Analyzing Single-Cell RNA-sequencing Data to Classify Cell States

Shreya Sharma[1], Samantha A. O'Connor[1], Christopher L. Plaisier[1]

[1]School of Biological and Health Systems Engineering, Arizona State University, Tempe AZ, USA;

Introduction

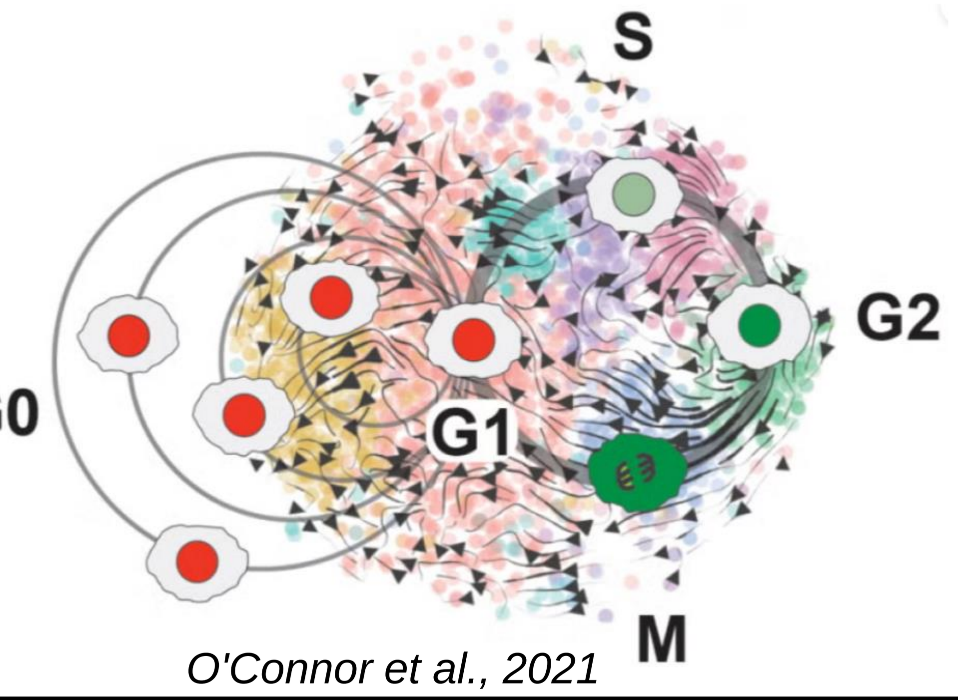
Single-cell RNA-sequencing (scRNA-seq) allows researchers to analyze gene expression at single-cell resolution, uncovering previously hidden subpopulations and dynamic cellular states.

As scRNA-seq use has grown, **computational classifiers** have been developed to efficiently annotate cell states across datasets. The **cell cycle** is a fundamental biological process often targeted by these classifiers.The **Plaisier Lab** developed **ccAF** (cell cycle classifier ASU/Fred Hutch), a cell cycle classifier trained on human neural stem cells, which predicts a novel **quiescent G0-like state** in neuroepithelial-derived glioma cells.

Cells in the G0-like state have exited the cell cycle and are found in neural stem and progenitor populations.

