

CiteFuse enables multi-modal analysis of CITE-seq data

Hani Jieun Kim^{1,2,3,**}, Yingxin Lin^{1,2,**}, Thomas A. Geddes⁴, Jean Yang^{1,2}, and Pengyi Yang^{1,2,3*}

¹School of Mathematics and Statistics, The University of Sydney, NSW, Australia

¹Charles Perkins Centre, The University of Sydney, NSW, Australia

²Computational Systems Biology Group, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, NSW, Australia

⁴School of Life and Environmental Sciences, The University of Sydney, NSW, Australia

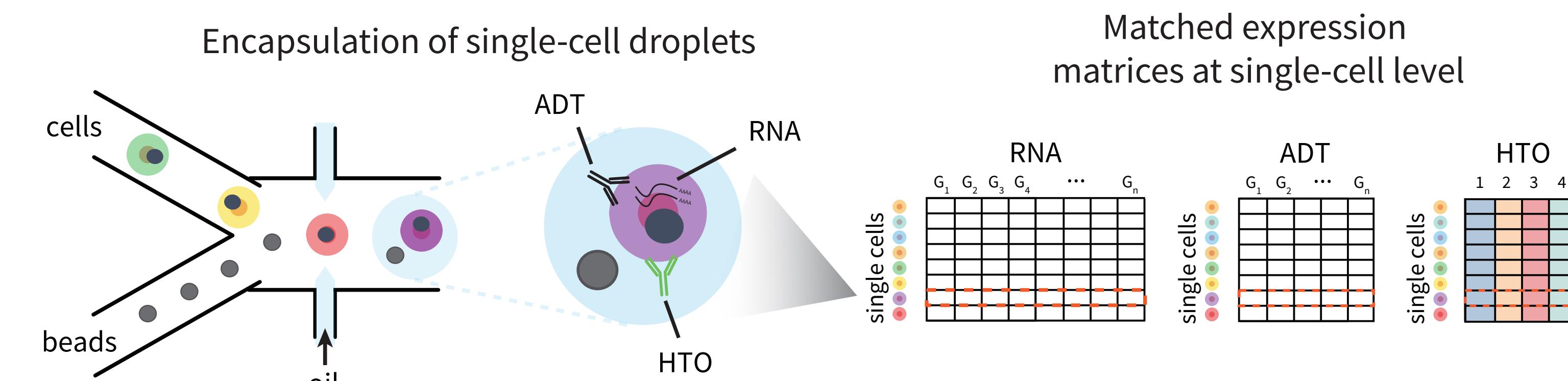
*Corresponding author (pengyi.yang@sydney.edu.au)

