Run this session using tmux or screen,
because after 15 minutes of inactivity, SCC interrupts the session.

You can use any directory names, but it will be better if for all your sources
you will use separate directory.

Create a directory which you are going to use for project

```
/projectnb2/wax-es
[scc4] >> mkdir -p Zach/G182
```

```
/projectnb2/wax-es
[scc4] >> cd Zach/G182
```

**Clone the last version of the DNAChIP pipeline (do not forget the last dot in
the command)**

```
/projectnb2/wax-es/Zach/G182
[scc4] >> git clone /projectnb2/wax-es/routines/DNAChIP_pipeline .
Cloning into '.'...
done.
```

```
/projectnb2/wax-es/Zach/G182
[scc4] >> ls -l
total 2
drwxr-sr-x  4 mpyatkov wax-es 4096 Sep 14 13:37 .
drwxr-sr-x  3 mpyatkov wax-es 4096 Sep 14 13:36 ..
drwxr-sr-x  8 mpyatkov wax-es 4096 Sep 14 13:37 .git
drwxr-sr-x 18 mpyatkov wax-es 4096 Sep 14 13:37 Scripts
```

**Copy fastq samples to this directory (actually the pipeline does not change
the fastq files, so we can just make a links. /projectnb2/wax-es/G182/ -
directory already contains the links so just copy it)**

```
/projectnb2/wax-es/Zach/G182
[scc4] >> cp -a /projectnb2/wax-es/G182/G182_M{1,2,3,7,8,9,10,11,12,13} ./
```

```
/projectnb2/wax-es/Zach/G182
[scc4] >> ls -l
total 7
```

```
drwxr-sr-x 14 mpyatkov wax-es 4096 Sep 14 13:41 .
drwxr-sr-x  3 mpyatkov wax-es 4096 Sep 14 13:36 ..
drwxr-sr-x  8 mpyatkov wax-es 4096 Sep 14 13:37 .git
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M1
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M10
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M11
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M12
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M13
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M2
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M3
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M7
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M8
drwxr-sr-x  2 mpyatkov wax-es 4096 Aug 31 14:22 G182_M9
drwxr-sr-x 18 mpyatkov wax-es 4096 Sep 14 13:37 Scripts
```

Inspect one of the directories (G182_M1). It contains links to the fastq files

```
/projectnb2/wax-es/Zach/G182
[scc4] >> tree G182_M1
G182_M1
|-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_1.fq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_1.fq.gz
`-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_2.fq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_2.fq.gz
```

The pipeline required that all fastq files should be located in the fastq directory and names should contain R{1,2}.gz identifiers for read1 and read2. Let's do it

Changing location

```
/projectnb2/wax-es/Zach/G182
[scc4] >> for i in G182*/ ; do pushd $i; mkdir fasta; mv ./*.gz ./fasta; popd;
done
```

/projectnb2/wax-es/Zach/G182

```
[scc4] >> tree G182_M1
G182_M1
`-- fasta
        |-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_1.fq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_1.fq.gz
        `-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_2.fq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_2.fq.gz
```

Renaming fastq files

```
/projectnb2/wax-es/Zach/G182
[scc4] >> find . -name '*_1.fq.gz' -exec bash -c 'mv $0
${0/_1.fq.gz/_R1.fastq.gz}' {} \;
```

```
/projectnb2/wax-es/Zach/G182
[scc4] >> find . -name '*_2.fq.gz' -exec bash -c 'mv $0
${0/_2.fq.gz/_R2.fastq.gz}' {} \;
```

```
/projectnb2/wax-es/Zach/G182
[scc4] >> tree G182_M1
G182_M1
`-- fasta
        |-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_R1.fastq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_1.fq.gz
        `-- G182_M1_CKDL200148791-1a-7UDI239-AK7213_HTMC7DSXX_L1_R2.fastq.gz ->
/net/waxman-server/mnt/data/volume2/Waxman_Illumina_HiSeq_Raw_Data_02/G174-G18
2/128.120.88.251/H202SC20020757/Rawdata/G182_M1/G182_M1_CKDL200148791-1a-7UDI2
39-AK7213_HTMC7DSXX_L1_2.fq.gz
```

