Running workflow "gatk-variant-call-paired-datasets-collection"

Step 1: Input dataset collection

Input Dataset Collection

type to filter

Step 2: Input dataset

Input Dataset

type to filter

Step 3: FASTQ Groomer(version 1.0.4)

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(version 2.2.6.2)

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)
Not available.

Write aligned reads (in fastq format) to separate file(s)
Not available.

Do you want to set paired-end options?
Yes

Set the minimum fragment length for valid paired-end alignments
Not available.

Set the maximum fragment length for valid paired-end alignments
500

Select the upstream/downstream mate orientations for a valid paired-end alignment against the forward reference strand
--fr

Disable no-mixed behavior
Not available.

Disable no-discordant behavior
Not available.

Allow mate dovetailing
Not available.

Disallow one mate alignment to contain another

Not available.

Disallow mate alignments to overlap
Not available.

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
Not available.

Read group identifier (ID)
group_id

Auto-assign
Not available.

Read group sample name (SM)
group_sample_name

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

Auto-assign
Not available.

Library name (LB)
Not available.

Sequencing center that produced the read (CN)
Not available.

Description (DS)
Not available.

Date that run was produced (DT)
Not available.

Flow order (FO)
Not available.

The array of nucleotide bases that correspond to the key sequence of each read (KS)
Not available.

Programs used for processing the read group (PG)
Not available.

Predicted median insert size (PI)
Not available.

Platform unit (PU)
Not available.

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
Not available.

Step 5: Realigner Target Creator(version 2.8.0)

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(version 2.8.0)

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

Known Variants

LOD threshold above which the realigner will proceed to realign
5.0

Use only known indels provided as RODs
Not available.

Basic or Advanced GATK options
Basic

Basic or Advanced Analysis options
Basic

Step 7: Unified Genotyper(version 2.8.0)

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
Set dbSNP

dbSNP ROD file

dbsnp ROD name
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

Step 8: SnpEff(version 4.0.0)

Sequence changes (SNPs, MNPs, InDels)
Output dataset 'output_vcf' from step 7

Input format
VCF

Output format
VCF (only if input is VCF)

Genome source
Named on demand

Snpff Genome Version Name (e.g. GRCh38.76)
hg19

Upstream / Downstream length
5000 bases

Set size for splice sites (donor and acceptor) in bases
2 bases

Annotation options
Nothing selected.

Use custom interval file for annotation
Selection is Optional

Only use the transcripts in this file.
Selection is Optional

Filter output
Nothing selected.

Filter out specific Effects
No

Chromosomal position
Use default (based on input type)

Text to prepend to chromosome name
Not available.

Produce Summary Stats
True

Do not report usage statistics to server
True

Send results to a new history

tools used in this workflow:

"tool_id": null,
"tool_id": null,
"tool_id": "toolshed.g2.bx.psu.edu/repos/devteam/fastq_groomer/fastq_groomer/1.0.4",
"tool_id": "toolshed.g2.bx.psu.edu/repos/devteam/bowtie2/bowtie2/2.2.6.2",
"tool_id": "toolshed.g2.bx.psu.edu/repos/iuc/gatk2/gatk2_realigner_target_creator/2.8.0",
"tool_id": "toolshed.g2.bx.psu.edu/repos/iuc/gatk2/gatk2_indel_realigner/2.8.0",
"tool_id": "toolshed.g2.bx.psu.edu/repos/iuc/gatk2/gatk2_unified_genotyper/2.8.0",
"tool_id": "toolshed.g2.bx.psu.edu/repos/iuc/snpeff/snpEff/4.0.0",