Published online 13 October 2021 *Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D771–D776***
https://doi.org/10.1093/nar/gkab895

ESC: a comprehensive resource for SARS-CoV-2 immune escape variants

Mercy Rophina [1,2], Kavita Pandhare [1,2], Afra Shamnath [1], Mohamed Imran [1,2], Bani Jolly [1,2] and Vinod Scaria [1,2,*]

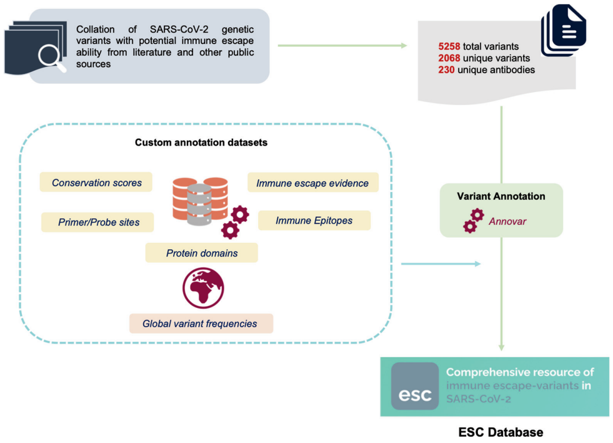
[1] CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, New Delhi, India and [2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Received August 15, 2021; Revised September 14, 2021; Editorial Decision September 20, 2021; Accepted October 10, 2021

ABSTRACT

Ever since the breakout of COVID-19 disease, ceaseless genomic research to inspect the epidemiology and evolution of the pathogen has been undertaken globally. Large scale viral genome sequencing and analysis have uncovered the functional impact of numerous genetic variants in disease pathogenesis and transmission. Emerging evidence of mutations in spike protein domains escaping antibody neutralization is reported. We have built a database with precise collation of manually curated variants in SARS-CoV-2 from literature with potential escape mechanisms from a range of neutralizing antibodies. This comprehensive repository encompasses a total of 5258 variants accounting for 2068 unique variants tested against 230 antibodies, patient convalescent plasma and vaccine breakthrough events. This resource enables the user to gain access to an extensive annotation of SARS-CoV-2 escape variants which would contribute to exploring and understanding the underlying mechanisms of immune response against the pathogen. The resource is available at http://clingen.igib.res.in/esc/.

GRAPHICAL ABSTRACT


Collation of SARS-CoV-2 genetic variants with potential immune escape ability from literature and other public sources
5258 total variants
2068 unique variants
230 unique antibodies
Custom annotation datasets
Conservation scores
Immune escape evidence
Primer/Probe sites
Immune Epitopes
Protein domains
Global variant frequencies
Variant Annotation
Annovar
esc
Comprehensive resource of immune escape-variants in SARS-CoV-2
ESC Database

INTRODUCTION

Genomic approaches have been instrumental in understanding the origin and evolution of SARS-CoV-2, the causative agent for the COVID-19 pandemic (1). Availability of the genome sequence of one of the earliest SARS-CoV-2 genomes from Wuhan province (2) and high throughput approaches to resequence and analyse viral genomes have facilitated the availability of numerous open genomic data sharing initiatives by the researchers worldwide. Pioneering public sources like GenBank (3) and Global Initiative on Sharing all Influenza Data (GISAID) (4) provide access to systematically organized genomes of SARS-CoV-2. The China National GeneBank DataBase (CNGBdb) (5), Genome Warehouse (GWH) (6) and Virus Pathogen Resource (ViPR) (7) are few other resources which provide access to viral genomes and perform analyses on phylogeny, sequence similarity and genomic variants.

There has been a significant interest in recent times in understanding the functional impact of genetic variants in SARS-CoV-2 apart from exploring the genetic epidemiology. The variant D614G present in spike protein has been one the earliest and prominent examples with

[*] To whom correspondence should be addressed. Tel: +91 9650466002; Fax: +91 11 27667471; Email: vinods@igib.in

potential implications associated with the infectivity of the virus (8). Studies explaining the possible impact of SARS-CoV-2 variants in diagnostic primers and probes have augmented the importance of analysing the variations and their underlying role in disease pathogenesis (9). Various resources have been made available to help comprehend the virus better and also to understand its evolution. Public sources exclusively documenting functionally relevant SARS-CoV-2 variants based on literature evidence are also available (10).

