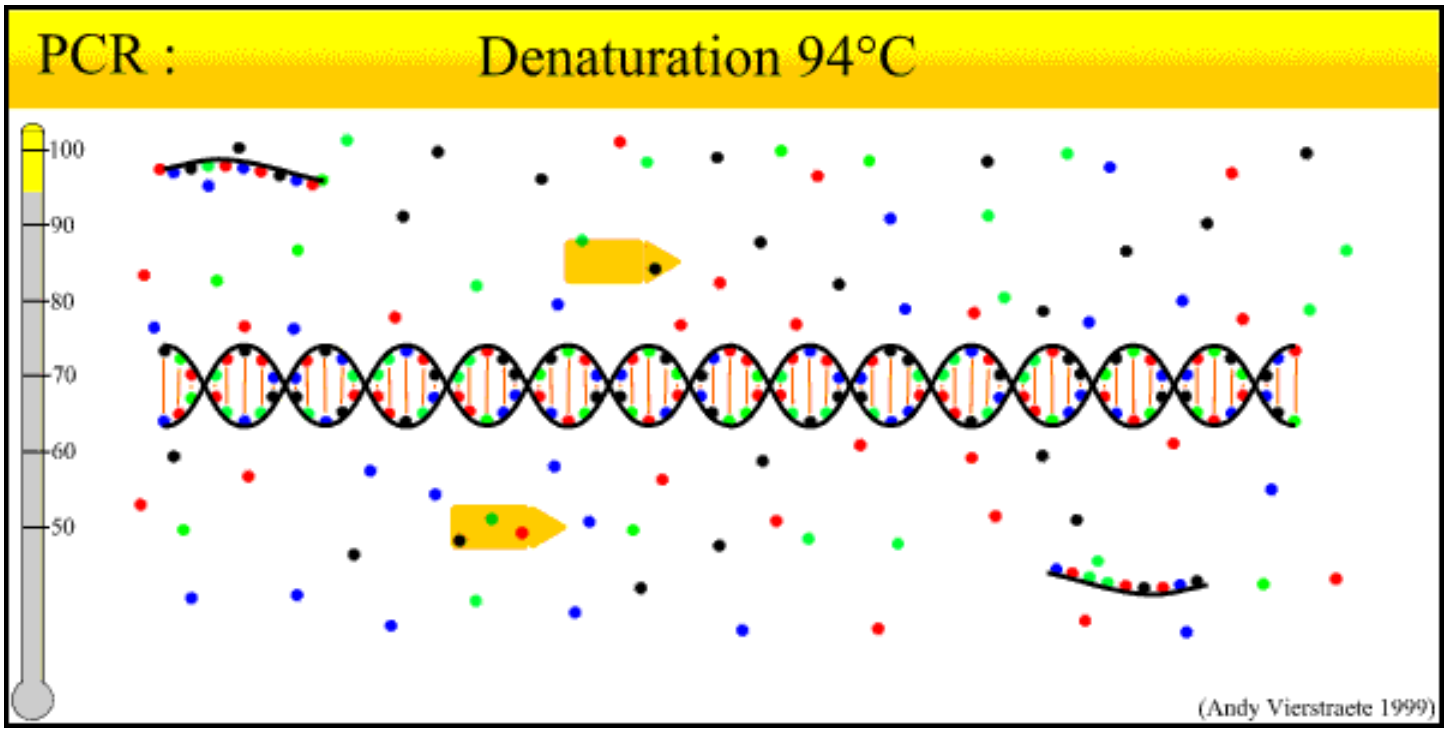
Course :Genomics

IDENTIFICATION OF DISEASE GENES

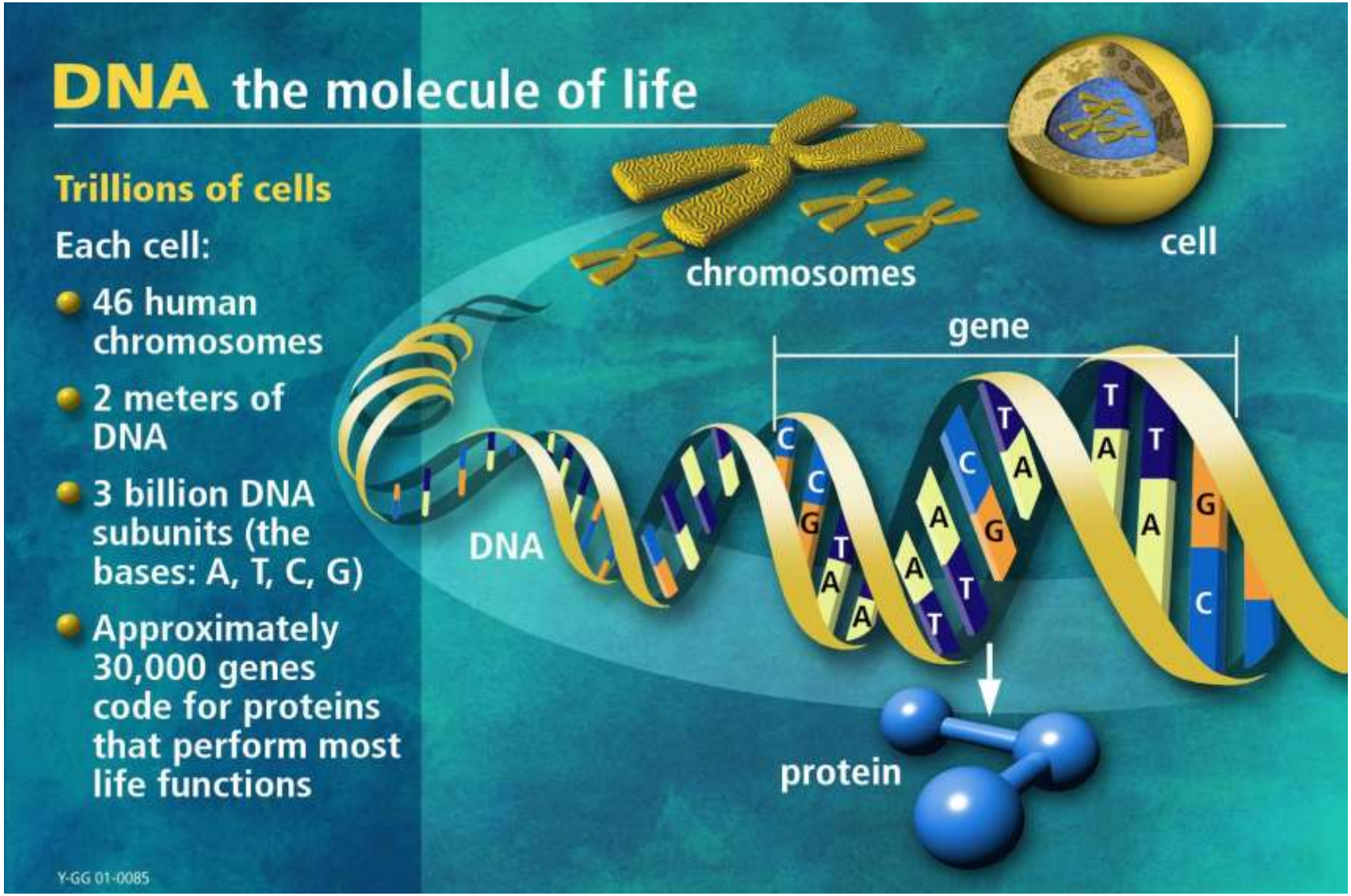
Dr. K. Premkumar
Associate Professor
Dept of Biomedical Science
Bharathidasan University

