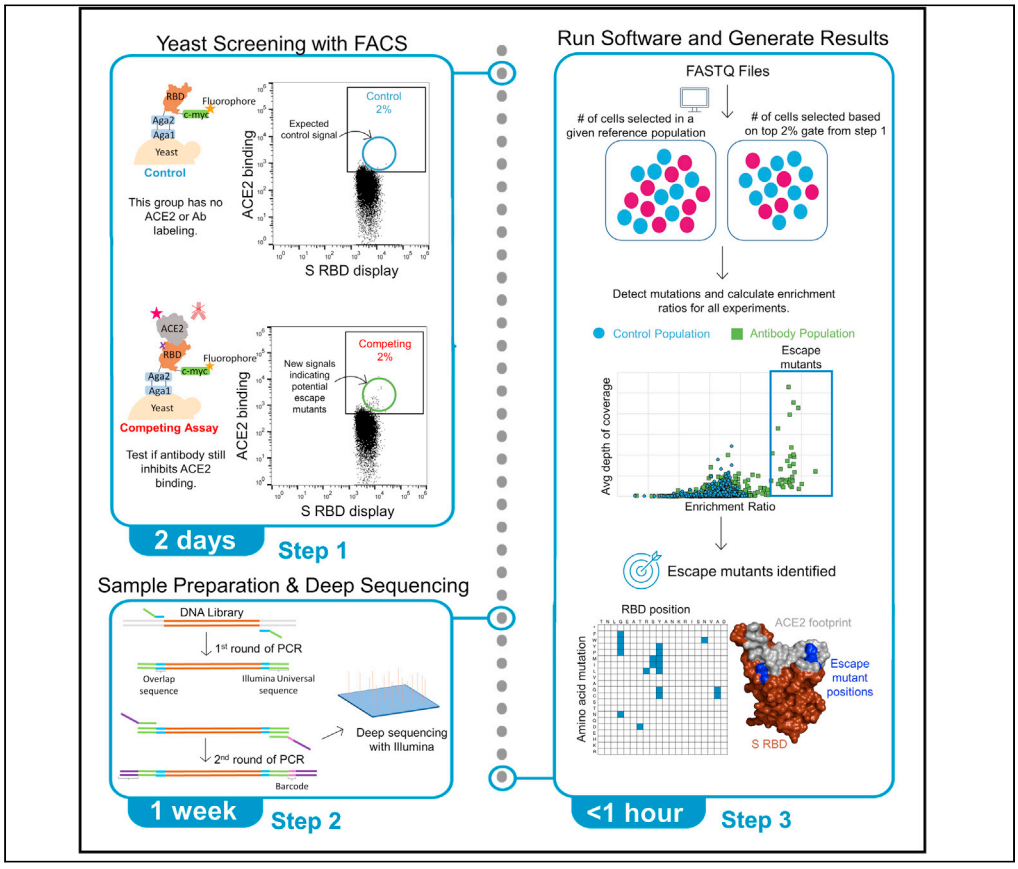
STAR Protocols

CellPress
OPEN ACCESS

Protocol

# Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning




Here, we describe a protocol to identify escape mutants on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) receptor-binding domain (RBD) using a yeast screen combined with deep mutational scanning. Over 90% of all potential single S RBD escape mutants can be identified for monoclonal antibodies that directly compete with angiotensin-converting enzyme 2 for binding. Six to 10 antibodies can be assessed in parallel. This approach has been shown to determine escape mutants that are consistent with more laborious SARS-CoV-2 pseudoneutralization assays.

Cyrus M. Haas, Irene M. Francino-Urdaniz, Paul J. Steiner, Timothy A. Whitehead

cyrus.haas@colorado.edu (C.M.H.)
timothy.whitehead@colorado.edu (T.A.W.)

## Highlights

A protocol for identifying S escape mutations for anti-S monoclonal antibodies

All anti-S antibodies that competitively inhibit ACE2 can be tested

Protocol uses yeast display screening of freely available S RBD libraries

Open-source software used to analyze sequencing data and identify escape mutants

Haas et al., STAR Protocols *24*, 100869
December 17, 2021 © 2021 The Authors.
https://doi.org/10.1016/j.xpro.2021.100869

STAR Protocols

CellPress
OPEN ACCESS

Protocol

# Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning

Cyrus M. Haas,[1,2,3,*] Irene M. Francino-Urdaniz,[1,2] Paul J. Steiner,[1] and Timothy A. Whitehead[1,4,*]

[1]Department of Chemical and Biological Engineering, University of Colorado, Boulder, Boulder, CO 80303, USA

[2]These authors contributed equally

[3]Technical contact

[4]Lead contact

*Correspondence: cyrus.haas@colorado.edu (C.M.H.), timothy.whitehead@colorado.edu (T.A.W.)
https://doi.org/10.1016/j.xpro.2021.100869

## SUMMARY

**Here, we describe a protocol to identify escape mutants on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) receptor-binding domain (RBD) using a yeast screen combined with deep mutational scanning. Over 90% of all potential single S RBD escape mutants can be identified for monoclonal antibodies that directly compete with angiotensin-converting enzyme 2 for binding. Six to 10 antibodies can be assessed in parallel. This approach has been shown to determine escape mutants that are consistent with more laborious SARS-CoV-2 pseudoneutralization assays.**

**For complete details on the use and execution of this protocol, please refer to Francino-Urdaniz et al. (2021).**

## BEFORE YOU BEGIN

### Protocol overview

This method identifies S RBD nAb escape mutations by identifying S RBD mutants that are not bound by a nAb but retain the ability to bind ACE2.

SARS-CoV-2 initiates infection with binding of the Spike receptor binding domain (S RBD) to angiotensin-converting enzyme 2 (ACE2). Many neutralizing antibodies (nAbs) prevent infection by binding to S RBD and blocking ACE2 binding. We have constructed three different S RBD libraries based on the original Wuhan-Hu-1 S RBD N343Q (333-537) (SRA: SAMN18250431-SAMN18250483). The first is in the original background, the second is in the E484K background, and the third is in the N501Y background. The E484K and N501Y mutations exist in currently circulating Alpha, Beta, and Gamma variants. These libraries are available from AddGene. Each is provided as two sub-libraries (Tile 1 and Tile 2) for compatibility with 250bp paired-end Illumina sequencing. Tile 1 contains mutations at positions 333-437, while tile 2 contains mutations at positions 438-537.

### Prepare chemically competent *S. cerevisiae* EBY100

Timing: 6 h

Competent yeast can be prepared in advance and stored at −80°C for at least 6 months (Medina-Cucurella and Whitehead, 2018).

1. Prepare chemically competent *S. cerevisiae* EBY100 (MYA-4941™)


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Table 1. List of plasmids that can be used for yeast display screening**

| Plasmid | Characteristics | Antibiotic resistance | AddGene collection # |
|---|---|---|---|
| pJS697 | Contains yeast display backbone | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699 | Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q (333-537) ("WT") | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001 | Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q, E484K (333-537) ("E484K") | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002 | Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q, N501Y (333-537) ("N501Y") | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699_L1 | Library of mutations on WT RBD between positions 333 and 437 | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699_L2 | Library of mutations on WT RBD between positions 438 and 537 | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001_T1 | Library of mutations on E484K RBD between positions 333 and 437 | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001_T2 | Library of mutations on E484K RBD between positions 438 and 537 | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002_T1 | Library of mutations on N501Y RBD between positions 333 and 437 | Kan | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002_T2 | Library of mutations on N501Y RBD between positions 438 and 537 | Kan | https://www.addgene.org/Timothy_Whitehead/ |

a. Grow EBY100 cells from glycerol stock in 500 mL YPD (recipe) at 30°C and 300rpm until culture reaches an $OD_{600}$ of 1.2 (about 4 h).
b. Harvest the EBY100 cells by centrifugation at 4000 × *g* for 5 min.
c. Wash the pellet by resuspending in 250 mL sterile $H_2O$ and repelleting by centrifugation at 4000 × *g* for 5 min.
d. Wash the pellet with 10 mL of 100 mM lithium acetate and repellet by centrifugation at 4000 × *g* for 5 min.
e. Resuspend cells in 3.5 mL of 100 mM lithium acetate and then add 1.5 mL of 50% v/v glycerol and mix.
f. Prepare 210 µL aliquots and store at −80°C. Do not use dry ice to freeze the cells.

***Note:*** Typical transformation efficiencies for these prepared cells are $5 \times 10^5$ colony forming units (cfu) per µg digested DNA.

## Plasmid preparation

**Timing: 1 day**

For this step, the plasmid containing yeast display backbone (pJS697) and a plasmid containing the S RBD insert (see Table 1 for variations) are digested. All yeast display plasmids are freely available via AddGene as *E. coli* glycerol cell stocks.

2. Digest backbone (pJS697; two BsaI restriction sites) and RBD-encoding (pJS699; two NotI restriction sites) plasmids. The pIFU-series plasmids follow the same procedure as pJS699.
   a. Set up digestion reactions as follows:

| Reagent | Final concentration | Amount |
|---|---|---|
| pJS697 (pJS697 for library) | n/a | 1 µg (5 µg) |
| 10× CutSmart Buffer | n/a | 5 µL |
| BsaI-HFv2 | n/a | 1 µL |
| Nuclease-free water | n/a | to 50 µL |
| **Total** | **n/a** | **50 µL** |

*Note:* Digestions are set up following NEB recommendations (NEB cloner).

| Reagent | Final concentration | Amount |
|---|---|---|
| pJS699 (or pJS699_L1, pJS699_L2) | n/a | 1 μg (5 μg) |
| 10× CutSmart Buffer | n/a | 5 μL |
| NotI-HF | n/a | 1 μL |
| Nuclease-free water | n/a | to 50 μL |
| **Total** | **n/a** | **50 μL** |

*Note:* Digestions are set up following NEB recommendations (NEB cloner).

b. Incubate the digestion reactions for one hour at 37°C.
c. Purify linear DNA from digested plasmids.
   i. Purify the backbone from the pJS697 digestion using a Monarch PCR & DNA purification kit. The digestion will result in two linear DNA pieces: the insert between BsaI sites is 26bp and the backbone is 6042bp in length. DNA smaller than 45bp will not stick to the column (Monarch PCR & DNA Cleanup kit) so the eluate will only contain the backbone. Thus, gel electrophoresis and extraction are not necessary with the pJS697 digestion.
   ii. Purify the linear insert from the pJS699 digestion containing S RBD by gel electrophoresis and gel extraction. The digestion will result in two linear DNA pieces: one at 832 bp and the other at 2692 bp. Excise the 832 bp fragment and purify using a Monarch Gel Extraction kit (NEB).
d. Quantify the linear DNA products by $A_{260}$ using a UV-Vis plate reader or nanodrop spectrophotometer. Typical yields for the digestion of 1 μg plasmid DNA are 600 ng (30 ng/μL) for backbone and 100 ng (5 ng/μL) for the gene encoding S RBD.

*Note:* Linearized plasmids can be stored frozen at -20°C for at least 3 weeks.

### Co-transformation of S RBD wild type into *S. cerevisiae* EBY100

Timing: 3–4 days

Here, the yeast display backbone from the digest reaction (pJS697) is recombined with the insert containing the S RBD by co-transformation into chemically competent *S. cerevisiae* EBY100 (ATCC MYA-4941) following Medina-Cucurella and Whitehead (2018). This step along with transformations for the S RBD libraries can be prepared in parallel.

**Day 1: Plasmid co-transformation for linear DNA of yeast display backbone**

3. Co-transform linear DNA corresponding to the yeast display backbone and S RBD insert into chemically competent *S. cerevisiae* EBY100 (MYA-4941™) (Figure 1A).
   a. Boil 10 μL salmon sperm DNA 10min at 97°C using a Thermal Cycler.
   b. Thaw chemically competent EBY100. Thawing can be performed either at 20°C–25°C or on ice.
   c. In a 1.5mL microcentrifuge tube add in order:

| Reagent | Final concentration | Amount |
|---|---|---|
| EBY100 cells | n/a | 70 μL |
| 50% w/v PEG 3350 | n/a | 240 μL |
| 1M lithium acetate | n/a | 36 μL |
| Salmon sperm DNA | n/a | 10 μL |
| **Total** | **n/a** | **356 μL** |

d. Swirl with pipette tip and vortex to mix thoroughly until solution looks homogeneous. Yeast cells are hardy and will remain viable after a hard mix.
e. To 50 μL of the cell mixture, add DNA in a 3:1 molar ratio of insert to backbone for a final 350 ng of DNA (100 ng linearized pJS699 insert and 250 ng linearized pJS697). The amount of DNA will typically result in $>10^4$ transformants.

*Alternatives:* The amount of DNA used can be reduced, and the molar ratio does not need to be strictly 3:1.

f. Incubate 30min at 30°C in a static incubator or Thermal Cycler.
g. Incubate 20min at 42°C in a water bath or Thermal Cycler.
h. Spin down 1 min at 16000 × *g* using a microcentrifuge. Remove supernatant.
i. Resuspend in 100 μL $SDCAA^+$ without antibiotics (recipe) and incubate for 5 min at 20°C–25°C.
j. Plate 3 serial dilutions into SDCAA agar plates (recipe) and incubate 2–3 days at 30°C.

