Tempest Technologies

Mathematical Modeling and Statistical Data Analysis: An Overview of Tempest Technologies' Capabilities

Prepared for Pasadena Bioscience Collaborative

25 September 2013

Ben Fitzpatrick, PhD
President
Kate Angelis, PhD
Analyst

Presentation Outline

- ***About Tempest Technologies***
- Tempest’s Service Offerings
- Track Record
- Mathematics and Biotechnology

Tempest Technologies (Tempest) *is a professional services firm devoted to advancing the objectives of government, non-governmental organizations, and private industry clients through quantitative research methods, mathematical and statistical modeling and analysis, and software engineering and development.*

Formed in 1998, Tempest is dedicated to meeting the growing needs of companies across the technological and scientific sectors

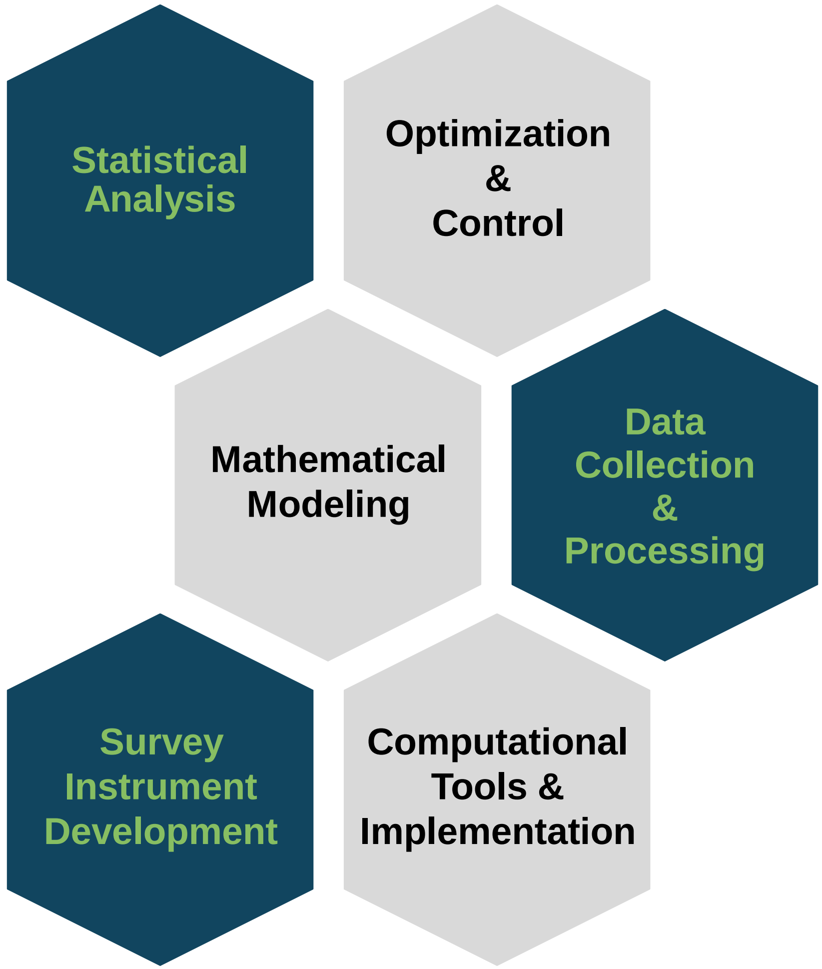
- Our ***focus*** is on delivering solutions to client problems using the right technologies for the project at hand to achieve globally competitive results and success
- We have deep ***expertise*** in mathematical and statistical modeling as well as a nuanced understanding of the health and life sciences landscape
- Our ***value*** is derived from our experienced staff of mathematicians, engineers, social and health scientists that offer diverse and wide-ranging experience in areas from survey instrument development and model development to data collection and processing, and computational implementation

Presentation Outline

- About Tempest Technologies
- ***Tempest's Service Offerings***
- Track Record
- Mathematics and Biotechnology

Tempest’s Service Offerings

Tempest’s unique skill set has been applied to tackle diverse projects; from physics- and engineering-related problems of mechanical vibration to social and public health modeling.



Core Competencies

Statistical Analysis

Experimental design, exploratory data analysis and descriptive statistics
Multivariate analysis, ANOVA, linear models, nonlinear models, longitudinal analysis
Time series analysis: model fitting and selection, time and frequency domain analysis
Information extraction: clustering, principal components, pattern recognition

Optimization & Control

Linear controls in the time and frequency domains
Real-time implementation and digital control
Optimal control, game theory and adversarial optimization

Mathematical Modeling

Systems modeling: compartmental models, epidemiological models, PBPK models
Distributed parameter models: electro-mechanical, fluids, diffusion and dispersion
Agent-based modeling

Data Collection & Processing

Database development: relational, hierarchical, networks and object-oriented databases
Qualitative research: coding and analysis
Sensor selection, hardware integration, data collection and storage

Survey Instrument Development

Survey instrument design
Survey data collection: web-based, telephone surveys and in-person surveys

Computational Tools & Implementation

Programming languages: C, C++, Java, Perl, python, Fortran, Visual Basic
Statistical analysis and programming environments: SPSS, R, SAS, Stata
High level computational environments: LabVIEW, MATLAB, Simulink, IDL.
Database tools: MySQL, PostgreSQL, MS Access

Presentation Outline

- About Tempest Technologies
- Tempest’s Service Offerings
- ***Track Record***
- Mathematics and Biotechnology

Track Record

We have had the opportunity to collaborate with diverse clients – acting as both an advisor and a partner – to achieve tangible results.

LSU Health New Orleans

DARPA

ARA expanding the realm of POSSIBILITY®

NIH National Institutes of Health
Turning Discovery Into Health

TREX ENTERPRISES CORPORATION

NATIONAL INSTITUTES OF HEALTH
NIAAA | NATIONAL INSTITUTE ON ALCOHOL ABUSE AND ALCOHOLISM
Understanding the impact of alcohol on human health and well-being

NORTHROP GRUMMAN

BOEING®

AIR FORCE RESEARCH LABORATORY

DRAPER LABORATORY

Success lies in leveraging our mathematics, engineering and scientific knowledge and expertise to develop innovative solutions that meet the unique needs of clients operating in increasingly complex environments.

Application Domains

Our specialized core competencies have been applied to a variety of projects across multiple sectors.

- *Population modeling, statistical data analysis, model tuning, and optimization in support of insurgency interdiction and influence operations*
- *Understanding social networks, social norms, and marketing impacts on drinking dynamics in college communities for the reduction of alcohol-related problems on campuses*
- *Inferring network structure from communications data for terrorist network disruption*
- *Developing optimal sensing plans for complex teams of recon troops and unmanned air vehicles in support of ground troop operations*
- *Gene expression network modeling for the understanding of stress response*
- *Real-time beam control for high energy laser tracking and weapon systems*
- *Ladar image processing and target extraction for smart munitions*
- *Modeling and statistical classification for an electromechanical arthritis probe*
- *Survival analysis for patients with Budd-Chiari Syndrome*

Based on our capabilities and extensive experience in applying mathematical tools to solve problems across engineering and defense disciplines, we are excited to strengthen our presence in the health and life sciences space.

- Bringing together expertise in mathematics, computation and the health and life sciences, Tempest is equipped to solve problems within the broader framework of public health
- In the "Big Data" era, where pharmaceutical and biotechnology players are dealing with massive amounts of data, it is critical that the appropriate analytical tools are utilized
- Tempest brings the necessary skills and critical insights related to proper data analysis:
  - Identifying appropriate analytic procedures
  - Drawing unbiased inference
  - Statistical significance
  - Data recording and presentation

"The healthcare sector would appear an obvious candidate for the use of computer modeling and for simulation to be employed as a valuable tool to provide sound evidence."

"Toward a New Comprehensive International Health and Health Care Policy Decision Support Tool." OECD, 2012

Tempest has the practical experience in data analysis, mathematical modeling and simulation to tackle your organization's most complex problems.

Presentation Outline

- About Tempest Technologies
- Tempest’s Service Offerings
- Track Record
- ***Mathematics and Biotechnology***

Modeling: What Is It?

- Modeling: an attempt to understand something by creating a likeness
  - Architectural scale model in cardboard of a skyscraper
  - Map of Pasadena
  - Mathematical equations describing gravity
- A model is a simplified representation of reality
- Models can be mental, verbal, graphical, physical, or mathematical
- Models can contain various levels of detail and provide various levels of accuracy
- Models may contain hypotheses about phenomena
  - In which case the model’s predictions provide a means of judging the validity of the hypothesis

Modeling: Why Do It?

- You already do:
  - Project-making requires a model
  - Often this model is implicit, without tested consistency or faithfulness to data, whose full consequences may be unknown
- Explicit models lead us to
  - Detail our assumptions
  - Examine their consequences
  - Integrate expertise across disciplines
- Statistical models provide exploratory and inferential structure
  - Search for similarities, patterns, relationships
  - Test for treatment improvements
- Mechanistic modeling provides additional tools
  - Extrapolation and projection
  - Decision and control

Modeling Issues: Objectives and Tools

- *The goal governs the approach*
  – Qualitative understanding
  – Quantitative understanding
  – Quantitative prediction
  – Design/decision making
  – Automated control/decisions

- *Empirical statistical approach*
  – Collect data
  – Fit a statistical model
    - Linear model, hierarchical linear model, multivariate/covariance, etc.
  – Test hypotheses
  – How to extrapolate?

- *The phenomena govern the approach*
  – Individual differences/similarities
  – Environmental effects
  – Dynamics
  – External/exogenous influences

- *Theoretical mathematical approach*
  – Quantify behaviors
  – Develop equations
  – Solve/simulate
  – Test hypotheses

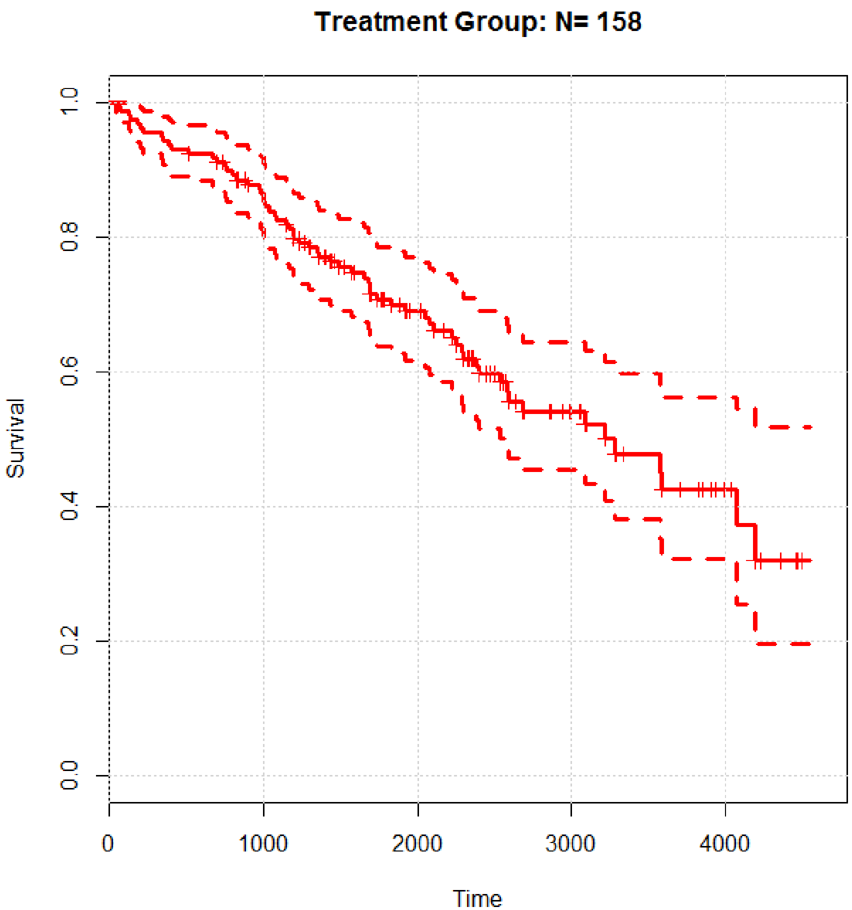
Modeling in the Health and Life Sciences