With the advent of therapies including monoclonal antibodies, convalescent plasma as well as the recent availability of vaccines, interest in genetic variants which could affect the efficacy of such modalities of therapy has accelerated. The targeting of spike proteins by broad-neutralizing antibodies against SARS-CoV-2 offers a potential means of treating and preventing further infections of COVID-19 (11). Evidence on immunodominant epitopes with significantly higher response rates have also been reported (12). Antibody response to SARS-CoV-2 is one of the key immune responses which is actively being pursued to develop therapeutic strategies as well as vaccines (13). The recent months have seen enormous research into the structural and molecular architecture of the interactions between the spike protein in SARS-CoV-2 and antibodies. Studies have also provided insights into the genetic variants which could confer partial or complete resistance to antibodies (14) as well as panels of convalescent plasma. With vaccines being widely available, the evidence on the effect of genetic variants on efficacy of vaccines is also emerging (15)

The lack of a systematic effort to compile genetic variants in SARS-CoV-2 associated with immune escape motivated us to compile the information in a relevant, searchable and accessible format. Towards this goal, we systematically evaluated publications for evidence on immune escape associated with genetic variants in SARS-CoV-2 and created a database named as ESC. User-friendly web interface is made available to retrieve information on immune escape variants as well as their extensive functional annotations. To the best of our knowledge, this is the first most comprehensive resource for immune escape variants for SARS-CoV-2. The resource can be accessed online at http://clingen.igib.res.in/esc/.

MATERIALS AND METHODS

Data and search strategy

Genetic variants in the SARS-CoV-2 genome and evidence suggesting association with immune escape were systematically catalogued. A significant number of variants were associated with escape or resistance to a range of neutralizing and monoclonal antibodies, while a subset was associated with resistance to convalescent plasma. The data was compiled by manual curation of literature available from peer-reviewed publications and preprints. Literature reports with relevant information on antibody escape variants were retrieved from sources including PubMed, LitCovid, Google Scholar and preprint servers. The reports were systematically checked for details pertaining to the variation, antibodies tested and experimental methods followed in the study. In addition, the variants were systematically categorized based on experimental validation and computational prediction. Collated data was organized in a pre-formatted template based on their protein positions. This comprehensive compendium was used for further functional annotations.

Variant information and annotations

The variant information and annotations were retrieved from annotation tables for individual features using ANNOVAR (16). Variant annotations broadly included genic features like the variant type and functional annotations related to deleteriousness and evolutionary conservation. Information on protein domains and immune epitopes was compiled and customized from various public sources. Variant sites reported to be potentially problematic including homoplasic regions, sites with recurrent sequencing errors and hypermutable sites were also labelled to enable quality check of the mutation site. Variants mapping back to sites of potential SARS-CoV-2 diagnostic primers/probes were also annotated.

Compilation of B-cell and T-cell epitope data

Details on B-cell and T-cell epitopes spanning the protein residues of SARS-CoV-2 were retrieved from Immune Epitope Database and Analysis Resource (IEDB) (17). All epitopes of SARS-CoV-2 (IEDB ID: 2697049) against human hosts with reported positive or negative assays and any type of MHC restriction were used for analysis. Epitope information pertaining to each amino acid residue including the epitope type (linear/discontinuous), epitope sequence with corresponding start and end positions and IEDB identifiers were systematically mapped back and documented.

Antibody details and annotation

Information pertaining to the list of antibodies associated with escape mechanisms was retrieved from available public sources. Compiled antibodies were systematically mapped back to the AntiBodies Chemically Defined (ABCD) database which provides integrated information regarding the antibodies along with its corresponding antigens and protein cross links to fetch unique antibody identifiers (18,19).

Database and web interface

The back-end of the web interface was implemented using Apache web server and MongoDB v3.4.10 in order to provide a user-friendly interface for variant search. The JavaScript Object Notation (JSON) file format was used to systematically store the data. PHP 7.0, AngularJS, HTML, Bootstrap 4 and CSS were used to code the web interface for querying. Highcharts javascript library was also used for improved presentation and interactivity. A Beacon API has been created using the PHP programming language. The ESC Beacon API v1.0.0 is a read-only API with specifications written in OpenAPI. It uses JSON in requests and responses and standard HTTPS for information transfer. The

