

ChIP-Seq Data Analysis: Beyond binding sites

Suraj Menon

Bioinformatics Core

CRUK Cambridge Institute

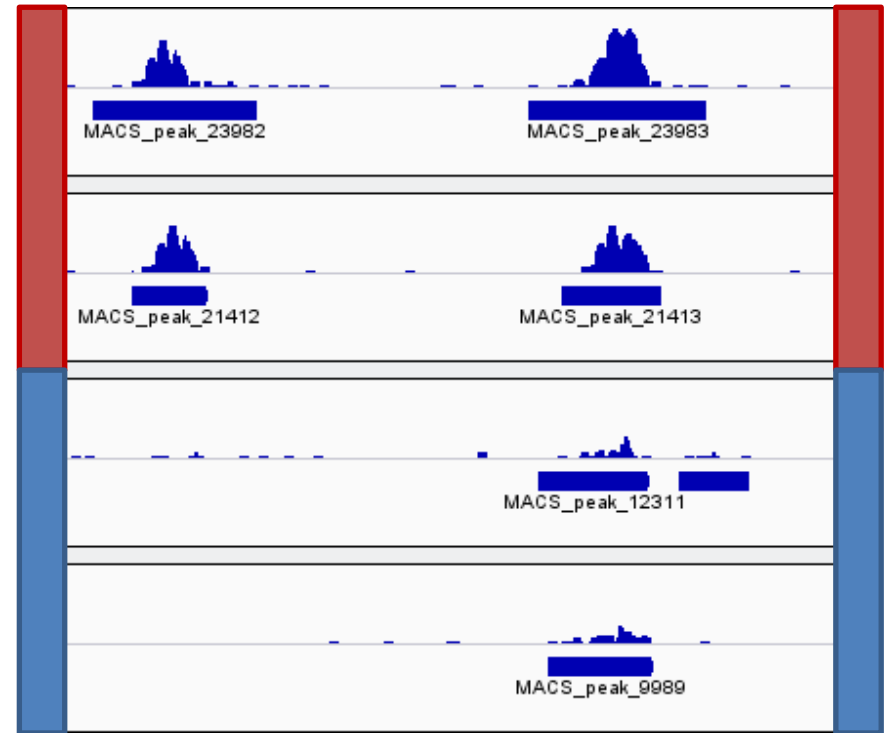
Acknowledgements: Rory Stark, Thomas Carroll

Analysis of Gene Regulation

- Majority of **functional** studies focus on transcript levels
- ChIP-Seq typically used for **structural** studies like mapping TF binding site (ENCODE!)
- Possible to study the **dynamics** of gene regulation using ChIP-Seq

Analysis of Gene Regulation

- Simple case: investigate changes in TF binding in two different conditions
- **Occupancy analysis**
 - binary outcome
 - qualitative
- **Affinity analysis**
 - quantitative
 - **differential binding** can have functional consequences



Differential oestrogen receptor binding is associated with clinical outcome in breast cancer

Caryn S. Ross-Innes, Rory Stark, Andrew E. Teschendorff, Kelly A. Holmes, H. Raza Ali, Mark J. Dunning, Gordon D. Brown, Ondrej Gojis, Ian O. Ellis, Andrew R. Green, Simak Ali, Suet-Feung Chin, Carlo Palmieri, Carlos Caldas & Jason S. Carroll

