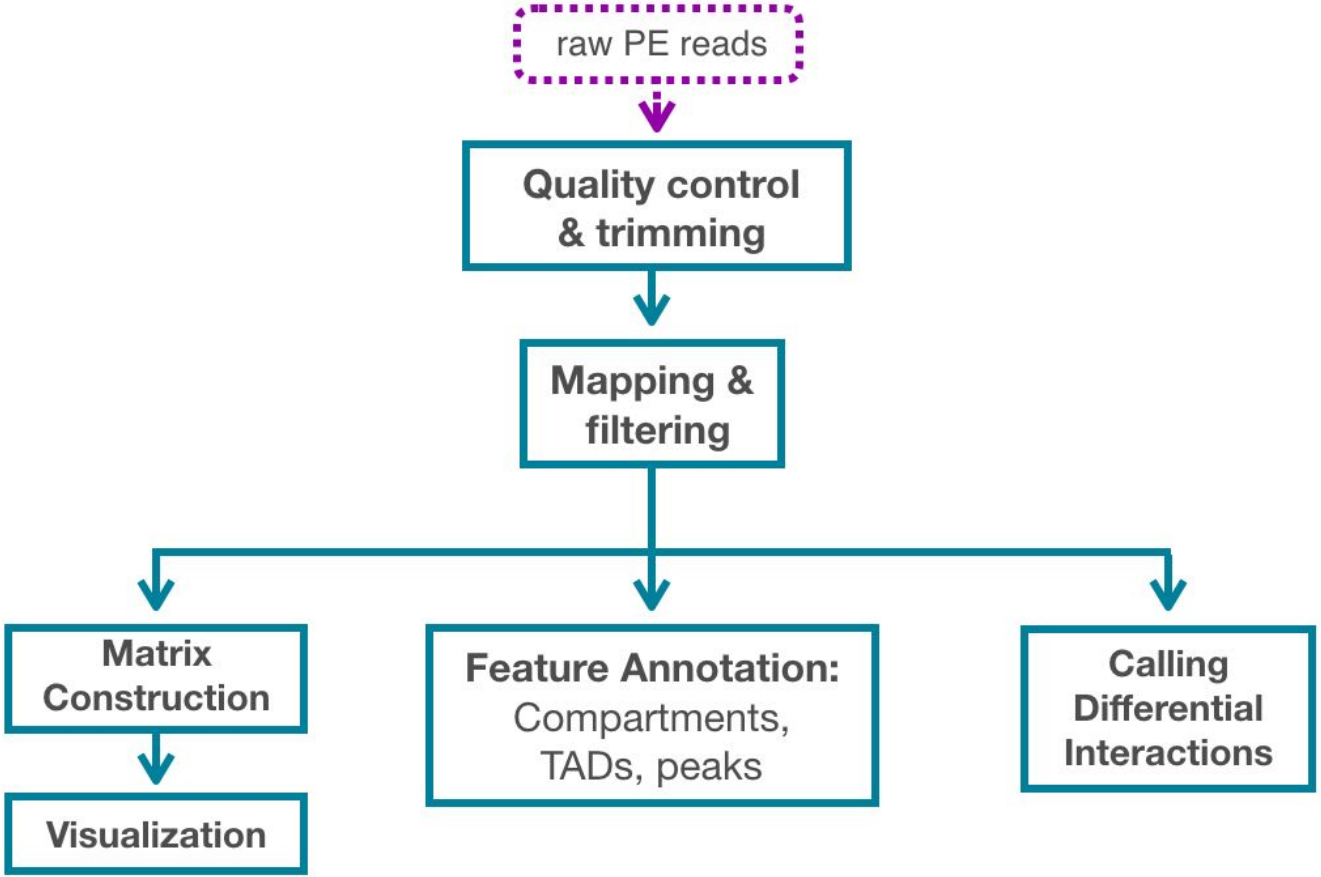
Detecting Differential genomic Interactions

Recap


raw PE reads
Quality control & trimming
Mapping & filtering
Matrix Construction
Visualization
Feature Annotation: Compartments, TADs, peaks
Calling Differential Interactions

Learning objectives

- Differential interactions on Hi-C data
- Between-sample normalization
- Modelling biological variability
- Testing for significant differential interactions

Complications with the identification of **biologically interesting interactions**

- False signals arise from technical causes
- Background signal depends on various biases (GC content, mappability, fragment length)
- Conserved features dominate the Interaction space

