Big Data and Information Analytics

https://www.aimspress.com/journal/BDIA

BDIA, 10: 338–349.
DOI:10.3934/bdia.2026017
Received: 03 October 2026
Accepted: 13 October 2026
Published: 04 November 2026

Research article

A web-based predictor for SARS-CoV-2 neutralizing antibody escape from genome sequences

Lanxin Zou[1], **Yixuan Miao**[1], **Rui Liu**[2] **and Daihai He**[1,*]

[1] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong SAR, China

[2] School of Mathematics, South China University of Technology, Guangzhou 510640, China

* **Correspondence:** Email: daihai.he@polyu.edu.hk.

Abstract: SARS-CoV-2 variants can carry spike mutations that reduce the activity of therapeutic antibodies. Laboratory neutralization assays provide direct measurements, but they are too slow and resource-intensive for rapid screening of every newly sequenced genome. We developed a web application that accepts FASTA genomes, checks SARS-CoV-2 identity, assigns one of six major variant categories, detects predefined spike substitutions, and predicts IC_{50} for Casirivimab and Imdevimab. Two ridge regression models were trained on an expanded CoV-DRDB dataset, and the arithmetic mean of their outputs is reported as the cocktail estimate. Spike mutations are called by protein alignment to Wuhan-Hu-1, with offset voting used as a fallback. On a held-out pipeline test set of seven NCBI genomes, both major-category identification and key-mutation recall were 100%. Identification was also 100% on seven external NCBI genomes that were not part of the website reference panel. For ten single-antibody measurements from genomes not used in training, excluding assay-ceiling values, the mean absolute error (MAE) on $\log_{10}$ fold-change was 0.45. On a separate machine-learning test set containing 40 antibody measurements from 20 spike isoforms, the overall MAE was 0.50 and Pearson's r was 0.87. The application is intended for research and teaching, not clinical decision-making.

Keywords: SARS-CoV-2; monoclonal antibody; IC_{50} prediction; spike mutations; ridge regression

1. Introduction

1.1. Background

SARS-CoV-2 was identified in Wuhan in late 2019 [1]. Since then, changes in the spike protein have altered viral transmission and recognition, even influencing antibody neutralization [2]. The D614G

substitution, for example, rapidly became globally dominant and was associated with greater infectivity in experimental systems [3]. Later lineages, including Omicron, accumulated many additional spike substitutions and spread rapidly [4].

Deep mutational scanning can measure antibody escape directly [5]. But such laboratory assays are not practical for screening every genome produced by routine sequencing.

Sequence databases and curated resistance data make a complementary computational approach possible. GenBank provides public viral genome records [6], while the Stanford Coronavirus Resistance Database links spike isoforms to experimental susceptibility measurements for monoclonal antibodies [7]. We used these resources to build a sequence-to-IC_{50} workflow. The model learns associations between spike substitutions and $\log_{10}$ fold-change in potency, then applies those associations to mutations found in an uploaded genome.Its output is only an estimate and cannot be used for clinical decisions.

1.2. Objectives

The application was designed to:

1) Accept SARS-CoV-2 genome sequences in FASTA format
2) Check sequence identity and assign the closest major variant category
3) Detect substitutions from a panel of 38 spike mutations in Wuhan-Hu-1 numbering
4) Estimate IC_{50} for Casirivimab and Imdevimab from models trained on expanded CoV-DRDB data (the author created a data table after cleaning and organizing various data from references)
5) Classify the cocktail estimate as High, Moderate, or Potentially Effective
6) Return the results through a web interface

1.3. Significance

The application gives researchers and students a quick way to explore how a spike mutation profile may affect two common antibodies. It does not establish clinical susceptibility and should not be used to guide treatment.

2. System architecture

2.1. Overall design

The application has four main components (Table 1).

Table 1. System components.

Component	Description
Frontend (HTML/CSS/JS)	FASTA upload and result display
Web server (Flask)	Request handling, file validation, and JSON responses
Sequence analysis (Python)	Variant matching and spike mutation detection
Prediction module	Antibody-specific ridge models and risk classification

2.2. Technology stack

The backend uses Python 3.12 and Flask 3.x. Biopython handles FASTA parsing, translation, and sequence alignment [8]. The ridge regression models were implemented with scikit-learn [9], following the standard ridge formulation [10], and stored as joblib files for the web application. The browser interface uses HTML, CSS, and JavaScript. Exact package requirements are listed in `requirements.txt`.