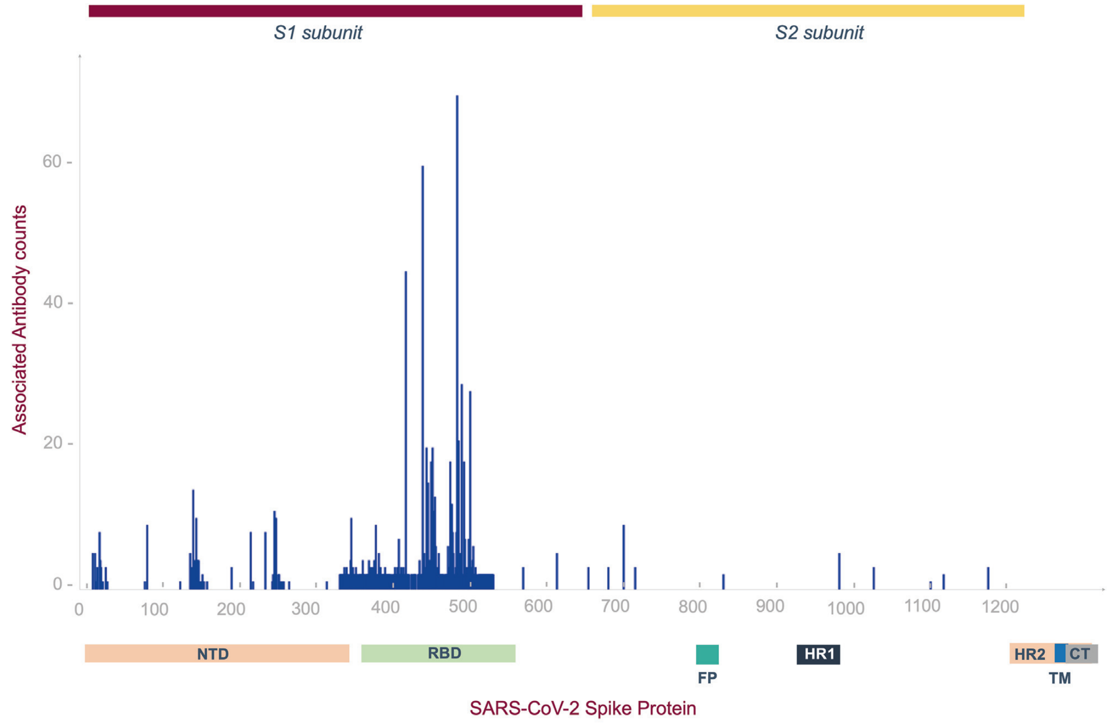


Figure 1. Illustration of mutation: antibody associations along the spike protein residues. Number of unique antibodies associated with potential immune evasion at each spike protein residue is marked along with domain annotations.

Beacon API has one endpoint: /beacon?Variant; query interface is provided by the beacon endpoint. Contents of the database gets updated every month by systematic literature curations and also made available to users for bulk download.

RESULTS AND DISCUSSION

Repository of SARS-CoV-2 escape variants

We compiled a total of 5258 variant entries from over 60 articles which studied SARS-CoV-2 variants and their effect on immune escape. This included a total of 2068 unique variants mapping to spike protein, ORF1ab and ORF3a. Out of the total unique variants, 2060 variants mapped to the gene coding for spike protein with potential immune escape mechanisms elucidated through experimental evidence as well as computational predictions. The remaining eight variants were found in ORF1ab and ORF3a genes, out of which, three were reported to confer potential epitope loss. The compiled list of variants was found associated with 230 unique SARS-CoV-2 antibodies and patient polyclonal sera. A handful of SARS-CoV-2 variations associated with vaccine breakthrough events have also been documented. A brief comparison of the curations in the ESC database with other publicly available resources is summarized in Supplementary Figure S1 and Supplementary Table S1a and b. Functional consequences of the variants were mapped from a total of 22 unique custom generated annotation datasets precisely including deleteriousness and conservation score predictions, protein domains and immune epitopes using ANNOVAR. The data features used in the study are summarised in Supplementary Table S2.

Antibody association mapping

By scanning through the spike protein residues and their associations with SARS-CoV-2 neutralizing and monoclonal antibodies, we were able to compile the exact count of antibodies reported to have potential associations with the residues. From our analysis we observed that spike protein residues ranging from 350 to 500 amino acid positions exhibited potential antibody associations with the possibility of immune escape against at least one antibody. A total of 22 hotspot residues (140, 144, 246, 248, 346, 417, 439, 444, 445, 446, 450, 452, 453, 455, 475, 477, 484, 485, 486, 490, 493, 501) were found to possess immune evasion capability against $>$ 10 monoclonal antibodies. A schematic representation of the number of antibodies associated with spike protein residues along with their domain annotations is shown in Figure 1. The cumulative frequencies of spike mutation sites associated with immune escape against $>$5 mAbs in the receptor binding domain is shown in Figure 2. Systematic categorization of mutation residues along with their localization in spike protein is depicted in Figure 3.

Overview of B cell epitopes and immunodominant epitope regions

With the aim of mapping back the known B and T cell epitopes encompassing the variant compendium, SARS-CoV-2 epitope details were extracted. There were a total of 310 and 472 experimentally validated B cell and T cell epitopes

