# SINGLE CELL
## DATA SCIENCE CONSORTIUM 2

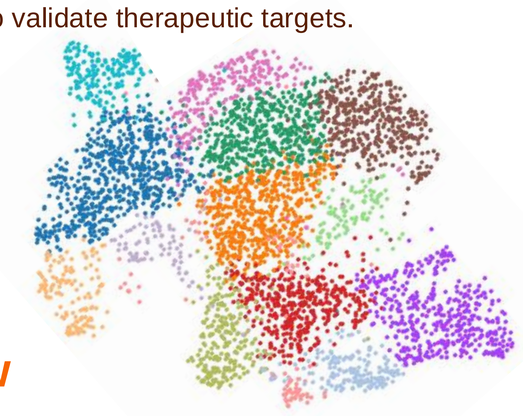
*Rancho BioSciences, LLC*
Anne Cooley, Dimitris Konstantopoulos, Panagiotis C. Agioutantis, Andrew Hill, Ben Zhang, Yang Hu, Mehmet Tekman, Domenico Cozzetto, Sondra Kopyscinski, Nicole Leyland, Andy Hope, Dan Rozelle

# Accelerating Pipeline Impact Across Therapeutic Areas!

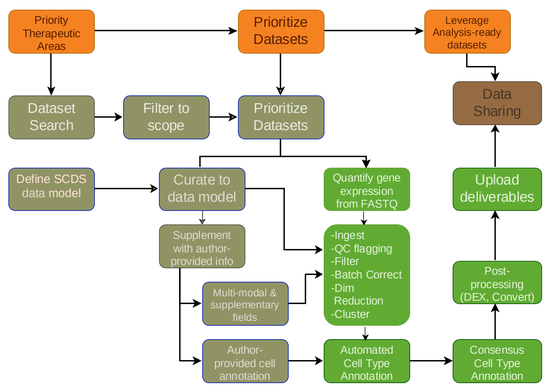
*Rancho Biosciences Single-Cell Data Science Consortium Enables Accelerated Discovery With 100-Million AI- and Analysis-Ready Single-Cells*

The Rancho Single-Cell Data Science Consortium is a member-led initiative to create resources which advance therapeutic discovery. This work has delivered 100 million AI-ready single-cells to 11 member groups. Distinguished by harmonized processing, comprehensive FAIR metadata, and advanced cell-typing, these datasets accelerate hypothesis-testing, target discovery, and model development. We demonstrate their utility by constructing a Dermatitis and IBD atlas to validate therapeutic targets.

- Shared ROI datasets
- Integrated reference atlases
- Collaborate with industry leaders
- ★★★ Highest quality, deeply annotated, fully reprocessed



## Dataset Ingestion Workflow




Identification, ingestion, and processing of single-cell datasets is a highly interdisciplinary project.
Curators are subject matter experts who maintain the data catalog, work with study authors to obtain supplementary information, and maintain the transcriptomic data model.
Bioinformatics scientists process transcriptomics data from sequencing files, filter and normalize data to optimize quality, and annotate diverse and nuanced cell types.

**100 million**
Reprocessed, and annotated single-cells

**35,932**
Batch-corrected samples

**16,591**
Biological donors

**1,000**
Analysis-ready datasets for Seurat and Scanpy

**100%**
Cell-type annotated with >1 method

**9**
Number of cell type atlases delivered

**43%**
Cells are from diseased donor samples

**48%**
of cells include author-provided annotations

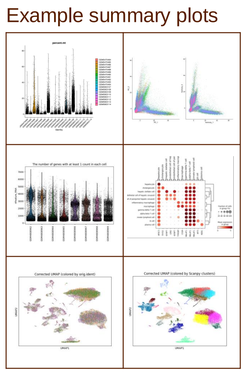
*final estimates for SCDS1 based on Feb 2025

