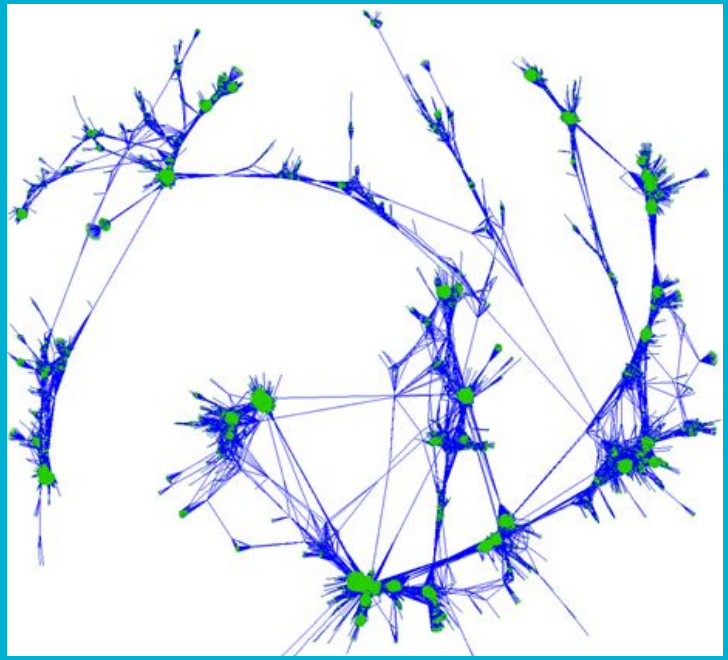
Crash Course in Genomics



Princeton SPLASH 2023

Sara Geraghty, Princeton Graduate Student

Introductions!

What's your name?

What year are you?

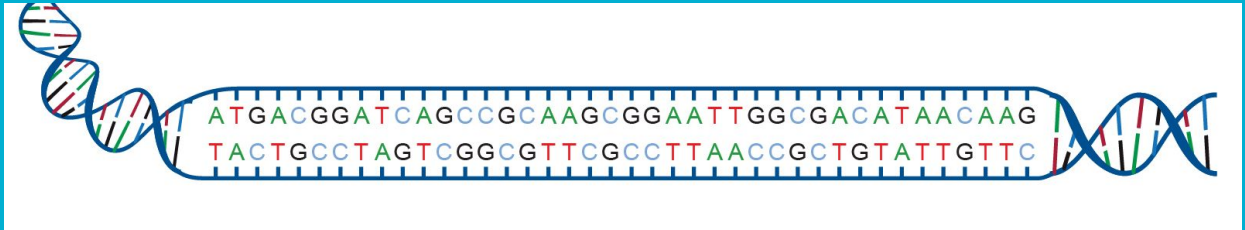
What interests you about genomics?

What's one weird thing that happened to you this week?

What exactly is genomics?

As defined by the NIH:

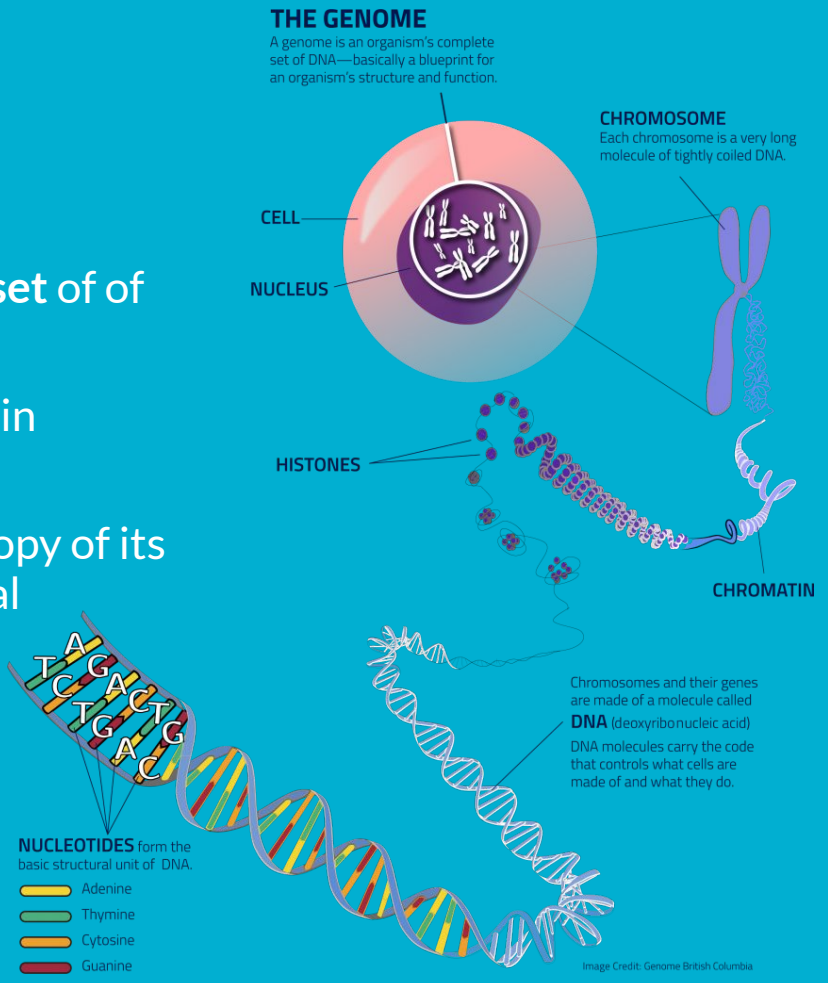
*"Genomics is the study of all of a person's genes (**the genome**), including interactions of those genes with each other and with the person's environment."*



