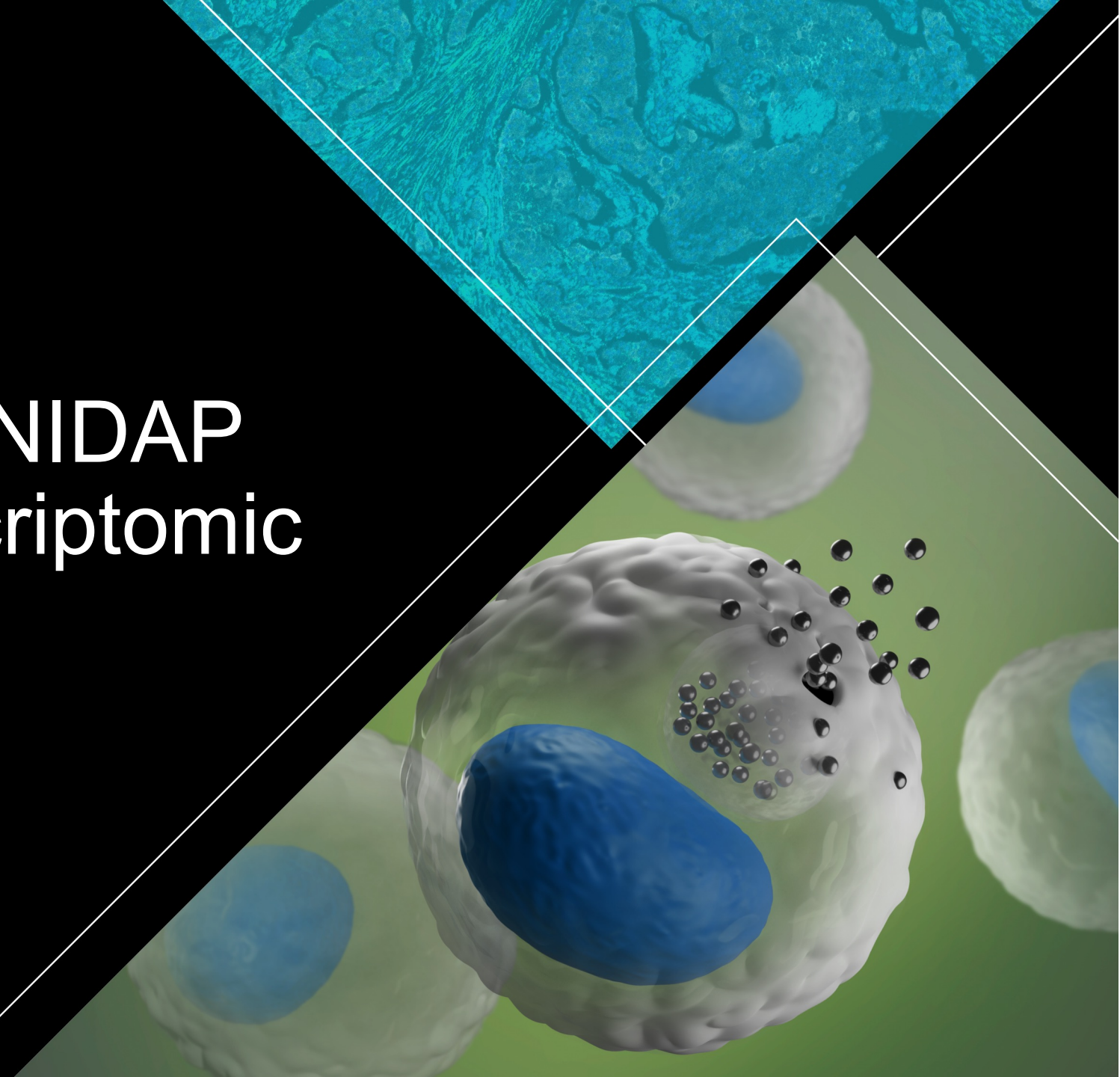
Capabilities of the NIDAP Platform for Transcriptomic Analysis

BTEP Webinar

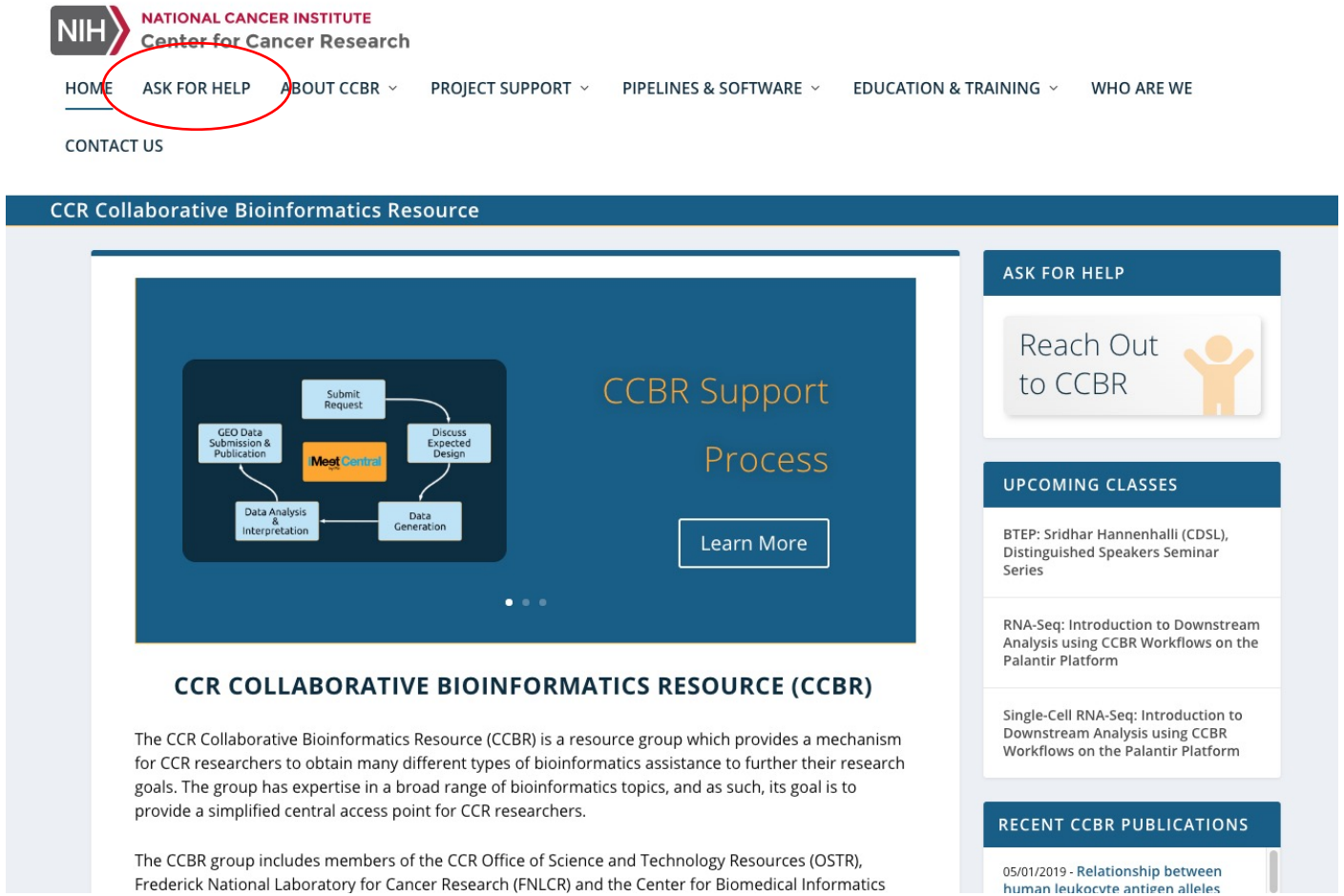
Dec. 14th, 2023

Joshua Meyer & Ned Cauley

SPONSORED BY THE NATIONAL CANCER INSTITUTE



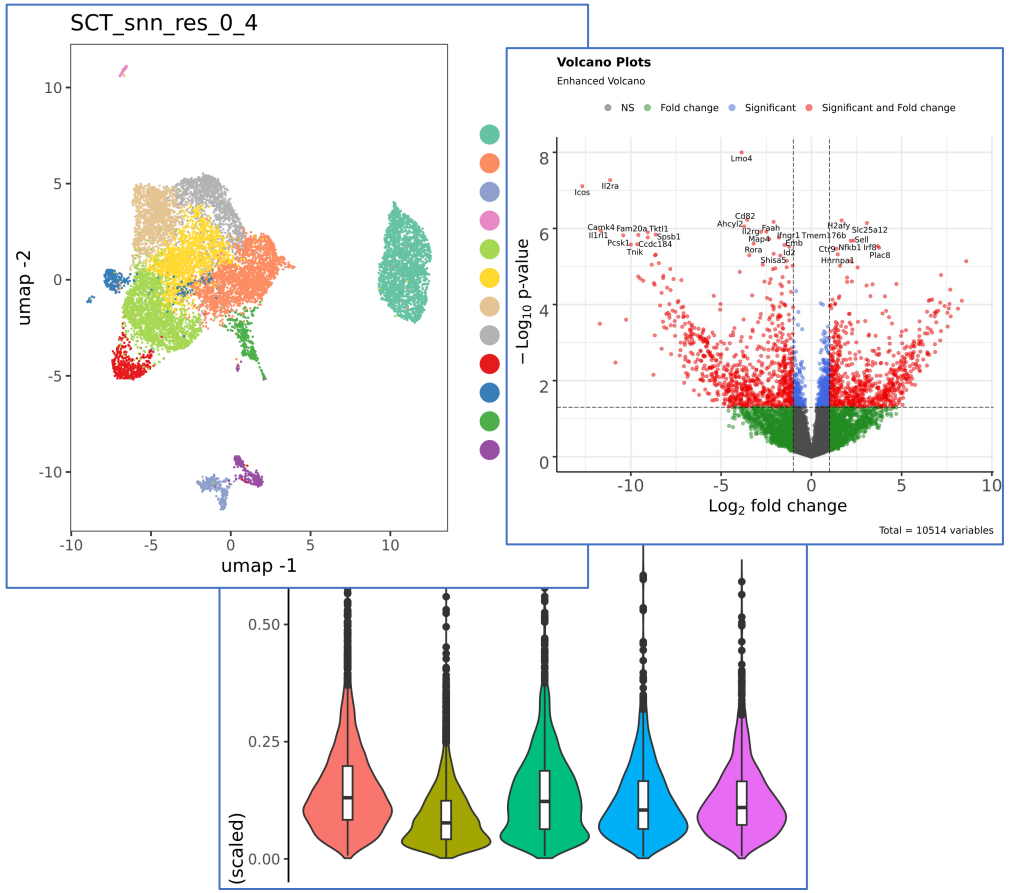
Who are we?
CCBR: CCR Collaborative Bioinformatics Resource



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

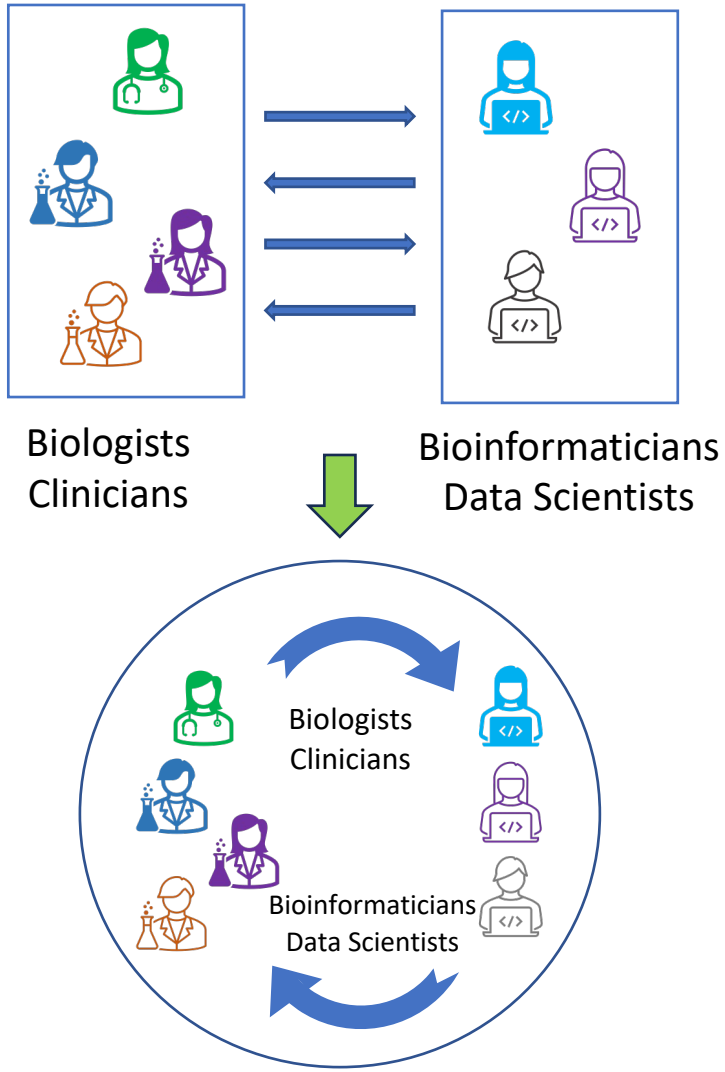
Overview

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



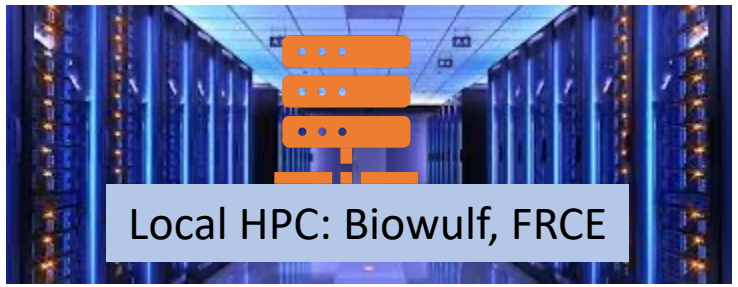
- **An interactive cloud-based interface for bioinformatics analyses, allowing:**
  - Collaborative development of up-to-date bioinformatics workflows by CCR bioinformaticians
  - Implementation of standardized, user-friendly code templates that are disseminated among collaborators and instructed to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