2.3. Data flow

1) The user uploads a FASTA file
2) Flask validates the file and passes the sequence to `SequenceAnalyzer`
3) The query is compared with reference spike sequences and assigned to one of six displayed categories: Wild Type, Alpha, Beta, Gamma, Delta, or Omicron
4) The spike region is extracted, translated, and checked for the 38 predefined substitutions
5) The detected substitutions are converted into model features and used to estimate antibody IC_{50}
6) Flask returns the analysis as JSON, and the browser displays the result

3. Methods

3.1. Sequence analysis module

Sequence processing is divided between two classes. `SequenceComparator` checks SARS-CoV-2 identity and finds the closest reference category. `SequenceAnalyzer` extracts the spike region, translates it, and calls substitutions for the prediction model.

3.1.1. Overall processing pipeline

For each uploaded file, `SequenceAnalyzer.analyze_sequence_file()` performs the following steps:

1) Read the first FASTA record and calculate sequence length and GC content
2) Compare the query with the reference panel. A sequence is rejected if its highest spike similarity is below 70%
3) Assign the category with the highest spike identity. Similarity $\geq$ 95% is reported as High confidence; 70% $\leq$ similarity $<$ 95% is reported as Medium confidence
4) Align the genome to Wuhan-Hu-1 (NC_045512.2) and map spike coordinates 21563–25384 onto the query
5) Translate the extracted sequence after locating the reading frame from the N-terminal MFV motif
6) Check the translated spike for 38 predefined substitutions, first by alignment-based coordinate mapping and then, if needed, by offset voting
7) Return the category, similarity, detected substitutions, and sequence metadata to the Flask route and prediction model

3.1.2. Sequence comparator

At startup, the comparator reads `manifest.json` and loads 14 NCBI reference assemblies: Wild Type, Alpha, Beta, Gamma, Delta, Omicron, BA.1, BA.2, BA.4, BA.5, BA.2.75.2, BQ.1.1, XBB.1.5, and JN.1. All references can participate in the alignment, but the interface reports only six categories. Omicron sub-lineages are therefore displayed as Omicron.

Variant matching uses global pairwise alignment with affine gap penalties, implemented with Biopython `PairwiseAligner` [11]. Only spike nucleotides are compared. The scores are match = 2, mismatch = −1, gap open = −2, and gap extension = −1. Identity is calculated as

$$\text{Similarity } (\%) = \frac{\text{aligned matching bases}}{\text{aligned non-gap positions}} \times 100. \tag{3.1}$$

3.1.3. Sequence analyzer and mutation calling

The mutation panel contains 38 spike substitutions associated with major SARS-CoV-2 lineages and antibody escape [2, 12, 13]. Each entry records the Wuhan-Hu-1 position, reference amino acid, and alternate amino acid. The panel includes receptor-binding domain substitutions such as N501Y and E484K, furin-cleavage-region changes such as P681R/H, and Omicron markers including G339D, S371F, R346T, and Q954H.

Tier 1 (primary)—alignment-based coordinates The translated query spike is aligned to the 1273-amino-acid Wuhan-Hu-1 spike. The alignment maps each Wuhan position to the corresponding query residue. A substitution is accepted only when at least 100 Wuhan positions can be mapped, the Wuhan residue agrees with the panel entry, and the mapped query residue matches the specified alternate amino acid. This approach allows for insertions and deletions without assuming a fixed offset.

Tier 2 (fallback)—anchor-weighted offset voting If Tier 1 maps fewer than 100 positions, the program searches a ±8 amino-acid window around each panel position. Each match votes for the offset between its observed and expected positions. D614G, N501Y, E484K, L452R, K417N/T, and P681R/H receive a weight of three; other substitutions receive a weight of one. The highest-scoring offset in $[-10, 2]$ is used when it has at least two weighted votes. Calls more than one residue from this consensus are discarded, and the reference and alternate amino acids are checked again at the shifted position. This fallback recovered calls from incomplete or shifted assemblies, including some Gamma/P.1 samples. Table 2 gives representative results.

Table 2. Mutation detection examples on held-out pipeline strains.

