Icahn School of Medicine at Mount Sinai

The Tisch Cancer Institute

NCI Designated Cancer Center

Black Family Stem Cell Institute

Skin Biology and Diseases Resource-based Center

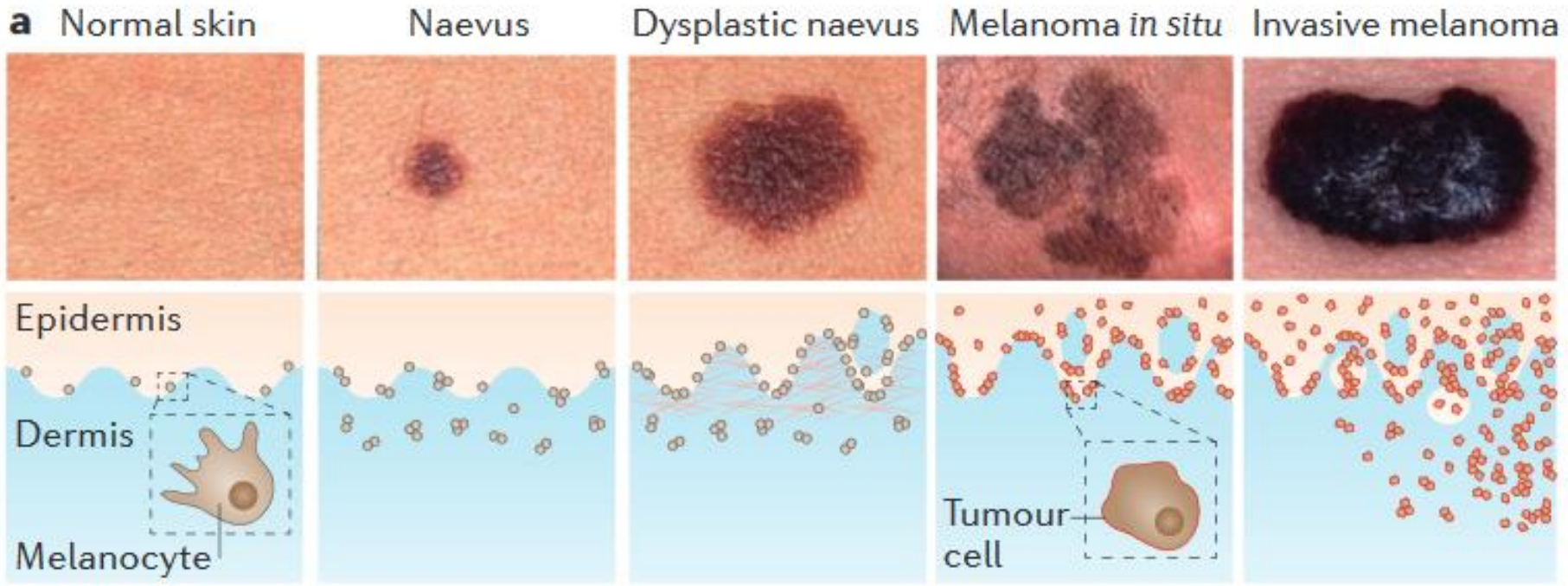
Department of Cell, Developmental and Regenerative Biology

Center for Advancement of Blood Cancer Therapies

BiNGS

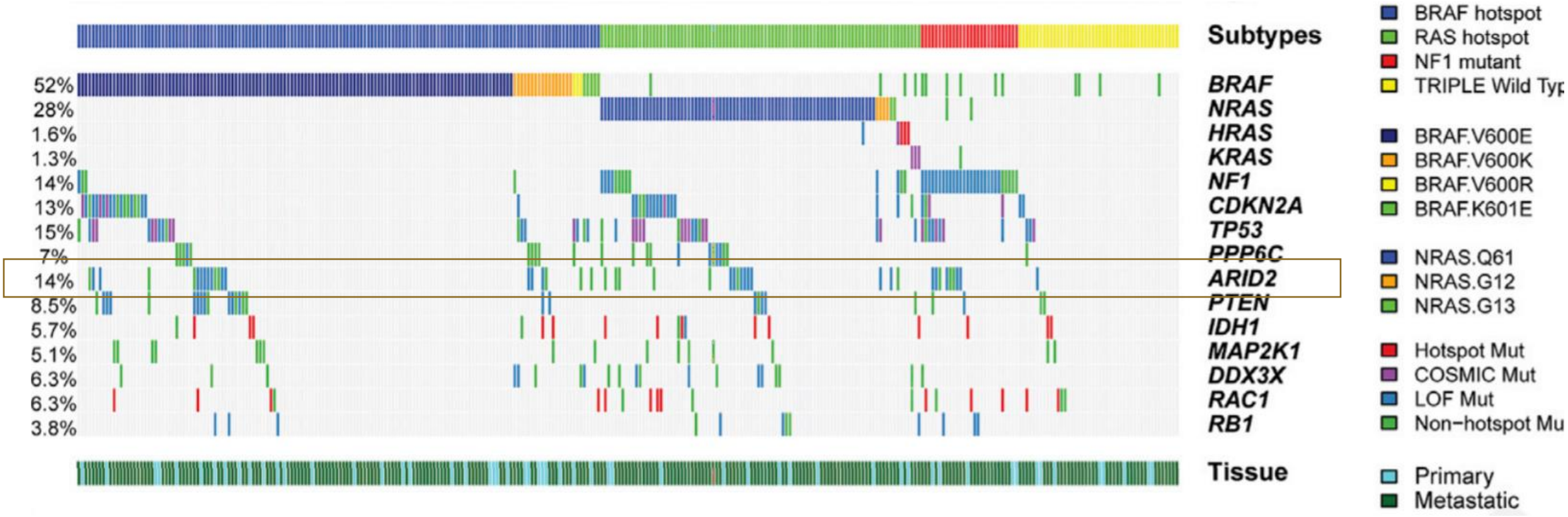
EMBL epigenetics Practicum Introduction

Melanoma is a malignancy of the pigment-producing cells melanocytes



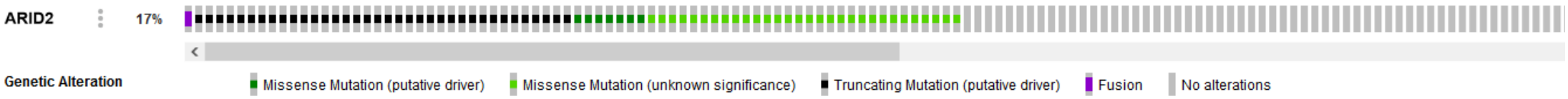
Shain AH, Bastian BC. From melanocytes to melanomas. Nat Rev Cancer. 2016

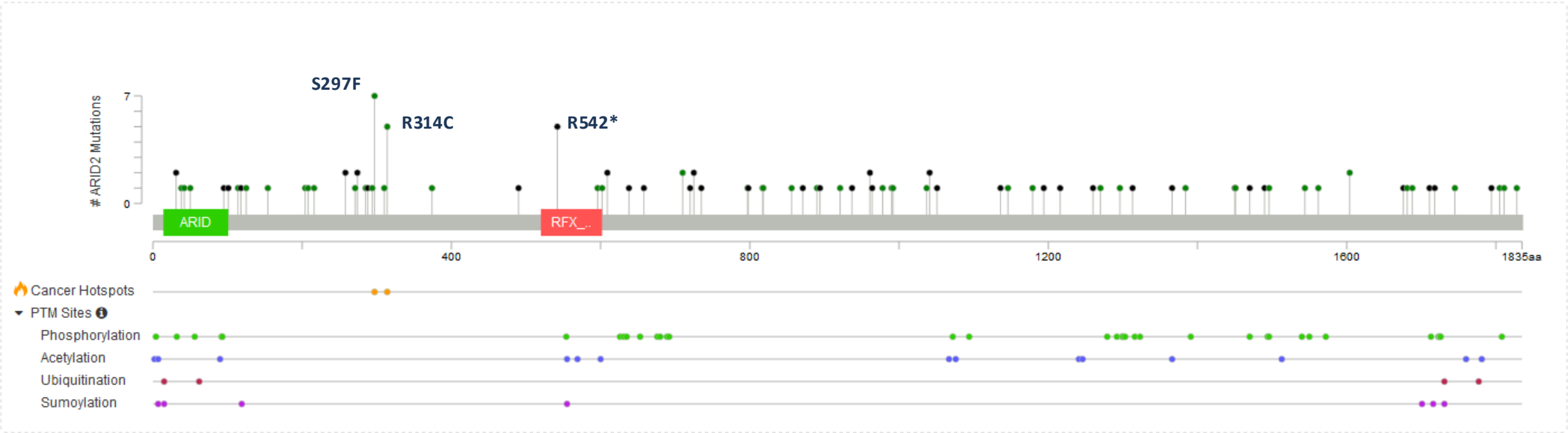
ARID2 is one of the most frequently mutated genes in melanoma



