



**Supplementary Fig. S10.** Results of finetuned SignalP 6.0 and finetuned ESM-2 model family with linear chain CRF in terms of MCC1 and MCC2 for SP prediction. The bold font in the x-axis represents the SP types with small training samples. The MCC1 and MCC2 scores are shown above the bars. The sorted mean for MCC1 and MCC2 are listed at the top. (A) MCC1 scores performance on negative class composed of soluble and transmembrane proteins. (B) MCC2 scores performance on negative class comprising soluble and transmembrane proteins and other SP types.