Run **proseqMapper.bsh** to:

* Collapse duplicate reads
* Trim 8 bp UMI from 5’ end of reads
* Remove 3’ adaptor (originally ligated to 5’ end of RNA)
* Align to genome (BWA)
* Convert bam to bigWig

CBSU location: /home/pr46\_0001/ChRO-seq/proseq/proseqMapper\_36threads\_randomOctomer.bsh (should run on machine with 36+ threads or change thread number in script)

To run:

bash proseqMapper\_36threads\_randomOctomer.bsh \

-i [bwa-index] \

-c [chromInfo] \

-b8 \ #flag to trim 8 bp UMI

-q \ #flag to perform QC

-O [output directory] \

-I [input fastq files] #if no input files are given, will use \*.fastq.gz

Outputs:

{sample}\_minus.bw

{sample}\_plus.bw

{sample}.sort.bam

Some QC (read length distribution) plots

Example:

nohup bash proseqMapper\_36threads\_randomOctomer.bsh -i /home/pr46\_0001/projects/genome/GRCh38.p7/GRCh38.primary\_assembly.genome -c /home/pr46\_0001/projects/genome/GRCh38.p7/GRCh38.chrom.sizes -b8 -q -O FLC07\_09 &> mapping\_FLC07\_09\_log.out &

Merge bigwigs with **mergeBigWigs\_4samples.bsh**:

* Note: we still have to compare differences if we merge here (before using dREG to call peaks) or after dREG (combining peaks from bed files).
* This script is currently really non-flexible and only merges a specific number of bigwigs. Currently, you have to modify the script if you want to merge a different number of bigwigs. Should be updated eventually.
* Must change chromInfo if using different species or genome build than specified
* Converts bigwigs to bedgraph files, merges them, and then converts back to bigwig

CBSU location: /home/pr46\_0001/ChRO-seq/proseq/mergeBigWigs\_4samples.bsh

To run:

bash mergeBigWigs\_4samples.bsh \

{output prefix} \

{sample1}\*.bw \

{sample2}\*.bw \

{sample3}\*.bw \

{sample4}\*.bw

Outputs:

{prefix}\_all-merge.minus.bw

{prefix}\_all-merge.plus.bw

Examples:

nohup bash mergeBigWigs\_4samples.bsh FLC 9007\_5598\_66054\_HY23MBGX3\_FLC07\_Fibrotic\_Tumor\_ATCACG\_R1\_\*bw 9007\_5598\_66055\_HY23MBGX3\_FLC09\_MUOD\_Tumor\_CGATGT\_R1\_\*bw 9008\_5598\_66051\_HWFCVBGX3\_FLC07\_Fibrotic\_Tumor\_ATCACG\_R1\_\*bw 9008\_5598\_66052\_HWFCVBGX3\_FLC09\_MUOD\_Tumor\_CGATGT\_R1\_\*bw &> merge\_FLC\_tumor\_bigWigs.log.out &

nohup bash mergeBigWigs\_2samples.bsh NML 9007\_5598\_66056\_HY23MBGX3\_FLC09\_MUOD\_Normal\_TTAGGC\_R1\_\*bw 9008\_5598\_66053\_HWFCVBGX3\_FLC09\_MUOD\_Normal\_TTAGGC\_R1\_\*bw &> merge\_NML\_bigWigs.log.out &

Merge dREG with **run\_dREG.bsh**:

* dREG uses a SVR model to find transcriptional regulatory elements
* run\_dREG.bsh predicts dREG scores (outputs bedgraph of scores)
* Use writeBedv2.bsh to pull out peaks with a score above a specified threshold

CBSU location: /home/pr46\_0001/ChRO-seq/dREG/ run\_dREG.bsh

To run:

bash run\_dREG.bsh \

{prefix}\_all-merge.plus.bw \

{prefix}\_all-merge.minus.bw \

{output prefix} \

{pre-trained SVM data} \

{number of threads}

bash writeBedv2.bsh \

{threshold} \ #on the Danko lab github, they recommend 0.8

{input bedgraph}

Outputs:

{prefix}.bedgraph (from run\_dREG.bsh)

{prefix}.bedgraph.bed (from writeBedv2.bsh)

Examples:

nohup bash /home/pr46\_0001/ChRO-seq/dREG/run\_dREG.bsh FLC\_all-merge.plus.bw FLC\_all-merge.minus.bw FLC /home/pr46\_0001/ChRO-seq/dREG-Model/asvm.mammal.RData 32 &> dREG.predScores.FLC.log.out &

nohup bash /home/pr46\_0001/ChRO-seq/dREG/run\_dREG.bsh NML\_all-merge.plus.bw NML\_all-merge.minus.bw NML /home/pr46\_0001/ChRO-seq/dREG-Model/asvm.mammal.RData 36 &> dREG.predScores.NML.log.out &

bash /home/pr46\_0001/ChRO-seq/dREG/writeBedv2.bsh 0.8 FLC.bedgraph

bash /home/pr46\_0001/ChRO-seq/dREG/writeBedv2.bsh 0.8 NML.bedgraph

Merge dREG-HD with **run\_dREG-HD.bsh**:

* dREG-HD uses a SVR model to impute DNaseI signal from dREG “peaks”

CBSU location: /home/pr46\_0001/ChRO-seq/dREG.HD/run\_dREG-HD.bsh

To run:

bash run\_dREG-HD.bsh \

{prefix}.bedgraph.bed \

{prefix}\_all-merge.plus.bw \

{prefix}\_all-merge.minus.bw \

{pre-trained dREG-HD SVM data} \ #different from dREG model

{number of threads}

Outputs:

{prefix}.bedgraph.bed\_dREG\_HD\_relaxed.bed (dREG-HD peaks, FDR=0.16, from Danko github)

{prefix}.bedgraph.bed\_dREG\_HD\_stringent.bed (dREG-HD peaks, FDR=0.1)

{prefix}.bedgraph.bed\_imputedDnase.bw

Examples:

nohup bash /home/pr46\_0001/ChRO-seq/dREG.HD/run\_dREG-HD.bsh FLC.bedgraph.bed FLC\_all-merge.plus.bw FLC\_all-merge.minus.bw /home/pr46\_0001/ChRO-seq/dREG-Model/dREG\_HD.model.rdata 36 &> dREG.HD.FLC.log.out & (outputs bed and bigwig files: FLC.bedgraph.bed\_dREG\_HD\_relaxed.bed, FLC.bedgraph.bed\_dREG\_HD\_stringent.bed, FLC.bedgraph.bed\_imputedDnase.bw)

nohup bash /home/pr46\_0001/ChRO-seq/dREG.HD/run\_dREG-HD.bsh NML.bedgraph.bed NML\_all-merge.plus.bw NML\_all-merge.minus.bw /home/pr46\_0001/ChRO-seq/dREG-Model/dREG\_HD.model.rdata 36 &> dREG.HD.NML.log.out & (outputs bed and bigwig files: NML.bedgraph.bed\_dREG\_HD\_relaxed.bed, NML.bedgraph.bed\_dREG\_HD\_stringent.bed, NML.bedgraph.bed\_imputedDnase.bw)