



Description of output files and fields

Release 202112.07

Sentieon, Inc

Apr 24, 2023

Contents

1	Introduction	1
2	TNsnv	2
2.1	Introduction	2
2.2	OUTPUT.VCF	2
2.3	CALL_STATS_OUTPUT.TXT	3
2.4	STD_COVERAGE.TXT	5
2.5	Q20_COVERAGE.TXT	5
2.6	POWER.TXT	5
2.7	TUMOR_DP.TXT	5
2.8	NORMAL_DP.TXT	5
3	TNhaplotyper	5
3.1	Introduction	5
3.2	OUTPUT.VCF	5
4	TNhaplotyper2	7
4.1	Introduction	7
4.2	OUTPUT.VCF	7
5	TNscope®	9
5.1	Introduction	9
5.2	OUTPUT.VCF	10

1 Introduction

This document describes the output files of Sentieon® TNsnv, TNhaplotyper, TNhaplotyper2 and TNscope® algorithms and the meaning of the fields in those files. You can use the information in this document to better understand the files produced by Sentieon® tumor-normal variant calling software.

2 TNSnv

2.1 Introduction

An example command with TNSnv is as follows

```
sentieon driver -t NUMBER_THREDS -r REFERENCE.FASTA \  
-i NORMAL_RECALED.BAM -i TUMOR_RECALED.BAM \  
--interval INTERVAL \  
--algo TNSnv --dbsnp DBSNP.VCF \  
--tumor_sample TUMOR_SM --normal_sample NORMAL_SM \  
-call_stats_out CALL_STATS_OUTPUT.TXT \  
--stdcov_out STD_COVERAGE.TXT \      Standard coverage output file \  
--q20cov_out Q20_COVERAGE.TXT \      Q20 coverage output file \  
--power_out POWER.TXT --tumor_depth_out TUMOR_DP.TXT \  
--normal_depth_out NORMAL_DP.TXT OUTPUT.VCF
```

This command line produces the following required output files:

- **OUTPUT.VCF**

In addition, the following optional output files are produced:

- **CALL_STATS_OUTPUT.TXT**
- **STD_COVERAGE.TXT**
- **Q20_COVERAGE.TXT**
- **POWER.TXT**
- **TUMOR_DP.TXT**
- **NORMAL_DP.TXT**

The **OUTPUT.VCF** of TNSnv contains only limited output information. Users who desired a more detailed output format should examine the **CALL_STATS_OUTPUT.TXT** file.

2.2 OUTPUT.VCF

The **OUTPUT.VCF** file conforms to the VCF 4.2 specification. More information on the VCF format can be found at <https://samtools.github.io/hts-specs/VCFv4.2.pdf>. The INFO field annotations are described in detail below.

INFO annotation	Description
DB	The variant is present in the VCF file supplied with the <code>-dbsnp</code> option
MQ0	Total number of reads with Mapping Quality equal to 0
SOMATIC	The variant occurs uniquely in the sample supplied with the <code>-tumor_sample</code> option
VT	Variant type, can be SNP, INS or DEL

TNSnv also populates the FILTER field of the output VCF file. Variants are filtered using TNSnv internal quality filters. More information on the applied filters can be found in the failure_reasons row in the table in section 2.3.

FILTER	Description
PASS	The variant passes TNSnv internal quality filters
REJECT	The variant fails TNSnv internal quality filters

Standard genotype fields are defined by the format specification. However, TNSnv also outputs the following non-standard fields.

GENOTYPE field	Description
BQ	Average base quality of bases supporting the alternate alleles
FA	Fraction of reads supporting the alternate allele
SS	Status of the variant. Not currently implemented, always set to 2

2.3 CALL_STATS_OUTPUT.TXT

The **CALL_STATS_OUTPUT.TXT** file is a tab-separated text file with the following columns for each candidate variant. The core statistic of the software is `t_lod_fstar` which is a measurement of the support for the mutation relative to the expected level of sequencing noise at the candidate site.

