



FQ2BAM Tutorial

This tutorial will show you how to run our core alignment tool, FQ2BAM, which allows you to align a FASTQ file according to GATK best practices at blazing speeds. This includes the gold-standard alignment tool BWA-MEM with inbuilt co-ordinate sorting of the output file, and optionally application of base-quality-score-recalibration and marking of duplicate reads.

The `fq2bam` tool aligns, sorts (by coordinate), and marks duplicates in paired-end FASTQ file data. The data files used in this example are taken from the sample data downloaded in the previous section.

If you execute the following command using the NVIDIA Parabricks sample data, you should get the same results as shown here.

Before executing this command, make sure your current directory is where you extracted the sample data; it should have a **parabricks_sample** sub-directory.

```
$ docker run \ --gpus all \ --rm \ --volume $(pwd):/workdir \ --volume
$(pwd):/outputdir \ nvcv.io/nvidia/clara/clara-parabricks:4.3.1-1 \ pbrun fq2bam \ --
ref /workdir/parabricks_sample/Ref/Homo_sapiens_assembly38.fasta \ --in-fq
/workdir/parabricks_sample/Data/sample_1.fq.gz
/workdir/parabricks_sample/Data/sample_2.fq.gz \ --out-bam
/outputdir/fq2bam_output.bam [Parabricks Options Mesg]: Checking argument
compatibility [Parabricks Options Mesg]: Automatically generating ID prefix
[Parabricks Options Mesg]: Read group created for
/workdir/parabricks_sample/Data/sample_1.fq.gz and
/workdir/parabricks_sample/Data/sample_2.fq.gz [Parabricks Options Mesg]:
@RG\tID:HK3TJBCX2.1\tLB:lib1\tPL:bar\tSM:sample\tPU:HK3TJBCX2.1 [PB Info 2022-
Sep-02 19:49:27] ----- [PB Info
2022-Sep-02 19:49:27] || Parabricks accelerated Genomics Pipeline || [PB Info
2022-Sep-02 19:49:27] || Version 4.0.0-1 || [PB Info 2022-Sep-02 19:49:27] || GPU-
BWA mem, Sorting Phase-I || [PB Info 2022-Sep-02 19:49:27] -----
----- [M::bwa_idx_load_from_disk] read 0 ALT contigs [PB
Warning 2022-Sep-02 19:50:02][ParaBricks/src/pbOpts.cu:325] WARNING The
system has 12 threads, however recommended number of threads with 1 GPU is
16. The run might not finish or might have less than expected performance. [PB Info
2022-Sep-02 19:50:02] GPU-BWA mem [PB Info 2022-Sep-02 19:50:02]
ProgressMeter Reads Base Pairs Aligned [PB Info 2022-Sep-02 19:50:45] 5043564
```

```

580000000 [PB Info 2022-Sep-02 19:51:21] 10087128 1160000000 [PB Info 2022-
Sep-02 19:51:59] 15130692 1740000000 [PB Info 2022-Sep-02 19:52:39] 20174256
2320000000 [PB Info 2022-Sep-02 19:53:20] 25217820 2900000000 [PB Info 2022-
Sep-02 19:53:58] 30261384 3480000000 [PB Info 2022-Sep-02 19:54:36] 35304948
4060000000 [PB Info 2022-Sep-02 19:55:13] 40348512 4640000000 [PB Info 2022-
Sep-02 19:55:53] 45392076 5220000000 [PB Info 2022-Sep-02 19:56:36] 50435640
5800000000 [PB Info 2022-Sep-02 19:57:02] GPU-BWA Mem time: 420.426442
seconds [PB Info 2022-Sep-02 19:57:02] GPU-BWA Mem is finished. [main] CMD:
/usr/local/parabricks/binaries/bin/bwa mem -Z ./pbOpts.txt
/workdir/parabricks_sample/Ref/Homo_sapiens_assembly38.fasta
/workdir/parabricks_sample/Data/sample_1.fq.gz
/workdir/parabricks_sample/Data/sample_2.fq.gz
@RG\tID:HK3TJBCX2.1\tLB:lib1\tPL:bar\tSM:sample\tPU:HK3TJBCX2.1 [main] Real
time: 455.468 sec; CPU: 4766.384 sec [PB Info 2022-Sep-02 19:57:02] -----
----- [PB Info 2022-Sep-02 19:57:02] || Program:
GPU-BWA mem, Sorting Phase-I || [PB Info 2022-Sep-02 19:57:02] || Version: 4.0.0-
1 || [PB Info 2022-Sep-02 19:57:02] || Start Time: Fri Sep 2 19:49:27 2022 || [PB
Info 2022-Sep-02 19:57:02] || End Time: Fri Sep 2 19:57:02 2022 || [PB Info 2022-
Sep-02 19:57:02] || Total Time: 7 minutes 35 seconds || [PB Info 2022-Sep-02
19:57:02] ----- [PB Info 2022-Sep-
02 19:57:03] ----- [PB Info 2022-
Sep-02 19:57:03] || Parabricks accelerated Genomics Pipeline || [PB Info 2022-Sep-
02 19:57:03] || Version 4.0.0-1 || [PB Info 2022-Sep-02 19:57:03] || Sorting Phase-II
|| [PB Info 2022-Sep-02 19:57:03] -----
----- [PB Info 2022-Sep-02 19:57:03] progressMeter - Percentage [PB Info 2022-Sep-
02 19:57:03] 0.0 0.00 GB [PB Info 2022-Sep-02 19:57:13] 72.8 0.00 GB [PB Info 2022-
Sep-02 19:57:23] Sorting and Marking: 20.001 seconds [PB Info 2022-Sep-02
19:57:23] ----- [PB Info 2022-Sep-
02 19:57:23] || Program: Sorting Phase-II || [PB Info 2022-Sep-02 19:57:23] ||
Version: 4.0.0-1 || [PB Info 2022-Sep-02 19:57:23] || Start Time: Fri Sep 2 19:57:03
2022 || [PB Info 2022-Sep-02 19:57:23] || End Time: Fri Sep 2 19:57:23 2022 || [PB
Info 2022-Sep-02 19:57:23] || Total Time: 20 seconds || [PB Info 2022-Sep-02
19:57:23] ----- [PB Info 2022-Sep-
02 19:57:23] ----- [PB Info 2022-
Sep-02 19:57:23] || Parabricks accelerated Genomics Pipeline || [PB Info 2022-Sep-
02 19:57:23] || Version 4.0.0-1 || [PB Info 2022-Sep-02 19:57:23] || Marking

```