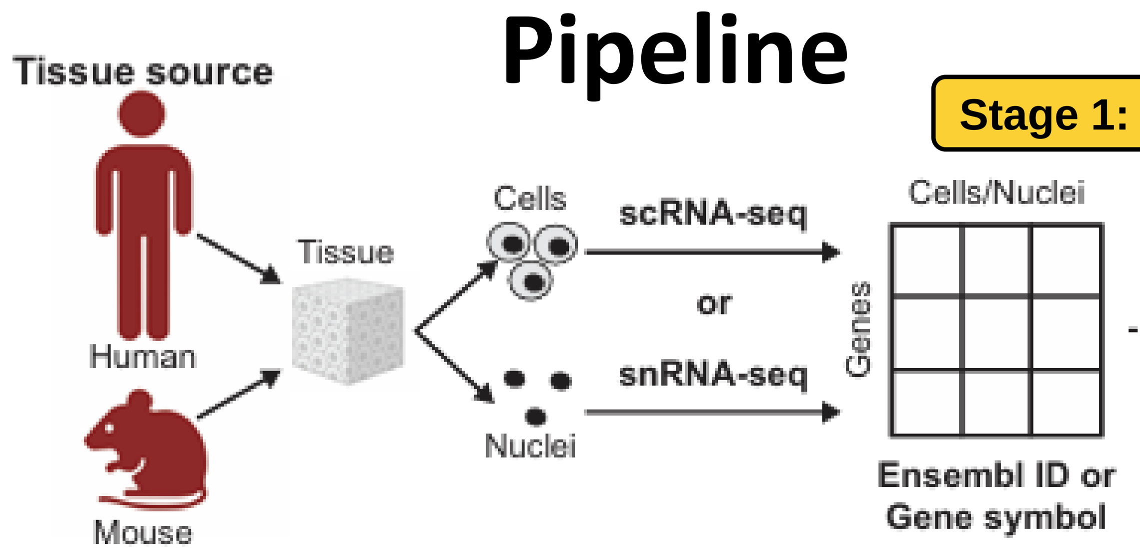
Higher expression of G0-like signature genes correlates with **better prognosis** in glioma patients.

G0-like states are present in multiple tissue types and are thought to represent **quiescence or dormancy**, making them promising targets for treating **tumors** and **neurodegenerative diseases**.



O'Connor et al., 2021

Pipeline

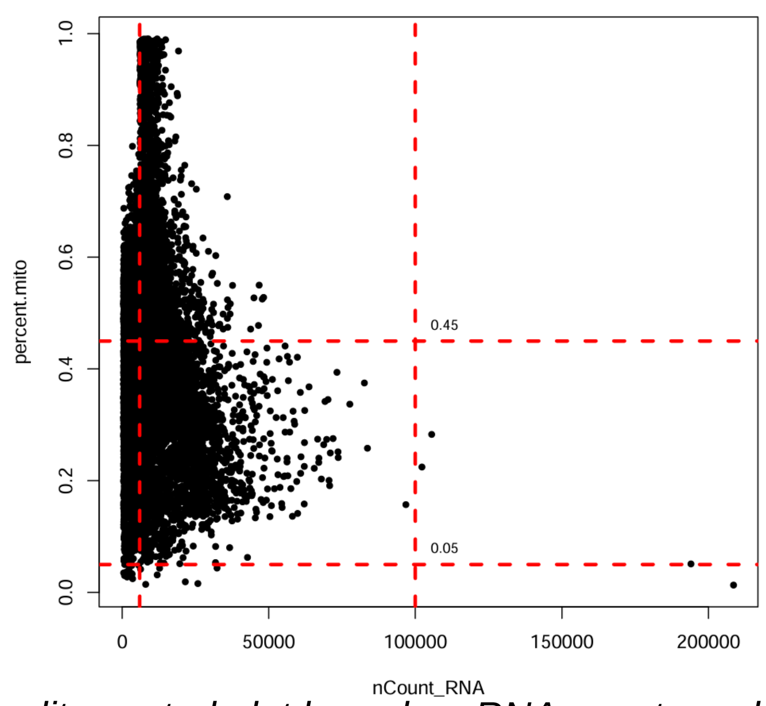


Overview of pipeline

- Prioritized datasets with downloadable metadata, including source tissue, clinical context, applied perturbations, and previously inferred cell types.

Stage 1: Test Dataset Discovery

- Identified 111 **publicly available 10x Genomics scRNA-seq datasets** from various published studies.
- Selected **datasets with over 1,000 cells** to ensure sufficient sampling for downstream analysis.

