CS 6824: New Directions in Computational Systems Biology

T. M. Murali

January 19, 2011
Course Structure

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

Discuss state-of-the-art research papers.

▶ Lectures
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures
- Class participation
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures
- Class participation
- Final project: either research project or term paper
Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%

Class participation ≠ attendance!
Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%
- Class participation is very important.
Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%
- Class participation is very important.
- Class participation ≠ attendance!
Student Groups For Projects

- Each group has 2-3 members.
- You can form your own groups.
- Try to form groups with students with different backgrounds.
Individual Presentations

- Number of papers: each student and I mutually decide a set of 1–2 (and perhaps 3) papers. You can either present one paper in detail (and summarise others) or give equal importance to all papers.
- Time: present for 45 minutes and expect 30 minutes of questions and discussion during the presentation. Be prepared for some discussions to take over your presentation.
- Prepare your presentation well in advance. Practise multiple times.
- Please give me PDF copies of slides (no Microsoft PowerPoint).
Suggestions on Reading and Presenting Papers

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

- 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?
  - Don’t have to give all details. You can just present the essential ideas.

Read supplementary information. Often has details about the assumptions, the techniques, and the results.
Suggestions on Reading and Presenting Papers

- 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?
  - Don’t have to give all details. You can just present the essential ideas.
- **Read supplementary information.** Often has details about the assumptions, the techniques, and the results.
Final Research Project

- Software + analysis project.
- We will define a project inspired by the papers you present.
- I will discuss list of projects in the next class.
- You can propose a project to me.
- I will meet each group once a month to monitor progress.
- You can use Perl, C, C++, Java, Python, R . . .
Final Research Project

- Software + analysis project.
- We will define a project inspired by the papers you present.
- I will discuss list of projects in the next class.
- You can propose a project to me.
- I will meet each group once a month to monitor progress.
- You can use Perl, C, C++, Java, Python, R . . .
- The software has to run on Linux!
Final Research Project

- Software + analysis project.
- We will define a project inspired by the papers you present.
- I will discuss list of projects in the next class.
- You can propose a project to me.
- I will meet each group once a month to monitor progress.
- You can use Perl, C, C++, Java, Python, R . . .
- The software has to run on Linux!
- If a life science student is part of a software project, biological analysis of the results must play a major role.
Final Term Paper

- Discuss a superset of the papers you present in class or a group of other papers.
- You can propose additional papers.
Final Term Paper

- Discuss a superset of the papers you present in class or a group of other papers.
- You can propose additional papers.
- The electronic document you submit must be in PDF format. Please do not give me a Microsoft Word document.
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

The structure of DNA was first described by James Watson and Francis Crick in 1953. The double helix model proposed by them explained how genetic information is stored in the form of DNA. The structure consists of two strands of nucleotides that are wound around each other in a helical fashion. Each strand is composed of four types of nucleotides: adenine (A), thymine (T), cytosine (C), and guanine (G). The A-T and C-G pairs are complementary, meaning that each nucleotide on one strand has a corresponding nucleotide on the other strand. This complementarity is crucial for the stability of the DNA double helix, as it ensures that the two strands are held together tightly.

In the context of the text provided, the emphasis is on the importance of understanding the structure of DNA and how it relates to the broader field of computational biology. The reference to rows 3,4, and 5 of Table 3 suggests that this is a scientific paper discussing the implications of DNA structure and function. The mention of the helix axis and the parallel planes of the nucleotides highlights the key features of the DNA double helix that are critical for its role in storing and transmitting genetic information.
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.
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.
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.
### Genome size comparison

<table>
<thead>
<tr>
<th>Species</th>
<th>Chromosomes</th>
<th>Genes</th>
<th>Base pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human (Homo sapiens)</td>
<td>46 (23 pairs)</td>
<td>28-35,000</td>
<td>3.1 billion</td>
</tr>
<tr>
<td>Mouse (Mus musculus)</td>
<td>40</td>
<td>22.5-30,000</td>
<td>2.7 billion</td>
</tr>
<tr>
<td>Puffer fish (Fugu rubripes)</td>
<td>44</td>
<td>31,000</td>
<td>365 million</td>
</tr>
<tr>
<td>Malaria mosquito (Anopheles gambiae)</td>
<td>6</td>
<td>14,000</td>
<td>289 million</td>
</tr>
<tr>
<td>Fruit fly (Drosophila melanogaster)</td>
<td>8</td>
<td>14,000</td>
<td>137 million</td>
</tr>
<tr>
<td>Roundworm (C. elegans)</td>
<td>12</td>
<td>19,000</td>
<td>97 million</td>
</tr>
<tr>
<td>Bacterium* (E. coli)</td>
<td>1</td>
<td>5,000</td>
<td>4.1 million</td>
</tr>
</tbody>
</table>

*Bacterial chromosomes are chromonemes, not true chromosomes.*

*John Blanchard / The Chronicle*
Chimps vs. Humans
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!
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,”
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.
Molecular Biology

- Genomes provide the parts lists (e.g., genes and proteins) but do not directly tell us how these parts fit.
Genomes provide the parts lists (e.g., genes and proteins) but do not directly tell us how these parts fit.

