



# A Novel Technology for Single Molecule Protein Sequencing

Swathi Ram<sup>1,2,3</sup>, Andy Mendoza<sup>1,2</sup>, Srivathsan Ranganthan<sup>1,2</sup>

Affiliations: 1- Cancer Early Detection Advanced Research (CEDAR); 2 – Knight Cancer Institute (KCI); 3 – Sunset High School



Partnership  
for  
Scientific  
Inquiry

Building tomorrow's researchers today

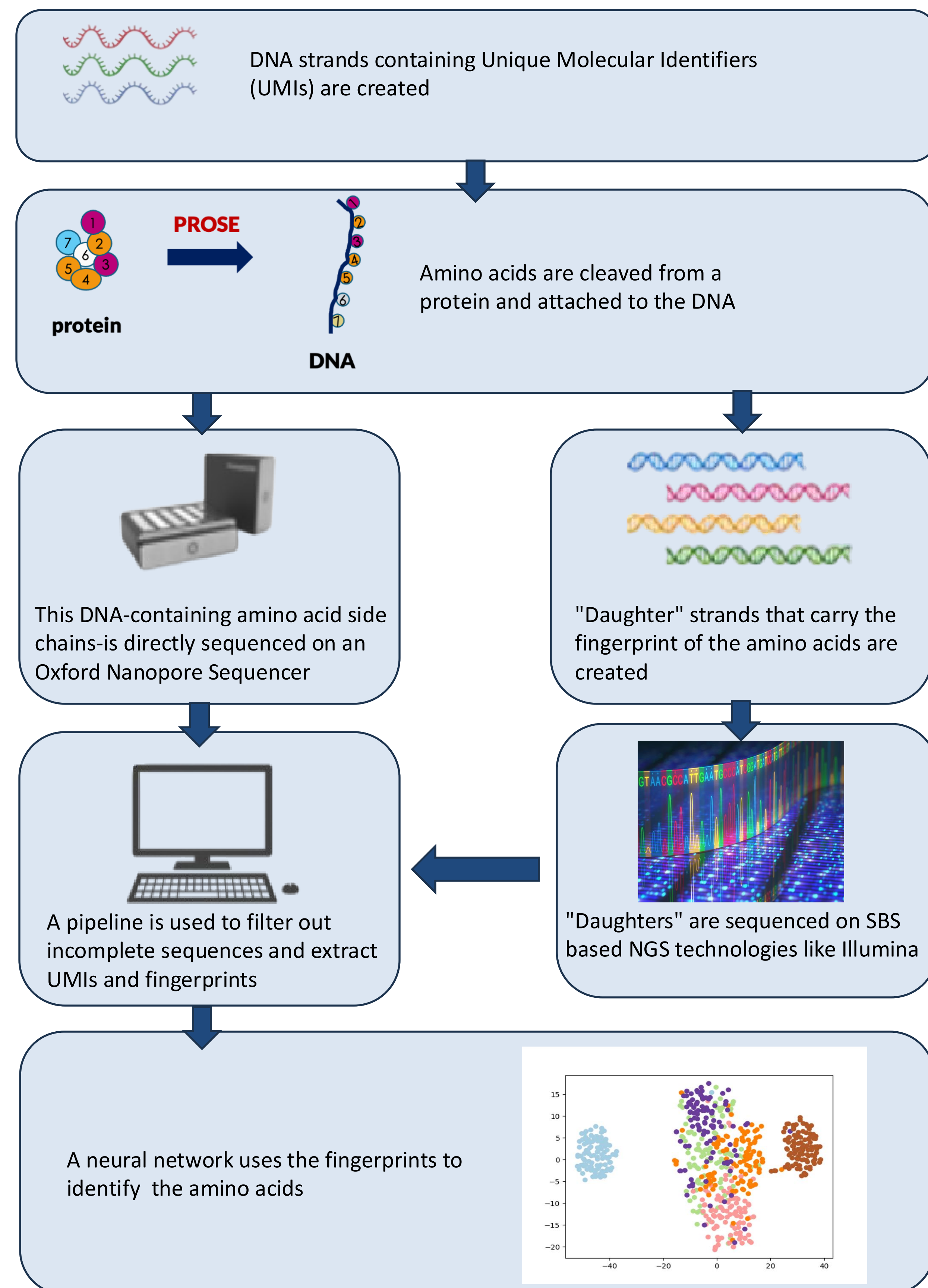
## Background

**Problem:** Accurate representation of protein expression is crucial for understanding the molecular mechanisms behind biological processes, and identification of novel biomarkers for diagnostics and therapeutics. Next-Generation Sequencing (NGS) of DNA and RNA, has made large strides towards that goal, but fails to account for **translational regulation**, **post translational modifications**, and **protein isoforms**, that can amount to millions of protein variants.<sup>1</sup> Current NGS platforms are not readily transferrable to direct protein sequencing.

**Current landscape:** Mass Spectrometry (MS) is the gold-standard for direct quantification of proteins. While MS has improved rapidly in the recent years, unlike NGS, it is not single-molecule and often relies on existing databases.<sup>2</sup>

**Solution:** This project aims to develop a **de novo single molecule protein sequencing** method by using a combination of biochemical processes that allow for fingerprinting of amino acids on a DNA molecule, which can then be sequenced on next-generation sequencing (NGS) platforms.

