

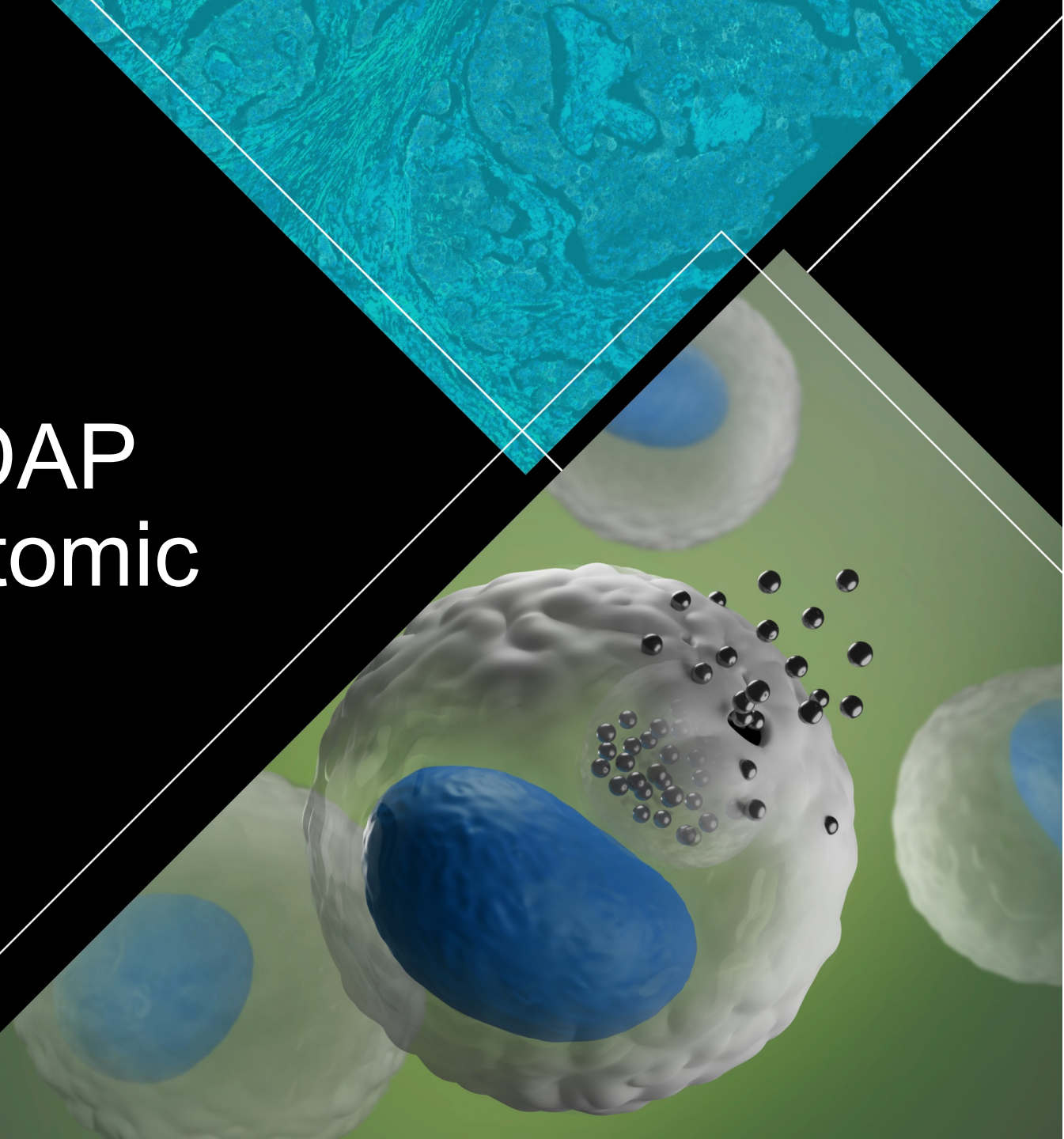
Capabilities of the NIDAP Platform for Transcriptomic Analysis

BTEP Webinar

Dec. 14th, 2023

Joshua Meyer & Ned Cauley

SPONSORED BY THE NATIONAL CANCER INSTITUTE



Who are we?

CCBR: CCR Collaborative Bioinformatics Resource

CCR Collaborative Bioinformatics Resource



CCR COLLABORATIVE BIOINFORMATICS RESOURCE (CCBR)

The CCR Collaborative Bioinformatics Resource (CCBR) is a resource group which provides a mechanism for CCR researchers to obtain many different types of bioinformatics assistance to further their research goals. The group has expertise in a broad range of bioinformatics topics, and as such, its goal is to provide a simplified central access point for CCR researchers.

The CCBR group includes members of the CCR Office of Science and Technology Resources (OSTR), Frederick National Laboratory for Cancer Research (FNLCR) and the Center for Biomedical Informatics

ASK FOR HELP

Reach Out
to CCBR



UPCOMING CLASSES

BTEP: Sridhar Hannehalli (CDSL),
Distinguished Speakers Seminar
Series

RNA-Seq: Introduction to Downstream
Analysis using CCBR Workflows on the
Palantir Platform

Single-Cell RNA-Seq: Introduction to
Downstream Analysis using CCBR
Workflows on the Palantir Platform

RECENT CCBR PUBLICATIONS

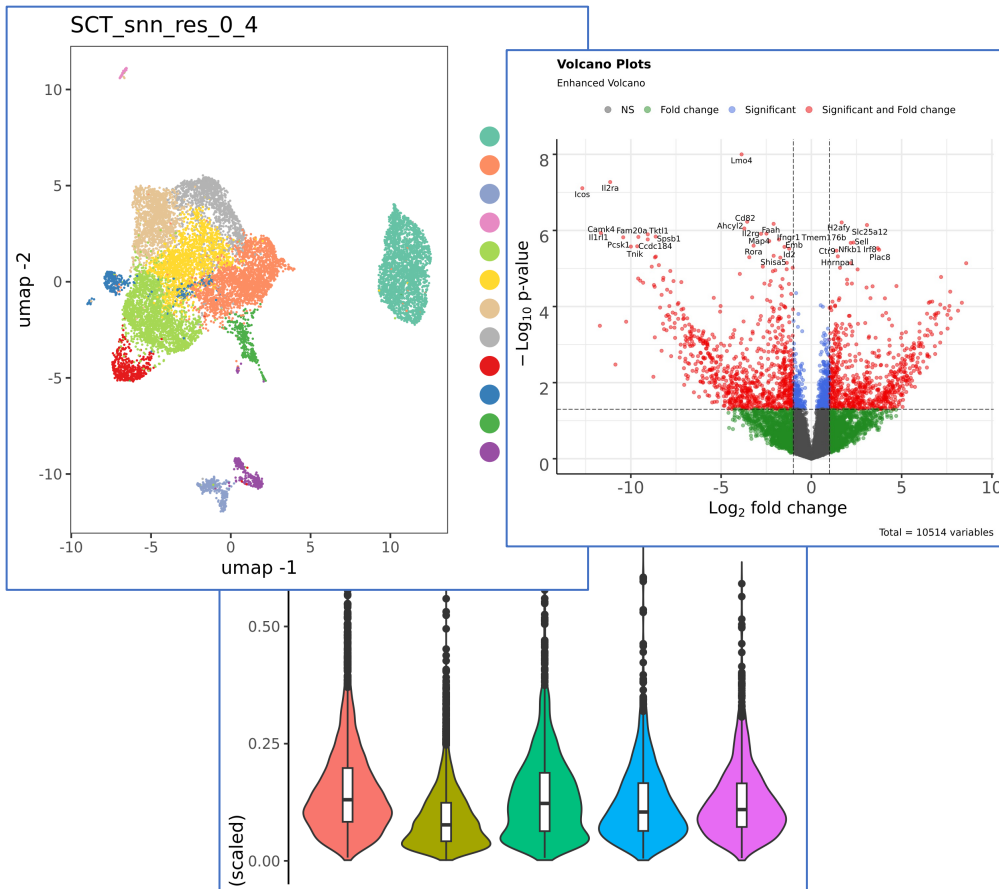
05/01/2019 - Relationship between
human leukocyte antigen alleles

- Experimental Design
- Training on genomics analysis
- Support on NIDAP platform for RNA-seq, single-cell RNA-seq, and spatial transcriptomics research projects
- Customized support for “non-standard” analysis
- <https://ccbr.ccr.cancer.gov/>

