

Differential binding analysis with diffbind

David Lewis

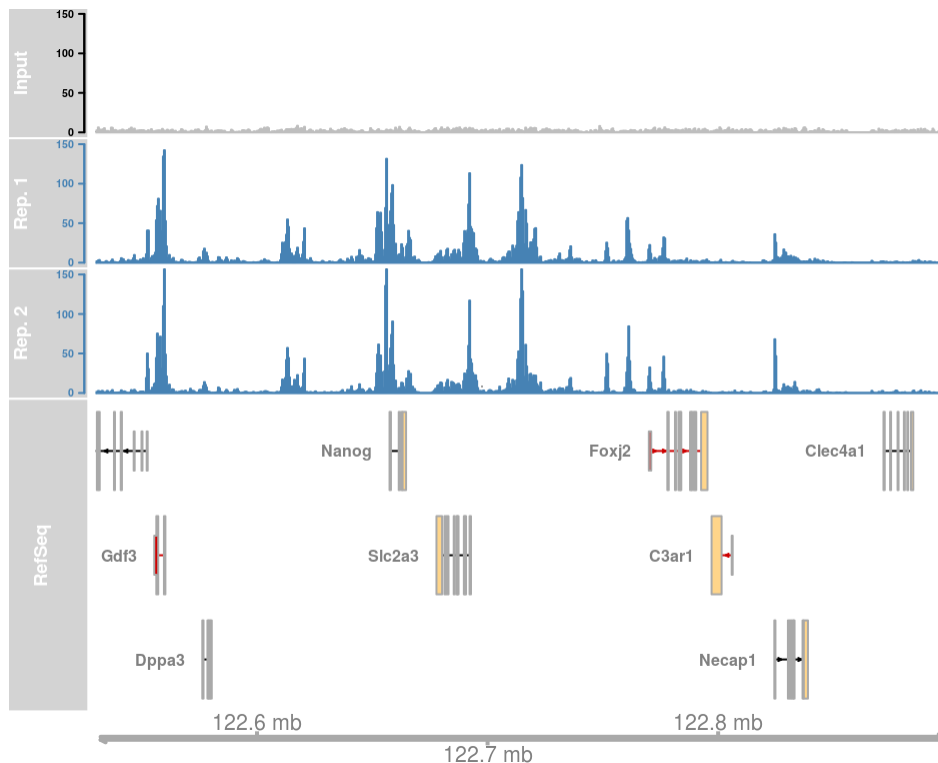
Introduction

- What is ChIP-seq data?
- How is Diffbind used?
- What do we get from Diffbind?

What is ChIP-seq data?

- ChIP-seq: Chromatin Immunoprecipitation sequencing
- Determine protein binding to genome

What is differential binding analysis?



- Difference between the binding site in condition vs control
- Very similar to other differential analysis (differential expression analysis)

Using Diffbind

Downloading the sample data from the vignette

```
url <- "https://content.cruk.cam.ac.uk/bioinformatics/software/DiffBind/DiffBind_vignette_data.tar.gz"
file <- basename(url)
options(timeout = 600)
download.file(url, file.path(".", file))
untar(file.path(".", file), exdir = ".")
```

Examining the metadata

```
head(read.csv("tamoxifen.csv"))[c("Condition", "bamReads")]
```

Condition	bamReads
Resistant	reads/Chr18_BT474_ER_1.bam
Resistant	reads/Chr18_BT474_ER_2.bam
Responsive	reads/Chr18_MCF7_ER_1.bam
Responsive	reads/Chr18_MCF7_ER_2.bam
Responsive	reads/Chr18_MCF7_ER_3.bam
Responsive	reads/Chr18_T47D_ER_1.bam

