

Example alignment work file

Below is an example "work" file.

```
#!/bin/bash

# -----
# This file shows steps taken to process October 2014 Science Park GBM data.
# It is a record of what was done, NOT meant to be run as an actual script.
# -----

# =====
# At TACC (Ls5)
# =====
# Set up environment
export PATH=/work/01063/abattenh/local/bin:$PATH

# Download NGS1140, ID19541 data from Science Park to $SCRATCH/seq/original/tmp_hold at TACC
cd $SCRATCH/seq/original/tmp_hold
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_CTCF_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_H3K4Me3_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_RNAP2_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_H3K4Me1_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_RNAP2_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_H3K27Ac_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_H3K9Ac_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_CTCF_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_H3K4Me3_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_input_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_H3K27Me3_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD01c_H3K9Me3_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_H3K27Ac_B2.tar
wget --user=viyer --password=SecretPassword https://spi.mdanderson.org/viyer2/20141007_NGS1140
/Sample_SD06_H3K9Ac_B2.tar

# on a login node
cd $SCRATCH/seq/original/tmp_hold
rm -rf tars
mkdir tars
for f in *.tar; do
    tar -xvf $f
    mv $f tars/
done

mkdir -p $SCRATCH/seq/original/Bc_Illumina/Oct14.SciPark/ID19541

cd $SCRATCH/seq/original/tmp_hold
mv Sample_* $SCRATCH/seq/original/Bc_Illumina/Oct14.SciPark/ID19541

# -----
# Alignment prep
# -----
mkdir -p $SCRATCH/seq/align/2014_10.sciPk/fq
mkdir -p $SCRATCH/seq/align/2014_10.sciPk/hg19/gbm

cd $SCRATCH/seq/align/2014_10.sciPk/fq
find $SCRATCH/seq/original/Bc_Illumina/Oct14.SciPark/ID19541 -name "*.gz" | xargs ln -s -f -t .
```

```

ls -1 *_R1*.gz
# SD01c_CTCF_B2_CTTGTA_L008_R1_001.fastq.gz
# SD01c_H3K27Ac_B2_TACAGC_L008_R1_001.fastq.gz
# SD01c_H3K27Me3_B2_CTATAC_L008_R1_001.fastq.gz
# SD01c_H3K4Me3_B2_ATGAGC_L008_R1_001.fastq.gz
# SD01c_H3K9Ac_B2_GCGCTA_L008_R1_001.fastq.gz
# SD01c_H3K9Me3_B2_CACCGG_L008_R1_001.fastq.gz
# SD01c_RNAP2_B2_CCGTCC_L008_R1_001.fastq.gz
# SD06_CTCF_B2_CAACTA_L007_R1_001.fastq.gz
# SD06_H3K27Ac_B2_TAATCG_L007_R1_001.fastq.gz
# SD06_H3K4Me1_B2_TATAAT_L007_R1_001.fastq.gz
# SD06_H3K4Me3_B2_ACTTGA_L007_R1_001.fastq.gz
# SD06_H3K9Ac_B2_CAGGCG_L007_R1_001.fastq.gz
# SD06_input_B2_TCATTTC_L007_R1_001.fastq.gz
# SD06_RNAP2_B2_CACGAT_L007_R1_001.fastq.gz

# count.cmds
for f in SD01c_[CR]*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done
for f in SD01c_H3K27*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done
for f in SD01c_H3K[49]*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done
for f in SD06_[CRi]*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done
for f in SD06_H3K[29]*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done
for f in SD06_H3K4*_R[12]_001.fastq.gz; do echo -e "$((`zcat $f | wc -l` / 4))\t$f" | tee ${f}_stats.txt; done

cd $SCRATCH/seq/align/2014_10.sciPk/fq
launcher_maker.py -t 1 -n count.cmds -w 8 -q dev -v -a CancerGenetics
sbatch count.slurm; showq -u

ls *.gz_stats.txt | xargs cat > fq_stats.txt
rm *.gz_stats.txt

# fqc.cmds
fastqc SD01c_[CR]*_R[12]_001.fastq.gz
fastqc SD01c_H3K27*_R[12]_001.fastq.gz
fastqc SD01c_H3K[49]*_R[12]_001.fastq.gz
fastqc SD06_[CRi]*_R[12]_001.fastq.gz
fastqc SD06_H3K[29]*_R[12]_001.fastq.gz
fastqc SD06_H3K4*_R[12]_001.fastq.gz
# done
cd $SCRATCH/seq/align/2014_10.sciPk/fq
launcher_maker.py -t 4 -n fqc.cmds -w 8 -v -a CancerGenetics -m fastqc
sbatch fqc.slurm; showq -u

rm *.zip

# -----
# Align Amelia's SciPk GBM ChIP
# -----
# sync code
rsync -avrP --delete --exclude=.git abattenh@iycerstor01.icmb.utexas.edu:/home/abattenhouse/gitdir/bioiteam/bioi/ $STOCKYARD/gitdir/bioiteam/bioi/
rsync -avrP --delete --exclude=CVS abattenh@iycerstor01.icmb.utexas.edu:/home/abattenhouse/sequencing/code/ $STOCKYARD/seq/code/

cd $SCRATCH/seq/align/2014_10.sciPk/hg19/gbm
ln -s -f ../../fq
ls -1 *_R1*.gz
# SD01c_CTCF_B2_CTTGTA_L008_R1_001.fastq.gz
# SD01c_H3K27Ac_B2_TACAGC_L008_R1_001.fastq.gz
# SD01c_H3K27Me3_B2_CTATAC_L008_R1_001.fastq.gz
# SD01c_H3K4Me3_B2_ATGAGC_L008_R1_001.fastq.gz
# SD01c_H3K9Ac_B2_GCGCTA_L008_R1_001.fastq.gz
# SD01c_H3K9Me3_B2_CACCGG_L008_R1_001.fastq.gz
# SD01c_RNAP2_B2_CCGTCC_L008_R1_001.fastq.gz
# SD06_CTCF_B2_CAACTA_L007_R1_001.fastq.gz

