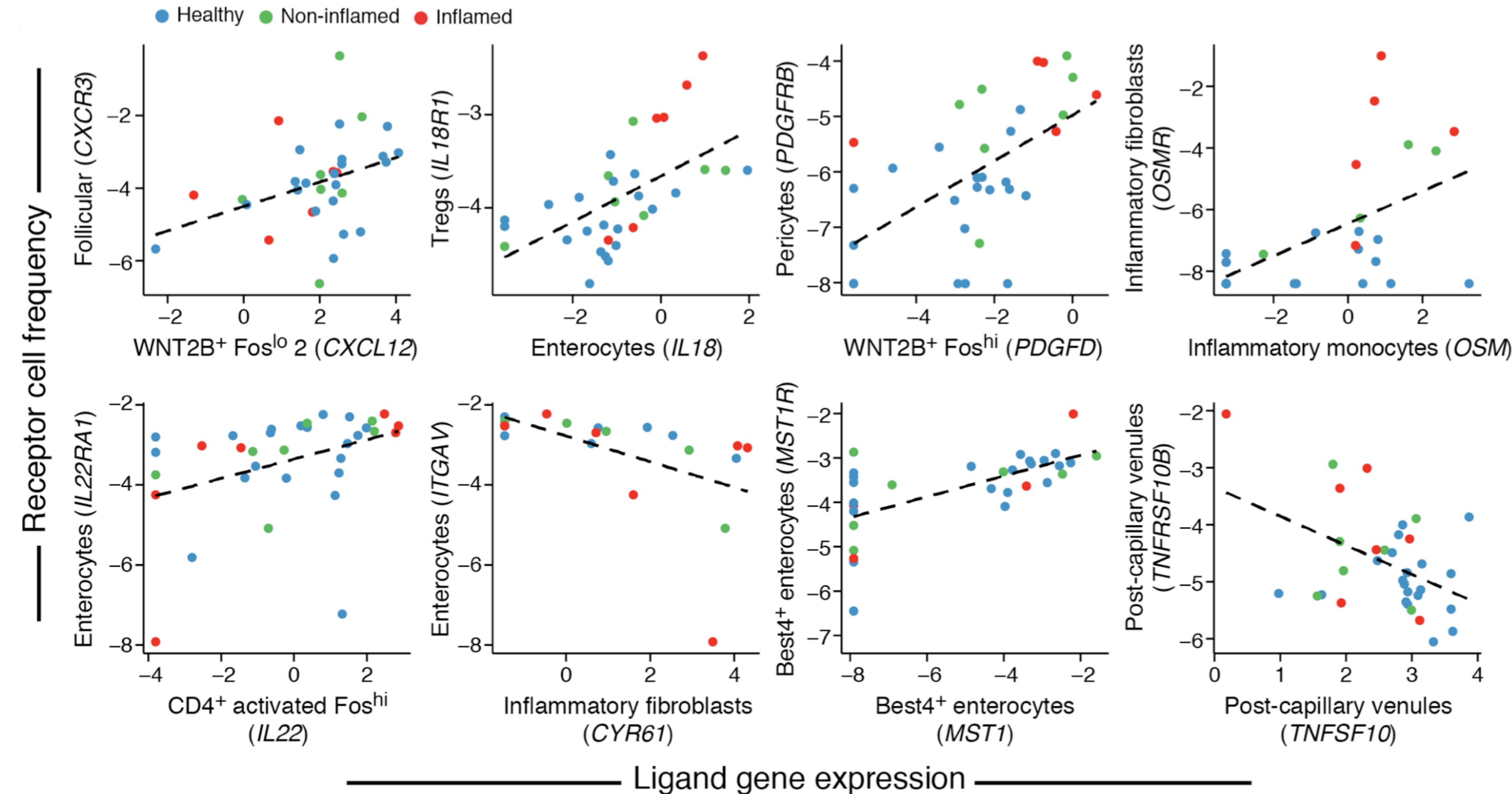
Inflamed



*Disease targets
B cell - T cell interactions*

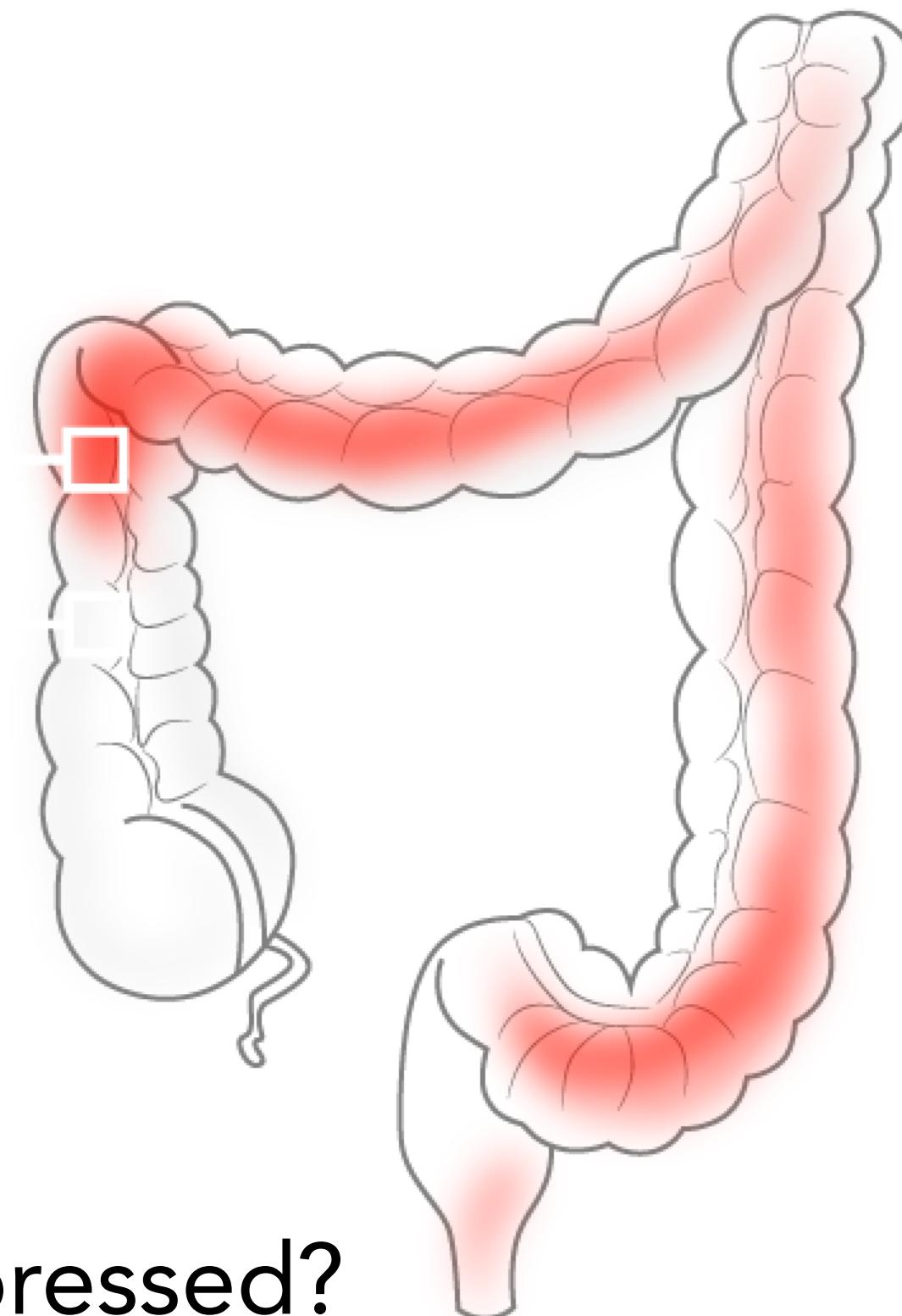
- Epithelial
- Endothelial
- Fibroblasts
- Glia
- Myeloid
- B cells
- T cells

Receptor-ligand interactions can explain changes in cell proportions



What can we learn from single cell data?

1) What **cell types** are present?



2) How do their **proportions** change with disease?

3) What **genes** are differentially expressed?

4) Mapping cell types onto published datasets

5) Which **cell-cell interactions** are perturbed during disease?

6) Can we use single cell data to understand IBD risk genes?