Run the analysis

```
tamoxifen <- DiffBind::dba.analyze("tamoxifen.csv")
```

```
## Loading sample sheet...
## BT4741 BT474 ER Resistant Full-Media 1 bed
## BT4742 BT474 ER Resistant Full-Media 2 bed
## MCF71 MCF7 ER Responsive Full-Media 1 bed
## MCF72 MCF7 ER Responsive Full-Media 2 bed
## MCF73 MCF7 ER Responsive Full-Media 3 bed
## T47D1 T47D ER Responsive Full-Media 1 bed
## T47D2 T47D ER Responsive Full-Media 2 bed
## MCF7r1 MCF7 ER Resistant Full-Media 1 bed
## MCF7r2 MCF7 ER Resistant Full-Media 2 bed
## ZR751 ZR75 ER Responsive Full-Media 1 bed
## ZR752 ZR75 ER Responsive Full-Media 2 bed
## Applying Blacklist/Greylists...
## Genome detected: Hsapiens.UCSC.hg19
```

```

## Applying blacklist...
## Removed: 3 of 14102 intervals.
## Counting control reads for greylist...
## Building greylist: reads/Chr18_BT474_input.bam
## coverage: 166912 bp (0.21%)
## Building greylist: reads/Chr18_MCF7_input.bam
## coverage: 106495 bp (0.14%)
## Building greylist: reads/Chr18_T47D_input.bam
## coverage: 56832 bp (0.07%)
## Building greylist: reads/Chr18_TAMR_input.bam
## coverage: 123903 bp (0.16%)
## Building greylist: reads/Chr18_ZR75_input.bam
## coverage: 68608 bp (0.09%)
## BT474c: 58 ranges, 166912 bases
## MCF7c: 14 ranges, 106495 bases
## T47Dc: 11 ranges, 56832 bases
## TAMRc: 11 ranges, 123903 bases
## ZR75c: 12 ranges, 68608 bases
## Master greylist: 70 ranges, 252415 bases
## Removed: 420 of 14099 intervals.
## Removed: 52 merged (of 3795) and 50 (of 2845) consensus.
## Forming consensus peakset and counting reads...
## Computing summits...
## Re-centering peaks...
## Normalize DESeq2 with defaults...
## Forming default model design and contrast(s)...
## Computing results names...
## Analyzing...
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates

```

```
DiffBind::dba.report(tamoxifen)
```

```

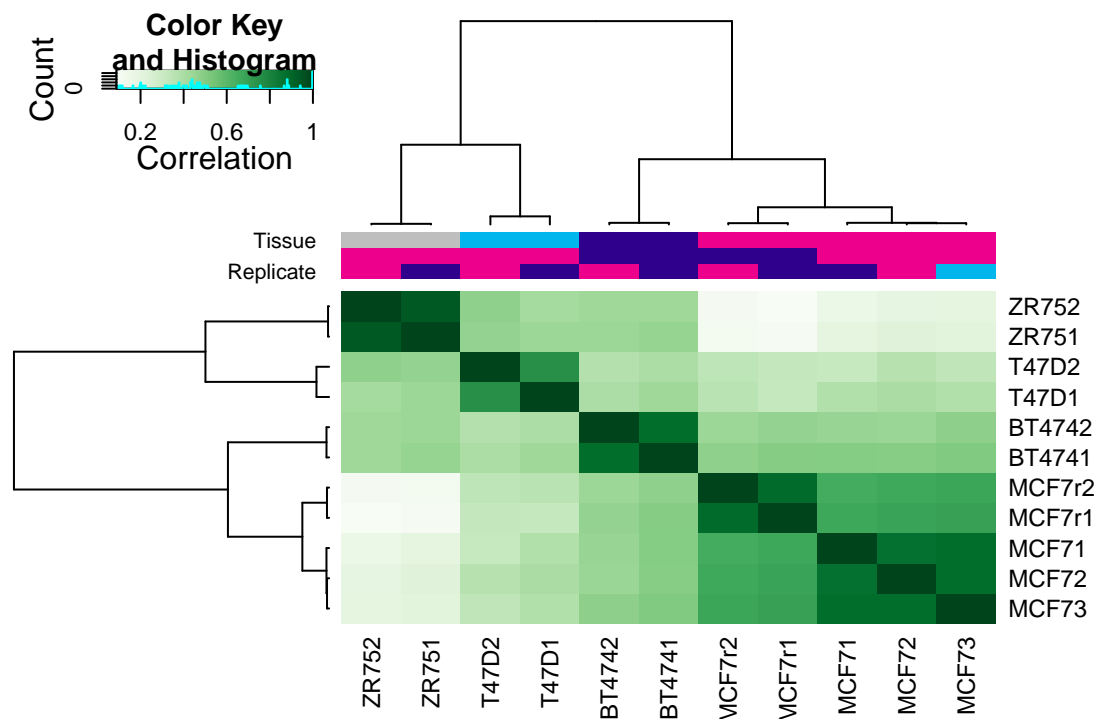
## GRanges object with 246 ranges and 6 metadata columns:
##      seqnames      ranges strand |      Conc Conc_Responsive
##      <Rle>         <IRanges> <Rle> | <numeric>      <numeric>
##    943   chr18 26861047-26861447   * |    8.37281        8.98991

