rnaSeqMap library in use

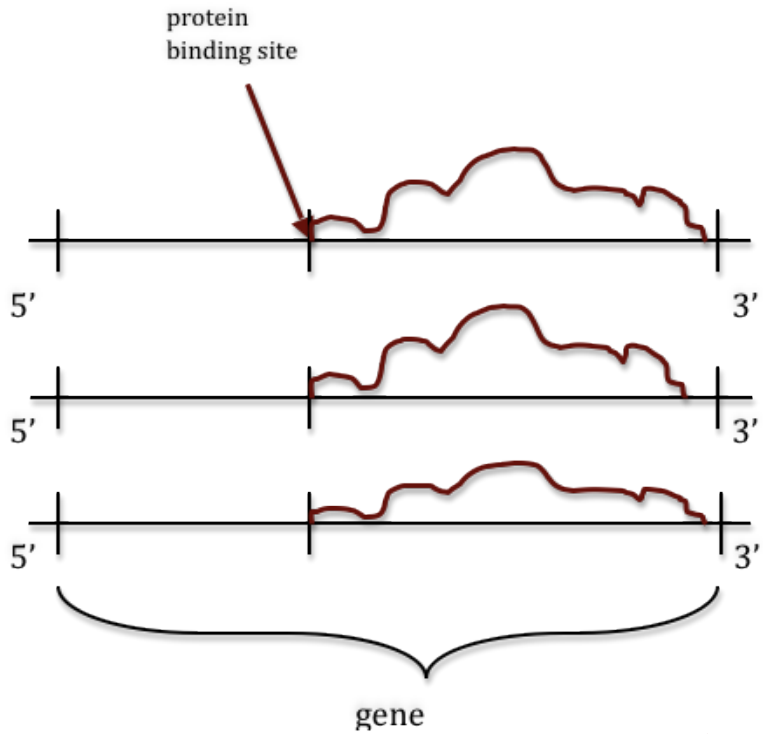
Solving particular biological problems in transcription regulation

Anna Leśniewska, Michał Okoniewski

BioC 2012 European Developers' Workshop Zurich

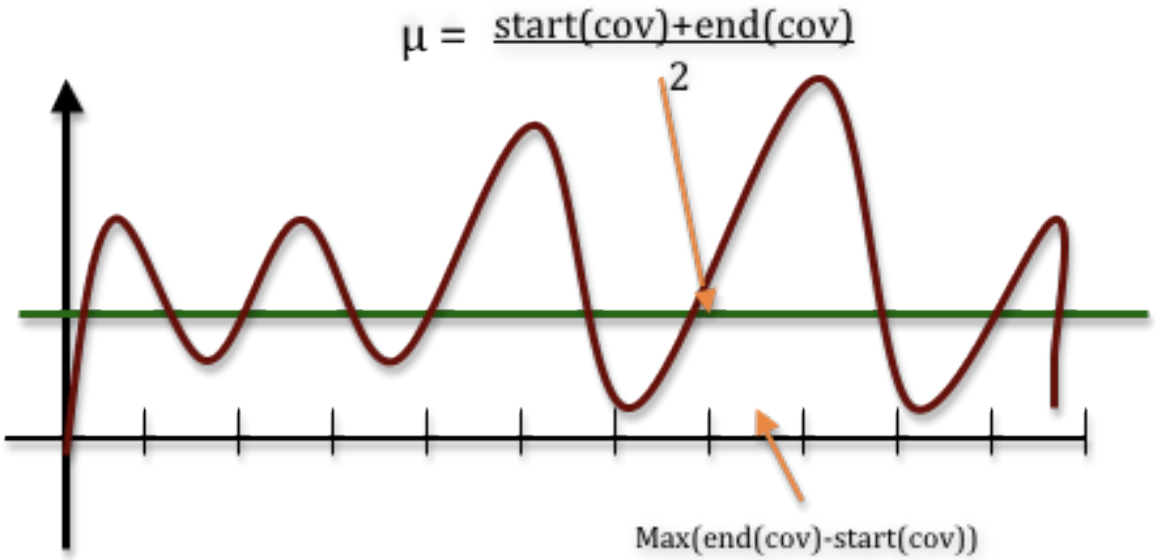
Biological question:

- show genes that have a particular expression pattern in one treatment group
- the patern is a result of a particular protein binding site
- the pattern is closing the transcription in rhybdosarcoma
- the pattern does not follow known exon boundaries



Sliding window approach

- Aumann-Lindell slinding window algorithm with bucketing
  - Split gene region on buckets
  - For each bucket check the coverage on the start and the end
  - Find bucket with $max(end(cov) - start(cov))$
  - Find μ (the threshold for coverage) as$(start(cov) + end(cov))/2$
  - Find region based on μ which start in choosing bucket

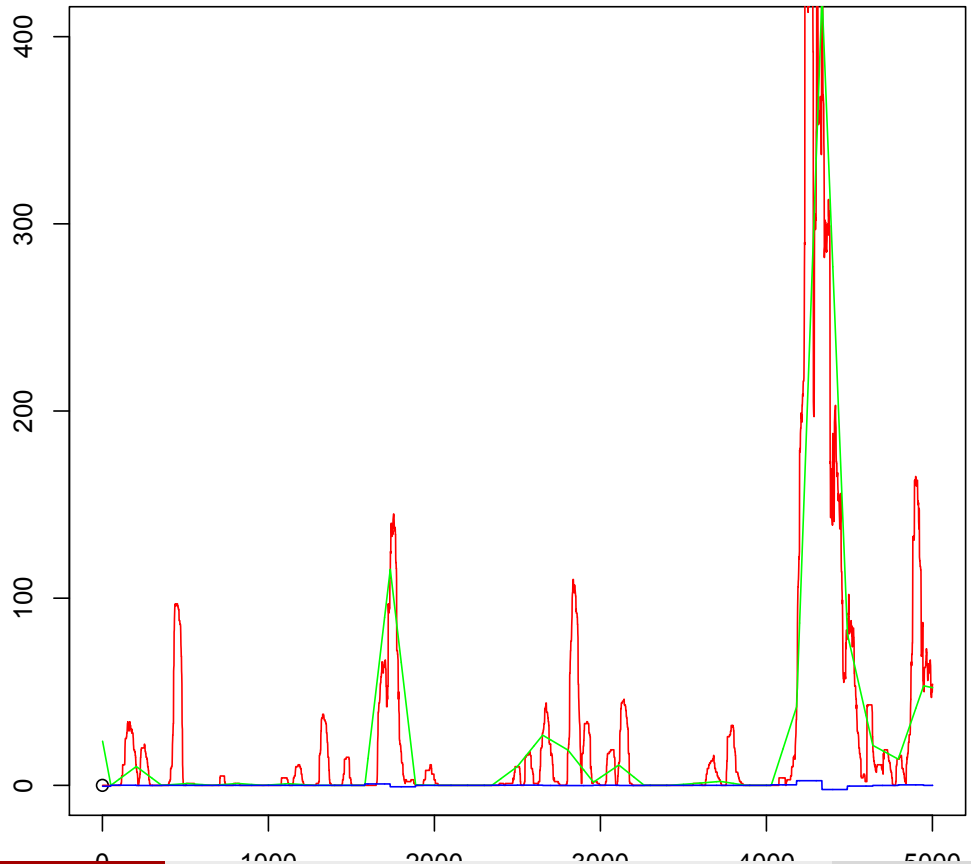


Experiment 1

```
rs<-newSeqReads('chr1',15783223,15798586,1)
rs<-getBamData(rs,1:6)
nd <- getCoverageFromRS(rs,1:6)
...
tab <- bucket(nd,exp,0.05)
...
reg2<-.Call("regionmining",as.integer(startR:nd@end),
as.integer(covR),as.integer(mi),as.integer(minl))
```

Lowess+derivative approach

- Based on the coverage profile:
  - Use lowess function to smooth the profile
  - Use derivative of the coverage profile
  - Find maximum of derivative of the coverage profile



