Week 2.2 Introduction to Genomics

Ted Laderas
2018-04-11

Reading Assignments

1. Why do genomics? https://discoveringthegenome.org/discovering-genome/what-genomics-why-do-genomics/health
2. What are proteins and what do they do? https://ghr.nlm.nih.gov/primer/howgeneswork/protein
3. How do genes direct the production of proteins? https://ghr.nlm.nih.gov/primer/howgeneswork/makingprotein
4. Interactive Presentation about Central Dogma and Genetic Medicine: http://www.hhmi.org/biointeractive/central-dogma-and-genetic-medicine
5. Intro to sequencing: https://discoveringthegenome.org/discovering-genome/dna-sequencing/sequencing-synthesis

Genomics

What is genomics?

As opposed to *genetics*, which is more concerned with single genes, the term *genomics* covers the entire genome

Last Time

- We talked about finding association of disease with SNPs
- This time, we're actually looking at understanding the effect and mechanism of mutations

What is the difference...

What is the difference between a mutation and a variant?

What is the difference...

What is the difference between a mutation and a variant?

Has to do with frequency - mutations are much more rare than variants

What is the difference...

What is the difference between a mutation and a variant?

Has to do with frequency - mutations are much more rare than variants

Our Motivation

Pamela's Story

Pamela's Story Pamela, an OHSU breast cancer patient, shares her story. Pamela, an OHSU breast cancer patient, shares her story.

Thriving with love, wisdom and gratitude

By Pamela Feidelson

My cancer journey started almost 10 years ago while at a routine check-up with my gynecologist. As she was doing the breast exam, she felt a lump. She instructed me to go immediately to the Breast Center at OHSU to have what would be my first mammogram at 38 years old. I'd say I was nervous but with not a huge amount of concern since it's not like breast cancer was prevalent in my family, nor did I know any woman my age with the disease. But my doctor said it was worth being checked out, and I was able to get an appointment right away. I went in for the mammogram and then an ultrasound, and it was after the ultrasound that the radiologist told me she was concerned and wanted to do a biopsy. It was at that moment that I became considerably emotional. I broke down and cried. I think I knew. Three agonizing days later, while sitting in my office at OHSU where I worked at the time, I got the phone call that changed my life with those three little words: You have cancer.

No stone unturned

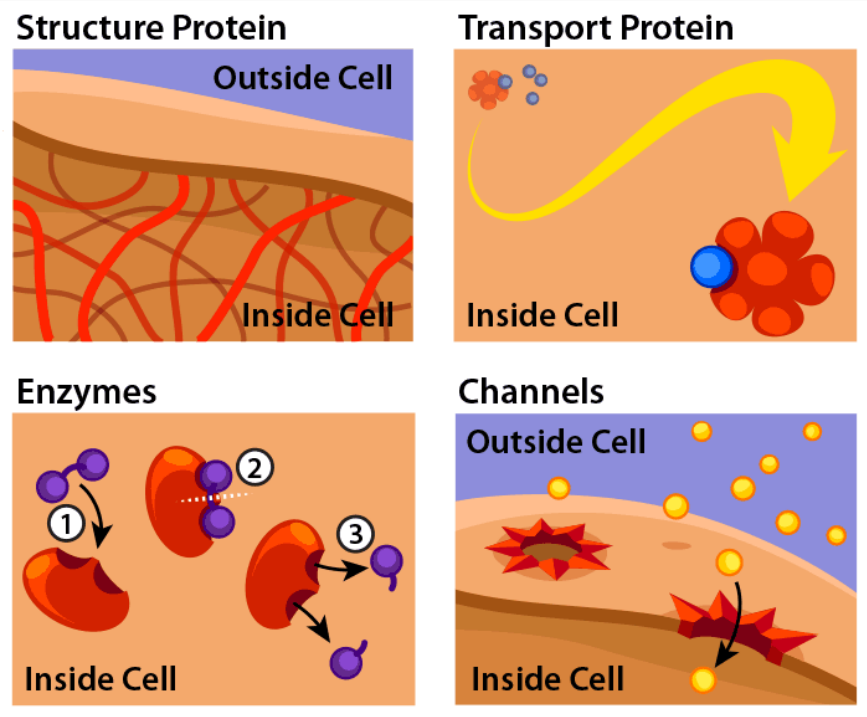
https://www.ohsu.edu/xd/health/services/cancer/patient-stories/pamelas-story.cfm

