Scan and Explore!

Strand's AI-Powered Single Cell RNA Portal:

Refining Precision Research, One Cell at a Time

Scan to log in to scRNA data curation portal.

We aim to establish an integrated platform that supports the entire drug discovery workflow, from the early stages of target discovery and validation through to the late stages of clinical trial support and patient stratification.

The Challenge: Data Heterogeneity Crisis

- Siloed data and inconsistent metadata make public scRNA-seq datasets difficult to compare or integrate, limiting target discovery.
- Lack of harmonization in terminologies, classifications, and identifiers across datasets complicates large-scale and integrative analyses.
- Integration across studies can dilute fine cell-type resolution, reducing accuracy and interpretability.
- Manual curation is slow and error-prone, delaying research onset as scientists must source datasets, fill gaps, and interpret raw data before moving forward.

Our Solution

Our proposed unified platform delivers AI-powered metadata harmonization and data analysis, bringing consistency across single-cell, bulk, and proteomics datasets. It provides secure pipelines for analyzing both public and in-house data, while integrating AI tools for annotation, integration, perturbation, and response prediction. The platform supports end-to-end discovery workflows, with applications in target discovery, biomarker identification, and translational research.

Metadata Harmonization

- Extracts key information
- Resolves inconsistencies in terms
- Standardizes with ontologies & vocabularies
- RAG-LLM Metadata Curation Tool

Analysis Pipelines

SCRNA-seq and **Proteomics** (MS/MS)

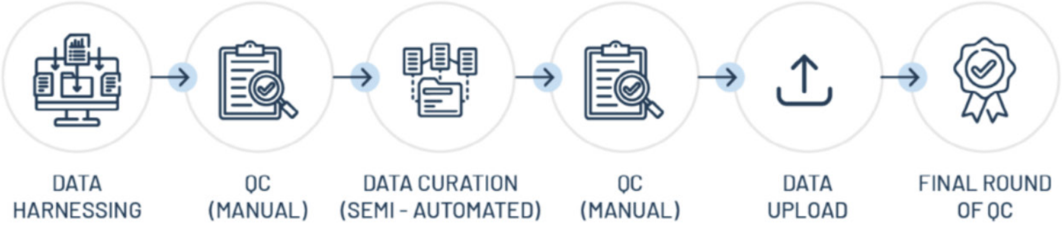
AI Toolkit

- **Cell-type** Annotation
- **Multimodal** Integration
- **Perturbation** Modeling
- **Drug Response Prediction**

1. Metadata Harmonization

i) Data Harnessing, Standardization and Harmonization:

- Strand has developed a data showcase for scRNAseq datasets with extensive metadata schema and standardized ontologies that can significantly cut down the time, effort, and error associated with each step.
- The showcase focuses on diseases with significant unmet clinical needs like Ulcerative Colitis, Alzheimer's Disease, Frontotemporal Dementia, etc.
- Public datasets, such as those from Gene Expression Omnibus (NCBI), are first harnessed and then meticulously curated and harmonized, with quality control processes applied separately to both stages, using a mix of automation and manual review.
- Turnaround time of 1-2 days.



Key Features:

- Our datasets have **80+** unique metadata fields organized at 3 levels of curation - study (11), experiment (35+), and sample (34+) levels.
- **19** metadata fields use controlled vocabulary and **21** metadata fields use defined ontologies.
- Our portal is equipped with **32** key filters that enable users to set attributes, categories, and values which can pick out relevant datasets within seconds.
- Well organized documentation section with clearly defined SOPs.
- Key metrics of the metadata can be visualized using bar graphs, pie charts and line graphs, and visualizations can be exported by users for easy and effective communication.
- Our showcase, updated monthly, currently contains datasets for Ulcerative Colitis, Alzheimer's disease and Frontotemporal dementia. Other disease types such as Parkinson's disease and Crohn's disease are in the pipeline.

Disease-focussed Portal

Inflammatory Bowel Disease

- Ulcerative Colitis
- Crohn's Disease

Neurodegenerative Disorders

- Alzheimer's Disease
- Parkinson's Disease
- Frontotemporal Dementia

Data Types

Single-cell Datasets

- Single-cell RNA Sequencing
- Single-cell TCR-Sequencing
- Single-cell BCR-Sequencing
- Single-nucleus ATAC Sequencing
- Single-cell ATAC Sequencing
- Single-nucleus RNA Sequencing
- CITE-Seq

Bulk RNA-seq Datasets

Proteomics Datasets (MS/MS)

Key Highlights

100+ fields curated, **40 with ontology** or manual standardization

32 filters that include all the metadata fields with ontologies (and a few more)

Search bar to navigate through metadata fields recorded as 'free-flow' texts

Example Showcase 1

Ulcerative Colitis (UC)

- Ulcerative Colitis is a chronic inflammatory disease with no available treatments for cure at this time. But RNAseq based studies are uncovering the cellular pathways that may be involved in disease pathogenesis.
- Our Ulcerative Colitis showcase currently hosts **43 million** cells, **1,814** samples, **106** distinct cell types, **35** different tissue types, **120+** metadata fields, and **89** datasets, sourced from NCBI's GEO platform.
- About **884** non-inflamed samples and **1002** inflamed samples are included, allowing meaningful comparisons between the two states.
- The samples are primarily from gut tissues, with a few from blood, altogether accounting for 106 unique cell types.

