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).

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
- *Protein Interaction Networks: Computational Analysis*, Aidong Zhang, Cambridge University Press, 2009
- *Computational Modeling of Genetic and Biochemical Networks*, James M. Bower and Hamid Bolouri, MIT Press, 2001

Rewind to 1953

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (*Stockholm*), **2** (11) (1905).

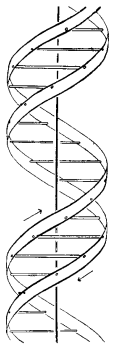
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

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 intertwined 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 what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear 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.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre 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 Furberg's[2] 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 near it is close to Furberg'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 *z*-direction. We have assumed an angle of 36° between adjacent residues in the same 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. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt 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 purine and pyrimidine 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 lie side by side with identical *z*-co-ordinates. One of the pair must be a purine 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 plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for 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 data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).

[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).

[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).

[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).

[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).

[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

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. Astbury[1]) 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 existing in this form when in the natural state. 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 considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline 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 sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the *n*th layer line being proportional to the square of J_n, the *n*th order Bessel function. A straight line may be drawn approximately through

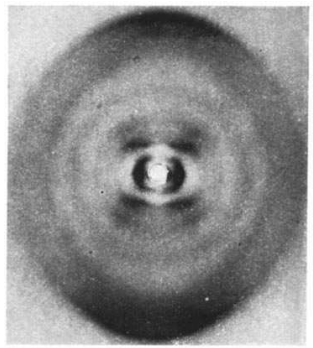


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats *n* times along the helix there will be a meridional reflexion (J_0^2) on the *n*th layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[5] being to reproduce the intensity distribution about the origin around the new origin, on the *n*th layer line, corresponding to *C* in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

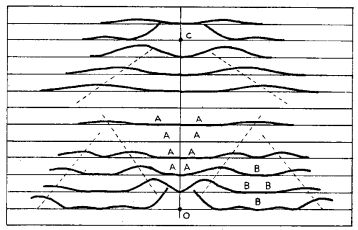
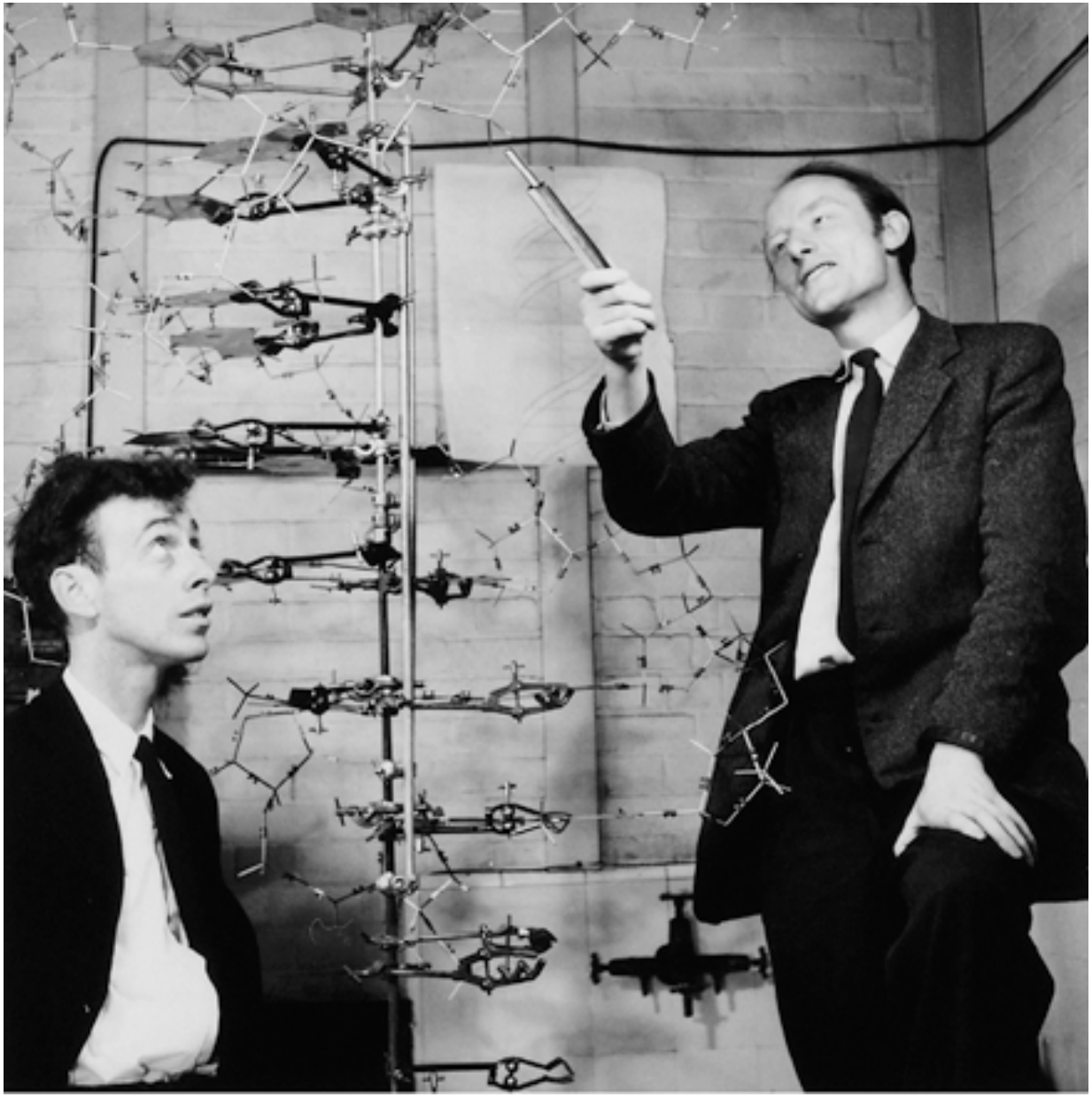
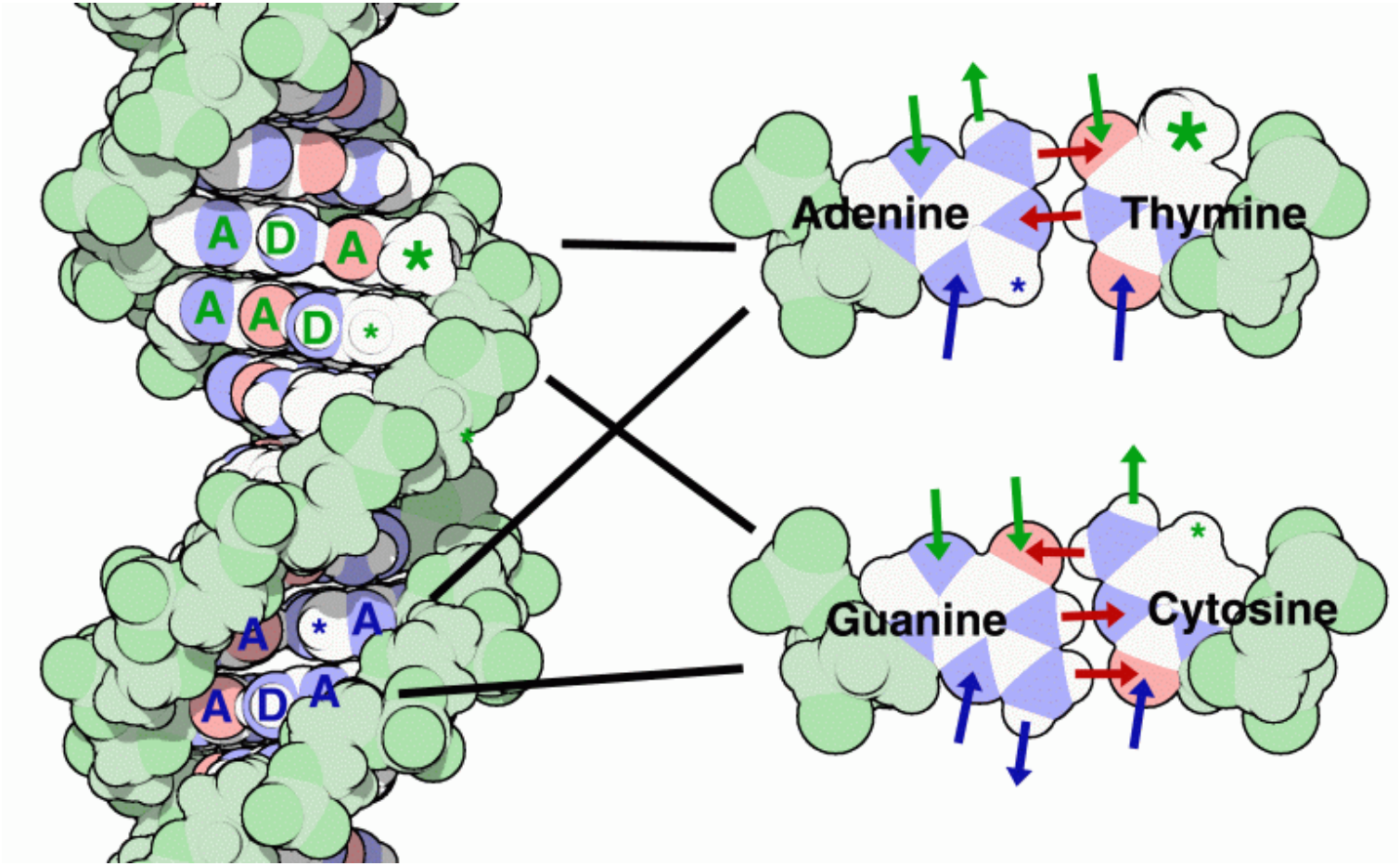


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About *C* on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

Rewind to 1953

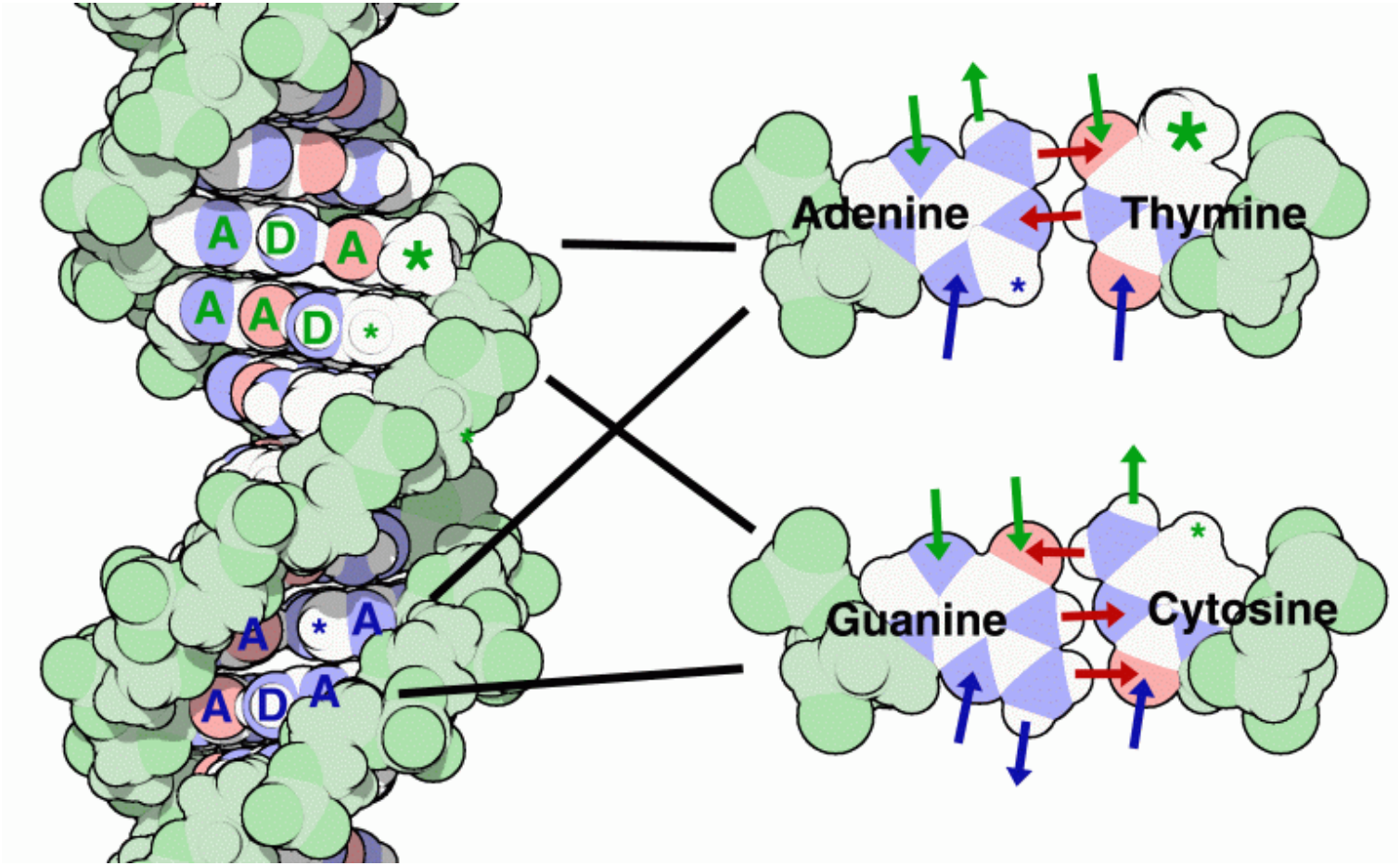


Properties of DNA



▸ Video on how DNA works (5 min 24 sec)

Properties of DNA

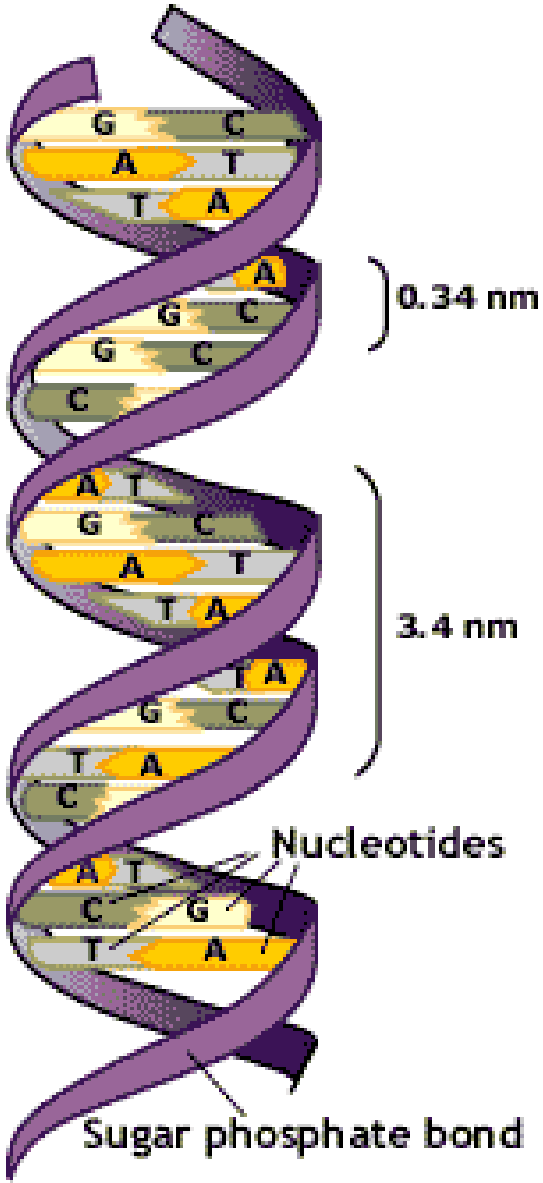


- DNA is a (very long) string made up of letters A, T, C, and G.

▸ Video on how DNA works (5 min 24 sec)

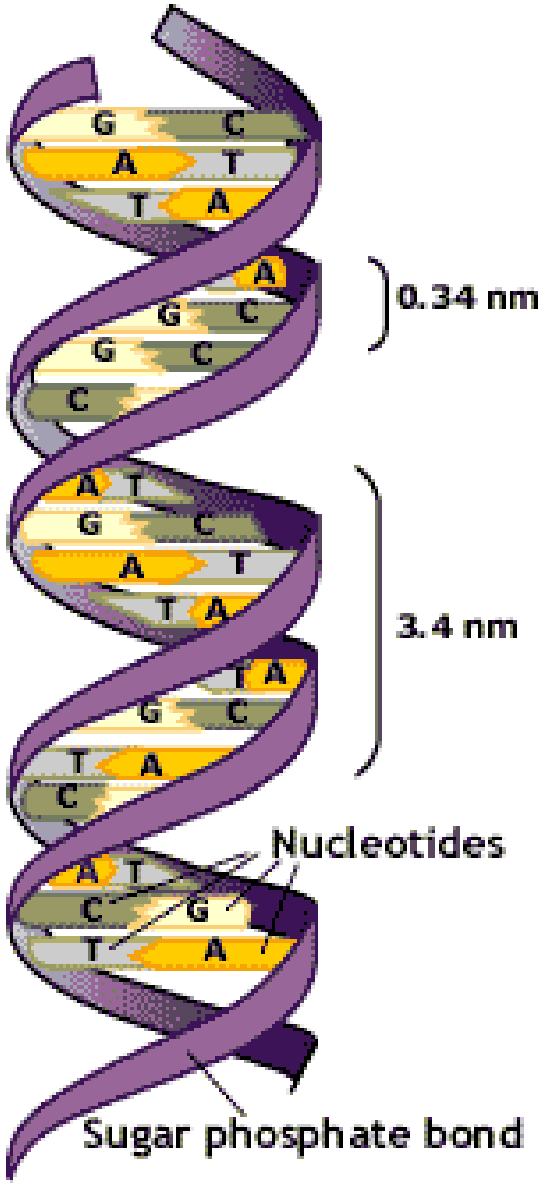
The Human Genome

- 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!



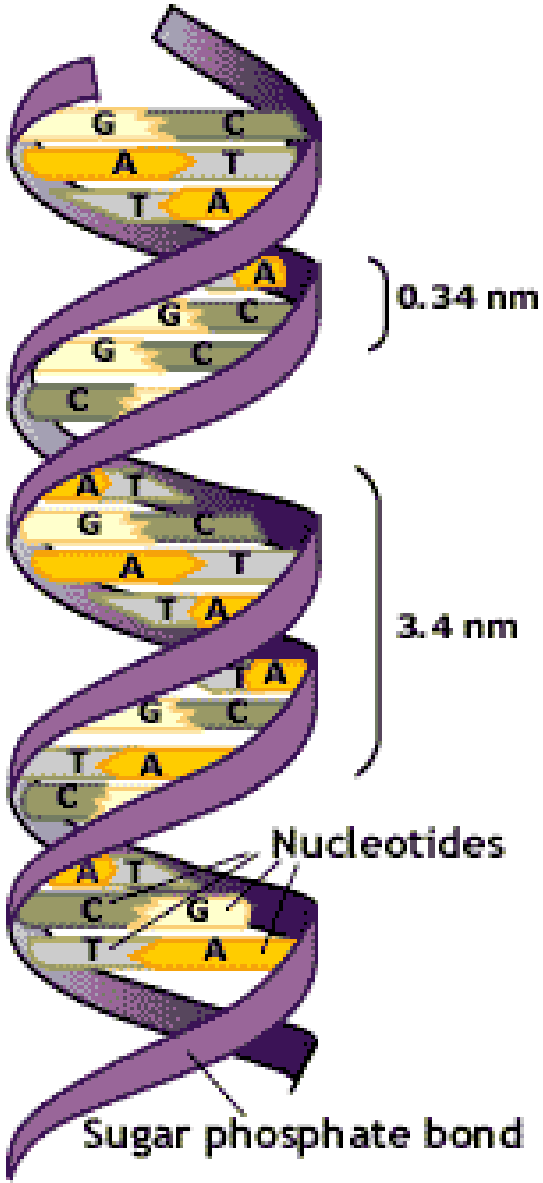
The Human Genome

- 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.



