

# AI-ready Perturb-seq and Spatial Transcriptomic Data for Drug Discovery Acceleration

Sondra Kopyscinski<sup>1,2</sup>, Amrita Bhattacharya<sup>2</sup>, Dzmitry Fedarovich<sup>2</sup>, Marissa Hirst<sup>2</sup>, Ben Ernest<sup>1</sup>, Corey Oravetz<sup>1</sup>, Nicole Leyland<sup>2</sup>, Dan Rozelle<sup>1,2</sup>

Rancho BioSciences Initiatives: <sup>1</sup>Perturb-seq and <sup>2</sup>Spatial Innovation Initiative

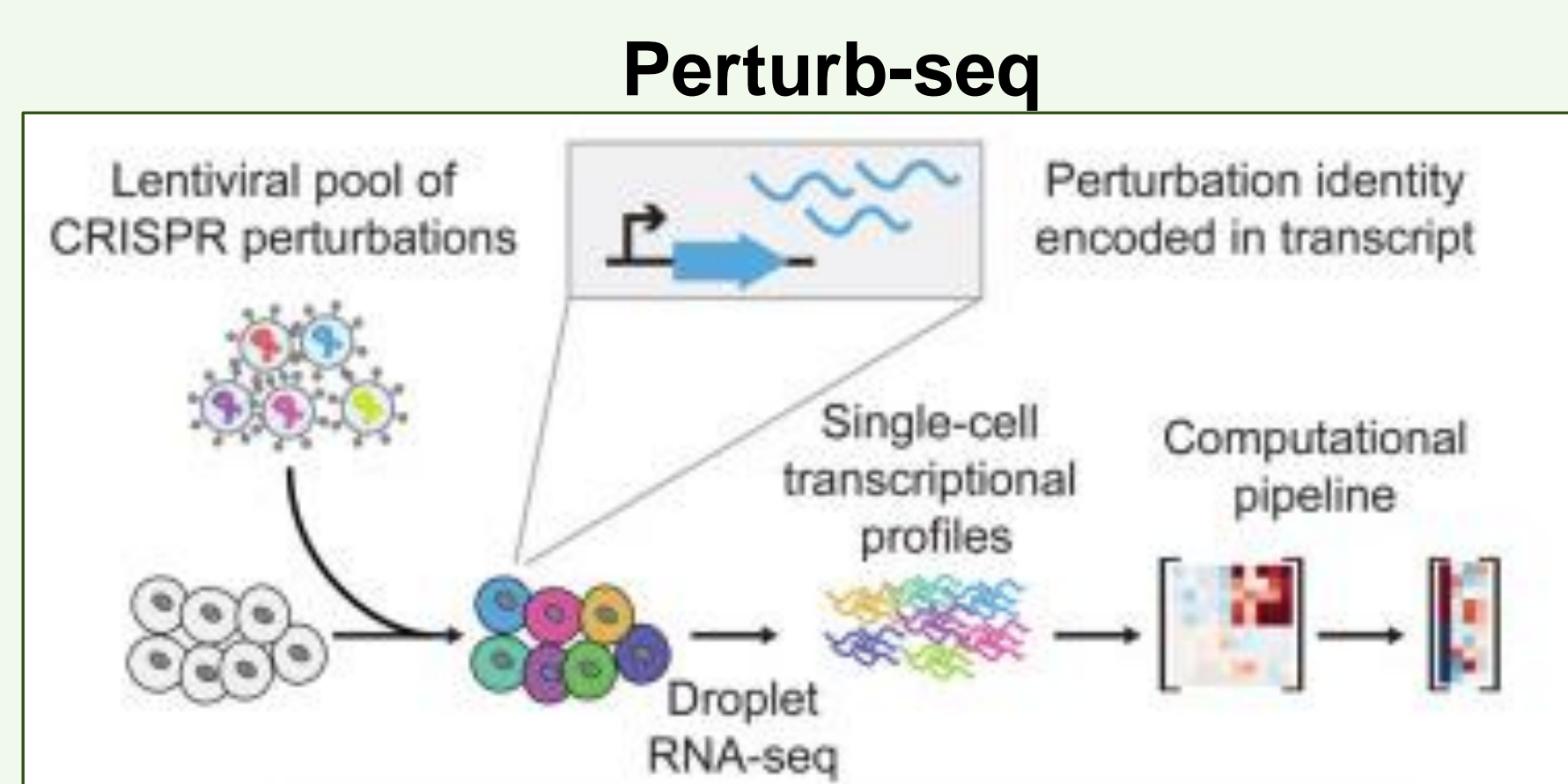


## Background

Drug discovery remains one of the most formidable challenges in the preclinical pipeline, particularly the identification of novel therapeutic targets. Breakthrough technologies such as Perturb-seq and Spatial Transcriptomics are reshaping this landscape by enabling unprecedented insights into cellular function and tissue organization. However, integrating and leveraging the vast datasets generated by these approaches is still complex and fragmented due to inconsistent formats, incomplete metadata, and the absence of standardized practices.

