

# Assignment7

Sonali Joshi

4/28/2017

From AnnotationHub get the H3K4me3 data for adult brain, fetal brain and adult liver.

```
ah = AnnotationHub()

## Warning: database may not be current
## database: '/Users/sonalijosbi//.AnnotationHub/annotationhub.sqlite3'
## reason: Timeout was reached: Resolving timed out after 10000 milliseconds
## snapshotDate(): 2017-10-27

ah <- subset(ah, species == "Homo sapiens")

ah_fetal <- query(ah, c("H3K4me3", "Brain", "narrowPeak", "Fetal"))
ah_fetal <- ah[["AH30479"]]

## loading from cache '/Users/sonalijosbi//.AnnotationHub/35919'
ah_adult <- query(ah, c("Homo sapiens", "H3K4me3", "Brain", "narrowPeak"))
ah_adult <- ah[["AH30413"]]

## loading from cache '/Users/sonalijosbi//.AnnotationHub/35853'
ah_liver <- query(ah, c("Homo sapiens", "H3K4me3", "Liver", "narrowPeak"))
ah_liver <- ah[["AH30367"]]

## loading from cache '/Users/sonalijosbi//.AnnotationHub/35807'
```

Get the up and down regulated genes from the previous exercise

Use mapIds to map the gene Symbols to UCSCCKG

```
#Get the annotation for the up and down regulated genes
#Convert Gene symbols to tx_id for list1 and list2

up_gene_list <- (as.matrix(up_genes$gene))
list1 <- up_gene_list[,1]
down_gene_list <- (as.matrix(down_genes$gene))
list2 <- down_gene_list[,1]
#head(list2)

orgs <- subset(ah, ah$rdataclass == "OrgDb")
human <- orgs[[1]]
upgenes_list <- (mapIds(human, keys=list1, column = "UCSCCKG", keytype = "SYMBOL"))
downgenes_list <- (mapIds(human, keys=list2, column = "UCSCCKG", keytype = "SYMBOL"))
```

Get the gene names for the genes associated with H3K4me3, from Txdb

```
transcriptsByGene = transcriptsBy(TxDb.Hsapiens.UCSC.hg19.knownGene, by="gene")
ul_transcriptsByGene <- unlist(transcriptsByGene)
overlaps_adult <- findOverlaps(ah_adult, ul_transcriptsByGene)
overlaps_fetal <- findOverlaps(ah_fetal, ul_transcriptsByGene)
overlaps_liver <- findOverlaps(ah_liver, ul_transcriptsByGene)
gene_names_adult <- ul_transcriptsByGene[subjectHits(overlaps_adult)]
gene_names_fetal <- ul_transcriptsByGene[subjectHits(overlaps_fetal)]
gene_names_liver <- ul_transcriptsByGene[subjectHits(overlaps_liver)]

list_ah2 = list(adult=gene_names_adult$tx_name ,
               fetal=gene_names_fetal$tx_name,
               liver=gene_names_liver$tx_name)

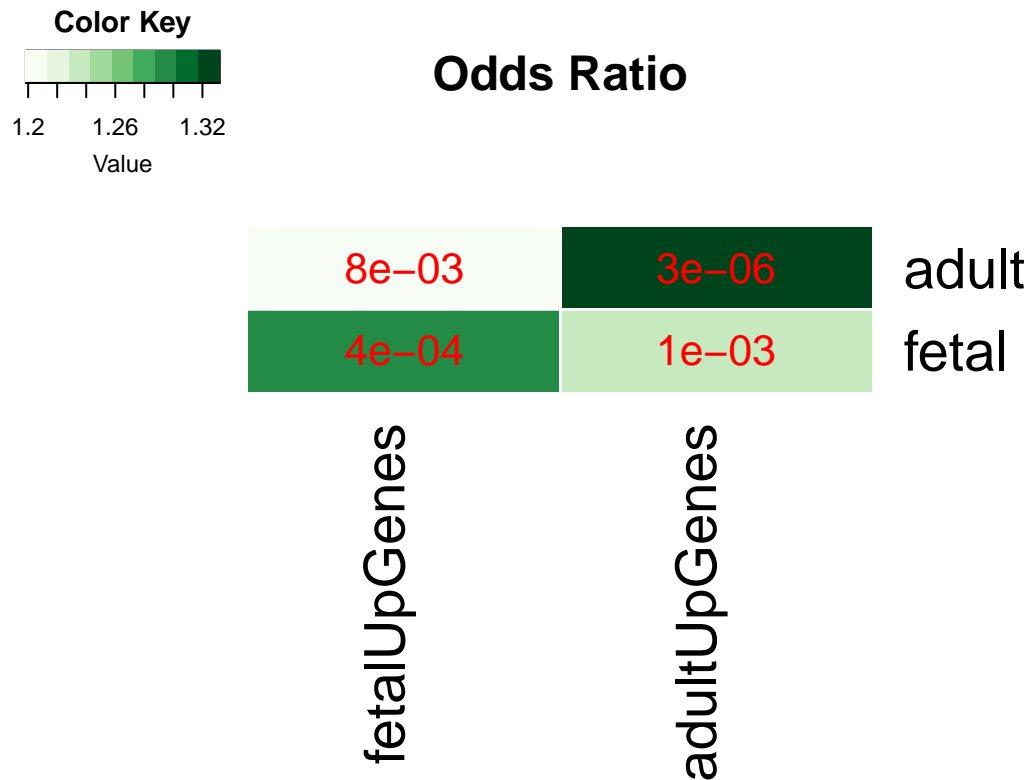
list_ah1 = list(adult=gene_names_adult$tx_name ,
               fetal=gene_names_fetal$tx_name)

list_gene_expr = list(fetalUpGenes = upgenes_list,
                     adultUpGenes = downgenes_list)
```

Test the first question “Are there changes in H3K4me3 between fetal and adult brain over promoters for genes differentially expressed between fetal and adult brain?”

We can see from the odds ratio/heatmap that fetal up regulated genes are indeed better associated with the H3Kme3 promoter regions and the same is true for the adult genes.

```
gom_obj1 <- newGOM(list_ah1, list_gene_expr, 10^6)
drawHeatmap(gom_obj1)
```

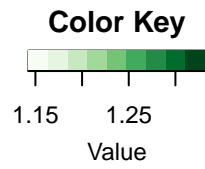


N.S.: Not Significant; —: Ignored

Test the second question “Are promoters of genes differentially expressed between adult and fetal brain marked by H3K4me3 in liver?”

We see from the heatmap that includes the liver data that the genes are marked in the liver to a lesser degree than in the brain.

```
gom_obj2 <- newGOM(list_ah2, list_gene_expr, 10^6)
drawHeatmap(gom_obj2)
```



## Odds Ratio

fetalUpGenes	8e-03	3e-06	adult
	4e-04	1e-03	fetal
	4e-02	1e-02	liver
	adultUpGenes		

N.S.: Not Significant; --: Ignored