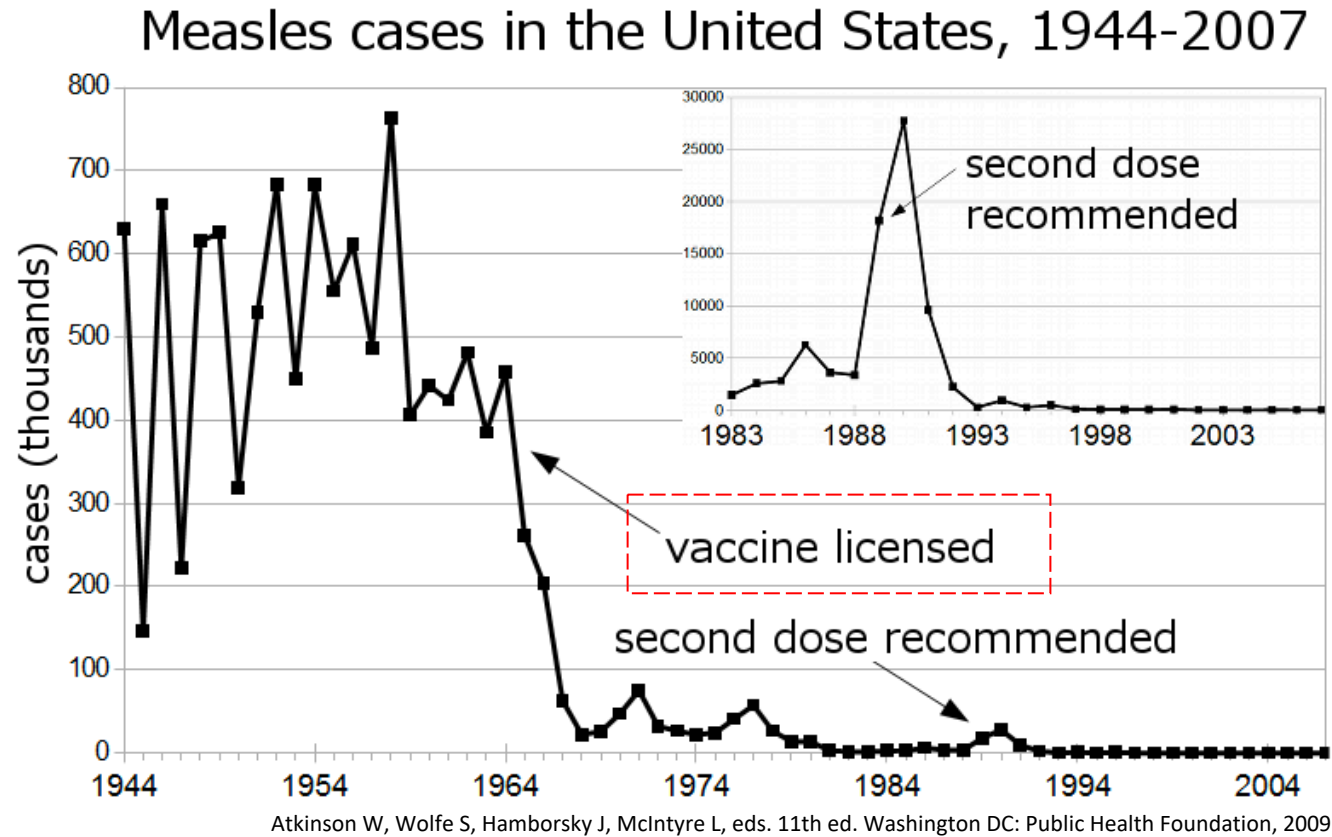


Identify SARS-CoV-2 vaccine candidates with ML

Binbin Chen, Ethan Fast, Russ Altman
Stanford University Department of Genetics

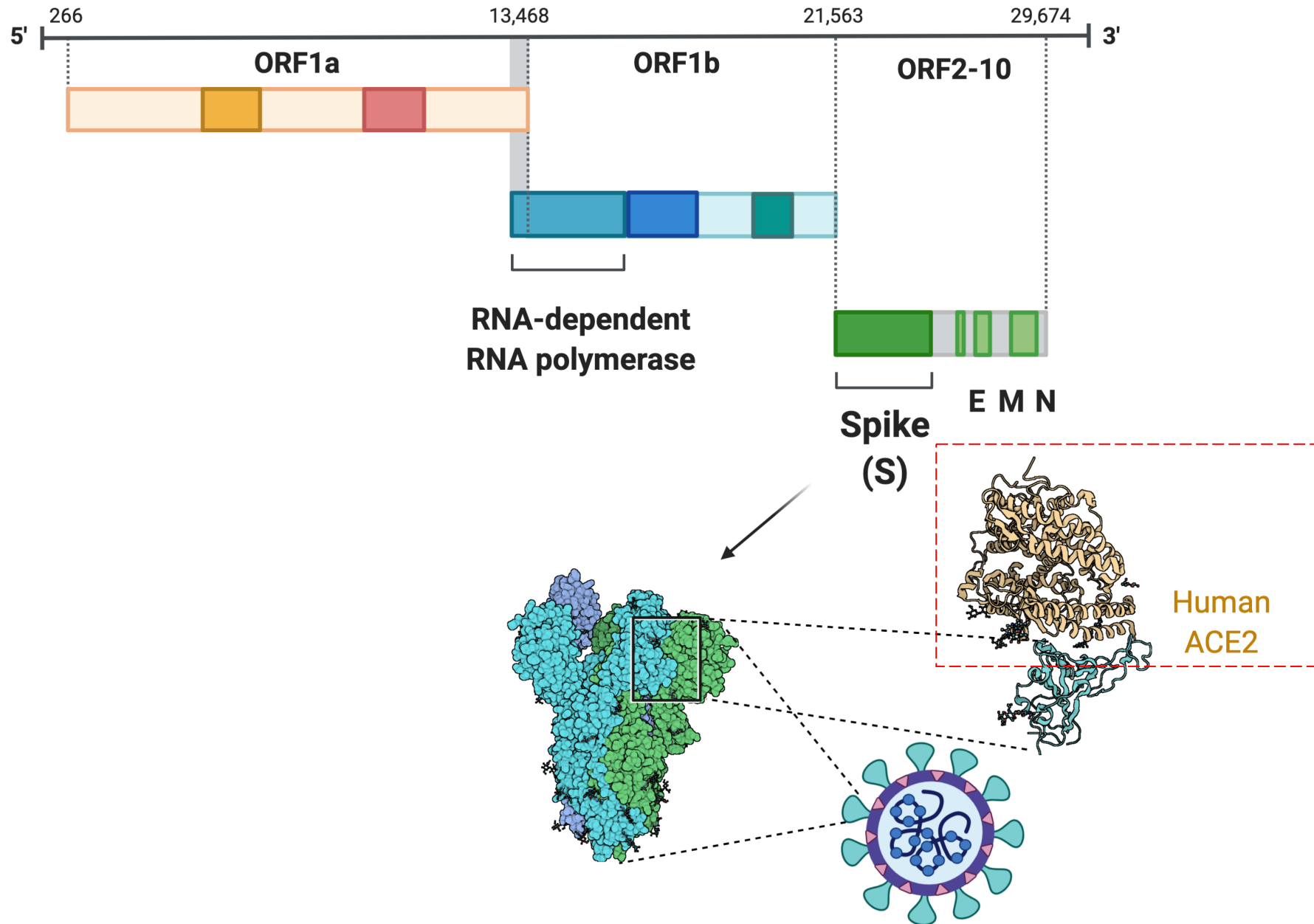


Vaccines: the most effective weapon against the pandemic

