**Supplementary materials**

**Deep sequencing approach in studying influenza virus quasispecies and vaccine ineffectiveness**

1. **Collecting of the information about the viral vaccine strains**In 2017-2018 there were four strains of influenza virus which the vaccine recommended to compose of. <https://www.who.int/influenza/vaccines/virus/recommendations/2017_18_north/en/>

It is recommended that trivalent vaccines for use in the 2017-2018 northern hemisphere influenza season contain the following:

* an A/Michigan/45/2015 (H1N1)pdm09-like virus;
* an A/Hong Kong/4801/2014 (H3N2)-like virus; and
* a B/Brisbane/60/2008-like virus.

It is recommended that quadrivalent vaccines containing two influenza B viruses contain the above three viruses and a B/Phuket/3073/2013-like virus.  
Virus from the roommate was close to A/Hong Kong/4801/2014 (H3N2).

1. **Downloading and inspecting the data**The data were obtained from the NCBI Sequence Read Archive:

[wget ftp://](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/001/SRR1705851/SRR1705851.fastq.gz)[ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/001/SRR1705851/SRR1705851.fastq.gz](http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/001/SRR1705851/SRR1705851.fastq.gz)

First, the data were unzipped and the quality of the reads was examined by the FASTQC program. The data represents Illumina reads from a single-end sequencing run.

gunzip SRR1705851.fastq.gz

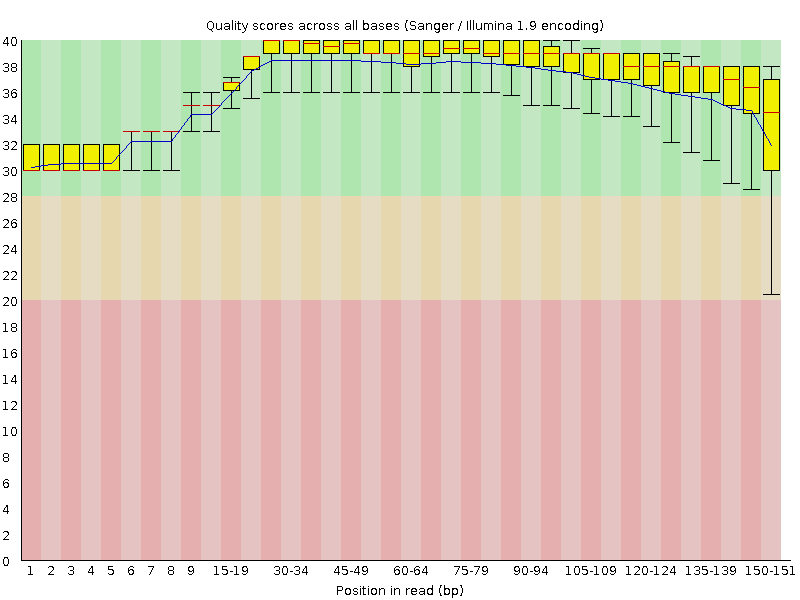
fastqc -o . ./SRR1705851.fastq

Results obtained from fastqc run are shown on the figures 1-9 and in table 1.

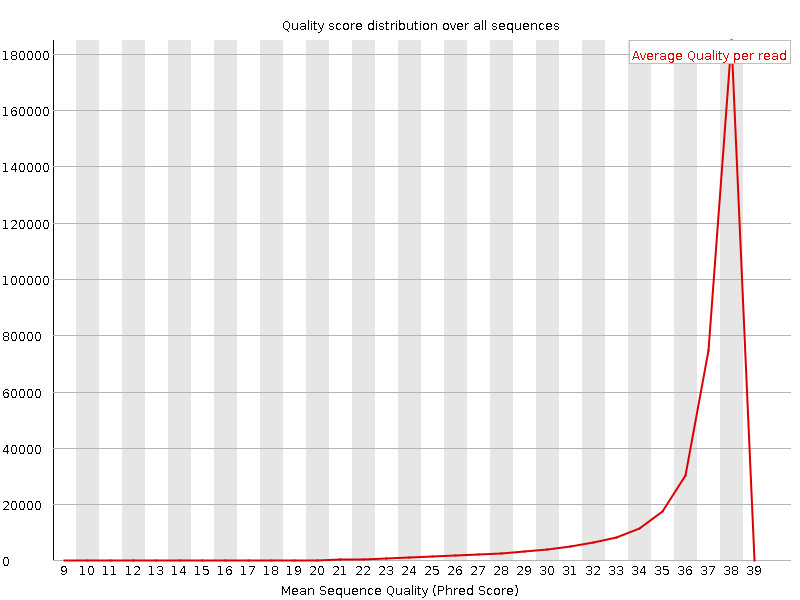
Table 1 - Basic statistics for roommate’s reads.

