

# **Current topic 1: DNA and Human genome project**

## **Lecture 2 – Mutations and human genome project**

Dr. Maggy Fostier.

[Maggy.fostier@manchester.ac.uk](mailto:Maggy.fostier@manchester.ac.uk)

# **Part A- Mutations**

# Definitions

**Mutation** = permanent alteration in the DNA sequence passed on into daughter cells and sometimes into gametes.

Size ranges from single base pair (**Single Nucleotide Polymorphism (SNP)** - pronounce 'snip') (**micro**) to large segment of a chromosome that includes multiple genes (**chromosome rearrangement** = **macro**).

**Heredity (germ line) mutation** = inherited from a parent gamete and present throughout a person's life in every cell in the body.

**Acquired (or somatic) mutation** = occurs at some time during a person's life and present only in certain cells (where it occurred) and their daughter cells (through mitosis).

If present in own gamete, it can be passed onto progeny (through meiosis).



# Causes of new mutations



## Environment

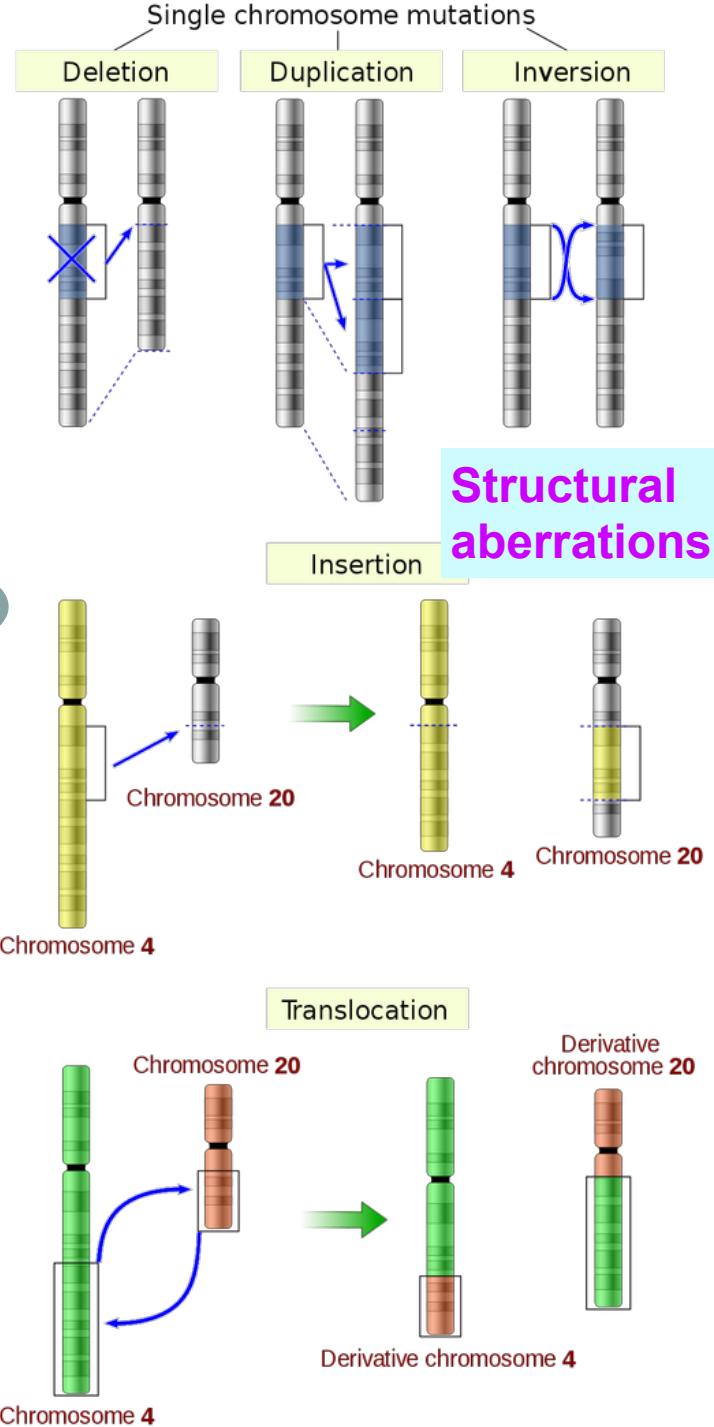
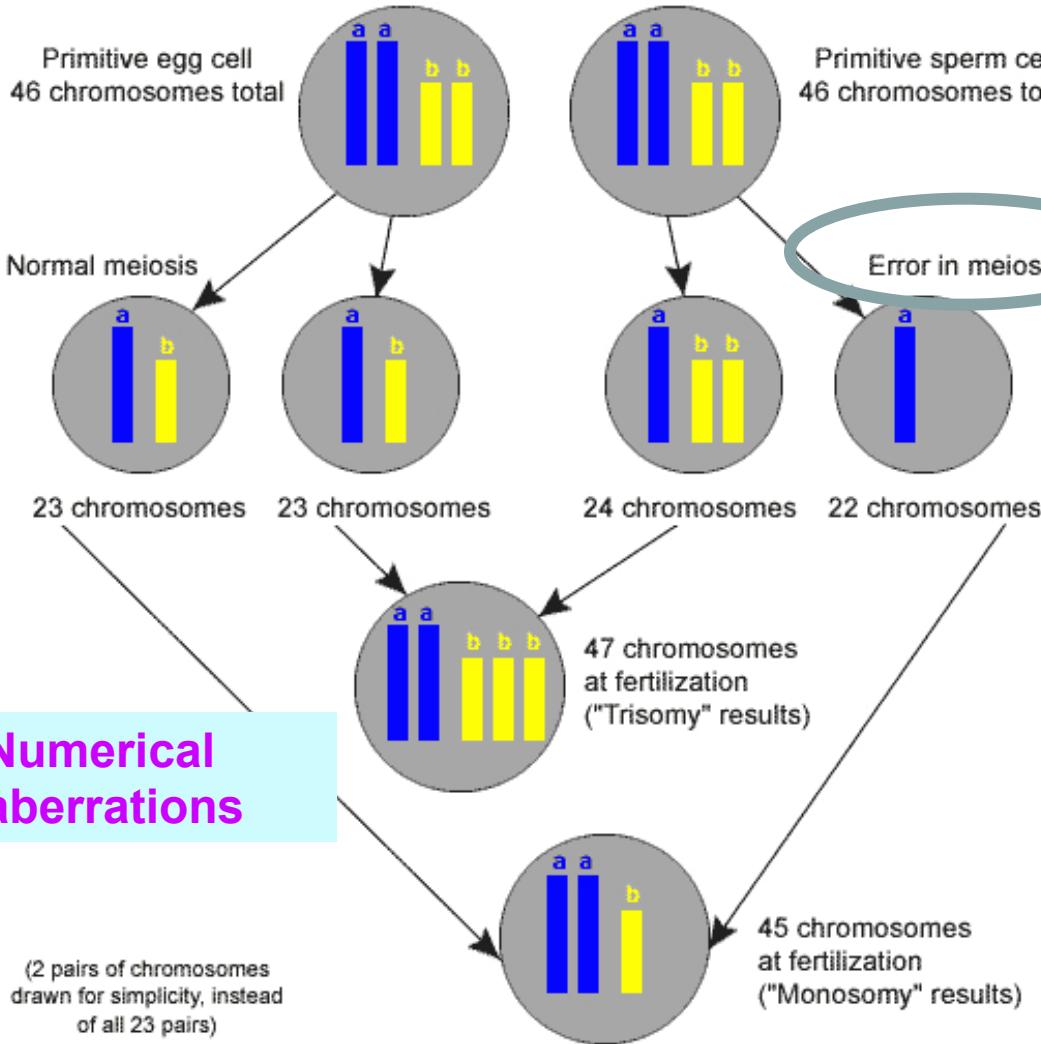
- **Mutagens or environmental factors** (e.g. radiations) causing breaks in between DNA bases or intercalating between bases
- **Biological factors** (e.g. viruses) that integrate in the genome and can cause a disturbance into an important DNA region.

