UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Differential Binding Analysis

Beyond peak calling

Suraj Menon

Bioinformatics Core

CRUK Cambridge Institute

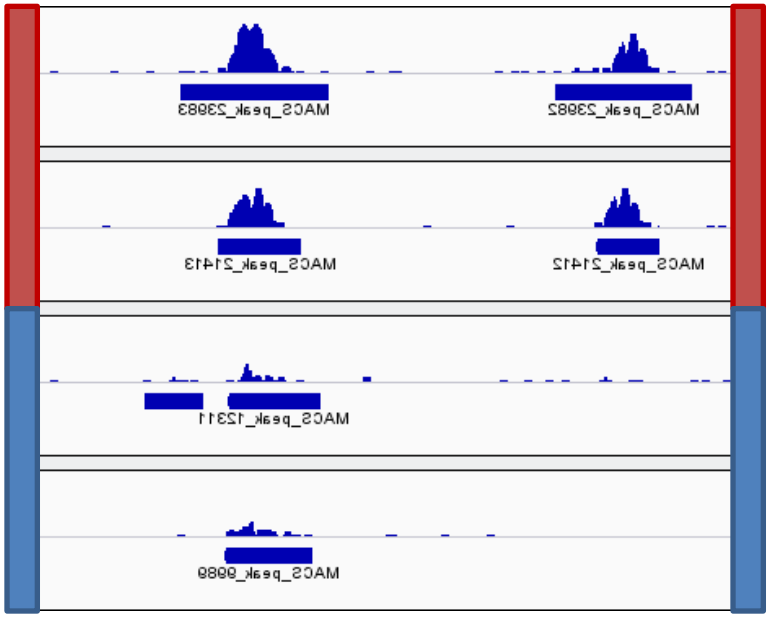
Acknowledgements: Rory Stark

Analysis of Gene Regulation

- Majority of **functional** studies focus on transcript levels
- ChIP-Seq typically used for **sturctural** studies like mapping TF binding sites (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



Case Study

NATURE | LETTER

previous article next article

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

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

The experiment

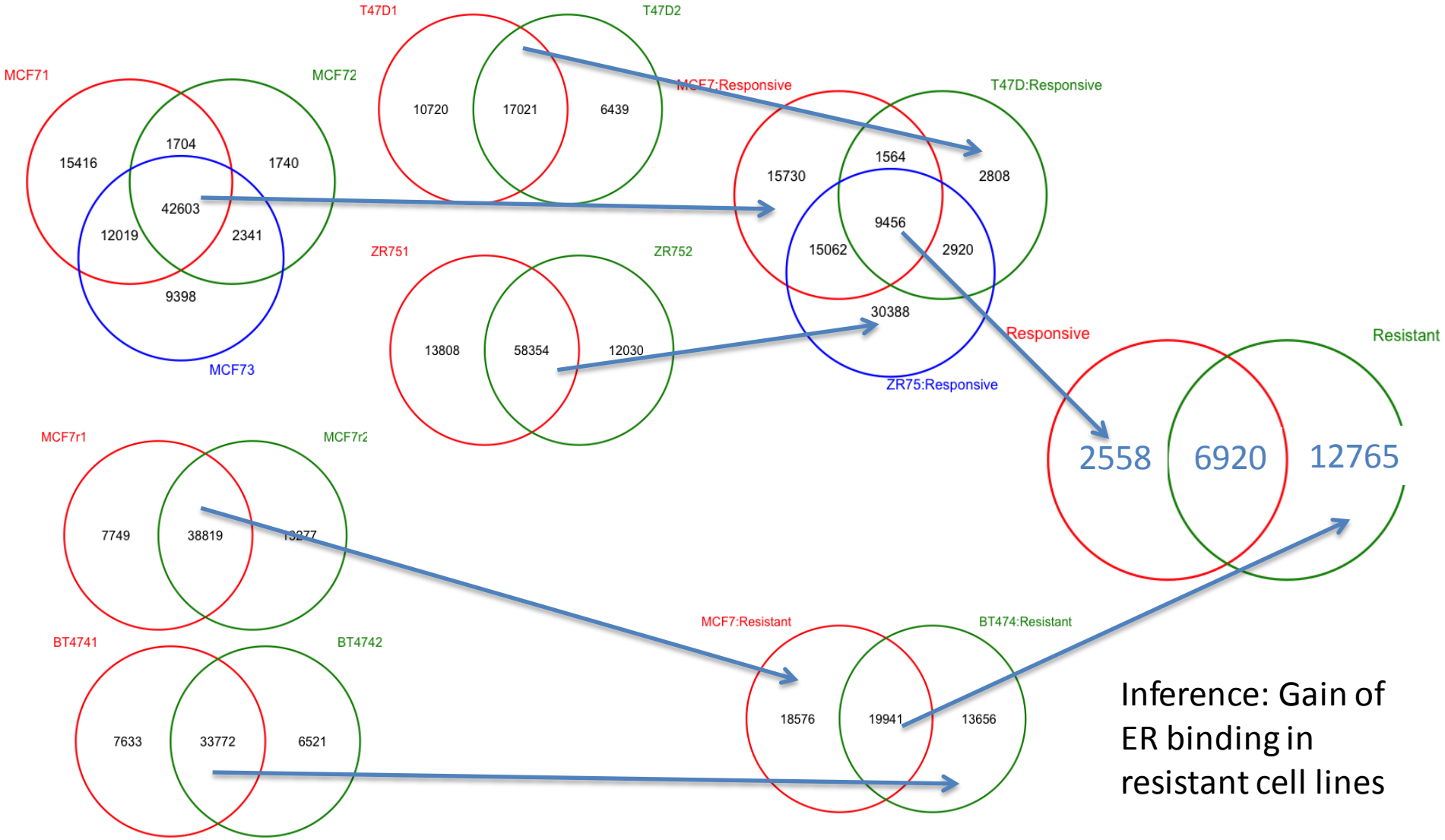
- Oestrogen receptor ChIP-Seq
- 4 drug-responsive breast cancer cell lines.
  - 2-3 replicates each
- 2 drug-resistant breast cancer cell lines
  - 2 replicates each
- "High-confidence" peaksets: called by at least two peak callers (MACS and Swembl)
- **What regions are differentially bound between the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