|  |  |
| --- | --- |
| **Measure** | **Value** |
| Filename | SRR1705851.fastq |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 358265 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 35-151 |
| %GC | 42 |

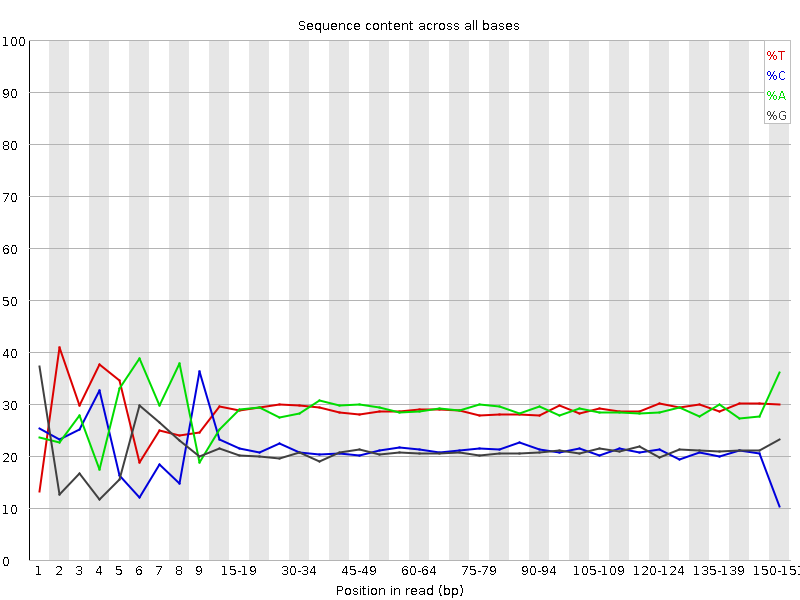
## [OK]**Figure 1 -** **Per base sequence quality**

****

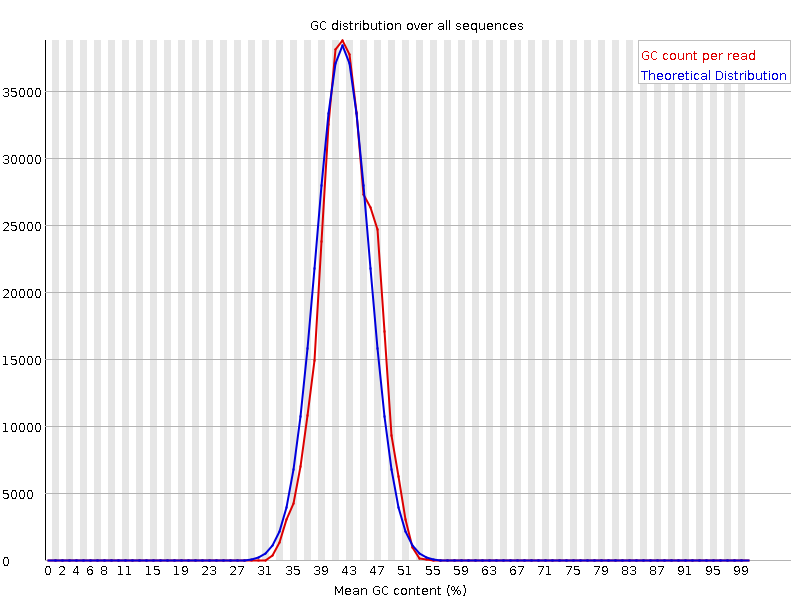
## **[OK]Figure 2 - Per sequence quality scores**

****

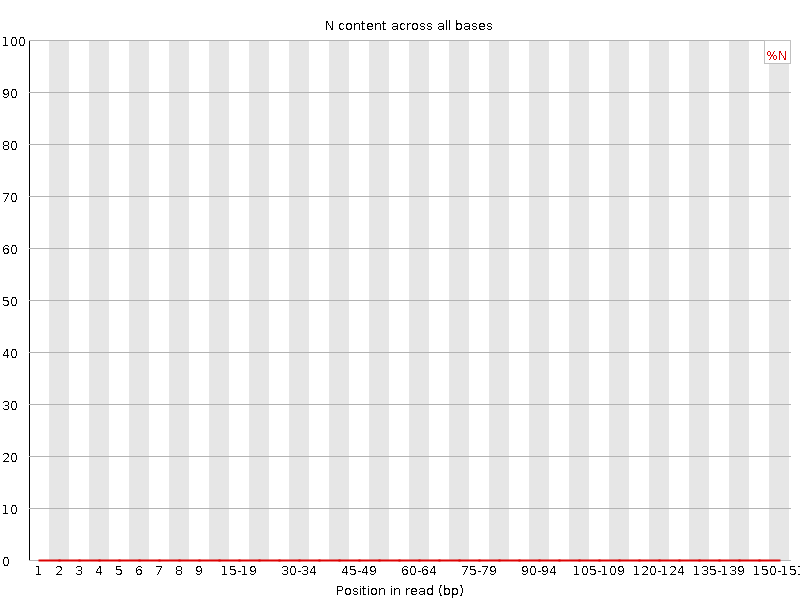
## **Figure 3 - Per base sequence content**