Overview

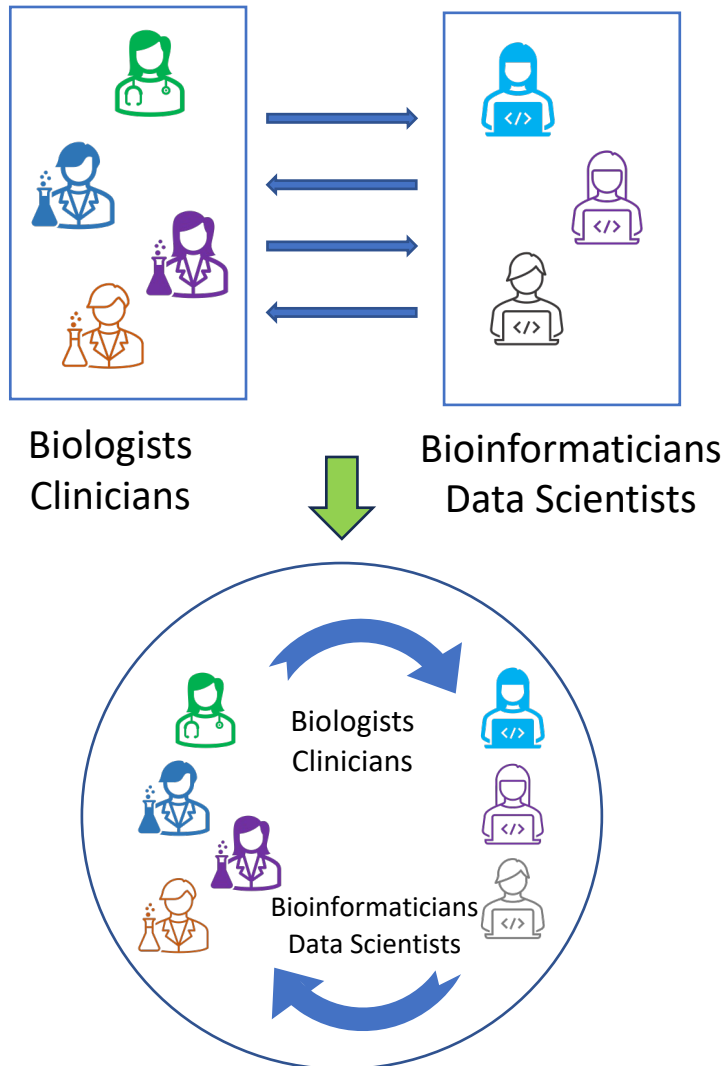


NATIONAL CANCER INSTITUTE
NIH Integrated Data Analysis Platform



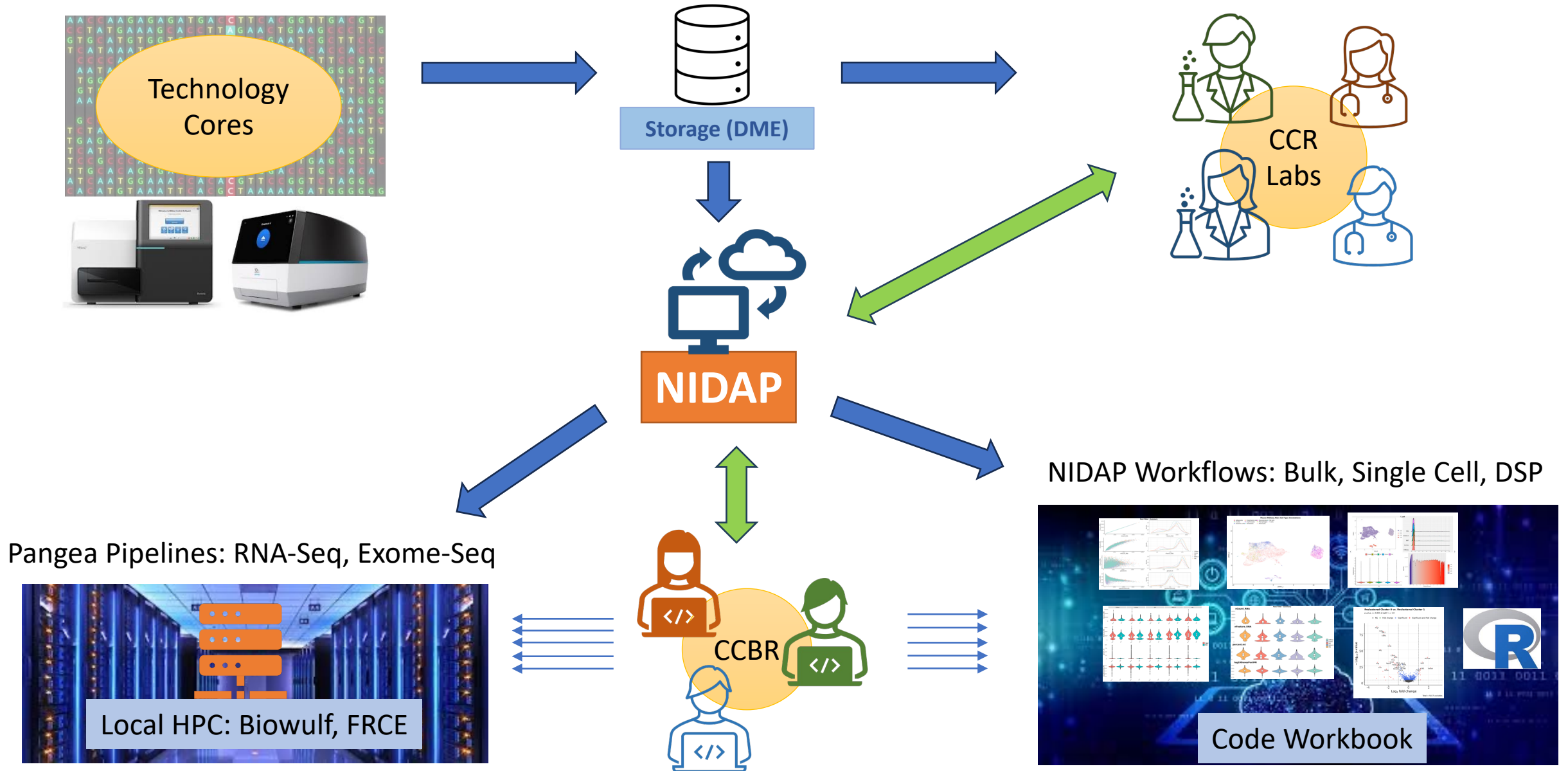
- Introduction to NIDAP
- Bulk RNA-seq Analysis on NIDAP
- Single-cell RNA-seq Analysis on NIDAP
- Spatial Transcriptomics on NIDAP
- Coming Soon on NIDAP
- Summary & Questions

NIDAP: The NIH Integrated Data Analysis Portal



- **An interactive cloud-based interface for bioinformatics analyses, allowing:**
 - Collaborative development of up-to-date bioinformatics workflows by CCR bioinformaticians
 - Implementation of standardized, user-friendly code templates that are disseminated among collaborators and instructed to laboratory scientists
 - Persistence and reuse of bioinformatics workflows enhances reproducibility

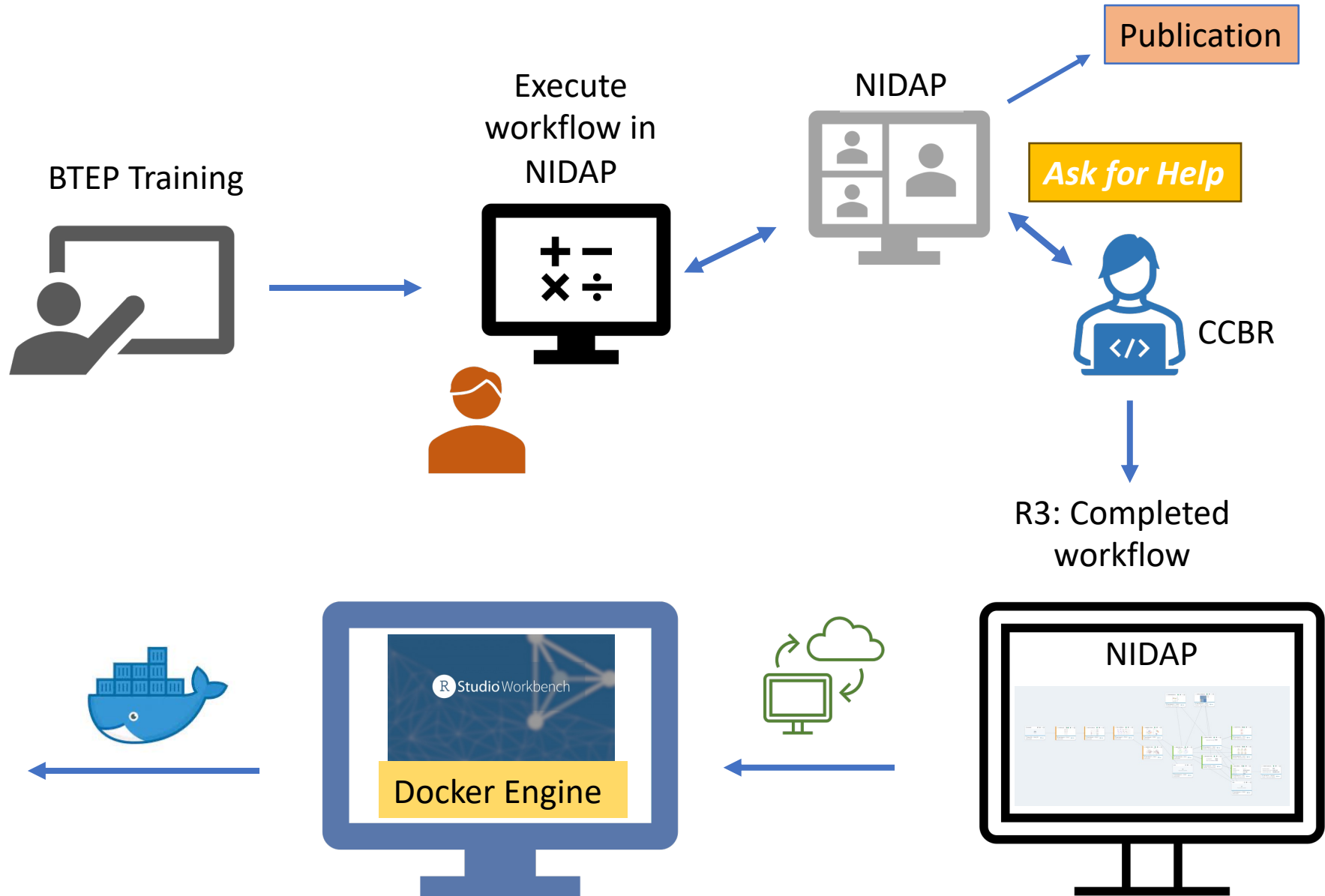
NIDAP facilitates access to and deployment of CCBR Data Analysis Workflows



Procedure for CCBR Projects using Automated Workflows on NIDAP


Request for:

- Bulk RNA-Seq
- Single cell RNA-Seq
- DSP
- Proteomics
- Exome-Seq
- Methylation
- Microarray
- Data mining
- sRNA-Seq/miR-Seq



Demo: The NIDAP Website

- The NIDAP Home Page:
<https://nidap.nih.gov/>
- Log-in with your NIH Credentials
- First time NIH users will be prompted to Register
- You can use NIDAP whether on- or off-VPN
- Google Chrome is currently the only supported browser



The screenshot shows the NIDAP website home page. At the top, the NIH logo and "NATIONAL CANCER INSTITUTE NIH Integrated Data Analysis Platform" are visible. The main header area features a large image of a microscope lens and the text "Hello Thomas" followed by "Welcome to NIDAP, the NIH Integrated Data Analysis Platform" and a "GET STARTED" button. Below this, a statistics bar displays: 83 RESEARCH GROUPS, 647,101 PATIENTS, 61,256,654 LAB TESTS, 240,851 PATHOLOGY REPORTS, and 51 BIOINFORMATICS TOOLS. The main content area includes three sections: "Connected Lab" (integrating genomic, clinical, and imaging data), "Bioinformatics Tools" (analytical templates supported by NCI experts), and "Research Funding Systems" (integrating investment). To the right, a section titled "NIDAP powers fundamental and translational research at the National Institutes of Health" lists three key capabilities: data integration, secure collaboration, and tool flexibility.

NIH NATIONAL CANCER INSTITUTE
NIH Integrated Data Analysis Platform

Hello Thomas
Welcome to NIDAP, the NIH Integrated Data Analysis Platform

GET STARTED

83 RESEARCH GROUPS 647,101 PATIENTS 61,256,654 LAB TESTS 240,851 PATHOLOGY REPORTS 51 BIOINFORMATICS TOOLS

Connected Lab
Connected Lab integrates genomic, clinical, and imaging data to create a Patient 360° view.

Bioinformatics Tools
Deploy these analytical templates on your data. All tools are fully supported by NCI Informatics experts.

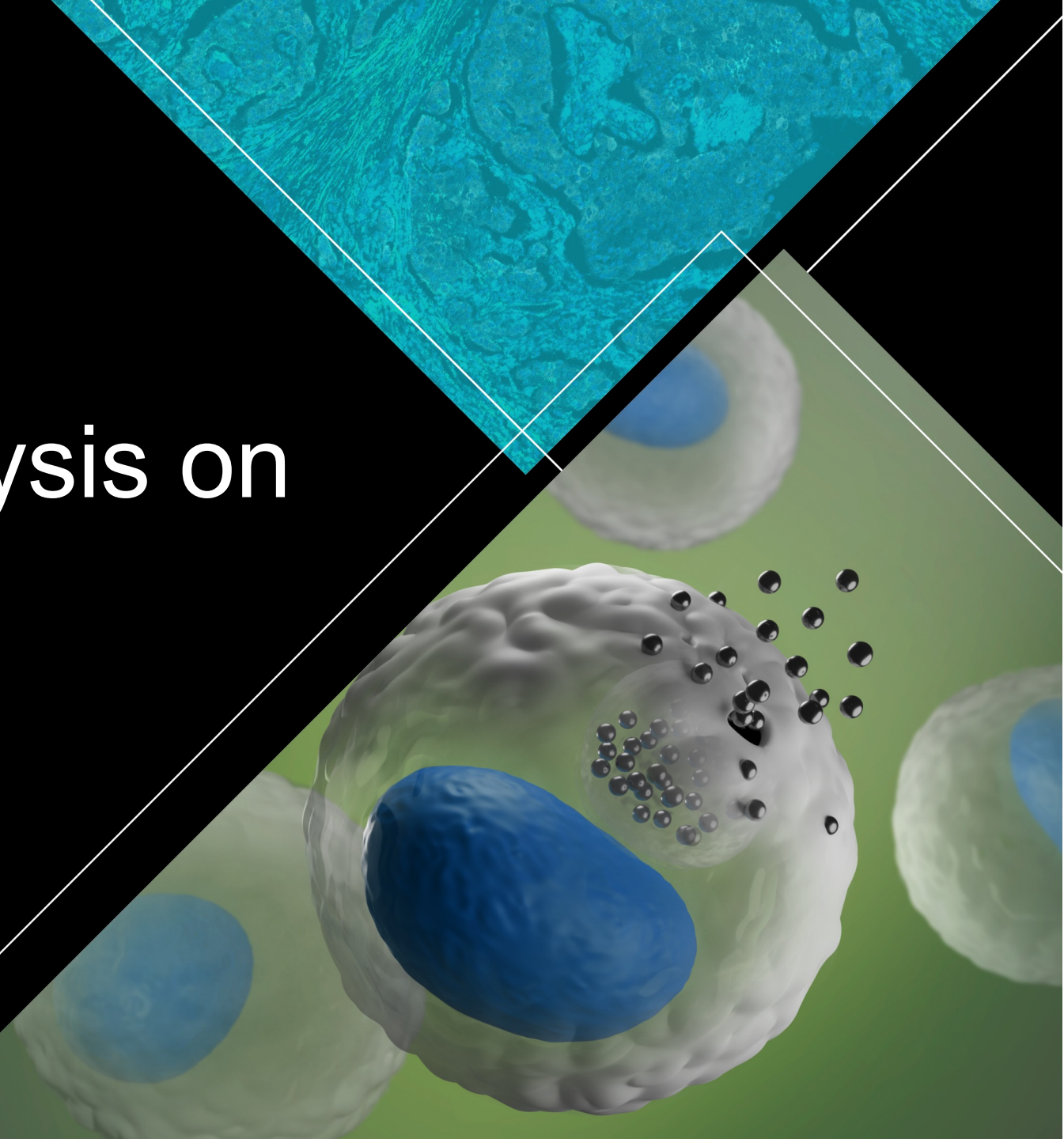
Research Funding Systems
Research Funding Systems integrate investment