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

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

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

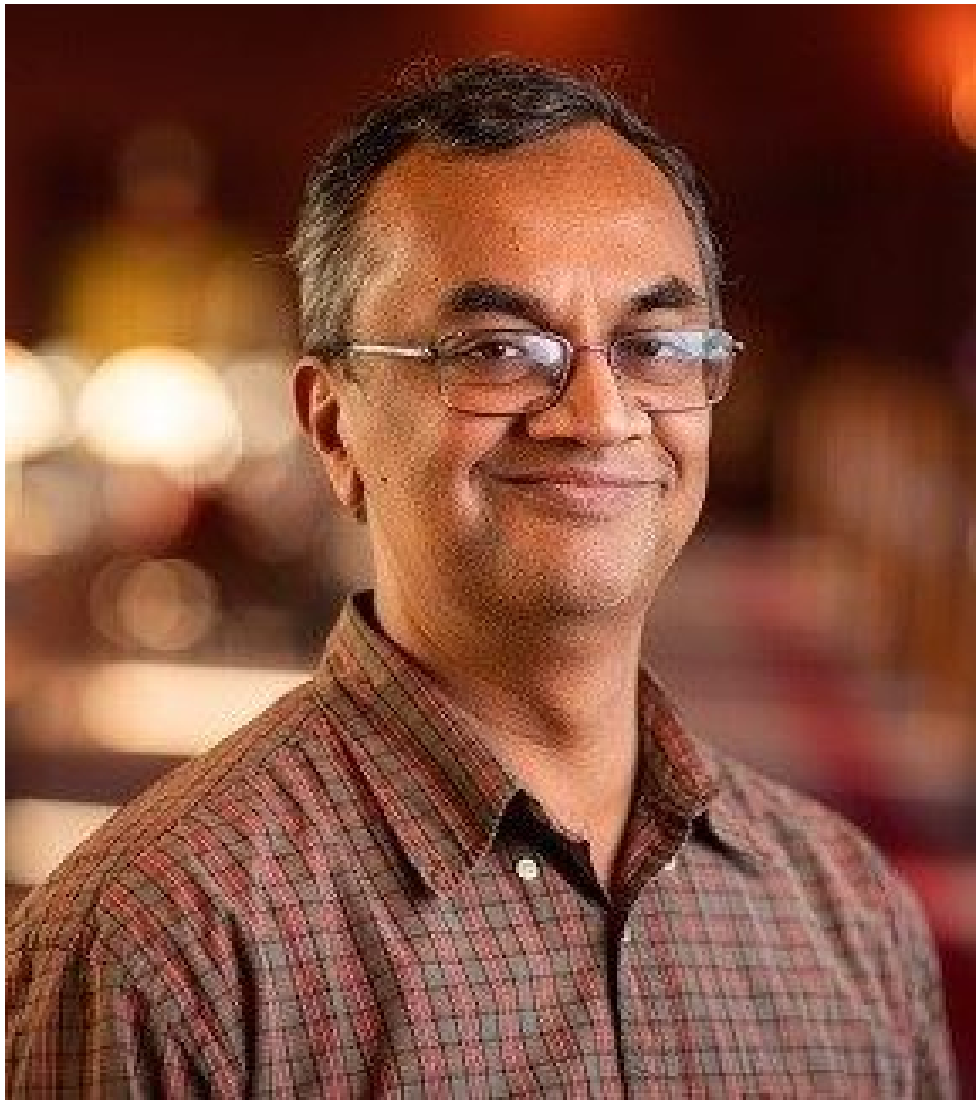
*Bacterial chromosomes are chromonemes, not true chromosomes

JOHN BLANCHARD / *The Chronicle*

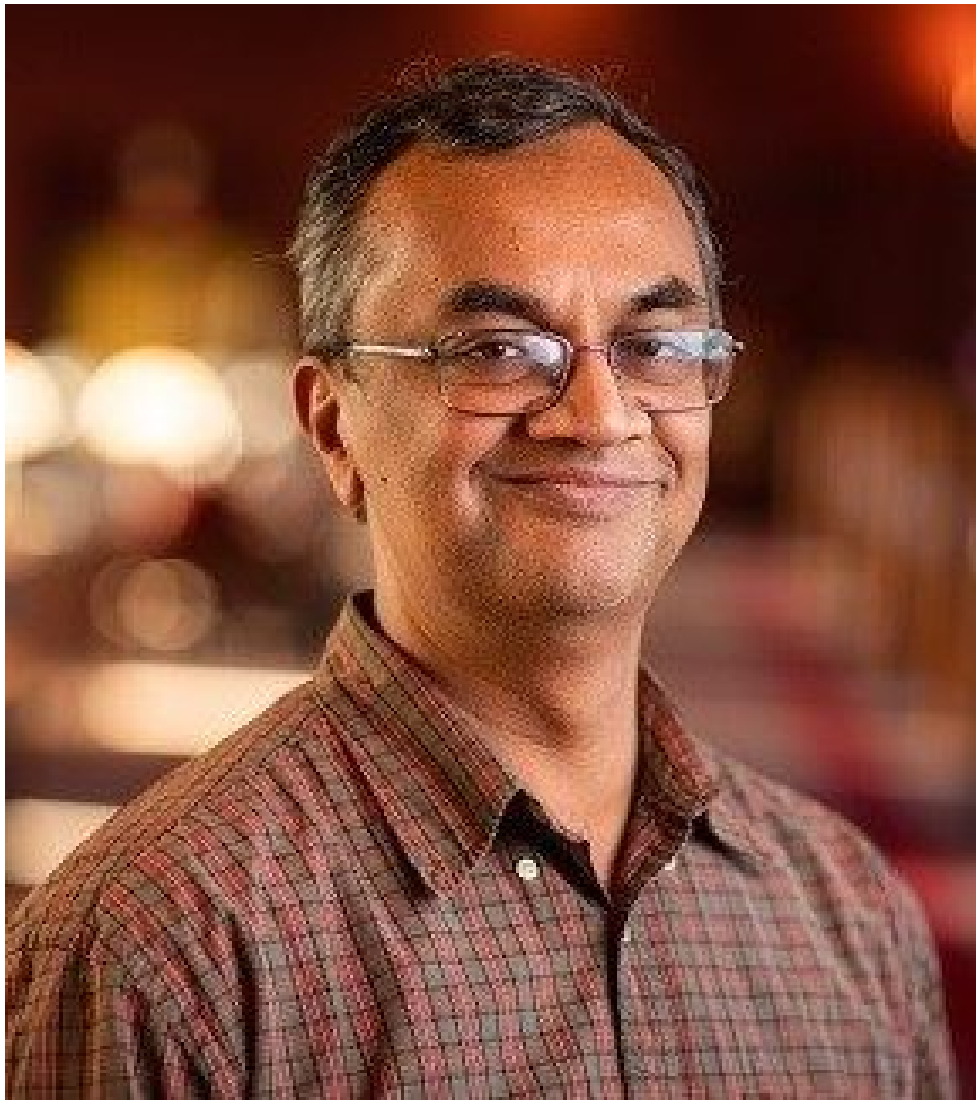
Chimps vs. Humans

Chimps vs. Humans

Chimps vs. Humans



Chimps vs. Humans



Chimp and chump genomes are only about 1.2% different!

What Factors Differentiate Various Species?

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.

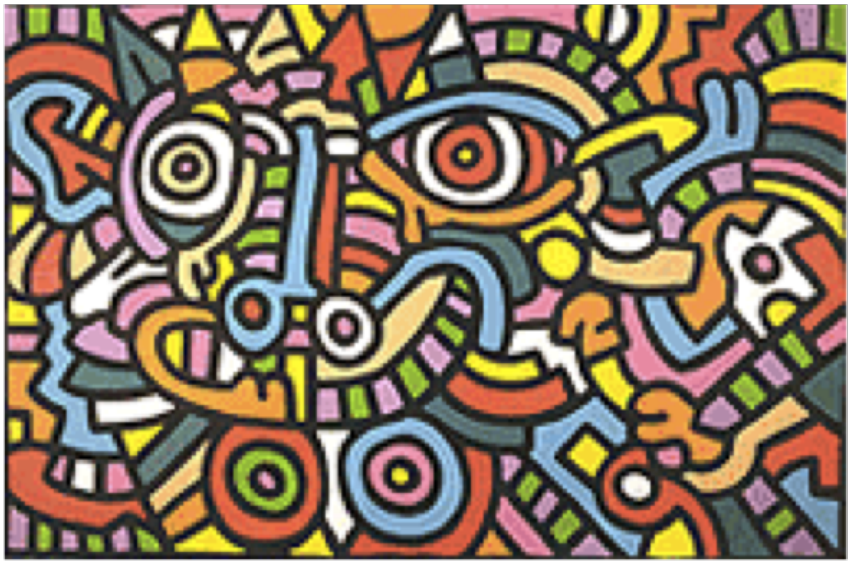
What Factors Differentiate Various Species?

- 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.

BACK to the System

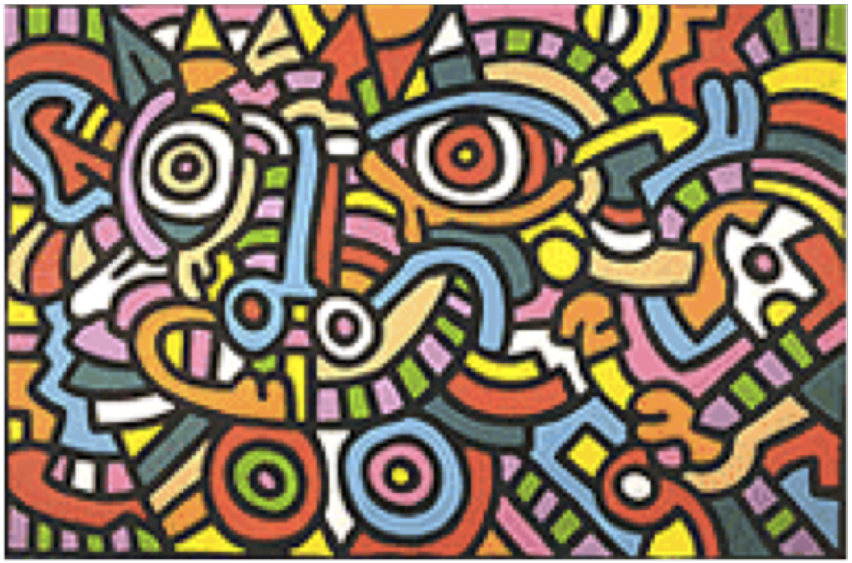


Keith Haring, *Untited*, 1986

Urs Wehrli, *Tidying Up Art*, 2003

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

BACK to the System



Keith Haring, *Untited*, 1986

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?

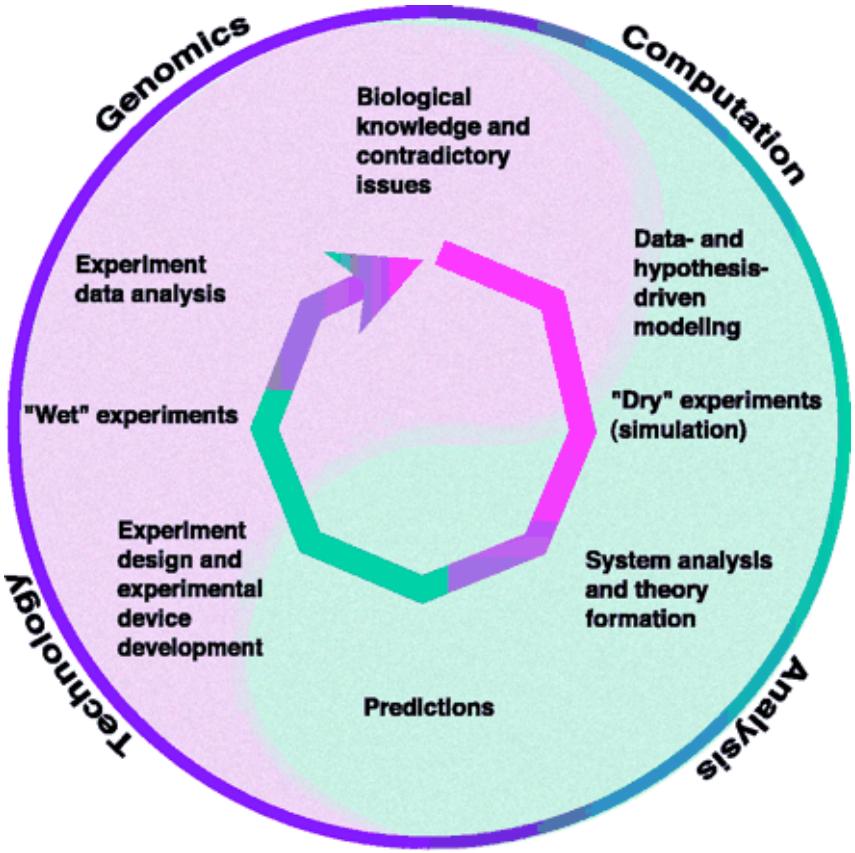
Systems Biology

- 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

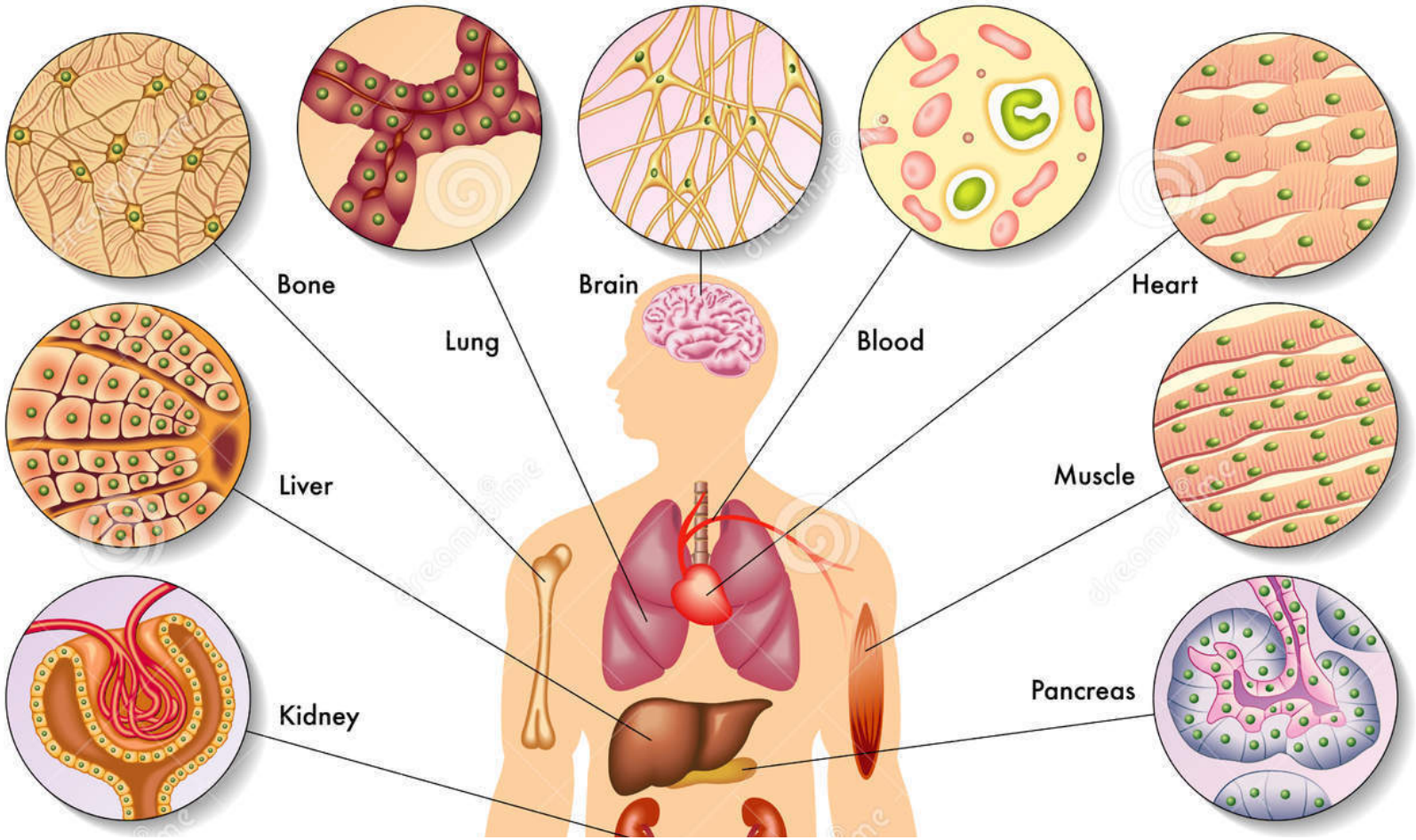


- Modular cell biology (rather than molecular).
- Discovery-driven *and* hypothesis-driven.
- Driven by high-throughput and accurate biological measurements.

- 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.

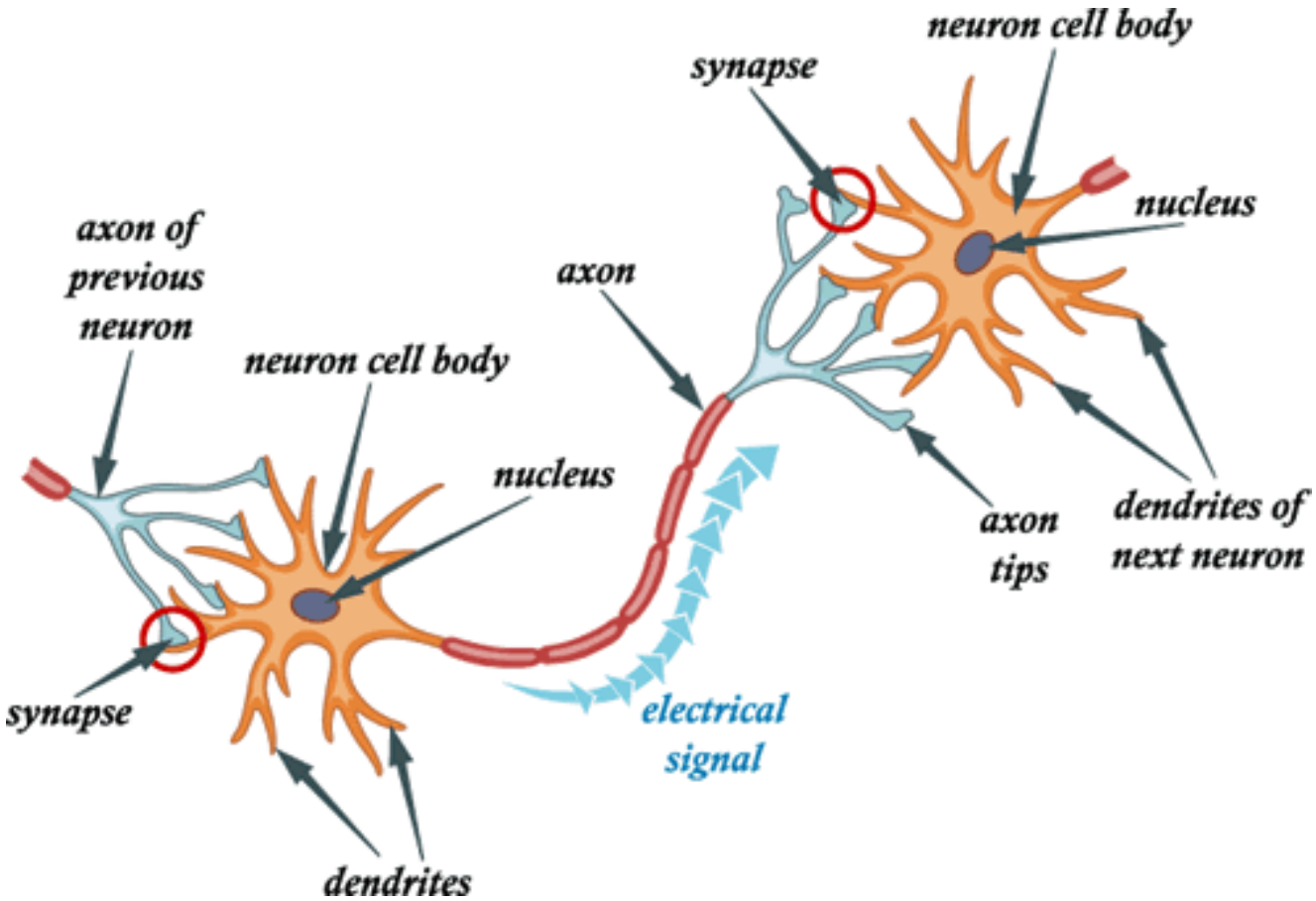
Kitano, Computational Systems Biology, Science, 1999.

