|  |  |
| --- | --- |
| |  | | --- | | **Young Shin Song** | |

1. Fastq QC  
  
2. Alignment using BWA mem  
  
3. sam to bam conversion using samtools  
  
4. sorting using samtools sort  
  
5. Duplicated read marking using Picard  
  
아래는 꼭 안해도 되기는 함.  
6. indel realignment using GATK  
  
7. base recalibration using GATK  
  
samtools idxstat / samtools flagstat 등으로 align이 잘 되었는지, depth는 얼마인지 등에 대한 정보를 얻을 수 있음 (일종의 bamQC)  
  
이후에는 snp, indel, sv, cnv calling 을 진행하면 됨.  
언제나 각 tool은 다른 option들이 있는 것이고 내가 쓰는 것은 그 중의 하나임.

**< Script >**

예시: HCNor1\_R1.fastq.gz / HCNor1\_R2.fastq.gz 샘플로 bam 만들기

**## 00\_GRCh37\_bamprocessing.sh 파일**

#!/bin/bash  
set -e  
fastq1=$1  
fastq2=$2  
sampleName=$3  
reference=$4  
  
echo create sam using bwa mem  
bwa mem -M -t 4 -R "@RG\tID:$sampleName\tLB:1\tSM:$sampleName\tPL:ILLUMINA" $reference $fastq1 $fastq2 > $sampleName.sam 2> $sampleName.bwamem.out  
  
echo sam to bam  
/usr/local/bin/samtools view -Sb -@ 3 -o $sampleName.bam $sampleName.sam > $sampleName.StoB.out 2>&1  
rm $sampleName.sam  
  
echo sort  
/usr/local/bin/samtools sort -@ 3 -o $sampleName.s.bam $sampleName.bam > $sampleName.sort.out 2>&1  
rm $sampleName.bam  
  
echo MarkDuplicate  
java -Xms8g -Xmx12g -jar /home/users/tools/picard/dist/picard.jar MarkDuplicates REMOVE\_DUPLICATES=true REMOVE\_SEQUENCING\_DUPLICATES=ture I=$sampleName.s.bam O=$sampleName.s.md.bam M=$sampleName.matrics.txt VALIDATION\_STRINGENCY=LENIENT CREATE\_INDEX=true > $sampleName.markdup.out 2>&1  
rm $sampleName.s.bam  
  
echo realignment  
java -Xms8g -Xmx12g -jar /home/users/tools/gatk/gatk-3.5/GenomeAnalysisTK.jar -T RealignerTargetCreator -R /home/users/yssong/99\_reference/human/GRCh37/human\_g1k\_v37.fasta -I $sampleName.s.md.bam --known /home/users/yssong/99\_reference/human/GRCh37/Mills\_and\_1000G\_gold\_standard.indels.b37.vcf -o $sampleName.s.md.bam.intervals -nt 4 > $sampleName.RTC.out 2>&1  
java -Xms8g -Xmx12g -jar /home/users/tools/gatk/gatk-3.5/GenomeAnalysisTK.jar -T IndelRealigner -R /home/users/yssong/99\_reference/human/GRCh37/human\_g1k\_v37.fasta -I $sampleName.s.md.bam -known /home/users/yssong/99\_reference/human/GRCh37/Mills\_and\_1000G\_gold\_standard.indels.b37.vcf -targetIntervals $sampleName.s.md.bam.intervals -o $sampleName.s.md.ir.bam > $sampleName.IR.out 2>&1  
rm $sampleName.s.md.bam  
  
echo recalibration  
java -Xms8g -Xmx12g -jar /home/users/tools/gatk/gatk-3.5/GenomeAnalysisTK.jar -T BaseRecalibrator -R /home/users/yssong/99\_reference/human/GRCh37/human\_g1k\_v37.fasta -I $sampleName.s.md.ir.bam --knownSites /home/users/yssong/99\_reference/human/GRCh37/Mills\_and\_1000G\_gold\_standard.indels.b37.vcf -knownSites /home/users/yssong/99\_reference/human/GRCh37/dbsnp\_138.b37.vcf -o $sampleName.s.md.ir.bam.table -nct 4 > $sampleName.BR.out 2>&1  
java -Xms8g -Xmx12g -jar /home/users/tools/gatk/gatk-3.5/GenomeAnalysisTK.jar -T PrintReads -R /home/users/yssong/99\_reference/human/GRCh37/human\_g1k\_v37.fasta -I $sampleName.s.md.ir.bam -BQSR $sampleName.s.md.ir.bam.table -o $sampleName.s.md.ir.br.bam -nct 4 > $sampleName.PR.out 2>&1  
  
rm $sampleName.s.md.bai  
rm $sampleName.s.md.ir.bai  
rm $sampleName.s.md.ir.bam  
rm $sampleName.s.md.ir.bam.table  
rm $sampleName.s.md.bam.intervals

**## 00\_1\_script\_bamprocessing.sh 파일**  
sh 00\_GRCh37\_bamprocessing.sh ../03\_fastq\_human\_sample/$1\_R1.fastq.gz ../03\_fastq\_human\_sample/$1\_R2.fastq.gz $1 /home/users/jhyouk/99\_reference/human/GRCh37/human\_g1k\_v37.fasta &> $1\_01.out

**sh 00\_1\_script\_bamprocessing.sh HCNor1**-->  실행하는 문장