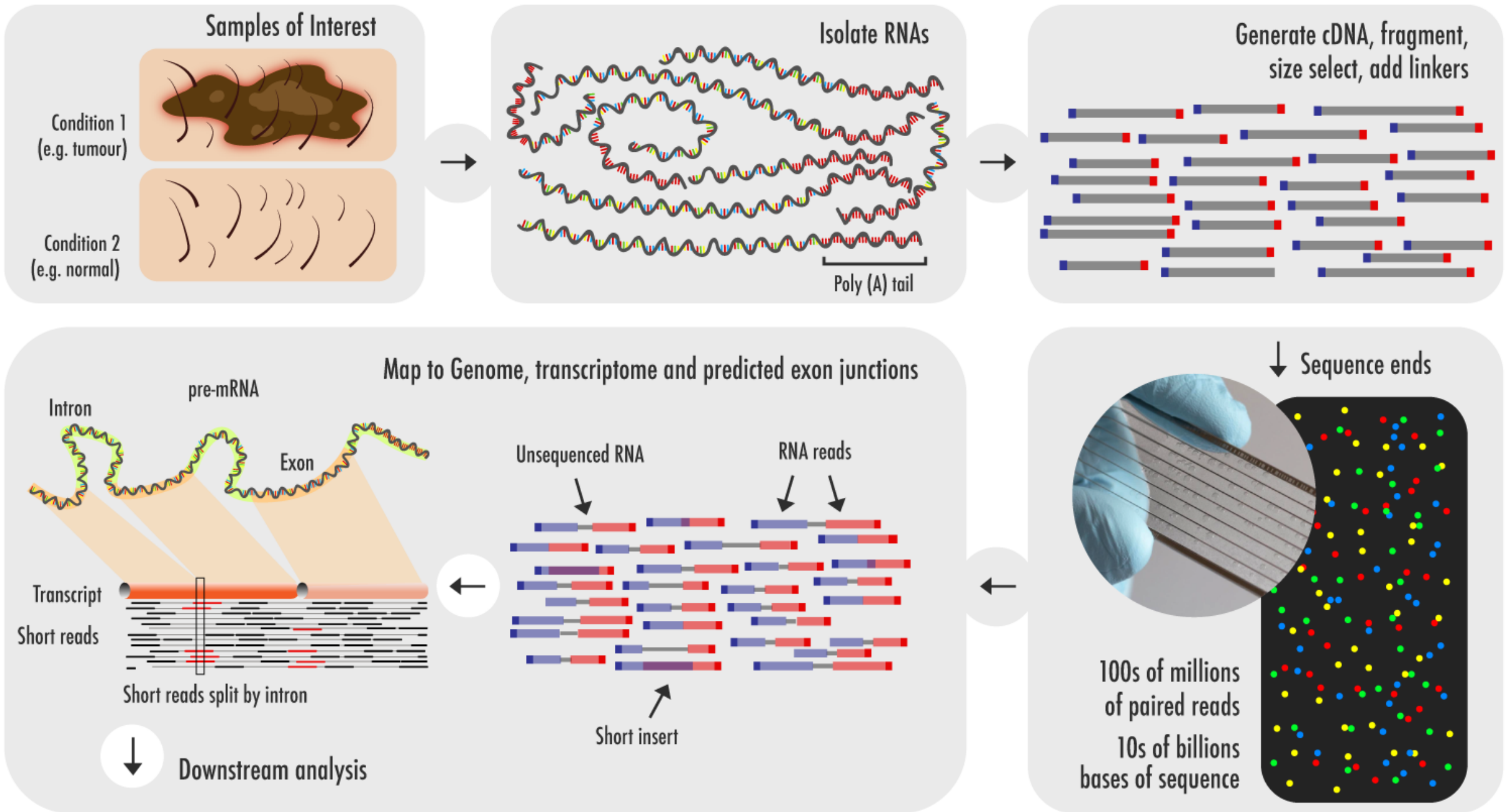
NIDAP powers fundamental and translational research at the National Institutes of Health

- Collect, integrate, and harmonize health data flexibly with bi-directional connectivity to external systems
- Securely work and collaborate with PII/PHI while maintaining full ownership
- Choose between point-and-click tools for cohorting and analysis or code-based tools for open languages

Bulk RNA-seq Analysis on NIDAP

**Thomas Joshua Meyer
(Josh)**

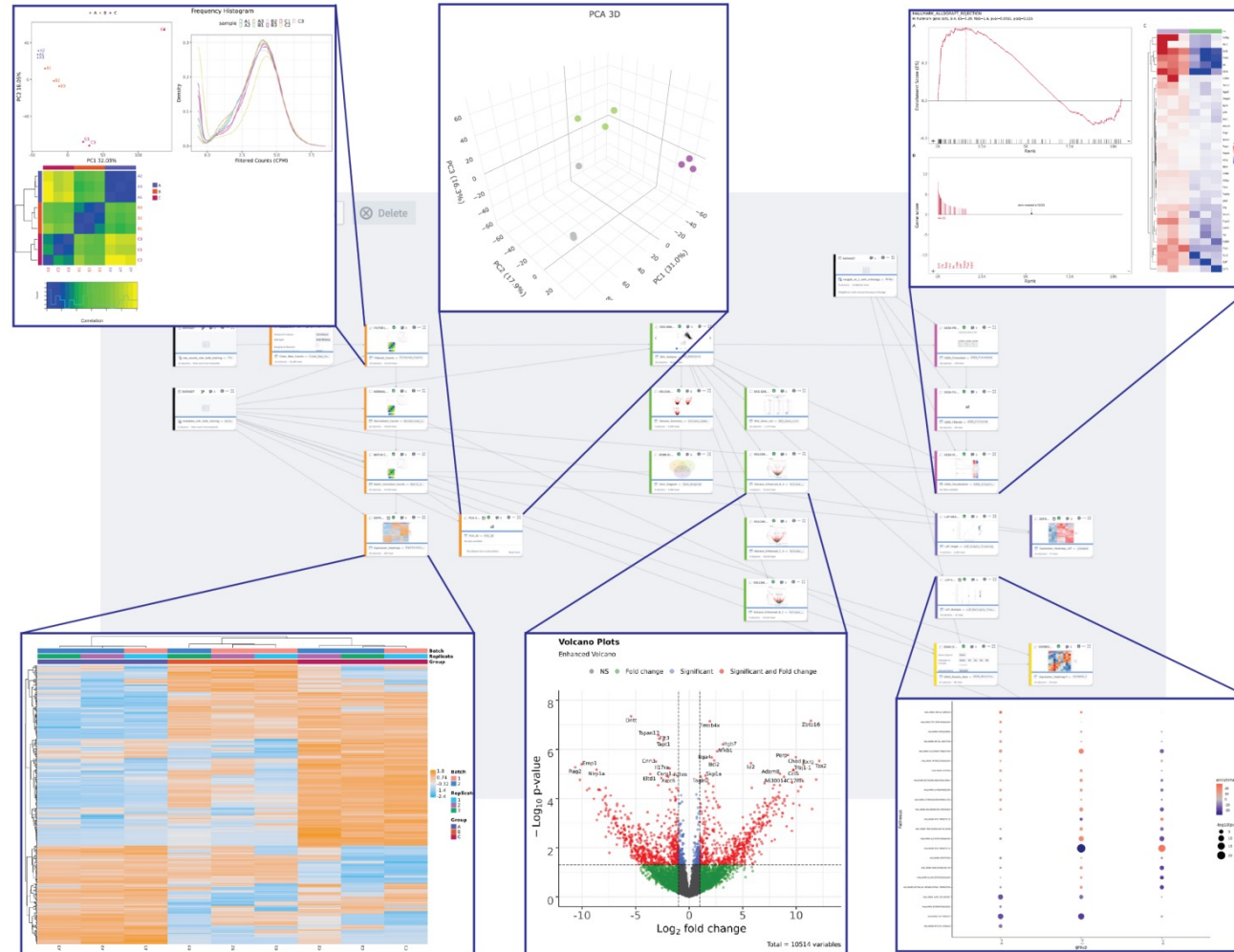




Bulk RNA-seq: Downstream Analysis

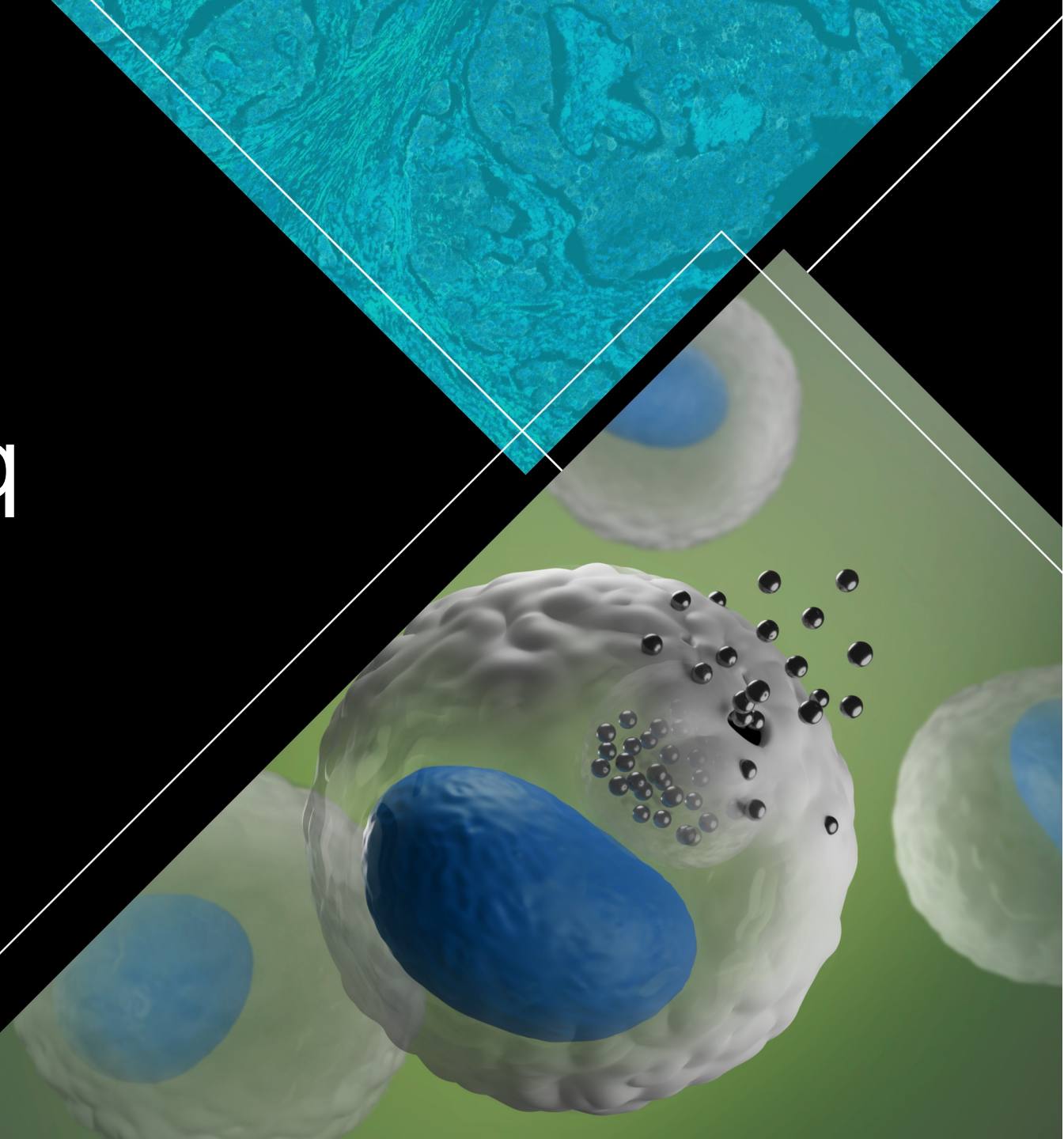
- Begins with raw counts of detected expression over annotated features (e.g. genes) in the reference genome to which reads are aligned
 - This is a large matrix of count values with rows for genes and columns for samples
 - Also need a metadata table, with sample names, groups, short labels, and any batch information
- Filter out low-count genes and convert to counts-per-million (CPM)
 - Those with fewer than X samples per group with at least Y count value
 - Defaults: $X = 3$, $Y = 1$
- Log₂ transformation of CPM values and Normalization to ensure comparisons between samples are valid
- Batch Correction (Optional) to identify and remove any batch effect
- Differential Expression of Genes (DEG) Analysis to look at relative expression between two groups of samples from different experimental conditions

