

MAS20-2020-000097

Abstract for an Invited Paper  
for the MAS20 Meeting of  
the American Physical Society

**Massively Parallel Reporter Assays, Machine Learning, and the Biophysics of Gene Regulation<sup>1</sup>**

JUSTIN KINNEY, Cold Spring Harbor Laboratory

Gene expression in all organisms is controlled by short DNA and RNA sequences called cis-regulatory elements (CREs). Proteins in the cellular milieu bind to nucleic acid sequences present within CREs, interact with one another, and thus form macromolecular complexes that modulate the expression of nearby genes. My lab uses a combination of experiments and mathematical modeling to study the biophysics of these regulatory processes in living cells. Our experiments are based on massively parallel reporter assays, which leverage the power of ultra-high-throughput DNA sequencing to measure the effects that thousands to millions of different CRE sequence perturbations have on gene expression. Our mathematical modeling efforts aim to infer biophysical models from the large DNA sequence datasets that these experiments yield. This machine learning task has proven to be remarkably fertile from a theoretical standpoint, tying together ideas in Bayesian field theory, information theory, dimensionality reduction, and deep learning.

<sup>1</sup>NIH grants R35GM133777 and P30CA045508, and the Cold Spring Harbor Laboratory / Northwell Health Alliance.