Day 3: Select colonies

k. Select 3-5 individual colonies from the SDCAA agar plate and inoculate in 1.5 mL $SDCAA^+$ in a 14 mL culture tube (VWR Cat# 10127-334). Grow for 16–20 h at 30°C and 300 rpm in a shaker incubator.

Day 4: Making yeast stocks

l. Use a spectrophotometer to record $OD_{600}$. In a centrifuge, spin cells down 30 s at 16000 × *g* and make 1 mL yeast stocks at $OD_{600}$=1 in yeast storage buffer (recipe). Saturated *S. cerevisiae* EBY100 cultures will have an $OD_{600}$ of 3-5. We typically perform a 10-fold dilution of culture to determine the $OD_{600}$. Each 1.5 mL culture will produce an average of six 1 mL yeast stocks.

## Co-transformation of S RBD libraries into *S. cerevisiae* EBY100

Timing: 4 days

Day 1: Plasmid co-transformation for S RBD libraries

4. Co-transform the libraries into chemically competent *S. cerevisiae* EBY100 (MYA-4941™) (Figure 1A).
   a. Boil 30 μL salmon sperm DNA 10min at 97°C using a Thermal Cycler.
   b. Thaw chemically competent EBY100 at 20°C–25°C or on ice.
   c. In a 1.5mL microcentrifuge tube add in order:

| Reagent | Final concentration | Amount |
|---|---|---|
| EBY100 cells | n/a | 210 μL |
| 50% w/v PEG 3350 | n/a | 720 μL |
| 1M lithium acetate | n/a | 108 μL |
| Salmon sperm DNA | n/a | 30 μL |
| **Total** | **n/a** | **1068 μL** |

   d. Swirl with pipette tip and vortex to mix thoroughly until solution looks homogeneous. Yeast cells are hardy and will remain viable.
   e. Add DNA in a 3:1 insert to backbone molar ratio to give a final concentration of 1.5 μg DNA (1.06 μg vector and 440 ng insert). Alternatively, to obtain more transformants, the

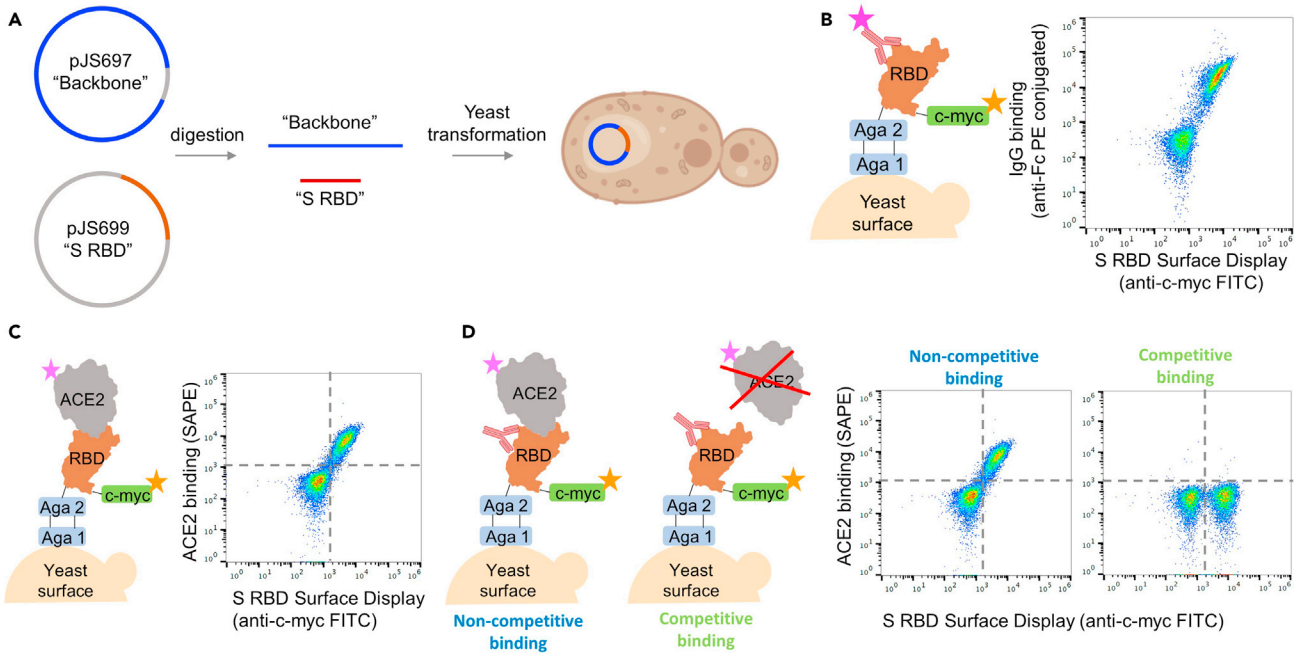



**Figure 1. Preparative steps for the escape mutant protocol**

(A) Homologous recombination in yeast. Left to right: the two plasmids (one containing the backbone and the other one the S RBD insert) are digested independently and the segments are collected. The linearized plasmids are co-transformed into yeast. Yeast cells generate a single plasmid through homologous recombination.

(B) S RBD displayed on the yeast surface bound to antibody. A PE fluorescence signal increase is observed when an antibody binds to S RBD. Two independent populations are detected when screening using FACS. The $IgG^{low}$/RBD $Display^{low}$ population are yeast cells not containing plasmid.

(C) S RBD labeled with biotinylated ACE2.

(D) Antibodies may bind competitively or non-competitively. Incubation of unlabeled antibody followed by co-incubation with biotinylated ACE2 results in loss of ACE2 binding signal only for the competitive inhibition case. This protocol is designed to identify escape mutants for neutralizing antibodies that function by competitive inhibition of ACE2.

amount of DNA transformed can increase up to 5 μg total without increasing the amount of other reagents.

f. Incubate 30min at 30°C in a static incubator (preferred) or Thermal Cycler.

g. Incubate 20min at 42°C in a water bath (preferred) or Thermal Cycler.

h. Spin down 30s at 16000 × *g* using a microcentrifuge. Remove supernatant.

i. Resuspend in 1mL SDCAA (recipe - make $SDCAA^+$ without antibiotics) and incubate for 5 min at 20°C–25°C.

j. Make 6 serial dilutions and plate 10 μL of each in an SDCAA agar plate to calculate transformation efficiency. Incubate 2–3 days at 30°C. Obtaining near-complete (>99.9%) coverage of the libraries requires $>1.2\times10^5$ transformants for each library (Steiner et al., 2020). The first tile (333-437) contains 1120 mutations and the second tile encodes 1260 mutations. The number of transformants obtained can be increased by scaling up the transformation, using more DNA, or using freshly prepared competent cells before freezing.

k. Transfer remaining culture into 100mL $SDCAA^+$ (recipe) in a 500 mL flask. Incubate at 30°C and 300rpm in the incubator shaker until saturation. This typically takes 48 h.

### Day 4: Making yeast stocks

l. Make 1 mL yeast stocks at $OD_{600}$=10

i. Centrifuge the cell culture 3 min at 3200 × *g*.

ii. Remove the supernatant and resuspend the cell pellet in yeast storage buffer (recipe) at $OD_{600}$=10. Stocks can be stored for over a year at -80°C.

### Confirm that neutralizing antibodies selected bind and inhibit the yeast displayed glycosylated WT RBD

Timing: 2 days

A recommended though not essential step is to confirm that your neutralizing antibody binds to displayed RBD and can competitively inhibit binding of surface-displayed S RBD to soluble ACE2. Here, yeast cells are grown from frozen stocks and induced for 16–20 h. For labeling, our lab prepares chemically biotinylated soluble ACE2-Fc produced by the Institute of Protein Design (Walls et al., 2020). However, biotinylated soluble ACE2 is now commercially available (e.g., R&D Systems Cat# BT933-020). We have not validated the use of other sources of ACE2 in this protocol. We biotinylate ACE2-Fc with NHS-Biotin following the vendor's protocol (User guide: EZ-link NHS Biotin) using a biotin to protein molar ratio from 20:1 to 80:1.

Day 1: Cell induction

5. Thaw a 1mL stock of yeast cells harboring the S RBD display plasmid.
6. Pellet the cells by centrifuging 1 min at 16000 × *g* using a microcentrifuge, remove supernatant, and resuspend in 1 mL $SDCAA^+$ (recipe). Transfer culture to a 14 mL culture tube (VWR Cat#10127-334).
7. Incubate at least 4h at 30°C and 300rpm in an incubator shaker. The cell concentration should double, but the culture should not be saturated. It is convenient to set this culture up in the morning so that cytometer screening can be performed during daytime hours on the following day.
8. Measure the culture $OD_{600}$ with a spectrophotometer. Centrifuge the cells 1 min at 16000 × *g* in a microcentrifuge and resuspend to $OD_{600}$=1 in $SGCAA^+$ (recipe).
9. Transfer 1 mL of the culture to a new 14 mL culture tube and incubate for 22 h at 22°C and 300 rpm in an incubator shaker.

Day 2: Test cells for (1) surface display of S RBD, (2) ability to bind biotinylated ACE2, (3) ability to bind antibody of interest, and (4) for antibody ability to competitively inhibit ACE2.

10. Transfer the culture to a 1.5 mL microcentrifuge tube and centrifuge for 1 min at 16000 × *g* using a microcentrifuge.
11. Measure the culture $OD_{600}$ and resuspend the cells in ice cold PBSF to an $OD_{600}$=2. From this point, keep the cells on ice at all times except when otherwise specified. We typically perform a 10-fold dilution of the culture with PBSF to determine the $OD_{600}$. The yeast culture does not usually grow during the 16–20 h induction with common resuspension volumes under 1 mL.

⚠ CRITICAL: On the same day, prepare PBSF (recipe) and keep on ice for the duration of the experiments. PBSF needs to be ice cold whenever added to cells.

Screen unlabeled cells

12. Add 5 μL cells with 45 μL of PBSF ('unlabeled' cells) in a 1.5mL microcentrifuge tube. Incubate 30 min at RT (20°C–25°C).
13. Centrifuge 1 min at 16000 × *g* using a microcentrifuge and resuspend the cells in 100μL ice cold PBSF. Screen 25,000 unlabeled cells. These unlabeled cells are used to set a FSC-A/$FITC^+$ gate to collect 0% of the events. Refer to FACS for more detail on setting the gates (Figure 2B).

Check display of S RBD

14. Label 5 μL cells with 44 μL of PBSF and 1 μL of anti-c-myc FITC in a 1.5mL microcentrifuge tube. Mix by pipetting up and down several times. Incubate the cells on ice for 10 min protected from light.

15. Centrifuge the cells 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 µL ice cold PBSF and remove supernatant. Cell pellets can be stored on ice until ready to screen.
16. Resuspend the cells in 100µL ice cold PBSF and screen 25,000 cells using FACS and analyze the FSC-H/FITC$^{+}$ gate.

⚠ CRITICAL: If less than 30% of cells display RBD, do not continue and induce a new cell stock. Refer to troubleshooting 1 for potential fixes.

### Check ACE2 binding, Ab binding, and competitive inhibition

17. Set up three separate labeling reactions in three different 1.5 mL microcentrifuge tubes as shown below. For tubes 1 and 3, a lower ACE2 concentration can be used if saturating conditions are maintained (we have also used 30 nM biotinylated ACE2).

| | Tube 1 | Tube 2 | Tube 3 |
|---|---|---|---|
| Cells | 5µL | 5µL | 5µL |
| Biotinylated ACE2 | 75nM | - | 75nM |
| Antibody | 10µg/mL | 10µg/mL | 10µg/mL |
| PBSF | To 50µL total volume | To 50µL total volume | To 50µL total volume |
| | Incubate for 30 min at 20°C–25°C. | | |
| Biotinylated ACE2 | - | - | 75nM |
| | - | - | Incubate for 30 min at 20°C–25°C. |

For example: In tube 1, if your biotinylated ACE2 concentration was 1 µM, you would add 5 µL of cells, 41.2 µL PBSF, and 3.8 µL antibody.

*Optional:* Yeast surface titration can be performed following Chao et al. (2006) to determine the effective dissociation constant. The observed dissociation constants of different ACE2-Fc preparations range from 200 pM to 2 nM.

18. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 µL PBSF and remove supernatant.
19. Resuspend each cell pellet in 50 µL fluorescent labeling mix (recipe). The set up for each tube is shown in the following table.

| | Tube 1 | Tube 2 | Tube 3 |
|---|---|---|---|
| anti-c-myc FITC | 0.6µL | 0.6µL | 0.6µL |
| Goat anti-Human IgG FC PE | - | 0.25µL | - |
| Streptavidin-Phycoerythrin (SAPE) | 0.25µL | - | 0.25µL |
| PBSF | 49.15 µL | 49.15 µL | 49.15 µL |

⚠ CRITICAL: Keep the fluorophores on ice and protected from light at all times.

*Alternatives:* Goat anti-Human IgG Fc PE conjugate is used as the secondary label for the antibody labeling reaction because our neutralizing antibodies are human IgGs. Different antibody platforms (e.g. Fabs, scFvs) or non-human antibodies would require a different secondary label.

20. Incubate 10 min on ice protected from light.
21. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 µL PBSF and remove supernatant.