Cancer Genome Atlas Network.. Genomic Classification of Cutaneous Melanoma. Cell. 2015

ARID2 mutations spectrum in melanoma





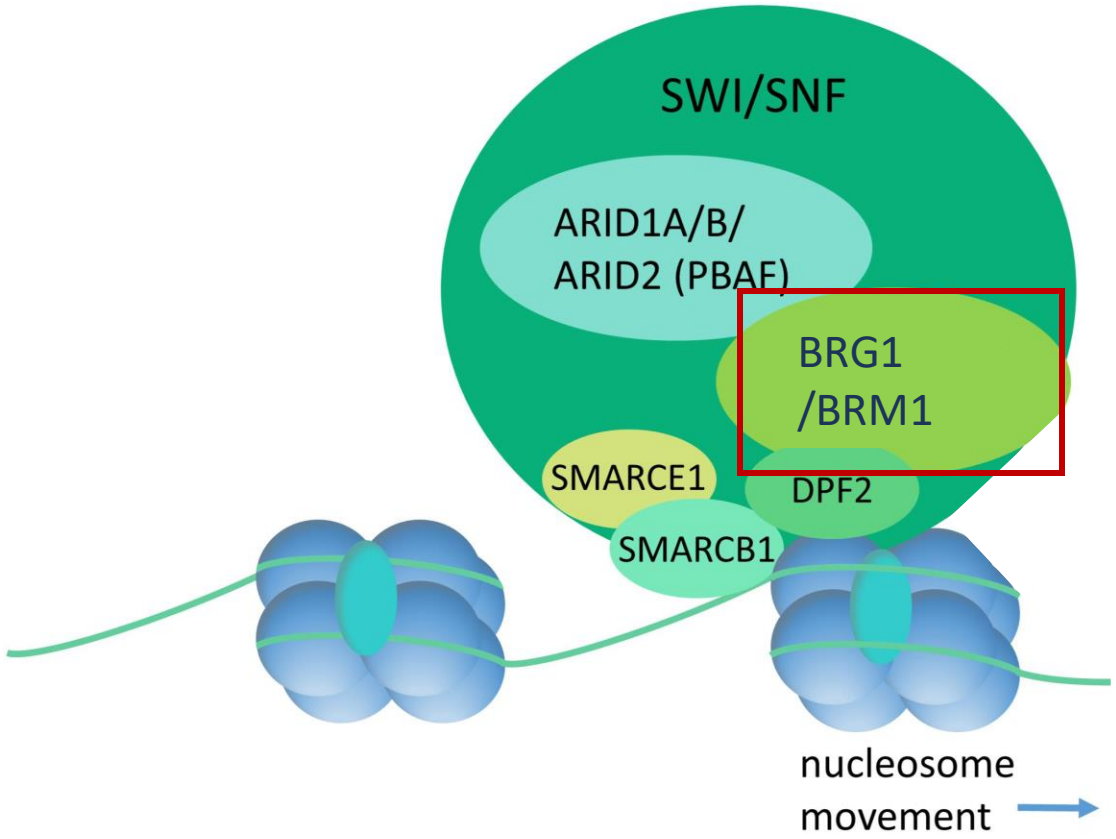
ARID2

RefSeq: NM_152641
Ensembl: ENST00000334344
CCDS: CCDS31783
UniProt: ARID2_HUMAN

Somatic Mutation Frequency: 16.8%

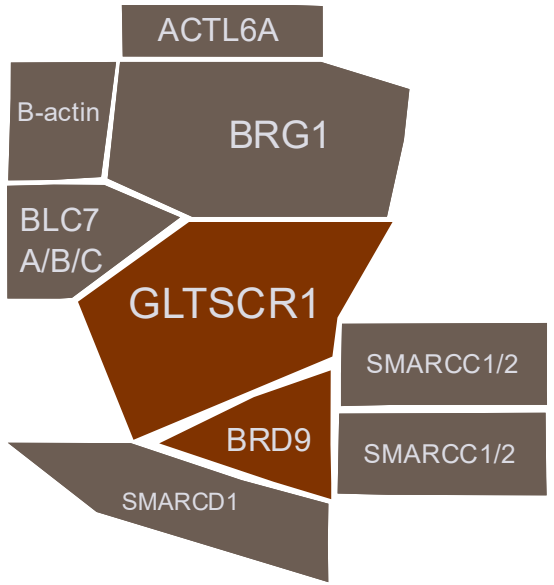
58 Missense 46 Truncating
0 Inframe 1 Other

The SWI/SNF complexes their and ATPase subunits

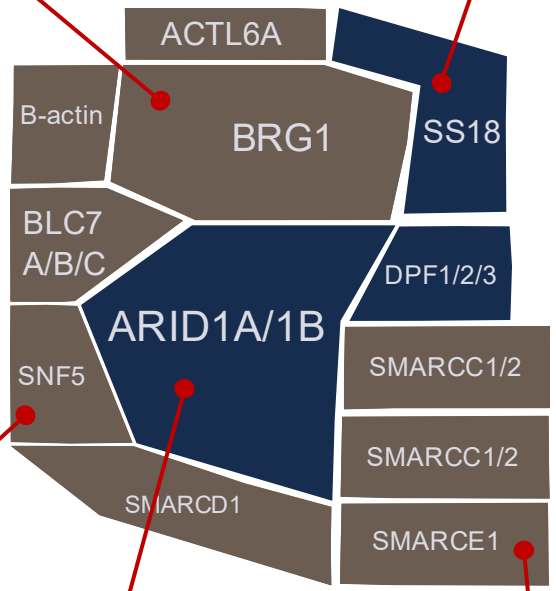


Bögershausen N, Wollnik B. Mutational Landscapes and Phenotypic Spectrum of SWI/SNF-Related Intellectual Disability Disorders. *Front Mol Neurosci*. 2018

SWI/SNF complex subunits are mutated in multiple cancers



Lung cancer
Breast cancer
Medulloblastoma



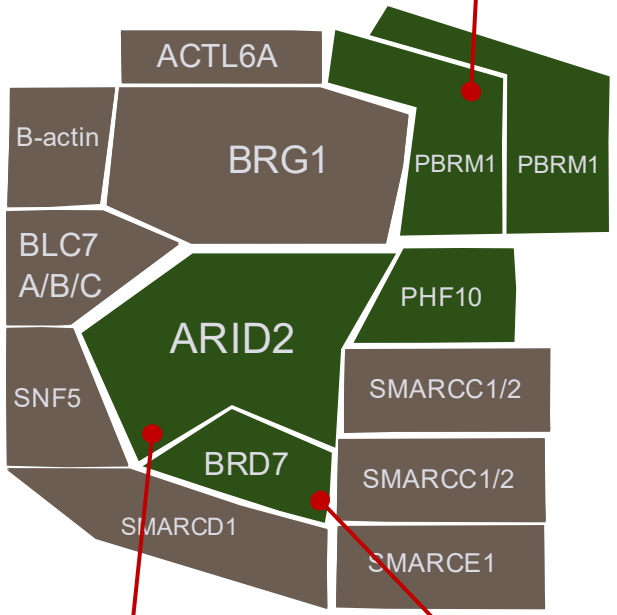
Synovial sarcoma

Malignant rhabdoid tumors

Ovaria clear cell carcinoma
Hepatocellular carcinoma
Endometroid cancer
Bladder cancer

