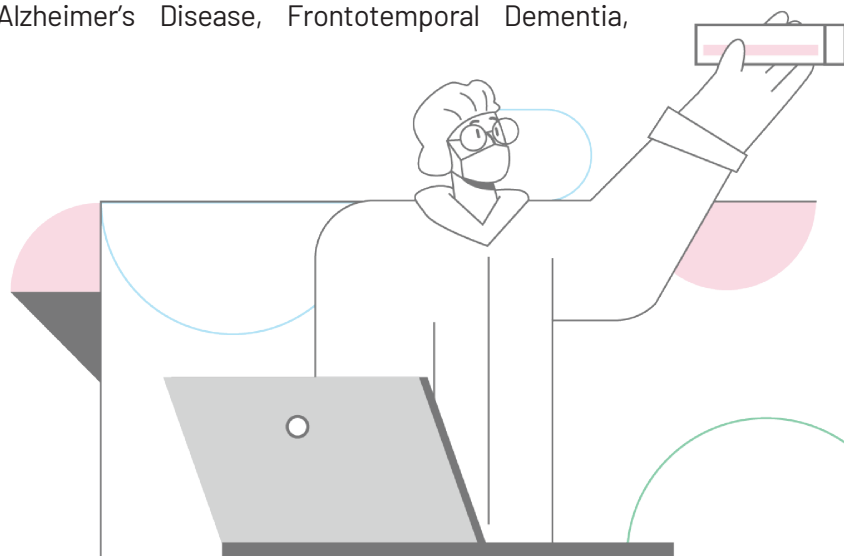


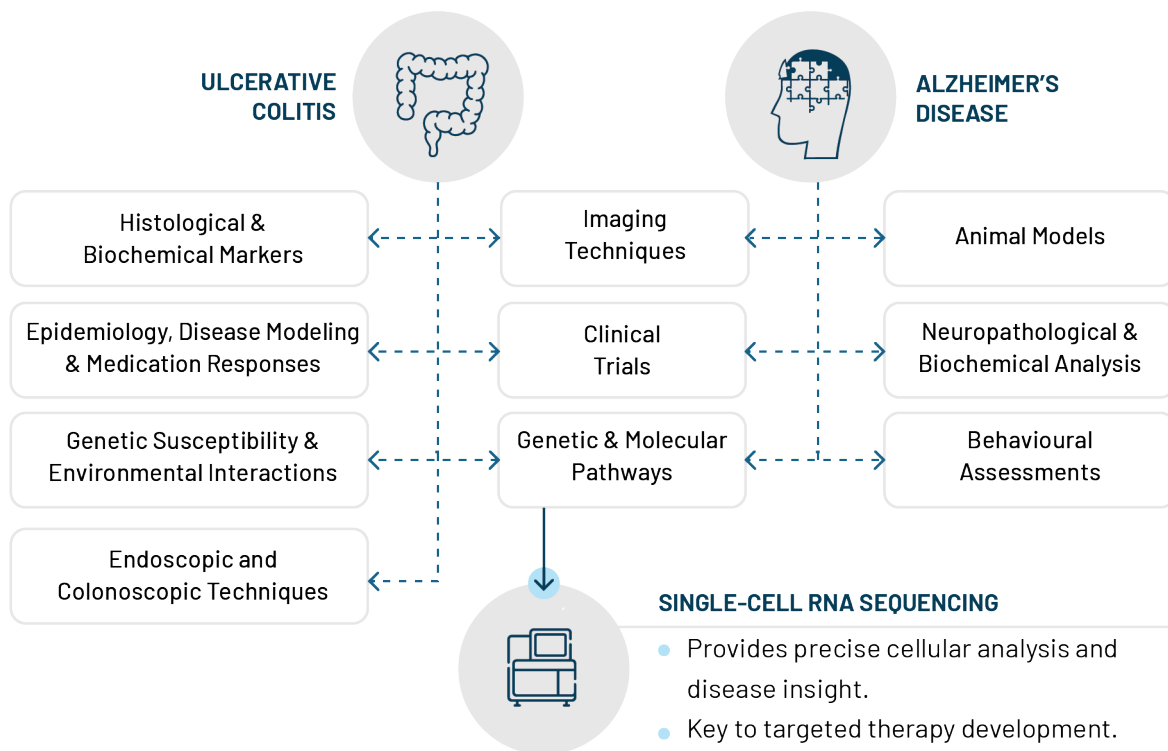
# Strand's scRNA Data Curation Portal: Bringing State-of-the-Art in Single Cell Transcriptomics to End Users

## sc-RNA Analysis Data Needs:

- Single-cell transcriptome analysis (scRNA sequencing) can reveal differences in gene expression between diseased and healthy cells at the single-cell level.
- The technique offers crucial insights into disease pathogenesis, and hence advances drug discovery.
- Increased automation and lower costs have driven a sharp rise in scRNA experiments, generating a huge volume of complex data.
- Researchers working on a specific disease need to first source relevant datasets, fill in missing pieces, and interpret raw data to extract useful information - all of which can delay a project at the outset.
- Strand has developed a curation service and 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.