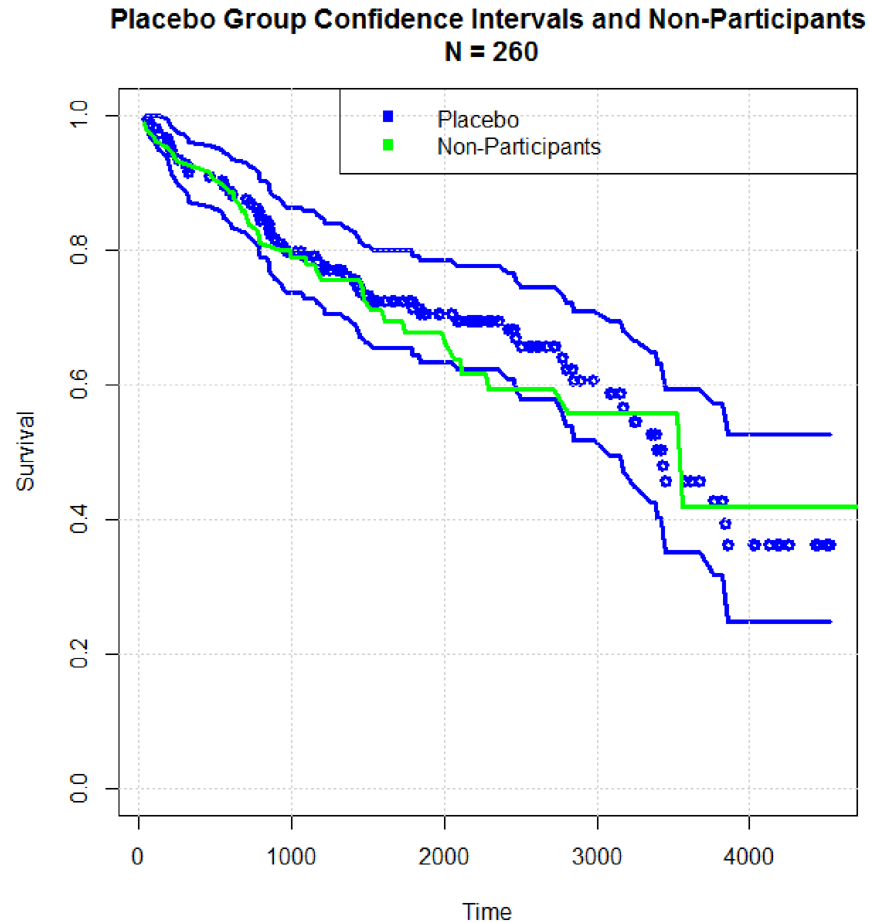
- ***Classical Statistical Problems***
  - Exploratory analysis: descriptives, clustering, pattern recognition, principal components
  - Clinical trials: sample size, effect size, power to detect true positives, significance tests
  - Survival analysis: survival and hazard rates, significance in treatments
  - Statistical analysis for publication and presentation
- ***Physiological, Biomedical, and Bioengineering***
  - System definition
  - Physical, chemical, biological phenomena quantification
  - Parameter identification
  - Risk assessment, treatment design
- ***Genomics, Bioinformatics and Drug Discovery***
  - Many common issues with classical statistics
  - High dimensionality of data creates new issues
  - Empirical statistical methods necessary to extract information from limited data
- ***NIH's Systems Science initiative***
  - Viewing public health in terms of complex systems
  - Spans the Institutes
- ***FDA's Quality by Design: Integrated and Systematic Development***
  - A priori risk analysis
  - Process analysis and control: feedback and feedforward

Survival Analysis of Primary Biliary Cirrhosis

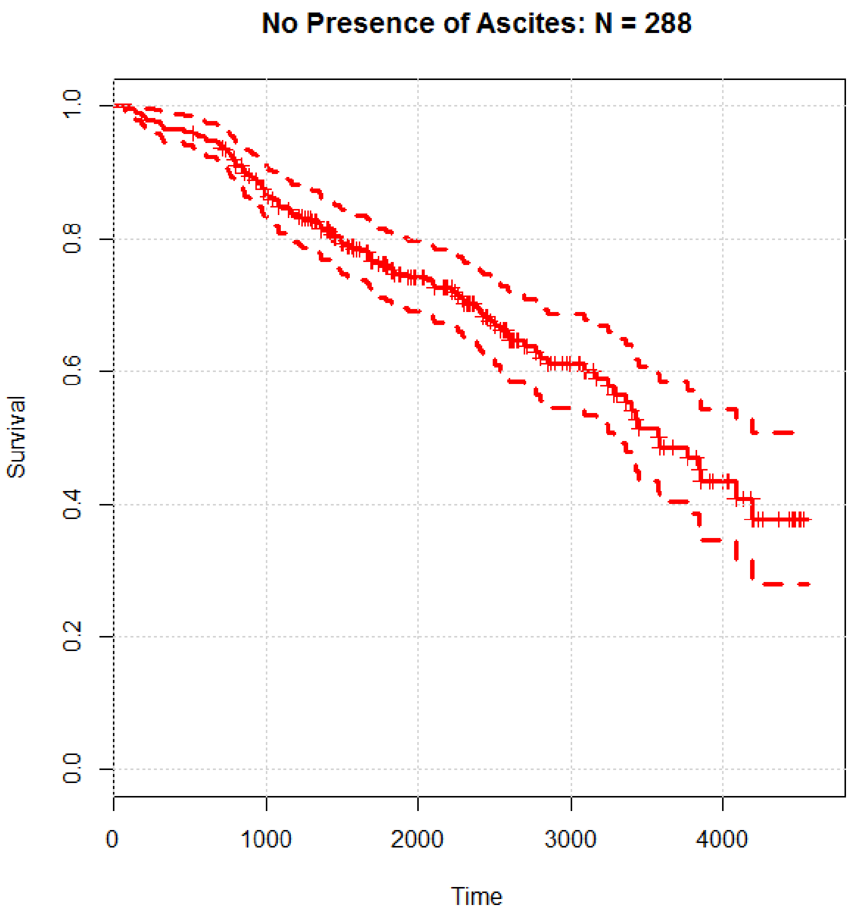
- ***PBC is an autoimmune disease of the liver***
  - Destruction of bile ducts in the liver
  - Bile builds up -> cirrhosis
  - Survival rates are low, treatment options few
- ***Control-treatment data from a Mayo Clinic study over 10 years***
  - Control = placebo
  - Treatment = D-penicillamine
  - Age, sex, edema (absent, present w/o diuretic or absent w, present in spite of diuretic), log bilirubin, log albumin, log prothrombin time, treatment, are the independent variables
    - Many other variable measured
  - Survival time is the independent variable
- ***Cox Proportional Hazard model is applied***
  - The hazard function is the marginal additional survival
  - Cox: log of the hazard rate is linear statistical model of the independent variables
    - “Proportional” bit: ratio of two patients removes the constant term, simplifies comps
  - Coefficients are estimated, tested for significance
- ***Sex, treatment were not significant to the model***
  - Differences can be seen, but variability is high

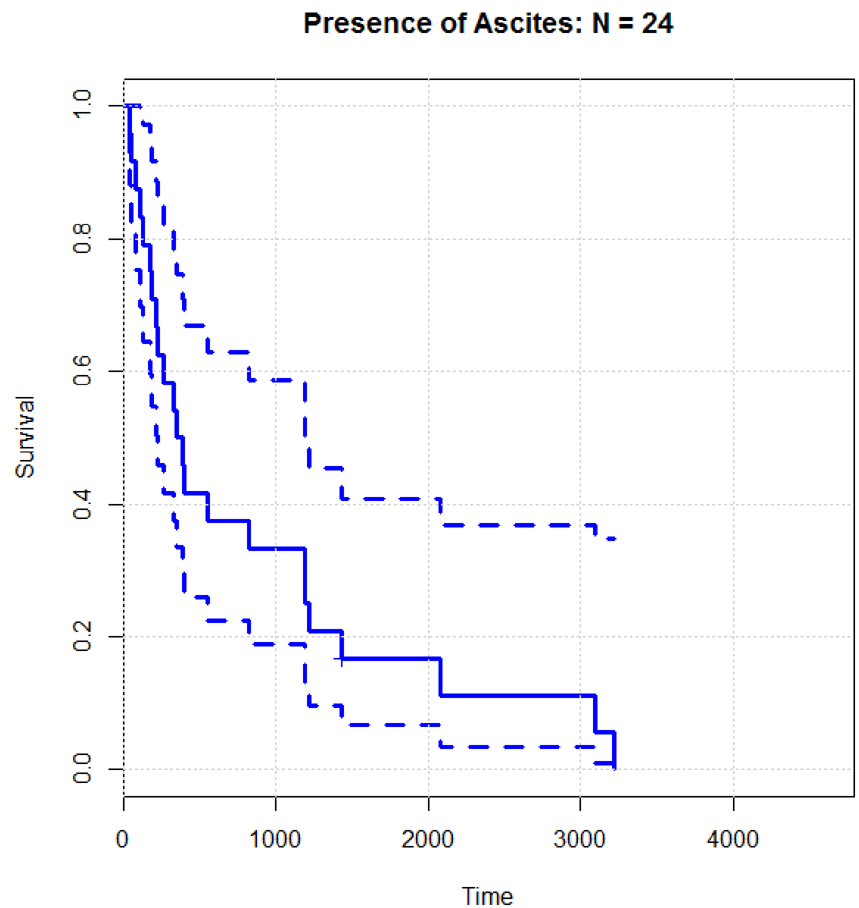
Treatment, Placebo, and Non-participant groups



