



---

# Revealing Transcriptomic Insights through Reference-Based Annotation of Skeletal Muscle single-nucleus RNA Sequencing Data

---

21800586 이진우

22000304 박지원

# **Index**

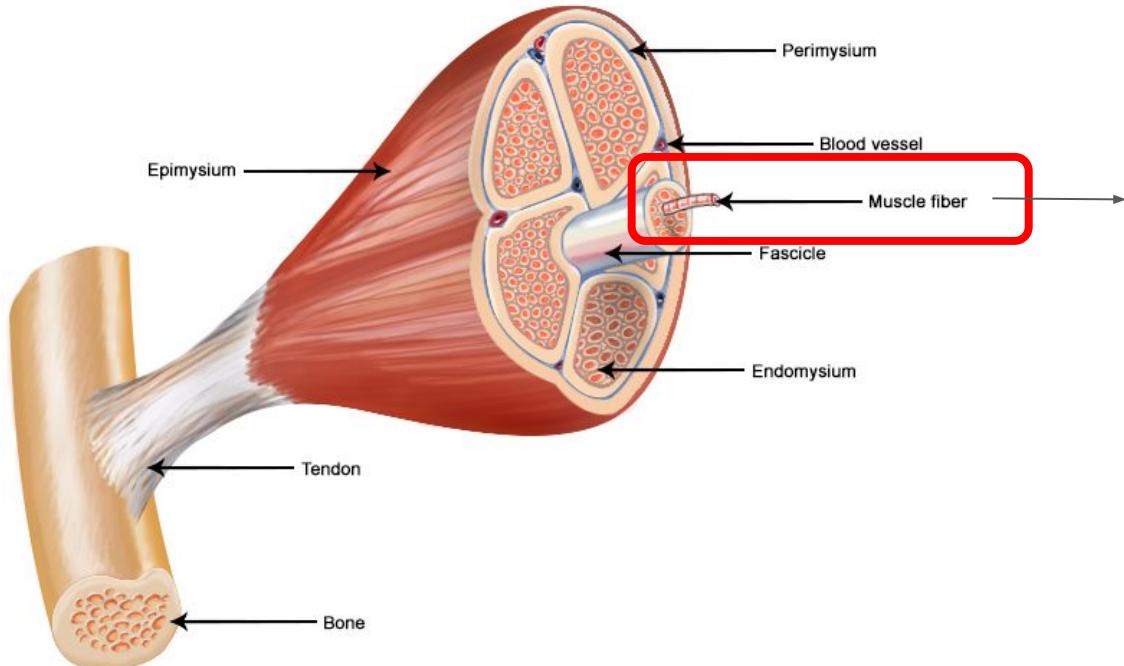
---

- **Introduction**
  - Background and purpose
  - snRNA seq vs scRNA seq
  - Mckellar dataset
  - Reference based annotation
- **Results**
  - Exploratory data analysis
  - Integration our data using harmony
  - Reference based annotation: scPred
  - Reference based annotation: SingleR + ImmGen
  - Functional annotation: clusterProfiler
- **Conclusion**

# Background and Purpose

"We used mouse skeletal muscle samples from d0 and d3 after cardiotoxin injection."

## Structure of a Skeletal Muscle

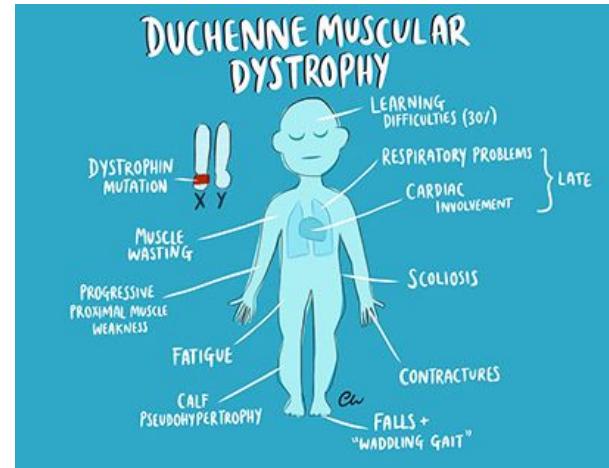
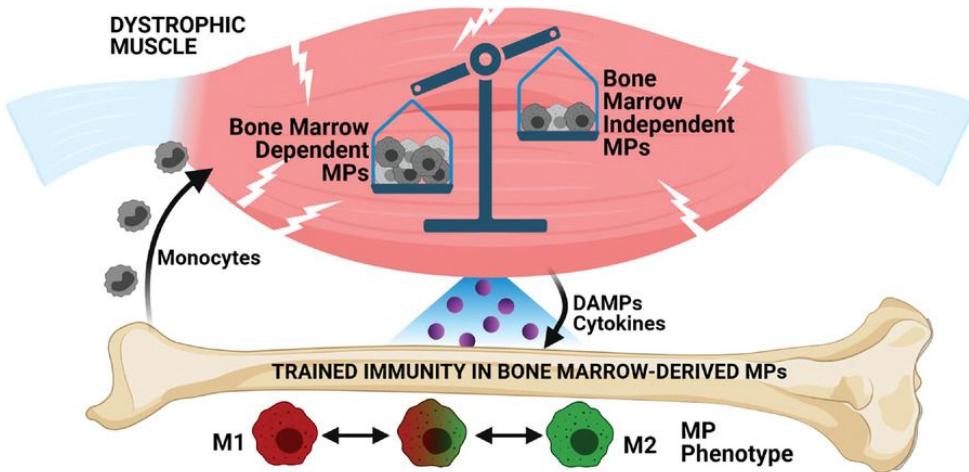


Myofiber is formed of fusion of multiple cells, unlike typical cells, it contains multiple nuclei

# Background and Purpose

- Macrophage

Why no macrophage clearance happening?



The Journal of  
Physiology

# Background and Purpose

- Macrophage

Why no macrophage clearance happening?



<Purpose>

We observe macrophage dynamics in the process of muscle regeneration, and through this, we want to understand the macrophage clearance process.

DUCHENNE MUSCULAR DYSTROPHY

The Journal of  
**Physiology**

# snRNA seq vs scRNA seq

- **scRNA seq**

Technique involving the isolation and sequencing of RNA from individual cells

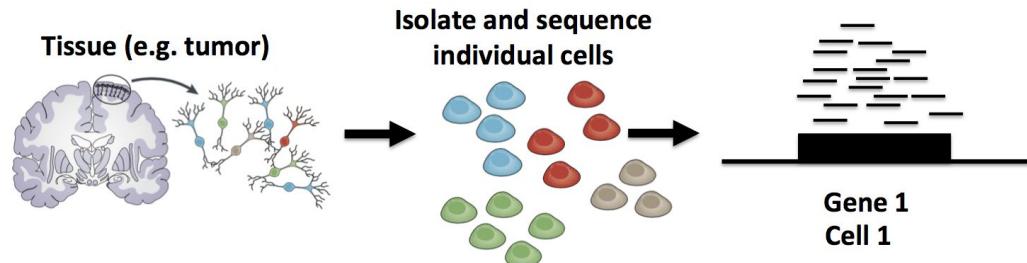
- **snRNA seq**

Technique involving the isolation and sequencing of RNA from intact nuclei of cells



An alternative for tissues where it is difficult to isolate individual cells

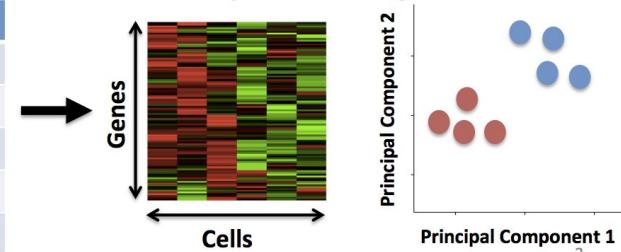
## Single-cell RNA-Seq (scRNA-Seq)



Read Counts

	Cell 1	Cell 2	...
Gene 1	18	0	
Gene 2	1010	506	
Gene 3	0	49	
Gene 4	22	0	
...			

