

Gurdon Bioinformatics Project proposal

GBP Id: GBP0030

Title: Analysis of CRAMP1 KO ATAC-seq data

Collaboration agreement:

The aim of this document is to define the scope of a project, the roles of collaborating partners and to clarify expectations for acknowledgement.

Where results have been used in, or advice has been instrumental to a piece of work the minimum expectation is for written acknowledgement of the Gurdon Institute Bioinformatics group and the analyst(s) involved in associated publications and presentations. If scientists from the team contributed more than routine techniques or advice, they should be co-authors of any publications. The scientists involved should be consulted on any manuscripts prior to publication to ensure all materials, methods and data analyses are correctly described.

Abstract (1. Background - why we are interested/wider project details, 2. what we hope to find with this GBP, 3. brief description of work i.e. how we are going to find it out):

This project is a continuation of the CRAMP1 project from Iva Lab. They want comparisons between shCRAMP1 and shControl, and also between shSuz12 and shControl. If there are significant changes in these, then they want to compare loci changing in shCRAMP1 and shSuz12 to establish the overlap. Next this data needs to be integrated with other NGS datasets like RNA-Seq, Cut&Run and Cut&Tag, which is also shared as a part of this project.

Group Leader: Iva Tchasonnikarova

Stakeholders:

Stakeholder name	Email	Role	Objectives, interests	Influence (high, medium, low)	Contribution
Rachael Matthews					

Primary analyst: Ankit Verma/Adam Reid

Start date: September 2023

End date:

GANTT chart:

Project Goal:

- Basic ATAC-Seq analysis
- Identifying Differential accessible regions (DAC)
- Annotate DAC
- Exploratory analysis of ATAC-Seq
- Compare shCRAMP1 and shSUZ12

Project Objectives:

- Correlation of differentially accessible regions with gene expressions and histone mark profile

Detailed description:

Notes:

- Initial sample sheet -  [atacseq_samplesheet_iva_lab](#)

Task list:

Linked documents:

E.g.

- Notes
-

Directories/filestores (e.g. on cluster):

- Raw data on cluster -
- [#ATAC-seq](#) KD
- Processed data: /mnt/home3/reid/av638/atacseq/iva_lab_gencode/boutfolder

#Cut&Run ([KO](#) and [KD](#))

Processed data: /mnt/home3/reid/av638/cutnrun/iva_lab_oct23/outfolder

[RNA-Seq](#) KD

Processed data: /mnt/home3/reid/av638/rnaseq/iva_lab_oct23/outfolder

[Cut&Tag](#) WT

Processed data: /mnt/home3/reid/av638/cutntag/iva_lab_oct23/outfolder

Cut&Tag KD and WT and FLAG

Processed data: /mnt/home3/reid/av638/cutntag/iva_lab_dec23/outfolder/

Notes: ATAC-Seq

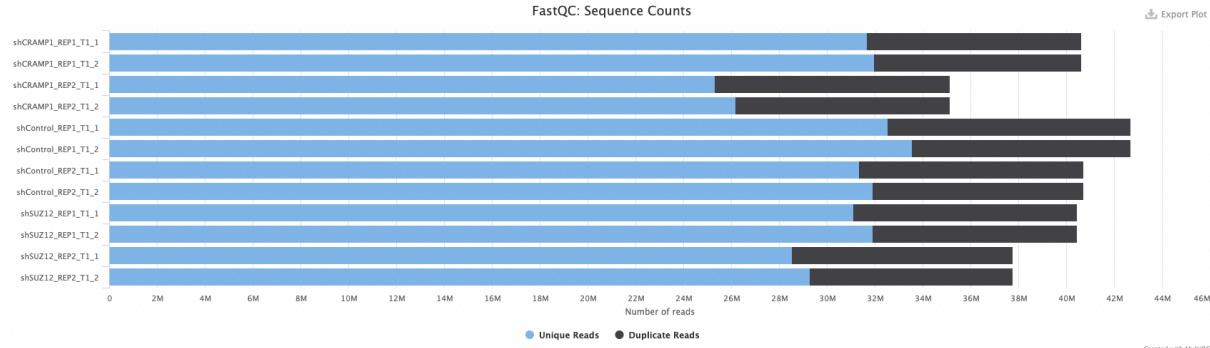
- ATAC-Seq Data was downloaded from Onedrive shared by Iva. The fastq.gz files were downloaded one by one as the zipped downloaded files were not the right zip format.

sample	fastq_1	fastq_2	replicate	control	control_replicate
shControl	1509-1_R1_001.fastq.gz	1509-1_R2_001.fastq.gz	1		
shControl	1509-2_R1_001.fastq.gz	1509-2_R2_001.fastq.gz	2		
shCRAMP1	shCR1-1_R1_001.fastq.gz	shCR1-1_R2_001.fastq.gz	1	shControl	1
shCRAMP1	shCR1-2_R1_001.fastq.gz	shCR1-2_R2_001.fastq.gz	2	shControl	2
shSUZ12	shSUZ-1_R1_001.fastq.gz	shSUZ-1_R2_001.fastq.gz	1	shControl	1
shSUZ12	shSUZ-2_R1_001.fastq.gz	shSUZ-2_R2_001.fastq.gz	2	shControl	2

2. Data was processed using Nextflow (v23): nf-core/atacseq v2.1.2

```
#using bowtie2 as aligner
/mnt/home3/slurm/slurm_sub.py nextflow run nf-core/atacseq -r 2.1.2 --profile singularity --input samplesheet.csv --outdir boutfolder --fasta GRCh38.p13.genome.fa --gtf GRCh38.gencode.v41.chr_patch_hapl_scaff.annotation.gtf --read_length 150 --blacklist ENCF356LFX.bed --mito_name chrM --save_reference -c /mnt/home3/nextflow/gurdon.config --aligner bowtie2
```

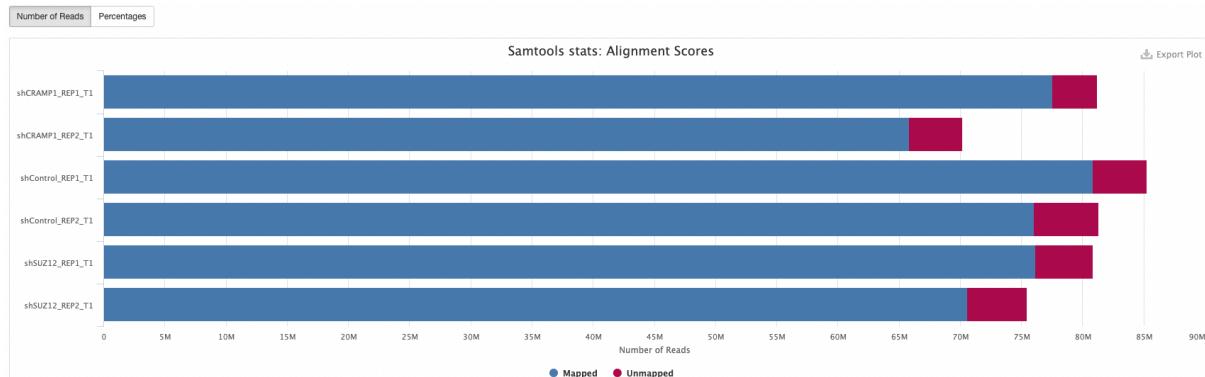
Raw read counts



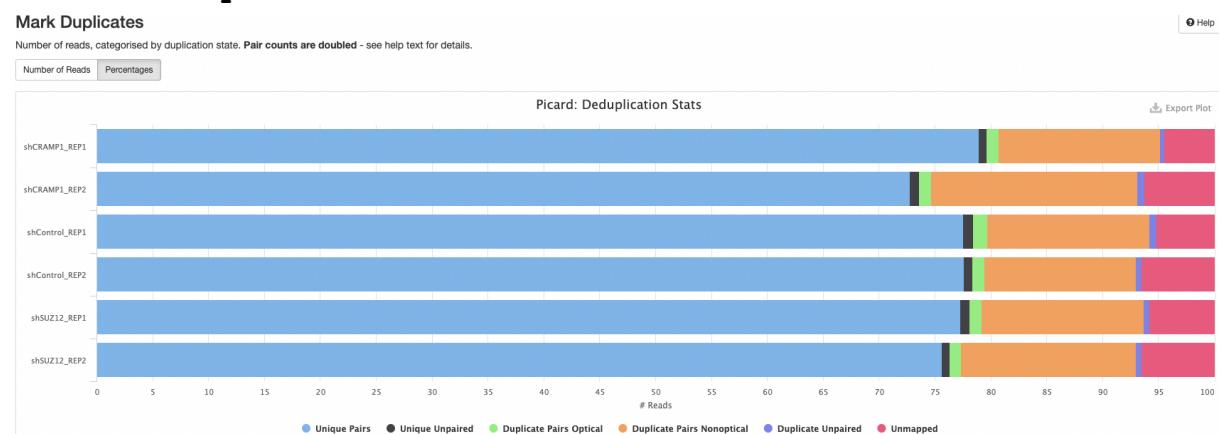
Reads Quality



Mapped

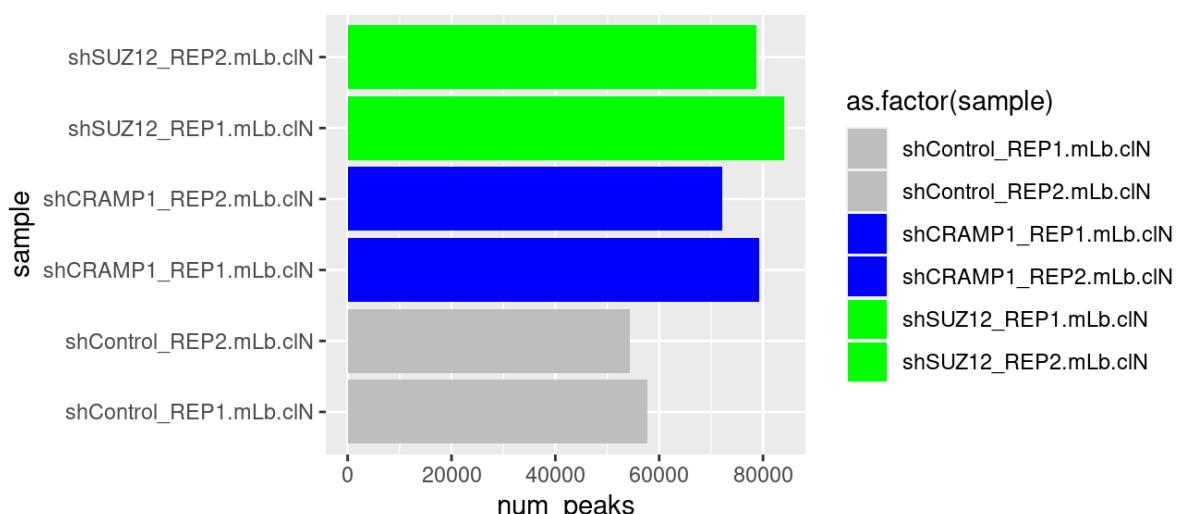


Marked duplicates



3. Downstream analysis is being performed in R

4. Peak count, nextflow output



5. Matrix Heatmap: DeepTools,

refTSS:https://reftss.riken.jp/datafiles/current/human/refTSS_v4.1_human_coordinate.hg38.bed.gz

```
grep start refTSS_v4.1_human_coordinate.hg38.bed.txt -v | awk
'{print $1"\t"$2"\t"$3}' >
refTSS_v4.1_human_coordinate.hg38.bed_chr.txt
```

```
computeMatrix reference-point -S shControl.mRp.clN.bigWig  
shCRAMP1.mRp.clN.bigWig shSUZ12.mRp.clN.bigWig -R  
refTSS_v4.1_human_coordinate.hg38.bed_chr.txt -o  
AroundTSSmatrix --outFileNameMatrix AroundTSSmatrix.txt  
--outFileSortedRegions AroundTSSmatrix_nosort.txt -a 3000  
-b 3000 -bs 25 -p 15 --missingDataAsZero
```

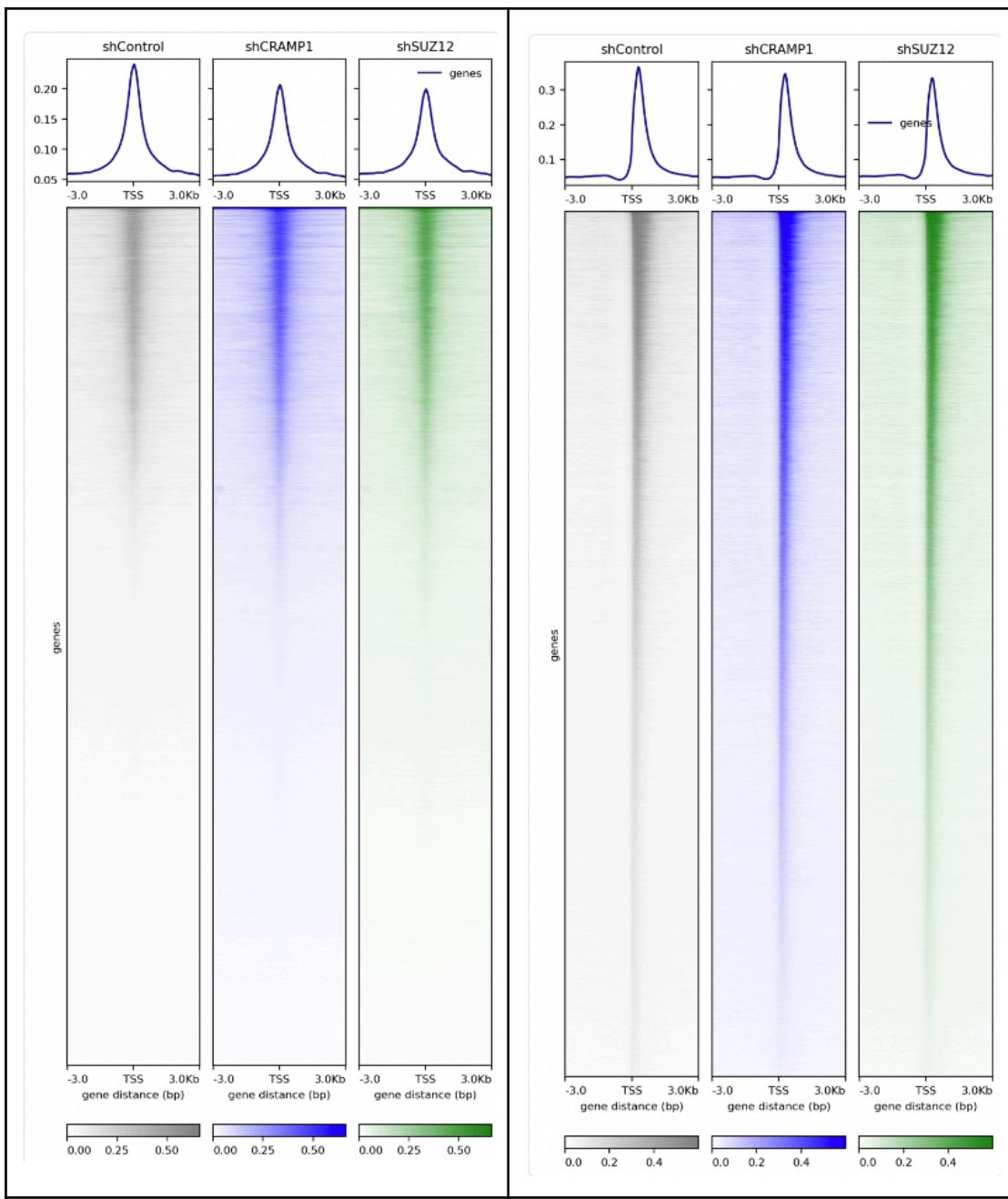
```
plotHeatmap -m AroundTSSmatrix --colorList "white, grey"  
"white, blue" "white, green" -out  
AroundTSS_computeMatrix1.png --sortUsing max  
--samplesLabel shControl shCRAMP1 shSUZ12
```

```
computeMatrix reference-point -S shControl.mRp.clN.bigWig  
shCRAMP1.mRp.clN.bigWig shSUZ12.mRp.clN.bigWig -R  
shControl.mRp.clN_peaks.broadPeak -o  
AroundshControlmatrix --outFileNameMatrix  
AroundshControlmatrix.txt --outFileSortedRegions  
AroundshControlmatrix_nosort.txt -a 3000 -b 3000 -bs 25  
-p 15 --missingDataAsZero
```

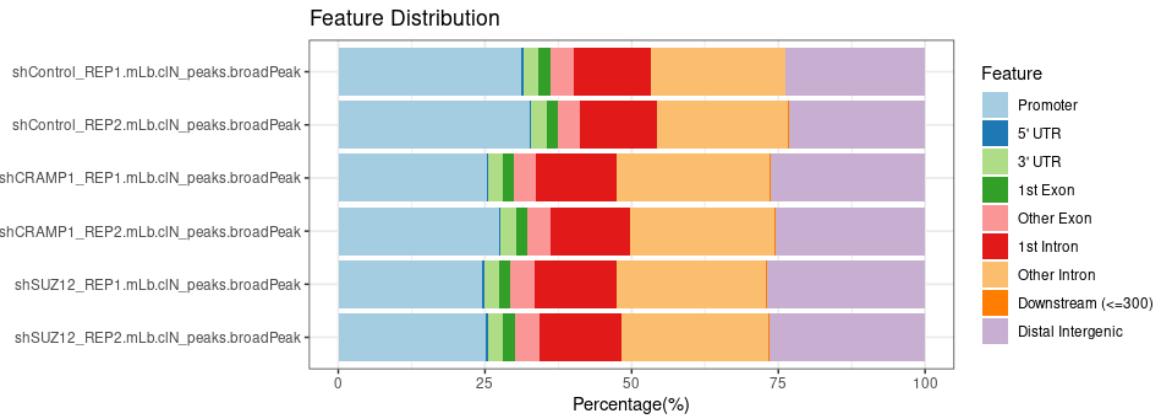
```
plotHeatmap -m AroundshControlmatrix --colorList  
"white, grey" "white, blue" "white, green" -out  
AroundshControl_computeMatrix1.png --sortUsing max  
--samplesLabel shControl shCRAMP1 shSUZ12
```

Around TSS

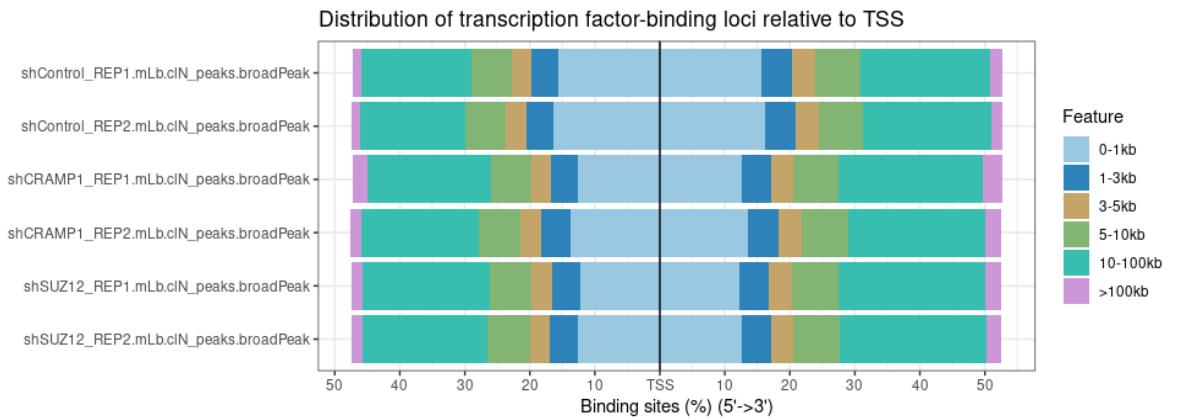
Around shControl peaks



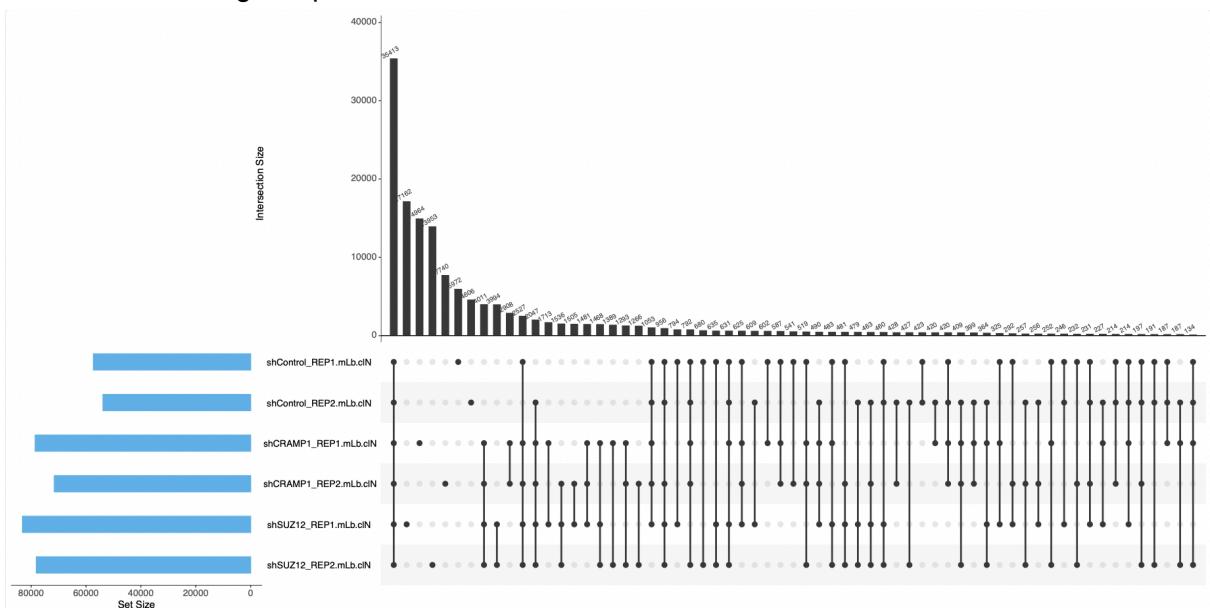
6. Annotation to genomic features



7. Annotation to TSS



8. Peak shared among samples



9. Motif

#meme

```
bedtools getfasta -fi
/mnt/home3/reid/av638/atacseq/iva_lab_gencode/GRCh38.p13.genome.fa
-bed shControl.mRp.cLN_peaks.broadPeak -fo shControl.mRp.cLN_peaks.fa
```

