

Differential binding analysis

Bori Mifsud

23.02.2018

Advanced ChIP-seq data analysis
course



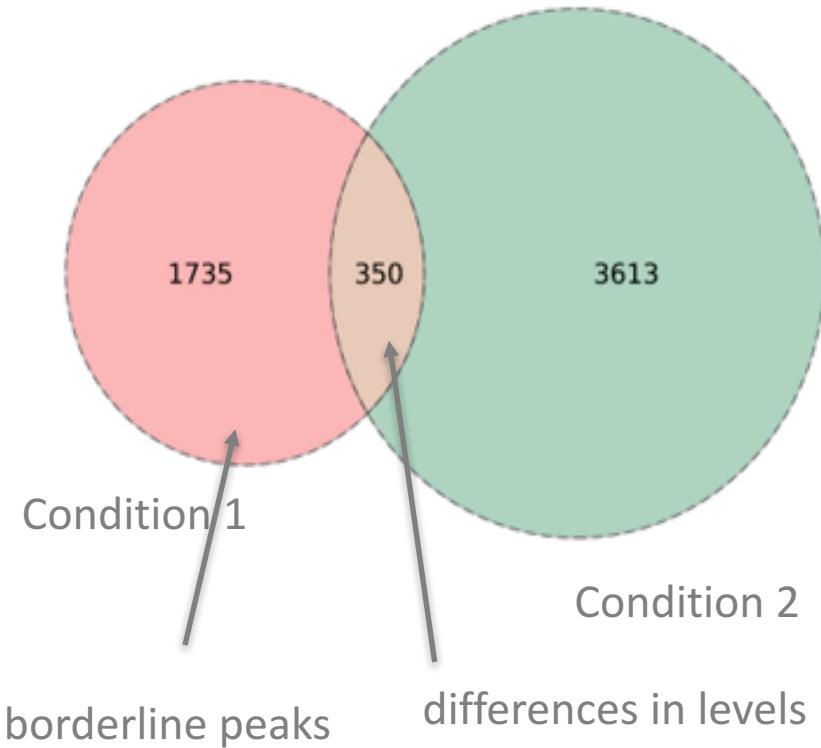
Queen Mary
University of London

Why do differential binding analysis?

- Which regions are regulated differently in two conditions (TFs)
- Which regions have different chromatin environment in two conditions (histone modifications)