Compare gene expression profiles of single cells



# Mckellar Dataset : Large Scale Reference Data

communications

biology

ARTICLE

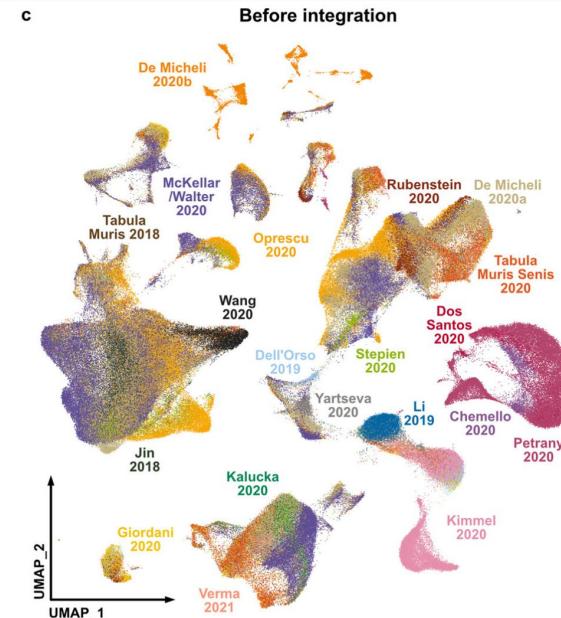
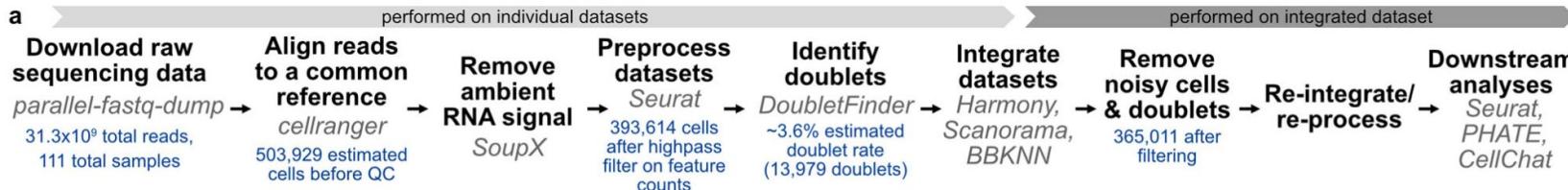


OPEN

## Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration

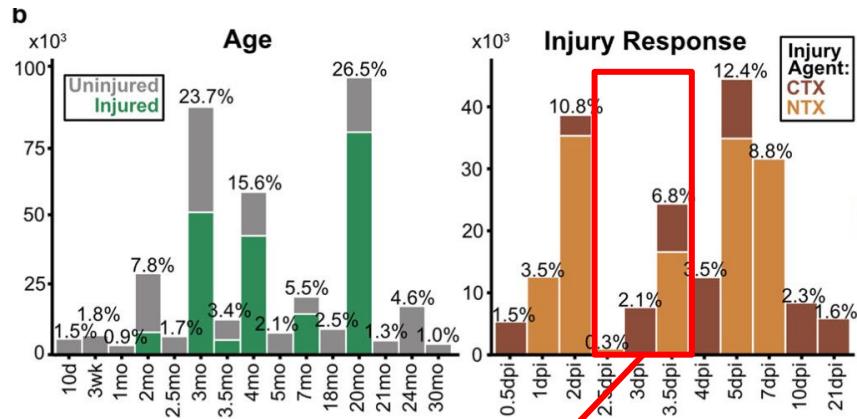
David W. McKellar<sup>1</sup>, Lauren D. Walter<sup>2</sup>, Leo T. Song<sup>1</sup>, Madhav Mantri<sup>1</sup>, Michael F. Z. Wang<sup>3</sup>, Iwijn De Vlaminck<sup>1,4</sup> & Benjamin D. Cosgrove<sup>1,4</sup>

- **Total samples: 111(88 + 23)**
- **365,000 cells after filtering**  
(18 prior studies + 23 new samples of regenerating mouse )

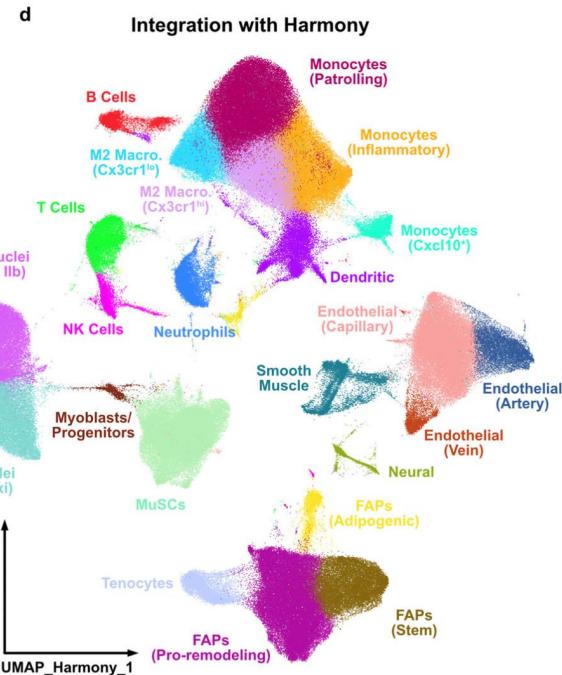


From: [Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration](#)

