Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons

Philip L. Tzou ^{1¶}, Kaiming Tao^{1¶}, Sergei L. Kosakovsky Pond², Robert W. Shafer^{1*}

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

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

* Corresponding author

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

[¶]These authors contributed equally to this work.

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

Beginning in late 2020, several SARS-CoV-2 variants sharing multiple spike mutations were reported from different parts of the world. These variants have been classified according to their phylogenetic lineage and component mutations. Variants that spread widely and displayed evidence for being more transmissible, causing more severe disease and/or reducing neutralization by antibodies generated during previous infection or vaccination have been classified as variants of concern (VOCs) by the World Health Organization and U.S. Centers for Disease Control and Prevention (reviewed in [1]). Variants that have spread less widely but share some of the key mutations present within VOCs have been classified as variants of interest (VOIs).

As novel SARS-CoV-2 variants have emerged, numerous investigations assessed the susceptibility of individual and combination spike mutations to neutralization by monoclonal antibodies (mAbs), convalescent plasma (CP), and vaccinee plasma (VP). The susceptibility of SARS-CoV-2 variants to mAbs is obviously relevant for preventing and treating SARS-CoV-2 infections with mAb regimens. The susceptibility of SARS-CoV-2 variants to CP provides insight into the likelihood that a variant will infect and cause illness in a person previously infected with and recovered from a different variant. The susceptibility of SARS-CoV-2 variants to VP provides insight into the risk that a variant will infect and cause illness in a previously vaccinated person.

The SARS-CoV-2 spike protein is a 1,273-amino acid trimeric glycoprotein responsible for entry into host cells. Each spike monomer has a largely exposed S1 attachment domain (residues 1–686) and a partially buried S2 fusion domain (residues 687–1,273) [2,3]. Part of S1, called the receptor-binding domain (RBD; residues 306–534) binds to the human angiotensin-converting enzyme 2 (ACE2) receptor [4,5]. Approximately 20 RBD residues bind the ACE2 receptor. The part of the RBD containing these residues (438–506) is referred to as the receptor-binding motif (RBM), whereas the remainder of the RBD is called the RBD core. The SARS-CoV-2 RBD is the main target of neutralizing antibodies [6,7]. Like the RBD,

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

The Stanford Coronavirus Resistance Database (CoV-RDB) provides open access to continuously curated neutralizing susceptibility data of SARS-CoV-2 mutations and variants to mAbs, CP, and VP. The database includes specific data on which spike mutations have been selected during both *in vitro* passage of SARS-CoV-2 in the presence of mAbs and *in vivo* emergence in persons receiving mAbs for treatment. The CoV-RDB website (<https://covdb.stanford.edu>) enables users to query its database and the associated Github repository enables users to download the entire database. The CoV-RDB website also contains a companion sequence analysis program that annotates SARS-CoV-2 user-submitted sequences using CoV-RDB data.

Methods and results

The CoV-RDB data model is made up of four major entities: published references, viral mutations and variants, antibodies (including mAbs, CP, and VP), and experimental results. As of October 2021, the database contains 186 publications: 18 (10%) with 380 results from in vitro mAb selection experiments, 64 (34%) with 7,328 neutralizing mAb susceptibility results, 96 (52%) containing 11,390 with CP susceptibility results, and 125 (68%) with 20,872 neutralizing VP results. As publications often contain more than one type of experiment, the sum of the percentages is greater than 100%. The complete database schema is available at <https://github.com/hivdb/covid-drdb/blob/master/schema.dbml>.

Curation of published references

Published references included in the CoV-RDB are obtained weekly from three literature sources: i) PubMed using the search term “SARS-CoV-2”; ii) BioRxiv/MedRxiv COVID-19 SARS-CoV-2 preprint servers; and iii) the Research Square SARS-CoV-2/COVID-19 preprint server. Publication titles and abstracts are reviewed manually to identify studies containing data on SARS-CoV-2 spike mutations and humoral immunity. Studies that pass initial review are downloaded to a Zotero reference database and full texts are reviewed to extract specific data on the selection of spike mutations *in vitro* and *in vivo*, and neutralizing susceptibility data for individual spike mutations or combinations of spike mutations, such as those present in SARS-CoV-2 variants to mAbs, CP, and VP. Curated studies initially published as preprints are re-evaluated following peer-review publication.

