**AR sequencing analysis pipeline – command line download**

**Purpose:** This document describes the bioinformatics analysis of antigen receptor sequencing data starting from the fastq file download from Basespace and including the Interrogate analysis.

The analysis can be done on a **local computer** or on a **Compute Canada cluster**. Transfer of fastq files from Basespace to a local computer using the ‘Download’ option on Basespace will create a different folder structure than the transferring the files using command line. The following **STANDARD** refers to the **direct file transfer from Basespace to Compute Canada via command line.**

The general **folder structure** should be as follows:

*Run<number and designation>*

*Data: contains large data files*

*InterrogateAnalysis: contains scripts; clone from Github*

*Results: contains results of analysis; small enough files to download to local computers*

Log into compute Canada account

cd /work/def-smkeller/Shared/SequencingData/

#create project folder (see name in Basepace)

$ mkdir Run<no>\_<RunName>

#Clone folder with analysis scripts

$ cd Run<no>\_<RunName>

$ git clone https://github.com/theKellerLab/IntrgAnalysis

#create folder for Basespace data

$ mkdir –p Data/Basespace

#========= Transfer fastq files from Basespace to Compute Canada ==============

$ cd Data/Basespace

See separate document:

0\_preClntab/0\_FastqBasespace2ComputeCanada.docx

$rm download.txt

The resulting folder structure looks like this:

Basespace/

ds.0baa9e37cb924a74aa950860a791ea2f/

18-069171-2D1P3C1L1P1\_L001\_ds.0baa9e37cb924a74aa950860a791ea2f.json

Parry-Lola\_S23\_L001\_R1\_001.fastq.gz

Parry-Lola\_S23\_L001\_R2\_001.fastq.gz

...

#========= Change file structure ==============

The subsequent trimming step with trimmomatic requires R1&R2 to be in one folder and each sample having a different folder. So the goal is: 1) to retain the general folder structure and 2) to concatenate the sampleId and sample name:

$ cd ../../IntrgAnalysis/0\_preClntab/

$ bash 0b\_moveFastq\_AfterCommandLineDownload\_separateFolders.sh

Original/

16-091599-1D1P1C1P1C1\_Anderson-Ditto\_S1/

16-091599-1D1P1C1P1C1\_Anderson-Ditto\_S1\_L001\_R1\_001.fastq.gz

16-091599-1D1P1C1P1C1\_Anderson-Ditto\_S1\_L001\_R2\_001.fastq.gz

#====== Combine fastq in one folder **OBSOLETE** =============

bash 0\_Move&QcFastqFiles/0b\_moveOriginal\_commandLineDownload.sh

#====== Option 1: using fastqc & multiqc & trimmomatic (see fastp approach below) ===========

Advantage:

* fastqc output is amenable to multiQC analysis

Disadvantage:

* QC & trimming are two separate steps
* trimmomatic is apparently slow compared to other programs (<https://academic.oup.com/bioinformatics/article/34/17/i884/5093234>)

#====== run QC & multiQC =============

bash 1\_qcOriginal.sh

#====== trim reads =====

bash 2\_trim30.sh

#====== run QC & multiQC on trimmed reads =============

bash 3\_qcTrimmed30

#====== Option 2: using fastp ===========

Advantage:

* QC and trimming are done in one step
* fastp is faster than trimmomatic

Disadvantages:

* fastp output seems not amenable to multiQC analysis

1\_qcAndTrim\_fastp.sh

#========= upload to Interrogate & run analysis ============

Upload instructions:

6a\_TrimmedFromComputeCanada2Interrogate.txt

Analysis instructions:

6b\_InterrogateAnalysis.docx

#========= Analyse Interrogate run report ============

Download run report from Interrogate and save under ‘Data/InterrogateRunReport/<file>’

#========= Transfer Clntab files to ComputeCanada ============

6c\_ClntabFromInterrogate2ComputeCanada.sh

#========= Optional: Transfer Clntab files from ComputCanada to local computer ============

If not done yet: go to ComputeCanada and Create tarball:

$ tar -czvf Clntab\_RDS.tar.gz Clntab\_RDS

Copy path – eg:

/work/def-smkeller/Shared/SequencingData/Run25\_JenkinsWylde\_v2/Data/ Clntab\_RDS.tar.gz

exit and go to local target directory; download:

rsync -ave ssh smkeller@graham.computecanada.ca:/work/def-smkeller/Shared/SequencingData/Run25\_JenkinsWylde\_v2/Data/Clntab\_RDS.tar.gz ./