Using single cell data to understand IBD GWAS

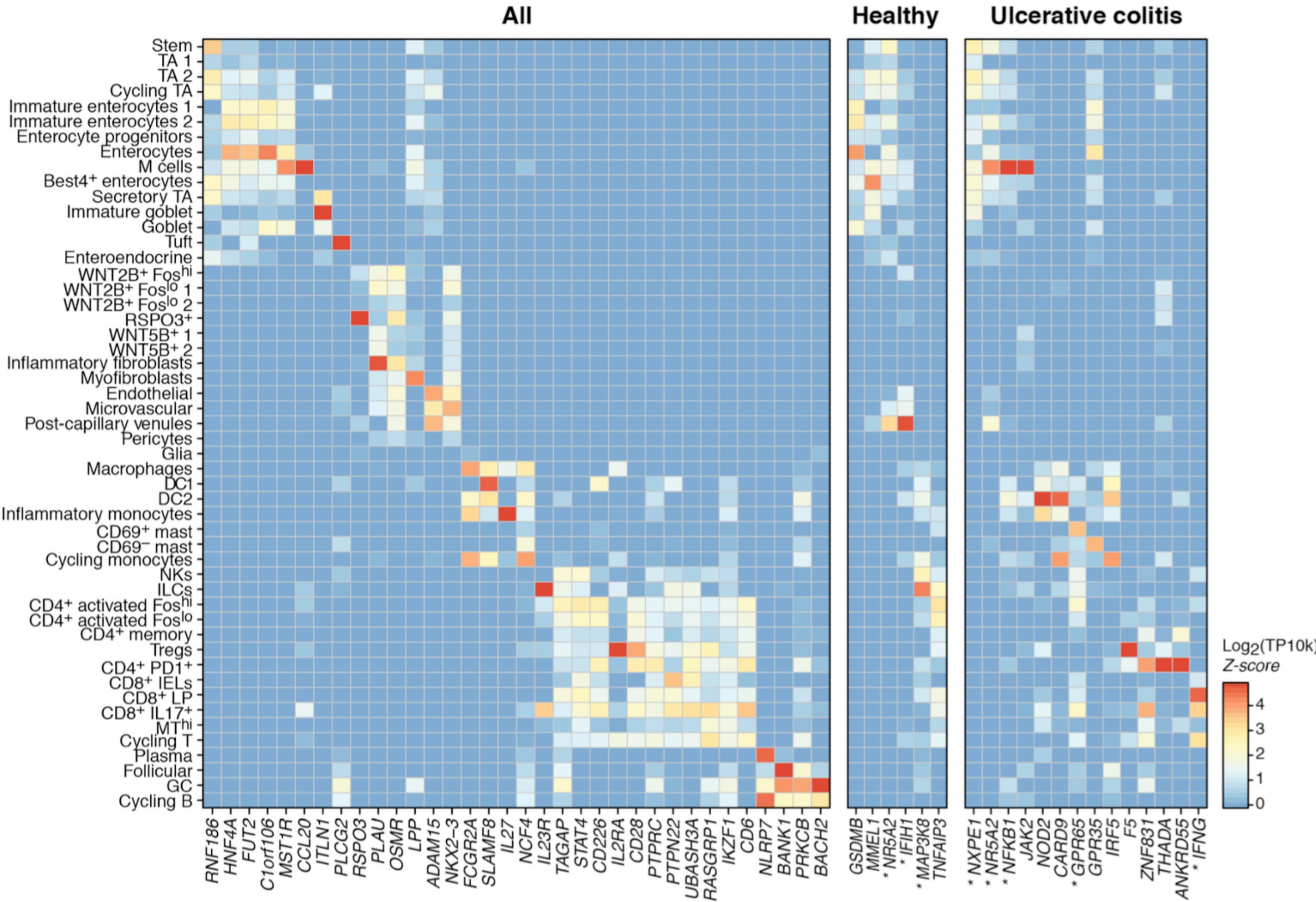
GWAS have uncovered many IBD risk variants, but:

- Their functions often remain unknown
- Many variants have not been mapped to single genes

Single cell data can help us to understand:

- What cell types express IBD risk genes?
- What pathways are they involved in?
- Can we predict “causal” genes across loci?

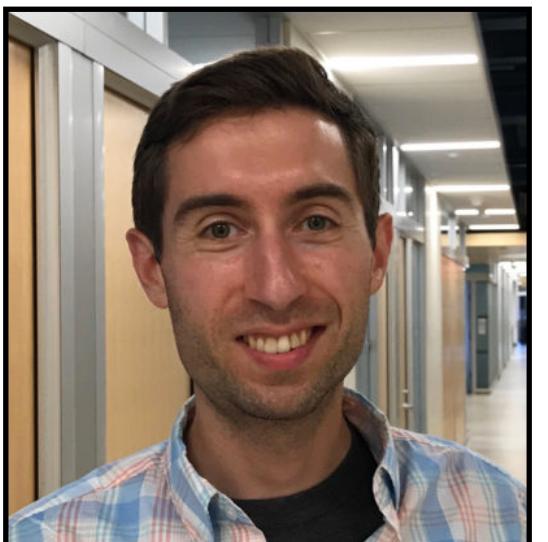
Mapping IBD risk genes onto cell types



Acknowledgments



Moshe Biton
(Regev/Xavier Labs)



Jose Ordovas
(Shalek Lab)

Aviv Regev
Ramnik Xavier
Alex Shalek

Regev Lab

Moshe Biton
Rebecca Herbst
Noga Rogel
Grace Burgin
Danielle Dionne
Orit Rozenblatt-Rosen
Jane Lee

Xavier Lab

Dan Graham

Shalek Lab

Jose Ordovas-Montanes

MGH Clinical Team

Thomas Cleland
Pamela Ribeiro
Keri Sullivan
Christina Wong



Aviv Regev



Ramnik Xavier



Alex Shalek