# Mckellar Dataset : Large Scale Reference Data



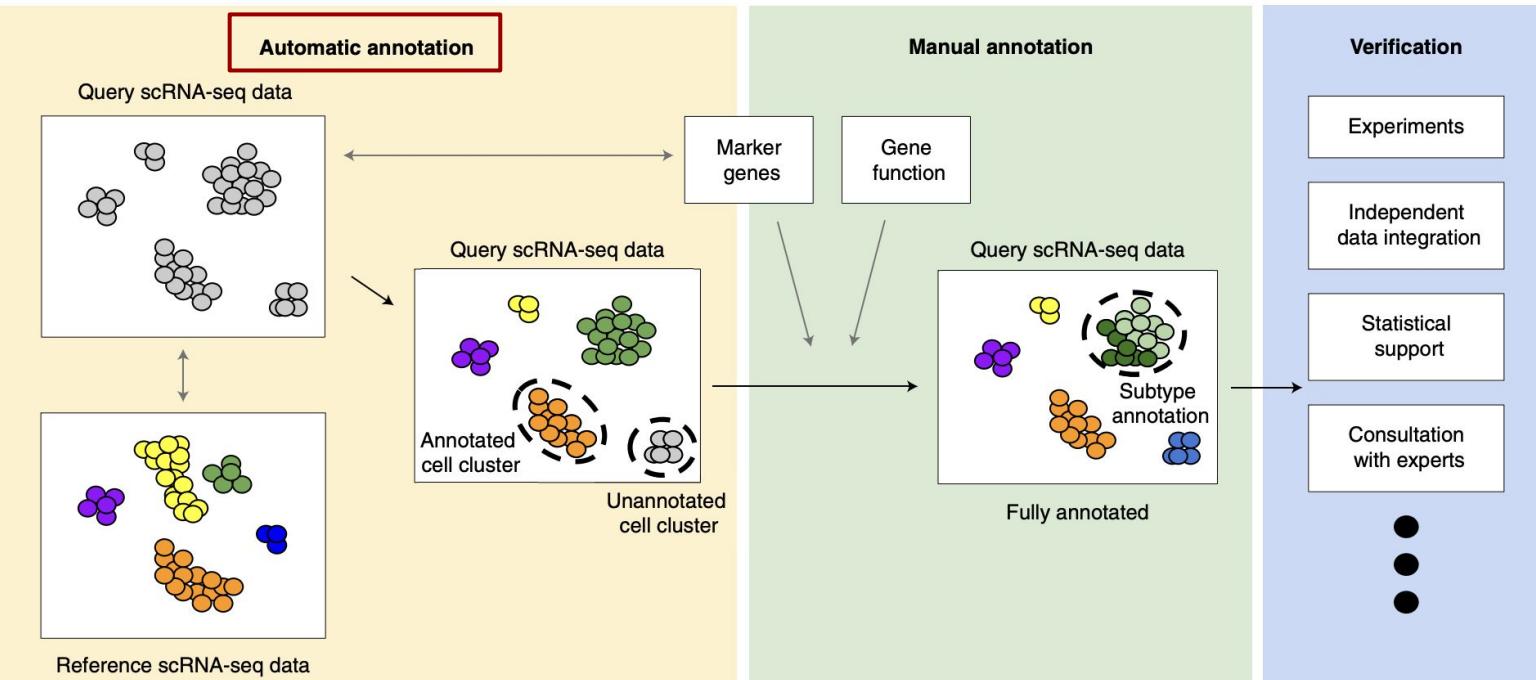
Our reference for 3 Day post injury



<UMAP of integrated data with Harmony >

# Reference based annotation

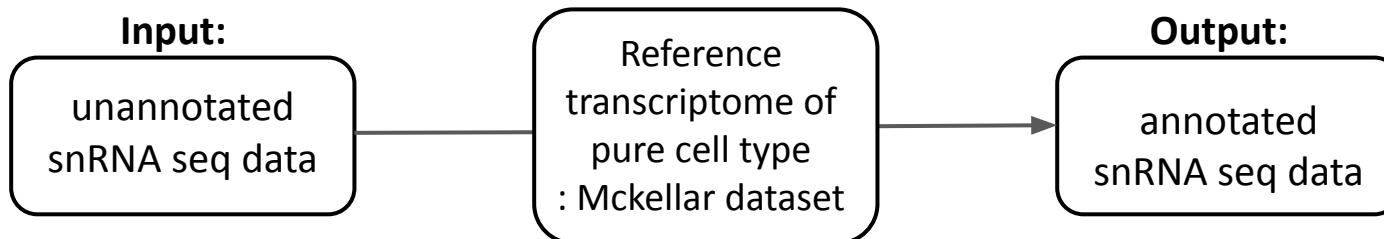
## How to Annotate Clusters or Cells?



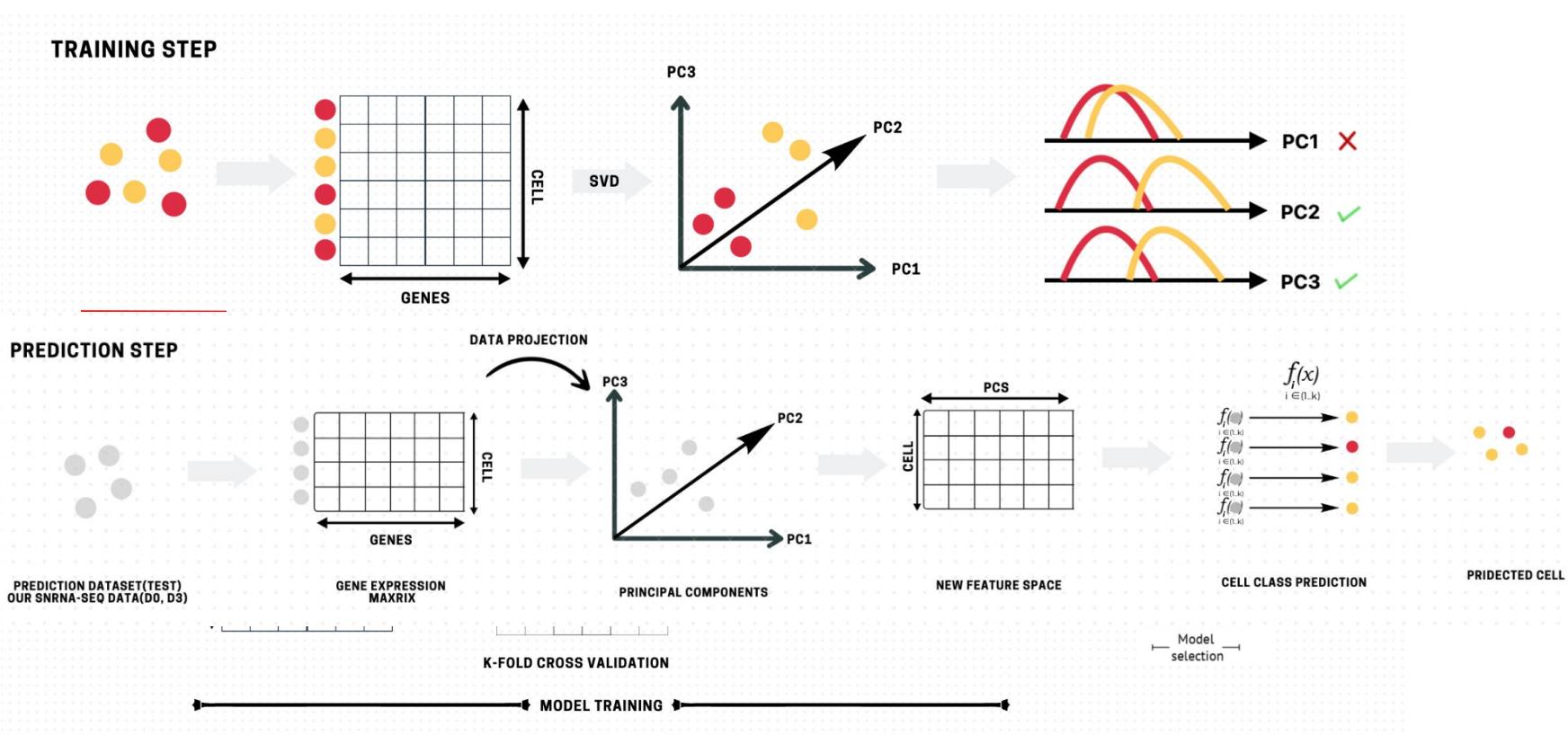
# Reference based annotation

## Why did we use reference based annotation method?

- The annotation process in snRNA seq analysis is **time-consuming** and requires **a lot of biological knowledge** → **Inefficient**
- For this reason, if we use **reference based annotation**, analysis can be easily performed using a reference dataset. → **Save time and effort**, Ensures **accurate annotation results**
- Process to **identify cell types** by using **known gene expression** patterns of **reference datasets**(Mckellar dataset)
- We performed automated cell type annotation using **scPred** and **SingleR**.



