Hackseq Project #5

Evaluating epigenetic modifications in ChIP-seq and methylation data across cell types and states

Tools of interest

- bedtools:
    - Intersection and set difference between genomic ranges (.bed files)
    - Extraction of FASTA sequences from bed files (requires a genome)
- FIMO: A motif finding tool for known motifs
    - Scores the occurence of motifs in sequences based on profile databases (e.g. JASPAR)
- TOMTOM: Motif comparison tool
    - Aligns query motifs to a profile database to compute the best possible match
- MotifGP: A discriminative de novo motif discovery algorithm
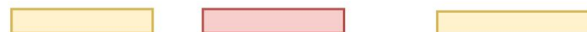- ChIPSeeker: ChIP-seq peak annotation software

Methylation 1

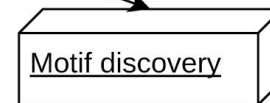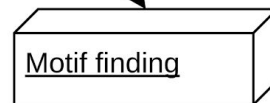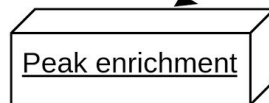**Methylation 2**

Differential Methylation

DMR

Intersect

Regions of interest (ENCODE ChIP-seq, known motif locations, generic genomic ranges.)

Peak enrichment

Motif finding

Motif discovery

-Filter by q-value
-Filter for expressed genes (RNA-seq data)
-Call enrichment

-Compute matched regions with FIMO and JASPAR
-Compare and count matches by frequency and significance

-Discover de novo motifs from DMR (hypermethylated v.s. hypomethylated)
-Validate predictions using TOMTOM and JASPAR

Motif finding in DMR

Top matches for known motifs in the down-methylated regions							
pattern	sequence name	start	stop	score	p-value	q-value	matched sequence
MA0528.1	chr7:108066562-108067006	206	226	29.4898	4.92E-13	6.52E-08	GGAGGAGGAGGAGGAGGAGGG
MA0528.1	chr4:56875392-56877360	233	253	29.449	5.95E-13	6.52E-08	ggaggaggaggaggaggagga
MA0528.1	chr4:56875392-56877360	236	256	29.449	5.95E-13	6.52E-08	ggaggaggaggaggaggagga
MA0528.1	chr4:56875392-56877360	239	259	29.449	5.95E-13	6.52E-08	ggaggaggaggaggaggagga
MA0528.1	chr4:56875392-56877360	242	262	29.449	5.95E-13	6.52E-08	ggaggaggaggaggaggagga
MA0528.1	chr4:56875392-56877360	245	265	29.449	5.95E-13	6.52E-08	ggaggaggaggaggaggagga

Top motif found by occurence (down-methylated regions)
5861 MA0528.1
2816 MA0073.1
2626 MA0149.1
2062 MA0472.1
1620 MA0516.1
1596 MA0050.2
1566 MA0079.3
1505 MA0474.1
1378 MA0098.2
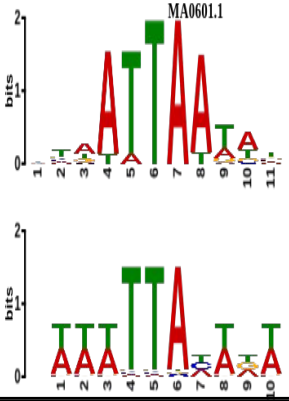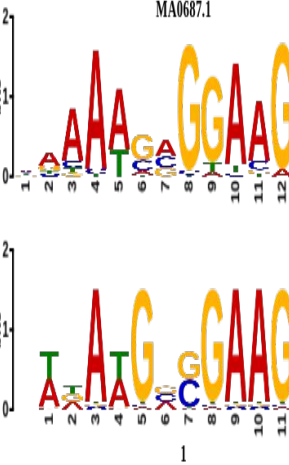
| Top motif found by occurence (up-methylated regions) |
|---|
| **5418 MA0528.1** |
| 3599 MA0554.1 |
| 2167 MA0073.1 |
| **2114 MA0149.1** |
| 1932 MA0543.1 |
| 1441 MA0079.3 |
| 1431 MA0516.1 |
| 1324 MA0753.1 |
| 1278 MA0537.1 |

De novo motif discovery in DMR

Top scoring motif in

1) Full DMR regions (Arid3b)

2) In hypermethylated regions discriminating against hypomethylated regions.

Arid3b - MA0601.1		Alignment
Database	JASPAR CORE 2016	
p-value	9.20e-05	
E-value	9.95e-02	
q-value	5.88e-02	
Overlap	10	
Offset	1	
Orientation	Reverse Complemen	

SPIC - MA0687.1		Alignment
Database	JASPAR CORE 2016	
p-value	8.59e-07	
E-value	9.30e-04	
q-value	1.86e-03	
Overlap	13	
Offset	1	
Orientation	Normal	

Peak enrichment analysis

See RMarkdown (meth_to_tf.rmd) document in the repository!

