



Spatial Innovation Initiative Delivering Harmonized Analysis-Ready Datasets

Marissa Hirst, Sondra Kopyscinski, Nicole Leyland, Andy Hope, Amrita Bhattacharya, Dzmitry Fedarovich, Kenneth Chan, Dimitris Konstantopoulos, Dan Rozelle

Rancho BioSciences, LLC

Spatial Innovation Initiative

Abstract

In 2024 the Spatial Innovation Initiative (SII) was formed to make spatial transcriptomic data accessible to the biomedical research community. As an emerging technology, many organizations have yet to leverage spatial transcriptomic data for gaining insight into tissue organization, cellular interaction, and the interpretation of spatial patterns in a clinically relevant manner.

This initiative focuses on providing analysis-ready datasets in a format optimized for use individually, ingested into a data lake, or combined into integrated collections. To accomplish this, data is consistently processed, metadata is aligned to a spatial-specific data model, and deliverables are validated.

With three charter members we have established technology, workflow, formatting priorities and have identified studies of interest. To date, this work has delivered 33 datasets composed of a mixture of organisms, diseases, technology and tissue types. Shortcomings of lower resolution technologies have been addressed using cell-type deconvolution to estimate individual cell type.

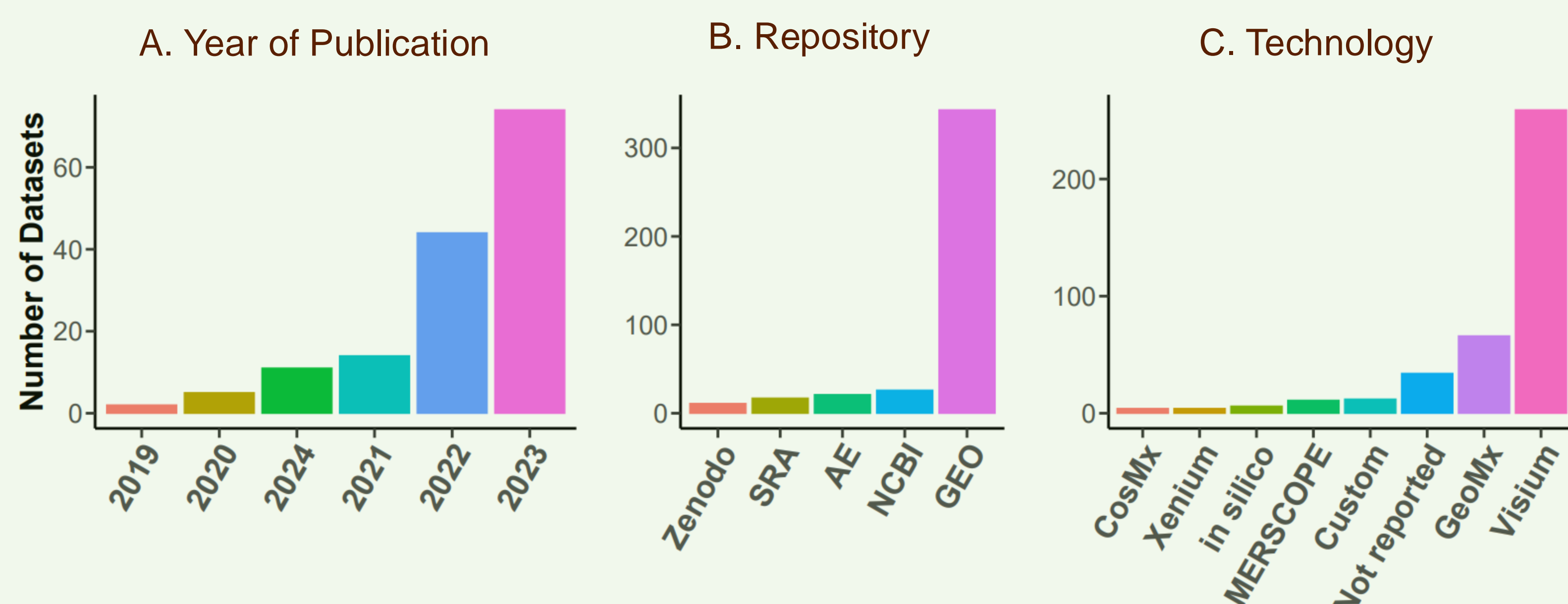
Based on the overwhelming success of the initial work, we anticipate expanding the SII to additional members and incorporating next generation technologies. The data produced from the SII will serve as a valuable resource in biotherapeutic target identification, high-quality training data for emerging AI models and accelerate insight generation across the biomedical community.

