COG-UK report on SARS-CoV-2 Spike mutations of interest in the UK

15th January 2021

Summary

This is a **research report** that provides information on mutations and associated variants of interest in the gene encoding the SARS-CoV-2 spike protein that have been identified from sequence data generated by the COVID-19 Genomics (COG-UK) Consortium [1]. We focus on SARS-CoV-2 **spike gene mutations** of potential or known importance based on epidemiological, clinical and/or experimental observations. We also include information on several variants first identified elsewhere that are of current concern in the UK; not all of these are in the UK dataset at the time of writing. In light of the very rapid rate at which new information is emerging, the information provided here will rapidly become outdated. The report details the following: 1) a list of high frequency individual mutations, a subset of which may be important; 2) highlighted mutations of potential or known clinical and public health importance based on current evidence; 3) a list of mutations known to lead to weaker neutralisation of the virus by convalescent plasma from people who have been infected with SARS-CoV-2, and/or some monoclonal antibodies (mAbs) that may be given to patients with COVID-19; and 4) structural analyses of mutations and variants described in 2 above. We also highlight **open access websites** that provide updated information on circulating lineages, variants of special interest, and amino acid replacement, insertion and deletion counts for all SARS-CoV-2 genes [2, 3]. **Appendix 1** provides an explanation of scientific terms.

Limitations

1. This report is for information only. The clinical and public health importance of any single mutation, or combination of mutations cannot be determined from sequence data alone.

2. Putative evidence for the importance of any single mutation, or combination of mutations can be derived from computational biology and further evaluated by laboratory experiments. Genomic and laboratory evidence then need to be combined with clinical datasets that are designed to allow detection of increased transmissibility, change in disease severity, drug resistance or altered vaccine efficacy. For this reason, surveillance and risk assessment of mutations and variants is a multi-agency process involving UK Public Health Agencies who have access to detailed information on patients and populations, and other groups including NERVTAG (New and Emerging Respiratory Virus Threats Advisory Group).

3. COG-UK generates around 10,000 genomes a week, which will rise to 20,000 per week by March 2021. When COVID-19 infection rates are high, not all viruses from infected people will be sequenced and some mutations at low frequency will not be detected, but COG-UK aims to take representative samples from across the UK.

Open access websites

A report of the geographic distribution and prevalence of SARS-CoV-2 lineages in general, and three global variants of interest (lineage B.1.1.7 first identified in the UK, B.1.351 first identified in South Africa, and P1 (a descendant of lineage B.1.1.28, first identified in Brazil), can be found here. Amino acid replacement, insertion and deletion counts for all SARS-CoV-2 genes in the global GISAID database can be found here.

Genome data used in this report

The analysis described in this report is based on 142,859 UK-derived genomes sequenced by COG-UK (complete data in MRC-CLIMB database to 29th December 2020). From this we identified 1,904 different **amino acid changing** (non-synonymous) mutations in the spike (S) gene (this does not include synonymous mutations that do not lead to an alteration of an amino acid, or mutations elsewhere in the genome). Of these non-synonymous changes, 36% (n=694) mutations were only observed in a single viral sequence, while 5% (n=101) were observed in at least 100 sequences.

1. High frequency spike gene mutations

Individual amino acid replacements detected at the fifteen highest frequencies in UK genomes are shown in **Table 1**. Selection in this table is based on frequency alone and does not imply clinical importance. The proportion of new sequences detected for each mutation in the 28 days up to 29th Dec 2020 is shown, as an indication of recent prevalence. **Table S1** (at the end of the document) shows all amino acid replacements detected in complete SARS-CoV-2 genomes in the COG-UK dataset when present in 5 or more sequences. Insertions or deletions are not shown.

Table 1. Spike mutations (top 15) present in the UK at high frequency[1,2]

Amino acid replace-ment	Cumulative sequences in UK	Sequences over the last 28 days in UK (%)	Sequences over the last 28 days in England	Sequences over the last 28 days in Scotland	Sequences over the last 28 days in Wales	Sequences over the last 28 days in Northern Ireland
D614G	135532	12087 (9%)	8767	455	2862	3
A222V	56870	6987 (12%)	4208	327	2450	2
L18F	31203	4435 (14%)	2175	274	1986	0
N501Y	6082	3650 (60%)	3388	91	170	1
P681H	5681	3638 (64%)	3429	90	118	1
T716I	5646	3627 (64%)	3417	90	119	1
D1118H	5597	3630 (65%)	3420	90	119	1
A570D	5591	3624 (65%)	3415	90	118	1
S982A	5590	3625 (65%)	3416	90	118	1
N439K	3669	207 (6%)	138	3	66	0
S477N	2928	181 (6%)	93	11	77	0
A262S	2882	296 (10%)	289	6	1	0
L5F	2278	153 (7%)	111	7	35	0
E583D	1991	243 (12%)	227	0	16	0
P272L	1898	173 (9%)	166	6	1	0

[1]As of 29th December 2020; the last 28 days spans the period from 2nd to 29th Dec 2020
[2]Number of genomes is not equal to number of COVID-19 cases as data have not been deduplicated

2. Spike gene mutations of potential importance

Single spike gene mutations of potential or clinical and public health importance based on current evidence are listed in **Table 2.** The table aims to provide information on individual mutations, but this is rapidly becoming an over-simplification because mutations are increasingly arising in a range of combinations. Several variants of concern or interest are examples of this (see Additional notes). Since the analysis was completed using 142,859 UK-derived genomes, several mutations and lineages have become of increasing global interest and have been included for completeness.

The term 'escape' is used in the table as shorthand to mean weaker neutralisation of the virus by convalescent plasma from people who have been infected with SARS-COV-2, and/or some monoclonal antibodies (mAbs) that may be given to patients with COVID-19. This refers to laboratory experiments only and does not imply reduction in the efficacy of vaccines in humans. At the time of writing, we are not aware of any evidence that the mutations or combination of mutations detected to date will reduce vaccine efficacy.

Table 2. Spike S gene mutations, lineage associations and reason for interest, lineages detected in the UK

Mutation	Lineage(s) in which it has been detected	Reason for interest	Cumulative sequences in UK[1]	Sequences over 28 days[1]
D614G	B.1	Moderate effect on transmissibility.	135,249[2]	12,080
A222V	B.1.177	Fast growing lineage but no evidence to date of mutation effect.	56,376[2]	6,964
N439K	B.1.141	Increased binding affinity to hACE2 receptor. Escape from some mAbs.	57	0
	B.1.258		3,067	187
	B.1.258 (+Δ69-70)	Effect of the combination to be determined.	2,125	140
Δ69-70	B.1.1	The deletion occurs alone, and with receptor binding motif (RBM) mutations including N439K (B.1.258), and N501Y (B.1.1.7). Δ69-70 has also been detected in immunocompromised patients.	5,395	3,538
	B.1.258		2,178	141
N501Y	B.1.1.7[3] (+Δ69-70)	Associated with fast growing lineages & increased binding affinity to hACE2 receptor.	5,325[4]	3,497[4]
	B.1.1.70		530	55
	B.1.351[3] (+E484K)	This lineage also has the **E484K** mutation, which is reported to escape several mAbs and some convalescent sera.	8[4]	8[4]
Y453F	B.1.1	Associated with increased binding affinity to hACE2 receptor. Escape to some mAbs. Human/mink associated.	0	0
	B.1.1.298		0	0
E484K	B.1.351	Reported as an escape mutation for several mAbs including C121, C144, REGN10933 and REGN10934, and for some convalescent sera.	9[4]	9[4]
	P.2		7[4]	7[4]
	P.1[3]		0[4]	0[4]

[1]As of 29th December 2020; the last 28 days spans the period from 2nd to 29th Dec 2020
[2]Numbers are lower than in Table 1 because Table 2 only considers specific lineages
[3]Global variants of interest, see Additional notes
[4]Refer to https://cov-lineages.org/global_report.html for updated numbers

Additional notes supporting Table 2

D614G was not present when the SARS-CoV-2 first emerged but is now almost ubiquitous. It is found in numerous lineages and appears to be associated with a moderate effect on SARS-CoV-2 transmissibility [4-6].

