

# TotalSeqB\_CELLRANGER

Our data was sequenced using TotalSeqB chemistry and therefore cell surface antibodies were used to multiplexed data. In our case 4 HTOs (antibody barcodes) were used for multiplexing samples in a batch of three. Totally 30 samples were sequenced in 10 lines using the combination of these 4 HTOs.

The data we received came in the following format:

```
[maider@login FASTQ]$ ll
total 6770
drwxr-xr-x 2 maider binf 8.0K Apr 21 14:34 .
drwxr-xr-x 4 maider binf 39 Apr 15 06:47 ..
-rwxr-xr-x 1 maider binf 6.2K Apr 15 06:47 md5sum.txt
-rwxr-xr-x 1 maider binf 568M Apr 14 16:58 MMC-125-135-144-C_Ab_S3_L001_I1_001.fastq.gz
-rwxr-xr-x 1 maider binf 651M Apr 14 16:59 MMC-125-135-144-C_Ab_S3_L001_I2_001.fastq.gz
-rwxr-xr-x 1 maider binf 1.9G Apr 14 17:02 MMC-125-135-144-C_Ab_S3_L001_R1_001.fastq.gz
-rwxr-xr-x 1 maider binf 3.3G Apr 14 17:07 MMC-125-135-144-C_Ab_S3_L001_R2_001.fastq.gz
-rwxr-xr-x 1 maider binf 3.9G Apr 14 17:13 MMC-125-135-144-C_GEX_S3_L003_I1_001.fastq.gz
-rwxr-xr-x 1 maider binf 4.3G Apr 14 17:19 MMC-125-135-144-C_GEX_S3_L003_I2_001.fastq.gz
-rwxr-xr-x 1 maider binf 16G Apr 14 17:42 MMC-125-135-144-C_GEX_S3_L003_R1_001.fastq.gz
-rwxr-xr-x 1 maider binf 34G Apr 14 18:33 MMC-125-135-144-C_GEX_S3_L003_R2_001.fastq.gz
-rwxr-xr-x 1 maider binf 568M Apr 14 18:34 MMC-150-160-166-D_Ab_S4_L002_I1_001.fastq.gz
-rwxr-xr-x 1 maider binf 690M Apr 14 18:35 MMC-150-160-166-D_Ab_S4_L002_I2_001.fastq.gz
-rwxr-xr-x 1 maider binf 1.9G Apr 14 18:38 MMC-150-160-166-D_Ab_S4_L002_R1_001.fastq.gz
-rwxr-xr-x 1 maider binf 3.4G Apr 14 18:43 MMC-150-160-166-D_Ab_S4_L002_R2_001.fastq.gz
-rwxr-xr-x 1 maider binf 3.7G Apr 14 18:49 MMC-150-160-166-D_GEX_S4_L003_I1_001.fastq.gz
-rwxr-xr-x 1 maider binf 4.3G Apr 14 18:55 MMC-150-160-166-D_GEX_S4_L003_I2_001.fastq.gz
-rwxr-xr-x 1 maider binf 15G Apr 14 19:17 MMC-150-160-166-D_GEX_S4_L003_R1_001.fastq.gz
-rwxr-xr-x 1 maider binf 32G Apr 14 20:06 MMC-150-160-166-D_GEX_S4_L003_R2_001.fastq.gz
-rwvvr-vvvv 1 maider binf 901M Apr 14 20:07 MMC-175-185-191-L_AH_SR_I100_T1_001.fastq.gz
```

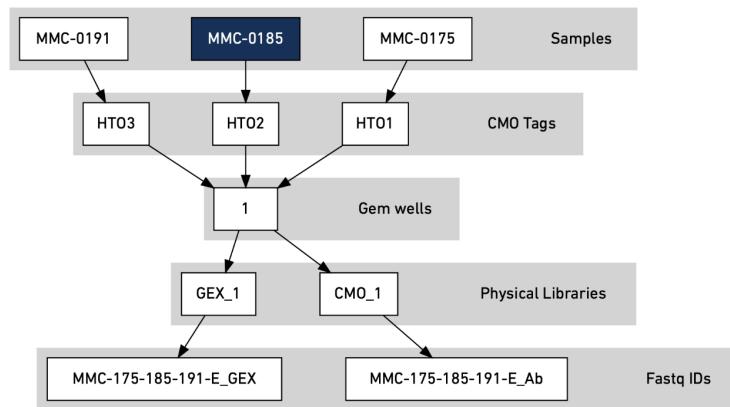
We received 8 files per sample. Four GEX files with expression data and 4 Abs files with antibody information.

