Imperial College London

Modelling Methods: High Throughput Time-Course Data

Jaroslav Stark
Centre for Integrative Systems Biology at Imperial College (CISBIC) and Department of Mathematics

Overview

- Dynamic v Static
- Forward and Inverse Systems Biology
- Model Identifiability
- Example: p53 Activity

Overview

- **Dynamic v Static**
- **Forward and Inverse Systems Biology**
- **Model Identifiability**
- **Example: p53 Activity**

Perturbations and Responses

- **To Understand Biological Systems We Need To:**
  - **Perturb the System**
  - **Observe Resulting Effect**

Static v Dynamic

- **Static Perturbation**
  - ***eg*** **Gene Knockout**
  - **Measure Long Term Effect**
- **Dynamic Perturbation**
  - ***eg*** **Immune Activation by Pathogen**
  - **Measure Dynamic Response**

Time Course Data

- **To Understand Dynamic Response We Need Time-Course Data**
- **Increasingly Can Obtain Many Simultaneous Measurements of Different Variables**
  - ***eg* Microarray Time-Coures**

Time Series

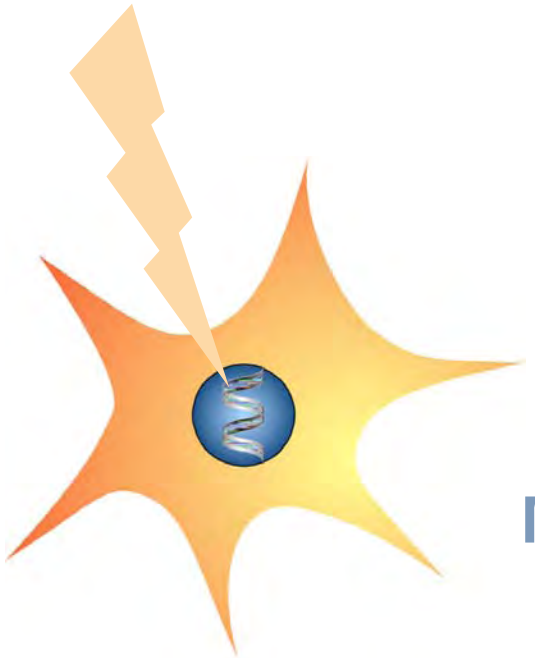
- **Vast Body of Methods:**
  - **Signal Processing**
  - **Statistics**
  - **Nonlinear Dynamics**
- **Systems Biology Presents a New Challenge**
  - **Very Few Time Points**
  - **Many Simultaneous Variables**

High-Throughput Time-Course Data

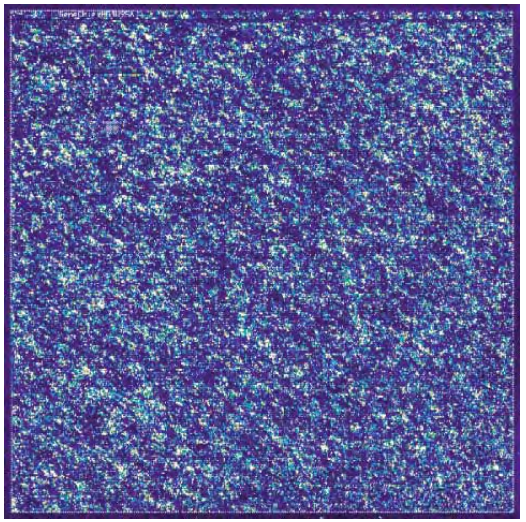
- **Methods in Use**
  - **Very Simplistic By Time Series Standards**
- **Traditional Time-Series Methods**
  - **Typically Do Not Work With "Short Fat" Data**