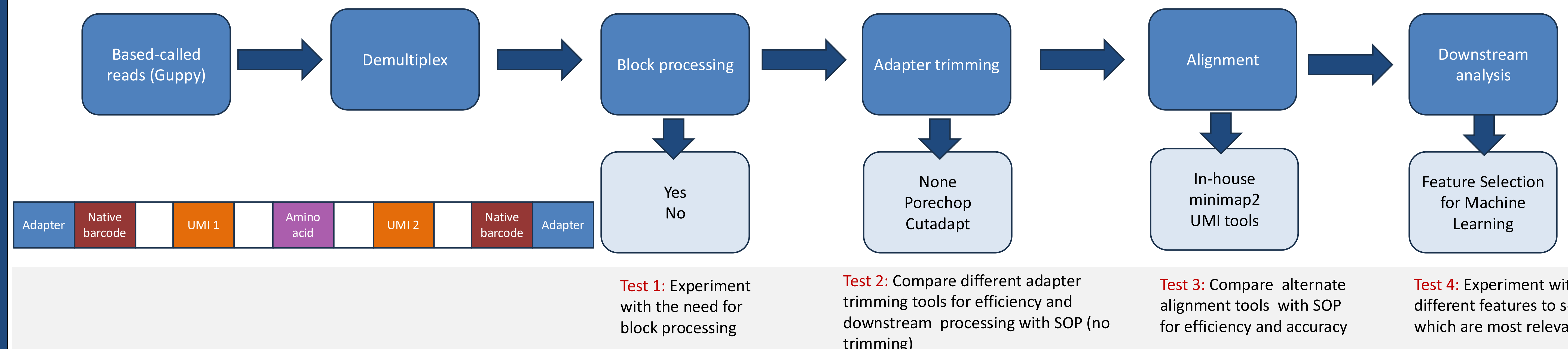
## Workflow



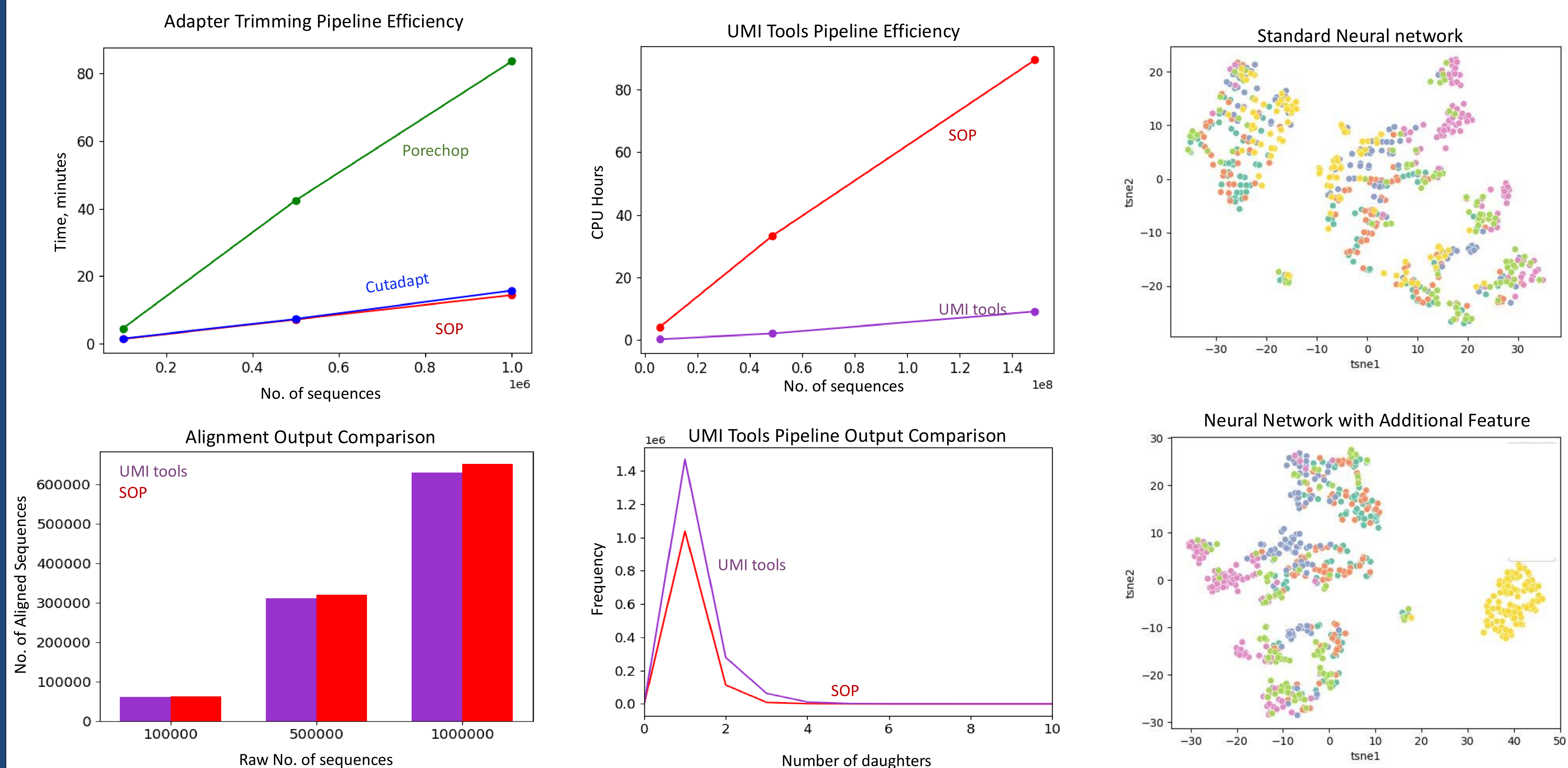
## Objectives

- To increase the efficiency of the pipeline (in terms of time and resources)
- To improve the accuracy of the neural network

## Methodology



## Results



## Conclusions and Future Directions

- Adapter trimming did not improve overall efficiency of the pipeline
- The new pipeline incorporating UMI tools was approximately 10 times faster than the SOP
- We significantly improved the accuracy of the neural network by identifying and adding a key feature to the dataset
- Next steps: Refine the biochemical process that controls the number of daughters created so that more daughters are created per mother

## Acknowledgments

- I would like to thank the PSI program managers and coordinators for this opportunity.
- I would like to thank Sofia Vignolo for guiding me through this process.
- I would like to thank the CEDAR and KCI staff for providing me with support and resources.

## References

- Alfaro, Javier Antonio, et al. "The emerging landscape of single-molecule protein sequencing technologies." *Nature methods* 18.6 (2021): 604-617.
- "Protein Sequencing: Methods and Applications | Abcam." *Abcam.com*, 2023, <https://www.abcam.com/en-us/knowledge-center/proteins-and-protein-analysis/protein-sequencing>