

Analysis Report (Part6): Epigenetics Analysis

(Last updated: 10/05/21)

The following R script gets the percentage of overlap between the promoters of differentially expressed genes found in Part5 of the Analysis Report (Differential Expression Analysis) and the epigenetically marked (**H3K4me3**) promoters in the fetal and adult brain cells, and in liver cell lines which are retrieved from Annotationhub (**EpigenomeRoadMap**).

1. Loading the Necessary Libraries and Data found in Part5 of the Analysis Report (Differential Expression Analysis)

```
library(AnnotationHub)
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
library(EnsDb.Hsapiens.v86)

# List of the Differentially EWxpressed Genes found in Part5 of the
# Analysis Report (Differential Expression Analysis)
sorted_significant_differ_res_deseq2_shrunk =
  read.table("../Differentially_Expressed_Genes-results.tsv", quote = "", sep = '\t')
```

2. Get the Promoter Associated Histone Modifications (H3k4me3) Narrow Peaks Data from Annotationhub (EpigenomeRoadMap)

```
# initialize AnnotationHub object
ah = AnnotationHub()
ah = subset(ah, species == "Homo sapiens")
# get the narrow peaks data (promoter associated histone modification H3K4me3) for the fetal brain cell.
fetal.brain = AnnotationHub::query(ah, c("EpigenomeRoadMap", "H3K4me3", "E081"))
fetal.brain.gr = fetal.brain[[2]]
# get the narrow peaks data (promoter associated histone modification H3K4me3) for the adult brain cell.
adult.brain = AnnotationHub::query(ah, c("EpigenomeRoadMap", "H3K4me3", "E073"))
adult.brain.gr = adult.brain[[2]]
# get the narrow peak data (promoter associated histone modification H3K4me3) for liver cell line
liver.line = AnnotationHub::query(ah, c("EpigenomeRoadMap", "H3K4me3", "Liver"))
liver.line.gr = liver.line[[2]]
```

3. Change ENSEMBL ID to ENTREZID that Matches gene_id in TXDB

```
sorted_significant_differ_res_deseq2_shrunk$gene = sub("\\.\\d+$", "", as.character(sorted_significant_differ_res_deseq2_shrunk$gene))
AnnotationDbi::keytypes(EnsDb.Hsapiens.v86)
```

```
## [1] "ENTREZID"          "EXONID"            "GENEBIOTYPE"
## [4] "GENEID"            "GENENAME"          "PROTDOMID"
## [7] "PROTEINDOMAINID"   "PROTEINDOMAINSOURCE" "PROTEINID"
## [10] "SEQNAME"           "SEQSTRAND"         "SYMBOL"
## [13] "TXBIOTYPE"         "TXID"              "TXNAME"
## [16] "UNIPROTID"
```

```
sorted_significant_differ_res_deseq2_shrunk.map = AnnotationDbi::select(EnsDb.Hsapiens.v86,
                                keys = as.character(sorted_significant_differ_res_deseq2_shrunk$gene), keytype = "GENEID")
```

4. Get the Promoters of DGEs Found in Part5 of the Analysis Report (Differential Expression Analysis)

```
# get the known genes from Tx database
txdb = TxDb.Hsapiens.UCSC.hg19.knownGene
txdb.genes = genes(txdb)
# get the promoters of differentially expressed genes
diff.expressed.gene.promoters = promoters(txdb.genes[txdb.genes$gene_id %in% sorted_significant_differ_res_deseq2_shrunk$gene_id])
```

5. Calculate the Percentage of Overlap Between the Promoters of DEGs Found in Part5 of the Analysis Report (Differential Expression Analysis) and the Epigenetically Marked (H3K4me3) Promoters in the 3 Different Kind of Cells

The percentage of overlap between the promoters of differentially expressed genes found in Part5 of the Analysis Report (Differential Expression Analysis) and the epigenetically marked (**H3K4me3**) promoters in the **fetal brain cells** is calculated as following:

```
# subsetByOverlaps() extracts the elements in the query (the first argument) that overlap at least one
fetal.brain.overlap.H3K4me3 = subsetByOverlaps(diff.expressed.gene.promoters, fetal.brain.gr)
(fetal.brain.overlap.percentage.H3K4me3 = length(fetal.brain.overlap.H3K4me3) / length(sorted_significant_differ_res_deseq2_shrunk$gene_id))
```

```
## [1] 34.77612
```

The percentage of overlap between the promoters of differentially expressed genes found in Part5 of the Analysis Report (Differential Expression Analysis) and the epigenetically marked (**H3K4me3**) promoters in the **adult brain cells** is calculated as following:

```
# subsetByOverlaps() extracts the elements in the query (the first argument) that overlap at least one
adult.brain.overlap.H3K4me3 = subsetByOverlaps(diff.expressed.gene.promoters, adult.brain.gr)
(adult.brain.overlap.percentage.H3K4me3 = length(adult.brain.overlap.H3K4me3) / length(sorted_significant
```

```
## [1] 52.60448
```

The percentage of overlap between the promoters of differentially expressed genes found in Part5 of the Analysis Report (Differential Expression Analysis) and the epigenetically marked (**H3K4me3**) promoters in the **liver cell line** is calculated as following:

```
# subsetByOverlaps() extracts the elements in the query (the first argument) that overlap at least one
liver.line.overlap.H3K4me3 = subsetByOverlaps(diff.expressed.gene.promoters, liver.line.gr)
(liver.line.overlap.percentage.H3K4me3 = length(liver.line.overlap.H3K4me3) / length(sorted_significant
```

```
## [1] 47.87313
```