**These authors contributed equally

The dawn of single-cell multi-omics

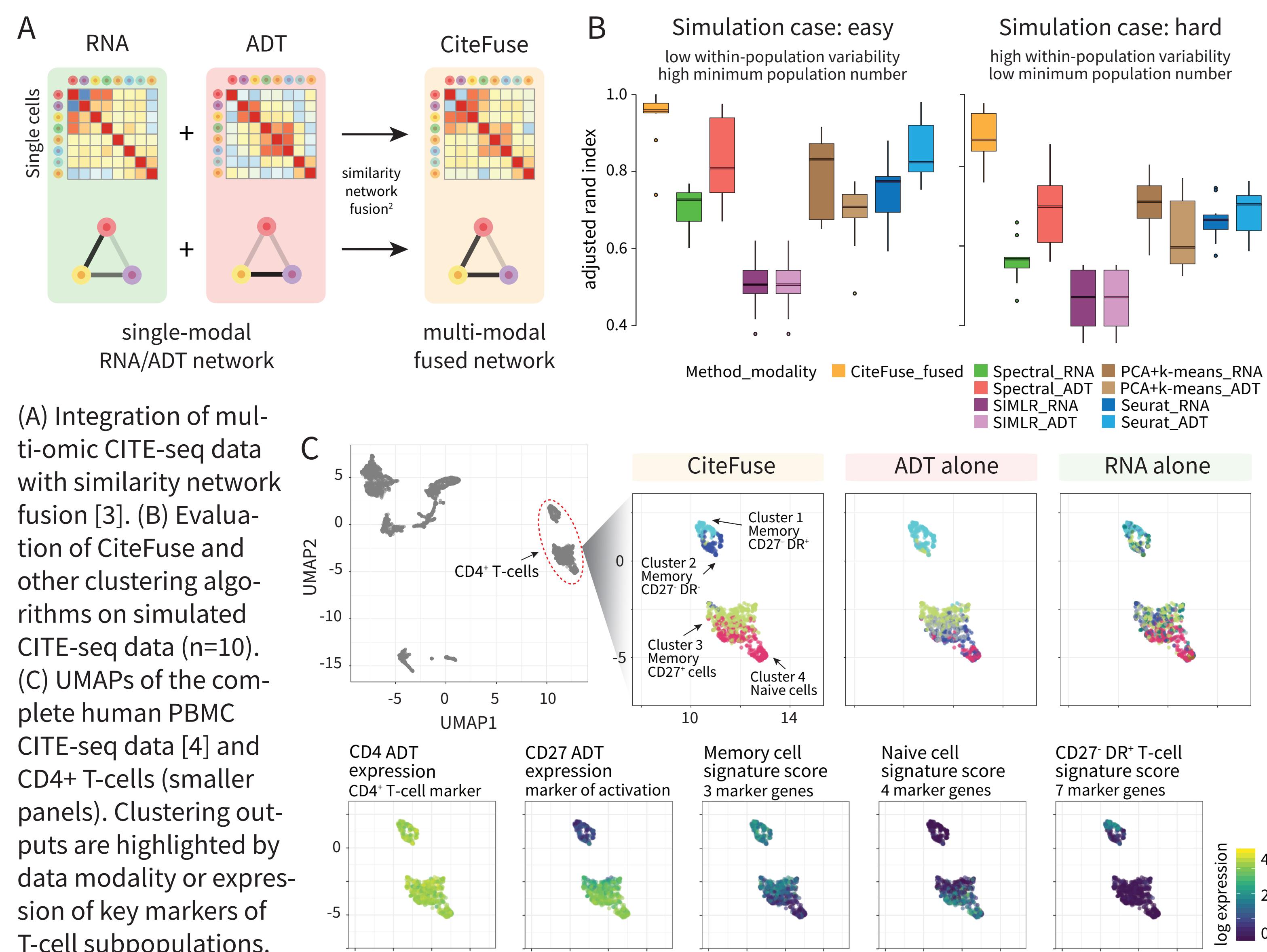
Single-cell technologies have transformed our understanding of cellular diversity and state by enabling high-throughput profiling of diverse information from individual cells [1]. Most of the technologies have focused on the measurement of a single modality. With fast-evolving technologies now measuring joint information from single cells, what more can we learn from multi-modal data?

