

# Genotyping and Copy Number Variation



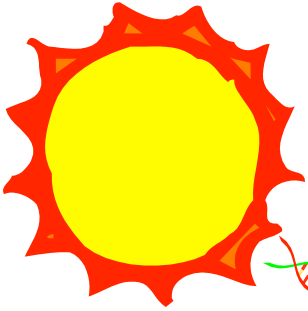
Molly Bray, PhD

and

Fady Mikhail, MD, PHD



**DNA (deoxyribonucleic acid)** The fundamental hereditary material of all living organisms. In eukaryotes, DNA is stored primarily in the cell nucleus.

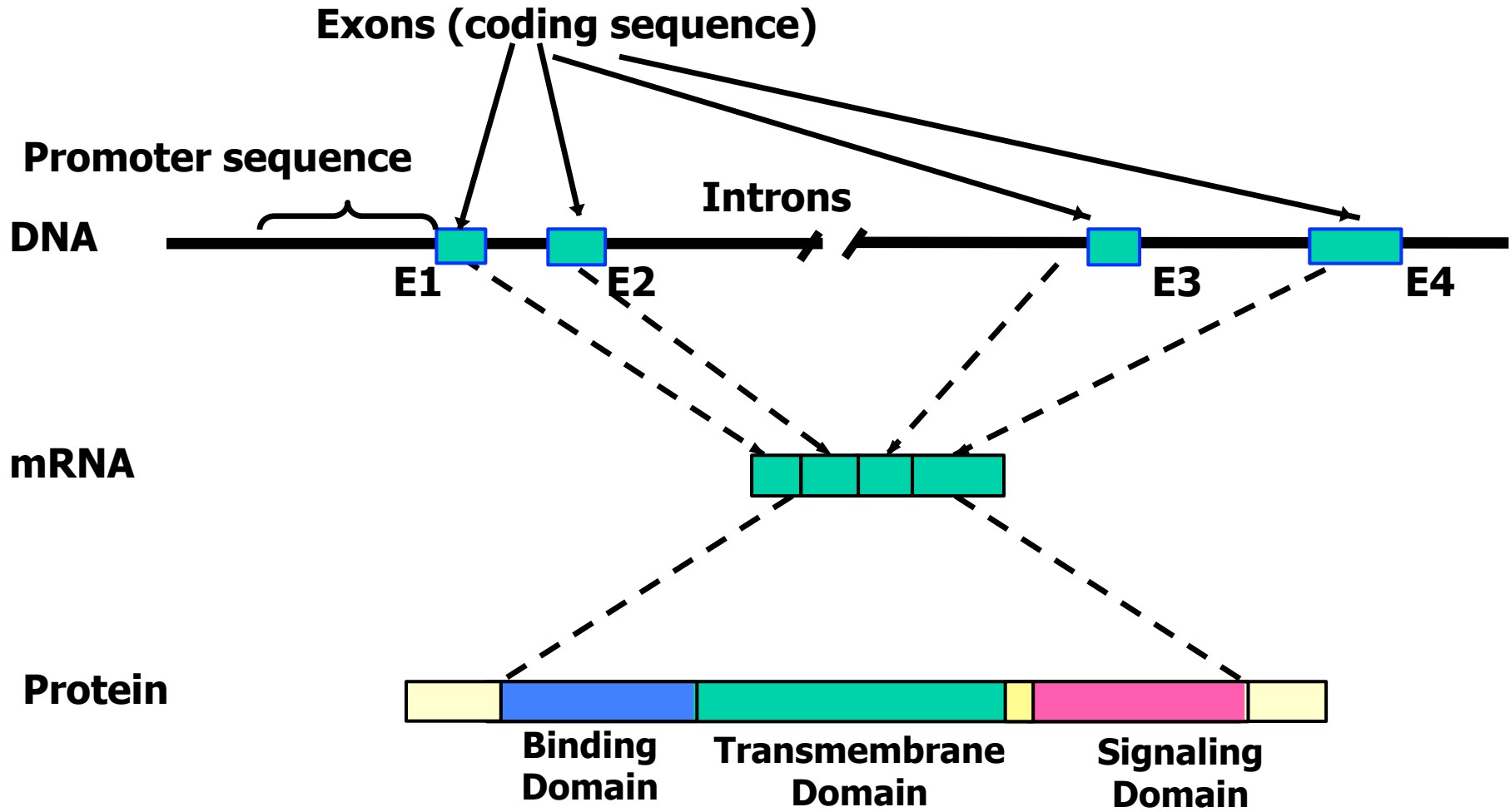


$6 \times 10^9$  base pairs/human cell  
Base pairs are  $.34 \times 10^{-9}$  m apart  
So the length of DNA/cell is  
 $6 \times 10^9 \times .34 \times 10^{-9} = 2$  meters

Each cell has 2 m of DNA  
Average person has 75 trillion cells =  $75 \times 10^{12}$   
Length of DNA in a person =  $150 \times 10^{12}$  m  
Distance from the earth to the sun =  $150 \times 10^9$  m  
Each person has enough DNA to  
go to the sun and back 500 times



# Gene Structure

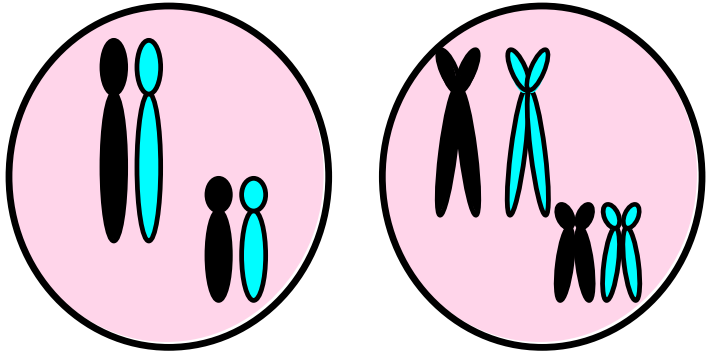




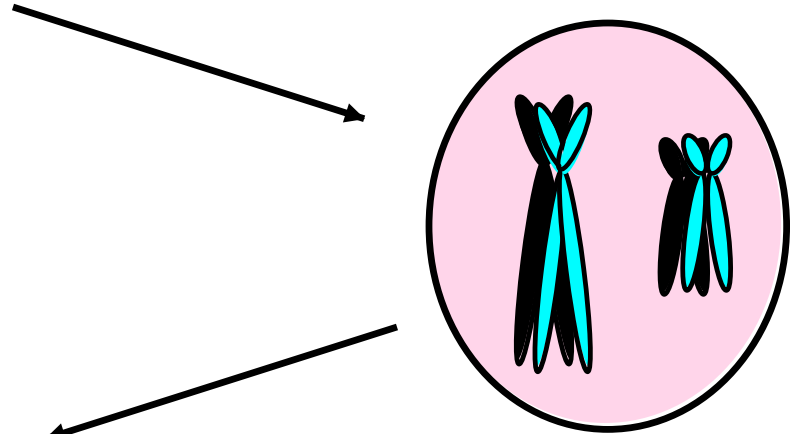
## Diversity in Human Populations



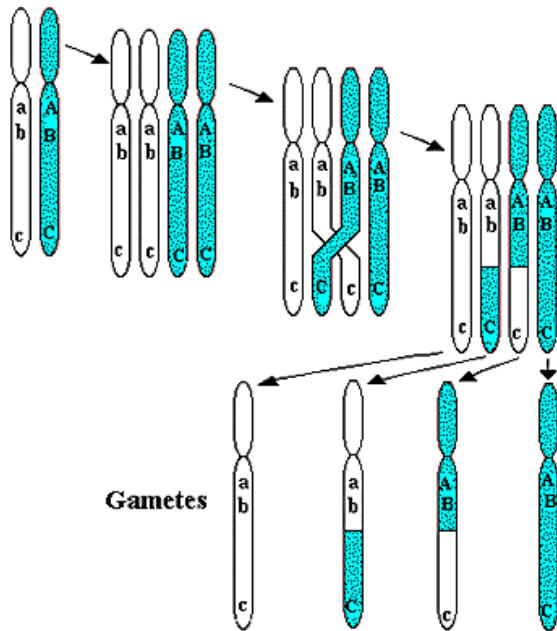
# Meiosis



DNA replication takes place prior to cell division exactly as in mitosis



Prior to cell division in meiosis I, homologous chromosomes align and crossing over takes place



Gametes

Crossing-over and recombination during meiosis

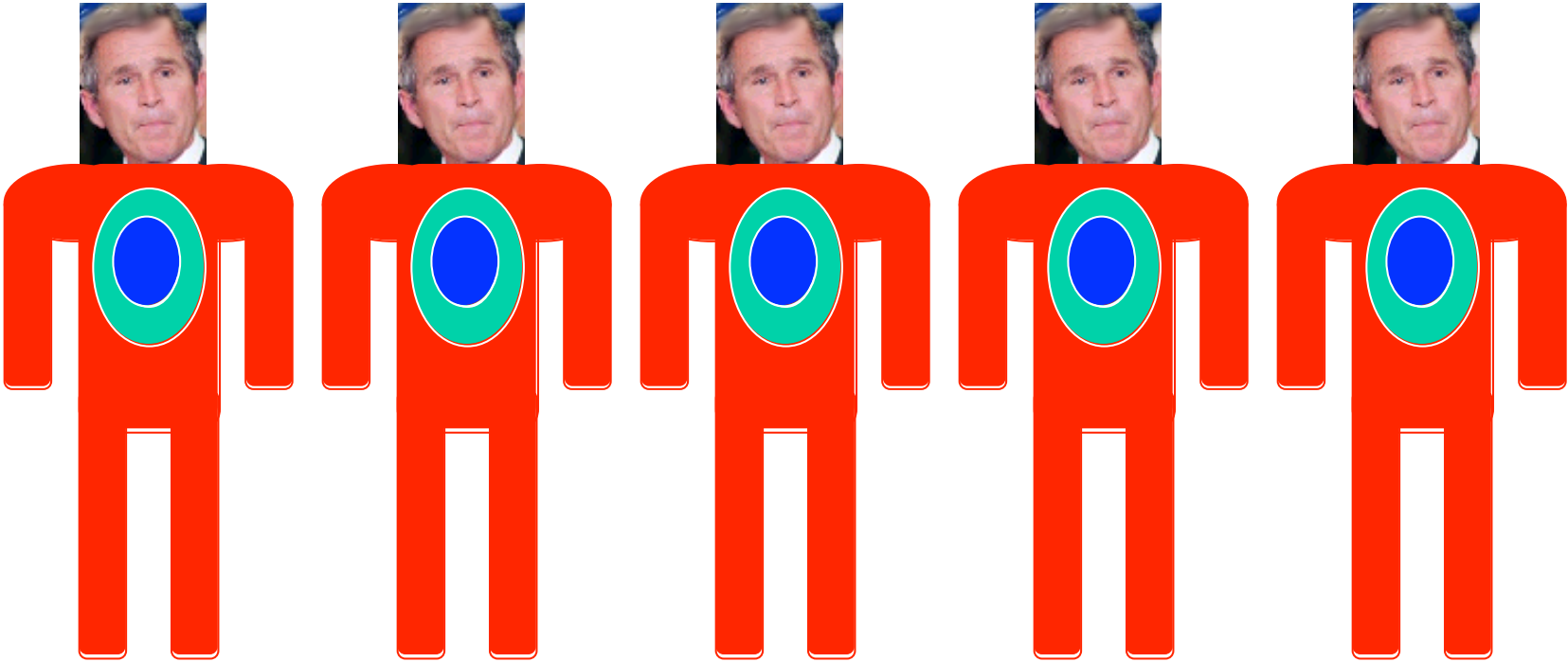
Crossing over involves the physical exchange of DNA between maternal and paternal chromosomes



