
德运康瑞
Dynamic Biosystems
空间转录组定量软件
DynamicSD

count 参数介绍

```
(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
```

--id：样本名前缀，用来再输入目录里匹配原始fastq数据
-I，--inputdir：原始fastq文件路径
--whitelist-fastq：包含白名单的fastq序列
--he-image: H&E图片
--probe-set：探针序列文件
--r2-length：在分析之前，将输入Read 2裁剪到此长度
--gtf：基因组注释文件

--transcriptome：包含转录组参考的文件夹的路径
-o，--outputdir：输出结果路径
--cellbin：是否进行cellbin分析
--manual-registration：是否进行手动umi&HE配准
--aligend-sdata：配准后的sdata对象
--CBposition：read1上细胞条形码的位置（从零开始计数）
--UMIposition：read1上UMI的位置（从零开始计数）
--cores：可以请求的最大cpu核数

输出文件

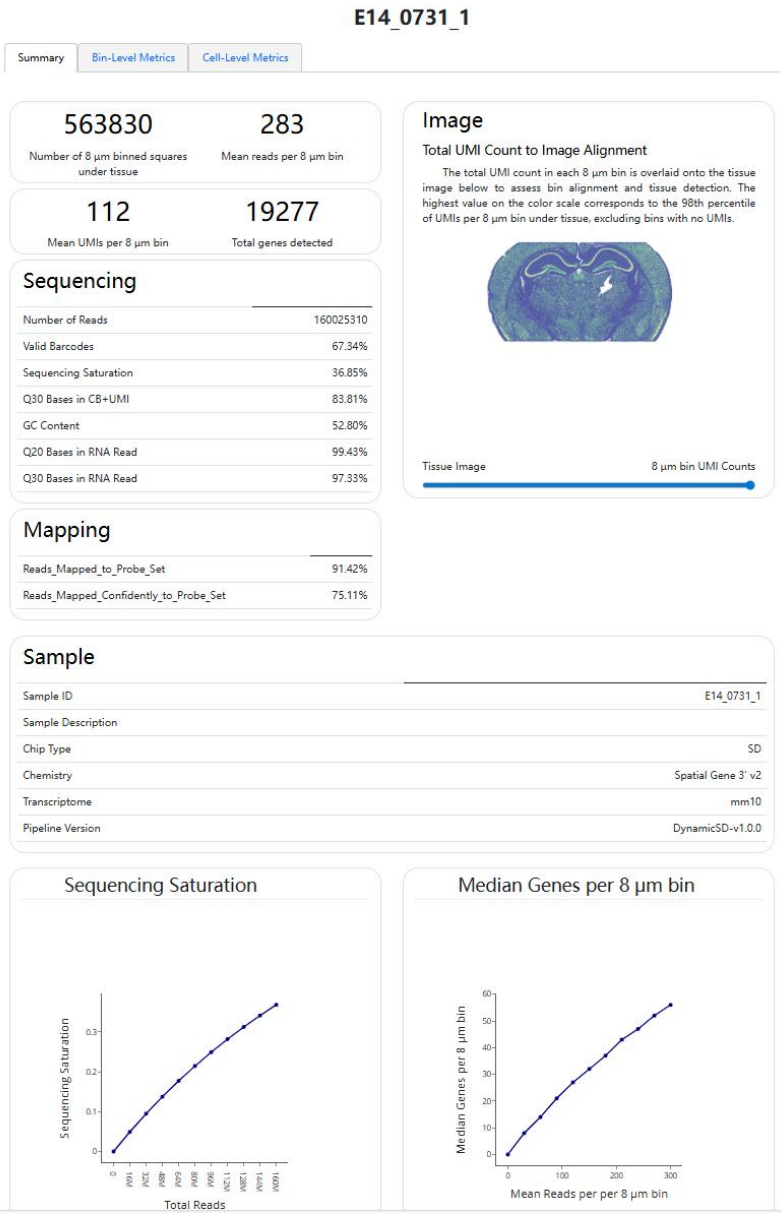
```
(DynamicSD) [pandunhuang@dykr outs]$ tree -L 2
.
├── 16um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── 50um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── spatial
├── 8um
│   ├── filtered_feature_bc_matrix
│   ├── filtered_feature_bc_matrix.h5
│   └── 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
```

不同binsize的表达矩阵&指标统计文件

cellbin的表达矩阵

芯片上所有spot的矩阵文件

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.

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.

