T2DSystems: Modelling

From Statistical Modelling to Mechanisms

Michael P.H. Stumpf

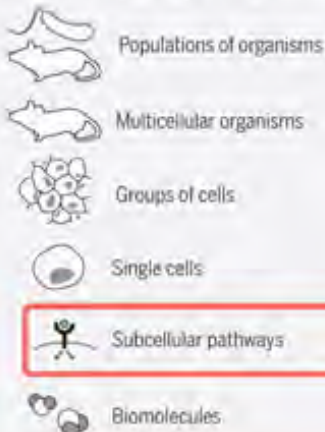Theoretical Systems Biology, Department of Life Sciences

Imperial College
London

Question

How does factor X affect proliferation of cell type Y?

Scales of biological systems

Populations of organisms

Multicellular organisms

Groups of cells

Single cells

Subcellular pathways

Biomolecules
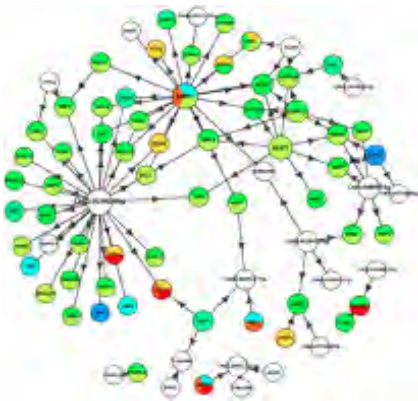
Models

Cell type Y

Kirk et al., Science (2015).

Network Inference: Use of transcriptomic data to learn the structure and dynamics of gene regulation networks (or co-regulation networks)

Data Fusion: Combine different types of (e.g. transcriptomic, proteomic, metabolomic, imaging, clinical) data into a single, coherent modelling framework.

Mechanistic Modelling: Develop and evaluate experimentally testable models (hypotheses) and explore their behaviour.

Networks: Mapping Processes and Understanding

Hypothesis Testing vs. Model Selection

Falsification — Karl Popper

We can *never* prove (or accept) a hypothesis to be correct. But a single counter example suffices to reject it.

Popper and two Hypotheses on FIFA

FIFA is a corrupt organization $\Leftrightarrow$ FIFA is not corrupt

The first hypothesis can never be falsified. But a single instance of corruption would be sufficient to reject the second hypothesis.
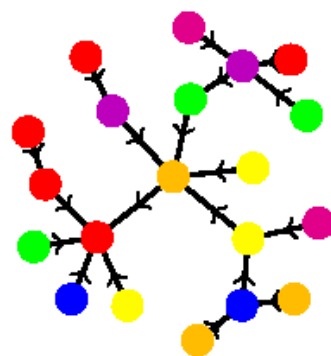
Bayesian Model Selection

In the Bayesian framework it makes sense to speak of a hypothesis to be true (conditional on the data and the proposed alternatives). We can also quantify the probability for other agencies (e.g. IAAF, IOC, Berlusconi's Italy, the Vatican or Oxfam) to be corrupt.
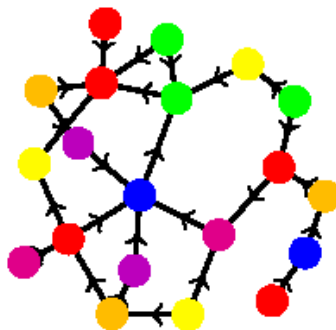
Networks: Mapping Processes and Understanding
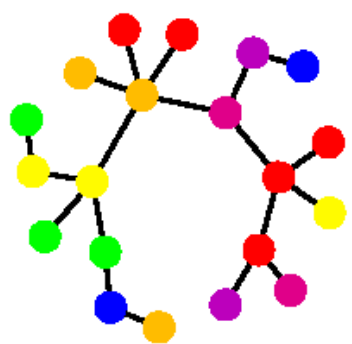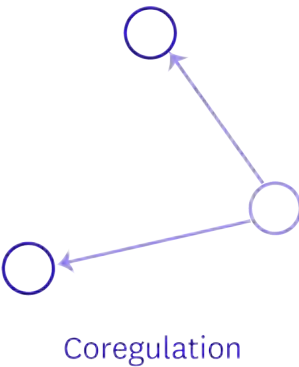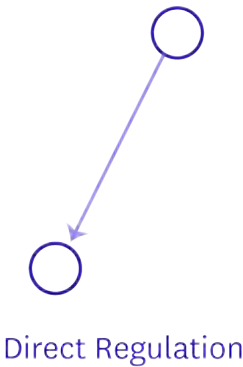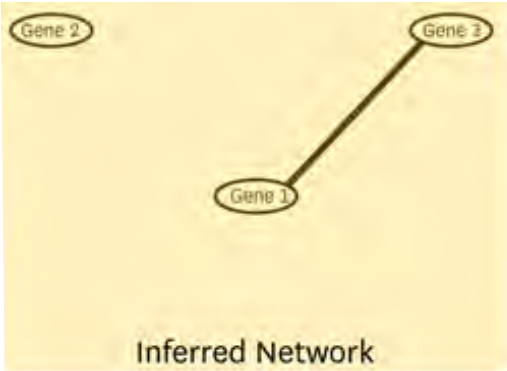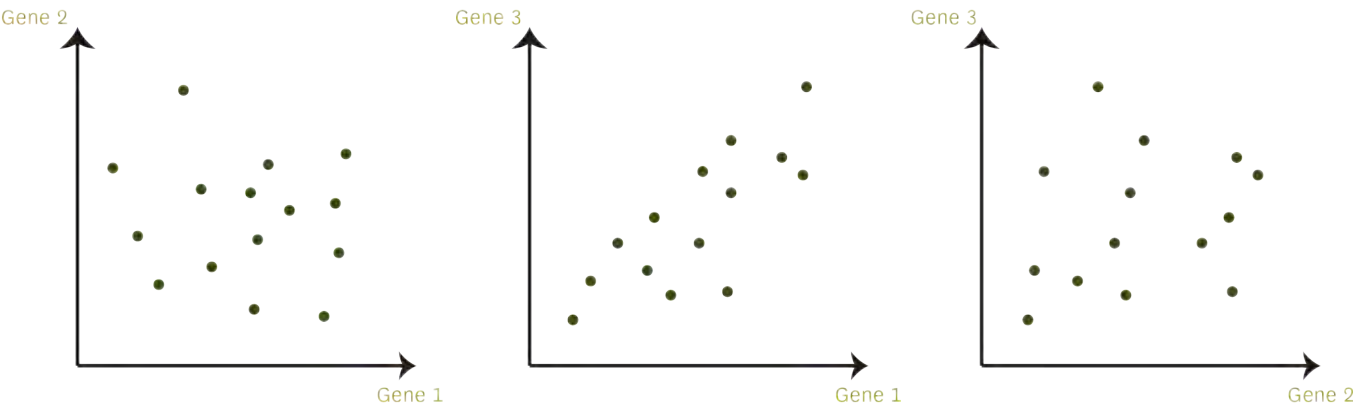

