

The Single Cell Data Alliance brings industry collaboration to single-cell research—delivering expertly curated and analyzed public datasets faster and more cost effectively than any one organization could alone.



Why it matters

Single-cell technology enables researchers to examine biological systems at an unprecedented resolution. While traditional bulk sequencing methods mask the natural heterogeneity within cell populations, single-cell technologies reveal distinct cell subtypes, rare cell populations, and transitional cellular states that are often critically involved in disease progression. In drug development, this granularity allows scientists to identify highly specific therapeutic targets, understand mechanisms of drug resistance at the cellular level, and predict off-target effects by mapping how different cell types respond to a compound.

Why Single Cell Data Alliance

While vast public datasets of single-cell data exist that can be used to further pharmaceutical research, the tasks of finding, acquiring, harmonizing, and curating single-cell data can be labor intensive and time consuming.

WHAT WE'VE ACHIEVED TOGETHER

Over the past several years, the Alliance, made up of the world's top pharma companies, has delivered:

115M+

reprocessed and
annotated single cells

38k+

samples from healthy donors,
470+ diseases & 670+ data cell types

14+

Disease and tissue
atlases delivered

The number that matters most isn't the cells—it's the consistency across all of them.

Consortium members:



REGENERON



Johnson&Johnson



Rapid acceleration of curation with technology

As a leader in advanced data science and bioinformatics analysis, our team continues to optimize the process of dataset quantification, high-resolution annotation, and harmonized metadata curation. Cloud-enabled pipelines provide reproducible data quantification. New modalities of annotation are incorporated in high-interest areas such as oncology, multimodal data, and model organisms. Agentic curation tools are rapidly becoming an essential part of our process, while continuing to optimize value of our subject matter expert in-the-loop systems. These technology improvements continue to increase the ROI for our alliance members, while maintaining gold-standard quality expected from Rancho.

BENEFITS OF MEMBERSHIP

New members receive all previous deliverables upon joining and will be able to participate in prioritization of future batches.

Mining public data



9000+ public single cell datasets available. Members choose the most relevant studies.

Flexible data delivery



Ready-to-use formats, such as, h5ad, Seurat, and TileDb objects. Data optimized for downstream analysis and integration.

Rich harmonized metadata



Comprehensive study, dataset, donor, sample, cell-level annotations, version-controlled ontologies.

Cell type atlases



Initiated by member-driven interest, new atlases focus on evolving disease models, updated tooling, and faster integration.

Advanced processing pipeline



Rancho handles end-to-end processing providing uniform mapping and quantification across datasets, standardized QC, and consensus cell-type annotation, and multi-modal data support.

Superior ROI with no subscription



In this true cost sharing approach, ROI continues to grow as new members join. Members keep all deliverables in perpetuity with no further fees due.

Community collaboration



Members get to enjoy the ability to interact regularly with like-minded peers in drug discovery in a pre-competitive manner, to discuss common challenges, share stories of practical work, discuss lessons learned, and demonstrate relevant methods that help advance the field together and deliver value to each member.

Nine cell-type reference atlases

- Hematologic Malignancie 610k cells, 12 datasets
- Hidradenitis Suppurativa 430k cells, 10 datasets
- General PBMC 1.07M cells, 26 datasets
- Kidney Diseases 210k cells, 6 datasets
- Cardiovascular Diseases 780k cells, 10 datasets



MEMBERSHIP

New members are welcome to join the Alliance.

Rapid acceleration of curation with technology

As a leader in advanced data science and bioinformatics analysis, our team continues to optimize the process of dataset quantification, high-resolution annotation, and harmonized metadata curation. Cloud-enabled pipelines provide reproducible data quantification. New modalities of annotation are incorporated in high-interest areas such as oncology, multimodal data, and model organisms. Agentic curation tools are rapidly becoming an essential part of our process, while continuing to optimize value of our subject matter expert in-the-loop systems. These technology improvements continue to increase the ROI for our alliance members, while maintaining gold-standard quality expected from Rancho.

BENEFITS OF MEMBERSHIP

New members receive all previous deliverables upon joining and will be able to participate in prioritization of future batches.

Mining public data



9000+ public single cell datasets available. Members choose the most relevant studies.

Flexible data delivery



Ready-to-use formats, such as, h5ad, Seurat, and TileDb objects. Data optimized for downstream analysis and integration.

Rich harmonized metadata



Comprehensive study, dataset, donor, sample, cell-level annotations, version-controlled ontologies.

Cell type atlases



Initiated by member-driven interest, new atlases focus on evolving disease models, updated tooling, and faster integration.

Advanced processing pipeline



Rancho handles end-to-end processing providing uniform mapping and quantification across datasets, standardized QC, and consensus cell-type annotation, and multi-modal data support.

Superior ROI with no subscription



In this true cost sharing approach, ROI continues to grow as new members join. Members keep all deliverables in perpetuity with no further fees due.

Community collaboration



Members get to enjoy the ability to interact regularly with like-minded peers in drug discovery in a pre-competitive manner, to discuss common challenges, share stories of practical work, discuss lessons learned, and demonstrate relevant methods that help advance the field together and deliver value to each member.

Nine cell-type reference atlases

- Hematologic Malignancies 610k cells, 12 datasets
- Hidradenitis Suppurativa 430k cells, 10 datasets
- General PBMC 1.07M cells, 26 datasets
- Kidney Diseases 210k cells, 6 datasets
- Cardiovascular Diseases 780k cells, 10 datasets



MEMBERSHIP

New members are welcome to join the Alliance.