*Note:* The cell pellet can be stored on ice until ready to screen. We do not recommend storing on ice for longer than 1h as fluorophores can dissociate, resulting in a lower signal.

22. Resuspend the cells in 100 µL PBSF. Screen 25,000 cells.

⚠ CRITICAL: The results of this experiment tell us the likelihood of success for the escape mutant protocol. First, the ACE2 binding population signal must be significantly higher and independent from the non-displaying population on the $PE^{+}/FITC^{+}$ (Figure 1C). Refer to troubleshooting 2 to help improve signal. Second, if no PE signal increase is observed for the biotinylated ACE2 labeling (tube 1) and antibody labeling (tube 2) yeast cells, then the antibody does not bind to the aglycosylated RBD and cannot be used with this method (see Figure 1B for expected results). Third, the competitive inhibition reaction should result in minimal PE signal corresponding to ACE2 binding (tube 3). If a PE signal increase is observed, the antibody does not compete with ACE2 for binding to the RBD and thus, cannot be used with this method (Figure 1D).

*Alternatives:* The fluorescent channel associated with the binding event will be dependent on the choice of secondary label. We use phycoerythrin (PE) as a fluorophore as it is compatible with anti-c-myc FITC and utilizes a 488 nm laser. Other fluorophores can be used.

23. If the antibody binds to RBD and blocks ACE2 binding, co-transform the libraries into yeast and make yeast stocks:
    a. Centrifuge the cell culture 3 min at 3200 × *g*.
    b. Remove the supernatant and resuspend the cell pellet in yeast storage buffer (recipe) at $OD_{600}$=10. Stocks can be stored for over a year at −80°C.

### Prepare operating system for running software with its dependencies

Timing: 20 min

To identify potential escape mutations from deep sequencing data, we developed Python software as outlined in Francino-Urdaniz et al. (2021). The steps below identify prerequisites necessary for running the software.

24. Download and install Python3 if it is not already installed. Instructions for this process can be found at https://www.python.org/downloads/. We recommend using Python v3.9 and higher. While older releases may work, the software has only been tested on recent versions of Python.
25. Install necessary packages.
    a. Ensure pip is installed with Python3 with the following command at the command line: pip3 –version. Pip is often installed with newer versions of Python. If pip is not downloaded, further instructions for installation can be found at https://pip.pypa.io/en/stable/installation/.
    b. Use pip to download the necessary packages: matplotlib, numpy, openpyxl, pandas, scipy, and statsmodels. These packages can be installed easily at the command line with pip3 –install package_name. More detailed instructions are at https://packaging.python.org/tutorials/installing-packages/. Download our software available on GitHub: https://github.com/WhiteheadGroup/SpikeRBDStabilization.git. A complete description of the software and its structure are provided in step-by-step methods.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Anti-c-myc FITC, 0.6 µL per 1e5 cells (80× dilution) | Miltenyi Biotec | Cat# 130-116-485 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Goat anti-human IgG Fc secondary antibody, PE, eBiosciences, 0.25 µL per 1e5 cells (200× dilution) | Invitrogen | Cat# 12-4998-82 |
| **Bacterial and virus strains** | | |
| *E. coli* Mach1™ | Thermo Scientific | Cat# C862003 |
| **Chemicals, peptides, and recombinant proteins** | | |
| ACE2-Fc | Institute of Protein Design, Laboratory of Professor Neil King | Walls et al., 2020 |
| NotI-HF | NEB | Cat# R3189 |
| BsaI-HFv2 | NEB | Cat# R3733 |
| CutSmart® Buffer | NEB | Cat# B7204S |
| Nuclease-Free Water | IDT | Cat# 11-05-01-14 |
| Ultrapure Agarose | Invitrogen | Cat# 16500-500 |
| TAE Buffer (Tris-acetate-EDTA) (50×) | Thermo Scientific | Cat# B49 |
| SYBR Safe DNA Gel Stain | Invitrogen | Cat# S33102 |
| Gel Loading Dye, Purple | NEB | Cat# B7024 |
| UltraPure™ Salmon Sperm DNA Solution | Invitrogen | Cat# 15632011 |
| Glycerol | Macron™ Chemicals | Cat# 5092-16 |
| HEPES-free acid | Millipore | Cat# 391338 |
| HEPES sodium salt | Amresco | Cat# 0485 |
| PEG 3350 | Spectrum | Cat #P0125 |
| Lithium acetate dihydrate | Sigma-Aldrich | Cat# L6883 |
| PBS - phosphate-buffered saline (10×) pH 7.4 | Invitrogen | Cat# AM9624 |
| Streptavidin phycoerythrin (SAPE) | Invitrogen | Cat# S866 |
| Bovine Serum Albumina (BSA), Fraction V, Fatty Acid-Free | VWR | Cat# 7907-25 |
| EZ-Link NHS-Biotin | Thermo Fisher | Cat# 20217 |
| Zymolyase | Zymo Research | Cat# E1005 |
| Sodium chloride | Sigma-Aldrich | Cat# 746398 |
| Difico yeast nitrogen base without amino acids | Sigma-Aldrich | Cat#Y026 |
| Bacto casamino acids, technical grade | Fisher | Cat# 223120 |
| Sodium phosphate dibasic anhydrous | Fisher Chemical | Cat# BP3321 |
| Sodium phosphate monobasic monohydrate | Fisher Chemical | Cat# S369-500 |
| D-Galactose | Fisher Bioreagents | Cat# BP656-500 |
| Dextrose | Fisher Chemical | Cat# D19212 |
| Peptone | Fisher | Cat# 211677 |
| Yeast extract | Fisher | Cat# 212750 |
| Agar | BD Biosciences | Cat# 214010 |
| Pen/Strep | Fisher | Cat# 15140-122 |
| Kanamycin | GoldBio | Cat# K-120-25 |
| Exonuclease I | NEB | Cat# M0293S |
| Lambda exonuclease | NEB | Cat# M0262S |
| Lambda exonuclease reaction buffer 10× | NEB | Cat# B0262S |
| Q5 HotStart 2× MasterMix | NEB | Cat# M0494L |
| rSAP | NEB | Cat# M0371L |
| 70% v/v Denatured ethanol solution | Fisher Bioreagents | Cat# BP82031GAL |
| IDTE pH 8.0 (1× TE Solution) | IDT | Cat# 11-05-01-13 |
| Quant-iT™ PicoGreen™ dsDNA Reagent | Thermo Scientific | Cat# T7581 |
| Lambda DNA | Thermo Scientific | Cat# SD0011 |
| **Critical commercial assays** | | |
| Monarch PCR & DNA Cleanup Kit | NEB | Cat# T1030 |
| Monarch DNA Gel Extraction Kit | NEB | Cat# T1020 |
| Monarch Plasmid Miniprep Kit | NEB | Cat# T1010 |
| Zymoprep Yeast Plasmid Miniprep II | Zymo Research | Cat# D2004 |
| Agencourt AMPure XP | Beckman Coulter | Cat# A63881 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| PhiX | Illumina | Cat# FC-110-3001 |
| Experimental models: Organisms/strains | | |
| *Saccharomyces cerevisiae* EBY100 | ATCC | MYA-4941™ |
| Oligonucleotides | | |
| IFU-104 | Francino-Urdaniz et al., 2021 | gttcagagttctacagtccgacgatctggaggaggctctgg |
| IFU-105 | Francino-Urdaniz et al., 2021 | ccttggcacccgagaattccaccaagctataacgcagcc |
| IFU-106 | Francino-Urdaniz et al., 2021 | gttcagagttctacagtccgacg atcggctgcgttatagcttgg |
| IFU-107 | Francino-Urdaniz et al., 2021 | ccttggcacccgagaattccagccccctttgtttttaaccaa |
| Forward outer primer | Kowalsky et al., 2015 | aatgatacggcgaccaccgaga tctacacgttcagagttctacagtccga |
| Reverse outer primer | Kowalsky et al., 2015 | caagcagaagacggcatacgagatnnnnnn gtgactggagttccttggcacccgagaattcca |
| Recombinant DNA | | |
| pJS697 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699 | Banach et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699_L1 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pJS699_L2 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001_T1 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU001_T2 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002_T1 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| pIFU002_T2 | Francino-Urdaniz et al., 2021 | https://www.addgene.org/Timothy_Whitehead/ |
| Software and algorithms | | |
| Python software (dms and analysis modules) | Francino-Urdaniz et al., 2021 | https://github.com/WhiteheadGroup/SpikeRBDStabilization.git |
| Python3 | N/A | https://www.python.org/ |
| Benchling | N/A | https://www.benchling.com |
| Other | | |
| −20°C Freezer | VWR | N/A |
| −80°C Freezer | Fisher Scientific, model: REVCO EXF | N/A |
| Pipettes | N/A | N/A |
| Centrifuge (does not need to be refrigerated) | Eppendorf, model: 5810R | Cat# 05-413-113 |
| Microcentrifuge (does not need to be refrigerated) | Fisher, model: accuSpin microcentrifuge 17 | Cat# 13-100-675 |
| Vortex mixer | Thermo Scientific, model: M37165 | Cat# M16710-33Q |
| Static incubator | VWR, model: Gr Con 4CF | Cat# 89511-420 |
| Incubator shaker | Eppendorf, model: New Brunswick I26 Inc Shaker | Cat# M1324-0000 |
| UV-Vis absorbance plate reader | BioTek, model: Synergy H1M | N/A |
| Spectrophotometer | Thermo Fisher Scientific, model: 4001/4 | Cat# 4001 |
| Cell sorter with a 488 nm laser | SONY, model: LE-SH800SAP | N/A |
| Thermal cycler | Eppendorf, model: Mastercycler™ pro | Cat# 950040025 |
| Horizontal Gel Electrophoresis System | Bio-Rad, model: Wide Mini-Sub Cell GT with PowerPac Basic Power Supply | Cat# 1640301 |

## MATERIALS AND EQUIPMENT

| Yeast extract-peptone-dextrose (YPD) media | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Yeast extract | 10 g/L | 10 g |
| Peptone | 20 g/L | 20 g |
| Dextrose | 20 g/L | 20 g |
| $ddH_2O$ | n/a | to 1 L |
| **Total** | **n/a** | **1 L** |

Filter sterilize using a 0.22 μm filter. Store at 20°C–25°C, protected from light for up to one month.

| Yeast storage buffer | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Glycerol ≥ 99.5% | 20% v/v | 200 mL |
| HEPES | 10 mM | 2.38 g |
| HEPES sodium salt | 10 mM | 2.60 g |
| NaCl | 200 mM | 11.69 g |
| $ddH_2O$ | n/a | to 1 L |
| **Total** | **n/a** | **1 L** |

Take to pH7.5 with NaOH.
Filter sterilize using a 0.22 μm filter and store at 20°C–25°C for up to one year.

| SCAA media | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Difico yeast nitrogen base | 6.7 g/L | 3.35 g |
| Bacto casamino acids | 5 g/L | 2.5 g |
| $NaH_2PO_4 \cdot H_2O$ | 62 mM | 4.28 g |
| $Na_2HPO_4$ | 38 mM | 2.7 g |
| $ddH_2O$ | n/a | to 450 mL |
| **Total** | **n/a** | **450 mL** |

Filter sterilize using a 0.22 μm filter and store at 20°C–25°C for up to six months.

⚠ CRITICAL: Add $NaH_2PO_4 \cdot H_2O$ before $Na_2HPO_4$ to ensure full dissolution of the phosphates.

| 20% w/v dextrose | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Dextrose | 20% w/v | 100 g |
| $ddH_2O$ | n/a | to 500 mL |
| **Total** | **n/a** | **500 mL** |

Filter sterilize using a 0.22μm filter and store at 20°C–25°C for up to six months.

| 20% w/v galactose | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Galactose | 20% w/v | 100 g |
| $ddH_2O$ | n/a | to 500 mL |
| **Total** | **n/a** | **500 mL** |

Filter sterilize using a 0.22μm filter and store at 20°C–25°C for up to six months.

| SDCAA+ | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| 20% w/v dextrose | 20g/L | 10 mL |
| SCAA | n/a | 90 mL |
| 100× Pen/Strep | 1× | 1 mL |
| 50 mg/mL Kanamycin | 50 μg/mL | 100 μL |
| **Total** | **n/a** | **101.1 mL** |
| Make on the same day of use and keep at 4°C. | | |

| SGCAA+ | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| 20% w/v galactose | 20 g/L | 10 mL |
| SCAA | n/a | 90 mL |
| 100× Pen/Strep | 1× | 1 mL |
| 50 mg/mL Kanamycin | 50 μg/mL | 100 μL |
| **Total** | **n/a** | **101.1 mL** |
| Make on the same day of use and keep at 4°C. | | |

## SDCAA agar plates

| Solution A | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| $NaH_2PO_4 \cdot H_2O$ | 62 mM | 4.28 g |
| $Na_2HPO_4$ | 38 mM | 2.7 g |
| Agar | 15 g/L | 7.5 g |
| $ddH_2O$ | n/a | to 450 mL |
| **Total** | **n/a** | **450 mL** |
| Autoclave 20 min at 121°C and 0.5 bar gauge. | | |

| Solution B | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| Difico yeast nitrogen base | 6.7 g/L | 3.35 g |
| Bacto casamino acids | 5 g/L | 2.5 g |
| Dextrose | 20% w/v | 10 g |
| $ddH_2O$ | n/a | to 50 mL |
| **Total** | **n/a** | **50 mL** |
| Filter sterilize using a 0.22μm filter. | | |

Cool autoclaved mixture (**Solution A**) with stirring until below 50°C, add filter-sterilized solution (**Solution B**), mix well, and pour plates. Plates can be stored for up to 6 months at 4°C.

| PBSF | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| bovine serum albumin frac V | 1 g/L | 100 mg |
| PBS | n/a | 100 mL |
| **Total** | **n/a** | **100 mL** |
| Filter sterilize using a 0.22μm filter and store at 4°C for no longer than 3 days. | | |

| Fluorescent labeling mix | | |
|---|---|---|
| Reagent | Final concentration | Amount |
| anti-c-myc FITC | n/a | 0.60 μL |
| SAPE | n/a | 0.25 μL |
| PBSF | n/a | 49.15 μL |
| **Total** | **n/a** | **50 μL** |

Make on same day of use and store on ice, protected from light. We recommend making a master mix, adding an extra 10% for each reaction.