Cells in the Human Body



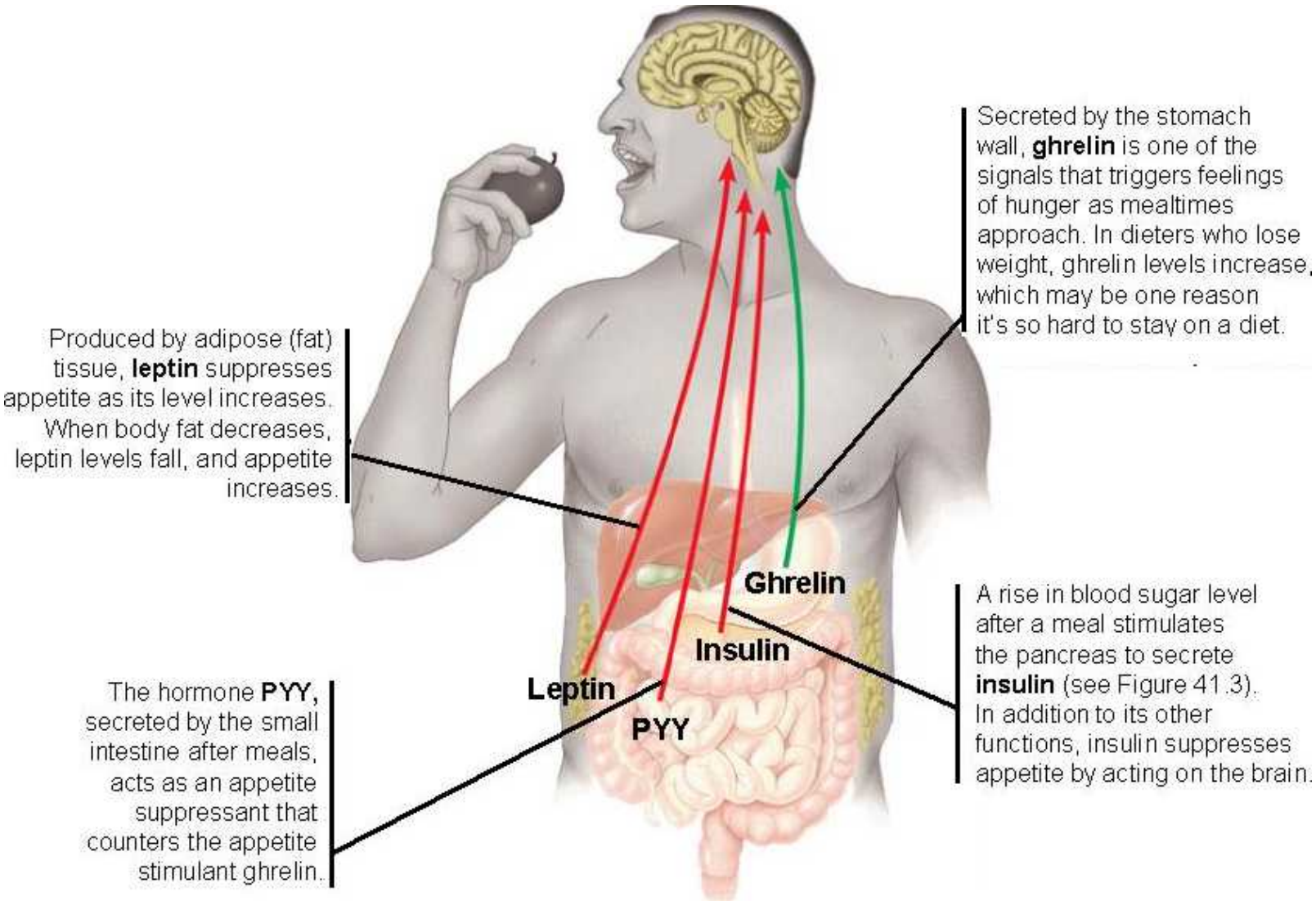
www.thetahealth.com

Cellular Communication: Neuron Firing



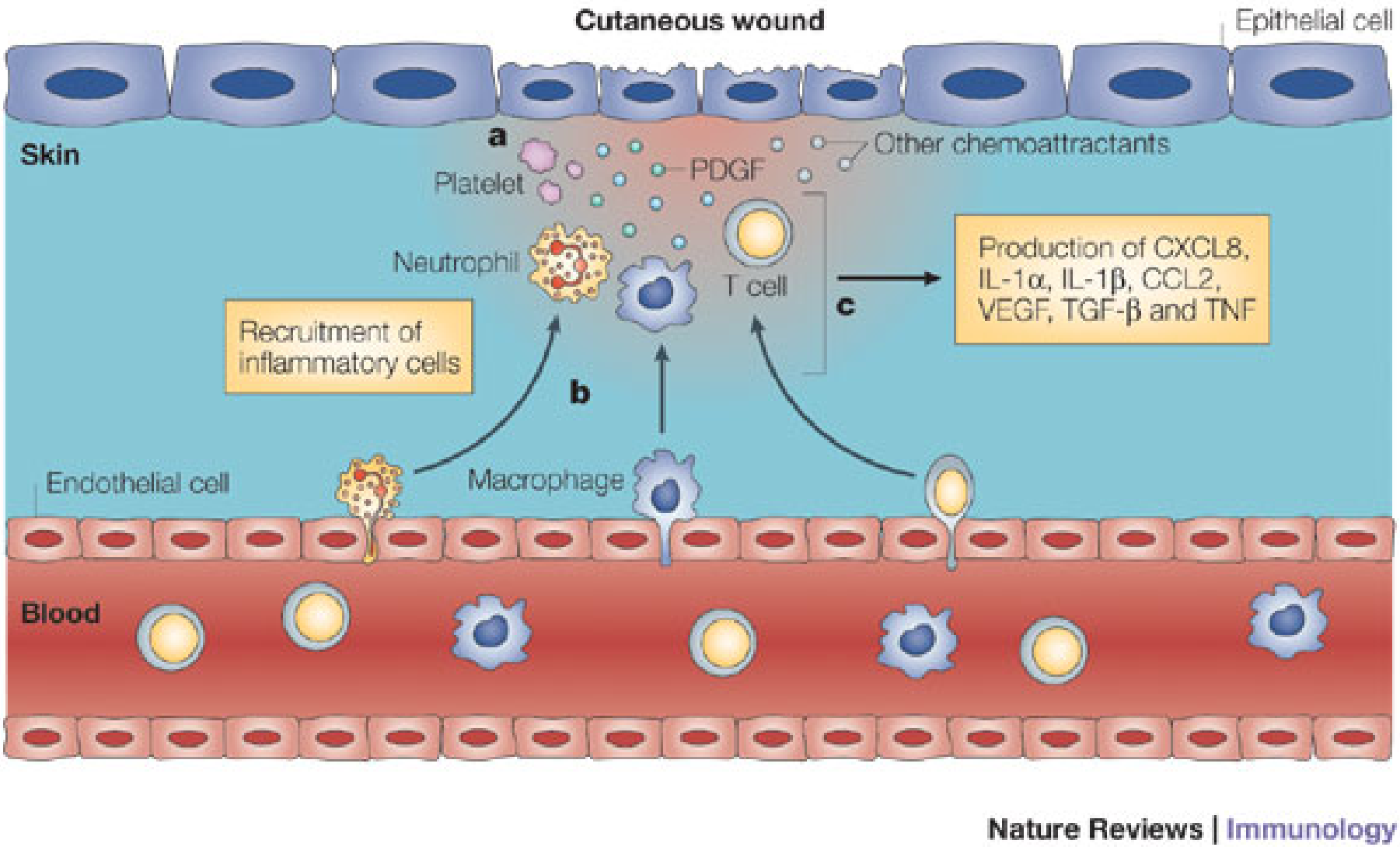
www.jasonshen.com

Cellular Communication: Hunger Response



www.barbellmedicine.com

Cellular Communication: Wound Healing

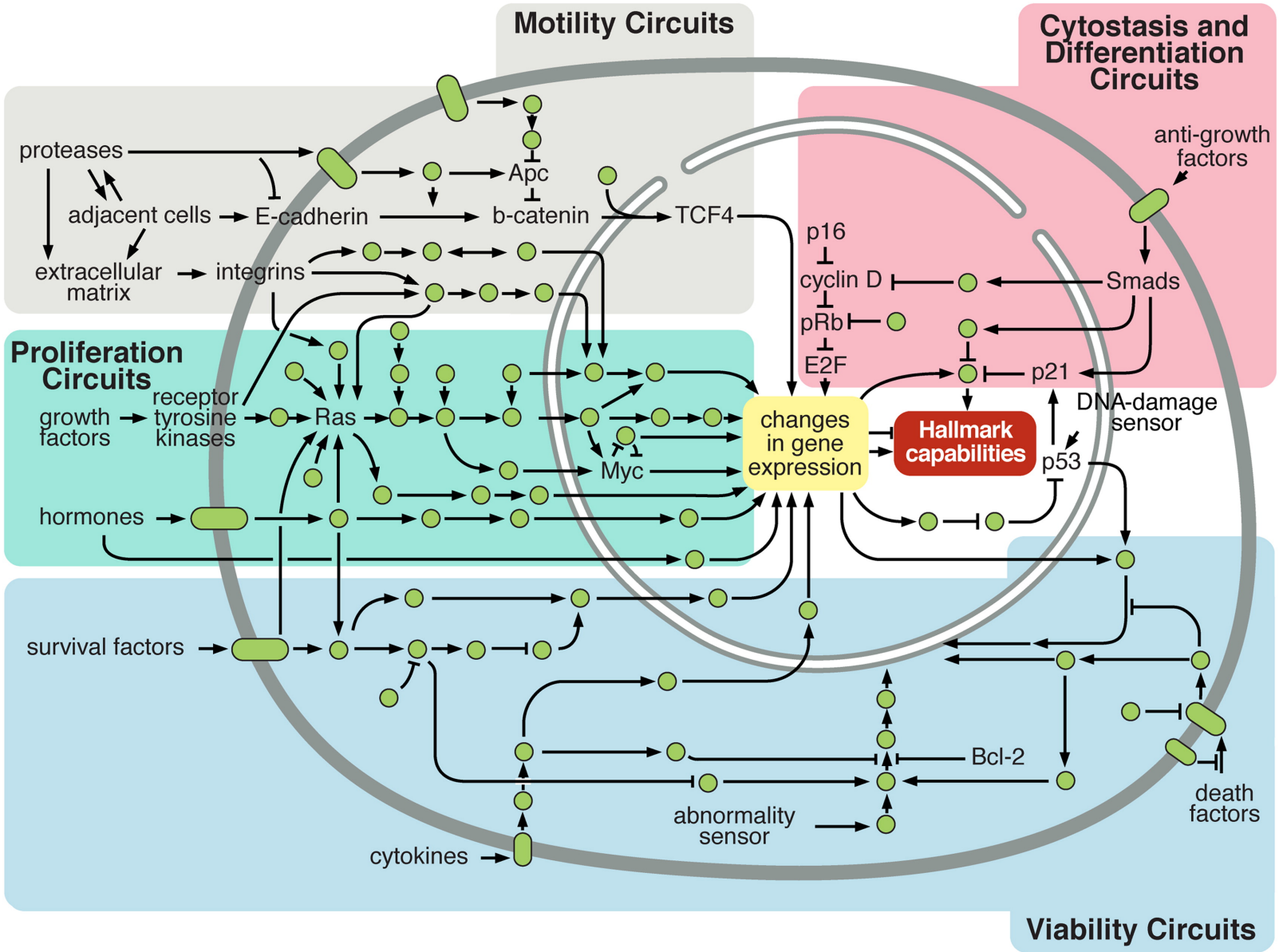


Glaser and Kiecolt-Glaser. Stress-induced immune dysfunction: implications for health. Nature Reviews Immunology 2005.

Cellular Signaling

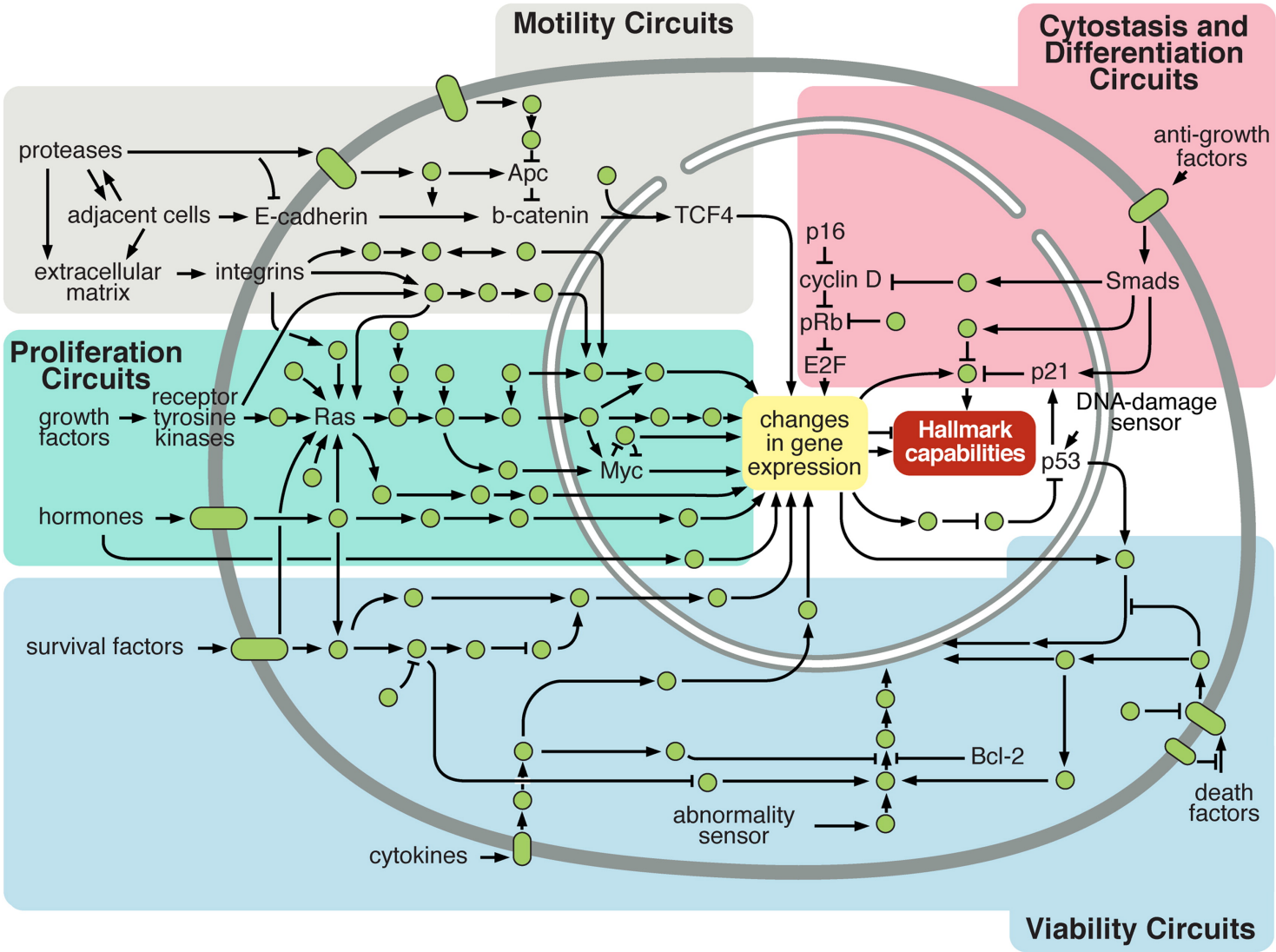
- Video on Cell Signals (14 min 15 sec)
- Video on Transcription and Translation (11 min 56 sec)

Cellular Response to External Signals

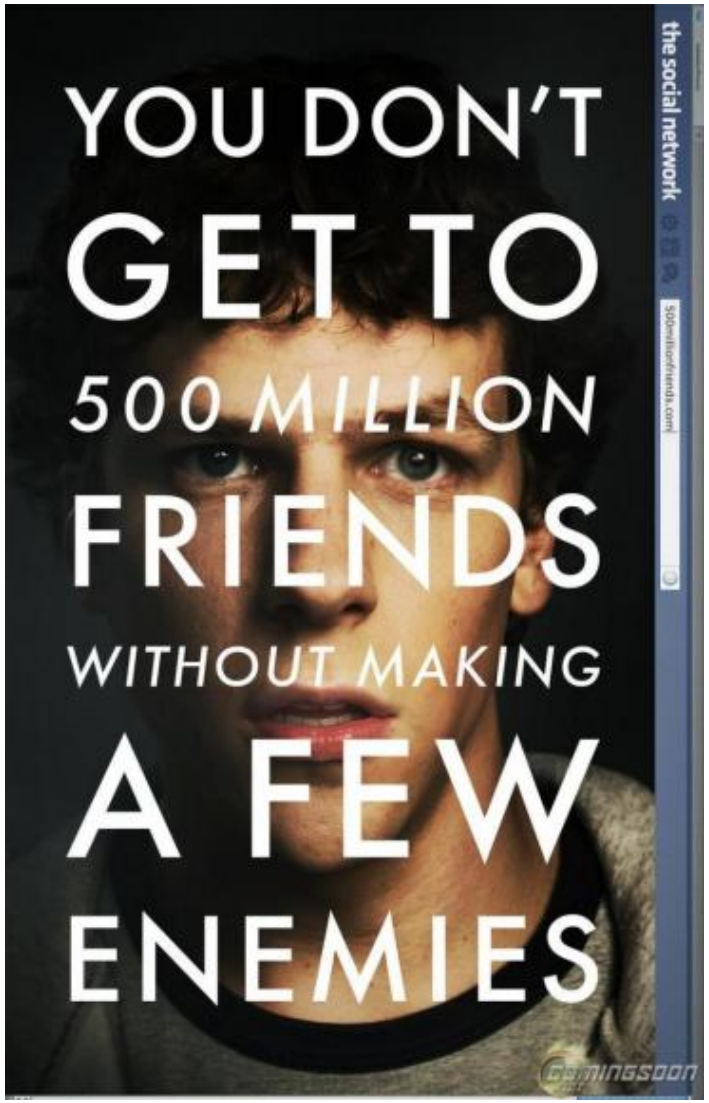


Hanahan and Wienberg. Hallmarks of cancer: the next generation. Cell, 2011.

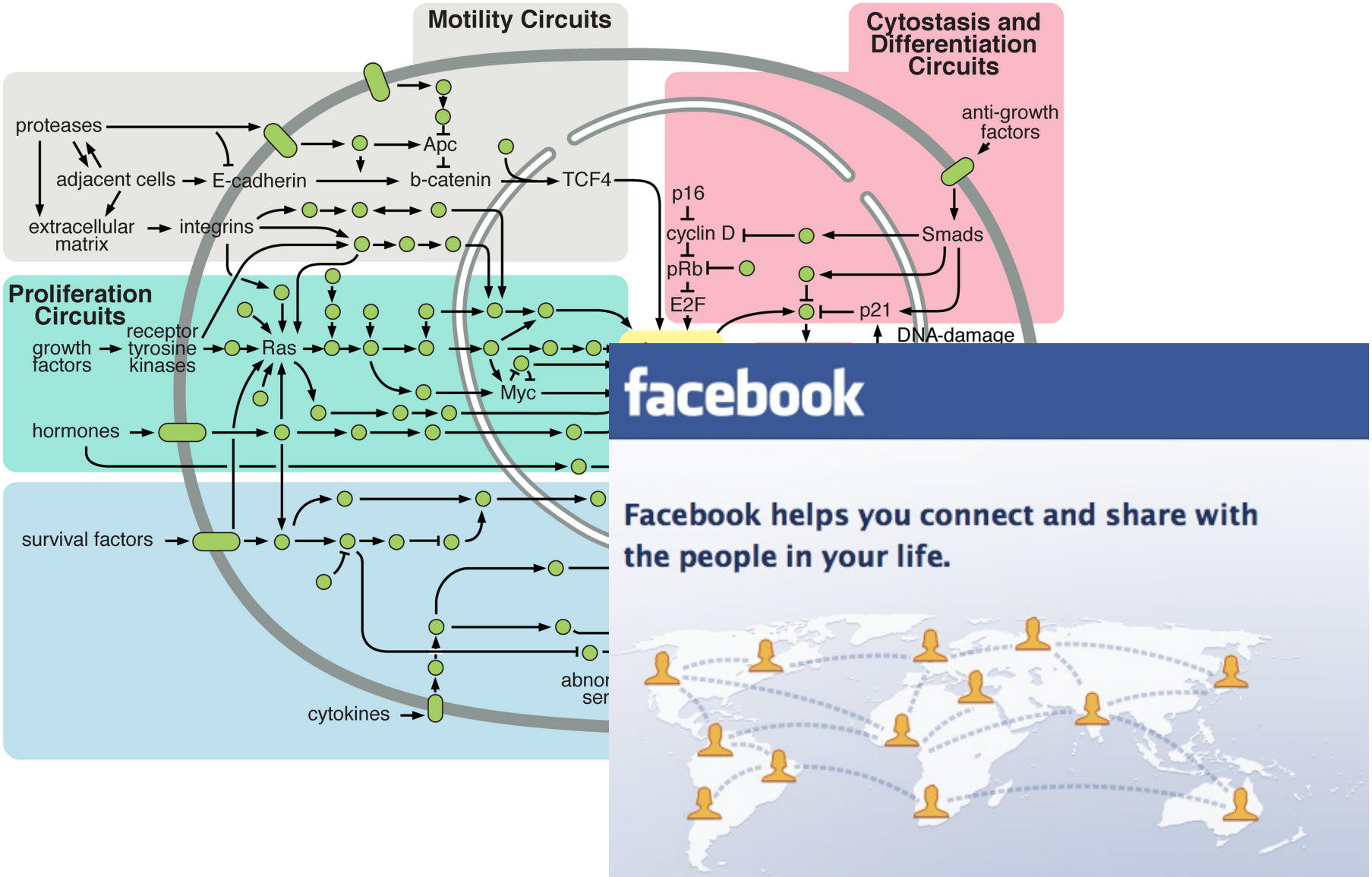
A Cell is Like



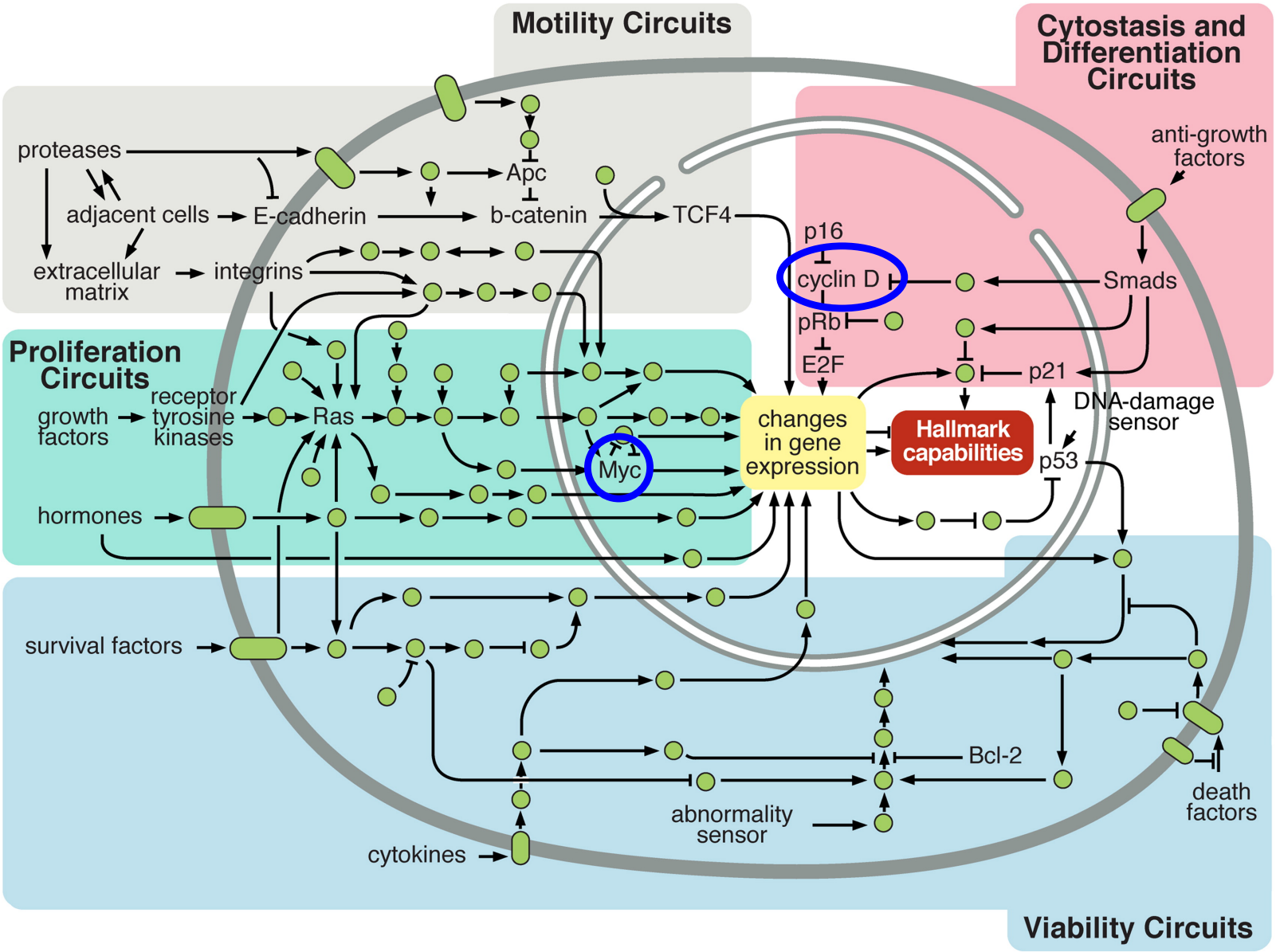
A Cell is Like



A Cell is Like facebook



Hallmarks of Cancer



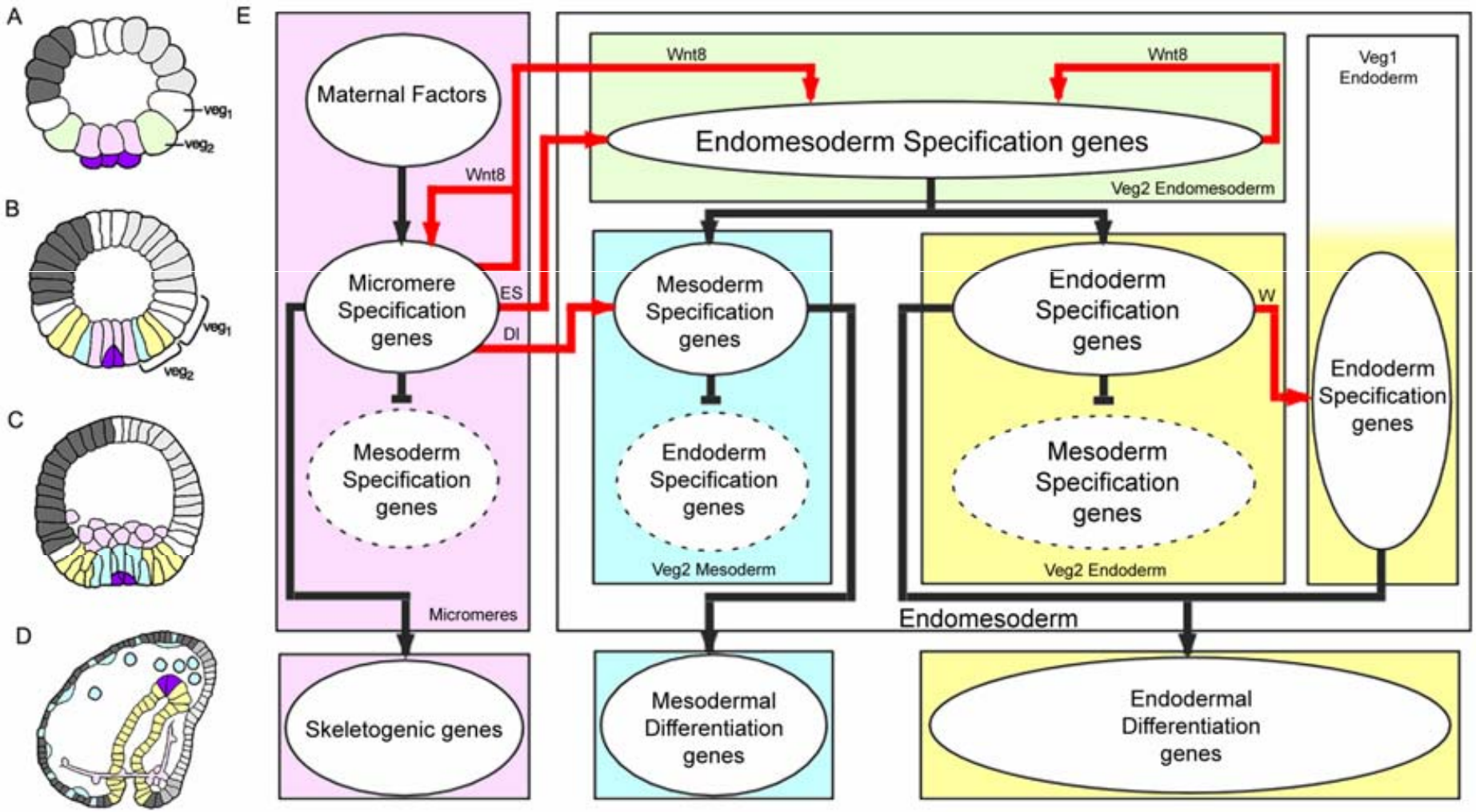
Sea Urchin (Strongylocentrotus purpuratus)