Learning Objectives

1. What are proteins and how are they made in the body?
2. What is a mutation? How can mutations disrupt protein functions?
3. What is a reference sequence and why is it important in sequencing?
4. What is Next Generation Sequencing and how do we use it to detect mutations?
5. Activity: Looking at a mutation in cancer
6. Biology of Cancer

Proteins are the machinery of the cell

- Proteins serve lots of different functions in the cell
  - Structure
  - Antibody (allergies)
  - Messenger (hormones)
  - Enzyme (assist with reactions)
  - Transport/Storage

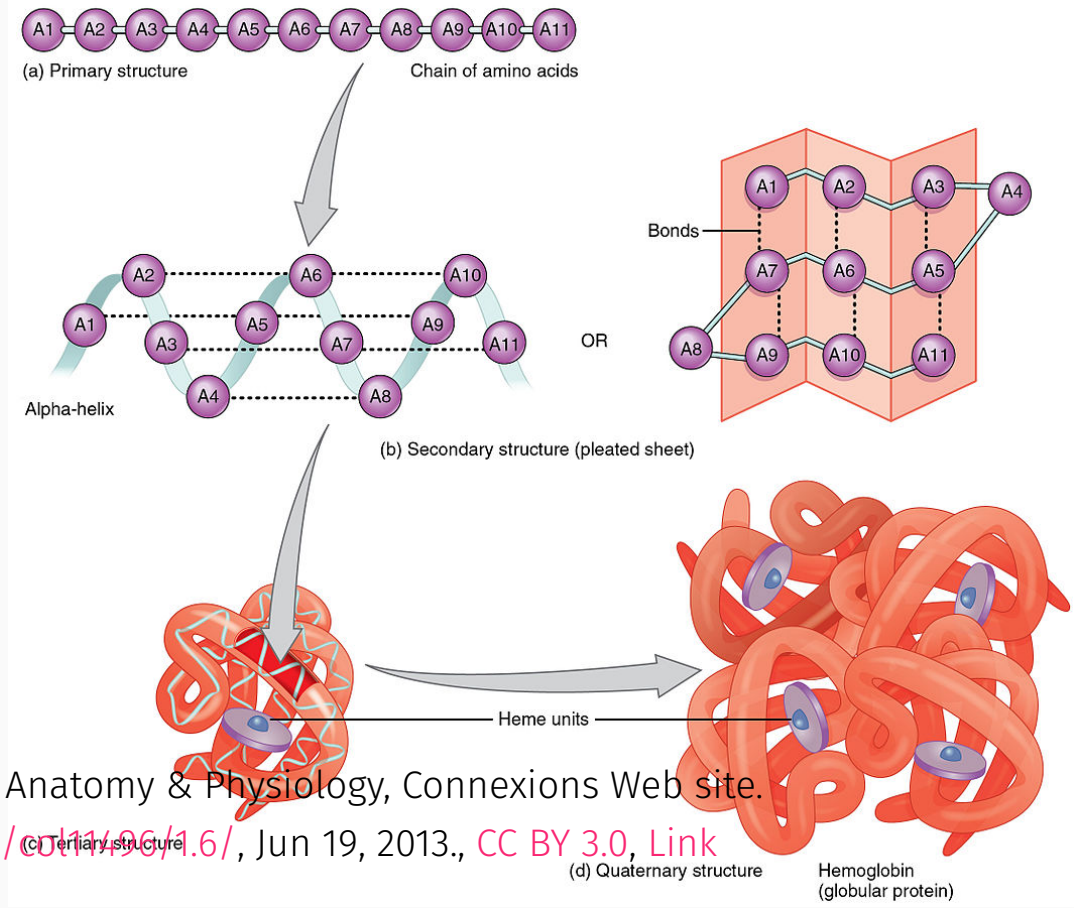


https://askabiologist.asu.edu/venom/what-are-proteins

8 / 31

How are proteins constructed?