CITE-seq generates multi-modal data from the same cell

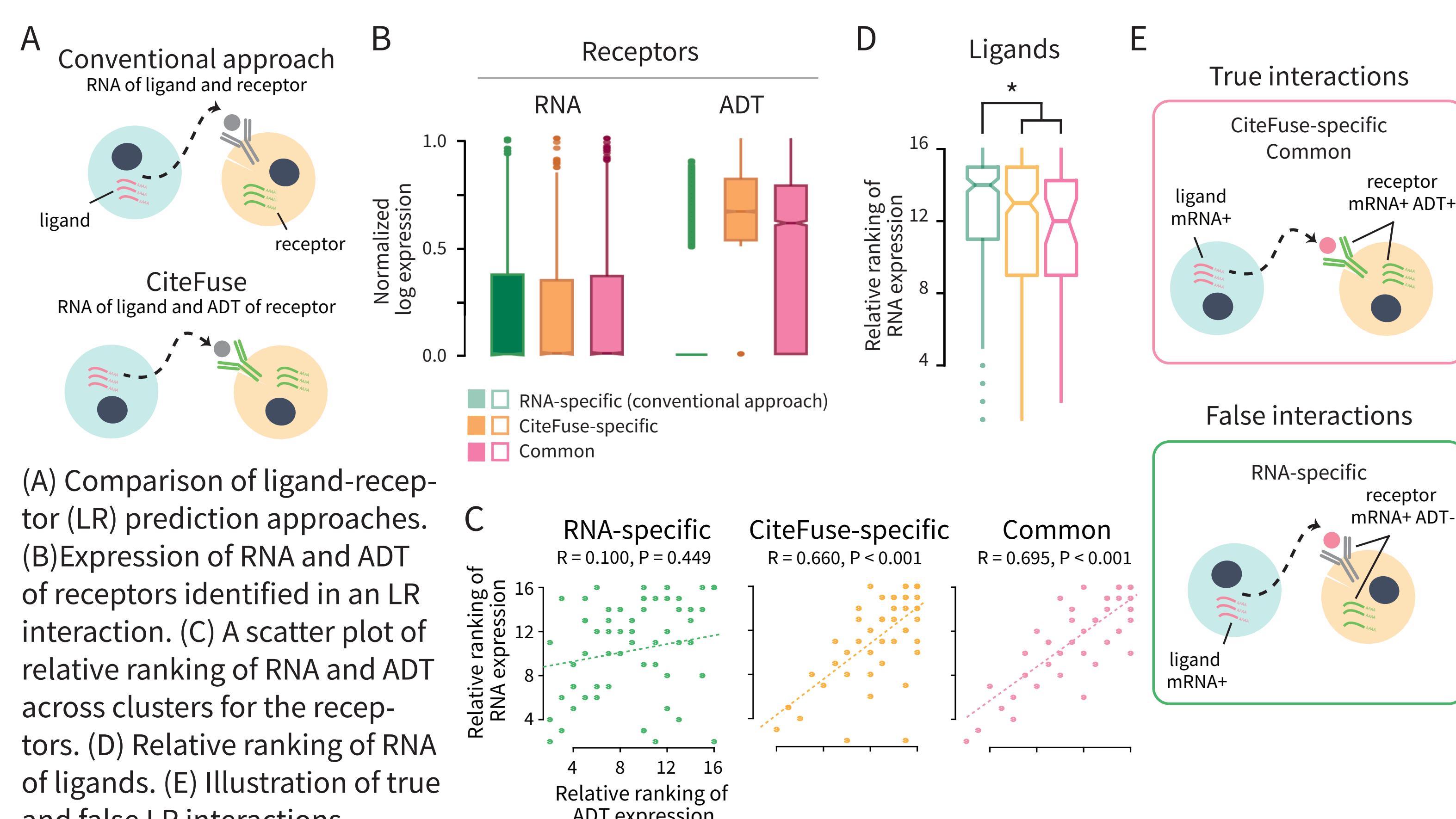


(A) A schematic of the micro-fluidic-based CITE-seq experiment [2] and the generation of matched RNA, antibody-derived tags (ADT), and hashtag oligonucleotide (HTO) expression matrices where rows indicate single cells and columns indicate RNA, ADTs, or HTOs, respectively.