A222V associated with **lineage B.1.177**. The very high prevalence and rapid European spread of this lineage does not appear to be associated with the presence of the A222V mutation based on current evidence [7]. Rather, the pattern of rapid spread appears to be associated with travel over the summer months followed by onward transmission [7].

N439K. The mutation first came to our attention in Scotland (in lineage B.1.141), but this lineage appears to have become extinct in the UK. N439K has also been detected in **lineage B.1.258** which is currently circulating in the UK. Phylodynamic analysis (analysis based on genomes) has not found the N439K mutation to have a faster rate of growth beyond that already determined for the D614G mutation (found in all variants carrying N439K) [8]. N439K enhances binding affinity to the hACE2 receptor and is associated with reduced neutralising activity of some mAbs, including one in clinical trials, and from some antibodies present in sera from a proportion of people who have recovered from infection [8]. By looking at the medical interventions required/outcomes for patients carrying either 439N or 439K, no increase in disease severity was detected [8]. There is no evidence that this mutation will allow the virus to impair the immunity triggered by vaccines. As N439K now co-occurs frequently with the spike 69-70 deletion, this importance of this combination variant will need to be assessed experimentally [9].

N501Y is associated with **lineage B.1.1.7** (VOC-202012/01) first detected in the UK [10], and with **lineage B.1.351 (501Y.V2)** first detected in South Africa [11] but since been detected in the UK (https://cov-lineages.org/global_report_B.1.351.html). See below for a more detailed description. N501Y has been identified in **lineage B.1.1.70** in Wales (https://cov-lineages.org/lineages.html). This lineage does not have the other novel mutations seen in B.1.1.7 or in B.1.351. Outside of the UK, **N501Y** is also associated with **lineage P1** (a descendant of B1.1.28).

Y453F has been identified in association with mink-human infections, but no cases have been detected in the UK. It was first identified in the Netherlands and Denmark associated with mink-human infections and in particular in Danish cluster 5 [12]. This lineage was defined as a cluster of variants carrying four mutations, the 69-70del and three amino acid replacements (Y453F, I692V and M1229I) in the spike protein, and observed to be spreading among farmed mink and a small number of people in Denmark. This lineage raised concerns because of a reported reduction of virus neutralisation activity of sera from people recovered from infection, but further studies are required [12]. The Y453F mutation, which occurs in the receptor binding motif of the spike protein has also been observed previously in SARS-CoV-2 genomes isolated from humans. This mutation has arisen independently multiple times in several countries [13]. Y453F has also been shown in laboratory studies to increase the affinity of spike protein binding to the hACE2 receptor [14].

E484K is present in **lineage B.1.351** and **lineages P.1** and **P.2.** This has been identified as an escape mutation emerging under exposure to mAbs C121 and C144 [15] and convalescent plasma [16], described to be able to reduce the neutralising activity of a combination of mAbs (REGN10989/10934) and lineage P.2 has been associated with two reinfection cases in Brazil [17, 18].

Three global variants of interest: VOC-202012/01, Lineage B.1.1.7, **Lineage B.1.351 (N501Y.V2)** and **Lineage P.1**. Much information is already available on B.1.1.7 and B.1.351, and the purpose here is to provide a brief review of salient features relating to mutations. We refer readers to published reports from PHE (here) and NERVTAG for on-going and updated information (here). A daily count and geographical distribution for each variant can be found here.

VOC-202012/01, Lineage B.1.1.7. First detected in England, this lineage has 23 mutations (6 are synonymous mutations, 14 are non-synonymous mutations and three are deletions) [10]. Two mutations have already been described that may alter SARS-CoV-2 biology: N501Y sits in the receptor binding motif of the spike protein, and increases binding affinity to the human ACE-2 receptor [14]; 69-70del has been identified in variants associated with immune escape in immunocompromised patients [19]. P681H is immediately juxtaposed to the amino acid 682-685, furin cleavage site, identified at the S1/S2 linkage site, which has been predicted to enhance systemic infection based on bioinformatic analysis [20, 21], and increased membrane fusion in laboratory experiments [10]. The relevance of this to human infection is not known. This lineage has also been indirectly associated with higher virus load in samples tested by an assay using RT-qPCR [22] and increased transmissibility [23]. It has been hypothesised (but not proven) that this lineage may have resulted from the transmission of the virus from a chronically infected individual [10]. This is based on observations that a high rate of mutations may accumulate in immunocompromised patients with chronic infections of SARS-COV-2 [19, 24, 25]. A recent study of the impact of B.1.1.7 variant mutations on antibody binding showed using bioinformatic analysis that for over 579 COVID-19 patients' sera collected between March and July 2020, mutations were predicted to affect binding in only 0.5% of samples, suggesting that these mutations would not result in immune evasion for a large majority of the investigated COVID-19 patients' sera [26].

Lineage B.1.351 (N501Y.V2). First detected in South Africa [11], this lineage has eight mutations (three of which are in the RBM, K417N, E484K and N501Y) in the spike protein gene and is associated with high numbers of infections and may be associated with increased transmissibility [11]. K417N and E484K have been shown to escape some mAbs [15, 16, 27, 28].

In order to test the effect of N501Y on vaccine efficacy, the neutralising activity of sera from vaccinated people against the virus carrying this mutation has to be assessed. Sera from 20 participants in a trial of the mRNA-based COVID-19 vaccine BNT162b2 showed equivalent neutralising titres to the N501 and Y501 viruses [29], suggesting that the efficacy of this vaccine will not be affected by the N501Y replacement. However, a limitation of this finding is that the Y501 virus used in this experiment did not include the full set of spike mutations found in the rapidly spreading strains in the UK or South Africa.

Lineage P.1. This lineage is a descendant of lineage B.1.1.28. Descendant lineages are usually suffixed with an additional sub-level of digits. However, lineage names cannot exceed three sub-levels in the lineage classification system [30]. Therefore, this descendant was not designated B.1.1.28.1 and instead given the next available top-level designation [30], which was P.1. First detected in Brazil [31] and in travellers from Brazil to Japan [32], it contains a unique constellation of lineage defining mutations (17 unique amino acid changes and 3 deletions), including mutations of putative or known biological importance such as E484K (present also in B.1.351 and at highest frequency in South Africa), K417T (possibly escaping some mAbs as observed for K417N in lineage B.1.351) and N501Y (present in lineages B.1.1.7 and B.1.351) [31]. More details of frequency and geographic distribution can be found here (https://cov-lineages.org/global_report_P.1.html). To date, lineage P.1. has not been seen in the COG-UK genome sequence data in the UK.

Another lineage has been described which is a different descendant of lineage B.1.1.28, and named P.2 according to the procedure described above. P.2 does not contain a constellation of mutations of interest, but does differ from its ancestral lineage, in that it carries the mutation E484K (this was initially called B.1.1.248 in a recent preprint [17], but has subsequently been renamed the P.2 lineage as described above). P.2 has been reported to be spreading in the Rio de Janeiro State [33] and associated with two independent reinfection cases in Brazil [17, 18]. Note, re-infection can occur without the requirement for novel mutations (e.g. if the immune response engendered by the initial infection was weak or wanes through time). Lineage P.2 does not contain the other important mutations carried by lineage P.1. We confirm that lineage P.2 has recently been detected in the UK.

The spike mutation E484K has been observed occurring independently of the other mutations seen in lineages B.1.1.7, B.1.351 and P.1, and its presence in a lineage (e.g. in lineage P.2) is not at present considered sufficient to designate it as a 'Variant of Concern'. Although some cases of the P.2 lineage have been detected in the UK, other variants with this mutation, but not related to this lineage, have also been observed. We note that no internationally agreed definition of a variant of concern has yet been agreed. The WHO has called for such a definition to be devised through consultation and consensus and we intend to contribute to that process. In the meantime, it is reasonable to expect some delay between the initial discovery of a new virus lineage and the decision as to its eventual nomenclature.

The recent emergence of variants with multiple shared mutations in spike raises concern about convergent evolution to a new phenotype, potentially associated with an increase in transmissibility or propensity for re-infection of individuals [31]. However, the effect of this mutation in relation to natural immune or immunity from vaccination has yet to be determined.