## Heritability of Facial Structure

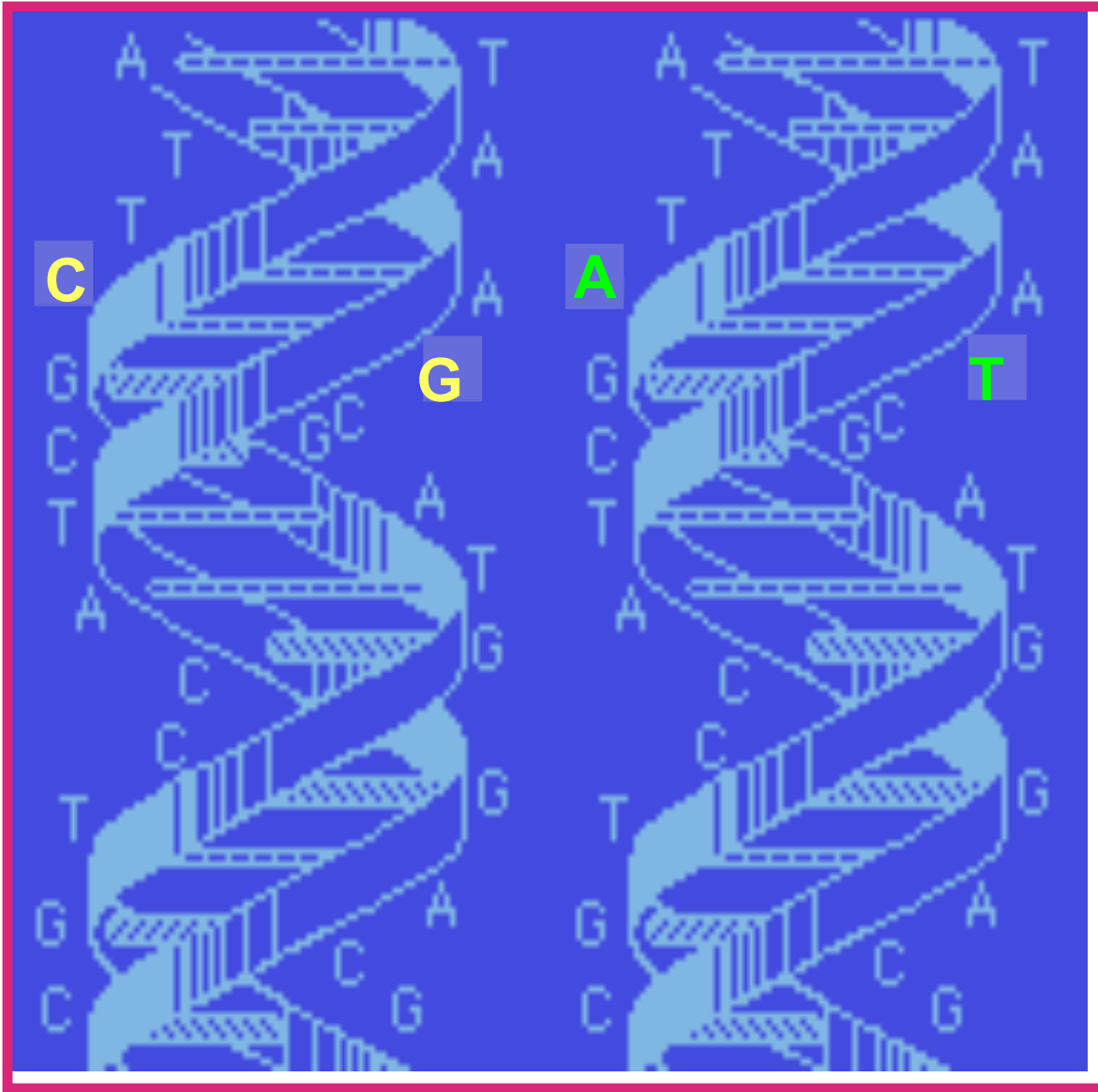


# A World without Genetic Diversity





# DNA Variation



Each form of a DNA sequence variant is called an **allele**.

**Homozygote:**  
2 copies of the same allele

**Heterozygote:**  
2 different alleles

# Types of DNA Sequence Variation

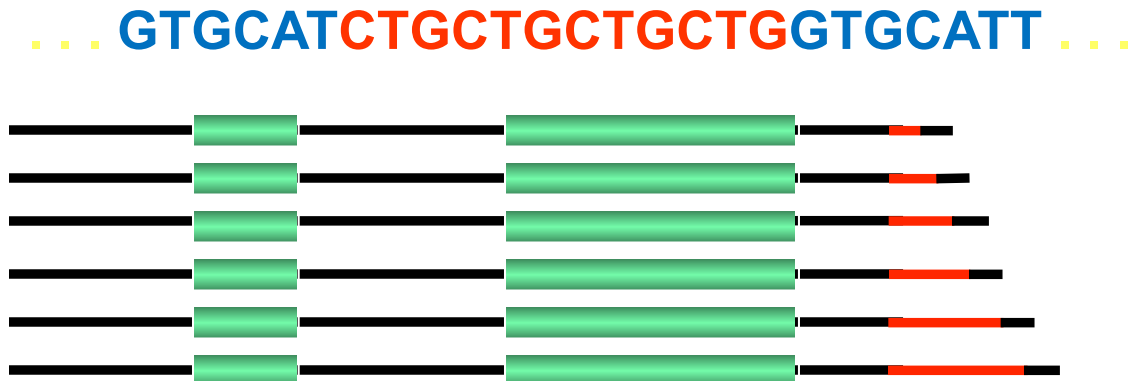
**Single nucleotide polymorphism (SNP)**



**Insertion/Deletion (Copy Number Variant)**



**Microsatellite repeats**

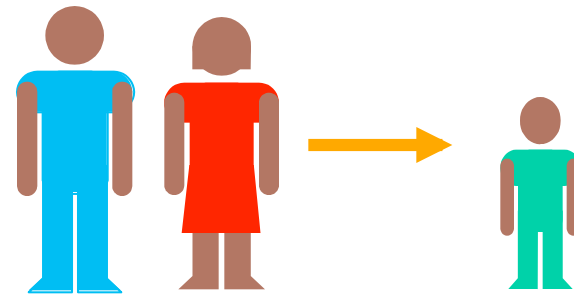


# Causes/Sources of DNA Mutation

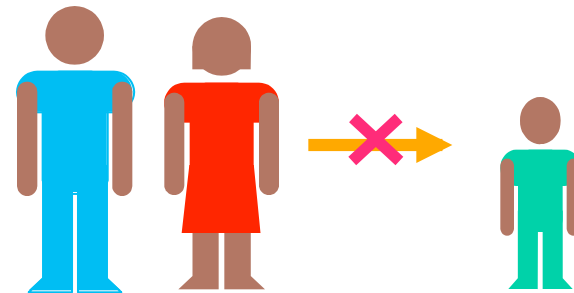
---

- Spontaneous replication error
- Aging
- Mutagenic agents
- Irradiation
- Viruses
- Others?

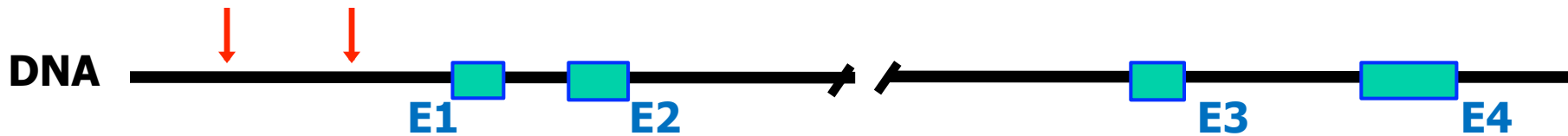
## Germ Cell Mutations



## Somatic Cell Mutations

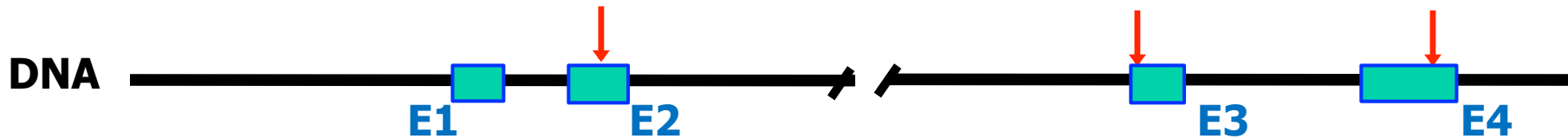


# Mutation/Variation in the Promoter Region



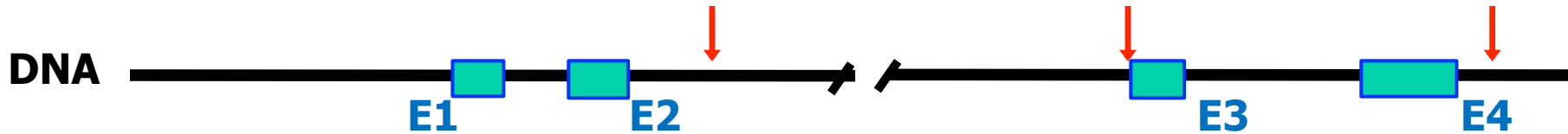
- May prevent or promote transcription factor (TF) binding
- May prevent or promote RNA polymerase binding
- May produce an unstable TF/RNA pol structure
- May alter (reduce or increase) gene expression
- May have no effect

# Mutation/Variation in the Exon (Coding) Region

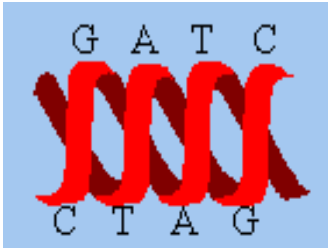


- May alter amino acid sequence
- May produce a shortened protein product
- May produce an unstable protein product
- May produce a non-functional protein by
  - Disrupting active domains
  - Disrupting protein folding
  - Disrupting dimer or other quaternary structure formation
- May have no effect

# Mutation/Variation in the Intronic Region and 3' UTR



- May alter RNA splicing and processing
- May affect stability of the nRNA
- May alter protein product
- May promote degradation of the RNA prior to processing
- May have no effect



