德运康瑞
Dynamic Biosystems

DynaSpatial SD
Operational Documentation

Environment

德运康瑞
Dynamic Biosystems

We recommend using Conda to manage project dependencies for ensuring environment consistency.

1.Use Conda to create an environment and install the dependencies. We provide an environment.yml file for quick environment setup. Run the following command, and Conda will automatically create a new environment named DynamicSD and install all required packages.

```
conda env create -f environment.yml -n DynamicSD
```

2.Activate the environment using the following command

```
conda activate DynamicSD
```

3.Download DynamicSD and extract the archive to your target folder.

```
tar -zxf dynamicsd.tar.gz
```

4.Add the folder path to your system's $PATH environment variable.

```
export PATH="/path/to/DynamicSD/folder:$PATH"
```

DYNAMICSD MKREF

德远康瑞
Dynamic Biosystems

```
usage: DynamicST mkref [-h] [-F] [-n] --genome_name GENOME_NAME --fasta FASTA --gtf GTF [--threads THREADS]

optional arguments:
  -h, --help            show this help message and exit
  -F, --forceall        Force the execution of the selected (or the first) rule and all rules it is dependent on regardless of already created output.
  -n, --dryrun          Do not execute anything, and display what would be done. If you have a very large workflow, use --dryrun --quiet to just print a summary of the DAG of jobs.
  --genome_name GENOME_NAME
                        output dir, used to build genome
  --fasta FASTA         path to FASTA file containing your genome reference
  --gtf GTF             Path to genes GTF file containing annotated genes for your genome reference
  --threads THREADS     Number of threads used during STAR genome index generation. Defaults to 8
```

*The primary function of **mkref** is to construct a genomic index for use in subsequent alignment and count steps.*

DYNAMICSD COUNT

德运康瑞
Dynamic Biosystems

```
(DynamicSD) [pandunhuang@dykr DynamicSD]$ ./bin/DynamicSD count -h
usage: DynamicSD count [-h] [-F] [-n] --id ID -I INPUTDIR --whitelist-fastq WHITELIST_FASTQ [--he-image HE_IMAGE] [--probe-set PROBE_SET] [--r2-length R2_LENGTH] [-g GTF]
                       [--transcriptome TRANSCRIPTOME] [-o OUTPUTDIR] [--cellbin] [--manual-registration] [--aligend-sdata ALIGEND_SDATA] [--CBposition CBPOSITION]
                       [--UMIposition UMIPOSITION] [--cores CORES]

optional arguments:
  -h, --help            show this help message and exit
  -F, --forceall        Force the execution of the selected (or the first) rule and all rules it is dependent on regardless of already created output.
  -n, --dryrun          Do not execute anything, and display what would be done. If you have a very large workflow, use --dryrun --quiet to just print a summary of the
                        DAG of jobs.
  --id ID               Final sample name,required
  -I INPUTDIR, --inputdir INPUTDIR
                        Raw data path,required
  --whitelist-fastq WHITELIST_FASTQ
                        FastQ sequence containing whitelist,required
  --he-image HE_IMAGE   Single H&E brightfield image
  --probe-set PROBE_SET
                        CSV file specifying the probe set used, if any
  --r2-length R2_LENGTH
                        Hard trim the input Read 2 to this length before analysis;default:50
  -g GTF, --gtf GTF     genome annotation file,default:/disk/reference/WPSRanger_ref/mm10/genes.gtf
  --transcriptome TRANSCRIPTOME
                        Path of folder containing transcriptome reference,default:/disk/reference/WPSRanger_ref/mm10/star/
  -o OUTPUTDIR, --outputdir OUTPUTDIR
                        Output file path,default ./
  --cellbin             Perform Cellbin analysis
  --manual-registration
                        Perform manual registration
  --aligend-sdata ALIGEND_SDATA
                        Sdata objects manually registered by napari,default None
  --CBposition CBPOSITION
                        position of Cell Barcode(s) on the barcode read(0-base).default:0_0_0_31
  --UMIposition UMIPOSITION
                        position of the UMI on the barcode read, same as CBposition(0-base).default:0_83_0_92
  --cores CORES         set max cores the pipeline may request at one time.;default 16
```

