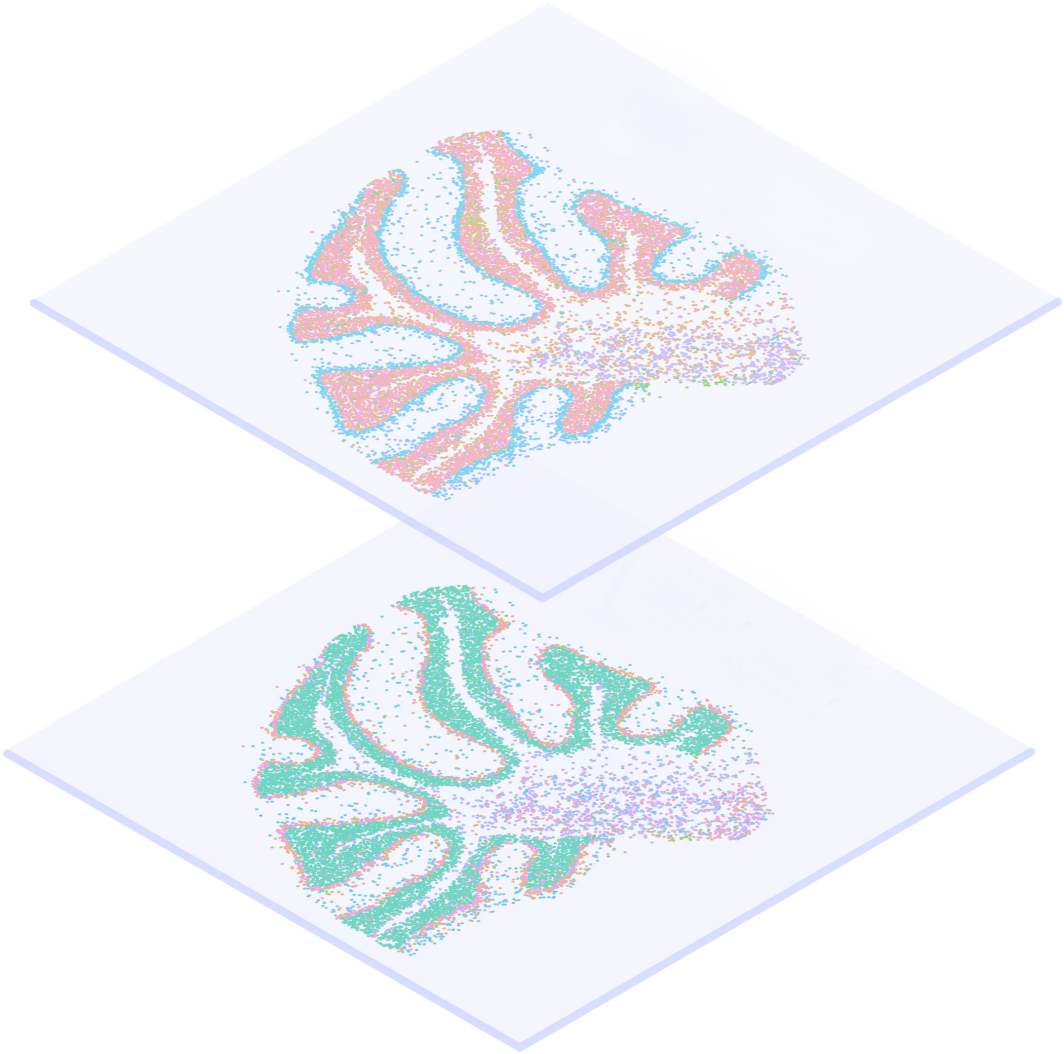
strand

Transcriptomics Data Analysis And Pipelines



#1 Slide-Seq Spatial Transcriptomics Workflow

The Problem

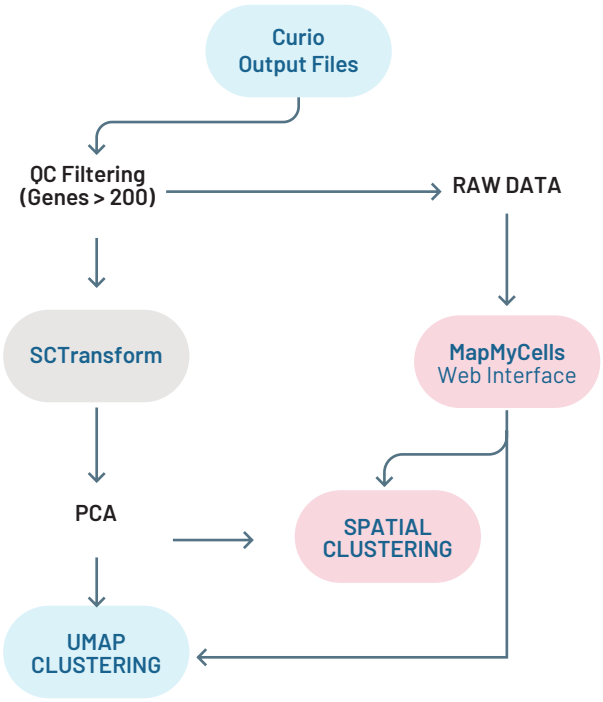
The impact of a microRNA linked to schizophrenia in the cerebellum

- Technology: Curio Seek (CS) Spatial Transcriptomics Mapping kit based on Slide-SeqV2
- The Challenge: The tools available could only analyze a single sample at a time
  - The Need: A method that compares different sample conditions (KO vs WT) and identifies common cell signatures/gene patterns