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.

web summary

E14_0731_1

Summary | Bin-Level Metrics | Cell-Level Metrics

Bin Metrics Overview

Bin Size (μm)	8 μm	16 μm	50 μm
Number of Bins Under Tissue	563830	156049	17708
Mean UMI Counts per Bin	112	408	3596
Mean Genes per Bin	103	340	2000

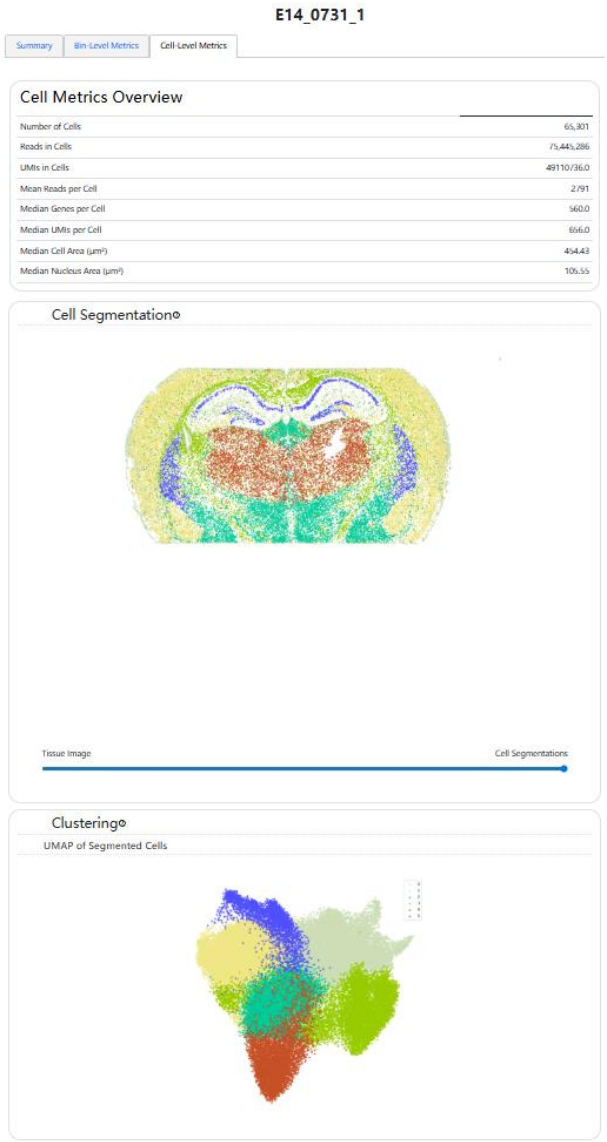
Clustering

8 μm | 16 μm | 50 μm

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.

mkref 参数介绍

```
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
```

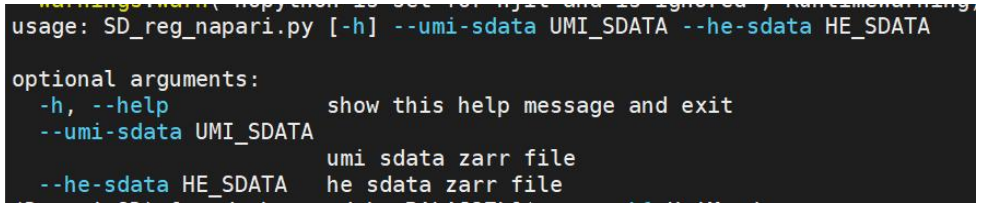
-h，--help:展示帮助信息
-F，--forceall：强制执行所选（或第一个）规则及其所依赖的所有规则，而不考虑已创建的输出。
-n，--dryrun：不要执行任何操作，并显示将要执行的操作。
--genome_name：输出参考索引的文件名
--fasta：需要建立索引的fasta文件路径
--gtf：需要建立索引的基因组注释文件路径
--threads：Star再建立索引时需要的线程数

```
GRCh38/star/
├── chrLength.txt
├── chrNameLength.txt
├── chrName.txt
├── chrStart.txt
├── exonGeTrInfo.tab
├── exonInfo.tab
├── geneInfo.tab
├── Genome
├── genomeParameters.txt
├── Log.out
├── SA
├── SAindex
├── sjdbInfo.txt
├── sjdbList.fromGTF.out.tab
├── sjdbList.out.tab
└── transcriptInfo.tab
```