Spinal meningioma

Renal clear cell carcinoma



Breast cancer

Melanoma
Hepatocellular carcinoma

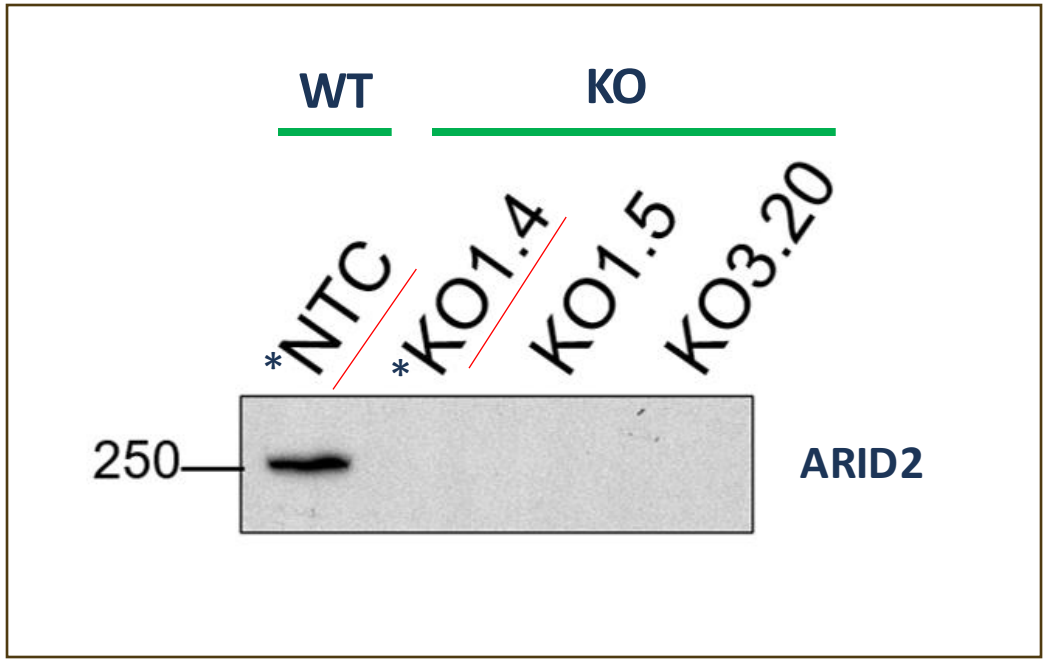
Hypothesis

ARID2 loss alters SWI/SNF occupancy and chromatin accessibility which alters transcription factor binding patterns, and results in aberrant transcriptional programs leading to increased metastases.

Model

- Cell line:
  - SKmel147 – melanoma cell line
  - NRAS Q61R
- Conditions:
  - WT (CRISPR sgControl)
  - KO (CRISPR sgARID2)
- Growing conditions:
  - 37 ℃, 5% CO_2
  - DMEM (Dulbecco Media), %5 penicillin/streptomycin, 10% Fetal bovine serum
- Assay
  - ATAC-seq

Western Blot



*Samples used for ATAC-seq analysis

Assignment expectations

- Perform quality control, alignment and filtering of data generate from ATAC-seq
- Perform differential chromatin accessibility analysis and integrate with RNAseq differential expression results.
- Visualize differential chromatin accessibility results and its integration with RNAseq results.
- Perform motif analysis of differentially accessible regions.
- Perform functional analysis of genes associated with differentially accessible regions
- Annotate chromatin accessible regions with their functional genomic region and their nearest gene (Homer)

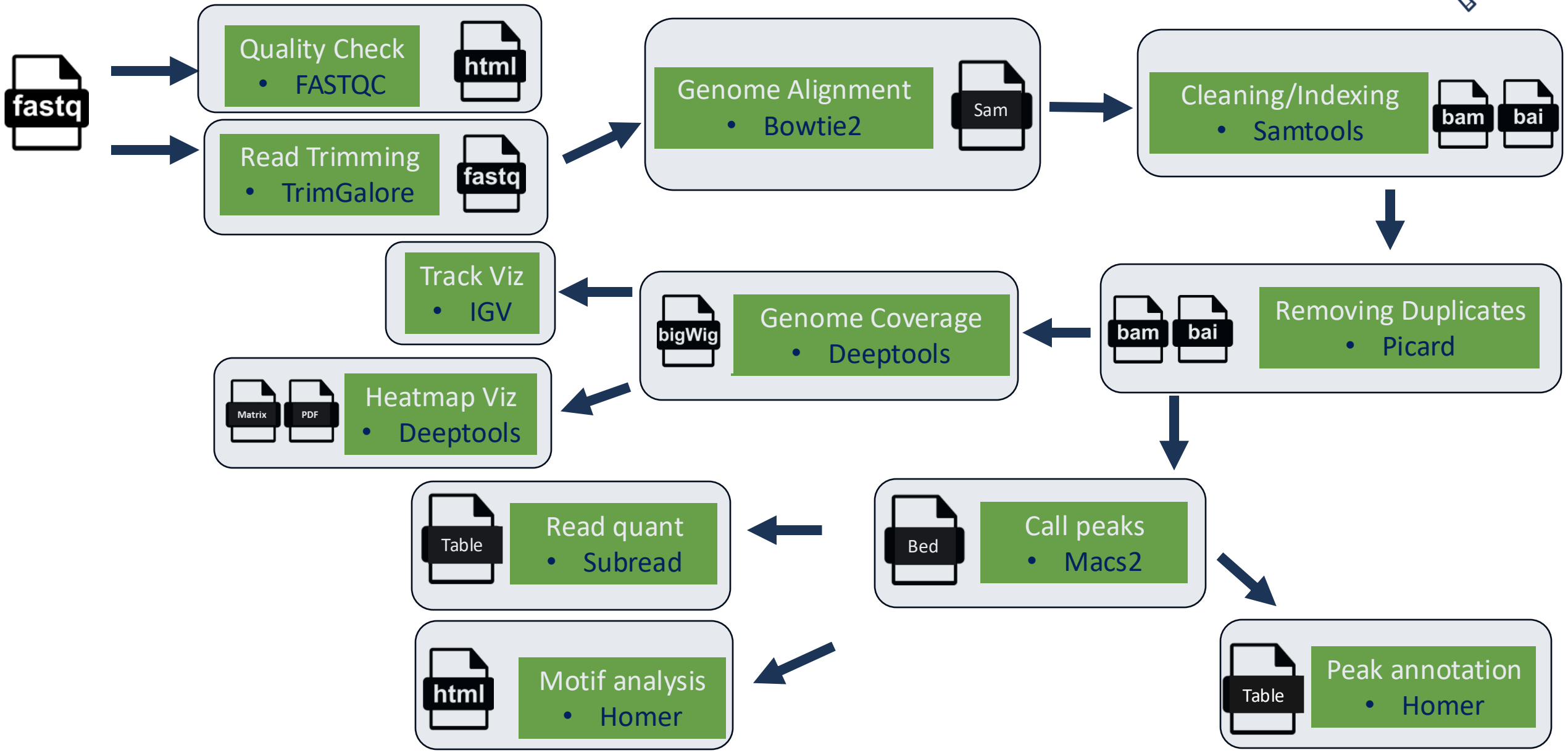
Module goals

- Develop a basic understanding on best practices for quality assessment of chromatin profiling methods.
- Become familiar with the processing steps of ATAC-seq analysis, the aim of each step and the specific arguments/functions used.
- Obtain a basic understanding of the biology of chromatin remodelers, transcriptions factor binding, chromatin profiling and their dynamic correlation.
- Familiarization with basic visualization practices for chromatin profiling methods.

Questions to drive the analysis

- Characterize the chromatin landscape at open regions?
- Which transcription factors are enriched at open regions? How many differentially accessible regions are between ARID2 WT and KO cells?
- What genes/pathways are enriched nearby the differential accessible regions?
- Which transcription factors are enriched at differential regions?
- What is the relationship between differentially open regions and differentially expressed genes?
- Hypothesize about the biological output (phenotype) produced by that changes that occur at the chromatin level.