Our showcase with deep metadata enables phenotype-specific exploration

Clinical Severity

Stratify by Mayo score, PDAI, colitis severity

Disease Extent

Map cellular programs by colon region

Treatment

Profile treatment-associated cellular changes

Disease Behavior

Compare remission vs active UC states

Inflammation Status

Identify inflammation-driven cell types & pathways

Treatment Response

Map response-associated cell states & pathways

Example Showcase 2

Alzheimer's Disease (AD)

- Alzheimer's disease is a progressive, neurodegenerative disorder that gradually impairs memory, cognition, and daily functioning as neurons are damaged and die. Over 416 million people globally are believed to be on the AD continuum.
- There is currently no cure, and disease pathogenesis is not clearly understood. Its genetic and molecular pathways are being increasingly investigated using scRNA seq studies.
- Our portal showcases **10 million** cells, **1,161** samples, **64** distinct cell types, **24** different tissue types, **120+** metadata fields, **78** studies, and **2** model organisms, sourced from NCBI's GEO platform.

	UC Showcase	AD Showcase	Parkinson's Disease (PD) Showcase:
DATASETS	89 sourced from GEO	78 sourced from GEO	13 sourced from GEO
SAMPLES	1814*	1161*	192
CELLS CURATED	43 million	10 million	0.63 million
UNIQUE CELL TYPES	106	64	25
TISSUE TYPES	35	24	6
METADATA TYPES	120+	120+	85+

**including 884 non-inflamed and 1002 inflamed samples*

ii) RAG-LLM Metadata Curation Tool: AI-Powered Automation for Large-Scale Metadata Harmonization

We developed a RAG-based LLM method for automated metadata ingestion, using our portal's curated metadata as the ground truth for evaluation. This tool streamlines the process by allowing users to curate metadata fields with a single click.

Highlights:

- **One-click curation** of complex metadata fields
- Achieves **95%+ accuracy** in disease term classification
- **3× faster** turnaround time vs. manual methods
- Adaptable across diverse disease areas

2. Analysis Pipelines

i) scRNA Sequencing Pipeline: Comprehensive End-to-End Pipeline for Transcriptomic Exploration

Our single-cell analysis platform enables quick onboarding of new datasets, with data showcases updated monthly. Users can bring their own datasets and analyze them alongside public datasets in a secure cloud environment, while ensuring that cellular granularity is preserved throughout integrative analysis.

Highlights:

- Raw data QC, normalization, and clustering with batch correction
- Cell-type annotation powered by AI and curated reference data
- Bring-your-own-data integration with curated public datasets
- Maintains cellular granularity for accurate disease modeling

ii) Post-Translational Modifications (PTM) Pipeline: A High-Resolution Proteomics Workflow for Deeper Biological Insights

Our platform supports PTM identification and annotation, enables differential analysis across conditions, and provides pathway analysis to explore potential targets.

Highlights:

- Identification and quantification of PTM peptides via tandem MS/MS data
- Differential analysis across disease and control conditions
- Pathway mapping to highlight potential druggable targets
- Complements scRNA-based insights with proteomic precision

Case Study: Ulcerative Colitis and Vedolizumab Response

Ulcerative Colitis (UC) is a chronic inflammatory disease that lowers quality of life and raises colorectal cancer risk. Vedolizumab (VDZ), a gut selective antibody to α4β7 integrin, is widely used, yet nearly half of patients do not respond, underscoring the need for predictive biomarkers.

Using Strand's scRNA portal, we integrated two published studies with over 170,000 cells from blood and intestinal biopsies. Harmonization, QC, clustering, annotation, and differential expression were performed entirely within the portal. The analysis revealed a few blood-based markers potentially associated with VDZ response, showing how the portal enables cross-study integration, high-resolution profiling and clinically-relevant insights.

3. AI Toolkit

A Suite of Tools Designed to Accelerate Drug Target Discovery

The AI toolkit is a suite of tools built on single-cell foundation models to address bottlenecks in drug target discovery. It automates metadata ingestion and harmonization, supports annotation and dataset integration, and includes perturbation and response prediction while preserving cellular granularity. It is designed for target discovery, biomarker identification, and translational research.

