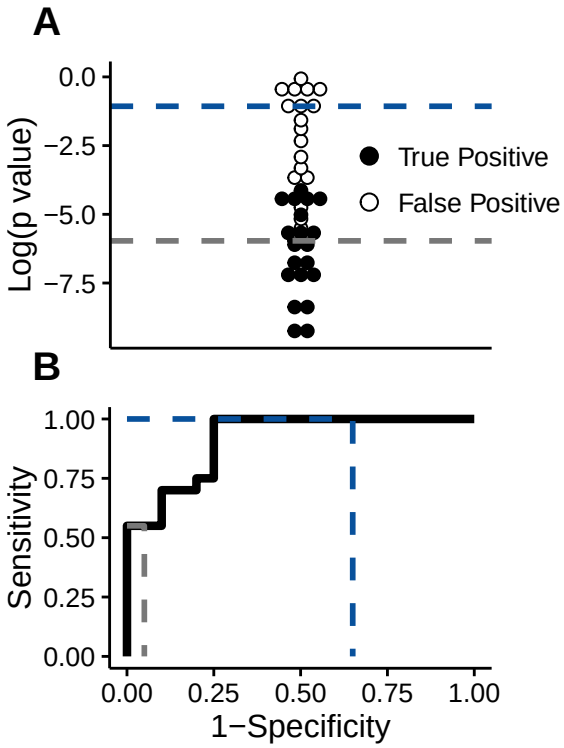
Benchmarking figures

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

Figure 1



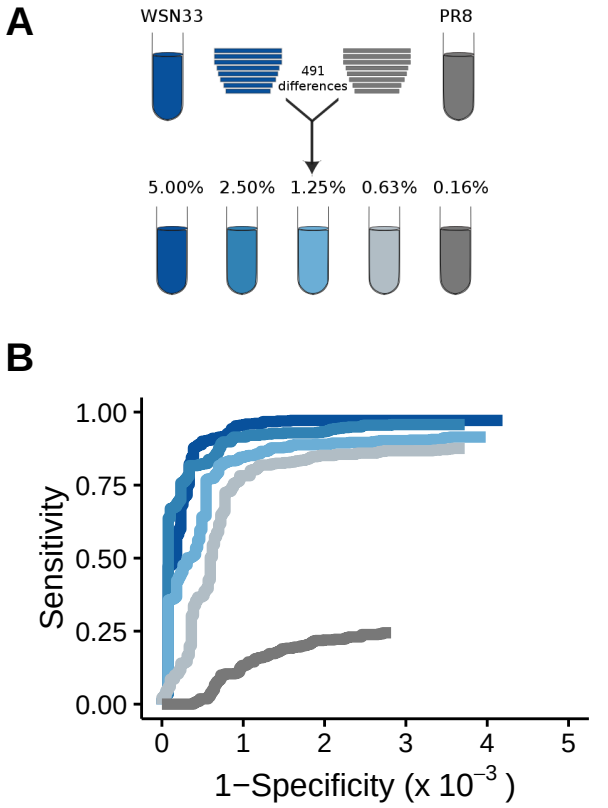

```
## pdf
##   2
```

```
## pdf
##   2
```

Figure 1. An example ROC curve. A) Hypothetical variants are stratified by log of the p value. True positives variants are shown as closed circles, while false positive variants are represented as open circles. Potential thresholds are indicated as colored lines. B) An ROC curve made from the hypothetical data shown in A. The colored lines indicate the points on the curve made by the thresholds in A.

