MobiVision®-M v1.3 使用手册

软件介绍

MobiVision®-M生信分析软件，可分析MobiMicrobe®高通量微生物单细胞转录组试剂盒来源的单细胞微生物转录组测序数据。 目前，MobiVision®-M v1.3共拥有三个子命令，分别为：

- quantify主要用于处理单细胞微生物转录组数据（.fastq或.fastq.gz格式），输出的数据分析报告、单微生物的基因表达矩阵（.mtx）等相关文件。
- mkindex用于构建quantify分析所需的参考基因文件。
- rcmicrobe用于对已经分析完成的quantify结果，重新call microbe，再次生成报告。

系统要求

- 8核心的Intel或AMD处理器，x86架构 (推荐16核以上)
- 32GB 内存 (建议64GB以上)
- 1TB的可用空间

- Linux操作系统，推荐64-bit CentOS 7、ubuntu:22.04或更高版本

软件安装

解压MobiVision-M_v1.3.tar.gz后，在shell命令行中执行以下source命令，即可使用MobiVision-M v1.3软件。 每次打开一个新的shell窗口或终端界面时都需要再次执行 “source” 命令。

```
###解压MobiVision-M
tar -zxvf MobiVision-M_v1.3.tar.gz
###激活MobiVision-M运行环境
source MobiVision-M_v1.3/source.sh
###测试MobiVision-M是否安装成功
mobivision-M --help
```

Quick start

构建reference

需要准备基因组序列文件，要求文件为fasta格式。以及基因注释文件，要求文件为gtf格式。
构建参考基因组的fasta文件中的contig名称必须与gtf文件中的contig名称对应。
构建参考基因组的gtf文件需满足以下要求：

1. 必须含有exon和transcript。虽然微生物一般不区分exon，但为了分析方便，可以将基因注释修改为exon+transcript的格式，两者的区域和原基因区域相同即可。
2. 除注释列外，其他列中都不能有空格。可将空格替换为”-”。
3. 基因注释中必须含有transcript_id， gene_id和gene_name。且注释在gtf中不重复。

4. 基因注释中可以含有gene_biotype。如果包含该项，则软件会统计结果中各gene biotype的占比（例如统计rRNA含量）

范例基因组序列文件： 范例基因组序列文件

范例基因注释文件： 范例基因注释文件

准备好输入文件后，运行以下代码：

```
mobivision-M mkindex \
-n speceis_name \ #species_name替换为物种名称
-f genome_file \ #genome_file替换为基因组序列文件的路径
-g gtf_file \ #gtf_file替换为基因注释文件的路径
-o reference_dir #reference_dir替换为reference输出路径
```

reference_dir整个文件夹的路径就是构建完成的reference。

基因定量分析

需要准备原始测序数据文件夹以及reference。
原始数据测序文件夹需要包含双端测序的下机文件，且已经经过了index拆分。

准备好输入文件后，运行以下代码：

```
mobivision-M quantify \
-i reference_dir \ #reference_dir替换为reference的路径
-t 12 \ #运行线程数
```

```
-f data_dir \ #data_dir替换为原始测序数据的文件夹路径
-o report_dir #report_dir替换为报告输出路径
```

report_dir中包含了数据分析报告（.html格式）以及表达矩阵（coo稀疏矩阵格式）。

子命令

---

quantify

quantify命令用于定量基因在每个单微生物中的表达量。生成质控结果，表达矩阵以及分析报告。

输入

quantify命令有两个必要输入：

1. 保存了测序下机数据的文件夹路径
   测序下机数据的文件夹中需要包含该样本的所有双端下机数据文件。 多批次的下机数据可以放在一个文件夹下，MobiVision-M会合并所有数据再进行分析。
   软件以prefix_flag_postfix.type的形式识别数据文件，其中:

- prefix为文件前缀，通常为样本名称。这部分必须存在。
- flag为双端测序的标记。目前支持R1|R2|1|2。这部分必须存在。
- postfix为文件后缀，通常标记了测序批次。这部分非必须存在。
- type为文件类型。目前支持fastq|fq|fastq.gz|fq.gz。
  MobiVision-M会将有相同prefix,postfix,type的样本视为同一批次，将不同批次的R1以及R2按顺序合并。之后再进行分析。