Gaussian
Graphical Models

Bayesian Networks

Dynamic Bayesian Networks

Network Inference Basics

Inferred Network

Direct Regulation

Coregulation
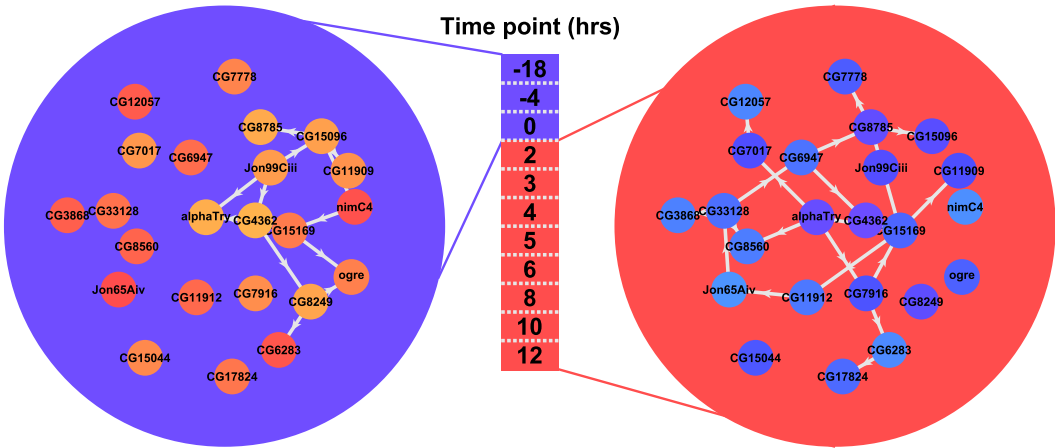
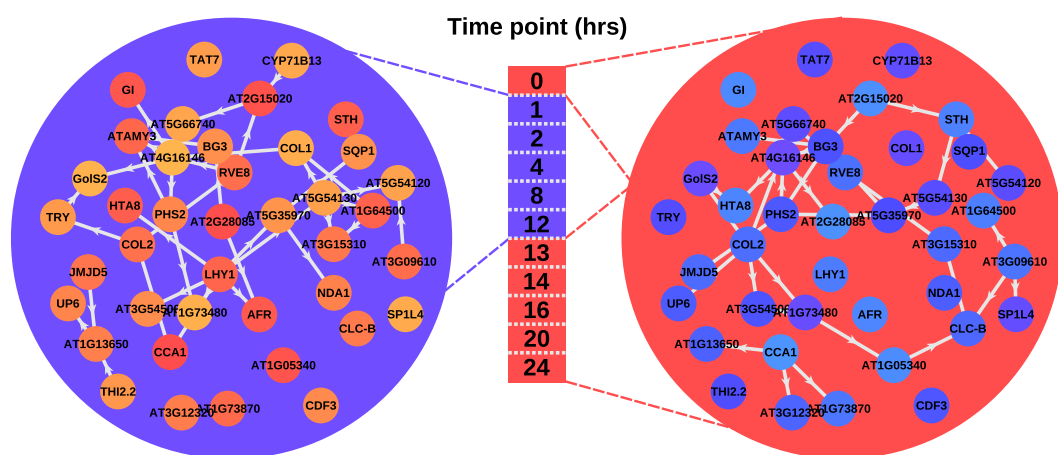# Biological Systems Over Time And Between Conditions
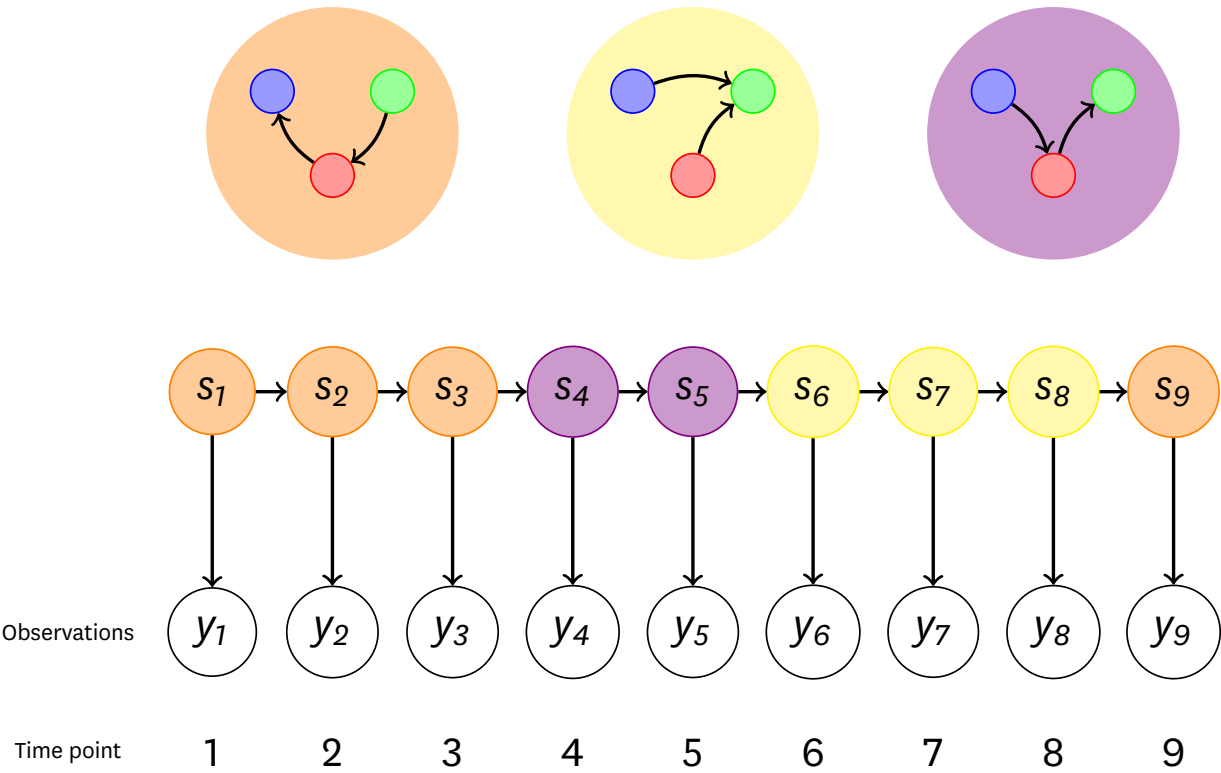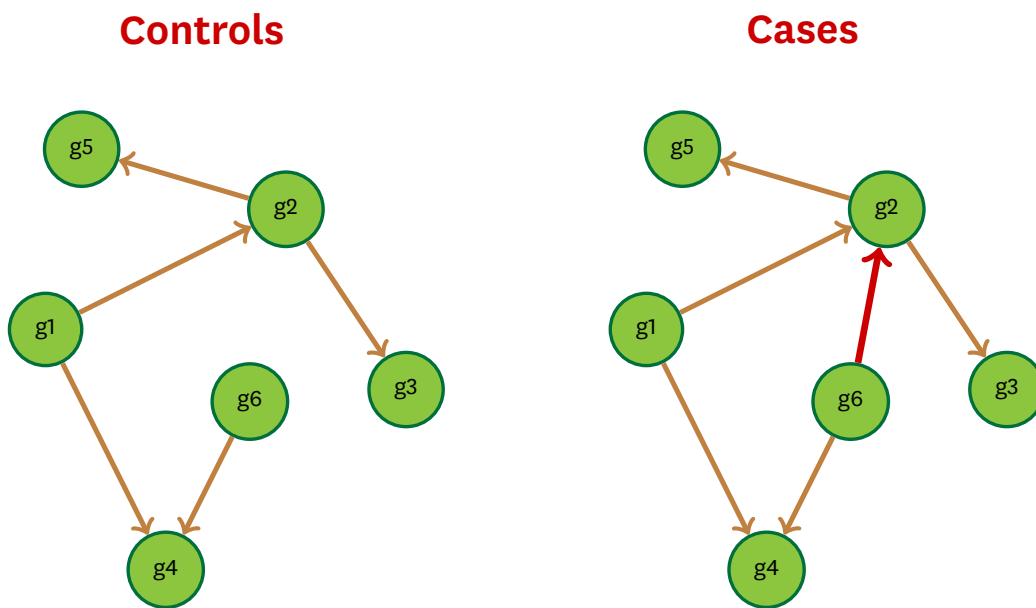
D. melanogaster development