From the National Human Genome Institute

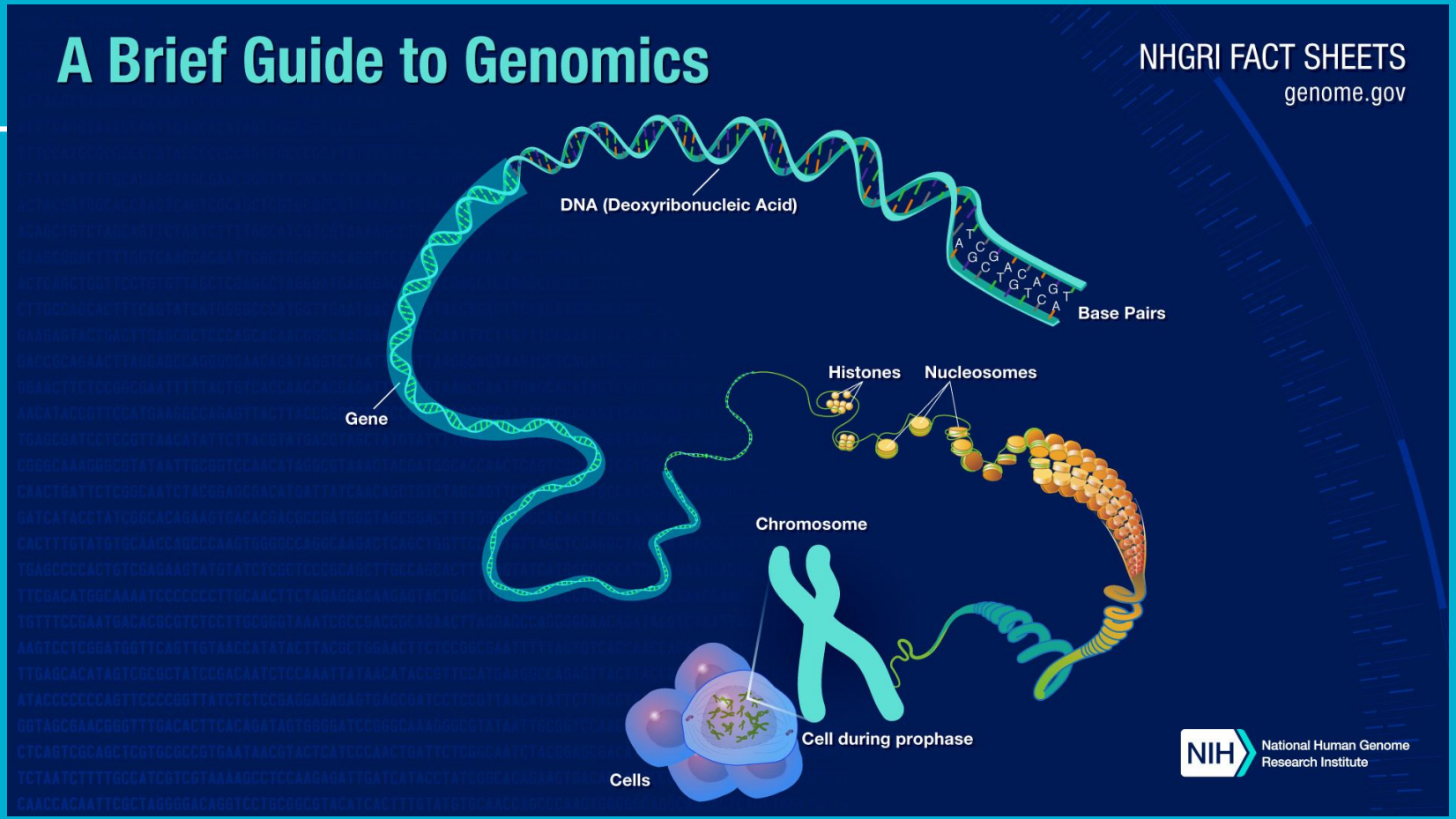
So, what's a genome?

- A genome is an organism's **complete set** of of DNA, which contains all of its genes
- Consists of ~3 billion DNA base pairs in humans (the 4-letter "code" of DNA)
- Every cell in an organism contains a copy of its full genome, like an instruction manual