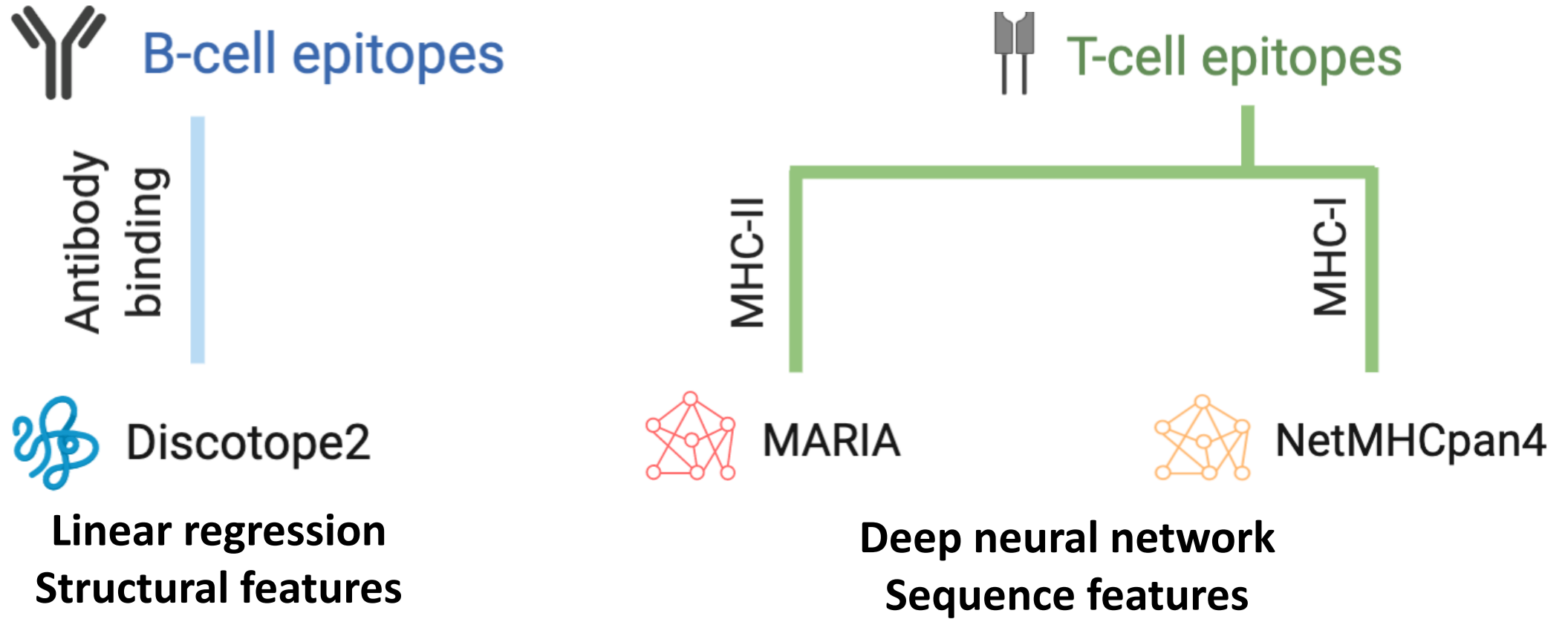


No vaccines available for any form of coronaviruses.

SARS-CoV-2 genome codes for key pathogenic proteins



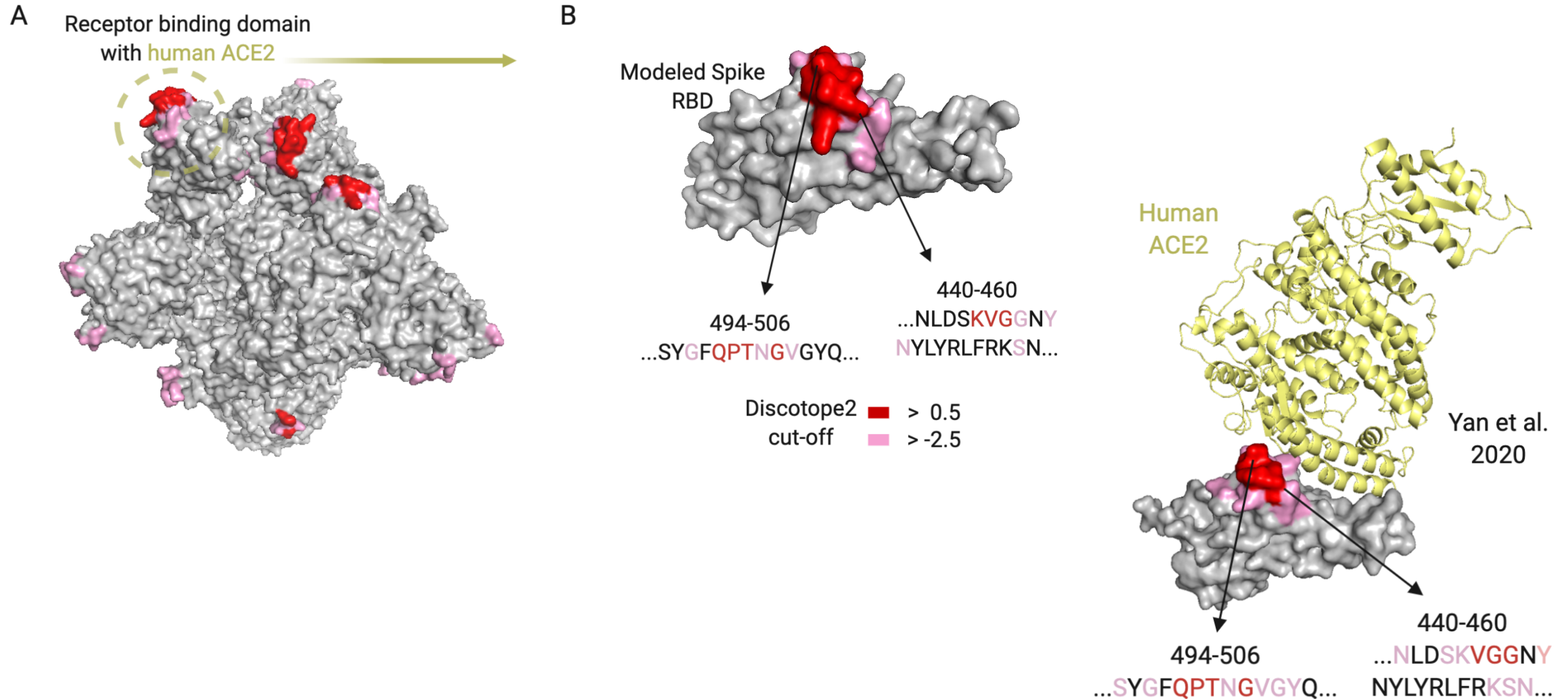
T-cell and B-cell responses are both essential for viral clearance



