An Assessment of Prediction Algorithms for Nucleosome Positioning

YOSHIAKI TANAKA[1,2] KENTA NAKAI[1,2,3]

[1] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[3] Institute for Bioinformatics Research and Development (BIRD), Japan Science and Technology Agency, 5-3 Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan

Nucleosome configuration in eukaryotic genome is one of important clues to clarify the mechanisms of regulation for various nuclear events. In the past few years, numerous computational tools have been developed for identification of nucleosome positioning, but there is no benchmark about their performances. Here we propose a performance evaluation test using genome-scale nucleosome maps in two vertebrates and three invertebrates. With our measurement, the recently updated versions of Segal's models and Gupta's SVM with RBF kernel showed high prediction accuracies. However, we also observed significantly different performances between these methods in medaka and candida nucleosomes. Additionally, we showed that Gupta's SVM exhibited the best performance in predicting genomic nucleosome locations with a model trained by samples from the same organism in medaka and yeast. With the analyses for over- and under-representation of DNA oligomers, we found several sharing and species-specific motifs in nucleosome and linker DNA and showed that medaka nucleosome DNA sequences are quite different from those in the other species. Surprisingly, oligomers commonly observed in all five eukaryotes are only CA/TG and AC/GT. Our results indicate that each prediction tool shows the different performance in each eukaryote by the difference of sequence dependencies in nucleosome positioning.