A Brief Guide to Genomics

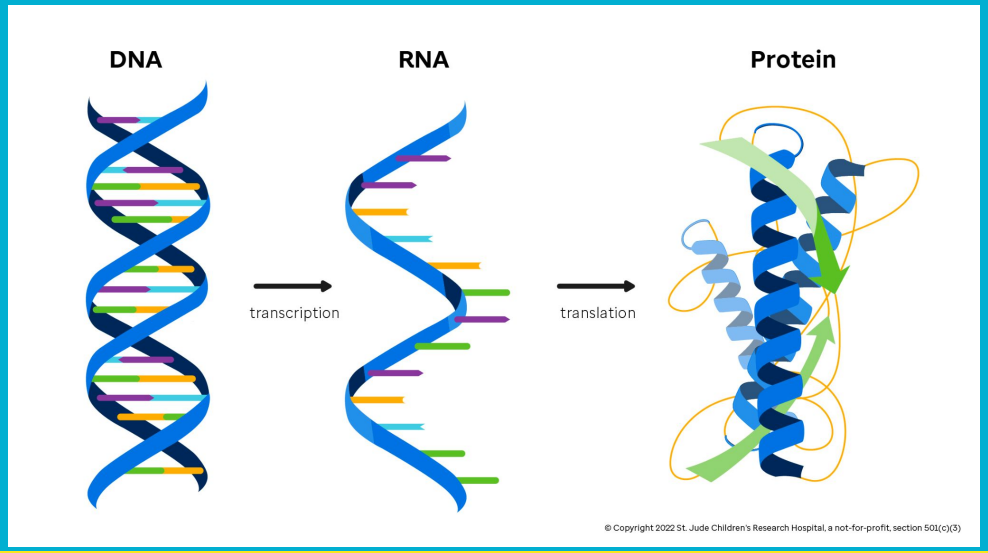
NHGRI FACT SHEETS
genome.gov



NIH National Human Genome Research Institute

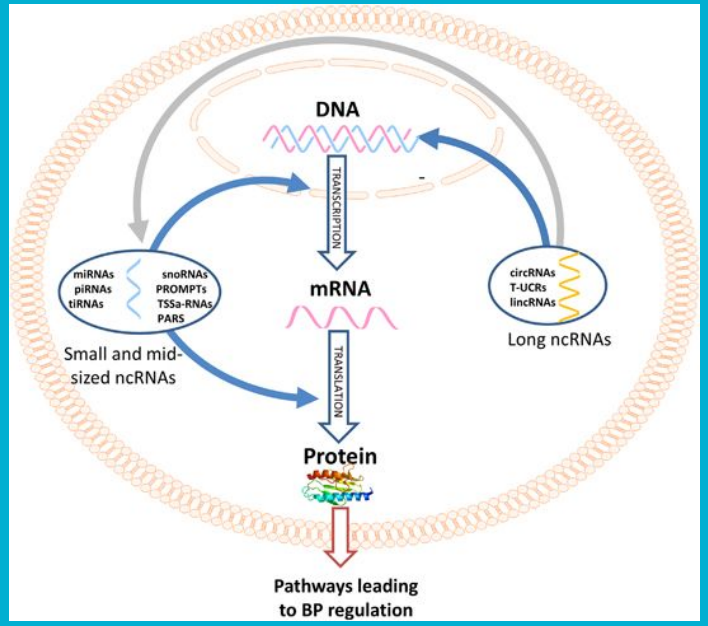
The central dogma of biology

- If the genome is an organism's instruction manual, the central dogma explains how that manual is read (DNA > messenger RNA > Protein)
- Proteins are the workhorses of the cell!



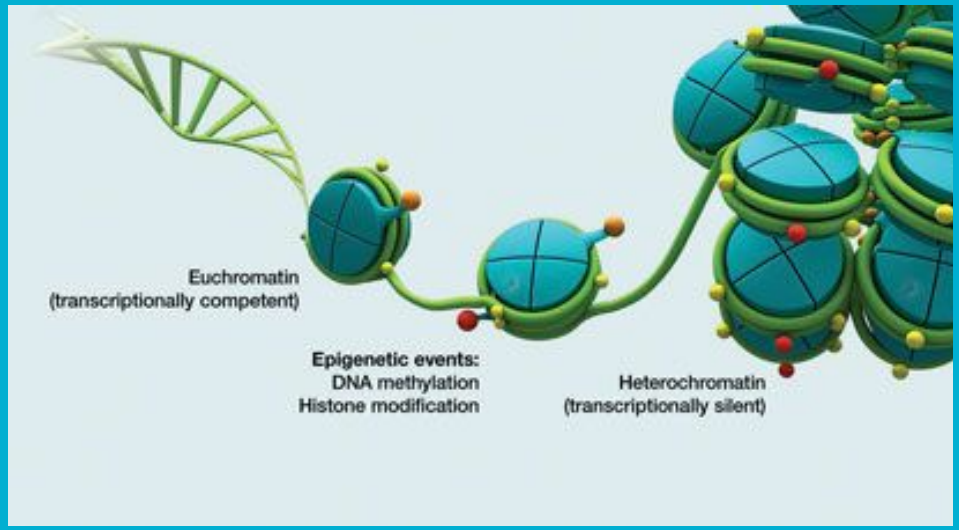
But, there's so much more!

Non-coding RNAs (ncRNAs) don't make protein, but they play a regulatory role in the cell



Link to source

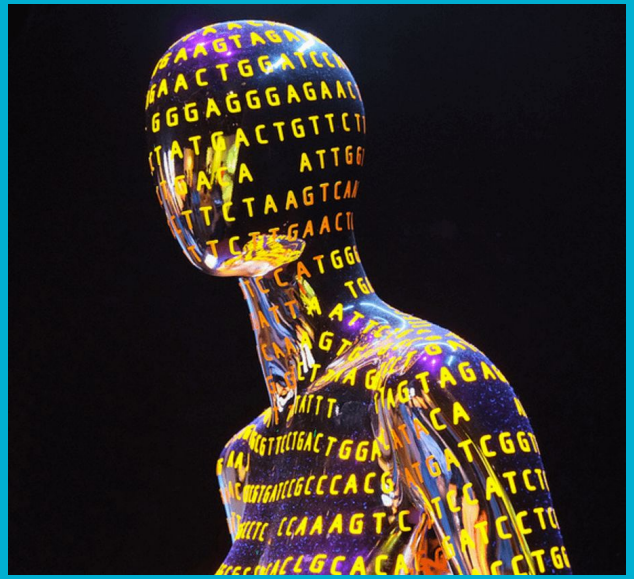
Chromatin folding and accessibility can change which genes can be transcribed