Figure 2

```
## pdf
##   2
```



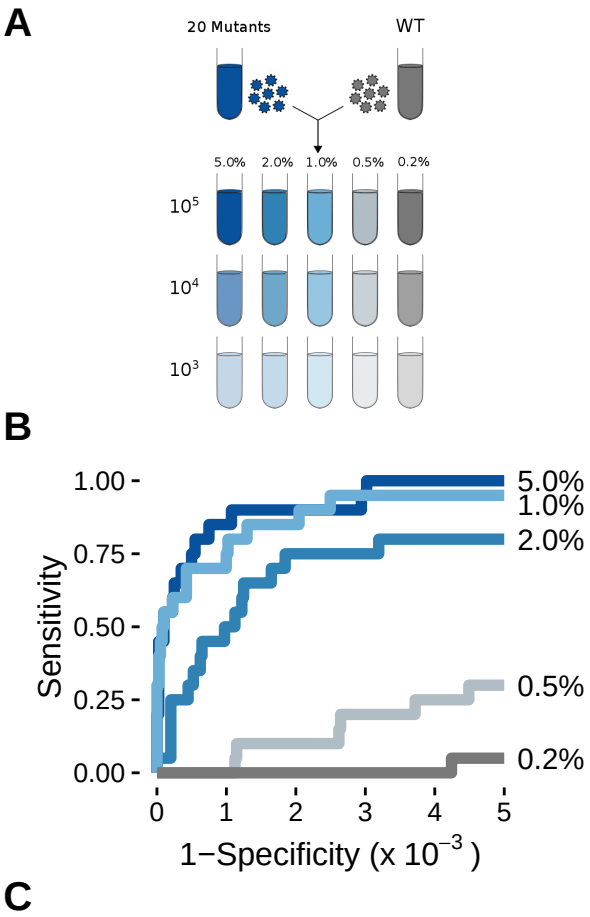
C

Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	0.9715	477	0.9982	68
2.5%	0.9348	459	0.9980	78
1.25%	0.8900	437	0.9981	75
0.63%	0.8330	409	0.9983	66
0.16%	0.1426	70	0.9989	41

Figure 2. Initial DeepSNV accuracy. A) Reconstitued cDNA genomes of influenza strain WSN33 were diluted serially into PR8 prior to sequencing on an Illumina Miseq machine. B) An ROC curve measuring the accuracy of DeepSNV in identifying WSN33 variants mixed with PR8 at the indicated frequencies. C) A summary the data in B at a p value threshold of 0.01.

Figure 3

```
## pdf
##   2
```



Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	1.00	20	0.9938	249
2.0%	0.85	17	0.9942	232
1.0%	0.95	19	0.9975	101
0.5%	0.35	7	0.9948	210
0.2%	0.05	1	0.9937	252

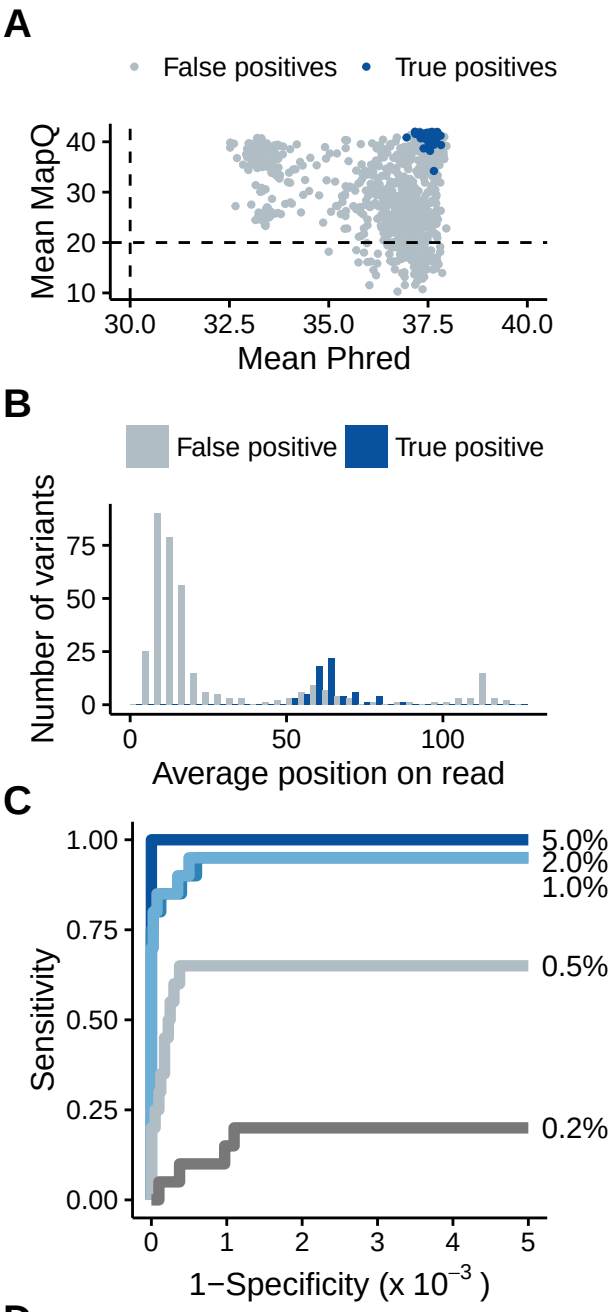
Figure 3. Decreased accuracy under patient-derived sample conditions A) 20 known single point mutants were diluted into WT WSN33 at decreasing frequencies. These populations were diluted further in basal media to match the genome concentrations found in patint-dervied samples (10^5-10^3 genomes/μl). B) The accuracy of DeepSNV in idenfying the 20 known true positive variants in the 10 ^{5} genomes/μl samples. C) A summary table of the accuracy of our method for the 10 ^{5} genomes/μl samples at a p value threshold of 0.01.

