Computational Systems Biology
Advanced Technologies in Bioscience 2008–2009
Chalmers Graduate School in Bioscience

T. M. Murali

August 18, 2008
Course Structure

Discuss state-of-the-art research papers.
Course Structure

Discuss state-of-the-art research papers.

- Reading assignments
Course Structure

Discuss state-of-the-art research papers.

- Reading assignments
- Lectures
Course Structure

Discuss state-of-the-art research papers.

- Reading assignments
- Lectures
- Exercises
Course Structure

Discuss state-of-the-art research papers.

- Reading assignments
- Lectures
- Exercises
- Class participation
Suggestions on Reading

- Be sceptical/critical: even papers in Nature, Science, or PNAS have errors or invalid thinking.
Suggestions on Reading

- Be sceptical/critical: even papers in Nature, Science, or PNAS have errors or invalid thinking.
- Algorithmic/computational papers:
  - Are the biological assumptions valid?
  - Is the algorithm good and computational efficient? Can you improve the technique?
  - Can you mathematically describe the output of the algorithm?
Suggestions on Reading

- Be sceptical/critical: even papers in Nature, Science, or PNAS have errors or invalid thinking.
- Algorithmic/computational papers:
  - Are the biological assumptions valid?
  - Is the algorithm good and computational efficient? Can you improve the technique?
  - Can you mathematically describe the output of the algorithm?
- Read supplementary information. Often has details about the assumptions, the techniques, and the results.
Sources of Information

- There is no textbook for the course.
- Useful/related books:
  - *Networks: From Biology to Theory*, Jianfeng Feng, Jürgen Jost, and Minping Qian, Springer-Verlag.
  - *Computational Modeling of Genetic and Biochemical Networks*, James M. Bower and Hamid Bolouri, MIT Press
More Sources of Information

- Conferences: ICSB, RECOMB, ISMB, PSB, KDD, machine learning conferences, discrete algorithms conferences.
- Journals (CS-oriented): Bioinformatics, Journal of Computational Biology, BMC Bioinformatics, TCBB, TKDE.
Rewind to 1953

1953

King's College, London. One of 12 (J.D. W.), has been awarded a fellowship from the National Foundation for Infantile Paralysis.

A. J. D. Waterson 1957.12

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cambridge, Massachusetts. 1949.

Molecular Structure of Nucleic Acids

We wish to suggest a structure for the salt of deoxyribonucleic acid (DNA). This structure has novel features which are of considerable biological interest. A structure for nucleic acid has already been proposed by Pauling and Corey. They kindly made available the drawings and discussion of their model to me, and I am indebted to them for this information.

A structure for nucleic acid must have a certain number of properties: (1) It must be capable of holding the two strands of DNA together to form a helix, with the bases on the inside and the sugar-phosphate backbone on the outside. (2) The structure must be capable of forming hydrogen bonds between the bases. (3) The structure must be able to unwind and reform when the DNA is transcribed or replicated.

A structure for nucleic acid has been suggested by Pauling and Corey. It consists of two anti-parallel strands of DNA, with the sugar-phosphate backbone on the outside and the bases on the inside. The bases are hydrogen bonded to form a helix, and the helix is stabilized by hydrogen bonds between the sugar-phosphate backbone and the bases.

We have suggested a new structure for DNA, which differs from the Pauling-Corey model in several important ways. Our structure consists of two anti-parallel strands of DNA, with the sugar-phosphate backbone on the outside and the bases on the inside. The bases are hydrogen bonded to form a helix, and the helix is stabilized by hydrogen bonds between the sugar-phosphate backbone and the bases.

In our structure, the two strands are anti-parallel, with the sugar-phosphate backbone on the outside and the bases on the inside. The bases are hydrogen bonded to form a helix, and the helix is stabilized by hydrogen bonds between the sugar-phosphate backbone and the bases.

