

[tile]DB



Accelerate your Journey to Discovery

Upscale your life sciences workflows to drive
therapeutics and precision medicine breakthroughs.

STARTED AT

intel 

Discovering the next therapy is challenging and expensive

With ultra-high costs exceeding \$2.6 billion and lengthy timelines, drug and target discovery has limited reach in finding new cures and tackling new therapeutic areas.

- Cost and innovation pressures higher than ever
- Ripple effects of Inflation Reduction Act have begun
- Greater expectations for time to market in post-COVID era
- New modalities beyond small molecules





To discover groundbreaking therapies,
**R&D teams are digging deeper into the
complexities of biology,**
**but the data powering that innovation
demands a high touch approach.**

Life sciences data is challenging

- ✓ complex life sciences data in custom formats, often "unstructured"
- ✓ extremely large individual records
- ✓ force fitting data into legacy systems that cannot handle it effectively
- ✓ the need for data to be FAIR: findable, accessible, interoperable, and reusable

BIOMEDICAL IMAGING

1+ PB/yr

per large pathology center*

SINGLE-CELL DATA

85 million

cells in atlas scale projects like CZI CellxGene

POPULATION SCALE OMICS

>12 PB

in biobank scale projects like 100,000 Genomes Project UK**

* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9577055/>

** <https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0>

CURRENT STATE

Data innovation but at great cost



Operational

Force-fitting complex data into tables requires a **patchwork of tools** and creates derivative data to do any kind of analysis. It creates **storage and compute costs** and results in **siloes data**.

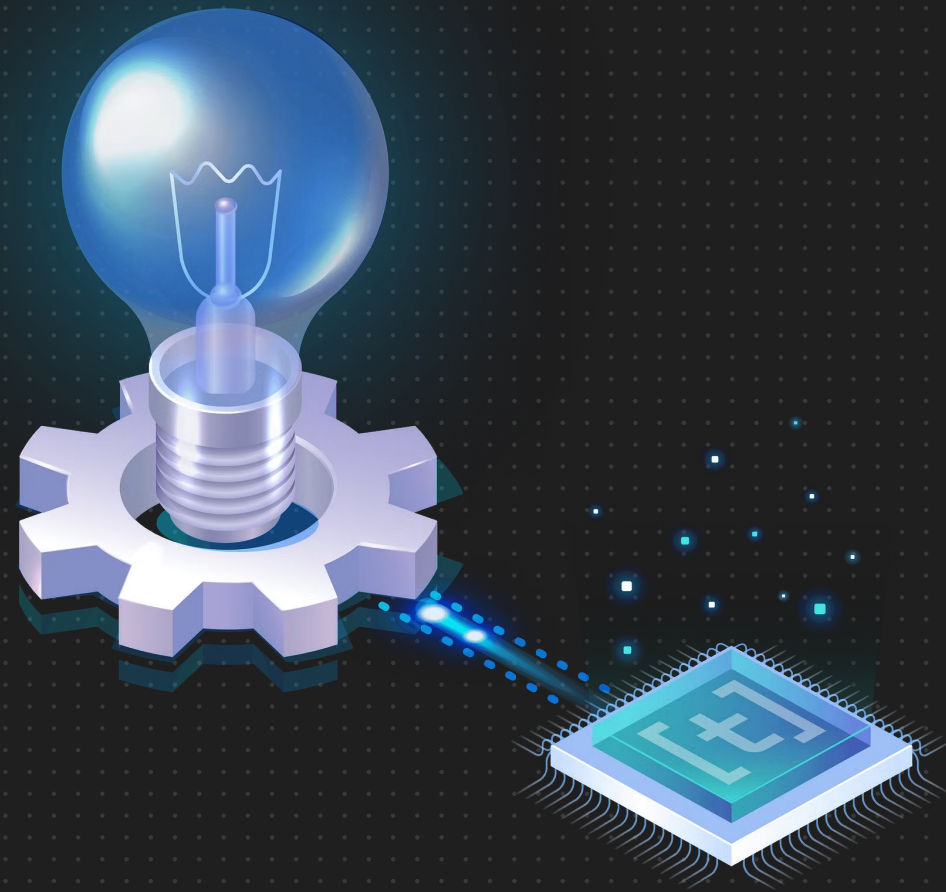
People

Low productivity of bioinformaticians and computational biologists who need to become **data wranglers**

Governance

Derivative datasets also create **traceability and governance overload**, hard to track original files and **data sharing** also gets impacted.