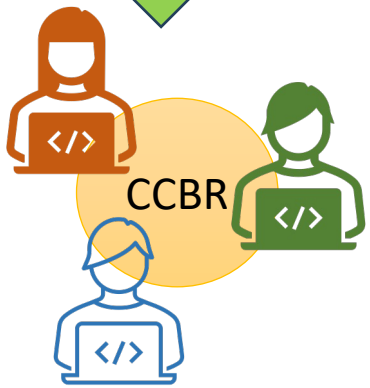
NIDAP facilitates access to and deployment of CCBR Data Analysis Workflows

Storage (DME)

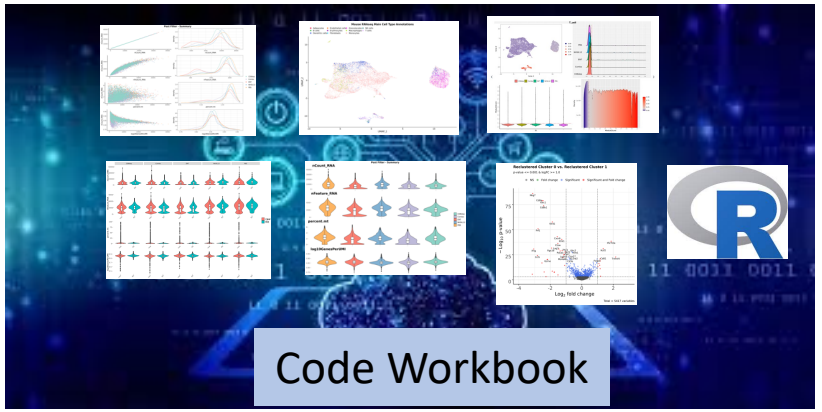
NIDAP

Pangea Pipelines: RNA-Seq, Exome-Seq





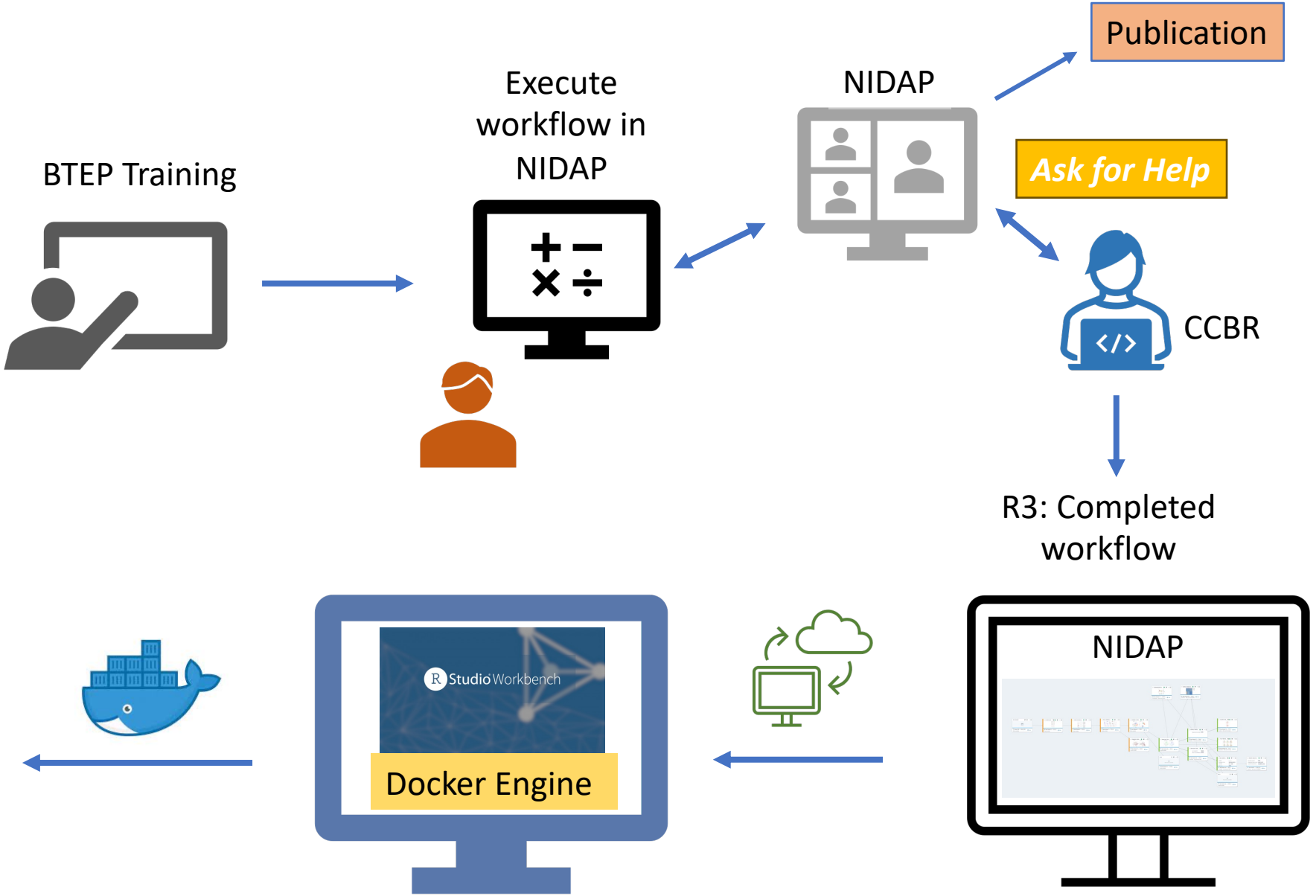
NIDAP Workflows: Bulk, Single Cell, DSP



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



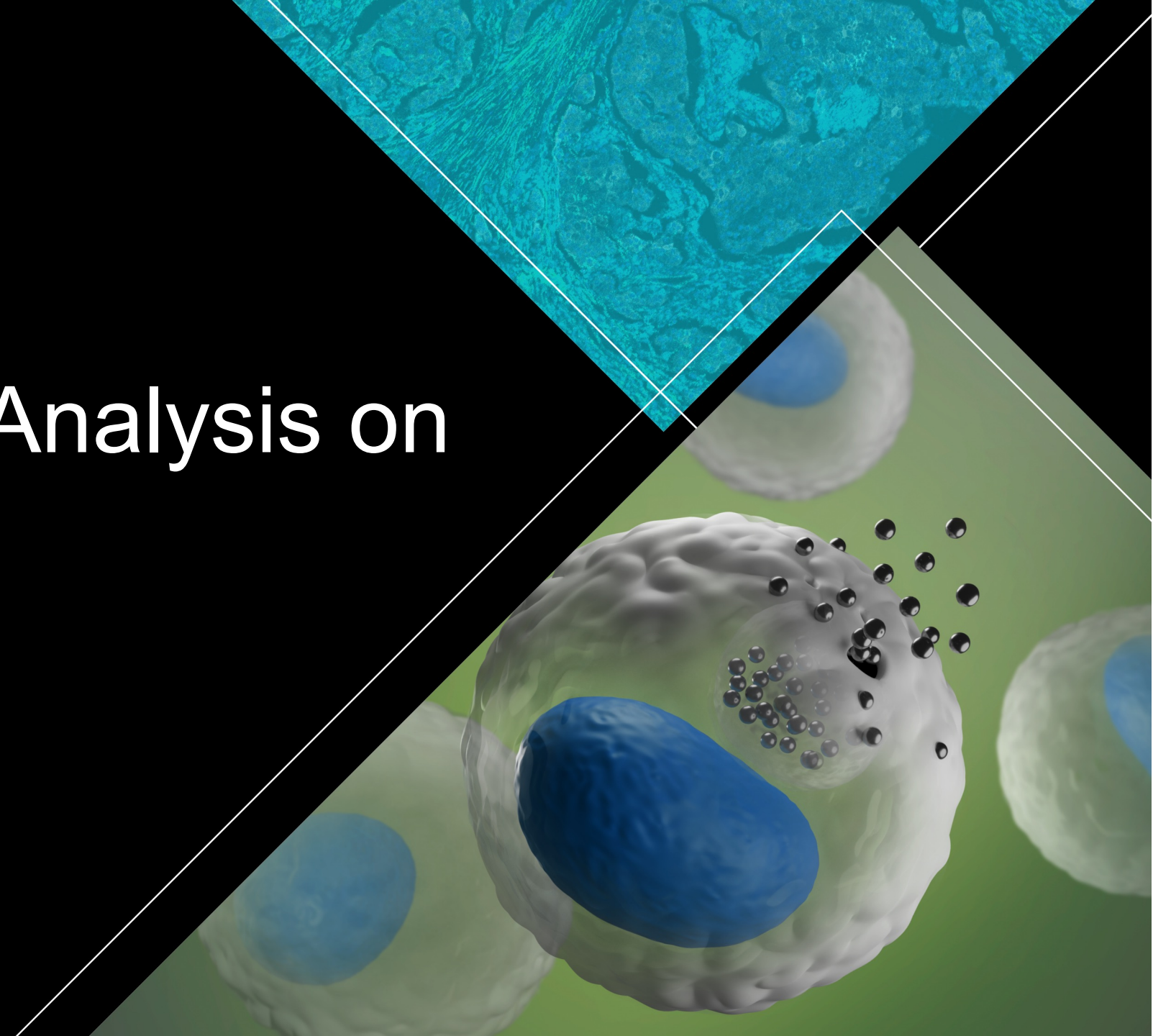
Demo:
The NIDAP Website

- 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

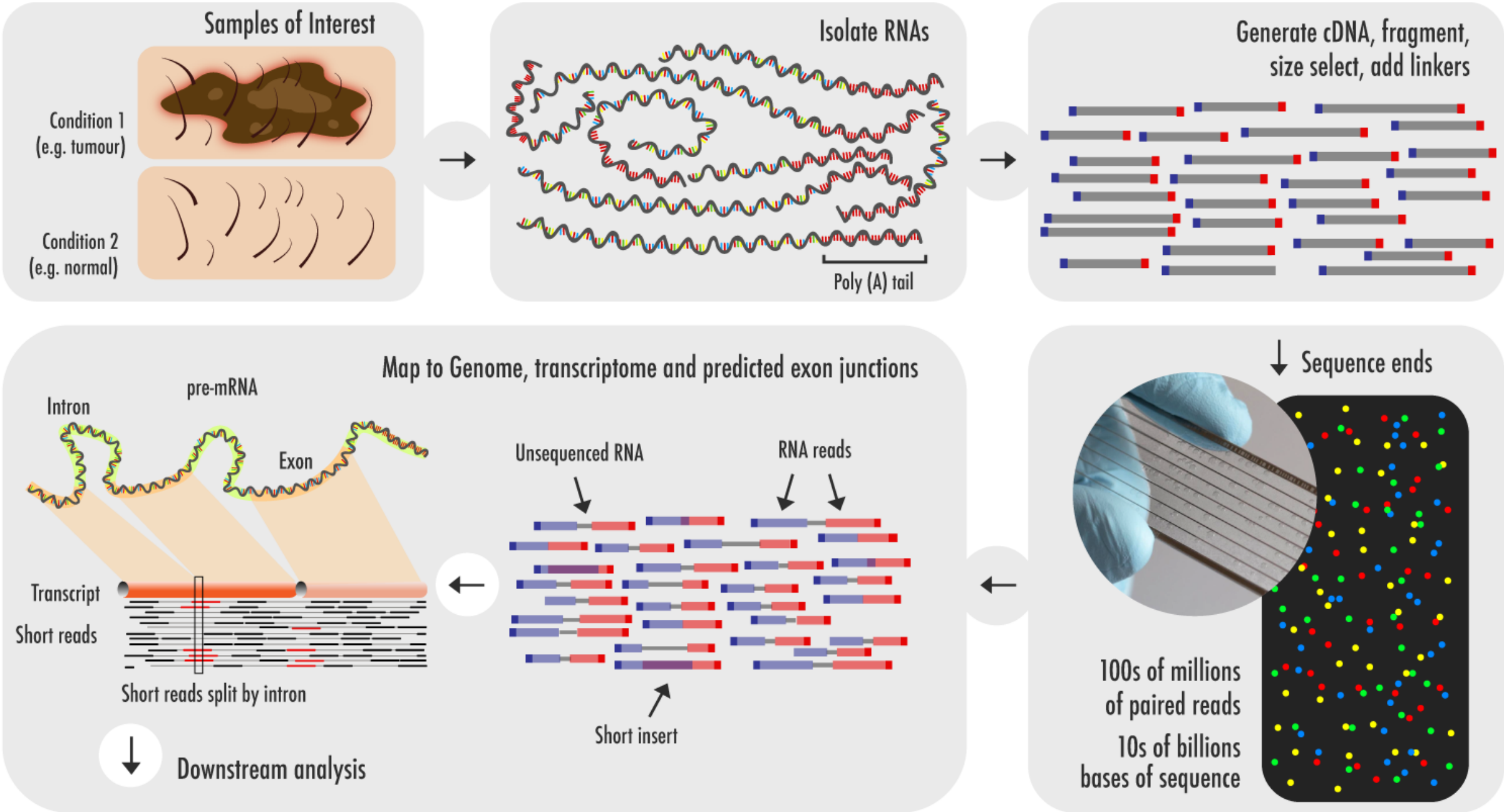
Bulk RNA-seq Analysis on NIDAP

Thomas Joshua Meyer

(Josh)