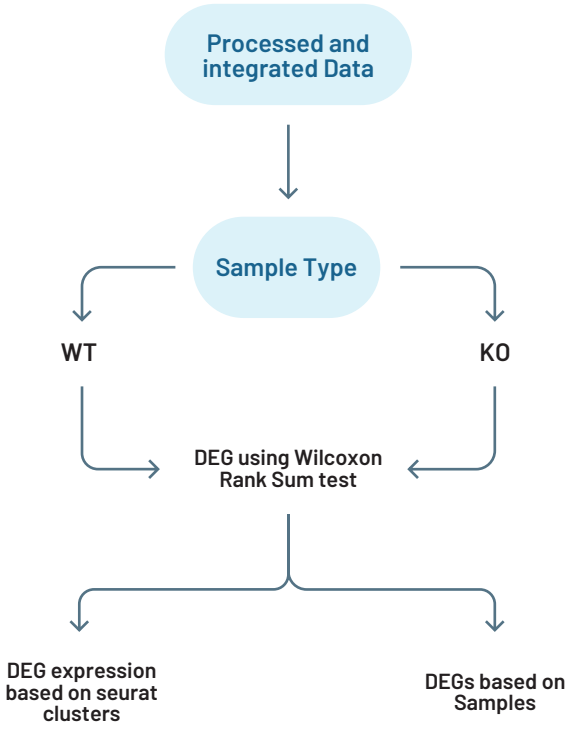
Our Process

- Development of an Integrated Slide-Seq Spatial transcriptomics analysis workflow to evaluate dataset heterogeneity and cross-condition comparative analyses

Implementation and tuning of Seurat workflows for Slide-Seq → **Standardization of cell type identification** → **Visualization**



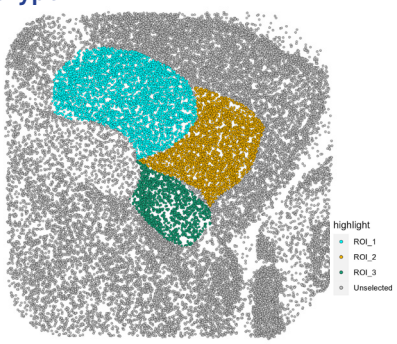
We developed a workflow for analyzing CS data including QC to filter out outliers and retain high-quality data, normalization and dimensional reduction. MayMyCells was employed to obtain cell-type annotations for the data.



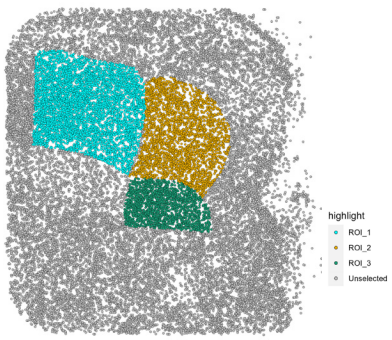
We combined data across sample conditions and identified the top differentially expressed genes within and between WT and KO sample brain regions

Webninar by Dr. Molly Heyer, Mt.Sinai

Wildtype



KO

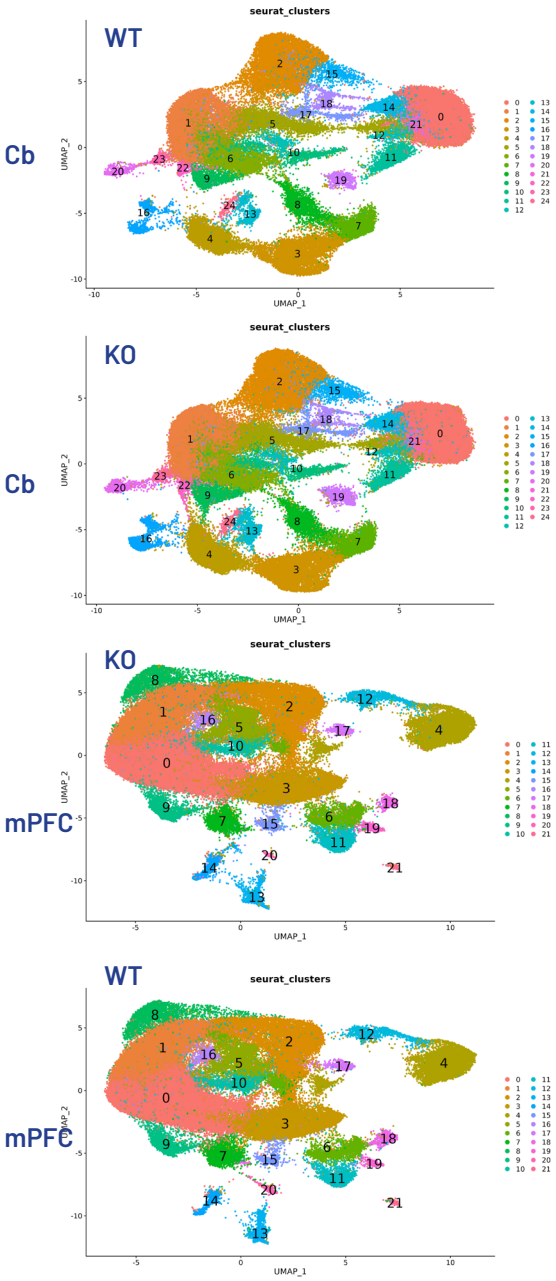


The team developed a novel ROI analysis pipeline for Slide-seq data, enabling precise gene expression analysis in specific brain regions, such as the deep cerebellar nuclei (DCN). This approach allows for detailed comparisons between experimental groups (WT vs. KO), assessment of specific marker effects in targeted regions, and clustering within ROIs.

Result / Impact

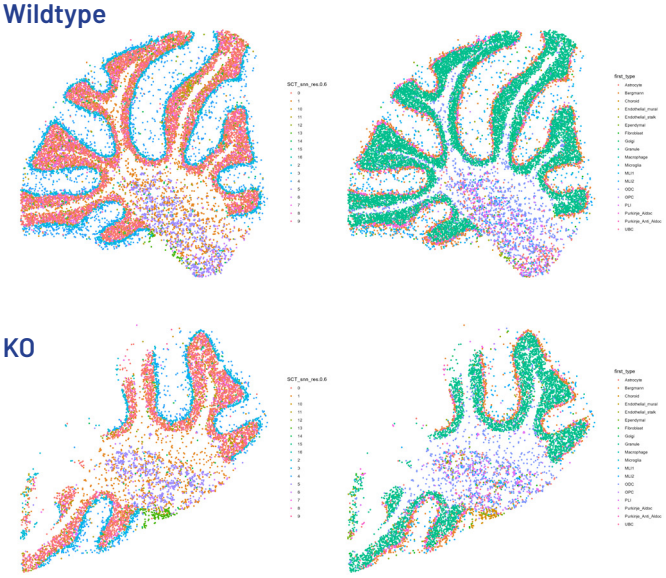
- Following preprocessing and integration, our UMAP-based clustering analysis revealed that data from all combined biological replicates of WT and KO medial prefrontal cortex (mPFC) samples from both tissues were distinctly separated into individual clusters
- ROI-specific gene expression analysis was developed for the first time for Slide-seq technology
- After miRNA KO, the cerebellum's Purkinje neurons showed upregulation of 3-4 genes that influence calcium signaling and may be linked to hyperactivity and schizophrenia-like characteristics

