Gene Sequencing

A bit of History

- 1866 - Gregor Mendel (1822-1884) published a paper on his “Laws of Inheritance” based on experiments involving the cultivation of 29,000 pea plants.
- 1869 - DNA was first isolated by the Swiss physician Friedrich Miescher.
- 1927 - Nikolai Koltsov proposed that inherited traits would be inherited via a "giant hereditary molecule"

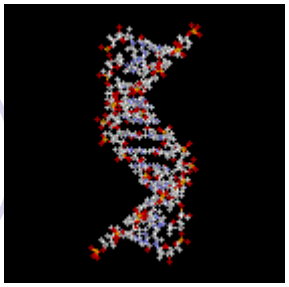
…history

- 1952 - a series of experiments conducted by Alfred Hershey and Martha Chase helped to confirm that DNA was the genetic material.
- 1953 - Structure of DNA discovered by Francis Crick & James Watson
  - They received the Nobel Prize in 1962 together with Maurice Wilkins.

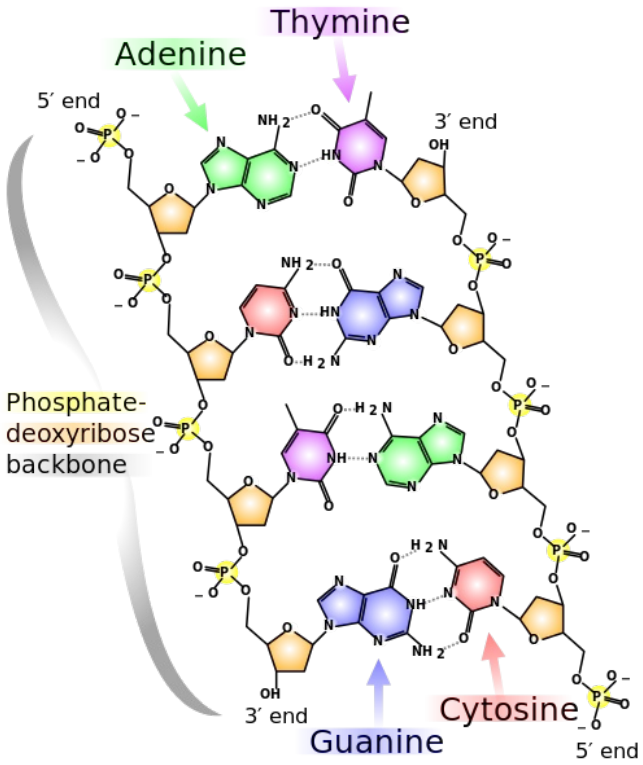
Gene Terminology

- DNA
- Genes
- Chromosomes
- Genomes

DNA Molecule



- Deoxyribonucleic acid
- A chain molecule made from a sugar backbone with sequences of the four bases :
  - Adenine
  - Thymine
  - Guanine
  - Cytosine
- Sequence of bases provides the blueprint for living organisms