In our structure, the two strands are anti-parallel, with the sugar-phosphate backbone on the outside and the bases on the inside. The bases are hydrogen bonded to form a helix, and the helix is stabilized by hydrogen bonds between the sugar-phosphate backbone and the bases.

In our structure, the two strands are anti-parallel, with the sugar-phosphate backbone on the outside and the bases on the inside. The bases are hydrogen bonded to form a helix, and the helix is stabilized by hydrogen bonds between the sugar-phosphate backbone and the bases.
The Human Genome Project

Before: human genome has about 100,000 genes.
After: human genome has about 30,000 genes.
The Human Genome Project

Before: human genome has about 100,000 genes.
The Human Genome Project

Before: human genome has about 100,000 genes.

After: human genome has about 30,000 genes.
The New York Times: **Genome Analysis Shows Humans Survive on Low Number of Genes** The two teams report that there are far fewer human genes than thought—probably a mere 30,000 or so—only a third more than those found in the roundworm. ... **The impact on human pride is another matter.**
Shock and Dismay

The New York Times: **Genome Analysis Shows Humans Survive on Low Number of Genes** The two teams report that there are far fewer human genes than thought—probably a mere 30,000 or so—only a third more than those found in the roundworm. ... The impact on human pride is another matter.

Washington Post: It also raises new and difficult questions, such as how human beings—with all their passions and fears, their capacity for art, music, culture and war—can be all that they are with just 30,000 or so genes, only five times as many as in baker’s yeast.
Shock and Dismay

- The New York Times: **Genome Analysis Shows Humans Survive on Low Number of Genes** The two teams report that there are far fewer human genes than thought—probably a mere 30,000 or so—only a third more than those found in the roundworm. ... The impact on human pride is another matter.

- Washington Post: It also raises new and difficult questions, such as how human beings—with all their passions and fears, their capacity for art, music, culture and war—can be all that they are with just 30,000 or so genes, only five times as many as in baker’s yeast.

- USA TODAY: Perhaps the biggest surprise since the code was deciphered in June is that it takes just 30,000 to 40,000 genes to make, maintain and repair a human. ... “If you’re judging the complexity of an organism by the number of genes it has, we’ve just taken a big hit in the pride department,” says the National Genome Research Institute's director, Francis Collins, who also heads the U.S. arm of the International Human Genome Project.
Shock and Dismay

- The New York Times: **Genome Analysis Shows Humans Survive on Low Number of Genes** The two teams report that there are far fewer human genes than thought—probably a mere 30,000 or so—only a third more than those found in the roundworm. ... The impact on human pride is another matter.

- Washington Post: It also raises new and difficult questions, such as how human beings—with all their passions and fears, their capacity for art, music, culture and war—can be all that they are with just 30,000 or so genes, only five times as many as in baker’s yeast.

- USA TODAY: Perhaps the biggest surprise since the code was deciphered in June is that it takes just 30,000 to 40,000 genes to make, maintain and repair a human. ... “If you’re judging the complexity of an organism by the number of genes it has, we’ve just taken a big hit in the pride department,” says the National Genome Research Institute’s director, Francis Collins, who also heads the U.S. arm of the International Human Genome Project.

- The New York Times (Aug 24, 2001): **Human Genome Now Appears More Complicated After All** After a humiliating deflation this February, human dignity is on the recovery path, at least as measured by the number of genes in the human genome.
Relative Genome Sizes

- Human: 31,000
- Thale cress: 26,000
- Nematode worm: 18,000
- Fruit fly: 13,000
- Yeast: 6,000
- Tuberculosis microbe: 4,000
Chimps vs. Humans

Chimp and chump genomes are only about 1.2% different!
Chimps vs. Humans

Chimp and chump genomes are only about 1.2% different!
Chimps vs. Humans

Chimp and chump genomes are only about 1.2% different!
Chimp and chump genomes are only about 1.2% different!
What Factors Differentiate Various Species?

- Genes are different (only dogs have the submaxillary mucin genes).
- Patterns of gene activity (gene expression) are different.
- Ways in which proteins interact with and regulate each other and other molecules are different.