****

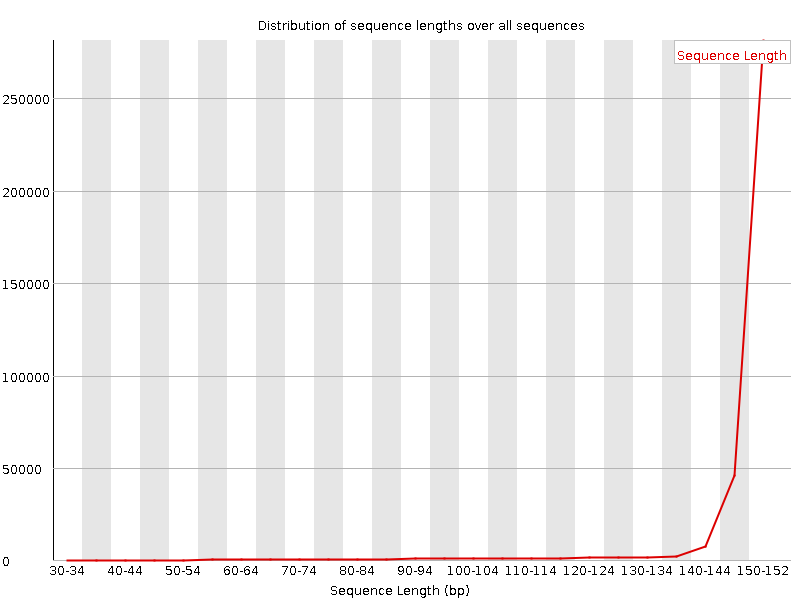
## **[WARN]Figure 4 - Per sequence GC content**



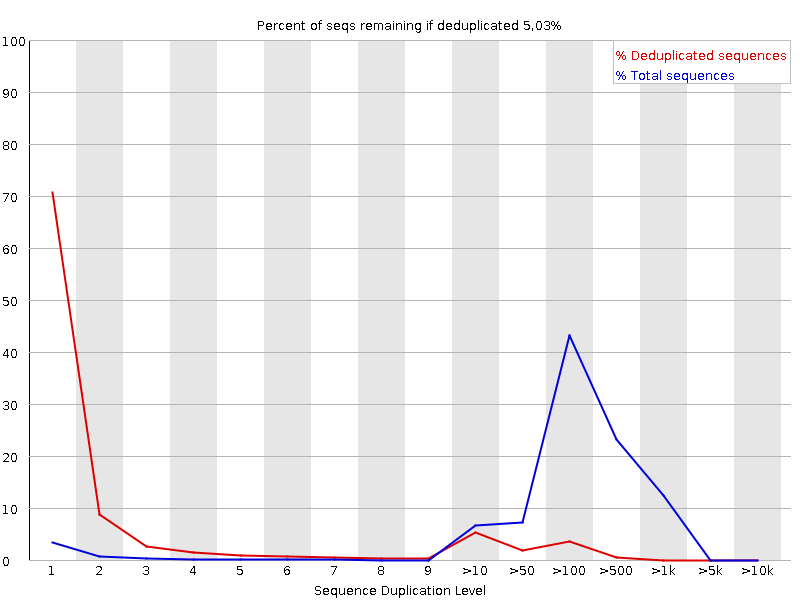
## [OK]**Figure 5 -** **Per base N content**



