Informatics on High-throughput Sequencing Data

(Summer Course 2020)

Day 17

Variant Calling

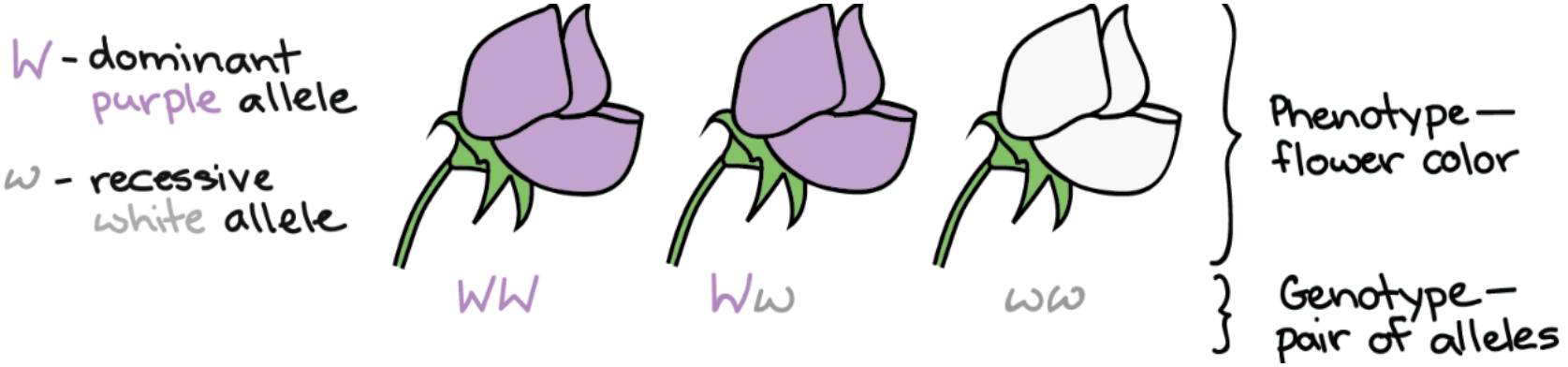
- Variant calling is the process by which we identify variants from sequence data.

1. Carry out whole genome or whole exome sequencing to create FASTQ files.
2. Align the sequences to a reference genome, creating BAM files.
3. Identify where the aligned reads differ from the reference genome and write to a VCF file.

https://www.ebi.ac.uk/training-beta/online/courses/human-genetic-variation-introduction/variant-identification-and-analysis/