SPONSORED BY THE NATIONAL CANCER INSTITUTE

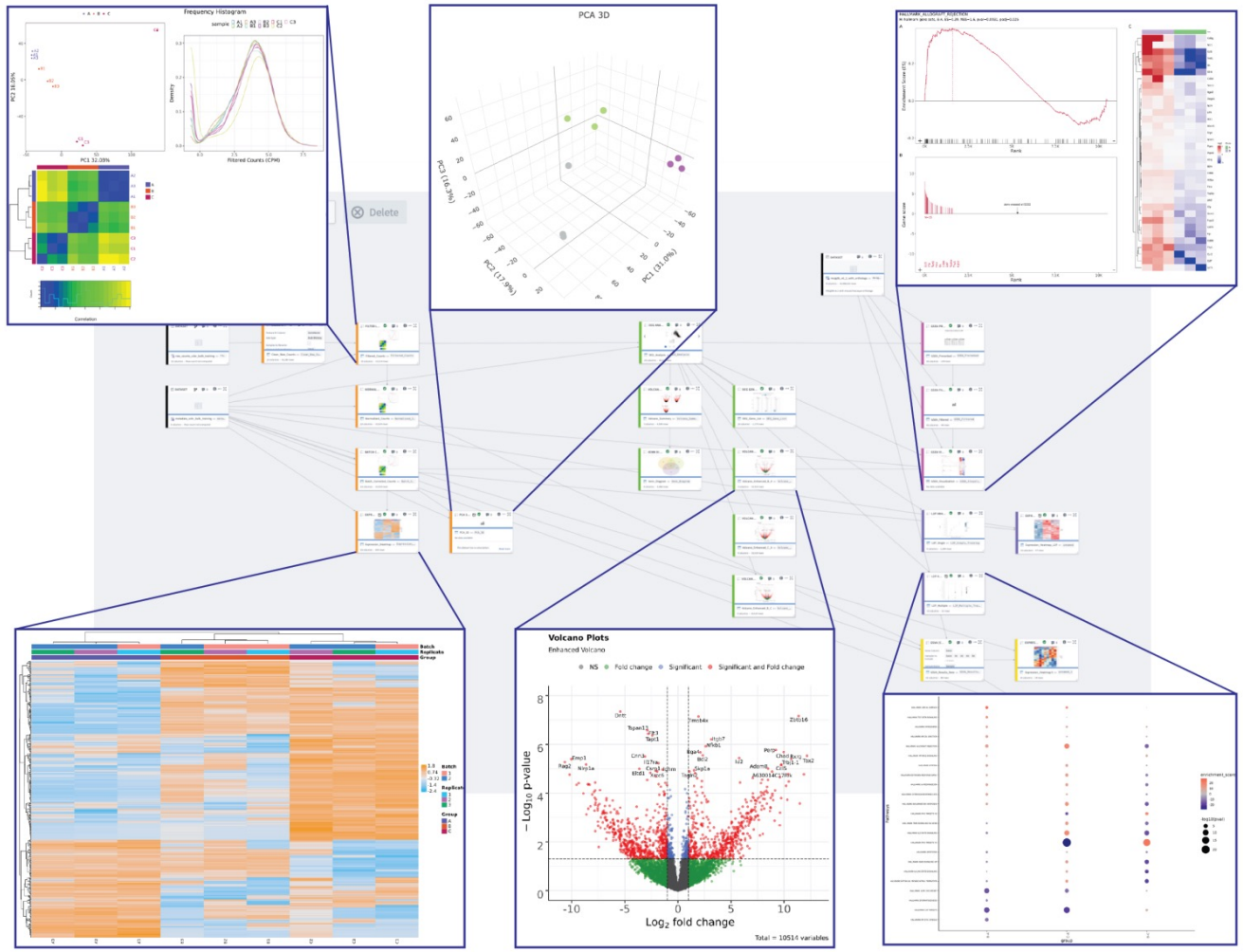


https://www.technologynetworks.com/genomics/articles/rna-seq-basics-applications-and-protocol-299461

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

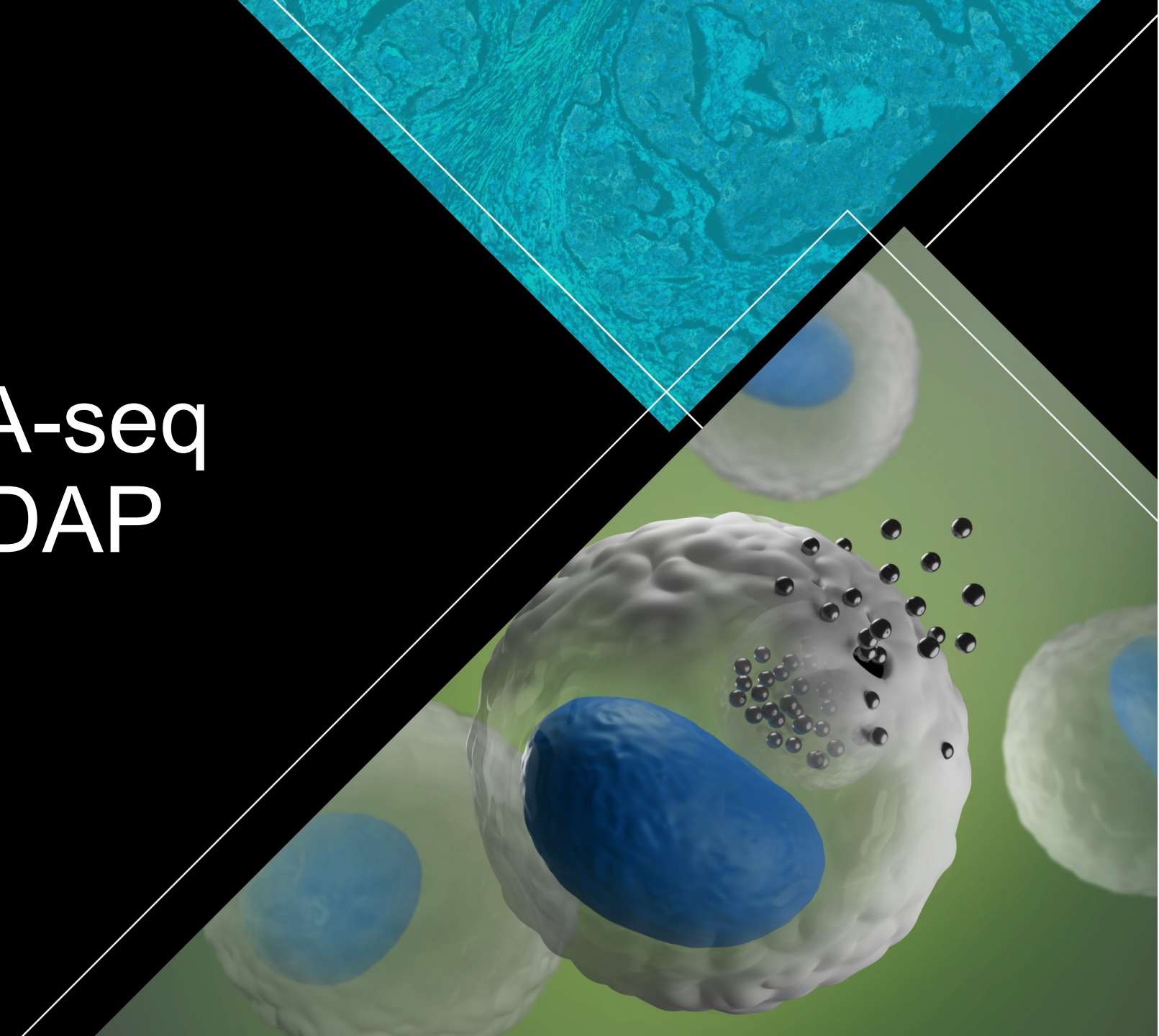
Demo:
The CCBR Bulk RNA-seq Workflow



Single-cell RNA-seq Analysis on NIDAP

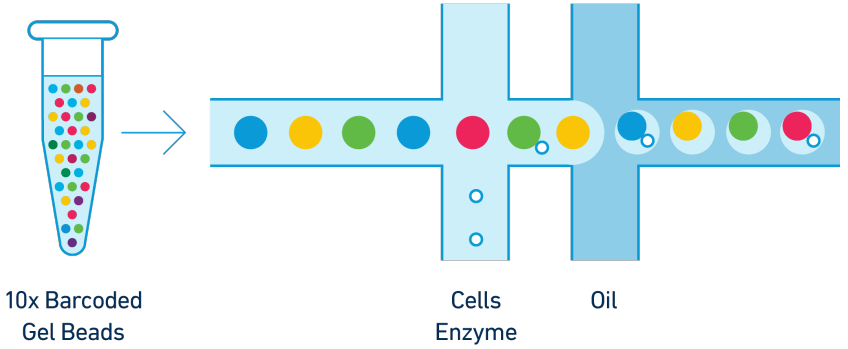
Thomas Joshua Meyer

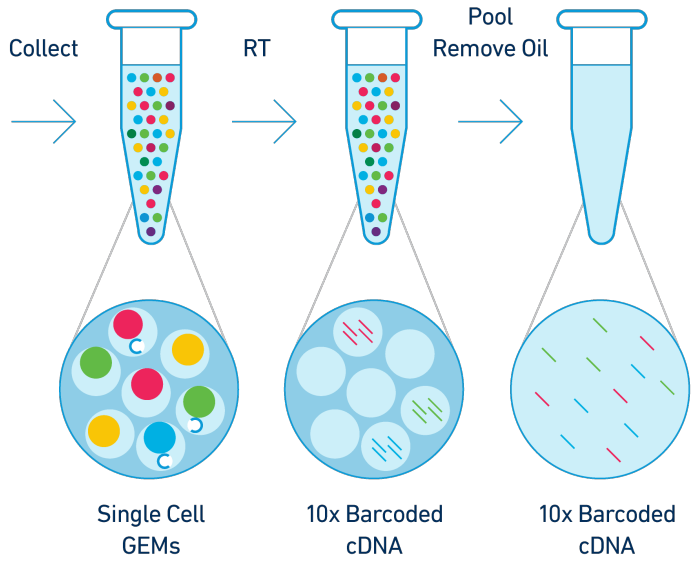
(Josh)

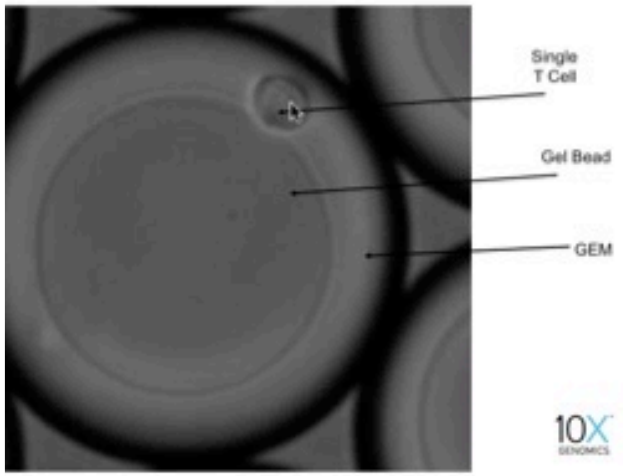


SPONSORED BY THE NATIONAL CANCER INSTITUTE

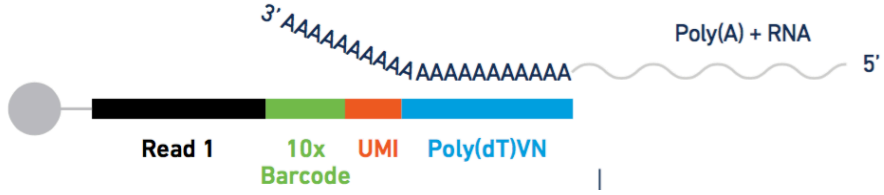
Droplet-based scRNA-Seq allows high-throughput profiling with unprecedented ease *(for good samples)*

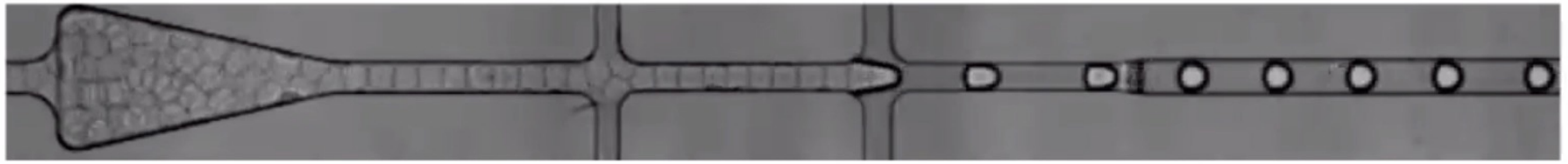






- Cell partitioned with barcoded gel bead
- Cell Barcode & Unique Molecular Identifiers (UMI) on end of cDNA
- ~2000 median genes detected on average with ~50,000 mean sequencing reads per cells
- Gene-level counts data



