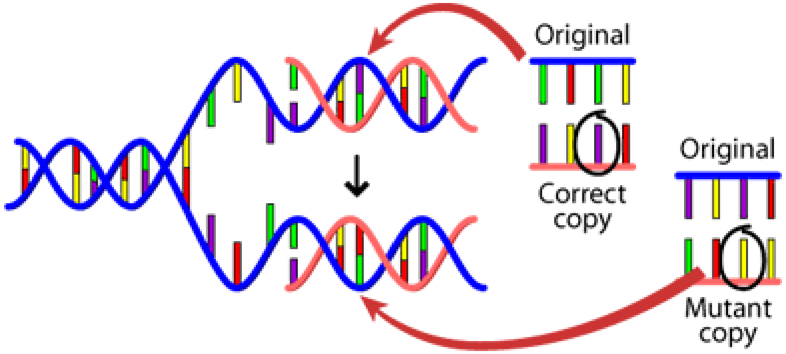
Human Genetic Variation

Ricardo Lebrón
rlebron@ugr.es
Dpto. Genética UGR

What is Genetic Variation?

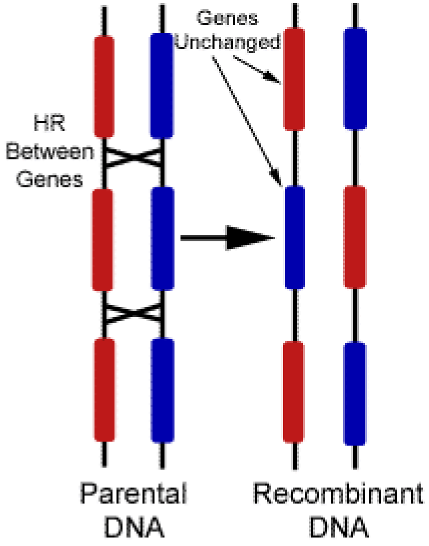
Origins of Genetic Variation

- ***Genetic Variation*** **is the difference in DNA sequences between individuals. Mutations and recombination are major sources of variation.**
- ***Mutations*** **are the original source of genetic variation. A mutation is a permanent alteration to a DNA sequence and it usually occurs when there is an error during DNA replication that is not corrected by DNA repair enzymes. Mutations may be beneficial, deleterious or neutral. They occur in germ and somatic cells, but only variations that arise in germ cells can be inherited.**



Origins of Genetic Variation

- ***Genetic Variation*** **is the difference in DNA sequences between individuals. Mutations and recombination are major sources of variation.**
- ***Recombination*** **is another major source of genetic variation and it occurs in meiosis when homologous DNA strands align and cross over. Recombination effectively shuffes maternal and paternal DNA, creating new combinations of variants in the daughter germ-cells.**



Variants, alleles and haplotypes

- **The term *variant* is used to refer to a specific position or region of the genome which differs between two or more genomes.**
- **Different versions of the same variant are called *alleles*.**
- **Alleles at variants close together on the same chromosome tend to occur together more often than is expected by chance. These blocks of alleles are called *haplotypes*.**
- ***Linkage disequilibrium* (LD) is a measure of how often two alleles are inherited together.**

Types of Genetic Variation

Single Nucleotide Polymorphism (SNP)

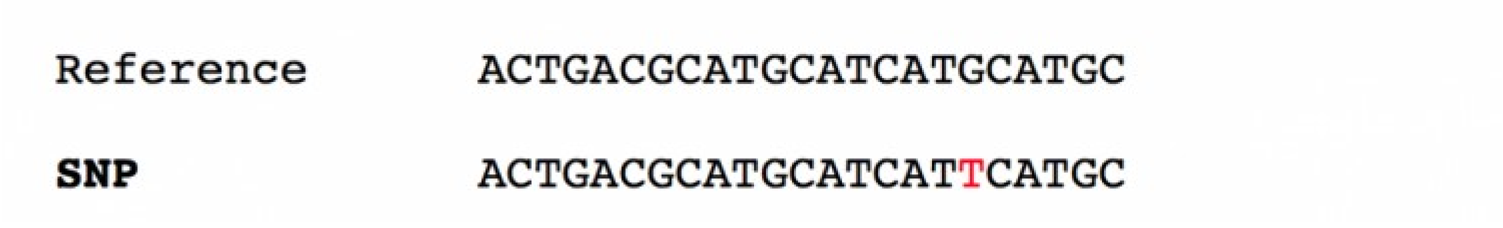
- It is a single base-pair substitution. There are two kinds of substitutions: ***transition*** and ***transversion***.
- Transitions:

  purine → purine, or pyrimidine → pyrimidine
- Transversions:

  purine → pyrimidine, or pyrimidine → purine

  purine: A and G.

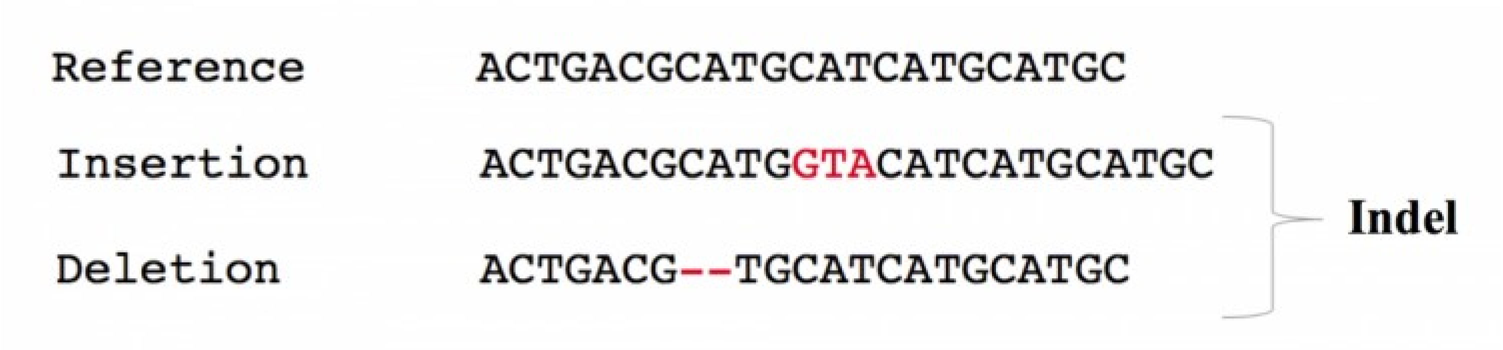
  pyrimidine: T and C.



Reference	ACTGACGCATGCATCATGCATGC
SNP	ACTGACGCATGCATCATTCATGC

Insertion or Deletion (indel)