# Reference based annotation: scPred



# Reference based annotation: SingleR + Immgen Dataset

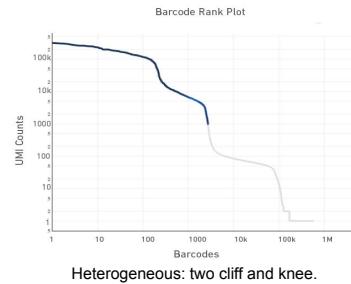
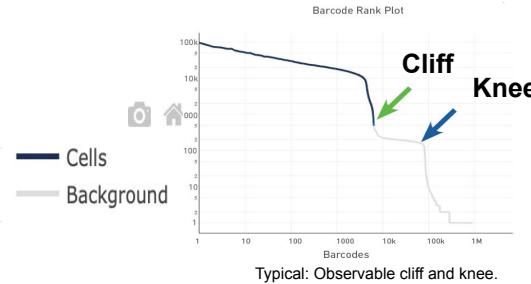
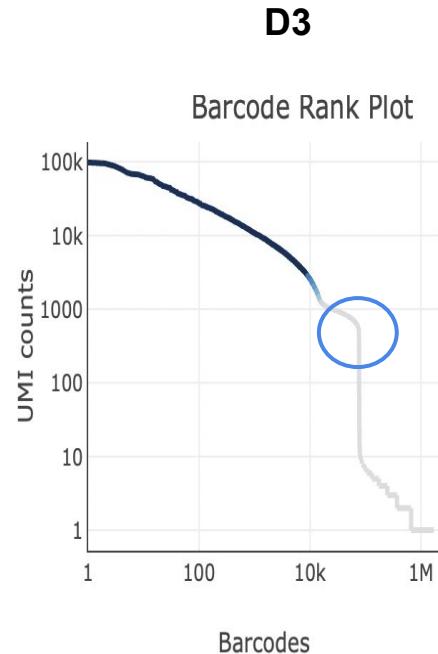
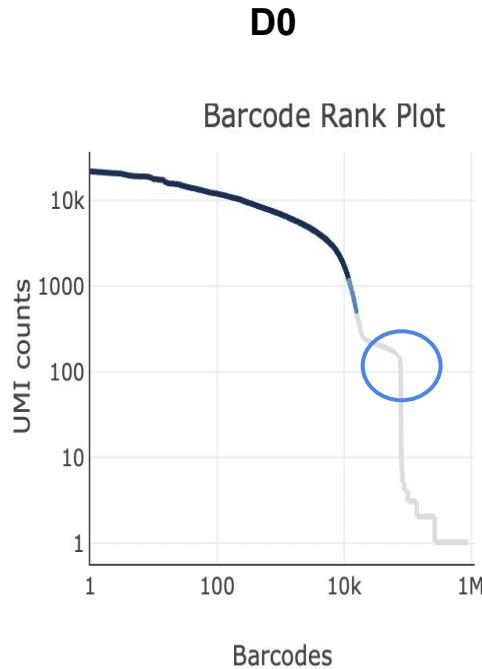
Name	GEO Accession	Description	Cell Types
ImmGen ULI: Deep RNA-seq data	GSE122597	Five highly purified immunocyte populations profiled to unusual depth as multiple replicates (8 to 16). Suitable for exploration of genes expressed at very low levels.	NK Cells, Follicular B, Naive CD4+ abT, gdT cells and peritoneal macrophages.
ImmGen ULI: Systemwide RNA-seq profiles (#1)	GSE109125	83 populations representing all lineages and several differentiation cascades prepared from unchallenged mice and after LPS, anti-CD3, viral infection cell activation.	B Cells, Stromal Cells, Dendritic Cells, Granulocytes, Innate Lymphocytes, Stem Cells, Macrophages, ab T Cells, gd T Cells
ImmGen ULI: OpenSource Mononuclear Phagocytes Project	GSE122108	130 populations comprising progenitors, residents, and stimulated (C.alb, LPS, injury, APAP+ starved overnight and pIC) mononuclear phagocytes for OpenSource MNP Project.	Macrophages, Kupffer Cell/Macrophages, Dendritic Cells, Microglia, Monocytes.
ImmGen ULI: Male-Female Immune Differences	GSE124829	11 diverse immunocyte populations from male and female mice of varying ages stimulated with different dose of IFN to understand the immune system's sexual differences.	B Cells, Dendritic Cells, Neutrophils, Macrophages, Natural Killer T Cells, ab T Cells, gd T Cells, Microglia, Regulatory T Cells.

- **ImmGen Dataset:**  
Large reference datasets specific to immune system cells
- **SingleR:**  
A computational method to predict cell types using reference scRNA-seq data

# **Result**

# Exploratory Data Analysis

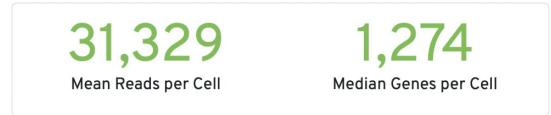
- CellRanger: Aligning D0 & D3 snRNA seq data



Alignment → QC / Filtering → Normalization → Batch-effect correction  
→ Integration → Clustering → Annotation → Downstream analysis

# Exploratory Data Analysis

- CellRanger: Aligning D0 & D3 snRNA seq data



**Sequencing** ?

Number of Reads	421,218,107
Number of Short Reads Skipped	0
Valid Barcodes	97.6%
Valid UMs	99.9%
Sequencing Saturation	81.4%
Q30 Bases in Barcode	97.6%
Q30 Bases in RNA Read	93.6%
Q30 Bases in UMI	97.6%



**Sequencing** ?

Number of Reads	385,327,095
Number of Short Reads Skipped	0
Valid Barcodes	97.3%
Valid UMs	99.9%
Sequencing Saturation	52.1%
Q30 Bases in Barcode	97.1%
Q30 Bases in RNA Read	92.4%
Q30 Bases in UMI	97.0%

**D0**

**Mapping** ?

Reads Mapped to Genome	94.8%
Reads Mapped Confidently to Genome	92.9%
Reads Mapped Confidently to Intergenic Regions	5.0%
Reads Mapped Confidently to Intronic Regions	58.1%
Reads Mapped Confidently to Exonic Regions	29.8%
Reads Mapped Confidently to Transcriptome	77.3%
Reads Mapped Antisense to Gene	10.4%

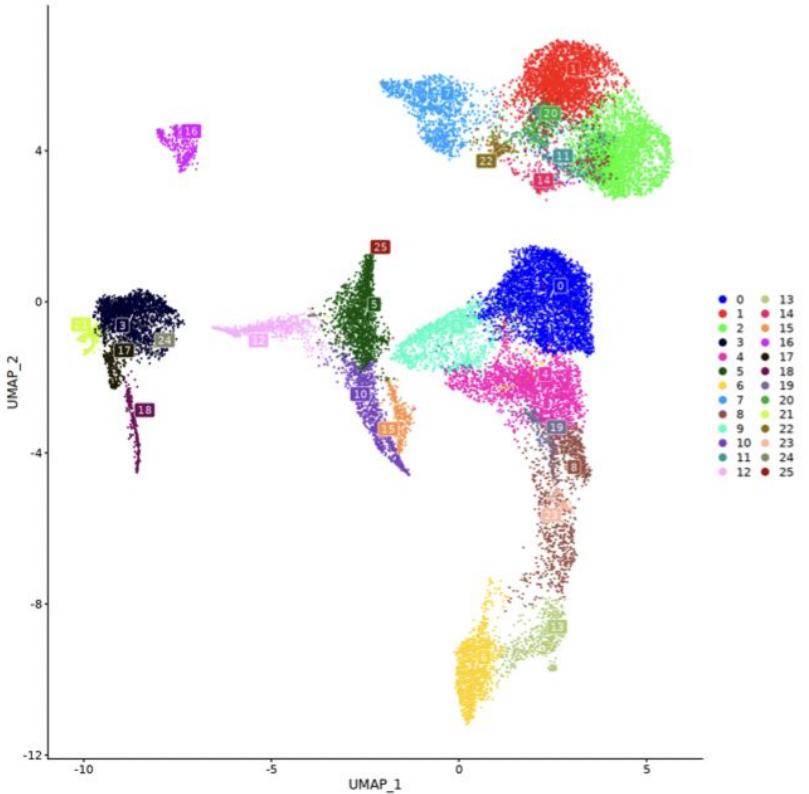
**D3**

**Mapping** ?

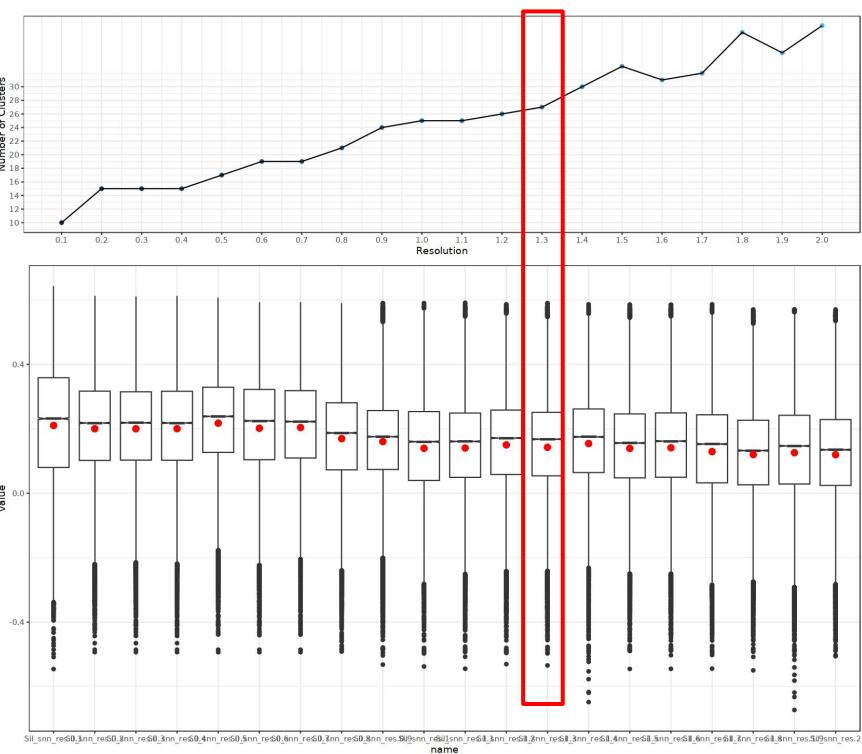
Reads Mapped to Genome	95.0%
Reads Mapped Confidently to Genome	93.1%
Reads Mapped Confidently to Intergenic Regions	4.1%
Reads Mapped Confidently to Intronic Regions	65.8%
Reads Mapped Confidently to Exonic Regions	23.2%
Reads Mapped Confidently to Transcriptome	69.4%
Reads Mapped Antisense to Gene	19.1%