#Cluster

```

/opt/meme-5.5.0/bin/meme-chip -meme-nmotifs 5
shControl.mRp.clN_peaks.fa -o shControl

#meme
bedtools getfasta -fi
/mnt/home3/reid/av638/atacseq/iva_lab_gencode/GRCh38.p13.genome.fa
-bed shCRAMP1.mRp.clN_peaks.broadPeak -fo shCRAMP1.mRp.clN_peaks.fa

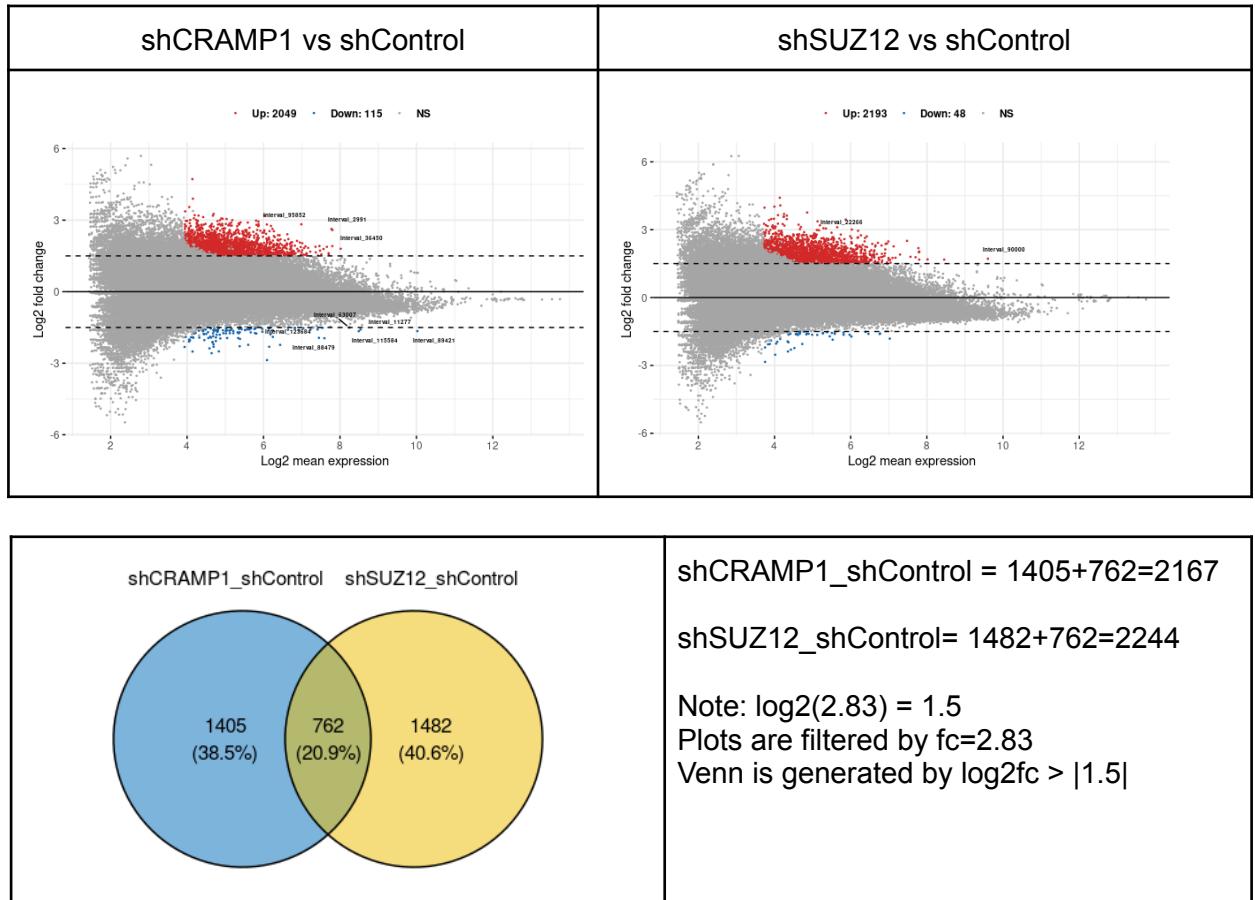
#Cluster
/opt/meme-5.5.0/bin/meme-chip -meme-nmotifs 5
shCRAMP1.mRp.clN_peaks.fa -o shCRAMP1

#meme
bedtools getfasta -fi
/mnt/home3/reid/av638/atacseq/iva_lab_gencode/GRCh38.p13.genome.fa
-bed shSUZ12.mRp.clN_peaks.broadPeak -fo shSUZ12.mRp.clN_peaks.fa

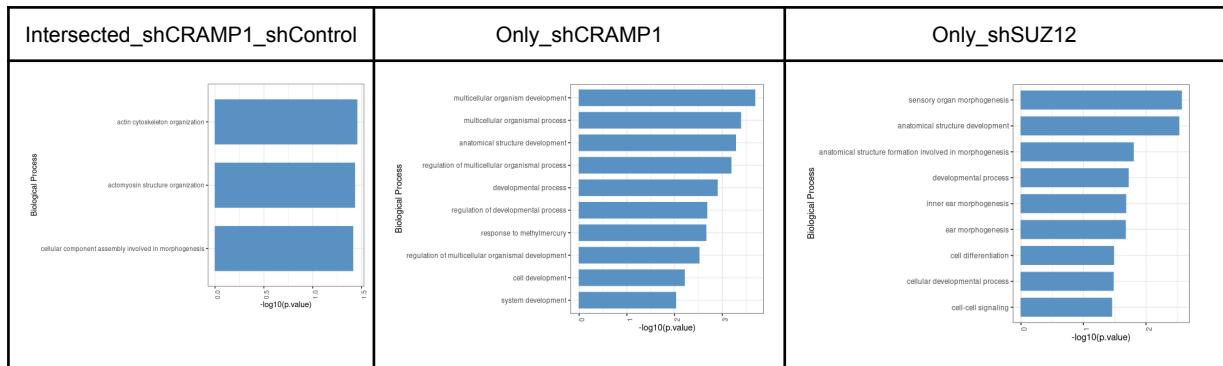
#Cluster
/opt/meme-5.5.0/bin/meme-chip -meme-nmotifs 5
shSUZ12.mRp.clN_peaks.fa -o shSUZ12

```

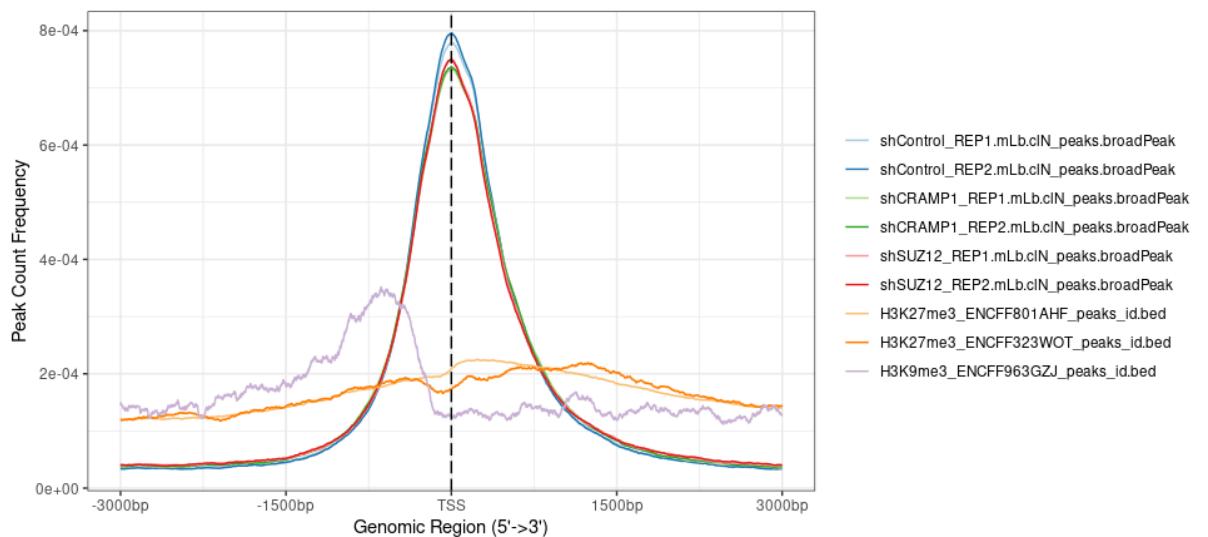
10. DAC analysis ($\log_{2}FC > |1.5|$, $p\text{-adj value} < 0.05$): DESeq2



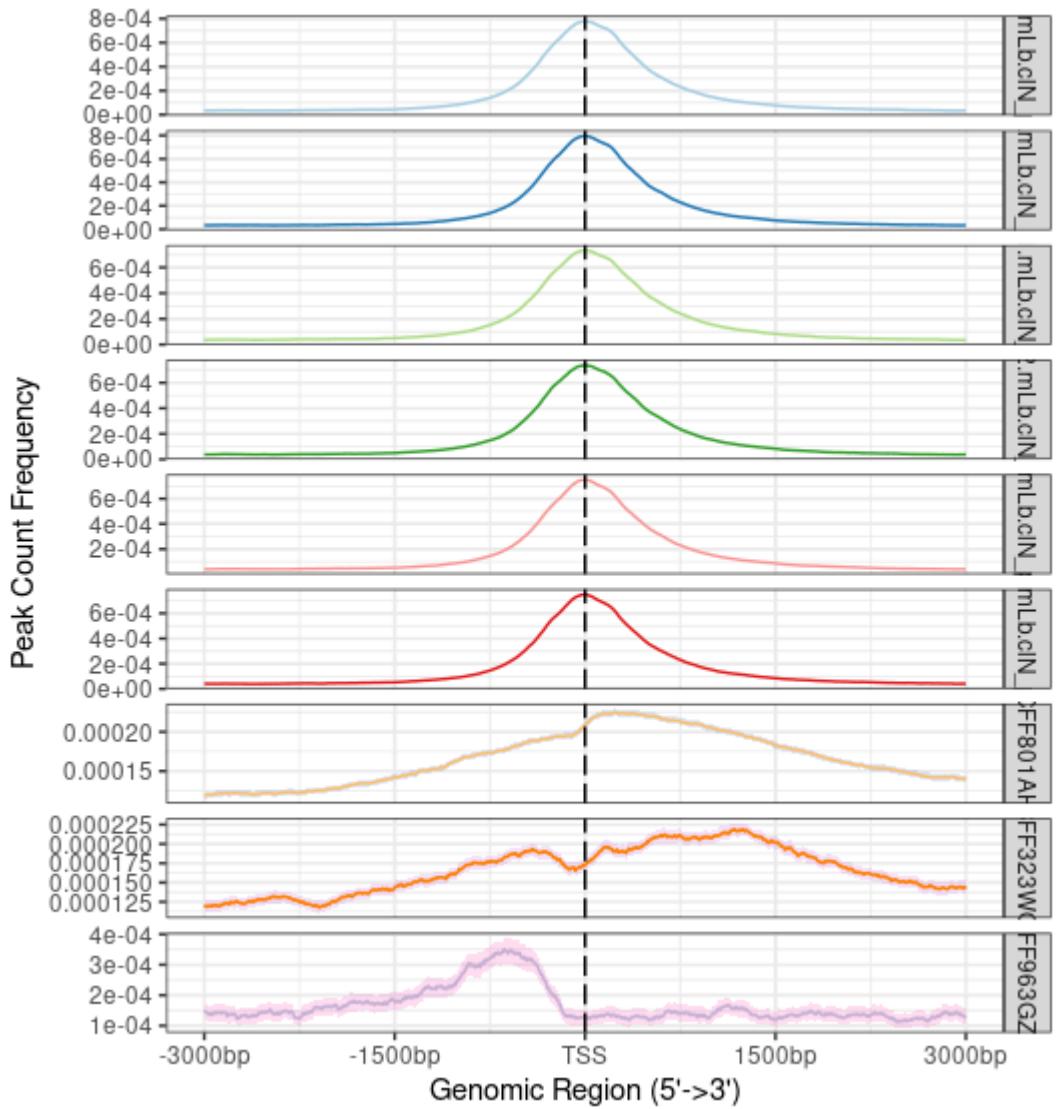
11. GO:BP TERM analysis



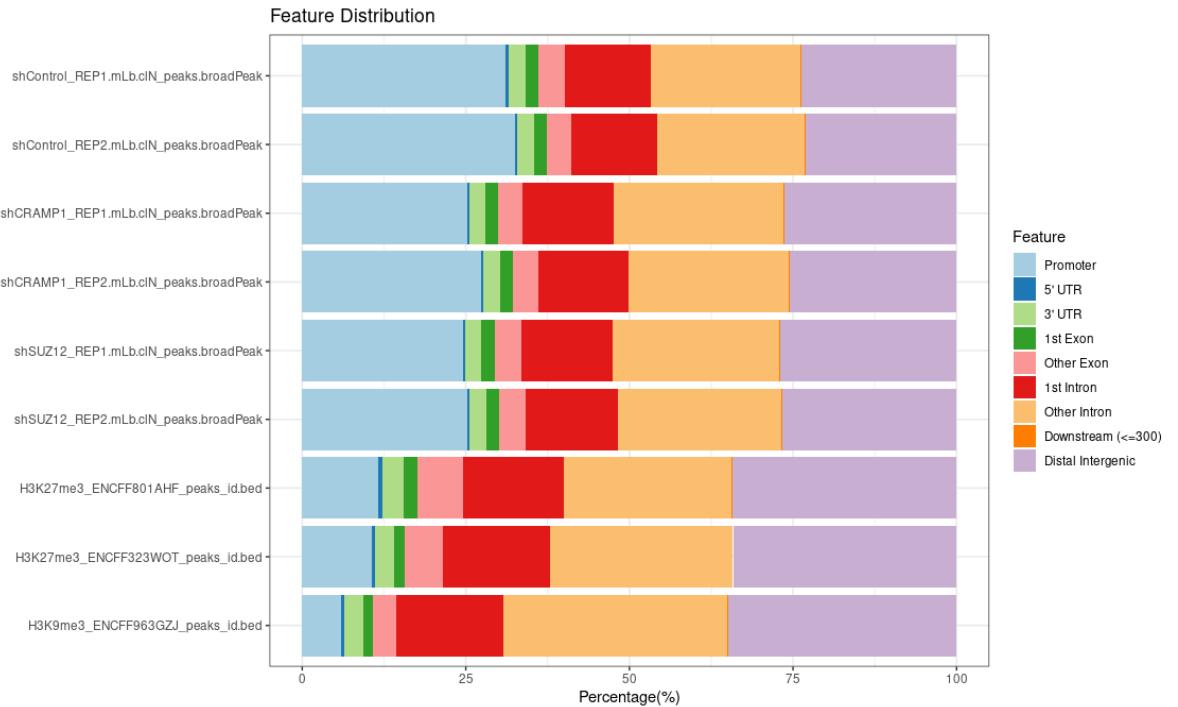
12. Average Profile: ChIP-Seq Peaks of H3K27me3 and H3K9me3 were downloaded from ENCODE



13. Average plot individual samples



14. Distribution of peaks around genomic features



15. CUT and RUN KO:

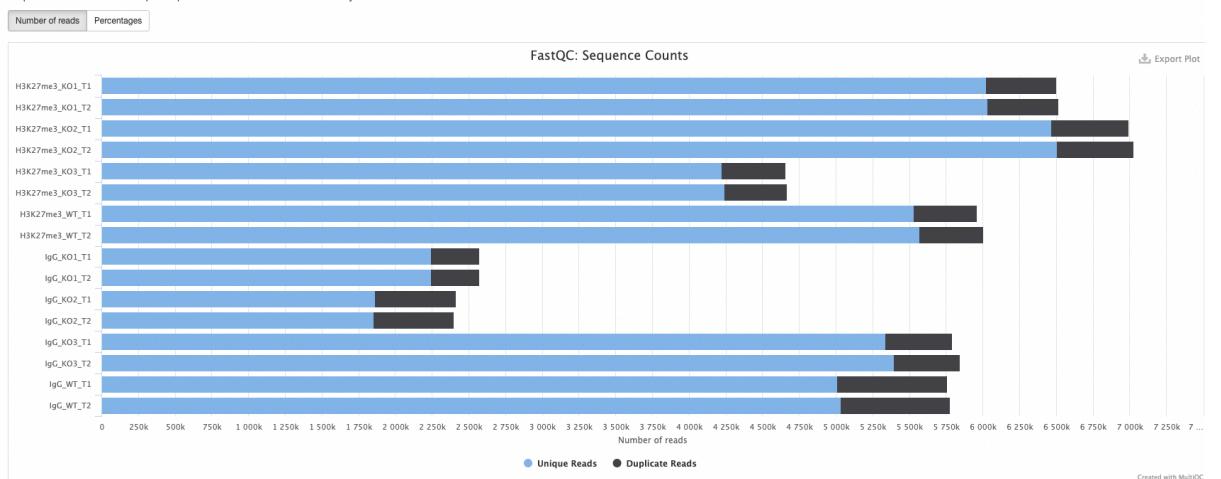
#samplesheet

sample	fastq_1	fastq_2	antibody	control	group
H3K27me3_KO3	ITit257-RM10_S34_L001_R1_001.fastq.gz		H3K27me3	IgG_KO3	H3K27me3_KO3
H3K27me3_KO3	ITit257-RM10_S34_L002_R1_001.fastq.gz		H3K27me3	IgG_KO3	H3K27me3_KO3
IgG_WT	ITit257-RM15_S39_L001_R1_001.fastq.gz				IgG_WT
IgG_WT	ITit257-RM15_S39_L002_R1_001.fastq.gz				IgG_WT
IgG_KO1	ITit257-RM16_S40_L001_R1_001.fastq.gz				IgG_KO1
IgG_KO1	ITit257-RM16_S40_L002_R1_001.fastq.gz				IgG_KO1
IgG_KO2	ITit257-RM17_S41_L001_R1_001.fastq.gz				IgG_KO2
IgG_KO2	ITit257-RM17_S41_L002_R1_001.fastq.gz				IgG_KO2
IgG_KO3	ITit257-RM18_S42_L001_R1_001.fastq.gz				IgG_KO3
IgG_KO3	ITit257-RM18_S42_L002_R1_001.fastq.gz				IgG_KO3
H3K27me3_WT	ITit257-RM7_S31_L001_R1_001.fastq.gz		H3K27me3	IgG_WT	H3K27me3_WT
H3K27me3_WT	ITit257-RM7_S31_L002_R1_001.fastq.gz		H3K27me3	IgG_WT	H3K27me3_WT
H3K27me3_KO1	ITit257-RM8_S32_L001_R1_001.fastq.gz		H3K27me3	IgG_KO1	H3K27me3_KO1
H3K27me3_KO1	ITit257-RM8_S32_L002_R1_001.fastq.gz		H3K27me3	IgG_KO1	H3K27me3_KO1
H3K27me3_KO2	ITit257-RM9_S33_L001_R1_001.fastq.gz		H3K27me3	IgG_KO2	H3K27me3_KO2
H3K27me3_KO2	ITit257-RM9_S33_L002_R1_001.fastq.gz		H3K27me3	IgG_KO2	H3K27me3_KO2

#FASTQC

Sequence Counts

Sequence counts for each sample. Duplicate read counts are an estimate only.

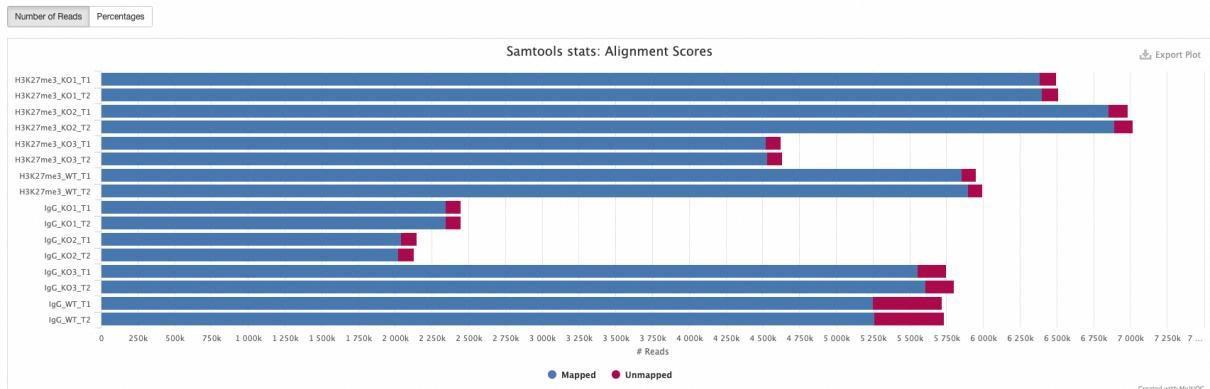




#Mapping

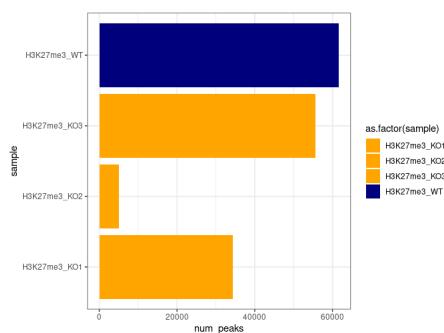
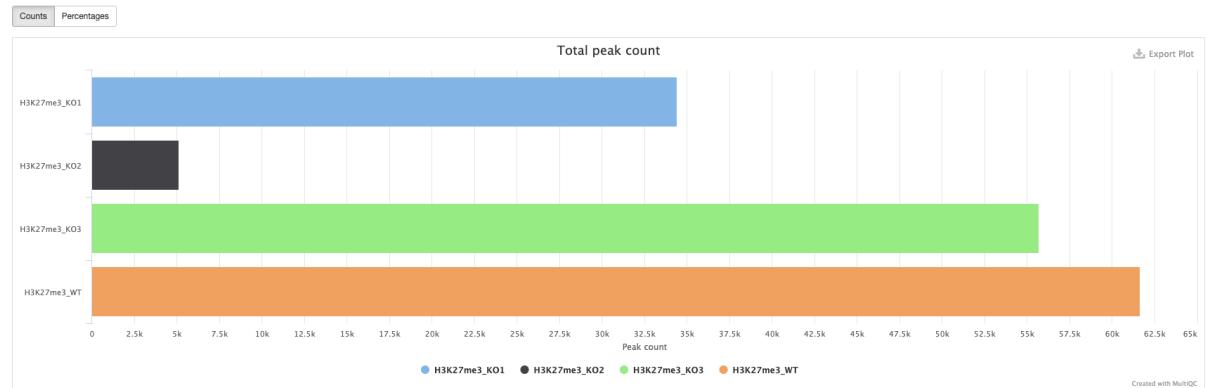
Percent Mapped

Alignment metrics from `samtools stats`; mapped vs. unmapped reads.

