nature portfolio

Corresponding author(s): Leo L.M. Poon

Last updated by author(s): 2022-08-03

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection in this study.
Data analysis	For NGS data analyses: BWA-MEM2 (v2.0pre2), iVar (v1.3.1) , SNPGenie (v1.0). For multiple other analysis: R (v 4.1.3) with custom codes (https://github.com/Leo-Poon-Lab/mutations-under-sarscov2-vaccination).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The sequencing data used in this study can be access through NCBI Sequence Read Archive (SRA) with accession ID: XXX. The anonymised metadata are deposited at https://github.com/Leo-Poon-Lab/mutations-under-sarscov2-vaccination/XXX.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	No information about sex and gender is disclosed
Population characteristics	All samples were collected from COVID-19 patients detected in Hong Kong
Recruitment	N/No patient was recruited in this study. These were archived SARS-CoV-2 samples confirmed by citywide public health screening programs.
Ethics oversight	This study was conducted under ethical approval from the Institutional Review Board of the University of Hong Kong (UW 20-168).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is a gnomic study that used all available genomic data (N=2053) to in fer the within-host diversity of SARS-CoV-2.
Data exclusions	The samples and data with insufficient quality (see Methods) were excluded from the analysis.
Replication	Amongst the tested samples, 182 samples were collected from 89 patients (i.e. serial samples)
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
	☐ Antibodies
	☐ Eukaryotic cell lines
	☐ Palaeontology and archaeology
	☐ Animals and other organisms
	☐ Clinical data
	☐ Dual use research of concern

Methods

n/a	Involved in the study
	☐ ChIP-seq
	☐ Flow cytometry
	☐ MRI-based neuroimaging