With great Parsl comes great portability:

Using Parsl through **CytoTable** *for harmonizing single-cell data*

CytoTable + Parsl = ➡️⚙️
❤️

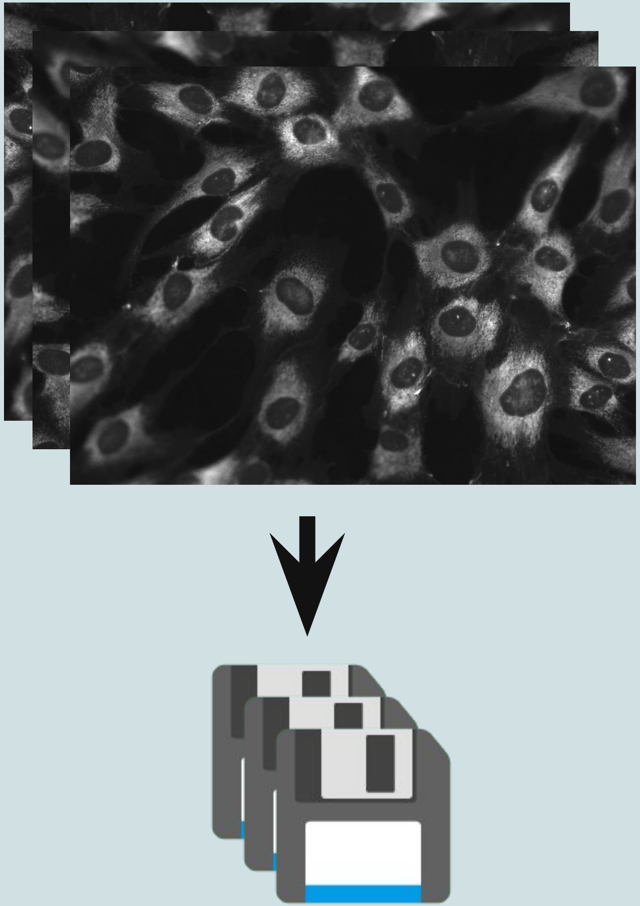
Introduction

- **Dave Bunten**
  Principal Research
  Software Engineer

- **Department of Biomedical Informatics**
  University of Colorado
  Anschutz Medical Campus
  School of Medicine

- **Way Lab (Gregory P. Way)**

Department of Biomedical Informatics
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**



Background

- *Bioinformatics:* **Image-based profiling**

- **Images of cells to numeric data (1000's of features)**

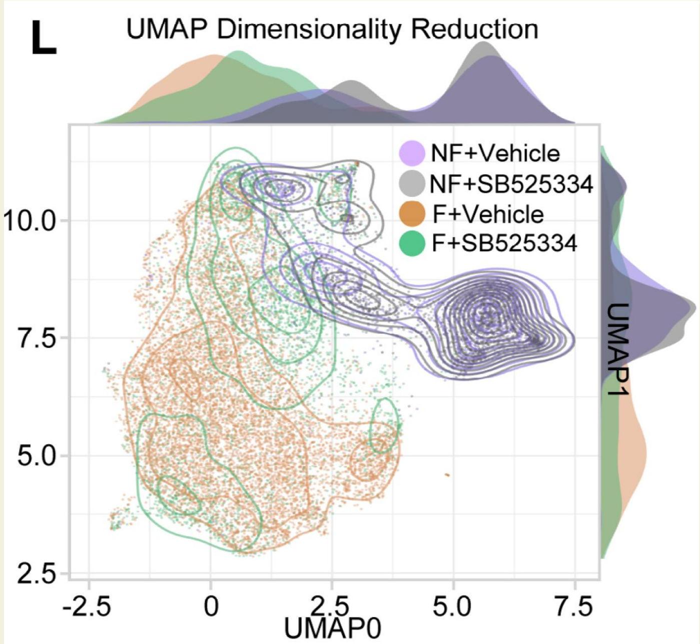
Images: Tomkinson et al., Plate 2 (Cell Painting images from Plate 2 for NF1_cellpainting_data project), 2023.

