Feature annotation
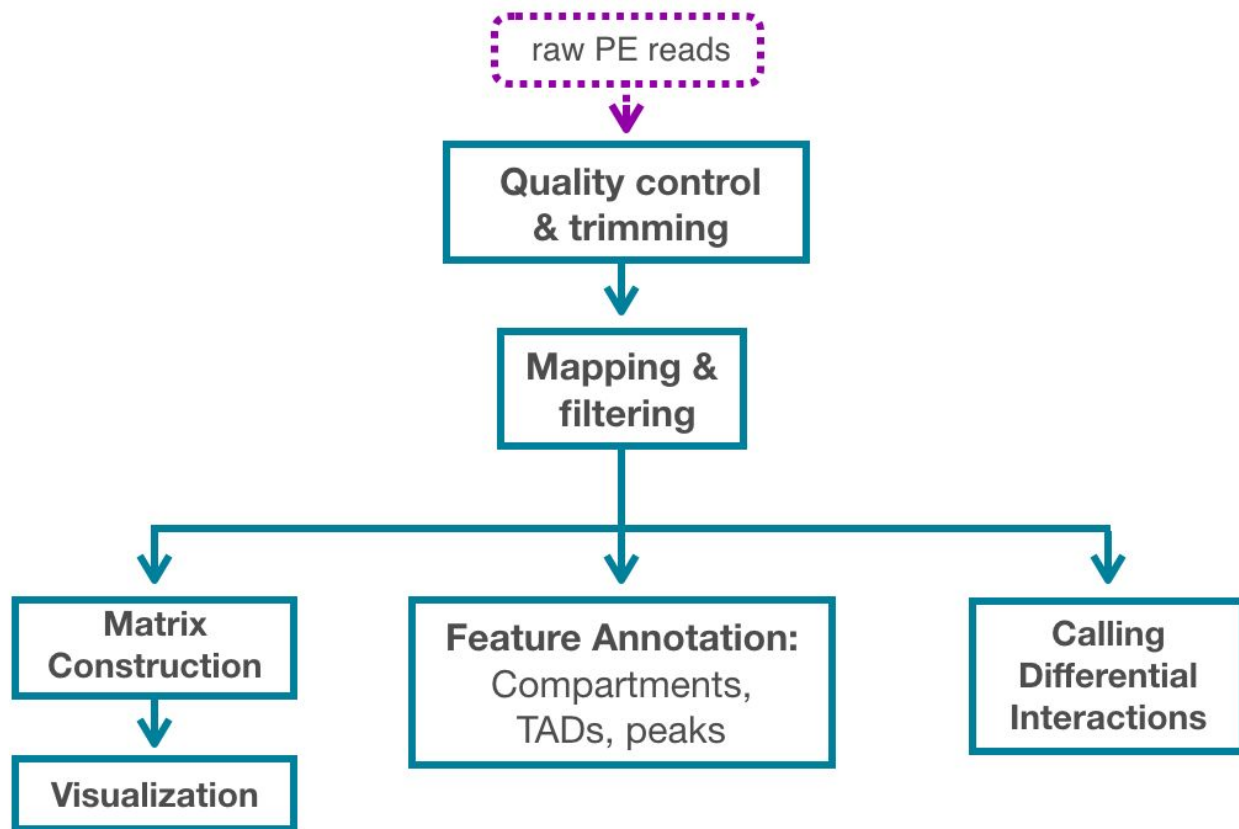
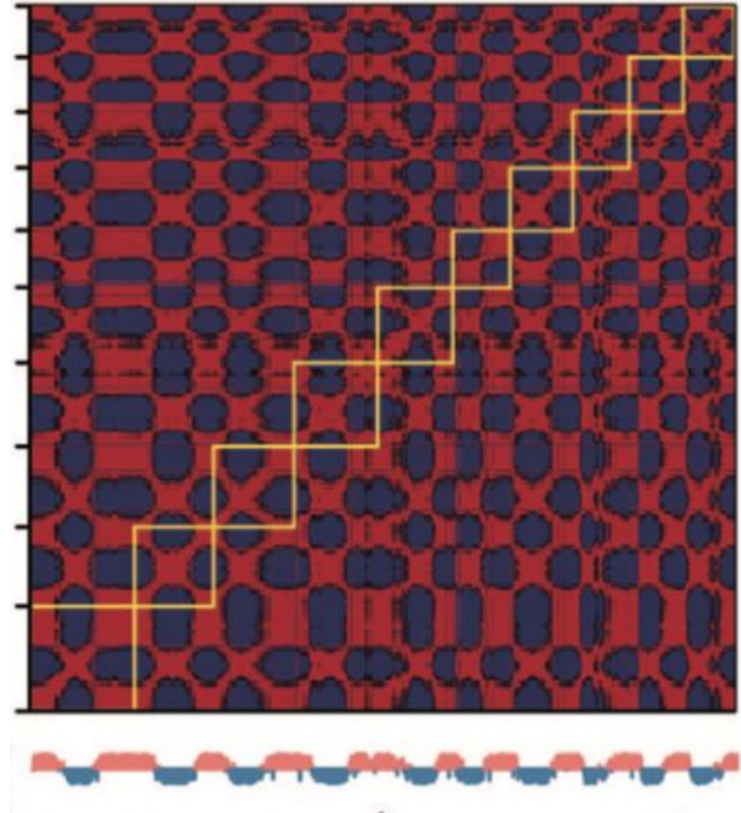Compartments, TADs and peaks