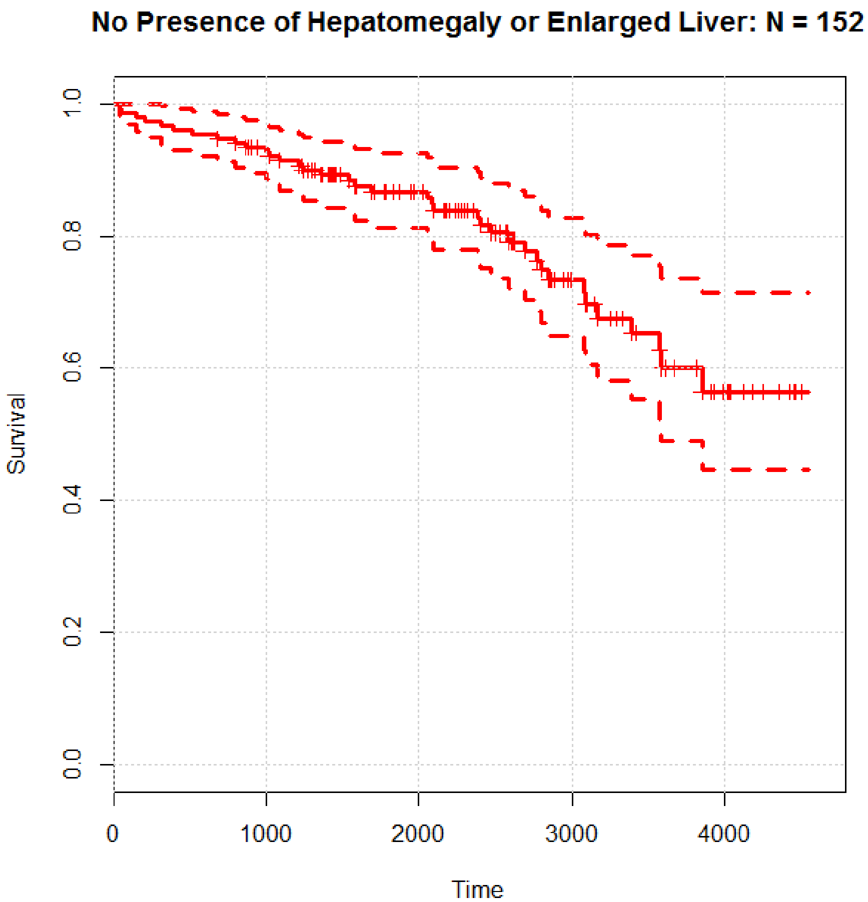


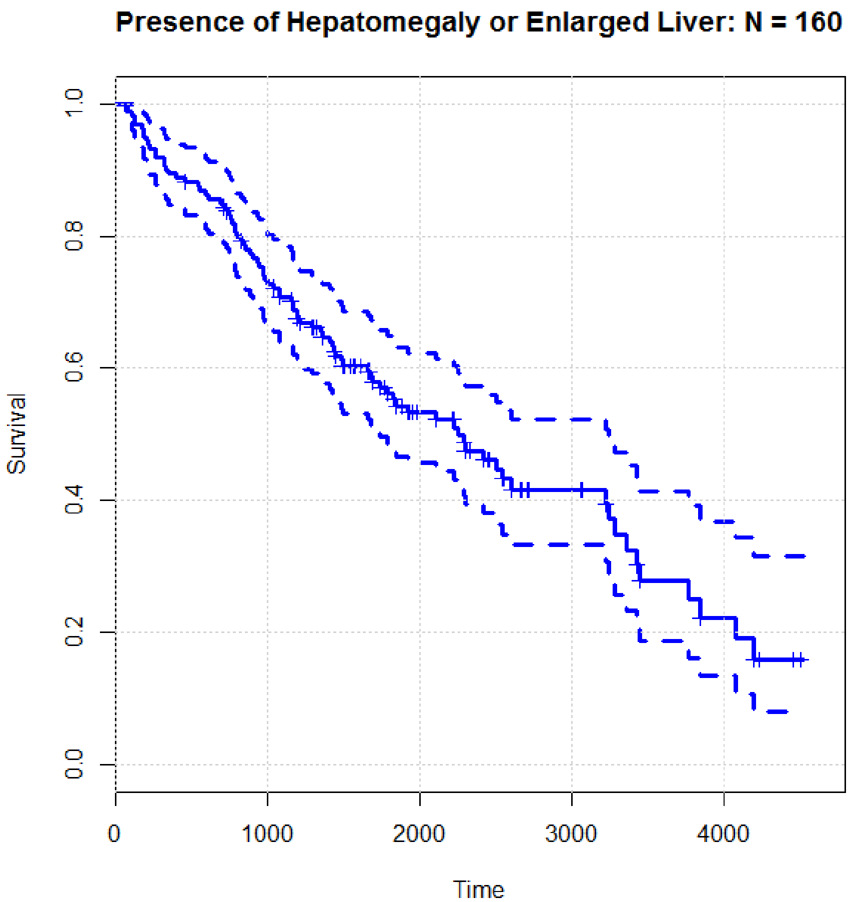
Survival Rate by Presence or Absence of Ascites



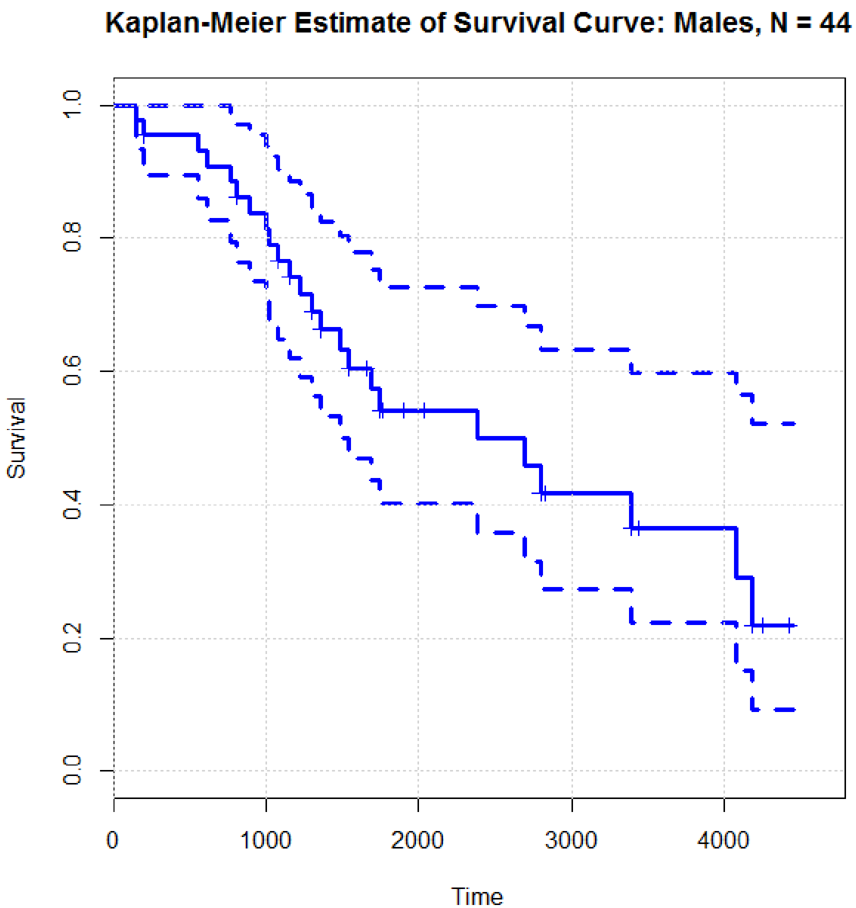


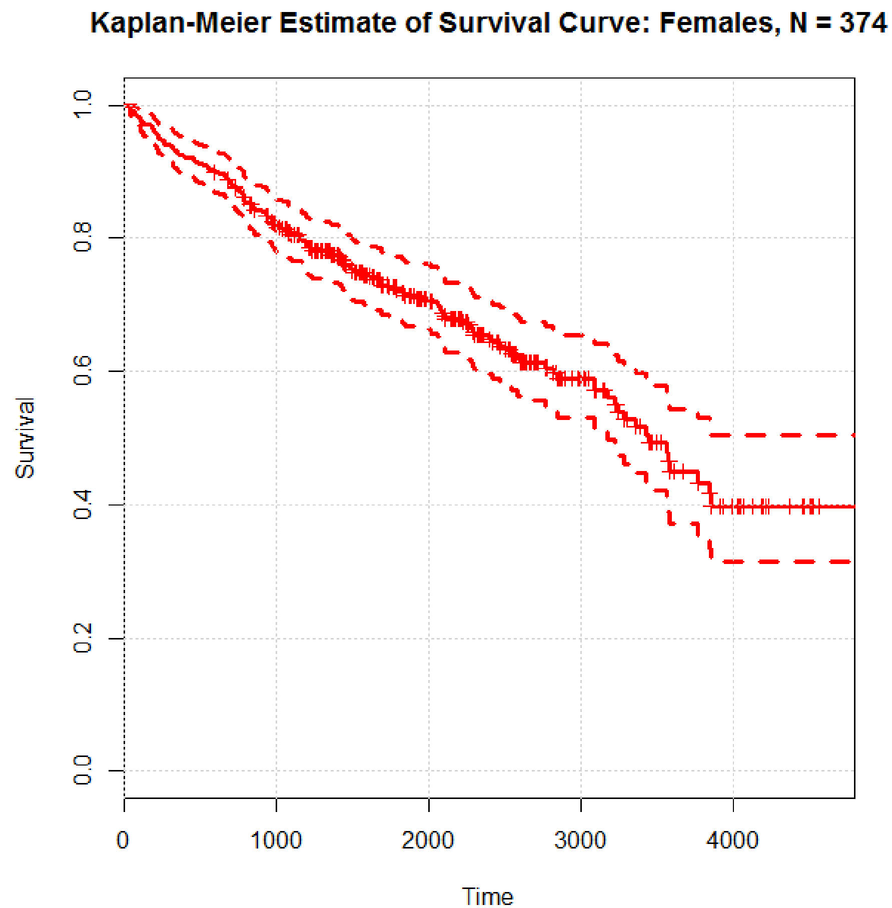
Presence or absence of Hepatomegaly or Enlarged Liver





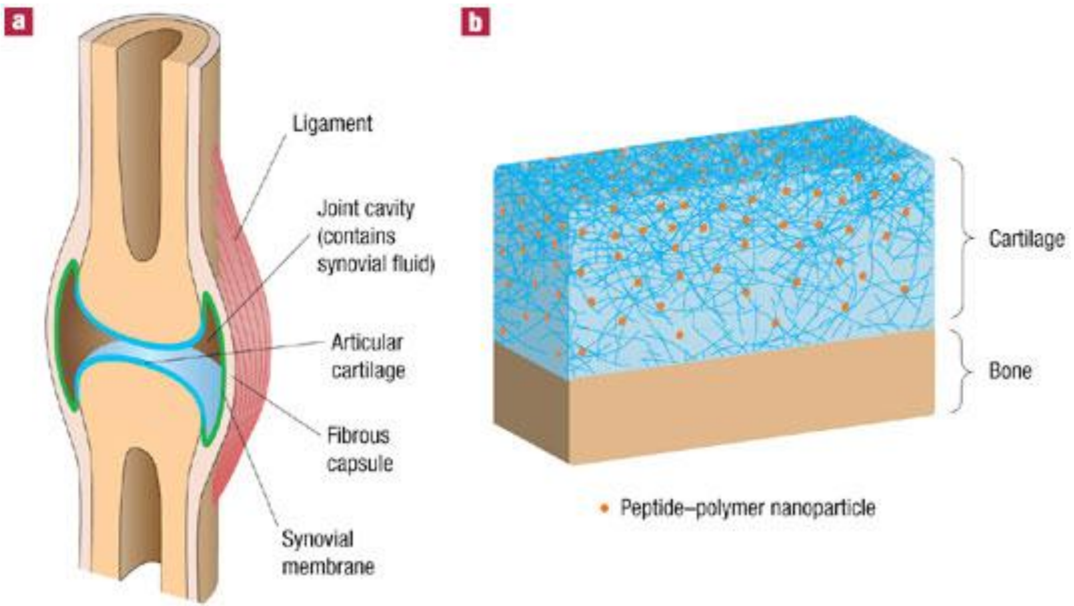
Survival Rate by Gender





Electromechanical Arthritis Detection Probe

- Osteoarthritis arises from degradation of cartilage and bone
- Treatment is pain mitigation, possibly followed by joint replacement
- Early detection offers some possibilities for therapies
- Cartilage is a deformable, electrically responsive solid
  - When strained, it produces current
  - When current is applied, it deforms
- An arthroscopic probe with both electrical and mechanical sensors and/or actuators can take advantage of these properties



