1

2

Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies,

convalescent plasma, and plasma from vaccinated persons

5

6

Philip L. Tzou [1][¶], Kaiming Tao[1][¶], Sergei L. Kosakovsky Pond[2], Robert W. Shafer[1][*]

8

9

10

[1]Division of Infectious Diseases, Stanford University School of Medicine, Stanford, CA, U.S.A.

[2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, U.S.A.

13

14

15

* Corresponding author

E-mail: rshafer@stanford.edu (RWS)

18

19

¶These authors contributed equally to this work.

21

Abstract

As novel SARS-CoV-2 variants with different patterns of spike mutations have emerged, the susceptibility of these variants to neutralization by antibodies has been rapidly assessed. However, neutralization data are generated using different approaches and are scattered across different publications making it difficult for these data to be located and synthesized. The Stanford Coronavirus Resistance Database (CoV-RDB; https://covdb.stanford.edu) is designed to house comprehensively curated published data on the neutralizing susceptibility of SARS-CoV-2 variants and spike mutations to monoclonal antibodies (mAbs), convalescent plasma (CP), and vaccinee plasma (VP). As of October 2021, CoV-RDB contains 186 publications including 64 (34%) containing 7,328 neutralizing mAb susceptibility results, 96 (52%) containing 11,390 neutralizing CP susceptibility results, and 125 (68%) containing 20,872 neutralizing VP results. The database also records which spike mutations are selected during *in vitro* passage of SARS-CoV-2 in the presence of mAbs and which emerge in persons receiving mAbs as treatment. The CoV-RDB interface interactively displays neutralizing susceptibility data at different levels of granularity by filtering and/or aggregating query results according to one or more experimental conditions. The CoV-RDB website provides a companion sequence analysis program that outputs information about mutations present in a submitted sequence and that also assists users in determining the appropriate mutation-detection thresholds for identifying non-consensus amino acids. The most recent data underlying the CoV-RDB can be downloaded in its entirety from a Github repository in a documented machine-readable format.

Introduction

41

42   Beginning in late 2020, several SARS-CoV-2 variants sharing multiple spike mutations were

43   reported from different parts of the world. These variants have been classified according to their

44   phylogenetic lineage and component mutations. Variants that spread widely and displayed evidence for

45   being more transmissible, causing more severe disease and/or reducing neutralization by antibodies

46   generated during previous infection or vaccination have been classified as variants of concern (VOCs) by

47   the World Health Organization and U.S. Centers for Disease Control and Prevention (reviewed in [1]).

48   Variants that have spread less widely but share some of the key mutations present within VOCs have been

49   classified as variants of interest (VOIs).

50   As novel SARS-CoV-2 variants have emerged, numerous investigations assessed the susceptibility

51   of individual and combination spike mutations to neutralization by monoclonal antibodies (mAbs),

52   convalescent plasma (CP), and vaccinee plasma (VP). The susceptibility of SARS-CoV-2 variants to mAbs

53   is obviously relevant for preventing and treating SARS-CoV-2 infections with mAb regimens. The

54   susceptibility of SARS-CoV-2 variants to CP provides insight into the likelihood that a variant will infect

55   and cause illness in a person previously infected with and recovered from a different variant. The

56   susceptibility of SARS-CoV-2 variants to VP provides insight into the risk that a variant will infect and

57   cause illness in a previously vaccinated person.

58   The SARS-CoV-2 spike protein is a 1,273-amino acid trimeric glycoprotein responsible for entry

59   into host cells. Each spike monomer has a largely exposed S1 attachment domain (residues 1–686) and a

60   partially buried S2 fusion domain (residues 687–1,273) [2,3]. Part of S1, called the receptor-binding domain

61   (RBD; residues 306–534) binds to the human angiotensin-converting enzyme 2 (ACE2) receptor [4,5].

62   Approximately 20 RBD residues bind the ACE2 receptor. The part of the RBD containing these residues

63   (438-506) is referred to as the receptor-binding motif (RBM), whereas the remainder of the RBD is called

64   the RBD core. The SARS-CoV-2 RBD is the main target of neutralizing antibodies [6,7]. Like the RBD,

65   much of the S1 amino-terminal domain (NTD) is also exposed on the spike trimer surface and is targeted

66   by neutralizing antibodies.

67   The Stanford Coronavirus Resistance Database (CoV-RDB) provides open access to continuously

68   curated neutralizing susceptibility data of SARS-CoV-2 mutations and variants to mAbs, CP, and VP. The

69   database includes specific data on which spike mutations have been selected during both i*n vitro* passage

70   of SARS-CoV-2 in the presence of mAbs and *in vivo* emergence in persons receiving mAbs for treatment.

71   The CoV-RDB website (https://covdb.stanford.edu) enables users to query its database and the associated

72   Github repository enables users to download the entire database. The CoV-RDB website also contains a

73   companion sequence analysis program that annotates SARS-CoV-2 user-submitted sequences using CoV-

74   RDB data.

75

Methods and results

76

77   The CoV-RDB data model is made up of four major entities: published references, viral

78   mutations and variants, antibodies (including mAbs, CP, and VP), and experimental results. As of

79   October 2021, the database contains 186 publications: 18 (10%) with 380 results from in vitro mAb

80   selection experiments, 64 (34%) with 7,328 neutralizing mAb susceptibility results, 96 (52%) containing

81   11,390 with CP susceptibility results, and 125 (68%) with 20,872 neutralizing VP results. As publications

82   often contain more than one type of experiment, the sum of the percentages is greater than 100%.

83   The   complete   database   schema   is   available   at   https://github.com/hivdb/covid-

84   drdb/blob/master/schema.dbml.

85

Curation of published references

86

87   Published references included in the CoV-RDB are obtained weekly from three literature sources:

88   i) PubMed using the search term "SARS-CoV-2"; ii) BioRxiv/MedRxiv COVID-19 SARS-CoV-2 preprint

89   servers; and iii) the Research Square SARS-CoV-2/COVID-19 preprint server. Publication titles and

90   abstracts are reviewed manually to identify studies containing data on SARS-CoV-2 spike mutations and

91   humoral immunity. Studies that pass initial review are downloaded to a Zotero reference database and full

92   texts are reviewed to extract specific data on the selection of spike mutations *in vitro* and *in vivo*, and

93   neutralizing susceptibility data for individual spike mutations or combinations of spike mutations, such as

94   those present in SARS-CoV-2 variants to mAbs, CP, and VP. Curated studies initially published as preprints

95   are re-evaluated following peer-review publication.

96   For each publication, data are first entered into linked comma-separated files (CSVs) contained in

97   a Github repository (https://github.com/hivdb/covid-drdb-payload). Data are then imported into a Postgres

98   database where they are validated for completeness and consistency. The data in the Postgres database are

99   then exported as a single SQLite database file that is both available for download under the CC BY-SA 4.0

100   open-source license and used as the source for regularly updated datasets and queries on the CoV-RDB

101   website. The entire workflow is summarized in S1 Fig.

102

Virus variants and mutations

103

104   In the CoV-RDB data model, the virus entity represents individual spike mutations, combinations

105   of spike mutations, and virus variants for which the full set of spike and genomic mutations is known

106   including VOCs and VOIs. Published *in vitro* neutralization experiments have been performed using (i)

107   pseudotyped viruses (vesicular stomatitis viruses [VSV] and lentiviruses) containing a SARS-CoV-2 spike

108   protein with specific mutations; (ii) chimeric SARS-CoV-2 viruses in a VSV genomic backbone containing

109   a spike protein with specific mutations; (iii) full-length cloned SARS-CoV-2 recombinant viruses generated

110   using multiple plasmids, or bacterial or yeast artificial chromosomes; and (iv) primary SARS-CoV-2

111   isolates [1]. For pseudotyped, chimeric, and recombinant viruses, the virus can be characterized entirely by

112   its spike mutations. For the primary SARS-CoV-2 isolates, the virus is also characterized by mutations in

113   other viral proteins that may influence viral replication kinetics but not neutralization susceptibility. Most

114   *in vitro* selection experiments have been performed using chimeric VSVs containing SARS-CoV-2 spike

115   proteins as these viral constructs can undergo multiple rounds of replication in cell culture.

116   Table 1 lists the VOCs, VOIs, and other full-genomic SARS-CoV-2 variants for which data are

117   available in CoV-RDB. Approximately 85% of results on variants are for VOCs, 12% are for VOIs, and 3%

118   are for other variants containing one or more RBD mutations. As SARS-CoV-2 variants are continually

119   evolving and variant definitions are regularly updated, individual VOCs and VOIs often differ slightly in

120   the mutations that they contain. S1 Table displays the variability in these patterns for each of the VOCs and

121   VOIs. For example, the alpha, beta, gamma, and delta variants in CoV-RDB contain 11, 20, 7, and 17

122   unique patterns of spike mutations, respectively, all closely related to an archetypal consensus variant.

123

Table 1. Variants of Concern, Variants of Interest, and Other Variants for which Neutralization Data are Available in CoV-RDB

Virus types	References	Results
Variants of concern		
Alpha	106	7439
Beta	114	7823
Gamma	61	3247
Delta	40	2612
Variants of interest		
Epsilon	22	926
Eta	9	216
Iota	14	554
Kappa	23	823
Lambda	8	324
Mu	4	32
Other variants containing >=1 RBD mutation (n=17)		
B.1.463 [F384L]	1	220
B.1.1.298 [Y453F]	6	192
B.1.617.3 [L452R, E484Q]	2	100
B.1.1.241 [S477N]	1	64
B.1/E484K	1	32
R.1 [E484K]	1	32
B.1.1.141 [T385I]	1	17
B.1.1.317 [S477N, A522S]	1	14
B.1.1.298 [Y453F]	6	192
B.1.1.318 [E484K]	1	25
B.1.618 [E484K]	1	20
A.23.1 [V367F, E484K]	3	27
B.1.1.519 [T478K]	2	12
A.30 [R346K, T478R, E484K]	1	8
A.27/A222V [L452R, N501Y]	1	6
B.1.619 [N440K, E484K]	1	1
R.2 [E484K]	1	1
Footnote: The most common ancestral variants lacking immune escape mutations used as controls include A variants (n=3899), B variants (Wuhan consensus; n=17136), B.1 variants (Wuhan consensus + D614G; n=16387).		

