

Table S6: The ROC-AUC performance of different methods on benchmark datasets with random scaffold split. All results are reported as mean \pm standard deviation. The yellow and blue backgrounds represent methods without and with pretraining, respectively. The light green background represents the results of our method. Rank represents the ranking of ImageMol in the comparison. Grover-10M and MPG-10M are our reproduced results based on the 10M pre-training data used by ImageMol.

Classification (Higher is better)						
	BBBP	SIDER	clintox	BACE	Tox21	toxcast
TF Robust	0.860 \pm 0.087	0.607 \pm 0.033	0.765 \pm 0.085	0.824 \pm 0.022	0.698 \pm 0.012	0.585 \pm 0.031
GraphConv	0.877 \pm 0.036	0.593 \pm 0.035	0.845 \pm 0.051	0.854 \pm 0.011	0.772 \pm 0.041	0.650 \pm 0.025
Weave	0.837 \pm 0.065	0.543 \pm 0.034	0.823 \pm 0.023	0.791 \pm 0.008	0.741 \pm 0.044	0.678 \pm 0.024
SchNet	0.847 \pm 0.024	0.545 \pm 0.038	0.717 \pm 0.042	0.750 \pm 0.033	0.767 \pm 0.025	0.679 \pm 0.021
MPNN	0.913 \pm 0.041	0.595 \pm 0.030	0.879 \pm 0.054	0.815 \pm 0.044	0.808 \pm 0.024	0.691 \pm 0.013
DMPNN	0.919 \pm 0.030	0.632 \pm 0.023	0.897 \pm 0.040	0.852 \pm 0.053	0.826 \pm 0.023	0.718 \pm 0.011
MGCN	0.850 \pm 0.064	0.552 \pm 0.018	0.634 \pm 0.042	0.734 \pm 0.030	0.707 \pm 0.016	0.663 \pm 0.009
AttentiveFP	0.908 \pm 0.050	0.605 \pm 0.060	0.933 \pm 0.020	0.863 \pm 0.015	0.807 \pm 0.020	0.579 \pm 0.001
N-GRAM	0.912 \pm 0.013	0.632 \pm 0.005	0.855 \pm 0.037	0.876 \pm 0.035	0.769 \pm 0.027	-
HU. et.al.	0.915 \pm 0.040	0.614 \pm 0.006	0.762 \pm 0.058	0.851 \pm 0.027	0.811 \pm 0.015	0.714 \pm 0.019
GROVER ^{large}	0.940 \pm 0.019	0.658 \pm 0.023	0.944 \pm 0.021	0.894 \pm 0.028	0.831 \pm 0.025	0.737 \pm 0.010
MPG	0.922 \pm 0.012	0.661 \pm 0.007	0.963 \pm 0.028	0.920 \pm 0.013	0.837 \pm 0.019	0.748 \pm 0.005
GROVER-10M	0.940 \pm 0.004	0.691 \pm 0.009	0.956 \pm 0.004	0.923 \pm 0.005	0.840 \pm 0.002	0.741 \pm 0.004
MPG-10M	0.943 \pm 0.0178	0.671 \pm 0.033	0.964 \pm 0.013	0.859 \pm 0.048	0.832 \pm 0.009	0.750 \pm 0.001
ImageMol	0.952\pm0.002	0.708\pm0.010	0.975\pm0.007	0.939\pm0.010	0.847\pm0.003	0.752\pm0.002
Rank	1	1	1	1	1	1
Regression (Lower is better)						
	FreeSolv	ESOL	Lipo	QM7		
TF Robust	4.122 \pm 0.085	1.722 \pm 0.038	0.909 \pm 0.060	120.6 \pm 9.6		
GraphConv	2.900 \pm 0.135	1.068 \pm 0.050	0.712 \pm 0.049	118.9 \pm 20.2		
Weave	2.398 \pm 0.250	1.158 \pm 0.055	0.813 \pm 0.042	94.7 \pm 2.7		
SchNet	3.215 \pm 0.755	1.045 \pm 0.064	0.909 \pm 0.098	74.2 \pm 6.0		
MPNN	2.185 \pm 0.952	1.167 \pm 0.430	0.672 \pm 0.051	113.0 \pm 17.2		
DMPNN	2.177 \pm 0.914	0.980 \pm 0.258	0.653 \pm 0.046	105.8 \pm 13.2		
MGCN	3.349 \pm 0.097	1.266 \pm 0.147	1.113 \pm 0.041	77.6 \pm 4.7		
AttentiveFP	2.030 \pm 0.420	0.853 \pm 0.060	0.650 \pm 0.030	126.7 \pm 4.0		
GROVER	1.544 \pm 0.397	0.831 \pm 0.120	0.560 \pm 0.035	72.6 \pm 3.8		
MPG	1.269 \pm 0.192	0.741 \pm 0.017	0.556 \pm 0.017	-		
GROVER-10M	1.366 \pm 0.175	0.730 \pm 0.026	0.556\pm0.012	72.1 \pm 1.1		
MPG-10M	1.246 \pm 0.131	0.716 \pm 0.062	0.648 \pm 0.003	67.4 \pm 2.9		
ImageMol	1.149\pm0.004	0.690\pm0.003	0.625 \pm 0.009	65.9\pm2.6		
Rank	1	1	4	1		