**ChIp-Seq analysis workflow using Singularity platform on NIH-Biowulf**

**Brief description about Singularity:**

Singularity is a container solution created by necessity for scientific and application drove workloads.

**Singularity features and limitations:**

* Admin of singularity container has full control over it.
* User runs Container in read-only mode
* By default, the home directory of user Biowulf account is shared with the container, so user has access to his home directory within container
* User should define INPUT/OUTPUT path before running the container if the path are different than his home directory with command named: --BIND

<http://singularity.lbl.gov/docs-mount>

* User can not edit anything in Singularity container at Biowulf (Biowulf admin disable this feature for safety reason) Quick fix for that is the user should install VAGRANT on his desktop machine and edit the container over there with command named: --writable using sudo.

<http://singularity.lbl.gov/install-mac>

The user can run singularity container in two modes:

**1) exec mode:** *(run the workflow with single command)*

The exec Singularity sub-command allows you to spawn an arbitrary command within your container image as if it were running directly on the host system.

**2) shell mode:** *you will enter the container bash mode and can explore around the container space (everything in here is read-only)*

The shell Singularity sub-command will automatically spawn a shell within a container. The default and the only requirement of any Singularity container are that /bin/sh must be present, and thus /bin/sh is also used as the default shell.

**Running ChIp-Seq workflow**

Where is my home directory?

Home Directory is the first directory when you log in the Biowulf

Home\_path: /home/$USER

\*The user should be particular about three directories path and one design file

**1) Input directory:**

This is the directory, which contains your fastq files.

*The ChIp-seq workflow will recognize fastq files with .fastq, and .fq extension; other data will not be processed; nothing will be written to this directory just reading from it*

**2) Output directory:**

This is the directory, which everything will be written to it,

*The size of this directory should be big enough two times of your Input directory should be sufficient; owner (user) should have read/write/execute privileges of this directory.*

*If you are not sure you can run chmod 0755 Output directory/*

**3) Index directory:**

This is the path for bowtie human genome index path on Biowulf

*For now we are using bowtie 1.0 so use the proper index version. Here is the path to the hg19: (/fdb/igenomes/Homo\_sapiens/UCSC/hg19/Sequence/BowtieIndex)*

**4) Design file**

Design file describes grouping information about each fastq file

*Use tab delimited format, not the excel one for now.*

*\*\*Please have this file available on your home directory*

**Command line:**

Please be at you home directory before running the workflow

1) cd /home/$USER

2) sinteractive --cpus-per-task=4 --mem=10g

3) module load singularity

4) singularity exec --bind your\_input\_directory:/INPUTDIR,your\_output\_directory:/OUTPUTDIR,bowtie\_1.0\_index\_path:/INDEXDIR /data/shamsaddinisha/chips\_singularity/singularity/chips.img python /chips.py --design chips\_design.txt

\*\*\*As you can see The container image is in my directory (/data/shamsaddinisha/chips\_singularity/singularity/) you can copy that however you can run that from mine too

**Checking Results:**

1) You can check the result files in your specified output directory path

2) ChIp-seq workflow will generate ChIps\_report.txt and ChIps\_result.xlsx