Figure 4

```
## pdf
##   2
```

```
## pdf
##   2
```

```
## pdf
##   2
```



D

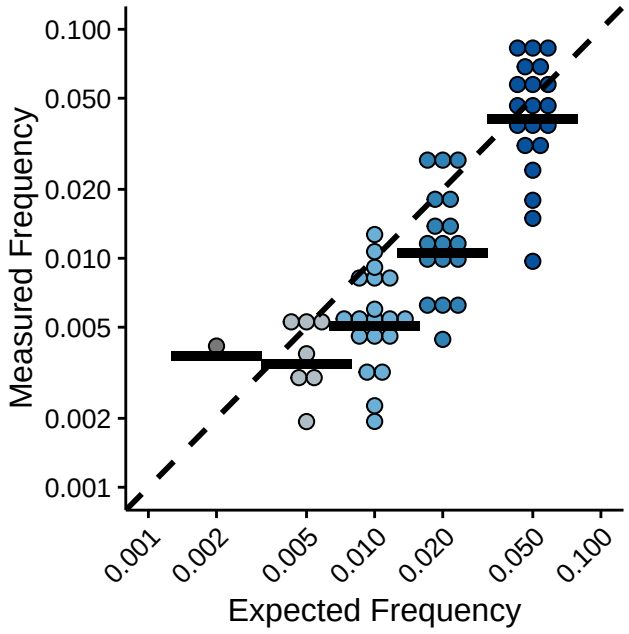
Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	1.00	20	>0.9999	1
2.0%	0.85	17	0.9999	6
1.0%	0.95	19	0.9995	20
0.5%	0.35	7	0.9999	5
0.2%	0.05	1	0.9997	11

Figure 4. Accuracy can be improved through quality thresholds. A) We stratified all called variants from the

10^{5} genomes/μl samples with p<0.01 stratified by the mean mapping quality of the reads containing the variant and the mean Phred scores of the variant bases. Dashed lines indicate common cutoffs of 20 and 30 for mapping quality and Phred respectively. B) A histogram of average position on a pair end read the variants that passed our mean MapQ threshold of 30 and mean Phred threshold of 35. C) The accuracy of our analyis after applying the following quality cut offs : mean MapQ> 30, mean Phred> 35, and an average read position between 32 and 94 (the middle 50% of the read). D) A table summarizing accuracy in C at a p value threshold of 0.01.

