SynEcoSys® - Bridging single cell data and clinical insights

SynEcoSys® is a comprehensive single cell knowledge base, integrating curated data from more than 60 millions single cells with clinical metadata. All datasets undergo uniform processing standards for data analysis and cell type annotation, ensuring unparalleled precision and comparability, enabling researchers to translate single cell data into clinical and drug discovery insights.

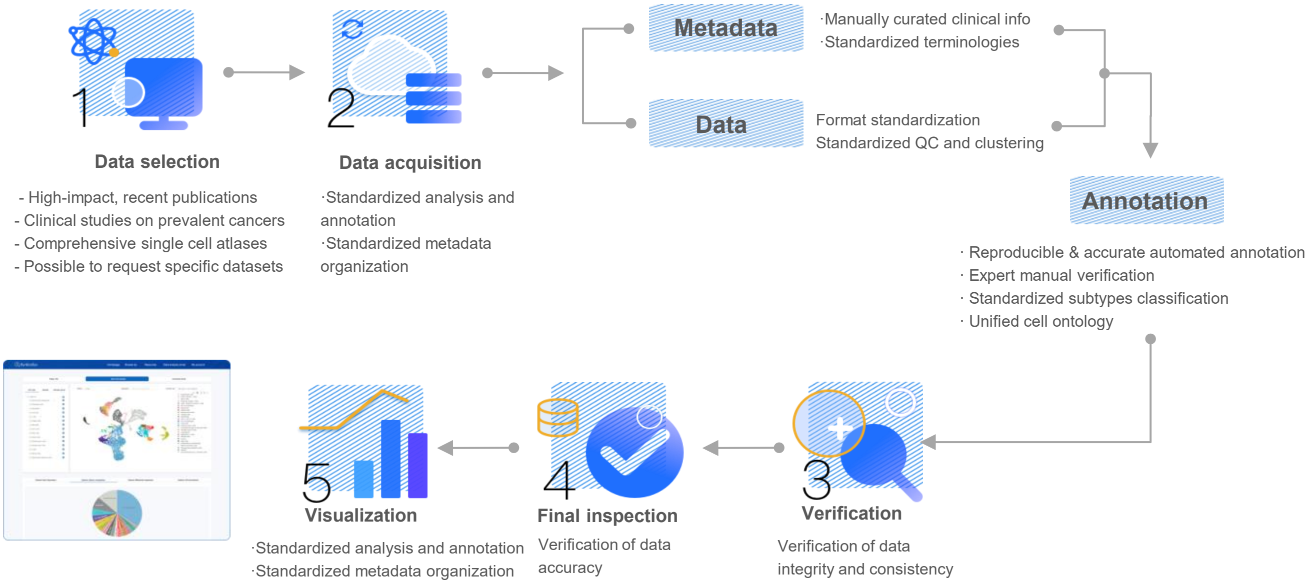
Highlights

- **Clinical relevance:** the first clinically-focused single cell database for translational research
- **Proprietary cell type markers references** curated by experienced scientists from literature
- **Core datasets for major diseases** for an in-depth dive into disease models

SynEcoSys harmonizes public single cell data for efficient data mining

SynEcoSys transforms public single cell data into a cohesive, mineable resource. This addresses the primary challenge in utilizing publicly available, clinically relevant single cell data: the lack of comparability due to diverse analysis methods, inconsistent cell type annotations, and the need for specialized bioinformatic expertise.

By integrating standardized, manually curated, high-impact single cell sequencing data with the CeleLens™ Cloud (page 3) analysis and visualization module, SynEcoSys enables robust mining of clinically relevant data to identify potential biomarkers and novel drug targets.



1

Data selection

- High-impact, recent publications
- Clinical studies on prevalent cancers
- Comprehensive single cell atlases
- Possible to request specific datasets

2

Data acquisition

·Standardized analysis and annotation
·Standardized metadata organization

Metadata

·Manually curated clinical info
·Standardized terminologies

Data

Format standardization
Standardized QC and clustering

Annotation

· Reproducible & accurate automated annotation
· Expert manual verification
· Standardized subtypes classification
· Unified cell ontology

3

Verification

Verification of data integrity and consistency

4

Final inspection

Verification of data accuracy

5

Visualization

·Standardized analysis and annotation
·Standardized metadata organization

Figure 1. Overview of standardized SynEcoSys data curation and presentation workflow.

Use Case 1

Geng, F. et al. identified thyroid adenoma as a potential risk factor for small cell lung cancer development in a retrospective study using SynEcoSys datasets from three different studies.

