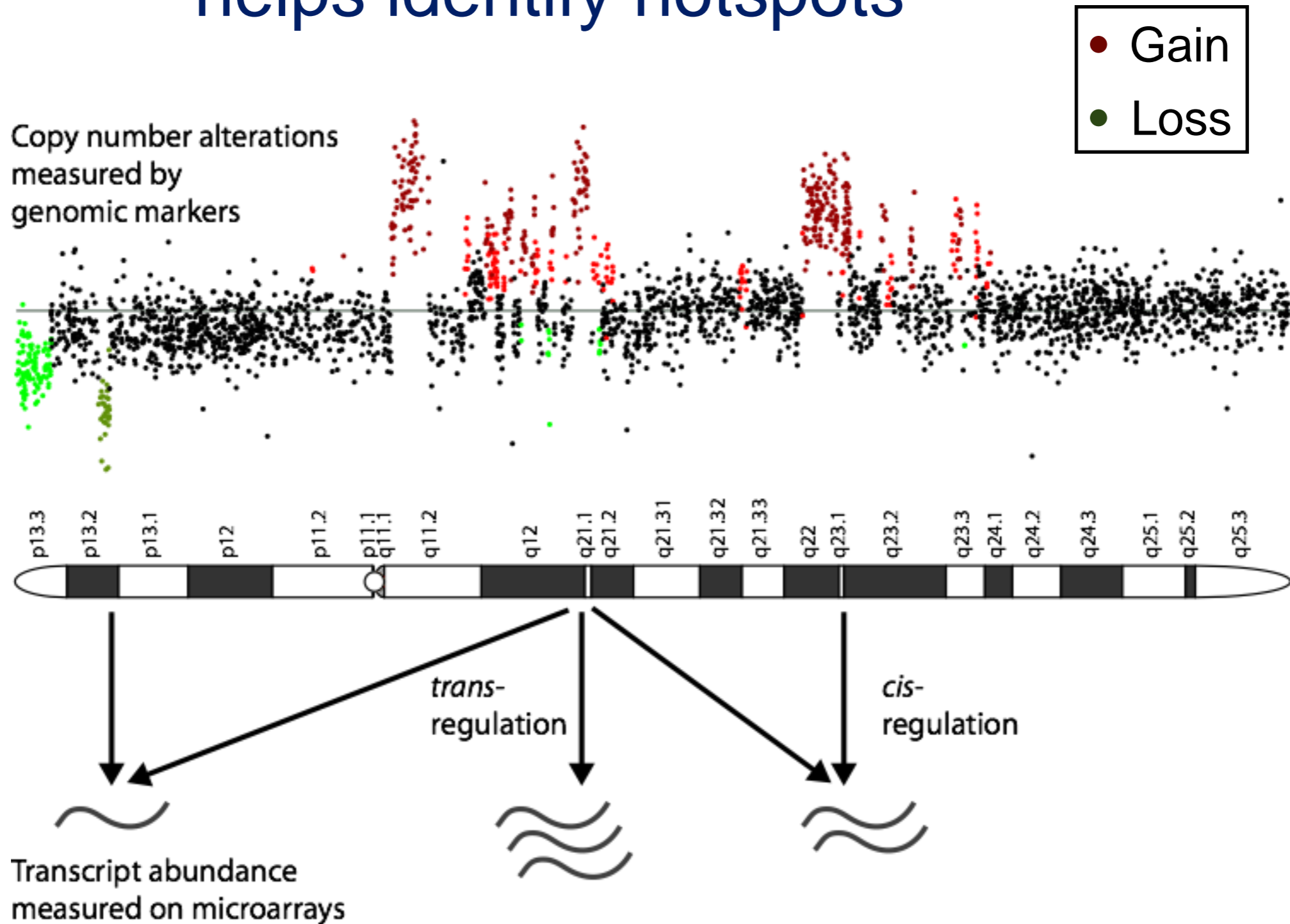# Inferring deregulation networks of copy-number driven expression in cancer sub-types

