CS 5854: Computational Systems Biology

T. M. Murali

January 23, 25, 2023
Course Information

- Meet on Mondays and Wednesdays, 4pm–5:15pm, NCB 110B.
- Office hours: By appointment.
- Course website: [http://courses.cs.vt.edu/~cs5854](http://courses.cs.vt.edu/~cs5854). Consult this website regularly. Course schedule is subject to change.
- I may use Canvas to post some lectures and some papers.
- Mailing list: class-cs-5854-20668-202301-g@vt.edu
Course Pre-requisites

- Conditioned on your background.
- Computer science or a quantitative science
  - Expect you to be proficient in algorithms and programming.
  - Taking “Biological Paradigms in Bioinformatics” will be very helpful.
- Life science
  - Expect you to be proficient in genetics, molecular and cell biology.
  - Taken “Computation for the Life Sciences” or an equivalent course that has taught basic programming.
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 (group)
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (group)
- Class participation
Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (group)
- Class participation
- Final project (group)
Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%
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!
Paper Presentations

- Each presentation group has two students but you are welcome to work in larger groups to read, understand, and discuss papers.
- Each group will present one–two papers.
- I will propose a slate of papers. Groups can vote on top choices.
- Many papers will require two full classes, i.e., a total of 150 minutes, including time for questions.
- 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) to post on the course web page.
- Papers can be complex: prepare reading notes for the other students to guide them through the papers you are presenting.
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.
- I am happy to help you with creating groups.
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 first few weeks.
- You can propose a project to me.
- I will meet each group once a month to monitor progress.
- You can use any programming language.
Course Structure  |  Introduction to CSB  |  Genomes to Networks  |  Focus  |  Papers

**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 first few weeks.**
- **You can propose a project to me.**
- **I will meet each group once a month to monitor progress.**
- **You can use any programming language.**
- **If a life science student is part of a software project, biological analysis of the results must play a major role.**

T. M. Murali  
January 23, 25, 2023  
CS 5854: Computational Systems Biology
Sources of Information

- I do not use a textbook for the course but there are several useful/related books:
Sources of Information

- I do not use a textbook for the course but there are several useful/related books:
  - *Computational Molecular Biology* series, MIT Press.
  - *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical and Computational Scientists 1st Edition*, Nataša Pržulj (Editor), Cambridge University Press, 2019
  - *Computational Modeling of Genetic and Biochemical Networks*, James M. Bower and Hamid Bolouri, MIT Press, 2001
More Sources of Information

- Conferences: ICSB, RECOMB, ISMB, PSB, KDD, machine learning conferences, discrete algorithms conferences.
Molecular Structure of Nucleic Acids

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey.1 They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fiber axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atom it is not clear when forces would hold the structure together, especially as the negatively charged phosphates near the axis will in general repel each other. This van der Waals distance appears to be too small. Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each rolled round the same axis (see diagram). We have made a number of assumptions, namely, that each chain consists of a sequence of phosphate-diester groups joined by 3'-deoxyribose residues with 3,5'-phosphodiesters. The two chains (but not their bases) are related by a dyad perpendicular to the fiber axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Fuerberg's model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms in the sugar chain is close to Fuerberg's "standard configuration," the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3 4 A., in the direction.

We have assumed an angle of 36° between adjacent residues in the two chains, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fiber axis is 10 A. As the bases are on the outside, excitons may have access to them. The structure is not an open one, as the van der Waals' contact is very high. At lower water contents we expect the bases to tilt so that their structure would become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the pyrimidine and purine bases. The planes of the bases are perpendicular to the fiber axis. The chains are then joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two sides of the fiber are linked by identical co-ordinates. One of the chains may be chosen as the primary base and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most probable forms (that is, with the hotter rather than the end configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenosine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenosine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. A single chain does not appear to be restricted in any way. However, if only sparsely ordered material is formed, it follows that if the sequence of bases on one chain is given, then the sequence of bases on the other chain is automatically determined. It has been found experimentally that the ratio of the amounts of adenosine to thymine, and the ratio of guanine to cytosine, are always in the same ratio for all deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data2 on deoxy-ribose nucleic acid are insufficient to confirm our structure. So far we can see, it is roughly compatible with the experimental data, and can be regarded as unproved until it has been checked against more exact results. Further these are given in the following communications. We were not aware of the details of the results presented by any other workers when we devised our structure, which rests mainly though not entirely on published experiments and on chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. Each detail of the structure, including the conditions assumed in building it, together with a set of base ratios that give the basic helix pattern to show how.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on intermolecular forces. We are indebted to Dr. J. D. Bernal for a knowledge of the general nature of the unphosphated deoxyribose as a long chain molecule. Finally, Dr. R. E. Branson and their co-workers at King's College, London, have been aided by a grant from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CHUCK