Demo: The CCBR Bulk RNA-seq Workflow

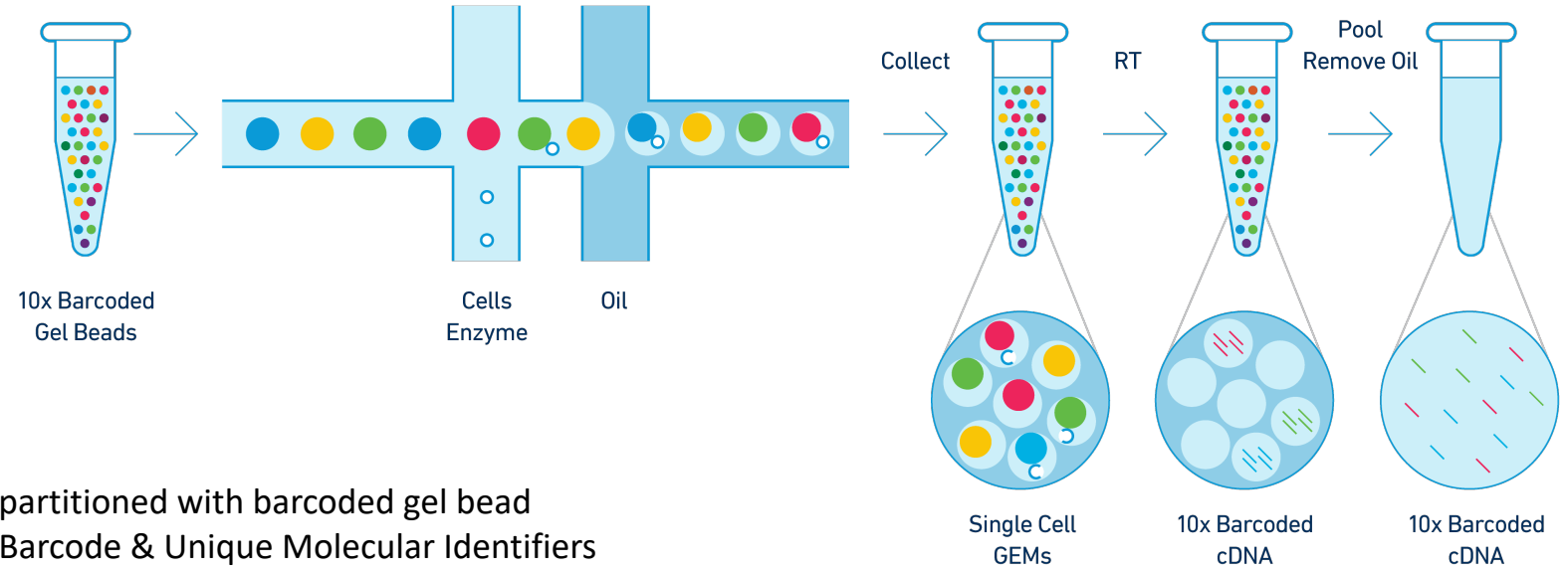


Single-cell RNA-seq Analysis on NIDAP

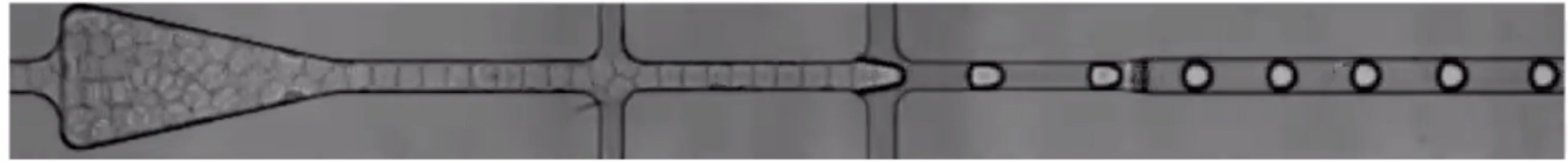
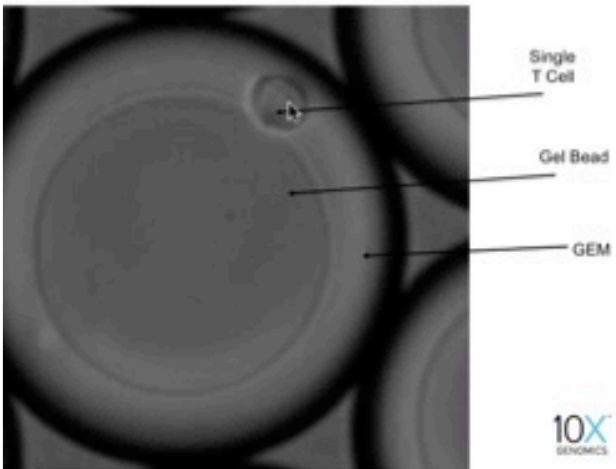
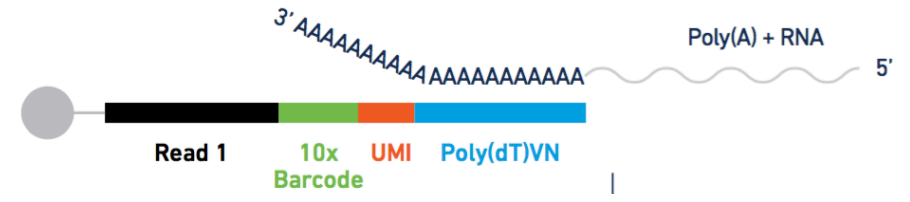
**Thomas Joshua Meyer
(Josh)**



Droplet-based scRNA-Seq allows high-throughput profiling with unprecedented ease *(for good samples)*



- Cell partitioned with barcoded gel bead
- Cell Barcode & Unique Molecular Identifiers (UMI) on end of cDNA
- ~2000 median genes detected on average with ~50,000 mean sequencing reads per cells
- Gene-level counts data

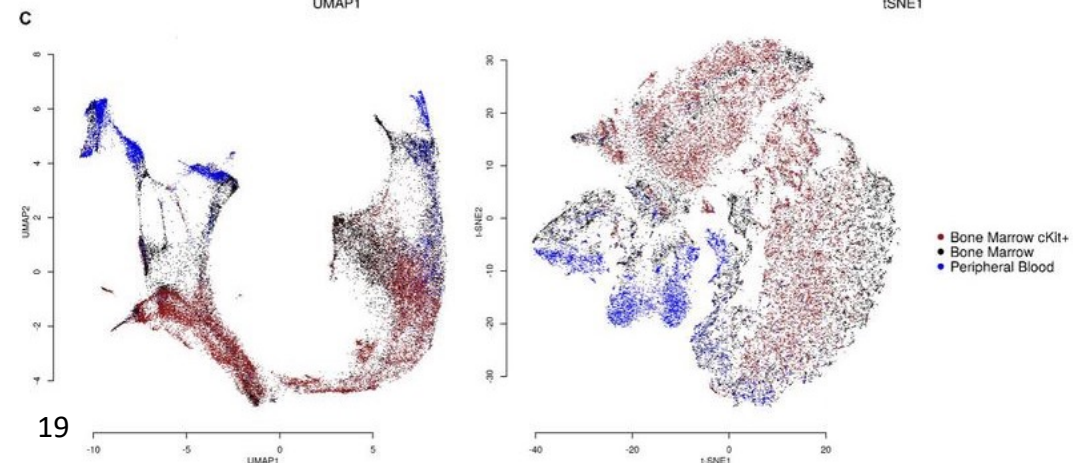
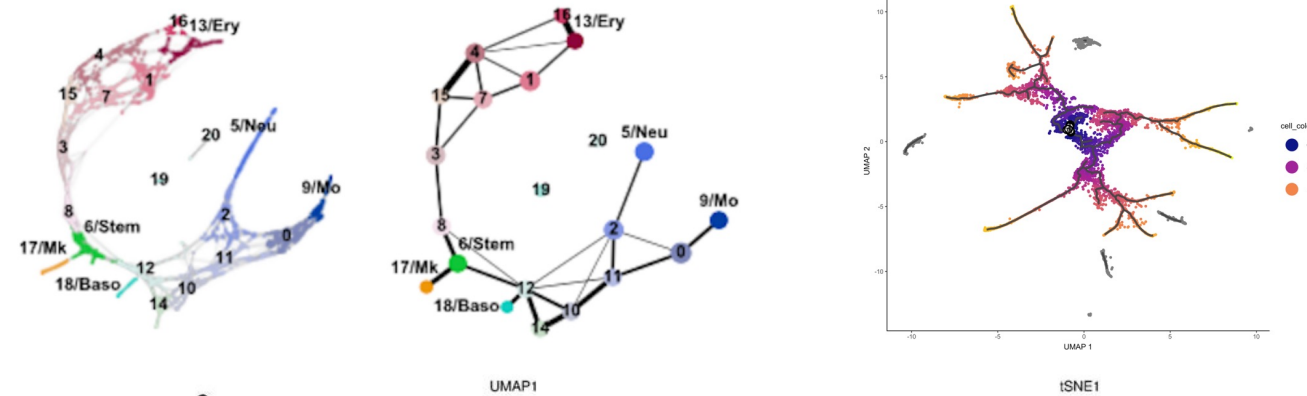
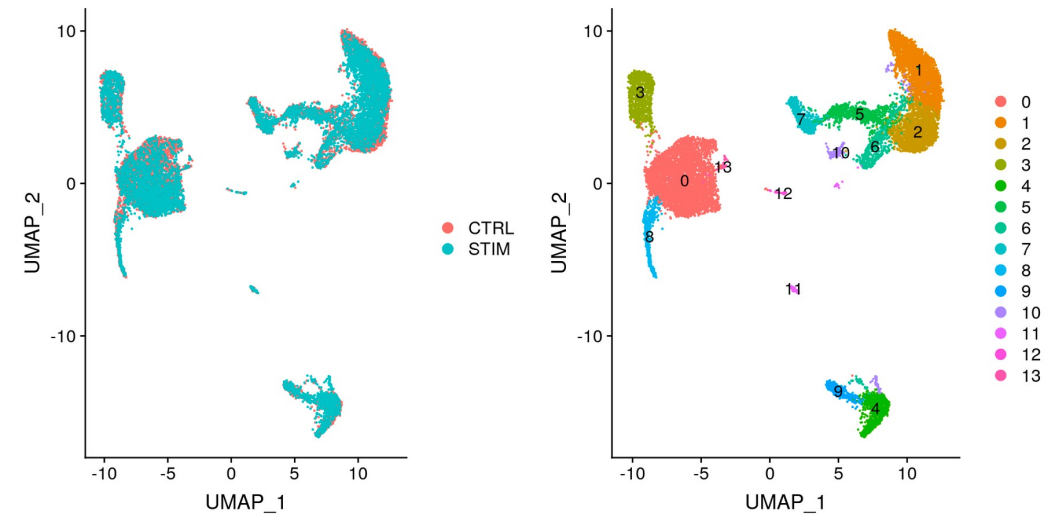


When you receive your data

- The data is delivered by the sequencing machines as BCL (raw base call) files
- The CellRanger software from 10x Genomics performs the following steps:
 - Make a FASTQ from the BCL files
 - Sort and count reads for each cell based on unique molecular identifiers (UMIs)
 - [Optional] Aggregate multiple samples into a single dataset and perform preliminary dimensionality reduction and clustering
 - These results can be viewed through the cLoupe Browser
- Counts files for each sample are provided as a set of tab separated files (.tsv) or as a single .h5 file