```

```

# SD06_H3K27Ac_B2_TAATCG_L007_R1_001.fastq.gz
# SD06_H3K4Me1_B2_TATAAT_L007_R1_001.fastq.gz
# SD06_H3K4Me3_B2_ACTTGA_L007_R1_001.fastq.gz
# SD06_H3K9Ac_B2_CAGGCG_L007_R1_001.fastq.gz
# SD06_input_B2_TCATTC_L007_R1_001.fastq.gz
# SD06_RNAP2_B2_CACGAT_L007_R1_001.fastq.gz

# aln.cmds
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq/SD01c_CTCF_B2_CTTGTA_L008_R1_001.
fastq.gz sd01_ctcf_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_H3K27Ac_B2_TACAGC_L008_R1_001.fastq.gz sd01_h3k27ac_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_H3K27Me3_B2_CTATAC_L008_R1_001.fastq.gz sd01_h3k27me3_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_H3K4Me3_B2_ATGAGC_L008_R1_001.fastq.gz sd01_h3k4me3_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_H3K9Ac_B2_GCGCTA_L008_R1_001.fastq.gz sd01_h3k9ac_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_H3K9Me3_B2_CACGCG_L008_R1_001.fastq.gz sd01_h3k9me3_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD01c_RNAP2_B2_CCGTCC_L008_R1_001.fastq.gz sd01_pol2_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq/SD06_CTCF_B2_CAACTA_L007_R1_001.
fastq.gz sd06_ctcf_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD06_H3K27Ac_B2_TAATCG_L007_R1_001.fastq.gz sd06_h3k27ac_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD06_H3K4Me1_B2_TATAAT_L007_R1_001.fastq.gz sd06_h3k4me1_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD06_H3K4Me3_B2_ACTTGA_L007_R1_001.fastq.gz sd06_h3k4me3_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq
/SD06_H3K9Ac_B2_CAGGCG_L007_R1_001.fastq.gz sd06_h3k9ac_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq/SD06_input_B2_TCATTC_L007_R1_001.
fastq.gz sd06_input_b2 hg19 1 50
/work/projects/BioITeam/common/script/script/align_bwa_illumina.sh global ./fq/SD06_RNAP2_B2_CACGAT_L007_R1_001.
fastq.gz sd06_pol2_b2 hg19 1 50

launcher_maker.py -t 24 -n aln.cmds -w 2 -v -a CancerGenetics
sbatch aln.slurm; showq -u

# combine sd06 inputs
mkdir -p $SCRATCH/seq/align/2014_10.sciPk/hg19/gbm/merge
cd $SCRATCH/seq/align/2014_10.sciPk/hg19/gbm/merge
ln -s ../sd06_input_b2.sort.dup.bam
cp -p $IYC/seq/align/2013_07.awh_sciPk/hg19/amelia/sd06a_input_b1.sort.dup.bam .

# merge.cmds
/work/projects/BioITeam/common/script/script/merge_sorted_bams.sh sd06_input_cmb sd06*_input_b[12].sort.dup.bam

launcher_maker.py -t 6 -n merge.cmds -w 2 -v -a CancerGenetics
sbatch merge.slurm; showq -u

# -----
# Fix file permissions then rsync back
# -----
cd $SCRATCH/seq/original/Bc_Illumina/Oct14.SciPark/
find . -type d | xargs chmod +x
chmod -R +r *
rsync -avrP $SCRATCH/seq/original/Bc_Illumina/Oct14.SciPark/ \
abattenh@iycercomp03.cccb.utexas.edu:/seq/original/Bc_Illumina/2014/2014_10.SciPark/

cd $SCRATCH/seq/align/2014_10.sciPk
find . -type d | xargs chmod +x
chmod -R +r *
rsync -avrP $SCRATCH/seq/align/2014_10.sciPk/ abattenh@iycercomp03.cccb.utexas.edu:/seq/align/2014_10.sciPk/

```