Napari - manual registration

Manual registration is used when the effect of automatic registration is unsatisfactory.
The Registration/debug.tif image can be used to view the registration effect.

```
UmiMap/
├── he_sdata.zarr
├── raw_adata.h5ad
├── raw_umi_heatmap_grayscale.tif
├── raw_umi_heatmap.tif
└── umi_sdata.zarr
```

```
usage: SD_reg_napari.py [-h] --umi-sdata UMI_SDATA --he-sdata HE_SDATA

optional arguments:
  -h, --help            show this help message and exit
  --umi-sdata UMI_SDATA
                        umi sdata zarr file
  --he-sdata HE_SDATA   he sdata zarr file
```

Step 1: Prepare Files

Download the following files from the data analysis directory to the local machine:
UmiMap/he_sdata.zarr
UmiMap/umi_sdata.zarr

Step 2: Launch the napari Interface

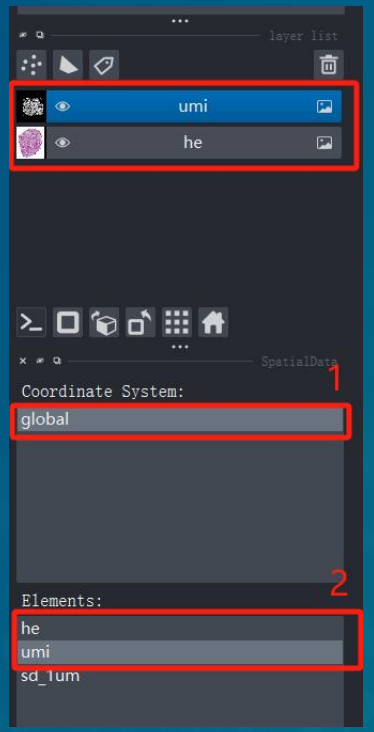
Run the SD_reg_napari.py script to open the napari interface for registration.

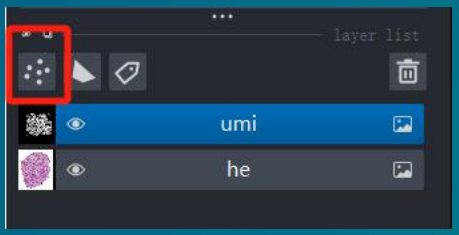
Napari - manual registration

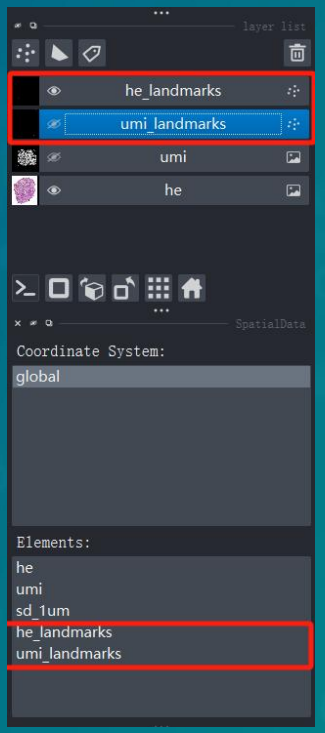
德运康瑞
Dynamic Biosystems

Step 3: Perform Registration in napari

1. **Select the coordinate system**: Choose the global coordinate system in the napari interface.
2. **Load layers**: Open the HE and umi layers respectively.
3. **Create landmark points**:
    1. Create two new points layers, named umi_landmarks and he_landmarks r espectively.
    2. Select at least 3 pairs of corresponding feature points (e.g., obvious tissue structures) and save the attributes by pressing Shift + E.