Pipeline required to create/edit several configuration files.
1. Sample_Labels.txt - to set up samples
2. 01_Pipeline_Setup.sh - to set up options
3. Addition setup for step 10 (consider latter)

go to the directory with configuration files

```
/projectnb2/wax-es/Zach/G182
[scc4] >> cd Scripts/00_Setup_Pipeline

/projectnb2/wax-es/Zach/G182/Scripts/00_Setup_Pipeline
[scc4] >> ls -l
total 58
-rwxr-xr-x  1 mpyatkov wax-es   7038 Sep 14 13:37 01_Pipeline_Setup.sh
-rwxr-xr-x  1 mpyatkov wax-es   5743 Sep 14 13:37
02_Review_Pipeline_Parameters.sh
-rwxr-xr-x  1 mpyatkov wax-es  13054 Sep 14 13:37 03_Run_Pipeline.sh
-rwxr-xr-x  1 mpyatkov wax-es  19761 Sep 14 13:37 Pipeline_Version_History.txt
-rw-r--r--  1 mpyatkov wax-es    121 Sep 14 13:37 Sample_Labels.txt
```

**Sample_Labels.txt tab separated file. 4 columns. First and second columns
(Sample_ID, Sample_ID) should be the same and equal to the directories with
your samples. Third column - description (without spaces, use underscore).
Fourth - color, it will be used on the last steps when the pipeline creates
tracks for UCSC browser. It is just easy to use MS Office to prepare this file
and export as TSV.**

```
[scc4] >> cat Sample_Labels.txt
Sample_DIR  Sample_ID   Description Color
G182_M1 G182_M1 CTCF_Control  128,0,0
G182_M2 G182_M2 CTCF_TCPOBOP_1mg    128,128,128
....
```

**01_Pipeline_Setup.sh - contains multiple options but to run the pipeline is
not required to change all of them, default setup is good. Change the
highlighted options, and it will be enough to start the pipeline.**

```
/projectnb2/wax-es/Zach/G182/Scripts/00_Setup_Pipeline
[scc4] >> head -n 20 01_Pipeline_Setup.sh

################################################################################
####
#Andy Rampersaud, 07.19.17
#This 01_Pipeline_Setup.sh file is used to initialize all variables used in
the pipeline
#The purpose of this file is to streamline the pipeline so that
#     1. Variables are initialized from a single file
#     2. Minimize redundancy between pipeline steps
```

```
#This file is organized by
#     1. "Global Variables"       = variables used by multiple steps
#     2. "Step-specific Variables" = variables used by a particular step
################################################################################
####
#------------------------------------------------------------------------------
----
#"Global Variables"  = variables used by multiple steps
#------------------------------------------------------------------------------
----
##changed the directory path, BU_User and Dataset_Label
Dataset_DIR=/projectnb/wax-es/test_DNAChIP # <-- /projectnb2/wax-es/Zach/G182
BU_User="mpyatkov"                          # <-- your SCC user name
Dataset_Label="G182"                        # <-- label
GTF_Files_DIR=/projectnb/wax-es/aramp10/GTF_Files
#------------------------------------------------------------------------------
----
################################################################################
####
```

To run the pipeline just use the following file it should go step by step and run apropriate scripts inside each directory (Run_Jobs.sh - preprocessing, file with qsub extension - processing, Summarize_Jobs.sh - post-processing)

Check the options:
/projectnb2/wax-es/Zach/G182/Scripts/00_Setup_Pipeline
[scc4] >> ./03_Run_Pipeline.sh
Usage: ./03_Run_Pipeline.sh <Start_Step>
<Start_Step> = Needs to be either "Full_Pipeline" or a specific pipeline step
See 03_Run_Pipeline.sh for details.

/projectnb2/wax-es/Zach/G182/Scripts/00_Setup_Pipeline
[scc4] >> ./03_Run_Pipeline.sh Full_Pipeline

Do not forget to run the previous command in tmux/screen.
If your setup is ok, pipeline should show you an error on Step10, because we did not configure it yet.