2. reference路径
   需要用mkindex命令输出的reference作为输入文件。

运行方法

```
mobivision-M quantify \
-f data_dir \ #data_dir替换为测序下机数据的文件夹的路径
-t 12 \ #运行线程数
-i reference \ #reference替换为reference文件夹的路径
-s ID \ #ID替换为样本ID
-o output_dir #output_dir替换为输出路径
```

可选参数

参数	默认值	描述
-f, --fastqDir	None	测序下机数据的文件夹路径。
-t, --threads	12	运行线程数。
-i, --index_dir	None	reference路径。

参数	默认值	描述
-o, --output_dir	None	输出路径。仅支持指定不存在的路径。软件不会覆写已存在的路径。
-s, --sample_ID	None	样本ID。
--cr2.2	False	指定CellRanger2.2算法作为call micorbe算法。若未指定其他call microbe算法，MobiVision-M默认使用EmptyDrop作为call microbe算法。
--cellnumber	None	使用指定细胞数的call microbe算法。软件根据指定的数目，选取UMI数量前n个barcode作为含有微生物的barcode。
--hard_filter	None	使用指定read或者UMI数最低阈值作为call microbe算法。软件根据指定数目，选取read数或UMI数高于阈值的barcode作为含有微生物的barcode。例如"min_UMI:2000"，"min_reads:5000"。
--UMI_adjust	no_adjust	指定UMI矫正算法。可选参数为"no_adjust", "step_1" and "step_1_and_2"。参数对应的算法详见软件介绍。
--multiplet_method	auto	指定判断多胞的算法。可选参数为"scaled_softmax", "majority" and "auto"。参数对应的算法详见软件介绍。
--nosecondary	False	若指定该参数，则不进行分群聚类分析。
--keep_bam	False	若指定该参数，则保留bam文件。
--keep_unmap_reads	False	若指定该参数，则保留未比对成功的reads。输出格式为fastq.gz。
--host_remove	False	若指定该参数，则进行去宿主流程。
--host_reference	None	宿主的reference路径。可以使用mkindex构建。
--qc_only	False	若指定该参数，则仅进行数据预处理，输出clean data的fastq文件、该样本的barcode whitelist以及其他。详见--qc_only参数。
--kit	v1.0	记录样本对应的试剂版本。
--test_run	False	运行小\demo数据，检验软件是否正常安装。
--config	Na	指定软件运行的config文件，其中可以设置更为详细的参数。详见config格式。
-h, --help	False	输出帮助信息。

输出

以--sample_ID为240131SW_240129B-S-SW-E01的样本为例

```
tree ./240131SW_240129B-S-SW-E01
./240131SW_240129B-S-SW-E01/
├── 240131SW_240129B-S-SW-E01
│   └── 240131SW_240129B-S-SW-E01_outs
│       ├── 240131SW_240129B-S-SW-E01.h5ad
│       ├── 240131SW_240129B-S-SW-E01.qlentille
│       ├── 240131SW_240129B-S-SW-E01_report.html
│       ├── barcode_info.tsv.gz
│       ├── filtered_cell_gene_matrix
│       │   ├── barcodes.tsv.gz
│       │   ├── features.tsv.gz
│       │   └── matrix.mtx.gz
│       ├── gene_type.json
│       ├── map_stat.tsv
│       ├── raw_cell_gene_matrix
│       │   ├── barcodes.tsv.gz
```

```
│       │   ├── features.tsv.gz
│       │   ├── matrix.mtx.gz
│       │   └── UniqueAndMult-Uniform.mtx.gz
│       ├── raw_re-assigned_cell_gene_matrix
│       │   ├── barcodes.tsv.gz
│       │   ├── features.tsv.gz
│       │   └── matrix.mtx.gz
│       └── report.json
├── Job_done.flag
├── logs
│   ├── MobiVision-M.log
│   ├── stderr.log
│   └── stdout.log
└── run_analysis_cmds.txt
```

