Variant calling and annotation

Alt-text Figure 27 - IGV visualisation of a SARS-CoV-2 variant site

Screenshot of IGV tool. It shows the SARS-CoV-2 reference genome (NC_045512.2) and the mapped query reads. Highlighted in orange in the middle of the page is the Variant site 23403 A>G. Details in the main text.

Alt-text Figure 28 - High-level overview of the variant calling process.

A flowchart showing the high-level overview of the variant calling process: Input = sample BAM/CRAM file > Identifying candidate variant sites > Ascertaining reference and alternate alleles > Assigning sample genotypes i.e. determining whether alleles are homozygous or heterozygous > (For SARS-CoV-2, all variant and reference alleles should be homozygous (i.e. single-stranded genome) > Variant quality control > Output = VCF file.

Alt-text Figure 29 - Example VCF output (simplified) after running a variant calling pipeline.

Screenshot of a VCF output. It highlights the reference contig: NC_045512.2, length=29903.. It also indicates homozygous alternate allele: format 1/1 in position 6517 and homozygous reference allele: format 0/0 in position 10029. Detains in the main text.

Alt-text Figure 30 - Example of a VCF file output after annotation by snpEff

Example of a VCF file output after annotation by snpEff. It shows the following information of each variant: chromosome, position, identification, reference, alteration, quality, filter, information format, sample. Details in the main text.