



deepsomatic

Table of contents

What is DeepSomatic?

Quick Start

Compatible Google DeepVariant Commands

deepsomatic Reference

GPU-accelerated DeepSomatic.

What is DeepSomatic?

DeepSomatic builds on the deep learning-based variant caller DeepVariant. It processes aligned reads from tumor and normal samples (in BAM or CRAM format), generates pileup image tensors, classifies these tensors using a convolutional neural network, and outputs somatic variants in standard VCF or gVCF files.

DeepSomatic is designed for somatic variant calling using tumor-normal sequencing data.

Parabricks has enhanced Google DeepSomatic to leverage GPUs extensively. The Parabricks version of DeepSomatic operates similarly to other common command line tools: it accepts two BAM files and a reference file as inputs and generates variants in a VCF file as output.

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 deepsomatic \ --ref /workdir/${REFERENCE_FILE} \ --in-tumor-bam /workdir/${INPUT_TUMOR_BAM} \ --in-normal-bam /workdir/${INPUT_NORMAL_BAM} \ --out-variants /outputdir/${OUTPUT_VCF}
```

Compatible Google DeepVariant Commands

The commands below are the Google 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.

```
docker run \ --interactive \ --tty \ --rm \ --volume ${INPUT_DIR}:${INPUT_DIR} \ --volume ${OUTPUT_DIR}:${OUTPUT_DIR} \ --workdir /workdir
```

```
google/deepvariant:1.6.1 \ run_deepvariant \ --ref ${REFERENCE_FILE} \ --
reads_tumor ${TUMOR_BAM} \ --reads_normal ${NORMAL_BAM} \ --
customized_model ${DEEPSOMATIC_WGS_MODEL_FILE} \ --output_vcf
${OUTPUT_VCF} \ --make_examples_extra_args
"ws_use_window_selector_model=true" \ --num_shards=$(nproc)
```

deepvariant Reference

Run DeepVariant to convert BAM/CRAM to VCF.

Input/Output file options

--ref REF

Path to the reference file. (default: None)

Option is required.

--in-tumor-bam IN_TUMOR_BAM

Path to the input tumor BAM/CRAM file for somatic variant calling. (default: None)

Option is required.

--in-normal-bam IN_NORMAL_BAM

Path to the input normal BAM/CRAM file for somatic variant calling. (default: None)

Option is required.

--interval-file INTERVAL_FILE

Path to a BED file (.bed) for selective access. This option can be used multiple times. (default: None)

--out-variants OUT_VARIANTS

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

Option is required.

`--pb-model-file PB_MODEL_FILE`

Path to a non-default parabricks model file for deepomatic. (default: None)

Tool Options:

`--disable-use-window-selector-model`

Change the window selector model from Allele Count Linear to Variant Reads. This option will increase the accuracy and runtime. (default: None)

`--no-channel-insert-size`

If True, don't add `insert_size` channel into the pileup image. (default: False)

`-L INTERVAL, --interval INTERVAL`

Interval within which to call the variants from the BAM/CRAM file. 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-cpu-threads-per-stream NUM_CPU_THREADS_PER_STREAM`

Number of CPU threads to use per stream. (default: 6)

`--num-streams-per-gpu NUM_STREAMS_PER_GPU`

Number of streams to use per GPU. (default: 2)

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-petogene-dir WITH_PETAGENE_DIR

Full path to the PetaGene installation directory. By default, this should have been installed at /opt/petogene. 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.