# Integrating Our Data Using Seurat (Harmony)

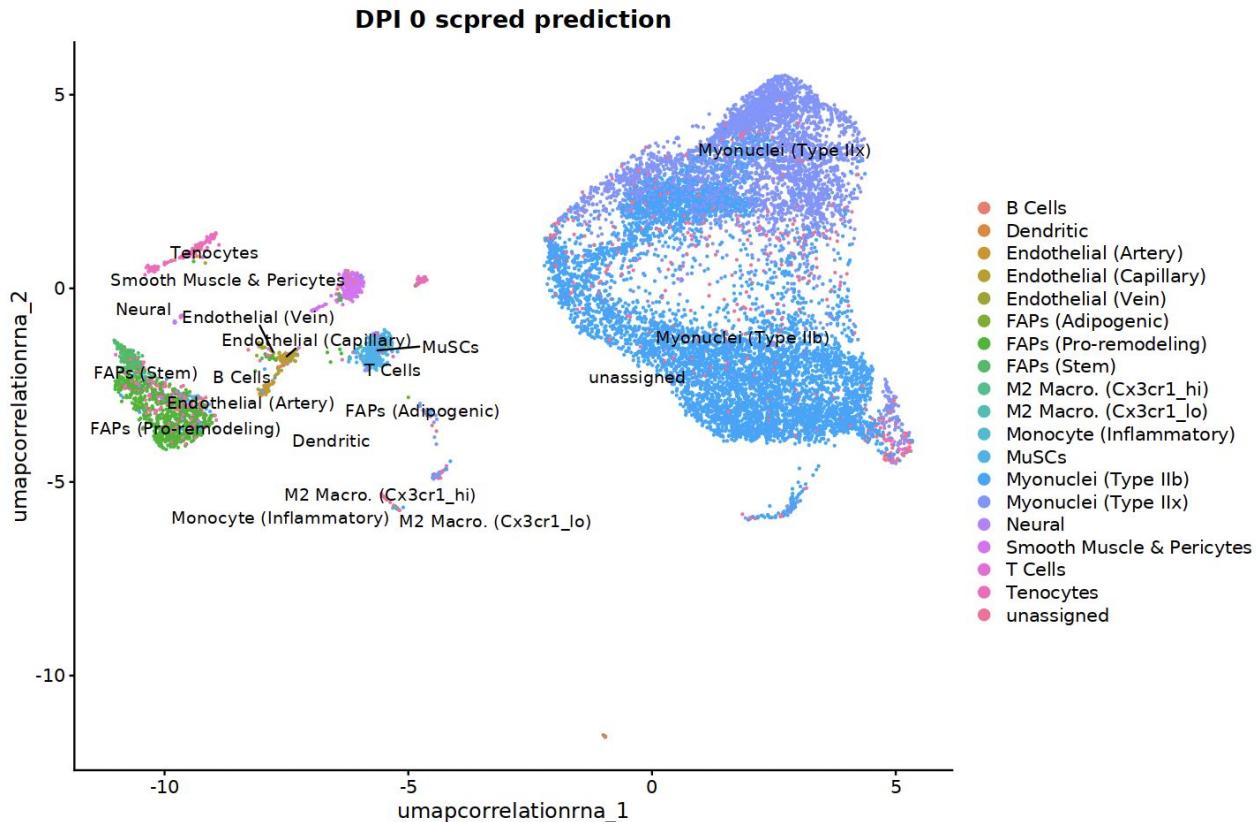
## - PCA after integration -



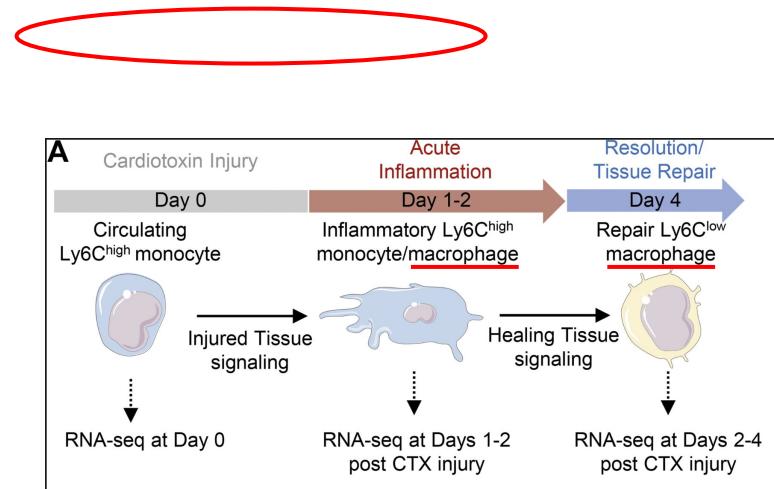
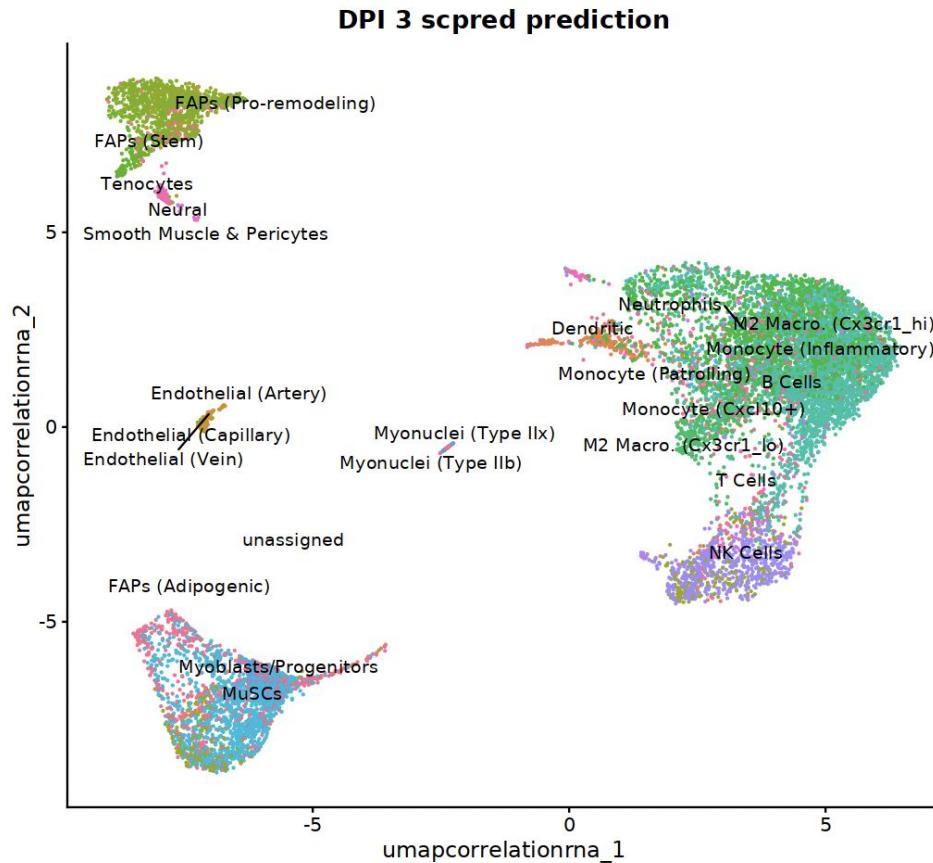
### - Select resolution using silhouette -



