

Viral Beacon Statistics part 2

1 Stats on variants

Statistics are calculated using [Query by region -by names/aliases](#) and [Filters](#).

I would keep separate this by sequencing technology (platform): Illumina vs Nanopore (for now only Illumina) and of course fastq vs consensus (for now only fastq)

1. Number of genomic positions with variants: 28754/ Number of genomic positions: 29903 (96.15758%)
2. Frequency of variants per position: [fig 1](#)
3. Number of variants per position: (3: 5769; 2: 11021; 1:11964) [fig 1](#)
4. Number of variants in database: 51313
5. Graph: Number of variants, split by options

option Split by variant frequency (groups) [fig 2](#)

option Split by variant type field: SNP, indels (only SNPs) [Mau to check pipeline?](#) [fig 3](#)

option Split by genomic region: coding: all with genomic region=CODING, non-coding: the rest [fig 3](#) (further UTRs, intergenic?)

option Split by molecular consequence (grouped: SYN: SILENT, NON-SYN: MISSENSE+NONSENSE, NONCODING:the rest) and further in all classes as in [fig 3](#)

option See Per region Statistics: Distribution in genomic regions. Upon click on genomic region graph, expand to see per number of variants distribution in genomic regions: select NON-CODING, and within CODING show options: gene, cds/mature peptide. This will show individual components of each class [fig 4](#) and [5](#).

6. Number of positions with aminoacid substitutions in database/ Number of coding positions
7. Number of variants producing unique aminoacid substitutions in database: 18308

option Filter upon cds/mature peptide. Number of variants producing aminoacid substitutions (aminoacid change) (eg. "G507C") are aggregated in mature protein region

