Cellranger 10X Work Flow

Table of Contents

[When is there data to process?](#_mps77ny4b3hy)

[Step 1: Check the Google Sheet:](#_yf5jrrzd7uqz)

[Check “Master 10X sample log” to see what kinds of runs are contained in the rawbcls](#_e3o5whpm6ol0)cd

[mRNA library](#_4c0i5j654rfe)

[Step 2: Make sample\_sheet.csv](#_hl41aakceqh4)

[Step 3: Run cellranger mkfastq](#_5mt6xrefphec)

[Step 4: Prep for running cellranger count](#_5c4lz1goutlb)

[Step 5: Run cellranger count](#_w1xprb1z61f6)

[mRNA library + ADT library](#_oj0l1ehgtg99)

[Step 2: Make sample\_sheet.csv](#_84u3jux857wi)

[Step 3: Run cellranger mkfastq](#_bmkm135tr67c)

[Step 4: Prep for running cellranger count](#_kgb948q2706s)

[Step 5: Run cellranger count](#_ejscaps07dx)

[mRNA library + hashing library](#_jm7a69o8wv54)

[Step 2: Make sample\_sheet.csv](#_v3cb9r8288yk)

[Step 3: Run cellranger mkfastq](#_5p0nar37ydb7)

[Step 4: Prep for running cellranger count](#_pcv0y0q6ymt8)

[Step 5: Run cellranger count](#_l7dw2axvc8ez)

[mRNA library + ADT library + hashing library (TBRU)](#_bax735sks62m)

[mRNA library + TCR/BCR library](#_p14rvtq2tm3r)

[ATACseq library](#_bzoy1k23t7bv)

[Transferring Data to Client](#_4ydy4n8l2u0z)

[Removal of Old Sequencing Data](#_xbasxu3t42yl)

# When is there data to process?

Rawbcl file comes through from sequencing (script checks and downloads automatically every night). Kevin might reach out and let us know some data is coming through (he gets an email from the sequencing core when it is coming through). If we start using other companies, then the data comes through an FTP transfer (Novogene) or on a hard drive (BGI)

# Step 1: Check the Google Sheet:

# Check “[Master 10X sample log](https://docs.google.com/spreadsheets/d/1LjuBvxrGOGqZ0AGqIOXSwK50BmCuEn-JTqTh3Zgm7uM/edit?usp=sharing)” to see what kinds of runs are contained in the rawbcls

The runs will be one of the following (click to see workflow):

1. mRNA library
2. mRNA library + ADT library
3. mRNA library + hashing library
4. mRNA library + ADT library + hashing library (TBRU)
5. mRNA library + TCR/BCR library
6. ATACseq library

# mRNA library

## Step 2: Make sample\_sheet.csv

Under the rawbcl file with matching work order number (KWXXXX) make sample\_sheet.csv with following format:

Lane,Sample,Index

1-4,BRI-136,SI-GA-A1

1-4,BRI-137,SI-GA-A2

1-4,BRI-138,SI-GA-A3

1-4,BRI-139,SI-GA-A4

Where Lane is the number of lanes used in sequencing (usually 4 unless otherwise specified), Sample is the BRI number of the runs contained in the work order/rawbcl folder and Index is the i7 index found in the “i7 Well” column of the [Master 10X sample log](https://docs.google.com/spreadsheets/d/1LjuBvxrGOGqZ0AGqIOXSwK50BmCuEn-JTqTh3Zgm7uM/edit?usp=sharing).

## Step 3: Run cellranger mkfastq

Copy the rawbcl folder name (ex. 190201\_KW6622\_10x\_rawbcl) into the file ```/data/srlab/bwh10x/lsf\_params\_mkfastq```

Run ```bash run\_lsf\_mkfast.sh```

Wait until you get an email that the fastqs have been made

## Step 4: Prep for running cellranger count

Under the rawbcl folder (190201\_KW6622\_10x\_rawbcl) make a directory called cellranger-3.1.0. Under that directory make a directory with the same name as the genome being used for the runs in the work order (ex. mm10 or GRCh38)

Make a directory called Logs under the rawbcl folder

## Step 5: Run cellranger count

Enter the parameters into the file ```/data/srlab/bwh10x/lsf\_params\_count``` according to the following format

Rawbcl folder (tab) BRI-number (tab) full\_path\_to\_alignment\_genome (tab) cellranger\_version (tab) genome

Some examples are:

190215\_KW6668\_10x\_rawbcl BRI-181 /data/srlab/external-data/10xgenomics/refdata-cellranger-mm10-3.0.0 cellranger-3.0.2 mm10

190201\_KW6622\_10x\_rawbcl BRI-158 /data/srlab/external-data/10xgenomics/refdata-cellranger-GRCh38-3.0.0 cellranger-3.0.2 GRCh38