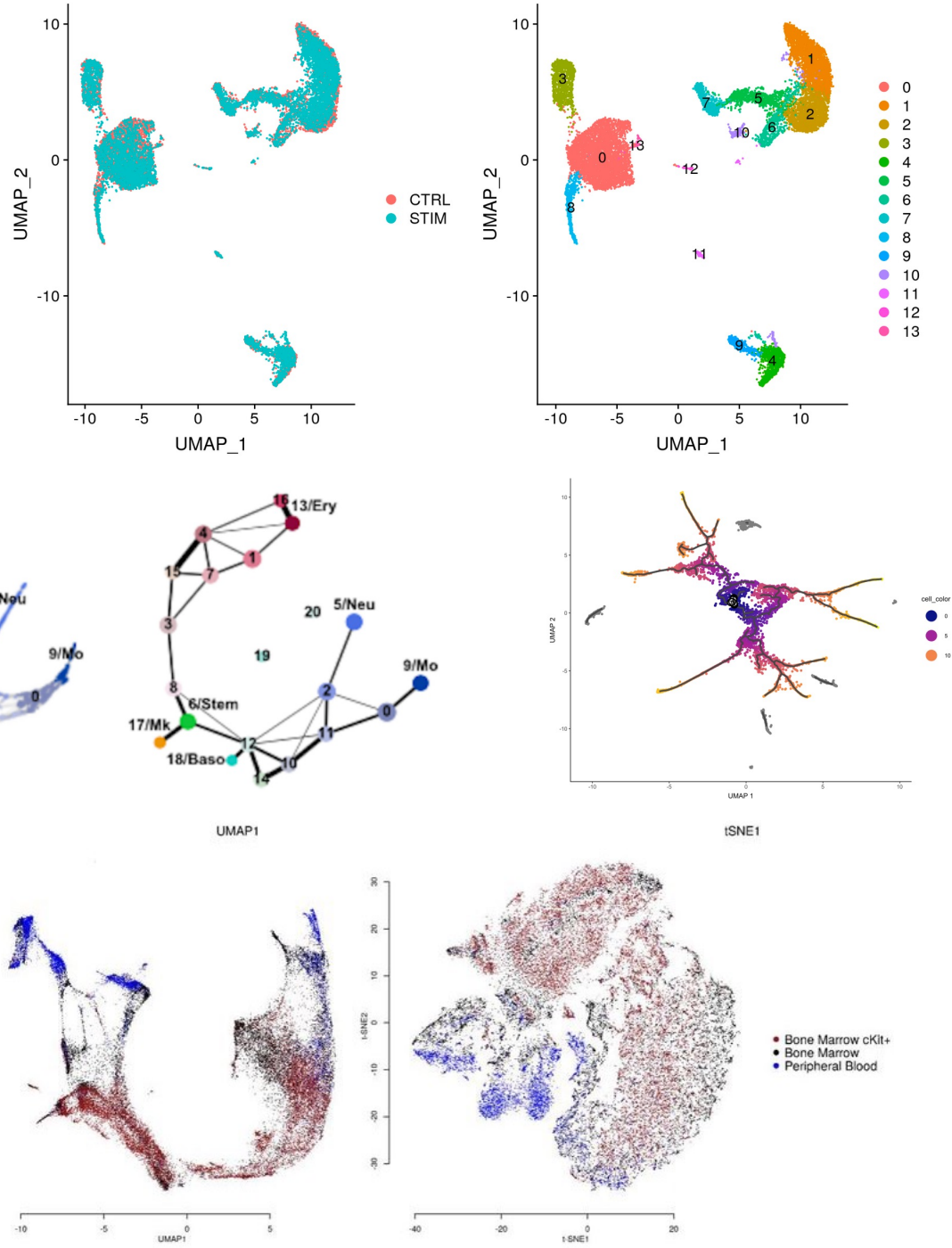


When you receive your data

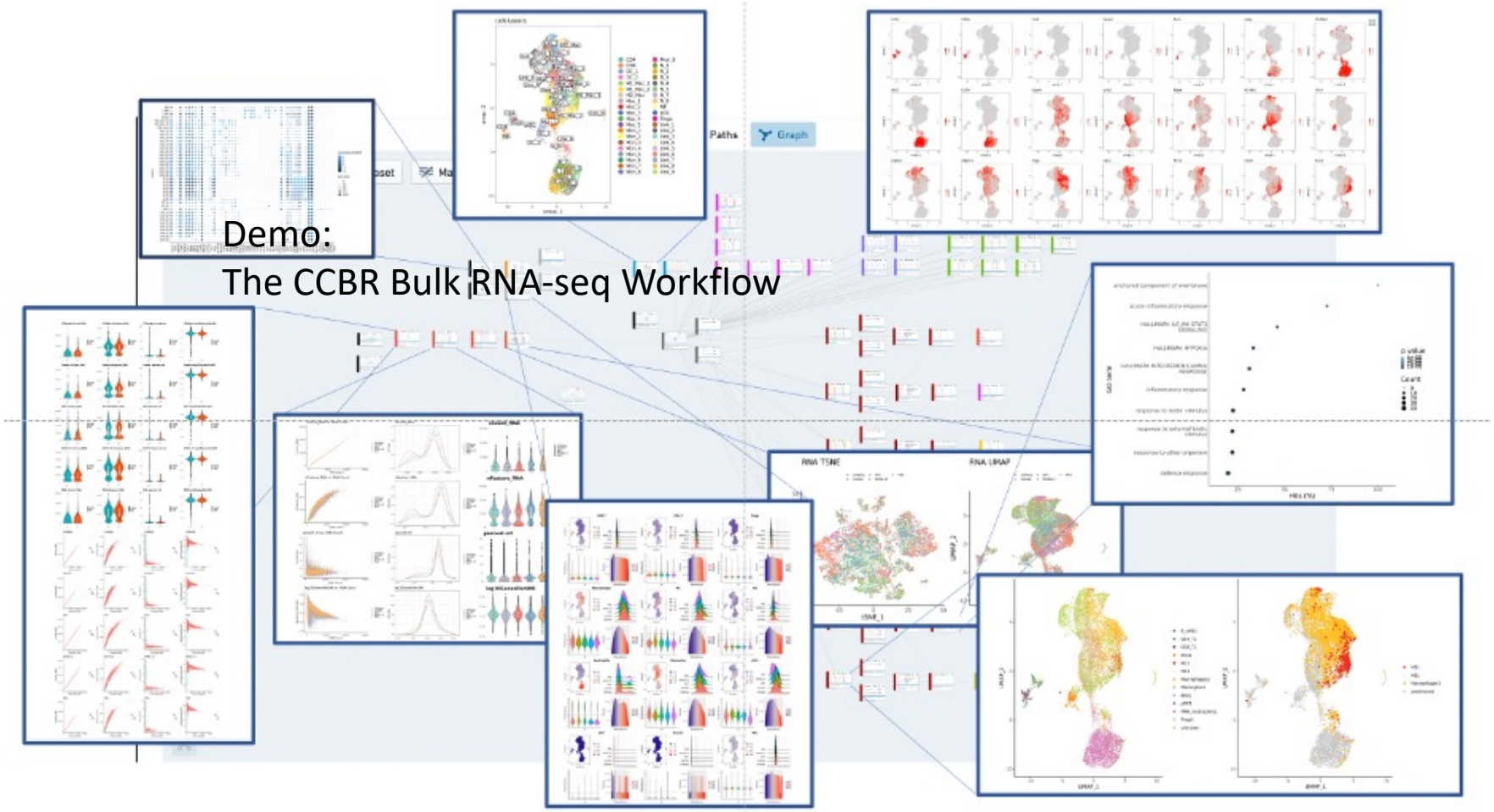
- The data is delivered by the sequencing machines as BCL (raw base call) files
- The CellRanger software from 10x Genomics performs the following steps:
  - Make a FASTQ from the BCL files
  - Sort and count reads for each cell based on unique molecular identifiers (UMIs)
  - [Optional] Aggregate multiple 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)