## Deliverable Format

Deliverables included

| Name | Format |
|---|---|
| Scanpy analysis object | h5ad |
| Seurat analysis object | RDS |
| Metadata table | csv |
| DGE top table | xlsx |
| Metadata workbook | xlsx |
| README | txt |
| QC directory | misc |
| Manifest | json |

Example summary plots



Datasets are delivered in modular collections and are provided to members in-perpetuity without subscription or limitation.
Analysis-ready datasets include both python and R compatible analysis objects, allowing researchers to work with the data in their language of choice.
Comprehensive QC and primary analysis results are included.

| **7.9M** cells<br>Autoimmune and Inflammatory Diseases | **21M cells**<br>Cancer and Neoplasms | **3.0M cells**<br>Cardiovascular and Blood Disorders | **435k cells**<br>Infectious Diseases |
|---|---|---|---|
| • Crohn's disease<br>• Ulcerative colitis<br>• Psoriatic arthritis<br>• Systemic lupus erythematosus<br>• Atopic dermatitis | • Chronic myeloid leukemia<br>• Pancreatic ductal adenocarcinoma<br>• Lung non-small carcinoma<br>• Glioblastoma<br>• Hepatocellular carcinoma | • Dilated cardiomyopathy<br>• Hypertrophic cardiomyopathy<br>• Cerebrovascular disease<br>• Cardiac arrest<br>• Acute myocardial infarction | • Hepatitis B<br>• Bacterial sepsis<br>• E. Coli infection<br>• HIV disease<br>• COVID-19 |
| **975k cells**<br>Metabolic and Endocrine Diseases | **9.5M cells**<br>Neurological and Psychiatric Disorders | **1.9M**<br>Respiratory Diseases (non-oncology) | **43M cells**<br>Other Conditions and Disorders |
| • Type 1 diabetes mellitus<br>• Type 2 diabetes mellitus<br>• Obesity<br>• Non-alcoholic fatty liver<br>• Acute pancreatitis | • Alzheimer's disease<br>• Parkinson's disease<br>• Multiple sclerosis<br>• Huntington's disease<br>• Diabetic neuropathy | • Idiopathic pulmonary fibrosis<br>• chronic obstructive pulmonary disease<br>• Allergic asthma<br>• Chronic asthma | • Healthy subject<br>• Unannotated<br>• Endometriosis<br>• Injury<br>• Stroke |

## Datasets are Analysis-Ready and AI-Ready

**Is the data prepared?**

- **Organized** — Systematic and accessible by AI algorithms.
- **Clean** — Errors and missing values are removed. Data is normalized.
- **Consistent** — Scoped and processed in a standardized way.
- **Interoperable** — Format and data should be easily integrated with external datasets.

**Is the data appropriate?**

- **Sufficient** — Enough data to produce meaningful results.
- **Representative** — Representative of the domain.
- **Descriptive** — Includes features that are informative for the task.

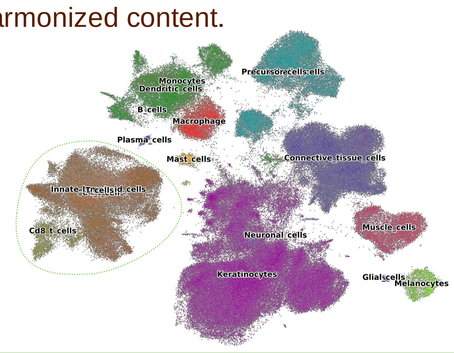
Significant effort has been spent documenting data governance, developing defined processes, and building validation tools to ensure data is analysis-ready. This includes elements of FAIR principles.
In addition to data preparation, before an analysis we also ensure data is appropriate to the task. In examples given below, we describe a task such as development of a disease atlas, but this also includes training a predictive model and tasks such as target validation.