## Intrinsic, usually linked to ageing

- **Errors during DNA replication** before mitosis
- **Errors during DNA repair**
- **Errors during meiosis** (e.g. error in chromosome separation, Trisomy 21)

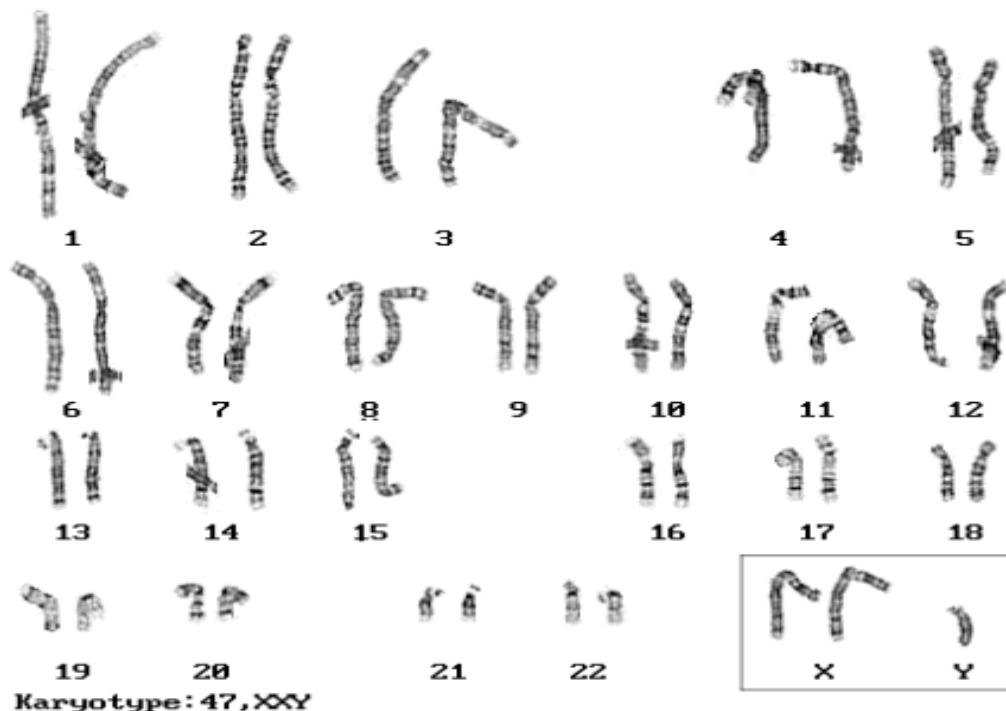
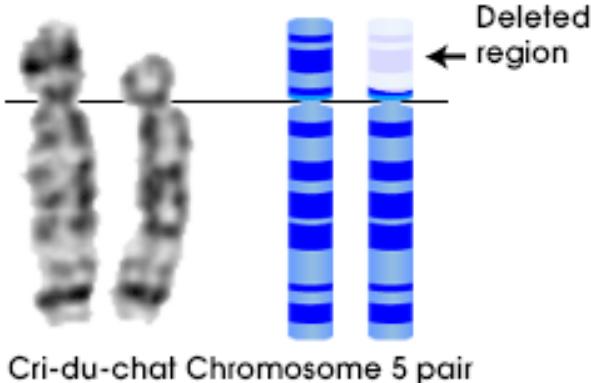
# Macro mutations – affect chromosomes

Occur during meiosis or in late stage of cancer



# Examples of diseases caused by macro mutations

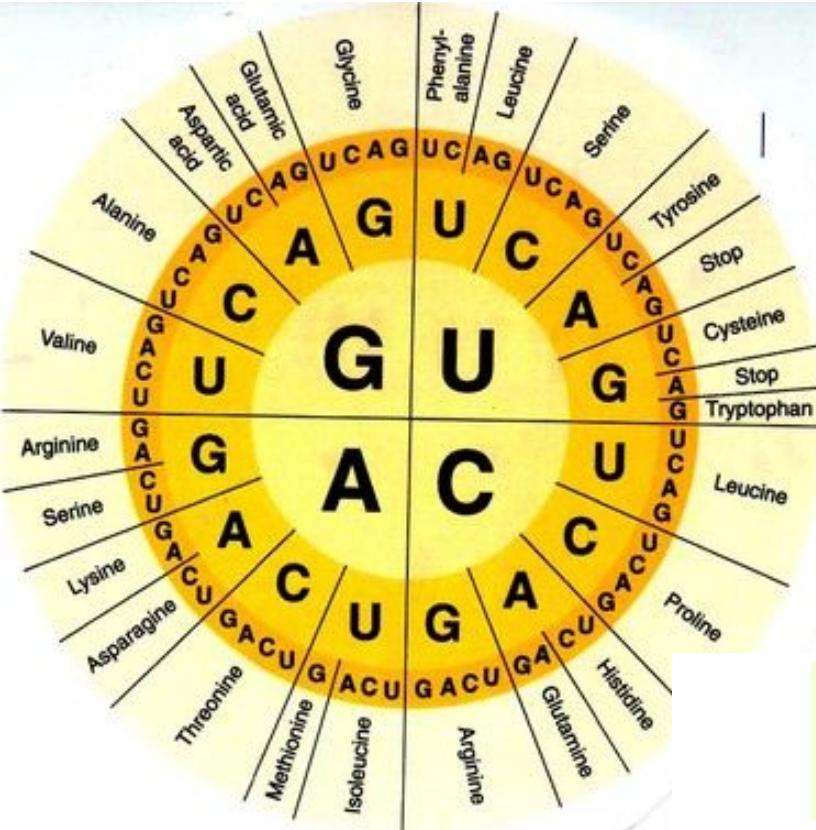
- Down syndrome (extra 21)
- Klinefelter syndrome (2X+Y)
- Cri du chat (deletion on 5)



Klinefelter syndrome (2X+Y)

<http://learn.genetics.utah.edu/content/disorders/chromosomal/>  
and <http://www.biology.iupui.edu/biocourses/N100/2k2humancsomaldisorders.html>