PCR Animation



Denaturation: DNA melts
Annealing: Primers bind
Extension: DNA is replicated

The Human Genome



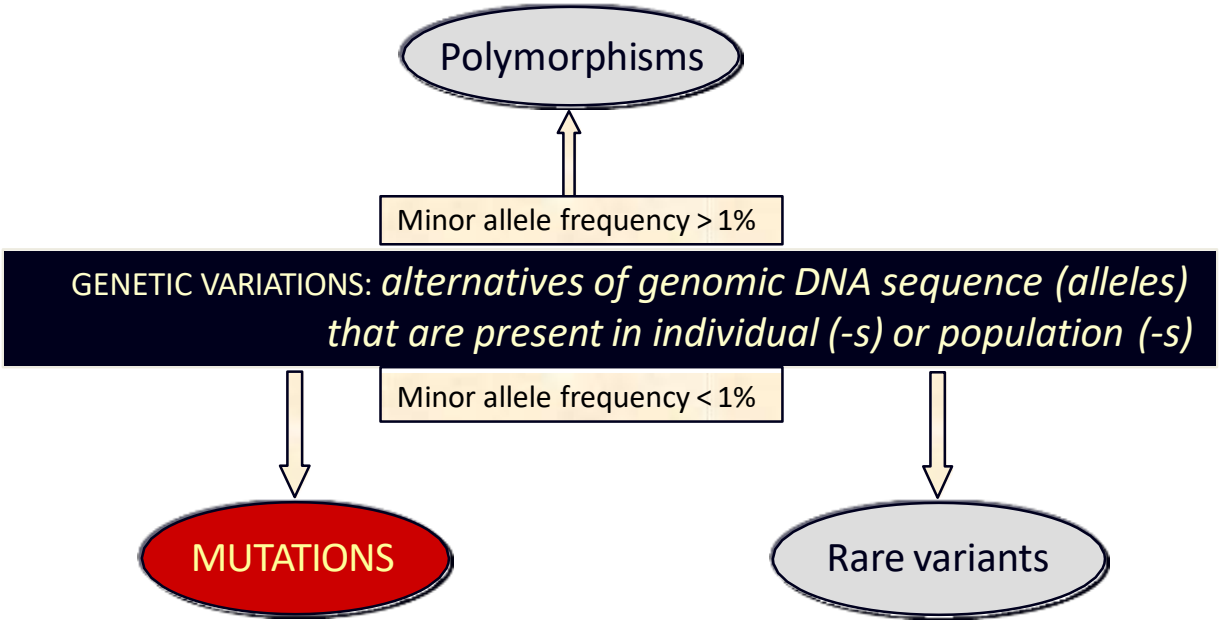
Genetic Polymorphisms

- Polymorphisms (common variation): majority – neutral
- The rest:

✓ slightly "bad" (predispose to disease)

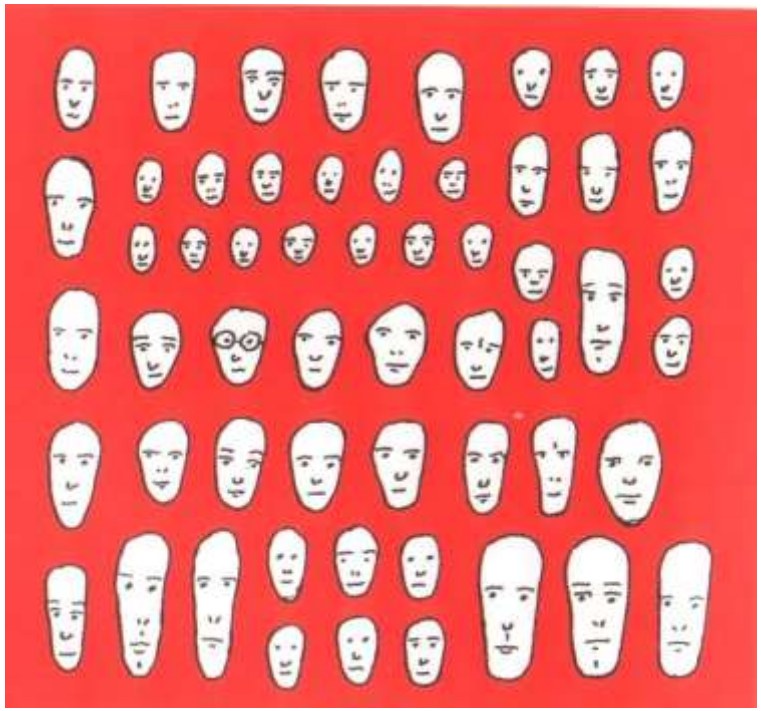
✓ slightly "good" (protect from disease)

✓ both slightly bad and good (predispose to and protect from certain conditions)



Genetic Variability

- Population is monomorphic at a locus - only one allele at the locus.
- Population is polymorphic at a locus - two or more alleles coexist in the population.

