MusVar: Output

Introduction

The output consists of the raw output files from the sarek/nf-core workflow along output from an additional caller (VarDict) plus some post processing of the mutation tables to filter for more stringent calls.

Directory Structure

```
{outdir}
├── sarek
│   ├── csv
│   ├── multiqc
│   ├── pipeline_info
│   ├── preprocessing
│   ├── reports
│   ├── tabix
│   └── variant_calling
│
└── post
    ├── reports
    └── variant_calling
```

Sarek output

A short description of some of the key files is given here. For a complete, detailed description of the Sarek/nf-core output, please see the nf-core webpage //nf-co.re/sarek/3.4.0/docs/output. First note: the `bam` files are not delivered by default to save space. If you require them, please contact the core. The `multiqc` folder contains an aggregated set of QC reports for the run. The `pipeline_info` folder includes the Nextflow execution reports, which are viewable as `html` files, along with a full manifest of all the software versions used in `software_versions.yml`. Additionally, here you will find the input manifest of the files used in the running of your project. The `variant_calling` folder contains the raw output for each sample from each of the three variant callers used: FreeBayes, Mutect2, and Strelka.

Custom Post-processing Output

The `post` folder contains the raw output of the fourth caller, VarDict, in the `variant_calling` folder, and a set of tables that are the combined results of all four callers in the reports folder. Currently, there are two files:

- `{ProjectNum}_MergedUnFiltered_MAF.txt`: This file is the full merge of all four callers in the standard MAF format. No filtering is applied, so any event called by at least one caller will be included. The output is de-duplicated, ensuring that only one row is present for mutations that were identified by multiple algorithms.
- `{ProjectNum}_mutationReport_MusVarV1.xlsx`: This Excel file contains two filtered tables. Initially, both tables have filters applied to minimize false positives:
    - `Variant Allele Frequency (VAF) >= 0.05`
    - `Mutation Allele Depth >= 8`
    - `Total Depth >= 20`
    - `Tumor VAF > 5* Normal VAF`

  Additionally, both tables exclude silent events, retaining only non-silent mutations.

  The tables differ in their levels of specificity/sensitivity:

  1. Sheet `HCEvents`: This table includes mutations that were marked `PASS` by at least two callers. It features an abbreviated layout with a minimal set of columns.
  2. Sheet `HighSensMAF`: This table includes mutations that were marked `PASS` by one or more callers. It follows the standard MAF format.