Column	Description
Contig	The contig (chromosome) with the candidate
Position	The genomic coordinate of the candidate along the contig
Context	The sequence 3bp to either side of the candidate
Ref_allele	The reference allele at the candidate site
Alt_allele	The alternate allele at the candidate site
Tumor_name	The name of the tumor sample with the candidate mutation
Normal_name	The name of the paired normal sample
Score	Variant score. Not currently implemented, always set to 0.0
Dbsnp_site	The variant is present in the VCF file supplied with the <code>-dbsnp</code> option (DBSNP) or is novel (NOVEL)
Covered	The site has sufficient read coverage to detect a variant with a 0.3 allele fraction at 80% power
Power	The product of tumor power and normal power, described below.
Tumor_power	The power to detect a mutation at a 0.3 allele fraction at the observed sequencing depth in the tumor sample
Normal_power	The power to detect a germline mutation at this site taking into account the presence of the site in dbSNP at the observed sequencing depth in the normal sample
Normal_power_nsp	The power to detect a germline mutation in the normal sample given that the mutation is not in dbSNP
Normal_power_wsp	The power to detect a germline mutation in the normal sample given that the mutation is in dbSNP
Total_reads	Total number of reads in both the tumor and normal samples at this site
Map_Q0_reads	Total number of reads in both the tumor and normal samples with mapping quality 0 at this site
Init_t_lod	Log odds of the likelihood that the candidate mutation is real over the likelihood that the candidate mutation is a sequencing error before any read-based filters are applied
t_lod_fstar	Log odds of the likelihood that the candidate mutation is real over the likelihood that the candidate mutation is a sequencing error
t_lod_fstar_forward	t_lod_fstar calculated using only reads on the forward strand
t_lod_fstar_reverse	t_lod_fstar calculated using only reads on the reverse strand
tumor_f	Estimated allele fraction of the candidate mutation in the tumor sample
Contaminant_fraction	Estimate of contamination of normal cells in the tumor sample
Contaminant_lod	Log odds of the likelihood that the candidate is contamination over the likelihood that the candidate is a sequencing error
t_q20_count	Count of the number of reads in the tumor sample with a base quality of at least 20
t_ref_count	Number of reads supporting the reference allele in the tumor sample
t_alt_count	Number of reads supporting the alternate allele in the tumor sample

Continued on next page

Table 2.1 – continued from previous page

Column	Description
t_ref_sum	Sum of the quality scores of the bases supporting the reference allele in the tumor sample
t_alt_sum	Sum of the quality scores of the bases supporting the alternate allele in the tumor sample
t_ref_max_mapq	The maximum mapping quality of tumor reads supporting the reference allele
t_alt_max_mapq	The maximum mapping quality of tumor reads supporting the alternate allele
t_ins_count	The number of reads in the tumor sample that have an insertion in the surrounding five bases
t_del_count	The number of reads in the tumor sample that have a deletion in the surrounding five bases
Normal_best_gt	The most likely genotype of the normal sample
Init_n_lod	Log odds of the likelihood that the normal sample is reference over the normal sample having the variant before any read-based filters are applied
normal_f	Estimated allele fraction of the candidate mutation in the normal sample
n_q20_count	Count of the number of reads in the normal sample with a base quality of at least 20
n_ref_count	Number of reads supporting the reference allele in the normal sample
n_alt_count	Number of reads supporting the alternate allele in the normal sample
n_ref_sum	Sum of the quality scores of the bases supporting the reference allele in the normal sample
n_alt_sum	Sum of the quality scores of the bases supporting the alternate allele in the normal sample
power_to_detect_positive_strand_bias	The power to detect strand bias to the positive strand at the given sequencing depth
power_to_detect_negative_strand_bias	The power to detect strand bias to the negative strand at the given sequencing depth
strand_bias_counts	A vector of counts for the tumor sample in the order of (tumor_ref_pos, tumor_ref_neg, tumor_alt_pos, tumor_alt_neg) where ref and alt specify the reference and alternate alleles and pos and neg specify the positive and negative strands. The numbers do not match those in earlier columns due to differential filtering
tumor_alt_fpir_median	Median position along forward strand reads for bases supporting the alternate allele in the tumor sample
tumor_alt_fpir_mad	Mean absolute deviation of the positions along forward strand reads for bases supporting the alternate allele in the tumor sample
tumor_alt_rpir_median	Median position along reverse strand reads for bases supporting the alternate allele in the tumor sample
tumor_alt_rpir_mad	Mean absolute deviation of the positions along reverse strand reads for bases supporting the alternate allele in the tumor sample
observed_in_normals_count	The number of reads supporting the candidate mutation in the normal sample
failure_reasons	Reasons for rejecting the candidate somatic mutation. Possibilities include: (1) alt_allele_in_normal - The alternate allele has significant support in the normal sample. (2) clustered_read_position - The alternate allele is not distributed evenly over the length of the read. (3) fstar_tumor_lod - the candidate does not have significant support above noise. (4) germline_risk - there is evidence for the mutation in the normal sample at a dbSNP site (5) nearby_gap_events - Insertion and deletion events were identified at the locus. (6) normal_lod - there is evidence for the mutation in the normal sample. (7) poor_mapping_region_alternate_allele_mapq - Low mapping quality for the alternate allele. (8) poor_mapping_region_mapq0 - Too many reads with a mapping quality of 0 at the locus. (9) possible_contamination - Possible contamination of the normal sample with tumor. (10) strand_artifact - The mutation is likely a strand bias artifact. (11) triallelic_site - The site is not biallelic.
judgement	The candidate is a true somatic variant (KEEP) or the candidate is not a likely somatic variant (REJECT).