Link to source

The Human Genome Project

- Large-scale international project led at the National Institutes of Health (NIH) by the National Human Genome Research Institute
- Produced the first high-quality version of the human genome sequence (finished in 2003), now freely available in public databases
- The sequence is not that of one person, but is meant to be a "representative" or generic sequence that was generated from many individuals



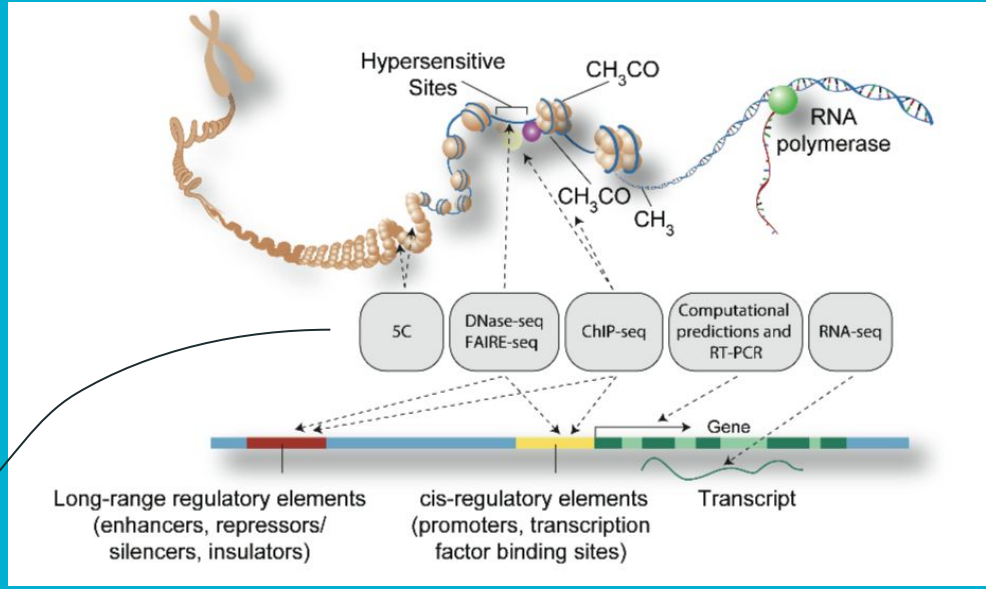
Link to source

So… now what?