Figure 5

```
## [1] 0.4072225
```

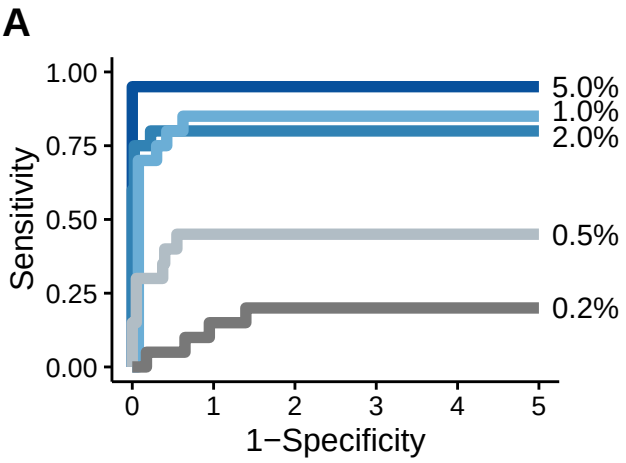



```
## pdf
##   2
```

Figure 5. The accuracy of the frequency measurements for the known variants in the 10^{5} genomes/μl samples. The median of each distribution is shown as a black bar.

Figure 6

```
## pdf
##   2
```



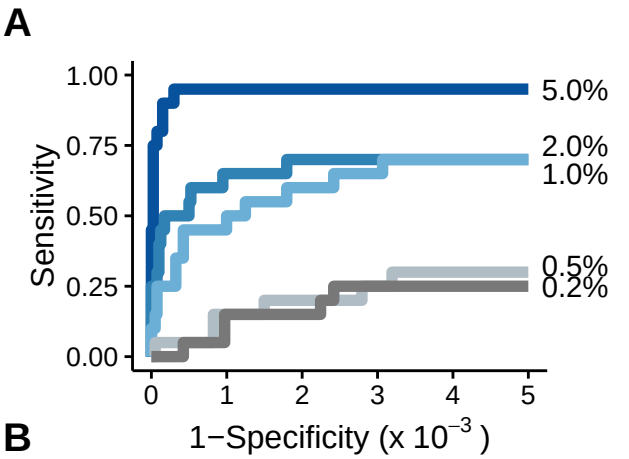
B

Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	0.95	19	0.9987	52
2.0%	0.80	16	0.9989	46
1.0%	0.85	17	0.9976	98
0.5%	0.45	9	0.9987	53
0.2%	0.20	4	0.9985	61

Figure 6. The accuracy of the Lofreq algorithm on our 10^5 genomes/μl populations. A) The accuracy of Lofreq using standard parameters. The specificity of the caller was scaled to account for the same number of tests as the DeepSNV algorithm. B) A table summarizing the accuracy of the ROC A using the standard cut offs applied by the algorithm.

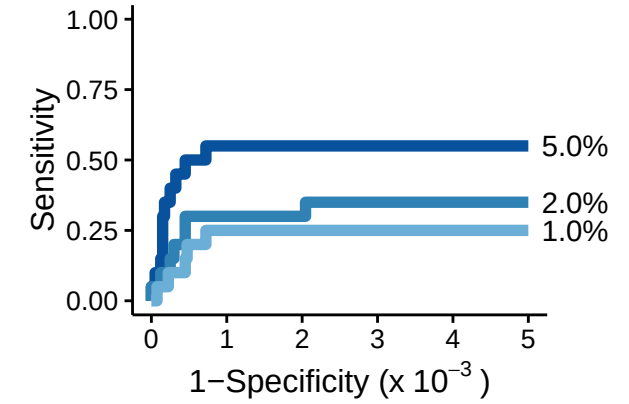
Figure 7

```
## pdf
##   2
```



B

Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	0.95	19	0.9990	41
2.0%	0.65	13	0.9985	60
1.0%	0.55	11	0.9983	67
0.5%	0.20	4	0.9983	69
0.2%	0.15	3	0.9983	69