Ideal solution



Interoperable analysis environment

Include a comfortable analysis environment with the tooling and libraries to easily load data and run experiments.

A single platform for all data

A single platform that can support all data and metadata through the entire data life cycle - one that stores raw data in a unified common data structure, as well as intermediate data products.

Secure collaboration

A collaborative environment to securely manage all data, metadata and custom algorithms in a single spot for growing research teams.

Scalability

Automatically scale as needed by the workload, allowing scientists and engineers to focus on analysis instead of infrastructure.

TileDB was purpose-built to solve this problem

Unified Data Model

Catalog & Governance

Low Maintenance

ETL & pipelines

Data warehousing

AI and ML

Vector search, ML models, LLMs

Unstructured data

Text (pdf, docx, etc), images, audio, video and more

Code management

Files

File management, semantic search

Applications

Notebooks, dashboards, web apps

Specialized multi-omics data

Genomics, Single-cell, Proteomics, ...

Scientific data

ND arrays



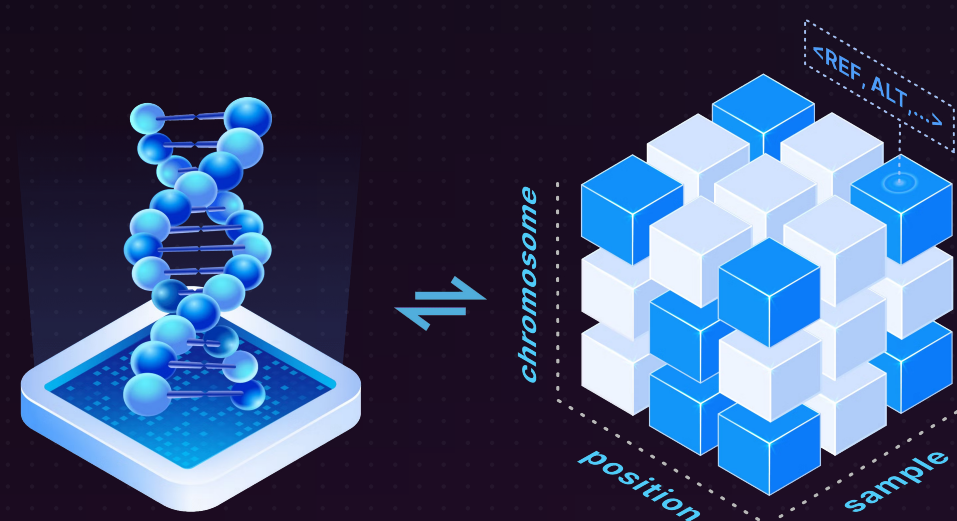
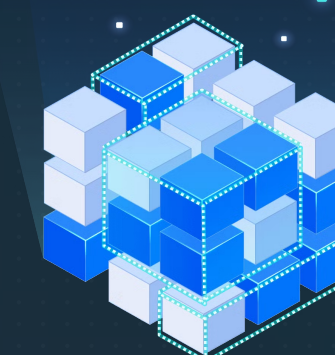
TileDB
the all-in-one
solution

No data is unstructured for shape-shifting arrays

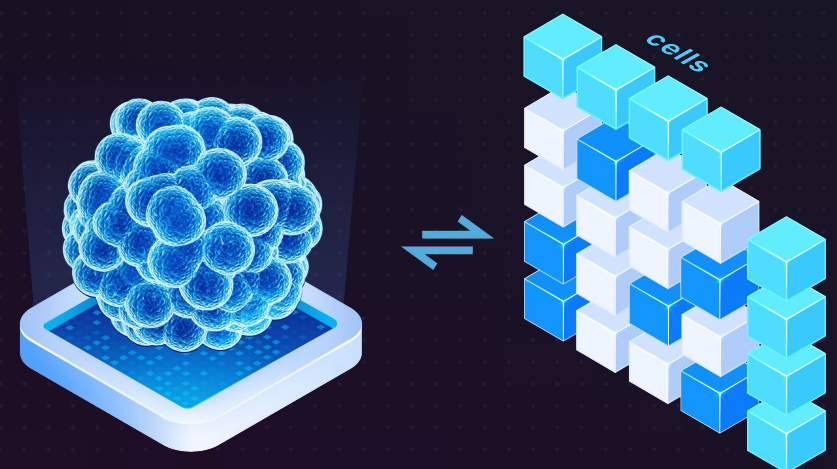
A **single platform** that can support **all data and metadata** through the entire data life cycle - one that stores raw data in a unified common data structure, as well as intermediate data products.