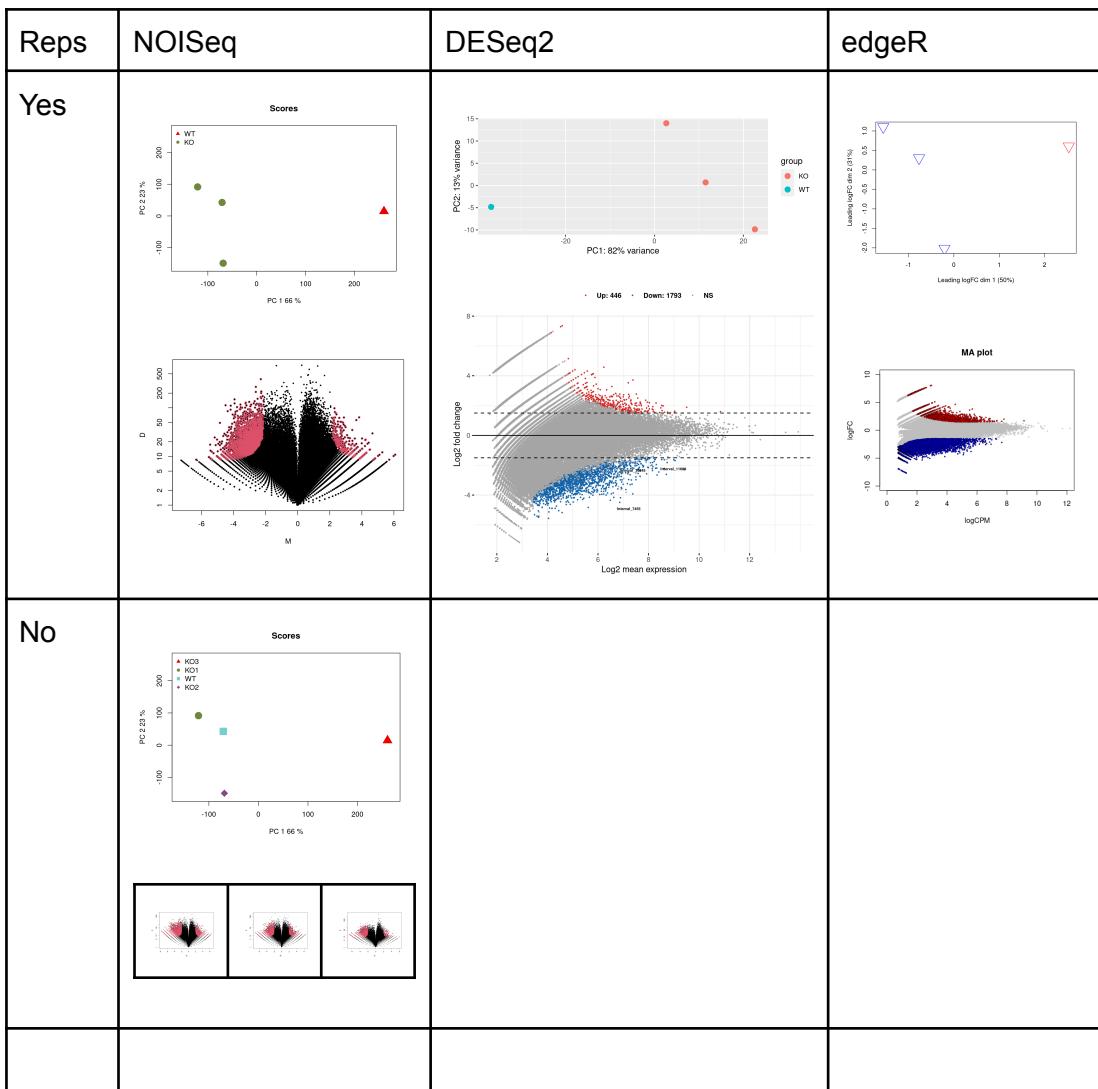


MERGED LIB: MACS2 peak count

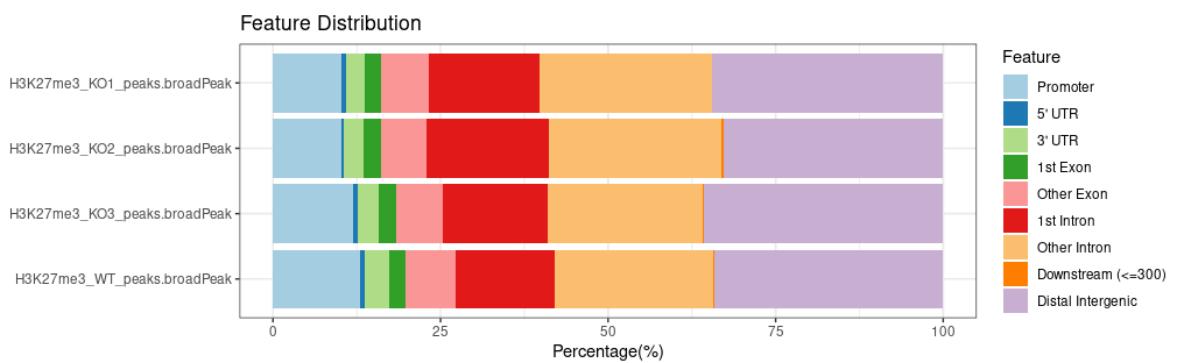
is calculated from total number of peaks called by [MACS2](#).



DE analysis PCA



Annotation



#

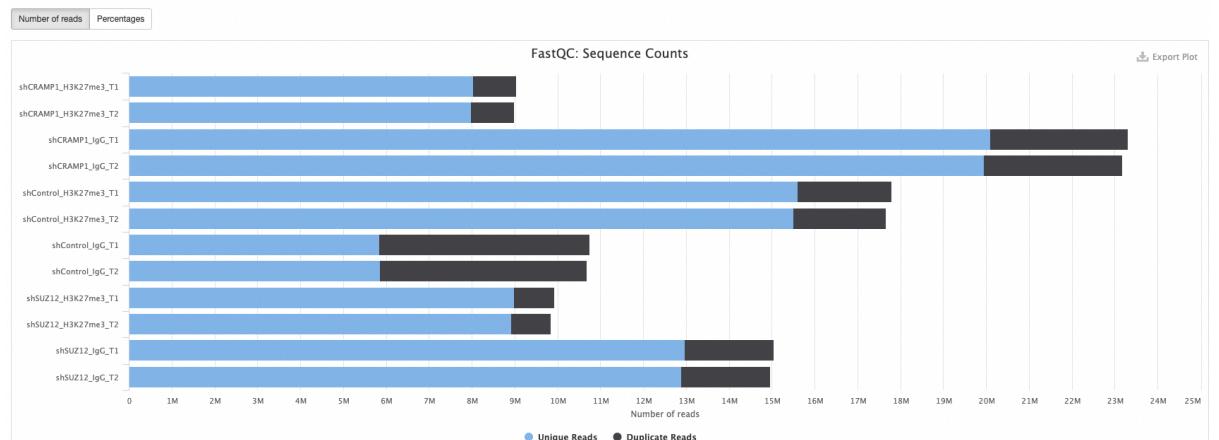
16. CUT and RUN KD: #samplesheet

sample	fastq_1	fastq_2	antibody	control
shControl_IgG	ITit257-shCon-IgG_S10_L001_R1_001.fastq.gz			
shControl_IgG	ITit257-shCon-IgG_S10_L002_R1_001.fastq.gz			
shControl_H3K27me3	ITit257-shCon-K27_S9_L001_R1_001.fastq.gz		H3K27me3	shControl_IgG
shControl_H3K27me3	ITit257-shCon-K27_S9_L002_R1_001.fastq.gz		H3K27me3	shControl_IgG
shCRAMP1_IgG	ITit257-shCR1-IgG_S18_L001_R1_001.fastq.gz			
shCRAMP1_IgG	ITit257-shCR1-IgG_S18_L002_R1_001.fastq.gz			
shCRAMP1_H3K27me3	ITit257-shCR1-K27_S17_L001_R1_001.fastq.gz		H3K27me3	shCRAMP1_IgG
shCRAMP1_H3K27me3	ITit257-shCR1-K27_S17_L002_R1_001.fastq.gz		H3K27me3	shCRAMP1_IgG
shSUZ12_IgG	ITit257-shSUZ-IgG_S14_L001_R1_001.fastq.gz			
shSUZ12_IgG	ITit257-shSUZ-IgG_S14_L002_R1_001.fastq.gz			
shSUZ12_H3K27me3	ITit257-shSUZ-K27_S13_L001_R1_001.fastq.gz		H3K27me3	shSUZ12_IgG
shSUZ12_H3K27me3	ITit257-shSUZ-K27_S13_L002_R1_001.fastq.gz		H3K27me3	shSUZ12_IgG

#FASTQC

Sequence Counts

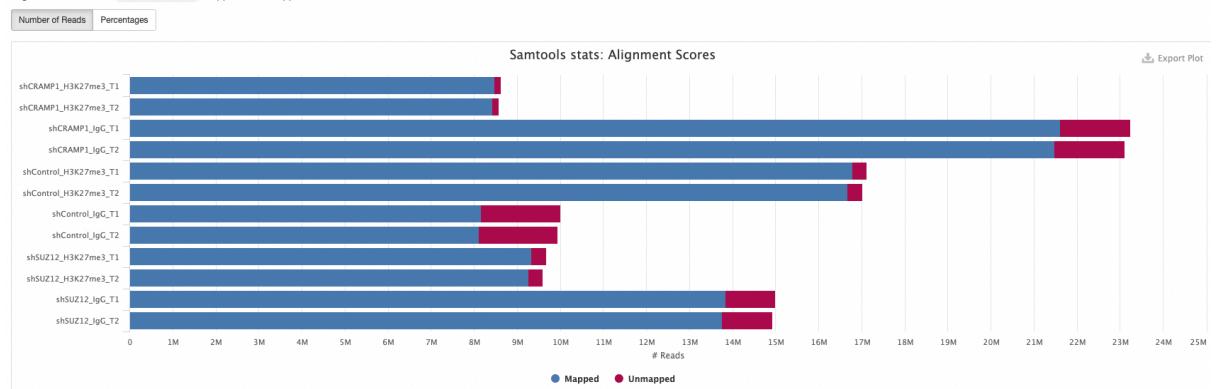
Sequence counts for each sample. Duplicate read counts are an estimate only.



#Mapping

Percent Mapped

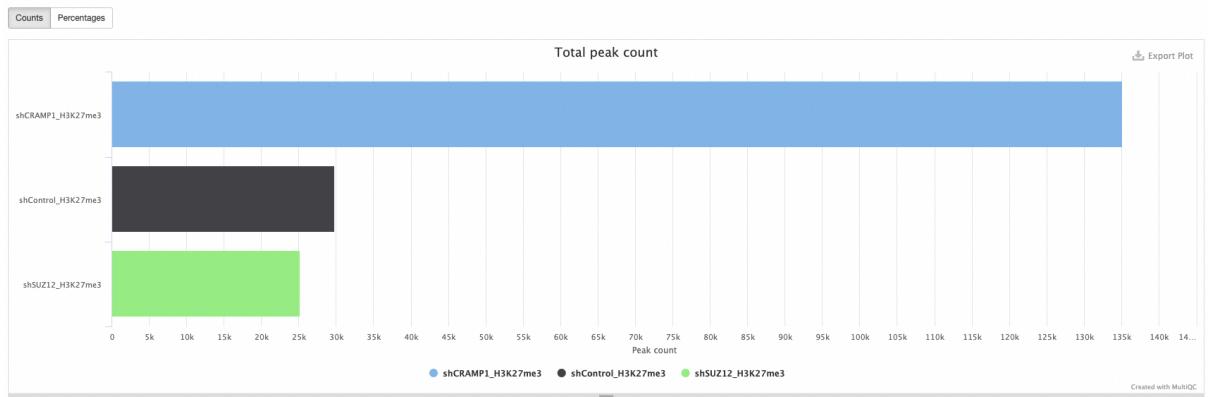
Alignment metrics from `samtools stats`; mapped vs. unmapped reads.



#MACS peaks

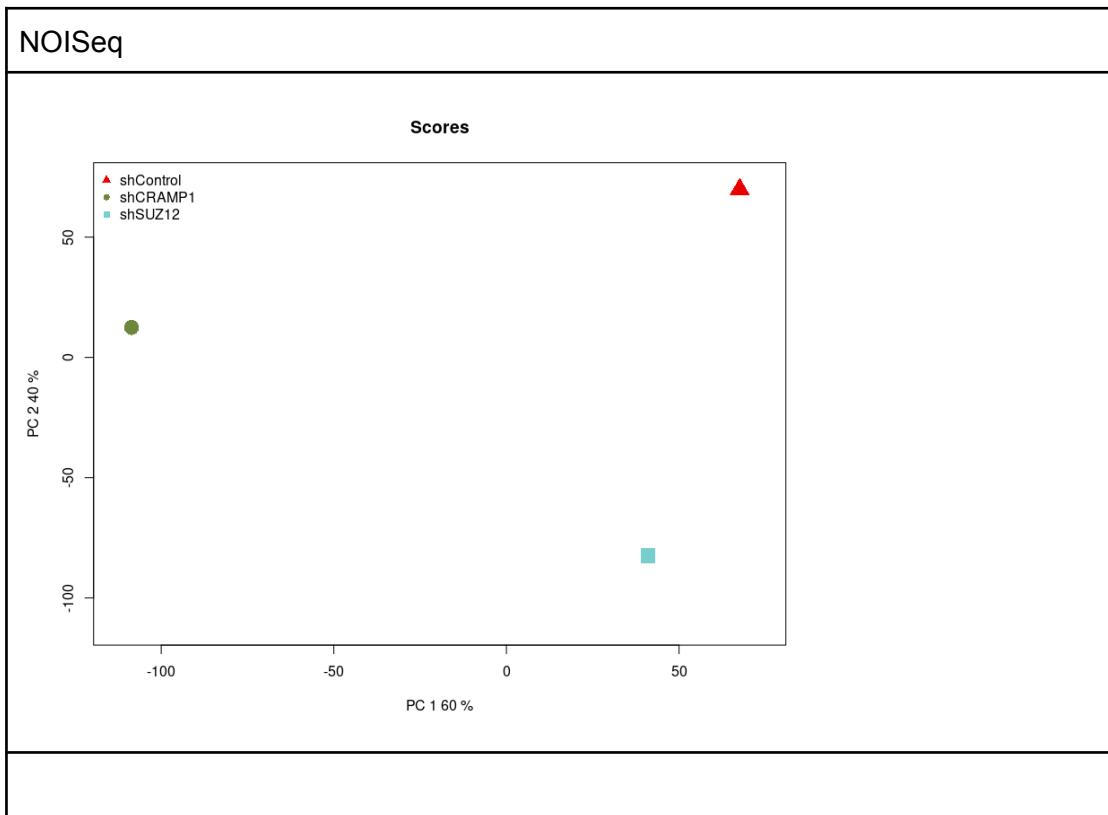
MERGED LIB: MACS2 peak count

is calculated from total number of peaks called by [MACS2](#).



#DE analysis:

PCA



17. CUT and TAG

#samplesheet

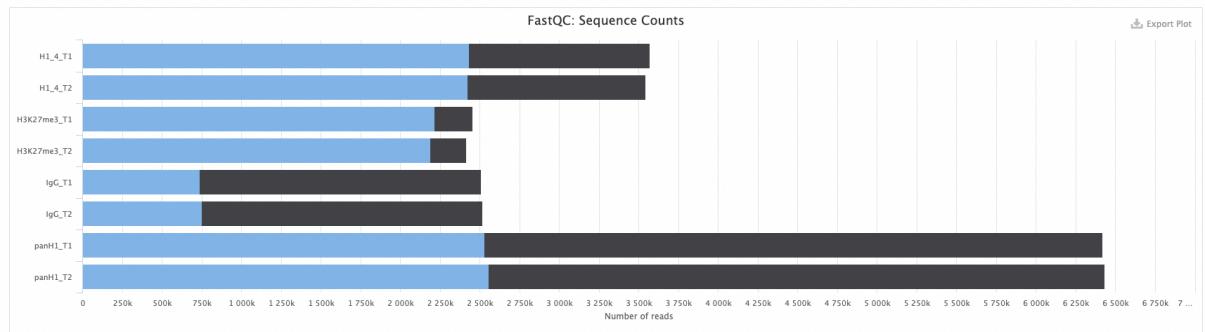
sample	fastq_1	fastq_2	antibody	control
H1_4	ITit257-RM1_S1_L001_R1_001.fastq.gz		H1.4	IgG
H1_4	ITit257-RM1_S1_L002_R1_001.fastq.gz		H1.4	IgG
panH1	ITit257-RM2_S2_L001_R1_001.fastq.gz		H1	IgG
panH1	ITit257-RM2_S2_L002_R1_001.fastq.gz		H1	IgG
H3K27me3	ITit257-RM3_S3_L001_R1_001.fastq.gz		H3K27me3	IgG
H3K27me3	ITit257-RM3_S3_L002_R1_001.fastq.gz		H3K27me3	IgG
IgG	ITit257-RM4_S4_L001_R1_001.fastq.gz			
IgG	ITit257-RM4_S4_L002_R1_001.fastq.gz			

#FASTQC

Sequence Counts

Sequence counts for each sample. Duplicate read counts are an estimate only.

Number of reads Percentages

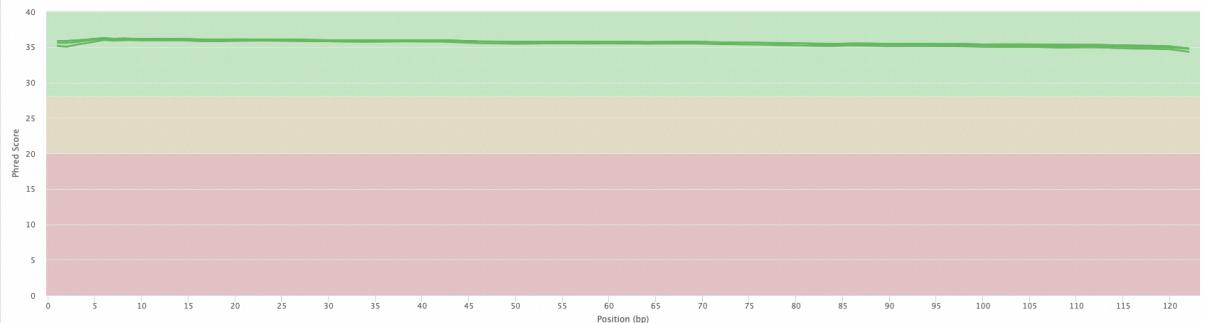


Help Export Plot

Created with MultiQC

FastQC: Mean Quality Scores

Export Plot



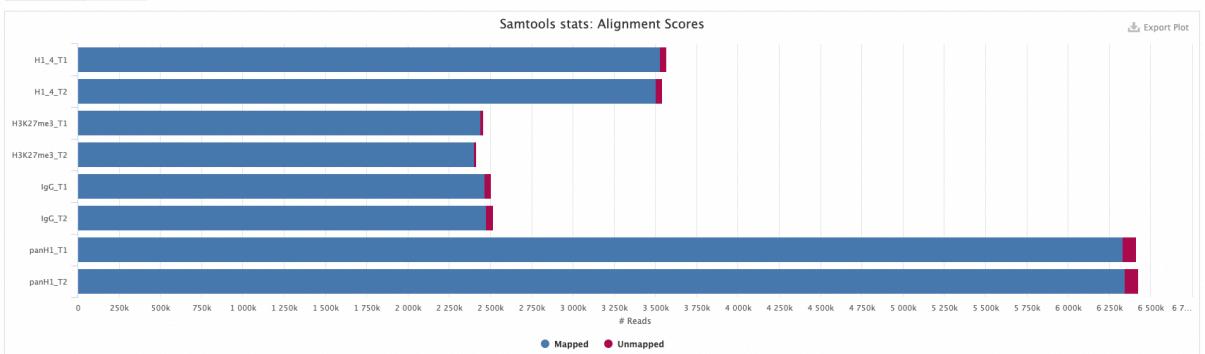
Created with MultiQC

#Mapping

Percent Mapped

Alignment metrics from samtools stats; mapped vs. unmapped reads.

Number of Reads Percentages



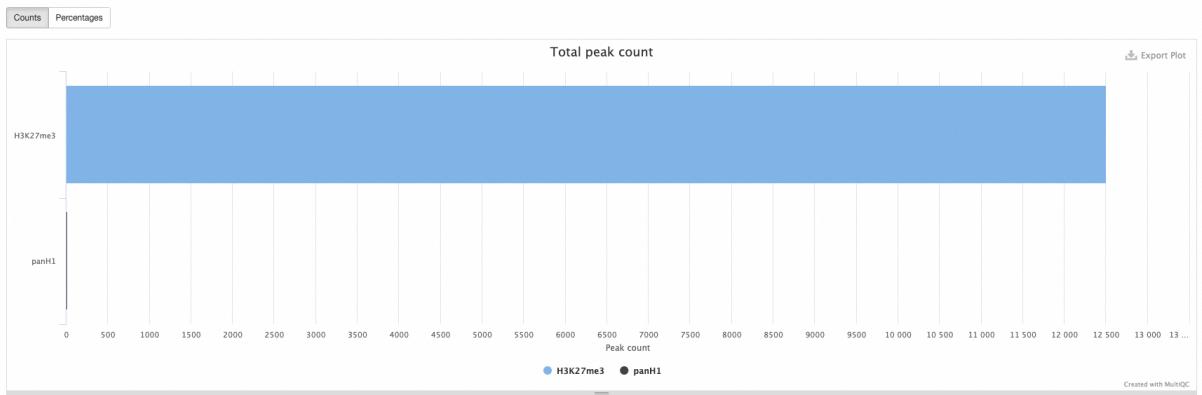
Help Export Plot

Created with MultiQC

#MACS peak call

MERGED LIB: MACS2 peak count

is calculated from total number of peaks called by [MACS2](#).