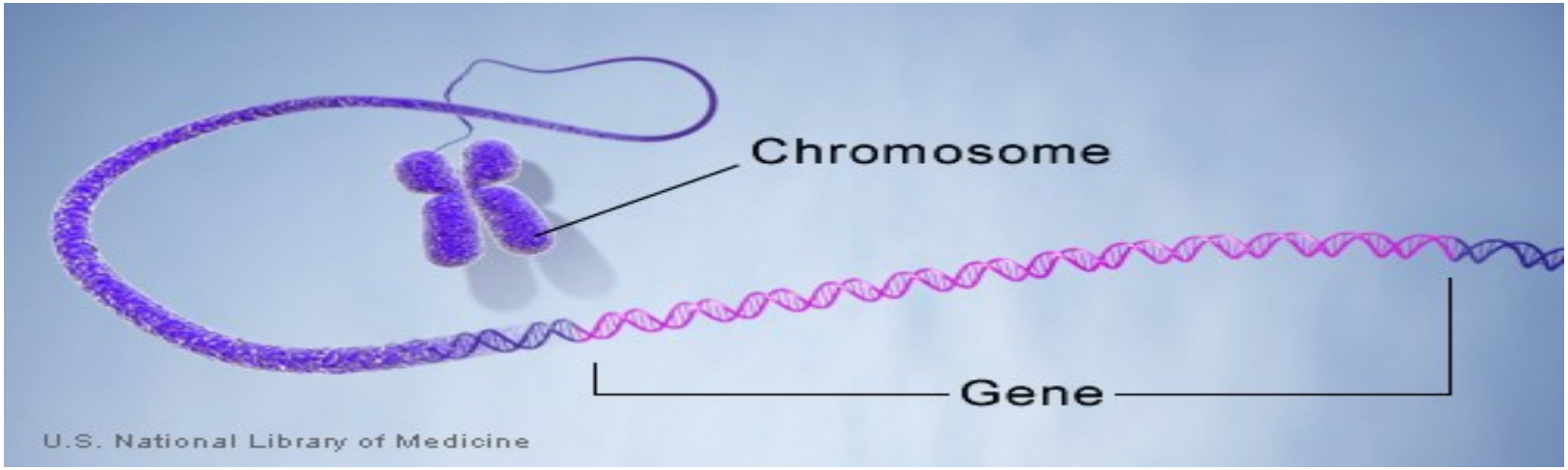


Some (big) DNA numbers…

- DNA is found in every human cell except red blood cells.
- Humans have roughly 100,000,000,000,000 (100 trillion cells).
- Each cell contains roughly 2 metres of DNA.
- There are approximately 3 billion (3,000,000,000) chemical letters (known as bases) in the DNA code in every cell in your body
- The DNA is tightly coiled up and structured into 46 chromosomes which are stored in the cell nucleus.
- Half life of DNA is around 520 years - DNA has been recovered from frozen 40,000 year old Woolly Mammoths

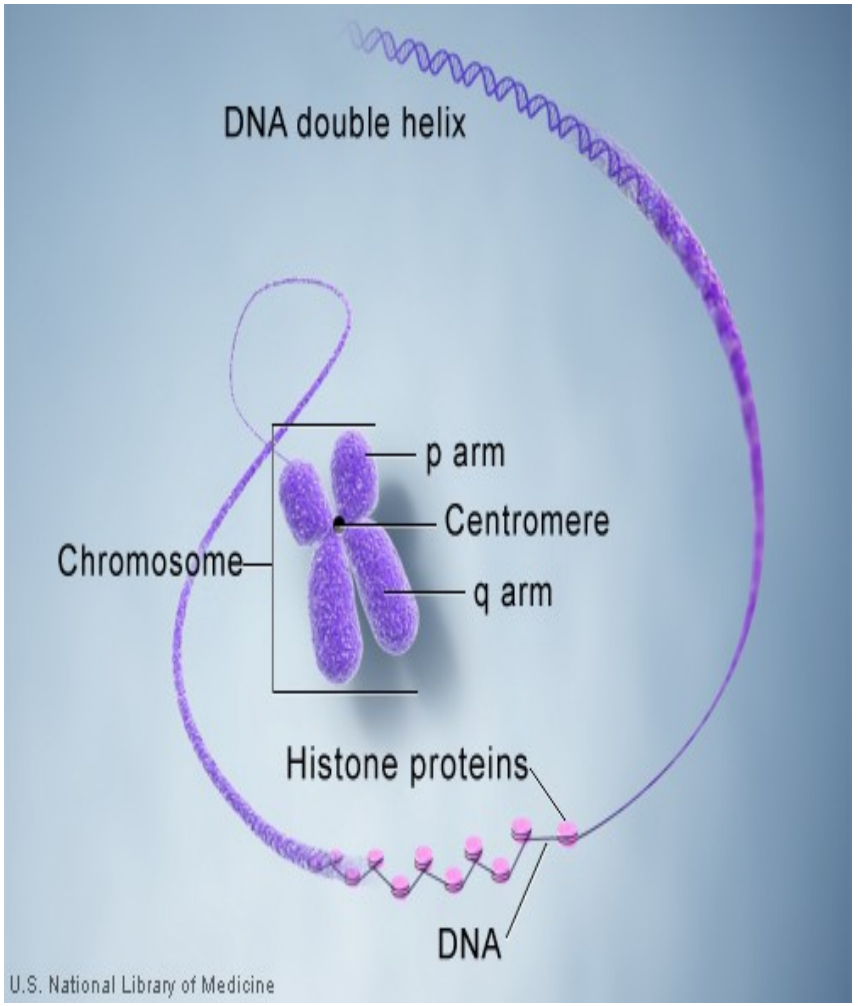
Genes

- A gene is the basic unit of heredity. Genes are pieces of DNA, which act as instructions to make and regulate the production of proteins.
- In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. Humans have between 20,000 and 25,000 genes.
- Every person has two copies of each gene, one inherited from each parent.
- Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) differ. These small differences contribute to each person's unique physical features.

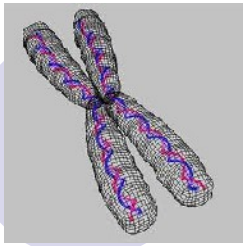


Chromosomes

- In humans DNA is not a single long chain.
- In the nucleus of each cell, the DNA molecule is packaged into thread-like structures called **chromosomes**.
- Each chromosome is made up of DNA tightly coiled many times around proteins called histones that support its structure.

