

**Additional Table 4.** Comparison among HTD and TPR ensemble variants considering only the best predictions for the newly annotated genes. Average AUROC and AUPRC across terms and average  $F_{max}$ , Precision and Recall across genes considering only HPO terms with AUROC > 0.7 (778 terms) and  $F_{max}$  > 0.3 (296 genes). Best results for each metric are highlighted in bold.

Meas. Methods	AUROC	AUPRC	$F_{max}$	Precision	Recall
<i>HTD</i>	0.8155	0.1551	0.4716	0.4429	0.5042
<i>TPR-T</i>	0.8155	0.1553	0.4714	0.4443	0.5021
<i>TPR-D</i>	0.8155	0.1554	0.4714	0.4436	0.5030
<i>TPR-TF</i>	0.8200	0.1577	0.4765	0.4419	<b>0.5171</b>
<i>TPR-W</i>	<b>0.8219</b>	<b>0.1594</b>	<b>0.4793</b>	<b>0.4572</b>	0.5037