"It is the evolution of the regulatory networks and not the genes themselves that play the critical role in making organisms different from one another," The Digital Code of DNA, Hood and Galas, Nature, vol 421, 2003.

- We need to understand how genes, proteins, and other molecules interact with other in different cell states, different tissues, and under different external conditions.
- Study only of individual elements is unlikely to reveal higher-order principles.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the *submaxillary mucin* genes).

> “It is the evolution of the regulatory networks and not the genes themselves that play the critical role in making organisms different from one another,” The Digital Code of DNA, Hood and Galas, *Nature*, vol 421, 2003.

- We need to understand how genes, proteins, and other molecules interact with each other in different cell states, different tissues, and under different external conditions.

- Study only of individual elements is unlikely to reveal higher-order principles.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the *submaxillary mucin* genes).
- Patterns of gene activity (gene expression) are different.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the *submaxillary mucin* genes).
- Patterns of gene activity (gene expression) are different.
- Ways in which proteins interact with and regulate each other and other molecules are different.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the *submaxillary mucin* genes).
- Patterns of gene activity (gene expression) are different.
- Ways in which proteins interact with and regulate each other and other molecules are different.

- "It is the evolution of the regulatory networks and not the genes themselves that play the critical role in making organisms different from one another," The Digital Code of DNA, Hood and Galas, *Nature*, vol 421, 2003.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the submaxillary mucin genes).
- Patterns of gene activity (gene expression) are different.
- Ways in which proteins interact with and regulate each other and other molecules are different.

- “It is the evolution of the regulatory networks and not the genes themselves that play the critical role in making organisms different from one another,” The Digital Code of DNA, Hood and Galas, Nature, vol 421, 2003.

- We need to understand how genes, proteins, and other molecules interact with other in different cell states, different tissues, and under different external conditions.
What Factors Differentiate Various Species?

- Genes are different (only dogs have the *submaxillary mucin* genes).
- Patterns of gene activity (gene expression) are different.
- Ways in which proteins interact with and regulate each other and other molecules are different.

- “It is the evolution of the regulatory networks and not the genes themselves that play the critical role in making organisms different from one another,” *The Digital Code of DNA*, Hood and Galas, *Nature*, vol 421, 2003.

- We need to understand how genes, proteins, and other molecules interact with other in different cell states, different tissues, and under different external conditions.
- Study only of individual elements is unlikely to reveal higher-order principles.
Enter Systems Biology

- Systems Biology is the study of the parts of the cell, their properties, and their relationships.
Enter Systems Biology

- Systems Biology is the study of the parts of the cell, their properties, and their relationships.
- What are the structures and modules that make up cellular networks?
Enter Systems Biology

- Systems Biology is the study of the parts of the cell, their properties, and their relationships.
- What are the structures and modules that make up cellular networks?
- How do these modules interact with each other over time and in different situations?
Enter Systems Biology

- Systems Biology is the study of the parts of the cell, their properties, and their relationships.
- What are the structures and modules that make up cellular networks?
- How do these modules interact with each other over time and in different situations?
- How can we interrogate the cell and iteratively refine our models of the cell?
Characteristics of Systems Biology

- Modular cell biology (rather than molecular).
- Discovery-driven \textit{and} hypothesis-driven.
- Driven by high-throughput and accurate biological measurements.
- Uses and needs sophisticated computational, mathematical, and statistical ideas.
- Requires close collaboration between biologists and quantitative scientists.
Continuum of Models in Systems Biology

Goals of the Course

- We will cover “high-level” models.
- Emphasise a data-driven approach to systems biology.
- Focus on large-scale properties of biological systems.
- Integrate massive quantities of different types of data.
- Learn techniques from clustering, data mining, and graph theory and apply them to solve specific biological questions.
Sources of Data

- Gene expression data
- Gene knockouts and external perturbations such as drugs.
- Samples belonging to various classes
- Time-series data.
- GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
- Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
- Affinity precipitation + mass spectometry (yeast).
- Literature (HPRD).