Strain	Count	Detected mutations
Wild Type	0	None
Gamma	5	D614G, N501Y, E484K, K417T, H655Y
Delta	6	D614G, L452R, T478K, P681R, T19R, G142D
JN.1	23	D614G, N501Y, K417N, T478K, P681H, H655Y, S371F, S373P, S375F, N440K, G446S, S477N, Q498R, Y505H, N679K, R346T, N460K, F486P, L455F, F456L, D796Y, Q954H, N969K

Detected labels such as "D614G" are converted to feature names such as "614G" and mapped onto the 95-dimensional training vector (sites outside the 38-site panel are set to 0 at inference).

3.2. Neutralization predictor

3.2.1. Antibody selection

Casirivimab and Imdevimab form the two-antibody REGN-COV2 combination. They bind distinct, non-overlapping epitopes in the spike receptor-binding domain [14, 15].

3.2.2. End-to-end prediction workflow

1) Receive the mutation labels from the sequence analyzer
2) Build a 95-element binary feature vector using `feature_columns.json`
3) Predict $\log_{10}$ fold-change $\hat{y}$ with one ridge model for each antibody
4) Clip $\hat{y}$ to $[-2, 4]$ and convert it to IC_{50} using $IC_{50}^{pred} = IC_{50}^{WT} \times 10^{\hat{y}}$
5) Take the arithmetic mean of the two antibody predictions as the cocktail estimate and assign a risk label

3.2.3. Training data and feature construction

Experimental IC_{50} measurements for Casirivimab and Imdevimab were obtained primarily from CoV-DRDB (the dataset which was preprocessed and cleaned from the dataset in reference) [7]. Records were restricted to IC_{50} values in ng/ml, linked to spike mutation lists, and aggregated by the median for each isoform using `scripts/expand_ic50_from_drdb.py`. The source payload also contains an older Arora laboratory table (`cov_unibind/raw/arora.txt`), which contributes early measurements. The resulting file, `processed_ml_features.csv`, contains 560 rows, 264 unique isoforms, and 95 binary mutation features.

The models were fitted to `ml_train.csv` ($n = 507$ rows). Twenty spike isoforms were held out for testing. The target was $\log_{10}$ fold-change from wild-type potency:

$$y = \log_{10}(IC_{50}^{variant}) - \log_{10}(IC_{50}^{WT}). \tag{3.2}$$

Inference and external evaluation use the same CoV-DRDB `B.1 Spike` median baselines: 0.00835 μg/ml (8.35 ng/ml) for Casirivimab and 0.0075 μg/ml (7.5 ng/ml) for Imdevimab. The displayed Wild Type category is Wuhan-Hu-1 (NC_045512.2); the numerical IC_{50} baseline is this `B.1 Spike` median, which typically includes D614G. These values are stored in `wildtype_ic50.csv` and can be regenerated with `scripts/derive_wildtype_ic50.py`.

3.2.4. Model architecture

Separate ridge regression models ($\alpha = 1.0$) were fitted for the two antibodies:

$$\hat{y} = \mathbf{w}^{\top}\mathbf{x} + b \tag{3.3}$$

where $\mathbf{x} \in \{0, 1\}^{95}$ is the binary mutation vector. The predicted IC_{50} is

$$IC_{50}^{pred} = IC_{50}^{WT} \times 10^{\hat{y}}, \tag{3.4}$$

with $\hat{y}$ clipped to $[-2, 4]$. The cocktail estimate is

$$\mathrm{IC}_{50}^{\mathrm{cocktail}} = \frac{\mathrm{IC}_{50}^{\mathrm{Casi}} + \mathrm{IC}_{50}^{\mathrm{Imdev}}}{2}. \tag{3.5}$$

3.2.5. Risk classification

The web application converts the cocktail estimate into the categories shown in Table 3. These are operational thresholds for this tool, not validated clinical breakpoints.

Table 3. Escape risk thresholds (IC_{50}, μg/ml).

IC_{50} range	Risk level	Interpretation
≥ 0.1	High	Substantially reduced neutralization
$0.03 \leq \mathrm{IC}_{50} < 0.1$	Moderate	Partial reduction in potency
< 0.03	Potentially Effective	Retained neutralization

3.3. Web interface

The Flask application exposes three endpoints: `/` for the main page, `/upload` for sequence analysis, and `/health` for service status. The result page shows sequence statistics, similarity, the assigned category, detected mutations, antibody-specific IC_{50} and fold-change estimates, and the risk label.