Example: DNA Damage Time-Course

Irradiate Cells and Take Microarray Time Series



Multiple Time Points

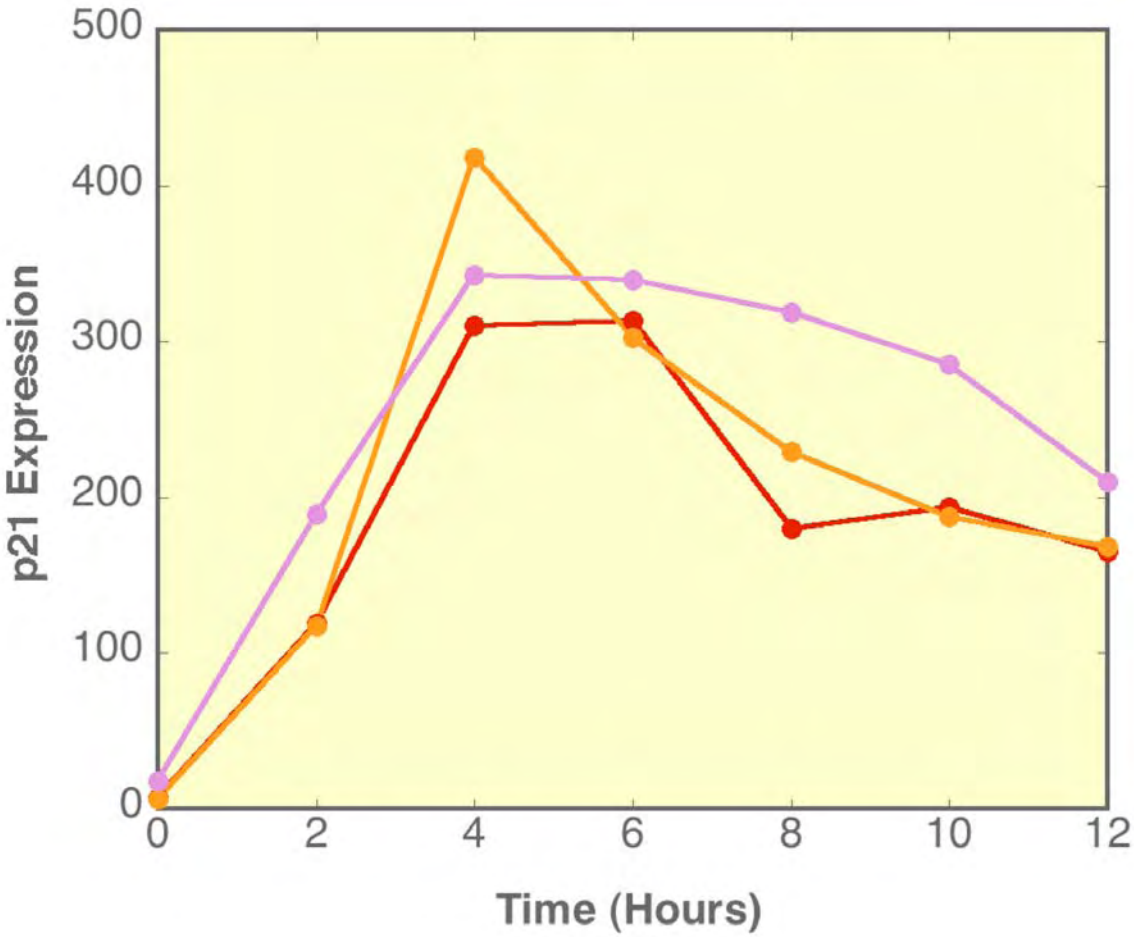


CISBIC

Imperial College London

Measured Data For One Gene (p21)