Why?

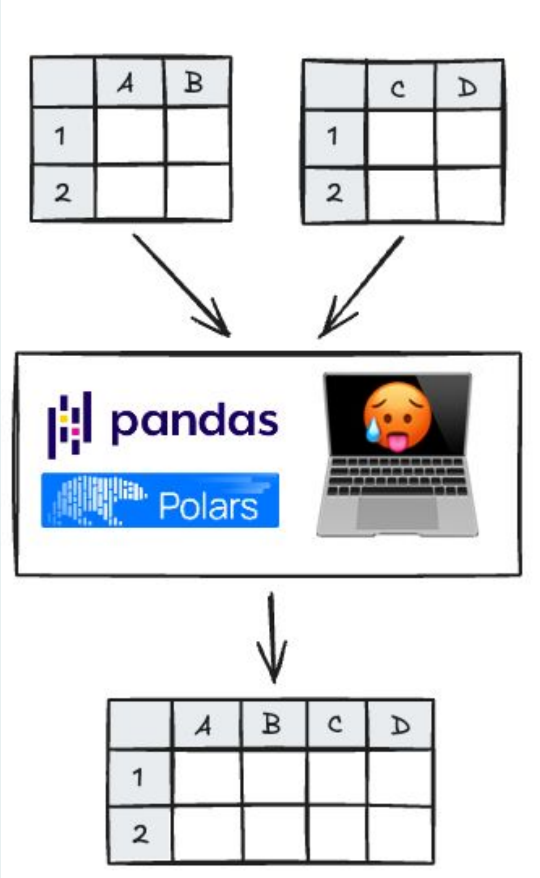
- Data derived from images help us understand **biological phenomena**.

- Information is used to help **cure diseases** and **improve lives**.

- Images are **cost-effective** to produce.



(*Travers et al., 2025, Cell Painting and Machine Learning Distinguish Fibroblasts From Nonfailing and Failing Human Hearts*)