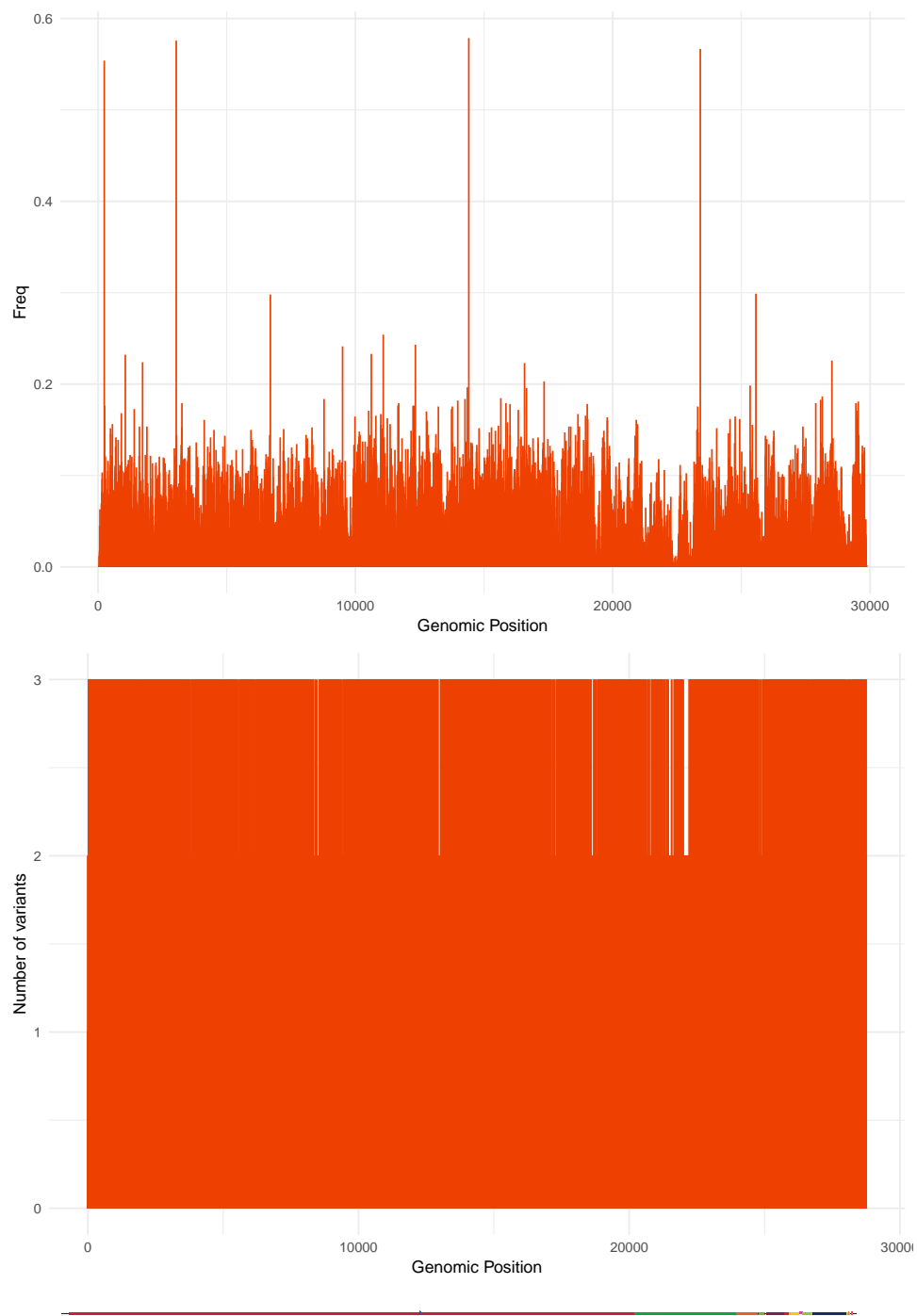


Figure 1: Needle plot: Top: Frequency of variants per position. Bottom: Number of alternates per position