Modeling: Mathematical And Statistical

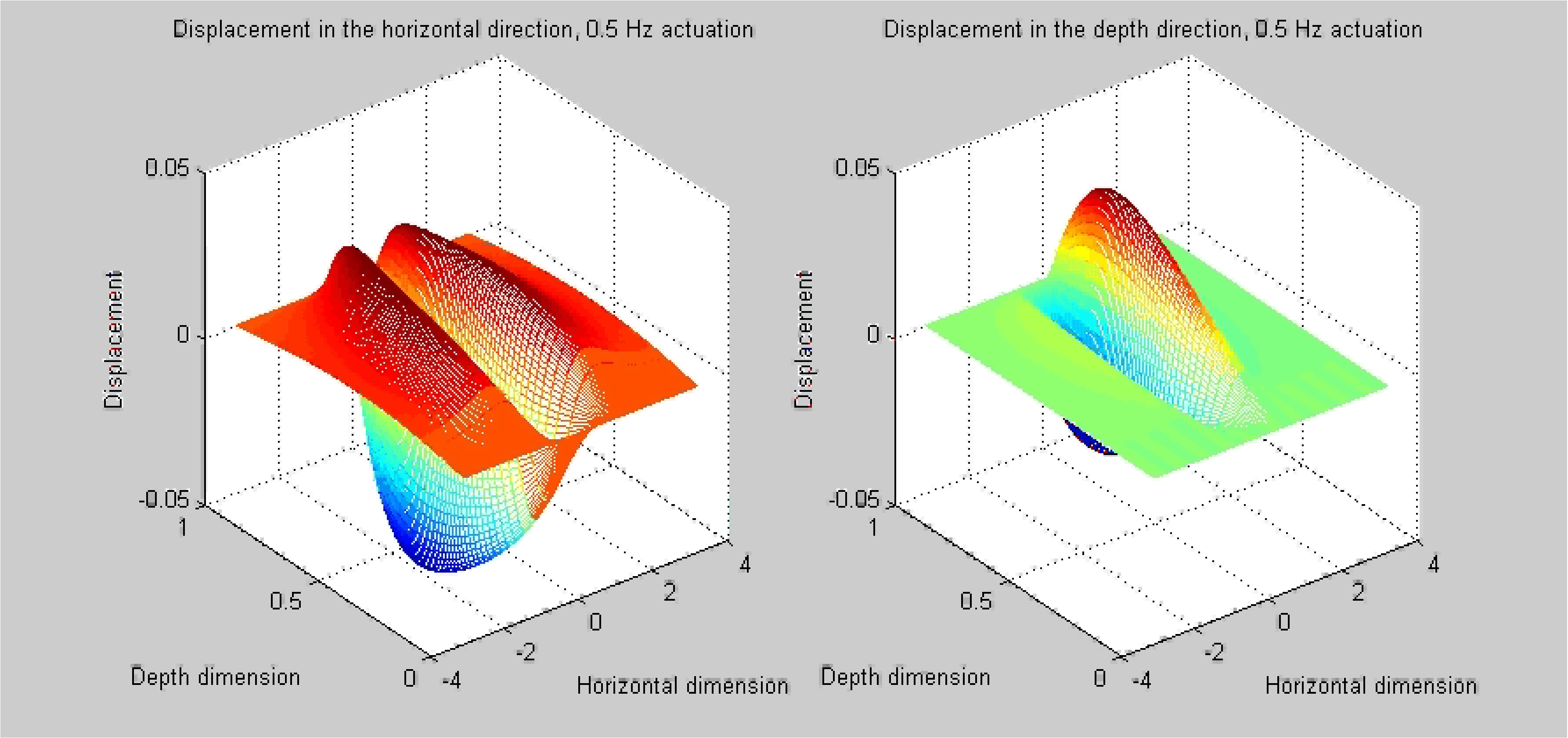
- One goal is understanding of cartilage dynamics under probing
- A second is a data discriminator to detect healthy and degraded cartilage
- For the first, we have the math model

$$0 = i\omega\nabla\cdot u - \nabla\cdot V$$
$$\nabla\cdot\sigma = 0$$
$$\sigma_{ij} = 2G\varepsilon_{ij} + \delta_{ij}(\lambda\varepsilon_{kk} - P)$$
$$\nabla\cdot J = 0$$
$$\begin{bmatrix} V \\ J \end{bmatrix} = \begin{bmatrix} -k_{11} & k_{12} \\ k_{21} & -k_{22} \end{bmatrix}\begin{bmatrix} \nabla P \\ \nabla\Phi \end{bmatrix}$$

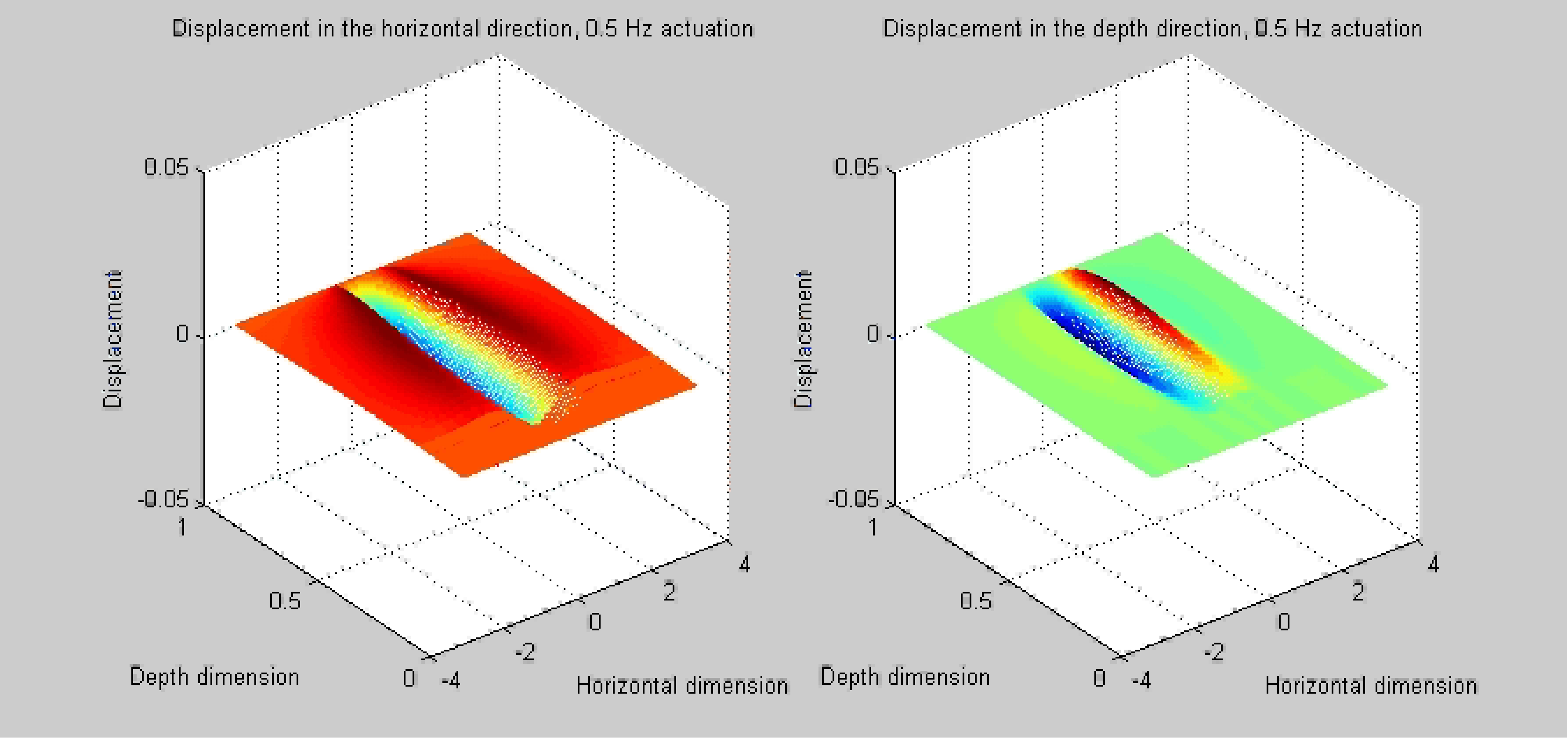
- u is the displacement (in periodic SS)
- P is the fluid pressure
- Φ is the electric potential
- σ, ε are stress and strain tensors
- J is current density
- V is fluid velocity

- For the second, we can identify the transfer function from actuator to sensor
  - The dynamics are linear, time invariant, even though they are spatially distributed
  - At a coarse level, then, the electromechanical model guides the empirical model development
- Displacement, pressure and potential to follow…

Simulation of Healthy Cartilage Response to Deformation Actuation



Simulation of Arthritic Cartilage Response to Deformation Actuation



Modeling: Mathematically Motivated Statistics

- Linear time invariant systems are often modeled with frequency response or transfer functions

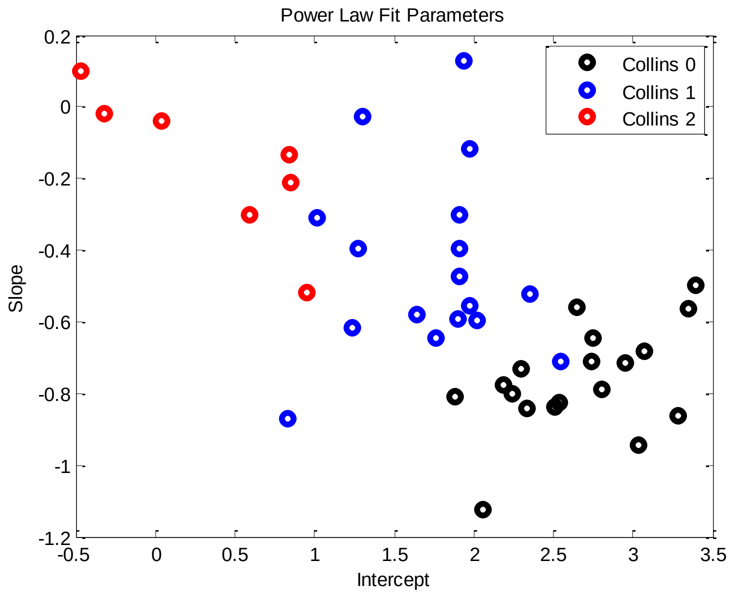
$$Y(\omega) = H(\omega)U(\omega)$$

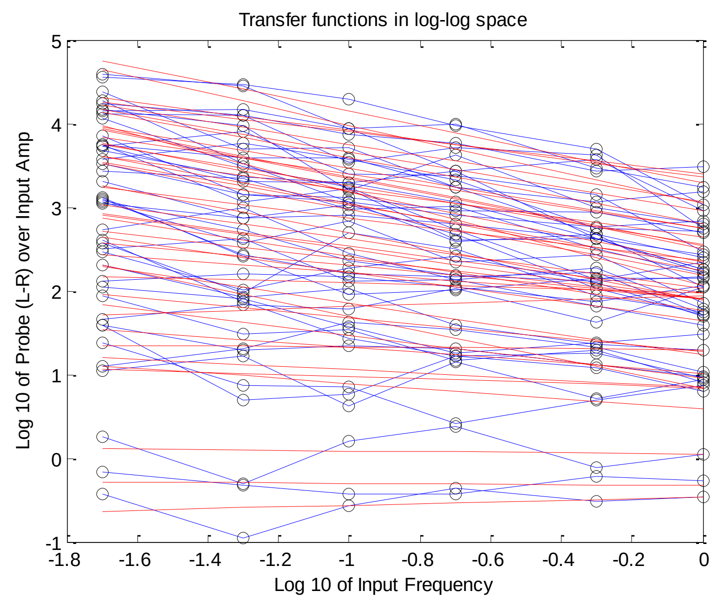
$U =$ Input forcing

$Y =$ Output response

$H =$ System transfer function (model)

