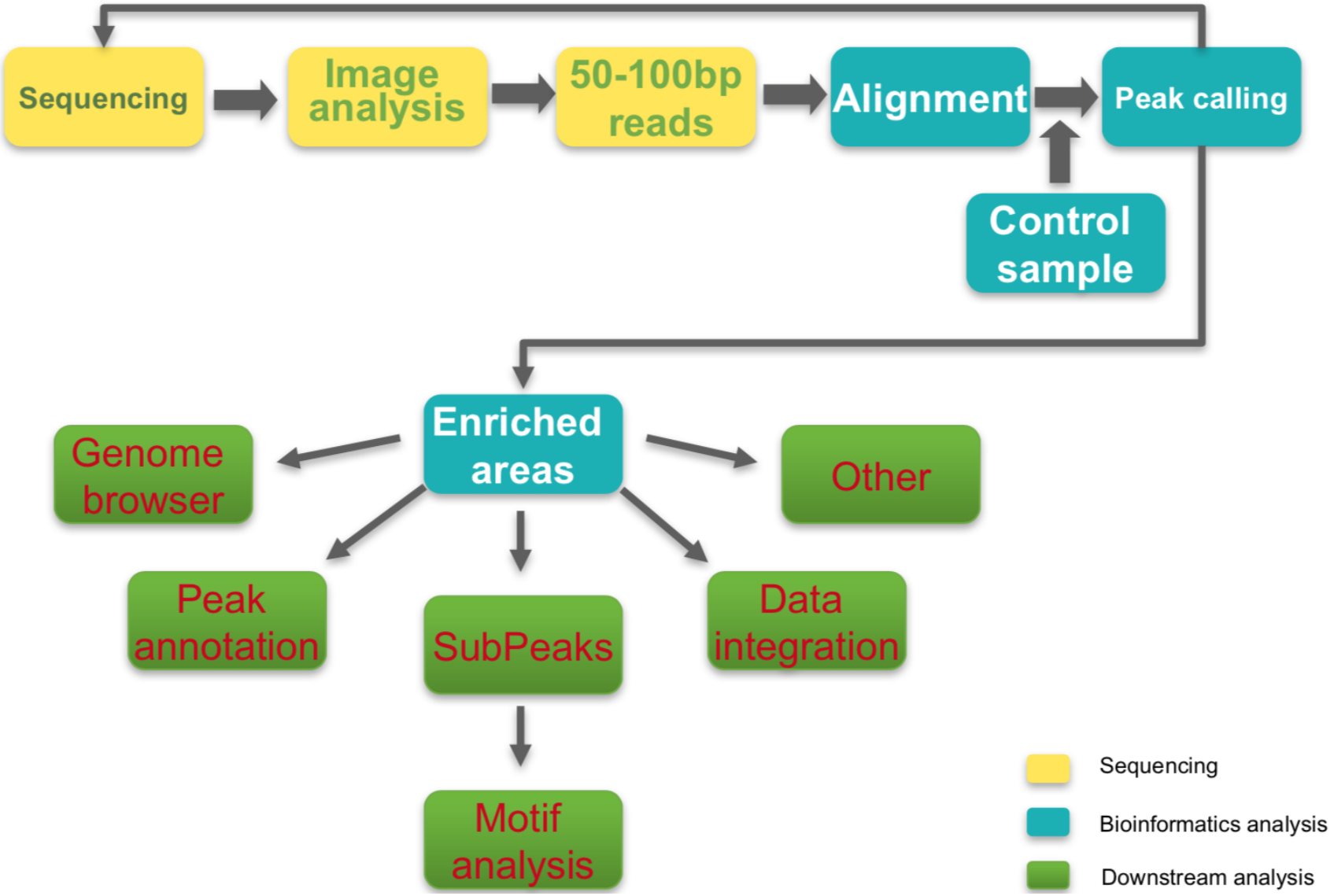
Recap and biological replicates

Sandra Cortijo sandra.cortijo@slcu.cam.ac.uk
Sergio Martinez Cuesta sermarcue@gmail.com
Ashley Sawle Ashley.Sawle@cruk.cam.ac.uk
Denis Seyres Denis.seyres@bioresource.nihr.ac.uk