3.4. Data sources

3.4.1. CoV-DRDB

CoV-DRDB records were processed into `expanded_raw_potency.csv`, `wildtype_ic50.csv`, and the model feature matrix. The expansion can be repeated with `scripts/expand_ic50_from_drdb.py`. External reference labels can be regenerated with `scripts/derive_external_ic50_labels.py`; their provenance is recorded in `pipeline_external_test_provenance.json`.

3.4.2. Reference genomes from NCBI

The 14 reference assemblies in `data/prediction/` are listed in Table 4. They were downloaded from NCBI with `scripts/download_variants.py` [6].

3.5. Train / test partition

The machine-learning split is summarized in Table 5, and the pipeline test genomes are listed in Table 6. A separate external set contains seven NCBI genomes that were never used as website references. Each genome was paired with the median CoV-DRDB IC_{50} for its spike isoform. Both inference and external fold-change calculations use the same `B.1 Spike` baselines: 8.35 ng/ml for Casirivimab and 7.5 ng/ml for Imdevimab.

All reported metrics can be regenerated with `scripts/run_evaluation.py`.

Table 4. Reference genome panel (14 variants).

Variant	Lineage	NCBI Accession
Wild Type	Wuhan-Hu-1	NC_045512.2
Alpha	B.1.1.7	OV054768.1
Beta	B.1.351	MZ433432.1
Gamma	P.1	OR701610.1
Delta	B.1.617.2	OK091006.1
Omicron	B.1.1.529 (early)	OP430898.1
BA.1	BA.1	OL672836.1
BA.2	BA.2	OM617939.1
BA.4	BA.4 (NCBI: BA.4.6)	OR325409.1
BA.5	BA.5	ON332989.1
BA.2.75.2	BA.2.75.2	OR275016.1
BQ.1.1	BQ.1.1	PP848077.1
XBB.1.5	XBB.1.5	PP848050.1
JN.1	JN.1	PP198484.1

Table 5. ML train/test split (by `iso_name`).

Split	File	n	Notes
ML train	`ml_train.csv`	507	244 isoforms in training
ML test	`ml_test.csv`	53	20 isoforms; MAE uses 40 single-antibody rows

Table 6. Held-out pipeline test genomes ($n = 7$). BA.4 and JN.1 are expected to match Omicron.

Variant	Lineage	FASTA file
Alpha	B.1.1.7	alpha_complete.fasta
Beta	B.1.351	beta_complete.fasta
Gamma	P.1	gamma_complete.fasta
Delta	B.1.617.2	delta_complete.fasta
Omicron	B.1.1.529 (early)	omicron_complete.fasta
BA.4	BA.4	ba4_complete.fasta
JN.1	JN.1	jn1_complete.fasta

4. Results

4.1. Pipeline test: variant identification and mutation detection

Seven genomes were processed from FASTA upload through prediction with `scripts/evaluate_pipeline_test_set.py`. All seven were assigned to the expected major category, and all predefined lineage-marker mutations used for scoring were recovered. Table 7 shows the antibody estimates, their arithmetic mean, and the detected key mutations.

Table 7. Pipeline end-to-end test results ($n = 7$). Cocktail IC_{50} in μg/ml.

Genome	Matched	IC_{50} Casi	IC_{50} Imdev	Cocktail	Key mutations
Alpha	Alpha	0.016	0.009	0.012	D614G, N501Y, P681H
Beta	Beta	0.877	0.009	0.443	D614G, N501Y, E484K, K417N
Gamma	Gamma	0.388	0.008	0.198	D614G, N501Y, E484K, K417T, H655Y
Delta	Delta	0.043	0.126	0.085	D614G, L452R, T478K, P681R, T19R, G142D
Omicron	Omicron	0.151	0.106	0.129	19 mutations
BA.4	Omicron	1.982	0.192	1.087	21 mutations
JN.1	Omicron	0.277	14.674	7.476	23 mutations

Alpha was classified as Potentially Effective and Delta as Moderate. Beta, Gamma, Omicron, BA.4, and JN.1 were classified as High. For Beta and Gamma, that High label comes from the cocktail mean: Casirivimab is elevated, while Imdevimab remains near the wild-type baseline.