For each publication, data are first entered into linked comma-separated files (CSVs) contained in a Github repository (<https://github.com/hivdb/covid-drdb-payload>). Data are then imported into a Postgres database where they are validated for completeness and consistency. The data in the Postgres database are 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

In the CoV-RDB data model, the virus entity represents individual spike mutations, combinations of spike mutations, and virus variants for which the full set of spike and genomic mutations is known including VOCs and VOIs. Published *in vitro* neutralization experiments have been performed using (i) pseudotyped viruses (vesicular stomatitis viruses [VSV] and lentiviruses) containing a SARS-CoV-2 spike protein with specific mutations; (ii) chimeric SARS-CoV-2 viruses in a VSV genomic backbone containing a spike protein with specific mutations; (iii) full-length cloned SARS-CoV-2 recombinant viruses generated using multiple plasmids, or bacterial or yeast artificial chromosomes; and (iv) primary SARS-CoV-2 isolates [1]. For pseudotyped, chimeric, and recombinant viruses, the virus can be characterized entirely by its spike mutations. For the primary SARS-CoV-2 isolates, the virus is also characterized by mutations in other viral proteins that may influence viral replication kinetics but not neutralization susceptibility. Most *in vitro* selection experiments have been performed using chimeric VSVs containing SARS-CoV-2 spike proteins as these viral constructs can undergo multiple rounds of replication in cell culture.

Table 1 lists the VOCs, VOIs, and other full-genomic SARS-CoV-2 variants for which data are available in CoV-RDB. Approximately 85% of results on variants are for VOCs, 12% are for VOIs, and 3% are for other variants containing one or more RBD mutations. As SARS-CoV-2 variants are continually evolving and variant definitions are regularly updated, individual VOCs and VOIs often differ slightly in the mutations that they contain. S1 Table displays the variability in these patterns for each of the VOCs and VOIs. For example, the alpha, beta, gamma, and delta variants in CoV-RDB contain 11, 20, 7, and 17 unique patterns of spike mutations, respectively, all closely related to an archetypal consensus variant.

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

Virus types	References	Results
<i>Variants of concern</i>		
Alpha	106	7439
Beta	114	7823
Gamma	61	3247
Delta	40	2612
<i>Variants of interest</i>		
Epsilon	22	926
Eta	9	216
Iota	14	554
Kappa	23	823
Lambda	8	324
Mu	4	32
<i>Other variants containing >=1 RBD mutation (n=17)</i>		
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

Table 2 lists the most common studied individual mutations within the spike RBD, NTD, C-terminal domain (CTD), and S2 domain. These mutations have generally been studied within an ancestral virus backbone. CoV-RDB contains experimental data on 296 different individual spike mutations at 159 positions including 197 RBD mutations at 76 positions, 55 NTD mutations at 43 positions, 32 S2 mutations at 29 positions, and 12 CTD mutations at 11 positions. CoV-RDB also contains results on 89 combinations of spike mutations in addition to VOCs and VOIs. These generally consist of sets of two or three mutations that are present within a VOC or VOI. Several of the most common mutation combinations are listed in the footnote of Table 2.

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

Mutations	References	Results
<i>Receptor-binding domain (RBD)</i>		
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
<i>N-terminal domain (NTD)</i>		
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
<i>C-terminal domain of S1 (CTD)</i>		
P681H	6	110
A570D	4	88
P681R	1	20
H655Y	1	12
Q675H	1	12
Other mutations	5	110
<i>S2</i>		
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
BR11-196	BR11 Biosciences	Completed phase III	7CDI	RBM I	4	146
BR11-198	BR11 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:¹RBM (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. ²The 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 Development¹

	BAM	ETE	CAS	IMD	CIL	TIX	SOT	REG	BR11-196	BR11-198	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: ¹The 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

Fig 1 shows the relationship between mAb epitopes, mutations selected during *in vitro* passage, and mutations that reduce mAb binding and/or susceptibility. S2 Fig shows that there is a strong correlation between RBD binding in a high throughput deep mutational scanning (DMS) assay and neutralization in that an escape fraction of >0.1 in the DMS assay is with few exceptions associated with a >10-fold reduction in neutralization in cell culture [8].

Fig 1. Monoclonal antibodies (mAbs) with EUAs or in advanced clinical development: receptor binding domain (RBD) epitopes and immune escape positions. For each mAb, the top of the RBD and two side views are depicted using coordinates from PDB 6M0J. ACE2 binding residues are shown in red; the mAb epitope defined as those residues within 4.5 angstroms of the RBD is shown in dark blue; and ACE2 binding residues within the mAb epitope are shown in purple. Those positions containing mutations that were either selected by the mAb *in vitro* (“SEL”), reduced binding in a deep mutational scanning assay (“DMS”), and/or reduced *in vitro* neutralizing susceptibility by a median of ≥ 4 -fold in CoV-RDB (drug resistance; “DR”) are also indicated. The mAb epitopes for BAM (bamlanivimab), ETE (etesevimab), CAS (casirivimab), IMD (imdevimab), SOT (sotrovimab), CIL (cilgavimab) and TIX (tixagevimab) were determined from their PDB structures.