- **Three Replicates**



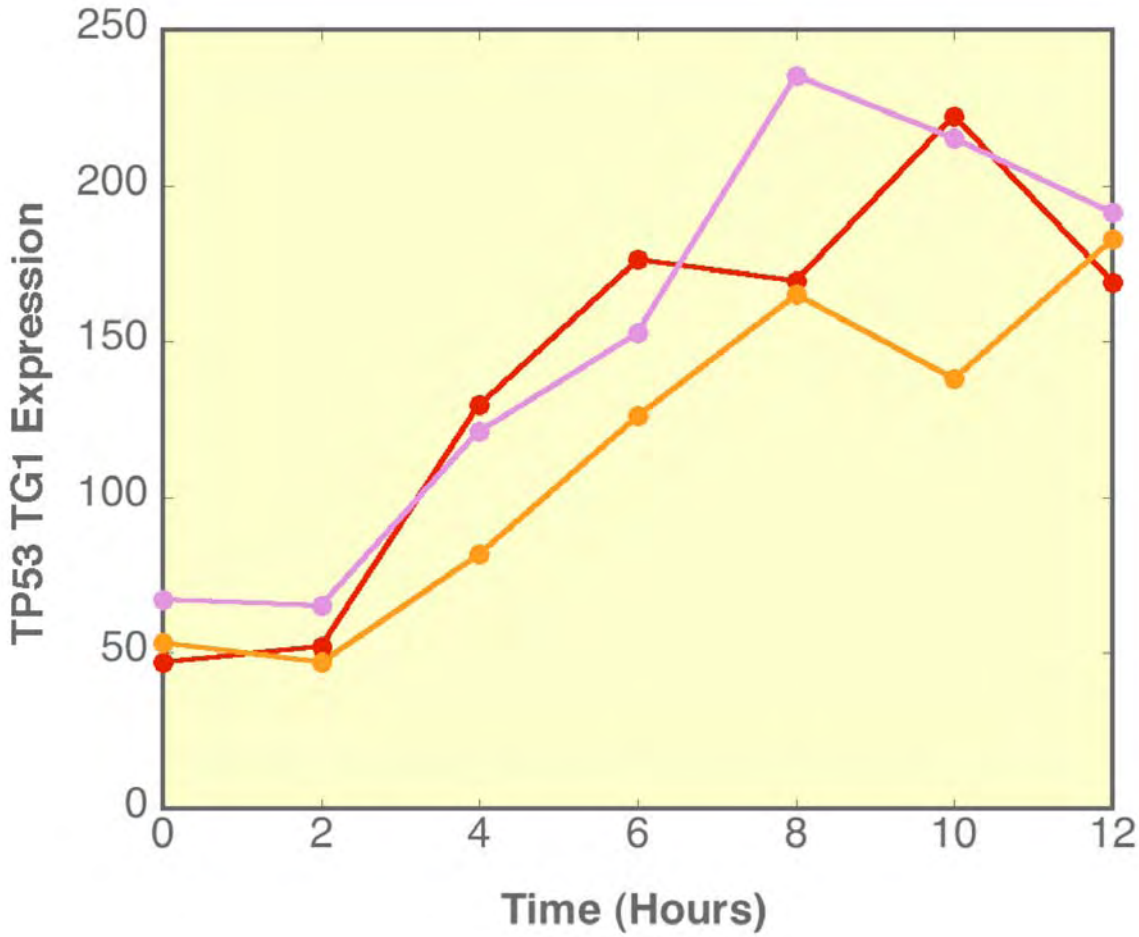
Barenco *et al*, *Genome Biology* 06

CISBIC

Imperial College London

We Measure 19,000 Genes Simultaneously

- **Different Genes Have Different Dynamics**



CISBIC

Imperial College London

Usual Analysis

- Clustering or Principal Component Analysis
- Basic Principle: Group Genes With Similar Patterns of Activity

Usual Analysis

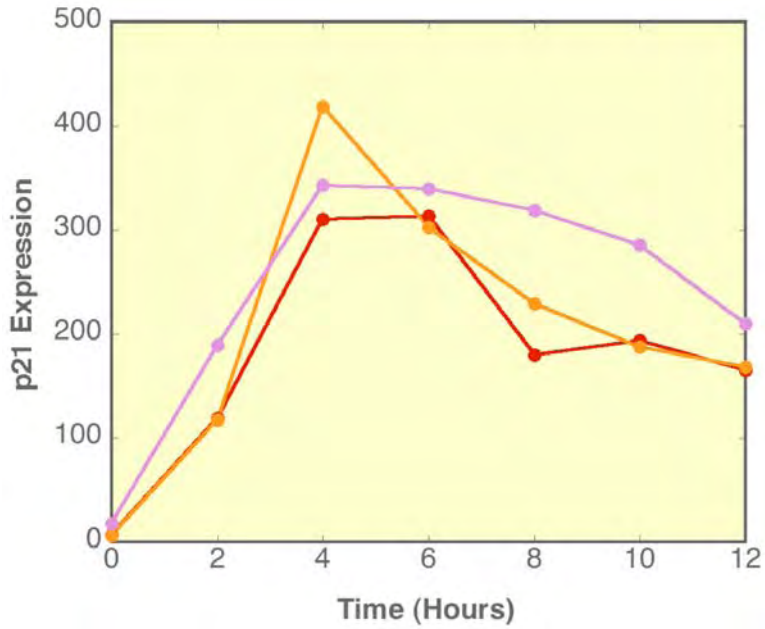
- Clustering or Principal Component Analysis
- Basic Principle: Group Genes With Similar Patterns of Activity
- Key Assumption

