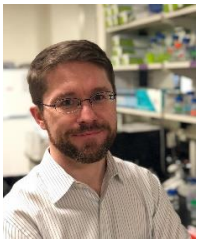
Biomedical Informatics Grand Rounds

Wednesday, April 20th, 2022 3:00 pm – 4:00 pm



Using neural networks to understand the DNA-binding specificity of transcription factors

Shaun Mahony, PHD
Associate Professor
The Department of Biochemistry & Molecular Biology and the Center for Eukaryotic Gene Regulation
The Pennsylvania State University

Remote Access
Join Zoom Meeting https://stonybrook.zoom.us/j/95617197636?pwd=KytzZ2pVRG9SZGpKZUtpNXJISjNjZz09
Meeting ID: 956 1719 7636 Passcode: 924293

Bio: Dr. Shaun Mahony received his B.E. in Electronic & Computer Engineering and his Ph.D. research in computational biology, at the National University of Ireland, Galway. He was a Postdoctoral Associate at the University of Pittsburgh (2 years), a Research Scientist in the Computer Science and Artificial Intelligence Laboratory at MIT (5 years), and a visiting scientist in the Department of Stem Cell and Regenerative Biology at Harvard University (1 year). Since 2012, he has been a faculty member in the Department of Biochemistry & Molecular Biology and the Center for Eukaryotic Gene Regulation at Penn State University. His research group focuses on building machine learning applications to understand transcription factor binding specificity, particularly in the context of cell fate decisions. He has published over 50 research articles, and his research is currently sponsored by a CAREER award from NSF and a MIRA award from NIH.

Abstract: The goal of my computational biology research is to understand how cellular identities are controlled by the dynamic regulatory activities of transcription factors (TFs). The typical vertebrate TF should have binding affinity for millions of sites along the genome, yet only a small fraction appears to be bound in a given cell type. Furthermore, a TF's binding sites can vary broadly across cell types. While TF regulatory specificity is key to understanding the establishment of cell fates, we still know little about how such specificity results from the interplay between a TF's sequence preference and cell-specific chromatin environments. My lab at Penn State aims to build machine learning applications that yield deeper insight into TF regulatory specificity.

To understand the determinants of TF binding specificity, we need to examine how newly activated TFs interact with sequence and pre-existing chromatin landscapes to determine their binding sites. In this talk, I'll present a principled approach to model the sequence and pre-existing chromatin determinants of TF binding. Specifically, we develop a bimodal neural network that jointly models sequence and prior chromatin data to interpret the binding specificity of TFs that have been induced in well-characterized chromatin environments. The bimodal network architecture allows us to quantify the degree to which sequence and prior chromatin features explain induced TF binding, both at individual sides and genome-wide. Our approach thus provides a framework for modeling, interpreting, and visualizing the joint sequence and chromatin landscapes that determine in vivo TF binding dynamics. I'll demonstrate our approach by using it to analyze the binding specificity of the two main proneural TFs in vertebrates, Ascl1 and Neurog2, when each is expressed in mouse embryonic stem cells. I'll show that the pre-existing chromatin landscape is an important determinant of these TFs' binding specificity. Furthermore, the binding patterns of the proneural TFs result in differential chromatin accessibility and activity landscapes that affect the genomic binding of shared downstream neurogenic TFs, and thereby have a profound impact on the subtypes of neurons that can be generated.

Educational Objects: Upon completion, participants should be able to:

- Introduce the challenges associated with characterizing transcription factor DNA-binding specificity.
- Illustrate how convolutional neural networks can be applied to characterize and predict functional elements in DNA sequences.
- Discuss the gene regulatory mechanisms that determine neuronal differentiation pathways.

Disclosure Statement: The faculty and planners have no relevant financial relationship with ineligible companies whose primary business is producing, marketing, selling, re-selling, or distributing health care products used by or on patients.

Continuing Medical Education Credits: The School of Medicine, State University of New York at Stony Brook, is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians. The School of Medicine, State University of New York at Stony Brook designates this live activity for a maximum of **1** ***AMA PRA Category 1 Credits™.*** Physicians should only claim credit commensurate with the extent of their participation in the activity.