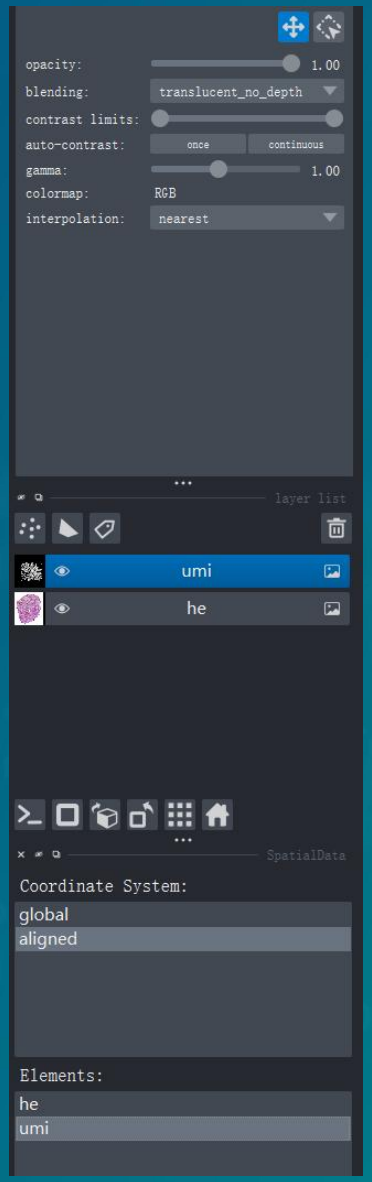






Napari - manual registration

4. **Verify registration effect**:
    1. After saving the landmarks, close the interface; napari will restart automatically.
    2. Select the aligned coordinate system, open the HE and umi layers again, and adjust the opacity to check the registration effect.
5. **Save and upload**: If the registration is correct, save the aligned_sdata.zarr file and upload it to the server.



Napari - manual registration

德运康瑞
Dynamic Biosystems

Step 4: Re-run the count Command

Specify the manual registration parameters to re-run the expression matrix analysis:
(specify the --manual-registration parameter (to determine whether to perform manual registration) and the --aligned-sdata (the aligned sdata object).)

DynamicSD count -n --id BALAJ27L --inputdir . --whitelist-fastq ../../2D250417005_D1.fq.gz --he-image ../BALAJ27L_HE.tif --probe-set /disk/pipeline/DynamicSD/db/probeV2_human.csv --transcriptome /disk/reference/WPSRanger_ref/GRCh38/star/ --cellbin --manual-registration --aligend-sdata aligned_sdata.zarr

The subsequent results are consistent with those of the automatic registration version.

OUTS

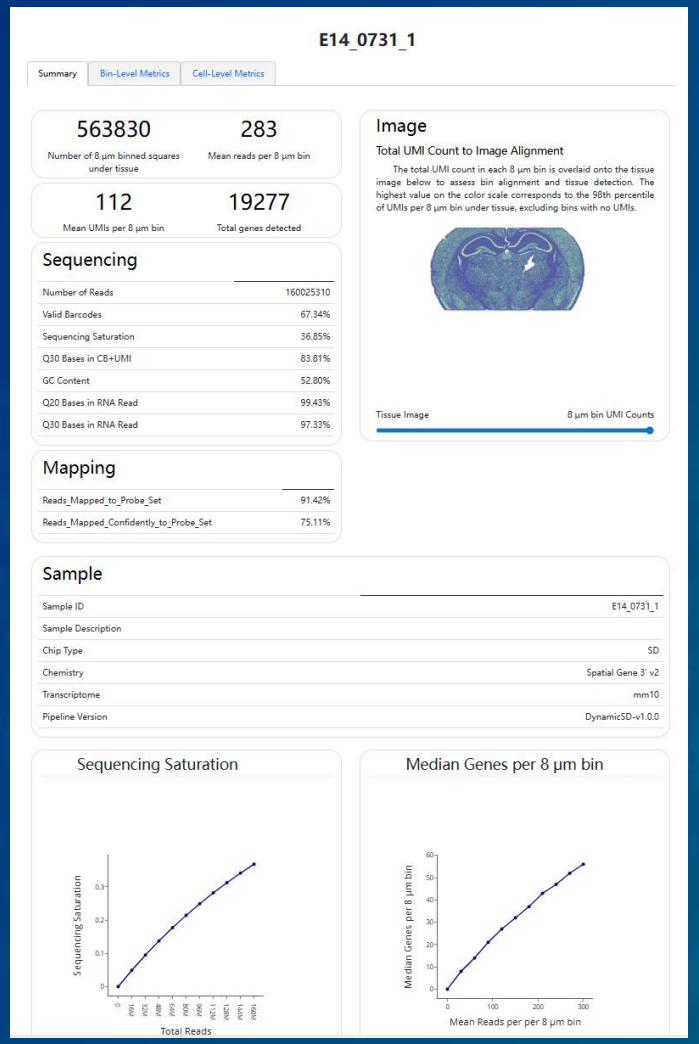
```
(DynamicSD) [pandunhuang@dykr outs]$ tree -L 2
.
├── 16um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── 50um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── 8um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── bin_summary.csv
├── CellBin
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── metrics_summary.csv
├── raw_feature_bc_matrix
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   └── matrix.mtx.gz
└── web_summary.html

13 directories, 10 files
```

