



haplotypcaller

Table of contents

Quick Start

Compatible GATK4 Command

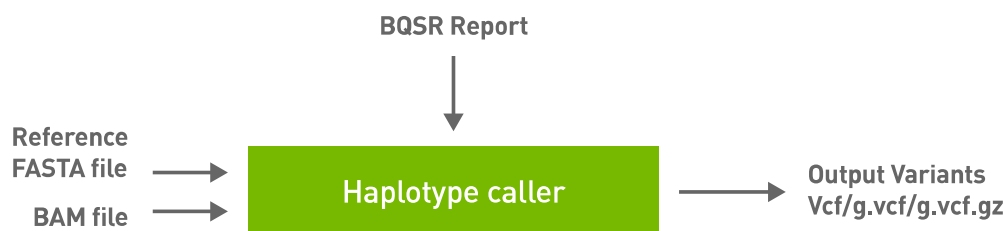
Specifying Haplotype Caller options

haplotypcaller Reference

Run a GPU-accelerated haplotypcaller.

This tool applies an accelerated GATK CollectMultipleMetrics for assessing the metrics of a BAM file, such as including alignment success, quality score distributions, GC bias, and sequencing artifacts. This functions as a 'meta-metrics' tool, and can run any combination of the available metrics tools in GATK to assess overall how well a sequencing run has performed. The available metrics tools (PROGRAMs) can be found in the command line example below.

You can provide an optional BQSR report to fix the BAM, similar to ApplyBQSR. In this case, the updated base qualities will be used.



Quick Start

```
# This command assumes all the inputs are in INPUT_DIR and all the outputs go to OUTPUT_DIR. docker run --rm --gpus all --volume INPUT_DIR:/workdir --volume OUTPUT_DIR:/outputdir \ --workdir /workdir \ nvc.io/nvidia/clara/clara-parabricks:4.3.1-1 \ pbrun haplotypcaller \ --ref /workdir/${REFERENCE_FILE} \ --in-bam /workdir/${INPUT_BAM} \ --in-recal-file /workdir/${INPUT_RECAL_FILE} \ --out-variants /outputdir/${OUTPUT_VCF}
```

Compatible GATK4 Command

The commands below are the GATK4 counterpart of the Parabricks command above. The output from these commands will be identical to the output from the above command. See the [Output Comparison](#) page for comparing the results.

```
# Run ApplyBQSR Step $ gatk ApplyBQSR \ --java-options -Xmx30g \ -R
Ref/Homo_sapiens_assembly38.fasta \ -l mark_dups_cpu.bam \ --bqsr-recal-file
recal_file.txt \ -O cpu_nodups_BQSR.bam #Run Haplotype Caller $ gatk
HaplotypeCaller \ --java-options -Xmx30g \ --input cpu_nodups_BQSR.bam \ --output
result_cpu.vcf \ --reference Ref/Homo_sapiens_assembly38.fasta \ --native-pair-
hmm-threads 16
```

Specifying Haplotype Caller options

Several original HaplotypeCaller options are supported by Parabricks. To specify the inclusion or exclusion of several haplotype caller annotations, use the

`--haplotypcaller-options` option:

```
$ # This command assumes all the inputs are in INPUT_DIR and all the outputs go to
OUTPUT_DIR. docker run --rm --gpus all --volume INPUT_DIR:/workdir --volume
OUTPUT_DIR:/outputdir \ --workdir /workdir \ nvr.io/nvidia/clara/clara-
parabricks:4.3.1-1 \ pbrun haplotypcaller \ ... --haplotypcaller-options '-min-
pruning 4 -A AS_BaseQualityRankSumTest -A TandemRepeat' ...
```

Annotations may be excluded in the same manner using the `-AX` option. There should be a space between the `-A` / `-AX` flag and its value.