2.4 STD_COVERAGE.TXT

A WIGGLE format file describing whether there is sufficient coverage to detect somatic variants at a 0.3 allele fraction in the tumor with 80% power. 1 indicates that the coverage at the locus passes this threshold, 0 otherwise.

2.5 Q20_COVERAGE.TXT

A WIGGLE format file describing whether there is sufficient coverage to detect somatic variants at a 0.3 allele fraction in the tumor with 80% power examining only bases with a quality of greater than 20. 1 indicates that the coverage at the locus passes this threshold, 0 otherwise.

2.6 POWER.TXT

A WIGGLE format file describing the power to detect a somatic variant at the observed coverage in the tumor and normal samples.

2.7 TUMOR_DP.TXT

A WIGGLE format file describing the observed sequence read depth in the tumor sample.

2.8 NORMAL_DP.TXT

A WIGGLE format file describing the observed sequence read depth in the normal sample.

3 TNhaplotyper

3.1 Introduction

An example command with TNhaplotyper is as follows

```
sentieon driver -t NUMBER_THREADS -r REFERENCE.FASTA \  
  -i NORMAL_RECALED.BAM -i TUMOR_RECALED.BAM \  
  --interval INTERVAL \  
  --algo TNhaplotyper --dbsnp DBSNP.VCF \  
  --tumor_sample TUMOR_SM --normal_sample NORMAL_SM \  
  OUTPUT.VCF
```

This command line produces the following required output files:

- **OUTPUT.VCF**

3.2 OUTPUT.VCF

The **OUTPUT.VCF** file conforms to the VCF 4.2 specification. More information on the VCF format can be found at <https://samtools.github.io/hts-specs/VCFv4.2.pdf>. The INFO field annotations are described in detail below.

The core statistics of the software are TLOD, which is a measure of the support for the mutation relative to the expected level of sequencing noise at the candidate site, and NLOD, which is a measure of the odds that the mutation is absent from the normal sample.

INFO annotation	Description
DB	The variant is present in the VCF file supplied with the <code>-dbsnp</code> option
ECNT	Number of candidate variants in the active region, typically the number of candidate variants in the +/- 50 to 300 bp region
HCNT	Number of haplotypes observed in the active region after assembly of the sequence reads
MAX_ED	Maximum edit distance between the observed haplotypes in the active region
MIN_ED	Minimum edit distance between the observed haplotypes in the active region
NLOD	Log odds that the variant is not present in the normal sample (confidence that the variant is not a germline variant)
PON	Number of times the variant is observed in the panel of normal samples
RPA	The number of times the repeat is present for each allele for an indel within a short tandem repeat
RU	The sequence of the repeated nucleotides for an indel within a short tandem repeat
STR	The variant is an expansion or contraction of a short tandem repeat
TLOD	Log odds that the variant is present in the tumor sample relative to the expected noise

TNhaplotyper also populates the FILTER field for the variants.

FILTER	Description
PASS	The variant is confidently a somatic mutation
alt_allele_in_normal	The alternate allele is present in the paired normal sample and is unlikely to be a somatic variant
clustered_events	Multiple events are present on the same haplotype as the variant which is indicative of a false-positive call
germline_risk	There is evidence that the variant is present in the normal sample given that the variant is present in supplied dbSNP VCF and not present in the supplied COSMIC vcf
homologous_mapping_event	More than three events are present at this locus in the tumor which is indicate of a false-positive call
low_t_alt_frac	The variant is filtered due to a low alternate allele fraction in the tumor sample
multi_event_alt_allele_in_normal	Multiple events are present in the tumor sample and the alternate allele appears in the normal sample
panel_of_normals	The mutation is present in at least two samples in the panel of normals.
str_contraction	The mutation is a contraction of a short tandem repeat
t_lod_fstar	The mutation does not have significant support above noise
triallelic_site	The mutation occurs at a triallelic site

Standard genotype fields are defined by the format specification. However, TNhaplotyper also outputs the following non-standard fields.

