Whole-Genome-Sequencing (WGS) Data Delivery Specifications

Output data info

- FASTQ.GZ files containing raw sequences.

  Reads will be provided with adapter sequences masked. No quality clipping is provided with raw reads delivery.

WGS bioinformatics analysis

- Alignments in BAM format.
- SNPs and INDELs in VCF format.
- Structural Variants in VCF format.
- Copy Number Variants (CNVs) in tabular format.
- SNP and INDEL functional annotation in tabular format. For human data, additional information will be provided such as dbSNP identifiers, population allele frequencies, clinical significance, SIFT scores, conservation scores, HPO terms, *etc*.
- HTML report describing overall experiment quality (library insert size, % of aligned and duplicated reads, on-target metrics, *etc.*).
- A REPORT file describing the library preparation and analysis flow.

FAQ

Do reads contain adapters?

Unless differently agreed, reads are provided with masking of adapters read-through. When a minimum of 5bp read-through is found with respect to sample-specific (barcode included) adapters, bases are masked with N character. Thus, read length is maintained to its original size. No quality clipping is applied on raw reads delivery, while regularly used in our standard bioinformatic pipelines.

Are reads quality trimmed?

Delivered raw data are not quality trimmed. However, our internal analysis pipelines always rely on a quality trimming step which will be described in the delivery REPORT.