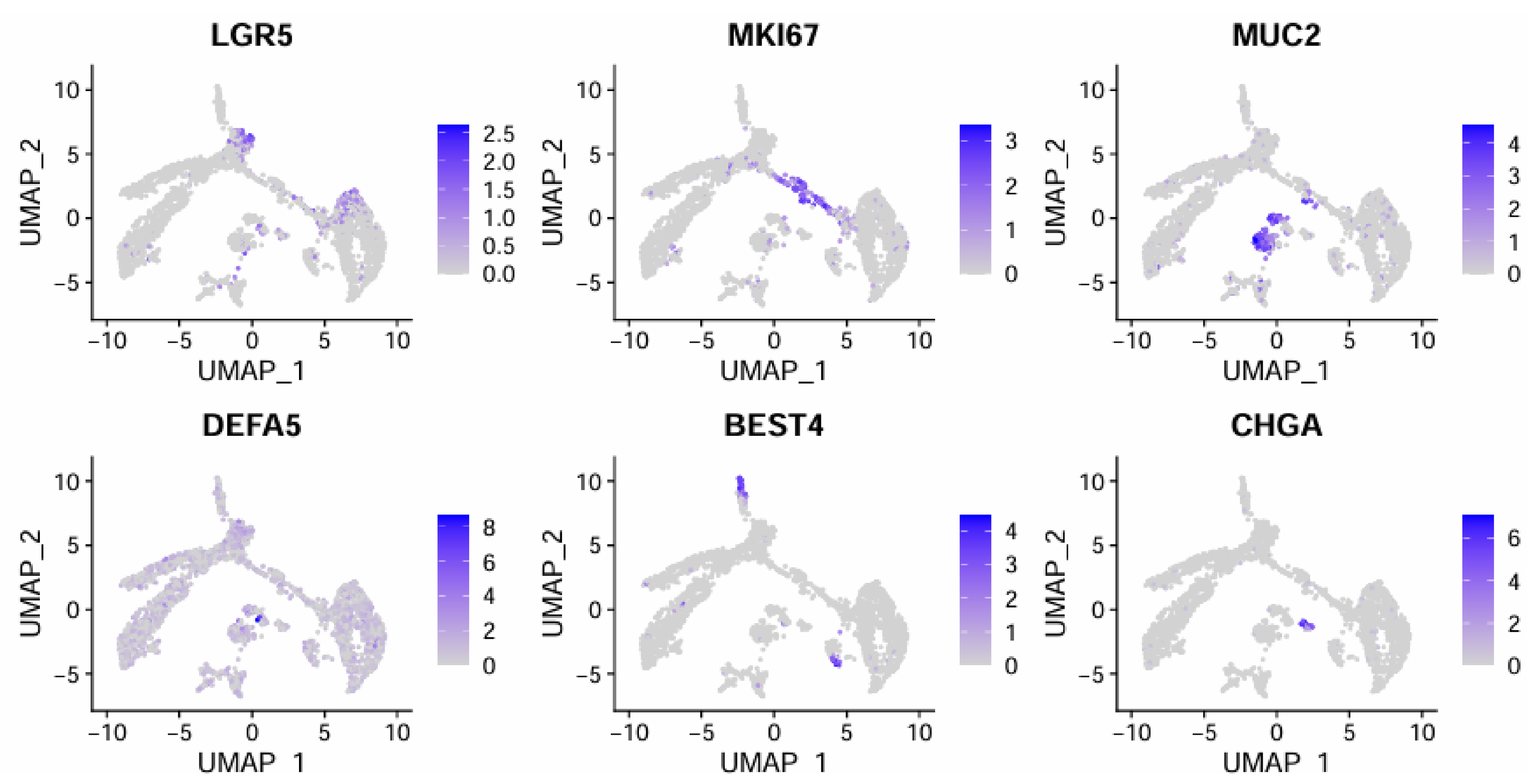


Quality control plot based on RNA counts and percent mitochondria, cutoffs in red.

Stage 2: Quality Control

- Assessed cell quality based on the **percentage of mitochondrial gene expression** and the number of unique RNA molecules.
- Filtered out low-quality cells that did not meet **thresholds** for **mitochondrial content** or **RNA complexity**.
- Applied **SCTransform normalization** to reduce technical variability while preserving biological signal.