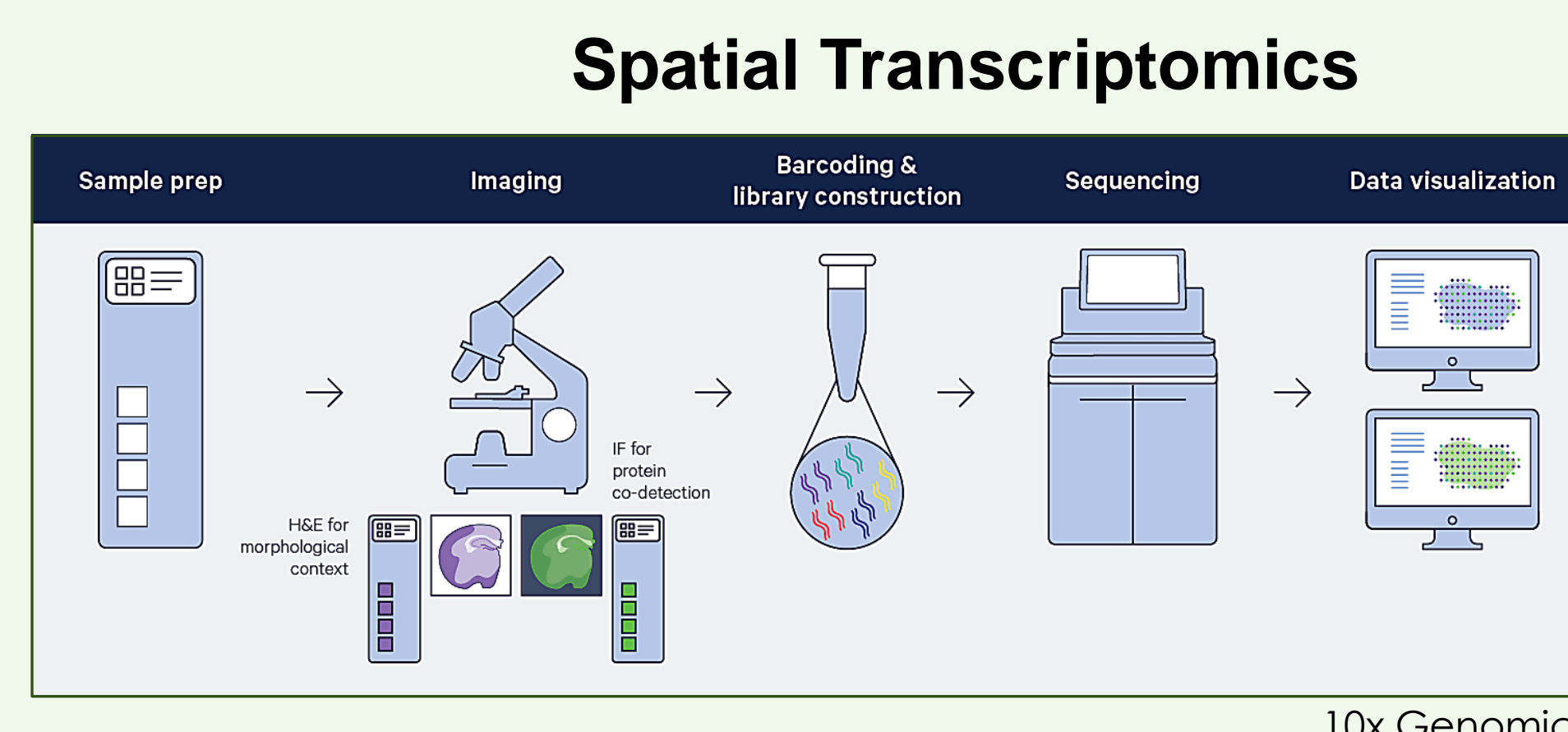
Rancho BioSciences combines deep expertise with proven experience in curating, harmonizing, and analyzing these cutting-edge datasets to accelerate discovery. Whether working with internal data or publicly available resources, Rancho BioSciences eliminates barriers to analysis, delivers AI- and ML-ready datasets, and speeds the path to actionable insights in drug development.