# Cell classification / Annotation - scPred



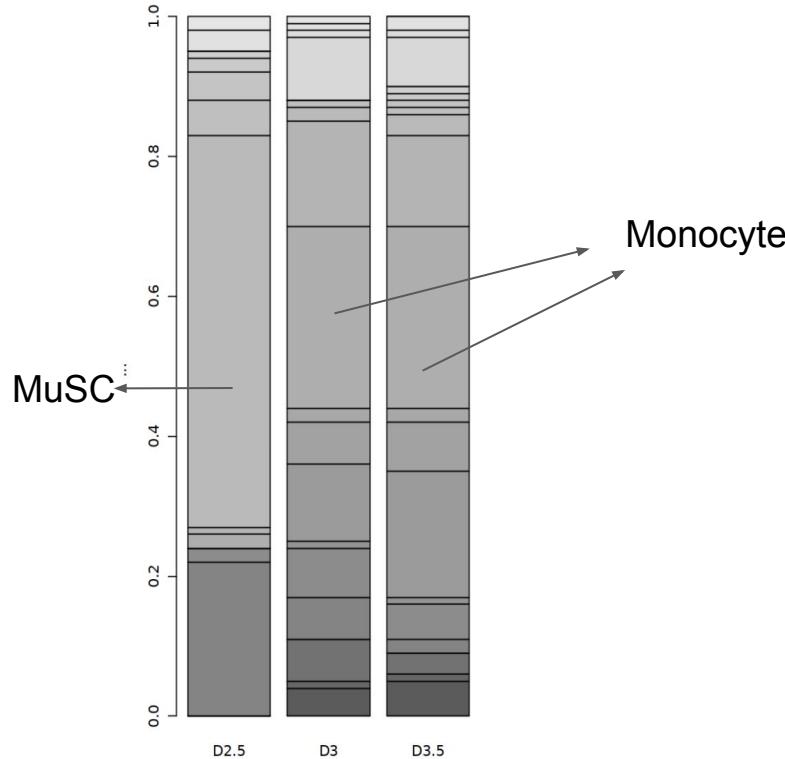
# Cell classification / Annotation - scPred



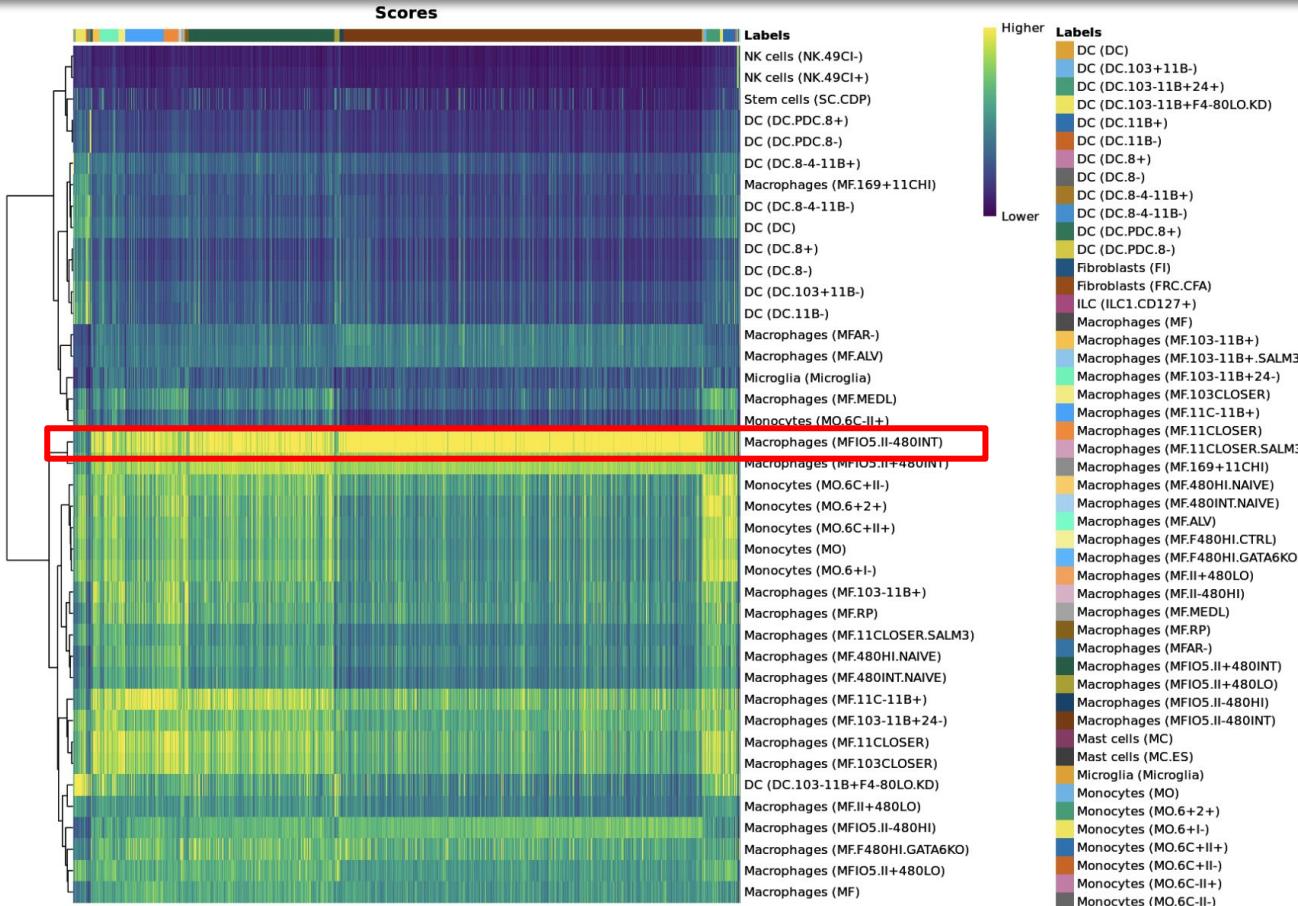
# Cell classification / Annotation - scPred

## Mckellar dataset cell type composition

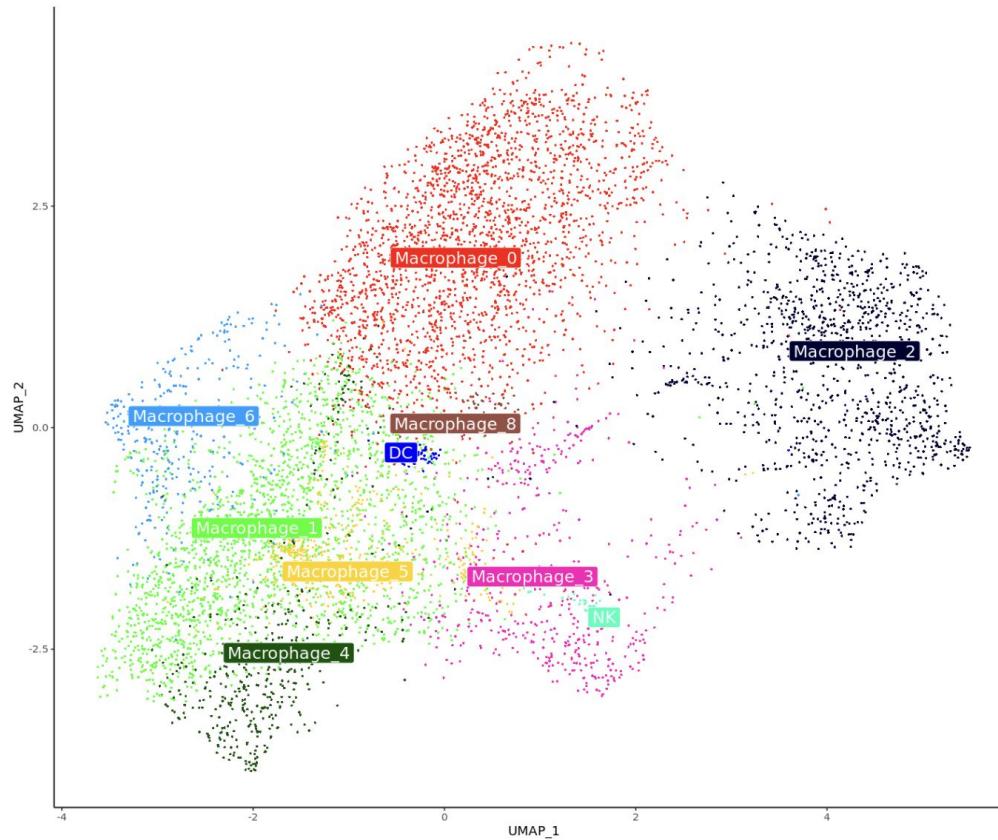
Cell Composition by Day Post Injury 2.5~3.5