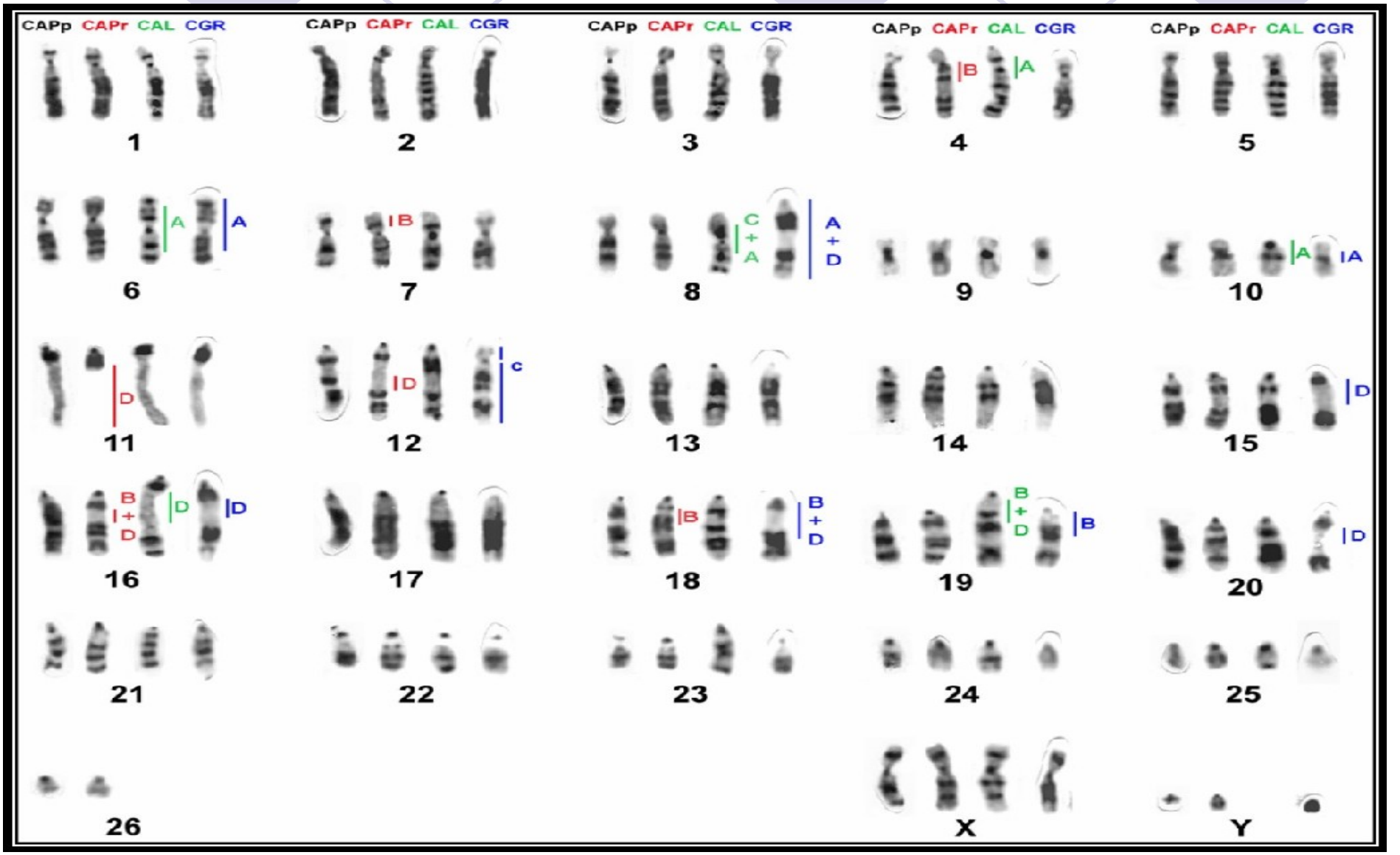


Chromosomes



- In humans, each cell contains 23 pairs of chromosomes, making a total of 46.
- 22 of these pairs, called autosomes, look the same in both males and females.
- The 23rd pair, the sex chromosomes, differ between males and females. Females have two copies of the X chromosome, while males have one X and one Y chromosome.
- When chromosomes are stained they can be quite easily recognised by their distinctive stripy patterns.

