SCP-Analyst: An End-to-End Single-Cell Proteomics Web Platform for Exploring Cellular Heterogeneity

Deeptarup Biswas, Ph.D.

Post-doctoral Fellow · Steen Lab · Accelerator Grant Awardee

Affiliation: Neurology, Boston Children's Hospital, Harvard Medical School

Project: Mapping the Thymic Microenvironment at Single-Cell Proteome Resolution via the SCP-Analyst Pipeline

1. Scientific Impact and Project Overview

While single-cell proteomics (SCP) offers the unprecedented ability to directly measure proteins at cellular resolution, the transition from raw mass spectrometry data to biological insight is severely bottlenecked by bioinformatics complexity. SCP datasets are inherently sparse, noisy, and require intricate normalization pipelines that can be challenging to implement. To remove these computational barriers and make SCP data analysis accessible to a broader community of researchers, we developed SCP-Analyst, an interactive, open-source web application. SCP-Analyst seamlessly ingests minimally modified outputs from search engines like DIA-NN and FragPipe, transforming them through a rigorous, user-guided pipeline into publication-ready figures. To demonstrate the platform's utility, we applied it to map the thymic microenvironment at single-cell resolution. The thymus gland is a central organ of the adaptive immune system, responsible for the maturation and education of T lymphocytes. We implemented an optimized SCP workflow to map essential cell-type heterogeneities, specifically focusing on the cortical and medullary thymic epithelial cells (cTECs and mTECs) that drive T cell selection. The inherently small size and low protein content of thymocytes presented a significant analytical challenge. However, by utilizing a targeted surface-marker sorting strategy, we successfully isolated high-purity mTEC and cTEC populations, yielding ~1,500 protein identifications per cell across 375 single cells.

2. Project Outcomes

SCP-Analyst mirrors the natural analytical workflow of a proteomics researcher, eliminating the need for advanced coding expertise. We utilized the platform's integrated modules to process our single-cell thymus dataset from raw intensity to biological divergence:

- **Data Assessment and Quality Control:** Upon uploading the raw intensity matrix and experimental metadata, users begin in the Data Assessment module. SCP-Analyst automatically calculates global dataset missingness and generates a diagnostic kNN (k-Nearest Neighbors) Elbow Plot (Figure 1). By simulating imputation on a dense subset of the data, this plot visually guides the identification of the optimal *k* value to minimize error, shifting imputation from guesswork to a data-driven decision.

- **Preprocessing and Normalization:** Once optimal parameters are identified, the platform executes kNN imputation to handle missing values, applies Total Ion Current (TIC) normalization to correct for technical run-to-run variability, and utilizes Log2 transformation to stabilize variance. The dashboard immediately updates to display the post-processed data distribution side-by-side with the raw data, ensuring normalization success prior to downstream evaluation.

- **Dimensionality Reduction and Feature Extraction:** Researchers can generate up to four independent UMAP projections simultaneously, colored dynamically by biological classification. Furthermore, we implemented Partial Least Squares Discriminant Analysis (PLS-DA) coupled with

Variable Importance in Projection (VIP) scoring. This allowed us to mathematically isolate the specific proteins driving the functional separation between cortical and medullary thymic states, turning clustering observations into actionable biological targets.