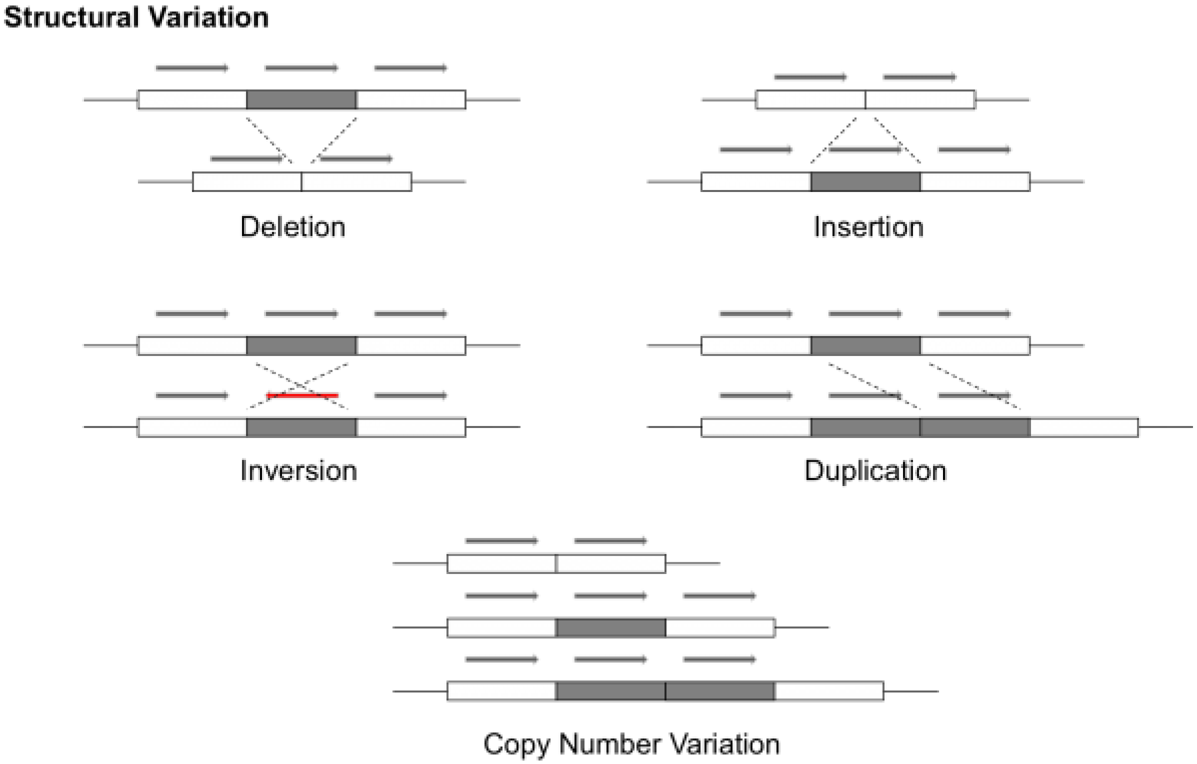
- It is an ***insertion*** or ***deletion*** of a single stretch of DNA sequence that can range from one to hundreds of base-pairs in length.



Reference	ACTGACGCATGCATCATGCATGC
Insertion	ACTGACGCATGGTACATCATGCATGC
Deletion	ACTGACG--TGCATCATGCATGC

Structural Variation

- **Typically used to describe genetic variation that occurs over a larger DNA sequence. This category of genetic variation includes both *copy number variation* and *chromosomal rearrangement events*.**



