

From Protein Sequence to Structure to Function

Xin Gao

King Abdullah University of Science and Technology

Proteins are basic building blocks of life. The amino acid sequence of a protein determines its 3D structure, which, in turn, provides indispensable information about its function. Sequence, structure and function can be considered as three modalities of proteins. Each of these modalities can be represented as a network in its corresponding space.

Despite the large amount of research on predicting protein structures from sequences or predicting functions from sequences/structures, not much work has been done on analyzing the homogeneity and heterogeneity of these networks. Even less has been done on simultaneously combining the information from different networks together.

For example, existing protein structure prediction methods have successfully used information in the sequence network, and that between the sequence and structure networks, but have not encoded information in the structure network.

In this talk, I will present our recent work on developing cross-modal methods to coherently combine information from different networks. Our results demonstrate that doing so can significantly improve the performance for protein structure and function prediction.

Speaker biography

Dr. Xin Gao is an assistant professor of computer science in Computational Bioscience Research Center in Computer, Electrical and Mathematical Sciences and Engineering Division at King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

Prior to joining KAUST, he was a Lane Fellow at Lane Center for Computational Biology in School of Computer Science at Carnegie Mellon University, U.S.. He earned his bachelor degree in Computer Science in 2004 from Computer Science and Technology Department at Tsinghua University, China, and his Ph.D. degree in Computer Science in 2009 from David R. Cheriton School of Computer Science at University of Waterloo, Canada.

His research interests are building computational models, designing efficient and effective algorithms, and developing machine learning techniques to solve key open problems in structural biology, systems biology and synthetic biology.