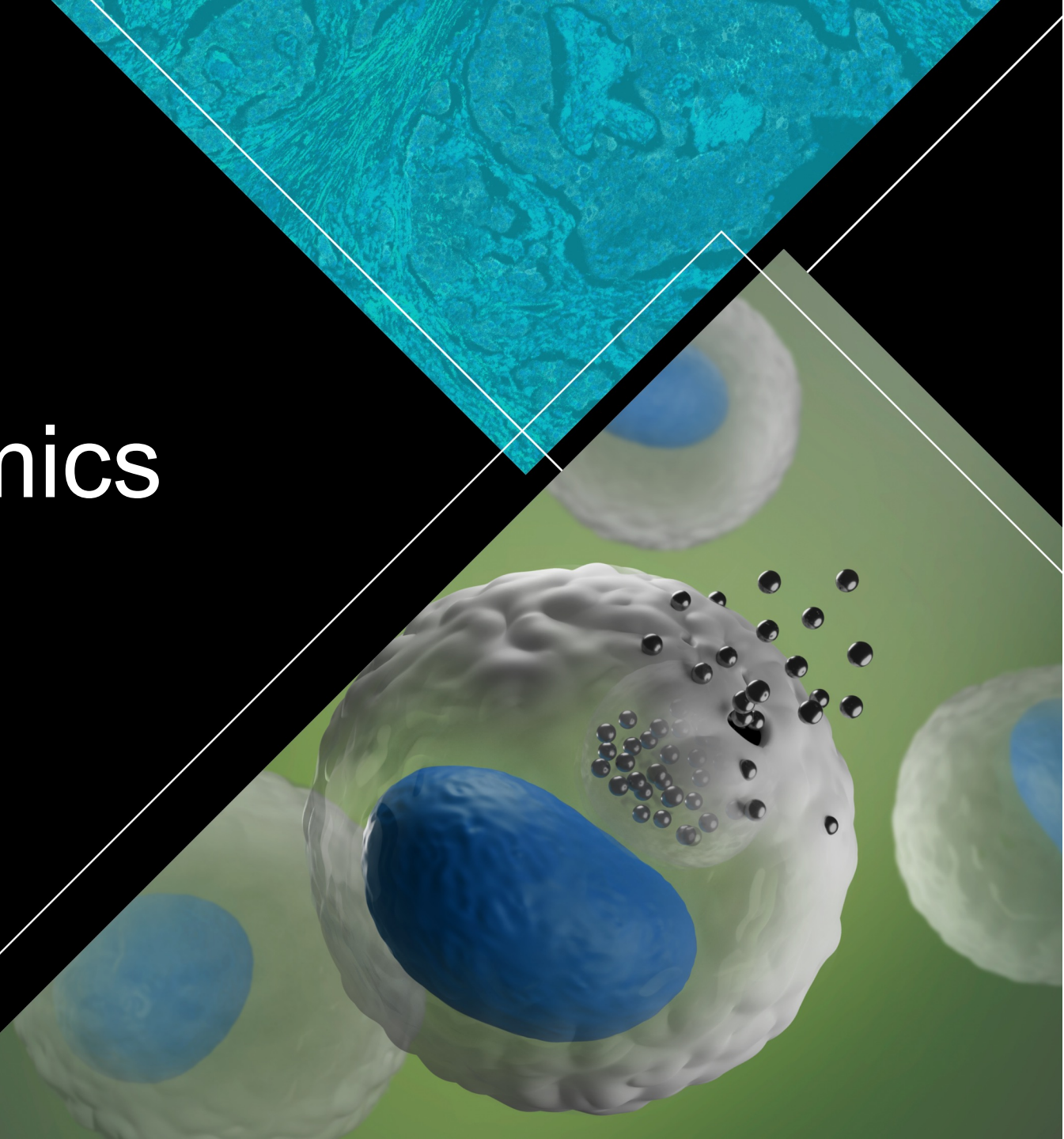
Downstream analysis tools

- By far, the most commonly cited tool is Seurat (from the Satija lab), which can perform quality control (QC), dimensionality reduction, and differential expression
- Additional applications:
 - Trajectory analysis (including TSCAN, Monocle, and PAGA)
 - Cell type identification (SingleR with reference sets like ImmGen)



Spatial Transcriptomics Analysis on NIDAP

Ned Cauley

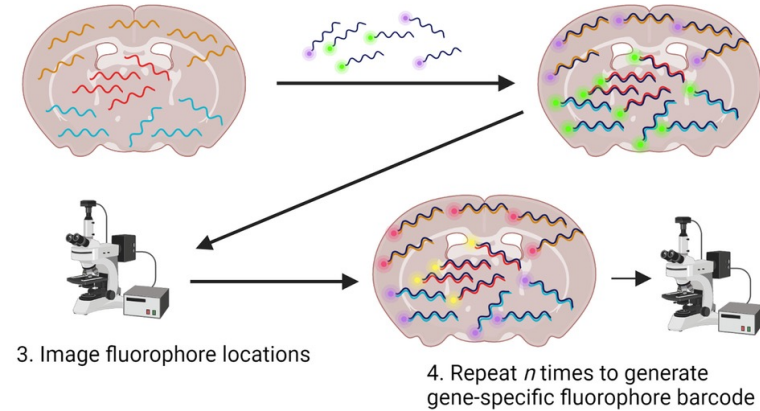


Spatial Transcriptomics on NIDAP

Imaging methods

ISH

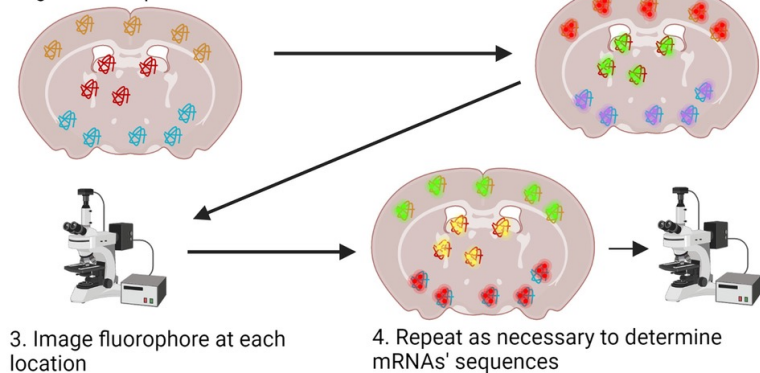
1. Hybridise labelled probes to target mRNA



ISS

1. Rolling circle amplification of target transcripts

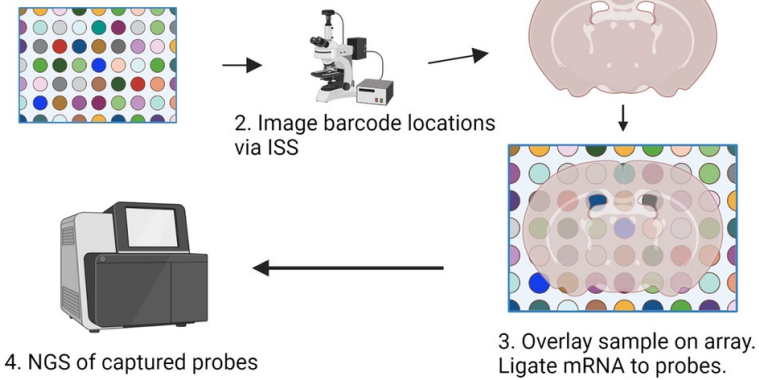
2. Hybridise short, labelled probes to determine 1-2nt of transcript's sequence



Sequencing methods

Arrays

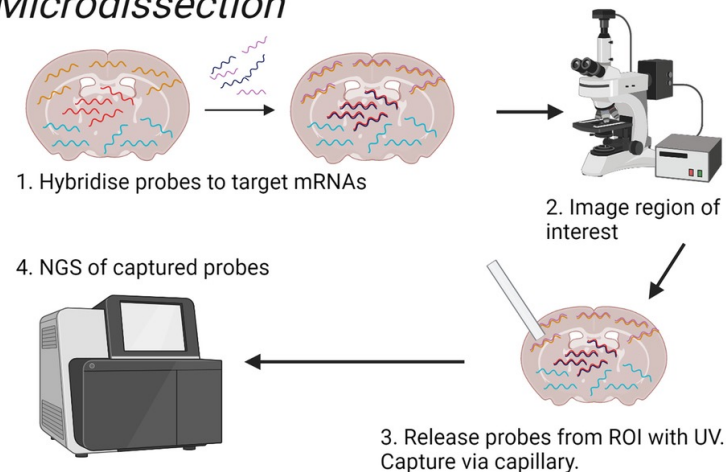
1. Array of spatially barcoded probes



Microdissection

1. Hybridise probes to target mRNAs

4. NGS of captured probes



- Workflows for the analysis of sequencing methods

Digital Spatial Profiling

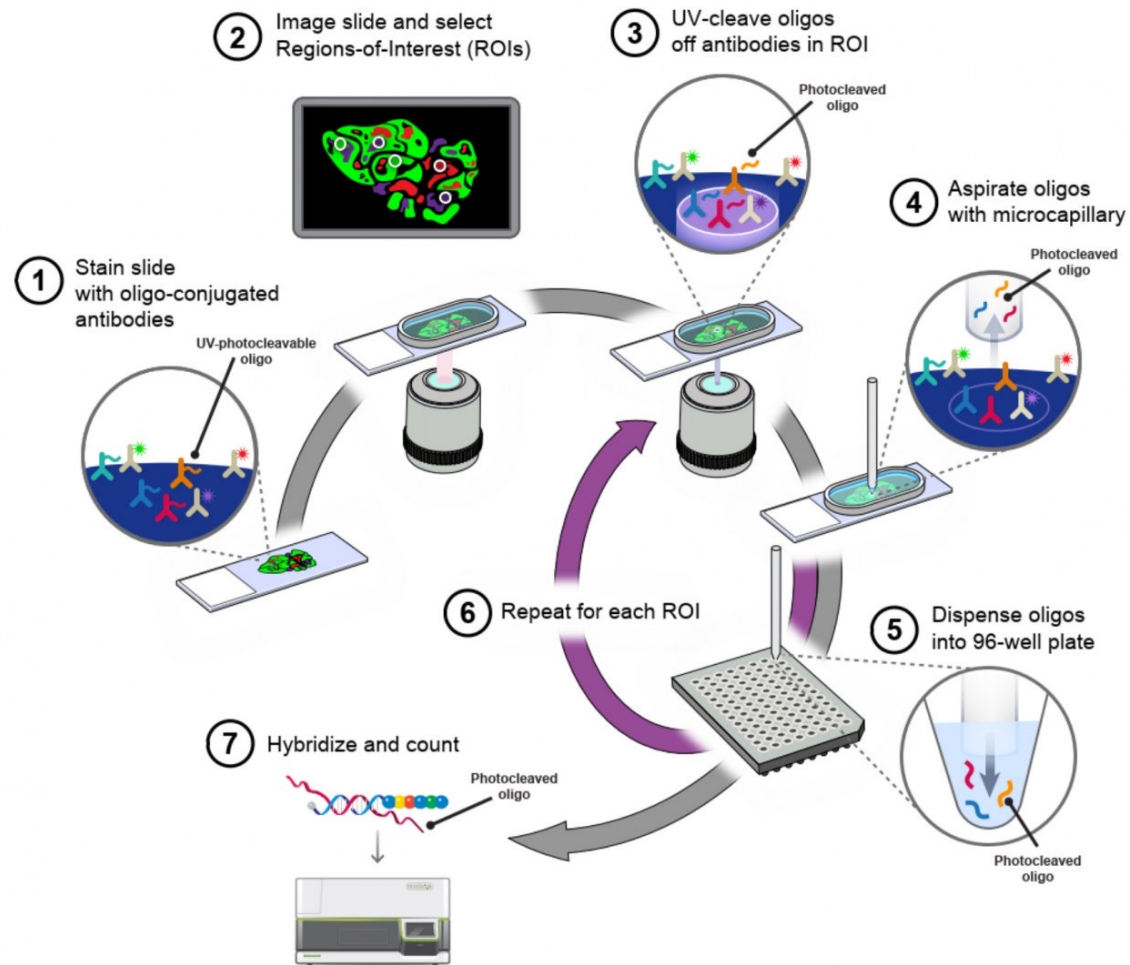
- Targeted regions of interest
- Complete workflow available on NIDAP
- Training videos
- Example analysis

Visium

- Uniform grid of regions of interest
- Beta Workflow available on NIDAP

Spatial Transcriptomics on NIDAP

Digital Spatial Profiling Workflow



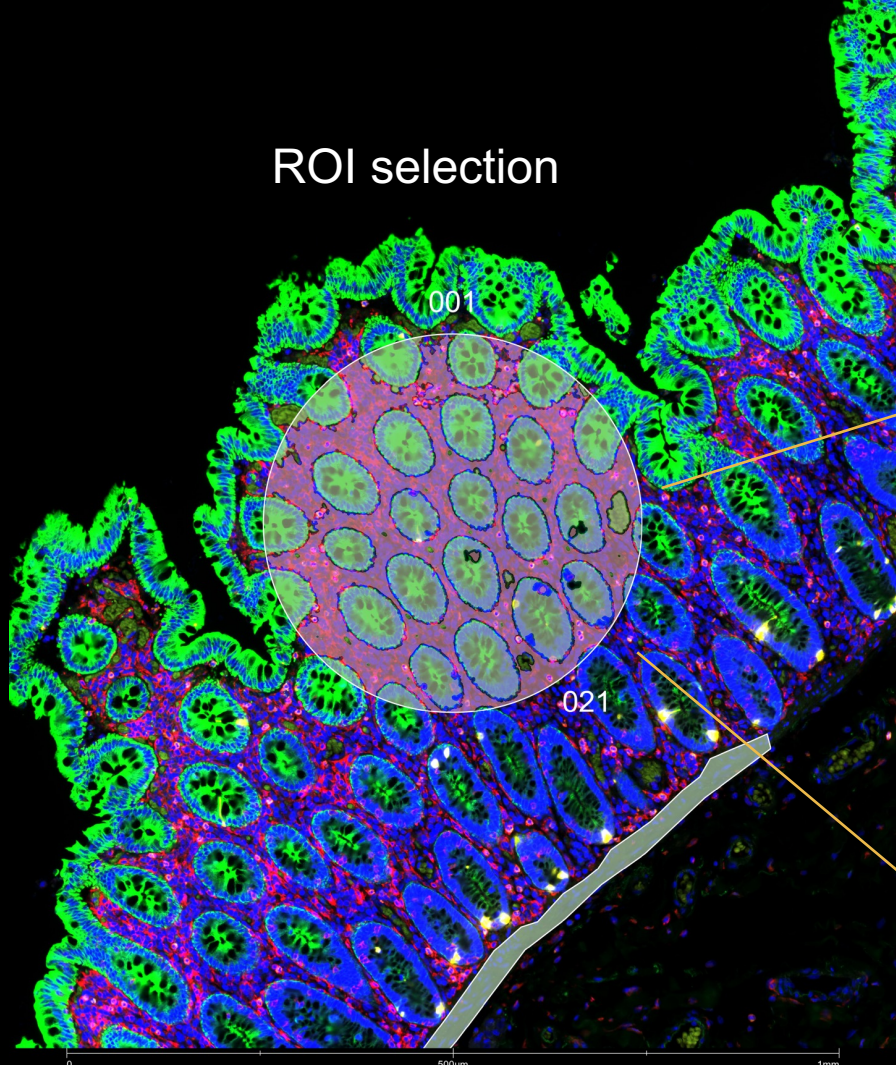
Workflows for the analysis of sequencing methods