Yinyin Yuan
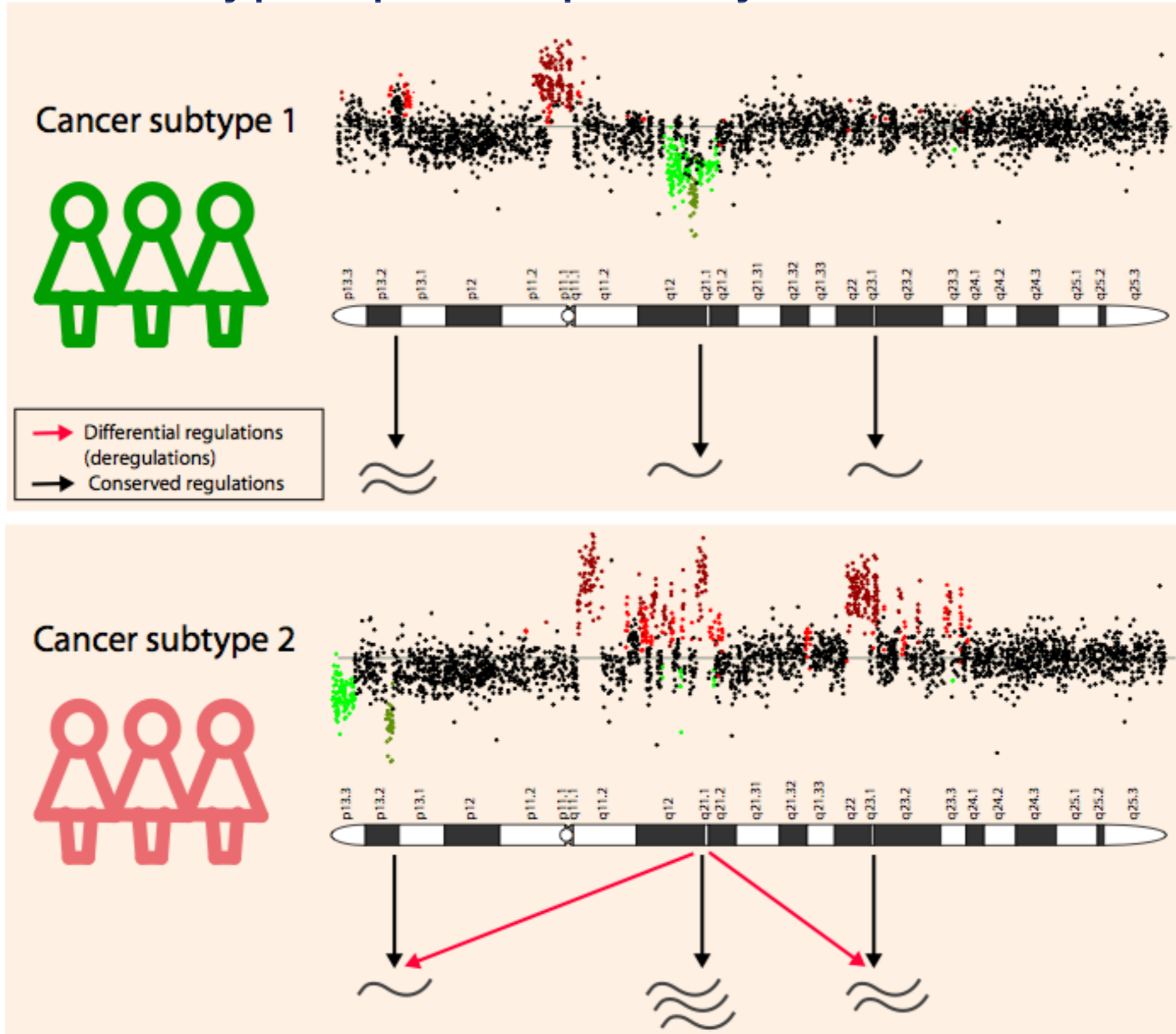
Florian Markowetz Lab
Cambridge Research Institute, Cancer Research UK
Department of Oncology, University of Cambridge

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK

Introduction

Regulatory network of CN-driving Expression helps identify hotspots

- Gain
- Loss

Copy number alterations measured by genomic markers

p13.3 p13.2 p13.1 p12 p11.2 p11.1 q11.1 q11.2 q12 q21.1 q21.2 q21.31 q21.32 q21.33 q22 q23.1 q23.2 q23.3 q24.1 q24.2 q24.3 q25.1 q25.2 q25.3