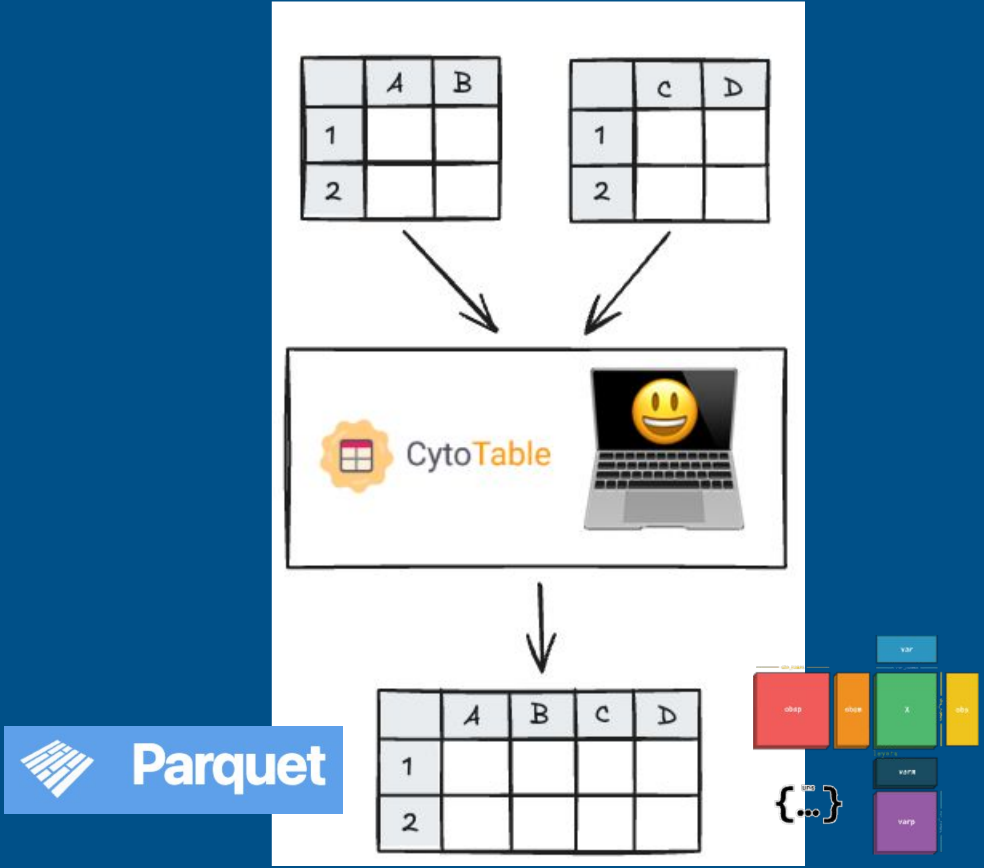


Challenge

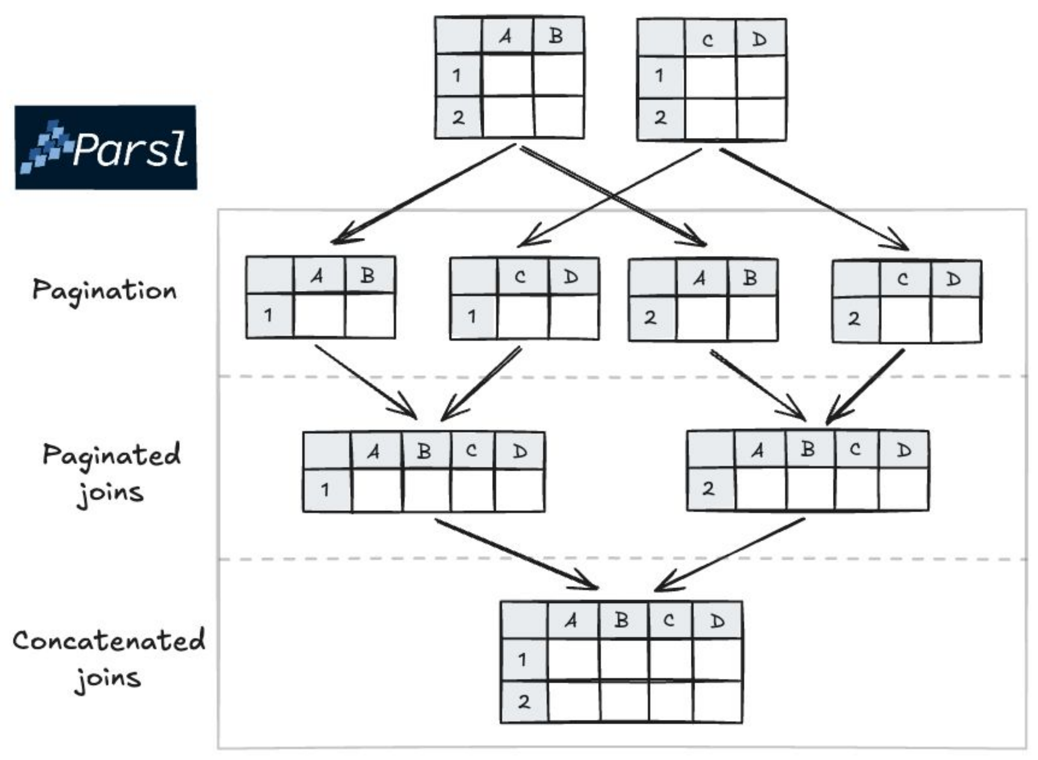
- Merging data is **expensive** and **expressed differently** in DataFrame libraries.
- Single-cell feature data entail **many different schema and file formats** from various ecosystems and microscopy products.

Solution

- **CytoTable** addresses these challenges through scalable **data harmonization**.
- We provide **Parquet** or **AnnData** output.



CytoTable + Parsl



- CytoTable uses Parsl to orchestrate **paginated data harmonization.**
- We implement map-reduce through **paginated maps and concatenated reduction**.

