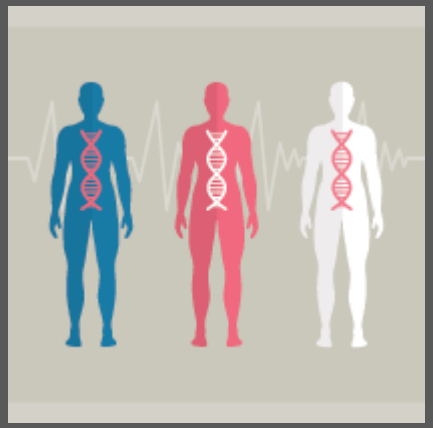
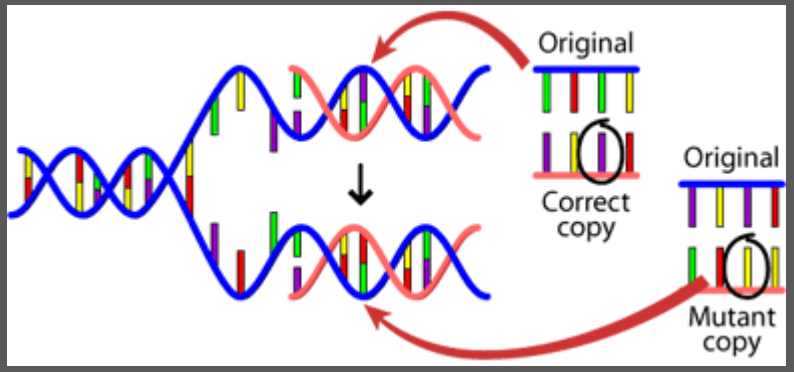
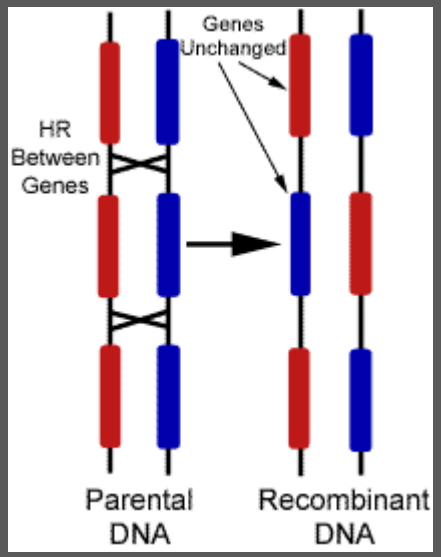
Detección de variantes





Mutations are the original source of genetic variation



Recombination contributes to human genetic variation by shuffling parental DNA and creating new combinations of variants

Variantes, alelos y haplotipos

- **Variantes**: Regiones genómicas diferentes entre dos genomas.

- **Alelos**: Diferentes versiones de una variante. Por ejemplo, un SNP puede tener dos bases alternativas, o alelos, C y T. El ***alelo de la referencia*** es el del genoma de referencia, y el otro el ***alelo alternativo***.

- **Haplotipos**: Los alelos de variantes cercanas en el cromosoma tienden a transmitirse juntos a la descendencia más a menudo de lo esperado por azar. Estos ***bloques de alelos*** se llaman haplotipos. Se dice que estos alelos están ***ligados*** o en '***desequilibrio de ligamiento (LD)***'.

