Nov 23, 2018

ADBS Whole Genome Sequencing (WGS) analysis pipeline for Genomic-QC Report

DOI
dx.doi.org/10.17504/protocols.io.vuae6se

Ravi Dr More[1]

[1]University of Cambridge

Ravi Dr More
University of Cambridge

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.vuae6se

Protocol Citation: Ravi Dr More 2018. ADBS Whole Genome Sequencing (WGS) analysis pipeline for Genomic-QC Report. **protocols.io** https://dx.doi.org/10.17504/protocols.io.vuae6se

Manuscript citation:

Protocol status: Working
We use this protocol and it's working

Created: November 23, 2018

Last Modified: November 23, 2018

Protocol Integer ID: 18018

Abstract

Whole Genome Sequencing (WGS) analysis pipeline devloped for generating Genomic-QC Report in Accelerator Program for Discovery in Brain Disorders Using Stem Cells (ADBS) program.

Define paths and directories

1

Command

```
SAMPLE_PATH="/path/to/sample"
SAMPLE_NAME="test_sample"
SOFTWARE_PATH="/path/to/software"
DATABASES_PATH="/path/to/databases"
TEMP_DIR="/path/to/temp"
```

Unzip the raw reads files from .gz to fastq format

2

Command

```
gunzip $SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME*.fastq.gz
```

QC check of R1 and R2 paired-end raw reads using FASTQC, Trimming poor quality reads using Prinseq-lite, and Adapter contimination removal using AfterQC

3 Software versions used:

FASTQC version 0.10.1
Prinseq-lite version 0.20.4
AfterQC version 0.9.6

Command

```
$SOFTWARE_PATH/FastQC/fastqc
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R1.fq

$SOFTWARE_PATH/FastQC/fastqc
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R2.fq

cd $SAMPLE_PATH/$SAMPLE_NAME/

python $SOFTWARE_PATH/AfterQC-master/after.py -f -1 -t -1 -q 30 -1
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R1.fq -2
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R2.fq

$SOFTWARE_PATH/prinseq-lite-0.20.4/prinseq-lite.pl -fastq
$SAMPLE_PATH/$SAMPLE_NAME/good/$SAMPLE_NAME\_R1.good.fq -fastq2
$SAMPLE_PATH/$SAMPLE_NAME/good/$SAMPLE_NAME\_R2.good.fq -out_good
$SAMPLE_PATH/$SAMPLE_NAME/cleaned -out_bad null -min_qual_mean 30

mv $SAMPLE_PATH/$SAMPLE_NAME/cleaned_1.fastq
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq

mv $SAMPLE_PATH/$SAMPLE_NAME/cleaned_2.fastq
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2.fastq

$SOFTWARE_PATH/FastQC/fastqc
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq

$SOFTWARE_PATH/FastQC/fastqc
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2.fastq

mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_4_FASTQC

mv $SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1_fastqc.zip
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_4_FASTQC/$SAMPLE_NAME\_
cleaned_R1_fastqc.zip

mv $SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2_fastqc.zip
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_4_FASTQC/$SAMPLE_NAME\_
cleaned_R2_fastqc.zip
```

Alignment of clened raw reads against Human Reference Genome hg19 GRCh37.p13 build using BWA and SAMTOOLS.

4 BWA version 0.5.9
Samtools version 1.3.1

Command

```
# Align cleaned R1 reads with hg19
/softwares/bwa-0.5.9/bwa aln -t 30 $DATABASES_PATH/hg19_fa-
chrMlast/hg19_chrM-last.fa
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R1.sai

# Align cleaned R2 reads with hg19
/softwares/bwa-0.5.9/bwa aln -t 30 $DATABASES_PATH/hg19_fa-
chrMlast/hg19_chrM-last.fa
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R2.sai

#convert sai to sam by using cleaned fastq reads
/softwares/bwa-0.5.9/bwa sampe $DATABASES_PATH/hg19_fa-
chrMlast/hg19_chrM-last.fa
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R1.sai
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_R2.sai
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME.sam

#convert sam to bam
/softwares/samtools1.3.1/bin/samtools view -bS
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME.sam >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME.bam

#bam to sort file
/softwares/samtools1.3.1/bin/samtools sort
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME.bam -o
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sorted.bam

#sort to flagstat
/softwares/samtools1.3.1/bin/samtools flagstat
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sorted.bam >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sorted_flagstat.txt

#index the sorted bam
/softwares/samtools1.3.1/bin/samtools index
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sorted.bam >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sortedbam.bai
```

