



# DeepSCore: a comprehensive, flexible, and user-friendly single-cell multimodal data annotation tool

Jessica Kanglin Li<sup>1</sup>, Elisabetta Mereu<sup>1</sup>

<sup>1</sup>Cellular Systems Genomics Group, Josep Carreras Leukemia Research Institute



## 1 INTRODUCTION

The development of single-cell multimodal sequencing technologies allows for the simultaneous profiling of transcriptomes and epigenomes within the same cells. This presents a distinctive opportunity to collectively analyse multimodal data at the single-cell level, facilitating the orthogonal characterization of distinct cell types and states. However, the mapping of different data sources remains a statistical and computational challenge.

In this project, we introduce deepSCore, a novel R package designed as a reference-based classifier for annotating single-cell multiomics data. Leveraging a flexible deep learning model, deepSCore classifies cell types based on the reference data, empowering users to adapt the model training process by adjusting parameters according to their specific input data.

## 2 METHODS

The use of deepSCore is based on three main steps, in the first step the user will set the reference data and the query data, and the tool will extract the common features between them. Then, by using this input data, a flexible deep learning model will be trained, in this step the user can configure the model training process adapting to the needs by changing the training parameters.

### APPLICATIONS:

- RNA → RNA
- RNA → ATAC (GA)
- ATAC (GA) → RNA
- ATAC → ATAC

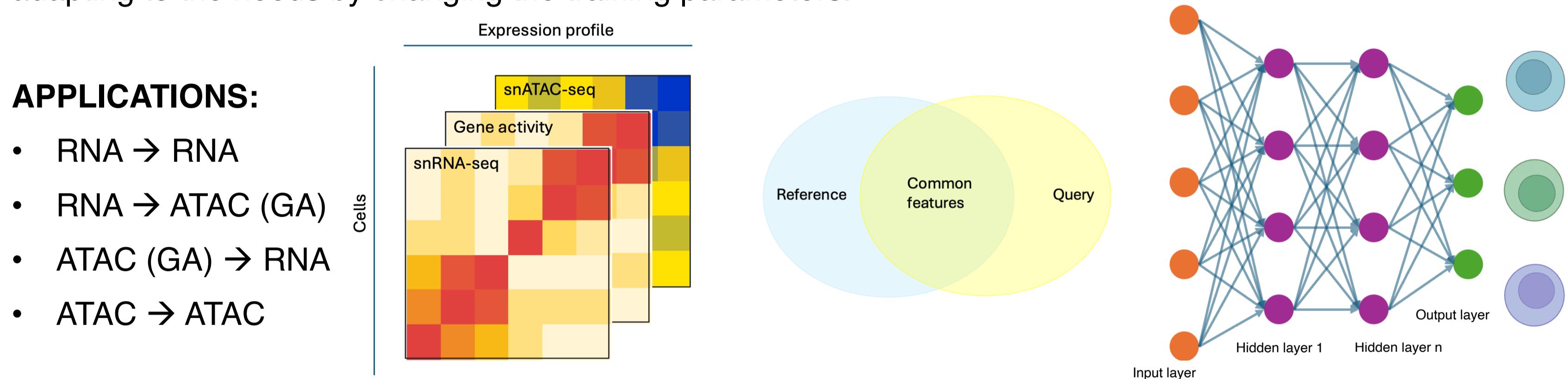


Figure 1. Scheme of deepSCore's workflow

## 3 RESULTS

Extensive testing using different human healthy datasets, including pancreas and kidney, the classifications demonstrated high-quality cell type annotations for each case. Comparative analysis against established cell type annotation tools such as Seurat, SCINA, scID, scMAP, and CHETAH revealed deepSCore's significantly high performance.

By using a single-cell multiome kidney data, we tested the performance of deepSCore in different annotation complexity levels, comparing also the cross-modality classification the ATAC to ATAC classification.

In general, the performance of ATAC to ATAC classification is better than the cross-modality classification, showing that predicting directly from peaks to peaks the noise produced by the difference between modalities is removed.