Digital Spatial Profiling

- ◇ Targeted regions of interest
- ◇ Complete workflow available on NIDAP
- ◇ Training videos
- ◇ Example analysis

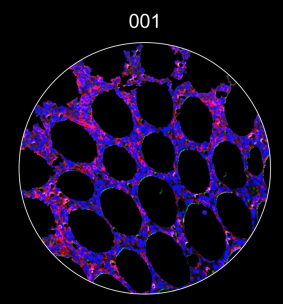
Visium

- ◇ Uniform grid of regions of interest
- ◇ Beta Workflow available on NIDAP

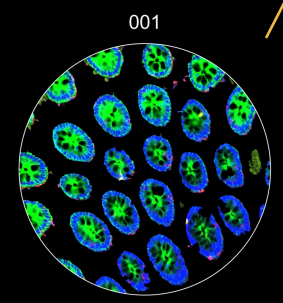


ROI selection

AOI selection



PanCK -



PanCK +

96 well plate

A01

A02

Each well contains probes captured from a single AOI or "segment"

NGS sequencing

FASTQ files

GeoMx NGS Pipeline

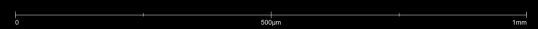
DCC files

DCC file for AOI 001 PanCK

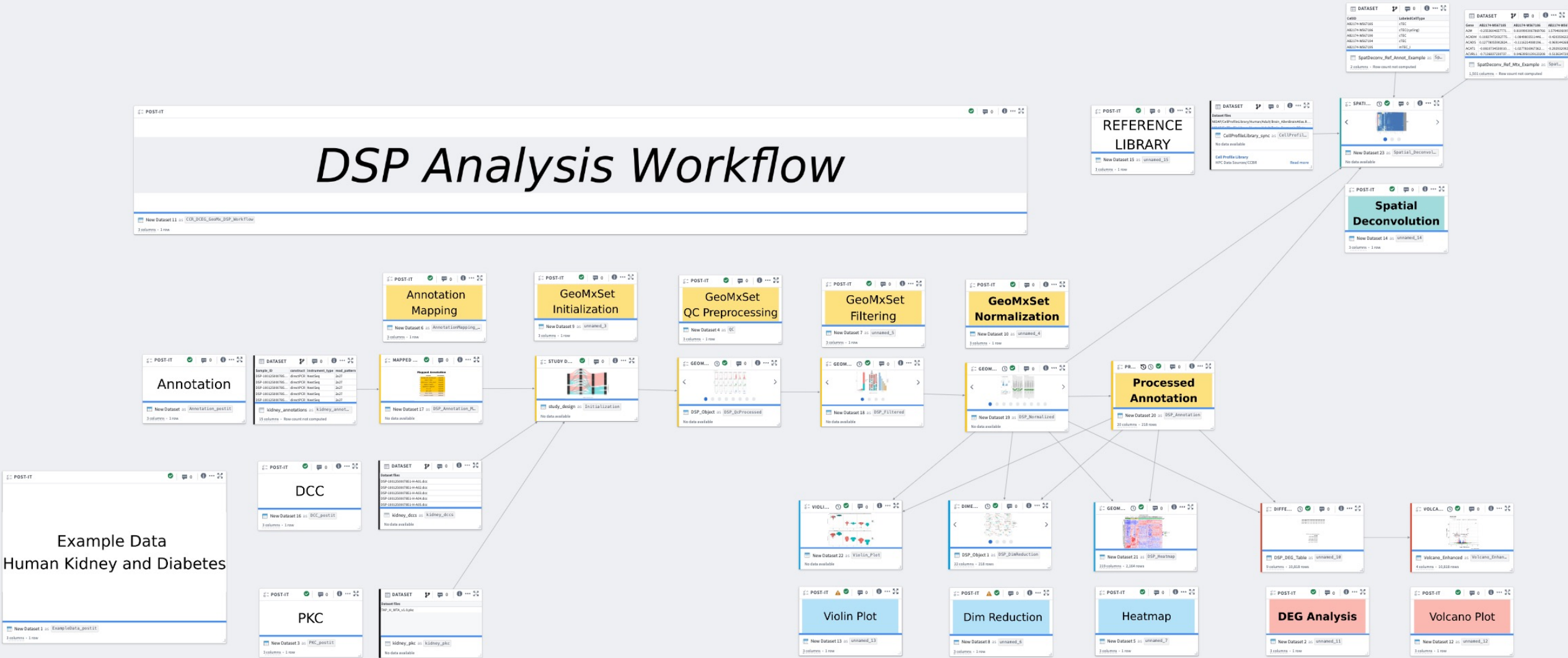
<u>ProbeID</u>	<u>Reads</u>
1111	20
2222	10
3333	35

Morphological Markers

- SYTO83 - Nuclei
- PanCK – Epithelial Cells
- CD45 – Immune Cells
- HT-5 – Endocrine Cells



DSP Analysis Workflow



DSP Analy

Example Data Human Kidney and Diabetes

Run Volcano_Enhanced as Volcano_Enhanced 4 columns - 10,818 rows Save as dataset </> Toggle view Edit template Actions

Created from (v54)
Volcano Plot - Enhanced [CCBR]
Implementation of Bioconductor's Enhanced Volcano Plot (v1.6.0, <https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html>). Template written by Matthew Angel and maintained by CCBP. Final Potomac Compatible Version: v52.

Basic parameters

- DEG Table**
Dataset containing differential expression of genes (DEG) analysis output columns. Usually, this includes columns for gene names, (log) fold changes, (adjusted) p-values, and t-statistics. Other columns may be present.
DSP_DEG_Table ✓
- Column with Feature ID**
Column from the input DEG table containing Feature ID (such as Gene Names, Isoform IDs, UniProt IDs, and so on). This is usually the first column (named "Feature_ID" or "Gene"). Only Text type columns will be allowed.
Gene ✓
- Significance Column**
Choose an unadjusted or adjusted p-value column from the input DEG table to use as the measure of significance in your Volcano plot. If your DEG analysis contained more than one contrast comparison, you will only be able to select one of these at a time. Make sure you select the same contrast that was selected for the "Log2 Fold Change Column" parameter.
normal_glomerulus-tubule_pval ✓
- P-Value Threshold**
0.001 ✓
- Log2 Fold Change Column**
Choose a log2 fold change column from the input DEG table. If your DEG analysis contained more than one contrast comparison, you will only be able to select one of these at a time. Make sure you select the same contrast that was selected for the "Significance Column" parameter.
normal_glomerulus-tubule_logFC ✓
- Log2 Fold Change Threshold**
1.0 ✓

Label parameters

Plot parameters

The screenshot shows a workflow interface with several interconnected components:

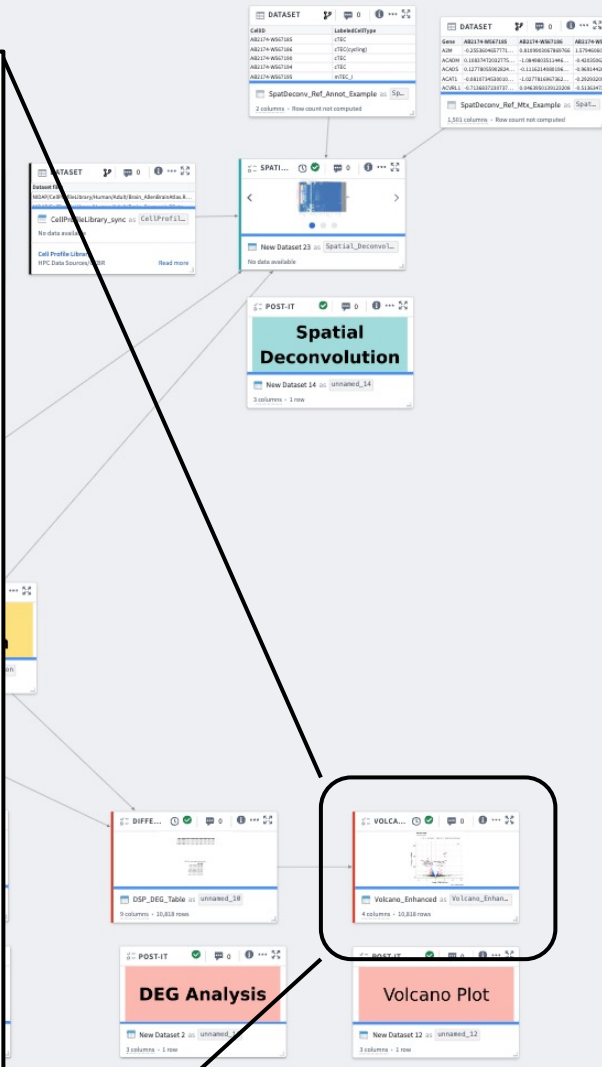
- Annotation**: A dataset containing gene annotations.
- Annotation Mapping**: A step that maps the annotation data to the DEG data.
- DCC** (Differential Expression Comparison): A dataset used for comparing different conditions.
- PKC**: A dataset used for protein kinase C analysis.
- DEG Analysis**: A step that performs differential expression analysis on the input data.
- Volcano Plot**: A visualization of the differential expression results.
- Spatial Deconvolution**: A step that performs spatial deconvolution on the data.

The interface also shows a list of datasets and their associated data sources, such as "New Dataset 11" (COL3CE1_Geom_Sp_WPKTLow) and "New Dataset 12" (Volcano_Enhanced).