Impact

1) SynEcoSys standardized datasets with disease annotation were used to validate the potential association.

2) The UMAP plot and gene expression plots generated in SynEcoSys were presented in the publication.

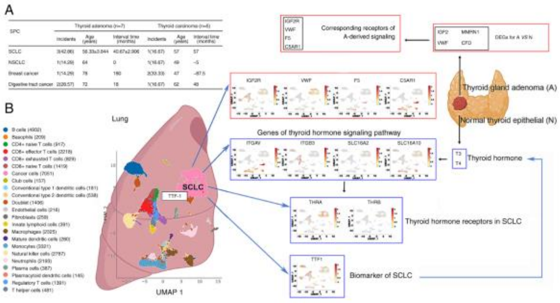


Figure 2. Potential association between thyroid gland adenoma and SCLC (Geng, F. et. al.) CC.BY

A comprehensive knowledge base with accurate cell annotation

The SynEcoSys database has standardized ontology, allowing for direct comparison between the different datasets.

An in-house transcriptome-based cell annotation reference was built through meticulous literature curation by experienced professionals. This cell annotation reference allows for accurate cell annotation to match the populations found in the publications or discover novel cellular subtypes.

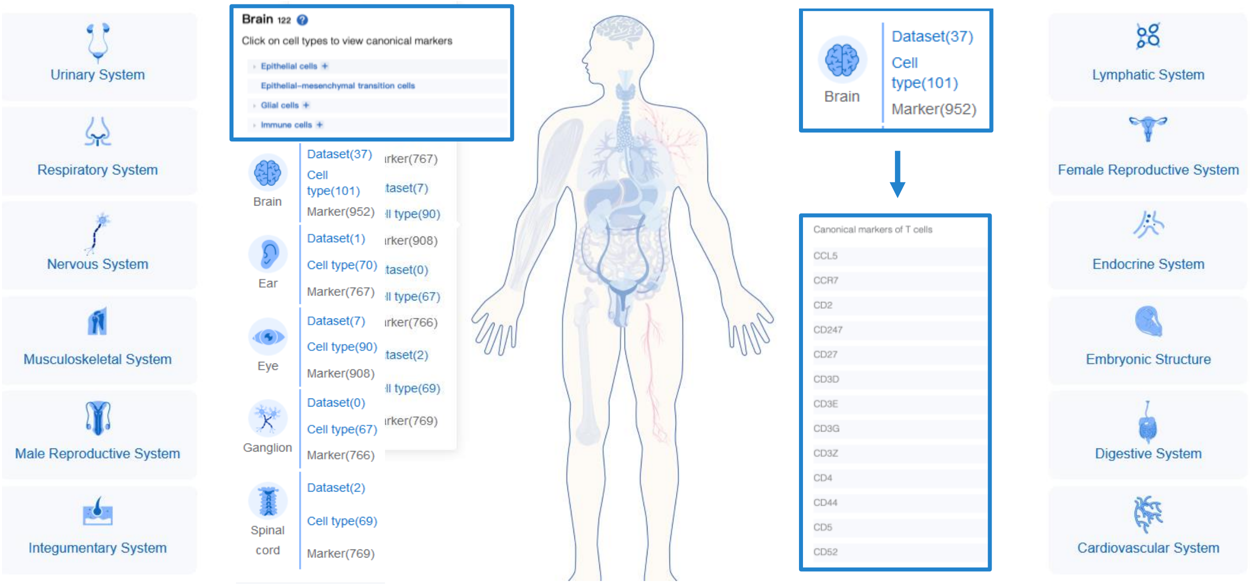


Figure 3. SynEcoSys database (online version) currently comprises 1,800+ single cell RNAseq datasets from 17,000+ samples; 600+ cell types; 8,000+ sets of marker genes, 130+ tissue types, and more than 60 million cells sequenced. In addition, SynEcoSys is regularly updated with new public data.

Use Case 2

CellFM – a large-scale foundation model pre-trained on transcriptomics of 100 million human cells

Leveraging SynEcoSys's vast data repository, Zeng, Y. et. al. developed CellFM, a powerful single cell foundation model. With 800 million parameters, it represents an eight-fold increase over existing single-species models.

Impact

CellFM's robust performance, built on SynEcoSys's standardized datasets, surpasses existing models in cell annotation, perturbation prediction, and gene function prediction.