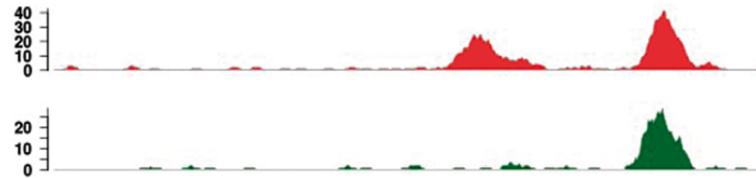


Occupancy based analysis

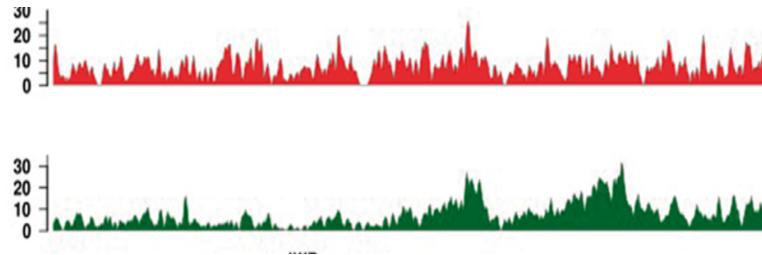


Affinity based analysis

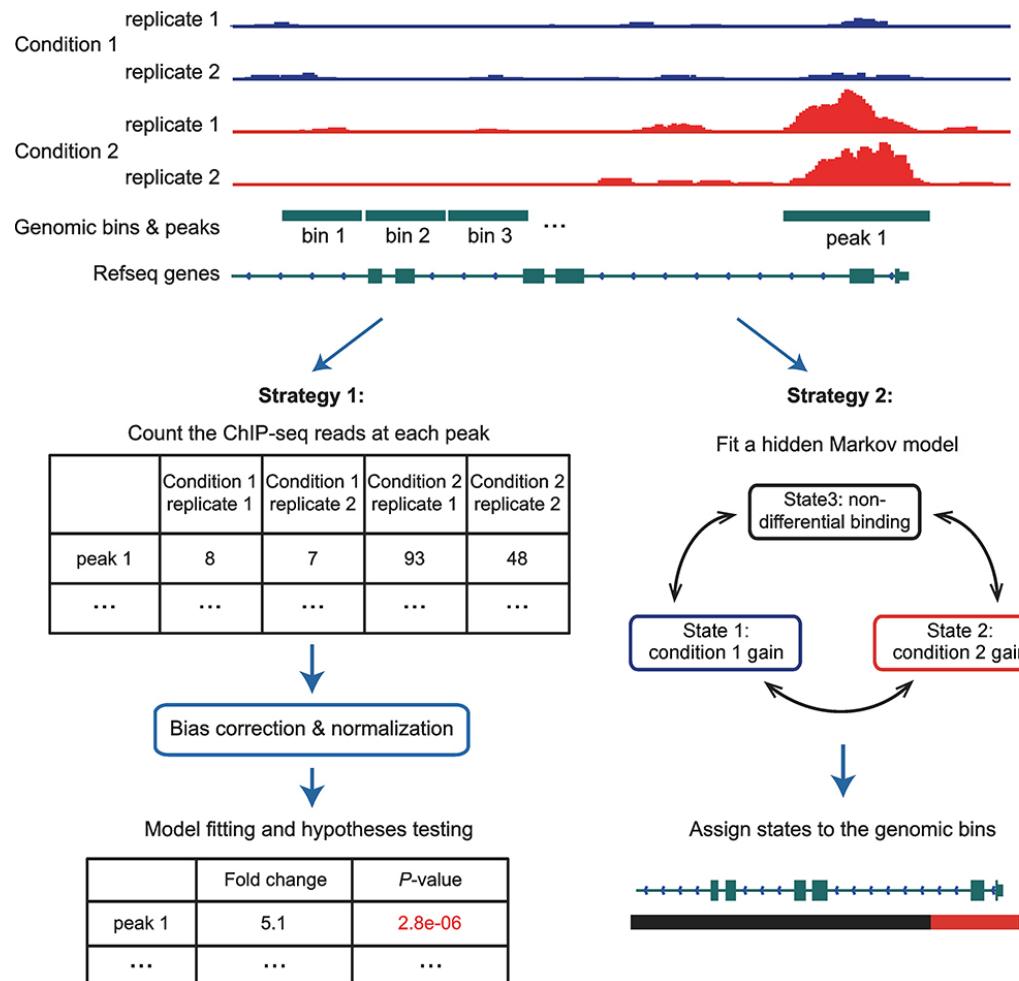
TF



H3K36me3

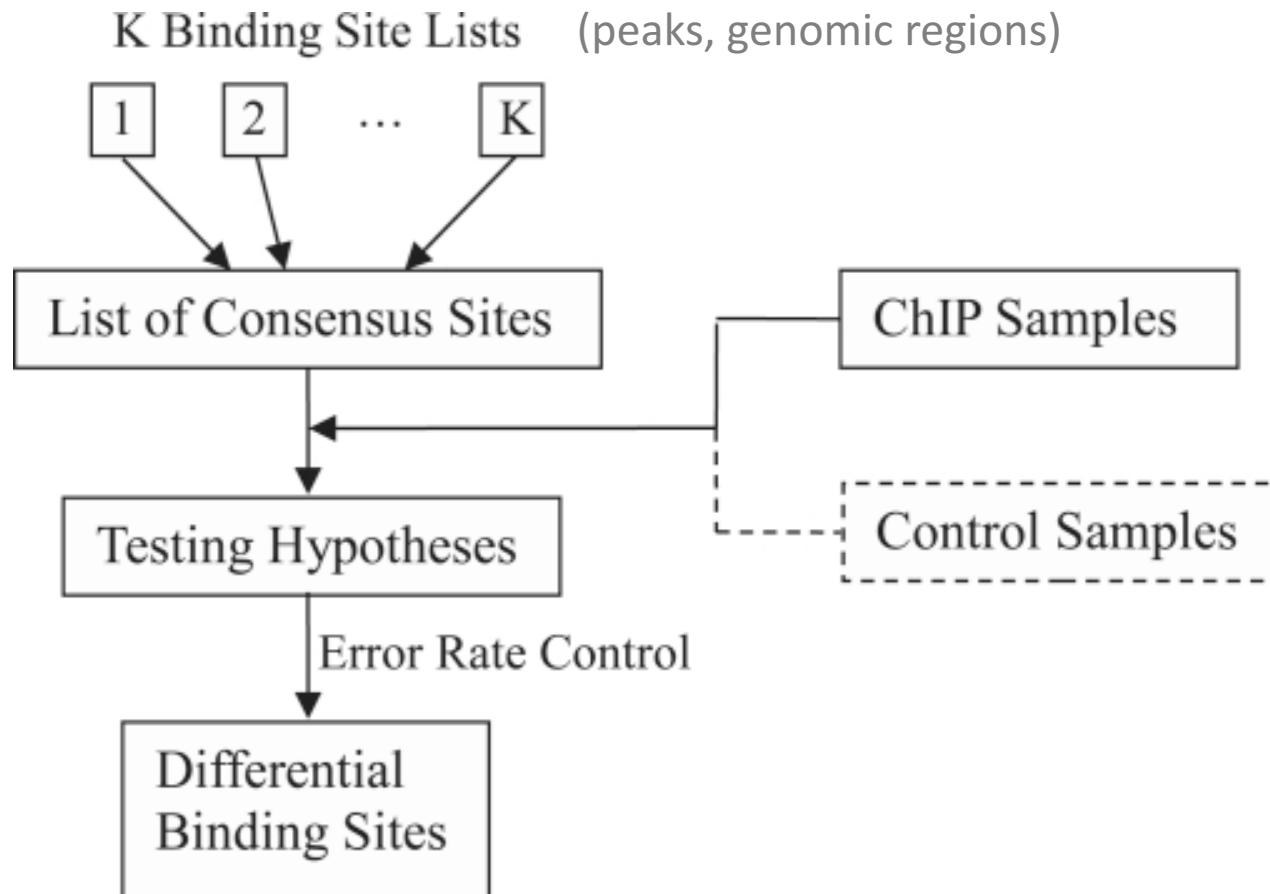


Two main strategies for differential binding analysis (counts at peaks – HMM at genomic bins)



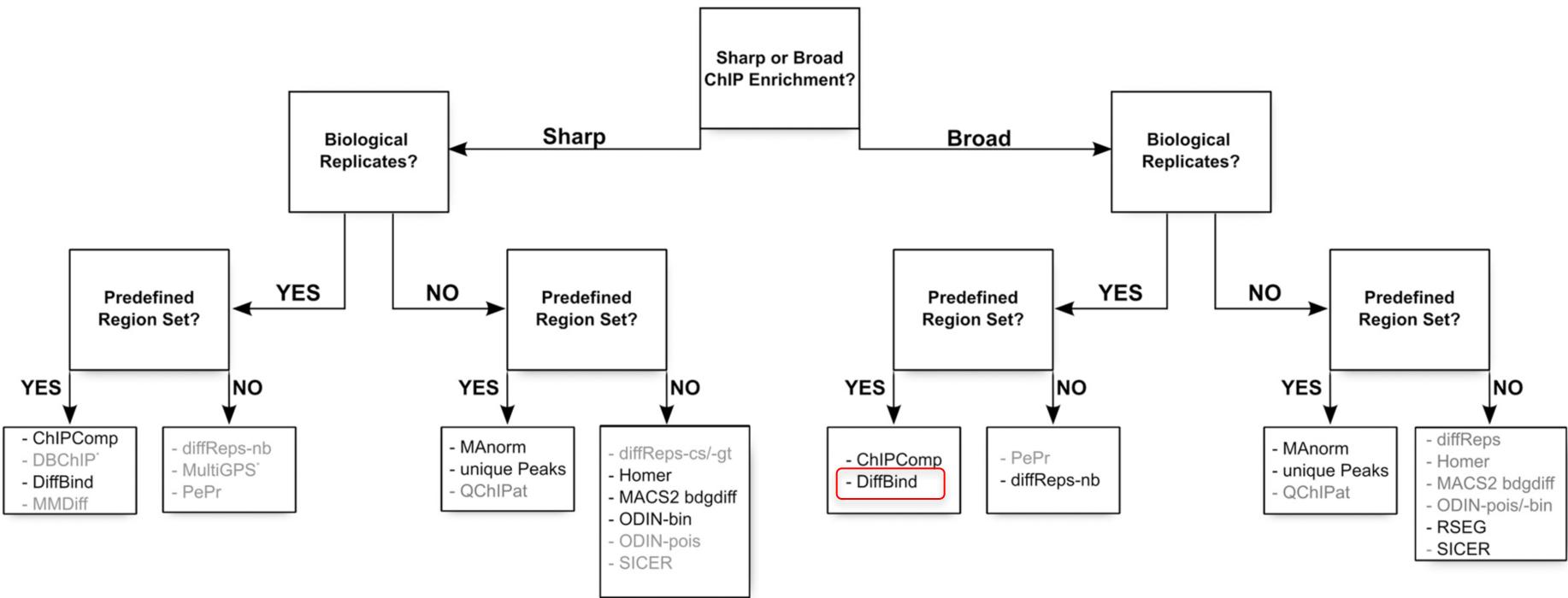
(Tu, 2017)