Mark PCR duplicates and sorting BAM using PICARD Tools

5 Picard version 2.0.1
Samtools version 1.3.1

Command

```
#Remove PCR duplicates
java -Djava.io.tmpdir=$TEMP_DIR/ -Xmx50g -jar
$SOFTWARE_PATH/picard/build/libs/picard.jar AddOrReplaceReadGroups
I="$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_sorted.bam"
O="$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_coordsort.bam" ID="1"
LB="libraryname" PL="Illumina" PU="platform unit" SM=samplenname
SO=coordinate VALIDATION_STRINGENCY=SILENT

java -Djava.io.tmpdir=$TEMP_DIR/ -Xmx50g -jar
$SOFTWARE_PATH/picard/build/libs/picard.jar MarkDuplicates
I="$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_coordsort.bam"
O="$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam" M="metrics"
REMOVE_DUPLICATES=true ASSUME_SORTED=true
VALIDATION_STRINGENCY=LENIENT

#Index the coordinate sorted bam file
/softwares/samtools1.3.1/bin/samtools index
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam
```

INDEL re-alignment using GATK tools

6 GATK version 3.6

Command

```
java -Xmx8g -jar $SOFTWARE_PATH/GenomeAnalysisTK-
3.6/GenomeAnalysisTK.jar -T RealignerTargetCreator -R
$DATABASES_PATH/hg19_fa-chrMlast/hg19_chrM-last.fa -I
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam --known
$DATABASES_PATH/REF_GENOME_hg19/1000G_phase1.indels.hg19.vcf -o
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_IndelRealigner.intervals
```

SNP and INDEL variant calling using Isaac Variant Caller tool and filter SNP and INDEL using rtg-tools

7 Isaac Variant Caller -- 1.0.7
rtg-tools version 3.7.1

Command

```
$SOFTWARE_PATH/isaac_variant_caller-master/bin/configureWorkflow.pl --
bam=$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam --
ref=$DATABASES_PATH/hg19_fa-chrMlast/hg19_chrM-last.fa --
config=$SAMPLE_PATH/config.ini --output-
dir=$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/

cd $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/
make -j 16

gzip -dc
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_RMDUP.genome.vcf.gz | $SOFTWARE_PATH/gvcftools-
0.16/bin/extract_variants | awk '/^#/ || $7 == "PASS"' >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_RMDUP_all_passed_variants.vcf

$SOFTWARE_PATH/rtg-tools-3.7.1/rtg vcffilter --snps-only -i
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_RMDUP_all_passed_variants.vcf -o
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_snp_issac.vcf

$SOFTWARE_PATH/rtg-tools-3.7.1/rtg vcffilter --non-snps-only -i
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_RMDUP_all_passed_variants.vcf -o
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_indel_issac.vcf

cp
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_snp_issac.vcf.gz
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_snp.vcf.gz

cp
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/resu
lts/$SAMPLE_NAME\_indel_issac.vcf.gz
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_indel.vcf.gz

gunzip
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
```

```
PLE_NAME\_snp.vcf.gz

gunzip
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_indel.vcf.gz
```

Check the alignment QC of the bam file using Qualimap

8 Qualimap version 2.2.1

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_5_ALIGNMENT_QC

$SOFTWARE_PATH/qualimap_v2.2.1/qualimap bamqc -bam
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam -gff
$DATABASES_PATH/trueseq1.bed -outdir
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_5_ALIGNMENT_QC/QualiMap
_$SAMPLE_NAME\_trueseq1_bed -outfile $SAMPLE_NAME\_trueseq1.pdf --
java-mem-size=500G
```

VCF QC of SNP and INDEL files using rtg-tools

9 rtg-tools version 3.7.1

Command

```
mkdir -p
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/VCF_
QC

$SOFTWARE_PATH/rtg-tools-3.7.1/rtg vcfstats
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_snp.vcf >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/VCF_
QC/$SAMPLE_NAME\_snp.vcf.stat

$SOFTWARE_PATH/rtg-tools-3.7.1/rtg vcfstats
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_indel.vcf >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/VCF_
QC/$SAMPLE_NAME\_indel.vcf.stat
```

SNP AND INDEL variant annotation using ANNOVAR

10 ANNOVAR reference assembly 65 with reference hg19

Command

```
mkdir -p
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar

perl $SOFTWARE_PATH/annovar/convert2annovar.pl -format vcf4
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_snp.vcf >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar/$SAMPLE_NAME\_snp.vcf.avinput

perl $SOFTWARE_PATH/annovar/convert2annovar.pl -format vcf4
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_indel.vcf >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar/$SAMPLE_NAME\_indel.vcf.avinput