Based on slides by Myrto Kostadima


Sequencing
Image analysis
50-100bp reads
Alignment
Peak calling
Control sample
Enriched areas
Genome browser
Other
Peak annotation
SubPeaks
Data integration
Motif analysis
Sequencing
Bioinformatics analysis
Downstream analysis

• Quality check	*fastqc*
• Alignment	*bwa, bowtie*
• Filtering	*samtools*
• Peak calling	*macs2*
• Peak annotation	*ChIPseeker*
• Sequence motif discovery	*MEME tools*
• Signal (non-peak) analysis	*deepTools*
• Visualisation	*IGV*

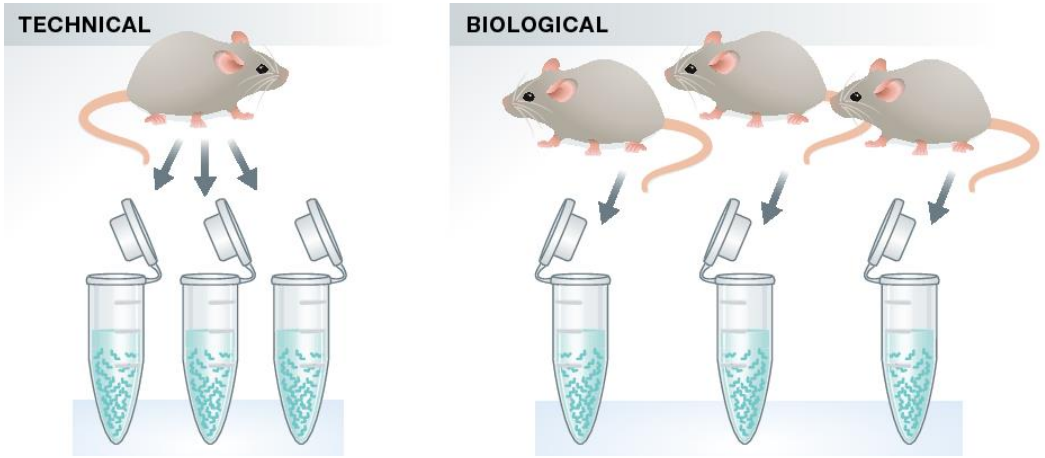
Do we need replicates?

To ensure that ChIP-seq experiments are reproducible and to test their validity we need **at least 2 replicates**

Do we need replicates?

To ensure that ChIP-seq experiments are reproducible and to test their validity we need **at least 2 replicates**

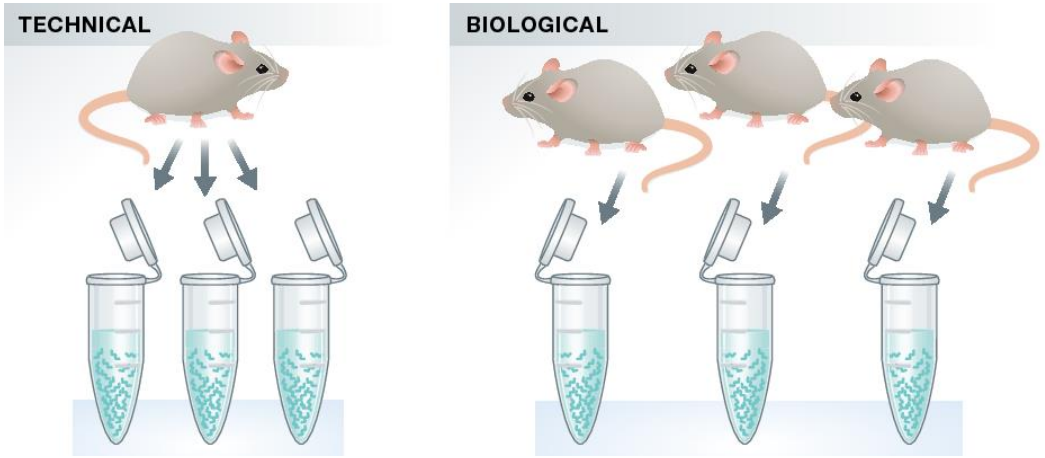
Technical vs. biological replicates



Do we need replicates?