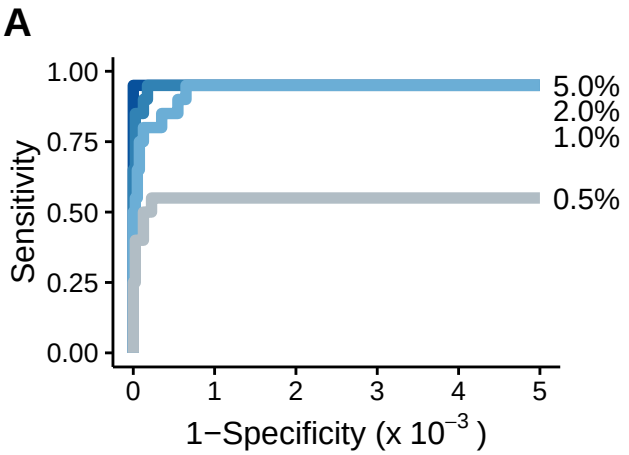
D

Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	0.55	11	0.9990	42
2.0%	0.30	6	0.9981	78
1.0%	0.25	5	0.9979	86

Figure 7. The accuracy of DeepSNV decreased when applied to samples with less input nucleic acid. A) An ROC curve for the 10^{4} genomes/μl samples. B) A table summarizing the accuracy at 10^{4} genomes/μl with a p value threshold of 0.01. C) An ROC curve depicting the accuracy of our method when applied to the 10^{3} genomes/μl samples. B) A table summarizing the accuracy at 10^{3} genomes/μl with a p value threshold of 0.01

Figure 8

```
## pdf
##   2
```



B

Frequency	Sensitivity	True Positives	Specificity	False Positives
5.0%	0.95	19	0.9999	4
2.0%	0.90	18	0.9999	6
1.0%	0.80	16	0.9998	10
0.5%	0.40	8	0.9999	3

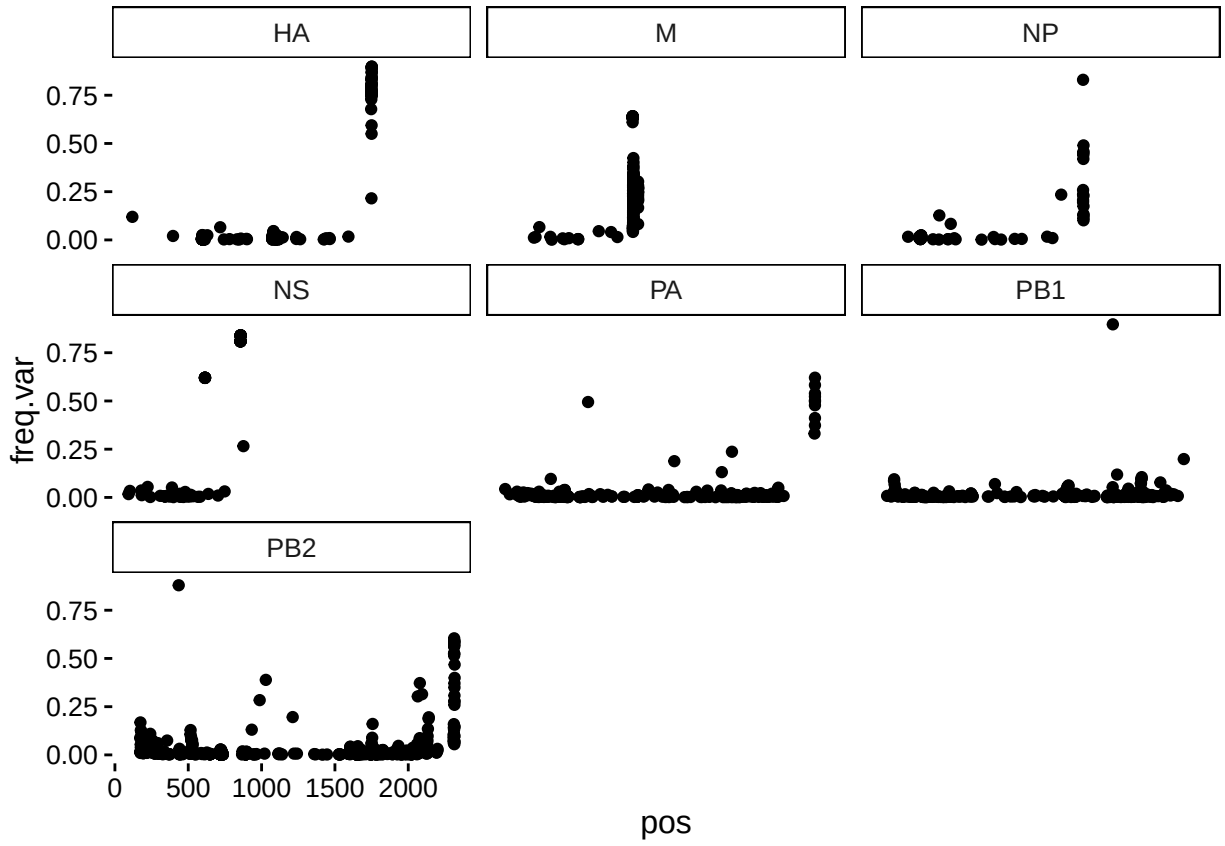
Figure 8. The accuracy of DeepSNV at lower input levels is improved by processing samples in duplicate. A) An ROC curve of the 10^{4} genomes/μl samples processed in duplicate. A variant was required to be found in both duplicates to be considered. B) A table summarizing the accuracy when applied to duplicate samples with 10^{4} genomes/μl input, $p<0.01$.