Run ```bash run\_lsf\_counts.sh```

Wait for email saying results are done

# mRNA library + ADT library

## Step 2: Make sample\_sheet.csv

Under the rawbcl file with matching work order number (KWXXXX) make sample\_sheet.csv with following format:

Lane,Sample,Index

1-4,BRI-136,SI-GA-A1

1-4,BRI-137,ADT\_index

1-4,BRI-138,SI-GA-A3

1-4,BRI-139,ADT\_index

ADT\_indes refers to “REAP-seq Primers”

Where Lane is the number of lanes used in sequencing (usually 4 unless otherwise specified), Sample is the BRI number of the runs contained in the work order/rawbcl folder and Index is the i7 index for the mRNA libraries, or the ADT index for the ADT libraries

## Step 3: Run cellranger mkfastq

Copy the rawbcl folder name (ex. 190201\_KW6622\_10x\_rawbcl) into the file ```/data/srlab/bwh10x/lsf\_params\_mkfastq```

Run ```bash run\_lsf\_mkfast.sh```

Wait until you get an email that the fastqs have been made

## Step 4: Prep for running cellranger count

Under the rawbcl folder (ex. 190201\_KW6622\_10x\_rawbcl) make a directory called cellranger-3.0.2. Under that directory make a directory with the same name as the genome being used for the runs in the work order (ex. mm10 or GRCh38)

Make a directory called logs under the rawbcl folder

Make a file called features\_BRI-\*\*\*.csv for each mRNA library with the following format:

id,name,read,pattern,sequence,feature\_type

CD314-NKG2D\_prot,CD314(NKG2D)\_prot,R2,5P(BC),CGTGTTTGTTCCTCA,Antibody Capture

IgM\_prot,IgM\_prot,R2,5P(BC),TAGCGAGCCCGTATA,Antibody Capture

…

Base this file off the corresponding ADT panel used (see the panel spreadsheet…)

Make a file called library\_BRI-\*\*\*.csv for each mRNA library with the following format:

fastqs,sample,library\_type

/data/srlab/bwh10x/190502\_KW6944\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCGNBGXB/BRI-252/,BRI-252,Gene Expression

/data/srlab/bwh10x/190503\_KW6945\_10x\_rawbcl/FASTQS/outs/fastq\_path/HGLVFAFXY/,BRI-255,Antibody Capture

## Step 5: Run cellranger count

Enter the parameters into the file ```/data/srlab/bwh10x/lsf\_params\_count``` according to the following format

Rawbcl folder (tab) BRI-number (tab) full\_path\_to\_alignment\_genome (tab) features-BRI-\*\*\*.csv (tab) libraries-BRI-\*\*\* (tab) cellranger\_version (tab) genome

Some examples are:

190517\_KW7000\_10x\_rawbcl BRI-239 /data/srlab/external-data/10xgenomics/refdata-cellranger-mm10-3.0.0 features-239.csv libraries-239.csv cellranger-3.0.2 mm10

190517\_KW7000\_10x\_rawbcl BRI-250 /data/srlab/external-data/10xgenomics/refdata-cellranger-mm10-3.0.0 features-250.csv libraries-250.csv cellranger-3.0.2 mm10

Run ```bash run\_lsf\_counts\_feature.sh```

Wait for email saying results are done

# mRNA library + hashing library

## Step 2: Make sample\_sheet.csv

Under the rawbcl file with matching work order number (KWXXXX) make sample\_sheet.csv with following format:

Lane,Sample,Index

1-4,BRI-136,SI-GA-A1

1-4,BRI-137,Hashing\_index

1-4,BRI-138,SI-GA-A3

1-4,BRI-139,Hashing\_index

\*,BRI-591,SI-GA-XX

\*,BRI-600,TTTCCGCT

Where Lane is the number of lanes used in sequencing (usually 4 unless otherwise specified), Sample is the BRI number of the runs contained in the work order/rawbcl folder and Index is the i7 index for the mRNA libraries, or the Hashing index for the Hashing libraries

## Step 3: Run cellranger mkfastq

Copy the rawbcl folder name (ex. 190201\_KW6622\_10x\_rawbcl) into the file ```/data/srlab/bwh10x/lsf\_params\_mkfastq```

Run ```bash run\_lsf\_mkfast.sh```

Wait until you get an email that the fastqs have been made

## Step 4: Prep for running cellranger count

Under the rawbcl folder (ex. 190201\_KW6622\_10x\_rawbcl) make a directory called cellranger-3.0.2. Under that directory make a directory with the same name as the genome being used for the runs in the work order (ex. mm10 or GRCh38)

Make a directory called log under the rawbcl folder

Make a file called features\_BRI-\*\*\*.csv for each mRNA library with the following format:

id,name,read,pattern,sequence,feature\_type

Hashtag1,Hashtag1,R2,5P(BC),ACCCACCAGTAAGAC,Antibody Capture

Hashtag2,Hashtag2,R2,5P(BC),GGTCGAGAGCATTCA,Antibody Capture

Base this file off the corresponding hashing antibodies used (see the 10X master log spreadsheet...)