To ensure that ChIP-seq experiments are reproducible and to test their validity we need **at least 2 replicates**

Technical vs. biological replicates



Increase the significance of peaks by using more replicates

Do we need replicates?

In ChIP-seq, what is the overlap between replicate peaks?

Do we need replicates?

In ChIP-seq, what is the overlap between replicate peaks?

To investigate the agreement between replicates: you can also use the method known as **Irreproducible Discovery Rate (IDR) analysis**

Do we need replicates?

In ChIP-seq, what is the overlap between replicate peaks?

To investigate the agreement between replicates: you can also use the method known as **Irreproducible Discovery Rate (IDR) analysis**

- Two replicates are similar if the ranking of significant peaks (by p-value) is consistent between the two replicates in comparison to the ranking of less significant peaks (noise)

Do we need replicates?

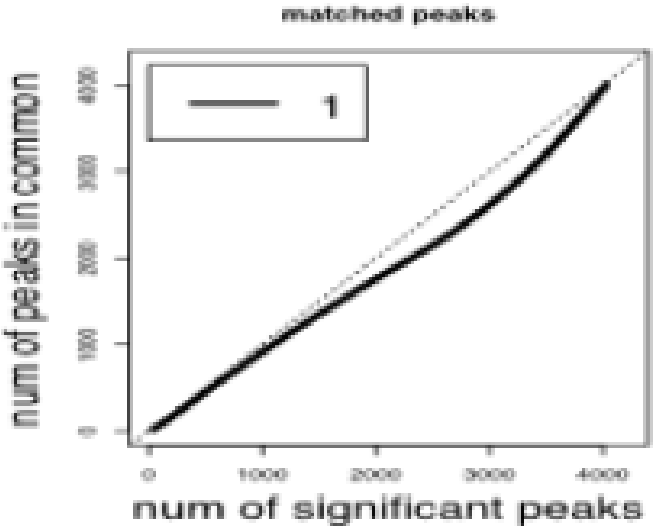
In ChIP-seq, what is the overlap between replicate peaks?

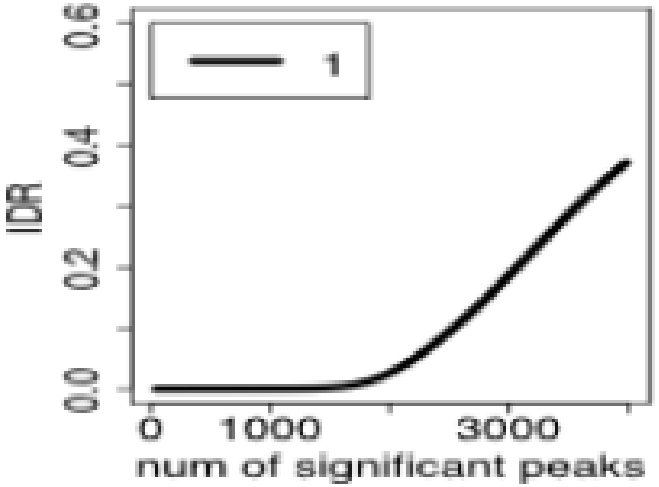
To investigate the agreement between replicates: you can also use the method known as **Irreproducible Discovery Rate (IDR) analysis**