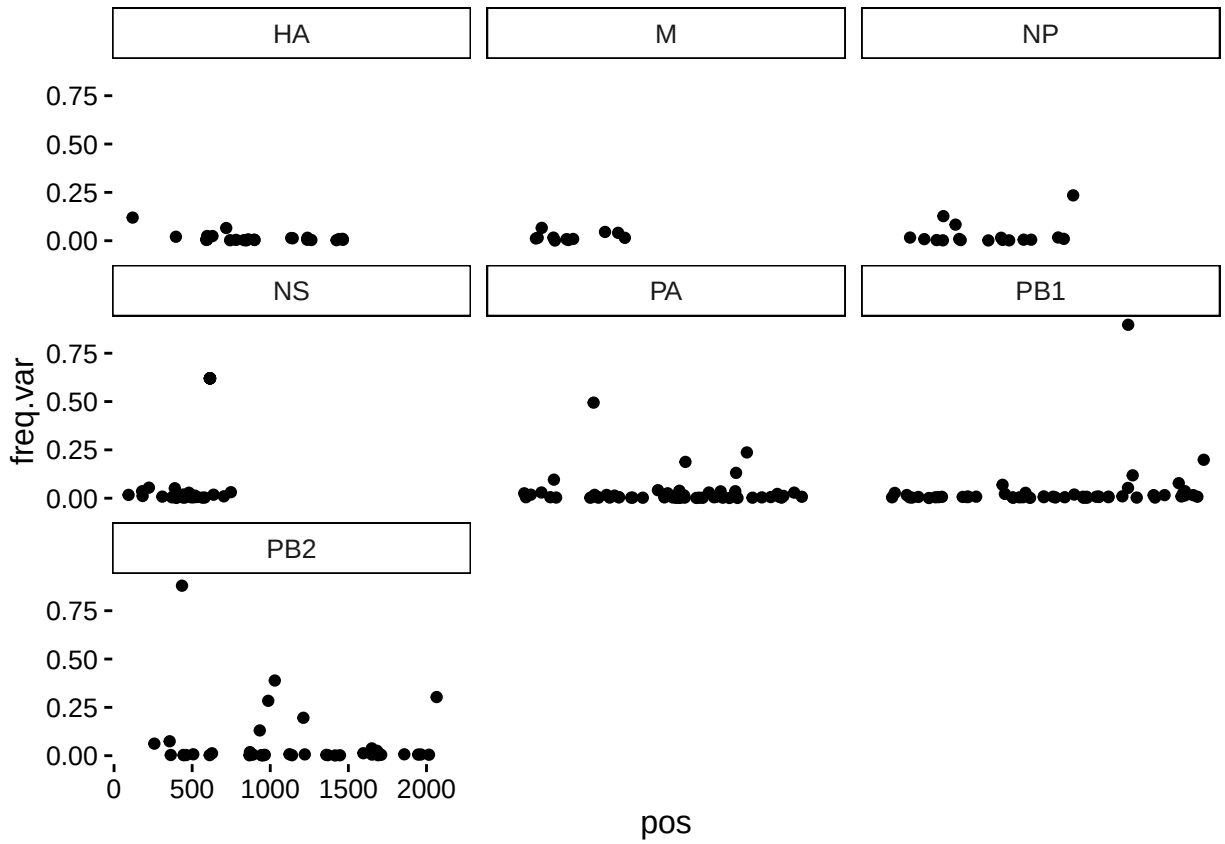
Table 1.