- Sequences of 20 different amino acids make up parts of the protein
- amino acid chains "fold" to make up parts
- sequence of amino acids determines the structure



By OpenStax College - Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013., CC BY 3.0, Link

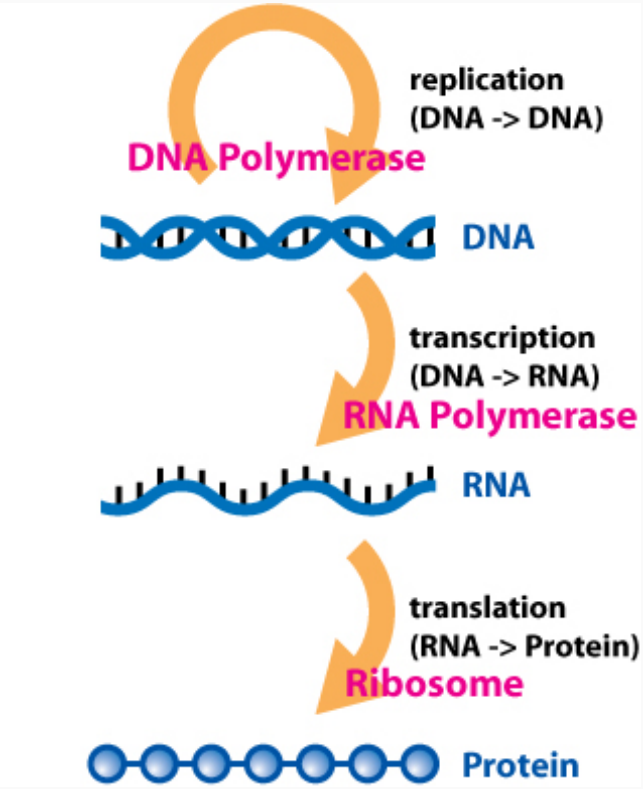
Proteins gone bad

- Specifying the wrong amino acid in a sequence can lead to changes in the protein structure
- Changes in protein structure can disrupt the function of proteins
- Can change phenotype!

We're going to be looking at BRAF, which is a protein that is altered in cancer.

How do we get from gene to protein?

Central Dogma:



https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

For right now, just think

Gene -> Complementary Strand -> Protein

3 bases = 1 codon

Codons: 3 nucleotide bases in the DNA = 1 amino acid in protein

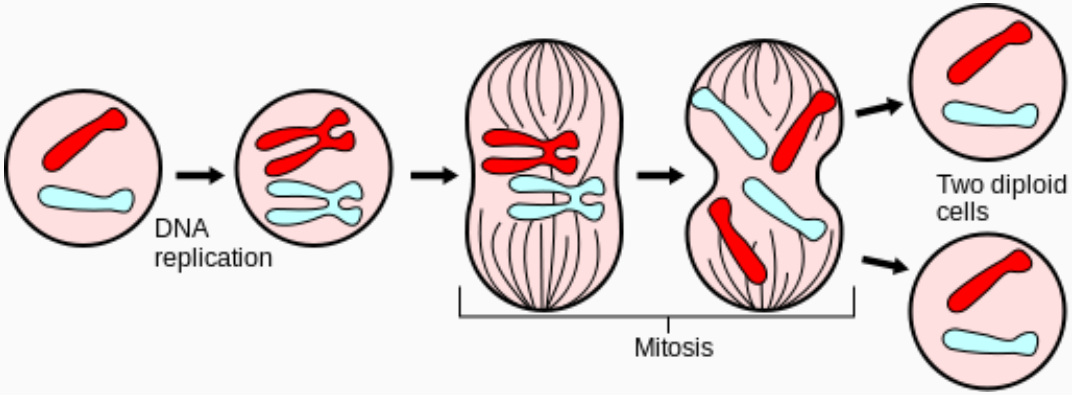
Example: GCT = Alanine

What AA is coded by CAC?