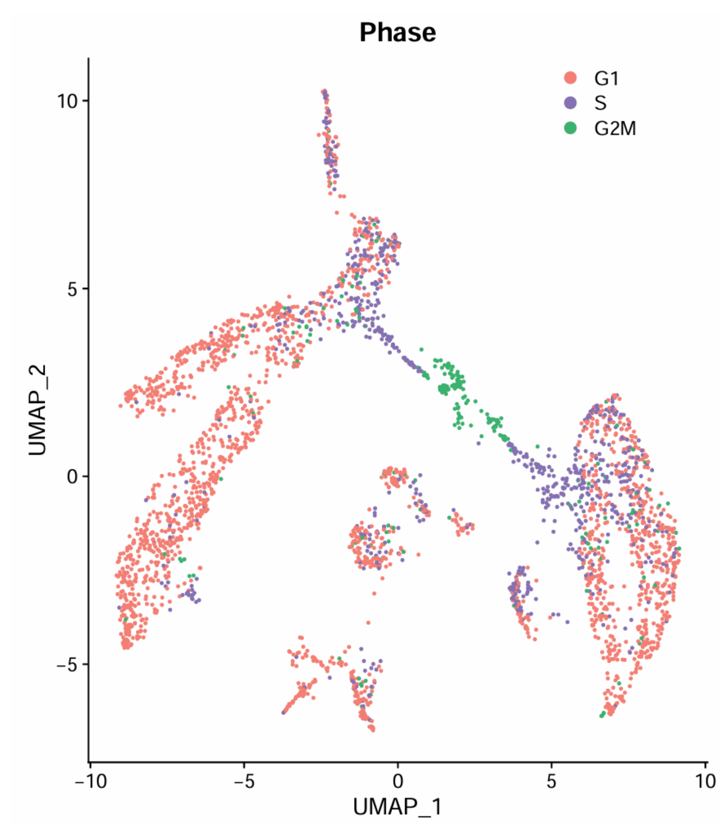
Stage 3: Clustering and Annotation



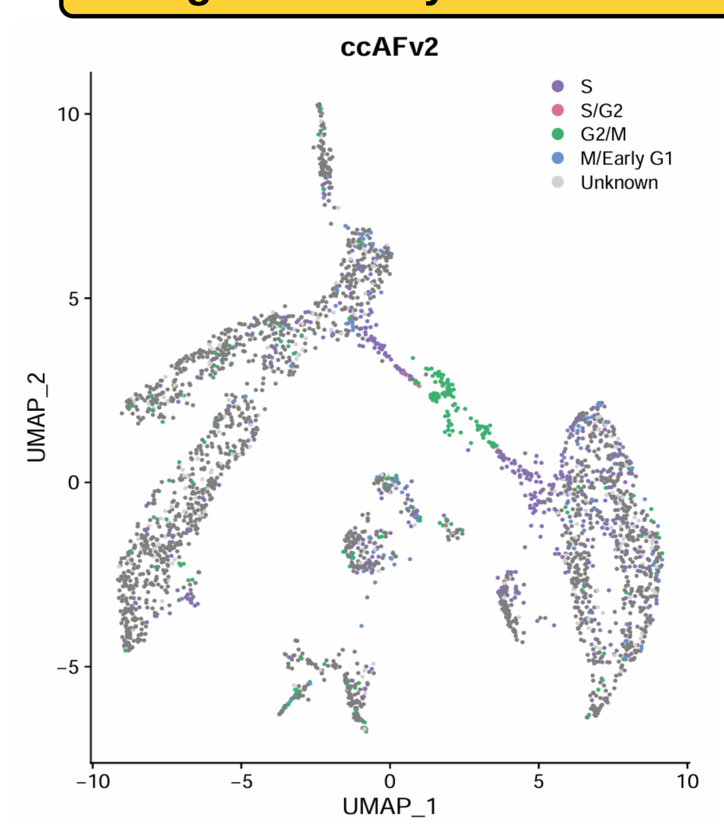
UMAPs showcasing upregulated gene expression

- **Visualized marker gene expression** to identify and distinguish key intestinal cell types (e.g., stem, goblet, Paneth, tuft).
- **Annotated clusters** by assigning biologically meaningful cell type labels based on known gene expression profiles.
- **Compared clustering with annotations** using dimensionality reduction plots to assess accuracy of cell type identities.

Stage 4: Cell Cycle Classification



UMAP projection of cells labelled by Seurat's cell cycle scoring.