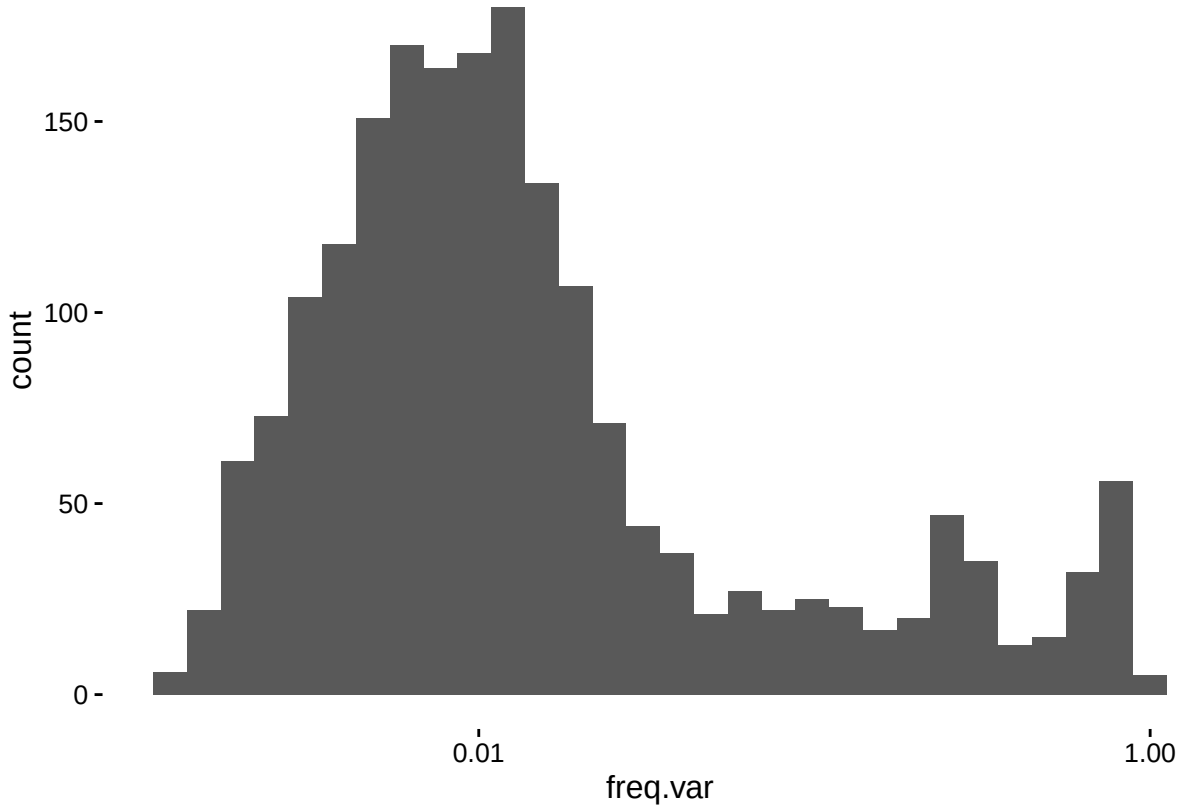
Table 1. Accurate SNV indentification is vital for accurate measure of intrahost diversity. We measured the richness and Shannon's Entropy of the designated 10^5 genomesμl populations at each stage of our benchmarking procedure and compared the results with what is expected. Additionaly we calculated the L1-norm of the populations relative to the expected population.