AA



Aa



aa

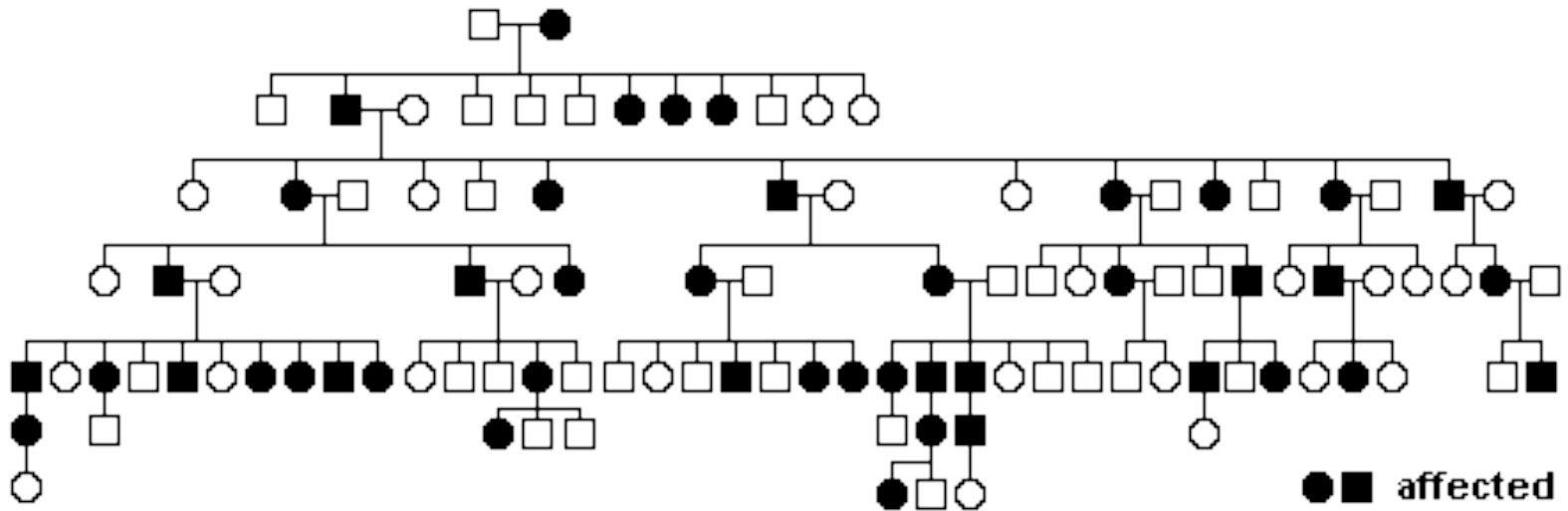


**Autosomal Recessive**  
(skips a generation)

**Autosomal Dominant**  
(observed in every generation)

**Co-Dominant**  
(linear range of phenotypes)

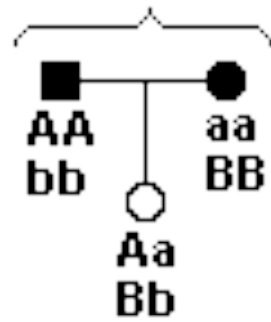
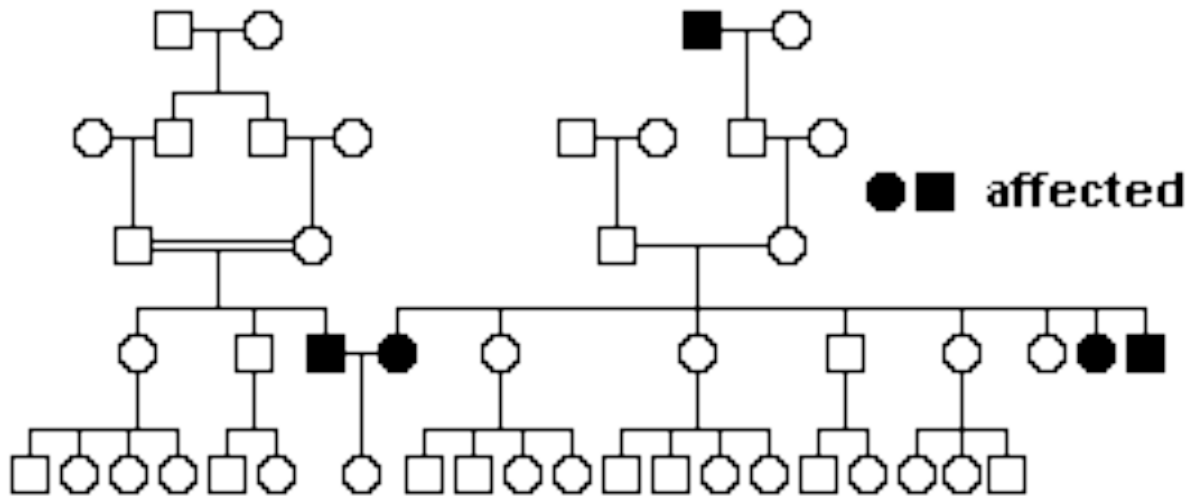

# Brachydactyly (shortened fingers and toes)



**Dominant Inheritance**

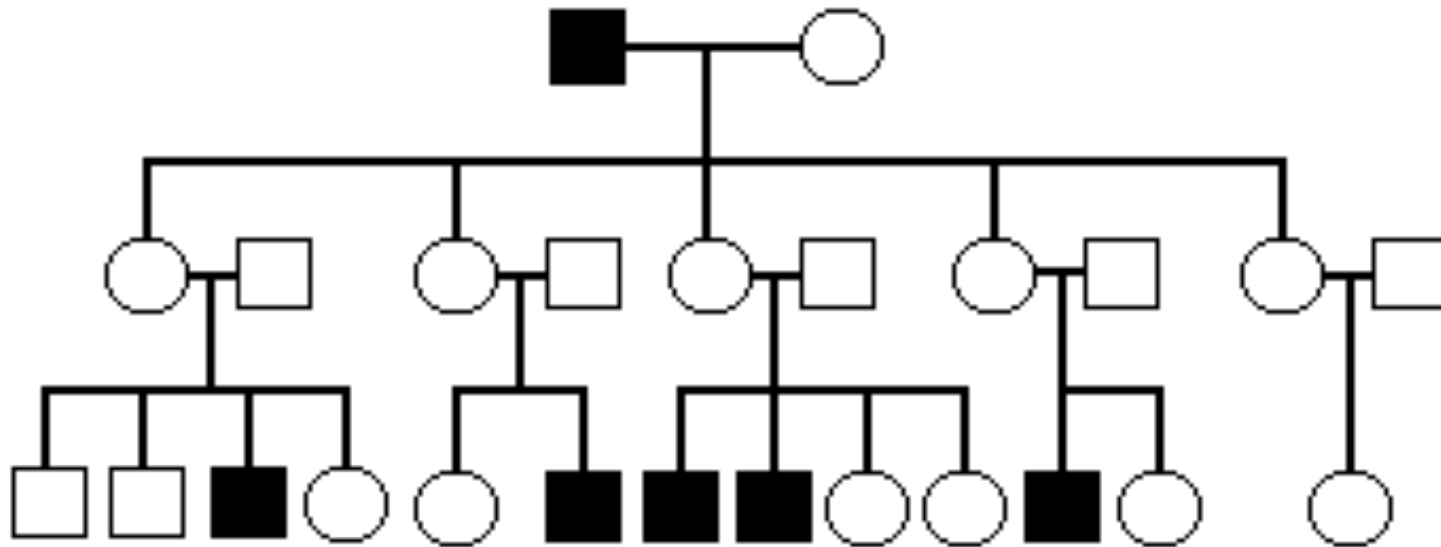


# Albinism (lack of pigmentation)



**Recessive Inheritance**

# Color blindness (inability to see red and green)



**X-linked Inheritance**

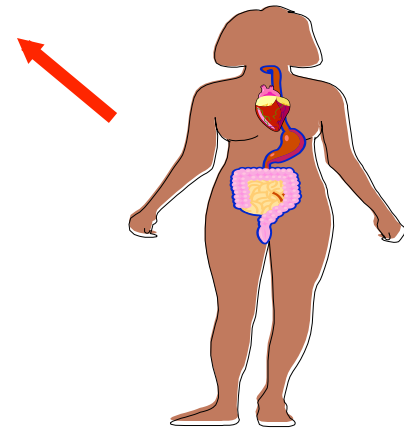
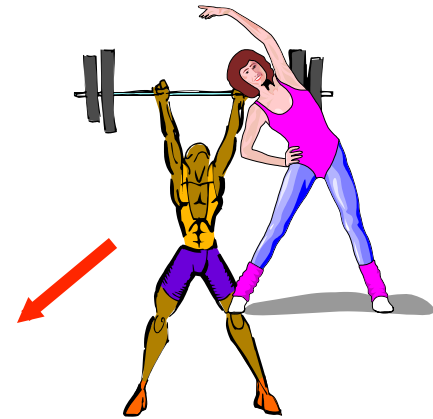
# Single Gene Mutations for Obesity in Humans



<b>Gene</b>	<b>Variant</b>	<b>Phenotype</b>
LEP	<b>G398Del-frameshift Arg105Trp</b>	<b>Early-onset morbid obesity Morbid obesity and hypogonadism Hyperphagis</b>
LEPR	<b>G→A exon 16 splice</b>	<b>Morbid obesity, hyperphagia and hypogonadism</b>
POMC	<b>G7013T C7133Del-frameshift C3804A</b>	<b>Obesity, red hair, ACTH insufficiency, hyperphagia</b>
MC4R	<b>Tyr35Stop CTCT<math>\Delta</math>, nt633 GATT insertion, nt732 Nt47-48, G insertion Codon 279, GT insertion</b>	<b>Absence of MC4R activity, early onset morbid obesity, hyperphagia</b>