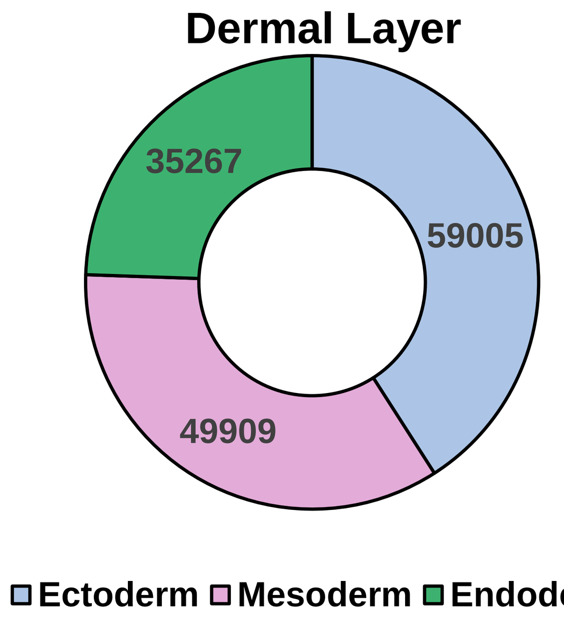


UMAP projection of cells labelled using ccAFv2.

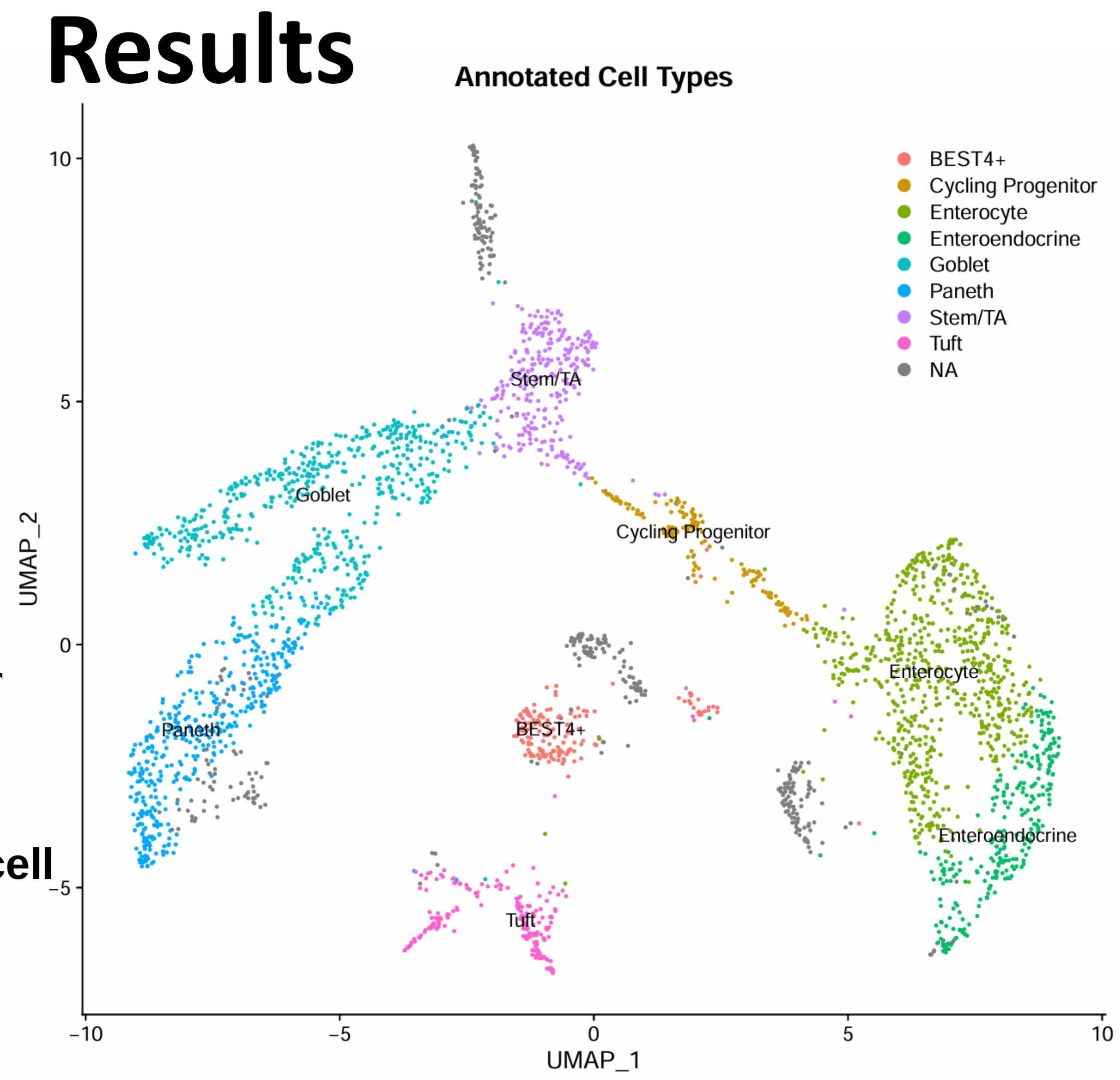
- Compared traditional Seurat annotations (Tirosh, I., et. al, (2016)) with ccAFv2-based classifications to evaluate prediction accuracy.