```

```
## 2422 chr18 65030123-65030523 * | 6.02482 6.62520
## 1449 chr18 41369550-41369950 * | 8.25529 8.84621
## 2404 chr18 64490736-64491136 * | 7.42827 8.05134
## 2290 chr18 60892950-60893350 * | 8.09133 8.72229
## ... ... ... ...
## 1187 chr18 33021825-33022225 * | 5.04267 5.57838
## 408 chr18 11320959-11321359 * | 4.29417 3.38332
## 1945 chr18 53054738-53055138 * | 6.44506 5.53915
## 218 chr18 7757228-7757628 * | 5.04773 3.56723
## 1370 chr18 38482793-38483193 * | 4.23057 4.76394
## Conc_Resistant Fold p-value FDR
## <numeric> <numeric> <numeric> <numeric>
## 943 4.44855 3.10593 1.95264e-08 3.58819e-05
## 2422 2.65613 2.87530 3.13216e-08 3.58819e-05
## 1449 5.12414 2.75925 4.18808e-08 3.58819e-05
## 2404 3.23712 3.09946 5.23823e-08 3.58819e-05
## 2290 3.44567 3.14110 8.21331e-08 4.50090e-05
## ... ... ... ...
## 1187 2.81227 1.47761 0.00426958 0.0483415
## 408 5.15749 -1.31788 0.00433276 0.0488549
## 1945 7.30585 -1.28749 0.00438440 0.0492346
## 218 6.13374 -1.46288 0.00440236 0.0492346
## 1370 2.02778 1.47251 0.00443285 0.0493740
## -----
## seqinfo: 1 sequence from an unspecified genome; no seqlengths
```

```
tamoxifen <- dba(sampleSheet = "tamoxifen.csv") %>%
+dba.blacklist() %>%
+dba.count() %>%
+dba.normalize() %>%
+dba.contrast() %>%
+dba.analyze()
```

```
plot(tamoxifen)
```