Workflow


fastq
Quality Check
• FASTQC
html
Read Trimming
• TrimGalore
fastq
Genome Alignment
• Bowtie2
Sam
Cleaning/Indexing
• Samtools
bam
bai
Removing Duplicates
• Picard
bam
bai
Genome Coverage
• Deeptools
bigWig
Track Viz
• IGV
Heatmap Viz
• Deeptools
Matrix
PDF
Call peaks
• Macs2
Bed
Read quant
• Subread
Table
Motif analysis
• Homer
html
Peak annotation
• Homer
Table

Pre-processing ATACseq Steps

- Adaptor removal: **TrimGalore**
- Alignment to GRCH38: **Bowtie2**
- Filtering/Sorting/Indexing sam files: **Samtools**
- Duplicate Removal: **Picard**
- Sorting/Indexing sam files: **Samtools**
- Generation of coverage tracks: **Deeptools**
- Calling significant peaks: **Macs2**
- Heatmap Visualization: **Deeptools**
- Coverage Visualization**: UCSC browser/IGV**
- Quantification of reads in peak: **Subread**
- Differential peak analysis: **AWK/other**
- Motif enrichment analysis: **Homer**
- Gene peak association (RNAseq results): **Bedtools**
- Pathway analysis: **Enrichr**

Dataset integration (RNAseq/ChIPseq/ATACseq)

- Association of differential ATACseq regions with differential genes
- Visualization of ChIPseq enrichment at differential regions

Files provided – Data to analyze

- Fastq files (ATACseq)
  - SKmel147 ARID2 WT ATAC-seq (2 files R1 and R2)
  - SKmel147 ARID2 KO ATAC-seq (2 files R1 and R2)

- Bigwig files (ChIPseq)
  - SKmel147 ARID2 WT ARID2
  - SKmel147 ARID2 WT FOSL2
  - SKmel147 ARID2 WT H3K4me3
  - SKmel147 ARID2 WT H3K27ac

- Table
  - RNAseq results for the differential gene expression analysis between SKmel147 ARID2 WT vs SKmel147 ARID2 KO

Files provided – Containers and software

#Containers

- bedtools_v2.31.1
- bowtie2_v2.4.5.sif
- deeptools_v3.5.1.sif
- fastqc_v0.11.8.sif
- homer.sif
- macs_v2.2.9.1.sif
- picard_v2.9.2.sif
- samtools_v1.15.sif
- subread_v2.0.1.sif
- trim-galore_v0.6.9.sif

#Other

- Singularity/3.11.0
- AWK
- IGV

Files provided – Supporting files

- GRCH38 blacklisted regions
  - Regions defined by the Encode consortium with challenging mapping characteristics.

- GRCH38 -\+ 1 Kb promoters
  - Bed files of the promoter of genes for the GRCH38 genome version

- Bowtie2 Index
  - Files need for alignment of FASTQs with Bowtie2 corresponding to the GRCH38 genome version
- Code
  - “Draft” version of code to be completed with the resources provided for the course.

Files – Paired end ATAC-seq

ATAC-seq

- Melanoma cell line SKmel147
- Reads only from chromosome 5
- CRISPR ARID2KO vs ARID2 WT cell lines

File name	Cell line	Genotype
SKMel147_ARID2WT_ATAC_chr5_R1.sam	SKmel147 (Melanoma)	ARID2 WT
SKMel147_ARID2KO_ATAC_chr5_R1.sam	SKmel147 (Melanoma)	ARID2 KO

Files – Single end ChIPseq

ChIPseq-seq

- Melanoma cell line SKmel147
- Reads only from chromosome 5
- Parental cell lines

Bigwigs

File name	Cell line	Genotype
SKmel147-H3K4me3_chr5.bw	SKmel147 (Melanoma)	ARID2 WT
SKmel147-H3K27ac_chr5.bw	SKmel147 (Melanoma)	ARID2 WT
SKmel147-FOSL2_chr5.bw	SKmel147 (Melanoma)	ARID2 WT
SKmel147-ARID2_chr5.bw	SKmel147 (Melanoma)	ARID2 WT

Files – RNAseq

ChIPseq-seq

- Melanoma cell line SKmel147
- Genes only from chromosome 5
- CRISPR ARID2KO vs ARID2 WT cell lines

Deseq results table

File name	Cell line
Deseq_Results_ARID2Ko_v_ARID2wt_chr5.csv	SKmel147 (Melanoma)

Script preparation

```
#define project directory
projectdir="/home/training/project_dir"

#practical data dir
practical_dir="/home/training/bulk_epigenetics_practical"

#container directory
container_dir=$practical_dir/"container/"

#blacklist_dir=
bl_regions=$practical_dir/"atacseq_data/supporting_files/hg38-blacklist.v2.bed"

#aligned_data
bowtie2_dir=$practical_dir/"atacseq_data/bowtie2"

# create directories for each packge in the data directory before running script
samtools_dir=$projectdir/"samtools"
picard_dir=$projectdir/"picard"
macs_dir=$projectdir/"macs"
deeptools_dir=$projectdir/"deeptools"
subread_dir=$projectdir/"subread"
homer_dir=$projectdir/"homer"
```

Alignment: Bowtie2

- Alignment to GRCH38 with bowtie2

```
bowtie2 -p 12 \
  -x /sc/arion/projects/BiNGS/bings_omics/engine/annotation/homo_sapiens/grch38_gencode_36/bowtie2/2.2.8/index/index \
  -1 ${trimmgalore_dir}/${samplename}_ATAC_chr5_R1_val_1.fq.gz \
  -2 ${trimmgalore_dir}/${samplename}_ATAC_chr5_R2_val_2.fq.gz \
  -S ${bowtie2_dir}/${samplename}_noA.sam \
  -X 2000 \
  2> ${bowtie2_dir}/${samplename}.log
```

Alignment: Bowtie2

BiNGS

ARID2 WT

```
5407385 reads; of these:
  5407385 (100.00%) were paired; of these:
    79578 (1.47%) aligned concordantly 0 times
    3250212 (60.11%) aligned concordantly exactly 1 time
    2077595 (38.42%) aligned concordantly >1 times
    ----
    79578 pairs aligned concordantly 0 times; of these:
      12484 (15.69%) aligned discordantly 1 time
    ----
    67094 pairs aligned 0 times concordantly or discordantly; of these:
      134188 mates make up the pairs; of these:
        82711 (61.64%) aligned 0 times
        28632 (21.34%) aligned exactly 1 time
        22845 (17.02%) aligned >1 times
99.24% overall alignment rate
```

ARID2 KO

```
5722715 reads; of these:
  5722715 (100.00%) were paired; of these:
    85307 (1.49%) aligned concordantly 0 times
    3211482 (56.12%) aligned concordantly exactly 1 time
    2425926 (42.39%) aligned concordantly >1 times
    ----
    85307 pairs aligned concordantly 0 times; of these:
      14640 (17.16%) aligned discordantly 1 time
    ----
    70667 pairs aligned 0 times concordantly or discordantly; of these:
      141334 mates make up the pairs; of these:
        88304 (62.48%) aligned 0 times
        29711 (21.02%) aligned exactly 1 time
        23319 (16.50%) aligned >1 times
99.23% overall alignment rate
```

Filtering/Sorting/Indexing and removing Dups

- Filtering/Sorting/Indexing

```
mkdir -p $samtools_dir
start_time=`date +%s`

#remove chrM and keep only "chr" chromosomes
singularity exec $container_dir/"samtools_v1.15.sif" \
samtools view -h ${bowtie2_dir}/${samplename}_noA.sam | awk  '($3 != "chrM" && $3 != "chrUn")' >  ${bowtie2_dir}/${samplename}_noA_noM_chr_q20.sam

#convert to bam
singularity exec $container_dir/"samtools_v1.15.sif" \
samtools view -h -q 20 ${bowtie2_dir}/${samplename}_noA_noM_chr_q20.sam -o ${samtools_dir}/${samplename}_noA_noM_chr_q20.bam

#sort
singularity exec $container_dir/"samtools_v1.15.sif" \
samtools sort  -@ 8 ${samtools_dir}/${samplename}_noA_noM_chr_q20.bam  -o ${samtools_dir}/${samplename}_noA_noM_chr_q20_sorted.bam

#index
singularity exec $container_dir/"samtools_v1.15.sif" \
samtools index -@ 8  -b ${samtools_dir}/${samplename}_noA_noM_chr_q20_sorted.bam

end_time=`date +%s`
echo execution time was `expr $end_time - $start_time` s.
```