Results

Cell Counts From Each Dermal Layer

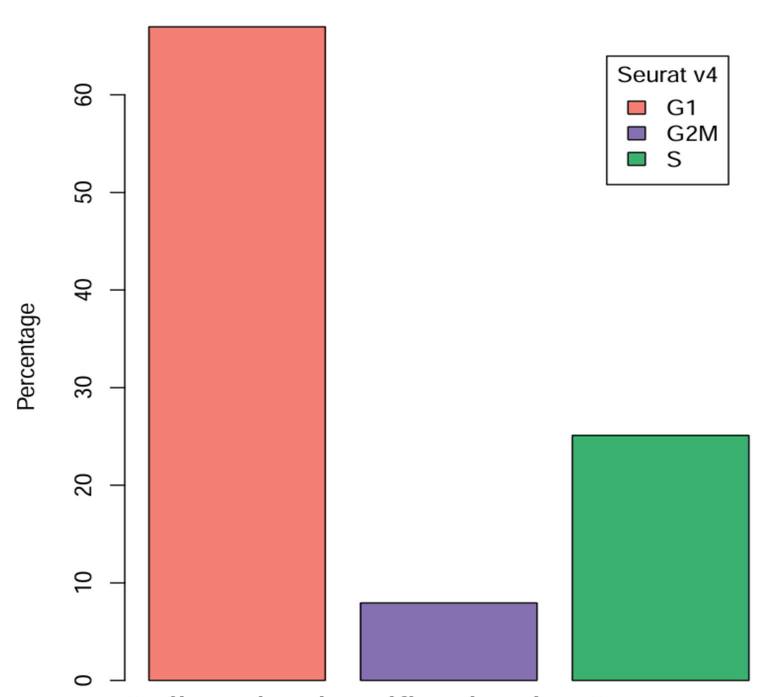


Cell count per dermal tissue layer across curated datasets.

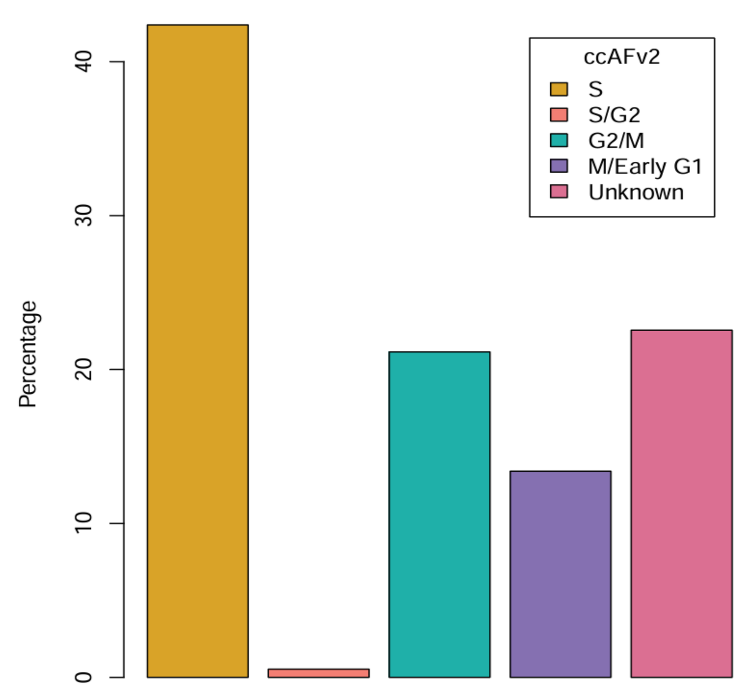


Annotated UMAP of cell types within a scRNA-seq sample.