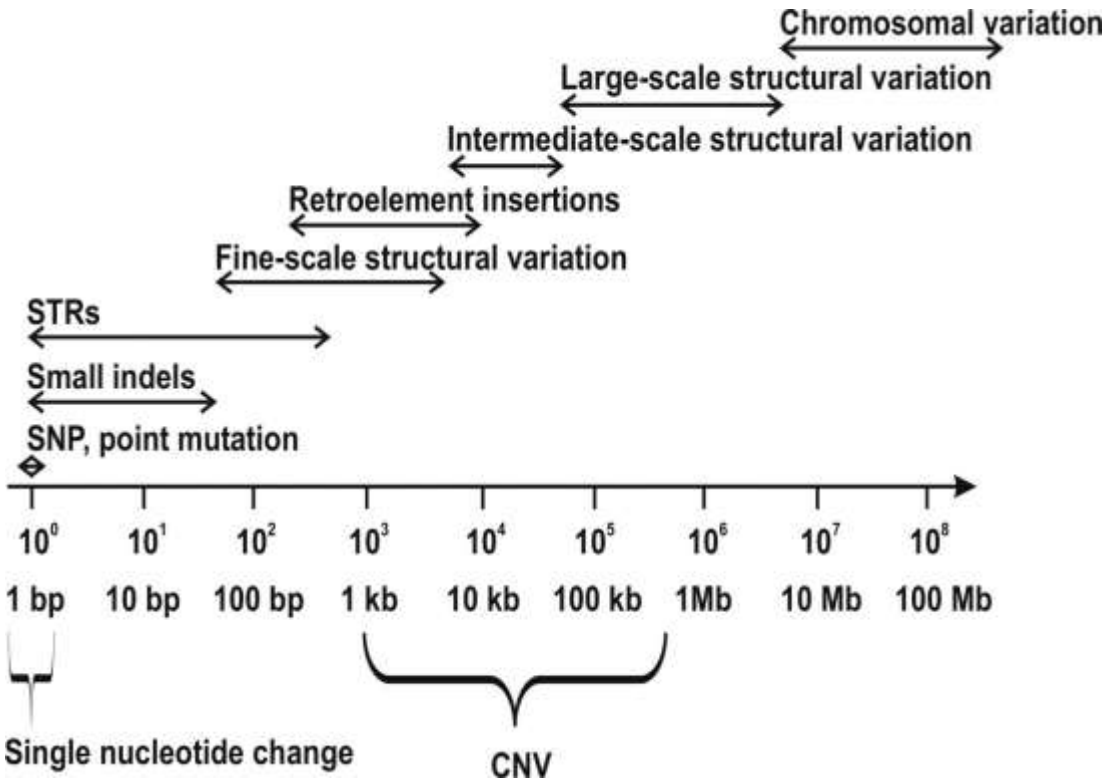


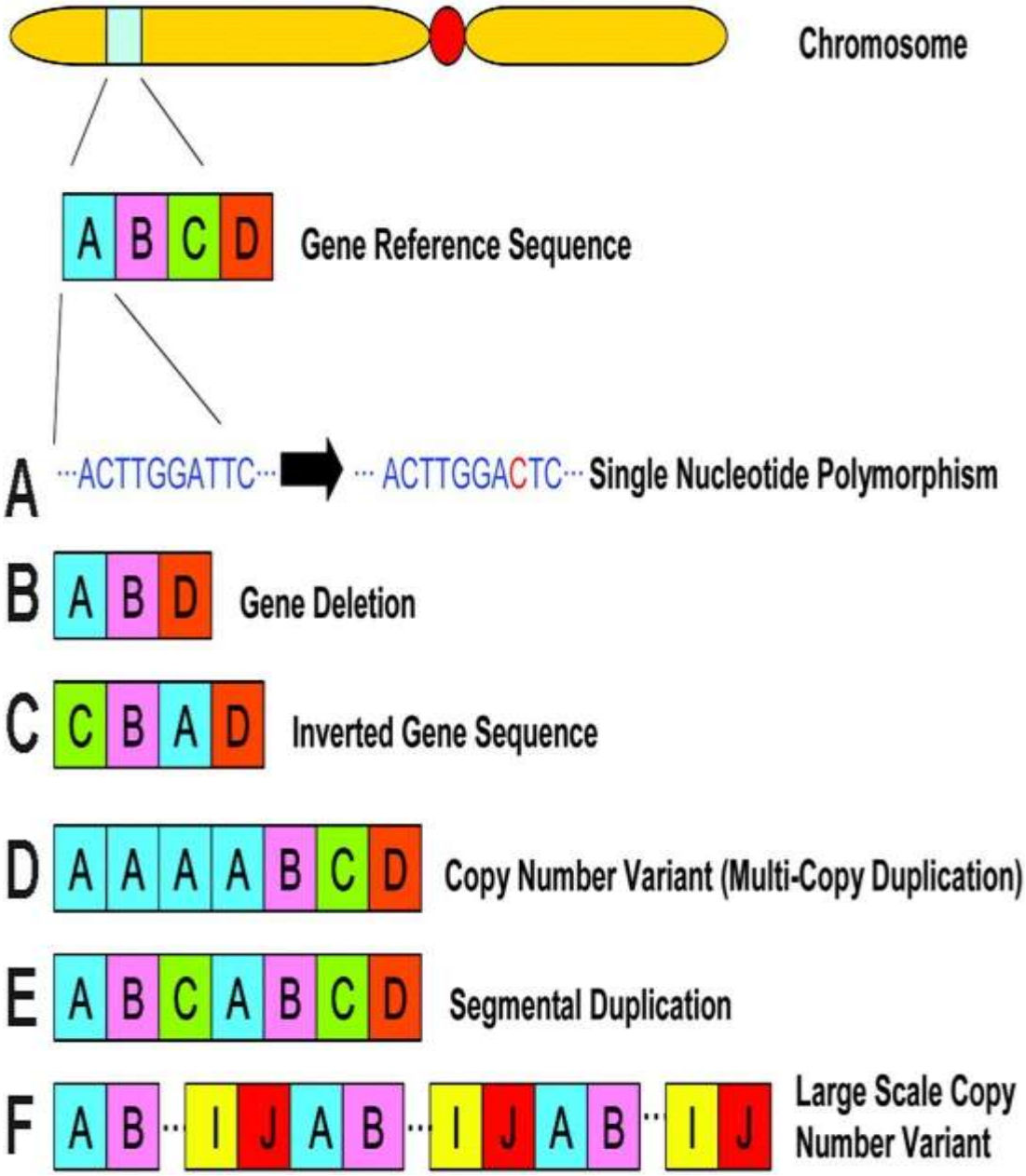
Types

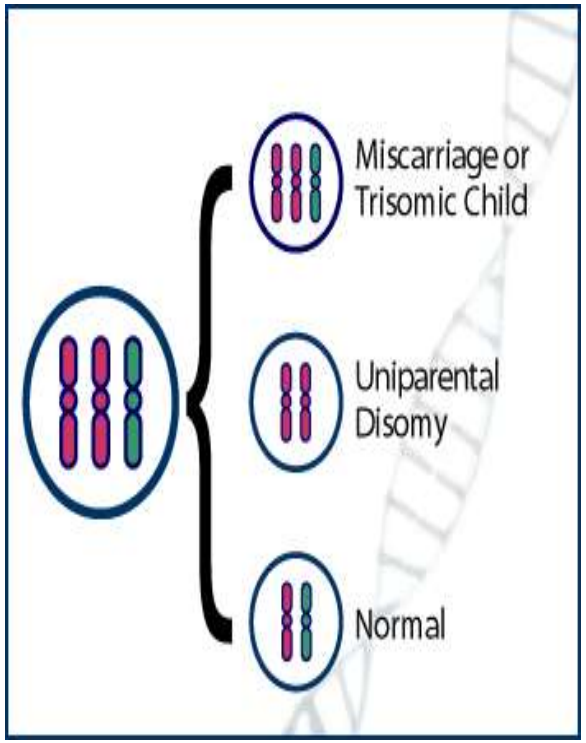
Single nucleotide polymorphisms (SNPs) due to point mutations.

Structural variation

- **Deletions**
- **Duplications**
- **Insertions**
- **Inversions**
- **Translocations**
- **Uniparental disomy**




Chromosome
A B C D
Gene Reference Sequence
A
···ACTTGGATTC···
··· ACTTGGACTC···
Single Nucleotide Polymorphism
B
A B D
Gene Deletion
C
C B A D
Inverted Gene Sequence
D
A A A A B C D
Copy Number Variant (Multi-Copy Duplication)
E
A B C A B C D
Segmental Duplication
F
A B
I J A B
I J A B
I J
Large Scale Copy Number Variant


Miscarriage or Trisomic Child
Uniparental Disomy
Normal

How SNPs are “born” in each generation?