trans- regulation

cis- regulation
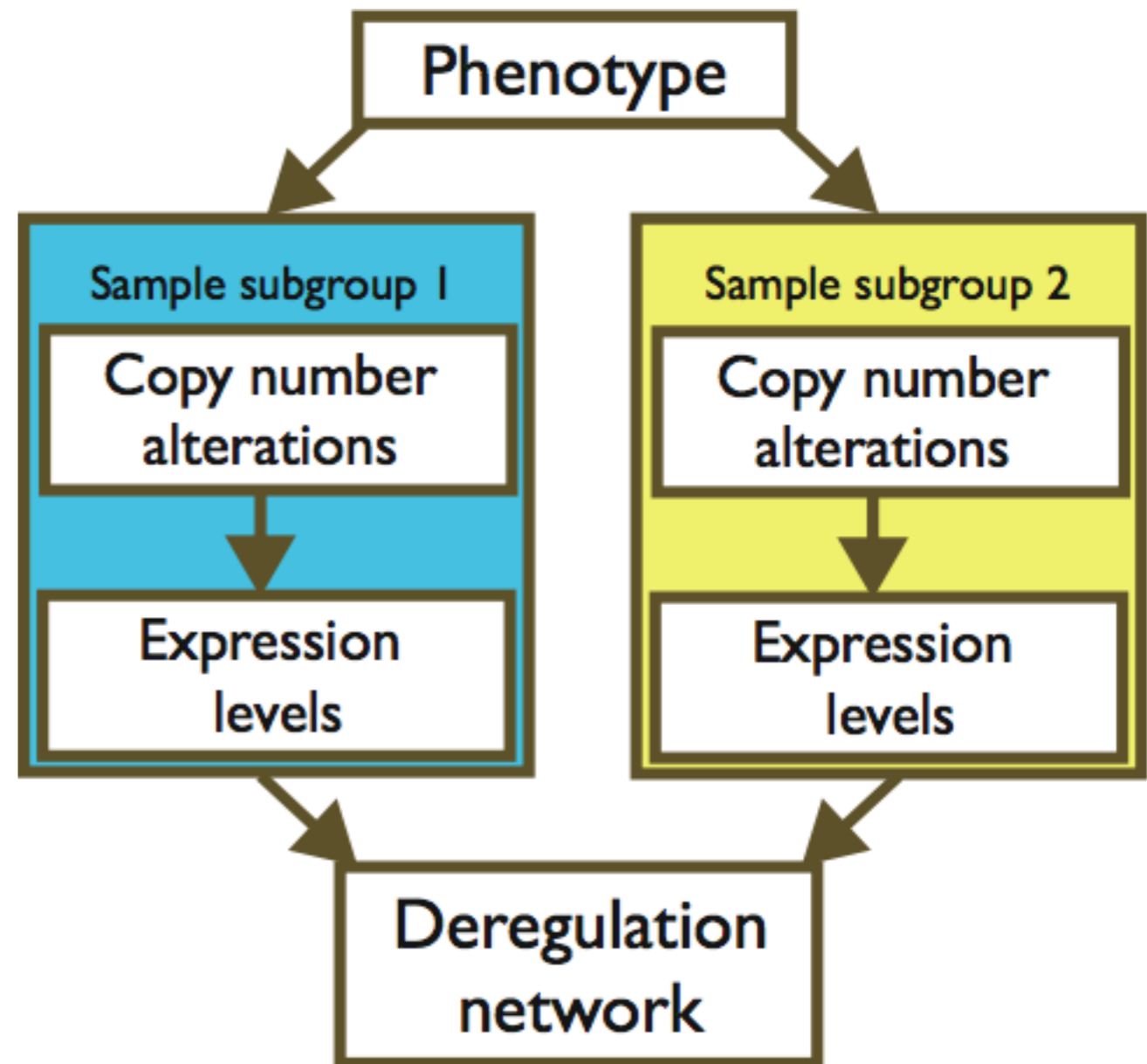
Transcript abundance measured on microarrays

Deregulation network leads to the discovery of subtype-specific primary aberrations
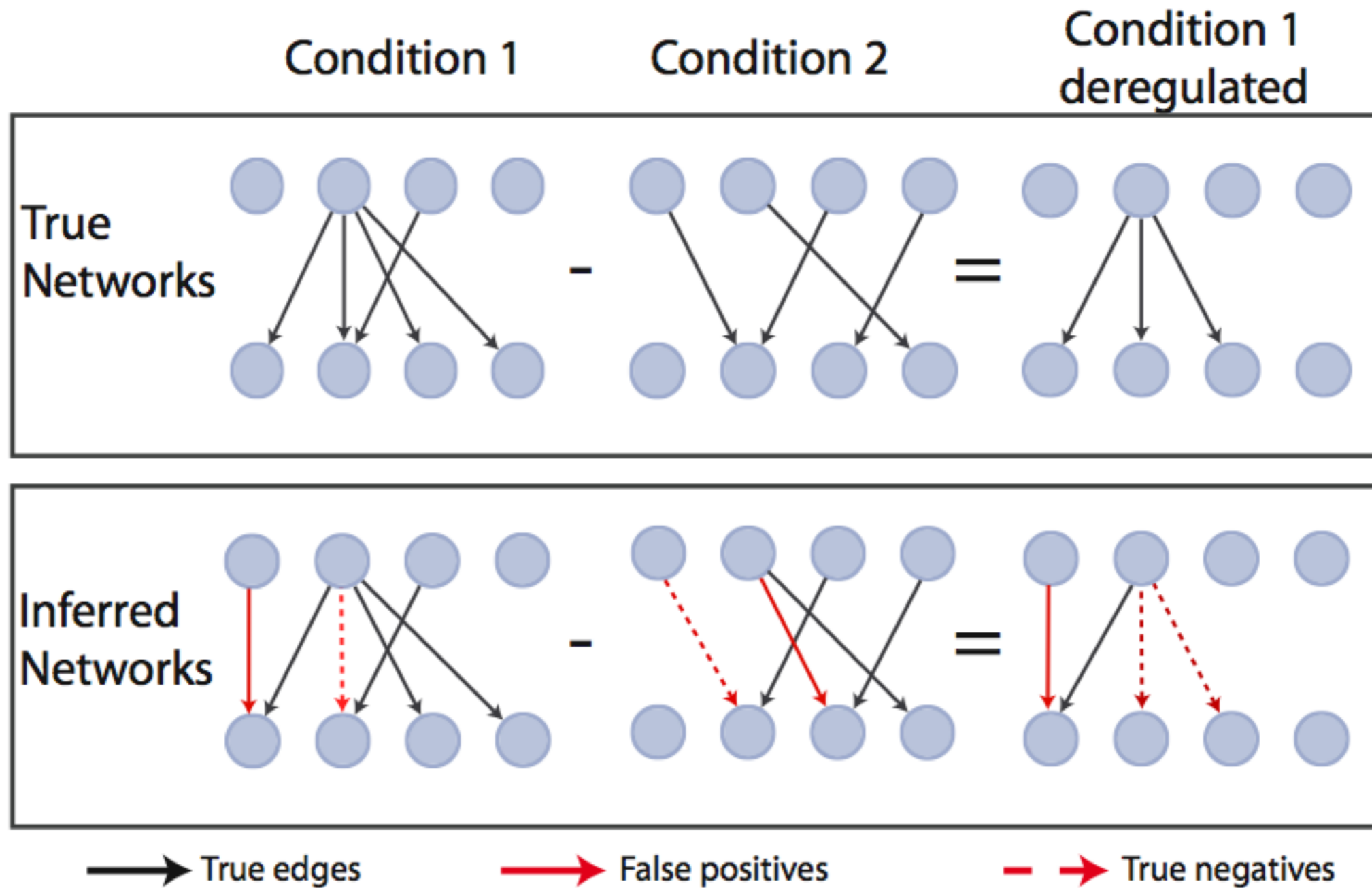
Aim