3. Spike gene mutations with a known antigenic role in laboratory experiments

Table 3 lists those mutations in the spike gene identified in the UK dataset that have been associated with weaker neutralisation of the virus by convalescent plasma from people who have been infected with SARS-COV-2, and/or some monoclonal antibodies (mAbs) that may be given to patients with COVID-19 (referred to below as 'escape'). **There is no evidence at the time of writing for this impacting on the efficacy of current vaccines or the immune response to natural SARS-CoV-2 infection**. Of note, mutations at site 484 of the spike gene have arisen repeatedly in the viral population; E484K mutation is associated with a known variant of concern (B.1.351) and a sub-lineage of B.1.1.28 in Brazil [33].

Table 3. Reported 'escape' mutations in the spike gene present in the UK

Amino acid replace-ment	Samples reported to have reduced neutralisation	References	Cumulative sequences in UK[1]	Sequences over the last 28 days in UK[1]
N439K	S2D19 S2D97 REGN10987 plasma_several	[8]	3669	207
G446V	plasma_subject_G_(day_18)	[28]	56	27
L452R	mAb_ X593 mAb_P2B-2F6	[34]	53	17
E484Q	plasma_subject_I_(day_102)	[28]	30	1
K444R	plasma_COV-NY	[15]	29	3
E484K	plasma_PT188 mAb_REGN10933 mAb_REGN10934 plasma_subject_I_(day_102) mAb_C121; mAb_C144	[16] [27] [27] [28] [15]	23	14
Y508H	mAb_H014	[34]	22	0
N440K	mAb_C135	[15]	21	2
L455F	mAb_REGN10933	[27]	19	1
A831V	mAb_B38; plasma	[34]	18	5
A475V	mAb_157 mAb_247 mAb_CB6 mAb_P2C-1F11 mAb_B38 mAb_CA1	[34]	17	3
F490S	mAb_REGN10989 mAb_REGN10934 plasma_subject_I_(day_102)	[27] [27] [28]	14	2
V483A	mAb_X593; mAb_P2B-2F6	[34]	12	7
R346K	mAb_C135	[15]	11	6
K378N	mAb_COV2-2082	[28]	6	0
K444N	plasma_COV-NY	[15]	3	0
G446S	plasma_subject_G_(day_18)	[28]	3	0
N450D	mAb_REGN10987 mAb_REGN10934	[27]	3	0
K150R	plasma_COV47	[15]	2	0
R346S	mAb_C135	[15]	2	1
K150T	plasma_COV47	[15]	1	0
V445A	mAb_REGN10987	[27]	1	0
G446A	plasma_subject_G_(day_18)	[28]	1	0
Y449H	plasma_subject_I_(day_102)	[28]	1	1
E484A	plasma_subject_E_(day_28) plasma_subject_E_(day_104) plasma_subject_F_(day_115) plasma_subject_I_(day_26) plasma_subject_I_(day_102)	[28]	1	0
E484R	plasma_subject_I_(day_102)	[28]	1	0
Q493R	mAb_C144	[15]	1	0

[1]As of 29th December 2020; the last 28 days spans the period from 2nd to 29th Dec 2020

4. Structural analyses of mutations and variants

The extent to which SARS-CoV-2 may evolve to escape immunity induced by infection or vaccination is not possible to assess from sequence data alone. Determining phenotype (how the virus could behave) from genetic data is a fundamental challenge requiring careful further experimentation. A computational biology/bioinformatics approach can be used to visualise the location of mutations on the virus protein structure. For example, human antibodies will target specific regions of the SARS-CoV-2 Spike protein and mutations proximal to these locations can be shown [34]. **Figure 1a** visualises the localisation of mutations listed in **Table 2** in a three-dimensional structure of the Spike protein (not included E4848K). A222V and the 69-70 deletion are localised relatively far from the receptor-binding site in comparison with amino acid residues 453, 439 and 501 which are in the RBD region. For each amino acid present in the Spike structure, an antibody accessibility score was calculated in **Figure 1b**. High antibody accessibility scores for 501, 439 and 70 correspond to sites that sit on the surface of the protein and that are more easily accessible to antibodies. Antibodies (Ab) are known to recognise specific regions of the Spike protein known as epitopes. Depending on the areas that antibodies target there are four classes for the RBD region and one class for the N-terminal domain (NTD) near to where 69-70 sit (**Figure 1c-d**).

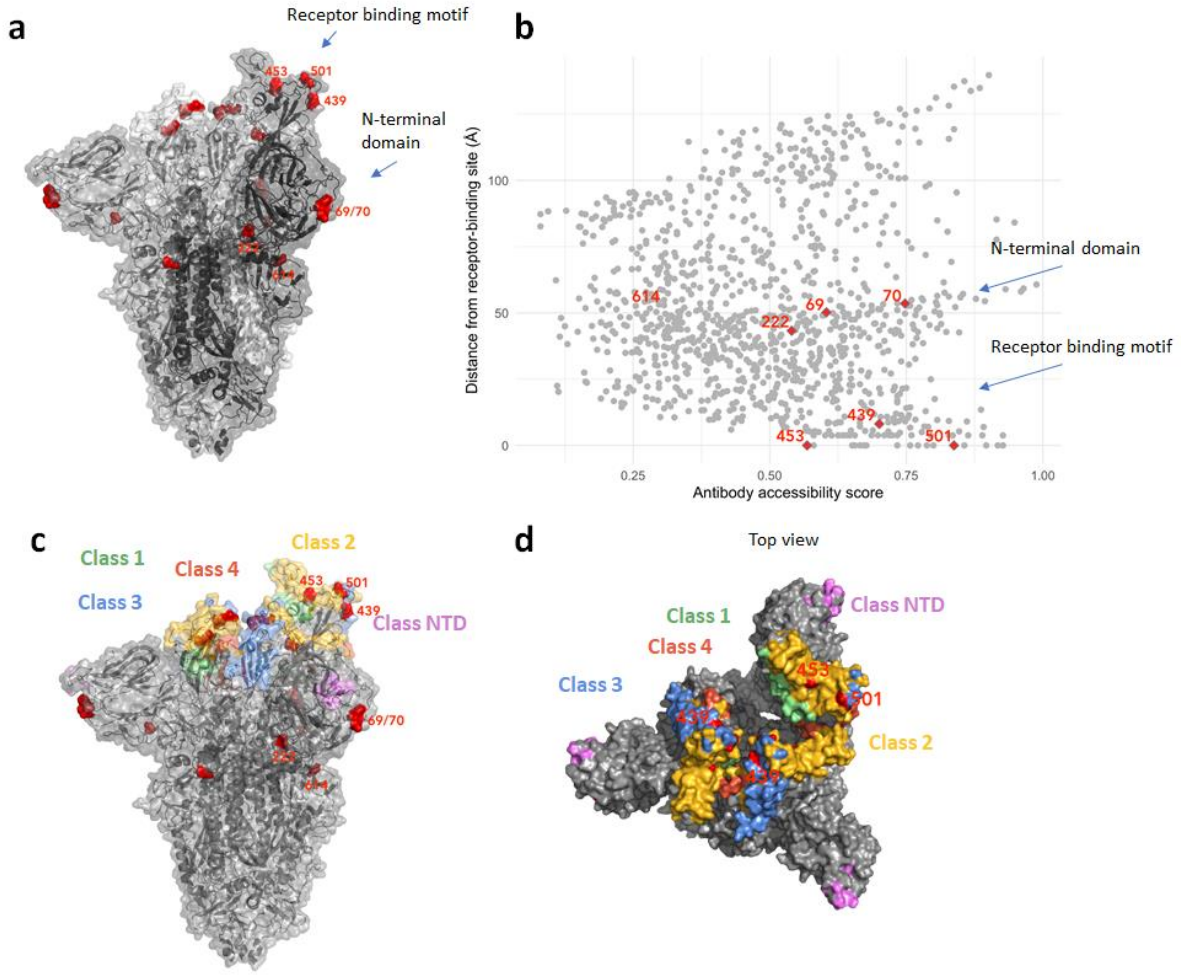

Figure 1. Localisation of mutations in the Spike structure. a) Spike heterotrimer in open conformation (PDB: 6ZGG, [17]). Locations of deleted residues His69 and Val70 and the residues involved in substitutions (A222V, N439K, Y453F, and N501Y) are highlighted in red; **b)** Each point represents a Spike protein amino

acid residue positioned according to distance from the hACE2 receptor-binding site and an antibody accessibility score. Residues associated with high interest amino acid substitution or deletions are highlighted with red diamonds. Residues belonging to the receptor-binding site defined as those with atoms within 4Å in Spike:hACE2 complex and distance to these residues based on closed conformation Spike. Antibody accessibility score represents surface accessibility and amino acid identity of target residue and weighted average of nearby residues and is scaled between minimum 0 and maximum 1, calculated across Spike in open and closed conformations; residues are positioned according to their maximum score across Spike in either open and closed conformations; **c-d)** Highlighted in colours regions target by different classes of antibodies. Residues 453, 501 and 439 are localised in the regions targeted by some classes of mAbs. 69-70 is near a region targeted by other mAbs and deletion might alter the structure of neighbouring amino acids. green = class 1: ACE2 blocking, bind open RBD only; yellow = class 2: ACE2 blocking, bind open and closed RBD; blue = class 3: non-ACE2 blocking, bind open and closed RBD; orange = class 4: non_ACE2 blocking, bind open RBD only). Epitope residues described in the NTD are coloured in magenta.