124

125

126   Table 2 lists the most common studied individual mutations within the spike RBD, NTD, C-

127   terminal domain (CTD), and S2 domain. These mutations have generally been studied within an ancestral

128   virus backbone. CoV-RDB contains experimental data on 296 different individual spike mutations at 159

129   positions including 197 RBD mutations at 76 positions, 55 NTD mutations at 43 positions, 32 S2 mutations

130   at 29 positions, and 12 CTD mutations at 11 positions. CoV-RDB also contains results on 89 combinations

131   of spike mutations in addition to VOCs and VOIs. These generally consist of sets of two or three mutations

132   that are present within a VOC or VOI. Several of the most common mutation combinations are listed in the

133   footnote of Table 2.

134

Table 2. Individual SARS-CoV-2 Spike Mutations for which Data are Available in CoV-RDB

Mutations	References	Results
Receptor-binding domain (RBD)		
E484K	37	687
N501Y	26	596
K417N	19	468
L452R	16	242
N439K	18	236
S477N	14	198
Y453F	13	168
F490S	14	149
E484Q	8	91
Other mutations	35	3713
N-terminal domain (NTD)		
Δ69/70	10	268
Δ242	6	172
L18F	6	133
D215G	4	126
S247R	1	33
R246I	4	116
D80A	3	114
Δ144	4	113
Other mutations	15	618
C-terminal domain of S1 (CTD)		
P681H	6	110
A570D	4	88
P681R	1	20
H655Y	1	12
Q675H	1	12
Other mutations	5	110
S2		
A701V	3	113
S982A	4	93
D1118H	4	88
T716I	4	87
D936Y	3	60
Other mutations	12	569
Footnote: The most commonly studied mutation combinations that are not VOCs or VOIs included K417N+E484K+N501Y (n=19 studies; 569 results), E484K+N501Y (n=5 studies; 189 results), Δ69/70+N501Y (n=6 studies; 135 results), Δ69/70+Y453F (n=2 studies; 74 results), L452R+E484Q+P681R (n=3 studies; 55 results) K417N+N501Y (n=2 studies; 75 results), K417N+E484K (n=2 studies; 75 results), and L452R+E484Q (n=4 studies; 55 results)		

135

136

Antibodies (mAbs, CP, and VP)

mAbs

In CoV-RDB, mAbs are characterized according to their stage of clinical development; spike target; and specific epitope, i.e., the list of amino acids within 4.5 angstroms of the mAb paratope, according to structural data obtained from the Protein DataBank (PDB). Table 3 lists those mAbs with FDA emergency use authorizations (EUAs) or in advanced clinical development along with their spike targets and numbers of experimental results including *in vitro* selection and neutralization susceptibility data. In addition to the mAbs in Table 3, CoV-RDB contains 340 results for five mAbs in phase I/II trials; and 4975 results for 295 mAbs that are not currently in clinical development. For 54 of these 295, additional mAbs, structural data are available in the PDB. Table 4 summarizes neutralization susceptibilities for the six most studied variants and the six most studied individual spike mutations to those mAbs that either have EUAs or are in advanced clinical development.

Table 3. Monoclonal Antibodies (mAbs) with Emergency Use Authorization (EUA) or in Advanced Clinical Trials

mAb	Company	Stage	PDB	Target[1]	Publication[2]	Results[2]
Bamlanivimab	Lilly	EUA	7KMG	RBM II	19	144
Etesevimab	Lilly	EUA	7C01	RBM I	22	279
Casirivimab	Regeneron	EUA	6XDG	RBM I	30	336
Imdevimab	Regeneron	EUA	6XDG	RBD core	30	335
Sotrovimab	Vir / GlaxoSmithKline	EUA	6WPS	RBD core	16	205
Tixagevimab	Vanderbilt / AstraZeneca	Completed phase III	N/A	RBM I	10	48
Cilgavimab	Vanderbilt / AstraZeneca	Completed phase III	N/A	N/A	10	48
BRII-196	BRII Biosciences	Completed phase III	7CDI	RBM I	4	146
BRII-198	BRII Biosciences	Completed phase III	N/A	RBD core	2	47
Regdanivimab	Celltrion	Phase III	7CM4	RBM I	4	20
C135	Rockefeller / Roche	Phase III	7K8Z	RBD core	4	105
C144	Rockefeller / Roche	Phase III	7K90	RBM II	4	138
Footnote:[1]RBM (receptor binding motif) is the part of the receptor binding domain (RBD) that contains the spike ACE2 binding residues. RBM class 1 mAbs bind epitopes dominated by ACE2-binding residues and as a result bind solely when the RBD is in the up/open position. RBM class 2 mAbs have a smaller ACE2-binding footprint and can often bind the RBD in down/closed position. RBD core mAbs target a surface accessible part of the RBD that is separate from the RBM. [2]The publications and results include those describing both *in vitro* selection and neutralization experiments.						

150

151

Table 4. Median Fold Reduced Neutralization Susceptibility of 6 Variants and 6 Spike Mutations to 12 mAbs in Advanced Clinical Development[1]

	BAM	ETE	CAS	IMD	CIL	TIX	SOT	REG	BRII-196	BRII-198	C135	C144
Alpha	1_{10}	10_8	1_{14}	0.7_{14}	0.7_4	1.6_4	$3^*{}_{12}$	1.4	0.5_3	0.2_3	0.9_2	-
Beta	$>100_{12}$	$>100_{10}$	73_{18}	0.7_{17}	0.7_4	4.9_4	1_{11}	27_2	0.6_4	6_3	0.9_2	-
Gamma	$>100_8$	$>100_8$	$>100_{12}$	0.4_{11}	0.5_3	3.7_3	1.3_9	81_2	0.3	0.7	-	-
Delta	$>100_8$	0.6_8	0.8_7	1.4_7	2.9	0.6	1_3	54_2	-	-	-	-
Iota	$>100_5$	1.4_5	11_4	1.2_4	0.9_2	8.1_2	0.8_4	-	-	-	-	-
Epsilon	$>100_4$	1_4	1.3_2	1.7_2	-	-	0.7_3	43_4	-	-	-	-
N501Y	1_3	2.8_6	1_8	0.8_8	1_3	1.1_3	$1.6^*{}_5$	5.5	1_3	1.8_2	1.1_4	1.3_4
E484K	$>100_4$	2.9_7	13_{13}	1_{13}	1.5_4	4.6_4	0.5_6	8.7	1.4_3	2.4_2	0.5_4	$>100_4$
K417N	0.3_2	$>100_7$	10_8	0.6_8	0.6_3	0.3_3	0.6_4	-	1.8_3	0.3_2	0.3_3	0.8_3
L452R	$>100_2$	1_5	1_5	2_6	-	-	0.6	35	1.4	-	-	-
F490S	$>100_2$	1.1_2	0.8_3	1.2_3	-	-	0.8_2	-	-	-	0.7_2	3.3_3
S494P	86_2	0.6_2	3.5_4	1.2_3	-	-	2_2	-	0.7	1.6	0.5_2	61_3

Footnote: [1]The subscript indicates the number of results from which the median fold-reduction in susceptibility was determined. Abbreviations: Bamlanivimab (BAM); Etesevimab (ETE); Casirivimab (CAS); Imdevimab (IMD); sotrovimab (SOT); Cilgavimab (CIL); Tixagevimab (TIX); Regdanivimab (REG).

152

153

154   Fig 1 shows the relationship between mAb epitopes, mutations selected during *in vitro* passage,

155   and mutations that reduce mAb binding and/or susceptibility. S2 Fig shows that there is a strong correlation

156   between RBD binding in a high throughput deep mutational scanning (DMS) assay and neutralization in

157   that an escape fraction of >0.1 in the DMS assay is with few exceptions associated with a >10-fold reduction

158   in neutralization in cell culture [8].

159

160 **Fig 1. Monoclonal antibodies (mAbs) with EUAs or in advanced clinical development: receptor**

161 **binding domain (RBD) epitopes and immune escape positions.** For each mAb, the top of the RBD and

162 two side views are depicted using coordinates from PDB 6M0J. ACE2 binding residues are shown in red;

163 the mAb epitope defined as those residues within 4.5 angstroms of the RBD is shown in dark blue; and

164 ACE2 binding residues within the mAb epitope are shown in purple. Those positions containing mutations

165 that were either selected by the mAb *in vitro* ("SEL"), reduced binding in a deep mutational scanning assay

166 ("DMS"), and/or reduced *in vitro* neutralizing susceptibility by a median of ≥4-fold in CoV-RDB (drug

167 resistance; "DR") are also indicated. The mAb epitopes for BAM (bamlanivimab), ETE (etesevimab), CAS

168 (casirivimab), IMD (imdevimab), SOT (sotrovirmab), CIL (cilgavimab) and TIX (tixagevimab) were

169 determined from their PDB structures.