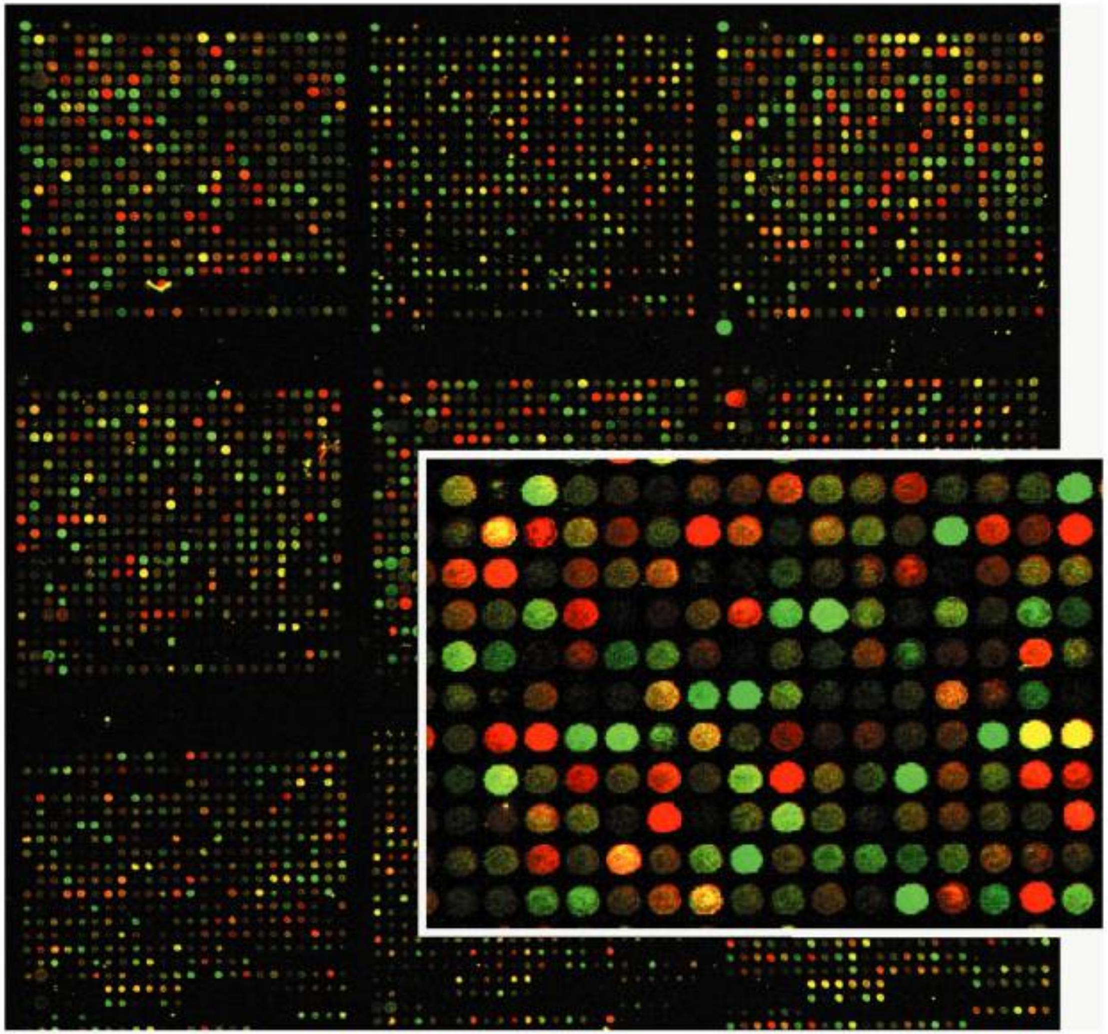
- First goal is to infer H from data
- Second goal is to discriminate "good" and "bad" H's
- Data suggest a power-law H will work
  - 42 Bovine "knee joints"





Modeling Gene Regulation of Cold Shock in Yeast

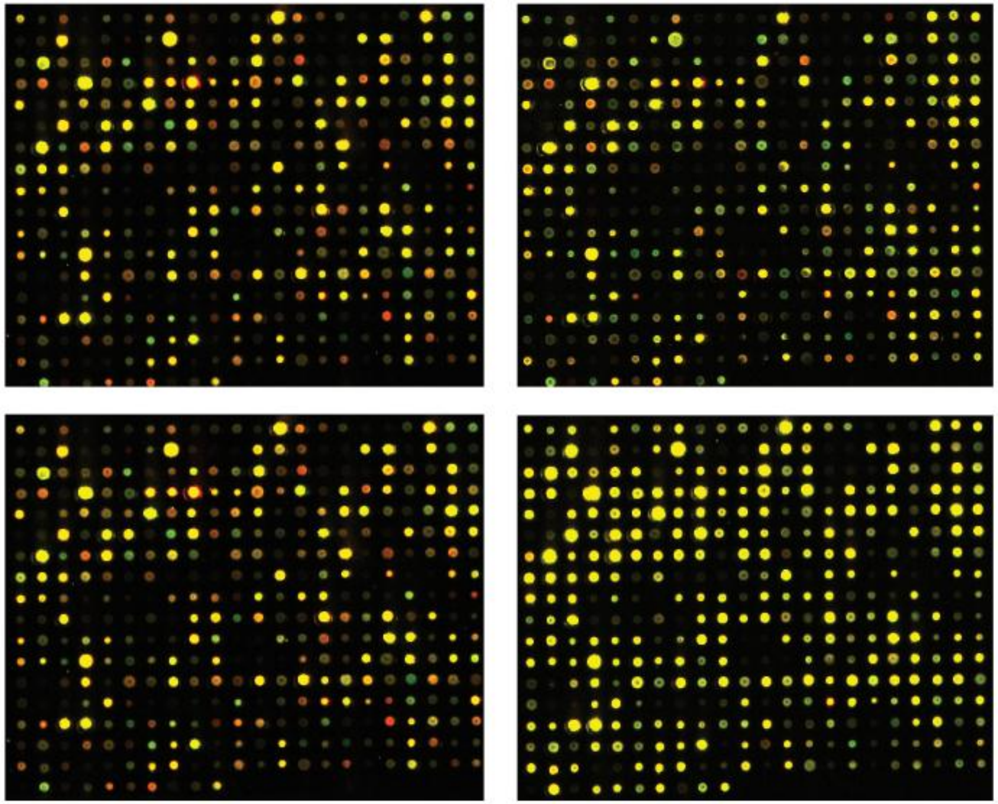
- DNA microarrays are used to measure gene expression for five strains of *S. cerevisiae* during cold shock (wild type and deletion strains)
  - Time course data of five time points of cold shock and recovery
  - Four to five replicates per time point
  - Five different strains
- Preliminary statistical analysis includes normalization and significance testing
  - Within and across chip normalization
  - Time course ANOVA-like testing for significant expression
- 21-gene network of transcription factors is thought to be involved in the cold shock response
  - From the literature and significance tests
- Gene expression is a dynamic balance between production and degradation of RNA and protein products
  - Leads to a conservation-based dynamic model



- One spot = one gene
- Green = decreased relative to control
- Red = increased
- Yellow = no change in gene expression

DNA Microarrays

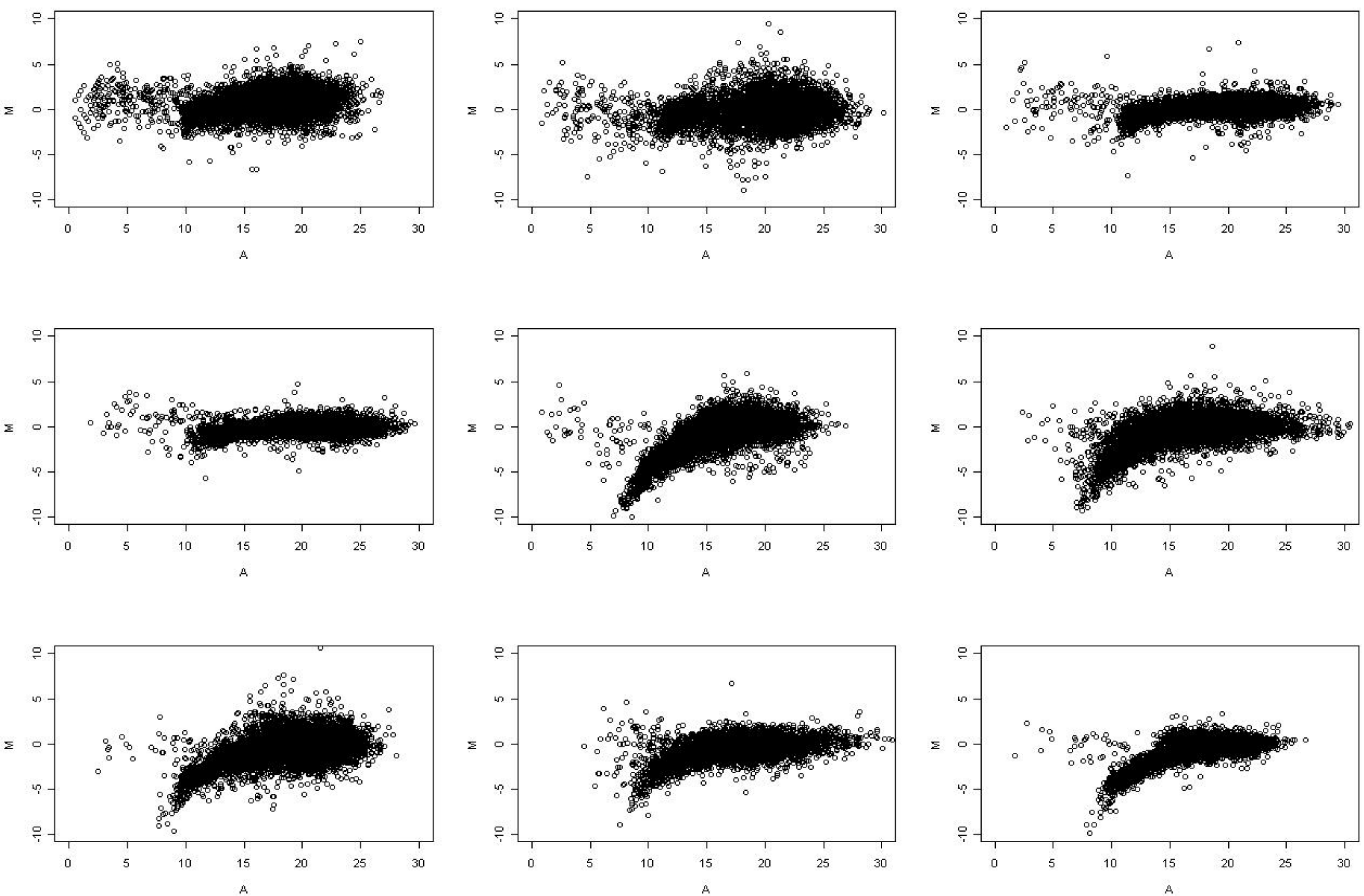
t_{30}/t_0 cold shock t_{60}/t_0 cold shock