#perl $SOFTWARE_PATH/annovar/table_annovar.pl
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_snp.vcf.avinput $SOFTWARE_PATH/annovar/humandb/ -buildver
hg19 -out
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar/$SAMPLE_NAME\_snp_annovar_annotation -remove -protocol
refGene,cytoBand,genomicSuperDups,esp6500siv2_all,1000g2015aug_all,100
0g2015aug_eur,exac03,avsnp147,dbnsfp30a -operation g,r,r,f,f,f,f,f,f -
nastring . -csvout

perl $SOFTWARE_PATH/annovar/table_annovar.pl
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_snp.vcf $SOFTWARE_PATH/annovar/humandb/ -buildver hg19 -out
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar/$SAMPLE_NAME\_snp_annovar_annotation -remove -protocol
refGene,cytoBand,genomicSuperDups,esp6500siv2_all,1000g2015aug_all,100
0g2015aug_eur,exac03,avsnp147,dbnsfp30a -operation g,r,r,f,f,f,f,f,f -
nastring . -vcfinput

perl $SOFTWARE_PATH/annovar/table_annovar.pl
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/$SAM
PLE_NAME\_indel.vcf $SOFTWARE_PATH/annovar/humandb/ -buildver hg19 -
out
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/anno
tated_annovar/$SAMPLE_NAME\_indel_annovar_annotation -remove -
protocol
```

```
refGene,cytoBand,genomicSuperDups,esp6500siv2_all,1000g2015aug_all,100
0g2015aug_eur,exac03,avsnp147,dbnsfp30a -operation g,r,r,f,f,f,f,f,f -
nastring . -vcfinput
```

Mitochondria analysis

11 Extracting mitochondrial reads from BAM file and creatiing anather BAM file to input mtDNA-Server tool for Mitochondria analysis
Samtools version 1.3

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_7_MITOCHONDRIA

/softwares/samtools1.3.1/bin/samtools view -b
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam chrM: -o
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_7_MITOCHONDRIA/$SAMPLE_
NAME\_MT.bam

/softwares/samtools1.3.1/bin/samtools index
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_7_MITOCHONDRIA/$SAMPLE_
NAME\_MT.bam
```

Blood Group Prediction

12 BOOGIE - Phenotype prediction from NGS data Version: 1.0

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_10_blood_group

perl $SAMPLE_PATH/rename_phase2_blood_group_detection.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_phase2_blood_group_summary.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_phase2_blood_group_genes_extracter.pl
$SAMPLE_NAME

chmod 755 $SAMPLE_PATH/$SAMPLE_NAME/*

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_blood_group_genes_extracter.pl

$SAMPLE_PATH/$SAMPLE_NAME/phase2_blood_group_detection.sh

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_blood_group_summary.pl
```

SNP-Chip rsID comparison with WGS rsID

13

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_14_VIRTUAL_SNP

perl $SAMPLE_PATH/rename_phase2_1rsid_get.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_phase2_2rsid_filter.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_phase2_3rsid_venn.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_phase2_4rsid_venn.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_common_indel.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_common_snp.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_rsid_indel.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_rsid_snp.pl $SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_unique_Illimina_snp.pl
$SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_unique_indel_Illimina.pl
$SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_unique_indel_sample.pl
$SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_extract_unique_sample_snp.pl
$SAMPLE_NAME

perl $SAMPLE_PATH/rename_exonic_venn_snp_indel.pl $SAMPLE_NAME

chmod 755 $SAMPLE_PATH/$SAMPLE_NAME/*

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_1rsid_get.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_2rsid_filter.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_3rsid_venn.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_4rsid_venn.pl
```

```
mkdir -p
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_14_VIRTUAL_SNP/exonic_r
sid

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_rsid_indel.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_rsid_snp.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_unique_indel_sample.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_unique_indel_Illimina.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_common_indel.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_unique_sample_snp.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_unique_Illimina_snp.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/exonic_extract_common_snp.pl
```

Extract Damaging Varaints (SIFT, PolyPhen) from SNP file

14

Command

```
mkdir -p
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/dama
ging

mkdir -p
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_13_VARIANT_CALLING/dama
ging/snp

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_damaging_1_get_snv_snp.pl

perl $SAMPLE_PATH/$SAMPLE_NAME/phase2_damaging_2_merge.pl
```

