A053 · CellFun: Decoding Cellular Functions from Single-Cell and Spatial Transcriptomics with Agentic AI

Presenter: Kulandaisamy Arulsamy — Department of Cardiology, Boston Children's Hospital, Boston, MA 02115, USA.

Authors:

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Abstract Functional enrichment analysis is central to interpreting transcriptomic data, yet its coverage and conclusions depend on gene selection and ranking, enrichment strategy, database choice, and prioritization of redundant functional and disease-associated terms. These decisions are made independently, with no unified framework for determining which functional profile best represents cellular biology. We developed CellFun, a multi-agent framework integrating gene prioritization, multidimensional pathway evaluation, adaptive optimization, and evidence-grounded interpretation. We used SCIG, a machine-learning framework for identifying and prioritizing cell identity genes (CIG), to derive CIG scores, previously shown to be useful in network analysis and single-cell clustering1. Here, we tested whether CIG-based gene prioritization improves functional interpretation beyond conventional expression-based selection. Across 462 cell types from human and mouse tissues, CIG-based gene sets1,2 recovered more significant functional and disease-associated terms than highly expressed or cell-type-specific gene sets, while showing greater specificity than high-expression gene sets and comparable specificity to cell-type-specific gene sets across multiple databases. We then developed a Q-score evaluating enrichment profiles across six dimensions: statistical significance, functional breadth, theme-to-term ratio, specificity, semantic coherence, and gene-support diversity after redundancy reduction. Cell type–database pair evaluations showed that CIG-based and cell-type-specific gene sets achieved the strongest Q-score performance, substantially exceeding expression-based strategies. CellFun-derived pathway profiles and functional summaries preserved transcriptome-level relationships, recovered cell functions, and supported marker-independent cell-identity retrieval. Together, CellFun enables functional annotation of single-cell transcriptomes, with extension to spatially resolved data, to map cellular functions across tissues, compare functional states in health and disease, and identify disease-associated pathways and candidate regulators.

References: 1. Arulsamy, K. et al. SCIG: Machine learning uncovers cell identity genes in single cells by genetic sequence codes. Nucleic Acids Res. 53, (2025). 2. Xia, B. et al. Machine learning uncovers cell identity regulator by histone code. Nat. Commun. 11, 2696 (2020).

CellFun: Decoding Cellular Functions from Single-Cell and Spatial Transcriptomics with Agentic AI

Kulandaisamy Arulsamy[#], Rongbin Zheng, Lili Zhang, and Kaifu Chen[*]

[1] *Department of Cardiology, Boston Children's Hospital, Boston, MA 02115, USA.*

[2] *Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA.*

** Corresponding author; # Presenting author*

Email: *Kaifu.Chen@childrens.harvard.edu; KulandaiSamy.Arulsamy@childrens.harvard.edu*

Abstract

Functional enrichment analysis is central to interpreting transcriptomic data, yet its coverage and conclusions depend on gene selection and ranking, enrichment strategy, database choice, and prioritization of redundant functional and disease-associated terms. These decisions are made independently, with no unified framework for determining which functional profile best represents cellular biology. We developed CellFun, a multi-agent framework integrating gene prioritization, multidimensional pathway evaluation, adaptive optimization, and evidence-grounded interpretation.

We used SCIG, a machine-learning framework for identifying and prioritizing cell identity genes (CIG), to derive CIG scores, previously shown to be useful in network analysis and single-cell clustering[1]. Here, we tested whether CIG-based gene prioritization improves functional interpretation beyond conventional expression-based selection. Across 462 cell types from human and mouse tissues**,** CIG-based gene sets[1,2] recovered more significant functional and disease-associated terms than highly expressed or cell-type-specific gene sets, while showing greater specificity than high-expression gene sets and comparable specificity to cell-type-specific gene sets across multiple databases. We then developed a Q-score evaluating enrichment profiles across six dimensions: statistical significance, functional breadth, theme-to-term ratio, specificity, semantic coherence, and gene-support diversity after redundancy reduction. Cell type–database pair evaluations showed that CIG-based and cell-type-specific gene sets achieved the strongest Q-score performance, substantially exceeding expression-based strategies.

CellFun-derived pathway profiles and functional summaries preserved transcriptome-level relationships, recovered cell functions, and supported marker-independent cell-identity retrieval. Together, CellFun enables functional annotation of single-cell transcriptomes, with extension to spatially resolved data, to map cellular functions across tissues, compare functional states in health and disease, and identify disease-associated pathways and candidate regulators.

References:

1. Arulsamy, K. *et al.* SCIG: Machine learning uncovers cell identity genes in single cells by genetic sequence codes. *Nucleic Acids Res.* **53**, (2025).
2. Xia, B. *et al.* Machine learning uncovers cell identity regulator by histone code. *Nat. Commun.* **11**, 2696 (2020).