Table S17: Results of statistical significance tests on the 13 SARS-CoV-2 datasets with balanced scaffold split, where numbers represent the p-values (one-sided significance level) of the McNemar's test between the models in the table (CHEM-BERT, GROVER, MolCLR$_{GIN}$ and MolCLR$_{GCN}$, RNN_LR, RNN_MLP, RNN_RF, TRFM_LR, TRFM_MLP and TRFM_RF) and ImageMol. The numbers in green background indicate statistically different models, using a significance threshold of 0.05.

	CHEM-BERT	GROVER	MolCLR$_{GIN}$	MolCLR$_{GCN}$	RNN_LR
3CL	0.221	0.683	0.221	0.221	1
ACE2	0.48	0.134	0.48	0.48	0.009
HEK293	0.024	0.576	0.073	0.001	0.014
hCYTOX	0.043	0.382	0.07	0.539	0.013
MERS-PPE	0.002	0.64	0.0001	0.0004	0.002
MERS-PPE_cs	0.062	0.232	0.185	0.845	0.016
CoV-PPE	0.012	0.562	0.05	0.185	0.057
CoV-PPE_cs	0.064	0.042	0.136	0.245	0.054
CPE	0.462	0.349	0.482	0.319	0.834
Cytotox	0.017	1.72E-06	1.22E-08	1.22E-08	0.07
AlphaLISA	1.35E-07	0.65	0.671	0.252	3.28E-05
TruHit	2.42E-06	0.456	0.511	0.002	0.0001
TMPRSS2	0.003	0.353	0.29	0.011	4.25E-05

——— continue ———

	RNN_MLP	RNN_RF	TRFM_LR	TRFM_MLP	TRFM_RF
3CL	0.505	0.371	0.289	0.45	0.371
ACE2	0.009	0.077	0.016	0.016	0.852
HEK293	3.12E-05	0.046	0.03	0.0004	0.523
hCYTOX	0.012	0.128	0.196	0.001	0.855
MERS-PPE	0.001	0.018	0.0008	0.0006	0.142
MERS-PPE_cs	0.0002	0.09	0.0271	0.011	0.775
CoV-PPE	0.011	0.002	0.022	2.20E-05	0.153
CoV-PPE_cs	2.33E-06	0.01	0.043	0.056	0.651
CPE	0.293	0.079	0.38	0.025	0.154
Cytotox	0.053	0.0003	1.39E-05	0.004	3.71E-07
AlphaLISA	4.96E-08	0.019	0.318	1.16E-05	0.043
TruHit	1.22E-07	0.0006	0.032	1.23E-04	0.014
TMPRSS2	1.53E-06	0.007	0.004	1.15E-06	0.246