# Micro mutations - SNPs



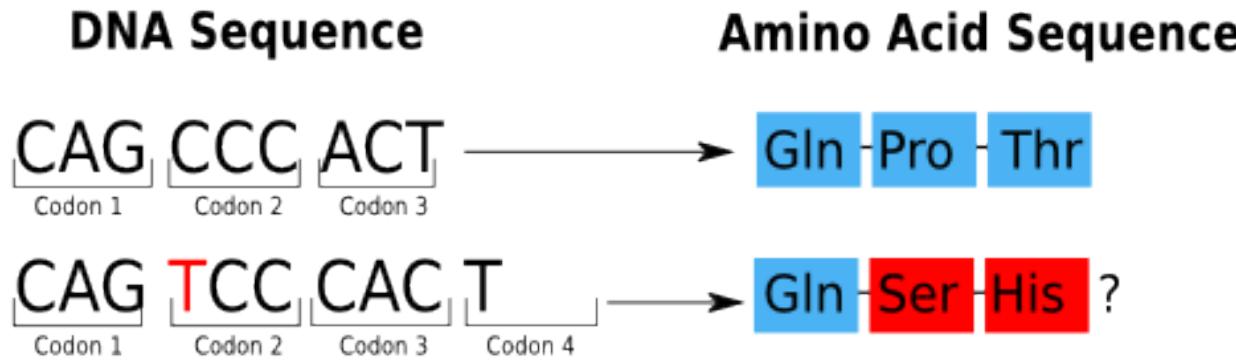
RNA codon wheel.  
Third letter is less  
discriminating.

mRNA level  
protein level

# Micro mutations: small deletions and insertions

**Insertions and deletions** can cause great disturbance to a protein through frameshift mutations, unless the number of bases deleted/inserted is a multiple of 3 (no frameshift).

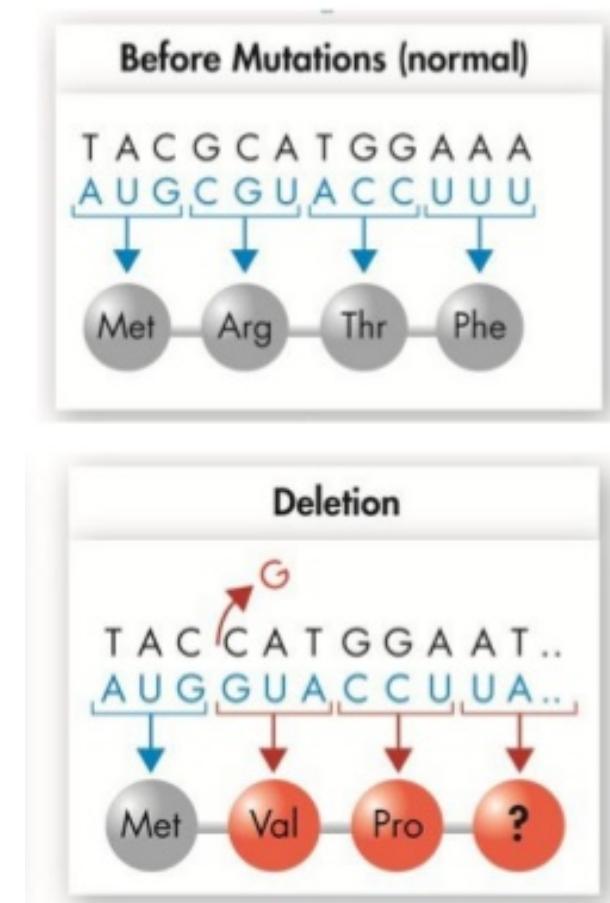
## Insertion



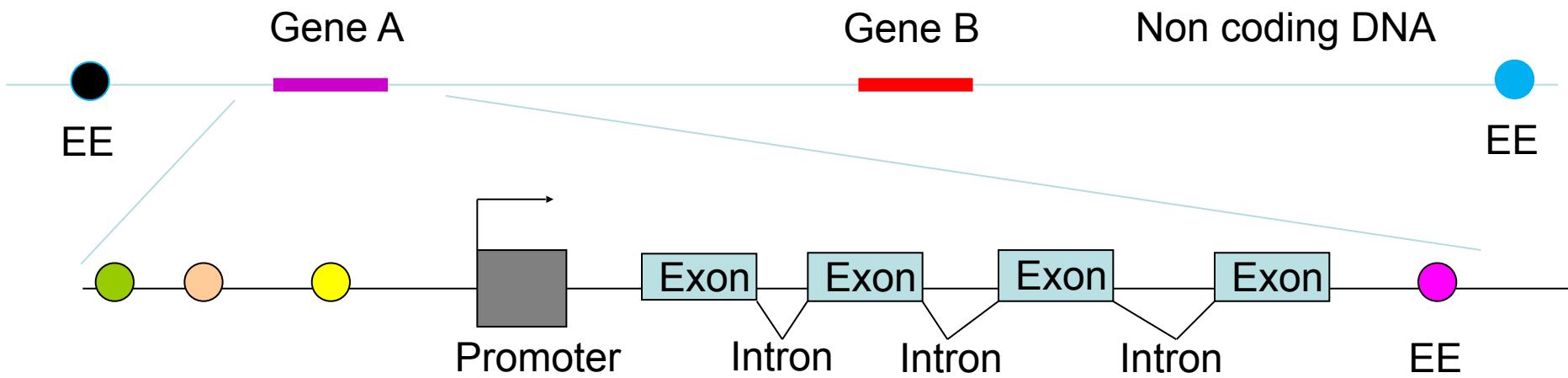
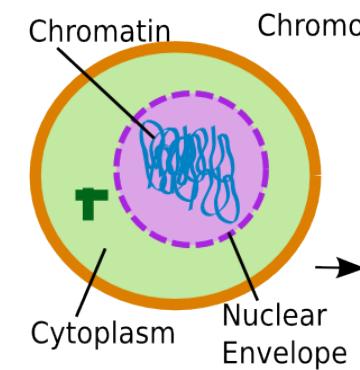
**Frameshift mutations** usually lead to an early STOP codon → usually very bad: shorter protein or no protein



**Insertion of 3 bases** only adds one codon.

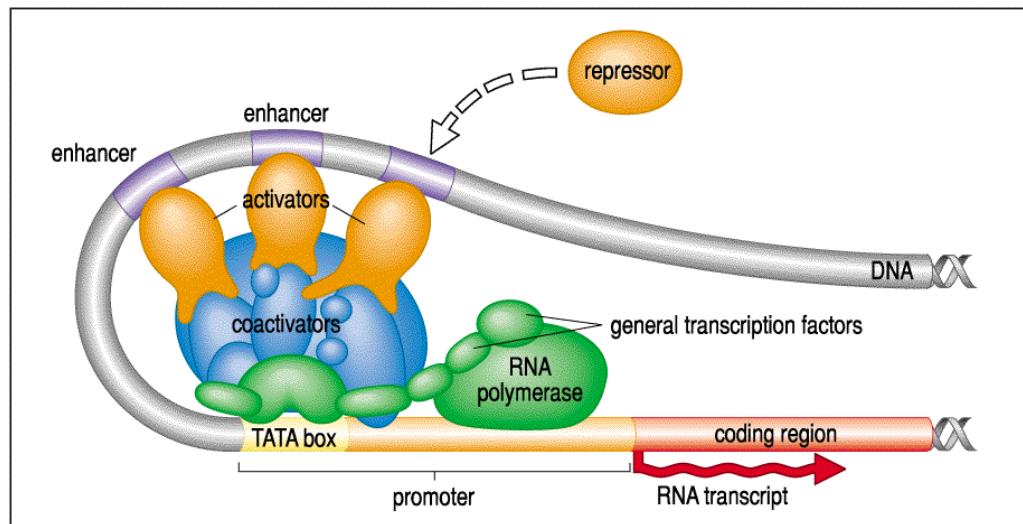


# The importance of mutation location



EE: Enhancer element:

- Sequence to which a protein that regulates transcription can bind.
- Can be close or very far away from target gene

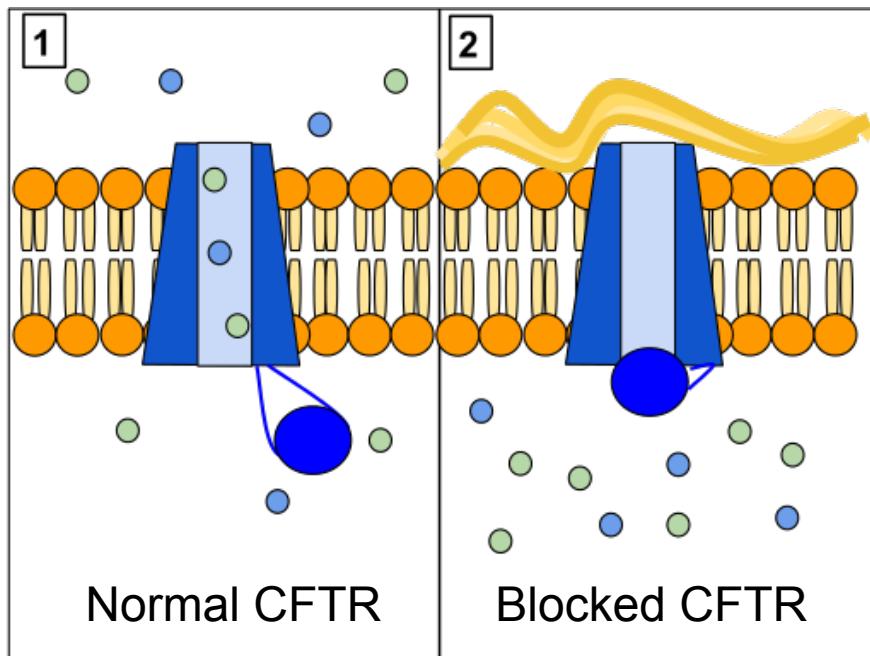


# Example 1: Cystic Fibrosis (CF)

Mutations in the CFTR gene can cause CF.

> 900 'bad' alleles exist -all recessive against the normal allele which produces enough protein to compensate.

CF patients are either homozygous for a given 'bad' allele or heterozygous for two different 'bad' alleles.



Mutation	Result
482 C G C ↓ C A C	Arg-117 His-117
1609 C A G ↓ T A G	Gln-493 STOP
<b>Insertion</b> of 2 nucleotides (AT) at 2566	<b>Frameshift</b>
<b>Deletion</b> of one C at 3659	<b>Frameshift</b>
<b>Deletion</b> of 3 nucleotides at 1654-1656	Deletion of Phe-508
	<b>Most common</b>

# Example 2: Trinucleotide repeat expansions (insertions) – genetic stutters

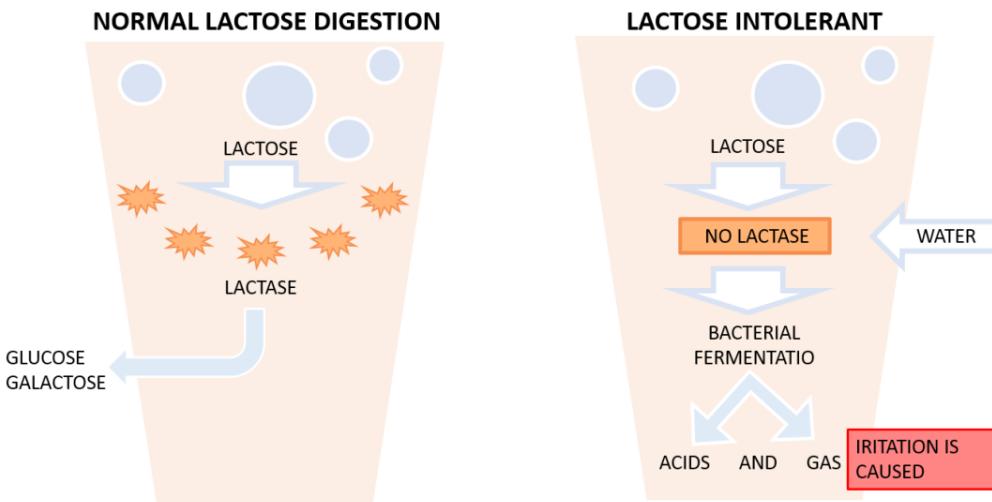
Disease	repeats	range	
		normal	mutant
*Huntington Disease	(CAG) $n$	11 - 34	36 - 120
*Spinocerebellar ataxia I	(CAG) $n$	6 - 40	40 - 82
*Machado-Joseph ataxia	(CAG) $n$	13 - 40	68-69
Friedrich's ataxia	(GAA) $n$	10 - 21	200 - 900
Myotonic dystrophy	(CTG) $n$	5 - 30	50 - 2000
Fragile X syndrome	(CGG) $n$	6 - 50	60-1000

\* The (CAG) $n$  is within an exon and produces polyglutamines in the protein product

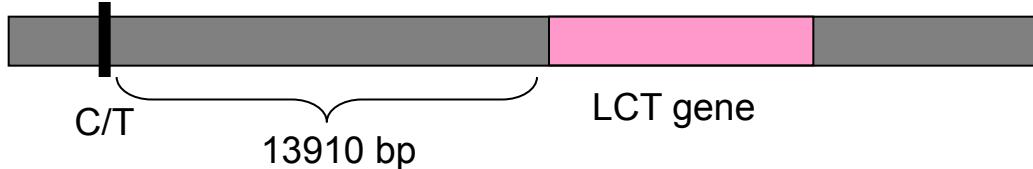
CAG codes for glutamine. When the polyglutamine stretch becomes too big, the proteins aggregate and get destroyed or prevent others to work properly.

Other triplets do not code for amino acid but their extra number interferes with the chromosome structure and the transcription of the nearby gene.

# Example 3: SNP in promoter region regulates lactose tolerance



Oct-1 binds when it is a T and activates LCT transcription.



Our ancestors no longer produced lactase from the LCT gene after weaning because a repressor prohibited its expression.

The ancestral allele, -13910C, does NOT allow for Oct-1 binding.

