CytoRxAI:Foundation Models for Scalable Single Cell Transcriptomic Analysis

Strand Life Sciences developed the CytoRx AI platform to support efficient and reliable analysis of single cell RNA sequencing data using finely tuned foundation models designed for transcriptomic applications.

The platform enables cell type annotation, gene function analysis, perturbation prediction, multimodal data integration, and cancer drug response modeling within a unified framework. CytoRx AI addresses common limitations of traditional workflows, including information loss, restricted scalability, and variability in annotation, and provides a valuable resource for researchers working with large and complex single cell datasets.

What are the Foundation Models in CytoRxAI?

scBERT

- Learns gene level relationships from unlabeled single cell RNA sequencing data and is fine tuned for cell type annotation in disease contexts.
- Used alongside scGPT to **improve annotation accuracy** in complex datasets such as inflammatory disease samples.

scGPT

- **Trained on millions of cells across tissues** using masked gene prediction to capture transcriptomic context.
- Provides robust representations for cell type annotation when combined with scBERT, particularly in heterogeneous disease data.

Geneformer

- Generates gene embeddings that support prediction of chromatin and network dynamics, disease modelling.
- Within CytoRx AI, Geneformer contributes to gene-function classification.

scLong

- A large scale model pretrained on **~48 million human cells, across ~27,874 genes** to capture long range transcriptomic dependencies.
- Supports perturbation effect prediction, gene regulatory network inference, and drug response modeling that require full transcriptome context.

CancerFoundation

- A cancer focused model trained on a selected set of highly variable genes relevant to tumor biology.
- Focuses on a subset of **~29,000 highly variable genes** relevant to tissues of interest.
- Enables cancer specific analyses such as malignant cell state identification and drug response prediction with improved interpretability.

Deep Learning Methods Using Foundation Model Embeddings

AttentionPert and GEARS

- Predicts gene expression outcomes under single and multiplex genetic perturbations, including unseen conditions.
- Used in ensemble configurations with embeddings from Gene2Vec, scLong, and scGPT to improve perturbation prediction accuracy.

Foundation Model	Number of Model Parameters	Tasks
scBERT	~ 8 million	Cell type annotation
scGPT	~ 53 million	Cell type annotation, Gene prediction
Geneformer	~ 10 million	Chromatin and network dynamics prediction, Disease modelling
scLong	~ 1 billion	Perturbation effect prediction, Drug-response modeling
CancerFoundation	~ 10 million	Cancer-drug response prediction, Resistance-associated transcriptional program identification

How are these foundation models integrated into CytoRx AI?

CytoRx AI uses ensemble methods (also called stacked generalisation) to combine outputs from several foundation models per task, which improves robustness across tissues, diseases, and experimental conditions, and is critical for biomarker discovery and target identification.

What were the performance gains when compared with internal benchmarks?

- **Cell-type annotation:** +7% accuracy, +33% improvement (macro-F1)
- **Perturbation prediction:** 10–20% reduction in mean squared error
- **Gene-function classification:** improved by 15% (macro-F1)
- **Cancer drug-response prediction:** improved by 15% (Pearson correlation)

How does CytoRx AI help researchers?

The CytoRx AI architecture

- supports accurate, and scalable analysis of single cell transcriptomic data, building on recent advances in AI foundation models for omics research.
- is structured to capture complex relationships between genes and between cells and genes across large and diverse datasets.

- supports more reliable cell type annotation, gene function inference, perturbation outcome prediction, multimodal data integration, and drug response analysis.

Through this design, CytoRx AI serves as a practical resource for both foundational biological research and pharmaceutical target discovery in high throughput single cell omics studies.