Compartments

Recap

Learning objectives

- Identify compartments in HiC matrix
- Compartments in plants

Compartments

- A compartment:
    - active regions (euchromatin)
- B compartment:
    - inactive regions (heterochromatin)
- Bin sizes 100kb - 1MB

- PC1

PCA

- Principal Component Analysis
- The first eigenvector will give the compartmentalization profile
- Positive values indicate one compartment, negative values indicate other compartment
- How to tell if A or B?

Correlation with other data

- Show first component together with other epigenomic tracks:
- Gene content / TE density
- Histone modifications:
  - A: H3K27Ac, H3K4me3
  - B: H3K27me3, H3K9me3
- Transcription
  - RNA seq tracks
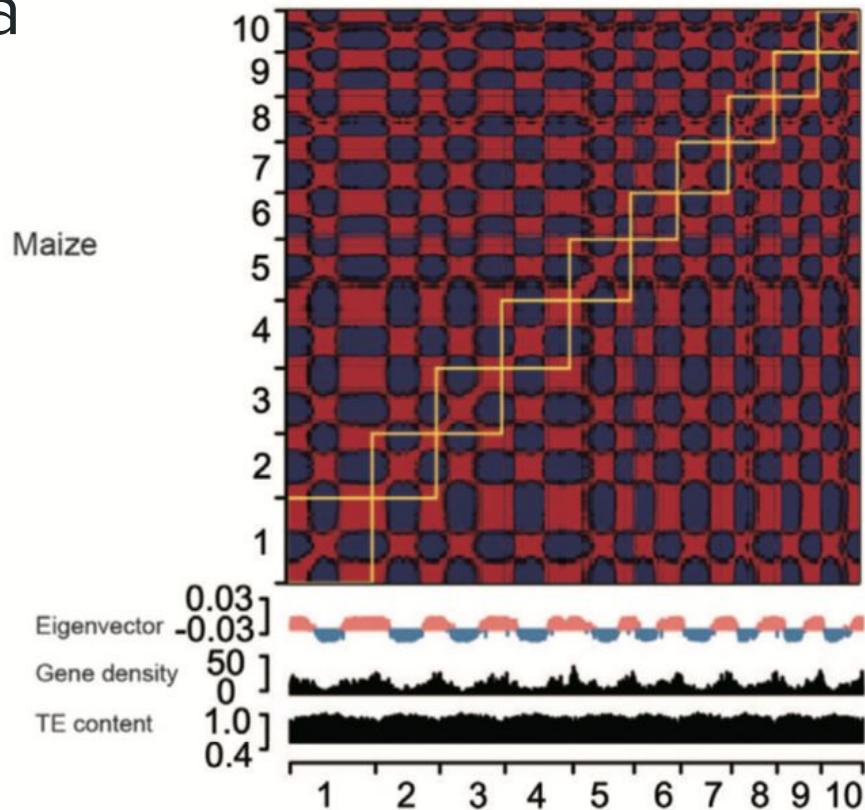
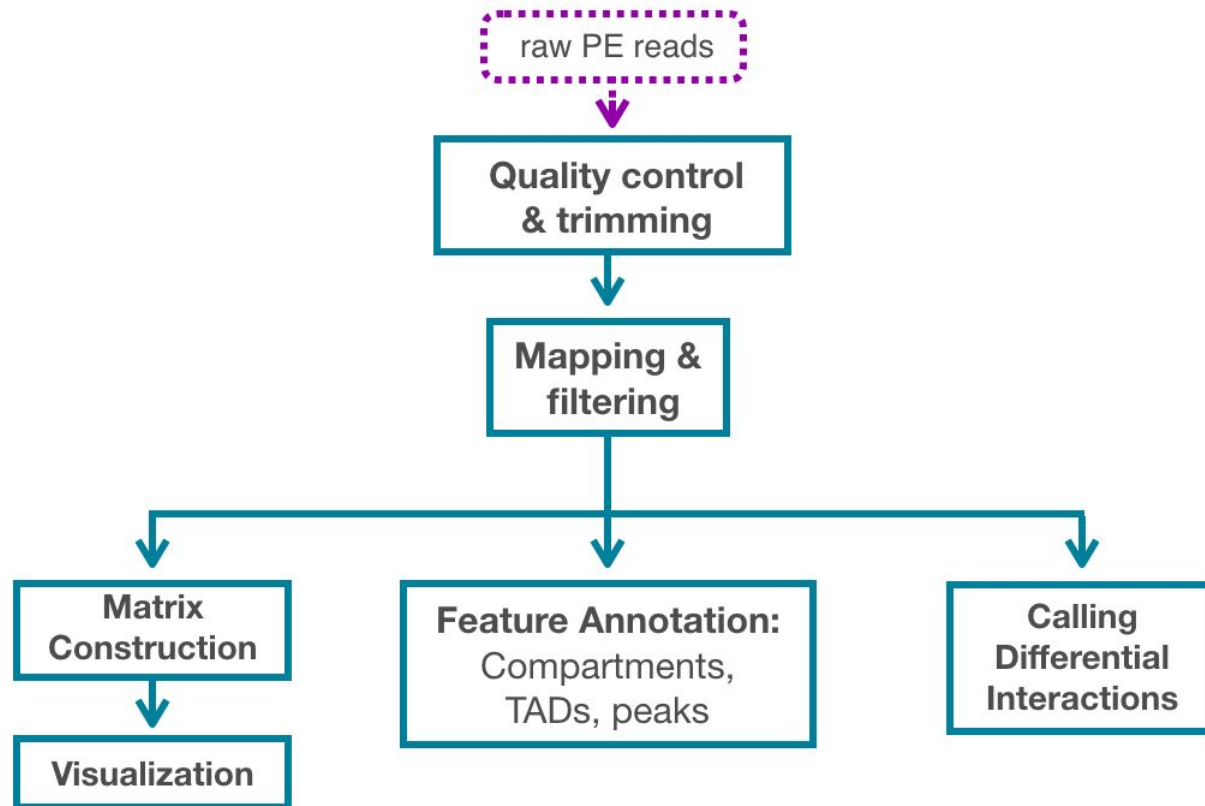# Practical