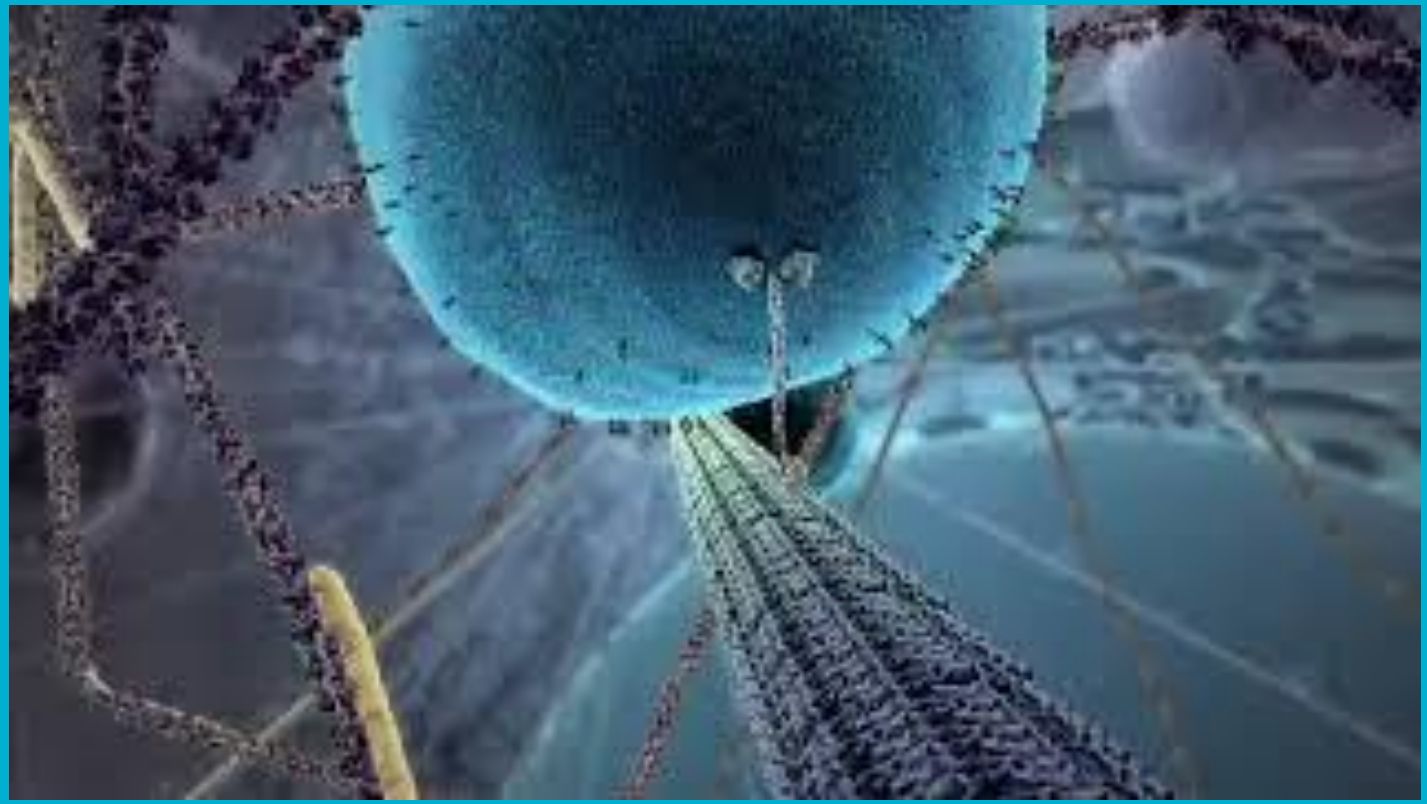

PHASE TWO : INTERPRETATION
I THINK I FOUND A CORNER PIECE.
3 BILLION PIECES
GENOME

Genome annotation and interpretation: at the heart of genomics

- Can think of a genome as a series of **functional elements** (like genes!)
  - What functions does each of these elements have in the cell? In what types of cells is it expressed? Is it related to disease?
- We use other types of sequencing techniques to figure out what these functional elements are are



The complexities of the cell



https://www.youtube.com/watch?v=Nnpl4mE-pX0

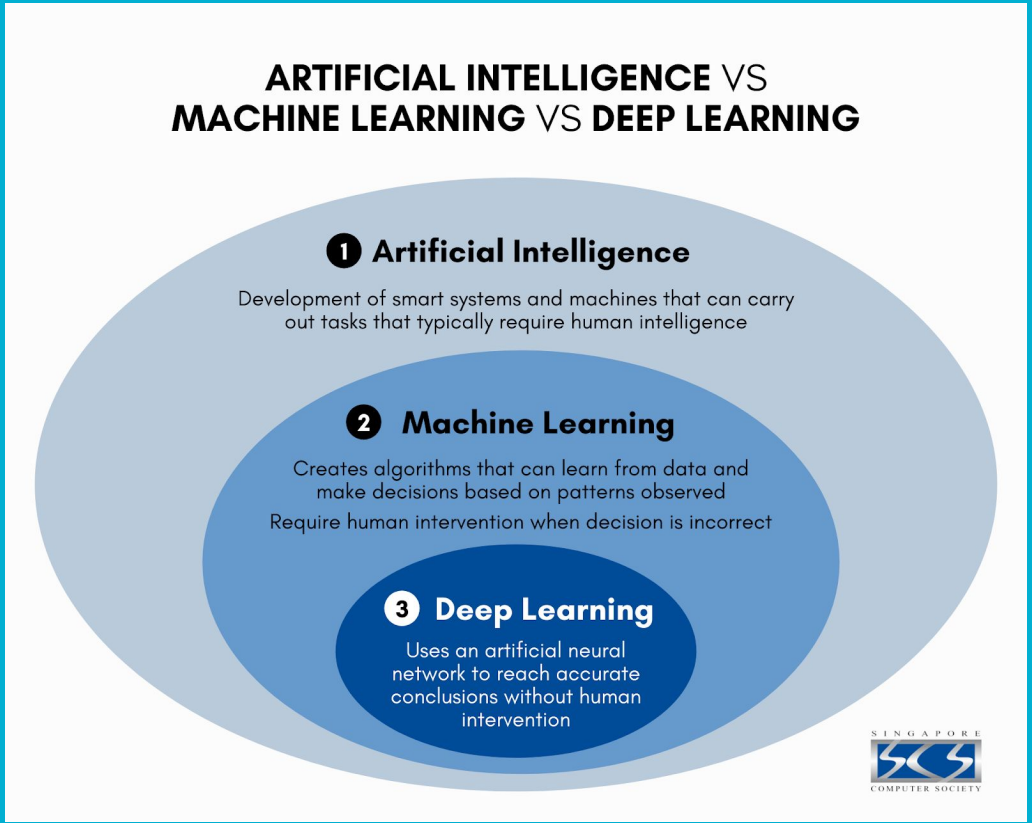
Genomics = Data

In genomics, we combine our genomic information with numerical readouts of what is going on in cells (like gene or protein expression).

Our goal is to use computer science and statistics to understand what precise changes in the genome *mean* for the cell.

How do we tackle data of this size?

If you can define inputs and outputs, and you have enough data, you can reformulate a biological problem as a computational problem!



[Link to source]

Natural language processing (NLP): For ChatGPT and for Biology

I fixed my dog a sandwich

A T C T T A G T

CHAPTER I

Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so VERY remarkable in that; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually TOOK A WATCH OUT OF ITS WAISTCOAT- POCKET, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again.

The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.

Either the well was very deep, or she fell very slowly, for she had plenty of time as she went down to look about her and to wonder what was going to happen next. First, she tried to look down and make out what she was coming to, but it was too dark to see anything; then she looked at the sides of the well, and noticed that they were filled with cupboards and book-shelves; here and there she saw maps and pictures hung upon pegs. She took down a jar from one of the shelves as she passed; it was labelled 'ORANGE MARMALADE', but to her great disappointment it was empty: she did not like to drop the jar for fear of killing somebody, so managed to put it into one of the cupboards as she fell past it.

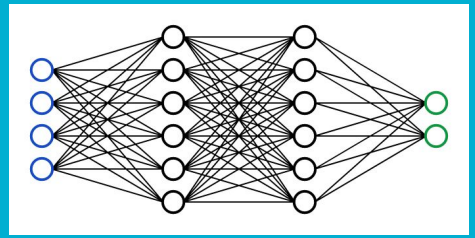
'Well!' thought Alice to herself, 'after such a fall as this, I shall think nothing of tumbling down stairs! How brave they'll all think me at home! Why, I wouldn't say anything about it, even if I fell off the top of the house!' (Which was very likely true.)