[WARN]**Figure 6 - Sequence Length Distribution**



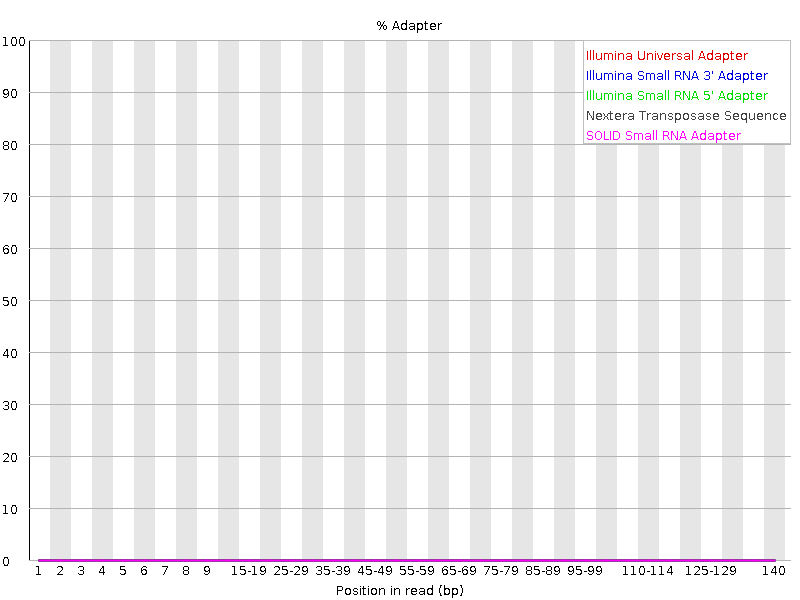
## **Figure 7 - Sequence Duplication Levels**



## [WARN]**Overrepresented sequences**

Presented, No hits

## [OK]**Figure 9 - Adapter Content**



The overall quality of the reads is good. The presence of overrepresented sequences and high level of duplication is normal for the deep sequencing data. The reads were pre-processed as we can see no adapters and no sufficient decrease of per base sequence quality on the graphs. There were 151 cycles in the sequencing run.

1. **Alignment of the roommate’s reads on reference gene**We used a bwa-mem tool to align the reads on the reference sequence of the hemagglutinin gene and a samtools package to convert SAM format to BAM format and sort it, all operations in one pipeline.

bwa index sequence.fasta

bwa mem sequence.fasta SRR1705851.fastq | samtools view -S -b - | samtools sort - -o roommate\_sorted.bam

samtools flagstat roommate\_sorted.bam

Mapping statistics:

361349 + 0 in total (QC-passed reads + QC-failed reads)

0 + 0 secondary

3084 + 0 supplementary

0 + 0 duplicates

361116 + 0 mapped (99.94% : N/A)

0 + 0 paired in sequencing

0 + 0 read1

0 + 0 read2

0 + 0 properly paired (N/A : N/A)

0 + 0 with itself and mate mapped

0 + 0 singletons (N/A : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

1. **Variant calling**We indexed the reference file using samtools.

samtools index roommate\_sorted.bam

Coverage of the reference sequence was calculated manually.

(358032 mapped reads \* 151 bp length\_reads ) / 1665 bp length\_reference = 32470

To find rare variants we set the d parameter more than our coverage while making a mpileup file.

d parameter for samtools mpileup = 33000

samtools mpileup -d 33000 -f sequence.fasta roommate\_sorted.bam > my.mpileup

To look for positions where most of the viruses infecting the roommate differ from the reference we used minimum variant frequency 95%.

varscan mpileup2snp my.mpileup --min-var-freq 0.95 --variants --output-vcf 1 > VarScan\_results.vcf

> VarScan\_results.vcf

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.95

Min avg qual: 15

P-value thresh: 0.01

Reading input from my.mpileup

1665 bases in pileup file

5 variant positions (5 SNP, 0 indel)

0 were failed by the strand-filter

5 variant positions reported (5 SNP, 0 indel)

Then we looked at obtained SNPs with the awk program. It is used for parsing .vcf files.

This is a template for awk command:

*awk '/search\_pattern/ {actiontotakeonmatches; another\_action;}' file\_to\_awk*

*cat VarScan\_results.vcf | awk 'NR>24 {print $1, $2, $4, $5}'*

Result:

KF848938.1 72 A G

KF848938.1 117 C T

KF848938.1 774 T C

KF848938.1 999 C T

KF848938.1 1260 A C

Next we watched the SNPs in IGV browser. List of SNPs represented in Table 2.

Table 2 - SNPs with high frequency in the roommate’s sample.

|  |  |  |
| --- | --- | --- |
| Position | Nucleotide | Amino acid |
| 72 | ACA > ACG | Threonine > Threonine |
| 117 | GCC > GCT | Alanine > Alanine |
| 774 | TTT > TTC | Phenylalanine > Phenylalanine |
| 999 | GGC > GGT | Glycine > Glycine |
| 1260 | CTA > CTC | Leucine > Leucine |

There was no SNP that changes the amino acid sequence.