Workflow of differential binding for count based methods



(Liang, 2012)

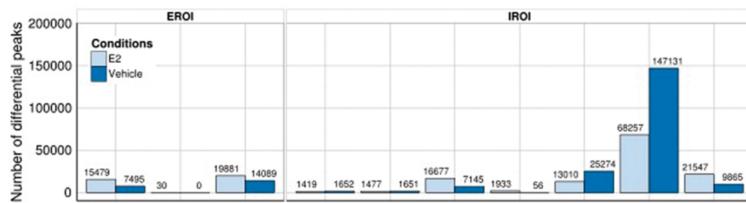
Which tool to use for differential binding?



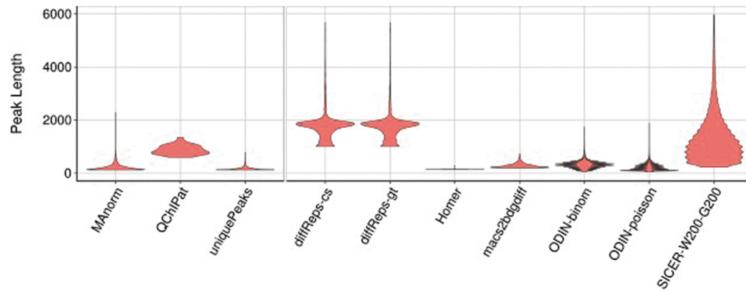
(Steinhauser, 2016)

Comparison of differential binding analysis tools (number, length)

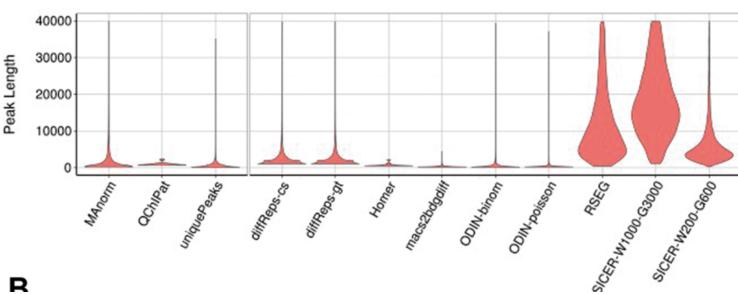
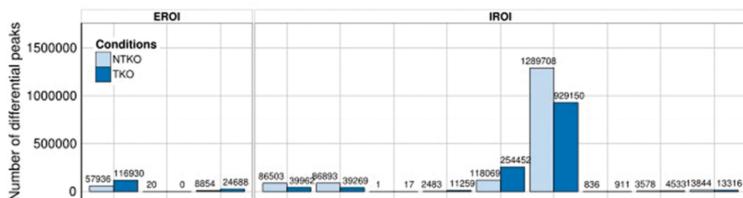
A



TF

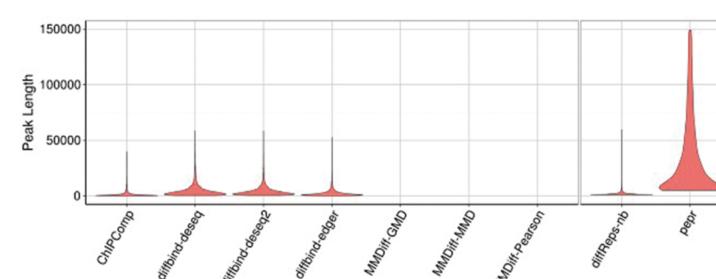
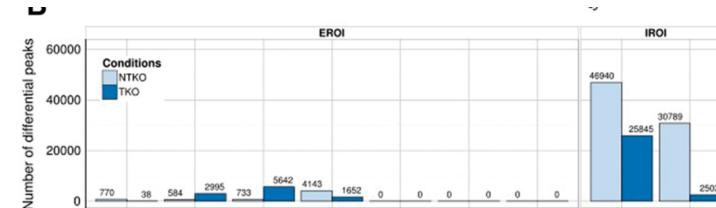
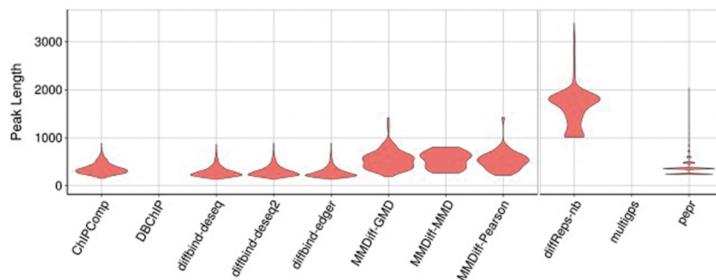
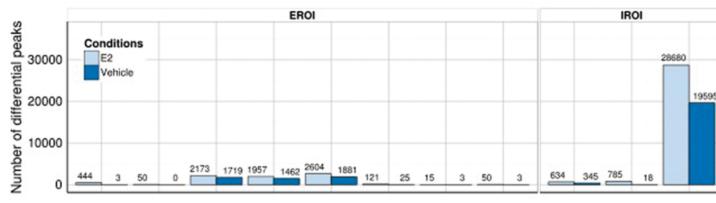