- Removing duplicates

```
82  mkdir -p $picard_dir
83  start_time=`date +%s`
84
85  singularity exec ${container_dir}/picard_v2.9.2.sif  \
86   java -jar /usr/local/share/picard-2.9.2-1/picard.jar MarkDuplicates \
87             I=${samtools_dir}/${samplename}_noA_noM_chr_q20_sorted.bam  \
88             O=${picard_dir}/${samplename}_noA_noM_chr_q20_sorted_nd.bam  \
89             REMOVE_DUPLICATES=true \
90             VALIDATION_STRINGENCY=LENIENT \
91             M=${picard_dir}/${samplename}_noA_noM_chr_q20_sorted_nd.txt
92
93  end_time=`date +%s`
94  echo execution time was `expr $end_time - $start_time` s.
95
```

Filtering/Sorting/Indexing and removing Dups

- Re-sort and re-index

```
start_time=`date +%s`

singularity exec $container_dir/"samtools_v1.15.sif" \
samtools sort  -@ 8 ${picard_dir}/${samplename}_noA_noM_chr_q20_sorted_nd.bam -o ${samtools_dir}/${samplename}_final.bam

singularity exec $container_dir/"samtools_v1.15.sif" \
samtools index -@ 8  -b ${samtools_dir}/${samplename}_final.bam

end_time=`date +%s`
echo execution time was `expr $end_time - $start_time` s.
```

- Generate bigwigs

```
mkdir -p $deeptools_dir
start_time=`date +%s`

singularity exec $container_dir/"deeptools_v3.5.1.sif" \
bamCoverage  --bam ${samtools_dir}/${samplename}_final.bam \
             --outFileName $deeptools_dir/${samplename}_final.bw \
             --outFileFormat bigwig  \
             --binSize=10 \
             --normalizeUsing RPKM \
             --extendReads=200  \
             --numberOfProcessors 8

end_time=`date +%s`
echo execution time was `expr $end_time - $start_time` s.
```

Merging Bams and calling peaks

- Merging Bam files

```
################################
#### Merging bam files##########
################################

echo "Merging bam files"

singularity exec ${container_dir}/samtools_v1.15.sif  \
 samtools merge ${samtools_dir}/master_atac.bam ${samtools_dir}/*_final.bam

singularity exec ${container_dir}/samtools_v1.15.sif  \
 samtools index -@ 12  -b ${samtools_dir}/master_atac.bam
```

- Calling peaks

```
singularity exec ${container_dir}/macs_v2.2.9.1.sif  \
 macs2 callpeak --nomodel \
      -t ${samtools_dir}/master_atac.bam  \
      --outdir ${macs_dir} \
      -n master_atac \
      -f BAMPE \
      -g hs \
      --keep-dup all \
      --slocal 1000 \
      2> ${macs_dir}/master_atac_macs2.log


end time=`date +%s`
```

- Master bam peaks = 8026 (after blacklisting

Peak filtering and coverage tracks

- Peak filtering/formatting

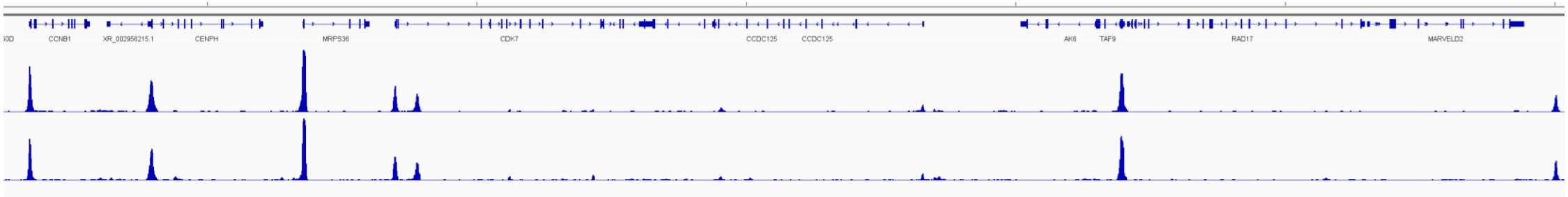
```
######################
# remove bl regions and generating bed file for gb
######################

echo "Peak filtering/formating"

singularity exec ${container_dir}/bedtools_v2.31.1.sif  \
 bedtools intersect -a ${macs_dir}/master_atac_peaks.narrowPeak -b $bl_regions -v > ${macs_dir}/master_atac_peaks_bl.narrowPeak

awk '{print $1,$2,$3,$4}' ${macs_dir}/master_atac_peaks_bl.narrowPeak > ${macs_dir}/temp.bed

echo "track name=\"SKmel147_ATAC_master_regions\" description=\"SKmel147_ATAC_master_regions\"" | cat - ${macs_dir}/temp.bed  > ${macs_dir}/master_atac_peaks_bl.bed
rm ${macs_dir}/temp.bed
```

- Generation of coverage tracks

```
################################################
#### run deeptools bamcoverage to generate bigwig
################################################
 echo "generating bigwig: master bam"

start_time=`date +%s`
singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
 bamCoverage  --bam ${samtools_dir}/master_atac.bam \
              --outFileName $deeptools_dir/master_atac.bw \
              --outFileFormat bigwig  \
              --binSize=10 \
              --normalizeUsing RPKM \
              --extendReads=200  \
              --numberOfProcessors 12


end_time=`date +%s`
echo execution time was `expr $end_time - $start_time` s.
```

Coverage Track Visualization: IGV

- Adding bigwigs to IGV



Quantification of RIPs

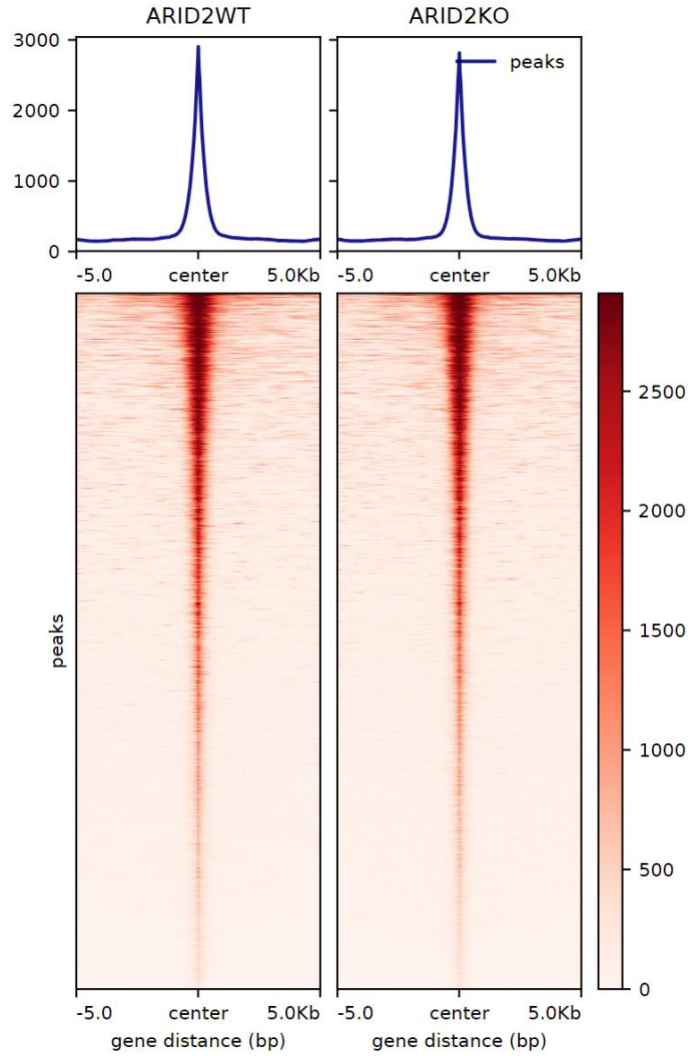
- Quantification of reads in peaks

```
# ##################################################
# #### run featurecounts to generate coun matrix
# ##################################################

echo "generating count matrix"
start_time=`date +%s`


awk 'OFS="\t" {print $4, $1, $2, $3, "."}' ${macs_dir}/master_atac_peaks_bl.narrowPeak > ${macs_dir}/master_atac_peaks_bl.saf

mkdir -p $subread_dir

singularity exec ${container_dir}/subread_v2.0.1.sif  \
 featureCounts -p -a ${macs_dir}/master_atac_peaks_bl.saf -F SAF -o ${subread_dir}/master_atac_peaks_bl_subread.txt ${samtools_dir}/*final.bam -T 12
```