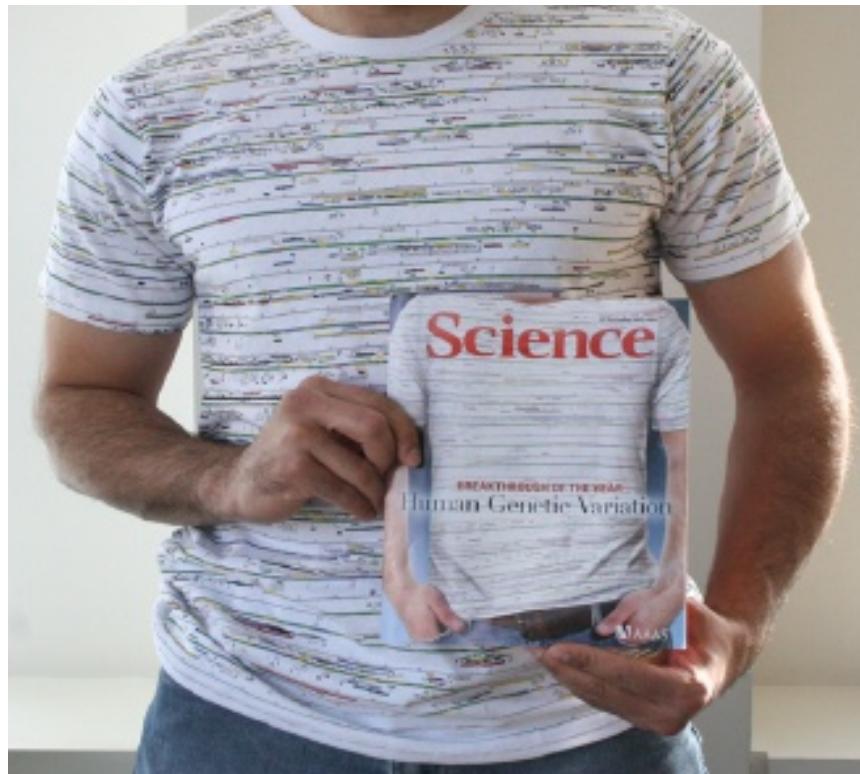
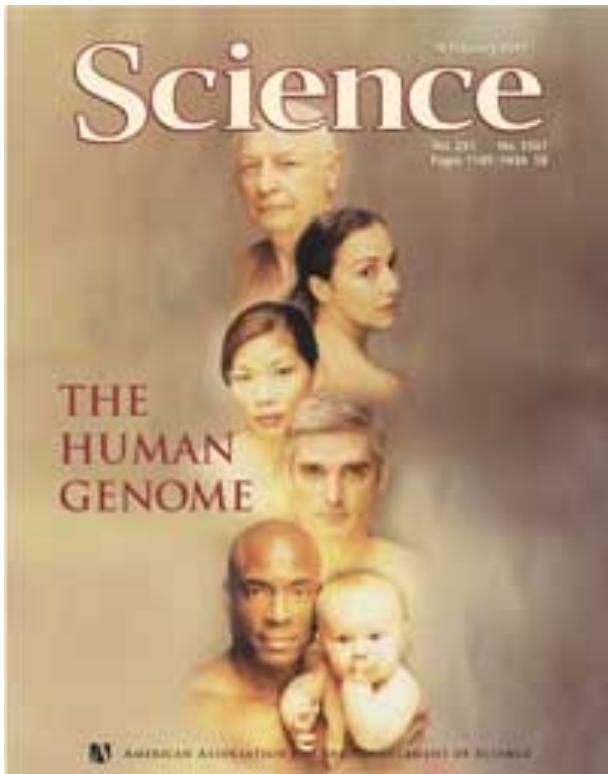
When the European SNP occurred (-13910 C→T), it was a new mutation creating a new function, favoured by the environment. The T allele has a dominant effect as just a bit more lactase production provides tolerance.

## **Part B –**

**What have we learnt from the Human Genome Project (HGP)?**

# Findings from the first draft

# 2004: First complete draft of human genome published.



It took 13 years to sequence 3 billion ( $10^9$ ) base pairs.  
DNA from 5 anonymous individuals of varying ethnicity.

# Finding 1 - Humans only have ~20,500 genes

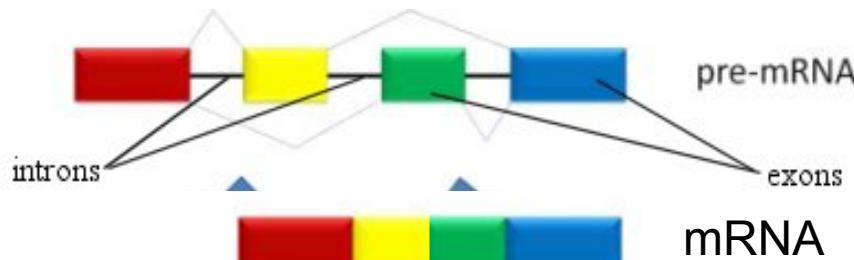
**Before HGP:** We thought different organisms must have very different genes, and complex organisms must have many more genes than simpler organisms. Flies have 13,000 genes so humans should hold > 100,000 genes.

**After HGP:**

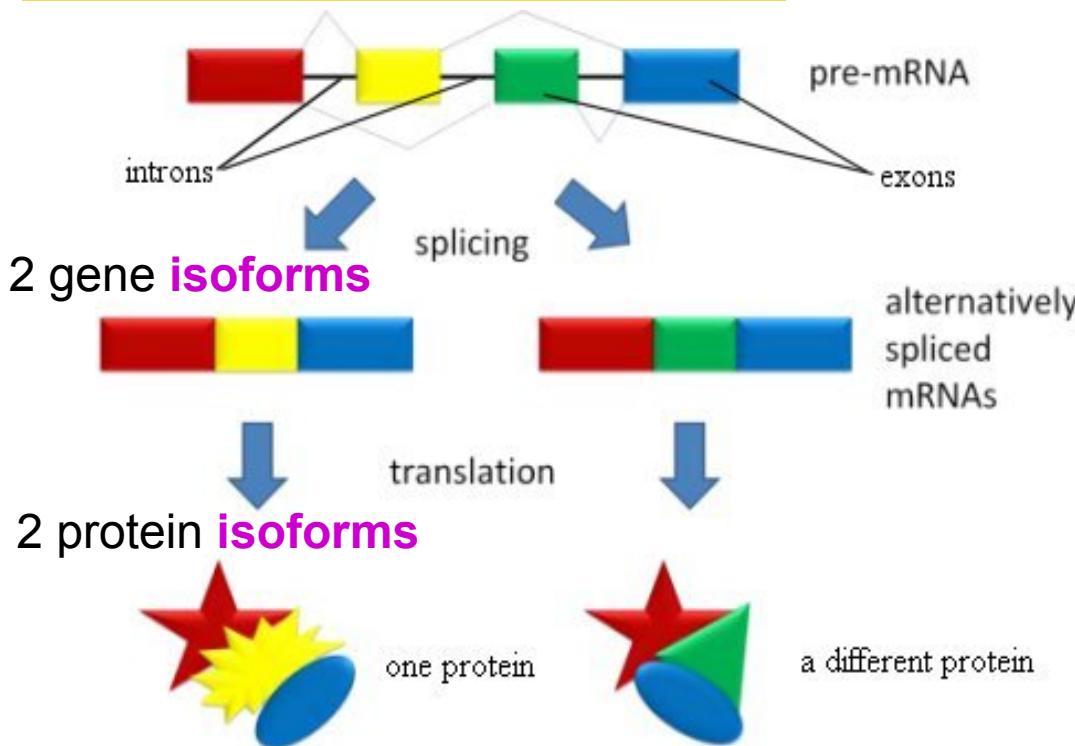
- Human ONLY have ~20,500 genes
- ~60% of these are similar between fly and human.
- ONLY ~2% of the human DNA codes for genes.

# Why so few genes?

## Normal splicing



## Alternative splicing



We now know that most genes code for more than one protein through **alternative splicing**

Scientists reckon that each of our ~20,000 genes has ~ 5 isoforms.