- Identify compartments with HiCexplorer
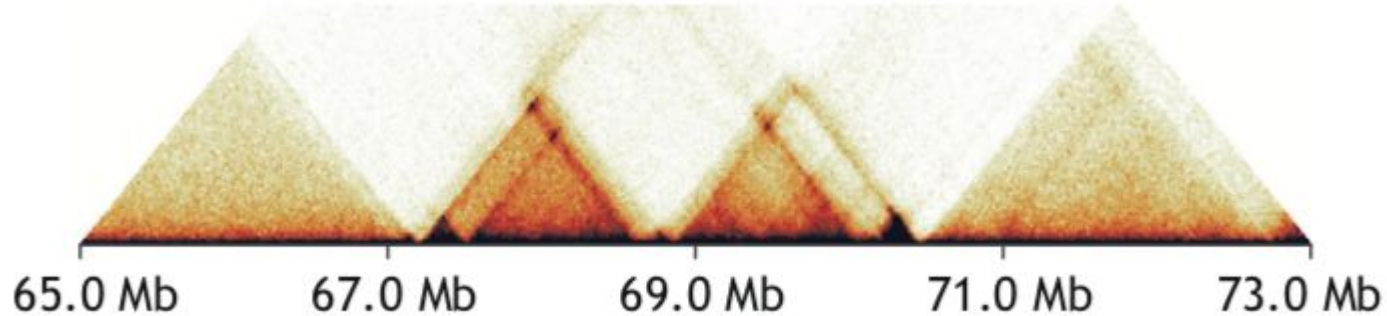
TADs

Recap

Learning objectives

- Identify TADs with HiCExplorer
- Discuss challenges in TAD calling

TADs

- TADs are regions with elevated self interaction frequencies
- TADs might act as an insulated genomic region that constrains regulatory interactions

TAD calling in HiCexplorer

1. Transform matrix to Z score (subtract mean contact frequency at each distance): make bins more comparable
2. For each bin, calculate average contacts between w upstream and downstrem bins.
3. Repeat for different values of w, and then take the average.
4. Local minima should be boundaries!
5. To double check: compare the distributions of upstream and downstream "diamonds".
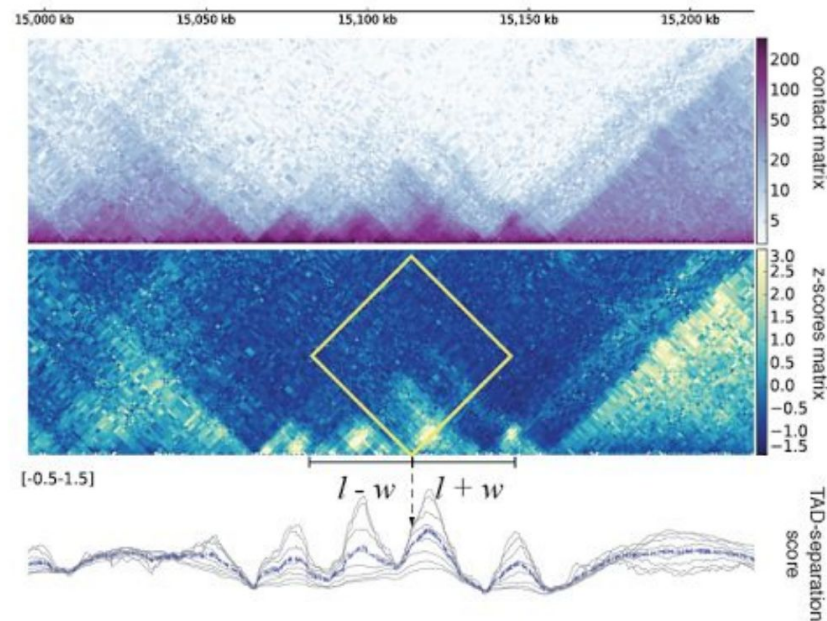
TAD calling challenges