GENO-TYPE	Description
AF	Fraction of reads supporting the alternate allele
ALT_F1R2	The number of reads in the F1R2 orientation supporting the alternate allele
ALT_F2R1	The number of reads in the F1R2 orientation supporting the alternate allele
FOXOG	The fraction of alt reads indicating OxoG error. OxoG error is induced by DNA oxidation during library preparation and is a frequent source of false-positive calls. See PMID: 23303777.
PGT	Physical phasing haplotype information describing how the alternate alleles are phased in relation to one another
PID	Physical phasing ID information, connecting records within a phasing group by using unique IDs within a given sample, but not across samples
QSS	Sum of base quality scores for each allele
REF_F1R2	The number of reads in the F1R2 orientation supporting the reference allele
REF_F2R1	The number of reads in the F2R1 orientation supporting the reference allele

4 TNhaplotyper2

4.1 Introduction

An example command with TNhaplotyper2 is as follows

```
sentieon driver -t NUMBER_THREADS -r REFERENCE.FASTA \
  -i NORMAL_RECALLED.BAM -i TUMOR_RECALLED.BAM \
  --algo TNhaplotyper2 --tumor_sample TUMOR_SM \
  --normal_sample NORMAL_SM \
  TMP.VCF \
  --algo OrientationBias --tumor_sample TUMOR_SM \
  ORIENTATION_DATA \
  --algo ContaminationModel --tumor_sample TUMOR_SM \
  --normal_sample NORMAL_SM \
  --vcf GERMLINE_RESOURCE \
  --tumor_segments CONTAMINATION_DATA.segments \
  CONTAMINATION_DATA

sentieon driver -r REFERENCE.FASTA \
  --algo TNfilter --tumor_sample TUMOR_SM \
  --normal_sample NORMAL_SM \
  -v TMP.VCF \
  --contamination CONTAMINATION_DATA \
  --tumor_segments CONTAMINATION_DATA.segments \
  --orientation_priors ORIENTATION_DATA \
  OUTPUT.VCF
```

This command line produces the following required output files:

- **OUTPUT.VCF**

4.2 OUTPUT.VCF

The **OUTPUT.VCF** file conforms to the VCF 4.2 specification. More information on the VCF format can be found at <https://samtools.github.io/hts-specs/VCFv4.2.pdf>. The INFO field annotations are described in detail below.

The core statistics of the software are TLOD, which is a measure of the support for the mutation relative to the expected level of sequencing noise at the candidate site, and NLOD, which is a measure of the odds that the mutation is absent from the normal sample.

INFO annotation	Description
AS_FilterStatus	The filter status of each allele, with alleles separated by the pipe character
AS_SB_TABLE	Forward and reverse read counts for each allele, with alleles separated by the pipe character
AS_UNIQ_ALT_READ_COUNT	Number of reads with unique start and mate-end positions for each alternate allele
CONTQ	Phred-scaled probability that the variant alleles are not due to contamination
DP	Approximate read depth
ECNT	Number of candidate variants in the active region, typically the number of candidate variants in the +/- 50 to 300 bp region
GERMQ	The phred-scaled posterior probability that the alternate allele(s) are not germline variants
MBQ	Median base quality of each allele
MFRL	Median fragment length of each allele
MMQ	Median mapping quality of each allele
MPOS	Median distance from the end of the read for each alternate allele
NALOD	Negative log 10 odds of the variant being an artifact in the normal sample with the same allele fraction as the tumor sample for each alternate allele
NCount	Count of N-bases in the read pileup
NLOD	Log odds that the variant is not present in the normal sample (confidence that the variant is not a germline variant) for each alternate allele
OCM	Number of reads supporting the alternate allele whose original alignment does not match the current contig
PON	The variant is found in the panel of normal samples
POPAF	Population allele frequency of the alternate alleles
ROQ	Phred-scaled probability that the variant alleles are not due to a read orientation artifact
RPA	The number of times the repeat is present for each allele for an indel within a short tandem repeat
RU	The sequence of the repeated nucleotides for an indel within a short tandem repeat
SEQQ	Phred-scaled probability that the variant alleles are not due to sequencing error
STR	The variant is an expansion or contraction of a short tandem repeat
STRANDQ	Phred-scaled probability of a strand-bias artifact
STRQ	Phred-scaled probability that the alternate alleles are errors due to polymerase slippage
TLOD	Log odds that the variant is present in the tumor sample relative to the expected noise

TNfilter also populates the FILTER field for the variants.