The following are supported options and their allowed values:

- -A
 - AS_BaseQualityRankSumTest
 - AS_FisherStrand
 - AS_InbreedingCoeff
 - AS_MappingQualityRankSumTest
 - AS_QualByDepth
 - AS_RMSMappingQuality
 - AS_ReadPosRankSumTest
 - AS_StrandOddsRatio
 - BaseQualityRankSumTest
 - ChromosomeCounts
 - ClippingRankSumTest
 - Coverage
 - DepthPerAlleleBySample
 - DepthPerSampleHC
 - ExcessHet
 - FisherStrand
 - InbreedingCoeff
 - MappingQualityRankSumTest
 - QualByDepth
 - RMSMappingQuality
 - ReadPosRankSumTest

- ReferenceBases
- StrandBiasBySample
- StrandOddsRatio
- TandemRepeat
- -AX
 - (same as for the -A option)
- --output-mode
 - EMIT_VARIANTS_ONLY
 - EMIT_ALL_CONFIDENT_SITES
 - EMIT_ALL_ACTIVE_SITES
- *-max-reads-per-alignment-start*
 - a positive integer
- *-min-dangling-branch-length*
 - a positive integer
- *-min-pruning*
 - a positive integer
- *-pcr-indel-model*
 - NONE
 - HOSTILE
 - AGGRESSIVE
 - CONSERVATIVE

- *-standard-min-confidence-threshold-for-calling*
 - a positive integer

haplotypcaller Reference

Run HaplotypeCaller to convert BAM/CRAM to VCF.

Input/Output file options

`--ref REF`

Path to the reference file. (default: None)

Option is required.

`--in-bam IN_BAM`

Path to the input BAM/CRAM file for variant calling. The argument may also be a local folder containing several BAM files. (default: None)

Option is required.

`--in-recal-file IN_RECAL_FILE`

Path to the input BQSR report. (default: None)

`--interval-file INTERVAL_FILE`

Path to an interval file in one of these formats: Picard-style (`.interval_list` or `.picard`), GATK-style (`.list` or `.intervals`), or BED file (`.bed`). This option can be used multiple times. (default: None)

`--htvc-bam-output HTVC_BAM_OUTPUT`

File to which assembled haplotypes should be written. (default: None)

`--out-variants OUT_VARIANTS`

Path of the `vcf/g.vcf/gvcf.gz` file after variant calling. (default: None)

Option is required.

Tool Options:

`--haplotypcaller-options HAPLOTYPECALLER_OPTIONS`

Pass supported haplotype caller options as one string. The following are currently supported original haplotypcaller options: `-A <AS_BaseQualityRankSumTest, AS_FisherStrand, AS_InbreedingCoeff, AS_MappingQualityRankSumTest, AS_QualByDepth, AS_RMSMappingQuality, AS_ReadPosRankSumTest, AS_StrandOddsRatio, BaseQualityRankSumTest, ChromosomeCounts, ClippingRankSumTest, Coverage, DepthPerAlleleBySample, DepthPerSampleHC, ExcessHet, FisherStrand, InbreedingCoeff, MappingQualityRankSumTest, QualByDepth, RMSMappingQuality, ReadPosRankSumTest, ReferenceBases, StrandBiasBySample, StrandOddsRatio, TandemRepeat>`, `-AX <same options as -A>`, `--output-mode <EMIT_VARIANTS_ONLY, EMIT_ALL_CONFIDENT_SITES, EMIT_ALL_ACTIVE_SITES>`, `-max-reads-per-alignment-start <int>`, `-min-dangling-branch-length <int>`, `-min-pruning <int>`, `-pcr-indel-model <NONE, HOSTILE, AGGRESSIVE, CONSERVATIVE>`, `-standard-min-confidence-threshold-for-calling <int>` (e.g. `--haplotypcaller-options="-min-pruning 4 -standard-min-confidence-threshold-for-calling 30"`). (default: None)

`--static-quantized-quals STATIC_QUANTIZED_QUALS`

Use static quantized quality scores to a given number of levels. Repeat this option multiple times for multiple bins. (default: None)

`--gvcf`

Generate variant calls in .gvcf Format. (default: None)

`--disable-read-filter DISABLE_READ_FILTER`

Disable the read filters for BAM entries. Currently, the supported read filters that can be disabled are `MappingQualityAvailableReadFilter`, `MappingQualityReadFilter`, `NotSecondaryAlignmentReadFilter`, and `WellformedReadFilter`. (default: None)

`--max-alternate-alleles MAX_ALTERNATE_ALLELES`

Maximum number of alternate alleles to genotype. (default: None)