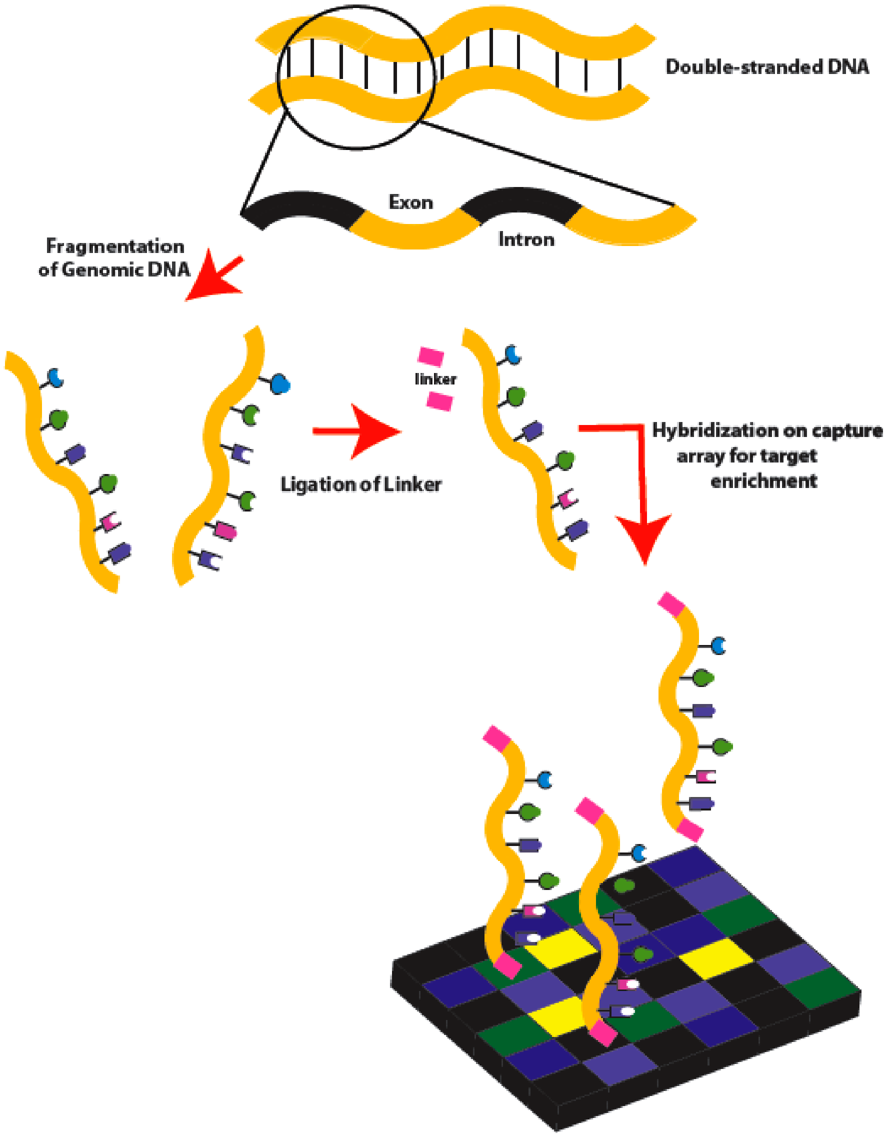
Exome Sequencing

Exome Sequencing

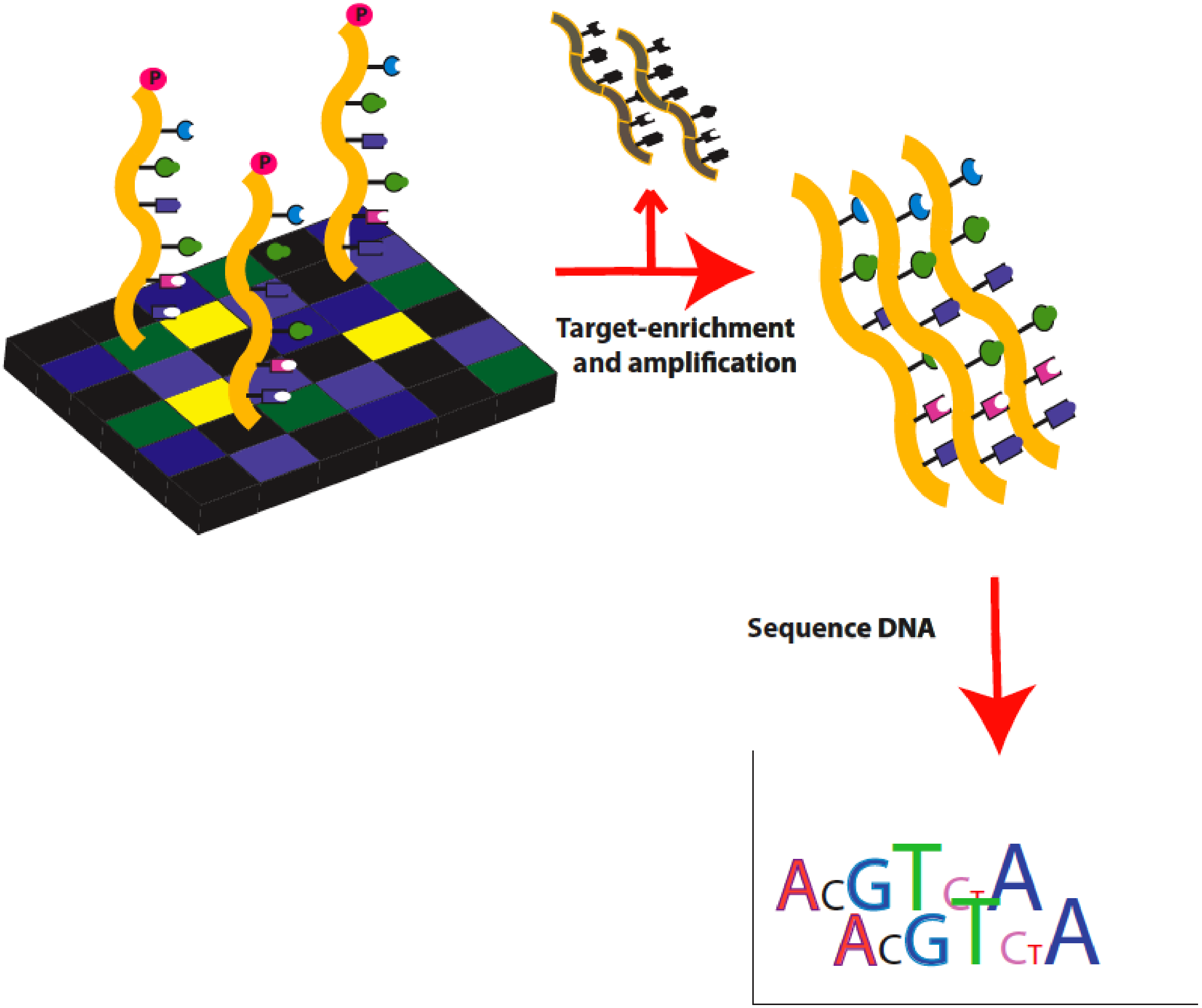
- **Select the subset of *DNA that encodes proteins*.**
- **Sequence the exonic DNA using any *high-throughput sequencing* technology.**