Furthermore, we received the following Excel sheet with HTO information to demultiplex the data. Meaning for Sample MMC-0079 HTO1 was used. Which is the first one in Name column.

A	B	C	D	E	F	G	H	I
Sample ID	HTO_TotalseqB			Name	Catalogue	Barcode Sequence		
1	TotalSeq™-B02	394631	GTCAACTTTAGCG					
2	TotalSeq™-B02	394633	TGATGGCCTATTGGG					
3	MMC-0079	1						
4	TotalSeq™-B02	394635	TTCGGCCTCTCTTG					
5	MMC-0089	2						
6	TotalSeq™-B02	394637	AGTAAGTTCAAGCGTA					
7	MMC-0093	3						
8								
9								
10								
11	MMC-0097	4						
12								
13	MMC-0107	1						
14								
15	MMC-0117	2						
16								
17								
18								
19	MMC-0125	3						
20								
21	MMC-0135	4						
22								
23	MMC-0144	1						
24								
25								
26								
27	MMC-0150	2						
28								
29	MMC-0160	3						
30								
31	MMC-0166	4						
32								
33								
34	MMC-0175	1						
35								
36	MMC-0185	2						
37								
38	MMC-0191	3						
39								
40								
41								
42								
43	MMC-0208	4						
44								
45	MMC-0218	1						
46								
47	MMC-0227	2						
48								
49								
50	MMC-0239	3						
51								
52	MMC-0254	4						
53								
54	MMC-0262	1						
55								
56								
57	MMC-0269	2						
58								
59	MMC-0276	3						
60								
61	MMC-0287	4						
62								
63								
64								
65	MMC-0295	1						
66								
67	MMC-0301	2						
68								
69	MMC-0310	3						
70								
71	MMC-0317	4						
72								
73	MMC-0325	1						
74								
75	MMC-0334	2						
76								

The experimental design was as follows:

## Experimental Design ②



We learned there are two different ways of demultiplexing/analyzing this data.

### First option: Use cellranger multi command

```
cellranger-multi
Analyze multiplexed data or combined gene expression/immune profiling/feature
barcode data

USAGE:
    cellranger multi [FLAGS] [OPTIONS] --id --csv

FLAGS:
    --dry          Do not execute the pipeline. Generate a pipeline
                  invocation (.mro) file and stop
    --disable-ui   Do not serve the web UI
    --noexit       Keep web UI running after pipeinstance completes or fails
    --nopreflight  Skip preflight checks
    -h, --help     Prints help information

OPTIONS:
    --id           A unique run id and output folder name [a-zA-Z0-
                  9_-]+
    --description  Sample description to embed in output files
                  [default: ]
    --csv          Path of CSV file enumerating input libraries and
                  analysis parameters
    ...
```

This will be run through the [TotalSeq\\_cellranger.py script](#).

```
[maider@login scripts]$ python cellranger_multi.py --help
usage: This script runs cellranger program from 10X genomics to map the single cell reads and generate the count
matrices. Exiting.....

Process command line arguments.

optional arguments:
  -h, --help            show this help message and exit
  -f FILE, --file_list FILE
                        fastq file list
  -c FILE, --config FILE
                        config file
```

To run this we need in our working directory the following CMO\_reference, R1, R2, I1 and I2 files for each sample and a config file for each sample.

A CMO\_reference is analogous to the Samplesheet.csv used in standard Illumina Sequencing and contains the barcode information for demultiplexing. This file works in conjunction with a configuration file (see the screenshot below) which contains the mapping between the sample and the **barcodes (HTO1, HTO2, HTO3, HTO4)** along with the reference genome and the library information. The HTO information was received from the sequencing company.

```
[komalj@login April19_2022]$ cat CMO_reference.csv
id,name,read,pattern,sequence,feature_type
HTO1,HTO1,R2,5PNNNNNNNNNN(BC),GTCAACTCTTAGCG,Multiplexing Capture
HTO2,HTO2,R2,5PNNNNNNNNNN(BC),TGATGGCCTATTGGG,Multiplexing Capture
HTO3,HTO3,R2,5PNNNNNNNNNN(BC),TTCCGCCTCTCTTG,Multiplexing Capture
HTO4,HTO4,R2,5PNNNNNNNNNN(BC),AGTAAGTTCAGCGTA,Multiplexing Capture
```