CiteFuse gains information from multi-modal analysis



CiteFuse facilitates prediction of accurate ligand-receptor interactions

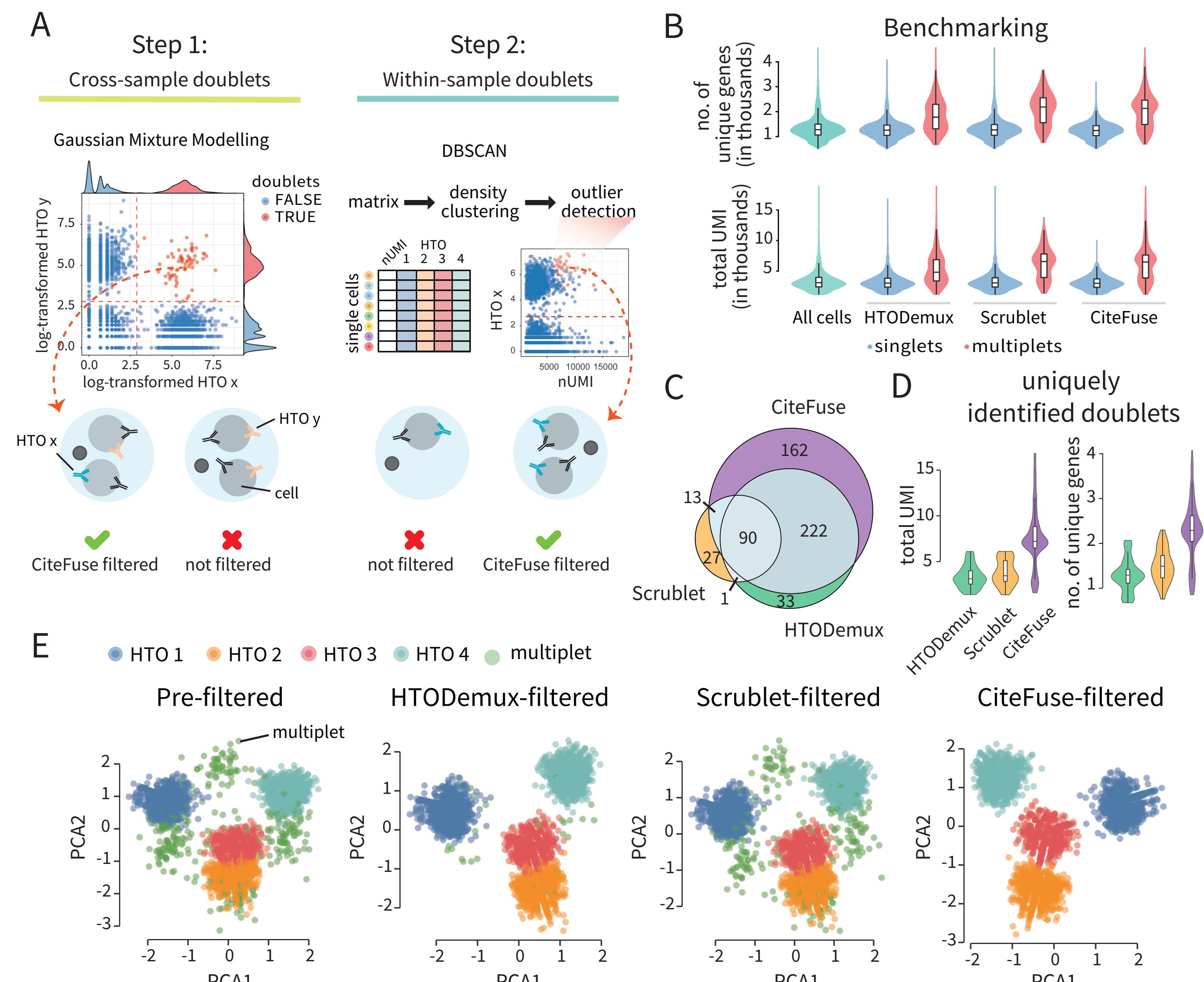


Acknowledgements and References

- [1] Stuart and Satija (2019) *Nature Review Genetics* [2] Stoeckius et al. (2017) *Nature Methods* [3] Wang et al. (2014) *Nature Methods* [4] Mimitou et al. (2019) *Nature Methods* [5] Stoeckius et al. (2019) *Genome Biology* [6] Wolock et al. (2019) *Cell Systems*