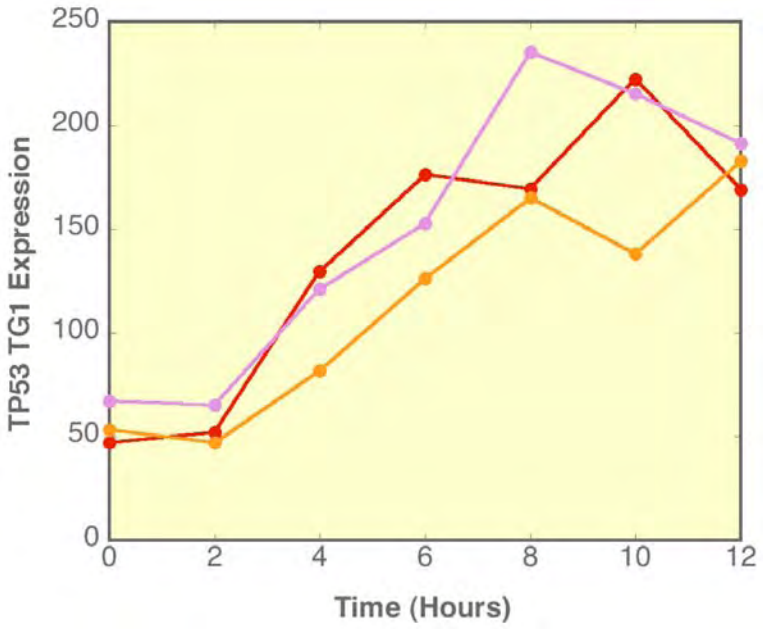
Similar Pattern = Similar Regulation

Different Pattern = Different Regulation

But …

- **p21 and TP53 TG1 Are Both p53 Targets**



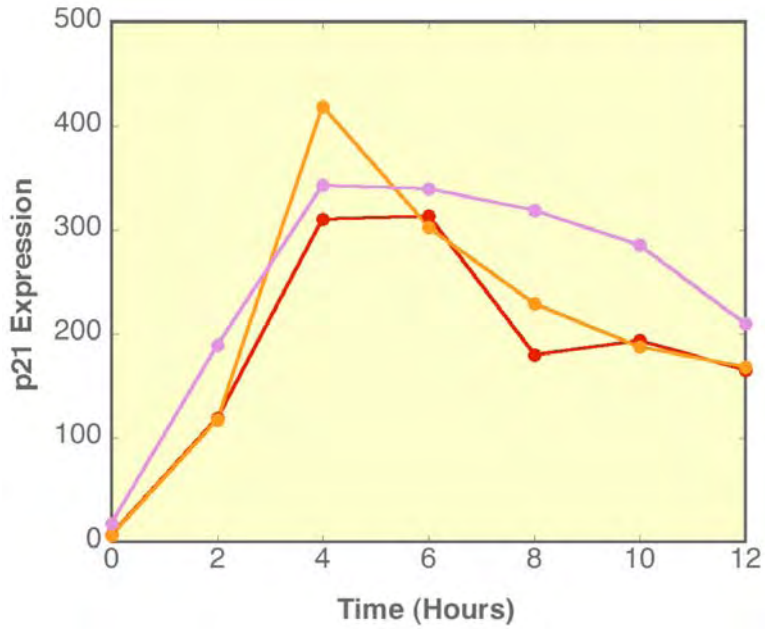


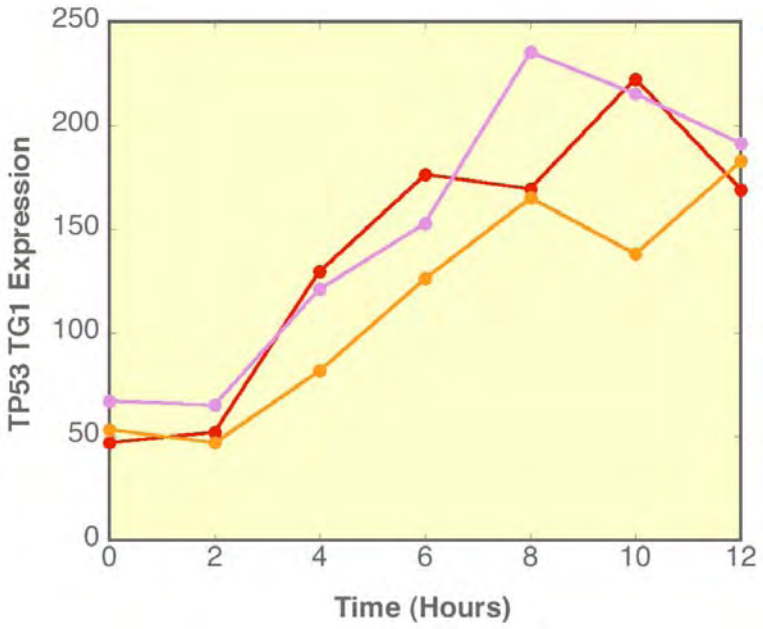
CISBIC

Imperial College London

Explanation

- **Similar Production, Different Degradation**





CISBIC

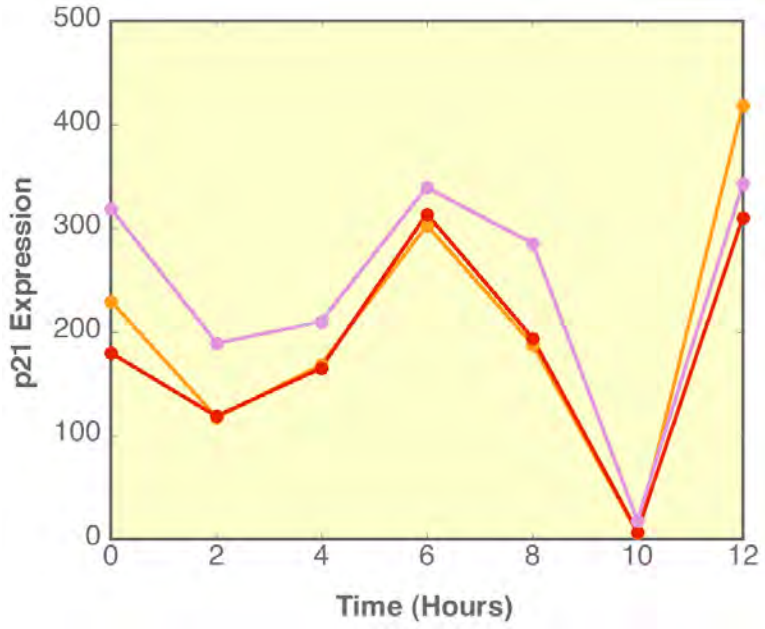
Imperial College London

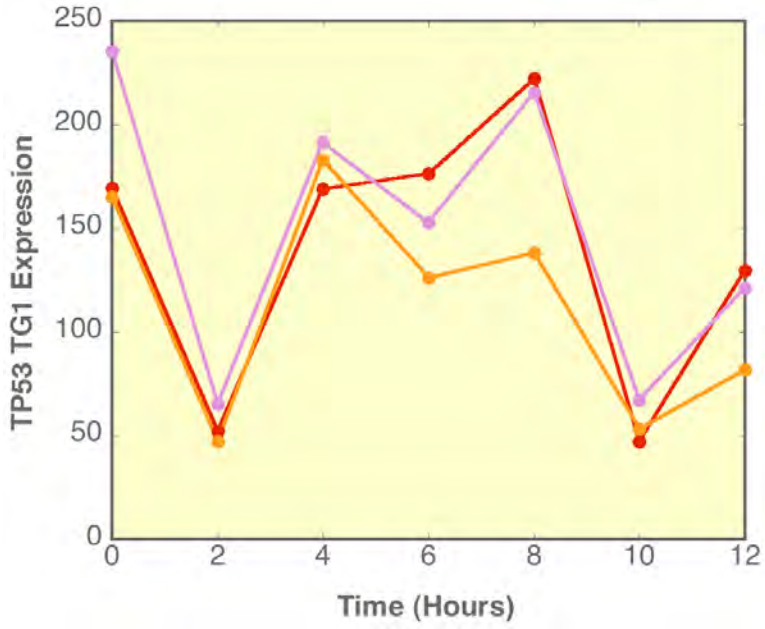
Usual Analysis: Major Flaw