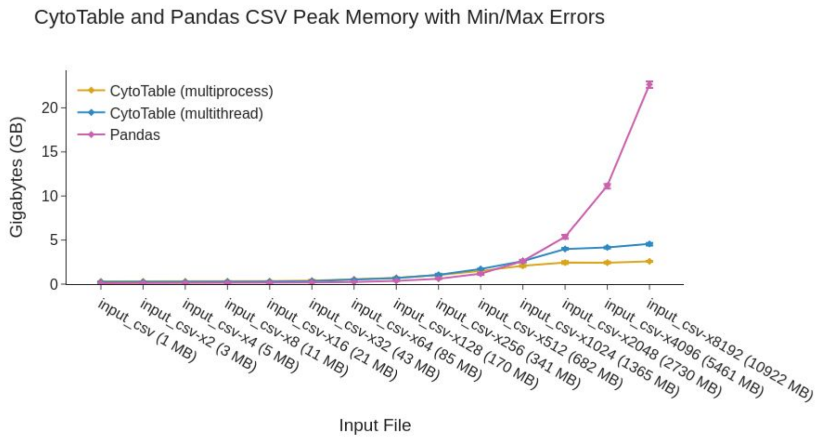
Portability + scale (!)

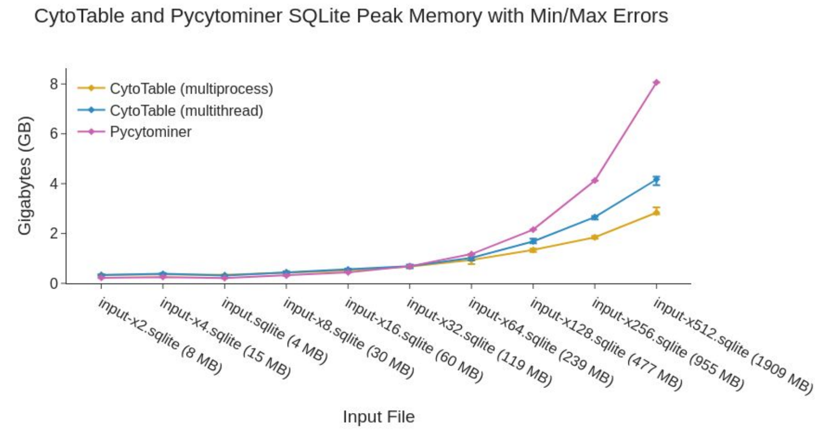
- CytoTable is managed using PyPI packages, including Parsl.

- Parsl config within CytoTable includes defaults with flexibility for overrides.

- Parsl is compatible with HPC, reducing compute barriers.

Solutions at Scale





- We find that CytoTable **enables scalable memory and time performance** when compared to existing methods.

Thank you!
Questions/comments?

Find out more below!


bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

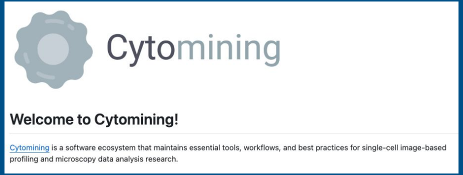
New Results

Scalable data harmonization for single-cell image-based profiling with CytoTable

Dave Bunten, Jenna Tomkinson, Erik Serrano, Michael J. Lippincott, Kenneth I. Brewer, Vince Rubinetti, Faisal Alquaddoomi, Gregory P. Way

doi: https://doi.org/10.1101/2025.06.19.660613

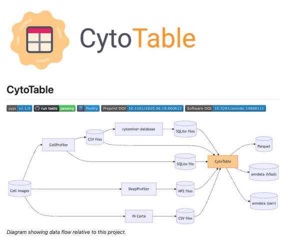
Preprint:
https://www.biorxiv.org/content/10.1101/2025.06.19.660613v1


Cytomining

Welcome to Cytomining!

Cytomining is a software ecosystem that maintains essential tools, workflows, and best practices for single-cell image-based profiling and microscopy data analysis research.

GitHub organization:
https://github.com/cytomining


CytoTable

CytoTable

Diagram showing data flow relative to this project.

GitHub repository:
https://github.com/cytomining/CytoTable