https://www.yourgenome.org/activities/kras-cancer-mutation

Our cells replicate

- During cell division (mitosis), chromosomes line up to be replicated in each cell
- Special enzymes (polymerases) produce copies of the cell DNA
- Each step has an opportunity to go wrong and introduce mistakes in the replication



https://commons.wikimedia.org/wiki/File:Major_events_in_mitosis.svg

Mistakes Happen

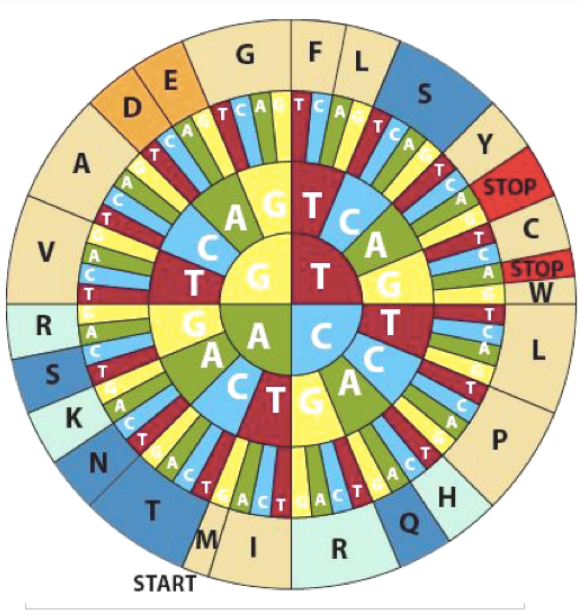
- Mistakes may happen in replication
- Introduce mutations in the sequence of replicated cells
- Enzymes can correct some mistakes, but not all

Mutations can alter protein function

- Mutations can change amino acid specified
- Genetic mutations can alter which amino acid is used in a protein sequence
- Alterations in Amino Acids

Not all Mutations are the same!

What do you notice about the coding wheel?



A Alanine
C Cysteine
D Aspartic Acid
E Glutamic Acid
F Phenylalanine
G Glycine
H Histidine
I Isoleucine
K Lysine
L Leucine
M Methionine
N Asparagine
P Proline
Q Glutamine
R Arginine
S Serine
T Threonine
V Valine
W Tryptophan
Y Tyrosine

- *synonymous mutations* (no change in AA)
- *non-synonymous mutations* (cause change in AA)

https://www.yourgenome.org/activities/kras-cancer-mutation

17 / 31

Reference Genomes are Important

We need reference genomes, which are derived from only a few individuals.

We need these to *align* our own sequencing samples to locations in the genome

This works because human genomes are 99.9% similar to each other.

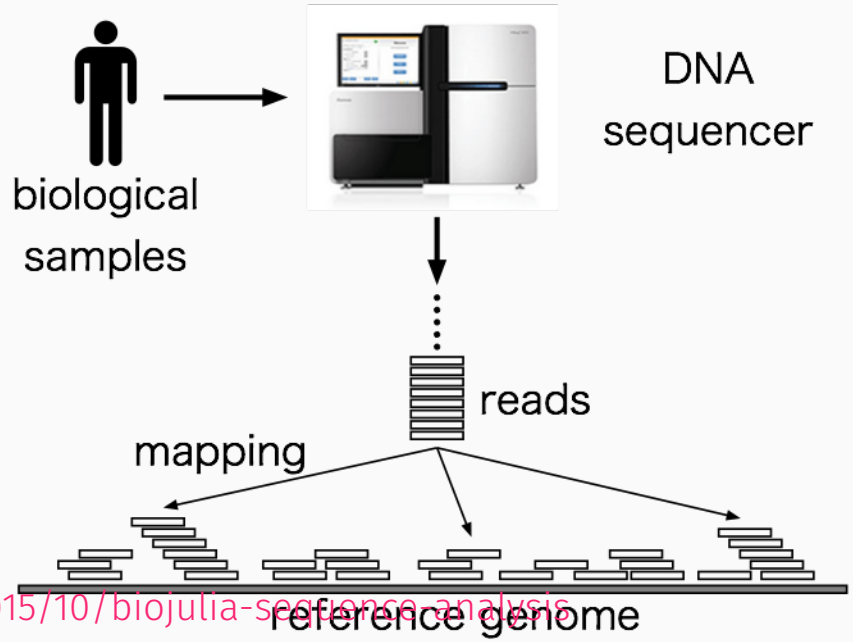
Lots of little pieces

- Extract DNA from tissue sample
- Fragment DNA into tiny pieces
- Sequence these tiny pieces all at once!
- *Sequence Read*: one of these tiny pieces