- **Clustering or Principal Component Analysis**
  Discards Important Information
- **If We Scramble Data in Time**
  - **Recover Same Groups**

Permuted Time Labels

- **Different Patterns, Same Groups**





CISBIC

Imperial College London

Usual Analysis: Major Flaw

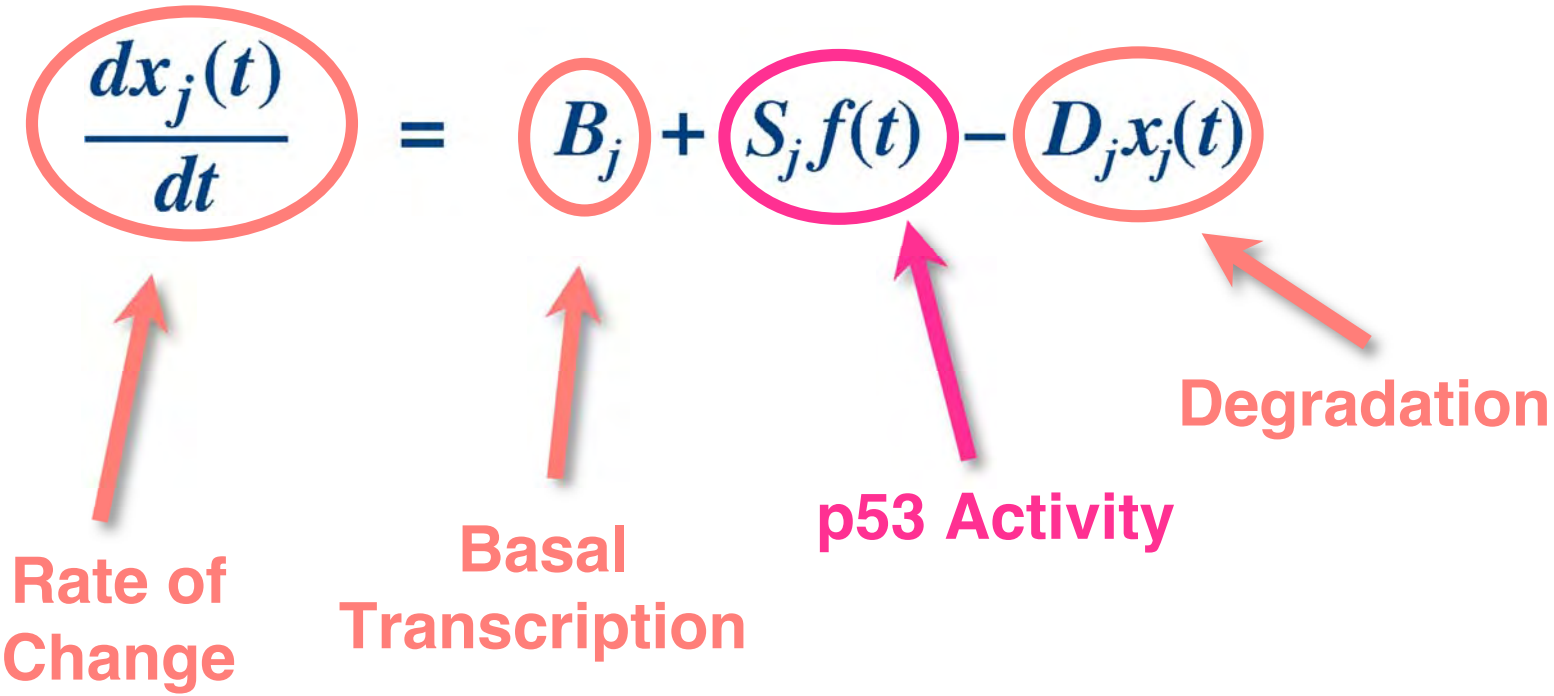
- Clustering or Principal Component Analysis Discards Important Information
- If We Scramble Data in Time
  - Recover Same Groups
- **Standard Analysis** Ignores Temporal Relationships