- Number of genomes: $N = 14 \times 10^{9}$ (twice the number of people)
- Mutation rate: $m = \sim 2 \times 10^{-8}$ per base-pair per generation

- New mutations = $Nm = 280$ per base-pair per generation

- Each nucleotide in the genome gets mutated on average in 280 individuals in each generation

- The overwhelming majority of these will never attain polymorphic status (arbitrarily set at 1% of the population)

Microsatellites

- Number of repeats varies greatly between individuals
- Make up to 10-15% of the mammalian genome
- Believed to have no function
- Have high mutation rates
- Used in forensic analysis
- Can be amplified by PCR – fragments that are generated have different length due to different number of repeats

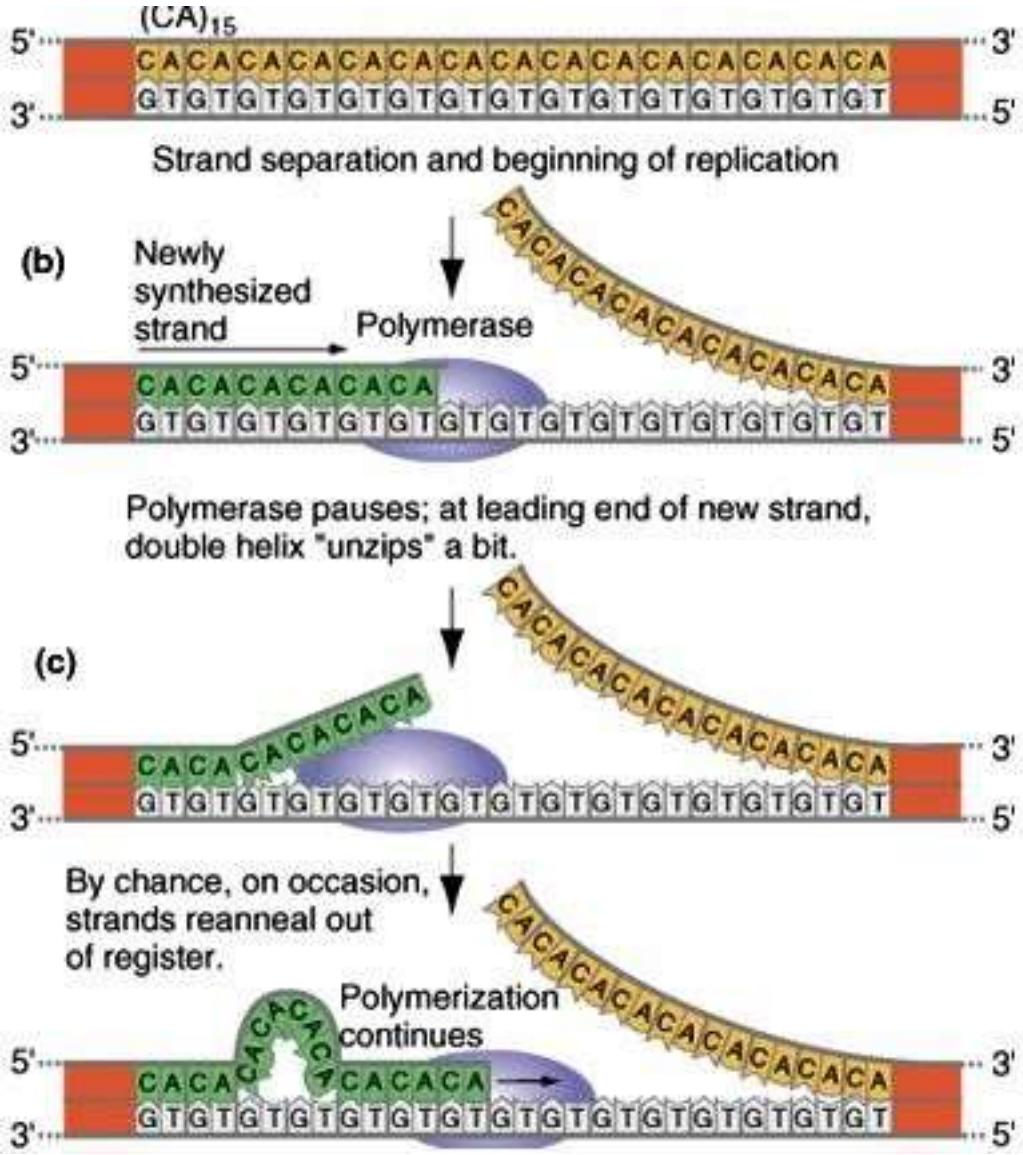
Examples:

Dinucleotide repeats: GTGTGTGTGTGT……

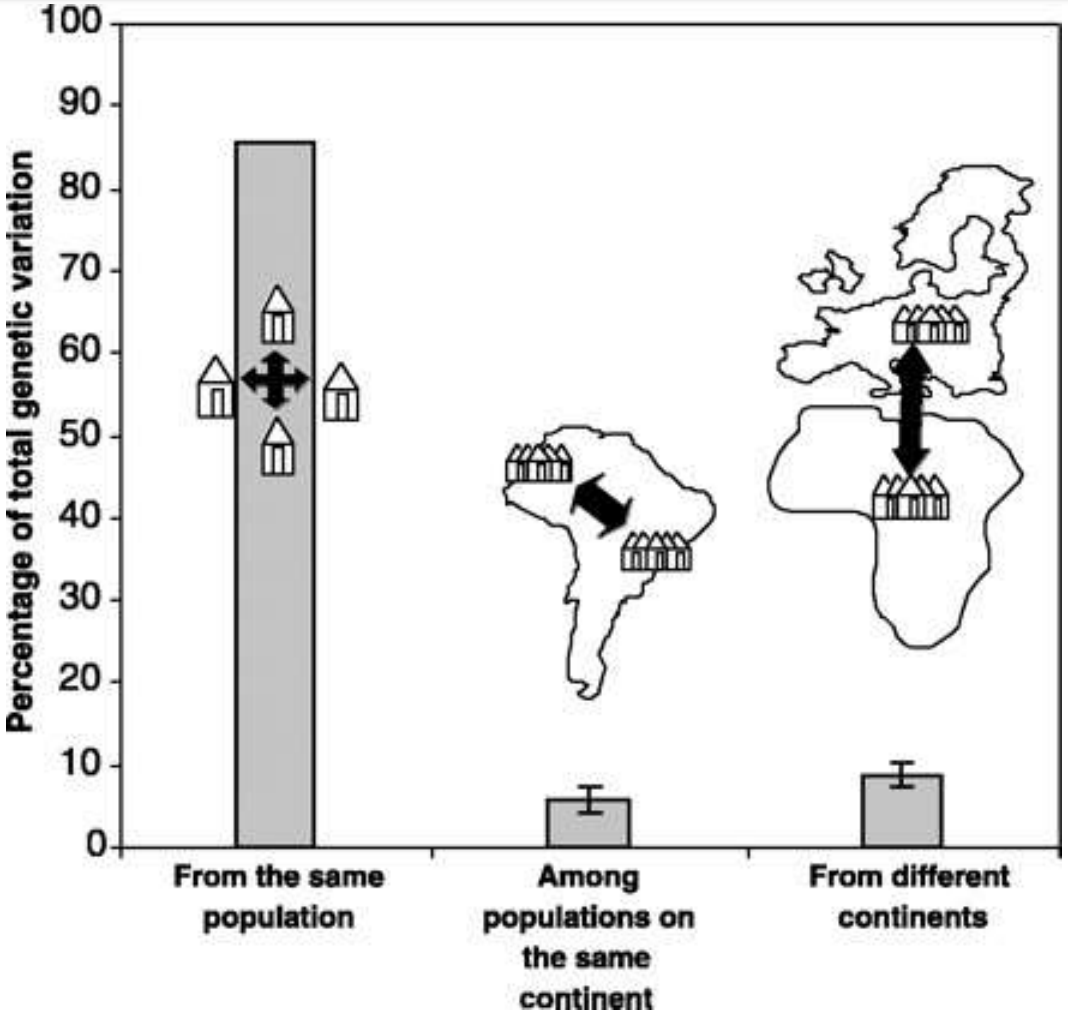
Trinucleotide repeats: ACGACGACGACG……

Tetranucleotide repeats: TATCTATCTATC……

Microsatellites are highly polymorphic due to potential for "skipping" during DNA replication