The experiment

11 Samples, 104051 sites in matrix (145586 total):

	ID	Tissue	Factor	Condition	Replicate	Caller	Intervals
1	MCF71	MCF7	ER	Responsive	1	macs	74029
2	MCF72	MCF7	ER	Responsive	2	macs	49075
3	MCF73	MCF7	ER	Responsive	3	macs	67130
4	T47D1	T47D	ER	Responsive	1	macs	28713
5	T47D2	T47D	ER	Responsive	2	macs	23575
6	ZR751	ZR75	ER	Responsive	1	macs	74971
7	ZR752	ZR75	ER	Responsive	2	macs	70560
8	MCF7r1	MCF7	ER	Resistant	1	macs	47023
9	MCF7r2	MCF7	ER	Resistant	2	macs	52517
10	BT4741	BT474	ER	Resistant	1	macs	41924
11	BT4742	BT474	ER	Resistant	2	macs	40783

Occupancy Analysis I: Strict consensus peaks

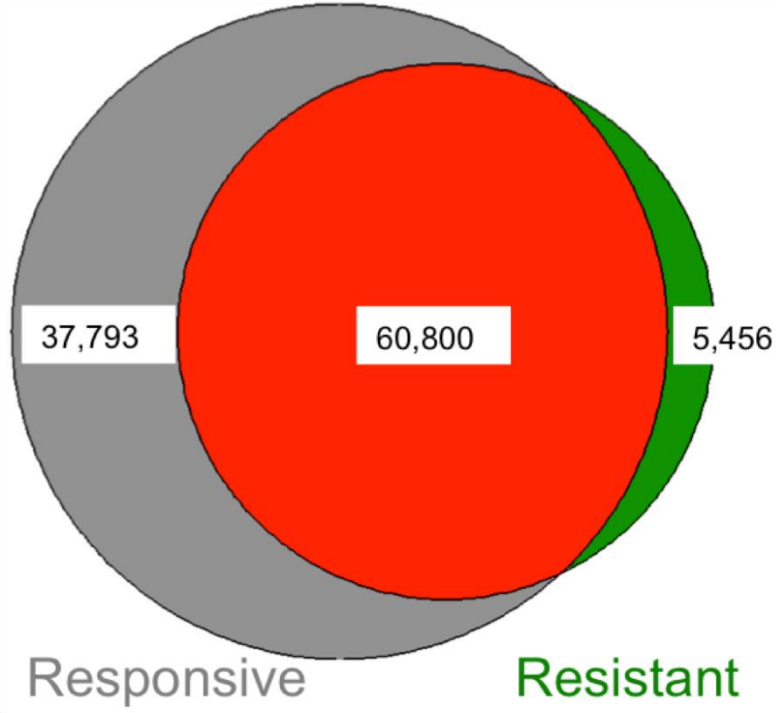


Inference: Gain of ER binding in resistant cell lines

Occupancy Analysis II

All peaks identified in at least 2 samples

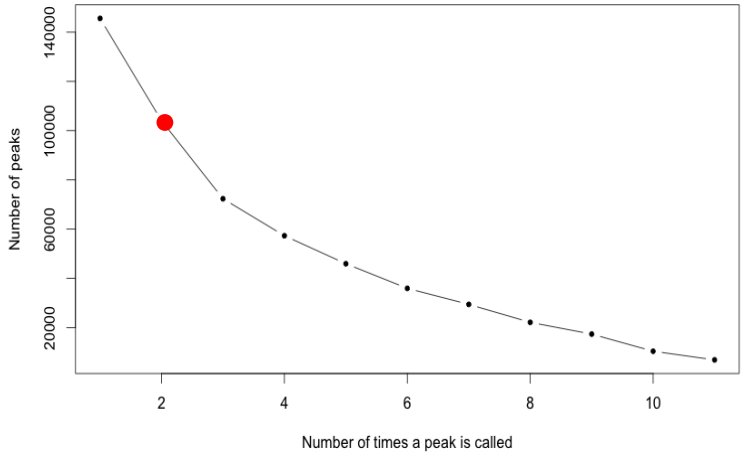
- Responsive only:
  - >= 2 Responsive samples
  - <2 Resistant samples
- Resistant only:
  - >= 2 Resistant samples,
  - <2 Responsive samples

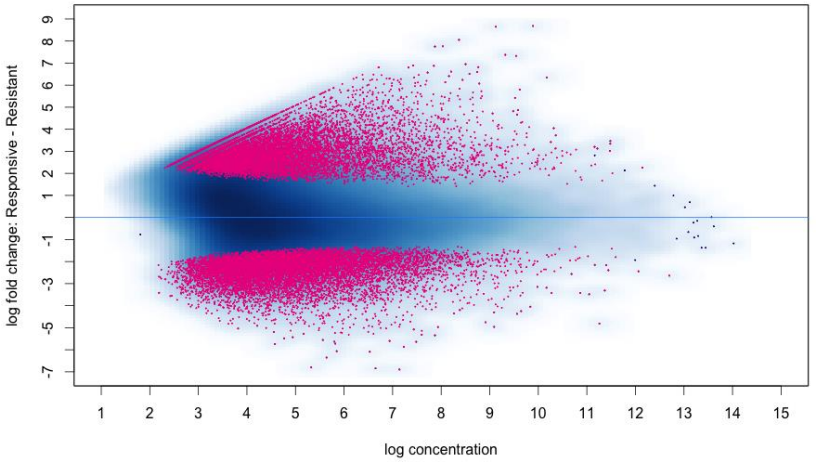


Inference: Loss of ER binding in resistant cell lines

Differential Binding Workflow (DiffBind)