Cerebellum and mPFC Clustering (integrated)



UMAP plots generated for Cb and mPFC show that tissue-wise gene expression is comparable

Wildtype



KO

Final images of WT and KO mPFC on tissue slides with annotated cell types, confirming high-quality integration and the presence of diverse cell types across the samples

Strand is improving the pipeline to analyze multiple datasets in parallel while incorporating additional visualization and selection modules

#2 Comparative Analysis Of The Nanostring Atomx Pipeline Vs. Strand Pipeline

The Problem

Immune response to viral infection in lungs, comparison between young and aged individuals.
Issues with the AtoMx pipeline:

- Missed cells - The cell segmentation pipeline missed a high percentage of cells (blurry tissue image/segmentation staining)
- False positives - Many cells that did not appear in the histopath images were called out within the boundaries marked by the cell segmentation algorithm, which turned out to be negative for transcripts (could be high amount of extracellular material)
- All tissue sections on a single slide are treated as one sample, making independent analysis of different tissue sections unfeasible

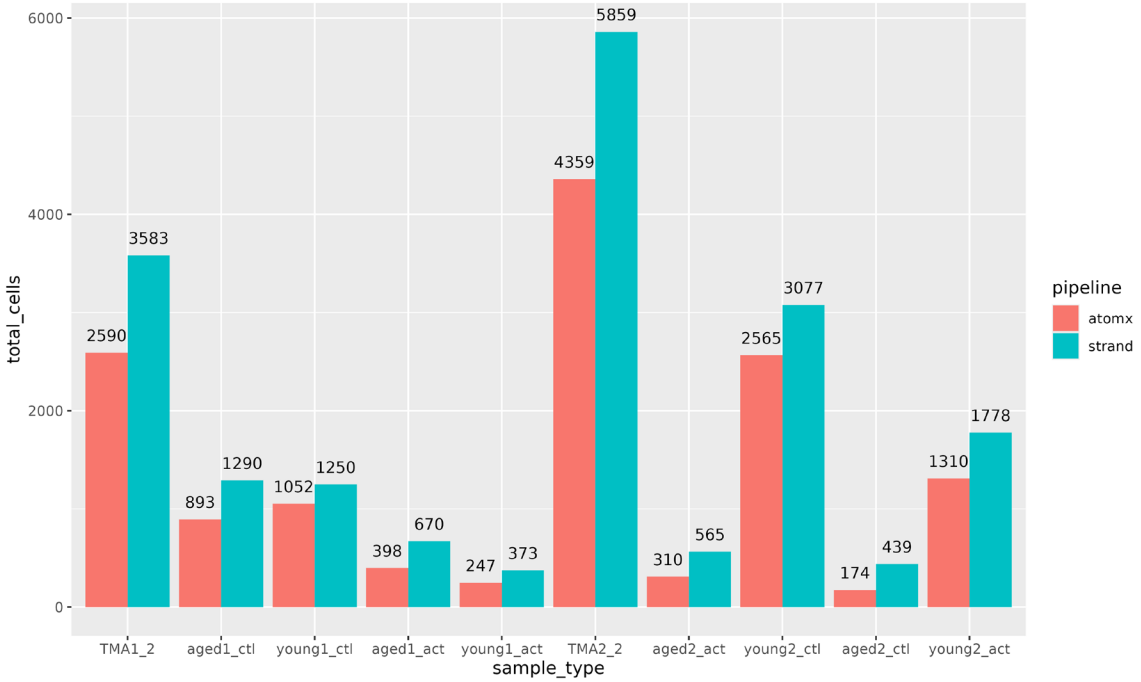
The Need : Compare the AtoMx and Strand pipeline outputs for CosMx spatial transcriptomics analysis on the same samples, to determine if the Strand pipeline offers improvements over the AtoMx pipeline

Our Process

- Cell segmentation enhancement using Cellpose3, addressing issues like missing cells and false positives (completed)
- Independent analysis of each tissue sample using the Strand Pipeline (completed)
- Cell type segmentation comparison between the AtoMx and Strand pipelines to verify accurate cell identification (completed)
- Comparative analysis of other module outputs such as cluster-wise DEG, pathway analysis and cell annotation, in AtoMx and Strand pipelines to assess similarities and differences in results (ongoing)

CosMx images for individual FOV → Cell Segmentation → Data → Quality Control → Downstream Analysis → Annotation & Visualization

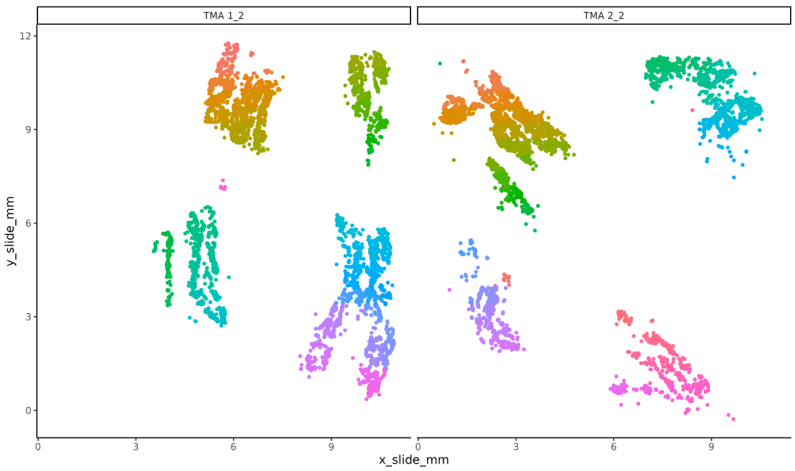
Strand's R-based Seurat workflow designed to process CosMx spatial data



Comparative analysis shows that the Strand pipeline captures a significantly higher number of cells for CosMx than the AtoMx pipeline.

