Cell, Volume *185*

Supplemental information

Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain

Joseph M. Taft, Cédric R. Weber, Beichen Gao, Roy A. Ehling, Jiami Han, Lester Frei, Sean W. Metcalfe, Max D. Overath, Alexander Yermanos, William Kelton, and Sai T. Reddy

SUPPLEMENTARY TABLES

Population	Antigen	Events Screened	Binding Events Collected	Non-Binding Events Collected
Library 2C	50 nM ACE2	5.00E+07	1.60E+06	6.40E+06
Library 2CE	50 nM ACE2	5.00E+07	5.90E+05	6.40E+06
Library 2C + 2CE, ACE2-binding	100 nM REGN10933	1.07E+07	1.21E+06	3.93E+05
Library 2C + 2CE, ACE2-binding	100n nM RENG10987	1.02E+07	9.54E+05	5.41E+05
Library 2C + 2CE, ACE2-binding	100 nM Ly-CoV16	8.12E+06	5.07E+05	6.92E+05
Library 2C + 2CE, ACE2-binding	100 nM Ly-CoV555	1.13E+07	2.36E+04	1.87E+06
Library 2T	50 nM ACE2	1.52E+07	8.53E+05	1.49E+06
Library 2T, ACE2-binding	100 nM REGN10933	1.94E+06	5.09E+05	4.11E+04
Library 2T, ACE2-binding	100n nM RENG10987	2.43E+06	5.01E+05	4.57E+04
Library 2T, ACE2-binding	100 nM Ly-CoV16	2.56E+06	5.66E+05	1.13E+05
Library 2T, ACE2-binding	100 nM Ly-CoV555	3.15E+06	7.61E+05	1.48E+05

Table S1. Sorting Statistics for RBD Library Sorting. Events screened correspond to all detected events above the forward scatter threshold, including doublet cells. Binding and non-binding events include cells which are not. See Figs 2C,E and S1B

		Paired and Filtered Reads	
Population	Antigen	Binding	Non-Binding
Library 2C	50 nM ACE2	1.20E+06	1.16E+06
Library 2CE	50 nM ACE2	1.01E+06	1.48E+06
Library 2C + 2CE, ACE2 binding	100 nM REGN10933	9.54E+05	7.55E+05
Library 2C + 2CE, ACE2 binding	100n nM RENG10987	2.27E+05	1.15E+04
Library 2C + 2CE, ACE2 binding	100 nM Ly-CoV16	1.85E+06	2.98E+05
Library 2C + 2CE, ACE2 binding	100 nM Ly-CoV555	1.16E+05	5.47E+05
Library 123T	50 nM ACE2	1.53E+06	1.88E+06
Library 123T, ACE2 binding	100 nM REGN10933	5.53E+06	2.55E+06
Library 123T, ACE2 binding	100n nM RENG10987	3.30E+06	2.87E+06
Library 123T, ACE2 binding	100 nM Ly-CoV16	1.59E+06	3.20E+06
Library 123T, ACE2 binding	100 nM Ly-CoV555	3.64E+06	2.56E+06

Table S2. Sequencing Statistics for RBD Library Sorting. Reads were paired, merged, and trimmed using Geneious Prime and the bbduk plugin. See Fig 2D.