Make a file called library\_BRI-\*\*\*.csv for each mRNA library with the following format:

fastqs,sample,library\_type

/data/srlab/bwh10x/190502\_KW6944\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCGNBGXB/BRI-252/,BRI-252,Gene Expression

/data/srlab/bwh10x/190503\_KW6945\_10x\_rawbcl/FASTQS/outs/fastq\_path/HGLVFAFXY/,BRI-255,Antibody Capture

## Step 5: Run cellranger count

Enter the parameters into the file ```/data/srlab/bwh10x/lsf\_params\_count``` according to the following format

Rawbcl folder (tab) BRI-number (tab) full\_path\_to\_alignment\_genome (tab) features-BRI-\*\*\*.csv (tab) libraries-BRI-\*\*\* (tab) cellranger\_version (tab) genome

Some examples are:

190517\_KW7000\_10x\_rawbcl BRI-239 /data/srlab/external-data/10xgenomics/refdata-cellranger-mm10-3.0.0 features-239.csv libraries-239.csv cellranger-3.0.2 mm10

190517\_KW7000\_10x\_rawbcl BRI-250 /data/srlab/external-data/10xgenomics/refdata-cellranger-mm10-3.0.0 features-250.csv libraries-250.csv cellranger-3.0.2 mm10

Run ```bash run\_lsf\_counts\_feature.sh```

Wait for email saying results are done

# mRNA library + ADT library + hashing library (TBRU)

Combine both ADT and hashing steps

# mRNA library + TCR/BCR library

Still to come...

# ATACseq library

## Step 2: Make sample\_sheet.csv

Under the rawbcl file with matching work order number (KWXXXX) make sample\_sheet.csv with following format:

Lane,Sample,Index

1-4,BRI-136,SI-NA-A1

1-4,BRI-137,SI-NA-A2

1-4,BRI-138,SI-NA-A3

1-4,BRI-139,SI-NA-A4

Where Lane is the number of lanes used in sequencing (usually 4 unless otherwise specified), Sample is the BRI number of the runs contained in the work order/rawbcl folder and Index is the i7 index found in the “Chromium-i7-Multiplex-Kit-N-Set” column of the [Master 10X sample log](https://docs.google.com/spreadsheets/d/1LjuBvxrGOGqZ0AGqIOXSwK50BmCuEn-JTqTh3Zgm7uM/edit?usp=sharing). Note that for ATAC-seq the indexes are SI-**NA**-XX

## Step 3: Run cellranger mkfastq-atac

Copy the rawbcl folder name (ex. 191107\_KW7547\_10x\_rawbcl) into the file ```/data/srlab/bwh10x/scripts/lsf\_params\_mkfastq\_atac```

Run ```bash run\_lsf\_mkfast\_atac.sh```

Wait until you get an email that the fastqs have been made

## Step 4: Prep for running cellranger count-atac

Under the rawbcl folder (191107\_KW7547\_10x\_rawbcl) make a directory called cellranger-3.1.0. Under that directory make a directory with the same name as the genome being used for the runs in the work order (ex. mm10 or GRCh38/hg19)

Make a directory called log under the rawbcl folder

## Step 5: Run cellranger count-atac

Enter the parameters into the file ```/data/srlab/bwh10x/scripts/lsf\_params\_count\_atac``` according to the following format

Rawbcl folder (tab) BRI-number (tab) full\_path\_to\_alignment\_genome (tab) cellranger\_version (tab) genome

Some examples are:

191107\_KW7547\_10x\_rawbcl BRI-639 /data/srlab/external-data/10xgenomics/refdata-cellranger-atac-hg19-1.1.0

cellranger-3.1.0 hg19mj

191107\_KW7547\_10x\_rawbcl BRI-639 /data/srlab/external-data/10xgenomics/refdata-cellranger-atac-GRCh38-1.1.0

cellranger-3.1.0 GRCh38

NOTE: for each atac-seq run we align to GRCh38 and hg19

Run ```bash run\_lsf\_counts\_atac.sh```

Wait for email saying results are done

# Transferring Data to Client

Go to /data/srlab/bwh10x/aspera/

Modify send\_rsync.txt to include the paths to the fastq files and cellranger outs files that you want to send. Ex:

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-266/

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-267/

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-268/

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-269/

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-270/

190502\_KW6937\_10x\_rawbcl/FASTQS/outs/fastq\_path/HCCJNBGXB/BRI-271/

190502\_KW6937\_10x\_rawbcl/cellranger-3.0.2/GRCh38/BRI-266/

190502\_KW6937\_10x\_rawbcl/cellranger-3.0.2/GRCh38/BRI-268/

190502\_KW6937\_10x\_rawbcl/cellranger-3.0.2/GRCh38/BRI-269/

190502\_KW6937\_10x\_rawbcl/cellranger-3.0.2/GRCh38/BRI-271/

ONE TIME: make a file called send\_rsync\_maria.sh/send\_rsync\_fan.sh or something and copy structure of send\_rsync.sh, but with your login for the broad and your own destination directory on broad

Bash send\_rsync\_maria.sh to send files over (Make a Maria/Fan alternate to send\_rsync\_custom.sh, and use that script if you wish to send bam files as well)

Once files have been transferred, make a folder for the client in the broad directory (I usually give it their name, ex. PBrennan or PeiTong) and move the appropriate fastq files and cellranger output files into the folder.

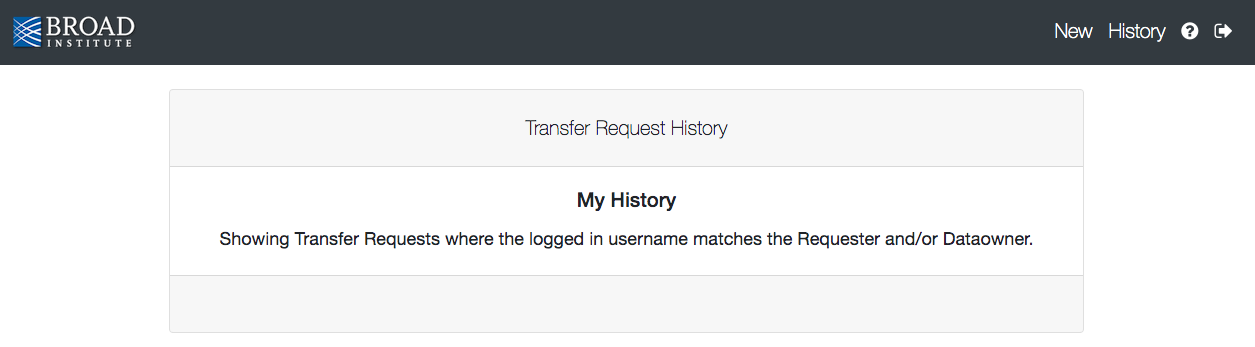
Tar the folder:

(tar -zcvf PBrennan.tar.gz PBrennan/)

(tar -zcvf PeiTong.tar.gz PeiTong/)

Go to the following website:

<https://transfer.broadinstitute.org/auth/login>

Log in and click New in the top right 

Enter your broad username as the Requester and the Data Owner in the folllowing form. I usually put something like “Count Matrices and Fastq files for X” in the Request name. Enter the full path on broad to the tar.gz file. Ex:

/broad/hptmp/jmears/bwh10x/PBrennan.tar.gz

And submit request.

You will get an email immediately saying that your request has been submitted. A few minutes later, you will get an email saying action required. Click the link in the email and click “activate website” or something (the instructions are pretty clear.) You will get three further emails. The final one will say “Thank you for using Aspera” and give a link to the data download. This is what we usually forward to the client!

Of course, if you want to streamline this process and do it your own way that’s fine! This is not the most efficient, but it is how I have been doing it so far. :)

# Removal of Old Sequencing Data

We let clients know that their data will be removed after 6 months. The detailed plan is as follows:

1. 6 months after the data is sent to the client, the rawbcl folder with all fastqs and output inside will be made into a tarball. The associated un-tarred folder will be removed
2. 1 year after the data is sent to the client, the folder will be un-tarred and all output **except** the fastq folder will be removed. The fastq folder will be tarred and placed in a folder called “Old\_FASTQ\_files”. The name of the fastq tarball will be as follows: KW###\_fastqs\_sequencingdate\_dateremoved

Addendum

To download data from DFCI ftp site:

Log into login node (like eris1n2)

Type screen or tmux

Connect to interactive file-move node with: bsub -Is -q filemove bash

Download the data witht wget or other command, for example:

wget --user bwh\_scRNAseq --password MBCFbwh\_scRNAseq18 --recursive --no-parent ftp://34.198.31.178/200731\_10X\_KW8115\_bcl/

Or log into ftp app:

ftp 34.198.31.178

bwh\_scRNAseq

MBCFbwh\_scRNAseq18