Better Analysis

- Need Models That Explicitly Incorporate Time
- Data Is Very Crude
  - Use Very Simple Models

Example: p53 Model

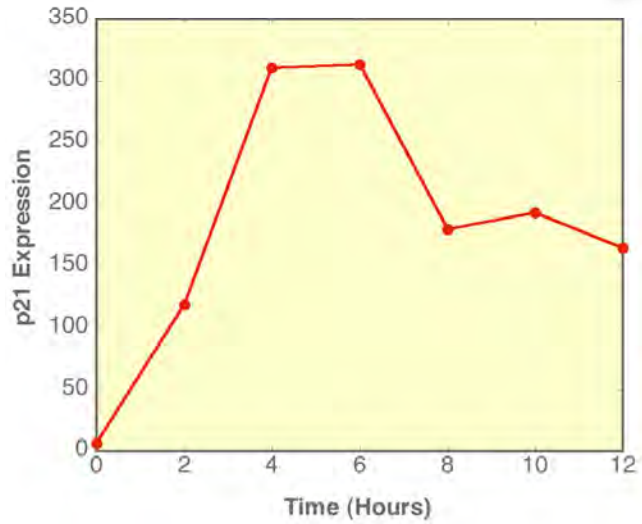
- **Ordinary Differential Equation for Expression Level $x_j(t)$ of Gene j**



Problem: p53 Activity Is Not Measured

- **Microarray Only Measures Expression** $x_j(t)$

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$



CISBIC

Imperial College London

Problem: p53 Activity Is Not Measured

- p53 Activity Not Directly Controlled By Transcription
- p53 Activity Controlled By
  - Phosphorylation
  - Tetramerization
  - Ubiquitination
  - Translocation to Nucleus
- This Will Not Show Up in Microarray Data

Overview

- Dynamic v Static
- **Forward and Inverse Systems Biology**
- Model Identifiability
- Example: p53 Activity

Paradox of Molecular Biology

- **It Possesses Vast Amounts of Information About Individual Molecules**
- **It Needs to Integrate This Into An Understanding of the Whole System**

Systems Biology: Forward Approach

- **Integration of Known Components**
  - **High Throughput Experiments**
    - **Measure All Potential Components**
    - **Measure All Potential Interactions**
  - **Modelling**
    - **Characterize Common Patterns**
    - **Suggest Functional Roles**
    - **“Understand System”**

Paradox of Molecular Biology

- It Possesses Vast Amounts of Information About Individual Molecules
- It Needs to Integrate This Into An Understanding of the Whole System
- **Many Potential Experiments or Treatments Are Difficult or Impossible**
  - **Combinatorics of Multiple Gene Knockouts**
  - **Dose-Time Course Experiments**

Three Approaches

- **Pretend Data is Correct and Complete**
- **Wait Until Data is Correct and Complete**
- **Develop Modelling Methods to Overcome Limitations**

Systems Biology: Inverse Approach

- **Discovery of Hidden Effects**
  - **Models Based on Incomplete Information**
  - **Comparison to Incomplete Data**
  - **Prediction of New Components or Interactions**

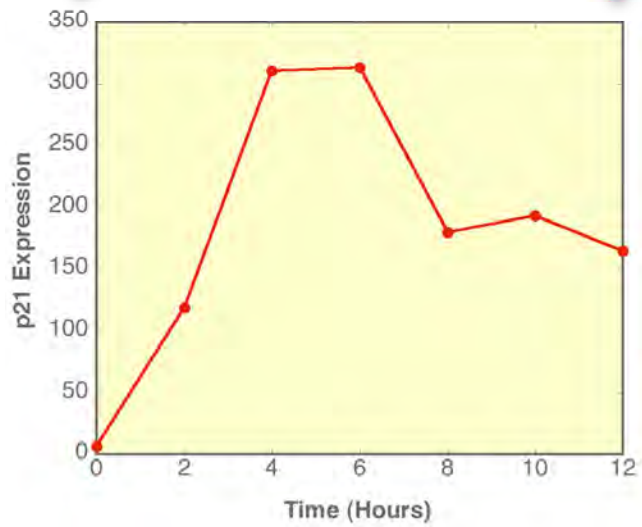
Systems Biology: Inverse Approach

- Discovery of Hidden Effects
  - Models Based on Incomplete Information
  - Comparison to Incomplete Data
  - Prediction of New Components or Interactions
- **Modelling is Useful With Limited Data**