The goal of this approach is to identify genetic variants that alter protein sequences, and to do this at a much lower cost than whole-genome sequencing.

Exome Sequencing



Exome Sequencing



The 1,000 Genomes Project

1,000 Genomes Project

- **2,504 individuals from 26 populations genotyped using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping.**

The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations.

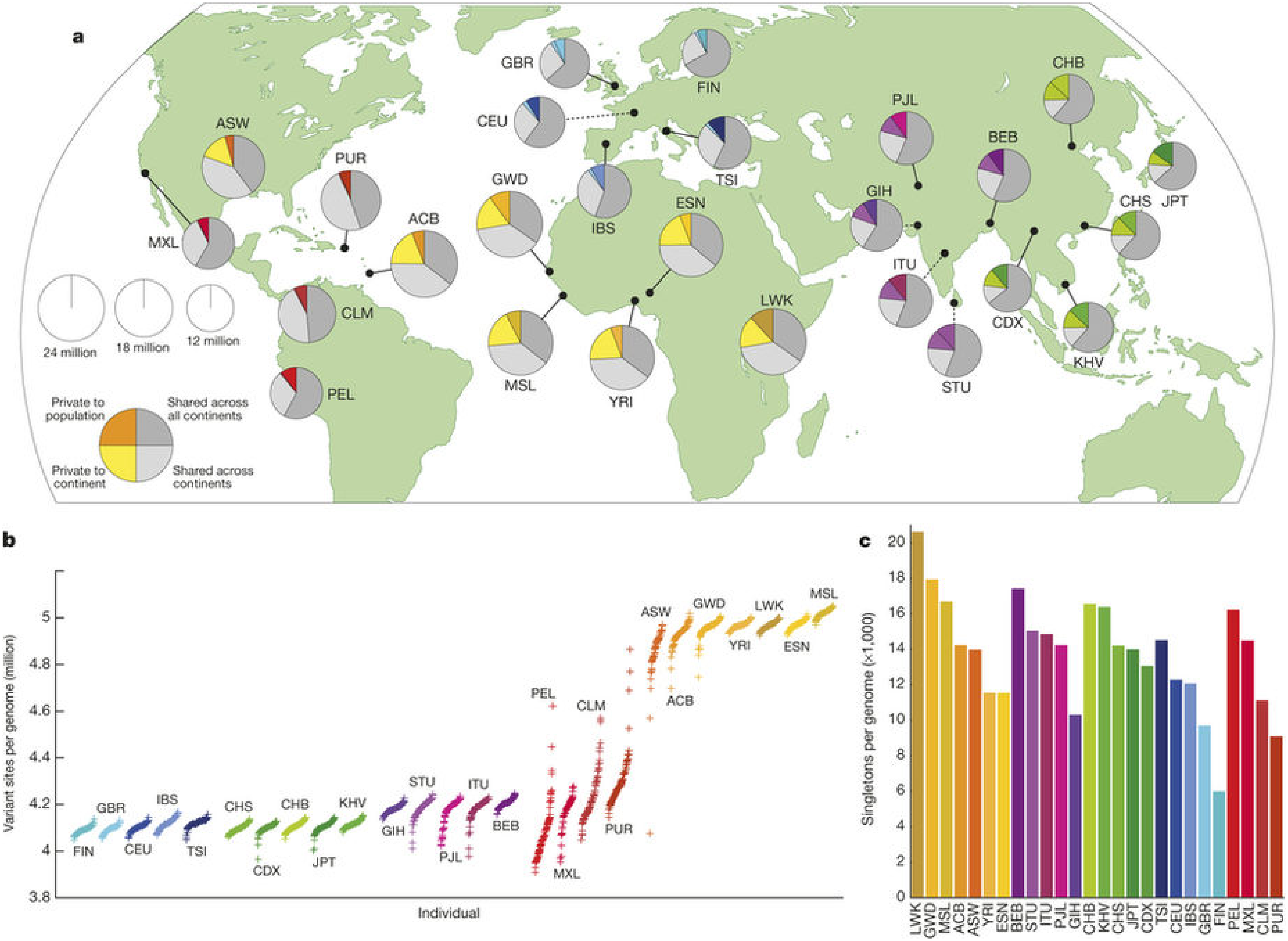
1,000 Genomes Project

88 million variants:

- **84.7 million single nucleotide polymorphisms (SNPs)**
- **3.6 million short insertions/deletions (indels)**
- **60,000 structural variants**