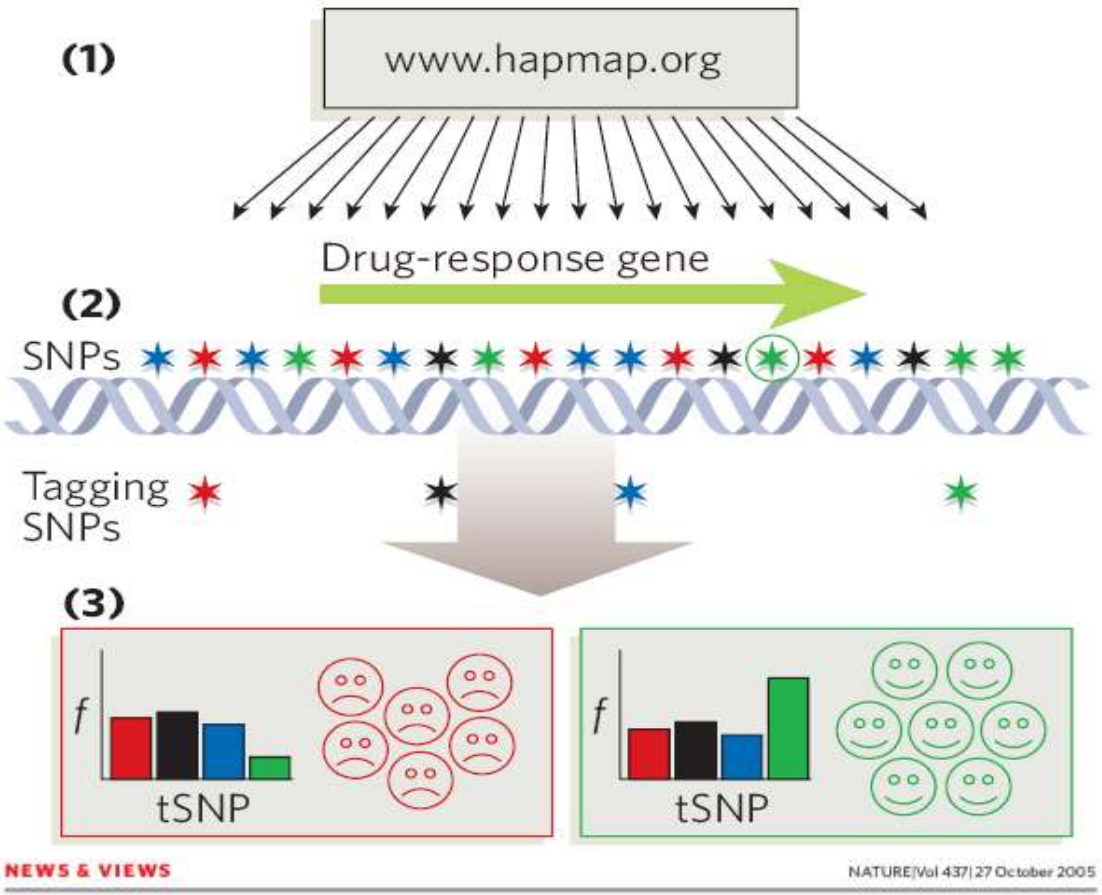


Percentage of Genetic Variation within and between populations



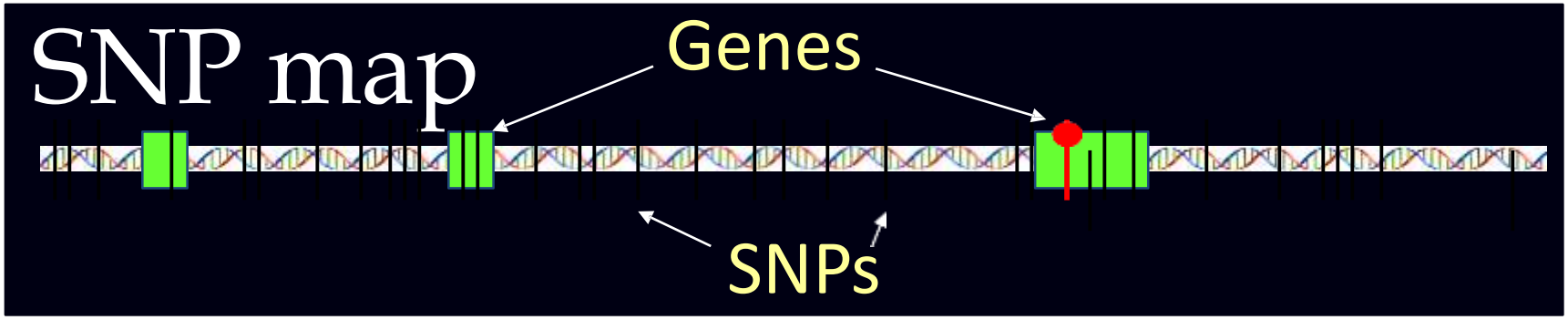
An average population from anywhere in the world contains 85% of all human variation at autosomal loci and 81% of all human variation in mtDNA sequences. Differences among populations from the same continent contribute another 6% of variation; only 9-13% of genetic variation differentiates populations from different continents.

Use of common variations in genetic association studies

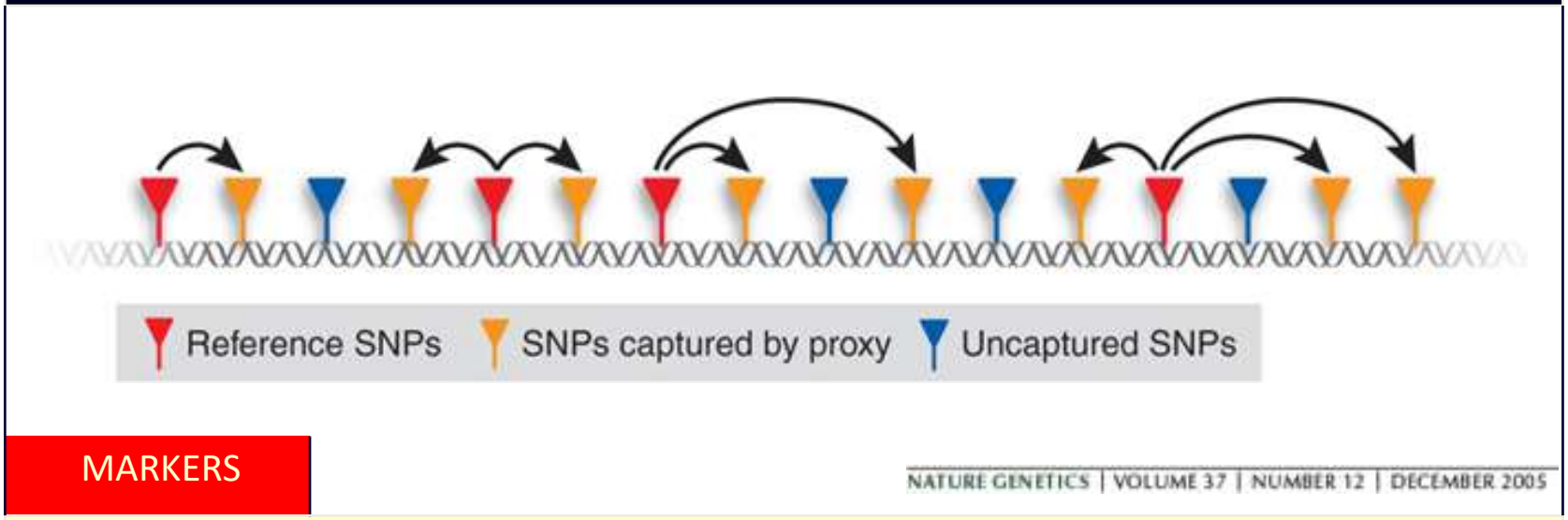


CD /CV (Common Disease/Common Variants) hypothesis:
Common diseases are underlied by a few common variants