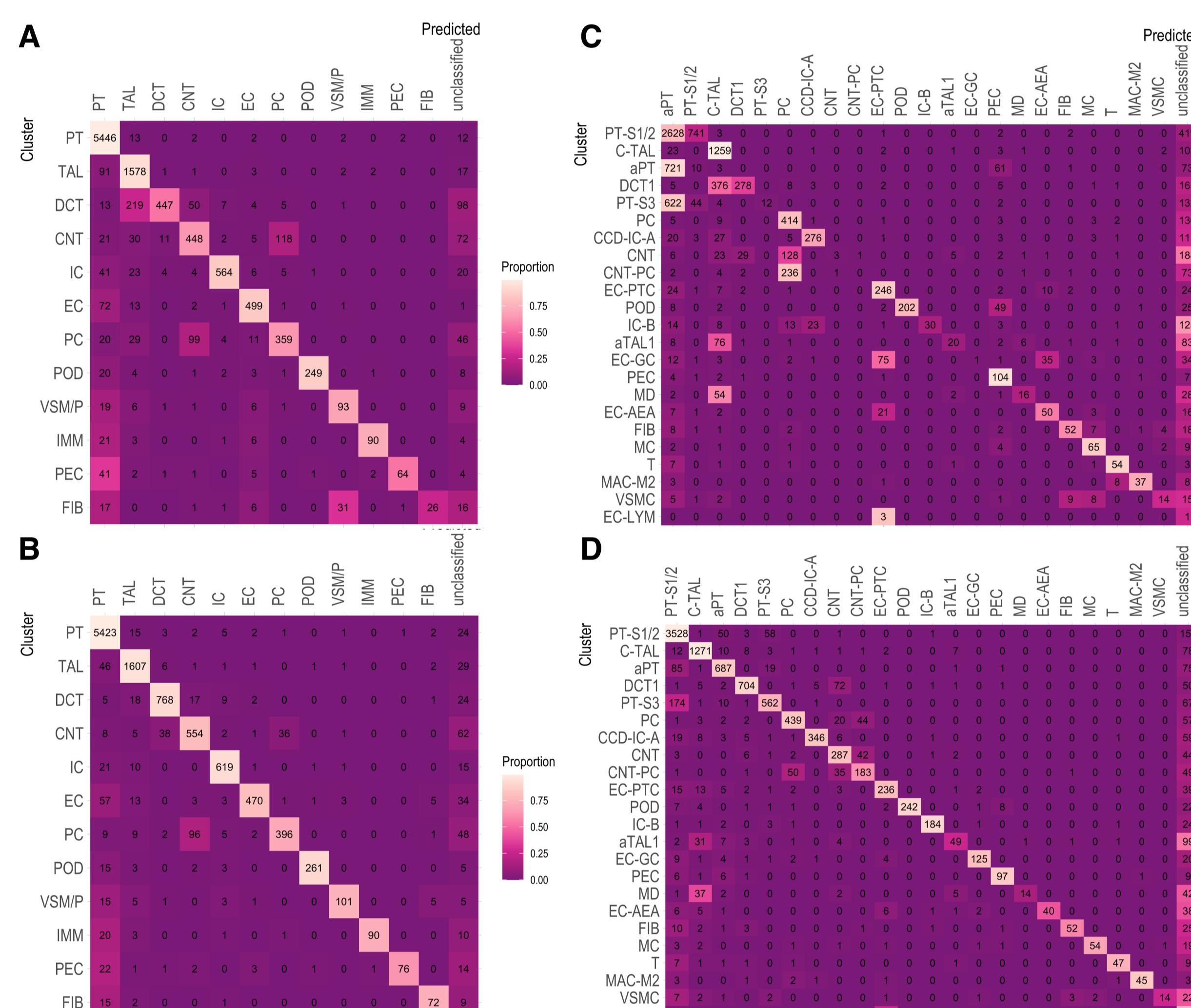


Figure 2. Performance of deepSCore when classifying kidney cells. **(A)** Cross-modality classification (RNA to ATAC) at simple annotation level L1. **(B)** ATAC to ATAC classification at simple annotation level L1. **(C)** Cross-modality classification at complex annotation level L2. **(D)** ATAC to ATAC classification at complex annotation level L2.