The localisation of the mutations in the Spike structure for the three global variants of interest (lineages B.1.1.7, B.1.351 (501Y.V2) and P.1) are shown in **figure 2**.

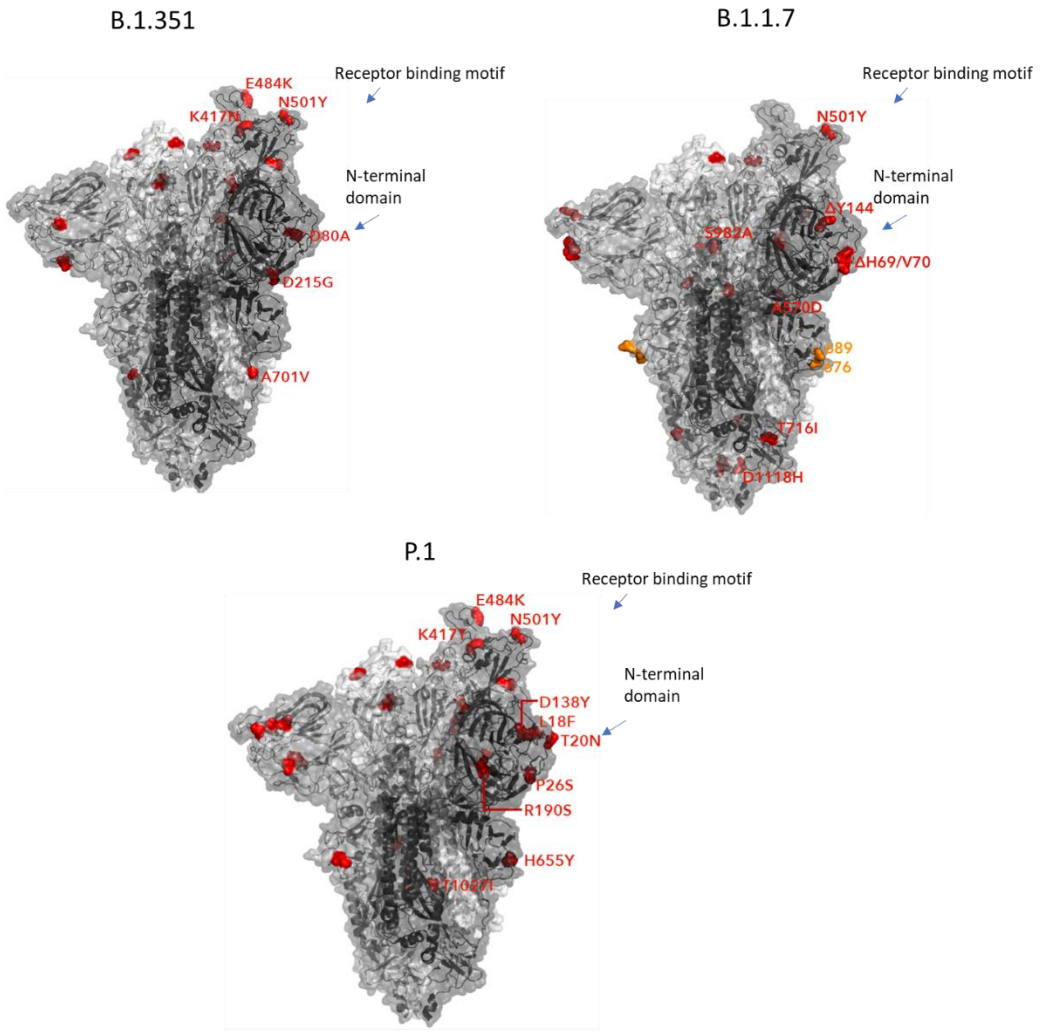

Figure 2. Localisation of mutations in the Spike structure. Spike heterotrimer in open conformation (PDB: 6ZGG, [17]). Locations of deleted residues and the residues involved in substitutions for each variant are highlighted in red. The location of an exposed loop containing the furin cleavage site and including residue 681 is absent from the structure, though modelled residues either side of this loop, 676 and 689, are coloured orange.

References

1. Covid-Genomics UK, *An integrated national scale SARS-CoV-2 genomic surveillance network.* Lancet Microbe, 2020. **1**(3): p. e99-e100.
2. Rambaut, A., et al. *SARS-CoV-2 lineages*. 2020; Available from: https://cov-lineages.org/lineages.html.
3. Robertson, D.L., et al. *CoV-GLUE*. 2020; Available from: doi: 10.20944/preprints202006.0225.v1.
4. Volz, E., et al., *Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity.* Cell, 2020.
5. Plante, J.A., et al., *Spike mutation D614G alters SARS-CoV-2 fitness.* Nature, 2020.
6. Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.* Cell, 2020. **182**(4): p. 812-827.e19.
7. Hodcroft, E.B., et al., *Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020.* medRxiv : the preprint server for health sciences, 2020: p. 2020.10.25.20219063.
8. Thomson, E.C., et al., *The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity.* bioRxiv, 2020: p. 2020.11.04.355842.
9. Kemp, S., et al., *Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/V70.* bioRxiv, 2020: p. 2020.12.14.422555.
10. Rambaut, A., et al. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*. 2020; Available from: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.
11. Tegally, H., et al., *Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa.* medRxiv, 2020: p. 2020.12.21.20248640.
12. Fonager J., et al., *Working paper on SARS-CoV-2 spike mutations arising in Danish mink, their spread to humans and neutralization data. https://files.ssi.dk/Mink-cluster-5-short-report_AFO2*. 2020.
13. van Dorp, L., et al., *Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation.* bioRxiv, 2020: p. 2020.11.16.384743.
14. Starr, T.N., et al., *Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.* Cell, 2020. **182**(5): p. 1295-1310.e20.
15. Weisblum, Y., et al., *Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants.* bioRxiv, 2020: p. 2020.07.21.214759.
16. Andreano, E., et al., *SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma.* bioRxiv, 2020: p. 2020.12.28.424451.
17. Nonaka, C.K.V., et al. *Genomic Evidence of a Sars-Cov-2 Reinfection Case With E484K Spike Mutation in Brazil*. 2021; Available from: https://www.preprints.org/manuscript/202101.0132/v1.
18. Resende, P.C., et al. *Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil, 2020*. 2021; Available from: https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584.
19. Kemp, S.A., et al., *Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion.* medRxiv, 2020: p. 2020.12.05.20241927.
20. Wang, Q., et al., *A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility.* Virol Sin, 2020. **35**(3): p. 337-339.
21. Hasan, A., et al., *A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin.* Journal of Biomolecular Structure and Dynamics, 2020: p. 1-9.