- 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



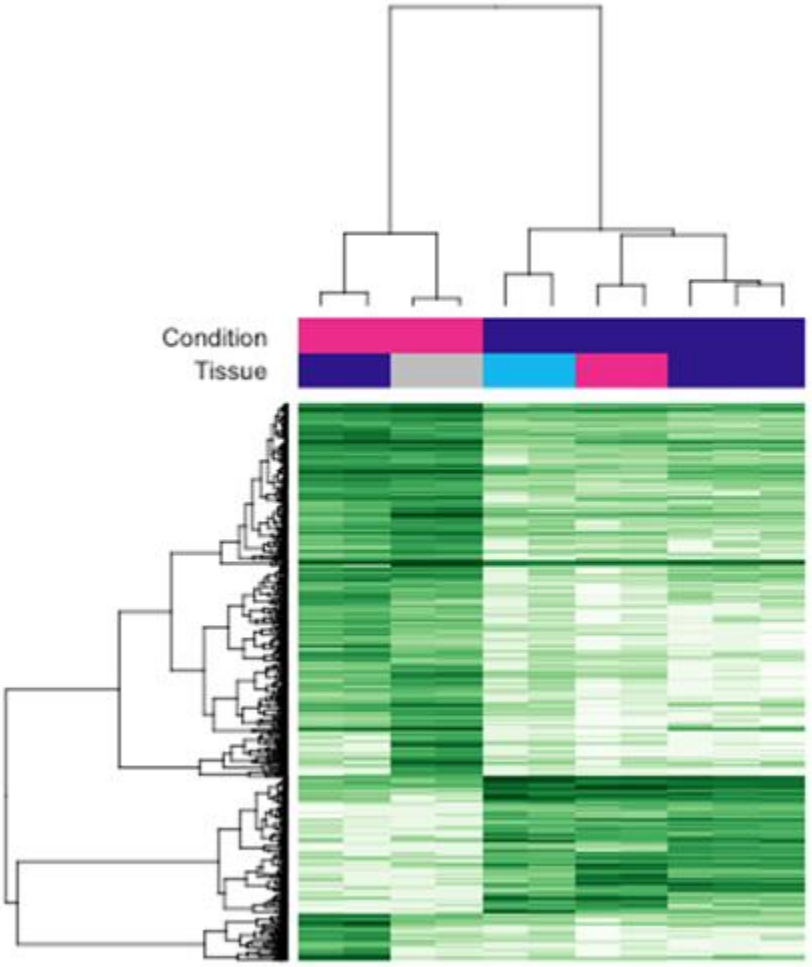


UNIVERSITY OF CAMBRIDGE

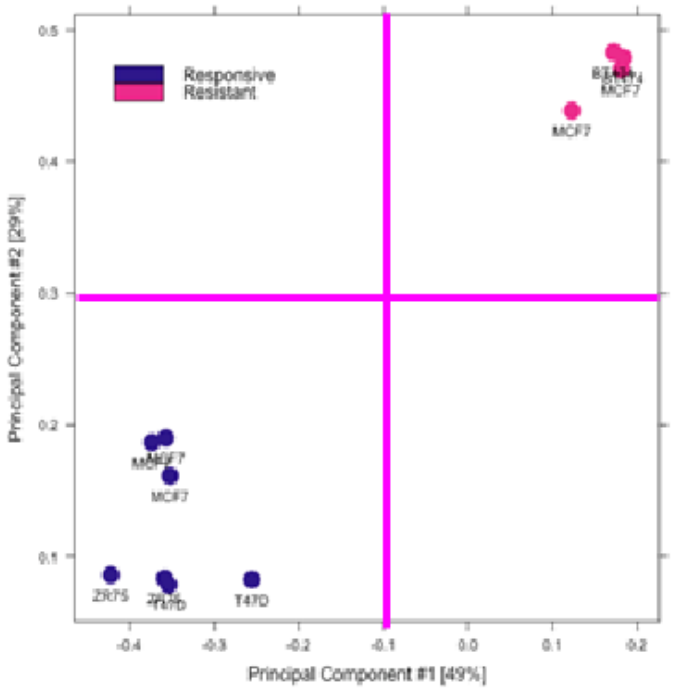
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

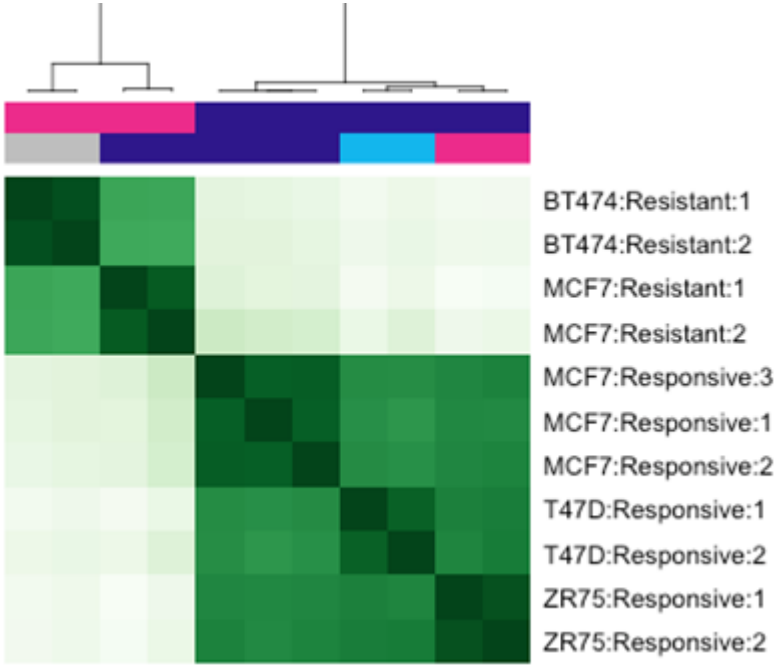
Differential Binding Workflow (DiffBind)