# Immune cell Subclustering / Annotation - SingleR +ImmGen

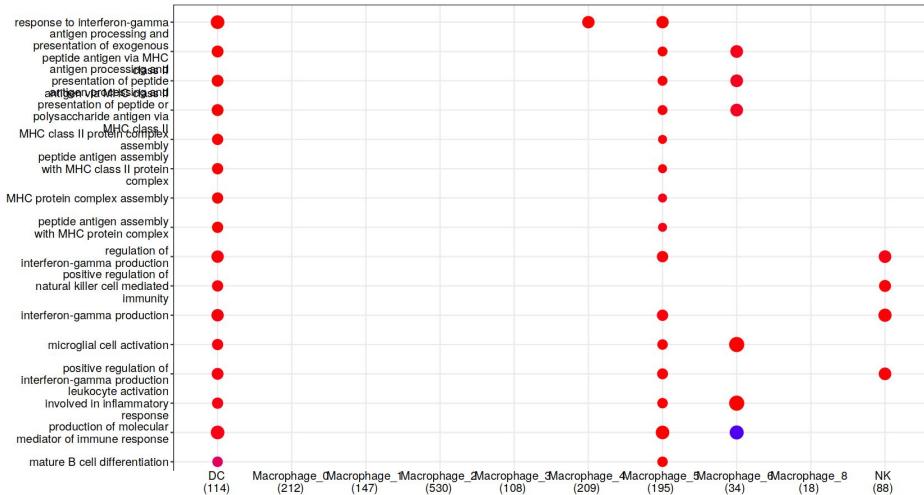


# Immune cell Subclustering / Annotation - SingleR +ImmGen



# Immune cell Functional Annotation - ClusterProfiler

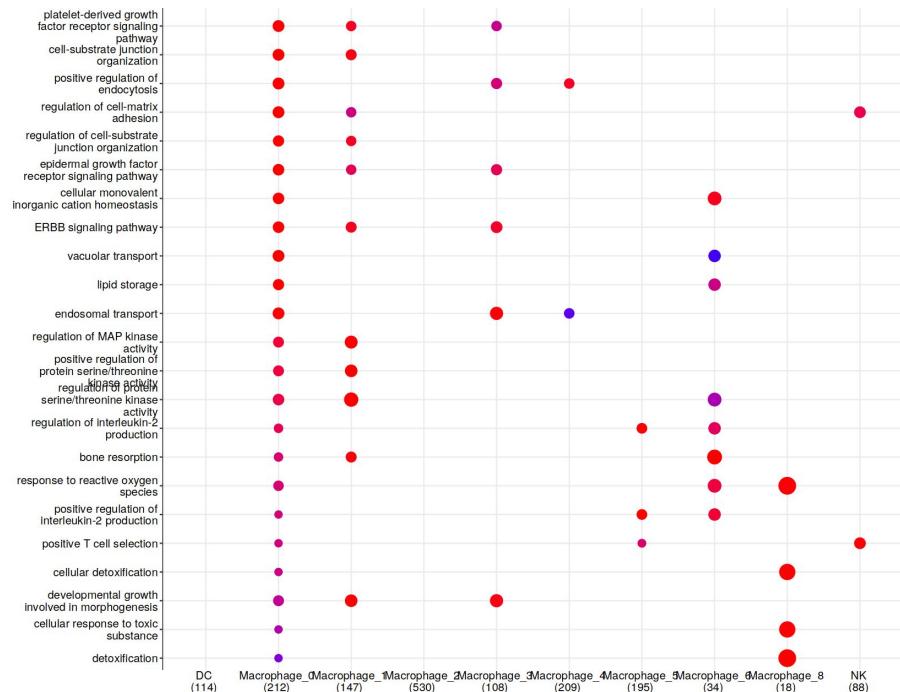
- Visualization of functional profile comparison



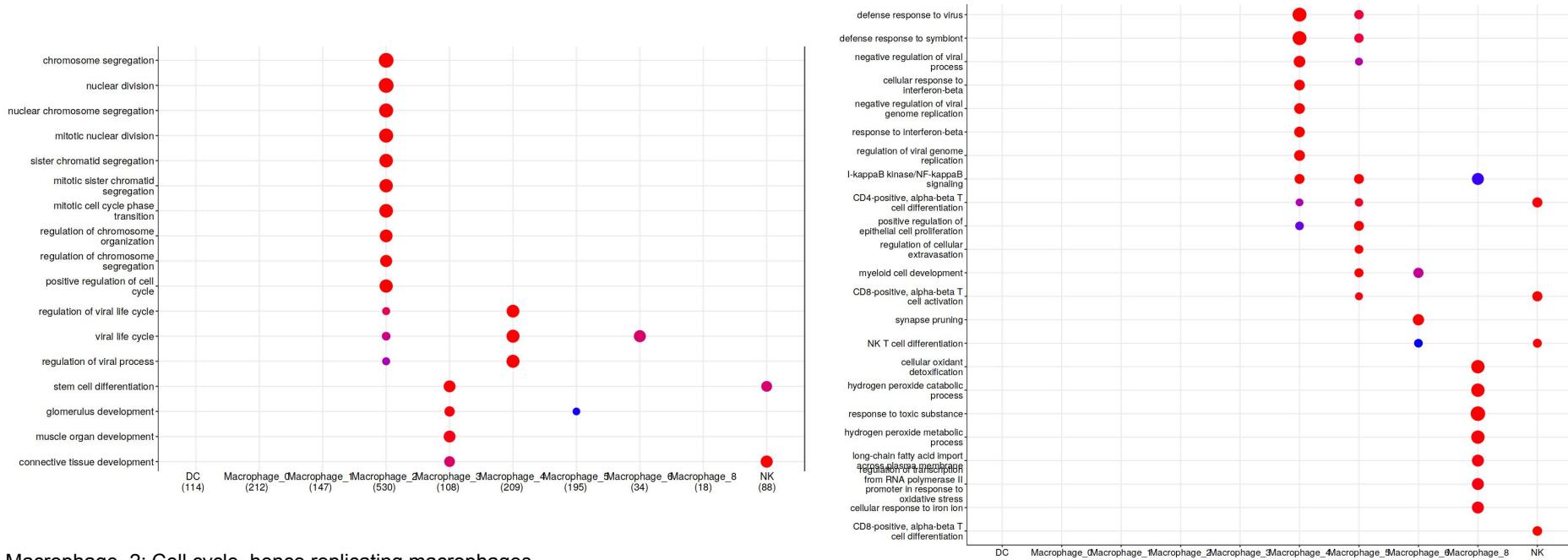
## Macrophage 5: MHCII terms.

Macrophage 0, 2, 8: ROS, detoxification and bone resorption terms,

hence some macrophages were similar with osteoclasts.