22. Kidd, M., et al., *S-variant SARS-CoV-2 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-QPCR.* medRxiv, 2020: p. 2020.12.24.20248834.
23. Volz, E., et al., *Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data.* medRxiv, 2021: p. 2020.12.30.20249034.
24. Choi, B., et al., *Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host.* New England Journal of Medicine, 2020. **383**(23): p. 2291-2293.
25. Avanzato, V.A., et al., *Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer.* Cell, 2020. **183**(7): p. 1901-1912 e9.
26. Haynes, W.A., et al., *Impact of B.1.1.7 variant mutations on antibody recognition of linear SARS-CoV-2 epitopes.* medRxiv, 2021: p. 2021.01.06.20248960.
27. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies.* Science, 2020: p. eabd0831.
28. Greaney, A.J., et al., *Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies.* bioRxiv, 2021: p. 2020.12.31.425021.
29. Xie, X., et al., *Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera.* bioRxiv, 2021: p. 2021.01.07.425740.
30. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology.* Nature Microbiology, 2020. **5**(11): p. 1403-1407.
31. Faria, N.R., et al. *Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings*. 2021; Available from: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.
32. Naveca, F., et al. *Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein*. 2021; Available from: https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585.
33. Voloch, C.M., et al., *Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil.* medRxiv, 2020: p. 2020.12.23.20248598.
34. Barnes, C.O., et al., *SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies.* Nature, 2020. **588**(7839): p. 682-687.

Useful links

https://cov-lineages.org/lineages.html
https://cov-lineages.org/global_report.html
http://cov-glue.cvr.gla.ac.uk/
https://www.gisaid.org
https://nextstrain.org/sars-cov-2/
https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201
https://app.box.com/s/3lkcbxepqixkg4mv640dpvvg978ixjtf/folder/111416414559

Appendix 1

Background

Mutations arise naturally in the SARS-CoV-2 genome as the virus replicates and circulates in the human population. As a result of this on-going process, many thousands of mutations have already arisen in the SARS-CoV-2 genome since the virus emerged in late 2019. As mutations continue to arise, novel combinations of mutations are increasingly observed. The vast majority of mutations have no apparent effect on the virus. Only a very small minority are likely to be important and change the virus in any appreciable way. This could include a change in the ability to infect/transmit between people; a change in disease severity; or a change in the way the virus interacts with the immune system (including the response generated by a vaccine). We pay most attention to mutations in the gene that encodes the Spike protein, which is associated with viral entry into cells and it is relevant in the context of immunity and vaccine efficacy.

Definitions

Mutation is used to describe a change of a nucleotide in the virus RNA genome, a subset of which results in a change in amino acid (sometimes referred to as a substitution or replacement), or a mutation can refer to a deletion or insertion event in the virus genome. By convention an amino acid change is written N501Y to denote the wildtype (N, asparagine) and replacement amino acid (Y, tyrosine) at site 501 in the amino acid sequence.

Viral variant refers to a genetically distinct virus with different mutations to other viruses. Variant can also refer to the founding virus of a cluster/lineage and used to refer collectively to the resulting variants that form the lineage.

Lineages are assigned combining genetic and, in the case of SARS-CoV-2 due to weak phylogenetic signals, also with epidemiological data. COG-UK uses the nomenclature system introduced by Rambaut et al. (2020), see https://cov-lineages.org.

VUI is used by Public Health England to indicate Variant Under Investigation.
VOC is used by Public Health England to indicate Variant of Concern. Lineage B.1.1.7 was initially named by Public Health England as VUI 202012/01 (Variant Under Investigation, year 2020, month 12, variant 01) and subsequently redesignated as VOC-202012/01 (Variant of Concern 202012/01).

Table S1. Spike mutations conferring an amino acid replacement present in the UK and detected in 5 or more sequences, as of 29th December 2020

Amino acid replace-ment	Cumulative sequences in UK	Sequences over the last 28 days in UK	Sequences over the last 28 days in England	Sequences over the last 28 days in Scotland	Sequences over the last 28 days in Wales	Sequences over the last 28 days in Northern Ireland
D614G	135532	12087	8767	455	2862	3
A222V	56870	6987	4208	327	2450	2
L18F	31203	4435	2175	274	1986	0
N501Y	6082	3650	3388	91	170	1
P681H	5681	3638	3429	90	118	1
T716I	5646	3627	3417	90	119	1
D1118H	5597	3630	3420	90	119	1
A570D	5591	3624	3415	90	118	1
S982A	5590	3625	3416	90	118	1
N439K	3669	207	138	3	66	0
S477N	2928	181	93	11	77	0
A262S	2882	296	289	6	1	0
L5F	2278	153	111	7	35	0
E583D	1991	243	227	0	16	0
P272L	1898	173	166	6	1	0
D1163Y	1731	161	140	15	6	0
G1167V	1705	167	145	16	6	0
R21I	1628	4	4	0	0	0
D80Y	1174	196	132	0	64	0
H655Y	1078	49	25	0	24	0
K1073N	893	38	30	0	8	0
L176F	886	82	52	4	26	0

D215H	886	64	62	0	2	0
A688V	833	32	29	1	2	0
S98F	809	53	42	2	9	0
T723I	760	10	9	0	1	0
S256L	746	94	65	8	21	0
N164T	692	132	132	0	0	0
S255F	596	23	22	0	1	0
Q23H	591	141	47	0	94	0
P1263L	583	6	5	0	1	0
D936Y	496	7	7	0	0	0
V622F	485	10	10	0	0	0
H146Y	483	5	5	0	0	0
S929T	443	27	27	0	0	0
P25S	417	3	3	0	0	0
G261V	415	10	5	0	5	0
L54F	402	34	30	4	0	0
C1253F	402	3	3	0	0	0
D574Y	389	3	2	1	0	0
V772I	333	15	15	0	0	0
L189F	332	15	15	0	0	0
P26S	307	16	15	0	1	0
P26L	298	5	5	0	0	0
T29I	285	0	0	0	0	0
V687L	282	11	11	0	0	0
R190M	269	33	33	0	0	0

G1219V	256	20	15	5	0	0
A879S	255	4	4	0	0	0
G181V	248	99	4	0	95	0
S939F	246	45	43	0	2	0
A1020S	227	45	33	1	11	0
G769V	219	12	10	1	1	0
Q675H	215	30	8	0	22	0
Q613H	210	53	24	0	29	0
A626S	210	3	3	0	0	0
T859N	207	8	8	0	0	0
Q677H	198	16	15	1	0	0
I1225F	191	0	0	0	0	0
V1228L	191	40	25	0	15	0
T95I	190	13	12	0	1	0
D1084Y	186	0	0	0	0	0
A845S	177	12	12	0	0	0
D198G	176	4	4	0	0	0
P1069S	173	15	15	0	0	0
C1250F	172	13	13	0	0	0
A846S	166	6	6	0	0	0
R847K	164	0	0	0	0	0
Q218H	161	14	5	9	0	0
V213L	155	0	0	0	0	0
M1229T	153	4	4	0	0	0
P812L	151	32	7	3	22	0

G842V	147	6	6	0	0	0
V1122L	145	8	0	0	8	0
A706V	143	46	46	0	0	0
M153I	140	12	12	0	0	0
R214L	140	5	2	0	3	0
S477I	136	1	1	0	0	0
E583Q	133	4	4	0	0	0
F1121L	133	0	0	0	0	0
T323I	130	17	16	0	1	0
T859I	127	17	16	0	1	0
A67V	124	19	18	1	0	0
S254F	124	34	3	1	30	0
I794L	119	33	30	0	3	0
T19K	118	2	2	0	0	0
T22I	118	31	13	18	0	0
T478I	117	5	5	0	0	0
V1068F	117	0	0	0	0	0
S12F	116	24	20	2	2	0
T20I	115	11	3	0	8	0
A522V	115	7	7	0	0	0
R102I	114	3	3	0	0	0
P330S	113	11	11	0	0	0
P479S	109	0	0	0	0	0
A1174V	109	5	5	0	0	0
S221L	104	34	31	0	3	0

T572I	103	18	6	0	12	0
E96D	102	21	21	0	0	0
N185S	101	65	0	0	65	0
S494P	101	46	45	0	1	0
P809L	99	2	1	0	0	1
T1117I	99	8	8	0	0	0
A1078S	98	10	9	0	1	0
C1235F	98	3	3	0	0	0
R21K	97	4	4	0	0	0
P1162R	97	8	8	0	0	0
G1219C	96	4	4	0	0	0
P1263T	95	33	33	0	0	0
S71P	94	1	1	0	0	0
H49Y	93	10	3	0	7	0
H245Y	93	10	8	2	0	0
P1162L	93	3	0	3	0	0
V308L	92	14	10	0	4	0
A771S	92	35	35	0	0	0
G1251V	92	1	1	0	0	0
G75V	91	3	1	2	0	0
M1229I	91	5	5	0	0	0
A520S	89	9	3	0	6	0
L841R	87	0	0	0	0	0
T51I	85	1	1	0	0	0
V1104L	85	3	3	0	0	0