Sea Urchin (Strongylocentrotus purpuratus)

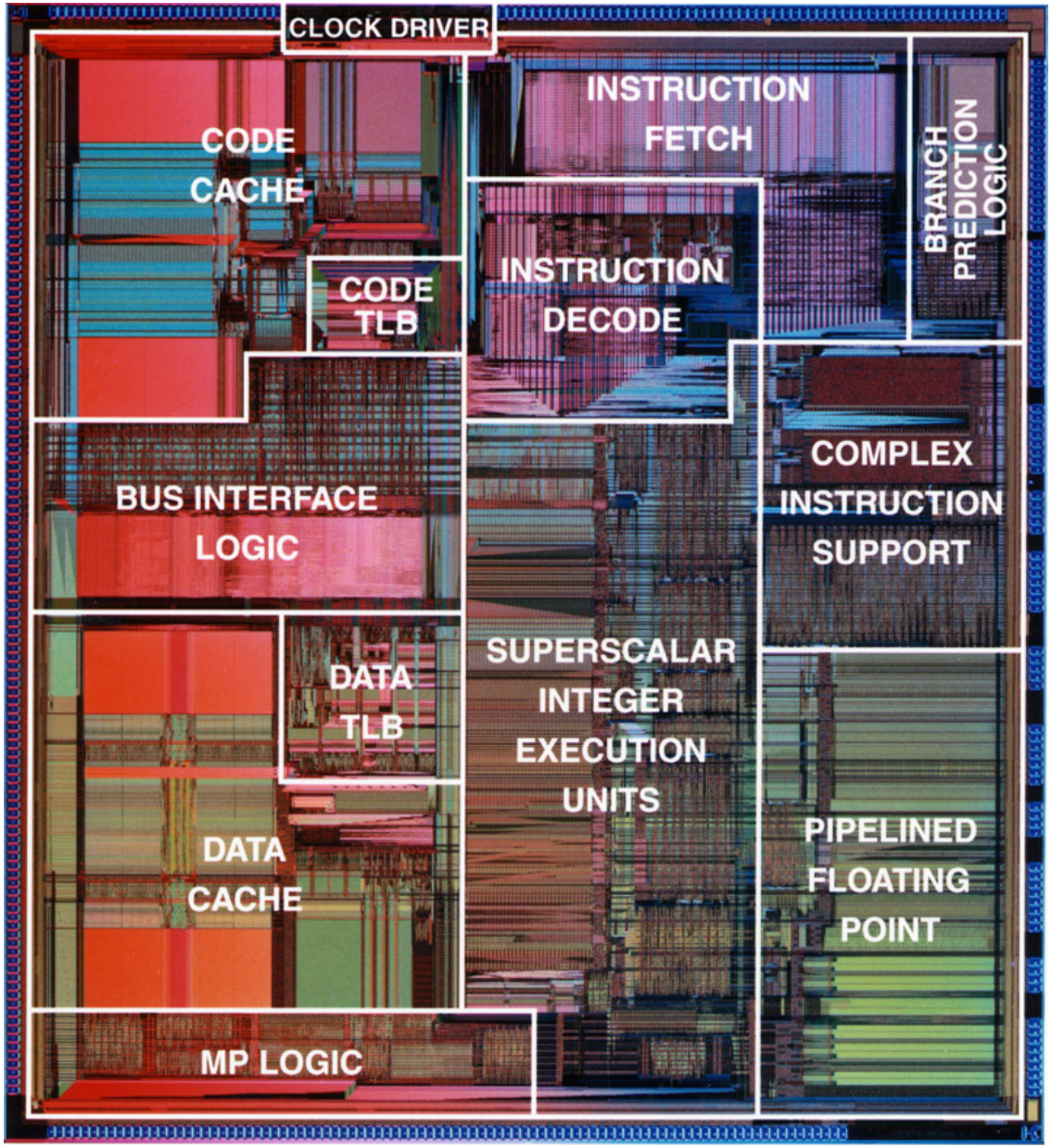
- Very important in developmental biology.
- Many principles of embryo development were discovered in the sea urchin.

A Cell

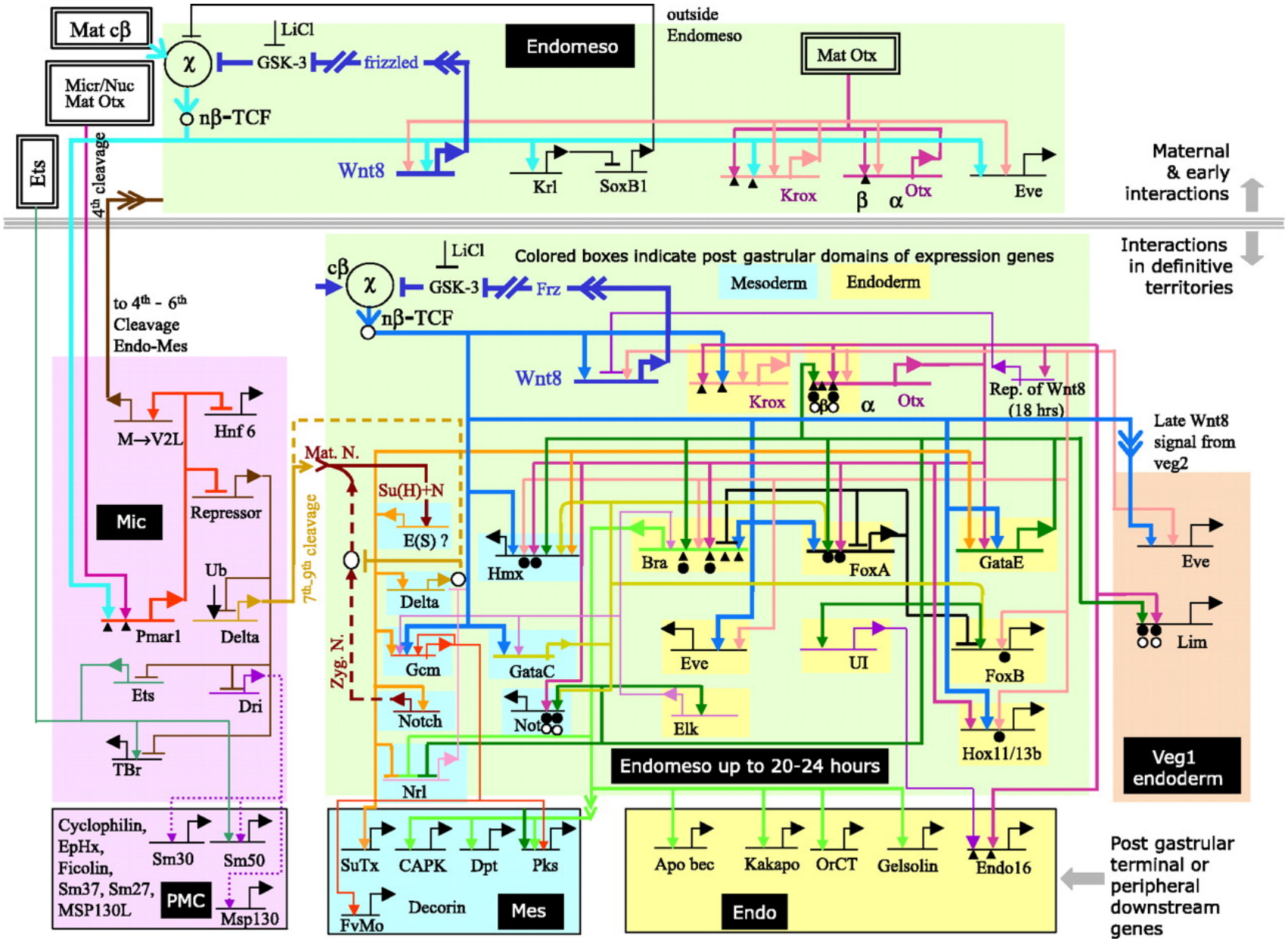
A Cell is a Modular



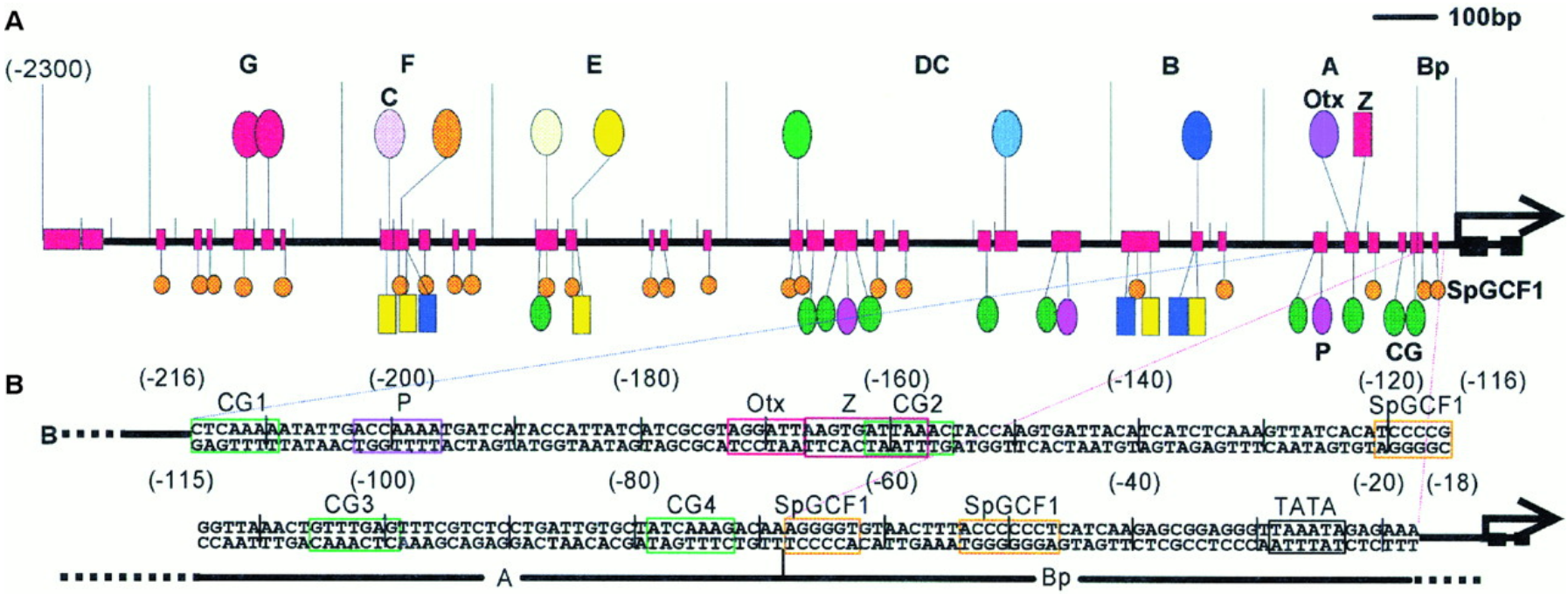
A Cell is a Modular



A Cell is a Modular Network



A Cell is a Modular Network



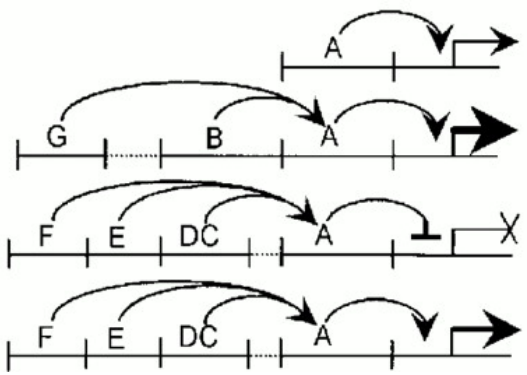
C Module A functions:

Vegetal plate expression in early development:

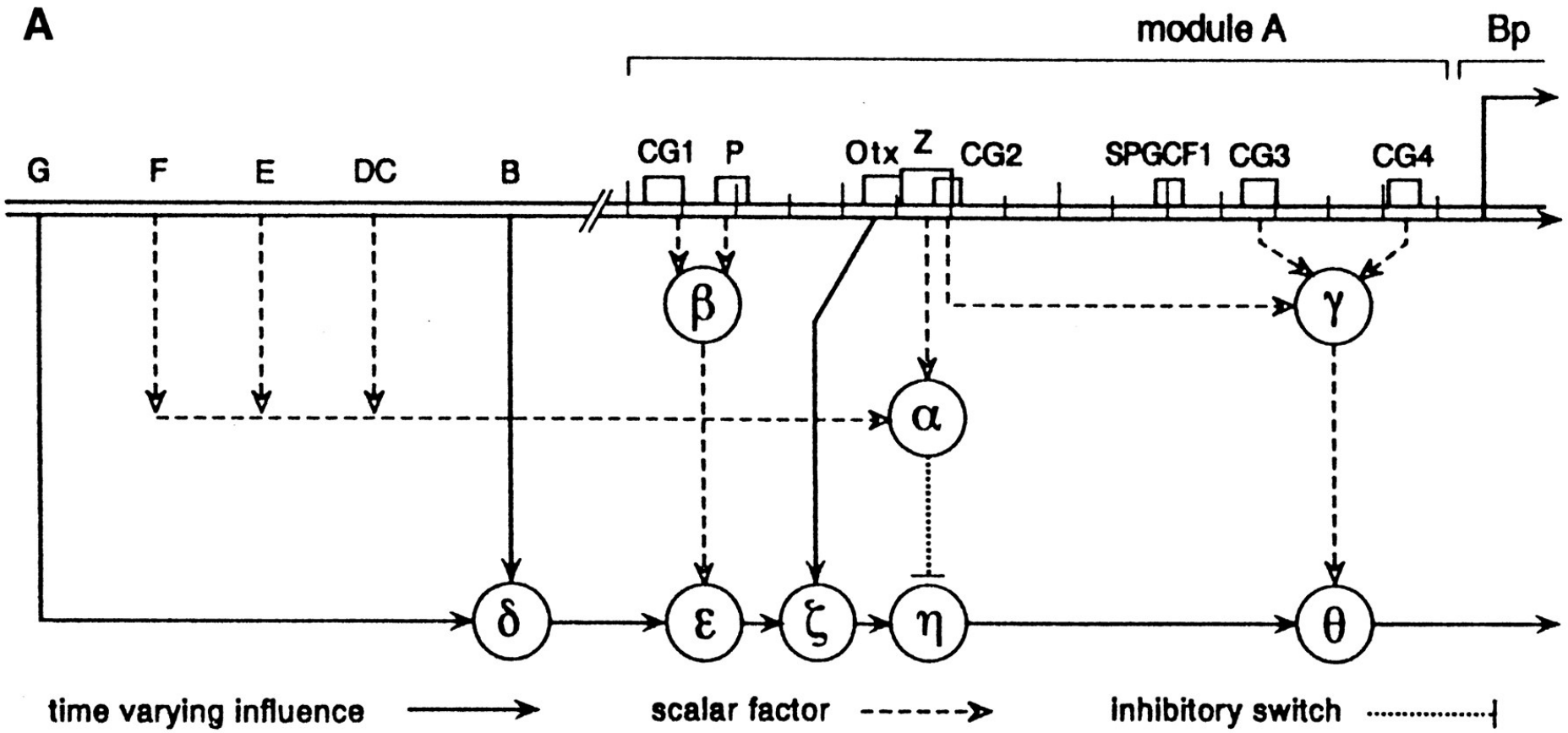
Synergism with modules B and G enhancing endoderm expression in later development:

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

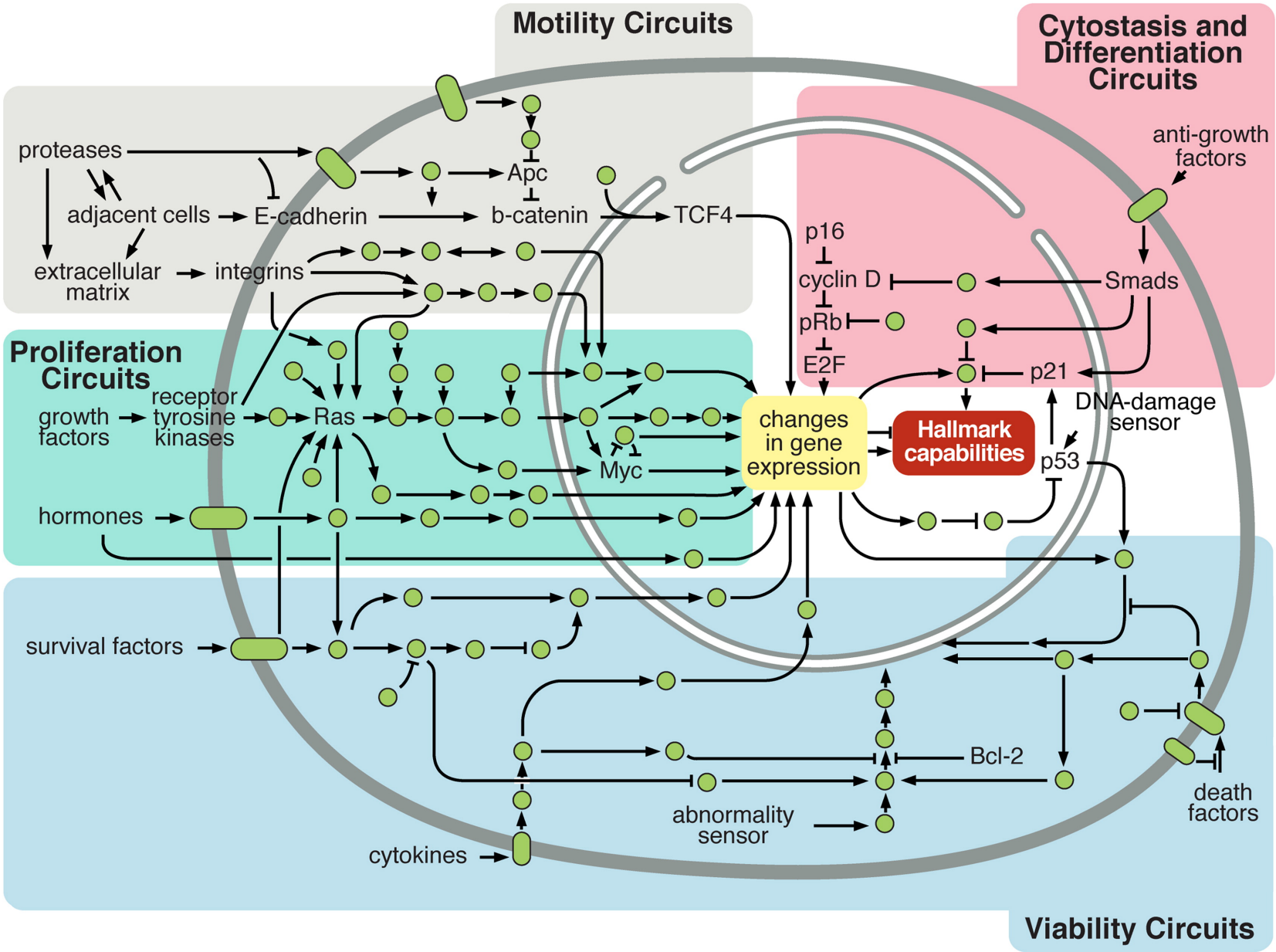


A Cell is a Modular Network that Computes

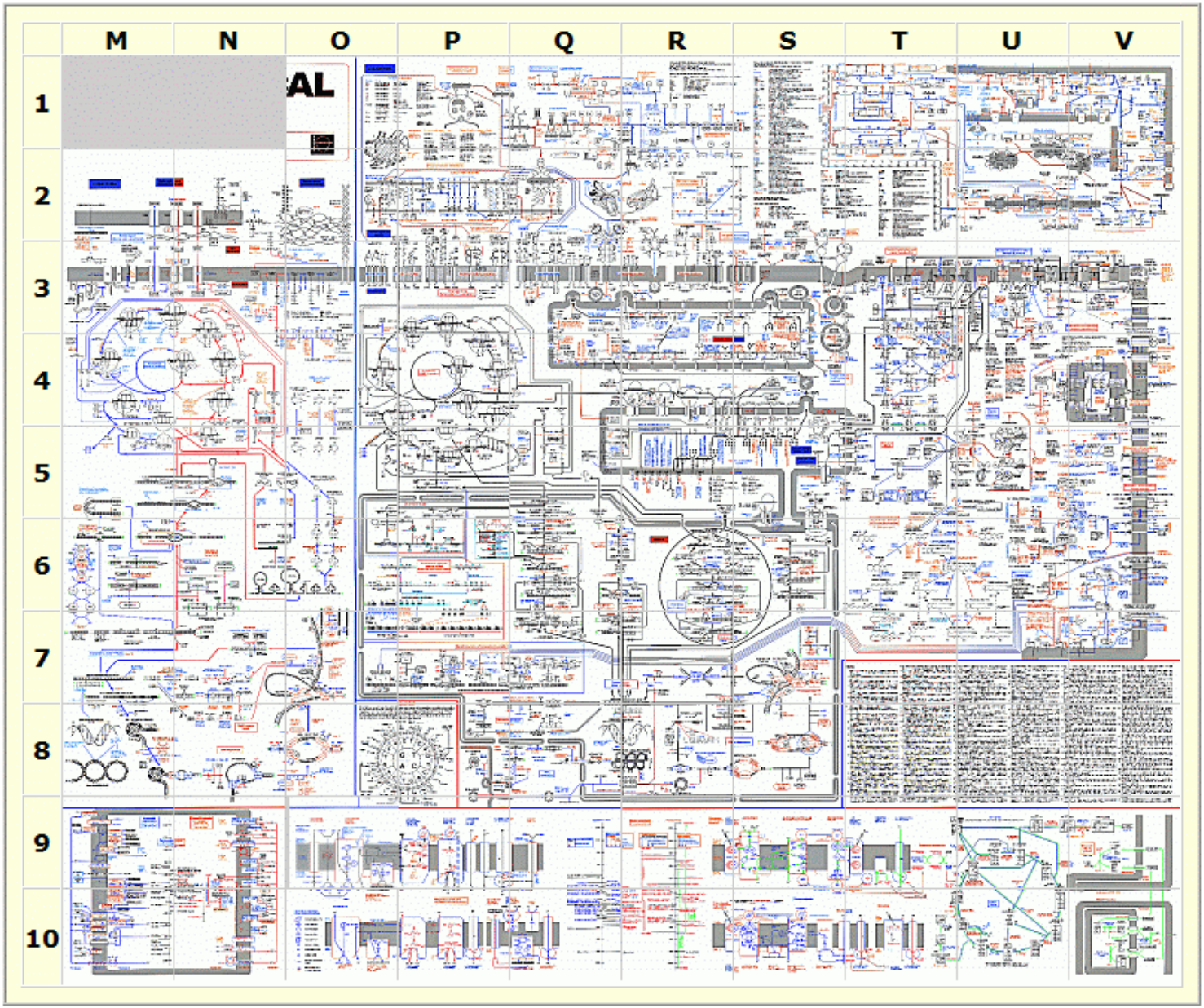
B

if (F = 1 or E = 1 or CD = 1) and (Z = 1)	Repression functions of modules F, E, and DC mediated by Z site
α = 1	
else α = 0	
if (P = 1 and CG_1 = 1)	Both P and CG_1 needed for synergistic link with module B
β = 2	
else β = 0	
if (CG_2 = 1 and CG_3 = 1 and CG_4 = 1)	Final step up of system output
γ = 2	
else γ = 1	
$\delta(t) = B(t) + G(t)$	Positive input from modules B and G
$\varepsilon(t) = \beta * \delta(t)$	Synergistic amplification of module B output by CG_1-P subsystem
if ($\varepsilon(t) = 0$)	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
$\xi(t) = Otx(t)$	
else $\xi(t) = \varepsilon(t)$	
if ($\alpha = 1$)	Repression function inoperative in endoderm but blocks activity elsewhere
$\eta(t) = 0$	
else $\eta(t) = \xi(t)$	
$\Theta(t) = \gamma * \eta(t)$	Final output communicated to BTA