Result / Impact

- Enhanced Cell Segmentation - the Strand pipeline captures significantly higher number of actual cells, improving analysis accuracy and providing deeper insights into cellular characteristics
- Quality Control - The Strand pipeline captured most cells positive for transcript expression and filtered out cells with fewer than 10 transcripts per cell, enhancing data accuracy
- Condition-Specific Analyses - The Strand pipeline was tailored to support condition-specific comparative analyses, enhancing the ability to address the client's specific research questions (young vs aged)



TMA-1 (slide 1)

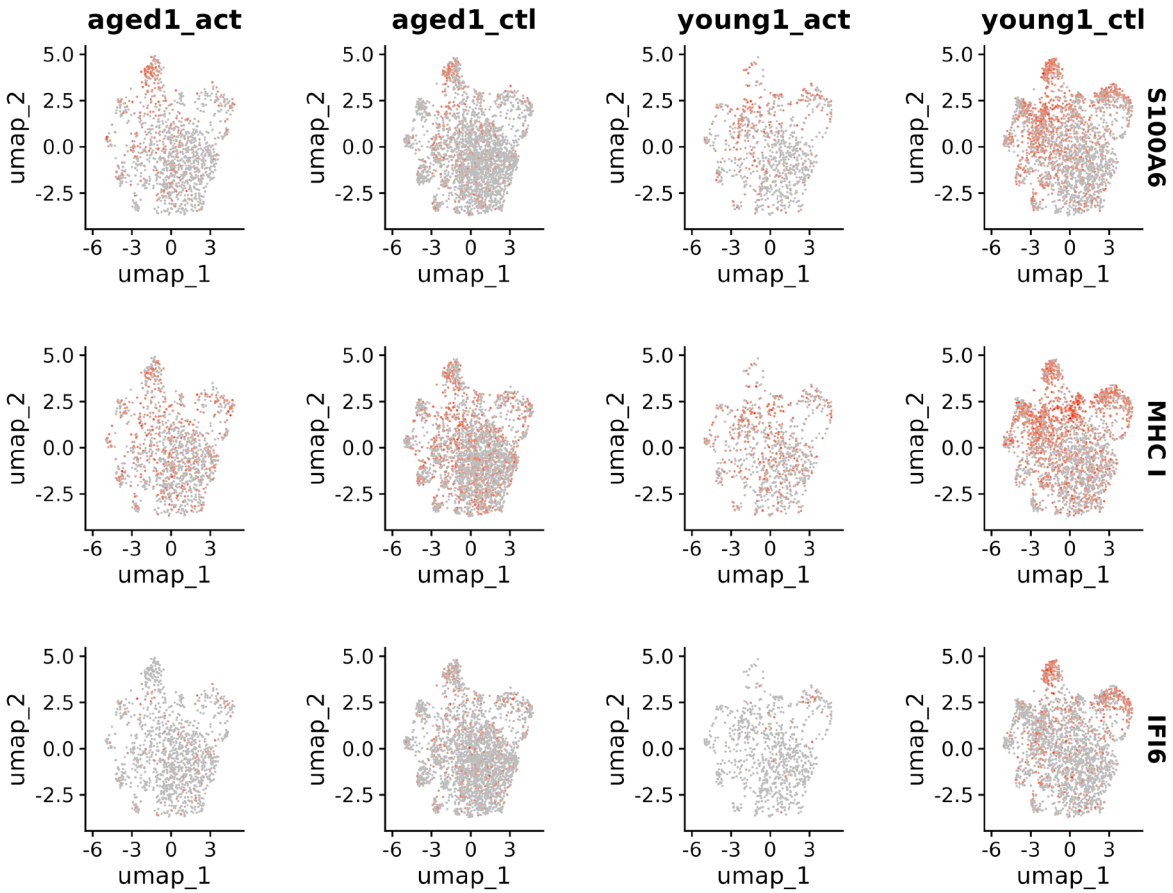
Young 1 Control	Young 1 activated
Aged 1 Control	Aged 1 activated

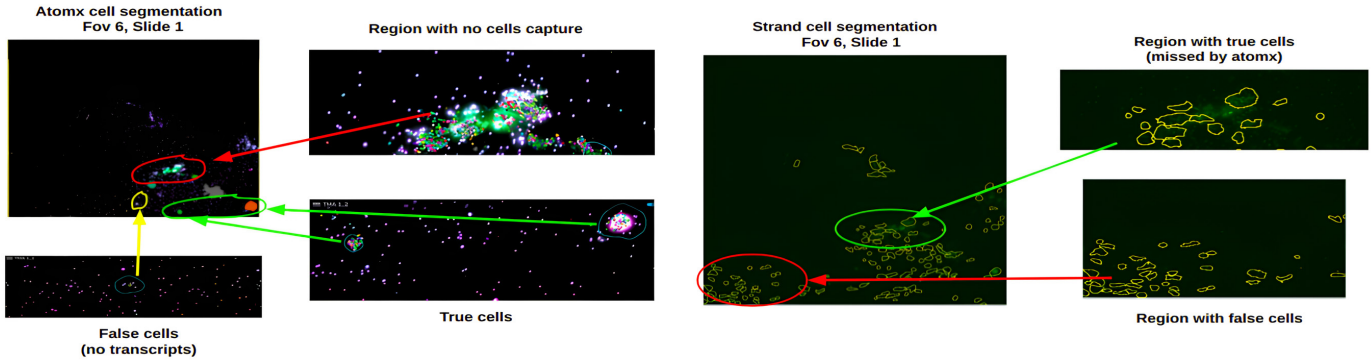
TMA-2 (slide 2)

Young 2 Control	Young 2 activated
Aged 2 Control	Aged 2 activated

The Atomx pipeline treats all four samples as a single entity during analysis. In contrast, the Strand pipeline offers the functionality to perform analyses on a sample-wise basis

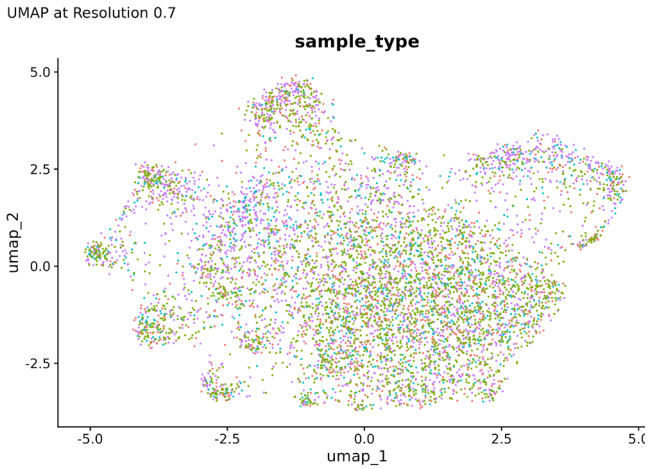
Cluster marker visualization using UMAP of individual samples



