

ABSTRACT

MACHINE LEARNING FOR THE STUDY OF GENE REGULATION AND COMPLEX TRAITS

By

Anne Sonnenschein

Functional elements are found in DNA outside of protein coding regions; an important class of these elements are 'enhancers', which govern when and where transcription occurs. Predicting the identity and function of potential enhancers based on DNA sequence remains a major goal of genomics. A number of features are associated with the enhancer state, but even combinations of these features in well-studied systems such as *Drosophila* have limited predictive accuracy. I have examined the current limits of computational enhancer prediction, and analyzed which features are most useful for this task, by applying machine-learning methods to an extensive set of genomic features.

Inferring the genetic underpinning of even well-characterized phenotypes is equally challenging, although similar analytical methods can be applied. Phenotypes are frequently defined based on a set of characteristic features; when images are used as specimens, these features are frequently based on morphometric landmarks, although computational pattern-recognition has been used as an alternative. I use *Drosophila* wing shape as a model for a complex phenotype, and use machine learning to predict underlying genotype using both traditional landmarks and features extracted using 'computer vision'.

Copyright by

ANNE SONNENSCHN

2017