To look at rare SNPs we set the minimum variant frequency 0,1%.

varscan mpileup2snp my.mpileup --min-var-freq 0.001 --variants --output-vcf 1 > VarScan\_results\_rare.vcf

Results:

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.001

Min avg qual: 15

P-value thresh: 0.01

Reading input from my.mpileup

1665 bases in pileup file

21 variant positions (19 SNP, 2 indel)

0 were failed by the strand-filter

19 variant positions reported (19 SNP, 0 indel)

To see the frequency of the SNPs we used awk again.

cat VarScan\_results\_rare.vcf | awk 'NR>24 {print $1, $2, $4, $5, $10}'

Results:

KF848938.1 72 A G 99,96%

KF848938.1 117 C T 99,82%

KF848938.1 254 A G 0,18%

KF848938.1 307 C T 0,96%

KF848938.1 340 T C 0,18%

KF848938.1 389 T C 0,22%

KF848938.1 722 A G 0,22%

KF848938.1 744 A G 0,18%

KF848938.1 774 T C 99,97%

KF848938.1 802 A G 0,24%

KF848938.1 915 T C 0,2%

KF848938.1 999 C T 99,86%

KF848938.1 1043 A G 0,19%

KF848938.1 1086 A G 0,21%

KF848938.1 1213 A G 0,22%

KF848938.1 1260 A C 99,94%

KF848938.1 1280 T C 0,18%

KF848938.1 1458 T C 0,83%

KF848938.1 1460 A G 0,18%

SNPs shown in green are those that we obtained in the previous run of VarScan with 95% minimum variant frequency threshold.

1. **Inspection and alignment of the control sample sequencing data**

To understand whether those rare SNPs that we had got from our roommate’s data were the real mutations and not the errors we examined the sequencing data of the control. The control was an isogenic viral sample cloned into the vector plasmids and sequenced three times on Illumina platform.

Downloading the data:

wget ftp://[ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/008/SRR1705858/SRR1705858.fastq.gz](http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/008/SRR1705858/SRR1705858.fastq.gz)

wget ftp://[ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/009/SRR1705859/SRR1705859.fastq.gz](http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/009/SRR1705859/SRR1705859.fastq.gz)

wget ftp://[ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/000/SRR1705860/SRR1705860.fastq.gz](http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR170/000/SRR1705860/SRR1705860.fastq.gz)

The fastqc results were almost the same as for the roommate’s sequence data. The coverage was calculated manually.

Control\_1, 256586 reads, coverage 23262

Control\_2 233327 reads, coverage 21153

Control\_3 249964 reads, coverage 22662

Then we performed the alignment with parameters same as for roommate’s reads and got mapping statistics for each control file.

bwa mem sequence.fasta SRR1705858.fastq | samtools view -S -b - | samtools sort - -o control\_1\_sorted.bam

bwa mem sequence.fasta SRR1705859.fastq | samtools view -S -b - | samtools sort - -o control\_2\_sorted.bam

bwa mem sequence.fasta SRR1705860.fastq | samtools view -S -b - | samtools sort - -o control\_3\_sorted.bam

samtools flagstat control\_1\_sorted.bam

Results for control\_1:

256744 + 0 in total (QC-passed reads + QC-failed reads)

0 + 0 secondary

158 + 0 supplementary

0 + 0 duplicates

256658 + 0 mapped (99.97% : N/A)

0 + 0 paired in sequencing

0 + 0 read1

0 + 0 read2

0 + 0 properly paired (N/A : N/A)

0 + 0 with itself and mate mapped

0 + 0 singletons (N/A : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

samtools flagstat control\_2\_sorted.bam

Results for control\_2:

233451 + 0 in total (QC-passed reads + QC-failed reads)

0 + 0 secondary

124 + 0 supplementary

0 + 0 duplicates

233375 + 0 mapped (99.97% : N/A)

0 + 0 paired in sequencing

0 + 0 read1

0 + 0 read2

0 + 0 properly paired (N/A : N/A)

0 + 0 with itself and mate mapped

0 + 0 singletons (N/A : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

samtools flagstat control\_3\_sorted.bam

Results for control\_3:

250184 + 0 in total (QC-passed reads + QC-failed reads)

0 + 0 secondary

220 + 0 supplementary

0 + 0 duplicates

250108 + 0 mapped (99.97% : N/A)

0 + 0 paired in sequencing

0 + 0 read1

0 + 0 read2

0 + 0 properly paired (N/A : N/A)

0 + 0 with itself and mate mapped

0 + 0 singletons (N/A : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

1. **Variant calling on control samples**

After alignment we indexed the sorted files and looked at rare variants with minimum variant frequency 0.1% to examine the sequencing errors.

samtools index control\_1\_sorted.bam

samtools index control\_2\_sorted.bam

samtools index control\_3\_sorted.bam

samtools mpileup -d 25000 -f sequence.fasta control\_1\_sorted.bam > my.mpileup\_1

samtools mpileup -d 25000 -f sequence.fasta control\_2\_sorted.bam > my.mpileup\_2

samtools mpileup -d 25000 -f sequence.fasta control\_3\_sorted.bam > my.mpileup\_3

varscan mpileup2snp my.mpileup\_1 --min-var-freq 0.001 --variants --output-vcf 1 > VarScan\_results\_control\_1.vcf