- 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



Affinity Analysis



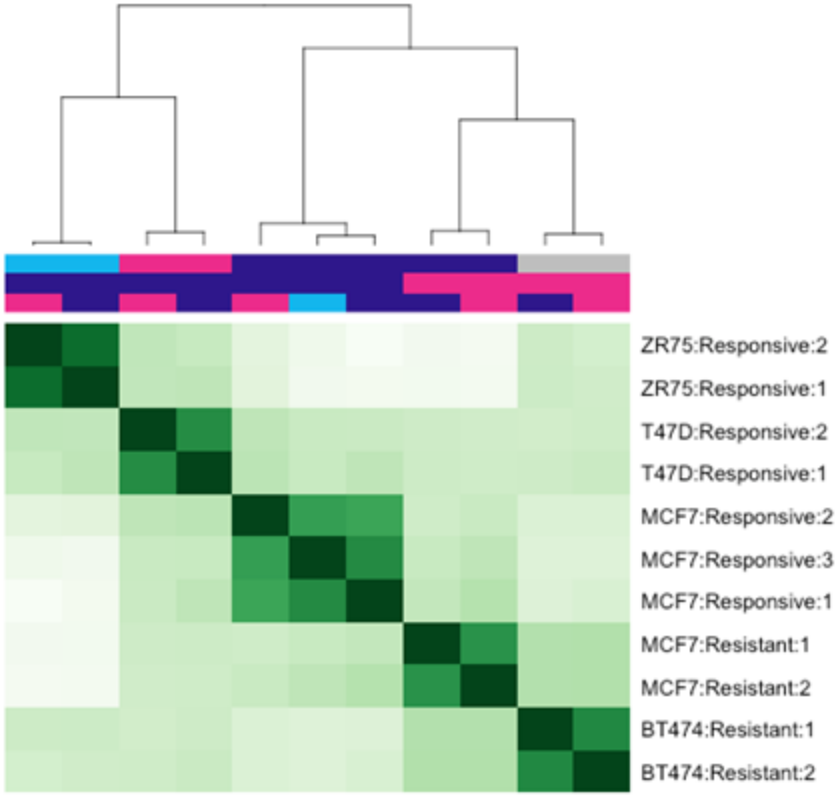


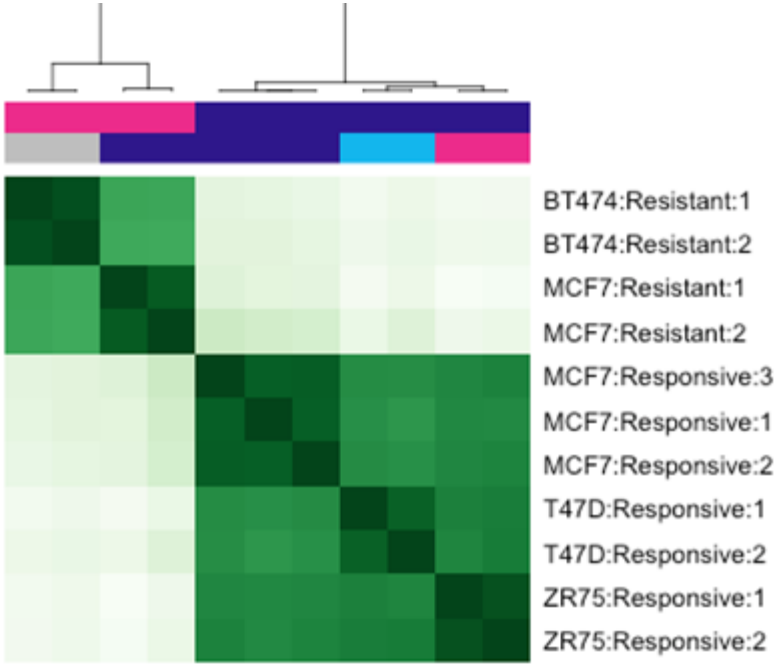
1,808 sites differentially bound at FDR <= 0.005

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Occupancy v Affinity Analysis