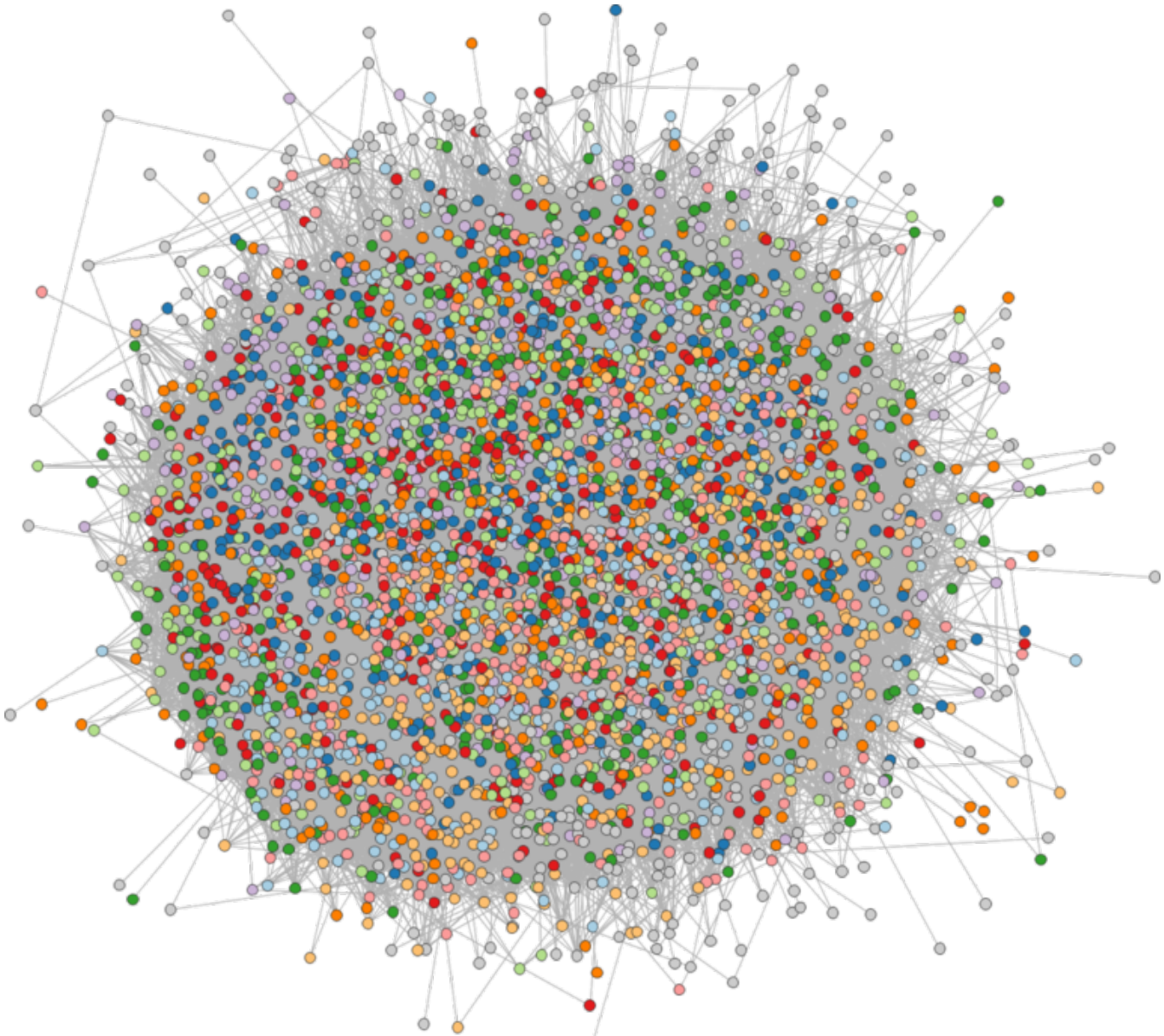
Network is Complex



Network is Complex

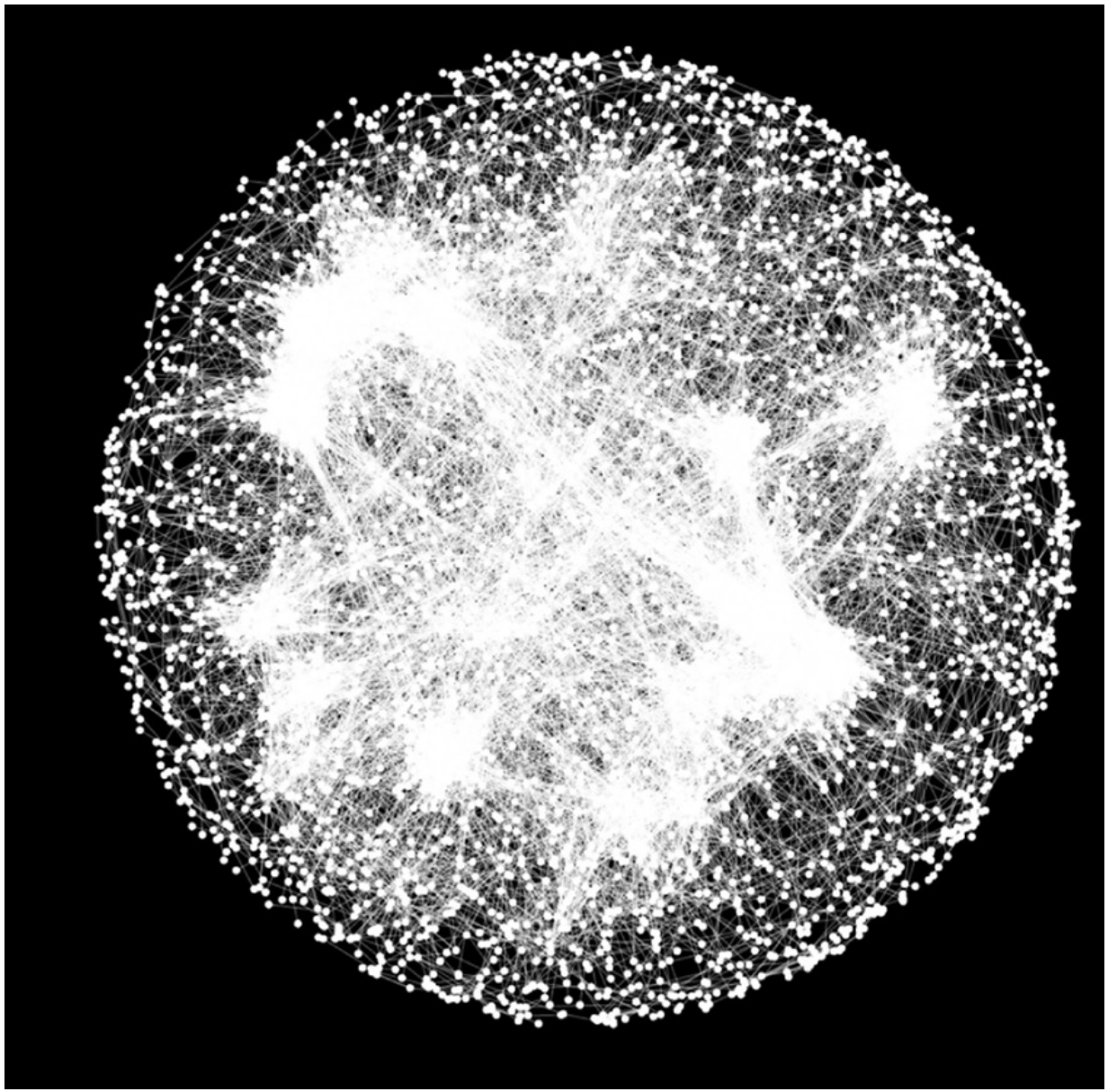


Network is Complex but Very Poorly Understood



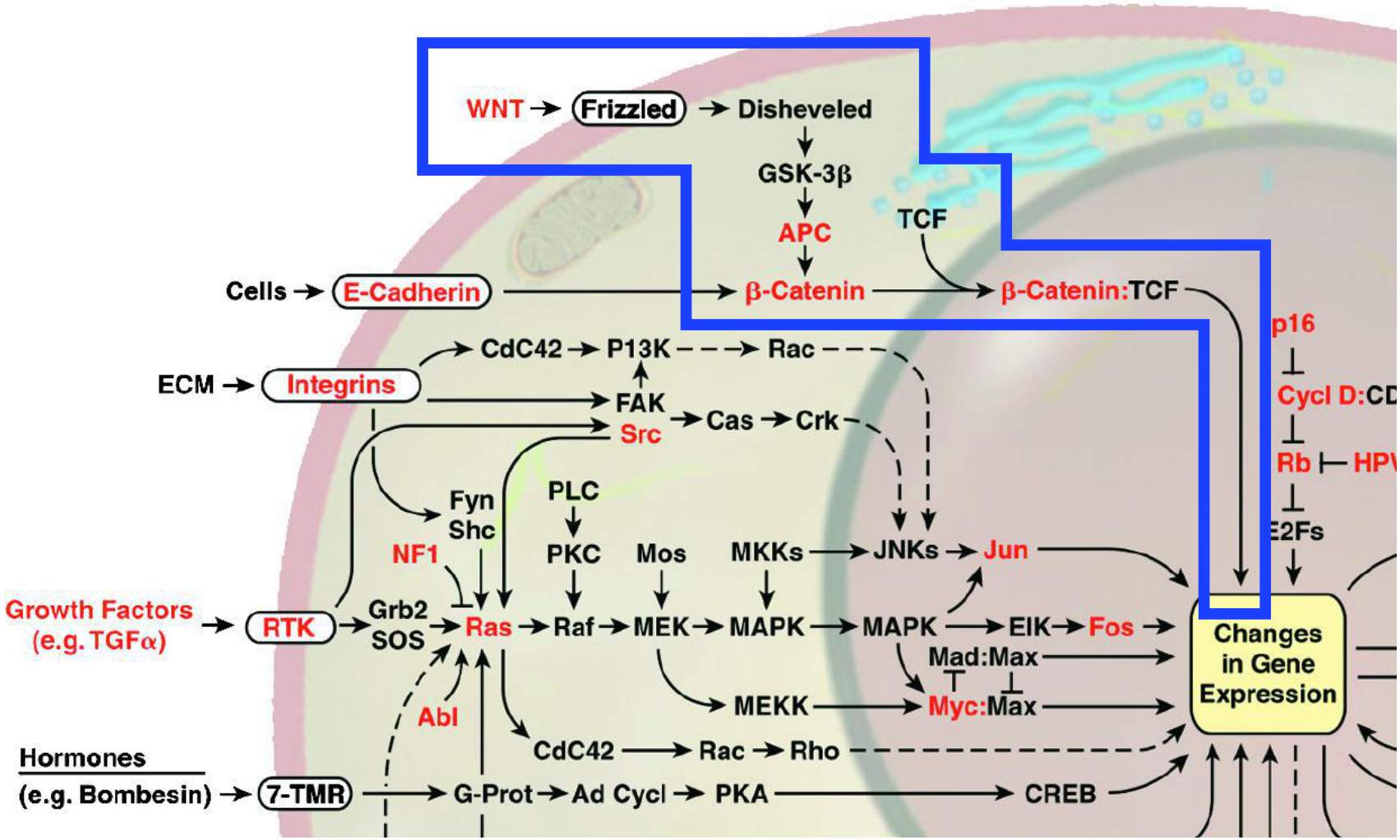
Arlind Nocaj, Untangling Networks: Focus on Less to See More, Ph.D. thesis, Universität Konstanz, 2015.

Network is Complex but Very Poorly Understood



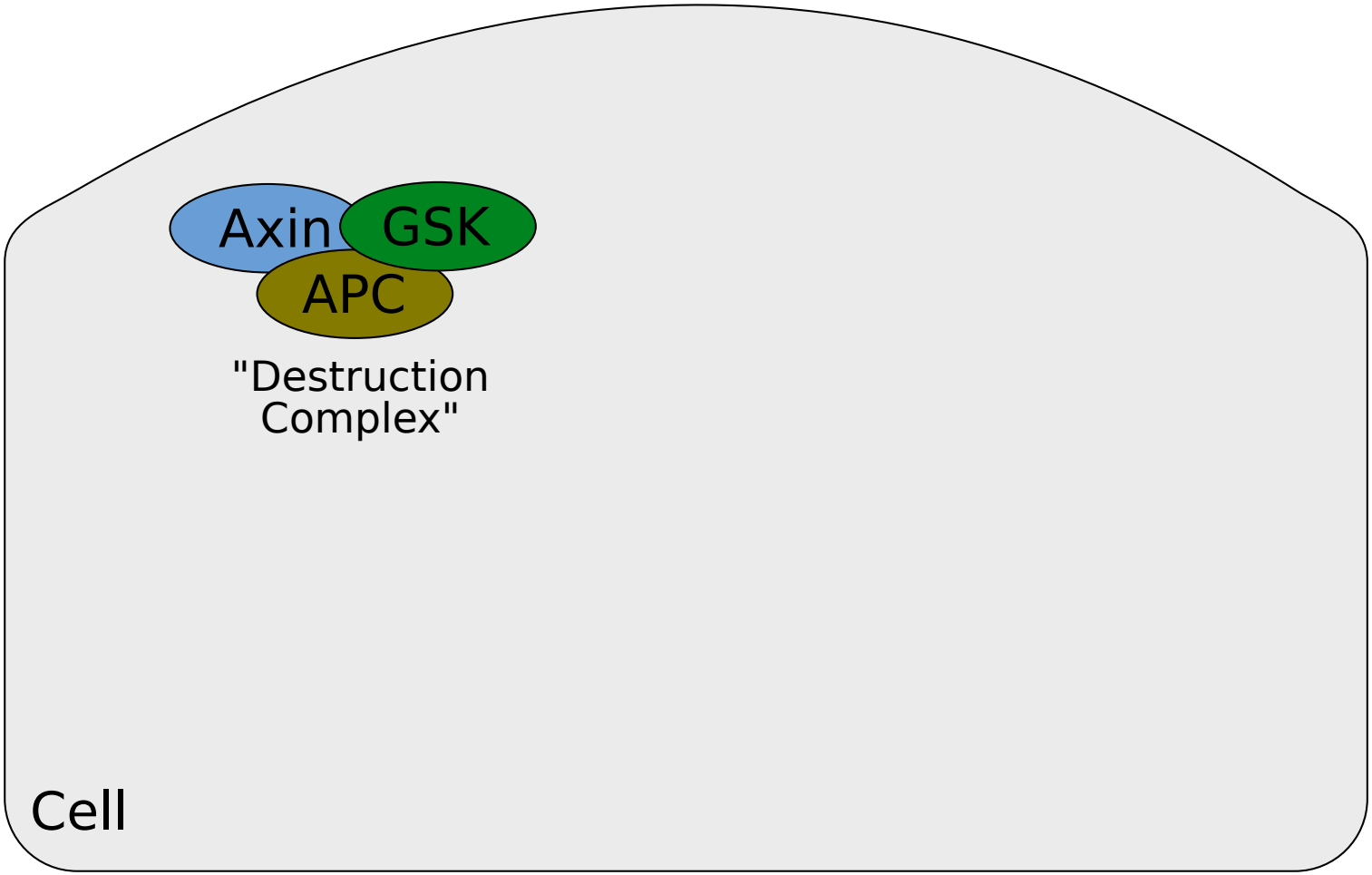
Costanzo et al., Cell, 2019.

Wnt Signaling

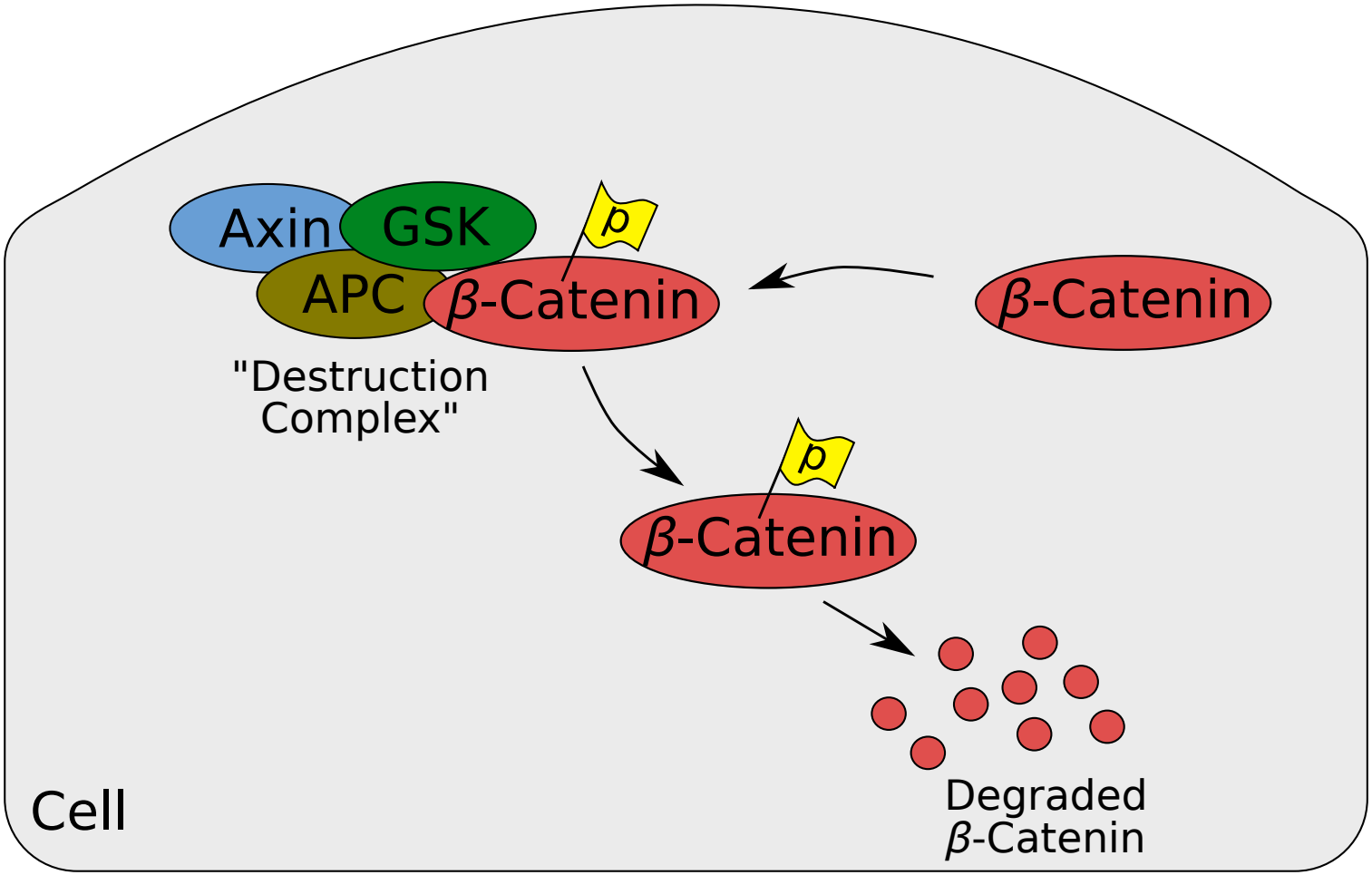


Hanahan and Wienberg. Hallmarks of cancer. Cell, 2000.