Experiment 2

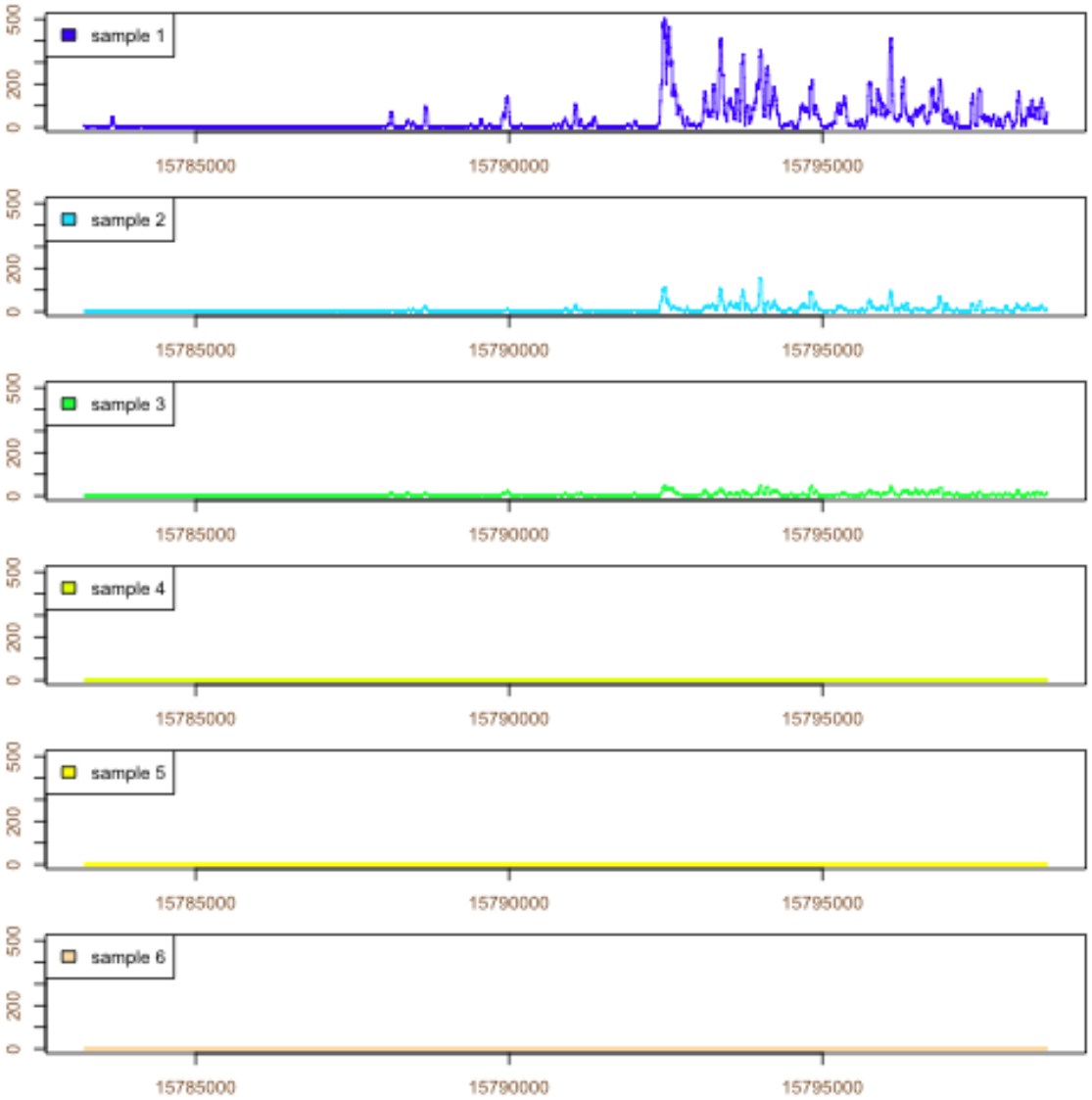
```
rs<-newSeqReads('chr1',15783223,15798586,1)
rs<-getBamData(rs,1:6)
nd <- getCoverageFromRS(rs,1:6)
lnd <- lowessND(nd,0.1,1:6)
...
poch <- derivative(dd1)
...
 tab_position <- poch[1,]

for(i in 1:length(exps)){
k <- which(poch[,i]==(max(poch[,i])))
tab_position[i] <- k[1]
 }
```

Results and conclusions

- With both methods we have found ca 500 candidates, overlap between them is 30%
- The coverage profiles will be visually inspected
- Biologists will check for proximity of the protein binding site
- We will look for dependence between the protein binding site and the pattern (chi-sq, Fisher-exact)

Examplary candidate



Examplary candidate with lowess function