HLA Analysis using HLAVBSeq

15 Read data aligned to GRCh37/hg19 using **HLA-VBSeq Software to predict HLA types**
BWA version 0.5.9

Command

```
erl $SAMPLE_PATH/rename_hla_calculation.pl $SAMPLE_NAME

$SOFTWARE_PATH/bwa.kit/bwa mem -t 8 -P -L 10000 -a
$SOFTWARE_PATH/HLA/hla_all.fasta
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R2.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_part.sam

mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/

java -jar $SOFTWARE_PATH/HLA/HLAVBSeq.jar
$SOFTWARE_PATH/HLA/hla_all.fasta
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_part.sam
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result --alpha_zero 0.01 --is_paired

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^A\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_A.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^B\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_B.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^C\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_C.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DMA\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DMA.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE PATH/HLA/Allelelist.txt
```

```
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DMB\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DMB.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DOA\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DOA.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DOB\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DOB.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DPA1\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DPA1.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DPB1\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DPB1.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DQA1\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DQA1.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DQB1\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DQB1.txt

perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/$SAMPLE_NAME\_pa
rt_result | grep "^DRA\*" | sort -k2 -n -r >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DRA.txt
```

```
perl $SOFTWARE_PATH/HLA/parse_result.pl
$SOFTWARE_PATH/HLA/Allelelist.txt
```

Structual Variants (SV) Analysis using GASV

16 Geometric Analysis of Structural Variants (GASV) Version: 2.0

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_15_SV

perl $SAMPLE_PATH/rename_SV_gasv.pl $SAMPLE_NAME

perl $SAMPLE_PATH/$SAMPLE_NAME/SV_gasv.sh

cp /home/odity/ravim/$SAMPLE_NAME\_RMDUP.bam.gasv.in.clusters
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_15_SV/$SAMPLE_NAME\_RMD
UP.bam.gasv.in.clusters

mv $SAMPLE_PATH/$SAMPLE_NAME/*_null*
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_15_SV/

mv $SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam.gasv.in
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_15_SV/

perl $SAMPLE_PATH/rename_SV_count_type.pl $SAMPLE_NAME

perl $SAMPLE_PATH/$SAMPLE_NAME/SV_count_type.pl
```

```
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_11_HLA/HLA_DRB6.txt
```

Gene Integration detection using string search

17 Samtools version 1.3

Command

```
#cmyc gene end (GE) 65
#TGTTGCGGAAACGACGAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACTCTTGTGCGTAA

#vector start (VS) 15
#GAATTCGCTAGCGAT

#cmyc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
TGTTGCGGAAACGACGAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACTCTTGTGCGTAAGAATT
CGCTAGCGAT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_cmyc_GE65-VS15.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
ATCGCTAGCGAATTCTTACGCACAAGAGTTCCGTAGCTGTTCAAGTTTGTGTTTCAACTGTTCTCGTCGT
TTCCGCAACA >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_cmyc_GE65-VS15_rc.sam

########################
#bmi
#gene end
CTTCTTTTGCCAATAGACCTCGAAAATCATCAGTAAATGGGTCATCAGCAACTTCTTCTGGTTGA
#vec start GAATTCGCTAGCGAT

/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
CTTCTTTTGCCAATAGACCTCGAAAATCATCAGTAAATGGGTCATCAGCAACTTCTTCTGGTTGAGAATT
CGCTAGCGAT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_bmi_GE65-VS15.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
ATCGCTAGCGAATTCTCAACCAGAAGAAGTTGCTGATGACCCATTTACTGATGATTTTCGAGGTCTATTG
GCAAAAGAAG >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_bmi_GE65-VS15_rc.sam
```

```
####################

#bclxl
#gene end
#GGTTCCTGACGGGCATGACTGTGGCCGGCGTGGTTCTGCTGGGCTCACTCTTCAGTCGGAAATGA

# vec start
#GAATTCGCTAGCGAT

/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
GGTTCCTGACGGGCATGACTGTGGCCGGCGTGGTTCTGCTGGGCTCACTCTTCAGTCGGAAATGAGAATT
CGCTAGCGAT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_bclxl_GE65-VS15.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
ATCGCTAGCGAATTCTCATTTCCGACTGAAGAGTGAGCCCAGCAGAACCACGCCGGCCACAGTCATGCCC
GTCAGGAACC >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_bclxl_GE65-VS15_rc.sam
####################

