Exome-seq 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.

Exome-Seq bioinformatics analysis

- Alignments in BAM format.
- Variants in VCF format.
- SNP and indel summary reports containing variants with functional annotation and other information helpful for variants prioritization (dbSNP identifiers, allele frequencies, clinical significance, SIFT scores, conservation scores, HPO terms, *etc.*).
- Metrics describing overall experiment quality computed from the BAM file (library insert size, % of aligned and duplicated reads, on-target metrics, *etc.*).
- A REPORT file describing the library preparation and analysis flow.

  For family trio, matched tumor/normal pair and mitochondrial DNA variants analyses please inquire to define parameters.

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.

Why target enrichment doesn't yield even coverage distribution?

High GC content in regions such as the 5'UTR, promoter regions and the first exons of genes affect enrichment efficiency as well as repeat elements, tandem repeats and pseudogenes resulting in uneven distribution of coverage. Finally, but not less importantly, a lower quantity or lower quality of DNA is often found to introduce bias in the downstream analysis.