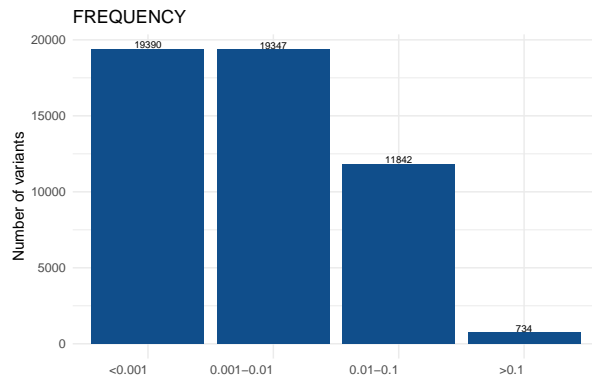


Figure 2: Number of variants per frequency group

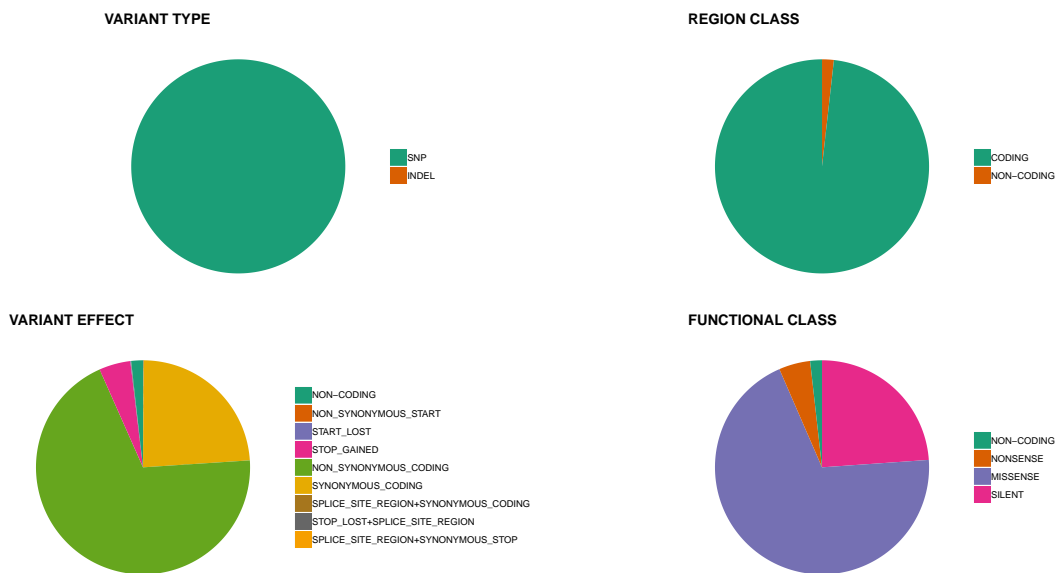


Figure 3: Number of variants, split by variant type, region class, variant effect and mol consequence

See variants distribution by region: [genes](#) [cds/mat_peptides](#) [noncoding](#) [stem loops](#)

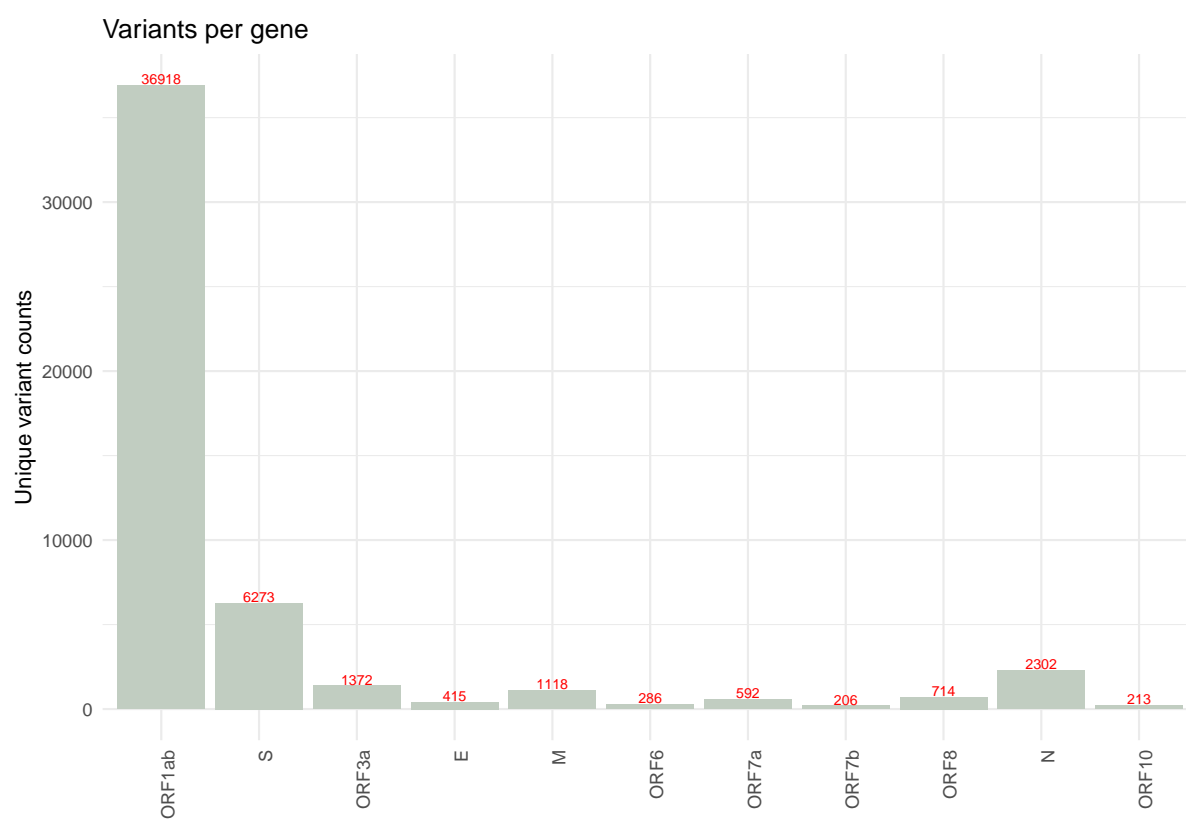


Figure 4: Number of variants in coding regions of SARS-CoV2, shown per genes, shown directly on clicking CODING area of genomic region graph or in option "See variants distribution by region" [genes](#).

