Bioinformatics Session

Kathryn Dempsey, Ph.D.
Research Associate
School of Interdisciplinary Informatics
University of Nebraska at Omaha
Email: kdempsey@unomaha.edu
Phone: 402-554-2562

OVERVIEW

- Introductions
- Part I: What is bioinformatics?
  - Activity 1: Extraction of DNA
- Part II: How is bioinformatics useful?
  - Activity 2: Finding Disease Genes
- Conclusions

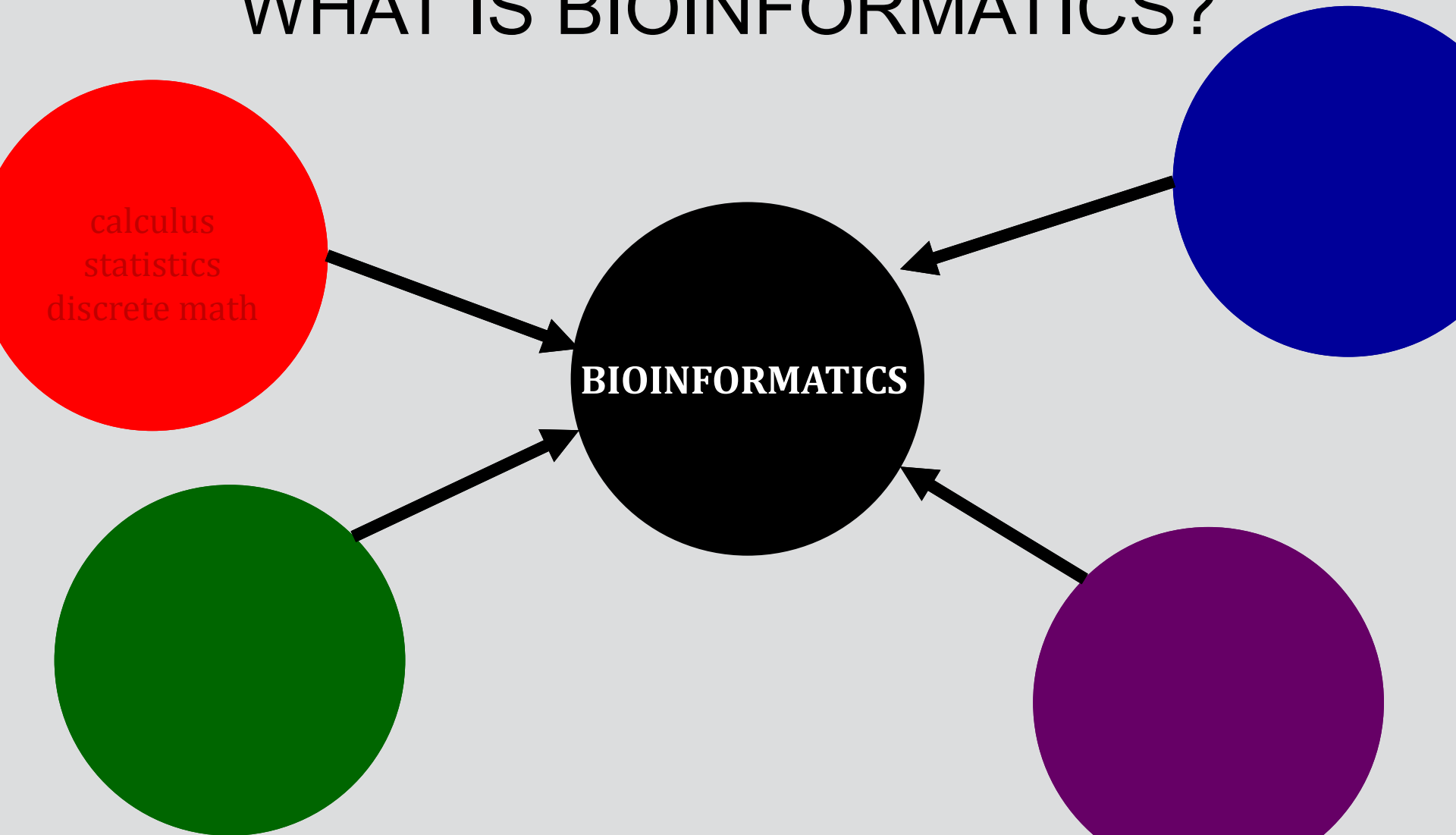
OVERVIEW

- Introductions
- Part I: What is bioinformatics?
  - Activity 1: Extraction of DNA
- Part II: How is bioinformatics useful?
  - Activity 2: Finding Disease Genes
- Conclusions

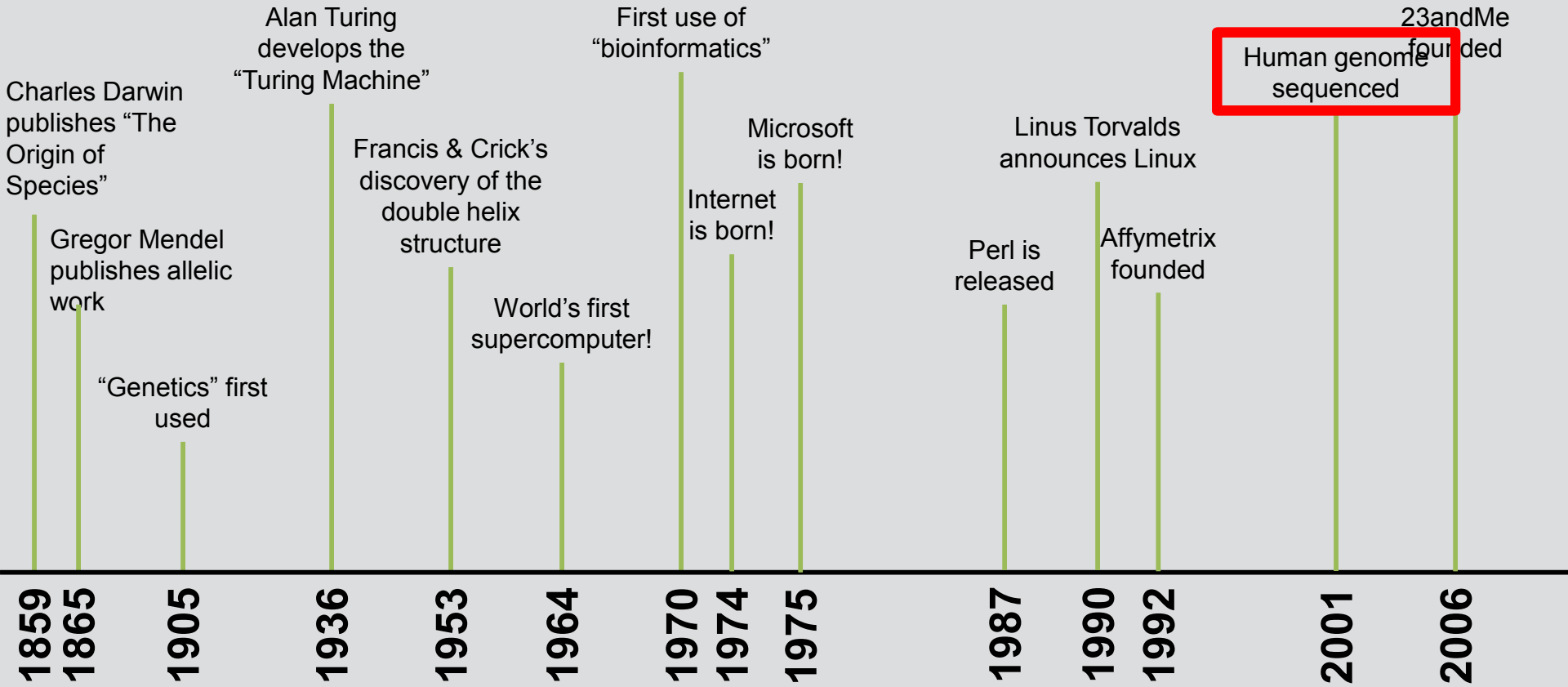
WHAT IS BIOINFORMATICS?

- The term was originally coined in the mid-1980s to refer to analysis of biological sequences*
- Later, used to describe all computer applications in biological sciences.
  - Definition varies
    - Bioinformatics is a new scientific discipline with foundations in computer science and molecular biology (and chemistry and mathematics and statistics and. . .)
    - Very few formally trained bioinformaticians—most have migrated from other fields (myself included)

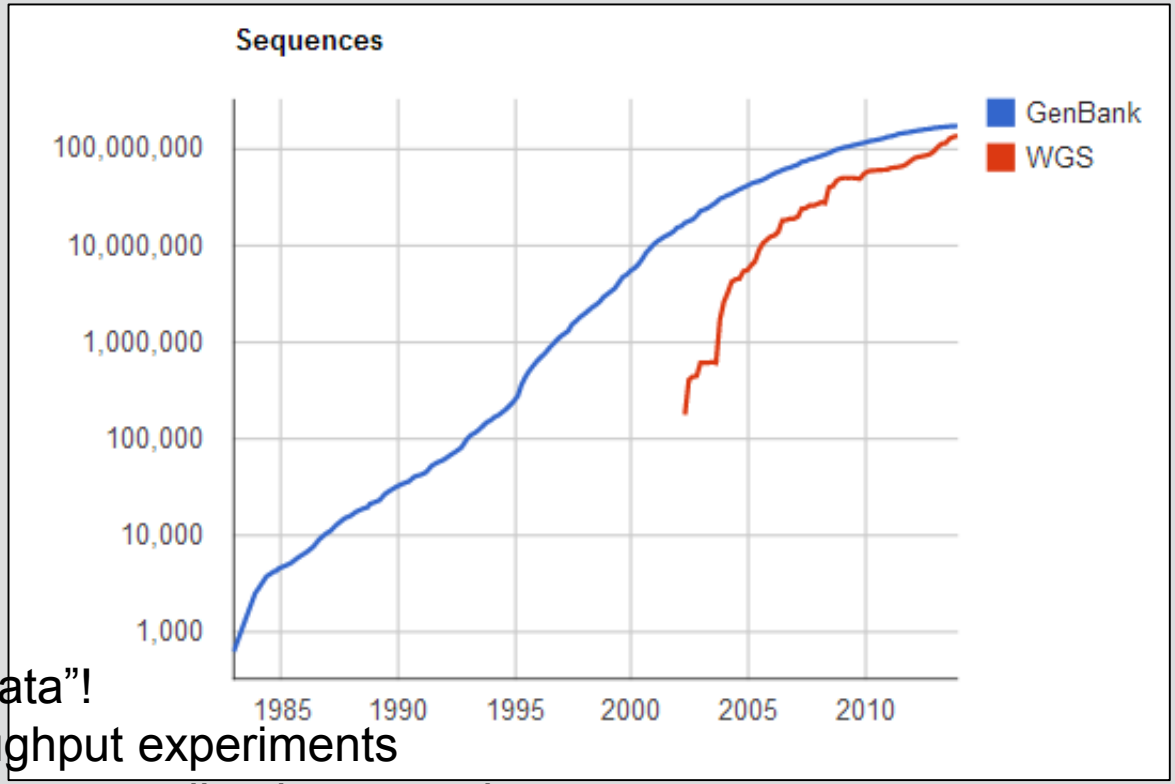
WHAT IS BIOINFORMATICS?



Brief History of Bioinformatics



WHY DO WE NEED BIOINFORMATICS?

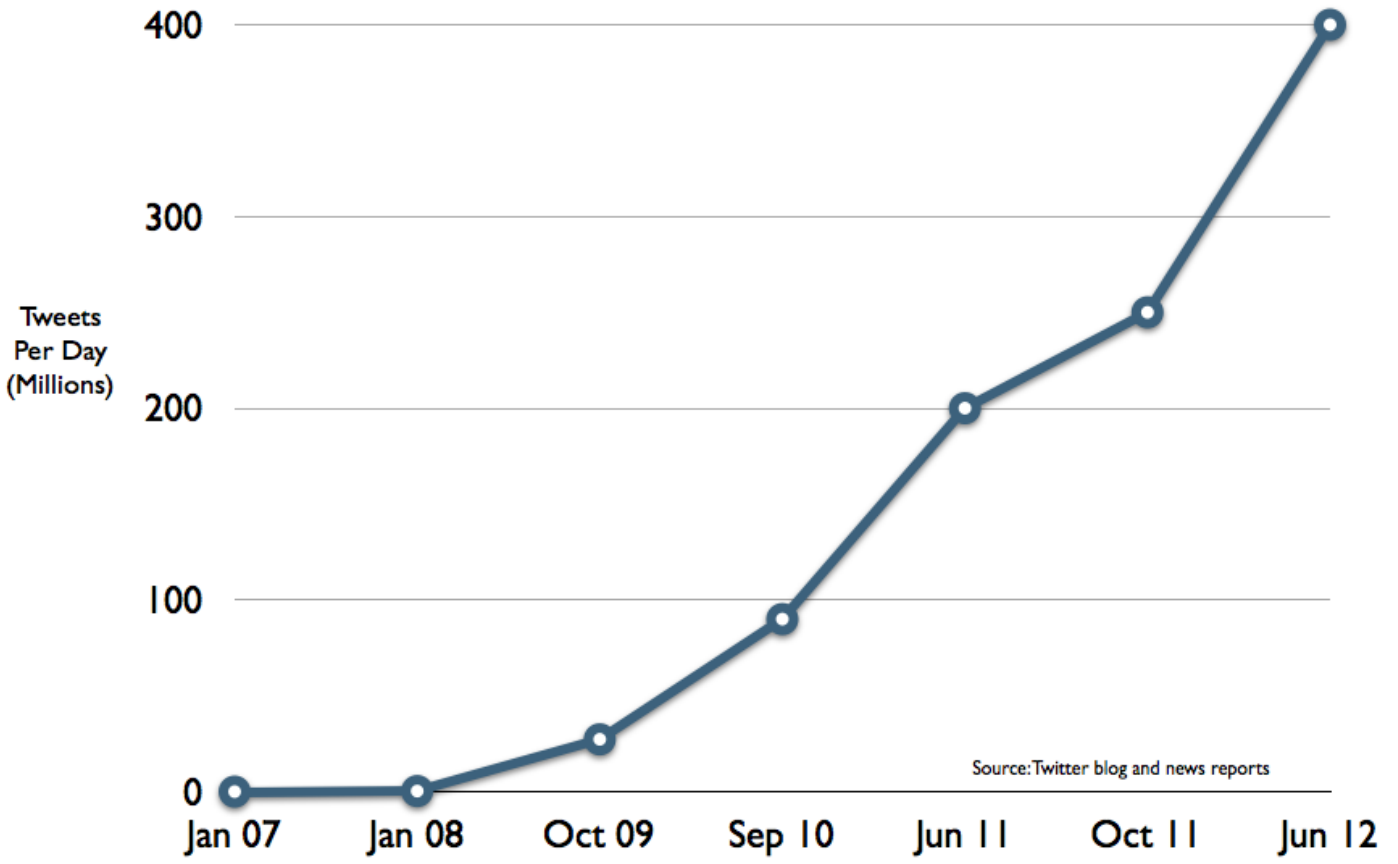


Biomedical “Big data”!