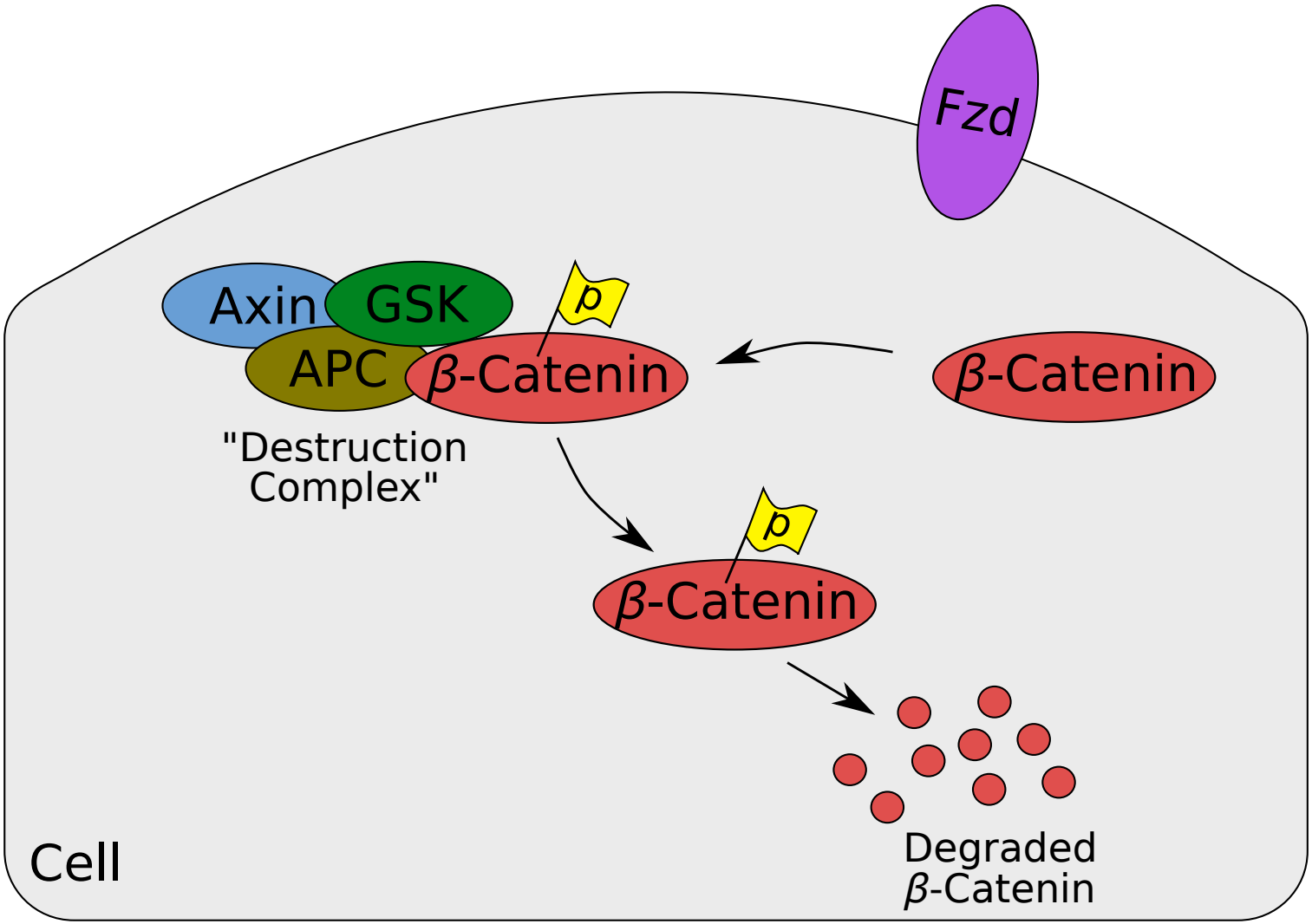
In the Absence of Wnt



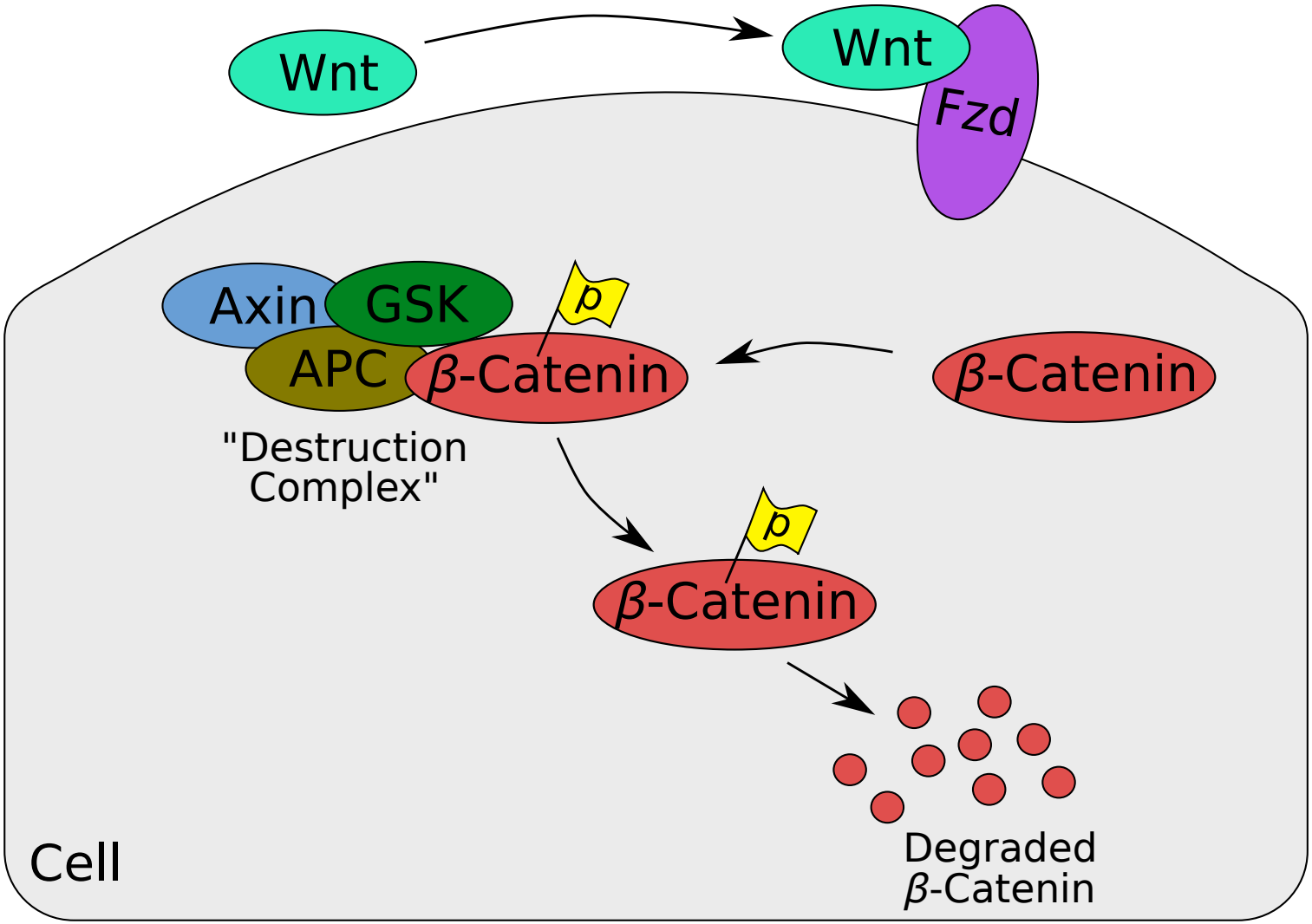
In the Absence of Wnt



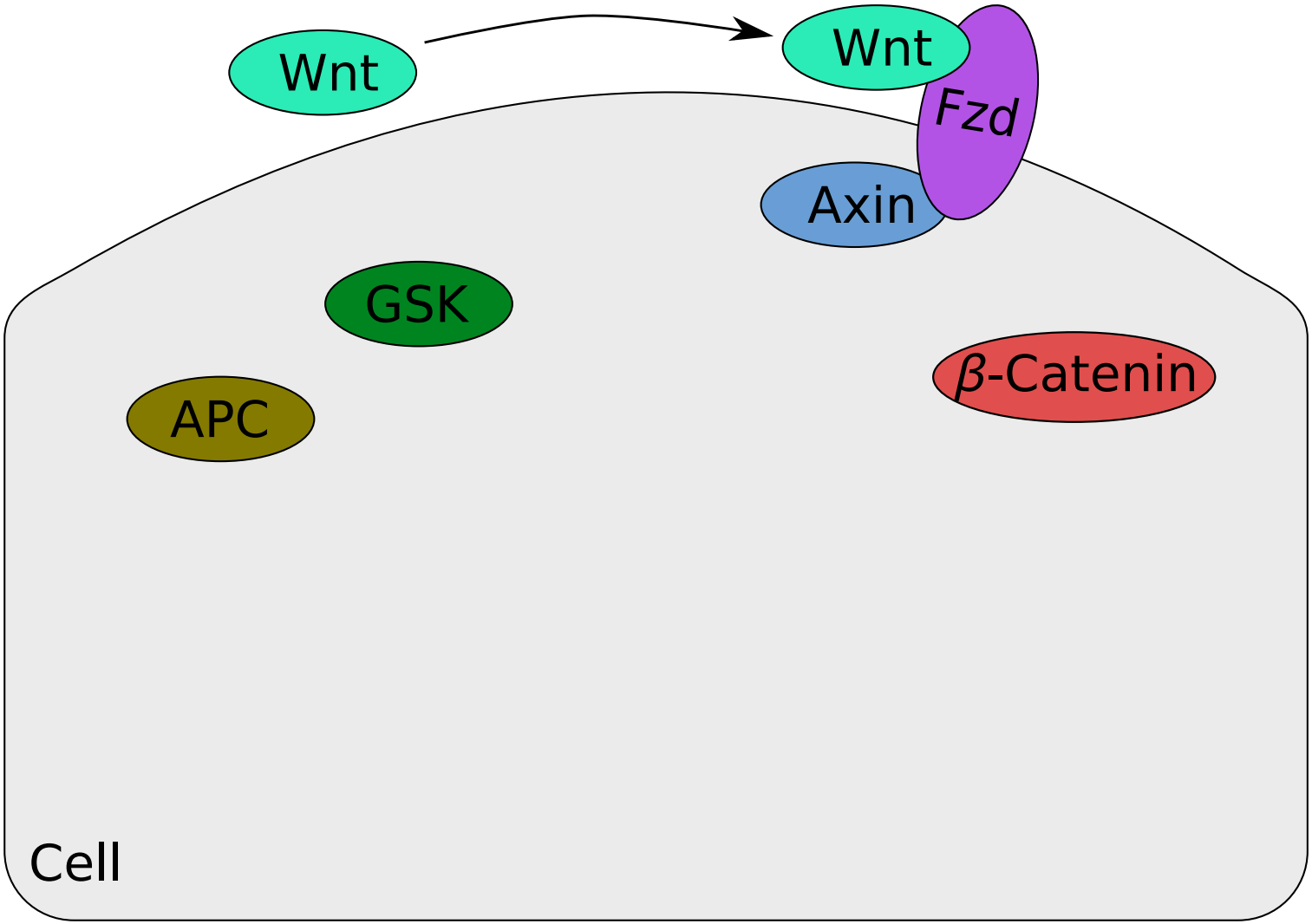
In the Presence of Extracellular Wnt



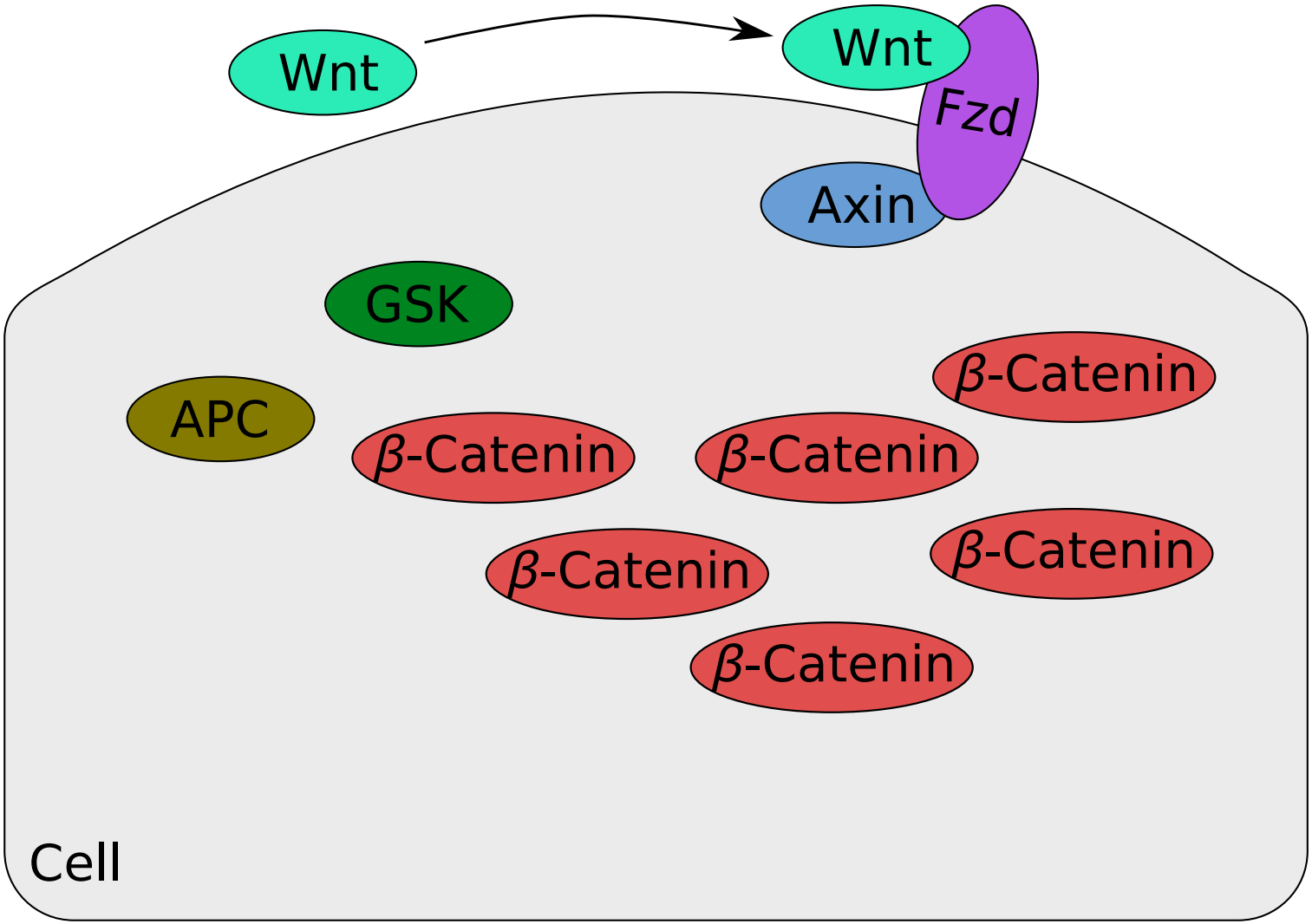
In the Presence of Extracellular Wnt



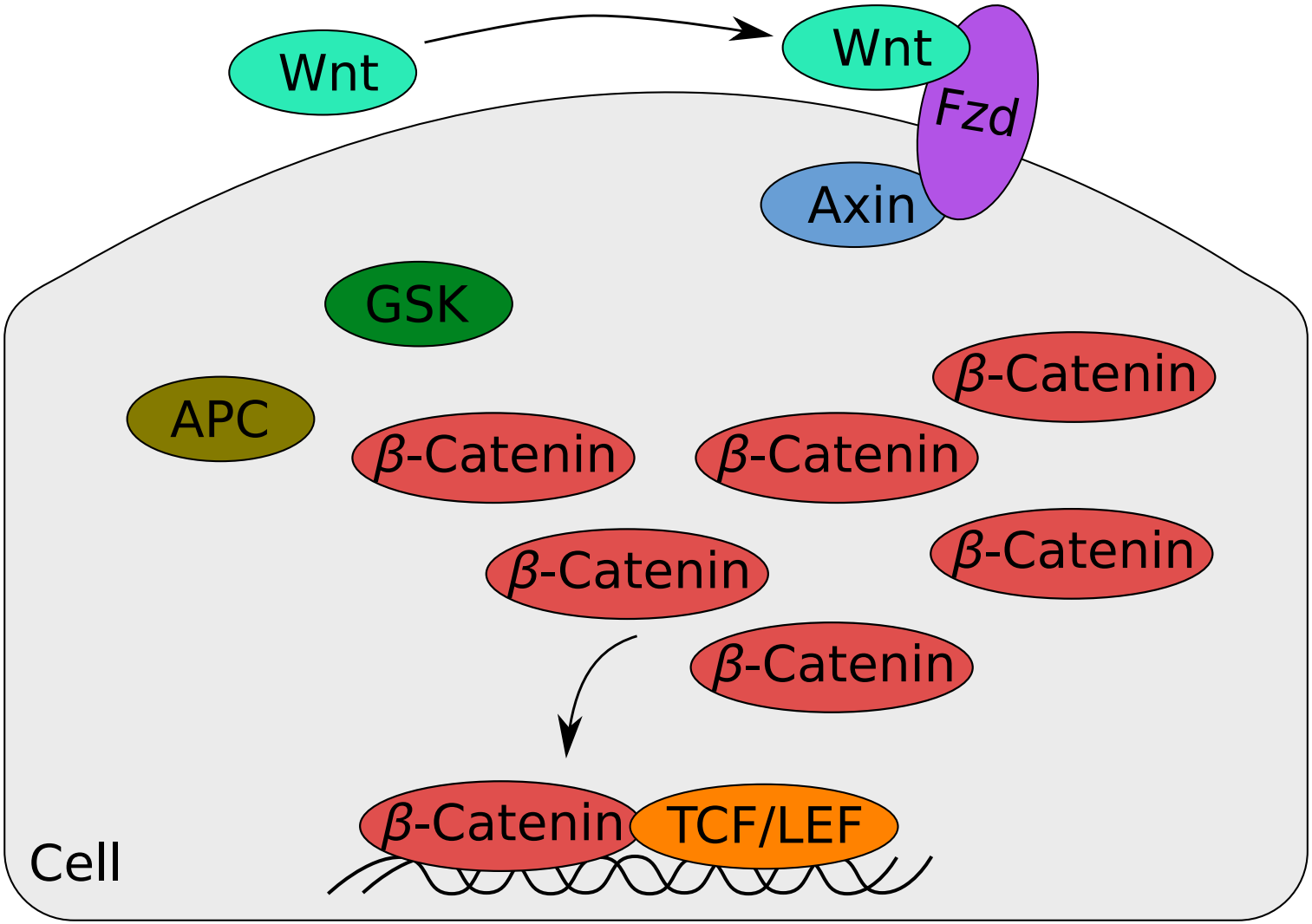
In the Presence of Extracellular Wnt



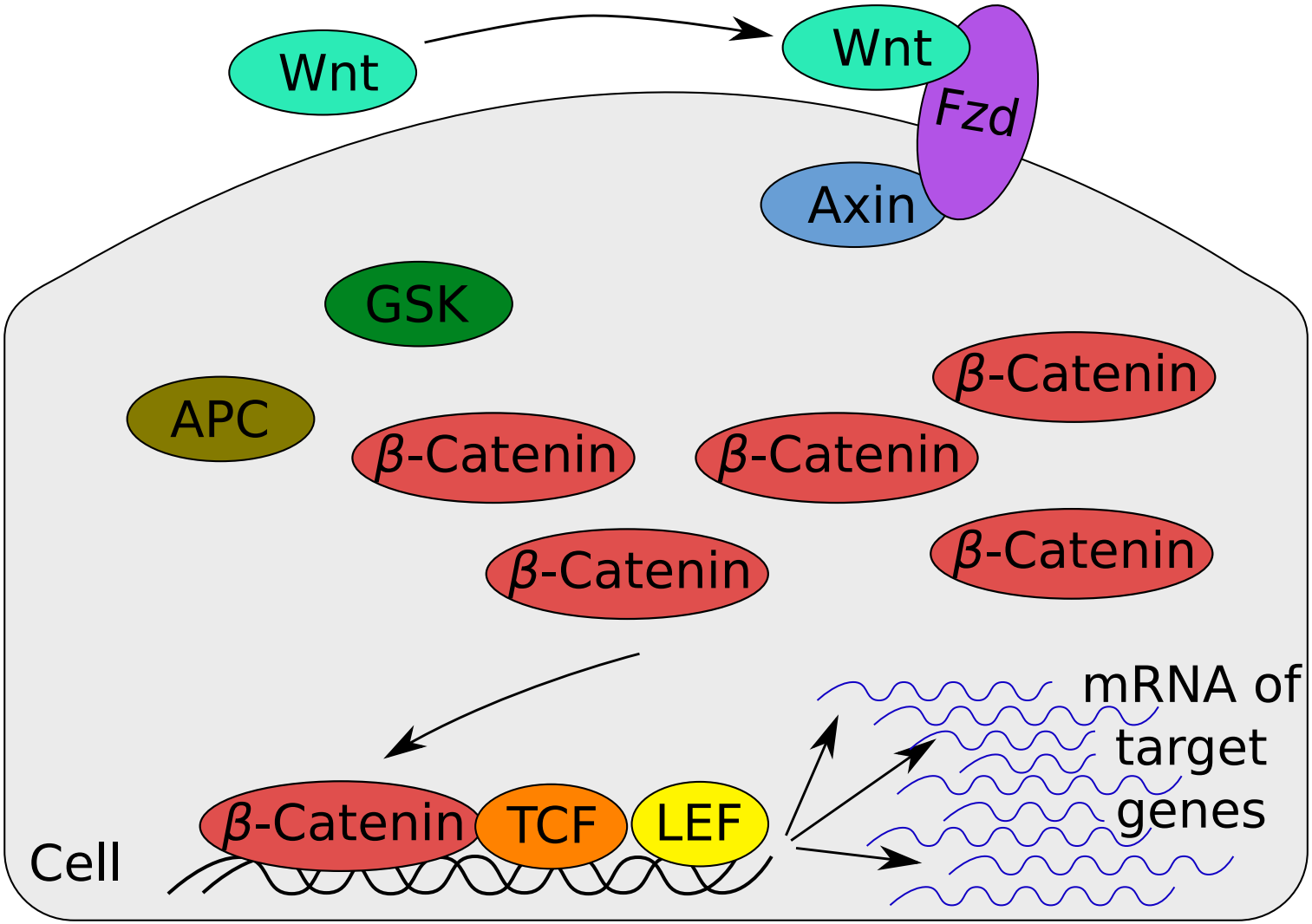
In the Presence of Extracellular Wnt



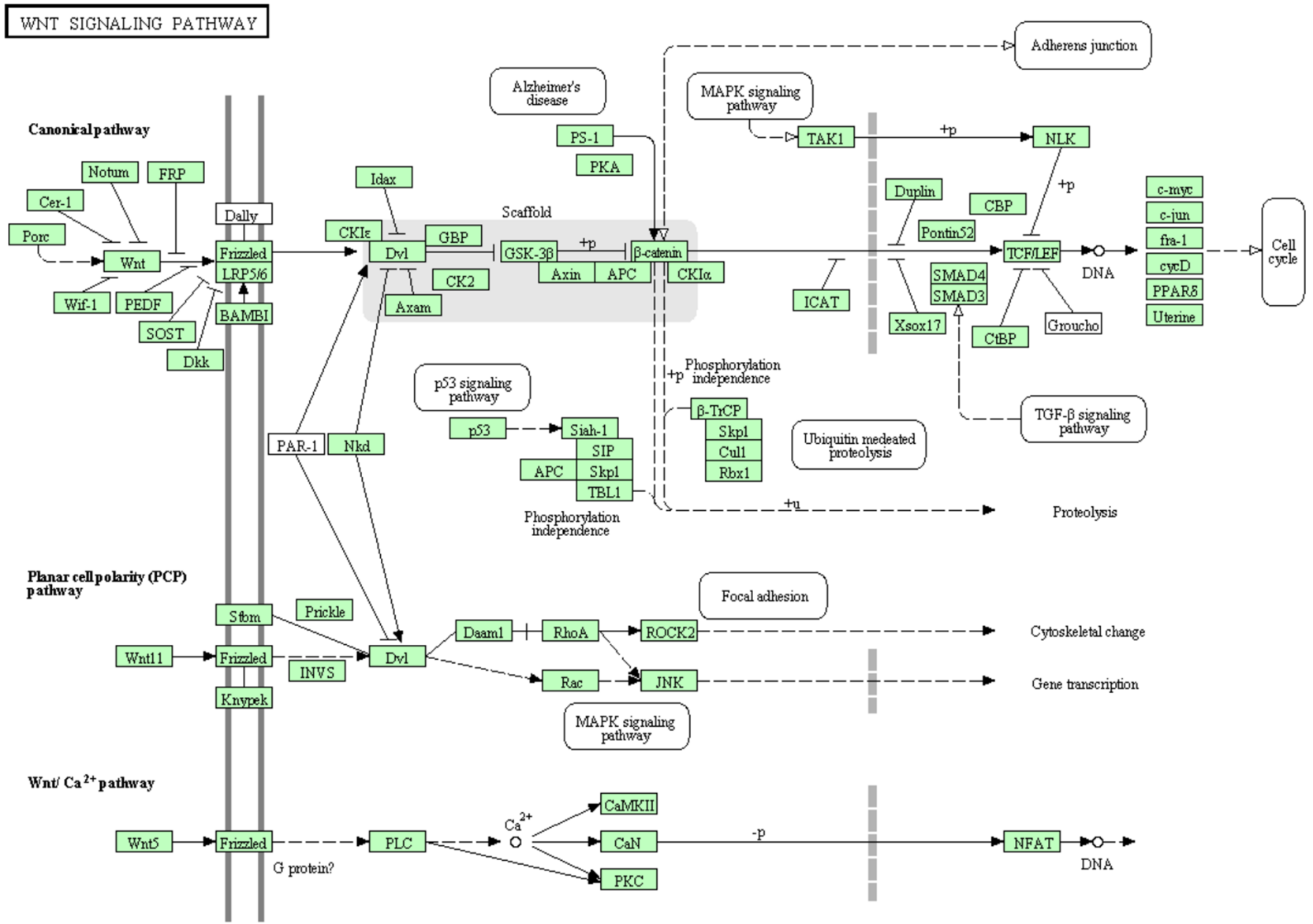
In the Presence of Extracellular Wnt



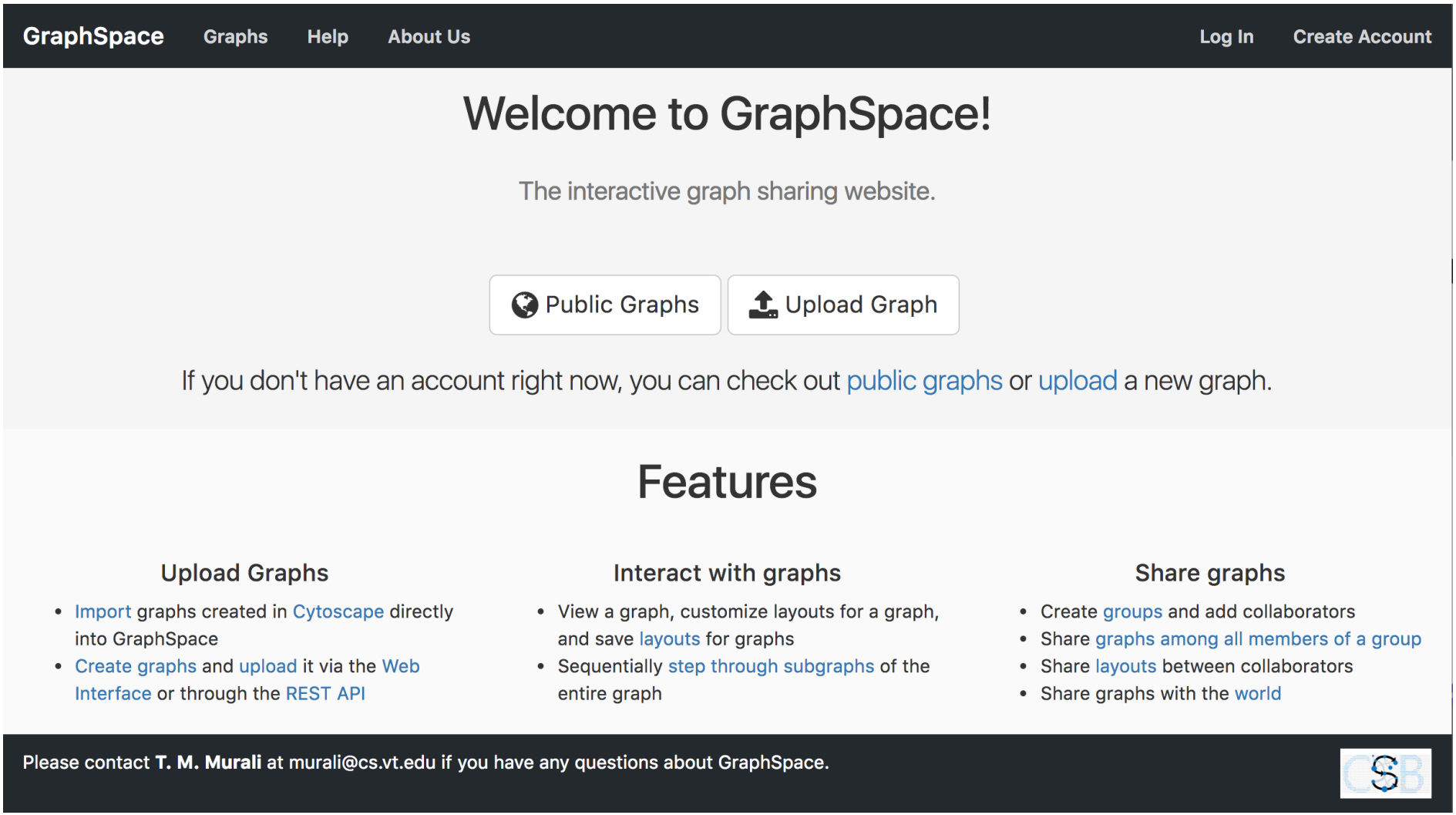
In the Presence of Extracellular Wnt