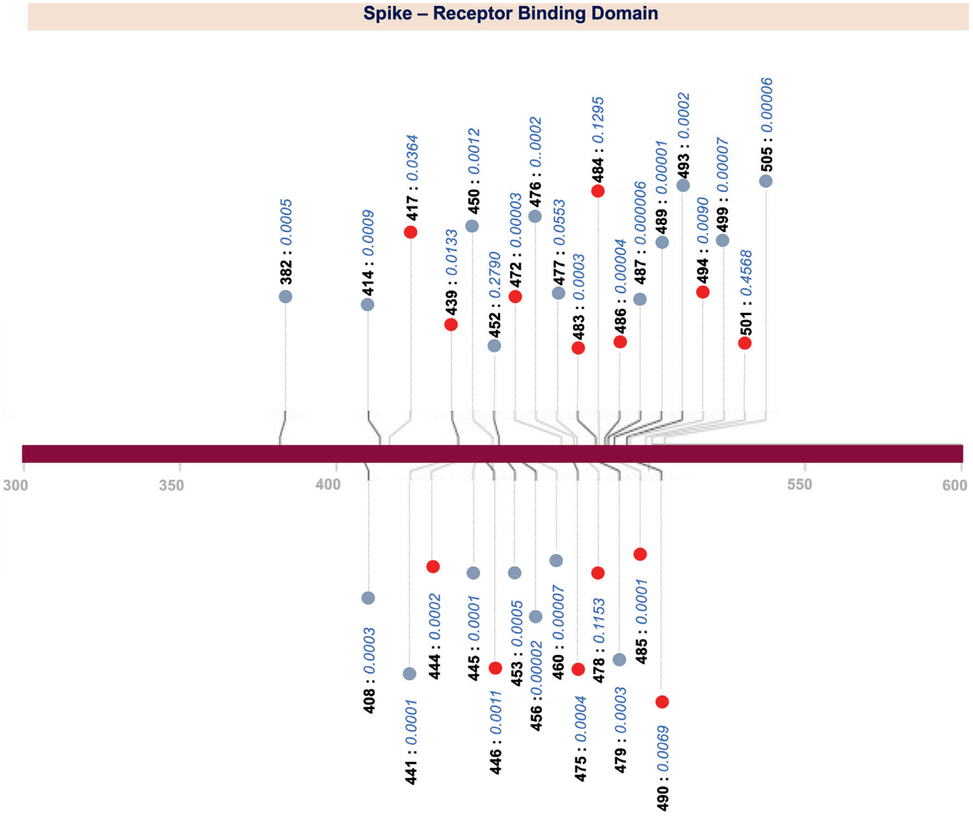


Figure 2. Distribution of mutation sites associated with >5 mAbs in the Receptor binding domain of SARS-CoV-2 spike protein. Variant sites with potential impact on neutralization of human polyclonal sera are represented in red. Cumulative frequencies of variants at RBD sites are mentioned alongside.