File Description

- **Expression matrices for different bin sizes**: Stored in the 8um/, 16um/, and 50um/ directories, including filtered matrices (filtered_feature_bc_matrix/, filtered_feature_bc_matrix.h5) and spatial information (spatial/).
- **CellBin expression matrix**: Stored in the CellBin/ directory, used for cell-level expression analysis.
- **Raw matrix files**: Stored in raw_feature_bc_matrix/, including barcodes (barcodes.tsv.gz), gene features (features.tsv.gz), and raw count matrix (matrix.mtx.gz).
- **Statistical files**:
  - bin_summary.csv: Summary of metrics for different bin sizes.
  - metrics_summary.csv: Overall analysis metrics summary.
- **Web report**: web_summary.html – Interactive HTML

WEB SUMMARY



Sequencing

Number of Reads
Total number of read pairs that were assigned to this library in demultiplexing.
Valid Barcodes
Fraction of reads with barcodes that match the whitelist after barcode correction.
Sequencing Saturation
The fraction of reads originating from an already-observed UMI. This is a function of library complexity and sequencing depth. More specifically, this is the fraction of confidently mapped, valid Spot-barcode, valid UMI reads that had a non-unique (Spot-barcode, UMI, gene).
GC Content
GC Content.
Q20 Bases in RNA Read
Fraction of RNA read bases with Q-score >= 20, excluding very low quality/no-call (Q <= 2) bases from the denominator.
Q30 Bases in RNA Read
Fraction of RNA read bases with Q-score >= 30, excluding very low quality/no-call (Q <= 2) bases from the denominator.
Sequencing Saturation：
This plot shows the Sequencing Saturation metric as a function of downsampled sequencing depth, up to the observed sequencing depth.
Median Genes per 8 µm bin:
This plot shows the Mean Genes per 8 µm bin as a function of downsampled sequencing depth in mean reads per 8 µm bin, up to the observed sequencing depth. The slope of the curve near the endpoint can be interpreted as an upper bound to the benefit to be gained from increasing the sequencing depth beyond this point.