To reveal the deregulation network between subgroups of samples

Outline

- One step inference of the deregulation network
- Introduce DANCE: a sparse inference model
- Experiments on simulated data and cancer datasets

Mistakes in network influence outcome get amplified with two networks

DANCE

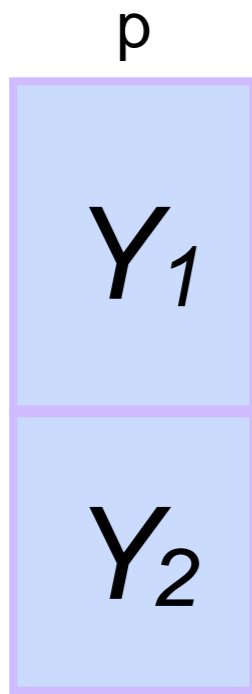## Deregulation Analysis of Networks between Copy-number and Expression
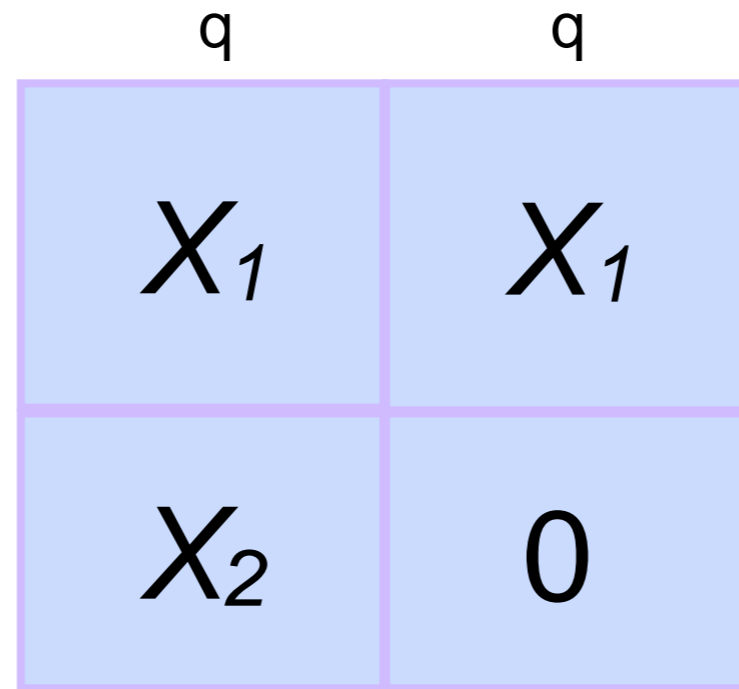
DANCE

Expression data
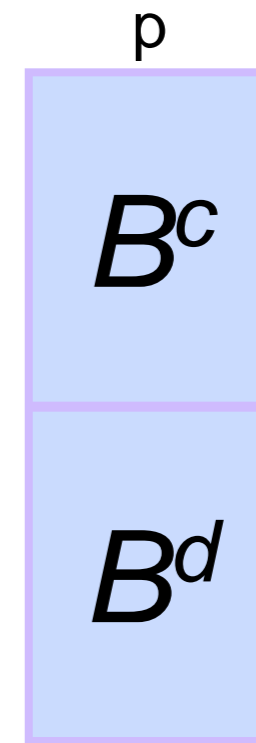
Copy number data

Sparse coefficients

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_1 \\ X_2 & 0 \end{bmatrix} \begin{bmatrix} B^c \\ B^d \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$\underbrace{\phantom{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}}}_{Y} \quad \underbrace{\phantom{\begin{bmatrix} X_1 & X_1 \\ X_2 & 0 \end{bmatrix}}}_{X'} \quad \underbrace{\phantom{\begin{bmatrix} B^c \\ B^d \end{bmatrix}}}_{B}$$

Objective function $\quad argmin_B \parallel Y - X'B \parallel_2 + \lambda \parallel B \parallel_1 ,$

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK

Experiments on HMM simulated data:
The sparse method outperforms others in most situations

- Data: HMM model for copy number simulation

- Sparse conserved and deregulation networks: coefficients from normal distributions

- Methods to compare: Gaussian Graphical Model, Partial Least Squares, DANCE.