Expression data for *D. melanogaster* midgut development. Taken at 11 time points, during which larval midgut becomes adult midgut.

A. thaliana diurnal cycle

Expression data for *A. thaliana* over 24 hours, with a light and dark phase.
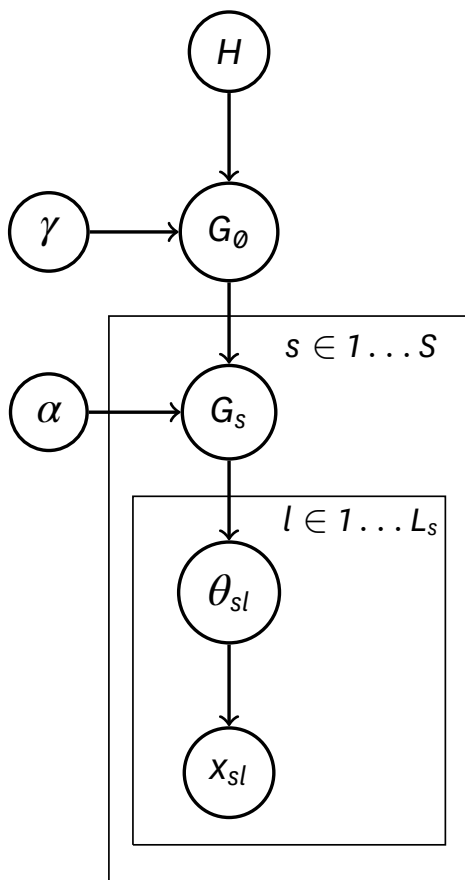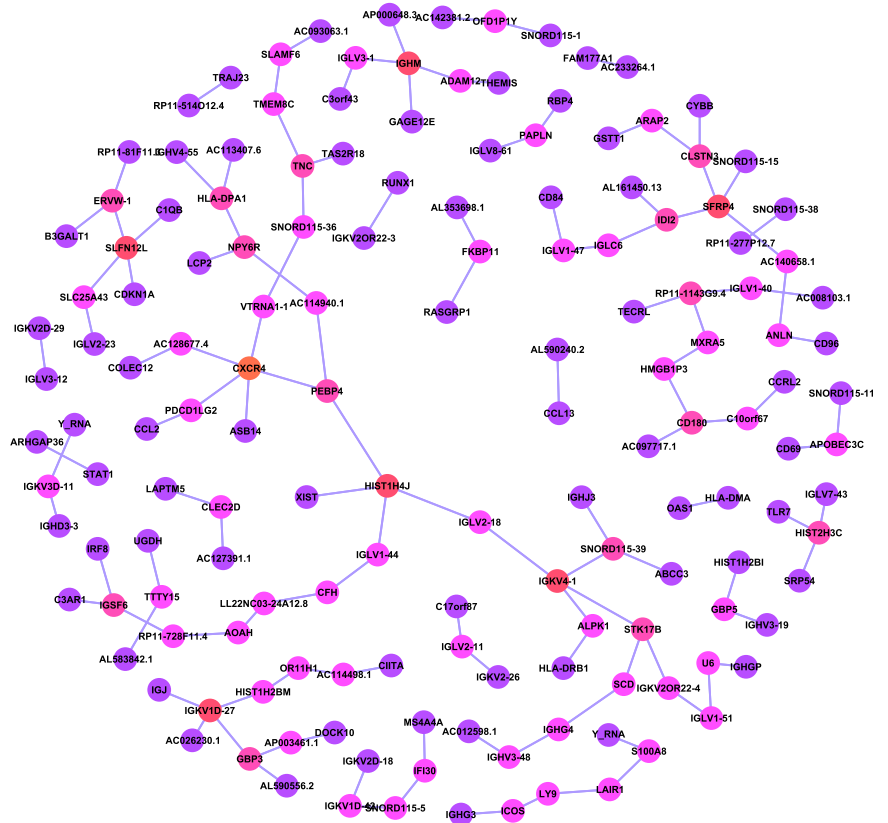
Networks from Disease and Control Cases

Models of Gene Regulation Networks

- Networks are assumed to be largely but not completely shared between cases and controls.
- We infer the networks and evidence for differences between them.