Identification of differential interactions (DI)
a like-for-like comparison

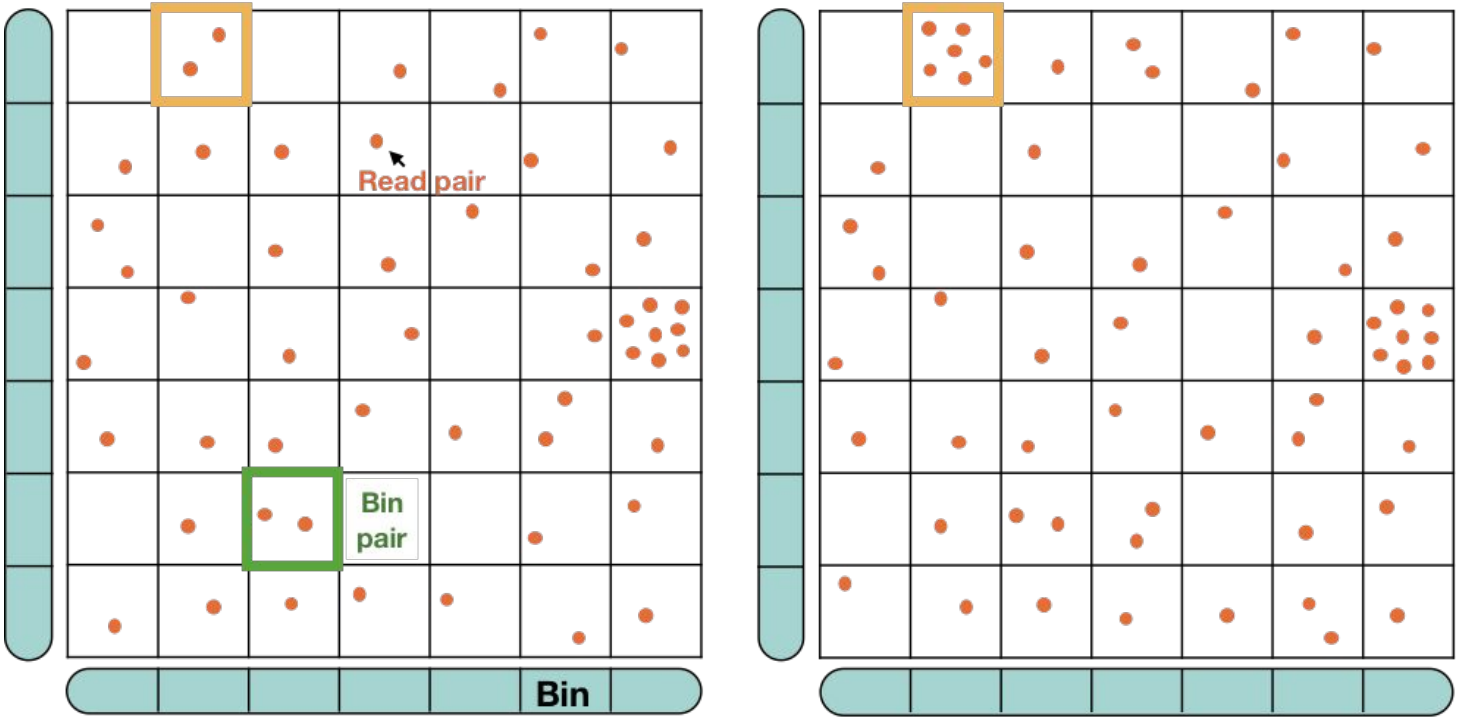
- For every interaction between two regions, the intensity is compared between samples and tested to obtain significant DI across biological conditions.

- Biases are constant between conditions

- These DI relate to the biological conditions studied

Identification of differential interactions (DI)

a like-for-like comparison



Based on Lun & Smyth (2015)

Software for differential analysis:

HOMER (Brenner *et al.*, 2014)

chromoR (Shavit *et al.*, 2010)

HiCcompare (Stansfield *et al.*, 2017)

FIND (Djekidel *et al.*, 2018)

diffHic (Lun & Smyth, 2015)