CD/RV (Common Disease /Rare Variants) – diseases are driven by many rare alleles

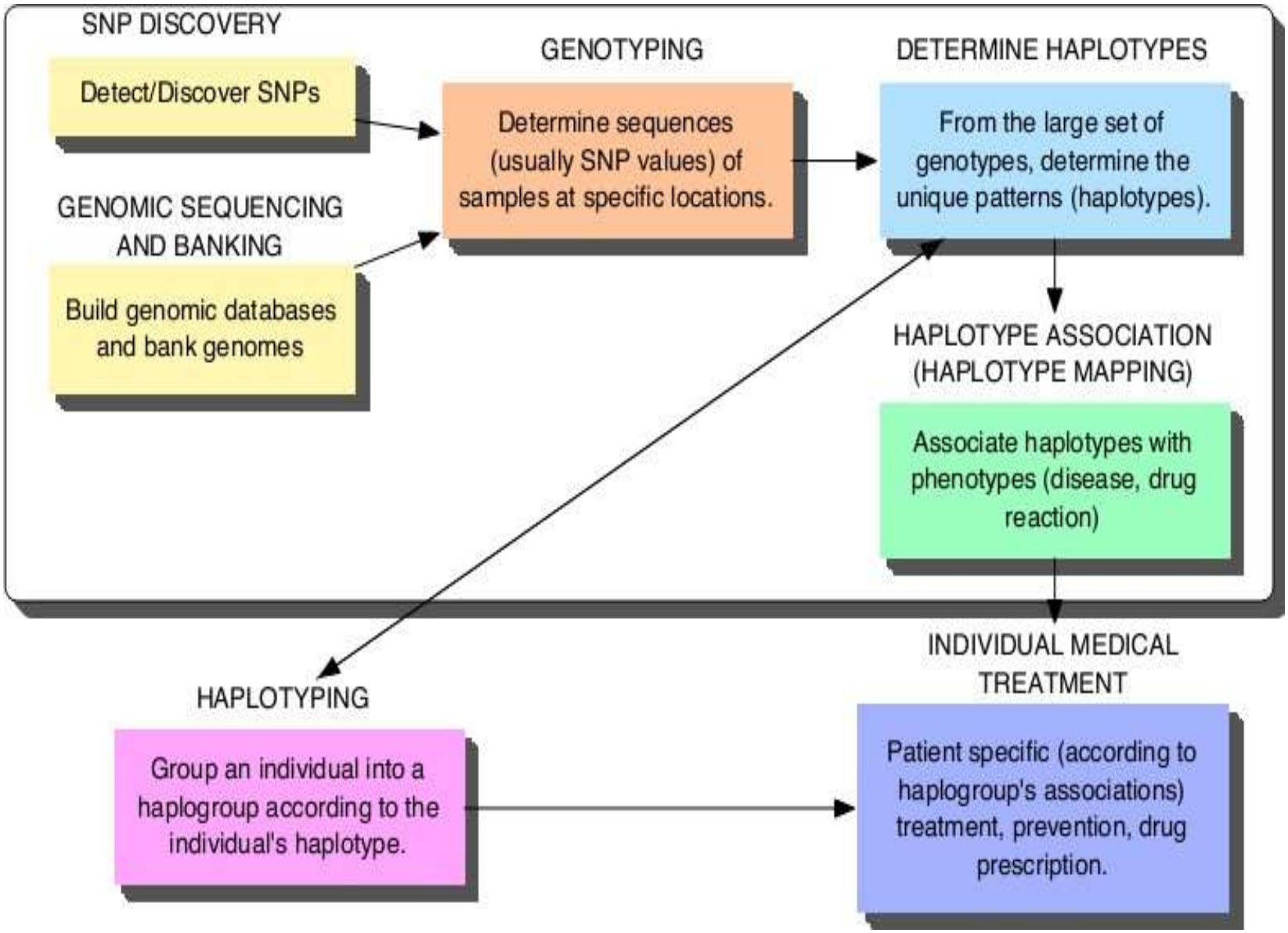


Analysis of some SNPs can capture effects of other SNPs



This is possible owing to Linkage Disequilibrium (LD)

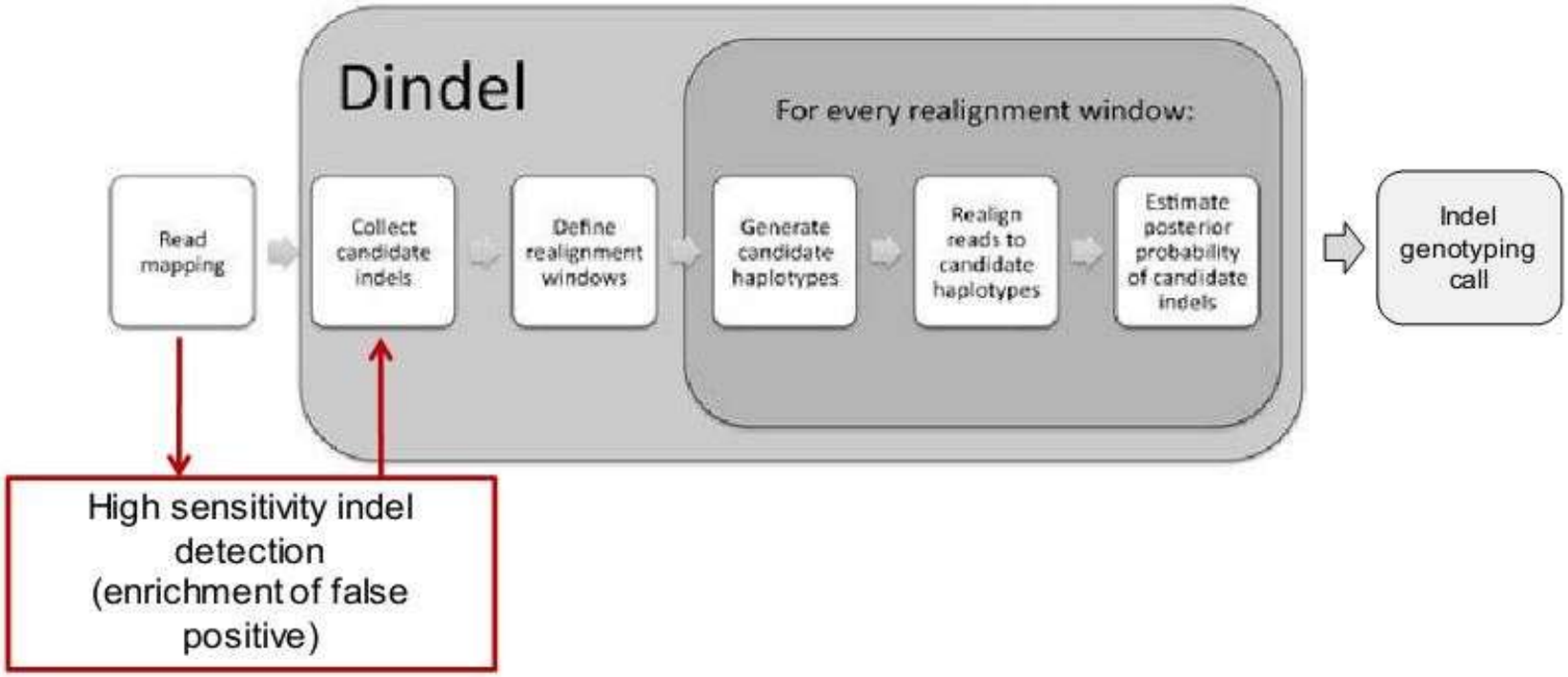
- **LD - Linkage Disequilibrium – non-random association among alleles at two or more loci in POPULATION (or a measure of co-segregation of alleles in population)**
- **Haplotype – combination of alleles on a chromosome (usually used with respect to a small region)**