Comparing the overall performance between the packages, deepSCore has shown a higher general performance compared to the other tools, specially in sensitivity and accuracy.

	Sensitivity	Specificity	Accuracy	Classification Rate	Error Rate
SCINA	0.347	0.963	0.655	0.826	0.479
CHETAH	0.471	0.986	0.728	0.600	0.129
scMAP	0.101	1.000	0.550	0.107	0.006
scID	0.521	0.971	0.746	0.700	0.179
Seurat	0.544	0.957	0.751	1.000	0.456
deepSCore	0.554	0.969	0.761	0.919	0.365

Table 1. Performance scores of different single-cell classifiers performing a cross-modality classification using the single-cell pancreas multiome dataset.

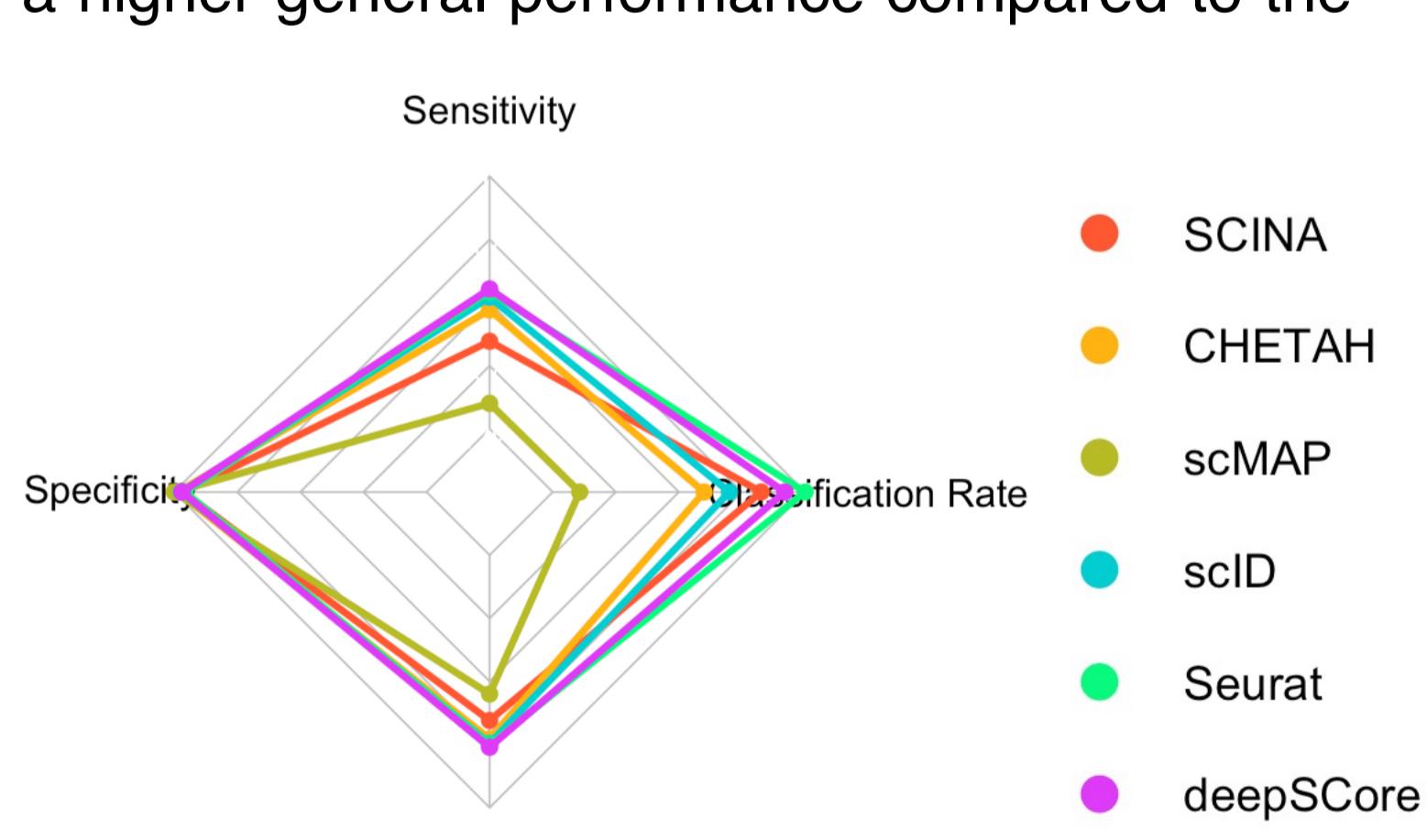


Figure 3. Radarchart of the performance of different single-cell classifiers.