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

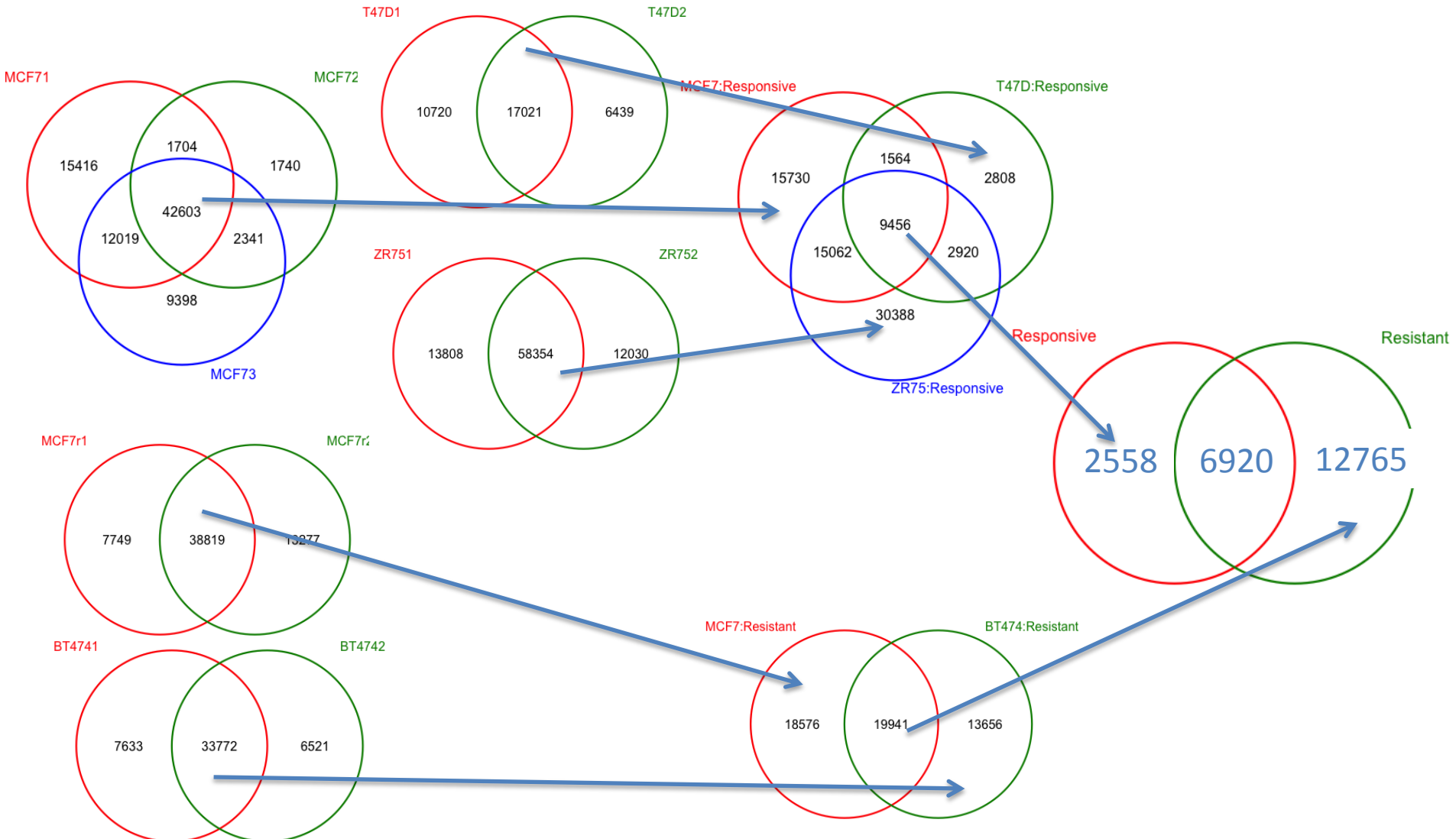
Nature **481**, 389–393 (19 January 2012) | doi:10.1038/nature10730

Received 19 May 2011 | Accepted 23 November 2011 | Published online 04 January 2012

