

Scale & Simplify Discovery with Single-Cell Omics

Unlock the power of all your
single-cell data for your next breakthrough

At a Glance

Large scale, high resolution multiomics data such as that generated by single-cell sequencing are becoming increasingly valuable in the drug discovery and development pipeline to find new targets and help progress fewer false positives through the pipeline. However, the field faces significant data management challenges due to the lack of standardization in storing omics data. Current toolkit-specific formats typically require loading the entire dataset into memory, which is increasingly infeasible as datasets grow in size. In addition, these formats are not interoperable or optimized for cloud storage, creating data silos that hinder data sharing and reproducibility. SOMA, a language-agnostic data model and API specification for single-cell data, offers a scalable, efficient, and user-friendly solution for storing and processing single-cell omics data. TileDB SOMA is optimized for cloud object stores, interoperable with popular tools and languages, and highly scalable for atlas-scale data. By leveraging TileDB, researchers can focus more on scientific discovery and less on data management.

[tile]DB Carrara offers a FAIR data catalog & compute for atlas-scale single-cell omics

TileDB Carrara offers a scalable, efficient, and user-friendly solution for single-cell omics data, enabling researchers to focus more on science and less on data management.

In partnership with the Chan Zuckerberg Initiative, TileDB built **SOMA (Stack Of Matrices, Annotated)**, a language-agnostic data model and API specification for single-cell data, which is SOMA's implementation with TileDB as the backend storage and processing engine, taking advantage of TileDB's powerful multi-dimensional array engine.

Why [tile]DB Carrara for Single-cell

[tile]DB
DESIGNED FOR DISCOVERY™

- ✓ Organize
- ✓ Structure
- ✓ Collaborate
- ✓ Analyze

Highly scalable

TileDB powers **100M+**
data from **cells**

in Chan Zuckerberg
Initiative CellxGene

Efficient FAIR data catalog

<1 hour to build
a single-cell atlas

which previously
took days

Interoperable

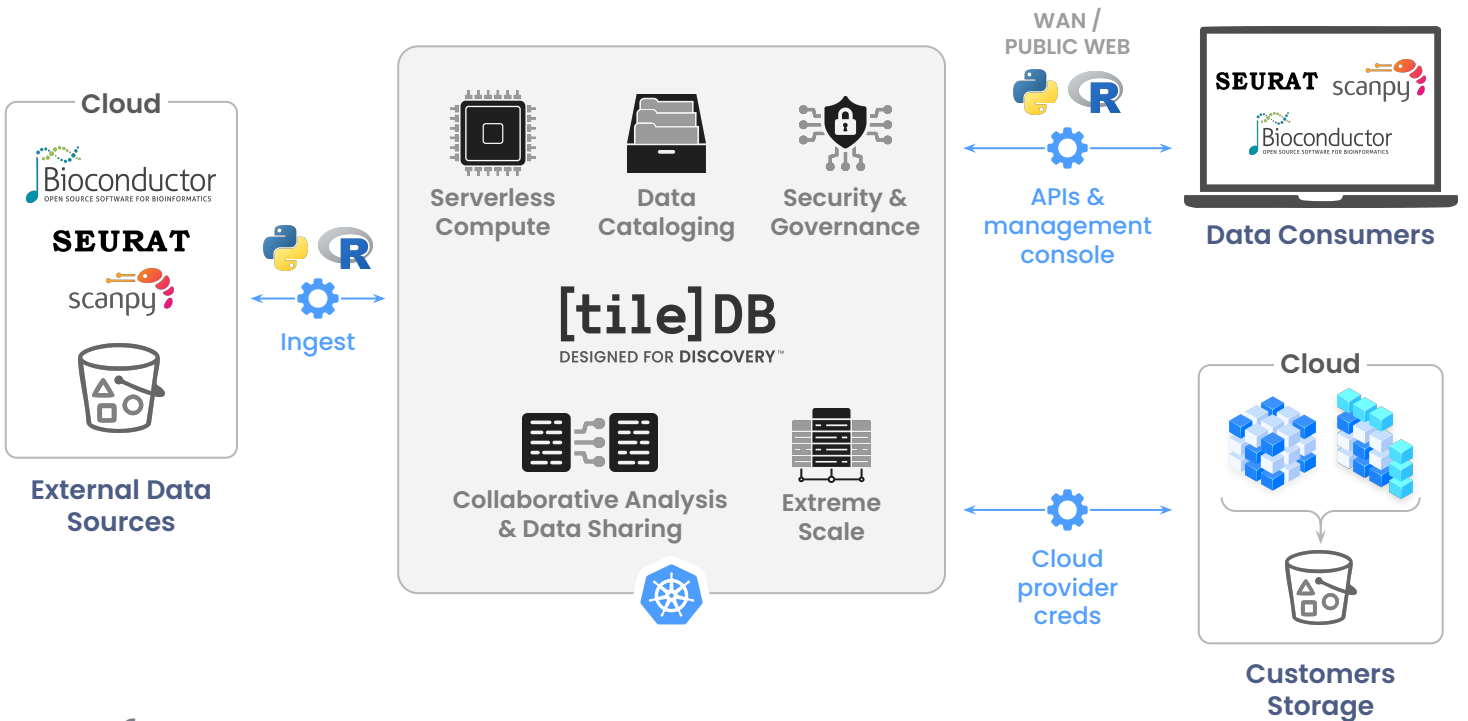
Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

SEURAT scanpy



Get started
with single-cell at:
tiledb.com

Advantage of TileDB's powerful multi-dimensional array engine



Key features

Optimized for object stores

TileDB inherits the cloud-native array format, particularly optimized for object stores (such as the popular AWS S3, Google Cloud Storage, Azure Blob Storage and MinIO).

Support for spatial transcriptomics

Efficiently write and access large spatial datasets both locally and in cloud storage through a centralized data store optimized for long-term cost effectiveness, and run low-latency queries with notebook-style analysis. Ingest data from the 10X Visium Space Ranger and integrate directly with the SpatialData python package.

Train ML models on single-cell data

With the TILEDB-SOMA-ML library, researchers can use PyTorch to train models at scale, with optimized data workflows for efficient shuffling, sampling, and analysis of complex datasets.

Vector search: Cell similarity & annotation

Enable automated cell type annotation and interactive analysis for deeper biological insights. These advances streamline reference mapping workflows and empower researchers to efficiently explore their data.

Customer Case Studies

Discover detailed customer case studies on our website at tiledb.com



Scale & speed target discovery by applying ML to large-scale single cell data on TileDB.



Power next generation drug discovery using single-cell data at scale.