Name

George F. Steinhardt

Email

george.steinhardt@gmail.com

Institutional Affiliation

e.g., Boston University, Harvard University, etc.

Campus

Charles River Campus

School

College of Engineering

Department

Bioinformatics

Position Held at Institution

Graduate

Poster Submissions

Poster Title

Limitations of inferring transcription factor binding using ATAC-Seq and DNase-Seq

Authors and their Affiliation

George F. Steinhardt and David J. Waxman, Department of Biology and Bioinformatics Program, Boston University

Please describe the extent of your work in this research

All data collection, data analysis.

Abstract Submission

- GSI-symposium-2015.doc

Would you like your abstract to be considered for an oral presentation (students and post docs only)?

Yes

Boston University Genome Science Institute Symposium, November 2015

Limitations of inferring transcription factor binding using ATAC-Seq and DNase-Seq

George F. Steinhardt and David J. Waxman, Department of Biology and Bioinformatics Program, Boston University

Digital genomic footprinting uses genome-wide DNase-seq data to infer sites of transcription factor (TF) occupancy in intact chromatin, based on the ability of a bound TF to inhibit local DNA digestion by DNase-I [1]. Comparisons of differential DNase-Seq cutting profiles when anchored at a TF binding motif. Further, each TF is expected to show a unique DNase-I cutting profile owing to its characteristic pattern of TF-DNA interactions [1]. We investigated this proposal using ChIP-seq data for 22 distinct TFs are available [2]. We found that digital DNase-I cutting profiles generated at individual TF motifs have distinct cutting patterns, however, these patterns are largely unchanged when TF bound genomic loci are compared with unbound loci, as defined by overlap with the TF's ChIP-seq peak set. Similar analyses were carried out using ATAC-seq-based chromatin accessibility data, which relies on transposase insertion for DNA fragment release from open chromatin sites, and is suggested to be well suited for digital genomic footprint calling [3]. Digital genomic footprinting profiles generated from ATAC fragment release data were also similar for TF-bound and TF-unbound motifs for the same set of 22 TFs. Substantial differences were seen between the DNase-I and ATAC cutting profiles at each TF motif, suggesting each method has its own unique intrinsic sequence bias for DNA cutting. Further, comparison of the cutting profiles generated from purified genomic DNA to those of intact chromatin showed little evidence for chromatin-associated TF-DNA interactions with either DNase-I or ATAC. Analysis of the purified genomic DNA cutting patterns for DNase-I and ATAC cutting confirmed that each cutting method has a unique intrinsic sequence bias, with ATAC showing ~1,000-fold range of cutting frequencies, 4-fold greater than the range seen with DNase-I. Previous efforts have been made to call footprints at individual genomic loci using DNase-seq data and the algorithm PIQ [4]. Here, PIQ intact chromatin for the set of 22 TFs, using both DNase-seq and ATAC-seq datasets. Digital genomic footprints were correctly called for a subset of the 22 FPs by PIQ using both datasets, with ATAC generally outperforming DNase-I, however, the true positive rate was < 20% for 14 of the 22 TFs, as determined by comparison to ChIP-seq data. Studies in progress will investigate whether this performance is improved using the algorithm msCentipede[5], which use a background for footprint calling.

Supported in part by NIH grant DK33765 (to DJW).

[1] J. R. Hesselberth, X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, and J. A. Stamatoyannopoulos, "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting," vol. 6, no. 4, pp. 283–289, 2009.

[2] The EENCODE Project Consortium, "An Integrated Encyclopedia of DNA Elements in the Human Genome," vol. 489, no. 7414, pp. 57–74, 2013.

[3] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of

native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.," *Nat. Methods*, vol. 10, no. 12, pp. 1213–8, Dec. 2013.

[4] R. I. Sherwood, T. Hashimoto, C. W. O'Donnell, S. Lewis, A. a Barkal, J. P. van Hoff, V. Karun, T. Jaakkola, and D. K. Gifford, "Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape.," *Nat. Biotechnol.*, vol. 32, no. 2, pp. 171–8, Feb. 2014.

[5] A. Raj, H. Shim, Y. Gilad, J. K. Pritchard, and M. Stephens, "msCentipede: Modeling Heterogeneity across Genomic Sites and Replicates Improves Accuracy in the Inference of Transcription Factor Binding," *PLoS One*, vol. 10, no. 9, p. e0138030, Sep. 2015.