- What is a TAD?
- TAD-like patterns are often hierarchical and overlapping (subTADs, gene mini domains)
- There are aprox. 22 different TAD calling algorithms
  - Current approaches often not reproducible
  - Sensible to normalization, bin size, sequencing depth

RESEARCH　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

Comparison of computational methods for the identification of topologically associating domains

Marie Zufferey[1,2†], Daniele Tavernari[1,2†], Elisa Oricchio[3] and Giovanni Ciriello[1,2*]

TAD calling algorithms

- Linear score
    - Directionality Index
    - Insulation score
        - TopDom
        - **HiCExplorer**
- Statistical models
    - TADbit
- Clustering
    - ClusterTAD
- Network analysis
    - spectral
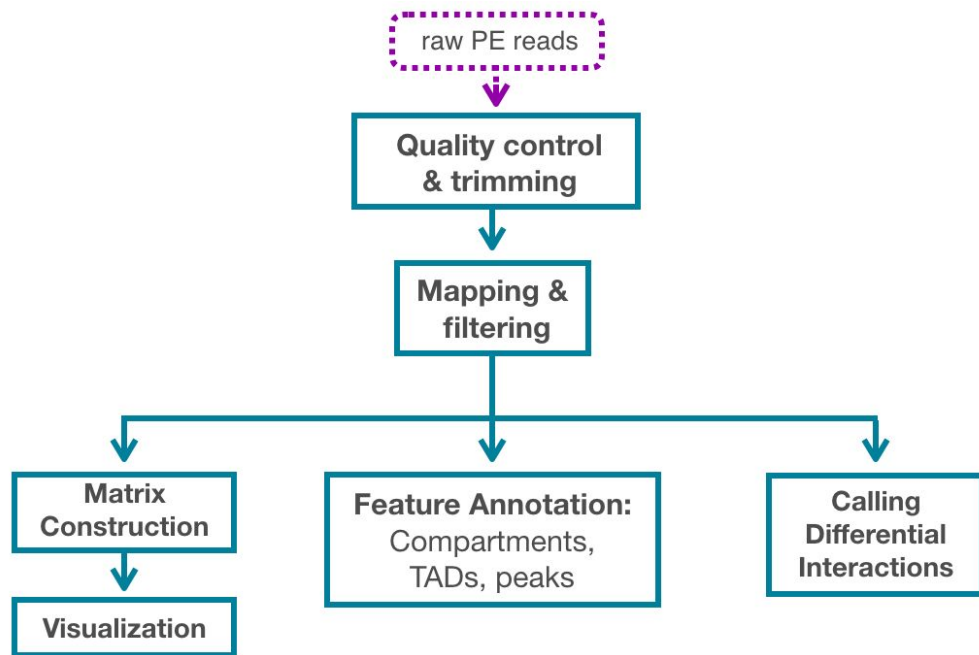
Practical:

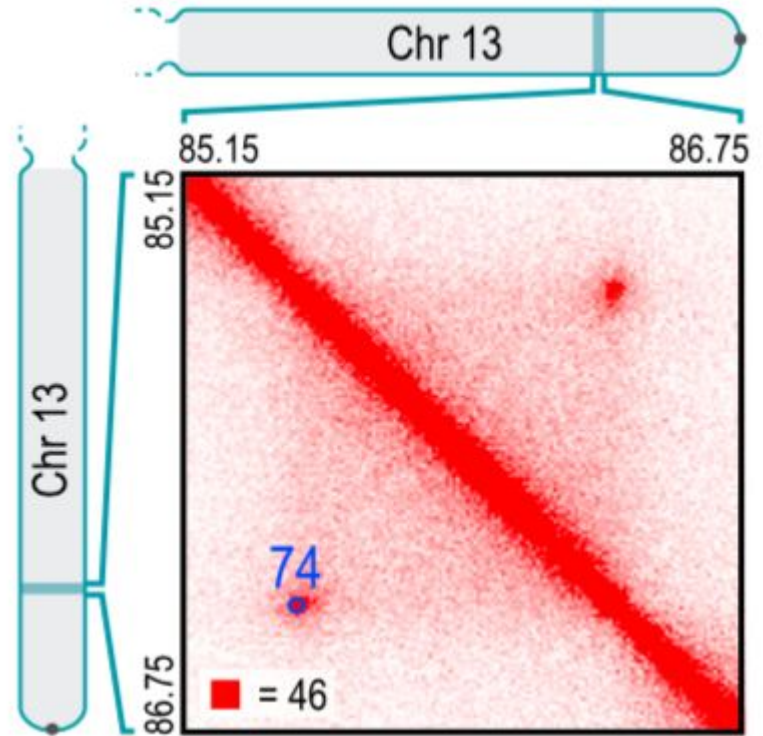- TAD calling with HiCexplorer

Interaction peaks

Recap

Learning objectives

- Identify peaks with HiCCUPS

Peaks and biological features

- Enhancer-promoter interactions
- CTCF-CTCF and cohesin binding sites
- Polycomb bodies
- KNOT region (Arabidopsis)
- Gene - Gene interactions (Maize)

HICCUPS

- Local enrichment over four backgrounds:
  - Donut
  - Horizontal
  - Vertical
  - Lower left corner
- Compare bin contact frequency with average contact frequency of each background
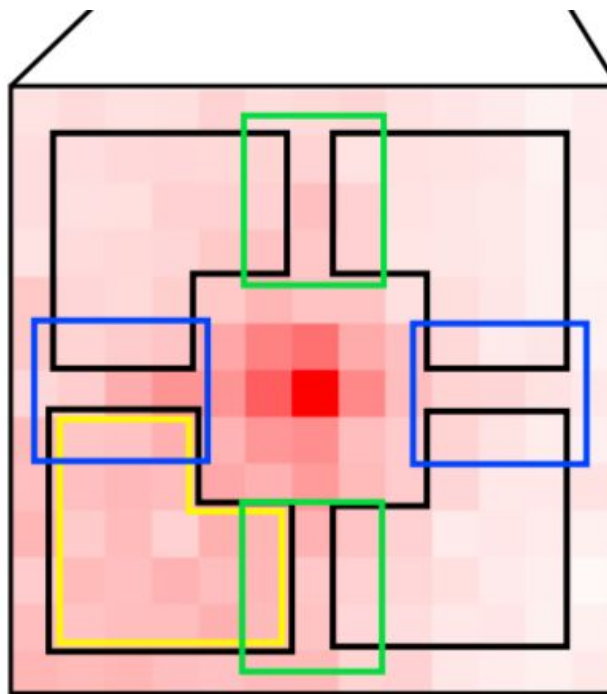
Exp(☐) = 83.3
Exp(☐) = 67.4
Exp(☐) = 78.9
Exp(☐) = 88.4
Peak   = 341

Aggregate Peak Analysis

- Analyse the average interaction profile for all peaks
- Can be used as QC
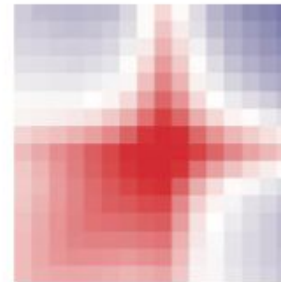
Loops

Peak calling algorithms

- Global enrichment
  - Fit-HiC
  - HOMER
  - HICCUPS
- Local enrichment
  - **HICCUPS**
  - **HiCExplorer**

Practical:

- Peak calling with HiCexplorer

Resources:

- HiCExplorer: https://hicexplorer.readthedocs.io/
- HiGlass: http://higlass.io/
- Deeptools: https://deeptools.readthedocs.io/en/develop/
- Juicer: https://github.com/aidenlab/juicer/wiki
- Collection of hic tools: https://github.com/mdozmorov/HiC_tools