histone modification



R

B



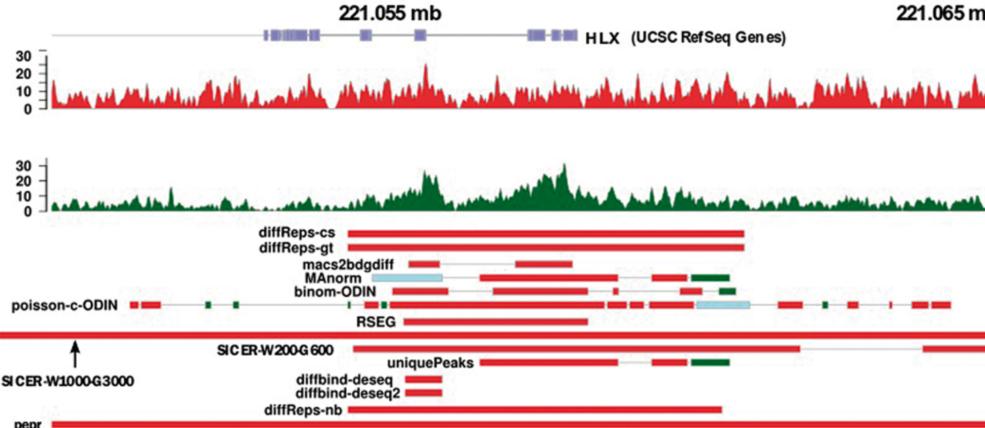
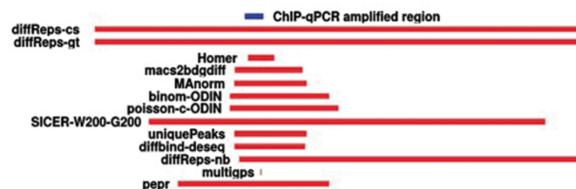
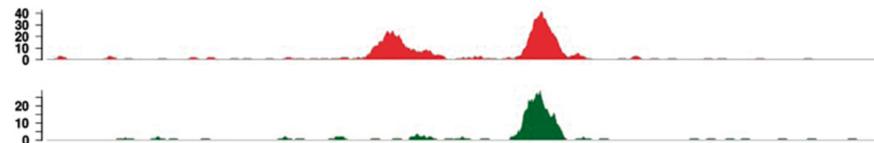
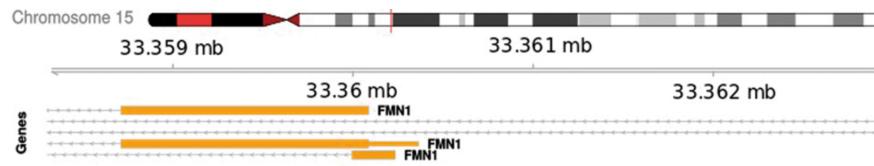
(Steinhauser,2016)



Comparison of differential binding analysis tools (different sizes)

TF

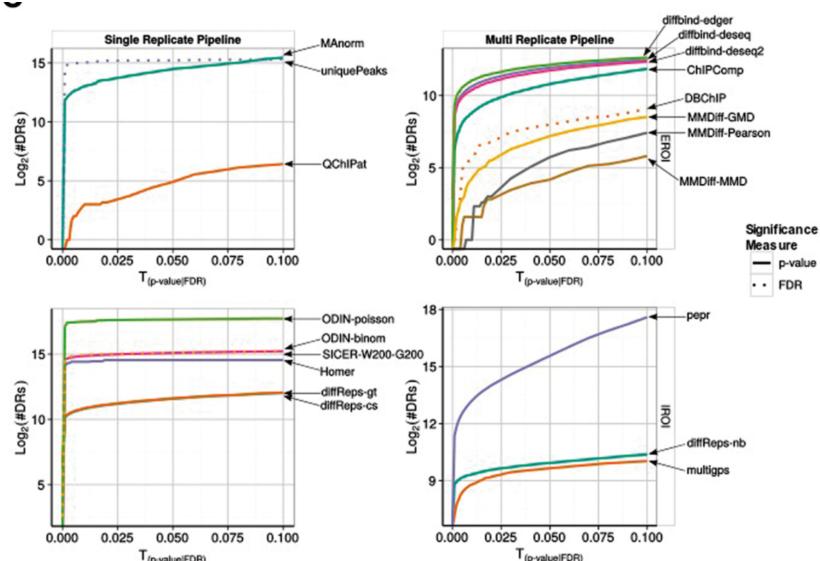
histone modification



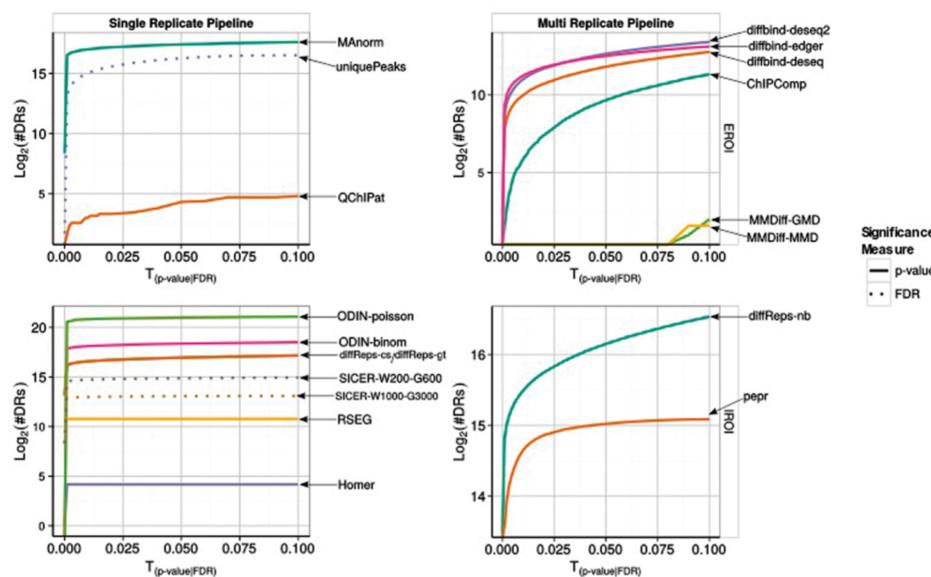
(Steinhauser, 2016)

Comparison of differential binding analysis tools (number vs. pvalue)

TF



histone modification



(Steinhauser, 2016)