Wnt Pathway in the KEGG Database

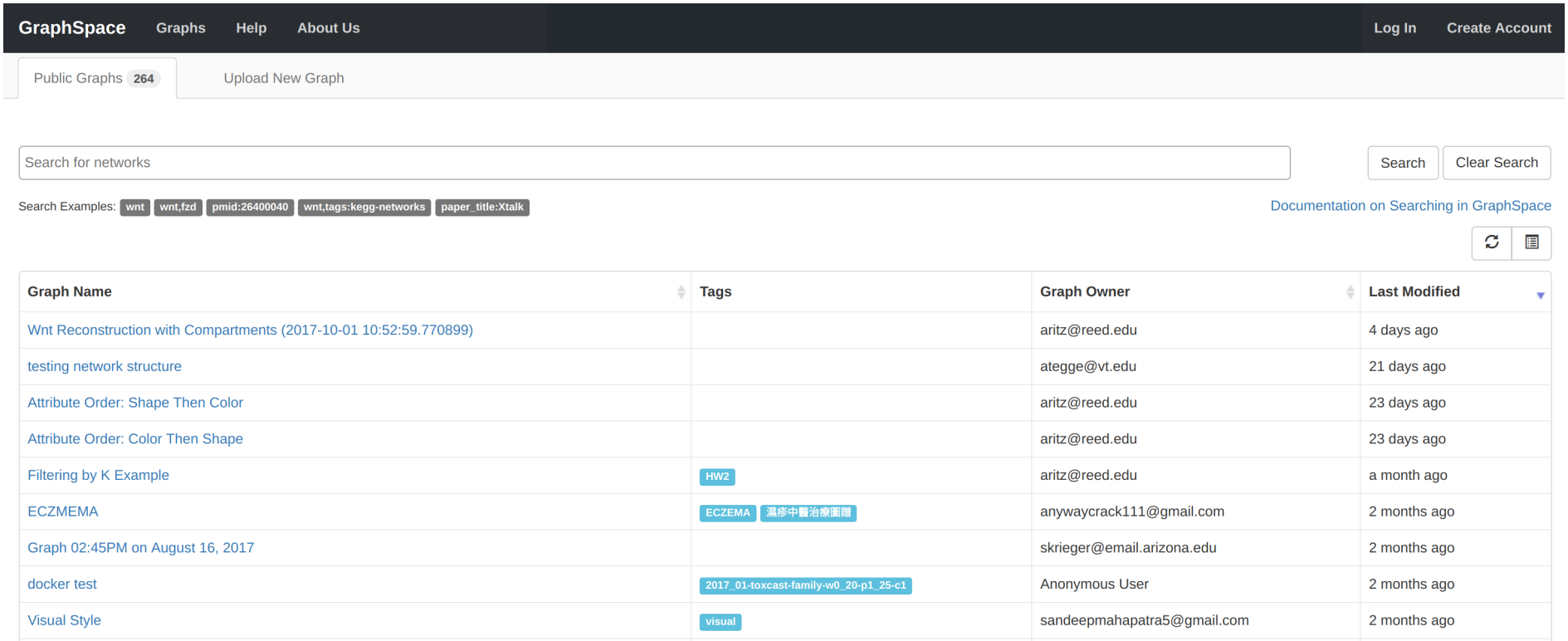


GraphSpace



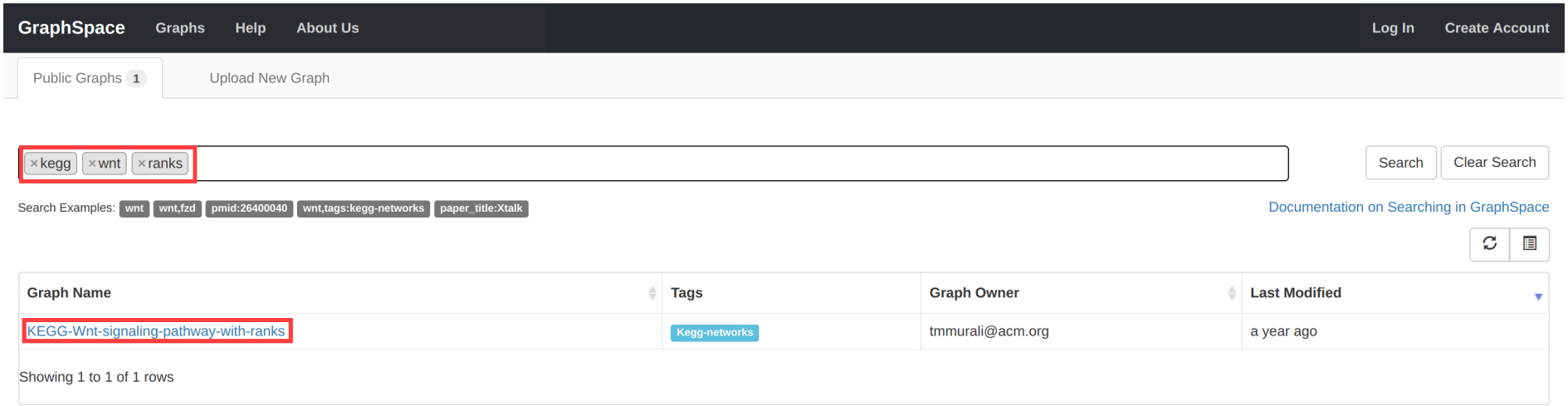
1. Click on "Public Graphs".

GraphSpace



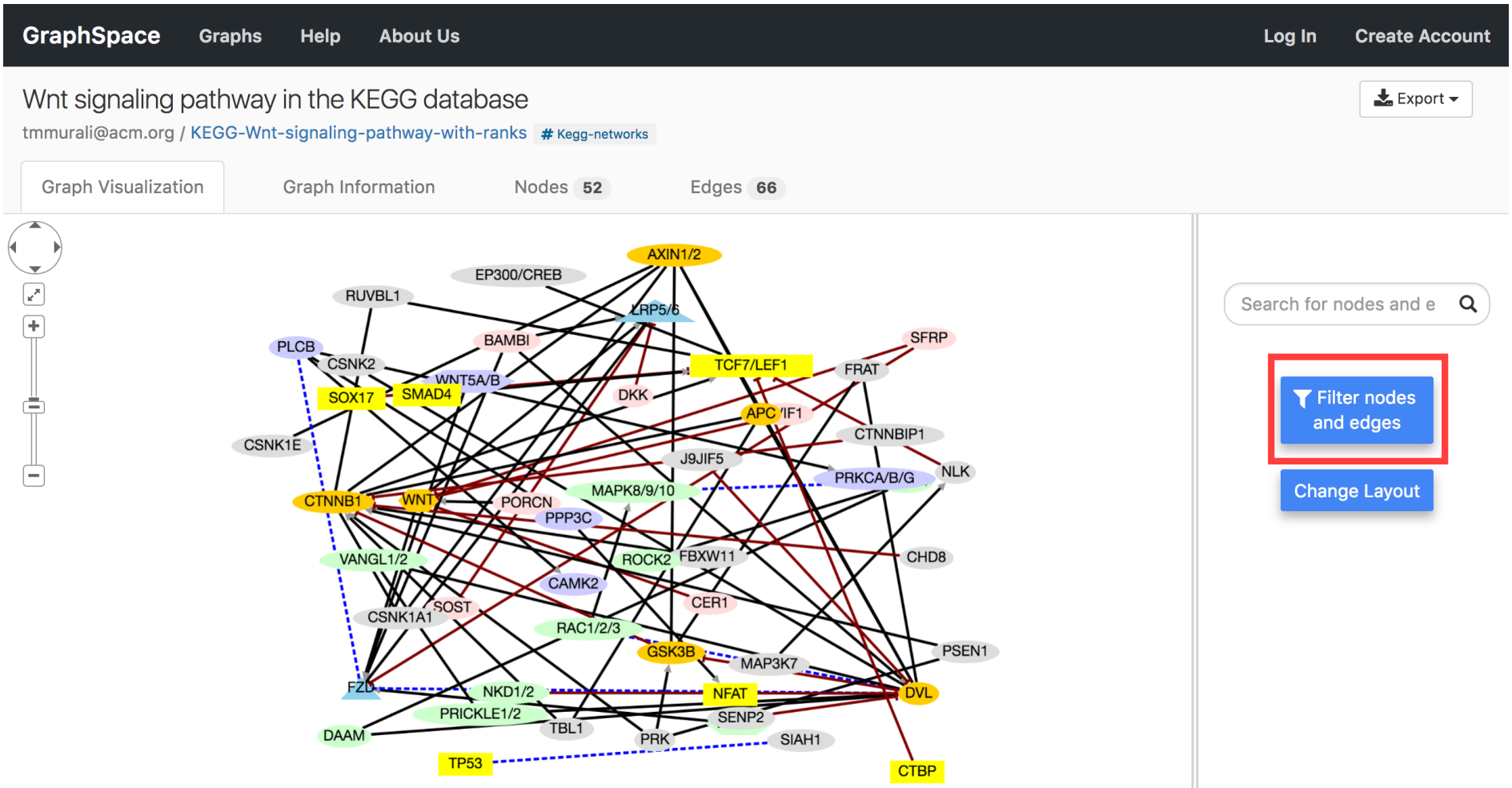
1. Click on "Public Graphs".
2. Search for "KEGG Wnt ranks".

GraphSpace



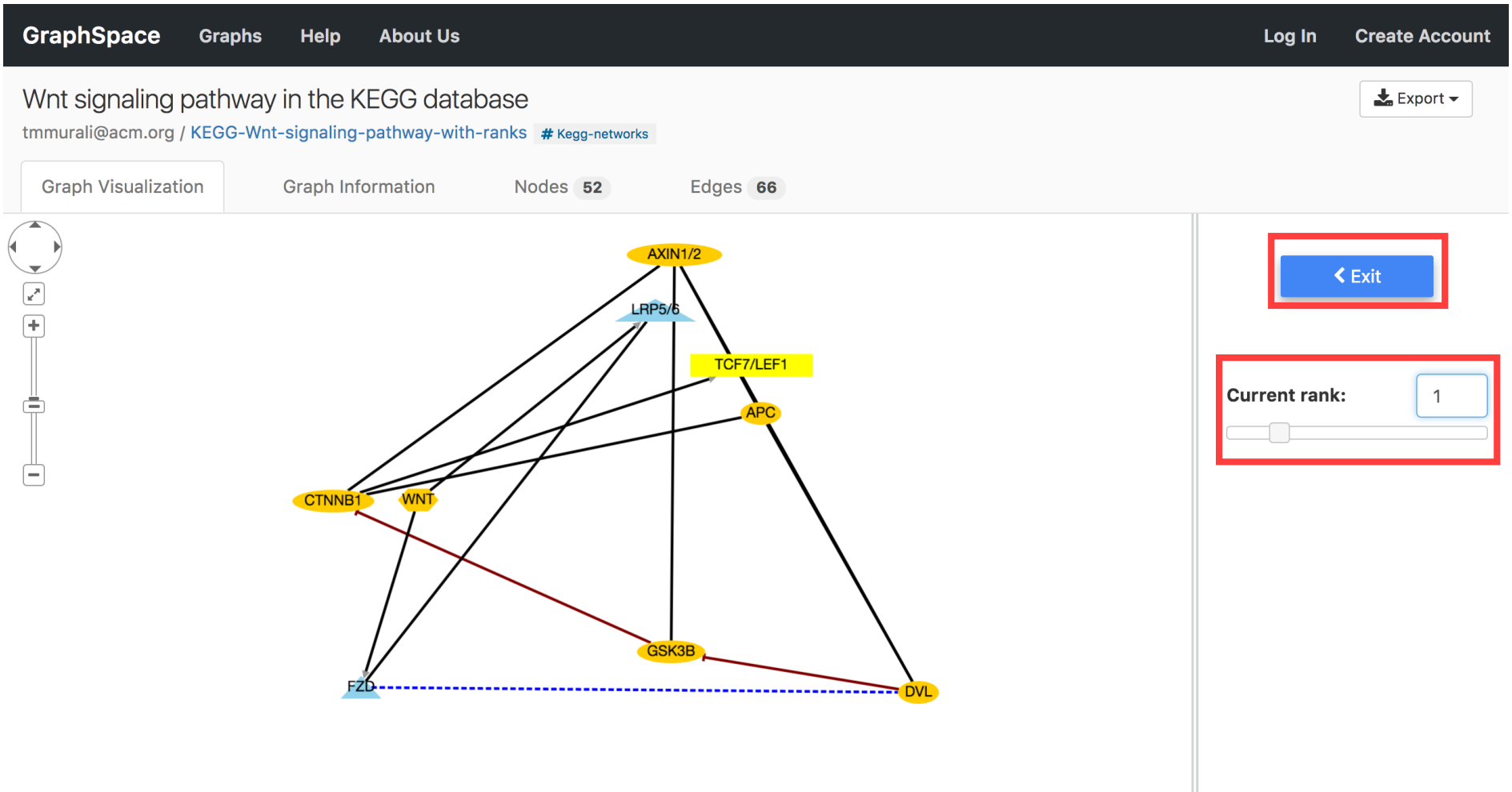
1. Click on "Public Graphs".
2. Search for "KEGG Wnt ranks".
3. Click on "KEGG-Wnt-signaling-pathway-with-ranks".

Wnt Pathway on GraphSpace

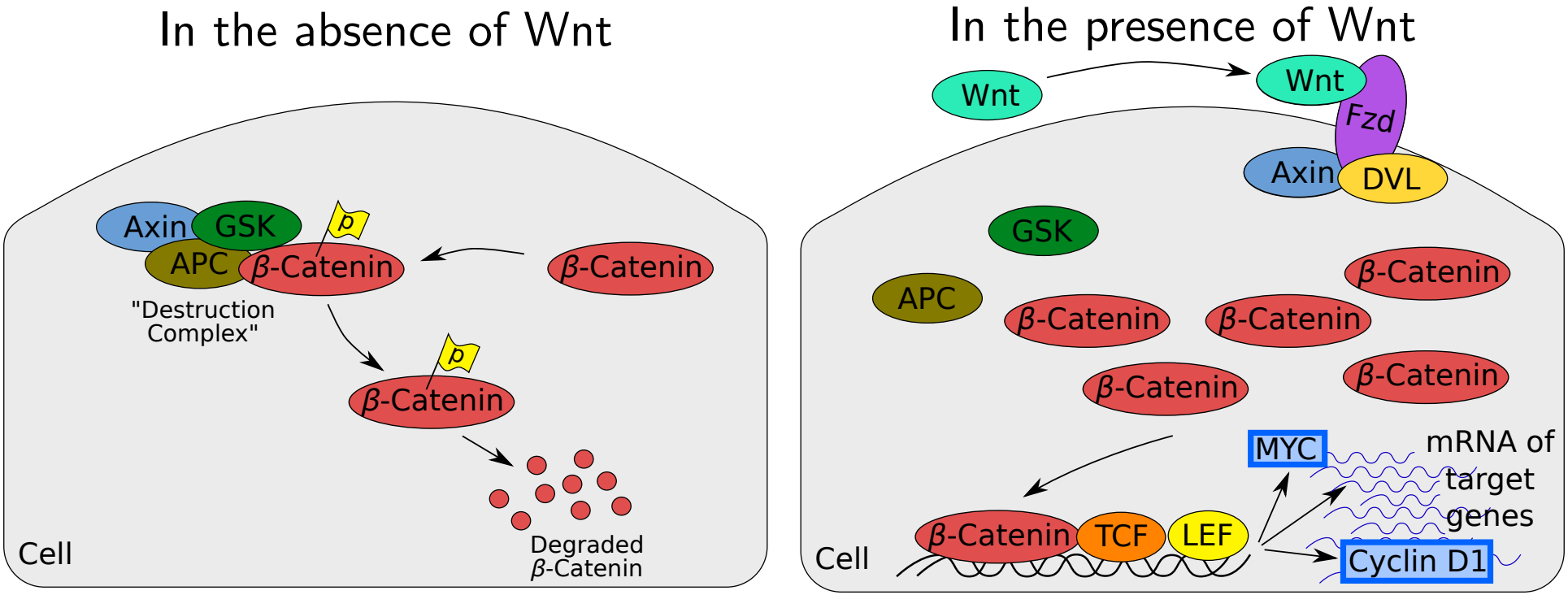


- Open the "Filter nodes and edges" panel on the right.
- Set the "Current rank" to "1" and then "Exit".