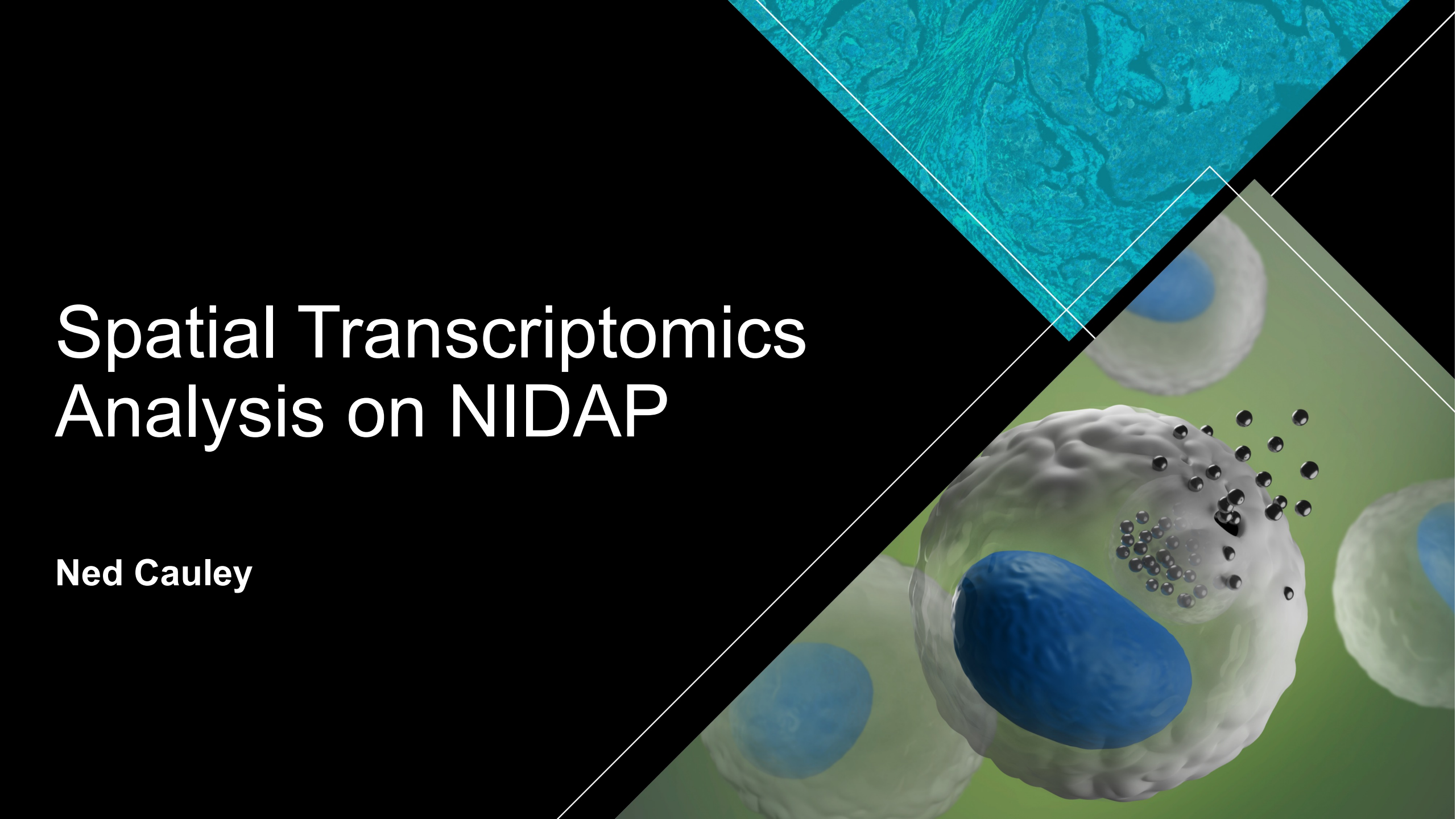


Demo:
The CCBR Single-cell RNA-seq Workflow

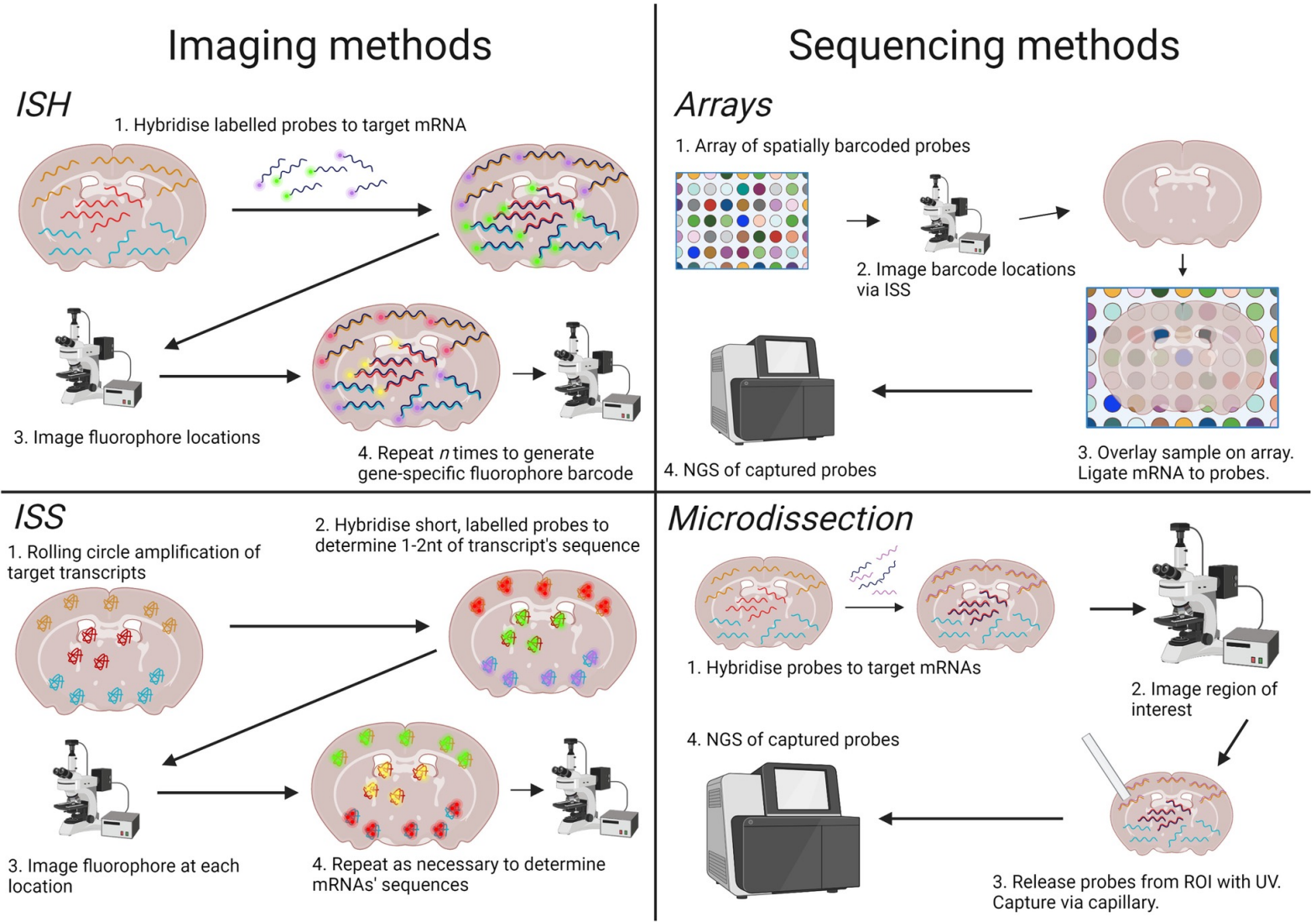


Spatial Transcriptomics Analysis on NIDAP

Ned Cauley

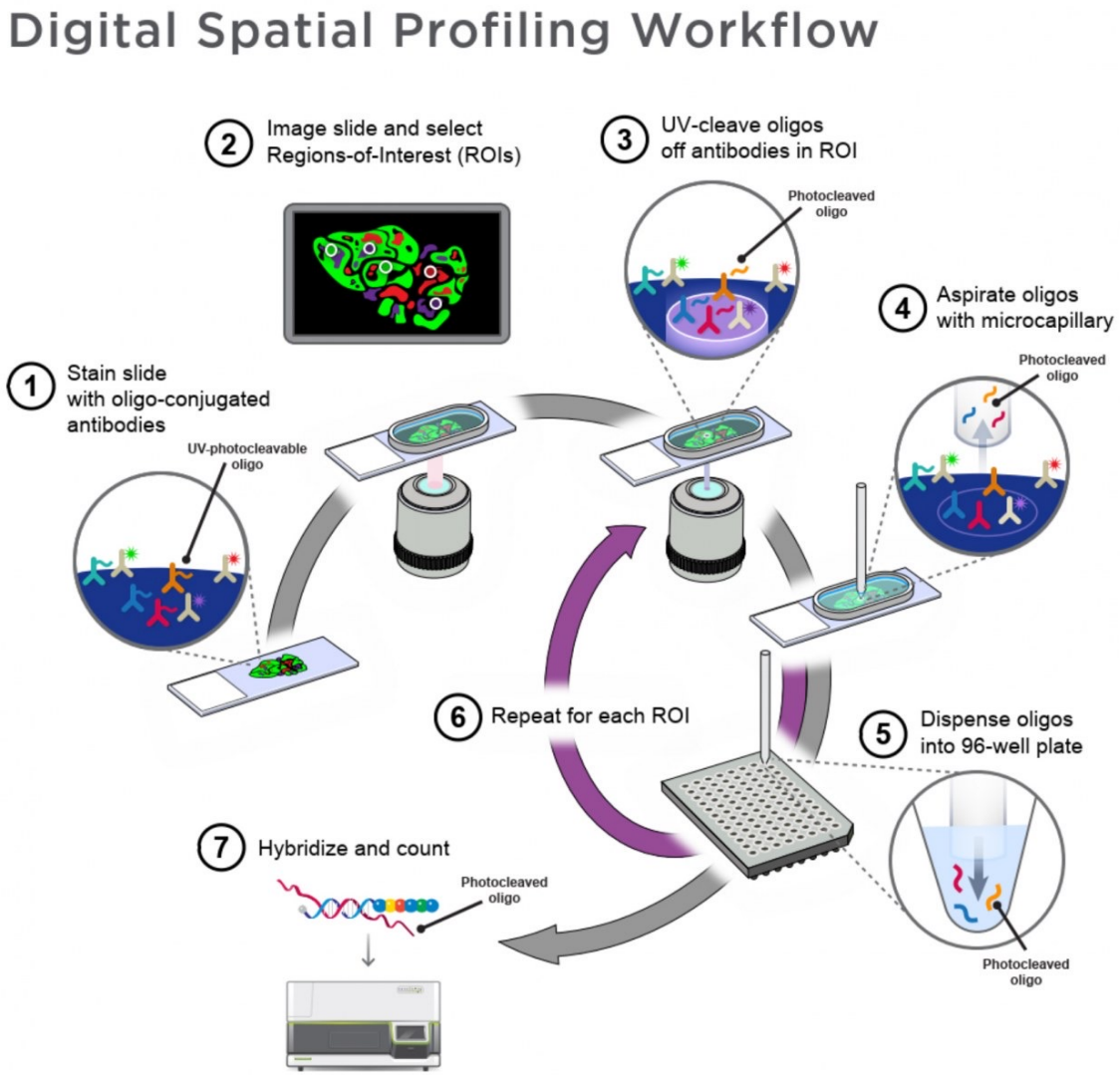


Spatial Transcriptomics on NIDAP



- **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

Spatial Transcriptomics on NIDAP

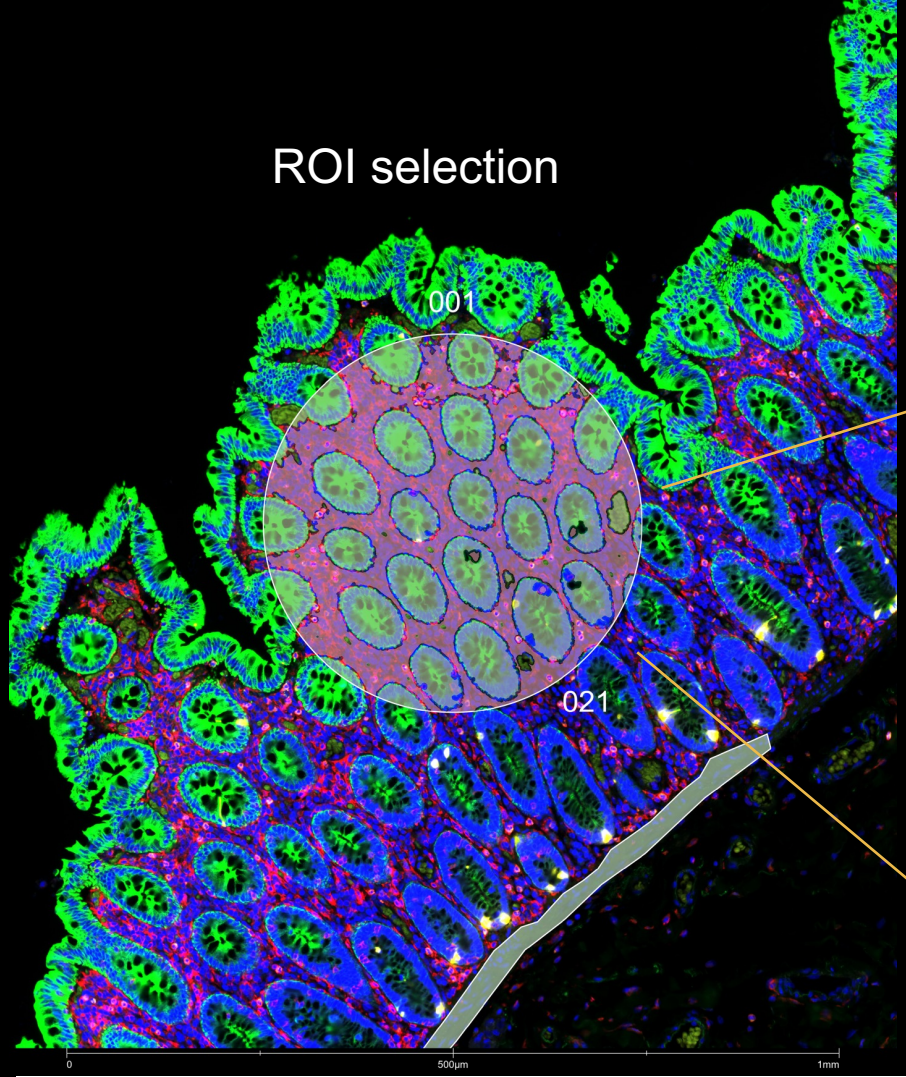


- **Workflows for the analysis of sequencing methods**

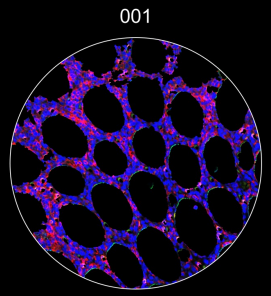
- **Digital Spatial Profiling**
  - Targeted regions of interest
  - Complete workflow available on NIDAP
  - Training videos
  - Example analysis

