Cellranger 10x Start Guide

Helpful documents/drives

[Cellranger 10x Google drive](https://drive.google.com/drive/folders/1cbGNWtu_hd4xuNPSx7TUhQheb74GsAIR?usp=sharing)

[Cellranger 10X Workflow](https://docs.google.com/document/d/1odI88zTQHMpFUDvTwtgm3iZzbHN34deL6LY-f6fm6X4/edit?usp=sharing)

[Cellranger Flowchart](https://drive.google.com/file/d/1auj0UU11JqatKMwEIGvAPDhB3_diUkc6/view?usp=sharing)

[Master 10x sample log](https://docs.google.com/spreadsheets/d/1LjuBvxrGOGqZ0AGqIOXSwK50BmCuEn-JTqTh3Zgm7uM/edit)

Scripts

Create cellranger input files such as sample sheet, feature files (non-ADT), library files, transfer files, lsf parameters: [Input\_files\_generation.ipynb](https://drive.google.com/file/d/1APueUYjAhYYwYbqcqtgfvimpQlh5yjZt/view?usp=sharing)

Create feature files for ADT libraries: [clean\_adt\_panel.ipynb](https://drive.google.com/file/d/15rKTwVrcvqR42aiSPvLTsj1W-QKjtne9/view?usp=sharing)

Make sure to download the csv files necessary in the scripts to create input files (ADT\_HTO\_primer\_reference.csv, cell\_multiplexing\_reference.csv, hashtag\_antibodies\_reference.csv)

Tar folders on the Broad: [tarscript\_ broad.sh](https://drive.google.com/file/d/1cgaIlArVNndtSMnzsIMQtWgkX-C9YeFw/view?usp=sharing). Start a tmux session, **use UGER**, then run script. May need to double memory/time limits for very large (>1TB) tarballs.

Checkpoints

* Compare produced sample sheet to expected sample sheet (i.e. check indices, sample names).
* Check that all feature and library files look correct.
* Check each ASPERA transfer file (produced by input\_files notebook), and compare to the spreadsheet for each use.
* After completing tarring of the files on the Broad, check all errors and skim through outputs, checking that the file paths look correct for each library for each user. In particular, run **for name in \*out; do echo $name; cat $name | grep -E "\outs/web\_summary.html$"; done;** checking that all the expected sample folders are present for each output file. Also run **for name in \*out; do echo $name; head $name; tail $name; done;** to confirm that the output looks correct.
* After tarring folders on srlab, make sure to check all outputs and errors before deleting the original data folder. The tarscript output should state “Successfully completed” and the list of tarred folders should include the following folders: Data/Intensities/, InterOp/, Logs/, Thumbnail\_Images/, FASTQs/. Different types of libraries may contain different folders.

Other

Transfer output folders to user (cellranger\_version/transcriptome/BRI-\*\*\*). This includes count matrices and BAMS. For VDJ libraries, also transfer FASTQs, as bam2fastq function does not support VDJ libraries.

Download data from the [google bucket](https://console.cloud.google.com/storage/browser/mbcf_bwh_sc_core;tab=objects?pli=1&prefix=&forceOnObjectsSortingFiltering=false&authuser=0) using the code as copied once you press download

gsutil -m cp -r \

"gs://mbcf\_bwh\_sc\_core/211212\_10X\_KW9562-2\_bcl/" \

.

Srlab cleanup: for libraries that are 6 months or more old, remove the cellranger folder (with count matrices and BAMs) and tar the folder using [tarscript\_srlab.sh](https://drive.google.com/file/d/1Qx614qL_K0xa0XrPZyKi2fM9L202b1dS/view?usp=sharing). Check the logs (as described above), then remove the folder, leaving only the tarball.

* Transfer to MAD3: Run script [cp\_mad3.sh](https://drive.google.com/file/d/1g2nsVm7NojmGsWSOkDApS2dfOSfTpZ88/view?usp=sharing) to copy the tarballs to the MAD3 server and print out file integrities using checksum.
* Compare checksum files via:   
  **diff (<(awk '{ print $1 }' /data/srlab/bwh10x/03\_01\_checksum.txt) <(awk '{ print $1 }' /external/BWH-SCARDATA/bwh10x/03\_01\_checksum\_mad3.txt)**
  + This will result in line-by-line output if there are differences in the file checksums. Also check visually that the filepaths in the second column of the checksum files are different.

Moving FASTQs for count:

* Sometimes, FASTQs will not be located in individual directories (ex. FASTQs/outs/fastq\_path/FLOWCELL/BRI-1200/). This is ok for running count, but can be a problem when transferring FASTQs to users, or when we want to merge FASTQs from multiple work orders. Also, it’s always good to check the FASTQ sizes with **du -sh \*** for each sample to confirm there are reads associated with the sample index (>1 GB). Create a new folder for each sample (check command output first) and move FASTQs to them using the following commands:

**ls \* | grep "\_L001\_I1\_001.fastq.gz" | awk 'BEGIN { FS="\_S" } /1/ { print $1 }' | xargs mkdir ### make directories**

**for dir in $(ls -d \*/ | cut -d "/" -f 1); do echo $dir; mv -n ${dir}\_S\* $dir; done ### move FASTQS to sample directories**

Identifying indexes of undetermined reads:

* When demultiplexing samples, if an index is incorrect, the reads will go to the undetermined FASTQs. As a last resort, figure out what top indexes are present with the script [count\_barcode\_frequency\_smallbc.py](https://drive.google.com/file/d/1CfHlfyHo3QOviQau9xvEDZnWpWpPFCSJ/view?usp=sharing) adapted from: <https://gencore.bio.nyu.edu/how-to-find-out-what-barcodes-are-in-your-undetermined-reads/>
  + Run the script: **python3 count\_barcode\_frequency\_smallbc.py fastq\_path/Undetermined\_S0\_L001\_I1\_001.fastq.gz >> input\_barcodes\_Undetermined\_S0\_L001\_I1\_001.txt** (takes ~1.5 hr for 9G FASTQ without parallelization). View the most common barcodes with **tail -30 input\_barcodes\_Undetermined\_S0\_L001\_I1\_001\_smallbc.txt** , compare to [index sequences](https://www.google.com/search?q=cellranger+tt+index+primers&rlz=1C5GCEM_enUS961US961&oq=cellrange&aqs=chrome.0.69i59l3j0i433i512j69i57j0i512l5.1365j0j7&sourceid=chrome&ie=UTF-8).

Transferring data to user via public dropbox: <https://rc.partners.org/kb/article/1383>

For users with accounts on erisone, data transfer directly on the server is much faster and easier than transferring to aspera. Refer to above link for general information. Below are sample commands

ssh mc1070@erisonexf.partners.org  
mkdir /pub/dropbox/2022-09-28\_for\_USERID

cd /pub/dropbox/2022-09-28\_for\_USERID

cp -ir /data/srlab/bwh10x/KW\*\*/cellranger-6.1.1/GRCh38/\* . ## transfer relevant files

setfacl -R -m u:USERID:rwX /pub/dropbox/2022-09-28\_for\_USERID ## set read/write permissions