See variants distribution by region:

[genes](#)

[cds/ mat peptides](#)

[noncoding](#)

[stem loops](#)

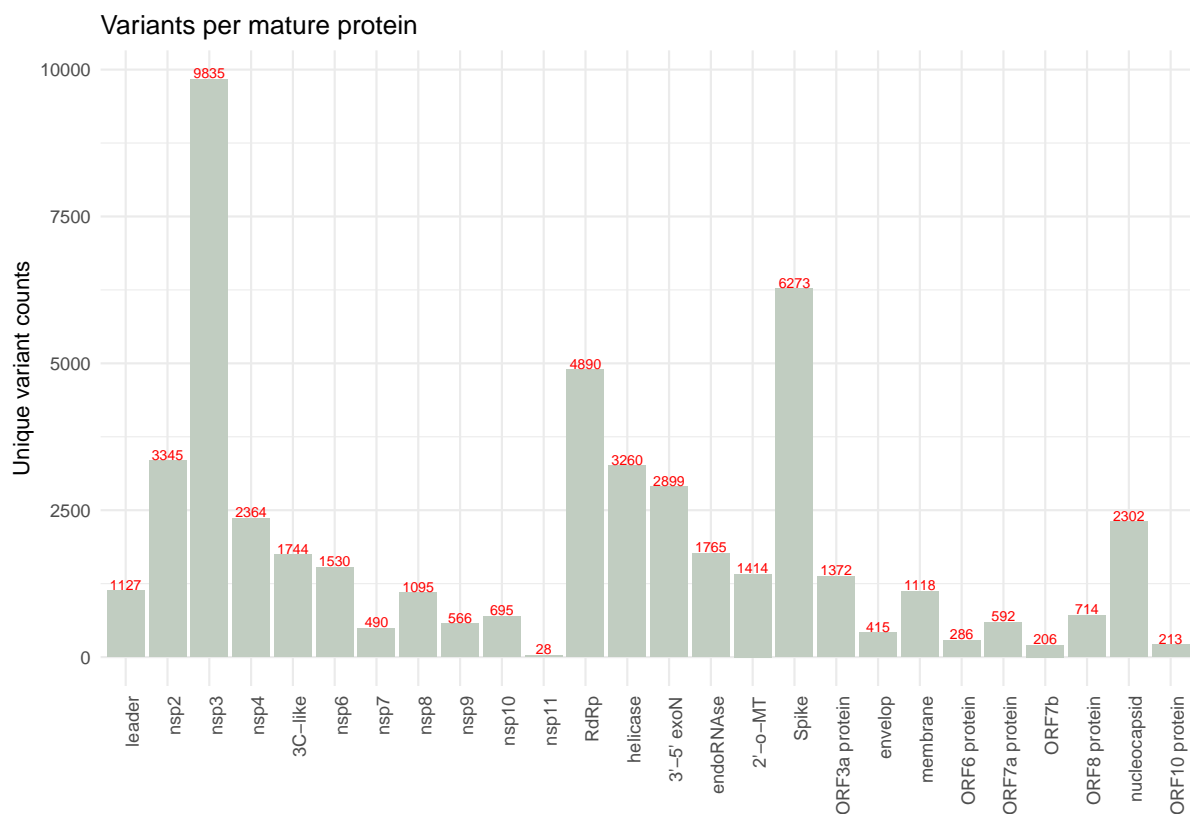


Figure 5: Number of variants in coding regions of SARS-CoV2, shown per mature proteins, shown on clicking cds/mature peptide option within CODING area of genomic region graph (which will show directly upon clicking the variants per gene graph 4 or in option "See variants distribution by region" [cds/mat.peptides](#). Additional options within this graph: [split by syn/non syn](#) (or default in stack bar)



Figure 6: Number of variants in non-coding regions of SARS-CoV2, shown per region, shown on clicking NON-CODING area of genomic region graph or in option "See variants distribution by region" [noncoding](#).

