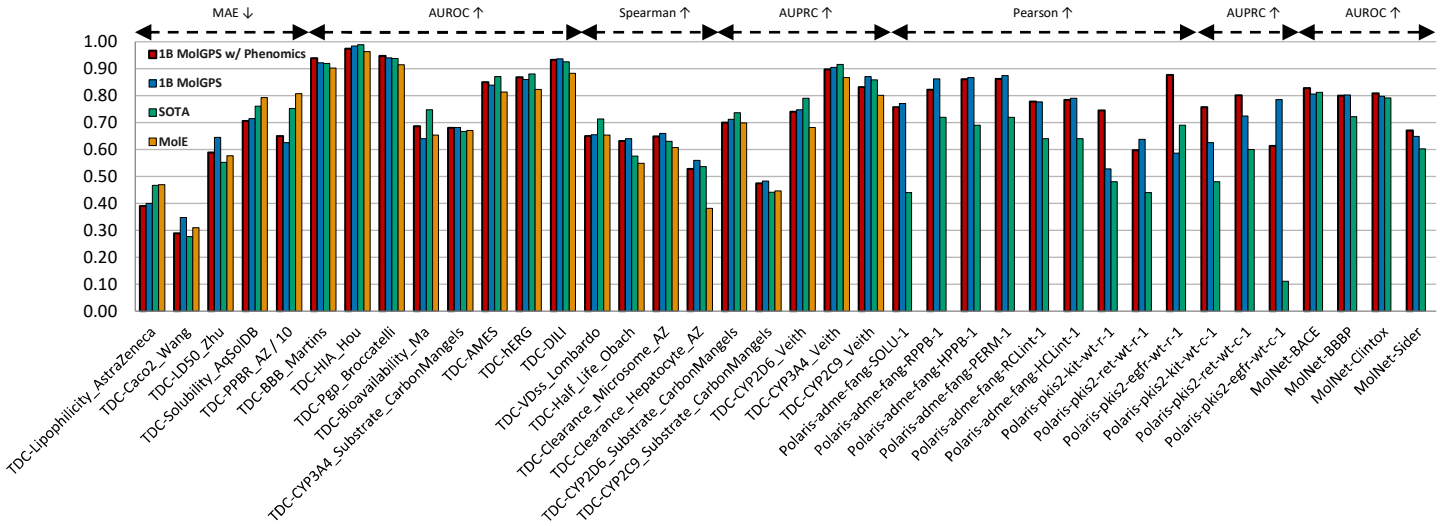

Figure 1: Updated Figure 3 from the submission with the added 1B MolGPS w/ Phenomics pretraining. We observe major improvements for the protein inhibition (pkis2-*) datasets.

