



Figure S3: **Heatmaps of p -values for Area Under Receiver Operating Characteristic curve metrics.** One-sided Wilcoxon rank-sum test with the alternative hypothesis of OneProt (vertical axis) outperforming baseline models (horizontal axis), striped pattern corresponding to values $p < 0.05$ (upper panel). Two-sided Wilcoxon rank-sum test with the alternative hypothesis of OneProt performing differently than baseline models, striped pattern corresponding to values $p \geq 0.05$ (middle panel). One-sided Wilcoxon rank-sum test for OneProt ablations with the alternative hypothesis of the models on the vertical axis outperforming the models on the horizontal axis (bottom panel). Striped pattern stands for the values $p < 0.05$.