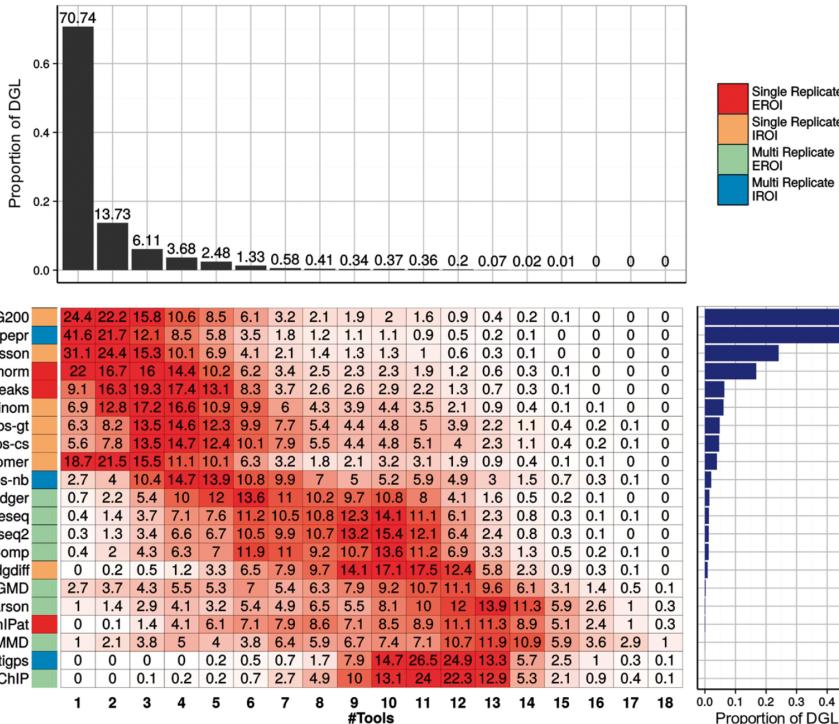
Comparison of differential binding analysis tools

TF

histone modification

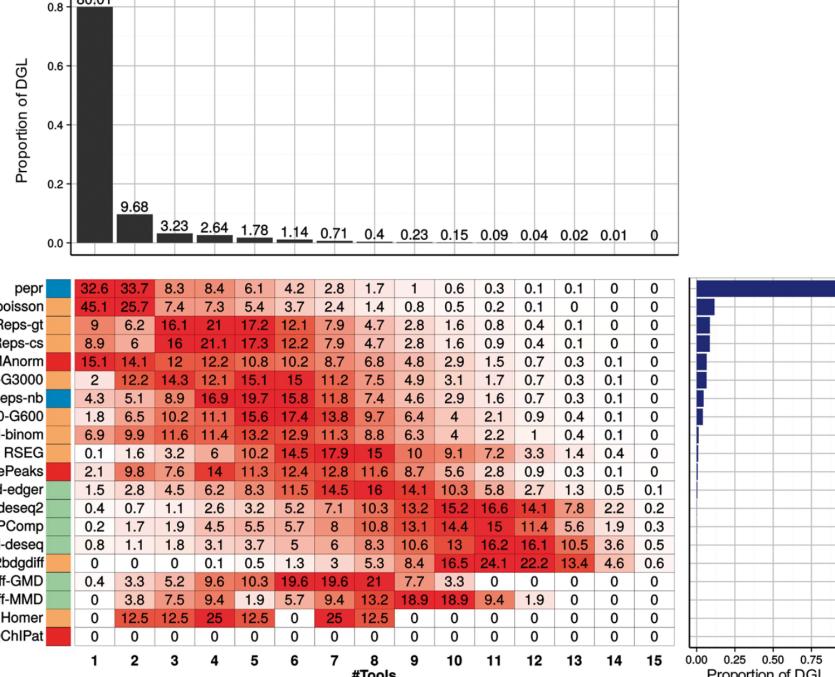
A

FoxA1 E2



B

H3K36me3 NTKO



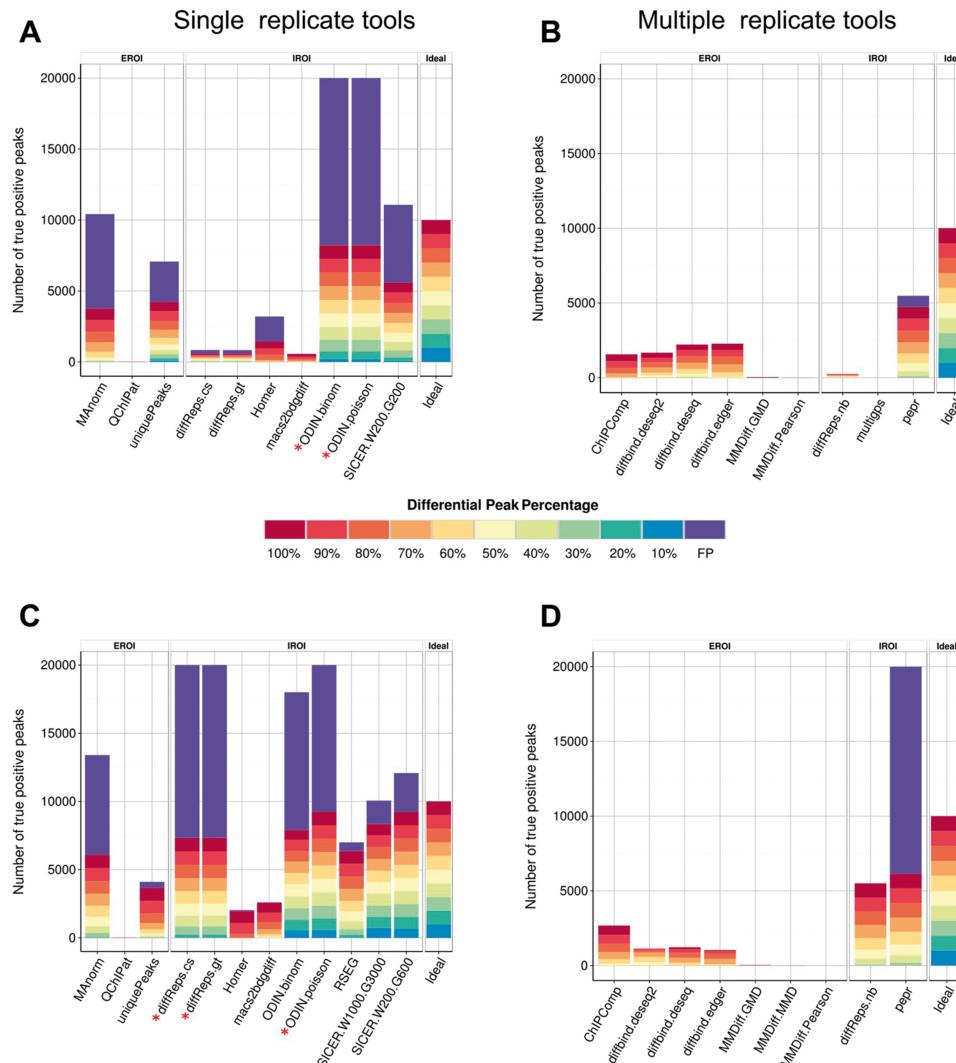
(Steinhauser,2016)

