



Figure S1: **Heatmaps of p -values according to one-sided Wilcoxon rank-sum test.** The alternative hypothesis of OneProt models (vertical axis) outperforming baseline models (horizontal axis) 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 OneProt being non-superior was rejected.