文件/文件夹	格式	描述
240131SW_240129B-S-SW-E01_outs	文件夹	主要输出目录。
240131SW_240129B-S-SW-E01.h5ad	h5ad	记录了报告中降维聚类结果。
240131SW_240129B-S-SW-E01.qlentille	qlentille	MobiBrowser输入文件。
240131SW_240129B-S-SW-E01_report.html	html	报告文件。
barcode_info.tsv.gz	gz压缩的tsv文件	记录了每个barcode的信息。
filtered_cell_gene_matrix	文件夹	包含过滤后的矩阵文件。
raw_cell_gene_matrix	文件夹	包含未过滤且未优化多位置比对reads的矩阵文件。
raw_re-assigned_cell_gene_matrix	文件夹	包含未过滤但优化了多位置比对reads的矩阵文件。
barcodes.tsv.gz	gz压缩的tsv文件	记录了矩阵的barcode信息。
features.tsv.gz	gz压缩的tsv文件	记录了矩阵的基因信息。
matrix.mtx.gz	gz压缩的coo稀疏矩阵	记录了矩阵的数值。
gene_type.json	json	记录了基因的分类信息。
map_stat.tsv	tsv	记录了每个物种的比对信息。
report.json	json	报告文件的json格式，包含了report.html中的全部信息。
Job_done.flag	txt	该文件存在表示运行成功。
run_analysis_cmds.txt	txt	记录了运行命令。
logs	文件夹	日志目录。
MobiVision-M.log	txt	软件主要运行日志。
stdout.log	txt	调用第三方软件的标准输出（Standard Output）。
stderr.log	txt	调用第三方软件的标准错误输出（Standard Error）。

矩阵文件格式

MobiVision-M输出coo格式的系数矩阵。其具体包含了三个文件：

- barcodes.tsv.gz: gz压缩的表格文件，记录了矩阵的barcode信息
- features.tsv.gz：gz压缩的表格文件，记录了矩阵的基因信息
- matrix.mtx.gz：gz压缩的矩阵文件，记录了矩阵的具体数值

barcodes.tsv.gz格式

将barcodes.tsv.gz解压后，其文件内容为以下格式：

AAACAGCAATAGGTCTCTGT
AAACAGCAATCGGACCAACA
AAACAGCAATGCCACGGTCT
AAACAGCAATGGACGGTATA
AAACAGCAATGTTGTTCGTT
AAACAGCAATTCATAAGTCG
AAACATTATGAACAGTCAAG
AAACATTATGAATAGATGAA
AAACATTATGCCAGGACGGC
AAACATTATGTACGGCACTA
每一行为一个barcode

features.tsv.gz格式

将features.tsv.gz解压后，其文件内容为以下格式：

gene-STM14_0002	thrA	Gene Expression
gene-STM14_0003	thrB	Gene Expression
gene-STM14_0004	thrC	Gene Expression
gene-STM14_0005	yaaA	Gene Expression
gene-STM14_0006	yaaJ	Gene Expression
gene-STM14_0007	talB	Gene Expression
gene-STM14_0008	mogA	Gene Expression
gene-STM14_0009	yaaH	Gene Expression
gene-STM14_0010	htgA	Gene Expression
gene-STM14_0011	yaaI	Gene Expression
gene-STM14_0012	STM14_0012	Gene Expression
gene-STM14_0013	dnaK	Gene Expression

gene-STM14_0014	dnaJ	Gene Expression
gene-STM14_0015	STM14_0015	Gene Expression

其中：

- 第一列为基因ID
- 第二列为基因名称
- 第三列为类型

matrix.mtx.gz格式

将matrix.mtx.gz解压后，其文件内容为以下格式：

%%MatrixMarket matrix coordinate integer general		
%		
5521	5985	1803635
59	1	1
160	1	1
186	1	1
198	1	1
203	1	1
258	1	1
259	1	2
266	1	1
272	1	1
359	1	1
360	1	1
369	1	1
392	1	1
459	1	1
523	1	1
556	1	1
627	1	1
655	1	1
693	1	1

其中：

- 第一行固定为：%%MatrixMarket matrix coordinate integer general
- 第二行固定为：%
- 第三行描述了整个矩阵的大小。第一个值为行数，也是基因数，与features.tsv.gz行数一致。第二个值为列数，也是barcode数，与barcodes.tsv.gz行数一致。第三个值为矩阵非0值个数。
- 接下来每一行描述了一个非0值。第一个值为行坐标，第二个值为列坐标，第三个值为该坐标的数值，及检测到的UMI数。