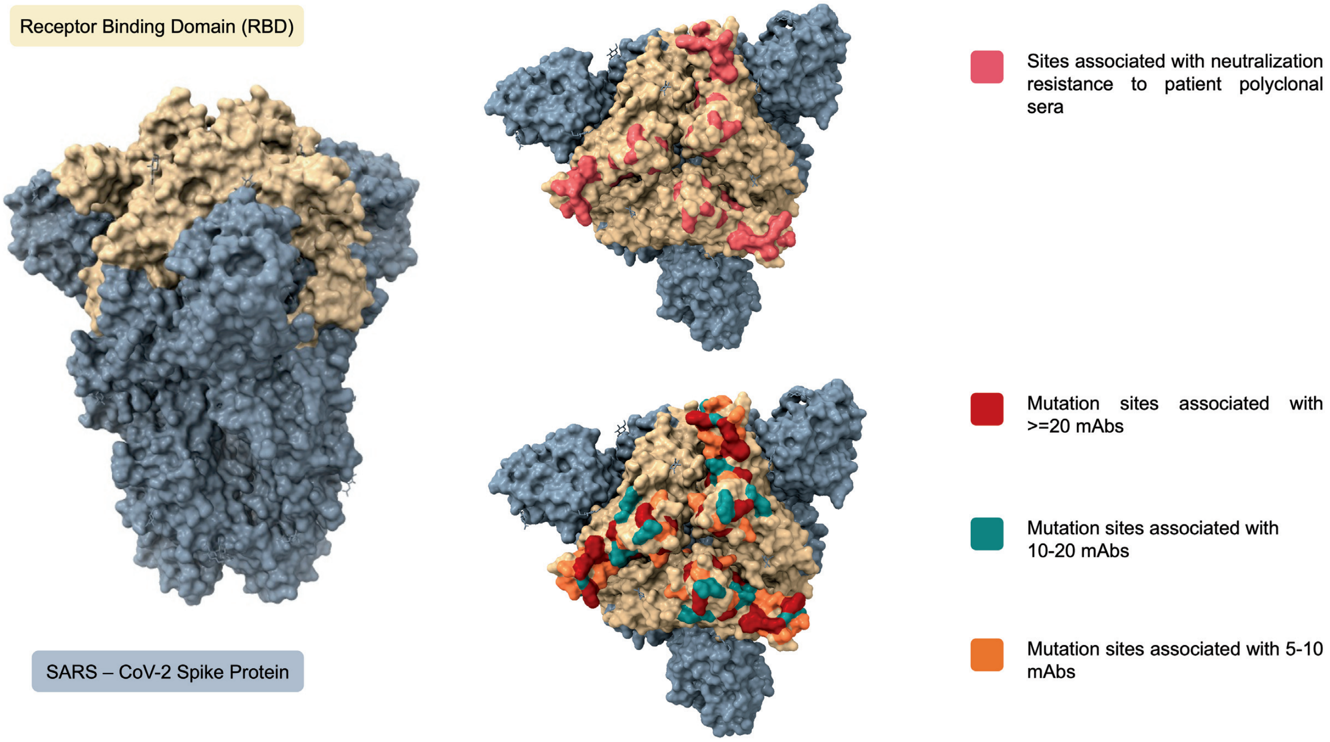
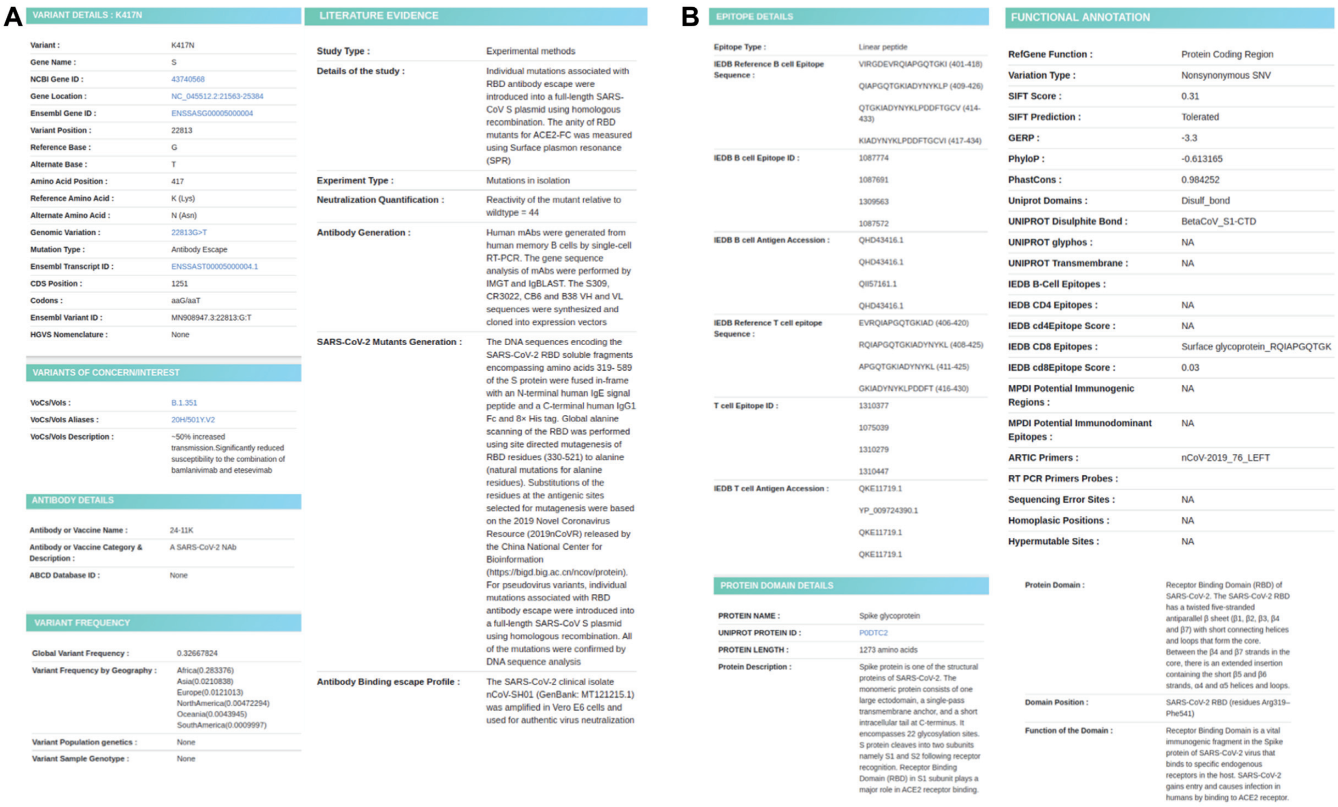

Receptor Binding Domain (RBD)
SARS – CoV-2 Spike Protein
Sites associated with neutralization resistance to patient polyclonal sera
Mutation sites associated with >=20 mAbs
Mutation sites associated with 10-20 mAbs
Mutation sites associated with 5-10 mAbs

Figure 3. Receptor binding domain of spike protein with mutation hotspot sites associated with decreased neutralization against patient polyclonal sera and monoclonal antibodies.