Network containing inferred interactions particular to sporadic inclusion body myositis (sIBM). From this set of hypotheses we can select suitable candidates for further analysis.
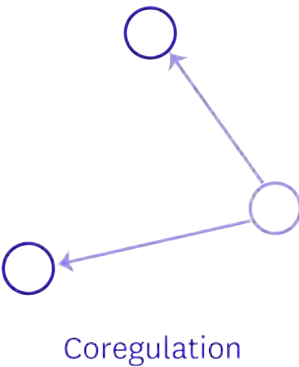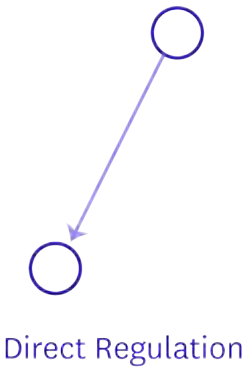
Thorne & Stumpf, Bioinformatics (2012); Thorne et al., MolBiosyst (2013).

Network Inference Basics

Inferred Network

Direct Regulation

Coregulation
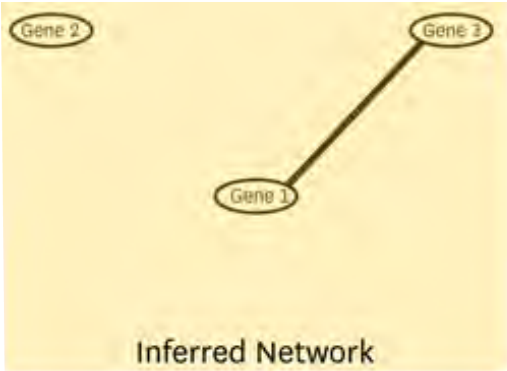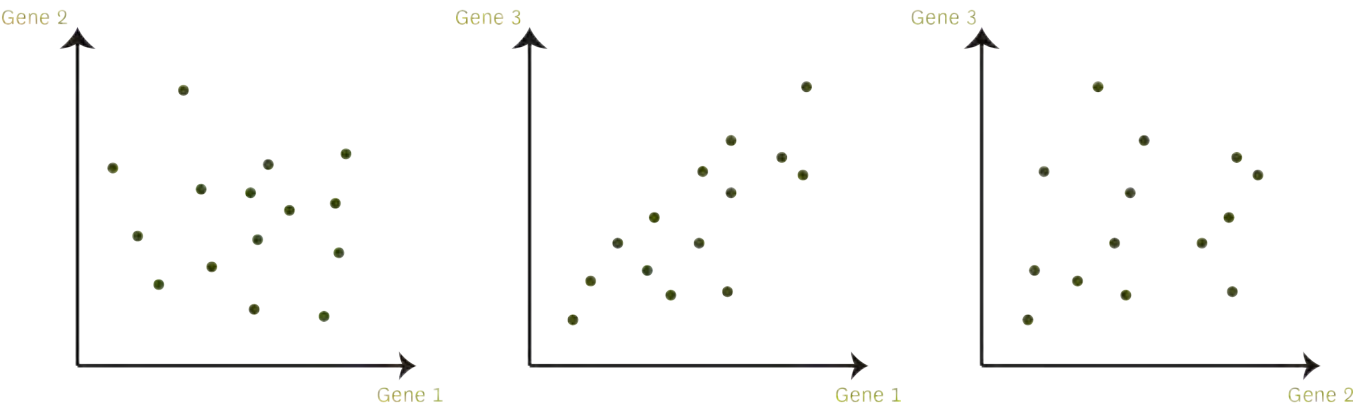
Beyond Correlation
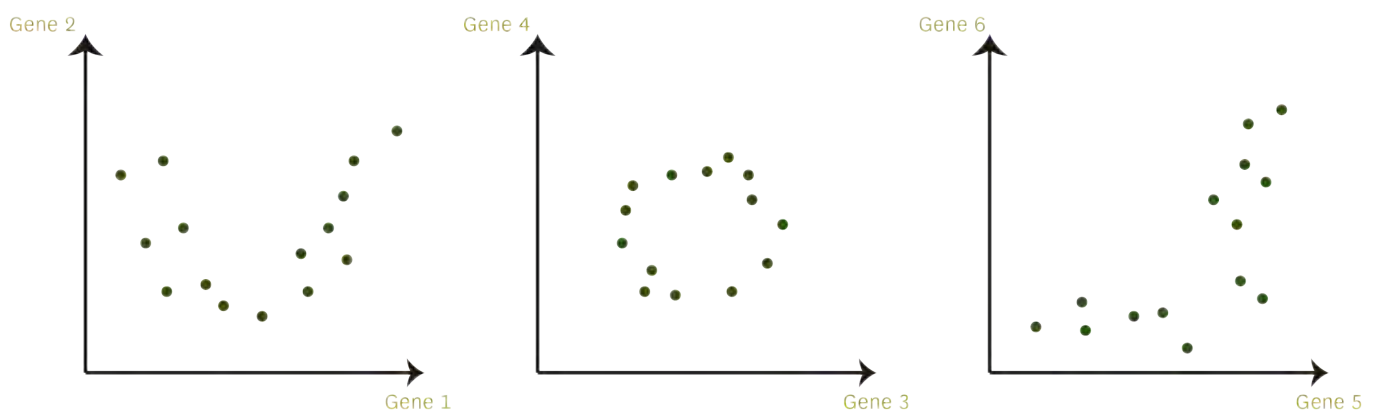
Correlation is Not Enough

- Correlation between g_1 and g_2 only measures *linear dependence*.
- Correlation is a good measure for *independence* ($C = 0$ is reliable) but not *dependence* ($C \neq 0$ is not reliable).
- Correlation cannot distinguish between direct and indirect interactions.