---

config格式

通过config可以设置更加详细的运行参数。config文件不是必须的，config内的每个参数也不是必须的。对于未指定的参数，MobiVision-M会读取内部的默认值进行分析。
config分为两个section，STAR和MobiVision-M。
STAR section均为STAR aligner软件的参数。
范例config文件: 范例config文件
config文件要求为ini格式，具体可选参数如下：

参数	section	默认值	描述
soloMultiMappers	STAR	Uniform	指定STAR的soloMultiMappers参数。
nExpectedCells	STAR	3000	指定STAR的soloCellFilter参数，EmptyDrops_CR或CellRanger2.2的第1个子参数。
maxPercentile	STAR	0.99	指定STAR的soloCellFilter参数，EmptyDrops_CR或CellRanger2.2的第2个子参数。
maxMinRatio	STAR	10	指定STAR的soloCellFilter参数，EmptyDrops_CR或CellRanger2.2的第3个子参数。
indMin	STAR	45000	指定STAR的soloCellFilter参数，EmptyDrops_CR的第4个子参数。
indMax	STAR	90000	指定STAR的soloCellFilter参数，EmptyDrops_CR的第5个子参数。
umiMin	STAR	150	指定STAR的soloCellFilter参数，EmptyDrops_CR的第6个子参数。
umiMinFracMedian	STAR	0.1	指定STAR的soloCellFilter参数，EmptyDrops_CR的第7个子参数。
candMaxN	STAR	20000	指定STAR的soloCellFilter参数，EmptyDrops_CR的第8个子参数。
FDR	STAR	0.1	指定STAR的soloCellFilter参数，EmptyDrops_CR的第9个子参数。
simN	STAR	10000	指定STAR的soloCellFilter参数，EmptyDrops_CR的第10个子参数。
soloUMIfiltering	STAR	-	指定STAR的soloUMIfiltering参数。
soloUMIdedup	STAR	1MM_CR	指定STAR的soloUMIdedup参数。
allow_multi_target_UMI	STAR	True	是否运行一个UMI比对到多个位置。
reclaim_UMI	STAR	True	是否将比对到次要位置的UMI纳入分析。
white_list_file	STAR	NA	指定使用的whitelist。
process_cutadapt	MobiVision-M	True	是否在预处理时运行cutadapt。
process_fastp	MobiVision-M	True	是否在预处理时运行fastp。
adpator_list_path	MobiVision-M	NA	自定义接头文件路径。

---

qc-only参数输出

如果指定了--qc_only，则MobiVision®-M会仅运行预处理流程，并保留全部预处理结果。 以输出目录为240321B-S-LLY-E10为例：

```
tree ./240321B-S-LLY-E10
./240321B-S-LLY-E10/
├── Job_done.flag
├── logs
│   ├── MobiVision-M.log
│   ├── stderr.log
│   └── stdout.log
├── pre_process
│   ├── 240321B-S-LLY-E10_filter_stat.tsv
│   ├── 240321B-S-LLY-E10_qc_stat.tsv
│   ├── 240321B-S-LLY-E10_sample_barcode_stat.tsv
│   ├── clean_data
│   │   └── 240321B-S-LLY-E10
│   │       ├── 240321B-S-LLY-E10_240321B-S-LLY-E10_S0_L001_R1_001.fastq.gz
│   │       ├── 240321B-S-LLY-E10_240321B-S-LLY-E10_S0_L001_R2_001.fastq.gz
│   │       └── white_list.tsv
│   ├── failed_reads.tsv
│   ├── Job_done.flag
│   ├── pre-process
│   │   └── 240321B-S-LLY-E10
│   │       ├── cutadapt_out
│   │       │   ├── 240321B-S-LLY-E10_combined_clean1_R1.fastq.gz
│   │       │   ├── 240321B-S-LLY-E10_combined_clean1_R2.fastq.gz
│   │       │   └── cutadapt_report.json
│   │       ├── fastp_out
│   │       │   ├── 240321B-S-LLY-E10_fastp-report.html
│   │       │   └── 240321B-S-LLY-E10_fastp-report.json
│   │       └── qc_data.tsv
│   ├── split_fastq
│   │   ├── 240321B-S-LLY-E10
│   │   │   ├── 240321B-S-LLY-E10_240321B-S-LLY-E10_S1_L001_R1_001.fastq.gz
│   │   │   ├── 240321B-S-LLY-E10_240321B-S-LLY-E10_S1_L001_R2_001.fastq.gz
│   │   │   └── white_list.tsv
│   │   └── unknown
│   │       ├── 240321B-S-LLY-E10_unknown_S1_L001_R1_001.fastq.gz
│   │       └── 240321B-S-LLY-E10_unknown_S1_L001_R2_001.fastq.gz
│   └── sub_process_annnote.tsv
└── run_analysis_cmds.txt
```