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
<i>Infecting virus</i>		
Ancestral (A, B, B.1, others)	86	8778
Alpha	10	825
Beta	7	214
Gamma	2	38
Delta	1	40
<i>Time since infection</i>		
1 month	74	6661
2-3 months	35	2743
4-6 months	23	1397
>6 months	14	589
<i>Variant / mutations undergoing neutralization</i>		
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

The time since infection was available for all CP sample results including 10,801 samples obtained within six months of infection and 589 samples obtained beyond six months. The severity of illness was described for 3,559 (31%) CP samples with 2,133 samples from persons with mild-to-moderate disease and 1,426 samples from persons with severe disease. Although the sequence of the infecting virus was rarely known, the vast majority were obtained prior to the emergence of VOCs and VOIs. Nonetheless, 7%, 2%, 1%, and 1% were obtained from persons known to be infected with the Alpha, Beta, Gamma, and Delta variants, respectively.

VP

In CoV-RDB, VP are characterized according to the vaccine received, number of vaccinations, time since vaccination, and whether the VP was obtained from a person who was also previously infected with SARS-CoV-2. Table 6 lists the numbers of experimental results according to each of the above characteristics and the variant or mutation(s) tested for neutralization. 70% of VP were obtained from persons receiving one of the two widely used mRNA vaccines (BNT162b and mRNA-1273) while about 20% were obtained from persons receiving the Coronovac, AZD1222, or Ad26.COV2s vaccines. Approximately 80%, 18%, and 2% of VP samples were obtained within 1 month, 2-6 months, and >6 months after vaccination, respectively. Nearly 5% of samples were obtained from persons with confirmed infection prior to vaccination.

Table 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
<i>Vaccine</i>		
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
<i>Previously SARS-CoV-2 infected</i>		
Yes	14	837
No	125	20035
<i>Time since vaccination</i>		
1 month	113	16789
2-3 months	21	3072
4-6 months	10	662
> 6 months	8	349
<i>Variant / mutations undergoing neutralization</i>		
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
<i>Number of vaccinations</i>		
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

Fig 2 shows the distribution of fold-reductions in neutralizing susceptibilities and absolute neutralizing antibody titers against each of the VOCs for previously uninfected persons completing the recommended course of vaccination for eight vaccines. A similar fold-reduction in neutralizing susceptibility against a given variant is likely to be more consequential for those vaccines that elicit lower titers of neutralizing antibodies. The figure shows that regardless of the vaccine, the fold reduction in neutralizing susceptibility was consistently greater for the Beta variant and that VP from persons receiving one of the two mRNA vaccines was usually more likely to retain neutralizing activity against each of the VOCs.

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

CoV-RDB Website

The CoV-RDB website contains four main features: (i) searchable tables containing SARS-CoV-2 variants, mAbs, vaccines, and references; (ii) regularly updated data summaries such as the data shown in Table 4 for mAbs and in Fig 2 for VP; (iii) user-defined queries; and (iv) a sequence analysis program. The user-defined queries and sequence analysis program are described here because they are interactive and contain multiple user options.

User-Defined queries

Query interface

The query interface allows users to search the database using one or more of the following three criteria: (i) published reference; (ii) antibody preparation (mAb, CP, or VP); and (iii) SARS-CoV-2 variant or spike mutation(s). If the “References” dropdown option is selected, then all the data associated with that reference is displayed in separate tables containing neutralization susceptibility data for mAbs, CP, and VP and/or *in vitro* selection data for mAbs. If the “Plasma / mAbs” dropdown option is selected, users must select either a specific mAb, CP from a person infected with a specific variant, or VP from persons who received a specific vaccine. If the “Variants / Mutations” dropdown option is selected, users must select a particular VOC, VOI, other variant, individual spike mutation, or combination of spike mutations. Selecting items from multiple dropdown boxes restricts the output to the data specified by the combination of dropdown items. Fig 3 shows the output returned by selecting E484K from the “Variants / Mutations” dropdown. Fig 4 shows the output returned by selecting BNT162b from the “Plasma / mAbs” dropdown and the Delta variant from the “Variant / Mutations” dropdown.

Fig 3. Query interface, sample query, and outline of query results for an example query for the SARS-CoV-2 RBD mutation E484K. The query interface containing three dropdown boxes is shown at the upper left (A). E484K is selected from the “Variants / Mutations” dropdown box. The upper right summarizes the data returned by the query, which in this case includes 244 mAb neutralizing susceptibility results from 28 publications, 232 convalescent plasma (CP) results from 15 publications, and 214 vaccinee plasma (VP) results from 11 publications (B). The summary distinguishes between results for which only aggregate (mean or median) data are provided and results for which individual data are provided. The sections below show the headers and first row of the tables containing mAb (C), VP (D), and CP (E) susceptibility results. The complete contents of these tables can be found on the web (<https://covdb.stanford.edu/search-drdb/?host=human&mutations=S%3A484K>).