#Quantification

```

awk    '{print  $1"\t"$2"\t"$3}'    H3K27me3_ENCFF801AHF_peaks_id.bed    >
H3K27me3_ENCFF801AHF_regions.bed

computeMatrix      scale-regions      -S      K27me3_ENCFF847BFA_Peggy.bigWig
H3K27me3.bigWig      H1_4.bigWig      panH1.bigWig      IgG.bigWig
K9me3_Bernstein.bigWig      -R      H3K27me3_ENCFF801AHF_regions.bed      -o
AroundK27matrix      --outFileNameMatrix      AroundK27matrix.txt
--outFileSortedRegions AroundK27matrix_nosort.txt -a 3000 -b 3000 -bs
25 -p 15 --missingDataAsZero

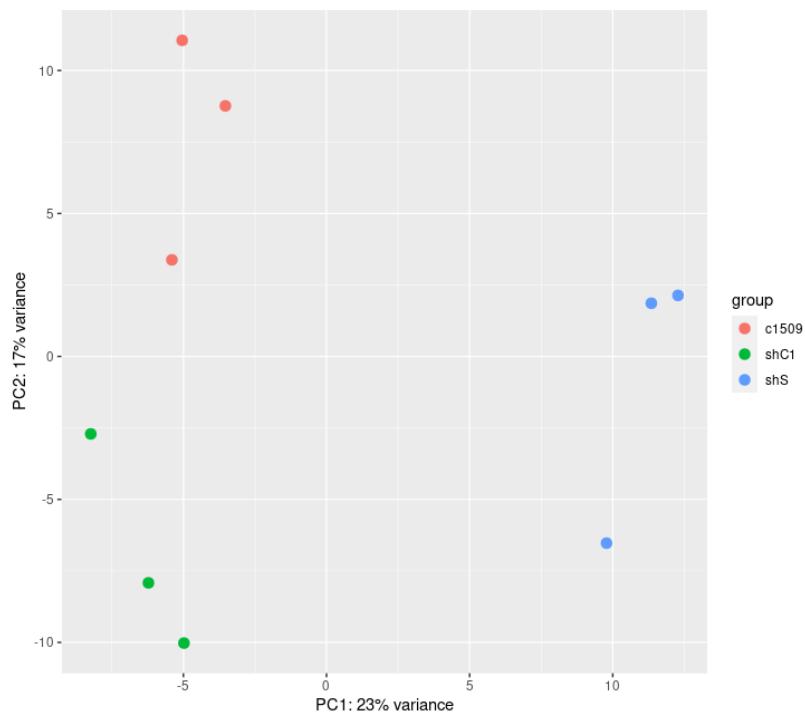
plotHeatmap      -m      AroundK27matrix      --colorList      "white,blue"
"white,blue" "white,blue" "white,blue" "white,blue" "white,blue"
-out      AroundK27_computeMatrix1.png      --sortUsing      max      --samplesLabel
K27_peggy H3K27me3 H1 panH1 IgG H3K9me3

bedtools bamtobed -i H1_4.mLb.clN.sorted.bam > H1_4.mLb.clN.sorted.bed
grep chr H1_4.mLb.clN.sorted.bed > H1_4.mLb.clN.sorted_chr.bed
bedtools coverage      -a      H3K27me3_ENCFF801AHF_regions.bed      -b
H1_4.mLb.clN.sorted_chr.bed | sort -k1,1 -k2,2n > H1_4_K27me3.bed

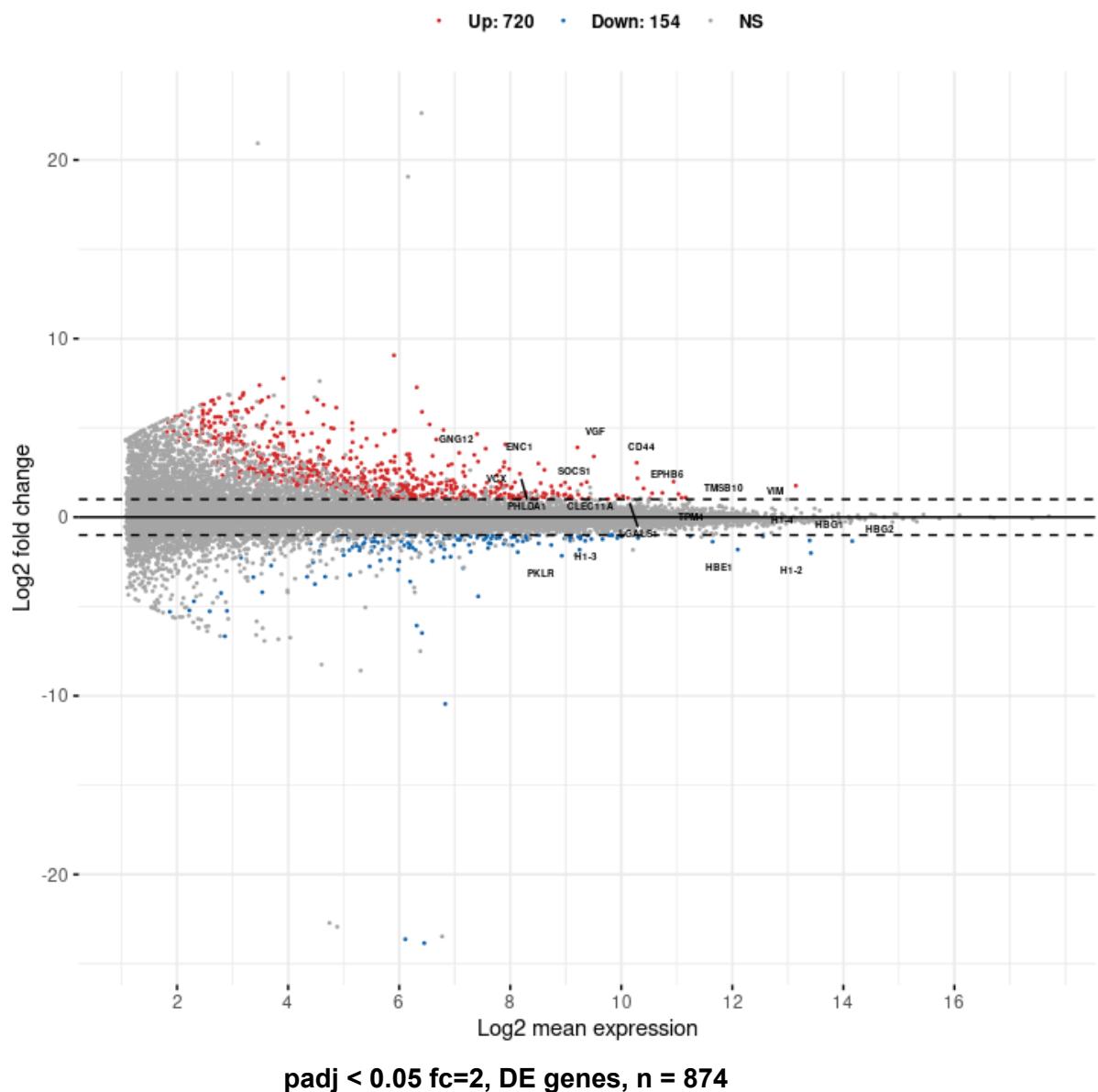
```

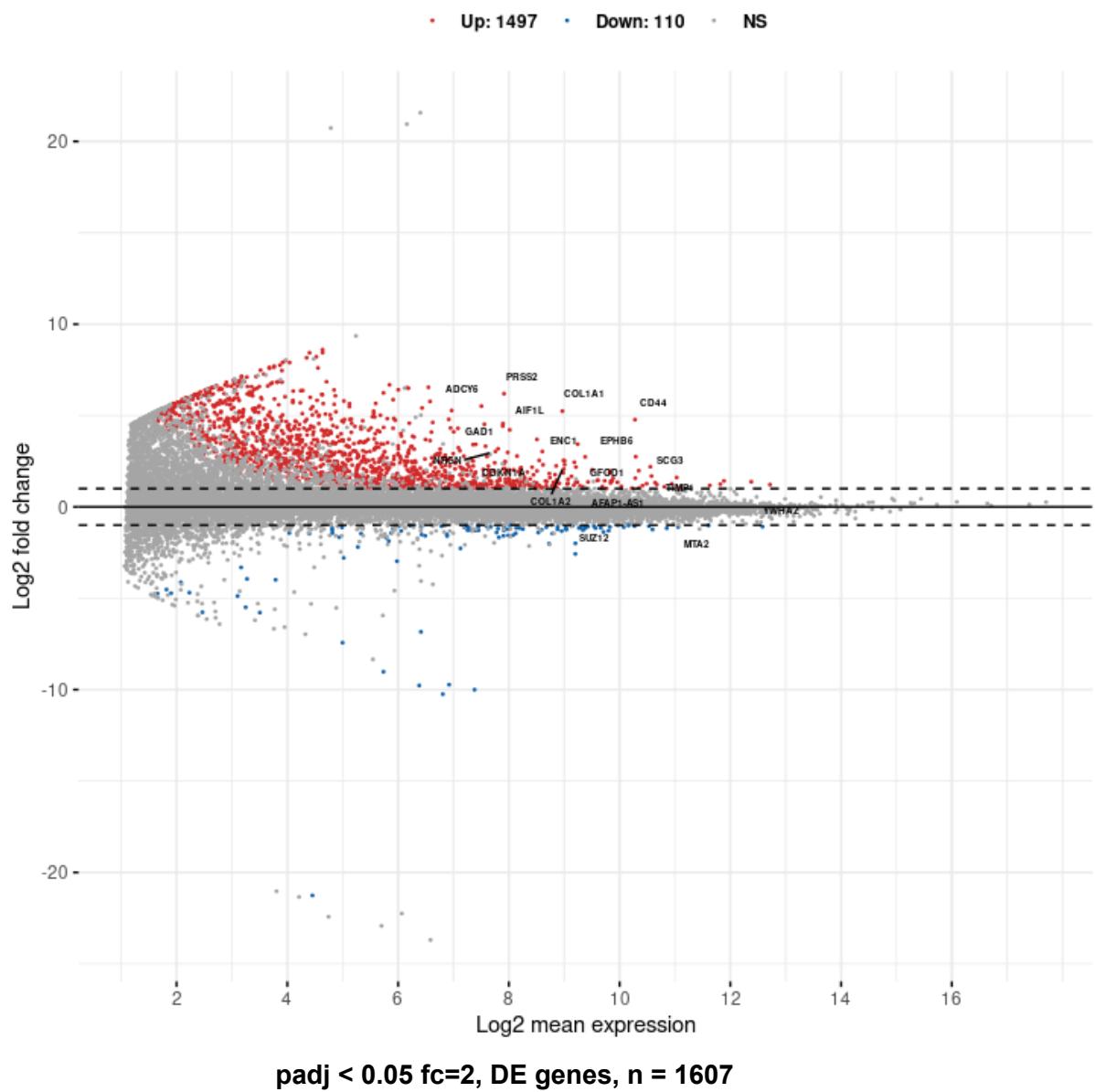
18. #RNA-Seq

PCA plot for all samples vst counts

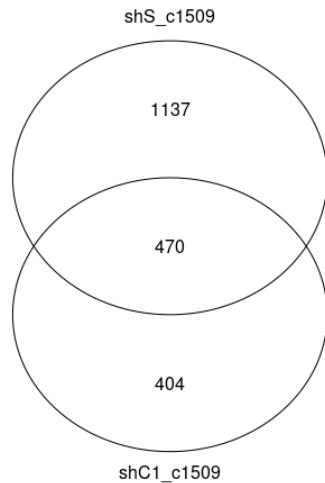


MA plot shC1 vs c1509

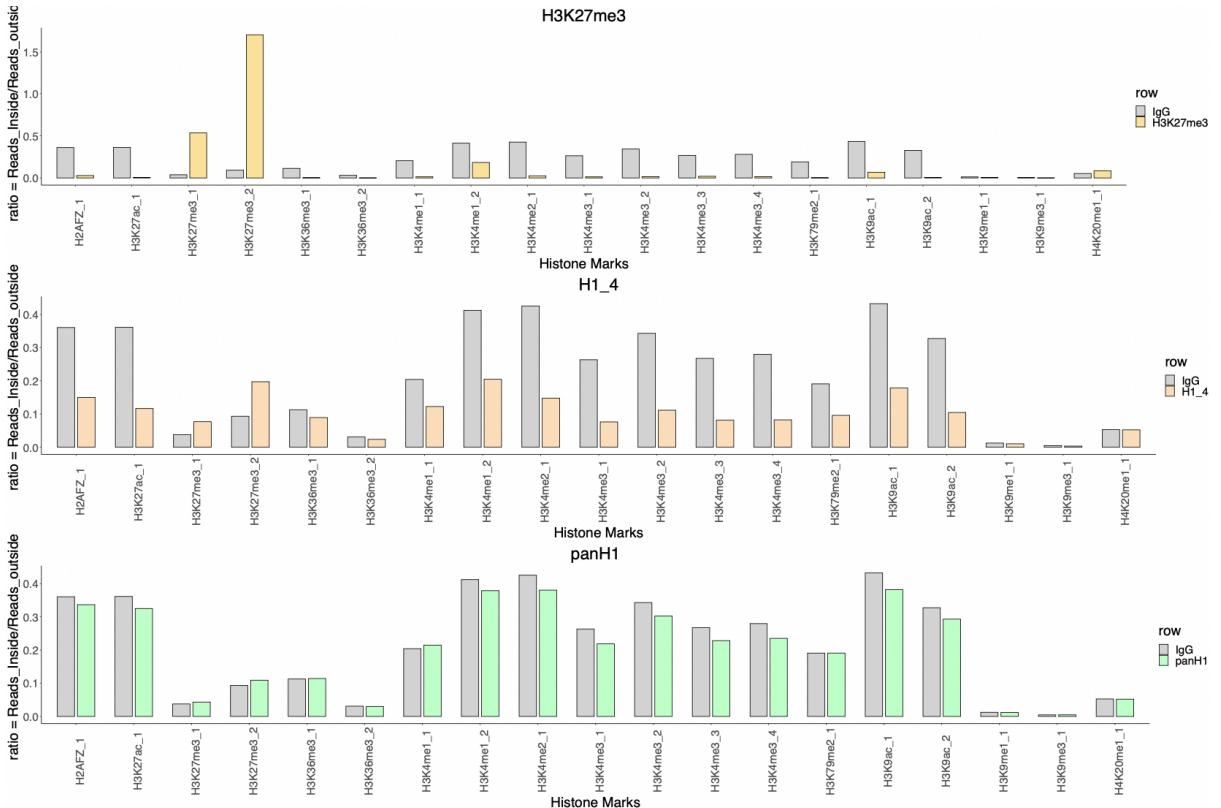




Venn shS_c1509 vs shC1_c1509



19. Enrichment analysis for H1, H3K27me3 and panH1



20.

Meeting with Iva and Rachael - 5th October

Adam's notes:

- Correlate with K27 peaks, SUZ12 peaks, H1 CUT&TAG and RNA-seq shCRAMP1 (genes).
- Do opening peaks overlap with enhancers, either active or general?

- We knock down CRAMP1, this knocks down H1, we expect to see opening chromatin at the genes which we see increase in expression when we knock down CRAMP1, and the same for K27. K27 ought to be maintained at these loci, but loss of H1 increases accessibility. We expect that K9 regions may not vary, although it is also possible that they do.
- Do we see differential peaks enriched for mononucleosomal reads i.e. they are not from promoter-like nucleosome-free regions.

Ankit notes

1. Complete all datasets analysis
 - a. ATAC-Seq: done
 - b. RNA-Seq: done
 - c. Cut&Run KO: ongoing, check with different packages, like edgeR, DESeq2, NOISeq, MA plot again
 - d. Cut&Run KD: ongoing, quantify over consensus peaks (peaks were called by nextflow) and annotate consensus peaks to genes (use annotation by nextflow)
 - e. Cut&Tag normal: statistical analysis, quantify over genes and nearby regions.
2. ENCODE K27 and K9 peaks and quantify cutntag data, prepare table as well as tornado
3. Integration table for all datasets, gene-based and bin-based on all consensus intervals (excluding rna-seq data)/ Complex heatmap for integration
4. H1 data is H1.4, yes
5. Use Peggy K27 data and available K9 data of ENCODE
6. Statistical analysis for cutntag data: overall enrichment
7. Check consensus of nextflow and manually

Adam's notes on analysis 19th October 2023

1. Make a heatmap with rows as genes, columns as:
 - RNA-seq KDs (reps, genes)
 - Log2fc (when q-value is significant) in RNA-seq of shCRAMP1 (blank otherwise)
 - Log2fc (when q-value is significant) in RNA-seq shSUZ12 (blank otherwise)
 - Log2fc (when q-value is significant) in RNA-seq shH1 (blank otherwise)???
 - ATAC-seq KDs (reps, peaks)
 - Log2fc (when q-value is significant) in ATAC-seq shCRAMP1 (blank otherwise)
 - Log2fc (when q-value is significant) in ATAC-seq shSUZ12 (blank otherwise)
 - CUT&RUN KDs (no reps, peaks)
 - Summed peak intensity of H3K27me3 CUT&RUN in shCramp1

- Summed peak intensity of H3K27me3 CUT&RUN in shSuz12
 - Summed peak intensity of H3K27me3 CUT&RUN in shControl
- CUT&RUN Cramp1 KO (reps, peaks)
 - Logfc H3K27me3 CUT&RUN in Cramp1 KO (reps)
- H1 CUT&TAG (no reps, no peaks)
 - norm reads counts of H1.4 CUT&TAG
 - norm reads counts of panH1 CUT&TAG
 - Summed peak intensity/norm reads counts of H3K27me3 CUT&TAG
- ENCODE Datasets (peaks, no reps)
 - summed peak intensity of H3K27me3 ChIP-seq
 - summed peak intensity of H3K9me3 ChIP-seq

2. Is there a significant enrichment of H1 in K27 or K9 peaks?
3. Do we see differential ATAC-seq peaks enriched for mononucleosomal reads i.e. they are not from promoter-like nucleosome-free regions?
4. Differential ATAC-seq peaks as bed files for Iva to send to their hi-c collaborators

Notes on October 27 2023, Ankit's:

1. I tested the overlap of differentially expressed genes (rnaseq) and genes near differentially accessible regions (atacseq) in two ways:
 - a. Perform annotation to gene by summing reads in those peaks which are near to the genes,
 - i. Homer (Distance.to.TSS < 5000 & Distance.to.TSS > -5000)


```
length(intersect(deseq2_atacseqkd_genic_de_shCRAMP1_shControl_0.05_fc_df$Gene, deseq2_rnaseqde_shC1_c1509_0.05_fc_df$gene))
[1] 8
```
 - ii. Manual (Distance < 5000)


```
>
length(intersect(deseq2_atacseqkd_mg_de_shCRAMP1_shControl_0.05_fc_df$Gene, deseq2_rnaseqde_shC1_c1509_0.05_fc_df$gene))
[1] 14
```
 - b. Perform DE analysis to identify DARs and then annotate to genes, which is again in two ways
 - i. Homer (Distance.to.TSS < 5000 & Distance.to.TSS > -5000)


```
>
length(intersect(deseq2_atacseqkd_de_shCRAMP1_shControl_0.05_an_n_nearest$Gene.Name,
deseq2_rnaseqde_shC1_c1509_0.05_fc_df$gene))
[1] 23
```

```

ii. Manual (Distance < 5000)
>
length(intersect(deseq2_atacseqkd_de_shCRAMP1_shControl_0.05_an
no_nearest$Gene, deseq2_rnaseqde_shC1_c1509_0.05_fc_df$gene))
[1] 117

```

2. Combining datasets of CUT&RUN KD and CUT&TAG with other datasets?
3. Overlap atac/rnaseq/cut&run KO with gene different approaches. There are NAs which means no genes are shared between datasets. So it is difficult to make interpretation
4. Genes present in both atacseq and rnaseq after integration:


```

> length(unique(na.omit(hgd_genes_akd_rkd)$Gene))
[1] 8
> length(unique(na.omit(mgd_genes_akd_rkd)$Gene))
[1] 14
> length(unique(na.omit(hpgd_genes_akd_rkd)$Gene))
[1] 23
> length(unique(na.omit(mpgd_genes_akd_rkd)$Gene))
[1] 117

```
5. Genes present in both atacseq and rnaseq after integration
6. RNA-Seq as such
7. ATAC-Seq KD, CUT&RUN KO merged and made one value by MEAN of log2FC, so that I get per gene log2FC
8. CUT&RUN KD: Combine and make one value, CPM, summing up, so that I get only one CPM count per gene
9. Cut&Tag count over gene+1500bp,, +start (-1500) , - (end + 1500), CPM
10. Merging for only Differential genes in (Atac seq , rnaseq, cutnrunkd)

Effect:

CRAMP1 loss → H1-4 loss → PRC2 target genes perturbed (up?)