→that way, you end up with >100,000 proteins.

→Cells have the same genome, but they do not express the same genes and isoforms

→When and where these protein isoforms are expressed in the body determines what each cells become.

# Finding 2 – Humans are 99.9% identical ([www.genome.gov](http://www.genome.gov)).

0.1% variation = 1 in 1000  
base pairs > 3 millions  
(genome is 3 billion) →  
mostly SNPs.

**Landmarks in genome:** ‘don’t know where I am on motorway, but am between exit 22 and service station.’

Can change timing, location or level of expression. E.g. lactase persistence

E.g. No longer tastes bitter substance of vegetables

**Linked SNPs outside of gene**  
no effect on protein production or function

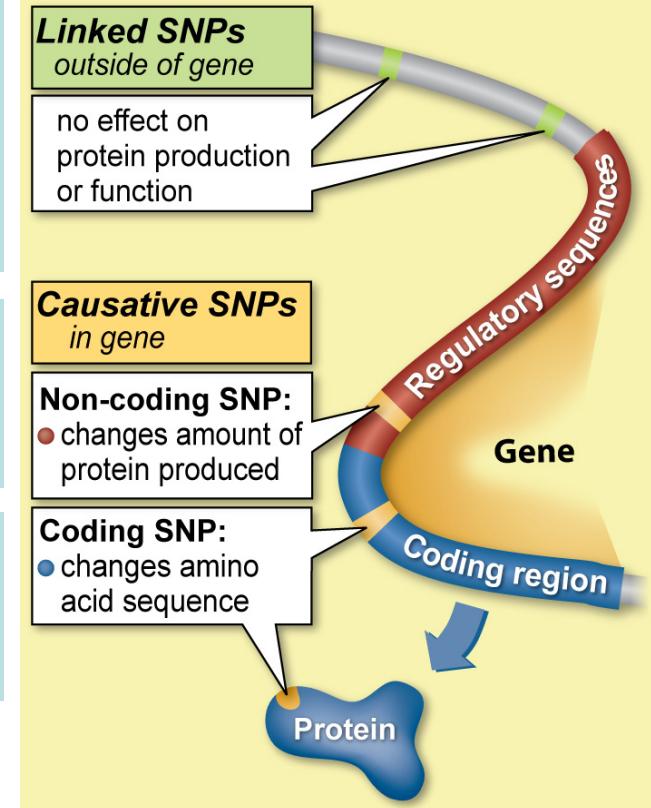
**Causative SNPs in gene**

**Non-coding SNP:**

- changes amount of protein produced

**Coding SNP:**

- changes amino acid sequence

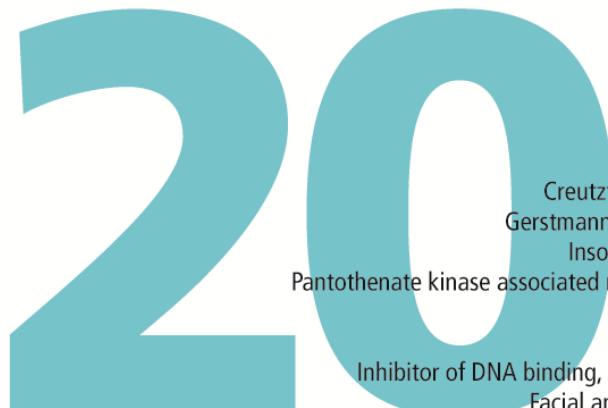


Most SNPs are outside genes (no effect on phenotype).

# Finding 3- location of 1,400 disease-causing mutations identified

Before HGP, we knew <100

Example: Below are the landmarks of chromosome 20.



63 million base pairs



Diabetes insipidus, neurohypophyseal  
McKusick-Kaufman syndrome  
Cerebral amyloid angiopathy  
Thrombophilia  
Myocardial infarction, susceptibility to  
Huntington-like neurodegenerative disorder  
Anemia, congenital dyserythropoietic  
Acromesomelic dysplasia, Hunter-Thompson type  
Brachydactyly, type C  
Chondrodysplasia, Grebe type  
Hemolytic anemia  
Myeloid tumor suppressor  
Breast cancer  
Maturity Onset Diabetes of the Young, type 1  
Diabetes mellitus, noninsulin-dependent  
Graves disease, susceptibility to  
Epilepsy, nocturnal frontal lobe and benign neonatal, type 1  
Epiphyseal dysplasia, multiple  
Electro-encephalographic variant pattern  
Pseudohypoparathyroidism, type IB

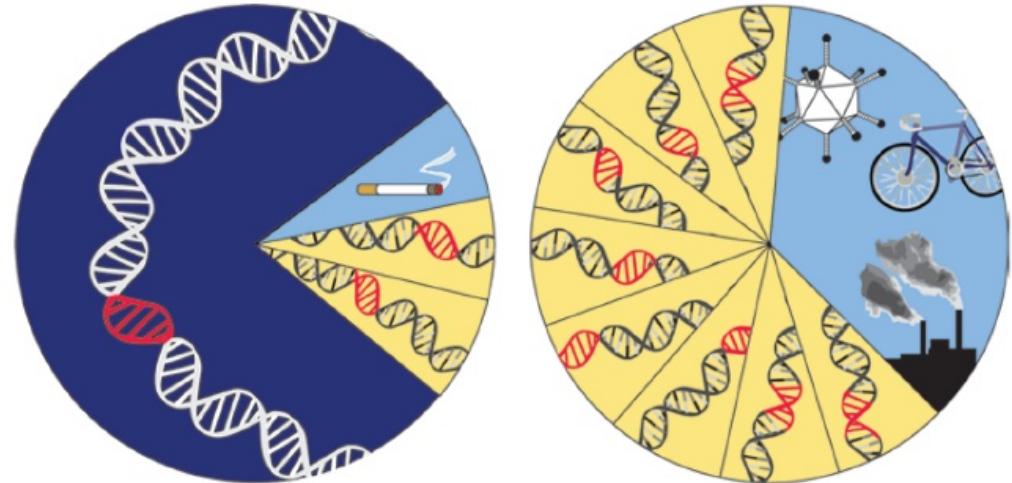
ONGOING

# Exploiting SNPs with GenomeWide Association Studies (GWAS)

# Exploiting SNPs

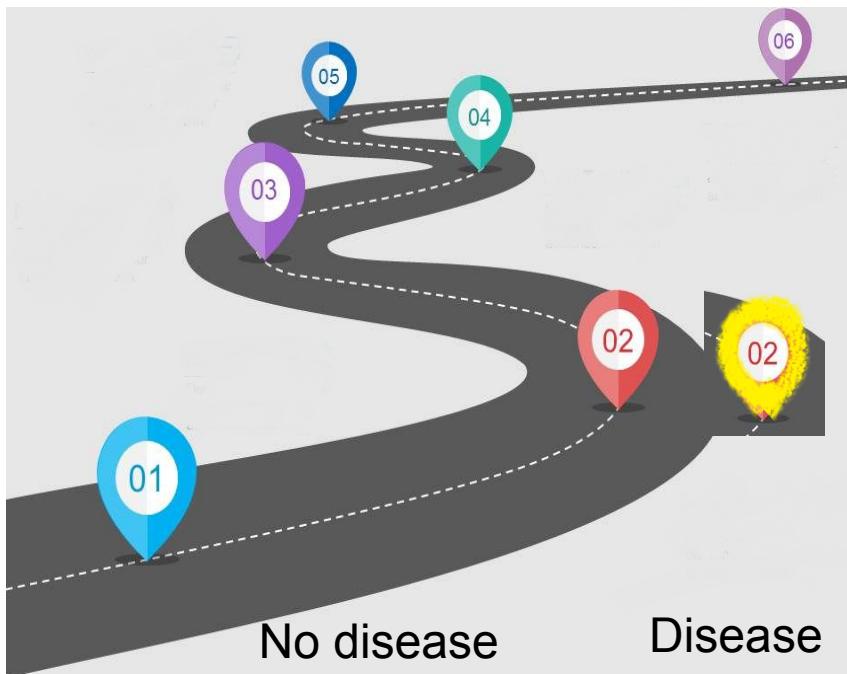
Most traits and diseases are multifactorial  
= polygenic + environmental interactions, e.g. cancers.