DSP Analy

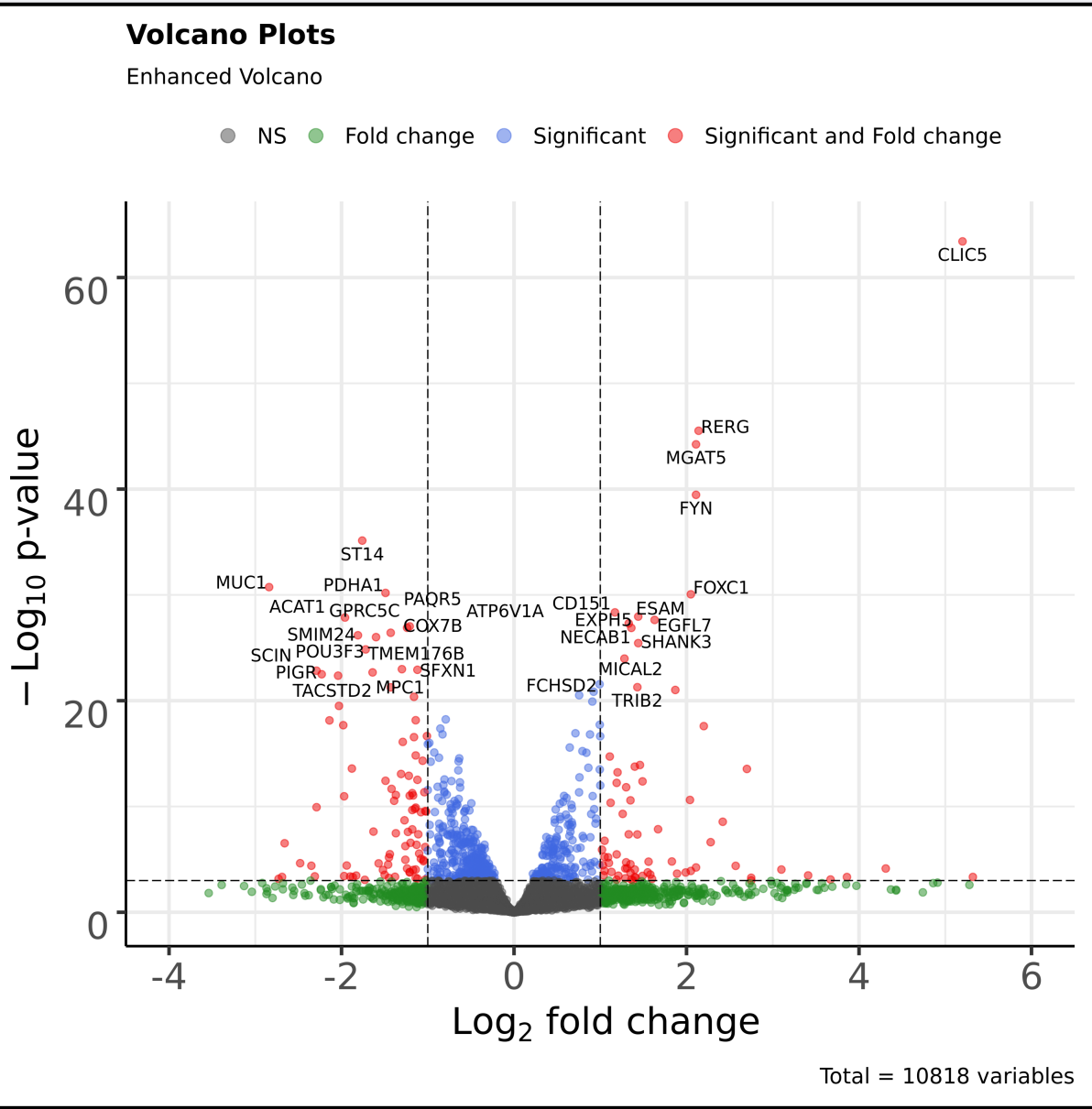
Example Data
Human Kidney and Diabetes

```
Run Volcano_Enhanced as Volcano_Enhanced Save as dataset Toggle view Edit template Actions 4 columns - 10,818 rows
Inputs DSP_DEG_Table
1 # Volcano Plot - Enhanced [CCBR] (0c91aa57-0f76-4513-a063-5f9263d65727): v54
2 Volcano_Enhanced <- function(DSP_DEG_Table) {
3   # image: png
4
5
6   # Changelog
7   # 2022-09-14 Rearranged structure and description
8   # 2020-10-29 Add support for pval == 0
9
10
11
12   ## ----- ##
13   ## Libraries ##
14   ## ----- ##
15
16   library(stringr)
17   library(ggplot2)
18   library(ggrepel)
19   library(dplyr)
20
21
22   ## ----- ##
23   ## User-Defined Template Parameters ##
24   ## ----- ##
25
26   #Basic Parameters:
27   df <- DSP_DEG_Table
28   label.col <- "Gene"
29   sig.col <- "normal_glomerulus-tubule_pval"
30   pCutoff = 0.001
31   lfc.col <- "normal_glomerulus-tubule_logFC"
32   FCcutoff = 1.0
33
34
35   #Label Parameters
36   value_to_sort_the_output_dataset <- "p-value"
37   no_genes_to_label <- 30
38   use_only_addition_labels <- FALSE
39   additional_labels <- ""
40   labSize <- 4
41
42
43   #Plot Parameters
44   change_sig_name <- "p-value"
45   change_lfc_name <- "log2FC"
46   title <- "Volcano Plots"
47   subtitle <- "Enhanced Volcano"
48   use_custom_lab <- FALSE
49   ylim <- 0
50   xlim_additional <- 0
51   ylim_additional <- 0
52   axisLabSize <- 24
53   pointSize <- 2
54
55
56   #Image Parameters
```



DSP A

Example Data
Human Kidney and Diabetes



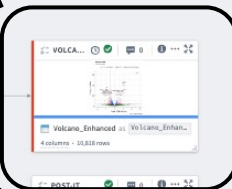
Gene	AB017486748	AB017486749	AB017486750
CLIC5	0.2226060775	0.2226060670	1.17966491
RERG	-0.1847670275	-0.1847670170	-0.4510022
MGAT5	0.2277802628	-0.2222488108	-0.9614428
FYN	-0.8481740288	-1.0277848232	-0.2202022
FOXC1	0.1318023872	0.1318023767	0.1318023662



Spatial Deconvolution



DEG Analysis



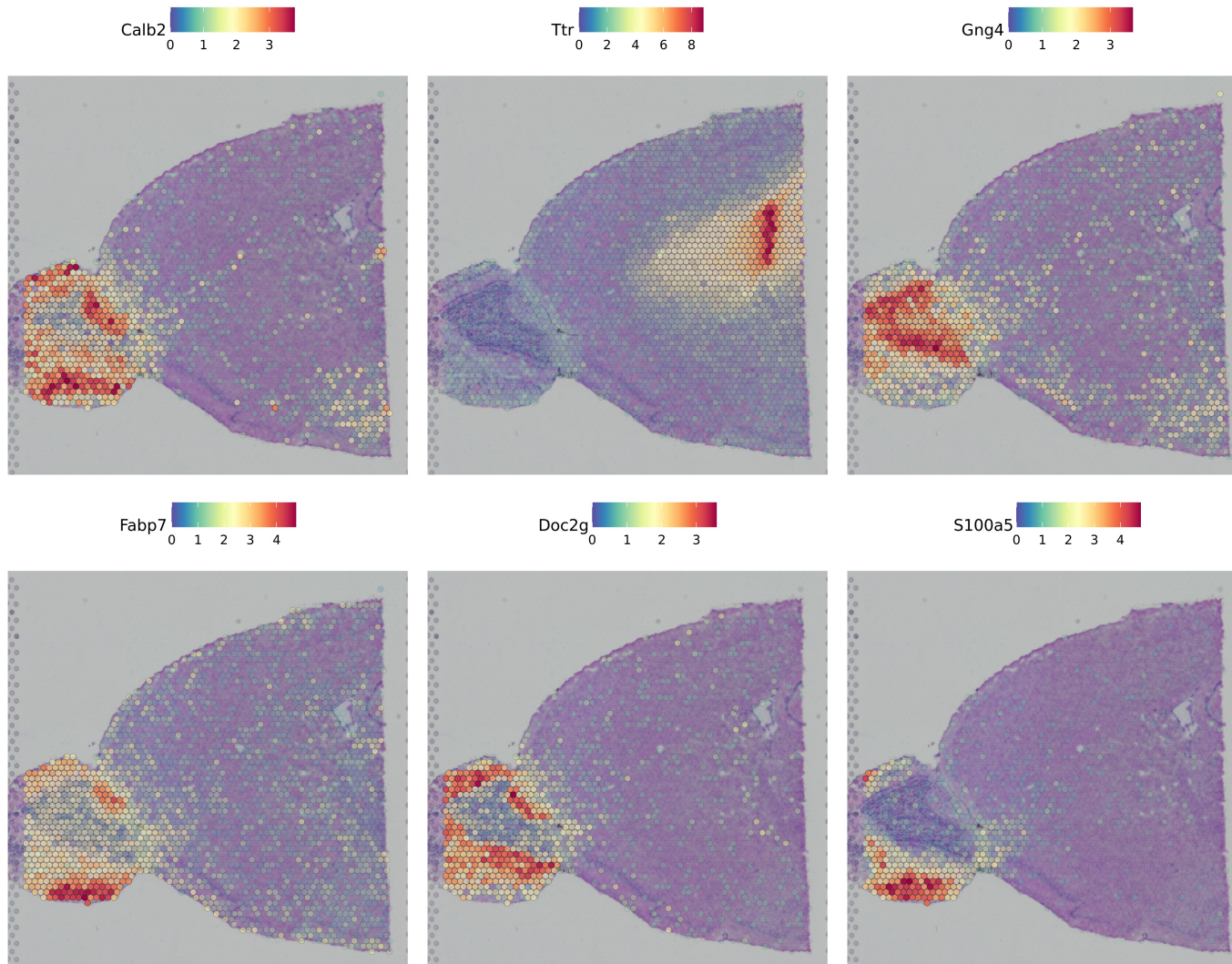
Volcano Plot



DSP Workflow on NIDAP

- <https://nidap.nih.gov/workspace/vector/view/ri.vector.main.workbook.67a84967-f588-496a-b81f-e8dbdd1dee6d?branch=test>

Visium Workflow on NIDAP



- **Workflows for the analysis of sequencing methods**

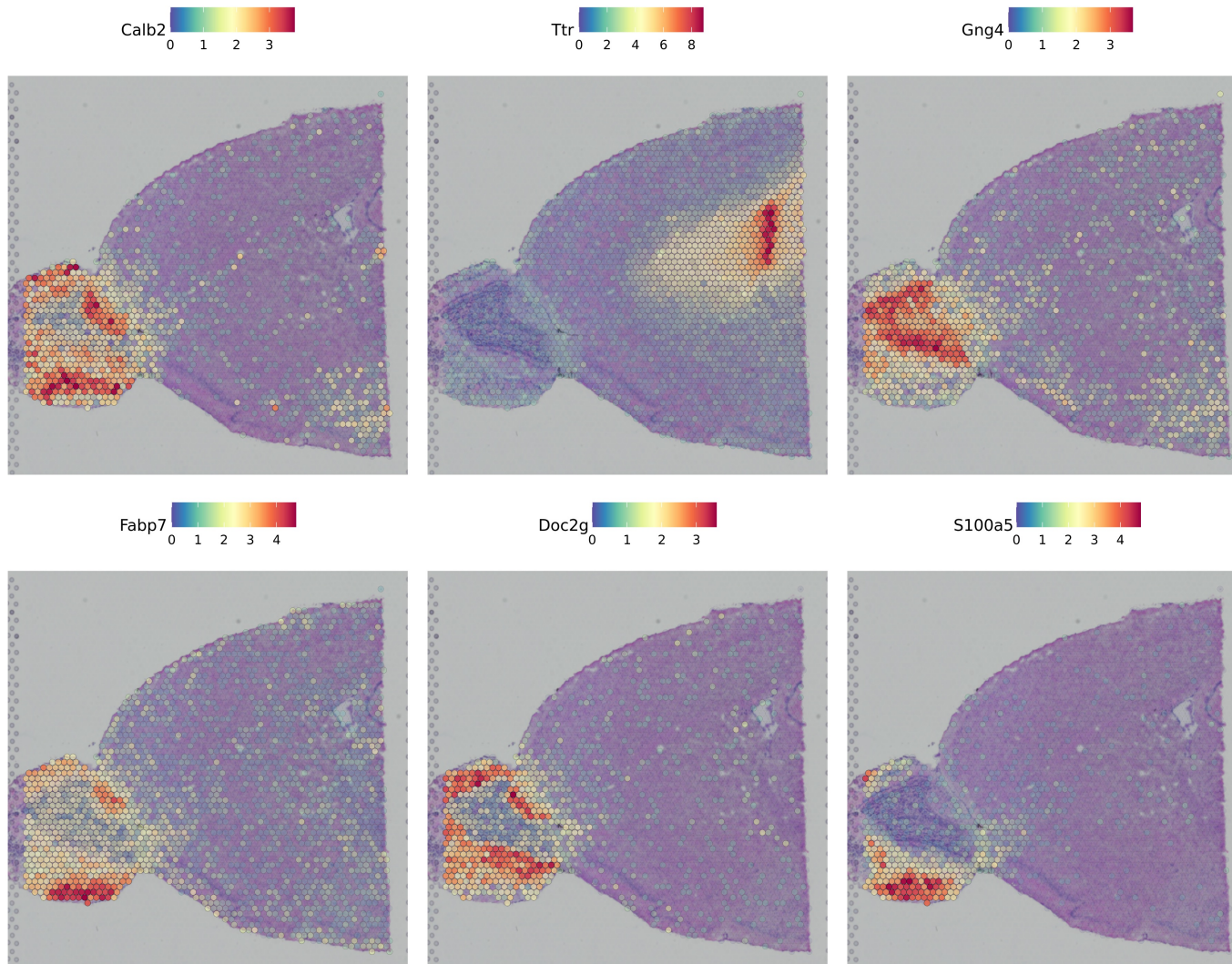
- **Digital Spatial Profiling**

- ◇ Targeted regions of interest
- ◇ Complete workflow available on NIDAP
- ◇ Training videos
- ◇ Example analysis