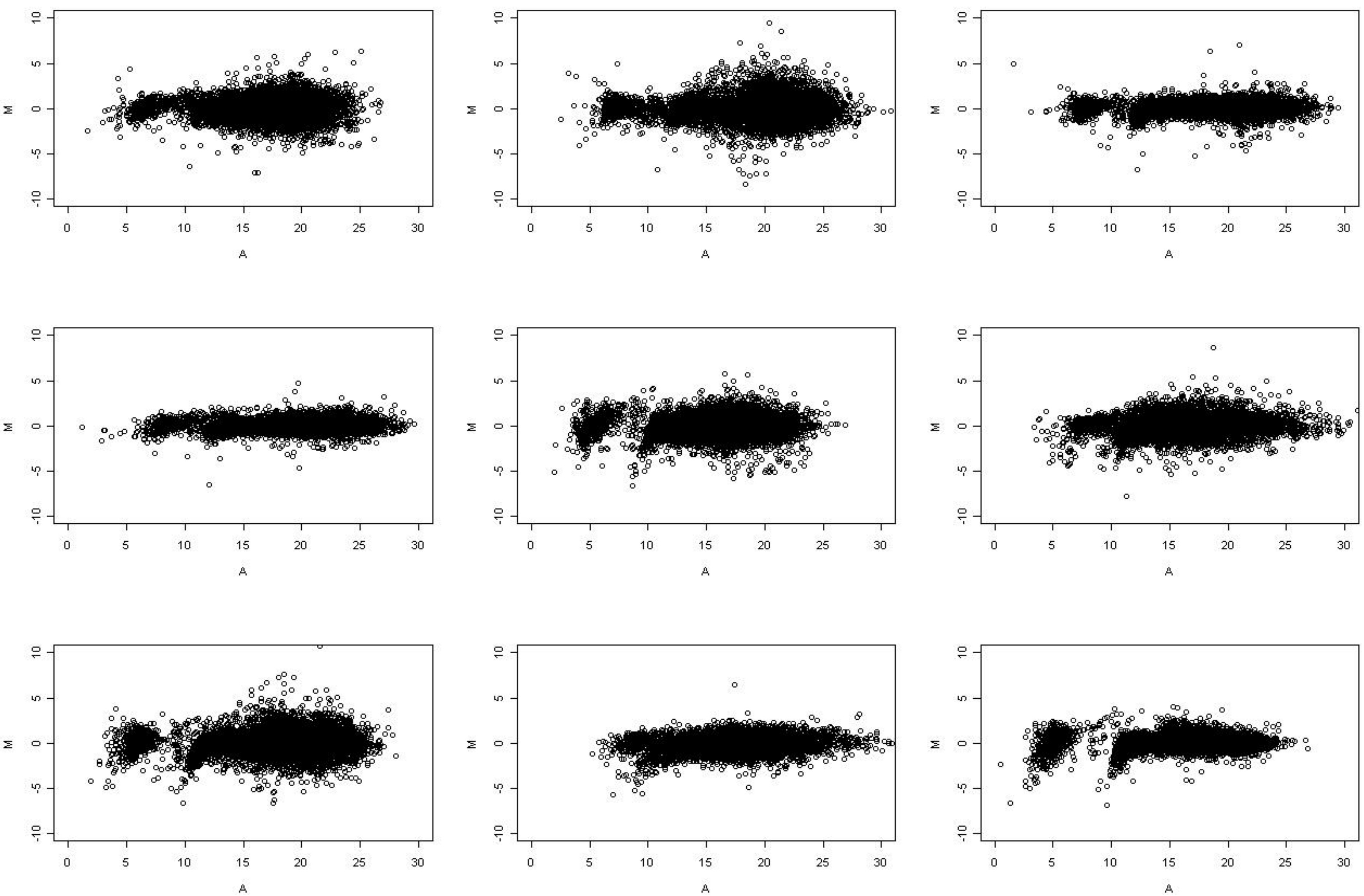
t_{90}/t_0 recovery t_{120}/t_0 recovery

- Gene expression changes in wild type yeast due to cold shock return to pre-shock levels during recovery
- One spot = one gene
- Green = expression decreased relative to control
- Red = expression increased relative to control
- Yellow = no change in gene expression

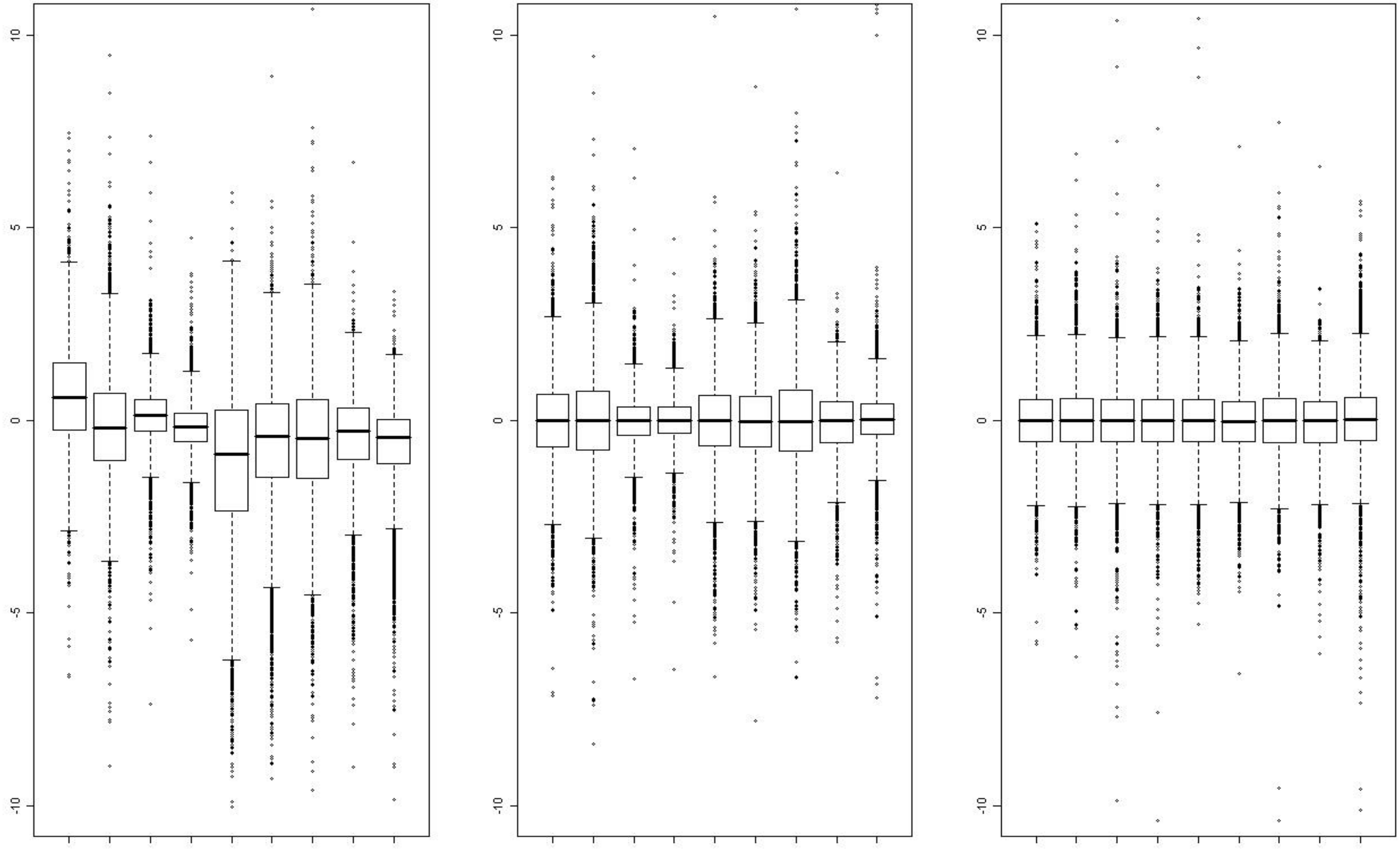
GCAT Chips, Wild Type Yeast, Before Within Array Normalization



GCAT Chips, Wild Type Yeast, After Within Array Normalization



Normalization Across a Set of GCAT Slides



No Normalization **Within Array Normalization** **Between Array Normalization**

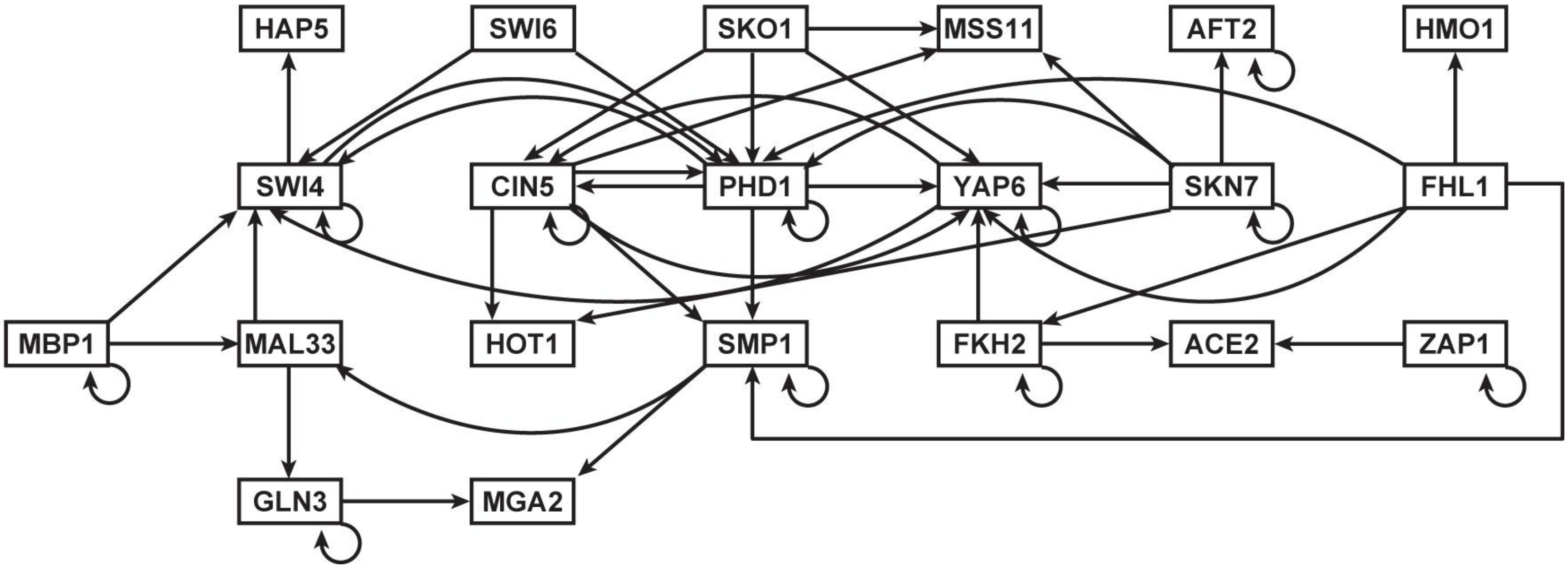
Statistical tests for the significance of gene expression changes

