Tim Tran Genomics Data Science

Gene Set analysis

The Bioconductor package AnnotationHub is used for this gene set analysis. The human hg19 annotation database is downloaded from this Bioconductor package.The Fetal brain, adult brain, and adult liver epigenetic narrow peaks were downloaded by using ‘H3K4me3’ and repective cell line ID as keywords.

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

Load the differentially expressed gene list:

```
de_genes <- readRDS("DE_genes.rds")
ah <- AnnotationHub()
```

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

Get the epigenetic narrow peaks data for the fetal brain cell line

```
ah_fetal <- query(ah, c("EpigenomeRoadMap", "H3K4me3", "E081"))
```

```
> head(ah_fetal)
AnnotationHub with 6 records
# snapshotDate(): 2016-10-11
# $dataprovider: BroadInstitute
# $species: Homo sapiens
# $rdataclass: BigWigFile, GRanges
# additional mcols(): taxonomyid, genome, description, tags,
#   sourceurl, sourcetype
# retrieve records with, e.g., 'object[["AH29448"]]'

            title
  AH29448 | E081-H3K4me3.broadPeak.gz
  AH30471 | E081-H3K4me3.narrowPeak.gz
  AH31464 | E081-H3K4me3.gappedPeak.gz
  AH32625 | E081-H3K4me3.fc.signal.bigwig
  AH33657 | E081-H3K4me3.pval.signal.bigwig
  AH40259 | E081-H3K4me3.imputed.pval.signal.bigwig
```

Get the epigenetic narrow peaks data for the adult brain cell line

```
ah_adult <- query(ah, c("EpigenomeRoadMap", "H3K4me3", "E073"))

> head(ah_adult)
AnnotationHub with 6 records
# snapshotDate(): 2016-10-11
# $dataprovider: BroadInstitute
# $species: Homo sapiens
# $rdataclass: BigWigFile, GRanges
# additional mcols(): taxonomyid, genome, description, tags,
#   sourceurl, sourcetype
# retrieve records with, e.g., 'object[["AH29392"]]'

            title
  AH29392 | E073-H3K4me3.broadPeak.gz
  AH30413 | E073-H3K4me3.narrowPeak.gz
  AH31412 | E073-H3K4me3.gappedPeak.gz
  AH32571 | E073-H3K4me3.fc.signal.bigwig
  AH33603 | E073-H3K4me3.pval.signal.bigwig
  AH40251 | E073-H3K4me3.imputed.pval.signal.bigwig
```

Get the epigenetic data peaks for the liver brain

```
ah_liver <- query(ah, c("EpigenomeRoadMap", "H3K4me3", "Liver"))


fetal_gr <- ah_fetal[[2]]
adult_gr <- ah_adult[[2]]
liver_gr <- ah_liver[[2]]
```

Since my differential gene list is in gene name. I need to convert the genename into gene_id using the mygene package.

```
library(mygene)
queryMany(gene_list, scopes = "symbol", fields = "entrezgene", species = "human" )
```

Loaded the differential gene list and use the annotation human database to generate a genomic range object DE_gr that contain all promoter ranges of the differentially expressed genes in adult and fetal brain. The subsetByOverlap function is use to compute the percentage overlapped between the Deg genomic range object with the fetal, adult and liver genomic ranges.

```
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
```

```
> gx <- genes(txdb)
> Deg.gr <- promoters(gx[DE_genes$ID %in% gx$gene_id])
> Deg.gr
```

Find the overlapped between the differential expressed gene Deg and the fetal, adult, And liver ranges.

```
fetal_brain_overlap <- length(subsetByOverlaps(Deg.gr, fetal_gr))/
length(de_genes) *100

> fetal_brain_overlap
[1] 13.2


adult_brain_overlap <- length(subsetByOverlaps(Deg.gr, adult_gr))/
length(DE_genes) *100

> adult_brain_overlap
[1] 19.3


liver_overlap <- length(subsetByOverlaps(Deg.gr, liver_gr))/length(DE_genes) *100

> liver_overlap
[1] 18.0
```

I found that about 13% of differentially expressed genes overlapped with the fetal brain and about 19% of differentially expressed genes are found in adult brain. There is about 18% differentially expressed genes are found in liver brain sample. The result showed that some of the genes are expressed in the fetal brain,but changes in the adult brain. It also showed that there are changes in the H3K4me3 in their promoters from fetal to adult brain.