**1000 Genomes Project Consortium (2015).
A global reference for human genetic variation.
Nature, 526, 68-74.**

1,000 Genomes Project



1,000 Genomes Project

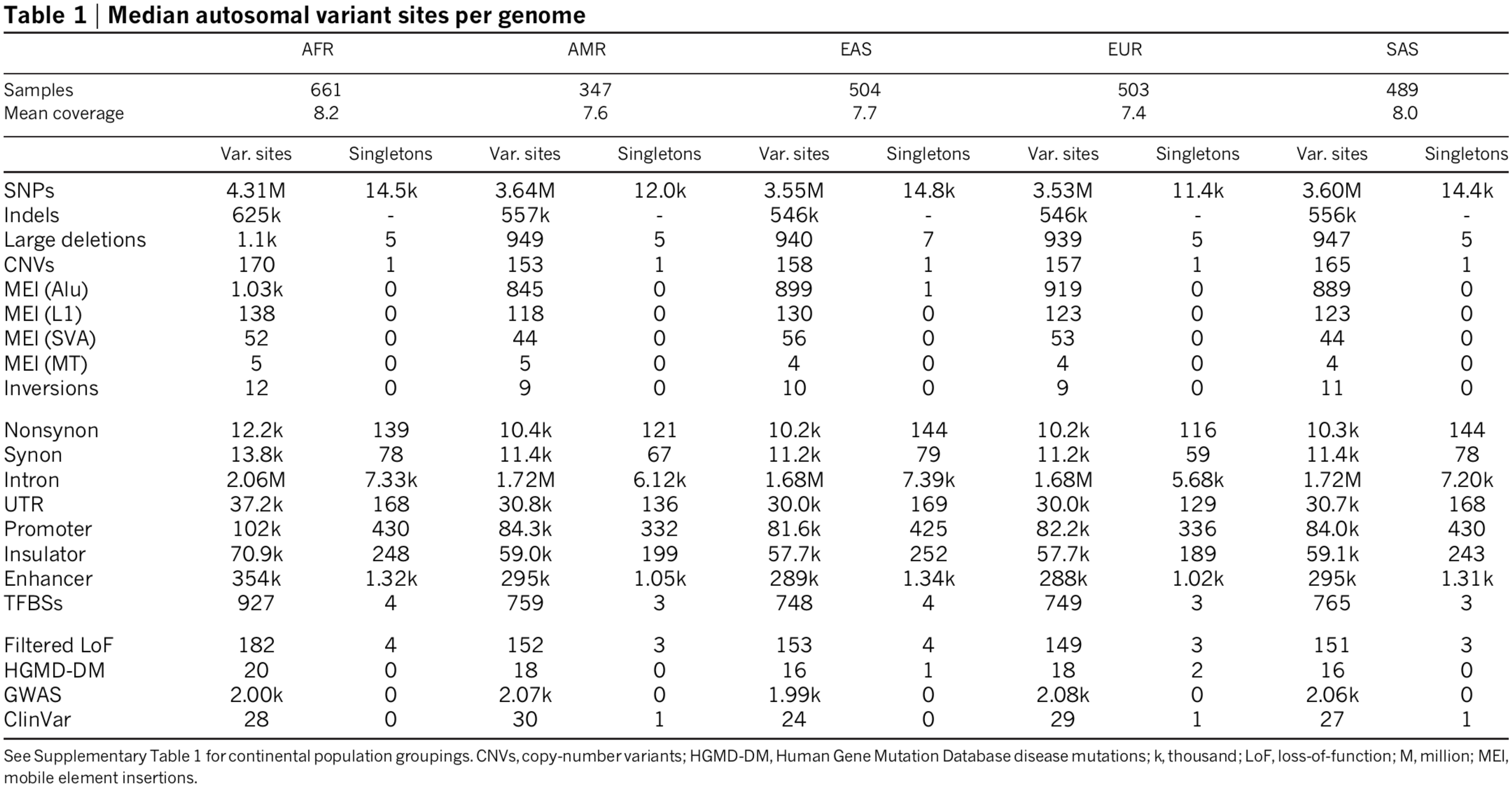


Table 1 | Median autosomal variant sites per genome

	AFR		AMR		EAS		EUR		SAS	
Samples	661		347		504		503		489	
Mean coverage	8.2		7.6		7.7		7.4		8.0	
	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons
SNPs	4.31M	14.5k	3.64M	12.0k	3.55M	14.8k	3.53M	11.4k	3.60M	14.4k
Indels	625k	-	557k	-	546k	-	546k	-	556k	-
Large deletions	1.1k	5	949	5	940	7	939	5	947	5
CNVs	170	1	153	1	158	1	157	1	165	1
MEI (Alu)	1.03k	0	845	0	899	1	919	0	889	0
MEI (L1)	138	0	118	0	130	0	123	0	123	0
MEI (SVA)	52	0	44	0	56	0	53	0	44	0
MEI (MT)	5	0	5	0	4	0	4	0	4	0
Inversions	12	0	9	0	10	0	9	0	11	0
Nonsynon	12.2k	139	10.4k	121	10.2k	144	10.2k	116	10.3k	144
Synon	13.8k	78	11.4k	67	11.2k	79	11.2k	59	11.4k	78
Intron	2.06M	7.33k	1.72M	6.12k	1.68M	7.39k	1.68M	5.68k	1.72M	7.20k
UTR	37.2k	168	30.8k	136	30.0k	169	30.0k	129	30.7k	168
Promoter	102k	430	84.3k	332	81.6k	425	82.2k	336	84.0k	430
Insulator	70.9k	248	59.0k	199	57.7k	252	57.7k	189	59.1k	243
Enhancer	354k	1.32k	295k	1.05k	289k	1.34k	288k	1.02k	295k	1.31k
TFBSs	927	4	759	3	748	4	749	3	765	3
Filtered LoF	182	4	152	3	153	4	149	3	151	3
HGMD-DM	20	0	18	0	16	1	18	2	16	0
GWAS	2.00k	0	2.07k	0	1.99k	0	2.08k	0	2.06k	0
ClinVar	28	0	30	1	24	0	29	1	27	1

See Supplementary Table 1 for continental population groupings. CNVs, copy-number variants; HGMD-DM, Human Gene Mutation Database disease mutations; k, thousand; LoF, loss-of-function; M, million; MEI, mobile element insertions.