Aligners

Heavy duty programs that take all the sequence pieces and align them to genome

- Aligning one read may take time
  - need to search over 3 Billion places in the genome
- Aligners try to align many reads at once
- Still takes a long time!
- Using them requires understanding sequencing technology



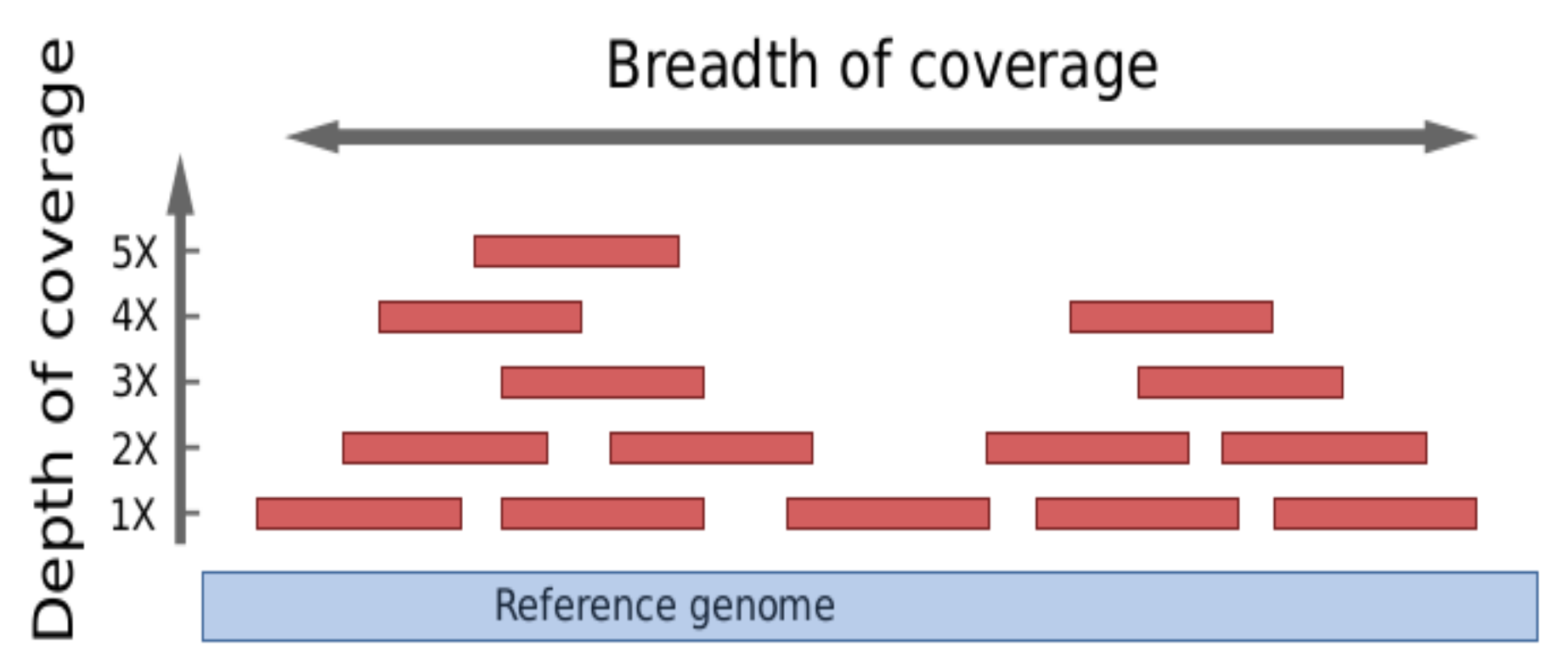
https://julialang.org/blog/2015/10/biojulia-sequence-analysis

20 / 31

What evidence is there?