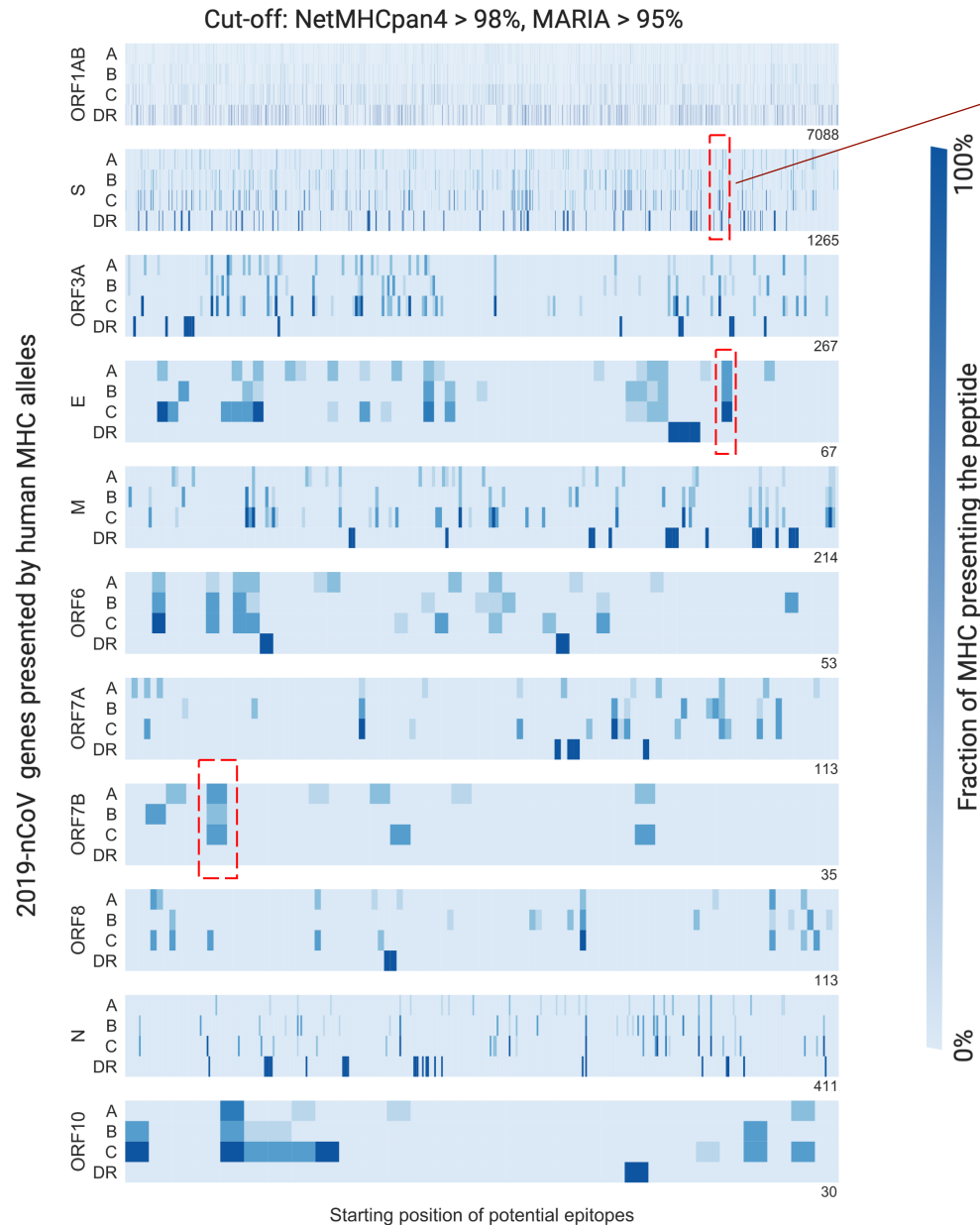
COVID-19 vaccines should contain both B-cell and T-cell epitopes

Discotope2 predicts antibody binding sites for S protein

Predicted B-cell epitopes on S protein structure up conformation



Some regions are highly presentable by human MHC



...VVFLHVTYV...
Presentable by 65% common MHCs

Top candidates for epitope based vaccines

Gene	Sequence	Position	MHC-I Cov.		MHC-II Cov.		Antibody
S	SYGFQPTNGVGYQPY	494	Yes	52%	Yes	100%	Predicted
	SQSIIAYTMSLGAEN	689	Yes	74%	Yes	100%	No
	IPTNFTISVTTEILP	714	Yes	70%	Yes	100%	No
	AAAYYVGYLQPRTFL	262	Yes	65%	Yes	100%	No
	APHGVVFLHVTYVPA	1056	Yes	65%	Yes	100%	SARS
ORF1ab	DGEVITFDNLKTLLS	1547	Yes	83%	Yes	100%	No
	EVRTIKVFTTVDNIN	1564	Yes	78%	Yes	100%	No
	IINLVQMAPISAMVR	2368	Yes	78%	Yes	100%	No
	NPTTFHLDGEVITFD	1540	Yes	74%	Yes	100%	No
	VAAIFYLITPVHVMS	2783	Yes	74%	Yes	100%	No
M	IASFRLFARTRSMWS	97	Yes	65%	Yes	100%	SARS
N	ATKAYNVTQAFGRRG	264	Yes	74%	Yes	100%	SARS
E	VKPSFYVYSRVKNLN	52	Yes	74%	Yes	100%	SARS

Conclusion

- Spike protein contains both T-cell and B-cell epitopes. Patients can likely mount antibodies blocking the viral entry.
- We have a comprehensive list of candidates for either vaccine development or understanding T-cell responses.

Thank you!