## What data is available?



Adamson et al. 2016

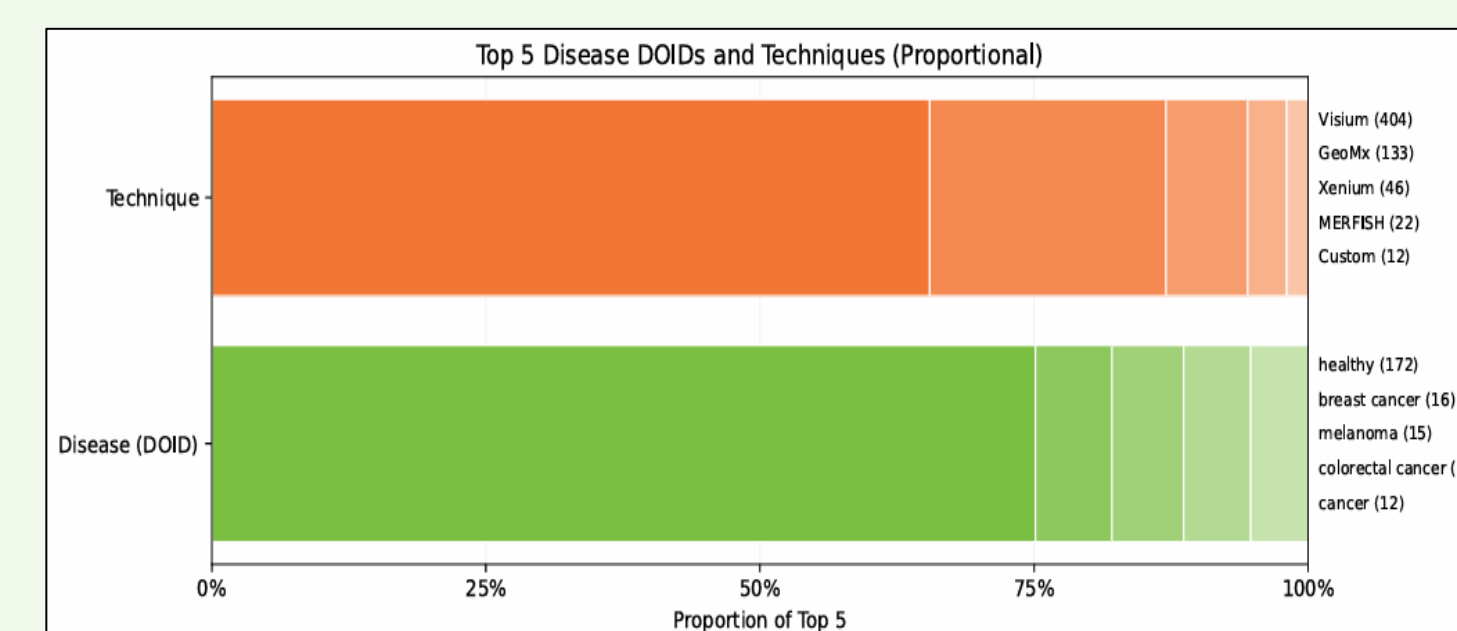
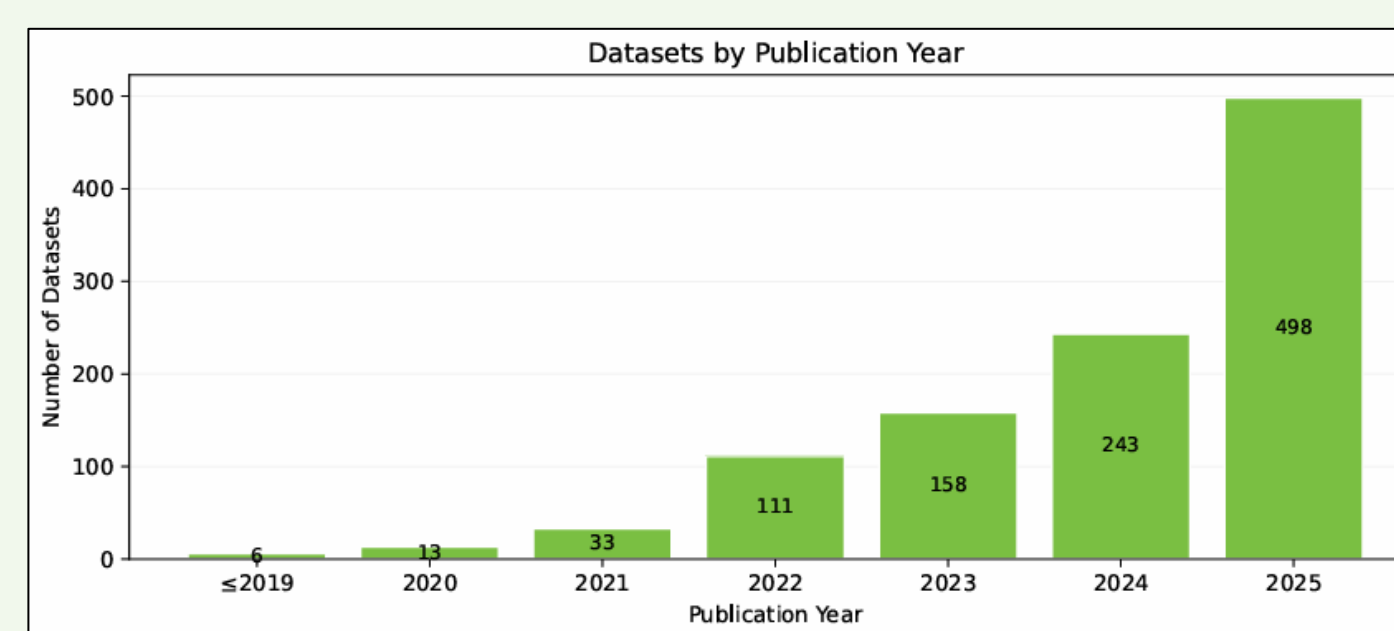
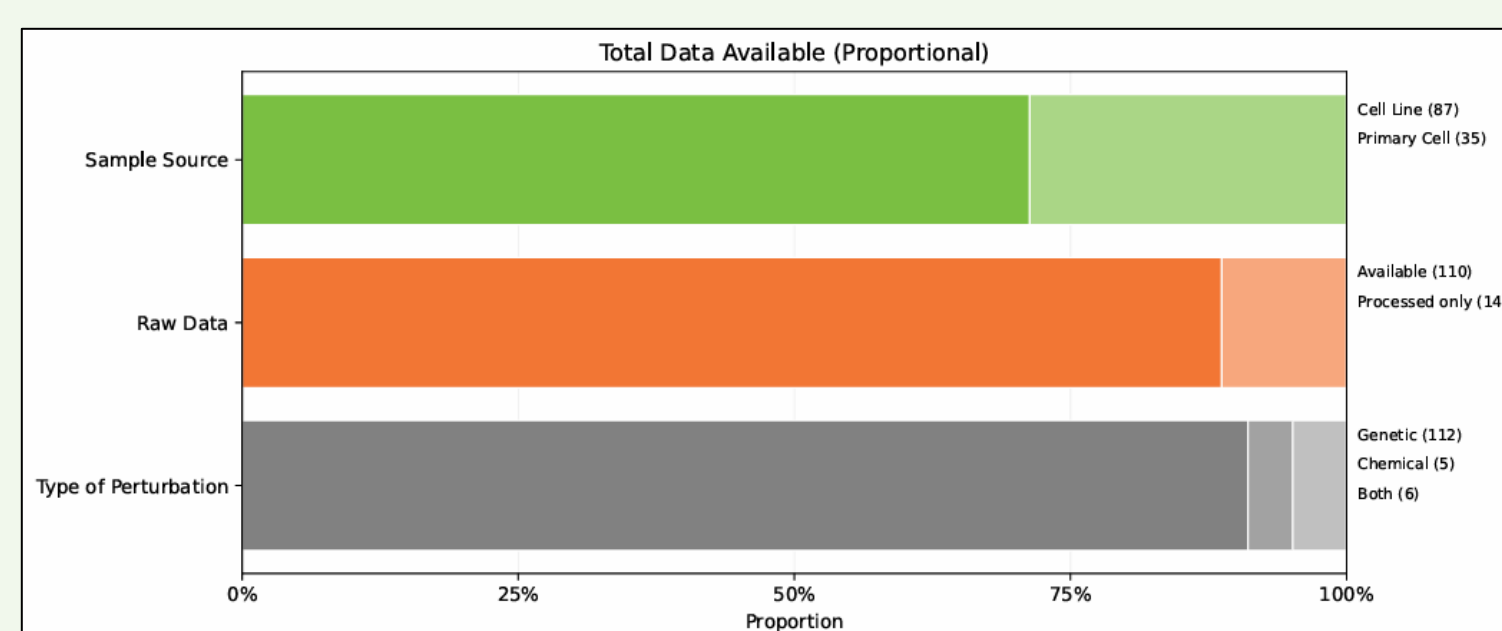
