

## Genomic data science Capstone Gene set enrichment - WEEK 9

### Analysis steps:

1. The results from DESEQ2 were further processed to stringent selection criteria of  $p_{adj} < 0.01$  and LogFoldChange threshold of 1 (corresponding to actual fold change = 2).
2. The list of ENTREZIDs of the Upregulated (LogFoldChange > 0) and downregulated genes (LogFoldChange < 0) were extracted from the DESEQ2 results object dataframe.
3. The narrow peak files were extracted for fetal brain(E081), adult brain(E073) and liver(E066) cell types from AnnotationHub.
4. The list of promoters present in the UP and DOWN regulated genes (obtained from DESEQ2 results) were extracted from the TxDb.Hsapiens.UCSC.hg19.knownGene database.
5. The overlaps between the promoters of the differentially expressed genes and the narrow peak files was checked.

The results are -

	Upregulated genes(fetal)	Downregulated genes(fetal)
Number of Genes	1221	1952
Matches of genes in Txdb promoters	3976	6274
Matches of promoters with H3K4ME3 Adult brain	1601 (40%)	2814 (44.8%)
Matches of promoters with H3K4ME3 Fetal brain	1773 (44%)	1271 (20.2)
Matches of promoters with H3K4ME3 liver	1658 (41.7%)	2235 (35.6%)

### Results:

H3K4me3 is an epigenetic mark that is involved in activation of transcription and is found associated with promoter sites. Regulation of gene expression through H3K4me3 plays a significant role in stem cell fate determination and early embryo development. Based on this information it was expected that H3K4me3 would show strong association with the upregulated promoters in Fetal brain, but will show low presence in adult samples. Similarly, a strong association was expected with downregulated promoters (upregulated in Adults) in adult samples and low in fetal conditions.

- In the fetal brain, overlap is seen in 20% DOWN promoters and 44% UP promoters, which is expected.
- In the adult brain, overlap is seen in 44.8% DOWN promoters which is also expected. But, overlap is seen in 40% UP promoters, which is unexpected, though the overall trend is as expected.
- In liver, 41% match is seen in UP genes and 35.6% match in DOWN genes.