The experiment

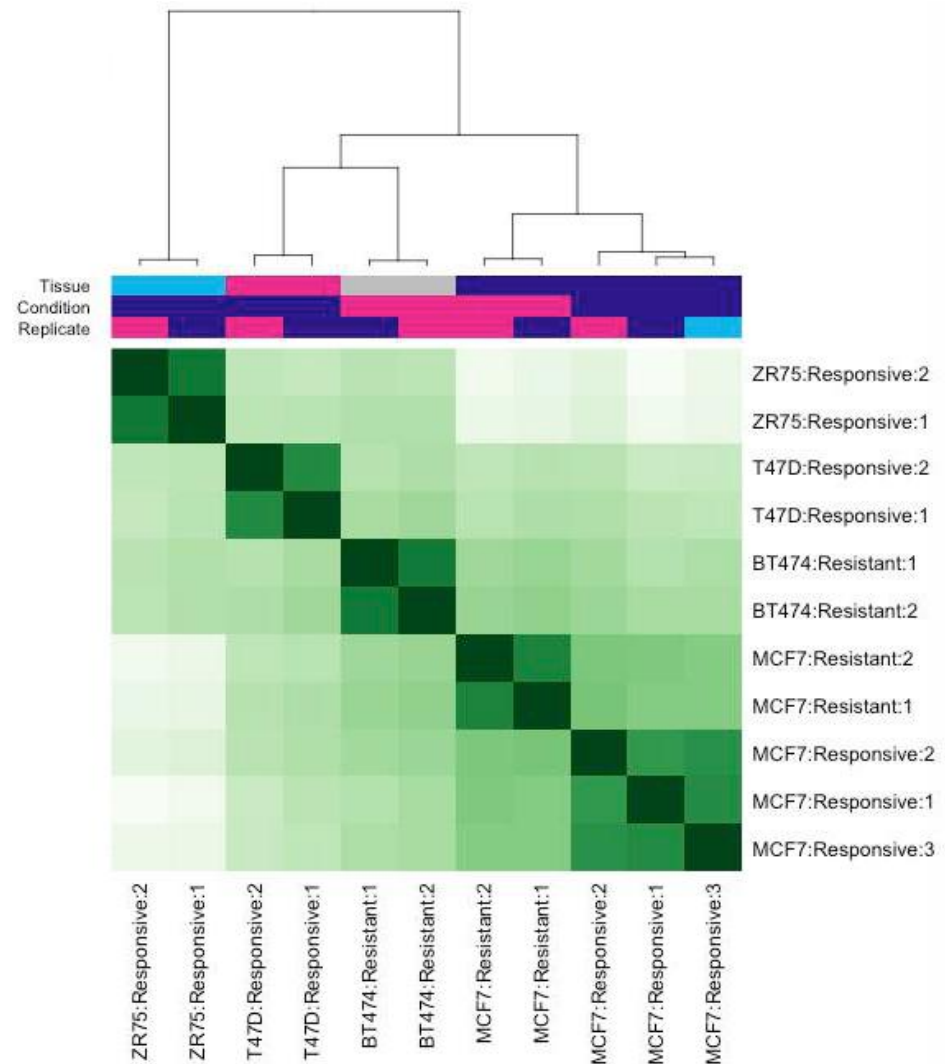
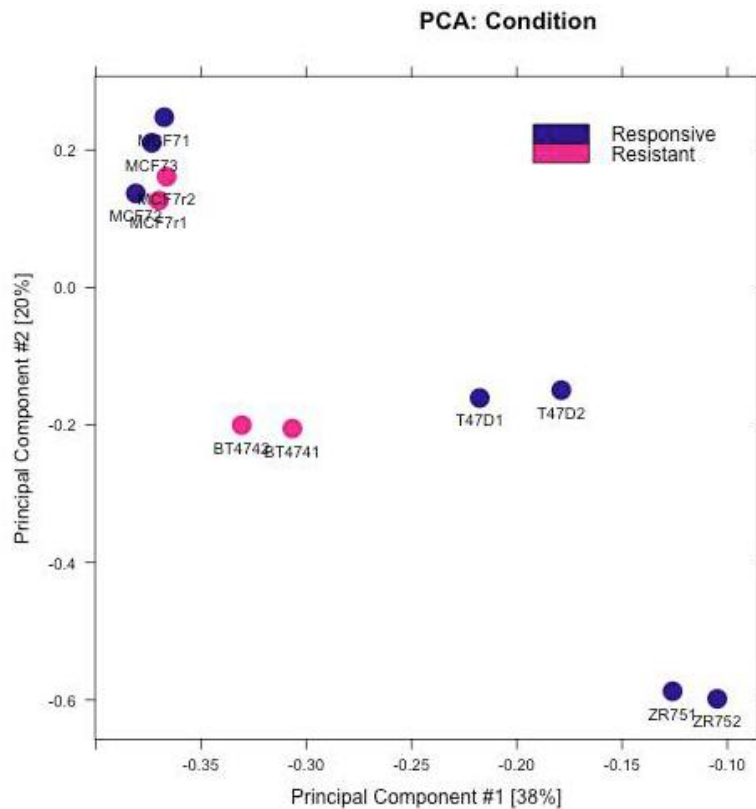
- Oestrogen receptor ChIP-Seq
- 4 drug-responsive breast cancer cell lines.
 - 2-3 replicates each
- 2 drug-resistant breast cancer cell lines
 - 2 replicates each
- “High-confidence” peaksets: called by at least two peak callers (MACS and Swembl)
- **What regions are differentially bound between the drug-resistant and drug-responsive cell lines?**

