CS 3824: Introduction to Computational Biology and Bioinformatics

T. M. Murali

August 23, 25, 2022
Course Information

- Meet on Tuesdays and Thursdays, 3:30pm–4:45pm, NCB 110A.
- Office hours: Mondays and Wednesdays, 3pm-5pm, TORG 3160A.
- GTA: Monjuri Rumi.
- Course website: http://bioinformatics.cs.vt.edu/~murali/teaching/2022-fall-cs3824/. Consult this website regularly. Course schedule is subject to change.
- Use Canvas mainly to submit assignments and grades.
- Use Piazza for questions and discussions.
Course Structure

- Lectures based on scientific research papers.
Course Structure

- Lectures based on scientific research papers.
- Programming Assignments (may include some non-programming problems).

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
Course Structure

- Lectures based on scientific research papers.
- Programming Assignments (may include some non-programming problems).
- Final group project (with presentations).
Course Structure

- Lectures based on scientific research papers.
- Programming Assignments (may include some non-programming problems).
- Final group project (with presentations).
- There will be no exams.
Grading

- Assignments: 60%
- Group project: 40%
Student Groups For Projects

- Each group has 2–3 members.
- You can form your own groups.
- I am happy to help you with creating groups.
Final Research Project

- Software + analysis project.
- We will define a project inspired by the lectures.
- I will discuss list of projects by the middle of the semester.
- You can propose a project to me.
- You can use any language you like.
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
Molecular Structure of Nucleic Acids

A Structure for Deoxyribonucleic Acid

We wish to suggest a structure for the salt of deoxyribonucleic 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 Corev. They laid down their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre 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 atoms it is not clear which forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. The van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Ferser (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 deoxyribonucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chain-like assumptions, namely, that each chain consists of a series of phosphate-diester groups joined by 3'-deoxyribose residues with 2',3'-bridges. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains run right-handed helices, but owing to the dyad, the sequences of the bases in the two chains run in opposite directions. Each chain loosely resembles Furbur'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 of phosphorus closely follows Furbur's "standard convention", 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 chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. All the phosphates are on the outside, Osborne may have access to them. The structure is in an open chain and its water content is rather high. At lower water contents we expect the bases to fold so that the structure could 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 fibre axis. They are 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 each side by side with identical co-ordinates. One of the adenine bases is near the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 3 to pyrimidine position 2.

If it is assumed that the bases only occur in the structure in the most probable hydrogen-bonding forms (that is, with the keto rather than the enol configuration) it is found that only the specific pairs of bases can bond together. Those pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine). In other words, if an adenine forms a member of a pair, either adenine then on these assumptions the other adenine must be thymine; similarly for thymine and guanine and cytosine and guanine. The single chain does not appear to be restricted in any way. However, if only specific pairs are 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 adenine to thymine, and the ratio of guanine to cytosine, are always the same for any one deoxyribonucleic acid. This is the same in all species (although the nitrogen base ratios alter with the nature of the substituent) and in cells, and in purified nucleic. The same linear group of polynucleotides may take together parallel in different ways to give crystalline, semi-crystalline or amorphous material. In all cases the X-ray diffraction photograph 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 units along the chains are not made visible.

Oriented paracrystalline deoxyribonucleic acid (structure X in the following communication by Pauling and Corey) gives a fibre diagram as shown in our structure. So far as we can tell, it is roughly compatible with the experimental results in the X-ray diagram, but it would be quite possible that the phase of the experimental results be reversed at the time, and that this provides the information available. There is, however, no evidence from our results that the structure of the X-ray diagram, although not necessarily restricted in the experimental results, is incompatible with the X-ray diagram.

In order to accommodate the X-ray diagram, one must, therefore, assume that the unique properties of the deoxyribonucleic acids are due to the particular ends of the molecule which are not made visible. If the ends of the molecule are allowed to be free, the structure of the X-ray diagram is not incompatible with the experimental results. However, if the ends of the molecule are not free, the structure of the X-ray diagram is incompatible with the experimental results. Therefore, in order to accommodate the X-ray diagram, one must, therefore, assume that the unique properties of the deoxyribonucleic acids are due to the particular ends of the molecule which are not made visible.
Rewind to 1953
Properties of DNA

Video on how DNA works (5 min 24 sec)
DNA is a (very long) string made up of letters A, T, C, and G.

Video on how DNA works (5 min 24 sec)
The Human Genome

- DNA is a (very long) string containing letters A, T, C, and G.
- 3 billion base pairs long.
- End to end length is 2 meters!
DNA is a (very long) string containing letters A, T, C, and G.

- 3 billion base pairs long.
- End to end length is 2 meters!
- The Human Genome Project sequenced the human genome: determined how to spell the genome.
The Human Genome

DNA is a (very long) string containing letters A, T, C, and G.

3 billion base pairs long.

End to end length is 2 meters!

The Human Genome Project *sequenced* the human genome: determined how to spell the genome.

Eric Lander (Nano-Lecture, 2003 Ig Nobel Prize Ceremony):

*Genome. Bought the book, hard to read.*
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.
## 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

Chimp and chimp 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!
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,"

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.
to the System

Keith Haring, *Untitled*, 1986
- Molecular biology: what are the parts of the cell? what functions does each part perform?

Urs Wehrli, *Tidying Up Art*, 2003
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.
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 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

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

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.

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.