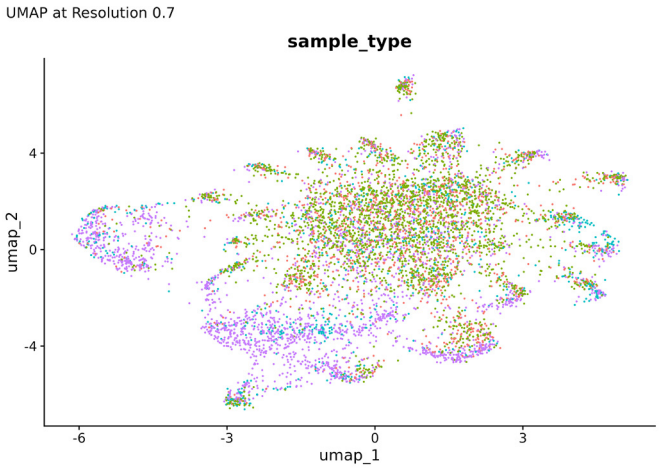


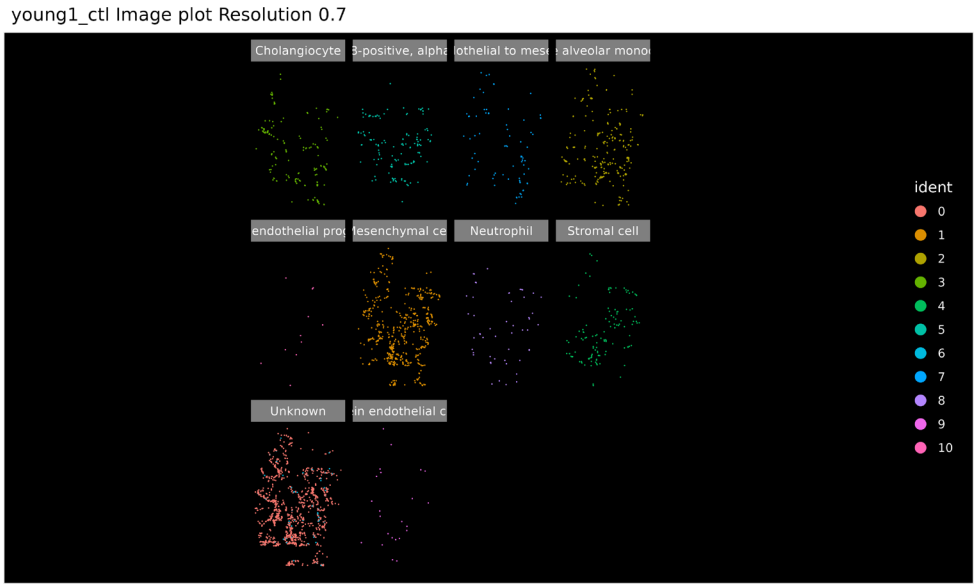
The Strand pipeline, powered by Cellpose3, captured more cells and successfully identified high-transcript cells missed by the AtoMx pipeline. Both pipelines initially detected low-transcript cells, which were filtered out by the Strand pipeline before further analysis.

Clustering before batch effect correction



Clustering after batch effect correction



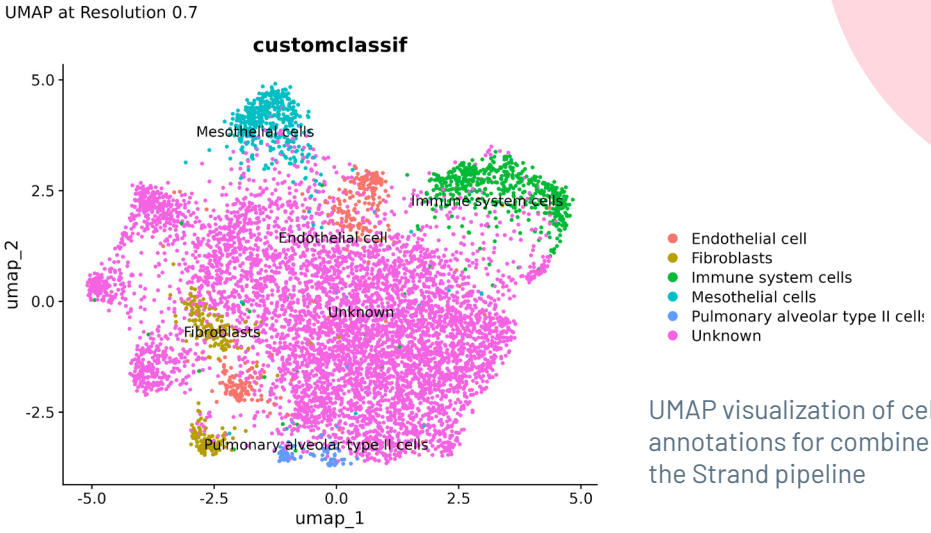


Spatial visualization of cell-type annotations on tissue slide image called out using the Strand pipeline