Our results can be summarized with two terms:

- *Coverage* - How much of the genome do our reads align to?
- *Depth* - How many reads are there for a particular position at the genome



http://www.metagenomics.wiki/pdf/definition/coverage-read-depth

21 / 31

Errors in sequencing

- Sequence calling process has errors
- Can impact alignment and evidence
  - Good aligners incorporate thinking about types of errors
- Solution: incorporate *quality* scores for each position
- Look for errors in the sequence data

Matched Normal/Tumor samples

In cancer, we try to compare *tumor* tissue with *normal* tissue when we sequence. Why?

Matched Normal/Tumor samples

In cancer, we try to compare *tumor* tissue with *normal* tissue when we sequence. Why?

We do this to separate genomic alterations into two types:

- *germline* - alterations that are inherited by birth
- *somatic* - alterations that are acquired after birth

In-class assignment: Genomics

Today we're going to look at a particular oncogene that is implicated in cancer: BRAF

Download the Codon Wheel and the Sample Sheet in `2.2/in-class-activity/` in D2L

Put your results here:
https://docs.google.com/spreadsheets/d/1QqQ4x2kBlkxA8kt9ookbXVq6otoeuqNM6CaQse22Wj8/edit?usp=sharing

https://www.yourgenome.org/sites/default/files/downloads/activities/braf-from-gene-to-cancer-therapy/

Translating

1. Find the consensus sequence first
2. Translate that into the complementary strand
    1. Remember, A pairs with T and C pairs with G
3. Working backwards, translate the complementary sequence into Amino Acids

```
consensus sequence
ATT TCT TAC
TAA AGA ATG
complementary sequence
I   R   V    <----
amino acid sequence (work backwards)
```

Some things to think about

What does an error look like, and what does a mutation look like?

When you translate to amino acids, read in the 5' to the 3' direction

Discussion: What are your findings?

Point mutations are only one kind

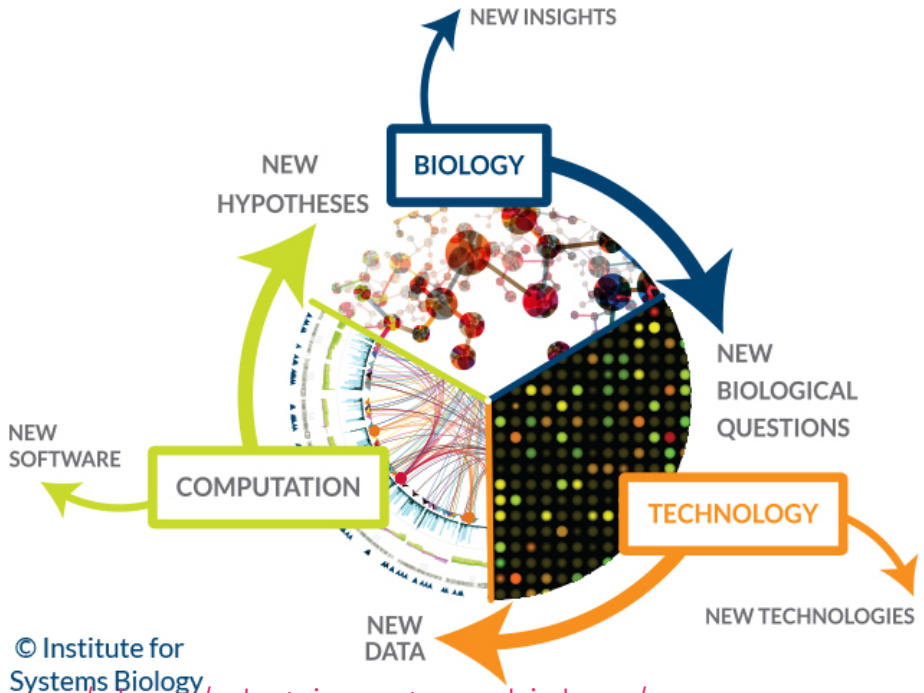
- Lots of other genomic alterations are associated with cancer
- Genomic Rearrangement
- Immune system
- Epigenome

"omics"

- quantitative measurements on cells and people
  - multiple things are measured at once ("high-throughput")
    - proteomics - measure levels of proteins
    - transcriptomics - measure levels of RNA
    - metabolomics - measure abundance of bacteria
    - immunomics - measure immune system cells