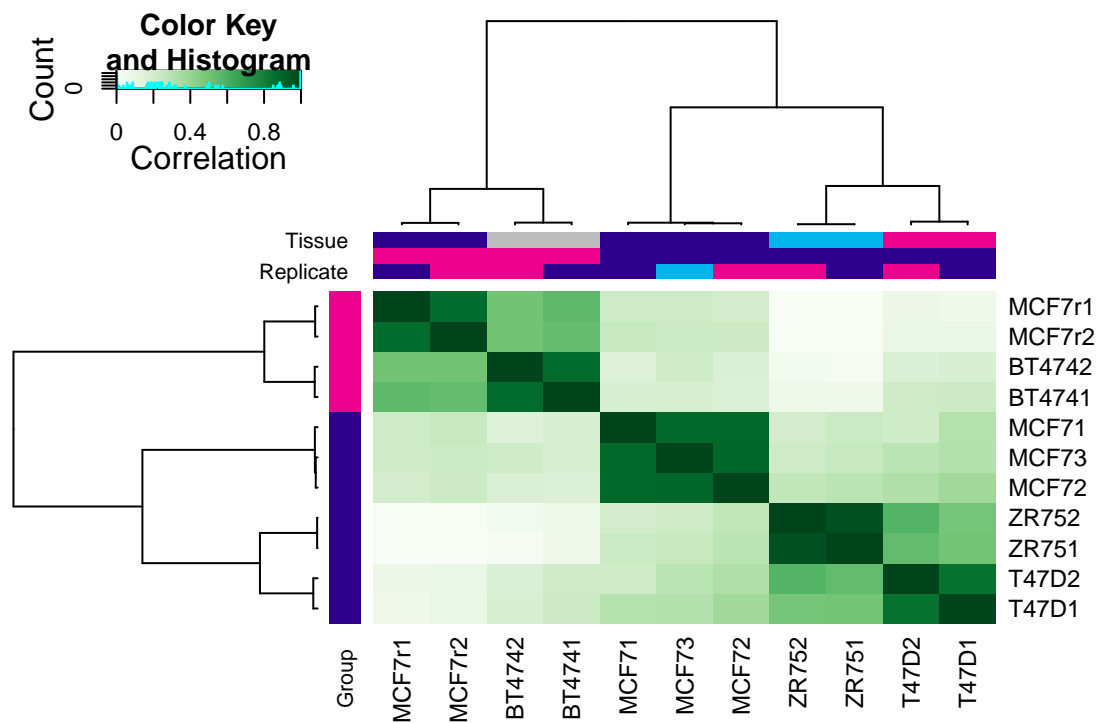


```
DiffBind::dba.show(tamoxifen, bContrasts = TRUE)
```

Factor	Group	Samples	Group2	Samples2	DB.DESeq2
Condition	Responsive	7	Resistant	4	246

Using only the 246 Genes.

```
plot(tamoxifen, contrast = 1)
```



Plotting With DiffBind

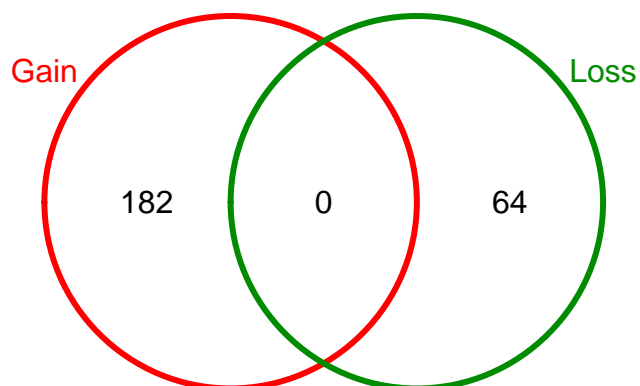
- DiffBind Provides functions to plot aspects of the data.

-
- Gain: increase binding enrichment in the Resistant condition (Positive Fold Change)
 - Loss: lower enrichment (Negative Fold Change)

```
DiffBind::dba.plotVenn(tamoxifen, contrast = 1, bDB = TRUE, bGain = TRUE, bLoss = TRUE, bAll = FALSE)
```

```
## Generating report-based DBA object...
```