Comparison of differential binding analysis tools – simulated data

histone modification

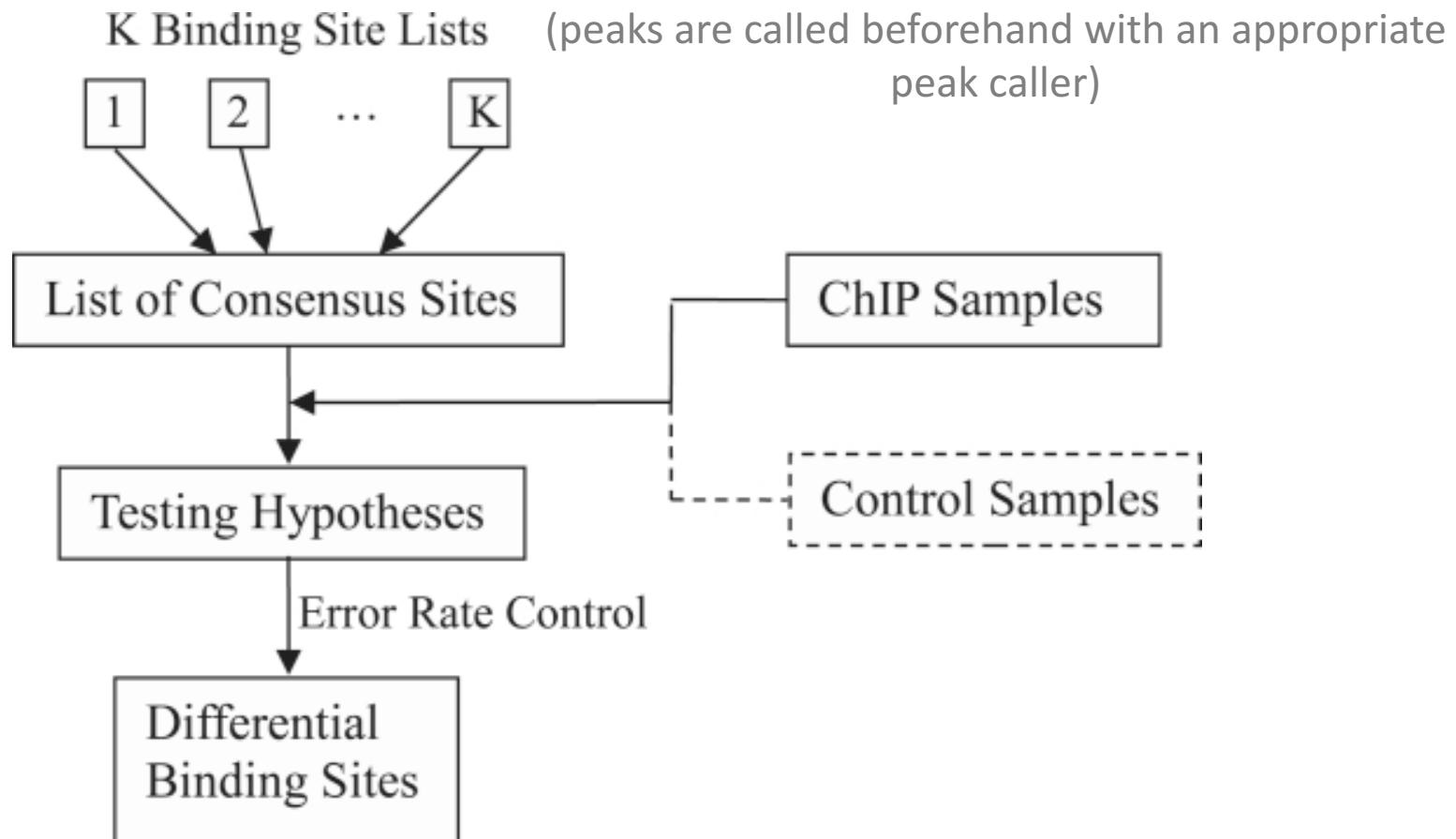
TF

Broad ChIP-seq signal: H3K36me3



(Steinhauser, 2016)