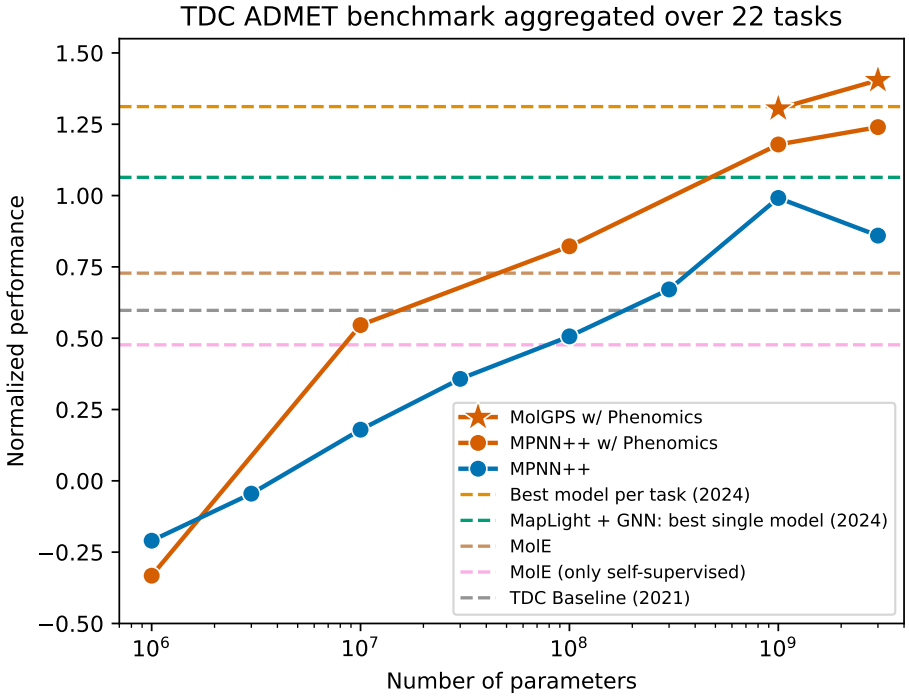

Figure 2: Updated Figure 2 (b) of the submission with additional scaling curve for MPNN++ (and MolGPS) w/ Phenomics pretraining and added additional 3B parameter results. Note that scaling curves slightly changed as we are now showing MPNN++ instead of GPS++ in the original submission (because our experiments with Phenomics pretraining were only conducted for MPNN++ for parameter scales below 1B).

Table 1: Comparison of MolGPS variants to the self-supervised GraphMVP model on the MoleculeNet dataset. Note that we have only the datasets that were not part of our pretraining.

Method	BACE	BBBP	Clintox	Sider
GraphMVP	0.812	0.724	0.775	0.639
1B MolGPS	0.806	0.802	0.797	0.649
1B MolGPS w/ Phenomics	**0.828**	**0.8**	0.**809**	**0.67**
3B MolGPS w/ Phenomics	**0.832**	**0.809**	**0.807**	**0.666**

Method	Norm.	Lipo. MAE	Caco2 MAE	LD50 MAE	Solu. MAE	PPBR MAE	BBB AUROC	HIA AUROC	Pgp AUROC	Bioavail. AUROC	CYP3A4_S AUROC	AMES AUROC	hERG AUROC	DILI AUROC	VDss Spear.	Half_Life Spear.	Clear_M Spear.	Clear_H Spear.	CYP2D6_S AUPRC	CYP2C9_S AUPRC	CYP2D6_V AUPRC	CYP3A4_V AUPRC	CYP2C9_V AUPRC
MolE (FuncEnv, only self-supervision)	0.47	0.46	0.355	0.597	0.799	8.57	0.895	0.951	0.873	0.638	0.612	0.831	**0.871**	0.89	0.622	0.579	0.567	0.373	**0.715**	0.411	0.678	0.857	0.759
MolE (AtomEnv, only self-supervision)	0.477	0.464	0.471	0.582	0.81	8.191	0.895	0.949	0.871	0.683	0.633	0.832	0.844	0.883	0.596	0.518	0.531	0.367	0.706	0.429	0.665	0.865	0.773
MolE	0.728	0.469	0.31	**0.577**	0.792	8.073	0.903	0.963	0.915	0.654	0.67	0.813	0.823	0.883	**0.654**	0.549	0.607	0.381	0.699	0.446	0.682	0.867	0.801
1B MolGPS ens	1.222	0.4	0.347	0.645	0.714	**6.249**	0.922	**0.984**	0.941	0.64	**0.681**	0.8389	0.86	**0.937**	**0.655**	**0.64**	**0.659**	**0.56**	0.712	**0.483**	**0.747**	**0.905**	**0.871**
1B MolGPS w/ Phenomics	**1.305**	**0.391**	**0.288**	0.589	**0.706**	6.497	**0.939**	0.975	**0.947**	**0.686**	**0.681**	**0.85**	**0.868**	0.933	0.649	**0.632**	**0.649**	0.527	0.700	**0.474**	0.741	0.898	0.832
3B MolGPS w/ Phenomics	**1.404**	**0.386**	**0.292**	**0.557**	**0.679**	**6.464**	**0.941**	**0.98**	**0.948**	**0.701**	0.68	**0.857**	0.864	**0.942**	0.649	0.631	0.633	**0.57**	**0.713**	0.464	**0.75**	**0.9**	**0.838**

Table 2: Comparison of MolGPS variants to the self-supervised variants of MolE and the standard MolE (supervised + self-supervised pretraining) on TDC benchmark collection. Supervised pretraining significantly improves MolE performance as seen in the normalized performance (left-most column). MolGPS models outperform all MolE models by a large margin, exhibiting the best performance in all but two tasks.