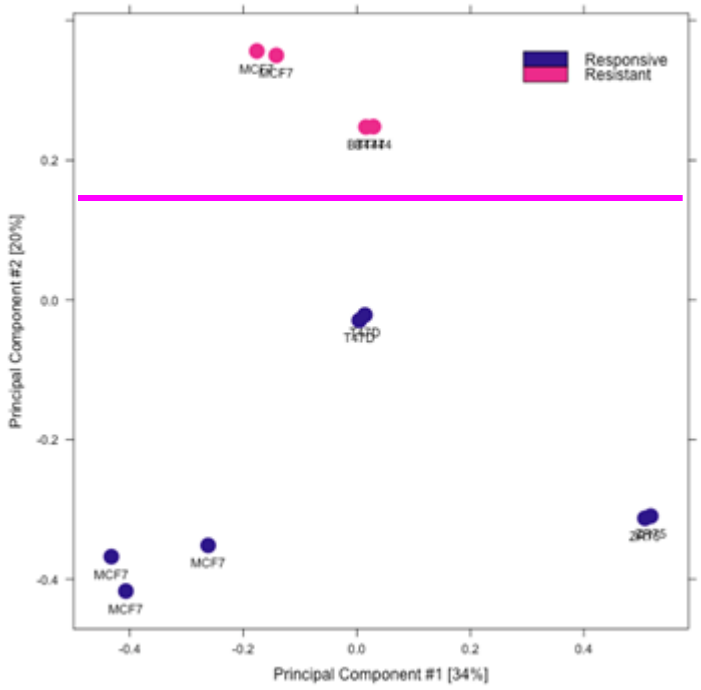


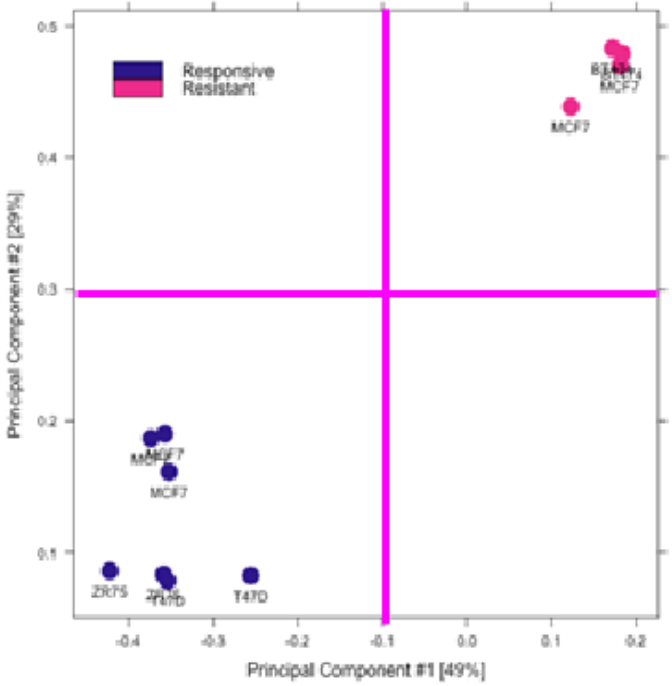


UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Occupancy v Affinity Analysis