Medical Research Council Unit for the Study of the Molecular Structure of Nucleic Acids, Cavendish Laboratory, Cambridge;

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Antibody) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and evidence for the van der Waals contact. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter), whether in living cells, in isolated nucleic acid, in animals or in plants, and in purified nucleic acid. The same linear groups of polymers which cohere in crystal parallel to preferential planes in different ways to give crystalline, semi-crystalline or amorphous material. In all cases the X-ray diffractogram consists of two regions, one, determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequences of different nitrogen bases along the chain are not resolved.

Figure 1. Fibers of deoxyribose nucleic acid from E. coli. Fibers axial vertical.

Fig. 1. Diffraction pattern of crystals of helical configuration in solution of deoxyribose nucleic acid. The open circles represent the positions of the nucleotides in the helix. The closed circles are placed on the straight lines parallel to the helix axis, and the horizontal lines are spaced by the base-stacking distance of 3 4 A.
The Human Genome Project

Before: human genome has about 100,000 genes.

After: human genome has about 30,000 genes.

T. M. Murali
January 23, 25, 2023
CS 5854: Computational Systems Biology
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.
<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

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,"

What Factors Differentiate Various Species?

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

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: what are the parts of the cell? what functions does each part perform?

Keith Haring, *Untitled*, 1986

→

Urs Wehrli, *Tidying Up Art*, 2003

Systems biology: how do the parts make up the whole? how do genes and their products collectively carry out complex cellular functions?

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
Molecular biology: what are the parts of the cell? what functions does each part perform?

Systems biology: how do the parts make up the whole? how do genes and their products collectively carry out complex cellular functions?

We need to understand how genes, proteins, and other molecules interact with other in different cell states, different tissues, and under different external conditions.
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 life and quantitative scientists.
- Computational analysis can suggest or prioritize wet-lab experiments.

Cells in the Human Body
Cellular Communication: Neuron Firing
**Cellular Communication: Hunger Response**

- **Leptin**: Produced by adipose (fat) tissue, leptin suppresses appetite as its level increases. When body fat decreases, leptin levels fall, and appetite increases.

- **Ghrelin**: Secreted by the stomach wall, ghrelin is one of the signals that triggers feelings of hunger as mealtimes approach. In dieters who lose weight, ghrelin levels increase, which may be one reason it's so hard to stay on a diet.

- **Insulin**: A rise in blood sugar level after a meal stimulates the pancreas to secrete insulin (see Figure 41.3). In addition to its other functions, insulin suppresses appetite by acting on the brain.

- **PYY**: The hormone PYY, secreted by the small intestine after meals, acts as an appetite suppressant that counters the appetite stimulant ghrelin.

---

www.barbellmedicine.com

T. M. Murali
January 23, 25, 2023

CS 5854: Computational Systems Biology
Cellular Communication: Wound Healing

Cellular Signaling

- Video on Cell Signals
- Video on Transcription and Translation
Wnt Signaling

In the Absence of Wnt

"Destruction Complex"

Cell

Axin
APC
GSK
In the Absence of Wnt

Axin  GSK  APC

β-Catenin

"Destruction Complex"

β-Catenin

β-Catenin

Degraded β-Catenin

Cell

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
In the Presence of Extracellular Wnt

Cell

Axin

APC

GSK

β-Catenin

"Destruction Complex"

β-Catenin

β-Catenin

β-Catenin

Degraded β-Catenin

Fzd

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
In the Presence of Extracellular Wnt

Wnt → β-Catenin

Axin, GSK, APC → "Destruction Complex" → β-Catenin

β-Catenin → Degraded β-Catenin

Fzd, Wnt → β-Catenin
In the Presence of Extracellular Wnt

Cell

- Wnt
- Fzd
- Axin
- GSK
- APC
- β-Catenin

Wnt
In the Presence of Extracellular Wnt

![Diagram showing the presence of Wnt and its interactions with Axin, APC, GSK, and β-Catenin in the cell.](image-url)
In the Presence of Extracellular Wnt

Cell

Wnt

Fzd

Axin

GSK

APC

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

TCF/LEF
In the Presence of Extracellular Wnt

Wnt → Axin → APC → GSK → β-Catenin → Fzd → Wnt

β-Catenin

mRNA of target genes

TCF, LEF

Cell
Wnt Pathway in the KEGG Database
Click on “Public Graphs”.

Click on “Public Graphs”.