As well as the CMO\_reference we need a csv config file with information regarding the location of reference transcriptome location, CMO\_reference locations, I1,I2,R1 and R2 files location and finally the link between HTO (barcodes) and samples. One example of config file is the following. We need to create these config files for each fastq ( containing 3 samples). The sample ids were generated by the project coordinator but can be any name.

```
[gene-expression]
reference,/share/data/RNA_Seq/reference_genomes/10X_genomics/human/refdata-cellranger-GRCh38-3.0.0/
cmo-set,/share/data/RNA_Seq/MECFS_scRNAseq/Working/April19_2022/CMO_reference.csv

[libraries]
fastq_id,fastqs,feature_types
MMC-208-218-227-F_GEX,/share/data/RNA_Seq/MECFS_scRNAseq/Working/April19_2022,Gene Expression
MMC-208-218-227-F_Ab,/share/data/RNA_Seq/MECFS_scRNAseq/Working/April19_2022,Multiplexing Capture #1

[samples]
sample_id,cmo_ids
MMC-0208,HT04
MMC-0218,HT01
MMC-0227,HT02
```

Example can be found on:

```
/share/data/RNA_Seq/MECFS_scRNAseq/Working/April19_2022/Cellranger
```

```

311 Apr 20 12:54 CMO_reference.csv
4.0K Apr 21 14:33 MMC-125-135-144-C
82 Apr 20 13:16 MMC-125-135-144-C_Ab_S3_L001_I1_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_Ab_S3_L001_I1_001.fastq.gz
82 Apr 20 13:16 MMC-125-135-144-C_Ab_S3_L001_I2_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_Ab_S3_L001_I2_001.fastq.gz
82 Apr 20 13:16 MMC-125-135-144-C_Ab_S3_L001_R1_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_Ab_S3_L001_R1_001.fastq.gz
82 Apr 20 13:16 MMC-125-135-144-C_Ab_S3_L001_R2_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_Ab_S3_L001_R2_001.fastq.gz
12K Apr 21 15:54 MMC-125-135-144-C_cellranger.o1384223
0 Apr 21 14:33 MMC-125-135-144-C_cellranger.po1384223
527 Apr 20 13:19 MMC-125-135-144-C.config
83 Apr 20 13:16 MMC-125-135-144-C_GEX_S3_L003_I1_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_GEX_S3_L003_I1_001.fastq.gz
83 Apr 20 13:16 MMC-125-135-144-C_GEX_S3_L003_I2_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_GEX_S3_L003_I2_001.fastq.gz
83 Apr 20 13:16 MMC-125-135-144-C_GEX_S3_L003_R1_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_GEX_S3_L003_R1_001.fastq.gz
83 Apr 20 13:16 MMC-125-135-144-C_GEX_S3_L003_R2_001.fastq.gz -> ../../RAW_DATA/singlecell_MECSFASTQ/MMC-125-135-144-C_GEX_S3_L003_R2_001.fastq.gz
214 Apr 21 14:33 MMC-125-135-144-C.mea

```

### **Use of script and working command will be:**

Note that cellranger will check for ids just before \_Ab or \_GEX.

```
/share/data/software/cellranger/cellranger-6.1.2/bin/cellranger multi --id=MMC-125-135-144-C --csv=MMC-125-135-144-C.config
```