# Exploring the Pathogenesis of Ulcerative Colitis & Alzheimer's Disease

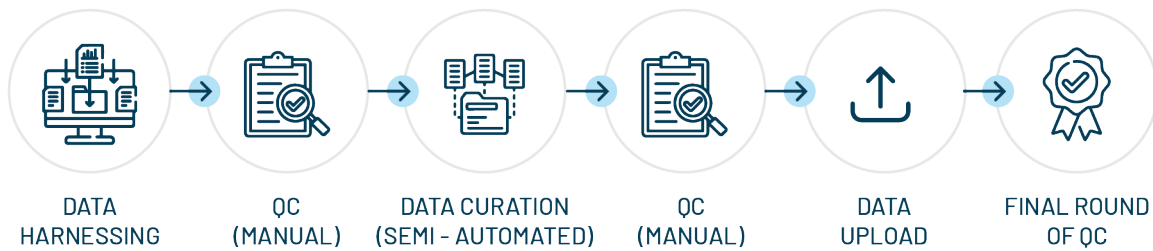


## Data Harnessing, Standardizing and Filtering:

- Largest open access scRNA display for Ulcerative Colitis and Alzheimer's disease.
- Datasets are first harnessed from publicly available resources like Gene Expression Omnibus (NCBI), then curated and harmonized, in two distinct steps.
- Independent quality control steps for both data harnessing and curation, with automated as well as manual components.
- **Turnaround time of 1-2 days.**
- **27 key filters** enable users to set attributes, categories, and values which can pick out relevant datasets within seconds.
- **54 unique metadata fields** organized at 3 levels of curation - study (11), experiment (35), and sample (54) levels, of which 28 are formatted to more rigorous internal standards.
- Controlled vocabulary using **16 metadata fields with internationally recognized ontologies.**
- 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.

## Our Workflow

### Data Curation and QC



#### 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 **44 datasets, 1144 samples, 1775138 cells, 65 distinct cell types, 23 different tissue types and 113 metadata fields**, sourced from NCBI's GEO platform.
- About **588 non-inflamed samples and 468 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 65 unique cell types.

#### 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 data from **48 studies, 607 samples, over 8 million cells, 115 different cell types, 19 different tissue types, 10 model organisms, and over 100 metadata fields**.



	UC Showcase	AD Showcase
 DATASETS	<b>44</b> sourced from GEO	<b>48</b> sourced from GEO
 SAMPLES	<b>1144</b> from 2 organisms*	<b>607</b> across 10 organisms
 CELLS CURATED	<b>&gt;1.7</b> million	<b>&gt;8.4</b> million
 UNIQUE CELL TYPES	<b>65</b>	<b>115</b>
 TISSUE TYPES	<b>23</b>	<b>19</b>
 METADATA TYPES	<b>&gt;100</b>	<b>&gt;100</b>

\*including 588 non-inflamed samples and 468 inflamed samples

## Applications:

- Enhances data discovery and accessibility.
- Facilitates data analysis and research for downstream applications including biomarker discovery, patient stratification, drug repurposing, etc.
- Supports machine learning, large learning models and AI applications.

Strand Life Sciences invites users to sign up for our portal and explore our data showcase at:

<https://scrna-curation.mystrand.org/>

We are

# Strand Life Sciences

## Our Genomics Solutions

We help you speed up the reporting of clinical NGS tests, both somatic and germline.

We enable pharmaceutical companies with clinical data and bioinformatics solutions for target discovery, translational research, and clinical research.

We help you build complex genomic software for managing and analyzing data and metadata from the entire gamut of omics modalities. We can help you accelerate releases and scale software operations.

# 24

Years of Experience

# 500+

Projects in Genomics

# 9M+

Lines of Code

# 20+

Countries



7th floor, M S R North Tower, Dr. Puneeth Rajkumar Road, Outer Ring Road, Bengaluru, Bengaluru Urban, Karnataka, 560045



+91-80-2309-5200