Spatial Technology

Technology	Type	Size	Features	Tissue Size
10x Visium	Sequencing	50um (10-25 cells)	Transcriptome (16k)	40 mm ²
Nanostring GeoMX DSP	Sequencing	Several cells	Transcriptome (11k)	Triplicate ROIs
Nanostring CosMx	Sequencing	Single cell	Up to 1000 genes	20 mm ²
Vizgen MERSCOPE/ MERFISH	Imaging	Single cell	300-500 gene panel	100 mm ²
10x Xenium	Imaging	Subcellular	200-400 gene panel	1400 mm ²

As an emerging technology, many technology solutions exist. The main types are defined as either sequencing or imaging-based techniques. In sequencing types, traditional high-throughput sequencing is performed on RNA located in a defined physical location which typically includes multiple cells. Imaging-based spatial technologies localize and enumerate gene-specific probes to a high degree of resolution but are limited to fewer genes than whole transcriptome.

Availability of Public Data



Rancho used their DataCrawler to identify publicly available spatial datasets. A., the number of publications is rapidly increasing each year. B., a majority of public spatial data is found in NCBI's Gene Expression Omnibus (GEO) repository C., 10x Visium is the most popular platform for spatial data.

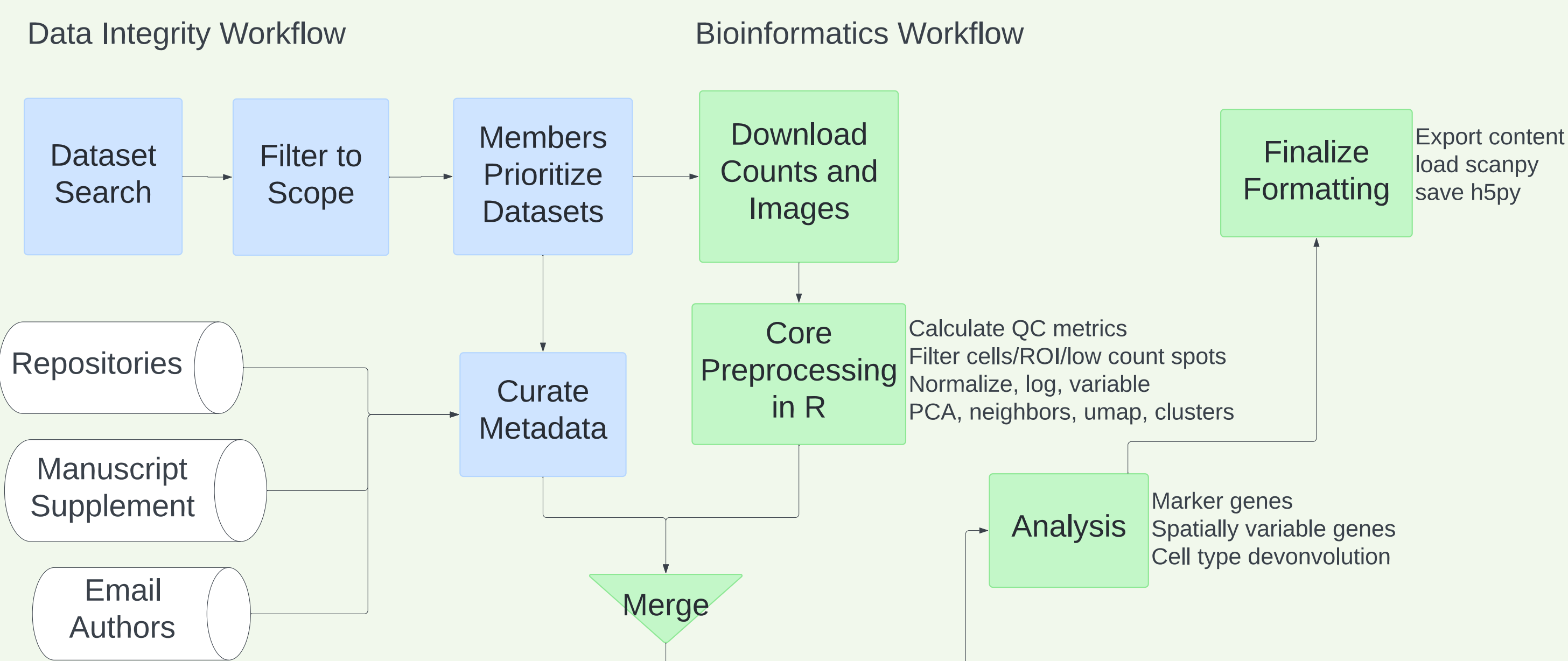
Rancho Initiatives and Consortia

As experienced consortium organizers Rancho plays a dynamic and rewarding role that requires leadership, organization, and a passion for bringing people together towards a shared vision.

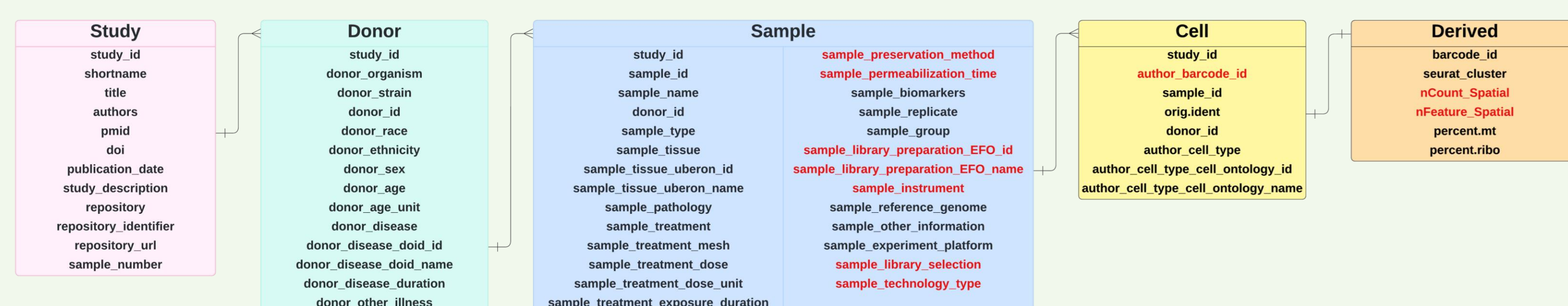
Rancho collaborates with Members to identify their goals and find solutions to ensure we are delivering in line with consortium needs. Through strong planning and communication, we collectively plan, manage timelines, budgets and the logistics for the consortium to keep everything running smoothly.

Ultimately, our goal is to see the collective efforts of the consortium produce impactful results, drive innovation of our Members, and save lives through data.