Fig 4. Query interface, sample query, and outline of query results for an example query for the susceptibility of the SARS-CoV-2 Delta variant to the vaccinee plasma (VP) of persons who received the BNT162b vaccine. The query interface containing three dropdown boxes is shown at the upper left (A). BNT162b is selected from the “Plasma / mAbs” dropdown box, and the Delta variant is selected from the “Variants / Mutations” dropdown box. The upper right summarizes the data returned by the query, which in this case includes 875 results from 12 publications (B). The summary distinguishes between results for which only aggregate (mean or median) data are provided and results for which individual data are provided. The section below the header shows the header and first few rows of the table entitled “Vaccinee Plasma Susceptibility Data”. The complete contents of this tables can be found on the web (<https://covdb.stanford.edu/search-drdb/?host=human&vaccine=BNT162b2&variant=Delta>).

Query results

The mAb, CP, and VP susceptibility query result tables contain between 10 to 13 column headers (Fig 3C-3E). The mAb table contains column headers indicating the reference and location of the data within each reference (e.g., figure or table); assay type (e.g., pseudotyped virus); mAb tested; variant tested; IC₅₀ in ng/ml; and fold-reduced susceptibility compared with a control virus (that is also present in the table; Fig 3C).

The VP table contains column headers that indicate the reference and location of the data within the reference, type of assay used, vaccine received, number of immunizations, number of months since immunization, whether the sample was obtained from a vaccinated person with confirmed prior infection, variant tested, geometric mean neutralizing titer, and median fold reduction in titer compared with the control virus (Fig 3D and Fig 4C).

The CP table contains column headers that indicate the reference and location of the data within the reference, assay type, lineage of the virus that infected the person from whom the CP was obtained, number of months between infection and the plasma sample, variant tested, geometric mean neutralizing titer; and median fold reduction in titer compared with the control virus (Fig 3E).

Each table also contains two additional columns: “# Results” and “Data Availability”. The number of results indicates the number of neutralizing experiments. The data availability contains a “√” if individual data are available or an “X” if data are only available in aggregate (i.e., as a geometric mean titer or a median fold reduction in susceptibility). Clicking on the spreadsheet icon in “Data Availability” column copies the contents of that study to the user’s clipboard. The “Download CSV” tab at the top of each table allows users to download all rows for analysis.

Interacting with the query results

Each table can be considered to have multiple dimensions represented by the table’s column headers. For example, Fig 5A indicates the table dimensions for the query output with BNT162b selected

from the “Plasma Abs / mAb” dropdown and the Delta variant selected from the “Variants / Mutations” dropdown. The table contains 22 rows (i.e., experimental conditions) summarizing 885 results obtained from 13 references. The table dimension check boxes enable users to aggregate the data in the table by deselecting one or more dimensions. For example, if the user deselects the “Reference”, “Assay”, and “Control” variant dimensions, the data are summarized with six rows that provide the median fold-reduction in susceptibility to BNT162b-associated VP for the Delta variant according to the number of vaccinations received, the time since vaccination, and whether the VP was obtained from a previously vaccinated person (Fig 5B).

Fig 5. Aggregating the query results for the example shown in Fig 4. The top part of the figure shows the first three of 22 experimental conditions defined by the column headers in which BNT162b-associated VP was tested for activity against the Delta variant (A). The results for each experimental condition are shown in different rows each containing the geometric mean of the neutralizing antibody titer and the median fold-reduction in titer compared with a control virus. The bottom part of the figure shows that the 22 rows can be displayed using 6 rows by aggregating those results obtained from different references using different assays or different control viruses by deselecting the “Reference”, “Assay”, and “Control” variant (checkboxes within red ovals). The neutralization data are now aggregated according to the number of BNT162b immunizations, time since immunization, and whether the VP was obtained from a person who had been infected prior to vaccination (shaded area superimposed on the first four columns). The number of experimental results in each row in Fig 5B is increased because each row may now contain data from more than one reference.

Sequence analysis program

The CoV-RDB website contains a SARS-CoV-2 sequence analysis program that leverages the same code base written for the widely used Stanford HIV Drug Resistance Database interpretation program [9,10]. The CoV-RDB sequence analysis program supports three types of input: (i) a list of spike mutations; (ii) one or more consensus FASTA sequences containing any part of the SARS-CoV-2 genome; and (iii) one or more codon frequency table (CodFreq) files containing the following seven columns: gene, amino acid position, number of reads at that position, codon, amino acid encoded by the codon, number of reads for that codon, and proportion of reads for that codon. An auxiliary program provided through the website or via download enables users to convert next generation sequencing (NGS) FASTQ files into human and machine-readable CodFreq files that are much faster to analyze. CodFreq files make it possible to estimate the extent of background noise resulting from sequencing or PCR artifact (S3 Fig) [10].