# Mapping DNA Variation to Complex Phenotype

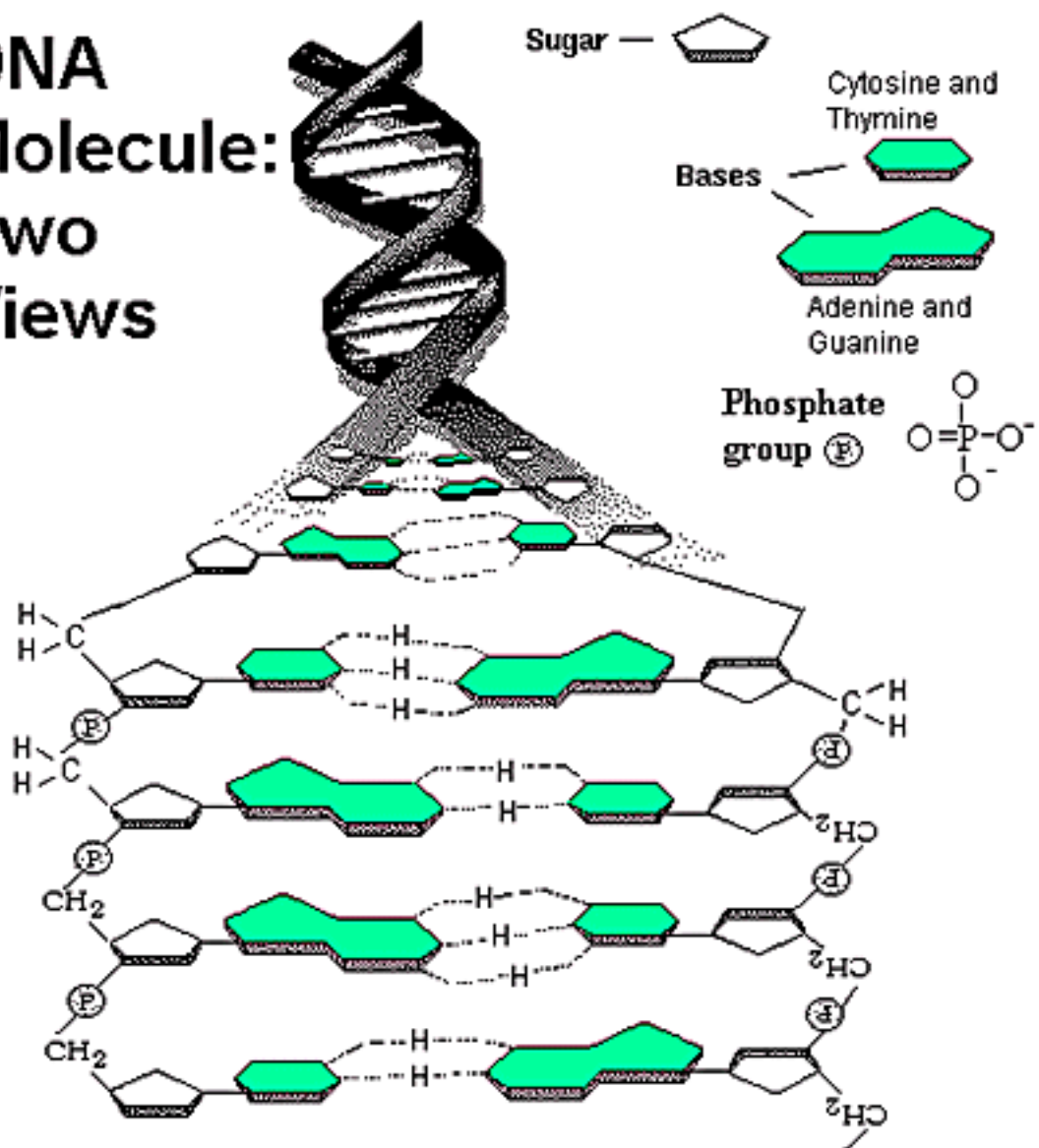
....TGCAT <sup>C</sup> TCATT....  
                  T



# Genotyping: Characterizing DNA Sequence Variation



# DNA Molecule: Two Views



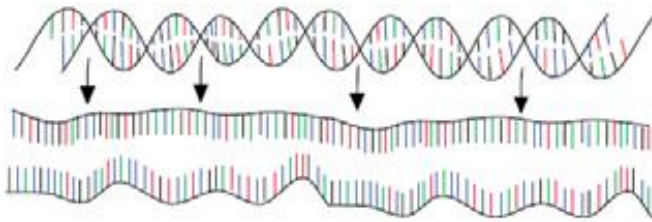
# Genotyping Using PCR

## PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :

### Step 1 : denaturation

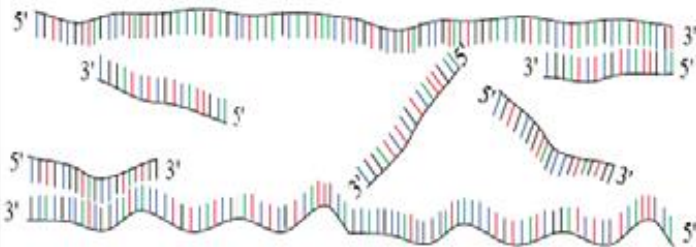
1 minut 94 °C



### Step 2 : annealing

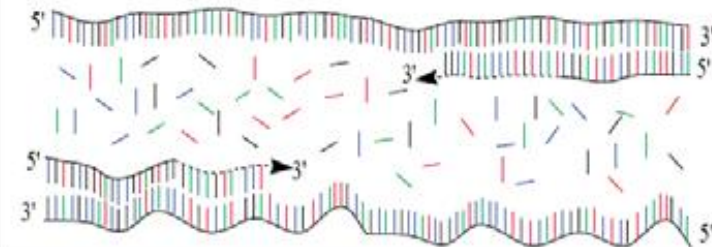
45 seconds 54 °C

forward and reverse primers !!!

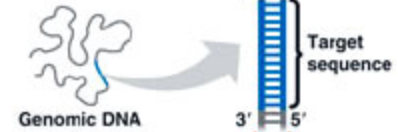


### Step 3 : extension

2 minutes 72 °C  
only dNTP's



(Aady Vierstraete 1999)



1 Denaturation:  
Heat briefly to separate DNA strands

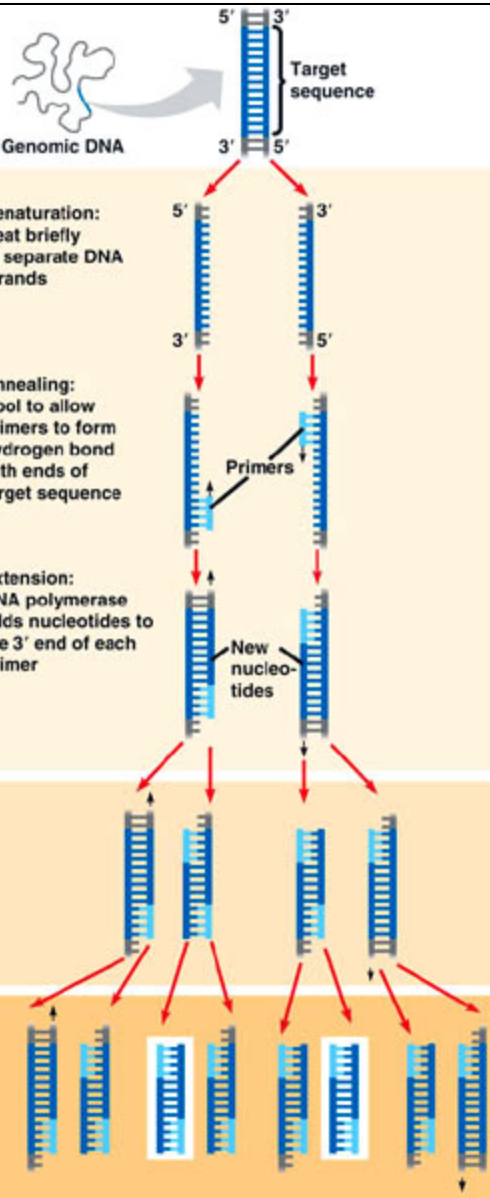
2 Annealing:  
Cool to allow primers to form hydrogen bond with ends of target sequence

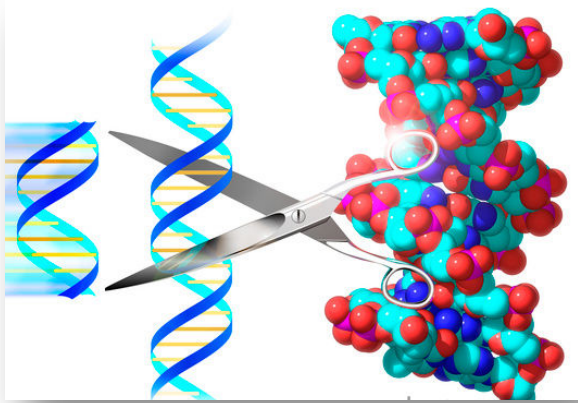
3 Extension:  
DNA polymerase adds nucleotides to the 3' end of each primer

Cycle 1 yields 2 molecules

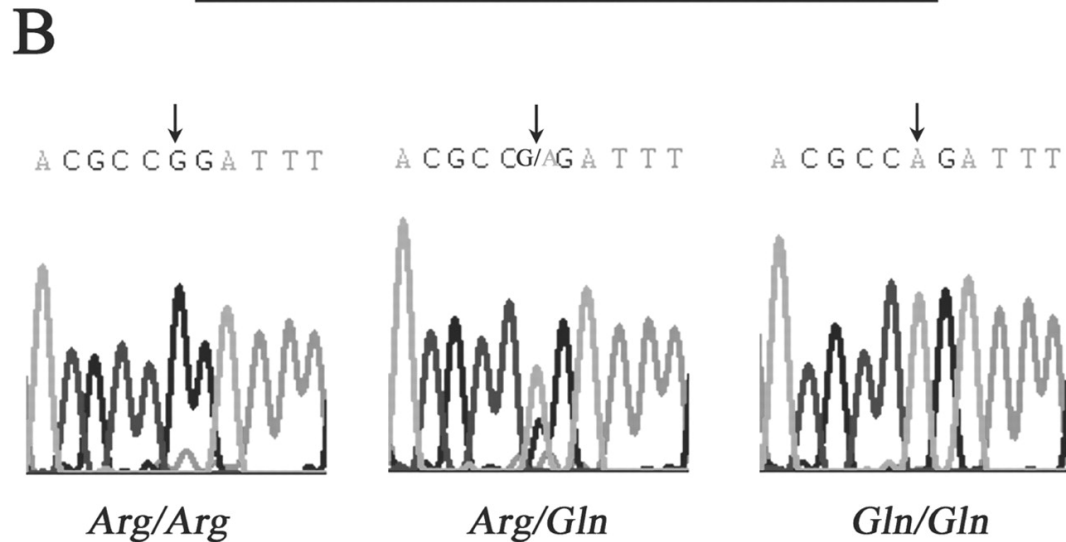
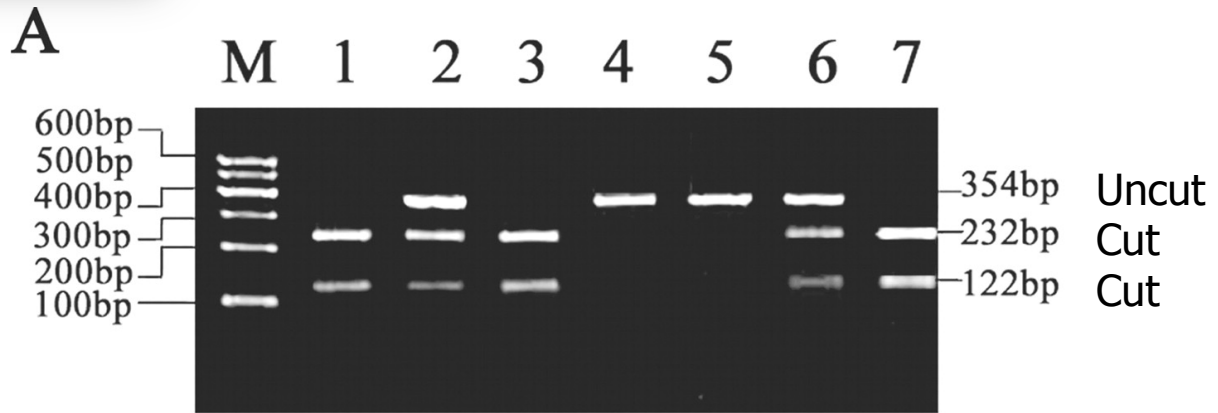
Cycle 2 yields 4 molecules