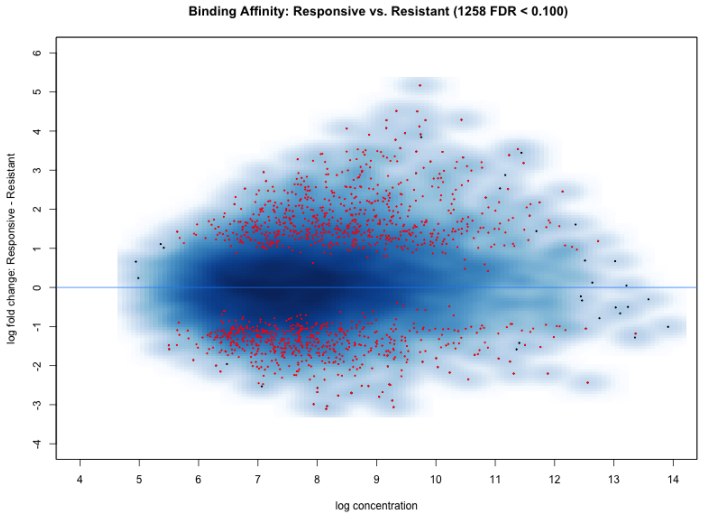


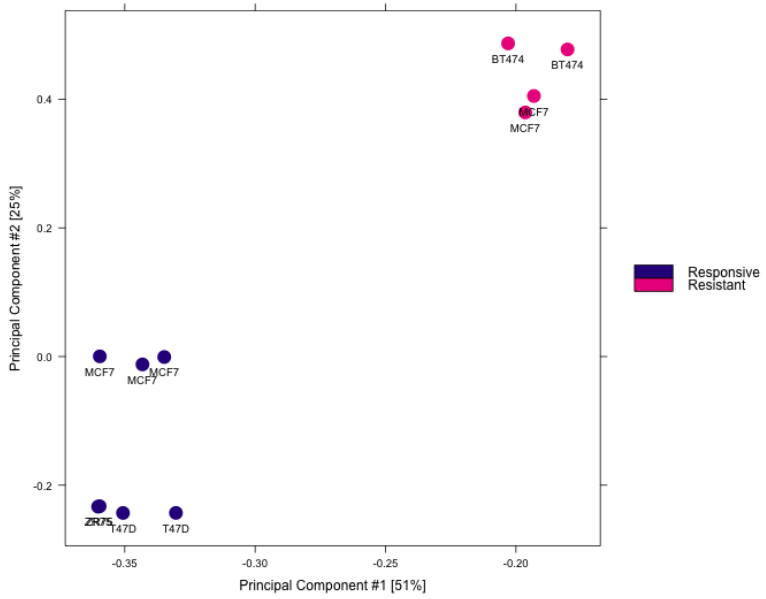
UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Occupancy v Affinity Analysis

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