Information Theoretical Network Inference
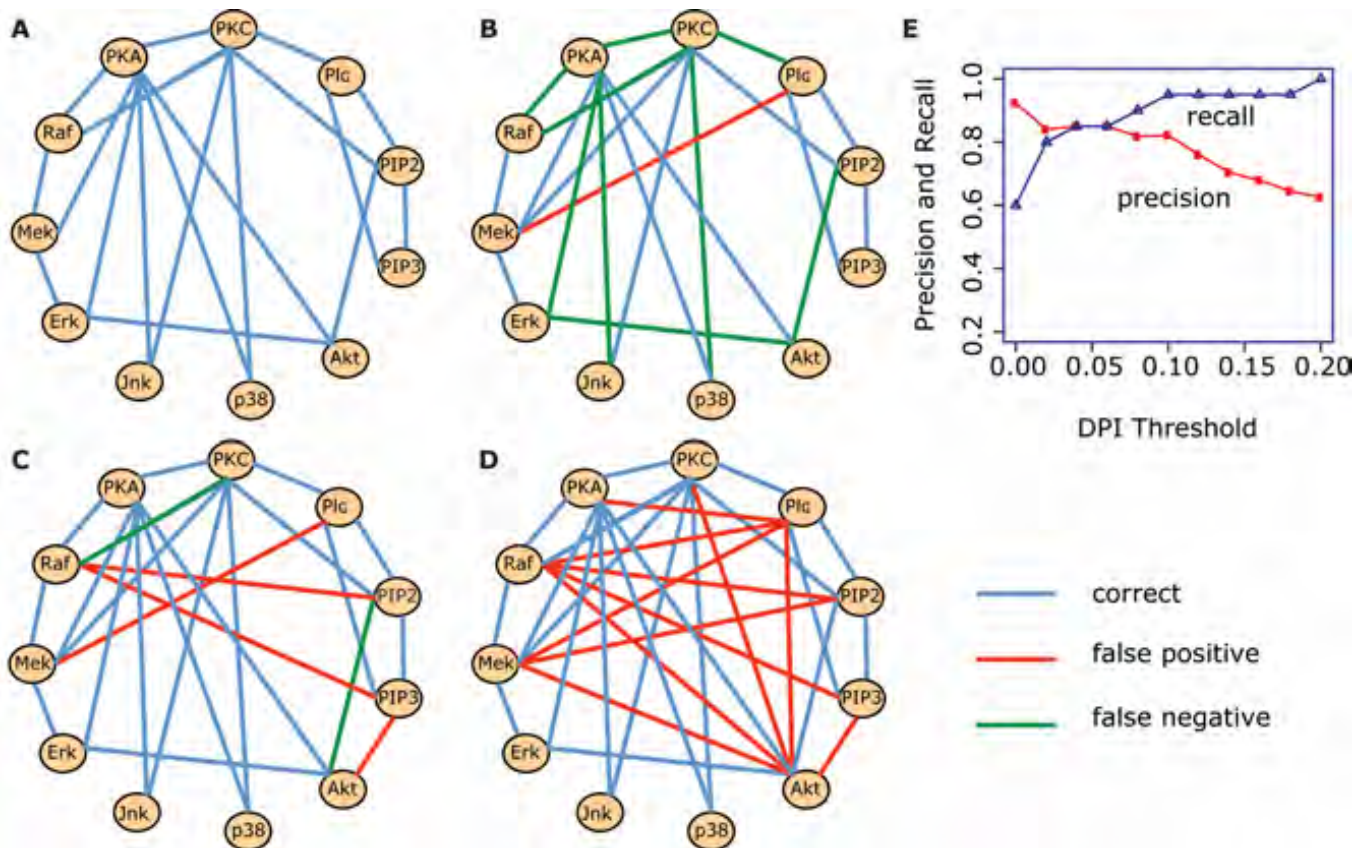
This approach can deal with non-linear interactions; distinguish between direct and indirect interactions; and provide clues as to underlying mechanisms (it can *imply causation*).

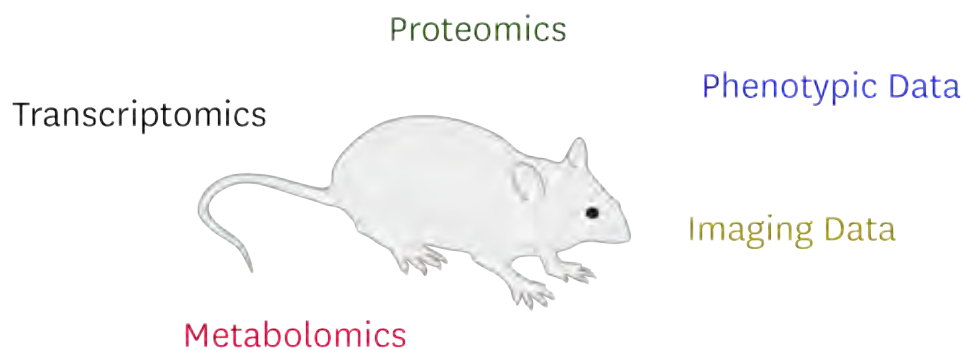# Data Fusion

Data Integration: Making all relevant data available in a convenient, computer-readable form.

Data Fusion: Putting all data into a single analysis/modelling/inference framework.