T33I	84	0	0	0	0	0
G261S	83	7	7	0	0	0
A701V	82	20	11	2	7	0
K1191N	80	11	9	0	2	0
L822F	79	1	1	0	0	0
P1162S	78	16	16	0	0	0
A846V	77	9	9	0	0	0
D138Y	76	33	9	0	24	0
W258L	76	1	1	0	0	0
T549A	76	0	0	0	0	0
A1020V	75	9	8	0	1	0
T676I	74	15	0	0	15	0
A684V	74	3	0	1	2	0
D839Y	74	0	0	0	0	0
G1124V	73	4	4	0	0	0
V3G	72	2	2	0	0	0
L938F	72	2	1	0	1	0
A942S	72	16	16	0	0	0
S1252P	72	1	0	0	1	0
D138H	71	11	11	0	0	0
A522S	71	8	7	0	1	0
G639S	71	4	0	0	4	0
A647S	71	2	2	0	0	0
A1087S	69	3	2	0	1	0
S940F	68	22	1	0	21	0

A871S	67	3	3	0	0	0
L1234I	67	8	0	0	8	0
A27T	65	1	1	0	0	0
K558N	65	3	3	0	0	0
C1247F	65	2	2	0	0	0
A27S	63	2	1	1	0	0
V1176F	63	8	8	0	0	0
A67S	62	18	18	0	0	0
M177I	62	4	4	0	0	0
S640F	62	8	8	0	0	0
I818V	62	7	7	0	0	0
V143F	61	17	0	2	15	0
D253G	61	15	15	0	0	0
A845V	61	9	6	3	0	0
T1238I	61	1	0	1	0	0
Q1208H	59	0	0	0	0	0
T76I	58	7	7	0	0	0
M1237I	58	1	1	0	0	0
G446V	56	27	1	0	26	0
V826L	56	6	6	0	0	0
A27V	54	22	22	0	0	0
V367F	54	13	12	0	1	0
P812S	54	14	12	0	2	0
L452R	53	17	17	0	0	0
S698L	53	4	4	0	0	0

D1146H	53	0	0	0	0	0
T73I	52	0	0	0	0	0
S151I	52	7	0	0	7	0
G257D	52	7	7	0	0	0
P384L	52	1	1	0	0	0
V83I	50	5	5	0	0	0
A623S	49	4	3	1	0	0
I909V	49	11	11	0	0	0
D936H	49	7	6	0	1	0
E1202Q	49	2	2	0	0	0
W152R	47	24	24	0	0	0
T299I	47	4	2	1	1	0
E654Q	47	18	18	0	0	0
S691A	47	1	1	0	0	0
D796H	47	27	0	0	27	0
T19I	46	0	0	0	0	0
E156A	46	7	7	0	0	0
T307I	45	4	3	0	1	0
Q675R	45	6	5	0	1	0
A1078V	45	2	1	0	1	0
F338L	44	11	1	1	9	0
G1167F	44	7	0	0	7	0
A892S	43	2	1	1	0	0
T573I	41	2	2	0	0	0
V483F	40	0	0	0	0	0

A575S	40	5	5	0	0	0
P631S	40	2	2	0	0	0
A1026S	40	0	0	0	0	0
H69Y	39	1	0	1	0	0
F565L	39	3	3	0	0	0
C1243F	39	3	3	0	0	0
D1259Y	39	1	1	0	0	0
T547I	38	5	5	0	0	0
V615I	38	0	0	0	0	0
E1207D	38	5	0	0	5	0
H146Q	36	7	3	0	4	0
T240I	36	13	11	2	0	0
T1027I	36	4	0	4	0	0
V1264L	36	3	3	0	0	0
E96G	35	2	1	1	0	0
D796Y	35	2	2	0	0	0
I834T	35	0	0	0	0	0
S1252F	35	3	2	0	1	0
G142V	34	1	1	0	0	0
F157L	34	14	14	0	0	0
A344S	33	1	1	0	0	0
I1227M	33	0	0	0	0	0
V289I	32	0	0	0	0	0
A626V	32	0	0	0	0	0
M731I	32	5	5	0	0	0

A892V	32	10	1	8	0	1
D1084G	32	0	0	0	0	0
H207R	31	0	0	0	0	0
D215Y	31	11	4	0	7	0
A570V	31	3	3	0	0	0
E484Q	30	1	1	0	0	0
A653V	30	4	4	0	0	0
S659L	30	9	9	0	0	0
Q787K	30	3	3	0	0	0
A879V	30	1	1	0	0	0
A879T	30	0	0	0	0	0
G1099D	30	20	0	0	20	0
E309Q	29	1	1	0	0	0
K444R	29	3	3	0	0	0
I119V	28	0	0	0	0	0
D215V	28	0	0	0	0	0
L216F	28	5	4	0	1	0
L242F	28	0	0	0	0	0
G257V	28	0	0	0	0	0
T385I	28	4	4	0	0	0
R765C	28	1	1	0	0	0
K986R	28	0	0	0	0	0
L1200F	28	1	1	0	0	0
D178N	27	0	0	0	0	0
L179F	27	6	6	0	0	0

G339D	27	4	3	1	0	0
A348S	27	2	1	0	1	0
V1133F	27	0	0	0	0	0
G75S	26	13	0	0	13	0
G184V	26	1	0	1	0	0
E654D	26	0	0	0	0	0
T678I	26	3	3	0	0	0
A1070V	26	3	0	0	3	0
D80A	25	10	8	2	0	0
V615F	25	12	12	0	0	0
N679K	25	2	2	0	0	0
V70F	24	1	1	0	0	0
D111N	24	0	0	0	0	0
G142S	24	3	3	0	0	0
S247G	24	1	1	0	0	0
A647V	24	2	2	0	0	0
Q690H	24	12	12	0	0	0
A701S	24	0	0	0	0	0
P1079S	24	1	1	0	0	0
S1147L	24	1	1	0	0	0
S1242I	24	0	0	0	0	0
C15F	23	0	0	0	0	0
V90F	23	0	0	0	0	0
R102S	23	4	4	0	0	0
W152L	23	2	2	0	0	0

M153T	23	5	5	0	0	0
E484K	23	14	12	2	0	0
A672V	23	5	5	0	0	0
P809S	23	6	0	0	6	0
M1050I	23	0	0	0	0	0
H1058Y	23	5	0	0	5	0
L1203F	23	0	0	0	0	0
S13I	22	0	0	0	0	0
W64L	22	1	1	0	0	0
E281Q	22	2	2	0	0	0
T478R	22	0	0	0	0	0
Y508H	22	0	0	0	0	0
P561S	22	4	2	2	0	0
A623V	22	2	2	0	0	0
H1101Y	22	5	5	0	0	0
N440K	21	2	1	1	0	0
P681R	21	16	16	0	0	0
V1068I	21	0	0	0	0	0
E1182Q	21	1	1	0	0	0
M1237T	21	0	0	0	0	0
H207Q	20	4	4	0	0	0
R408I	20	0	0	0	0	0
D745G	20	0	0	0	0	0
Q779H	20	0	0	0	0	0
Q957L	20	0	0	0	0	0

K1205N	20	0	0	0	0	0
C1254F	20	0	0	0	0	0
S71F	19	7	7	0	0	0
G72R	19	1	1	0	0	0
P272H	19	4	4	0	0	0
P384S	19	5	5	0	0	0
L455F	19	1	0	0	1	0
T747I	19	3	3	0	0	0
T778I	19	1	1	0	0	0
S813I	19	0	0	0	0	0
A1022S	19	1	1	0	0	0
F2L	18	1	0	0	1	0
Q52L	18	0	0	0	0	0
S98P	18	0	0	0	0	0
W104F	18	10	10	0	0	0
M153V	18	0	0	0	0	0
A243S	18	3	3	0	0	0
S359N	18	3	3	0	0	0
A831V	18	5	2	0	3	0
Q836H	18	1	1	0	0	0
K854N	18	0	0	0	0	0
E1111Q	18	1	1	0	0	0
P9S	17	0	0	0	0	0
T33K	17	17	17	0	0	0
S50L	17	2	2	0	0	0

