Learning nucleosome binding energies

Mark Heron[1,2], Johannes Söding[2]
[1]Gene Center, LMU Munich; [2]Max Planck Institute for Biophysical Chemistry, Göttingen

Genomic DNA is wrapped around histones creating so called nucleosomes. This chromatin packing affects the local accessibility, which is important for most DNA binding factors. While nucleosomes have no strong binding motif, the physical task of wrapping the 147 DNA base pairs around the histone core does lead to preferred sequence features. These sequence features have been used extensively by others, directly or as basis for binding energies in thermodynamic models, to predict genome wide nucleosome positioning.

We optimize the sequence-based nucleosome binding energies in a maximum likelihood approach. We compute the likelihood of the genome-wide nucleosome position measurements given a thermodynamic model with steric hindrance effects between nucleosomes. Our predictions only show minor improvements compared to other methods for predicting genome-wide nucleosome occupancy. However, the validation is limited by comparison to nucleosome measurements, which we show are influenced by biases that distort the validation. (MNase-Seq has low positional resolution; the chemical map has a sequence based cut site preferences; and both have issues with their occupancy values.) By accounting for biases specific to either measurement type, we can partially compensate for them, allowing us to learn similar binding energy models from nucleosome data of distinct experimental techniques.