170

171

172 **CP**

173   In CoV-RDB, CP are characterized by the sequence of the infecting variant, severity of illness, and

174 time since infection. Table 5 lists the numbers of experimental results in CoV-RDB according to these

175 characteristics and the SARS-CoV-2 variant or mutation(s) tested for neutralization. Overall, 11,390

176 neutralization experiments were performed using CP samples (96 studies) including 84 studies that

177 provided data for individual samples and 12 studies that provided only aggregate data.

178

Table 5. Numbers of Convalescent Plasma (CP) Neutralization Experiments in CoV-RDB According to the Infecting Virus, Time Since Infection, and SARS-CoV-2 Variant Tested

Characteristic	References	Results
Infecting virus		
Ancestral (A, B, B.1, others)	86	8778
Alpha	10	825
Beta	7	214
Gamma	2	38
Delta	1	40
Time since infection		
1 month	74	6661
2-3 months	35	2743
4-6 months	23	1397
>6 months	14	589
Variant / mutations undergoing neutralization		
Alpha	55	2294
Beta	58	2416
Gamma	26	964
Delta	13	444
VOIs (kappa, iota, epsilon, lambda)	9	161
Other variants	19	655
Individual mutations and mutation combinations	28	3795

179

180

181   The time since infection was available for all CP sample results including 10,801 samples obtained

182   within six months of infection and 589 samples obtained beyond six months. The severity of illness was

183   described for 3,559 (31%) CP samples with 2,133 samples from persons with mild-to-moderate disease and

184   1,426 samples from persons with severe disease. Although the sequence of the infecting virus was rarely

185   known, the vast majority were obtained prior to the emergence of VOCs and VOIs. Nonetheless, 7%, 2%,

186   1%, and 1% were obtained from persons known to be infected with the Alpha, Beta, Gamma, and Delta

187   variants, respectively.

188

189   **VP**

190   In CoV-RDB, VP are characterized according to the vaccine received, number of vaccinations, time

191   since vaccination, and whether the VP was obtained from a person who was also previously infected with

192   SARS-CoV-2. Table 6 lists the numbers of experimental results according to each of the above

193   characteristics and the variant or mutation(s) tested for neutralization. 70% of VP were obtained from

194   persons receiving one of the two widely used mRNA vaccines (BNT162b and mRNA-1273) while about

195   20% were obtained from persons receiving the Coronovac, AZD1222, or Ad26.COV2s vaccines.

196   Approximately 80%, 18%, and 2% of VP samples were obtained within 1 month, 2-6 months, and >6

197   months after vaccination, respectively. Nearly 5% of samples were obtained from persons with confirmed

198   infection prior to vaccination.

199

able 6. Numbers of Vaccinee Plasma (VP) Neutralization Experiments in CoV-RDB According to the Vaccine, History of Previous Infection, Time Since Infection, and SARS-CoV-2 Variant Tested		
Characteristic	References	Results
Vaccine		
BNT162b2	82	10608
mRNA-1273	31	4048
CoronaVac	8	1607
AZD1222	15	1527
Ad26.COV2.S	8	828
BBV152	5	449
Sputnik V	3	408
AZD1222 + BNT162b2	3	258
MVC-COV1901	1	342
BBV154	1	120
BBIBP-CorV	2	111
NVX-CoV2373	2	148
AZD1222 + AZD2816	1	92
ZF2001	2	56
mRNA-1273 + mRNA-1273.351	1	40
AZD2816	1	20
Previously SARS-CoV-2 infected		
Yes	14	837
No	125	20035
Time since vaccination		
1 month	113	16789
2-3 months	21	3072
4-6 months	10	662
> 6 months	8	349
Variant / mutations undergoing neutralization		
Alpha	73	4743
Beta	79	5090
Gamma	40	2114
Delta	32	2063
VOIs (kappa, iota, epsilon, lambda)	21	978
Other variants	38	1898
Individual mutations and mutation combinations	23	2327
Number of vaccinations		
1	34	4438
2	118	15979
3	6	455
Footnote: The numbers of references and results obtained from persons receiving uncommonly studied vaccines or vaccine combinations are not shown.		

200

201

202   Fig 2 shows the distribution of fold-reductions in neutralizing susceptibilities and absolute

203   neutralizing antibody titers against each of the VOCs for previously uninfected persons completing the

204   recommended course of vaccination for eight vaccines. A similar fold-reduction in neutralizing

205   susceptibility against a given variant is likely to be more consequential for those vaccines that elicit lower

206   titers of neutralizing antibodies. The figure shows that regardless of the vaccine, the fold reduction in

207   neutralizing susceptibility was consistently greater for the Beta variant and that VP from persons receiving

208   one of the two mRNA vaccines was usually more likely to retain neutralizing activity against each of the

209   VOCs.

210

211   **Fig 2. Distribution of fold-reduction in susceptibilities (A) and of absolute neutralizing titers (B) of**

212   **vaccinee plasma (VP) associated with eight vaccines to the four variants of concern (VOC): Alpha,**

213   **Beta, Gamma, and Delta.** The X axes indicate the Greek letter associated with each VOC. The Y axes

214   indicate the number of neutralizing assays. The numbers above the stacked bars indicate the number of

215   studies reporting the experimental results. The figure includes VP obtained solely from previously

216   uninfected persons receiving a complete vaccination schedule.

217

218

219 CoV-RDB Website

220   The CoV-RDB website contains four main features: (i) searchable tables containing SARS-CoV-2

221 variants, mAbs, vaccines, and references; (ii) regularly updated data summaries such as the data shown in

222 Table 4 for mAbs and in Fig 2 for VP; (iii) user-defined queries; and (iv) a sequence analysis program. The

223 user-defined queries and sequence analysis program are described here because they are interactive and

224 contain multiple user options.

225

226 User-Defined queries

227 Query interface

228   The query interface allows users to search the database using one or more of the following three

229 criteria: (i) published reference; (ii) antibody preparation (mAb, CP, or VP); and (iii) SARS-CoV-2 variant

230 or spike mutation(s). If the "References" dropdown option is selected, then all the data associated with that

231 reference is displayed in separate tables containing neutralization susceptibility data for mAbs, CP, and VP

232 and/or *in vitro* selection data for mAbs. If the "Plasma / mAbs" dropdown option is selected, users must

233 select either a specific mAb, CP from a person infected with a specific variant, or VP from persons who

234 received a specific vaccine. If the "Variants / Mutations" dropdown option is selected, users must select a

235 particular VOC, VOI, other variant, individual spike mutation, or combination of spike mutations. Selecting

236 items from multiple dropdown boxes restricts the output to the data specified by the combination of

237 dropdown items. Fig 3 shows the output returned by selecting E484K from the "Variants / Mutations"

238 dropdown. Fig 4 shows the output returned by selecting BNT162b from the "Plasma / mAbs" dropdown

239 and the Delta variant from the "Variant / Mutations" dropdown.

240

241

242   **Fig 3. Query interface, sample query, and outline of query results for an example query for the SARS-**

243   **CoV-2 RBD mutation E484K.** The query interface containing three dropdown boxes is shown at the upper

244   left (A). E484K is selected from the "Variants / Mutations" dropdown box. The upper right summarizes the

245   data returned by the query, which in this case includes 244 mAb neutralizing susceptibility results from 28

246   publications, 232 convalescent plasma (CP) results from 15 publications, and 214 vaccinee plasma (VP)

247   results from 11 publications (B). The summary distinguishes between results for which only aggregate

248   (mean or median) data are provided and results for which individual data are provided. The sections below

249   show the headers and first row of the tables containing mAb (C), VP (D), and CP (E) susceptibility results.

250   The complete contents of these tables can be found on the web (https://covdb.stanford.edu/search-

251   drdb/?host=human&mutations=S%3A484K).

252

253 **Fig 4. Query interface, sample query, and outline of query results for an example query for the**

254 **susceptibility of the SARS-CoV-2 Delta variant to the vaccinee plasma (VP) of persons who received**

255 **the BNT162b vaccine.** The query interface containing three dropdown boxes is shown at the upper left

256 (A). BNT162b is selected from the "Plasma / mAbs" dropdown box, and the Delta variant is selected from

257 the "Variants / Mutations" dropdown box. The upper right summarizes the data returned by the query,

258 which in this case includes 875 results from 12 publications (B). The summary distinguishes between results

259 for which only aggregate (mean or median) data are provided and results for which individual data are

260 provided. The section below the header shows the header and first few rows of the table entitled "Vaccinee

261 Plasma Susceptibility Data". The complete contents of this tables can be found on the web

262 (https://covdb.stanford.edu/search-drdb/?host=human&vaccine=BNT162b2&variant=Delta).

263

Query results

264

265 The mAb, CP, and VP susceptibility query result tables contain between 10 to 13 column headers

266 (Fig 3C-3E). The mAb table contains column headers indicating the reference and location of the data

267 within each reference (e.g., figure or table); assay type (e.g., pseudotyped virus); mAb tested; variant tested;

268 IC_{50} in ng/ml; and fold-reduced susceptibility compared with a control virus (that is also present in the table;

269 Fig 3C).

270 The VP table contains column headers that indicate the reference and location of the data within

271 the reference, type of assay used, vaccine received, number of immunizations, number of months since

272 immunization, whether the sample was obtained from a vaccinated person with confirmed prior infection,

273 variant tested, geometric mean neutralizing titer, and median fold reduction in titer compared with the

274 control virus (Fig 3D and Fig 4C).

275 The CP table contains column headers that indicate the reference and location of the data within

276 the reference, assay type, lineage of the virus that infected the person from whom the CP was obtained,

277 number of months between infection and the plasma sample, variant tested, geometric mean neutralizing

278 titer; and median fold reduction in titer compared with the control virus (Fig 3E).