Proteomics

Phenotypic Data

Transcriptomics
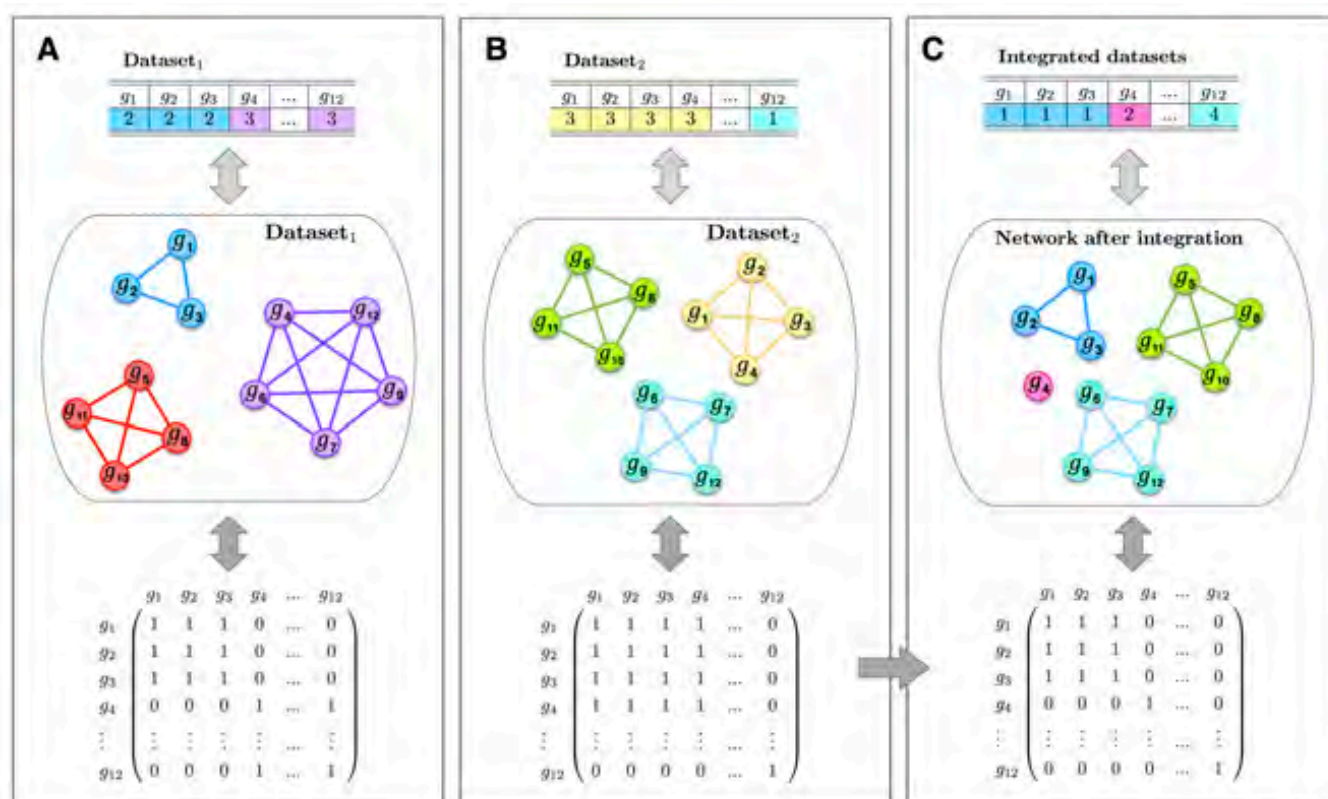
Imaging Data

Metabolomics

Model-Based
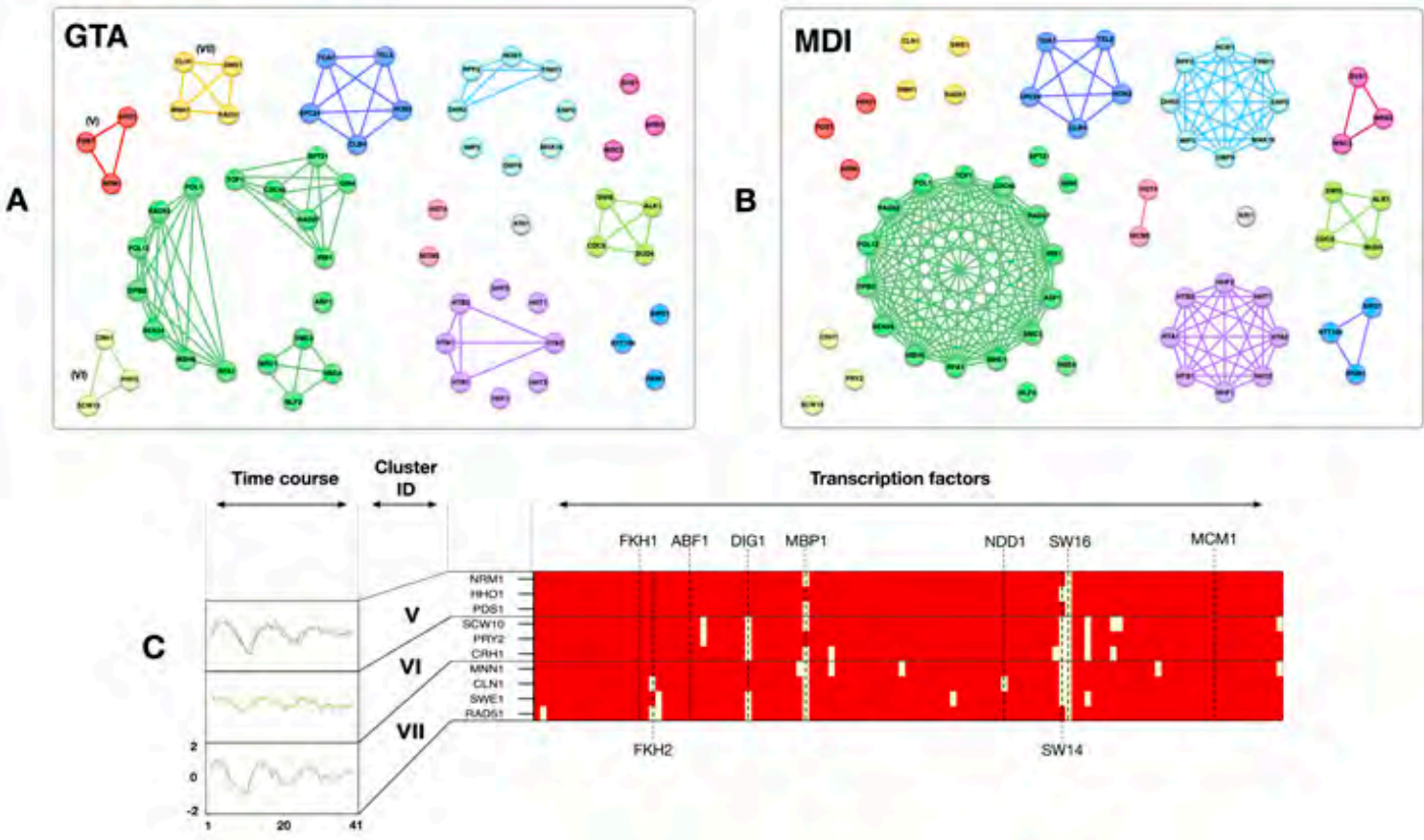Requires a mathematical model.

Data-Driven
Relies solely/mostly on the data.

