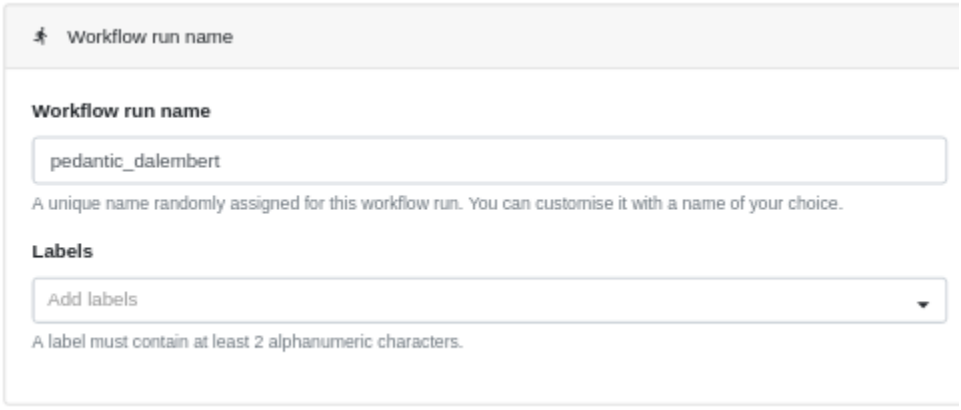
Nextflow tower pipelines - HGI_sarek

This pipeline is a fork from nf-core sarek https://github.com/nf-core/sarek modified to run on the WTSI LSF farm5 cluster.

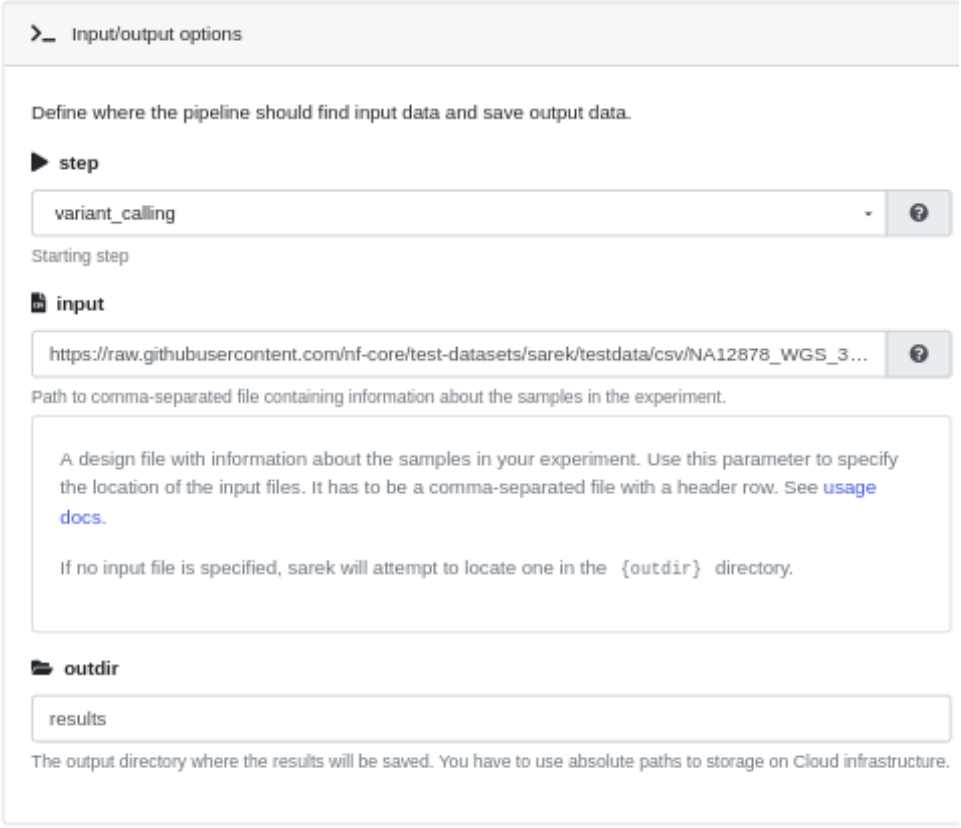
Inputs

The inputs to the pipeline can be defined on the launch page via a form generated by the nextflow_schema.json file in the repo.