	KO	WT
Assigned	977981	1161417
Unassigned_Unmapped	0	0
Unassigned_Read_Type	0	0
Unassigned_Singleton	0	0
Unassigned_MappingQuality	0	0
Unassigned_Chimera	0	0
Unassigned_FragmentLength	0	0
Unassigned_Duplicate	0	0
Unassigned_MultiMapping	0	0
Unassigned_Secondary	0	0
Unassigned_NonSplit	0	0
Unassigned_NoFeatures	951625	1050525
Unassigned_Overlapping_Length	0	0
Unassigned_Ambiguity	307	471

Heatmap Visualization: Deeptools

```
################################################
#### run deeptools to generate heatmap (split by sample)
################################################
echo "generating deeptool heatmap - individual sample"

start_time=`date +%s`


awk 'FNR==NR {a[$4]; next} FNR> 1 && $4 in a' ${macs_dir}/master_atac_peaks_bl.bed ${macs_dir}/master_atac_summits.bed > ${macs_dir}/master_atac_summits_bl.bed

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
 computeMatrix reference-point \
    -R ${macs_dir}/master_atac_summits_bl.bed \
    --skipZeros \
    -S ${deeptools_dir}/SKMel147_ARID2WT_final.bw ${deeptools_dir}/SKMel147_ARID2KO_final.bw \
    -o ${deeptools_dir}/master_atac_summits_bl_2_samples.gz \
    -b 5000 -a 5000 \
    --referencePoint center \
    -p 4 \
    --samplesLabel ARID2WT ARID2KO

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
 plotHeatmap -m ${deeptools_dir}/master_atac_summits_bl_2_samples.gz \
    --plotFileFormat pdf \
    -out ${deeptools_dir}/master_atac_summits_bl_2_samples.pdf \
    --outFileSortedRegions ${deeptools_dir}/master_atac_summits_bl_2_samples_sorted.bed \
    --dpi 720 \
    --missingDataColor White \
    --colorMap Reds \
    --regionsLabel peaks \
    --heatmapHeight 13
```



Heatmap Visualization: Deeptools

```
################################################
#### run deeptools to generate heatmap (merged)
################################################


echo "generating deeptool heatmap - merged sample"

start_time=`date +%s`


awk 'FNR==NR {a[$4]; next} FNR> 1 && $4 in a' ${macs_dir}/master_atac_peaks_bl.bed ${macs_dir}/master_atac_summits.bed > ${macs_dir}/master_atac_summits_bl.bed

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
 computeMatrix reference-point \
    -R ${macs_dir}/master_atac_summits_bl.bed \
    --skipZeros \
    -S ${deeptools_dir}/master_atac.bw  \
    -o ${deeptools_dir}/master_atac_summits_bl_merged.gz \
    -b 5000 -a 5000 \
    --referencePoint center \
    -p 4 \
    --samplesLabel master_atac

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
 plotHeatmap -m ${deeptools_dir}/master_atac_summits_bl_merged.gz \
    --plotFileFormat pdf \
    -out ${deeptools_dir}/master_atac_summits_bl_merged.pdf \
    --outFileSortedRegions ${deeptools_dir}/master_atac_summits_bl_merged_sorted.bed \
    --dpi 720 \
    --missingDataColor White \
    --colorMap Reds \
    --regionsLabel peaks \
    --heatmapHeight 13
```

Normalization and differential peak analysis

- With AWK

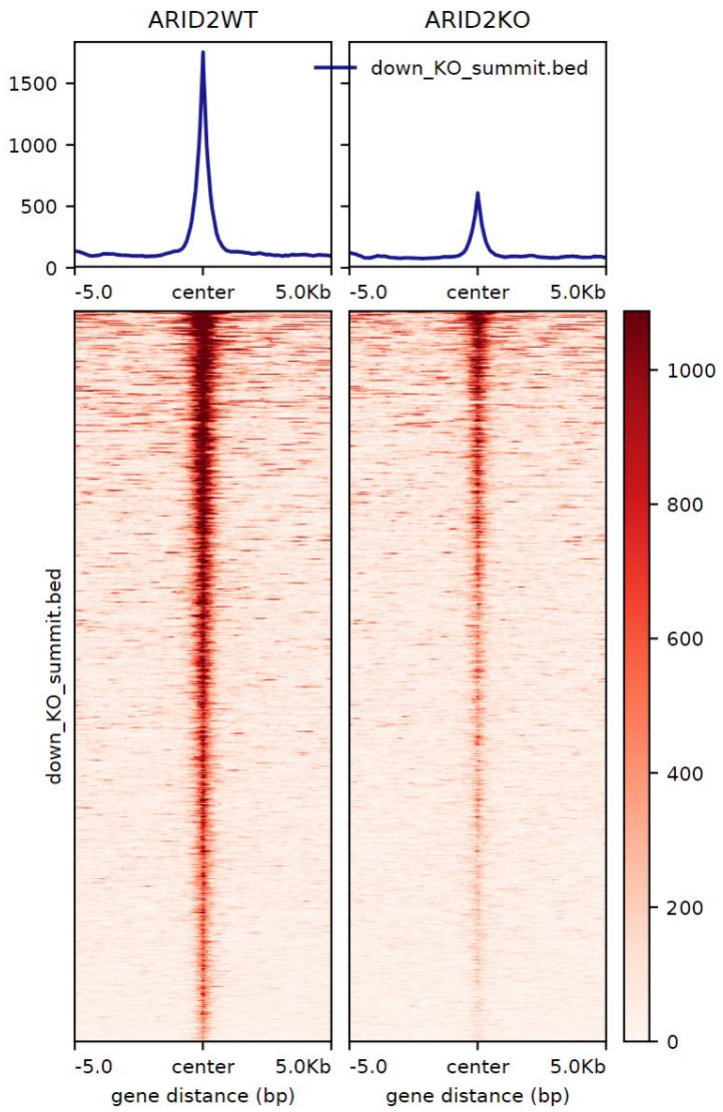
```
#################################################################
#### Normalize and compute log2(KO/WT) add differential status
#################################################################

tail -n +3 ${subread_dir}/master_atac_peaks_bl_subread.txt > ${subread_dir}/master_atac_peaks_bl_subread_no_header.txt

#calculate total reads per sample
total_reads_wt=$(awk '{ sum += $8 } END { print sum }' ${subread_dir}/master_atac_peaks_bl_subread_no_header.txt)
total_reads_ko=$(awk '{ sum += $7 } END { print sum }' ${subread_dir}/master_atac_peaks_bl_subread_no_header.txt)

#replace 0s
awk '{ if ($7 == 0) $7 = 1; print }' ${subread_dir}/master_atac_peaks_bl_subread_no_header.txt > ${subread_dir}/master_atac_peaks_bl_subread_no_header_nozero.txt

#calculate rpkm for wt
awk -v total_reads_wt="$total_reads_wt" '{ printf "%s\t%s\t%.2f\t%.2f\n", $1, $6,  ($8 / ($6 / 1000)) / (total_reads_wt / 1000000), $7 }' ${subread_dir}/master_atac_peaks_bl_subread_no_header_nozero.txt > ${subread_dir}/master_atac_peaks_bl_subread_norm.txt
awk -v total_reads_ko="$total_reads_ko" '{ printf "%s\t%s\t%.2f\t%.2f\n", $1, $2, $3, ($4 / ($2 / 1000)) / (total_reads_ko / 1000000) }' ${subread_dir}/master_atac_peaks_bl_subread_norm.txt > ${subread_dir}/master_atac_peaks_bl_subread_norm_final.txt

#calculate log2fc
awk '{ ($5 = log($4/$3)/log(2)) print }' ${subread_dir}/master_atac_peaks_bl_subread_norm_final.txt > ${subread_dir}/master_atac_peaks_bl_subread_no_header_log2diff.txt


# save up in KO peaks
awk '$5 >= 1' ${subread_dir}/master_atac_peaks_bl_subread_no_header_log2diff.txt > ${subread_dir}/master_atac_peaks_bl_up_KO.txt
awk 'FNR==NR {a[$1]; next} FNR> 1 && $4 in a' ${subread_dir}/master_atac_peaks_bl_up_KO.txt ${macs_dir}/master_atac_summits.bed > ${subread_dir}/master_atac_peaks_bl_up_KO_summit.bed

# save down in KO peaks
awk '$5 <= -1' ${subread_dir}/master_atac_peaks_bl_subread_no_header_log2diff.txt > ${subread_dir}/master_atac_peaks_bl_down_KO.txt
awk 'FNR==NR {a[$1]; next} FNR> 1 && $4 in a' ${subread_dir}/master_atac_peaks_bl_down_KO.txt ${macs_dir}/master_atac_summits.bed > ${subread_dir}/master_atac_peaks_bl_down_KO_summit.bed

# save static peaks
awk '$5 < 1 && $5 > -1 ' ${subread_dir}/master_atac_peaks_bl_subread_no_header_log2diff.txt > ${subread_dir}/master_atac_peaks_bl_static.txt
awk 'FNR==NR {a[$1]; next} FNR> 1 && $4 in a' ${subread_dir}/master_atac_peaks_bl_static.txt ${macs_dir}/master_atac_summits.bed > ${subread_dir}/master_atac_peaks_bl_static_summit.bed
```