279 Each table also contains two additional columns: "# Results" and "Data Availability". The number

280 of results indicates the number of neutralizing experiments. The data availability contains a "√" if individual

281 data are available or an "X" if data are only available in aggregate (i.e., as a geometric mean titer or a

282 median fold reduction in susceptibility). Clicking on the spreadsheet icon in "Data Availability" column

283 copies the contents of that study to the user's clipboard. The "Download CSV" tab at the top of each table

284 allows users to download all rows for analysis.

285

Interacting with the query results

286

287 Each table can be considered to have multiple dimensions represented by the table's column

288 headers. For example, Fig 5A indicates the table dimensions for the query output with BNT162b selected

289   from the "Plasma Abs / mAb" dropdown and the Delta variant selected from the "Variants / Mutations"

290   dropdown. The table contains 22 rows (i.e., experimental conditions) summarizing 885 results obtained

291   from 13 references. The table dimension check boxes enable users to aggregate the data in the table by

292   deselecting one or more dimensions. For example, if the user deselects the "Reference", "Assay", and

293   "Control" variant dimensions, the data are summarized with six rows that provide the median fold-reduction

294   in susceptibility to BNT162b-associated VP for the Delta variant according to the number of vaccinations

295   received, the time since vaccination, and whether the VP was obtained from a previously vaccinated person

296   (Fig 5B).

297

298   **Fig 5. Aggregating the query results for the example shown in Fig 4.** The top part of the figure shows

299   the first three of 22 experimental conditions defined by the column headers in which BNT162b-associated

300   VP was tested for activity against the Delta variant (A). The results for each experimental condition are

301   shown in different rows each containing the geometric mean of the neutralizing antibody titer and the

302   median fold-reduction in titer compared with a control virus. The bottom part of the figure shows that the

303   22 rows can be displayed using 6 rows by aggregating those results obtained from different references using

304   different assays or different control viruses by deselecting the "Reference", "Assay", and "Control" variant

305   (checkboxes within red ovals). The neutralization data are now aggregated according to the number of

306   BNT162b immunizations, time since immunization, and whether the VP was obtained from a person who

307   had been infected prior to vaccination (shaded area superimposed on the first four columns). The number

308   of experimental results in each row in Fig 5B is increased because each row may now contain data from

309   more than one reference.

310

Sequence analysis program

311

312   The CoV-RDB website contains a SARS-CoV-2 sequence analysis program that leverages the

313   same code base written for the widely used Stanford HIV Drug Resistance Database interpretation program

314   [9,10]. The CoV-RDB sequence analysis program supports three types of input: (i) a list of spike mutations;

315   (ii) one or more consensus FASTA sequences containing any part of the SARS-CoV-2 genome; and (iii)

316   one or more codon frequency table (CodFreq) files containing the following seven columns: gene, amino

317   acid position, number of reads at that position, codon, amino acid encoded by the codon, number of reads

318   for that codon, and proportion of reads for that codon. An auxiliary program provided through the website

319   or via download enables users to convert next generation sequencing (NGS) FASTQ files into human and

320   machine-readable CodFreq files that are much faster to analyze. CodFreq files make it possible to estimate

321   the extent of background noise resulting from sequencing or PCR artifact (S3 Fig) [10].

322   If one or more spike mutations is submitted, the CoV-RDB sequence analysis program reports

323   information on specific mutations and generates summary tables containing the susceptibility of viruses

324   with these mutations to mAbs, CP, and VP (S4 Fig). If a FASTA sequence is submitted, the program returns

325   comments, summary tables, a list of the SARS-CoV-2 genes in the submitted sequence, a list of each of the

326   amino acid mutations in each virus gene, and the sequence's PANGO lineage (S4 Fig). If a codon frequency

327   table is submitted, the program returns all of the above information and also summarizes the read coverage

328   for each position along the genome and provides users with options to select read depth and mutation

329   detection thresholds below which mutations will not be reported (S4 Fig). If multiple sequences are

330   submitted, users have the option of obtaining all the results in a single downloadable CSV file.

331

332

333

334   Fig 6A-6C show three parts of the output generated when either a FASTQ sequence or CodFreq

335   file is submitted to the sequence analysis program: sequence summary (Fig 6A), sequence quality

336   assessment (Fig 6B), and mutation list (Fig 6C). The mutation comments and neutralization susceptibility

337   data, which are also included in the output, are not shown. The sequence summary section (Fig 6A) lists

338   the genes that underwent sequencing, the median read depth, and the Pango lineage. This section also

339   contains three dropdown boxes that help users select the appropriate threshold for identifying sub-consensus

340   mutations. The read depth and mutation detection thresholds are used to select the minimum number and

341   proportion of reads required for a mutation to be considered viral in origin rather than a sequence or PCR

342   artifact. The nucleotide mixture threshold allows users to select a threshold which minimizes the number

343   of nucleotide ambiguities present in the sequence.

344

345   **Fig 6. SARS-CoV-2 sequence analysis program output for FASTQ sequences or codon frequency**

346   **tables.** The sequence summary section (A) lists the genes that were sequenced, the median read depth, and

347   the Pango lineage. This section also contains dropdown boxes that enable the user to select the minimum

348   number and proportion of reads, respectively, required to identify a mutation. The threshold that minimizes

349   the proportion of positions with nucleotide ambiguities can also be selected. The sequence quality

350   assessment section displays the read depth across the genome and lists only those amino acid mutations

351   that meet the user-defined criteria outlined in the sequence summary section (B). The mutation list section

352   lists those amino acid mutations that meet the user-defined criteria and shows the proportion of reads

353   containing the mutation (C). The output shown in this figure can be regenerated by loading the example

354   file B.1.1.7 (ERR5026962) at this URL: https://covdb.stanford.edu/sierra/sars2/by-reads/.

355

Discussion

356

357   Because SARS-CoV-2 is a rapidly evolving RNA virus infecting millions of people worldwide,

358   new variants will continue to emerge and influence regional and global pandemic trajectories. Ongoing

359   epidemiological and clinical studies are needed to understand the risk of re-infection and vaccine failure

360   posed by different variants. However, because the spectrum of SARS-CoV-2 variants is expanding and

361   shifting faster than epidemiological studies can be conducted, laboratory-based markers will increasingly

362   be used to identify those variants that could predispose to re-infection and vaccine failure. SARS-CoV-2

363   neutralizing antibody titers are correlates of immune protection and are likely to be useful as an endpoint

364   in vaccine trials and to determine the need for vaccine boosters and immunogen updates.

365   CoV-RDB is the only database to comprehensively curate published data on the neutralizing

366   susceptibility of SARS-CoV-2 variants and spike mutations to mAbs, CP, and VP. Although non-

367   neutralizing Abs and cellular immune responses also contribute to protection from infection, the presence

368   of neutralizing antibodies targeting the spike protein has correlated most strongly with protection from

369   infection in animal models and in previously infected and vaccinated persons [11–14], Neutralizing

370   susceptibility assays are also better suited to standardization compared with assays of other immunological

371   defense mechanisms.

372   CoV-RDB can be downloaded in its entirety without restrictions. This is accomplished using a dual

373   database pipeline that combines the full-fledged Postgres database system to enforce relational data

374   integrity and the simplicity of the SQLite database system to enable users to download and query the

375   database without the overhead of accessing a host server. By making the database fully available to all

376   users, we aim to encourage data sharing and the editing of the underlying CSV files by the authors of

377   published studies.

378   Cov-RDB neutralizing susceptibility query output can be considered multidimensional tables in

379   which the rightmost columns contain numerical results (e.g., titers and fold-reductions in susceptibility)

380   while the leftmost columns contain experimental conditions. The experimental conditions are explanatory

381 variables that either directly influence neutralizing susceptibility (e.g., specific vaccine, time since

382 vaccination, and SARS-CoV-2 variant) or have a more subtle effect on susceptibility (e.g., type of

383 neutralizing assay and control virus). The CoV-RDB query interface enables users to explore the query

384 results at different levels of granularity by filtering or aggregating them according to one or more

385 experimental conditions without making additional calls to the web server.

386   The sequence analysis program shares many features with the Sierra HIV Drug Resistance

387 Database sequence analysis program. For the analysis of NGS data, both programs use Minimap2 [15] to

388 align individual reads to a reference sequence resulting in SAM/BAM files and SAM2CodFreq, a program

389 we wrote to create CodFreq files. The CodFreq format has several advantages over the commonly used

390 variant call format (VCF) because it can be interpreted without a reference sequence and can be used

391 independently from the accompanying SAM file. CodFreq files have a simple tabular format enabling them

392 to be viewed and manipulated using a spreadsheet. The sequence analysis program uses the CodFreq files

393 to assist users determine the appropriate threshold for distinguishing background sequence artifact from

394 authentic sub-consensus amino acids.

395

Study limitations

397   Neutralizing susceptibility data are highly heterogeneous often resulting in discordant results across

398 studies [16,17]. There are three main sources for this heterogeneity. First, the composition of neutralizing

399 antibodies among previously infected and vaccinated individuals is heterogeneous [6,18]. Second, there are

400 different types of neutralizing assays including those performed in cell culture using pseudotyped viruses,

401 chimeric viruses, recombinant viruses, and clinical isolates and, more recently, surrogate neutralizing

402 assays that assess the ability of antibodies to block the interaction between the SARS-CoV-2 RBD and

403 ACE2. Third, results for the same sample against a virus variant can differ even among laboratories using

404 the same type of assay as a result of differences in virus inoculum size, the cells used for culture, and viral

405 replication endpoints [16,17,19]. As neutralizing assays become more standardized and as external controls

406 such as those provided by the WHO [20] are increasingly used, it is likely that reproducibility across studies

407 will improve.