Gene annotations

Genes Predicted Downregulated in Aged Mice
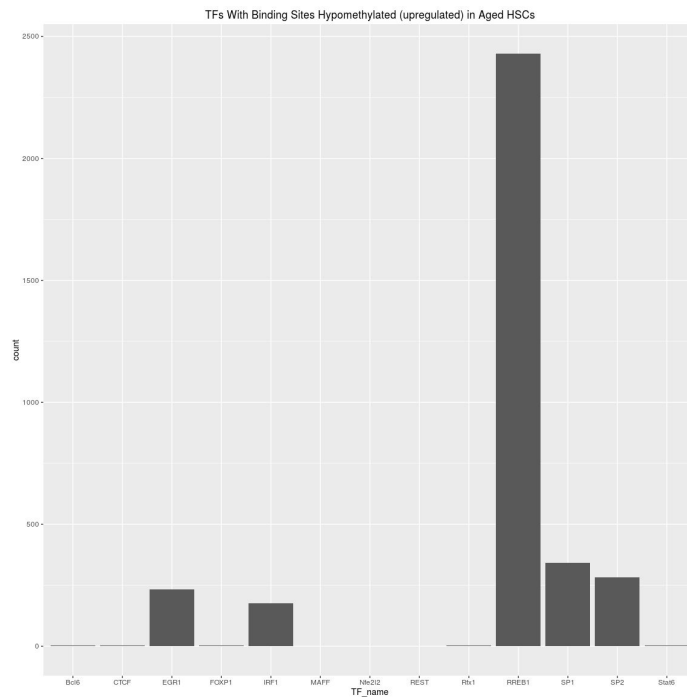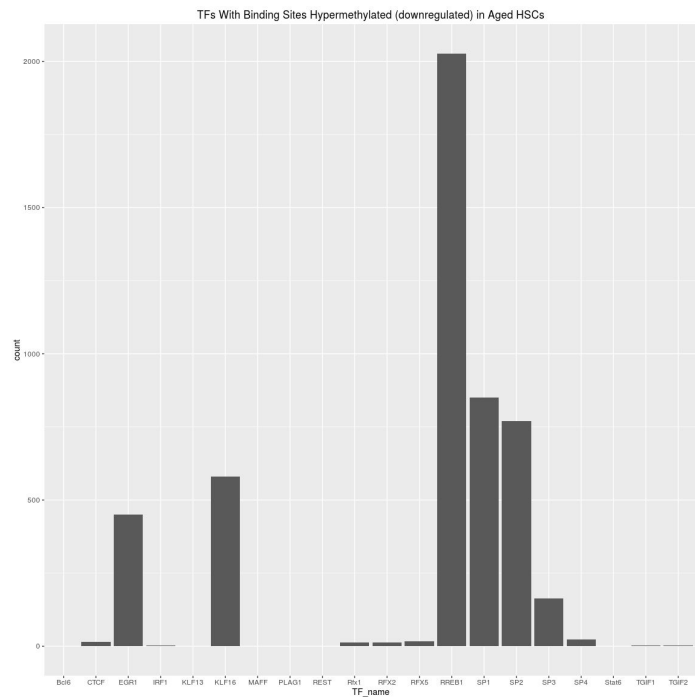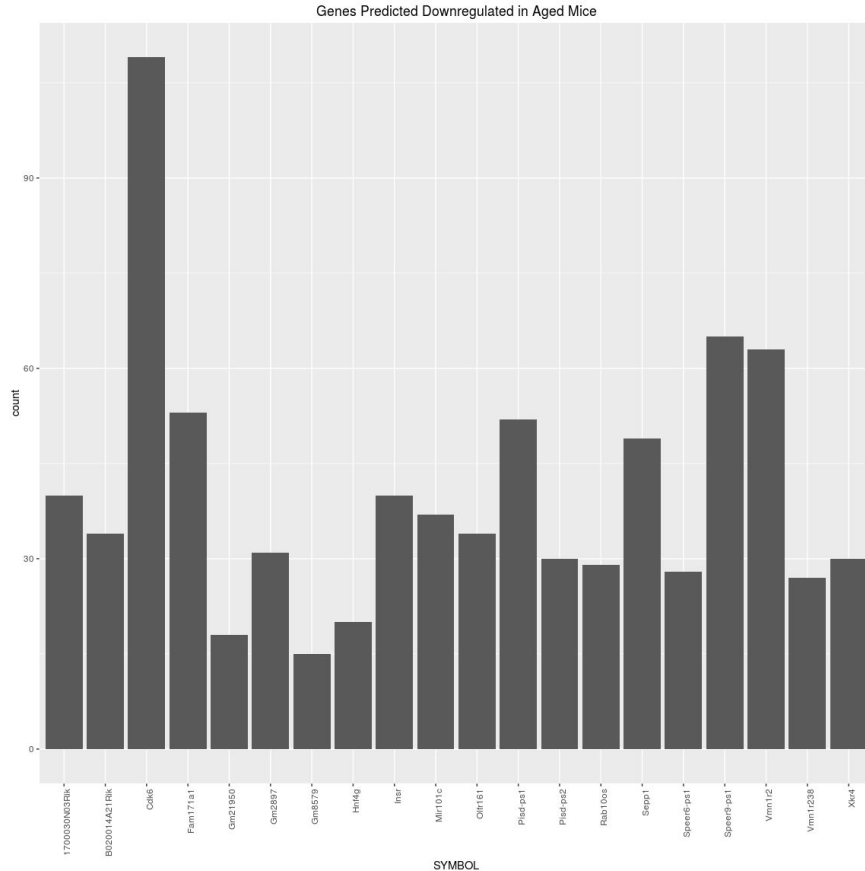
CDK6 known to be involved in HSC quiescence

Checked in RNA-seq: 1.45-fold higher expression in young mice

Mir-101c also implicated in leukemias

Gene annotations

Genes Predicted Downregulated in Young Mice