Structure variation detection

Haoxiang Lin

Abstract

Citation: Haoxiang Lin Structure variation detection. **protocols.io**
https://www.protocols.io/view/structure-variation-detection-gr4bv8w
Published: 22 Jun 2017

Protocol

Quick align by BWA

Step 1.

Quick align by BWA

SOFTWARE PACKAGE (Linux)
BWA, 0.6.1
DATASET
Contig fasta
COMMAND
```
bwa bwasw -t 4 $HG19 $CONTIG
```
Only use contig which length > 100bp

Exact align by LASTZ

Step 2.

Exact align by LASTZ

SOFTWARE PACKAGE (LINUX)
LASTZ, 1.02
DATASET
Contig fasta
COMMAND
```
lastz --targetcapsule=$CAPSULE $FASTA[nameparse=darkspace] --strand=both --chain --
ambiguous=iupac --gapped --ydrop=50000 --gap=1000,1 --format=axt --output=$AXT --markend
```
EXPECTED RESULTS

LASTZ aligment

Call SV by SOAPSV

Step 3.

Call SV by SOAPSV. SOAPSV is a huge pipeline, include many programes and scripts. Several commands lines of key steps are showed.

SOFTWARE PACKAGE (LINUX)

SOAPsv, 1.02

DATASET

LASTZ aligment

COMMAND

```
axtSort $AXT > $SORT_AXT
...
# find best hit in alignments, alignment linearization
best_hit $SORT_AXT > $BEST_AXT
...
intro_indel_1.3 $FINAL_AXT > $SV
```

Call SV by Pindel

Step 4.

Call SV by Pindel

SOFTWARE PACKAGE (LINUX)

PINDEL, 0.2.4t

DATASET

Merged BAM

COMMAND

```
pindel -f   $HG19 -i $CFG -o $OUT_PREFIX
pindel2vcf -P $PREFIX -r $HG19 -R hg19 -d hg19 -v $VCF
```

EXPECTED RESULTS

SV result

Call SV by CNVnator

Step 5.

Call SV by CNVnator

SOFTWARE PACKAGE (LINUX)

cnvnator, 0.2.7

DATASET

Merged BAM

COMMAND

```
./cnvnator -genome hg19 -root out.root  -tree $BAM
./cnvnator -genome hg19 -root out.root -his 100
./cnvnator -root out.root  -stat 100
./cnvnator -root out.root -partition 100
./cnvnator -root out.root  -call 100
```

EXPECTED RESULTS

SV result

Call SV by Breakdancer

Step 6.

Call SV by Breakdancer

SOFTWARE PACKAGE (LINUX)

Breakdancer-max, 1.2

DATASET

Merged BAM

cmd COMMAND

```
bam2cfg.pl -q 20 -c 3 -g -h $BAM > $CFG
breakdancer -o $PREFIX -q 20 -d $CTX -a -y 30 $CFG
```

EXPECTED RESULTS

SV result

Call SV by Genome STRIP

Step 7.

Call SV by Genome STRIP

SOFTWARE PACKAGE (LINUX)

Genome STRIP, v1.0

DATASET

Merged BAM

cmd COMMAND

```
java -cp ${classpath} ${mx} \
    org.broadinstitute.sting.queue.QCommandLine \
    -S ${SV_DIR}/qscript/SVPreprocess.q \
    -S ${SV_DIR}/qscript/SVQScript.q \
    -gatk ${SV_DIR}/lib/gatk/GenomeAnalysisTK.jar \
    -cp ${classpath} \
    -configFile conf/genstrip_parameters.txt \
    -disableGATKTraversal \
    -tempDir ${SV_TMPDIR} \
    -R $HG19 \
    -computeGCProfiles \
    -genomeMaskFile hg19.mask.101.fasta \
    -ploidyMapFile hg19.ploidy.map \
    -copyNumberMaskFile cn2_mask_hg19.fasta \
    -genderMapFile gender.list \
    -runDirectory ${runDir} \
    -computeGCProfiles \
    -md ${runDir}/metadata \
    -jobLogDir ${runDir}/logs \
    -I ${bam} \
    --disableJobReport \
    -run     || exit 1

java -cp ${classpath} ${mx} \
    org.broadinstitute.sting.queue.QCommandLine \
    -S ${SV_DIR}/qscript/SVDiscovery.q \
    -S ${SV_DIR}/qscript/SVQScript.q \
    -gatk ${SV_DIR}/lib/gatk/GenomeAnalysisTK.jar \
    --disableJobReport \
    -cp ${classpath} \
    -configFile ./genstrip_parameters.txt \
    -tempDir ${SV_TMPDIR} \
    -R $HG19 \
    -genomeMaskFile hg19.mask.101.fasta \
```

```
    -genderMapFile gender.list \
    -runDirectory ${runDir} \
    -md ${runDir}/metadata \
    -disableGATKTraversal \
    -jobLogDir ${runDir}/logs \
    -minimumSize 50 \
    -maximumSize 1000000 \
    -windowSize 20000000 \
    -windowPadding 10000 \
    -I ${bam} \
    -O ${sites} \
    -P select.validateReadPairs:false \
    -run     || exit 1
```

EXPECTED RESULTS

SV Result

Combine Deletion

Step 8.

Combine deletion in individual level between different methods and Combine SV in popluation leverl in different individuals with using in-house scrtips. The methods are similar in 1000 genome paper. Merge exact breakpoint by locations and merge imprecise breakpoint by confident region.

Genotyping by Genome STRIP

Step 9.

Genotyping deletions by Genome STRIP

SOFTWARE PACKAGE (LINUX)

Genome STRIP, v1.0

DATASET

Merged BAM and site VCF

COMMAND

```
java  -cp ${classpath} ${mx} \
    org.broadinstitute.sting.queue.QCommandLine \
    -S ${SV_DIR}/qscript/SVGenotyper.q \
    -S ${SV_DIR}/qscript/SVQScript.q \
    -gatk ${SV_DIR}/lib/gatk/GenomeAnalysisTK.jar \
    --disableJobReport \
    -cp ${classpath} \
    -configFile genstrip_parameters.txt \
    -tempDir ${SV_TMPDIR} \
    -R $HG19 \
    -genomeMaskFile hg19.mask.101.fasta \
    -genderMapFile gender.list \
    -runDirectory ${runDir} \
    -md ${runDir}/metadata \
    -jobLogDir ${runDir}/logs \
    -I ${bam} \
    -vcf ${sites} \
    -disableGATKTraversal \
    -O ${genotypes} \
    -run     || exit 1
```

EXPECTED RESULTS

SV genotypes in VCF format