## STEP-BY-STEP METHOD DETAILS

### Cell induction

Timing: 26 h

S RBD is expressed in yeast from a galactose-inducible promoter. The first step is to expand cells from a frozen stock and then induce S RBD display by growth on galactose.

1. Thaw 1mL stock of yeast cells harboring the S RBD display plasmid of the desired library for identifying escape mutants and WT for controls (Confirm that antibodies selected bind and inhibit the displayed WT RBD). It is convenient to start this growth step in the morning so that cytometer screening can be performed during daytime hours on the following day.
2. Pellet the cells by centrifuging 1 min at 16000 × *g* using a microcentrifuge, remove supernatant, and resuspend in $SDCAA^+$ (recipe).
3. Transfer culture to a 14 mL culture tube for volumes lower than 1.5 mL (VWR Cat#10127-334) or a 250 mL shaker flask for large volumes (Fisher Cat# 10040F). Incubate cells at 30°C and 300rpm for 4–6 h.

   *Note:* Do not grow cultures of more than 1.5 mL volume in 14 mL tubes, as they are insufficiently aerated for optimal display.

4. Measure the culture $OD_{600}$ with a spectrophotometer. If $1.2 < OD_{600} < 5$, harvest the cells by centrifugation for 1 min at 16000 × *g* in a microcentrifuge and resuspend to $OD_{600}$=1 in $SGCAA^+$ (recipe).
5. Transfer the appropriate amount of the culture to a new 14mL culture tube or a 250mL shaker flask as necessary and incubate 22h at 22°C and 300rpm in an incubator shaker. As previously stated, we recommend using 14 mL culture tubes for the volumes lower than 1.5 mL and 250 mL shaker flasks for larger volumes.

### Competitive binding reactions

Timing: 4 h

Yeast cells displaying nearly all possible single point mutants on surface exposed positions of S RBD are pre-incubated with a neutralizing antibody before co-incubation with biotinylated ACE2. These cells are then labeled with secondary fluorophores to detect S RBD expression and ACE2 binding. Fluorescence activated cell sorting (FACS) is used to isolate the population of cells still able to bind ACE2.

⚠ CRITICAL: The binding reactions cannot be stored to sort another day. If multiple antibodies are being tested, we recommend splitting the samples into multiple days.

6. Transfer the culture to a 1.5 mL microcentrifuge tube for low volumes or a 15 mL conical tube for larger volumes and centrifuge for 1 min at 16000 × *g* using a centrifuge.
7. Measure the culture $OD_{600}$ and resuspend the cells in ice cold PBSF to an $OD_{600}$=2 for the controls and to an $OD_{600}$=10 for sorting. We typically perform a 10-fold dilution of culture to determine the $OD_{600}$. The yeast culture will not necessarily increase in $OD_{600}$ during this 16–24 h induction.

*Note:* Keep this suspension on ice during the whole procedure except when otherwise specified.

### Screen unlabeled cells

8. Combine 5 μL cells with 45 μL of PBSF ('unlabeled' cells) in a 1.5 mL microcentrifuge tube. Incubate 30 min at 20°C–25°C.
9. Centrifuge 1 min at 16000 × *g* using a microcentrifuge and resuspend the cells in 100μL ice cold PBSF. Screen 25,000 unlabeled cells and set the FSC-A/$FITC^+$ gate as in Figure 2B.

### Check display of S RBD

10. Label 5 μL cells with 44 μL of PBSF and 1 μL of anti-c-myc FITC in a 1.5mL microcentrifuge tube. Mix by pipetting up and down several times. Incubate the cells on ice for 10 min protected from light.
11. Centrifuge the cells 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 μL ice cold PBSF and remove supernatant. Cell pellets can be stored on ice until ready to screen.
12. Resuspend the cells in 100μL ice cold PBSF and screen 25,000 cells. If the yeast does not display the S RBD, do not prepare the binding reactions and refer to troubleshooting 1.

**Check unlabeled cells, ACE2 binding, and competitive inhibition.** These are the recommended controls for WT and each library sorted. The unlabeled cells will be used to set the correct gates on FACS for each different library sorted. The ACE2 binding is the positive control and the competitive inhibition the negative control.

13. Set up three separate labeling reactions in three different 1.5 mL microcentrifuge tubes using the cells at $OD_{600}$=2. Incubate 30 min at 20°C–25°C. Ensure that the biotin:protein molar ratio has been adjusted before starting the escape mutant identification, else, refer to 'Confirm that neutralizing antibodies selected bind and inhibit the yeast displayed glycosylated WT RBD'.

| | Tube 1 | Tube 2 | Tube 3 |
|---|---|---|---|
| Library Cells | 5μL | 5μL | - |
| WT Cells | - | - | 5μL |
| Biotinylated ACE2 | - | 75nM | - |
| Free ACE2 | - | - | 10μg/mL |
| PBSF | To 50μL total volume | To 50μL total volume | To 50μL total volume |
| | Incubate for 30 min at 20°C–25°C. | | |
| Biotinylated ACE2 | - | - | 75nM |
| | - | - | Incubate for 30 min at 20°C–25°C. |

a. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 μL PBSF and remove supernatant.
b. Resuspend the second and third pellet with 50 μL fluorescent labeling mix (recipe). The first tube must not be labeled. Keep the fluorophores on ice and protected from light at all times.

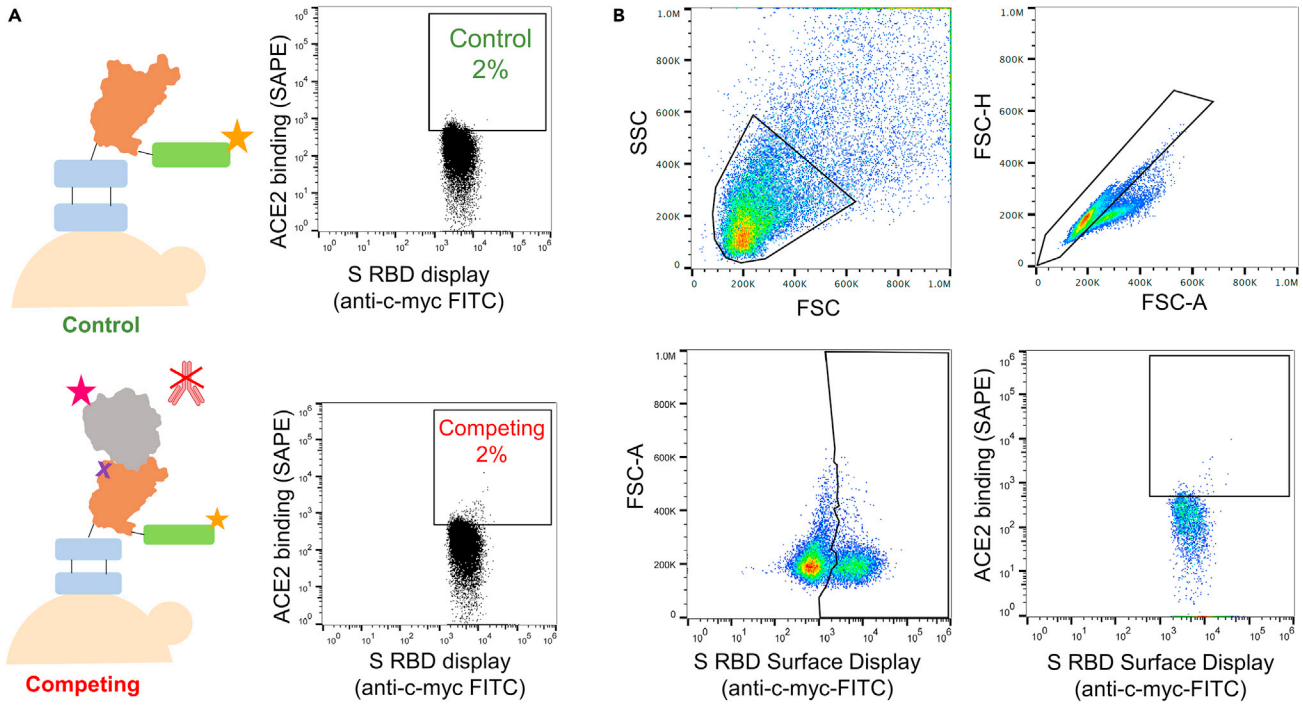



**Figure 2. Sort gates needed for FACS screening of potential S RBD escape mutants**

(A) Competitive binding reaction cartoons between neutralizing antibody and ACE2. Top is the control where the displayed S RBD is not labeled, top right shows the $SAPE^+/FITC^+$ gate collecting the top 2%. Bottom shows escape mutants that no longer bind to nAb but keep the affinity for ACE2. Bottom right shows the $SAPE^+/FITC^+$ gate collecting the top 2%.

(B) Collection gates on FACS. FSC/SSC+ gate for isolation of yeast cells, FSC-H/FSC-A gate to discriminate single cells, an FSC-A/$FITC^+$ gate selects the cells displaying the RBD on their surface and from this last gate, the top 2% by a $PE^+/FITC^+$ is collected.

c. Incubate 10 min on ice and protect from light.
d. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 200 µL PBSF and remove supernatant. The cell pellet can be stored on ice until ready to screen. We do not recommend storing on ice for longer than 1h as fluorophores might dissociate, resulting in a lower signal.
e. Resuspend the cells in 100 µL PBSF and screen 25,000 cells.

### Competitive binding reactions to identify escape mutants

14. In a 1.5mL microcentrifuge tube add 150 µL of cells, PBSF, and 10 µg/mL antibody to a final volume of 400 µL. Incubate 30 min at 20°C–25°C. Then, add biotinylated ACE2 to a final concentration of 75nM. Again, incubate 30 min at 20°C–25°C (Figure 2A).

⚠ CRITICAL: If multiple samples are sorted, it is better to incubate pellets on ice before the next step. Fluorescent label right before sorting to ensure maximum fluorescence. To minimize binding loss, stagger the binding reactions to minimize the storage time so that when the reaction is completed, the sample can be sorted.

a. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 1 mL PBSF and remove supernatant. Cell pellets can be stored on ice for ~4 h without significant binding loss if all the supernatant is removed.
b. Resuspend each cell pellet with 200 µL fluorescent labeling mix (recipe). Keep the fluorescent labeling mix on ice and always protected from light.

⚠ CRITICAL: Better results are obtained if the cells are fluorescently labelled right before sorting.

c. Incubate 10 min on ice and protect from light.
d. Centrifuge 1 min at 16000 × *g* using a microcentrifuge. Wash twice with 1 mL PBSF and remove supernatant. Cell pellets can be stored covered and on ice until ready to screen.

*Optional:* While sorting, the fluorescence signal might decrease over time. If the decrease is significant, divide the sample into two 500 μL samples and spin down again, remove supernatant and spin down to make sure no liquid is left in the sample. Store each pellet on ice until ready to use.

### Fluorescence-activated cell sorting (FACS)

Timing: approx. 40 min per sorted sample

Cells are sorted by FACS using four separate gates (Figure 2). We first set an SSC/FSC gate for isolation of yeast cells. The second gate set is on FSC-H/FSC-A to discriminate single yeast cells from budding cells and/or aggregates. The third gate is an FSC-A/FITC$^+$ gate that selects the cells displaying the RBD on their surface and excludes cells not displaying RBD. The fourth gate is a square gate collecting the top 2% by a PE$^+$/FITC$^+$ (Figure 2B). The first three gates are set using unlabeled cells as described below. The collection PE$^+$/FITC$^+$ gate is adjusted for each sorted sample to collect the cells giving the top 2% PE fluorescence signal. We set a 2% limit for collecting the population of the PE fluorescence signal for two reasons. First, collecting the entire population would require collecting considerably more cells to obtain the same quality in results since different PE signal increases correlate with varying affinity. In addition, collecting the top 2% allows us to set a false discovery rate (FDR) given that cells with higher signal are likely to contain escape mutants.

