File S2: Code


```
module load bedtools
module load python/3.7.6
module load meme_suite
email='exampleEmail@test.com'
qsubbingH="qsub -V -b y -q 1-hour -m ae -M $email -pe threaded 8 -cwd"
### Threaded hour queue
qsubbing="qsub -V -b y -q 1-day -m ae -M $email -pe threaded 8 -cwd"
### Threaded hour queue
cLoc=`pwd`/
bws=${cLoc}bigwigs/
peaks=${cLoc}peaks/

#Download bigwigs/beds from browser
for i in `cat idList `; do wget ftp://ftp.ncbi.nlm.nih.gov/geo/series/
GSE159nnn/GSE159503/suppl/GSE159503_${i}.mm10.broadpeak.gz; done
#remember to unzip

#Download Hocomoco v11 motifs
#Hand make mouseTF.meme file containing TAL1 and CTCF core motifs

#Download optimal IDR peak beds ENCODE
#Rename for safety:
mv ENCFF972VNL.bed ENCFF972VNL_optimalTAL1.bed
mv ENCFF929RJU.bed ENCFF929RJU_optimalCTCF.bed

#Produce fasta files from peak beds
for i in ${peaks}*peak; do $qsubbingH fastaFromBed -name+ -fi ${cLoc}
mm10.fa -bed ${i} -fo ${i}.fa; done

#Run FIMO on all peak fastas
for i in *fa; do name=${i%%.*}; name1=${name#*_}; fimo --max-stored-
scores 10000000 --o ${name1}_pVal ../mouseTF.meme $i ; done


# fgrep TAL1 fimo.tsv | cut -f3 | sort -u | wc -l
#Get counts for each motif per sequence
for i in *pVal/fimo.tsv; do fgrep TAL1 $i | cut -f3 | sort -u | wc -
l ; done
for i in *pVal/fimo.tsv; do fgrep CTCF $i | cut -f3 | sort -u | wc -
l ; done
#Get names for each sequence with motif
for i in *pVal/fimo.tsv; do fgrep TAL1 $i | cut -f3 | sort -u > $
{i%.tsv}TAL1seq.names ; done
for i in *pVal/fimo.tsv; do fgrep CTCF $i | cut -f3 | sort -u > $
{i%.tsv}CTCFseq.names ; done

### Create bed files of sequences with motifs
```

```
for i in *pVal/*names; do sed 's/[:-]/\t/g' ${i} > ${i}.bed ; done

### Make fastas for optimal IDR hc peaks and test for motifs
for i in optimal*bed; do $qsubbingH fastaFromBed -name+ -fi ${cLoc}
mm10.fa -bed ${cLoc}$i -fo ${cLoc}${i%bed}fa ; done
for i in optimal*fa; do fimo --max-stored-scores 10000000 --o ${i%.fa}
mouseTF.meme ${i} ; done

### Create name files for optimal peak IDR
grep CTCF fimo.tsv | cut -f3 | sort -u > strongCTCF.txt
grep CTCF fimo.tsv | cut -f3 | cut -f1 -d: | sort -u >
strongCTCF_onlyNames.txt
grep '>' ../optimalCTCF_wNames.fa > allnames
fgrep -v -f test2 allnames > noMotif.txt

grep TAL1 fimo.tsv | cut -f3 | sort -u > strongTAL1.txt
grep TAL1 fimo.tsv | cut -f3 | cut -f1 -d: | sort -u >
strongTAL1_onlyNames.txt
grep '>' ../optimalTAL1_wNames.fa > allnames
sed 's/^/>/' strongTAL1_onlyNames.txt > test
sed 's/$/:/' test > test2
fgrep -v -f test2 allnames > noMotif.txt

### Create motif or no motif files for optimal IDR hc peaks
cut -f2 -d'>' noMotif.txt | cut -f1 -d: > noMotif_TAL1_onlyNames.txt
awk -F'\t' 'NR==FNR{c[$1]++;next};c[$4] > 0'
noMotif_TAL1_onlyNames.txt ../optimalTAL1_wNames.bed >
noMotif_TAL1.bed
awk -F'\t' 'NR==FNR{c[$1]++;next};c[$4] > 0'
strongTAL1_onlyNames.txt ../optimalTAL1_wNames.bed >
strongTAL1_onlyNames.bed


cut -f2 -d'>' noMotif.txt | cut -f1 -d: > noMotif_CTCF_onlyNames.txt
awk -F'\t' 'NR==FNR{c[$1]++;next};c[$4] > 0'
noMotif_CTCF_onlyNames.txt ../optimalCTCF_wNames.bed >
noMotif_CTCF.bed
awk -F'\t' 'NR==FNR{c[$1]++;next};c[$4] > 0'
strongCTCF_onlyNames.txt ../optimalCTCF_wNames.bed >
strongCTCF_onlyNames.bed



###getting peak stats for motif/no motif overlaps with optimal
for i in $tal1; do intersectBed -wa -a ../optimalTAL1_wNames/