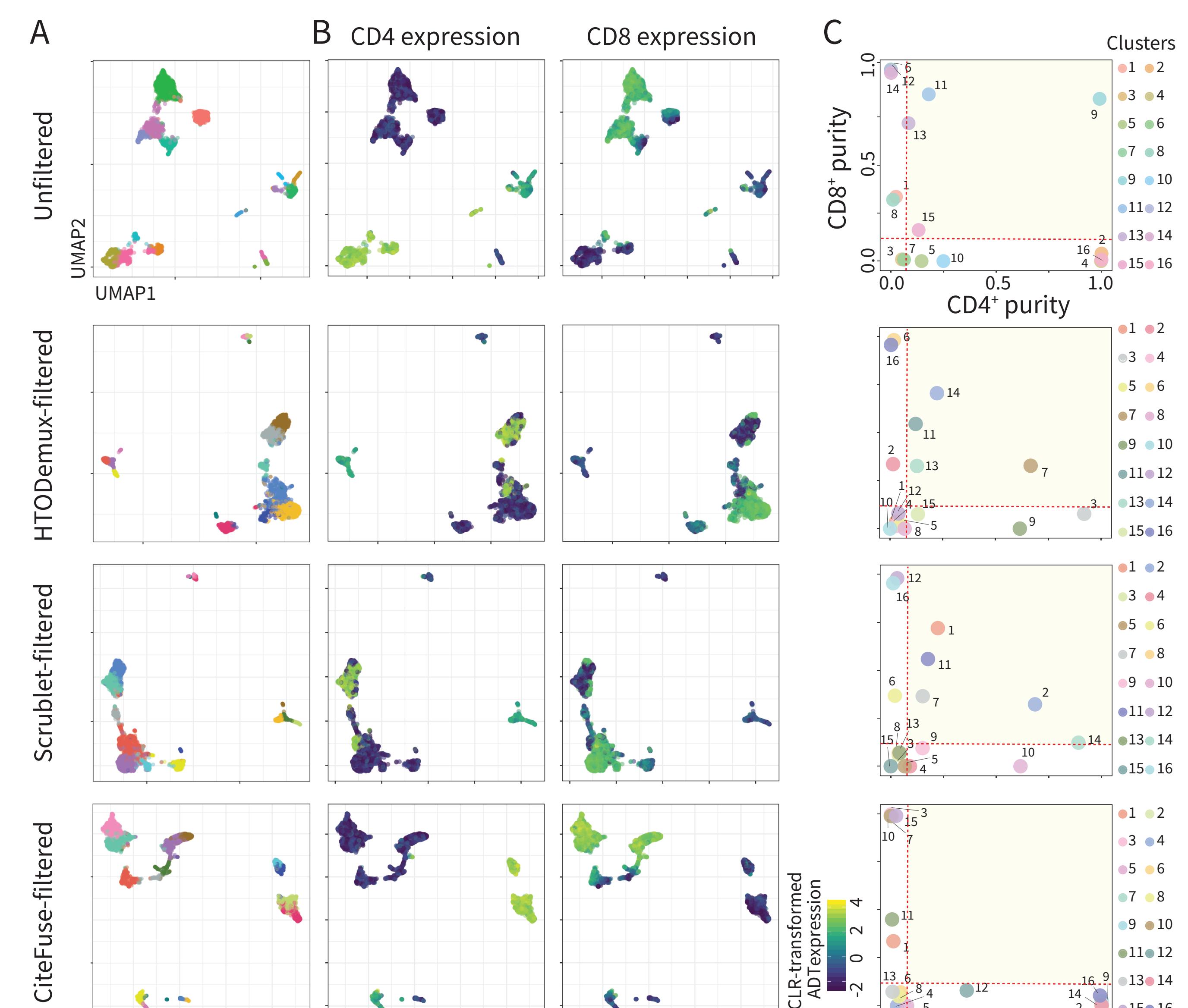
This work was supported by the Australian Research Council Discovery Project grant (DP170100654) to PY and JY; Discovery Early Career Researcher Award (DE170100759) to PY; Australia NHMRC Career Developmental Fellowship (APP1111338) to JY. Research Training Program Tuition Fee Offset and Stipend Scholarship to HK and YL. Chen Family Research Scholarship to YL. Children's Medical Research Institute Scholarship to HK.

A stepwise doublet detection method filters both within- and cross-sample doublets



(A) An overview of the doublet detection approach in CiteFuse. (B) Total number of unique molecular identifiers (UMIs) and number of unique genes expressed in all cells and in HTODemux- [5], Scrublet- [6] and CiteFuse-identified singlets and doublets. (C) Venn diagram depicting the overlap in predicted doublets. (D) Evaluation of doublets uniquely identified by each method. (E) PCA visualization of HTO expression before and after filtering doublets.

CiteFuse-filtered dataset best preserves the separation between CD4⁺ and CD8⁺ T-cells



(A) UMAPs of the unfiltered (4292 cells), HTODemux- (3753 cells), Scrublet- (3968 cells), and CiteFuse-filtered (3612 cells) human PBMC datasets. CLR-transformed ADT expression of (B) CD4 and (C) CD8 expression. (C) Plots of purity scores of CD8+ cells (y-axis) against CD4+ cells (x-axis).

Take home messages for CiteFuse

- CiteFuse gains information from multi-modal integration of CITE-seq data
- Doublet detection approach in CiteFuse effectively removes both within- and cross-sample doublets
- CiteFuse facilitates accurate ligand-receptor interaction prediction



THE UNIVERSITY OF
SYDNEY

<https://bioRxiv.org/cgi/content/short/854299v1>