Possibilities:

1. H1.4 helps in deposition of H3K27me3 → X
2. H1.4 helps in recruitment of PRC2 → need to check

Notes on 30 Oct, 2023

```

# cpm -> log2 cpm +1
# NA -> 0 -> make it white color instead of yellow
# Give uniform color to ckd and cutntag
# cutnrunko calculate differential by subtract of log2CPM shCRAMP1 - shControl

```

```

# cutntag calculate enrichment over IgG # or cutandtag convert to z-scores of CPM
# add Suz12 kd rnaseq, atacseq,
# add H1 rnaseq kd
# do the matrix for all genes without removing anything and make master matrix and then
reduce to subsets

```

Ankit's notes Nov 1 , 2023

1. Count over only WT Peaks
2. ATAC-Seq:
3. CUT&RUN KO:
4. CUT&RUN KD:
5. CUT&TAG:
6. Group into different categories

1. Bin size (5,10 Kb) Based analysis, ATAC-Seq, CUT&RUN KO, KD, CUT&TAG
 - a. Narrow down to ATAC-Seq Control Peaks +/- 1Kb
 - b. Take H3K27me3 Peaks too
2. DARs ATAC-Seq peaks +/- 1Kb: (to capture K27me3 info)
 - a. DARs, shCRAMP1-shControl,
 - b. DARs, shSUZ12-shControl
3. Gene Based:Differential based analysis: DEGs, done prior
4. PCA plot for better representation of combined data
5. Categorize the data

Ankit Notes:

Nov 16

1. Divide Genes into cluster based on clustering across datasets
2. Get Genes with cluster 1,2,3 and get their bed files,
3. Run computematrix

```

# deepTools
computeMatrix scale-regions -S H3K27me3.bigWig H1_4.bigWig panH1.bigWig IgG.bigWig
-R gene_v41_human_out_re_cluster1.txt gene_v41_human_out_re_cluster2.txt
gene_v41_human_out_re_cluster3.txt -a 1000 -b 1000 --regionBodyLength 5000
--skipZeros -o gene_cluster_matrix.mat.gz

```

Since IgG was having intensity to high scaling factor which is due to low read coverage I decided to compare bigwig and subtract

```

bigwigCompare -b1 H3K27me3.bigWig -b2 IgG.bigWig --operation subtract -o
H3K27me3_minus_IgG.bw -p 12 -v

```

```

bigwigCompare -b1 H1_4.bigWig -b2 IgG.bigWig --operation subtract -o
H1_4_minus_IgG.bw -p 12 -v
bigwigCompare -b1 panH1.bigWig -b2 IgG.bigWig --operation subtract -o
panH1_minus_IgG.bw -p 12 -v
bigwigCompare -b1 IgG.bigWig -b2 IgG.bigWig --operation subtract -o
IgG_minus_IgG.bw -p 12 -v

computeMatrix scale-regions -S H3K27me3_minus_IgG.bw H1_4_minus_IgG.bw
panH1_minus_IgG.bw -R gene_v41_human_out_re_cluster1.txt
gene_v41_human_out_re_cluster2.txt gene_v41_human_out_re_cluster3.txt -a 1000 -b
1000 --regionBodyLength 5000 --skipZeros -o gene_cluster_matrix.mat.gz -p 12

plotProfile -m gene_cluster_matrix.mat.gz -out gene_cluster_matrix.mat.png
--numPlotsPerRow 3 --plotTitle "CUT&TAG data profile around gene cluster(1-3)"
--samplesLabel

plotHeatmap -m gene_cluster_matrix.mat.gz -out heatmap_gene_cluster_matrix.png
--sortUsing max

```

4. Check what is scale in plotProfile? [Mean of score from bigwig in a given bin](#)
5. See why IgG is coming highest. [Because the coverage is low and that leads to high scale factor](#)
6. Select appropriate clustering . It is better to use ComplexHeatmap clustering. Draw a line to get three different clusters
7. Perform Bin based DARs, DBRs analysis and combine the data like gene based

DARs

CUTAND RUN KO

Meeting with Rachael and Iva 20th November 2023

We came up with a plan for the main analysis - leading us to the final hypothesis of how cramp1 affects the expression of PRC2 target genes:

1. Overlap between cramp1 and suz12 target genes
 - a. Venn diagram
 - b. Heatmap with no p-value cutoff, but logFC cutoff!!!
2. H3K27me3 (CUT&RUN WT) including 1,2,3kb upstream
 - a. Look at commonly regulated and individually regulated ([UP only, add rest of the genes](#))
 - i. Heatmap?
 - ii. Histograms/density
 - iii. H3K27me3 Differentially enriched Genes for ([UP only, add rest of the genes](#))

3. H3K27me3 changes when knock out cramp1, and down Cramp1, or SUZ12 in these different categories of genes (CUT&RUN) (UP only, add rest of the genes)
 - a. Same as (2) but for KD/KO
4. ATAC-seq DAR at the three categories of genes – do they become more accessible at TSS/upstream? (UP only, add rest of the genes)
 - a. 2 histograms with each of the 3 gene categories
5. Do the genes in the three categories have enrichment for H1.4 (CUT&TAG)?
 - a. Histogram for three gene categories (UP only, add rest of the genes)

We also discussed:

- ATAC-seq – should only include upstream of start site, by 1-3kb. But will this result in too few peaks?
- CUT&TAG artefact of integration near TSSs might be something other people have worked with. Could be that we use the flag control sample rather than the IgG which has much lower coverage (Ankit/Iva will check for this).

New data on the way soon:

- CUT&TAG for all H1 variants, endogenous with antibody and with FLAG.
- Cramp1 and SUZ12 KD with H1 and panH1 (CUT&TAG), no replicates, paired-end.

Notes

ATAC seq data

1. Run Deseq2, limma and edgeR individually and compare
2. compare deseq2 and limma and edgeR overlap
3. Export counts from Diffbind and run limma on it and compare with diffbind, deseq2, edgeR and limma

CUT&Run seq data

Run Deseq2, limma and edgeR individually
 compare deseq2 and limma and edgeR overlap

Notes Dec 6

CUT&TAG new data paired end

Check enrichment for all H1 variants across all Histone marks

Use old H1.4 and new H1.4

Send the bed files

Iva's email Dec 6

Hi both,

Thank you for joining us this morning for an impromptu meeting! - we're quite excited about the H3K27me3 data you showed us, and think it will make an excellent addition to the paper. Fingers crossed for the profile and tornado plots for these!

As promised here is a link to the new H1 CUT&Tag data:

[RM_CUT&Tag](#)

In there is a sample sheet with info, but briefly the samples are split into three groups:

1) Endogenous H1 variants in wild-type cells: just like with the previous H1.4 data, it would be good to look at the enrichment of these across the different histone modification peaks from ENCODE. We think H3 and H5 haven't worked, but I guess they can be good controls if not anything else! We can also compare the localisation of the variants that have worked to each other?

2) FLAG-tagged H1 variants: this is the equivalent to the above but with exogenously expressed FLAG-tagged versions - so the same analysis as above can be performed. Along with these are the FLAG-H1.4 and FLAG control (FLAG CUT&Tag in non-FLAG-expressing cells) from last time but just sequenced again to try and get more coverage.

3) Endogenous H1.4 and pan-H1 in shControl and shCRAMP1 cells: you can use the data in the shControl cells as replicates for H1.4 and pan-H1 if you'd like, but the main thing we'd like to test with this experiment is whether H1.4 and/or pan-H1 levels decrease uniformly across the genome, or whether they decrease at specific sites upon shCRAMP1;

Also if you could share with us the bed files for the ATAC-seq data using the different packages, that would be great.

samplesheet					
group	replicates	fastq_1	fastq_2	control	
FLAG_H12 ✓	1	ITit257_RM16_S16_L001_R1_001.fastq.gz	ITit257_RM16_S16_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H12 ✓	1	ITit257_RM16_S16_L002_R1_001.fastq.gz	ITit257_RM16_S16_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H13 ✓	1	ITit257_RM17_S17_L001_R1_001.fastq.gz	ITit257_RM17_S17_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H13 ✓	1	ITit257_RM17_S17_L002_R1_001.fastq.gz	ITit257_RM17_S17_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H14 ✓	1	ITit257_RM18_S18_L002_R1_001.fastq.gz	ITit257_RM18_S18_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H14 ✓	1	ITit257_RM18_S18_L001_R1_001.fastq.gz	ITit257_RM18_S18_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H14_old ✓	1	ITit257_RM22_S22_L002_R1_001.fastq.gz	ITit257_RM22_S22_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H14_old ✓	1	ITit257_RM22_S22_L001_R1_001.fastq.gz	ITit257_RM22_S22_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H15 ✓	1	ITit257_RM19_S19_L002_R1_001.fastq.gz	ITit257_RM19_S19_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H15 ✓	1	ITit257_RM19_S19_L001_R1_001.fastq.gz	ITit257_RM19_S19_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H1X ✓	1	ITit257_RM20_S20_L002_R1_001.fastq.gz	ITit257_RM20_S20_L002_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_H1X ✓	1	ITit257_RM20_S20_L001_R1_001.fastq.gz	ITit257_RM20_S20_L001_R2_001.fastq.gz	FLAG_old ✓	✓
FLAG_old ✓	1	ITit257_RM21_S21_L001_R1_001.fastq.gz	ITit257_RM21_S21_L001_R2_001.fastq.gz	✓	✓
FLAG_old ✓	1	ITit257_RM21_S21_L002_R1_001.fastq.gz	ITit257_RM21_S21_L002_R2_001.fastq.gz	✓	✓
shControl_H1 ✓	1	ITit257_RM2_S2_L001_R1_001.fastq.gz	ITit257_RM2_S2_L001_R2_001.fastq.gz	shControl_IgG ✓	✓
shControl_H1 ✓	1	ITit257_RM2_S2_L002_R1_001.fastq.gz	ITit257_RM2_S2_L002_R2_001.fastq.gz	shControl_IgG ✓	✓
shControl_H14 ✓	1	ITit257_RM1_S1_L002_R1_001.fastq.gz	ITit257_RM1_S1_L002_R2_001.fastq.gz	shControl_IgG ✓	✓
shControl_H14 ✓	1	ITit257_RM1_S1_L001_R1_001.fastq.gz	ITit257_RM1_S1_L001_R2_001.fastq.gz	shControl_IgG ✓	✓
shControl_IgG ✓	1	ITit257_RM3_S3_L002_R1_001.fastq.gz	ITit257_RM3_S3_L002_R2_001.fastq.gz	✓	✓
shControl_IgG ✓	1	ITit257_RM3_S3_L001_R1_001.fastq.gz	ITit257_RM3_S3_L001_R2_001.fastq.gz	✓	✓
shCRAMP1_H1 ✓	1	ITit257_RM5_S5_L002_R1_001.fastq.gz	ITit257_RM5_S5_L002_R2_001.fastq.gz	shCRAMP1_IgG ✓	✓
shCRAMP1_H1 ✓	1	ITit257_RM5_S5_L001_R1_001.fastq.gz	ITit257_RM5_S5_L001_R2_001.fastq.gz	shCRAMP1_IgG ✓	✓
shCRAMP1_H14 ✓	1	ITit257_RM4_S4_L001_R1_001.fastq.gz	ITit257_RM4_S4_L001_R2_001.fastq.gz	shCRAMP1_IgG ✓	✓
shCRAMP1_H14 ✓	1	ITit257_RM4_S4_L002_R1_001.fastq.gz	ITit257_RM4_S4_L002_R2_001.fastq.gz	shCRAMP1_IgG ✓	✓
shCRAMP1_IgG ✓	1	ITit257_RM6_S6_L001_R1_001.fastq.gz	ITit257_RM6_S6_L001_R2_001.fastq.gz	✓	✓
shCRAMP1_IgG ✓	1	ITit257_RM6_S6_L002_R1_001.fastq.gz	ITit257_RM6_S6_L002_R2_001.fastq.gz	✓	✓
shSUZ12_H1 ✓	1	ITit257_RM8_S8_L001_R1_001.fastq.gz	ITit257_RM8_S8_L001_R2_001.fastq.gz	shSUZ12_IgG ✓	✓
shSUZ12_H1 ✓	1	ITit257_RM8_S8_L002_R1_001.fastq.gz	ITit257_RM8_S8_L002_R2_001.fastq.gz	shSUZ12_IgG ✓	✓
shSUZ12_H14 ✓	1	ITit257_RM7_S7_L002_R1_001.fastq.gz	ITit257_RM7_S7_L002_R2_001.fastq.gz	shSUZ12_IgG ✓	✓
shSUZ12_H14 ✓	1	ITit257_RM7_S7_L001_R1_001.fastq.gz	ITit257_RM7_S7_L001_R2_001.fastq.gz	shSUZ12_IgG ✓	✓
shSUZ12_IgG ✓	1	ITit257_RM9_S9_L002_R1_001.fastq.gz	ITit257_RM9_S9_L002_R2_001.fastq.gz	✓	✓
shSUZ12_IgG ✓	1	ITit257_RM9_S9_L001_R1_001.fastq.gz	ITit257_RM9_S9_L001_R2_001.fastq.gz	✓	✓
WT_H12 ✓	1	ITit257_RM10_S10_L002_R1_001.fastq.gz	ITit257_RM10_S10_L002_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H12 ✓	1	ITit257_RM10_S10_L001_R1_001.fastq.gz	ITit257_RM10_S10_L001_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H13 ✓	1	ITit257_RM11_S11_L001_R1_001.fastq.gz	ITit257_RM11_S11_L001_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H13 ✓	1	ITit257_RM11_S11_L002_R1_001.fastq.gz	ITit257_RM11_S11_L002_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H14 ✓	1	ITit257_RM12_S12_L002_R1_001.fastq.gz	ITit257_RM12_S12_L002_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H14 ✓	1	ITit257_RM12_S12_L001_R1_001.fastq.gz	ITit257_RM12_S12_L001_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H15 ✓	1	ITit257_RM13_S13_L001_R1_001.fastq.gz	ITit257_RM13_S13_L001_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H15 ✓	1	ITit257_RM13_S13_L002_R1_001.fastq.gz	ITit257_RM13_S13_L002_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H1X ✓	1	ITit257_RM14_S14_L002_R1_001.fastq.gz	ITit257_RM14_S14_L002_R2_001.fastq.gz	WT_IgG ✓	✓
WT_H1X ✓	1	ITit257_RM14_S14_L001_R1_001.fastq.gz	ITit257_RM14_S14_L001_R2_001.fastq.gz	WT_IgG ✓	✓
WT_IgG ✓	1	ITit257_RM15_S15_L001_R1_001.fastq.gz	ITit257_RM15_S15_L001_R2_001.fastq.gz	✓	✓
WT_IgG ✓	1	ITit257_RM15_S15_L002_R1_001.fastq.gz	ITit257_RM15_S15_L002_R2_001.fastq.gz	✓	✓

```
> dba.show(atacseqkd_diffbind_list$dar_analysis$edgeR$norm, bContrasts=TRUE)
Factor      Group Samples   Group2 Samples2 DB.edgeR DB.DESeq2
1 Condition shControl      2 shCRAMP1      2       13401      7666
```

```

2 Condition shControl      2 shSUZ12      2    13914      7275
3 Condition shSUZ12       2 shCRAMP1     2    5526       1252
> dba.show(atacseqkd_diffbind_list$dar_analysis$deseq2$norm, bContrasts=TRUE)
  Factor      Group Samples   Group2 Samples2 DB.edgeR DB.DESeq2
1 Condition shControl      2 shCRAMP1     2    13401      7666
2 Condition shControl      2 shSUZ12      2    13914      7275
3 Condition shSUZ12       2 shCRAMP1     2    5526       1252

```

Dec 13

1. Perform Deseq2, edgeR and limma on Consensus ATACseq peaks and plot and MA and Upset plot
2. Duplicates don't remove and peak calls using SEACR and MACS2
3. Histone variant data with encode marks
4. PCA for new Cut&Tag data: Take both duplicate and non-duplicate and count the bam and perform PCA
5. Markup -> Dedup (By nextflow only control/IgG samples were deduplicated)

Dec 19- Plan

1. How does Diffbind run Deseq2 and edgeR? Which one to choose?
2. How SEACR works?
3. Annotate to genes based on 2kb upstream of genes + gene bodies and use bedtools closest ad -d 0 for all genes overlap to say it as nearest genes.
4. Chipseeker based annotation
5. Keep FLAG sample codes and re-process Cut&Tag. Remove FLAGs except FLAG-H1X and also remove WT_H13.
6. Visualise the Cut&Tag data on IGV. (Integration with old Cut&Tag data). Flag old seems similar to H1s and Flag_H14_old looks low coverage and more like IgG.(Why??)
7. At counts level merge the old and new Cut&Tag data.
 - a. Histone Marks
8. Quantification:
 - a. Requantify ATACseq with Diffbind and assign genes
 - b. Gene-based: DE genes or Genes near the DARs (atacseq: mean log2FC) or Genes near the DBRs (cutnrun KO: mean) Genes near all consensus Peaks (cutnrun KD: sum) , Genes quantification Cut&Tag old and new: sum.
 - c. Extended Gene coordinates:1.5Kb upstream: all except RNA-Seq.
 - d. Bin size (5kb/10kb) quantification for data and process as gene-based data. (To see the pattern of changes independent of genes and annotate these regions to genomic features to tell if any specific type genomic feature has biasness): all except RNA-Seq
 - i. PCA: Global wide samples (CUT&TAG new only)

- e. Use CUT&TAG H3K27me3 as one more control to check PCA and if ok use it in Diffbind
- f. Four categories of genes promoter regions (CRAMP1/SUZ12 DE and overlap and unchanged) and take only **upstream of the start site**, by 1,2,3kb (do they become more accessible at TSS/upstream).: all except RNA-Seq

9. Integration:

- a. Gene-based and subset
 - i. Heatmap (Keep the same pattern: NA removed from CRAMP1 KD RNASeq only)
 - 1. Categorise integrated data into groups based on clustering (Keep DE genes as a base)
 - a. Categories of genes/bins obtained after **clustering** all datatypes (from Integration). Evaluate the evidence on IGV.
 - b. Explore the function of these genes /Pathways.
 - ii. PCA (How?? Some Genes values represent log2FC and some have sum)
- b. Bin-based and subset
 - i. PCA: All data together (control subtracted) : bins as points (<https://www.nature.com/articles/s41586-020-2023-4/figures/1>):
 - ii. DARs from ATAC-Seq data and CUT&RUN KO and KD: Heatmap cluster and Tornado and Average Plots.
 - iii. DARs together both /uniq shCRAMP1/shSUZ12
- c. DARs from ATAC-Seq data and CUT&TAG: Heatmap cluster and Tornado and Average Plots.
- d. DEGs list (four categories) from RNA-Seq match with **nearest genes** to DARs ATAC-Seq data (log2FC).
 - i. Density Plot (log2FC) per gene category (four categories).
- e. DEGs list (four categories) from RNA-Seq match with **nearest genes** DBRs CUT&RUN KO (log2FC)
 - i. Density Plot (log2FC) per gene category (four categories).
 - ii. Density plot for quantification in DEG genes RNASeq.
- f. DEGs list (four categories) from RNA-Seq match with **nearest genes** CUT&RUN KD.
- g. DEGs from RNA-Seq and Genes quantification CUT&TAG.
 - i. Density plot

10. Heatmap plot/Tornado Plots/Meta Plots for All histone variants around TSS.

11. Correlation of H3K27me3 and other histone marks with all Histone variants Cut&Tag datasets using Bins based matrix (5Kb).

12. Extract differentially enriched regions for each Histone variant and annotate to genomic features. DBRs? The data has a very low coverage.
13. Learn how to generate Metagene Profile plots. Which approach is better bigwig compare ? or bam compare and then convert to bigwig and then plot? What is the best method of normalisation: divide or subtract? Plot all 6 data types on profile plot.
 - a. Plot for Cut&Tag data and Cut&Run around TSS, TES
 - b. Plot for Cut&Tag data and Cut&Run KO, KD around ATAC-Seq DARs in both CRAMP1, SUZ12

Dec 21

Standardisation of commands for Metagene plots

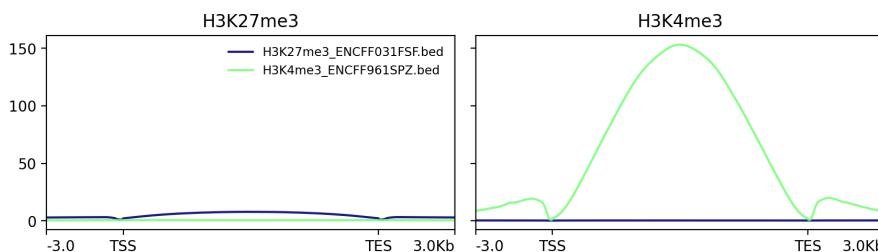
Prepare file

```
samtools sort -o hesc.Input.500K.sorted.bam hesc.Input.500K.bam
bedtools genomecov -ibam hesc.Input.500K.sorted.bam -bg > hesc.Input.500K.sorted.bdg
sort -k1,1 -k2,2n hesc.Input.500K.sorted.bdg > hesc.Input.500K.resorted.bdg
~/tools/av/bedGraphToBigWig hesc.Input.500K.resorted.bdg hg19.chrom.sizes hesc.Input.500K.resorted.bw
bigwigCompare -b1 hesc.H3k27me3.1M.resorted.bw -b2 hesc.Input.500K.resorted.bw -o
hesc.H3k27me3.1M.resorted input frombigwig.bw -p 20

samtools sort -o hesc.H3k4me3.1M.sorted.bam hesc.H3k4me3.1M.bam
bedtools genomecov -ibam hesc.H3k4me3.1M.sorted.bam -bg > hesc.H3k4me3.1M.sorted.bdg
sort -k1,1 -k2,2n hesc.H3k4me3.1M.sorted.bdg > hesc.H3k4me3.1M.resorted.bdg
~/tools/av/bedGraphToBigWig hesc.H3k4me3.1M.resorted.bdg hg19.chrom.sizes hesc.H3k4me3.1M.resorted.bw
bigwigCompare -b1 hesc.H3k4me3.1M.resorted.bw -b2 hesc.Input.500K.resorted.bw -o hesc.H3k4me3.1M.resorted input frombigwig.bw
```

Deeptools

```
computeMatrix scale-regions -S H3K27me3 ENCFF470HOG.bigWig H3K4me3 ENCFF144MRB.bigWig -R H3K27me3 ENCFF031FSF.bed
H3K4me3 ENCFF961SPZ.bed -a 3000 -b 3000 --regionBodyLength 10000 --sortRegions keep --missingDataAsZero -o
over_histone_matrix.gz -p 20 --skipZeros
plotProfile -m over_histone_matrix.gz -out over_histone_matrix.png --samplesLabel H3K27me3 H3K4me3
```



```
computeMatrix scale-regions -S H3K27me3 ENCFF470HOG.bigWig H3K4me3 ENCFF144MRB.bigWig -R H3K27me3 ENCFF031FSF.bed
H3K4me3_ENCFF961SPZ.bed -a 5000 -b 5000 --regionBodyLength 1500 --sortRegions keep --missingDataAsZero -o
over_histone_matrix.gz -p 20 --skipZeros
plotProfile -m over_histone_matrix.gz -out over_histone_matrix.png --samplesLabel H3K27me3 H3K4me3
```