Visualization: Differential peak analysis

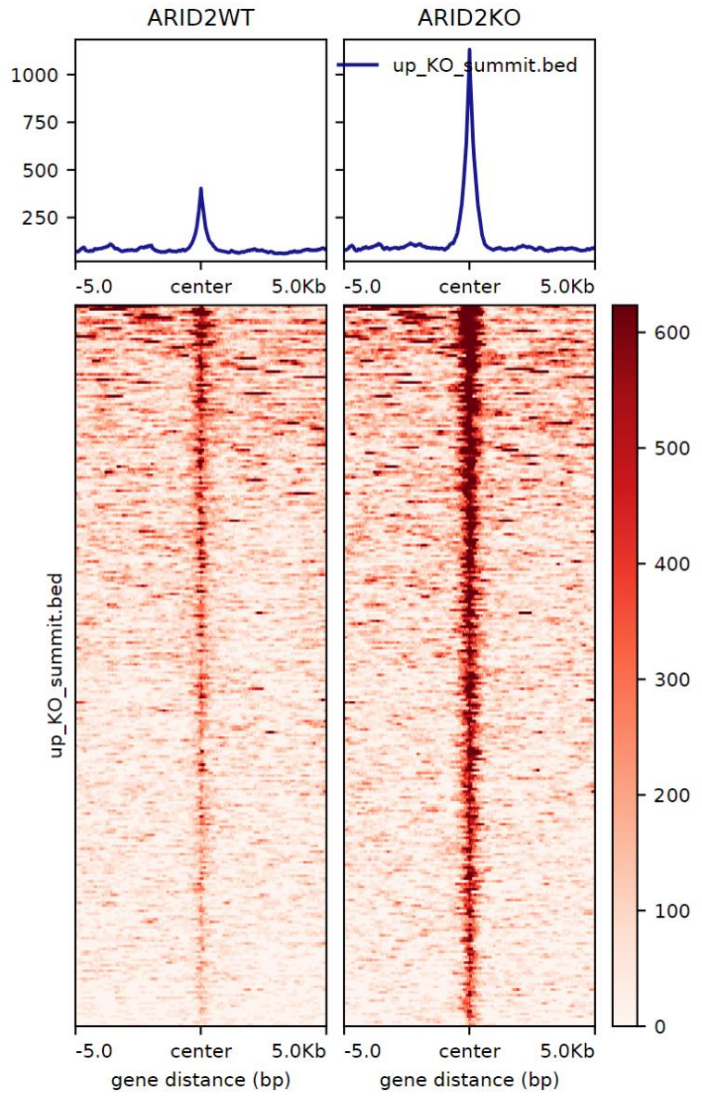
```
#####################################################
### plot heatmap on differential and static regions
#####################################################



echo "generating deeptool heatmap - differential and static peaks"

start_time=`date +%s`

for file in ${subread_dir}/*_summit.bed;

  do
    # file="$(basename $file)"
    # samplename=$(echo "$file" | sed 's/.*\(_bl_\)/\1/g')
    samplename=$(echo "$file" | sed 's/.*bl.//g')

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
    computeMatrix reference-point \
        -R ${file} \
        --skipZeros \
        -S ${deeptools_dir}/SKMel147_ARID2WT_final.bw ${deeptools_dir}/SKMel147_ARID2KO_final.bw \
        -o ${deeptools_dir}/${samplename}.gz \
        -b 5000 -a 5000 \
        --referencePoint center \
        -p 4 \
        --samplesLabel ARID2WT ARID2KO

singularity exec ${container_dir}/deeptools_v3.5.1.sif  \
    plotHeatmap -m ${deeptools_dir}/${samplename}.gz \
        --plotFileFormat pdf \
        -out ${deeptools_dir}/${samplename}.pdf \
        --dpi 720 \
        --missingDataColor White \
        --colorMap Reds \
        --regionsLabel ${samplename} \
        --heatmapHeight 13
done



end_time=`date +%s`
echo execution time was `expr $end_time - $start_time` s.
```

Visualization: Differential peak analysis

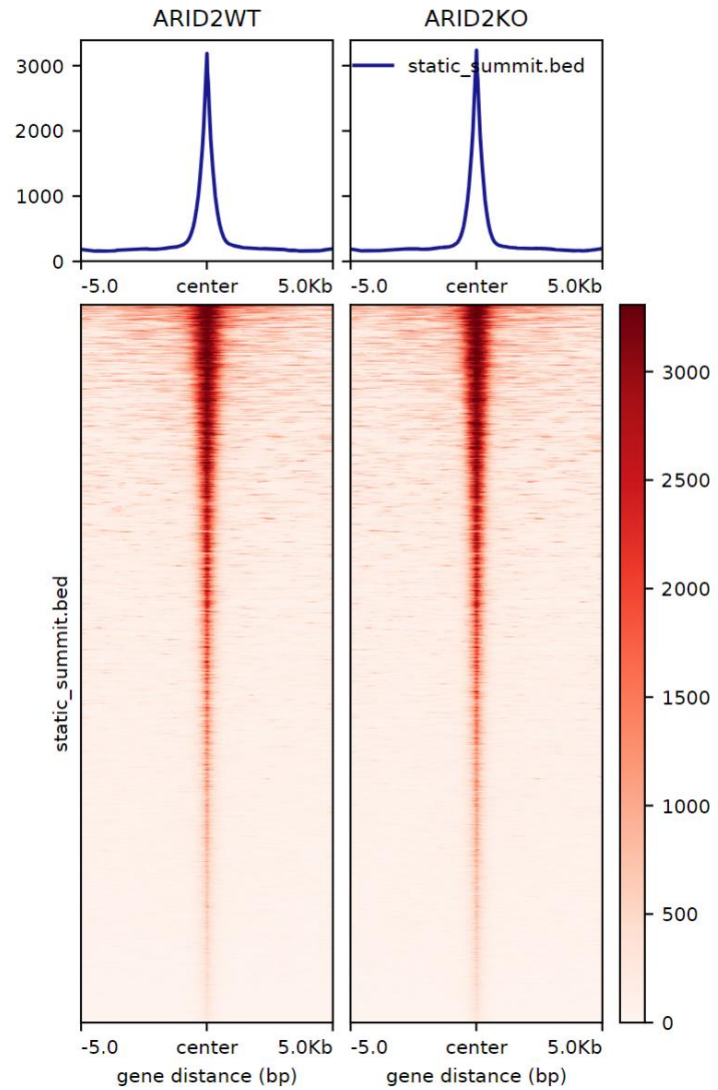
N = 762



N = 496



N = 6765



Motif analysis: Differential peak analysis

```
#####################################################################
#### homer motif analysis - split by differential and static peaks
#####################################################################

echo "motif enrichment analysis using homer"


preparsedDir=${practical_dir}/"atacseq_data/supporting_files/preparsed_dir"


start_time=`date +%s`


for file in ${subread_dir}/*_summit.bed;

  do

    samplename=$(echo "$file" | sed 's/.*bl.//g' | sed 's/_summit.bed//')
    output_dir=${homer_dir}/$samplename
    mkdir -p $output_dir

    bed_path=$file

    singularity exec ${container_dir}/homer_v4.11_hg38.sif    \
    findMotifsGenome.pl   $bed_path   hg38 $output_dir -preparsedDir $preparsedDir   -size 200   -p 6

done
```

