



Germline Copy-Number Variant Calling with CNVscope

Release 202308.03

Sentieon, Inc

Oct 23, 2024

Contents

1	Introduction	1
2	Germline CNV calling with CNVscope	1
2.1	Basic usage of CNVscope	1
2.2	Limitations of CNVscope machine learning model	2

1 Introduction

This document describes the capabilities of CNVscope for germline copy number variation (CNV) calling for whole-genome sequencing (WGS). If you have any additional questions, please contact the technical support at Sentieon® Inc. at support@sentieon.com.

2 Germline CNV calling with CNVscope

2.1 Basic usage of CNVscope

Two individual commands are run to call CNV and to apply the machine learning model. The input BAM file should come from a pipeline where alignment and deduplication have been performed.

```
sentieon driver -t NUMBER_THREADS -r REFERENCE -i DEDUPED_BAM \  
--algo CNVscope --model ML_MODEL/cnv.model TMP_VARIANT_VCF  
  
sentieon driver -t NUMBER_THREADS -r REFERENCE --algo CNVModelApply \  
--model ML_MODEL/cnv.model -v TMP_VARIANT_VCF VARIANT_VCF
```

Reminder

It is important to use the same model for CNVscope and CNVModelApply. If different models are used, CNVModelApply will give an error.

The following inputs are required for the command:

- **NUMBER_THREADS**: the number of computer threads that will be used in the calculation. We recommend that the number does not exceed the number of computing cores available in your system.
- **REFERENCE**: the location of the reference FASTA file. You should make sure that the reference is the same as the one used in the mapping stage.
- **DEDUPED_BAM**: the location of the input BAM file.
- **TMP_VARIANT_VCF**: the location and filename of the variant calling output of CNVscope. This is a temporary file.
- **VARIANT_VCF**: the location and filename of the variant calling output. A corresponding index file will be created. The tool will output a compressed file by using .gz extension.
- **ML_MODEL**: the location of the machine learning model file. In the CNVscope command the model will be used to determine the settings used in variant calling.

The final output VCF file uses CN annotation for the copy-number state for each region called by CNVscope machine-learning model. Possible copy-number states called by CNVscope are from 0 to 4, with CN=4 representing copy number states equal to or larger than 4.

2.2 Limitations of CNVscope machine learning model

Currently, CNVscope is trained on diploid WGS samples with CNV event size above 5000 base pairs, and thus should only be used for these scenarios.