Finding the genetic risk factors is difficult



# Exploiting SNPs

~10 million SNPs constitute 90% of the variation in the population. They are interspersed in the genome and provide mapping references.





# Genome Wide Association Studies = GWAS

## Affected Individuals

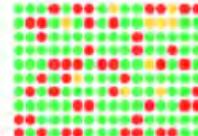
Profile 1

Profile 2

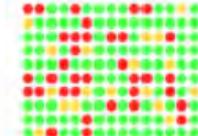
Profile 3



## Unaffected Individuals



**SNPs analyzed  
and compared  
statistically**



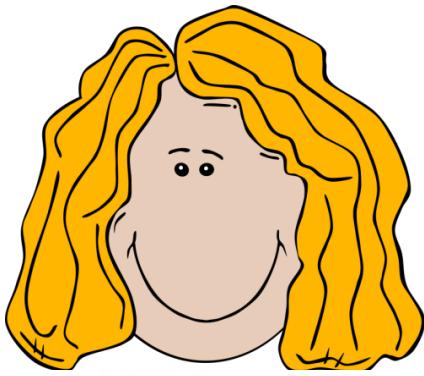
GWAS aim to identify the common genetic variants (SNPs) associated with complex diseases and traits, by testing a minimum of hundreds of thousands of SNPs in large population samples.

# What can GWAS tell us? 1

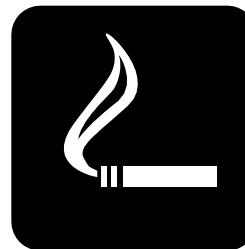
- When particular landmark SNPs are seen in greater proportions in diseased patients compared to controls, we say these SNPs are **associated** with the disease.
- It helps **directing further studies** to find genetic risk factors.
- It can help **establishing predisposition** to a disease, but it is very rare to have a 100% association.
- Having an SNP associated with a complex disease does not mean you will develop the disease, but you will have a higher risk.

# What can GWAS tell us? 2

Some SNPs may help us establish the risk of environmental factors.



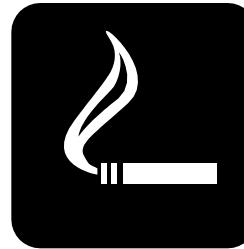
SNP PROFILE A +



= No development of disease



SNP PROFILE B +



= Development of disease

# What can GWAS tell us? 3

**Pharmacogenomics** = How do patients genomes affect response to treatment?

A SNP profile can be used to stratify patients

Drug treatment worked



Drug treatment didn't work



SNPs predictive of efficacy



SNPs predictive of NO efficacy

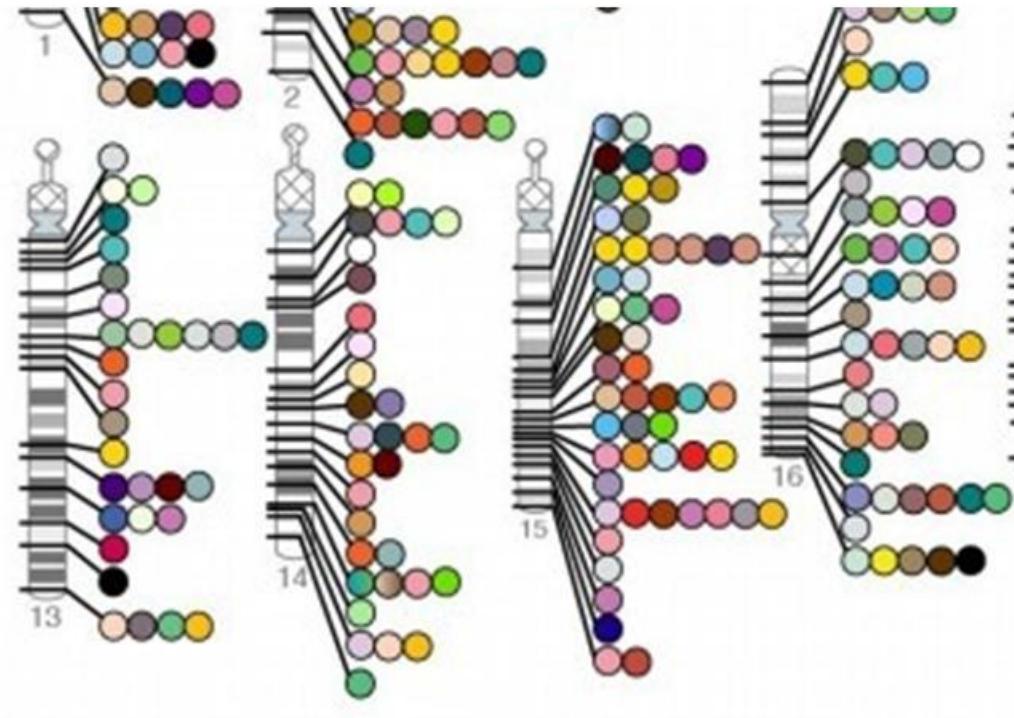


See Dr  
Cochran's  
lectures

# Results of GWAS

In 2005, <20 SNPs were known to be significantly associated to diseases.

