

Abstract

Due to their enormous potential for advancing drug discovery, there continues to be an exponential growth in the use of single cell sequencing methods, and a corresponding increase in datasets in publicly available repositories. While these datasets are freely available, they come with **hidden costs** that hinder the ability of companies to exploit them to their maximum potential. These costs typically result from a **lack of metadata standards** and **significant variation in the processing** approach.

The Single Cell Data Science (SCDS) Consortium was formed in 2022 with four Charter Members (3 large Pharma and 1 Biotech) as a multi-year effort to harmonize single cell experiments more quickly and cost effectively. This **pre-competitive organization is being led by Rancho BioSciences**, with expertise in single cell data curation, processing, and analysis. To date, SCDS has successfully delivered 115 high-quality datasets with metadata harmonized to a 4 entity, 79 attribute data model.

In 2023 the consortium has increased to six member companies and added several defined functions to the scope. Updates to the ingestion pipeline to adapt to these changing needs is currently in progress and seeks to increase both the processing capacity and features provided to analysts. Tissue, disease and organ-specific reference atlases will also be constructed. **Curated datasets delivered as part of this consortium are already accelerating reproducible science, rapid discovery, and joint analysis of valuable public data.**