15. Using the unlabeled cells, set the FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$ gates as shown in Figure 2B.
16. Using the same unlabeled cells, obtain the reference population by collecting at least 200,000 cells by sorting only for passing through the FSC/SSC gate. This reference population contains all the mutations in the library. It will be the reference to calculate the enrichment ratio for each mutation. For this and all other cell populations, collect cells into a 14mL culture tube (VWR Cat#10127-334).
17. Sort the control population. Here we use cells that have not labeled with antibody or ACE2 but have been fluorescently labeled with SAPE and anti-c-myc FITC. Collect 200,000 cells that pass through FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$, and the PE$^+$/FITC$^+$ gates.
18. For every other sample, collect 200,000 cells that pass through FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$, and the PE$^+$/FITC$^+$ gates. The square PE$^+$/FITC$^+$ gates will have to be reset for each sample to ensure 2% collection.

*Note:* Using our Sony SH-800 cytometer, it typically takes 40min to collect 200,000 cells in the control and sample populations. This time results from a sorting rate of 5-10,000 cells per sec and maintaining a 70% sorting efficiency.

*Note:* To avoid complexity bottlenecks, it is important that the number of collected cells is at least 150× the theoretical library size. For a library with NNK mutations on 56 and 63 mutations for library 1 and 2 respectively, 168,000 cells for library 1 and 189,000 cells for library 2 must be collected. Results are improved with more collected cells.

### Cell recovery

Timing: 3 days

Cells are recovered and yeast stocks are prepared.

Day 1

19. Centrifuge collection tube containing the collected cells 5 min at 3200 × *g* using a centrifuge. Collected cells are smaller than typical yeast bulk population, and therefore a longer centrifugation time of 5 min is required.
20. Remove approx. 80% of the supernatant and resuspend in 1.5mL SDCAA$^+$ at 30°C for 16–20 h at 300rpm in an incubator shaker in the 14mL culture tubes.

*Note:* 14mL culture tubes cannot hold more than 1.5mL. Larger volumes are not well aerated. To use a larger resuspension volume, grow the culture in a 25 mL shaker flask.

Day 2

21. Passage cells into 3 culture tubes using 1.5mL SDCAA$^+$ on each. Incubate at 30°C and 300rpm for 16–24 h.

Day 3

22. Prepare 1mL cell stocks at $OD_{600}$=4 in yeast storage buffer at −80°C.
    a. Centrifuge the cell culture 2 min at 3200 × *g*.
    b. Remove the supernatant and resuspend the cell pellet in yeast storage buffer (recipe) at $OD_{600}$=4. Stocks can be stored for over a year at −80°C. In total, at least 1 mL cell stock at $OD_{600}$=4 is needed for deep sequencing preparation.

## Deep sequencing library preparation

Timing: 1–2 days

After 16–24 h of growth, plasmid DNA from this population is isolated and prepared for deep sequencing. Plasmid DNA from cells is extracted and amplicons prepared for Illumina sequencing (Figure 3). The protocol used is adapted from Method B detailed in Kowalsky et al. (2015).

Day 1: Yeast miniprep

23. Perform a yeast miniprep.
    a. On ice, thaw a cell stock prepared in step 23.
    b. Pellet the cells by centrifuging 1 min at 16000 × *g* using a microcentrifuge and resuspend in 200 µL Solution 1 from Zymo Yeast Plasmid Miniprep II kit.
    c. Add 5µL Zymolyase (5U/µL).
    d. Incubate in a 37°C incubator for 4h on an end-to-end mixer or incubate at 37°C with mixing by pipetting up and down (ten times) every hour.
    e. Perform 1 freeze-thaw cycle using a dry ice/EtOH bath followed by a 42°C incubation for 10 min. Alternatively, freeze at −80°C for 20 min and thaw at 42°C in a temperature bath.
    f. Add 200 µL Solution 2 from Zymo Yeast Plasmid Miniprep kit, mix end-over-end and let sit for 5 min. The solution will be transparent.
    g. Add 400 µL Solution 3 from Zymo Yeast Plasmid Miniprep kit mix end-over-end centrifuge 5 min at 16000 × *g*. The solution will turn yellow, and precipitate will form.
    h. Transfer supernatant to a NEB miniprep column.

*Note:* The NEB miniprep column has a higher DNA binding capacity (15 µg) compared to the Zymo miniprep columns (5 µg). The rationale for using NEB miniprep columns is that lysed

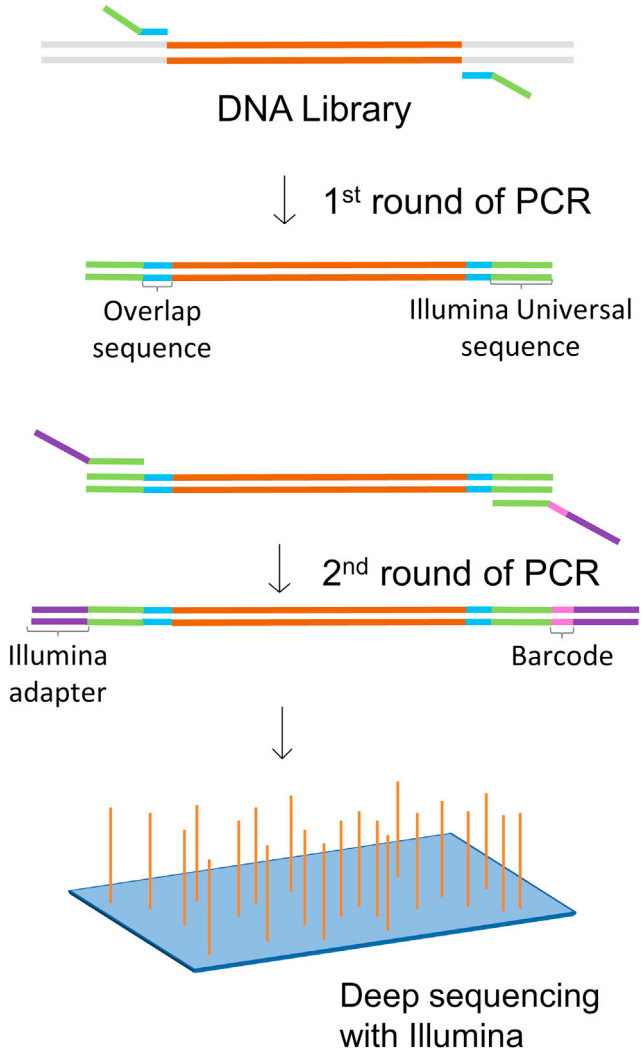



**Figure 3. Schematic of deep sequencing library preparation**

Barcodes and Illumina adapters are added to the DNA library through two rounds of PCR. The amplicon is sequenced using an Illumina MiSeq.

cells contain large amounts of sheared yeast genomic DNA in addition to the desired plasmid DNA.

i. Centrifuge 1 min at 16000 × *g* using a microcentrifuge.
j. Add 700 µL Wash 1 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 × *g*.
k. Add 700 µL Wash 2 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 × *g*.
l. Add again 700 µL Wash 2 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 × *g*.
m. Decant the supernatant and centrifuge 1 min at 16000 × *g* to dry the column.
n. Add 30 µL elution buffer from NEB mini-prep kit to elute the DNA and centrifuge 1 min at 16000 × *g*.
o. Reload column with eluate and spin again.

24. Enzymatic cleanup of yeast genomic DNA. In PCR tubes add in order:

| Reagent | Final concentration | Amount |
|---|---|---|
| Lambda buffer 10× | n/a | 2 μL |
| Miniprep DNA | n/a | 15 μL |
| Exonuclease I | n/a | 2 μL |
| Lambda exonuclease | n/a | 1 μL |
| **Total** | **n/a** | **20 μL** |

Thermocycler conditions

| Steps | Temperature | Time | Cycles |
|---|---|---|---|
| Step 1 | 30°C | 1:30 h | 1 |
| Step 2 | 80°C | 20 min | 1 |
| Hold | 4°C | Forever | |

25. Plasmid cleanup. Using Monarch PCR & DNA cleanup kit according to the manufacturer instructions. Elute in 30 μL elution buffer.

Can divide into two days or continue same day from here.

### Day 2: Sequencing sample preparation

26. The Illumina universal sequences are attached to the amplicon with an overlapping region included in the primers (Figure 3) using a first PCR.

Primers used for first PCR (corresponding to "Overlap sequence" on Figure 3):

| Name | Description | Sequence | Section of RBD amplification | Library compatibility |
|---|---|---|---|---|
| IFU-104 | L1_Inner_FWD | gttcagagttctacagtccgacgatctggaggaggctctgg | 333-437 | pJS699_L1, pIFU001_T1 and pIFU002_T1 |
| IFU-105 | L1_Inner_REV | ccttggcacccgagaattccaccaagctataacgcagcc | | |
| IFU-106 | L2_Inner_FWD | gttcagagttctacagtccg acgatcggctgcgttatagcttgg | 438-537 | pJS699_L2, pIFU001_T2 and pIFU002_T2 |
| IFU-107 | L2_Inner_REV | ccttggcacccgagaattcca gccccctttgtttttaaccaa | | |

| Reagent | Final concentration | Amount |
|---|---|---|
| DNA (from PCR cleanup) | n/a | 15 μL |
| Fwd inner primer (10μM) | n/a | 2.5 μL |
| Rev inner primer (10μM) | n/a | 2.5 μL |
| 2× Q5 Master mix | n/a | 25 μL |
| $ddH_2O$ | n/a | 5 μL |
| **Total** | **n/a** | **50 μL** |

1st PCR cycling conditions

| Steps | Temperature | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 98°C | 1 min | 1 |
| Denaturation | 98°C | 10 s | 25 cycles |
| Annealing | 64°C | 20 s | |
| Extension | 72°C | 1 min | |
| Final extension | 72°C | 2 min | 1 |
| Hold | 4°C | forever | |

27. Perform a PCR cleanup to remove unwanted single stranded DNA.

To the PCR products add:

| Reagent | Final concentration | Amount |
|---|---|---|
| PCR Products | n/a | 50 μL |
| Exonuclease I | n/a | 5 μL |
| rSAP | n/a | 10 μL |
| **Total** | **n/a** | **65 μL** |

Thermocycler conditions (PCR cleanup)

| Steps | Temperature | Time | Cycles |
|---|---|---|---|
| Step 1 | 37°C | 20 min | 1 |
| Step 2 | 80°C | 20 min | 1 |
| Hold | 4°C | forever | |

28. Attach the barcodes and the Illumina adapter with a 2nd PCR reaction. The barcode must be unique for each sample to allow the analysis of different samples on a single MiSeq run.

Primers used on 2nd PCR (corresponding to the Illumina primers – "Illumina adapter" and "Barcode" – in Figure 3):

| Description | Sequence |
|---|---|
| Illumina_FWD (RP1) | aatgatacggcgaccaccgagatcta cacgttcagagttctacagtccga |
| Illumina_REV | caagcagaagacggcatacgagat**nnnnnn** gtgactggagttccttggcacccgagaattcca |

*Note:* **nnnnnn** refers to unique RPI Illumina barcode (full list: Illumina Adapter Sequences and Kowalsky et al., 2015). In this study, we used adapters #1–20 but any set of adapters can be used.

| Reagent | Final concentration | Amount |
|---|---|---|
| $H_2O$ | n/a | 18 μL |
| DNA (from cleaning step) | n/a | 2 μL |
| Fwd outer primer (10μM) | n/a | 2.5 μL |
| Rev outer primer (10μM) | n/a | 2.5 μL |
| 2× Q5 MM | n/a | 25 μL |
| **Total** | **n/a** | **50 μL** |

2nd PCR cycling conditions

| Steps | Temperature | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 98°C | 1 min | 1 |
| Denaturation | 98°C | 10 s | 25 cycles |
| Annealing | 64°C | 20 s | |
| Extension | 72°C | 1 min | |
| Final extension | 72°C | 2 min | 1 |
| Hold | 4°C | forever | |

29. Gel extract the amplified DNA using Monarch Gel Extraction kit. Select the approx. 500bp band. A clear band should be seen with the correct size (515 bp for Tile 1 and 487bp for Tile 2) on the SYBR safe gel. If this band is not seen, refer to troubleshooting 3. If products cannot be seen in a SYBR safe gel, there will not be enough DNA for sequencing (minimum 2ng DNA is required).
30. Clean up with Agencourt AMPure XP following standard procedure on a 96 well format (Instructions For Use).

31. Quantification with Quant-iT™ PicoGreen™ dsDNA Reagent following standard procedure (Quant-iT™ PicoGreen ® dsDNA Reagent and Kits).
    a. Prepare Quant-iT$^{TM}$ reagent.
        i. Thaw Quant-iT$^{TM}$ reagent at 20°C–25°C covered in foil.
        ii. Prepare a working solution by diluting the Quant-iT$^{TM}$ reagent 200-fold in TE.

⚠ CRITICAL: The working solution must be prepared on the same day of the quantification and stay protected from the light, preferably wrapped in aluminum foil.

    b. In a black 96-well plate (Product # 3792), prepare a 1:2 standard curve starting at 100ng/mL of Lambda DNA using the first row of the plate. Add 100 µL in each well and include a blank with 100 µL TE.
    c. In the same plate, add 1 µL of purified sample in 99 µL TE.

*Note:* If the sample concentration is < 40 ng/µL, up to 2.5 µL of sample can be used. On the contrary, if the concentration is too high, necessary dilutions in IDTE can be carried out.

    d. Add 100 µL Quant-iT$^{TM}$ working solution to each well and incubate for 5 min at 20°C–25°C while covered in foil.
    e. Measure the fluorescence using a plate reader (ex: 480nm, em: 520 nm). Subtract the blank from all the samples and use the Lambda DNA serial dilutions to generate a standard curve of fluorescence vs DNA concentration. With the curve, determine the concentration of each sample. In our hands the concentrations are between 6 ng/µL and 60 ng/µL.