Estimate p53 Activity

- **Use 5 Known Targets to Fit This Equation**

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$



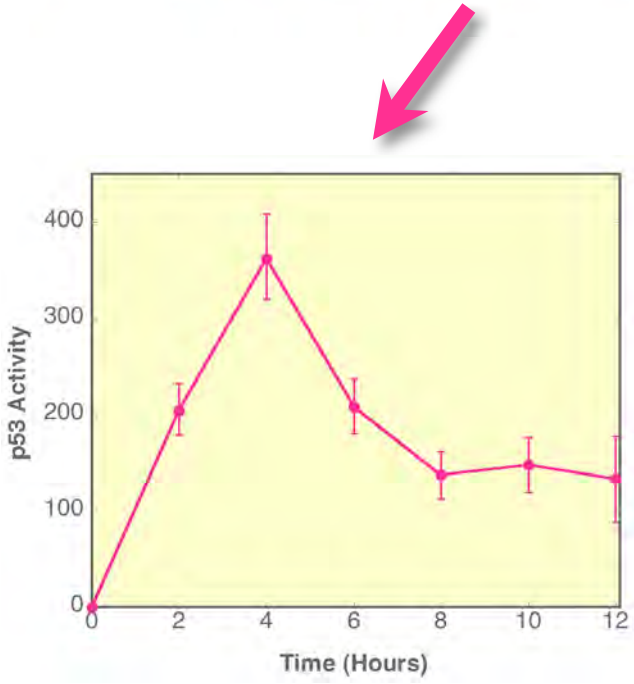
CISBIC

Imperial College London

Estimate p53 Activity

- **Estimate Unknown p53 Profile**

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$



CISBIC

Imperial College London

Overview

- Dynamic v Static
- Forward and Inverse Systems Biology
- **Model Identifiability**
- Example: p53 Activity