- ANOVA-like test using a General Linear Model to see if at least one time point has a significant expression level, for each gene
- ANOVA-like test using a General Linear Model to see if all strains have the same dynamic profile, for each gene
- Benjamini-Hochsberg correction used for multiple testing control

Strains Compared	GLM ($p<0.05$)	B&H ($p_{B\&H}<0.05$)
All	677	64
Wt vs dCIN5	507	4
Wt vs dGLN3	697	38
Wt vs dHMO1	520	5
Wt vs dZAP1	545	30

Regulatory Network Model

- From the literature, we took the largest connected component of the graph of environmental stress response transcription factors
- We added significantly expressed transcription factor producers from experiments
- The following network emerged from this analysis



Dynamic Model of Expression

- RNA is produced, RNA degrades

$$\dot{x}_i(t) = p_i(x(t)) - d_i x_i(t)$$

- This form represents a system of 21 differential equations
- Production rate function models the network
- Functional form has many possible choices
  - Linear
  - Michaelis-Menten, Hill
  - Sigmoidal

$$p_i(x(t)) = \frac{P_i}{1 + \exp\left(-\sum_j w_{ij}(x_j(t) - \tau_{ij}) \right)}$$

"On/Off" Model of RNA production

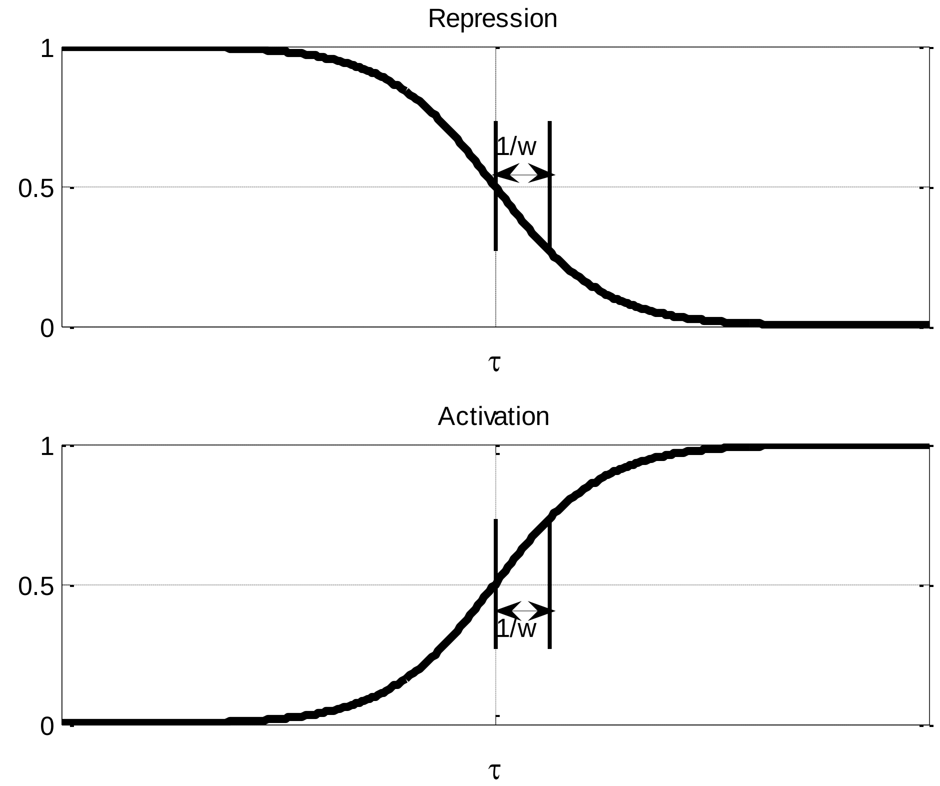
- Standard sigmoid

$$S(u)=\frac{1}{1+e^{-u}}$$

- Scaling and positioning

$$P\cdot S(w(u-\tau))=\frac{P}{1+e^{-w(u-\tau)}}$$

- Sign of w:
  - + = up-regulation (on)
  - - = down-regulation (off)



Fitting Considerations and Inverse Problems

- Natural fit criterion is log ratio least squares:

$$J_0(\theta) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_g} \left(\log_2 (x_j(t_i;\theta)) - m_j(t_i) \right)^2$$

$$m_j(t_i) = \text{Observed (normalized) log ratio data}$$

$$x_j(t_i;\theta) = \text{Parameter-dependent solution of DE}$$

$$\theta = (w, \tau, P)$$

$$\dot{x}_j(t;\theta) = \frac{P_j}{1 + \exp\left(-\sum_k w_{jk}(x_k - \tau_{jk}) \right)} - d_j x_j$$

Penalty Terms

- Typically the number of observations is small:
  - Very few time points, that is.
  - Model contains many parameters

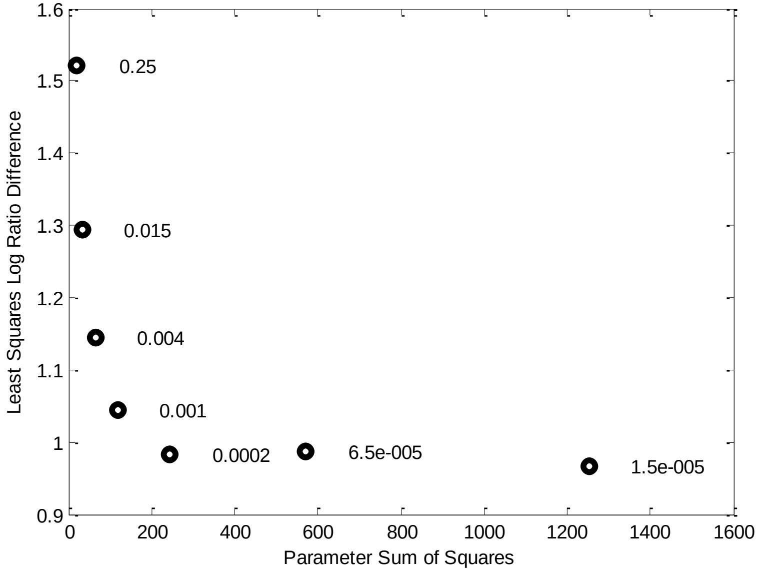
$$J_\alpha(\theta) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_g} \left(\log_2(x_j(t_i;\theta)) - m_j(t_i)\right)^2$$

$$+\alpha Q(\theta)$$

$$Q = \text{Positive definite quadratic form}$$

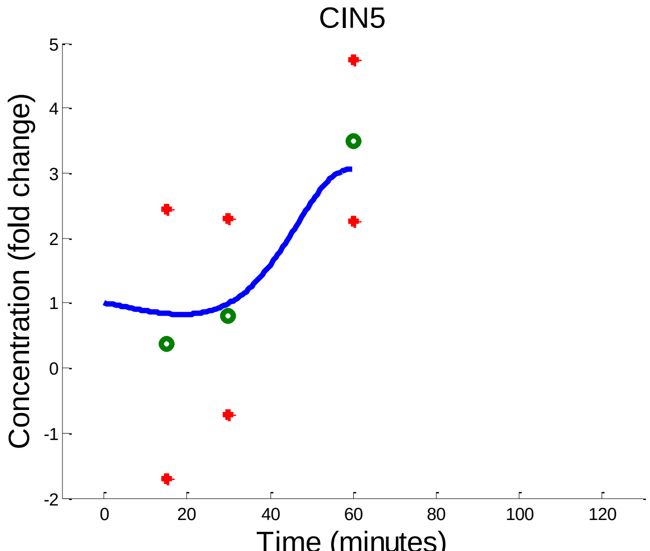
$$\alpha = \text{Penalty parameter}$$

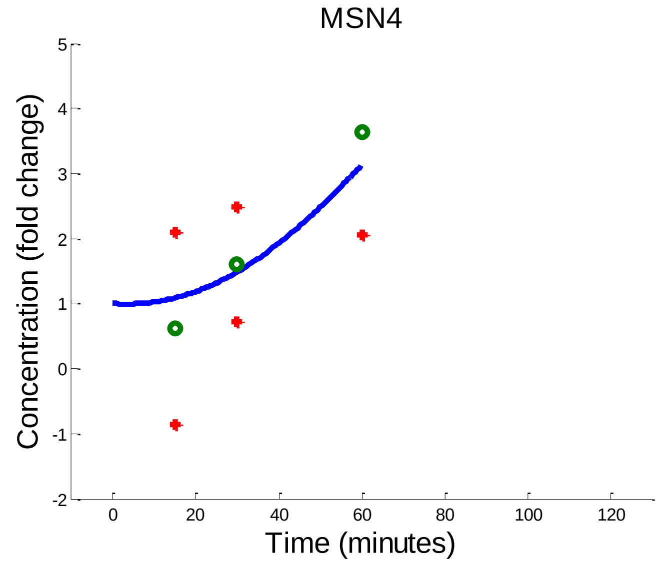
- L-curve analysis helps select the penalty parameter

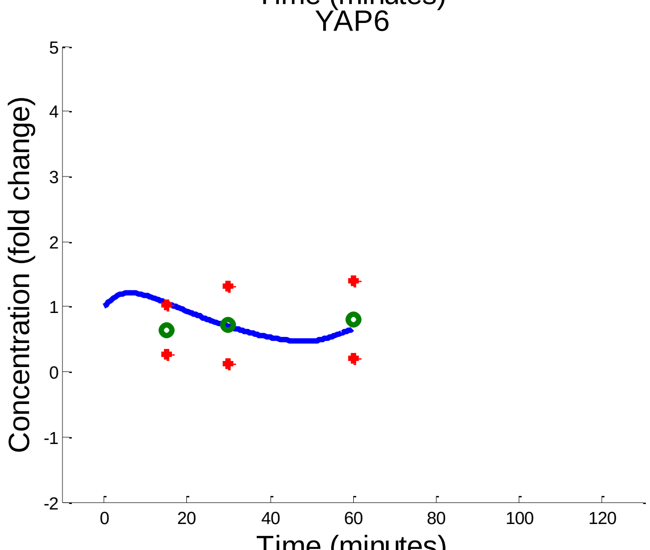


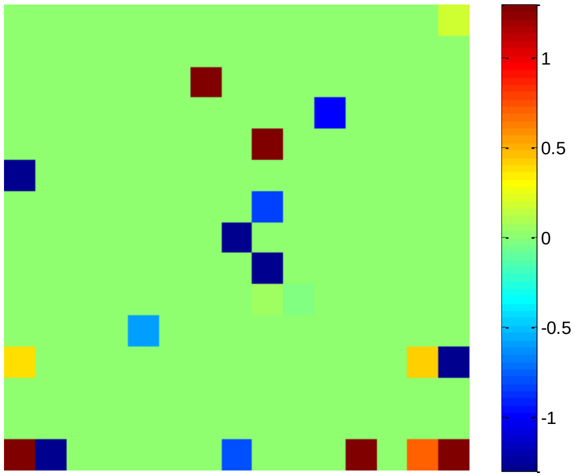
Data from Our Lab

- Data collected at 15, 30, 60 minutes
- Cold shock from 30°C to 13°C









NIH Systems Science Initiatives

- Seeks proposals on basic and applied projects using systems science methods
- Systems science refers to the leverage gained by considering interacting components and individuals rather than taking a reductionist approach
- Systems science usually requires computational, mathematical, and/or statistical models
- Of particular interest to NIH are public health problems ranging from heart, lung, and blood disease to substance abuse and obesity, coupling socioeconomic, cultural, behavioral, and policy issues
- Institutes including NCI, NHLBI, NIBIB, NIGMS are participating
- R01 and R21 funding is available