- Transcriptional regulation
- Protein-DNA binding data (yeast, human liver TFs).
- Binding profiles for known TFs (SCPD, TRANSFAC).

- Protein abundance and activity
- Metabonomics
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.
- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).
- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).
- Protein abundance and activity
  - Metabonomics
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).

- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).

- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).

- Protein abundance and activity
Sources of Data

- Gene expression data
  - Gene knockouts and external perturbations such as drugs.
  - Samples belonging to various classes
  - Time-series data.
  - GEO, SGD, the Whitehead institute.

- Protein-protein interaction data
  - Large-scale Yeast 2-hybrid assays (yeast, worm, fruitfly).
  - Affinity precipitation + mass spectometry (yeast).
  - Literature (HPRD).

- Transcriptional regulation
  - Protein-DNA binding data (yeast, human liver TFs).
  - Binding profiles for known TFs (SCPD, TRANSFAC).

- Protein abundance and activity

- Metabonomics
Sources of Data

- Literature, Computation, Databases
  - Transcriptional regulators (TRANSFAC)
  - Protein-protein interactions (DIP, GRID, Predictome, MIPS)
  - Metabolic networks (KEGG, EcoCyC, BioCarta, GenMAPP)
  - Functional annotations (GO, MIPS, species-specific databases)
  - Genetic Associations with Disease (GAD, MEDGENE, i-HOP).
Specific Topics

Monday  Clustering gene expression data; application to find cancer gene modules.

Tuesday  Biclustering, application to data integration in *S. cerevisiae*.

Wednesday  Response networks and network legos.

Thursday  Gene function prediction.

Friday  Host-pathogen interaction networks (*ICSB tutorial*).
Gene Regulation

Diagram showing the interaction between activators, co-activators, and general transcription factors. The diagram illustrates the process of gene regulation with an emphasis on enhancers, promoter-proximal regions, and transcriptional machinery.

T. M. Murali  
August 18, 2008  
CSB@Chalmers
Regulatory Networks
Regulatory Networks

Module A functions:

Vegetal plate expression in early development:

Synergism with modules B and G enhancing endoderm expression in later development:

Repression in ectoderm (modules E and F) and skeletogenic mesenchyme (module DC):

Modules E, F and DC with LiCl treatment:
Regulatory Networks

B

if (F = 1 or E = 1 or CD = 1) and (Z = 1)
    \( \alpha = 1 \)
else
    \( \alpha = 0 \)
if (P = 1 and CG = 1)
    \( \beta = 2 \)
else
    \( \beta = 0 \)
if (CG = 1 and CG = 1 and CG = 1)
    \( \gamma = 2 \)
else
    \( \gamma = 1 \)
\( \delta(t) = B(t) + G(t) \)
\( \epsilon(t) = \beta \cdot \delta(t) \)
if (\( \epsilon(t) = 0 \))
    \( \zeta(t) = \text{Otx}(t) \)
else
    \( \zeta(t) = \epsilon(t) \)
if (\( \alpha = 1 \))
    \( \eta(t) = 0 \)
else
    \( \eta(t) = \zeta(t) \)
\( \Theta(t) = \gamma \cdot \eta(t) \)

Repression functions of modules F, E, and DC mediated by Z site
Both P and CG, needed for synergistic link with module B
Final step up of system output
Positive input from modules B and G
Synergistic amplification of module B output by CG,P subsystem
Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
Repression function inoperative in endoderm but blocks activity elsewhere
Final output communicated to BTA
Signal Transduction Cascades
Protein-Protein Interaction Networks
Protein-Protein Interaction Networks
Host-Pathogen Interactions

Regulation of cell cycle

Non-specific

G1 phase

Mitosis

G2 phase

S phase

Course Structure

Introduction to CSB

Focus of the Course

T. M. Murali

August 18, 2008

CSB@Chalmers