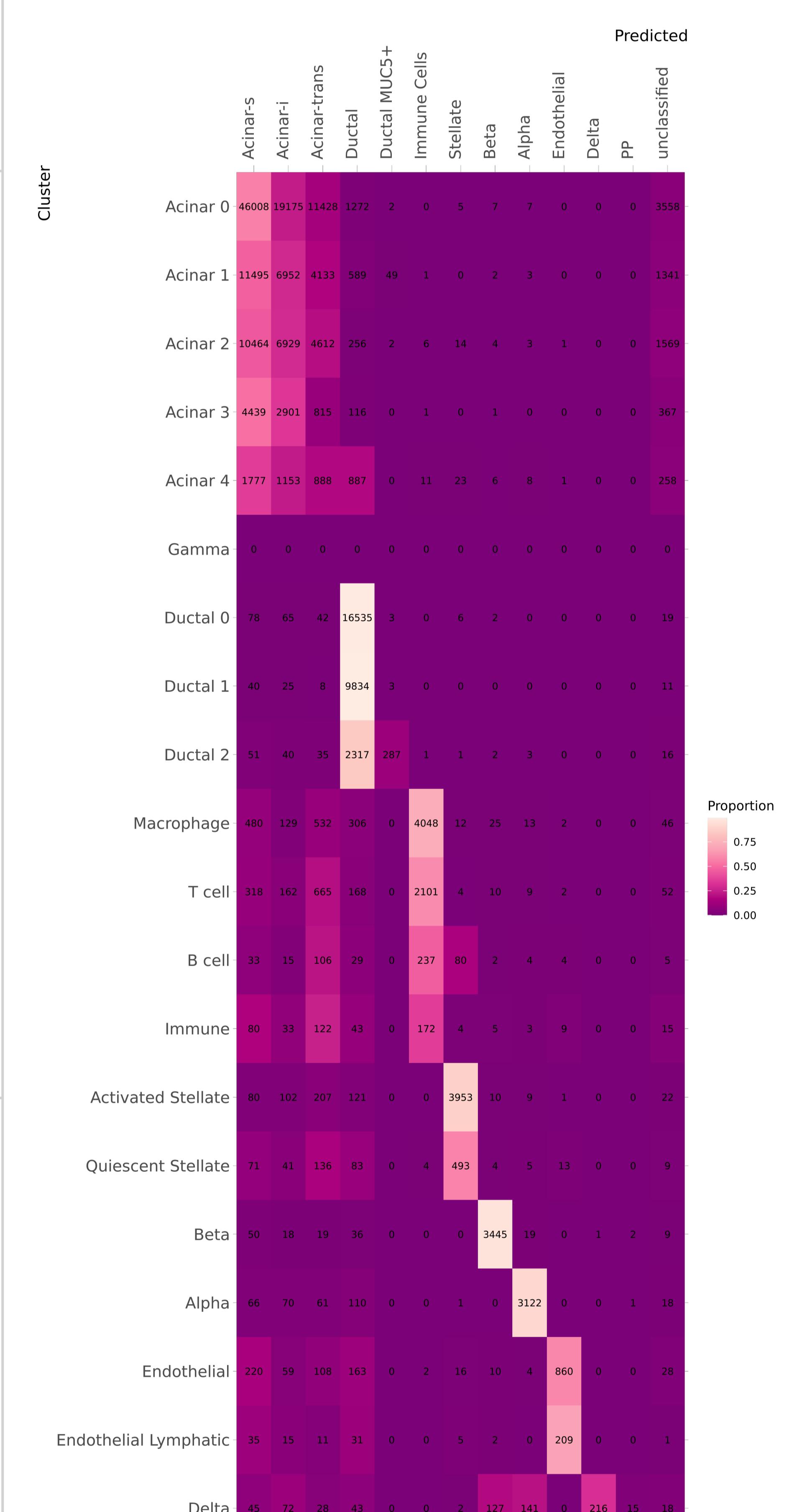


Figure 2. Performance of deepSCore when classifying pancreas' cells. Shown in presentation of Gonzales et.al.

## 4 FUTURE STEPS

In conclusion, deepSCore emerges as a highly versatile tool, allowing users to tailor parameters to their need and effectively handling distinct single-cell modalities, thereby exhibiting performance across challenging substrates.

Ongoing efforts involve further testing in diverse biological scenarios like distinguishing between healthy and pathological cells, searching the classification limits of the tool.

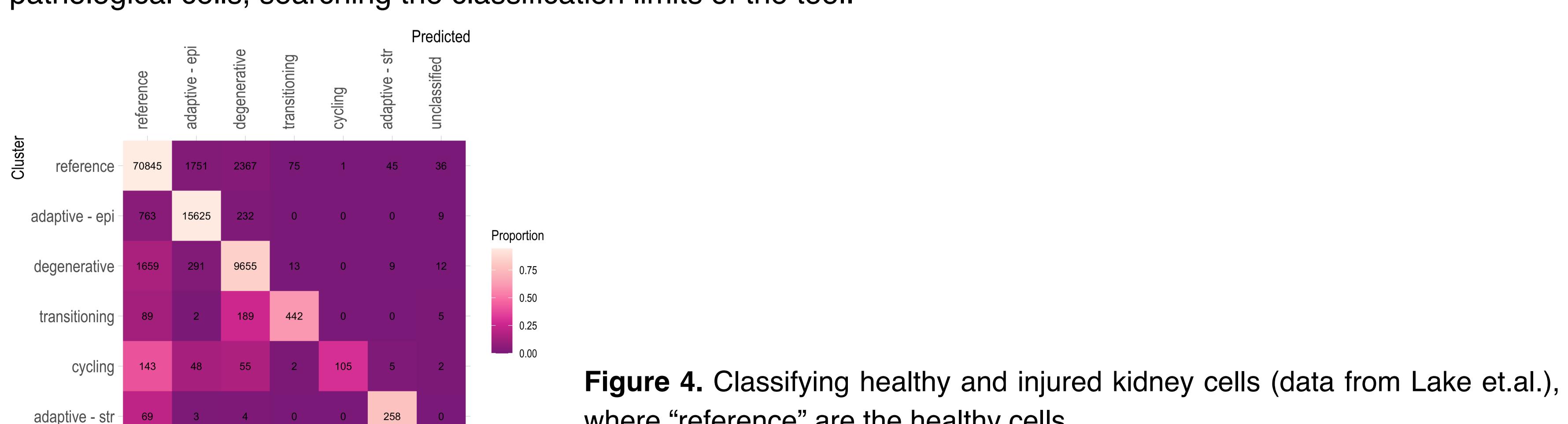


Figure 4. Classifying healthy and injured kidney cells (data from Lake et.al.), where “reference” are the healthy cells.

## REFERENCES

- Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* **24**, 550–572 (2023). <https://doi.org/10.1038/s41576-023-00586-w>
- Lake, B.B., Menon, R., Winfree, S. *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *Nature* **619**, 585–594 (2023). <https://doi.org/10.1038/s41586-023-05769-3>
- Wang, G., Chiou, J., Zeng, C. *et al.* Integrating genetics with single-cell multiomic measurements across disease states identifies mechanisms of beta cell dysfunction in type 2 diabetes. *Nat Genet* **55**, 984–994 (2023). <https://doi.org/10.1038/s41588-023-01397-9>