



Figure S4: Heatmaps of p -values according to one-sided Wilcoxon rank-sum test for OneProt ablations. The alternative hypothesis of the models on the vertical axis outperforming the models on the horizontal according to metric values from Tables 3 and 4 of the main text, where for binary classification accuracy was compared. Striped pattern stands for the values $p < 0.05$, when the null hypothesis of model on the vertical axis being non-superior was rejected.