bammetrics

Table of contents

Accelerated GATK4 CollectWGSMetrics.

This tool applies an accelerated version of the GATK CollectWGSMetrics for assessing coverage and quality of an aligned whole-genome BAM file. This includes metrics such as the fraction of reads that pass the base and mapping quality filters, and the coverage levels (read-depth) across the genome. These act as an overall quality check for the user, allowing assessment of how well a sequencing run has performed.

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 bammetrics \ --ref /workdir/${REFERENCE_FILE} \ --bam /workdir/${INPUT_BAM} \ --out-metrics-file /outputdir/${METRICS_FILE}
```

Compatible GATK4 Command

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

```
$ gatk CollectWgsMetrics \ -R <INPUT_DIR>/${REFERENCE_FILE} \ -I <INPUT_DIR>/${INPUT_BAM} \ -O <OUTPUT_DIR>/${METRICS_FILE}
```

bammetrics Reference

Run bammetrics on a BAM file to generate a metrics file.

Input/Output file options

--ref REF

Path to the reference file. (default: None)

Option is required.

--bam BAM

Path to the BAM 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-metrics-file OUT_METRICS_FILE

Output Metrics File. (default: None)

Option is required.

Tool Options:

--minimum-base-quality MINIMUM_BASE_QUALITY

Minimum base quality for a base to contribute coverage. (default: 20)

--minimum-mapping-quality MINIMUM_MAPPING_QUALITY

Minimum mapping quality for a read to contribute coverage. (default: 20)

--count-unpaired

If true, count unpaired reads and paired reads with one end unmapped. (default: None)

--coverage-cap COVERAGE_CAP

Treat positions with coverage exceeding this value as if they had coverage at this value (but calculate the difference for PCT_EXC_CAPPED). (default: 250)

-L INTERVAL, --interval INTERVAL

Interval within which to collect metrics from the BAM/CRAM file. All intervals will have a padding of 0 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)

Performance Options:

--num-threads NUM_THREADS

Number of threads to run. (default: 12)

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.