

Ideas for Variants Statistics in Beacon

1 Number of variants per functional region (proteins, utrs, stem loops)

By using `Query by region -by names/aliases` we could calculate and show in Statistics in Beacon mainpage the number of unique variants that have been found per protein/other genomic regions and their frequencies in terms of number of sequences having them.

1.1 Beacon data on mutations on coding regions - all mutation including synonymous

Figure 1 shows all unique mutations found in coding regions by mature protein. Although not the same info as in data from [CoV-GLUE data \(CVR Bioinformatics\)](#), since both synonymous and non-synonymous mutations data are included here (and unique variants are shown regardless of aminoacid change), the distribution looks pretty much the same, in agreement with what they say that selective pressure shifting non-synonymous/synonymous ratio hasn't been observed so far. Anyway, I would do the separated graph with non-synonymous variants.

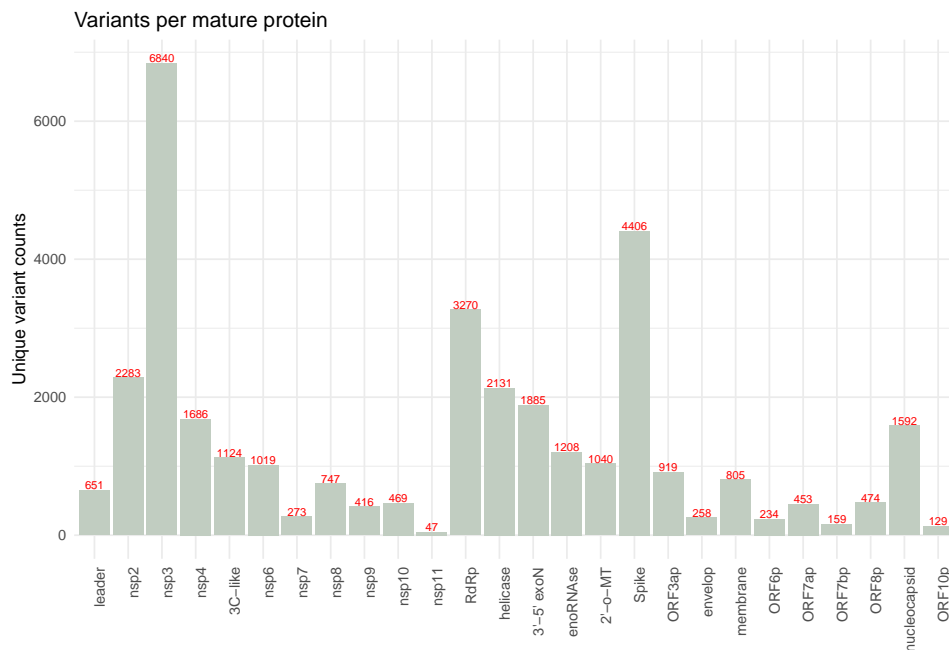


Figure 1: Unique variants in coding regions of SARS-CoV2, shown per mature proteins.

1.2 Beacon data on mutations on non-coding regions: all mutation including synonymous

Figure 2 shows non-coding regions such as UTRs and stem loops. Intergenic regions should be added as well.

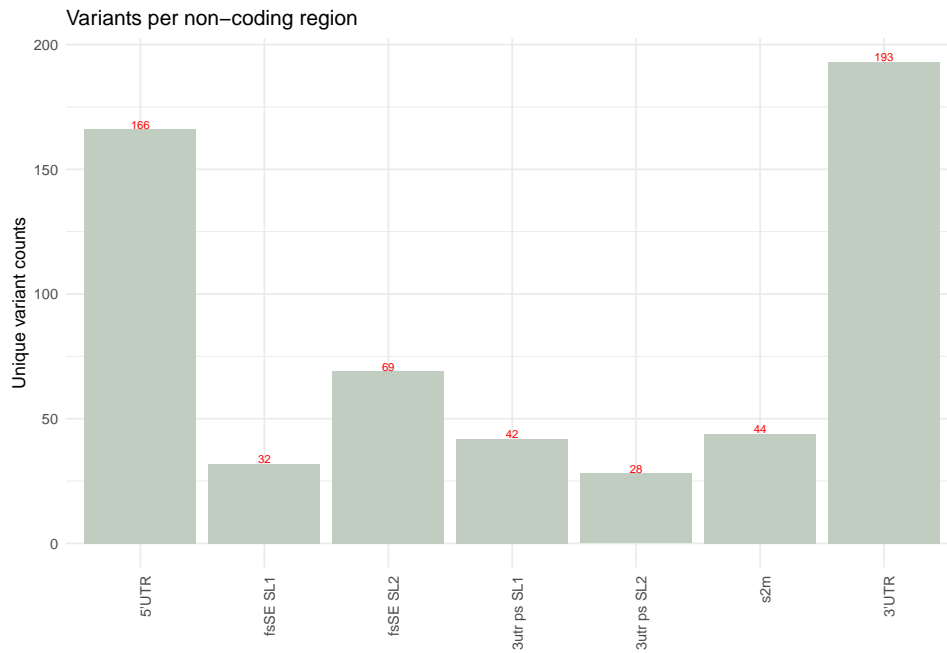


Figure 2: Unique variants in non-coding regions of SARS-CoV2. 5'UTR, stem loops (Coronavirus frameshifting stimulation element stem-loop 1, Coronavirus frameshifting stimulation element stem-loop 2, Coronavirus 3' UTR pseudoknot stem-loop 1, Coronavirus 3' UTR pseudoknot stem-loop 2, Coronavirus 3' stem-loop II-like motif (s2m), 3'UTR.

2 Other ideas for statistics on variants?

2.1 Statistics in genomic region

It would be interesting to calculate co-occurrence / mutual exclusivity patterns in mutations.

It would be interesting to calculate and show behind needle of freq per position the calculated expected frequency of variants per region under null model, so possible selection pressure on regions can be easily spotted.

- Freq variants per variantType
- dN/dS in region

2.2 Statistics per position within region

- Needle for region freq alternate per position
- Freq variants per variantType
- dN/dS per position