文件/文件夹	格式	描述
Job_done.flag	txt	该文件存在表示运行成功。
run_analysis_cmds.txt	txt	记录了运行命令。
logs	文件夹	日志目录。
MobiVision-M.log	txt	软件主要运行日志。
stdout.log	txt	调用第三方软件的标准输出（Standard Output）。

文件/文件夹	格式	描述
stderr.log	txt	调用第三方软件的标准错误输出（Standard Error）。
pre_process	文件夹	预处理结果目录。
240321B-S-LLY-E10_filter_stat.tsv	tsv	reads通过barcode检测的统计结果。
240321B-S-LLY-E10_qc_stat.tsv	tsv	reads通过cutadapt、fastp或去宿主的统计结果。
240321B-S-LLY-E10_sample_barcode_stat.tsv	tsv	样本结果文件记录。
failed_reads.tsv	tsv	未通过barcode检测的reads name。
sub_process_annote.tsv	tsv	运行各流程的时间记录。
split_fastq	文件夹	barcode检测输出文件夹。
240321B-S-LLY-E10	文件夹	含有正确barcode和UMI的reads的文件夹。
240321B-S-LLY-E10_240321B-S-LLY-E10_S1_L001_R1_001.fastq.gz	gzip压缩的fastq文件	仅含有reads的barcode+UMI序列。
240321B-S-LLY-E10_240321B-S-LLY-E10_S1_L001_R2_001.fastq.gz	gzip压缩的fastq文件	含有raw_data R2端的序列，未做处理。
white_list.tsv	tsv	该样本的barcode列表。
unknown	文件夹	含有不正确的barcode或者UMI的reads的文件夹。
240321B-S-LLY-E10_unknown_S1_L001_R1_001.fastq.gz	gzip压缩的fastq文件	未通过barcode或UMI检测的reads的R1端序列。
240321B-S-LLY-E10_unknown_S1_L002_R1_001.fastq.gz	gzip压缩的fastq文件	未通过barcode或UMI检测的reads的R2端序列。
pre-process	文件夹	调用第三方软件的处理结果文件夹。
cutadapt_out	文件夹	cutadapt输出结果文件夹。
fastp_out	文件夹	fastp输出结果文件夹。
qc_data.tsv	tsv	第三方软件的数据处理结果统计。

mkindex

mkindex命令用于构建MobiVision-M quantify命令运行需要的refernce文件。

输入

为了构建MobiVision-M quantify需要的reference，每个物种需要两个文件：基因组序列文件以及基因注释文件。
基因序列文件要求为fasta格式。
基因注释文件要求为gtf格式，并满足以下要求：

1. 必须含有exon和transcript。虽然微生物一般不区分exon，但为了分析方便，可以将基因注释修改为exon+transcript的格式，两者的区域和原基因区域相同即可。
2. 除注释列外，其他列中都不能有空格。
3. 基因注释中必须含有transcript_id， gene_id和gene_name。且每个基因的这三个注释的值在gtf中唯一。

4. 基因注释中可以含有gene_biotype。如果包含该项，则软件会统计结果中各gene biotype的占比（例如统计rRNA含量）

运行方法

构建单物种reference

```
mobivision-M mkindex \
-n speceis_name \ #species_name替换为reference名称
-f genome_file \ #genome_file替换为基因组序列文件的路径
-g gtf_file \ #gtf_file替换为基因注释文件的路径
-o reference_dir #reference_dir替换为reference输出路径
```