- High-throughput experiments
- Genomes, personalized sequencing
- Complexity of disease
- Health records, public health
- Disaster preparedness

http://www.ncbi.nlm.nih.gov/genbank/statistics

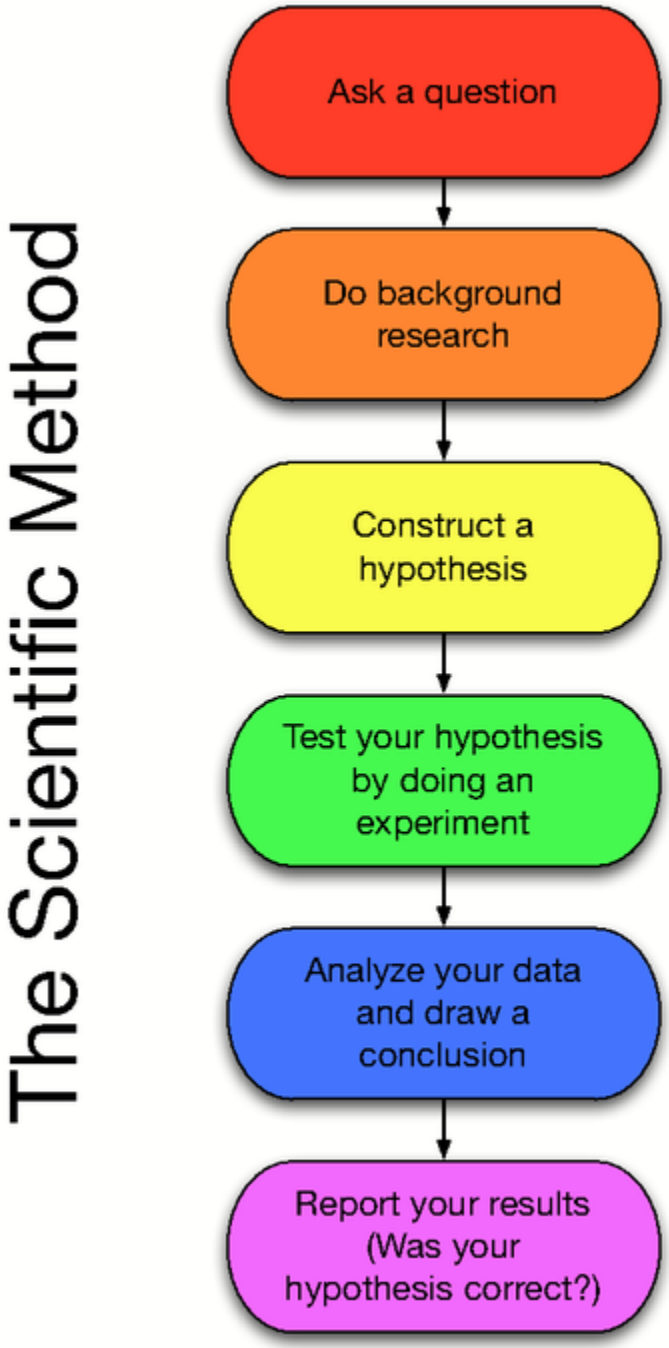
Twitter: Tweets Per Day



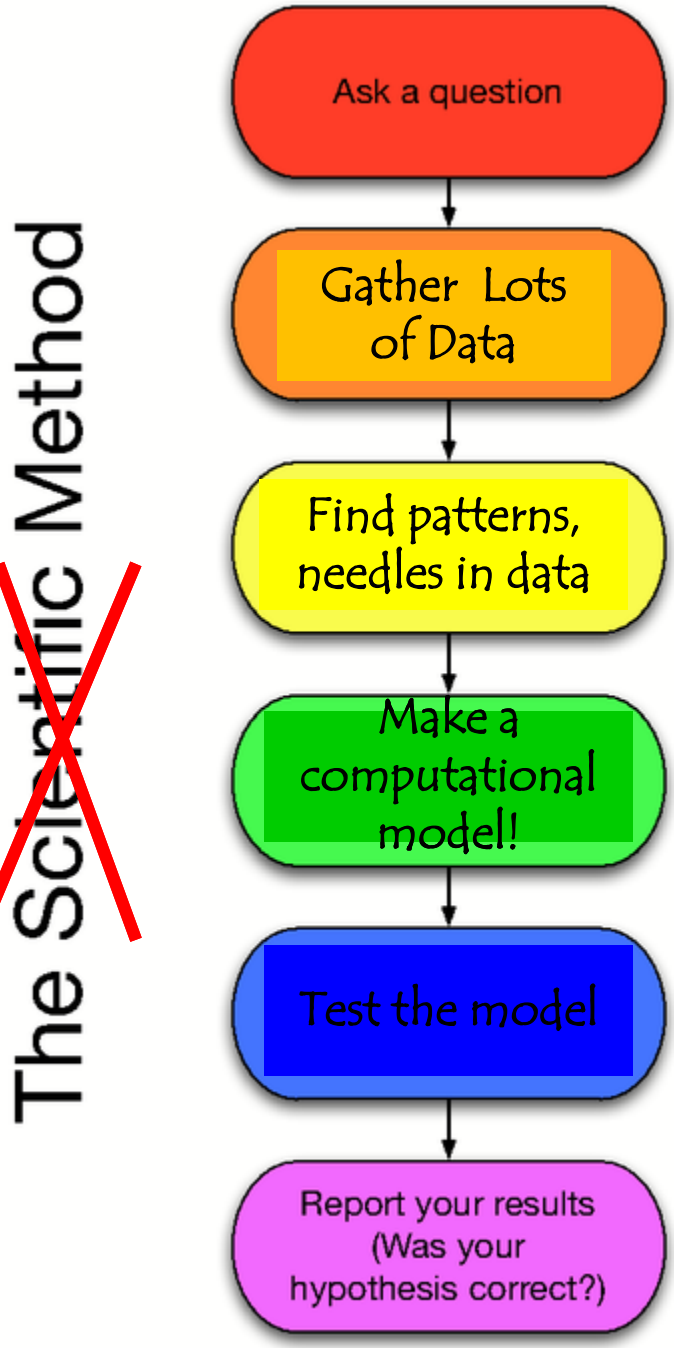
dave@vcdave.com

blogs.forbes.com/davefeinleib

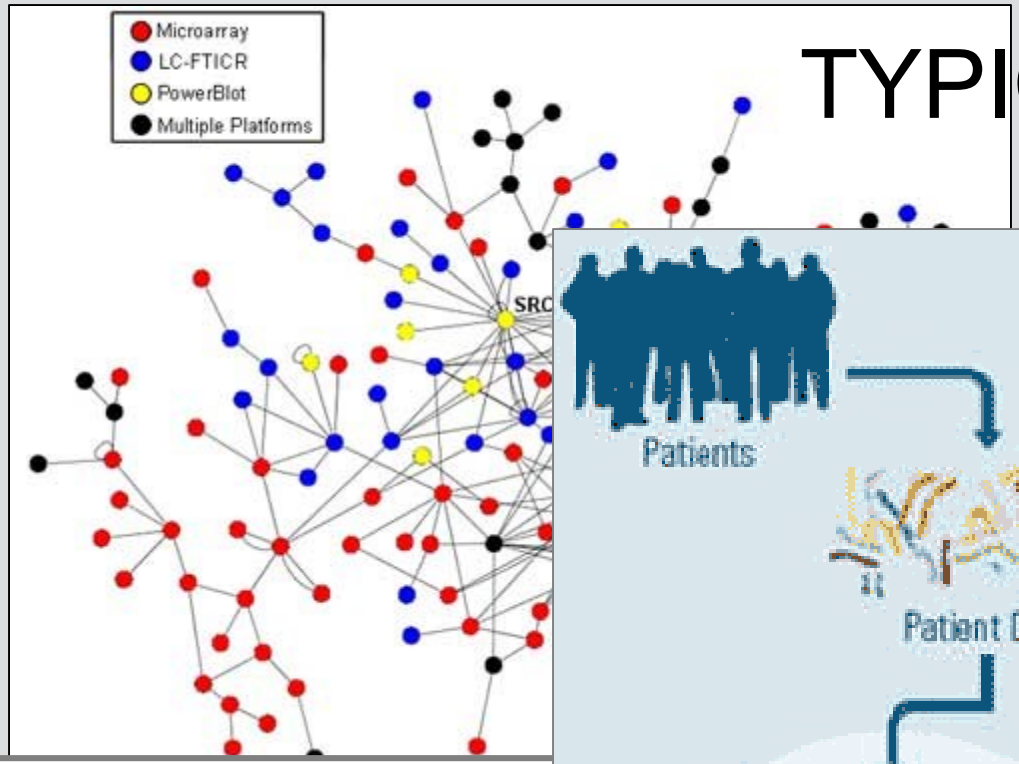
Image Source: http://api.ning.com/files/O6-JQcfS6sxRuZ8I2i5nJVa59xL-krT-a6UqeoLNaHwL2w-JSR-Cy56PmikOywRQgy2gDfYxLAb0Hs*VFr8IePv5QFBJdhDH/BigData.001.jpg