The id generates an output directory containing all of the demultiplexing data and output explanation can be found in here: <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/summary-multi>

```

[[singlecell-RNAseq] [maider@login MMC-175-185-191-E]$ ll -l *
-rw-r--r-- 1 maider binf 124 Apr 20 19:33 cmdline
-rw-r--r-- 1 maider binf 262K Apr 20 19:34 filelist
-rw-r--r-- 1 maider binf 7.0M Apr 20 19:33 finalstate
-rw-r--r-- 1 maider binf 346 Apr 20 14:19 invocation
-rw-r--r-- 1 maider binf 5 Apr 20 14:19 jobmode
-rw-r--r-- 1 maider binf 239K Apr 20 19:33 log
-rw-r--r-- 1 maider binf 13M Apr 20 19:34 MMC-175-185-191-E.mri.tgz
-rw-r--r-- 1 maider binf 236K Apr 20 14:19 mrosource
-rw-r--r-- 1 maider binf 1.7M Apr 20 19:33 perf
-rw-r--r-- 1 maider binf 10K Apr 20 19:34 perf._truncated_
-rw-r--r-- 1 maider binf 25K Apr 20 19:34 sitecheck
-rw-r--r-- 1 maider binf 2 Apr 20 14:19 tags
-rw-r--r-- 1 maider binf 51 Apr 20 19:33 timestamp
-rw-r--r-- 1 maider binf 36 Apr 20 14:19 uuid
-rw-r--r-- 1 maider binf 360K Apr 20 19:33 vdrkill
-rw-r--r-- 1 maider binf 64 Apr 20 14:19 versions

outs:
total 8.0K
drwxr-xr-x 4 maider binf 73 Apr 20 19:33 .
drwxr-xr-x 4 maider binf 4.0K Apr 20 19:34 ..
-rw-r--r-- 1 maider binf 527 Apr 20 14:19 config.csv
drwxr-xr-x 4 maider binf 58 Apr 20 19:33 multi
drwxr-xr-x 5 maider binf 67 Apr 20 19:33 per_sample_outs

SC_MULTI_CS:
total 12K
drwxr-xr-x 12 maider binf 4.0K Apr 20 19:33 .
drwxr-xr-x 4 maider binf 4.0K Apr 20 19:34 ..
drwxr-xr-x 5 maider binf 82 Apr 20 19:33 BUILD_SAMPLE_OUTS
drwxr-xr-x 3 maider binf 26 Apr 20 19:33 BUILD_VDJ_OUTPUTS_CS
drwxr-xr-x 5 maider binf 134 Apr 20 19:33 fork0
drwxr-xr-x 4 maider binf 53 Apr 20 14:20 FULL_COUNT_INPUTS
drwxr-xr-x 3 maider binf 26 Apr 20 14:20 FULL_VDJ_INPUTS
drwxr-xr-x 3 maider binf 26 Apr 20 14:19 MULTI_PREFLIGHT
drwxr-xr-x 3 maider binf 26 Apr 20 14:19 MULTI_PREFLIGHT_LOCAL
drwxr-xr-x 3 maider binf 26 Apr 20 14:19 PARSE_MULTI_CONFIG
drwxr-xr-x 3 maider binf 26 Apr 20 19:32 SANITIZE_MAP_CALLS
drwxr-xr-x 18 maider binf 4.0K Apr 20 19:33 SC_MULTI_CORE
(singlecell-RNAseq) [maider@login MMC-175-185-191-E]$ 
```

```

[[singlecell-RNAseq] [maider@login outs]$ ll -l *
-rw-r--r-- 1 maider binf 527 Apr 20 14:19 config.csv

multi:
total 4.0K
drwxr-xr-x 4 maider binf 58 Apr 20 19:33 .
drwxr-xr-x 4 maider binf 73 Apr 20 19:33 ..
drwxr-xr-x 3 maider binf 4.0K Apr 20 19:33 count
drwxr-xr-x 2 maider binf 146 Apr 20 19:33 multiplexing_analysis

per_sample_outs:
total 0
drwxr-xr-x 5 maider binf 67 Apr 20 19:33 .
drwxr-xr-x 4 maider binf 73 Apr 20 19:33 ..
drwxr-xr-x 3 maider binf 83 Apr 20 19:33 MMC-0175
drwxr-xr-x 3 maider binf 83 Apr 20 19:33 MMC-0185
drwxr-xr-x 3 maider binf 83 Apr 20 19:33 MMC-0191
(singlecell-RNAseq) [maider@login outs]$ 
```

If I go to per\_sample\_outs and then to one of the sample folders:

```
[singlecell-RNAseq] [maider@login MMC-0175]$ ll -l *
-rw-r--r-- 1 maider binf 6.2K Apr 20 19:33 metrics_summary.csv
-rw-r--r-- 1 maider binf 5.0M Apr 20 19:33 web_summary.html

count:
total 7.0G
drwxr-xr-x 4 maider binf 4.0K Apr 20 19:33 .
drwxr-xr-x 3 maider binf 83 Apr 20 19:33 ..
drwxr-xr-x 7 maider binf 93 Apr 20 19:32 analysis
-rw-r--r-- 1 maider binf 31M Apr 20 19:32 cloupe.cloupe
-rw-r--r-- 1 maider binf 311 Apr 20 19:28 feature_reference.csv
-rw-r--r-- 1 maider binf 6.9G Apr 20 19:14 sample_alignments.bam
-rw-r--r-- 1 maider binf 4.3M Apr 20 19:14 sample_alignments.bam.bai
-rw-r--r-- 1 maider binf 95K Apr 20 19:08 sample_barcodes.csv
drwxr-xr-x 2 maider binf 86 Apr 20 19:28 sample_feature_bc_matrix
-rw-r--r-- 1 maider binf 6.8M Apr 20 19:07 sample_feature_bc_matrix.h5
-rw-r--r-- 1 maider binf 108M Apr 20 19:14 sample_molecule_info.h5
[singlecell-RNAseq] [maider@login MMC-0175]$ 

[singlecell-RNAseq] [maider@login MMC-0175]$ cd count/sample_feature_bc_matrix
[singlecell-RNAseq] [maider@login sample_feature_bc_matrix]$ ll
total 17M
drwxr-xr-x 2 maider binf 86 Apr 20 19:28 .
drwxr-xr-x 4 maider binf 4.0K Apr 20 19:33 ..
-rw-r--r-- 1 maider binf 20K Apr 20 19:07 barcodes.tsv.gz
-rw-r--r-- 1 maider binf 298K Apr 20 19:07 features.tsv.gz
-rw-r--r-- 1 maider binf 17M Apr 20 19:07 matrix.mtx.gz
[singlecell-RNAseq] [maider@login sample_feature_bc_matrix]$ 
```