Occupancy Analysis



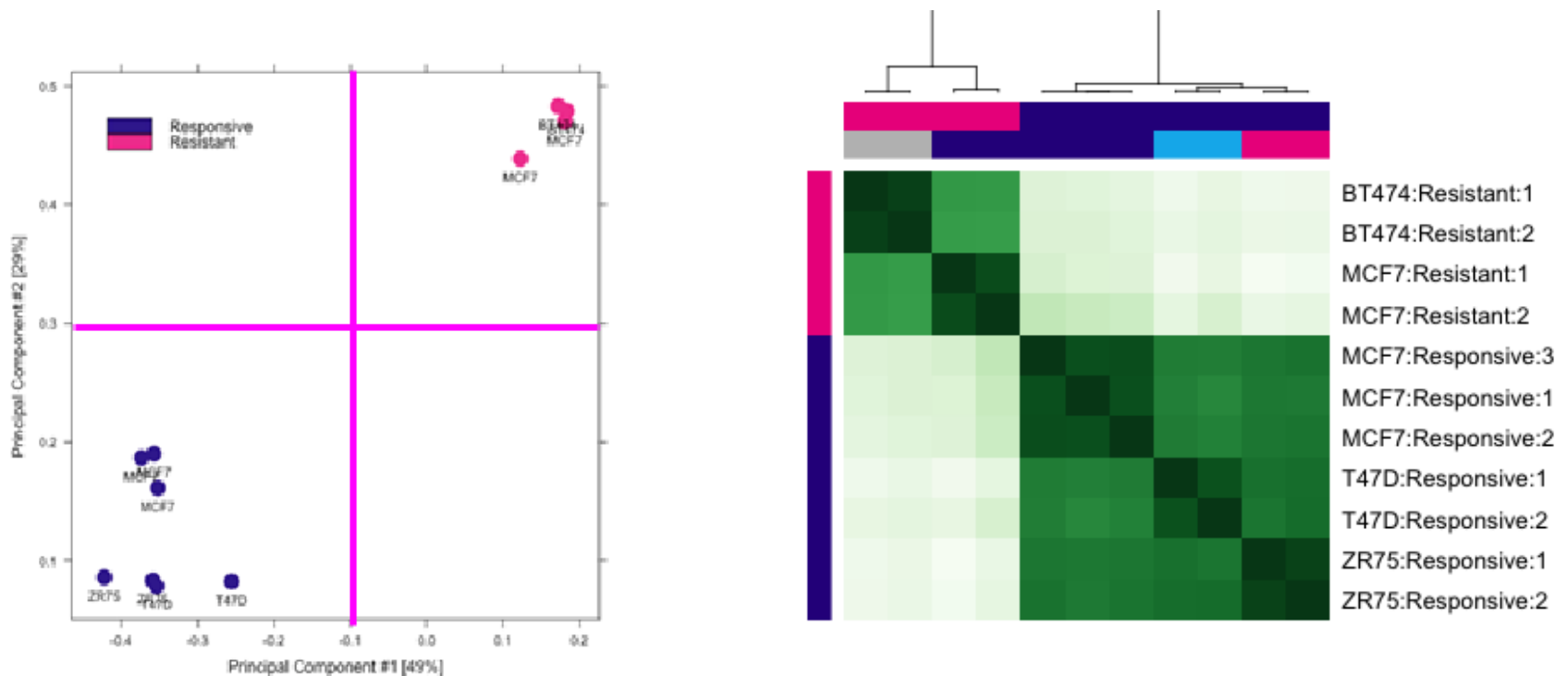
Occupancy Analysis

No good separation of Responsive and Resistant samples



Affinity Analysis

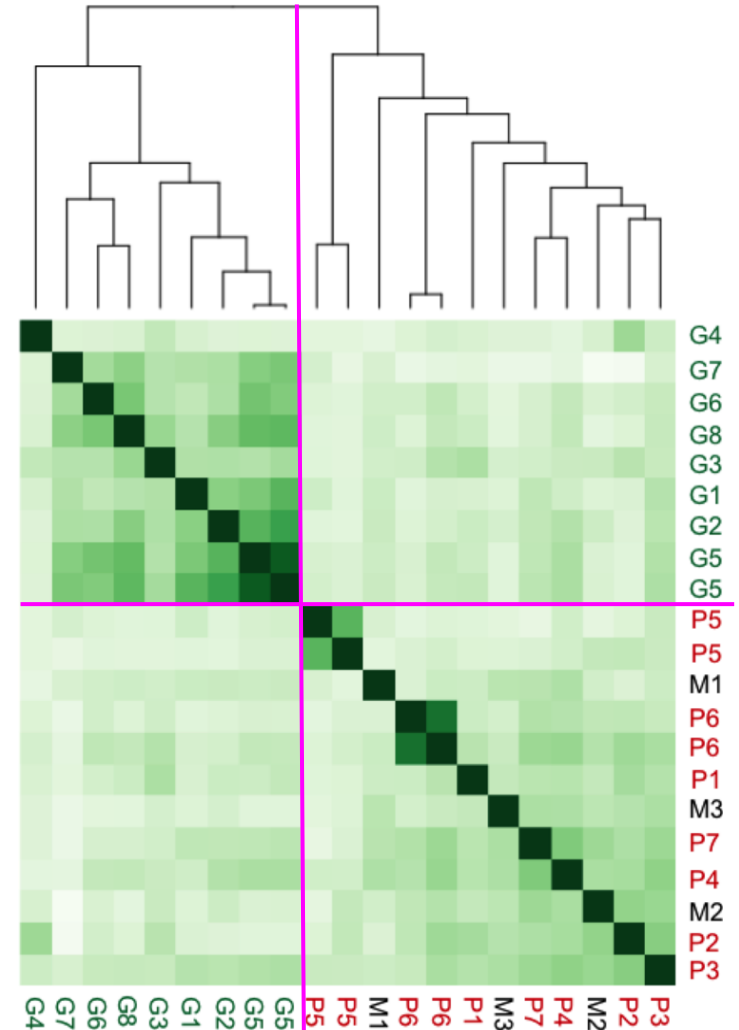
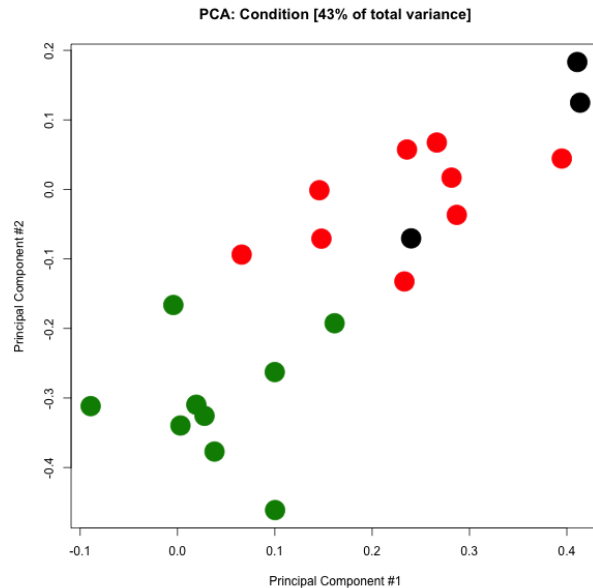