408 Data on the clinical significance of SARS-CoV-2 neutralizing susceptibility is continually evolving

409 [11,21–26]. The utility of neutralizing antibodies as a correlate of immune protection is ultimately

410 determined in epidemiological studies. Therefore, a database devoted to protective immunity should ideally

411 contain both laboratory and epidemiological data. The main obstacle to expanding CoV-RDB to also

412 include epidemiological studies of vaccine efficacy is that such studies are much more complex than those

413 reporting *in vitro* neutralizing data. For example, vaccine efficacy data depends not just on the vaccine, the

414 variant, and the time since vaccination but also on the study design and the age and immune status of the

415 study population. Moreover, in many vaccine efficacy studies, the proportion of individuals infected with

416 different variants is not known.

417

Future directions

419 Although the CoV-RDB is centered around just four main entities (references, viruses, antibodies,

420 and experiments), differences among the types of viruses, types of antibody preparations, and types of

421 experiments has necessitated a sophisticated database design. Nonetheless, as new types of experiments are

422 being published, we have been expanding the database schema. For example, an increasing number of

423 studies report the results of surrogate neutralizing assays based on antibody-mediated blocking of the

424 interaction of RBD with ACE2 [19,27,28]. These assays appear to correlate strongly with plasmid and

425 conventional neutralization test results, although they do not assess the effects of NTD-binding and other

426 non-RBD antibodies that synergistically inhibit virus replication [19]. In addition, the use of an international

427 external standard for calibration such as the one developed by the WHO will increase concordance across

428 different assays and will be reported using international units rather than as an IC_{50} (for mAbs) or a plasma

429 dilution [20,29].

430   We have also added the comprehensive deep mutational scanning data published by the Bloom

431   laboratory to the database [8,30–34]. The data from these studies are displayed only for those mAbs in

432   advanced clinical development and those mutations that have been reported to occur at a frequency above

433   0.001%. In addition to reporting the escape fraction associated with a mutation to an mAb, we report the

434   level of protein expression within yeast of RBDs containing the mutation (a measure of protein stability)

435   and the ACE2 binding of RBDs containing the mutation as mutations that bind poorly to ACE2 are less

436   likely to be selected *in vivo*. Although binding data have been reported for many other mAbs using enzyme

437   linked immunoassays, surface plasmon resonance, and biolayer interferometry, we have not curated these

438   data as they have not been as comprehensive as the deep mutational scanning data.

439   Commercial total binding assays do not differentiate between binding and neutralizing antibodies.

440   They also do not measure binding or neutralization of multiple variants but rather assess binding to pre-

441   variant spike proteins. However, total binding assays often display moderately strong correlations with

442   neutralizing assays [29,35,36] and activity against specific variants may eventually be assessed using

443   variant-specific reagents. We may eventually add such data to CoV-RDB if they will provide insights that

444   cannot be obtained solely from neutralizing antibody studies.

445   Although CoV-RDB contains neutralizing susceptibility data obtained using the plasma from

446   infected animal experiments (e.g., non-human primates, hamsters, and mice), we have not included data

447   from animal model challenge studies as such studies would require extensive modifications to our current

448   database schema. Therefore, we will continue to monitor these studies and consider adding top-line data

449   from these studies to alert database users to the existence of these studies without rigorously representing

450   study details. Finally, a similar approach will be considered for studies of vaccine efficacy. We are therefore

451   exploring the possibility of adding the top-line data of these studies so that the findings of these studies can

452   be correlated with the *in vitro* neutralizing data in CoV-RDB.

453

454

455

References

457   1.   Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al. The biological

458   and clinical significance of emerging SARS-CoV-2 variants. Nat Rev Genet. 2021.

459   doi:10.1038/s41576-021-00408-x

460   2.   Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of

461   the 2019-nCoV spike in the prefusion conformation. Science. 2020;367: 1260–1263.

462   doi:10.1126/science.abb2507

463   3.   Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and

464   Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181: 281-292.e6.

465   doi:10.1016/j.cell.2020.02.058

466   4.   Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-

467   binding domain bound to the ACE2 receptor. Nature. 2020;581: 215–220. doi:10.1038/s41586-020-

468   2180-5

469   5.   Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by

470   SARS-CoV-2. Nature. 2020;581: 221–224. doi:10.1038/s41586-020-2179-y

471   6.   Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive

472   mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by

473   polyclonal human plasma antibodies. Cell Host Microbe. 2021;29: 463-476.e6.

474   doi:10.1016/j.chom.2021.02.003

475   7.   Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2

476   variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19: 409–424.

477   doi:10.1038/s41579-021-00573-0

478   8.   Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational

479   Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2

480   Binding. Cell. 2020;182: 1295-1310.e20. doi:10.1016/j.cell.2020.08.012

481   9.   Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus

482   reverse transcriptase and protease sequence database. Nucleic Acids Res. 2003;31: 298–303.

483   doi:10.1093/nar/gkg100

484   10.   Tzou PL, Kosakovsky Pond SL, Avila-Rios S, Holmes SP, Kantor R, Shafer RW. Analysis of

485   unusual and signature APOBEC-mutations in HIV-1 pol next-generation sequences. PLoS One.

486   2020;15: e0225352. doi:10.1371/journal.pone.0225352

487   11.   Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody

488   levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. Nature

489   Medicine. 2021 [cited 23 May 2021]. doi:10.1038/s41591-021-01377-8

490   12.   Earle KA, Ambrosino DM, Fiore-Gartland A, Goldblatt D, Gilbert PB, Siber GR, et al. Evidence for

491   antibody as a protective correlate for COVID-19 vaccines. Vaccine. 2021.

492   doi:10.1016/j.vaccine.2021.05.063

493   13.   McMahan K, Yu J, Mercado NB, Loos C, Tostanoski LH, Chandrashekar A, et al. Correlates of

494   protection against SARS-CoV-2 in rhesus macaques. Nature. 2020;590: 630–634.

495   doi:10.1038/s41586-020-03041-6

496   14.   Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, et al. Antibodies to

497   SARS-CoV-2 are associated with protection against reinfection. medRxiv. 2020;

498   2020.11.18.20234369. doi:10.1101/2020.11.18.20234369

499 15. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–

500   3100. doi:10.1093/bioinformatics/bty191

501 16. Chen X, Chen Z, Azman AS, Sun R, Lu W, Zheng N, et al. Neutralizing Antibodies Against Severe

502   Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants Induced by Natural Infection

503   or Vaccination: A Systematic Review and Pooled Analysis. Clinical Infectious Diseases. 2021

504   [cited 9 Sep 2021]. doi:10.1093/cid/ciab646

505 17. Huang Y, Borisov O, Kee JJ, Carpp LN, Wrin T, Cai S, et al. Calibration of Two Validated SARS-

506   CoV-2 Pseudovirus Neutralization Assays for COVID-19 Vaccine Evaluation. 2021 Sep p.

507   2021.09.09.21263049. doi:10.1101/2021.09.09.21263049

508 18. Greaney AJ, Starr TN, Eguia RT, Loes AN, Khan K, Karim F, et al. A SARS-CoV-2 variant elicits

509   an antibody response with a shifted immunodominance hierarchy. 2021 Oct p. 2021.10.12.464114.

510   doi:10.1101/2021.10.12.464114

511 19. Sholukh AM, Fiore-Gartland A, Ford ES, Miner MD, Hou YJ, Tse LV, et al. Evaluation of Cell-

512   Based and Surrogate SARS-CoV-2 Neutralization Assays. Journal of Clinical Microbiology. 59:

513   e00527-21. doi:10.1128/JCM.00527-21

514 20. Kristiansen PA, Page M, Bernasconi V, Mattiuzzo G, Dull P, Makar K, et al. WHO International

515   Standard for anti-SARS-CoV-2 immunoglobulin. The Lancet. 2021;397: 1347–1348.

516   doi:10.1016/S0140-6736(21)00527-4

517 21. Earle KA, Ambrosino DM, Fiore-Gartland A, Goldblatt D, Gilbert PB, Siber GR, et al. Evidence for

518   antibody as a protective correlate for COVID-19 vaccines. Vaccine. 2021;39: 4423–4428.

519   doi:10.1016/j.vaccine.2021.05.063

520   22.   Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta K-D, et al. Effect of Delta

521   variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK.

522   Nat Med. 2021; 1–9. doi:10.1038/s41591-021-01548-7

523   23.   Clemens SAC, Folegatti PM, Emary KRW, Weckx LY, Ratcliff J, Bibi S, et al. Efficacy of

524   ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 lineages circulating in Brazil. Nat

525   Commun. 2021;12: 5861. doi:10.1038/s41467-021-25982-w

526   24.   Levin EG, Lustig Y, Cohen C, Fluss R, Indenbaum V, Amit S, et al. Waning Immune Humoral

527   Response to BNT162b2 Covid-19 Vaccine over 6 Months. New England Journal of Medicine.

528   2021;0: null. doi:10.1056/NEJMoa2114583

529   25.   Bruxvoort K, Sy LS, Qian L, Ackerson BK, Luo Y, Lee GS, et al. Effectiveness of mRNA-1273

530   against Delta, Mu, and other emerging variants. 2021 Oct p. 2021.09.29.21264199.

531   doi:10.1101/2021.09.29.21264199

532   26.   Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, et al. Correlates of protection against

533   symptomatic and asymptomatic SARS-CoV-2 infection. Nat Med. 2021; 1–9. doi:10.1038/s41591-

534   021-01540-1

535   27.   Abe KT, Li Z, Samson R, Samavarchi-Tehrani P, Valcourt EJ, Wood H, et al. A simple protein-

536   based surrogate neutralization assay for SARS-CoV-2. JCI Insight. 2020;5: e142362.

