Materials Design Analysis Reporting (MDAR) Checklist for Authors

The MDAR framework establishes a minimum set of requirements in transparent reporting applicable to studies in the life sciences (see Statement of Task: doi:10.31222/osf.io/9sm4x.). The MDAR checklist is a tool for authors, editors and others seeking to adopt the MDAR framework for transparent reporting in manuscripts and other outputs. Please refer to the MDAR Elaboration Document for additional context for the MDAR framework.

Materials

Antibodies	Yes (indicate where provided: page no/section/legend)	n/a
For commercial reagents, provide supplier name, catalogue number and RRID, if available.		X

Cell materials	Yes (indicate where provided: page no/section/legend)	n/a
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID		X
Primary cultures: Provide species, strain, sex of origin, genetic modification status.		X

Experimental animals	Yes (indicate where provided: page no/section/legend)	n/a
Laboratory animals: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID		X
Animal observed in or captured from the field: Provide species, sex and age where possible		X
Model organisms: Provide Accession number in repository (where relevant) **OR** RRID		X

Plants and microbes	Yes (indicate where provided: page no/section/legend)	n/a
Plants: provide species and strain, unique accession number if available, and source (including location for collected wild specimens)		X
Microbes: provide species and strain, unique accession number if available, and source		X

Human research participants	Yes (indicate where provided: page no/section/legend)	n/a
Identify authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		X
Provide statement confirming informed consent obtained from study participants.		X
Report on age and sex for all study participants.		X

Design

Study protocol	Yes (indicate where provided: page no/section/legend)	n/a
For clinical trials, provide the trial registration number **OR** cite DOI in manuscript.		X

Laboratory protocol	Yes (indicate where provided: page no/section/legend)	n/a
Provide DOI or other citation details if detailed step-by-step protocols are available.		X

Experimental study design (statistics details)	Yes (indicate where provided: page no/section/legend)	n/a
State whether and how the following have been done**, or** if they were not carried out.	All analysis used publicly available, published datasets with sample sizes and criteria determined by previous study authors.	
Sample size determination		X
Randomisation		X
Blinding		X
Inclusion/exclusion criteria		X

Sample definition and in-laboratory replication	Yes (indicate where provided: page no/section/legend)	n/a
State number of times the experiment was replicated in laboratory		X
Define whether data describe technical or biological replicates		X

Ethics	Yes (indicate where provided: page no/section/legend)	n/a
Studies involving human participants: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		X
Studies involving experimental animals: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		X
Studies involving specimen and field samples: State if relevant permits obtained, provide details of authority approving study; if none were required, explain why.		X

Dual Use Research of Concern (DURC)	Yes (indicate where provided: page no/section/legend)	n/a
If study is subject to dual use research of concern, state the authority granting approval and reference number for the regulatory approval		X

Analysis

Attrition	Yes (indicate where provided: page no/section/legend)	n/a
State if sample or data point from the analysis is excluded, and whether the criteria for exclusion were determined and specified in advance.	Protein sequences from public databases were excluded based on predetermined criteria meant to remove incomplete or low-quality sequences similar to criteria used in previous studies; see "Viral protein sequence datasets and model training" in **Materials and Methods** for more details.	

Statistics	Yes (indicate where provided: page no/section/legend)	n/a
Describe statistical tests used and justify choice of tests.	For statistical significance of correlation coefficients, we make a standard assumption that the null distribution is *t*-distributed with $N - 2$ degrees of freedom. For statistical significance of AUC values, combinatorial escape potential, and regional escape enrichment, in which the null distribution is not clear, we use randomized, permutation-based methods to estimate the null distribution. Additional statistical details described in **Materials and Methods**.	

Data Availability	Yes (indicate where provided: page no/section/legend)	n/a
State whether newly created datasets are available, including protocols for access or restriction on access.	All data in analysis is reused and publicly available; see **Data and materials availability**.	
If data are publicly available, provide accession number in repository or DOI or URL.	Code and pretrained models are deposited to Zenodo at doi:10.5281/zenodo.4033175. Training and validation datasets used in this study are deposited to Zenodo at doi:10.5281/zenodo.4029296.	
If publicly available data are reused, provide accession number in repository or DOI or URL, where possible.	See **Data and materials availability**. We used the following publicly available datasets for model training: • Influenza A HA protein sequences from the NIAID Influenza Research Database (IRD) (http://www.fludb.org) • HIV-1 Env protein sequences from the Los Alamos National Laboratory (LANL) HIV database (https://www.hiv.lanl.gov) • *Coronavidae* spike protein sequences from the Virus Pathogen Resource (ViPR) database (https://www.viprbrc.org/brc/home.spg?decorator=corona) • SARS-CoV-2 Spike protein sequences from NCBI Virus (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/) • SARS-CoV-2 Spike and other Betacoronavirus spike protein sequences from GISAID (https://www.gisaid.org/) We used the following publicly available datasets for fitness and escape validation: • Fitness single-residue DMS of HA H1 WSN33 from Doud and Bloom (2016) (https://www.mdpi.com/1999-4915/8/6/155/htm) • Fitness combinatorial DMS of antigenic site B in six HA H3 strains from Wu et al. (https://github.com/wchnicholas/site_B_landscape) • Fitness single-residue DMS of Env BF520 and BG505 from Haddox et al. (https://github.com/jbloomlab/EnvMutationalShiftsPaper)	

	• ACE2 binding affinity combinatorial DMS of Spike from Starr et al. (https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS) • Escape single-residue DMS of HA H1 WSN33 from Doud et al. (2018) (https://github.com/jbloomlab/HA_antibody_ease_of_escape) • Escape single-residue DMS of HA H3 Perth09 from Lee et al. (https://github.com/jbloomlab/map_flu_serum_Perth2009_H3_HA) • Escape single-residue DMS of Env BG505 from Dingens et al. (https://github.com/jbloomlab/EnvsAntigenicAtlas) • Escape mutations of Spike from Baum et al. (https://science.sciencemag.org/content/early/2020/06/15/science.abd0831) • Escape single-residue DMS of Spike RBD from Greaney et al. (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies)	

Code Availability	**Yes (indicate where provided: page no/section/legend)**	**n/a**
For all newly generated code and software essential for replicating the main findings of the study:		
State whether the code or software is available.	Yes; see **Data and materials availability**.	
If code is publicly available, provide accession number in repository, or DOI or URL.	Code and pretrained models are deposited to Zenodo at doi:10.5281/zenodo.4034681 and are also available at https://github.com/brianhie/viral-mutation	

Reporting

Adherence to community standards	**Yes (indicate where provided: page no/section/legend)**	**n/a**
MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.		
State if relevant guidelines (eg., ICMJE, MIBBI, ARRIVE) have been followed, and whether a checklist (eg., CONSORT, PRISMA, ARRIVE) is provided with the manuscript.	We adhere to all other journal-specific policies.	