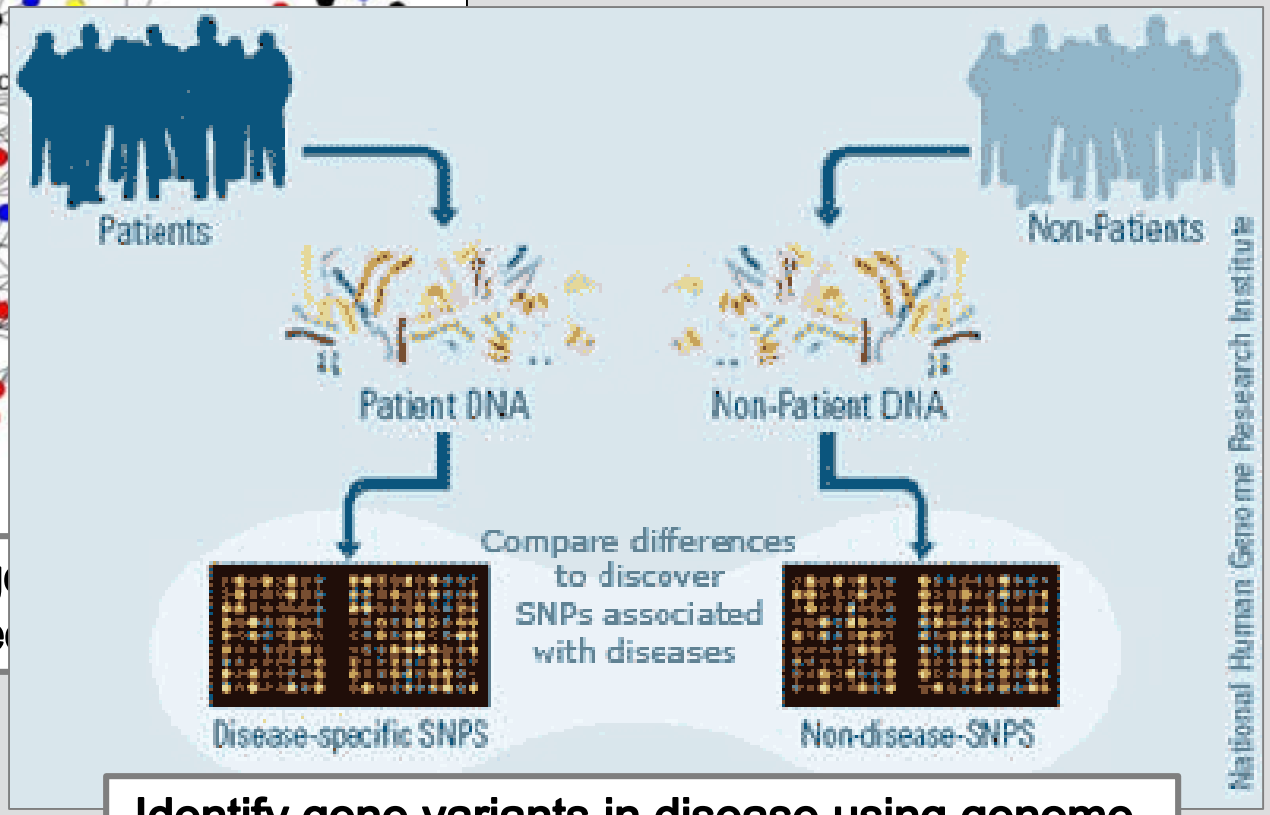
Bioinformatics



TYPICAL PROJECTS



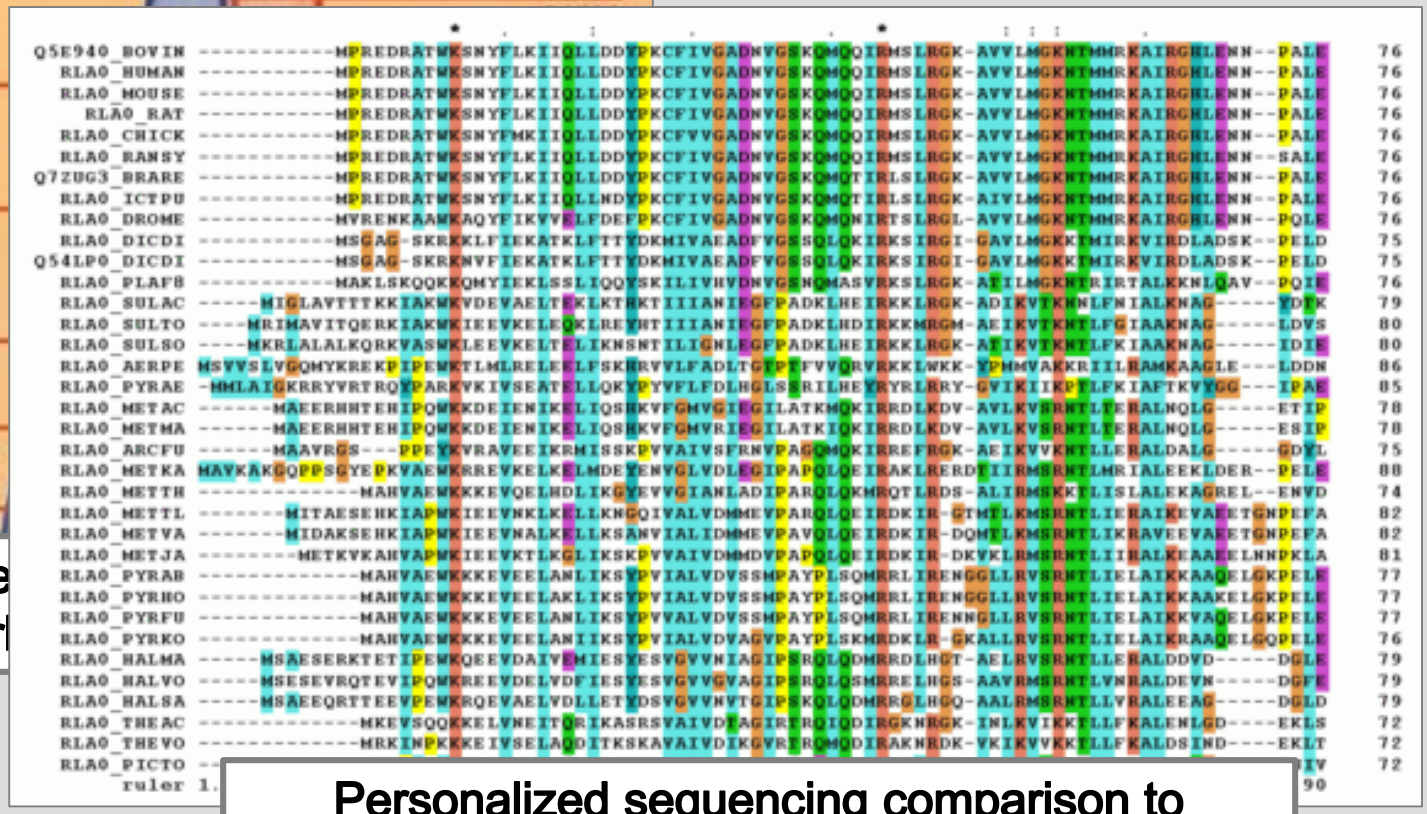
Model co-expression of g
networks & graph the



Identify gene variants in disease using genome-wise association studies (GWAS)

TYPICAL PROJECTS

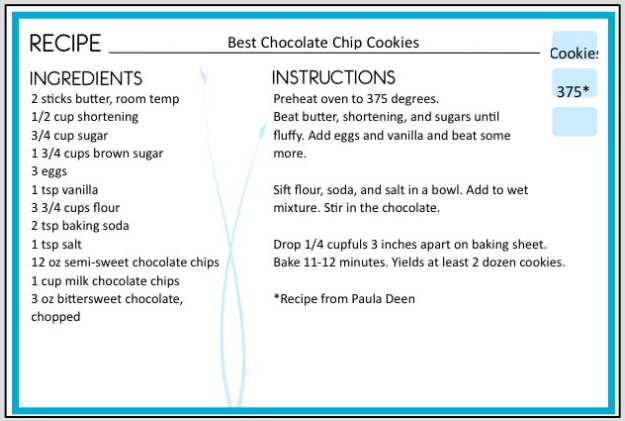
Identify biomarke
for use in earl



Personalized sequencing comparison to reference sequences toward personalized medicine

THE CENTRAL DOGMA OF BIOINFORMATICS

DNA —transcription→ **RNA** —translation→ **PROTEIN**



RECIPE — Best Chocolate Chip Cookies — Cookies

INGREDIENTS
2 sticks butter, room temp
1/2 cup shortening
3/4 cup sugar
1 3/4 cups brown sugar
3 eggs
1 tsp vanilla
3 3/4 cups flour
2 tsp baking soda
1 tsp salt
12 oz semi-sweet chocolate chips
1 cup milk chocolate chips
3 oz bittersweet chocolate, chopped

INSTRUCTIONS
Preheat oven to 375 degrees.
Beat butter, shortening, and sugars until fluffy. Add eggs and vanilla and beat some more.

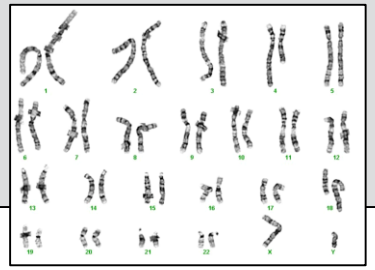
Sift flour, soda, and salt in a bowl. Add to wet mixture. Stir in the chocolate.

Drop 1/4 cupfuls 3 inches apart on baking sheet. Bake 11-12 minutes. Yields at least 2 dozen cookies.

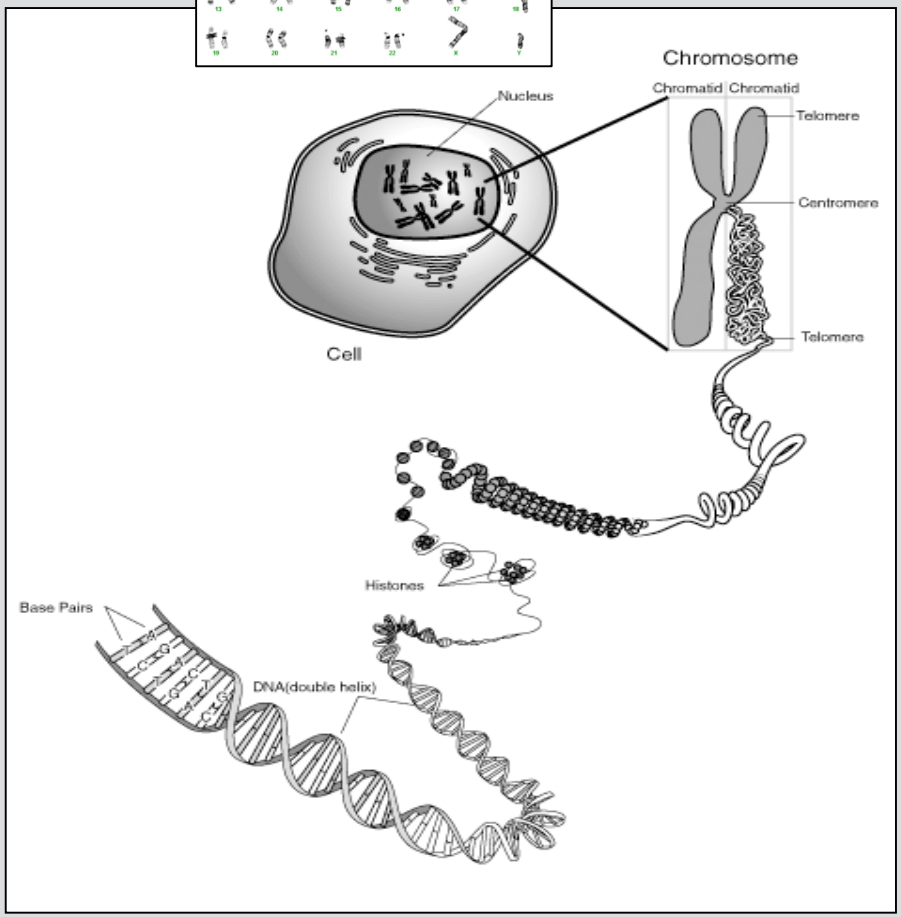
*Recipe from Paula Deen

375*

DNA (Deoxyribose Nucleic Acid)

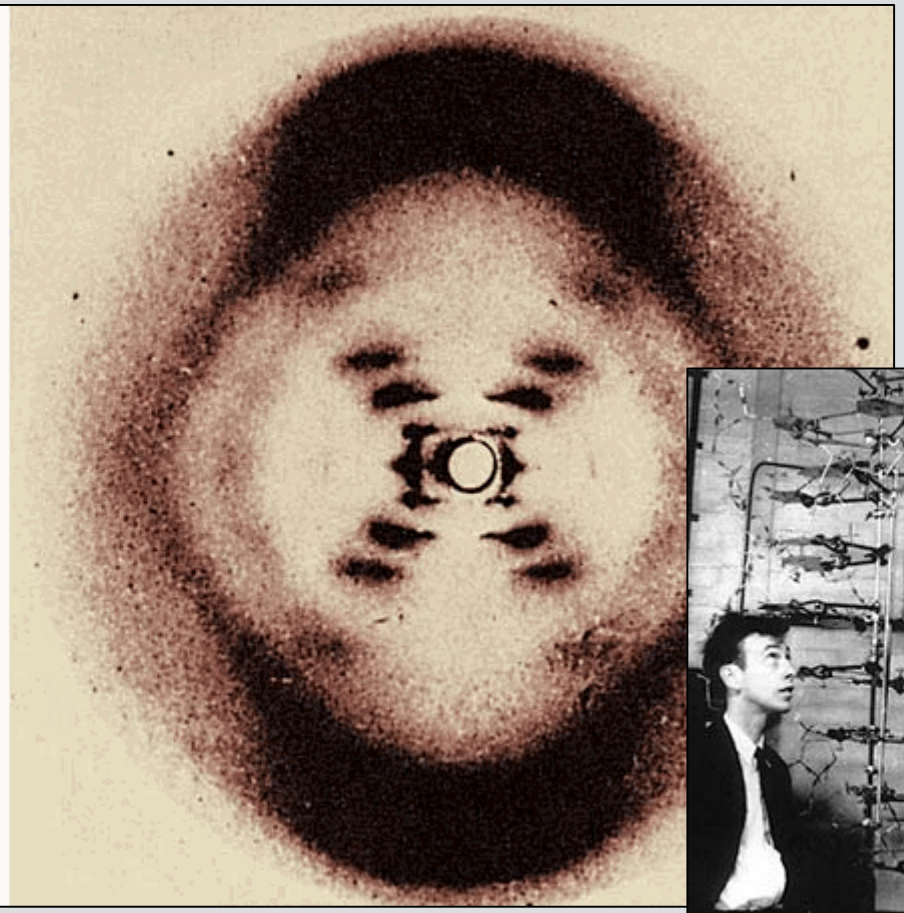


- Genetic material
- Polymer of nitrogenous bases (A, **T**, G & C)
- Contains hereditary information (Genes) in the chromosome
- Chromosome is a thread like linear strand of DNA and associated proteins (histones)
- Chromosomes constitute the genome of an organism



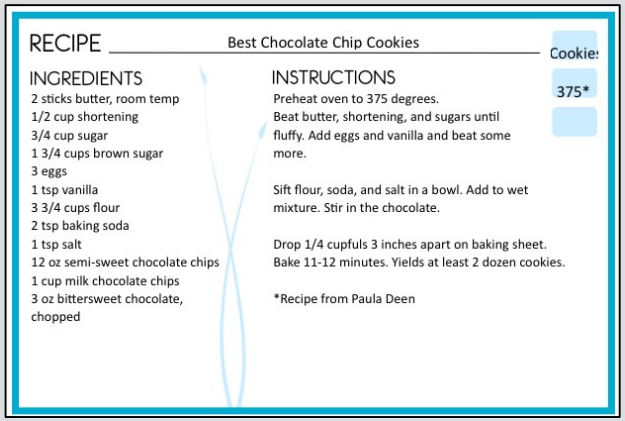

6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X Y

Rosalind Franklin!



THE CENTRAL DOGMA OF BIOINFORMATICS

DNA —transcription→ **RNA** —translation→ **PROTEIN**



RECIPE — Best Chocolate Chip Cookies — Cookies

INGREDIENTS
2 sticks butter, room temp
1/2 cup shortening
3/4 cup sugar
1 3/4 cups brown sugar
3 eggs
1 tsp vanilla
3 3/4 cups flour
2 tsp baking soda
1 tsp salt
12 oz semi-sweet chocolate chips
1 cup milk chocolate chips
3 oz bittersweet chocolate, chopped

INSTRUCTIONS
Preheat oven to 375 degrees.
Beat butter, shortening, and sugars until fluffy. Add eggs and vanilla and beat some more.

Sift flour, soda, and salt in a bowl. Add to wet mixture. Stir in the chocolate.

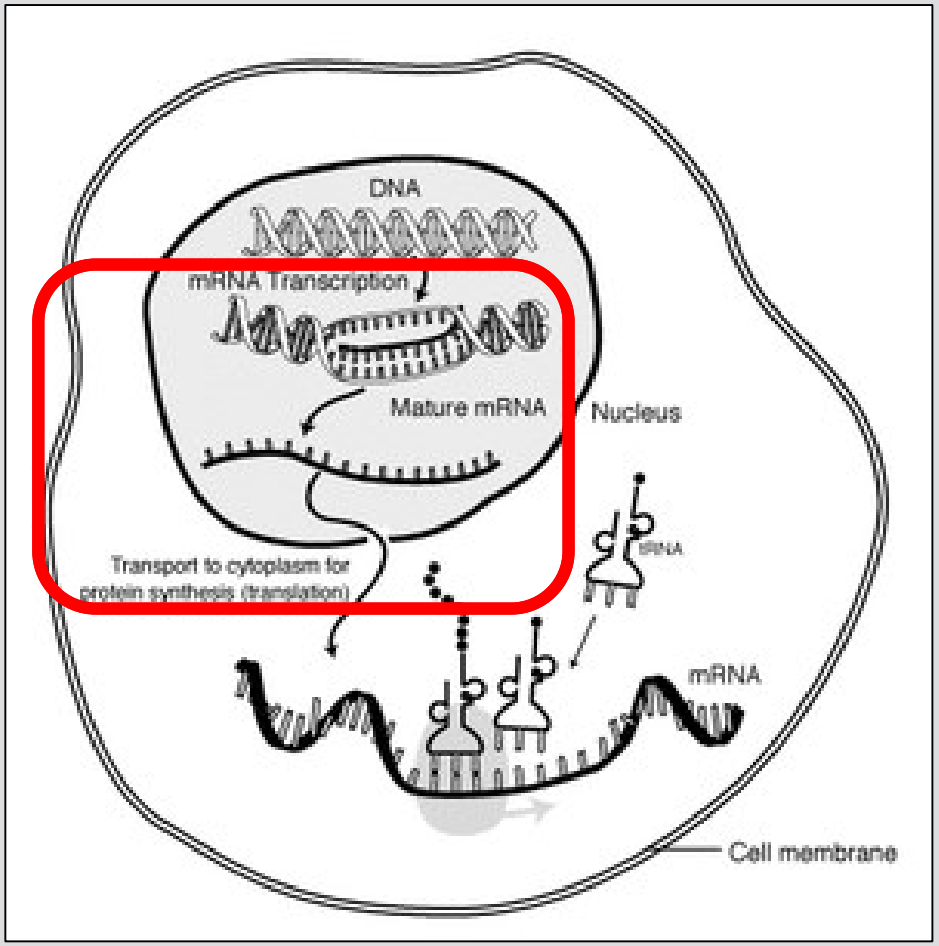
Drop 1/4 cupfuls 3 inches apart on baking sheet. Bake 11-12 minutes. Yields at least 2 dozen cookies.