Chromosomes

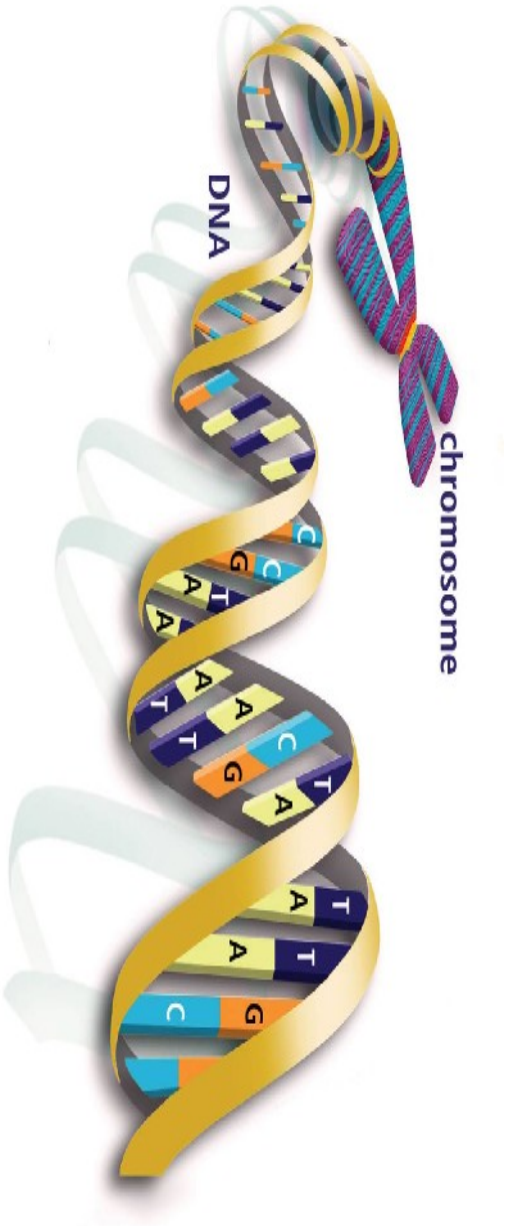


Genome

- A genome is a living creatures entire genetic information. It is the living creature's genetic material which contains the hereditary instructions for constructing, operating, and maintaining an organism, and transferring life on to the next generation
- In humans, a copy of the entire genome—more than 3 billion DNA base pairs—is contained in all cells that have a nucleus.

Gene Terminology

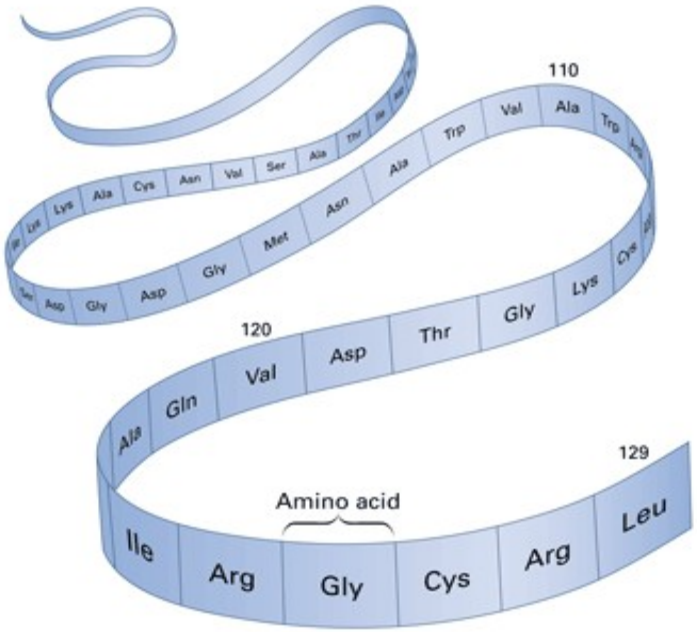
- DNA – long chain molecule comprising a sequence of bases
- Genes – basic units of heredity within a DNA strand which code for proteins and their regulation
- Chromosomes – coiled strands of DNA within the nucleus usually paired.
- Genomes – the entire hereditary material for an organism – all the chromosomes.



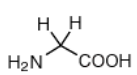
What do Genes do?

- Store the information needed to build and maintain an organism and pass it onto the next generation
- Code for proteins and their regulation
  - Store information needed to make proteins
  - Regulate when proteins are made
  - Regulate the overall cell construction process
  - Control cell differentiation
  - Other things we don’t know about !