If one or more spike mutations is submitted, the CoV-RDB sequence analysis program reports information on specific mutations and generates summary tables containing the susceptibility of viruses with these mutations to mAbs, CP, and VP (S4 Fig). If a FASTA sequence is submitted, the program returns comments, summary tables, a list of the SARS-CoV-2 genes in the submitted sequence, a list of each of the amino acid mutations in each virus gene, and the sequence's PANGO lineage (S4 Fig). If a codon frequency table is submitted, the program returns all of the above information and also summarizes the read coverage for each position along the genome and provides users with options to select read depth and mutation detection thresholds below which mutations will not be reported (S4 Fig). If multiple sequences are submitted, users have the option of obtaining all the results in a single downloadable CSV file.

Fig 6A-6C show three parts of the output generated when either a FASTQ sequence or CodFreq file is submitted to the sequence analysis program: sequence summary (Fig 6A), sequence quality assessment (Fig 6B), and mutation list (Fig 6C). The mutation comments and neutralization susceptibility data, which are also included in the output, are not shown. The sequence summary section (Fig 6A) lists the genes that underwent sequencing, the median read depth, and the Pango lineage. This section also contains three dropdown boxes that help users select the appropriate threshold for identifying sub-consensus mutations. The read depth and mutation detection thresholds are used to select the minimum number and proportion of reads required for a mutation to be considered viral in origin rather than a sequence or PCR artifact. The nucleotide mixture threshold allows users to select a threshold which minimizes the number of nucleotide ambiguities present in the sequence.

Fig 6. SARS-CoV-2 sequence analysis program output for FASTQ sequences or codon frequency tables. The sequence summary section (A) lists the genes that were sequenced, the median read depth, and the Pango lineage. This section also contains dropdown boxes that enable the user to select the minimum number and proportion of reads, respectively, required to identify a mutation. The threshold that minimizes the proportion of positions with nucleotide ambiguities can also be selected. The sequence quality assessment section displays the read depth across the genome and lists only those amino acid mutations that meet the user-defined criteria outlined in the sequence summary section (B). The mutation list section lists those amino acid mutations that meet the user-defined criteria and shows the proportion of reads containing the mutation (C). The output shown in this figure can be regenerated by loading the example file B.1.1.7 (ERR5026962) at this URL: <https://covdb.stanford.edu/sierra/sars2/by-reads/>.

Discussion

Because SARS-CoV-2 is a rapidly evolving RNA virus infecting millions of people worldwide, new variants will continue to emerge and influence regional and global pandemic trajectories. Ongoing epidemiological and clinical studies are needed to understand the risk of re-infection and vaccine failure posed by different variants. However, because the spectrum of SARS-CoV-2 variants is expanding and shifting faster than epidemiological studies can be conducted, laboratory-based markers will increasingly be used to identify those variants that could predispose to re-infection and vaccine failure. SARS-CoV-2 neutralizing antibody titers are correlates of immune protection and are likely to be useful as an endpoint in vaccine trials and to determine the need for vaccine boosters and immunogen updates.

CoV-RDB is the only database to comprehensively curate published data on the neutralizing susceptibility of SARS-CoV-2 variants and spike mutations to mAbs, CP, and VP. Although non-neutralizing Abs and cellular immune responses also contribute to protection from infection, the presence of neutralizing antibodies targeting the spike protein has correlated most strongly with protection from infection in animal models and in previously infected and vaccinated persons [11–14]. Neutralizing susceptibility assays are also better suited to standardization compared with assays of other immunological defense mechanisms.

CoV-RDB can be downloaded in its entirety without restrictions. This is accomplished using a dual database pipeline that combines the full-fledged Postgres database system to enforce relational data integrity and the simplicity of the SQLite database system to enable users to download and query the database without the overhead of accessing a host server. By making the database fully available to all users, we aim to encourage data sharing and the editing of the underlying CSV files by the authors of published studies.

Cov-RDB neutralizing susceptibility query output can be considered multidimensional tables in which the rightmost columns contain numerical results (e.g., titers and fold-reductions in susceptibility) while the leftmost columns contain experimental conditions. The experimental conditions are explanatory

variables that either directly influence neutralizing susceptibility (e.g., specific vaccine, time since vaccination, and SARS-CoV-2 variant) or have a more subtle effect on susceptibility (e.g., type of neutralizing assay and control virus). The CoV-RDB query interface enables users to explore the query results at different levels of granularity by filtering or aggregating them according to one or more experimental conditions without making additional calls to the web server.

The sequence analysis program shares many features with the Sierra HIV Drug Resistance Database sequence analysis program. For the analysis of NGS data, both programs use Minimap2 [15] to align individual reads to a reference sequence resulting in SAM/BAM files and SAM2CodFreq, a program we wrote to create CodFreq files. The CodFreq format has several advantages over the commonly used variant call format (VCF) because it can be interpreted without a reference sequence and can be used independently from the accompanying SAM file. CodFreq files have a simple tabular format enabling them to be viewed and manipulated using a spreadsheet. The sequence analysis program uses the CodFreq files to assist users determine the appropriate threshold for distinguishing background sequence artifact from authentic sub-consensus amino acids.