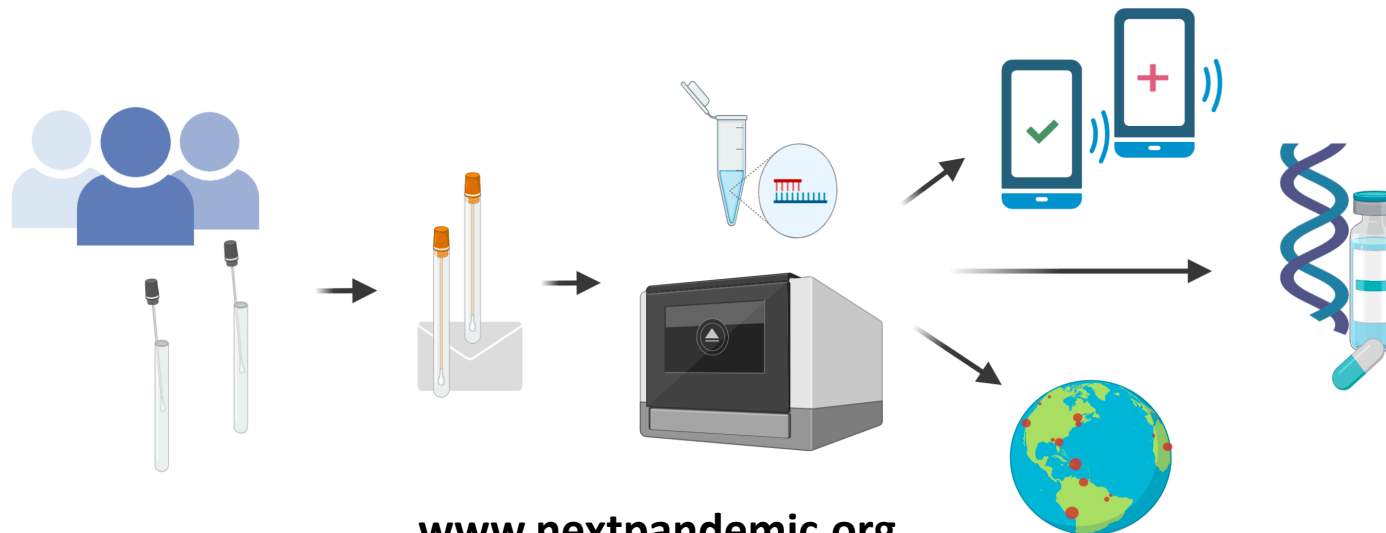
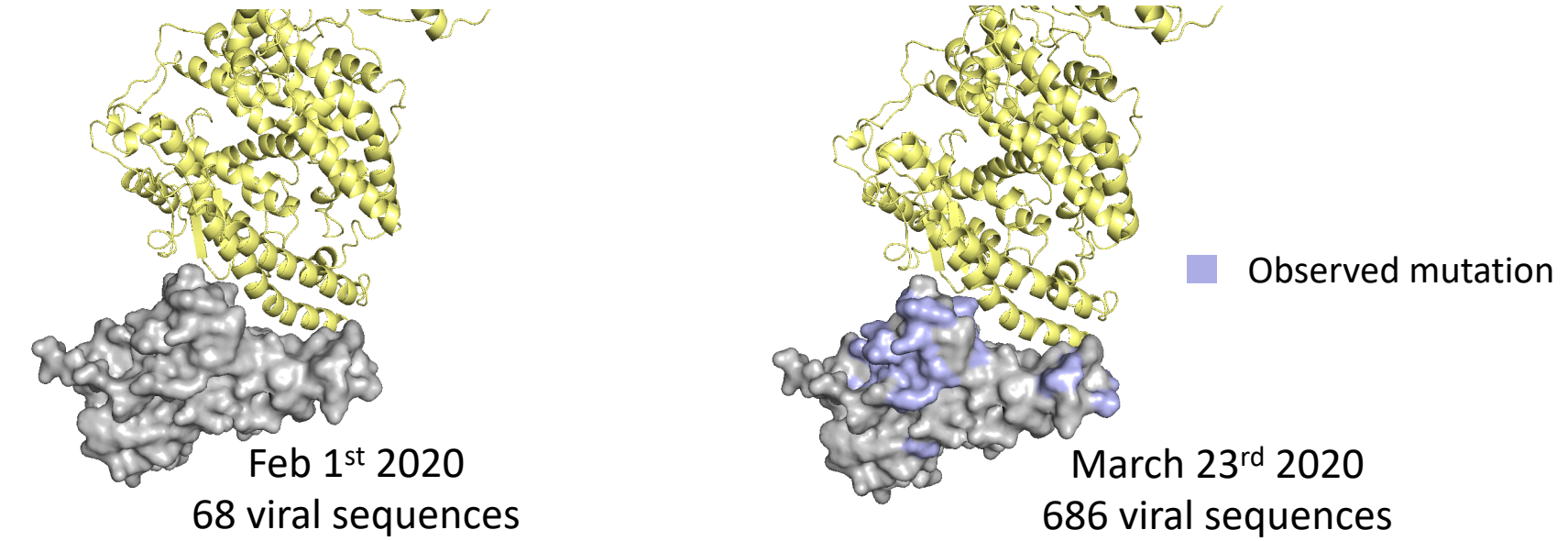
PAUL & DAISY
SOROS
FELLOWSHIPS FOR
NEW AMERICANS



