Processing 10X data

5' Single Cell gene expression with immune profiling

Demultiplexing (BCL -> FASTQ)

The reads from the Clinical Genomics Core are delivered in one of two *raw* formats: BCL or FASTQs. If they are BCLs, the data will need to be converted to FASTQ before mapping and counting by CellRanger.

Sample Sheet

Start by preparing the sample sheet. Likely, you have already prepared this for submitting samples to sequencing. The samplesheet should be in comma delimited format (i.e. .csv) and it its most basic form, should have three sections - Header, Reads, and Data - like so:

```
[Header],,,,
EMFileVersion,4,,,
,,,,
[Reads],,,,
26,,,,
90,,,,
,,,,
[Data],,,,
Sample_ID,Sample_Name,index,index2,Sample_Project
```

Additionally, if the sequencing run was divided into two or more lanes, a "lane" column can be added to the [Data] section.

> [!NOTE] If editing using a text editor, you need to ensure that all lines have the same number of columns (i.e. has the same number of commas)

The information in the sample sheet is used to separate reads belonging to each sample and to name the resulting FASTQs: * Sample_ID: this will be prepended to the name of the resulting files matching the two indices below * Sample_Name: not used * index: the i7 index * index2: the i5 index * Sample_Project: this will be used to group output files into folders

bcl-convert

Illumina makes it difficult to install bcl-convert, so it is necessary to run it from within a container using either Singularity or Apptainer. At current (2023-09-11), there is a Singularity container with bcl-convert version 4.1.7 in `/Volumes/guth_aci_informatics/software`. To run bcl-convert:

```
apptainer run \
    --bind /s/guth-aci/var:/var \
    --bind /s/guth-aci:/s/guth-aci \
    /Volumes/guth_aci_informatics/software/bclconvert-4.1.7.sif \
    bcl-convert \
        --output-directory /s/guth-aci/{PROJECT}/data/fastqs/ \
        --bcl-input-directory /s/guth-aci/{PROJECT}/data/bcls/ \
        --sample-sheet /s/guth-aci/{PROJECT}/metadata/samplesheet.csv \
        --force \
        --no-lane-splitting true \
        --bcl-sampleproject-subdirectories true
```

> [!NOTE] Make sure that the `/s/guth-aci/var` directory exists.

The first two lines that start with `--bind` map a directory outside to a location inside the container. You will need to adjust the `--output-directory`, `--bcl-input-directory`, and `--sample-sheet` arguments to match the desired destination for the fastqs, the location of the bcls, and location of the sample sheet, respectively. If your data was split by lane, set `--no-lane-splitting` to `false`.

Count