VARIANT DETAILS : K417N

Variant :	K417N
Gene Name :	S
NCBI Gene ID :	43740568
Gene Location :	NC_045512.2:21563-25384
Ensembl Gene ID :	ENSSASG00005000004
Variant Position :	22813
Reference Base :	G
Alternate Base :	T
Amino Acid Position :	417
Reference Amino Acid :	K (Lys)
Alternate Amino Acid :	N (Asn)
Genomic Variation :	22813G>T
Mutation Type :	Antibody Escape
Ensembl Transcript ID :	ENSSAST00005000004.1
CDS Position :	1251
Codons :	aaG/aaT
Ensembl Variant ID :	MN908947.3:22813:G:T
HGVS Nomenclature :	None

VARIANTS OF CONCERN/INTEREST

VoCs/VoIs :	B.1.351
VoCs/VoIs Aliases :	20H/501Y.V2
VoCs/VoIs Description :	~50% increased transmission.Significantly reduced susceptibility to the combination of bamlanivimab and etesevimab

ANTIBODY DETAILS

Antibody or Vaccine Name :	24-11K
Antibody or Vaccine Category & Description :	A SARS-CoV-2 NAb
ABCD Database ID :	None

VARIANT FREQUENCY

Global Variant Frequency :	0.32667824

Variant Frequency by Geography :	Africa(0.283376)
	Asia(0.0210838)
	Europe(0.0121013)
	NorthAmerica(0.00472294)
	Oceania(0.0043945)
	SouthAmerica(0.0009997)

Variant Population genetics :	None
Variant Sample Genotype :	None

LITERATURE EVIDENCE

Study Type :	Experimental methods
Details of the study :	Individual mutations associated with RBD antibody escape were introduced into a full-length SARS-CoV S plasmid using homologous recombination. The anity of RBD mutants for ACE2-FC was measured using Surface plasmon resonance (SPR)
Experiment Type :	Mutations in isolation
Neutralization Quantification :	Reactivity of the mutant relative to wildtype = 44
Antibody Generation :	Human mAbs were generated from human memory B cells by single-cell RT-PCR. The gene sequence analysis of mAbs were performed by IMGT and IgBLAST. The S309, CR3022, CB6 and B38 VH and VL sequences were synthesized and cloned into expression vectors
SARS-CoV-2 Mutants Generation :	The DNA sequences encoding the SARS-CoV-2 RBD soluble fragments encompassing amino acids 319- 589 of the S protein were fused in-frame with an N-terminal human IgE signal peptide and a C-terminal human IgG1 Fc and 8× His tag. Global alanine scanning of the RBD was performed using site directed mutagenesis of RBD residues (330-521) to alanine (natural mutations for alanine residues). Substitutions of the residues at the antigenic sites selected for mutagenesis were based on the 2019 Novel Coronavirus Resource (2019nCoVR) released by the China National Center for Bioinformation (https://bigd.big.ac.cn/ncov/protein). For pseudovirus variants, individual mutations associated with RBD antibody escape were introduced into a full-length SARS-CoV S plasmid using homologous recombination. All of the mutations were confirmed by DNA sequence analysis
Antibody Binding escape Profile :	The SARS-CoV-2 clinical isolate nCoV-SH01 (GenBank: MT121215.1) was amplified in Vero E6 cells and used for authentic virus neutralization

EPITOPE DETAILS

Epitope Type :	Linear peptide
IEDB Reference B cell Epitope Sequence :	VIRGDEVRQIAPGQTGKI (401-418) QIAPGQTGKIADYNYKLP (409-426) QTGKIADYNYKLPDDFTGCV (414-433) KIADYNYKLPDDFTGCVI (417-434)
IEDB B cell Epitope ID :	1087774 1087691 1309563 1087572
IEDB B cell Antigen Accession :	QHD43416.1 QHD43416.1 QII57161.1 QHD43416.1
IEDB Reference T cell epitope Sequence :	EVRQIAPGQTGKIAD (406-420) RQIAPGQTGKIADYNYKL (408-425) APGQTGKIADYNYKL (411-425) GKIADYNYKLPDDFT (416-430)
T cell Epitope ID :	1310377 1075039 1310279 1310447
IEDB T cell Antigen Accession :	QKE11719.1 YP_009724390.1 QKE11719.1 QKE11719.1

PROTEIN DOMAIN DETAILS

PROTEIN NAME :	Spike glycoprotein
UNIPROT PROTEIN ID :	P0DTC2
PROTEIN LENGTH :	1273 amino acids
Protein Description :	Spike protein is one of the structural proteins of SARS-CoV-2. The monomeric protein consists of one large ectodomain, a single-pass transmembrane anchor, and a short intracellular tail at C-terminus. It encompasses 22 glycosylation sites. S protein cleaves into two subunits namely S1 and S2 following receptor recognition. Receptor Binding Domain (RBD) in S1 subunit plays a major role in ACE2 receptor binding.

