

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

## Data Sources

~29M cells + ~31K bulk samples

### Single-Cell

**CIMA** **6.5M cells**

421 donors, healthy population

**Inflammation Atlas** **6.3M cells**

main/val/ext, 20 diseases

**scAtlas** **6.4M cells**

Normal + Cancer, multi-organ

**parse\_10M** **9.7M cells**

12 donors × 90 cytokines (+PBS)

### Bulk RNA-seq

**GTEx** **19.8K samples**

~50 normal tissues, TPM

**TCGA** **11.1K samples**

33 cancer types, RSEM

## 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

### Overall Performance

Spearman  $\rho$  across all targets

### Per-Tissue Stratified

GTEx tissues / TCGA cancer types

### Cross-Dataset Comparison

CytoSig vs SecAct across 6 datasets

### Best/Worst Targets

Top and bottom correlated targets

### Cross-Dataset Consistency

Same targets across datasets

### Aggregation Levels

Donor → celltype → single-cell

### Bulk RNA-seq Validation

GTEx + TCGA concordance