Results:

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.001

Min avg qual: 15

P-value thresh: 0.01

Reading input from my.mpileup\_1

1665 bases in pileup file

56 variant positions (56 SNP, 0 indel)

1 were failed by the strand-filter

55 variant positions reported (55 SNP, 0 indel)

varscan mpileup2snp my.mpileup\_2 --min-var-freq 0.001 --variants --output-vcf 1 > VarScan\_results\_control\_2.vcf

Results:

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.001

Min avg qual: 15

P-value thresh: 0.01

Reading input from my.mpileup\_2

1665 bases in pileup file

54 variant positions (54 SNP, 0 indel)

2 were failed by the strand-filter

52 variant positions reported (52 SNP, 0 indel)

varscan mpileup2snp my.mpileup\_3 --min-var-freq 0.001 --variants --output-vcf 1 > VarScan\_results\_control\_3.vcf

Results:

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.001

Min avg qual: 15

P-value thresh: 0.01

Reading input from my.mpileup\_3

1665 bases in pileup file

61 variant positions (61 SNP, 0 indel)

0 were failed by the strand-filter

61 variant positions reported (61 SNP, 0 indel)

cat VarScan\_results\_control\_1.vcf | awk 'NR>24 {print $1, $2, $4, $5, $10}'

cat VarScan\_results\_control\_2.vcf | awk 'NR>24 {print $1, $2, $4, $5, $10}'

cat VarScan\_results\_control\_3.vcf | awk 'NR>24 {print $1, $2, $4, $5, $10}'

We obtained information about positions, nucleotides and frequency of the SNPs.