537   doi:10.1172/jci.insight.142362

538   28.   Tan CW, Chia WN, Qin X, Liu P, Chen MI-C, Tiu C, et al. A SARS-CoV-2 surrogate virus

539   neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction.

540   Nat Biotechnol. 2020;38: 1073–1078. doi:10.1038/s41587-020-0631-z

541   29.   Kung Y-A, Huang C-G, Huang S-Y, Liu K-T, Huang P-N, Yu K-Y, et al. Antibody titers measured

542   by commercial assays are correlated with neutralizing antibody titers calibrated by international

543   standards. medRxiv. 2021; 2021.07.16.21260618. doi:10.1101/2021.07.16.21260618

544   30.   Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective

545   mapping of viral mutations that escape antibodies used to treat COVID-19. Science. 2021;371: 850–

546   854. doi:10.1126/science.abf9302

547   31.   Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that

548   escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Reports

549   Medicine. 2021;2: 100255. doi:10.1016/j.xcrm.2021.100255

550   32.   Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete Mapping of

551   Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition.

552   Cell Host Microbe. 2021;29: 44-57.e9. doi:10.1016/j.chom.2020.11.007

553   33.   Greaney AJ, Loes AN, Gentles LE, Crawford KHD, Starr TN, Malone KD, et al. Antibodies elicited

554   by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from

555   SARS-CoV-2 infection. Science Translational Medicine. 2021;13: eabi9915.

556   doi:10.1126/scitranslmed.abi9915

557   34.   Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, et al. Mapping mutations to

558   the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun.

559   2021;12: 4196. doi:10.1038/s41467-021-24435-8

560   35.   Peluso MJ, Takahashi S, Hakim J, Kelly JD, Torres L, Iyer NS, et al. SARS-CoV-2 antibody

561   magnitude and detectability are driven by disease severity, timing, and assay. Science Advances.

562   2021;7: eabh3409. doi:10.1126/sciadv.abh3409

563   36.   Savage HR, Santos VS, Edwards T, Giorgi E, Krishna S, Planche TD, et al. Prevalence of

564   neutralising antibodies against SARS-CoV-2 in acute infection and convalescence: A systematic

565   review and meta-analysis. PLoS Negl Trop Dis. 2021;15: e0009551.

566   doi:10.1371/journal.pntd.0009551

567

568

Supporting information

569

570

571 **S1 Table.** Slightly different patterns of spike mutations within each of the SARS-CoV-2 variants of

572 concern (VOCs) and variants of interest (VOIs).

573

574 **S1 Fig. Data management workflow for CoV-RDB.** Weekly incremental searches of PubMed and

575 preprint servers (BioRxiv/MedRxiv and Research Square) are performed. Publications that appear to have

576 data pertinent to SARS-CoV-2 variants and their susceptibility to mAbs, convalescent plasma (CP), and

577 vaccinee plasma (VP) are downloaded to a Zotero reference database folder to enable full-text review and

578 data curation. Extracted data are exported into a set of linked CSV files in an open-source Github repository

579 (https://github.com/hivdb/covid-drdb-payload). Extracted data are then imported into a Postgres database

580 where the data are validated for completeness and consistency before being exported as a single SQLite

581 database file that serves as the back end for the CoV-RDB website and is available to users for download.

582

583 **S2 Fig.** Effects of spike receptor binding domain (RBD) mutations on monoclonal antibody (mAb) binding

584 and neutralization. The X-axis shows the escape fraction as determined using a deep mutational scanning

585 (DMS) assay that measures the binding of various mAbs to RBD variants produced in yeast. The Y-axis

586 shows the results of in vitro neutralization using the same mAbs and spike proteins with the same RBD

587 mutations as those used in the DMS assay. Discordant results are shown in the figure. Abbreviations: SOT:

588 sotrovimab; BAM: bamlanivimab; ETE: etesevimab; CAS: casirivimab; IMD: imdevimab.

589

590 **S3 Fig. Creating a codon frequency file from a FASTQ file.** FASTQ files are aligned to the consensus

591 Wuhan reference sequence using the Minimap2 alignment program. The resulting BAM/SAM files are then

592 processed by a program SAM2CodFreq that we wrote to generate a codon frequency file containing seven

593 columns as shown on the right. The table here shows the results from three codons (spike positions 500 to

594   502). The observation that many codons shown in this (and other parts of the same file which are not shown)

595   are present at levels between 0.2% and about 2% suggests that codons present at these low proportions

596   likely represent sequencing or PCR artifact (i.e., "background noise"). However, as the mutation N501Y

597   occurs at a considerably higher proportion (34.3%), it is likely to be present in the infecting virus population.

598

599   **S4 Fig. Functions of the SARS-CoV-2 sequence analysis program.** The program supports three types of

600   input: a list of spike mutations; one or more consensus FASTA sequences containing any part of the SARS-

601   CoV-2 genome; and one or more FASTQ sequences. However, because a FASTQ sequence can take several

602   minutes to analyze, users are advised to first convert them to a codon frequency file through an auxiliary

603   program. If a list of spike mutations is submitted, the program returns comments about notable mutations

604   and summary tables reporting the susceptibility of viruses with these mutations to mAbs, CP, and VP. If a

605   FASTA sequence is submitted, the program returns the preceding information plus a list of the SARS-CoV-

606   2 genes, the amino acid mutations in the sequence, and the sequence's PANGO lineage. If a FASTQ

607   sequence or codon frequency table is submitted, the program provides the preceding information and the

608   read coverage for each position along the genome. It also provides users with the options to select read

609   depth and mutation detection thresholds below which mutations will not be reported.