Motif analysis: Differential peak analysis

- Homer de-novo UP in KO

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-40	-9.412e+01	29.32%	4.39%	47.9bp (65.9bp)	BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer(0.975) More Information | Similar Motifs Found	motif file (matrix)
2		1e-26	-6.212e+01	14.66%	1.37%	42.5bp (58.7bp)	BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer(0.881) More Information | Similar Motifs Found	motif file (matrix)
3		1e-12	-2.870e+01	16.54%	4.65%	56.5bp (62.8bp)	PB0086.1_Tcfap2b_1/Jaspar(0.722) More Information | Similar Motifs Found	motif file (matrix)
4		1e-12	-2.834e+01	2.26%	0.01%	52.9bp (28.7bp)	PB0051.1_Osr2_1/Jaspar(0.646) More Information | Similar Motifs Found	motif file (matrix)

- Homer de-novo Down in KO

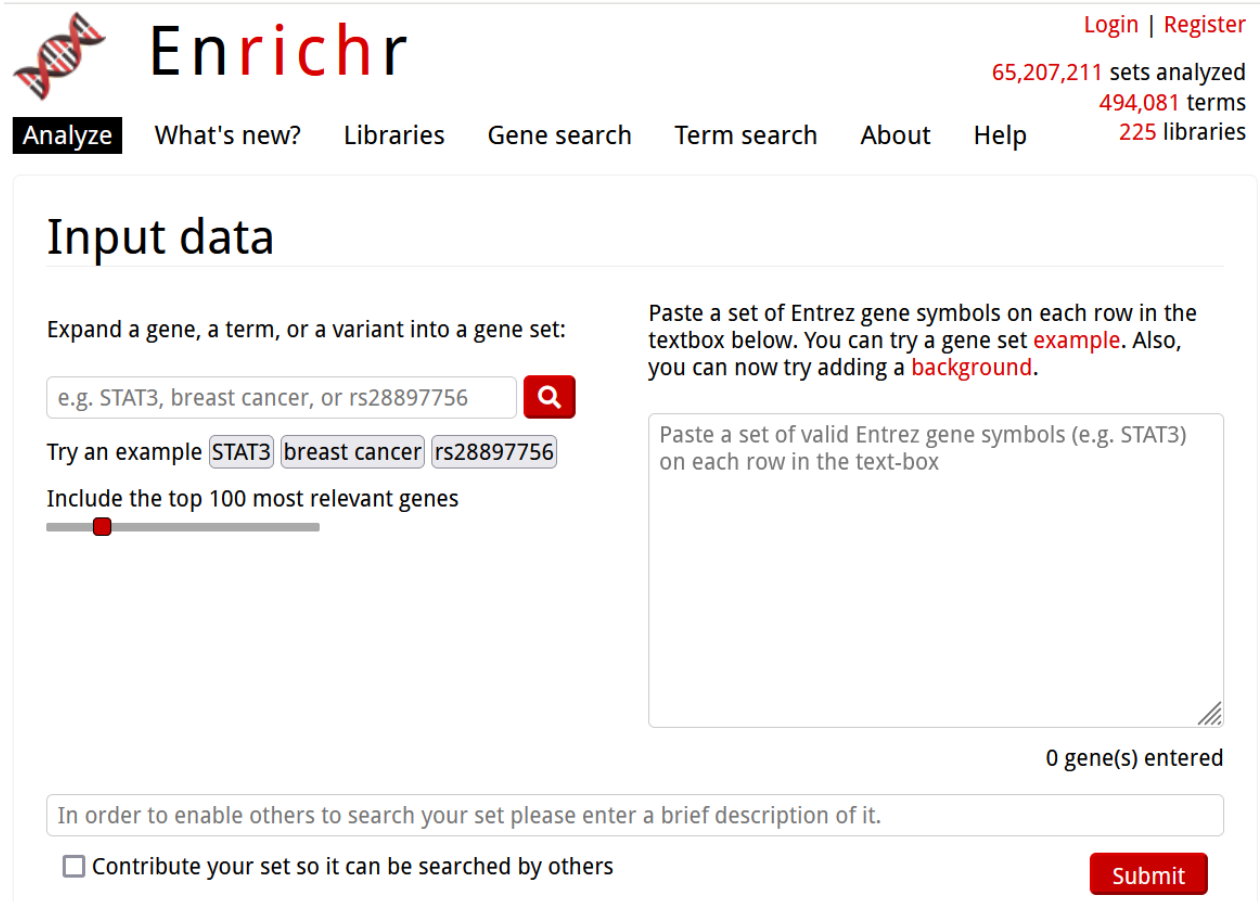
Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-211	-4.880e+02	36.88%	5.41%	43.7bp (65.8bp)	Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer(0.993) More Information | Similar Motifs Found	motif file (matrix
2		1e-29	-6.809e+01	7.19%	1.41%	51.8bp (59.4bp)	RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer(0.943) More Information | Similar Motifs Found	motif file (matrix
3		1e-25	-5.804e+01	38.10%	23.70%	52.2bp (62.8bp)	EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer(0.945) More Information | Similar Motifs Found	motif file (matrix
4		1e-18	-4.249e+01	3.45%	0.51%	51.0bp (61.9bp)	IRF1(IRF)/PBMC-IRF1-ChIP-Seq(GSE43036)/Homer(0.950) More Information | Similar Motifs Found	motif file (matrix

Peak annotation

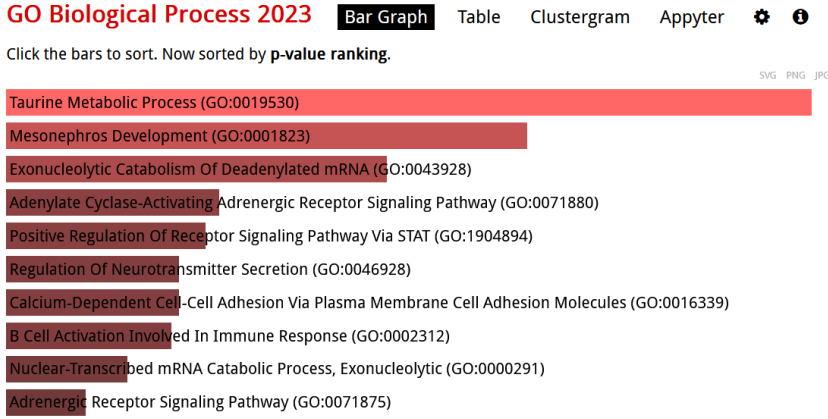
```
#############################################################
#### associate peaks with nearest gene promoter
#############################################################

##Peak annotation
echo "annotate peaks with nearest gene and genomic localization"

for file in ${subread_dir}/*_summit.bed;

  do

    samplename=$(echo "$file" | sed 's/.*bl.//g' | sed 's/_summit.bed//')

    singularity exec ${container_dir}/homer_v4.11_hg38.sif  \
    annotatePeaks.pl  ${file} hg38 -size 200 > ${subread_dir}/${samplename}_annotated.txt

done
```

Pathway analysis

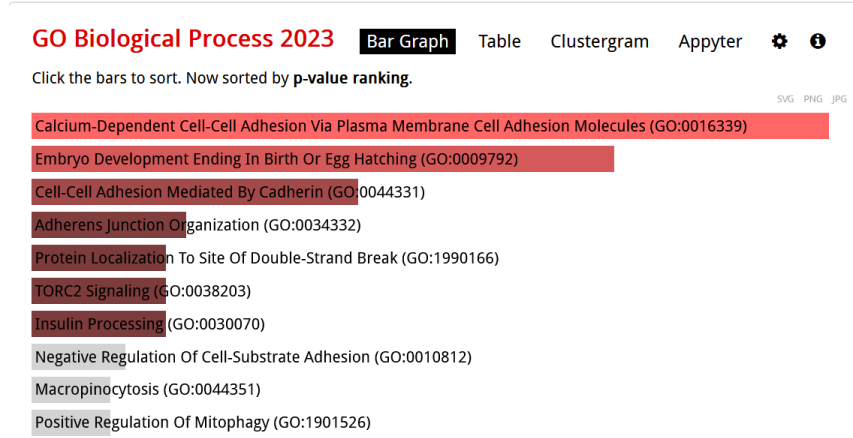
https://maayanlab.cloud/Enrichr/



Down in KO



Up in KO



Heatmap with ChIPs - Kmeans