

1 **To improve the predictions of binding residues with DNA, RNA,**  
2 **carbohydrate, and peptide via multiple-task deep neural networks**

3 **Zhe Sun<sup>1</sup>, Shuangjia Zheng<sup>1</sup>, Huiying Zhao<sup>2</sup>, Zhangming Niu<sup>3</sup>, Yutong Lu<sup>1</sup>, Yi Pan<sup>5</sup> and**  
4 **Yuedong Yang<sup>1,4\*</sup>**

5 <sup>1</sup>School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510000, China

6 <sup>2</sup>Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510000, China

7 <sup>3</sup>Aladdin Healthcare Technologies Ltd. 24-26 Baltic Street West, London EC1Y 0UR, UK

8 <sup>4</sup>Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of  
9 Education, China

10 <sup>5</sup>Department of Computer Science, Georgia State University, USA

11

12 **Abstract**

13 **Motivation:** The interactions of proteins with DNA, RNA, peptide, and carbohydrate play key roles in  
14 various biological processes. The studies of uncharacterized protein–molecules interactions could be  
15 aided by accurate predictions of residues that bind with partner molecules. However, the existing  
16 methods for predicting binding residues on proteins remain of relatively low accuracies due to the  
17 limited number of complex structures in databases. As different types of molecules partially share  
18 chemical mechanisms, the predictions for each molecular type should benefit from the binding  
19 information with other molecules types.

20 **Results:** In this study, we employed a multiple task deep learning strategy to develop a new  
21 sequence-based method for simultaneously predicting binding residues/sites with multiple important  
22 molecule types named MTDsite. By combining four training sets for DNA, RNA, peptide, and  
23 carbohydrate-binding proteins, our method yielded accurate and robust predictions with AUC values of  
24 0.852, 0.836, 0.758, and 0.776 on their respective independent test sets, which are 0.52 to 6.6% better  
25 than other state-of-the-art methods. More importantly, this study provides a new strategy to improve  
26 predictions by combining multiple similar tasks.

27 **Availability:** <http://biomed.nscg-gz.cn/server/MTDsite/>

28 **Contact:** [yangyd25@mail.sysu.edu.cn](mailto:yangyd25@mail.sysu.edu.cn)

29