CASE STUDY:

Pharma-AI Collaboration Drives ~$3M Cost Reduction by Using Highly Curated Public Data

THE PROBLEM:

A **leading genomics-based drug discovery company** located in California, USA wanted to accelerate the identification of putative targets across immunological diseases and cancer by complementing their research with **publicly available data** but were unable to leverage the full potential of available data because the data from various sources were not FAIR - **Findable Accessible Interoperable Reusable**.

THE CHALLENGE:

Valuable research data is stored in different repositories in diverse formats. The semi-structured form in which the data is shared and stored makes it extremely difficult to derive value out of it.

Specific challenges were around:

Finding relevant data because of incomplete metadata, lack of a standard ontology across the names of genes, tissues, diseases, etc.

Non-standardized and non-annotated data which was not fit for downstream analyses.

Lack of analysis tools which could directly perform downstream analysis on the available data.

Reports state that "Each **one-day delay** in getting a drug or therapeutic to market costs their company an average of **$1.26M** in lost revenue and/or loss of first-mover advantage."

THE SOLUTION:

Public OmixAtlas which hosts highly curated data that is easily findable, and is structured in an ML-ready tabular format. The major value drivers were:

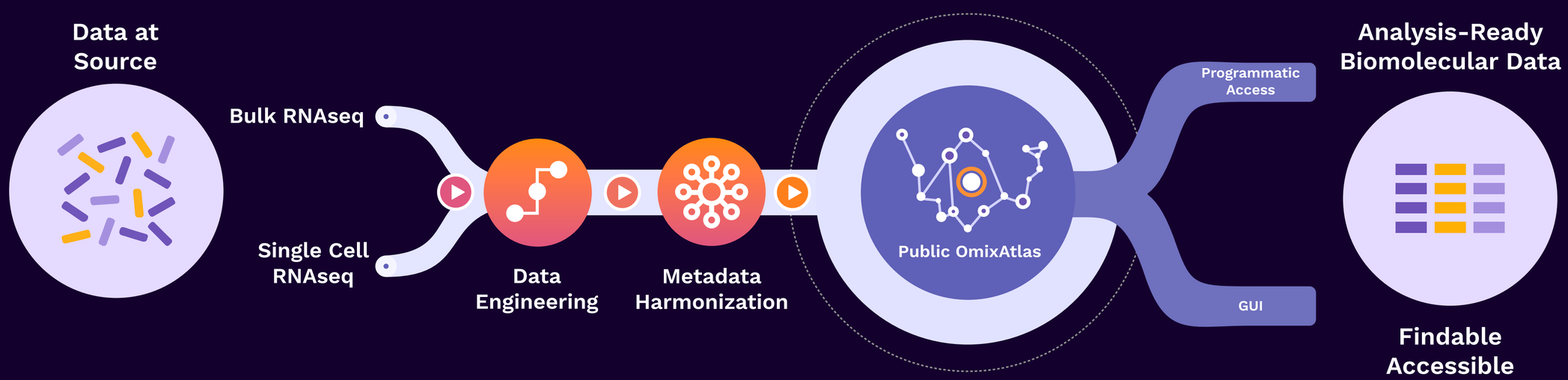
- **Standardized, Annotated** Data
- **Accurate, Harmonized, Complete** Metadata
- **Standard Ontology**
- **Custom Curated Fields**
- **Integrated Tools** for Data Exploration

(Check out the next page for more details on the process)

THE OUTCOME:

- Integrated large amounts for public transcriptomics (bulk and sc-RNAseq) from **GEO** and **TCGA** to prioritize putative targets across immunological diseases and cancer.
- Created a curated **Pan-cancer Immune Atlas** to understand the tumor infiltration by immune cells across 33 cancer types using transcriptomics data from TCGA and identified **1+ target for immunology group.**

THE WORKFLOW



THE IMPACT

De-risk Advancement to Phase II Leading to Lower Trial Cost	**$3M** Cost savings
Time savings for R&D personnel and Bioinformatics	**2,000** Hours freed up annually

To know more about Polly, contact us at **info@elucidata.io**

elucidata

EXECUTION DETAILS

The Ask

- **Finding** relevant datasets
- **Deeper curation** to drive discovery and more granular comparisons / analysis
- Data **visualization tools**

Technology

- **Customised pipelines** to process data
- **Ontology recommendations** for better search results
- **Polly Python** to search for datasets across repositories
- **Integrated tools** such as Spotfire and Cellxgene