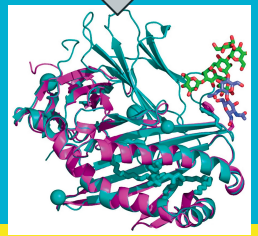
Down, down, down. Would the fall NEVER come to an end! 'I wonder how many miles I've fallen by this time?' she said aloud. 'I must be getting somewhere near the centre of the earth. Let me see: that would be four thousand miles down, I think--' (for, you see, Alice had learnt several things of this sort in her lessons in the schoolroom, and though this was not a VERY good opportunity for showing off her knowledge, as there was no one to listen to her, still it was good practice to say it over) '--yes, that's about the right distance--but then I wonder what Latitude or Longitude I've got to?' (Alice had no idea what Latitude was, or Longitude either, but thought they were nice grand words to say.)

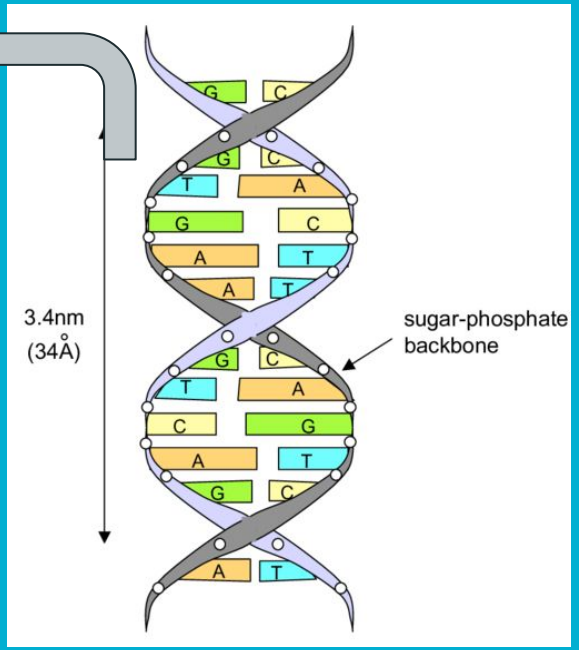
Presently she began again. 'I wonder if I shall fall right THROUGH the earth! How funny it'll seem to come out among the people that walk with their heads downward! The Antipathies, I think--' (she was rather glad there WAS no one listening, this time, as it didn't sound at all the right word) '--but I shall have to ask them what the name of the country is, you know. Please, Ma'am, is this New Zealand or Australia?' (and she tried to curtsey as she spoke--fancy CURTSEYING as you're falling through the air! Do you think you could manage it?) 'And what an ignorant little girl she'll think me for asking! No, it'll never do to ask: perhaps I shall see it written up somewhere.'

[Link to source](#)

Neural network algorithms

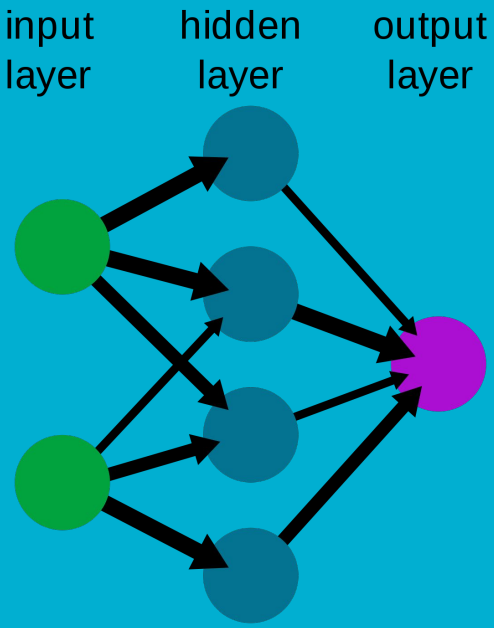






... can also use biology to inform computer science!

A simple neural network



Inspired by neurons in the brain!

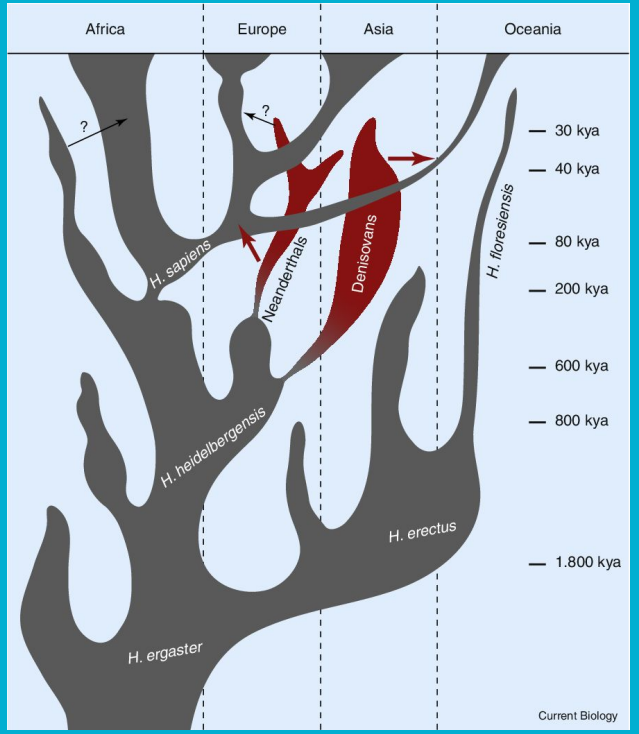
Link to source

5 minute break!