Our AI toolkit brings together:

- RAG-based LLM method for automated metadata ingestion
- AI tools for annotation, integration, perturbation, and response prediction
- Standardized metadata across single-cell, bulk, and proteomics datasets
- Secure pipelines for analyzing public and in-house data

AI Toolkit in Action: From Annotation to Actionable Insights

01 Automated Cell-Type Annotation

- Quickly label datasets
- Identify rare/ novel cell types

02 Multimodal Integration

Link transcriptomics, epigenomics & proteomics

03 Gene Perturbation Modeling

Simulate single/two-gene knockouts for validation

04 Drug Response Prediction

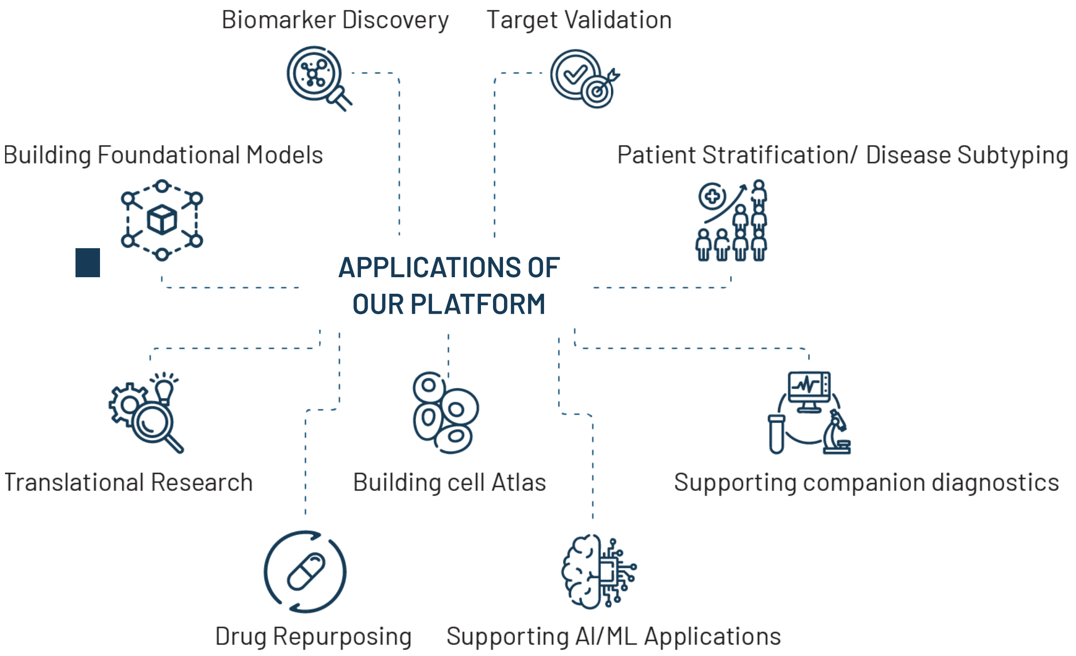
Predict responders, identify resistance mechanisms

Highlights of our AI toolkit include:

- Fine-tuned scBERT and scGPT models for accurate cell-type annotation
- Ensemble predictions that flag rare or novel cell types
- Supports hypothesis generation, biomarker discovery, and target validation
- Seamless integration with curated scRNA datasets

Applications

- Enhances data discovery and accessibility across diverse datasets
- Facilitates analysis for downstream research including biomarker discovery, patient stratification, and drug repurposing
- Supports machine learning, large language models, and AI-driven applications



APPLICATIONS OF OUR PLATFORM

Biomarker Discovery

Target Validation

Building Foundational Models

Patient Stratification/ Disease Subtyping

Translational Research

Building cell Atlas

Supporting companion diagnostics

Drug Repurposing

Supporting AI/ML Applications

Key use cases include:

- Patient stratification
- Clinical trial design
- Target discovery
- Drug response prediction

What Sets Us Apart:

- Showcase of disease-specific public datasets with **deep metadata**
- Metadata curation pipeline with **LLM automation**
- **Comprehensive AI tools** purpose-built for target discovery
- Proteomics, workflows for PTM and **differential expression analysis**
- **Quick** dataset onboarding
- **Monthly updation** of our data showcase

25+ YEARS OF EXPERIENCE

80,000+ Genetic Tests Reported

500+ Projects Executed for Genomics Majors Globally

Presence in **20+** Countries

illumina
Illumina Connected Analytics (ICA) Training and Onboarding

aws
PARTNER

strand

7th Floor, MSR North Tower, #144, Outer Ring Road, Nagavara, Bengaluru - 560045
+91 9980448044
hello@strandls.com
strandls.com