Quantitative data for differentially bound site cleanly separates groups



Clean separation even seen from DB testing of sites common to all samples

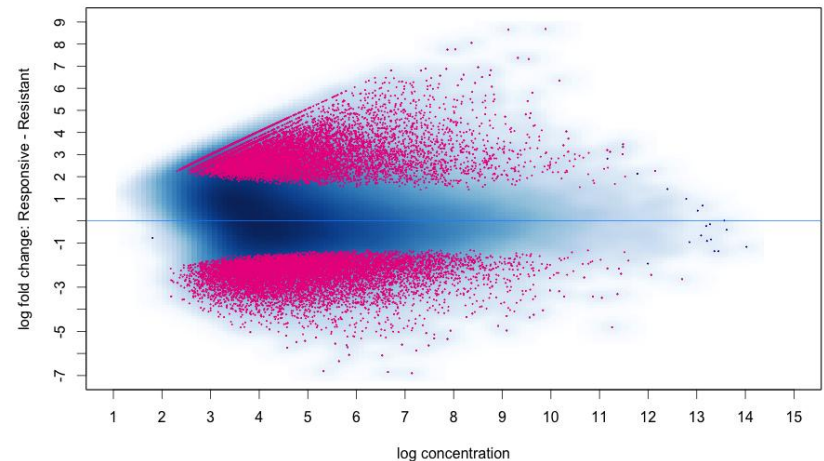
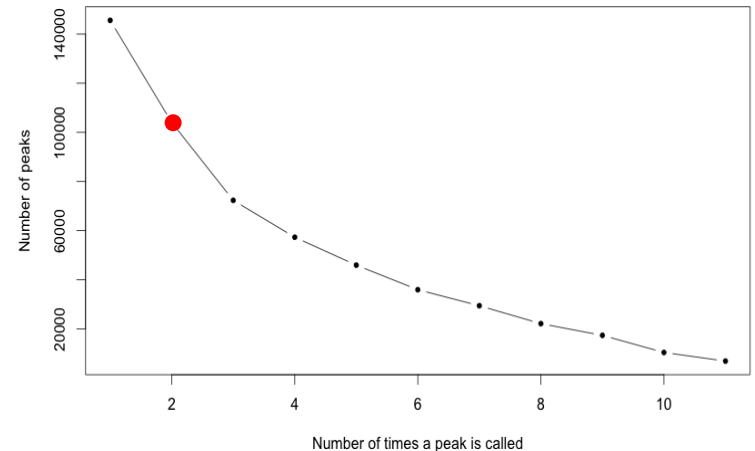
Affinity Analysis