Applications of Genomics

Paleogenomics

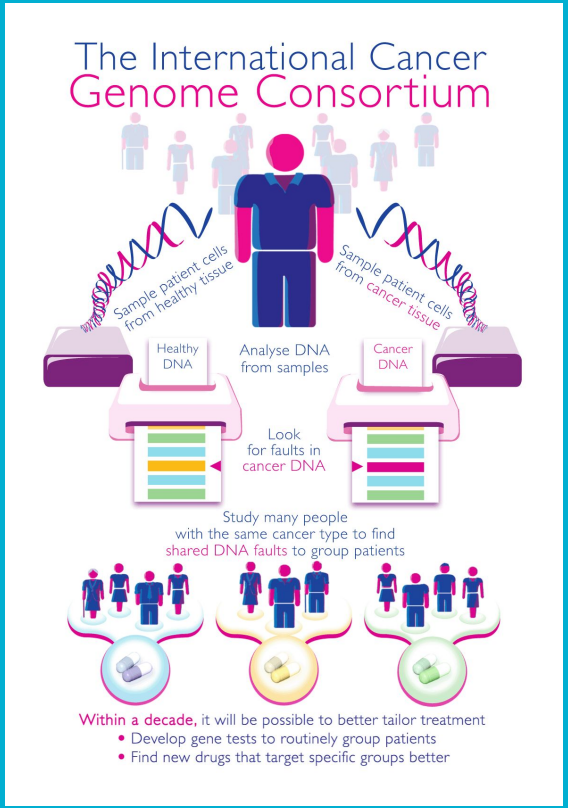
- We can extract **ancient DNA** (aDNA) samples from fossils, and use them to reconstruct whole genomes of extinct species
- Using these genomes, we can create species trees, which can help us understand:
  - How modern species are related to each other
  - How complex traits, like opposable thumbs or color vision, evolved
  - What distinguishes us as humans from other hominids?
- Projects to “bring back” extinct species, like the woolly mammoth



[Link to source]

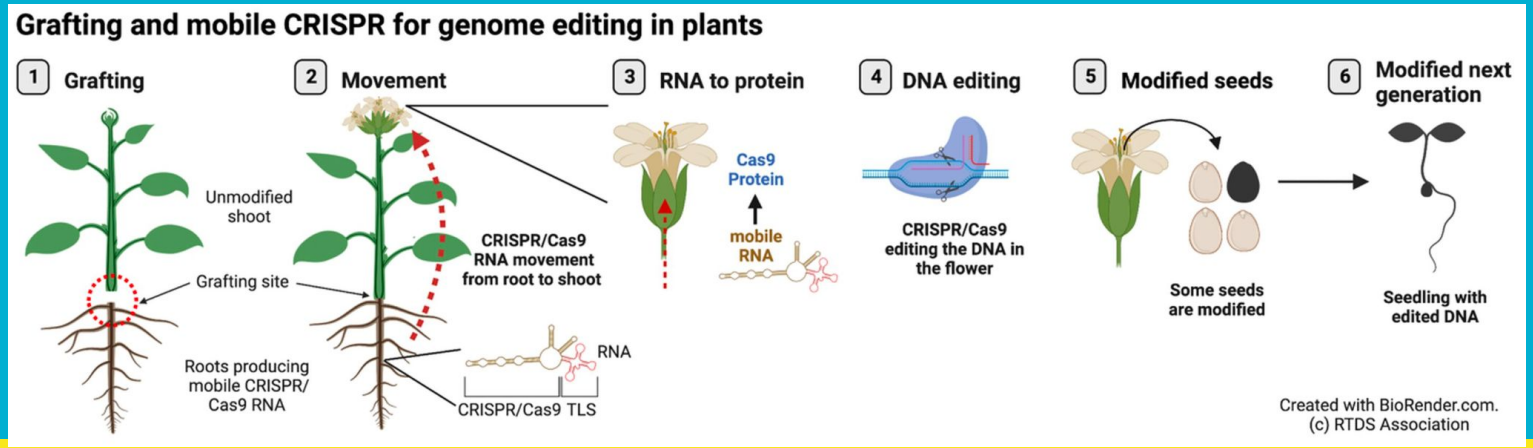
Cancer genomics (my specialty!)

- Our cells develop **mutations** (changes in the sequence of our DNA), along with other changes
  - This happens naturally, and in response to carcinogens, like UV radiation and cigarette smoke
  - Some of these mutations lead to cancer development
- Over the last few decades, we've collected **thousands of healthy and tumor genomes**, along with information about gene expression, chromatin structure, etc.
- Can we use this information to figure out which mutations cause cancer? Then, can we figure out *how* they cause cancer, in order to treat them?
- The goal of cancer genomics = p**ersonalized medicine**
  - Using a patient's genome to understand their tumor and treat them accordingly



