Request Information

Permalink

# Robust Single Cell Classification Methods and System

Tech ID: 33025 / UC Case 2023-902-0

**CONTACT**

Marc Oettinger
marc.oettinger@ucsc.edu
tel: 831-502-0253.

**INVENTORS**

▶ Jonsson, Vanessa

▶ Lehrer, Julian

▶ Mostajo-Radji, Mohammed

**OTHER INFORMATION**

**KEYWORDS**

transcriptomics, cell-type, cell type, classification, cell atlas, sequencing, RNA-seq, single-cell, transcriptome, expression matrix, clusters, clustering, classifier, class distribution, next-generation sequencing, NGS

**CATEGORIZED AS**

▶ **Biotechnology**

   ▶ Bioinformatics

   ▶ Genomics

▶ **Computer**

   ▶ Software

▶ **Research Tools**

   ▶ Bioinformatics

**RELATED CASES**

2023-902-0

## BACKGROUND

High-throughput next-generation sequencing (NGS) systems have allowed for large scale collection of transcriptomic data with single cell resolution. Within this data lies variability allowing researchers to characterize and/or infer certain morphological aspects of interest, such as single cell type, cell state, cell growth trajectories, and inter-cellular gene regulatory networks. All of these qualities are important parts of understanding how cells interact with one another, both for building better cellular models *in vitro* and for understanding biological processes *in vivo*. While the size of single cell data has increased massively, NGS techniques for key pieces of analysis have not kept pace, using slow, manual pipelines of domain experts for initial clustering. Attempts to improve NGS classification performance have fallen short as the numbers of cell types (often asymmetric) and cell subtypes have increased while the number of samples per label has become small. The technical variability between NGS experiments can make robust classification between multiple tissue samples difficult. Moreover, the high-dimensional nature of NGS transcriptomic data makes this type of analysis statistically and computationally intractable.

## TECHNOLOGY DESCRIPTION

To help address these challenges in NGS, investigators at UC Santa Cruz (UCSC) have developed a new approach based on deep learning and transformer-based neural network. UCSC's Scalable, Interpretable Machine Learning for Single-Cell (SIMS) targeting traditional NGS classification methods though an end-to-end modeling pipeline for discrete morphological prediction of single-cell data, with minimal overhead and high classification accuracy. SIMS takes an expression matrix with associated labels and learns a mapping between transcriptome and cell type. This mapping can be used to automatically infer cell types in new single-cell data. SIMS can handle both small and large transcriptomic datasets effectively while having direct interpretability at individual sample level.

## APPLICATIONS

- ▶ cell research
- ▶ drug discovery
- ▶ software

## ADVANTAGES

- ▶ modular and fast to develop with
- ▶ no custom hardware or specialized computer
- ▶ can be used with a broad range of cell types
- ▶ outperforms peer tools including scANVI

## INTELLECTUAL PROPERTY INFORMATION

| Country | Type | Number | Dated | Case |
|---------|------|--------|-------|------|
| Patent Cooperation Treaty | Published Application | WO2024/103057 A | 05/16/2024 | 2023-902 |

Additional Patent Pending

## RELATED MATERIALS

- ▶ SIMS: A deep-learning label transfer tool for single-cell RNA sequencing analysis

## STATE OF DEVELOPMENT

- ▶ patent pending
- ▶ software

## ADDITIONAL TECHNOLOGIES BY THESE INVENTORS

- ▶ METHODS OF PRODUCING AND USING AVIAN EMBRYONIC STEM CELLS AND AVIAN TELENCEPHALIC ORGANOIDS