Comparison with Baseline Tools

Table of contents

Many users want to compare output generated by NVIDIA Parabricks software with other standard tools. We recommend the following way to compare output generated by Parabricks software and the counterpart non-accelerated software.

Comparing BAM Files

GATK4 sorts the BAM files based on QNAME, FLAG, RNAME, POS, MAPQ, MRNM/RNEXT, MPOS/PNEXT, and ISIZE. If all these fields are the same for two different BAMs, they are considered equal for sorting purposes. Therefore, to compare two sorted BAMs, use the BamUtil diff tool to compare these fields:.

```
$ bam diff --in1 mark_dups_gpu.bam --in2 mark_dups_cpu.bam --noCigar --isize --flag --mate --mapQual
```

The output of this comparison should result in no differences.

Comparing BQSR Reports

The files generated by Parabricks and GATK4 should be exactly the same. There should be no output from the following command:

```
$ diff -w recal_gpu.txt recal_cpu.txt
```

Comparing VCF Files

To compare VCF files, use the GATK Concordance tools to get sensitivity and specificity of SNPs and INDELs. When the following command is run, variant accuracy results will be stored in out.txt .

```
$ gatk Concordance \ --evaluation result_gpu.vcf \ --truth result_cpu.vcf \ --summary out.txt
```