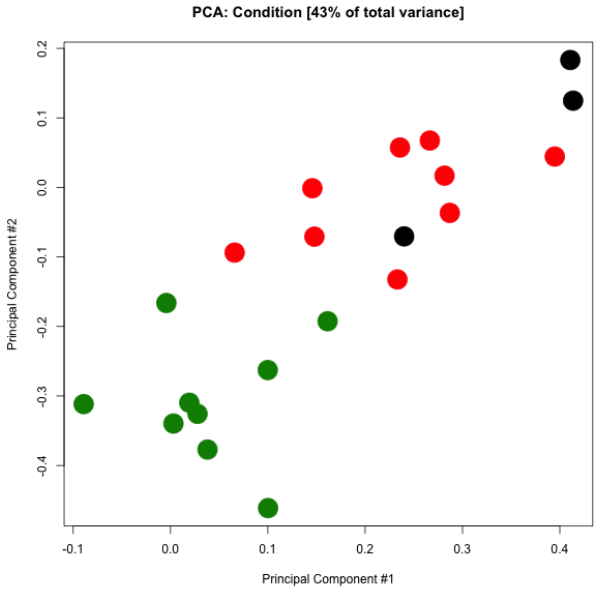


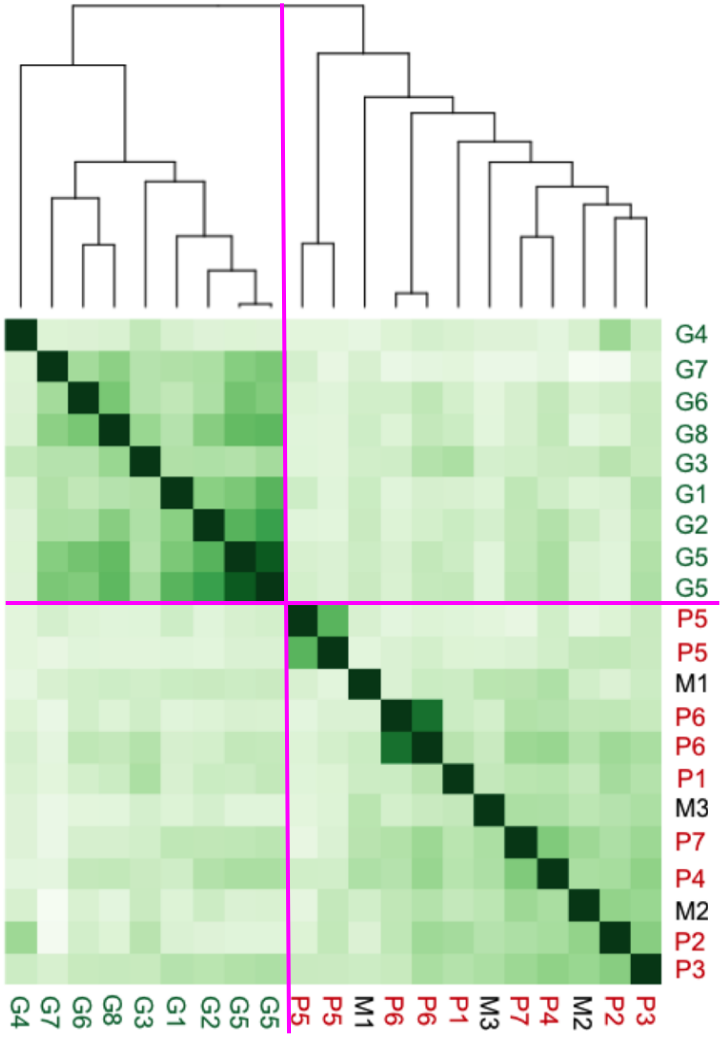
UNIVERSITY OF CAMBRIDGE

Affinity Analysis: Tumour Data

1,791 sites identified as differentially bound between good and poor prognosis (based on PR/HER2 status)