构建多物种reference
可以通过多次输入-n、-f、-g，输入多物种的信息。 同一物种的参数输入顺序必须一致。

```
mobivision-M mkindex \
-n speceis_name1 \ #species_name1替换为物种1的名称
-f genome_file1 \ #genome_file1替换为物种1的基因组序列文件的路径
-g gtf_file1 \ #gtf_file1替换为物种1的基因注释文件的路径
-n speceis_name2 \ #species_name2替换为物种2的名称
-f genome_file2 \ #genome_file2替换为物种2的基因组序列文件的路径
-g gtf_file2 \ #gtf_file2替换为物种2的基因注释文件的路径
-o reference_dir #reference_dir替换为reference输出路径
```

也可以将多个物种的基因序列文件和gtf文件做成表格，通过--input_file输入
表格需要name、gtf和fasta散列，并以制表符分隔。例如：

name	gtf	fasta
E.coi	/share/home/sc/Projects/microbeRNA-seq/reference_20230808/GCF_000005845.2_modified.gtf	/share/home/sc/Projects/microbeRNA-seq/reference_20230808/GCF_000005845.2_ASM584v2_genomic.fna
B.sub	/share/home/sc/Projects/microbeRNA-seq/reference_20230808/GCF_000009045.1_modified.gtf	/share/home/sc/Projects/microbeRNA-seq/reference_20230808/GCF_000009045.1_ASM904v1_genomic.fna

```
mobivision-M mkindex \
--input_file genomes_metadata.tsv \
-o combine_ref
```

可选参数

参数	默认值	描述
-n, --nameOfSpecies	None	物种名称。可以重复输入以对应多物种，但顺序必须与-f和-g一致。
-f, --fastaPath	None	物种的基因组序列文件。要求文件为fasta格式。可以重复输入以对应多物种，但顺序必须与-n和-g一致。
-g, --gtfPath	None	物种的基因注释文件。要求文件为gtf格式。可以重复输入以对应多物种，但顺序必须与-n和-f一致。
--inut_file	NA	使用制表符分隔的物种信息表格构建reference。默认值为NA，即不使用本参数的输入。
-r, --referenceVerString	unknonw	指定reference的版本
-m, --memoryUsed	64	限制构建STAR reference时的最大内存。单位为GB。
-o, --output_dir	None	指定reference的输出路径。
--test_run	False	运行小demo数据，检验软件是否正常安装。
-h, --help	False	输出帮助信息。

输出

以输出路径为E.coil_and_B.sub为例。

```
tree ./E.coil_and_B.sub
./E.coil_and_B.sub
├── fasta
│   ├── genome.fa
│   └── genome.fa.fai
├── genes
│   ├── gene_info.json
│   └── genes.gtf
├── Job_done.flag
├── logs
│   ├── MobiVision-M.log
│   ├── stderr.log
│   └── stdout.log
```

```
├── reference.json
└── 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
```

文件/文件夹	格式	描述
fasta	文件夹	保存构建完成的序列信息。
genome.fa	fasta	基因组序列文件。
genome.fa.fai	fai	基因组序列的index文件。
genes	文件夹	保存构建完成的基因信息。
gene_info.json	json	记录基因分类信息。
genes.gtf	gtf	记录基因的注释信息。
Job_done.flag	txt	该文件存在表示运行成功。
logs	文件夹	日志目录。
MobiVision-M.log	txt	软件主要运行日志 。
stdout.log	txt	调用第三方软件的标准输出（Standard Output）。
stderr.log	txt	调用第三方软件的标准错误输出（Standard Error）。
star	文件夹	STAR reference 目录。

---

rcmicrobe

rcmicrobe用于从已运行完成的quantify结果中，使用不同的call microbe 方法，得到新的分析结果。

输入

rcmicrobe需要输入已运行完成的quantify分析结果。
该输入为一个文件夹，包含了report.json，barcode_info.tsc.gz等文件。一般以_outs结尾。