*Recipe from Paula Deen

375*

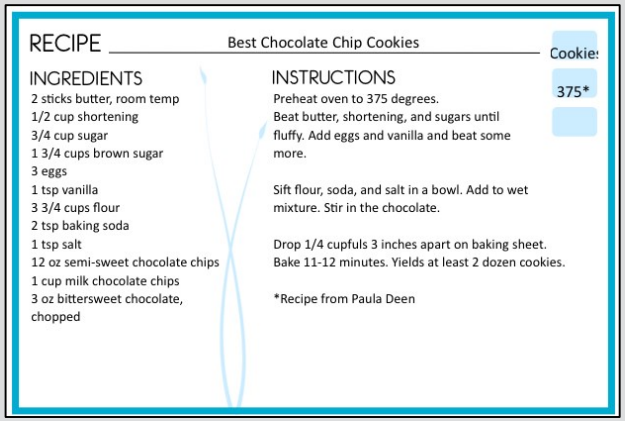
RNA (RiboNucleic Acid)

- Polymer of nucleotides (A, **U,** G & C)
- The temporary copy of a gene
- Copied in the nucleus, transported to cytoplasm to become protein



THE CENTRAL DOGMA OF BIOINFORMATICS

DNA —transcription→ **RNA** —translation→ **PROTEIN**



RECIPE — Best Chocolate Chip Cookies — Cookies

INGREDIENTS
2 sticks butter, room temp
1/2 cup shortening
3/4 cup sugar
1 3/4 cups brown sugar
3 eggs
1 tsp vanilla
3 3/4 cups flour
2 tsp baking soda
1 tsp salt
12 oz semi-sweet chocolate chips
1 cup milk chocolate chips
3 oz bittersweet chocolate, chopped

INSTRUCTIONS
Preheat oven to 375 degrees.
Beat butter, shortening, and sugars until fluffy. Add eggs and vanilla and beat some more.

Sift flour, soda, and salt in a bowl. Add to wet mixture. Stir in the chocolate.

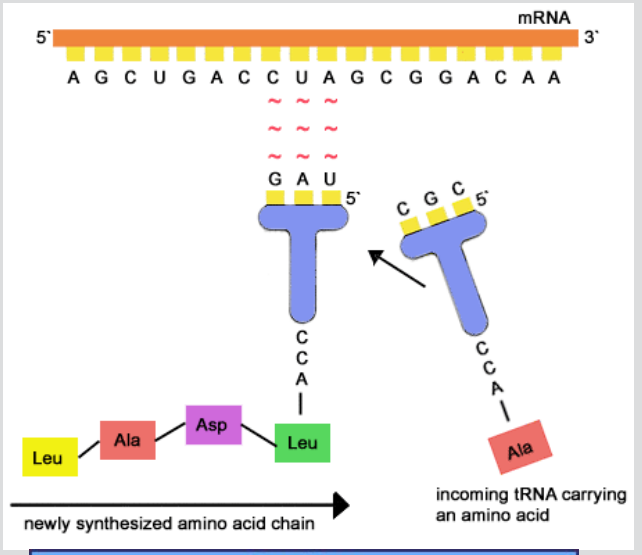
Drop 1/4 cupfuls 3 inches apart on baking sheet. Bake 11-12 minutes. Yields at least 2 dozen cookies.

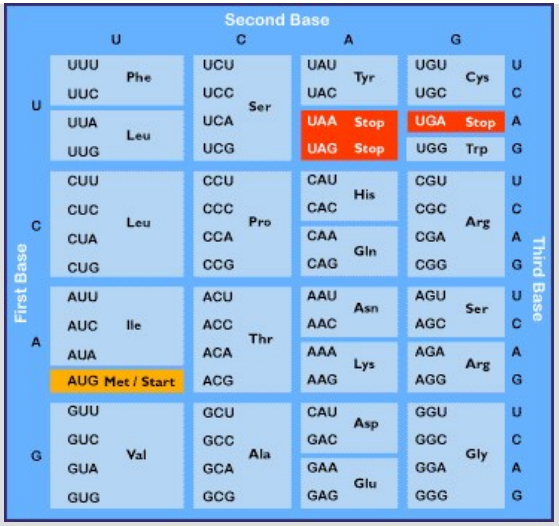
*Recipe from Paula Deen

375*

Proteins

- Functional unit of life
- Polymer of 20 naturally occurring amino acids
- Made from RNA molecules during **translation** by ribosome





Onward to Activity 1

- Central Dogma
  - If you remember one thing, remember this!
- Bioinformatics has roots in biology
- To learn what the human genome is, we must first get the genome out of the cells!

ACTIVITY 1:

Extracting DNA from a Strawberry

http://www.youtube.com/watch?v=hOpu4iN5Bh4&noredirect=1

http://www.fanpop.com/clubs/fruit/images/34914851/title/strawberry-photo
http://img.wonderhowto.com/img/97/38/63488691134384/0/extract-dna-from-strawberry-with-basic-kitchen-items.w654.jpg

OVERVIEW

- Introductions
- Part I: What is bioinformatics?
  - Activity 1: Extraction of DNA
- Part II: How is bioinformatics useful?
  - Activity 2: Finding Disease Genes
- Conclusions

OVERVIEW

- Introductions
- Part I: What is bioinformatics?
  - Activity 1: Extraction of DNA
- Part II: How is bioinformatics useful?
  - Activity 2: Finding Disease Genes
- Conclusions

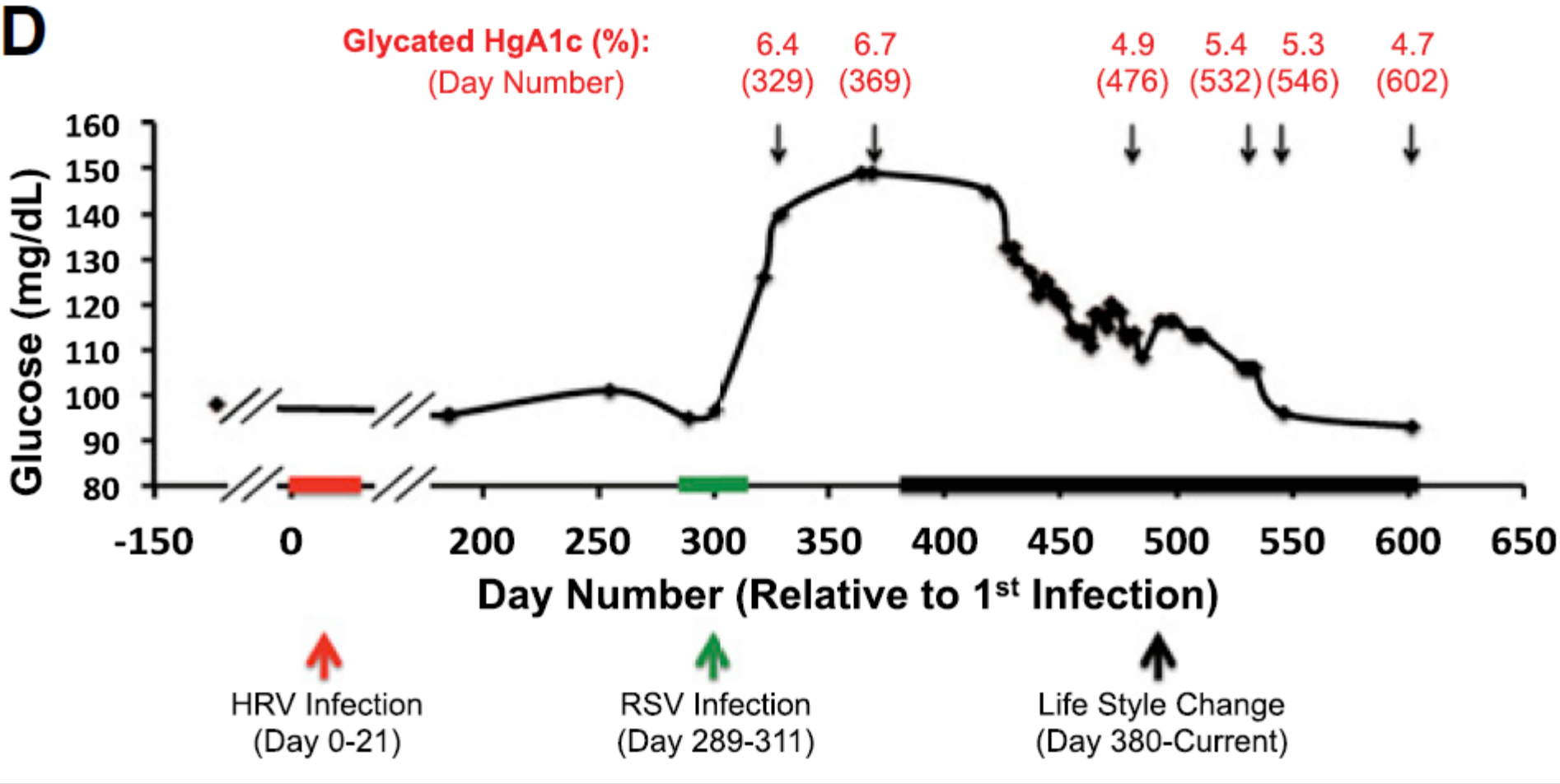
Why don’t we have personalized medicine?

Where is the cure for cancer?

Why is AIDS still misunderstood?

Personalized Medicine Study

- 54yr old male volunteer
- Plasma and serum used for testing
- 14 month time course
- Complete medical exams and labs at each meeting (20 time points total)
- Extensive sampling at 2 periods of viral infection:
  - HRV (human rhinovirus) – common cold
  - RSV (respirtatory synticial) - bronchitis


D
Glycated HgA1c (%):
(Day Number)
6.4 (329)
6.7 (369)
4.9 (476)
5.4 (532)
5.3 (546)
4.7 (602)
Glucose (mg/dL)
160
150
140
130
120
110
100
90
80
-150
0
200
250
300
350
400
450
500
550
600
650
Day Number (Relative to 1st Infection)
HRV Infection
(Day 0-21)
RSV Infection
(Day 289-311)
Life Style Change
(Day 380-Current)

Resource

Cell

Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes

Rui Chen,[1,11] George I. Mias,[1,11] Jennifer Li-Pook-Than,[1,11] Lihua Jiang,[1,11] Hugo Y.K. Lam,[1,12] Rong Chen,[2,12] Elana Miriami,[1] Konrad J. Karczewski,[1] Manoj Hariharan,[1] Frederick E. Dewey,[3] Yong Cheng,[1] Michael J. Clark,[1] Hogune Im,[1] Lukas Habegger,[6,7] Suganthi Balasubramanian,[6,7] Maeve O'Huallachain,[1] Joel T. Dudley,[2] Sara Hillenmeyer,[1] Rajini Haraksingh,[1] Donald Sharon,[1] Ghia Euskirchen,[1] Phil Lacroute,[1] Keith Bettinger,[1] Alan P. Boyle,[1] Maya Kasowski,[1] Fabian Grubert,[1] Scott Seki,[2] Marco Garcia,[2] Michelle Whirl-Carrillo,[1] Mercedes Gallardo,[9,10] Maria A. Blasco,[9] Peter L. Greenberg,[4] Phyllis Snyder,[1] Teri E. Klein,[1] Russ B. Altman,[1,5] Atul J. Butte,[2] Euan A. Ashley,[3] Mark Gerstein,[6,7,8] Kari C. Nadeau,[2] Hua Tang,[1] and Michael Snyder[1,*]

[1]Department of Genetics, Stanford University School of Medicine

[2]Division of Systems Medicine and Division of Immunology and Allergy, Department of Pediatrics

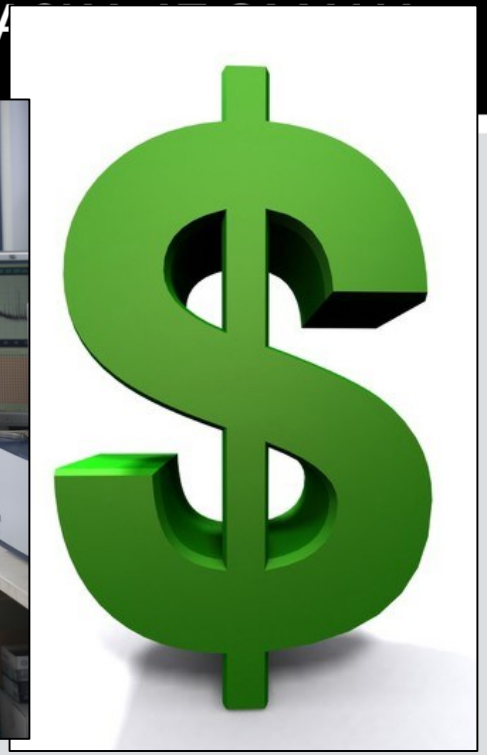
[3]Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine

[4]Division of Hematology, Department of Medicine

[5]Department of Bioengineering

Stanford University, Stanford, CA 94305, USA

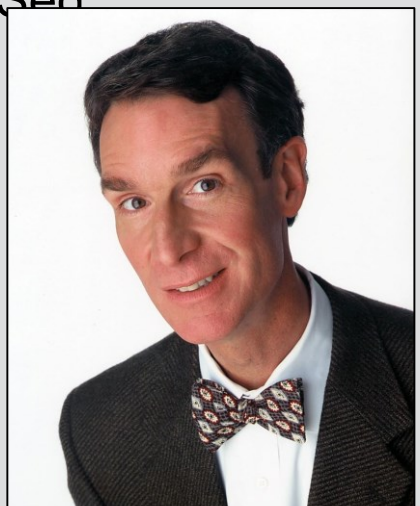
Techniqu



chniques used:

- HRV and RSV
- Whole-genom
- Whole-exome
- Sanger-DNA s
- Whole-transcr mRNA-Seq
- Small RNA se Seq

teome Profiling
Profiling
ofiling
rofiling
ssay
s



Why don't we have personalized medicine?

