

# MetaGOPlus: Improving Gene Ontology Prediction of Proteins Using Deep Residual Network with Hierarchical Classification

Yi-Heng Zhu <sup>1,2</sup>, Chengxin Zhang <sup>2</sup>, Rucheng Diao <sup>3</sup>, Xiaogen Zhou <sup>2</sup>, Peter L. Freddolino <sup>3</sup>,  
Dong-Jun Yu <sup>1</sup>, Yang Zhang <sup>2,3</sup>

<sup>1</sup> School of Computer Science and Engineering, Nanjing University of Science and Technology,  
200 Xiaolingwei, Nanjing, 210094, China

<sup>2</sup> Department of Computational Medicine and Bioinformatics, University of Michigan,  
100 Washtenaw, Ann Arbor, Michigan, 48109-2218, United States

<sup>3</sup> Department of Biological Chemistry, University of Michigan,  
100 Washtenaw, Ann Arbor, Michigan, 48109-2218, United States

## Abstract

Gene ontology (GO) has been widely used to annotate functions of proteins. Accurate identification of GO attributes from proteins provides critical help in understanding the biological activities of proteins. We proposed a new pipeline, MetaGOPlus, to predict the GO attributes of proteins. MetaGOPlus consists of five sub-pipelines whose predictions are effectively ensembled through logistic regression. These sub-pipelines include a new deep-learning-based model and four sub-pipelines inherited from MetaGO using structure alignment, sequence profile comparison, protein-protein interaction, and naïve probability. The newly proposed deep-learning-based sub-pipeline uses deep residual network (ResNet) as the basic model and extracts multiple sequence-based features as the input of model. Moreover, a hierarchical classification layer at the end of ResNet can fully utilize inter-relationships between GO attributes. The proposed pipeline was tested on a large-scale set of 1000 non-redundant proteins from CAFA3 experiment. Computational experimental results show that MetaGOPlus achieves better performance than other state-of-the-art function annotations methods. Detailed analyses show that the major advantages of MetaGOPlus stem from two points. First, the newly deep-learning-based sub-pipeline can effectively learn hidden relationships between proteins and GO attributes from large-scale benchmark dataset. Second, five sub-pipelines can effectively learn the knowledge of GO attributes from different and complementary views.



# Certificate of Participation & Presentation

INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

July 16, 2020

Yiheng Zhu  
1040 Island Drive  
Ann Arbor, Michigan  
48105  
United States

**Subject: CERTIFICATION OF PARTICIPATION AND PRESENTATION**

On behalf of the Organizing Committee for ISMB 2020 held as virtual conference, thank you for your participation and presentation of "*MetaGOPlus: Improving Gene Ontology Prediction of Proteins Using Deep Residual Network with Hierarchical Classification*"

This letter certifies that Yiheng Zhu was a participant at ISMB 2020 and presented the above research.

ISMB 2020, the 28th annual conference, was hosted by the International Society for Computational Biology (ISCB) July 13 - 16, 2020. The ISMB conference has grown to become the world's largest bioinformatics/computational biology conference.

Please visit <https://www.iscb.org/ismb2020> for additional conference information.

Yours sincerely,

A handwritten signature in black ink that reads 'Diane E Kovats'.

Diane E. Kovats, CAE, CMP, DES  
Executive Director  
International Society for Computational Biology (ISCB)  
e: [dkovats@iscb.org](mailto:dkovats@iscb.org)  
[www.iscb.org](http://www.iscb.org)