Alignment and Preprocessing

Alignment to the reference genome with BWA

```
bwa mem \
    -K 10000000 \
    -M \
    -R '${rg}' \
    -t ${nt} \
    hs37d5.fa ${fq1} ${fq2} | \
sentieon util sort \
    -o ${aligned_bam} \
    -t ${nt} \
    --sam2bam \
    -i -
```

Duplicate marking

```
sentieon driver \
    -t ${nt} \
    -i ${aligned_bam1} \
    -i ${aligned_bam2} \
    … \
    --algo LocusCollector \
    score.txt

sentieon driver \
    -t {nt} \
    -i ${aligned_bam1} \
    -i ${aligned_bam2} \
    … \
    --algo Dedup \
    --rmdup \
    --score_info score.txt \
    ${deduped_bam}
```

Indel Realignemnt

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
    -i ${deduped_bam} \
    --algo Realigner \
    ${realign_sites} \
    ${realigned_bam}
```

Base Quality Score Recalibration

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
```

```
    -i ${realigned_bam} \
    --algo QualCal \
    ${bqsr_sites} \
    ${qcal_table}
```

Variant Calling

TNsnv

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
    -i ${tumor_realigned_bam} \
    -i ${normal_realigned_bam} \
    -q ${tumor_qcal_table} \
    -q ${normal_qcal_table} \
    --algo TNsnv \
    --tumor_sample ${tumor_samplename} \
    --normal_sample ${normal_samplename} \
    --call_stats_out ${out_call_stats} \
    --min_init_tumor_lod 1.0 \
    ${tnsnv_vcf}
```

TNhaplotyper

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
    -i ${tumor_realigned_bam} \
    -i ${normal_realigned_bam} \
    -q ${tumor_qcal_table} \
    -q ${normal_qcal_table} \
    --algo TNhaplotyper \
    --tumor_sample ${tumor_samplename} \
    --normal_sample ${normal_samplename} \
    --min_init_tumor_lod 1.0 \
    ${tnhaplotyper_vcf}
```

TNscope

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
    -i ${tumor_realigned_bam} \
    -i ${normal_realigned_bam} \
    -q ${tumor_qcal_table} \
    -q ${normal_qcal_table} \
    --algo TNscope \
    --tumor_sample ${tumor_samplename} \
```

```
    --normal_sample ${normal_samplename} \
    --min_init_tumor_lod 1.0 \
    ${tnscope_vcf}
```

TNscope-model

```
sentieon driver \
    -t ${nt} \
    -r hs37d5.fa \
    -i ${tumor_realigned_bam} \
    -i ${normal_realigned_bam} \
    -q ${tumor_qcal_table} \
    -q ${normal_qcal_table} \
    --algo TNscope \
    --tumor_sample ${tumor_samplename} \
    --normal_sample ${normal_samplename} \
    --no_mapq_cap 1 \
    --clip_by_minbq 1 \
    --assemble_mode 4 \
    --max_error_per_read 3 \
    --min_init_tumor_lod 1.0 \
    --min_normal_lod 1.0 \
    --min_tumor_allele_frac 0.006 \
    ${tnscope_raw_vcf}

# apply the machine learning model
sentieon driver \
    -r hs37d5.fa \
    --algo TNModelApply \
    -v ${tnscope_raw_vcf} \
    -m Sentieon_GiAB_201711.model \
    ${tnscope_model_vcf}

# remove default hard filters
bcftools annotate \
    -o ${tnscope_filter_vcf} \
    -O z \
    -x "^FILTER/MLrejected,FILTER/PASS" \
    ${tnscope_model_vcf}
```

Truthset Generation

Merge the two high-confidence regions

```
bedtools intersect \
    -a <(sort -k1,3V $tumor_bed) \
    -b <(sort -k1,3V $normal_bed) \
    | sort -k1,3V \
    | bedtools merge -i - \
```

```
    | sort -k1,3V \
    > $merged_bed
```

Generate the unique truthset for the tumor sample

```
bcftools isec
    -C \
    -o ${unique_truth} \
    -O z \
    -w 1 \
    -R $merged_bed \
    $tumor_truth \
    $normal_truth
sentieon util vcfindex $unique_truth
```

Call the unique variants to the tumor sample in the normal sample

```
sentieon driver \
    -i $normal_bam \
    -q $normal_recal_table \
    -t ${nt} \
    -r $ref \
    --interval $autosomes \
    --algo TNscope \
    --tumor_sample $normal_sample \
    --given $unique_truth \
    $given_vcf
```

Filter the given VCF

```
bcftools view \
    -O z \
    -i 'FORMAT/AD[1] >= 10 | (FORMAT/AD[1] / (FORMAT/AD[0] +
FORMAT/AD[1])) > 0.1' \
    $given_vcf \
    > $mask_vcf
```

Mask variants present in the normal sample

```
bedtools subtract \
    -a $merged_bed \
    -b $mask_vcf \
    > $mask_bed
```

Variant Evaluation

Generating SNV and indel callsets

```
# SNPs VCF
bcftools view \
    -o ${snps_vcf} \
```

```
    -O z \
    -v snps \
    ${raw_vcf}
sentieon util vcfindex ${snps_vcf}

# indels VCF
bcftools view \
    -o ${indels_vcf} \
    -O z \
    -V snps \
    ${raw_vcf}
sentieon util vcfindex ${indels_vcf}
```

TNsnv, TNhaplotyper, and TNscope

```
java -jar RTG.jar vcfeval \
    --baseline=${truth_vcf} \
    --calls=${calls_vcf} \
    --template=${ref} \
    --bed-regions=${truth_bed} \
    --output=${output} \
    --squash-ploidy \
    --sample=${sample_name} \
    --vcf-score-field=INFO.TLOD
```

Plotting - TNscope-model

```
java -jar RTG.jar vcfeval \
    --baseline=${truth_vcf} \
    --calls=${calls_vcf} \
    --template=${ref} \
    --bed-regions=${truth_bed} \
    --output=${output} \
    --squash-ploidy \
    --sample=${sample_name} \
    --vcf-score-field=INFO.ROC_VAL \
    --all-records
```

Scoring - TNscope-model

```
java -jar RTG.jar vcfeval \
    --baseline=${truth_vcf} \
    --calls=${calls_vcf} \
    --template=${ref} \
    --bed-regions=${truth_bed} \
    --output=${output} \
    --squash-ploidy \
    --sample=${sample_name} \
    --vcf-score-field=INFO.ROC_VAL
```