- **Visium**

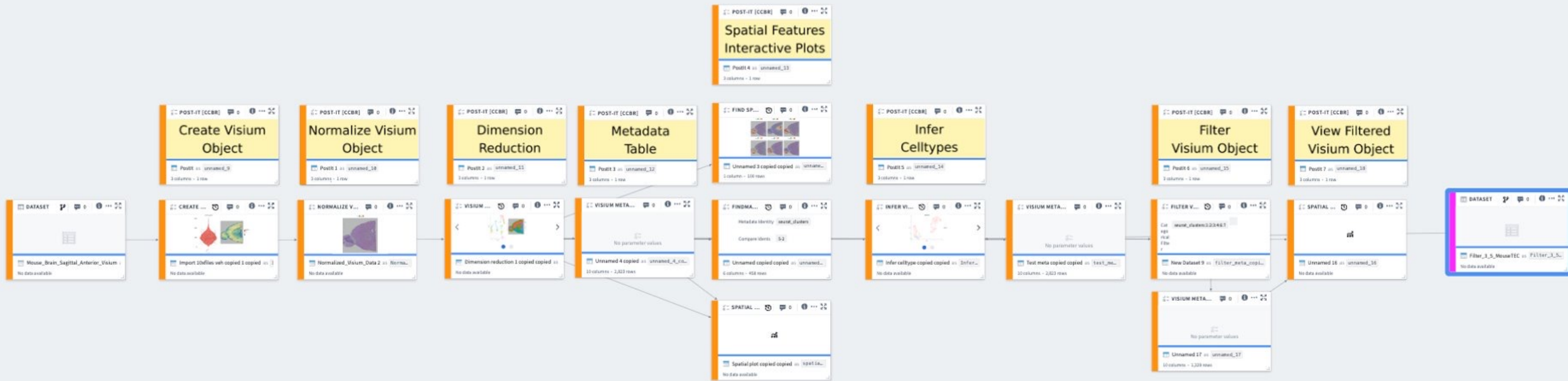
- ◇ Uniform grid of regions of interest
- ◇ Beta Workflow available on NIDAP

Visium Workflow on NIDAP

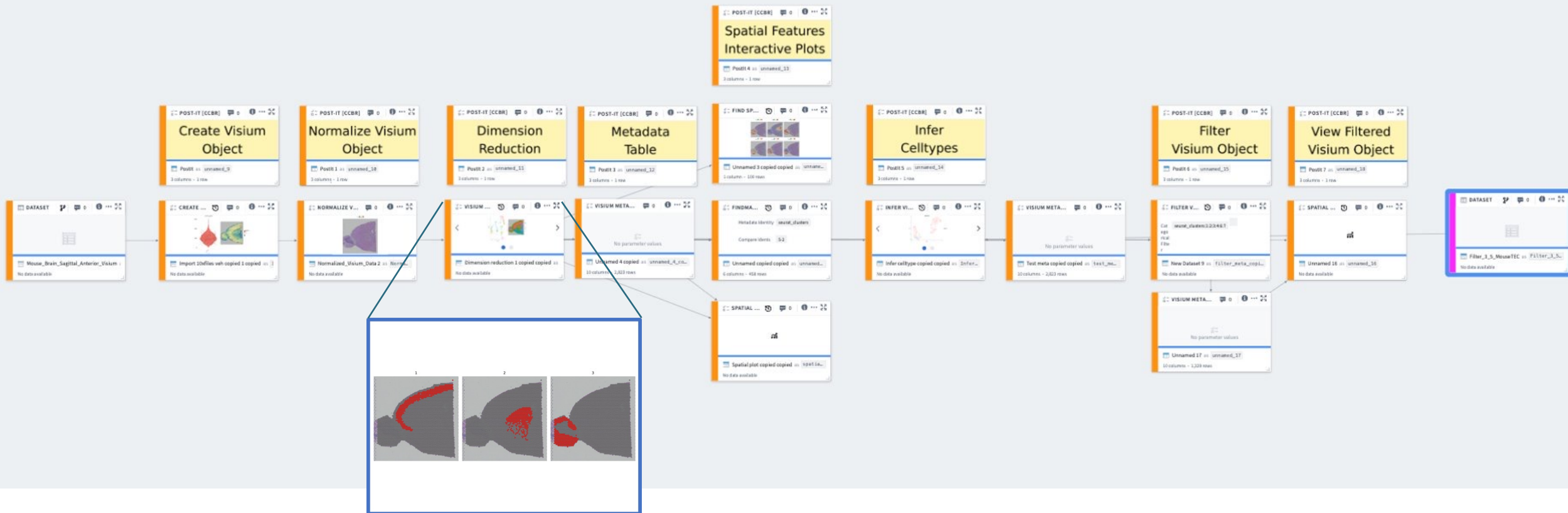


- Based on the Seurat Visium vignette
- Workflow currently in “beta”
 - ◇ Functional workflow that has been tested successfully
 - ◇ No video tutorials (yet!)
- Currently available for any NIDAP user to use

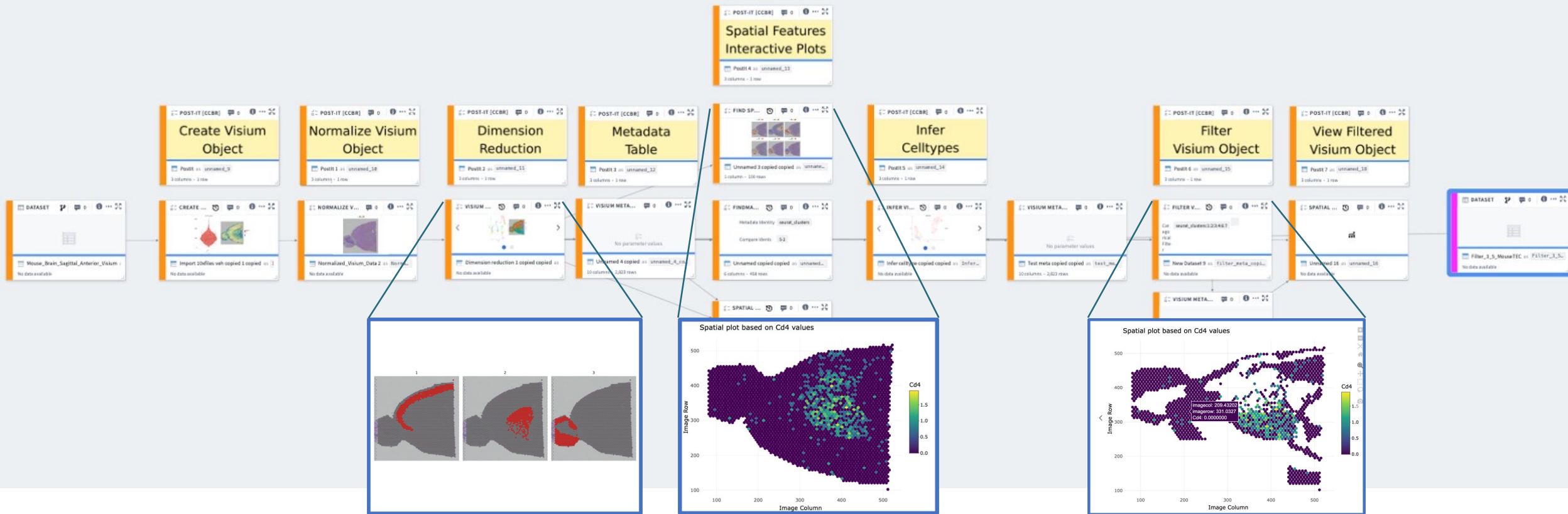
Visium Analysis Workflow



Visium Analysis Workflow

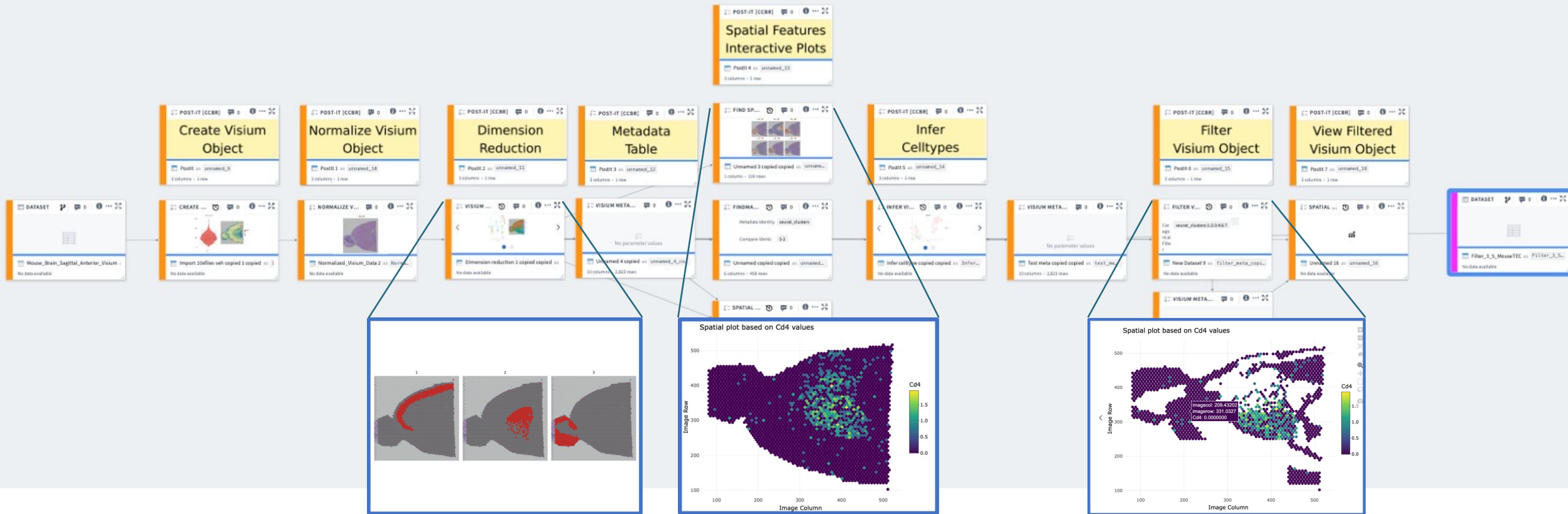


Visium Analysis Workflow



Visium Analysis Workflow

https://nidap.nih.gov/workspace/vector/view/ri.vector.main.workbook.28d315c3-cc23-4da5-bd97-3629afdd56a0?branch=Mouse_Brain_Sagittal_Anterior



Summary

- NIDAP allows NIH researchers to collaborate on analyses, develop new workflows, and speed research
- CCBR currently supports with full training courses:
 - Bulk RNA-seq on NIDAP
 - Single-cell RNA-seq on NIDAP (with scTCR-seq)
 - Digital Spatial Transcriptomics
- Additional courses and workflows soon:
 - Visium, Proteomics, CITE-seq, and more!
 - New Shiny Apps allowing even more interactive exploration of your data!
- If you are an NCI researcher, you can begin using NIDAP today

Bioinformatics support Provided to CCR through CCBR Workflows on NIDAP

- CCBR: CCR Collaborative Bioinformatics Resource:
- Collaborative Core (no fee)
 - Bioinformatics group operational at NCI since 2014
 - 10-12 Bioinformatics Analysts
- Our Mission:
 - Bioinformatics support for projects resulting in publications
 - Education and training of NCI researchers
- <https://bioinformatics.ccr.cancer.gov/ccbr/>
- The group of NIDAP developers include those from CCBR and several allied groups across NCI:
 - CCBR Developers:
 - Josh Meyer
 - Ned Cauley
 - Phil Homan
 - Alexei Lobanov
 - Jing Bian
 - Maggie Cam
 - Other Co-Developers and Collaborators:
 - Aleksandra Michalowski (CCR)
 - Chad Highfill (DCEG)
 - Difei Wang (DCEG)
 - Rui He (CBIIT)
 - George Zaki (CBIIT)

CCR Collaborative Bioinformatics Resource

Bioinformatics assistance to further CCR researchers' goals.

[Support Process →](#)



NIDAP Training

Online training for interactive CCBR workflows for bioinformatics analyses on NIDAP. Currently released workflows include: Bulk RNA-seq, Single-cell RNA-seq, and Digital Spatial Profiling (DSP).

[NIDAP Trainings](#)



Project Support

Learn how CCBR can assist with CCR Researchers with their projects.

[Explore Process](#)



Pipelines & Workflows

Workflows & Pipeline development. View whole exome & genome, single cell RNA-Seq, and ChIP-Seq pipeline examples.

[Workflows & Pipelines](#)

Questions?

Thank You!

NCICCBRNIDAP@mail.nih.gov

CCR COLLABORATIVE BIOINFORMATICS RESOURCE (CCBR)

The CCR Collaborative Bioinformatics Resource (CCBR) is a resource group which provides a mechanism for CCR researchers to obtain many different types of bioinformatics assistance to further their research goals. The group has expertise in a broad range of bioinformatics topics, and as such, its goal is to provide a simplified central access point for CCR researchers. The CCBR group includes members of the CCR Office of Science and Technology Resources (OSTR), Frederick National Laboratory for Cancer Research (FNLCR) and the Center for Biomedical Informatics and Information Technology (CBIIT). The CCBR may also direct projects to other available CCR bioinformaticians as needs demand. Requests for any type of Bioinformatics support should be through the CCBR Project Submission Form. Our main office is in Bethesda, Bldg 37, Rm 3041. Please contact Parthav Jailwala (parthav.jailwala@nih.gov, 240-760-6629) or Maggie Cam (maggie.cam@nih.gov, 240-760-7179) for any assistance.

[Weekly CCBR Walk-in Consultation Hours](#)

ASK FOR HELP

[Reach Out to CCBR](#)



[Check on Project Status](#)



PUBLICATIONS