4.2. External validation on unseen NCBI genomes

The external test used seven NCBI genomes that do not appear in the website manifest. Reference IC_{50} values were recalculated from `expanded_raw_potency.csv` as the median cleaned value for each antibody–isoform pair. The literature cocktail value in Table 8 is the arithmetic mean of the two antibody medians. Measurements at the assay ceiling ($\geq 10,000$ ng/ml) were treated as censored. All seven genomes were assigned to the expected major category. Across the ten uncensored single-antibody measurements, the MAE on $\log_{10}$ fold-change from the `B.1 Spike` baselines was 0.45. This MAE is not computed on the cocktail means in Table 8.

The suffix "-alt" marks an alternate NCBI accession, not a separate lineage. For example, Wuhan-alt is MN908947.3, while the website uses NC_045512.2 as its Wild Type reference. Likewise, Beta-alt (OR353131.1) is different from the Beta accession in the manifest (MZ433432.1). BA.2.12.1, BF.7, and CH.1.1 are Omicron sub-lineages that are also absent from the reference panel.

4.3. ML test: IC_{50} model validation

The held-out model evaluation contained 40 measurements from 20 spike isoforms. As shown in Table 9, the overall $\log_{10}$ fold-change MAE was 0.50 and Pearson's r was 0.87. Table 10 gives several examples.

5. Discussion

5.1. Variant identification and mutation detection

The pipeline correctly assigned all seven test genomes and recovered every predefined key mutation. This result shows that the reference matching and mutation-calling steps worked on this small panel. It

Table 8. External validation using NCBI genomes not present in the reference manifest. "-alt" denotes an alternate accession for the same major category. Cocktail IC_{50} is in ng/ml. Literature and predicted values are component means.

Case	Accession	Matched	Literature (ng/ml)	Predicted (ng/ml)
Wuhan-alt	MN908947.3	Wild Type	7.9	15.0
Delta-alt	OR323381.1	Delta	5.8	46.9
Beta-alt	OR353131.1	Beta	182	443
Gamma-alt	MW642250.1	Gamma	433	198
BA.2.12.1	OR326585.1	Omicron	5375	2205
BF.7	PP847973.1	Omicron	6098	1087
CH.1.1	PP847896.1	Omicron	$> 10,000$	1747

Table 9. ML test set performance ($\log_{10}$ fold-change).

Antibody	n	MAE	Pearson r
Casirivimab	20	0.59	0.75
Imdevimab	20	0.40	0.95
Overall	40	0.50	0.87

Table 10. Selected ML test predictions vs. experimental $\log_{10}$(fold-change).

Isoform	Antibody	Observed	Predicted
BA.1 Spike	Casirivimab	3.08	3.02
BQ.1.1 Spike	Casirivimab	3.78	2.87
CH.1.1 Spike	Casirivimab	3.08	2.70
XBB.1.16 Spike	Imdevimab	2.82	3.21
B.1.617.2 Spike	Casirivimab	−0.18	−0.14

does not establish performance across all SARS-CoV-2 diversity. Reporting six broad categories keeps the output simple, while the mutation list preserves more detailed sequence information.

5.2. IC_{50} prediction

The held-out correlation ($r = 0.87$) shows that the two ridge models captured broad differences in potency across the tested isoforms. The MAE was higher for Casirivimab (0.59) than for Imdevimab (0.40), meaning that the present feature set and linear model fitted Casirivimab less accurately. This difference should not be read as a general comparison of the two antibodies.

The higher estimates for several Beta-, Delta-, and Omicron-derived genomes are broadly consistent with laboratory reports of reduced monoclonal-antibody activity against these lineages [12, 13, 16, 17]. In Table 10, BA.1 was close to its experimental value. Extreme escape was often underestimated for BQ.1.1 and CH.1.1 in that table, and for BF.7 in Table 8, because values reported only at an assay ceiling were treated as exact values during model fitting.

5.3. Limitations

The training split contains 507 rows for 95 mutation features, and the ridge penalty was fixed at $\alpha = 1.0$ rather than selected by cross-validation. Measurements reported at an upper assay limit were treated as exact, which compresses very large fold-changes. The cocktail estimate is only the arithmetic mean of the two component predictions and does not model interaction or synergy, so a genome may be labelled High even if one antibody remains potent. Finally, the pipeline and external genome tests each contain only seven cases.