strongTAL1_onlyNames.bed -b ${i}_pVal/fimoTAL1seq.names.bed > ${i}
_pVal/optimalOverlap_${i}_TAL1.bed; done
for i in $ctcf; do intersectBed -wa -a ../optimalCTCF_wNames/
strongCTCF_onlyNames.bed -b ${i}_pVal/fimoCTCFseq.names.bed > ${i}
_pVal/optimalOverlap_${i}_CTCF.bed; done
```

```
for i in $tal1; do intersectBed -wa -a ../optimalTAL1_wNames/
noMotif_TAL1.bed -b ${i}_pVal/fimoTAL1seq.names.bed > ${i}_pVal/
optimalOverlap_noTAL1motif_${i}_TAL1.bed; done
for i in $ctcf; do intersectBed -wa -a ../optimalCTCF_wNames/
noMotif_CTCF.bed -b ${i}_pVal/fimoCTCFseq.names.bed > ${i}_pVal/
optimalOverlap_noCTCFmotif_${i}_CTCF.bed; done

### TAL1 bigwigs
tal1='670
671
672
673
674
702
703
704
705
706
'

tal1BW=`for i in $tal1 ; do echo /research/labs/gastroenterology/
sjohnsen/m219097/motifAnalysis/bigwigs/meryAbTest${i}.bw; done `


### CTCF bigwigs
ctcf='692
693
694
695
696
732
733
734
735
736
'

ctcfBW=`for i in $ctcf ; do echo motifAnalysis/bigwigs/meryAbTest$
{i}.bw; done `

labels='3 5 7 10 15 3 5 7 10 15'

### create centered bed files
bedtools slop -i optimalTAL1_wNames/noMotif_TAL1.bed -g
mm10.chrom.sizes -pct -l -0.5 -r -0.5 > optimalTAL1_wNames/
noMotif_TAL1_centered.bed
bedtools slop -i optimalTAL1_wNames/strongTAL1_onlyNames.bed -g
mm10.chrom.sizes -pct -l -0.5 -r -0.5 > optimalTAL1_wNames/
strongTAL1_onlyNames_centered.bed
```

```
bedtools slop -i optimalCTCF_wNames/noMotif_CTCF.bed -g
mm10.chrom.sizes -pct -l -0.5 -r -0.5 > optimalCTCF_wNames/
noMotif_CTCF_centered.bed
bedtools slop -i optimalCTCF_wNames/strongCTCF_onlyNames.bed -g
mm10.chrom.sizes -pct -l -0.5 -r -0.5 > optimalCTCF_wNames/
strongCTCF_onlyNames_centered.bed

### Deeptools visualization using centered beds
$qsubbingH computeMatrix reference-point -S $ctcfBW -R ${cLoc}
optimalCTCF_wNames/strongCTCF_onlyNames_centered.bed ${cLoc}
optimalCTCF_wNames/noMotif_CTCF_centered.bed -a 1000 -b 1000 -o $
{cLoc}CTCF_motif_centered.matrix.gz

$qsubbingH plotHeatmap -m ${cLoc}CTCF_motif_centered.matrix.gz --
refPointLabel "Peak" -out
CTCF_optimalPeakComparison_motif_Blues_centered_022421.pdf --
samplesLabel $ctcf --colorMap 'Blues' --missingDataColor white -z
"Motif" "No\ motif" --whatToShow "heatmap\ and\ colorbar" -x
"Distance\ \(bp\)"

$qsubbingH plotProfile -m ${cLoc}CTCF_motif_centered.matrix.gz --
refPointLabel "Peak" -out
CTCF_optimalPeakProfile_motif_matchedColors_centered_030121.pdf --
samplesLabel $labels -z "Motif" "No\ motif" --colors '\#0300b0'
'\#6dddfc' --numPlotsPerRow 10 --plotType fill --plotWidth 4


$qsubbingH computeMatrix reference-point -S $tal1BW -R ${cLoc}
optimalTAL1_wNames/strongTAL1_onlyNames_centered.bed ${cLoc}
optimalTAL1_wNames/noMotif_TAL1_centered.bed  -a 1000 -b 1000 -o $
{cLoc}TAL1_motif_centered.matrix.gz

$qsubbingH plotHeatmap -m ${cLoc}TAL1_motif_centered.matrix.gz --
refPointLabel "Peak" -out
TAL1_optimalPeakComparison_motif_Oranges_centered_022421.pdf --
samplesLabel $tal1 --colorMap 'Oranges' --missingDataColor white -z
"Motif" "No\ motif" --whatToShow "heatmap\ and\ colorbar" -x
"Distance\ \(bp\)"

$qsubbingH plotProfile -m ${cLoc}TAL1_motif_centered.matrix.gz --
refPointLabel "Peak" -out
TAL1_optimalPeakProfile_motif_matchedColors_centered_030121.pdf --
samplesLabel $labels -z "Motif" "No\ motif" --colors '\#c45400'
'\#FCDCC4' --numPlotsPerRow 10 --plotType fill --plotWidth 4
```