索引文件展示

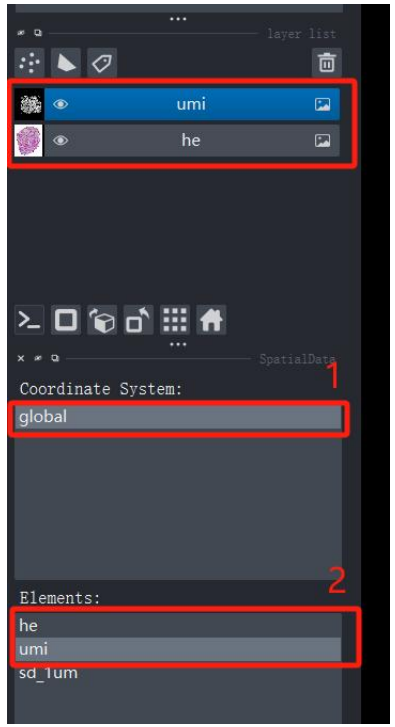
手动配准

将数据分析目录下，
UmiMap/he_sdata.zarr
UmiMap/umi_sdata.zarr
下载至本地

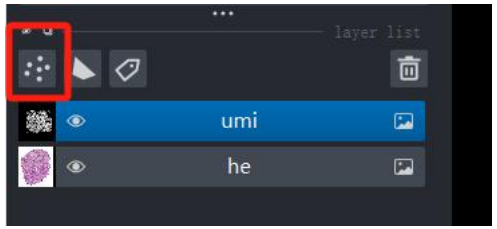


运行脚本SD_reg_napari.py脚本打开napari界面，进行配准

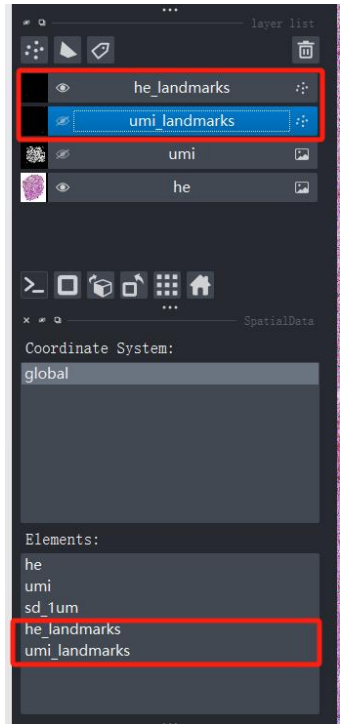
手动配准



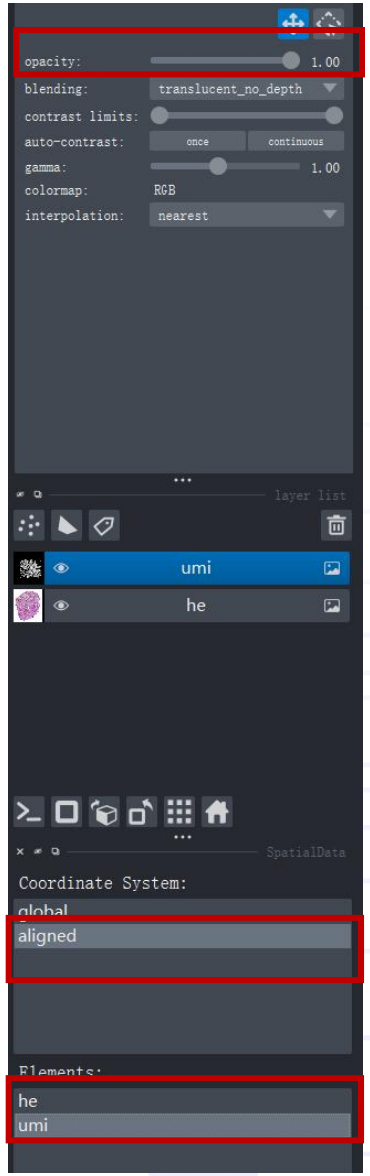
选取global坐标系
分别打开HE和umi图层



新建两个points
分别命名为
umi_landmarks
和
he_landmakrs



选取至少3对对应的特征点，同时按下Shift&E保存属性



保存特征点后关闭界面会重新启动napari，选择aligned坐标系，分别打开HE和umi图层，调节opacity查看配准效果，配准无误后保存aligned_sdata.zarr文件并上传服务器

手动配准

```
(DynamicSD) [pandunhuang@dykr BALAJ27L]$ /disk/pipeline/DynamicSD/bin/DynamicSD count -n --id BALAJ27L --inputdir ./ --wh
itelist-fastq ../../2D250417005_D1.fq.gz --he-image ../BALAJ27L_HE.tif --probe-set  /disk/pipeline/DynamicSD/db/probeV2_h
uman.csv --transcriptome /disk/reference/WPSRanger_ref/GRCh38/star/ --cellbin --manual-registration --aligend-sdata align
ed_sdata.zarr
```

需要指定--manual-registration参数（判断是否进行手动配准）以及--aligend-sdata（配准后的sdata对象）

后续结果与自动配准版一致

德运康瑞
Dynamic Biosystems

Thanks！

苏州德运康瑞生物科技有限公司

http://www.dynamic-biosystems.com