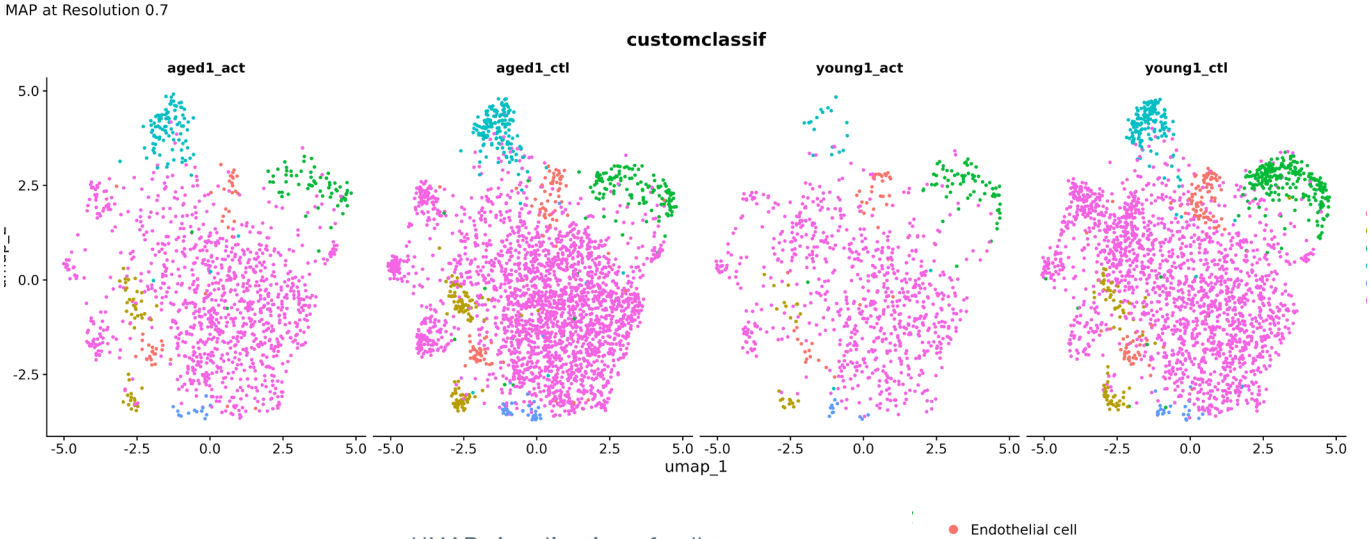
Result / Impact

- Robust Analysis and visualization:
- Batch Correction - Integration of algorithms for batch effect correction (CCA, Harmony and Liger) ensured that differences reflected true biological variation (rather than technical artifacts), enhancing pipeline robustness
- Differentially Expressed Genes (DEGs) - The pipeline was tailored to identify DEGs within and across clusters, providing heatmaps, violin plots, and UMAP visualizations.
- Cell Type Annotation - Integration of advanced annotation algorithms like ScType, SingleR, and Seurat label transfer clarified cell types, while UMAP plots and spatial images enabled precise localization on tissue slides.

The Strand pipeline effectively addresses the limitations of the AtoMx pipeline



UMAP visualization of cell-type annotations for combined samples with the Strand pipeline



UMAP visualization of cell-type annotations across different sample conditions with the Strand pipeline

#3 CRISPR Assay Refinement

The Problem

The Need: Develop/Improve Cas9 off-target specificity for RUO CRISPR assays by running in-house assays and analyzing sequenced data

Our Process

- Benchmarking, designing and execution of pipelines for
  - Guide RNA design
  - Off-target prediction
- Comparison of the performance of 12 publicly available Cas9 off-target prediction tools, which included the client's in-house developed tool, against empirically generated GUIDE-Seq assay data to identify the tool with the best sensitivity

#4 Companion Diagnostics insights from publicly available data

The Problem

The company's pipeline includes microbiota-activated gut-bound 'target' inhibitor drugs for moderate-to- severe Ulcerative Colitis (UC) and fibrostenotic Crohn's Diseases

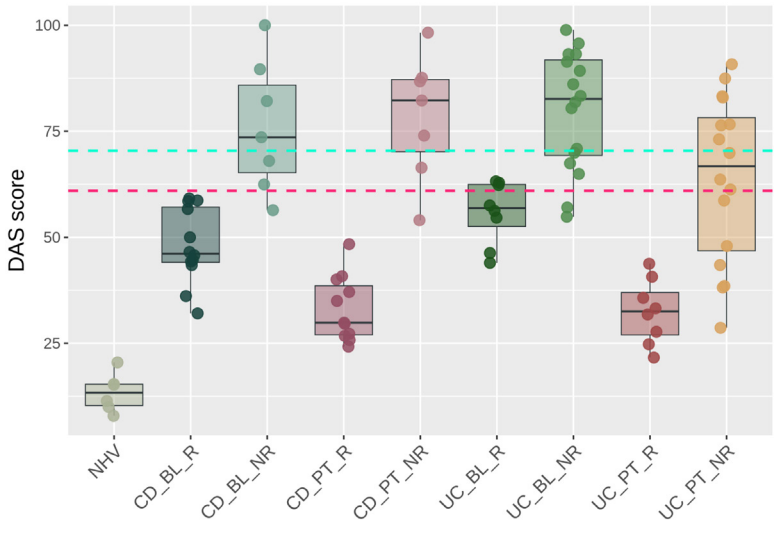
- The Need: Biomarker identification from publicly available datasets to identify patients receptive to therapy, with the goal of developing a PCR assay .

Our Process

- Integrated workflow that involved:
  - Automation of the search and analysis of ~2,000 RNA-Seq samples and 950 microarray samples, including the client's in-house samples
  - Building and refining a Disease Activity Score (DAS) through ML on responder/non-responder data, and verifying it using a published method

Result / Impact

- The automated search and filtering process began with 2,600 datasets and narrowed down to 7 that met all criteria for meta-analysis.
- An initial set of 40 markers from the meta-analysis was used to create a Disease Activity Score (DAS) with 9 markers, achieving up to 70% specificity (adult dataset).
- Analysis of the microarray dataset identified 53 markers, contributing to marker-list refinement.
- The DAS was refined using 5 markers (70% sensitivity).
- Employing an independent method published by a large pharma company achieved 84% sensitivity and 70% specificity (i_MDS score).