Table 3 - Results of variant calling for control samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample | Position | Nucleotide\_1 | Nucleotide\_2 | Frequency |
| Control\_1 | 38 | T | C | 0,66 |
| Control\_1 | 54 | T | C | 0,3 |
| Control\_1 | 72 | A | G | 0,3 |
| Control\_1 | 95 | A | G | 0,24 |
| Control\_1 | 117 | C | T | 0,3 |
| Control\_1 | 165 | T | C | 0,24 |
| Control\_1 | 183 | A | G | 0,3 |
| Control\_1 | 216 | A | G | 0,22 |
| Control\_1 | 218 | A | G | 0,28 |
| Control\_1 | 222 | T | C | 0,26 |
| Control\_1 | 235 | T | C | 0,25 |
| Control\_1 | 254 | A | G | 0,25 |
| Control\_1 | 276 | A | G | 0,22 |
| Control\_1 | 297 | T | C | 0,19 |
| Control\_1 | 319 | T | C | 0,18 |
| Control\_1 | 328 | T | C | 0,19 |
| Control\_1 | 340 | T | C | 0,25 |
| Control\_1 | 356 | A | G | 0,22 |
| Control\_1 | 370 | A | G | 0,23 |
| Control\_1 | 389 | T | C | 0,27 |
| Control\_1 | 409 | T | C | 0,21 |
| Control\_1 | 414 | T | C | 0,28 |
| Control\_1 | 463 | A | G | 0,2 |
| Control\_1 | 516 | A | G | 0,2 |
| Control\_1 | 566 | A | G | 0,26 |
| Control\_1 | 595 | G | T | 0,33 |
| Control\_1 | 660 | A | G | 0,22 |
| Control\_1 | 670 | A | G | 0,24 |
| Control\_1 | 691 | A | G | 0,21 |
| Control\_1 | 722 | A | G | 0,25 |
| Control\_1 | 744 | A | G | 0,23 |
| Control\_1 | 774 | T | C | 0,3 |
| Control\_1 | 859 | A | G | 0,26 |
| Control\_1 | 915 | T | C | 0,28 |
| Control\_1 | 987 | A | G | 0,21 |
| Control\_1 | 1008 | T | G | 0,26 |
| Control\_1 | 1031 | A | G | 0,29 |
| Control\_1 | 1043 | A | G | 0,25 |
| Control\_1 | 1056 | T | C | 0,2 |
| Control\_1 | 1086 | A | G | 0,33 |
| Control\_1 | 1089 | A | G | 0,23 |
| Control\_1 | 1213 | A | G | 0,24 |
| Control\_1 | 1260 | A | C | 0,3 |
| Control\_1 | 1264 | T | C | 0,26 |
| Control\_1 | 1280 | T | C | 0,25 |
| Control\_1 | 1281 | T | C | 0,22 |
| Control\_1 | 1286 | T | C | 0,2 |
| Control\_1 | 1339 | T | C | 0,41 |
| Control\_1 | 1358 | A | G | 0,26 |
| Control\_1 | 1398 | T | C | 0,2 |
| Control\_1 | 1421 | A | G | 0,31 |
| Control\_1 | 1460 | A | G | 0,34 |
| Control\_1 | 1482 | A | G | 0,24 |
| Control\_1 | 1580 | T | C | 0,25 |
| Control\_1 | 1591 | T | C | 0,29 |
| Control\_2 | 44 | T | C | 0,47 |
| Control\_2 | 158 | A | G | 0,24 |
| Control\_2 | 165 | T | C | 0,27 |
| Control\_2 | 183 | A | G | 0,22 |
| Control\_2 | 193 | A | G | 0,22 |
| Control\_2 | 216 | A | G | 0,24 |
| Control\_2 | 218 | A | G | 0,29 |
| Control\_2 | 222 | T | C | 0,25 |
| Control\_2 | 254 | A | G | 0,19 |
| Control\_2 | 276 | A | G | 0,24 |
| Control\_2 | 319 | T | C | 0,23 |
| Control\_2 | 340 | T | C | 0,21 |
| Control\_2 | 356 | A | G | 0,24 |
| Control\_2 | 370 | A | G | 0,21 |
| Control\_2 | 389 | T | C | 0,2 |
| Control\_2 | 398 | A | G | 0,22 |
| Control\_2 | 409 | T | C | 0,19 |
| Control\_2 | 414 | T | C | 0,22 |
| Control\_2 | 421 | A | G | 0,19 |
| Control\_2 | 463 | A | G | 0,19 |
| Control\_2 | 499 | A | G | 0,21 |
| Control\_2 | 516 | A | G | 0,2 |
| Control\_2 | 548 | A | G | 0,2 |
| Control\_2 | 591 | A | G | 0,19 |
| Control\_2 | 607 | A | G | 0,18 |
| Control\_2 | 660 | A | G | 0,28 |
| Control\_2 | 670 | A | G | 0,32 |
| Control\_2 | 691 | A | G | 0,19 |
| Control\_2 | 722 | A | G | 0,24 |
| Control\_2 | 744 | A | G | 0,26 |
| Control\_2 | 793 | A | G | 0,2 |
| Control\_2 | 859 | A | G | 0,29 |
| Control\_2 | 898 | A | G | 0,2 |
| Control\_2 | 915 | T | C | 0,22 |
| Control\_2 | 987 | A | G | 0,22 |
| Control\_2 | 1031 | A | G | 0,28 |
| Control\_2 | 1056 | T | C | 0,19 |
| Control\_2 | 1086 | A | G | 0,21 |
| Control\_2 | 1100 | T | C | 0,21 |
| Control\_2 | 1213 | A | G | 0,22 |
| Control\_2 | 1264 | T | C | 0,21 |
| Control\_2 | 1280 | T | C | 0,24 |
| Control\_2 | 1358 | A | G | 0,25 |
| Control\_2 | 1366 | A | G | 0,22 |
| Control\_2 | 1398 | T | C | 0,23 |
| Control\_2 | 1421 | A | G | 0,24 |
| Control\_2 | 1460 | A | G | 0,37 |
| Control\_2 | 1482 | A | G | 0,25 |
| Control\_2 | 1517 | A | G | 0,24 |
| Control\_2 | 1520 | T | C | 0,27 |
| Control\_2 | 1600 | T | C | 0,35 |
| Control\_2 | 1604 | T | C | 0,31 |
| Control\_3 | 38 | T | C | 0,7 |
| Control\_3 | 44 | T | C | 0,5 |
| Control\_3 | 95 | A | G | 0,24 |
| Control\_3 | 105 | A | G | 0,25 |
| Control\_3 | 133 | A | G | 0,22 |
| Control\_3 | 158 | A | G | 0,26 |
| Control\_3 | 165 | T | C | 0,25 |
| Control\_3 | 183 | A | G | 0,23 |
| Control\_3 | 199 | A | G | 0,19 |
| Control\_3 | 216 | A | G | 0,24 |
| Control\_3 | 218 | A | G | 0,23 |
| Control\_3 | 222 | T | C | 0,3 |
| Control\_3 | 228 | T | C | 0,19 |
| Control\_3 | 230 | A | G | 0,19 |
| Control\_3 | 235 | T | C | 0,25 |
| Control\_3 | 254 | A | G | 0,23 |
| Control\_3 | 271 | A | G | 0,21 |
| Control\_3 | 276 | A | G | 0,33 |
| Control\_3 | 297 | T | C | 0,23 |
| Control\_3 | 319 | T | C | 0,2 |
| Control\_3 | 340 | T | C | 0,21 |
| Control\_3 | 356 | A | G | 0,22 |
| Control\_3 | 370 | A | G | 0,24 |
| Control\_3 | 389 | T | C | 0,2 |
| Control\_3 | 409 | T | C | 0,2 |
| Control\_3 | 414 | T | C | 0,32 |
| Control\_3 | 421 | A | G | 0,21 |
| Control\_3 | 463 | A | G | 0,2 |
| Control\_3 | 499 | A | G | 0,19 |
| Control\_3 | 566 | A | G | 0,29 |
| Control\_3 | 597 | A | G | 0,2 |
| Control\_3 | 607 | A | G | 0,19 |
| Control\_3 | 660 | A | G | 0,28 |
| Control\_3 | 670 | A | G | 0,26 |
| Control\_3 | 691 | A | G | 0,2 |
| Control\_3 | 722 | A | G | 0,3 |
| Control\_3 | 744 | A | G | 0,25 |
| Control\_3 | 759 | T | C | 0,21 |
| Control\_3 | 859 | A | G | 0,19 |
| Control\_3 | 915 | T | C | 0,28 |
| Control\_3 | 987 | A | G | 0,22 |
| Control\_3 | 1031 | A | G | 0,24 |
| Control\_3 | 1043 | A | G | 0,22 |
| Control\_3 | 1056 | T | C | 0,2 |
| Control\_3 | 1086 | A | G | 0,29 |
| Control\_3 | 1089 | A | G | 0,22 |
| Control\_3 | 1105 | A | G | 0,22 |
| Control\_3 | 1209 | A | G | 0,27 |
| Control\_3 | 1213 | A | G | 0,24 |
| Control\_3 | 1264 | T | C | 0,27 |
| Control\_3 | 1280 | T | C | 0,25 |
| Control\_3 | 1281 | T | C | 0,21 |
| Control\_3 | 1301 | A | G | 0,22 |
| Control\_3 | 1358 | A | G | 0,29 |
| Control\_3 | 1366 | A | G | 0,21 |
| Control\_3 | 1398 | T | C | 0,23 |
| Control\_3 | 1421 | A | G | 0,37 |
| Control\_3 | 1460 | A | G | 0,26 |
| Control\_3 | 1482 | A | G | 0,23 |
| Control\_3 | 1580 | T | C | 0,27 |
| Control\_3 | 1604 | T | C | 0,3 |