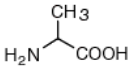
Proteins are made from chains of amino acids



Small

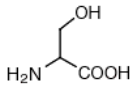


Glycine (Gly, G)
MW: 57.05

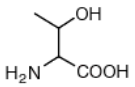


Alanine (Ala, A)
MW: 71.09

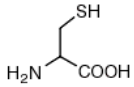
Nucleophilic



Serine (Ser, S)
MW: 87.08, pK_a ~ 16

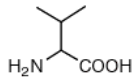


Threonine (Thr, T)
MW: 101.11, pK_a ~ 16

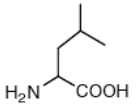


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

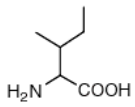
Hydrophobic



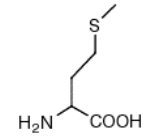
Valine (Val, V)
MW: 99.14



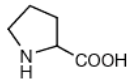
Leucine (Leu, L)
MW: 113.16



Isoleucine (Ile, I)
MW: 113.16

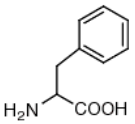


Methionine (Met, M)
MW: 131.19

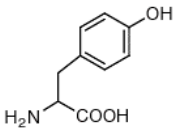


Proline (Pro, P)
MW: 97.12

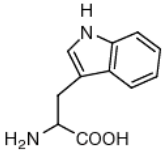
Aromatic



Phenylalanine (Phe, F)
MW: 147.18

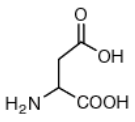


Tyrosine (Tyr, Y)
MW: 163.18

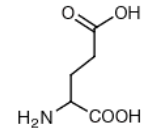


Tryptophan (Trp, W)
MW: 186.21

Acidic

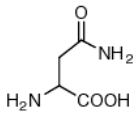


Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

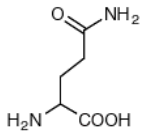


Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

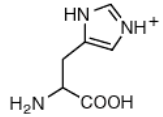


Asparagine (Asn, N)
MW: 114.11

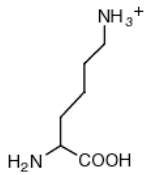


Glutamine (Gln, Q)
MW: 128.14

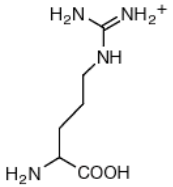
Basic



Histidine (His, H)
MW: 137.14, pK_a = 6.04



Lysine (Lys, K)
MW: 128.17, pK_a = 10.79



Arginine (Arg, R)
MW: 156.19, pK_a = 12.48

How are proteins made from DNA?

- The genetic code defined by the sequence of bases in DNA defines the sequence of amino acids in the protein
- Sequences of three DNA bases (called codons) code for each amino acid
- A complex process builds the protein from the information in this code.