Where is the cure for cancer?

Why is AIDS still misunderstood?

We don't know everything.

There's lots and lots of data.

Life is complex.

Everyone is unique.

Databases

- Pubmed – Journal articles on biomedical research
- OMIM – Disease genes in humans
- Genbank – All known data on genes and their proteins, and their DNA sequence
- PDB – 3-D proteins structures
- MGD – Organism specifc (mouse)

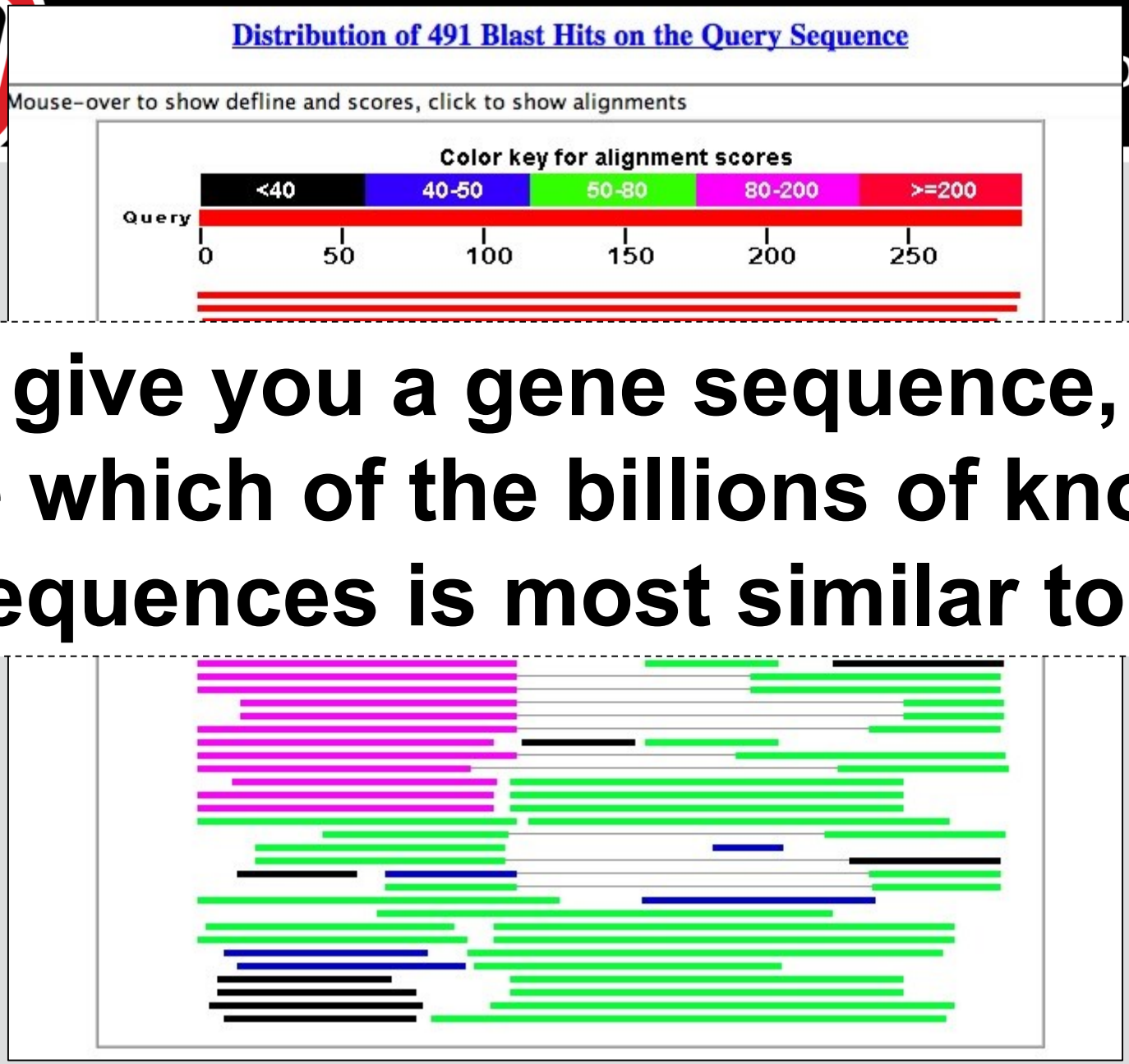
Phenotypic Alleles

Query Results -- Summary

Symbol	Kl
Name	klotho
ID	MGI:1101771

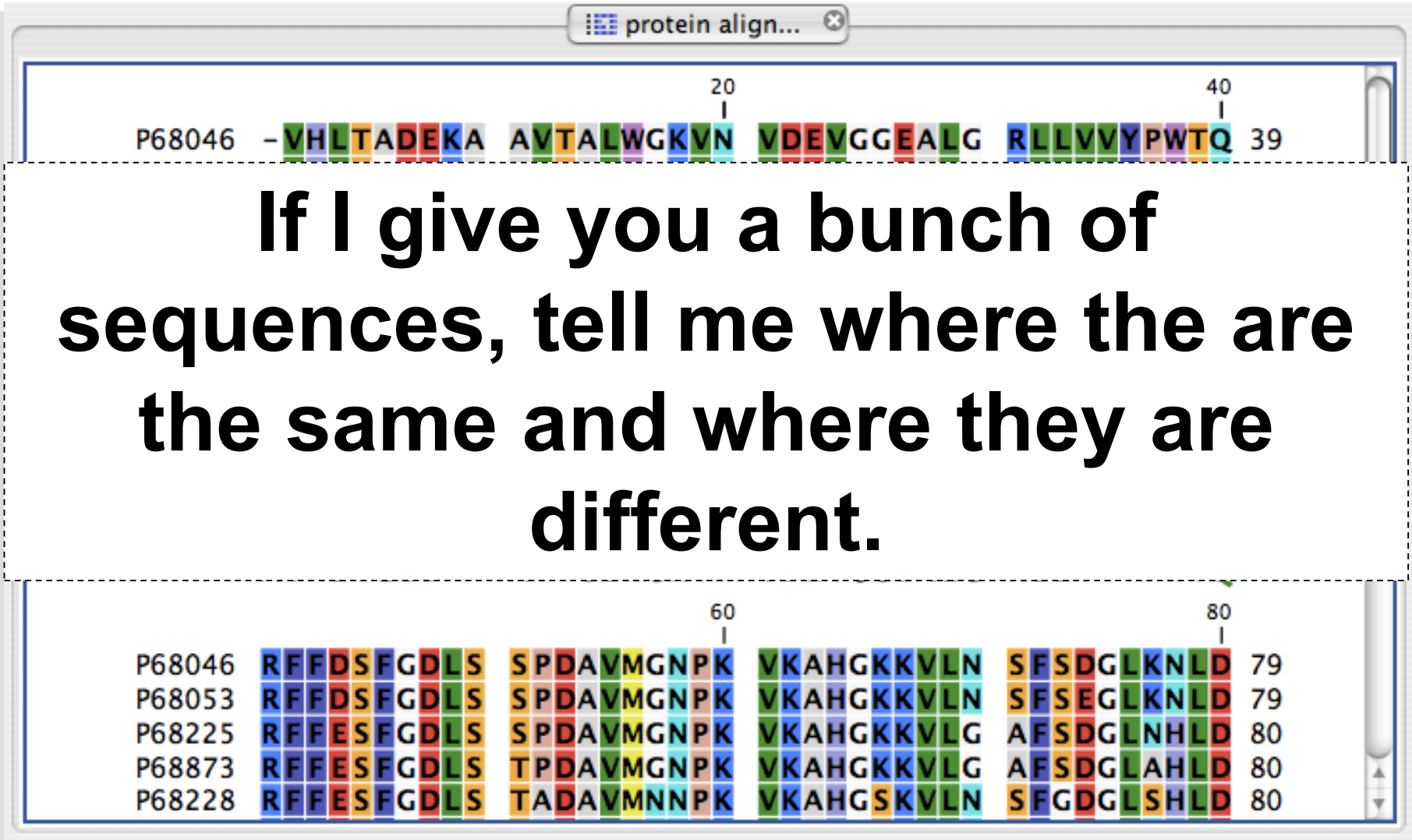
9 matching Alleles (1 Gene/Marker represented)

Allele Symbol Gene; Allele Name	Chr	Synonyms	Category	Abnormal Phenotypes Reported in these Systems	Human Disease Models
Kl^{kl} klotho; klotho	5	alpha-kl-, kl, Kl-, $klotho^{hm}$	Transgenic (random, gene disruption)	adipose, behavior, cardiovascular, cellular, endocrine/exocrine, growth/size, hematopoietic, homeostasis, immune, integument, limbs/digits/tail, liver/biliary, mortality/aging, muscle, nervous system, other, reproductive, respiratory, skeleton	Emphysema, Hereditary Pulmonary 130700; Klotho; KL 604824
$Kl^{tm1.1Tel}$ klotho; targeted mutation 1.1, Tobias E Larsson	5	$Klotho^{flox}$	Targeted (Floxed/Frt)	homeostasis, renal/urinary	
$Kl^{tm1.2Tel}$ klotho; targeted mutation 1.2, Tobias E Larsson	5	beta-KL-	Targeted (knock-out)	behavior, growth/size, homeostasis, mortality/aging, renal/urinary, skeleton	
Kl^{tm1Lex} klotho; targeted mutation 1, Lexicon Genetics	5		Targeted (knock-out)	growth/size, homeostasis, limbs/digits/tail, mortality/aging, skeleton	
Kl^{tm1Yin} klotho; targeted mutation 1, Yo-ichi Nabeshima	5	kl-	Targeted (knock-out)	adipose, growth/size, homeostasis, mortality/aging, skeleton	
Kl^{tm2Yin} klotho; targeted mutation 2, Yo-ichi Nabeshima	5	kl-geo	Targeted (Reporter)	cardiovascular, mortality/aging	
$Kl^{Gt(522F6)Cmhd}$ klotho; gene trap 522F6, Centre for Modeling Human Disease	5		Gene trapped ***(Cell Line)***		
$Kl^{tm1a(EUCOMM)Hmgu}$ klotho; targeted mutation 1a, Helmholtz Zentrum Muenchen GmbH	5		Targeted (Floxed/Frt) ***(Cell Line)***		
$Kl^{tm1e(EUCOMM)Hmgu}$ klotho; targeted mutation 1e, Helmholtz Zentrum Muenchen GmbH	5		Targeted (Reporter) ***(Cell Line)***		



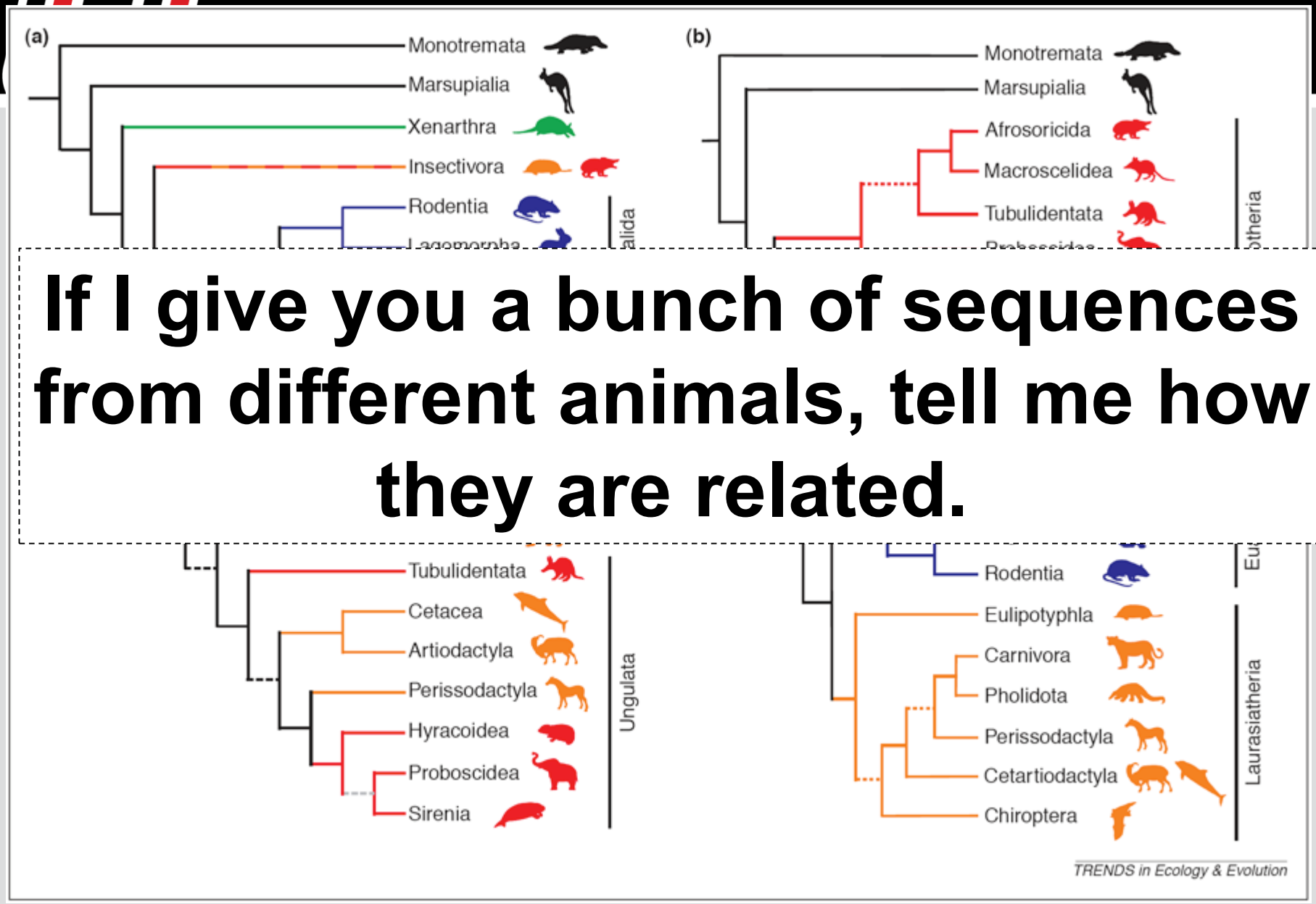
If I give you a gene sequence, tell me which of the billions of known sequences is most similar to it.

BLAST



If I give you a bunch of sequences, tell me where the are the same and where they are different.