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 organisation of cellular interaction networks.
Systems Biology

- Systems Biology is the study of the parts of the cell, their properties, and their relationships.
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?
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?
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 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.
- Computational analysis can suggest or prioritize wet-lab experiments.
Very important in developmental biology. Many principles of embryo development were discovered in the sea urchin.
Sea Urchin (*Strongylocentrotus purpuratus*)

Very important in developmental biology.

Many principles of embryo development were discovered in the sea urchin.
Sea Urchin (Strongylocentrotus purpuratus)

- Very important in developmental biology.
- Many principles of embryo development were discovered in the sea urchin.
A Cell
A Cell is a Modular
A Cell is a Modular
A Cell is a Modular Network

- **Mat cβ**
- **Micr/Nuc Mat Otx**
- **Ets**

Colored boxes indicate post gastrular domains of expression genes

- **Mesoderm**
- **Endoderm**

Endomeso up to 20-24 hours

Maternal & early interactions

Interactions in definitive territories

Late Wnt8 signal from veg2

Post gastrular terminal or peripheral downstream genes

T. M. Murali  January 19, 2011  CS 6824: New Directions in Computational Systems Biology
A Cell is a Modular Network

C 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:
A Cell is a Modular Network that Computes

B

if (F = 1 or E = 1 or CD = 1) and (Z = 1)
   \[ \alpha = 1 \]
else \[ \alpha = 0 \]
if (P = 1 and CG₁ = 1)
   \[ \beta = 2 \]
else \[ \beta = 0 \]
if (CG₂ = 1 and CG₃ = 1 and CG₄ = 1)
   \[ \gamma = 2 \]
else \[ \gamma = 1 \]
\[ \delta(t) = B(t) + G(t) \]
\[ \varepsilon(t) = \beta^\ast \delta(t) \]
if (\varepsilon(t) = 0)
   \[ \xi(t) = \text{Otx}(t) \]
else \[ \xi(t) = \varepsilon(t) \]
if (\alpha = 1)
   \[ \eta(t) = 0 \]
else \[ \eta(t) = \xi(t) \]
\[ \Theta(t) = \gamma^\ast \eta(t) \]

Repression functions of modules F, E, and DC mediated by Z site
Both P and CG₁ 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 CG₁-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
Network is Complex

Survival Factors (e.g., IGF1)

Chemokines, Hormones, Transmitters (e.g., interleukins, serotonin, etc.)

Growth Factors (e.g., TGFβ, EGF)

Extracellular Matrix

Integrins

cdc42

Fyn/Shc

Ras

Raf

FAK

Src

Dishevelled

GSK-3β

APC

β-catenin

TCF

Wnt

Hedgehog

Fz

Patched

Smo

Cytokines (e.g., EPC)

Cytokine Receptor

JAKs

STAT3,5

Bcl-xL

Cytochrome C

Apoptosis

Caspase 9

Caspase 8

FADD

Bcl-2

Bad

Abnormality Sensor

Bim

Death factors (e.g., Fasl, Tnf)

G protein

G-Protein

PLC

PKC

MEK

MEKK

MAPK

MKK

PKA

G-Protein

Akt

Growth Factors

Myc → Mad: Max ← Max

ERK

JNKs

β-catenin:TCF

Gene Regulation

Cell Proliferation

ACRE

Fos

Jun

ARF

mdm2

p53

p21

Smads

CyclE

CDK2

p27

E2F

CDK4

p15

p16

Rb
Network is Complex
Network is Complex but Very Poorly Understood
Challenges with Molecular Interaction Networks

- Biological data sets and networks are large.
- They are intricate and of very diverse types.
- They are noisy: experiments are error-prone.
- They are highly incomplete. We barely know which genes interact, let alone the detailed kinetics of each interaction.
Promises of Human Genome Project

- Identify numerous novel targets for drug therapy.
- Determine the physiological functions of many proteins.
- Enhance knowledge of the genetic basis of various complex diseases.
- Knowledge of all human genes and haplotypes will lead to a better understanding of individual drug responses.
Challenges in Achieving these Promises

- What are the pathways and genetic programmes that cause diseases?
- What are the functions of human genes and how are they involved in disease processes?
- What are the effects of administering a drug “downstream” of the drug target?
- What genetic and environmental factors cause differences in an individual’s susceptibility to a disease or response to a drug?