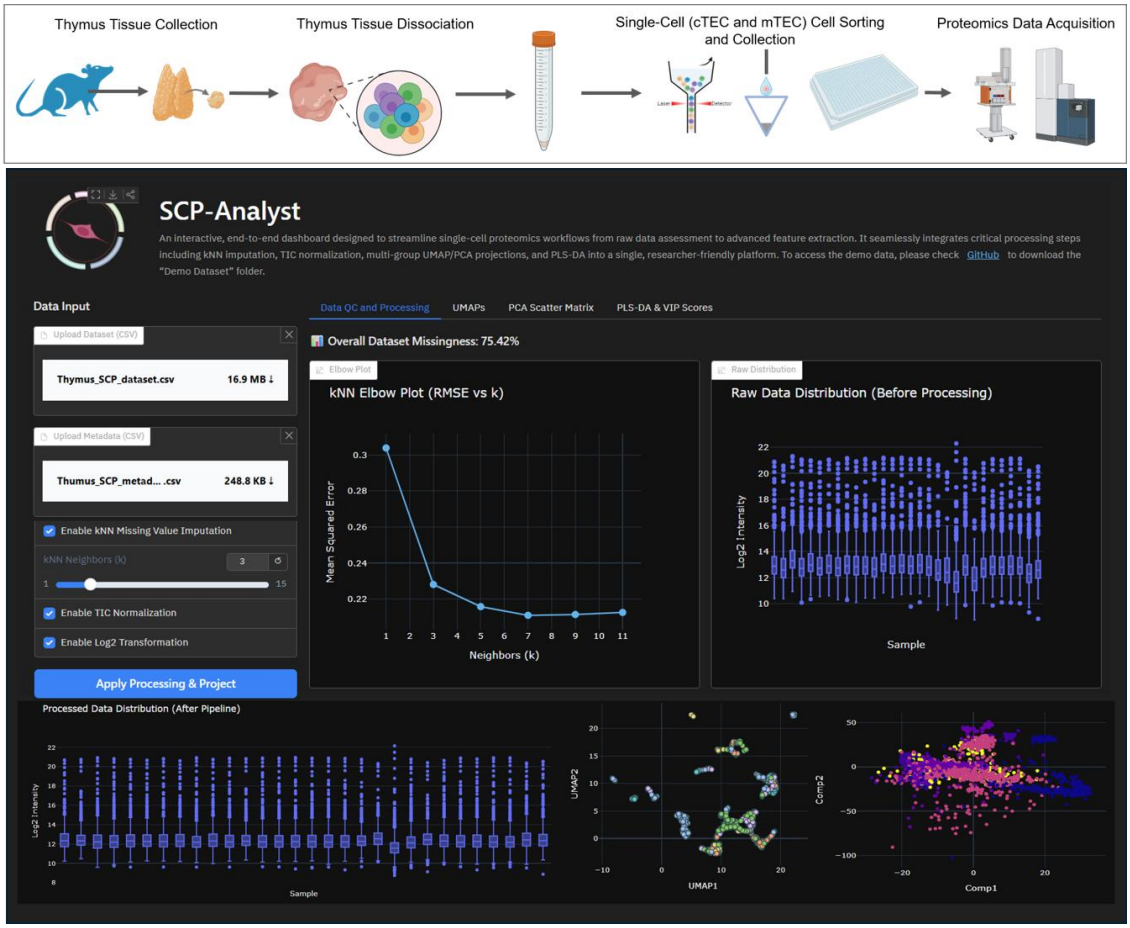


Figure 1: End-to-end experimental and computational pipeline utilizing SCP-Analyst. (Top) Overview of the biological workflow, detailing thymus tissue collection from a murine model, tissue dissociation, isolation of specific thymic epithelial cells (cTECs and mTECs) via targeted cell sorting, and subsequent high-throughput proteomics data acquisition. (Bottom) The SCP-Analyst graphical user interface processing the resulting dataset. The interface features a configuration sidebar (left) for data ingestion and parameter tuning, including toggles for kNN missing value imputation, TIC normalization, and Log2 transformation. The primary visualization workspace (right) dynamically generates a diagnostic kNN Elbow Plot to guide imputation, and displays side-by-side box plots for immediate visual validation of raw versus post-processed data distributions before downstream dimensionality reduction (UMAP/PCA).

3. Dissemination, Data Sharing, and Future Plans

To ensure widespread community adoption of this computational framework, we have prioritized seamless accessibility:

- **Live Application:** The tool is open-source, cloud-hosted web dashboard via Hugging Face Spaces: https://deepbiswas-scp-analyst-v1.hf.space/
- **Open-Source Repository:** The complete documentation and instructions, is publicly maintained on GitHub: https://github.com/DeepHMS/SCP-Analyst

This project is currently ongoing, and the associated publication detailing the findings will be submitted soon. Upon publication, the complete single-cell proteomics dataset will be deposited and publicly released in established repositories such as ProteomeXchange to support the broader scientific community.