Revisiting digital footprinting of DNase-seq for detection of cell-specific binding sites

Eduardo G. Gusmao, Manuel Allhoff, Martin Zenke and Ivan G. Costa

IZKF Research Group Bioinformatics, RWTH Aachen Medical Faculty

Abstract

DNase-seq is a powerful technique for detection of cell-specific binding sites in a genome-wide manner. Computational footprinting methods, which search for footprint-like DNase I cleavage patterns on the DNA, allow the detection of binding sites in a base pair resolution. There is, however, a debate in the literature on the influence of experimental artifacts as DNase I cleavage bias and transcription factor residence time on computational footprint methods. We investigated these artifacts in a comprehensive panel of DNase-seq data sets employing 10 footprinting methods and 88 transcription factors (Gusmao *et al.*, 2016). Our comparative analysis indicates the advantage of three methods (HINT, DNase2TF and PIQ) in relation to other competing methods. We demonstrate that correcting the DNase-seq signal based on cleavage bias estimation significantly improves accuracy of computational footprinting. We also propose a score to detect footprints arising from transcription factors with short residence time, as footprints of such factors have low predictive performance.

Gusmao E.G., Allhoff, M., Zenke, M., Costa, I.G. (2016), Analysis of computational footprinting methods for DNase sequencing experiments, *Nature Methods*, 13, 303–309.