

# CytoAtlas: Pan-Disease Single-Cell Cytokine Activity Atlas

## Data Sources

~29M cells + ~31K bulk samples

### Single-Cell

#### CIMA

421 donors, healthy population

6.5M cells

#### Inflammation Atlas

main/val/ext, 20 diseases

6.3M cells

#### scAtlas

Normal + Cancer, multi-organ

6.4M cells

#### parse\_10M

12 donors × 90 cytokines (+PBS)

9.7M cells

### Bulk RNA-seq

#### GTEx

~50 normal tissues, TPM

19.8K samples

#### TCGA

33 cancer types, RSEM

11.1K samples

## Activity Inference

Ridge regression – GPU-accelerated

### 3 Signature Matrices

#### CytoSig

43 cytokines

#### LinCytoSig

178 cell-type specific

#### SecAct

1,170 secreted proteins

### Pipeline

1. Pseudobulk aggregation  
(donor, donor×celltype)
2. Gene expression matrix  
(genes × samples)
3. Ridge regression  
(signature × samples → activity)
4. Cross-sample correlation  
(predicted vs observed)

## Validation

6 datasets × 3 signatures

### 7 Validation Analyses

#### Overall Performance

Spearman ρ across all targets

\$4.1

#### Per-Tissue Stratified

GTEx tissues / TCGA cancer types

\$4.2

#### Cross-Dataset Comparison

CytoSig vs SecAct across 6 datasets

\$4.3

#### Best/Worst Targets

Top and bottom correlated targets

\$4.4

#### Cross-Dataset Consistency

Same targets across datasets

\$4.5

#### Aggregation Levels

Donor → celltype → single-cell

\$4.6

#### Bulk RNA-seq Validation

GTEx + TCGA concordance

\$4.7