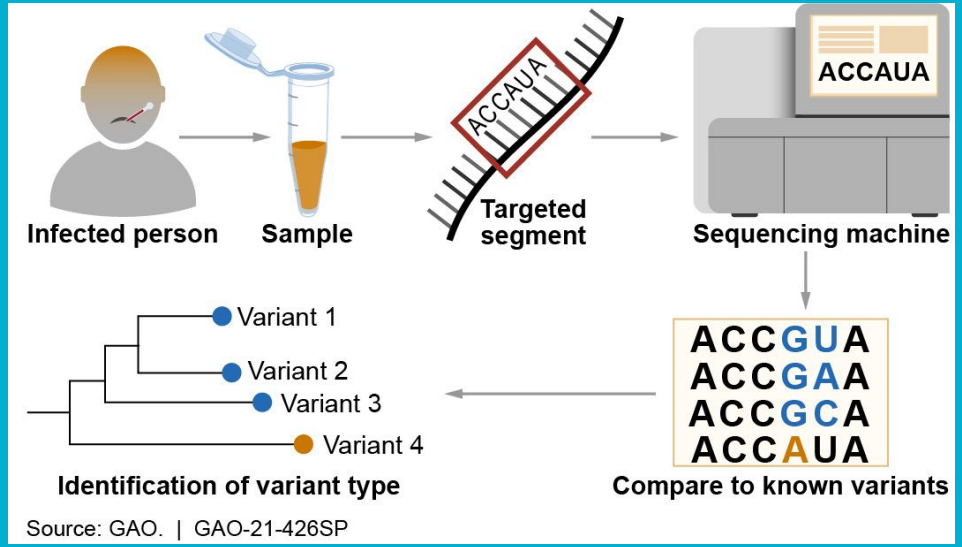
Plant genomics and agricultural biotechnology

- Big movement to sequence and annotate the genomes of important plants, particularly plants we eat!
- Once we understand their genomes, we can edit them using tools like CRISPR
  - Drought resistance, pest tolerance, more nutritional value, better flavor, etc.
- Helping crops and ecosystems be more resilient to the effects of climate change



Infectious disease/ pathogen genomics

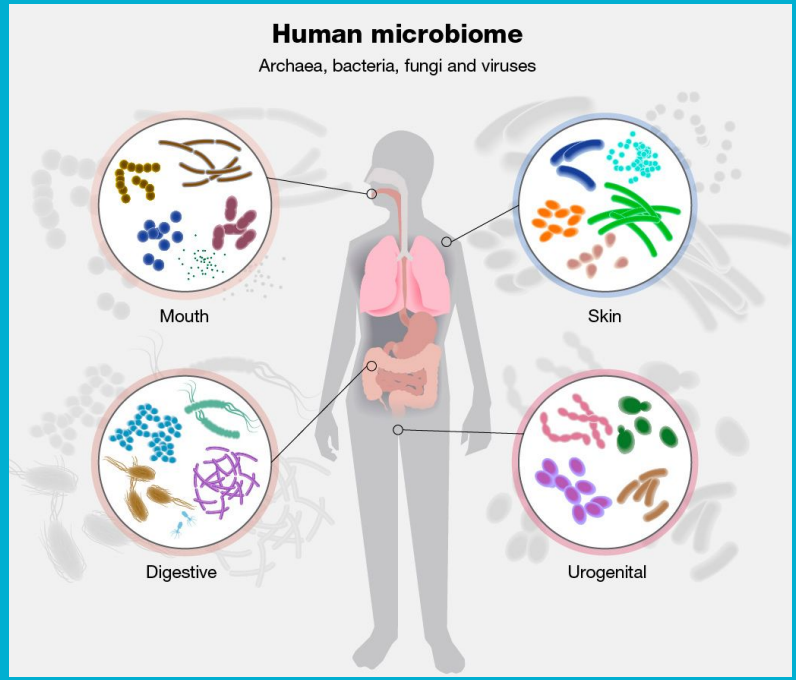
- Sequencing the genomes of bacteria, fungi, and viruses that make us sick
- We can use pathogen genomics to:
  - Understand the molecular mechanism of how bugs make us sick, to inform development of drugs, treatments, and vaccines
  - Track how these microorganisms change over time in response to environmental pressures (e.g. ‘variants’)
  - Diagnosing infections, in order to treat and to track transmission and outbreaks



Link to source

Genomics for understanding microbiomes

- **Microbiome** = the community of microorganisms (such as fungi, bacteria and viruses) that exists in a particular environment
- Sequencing the genomes of the microorganisms that live in and on us can help us understand what species live there and how they interact with our cells
  - We are still uncovering all the many ways our microbiome impacts our health!
- We can also sequence microbiomes from other environments (like soil, or tree roots) to better understand how human activity affects the diversity of microbes in these environments



Link to source

Some other fun examples...

- Making better cheeses, yogurts, breads, beers, wines...
  - All of these products involve the use of microorganisms; we can use genomics to understand how these microorganisms work, and modify them to make them more effective
  - Fun fact: the science behind CRISPR was invented by a team searching for a better way to make yogurt!
- DNA sequencing is often used in forensics, to match the DNA of site samples to suspects
  - Not much AI involved here, but still genomics!
- Companies like 23andMe and Ancestry.com use genomics to uncover the genetics of human traits, like whether you like chocolate or sneeze multiple times in a row, and where your family came from

23andMe®

What did we learn?

1. Genomics is the study of **genomes**, which is an organism's complete set of DNA
2. The field relies on the **central dogma of biology,** which states that the genome (the "blueprint") is transcribed into mRNA, which is translated into protein (which makes things happen!)
3. The Human Genome Project sequenced the human genome, but we are still working on annotating and interpreting that genome (defining "**functional elements**" and their role in the cell and the body)
4. We can combine genomic information with other information in the cell, like gene expression, but that data gets very large and complex very quickly!
5. Computation, like artificial intelligence (AI), can help us make sense and draw meaning from this data
    a. Likewise, the more we learn about biology, the more we can use that biology to design new computer science techniques
6. There's so many applications of genomics that spans all kinds of fields!

Time for an activity!

Choose your favorite topic to the right. We're going to dig a little deeper…

- Paleogenomics
- Cancer genomics
- Plant genomics and agricultural biotechnology
- Pathogen genomics
- Genomics for understanding microbiomes

Thank you! (Any questions?)

My email: scamilli@princeton.edu