Files for this analysis can be found in **cellranger\_multi folder**

Following scRNASeq scripts have been developed with this output. Information about the following steps are in:

[https://drive.google.com/drive/folders/1fGJgqbToiJtsH9wwXI4z2l9Z2TdzD\\_55](https://drive.google.com/drive/folders/1fGJgqbToiJtsH9wwXI4z2l9Z2TdzD_55)

### Second option: Use cellranger count command and then run Seurat

For the same sample format we have received there is a second way of analyzing this data.

This will be done by **TotalSeq\_cellranger.py script**

```
[maider@login scripts]$ python cellranger_count_Abs.py --help
usage: This script runs cellranger program from 10X genomics to map the single cell reads and generate the count
matrices. Exiting.....

Process command line arguments.

optional arguments:
  -h, --help            show this help message and exit
  -f FILE, --file_list FILE
                        fastq file list
  -c FILE, --config FILE
                        config file
```

<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/feature-bc-analysis>

Following this vignette we will follow this example.

```
cd /home/jdoe/runs  
cellranger count --id=sample345 \  
    --libraries=library.csv \  
    --transcriptome=/opt/refdata-gex-GRCh38-2020-A \  
    --feature-ref=feature_ref.csv \  
    --expect-cells=1000
```

As previously we need to create two csv files: CMO reference and config file with fastq path, sample name and library type. The difference with the first option relies in the feature\_type for this CMO reference file.

```
...~/pipelines/scRNASeq/cellranger_run — ssh maider@10.110.14.137 ~/Desktop/SC...  
id,name,read,pattern,sequence,feature_type  
HT01,HT01,R2,5PNNNNNNNNNN(BC),GTCAACTCTTAGCG,Antibody Capture  
HT02,HT02,R2,5PNNNNNNNNNN(BC),TGATGGCCTATTGGG,Antibody Capture  
HT03,HT03,R2,5PNNNNNNNNNN(BC),TTCCGCCTCTCTTG,Antibody Capture  
HT04,HT04,R2,5PNNNNNNNNNN(BC),AGTAAGTTCAGCGTA,Antibody Capture  
~
```

While the config file for each fastq file needs to be as follows:

```
...~/pipelines/scRNASeq/cellranger_run — ssh maider@10.110.14.137 ~/Desktop/Scripts/COLUMBIA — zsh  
fastq,sample,library_type  
/share/data/RNA_Seq/MECFS_scRNAseq/Working/test_April25_2022,MMC-79-89-93-A_GEX,Gene Expression  
/share/data/RNA_Seq/MECFS_scRNAseq/Working/test_April25_2022,MMC-79-89-93-A_Ab,Antibody Capture
```

Example can be found in:

```
[maider@login test_April25_2022]$ pwd  
/share/data/RNA_Seq/MECFS_scRNAseq/Working/test_April25_2022
```

### Use of script and working command will be:

The output file os this run will be different from the first Option. Here there will be only one matrix for the three sequenced together samples. Therefore we will need an additional step run with a R script.

Using the the three files in output filtered\_feature\_bc\_matrix folder we can run HTDemult from Seurat:

[https://satijalab.org/seurat/archive/v3.1/hashing\\_vignette.html](https://satijalab.org/seurat/archive/v3.1/hashing_vignette.html)

We load the three matrix files and follow the R script.

The R script for demultiplexing ([HTO\\_demult.html](#)) as well as examples of CMO

and config file can be found in [cellranger\\_count\\_Abs folder](#)