- Name your workflow using a name which indicates which data is being processed

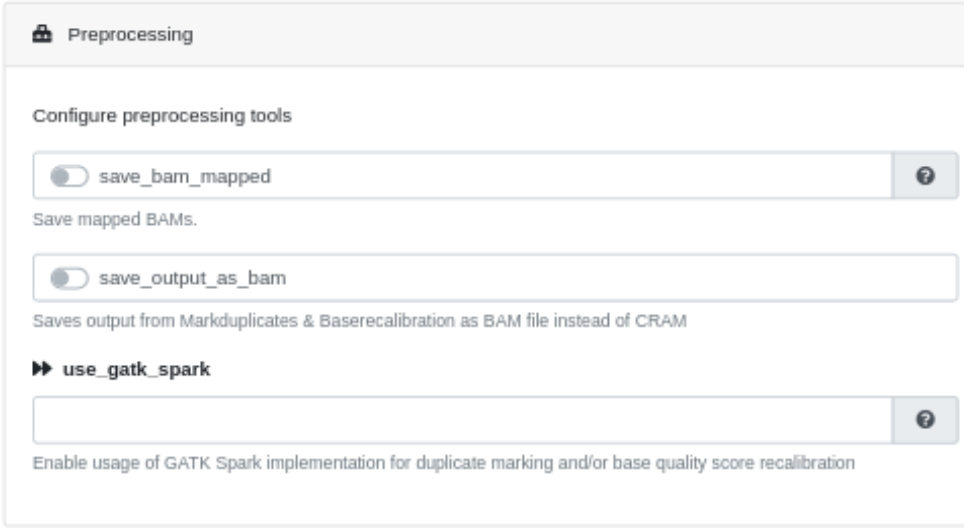


- There are several entry points (step) to the pipeline, but as of June 2023, only the variant calling entry has been tested
- Specify the CSV file containing details of the input CRAM files to be using in the variant calling process
  - CSV should have four fields "patient,sample,cram,crai" (patient and sample can be the same)

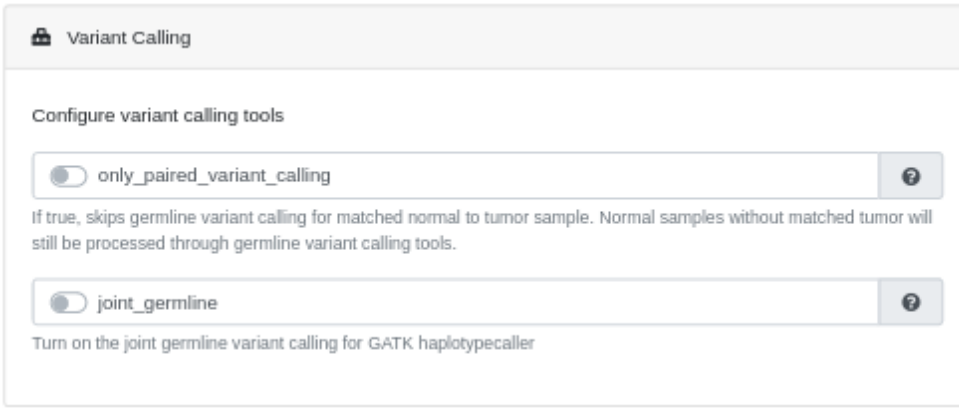


- Most of the main options are self explanatory or not important
  - split_fastq is not used if entry is at the variant calling step
  - wes should be selected if CRAMs are from exome sequencing
  - intervals is a bed format file indicating your variant calling regions
  - nucleotides_per_second is used to determine the size of the regions called so that the jobs complete in an appropriate time (40000 is appropriate for WGS and results in approx. 139 fragments for WGS calling) This number should be reduced depending on the number of input samples 40000 results in a genomicDB import time of around 6hr for 200 WGS samples and a slightly shorter time for the genotypeGVCF step. The job time is not quite linear by sample number but close.
  - the tools tested so fare are "haplotypecaller" and "deepvariant" (GATK gVCFs are only produced if joint calling is selected in the next list of parameters.)

- Ignore this section of parameters as they refer to steps in the pipeline before "variant calling"



- Turn on the joint_germline option to enable the creation of GATK gVCF files and run GATK joint calling



- Reference parameters, most of these can be left blank and the pipeline will use appropriate public reference data but enter local files for the genome fasta and fasta index e.g. /lustre/scratch125/humgen/resources/ref/Homo_sapiens/HS38DH/hs38DH.fa and /lustre /scratch125/humgen/resources/ref/Homo_sapiens/HS38DH/hs38DH.fa.fai