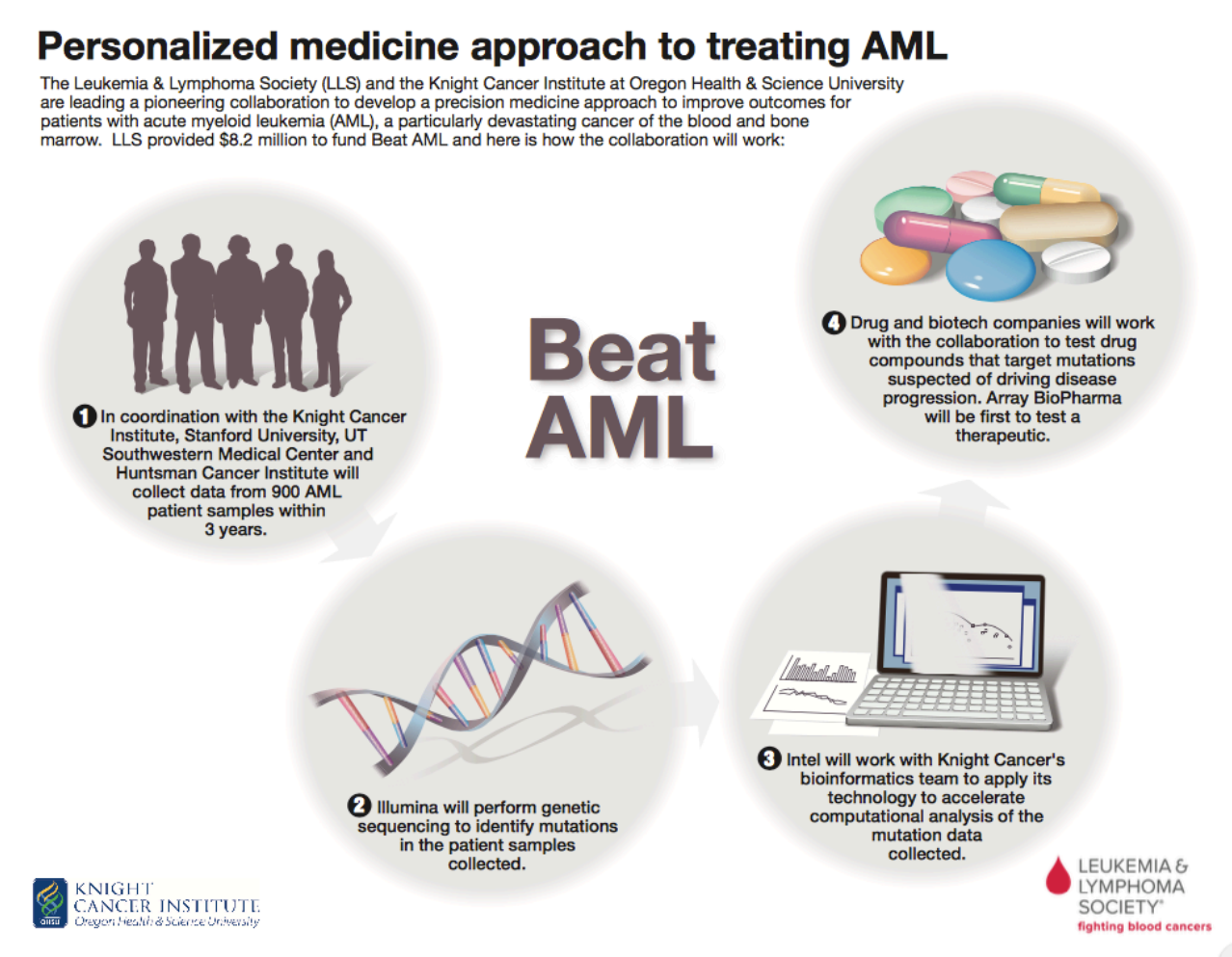
Need the bigger picture!

- Systems Biology of Cancer
- Understand how cell function is dysregulated
- Requires integrating across omics measurements
- Requires previous knowledge of low-throughput biology



https://www.systemsbiology.org/about/what-is-systems-biology/

BeatAML



http://www.ohsu.edu/xd/health/services/cancer/about-us/druker/beat-acute-myeloid-leukemia.cfm