32. Pool individual barcoded amplicons into a single test tube. In a 1.5 mL tube, mix equivalent molar amounts of each sample to have the same number of reads and submit to a sequencing facility.

### Illumina sequencing

Libraries can be sequenced on an Illumina MiSeq using the MiSeq V2-500 cycle kit. While we have used the University of Colorado sequencing core, there are commercial and academic cores that offer sequencing as a fee-for-service (NovoGene or Rush University Medical Center, for example). The libraries have been split into two tiles assuming the use of 250 bp paired end sequencing, but our software supports longer paired-end reads as well. One critical detail essential for sequencing is dealing with the low nucleotide diversity of the RBD libraries. We circumvent this issue primarily by instructing the sequencing core to add PhiX in the pooled library, which is a crucial step for correct clustering of the sequences. PhiX is a well characterized adapter-ligated library that helps balance nucleotide diversity. We have obtained good sequencing results with 35% PhiX. Refer to troubleshooting 4 if the DNA is not clustering correctly.

We typically obtain around 90% coverage for all the libraries screened.

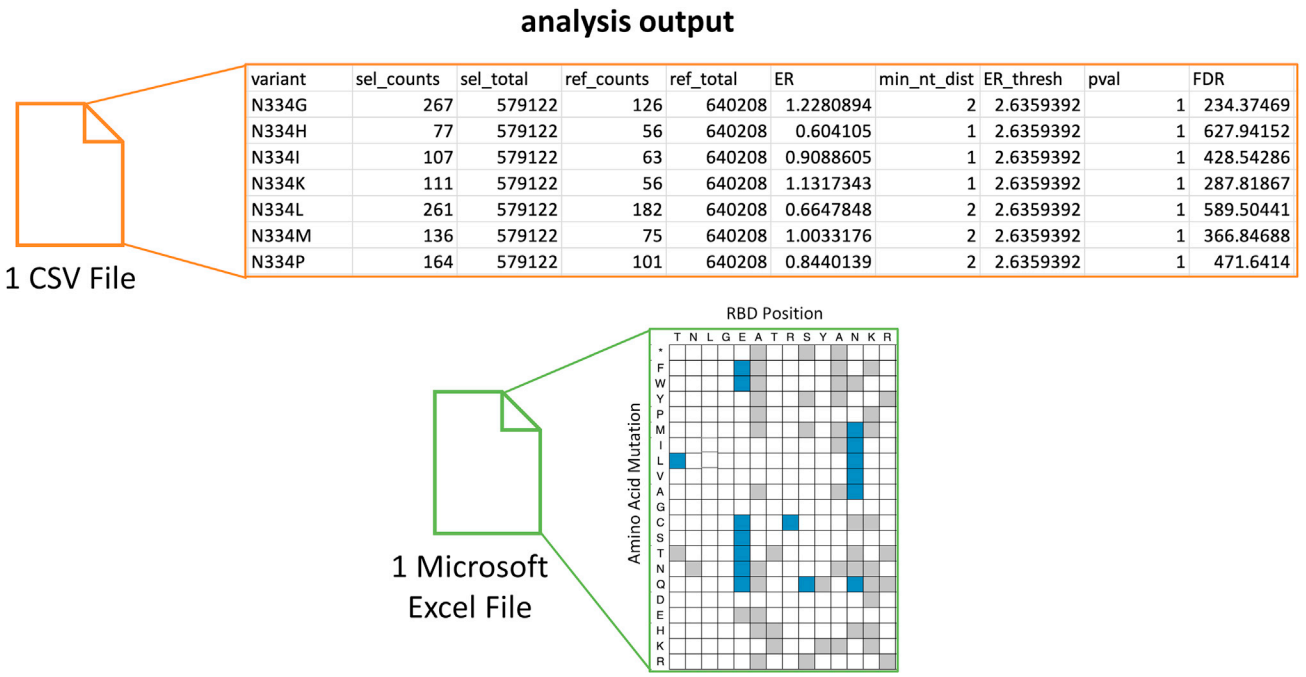
### Run software and generate results

Timing: <1 h

This is the primary computational step in the protocol. This software package can perform all necessary analysis and interpretation of results. With the completion of this step, potential escape mutations are identified for a given antibody.

#### Software description

This software is divided into two separate modules. The first module manages the input FASTQ files and is abbreviated 'dms' for deep mutational scanning. The second module, titled 'analysis', is




| variant | sel_counts | sel_total | ref_counts | ref_total | ER | min_nt_dist | ER_thresh | pval | FDR |
|---|---|---|---|---|---|---|---|---|---|
| N334G | 267 | 579122 | 126 | 640208 | 1.2280894 | 2 | 2.6359392 | 1 | 234.37469 |
| N334H | 77 | 579122 | 56 | 640208 | 0.604105 | 1 | 2.6359392 | 1 | 627.94152 |
| N334I | 107 | 579122 | 63 | 640208 | 0.9088605 | 1 | 2.6359392 | 1 | 428.54286 |
| N334K | 111 | 579122 | 56 | 640208 | 1.1317343 | 1 | 2.6359392 | 1 | 287.81867 |
| N334L | 261 | 579122 | 182 | 640208 | 0.6647848 | 2 | 2.6359392 | 1 | 589.50441 |
| N334M | 136 | 579122 | 75 | 640208 | 1.0033176 | 2 | 2.6359392 | 1 | 366.84688 |
| N334P | 164 | 579122 | 101 | 640208 | 0.8440139 | 2 | 2.6359392 | 1 | 471.6414 |

**Figure 4. An example of the analysis output from the software**

The CSV file is first generated, adding the results of statistical calculations. The second output from the analysis module is the Microsoft Excel document containing a heatmap with escape mutant residues identified. In the heatmap, wild type residues and positions are listed in the top two rows of the file and each possible mutation is listed in the left column. Three colors exist in each Excel output file: gray cells indicate that the mutation was not observed in the population; white cells indicate mutations that were observed in the population but did not meet the criteria for escape mutant; and, blue cells represent mutations that meet the necessary criteria to be classified as potential escape mutants for the given antibody.

responsible for interpreting the output generated from the dms module. Apart from specifying which experiment to analyze before the analysis module is run, the transition between these two modules is seamless.

This protocol results in a total of three output files. The dms module generates one of these files and the analysis module identifies the other two. The output from the dms module is a comma-separated values (CSV) file that identifies the single mutations, their corresponding counts in each population, and an enrichment ratio (ER) that compares these counts. Next, the analysis module generates a CSV file with additional columns to the file generated in the dms module. The new columns identify the following for each observed mutation: the threshold ER value determined from the control data for a given tile, the p-value calculated by the one-sided exact Poisson rate ratio test, the FDR value, and the minimum number of nucleotides away the given mutation could be from the wild type sequence. The second output of the analysis module is a heatmap in Microsoft Excel that identifies potential escape mutations in blue. Refer to Figure 4 for details on the outputs mentioned above.

The dms module assumes that amplicons are of the same fixed length and that each amplicon is read in both the forward and reverse directions, leading to a pair of FASTQ files. When the module is run, it merges the paired reads of the FASTQ files, collapses identical synonymous mutations into single counts, and subsequently calculates enrichment ratios between specified populations (see 'quantification and statistical analysis section'). A detailed explanation of this process has already been published (Fowler et al., 2011; Klesmith and Hackel, 2019).

The analysis module performs the necessary analytical steps to identify the escape mutations. While a thorough explanation of the statistical approach is detailed in Quantification and statistical analysis, please refer to Figure 5 for an overview of the entire process.

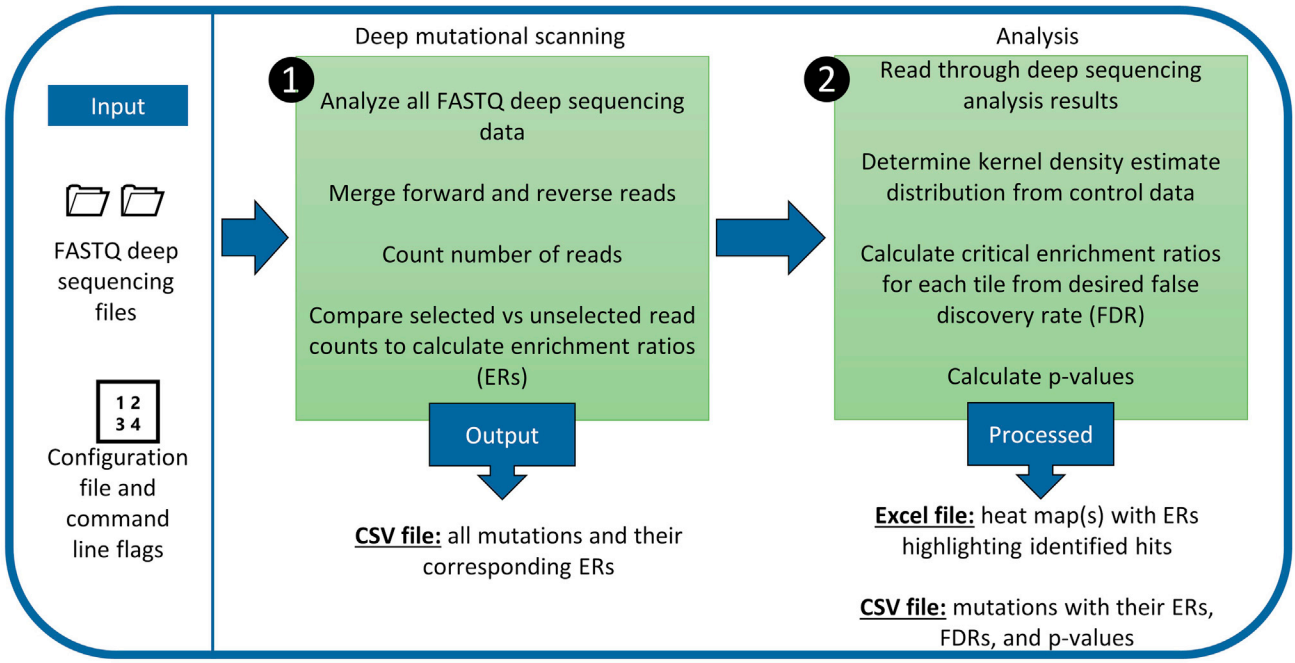



**Figure 5. The flow structure of the software**
For each module (Deep mutational scanning; Analysis) the actions taken and output are listed.

33. Complete the configuration file to include the details of the experiment.
    For the dms module, the user should fill out the configuration file to include all experiments. Therefore, the module can be run once and the analysis can be streamlined. An example configuration file is on Github. We recommend that this example file is copied and that the original document remains available for reference. For an explanation of each section and option in the configuration file, please refer to Tables 2 and 3.
    a. If the experiment of interest only uses one tile, refer to troubleshooting 5.

⚠ CRITICAL: The example configuration file is already filled out so that users only need to complete the file names, the file paths, and create names for their experiments. If the original libraries are used, no additional changes are needed to the configuration file. All other parameters and details are set as default and do not need to be changed for a successful run of the software.

34. Run the software by execution of commands at the command line. Some command line flags can be used here to override information set in the configuration file. Details about these command line arguments are in Table 4.

    a. Run the dms module. This module can be run with following command (assume the configuration file is titled *example.config* and that it is found in the root directory):

```
python3 -m dms -config example.config
```

    The dms module will deposit output files in a folder found in the root directory of the software unless the output directory is intentionally changed. The files are always deposited into the folder titled "Output". If that folder does not exist in the directory, an output folder will be created.

*Note:* The dms package is a general deep mutational scanning tool. It can be run independently of any other analysis and does not depend on the other module described here.