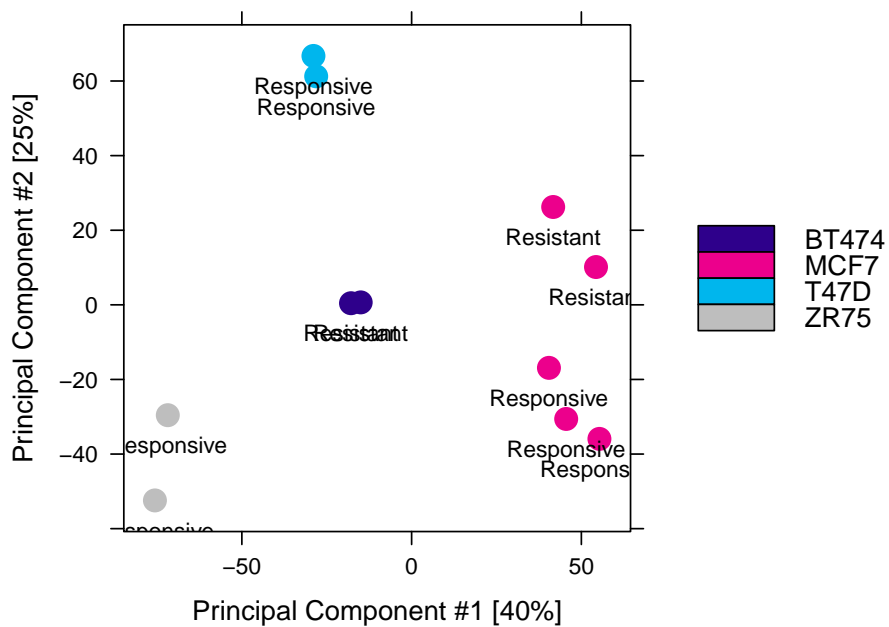
Binding Site Overlaps



Responsive vs. Resistant:DB:DESeq2

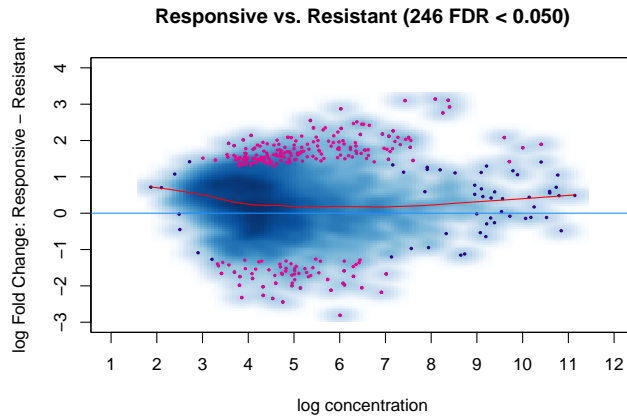
```
DiffBind::dba.plotPCA(tamoxifen, DiffBind::DBA_TISSUE, label = DiffBind::DBA_CONDITION)
```

PCA: Tissue

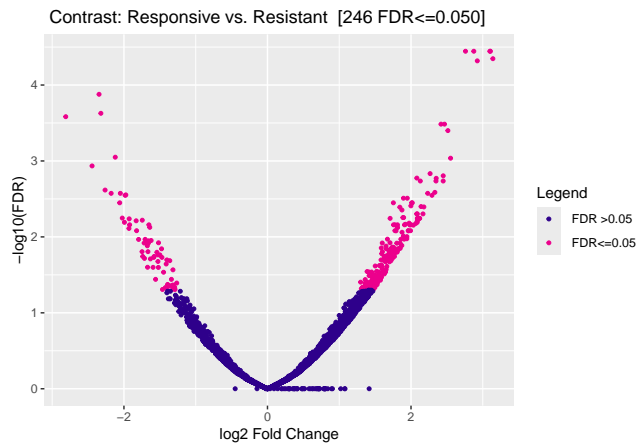


MA Plot

```
DiffBind::dba.plotMA(tamoxifen)
```



```
DiffBind::dba.plotVolcano(tamoxifen)
```



Deliverables

- Table of differentially bound regions of the genome (binding sites)
- Plots

Under the hood

- Either Deseq2 or EdgeR is used for differential binding analysis
- These tools are usually used for differentially expression analysis

Deseq2

- Fits a negative binomial distribution to the data
- Quantifies expected gene (or “read”) variance from mean
- Genes that defy this expectation are considered “significant”

Conclusion

- What is ChIP-seq data?
- How is Diffbind used?
- What do we get from Diffbind?

Diffbind utilizes commonly used differential expression analysis tools to determine differentially bound sites on the genome.