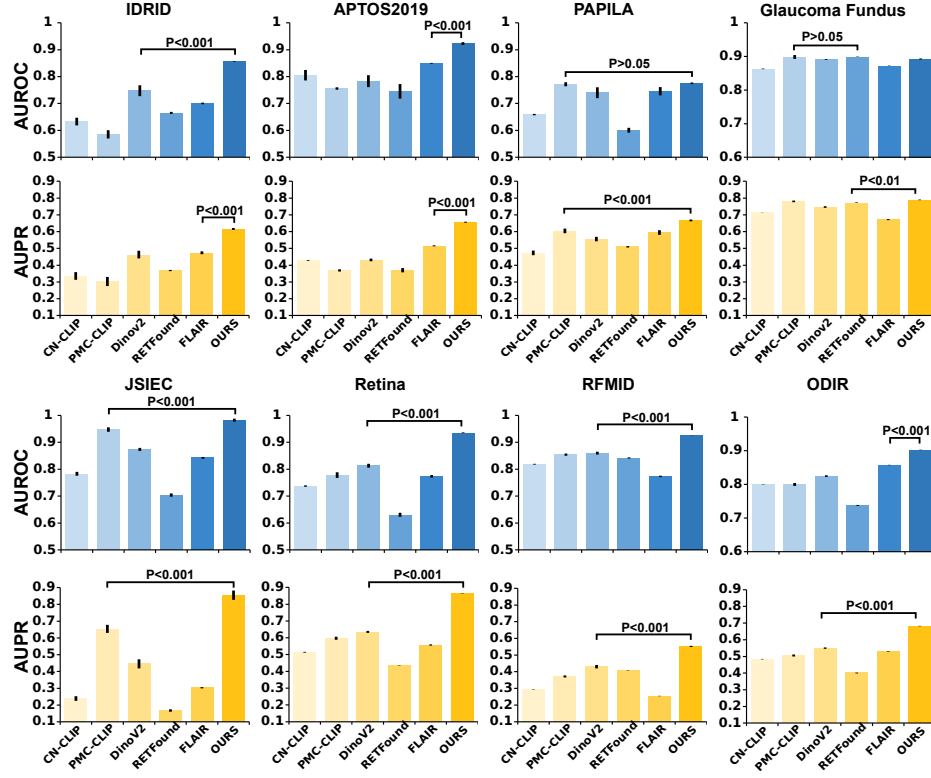# Supplementary Materials

Fig. 1. More details beyond the mean values for linear probing. The error bars show 95% confidence intervals and the bar centre represents the mean value of the AUROC and AUPR. P value is calculated with the two-sided t-test on the best and second-best results to determine the significance level at which the top-performing method surpasses the others. The proposed RET-CLIP consistently outperforms the second-best methods across nearly all downstream datasets, with statistically significant superiority.
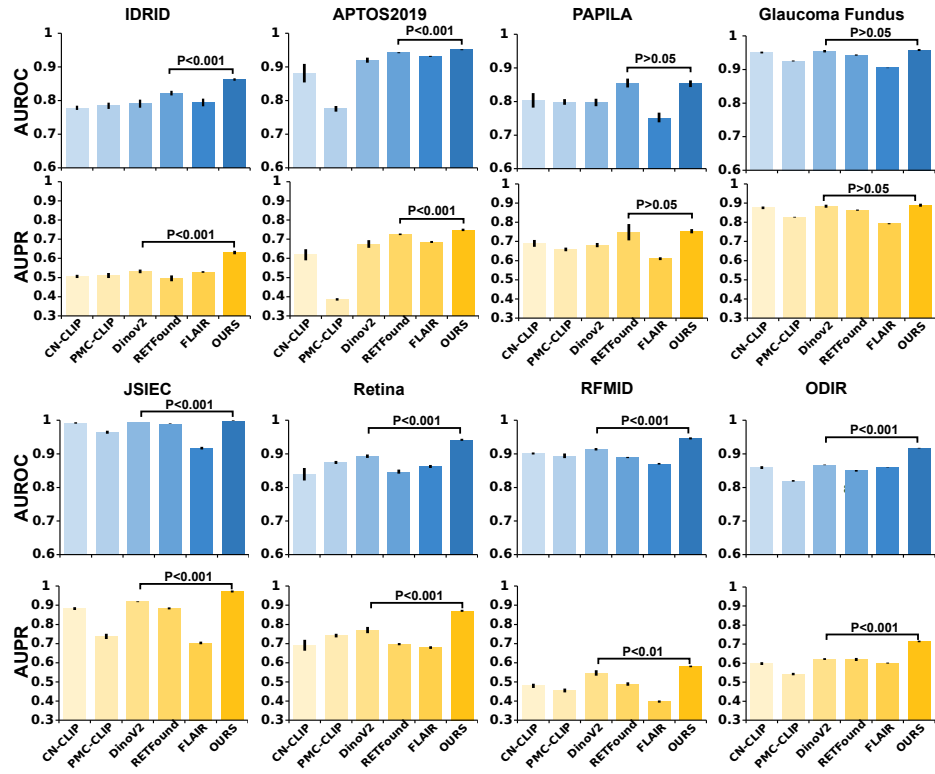
Fig. 2. More details beyond the mean values for fine tuning.