Tipos de variación genética

SNPs

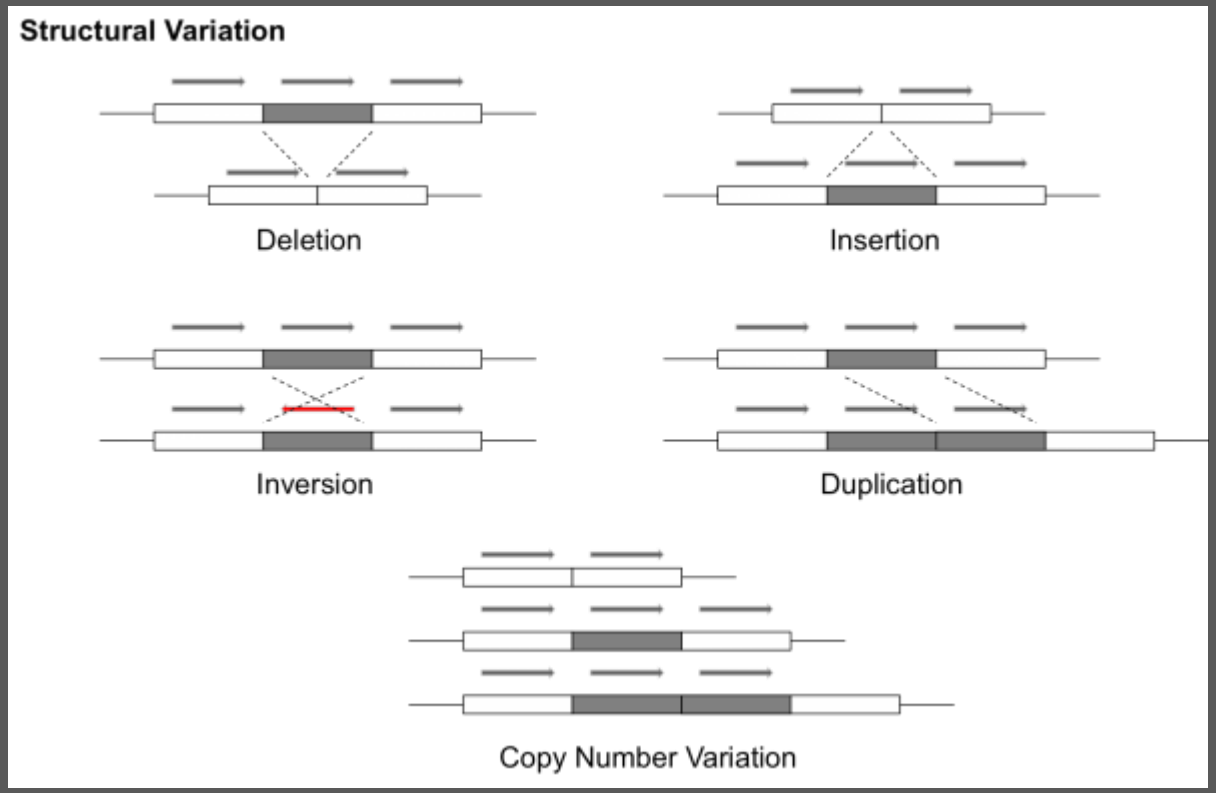
Reference	ACTGACGCATGCATCATGCATGC
SNP	ACTGACGCATGCATCATTCATGC

Indels

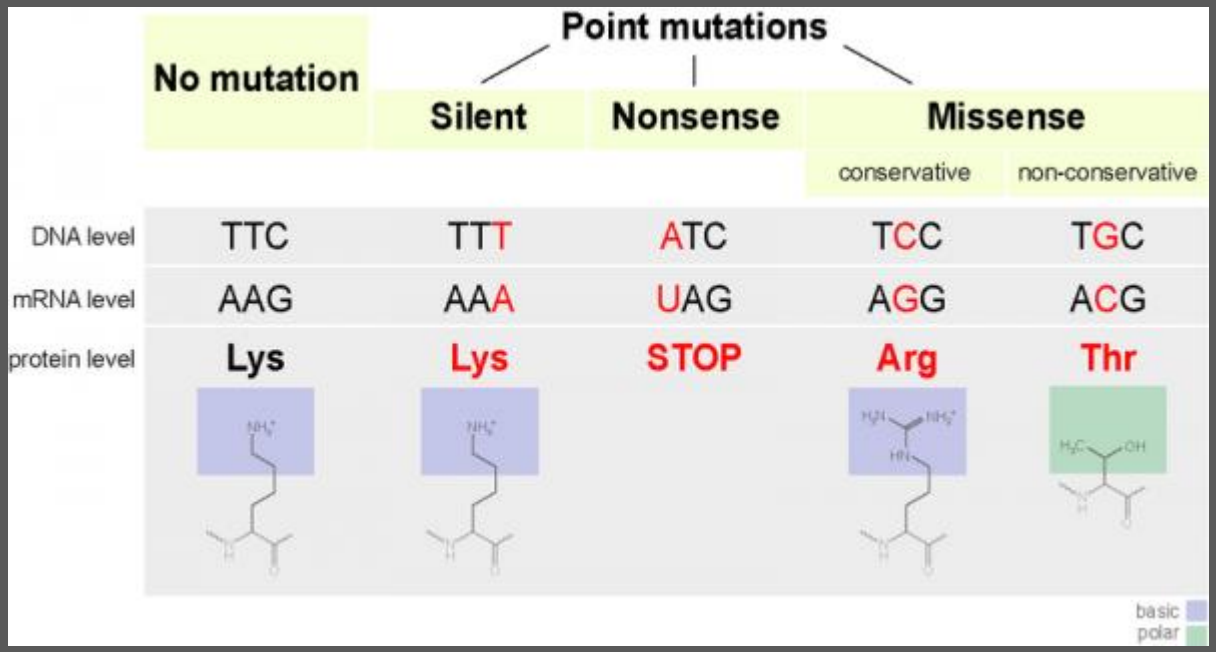
Reference	ACTGACGCATGCATCATGCATGC
Insertion	ACTGACGCATGGTACATCATGCATGC
Deletion	ACTGACG--TGCATCATGCATGC