Dec 31

Cutnrun ko atacseq DARs

```
grep chr atacseqkd_diffbind_d_shCRAMP1_shControl_de_re.bed >
atacseqkd_diffbind_d_shCRAMP1_shControl_de_rechr.bed
```

```

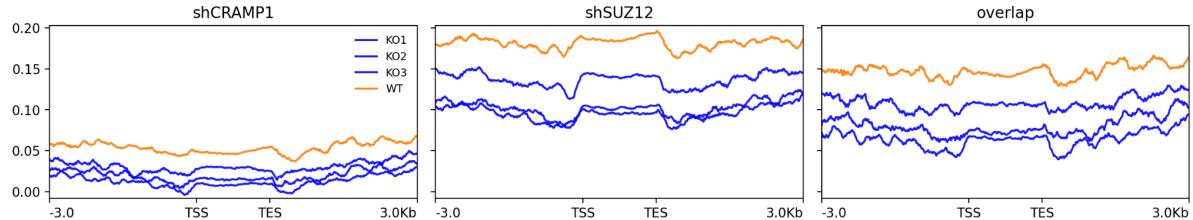
grep chr atacseqkd_diffbind_d_shSUZ12_shControl_de_re.bed >
atacseqkd_diffbind_d_shSUZ12_shControl_de_rechr.bed
bedtools intersect -a atacseqkd_diffbind_d_shCRAMP1_shControl_de_rechr.bed -b
atacseqkd_diffbind_d_shSUZ12_shControl_de_rechr.bed >
atacseqkd_diffbind_d_overlap_shCRAMP1_shSUZ12_shControl_de_rechr.bed

bigwigCompare -b1 H3K27me3_KO1.bigWig -b2 IgG_KO1.bigWig -o H3K27me3_KO1_IgG.bw
-p 20
bigwigCompare -b1 H3K27me3_KO2.bigWig -b2 IgG_KO2.bigWig -o H3K27me3_KO2_IgG.bw
-p 20
bigwigCompare -b1 H3K27me3_KO3.bigWig -b2 IgG_KO3.bigWig -o H3K27me3_KO3_IgG.bw
-p 20
bigwigCompare -b1 H3K27me3_WT.bigWig -b2 IgG_WT.bigWig -o H3K27me3_WT_IgG.bw -p
20

computeMatrix scale-regions -S H3K27me3_KO1_IgG.bw H3K27me3_KO2_IgG.bw
H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
atacseqkd_diffbind_d_shCRAMP1_shControl_de_rechr.bed
atacseqkd_diffbind_d_shSUZ12_shControl_de_rechr.bed
atacseqkd_diffbind_d_overlap_shCRAMP1_shSUZ12_shControl_de_rechr.bed -a
3000 -b 3000 --regionBodyLength 1500 --sortRegions keep --missingDataAsZero
-o matrix.gz -p 20 --skipZeros

plotProfile -m matrix.gz -out cutnrunko_matrix.png --samplesLabel
KO1 KO2 KO3 WT --perGroup --colors "#0000FF" "#0000FF" "#0000FF"
"#ff7f00" --regionsLabel shCRAMP1 shSUZ12 overlap

```



Cutnrun kd atacseq DARs

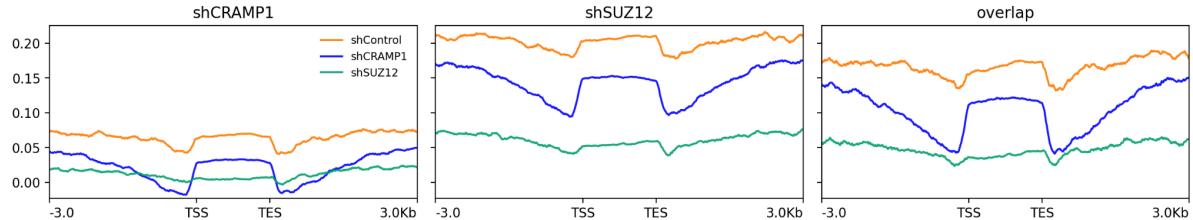
```

bigwigCompare -b1 shControl_H3K27me3.bigWig -b2
shControl_IgG.bigWig -o shControl_H3K27me3_IgG.bw -p 20
bigwigCompare -b1 shCRAMP1_H3K27me3.bigWig -b2 shCRAMP1_IgG.bigWig
-o shCRAMP1_H3K27me3_IgG.bw -p 20
bigwigCompare -b1 shSUZ12_H3K27me3.bigWig -b2 shSUZ12_IgG.bigWig -o
shSUZ12_H3K27me3_IgG.bw -p 20

computeMatrix scale-regions -S shControl_H3K27me3_IgG.bw
shCRAMP1_H3K27me3_IgG.bw shSUZ12_H3K27me3_IgG.bw -R
atacseqkd_diffbind_d_shCRAMP1_shControl_de_rechr.bed
atacseqkd_diffbind_d_shSUZ12_shControl_de_rechr.bed
atacseqkd_diffbind_d_overlap_shCRAMP1_shSUZ12_shControl_de_rechr.bed -a
3000 -b 3000 --regionBodyLength 1500 --sortRegions keep --missingDataAsZero
-o matrix.gz -p 20 --skipZeros

```

```
plotProfile -m matrix.gz -out cutnrunkd_matrix_atacseqkd_dars.png
--samplesLabel shControl shCRAMP1 shSUZ12 --perGroup --colors
"#ff7f00" "#0000FF" "#00A36C" --regionsLabel shCRAMP1 shSUZ12
overlap
```

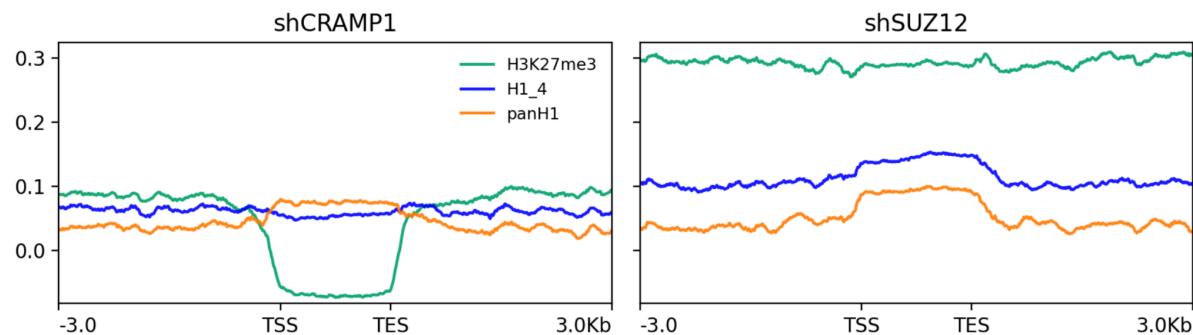


Cutntag wt atacseq DARs

```
bigwigCompare -b1 H3K27me3.bigWig -b2 IgG.bigWig -o H3K27me3_IgG.bw -p 20
bigwigCompare -b1 H1_4.bigWig -b2 IgG.bigWig -o H1_4_IgG.bw -p 20
bigwigCompare -b1 panH1.bigWig -b2 IgG.bigWig -o panH1_IgG.bw -p 20
```

```
computeMatrix scale-regions -S H3K27me3_IgG.bw H1_4_IgG.bw panH1_IgG.bw -R
atacseqkd_diffbind_d_shCRAMP1_shControl_de_re.bed
atacseqkd_diffbind_d_shSUZ12_shControl_de_re.bed -a 3000 -b 3000
--regionBodyLength 1500 --sortRegions keep --missingDataAsZero -o
atacseqkd_dars_matrix.gz -p 20 --skipZeros
```

```
plotProfile -m atacseqkd_dars_matrix.gz -out
cutntag_atacseqkd_dars_matrix.png --samplesLabel H3K27me3 H1_4
panH1 --perGroup --colors "#00A36C" "#0000FF" "#ff7f00"
--regionsLabel shCRAMP1 shSUZ12
```

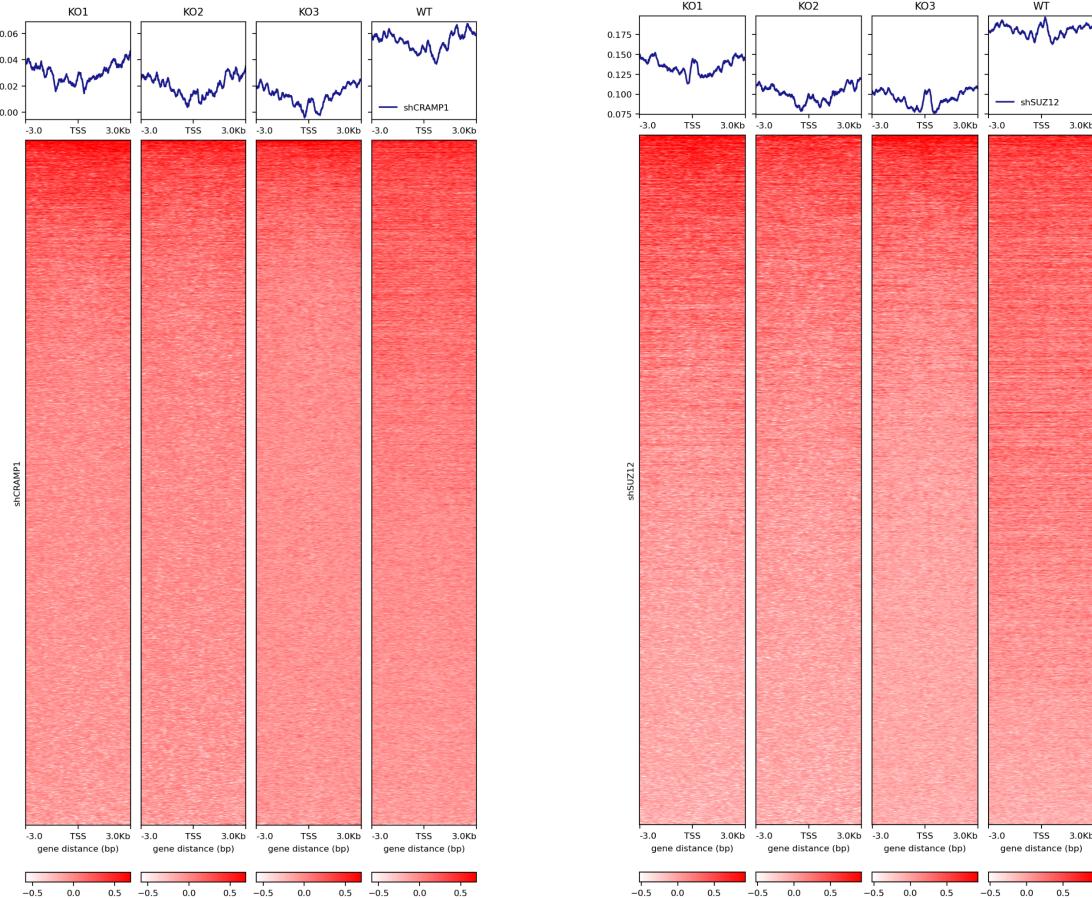


Enrichment plot # CUT&RUN KO atacseq DARs

```
computeMatrix reference-point -S H3K27me3_KO1_IgG.bw
H3K27me3_KO2_IgG.bw H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
atacseqkd_diffbind_d_shCRAMP1_shControl_de_re.bed -a 3000 -b 3000
```

```
--missingDataAsZero -o shCRAMP1_refpoint_matrix.gz -p 20
--skipZeros
```

```
plotHeatmap -m shCRAMP1_refpoint_matrix.gz -out
cutnrunko_enrichmentplot_shCRAMP1.png --samplesLabel KO1 KO2 KO3
WT --colorList "white,red" "white,red" "white,red" "white,red"
--regionsLabel shCRAMP1
```



```
computeMatrix reference-point -S H3K27me3_KO1_IgG.bw
H3K27me3_KO2_IgG.bw H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
atacseqkd_diffbind_d_shSUZ12_shControl_de_re.bed -a 3000 -b 3000
--missingDataAsZero -o shSUZ12_refpoint_matrix.gz -p 20
--skipZeros
```

```
plotHeatmap -m shSUZ12_refpoint_matrix.gz -out
cutnrunko_enrichmentplot_shSUZ12.png --samplesLabel KO1 KO2 KO3 WT
--colorList "white,red" "white,red" "white,red" "white,red"
--regionsLabel shSUZ12
```

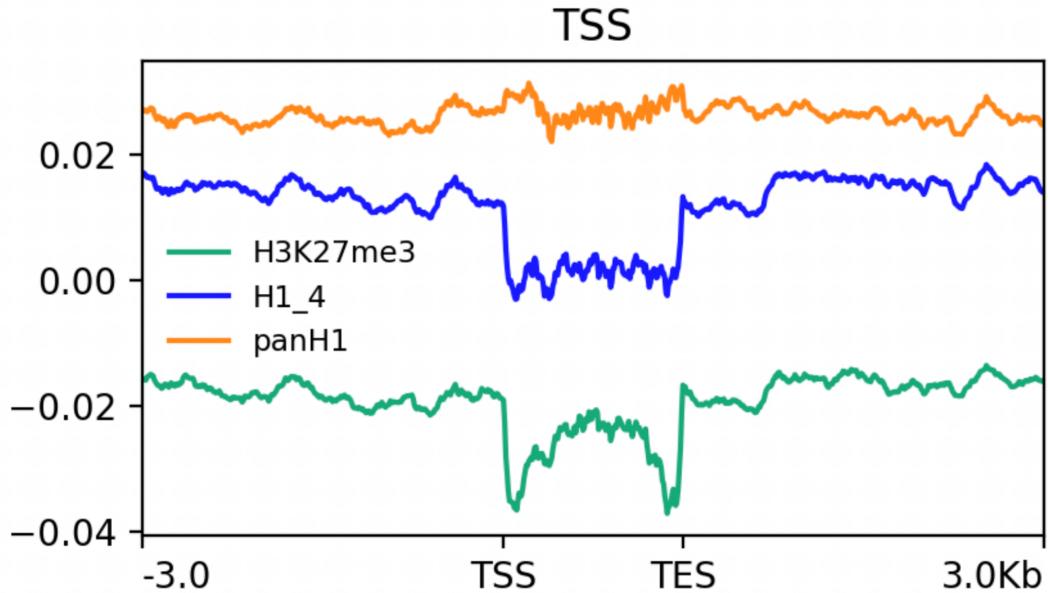
Cut&Tag wt over gene body

```
computeMatrix scale-regions -S H3K27me3_IgG.bw H1_4_IgG.bw panH1_IgG.bw -R
gene_gencodev41_gene.txt -a 3000 -b 3000 --regionBodyLength 1500
--sortRegions keep --missingDataAsZero -o cutntagwt_gene_matrix.gz -p 20
--skipZeros
```

```

plotProfile -m cutntagwt_gene_matrix.gz -out
cutntagwt_gene_matrix.png --samplesLabel H3K27me3 H1_4 panH1
--perGroup --colors "#00A36C" "#0000FF" "#ff7f00" --regionsLabel
TSS

```



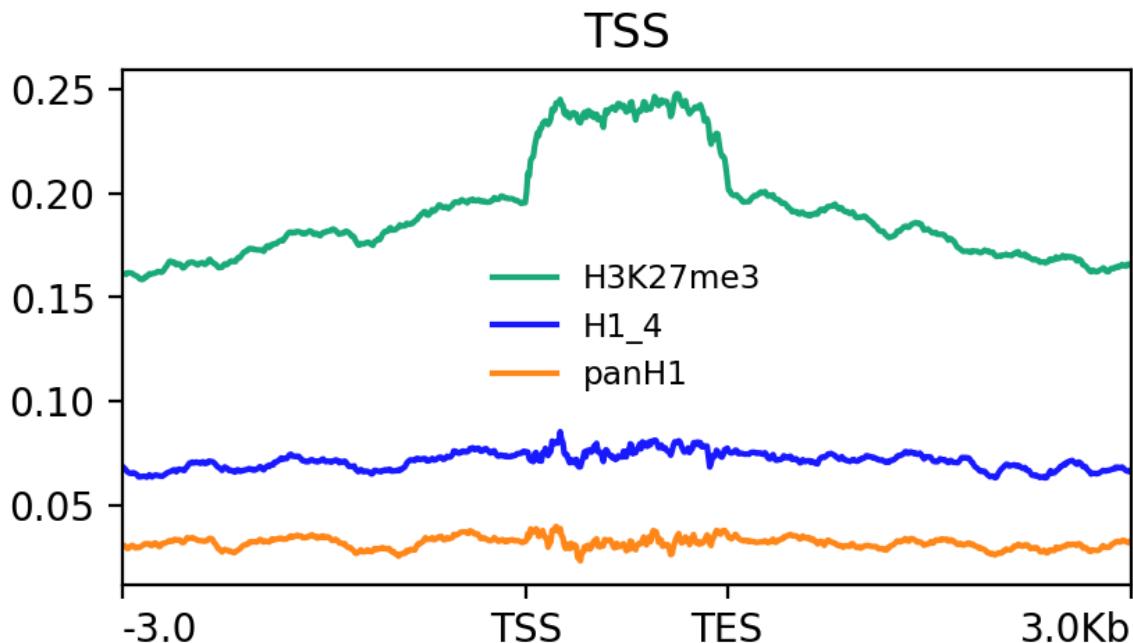
Plot for Cut&Tag For H3K27me3 from encode marked genes

```

computeMatrix scale-regions -S H3K27me3_IgG.bw H1_4_IgG.bw panH1_IgG.bw -R
H3K27me3_ENCFF801AHF_peaks_id_nearest_anno.txt -a 3000 -b 3000
--regionBodyLength 1500 --sortRegions keep --missingDataAsZero -o
cutntagwt_H3K27me3markedgegene_matrix.gz -p 20 --skipZeros

plotProfile -m cutntagwt_H3K27me3markedgegene_matrix.gz -out
cutntagwt_H3K27me3markedgegene_matrix.png --samplesLabel H3K27me3 H1_4
panH1 --perGroup --colors "#00A36C" "#0000FF" "#ff7f00"
--regionsLabel H3k37me3markedges

```



Cutnrun ko H3k27me3 marked genes

```
computeMatrix scale-regions -S H3K27me3_KO1_IgG.bw H3K27me3_KO2_IgG.bw
H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
H3K27me3_ENCFF801AHF_peaks_id_nearest_anno.txt -a 3000 -b 3000
--regionBodyLength 1500 --sortRegions keep --missingDataAsZero -o
cutntagko_H3K27me3markedgene_matrix.gz -p 20 --skipZeros

plotProfile -m cutntagko_H3K27me3markedgene_matrix.gz -out
cutntagko_H3K27me3markedgene_matrix.png --samplesLabel KO1 KO2 KO3
WT --perGroup --colors "#0000FF" "#0000FF" "#0000FF" "#ff7f00"
--regionsLabel H3k37me3markedgenes
```

Cutnrun kd H3k27me3 marked genes

```
computeMatrix scale-regions -S shControl_H3K27me3_IgG.bw
shCRAMP1_H3K27me3_IgG.bw shSUZ12_H3K27me3_IgG.bw -R
H3K27me3_ENCFF801AHF_peaks_id_nearest_anno.txt -a 3000 -b 3000
--regionBodyLength 1500 --sortRegions keep --missingDataAsZero -o
cutntagkd_H3K27me3markedgene_matrix.gz -p 20 --skipZeros

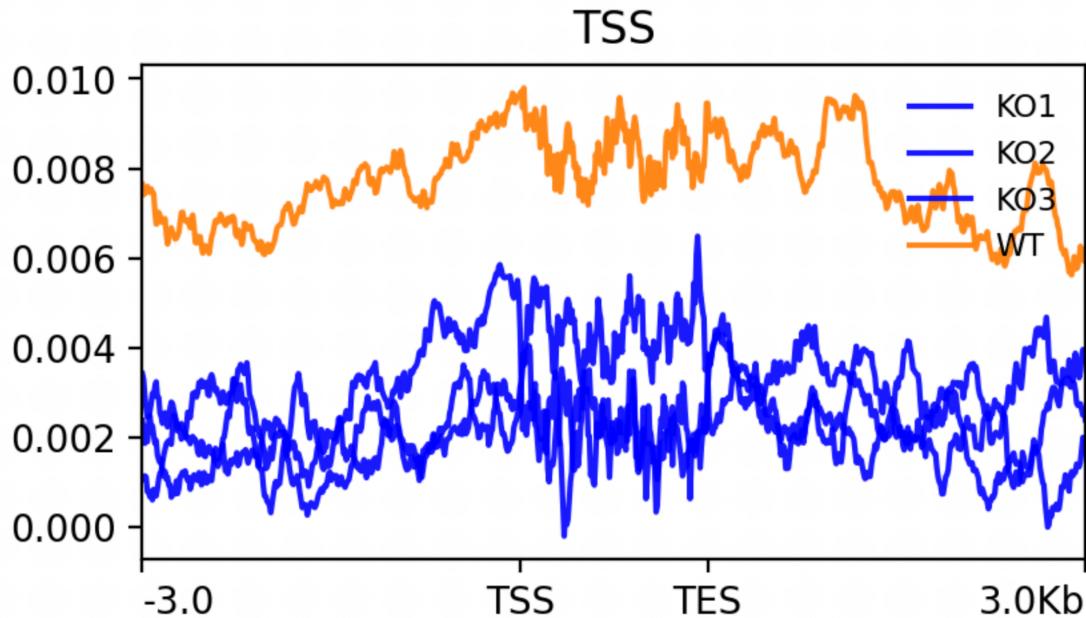
plotProfile -m cutntagkd_H3K27me3markedgene_matrix.gz -out
cutntagkd_H3K27me3markedgene_matrix.png --samplesLabel shControl
shCRAMP1 shSUZ12 --perGroup --colors "#ff7f00" "#0000FF"
"#00A36C" --regionsLabel H3k37me3markedgenes
```

Cutnrun ko gene body

```
computeMatrix scale-regions -S H3K27me3_KO1_IgG.bw
H3K27me3_KO2_IgG.bw H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
```

```
gene_gencodev41_gene.txt -a 3000 -b 3000 --regionBodyLength  
1500 --sortRegions keep --missingDataAsZero -o  
cutnrunko_gene_matrix.gz -p 20 --skipZeros
```