**Table 2. Descriptions of sections in the configuration file**

| Configuration file section | Description |
|---|---|
| Parameters | This section defines the basic parameters for the dms module.<br>Available options (described in Table 2): max_mismatches, min_quality, and fastq_file_dir. |
| Tile:Tx | This section defines a tile that will be analyzed. For every tile in an experiment, a new tile section should be created. To define a tile, change the name after the colon.<br>Available options (described in Table 2): wt_seq, first_aa, cds_start, cds_end, and positions.<br>**Note:** this section should generally remain unchanged from the example configuration file. The option to change these parameters is provided as a convenience for anyone following these protocols with different libraries. |
| Samples | This section defines the different samples in the experiment, connecting the tiles with the file names for the paired-end reads. All samples are in this section, no other [Samples] section is necessary to define multiple samples.<br>The title for each sample is chosen by the user and should be entered without quotation marks. Each sample must be unique, and all samples should be options under [Samples].<br>The samples must be defined as follows: **sample_name: 'Tx', 'FASTQ_filename1', 'FASTQ_filename2'**<br>'Tx' - the tile identifier that must match the tile name provided in the tile section.<br>'FASTQ_filename1' - the name of the FASTQ file that contains the sample's forward reads.<br>'FASTQ_filename2' - the name of the FASTQ file that contains the sample's reverse reads |
| Experiments | This section defines each experiment by giving it a name and identifying the reference and selected populations. All experiments are in this section, no other [Experiments] section is necessary to define multiple experiments.<br>The title for each experiment is chosen by the user and should be entered without quotation marks. Each experiment must be unique, and all experiments should be options under [Experiments].<br>The experiments must be defined as follows: **experiment_name: 'reference_sample', 'selected_sample'**<br>'reference_sample' - the name of the sample that contains the reference population<br>'selected_sample' - the name of the sample that contains the desired selected population<br>*Both 'reference_sample' and 'selected_sample' must be defined in the [Samples] section as options. The name of these samples must be copied exactly into quotes in the [Experiments] section. |
| Proteins | This section combines different tiles that correspond to the same protein. This is so the information across all tiles for a given protein can be consolidated into single concise files.<br>The title for each protein is chosen by the user and should be entered without quotation marks. Each protein must be unique, and all proteins should be defined under this section.<br>The proteins must be defined as follows: **protein_name: 'experiment_name1', 'experiment_name2', …, 'experiment_nameN'**<br>'experiment_name1' - the name of the experiment that is the first tile for the protein<br>'experiment_name2' - the name of the experiment that is the second tile for the protein<br>*All experiment_names must be defined in the [Experiments] section as options. The name of these samples must be compiled exactly into quotes in the [Proteins] section.<br>It is possible to define a protein with a single experiment, but the single quoted experiment name must be followed by a comma in the definition. |
| Analysis | This section defines statistical parameters for the analysis module. The information provided here results in the final escape mutant hits identified.<br>Available options (described in Table 2): control_filepath, antibody_filepath, output_title, FDR, significance. |

b. Run the analysis module with the following command (again, assume the configuration file is titled *example.config* and that it is found in the root directory). The [Analysis] section of the configuration file must be updated before every run to represent accurate experiments. However, all other sections of the configuration file should remain the same for both modules:

**Table 3. Descriptions of options for each section in the configuration file**

| Configuration file section | Configuration file options | Description |
|---|---|---|
| Parameters | max_ mismatches | The maximum number of mismatches allowed when reading overlapping sequences of paired-end reads from FASTQ deep sequencing files. Reads with a higher number of mismatches are discarded. The input must be a positive integer or None. If None, then no reads will be discarded due to mismatches. The default value used in our analysis is 10. |
| Parameters | min_quality | The minimum sequencing quality score required to keep a read in the analysis. A read with a score lower than this minimum benchmark will be discarded. The input must be a positive integer or None. If None, then no reads will be discarded due to quality. The default value used in our analysis is 10. |
| Parameters | fastq_file_dir | This option allows the user to specify the directory that contains the FASTQ files to be analyzed. The file path should be a folder that contains FASTQ files with the file names provided in the configuration file. An absolute file path should be used. The input must be a string. An example for this value is: '/Users/lab/Documents/fastq_files' |
| Tile:Tx | wt_seq | The nucleotide DNA sequence of the tile which includes the entire sequence of a read from deep sequencing. This generally includes the primers used. The input must be a string of capital letters. The default values used in our analysis is different for each Tile and shown below:<br>Tile 1: 'TGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGACAAACTTG<br>TGCCCTTTTGGTGAAGTTTTTCAAGCCACCAGATTTGCATCTGTTTATGCTTGGAAC<br>AGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCATTT<br>TCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTT<br>ACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAAT<br>CGCTCCAGGGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGA<br>TGATTTTACAGGCTGCGTTATAGCTTGG'<br>Tile 2: 'GGCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGT<br>GGTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAA<br>CCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCAC<br>ACCTTGTAATGGTGTTGAAGGTTTTAATTGTTACTTTCCTTTACAAT<br>CATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCATACAGAG<br>TAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGG<br>ACCTAAAAAGTCTACTAATTTGGTTAAAAACAAAGGGGGC' |
| Tile:Tx | first_aa | The residue number of the first amino acid in the protein sequence. The defaults used in our analysis are:<br>Tile 1: 333<br>Tile 2: 437 |
| Tile:Tx | cds_start | The nucleotide position in a merged read is where the coding sequence starts (after primers or any other miscellaneous nucleotides at the start of the read). Note: this index is given using Python slicing conventions, so it is zero-indexed. For example, the sequence 'GATC' is numbered 0123. If the sequence starts at 'A', cds_start would be 1. The input must be a nonnegative integer. The default used in our analysis is different for each Tile and shown below:<br>Tile 1: 43<br>Tile 2: 18 |

(*Continued on next page*)

**Table 3. *Continued***

| Configuration file section | Configuration file options | Description |
|---|---|---|
| Tile:Tx | cds_end | The nucleotide position in a merged read is where the coding sequence ends (before primers or any other miscellaneous nucleotides at the end of the read). Note: this index is given using Python slicing conventions, and so is the index of the first nucleotide that is not part of the coding sequence. For example, if the CDS is 'AT' in 'GATC', cds_end would be 3. The input must be a nonnegative integer.<br>The default values used in our analysis is different for each Tile and shown below:<br>Tile 1: 355<br>Tile 2: 321 |
| Tile:Tx | positions | The amino acid positions of interest in the experiment. If the user only desires certain positions in the output files, this is where those positions should be specified.<br>The input must be a list of integers. For example: [22, 42, 44, 51]. If you desire only one residue, enter the integer as a list (e.g., [31]).<br>The default positions used in our analysis is different for each Tile and shown below:<br>Tile 1: 333, 334, 335, 339, 340, 344, 345, 346, 349, 351, 352, 354, 356, 357, 358, 359, 360, 362, 363, 364, 366, 367, 370, 372, 373, 375, 376, 377, 378, 380, 381, 382, 383, 384, 385, 386, 388, 389, 390, 394, 396, 405, 408, 411, 412, 413, 414, 415, 417, 420, 421, 424, 426, 427, 428, 430<br>Tile 2: 437, 439, 440, 441, 444, 445, 446, 447, 448, 449, 450, 452, 455, 456, 457, 458, 459, 460, 462, 463, 464, 465, 466, 468, 469, 470, 471, 473, 474, 475, 477, 478, 479, 481, 482, 483, 484, 485, 486, 487, 489, 490, 492, 493, 494, 498, 499, 500, 501, 503, 504, 505, 506, 508, 516, 517, 518, 519, 520, 521, 522, 523, 527 |
| Analysis | control_ filepath | The file path, including the file name, to the output file from the initial deep mutational scanning module for the control experiment. Absolute file paths should be used. Include extensions, if applicable.<br>An example of this value is: '/Users/lab/Documents/rbd/Output/example_control.csv'. |
| Analysis | antibody_ filepath | The file path, including the file name, to the output from the initial deep mutational scanning module for the desired antibody experiment. Absolute file paths should be used. Include extensions, if applicable.<br>An example of this value is: '/Users/lab/Documents/rbd/Output/example_CC12.1.csv'. |
| Analysis | output_ title | The prefix to be used in the processed file names at the end of running the analysis module.<br>An example of this value is: 'CC12-1' |
| Analysis | FDR | The target false discovery rate (FDR) for identifying an enrichment ratio threshold. This threshold is then used to determine escape mutant hits.<br>The default used in our analysis is 1. |
| Analysis | significance | The p-value cutoff used to determine if a given enrichment ratio is significantly greater than the threshold value.<br>The default used in our analysis is 0.01. |

```
python3 -m analysis -config example.config
```

i. Details on the approach for this analysis can be found in the Quantification and statistical analysis section.

The analysis module will deposit output files in a folder found in the root directory. The files are always deposited into the folder titled "Processed". If that folder does not exist in the directory, the folder will be created.

***Note:*** We provide the CSV outputs of the dms module on Github in the "Output" folder. There are two files: one representing the control population (example_control.csv), cells without ACE2 labeling, and the other representing an antibody population for CC12.1 (Francino-Urdaniz et al., 2021), titled example_CC12.1.csv. In addition, the example configuration file provided is already configured to properly run the analysis module with these example files. The only required change is to update the file paths to the Output folder containing the aforementioned example files. To run the analysis module with this data, follow the instructions in step 34b.

⚠ CRITICAL: The analysis module is written with the intention that it will be executed shortly after running the dms module. Therefore, if output files of the dms module are significantly edited, the analysis module may not run as expected.

**Table 4. Command line flags that can be used when running dms module**

| Command Line flag | Description |
|---|---|
| '–use_multiprocessing' [boolean] | This flag provides an option for the user to use the multiprocessing module or to turn it off. The multiprocessing module reads FASTQ files using multiple cores and can speed up the run time.<br>The default is set to True and the argument parser accepts yes, no, true, or false as options. |
| '–fastq_file_dir' [file path] | This flag allows the user to specify the directory that contains the FASTQ files to be analyzed. The file path should be a folder that contains FASTQ files with the file names provided in the configuration file.<br>The default is an empty string. |
| '–output_dir' [file path] | This flag allows the user to specify the output directory where all CSV and Microsoft Excel files will be saved when the run is complete.<br>The default is an empty string. |
| '–max_mismatches' [number] | The maximum number of mismatches allowed when reading overlapping sequences of paired-end reads from FASTQ deep sequencing files. Reads with a higher number of mismatches are discarded.<br>The default is set to None which sets no limit to the number of mismatches and does not discard reads due to mismatches. Any other input should be a positive integer. |
| '–min_quality' [number] | The minimum deep sequencing quality score required to keep a read in the analysis. A read with a score lower than this minimum benchmark will be discarded.<br>The default is set to None which sets no lower limit on the quality of a read and accepts all scores. Any alternative input must be a positive integer. |
| '–min_ref_counts' [number] | The cutoff used to discard variants that have a low number of counts in the reference population. Variants with reference counts below this cutoff will be discarded.<br>The default is set to 5 and inputs must be a nonnegative integer. |
| '–pseudocount' [number] | The pseudocount used when a variant contains counts observed in the reference population but not in the selected population. This is required because observing no counts in the selected population while having counts in the reference population will lead to an undefined enrichment ratio.<br>The default is set to 1 and inputs must be a positive integer. |

## EXPECTED OUTCOMES

There are several expected outcomes from the experimental procedure. These outcomes are included in the protocol, but they are also written here for completeness. First, in the preparation steps, co-transformation into yeast should result in $>10^4$ transformants for the wild type S RBD (Before you begin: #3e) and $>1.2\times10^5$ transformants for each S RBD library (Steiner et al., 2020) (Before you begin: #4j). Second, when making yeast stocks for the wild type S RBD (Before you begin: #3l), each 1.5 mL *S. cerevisiae* EBY100 culture will produce an average of six 1 mL yeast stocks. Third, the FACS step sorting 200,000 cells from the competing assay (Step-by-step method details: #18) should take around 40 min to collect if using our Sony SH-800 system. This time results from a sorting rate of 5-10,000 cells per sec and maintaining a 70% sorting efficiency. Finally, gel extraction in the sample preparation stage before sequencing (Step-by-step method details: #29) should always give one clear band with the correct size on the SYBR safe gel (515 bp for Tile 1 and 487bp for Tile 2). Any other expected outcomes during the experimental stages are very small and listed with their associated steps.

After the experimental work, the major outcome of the computational step is a statistical measure of how likely each single point mutant present in the library is of being an escape mutant for a specific neutralizing antibody. This information is captured in a comma-separated values (CSV) file generated by the analysis module. The most important information in this CSV file is a false discovery rate (FDR) for each mutation. The FDR for a mutation quantifies how many mutations we would expect to have similar, or more significant, enrichment ratios and depths of coverage by random chance. Other

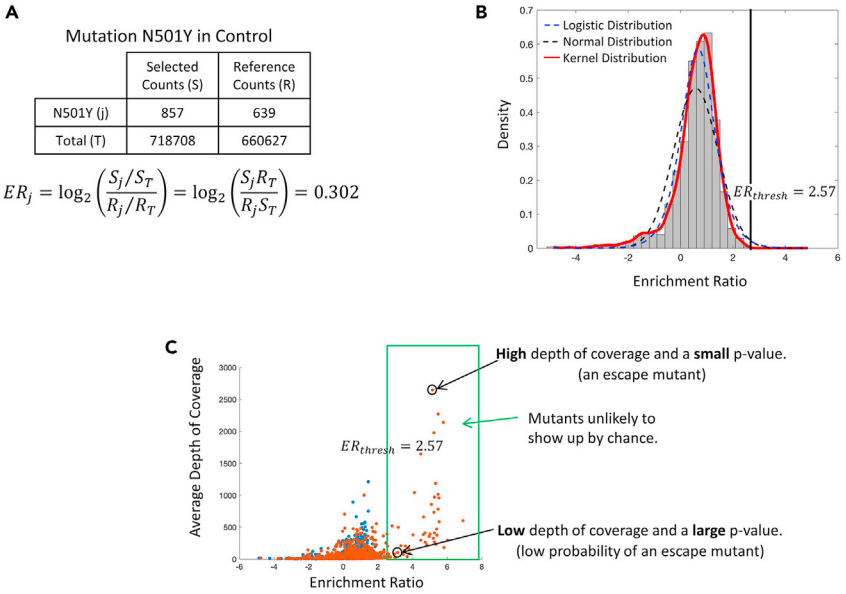



**Figure 6. Statistical approach used to determine potential escape mutants**

(A) An example of an enrichment ratio calculation. This calculation uses the total counts in the selected population, $S_T$, the total counts in the reference population, $R_T$, and the total counts of the given N501Y mutation in the selected and reference populations, $S_j$ and $R_j$, respectively.