FUNCTIONAL ANNOTATION

RefGene Function :	Protein Coding Region
Variation Type :	Nonsynonymous SNV
SIFT Score :	0.31
SIFT Prediction :	Tolerated
GERP :	-3.3
PhyloP :	-0.613165
PhastCons :	0.984252
Uniprot Domains :	Disulf_bond
UNIPROT Disulphite Bond :	BetaCoV_S1-CTD
UNIPROT glyphos :	NA
UNIPROT Transmembrane :	NA
IEDB B-Cell Epitopes :	
IEDB CD4 Epitopes :	NA
IEDB cd4Epitope Score :	NA
IEDB CD8 Epitopes :	Surface glycoprotein_RQIAPGQTGK
IEDB cd8Epitope Score :	0.03
MPDI Potential Immunogenic Regions :	NA
MPDI Potential Immunodominant Epitopes :	NA
ARTIC Primers :	nCoV-2019_76_LEFT
RT PCR Primers Probes :	
Sequencing Error Sites :	NA
Homoplasic Positions :	NA
Hypermutable Sites :	NA

Protein Domain :	Receptor Binding Domain (RBD) of SARS-CoV-2. The SARS-CoV-2 RBD has a twisted five-stranded antiparallel β sheet (β1, β2, β3, β4 and β7) with short connecting helices and loops that form the core. Between the β4 and β7 strands in the core, there is an extended insertion containing the short β5 and β6 strands, α4 and α5 helices and loops.
Domain Position :	SARS-CoV-2 RBD (residues Arg319–Phe541)
Function of the Domain :	Receptor Binding Domain is a vital immunogenic fragment in the Spike protein of SARS-CoV-2 virus that binds to specific endogenous receptors in the host. SARS-CoV-2 gains entry and causes infection in humans by binding to ACE2 receptor.

Figure 4. Panel illustrating the query search and display features in ESC database.

respectively. This precisely included 263 linear and 47 discontinuous B cell epitopes in spike protein. Reported B cell and T cell epitope information was mapped back to residues possessing antibody escape mutations, which provided a brief insight on the potential impact of these variations in immune recognition and responses.

Database features

The database offers a user-friendly interface enabling the users to search for variants based on their amino acid change, gene name or the antibody name as per the specified format. The search query returns a list of matching results, whose complete functional annotations can be viewed by clicking on the displayed elements. The resource provides a list of annotation features for each variant precisely organized into eight major sections namely Variant details, Antibody details, Variants of Concern/Interest, Protein domain details, Epitope details, Functional annotation, Literature Evidence and Variant frequency. Figure 4A and B portrays the query search and the result display section of the resource.

Basic details pertaining to the variant like the amino acid change, genomic coordinates and the variant type have been enlisted in the Variant details section. Information on the associated neutralizing antibodies and their identifiers are provided in the Antibody details section. Domain and epitope details section exclusively comprises details on the protein domain, epitopes reported to span the protein residue through experimental validations. Computationally predicted functional annotations on deleteriousness from SIFT (20), evolutionary conservation scores provided by PhastCons (21), GERP (22) and PhyloP (23) are included in the Functional annotation section. This section also enlists protein domain information retrieved from UniProt and immune epitopes documented from IEDB (17,24), UCSC and predictions from different software packages (B cells- BepiPred 2.0, CD4-IEDB Tepitool, CD8-NetMHCpan4). Annotations of potential error prone sites including sites of sequencing errors, homoplasic and hypermutable regions (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473) and diagnostic primer/probe sites are also mapped. Extensive evidence from literature curation including the methods of the study, neutralization quantification profiles and details of antibody/mutant generation are summarized in the literature evidence section. Variant frequency section exclusively summarizes the estimated frequencies of the variant on a global scale as well as by its geography. In addition, characteristic mutations of VoCs and VoIs have also been annotated with brief descriptions in the Variants of Concern/Interest section.

CONCLUSIONS

With evidence emerging on genetic variants in SARS-CoV-2 associated with resistance to monoclonal antibodies and convalescent plasma using *in-vitro* assays, unique insights into the structural and functional mechanisms whereby the pathogen could evolve and evade antibodies have become possible. These insights could have enormous implications in efficacy of vaccines currently being used as well as under trials. One of the recent studies has reported the impact of a few immune escape variants on the efficacy of vaccines (25). It is expected that similar studies would be extended for a wider number of variants as well as vaccines. In order to keep pace with rapid discoveries regarding SARS-CoV-2 escape variations and mechanisms, the database and the associated Github repository is being updated every month from peer reviewed publications and pre-print articles with complete annotations. We therefore foresee that the ESC resource would be a central resource to enable such studies and provide a ready reference to the emerging evidence on immune escape.

DATA AVAILABILITY

The completed data curated from various literature sources are collated and made available for access and bulk download at https://github.com/mercywilliams160896/ESC_COVID19.

An API for ESC is also made available for ease of access to data. Example search: https://clingen.igib.res.in/esc/api/beacon?Variant=A475V. A detailed overview of The ESC Beacon API v1.0.0 has been documented and linked to the webpage duly for user interests.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Authors acknowledge Anjali Bajaj for her constructive comments and suggestions. The funders have no role in the analysis or decision to publish.

Author's contributions: M.R., A.S. and M.I. systematically curated data for the database. K.P. designed the database. B.J. provided annotation details and helped in fact checks. V.S. conceived and designed the project. All authors approved the final manuscript.