- **599** enriched in good prognosis
- **1,192** enriched in poor pronosis





UNIVERSITY OF CAMBRIDGE

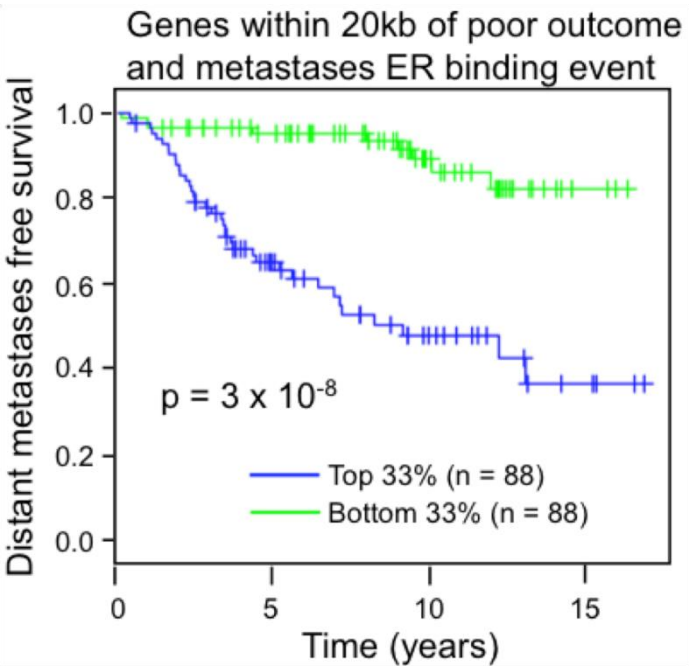
Affinity Analysis: Tumour Data

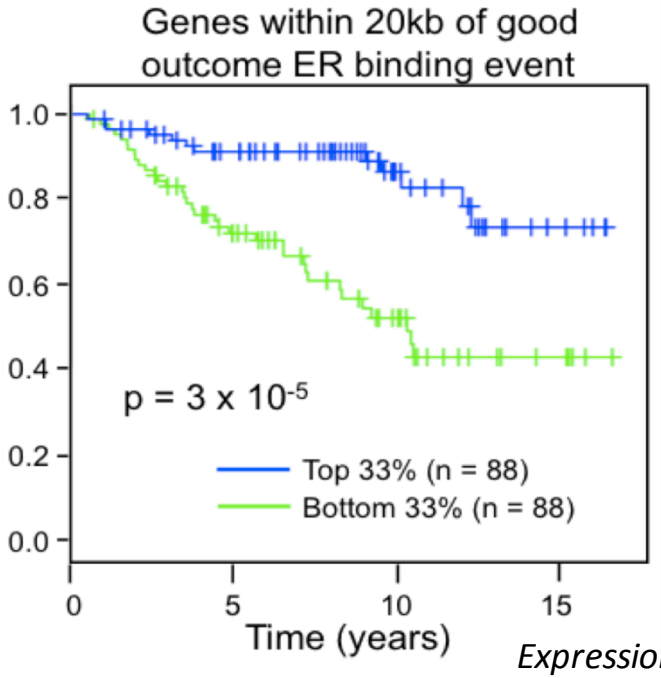
Differentially enriched co-factor motifs

	Tumour Prognosis	Tamoxifen Resistance	Mitogenic Cocktail
Poor/Metastatic tumours Tamoxifen Resistant Mitogenic MCF7	ERE FoxA1	Pax2 AP-1 FoxA1 ERE	Pax2 NFE2L2 FoxA1 ERE
Good tumours Tamoxifen Responsive Normal MCF7	ERE	ERE GATA	GATA ERE

Affinity Analysis: Tumour Data

Genes near DB sites form prognostic gene signatures





Expression data: Loi et al.

- Signature composed of genes within 20kb of DB sites
  - **265** genes in Poor outcome signature
  - **109** genes in Good outcome signature
- Classifier based on up/down regulation in mRNA expression sets
- Validated in 7 publicly available BC expression datasets

UNIVERSITY OF CAMBRIDGE

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

Further reading

Nucleic Acids Research Advance Access published May 22, 2014
Nucleic Acids Research, 2014 **1**
doi: 10.1093/nar/gku351

De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly

Aaron T.L. Lun[1,2] and Gordon K. Smyth[1,3,*]