(B) Demonstrating that the best fit for this data is a kernel density estimate (KDE). The KDE fits very closely to the data, while other common probability distributions, such as the normal and logistic curves, do not.

(C) Showing the final statistical step that considers the depth of coverage when deciding to identify an escape mutation. For mutations with high depth of coverage, the returned p-value will be much smaller than for mutations with lower depths of coverage.

information contained in the CSV file for the set of mutants includes their corresponding counts in each population, an enrichment ratio (ER) that compares these counts, a p-value calculated by the one-sided exact Poisson rate ratio test, and the minimum number of nucleotides away the given mutation could be from the wild type sequence. The second output of this software is a heatmap in Microsoft Excel that identifies potential escape mutations in blue. Refer to Figure 4 for details on the outputs mentioned above.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis is performed by the custom software developed for this protocol. This occurs entirely in Step 3 from the procedures above. To determine escape mutations, the antibody population must be compared against the control group. A control was used to help identify which mutations are randomly expected to appear in the given library. With this data, we can determine the mutations present in the selected antibody population that occur more frequently than would typically be expected when looking at the control data. This information is captured as the FDR value. If these specific mutations have enough counts to be confident of their presence in the antibody library, then we call those variants escape mutations. A more detailed and formal explanation of the statistical approach follows.

First, the critical metric for determining escape mutants in this analysis is the enrichment ratio at each position. An enrichment ratio is the $\log_2$ transform of the frequency of an observed mutation in the selected population divided by the frequency of that mutation in the reference population (Figure 6A). This calculation is performed by the dms module prior to running any statistical analysis.

When running the analysis module, a kernel density estimate (KDE) is first fit to the control data. This provides an empirical ER distribution for cells sorted without antibody or ACE2 labeling, which

serves as a null hypothesis in which observed ERs are due to stochasticity in sorting and other non-relevant differences between variants. In determining which distribution to use for this data, we initially fit common probability distributions (Figure 6B). However, KDE consistently provided a better fit to the data. In addition, a KDE eliminates the need to make the underlying assumptions common with parametric functions, yet still provides an accurate distribution (Kvam and Vidakovic, 2007). Thus, given the unpredictable nature of the control data, a nonparametric probability distribution such as the KDE was chosen to ensure a reliable distribution for each given control group without making any assumptions about the distribution of data.

Using this KDE from the null distribution, a threshold value for enrichment ratios can be determined. We estimate the false discovery rate (FDR) for a given threshold $ER_t$ as the number ERs we expect to find above that threshold in the null experiment. This is given by $\mathrm{FDR} = (1 - \mathrm{F}(\mathrm{ER}_t)) * \mathrm{N}$, where F is the experimentally determined cumulative distribution function (CDF) of the control experiment ERs and is the number of variants in the library. Rearranging, we find $\mathrm{ER}_t = \mathrm{F}^{-1}\left(1 - \frac{\mathrm{FDR}}{\mathrm{N}}\right)$, where $\mathrm{F}^{-1}$ is the inverse CDF. For a library size using these experiments (N = 1120 for tile 1 or N = 1260 for tile 2) and with an FDR = 1, the typical range for this threshold is $2.2 < \mathrm{ER}_t < 2.6$.

We then compare each observed ER in an experiment to this threshold to see which variants have enrichment ratios larger than $\mathrm{ER}_t$. However, comparing just the observed ER to the threshold would ignore the depth of coverage for each mutation. Since the depth of coverage provides information about our confidence in the observed ER, we use it directly in the comparison. This is explained in detail below.

For a given experiment, we have two sets of reads $\{k_i^p\}$ and two totals $n^p$, where $p \in \{\mathrm{ref},\ \mathrm{sel}\}$ is the reference or selected population. Note that, since we discard reads that don't encode single mutations, it is not generally true that $\sum_i k_i^p = n^p$. We approximate each read count $k_i^p$ as being distributed as $k_i^p \sim Pois(r_i^p n_i^p)$ for some unknown rate $r_i^p$. If we knew this rate, the true enrichment ratio would be given by $\mathrm{ER}_i = \log_2\left(\frac{r_i^{sel}}{r_i^{ref}}\right)$. Therefore, testing the hypothesis that $\mathrm{ER}_i \geq \mathrm{ER}_t$ amounts to testing that $\frac{r_i^{sel}}{r_i^{ref}} \geq 2^{\mathrm{ER}_t}$, which we do with $k_i^{ref}, n^{ref}, k_i^{sel}$, and $n^{sel}$ using a one-sided Poisson rate ratio test implemented in the Python package statsmodels (Seabold and Perktold, 2010). The significance level of the test, $\alpha$, is determined by the user and can be changed in the configuration file. Our analysis used $\alpha = 0.01$. Please refer to Figure 6C for additional details.

## LIMITATIONS

The method described has several experimental parts that introduce limitations to the protocol. First, this protocol identifies only the subset of neutralizing antibodies that directly compete with ACE2 for binding on S RBD. Neutralizing antibodies for SARS-CoV-2 can bind on Spike at locations other than the RBD (Cerutti et al., 2021; Li et al., 2021; Mccallum et al., 2021), and some neutralizing antibodies recognize the N343 glycan that we have removed from the displayed S RBD to avoid hyper-mannosylation (Jigami, 2014). A different approach taken by the Bloom research group can be considered if non-competing antibodies must be analyzed (Greaney et al., 2020; Starr et al., 2021) that also bind on the S RBD. We have previously tried to display the S ectodomain in yeast but were not successful (Francino-Urdaniz et al., 2021).

Second, this procedure is only designed to identify escape mutants resulting from single amino acid mutations. The smaller single mutation library limits our ability to test the possibility that a combination of amino acid mutations could work in conjunction and ultimately lead to escaping antibody binding.

Larger and more complex libraries would be needed to test this type of escape mutant. In addition, the custom software developed with this protocol is not written to identify such scenarios.

Third, the residues examined in this study are all located on the surface of S RBD. While ACE2 and other antibodies will only interact directly with these surface residues, it is possible that buried residues could have indirect impacts on binding. Furthermore, since we test monomeric S RBD some surface mutations identified as putative escapes by this method could produce steric clashes in the full trimeric S. Each mutation is pleiotropic with unpredictable potential changes in the fitness of the virus. Thus, the escape mutants identified by this protocol will necessarily be a subset of the escape mutants for these same positions observed in nature.

Finally, given that a mutation was observed in the population, the output is binary (yes/no for an escape mutant). Some true escape mutants may be missed if the mutation does not appear with high enough frequency in the antibody population. This is especially true if ER values are close to the cutoff, if the mutation is not sampled in the reference library, or if the control population has some aberrations that skew the ER threshold higher.

## TROUBLESHOOTING

### Problem 1

S RBD does not display on the yeast surface or has a very low display. This problem may rise in Before you begin step 16 or Step-by-step method details step 12.

### Potential solution

Low display percentages occur with insufficient culture aeration during growth. Ensure the 14 mL culture tube is shaking and that the culture volume is 1.5 mL or less. Induction time and temperature were optimized for our laboratory settings and may need to be altered to optimize display. Other common temperatures are 18°C and 30°C.

If necessary, a 0.2 w/v% galactose can be added to the $SDCAA^+$ and 0.2 w/v% dextrose can be added to the $SGCAA^+$. If the RBD is still not displaying, co-transform again.

### Problem 2

No or low PE signal increase when binding biotinylated ACE2 to the displayed RBD. This problem may arise in Before you begin step 22 or Step-by-step method details step 13 Tube 2.

### Potential solution

Increase the biotin to protein ratio. If the signal increases as the ratio increases, use the necessary ratio to see a good PE signal increase.

### Problem 3

Product from the second PCR when preparing the libraries for deep sequencing does not show on a SYBR safe gel in Step-by-step method details step 29.

### Potential solution

First check that there is some product on the right band size using the more sensitive SYBR gold stain. If so, rerun the PCR steps using more cycles. Else, cells could have been contaminated during cell recovery or the yeast miniprep did not have a high enough yield.

### Problem 4

Illumina deep sequencing does not cluster correctly. This would occur after preparing for sequencing in Step-by-step method details Illumina sequencing.

### Potential solution

Increase PhiX percentage to increase nucleotide diversity to promote correct clustering.

### Problem 5

The example configuration file is not set up by default for running with only one tile. Thus, using the software for an experiment that only has a single tile may result in an error if the configuration file is not formatted correctly through the initial setup in Step-by-step method details step 33.

### Potential solution

To run a single tile, continue to add the experiment in the [Proteins] section. However, after identifying the first tile, add a comma and then a space (leaving the rest of the line blank). For example, if the experiment name is '6-29' and the chosen title for the protein is '1_6–29' then use the following format in the [Proteins] section.

Control_1: '1_Control',

This should fix any issues.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tim Whitehead timothy.whitehead@colorado.edu.

### Materials availability

All plasmids and mutational libraries used in this work are available from AddGene (https://www.addgene.org/Timothy_Whitehead/).

### Data and code availability

Original software has been deposited on Github (https://github.com/WhiteheadGroup/SpikeRBDStabilization.git).

## ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI141452 to T.A.W. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was also supported by a Balsells Fellowship to I.M.F.-U. and a CU Boulder BSI fellowship to C.M.H.

## AUTHOR CONTRIBUTIONS

C.M.H. and P.J.S. developed computational software; I.M.F.-U., P.J.S., and T.A.W. developed experimental protocol; C.M.H., I.M.F.-U., P.J.S., and T.A.W. wrote the protocol.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

Banach, B.B., Cerutti, G., Fahad, A.S., Shen, C.-H., Olivera de Souza, M., Katsamba, P.S., Tsybovsky, Y., Wang, P., Nair, M.S., Huang, Y., et al. (2021). Paired Heavy and Light Chain Signatures Contribute to Potent SARS-CoV-2 Neutralization in Public Antibody Responses. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3754549.

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E.R., Yu, J., Bahna, F., Bimela, J., et al. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host and Microbe *29*, 819–833.e7. https://doi.org/10.1016/j.chom.2021.03.005.

Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat. Protoc. *1*, 755–768. https://doi.org/10.1038/nprot.2006.94.

Fowler, D.M., Araya, C.L., Gerard, W., and Fields, S. (2011). Enrich: software for analysis of protein

function by enrichment and depletion of variants. Bioinformatics *27*. https://doi.org/10.1093/bioinformatics/btr577.

Francino-Urdaniz, I.M., Steiner, P.J., Kirby, M.B., Zhao, F., Haas, C.M., Barman, S., Rhodes, E.R., Leonard, A.C., Peng, L., Sprenger, K.G., et al. (2021). One-shot identification of SARS-CoV-2 S RBD escape mutants using yeast screening. Cell Rep. *36*. https://doi.org/10.1016/j.celrep.2021.109627.

Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., et al. (2020). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host and Microbe *29*, 44–57.e9. https://doi.org/10.1016/j.chom.2020.11.007.

Jigami, Y. (2014). Yeast glycobiology and its application. Biosci. Biotechnol. Biochem. 8451. https://doi.org/10.1271/bbb.70725.

Kowalsky, C.A., Klesmith, J.R., Stapleton, J.A., Kelly, V., Reichkitzer, N., and Whitehead, T.A. (2015). High-resolution sequence-function mapping of full-length proteins. PLoS One *10*, 1–23. https://doi.org/10.1371/journal.pone.0118193.

Klesmith, J.R., and Hackel, B.J. (2019). Improved mutant function prediction via PACT: protein analysis and classifier toolkit. Bioinformatics *35*. https://doi.org/10.1093/bioinformatics/bty1042.

Kvam, P.H., and Vidakovic, B. (2007). Nonparametric Statistics with Applications to Science and Engineering (John Wiley & Sons). https://doi.org/10.1002/9780470168707.

Li, Y., Ma, M., Lei, Q., Li, Y., Ma, M., Lei, Q., Wang, F., Hong, W., Lai, D., Hou, H., and Xu, Z. (2021). Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients. CellReports *34* (13), 108915. https://doi.org/10.1016/j.celrep.2021.108915.

Mccallum, M., MarcoDe, A., Lempp, F.A., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell *184*, 2332–2347.e16. https://doi.org/10.1016/j.cell.2021.03.028.

Medina-Cucurella, A.V., and Whitehead, T.A. (2018). Characterizing protein-protein interactions using deep sequencing coupled to yeast surface display. Methods Mol. Biol. https://doi.org/10.1007/978-1-4939-7759-8_7.

Seabold, S., and Perktold, J. (2010). Statsmodels : Econometric and statistical modeling with python. Proc. 9th Python Sci. Conf. *57* (Scipy), 92–96.

Starr, T.N., Greaney, A.J., Dingens, A.S., and Bloom, J.D. (2021). Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Rep. Med. *2*, 100255. https://doi.org/10.1016/j.xcrm.2021.100255.

Steiner, P.J., Baumer, Z.T., and Whitehead, T.A. (2020). A method for user-defined mutagenesis by integrating oligo pool synthesis technology with nicking mutagenesis. Bio-protocol *10*, e3697.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell *180*, 1–12. https://doi.org/10.1016/j.cell.2020.02.058.