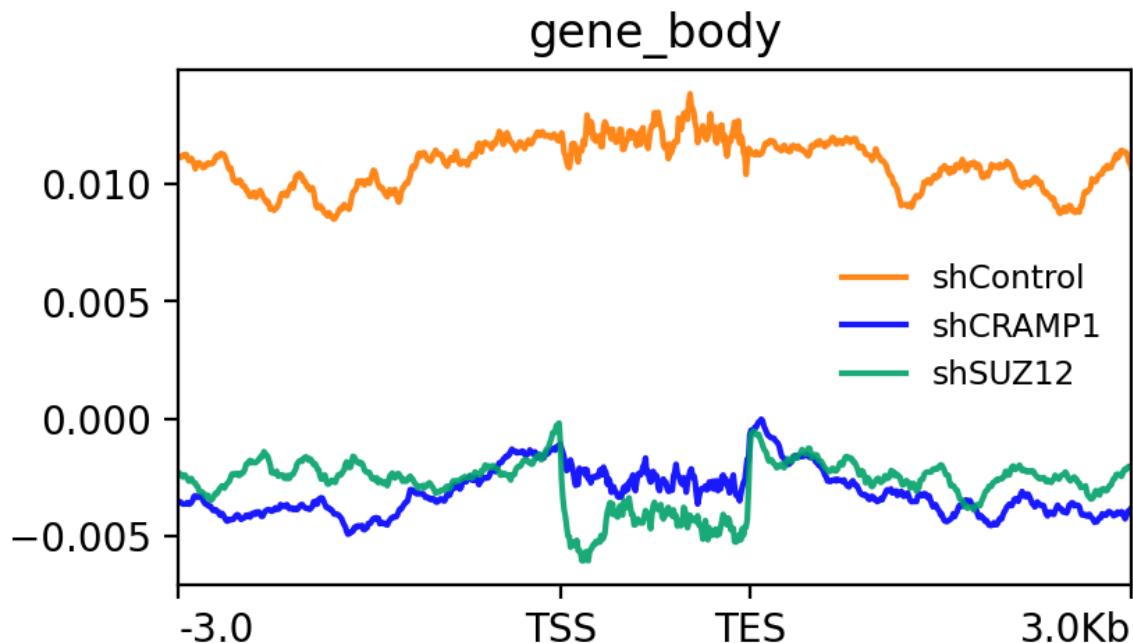
```
plotProfile -m cutnrunko_gene_matrix.gz -out  
cutnrunko_gene_matrix.png --samplesLabel KO1 KO2 KO3 WT  
--perGroup --colors "#0000FF" "#0000FF" "#0000FF" "#ff7f00"  
--regionsLabel TSS
```



Cutnrun kd gene body

```
computeMatrix scale-regions -S shControl_H3K27me3_IgG.bw  
shCRAMP1_H3K27me3_IgG.bw shSUZ12_H3K27me3_IgG.bw -R  
gene_gencodev41_gene.txt -a 3000 -b 3000 --regionBodyLength  
1500 --sortRegions keep --missingDataAsZero -o  
cutnrunkd_gene_matrix.gz -p 20 --skipZeros
```

```
plotProfile -m cutnrunkd_gene_matrix.gz -out  
cutnrunkd_geneTSS.png --samplesLabel shControl shCRAMP1  
shSUZ12 --perGroup --colors "#ff7f00" "#0000FF" "#00A36C"  
--regionsLabel gene_body
```



Cutnrun ko DEGs

```
computeMatrix scale-regions -S H3K27me3_KO1_IgG.bw
H3K27me3_KO2_IgG.bw H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
set_rnaseqde_merged_shC1_shS_up_pos_both_shC1_shS_up.txt
set_rnaseqde_merged_shC1_shS_up_pos_none_up.txt
set_rnaseqde_merged_shC1_shS_up_pos_only_shC1_up.txt
set_rnaseqde_merged_shC1_shS_up_pos_only_shS_up.txt -a 3000 -b
3000 --regionBodyLength 1500 --sortRegions keep
--missingDataAsZero -o cutnrunko_degene_matrix.gz -p 20
--skipZeros
```

```
plotProfile -m cutnrunko_degene_matrix.gz -out
cutnrunko_degene_matrix.png --samplesLabel KO1 KO2 KO3 WT
--perGroup --colors "#0000FF" "#0000FF" "#0000FF" "#ff7f00"
--regionsLabel both_shC1_shS_up none_up only_shC1_up
only_shS_up
```

```
computeMatrix scale-regions -S H3K27me3_KO1_IgG.bw
H3K27me3_KO2_IgG.bw H3K27me3_KO3_IgG.bw H3K27me3_WT_IgG.bw -R
set_rnaseqde_merged_shC1_shS_up_pos_both_shC1_shS_up.txt
set_rnaseqde_merged_shC1_shS_up_pos_none_up.txt
set_rnaseqde_merged_shC1_shS_up_pos_only_shS_up.txt -a 3000 -b
3000 --regionBodyLength 1500 --sortRegions keep
```

```
--missingDataAsZero -o cutnrunko_degene_matrix.gz -p 20
--skipZeros

plotProfile -m cutnrunko_degene_matrix.gz -out
cutnrunko_degene_matrix.png --samplesLabel KO1 KO2 KO3 WT
--perGroup --colors "#0000FF" "#0000FF" "#0000FF" "#ff7f00"
--regionsLabel both_shC1_shS_up none_up only_shS_up
-----

computeMatrix scale-regions -S FLAG_H12_R1.bigWig
FLAG_H15_R1.bigWig FLAG_H13_R1.bigWig FLAG_H1X_R1.bigWig
FLAG_H14_old_R1.bigWig FLAG_H14_R1.bigWig FLAG_old_R1.bigWig
shCRAMP1_H14_R1.bigWig shControl_H1_R1.bigWig
shSUZ12_IgG_R1.bigWig shCRAMP1_H1_R1.bigWig
shControl_IgG_R1.bigWig shCRAMP1_IgG_R1.bigWig
shSUZ12_H14_R1.bigWig shControl_H14_R1.bigWig
shSUZ12_H1_R1.bigWig WT_H13_R1.bigWig WT_IgG_R1.bigWig
WT_H12_R1.bigWig WT_H1X_R1.bigWig WT_H14_R1.bigWig
WT_H15_R1.bigWig -R gene_gencodev41_gene.txt -a 3000 -b 3000
--regionBodyLength 1500 --sortRegions keep --missingDataAsZero
-o cutntagwd_gene_matrix.gz -p 20 --skipZeros

plotProfile -m cutntagwd_gene_matrix.gz -out
cutntagwd_gene_matrix.png --samplesLabel FLAG_H12 FLAG_H15
FLAG_H13 FLAG_H1X FLAG_H14_old FLAG_H14 FLAG_old shCRAMP1_H14
shControl_H1 shSUZ12_IgG shCRAMP1_H1 shControl_IgG
shCRAMP1_IgG shSUZ12_H14 shControl_H14 shSUZ12_H1 WT_H13
WT_IgG WT_H12 WT_H1X WT_H14 WT_H15 --perGroup --colors
"#0000FF" "#0000FF" "#0000FF" "#0000FF" "#0000FF" "#0000FF"
--regionsLabel TSS
```

JAN 11

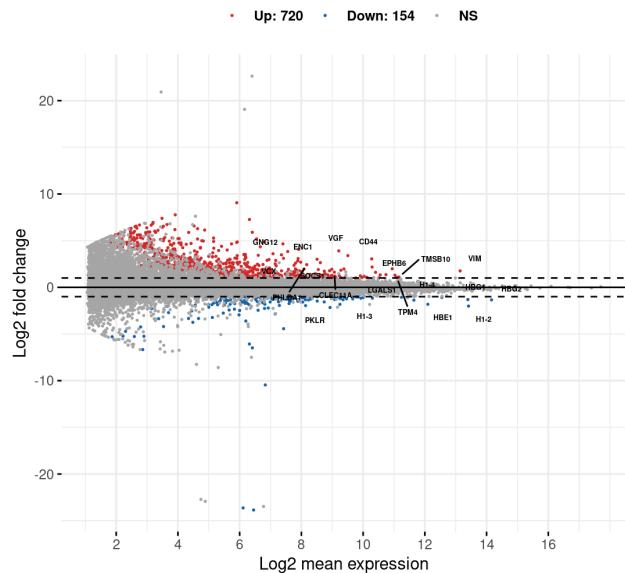
1. Repeat this with Cutnrun KD to see profile over ATAC-seq DARs. Also the overlap b/w cramp1 ad suz12 dars
2. CUTNTAG WT: Take genes marked by only H3K27me3
3. Repeat this with Cutnrun KD to see profile over gene body

31st January 2024

Key bits of hypothesis we need to evidence

- What analysis/figure needed to show each point.
- Is it complete? What was the answer?

1. What genes are regulated by cramp1?



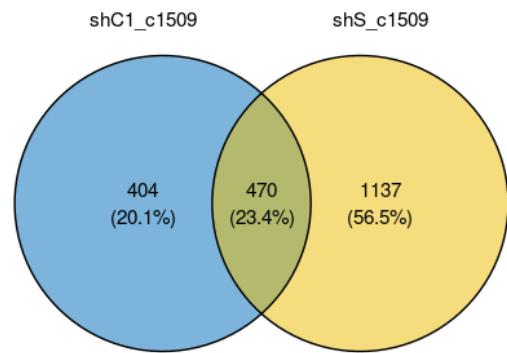
What are those genes? [Checked see the g profiler results and interpret](#)

Categorize what are Up and Down regulated genes? [Checked see the g profiler results and interpret](#)

Use Ifc shrinkage, to plot MA?? IfcSE are from 0,1

[Upload the values of RNASeq KD , in excel sheet in one drive](#)

2. Do these overlap with SUZ12 genes? Yes 470



Q. Venn diagram between Upregulated shC1 and Upregulated shS

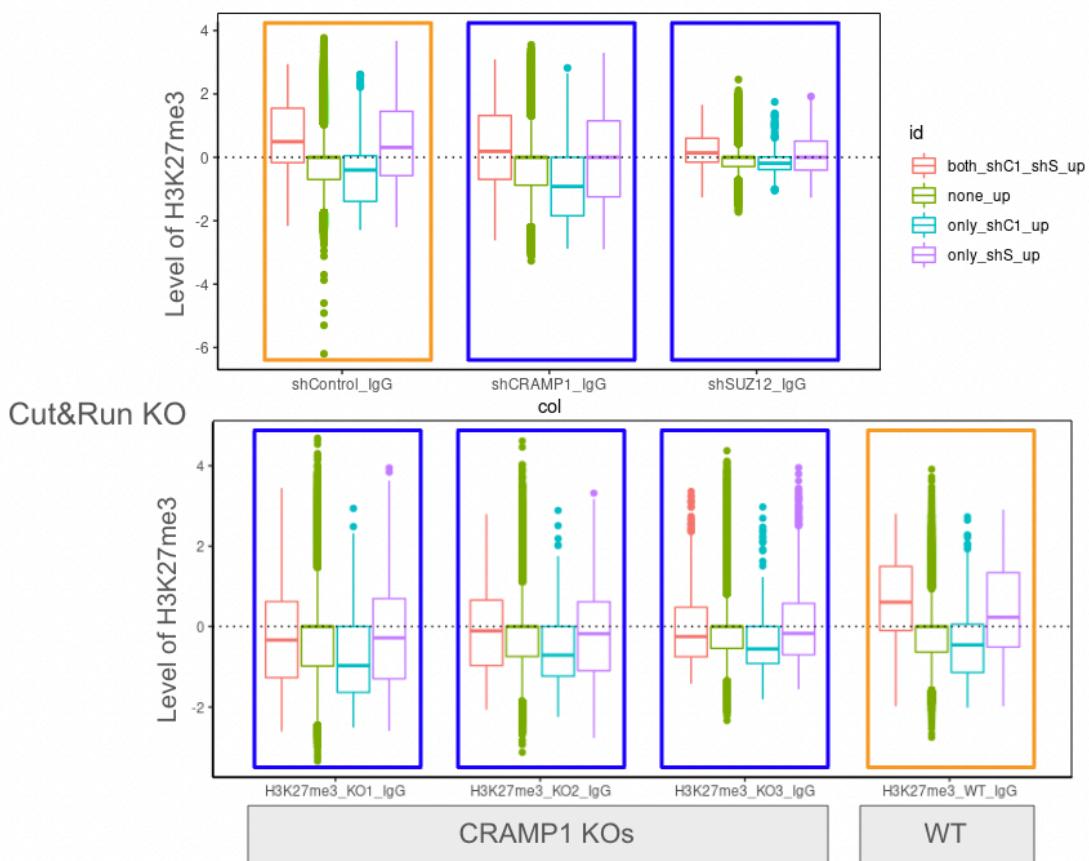
Q. Venn diagram between Deregulated shC1 and Deregulated shS

Q. Venn diagram between Downregulated shC1 and Downregulated shS

3. Are cramp1-regulated genes enriched for H3K27me3?

Remove none_up category

Cut&Run KD



Use density plot

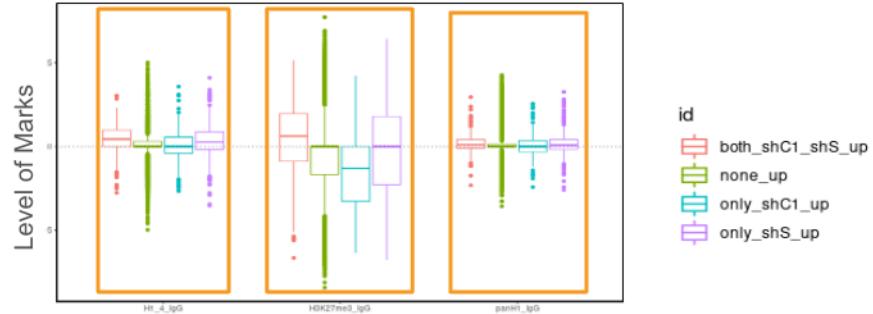
Hypergeometric Test		CUT&RUN KO Genes near the DBR Peaks		
RNA-Seq KD Genes	Conditions	DE	UP	DOWN
	shC1_up	0.9359138	0.6406263	0.9125469
	shS_up	0.0005829921	0.7364216	1.620076e-05
	shC1_shS_up	5.662137e-15	0.3929884	0

4. What genes does Cramp1 bind?

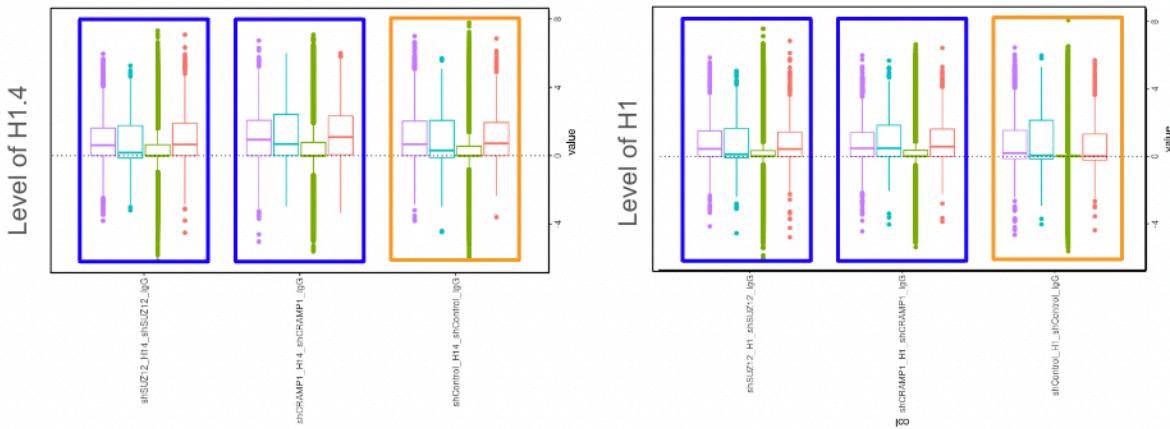
Many Histone genes, already done

5. Is H1/H1.4/PanH1 (WT) enriched at SUZ12/Cramp1 regulated genes?

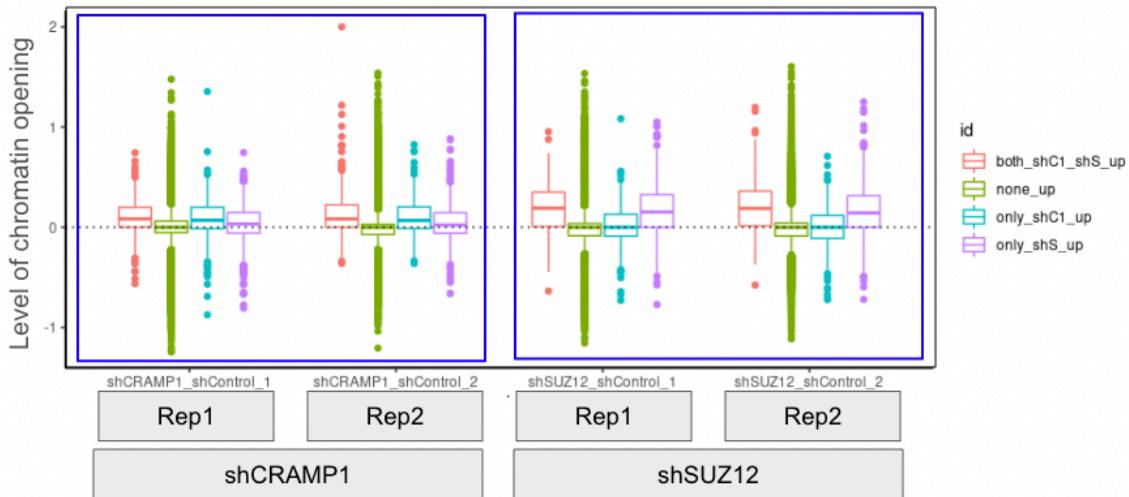
Cut&Tag WT



6. Does H1/H1.4/PanH1 go down at SUZ12/Cramp1 regulated genes when you knockdown Cramp1/SUZ12?



7. When we knockdown cramp1 do we see increased accessibility at SUZ12/Cramp1 regulated genes?



Feb 8 2024

- ATAC-Seq reads in H3K27me3 Peaks (+/- 500 bp) plot by my data, RM already did it but repeat
- Histone gene expression
- Obtain BAM files from RM for Cut&Tag data
 - Heatmap against histone marks for new CUT&TAG Data
 - Pearson correlation scatter plot (from the paper I presented)
 - Correlation plot corr: Genome wide, H3K27m3 peaks
 - PCA Plot: Genome wide, H3K27m3 peaks
 - Enrichment in H1.4 in different categories, ChIPseeker (Peak calling?)

What data need two-way transfer

- RNA-Seq: AV
- ATAC-Seq: AV BAMs and AV peak calling
- CUT&RUN KO: AV BAMs and RM will call the peaks
- CUT&RUN KD: AV BAMs and RM will call the peaks
- CUT&TAG WT: RM BAMs and RM will call the peaks (data not in use)

- CUT&TAG KD, & others: RM Mapping and RM will call the peaks (data not in use)
- CUT&TAG WT, & others: RM Mapping and RM will call the peaks

```
(base)
av638@cb-milan1:/mnt/bioinfo_sharing/sharing/tchasovnikarova/GBP0030/AV
/atacseq/kd$ cp -r
~/atacseq/iva_lab_gencode/boutfolder/bowtie2/merged_library/*bam .

(base)
av638@cb-milan1:/mnt/bioinfo_sharing/sharing/tchasovnikarova/GBP0030/AV
$ cp -r
~/cutnrun/iva_lab_oct23/cutnrun_k27_ko/outfolder/bowtie2/mergedLibrary/
*bam cutnrun/cutnrun_k27_ko/

(base)
av638@cb-milan1:/mnt/bioinfo_sharing/sharing/tchasovnikarova/GBP0030/AV
$ cp -r
~/cutnrun/iva_lab_oct23/cutnrun_k27_kd/outfolder/bowtie2/mergedLibrary/
*bam cutnrun/cutnrun_k27_kd/

(base)
av638@cb-milan1:/mnt/bioinfo_sharing/sharing/tchasovnikarova/GBP0030/AV
$ cp -r ~/rnaseq/iva_lab_oct23/outfolder/star_salmon/*.ba* rnaseq/
```

RNA-Seq_Oct2023

```
# ENCODE bams
# H3K9me3

samtools merge H3K9me3_ENCAN967RGD.bam ENCFF155UQU.bam ENCFF104THG.bam
samtools index H3K9me3_ENCAN967RGD.bam
bamCoverage -b H3K9me3_ENCAN967RGD.bam -o H3K9me3_ENCAN967RGD.bam.bw --normalizeUsing RPKM -p 20

samtools merge Control_ENCAN410JBG.bam ENCFF226FKB.bam ENCFF666YXZ.bam ENCFF420ARY.bam
samtools index Control_ENCAN410JBG.bam
bamCoverage -b Control_ENCAN410JBG.bam -o Control_ENCAN410JBG.bam.bw --normalizeUsing RPKM
-p 20

# H3K27me3

samtools merge H3K27me3_ENCAN563KAP.bam ENCFF652TXG.bam ENCFF508LLH.bam
samtools index H3K27me3_ENCAN563KAP.bam
bamCoverage -b H3K27me3_ENCAN563KAP.bam -o H3K27me3_ENCAN563KAP.bam.bw --normalizeUsing RPKM -p 20

cp ENCFF355LXQ.bam Control_ENCFF355LXQ.bam
samtools index Control_ENCFF355LXQ.bam
bamCoverage -b Control_ENCFF355LXQ.bam -o Control_ENCFF355LXQ.bam.bw --normalizeUsing RPKM
-p 20

# H3K4me3

samtools merge H3K4me3_ENCAN646APS.bam ENCFF855ZMQ.bam ENCFF752MYF.bam
samtools index H3K4me3_ENCAN646APS.bam
bamCoverage -b H3K4me3_ENCAN646APS.bam -o H3K4me3_ENCAN646APS.bam.bw --normalizeUsing RPKM
-p 20
```

```
# control is same for bradley Control_ENCAN410JBG.bam.bw (see below for comparison)

samtools merge H3K4me3_ENCAN750ONT.bam ENCF181ANT.bam ENCF747HEB.bam
samtools index H3K4me3_ENCAN750ONT.bam
bamCoverage -b H3K4me3_ENCAN750ONT.bam -o H3K4me3_ENCAN750ONT.bam.bw --normalizeUsing RPKM
-p 20

# control is same for peggy Control_ENCFF355LXQ.bam.bw (see below for comparison)
```