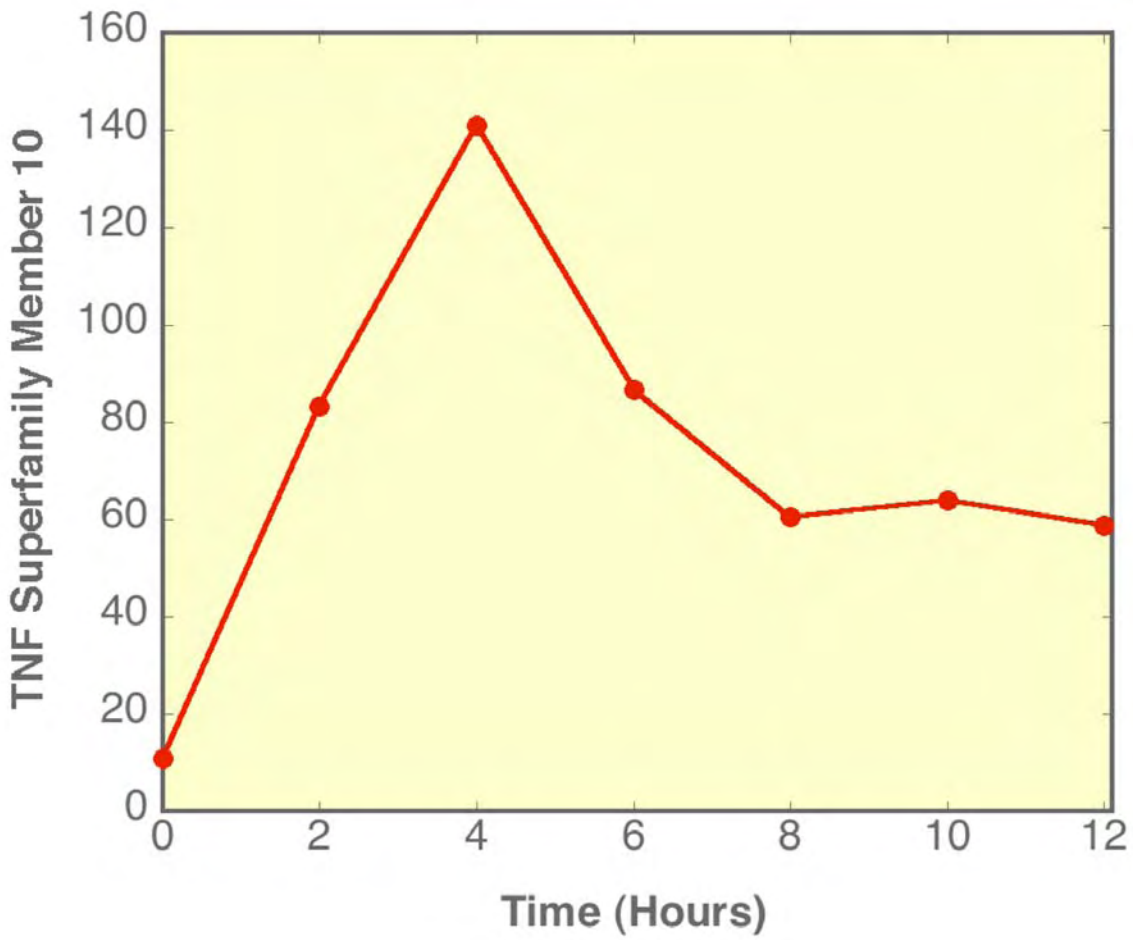
p53 Model: Problem 2

- **Run Model at Two Different Parameter Combinations**

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$

p53 Model: Problem 2

- **Parameter Combination 1**

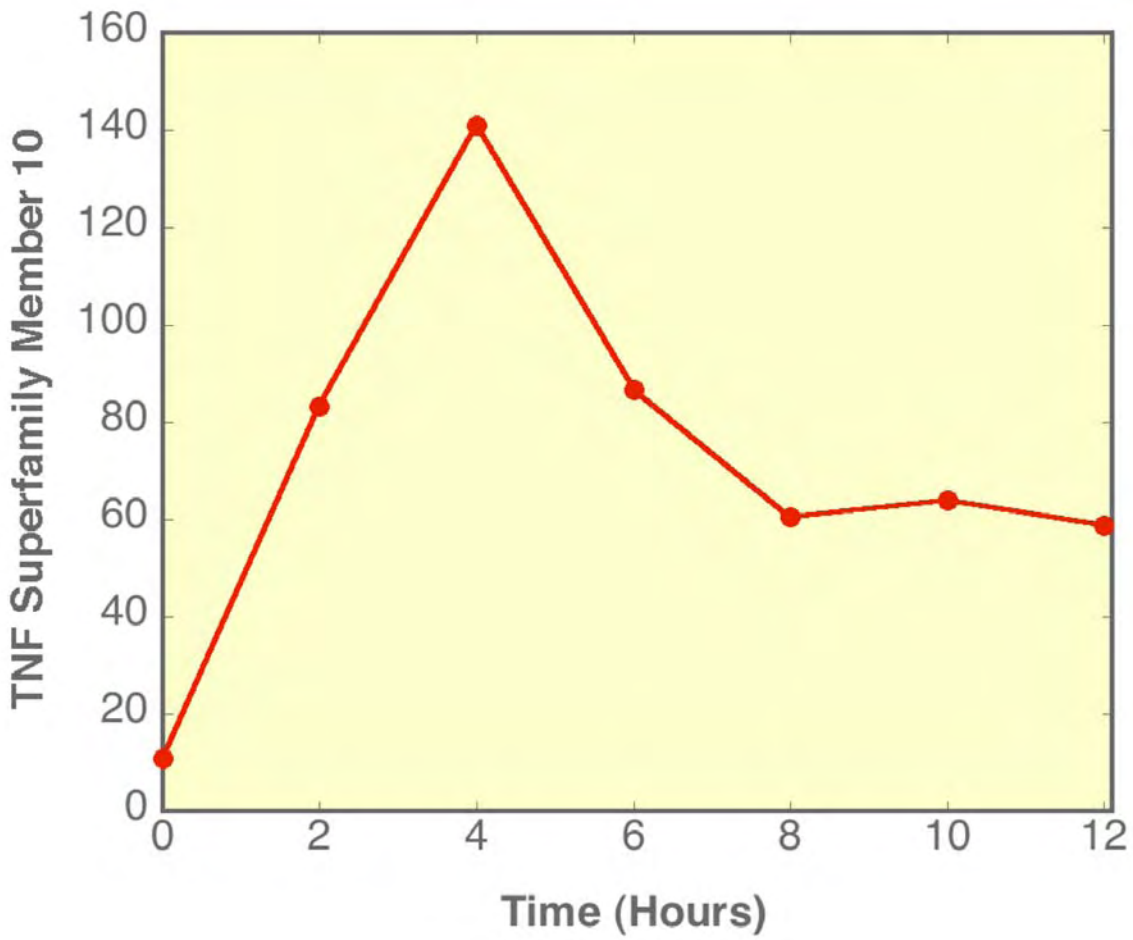


CISBIC

Imperial College London

p53 Model: Problem 2

- **Parameter Combination 2**



CISBIC

Imperial College London

p53 Model: Problem 2

- Different **Parameter Combinations Give Same Solution**

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$

Explanation

- **Trade-Off Between Basal and Induced Activity**

$$\frac{dx_j(t)}{dt} = B_j + S_j f(t) - D_j x_j(t)$$

Explanation

- **Trade-Off Between Basal and Induced Activity**

$$B_j + S_j f(t)$$

Explanation

- **Trade-Off Between Basal and Induced Activity**

$$(B_j + 1) + S_j f(t)$$

Explanation

- **Trade-Off Between Basal and Induced Activity**

$$(B_j + 1) + S_j\,(f(t) - 1/S_j)$$

Model Non-Identifiability

- **Well Known Problem in Statistics and Control Theory**
- **Sontag, *J Nonlin Sci*, 12, (2002), 553-583.**
  - **Classic Operon Model**
- **Can Be Overcome by Careful Choices of Normalization**
  - **Reduces Number of Parameters to Be Estimated**
  - **Tutorial**