Graph-Theoretical Data Fusion
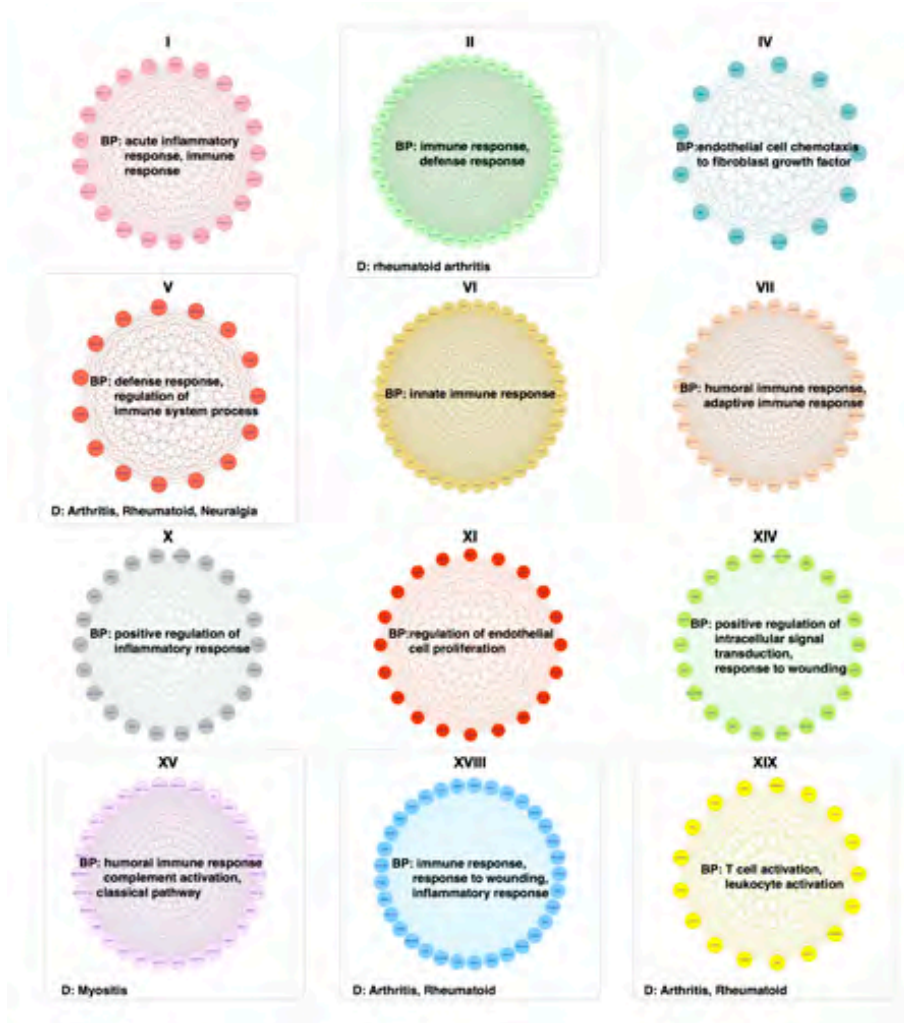

This approach tries to identify the modules that are in agreement with all the different experimental datasets.

Graph-Theoretical Data Fusion

Data Fusion for sIBM

Scope and Extensions

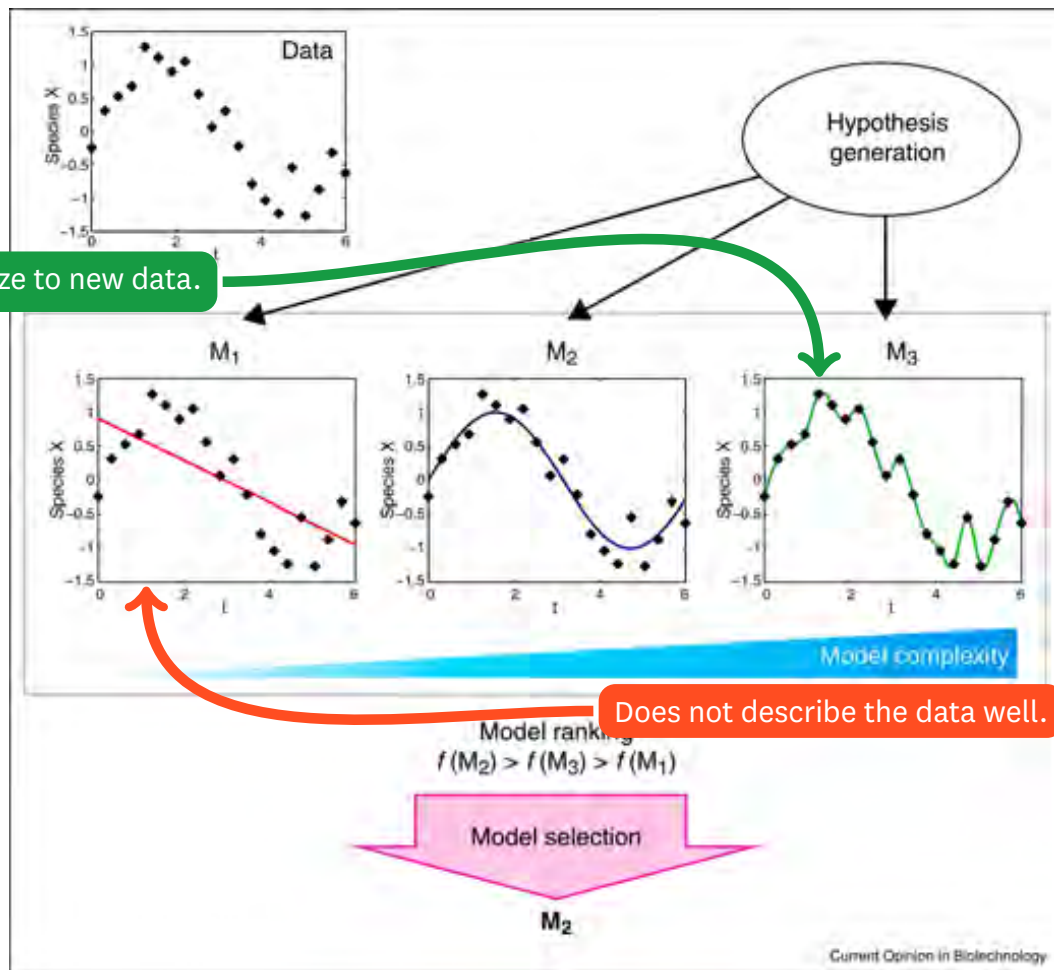- GTA is cheap, flexible and suited for *on-the-fly* (genome-scale) hypothesis-generation.

- The aim is to **distill motifs or compact sets of genes** for further analysis (or model development).

