Gene315_chip

Dr Megan Wilson

Semester 1 2018

Using GenomicRanges to find common and unique peak regions

First you need to make sure you have the required packages installed. If not, you can download them from bioconductor using the code below:

```
source("http://bioconductor.org/biocLite.R")
biocLite("GenomicRanges")
biocLite("RIPSeeker")
```

1. Load the required packages into R

```
library("RIPSeeker")
library(GenomicRanges)
```

2. Read the peak region bed files into R for each data set

```
pk1=read.table("H3K4me1_all_peaks.bed")
pk2=read.table("H3K27Ac_all_peaks.bed")
```

3. Convert these into an appropriate data frame

```
pk1.gr=makeGRangesFromDataFrame(pk1, seqnames.field=c("V1"),start.field=c("V2"), end.field=c("V3"))
pk2.gr=makeGRangesFromDataFrame(pk2, seqnames.field=c("V1"),start.field=c("V2"), end.field=c("V3"))
```

4. Generate a subset of peak regions: peak regions that a common to both data sets. In this example, this function will give me a list of the genome regions enriched for both the H3Kme1 and K27me3 modifications in ES cells.

```
subset <- subsetByOverlaps(pk1.gr,pk2.gr)
#How many peak regions were common to both datasets?
length(subset)
```

```
## [1] 41658
```

```
#save the results to a new bed file
export.bed(subset, "subset.bed")
```

5. How many peak regions are unique to the first data set (in my case, H3K4me1, loaded into R as pk1 (above))

```
setdiff <- setdiff(pk1.gr, pk2.gr)
length(setdiff)
```

```
## [1] 124841
```

```
export.bed(setdiff, "setdiff1.bed")
```

6. What about for the second dataset (in this case pk2=H3K27ac)?

```
setdiff <- setdiff(pk2.gr, pk1.gr)
length(setdiff)
```

```
## [1] 109541
```

```
export.bed(setdiff, "setdiff2.bed")
```