Delivered

Data Audit

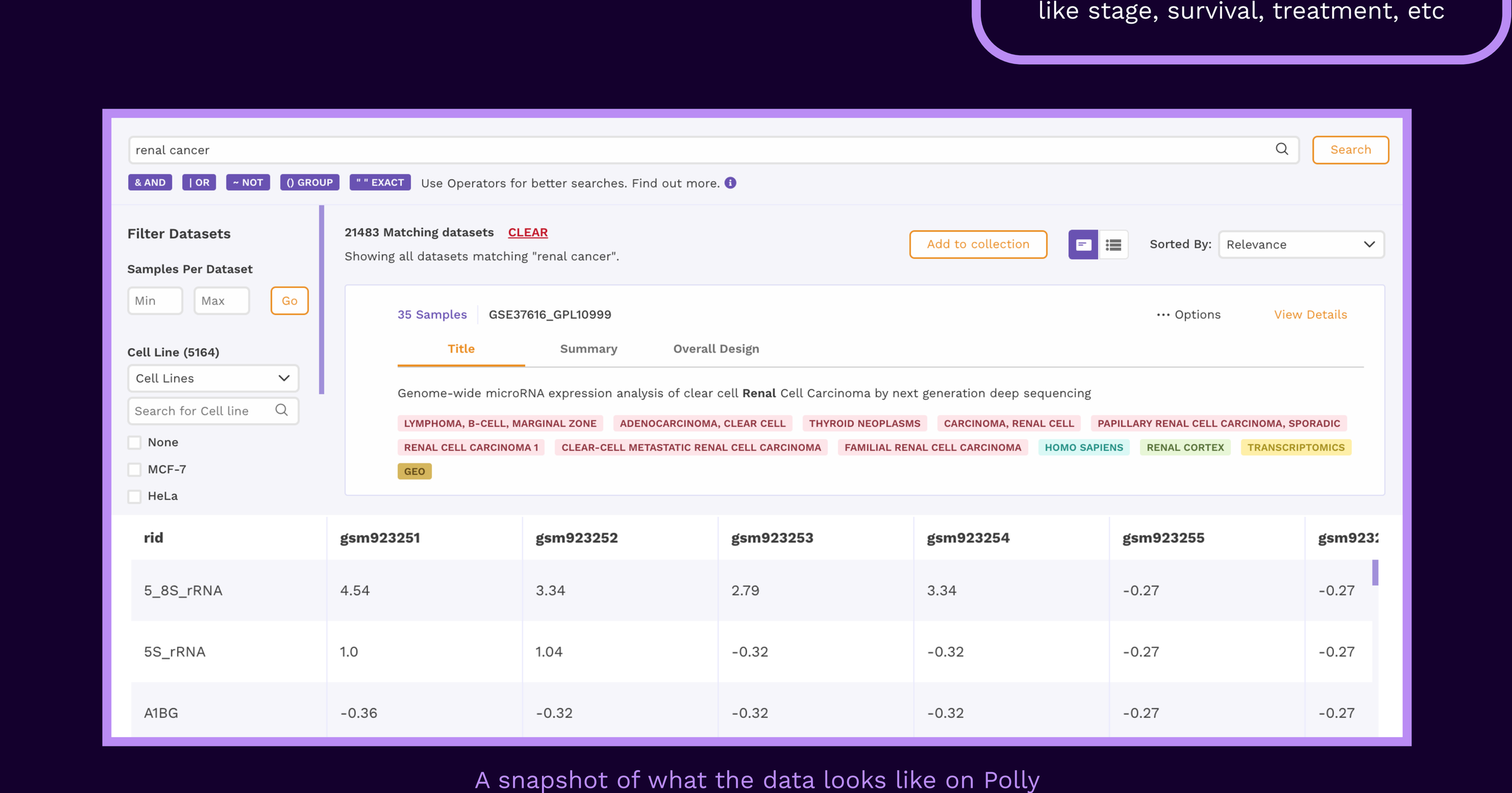
Identified **4000+** relevant datasets

ML-ready Datasets

500 datasets with **10,000** samples

Highly curated data

6 standard fields like cell line, organism, etc. **10** custom curation fields like stage, survival, treatment, etc



A snapshot of what the data looks like on Polly

Industry reports state that there is approximately a **65% ACCELERATION** in time-to-market if data scientists, scientists, QA, and other stakeholders had access to FAIR scientific data.

In alignment with this statement, the highly curated data on Elucidata's Public OmixAtlas played a pivotal role in helping the pharma company identify **1 target** for an immunology group in just **6 months**, as compared to the usual time period of 2-3 years.

WHY ELUCIDATA

Infrastructure

Polly Connectors to ingest and harmonize data at scale.
Data velocity:
25,000 datasets/month

Discovery

Searching for datasets with harmonized metadata annotations on Polly lends **~300%** more results across sources

Expertise

Multi-disciplinary team of **150+** experts based across the US and India

ABOUT US

Elucidata's mission is to transform the way data is used to drive decisions in R&D labs. Polly, our proprietary an end to end vendor-neutral, cloud-based platform standardizes and streamlines the biological big data analysis workflows and is currently used by over 30 MAUs in labs and organizations including Pfizer and Yale. Polly has enabled 10x faster identification of therapeutic assets with high odds of success in the clinic, using ML-ready data. It has enabled the detection of multiple validated drug targets across immunology, oncology, and metabolic disorders. We are a 150+ multi-disciplinary team of experts based across the US, Canada, and India.

Book a call with us to learn more

To know more about Polly, contact us at info@elucidata.io

elucidata