Galaxy Workflow ' gatk-variant-call-pe-dataset-collection'

Step **Annotation**

Step 1: Input dataset collection

input
select at runtime

Step 2: Input dataset

input
select at runtime

Step 3: FASTQ Groomer

File to groom
Output dataset 'output' from step 1

Input FASTQ quality scores type
Sanger & Illumina 1.8+

Advanced Options
Hide Advanced Options

Step 4: Bowtie2

Is this single or paired library
Paired-end Dataset Collection

FASTQ Paired Dataset
Output dataset 'output_file' from step 3

Write unaligned reads (in fastq format) to separate file(s)
False

Write aligned reads (in fastq format) to separate file(s)
False

Do you want to set paired-end options?
No

Will you select a reference genome from your history or use a built-in index?
Use a genome from the history and build index

Select reference genome
Output dataset 'output' from step 2

Set read groups information?
Set read groups (SAM/BAM specification)

Auto-assign
False

Read group identifier (ID)
gid

Auto-assign
False

Read group sample name (SM)
gsn

Platform/technology used to produce the reads (PL)
ILLUMINA

Auto-assign
False

Library name (LB)
Empty.

Sequencing center that produced the read (CN)
Empty.

Description (DS)
Empty.

Date that run was produced (DT)
Empty.

Flow order (FO)
Empty.

The array of nucleotide bases that correspond to the key sequence of each read (KS)
Empty.

Programs used for processing the read group (PG)
Empty.

Predicted median insert size (PI)
Not available.

Platform unit (PU)
Empty.

Select analysis mode
1: Default setting only

Do you want to use presets?
No, just use defaults

Save the bowtie2 mapping statistics to the history
False

Step 5: Realigner Target Creator

Choose the source for the reference list

History

BAM file
Output dataset 'output' from step 4

Using reference file
Output dataset 'output' from step 2

Known Variants

Basic or Advanced GATK options
Basic

Basic or Advanced Analysis options
Basic

Step 6: Indel Realigner

Choose the source for the reference list
History

BAM file
Output dataset 'output' from step 4

Using reference file
Output dataset 'output' from step 2

Restrict realignment to provided intervals
select at runtime

Known Variants

LOD threshold above which the realigner will proceed to realign
5.0

Use only known indels provided as RODs
False

Basic or Advanced GATK options
Basic

Basic or Advanced Analysis options
Basic

Step 7: Unified Genotyper

Choose the source for the reference list
History

BAM files

> **BAM file 1**
>
> **BAM file**
> Output dataset 'output_bam' from step 6

Using reference file
Output dataset 'output' from step 2

Provide a dbSNP Reference-Ordered Data (ROD) file

Don't set dbSNP

Genotype likelihoods calculation model to employ
BOTH

The minimum phred-scaled confidence threshold at which variants not at 'trigger' track sites should be called
30.0

The minimum phred-scaled confidence threshold at which variants not at 'trigger' track sites should be emitted (and filtered if less than the calling threshold)
30.0

Basic or Advanced GATK options
Basic

Basic or Advanced Analysis options
Basic