Cycle 3 yields 8 molecules; 2 molecules (in white boxes) match target sequence



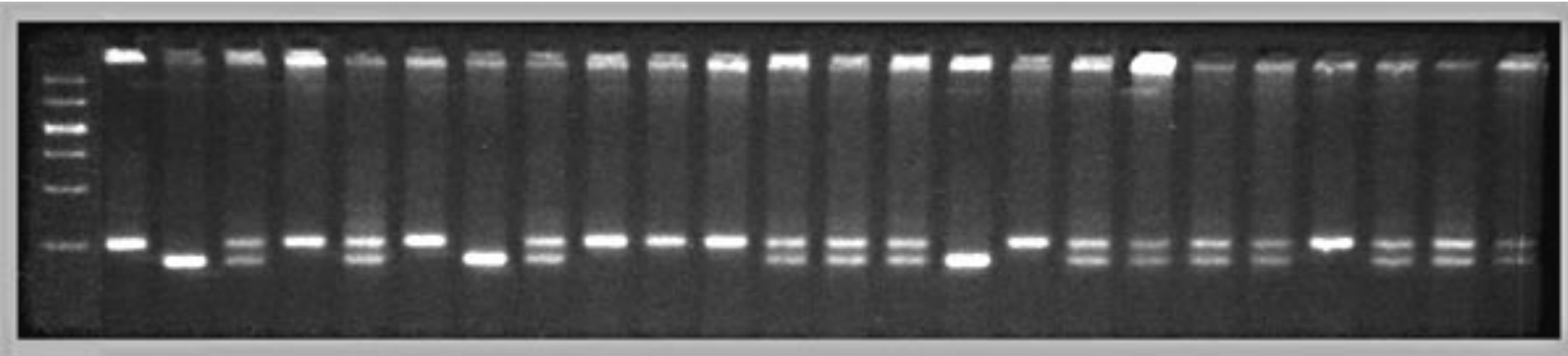


**PCR product  
cut with *Msp*I  
restriction  
enzyme:**



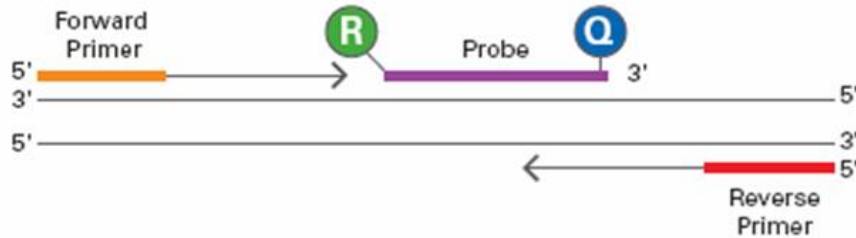


# Microsatellite Genotyping



Genotypes at a microsatellite locus on chromosome 5 (D5Mit294) include bands at 198 and 176 bp

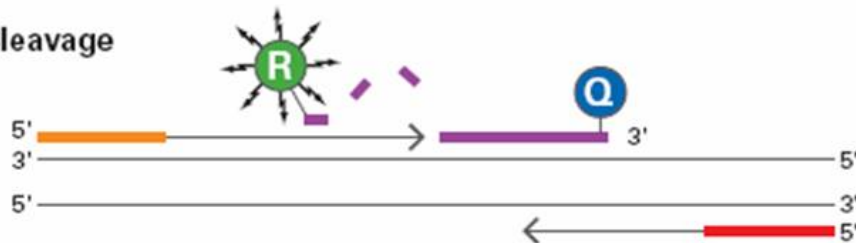
## Polymerization



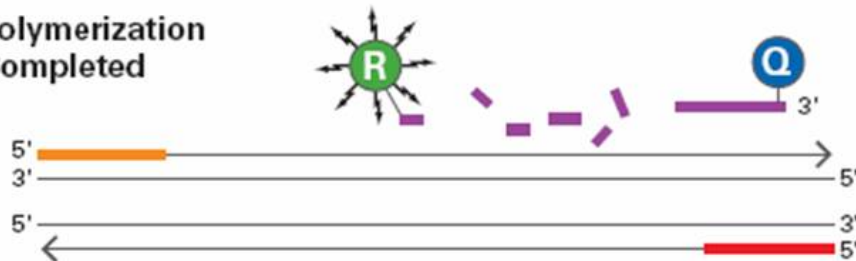
## Strand Displacement



## Cleavage



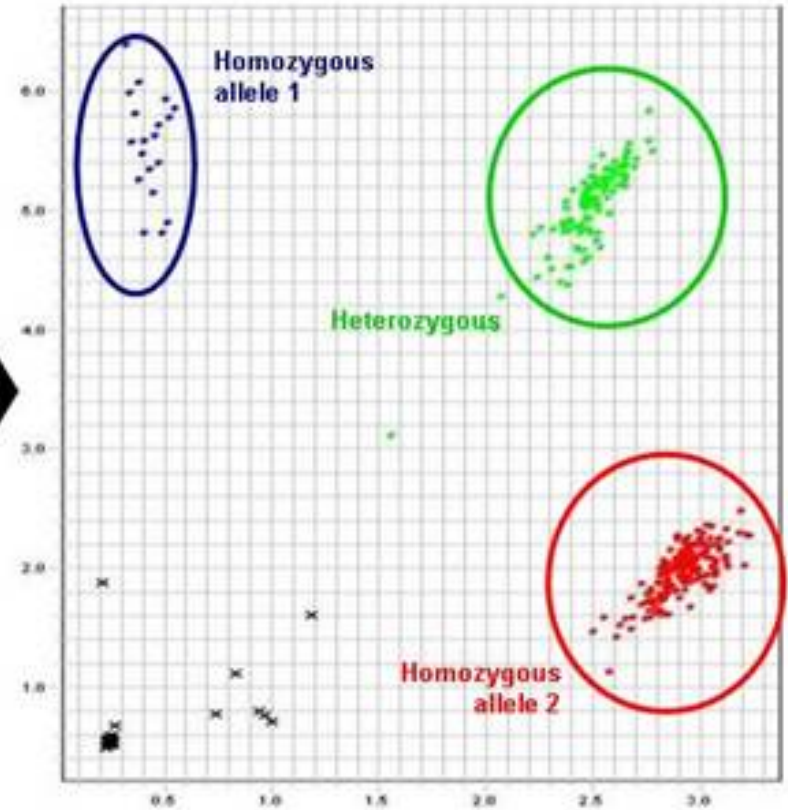
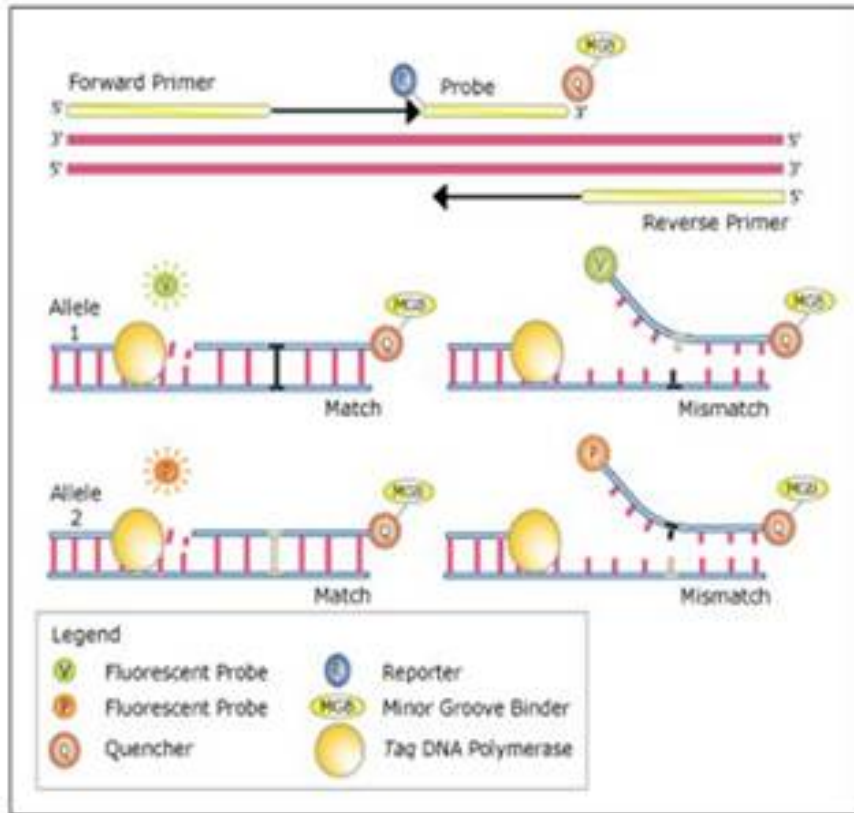
## Polymerization Completed



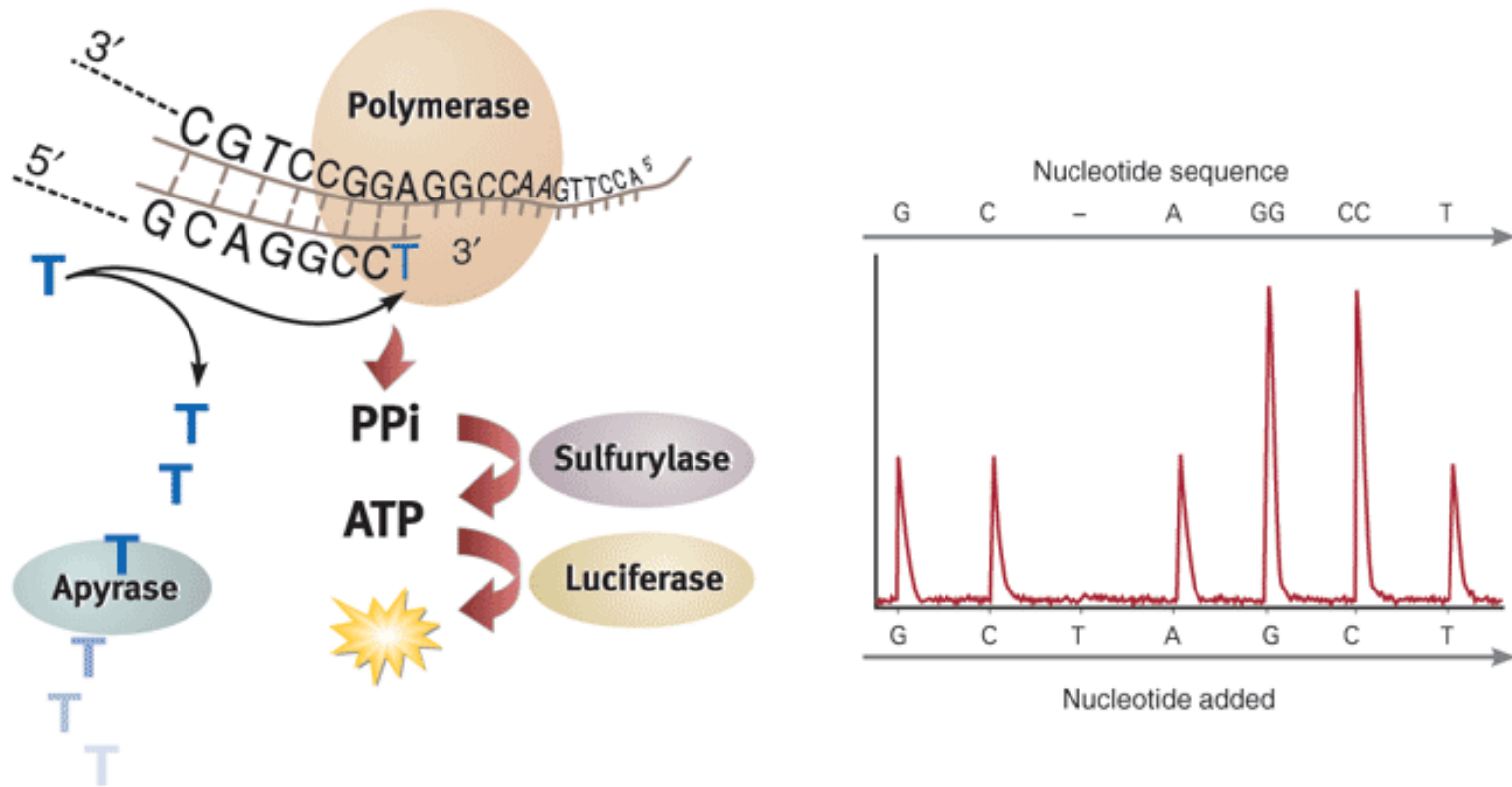
# The Taqman Assay

- PCR primers bind to target region
- Allele-specific oligonucleotide probes, labeled at each end with either a fluorescent dye or a quencher molecule, bind to the variant site
- Proofreading polymerase degrades the oligonucleotide probe, releasing the dye molecule resulting in a fluorescent signal
- PCR reaction is repeated

