\*\*This protocol is optimized for influenza A virus genome data

|  |  |
| --- | --- |
| pre-STEP 1 | Index reference genome |
| Tool | bwa/intel/0.7.15 |
| INPUT | reference.fasta (reference genome file) |
| OUTPUT | reference.fasta(or fa).ann; reference.fasta.pac; reference.fasta.amb; reference.fasta.bwt; reference.fasta.sa |
| Command | ① load module  module load bwa/intel/0.7.15  ② Indexing  bwa index reference.fasta |
| note |  |

|  |  |
| --- | --- |
| STEP1 | Alignment |
| Tool | bwa/intel/0.7.15 |
| INPUT | $i.trimmed.r1.fastq; $i.trimmed.r2.fastq; reference.fasta (reference genome file) |
| OUTPUT | $i.aligned\_reads.sam |
| Command | bwa mem -M reference.fasta $i.trimmed.r1.fastq $i.trimmed.r2.fastq > $i.aligned\_reads.sam |
| Loop  command | ① creating a shell script file  vim SAM-STEP1.sh  ② writing commands (press ‘I’ and input below)  \*\* This part is only valid on NYU HPC server (Prince)  \*\* Replace this part with the job-submission commands for your own HPC server  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP1  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP1.err  /path/to/the/files  module purge  module load bwa/intel/0.7.15  cd /path/to/the/files  for i in {start#..end#}; do  bwa mem -M reference.fasta $i.trimmed.r1.fastq $i.trimmed.r2.fastq > $i.aligned\_reads.sam  done |
| note | submitting the job is recommended  commands  sbatch <filename.sh> |

|  |  |
| --- | --- |
| STEP2 | SAM to BAM |
| Tool | samtools/intel/1.3.1 |
| INPUT | $i.aligned\_reads.sam |
| OUTPUT | $i.SAM.bam |
| Command | samtools view -S -b $i.aligned\_reads.sam > $i.SAM.bam |
| Loop  Command | ① creating a shell script file  vim SAM-STEP2.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP2  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP2.err  /bin/scratch/mk6134/mirella107  module purge  module load samtools/intel/1.3.1  cd /path/to/the/files  for i in {start#..end#}; do  samtools view -S -b $i.aligned\_reads.sam > $i.SAM.bam  done |
| note | submitting the job is recommended  commands  sbatch <filename.sh> |

|  |  |
| --- | --- |
| STEP3 | Sorting BAM ( for SNV calling) |
| Tool | samtools/intel/1.3.1 |
| INPUT | $i.SAM.bam |
| OUTPUT | $i.SAM.sorted.bam |
| Command | samtools sort -T/tmp/$iSAM.sorted -o $i.SAM.sorted.bam $i.SAM.bam |
| Loop  Command | ① creating a shell script file  vim SAM-STEP3.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP3  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP3.err  /bin/scratch/mk6134/mirella107  module purge  module load samtools/intel/1.3.1  cd /path/to/the/files  for i in {start#..end#}; do  samtools sort -T/tmp/$iSAM.sorted -o $i.SAM.sorted.bam $i.SAM.bam  done |
| note | submitting the job is recommended  commands  sbatch <filename.sh> |

|  |  |
| --- | --- |
| STEP4 | Index BAM |
| Tool | samtools/intel/1.3.1 |
| INPUT | $i.SAM.sorted.bam |
| OUTPUT | $i.SAM.sorted.bam.bai |
| Command | samtools index $i.SAM.sorted.bam |
| Loop  Command | ① creating a shell script file  vim SAM-STEP4.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP4  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP4.err  /bin/scratch/mk6134/mirella107  module purge  module load samtools/intel/1.3.1  cd /path/to/the/files  for i in {start#..end#}; do  samtools index $i.SAM.sorted.bam  done |
| note | submitting the job is recommended  commands  sbatch <filename.sh> |

|  |  |
| --- | --- |
| STEP 5 | Index the reference genome |
| Tool | samtools/intel/1.3.1 |
| INPUT | reference.fasta (reference genome) |
| OUTPUT | reference.fasta.fai |
| Command | samtools faidx reference.fasta |
| Loop  Command |  |
| Note |  |

|  |  |
| --- | --- |
| STEP 6 | Run ‘mpileup’ as a prestep for generating VCF format (rawVCF) |
| Tool | samtools/intel/1.3.1 |
| INPUT | $i.SAM.sorted.bam; reference.fasta (reference genome) |
| OUTPUT | $i.bcf |
| Command | samtools mpileup -g -f reference.fasta $i.SAM.sorted.bam > $i.bcf |
| Loop  Command | ① creating a shell script file  vim SAM-STEP6.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP6  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP6.err  /bin/scratch/mk6134/mirella107  module purge  module load samtools/intel/1.3.1  cd /path/to/the/files  for i in {start#..end#}; do  samtools mpileup -g -f reference.fasta $i.SAM.sorted.bam > $i.bcf  done |
| Note | submitting the job is recommended  commands  sbatch <filename.sh> |

|  |  |
| --- | --- |
| STEP 7 | Variant Calling ( from bcf format to vcf) |
| Tool | bcftools/intel/1.3.1 |
| INPUT | $i.bcf |
| OUTPUT | $i.SAM-BCF.raw.vcf |
| Command | bcftools call -c -v $i.bcf > $i.SAM-BCF.raw.vcf |
| Loop  Command | ① creating a shell script file  vim SAM-STEP7.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP7A  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP7.err  /bin/scratch/mk6134/mirella107  module purge  module load samtools/intel/1.3.1  cd /path/to/the/files  for i in {start#..end#}; do  bcftools call -c -v $i.bcf > $i.SAM-BCF.raw.vcf  done |
| Note | submitting the job is recommended  commands  sbatch <filename.sh> |

<filtering>

|  |  |
| --- | --- |
| STEP 8 | VCF filtering |
| Tool | bcftools/intel/1.3.1 |
| INPUT | $i.SAM-BCF.raw.vcf |
| OUTPUT | $i.SAM-BCF.flt.vcf |
| Command | bcftools filter -s LowQual -e ‘%QUAL<20 || DP>200’ $i.SAM-BCF.raw.vcf > $i.SAM-BCF.flt.vcf |
| Loop  Command | ① creating a shell script file  vim SAM-STEP8.sh  ② writing commands (press ‘I’ and input below)  #!/bin/bash  #  ##SBATCH --nodes=1  #SBATCH --nodes=1  #SBATCH --ntasks-per-node=1  #SBATCH --cpus-per-task=2  #SBATCH --time=24:00:00  #SBATCH --mem=20GB  #SBATCH --job-name=SAM-STEP8  #SBATCH --mail-type=END  #SBATCH --mail-user=mk6134@nyu.edu  #SBATCH --output=slurm\_%j.out  #SBATCH --error=SNVer-SAM-STEP8.err  cd /path/to/the/files  module load bcftools/intel/1.3.1  for i in {1..107}; do  bcftools filter -s HighQual -e ‘%QUAL<20 || DP>200’ $i.SAM-BCF.raw.vcf > $i.SAM-BCF.flt.vcf  done |
| Note | ① variants with lower quality score(<20) and low depth of coverage(<200) are filtered out  ② -s HighQual : filter name  ③ submitting the job is recommended  commands: sbatch <filename.sh> |

Reference

1. Samtools document (<http://samtools.sourceforge.net/mpileup.shtml>; http://www.htslib.org/workflow/)

2. BCFtools manual page(https://samtools.github.io/bcftools/bcftools.html)