# Immune cell Functional Annotation - ClusterProfiler



Macrophage\_2: Cell cycle, hence replicating macrophages.

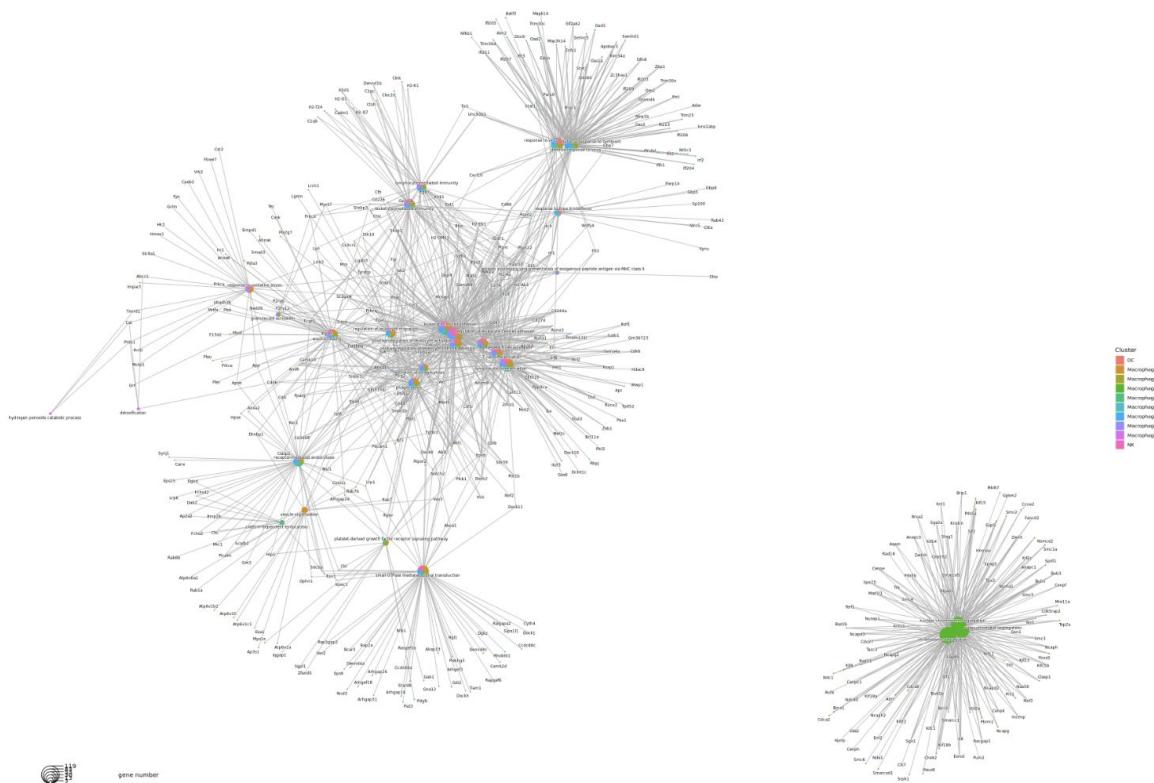
Macrophage\_3: Cell differentiation and development.

Macrophage\_4: IFN-beta, NF-kb

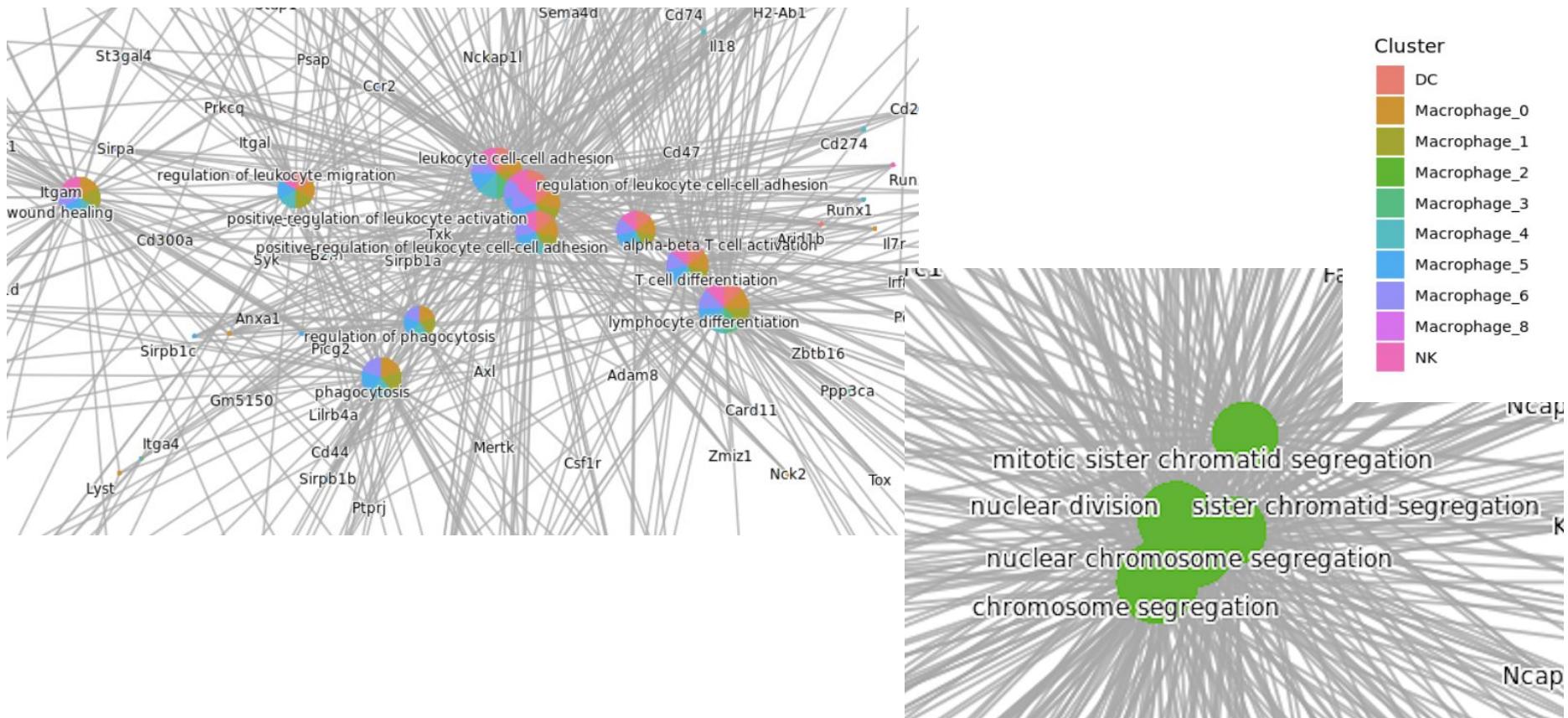
Macrophage\_8: Detoxification

# Immune cell Functional Annotation - ClusterProfiler

## • Gene-Concept Network



# Immune cell Functional Annotation - ClusterProfiler



# Conclusion

---

1. We **integrate** our skeletal muscle snRNA-seq data  **uninjury and 3 day post injury** by using harmony
2. Annotate the **cell type** of D0 and D3 of snRNA seq data using **reference based annotation tools** such as **scPred** and **SingleR**.
3. After extracting and **subclustering immune cells**, the immune cell type was newly defined, the **overexpressed gene set(DEG)** was found among the subtypes, and the **functional profile** related to it was examined using **clusterProfiler**.

# Conclusion

1. We integrate our skeletal muscle snRNA-seq data uninjury and Through snRNA-seq analysis, we identified an immune subcluster about 3 Day post-injury in skeletal muscle. We obtained around 1500 unique gene ontology biological pathway terms across all subclusters, and we intend to conduct further investigation.
2. Annotate the cell type of D0 and D3 of snRNA-seq data using reference based annotation tools such as scPred and SingleR.
3. After extracting and subclustering immune cells, the immune cell type was newly defined, the overexpressed gene set(DEG) was found among the subtypes, and the functional profile related to it was examined using clusterProfiler.

---

**Thank you**

---

# Question?

---