FILTER	Description
PASS	The site contains at least one allele that passes all filters
FAIL	All variant alleles are filtered, but for different reasons
base_qual	The median base quality of bases supporting the alternate allele is too low
clustered_events	Multiple events are present on the same haplotype as the variant which is indicative of a false-positive call
contamination	The alternate allele is present due to contamination
duplicate	The alternate allele is overrepresented by apparent sequencing duplicates
fragment	A large difference is observed in the median fragment length for reads supporting the reference and alternate alleles
germline	There is evidence that the variant is germline
haplotype	Variant is on the same haplotype as other filtered variants
low_allele_frac	The variant allele fraction is below the threshold
map_qual	The median mapping quality of reads supporting the alternate allele is too low
multiallelic	The mutation occurs at a multiallelic site
n_ratio	Too many 'N' bases at the site
normal_artifact	The variant is likely an artifact in the normal sample
orientation	The variant is likely an artifact due to orientation bias
panel_of_normals	The site is present in the panel of normals
position	The allele is close to the ends of the reads
slippage	The variant is likely an artifact due to polymerase slippage
strand_bias	Evidence for the alternate allele comes from only one read direction
strict_strand	Evidence for the alternate allele is not significant on both directions
weak_evidence	The mutation does not have significant support above noise

Standard genotype fields are defined by the format specification. However, TNhaplotyper2 also outputs the following non-standard fields.

GENO-TYPE	Description
AF	Fraction of reads supporting the alternate allele
AD	Allelic depths for the reference and alternate alleles
DP	Approximate read depth
F1R2	The number of reads in the F1R2 orientation supporting each allele
F2R1	The number of reads in the F2R1 orientation supporting each allele
PGT	Physical phasing haplotype information describing how the alternate alleles are phased in relation to one another
PID	Physical phasing ID information, connecting records within a phasing group by using unique IDs within a given sample, but not across samples
PS	Phasing set; typically the position of the first variant in the set
SB	The forward and reverse read counts for the reference and alternate alleles

5 TNscope®

5.1 Introduction

An example command with TNscope® is as follows

```
sentieon driver -t NUMBER_THREADS -r REFERENCE.FASTA \  
-i NORMAL_RECALLED.BAM -i TUMOR_RECALLED.BAM \  
--interval INTERVAL \  
--algo TNscope --tumor_sample TUMOR_SM \  
--normal_sample NORMAL_SM --dbsnp DBSNP.VCF OUTPUT.VCF
```

This command line produces the following required output files:

- **OUTPUT.VCF**

5.2 OUTPUT.VCF

The **OUTPUT.VCF** file conforms to the VCF 4.2 specification. More information on the VCF format can be found at <https://samtools.github.io/hts-specs/VCFv4.2.pdf>. The INFO field annotations are described in detail below.

The core statistics of the software are TLOD, which is a measure of the support for the mutation relative to the expected level of sequencing noise at the candidate site, and NLOD, which is a measure of the odds that the mutation is absent from the normal sample.