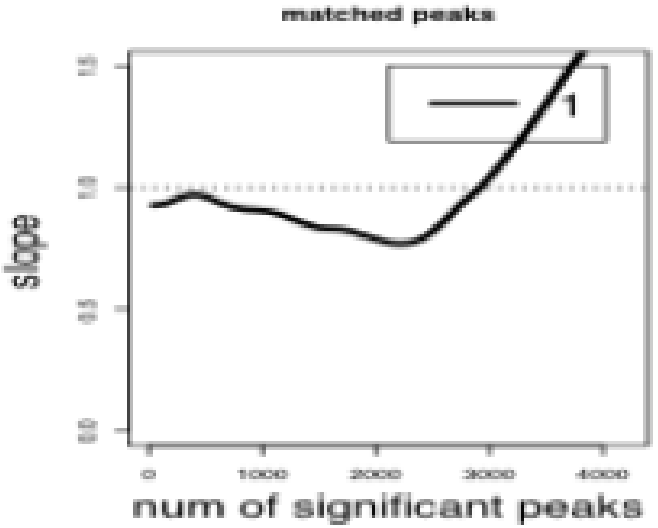
- Two replicates are similar if the ranking of significant peaks (by p-value) is consistent between the two replicates in comparison to the ranking of less significant peaks (noise)