Search for “KEGG Wnt ranks”.

Click on “KEGG-Wnt-signaling-pathway-with-ranks”.

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
Click on “Public Graphs”.

Search for “KEGG Wnt ranks”.

T. M. Murali
January 23, 25, 2023

CS 5854: Computational Systems Biology
1. Click on “Public Graphs”.
2. Search for “KEGG Wnt ranks”.
3. Click on “KEGG-Wnt-signaling-pathway-with-ranks”.
Open the “Filter nodes and edges” panel on the right.

Set the “Current rank” to “1” and then “Exit”.
Set the “Current rank” to “1” and then “Exit”.
Move the nodes in the network so that you can arrange them similar to the presentation.
In the absence of Wnt:
- Axin
- GSK
- APC
- β-Catenin

"Destruction Complex"

β-Catenin

Degraded β-Catenin

In the presence of Wnt:
- Wnt
- Fzd
- Axin
- DVL
- GSK
- APC
- β-Catenin

MYC

mRNA of target genes

Cyclin D1

β-Catenin

TCF

LEF

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
Interpreting the Wnt Pathway

What do the arrows mean?
- Black arrowhead: Wnt activates LRP5/6
- Red blunt head: DVL inhibits GSK3B
- Black no head: DVL binds to Axin1/2
- Dashed blue: Fzd indirectly binds to DVL

What may happen if the cell makes lots of DVL, e.g., due to a mutation?
- ▶ β-catenin constantly activates TCF/LEF.
- ▶ Cell behaves as if the Wnt pathway is always activated. Can lead to cancer.

Now suppose you want to develop a drug that binds to the Frizzled (FZD) protein. Should the drug activate or inhibit FZD?
- ▶ It should activate FZD.
- ▶ Then FZD will bind to DVL and prevent DVL from inactivating GSK3B.
Interpreting the Wnt Pathway

What do the arrows mean?
- Black arrowhead: Wnt activates LRP5/6
- Red blunt head: DVL inhibits GSK3B
- Black no head: DVL binds to Axin1/2
- Dashed blue: Fzd indirectly binds to DVL

What may happen if the cell makes lots of DVL, e.g., due to a mutation?
- β-catenin constantly activates TCF/LEF.
- Cell behaves as if the Wnt pathway is always activated. Can lead to cancer.
- If you want to develop a drug that binds to the Frizzled (FZD) protein, the drug should activate FZD.
  - Then FZD will bind to DVL and prevent DVL from inactivating GSK3B.
Interpreting the Wnt Pathway

What do the arrows mean?
- Black arrowhead: Wnt activates LRP5/6
- Red blunt head: DVL inhibits GSK3B
- Black no head: DVL binds to Axin1/2
- Dashed blue: Fzd indirectly binds to DVL

What may happen if the cell makes lots of DVL, e.g., due to a mutation?
- $\beta$-catenin constantly activates TCF/LEF.
- Cell behaves as if the Wnt pathway is always activated. Can lead to cancer.
Interpreting the Wnt Pathway

What do the arrows mean?
- Black arrowhead: Wnt activates LRP5/6
- Red blunt head: DVL inhibits GSK3B
- Black no head: DVL binds to Axin1/2
- Dashed blue: Fzd indirectly binds to DVL

What may happen if the cell makes lots of DVL, e.g., due to a mutation?
- β-catenin constantly activates TCF/LEF.
- Cell behaves as if the Wnt pathway is always activated. Can lead to cancer.

Now suppose you want to develop a drug that binds to the Frizzled (FZD) protein. Should the drug activate or inhibit FZD?
Interpreting the Wnt Pathway

What do the arrows mean?
- Black arrowhead: **Wnt activates LRP5/6**
- Red blunt head: **DVL inhibits GSK3B**
- Black no head: **DVL binds to Axin1/2**
- Dashed blue: **Fzd indirectly binds to DVL**

- What may happen if the cell makes lots of DVL, e.g., due to a mutation?
  - β-catenin constantly activates TCF/LEF.
  - Cell behaves as if the Wnt pathway is always activated. Can lead to cancer.