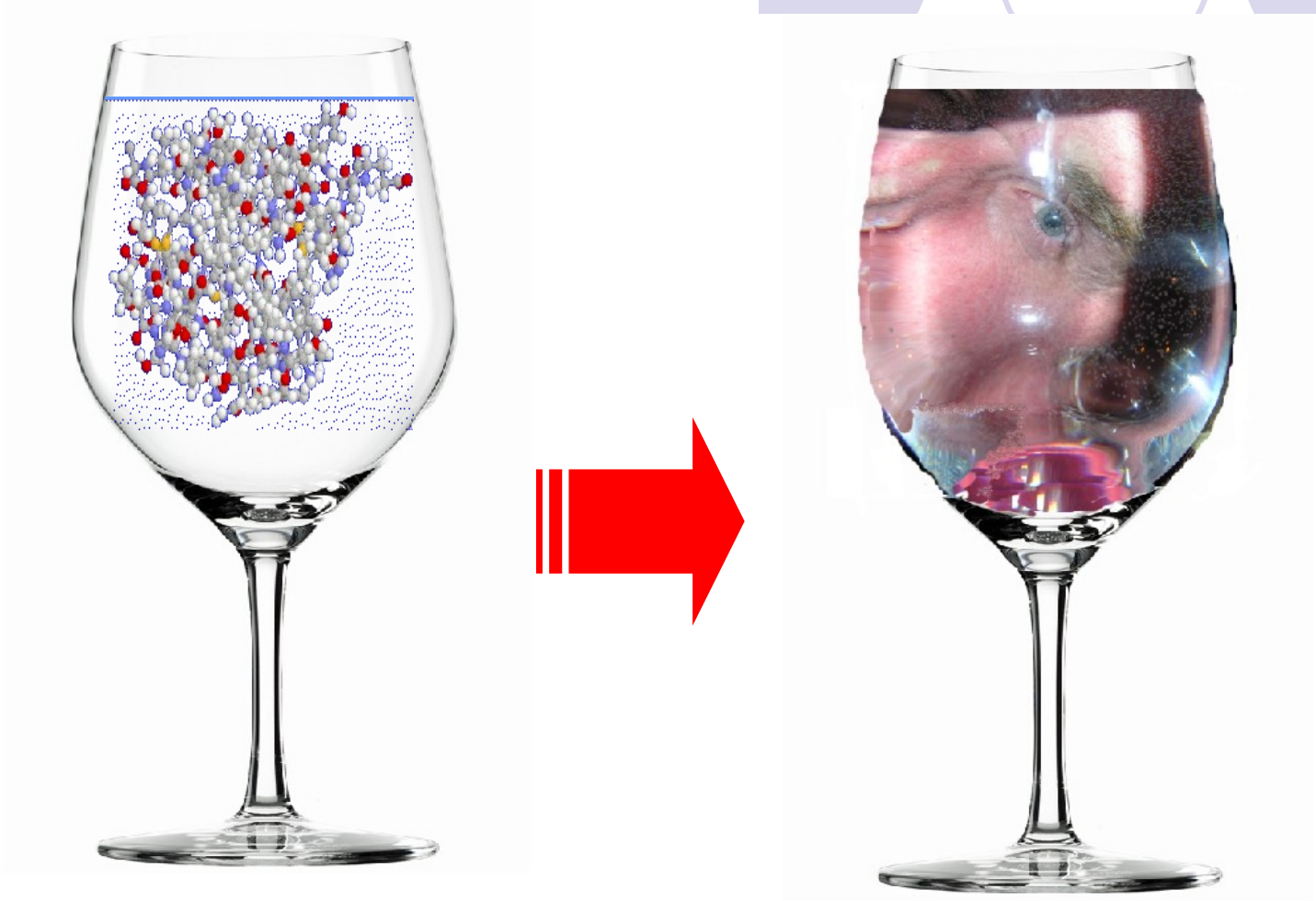
20 Amino Acids and their DNACodons

Figure 1.3 DNA Codes for Amino Acids

	T		C		A		G	
T	TTT	phenylalanine	TCT	serine	TAT	tyrosine	TGT	cysteine
	TTC	phenylalanine	TCC	serine	TAC	tyrosine	TGC	cysteine
	TTA	leucine	TCA	serine	TAA	stop	TGA	stop
	TTG	leucine	TCG	serine	TAG	stop	TGG	tryptophan
C	CTT	leucine	CCT	proline	CAT	histidine	CGT	arginine
	CTC	leucine	CCC	proline	CAC	histidine	CGC	arginine
	CTA	leucine	CCA	proline	CAA	glutamine	CGA	arginine
	CTG	leucine	CCG	proline	CAG	glutamine	CGG	arginine
A	ATT	isoleucine	ACT	threonine	AAT	asparagine	AGT	serine
	ATC	isoleucine	ACC	threonine	AAC	asparagine	AGC	serine
	ATA	methionine	ACA	threonine	AAA	lysine	AGA	arginine
	ATG	methionine	ACG	threonine	AAG	lysine	AGG	arginine
G	GTT	valine	GCT	alanine	GAT	aspartic acid	GGT	glycine
	GTC	valine	GCC	alanine	GAC	aspartic acid	GGC	glycine
	GTA	valine	GCA	alanine	GAA	glutamic acid	GGA	glycine
	GTG	valine	GCG	alanine	GAG	glutamic acid	GGG	glycine

However…..

- Less than 2% of DNA codes for protein amino acid sequence
- What does the rest do?



Mixing together all the ingredients doesn’t make a person

Why do we need to know the gene sequence?

- The basis for how our cells function
- The basis for understanding disease
- The basis for drug development
- …

The Human Genome Project

- Objectives
- History

Human Genome Project Objectives

- To identify all the approximately 20,000-25,000 genes in human DNA
- To determine the sequences of the 3 billion chemical base pairs that make up human DNA
- To store this information in databases
- To improve tools for data analysis
- To transfer related technologies to the private sector
- To address the ethical, legal, and social issues that may arise from the project.

The Human Genome Project History

- The $3-billion project was formally founded in 1990 by the US Department of Energy and the National Institutes of Health
- a 'rough draft' of the genome was finished in 2000
- The Human Genome Project was declared complete in April 2003
- major quality assessment of the human genome sequence was published in May 27, 2004 indicating over 92% of sampling exceeded 99.99% accuracy