The FDA's Quality by Design Approach

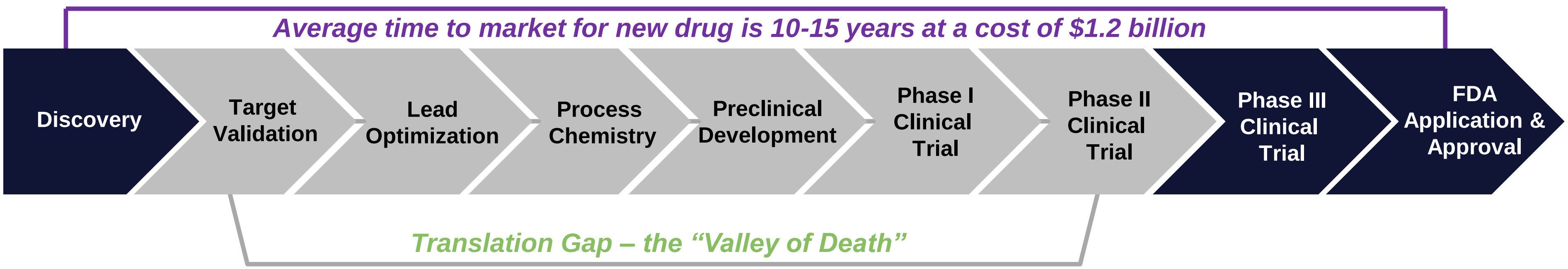
"Quality by Design (QbD) is a systematic approach to pharmaceutical development that begins with predefined objectives. It emphasizes product and process understanding and process control, based on sound science and quality risk management."

- This systematic approach forces manufacturers to develop a process to produce the desired quality of the product and help the FDA to better understand the manufacturing strategy
- The **QbD strategy** can include a variety of components and will vary depending on the drug candidate, the patient needs and the manufacturing company:
  - Incorporation of prior knowledge (public domain or internally documented)
  - Results of studies using statistically designed experiments
  - Use of quality risk management
  - Use of knowledge management throughout the lifecycle (Knowledge management is a systematic approach to acquiring, analyzing, storing, and disseminating information related to products, manufacturing processes, and components of the product)

Pharmaceutical Development

- The purpose of drug development is to design a quality product whereby the manufacturing process will **consistently** create the intended performance of the product
- Despite advancements in genomics, computerized molecular modeling and screening, there is a significant decrease in the number of new drugs entering the market
- Several factors at work:
  - Incorrect drug candidates
  - Sloppy manufacturing
  - Funding to bring through the pipeline

Therapeutic Drug Discovery Pipeline



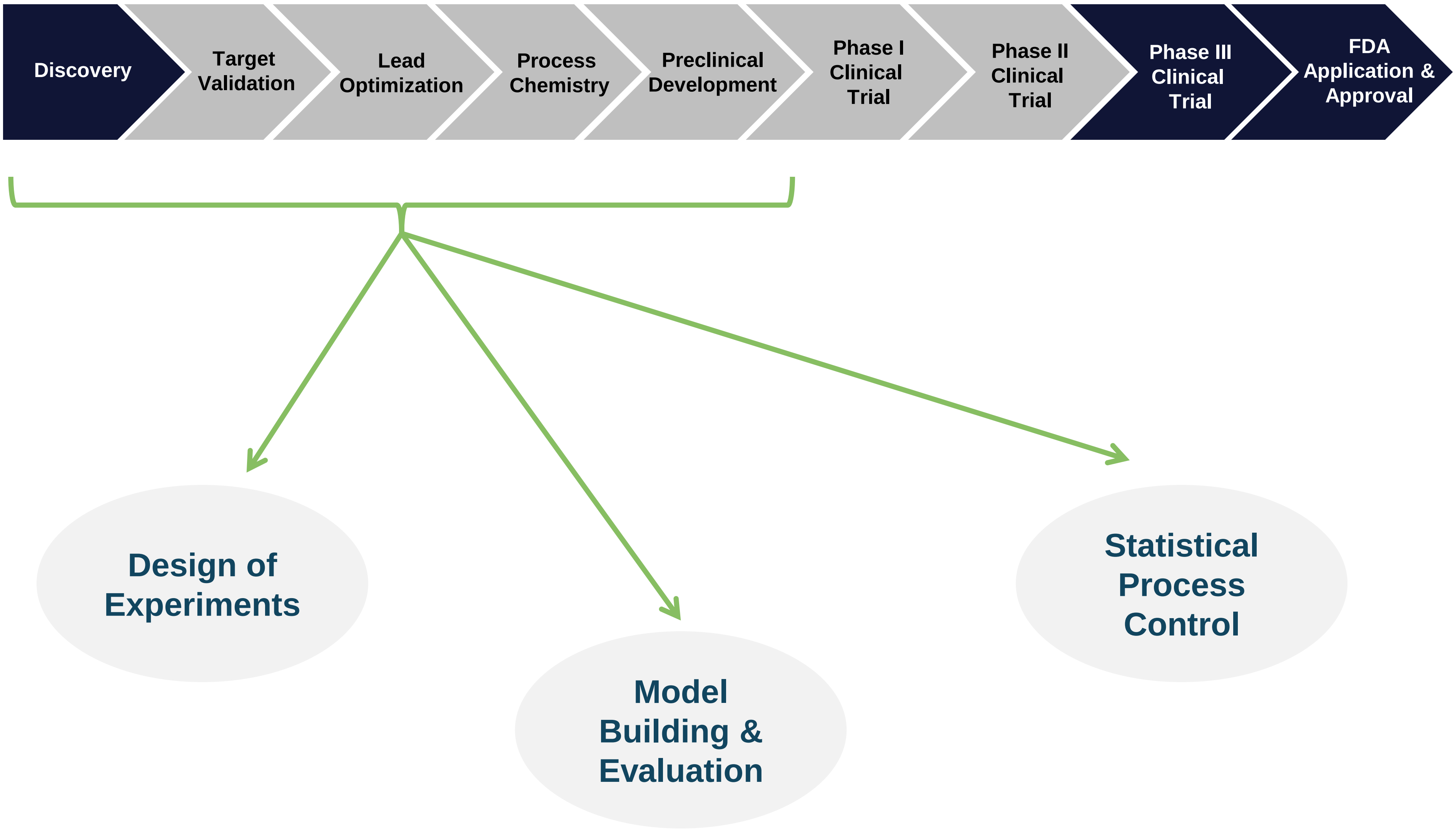
FDA is dedicated to ensuring that manufacturers improve their production processes to enable product durability and embrace the concept that quality cannot be tested into a product – it should be built in by design.

The Quality by Design Initiative – Cont.

- As outlined by the FDA, the following components are included in a QbD approach:
  - The product is designed to meet patient requirements
  - The process is designed to consistently meet product critical quality attributes (CQA, physical, chemical, biological or microbiological properties or characteristics that need to be controlled (directly or indirectly) to ensure product quality)
  - The impact of formulation components and process parameters on product quality is understood
  - Critical sources of process variability are identified and controlled
  - The process is continually monitored and updated to assure consistent quality over time

Current Approach	***QbD Approach***
Quality assured by testing and inspection	Quality built into product & process by design, based on scientific understanding
Data intensive submission – disjointed information without "big picture"	Knowledge rich submission – showing product knowledge & process understanding
Specifications based on batch history	Specifications based on product performance requirements
'Frozen process," discouraging changes	Flexible process within design space, allowing continuous improvement
Focus on reproducibility – often avoiding or ignoring variation	Focus on robustness – understanding and controlling variation

Systems Science and QbD

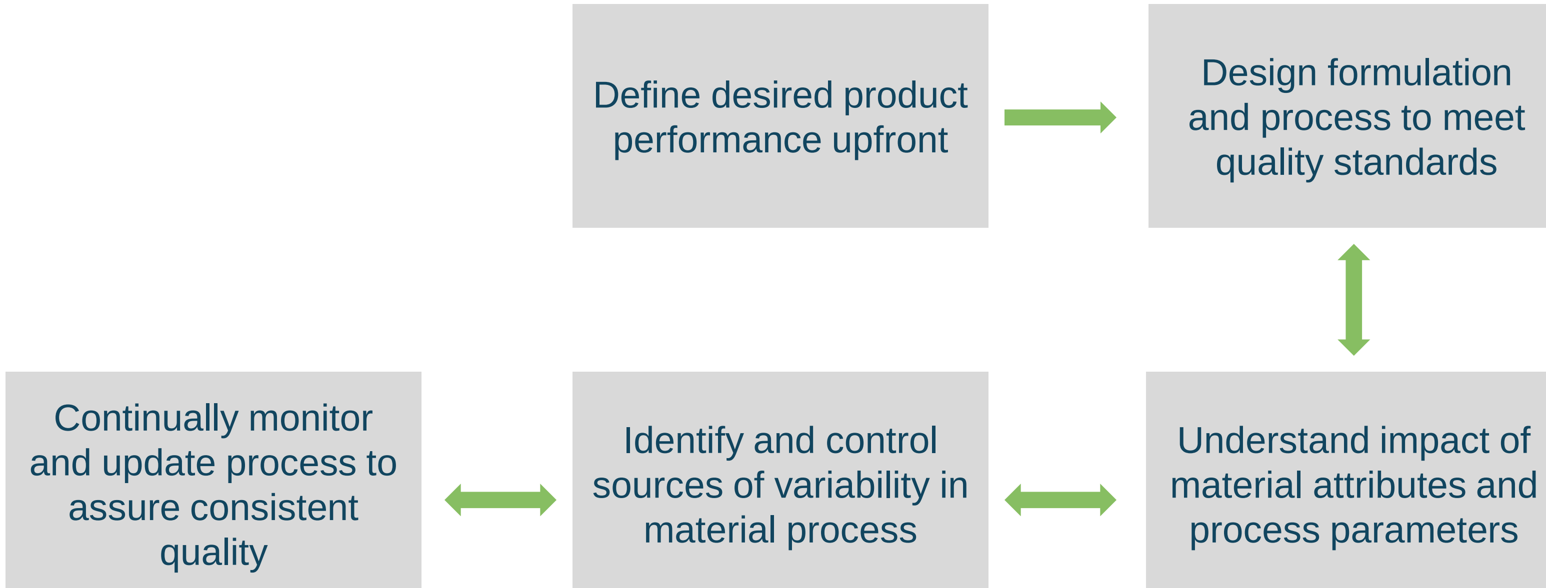


QbD in A Nutshell

- Scientific, risk-based, holistic and proactive approach to pharmaceutical development
- Deliberate design effort from product conception through commercialization„

  *(Developing **product** to meet predefined product quality, safety and efficacy)*
- Full understanding of how product attributes and process relate to product performance

  *(Designing manufacturing **process** to meet predefined product quality, safety and efficacy)*
- QbD should be customized to each product
- Use of statistics and models are key factor of success



*Adapted from RDD materials, 2007

Questions and Discussion

- Tempest is a full service quantitative analysis company

- We support scientific and engineering R&D from proposal writing and early stage exploratory analysis to prototyping to reporting and product manufacture

Thank you!

Tempest Technologies

8939 South Sepulveda Blvd, Suite 506
Los Angeles, CA
USA 90045

T +1.310.216.1677 F +1.310.216.1628

www.tempest-tech.com

@tempestmath