`-G ANNOTATION_GROUP, --annotation-group ANNOTATION_GROUP`

The groups of annotations to add to the output variant calls. Currently supported annotation groups are StandardAnnotation, StandardHCAnnotation, and AS_StandardAnnotation. (default: None)

`-GQB GVCF_GQ_BANDS, --gvcf-gq-bands GVCF_GQ_BANDS`

Exclusive upper bounds for reference confidence GQ bands. Must be in the range [1, 100] and specified in increasing order. (default: None)

`--rna`

Run haplotypcaller optimized for RNA data. (default: None)

`--dont-use-soft-clipped-bases`

Don't use soft clipped bases for variant calling. (default: None)

`--minimum-mapping-quality MINIMUM_MAPPING_QUALITY`

Minimum mapping quality to keep (inclusive). (default: None)

`--mapping-quality-threshold-for-genotyping
MAPPING_QUALITY_THRESHOLD_FOR_GENOTYPING`

Control the threshold for discounting reads from the genotyper due to mapping quality after the active region detection and assembly steps but before genotyping. (default: None)

`--enable-dynamic-read-disqualification-for-genotyping`

Will enable less strict read disqualification low base quality reads. (default: None)

`--no-alt-contigs`

Get rid of output records for alternate contigs. (default: None)

`--ploidy PLOIDY`

Ploidy assumed for the BAM file. Currently only haploid (ploidy 1) and diploid (ploidy 2) are supported. (default: 2)

`-L INTERVAL, --interval INTERVAL`

Interval within which to call the variants from the BAM/CRAM file. All intervals will have a padding of 100 to get read records, and overlapping intervals will be combined. Interval files should be passed using the `--interval-file` option. This option can be used multiple times (e.g. "`-L chr1 -L chr2:10000 -L chr3:20000+ -L chr4:10000-20000`"). (default: None)

`-ip INTERVAL_PADDING, --interval-padding INTERVAL_PADDING`

Amount of padding (in base pairs) to add to each interval you are including. (default: None)

`--sample-sex SAMPLE_SEX`

Sex of the sample input. This option will override the sex determined from any X/Y read ratio range. Must be either male or female. (default: None)

`--range-male RANGE_MALE`

Inclusive male range for the X/Y read ratio. The sex is declared male if the actual ratio falls in the specified range. Syntax is "`<min>-<max>`" (e.g. "`--range-male 1-10`"). (default: None)

`--range-female RANGE_FEMALE`

Inclusive female range for the X/Y read ratio. The sex is declared female if the actual ratio falls in the specified range. Syntax is "`<min>-<max>`" (e.g. "`--range-female 150-250`"). (default: None)

`--use-GRCh37-regions`

Use the pseudoautosomal regions for GRCh37 reference types. This flag should be used for GRCh37 and UCSC hg19 references. By default, GRCh38 regions are used.

(default: None)

Performance Options:

`--htvc-low-memory`

Use low memory mode in htvc. (default: None)

--num-htvc-threads NUM_HTVC_THREADS

Number of CPU threads. (default: 5)

--run-partition

Divide the whole genome into multiple partitions and run multiple processes at the same time, each on one partition. (default: None)

Common options:

--logfile LOGFILE

Path to the log file. If not specified, messages will only be written to the standard error output. (default: None)

--tmp-dir TMP_DIR

Full path to the directory where temporary files will be stored.

--with-petagene-dir WITH_PETAGENE_DIR

Full path to the PetaGene installation directory. By default, this should have been installed at /opt/petagene. Use of this option also requires that the PetaLink library has been preloaded by setting the LD_PRELOAD environment variable. Optionally set the PETASUITE_REFPATH and PGCLOUD_CREDPATH environment variables that are used for data and credentials (default: None)

--keep-tmp

Do not delete the directory storing temporary files after completion.

--no-seccomp-override

Do not override seccomp options for docker (default: None).

--version

View compatible software versions.

GPU options:

`--num-gpus NUM_GPUS`

Number of GPUs to use for a run. GPUs 0..(NUM_GPUS-1) will be used.

Note

In the values provided to `--haplotypcaller-options` `--output-mode` requires two leading hyphens, while all other values take a single hyphen.

© Copyright 2024, Nvidia.. PDF Generated on 06/05/2024