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

Cellular Signaling

- Video on Cell Signals (14 min 15 sec)
- Video on Transcription and Translation (11 min 56 sec)
Cellular Response to External Signals

A Cell is Like

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

Cytostasis and Differentiation Circuits
- p16
- cyclin D
- pRb
- E2F
- p21

Proliferation Circuits
- growth factors
- tyrosine kinases
- receptors
- Ras
- Myc
- changes in gene expression

Viability Circuits
- cytokines
- abnormality sensor
- Bcl-2
- DNA damage sensor
- death factors
- hallmark capabilities

T. M. Murali
August 23, 25, 2022
CS 3824: Introduction to Computational Biology and Bioinformatics
A Cell is Like

YOU DON’T GET TO 500 MILLION FRIENDS WITHOUT MAKING A FEW ENEMIES
A Cell is Like

Proteases → adjacent cells → E-cadherin → integrins → extracellular matrix

Motility Circuits:
- Apc → b-catenin → TCF4

Cytostasis and Differentiation Circuits:
- p16 → cyclin D → pRb → E2F → Smads

Proliferation Circuits:
- Growth factors → receptor tyrosine kinases → Ras
- Hormones → Myc
- Survival factors → cytokines → abnormal senescence

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

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

Cytostasis and Differentiation Circuits
- anti-growth factors

Proliferation Circuits
- growth factors
- receptor tyrosine kinases
- Ras
- Myc
- cyclin D
- pRb
- p21

Viability Circuits
- death factors
- abnormality sensor
- Bcl-2

DNA-damage sensor

changes in gene expression
Hallmark capabilities

T. M. Murali
August 23, 25, 2022
CS 3824: Introduction to Computational Biology and Bioinformatics
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.
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
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

T. M. Murali
August 23, 25, 2022
CS 3824: Introduction to Computational Biology and Bioinformatics
A Cell is a Modular Network that Computes

if (F = 1 or E = 1 or CD = 1) and (Z = 1)  
  α = 1  
else  α = 0

if (P = 1 and CG₁ = 1)  
  β = 2  
else  β = 0

if (CG₂ = 1 and CG₃ = 1 and CG₄ = 1)  
  γ = 2  
else  γ = 1

δ(t) = B(t) + G(t)  
ε(t) = β*δ(t)  

if (ε(t) = 0)  
  ξ(t) = Otx(t)  
else  ξ(t) = ε(t)

if (α = 1)  
  η(t) = 0  
else  η(t) = ξ(t)

θ(t) = γ*η(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

Proliferation Circuits
- Growth factors
  - Receptor tyrosine kinases
  - Hormones
- Survival factors
  - Cytokines

Motility Circuits
- Proteases
  - Adjacent cells
    - E-cadherin
  - Extracellular matrix
  - Integrins

Cytostasis and Differentiation Circuits
- Anti-growth factors
  - Smads

Hallmark capabilities
- Changes in gene expression
  - DNA-damage sensor
  - p53

Viability Circuits
- Bcl-2
  - Abnormality sensor
  - Death factors

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
Network is Complex
Network is Complex but Very Poorly Understood

Network is Complex but Very Poorly Understood

Costanzo et al., Cell, 2019.
Wnt Signaling


T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
In the Absence of Wnt

Cell

"Destruction Complex"

Axin, GSK, APC
In the Absence of Wnt

Cell

Axin
GSK
APC
"Destruction Complex"

β-Catenin

β-Catenin

Degraded β-Catenin

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
In the Presence of Extracellular Wnt

Cell

Axin

GSK

APC

"Destruction Complex"

β-Catenin

β-Catenin

Fzd

Degraded β-Catenin
In the Presence of Extracellular Wnt

Cell

Axin
APC
GSK
"Destruction Complex"

β-Catenin

Fzd
Wnt

Wnt

β-Catenin

β-Catenin

Degraded β-Catenin

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
In the Presence of Extracellular Wnt
In the Presence of Extracellular Wnt

Cell

Wnt

Axin

Fzd

GSK

APC

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin
In the Presence of Extracellular Wnt

Wnt → Fzd

Axin

GSK

β-Catenin

APC

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

β-Catenin

TCF/LEF

β-Catenin

Cell
In the Presence of Extracellular Wnt

Cell
Axin
APC
GSK
β-Catenin
Fzd
Wnt
β-Catenin
mRNA of target genes

β-Catenin
TCF
LEF
Wnt Pathway in the KEGG Database
GraphSpace

Welcome to GraphSpace!

The interactive graph sharing website.

- Click on “Public Graphs”.
- Search for “KEGG Wnt ranks”.
- Click on “KEGG-Wnt-signaling-pathway-with-ranks”.

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
1. Click on “Public Graphs”.
2. Search for “KEGG Wnt ranks”.

T. M. Murali
August 23, 25, 2022
CS 3824: Introduction to Computational Biology and Bioinformatics
1. Click on “Public Graphs”.
2. Search for “KEGG Wnt ranks”.
3. Click on “KEGG-Wnt-signaling-pathway-with-ranks”.
Wnt Pathway on GraphSpace

- Open the “Filter nodes and edges” panel on the right.
- Set the “Current rank” to “1” and then “Exit”.

T. M. Murali  August 23, 25, 2022  CS 3824: Introduction to Computational Biology and Bioinformatics
Wnt Pathway on GraphSpace

- 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

- β-Catenin
- β-Catenin
- β-Catenin
- β-Catenin

- TCF
- LEF
- MYC

mRNA of target genes

- Cyclin D1

T. M. Murali August 23, 25, 2022 CS 3824: Introduction to Computational Biology and Bioinformatics
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, 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.
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!
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 Network Biology

Goals of the Course

- Emphasise a data-driven approach to biology.
- Take a network-level view of cellular processes.
- Abstract biological questions into computer science problems.
- Describe graph algorithms to solve these problems.