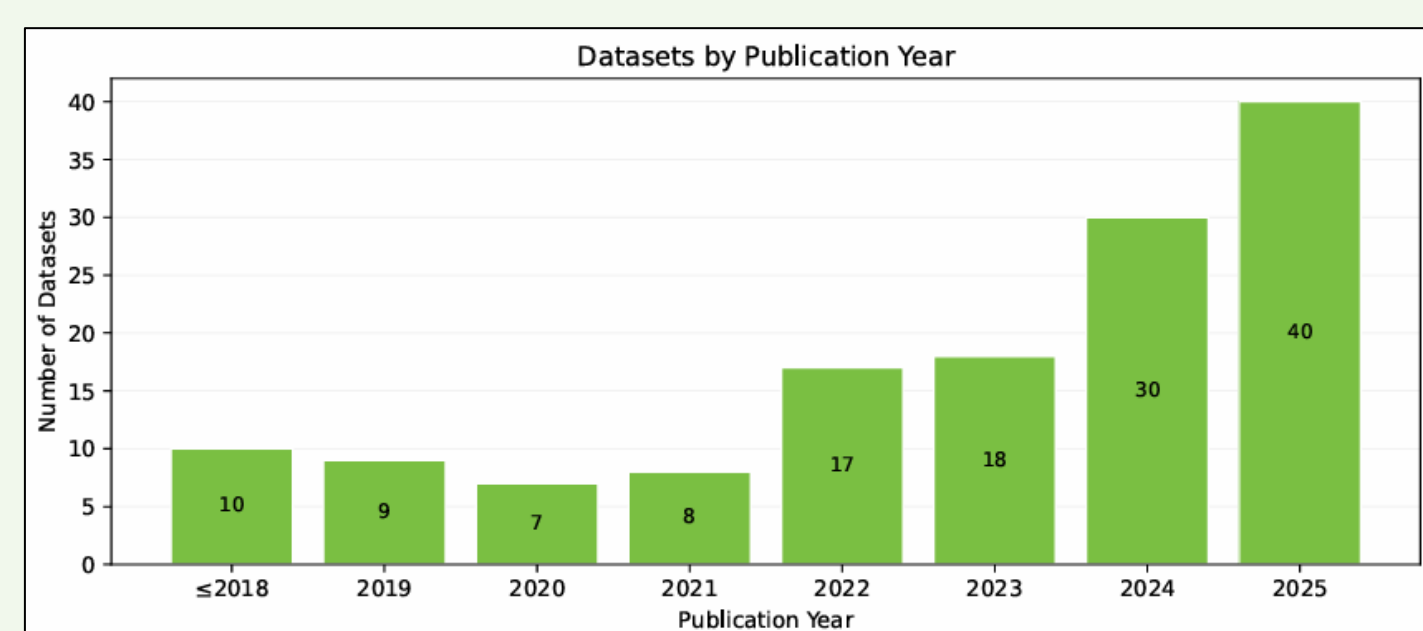
Perturb-seq is a powerful high-throughput functional genomics technique that links pooled genetic perturbations with single-cell gene expression profiling. This method transforms chemical- or CRISPR-based perturbations into detailed, single-cell transcriptomic maps, enabling comprehensive phenotypic profiling across genetic perturbations.