ER differentially bound sites allow for separation of good and bad prognosis patients!



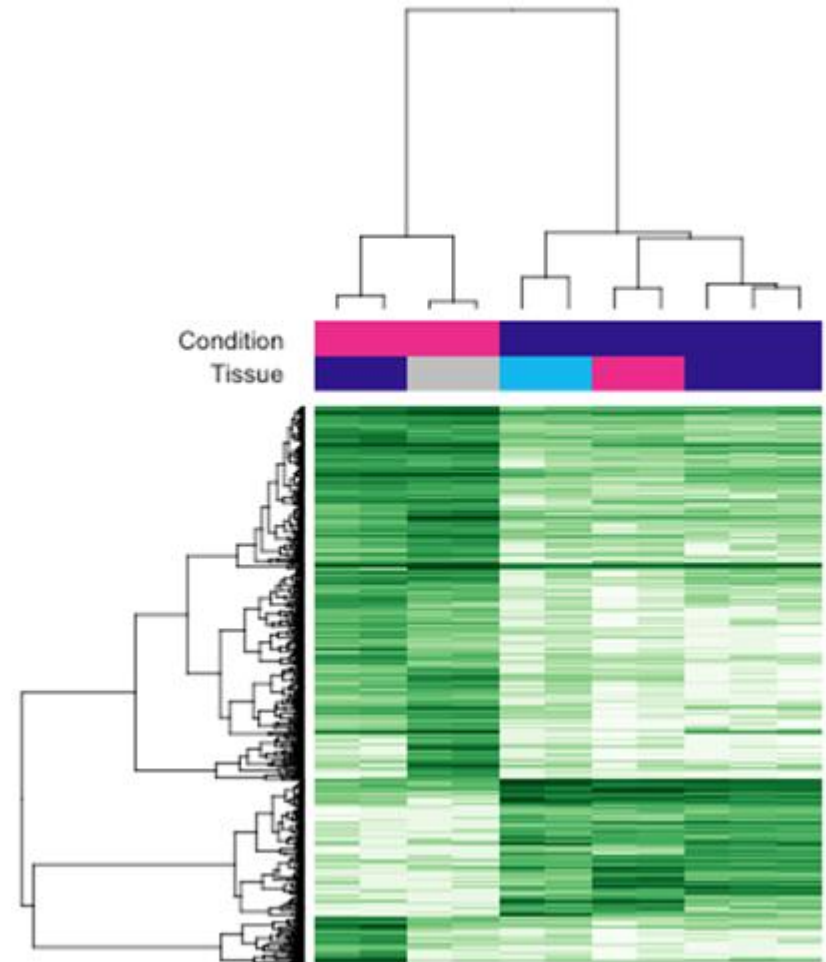
Differential Binding Workflow

- Aligned reads including duplicates
- Select regions of interest
 - select from consensus peakset (e.g. peaks in 2 of 8 samples)
 - promoters
 - pre-defined genomic windows
- Count reads
- Normalise
 - RNA-Seq strategies
 - Careful when assumptions don't hold true e.g. when global ChIP signal changes



Differential Binding Workflow

- Set up contrasts for comparisons of interest
- Statistical assessment of differential binding
 - set up appropriate statistical model
 - edgeR, DESeq2
- Visualise data and results
 - MA plots
 - Heatmaps
 - Clustering and PCA



Practical: DiffBind package vignette

- Bioconductor package
- Full integrated workflow for differential binding analysis
 - start from aligned BAMs and peak files
 - count, normalise, analyse binding, visualise
- Includes occupancy analysis
- Includes useful functions for peakset manipulation
 - building consensus peaksets
 - Venn diagrams