610

Fig 1

Fig 2

A
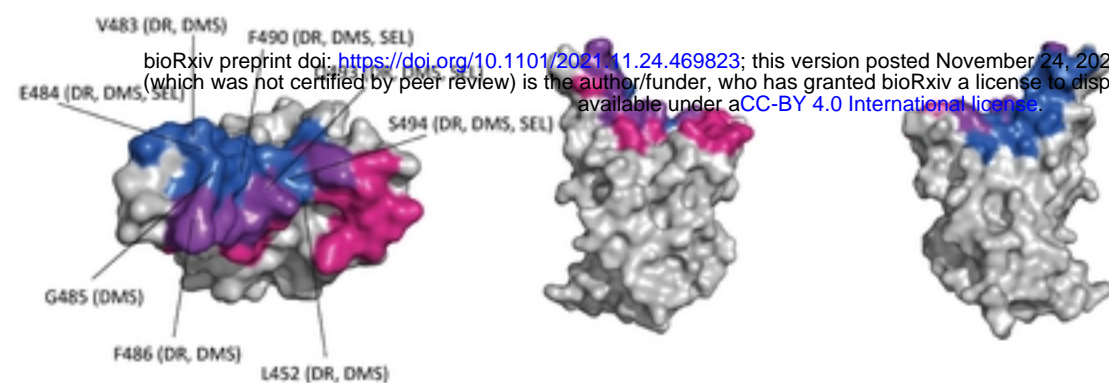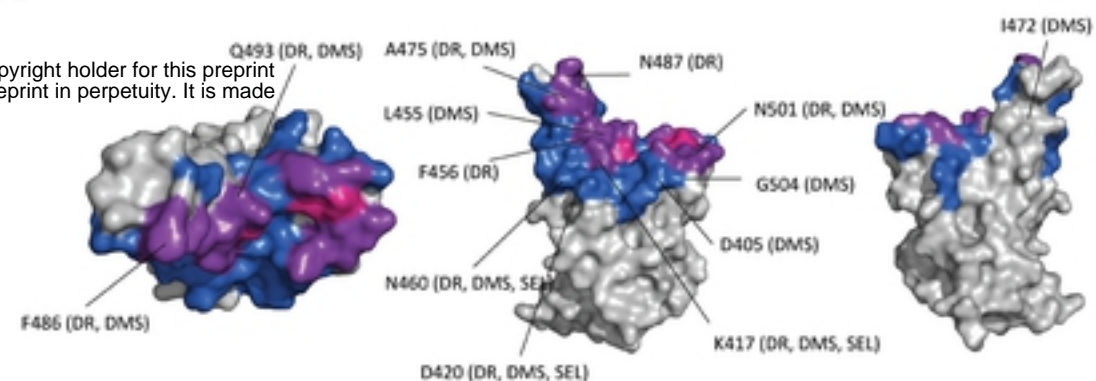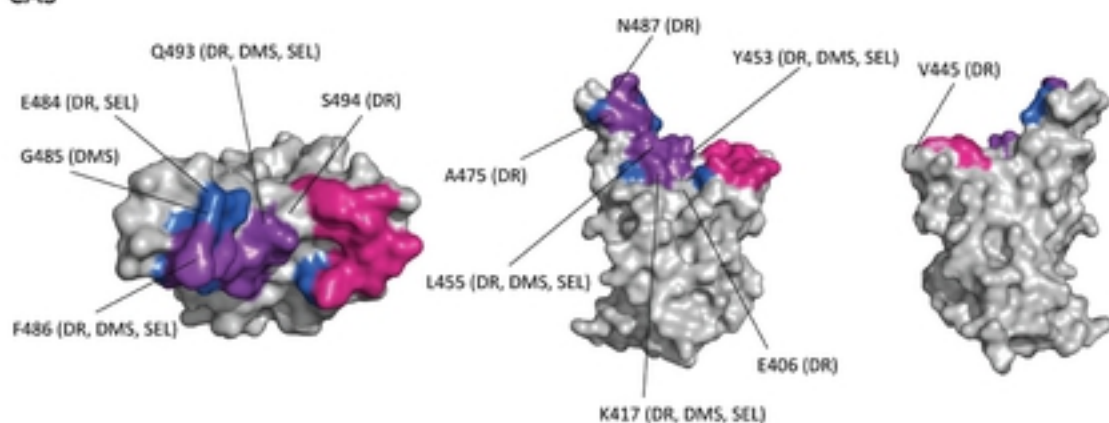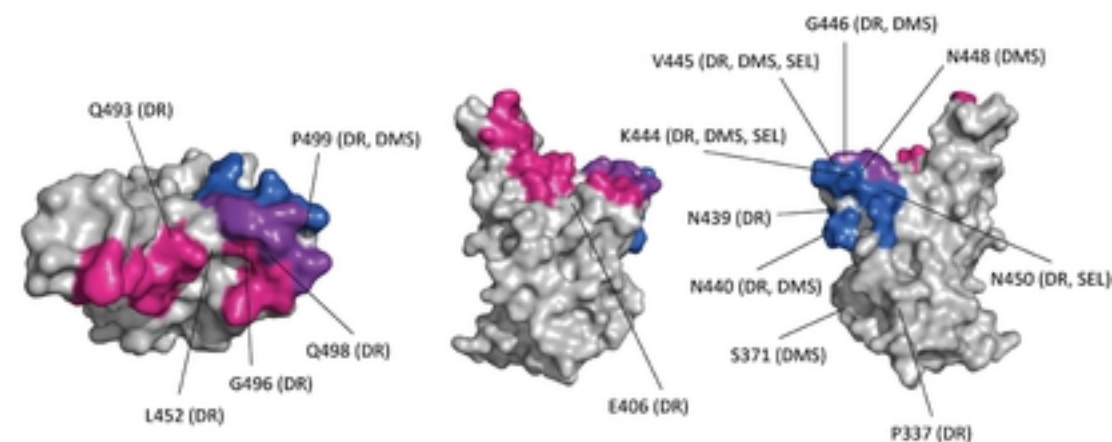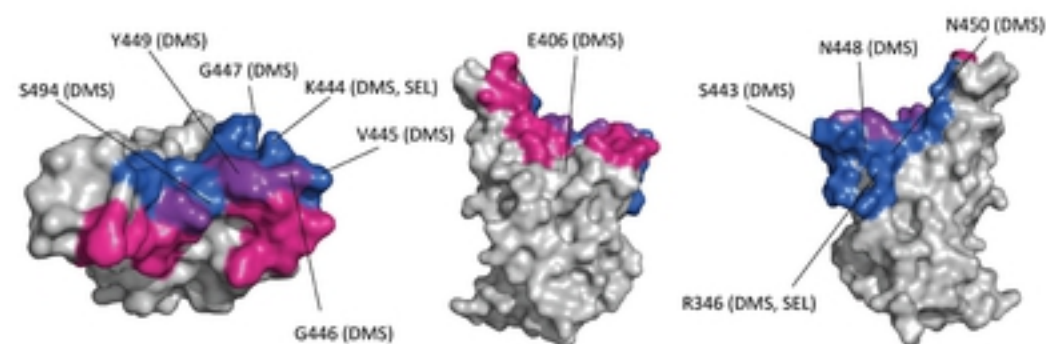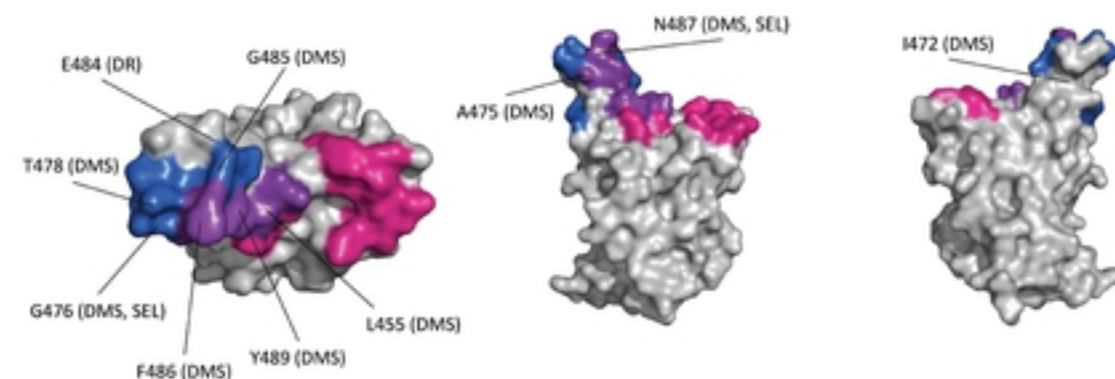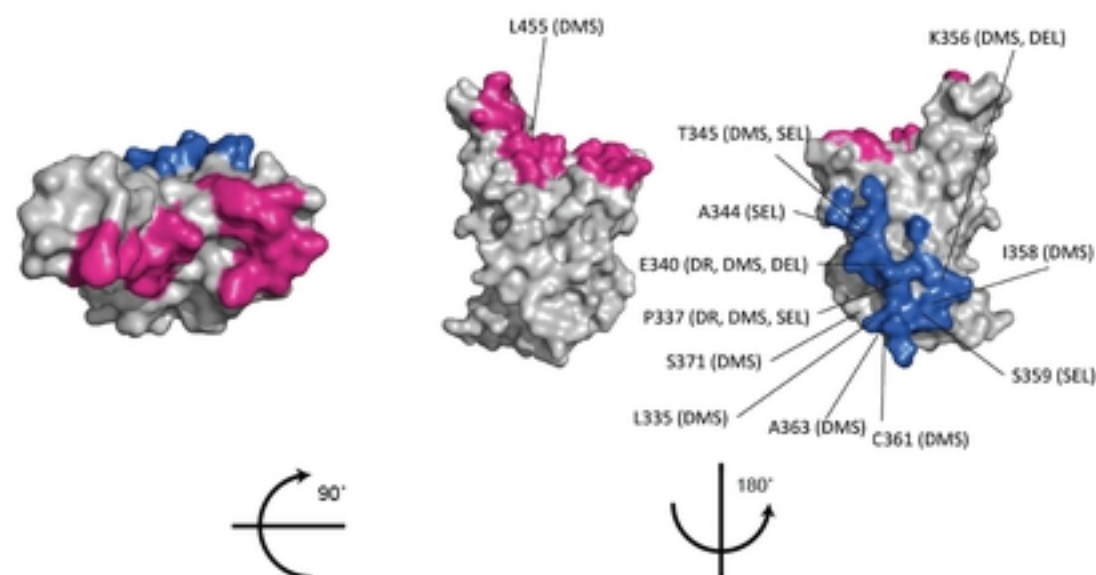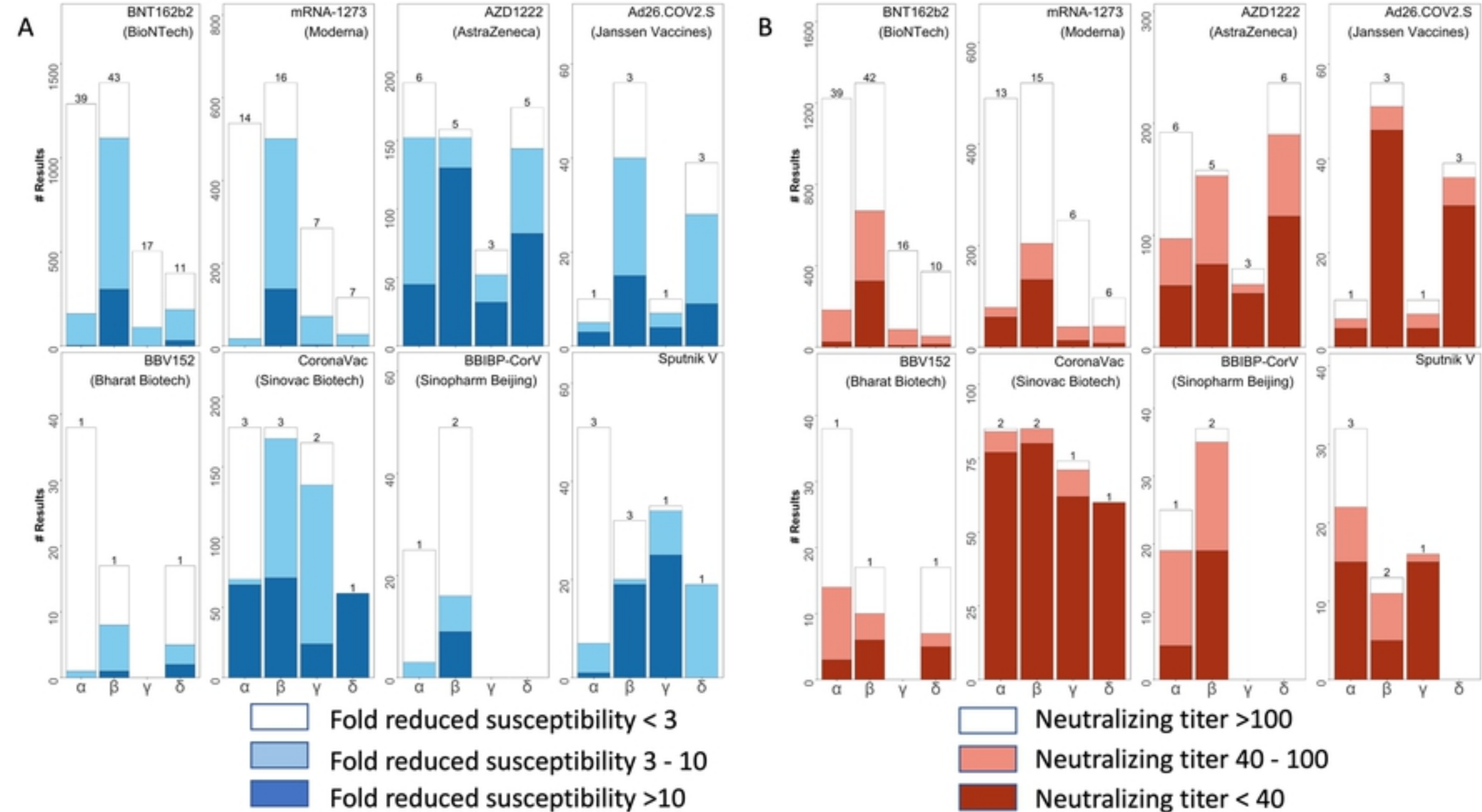
References