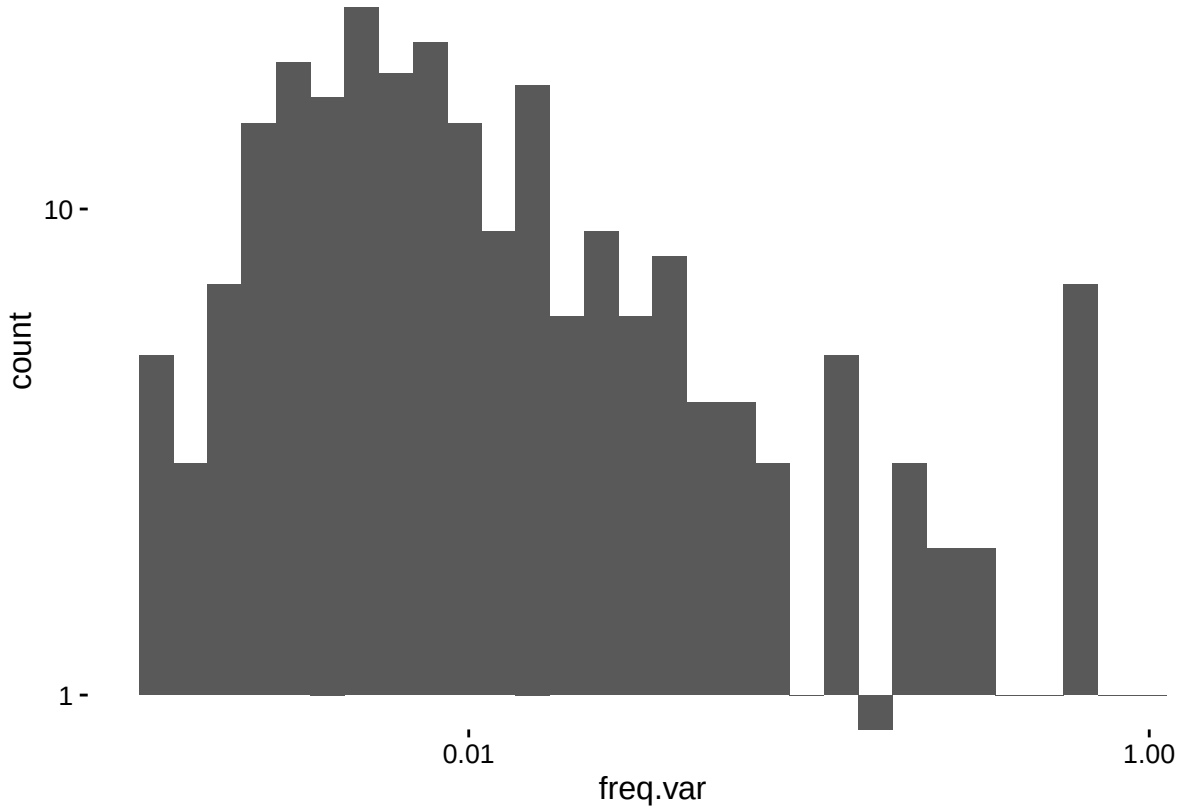
% latex table generated in R 3.2.3 by xtable 1.8-2 package % Fri Apr 8 13:15:51 2016

Frequency		Expected	DeepSNV modified	Lofreq	DeepSNV
	Richness	20	21	71	269
5.0%	Entropy	2.97e-04	2.77e-04	3.35e-04	1.6e-03
	L1-norm	0	0.378	0.519	4.006
	Richness	20	39	115	120
1.0%	Entropy	8.38e-05	7.3e-05	2.78e-04	3.14e-04
	L1-norm	0	0.133	2.702	0.704
	Richness	20	12	62	217
0.5%	Entropy	4.71e-05	2.1e-05	8.47e-05	1.12e-03
	L1-norm	0	0.089	0.196	3.156




HA
M
NP
NS
PA
PB1
PB2
freq.var
0.75
0.50
0.25
0.00
0
500
1000
1500
2000
pos


count
150
100
50
0
0.01
1.00
freq.var


10
1
count
0.01
1.00
freq.var