SNP DISCOVERY
Detect/Discover SNPs
GENOTYPING
Determine sequences (usually SNP values) of samples at specific locations.
DETERMINE HAPLOTYPES
From the large set of genotypes, determine the unique patterns (haplotypes).
GENOMIC SEQUENCING AND BANKING
Build genomic databases and bank genomes
HAPLOTYPE ASSOCIATION (HAPLOTYPE MAPPING)
Associate haplotypes with phenotypes (disease, drug reaction)
INDIVIDUAL MEDICAL TREATMENT
HAPLOTYPING
Group an individual into a haplogroup according to the individual's haplotype.
Patient specific (according to haplogroup's associations) treatment, prevention, drug prescription.

InDel Detection

STEPS

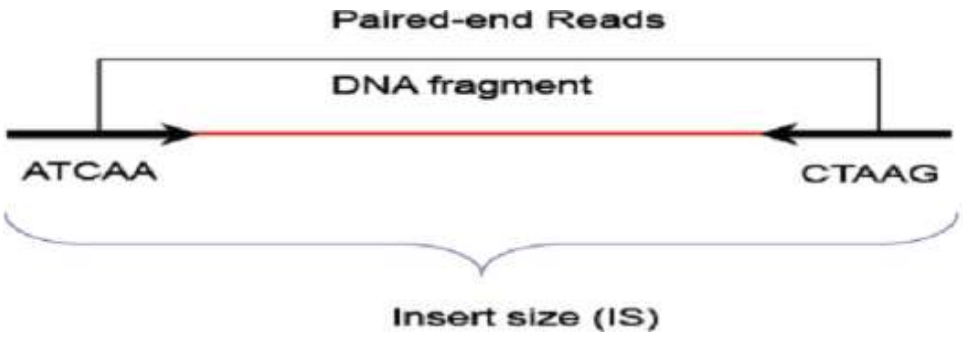
1. Candidate indel identification
2. Calculation of genotype likelihood through local re-alignment
3. LD-based genotype inference and calling



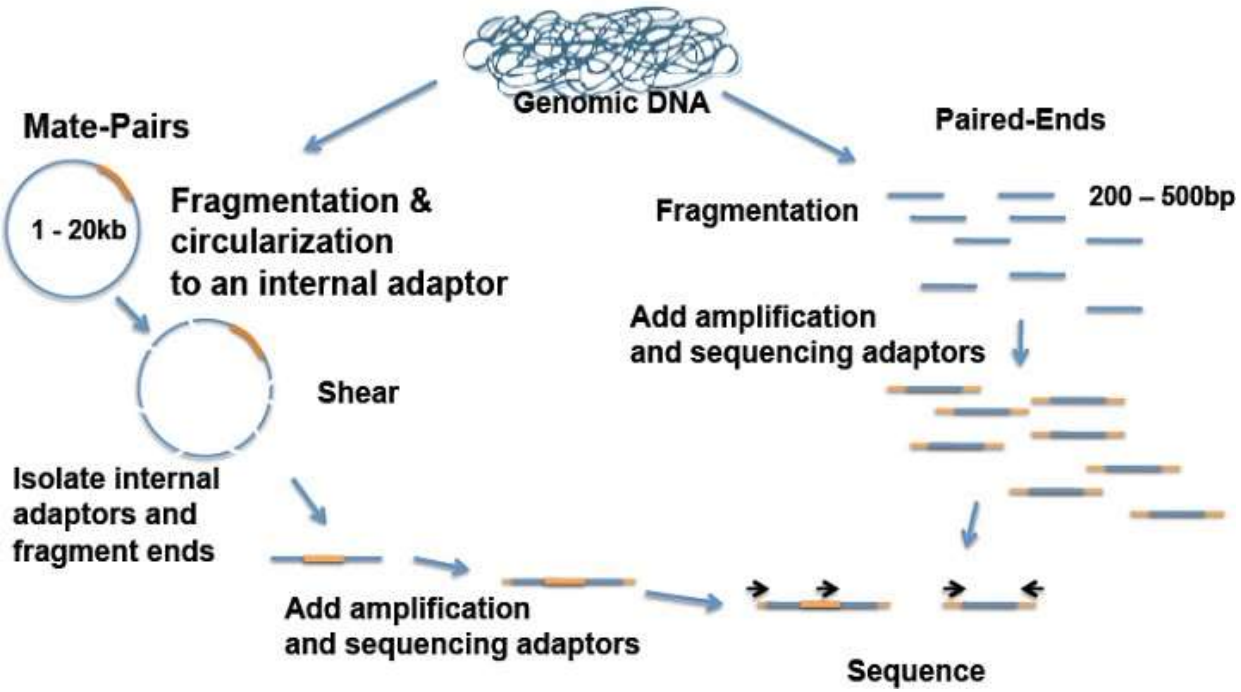
Modified from : Albers, CA et al. 2011 Genome Res

Structural Variations Detection

Next Generation Sequencing



Mate-pair and paired-end reads can be used to detect structural variants



Microsatellite Detection

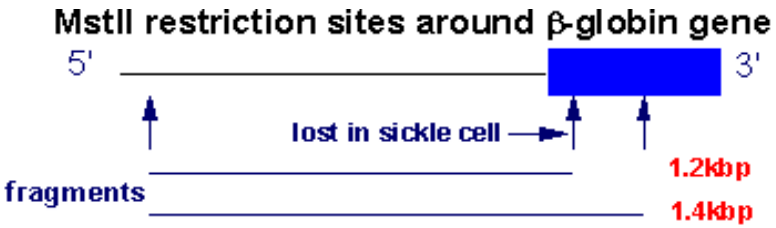
Restriction Fragment Length Polymorphisms (RFLPs)