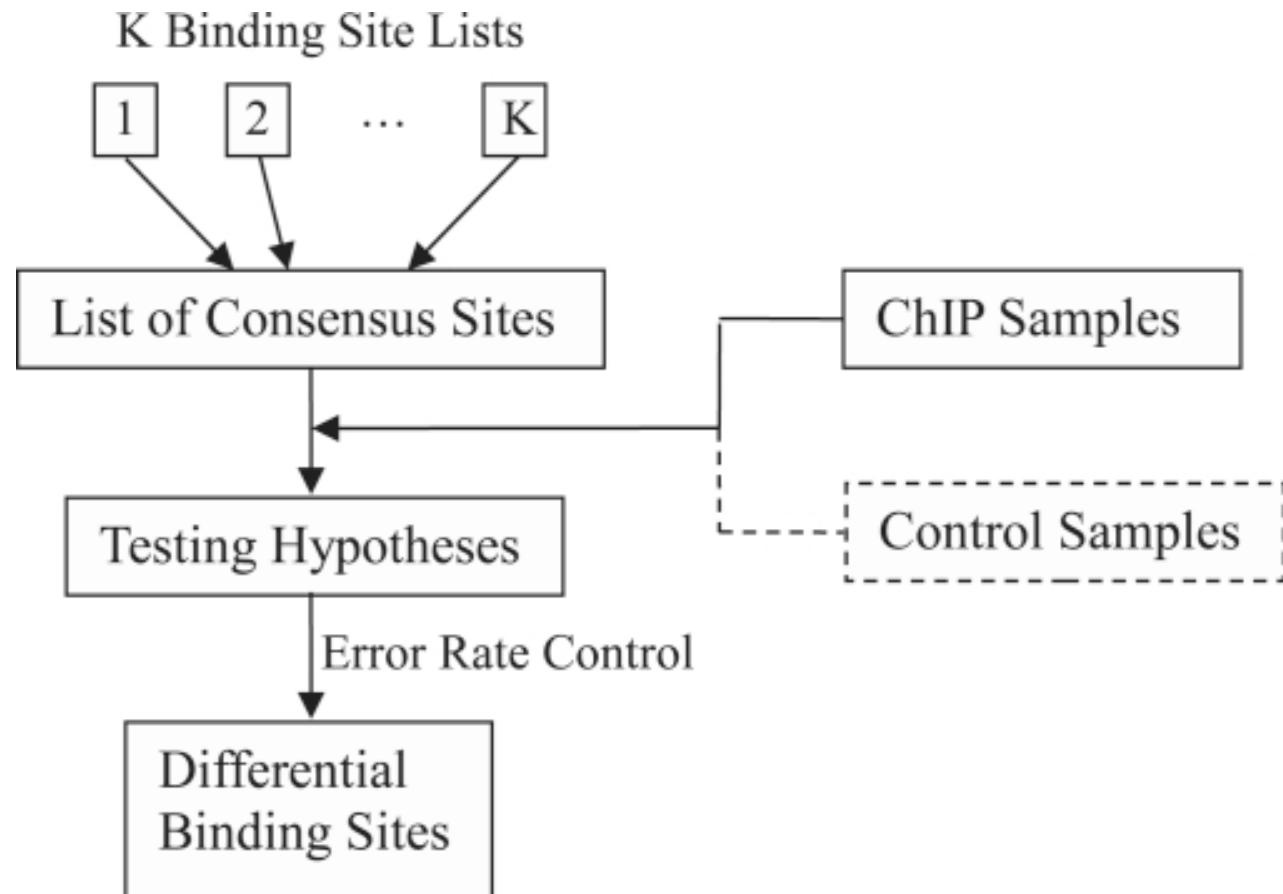
Workflow of DiffBind



(Liang, 2012)

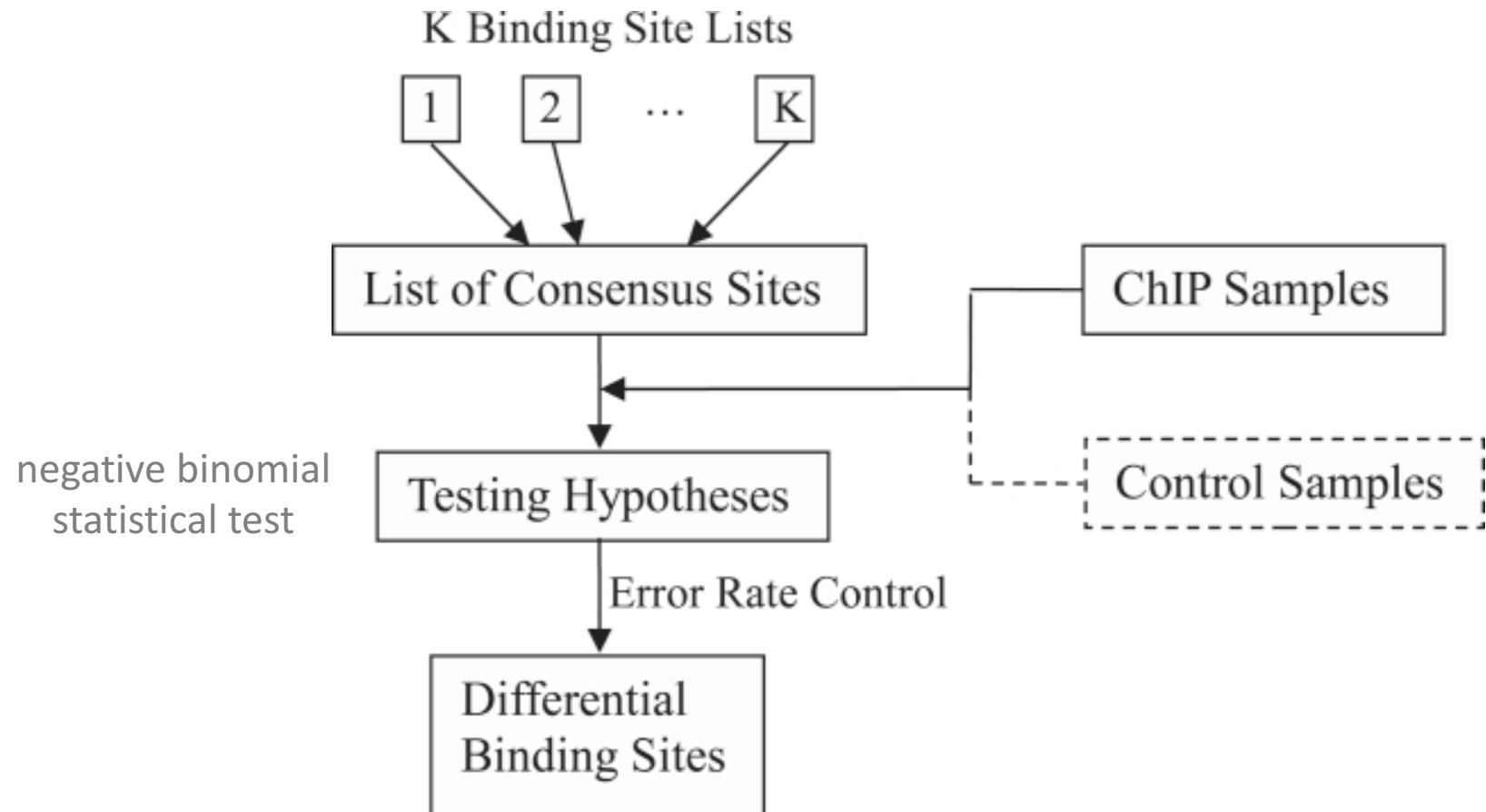
Workflow of DiffBind

Consensus sites are chosen within DiffBind (number of replicates a peak is present)



(Liang, 2012)