- Now suppose you want to develop a drug that binds to the Frizzled (FZD) protein. Should the drug activate or inhibit FZD?
  - It should activate FZD.
  - Then FZD will bind to DVL and prevent DVL from inactivating GSK3B.
Imagine the Difficulty of Interpreting this Network!
Cellular Response to External Signals


T. M. Murali
January 23, 25, 2023
CS 5854: Computational Systems Biology
A Cell is Like

YOU DON'T GET TO
500 MILLION FRIENDS
WITHOUT MAKING
A FEW ENEMIES

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
A Cell is Like Facebook

Motility Circuits
- proteases
- adjacent cells
- E-cadherin
- b-catenin
- TCF4

Proliferation Circuits
- growth factors
- tyrosine kinases
- receptor
- Ras
- Myc
- hormones
- survival factors
- cytokines
- abnormal serine

Cytostasis and Differentiation Circuits
- anti-growth factors
- p16
- cyclin D
- pRb
- Smads
- DNA damage
- E2F
- p21

Facebook helps you connect and share with the people in your life.

T. M. Murali
January 23, 25, 2023
CS 5854: Computational Systems Biology
Sea Urchin (Strongylocentrotus purpuratus)

Very important in developmental biology. Many principles of embryo development were discovered in the sea urchin.

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
Sea Urchin (*Strongylocentrotus purpuratus*)

Very important in developmental biology. Many principles of embryo development were discovered in the sea urchin.

T. M. Murali January 23, 25, 2023 CS 5854: Computational Systems Biology
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

[Diagram of cell biology processes]
A Cell is a Modular

T. M. Murali
January 23, 25, 2023
CS 5854: Computational Systems Biology
A Cell is a Modular Network

Mat cβ
Micr/Nuc Mat Otx

4th cleavage
M→V2L
Hnf 6
Repressor

Ets
Dri

Cyclophilin, EpHx, Ficolin, Sm37, Sm27, MSP130L

Mat N.

Reprod. of Wnt8 (18 hrs)

LiCl
GSK-3
frizzled

Endomeso
outside Endomeso
Mat Otx

β α Otx
Eve

Colored boxes indicate post gastrular domains of expression genes

Mesoderm
Endoderm

Wnt8
Krl
SoxB1
Krox

Interactions in definitive territories

Maternal & early interactions

Eve

SU Lim

Late Wnt8 signal from veg2

Post gastrular terminal or peripheral downstream genes

Endo16
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
A Cell is a Modular Network that Computes

\begin{align*}
\text{B} & \\
\text{if } (F = 1 \text{ or } E = 1 \text{ or } CD = 1) \text{ and } (Z = 1) & \quad \alpha = 1 \\
\quad \text{else} & \quad \alpha = 0 \\
\text{if } (P = 1 \text{ and } CG_1 = 1) & \quad \beta = 2 \\
\quad \text{else} & \quad \beta = 0 \\
\text{if } (CG_2 = 1 \text{ and } CG_3 = 1 \text{ and } CG_4 = 1) & \quad \gamma = 2 \\
\quad \text{else} & \quad \gamma = 1 \\
\delta(t) = B(t) + G(t) & \quad \text{Positive input from modules B and G} \\
\epsilon(t) = \beta \delta(t) & \quad \text{Synergistic amplification of module B output by CG}_1 \text{-P subsystem} \\
\text{if } (\epsilon(t) = 0) & \quad \xi(t) = Otx(t) \\
\quad \text{else} & \quad \xi(t) = \epsilon(t) \\
\text{if } (\alpha = 1) & \quad \eta(t) = 0 \\
\quad \text{else} & \quad \eta(t) = \xi(t) \\
\Theta(t) = \gamma \eta(t) & \quad \text{Final output communicated to BTA}
\end{align*}
Network is Complex
Network is Complex
Network is Complex but Very Poorly Understood

Network is Complex but Very Poorly Understood

Costanzo et al., Cell, 2019.
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.
Continuum of Models in Systems Biology

Goals of the Course

- Emphasise a data-driven approach to systems biology.
- Integrate massive quantities of different types of data.
- Stress methods that can prioritise experiments.
- Learn techniques from clustering, data mining, and graph theory and apply them to solve specific biological questions.
- Focus on analysis of COVID-19, network, and single-cell RNA-seq data.
Papers to be Presented

Three broad topics:

1. Pathway reconstruction
2. Gene regulatory network inference
3. Single-cell data analysis
Wnt Signaling in a Pathway Database

www.netpath.org/netslim
Signaling Pathways as Directed Graphs

Bidirected Physical Interactions

Directed Regulatory Interactions

GSK

β-Catenin

CTNNB1
Signaling Pathways as Directed Graphs
Reconstructing Signaling Pathways

Human protein-protein interactome
Reconstructing Signaling Pathways

Curated pathway as a subnetwork of the interactome
Question: Can we reconstruct the curated pathway given only receptors and transcriptional regulators?
Reconstructing Signaling Pathways

Proposed pathway reconstruction


Papers on Pathway Reconstruction

1. **Pathways on demand: automated reconstruction of human signaling networks**, Ritz *et al.*, npj *Systems Biology and Applications*, 2016. Murali will present this paper.