运行方法

```
mobivision-M rcmicrobe \
-i analysis_dir \ #替换为已分析完成的MobiVision-M结果路径，一般以_outs结尾
-o output_dir \ #替换为输出路径
-t 8 \ #运行线程数
--cr2.2 \ #指定call microbe算法
```

可选参数

参数	默认值	描述
-i, --analysis_dir	None	已分析完成的MobiVision-M结果路径。一般以_outs结尾。
-o, --output_dir	None	输出路径。
-c, --call_mtx	None	基于哪个矩阵文件重新分析。如果该参数未指定，则会依次检查raw_re-assigned_cell_gene_matrix和raw_cell_gene_matrix。
-t, --threads	12	运行线程数。
-s, --sample_ID	None	报告中的样本名称。如果该参数未指定，则会沿用analysis_dir的样本名称。
--cr2.2	False	指定CellRanger2.2算法作为call micorbe算法。若未指定其他call microbe算法，MobiVision默认使用EmptyDrop作为call microbe算法。
--cellnumber	None	使用指定细胞数的call microbe算法。软件根据指定的数目，选取UMI数量前n个barcode作为含有微生物的barcode。
--hard_filter	None	使用指定read或者UMI数最低阈值作为call microbe算法。软件根据指定数目，选取read数或UMI数高于阈值的barcode作为含有微生物的barcode。例如“min_UMI:2000”，”min_reads:5000"。
--UMI_adjust	no_adjust	指定UMI矫正算法。可选参数为“no_adjust”, “step_1” and “step_1_and_2”。参数对应的算法详见软件介绍。
--multiplet_method	auto	指定判断多胞的算法。可选参数为"scaled_softmax”, “majority” and “auto”。参数对应的算法详见软件介绍。
--nosecondary	False	若指定该参数，则不进行分群聚类分析。
--keep_bam	False	若指定该参数，则保留bam文件。
--kit	v1.0	记录样本对应的试剂版本。
-h, --help	False	输出帮助信息。

输出

以输出路径为demo为例

```
tree ./demo
./demo/
├── barcode_info.tsv.gz
├── demo.h5ad
├── demo.qlentille
├── demo_report.html
├── filtered_cell_gene_matrix
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
```

```
│   └── matrix.mtx.gz
├── gene_type.json
├── high_variable_genes.tsv
├── Job_done.flag
├── logs
│   ├── MobiVision-M.log
│   ├── stderr.log
│   └── stdout.log
├── map_stat.tsv
├── raw_cell_gene_matrix
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   ├── matrix.mtx.gz
│   └── UniqueAndMult-Uniform.mtx
├── raw_re-assigned_cell_gene_matrix
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   └── matrix.mtx.gz
└── report.json
```

文件/文件夹	格式	描述
demo.h5ad	h5ad	记录了报告中降维聚类结果。
demo.qlentille	qlentille	MobiBrowser输入文件。
demo_report.html	html	报告文件。
barcode_info.tsv.gz	gz压缩的tsv文件	记录了每个barcode的信息。
filtered_cell_gene_matrix	文件夹	包含过滤后的矩阵文件。
raw_cell_gene_matrix	文件夹	包含未过滤且未优化多位置比对reads的矩阵文件。
raw_re-assigned_cell_gene_matrix	文件夹	包含未过滤但优化了多位置比对reads的矩阵文件。
barcodes.tsv.gz	gz压缩的tsv文件	记录了矩阵的barcode信息。
features.tsv.gz	gz压缩的tsv文件	记录了矩阵的基因信息。
matrix.mtx.gz	gz压缩的coo稀疏矩阵	记录了矩阵的数值。
gene_type.json	json	记录了基因的分类信息。
map_stat.tsv	tsv	记录了每个物种的比对信息。
high_variable_genes.tsv	tsv	报告中降维聚类结果使用的高变基因。
report.json	json	报告文件的json格式，包含了report.html中的全部信息。
Job_done.flag	txt	该文件存在表示运行成功。
run_analysis_cmds.txt	txt	记录了运行命令。
logs	文件夹	日志目录。
MobiVision-M.log	txt	软件主要运行日志。

文件/文件夹	格式	描述
stdout.log	txt	调用第三方软件的标准输出（Standard Output）。
stderr.log	txt	调用第三方软件的标准错误输出（Standard Error）。