- Prior knowledge: number of edges in both conserved and deregulation networks.
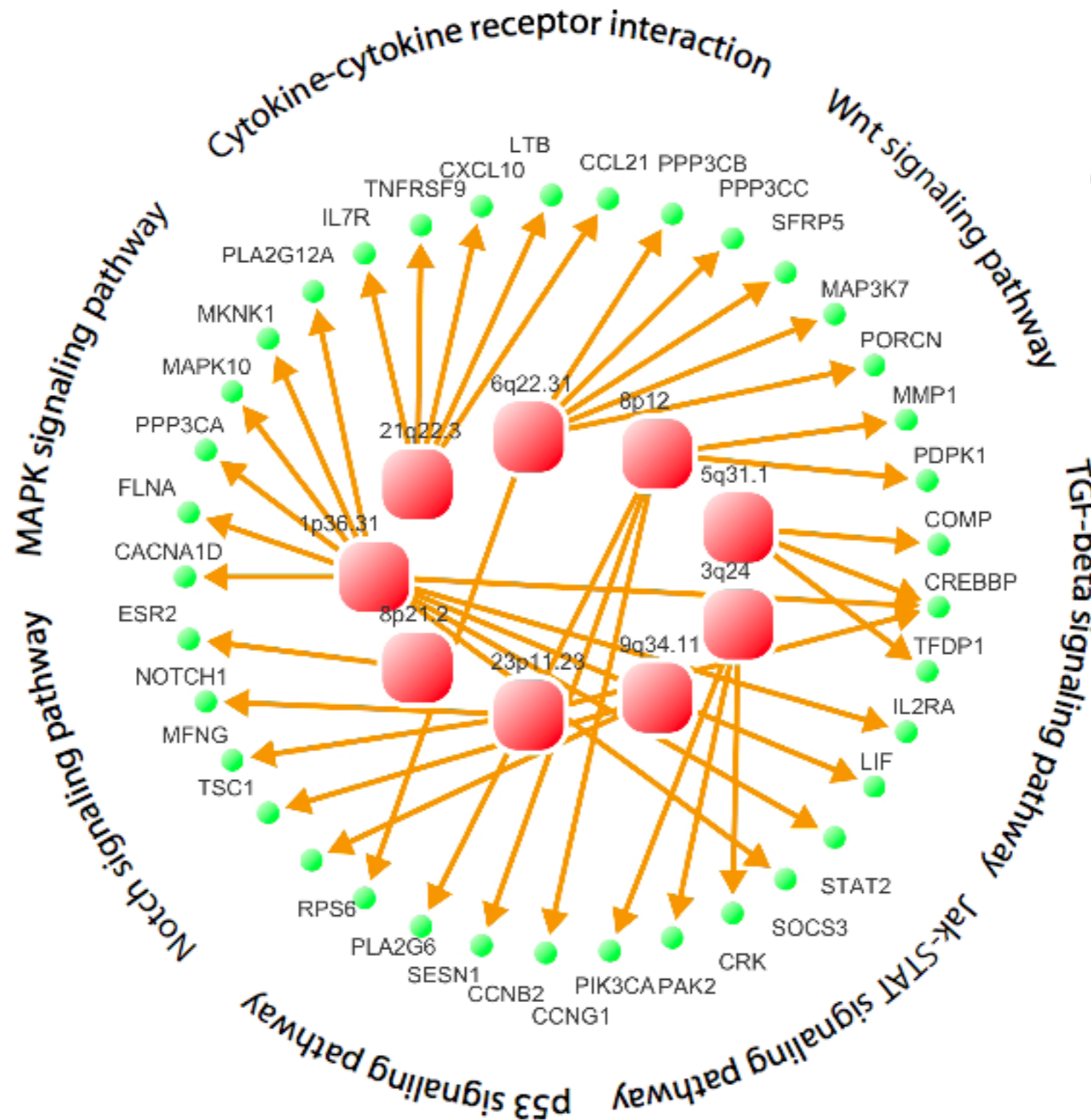
True positive rate
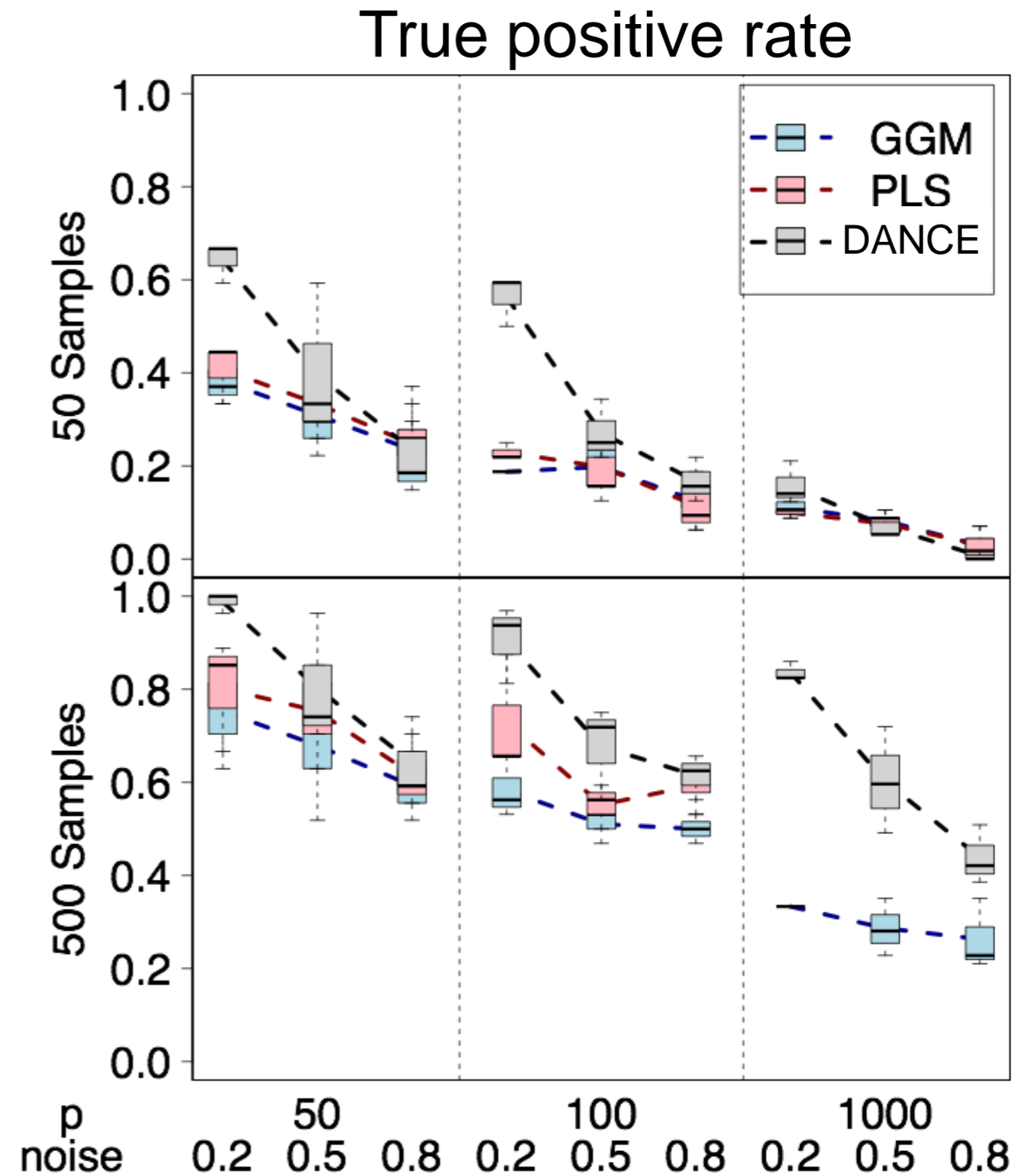
# Hotspots responsible for pathway deregulation

