Output description Human Exome Analysis for Epidermiolisis Bullosa genes.

Table of Contents

Pipeline overview

- Sarek v0.11.8 - Preprocessing and variant calling pipeline.
- GATK - Variant Filtering v.3.8 - variant filtering and post-processing.
- Exomiser v.12.1.0 - Variant annotation and priorization.
- Vep v.101.0 - Variant calling annotation and effect prediction
- Picard HsMetrics v1.140 - Mapping statistics.

Note:

Depending on the analysis, we could have some ANALYSIS_IDs. This ANALYSIS_IDs are going to be composed of the date of the analysis, and an analysis identification. You can find a README in de ANALYSIS folder with a brief description of the different analysis.

Sarek: preprocessing, mapping and variant calling pipeline.

Sarek is a workflow designed to detect variants on whole genome or targeted sequencing data. Initially designed for Human, and Mouse, it can work on any species with a reference genome. Sarek can also handle tumour / normal pairs and could include additional relapses.

Output directory: `01-sarek`

- `PipelineInfo/results_description.html`
  - html report. This file can be opened in your favourite web browser (Firefox/chrome preferable) and it contains the description of all Sarek outputs.
- `Preprocessing/Recalibration/{sample_id}/{sample_id}.rc.bam`
  - bam file including mapping information that can be loaded into IGV.

Variant calling post-processing: hard-filtering

GATK is used for hard filtering following GATK best practices

Output directory: `02-postprocessing`

- `{samples_id}_variants_fil.vcf`
  - vcf file with variants tagged with filter information.

Annotation and priorization

Exomiser

Exomiser(https://www.sanger.ac.uk/tool/exomiser/) annotates and priorizates variants according to phenotype, inheritance, patogenity, etc. In this analysis variants are prioritaze based on EB gene panel (COL7A1, KRT5, KRT14, PLEC, ITGB4, LAMC2, LAMB3, LAMA3, COL17A1, FERMT1, KLHL24, DST, EXPH5, CD151, TGM5, PKP1, DSP, JUP, LAMA3A, ITGA6, ITGA3, PLOD3.TGM5, CSTA, CTSB, SERPINB8, FLG2, CDSN, CAST, DSG1, SPINK5, DSC3, DSG3, KRT1, KRT10, KRT2, KRT6A, KRT6B, KRT6C, KRT16, KRT17).

Output directory: `03-annotation/exomizer`

- `{samples_id}_exomiser.html`
  - html exomiser output.
- `{samples_id}_exomiser.json`
  - json exomiser output.
- `{samples_id}_exomiser_AD.[genes,variants].[tsv,vcf]`

    - variants/genes for Autosomal Dominant inheritance model annotated in tsv/vcf format.
- `{samples_id}_exomiser_AR.[genes,variants].[tsv,vcf]`
    - variants/genes for Autosomal Recesive inheritance model annotated in tsv/vcf format.
- `{samples_id}_exomiser_MT.[genes,variants].[tsv,vcf]`
    - variants/genes for Mitocondrial inheritance model annotated in tsv/vcf format.
- `{samples_id}_exomiser_XD.[genes,variants].[tsv,vcf]`
    - variants/genes for Sex-associated Dominant inheritance model annotated in tsv/vcf format.
- `{samples_id}_exomiser_XR.[genes,variants].[tsv,vcf]`
    - variants/genes for Sex-associated Recesive inheritance model annotated in tsv/vcf format.

VEP

VEP (Variant Effect Predictor) determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions. **Output directory:** **`03-annotation/vep`**

- `{samples_id}_final_annot.txt`
- `{samples_id}_variants_fil_mod.vcf`
- `{samples_id}_variants.table`
- `HaplotypeCaller_vep_anno_{samples_id}.vcf`
- `HaplotypeCaller_vep_anno_{samples_id}.vcf_summary.html`
- `variants_annot_all.tab`
- **`{samples_id}_variants_annot_highModerate.xlsx`**
    - Variants filtrated by high and moderate effect

Quality control stats.

Picard HsMetrics

Picard is used for coverage metrics calculation.

Output directory: **`99-stats`**

- `hs_metrics_all.csv`
    - Mean depth of coverage, capture enrichment and coverage stats.