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

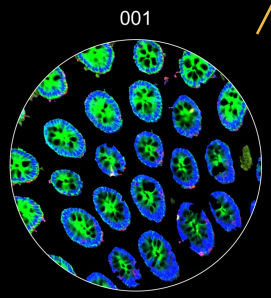
ROI selection



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

ProbeID	Reads
1111	20
2222	10
3333	35

Morphological Markers

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

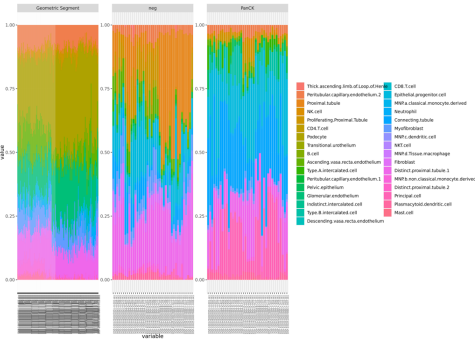
DSP Analysis on NIDAP

- **DCC files**
- **Annotation**
- **PKC mapping**

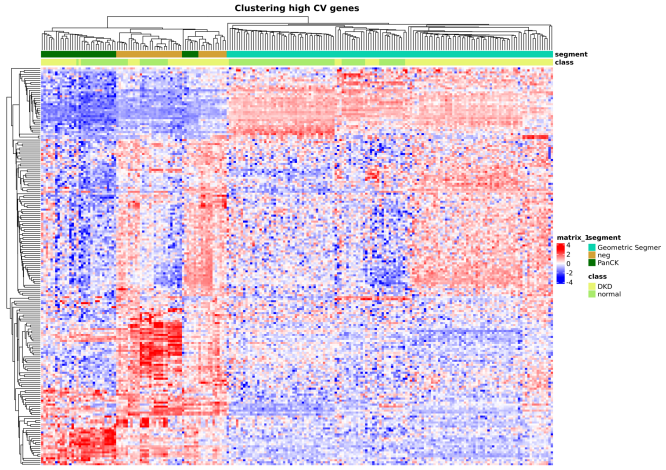
NIDAP

QC
Filtering
Normalization

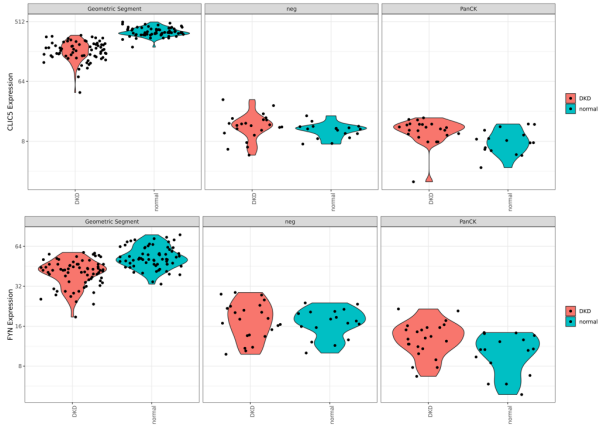
Spatial deconvolution



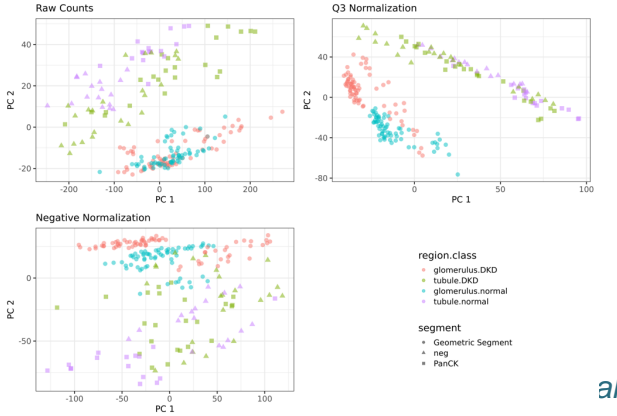
Heatmap



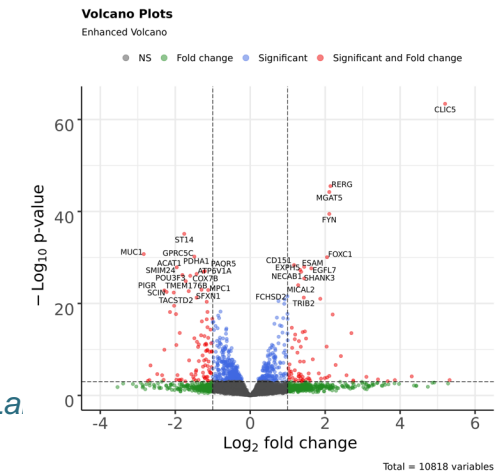
Violin Plot



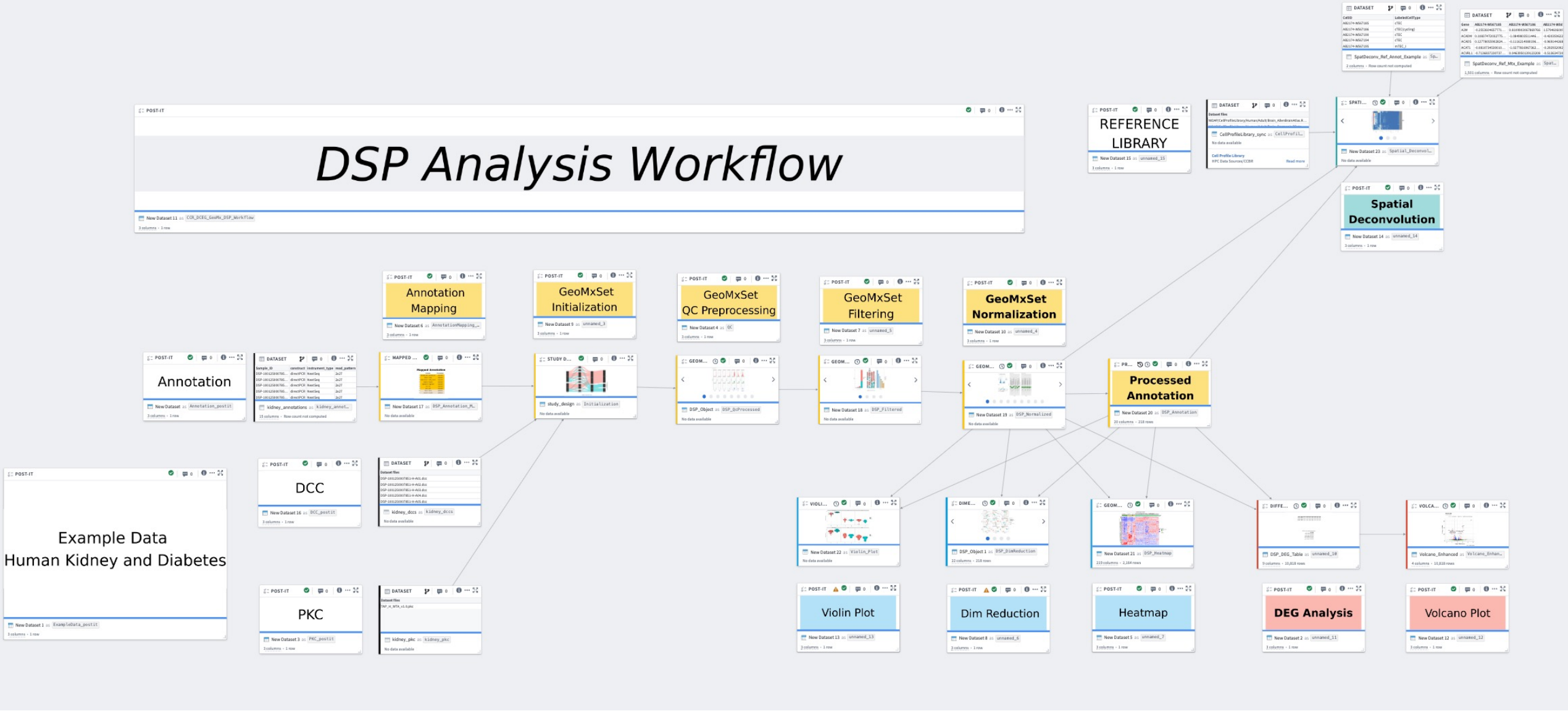
PCA

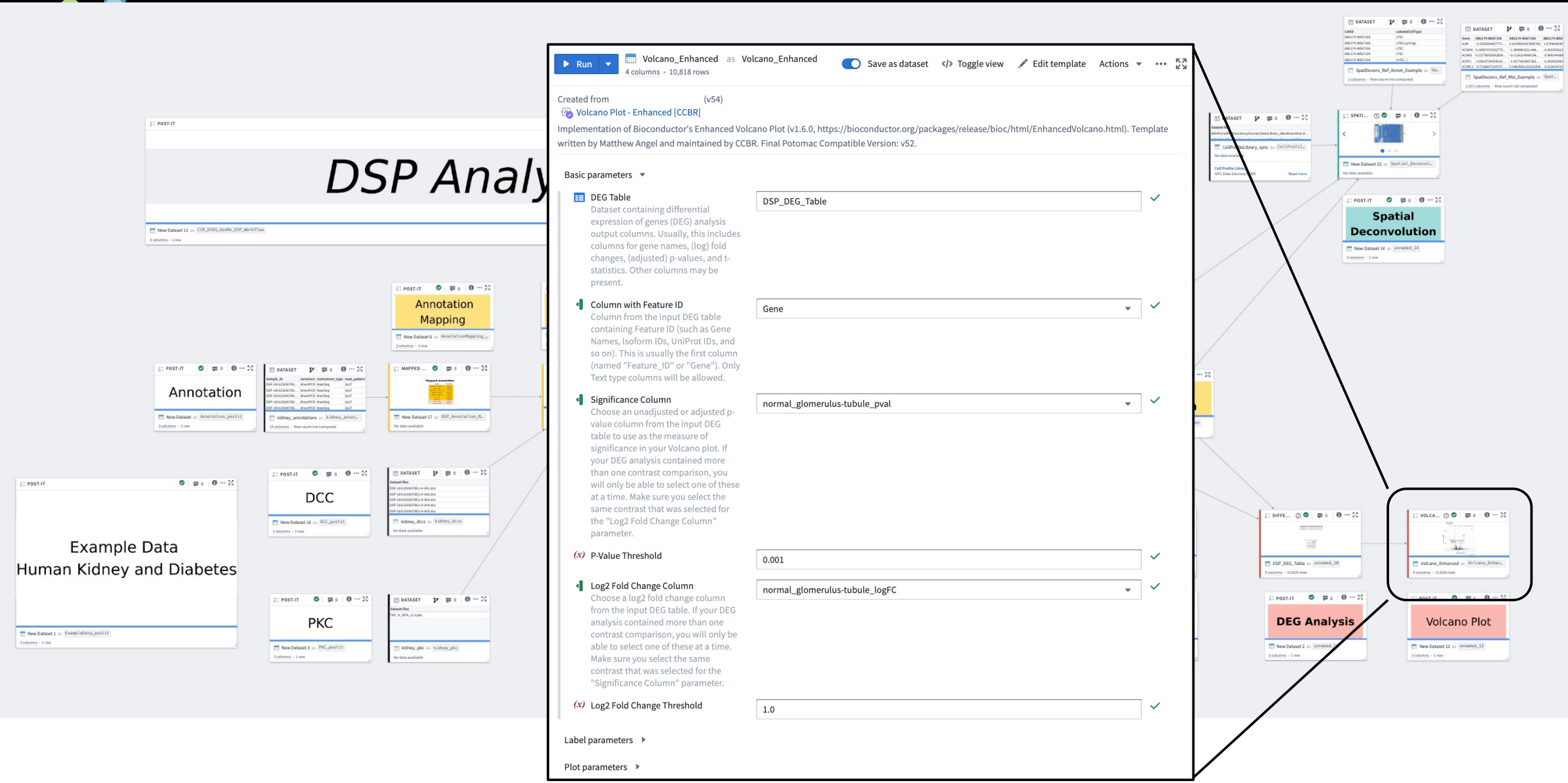


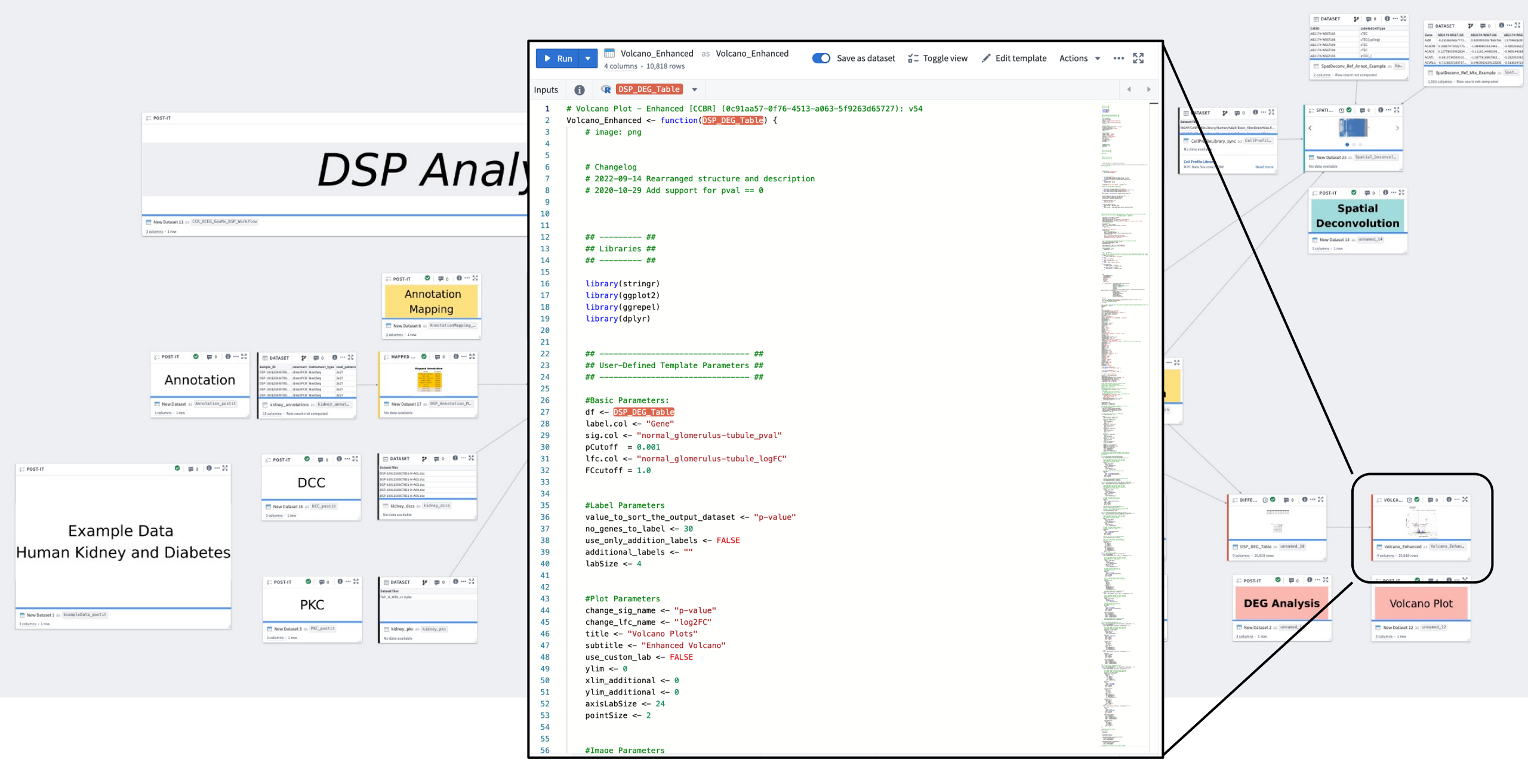
Differential Expression



DSP Analysis Workflow



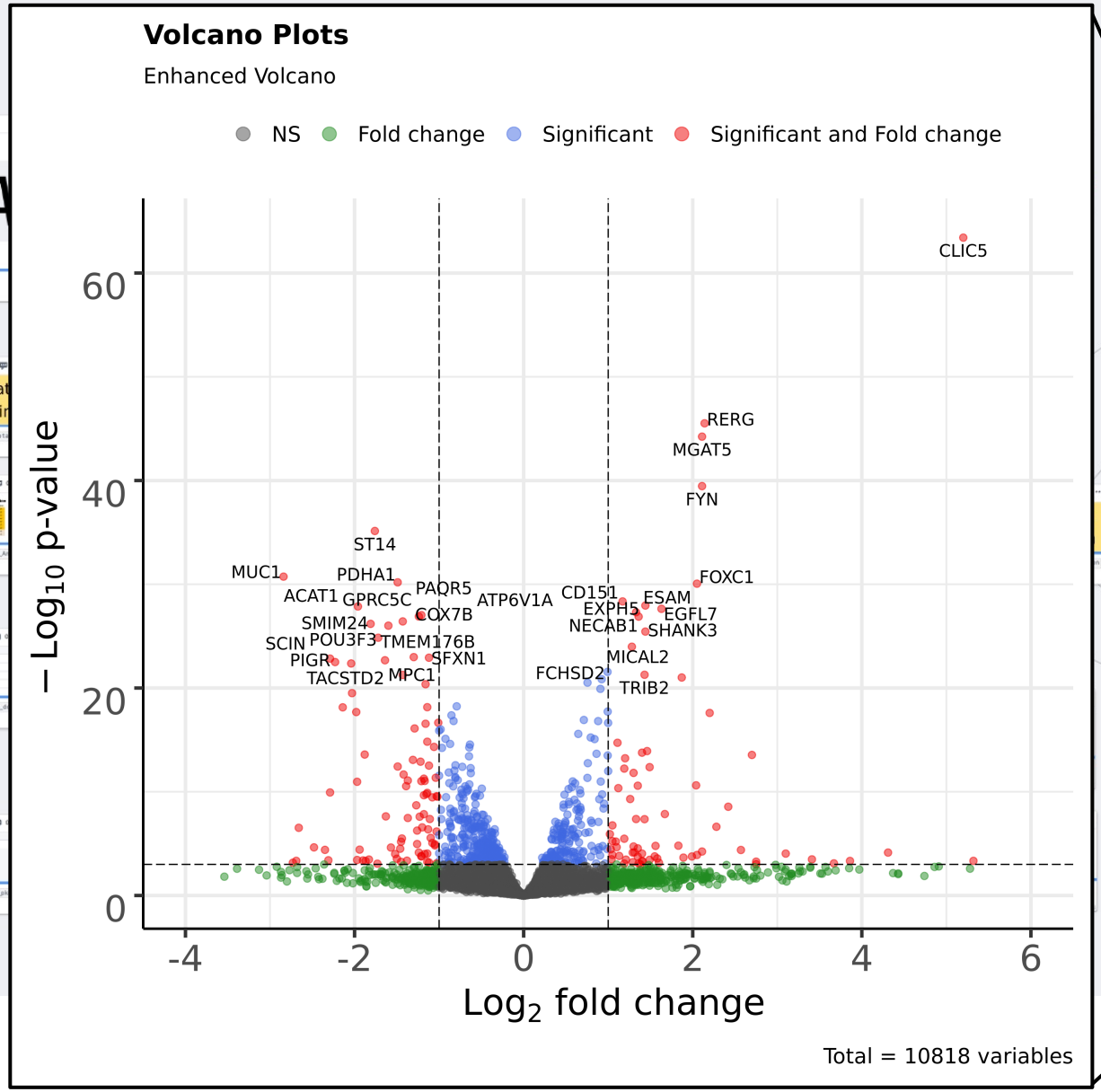

DSP Analy
Annotation Mapping
Annotation
DCC
PKC
Example Data
Human Kidney and Diabetes
Run
Volcano_Enhanced as Volcano_Enhanced
4 columns • 10,818 rows
Save as dataset
Toggle view
Edit template
Actions
Created from (v54)
Volcano Plot - Enhanced [CCBR]
Implementation of Bioconductor's Enhanced Volcano Plot (v1.6.0, https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html). Template written by Matthew Angel and maintained by CCBR. 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
Spatial Deconvolution
DEG Analysis
Volcano Plot


Run
Volcano_Enhanced as Volcano_Enhanced
4 columns · 10,818 rows
Save as dataset
Toggle view
Edit template
Actions
Inputs
DSP_DEG_Table
# Volcano Plot - Enhanced [CCBR] (0c91aa57-0f76-4513-a063-5f9263d65727): v54
Volcano_Enhanced <- function(DSP_DEG_Table) {
# image: png
# Changelog
# 2022-09-14 Rearranged structure and description
# 2020-10-29 Add support for pval == 0

Libraries