Developed in Qunhua Li and Peter Bickel as part of the ENCODE project

IDR

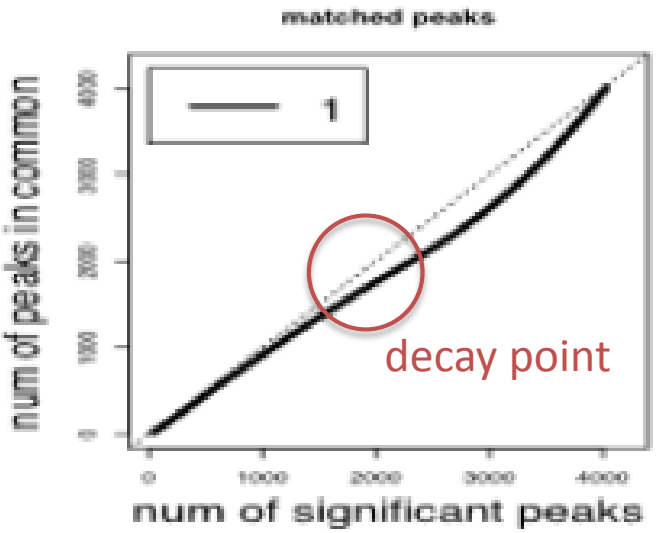


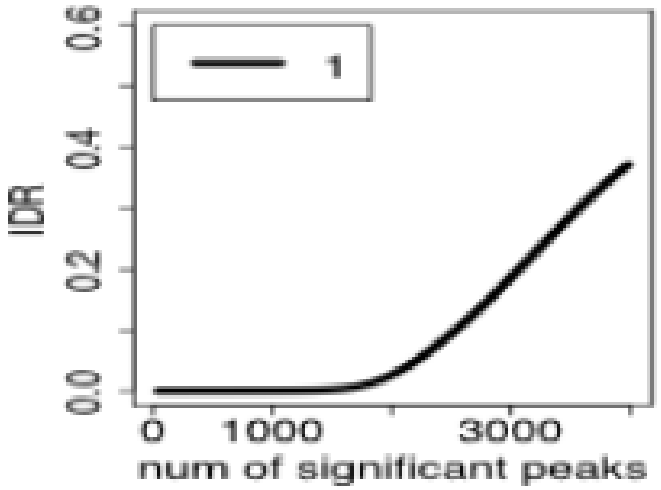


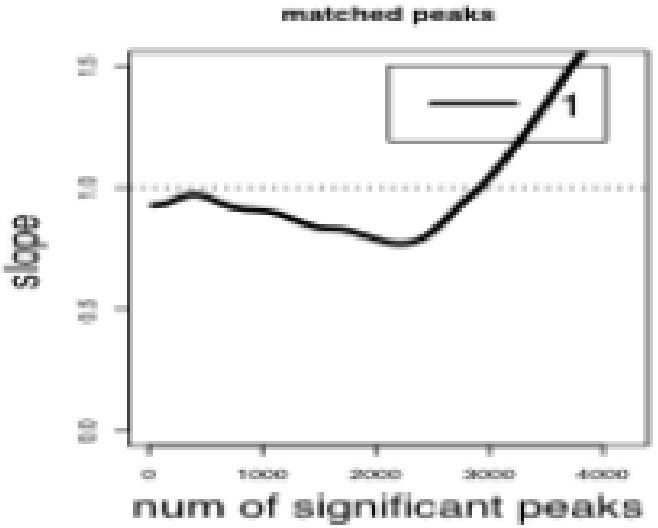


reasonable consistency

IDR

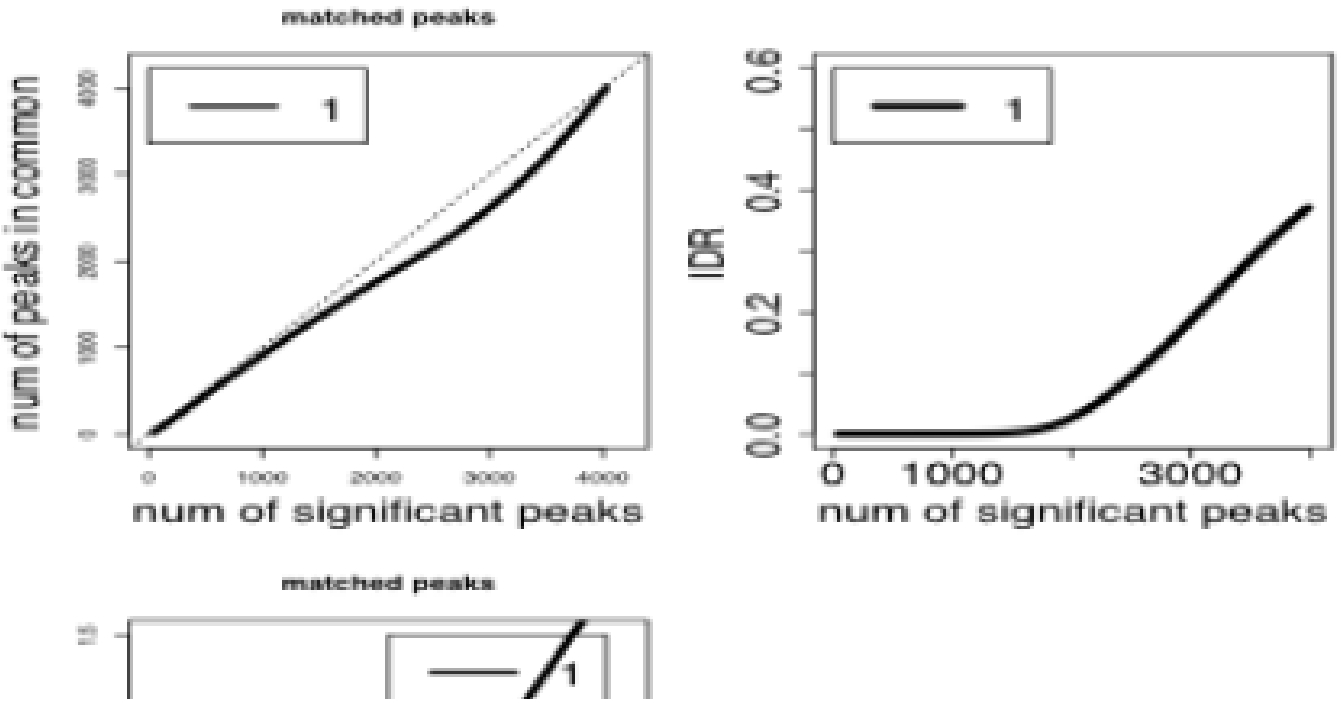






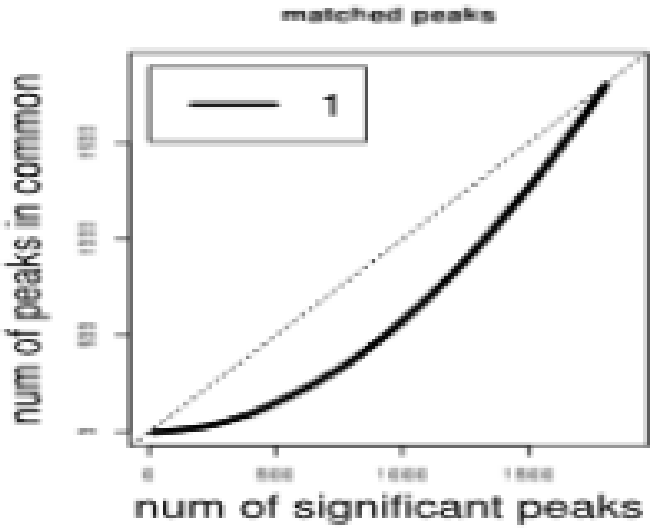
reasonable consistency

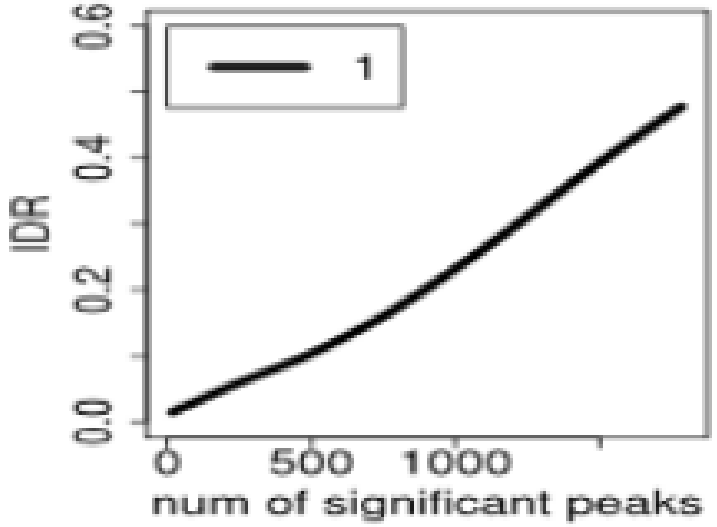
IDR

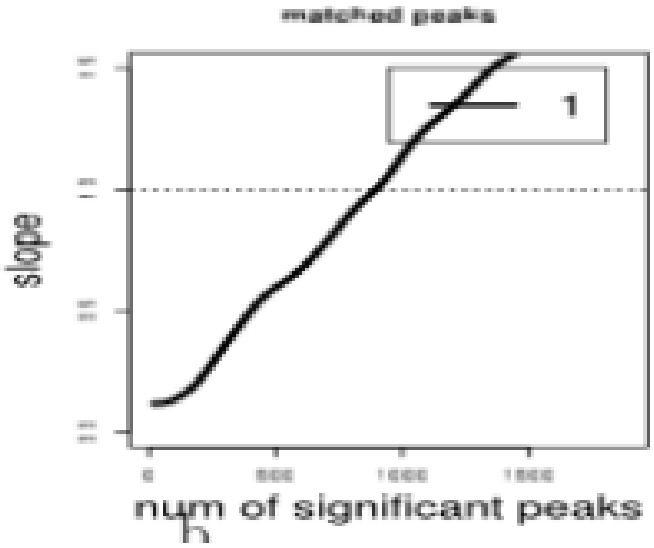


The IDR score represents the *probability of a peak to belong to the irreproducible group*

IDR







poor consistency

IDR

The method compares two lists of ChIP-seq peaks, and statistically assesses the point where the ranking in the list is no longer conserved between the replicates (**decay point**).

IDR

The method compares two lists of ChIP-seq peaks, and statistically assesses the point where the ranking in the list is no longer conserved between the replicates (**decay point**).

Li and Bickel have used this principle to define numerical cutoffs on peaks called after merging the replicates (**IDR thresholds**) where an IDR of 0.05 means that there is a 5% chance that a called peak is irreproducible.

Li et al., The Annals of Applied Statistics (2011)