## Rapid Single-Cell Atlas Development for Therapeutic Target Discovery

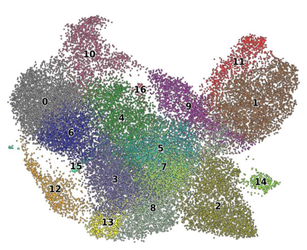
Single-cell RNA sequencing offers a powerful tool to identify therapeutic targets by providing insights into cellular heterogeneity in various diseases. Datasets from diverse sources are difficult to integrate due to inconsistent metadata, formats and batch effects. Analysis-ready SCDS datasets accelerate this process by providing harmonized content.

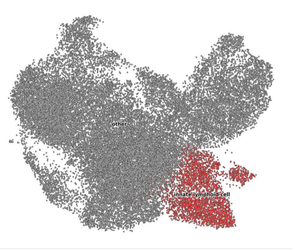
### Autoimmune Skin Atlas

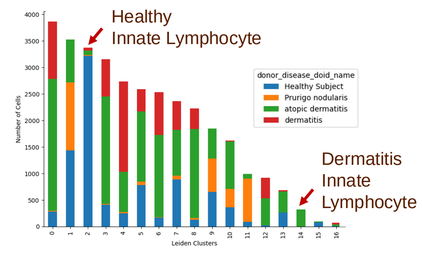
By integrating four single-cell datasets, we generated a 250k skin cell atlas with good representation in dermatitis, an inflammatory skin disease. Datasets were integrated using the scVI Variable Autoencoder and highly variable genes were used to inform scVI embeddings and Minimum-Distortion Embedding representation. The resulting atlas showcases the diverse cellular populations found in the skin.




**Immune sub-clustering reveals distinct AD population**





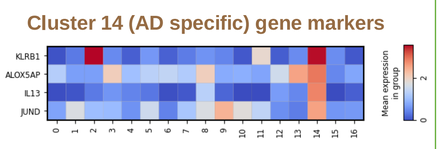



### Therapeutic Targets in Dermatitis

Dermatitis is characterized by chronic inflammation and skin barrier dysfunction. We isolated and re-clustered a collection of immune-related clusters to identify disease-specific subpopulations. Several key biomarkers are differentially expressed in the inflammatory context of dermatitis. Notably, we validated existing therapeutic targets (IL13, PDE4D, IL4R, IL31).

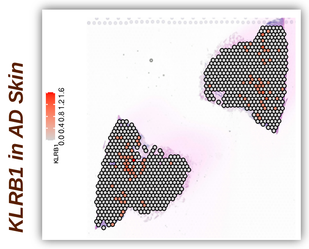
Additionally, we identified novel therapeutic targets (IL26, IL32) not currently addressed by treatments.

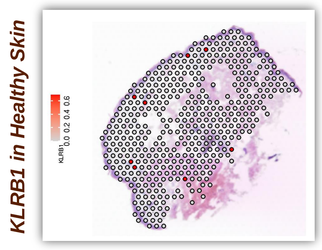
**Cluster 14 (AD specific) gene markers**




**Therapeutic targets from Cluster 14**

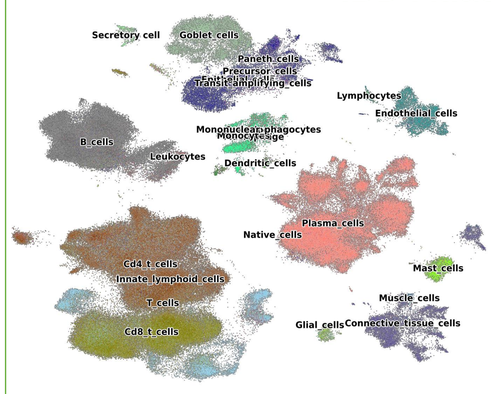
| Gene | DEG Score | Adj. p-value | Drug |
|---|---|---|---|
| ALOX5AP | 53.87 | 5.45E-205 | |
| S100A11 | 37.96 | 7.22E-147 | |
| PTPRC | 37.70 | 3.91E-148 | |
| JUND | 33.24 | 5.37E-122 | |
| IL13 | 30.49 | 1.64E-97 | Tralokinumab, Lebrikizumab |
| **IL26** | 26.26 | 1.58E-79 | Potential target |
| PDE4D | 12.18 | 4.11E-27 | Roflumilast, Crisaborole |
| **IL32** | 11.83 | 1.03E-26 | Potential target |
| IL4R | 10.69 | 1.65E-21 | Dupilumab |
| IL31 | 6.769 | 1.41E-09 | Nemolizumab |