References

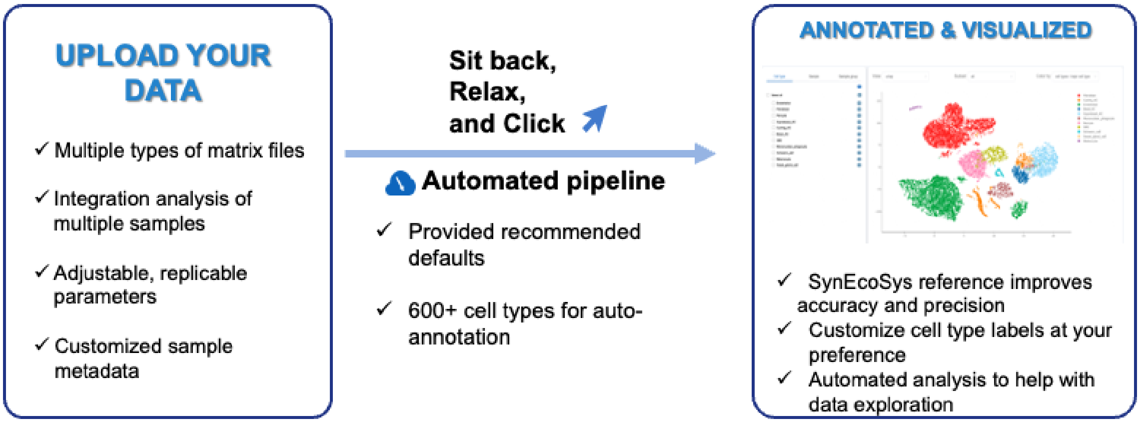
- Zhang, Y., Li, B., Duan, J., Cheng, X., Zhang, X., Ye, J. … Fang, N. (2023). SynEcoSys: a multifunctional platform of large-scale single-cell omics data analysis. bioRxiv 2023.02.14.528566; doi: https://doi.org/10.1101/2023.02.14.528566
- Geng, F., Liu, M., Chen, J., Ge, Y., Wei, S., Li, F. ... Zhang, J. (2023). Clinical characteristics of second primary malignancies among first primary malignancy survivors: A single-center study, 2005-2020. Oncology Letters, 25, 24. https://doi.org/10.3892/ol.2022.13610
- Zeng, Y., Xie, J., Wei, Z., Su, Y., Shangquan, N., Yang, S. … Yang, Y. (2024). CellFM: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. bioRxiv 2024.06.04.597369; doi: https://doi.org/10.1101/2024.06.04.597369

Celelens^{TM} Cloud: A code-free single cell data analysis & annotation tool

CeleLens Cloud offers intuitive, automated single cell analysis and annotation pipelines with adjustable parameters. Its visualization tools generate publication-quality figures and provide detailed methodological descriptions compatible with leading scientific journals.

Highlights

- **Versatile integration** of multi-format data matrices from diverse single-cell platforms, with advanced batch effect correction
- **Automated cell type annotation** using proprietary markers and Singleron pre-trained reference
- **Generation of publication-ready figures** with comprehensive analysis method descriptions



UPLOAD YOUR DATA

- ✓ Multiple types of matrix files
- ✓ Integration analysis of multiple samples
- ✓ Adjustable, replicable parameters
- ✓ Customized sample metadata

Sit back, Relax, and Click

Automated pipeline

- ✓ Provided recommended defaults
- ✓ 600+ cell types for auto-annotation

ANNOTATED & VISUALIZED

- ✓ SynEcoSys reference improves accuracy and precision
- ✓ Customize cell type labels at your preference
- ✓ Automated analysis to help with data exploration

Figure 4. Integrate data from multiple studies and annotate cell types easily with Celelens Cloud's built-in cell marker database.