Study limitations

Neutralizing susceptibility data are highly heterogeneous often resulting in discordant results across studies [16,17]. There are three main sources for this heterogeneity. First, the composition of neutralizing antibodies among previously infected and vaccinated individuals is heterogeneous [6,18]. Second, there are different types of neutralizing assays including those performed in cell culture using pseudotyped viruses, chimeric viruses, recombinant viruses, and clinical isolates and, more recently, surrogate neutralizing assays that assess the ability of antibodies to block the interaction between the SARS-CoV-2 RBD and ACE2. Third, results for the same sample against a virus variant can differ even among laboratories using the same type of assay as a result of differences in virus inoculum size, the cells used for culture, and viral replication endpoints [16,17,19]. As neutralizing assays become more standardized and as external controls

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

Data on the clinical significance of SARS-CoV-2 neutralizing susceptibility is continually evolving [11,21–26]. The utility of neutralizing antibodies as a correlate of immune protection is ultimately determined in epidemiological studies. Therefore, a database devoted to protective immunity should ideally contain both laboratory and epidemiological data. The main obstacle to expanding CoV-RDB to also include epidemiological studies of vaccine efficacy is that such studies are much more complex than those reporting *in vitro* neutralizing data. For example, vaccine efficacy data depends not just on the vaccine, the variant, and the time since vaccination but also on the study design and the age and immune status of the study population. Moreover, in many vaccine efficacy studies, the proportion of individuals infected with different variants is not known.

Future directions

Although the CoV-RDB is centered around just four main entities (references, viruses, antibodies, and experiments), differences among the types of viruses, types of antibody preparations, and types of experiments has necessitated a sophisticated database design. Nonetheless, as new types of experiments are being published, we have been expanding the database schema. For example, an increasing number of studies report the results of surrogate neutralizing assays based on antibody-mediated blocking of the interaction of RBD with ACE2 [19,27,28]. These assays appear to correlate strongly with plasmid and conventional neutralization test results, although they do not assess the effects of NTD-binding and other non-RBD antibodies that synergistically inhibit virus replication [19]. In addition, the use of an international external standard for calibration such as the one developed by the WHO will increase concordance across different assays and will be reported using international units rather than as an IC_{50} (for mAbs) or a plasma dilution [20,29].

We have also added the comprehensive deep mutational scanning data published by the Bloom laboratory to the database [8,30–34]. The data from these studies are displayed only for those mAbs in advanced clinical development and those mutations that have been reported to occur at a frequency above 0.001%. In addition to reporting the escape fraction associated with a mutation to an mAb, we report the level of protein expression within yeast of RBDs containing the mutation (a measure of protein stability) and the ACE2 binding of RBDs containing the mutation as mutations that bind poorly to ACE2 are less likely to be selected *in vivo*. Although binding data have been reported for many other mAbs using enzyme linked immunoassays, surface plasmon resonance, and biolayer interferometry, we have not curated these data as they have not been as comprehensive as the deep mutational scanning data.

Commercial total binding assays do not differentiate between binding and neutralizing antibodies. They also do not measure binding or neutralization of multiple variants but rather assess binding to pre-variant spike proteins. However, total binding assays often display moderately strong correlations with neutralizing assays [29,35,36] and activity against specific variants may eventually be assessed using variant-specific reagents. We may eventually add such data to CoV-RDB if they will provide insights that cannot be obtained solely from neutralizing antibody studies.

Although CoV-RDB contains neutralizing susceptibility data obtained using the plasma from infected animal experiments (e.g., non-human primates, hamsters, and mice), we have not included data from animal model challenge studies as such studies would require extensive modifications to our current database schema. Therefore, we will continue to monitor these studies and consider adding top-line data from these studies to alert database users to the existence of these studies without rigorously representing study details. Finally, a similar approach will be considered for studies of vaccine efficacy. We are therefore exploring the possibility of adding the top-line data of these studies so that the findings of these studies can be correlated with the *in vitro* neutralizing data in CoV-RDB.