Data Ingestion Workflow

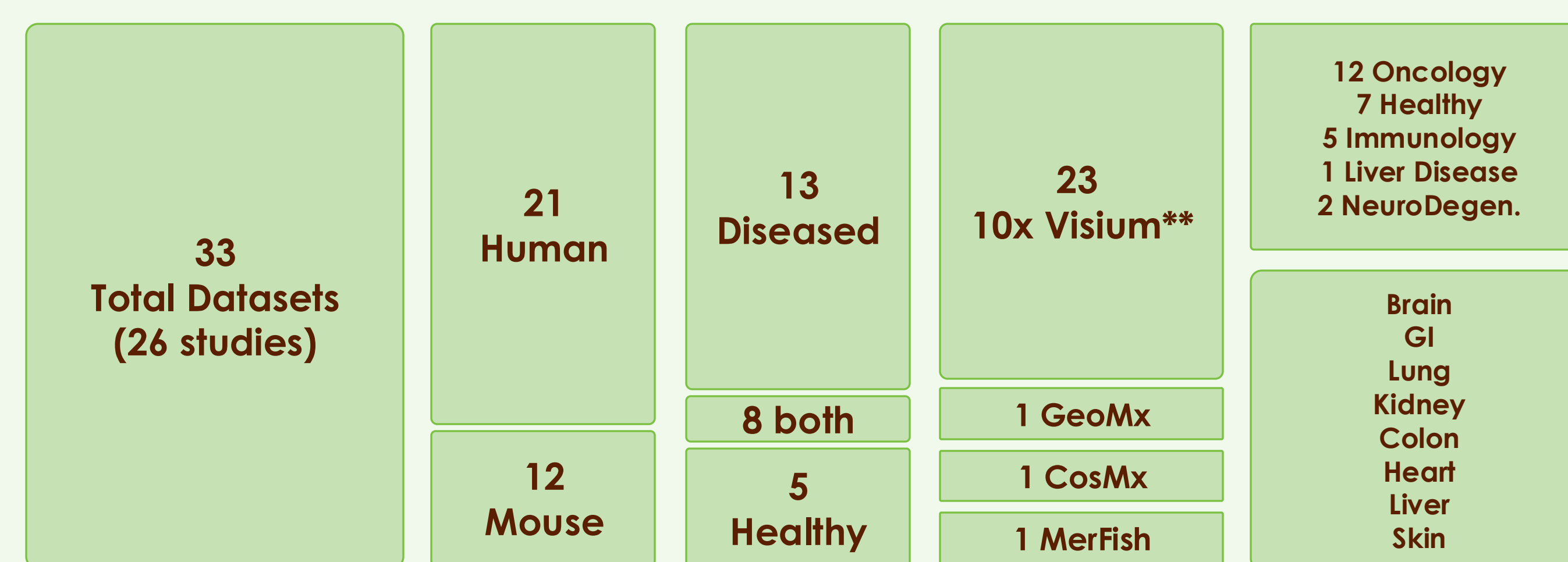


Spatial Transcriptomic Data Model



The 76-attribute Rancho data model is divided into 5 entities. Spatial-specific fields are highlighted in red. Metadata is derived from the publication, supplemental data, and repository. Values are standardized to public ontologies or a controlled vocabulary.

Delivered Datasets (July 2024)



**cell deconvolution completed for all Visium datasets

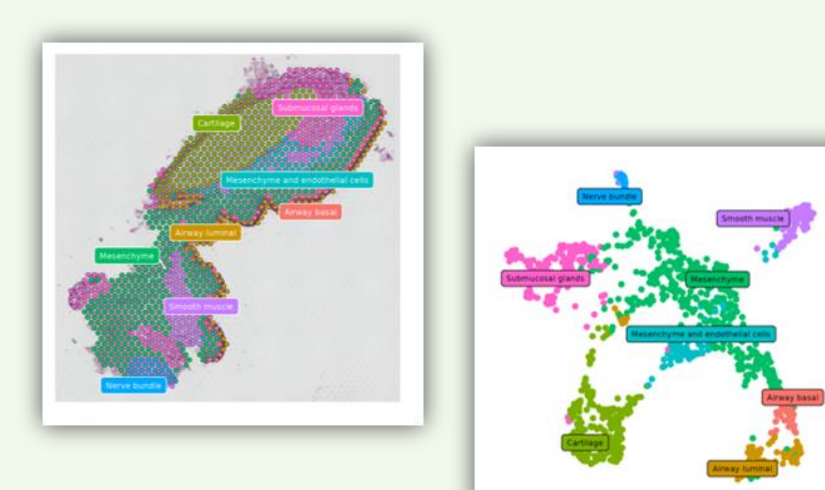
Primary Deliverables

- Catalog of spatial data**
 - Updated list of available datasets with key metadata categories. This is used to select datasets but can also inform state of the industry.
- Extended spatial data model**
 - Create a spatial data model to organize concepts and ensure consistent terminology.
 - Special emphasis on spatial-specific data features.
- Dataset deliverables**
 - 10 spatial datasets you designate.
 - + 33 spatial datasets selected by previous members.

New Members

A collective value initiative, designed to deliver ready-to-use high-quality spatial transcriptomic datasets. Valuable data to enrich existing biomedical research programs using public domain data.

- New 4th member will receive**
- 10 dataset selections
 - 33 existing datasets
 - Spatial data model
 - Pipeline code
 - Keep ALL deliverables in perpetuity.



Recapitulate published findings, and interrogate results to align with your internal biomarker programs

Join today and accelerate your research

Deliverable	Format
Scanpy analysis object	h5ad
Seurat analysis object	RDS
Spatially Variable Genes	csv
README	txt
Metadata workbook	xlsx
QC directory	misc
manifest	json

Contact Us



Rancho
bioSciences