Mechanistic Modelling

substrate products active site regulator sites gate
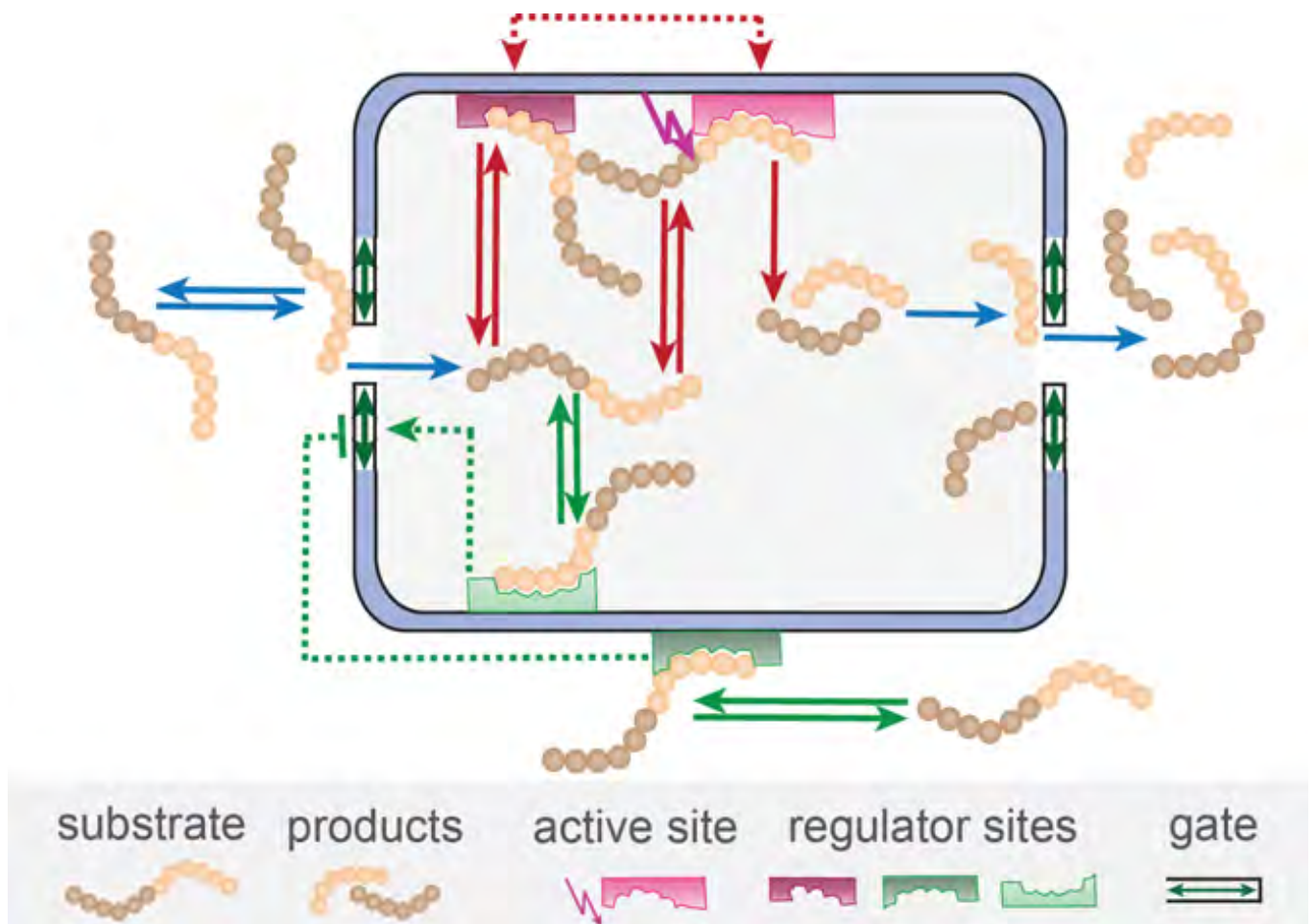
Liepe et al., *Nature Protocols*, 2014; Liepe et al., *eLife*, 2015.

Model Selection
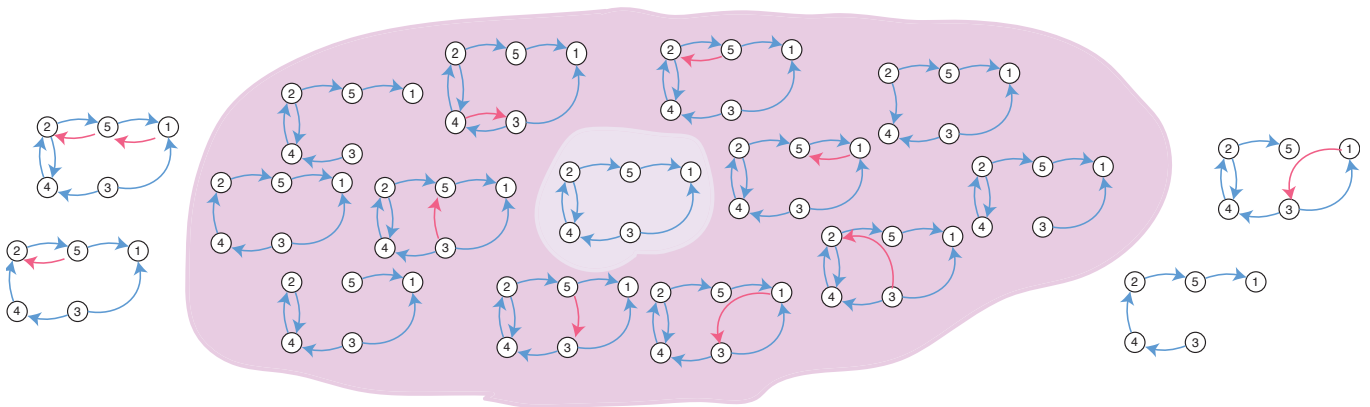

Kirk et al., *Curr.Opin.Biotech.*, 2013, **24**:767-774.

Model Neighbourhoods

We start from a seed-model, M_0 and define a set of models

$$\mathcal{M} = \{M_0, M_1, \ldots, M_n\}$$

that are structurally or functionally related to M_0.

We then compare all of these models in light of the data.

The main problem is that the number of models is enormous.

Babtie et al., PNAS (2014).

How Will it Work

- For *network inference* we can help relatively straightforwardly for both transcriptomic and proteomic data.

- *Data-fusion* is best done in close collaboration with the domain experts — you and your RAs.

- *Mechanistic modelling* relies on close collaboration and rapid as well as frequent iteration between experiment and theory from the conception of experiments all the way to the analysis and validation.

- Please feel free to visit us at Imperial College, or have your RAs visit us in London: for any meaningful analysis nothing substitutes for personal links.

- Equally, the Imperial RA and other people associated with T2D will be interested in visiting your groups to help with experimental design and analysis.

- In the most successful collaborations we had collaborating RAs visit us to become familiar with basic programming, data processing etc , and we would be happy to do this for T2D as well.