Wnt Pathway on GraphSpace

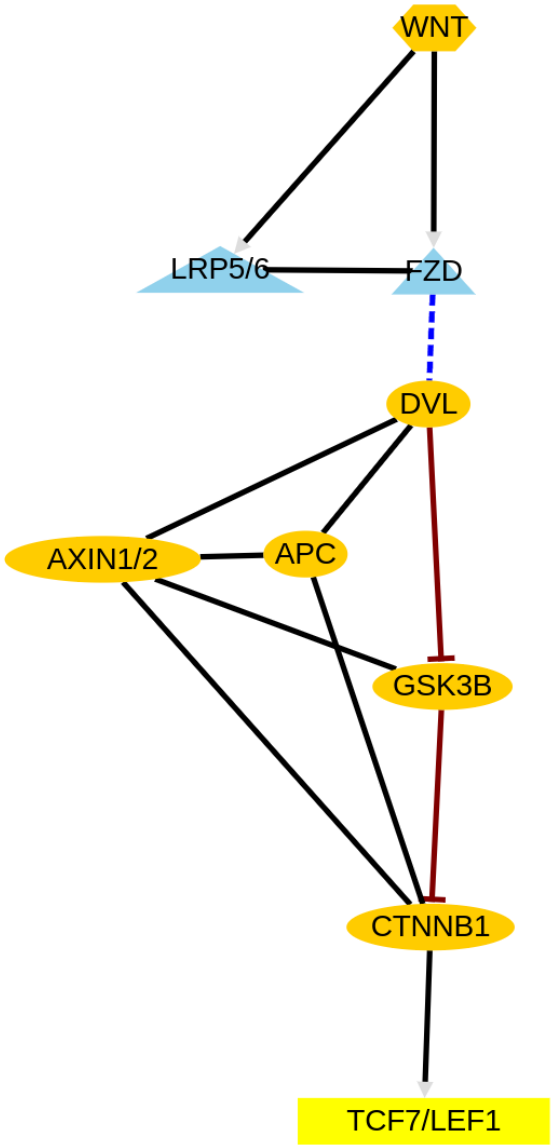


- Set the "Current rank" to "1" and then "Exit".
- Move the nodes in the network so that you can arrange them similar to the presentation.


In the absence of Wnt
Axin
GSK
APC
β-Catenin
p
"Destruction Complex"
β-Catenin
β-Catenin
p
Degraded β-Catenin
Cell
In the presence of Wnt
Wnt
Wnt
Fzd
Axin
DVL
GSK
APC
β-Catenin
β-Catenin
β-Catenin
β-Catenin
β-Catenin
MYC
mRNA of target genes
β-Catenin
TCF
LEF
Cyclin D1
Cell

Interpreting the Wnt Pathway

Wnt signaling pathway in the KEGG database

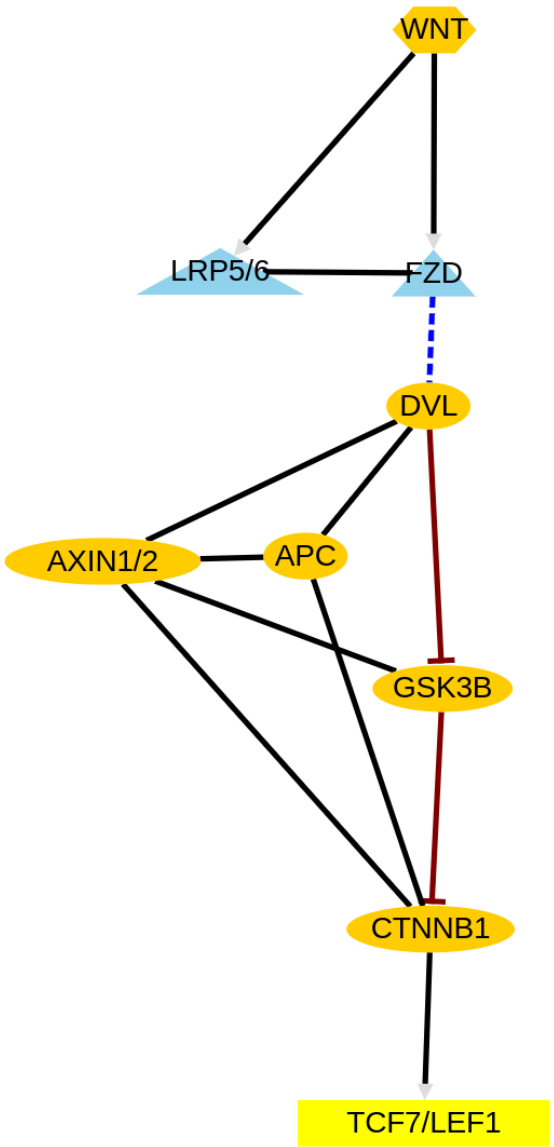


What do the arrows mean?

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

Interpreting the Wnt Pathway

Wnt signaling pathway in the KEGG database

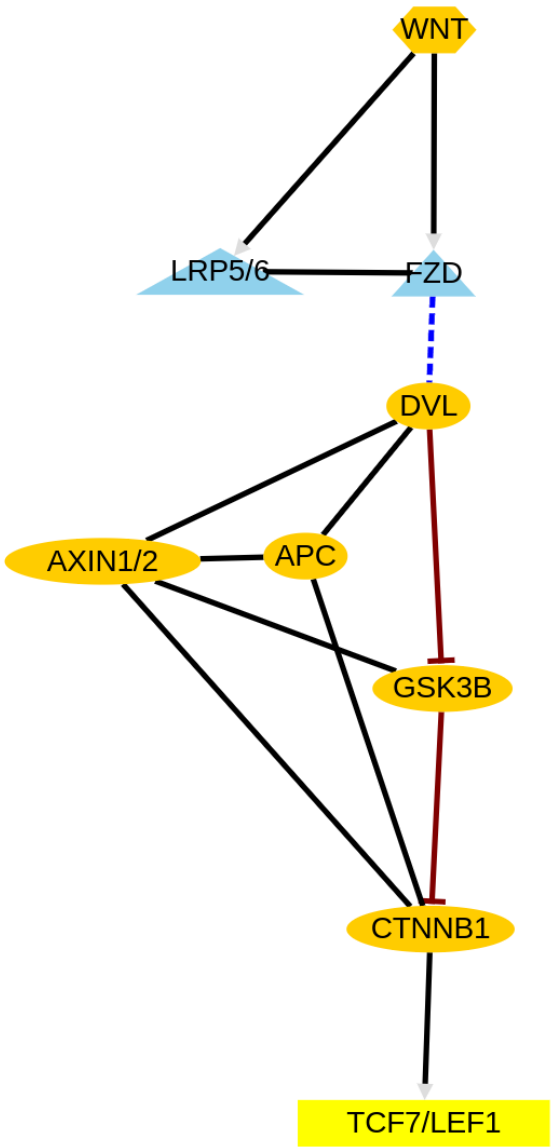


What do the arrows mean?	
Black arrowhead	Wnt activates LRP5/6
Red blunt head	DVL inhibits GSK3B
Black no head	DVL binds to Axin1/2
Dashed blue	Fzd indirectly binds to DVL

- What may happen if the cell makes lots of DVL, e.g., due to a mutation?

Interpreting the Wnt Pathway

Wnt signaling pathway in the KEGG database

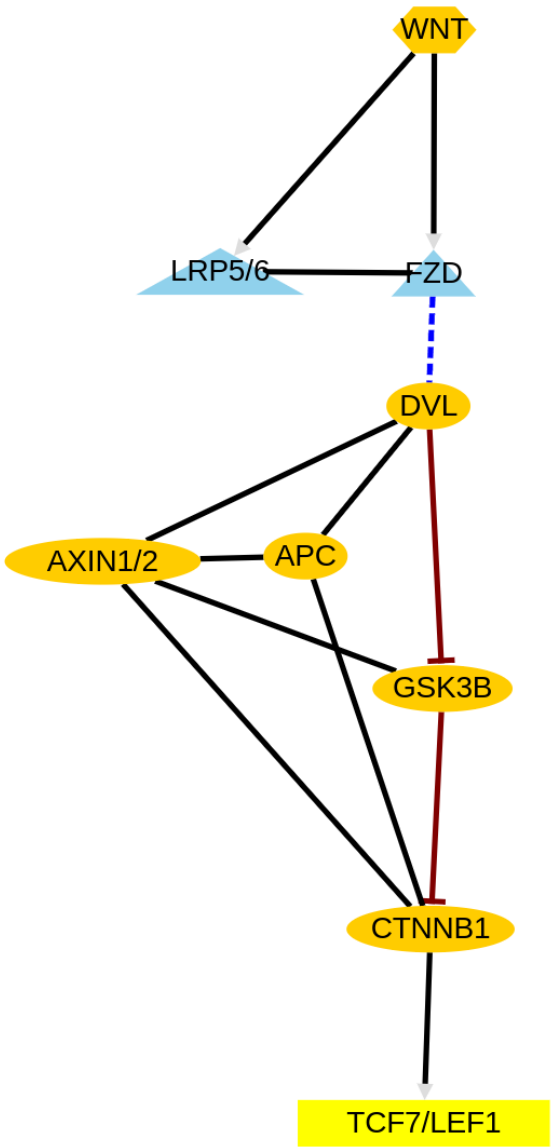


What do the arrows mean?	
Black arrowhead	Wnt activates LRP5/6
Red blunt head	DVL inhibits GSK3B
Black no head	DVL binds to Axin1/2
Dashed blue	Fzd indirectly binds to DVL

- 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

Wnt signaling pathway in the KEGG database

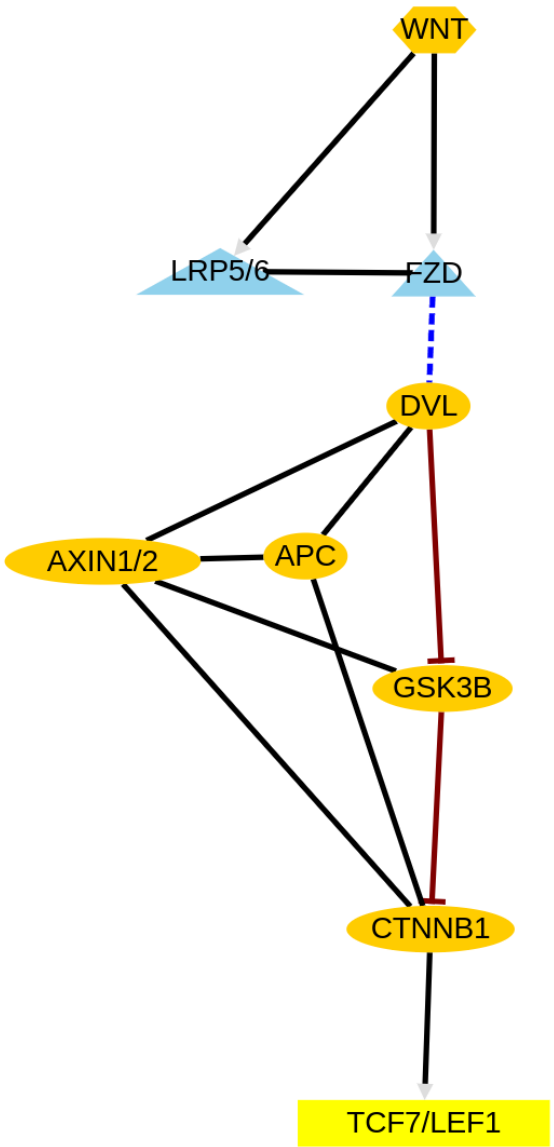


What do the arrows mean?	
Black arrowhead	Wnt activates LRP5/6
Red blunt head	DVL inhibits GSK3B
Black no head	DVL binds to Axin1/2
Dashed blue	Fzd indirectly binds to DVL

- 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

Wnt signaling pathway in the KEGG database

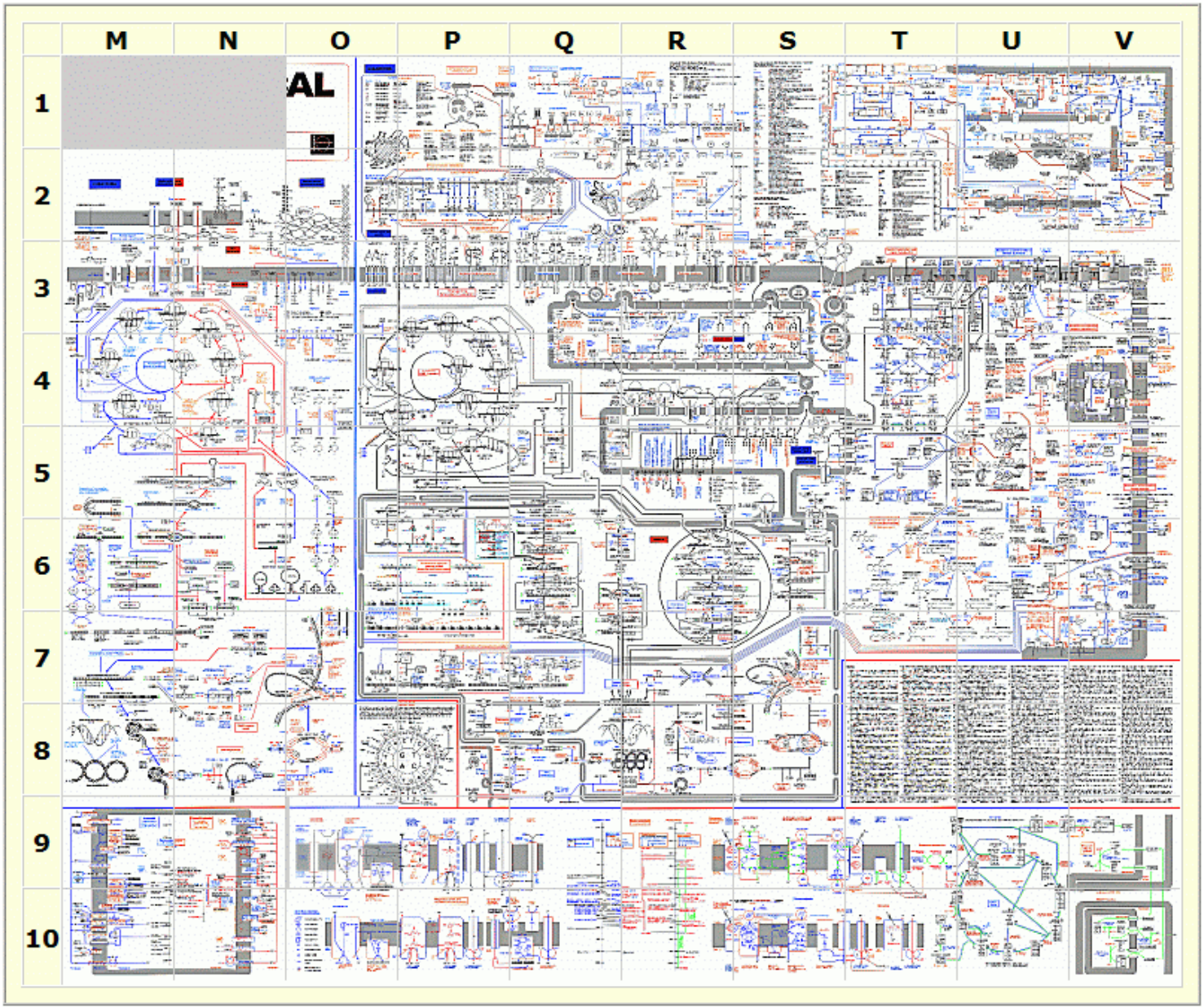


What do the arrows mean?

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.

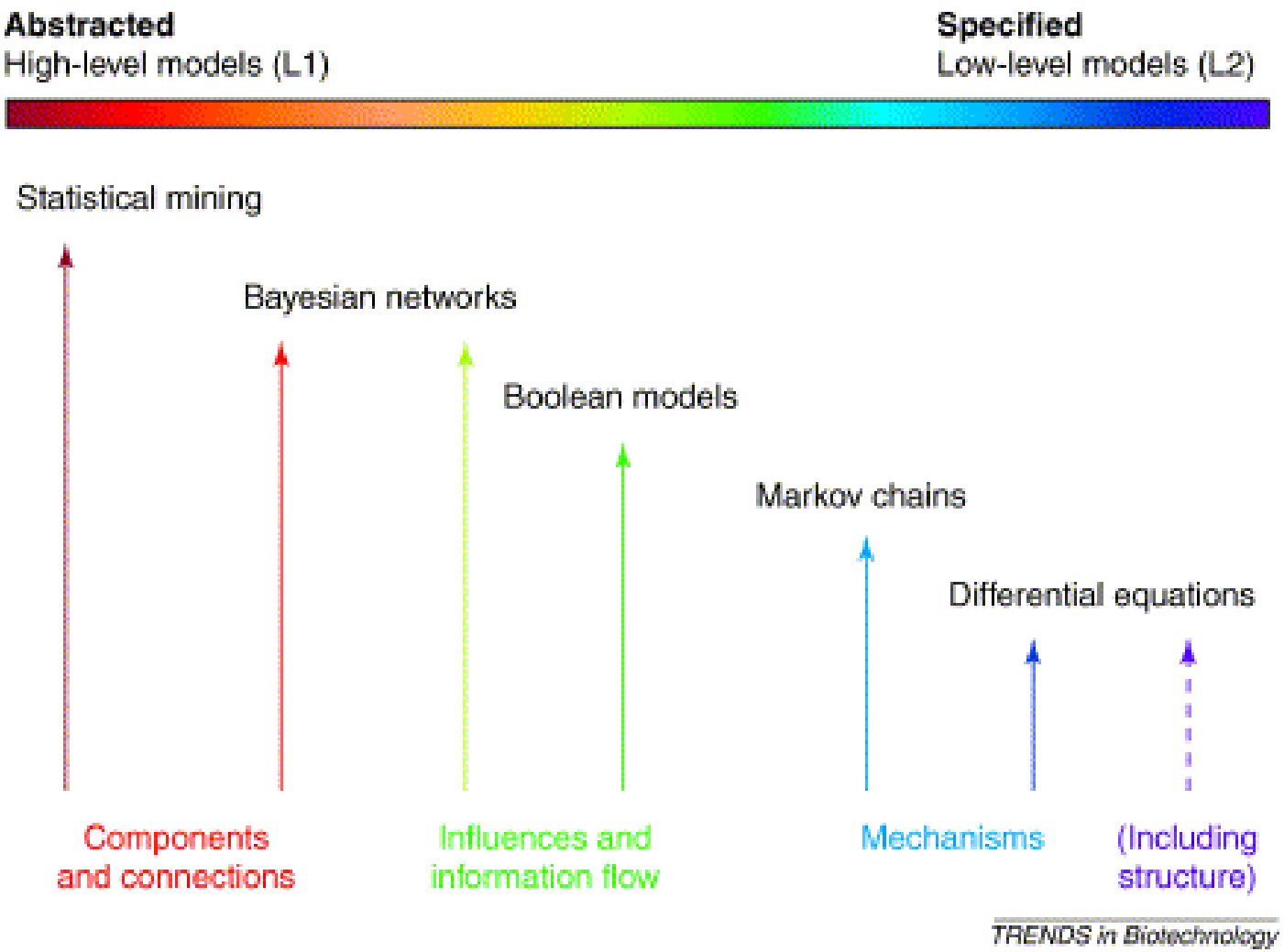
Imagine the Difficulty of Interpreting this Network!



Challenges with Molecular Interaction Networks

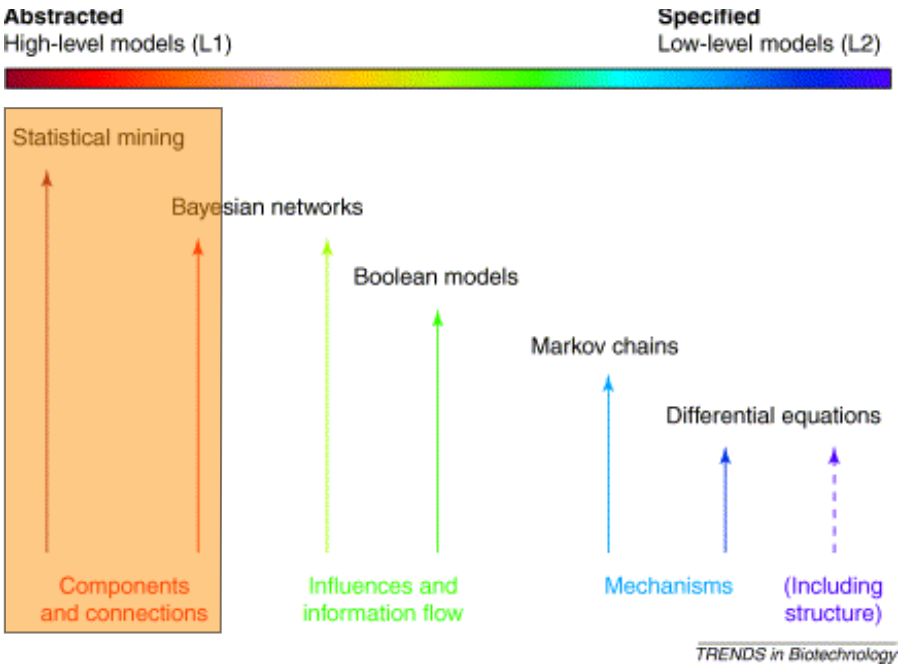
- Biological data sets and networks are large.
- They are intricate and of very diverse types.
- They are noisy: experiments are error-prone.
- They are highly incomplete. We barely know which genes interact, let alone the detailed kinetics of each interaction.

Continuum of Models in Network Biology



From *Building with a scaffold: emerging strategies for high- to low-level cellular modeling*, Ideker and Lauffenburger, Trends in Biotechnology Volume 21, Issue 6 , June 2003, Pages 255-262.

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.