ER pos
vs.
ER neg

Concomitant
CN signatures

21q22
Amp with
few *cis-*
changes

8p12
Amp in 32%
tumours
Neutral in ER-

Data from Chin *et al*, 07

UNIVERSITY OF CAMBRIDGE

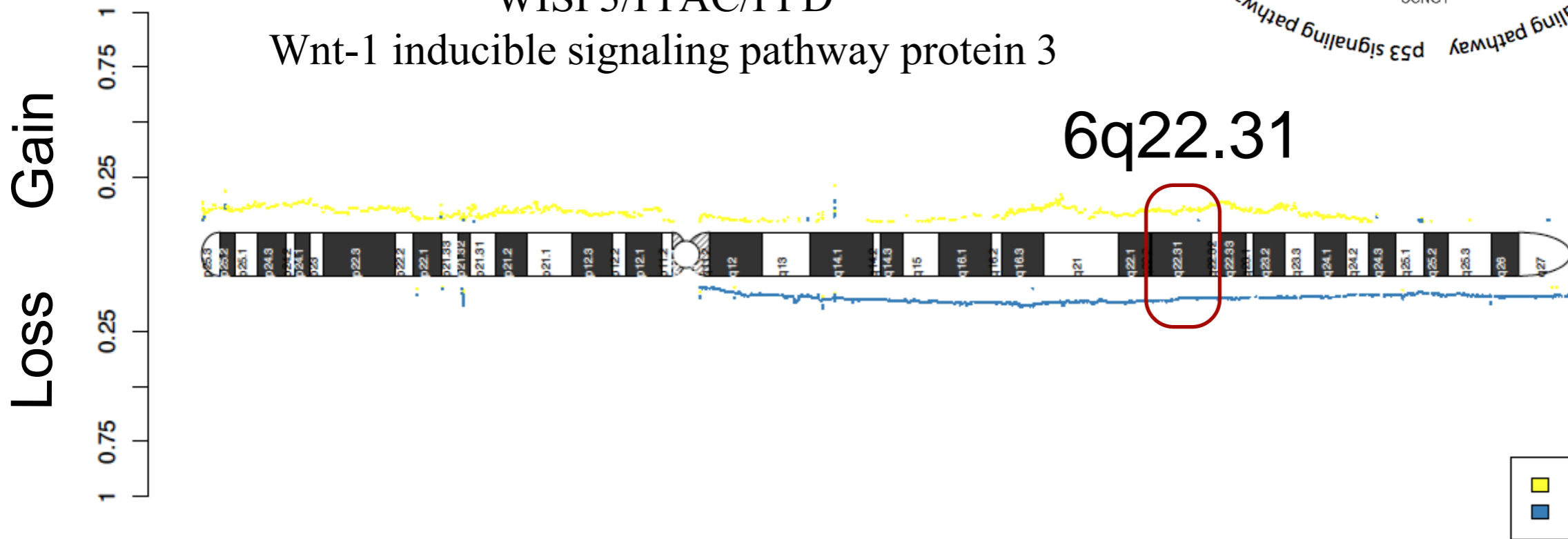CANCER RESEARCH UK

Hotspots responsible for pathway deregulation