----
library(stringr)
library(ggplot2)
library(ggrepel)
library(dplyr)

User-Defined Template Parameters

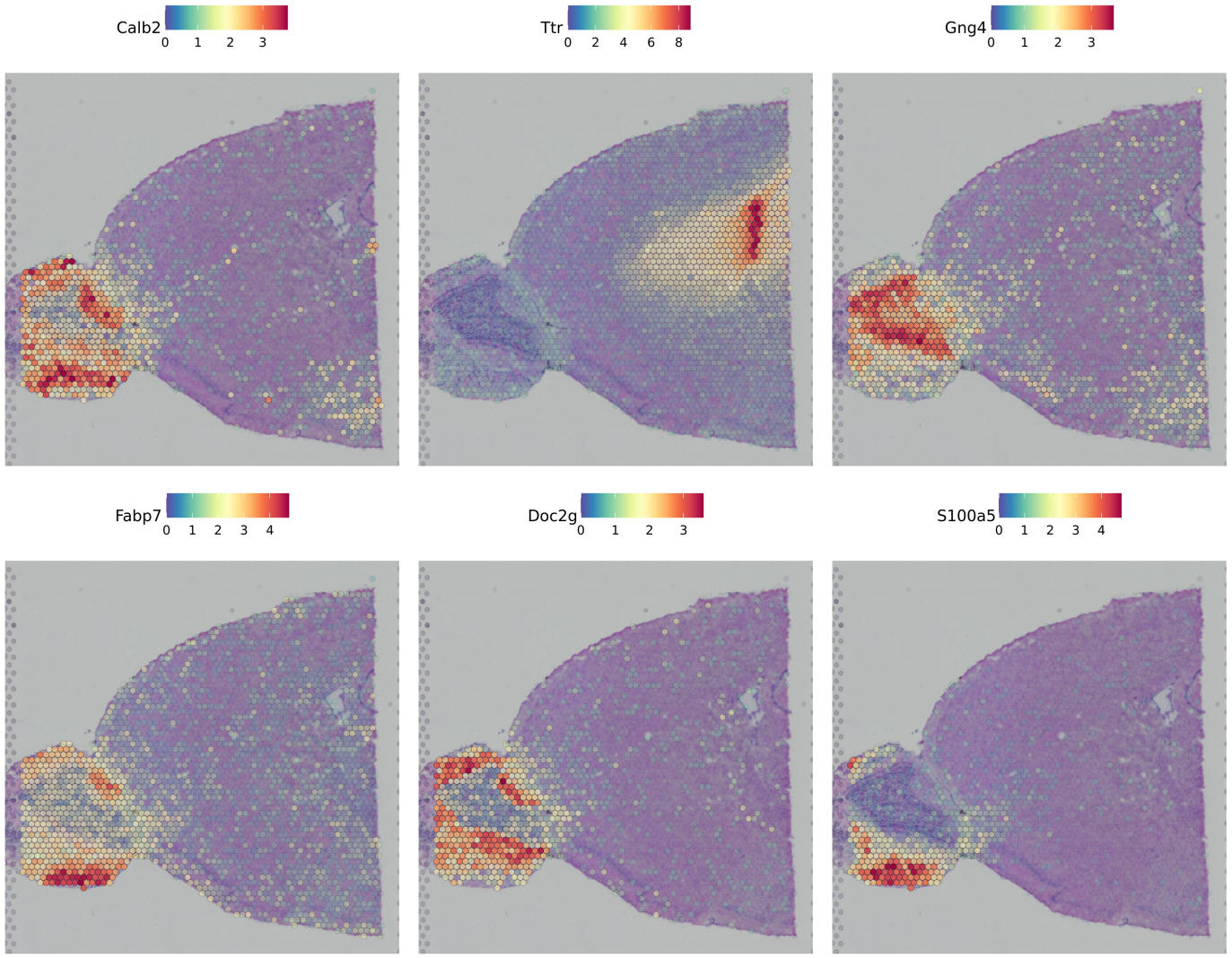
#Basic Parameters:
df <- DSP_DEG_Table
label.col <- "Gene"
sig.col <- "normal_glomerulus-tubule_pval"
pCutoff = 0.001
lfc.col <- "normal_glomerulus-tubule_logFC"
FCcutoff = 1.0
#Label Parameters
value_to_sort_the_output_dataset <- "p-value"
no_genes_to_label <- 30
use_only_addition_labels <- FALSE
additional_labels <- ""
labSize <- 4
#Plot Parameters
change_sig_name <- "p-value"
change_lfc_name <- "log2FC"
title <- "Volcano Plots"
subtitle <- "Enhanced Volcano"
use_custom_lab <- FALSE
ylim <- 0
xlim_additional <- 0
ylim_additional <- 0
axisLabSize <- 24
pointSize <- 2
#Image Parameters
DSP Analy
Annotation Mapping
Annotation
DCC
PKC
Example Data
Human Kidney and Diabetes
Spatial Deconvolution
DEG Analysis
Volcano Plot


Volcano Plots
Enhanced Volcano
NS
Fold change
Significant
Significant and Fold change
CLIC5
RERG
MGAT5
FYN
FOXC1
ST14
MUC1
PDHA1
PAQR5
ACAT1
GPRC5C
ATP6V1A
CD151
ESAM
EXPH5
EGFL7
COX7B
NECAB1
SHANK3
SMIM24
SCIN
POU3F3
TMEM176B
PIGR
SFXN1
MICAL2
TACSTD2
MPC1
FCHSD2
TRIB2
60
40
20
0
−Log10 p-value
-4
-2
0
2
4
6
Log2 fold change
Total = 10818 variables

DSP Workflow on NIDAP

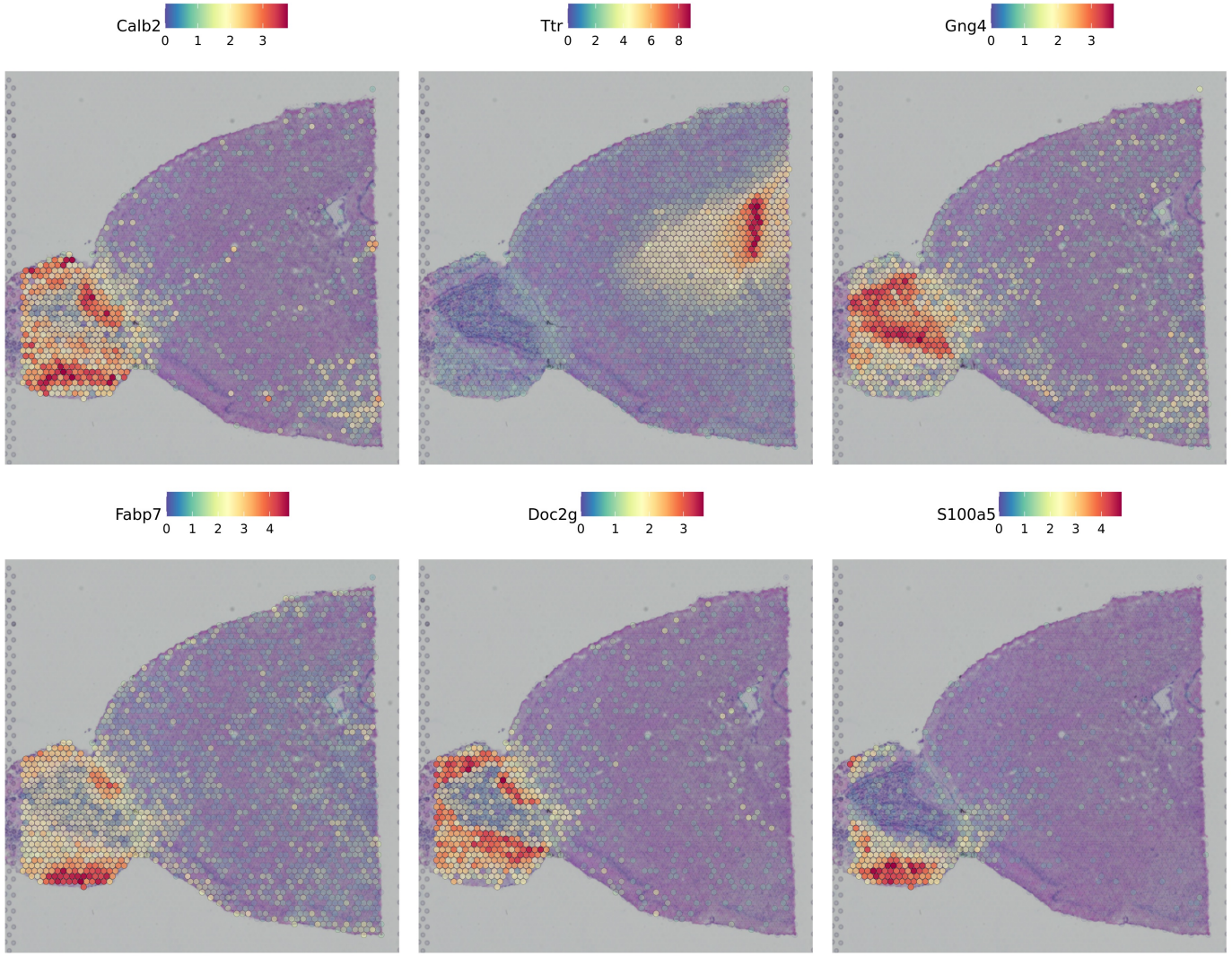
- https://nidap.nih.gov/workspace/vector/view/ri.vector.main.workbook.67a84967-f588-496a-b81f-e8dbdd1dee6d?branch=test

Visium Workflow on NIDAP



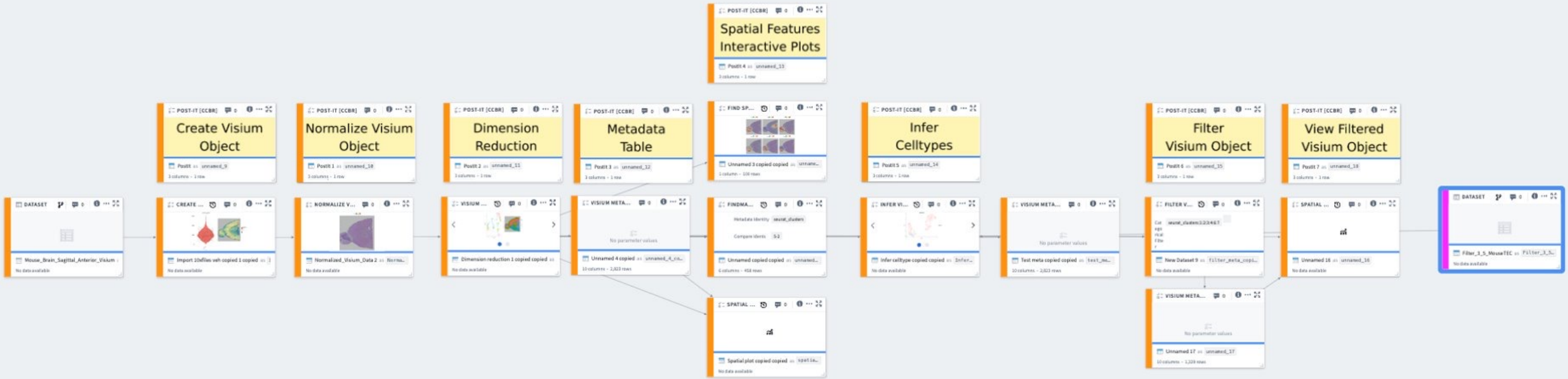
- **Workflows for the analysis of sequencing methods**
- **Digital Spatial Profiling**
  - Targeted regions of interest
  - Complete workflow available on NIDAP
  - Training videos
  - Example analysis
- **Visium**
  - Uniform grid of regions of interest
  - Beta Workflow available on NIDAP