INFO annotation	Description
CIEND	The confidence interval around the END position for imprecise structural variants
CIPOS	Confidence interval around POS for imprecise structural variants
DB	The variant is present in the VCF file supplied with the <code>-dbsnp</code> option
DPR	Average depth in the region surrounding the variant (± 1 bp)
ECNT	Number of candidate variants in the active region, typically the number of candidate variants in the ± 50 to 300 bp region
END	The end position of the structural variant
FS	Phred-scale p-value using Fisher's exact test to detect strand bias
HCNT	The number of haplotypes observed in the active region after assembly of the sequence reads
IMPRECISE	The breakpoints of the structural variant are not precisely known
MATEID	Breakend mate
MAX_ED	Maximum edit distance between the observed haplotypes in the active region
MIN_ED	Minimum edit distance between the observed haplotypes in the active region
NLOD	Log odds that the variant is not present in the normal sample (confidence that the variant is not a germline variant)
NLODF	Log odds that the variant is not present in the normal sample (confidence that the variant is not a germline variant) given the allele fraction in the tumor sample
PON	Number of times the variant is observed in the panel of normal samples
PV	The p-value from a Fisher's exact test of the number of reads supporting the reference and alternate alleles in the tumor and normal samples
PV2	The p-value from a Fisher's exact test of the number of reads supporting the reference and alternate alleles in the tumor and normal samples using only high-confidence reads
RPA	The number of times the repeat is present for each allele for an indel within a short tandem repeat
RU	The sequence of the repeated nucleotides for an indel within a short tandem repeat
SOMATIC	The variant occurs uniquely in the sample supplied with the <code>-tumor_sample</code> option
SOR	Symmetric Odds Ratio of 2x2 contingency table to detect strand bias
STR	The variant is an expansion or contraction of a short tandem repeat
SVLEN	The difference in length between REF and ALT alleles of structural variants
SVTYPE	The type of structural variant
TLOD	Log odds that the variant is present in the tumor sample relative to expected noise
VAF	The variant allele frequency. The fraction of reads supporting the alternate allele in the tumor sample.

TNscope® also populates the FILTER field for the variants.

FILTER	Description
PASS	The variant is confidently a somatic mutation
alt_allele_in_normal	The alternate allele is present in the paired normal sample and is unlikely to be a somatic variant
clustered_events	Multiple events are present on the same haplotype as the variant which is indicative of a false-positive call
germline_risk	There is evidence that the variant is present in the normal sample given that the variant is present in supplied dbSNP VCF and not present in the supplied COSMIC vcf
homologous_mapping_event	More than three events are present at this locus in the tumor which is indicate of a false-positive call
low_t_alt_frac	The variant is filtered due to a low alternate allele fraction in the tumor sample
multi_event_alt_allele_in_normal	Multiple events are present in the tumor sample and the alternate allele appears in the normal sample
panel_of_normals	The mutation is observed in at least two samples in the panel of normals
str_contraction	The mutation is a contraction of a short tandem repeat
t_lod_fstar	The mutation does not have significant support above noise
trialelic_site	The mutation occurs at a triallelic site

Standard genotype fields are defined by the format specification. However, TNscope® also outputs the following non-standard fields.

GENOTYPE field	Description
AF	Fraction of reads supporting the alternate allele
AFDP	Read depth used to calculate AF
AFDPLOWMQ	Read depth used to calculate AF including reads with low mapping quality
AFLOWMQ	Allele fraction of the event in the tumor including low mapq reads
ALT_F1R2	The number of reads in the F1R2 orientation supporting the alternate allele
ALT_F2R1	The number of reads in the F2R1 orientation supporting the alternate allele
ALTHC	Depth of reads supporting the highest confidence alternate allele
ALTHCLOWMQ	Depth of reads supporting the highest confidence alternate allele including reads with low mapping quality
BaseQRankSumPS	Z-score from Wilcoxon rank sum test of Alt vs. Ref base qualities per sample
ClippingRankSumPS	Z-score from Wilcoxon rank sum test of Alt vs. Ref number of hard clipped bases per sample
DPHC	Depth of high-confidence reads supporting the reference or alternate allele
DPHCLOWMQ	Depth of high-confidence reads supporting the reference or alternate allele including reads with low mapping quality
FOXOG	The fraction of alt reads indicating OxoG error. OxoG error is induced by DNA oxidation during library preparation and is a frequent source of false-positive calls. See PMID: 23303777.
MQRankSumPS	Z-score from Wilcoxon rank sum test of Alt vs. Ref read mapping qualities per sample
NBQPS	Mean Neighboring Base Quality, including 5bp on both sides per sample
PGT	Physical phasing haplotype information, describing how the alternate alleles are phased in relation to one another
PID	Physical phasing ID information, connecting records within a phasing group by using unique IDs within a given sample, but not across samples
QSS	Sum of base quality scores for each allele
ReadPosEndDistPS	Z-score from Wilcoxon rank sum test of mean distance from either end of read per sample
ReadPosRankSumPS	Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias per sample
REF_F1R2	The number of reads in the F1R2 orientation supporting the reference allele
REF_F2R1	The number of reads in the F2R1 orientation supporting the reference allele

©Sentieon Inc.
160 E Tasman Dr STE 208, San Jose, CA 95134-1619
www.sentieon.com