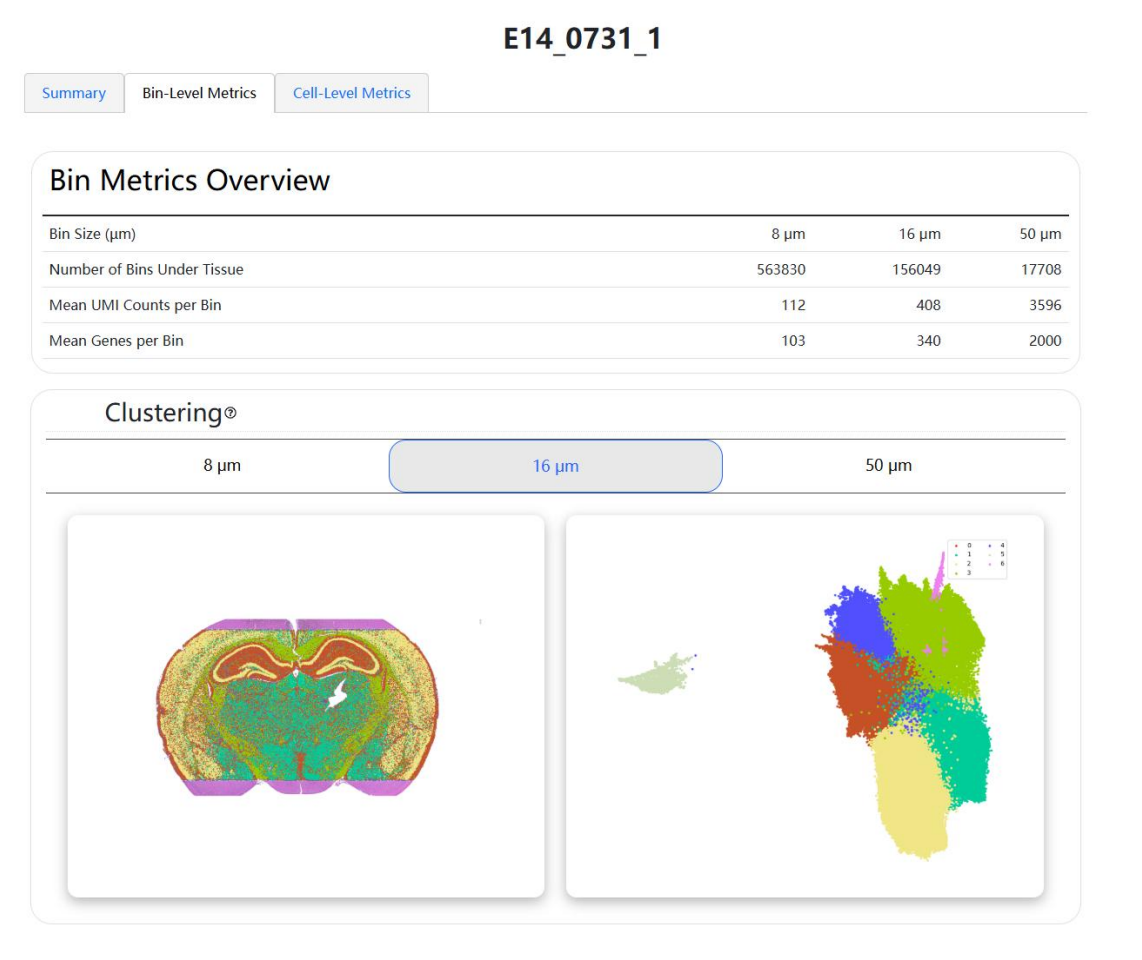
Image

The total UMI count in each 8 µm bin is overlaid onto the tissue image below to assess bin alignment and tissue detection. The highest value on the color scale corresponds to the 98th percentile of UMIs per 8 µm bin under tissue, excluding bins with no UMIs.

Mapping

Reads_Mapped_to_Probe_Set
Fraction of valid-barcode reads that mapped to the Probe Set.
Reads_Mapped_Confidently_to_Probe_Set
Fraction of valid-barcode reads that mapped uniquely to the Probe Set.

WEB SUMMARY



Bin Metrics Overview

Bin Size
The physical size of bin
Number of Bins Under Tissue
Number of bins under the Tissue.
Mean UMI Counts per Bin
Mean UMI Counts per Bin
Mean Genes per Bin
Mean Genes per Bin

Clustering

(left) Tissue plot with bins colored by Graph-based clustering.
(right) UMAP Projection of bins colored by Graph-based clustering.

WEB SUMMARY

Cell Metrics Overview

Number of Cells

The total number of cells with >= 1 unique molecular identifier (UMI).

Reads in Cells

The total number of reads assigned to cells divided by the total number of reads in the experiment.

UMIs in Cells

The percentage of UMIs under tissue within cells.

Mean Reads per Cell

The total number of reads assigned to cells divided by the number of cells.

Median Genes per Cell

Median number of genes detected per cell. Cells with zero genes detected are excluded from the calculation.

Median UMIs per Cell

Median number of unique molecular identifiers (UMIs) detected per cell. Cells with zero UMIs are excluded from the calculation.

Median Cell Area (µm²)

Each cell area is calculated by the sum area of 2 µm² squares within the segmented cell.

Cell Segmentation

Cells segmented by nucleus expansion. Cell boundaries are colored and filled by graph based clustering assignment. Black box represents the capture area. The tissue image is cropped to the capture area + 30 pixels.

Clustering

UMAP representation of segmented cells labeled with graph based clustering. Plot is sampled to 20,000 cells for visualization purposes.

德运康瑞
Dynamic Biosystems

Scan the QR code to recognize and follow the official account.

Thanks !

苏州德运康瑞生物科技有限公司

http://www.dynamic-biosystems.com

单细胞空间组学技术平台型企业
Provider

A Single Cell and Spatial Multiomics Solution