www.nextpandemic.org

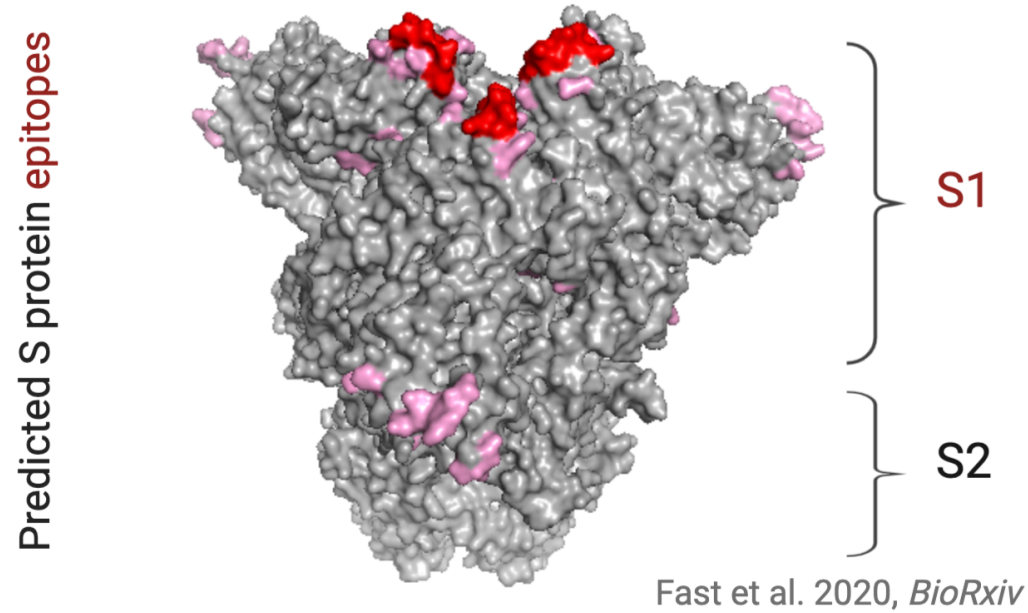
Additional slides

Tracking viral mutations can be informative for vaccine design



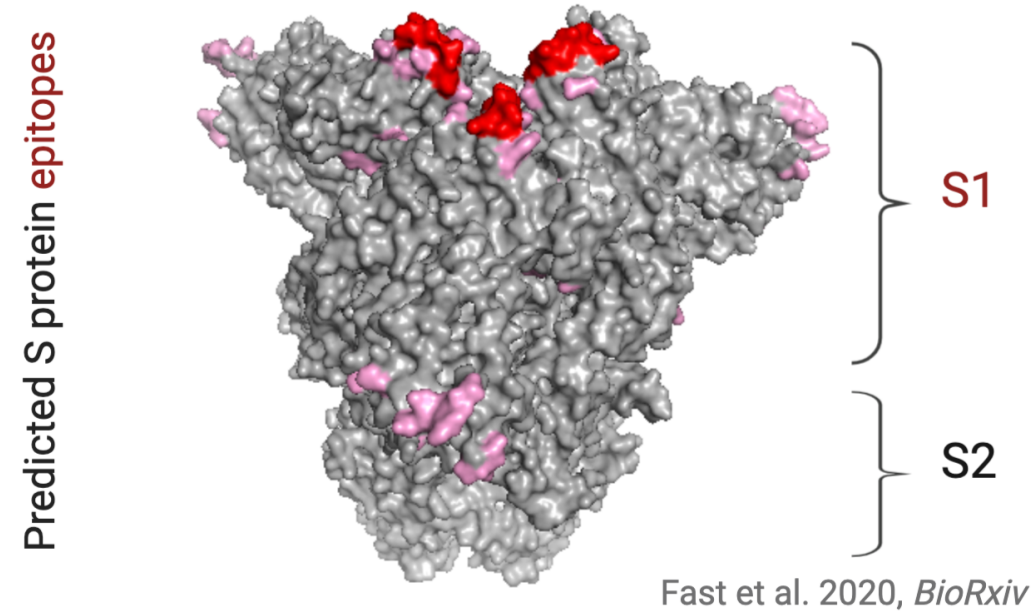
Antibody prediction validated by the patient blood sample

Predicted vs. detected SARS-CoV-2 antibody epitopes

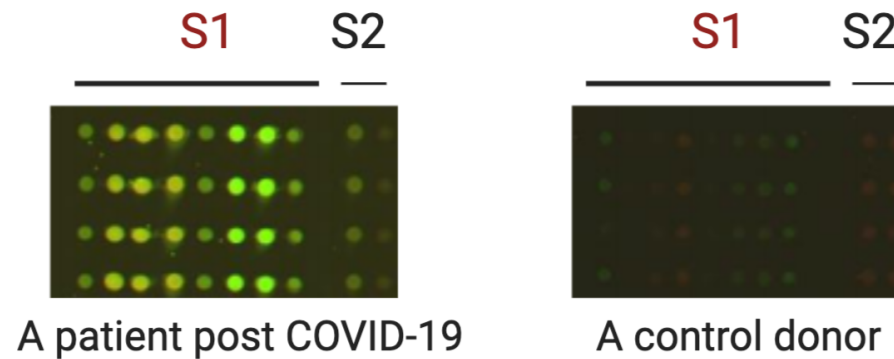


Antibody prediction validated by the patient blood sample

Predicted vs. detected SARS-CoV-2 antibody epitopes



Detected epitopes
in COVID-19 patient

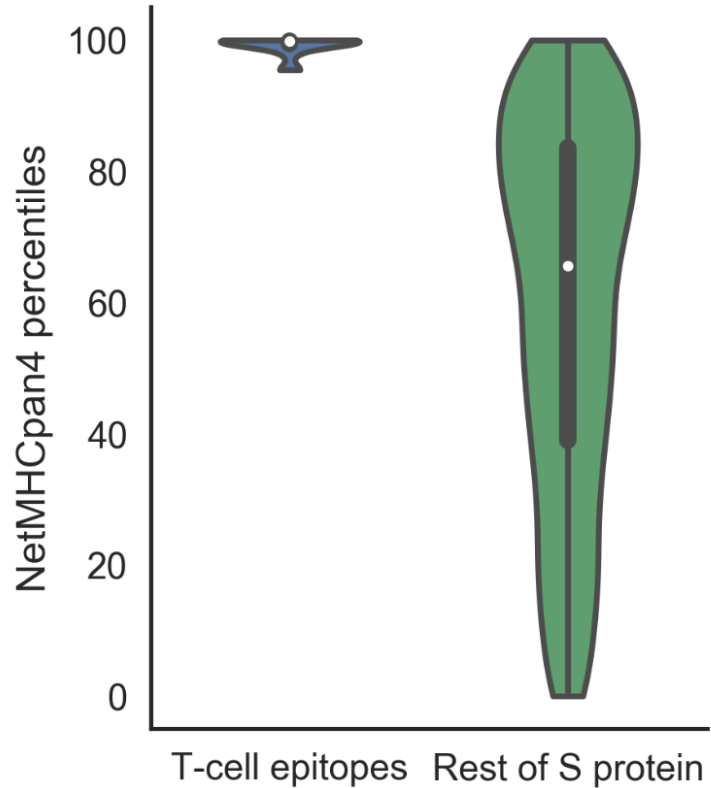


Jiang et al. 2020, *MedRxiv*

Our computational pipeline predicts historical SARS-CoV T-cell epitopes

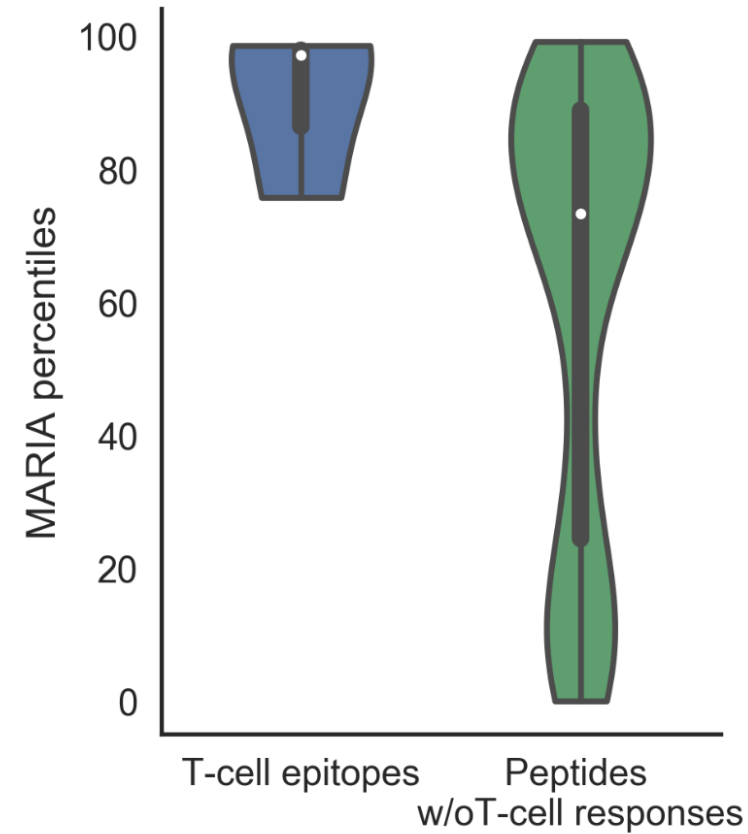
A

Predicted scores for SARS CD8 T epitopes
(n=1253, AUC=0.98)