See variants distribution by region: [genes](#) [cds/mat_peptides](#) [noncoding](#) [stem loops](#)

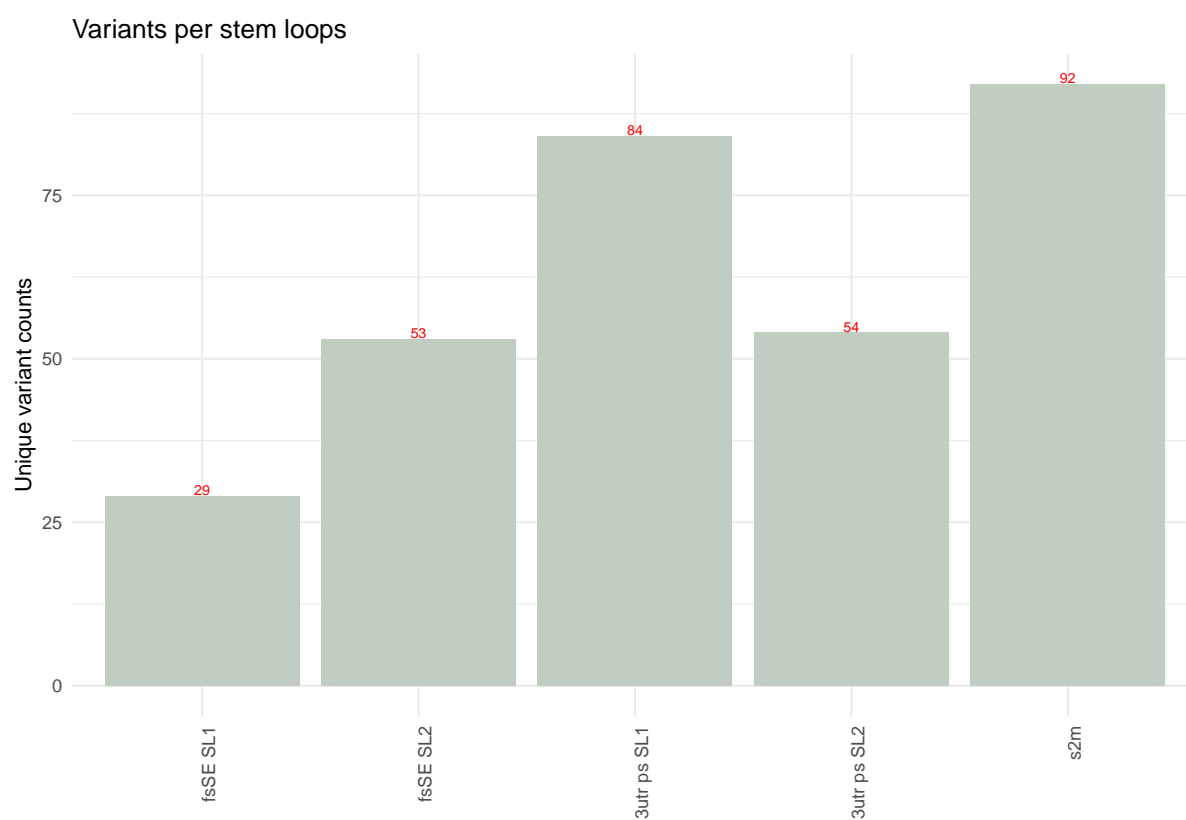


Figure 7: Number of variants in stem loops regions of SARS-CoV2, shown per region, shown on clicking in option "See variants distribution by region" [stem loops](#).