10x Genomics



Spatial transcriptomics is an advanced technique that measures gene expression while preserving the spatial context of cells within a tissue. Unlike traditional RNA sequencing, which loses positional information when tissue is dissociated, spatial transcriptomics maps transcript locations in the tissue architecture. This spatial dimension provides critical context for understanding disease mechanisms and therapeutic opportunities.

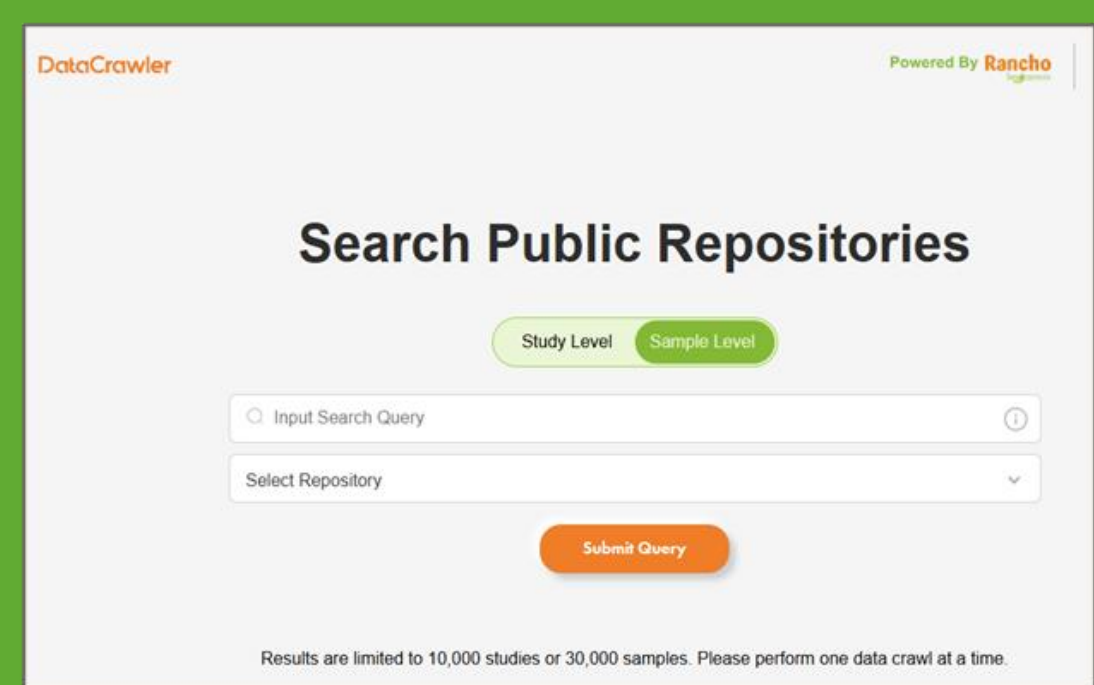


There is a vast wealth of Perturb-seq and spatial transcriptomic data available in public repositories.

## How do you get AI- and ML-ready datasets?

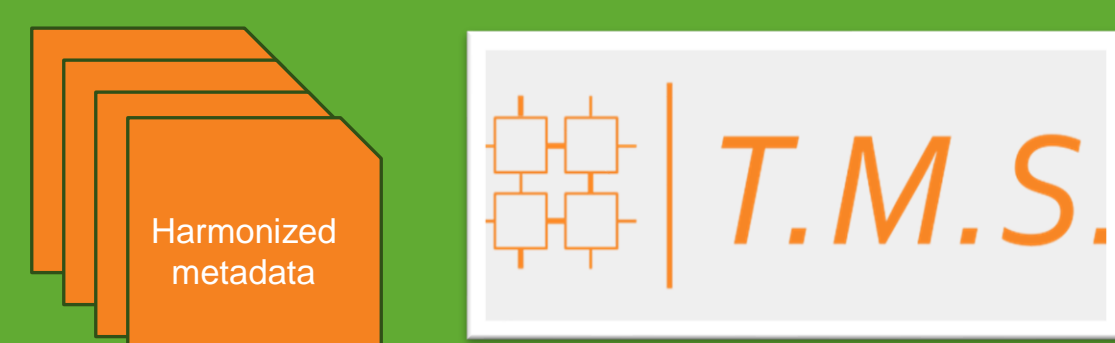
### Find Relevant Data

Rancho Data Integrity Specialists use our **DataCrawler** tool to search 12 public repositories for available data.