Any ▼

Plasma Abs / mAbs

Any ▼

Variants / Mutations

E484K ▼

B

| Wild type infection (n=209) | Alpha infection... | Beta infection (... | mRNA-1273 (n=107) | BNT162b2 (n=95) | Sputnik V (n=12) | MAbs (n=239) |

E484K (n=685) ⊗

Monoclonal antibody - Individual mutation:	244 results, 28 publications
└ Individual sample data are available:	244 results, 28 publications
Convalescent plasma - Individual mutation:	232 results, 15 publications
└ Individual sample data are available:	202 results, 14 publications
└ Only aggregate sample data are available:	30 results, 2 publications
Vaccinee plasma - Individual mutation:	214 results, 11 publications
└ Individual sample data are available:	204 results, 10 publications
└ Only aggregate sample data are available:	10 results, 1 publication

C

MAb Susceptibility Data

Table dimensions: ☑ Reference ☑ Assay ☑ Antibodies ☑ Control ☑ Mutation

244 results; 233 experimental conditions (rows); 28 publications.

Download Excel ▼

Reference ⇕	Assay ⇕	Section ⇕	Antibodies ⇕	Control ⇕	Mutation ⇕	# Results ⇕	IC50 (ng/ml) ⇕	Fold Reduction ⇕	Data Availability ⇕
Wang, P (2021b)	Pseudovirus (VSV)	Figure 2a	Bamlanivimab	B.1 (D614G)	E484K	1	N.A.	>100.0	▦✓

D

Vaccinee Plasma Susceptibility Data

Table dimensions: ☑ Reference ☑ Assay ☑ Pre-vaccine Infection ☑ Vaccine ☑ # Shots ☑ Months ☐ Host ☑ Control ☑ Mutation

214 results; 18 experimental conditions (rows); 11 publications.

Download Excel ▼

Reference ▲	Assay ⇕	Section ⇕	Pre-vaccine Infection	Vaccine ⇕	# Shots ⇕	Months ⇕	Control ⇕	Mutation ⇕	# Results ⇕	NT50 Dilution GeoMean⚡GSD ⇕	Fold Reduction Median (IQR) ⇕	Data Availability ⇕
Ferreira, I (2021)	Pseudovirus (HIV)	Supplementary Figure 1A; Supplementary Figure 1C	None	BNT162b2	2	2-6	B.1 (D614G) NT50: 582	E484K	34	228⚡3.3	2.3 (1.7-3.2)	▦✓

E

🔗 **Convalescent Plasma Susceptibility Data**

Table dimensions: ☑ Reference ☑ Assay ☑ Infection (CP) ☑ Months ☐ Host ☑ Control ☑ Mutation

232 results; 21 experimental conditions (rows); 15 publications.
7 results had titers against the control virus below the experimental detection threshold and are not shown. (unhide)

Download Excel ▼

Reference ⇕	Assay ⇕	Section ⇕	Infection (CP) ⇕	Months ⇕	Control ⇕	Mutation ⇕	# Results ⇕	NT50 Dilution GeoMean⚡GSD ⇕	Fold Reduction Median (IQR) ▼	Data Availability ⇕
Greaney, AJ (2021)	Pseudovirus (HIV)	Figure 5; git.io/JCEvr	Wild Type	2-6	B.1 (D614G) NT50: 1,548	E484K	1	65	23.8	▦✓

Fig 3

A

References

Any ▼

Plasma Abs / mAbs

BNT162b2 ▼

Variants / Mutations

Delta (B.1.617.2) ▼

mRNA vaccine ⊗

BNT162b2

Developed by the German biotechnology company BioNTech, **BNT162b2** is an mRNA-based COVID-19 vaccine authorized by FDA and recommended by the CDC for use in the US for a limited population.

Other names: Comirnaty, Tozinameran
Legal status: FDA Full Approval; WHO Emergency Use

Vaccine type: mRNA
Dosage: 2 shots, 21 days apart

Manufacturer: Pfizer, Inc. and BioNTech
Links: FDA, CDC, Vaccine Tracker, MedlinePlus, Wikipedia

BNT162b2 (n=875) ⊗

Delta (n=875) ⊗

Vaccinee plasma - Variant / mutation combination:	875 results, 12 publications
Individual sample data are available:	802 results, 9 publications
Only aggregate sample data are available:	73 results, 3 publications

B

🔗 Vaccinee Plasma Susceptibility Data

Table dimensions: ☑ Reference ☑ Assay ☑ Pre-vaccine Infection ☑ Vaccine ☑ # Shots ☑ Months ☑ Host ☑ Control ☑ Variant

875 results; *23* experimental conditions (rows); *12* publications.

Copy to clipboard ▼

Reference ⇕	Assay ⇕	Section ⇕	Pre-vaccine Infection	Vaccine ⇕	# Shots ⇕	Months ⇕	Host ⇕	Control ⇕	Variant ⇕	# Results	NT50 Dilution GeoMean⚥GSD	Fold Reduction Median (IQR)	Data Availability
Edara, VV (2021c)	Virus isolate	Supplementary Table S6	None	BNT162b2	2	1	Human	A (WA1) NT50: 777	Delta (B.1.617.2)	10	235⚥1.6	3.4 (2.5-4.5)	▦ ✓
Wilhelm, A (2021)	Virus isolate	Figure 1B	None	BNT162b2	2	1	Human	B (Wuhan-Hu-1)	Delta (B.1.617.2)	8	N.A.	3.5	▦ ✗

Fig 4

A

Table dimensions: ☑ Reference ☑ Assay ☑ Pre-vaccine Infection (None, Infected) ☑ Vaccine ☑ # Shots (1, 2, 3) ☑ Months (1, 2-6, ≥6) ☐ Host (Human ▼, Animal) ☑ Control ☑ Variant (Select all · Reset)

885 results; 22 experimental conditions (rows); 13 publications.
107 results had titers against the control virus below the experimental detection threshold and are not shown. (unhide)

Download CSV

Reference ⇕	Assay ⇕	Section ⇕	Pre-vaccine Infection	Vaccine ⇕	# Shots ⇕	Months ⇕	Control ⇕	Variant ⇕	# Results ⇕	NT50 Dilution GeoMean⚖GSD ⇕	Fold Reduction Median (IQR)	Data Availability ⇕
Liu, J (2021e)	SARS-CoV-2 recombinant	Extended Data Table 1	None	BNT162b2	2	1	A (WA1) NT50: 502	Delta (B.1.617.2)	40	349⚖1.7	1.4 (1.0-2.0)	▦ ✓
Arora, P (2021)	Pseudovirus (VSV)	Figure S1E	None	BNT162b2	2	1	B (Wuhan-Hu-1) NT50: 654	Delta (B.1.617.2)	15	195⚖3.3	2.8 (2.0-6.7)	▦ ✓
McCallum, M (2021c)	Pseudovirus (VSV)	Figure 1A	None	BNT162b2	2	1	B.1 (D614G) NT50: 302	Delta (B.1.617.2)	15	111	2.7	▦ ✗

B

Table dimensions: ☐ Reference ☐ Assay ☑ Pre-vaccine Infection (None, Infected) ☑ Vaccine ☑ # Shots (1, 2, 3) ☑ Months (1, 2-6, ≥6) ☐ Host (Human ▼, Animal) ☐ Control ☑ Variant (Select all · Reset)

885 results; 5 experimental conditions (rows); 13 publications.
107 results had titers against the control virus below the experimental detection threshold and are not shown. (unhide)

Download CSV

Pre-vaccine Infection ⇕	Vaccine ⇕	# Shots ⇕	Months ⇕	Variant ⇕	# Publications ⇕	# Results ⇕	NT50 Dilution GeoMean⚖GSD ⇕	Fold Reduction Median (IQR) ⇕	Data Availability ⇕
None	BNT162b2	2	1	Delta (B.1.617.2)	12	395	277⚖3.3	2.9 (2.0-5.6)	▦ ✓
None	BNT162b2	2	2-6	Delta (B.1.617.2)	4	127	145⚖3.5	4.8 (3.3-6.7)	▦ ✓
None	BNT162b2	1	1	Delta (B.1.617.2)	2	162	19⚖4.6	9.6 (4.6-13.4)	▦ ✓
None	BNT162b2	1	2-6	Delta (B.1.617.2)	1	73	26⚖5.2	8.8 (5.0-12.8)	▦ ✓
Wild Type	BNT162b2	1	1	Delta (B.1.617.2)	1	21	2,096⚖2.3	1.7 (1.3-1.9)	▦ ✓

Fig 5

Fig 6