

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 \times 90 cytokines (+PBS)

Bulk RNA-seq

GTEX **19.8K samples**

\sim 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 \times celltype)
2. Gene expression matrix
(genes \times samples)
3. Ridge regression
(signature \times samples \rightarrow activity)
4. Cross-sample correlation
(predicted vs observed)

Validation

6 datasets \times 3 signatures

7 Validation Analyses

Overall Performance §4.1

Spearman p across all targets

Per-Tissue Stratified §4.2

GTEx tissues / TCGA cancer types

Cross-Dataset Comparison §4.3

CytoSig vs SecAct across 6 datasets

Best/Worst Targets §4.4

Top and bottom correlated targets

Cross-Dataset Consistency §4.5

Same targets across datasets

Aggregation Levels §4.6

Donor \rightarrow celltype \rightarrow single-cell

Bulk RNA-seq Validation §4.7

GTEx + TCGA concordance