applybqsr

Table of contents

Updates the Base Quality Scores using the BQSR report.

This tool recalibrates quality scores in a BAM file using the report generated by the bqsr [link] tool. This should be applied after alignment but before variant calling to maximize final accuracy in variant calling, as recommended by GATK best practices.

Please note that the **applybqsr** tool will use at most two GPUs.

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 \ nvcr.io/nvidia/clara/clara-parabricks:4.3.1-1 \ pbrun applybqsr \ --ref /workdir/${REFERENCE_FILE} \ --in-bam /workdir/${INPUT_BAM} \ --in-recal-file /workdir/${INPUT_RECAL_FILE} \ --out-bam /outputdir/${OUTPUT_BAM}
```

Compatible GATK4 Command

The command below is the GATK4 counterpart of the Parabricks command above. The output from these commands will be identical to the output from the above command.

```
$ gatk ApplyBQSR \ --java-options -Xmx30g \ -R <INPUT_DIR>/${REFERENCE_FILE} \ -I <INPUT_DIR>/${INPUT_BAM} \ --bqsr-recal-file <INPUT_DIR>/${INPUT_RECAL_FILE} \ -O <OUTPUT_DIR>/${OUTPUT_BAM}
```

applybqsr Reference

Update the Base Quality Scores using the BQSR report.

Input/Output file options

--ref REF

Path to the reference file. (default: None)

Option is required.

--in-bam IN_BAM

Path to the BAM file. (default: None)

Option is required.

--in-recal-file IN_RECAL_FILE

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

Option is required.

--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)

--out-bam OUT_BAM

Output BAM file. (default: None)

Option is required.

Tool Options:

-L INTERVAL, --interval INTERVAL

Interval within which to call applyBQSR from the input reads. 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)

Performance Options:

--num-threads NUM_THREADS

Number of threads for worker. (default: 8)

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.