- **Quantified and visualized cell counts** across annotated dermal layers from the 111 dataset to evaluate tissue representation and sampling consistency.
- **Identified clusters** of cell types based on gene expression.
- Defined **more specific cell states than current** publicly available **classifiers.**
- Identified "unknown" states that may hold relevance to a **possible G0/ quiescent state**.



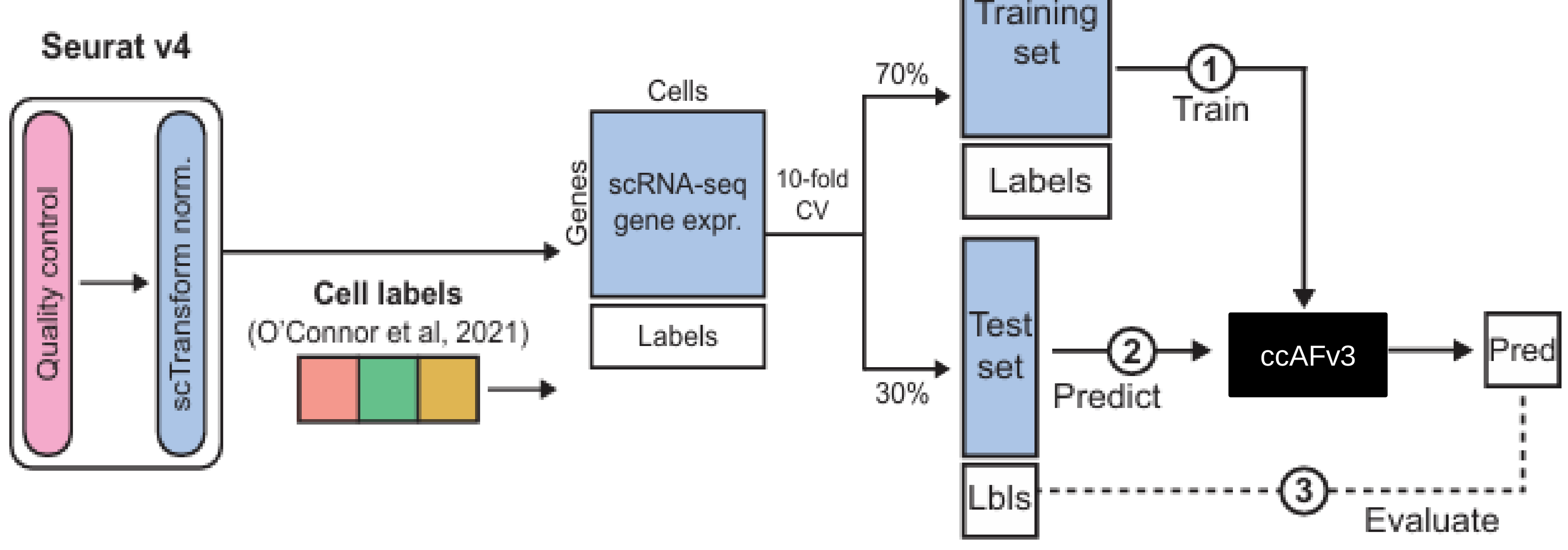
Cell cycle classification by Seurat.



Cell cycle classification by ccAFv2.

Future Directions and Acknowledgements

Stage 5: Testing and Training



Visualization of the integration of more markers for G0 in ccAFv3, found through datasets.

- Split cells into **testing and training datasets**.
- Analyze datasets by applying ccAFv3, **labelling** with previously **established cell cycle labels** (O'Connor et al. 2021)
- **Assess accuracy** and compared with publicly available cell cycle classifiers (Seurat v5).

SBHSE School of Biological and Health Systems Engineering | ASU Ira A. Fulton Schools of Engineering Arizona State University | FRED HUTCH CURES START HERE™

FURI

ASU Ira A. Fulton Schools of Engineering Arizona State University