Hotspots either exhibit imbalance between subtypes or a phenotypical consequence on expression
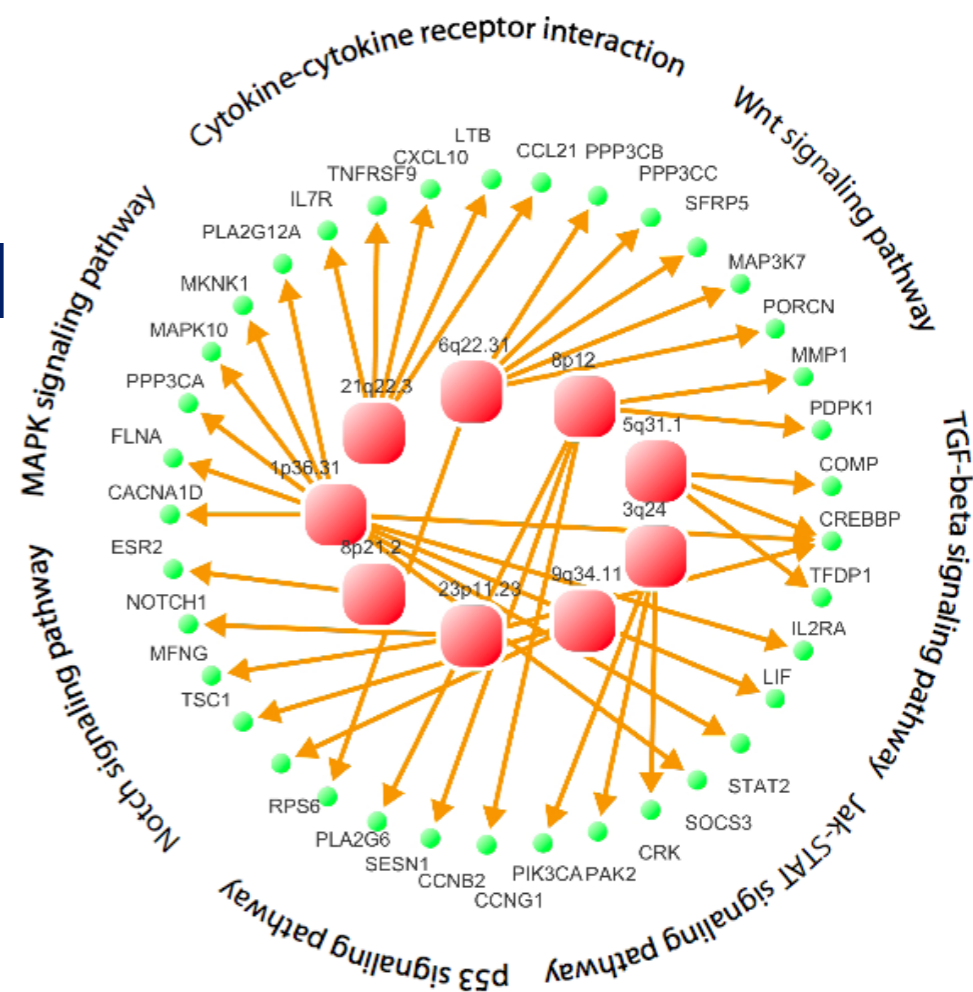
WISP3/PPAC/PPD
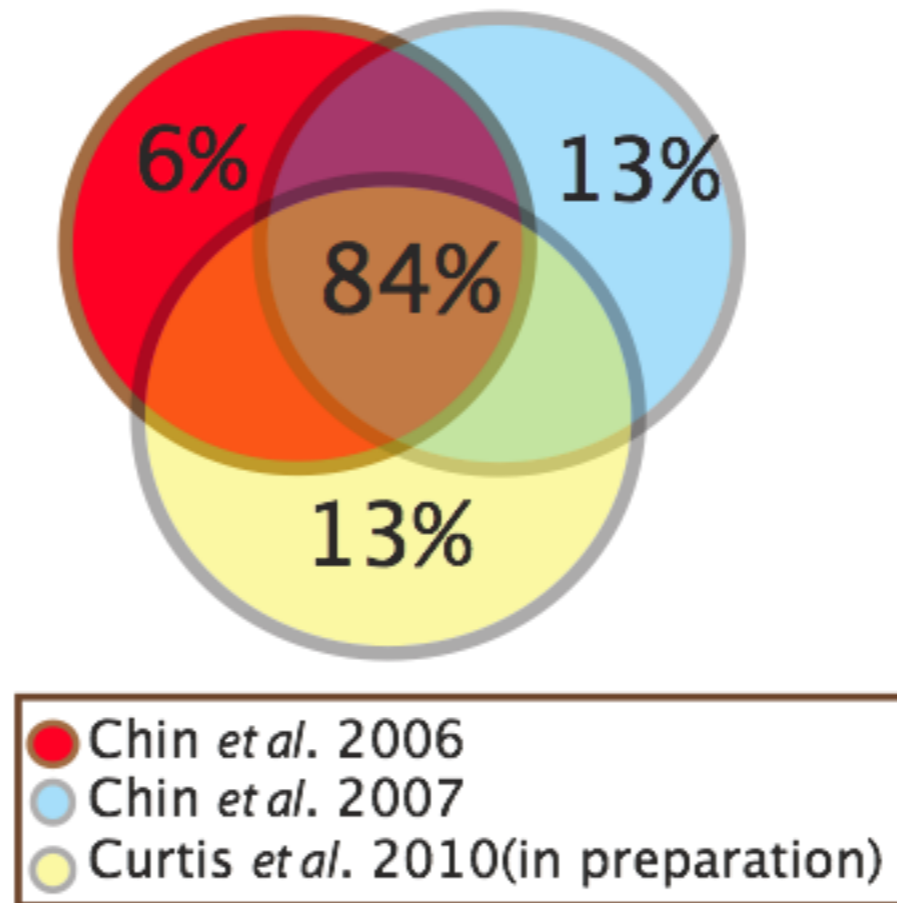Wnt-1 inducible signaling pathway protein 3