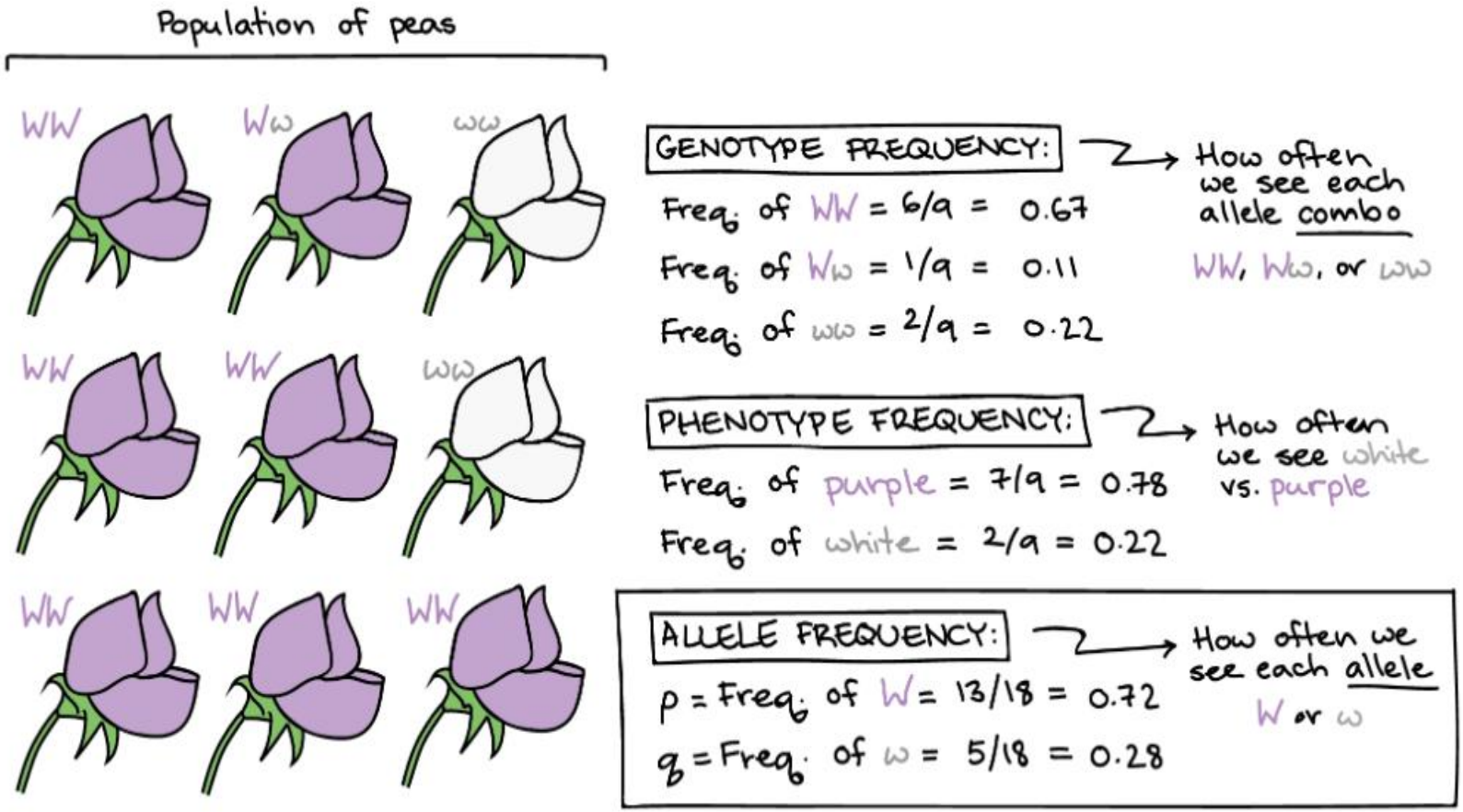
Notes !

- A genotype is an individual's collection of genes. The term also can refer to the two alleles inherited for a particular gene.
- Allele is a variant form of a given gene. a heritable unit that controls a particular feature of an organism.



https://www.khanacademy.org/science/biology/her/heredity-and-genetics/a/allele-frequency-the-gene-pool

Notes !



https://www.khanacademy.org/science/biology/her/heredity-and-genetics/a/allele-frequency-the-gene-pool

Notes !

- Unphased data are simply the genotypes without regard to which one of the pair of chromosomes holds that allele.
- / : genotype unphased

- Phased data are ordered along one chromosome and so from these data you know the haplotype.
- | : genotype phased

- A haplotype can refer to a combination of alleles or to a set of single nucleotide polymorphisms (SNPs) found on the same chromosome.
- Information about haplotypes is being collected by the International HapMap Project and is used to investigate the influence of genes on disease.

https://support.10xgenomics.com/genome-exome/software/pipelines/latest/output/vcf

VCF/BCF format

- VCF is the standard file format for storing variation data.
- VCF files are tab delimited text files.

Types of variants

SNPs

Alignment	*VCF representation*			
ACGT	POS	REF	ALT
ATGT	2	C	T

Insertions

Alignment	*VCF representation*		
AC-GT	POS	REF	ALT
ACTGT	2	C	CT

Deletions

Alignment	*VCF representation*		
ACGT	POS	REF	ALT
A--T	1	ACG	A

Complex events

Alignment	*VCF representation*		
ACGT	POS	REF	ALT
A-TT	1	ACG	AT

Large structural variants

VCF representation

POS	REF	ALT	INFO
100	T	<DEL>	SVTYPE=DEL;END=300

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

http://digitheadslabnotebook.blogspot.com/2013/01/vcf-variant-call-format.html

Thanks!

// | ?