

Supplementary Information for:

Protein Functional Annotation of Simultaneously Improved Stability, Accuracy and False Discovery Rate Achieved by a Sequence-based Deep Learning

Jiajun HONG^{1,2}, Yongchao LUO², Yang ZHANG³, Junbiao YING², Weiwei XUE³, Tian XIE¹, Lin TAO^{1,*},
Feng ZHU^{1,2,*}

¹Key Laboratory of Elemene Class Anti-cancer Chinese Medicine of Zhejiang Province, School of
Medicine, Hangzhou Normal University, Hangzhou 310036, China

²College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

³School of Pharmaceutical Sciences, Chongqing University, Chongqing 401331, China

*To whom the correspondence should be addressed: Prof. Feng ZHU, College of Pharmaceutical Sciences,
Zhejiang University, Hangzhou 310058, China. E-mail: zhufeng@zju.edu.cn; Prof. Lin TAO, School of
Medicine, Hangzhou Normal University, Hangzhou 310036, China. E-mail: taolin@hznu.edu.cn.

Jiajun HONG, Yang ZHANG and Junbiao YING are the doctoral, Master and undergraduate students of the
College of Pharmaceutical Sciences in Zhejiang University, China and jointly cultivated by the School of
Pharmaceutical Sciences in Chongqing University, China. They are interested in Artificial Intelligence.

Weiwei XUE is professor of the School of Pharmaceutical Sciences in Chongqing University, China. He is
interested in the area of computer-based drug design and molecular dynamics simulation.

Tian XIE and Lin TAO are professors of the School of Medicine in Hangzhou Normal University, China.
They are interested in the area of Traditional Chinese Medicine, bioinformatics and machine learning.

Feng ZHU is professor of the College of Pharmaceutical Sciences in Zhejiang University, China. He got his
Ph.D. degree from the National University of Singapore, Singapore. His research group (<https://idrblab.org/>)
has been working in the fields of bioinformatics, OMIC-based drug discovery, system biology and medicinal
chemistry. Welcome to visit his personal website at: <https://idrblab.org/Peoples.php>

Running Title: Protein Function Annotation of Improved Stability, Accuracy and False Discovery Rate

Supplementary Table S1. The statistical differences of four measurements between any two methods. *P*-values with Bold Font indicate the statistical difference.

Methods	CNN-SVM	CNN-KNN	CNN-PNN	SVM-KNN	SVM-PNN	KNN-PNN
MCC	0.000	0.000	0.000	0.210	0.557	0.463
Accuracy	0.046	0.006	0.015	0.437	0.755	0.613
Sensitivity	0.007	0.313	0.117	0.077	0.288	0.526
Specificity	0.082	0.000	0.002	0.001	0.044	0.197

Supplementary Table S2. The statistical differences of the EFs between any two methods. *P*-values with Bold Font indicate the statistical difference.

Methods	CNN- BLAST	CNN- PHMM	CNN- SVM	CNN- KNN	CNN- PNN	BLAST- PHMM	BLAST -SVM	BLAST -KNN	BLAST -PNN	PHMM -SVM	PHMM -KNN	PHMM -PNN	SVM- KNN	SVM- PNN	KNN- PNN
Enrichment Factor	0.020	0.054	0.024	0.007	0.025	0.275	0.998	0.223	0.882	0.377	0.040	0.403	0.338	0.906	0.233