FUNDING

Council of Scientific and Industrial Research (CSIR), India through CODEST grant. Funding for open access charge: Council of Scientific and Industrial Research (CSIR), India.

Conflict of interest statement. None declared.

REFERENCES

1. Zhang,Y.-Z. and Holmes,E.C. (2020) A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell.*, **181**, 223–227.
2. Tang,X., Wu,C., Li,X., Song,Y., Yao,X., Wu,X., Duan,Y., Zhang,H., Wang,Y., Qian,Z. *et al.* (2020) On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.*, **7**, 1012–1023.
3. Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2016) GenBank. *Nucleic Acids Res.*, **44**, D67–D72.
4. Shu,Y. and McCauley,J. (2017) GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.*, **22**, 30494.
5. Chen,F.Z., You,L.J., Yang,F., Wang,L.N., Guo,X.Q., Gao,F., Hua,C., Tan,C., Fang,L., Shan,R.Q. *et al.* (2020) CNGBdb: China National GeneBank DataBase. *Yi Chuan*, **42**, 799–809.
6. CNCB-NGDC Members and Partners (2021) Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2021. *Nucleic Acids Res.*, **49**, D18–D28.
7. Pickett,B., Greer,D., Zhang,Y., Stewart,L., Zhou,L., Sun,G., Gu,Z., Kumar,S., Zaremba,S., Larsen,C. *et al.* (2012) Virus Pathogen Database and Analysis Resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses*, **4**, 3209–3226.
8. Korber,B., Fischer,W.M., Gnanakaran,S., Yoon,H., Theiler,J., Abfalterer,W., Hengartner,N., Giorgi,E.E., Bhattacharya,T., Foley,B. *et al.* (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, **182**, 812–827.
9. Wang,R., Hozumi,Y., Yin,C. and Wei,G.-W. (2020) Mutations on COVID-19 diagnostic targets. *Genomics*, **112**, 5204–5213.
10. Tzou,P., Tao,K., Nouhin,J., Rhee,S.-Y., Hu,B., Pai,S., Parkin,N. and Shafer,R. (2020) Coronavirus Antiviral Research Database (CoV-RDB): an online database designed to facilitate comparisons between candidate anti-coronavirus compounds. *Viruses*, **12**, 1006.
11. Jiang,S., Hillyer,C. and Du,L. (2020) Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol.*, **41**, 545.
12. Farrera-Soler,L., Daguer,J.-P., Barluenga,S., Vadas,O., Cohen,P., Pagano,S., Yerly,S., Kaiser,L., Vuilleumier,N. and Winssinger,N. (2020) Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS One*, **15**, e0238089.
13. Biswas,A., Bhattacharjee,U., Chakrabarti,A.K., Tewari,D.N., Banu,H. and Dutta,S. (2020) Emergence of novel coronavirus and COVID-19: whether to stay or die out? *Crit. Rev. Microbiol.*, **46**, 182–193.
14. Weisblum,Y., Schmidt,F., Zhang,F., DaSilva,J., Poston,D., Lorenzi,J.C.C., Muecksch,F., Rutkowska,M., Hoffmann,H.-H., Michailidis,E. *et al.* (2020) Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife*, **9**, e61312.
15. Williams,T.C. and Burgers,W.A. (2021) SARS-CoV-2 evolution and vaccines: cause for concern? *Lancet Respir Med*, **9**, 333–335.
16. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
17. Vita,R., Mahajan,S., Overton,J.A., Dhanda,S.K., Martini,S., Cantrell,J.R., Wheeler,D.K., Sette,A. and Peters,B. (2019) The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.*, **47**, D339–D343.
18. Page,A.J., Taylor,B., Delaney,A.J., Soares,J., Seemann,T., Keane,J.A. and Harris,S.R. (2016) : rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*, **2**, e000056.
19. Lima,W.C., Gasteiger,E., Marcatili,P., Duek,P., Bairoch,A. and Cosson,P. (2020) The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.*, **48**, D261–D264.
20. Ng,P.C. and Henikoff,S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.
21. Siepel,A. and Haussler,D. (2005) Phylogenetic Hidden Markov models. In: *Statistical Methods in Molecular Evolution. Statistics for Biology and Health*. Springer, NY.
22. Cooper,G.M. (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*, **15**, 901–913.
23. Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R. and Siepel,A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
24. Vita,R., Mahajan,S., Overton,J.A., Dhanda,S.K., Martini,S., Cantrell,J.R., Wheeler,D.K., Sette,A. and Peters,B. (2019) The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.*, **47**, D339–D343.
25. Nelson,G., Buzko,O., Spilman,P., Niazi,K., Rabizadeh,S. and Soon-Shiong,P. (2021) Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. bioRxiv doi: https://doi.org/10.1101/2021.01.13.426558, 13 January 2021, preprint: not peer reviewed.