Early DNA Sequencing

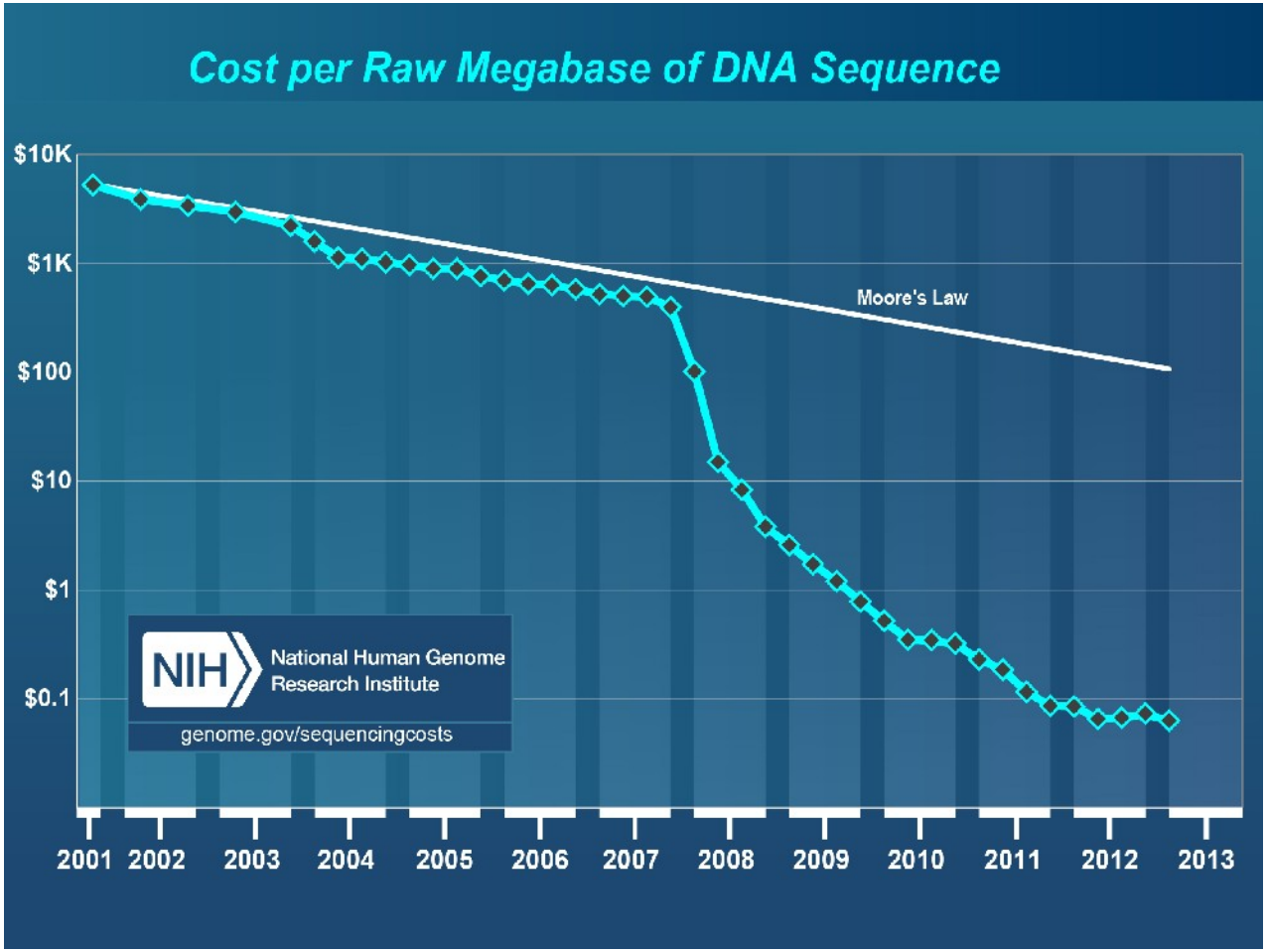
- DNA fragments created by selective cleavage at known points
- Fragments tagged with radioactive markers
- Fragments separated by gel electrophoresis
- DNA sequence assembled by hand

Tedious, repetitive, time-consuming, costly, error prone.