To corroborate the findings from our scRNA-seq analysis, we checked an unrelated spatial transcriptomics dataset with Dermatitis skin samples. This comparison reveals a clear, disease-specific differences in lymphocyte populations between dermatitis and healthy samples.

### IBD Atlas: Therapeutic Targets in Crohn's and UC



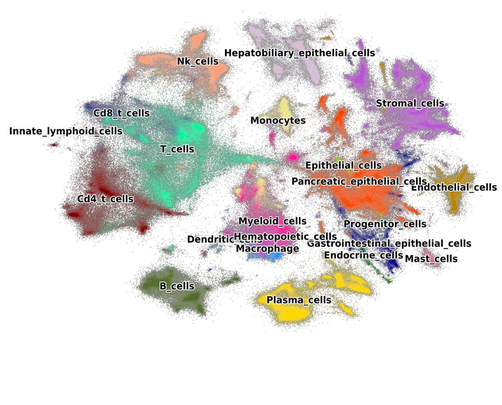

In addition to dermatologic diseases, we demonstrate similar value for inflammatory bowel diseases (IBD), including Crohn's disease and ulcerative colitis (UC). Utilizing similar methods, we integrated 8 single-cell datasets from healthy and diseased gastrointestinal tissues to construct a 360k atlas.

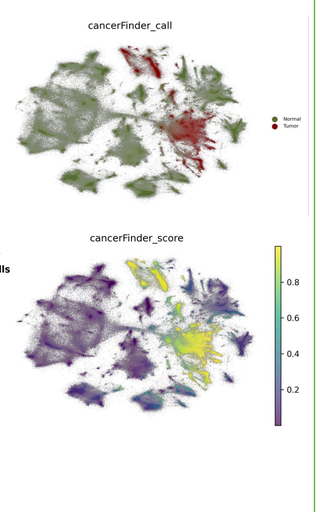
In the context of IBD, we identified several therapeutic targets, including current therapeutic targets.

| Gene | DEG Score | Adj p-value | Drug targeting gene |
|---|---|---|---|
| SKAP1 | 26.84 | 4.22E-156 | Potential target |
| SIK3 | 25.79 | 3.36E-144 | Potential target |
| PDE4D | 19.56 | 6.68E-83 | Potential target |
| IL12RB2 | 9.96 | 1.03E-21 | Ustekinumab (Stelara), Briakinumab |

### 1.6M Cell Digestive System Cancer Atlas

Following our investigation of inflammatory diseases, we expanded our focus to other disease areas by creating a comprehensive digestive system cancer atlas. This atlas integrates data from 18 datasets and 303 samples, encompassing 10 cancer types, 6 tissue types, and a total of 1.6 million cells.
In addition to sample-level disease status, we enriched the atlas for cell-level malignancy score. Using an ML marker-based approach (Cancer-Finder) we scored each cell and identified distinct cancer clusters.




This comprehensive approach demonstrates the power of scRNA-seq in the discovery of biomarkers and therapeutic targets for complex diseases. By leveraging large-scale, integrated datasets from both dermatitis and IBD, we not only confirmed existing therapeutic targets but also identified novel candidates that hold promise for improving patient outcomes. Our findings underscore the potential of single-cell data as indispensable tools in precision medicine and drug discovery.

Rancho BioSciences

Learn More