# Genotype Calling with the Taqman Assay

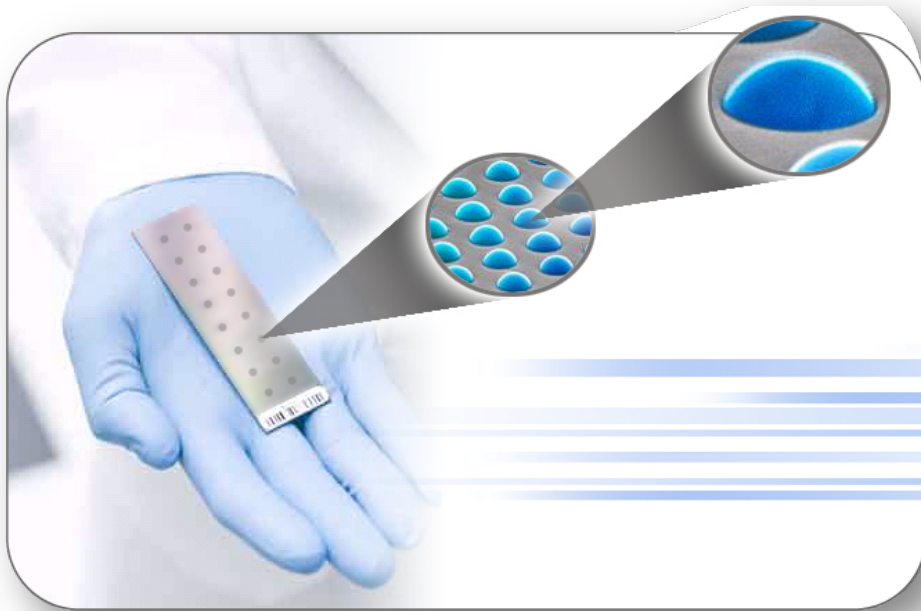


# Genotyping with Pyrosequencing



**Figure 1** | The principle of Pyrosequencing and the output Pyrogram™. Double peak heights indicate incorporations of two nucleotides in a row.

# ILLUMINA'S CORE TECHNOLOGIES

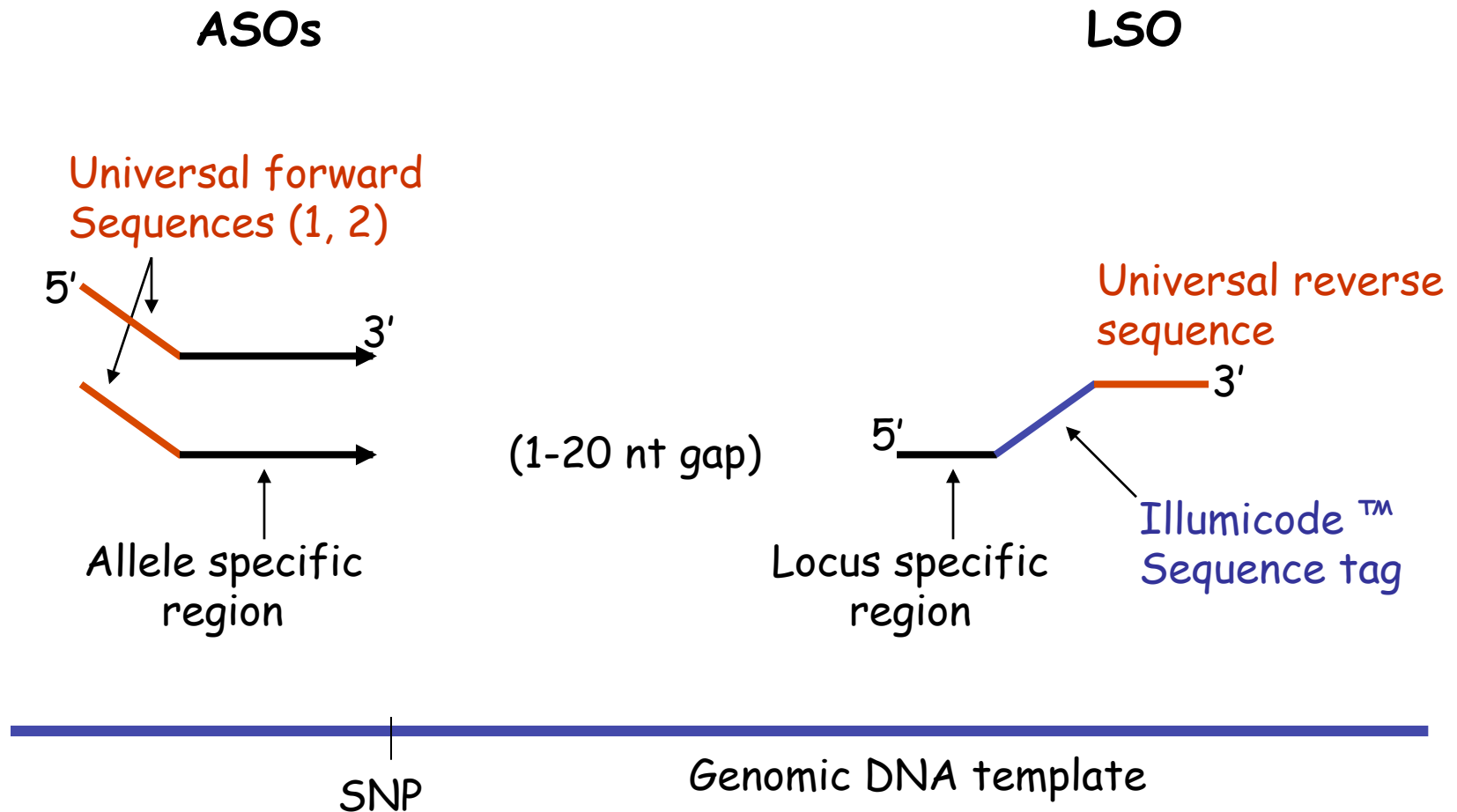


**BeadArray**

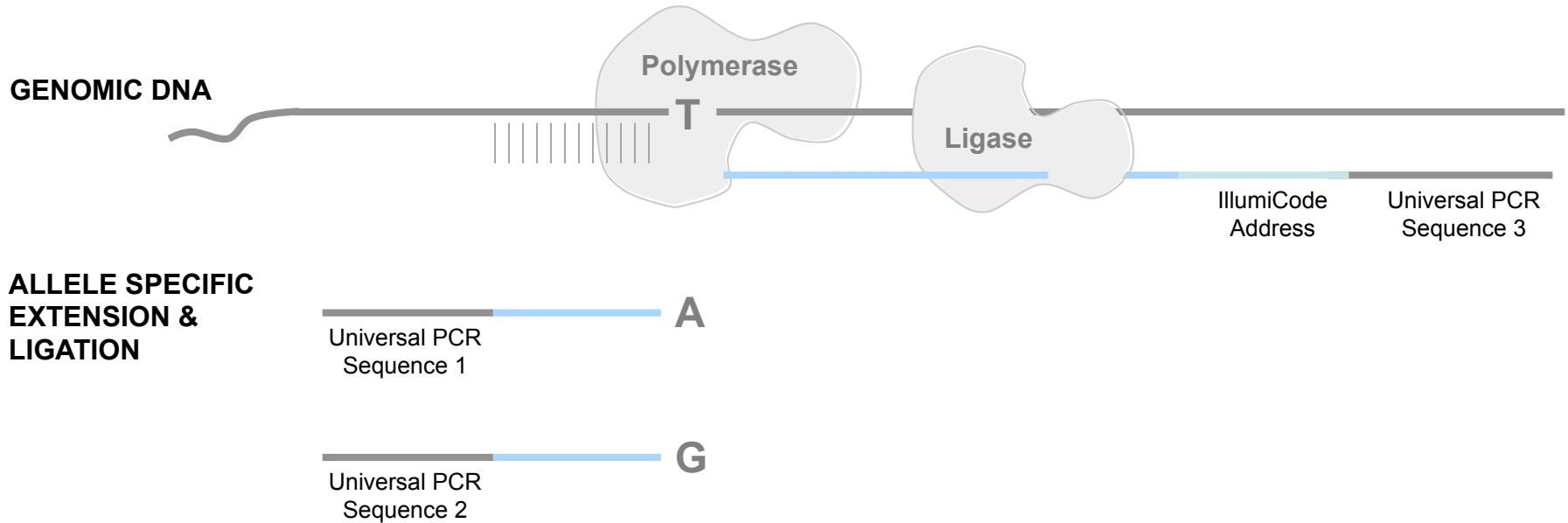
**Veracode**



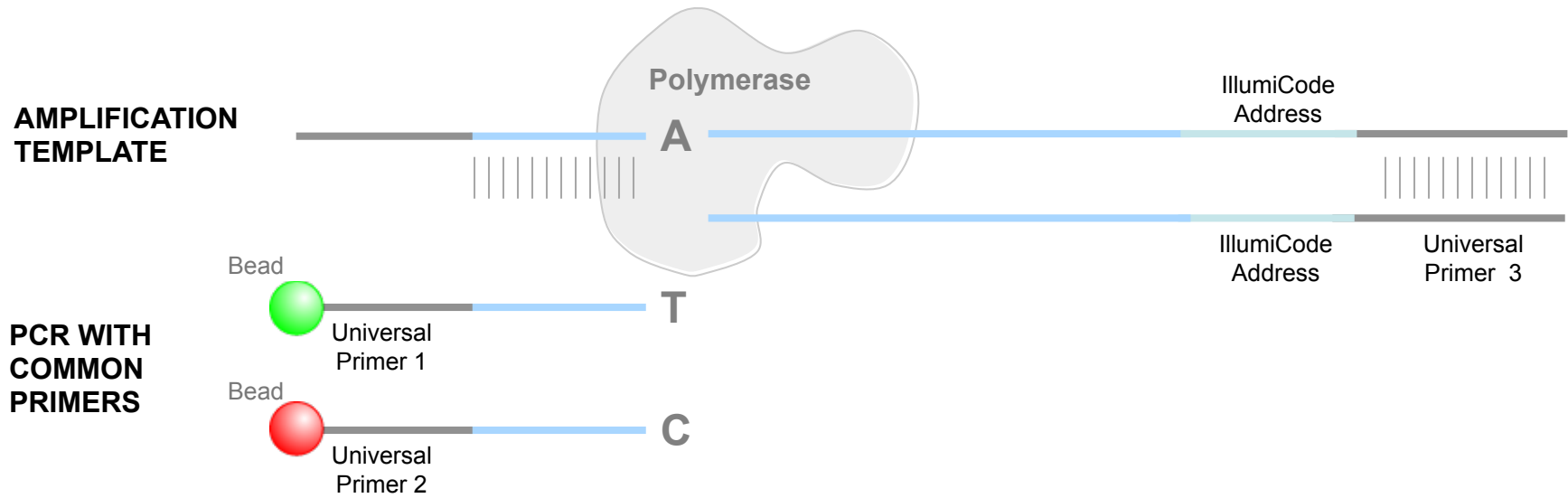
# High Throughput Genotyping (48-1536 SNPs): GoldenGate and VeraCode Assays