References

1. Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet.* 2021. doi:10.1038/s41576-021-00408-x
2. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020;367: 1260–1263. doi:10.1126/science.abb2507
3. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veasler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020;181: 281-292.e6. doi:10.1016/j.cell.2020.02.058
4. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581: 215–220. doi:10.1038/s41586-020-2180-5
5. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature.* 2020;581: 221–224. doi:10.1038/s41586-020-2179-y
6. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe.* 2021;29: 463-476.e6. doi:10.1016/j.chom.2021.02.003
7. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol.* 2021;19: 409–424. 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

22. Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta K-D, et al. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nat Med.* 2021; 1–9. doi:10.1038/s41591-021-01548-7
23. Clemens SAC, Folegatti PM, Emary KRW, Weckx LY, Ratcliff J, Bibi S, et al. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 lineages circulating in Brazil. *Nat Commun.* 2021;12: 5861. doi:10.1038/s41467-021-25982-w
24. Levin EG, Lustig Y, Cohen C, Fluss R, Indenbaum V, Amit S, et al. Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. *New England Journal of Medicine.* 2021;0: null. doi:10.1056/NEJMoa2114583
25. Bruxvoort K, Sy LS, Qian L, Ackerson BK, Luo Y, Lee GS, et al. Effectiveness of mRNA-1273 against Delta, Mu, and other emerging variants. 2021 Oct p. 2021.09.29.21264199. doi:10.1101/2021.09.29.21264199
26. Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat Med.* 2021; 1–9. doi:10.1038/s41591-021-01540-1
27. Abe KT, Li Z, Samson R, Samavarchi-Tehrani P, Valcourt EJ, Wood H, et al. A simple protein-based surrogate neutralization assay for SARS-CoV-2. *JCI Insight.* 2020;5: e142362. doi:10.1172/jci.insight.142362
28. Tan CW, Chia WN, Qin X, Liu P, Chen MI-C, Tiu C, et al. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction. *Nat Biotechnol.* 2020;38: 1073–1078. doi:10.1038/s41587-020-0631-z

29. Kung Y-A, Huang C-G, Huang S-Y, Liu K-T, Huang P-N, Yu K-Y, et al. Antibody titers measured by commercial assays are correlated with neutralizing antibody titers calibrated by international standards. medRxiv. 2021; 2021.07.16.21260618. doi:10.1101/2021.07.16.21260618
30. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science. 2021;371: 850–854. doi:10.1126/science.abf9302
31. Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Reports Medicine. 2021;2: 100255. doi:10.1016/j.xcrm.2021.100255
32. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host Microbe. 2021;29: 44-57.e9. doi:10.1016/j.chom.2020.11.007
33. Greaney AJ, Loes AN, Gentles LE, Crawford KHD, Starr TN, Malone KD, et al. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. Science Translational Medicine. 2021;13: eabi9915. doi:10.1126/scitranslmed.abi9915
34. Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun. 2021;12: 4196. doi:10.1038/s41467-021-24435-8
35. Peluso MJ, Takahashi S, Hakim J, Kelly JD, Torres L, Iyer NS, et al. SARS-CoV-2 antibody magnitude and detectability are driven by disease severity, timing, and assay. Science Advances. 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

S1 Table. Slightly different patterns of spike mutations within each of the SARS-CoV-2 variants of concern (VOCs) and variants of interest (VOIs).

S1 Fig. Data management workflow for CoV-RDB. Weekly incremental searches of PubMed and preprint servers (BioRxiv/MedRxiv and Research Square) are performed. Publications that appear to have data pertinent to SARS-CoV-2 variants and their susceptibility to mAbs, convalescent plasma (CP), and vaccinee plasma (VP) are downloaded to a Zotero reference database folder to enable full-text review and data curation. Extracted data are exported into a set of linked CSV files in an open-source Github repository (<https://github.com/hivdb/covid-drdb-payload>). Extracted data are then imported into a Postgres database where the data are validated for completeness and consistency before being exported as a single SQLite database file that serves as the back end for the CoV-RDB website and is available to users for download.

S2 Fig. Effects of spike receptor binding domain (RBD) mutations on monoclonal antibody (mAb) binding and neutralization. The X-axis shows the escape fraction as determined using a deep mutational scanning (DMS) assay that measures the binding of various mAbs to RBD variants produced in yeast. The Y-axis shows the results of in vitro neutralization using the same mAbs and spike proteins with the same RBD mutations as those used in the DMS assay. Discordant results are shown in the figure. Abbreviations: SOT: sotrovimab; BAM: bamlanivimab; ETE: etesevimab; CAS: casirivimab; IMD: imdevimab.

S3 Fig. Creating a codon frequency file from a FASTQ file. FASTQ files are aligned to the consensus Wuhan reference sequence using the Minimap2 alignment program. The resulting BAM/SAM files are then processed by a program SAM2CodFreq that we wrote to generate a codon frequency file containing seven columns as shown on the right. The table here shows the results from three codons (spike positions 500 to

502). The observation that many codons shown in this (and other parts of the same file which are not shown) are present at levels between 0.2% and about 2% suggests that codons present at these low proportions likely represent sequencing or PCR artifact (i.e., “background noise”). However, as the mutation N501Y occurs at a considerably higher proportion (34.3%), it is likely to be present in the infecting virus population.