2. **Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data**, Köksal *et al.*, *Cell Reports*, 2018

3. **Growing DAGs: Optimization Functions for Pathway Reconstruction Algorithms**, Köse, Li, and Ritz, bioRxiv, 2022


5. **Mutual interactors as a principle for phenotype discovery in molecular interaction networks**, *Pacific Symp. Biocomputing*, PSB 2023
Signaling Pathways and Gene Expression

Motility Circuits

Cytostasis and Differentiation Circuits

antigrowth factors

Proliferation Circuits

growth factors

receptor tyrosine kinases

hormones

survival factors

cytokines

abnormality sensor

Viability Circuits

changes in gene expression

Hallmark capabilities

T. M. Murali
January 23, 25, 2023

CS 5854: Computational Systems Biology
Gene Expression is a Dynamic Process

\[ B \]

if \((F = 1 \text{ or } E = 1 \text{ or } CD = 1) \text{ and } (Z = 1)\)
\[ \alpha = 1 \]
else \[ \alpha = 0 \]
if \((P = 1 \text{ and } CG_1 = 1)\)
\[ \beta = 2 \]
else \[ \beta = 0 \]
if \((CG_2 = 1 \text{ and } CG_3 = 1 \text{ and } CG_4 = 1)\)
\[ \gamma = 2 \]
else \[ \gamma = 1 \]
\[ \delta(t) = B(t) + G(t) \]
\[ \epsilon(t) = \beta * \delta(t) \]
if \((\epsilon(t) = 0)\)
\[ \xi(t) = Otx(t) \]
else \[ \xi(t) = \epsilon(t) \]
if \((\alpha = 1)\)
\[ \eta(t) = 0 \]
else \[ \eta(t) = \xi(t) \]
\[ \Theta(t) = \gamma * \eta(t) \]

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 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
Cellular Differentiation

- Sperm
- Egg
- Fertilized egg
- Stem cell
- Progenitor cell
- Progenitor cell
- Progenitor cell
- Progenitor cell
- Neuron
- Astrocyte
- Nervous tissue cells
- Skin cells
- Sebaceous gland cell

Cells in different states express different sets of genes.

Cells move from one “state” to another.

*Krumsiek et al. (2010). “Hierarchical Differentiation of Myeloid Progenitors...” PLoS ONE*
Transcription factors activate/inhibit genes to effect cell transition from one state to another.

Gene Regulatory Network (GRN)

Gene Regulatory Network (GRN)

How do we build GRNs using computational techniques?


T. M. Murali
January 23, 25, 2023
Bulk RNA Sequencing

- A population of cells isolated at the same time may correspond to multiple, distinct intermediate differentiation states.
- Averages gene expression and masks cellular heterogeneity.
- Difficult to experimentally purify cells in intermediate states.
Single-cell RNA Sequencing (scRNA-seq)

- Produce thousands of independent measurements.
- Computational ordering of cells along “lineages” provide a high-resolution “pseudotemporal” view of gene expression kinetics.
- Richness of these datasets may facilitate inference.

10x Genomics
scRNA-seq Method

(A) Dissolve complex tissue into single cells

Trichoblast
Atrichoblast
Cortex
Endodermis
Pericycle
Procambium

Phloem
Xylem
Lateral root cap
Columnella root cap
Stem cell niche
Quiescent centre

scRNA-seq Method

Droplet generation inside the microfluidic device

scRNA-seq Method

(D) Cell lysis

(E) cDNA hybridised on beads

(F) Barcoded cDNA of thousands of single cells

(G) DGE matrix

Papers on Gene Regulatory Network Inference


2. scSGL: kernelized signed graph learning for single-cell gene regulatory network inference, Karaaslanli et al., *Bioinformatics*, 2022

3. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data, Livnat Jerby-Arnon and Aviv Regev, *Nature Biotechnology*, 2022


Applications of scRNA-seq Data


T. M. Murali

January 23, 25, 2023

CS 5854: Computational Systems Biology
Papers on Single-cell Data Analysis

- Search the literature.
- Find two papers or pre-prints on this topic.
- Select “good journals” and papers by “reputable” authors.
- Send me your list by 5pm, Tuesday, Jan 31, 2023.
Papers on Single-cell Data Analysis

Selected from your choices.


2. Identifying tumor cells at the single-cell level using machine learning, Dohmen *et al.*, *Genome Biology*, 2021


5. Spatial charting of single-cell transcriptomes in tissues, Wei *et al.*, *Nat. Biotech.*, 2022