# Allele Specific Extension and Ligation

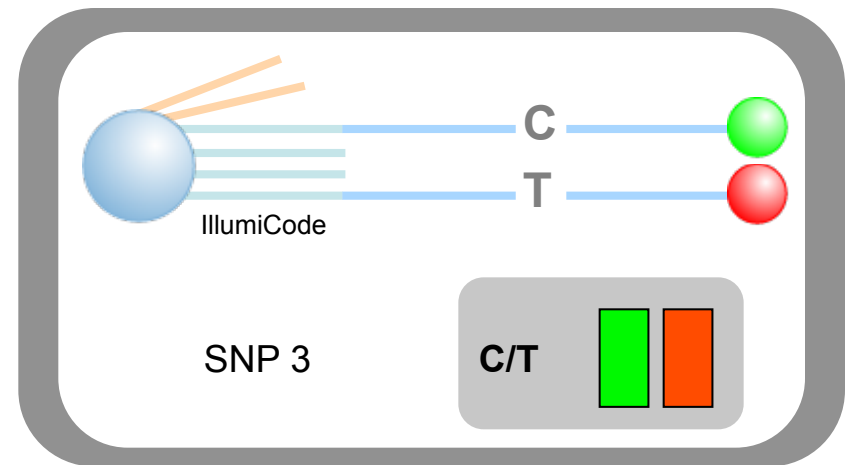
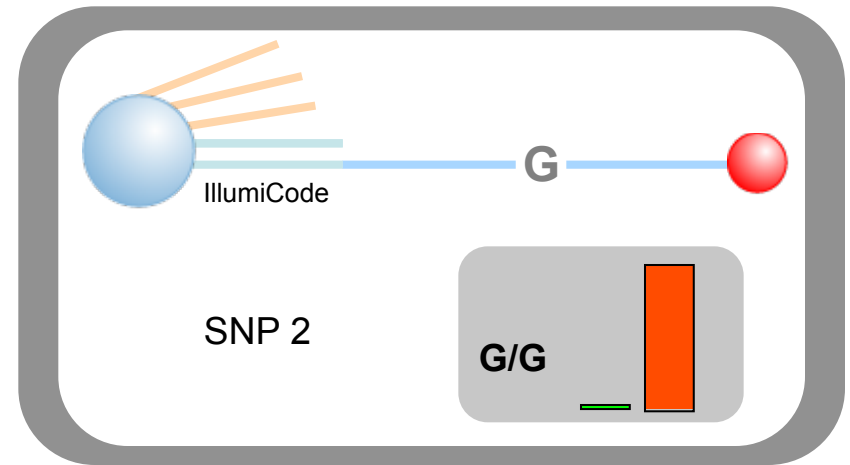
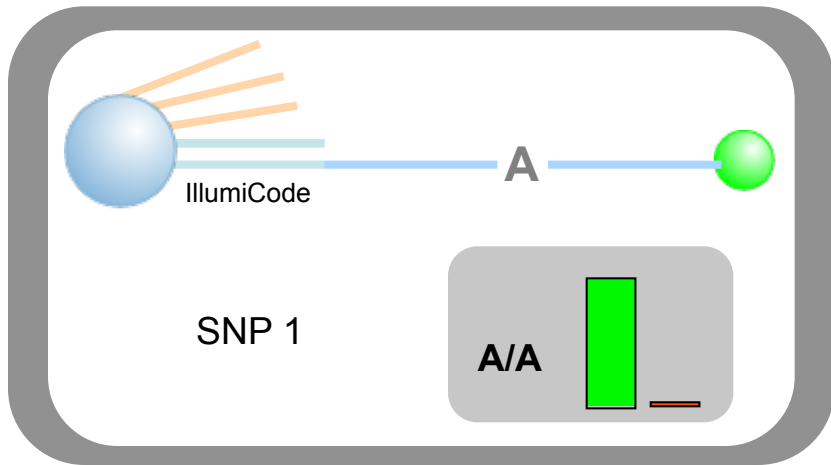


# Amplification





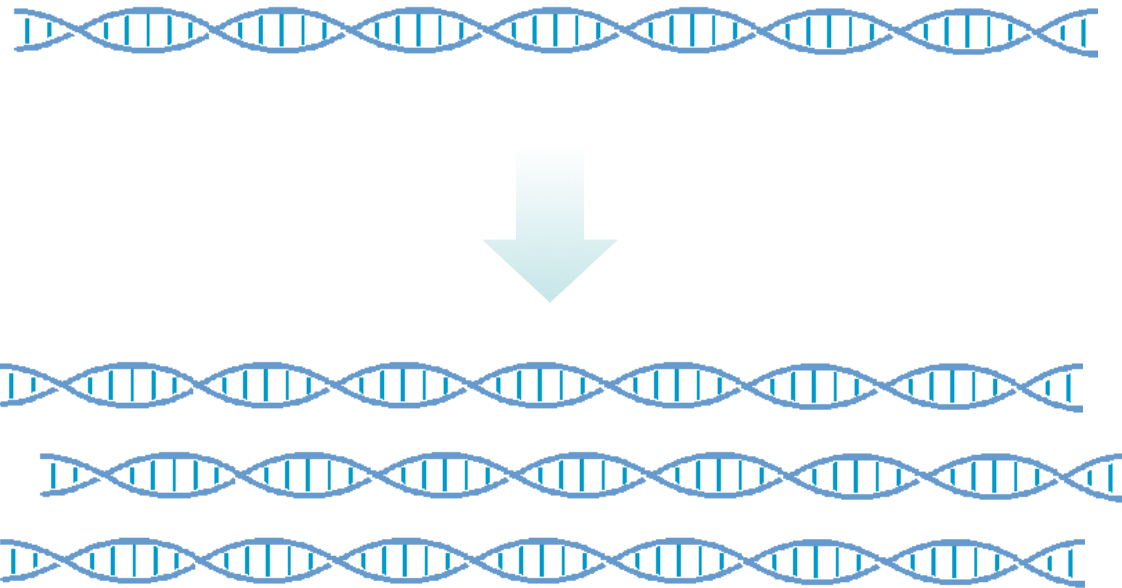
# Hybridization to Universal BeadArray or VeraCode Pool