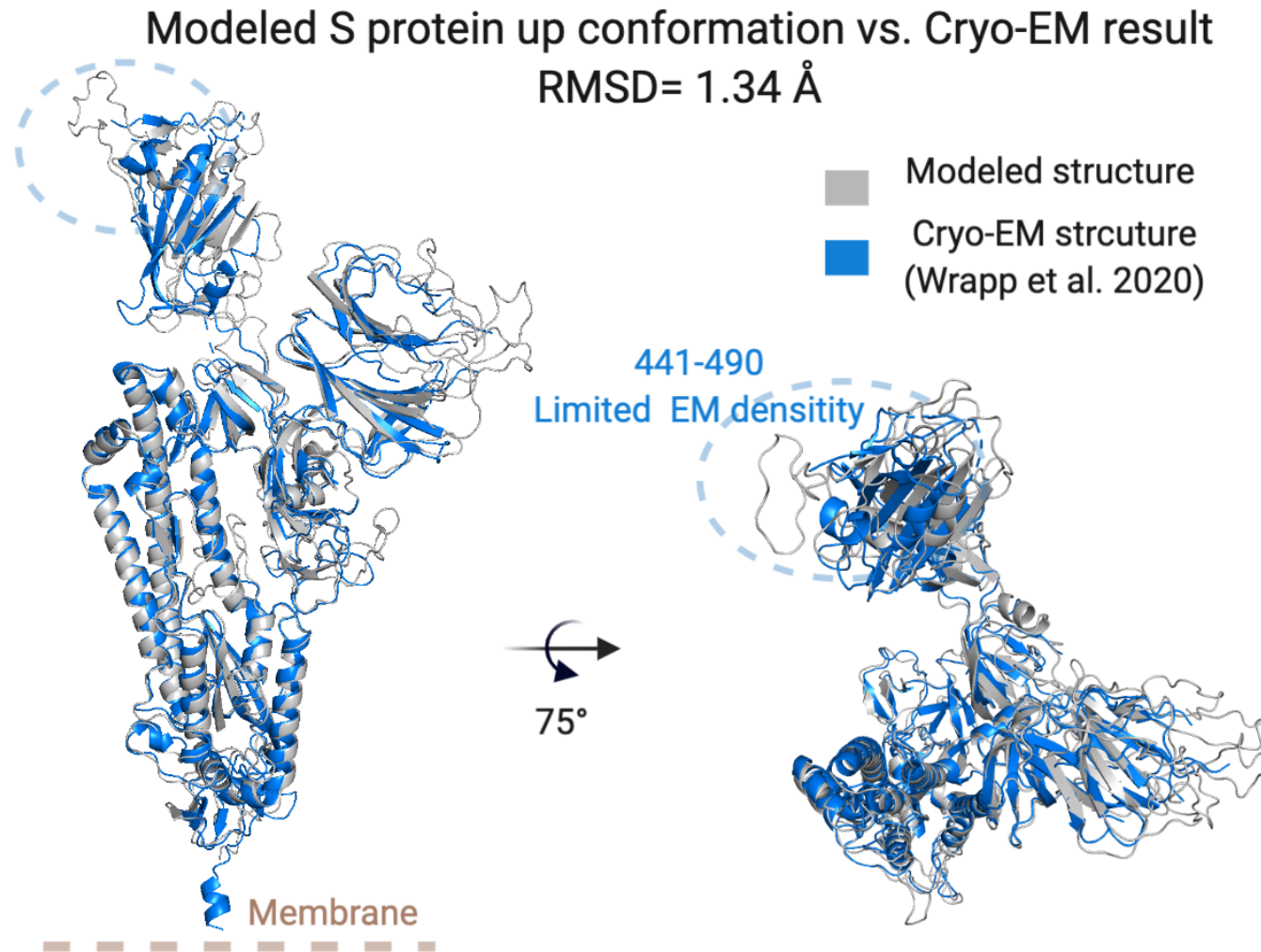


B

Predicted scores for SARS CD4 T epitopes
(n=249, AUC=0.83)



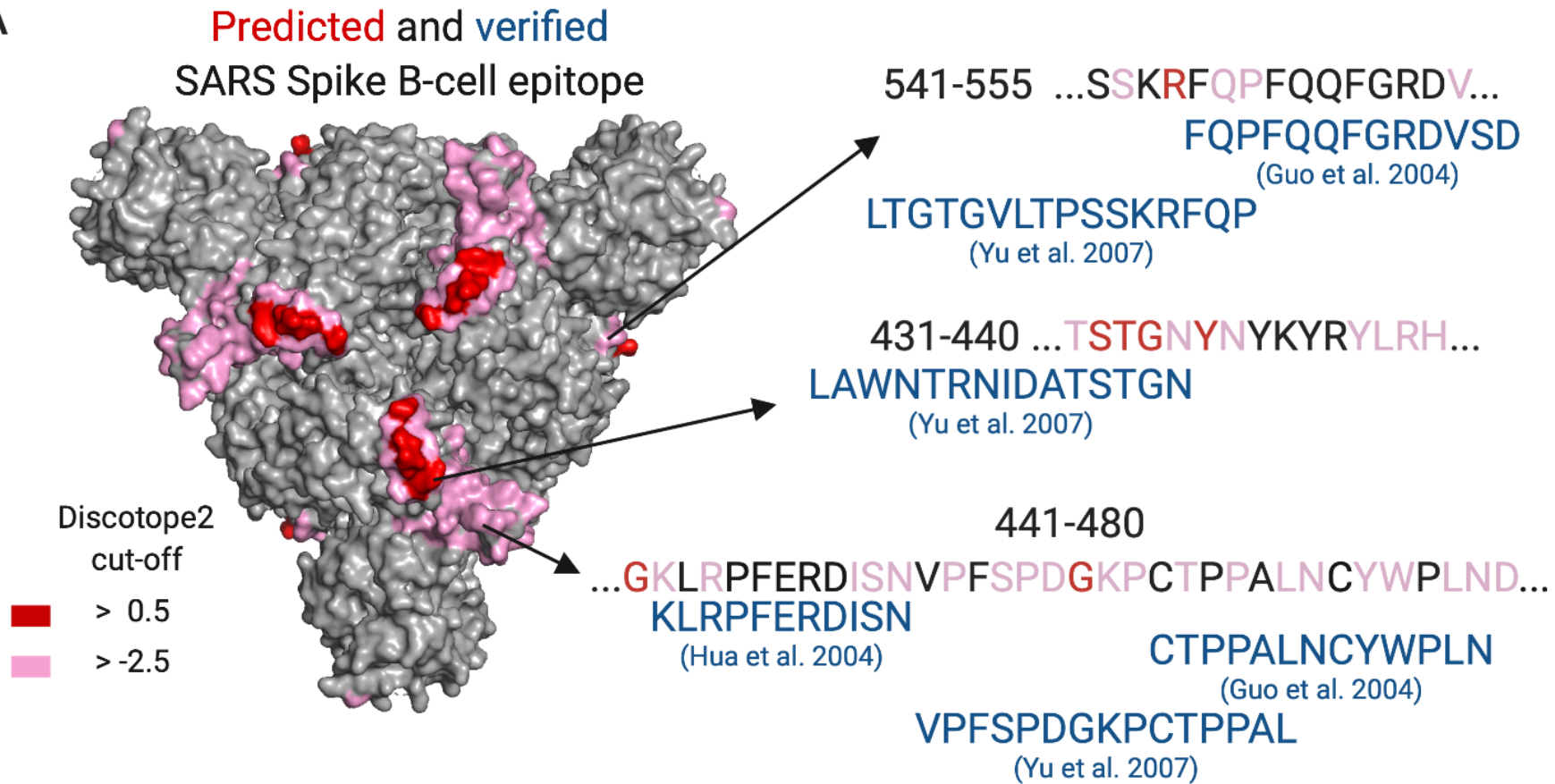
Structural homology modeling shows consistency with Cryo-EM results