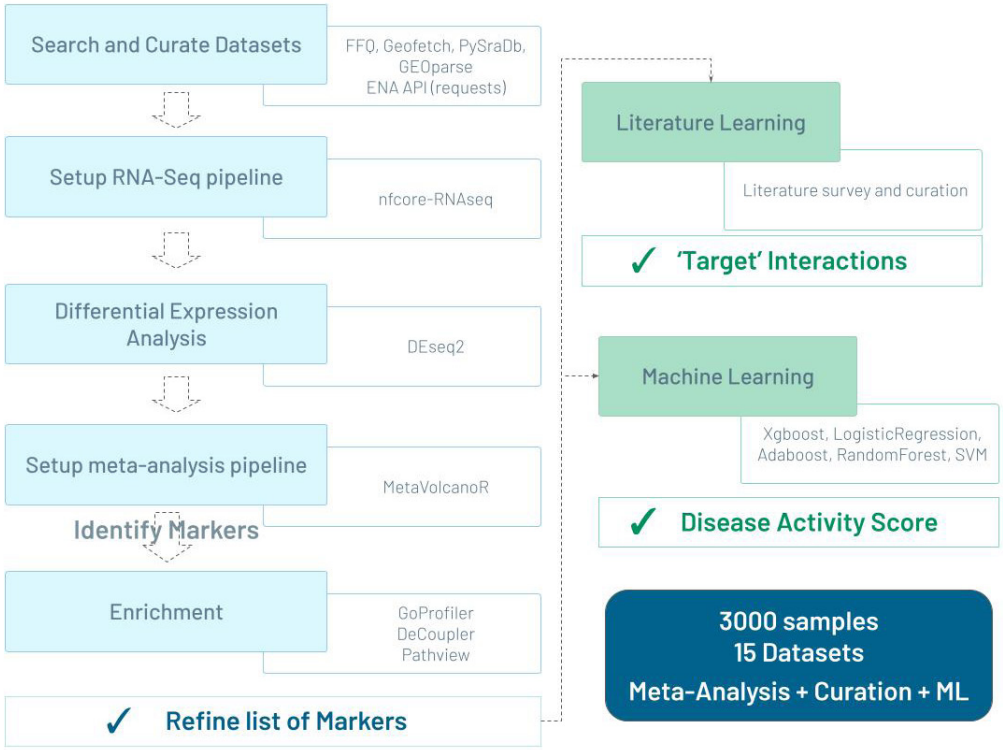
For a large manufacturer of sequencing reagents

Result / Impact

- Confirmed that the client's tool showed 96% sensitivity with GUIDE-Seq
- Helped build confidence in the client's lab science department to use the in-house computation tool as a reliable surrogate for GUIDE-Seq
- Helped refine the lab fine tune their strategy towards GUIDE-Seq experimental protocol for improved off-target prediction specificity

Strand has built and improved the client's CRISPR/Cas9 assays with validated pipelines and highly performing code.

For a therapeutic biotech investigating intestinal disorder drug targets



The client is interested in the gene cluster with strong expression correlations and marker potential to boost the chances of positive PCR hits for patient segregation.

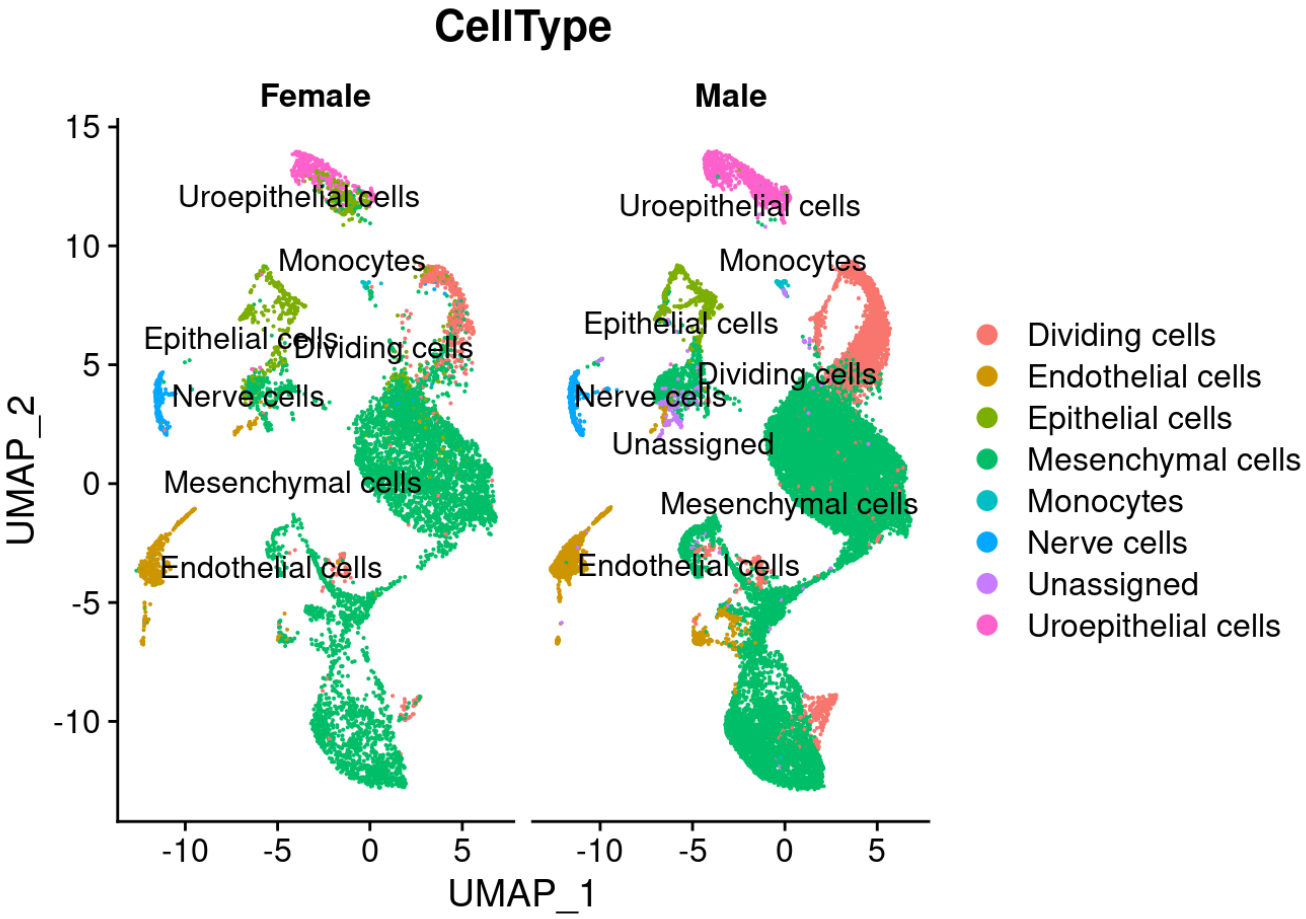
#5 Single Cell Data Analysis Using Open Source Tools In Strandngs

The Problem

- Analysis of single nuclei data generated from human embryos at different developmental stages
- The Need: An in-depth analysis of the crucial human embryo dataset, focusing on cell progression through different stages of development, including cell division, differentiation, and cell-type lineage evolution