Calling differential interactions with diffHic

Read alignment and processing

Binning

Bin pairs filtering

Normalization

Estimating biological variability

Testing for significant differential interactions

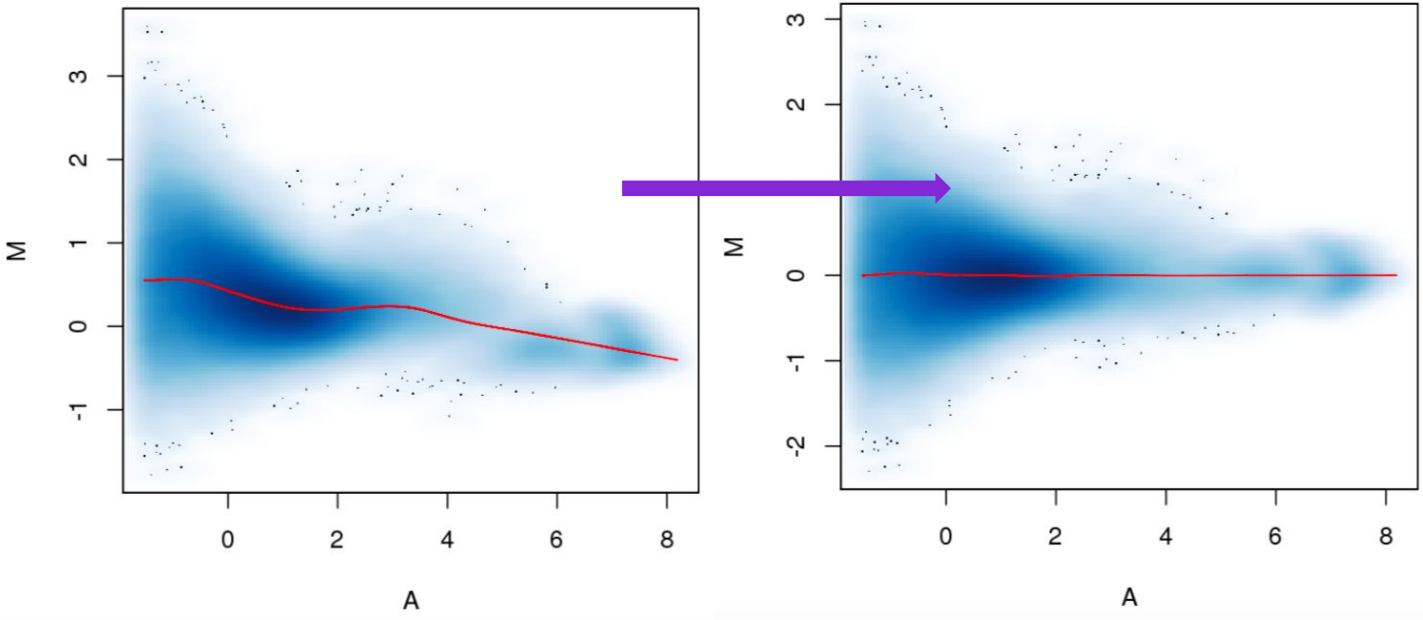
- Binning and filtering

Each pair of bins represents an interaction
Bin size determines the resolution:
larger bins = More robust
smaller bins = Increases spatial resolution
*The boundary bin is rounded to the nearest restriction site

Filtering bin pairs

- Average abundance
- Median abundance across inter-chromosomal bin pairs.
- By distance
- Peak calling

- Between-samples Non-linear normalization with Local weighted regression (loess) method



With NLN, a **matrix of offsets** is generated, containing the log-transformed scaling factors necessary for normalizing each entry of the count matrix

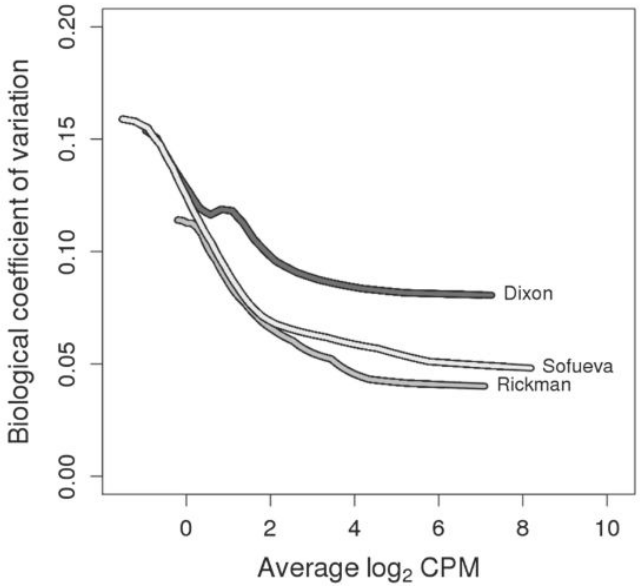
*Other methods can be implemented in diffHic

Estimating biological variation

Counts are modeled under a **negative binomial (NB) distribution**

Variability of the bin-pairs counts between replicates is modeled with the **quasi-likelihood (QL) method**

Sharing information across bins accounts for limited replicates (Bayes)



- Testing for significant interactions

QL F-test for each bin pair

Identify significant differences between samples

Clustering to reduce redundancy

Based on significant bin pairs
Controlling FDR