Antibody	Class	Description	Reference
LY-CoV16 (Etesevimab)	1	Neutralizing antibody used as part of the Eli-Lilly therapeutic antibody cocktail. Derived from convalescent COVID-19 patient samples. LY-CoV16 is no longer authorized for clinical use due to loss of neutralization to Omicron BA.1 or sublineages.	Shi et al. Nature. 2020
LY-CoV555 (Bamlanivimab)	2	Neutralizing antibody used as part of the Eli-Lilly therapeutic antibody cocktail. Identified through high-throughput screening of B cells from convalescent COVID-19 patient samples. LY-CoV555 is no longer authorized for clinical use due to loss of neutralization to Omicron BA.1 or sublineages.	Jones et al. Science Transl. Med. 2021
REGN10933 (Casirivimab)	1	Neutralizing antibody used as part of the Regeneron COVID-19 therapeutic antibody cocktail. Derived from genetically humanized mice or convalescent human COVID-19 patient samples. REGN10933 is no longer authorized for clinical use due to loss of neutralization to Omicron BA.1 or sublineages.	Hansen et al. Science, 2020
REGN10987 (Imdevimab)	3	Neutralizing antibody used as part of the Regeneron COVID-19 therapeutic antibody cocktail. Derived from genetically humanized mice or convalescent human COVID-19 patient samples. REGN10987 is no longer authorized for clinical use due to loss of neutralization to Omicron BA.1 or sublineages.	Hansen et al. Science, 2020
mAb-50	3	Neutralizing antibody isolated from expanded plasma cells of convalescent COVID-19 patients.	Ehling et al. Cell Rep, 2021.
mAb-64	1	Neutralizing antibody isolated from expanded plasma cells of convalescent COVID-19 patients.	Ehling et al. Cell Rep, 2021.
mAb-82	1	Neutralizing antibody isolated from expanded plasma cells of convalescent COVID-19 patients.	Ehling et al. Cell Rep, 2021.
S309 (Sotrovimab)	3	Neutralizing antibody used for treatment in the Vir/GSK therapeutic treatment. Originally isolated from B cells of a convalescent patient recovered from SARS-CoV-1 and with cross-reactivity to SARS-CoV-2. S309 is no longer authorized for clinical use due to loss of neutralization to the Omicron BA.2 variant.	Pinto et al. Nature, 2020
S2E12	1	Neutralizing antibody with broad neutralization against a diverse clade of sabrecoviruses.	Starr et al. Nature, 2021
S2H97	4	Neutralizing antibody with broad neutralization against a diverse clade of sabrecoviruses.	Starr et al. Nature, 2021
LY-CoV1404 (Bebtelovimab)	3	Neutralizing antibody used in the Eli-Lilly therapeutic antibody treatment. Identified through high-throughput screening of B cells from convalescent COVID-19 patient samples. LY-CoV1404 remains authorized for clinical use due to effective neutralization to Omicron BA.1 and BA.2, BA.4 and BA.5 sublineages.	Westendorf et al. Cell Rep, 2022
A23-58.1	1	Neutralizing antibody isolated from B cells of convalescent COVID-19 patients and shown to have broad neutralization.	Wang et al. Science, 2021
G32A4	1	Neutralizing antibody isolated from B cells of convalescent COVID-19 patients and shown to have broad neutralization.	Tong et al. Cell, 2021

Table S3. List of antibodies screened by DML in this study. See Fig. 2.

Random Forest	**RNN**
n_estimators = 500,	
min_samples_split = 2,	3 sequential LSTM/Dropout layers
min_samples_leaf = 1,	Dense layer 50 units
max_depth = 150,	Sigmoid activation output
max_features = ‘sqrt’,	Cross-entropy loss
criterion = gini	
HP tuning parameters:	HP tuning parameters:
n_estimators: [50-500]	batch size: [16, 32]
min_samples_split: [2,5,10]	epochs: [10, 20]
min_samples_leaf: [1,2,5]	dropout rate: [0.1,0.2],
max_features: [‘auto’, ‘sqrt’]	LSTM units: [40, 80],
max_depth: [20-150,None]	optimizer: [adam, rmsprop]

Table S5. Key Metrics for Random Forest and RNN Models. Hyperparameters were tuned through RandomSearchCV, scored for “precision”, using Sci-kit Learn for 50 rounds. For baseline models trained on ACE2 sequences, metrics evaluated include: total accuracy and F1, precision, and recall for both classes. Models were evaluated on the entirety of the held-out test set, without hyperparameter optimization. For final models, metrics evaluated include: total accuracy and F1, precision, and recall for the positive class. Models were evaluated on the entirety of the held-out test set, or “Low Distance” and “High Distance” subsets of the held-out test set. See Fig. 3.