Indel

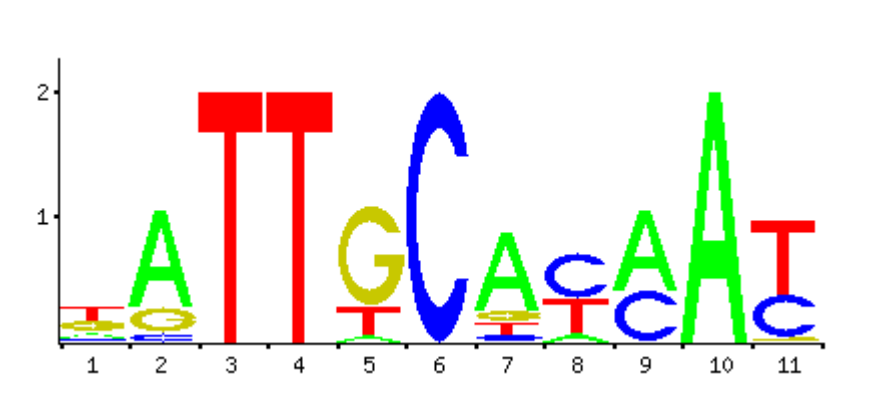
Variantes estructurales



Variantes en regiones codificadoras



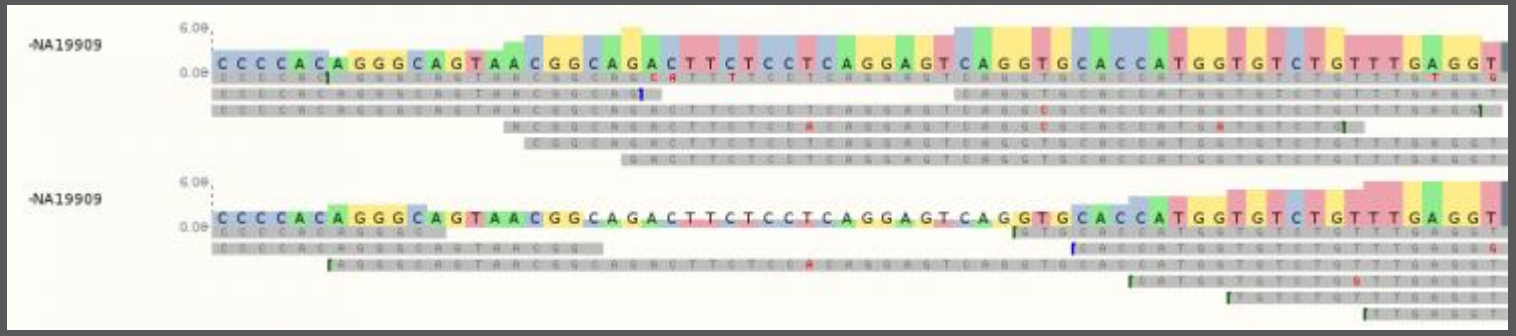
Variantes en sitios de unión de factores de transcripción (TFBSs)



A CEBPB binding motif sequence logo from JASPAR

Detección de variantes (variant calling):

1. WGS -> Lecturas cortas (reads) -> FASTQ files
2. Alineamiento con el genoma de referencia -> BAM files
3. Identificación de las diferencias -> VCF files
4. Filtrado



A CRAM file aligned to a reference genomic region as visualised in Ensembl.
Differences are highlighted in red in the reads, and will be called as variants.

Formato VCF

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA19909
11	5248232	rs334	T	A	100	PASS	AA=T|||;AC=1;AF=0.0273562;AFR_AF=0.0998;AMR_AF=0.0072;AN=2;DP=22876;EAS_AF=0;EUR_AF=0;EX_TARGET;NS=2504;SAS_AF=0;VT=SNP	GT	0|1

Column	Content	Description
#CHROM	Chromosome	
POS	Co-ordinate	The start coordinate of the variant.
ID	Identifier	
REF	Reference allele	The reference allele is whatever is found in the reference genome. It is not necessarily the major allele.
ALT	Alternative allele	The alternative allele is the allele found in the sample you are studying.
QUAL	Score	Quality score out of 100.
FILTER	Pass/fail	If it passed quality filters.
INFO	Further information	Allows you to provide further information on the variants. Keys in the INFO field can be defined in header lines above the table.
FORMAT	Information about the following columns	The GT in the FORMAT column tells us to expect genotypes in the following columns.
NA19909	Individual identifier (optional)	The previous column told us to expect to see genotypes here. The genotype is in the form 0|1, where 0 indicates the reference allele and 1 indicates the alternative allele, i.e it is heterozygous. The vertical pipe | indicates that the genotype is phased, and is used to indicate which chromosome the alleles are on. If this is a slash / rather than a vertical pipe, it means we don't know which chromosome they are on.

Efectos biológicos de las variantes

- SnpEff
- Ensembl VEP
- …