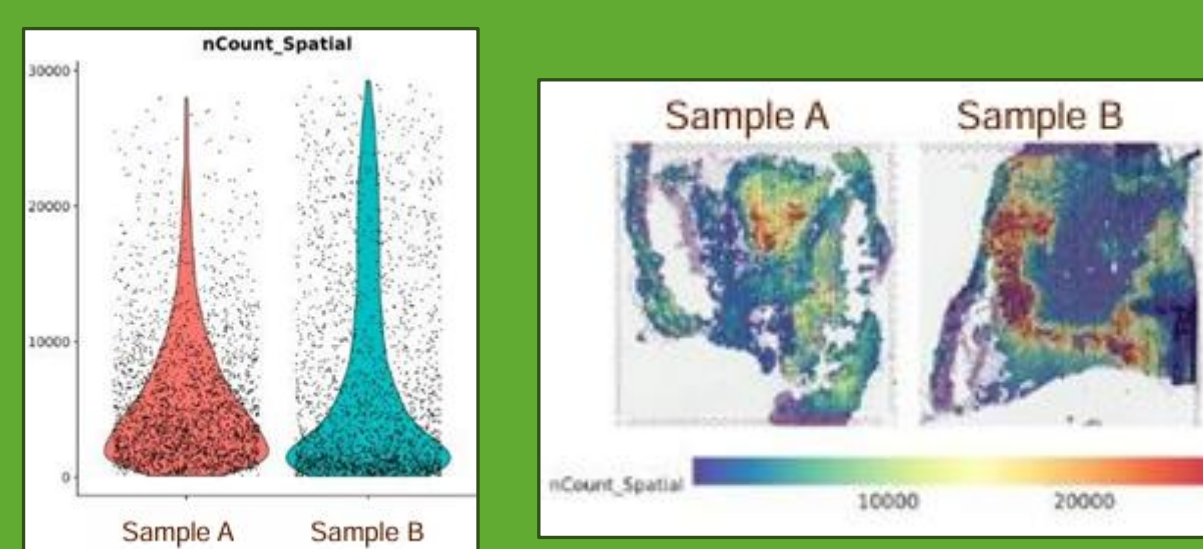
### Standardize Metadata

Rancho Data Integrity Specialists harmonize and standardize metadata from the repository, publications, and other sources to a desired data model and desired ontologies. Rancho uses our **Terminology Management Solution** to map values to ontologies in less time.



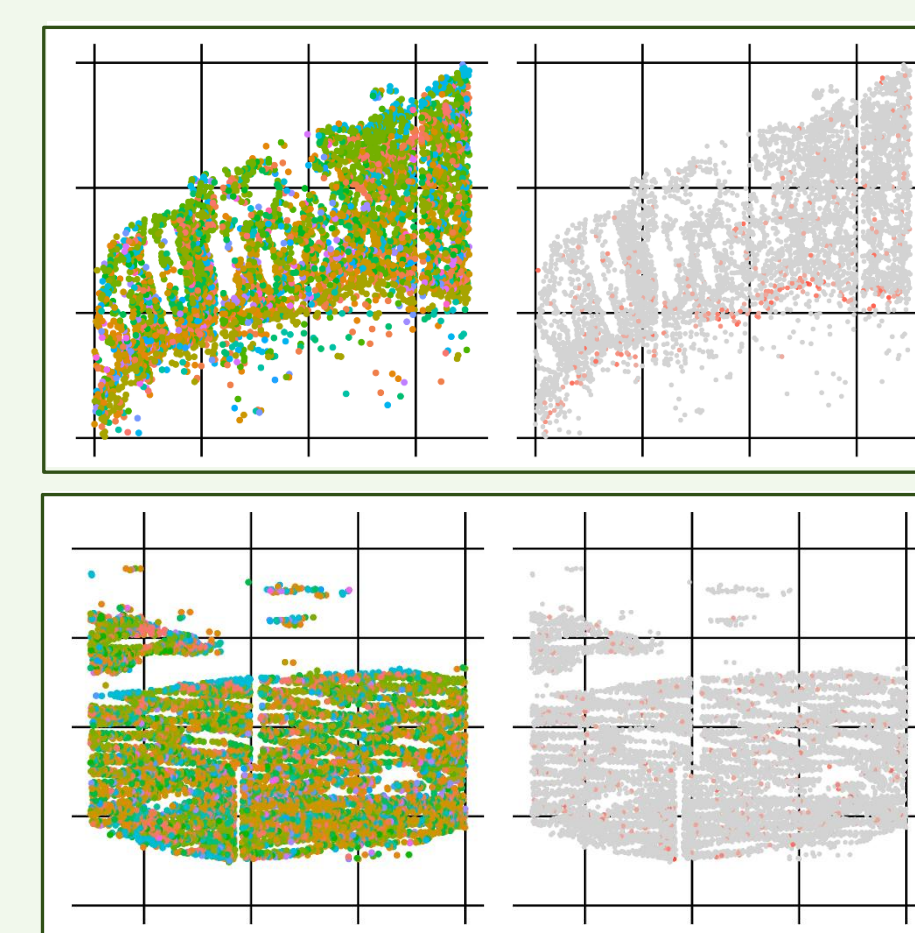
### Bioinformatics Processing

Rancho Bioinformaticians download and process molecular data uniformly, better enabling cross-study comparison.