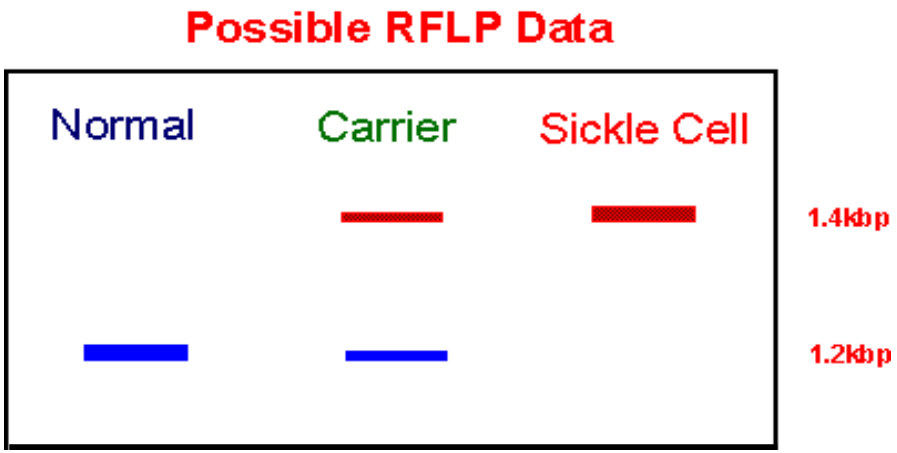
- Consider two alleles having slightly different sequences

GAATTC GCATTC

CTTAAG CGTAAG

*Eco*RI will cut the first but not the second





Genetic Polymorphisms & Drug targets

Genetic variations induce differential drug efficacy

Toxic responders

Non-responders

Responders

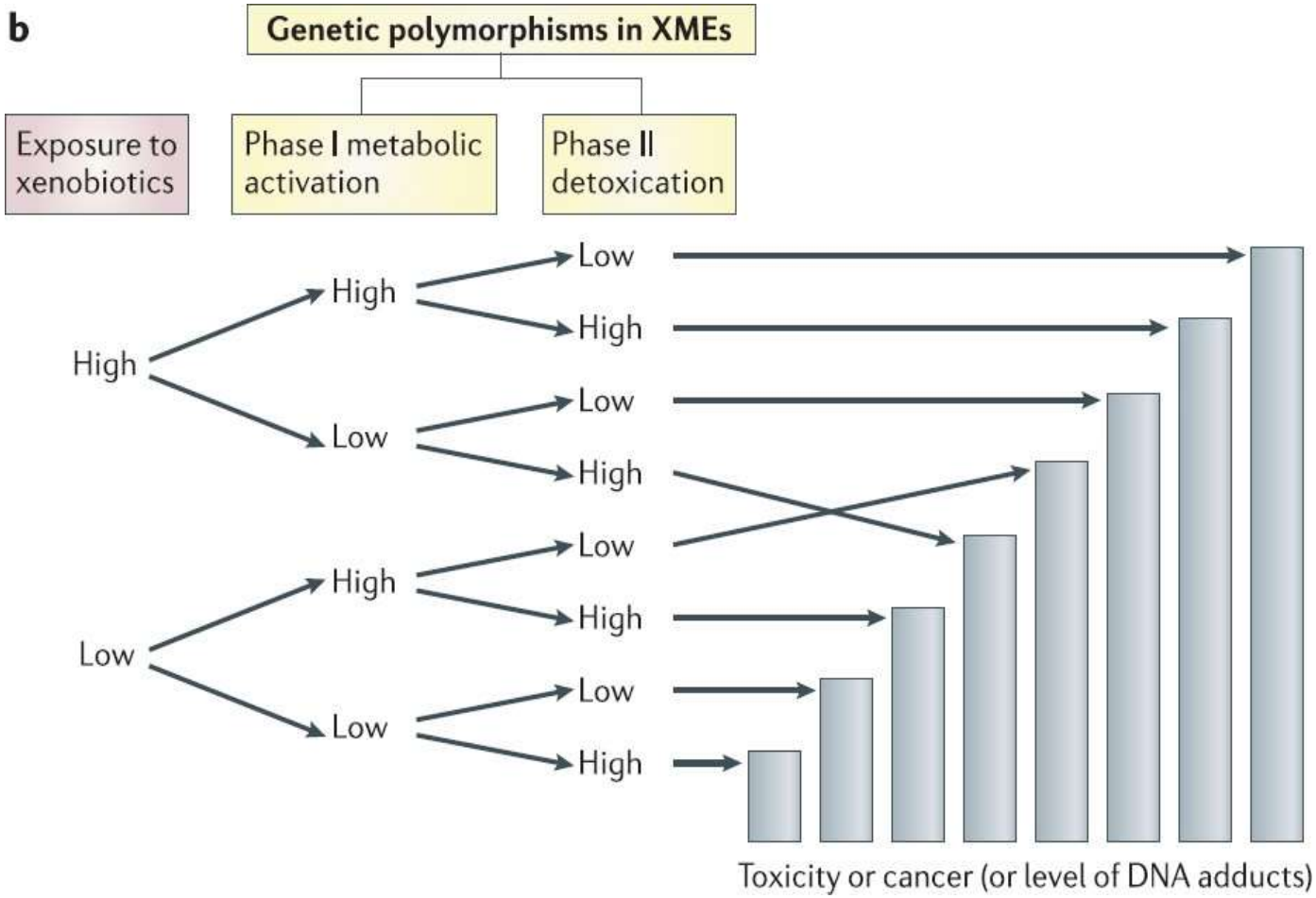
Patient population with same disease phenotype

Genotyping

Patients with drug toxicity

Patients with non-response to drug therapy

Patients with normal response to drug therapy



"The classical interaction of exposure with phase I and phase II XME metabolism, and risk of developing cancer. High exposure to a foreign chemical, combined with rapid metabolic activation and slow conjugation, should put an individual at a high risk of developing cancer. Low or negligible exposure, in combination with slow rates of activation and rapid rates of conjugation, should lead to a low risk of developing environmentally caused cancer."

TABLE 1
Importance of Polymorphic CYP for the Metabolism of Drugs and Carcinogens

Enzyme	Substrates	Polymorphism frequency	Functional effects	Most important polymorphic variants
CYP1A1	Carcinogens	Relatively high	Unproven	No important functional variant alleles
CYP1A2	Drugs, carcinogens	High	Rare	*CYP1A2*1F*, *CYP1A2*1K*
CYP1B1	Carcinogens, estrogens	Rare null alleles, frequent missense mutations	At least seven haplotypes with similar activity	*CYP1B1*7*
CYP2A6	Nicotine, drugs, carcinogens	High in Orientals, less frequent in Caucasians	Important for nicotine metabolism	*CYP2A6*1B*, *CYP2A6*4*, *CYP2A6*9*, *CYP2A6*12*
CYP2B6	Drugs	High	Reduced drug metabolism	*CYP2B6*5*, *CYP2B6*6*, *CYP2B6*16*
CYP2C8	Some drugs	High	Reduced drug metabolism	*CYP2C8*3*
CYP2C9	Drugs	Relatively rare in Caucasians	Very significant	*CYP2C9*2*, *CYP2C9*3*
CYP2C19	Drugs	High	Very significant	*CYP2C19*2*, *CYP2C19*3*, *CYP2C19*17*
CYP2D6	Drugs	Very high	Very significant	*CYP2D6*2xn* *CYP2D6*4*, *CYP2D6*5*, *CYP2D6*10*, *CYP2D6*17*
CYP2E1	Carcinogens, solvents, few drugs	Low	No significant cases have been reported	No important functional variant alleles
CYP3A4	Drugs, carcinogens	Low	No or small	*CYP3A4*1B*
CYP3A5	Drugs, carcinogens	High	Significant	*CYP3A5*3*, *CYP3A5*6*, *CYP3A5*7*
CYP3A7	Drugs, carcinogens	Low	Some	*CYP3A7*2*