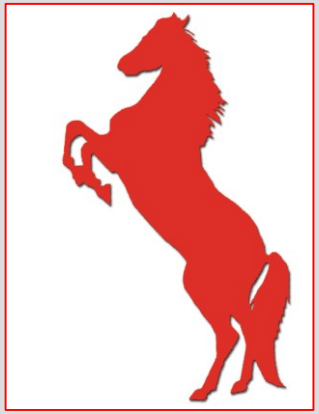
Alignment (Clustal, MUSCLE, Tcoffee)



If I give you a bunch of sequences from different animals, tell me how they are related.

Phylogenetics

```
>Mouse
ATTCAGATCA
>Rat
TTTCAGATCG
>Horse
TACCAATCGC
```



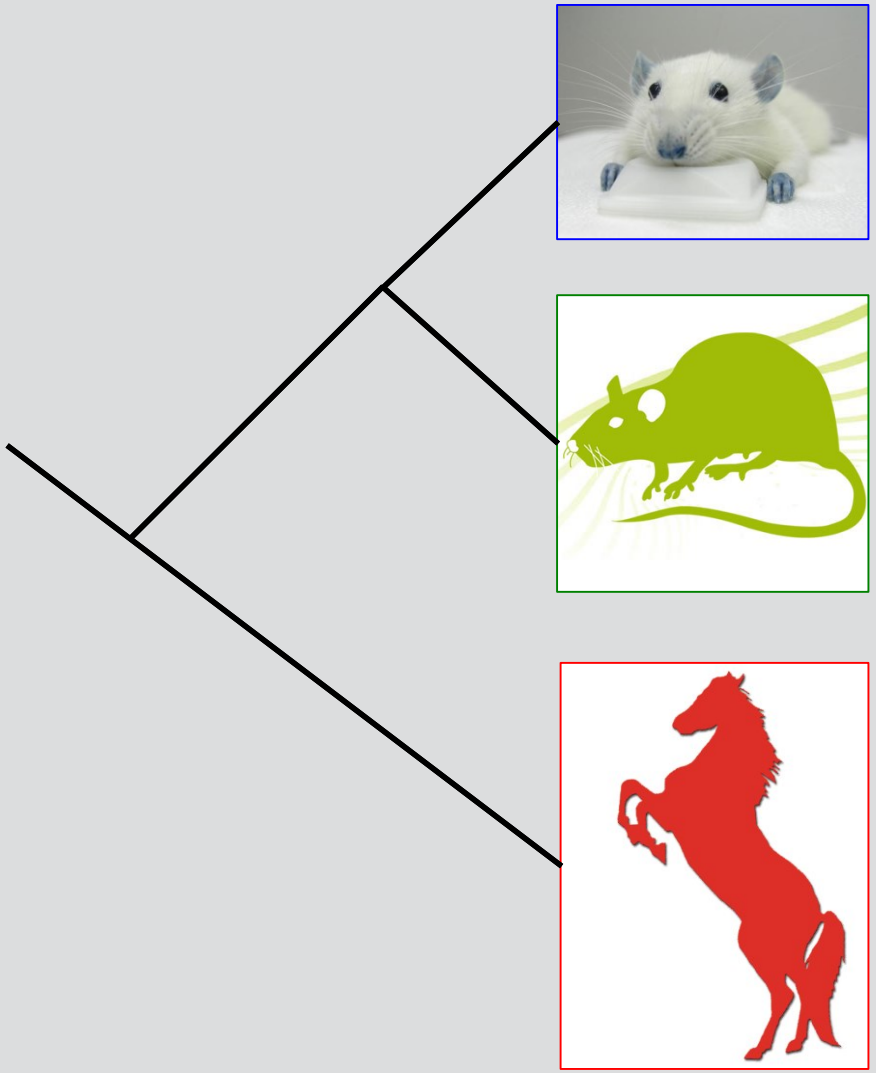
ATTCAGATCA

80% Similar

TTTCAGATCG

20% Similar

TACCAATCGC




99.9%


96-99.4%


94%


50-60%

We already have some databases and tools .

- ….but we need more to solve those questions.

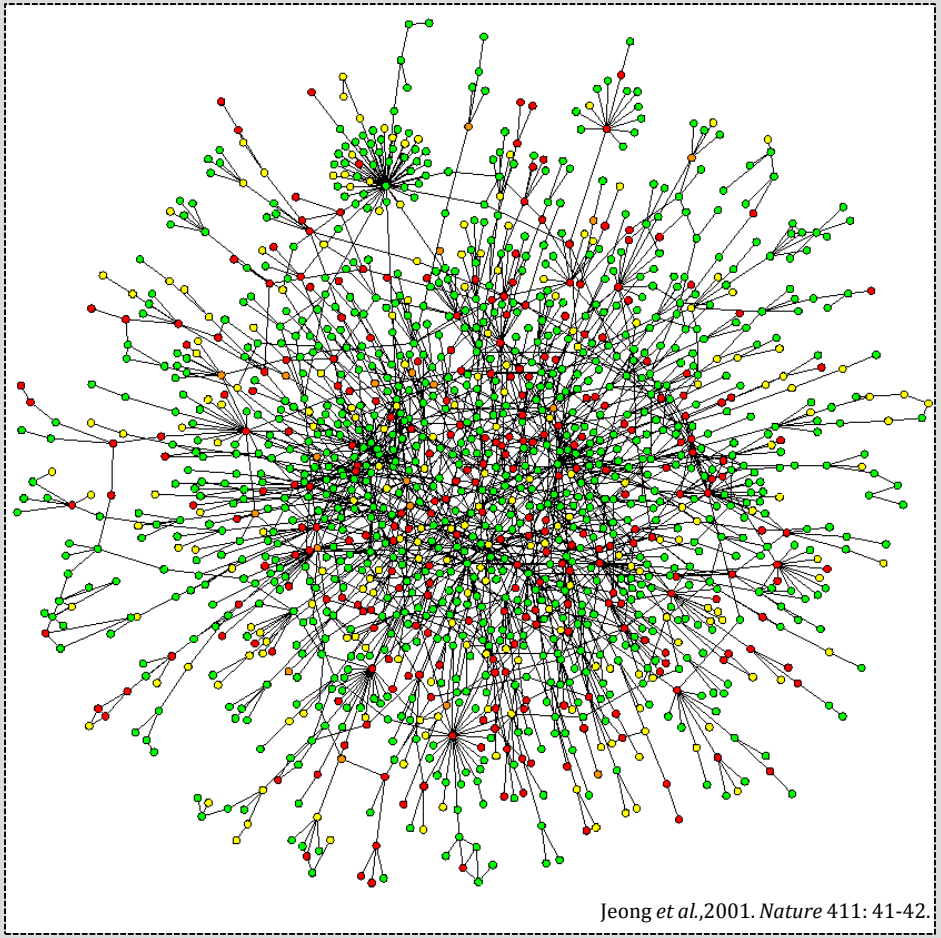
- Example: A disease where only *one set of 3 DNA bases is missing*.
  - Do you know what this disease is?

- Activity 2:
  - The knowledge in bioinformatics databases
  - How to use some tools: BLAST, Alignment, Translation!

OVERVIEW

- Introductions
- Part I: What is bioinformatics?
  - Activity 1: Extraction of DNA
- Part II: How is bioinformatics useful?
  - Activity 2: Finding Disease Genes
- Conclusions

- Massive amounts of data
- Many generation methods
- FEW ANALYSIS methods
- “Signal corruption”
- *How to model data??*
- *How to extract knowledge?*



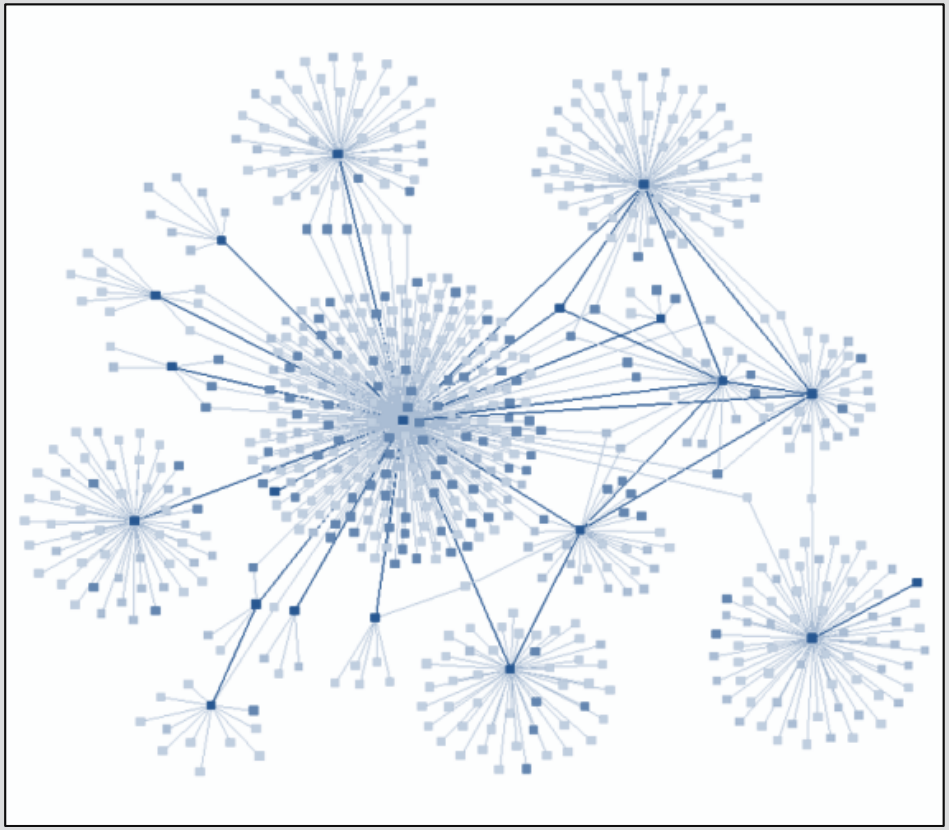
Jeong *et al.*,2001. *Nature* 411: 41-42.

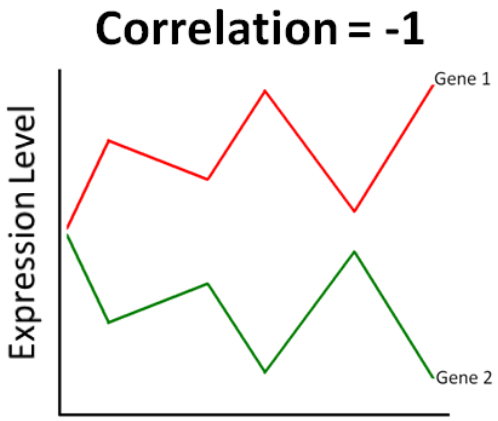
The Network Model

- A network:
- Elements and their interactions.

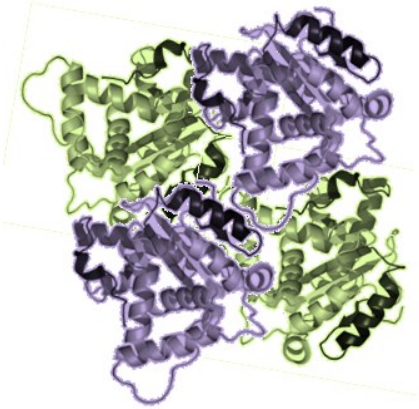
- Nodes → elements
- Edges → interactions