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

### Example Spatial Transcriptomics Output



Healthy gut tissue with polarized expression

Diseased gut tissue with random expression

Together, these data products deliver standardized, high quality, multi study datasets that scientists can plug directly into their own discovery and AI/ML workflows.

## What can you do with AI- and ML-ready datasets ?

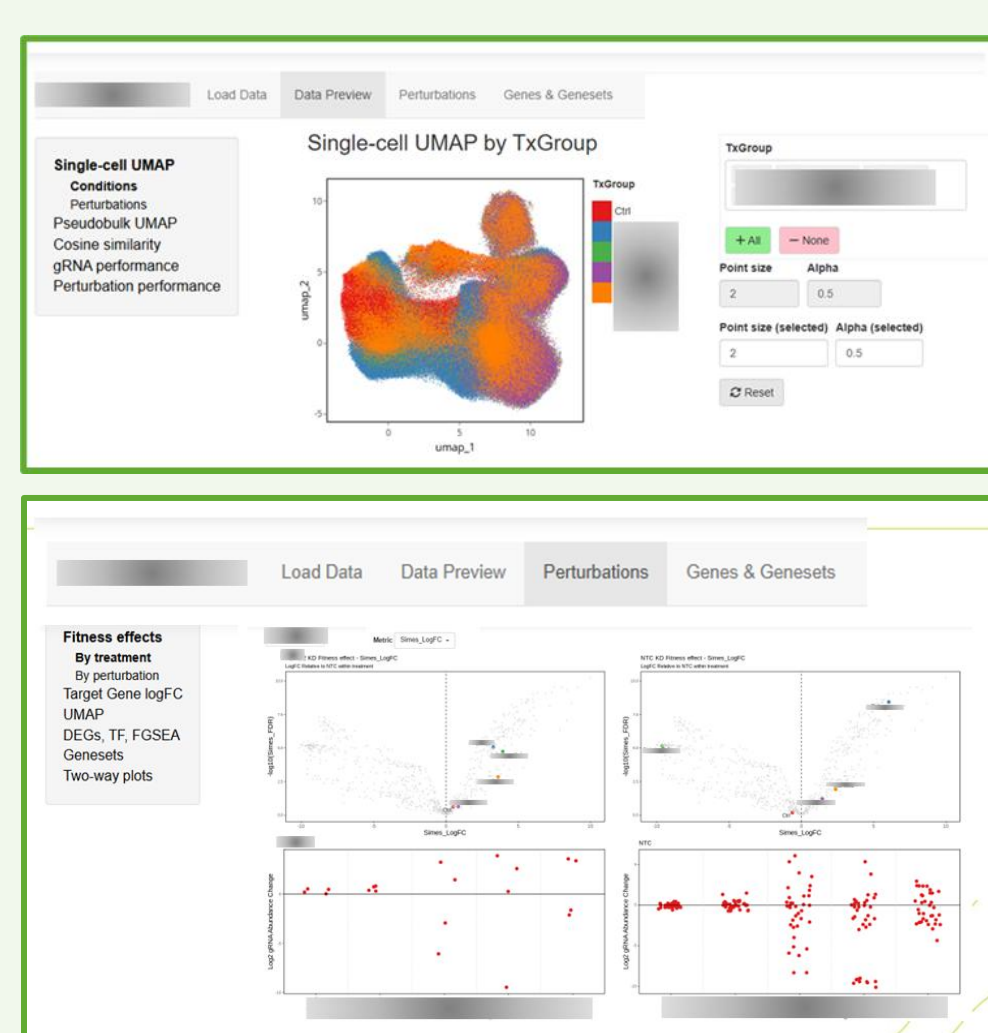
### Visualization App

#### Core Capabilities

- Explore data from genome-wide to guide RNA level
- Dozens of visualizations with intuitive layout
- Highly interactive plots with large data
- Publication-quality plot customization

#### Analysis Features

- Dimensional reduction & cell clustering
- Gene & gene set exploration
- gRNA performance tracking
- Fitness effects & correlation analysis



### Downstream Analysis

#### Drug Discovery

- Target identification
- Drug repurposing
- Toxicity prediction

#### Predictive Modeling

- Gene expression imputation
- Super-resolution
- Spatial trajectory inference
- Future state prediction

#### Generative Modeling

- Synthetic tissue generation
- Data augmentation
- Counterfactual analysis
- Missing data imputation

AND so much more!!

Want to learn more?  
Come meet us at  
**Booth 136A!**