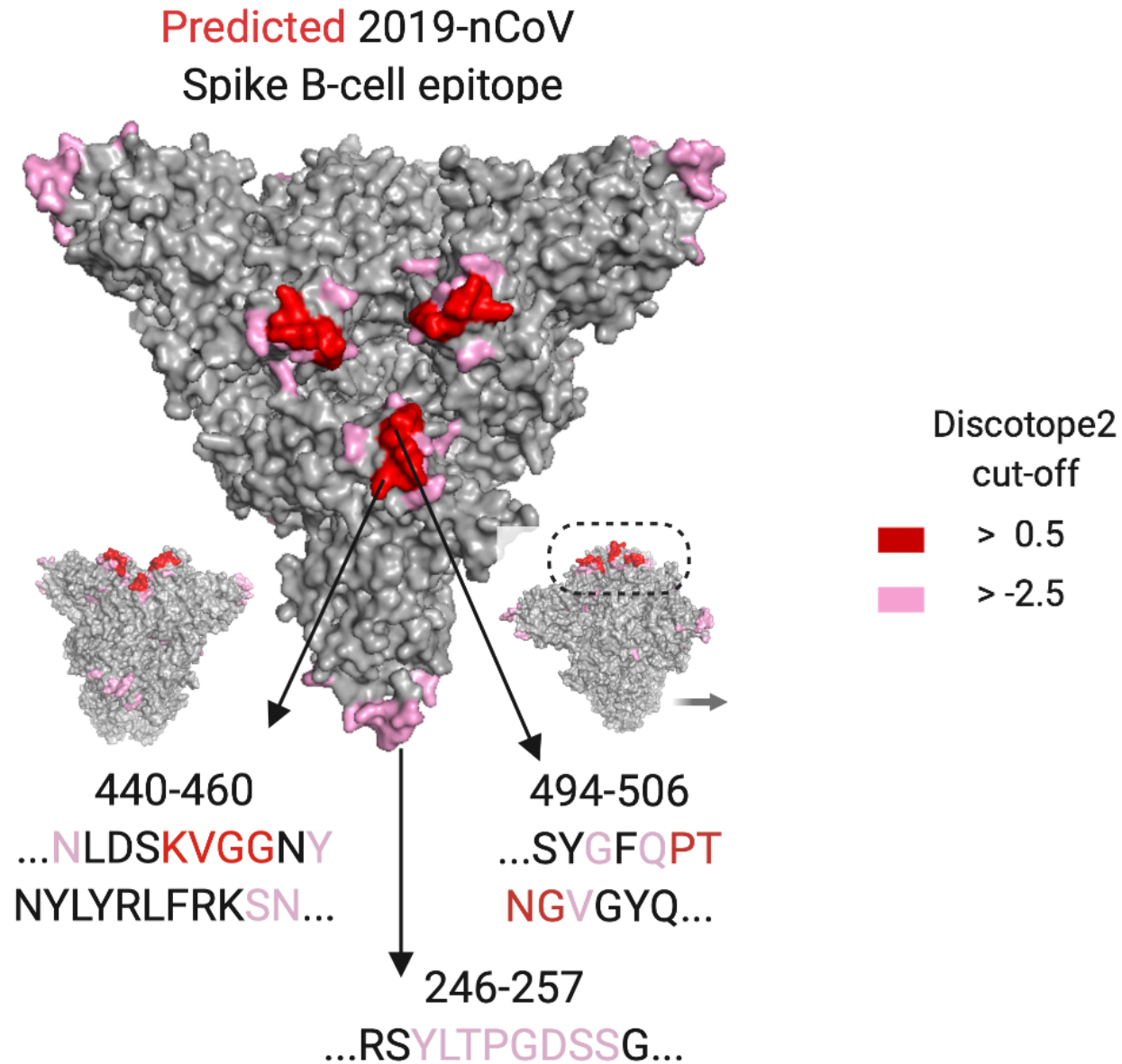
Discotope2 prediction on SARS spike protein consistent with experiments

Predicted B-cell epitopes on S protein structure down conformation

A



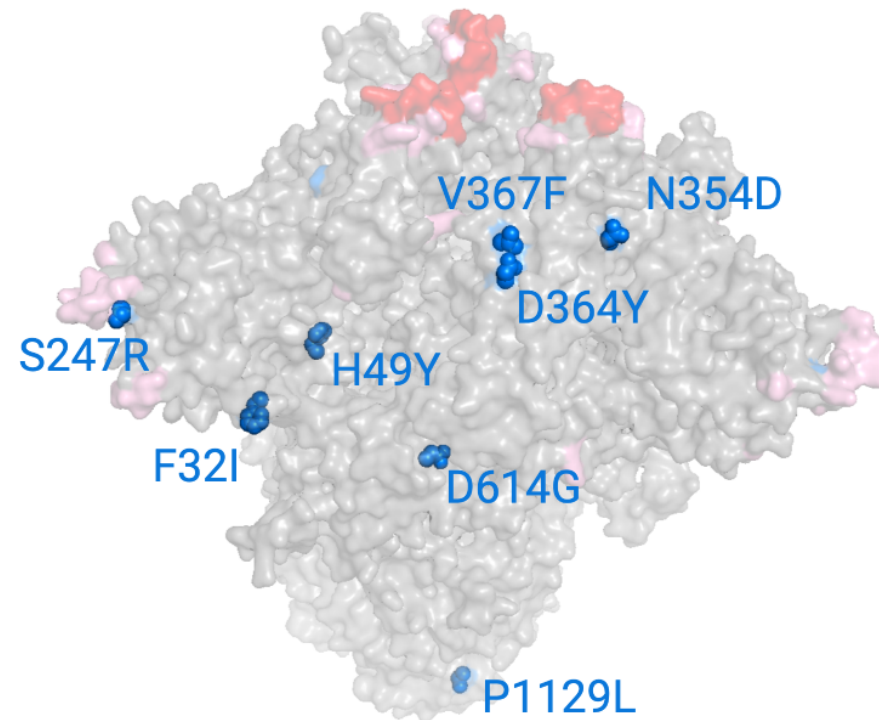
Similar epitopes identified in SARS-CoV-2