- Millions of specific probes
- Average 30-fold redundancy

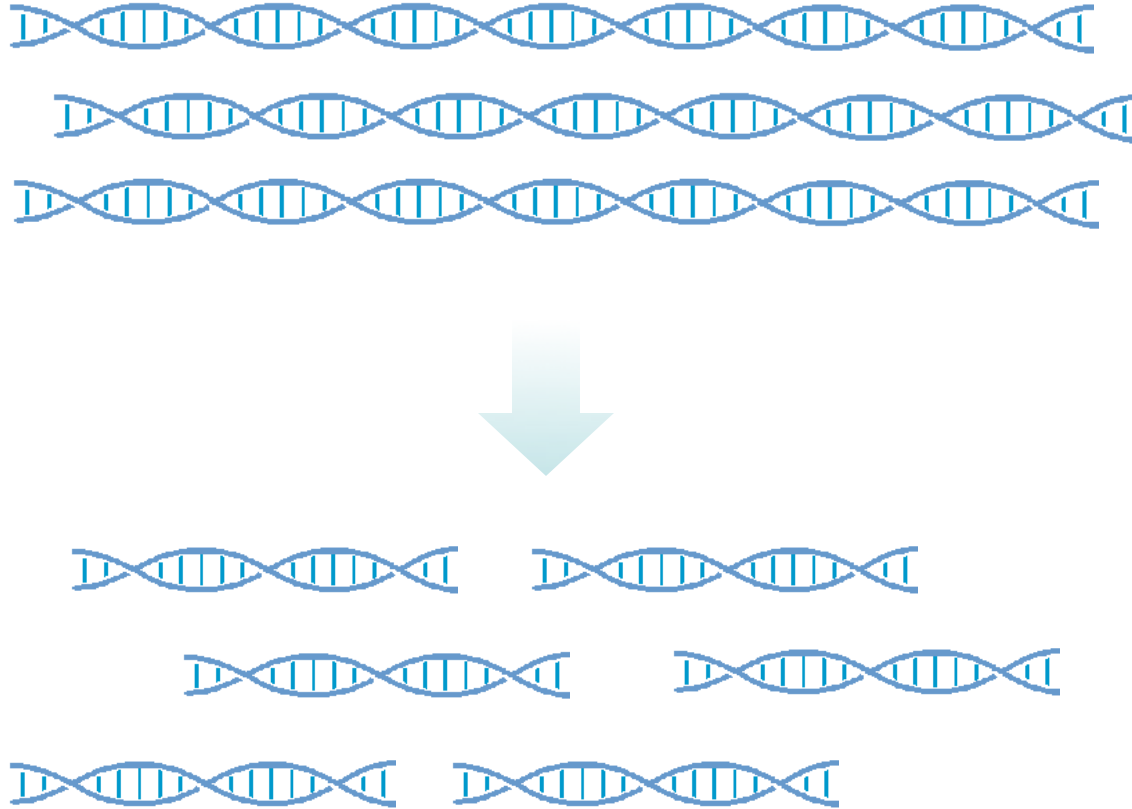
# Infinium Assay

## Very High Density (200K-5M SNPs)



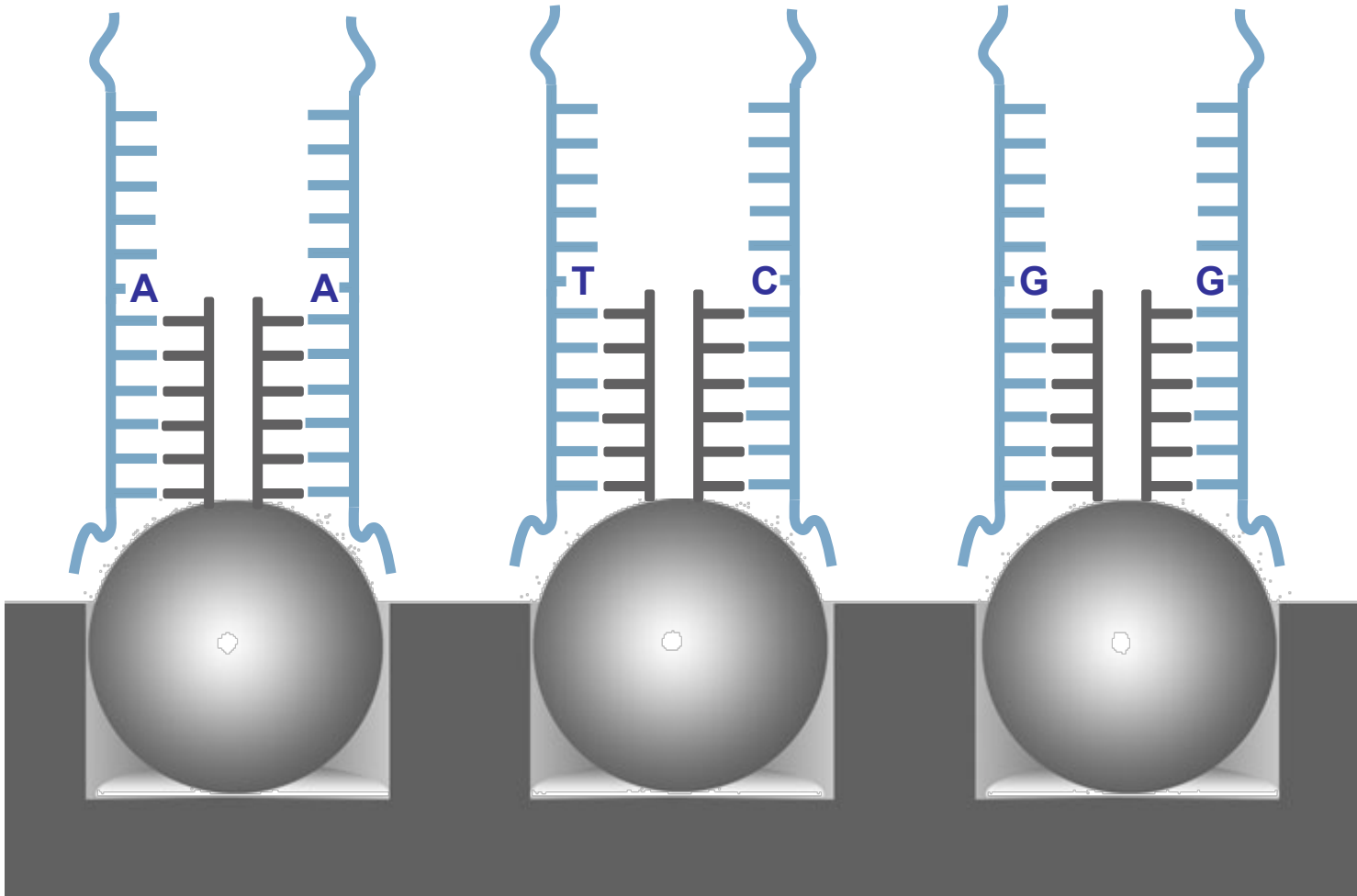
- **Optimized whole genome amplification reaction reduces GC bias**

# Fragmentation

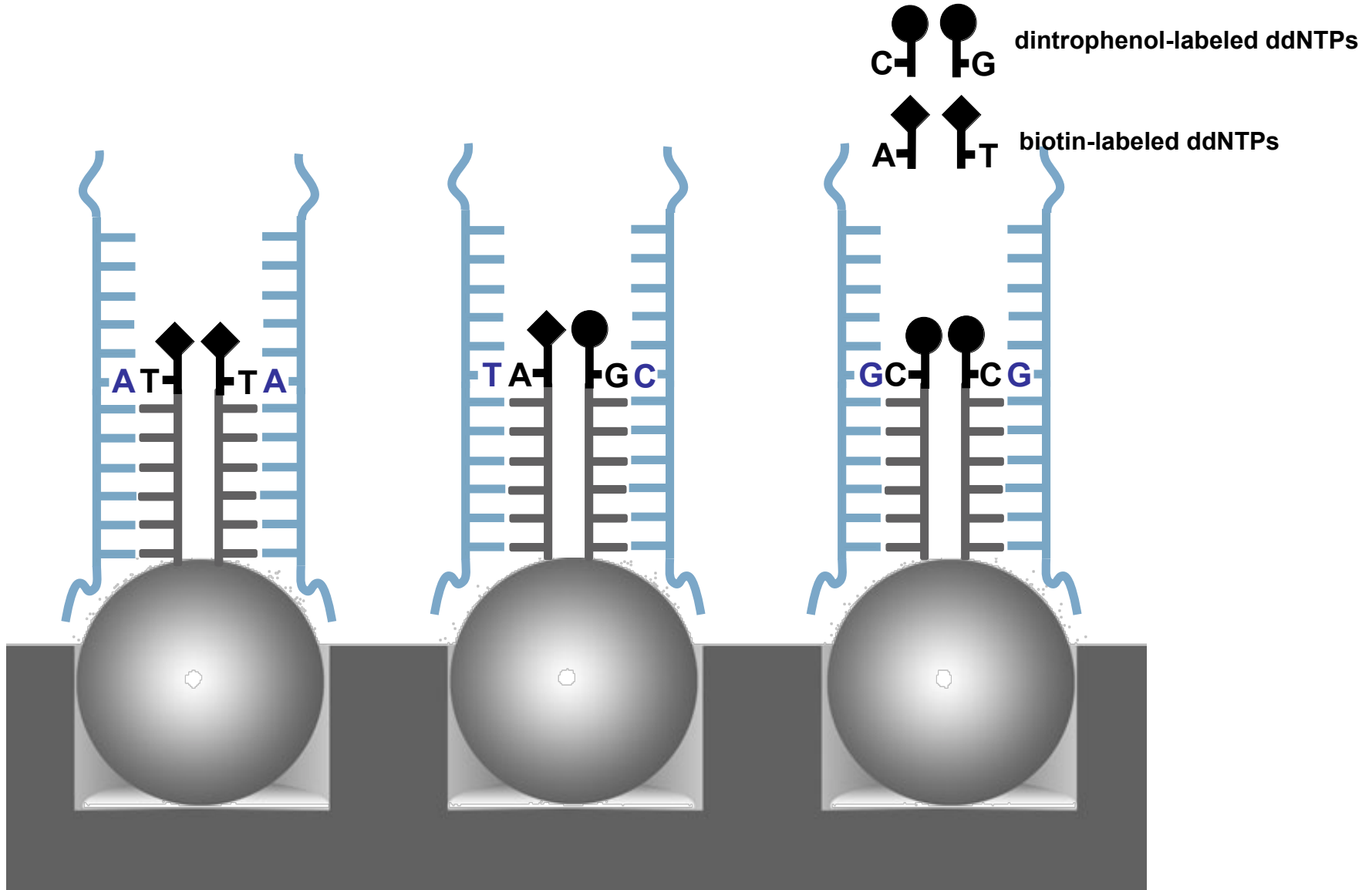


- Robust endpoint fragmentation
- Allows access to vast majority of genome

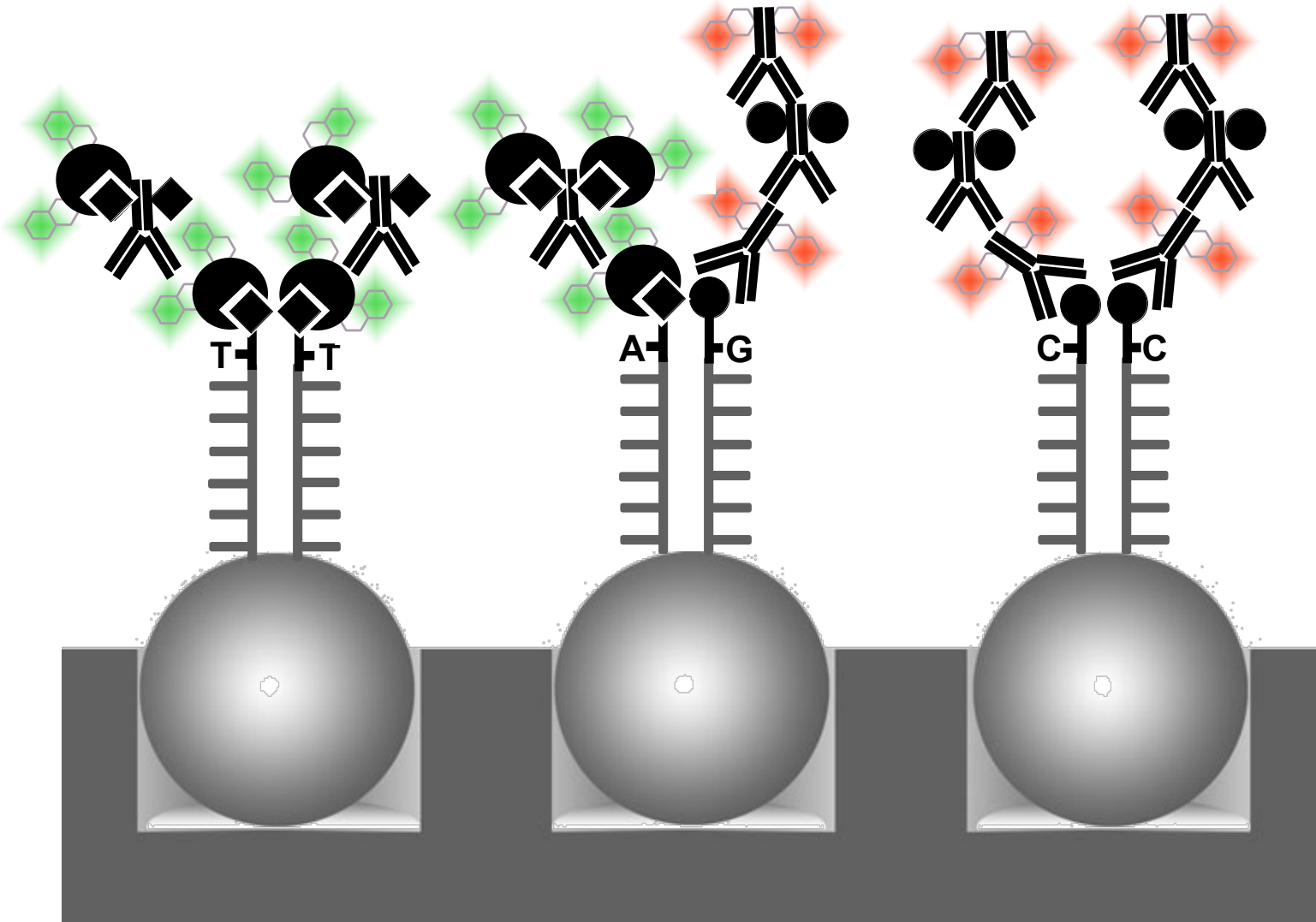
# Hybridization




# Single Base Extension



# Staining



  
streptavidin-  
green

  
anti-DNP-  
red

  
anti-streptavidin-  
biotin

  
anti-Ab-  
DNP

# Image

