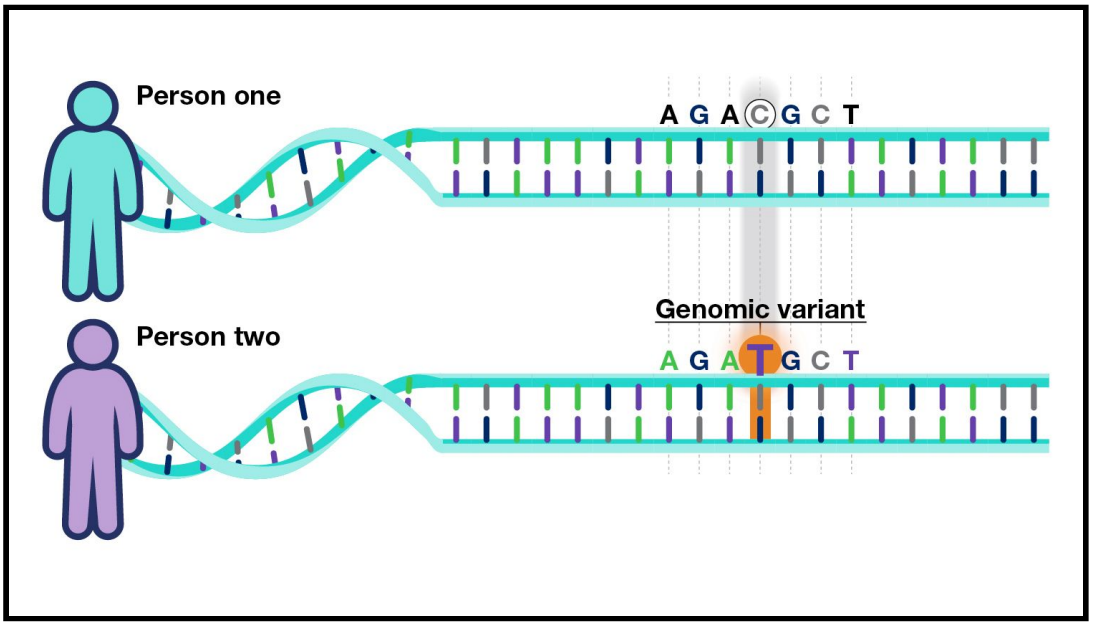
Fiche sur les données de Variations Génomiques

1. Qu'est ce que c'est une variation génomique ?

Une variation génomique est un changement, d'une ou plusieurs bases nucléotides, dans une séquence d'ADN particulière en comparaison avec une séquence d'ADN (un génome) de référence (1). Les variations génomiques se distinguent en deux catégories : polymorphismes et mutations.



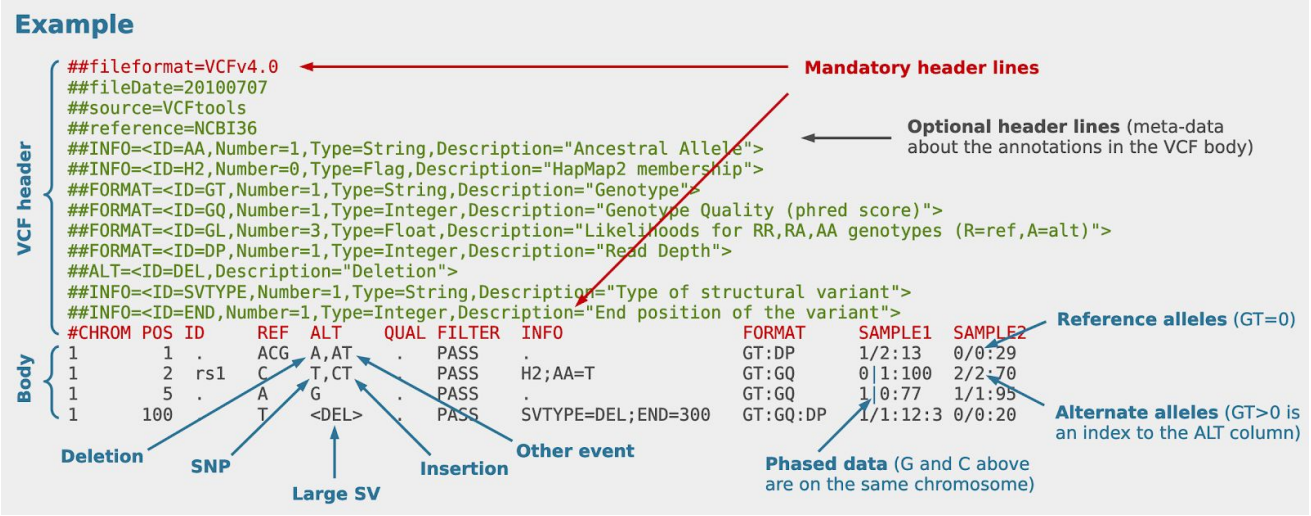
(https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores#one
)

Il existe différentes types de variations :

- SNV (Single Nucleotide Variation).
- InDel (Insertion-Deletion).
- SV (Structural Variation).
- CNV (Copy Number Variation).
-

2. Description de la donnée

Type de données	qualitatif	Présence / absence des variations
Technique/source	Variant calling workflow	Détection automatisée des variations à partir de données DNAseq ou RNAseq (moins fréquent). Cela inclut une étape d'alignement des séquences DNAseq/RNAseq sur un génome de référence.
Format standard	VCF (Variant Call Format)	Example (figure)
	Matrice binaire	
Métadonnées requises	Espèce/variété	
	Version du génome de référence	utilisé dans l'alignement de séquences
	Type de séquençage (individu ou bulk)	



(http://vcftools.sourceforge.net/VCF-poster.pdf)

3. Bases de données et ontologies

Coté standard, le VCF est un format standard décrivant tous les variants et repris par tous les outils

Base de données	Description	download/submit
dbVar (Database of genomic structural VARiation)	dbVar is a database of genomic structural variation. It accepts data from all species and includes clinical data. It can accept diverse types of events, including inversions, insertions and translocations. Additionally, both germline and somatic variants are accepted.	download/submit
EVA (European Variation Archive)	The European Variation Archive is an open-access database of all types of genetic variation data from all species. https://www.ebi.ac.uk/eva/	download/submit
1001 Genomes	A Catalog of Arabidopsis thaliana Genetic Variation.	download
GnpIS	GnpIS handles different types of data in the scope of genetics and genomics for plants including forest trees and fungi: genetic resources, polymorphisms and genotyping data, phenotyping data, association data, genetic maps and QTLs, synteny data.	download/submit

Ontologies		
Variant Ontology (VariO)	is an ontology for standardized, systematic description of effects, consequences and mechanisms of variations	Vihinen, M., 2014
The SNP Ontology	SNP-Ontology is a domain ontology that provides a formal representation (OWL-DL) of genomic variations.	
The Sequence Ontology (SO)	The Sequence Ontology is a set of terms and relationships used to describe the features and attributes of biological sequence. SO includes different kinds of features which can be located on the sequence.	Eilbeck et al. 2005