I68T	17	0	0	0	0	0
V90L	17	0	0	0	0	0
S256P	17	0	0	0	0	0
I402V	17	0	0	0	0	0
A475V	17	3	2	0	1	0
S673T	17	0	0	0	0	0
P681L	17	1	0	0	1	0
T719I	17	0	0	0	0	0
S884F	17	1	1	0	0	0
D1146Y	17	1	0	0	1	0
I1216V	17	3	3	0	0	0
G1246S	17	0	0	0	0	0
Y28H	16	2	2	0	0	0
R78M	16	0	0	0	0	0
N125S	16	1	1	0	0	0
Y144F	16	13	13	0	0	0
G181R	16	0	0	0	0	0
P251H	16	9	0	0	9	0
F306L	16	16	16	0	0	0
T376I	16	1	0	1	0	0
D427G	16	0	0	0	0	0
A570S	16	3	3	0	0	0
I670L	16	0	0	0	0	0
A893V	16	4	0	0	4	0
S937L	16	1	1	0	0	0

T1006I	16	1	0	0	1	0
D1118Y	16	1	1	0	0	0
V1164I	16	0	0	0	0	0
P9L	15	0	0	0	0	0
V62I	15	0	0	0	0	0
G72E	15	0	0	0	0	0
E132Q	15	0	0	0	0	0
L141F	15	9	8	0	1	0
K182R	15	0	0	0	0	0
G184S	15	2	2	0	0	0
T208M	15	0	0	0	0	0
L249S	15	1	1	0	0	0
P251L	15	0	0	0	0	0
Q414K	15	3	2	1	0	0
S477R	15	0	0	0	0	0
V539I	15	1	1	0	0	0
R765L	15	4	4	0	0	0
T791P	15	0	0	0	0	0
S810L	15	5	5	0	0	0
S943I	15	1	1	0	0	0
E1150D	15	0	0	0	0	0
G1167D	15	0	0	0	0	0
D1260Y	15	1	0	0	1	0
V6F	14	4	4	0	0	0
Q14H	14	4	2	2	0	0

R21T	14	3	3	0	0	0
V159I	14	0	0	0	0	0
G181A	14	4	4	0	0	0
V193L	14	0	0	0	0	0
H245R	14	0	0	0	0	0
T250N	14	0	0	0	0	0
A352S	14	2	2	0	0	0
A411S	14	2	2	0	0	0
K417N	14	10	8	2	0	0
S459Y	14	0	0	0	0	0
F490S	14	2	2	0	0	0
S689I	14	3	0	3	0	0
Q787H	14	0	0	0	0	0
A871V	14	3	3	0	0	0
A924S	14	1	1	0	0	0
G946V	14	12	12	0	0	0
G1171D	14	0	0	0	0	0
E1202D	14	0	0	0	0	0
C1240F	14	0	0	0	0	0
V11F	13	5	5	0	0	0
R78S	13	0	0	0	0	0
L84I	13	0	0	0	0	0
I100T	13	0	0	0	0	0
S112L	13	0	0	0	0	0
D215G	13	10	8	2	0	0

A243V	13	2	2	0	0	0
D287G	13	6	6	0	0	0
N481D	13	0	0	0	0	0
K529R	13	0	0	0	0	0
I670M	13	0	0	0	0	0
M740I	13	2	2	0	0	0
V915I	13	5	5	0	0	0
L922F	13	0	0	0	0	0
N1074D	13	0	0	0	0	0
E1195G	13	0	0	0	0	0
Q1201K	13	0	0	0	0	0
L8F	12	0	0	0	0	0
S12Y	12	1	1	0	0	0
V16F	12	1	1	0	0	0
P25L	12	1	1	0	0	0
G142D	12	0	0	0	0	0
T307S	12	1	1	0	0	0
V483A	12	7	7	0	0	0
T618I	12	0	0	0	0	0
A706S	12	0	0	0	0	0
T732A	12	3	3	0	0	0
A783S	12	0	0	0	0	0
F797L	12	1	1	0	0	0
D867G	12	0	0	0	0	0
I870V	12	0	0	0	0	0

Q913H	12	0	0	0	0	0
A924V	12	1	1	0	0	0
N955H	12	0	0	0	0	0
S1021Y	12	0	0	0	0	0
D1146G	12	1	1	0	0	0
V11I	11	1	1	0	0	0
Q52R	11	2	2	0	0	0
H146R	11	2	2	0	0	0
R237K	11	0	0	0	0	0
L244F	11	0	0	0	0	0
V289L	11	0	0	0	0	0
R346K	11	6	0	0	6	0
N354S	11	0	0	0	0	0
L517F	11	1	1	0	0	0
T549I	11	1	1	0	0	0
P681S	11	1	1	0	0	0
L849F	11	1	1	0	0	0
L948I	11	0	0	0	0	0
G1085R	11	1	1	0	0	0
D1153Y	11	0	0	0	0	0
D1199Y	11	0	0	0	0	0
Q52H	10	2	2	0	0	0
R102G	10	0	0	0	0	0
S162I	10	4	4	0	0	0
R190K	10	4	4	0	0	0

E224Q	10	0	0	0	0	0
G252S	10	1	1	0	0	0
W258R	10	0	0	0	0	0
A263V	10	0	0	0	0	0
A292S	10	0	0	0	0	0
G476S	10	1	0	0	1	0
T478K	10	1	1	0	0	0
N481K	10	0	0	0	0	0
T553I	10	2	2	0	0	0
S691F	10	3	3	0	0	0
V705L	10	1	1	0	0	0
T791I	10	0	0	0	0	0
S929I	10	0	0	0	0	0
A1016T	10	0	0	0	0	0
T1027S	10	0	0	0	0	0
A1056V	10	1	1	0	0	0
V1065L	10	0	0	0	0	0
A1070S	10	0	0	0	0	0
K1073E	10	2	1	0	1	0
R1091L	10	2	0	2	0	0
R1107S	10	1	1	0	0	0
F1109L	10	0	0	0	0	0
I1130M	10	2	2	0	0	0
D1199N	10	1	1	0	0	0
V1230L	10	1	1	0	0	0

V6A	9	9	9	0	0	0
R21G	9	2	2	0	0	0
G75R	9	1	1	0	0	0
K97N	9	0	0	0	0	0
I100L	9	0	0	0	0	0
S155I	9	3	0	0	3	0
Q173K	9	0	0	0	0	0
D178H	9	1	1	0	0	0
T286I	9	0	0	0	0	0
D427V	9	0	0	0	0	0
N440Y	9	1	1	0	0	0
T470A	9	0	0	0	0	0
V551I	9	5	0	0	5	0
E554G	9	0	0	0	0	0
G566S	9	2	0	2	0	0
D614N	9	1	1	0	0	0
V622L	9	0	0	0	0	0
A694F	9	0	0	0	0	0
N717S	9	3	3	0	0	0
E780D	9	0	0	0	0	0
N856S	9	0	0	0	0	0
M869I	9	0	0	0	0	0
L938I	9	3	3	0	0	0
W1102L	9	0	0	0	0	0
D1139H	9	1	0	0	1	0

E1195Q	9	0	0	0	0	0
C1236S	9	0	0	0	0	0
C1236F	9	0	0	0	0	0
C1248S	9	0	0	0	0	0
V62L	8	0	0	0	0	0
G72V	8	0	0	0	0	0
P82H	8	0	0	0	0	0
V127F	8	1	1	0	0	0
G142A	8	0	0	0	0	0
K150N	8	1	1	0	0	0
E156D	8	0	0	0	0	0
G252V	8	2	2	0	0	0
A263S	8	1	1	0	0	0
N354D	8	2	2	0	0	0
V367I	8	0	0	0	0	0
I468V	8	1	0	0	1	0
I472V	8	0	0	0	0	0
S514F	8	2	2	0	0	0
N536S	8	0	0	0	0	0
L552F	8	1	1	0	0	0
T638I	8	2	0	0	2	0
N658Y	8	0	0	0	0	0
A684T	8	0	0	0	0	0
Q690R	8	1	1	0	0	0
A694V	8	0	0	0	0	0