Arrays

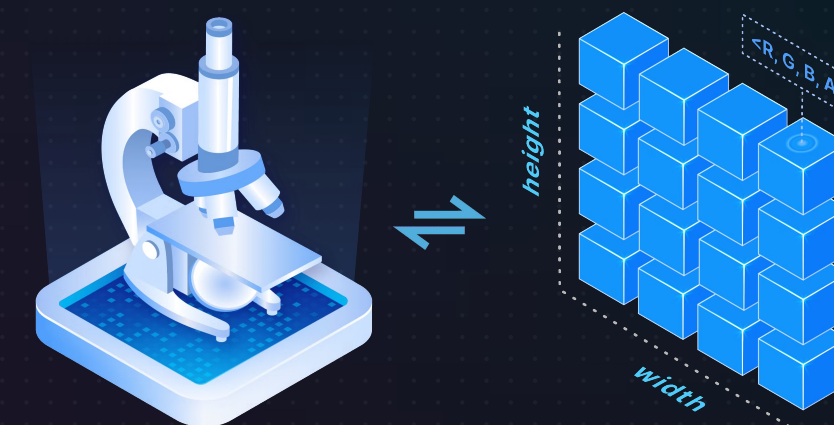
A fundamental data structure that shape-shifts to perfectly capture any diverse data type and workload. Optimized for cloud object stores.



Population genomics



Single-cell



Biomedical imaging

Serverless cloud architecture to allow experts to focus on scientific analysis and ML.

A collaborative environment to securely manage all data, metadata and custom algorithms in a single spot for growing research teams

Usability and ease of extracting, filtering and downsampling subsets of large datasets to accommodate complex analyses

Compute

A flexible distributed computing platform based on task graphs, 100% serverless, supporting heterogeneous resources and multiple languages.



Governance

All data, code and integration interactions are access controlled and logged. A unified catalog for holistic view over all assets.



Code

Numerous APIs, Jupyter notebooks, visualization tools, versatile dashboards, and integrations with a vast open-source ecosystem.



TileDB for Life Sciences

Population genomics

Power variant data analysis at biobank scale

Scaling genomic variant data analysis for rare disease detection in newborns and advancing cancer treatment development through collaboration with Rady Children's Hospital and the BeginNGS consortium.

Single-cell analysis

Unify single-cell data at atlas scale

TileDB-SOMA powers Chan Zuckerberg Institute CELLxGENE Census, providing cloud-native scale, extreme performance, and data interoperability for accelerated drug discovery in single-cell transcriptomics.

Biomedical imaging

Securely manage microscopy data

TileDB enables ML-based segmentation for imaging in diagnostics and biopharma and also enables rapid ML training and search in oncology and pathology imaging for swift diagnoses.



TileDB architecture

Interoperability

APIs and ecosystem integrations

Domain-specific applications



APIs



ML & data science



Efficient APIs and tool integrations with zero-copy techniques

[tile]DB CLOUD

- Access control and logging
- Serverless SQL, UDFs, task graphs
- Jupyter notebooks and dashboards

Unified data management and easy serverless compute at global scale

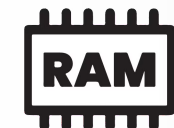
[tile]DB OPEN-SOURCE

- Rapid reads and writes
- Columnar, cloud-optimized
- Data versioning and time traveling

Open-source interoperable storage with a universal open-spec array format

Compute & governance

Powered by k8S
SaaS or self-hosted



Storage

You always own your data

TILEDB CLOUD

Designed for Discovery

Request a demo →



tiledb.com