- Any relationship can be modeled using the network model



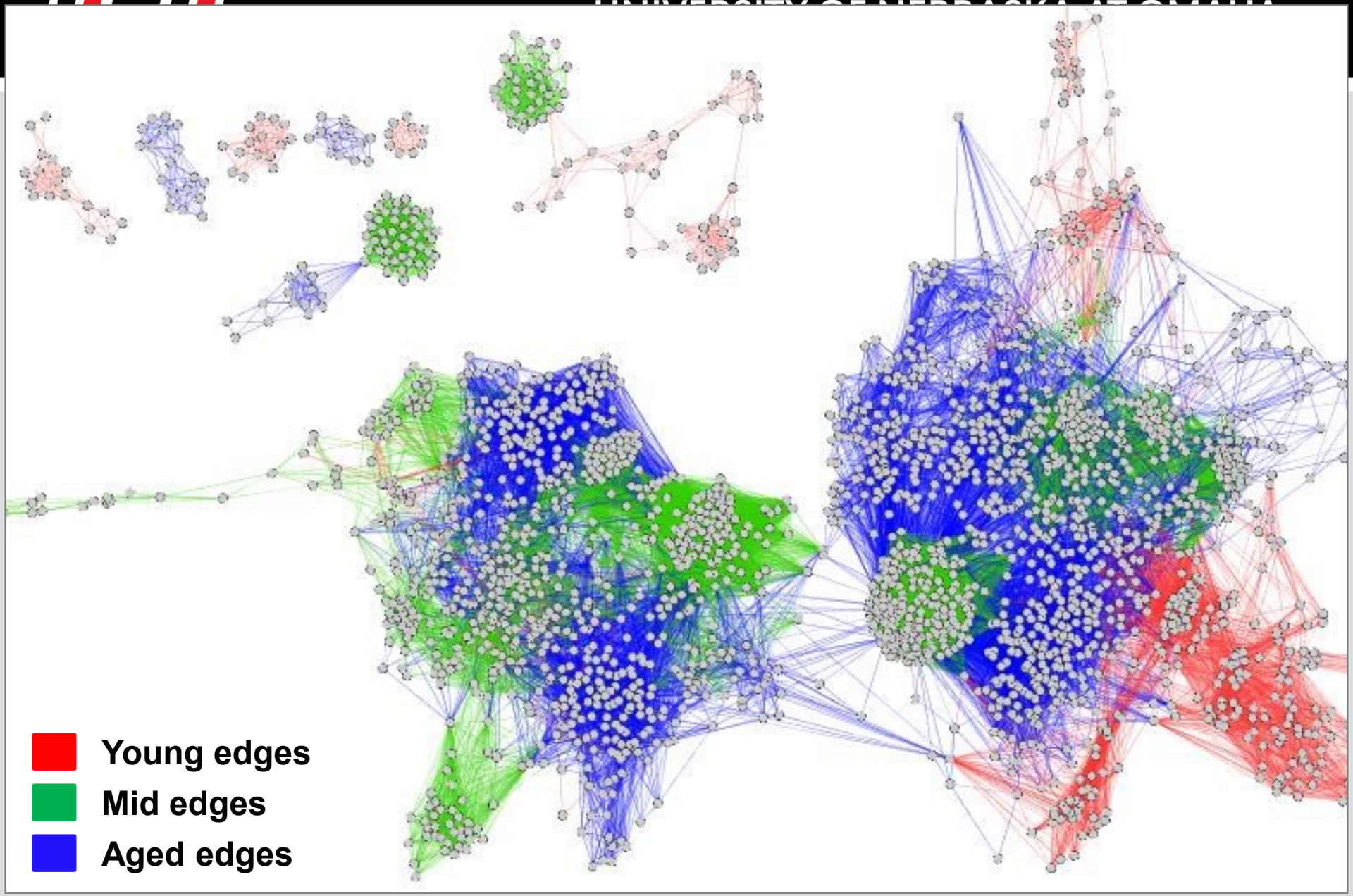

Correlation = -1
Expression Level
Gene 1
Gene 2


Gene 1
Gene 2

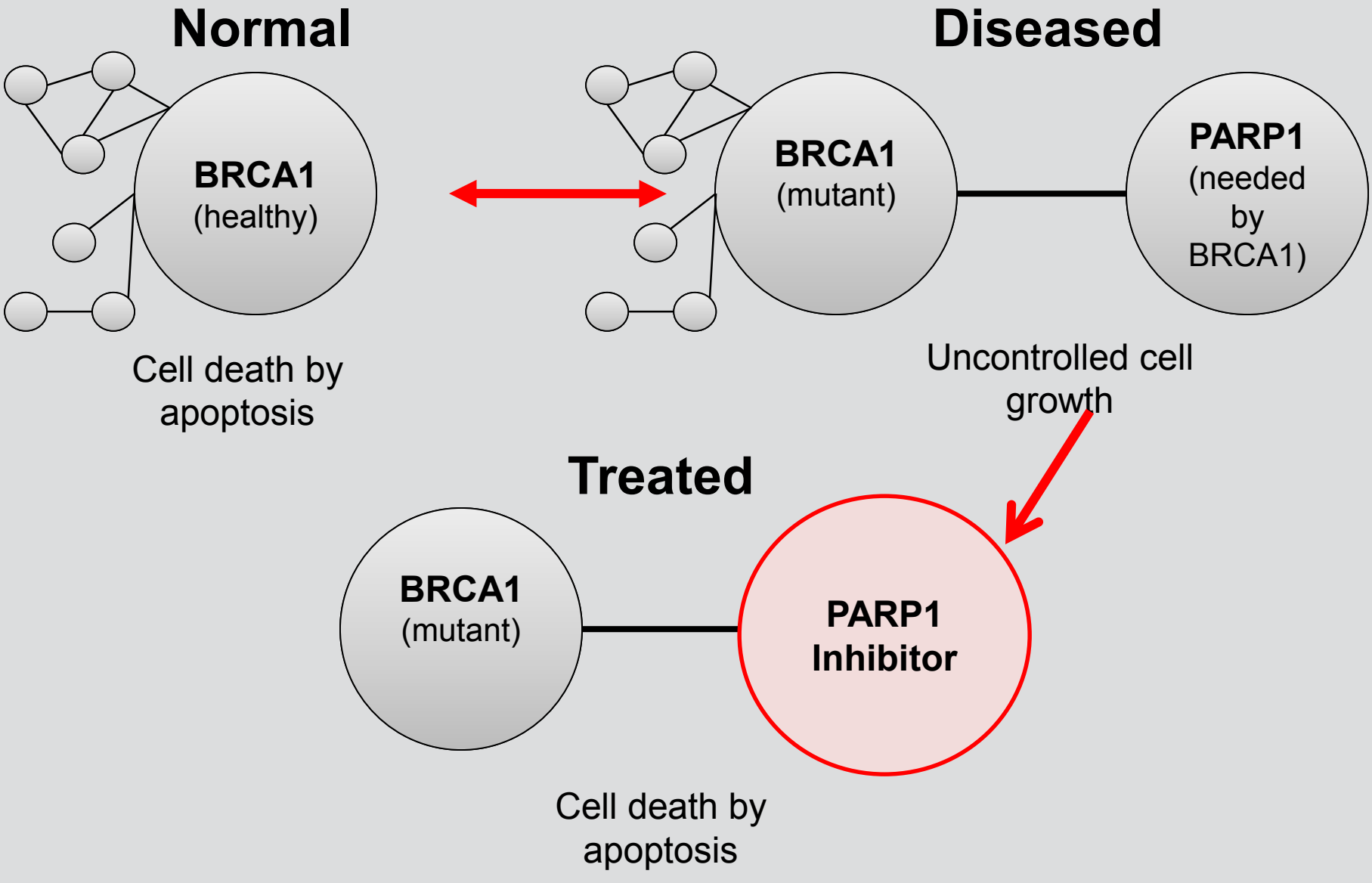



Prot A
Prot B

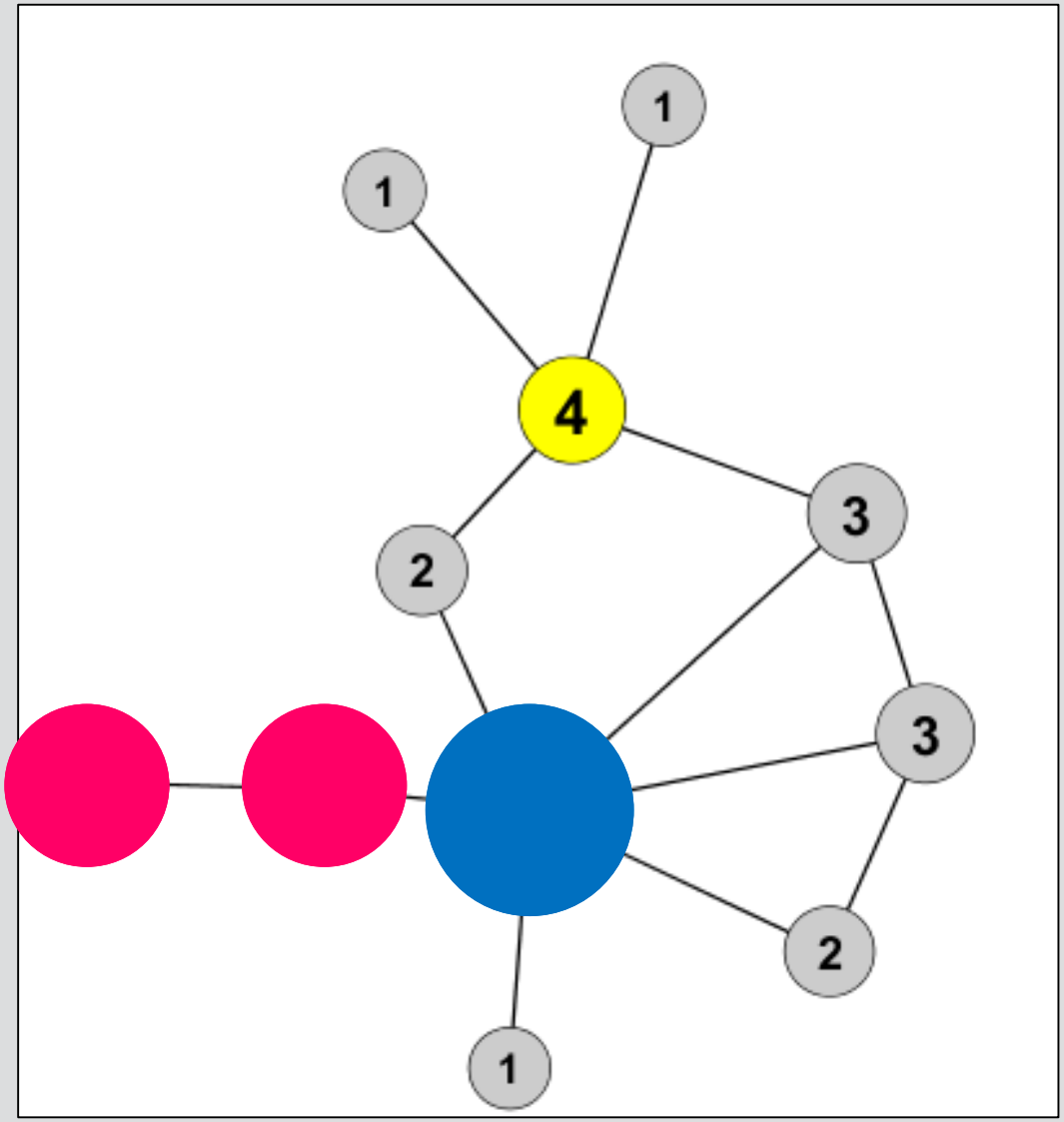

John
Bob

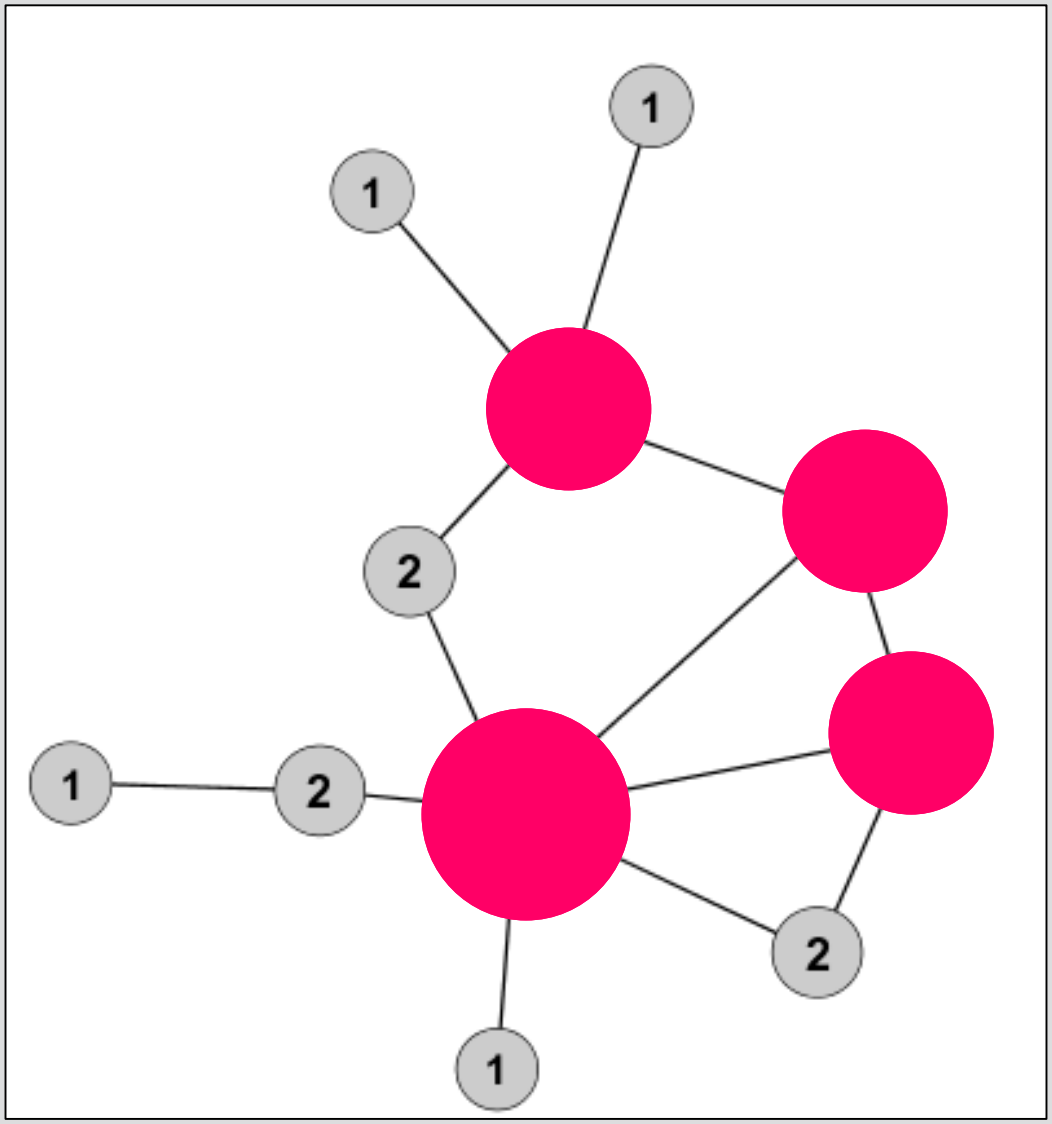

Young edges
Mid edges
Aged edges

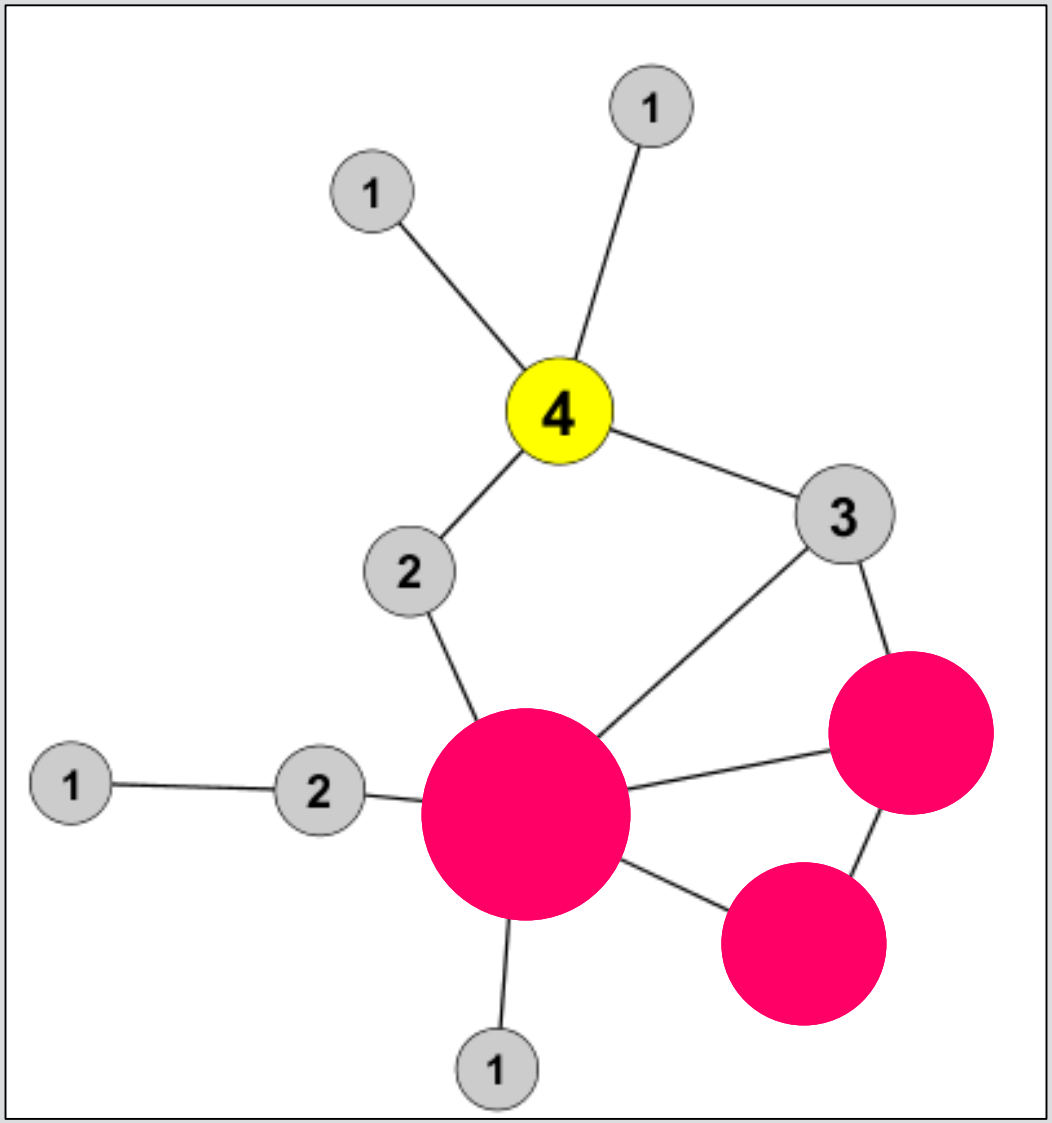
EXAMPLE: SYNTHETIC LETHALITY NETWORKS

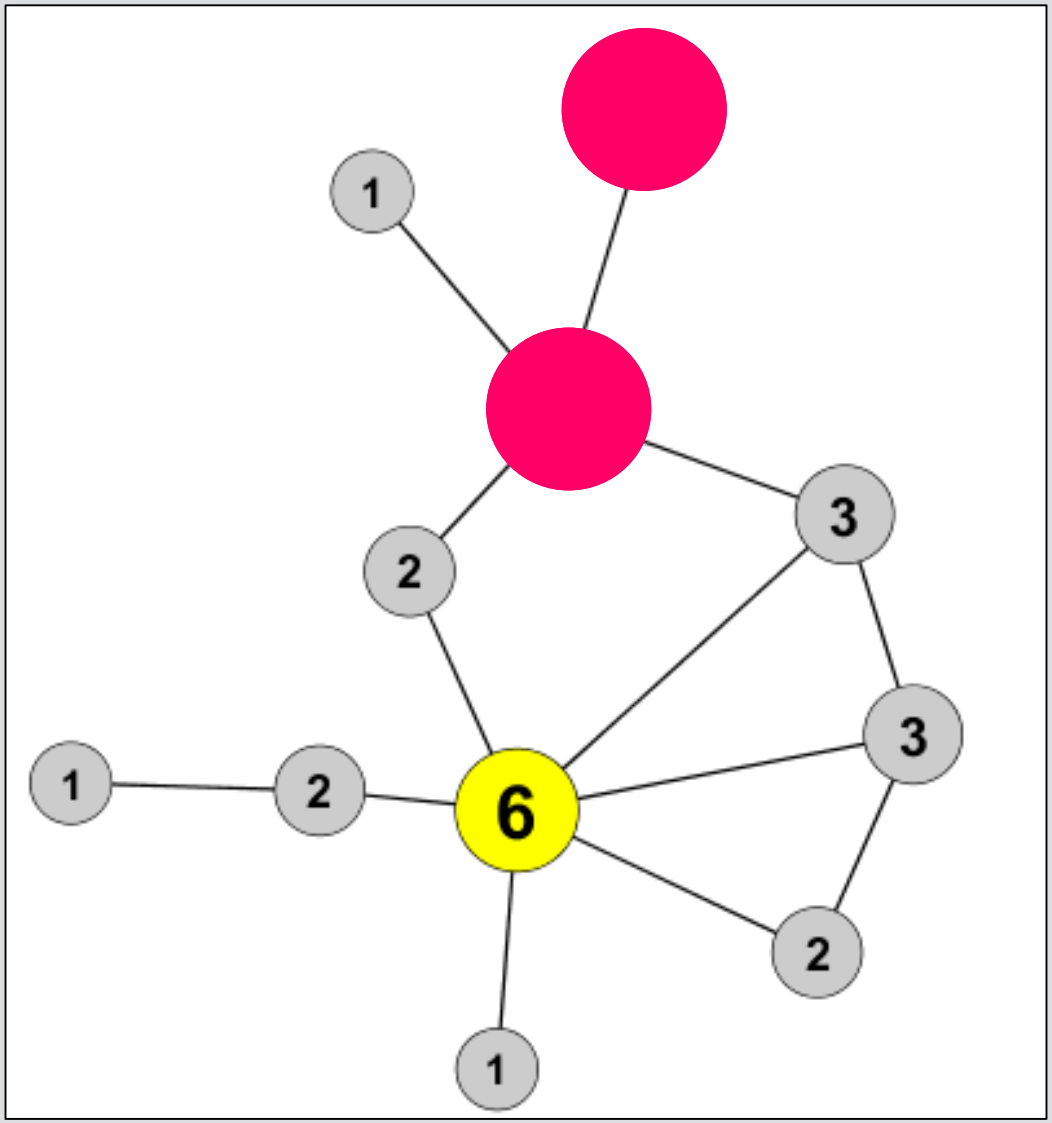


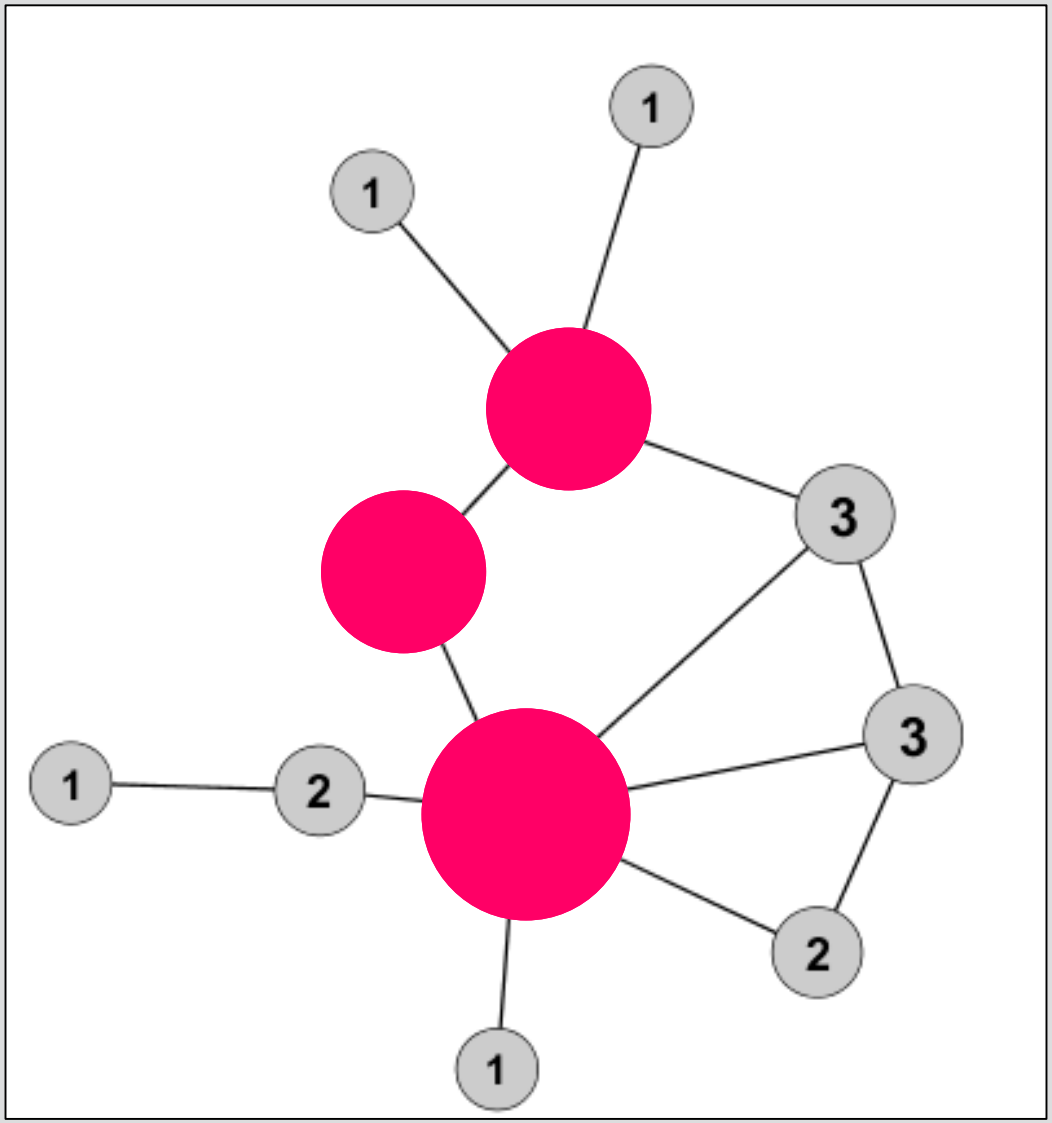
Activity 3: Networks

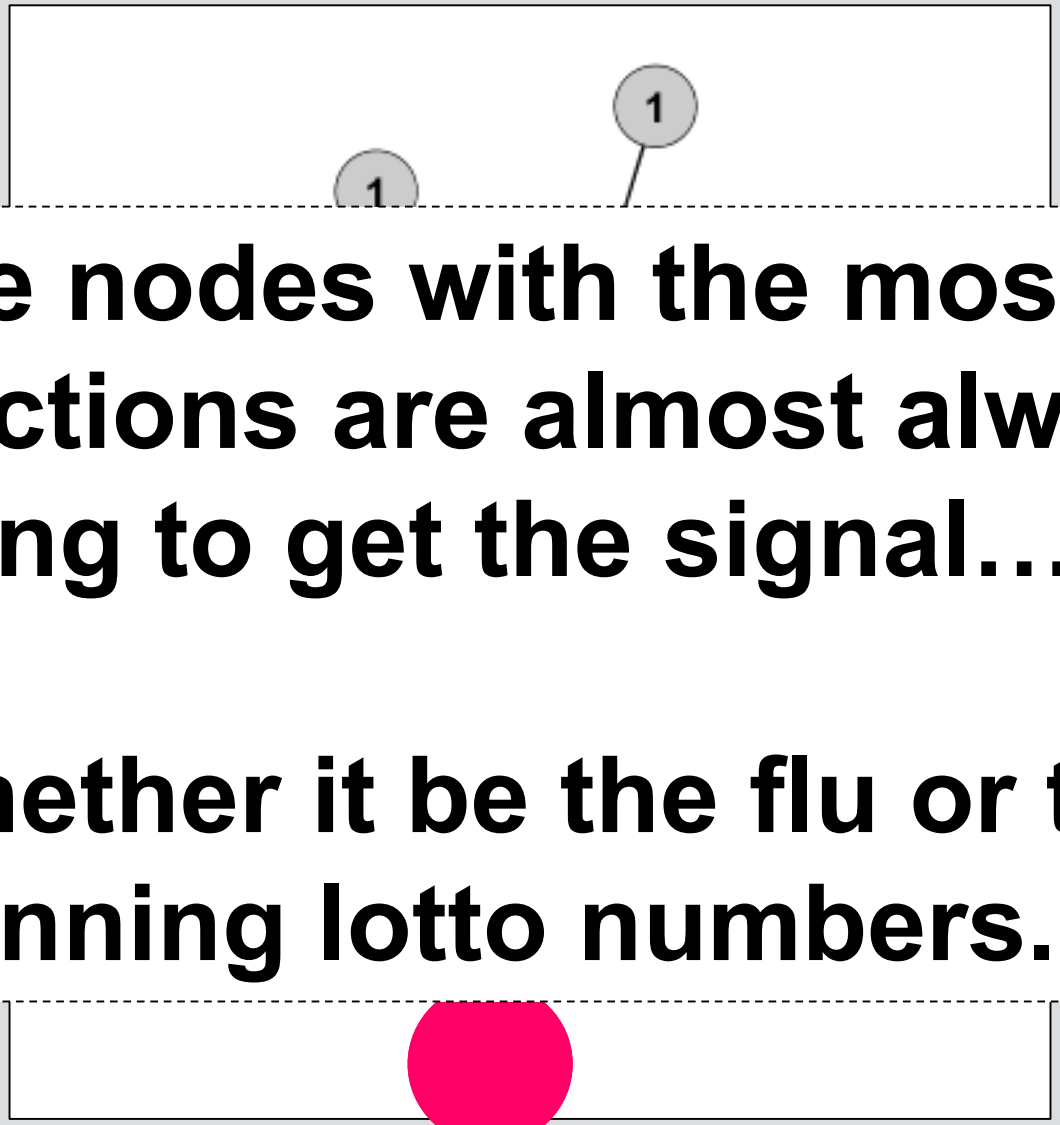

1
1
4
3
2
3
2
1


1
1
2
1
2
2
1


1
1
4
3
2
1
2
1


1
3
2
3
1
2
6
2
1


1
1
3
3
1
2
2
1



The nodes with the most interactions are almost always going to get the signal….

….whether it be the flu or the winning lotto numbers.

Networks!

- Human social networks work this way
- Cell networks work this way
- Many other networks act this way

- Flu pandemic planning
- Vaccination planning
- Drug targets for the cell
- National security planning

Conclusions

- DNA → RNA → Protein
- We *need* bioinformatics
  - Understanding cellular systems
  - Personalized medicine
  - Prevention vs. treatment
- Many skills gained
  - Biomedical research
  - Computer science
  - Mathematics
  - Team science
  - …. & many more!

A Career in Bioinformatics

- Skills needed
  - Programming (e.g., Perl, Python, Java, C++, PHP)
  - Database administration (e.g. MySQL, Oracle)
  - UNIX/Linux Operating System
  - Information Management
- Types of Jobs
  - Scientific curators, Software Developer, Network Engineering, Administrator/analyst, Bio-Statistics or any jobs where biologists are currently hired.
- Types of Employers
  - Pharmaceutical, Biotech and Software development companies
  - Academic Institutes and Hospitals
  - Research Institutes (JCVI, Broad Institute, JGI, Tgen,)

Contact us!

Kdempsey@unomaha.edu
@Science_Kate

Dkbastola@unomaha.edu