Visium Workflow on NIDAP

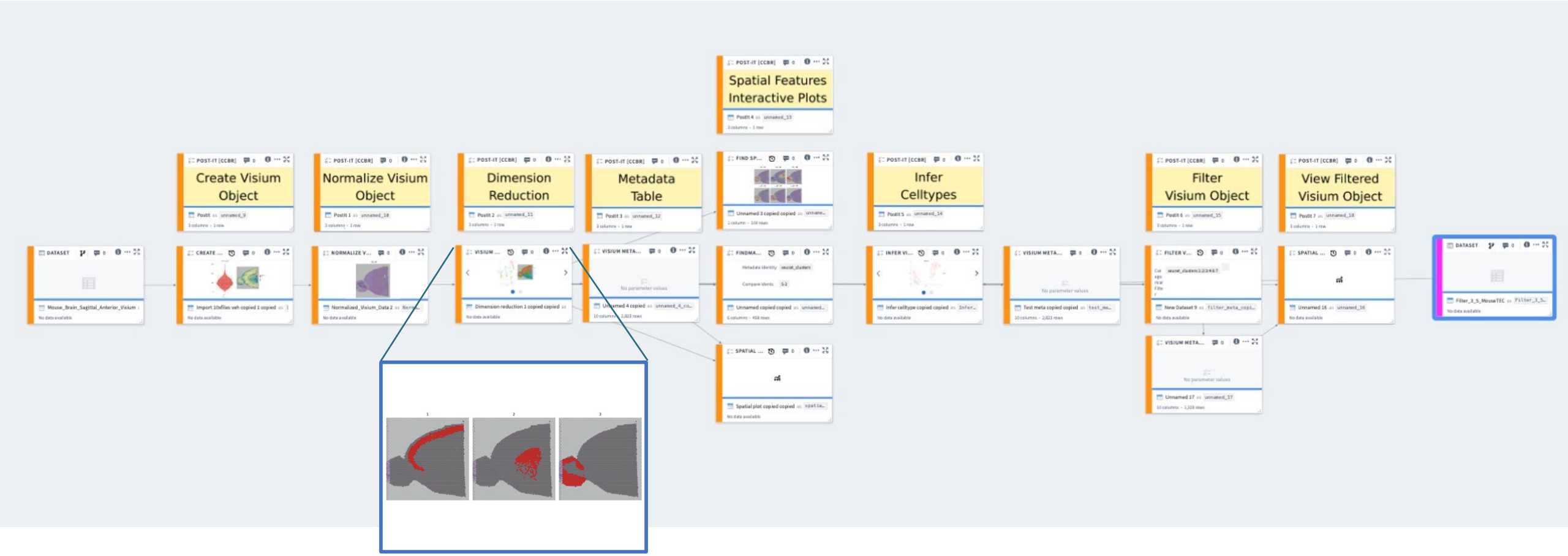


- Based on the Seurat Visium vignette
- Workflow currently in “beta”
  - Functional workflow that has been tested successfully
  - No video tutorials (yet!)
- Currently available for any NIDAP user to use

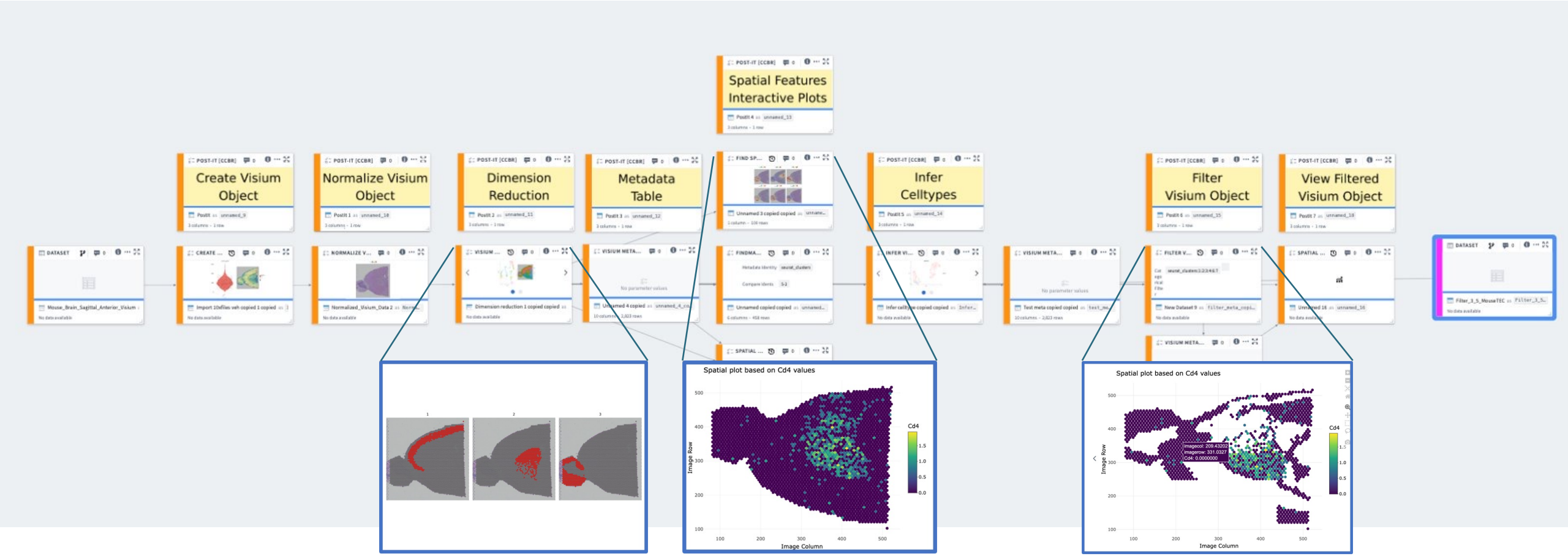
Visium Analysis Workflow



Visium Analysis Workflow

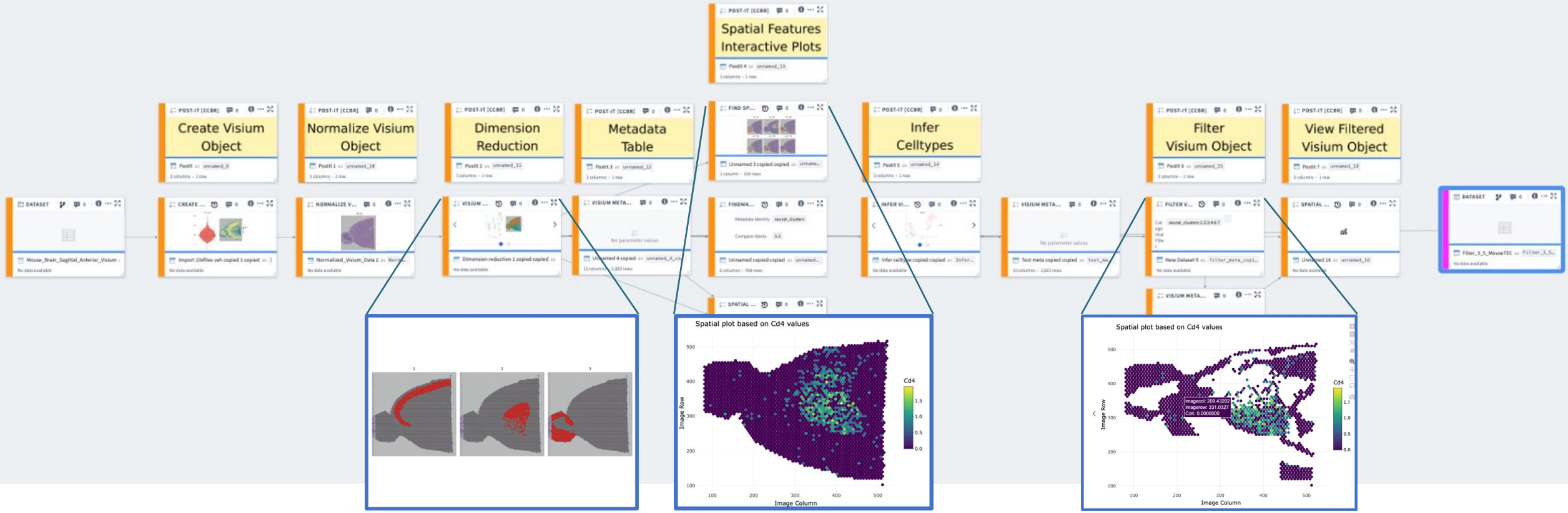


Visium Analysis Workflow



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/

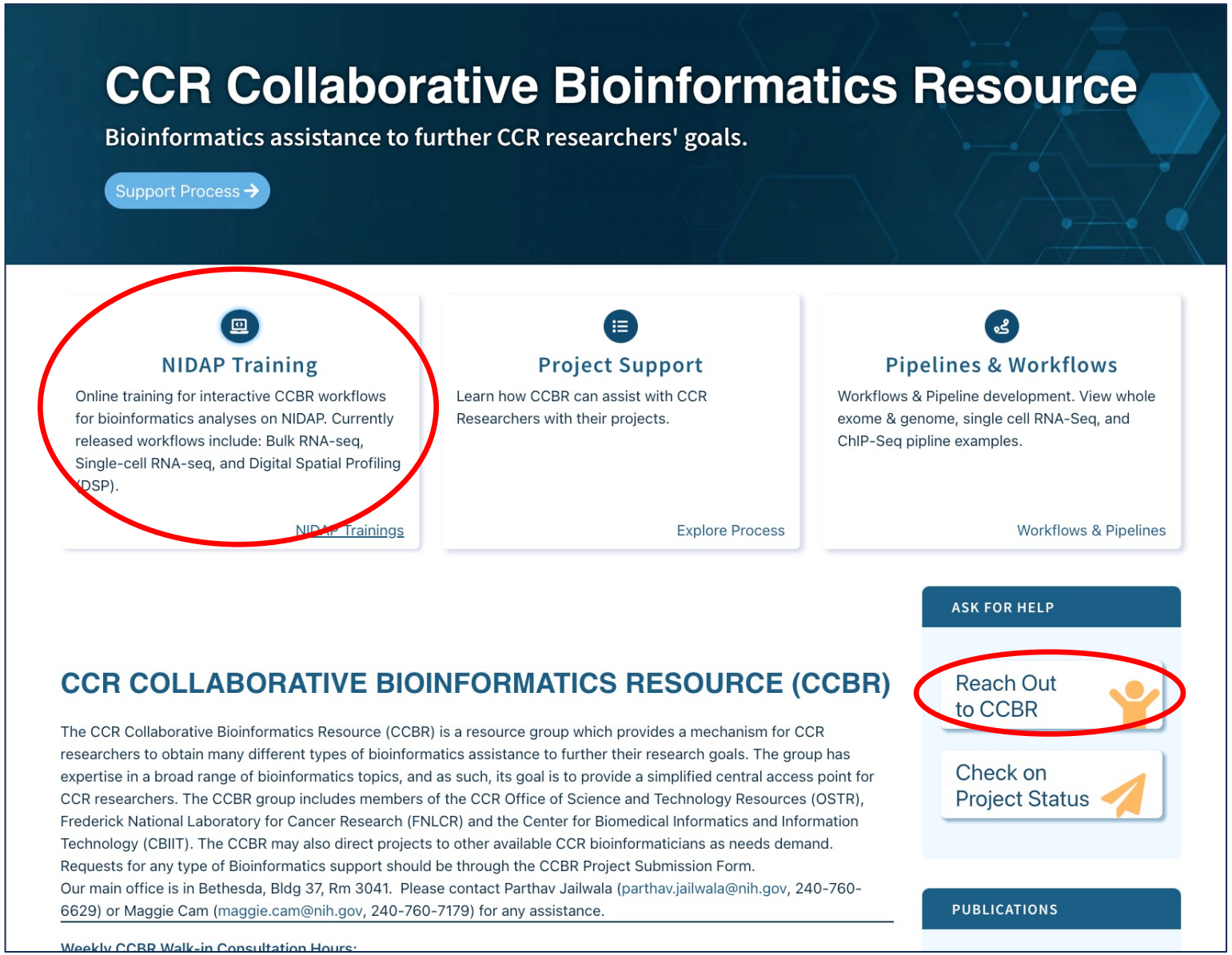
- The group of NIDAP developers include those from CCBR and several allied groups across NCI:
  - CCBR Developers:
    - Josh Meyer
    - Ned Cauley
    - Phil Homan
    - Alexei Lobanov
    - Jing Bian
    - Maggie Cam
  - Other Co-Developers and Collaborators:
    - Aleksandra Michalowski (CCR)
    - Chad Highfill (DCEG)
    - Difei Wang (DCEG)
    - Rui He (CBIIT)
    - George Zaki (CBIIT)

https://bioinformatics.ccr.cancer.gov/ccbr/

Questions?

Thank You!

NCICCBRNIDAP@mail.nih.gov



https://nidap.nih.gov/