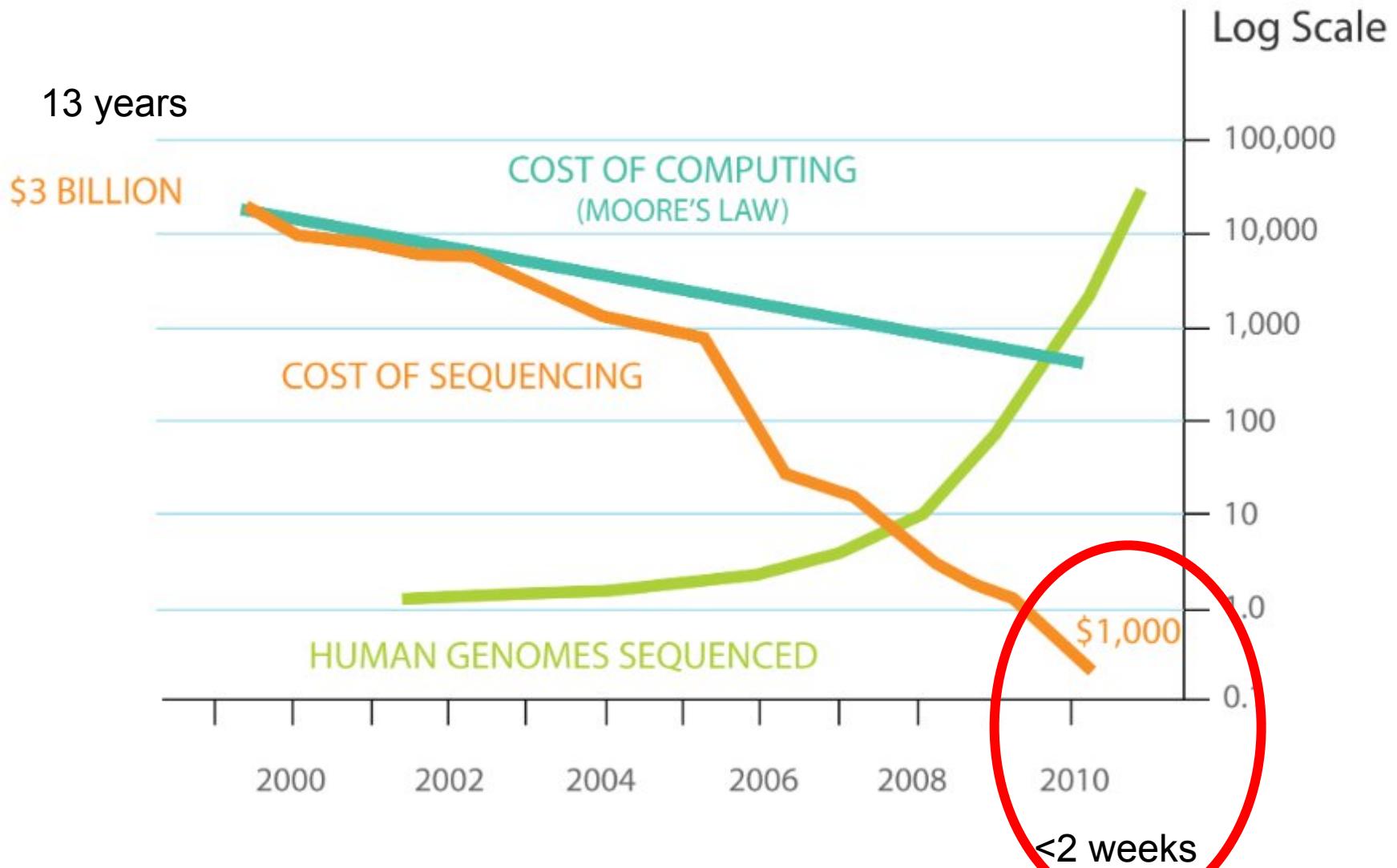
**In 2015, >14,000 SNPs significantly associated to >1500 traits and put into a public catalog (right)-ONGOING**

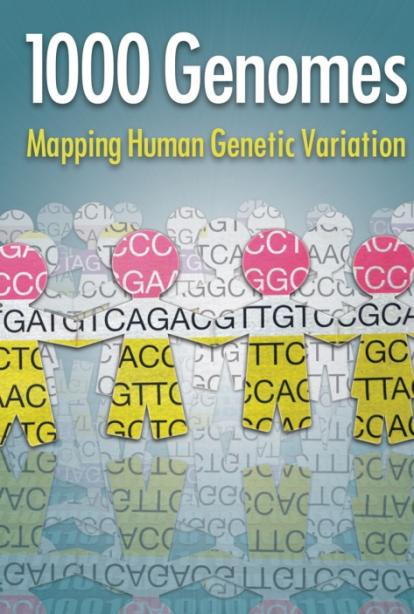


<https://www.ebi.ac.uk/gwas/diagram>

More HGPs and  
private genome  
sequencing

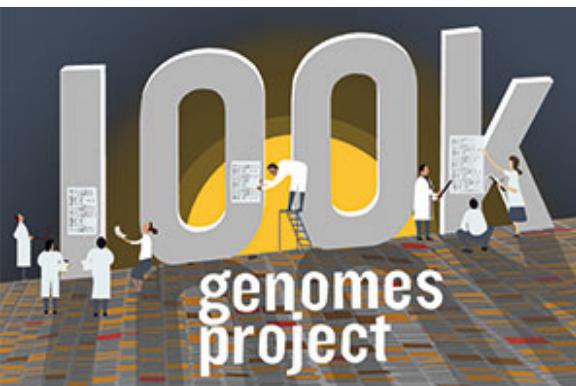
# Sequencing technology has evolved





# 2008-15: The 1000 genomes project (International)

Aim: To establish the most detailed catalogue of human genetic variations - worldwide representation.



# 2014-18: The 100,000 genomes project (UK)

Aim: To help research and transform diagnosis and treatment for patients with cancer and rare genetic diseases.

“To date, actionable findings have been found for **1 in 4**/**1 in 5** rare disease patients, and around **50%** of cancer cases contain the potential for a therapy or a clinical trial.”

# You can now get your genome privately sequenced

The screenshot shows the 23andMe website. At the top, there's a navigation bar with links: welcome (highlighted in green), health, ancestry, how it works, research, buy, help, and a search icon. On the left, there's a large image of a DNA collection kit with a colorful, abstract design. A circular badge on the kit says "buy one, get 10% off each additional kit". To the right of the kit, a teal button says "Learn more about yourself this new year." Below this, a list of benefits includes: View reports on over 100 health conditions and traits, Find out about your inherited risk factors and how you might respond to certain medications, and Discover your lineage and find DNA relatives. A pink button at the bottom says "order now" and next to it is the price "£125 Shipping included".

<https://www.23andme.com/en-gb/>

Partial genome, focused on health and ancestry. Limited advice/counselling, so not complete health screen and cheaper.

[http://www.illumina.com/clinical/illumina\\_clinical\\_laboratory.html](http://www.illumina.com/clinical/illumina_clinical_laboratory.html)

Whole genome with full advice. Focused on health. More expensive.

Human Genome Project The Observer

What happened when I had my genome sequenced

<http://www.theguardian.com/science/2013/jun/08/genome-sequenced>, June 2013

The screenshot shows the Illumina Clinical Laboratory website. At the top, there's a navigation bar with a search icon and a "QUESTIONS" link. Below the navigation, a testimonial reads: "Thanks to answers found through genome sequencing, Shelby Valint is now able to walk and talk." A "VIEW STORY" button is located below the text. To the right, there's a photo of a young girl smiling. The background features a decorative graphic of colored DNA helixes.

Genomics-based health care is available now for you and your patients.

We are in the beginning of a new era of using genomic information to make critical health care decisions. A proven and trusted partner, the Illumina Clinical Services Laboratory offers the most comprehensive clinical whole-genome sequencing tests today.

# Check-list

- Do they sequence whole genome or bunch of SNPs
- What do they claim you will learn?
- How do they deliver results? In a clinical setting, this will be via a genetic counsellor: do they offer this? – think of this carefully – the results could cause you or your family distress especially if you do not know how to interpret them.
- Remember, we only talk about risk - not certainty- for complex diseases.
- Do they ask you to share your data for research purposes? Read the small prints carefully and if an American company, I would have serious concerns with data privacy.