5.4. Future work

Future versions could include more antibodies and a mutation panel that can be updated as new lineages appear. Cross-validated model selection and methods that handle censored measurements would strengthen the statistical analysis. A cocktail model trained on combination data would also be preferable to averaging two single-antibody predictions. Automated updates from NCBI and CoV-DRDB would reduce manual maintenance.

6. Conclusions

We developed a working web application that accepts SARS-CoV-2 FASTA genomes, assigns a major variant category, detects predefined spike substitutions, and estimates IC_{50} for Casirivimab and Imdevimab from models trained on expanded CoV-DRDB data.

Major-category identification and key-mutation recall were both 100% on the seven-genome pipeline test, and identification was 100% on the seven external NCBI genomes. On the held-out model test, the overall $\log_{10}$ fold-change MAE was 0.50 ($r = 0.87$) across 40 measurements from 20 isoforms. The source code is available at `https://github.com/coohzou/sars-cov-2-antibody-escape-predictor`, and the web application is available at `https://sars-cov-2-antibody-escape-predictor.onrender.com`. The reported values are research estimates and must not be used for clinical decisions.

Data availability

Evaluation outputs are generated by `scripts/run_evaluation.py`. External IC_{50} label provenance is stored in `data/evaluation/pipeline_external_test_provenance.json` and can be regenerated with `scripts/derive_external_ic50_labels.py`. Trained ridge models are included under `data/training/models/`.

Use of AI tools declaration

Generative AI tools (mainly OpenAI's GPT-5.6) were used to assist with coding, formatting, and language editing. The authors reviewed the final manuscript and take responsibility for its content. Experimental IC_{50} data were obtained from CoV-DRDB, and reference genomes were obtained from NCBI GenBank. .

Acknowledgments

The authors thank the University Research Facility in Big Data Analytics (UBDA) for computational resources, and the Smart Cities Research Institute and the Hong Kong Polytechnic University for financial support. The work was partially supported by the Hong Kong Research Grants Council Theme-based Research Scheme fund (T24-508/22-N), the Hong Kong Research Grants Council General Research Fund (15307626), and the Hong Kong Research Grants Council Collaborative Research Fund (C5079-21G).

Conflict of interest

The authors declare no conflicts of interest.

References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579: 265–269. https://doi.org/10.1038/s41586-020-2008-3
2. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. (2021) SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19: 409–424. https://doi.org/10.1038/s41579-021-00573-0
3. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. (2020) Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182: 812–827.e19. https://doi.org/10.1016/j.cell.2020.06.043
4. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. (2022) Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 603: 679–686. https://doi.org/10.1038/s41586-022-04411-y
5. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. (2021) Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29: 44–57.e9. https://doi.org/10.1016/j.chom.2020.11.007
6. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, et al. (2022) GenBank. *Nucleic Acids Res* 50: D161–D164.https://doi.org/10.1093/nar/gkab1135
7. Tzou PL, Tao K, Pond SLK, Shafer RW.(2022) Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. *PLoS ONE* 17: e0261045. https://doi.org/10.1371/journal.pone.0261045
8. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423. https://doi.org/10.1093/bioinformatics/btp163
9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12: 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

10. Hoerl AE, Kennard RW, (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12: 55–67. https://doi.org/10.1080/00401706.1970.10488634

11. Gotoh O, (1982) An improved algorithm for matching biological sequences. *J Mol Biol*, 162: 705–708. https://doi.org/10.1016/0022-2836(82)90398-9

12. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. (2021) Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593: 130–135. https://doi.org/10.1038/s41586-021-03398-2

13. Cao Y, J Wang, F Jian, T Xiao, W Song, A Yisimayi, et al. (2022) Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* 602: 657–663. https://doi.org/10.1038/s41586-021-04385-3

14. Hansen J, Baum A, Pascal KE, Russo V, Giordano S, Wloga E, et al. (2020) Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* 369: 1010–1014. https://doi.org/10.1126/science.abd0827

15. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. (2020) Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 369: 1014–1018. https://doi.org/10.1126/science.abd0831

16. Planas D, et al. (2021) Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 596 : 276–280. https://doi.org/10.1038/s41586-021-03777-9

17. VanBlargan LA, Errico JM, Halfmann PJ, Zost SJ, Crowe Jr JE, Purcell LA, et al. (2022) An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. *Nat Med* 28: 490–495. https://doi.org/10.1038/s41591-021-01678-y



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)