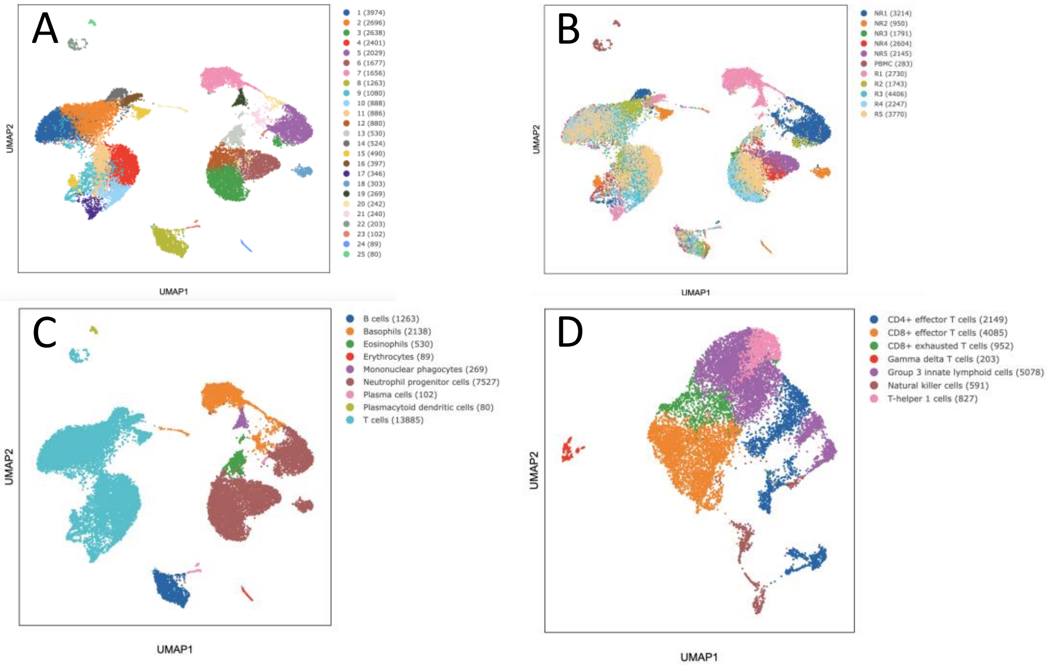


Figure 5. Overview of standardized data integration, automated annotation and sub-clustering of selected subset. (A) UMAP projection for clusters from multiple samples after standardized data integration (B) Colored by different samples (C) Automated annotations for multiple cell types from different samples (D) Subtypes of T cells from multiple samples

Perform data visualization and exploratory analysis easily with Celelens Cloud's user-friendly interface. No previous coding knowledge is required.

Celelens Cloud embeds commonly used single cell data inference pipelines, including:

- Gene expression
- Gene set scoring
- Cellular composition
- Differential expression
- Pathway enrichment
- Cell-cell interaction
- Trajectory analysis
- CNV/ITH
- TF regulation

In every module, the clinical metadata can be used to perform comparison analysis, in order to reveal the impact of disease conditions or treatment on transcriptome.

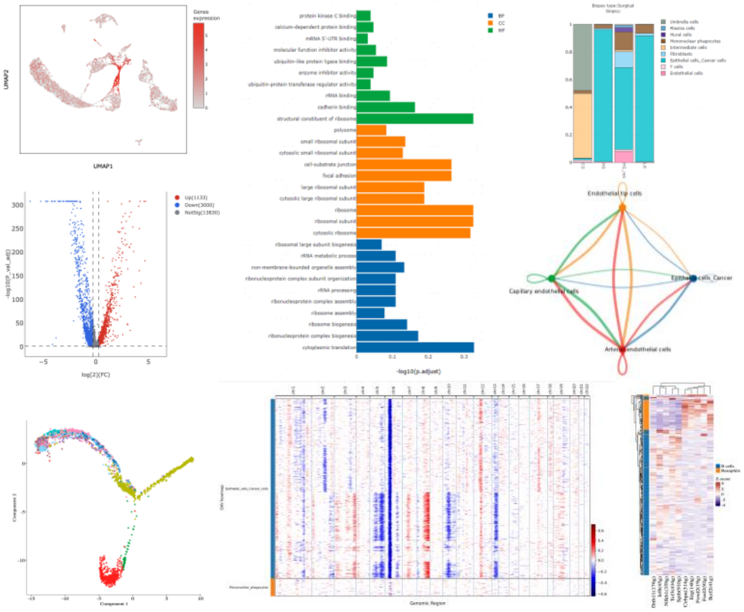


Figure 6. Celelens Cloud automated analysis can be used for exploration of user data and SynEcoSys database with variety of commonly used single cell data inference pipelines.

Ordering information

Product	Catalog Number
SynEcoSys® Clinical Single Cell Database (1 year subscription)	8270060
CeleLens^{TM} Cloud - Platform	PR005

Resources

- **Tutorial videos** https://www.youtube.com/channel/UCZeGRq83pQv4ebiy16Nc-Rw
- **User guide:** www.synecosys.com
- **Sign up and free trial inquiry:** https://celelenscloud.singleron.bio/#/login
- **Support:** sessupport@singleronbio.com

Singleron Biotechnologies GmbH
Phone: +49 (0) 221 16824777
Email: info@singleron.bio
Address: Gottfried-Hagen-Strasse 60, 51105 Cologne, Germany

Singleron Biotechnologies Pte. Ltd (UEN 201718919Z)
Phone: +65 6990 5987
Email: sales.sg@singleron.bio
Address: 61 Science Park Rd, the Galen 02-05 Singapore 117525