Some statistics were calculated using R version 3.6.3 and dplyr package.

Table 4 - Statistics for control variant frequencies.

|  |  |  |
| --- | --- | --- |
| Sample | mean | sd |
| Control\_1 | 0.261090909090909 | 0.0705710235172295 |
| Control\_2 | 0.238846153846154 | 0.0527165504448024 |
| Control\_3 | 0.250983606557377 | 0.077903892030133 |

1. **Assessment of the roommate’s data using control frequencies**

We standardized the frequencies of variants in the roommate's data using mean values of control samples.

Table 5 - Results of variant calling for roommate’s sample with standardized values of variant frequencies.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Sample | Position | Nucleotide\_1 | Nucleotide\_2 | Frequency | Frequency\_std |
| 1 | Neighbour | 72 | A | G | 99.96 | 1486.788 |
| 2 | Neighbour | 117 | C | T | 99.82 | 1484.701 |
| 3 | Neighbour | 254 | A | G | 0.18 | -1.048 |
| 4 | Neighbour | 307 | C | T | 0.96 | 10.582 |
| 5 | Neighbour | 340 | T | C | 0.18 | -1.048 |
| 6 | Neighbour | 389 | T | C | 0.22 | -0.452 |
| 7 | Neighbour | 722 | A | G | 0.22 | -0.452 |
| 8 | Neighbour | 744 | A | G | 0.18 | -1.048 |
| 9 | Neighbour | 774 | T | C | 99.97 | 1486.937 |
| 10 | Neighbour | 802 | A | G | 0.24 | -0.154 |
| 11 | Neighbour | 915 | T | C | 0.2 | -0.75 |
| 12 | Neighbour | 999 | C | T | 99.86 | 1485.297 |
| 13 | Neighbour | 1043 | A | G | 0.19 | -0.899 |
| 14 | Neighbour | 1086 | A | G | 0.21 | -0.601 |
| 15 | Neighbour | 1213 | A | G | 0.22 | -0.452 |
| 16 | Neighbour | 1260 | A | C | 99.94 | 1486.49 |
| 17 | Neighbour | 1280 | T | C | 0.18 | -1.048 |
| 18 | Neighbour | 1458 | T | C | 0.83 | 8.644 |
| 19 | Neighbour | 1460 | A | G | 0.18 | -1.048 |

Mutations shown in red in table 5 represent low frequency variants that can belong to quasispecies.

1. **Epitope mapping**

According to Munoz et al position 307 is located in D-epitope of hemagglutinin protein and SNP in this position can cause lower affinity to antibodies produced after vaccination.