Significance analysis of the the performance of the 0% and 30% pre-training dataset size setups / split = random Permeability Solubility HLM - NS - NS - NS MLM\_0% | MLM\_0% MLM\_0% 0.095 0.095 0.222 - p < 0.05 - p < 0.05 - p < 0.05p < 0.01 p < 0.01 MLM\_30% MLM\_30% 0.095 0.095 0.222 - p < 0.001 MLM\_30% WFW \_00/0 MLM\_30% MLM\_30% MLM\_0% MTW \_00/0 hPPB rPPB **RLM** - NS - NS - NS MLM\_0% 1 MLM\_0% , MLM\_0% 1 0.841 0.841 0.032 - p < 0.05 - p < 0.05 -p < 0.05p < 0.01 p < 0.01 - p < 0.01 MLM\_30% MLM\_30% 0.841 0.841 0.032 p < 0.001 MLM\_30% MLM\_30% MLM\_30% MFW \_00/0 MLM\_0% MLM\_0%