#KLF4
#gene end
GTTTGTATTTTGCATACTCAAGGTGAGAATTAAGTTTTAAATAAACCTATAATATTTTATCTGAA
#vec start GAATTCGCTAGCGAT

/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
GTTTGTATTTTGCATACTCAAGGTGAGAATTAAGTTTTAAATAAACCTATAATATTTTATCTGAAGAATT
CGCTAGCGAT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_klf4_GE65-VS15.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
ATCGCTAGCGAATTCTTCAGATAAAATATTATAGGTTTATTTAAAACTTAATTCTCACCTTGAGTATGCA
AAATACAAAC >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_klf4_GE65-VS15.sam
####################
```

```
#Lin28
```

Mycoplasma Contimination detection using BWA

18 BWA version 0.5.9
Samtools version 1.3.1

Command

```
mkdir -p $SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma

#####Alaidlawii
/softwares/bwa-0.5.9/bwa aln -t 30
$SOFTWARE_PATH/Mycoplasma/Alaidlawii.fa
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.sai

/softwares/bwa-0.5.9/bwa samse
$SOFTWARE_PATH/Mycoplasma/Alaidlawii.fa
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.sai
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_cleaned_R1.fastq >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.sam

/softwares/samtools1.3.1/bin/samtools view -bS
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.sam >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.bam

/softwares/samtools1.3.1/bin/samtools sort
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii.bam -o
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sorted.bam

/softwares/samtools1.3.1/bin/samtools flagstat
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sorted.bam >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sorted_flagstat.txt

/softwares/samtools1.3.1/bin/samtools index
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sorted.bam >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sortedbam.bai

/softwares/samtools1.3.1/bin/samtools idxstats
```

```
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/$SAMPLE_N
AME\_R1_Alaidlawii_sorted.bam

for BAM in
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_18_Mycoplasma/*bam ; do

  CNT=`/softwares/samtools1.3.1/bin/samtools view -c -q20 $BAM`

  echo $BAM $CNT

done
```

```
#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
CTGGGGCCTCAGTCCTGTTCTCTTCCACATCACTAAACTGACTCCAGCTGTATCCTTTCTGGGAAAGCTT
GTAGGAGAGA >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_bclxl_VE15-GS65_rc.sam

#KLF4
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
TTGCGTACGGCCAGCAGTTTCCCGACCAGAGAGAACGAACGTGTCTGCGGGCGCGCGGGGAGCAGAGGCG
GTGGCGGGCG >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_klf4_VE15-GS65.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
GGGGCCAGAGGGGCGGGGGAGGGTCACTCGGCGGCTCCCGGTGCCGCCGCCGCCCGCCACCGCCTCTGCT
CCCCGCGCGC >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_klf4_VE15-GS65.sam

#Lin28
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
TTGCGTACGGCCAGCGTGCGGGGGAAGATGTAGCAGCTTCTTCTCCGAACCAACCCTTTGCCTTCGGACT
TCTCCGGGGC >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
```

```
MPLE_NAME\_lin28_VE15-GS65.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
GCCCCGGAGAAGTCCGAAGGCAAAGGGTTGGTTCGGAGAAGAAGCTGCTACATCTTCCCCCGCACGCTGG
CCGTACGCAA >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_lin28_VE15-GS65_rc.sam

#oct
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
TTGCGTACGGCCAGCTTGCTTTTGCAGATGTACCTTCTTAAAGTTTTTTCTTAAAGTTTGGGAAATATTG
AAATACGCTT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_oct_VE15-GS65.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
AAGCGTATTTCAATATTTCCCAAACTTTAAGAAAAAACTTTAAGAAGGTACATCTGCAAAAGCAAGCTGG
CCGTACGCAA >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_oct_VE15-GS65.sam

#sox2
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
TTGCGTACGGCCAGCGGATGGTTGTCTATTAACTTGTTCAAAAAAGTATCAGGAGTTGTCAAGGCAGAGA
AGAGAGTGTT >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_sox2_VE15-GS65.sam

#rc
/softwares/samtools1.3/bin/samtools view
$SAMPLE_PATH/$SAMPLE_NAME/$SAMPLE_NAME\_RMDUP.bam | grep
AACACTCTCTTCTCTGCCTTGACAACTCCTGATACTTTTTTGAACAAGTTAATAGACAACCATCCGCTGG
CCGTACGCAA >
$SAMPLE_PATH/$SAMPLE_NAME/Report_$SAMPLE_NAME\_16_GENE_INTEGRATION/$SA
MPLE_NAME\_sox2_VE15-GS65_rc.sam
```