6q22.31

Experiments on three breast cancer datasets:

Good overlap among three datasets shows the robustness of the algorithm

Overlap of hotspots

6%   13%

84%

13%

- Chin *et al.* 2006
- Chin *et al.* 2007
- Curtis *et al.* 2010 (in preparation)

Deregulation in ER- breast cancer based on 12 KEGG pathways

Summary

- Introduce DANCE, a statistical package for the inference of deregulation network in disease subtypes

- The sparse model is efficient in large-scale data modelling

- DANCE shows good accuracy in simulated experiment and robustness on real biological data

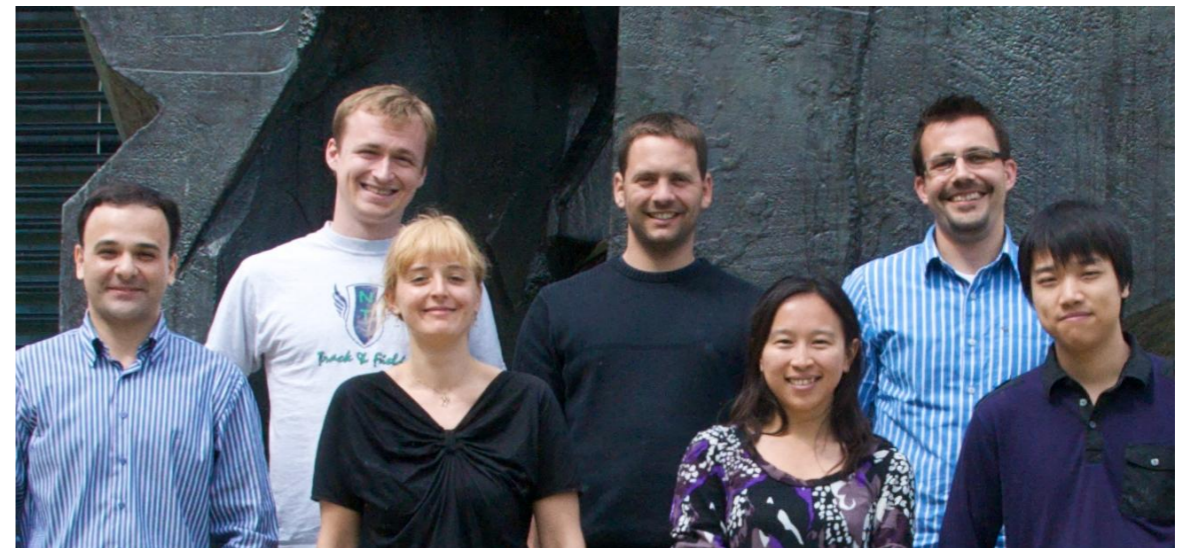- Hotspots leading to deregulated pathways exhibit genomic imbalance or imply *trans*-regulations

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK

Acknowledgement

UNIVERSITY OF CAMBRIDGE

HWL Hutchison Whampoa

CANCER RESEARCH UK

Thank you!

Questions?