As Feb 2020, no spike mutations occur near RBD

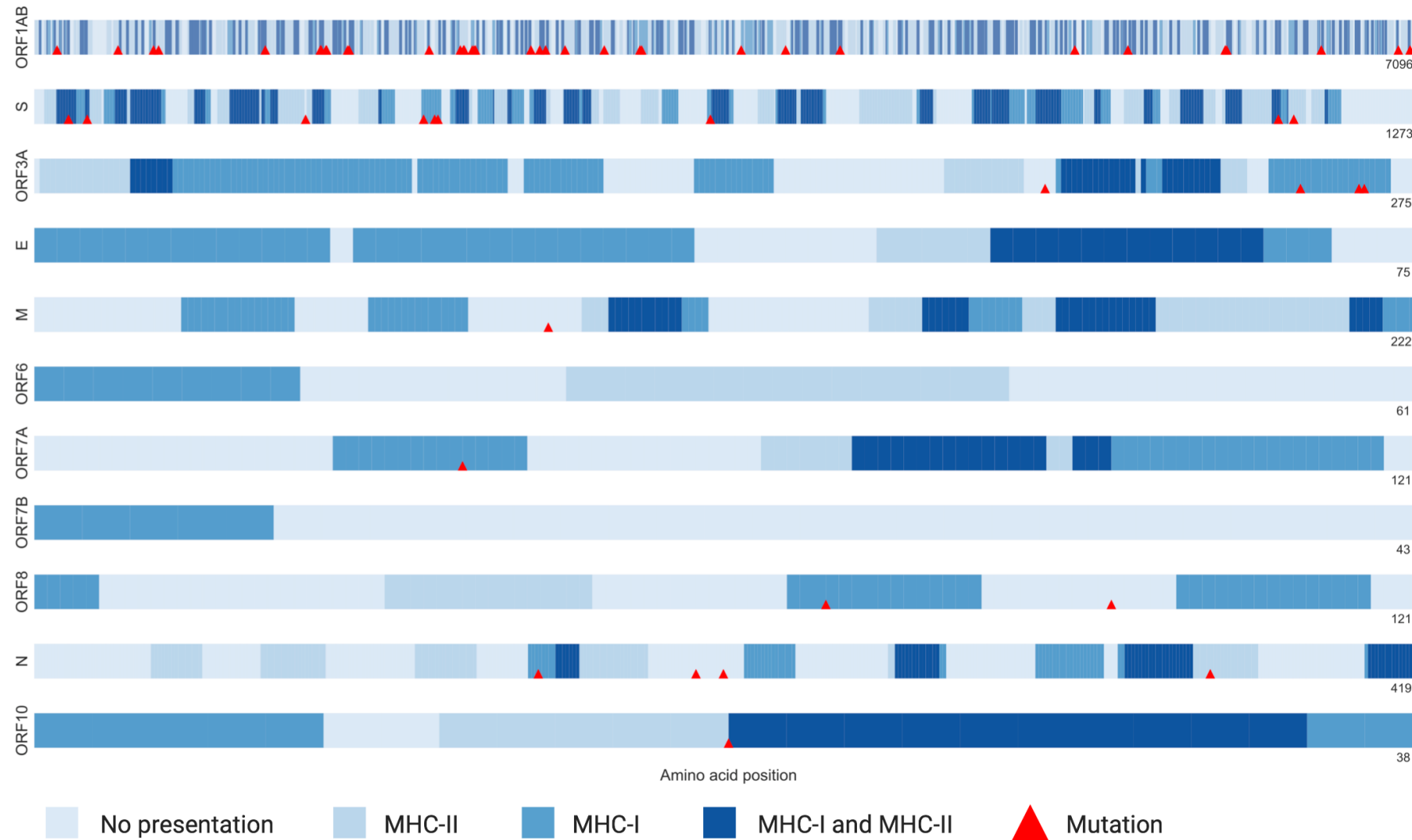
C

Sipke mutations from 68 samples



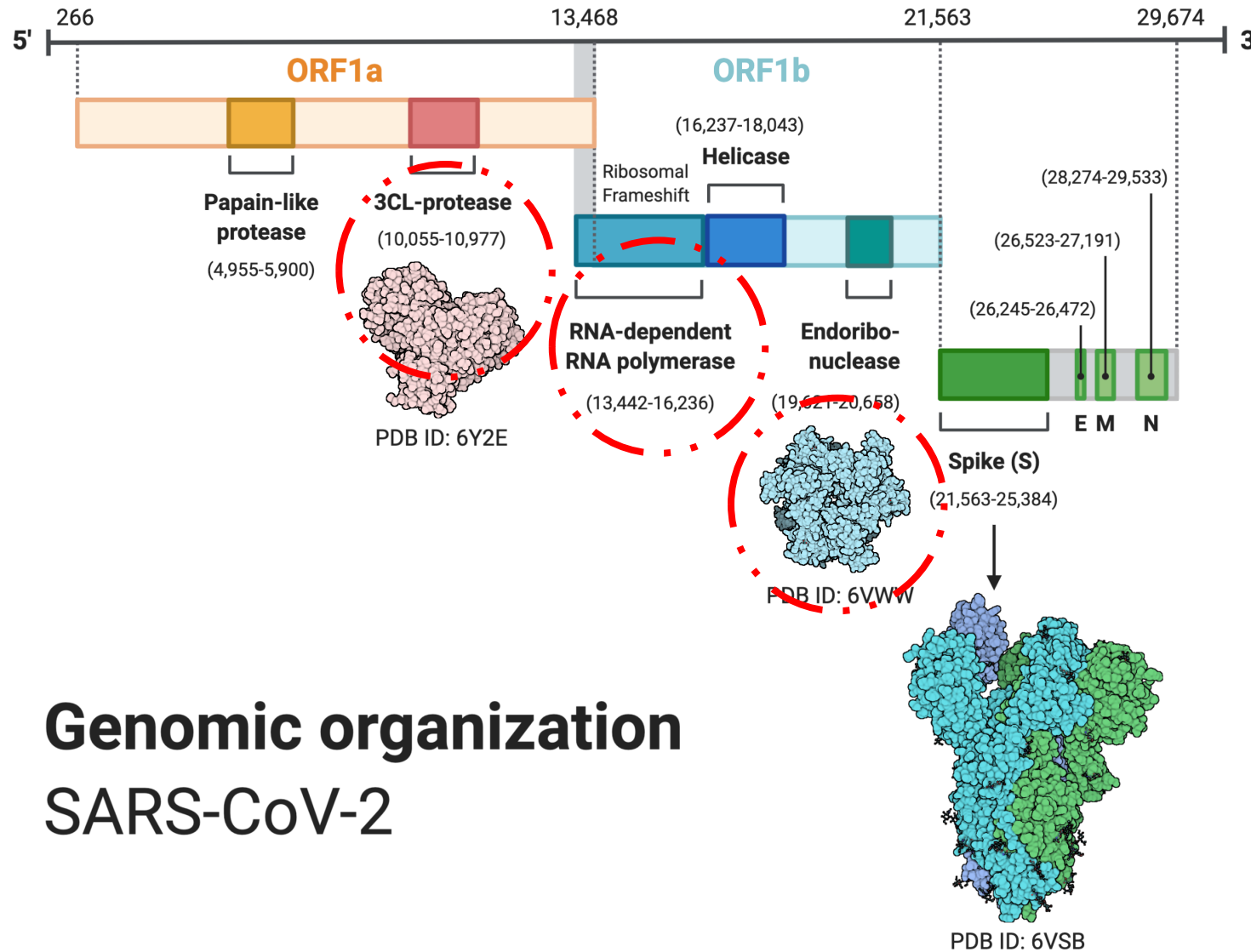
Discotope2 ■ > 0.5
cut-off ■ > -2.5

Mutations are more likely to occur in a region with good MHC-I presentation



Fisher exact
 $P = 0.02$

Other drug targets



Genomic organization
SARS-CoV-2