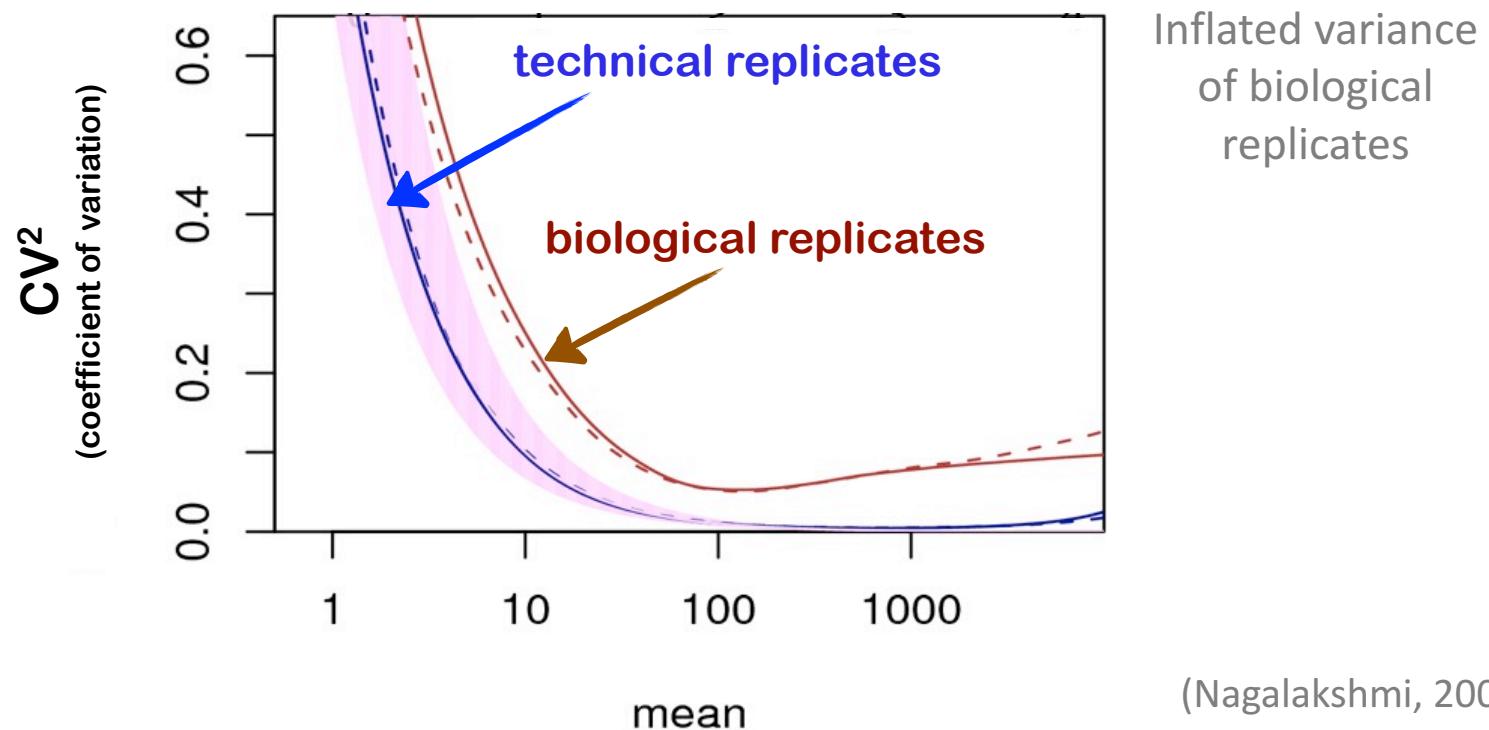
Workflow of DiffBind



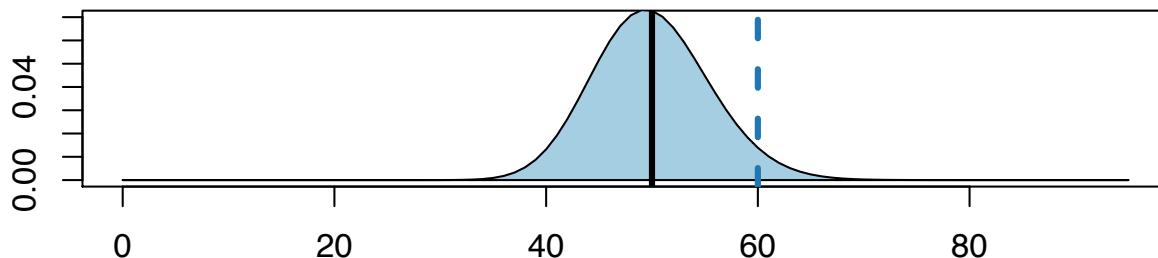
(Liang, 2012)

Negative binomial test accounts for biological variability

- Sequencing data can be modelled with the Poisson distribution
BUT
In the the Poisson distribution the variance=mean



Negative binomial (NB) test accounts for biological variability



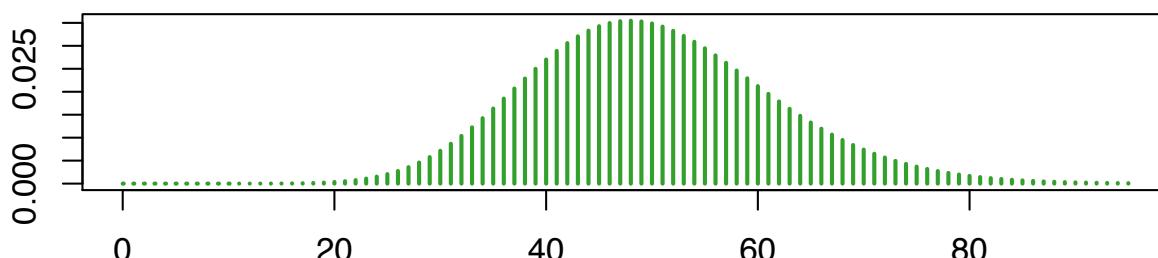
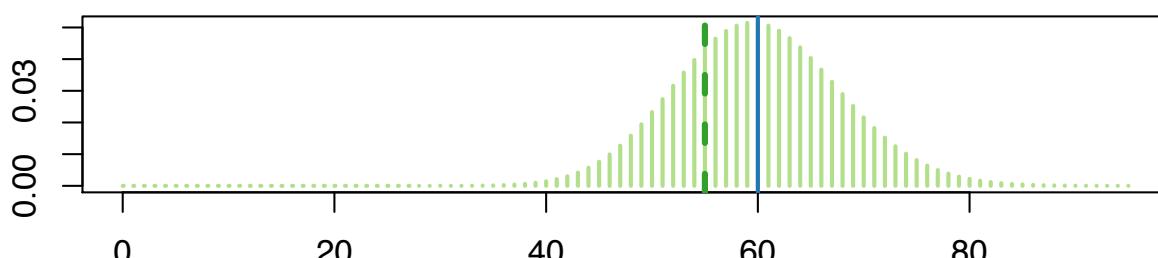
Biological sample to sample variability Γ



Poisson counting statistics Λ



Overall distribution NB



$$NB(\mu, \sigma^2 + \mu) = \Lambda(\Gamma(\mu, \sigma^2))$$

Used both in DESeq and EdgeR R packages, that provide the statistical test in DiffBind

Important considerations for differential binding analysis

- Compare samples where the sequencing reads are the same length -mappability (avoid trimming unless necessary, treat samples the same way)
- Use uniquely aligning reads
- Defining the regions to test is vital, good peak calling is needed (deduplicate reads before)
- To keep full dynamic range put back duplicate reads for the differential analysis