New CutnTag data Feb 2024

```
# Convert bam to bigwig
for var in *.bam
do
    echo ${var}
    bamCoverage -b ${var} -o ${var}.bw --normalizeUsing RPKM -p 20
done

# Normalize IP by control/IgG
-----
# subtraction normalised
# bigwig normalise against control, IgG
-----

bigwigCompare -b1 ITit257-RM-FH12_S1_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH12_S1_F_S6.bw --operation subtract
bigwigCompare -b1 ITit257-RM-FH13_S2_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH13_S2_F_S6.bw --operation subtract
bigwigCompare -b1 ITit257-RM-FH14_S3_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH14_S3_F_S6.bw --operation subtract
bigwigCompare -b1 ITit257-RM-FH15_S4_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH15_S4_F_S6.bw --operation subtract
bigwigCompare -b1 ITit257-RM-FH1X_S5_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH1X_S5_F_S6.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H12_S15_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H12_S15_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H13_S16_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H13_S16_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H14_S17_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H14_S17_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H15new_S19_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H15new_S19_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H15old_S18_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H15old_S18_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-H1X_S20_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H1X_S20_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-K27-1_S7_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o K27-1_S7_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-K27-2_S14_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o K27-2_S14_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-K27-3_S21_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o K27-3_S21_IgG_S22.bw --operation subtract
bigwigCompare -b1 ITit257-RM-V5H12_S8_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H12_S8_V5_S13.bw --operation subtract
bigwigCompare -b1 ITit257-RM-V5H13_S9_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H13_S9_V5_S13.bw --operation subtract
bigwigCompare -b1 ITit257-RM-V5H14_S10_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H14_S10_V5_S13.bw --operation subtract
bigwigCompare -b1 ITit257-RM-V5H15_S11_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H15_S11_V5_S13.bw --operation subtract
bigwigCompare -b1 ITit257-RM-V5H1X_S12_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H1X_S12_V5_S13.bw --operation subtract

Saved as performbigwigcompare_subtract.sh

Calculate correlation using deepTools
-----
All samples
multibigwigSummary bins -b FH12_S1_F_S6.bw FH13_S2_F_S6.bw FH14_S3_F_S6.bw FH15_S4_F_S6.bw FH1X_S5_F_S6.bw H12_S15_IgG_S22.bw
H13_S16_IgG_S22.bw H14_S17_IgG_S22.bw H15new_S19_IgG_S22.bw H15old_S18_IgG_S22.bw H1X_S20_IgG_S22.bw K27-1_S7_IgG_S22.bw K27-2_S14_IgG_S22.bw
K27-3_S21_IgG_S22.bw V5H12_S8_V5_S13.bw V5H13_S9_V5_S13.bw V5H14_S10_V5_S13.bw V5H15_S11_V5_S13.bw V5H1X_S12_V5_S13.bw -
cutntagrmwt_correlation_matrix_sub.npz -p 20 -bs 80000 -1 FH12 FH13 FH14 FH15 FH1X H12 H13 H14 H15new H15old H1X K27_1 K27_2 K27_3 V5H12 V5H13
V5H14 V5H15 V5H1X -v --outRawCounts cutntagrmwt_correlation_scores_per_bin_sub.tab

plotCorrelation -in cutntagrmwt_correlation_matrix_sub.npz -c pearson -p scatterplot -o cutntagrmwt_correlation_scatterplot_sub.png --skipZeros
--outFileCorMatrix cutntagrmwt_correlation_scatterplot_sub.txt

plotCorrelation -in cutntagrmwt_correlation_matrix_sub.npz -c pearson -p heatmap -o cutntagrmwt_correlation_heatmap_sub.png --skipZeros
--colorMap RdYlBu --outFileCorMatrix cutntagrmwt_correlation_heatmap_sub.txt

# Only H1s endogenous
multibigwigSummary bins -b H12_S15_IgG_S22.bw H13_S16_IgG_S22.bw H14_S17_IgG_S22.bw H15old_S18_IgG_S22.bw H1X_S20_IgG_S22.bw
K27-1_S7_IgG_S22.bw -o Hlendo_cutntagrmwt_correlation_matrix_sub.npz -p 20 -bs 80000 -1 H12 H13 H14 H15old H1X K27_1 -v --outRawCounts
Hlendo_cutntagrmwt_correlation_scores_per_bin_sub.tab

plotCorrelation -in Hlendo_cutntagrmwt_correlation_matrix_sub.npz -c pearson -p scatterplot -o
Hlendo_cutntagrmwt_correlation_scatterplot_sub.png --skipZeros --outFileCorMatrix Hlendo_cutntagrmwt_correlation_scatterplot_sub.txt

plotCorrelation -in Hlendo_cutntagrmwt_correlation_matrix_sub.npz -c pearson -p heatmap -o Hlendo_cutntagrmwt_correlation_heatmap_sub.png
--skipZeros --colorMap RdYlBu --outFileCorMatrix Hlendo_cutntagrmwt_correlation_heatmap_sub.txt

-----
# log2(IP/Input) normalised
# bigwig log2(IP/Input) normalise against control, IgG
-----
```

```
bigwigCompare -b1 ITit257-RM-FH12_S1_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH12_S1_F_S6.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-FH13_S2_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH13_S2_F_S6.divlog2.bw --operation log2
```

```

bigwigCompare -b1 ITit257-RM-FH14_S3_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH14_S3_F_S6.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-FH15_S4_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH15_S4_F_S6.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-FH15_S5_sorted.bam.bw -b2 ITit257-RM-F_S6_sorted.bam.bw -p 20 -o FH15_S5_F_S6.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H12_S15_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H12_S15_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H12_S16_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H12_S16_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H14_S17_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H14_S17_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H15new_S19_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H15new_S19_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H15old_S18_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H15old_S18_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-H1X_S20_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o H1X_S20_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-K27-1_S7_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o K27-1_S7_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-K27-2_S14_sorted.bam.bw -b2 ITit257-RM-IgG_S22_sorted.bam.bw -p 20 -o K27-2_S14_IgG_S22.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-V5H12_S8_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H12_S8_V5_S13.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-V5H13_S9_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H13_S9_V5_S13.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-V5H14_S10_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H14_S10_V5_S13.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-V5H15_S11_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H15_S11_V5_S13.divlog2.bw --operation log2
bigwigCompare -b1 ITit257-RM-V5H1X_S12_sorted.bam.bw -b2 ITit257-RM-V5_S13_sorted.bam.bw -p 20 -o V5H1X_S12_V5_S13.divlog2.bw --operation log2

Saved as performbigwigcompare_divlog2.sh

bigwigCompare -b1 H3K27me3_ENCANN563KAP.bam.bw -b2 Control_ENCFF355LXQ.bam.bw -p 20 -o H3K27me3_ENCANN563KAP_Control_ENCFF355LXQ.divlog2.bw
--operation log2
bigwigCompare -b1 H3K9me3_ENCANN967RGD.bam.bw -b2 Control_ENCANN410JBG.bam.bw -p 20 -o H3K9me3_ENCANN967RGD_Control_ENCANN410JBG.divlog2.bw
--operation log2
bigwigCompare -b1 H3K4me3_ENCANN646APS.bam.bw -b2 Control_ENCANN410JBG.bam.bw -p 20 -o H3K4me3_ENCANN646APS_Control_ENCANN410JBG.divlog2.bw
--operation log2
bigwigCompare -b1 H3K4me3_ENCANN750ONT.bam.bw -b2 Control_ENCFF355LXQ.bam.bw -p 20 -o H3K4me3_ENCANN750ONT_Control_ENCFF355LXQ.divlog2.bw
--operation log2

Calculate correlation using deepTools
-----
All samples
multibigwigSummary bins -b FH12_S1_F_S6.divlog2.bw FH13_S2_F_S6.divlog2.bw FH14_S3_F_S6.divlog2.bw FH15_S4_F_S6.divlog2.bw
FH15_S5_F_S6.divlog2.bw H12_S15_IgG_S22.divlog2.bw H13_S16_IgG_S22.divlog2.bw H14_S17_IgG_S22.divlog2.bw H15new_S19_IgG_S22.divlog2.bw
H15old_S18_IgG_S22.divlog2.bw H1X_S20_IgG_S22.divlog2.bw K27-1_S7_IgG_S22.divlog2.bw K27-2_S14_IgG_S22.divlog2.bw K27-3_S21_IgG_S22.divlog2.bw
V5H12_S8_V5_S13.divlog2.bw V5H13_S9_V5_S13.divlog2.bw V5H14_S10_V5_S13.divlog2.bw V5H15_S11_V5_S13.divlog2.bw V5H1X_S12_V5_S13.divlog2.bw -o
cutntagrmwt_correlation_matrix_divlog2.npz -p 20 -bs 80000 -l FH12 FH13 FH14 FH15 FH16 H13 H14 H15new H15old H1X K27_1_K27_2_K27_3 V5H12
V5H13 V5H14 V5H15 V5H1X -v --outRawCounts cutntagrmwt_correlation_scores_per_bin_divlog2.tab

plotCorrelation -in cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p scatterplot -o cutntagrmwt_correlation_scatterplot_divlog2.png
--skipZeros --outFileCorMatrix cutntagrmwt_correlation_scatterplot_divlog2.txt

plotCorrelation -in cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p heatmap -o cutntagrmwt_correlation_heatmap_divlog2.png --skipZeros
--colorMap RdYlBu --outFileCorMatrix cutntagrmwt_correlation_heatmap_divlog2.txt

-----
Only H1s endogenous
multibigwigSummary bins -b H12_S15_IgG_S22.divlog2.bw H13_S16_IgG_S22.divlog2.bw H14_S17_IgG_S22.divlog2.bw H15old_S18_IgG_S22.divlog2.bw
H1X_S20_IgG_S22.divlog2.bw K27-1_S7_IgG_S22.divlog2.bw -o H1endo_cutntagrmwt_correlation_matrix_divlog2.npz -p 20 -bs 80000 -l H12 H13 H14
H15old H1X K27_1 -v --outRawCounts H1endo_cutntagrmwt_correlation_scores_per_bin_divlog2.tab

plotCorrelation -in H1endo_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p scatterplot -o
H1endo_cutntagrmwt_correlation_scatterplot_divlog2.png --skipZeros --outFileCorMatrix H1endo_cutntagrmwt_correlation_scatterplot_divlog2.txt

plotCorrelation -in H1endo_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p heatmap -o
H1endo_cutntagrmwt_correlation_heatmap_divlog2.png --skipZeros --colorMap RdYlBu --outFileCorMatrix
H1endo_cutntagrmwt_correlation_heatmap_divlog2.txt

```

13 Feb 2024

1. Remove **none_up** category, Violin plots for all data.
2. Export data from **multibigwigsummary** and plot scatters one by one, V/Flag vs H1 endo.
3. Heatmap for only H1 endogenous
4. Do it for subtraction and division correlation
5. Send bigwigs to Rachael
(/mnt/bioinfo_sharing/sharing/tchasovnikarova/GBP0030/AV/cutntag/feb_2024)
6. Values for correlation for both subtraction and division coming exactly the same????? Check again. Solution: Just add operation subtraction and run all again: Resolved
7. For each endogenous H1 there should be 2 scatter plots, one for the FLAG and the other for the V5. E.g. for H1.2 there should be one scatterplot with H1.2 plotted against FLAG-H1.4 and another scatterplot with H1.2 plotted against V5-H1.2.
8. Annotate Peaks or Enrichment profile for each H1s with genomic features as suggested by RM (Rachael will send the peaks)

14 Feb 2024

1. Try spearmon rather than pearson correlation

- plotCorrelation -in cutntagrmwt_correlation_matrix_divlog2.npz -c spearman -p heatmap -o cutntagrmwt_correlation_heatmap_divlog2_sp.png --skipZeros --colorMap RdYlBu --outFileCorMatrix cutntagrmwt_correlation_heatmap_divlog2_sp.txt
- plotCorrelation -in Hlendo_cutntagrmwt_correlation_matrix_divlog2.npz -c spearman -p heatmap -o Hlendo_cutntagrmwt_correlation_heatmap_divlog2_sp.png --skipZeros --colorMap RdYlBu --outFileCorMatrix Hlendo_cutntagrmwt_correlation_heatmap_divlog2_sp.txt

2. Mean the histone marks value for heatmap

15 Feb 2024

1. How Peak calling performed? Macs2 broad with some cutoff (RM)

2. Repressive and active marks

- H2AFZ: repressive
- H3K27ac: active
- H3K27me3: repressive
- H3K36me3: elongation
- H3K4me1: enhancer active
- H3K4me2: 5'UTR active
- H3K4me3: active
- H3K79me2: active
- H3K9ac: active
- H3K9me1: repressive
- H3K9me3: repressive
- H4K20me1: active

3. Remove H1.5 new

4. Keep FLAG/V5?Endo together. Later on as it is a representation

5. Enhancer annotation K562

6. ATAC-Seq DARs : See which one gives differences from DESeq2, EdgeR, Diffbind results

7. ATAC-Seq reads quantification with ENCODE H3K27me3 peaks and Other histone marks USE RPKM because length of each peaks is different within histone marks

```
# Convert bam to bigwig
for var in *.bam
do
echo ${var}
bamCoverage -b ${var} -o ${var}.bw --normalizeUsing RPKM -p 20
done
```

```
# log2(IP/Input) normalised
# bigwig log2(IP/Input) normalise against control, IgG
```

```
shControl_REPO.mLb.clN.sorted.bam.bw
shControl_REPO2.mLb.clN.sorted.bam.bw
shCRAMP1_REPO.mLb.clN.sorted.bam.bw
shCRAMP1_REPO2.mLb.clN.sorted.bam.bw
```

```
shSUZ12_REPO1.mLb.clN.sorted.bam.bw  
shSUZ12_REPO2.mLb.clN.sorted.bam.bw
```

```
bigwigCompare -b1 shCRAMP1_REPO1.mLb.clN.sorted.bam.bw -b2  
shControl_REPO1.mLb.clN.sorted.bam.bw -p 20 -o shCRAMP1_shControl_REPO1.divlog2.bw  
--operation log2  
bigwigCompare -b1 shCRAMP1_REPO2.mLb.clN.sorted.bam.bw -b2  
shControl_REPO2.mLb.clN.sorted.bam.bw -p 20 -o shCRAMP1_shControl_REPO2.divlog2.bw  
--operation log2  
bigwigCompare -b1 shSUZ12_REPO1.mLb.clN.sorted.bam.bw -b2  
shControl_REPO1.mLb.clN.sorted.bam.bw -p 20 -o shSUZ12_shControl_REPO1.divlog2.bw  
--operation log2  
bigwigCompare -b1 shSUZ12_REPO2.mLb.clN.sorted.bam.bw -b2  
shControl_REPO2.mLb.clN.sorted.bam.bw -p 20 -o shSUZ12_shControl_REPO2.divlog2.bw  
--operation log2
```

Saved as performbigwigcompare_divlog2.sh

Calculate correlation using deepTools

```
computeMatrix scale-regions -S shCRAMP1_shControl_REPO1.divlog2.bw  
shCRAMP1_shControl_REPO2.divlog2.bw shSUZ12_shControl_REPO1.divlog2.bw  
shSUZ12_shControl_REPO2.divlog2.bw -R H2AFZ_ENCFF213OTI_peaks_id.bed  
H3K27ac_ENCFF544LXB_peaks_id.bed H3K27me3_ENCFF323WOT_peaks_id.bed  
H3K27me3_ENCFF801AHF_peaks_id.bed H3K36me3_ENCFF193ERO_peaks_id.bed  
H3K36me3_ENCFF561OUZ_peaks_id.bed H3K4me1_ENCFF135ZLM_peaks_id.bed  
H3K4me1_ENCFF540NGG_peaks_id.bed H3K4me2_ENCFF749KLQ_peaks_id.bed  
H3K4me3_ENCFF122CSI_peaks_id.bed H3K4me3_ENCFF689QIJ_peaks_id.bed  
H3K4me3_ENCFF706WUF_peaks_id.bed H3K4me3_ENCFF885FQN_peaks_id.bed  
H3K79me2_ENCFF209OQD_peaks_id.bed H3K9ac_ENCFF148UQI_peaks_id.bed  
H3K9ac_ENCFF891CHI_peaks_id.bed H3K9me1_ENCFF462AVD_peaks_id.bed  
H3K9me3_ENCFF963GZJ_peaks_id.bed H4K20me1_ENCFF909RKY_peaks_id.bed -a 0  
-b 0 --skipZeros -o atacseqkd_histonemarks_mat.gz -p 20 --sortRegions  
keep --missingDataAsZero --outFileNameMatrix  
atacseqkd_histonemarks_mat.tab --outFileSortedRegions  
atacseqkd_histonemarks_mat_sortedRegions.bed  
  
plotProfile -m atacseqkd_histonemarks_mat.gz -out  
profile_atacseqkd_histonemarks_mat.png --numPlotsPerRow 1  
--samplesLabel shCRAMP1_1 shCRAMP1_2 shSUZ12_1 shSUZ12_2  
--regionsLabel H2AFZ_1 H3K27ac_1 H3K27me3_1 H3K27me3_2 H3K36me3_1  
H3K36me3_2 H3K4me1_1 H3K4me1_2 H3K4me2_1 H3K4me3_1 H3K4me3_2 H3K4me3_3  
H3K4me3_4 H3K79me2_1 H3K9ac_1 H3K9ac_2 H3K9me1_1 H3K9me3_1 H4K20me1_1  
--plotHeight 20 --plotWidth 15  
  
plotHeatmap -m atacseqkd_histonemarks_mat.gz -out  
heatmap_atacseqkd_histonemarks_mat.png --sortUsing max
```

```
--samplesLabel shCRAMP1_1 shCRAMP1_2 shSUZ12_1 shSUZ12_2
--regionsLabel H2AFZ_1 H3K27ac_1 H3K27me3_1 H3K27me3_2 H3K36me3_1
H3K36me3_2 H3K4me1_1 H3K4me1_2 H3K4me2_1 H3K4me3_1 H3K4me3_2 H3K4me3_3
H3K4me3_4 H3K79me2_1 H3K9ac_1 H3K9ac_2 H3K9me1_1 H3K9me3_1 H4K20me1_1
```

20 Feb 2024

1. Crispr screen genome wide
2. Reporter assay Iva lab?
3. Depmap
- 4. Remove H3k27me3 from histone mark**
- 5. Add H3K9me3 and H3K4me3 in correlation plot**

a. **Only H1s endogenous**

```
multibigwigSummary bins -b H12_S15_IgG_S22.divlog2.bw H13_S16_IgG_S22.divlog2.bw H14_S17_IgG_S22.divlog2.bw
H15old_S18_IgG_S22.divlog2.bw H1X_S20_IgG_S22.divlog2.bw K27-1_S7_IgG_S22.divlog2.bw
H3K27me3_ENCAN563KAP_Control_ENCFF355LXQ.divlog2.bw H3K9me3_ENCAN967RGD_Control_ENCAN410JBG.divlog2.bw
H3K4me3_ENCAN64GAPS_Control_ENCAN410JBG.divlog2.bw H3K4me3_ENCAN750ONT_Control_ENCFF355LXQ.divlog2.bw -o
Hlendo_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -p 20 -bs 80000 -1 H12 H13 H14 H15old H1X K27_1 enK27 enK9
enK4me3_1 enK4me3_2 -v --outRawCounts Hlendo_ENC_cutntagrmwt_correlation_scores_per_bin_divlog2.tab

plotCorrelation -in Hlendo_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p scatterplot -o
Hlendo_ENC_cutntagrmwt_correlation_scatterplot_divlog2.png --skipZeros --outFileCorMatrix
Hlendo_ENC_cutntagrmwt_correlation_scatterplot_divlog2.txt

plotCorrelation -in Hlendo_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p heatmap -o
Hlendo_ENC_cutntagrmwt_correlation_heatmap_divlog2.png --colorMap RdYlBu --outFileCorMatrix
Hlendo_ENC_cutntagrmwt_correlation_heatmap_divlog2.txt --skipZeros

All samples
multibigwigSummary bins -b FH12_S1_F_S6.divlog2.bw FH13_S2_F_S6.divlog2.bw FH14_S3_F_S6.divlog2.bw
FH15_S4_F_S6.divlog2.bw FH1X_S5_F_S6.divlog2.bw H12_S15_IgG_S22.divlog2.bw H13_S16_IgG_S22.divlog2.bw
H14_S17_IgG_S22.divlog2.bw H15new_S19_IgG_S22.divlog2.bw H15old_S18_IgG_S22.divlog2.bw H1X_S20_IgG_S22.divlog2.bw
K27-1_S7_IgG_S22.divlog2.bw K27-2_S14_IgG_S22.divlog2.bw K27-3_S21_IgG_S22.divlog2.bw V5H12_S8_V5_S13.divlog2.bw
V5H13_S9_V5_S13.divlog2.bw V5H14_S10_V5_S13.divlog2.bw V5H15_S11_V5_S13.divlog2.bw V5H1X_S12_V5_S13.divlog2.bw
H3K27me3_ENCAN563KAP_Control_ENCFF355LXQ.divlog2.bw H3K9me3_ENCAN967RGD_Control_ENCAN410JBG.divlog2.bw
H3K4me3_ENCAN64GAPS_Control_ENCAN410JBG.divlog2.bw H3K4me3_ENCAN750ONT_Control_ENCFF355LXQ.divlog2.bw -o
all_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -p 20 -bs 80000 -1 FH12 FH13 FH14 FH15 FH1X H12 H13 H14 H15new H15old
H1X K27_1 K27_2 K27_3 V5H12 V5H13 V5H14 V5H15 V5H1X enK27 enK9 enK4me3_1 enK4me3_2 -v --outRawCounts
all_ENC_cutntagrmwt_correlation_scores_per_bin_divlog2.tab

plotCorrelation -in all_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p scatterplot -o
all_ENC_cutntagrmwt_correlation_scatterplot_divlog2.png --skipZeros --outFileCorMatrix
all_ENC_cutntagrmwt_correlation_scatterplot_divlog2.txt

plotCorrelation -in all_ENC_cutntagrmwt_correlation_matrix_divlog2.npz -c pearson -p heatmap -o
all_ENC_cutntagrmwt_correlation_heatmap_divlog2.png --skipZeros --colorMap RdYlBu --outFileCorMatrix
all_ENC_cutntagrmwt_correlation_heatmap_divlog2.txt
```

6. Try these H1 figures with old H1 data
- 7. Represent violin gene by gene category (rnaseqkd)**
8. Merge violin plots reps in atacseqkd
9. Did you get perturbation of H1-2, H1-4, H1-3 in shCRAMP1?