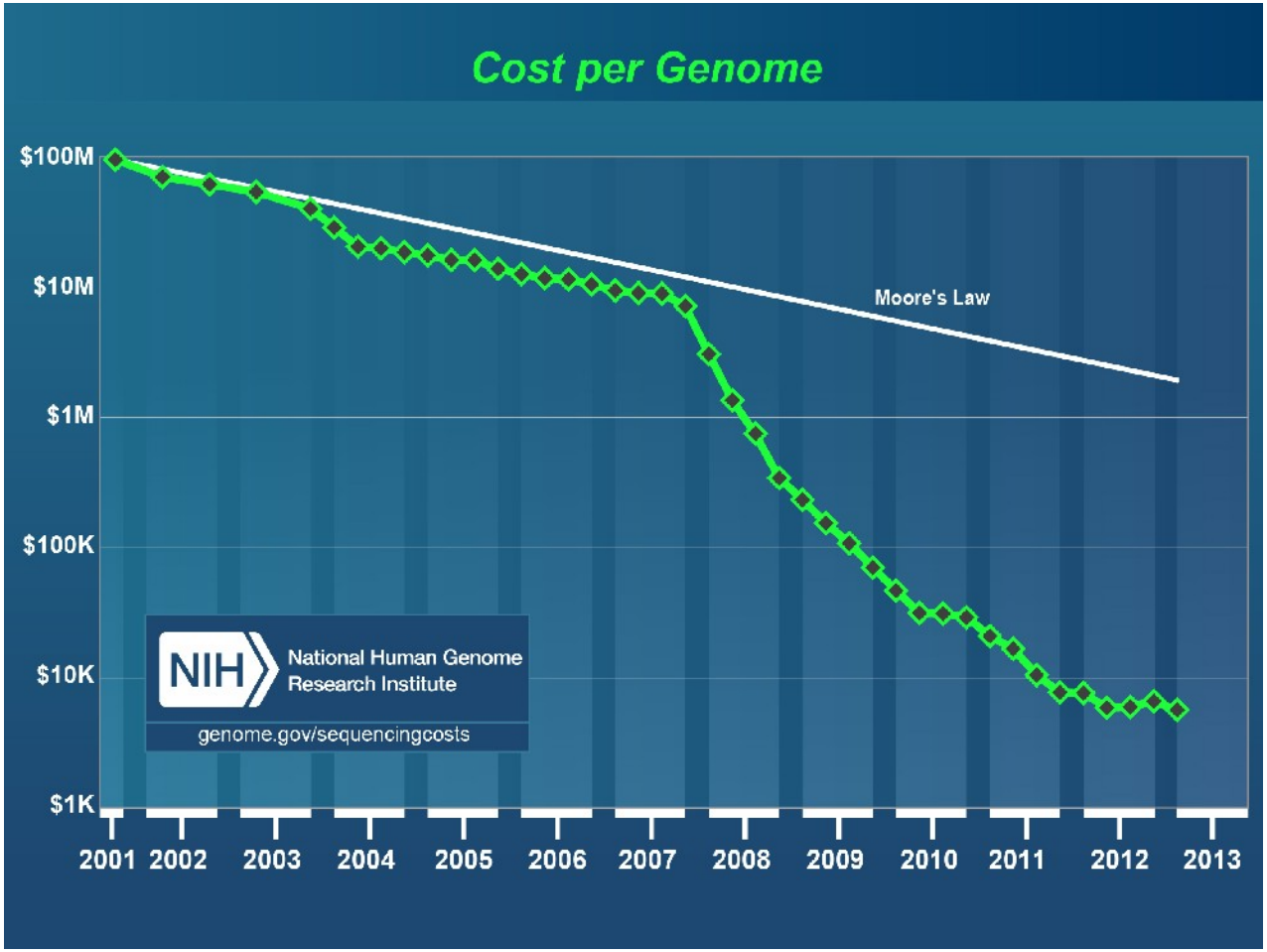
Major Improvements

- Polymerase Chain Reaction process which allows the production of millions of copies of a DNA sample
- Creation of DNA sequencing machines that replace gel electrophoresis with capillary electrophoresis

Gene Sequencing Costs



Gene Sequencing Costs



DNA Sequencers 2005

USB stick can sequence DNA in seconds

Updated 19:14 22 February 2012 by **Duncan Graham-Rowe**

Read more: Click here to read a later version of this story

PRINT SEND SHARE

It may look like an ordinary USB memory stick, but a little gadget that can sequence DNA while plugged into your laptop could have far-reaching effects on medicine and genetic research.

The UK firm Oxford Nanopore built the device, called MinION, and claims it can sequence simple genomes – like those of some viruses and bacteria – in a matter of seconds. More complex genomes would take longer, but MinION could also be useful for obtaining quick results in sequencing DNA from cells in a biopsy to look for cancer, for example, or to determine the genetic identity of bone fragments at an archaeological dig.

The company demonstrated today at the Advances in Genome Biology and Technology (AGBT) conference in Marco Island, Florida, that MinION has sequenced a simple virus called Phi X, which contains 5000 genetic base pairs.

(Image: Oxford Nanopore Technologies)

Benefits

- Focus on looking to the fundamental causes of disease rather than by treating symptoms
- Rapid and more specific diagnostic tests
- Organ donor matching in transplant programs
- Identification of susceptible individuals
- Augmentation or even replacement of defective genes through gene therapy