L699I	8	0	0	0	0	0
S711F	8	0	0	0	0	0
G769R	8	0	0	0	0	0
A829V	8	1	1	0	0	0
I834S	8	0	0	0	0	0
Y837H	8	0	0	0	0	0
G842D	8	0	0	0	0	0
P862S	8	0	0	0	0	0
L864V	8	0	0	0	0	0
F888L	8	0	0	0	0	0
D936V	8	0	0	0	0	0
G946E	8	0	0	0	0	0
D950H	8	1	1	0	0	0
V1104I	8	0	0	0	0	0
I1132V	8	2	2	0	0	0
P1143L	8	0	0	0	0	0
D1165Y	8	0	0	0	0	0
Q1201H	8	0	0	0	0	0
L1224F	8	0	0	0	0	0
C1247Y	8	0	0	0	0	0
E1258Q	8	0	0	0	0	0
L10I	7	0	0	0	0	0
T33A	7	0	0	0	0	0
A67T	7	1	1	0	0	0
V70I	7	1	1	0	0	0

K77N	7	0	0	0	0	0
V83F	7	0	0	0	0	0
P85S	7	0	0	0	0	0
V120F	7	1	1	0	0	0
L141V	7	1	1	0	0	0
A163V	7	1	1	0	0	0
Q183R	7	0	0	0	0	0
L212S	7	0	0	0	0	0
G232C	7	0	0	0	0	0
I233V	7	0	0	0	0	0
S247N	7	1	1	0	0	0
L303F	7	0	0	0	0	0
K310R	7	6	6	0	0	0
L335F	7	0	0	0	0	0
Q414R	7	0	0	0	0	0
A419S	7	0	0	0	0	0
E471Q	7	1	1	0	0	0
G482S	7	0	0	0	0	0
N501S	7	0	0	0	0	0
P521S	7	0	0	0	0	0
T547K	7	0	0	0	0	0
T618R	7	0	0	0	0	0
P621S	7	3	3	0	0	0
A688S	7	0	0	0	0	0
S689R	7	0	0	0	0	0

S698A	7	0	0	0	0	0
S708F	7	0	0	0	0	0
L754F	7	3	3	0	0	0
A845T	7	0	0	0	0	0
R847T	7	2	2	0	0	0
I850T	7	4	4	0	0	0
T1009I	7	0	0	0	0	0
Q1071L	7	0	0	0	0	0
R1091C	7	0	0	0	0	0
D1163A	7	4	4	0	0	0
D1168G	7	4	4	0	0	0
N1194H	7	2	2	0	0	0
D1199H	7	2	2	0	0	0
K1205Q	7	0	0	0	0	0
K1205R	7	0	0	0	0	0
A1226V	7	1	1	0	0	0
I1232T	7	0	0	0	0	0
M1237V	7	3	2	1	0	0
C1247G	7	0	0	0	0	0
P1263S	7	2	2	0	0	0
L1034I	7	0	0	0	0	0
S13N	6	0	0	0	0	0
F32L	6	1	1	0	0	0
V90I	6	0	0	0	0	0
A93S	6	1	1	0	0	0

L118F	6	1	1	0	0	0
E154K	6	1	1	0	0	0
V171F	6	0	0	0	0	0
N185D	6	0	0	0	0	0
N211S	6	0	0	0	0	0
I231L	6	0	0	0	0	0
R246K	6	1	1	0	0	0
P322A	6	0	0	0	0	0
P322S	6	2	2	0	0	0
S325F	6	0	0	0	0	0
K378N	6	0	0	0	0	0
I410V	6	0	0	0	0	0
D427Y	6	0	0	0	0	0
E471G	6	0	0	0	0	0
S494L	6	0	0	0	0	0
E516Q	6	0	0	0	0	0
K529M	6	0	0	0	0	0
P589S	6	0	0	0	0	0
T604Y	6	2	2	0	0	0
E661D	6	1	1	0	0	0
S680P	6	0	0	0	0	0
E702D	6	0	0	0	0	0
D737Y	6	0	0	0	0	0
R765H	6	0	0	0	0	0
Q779L	6	1	1	0	0	0

Q779K	6	0	0	0	0	0
P809T	6	0	0	0	0	0
E868G	6	0	0	0	0	0
A899S	6	0	0	0	0	0
V952F	6	0	0	0	0	0
L1063F	6	1	1	0	0	0
I1115V	6	0	0	0	0	0
V1133L	6	1	1	0	0	0
D1146E	6	0	0	0	0	0
K1149R	6	0	0	0	0	0
H1159Y	6	0	0	0	0	0
D1165G	6	0	0	0	0	0
R1185C	6	1	1	0	0	0
K1245N	6	0	0	0	0	0
D1257E	6	0	0	0	0	0
S1261F	6	0	0	0	0	0
K1266R	6	4	4	0	0	0
V1268A	6	0	0	0	0	0
Y1272H	6	1	1	0	0	0
L24V	6	0	0	0	0	0
N30T	5	0	0	0	0	0
P57S	5	0	0	0	0	0
G72W	5	0	0	0	0	0
G75D	5	2	2	0	0	0
T124I	5	0	0	0	0	0

Y144H	5	0	0	0	0	0
N148D	5	0	0	0	0	0
S151G	5	3	2	1	0	0
E154Q	5	0	0	0	0	0
Y170H	5	3	0	0	3	0
P174S	5	1	1	0	0	0
L179P	5	0	0	0	0	0
A222S	5	1	1	0	0	0
L226F	5	0	0	0	0	0
R246T	5	2	2	0	0	0
Y248S	5	1	1	0	0	0
D253N	5	0	0	0	0	0
S255P	5	0	0	0	0	0
G257S	5	1	1	0	0	0
A260V	5	0	0	0	0	0
A260S	5	1	0	0	1	0
Y266H	5	0	0	0	0	0
Y269H	5	0	0	0	0	0
T299A	5	0	0	0	0	0
I312L	5	0	0	0	0	0
F329L	5	5	5	0	0	0
V341I	5	0	0	0	0	0
R357K	5	5	5	0	0	0
V362F	5	1	1	0	0	0
V367L	5	0	0	0	0	0

S371T	5	1	1	0	0	0
F377L	5	0	0	0	0	0
K462T	5	0	0	0	0	0
P463S	5	0	0	0	0	0
P479L	5	2	2	0	0	0
H625Y	5	0	0	0	0	0
G639V	5	2	2	0	0	0
F643L	5	0	0	0	0	0
I670V	5	0	0	0	0	0
Q677R	5	2	2	0	0	0
R683W	5	0	0	0	0	0
E702Q	5	0	0	0	0	0
S704L	5	0	0	0	0	0
V705F	5	0	0	0	0	0
I714L	5	0	0	0	0	0
T732I	5	0	0	0	0	0
T778N	5	0	0	0	0	0
E780Q	5	0	0	0	0	0
G798S	5	0	0	0	0	0
P807L	5	0	0	0	0	0
D830Y	5	0	0	0	0	0
G842S	5	2	2	0	0	0
L858F	5	1	1	0	0	0
V860F	5	0	0	0	0	0
V911I	5	0	0	0	0	0

L1024F	5	0	0	0	0	0
K1045N	5	0	0	0	0	0
T1066N	5	0	0	0	0	0
T1066I	5	0	0	0	0	0
A1070T	5	0	0	0	0	0
P1112L	5	1	1	0	0	0
T1120I	5	0	0	0	0	0
D1153G	5	0	0	0	0	0
H1159R	5	0	0	0	0	0
D1184Y	5	0	0	0	0	0
N1194Y	5	0	0	0	0	0
E1202K	5	0	0	0	0	0
L1265F	5	0	0	0	0	0
G1267V	5	1	1	0	0	0
T1273I	5	0	0	0	0	0
E324K	5	0	0	0	0	0
I101M	5	5	5	0	0	0
T376N	5	0	0	0	0	0