S4 Fig. Functions of the SARS-CoV-2 sequence analysis program. The program supports three types of input: a list of spike mutations; one or more consensus FASTA sequences containing any part of the SARS-CoV-2 genome; and one or more FASTQ sequences. However, because a FASTQ sequence can take several minutes to analyze, users are advised to first convert them to a codon frequency file through an auxiliary program. If a list of spike mutations is submitted, the program returns comments about notable mutations and summary tables reporting the susceptibility of viruses with these mutations to mAbs, CP, and VP. If a FASTA sequence is submitted, the program returns the preceding information plus a list of the SARS-CoV-2 genes, the amino acid mutations in the sequence, and the sequence’s PANGO lineage. If a FASTQ sequence or codon frequency table is submitted, the program provides the preceding information and the read coverage for each position along the genome. It also provides users with the options to select read depth and mutation detection thresholds below which mutations will not be reported.

BAM

ETE

CAS

IMD

CIL

TIX

SOT

Fig 1

Fig 2

A

References

Any

Plasma Abs / mAbs

Any

Variants / Mutations

E484K

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:

Individual sample data are available:244 results, 28 publications

Convalescent plasma - Individual mutation:

Individual sample data are available:232 results, 15 publications

Only aggregate sample data are available:202 results, 14 publications

Vaccinee plasma - Individual mutation:

Individual sample data are available:214 results, 11 publications

Only aggregate sample data are available:204 results, 10 publications

C

MAb Susceptibility Data

bioRxiv preprint doi: <https://doi.org/10.1101/2021.11.24.469823>; this version posted November 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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	<div></div> ✓

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)	<div></div> ✓

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	<div></div> ✓

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:

Individual sample data are available:

Only aggregate sample data are available:

875 results, 12 publications

802 results, 9 publications

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 $\pm$ GSD	Fold Reduction Median (IQR)	Data Availability
Edara, VV (2021c)	Virus isolate	Supplementary Table S6	None	BNT162b2	2	1	Human	A (WA1) <small>NT50: 777</small>	Delta (B.1.617.2)	10	235 $\pm$ 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 $\pm$ 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 $\pm$ 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 $\pm$ 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 $\pm$ GSD	Fold Reduction Median (IQR)	Data Availability
None	BNT162b2	2	1	Delta (B.1.617.2)	12	395	277 $\pm$ 3.3	2.9 (2.0-5.6)
None	BNT162b2	2	2-6	Delta (B.1.617.2)	4	127	145 $\pm$ 3.5	4.8 (3.3-6.7)
None	BNT162b2	1	1	Delta (B.1.617.2)	2	162	19 $\pm$ 4.6	9.6 (4.6-13.4)
None	BNT162b2	1	2-6	Delta (B.1.617.2)	1	73	26 $\pm$ 5.2	8.8 (5.0-12.8)
Wild Type	BNT162b2	1	1	Delta (B.1.617.2)	1	21	2,096 $\pm$ 2.3	1.7 (1.3-1.9)

Fig 5

A

Sequence summary

Sequence includes following genes:

nsp1 (missing: 1-19) • nsp2 • PLpro • nsp4 (missing: 240-308) • 3CLpro • nsp6 • nsp7 • nsp8 • nsp9 • nsp10 • RdRP • nsp13 • nsp14 • nsp15
• nsp16 • Spike • ORF3a • E • M • ORF6 • ORF7a • ORF7b • ORF8 • N • ORF10

[Download consensus sequence](#)

Median read depth:

1,232

PANGO lineage:

B.1.1.7 (Prob=1.0; pangolin: 3.1.11; pangoLEARN: 2021-08-24)

Spike Variant:

[+](#) Alpha (0.00%)

Outbreak.info:

B.1.1.7 (n=1,078,651)

Nucleotide mixture threshold:

 (actual: 0.024%)

Mutation detection threshold:

 (actual: ≥14%)

Read depth threshold by codon:

B

Sequence quality assessment

C

Mutation list

PLpro: T183I 100% cov=304 • C285CG C: 75%, G: 25% cov=500 • A890D 99% cov=192 • I1412T 99% cov=841

nsp6: Δ106-108 100% cov=1,631

RdRP: P323L 99% cov=6,207

nsp13: K460R 99% cov=2,088

Spike: Δ69-70 100% cov=1,573 • Δ144 100% cov=2,546 • N501Y 99% cov=1,231 • A570D 100% cov=4,770 • D614G 100% cov=4,766 • P681H 100% cov=1,132

• T716I 100% cov=1,131 • S982A 99% cov=318 • D1118H 99% cov=9,114 • W1214WG W: 81%, G: 19% cov=473

ORF8: Q27* 99% cov=422 • R52I 100% cov=420 • Y73C 100% cov=3,548

N: D3L 98% cov=10,118 • R203K 97% cov=470 • G204R 99% cov=470 • S235F 100% cov=233

Fig 6