Our Process

- Evaluation of mitochondrial/ribosomal content discrepancies to verify data quality
- Clustering analysis with reference to cell type atlas scores and cell cycle gene patterns (Seurat, SingleR)
- Comparative evaluation across various groups to identify conserved markers and differential expression patterns (R Seurat, Conos)
- Analysis of the trajectory/lineage of the cells in the sample based on the cell stage (Monocle)

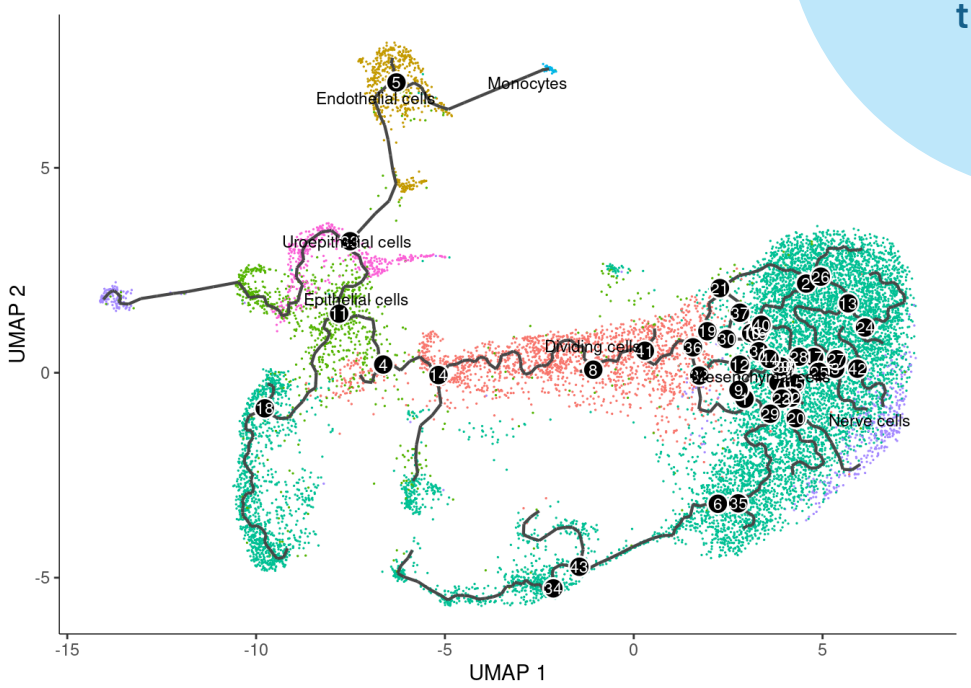


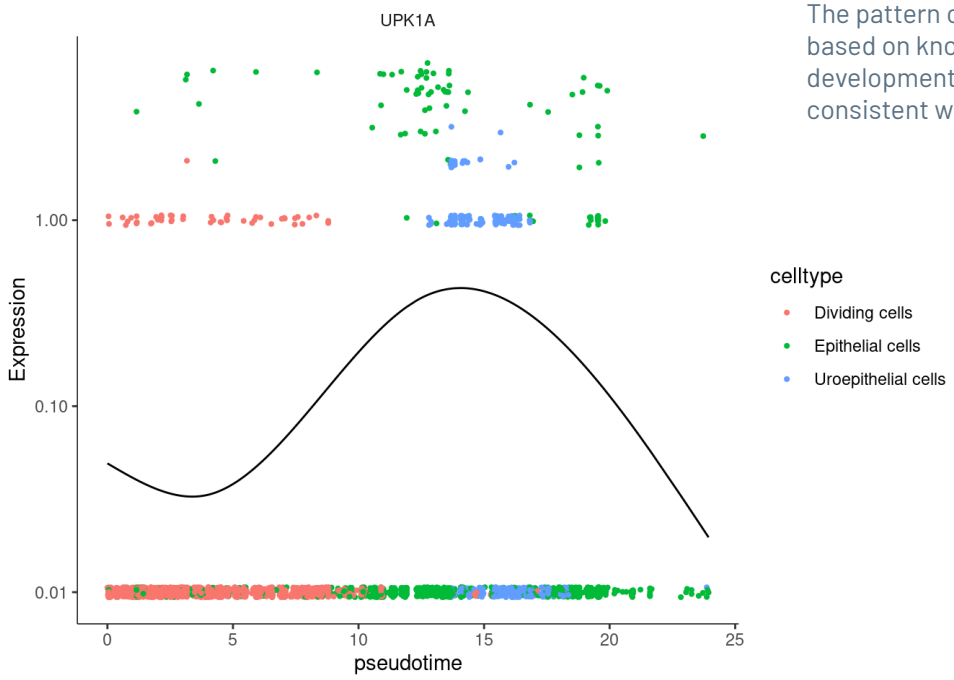
Cell type-based clustering by sample group (female vs. male human embryos at week 11 of development) aligned with previous experimental findings.

Result / Impact

- Identified one sample that had to be resequenced, which was later added to the dataset
- The markers identified in previous experimental datasets correlated with the clustering analysis findings
- Clustering patterns were consistent with the cell type lineage within the developmental stage-wise trajectory observed

The client chose Strand for the in-depth analysis of their crucial single nuclei sequencing dataset from human embryos, following the pilot study.





The pattern of cell lineage and development, based on known marker genes across developmental stages, followed a trajectory consistent with earlier research findings

Omics CRO

Curation

15 years of experience curating variants, genes, pathways and diseases for clinical reporting and pharma/biotech custom solutions

~50
Molecular Biologists

Bioinformatics and Software

22 years of experience providing bioinformatics solutions to global instrument, diagnostic and pharma companies

~220
SW Engineers, Bioinformaticians

Omics Assays

11 years of experience with sequencing-based diagnostics across oncology and genetics, at our CAP lab in India

~90
Lab Scientists, Clin. Res. Scientists

24+ YEARS OF EXPERIENCE

80,000+ Genetic Tests Reported

500+ Projects Executed for Genomics Majors Globally

Presence in **20+** Countries

strand

7th Floor, MSR North Tower, #144, Outer Ring Road, Nagavara, Bengaluru - 560045

+91 9980448044 hello@strandls.com strandls.com