Developing a software suite to analyze the interplay between nucleosome arrangement, DNA methylation and transcription factor binding

Vladimir B. Teif, Daria A. Beshnova, Yevhen Vainshtein, Thomas Höfer, Karsten Rippe ✉

German Cancer Research Center, Heidelberg. Germany

Motivation and Objectives

Many of the underlying cell fate decisions occur via changes of chromatin features that affect gene expression. The specific location of nucleosomes on the DNA has important functions in controlling access to the DNA. Binding of protein factors to the 145-147 bp of DNA wrapped around the histone octamer core is frequently impeded, while the linker DNA between nucleosomes is more easily accessible. Recent advancements in high-throughput sequencing methods allowed genome-wide mapping of individual nucleosomes at single base pair resolution, with yeast serving as a model system for the initial pioneering studies. More recently, tissue- and disease-specific features of nucleosome positions in higher organisms were reported. These include for example studies of human and mouse cells. However, the large amount of data, as well as complex relation between different data types, poses a challenge to the understanding the underlying biology. Here we set to develop computational software capable to tackle this problem and applied it to the analyze the interplay between nucleosome arrangement, DNA methylation and transcription factor binding.

Methods

In a recent work (Teif *et al.*, 2012), we determined genome-wide nucleosome occupancies in mouse embryonic stem cells and their neural progenitor and embryonic fibroblast counterparts to assess features associated with nucleosome positioning during lineage commitment. Cell type and protein specific binding preferences of transcription factors to sites with either low (e.g. Myc, Klf4, Zfx) or high (e.g. Nanog, Oct4 and Sox2) nucleosome occupancy as well as complex patterns for CTCF were identified. Nucleosome depleted regions around transcription start and termination sites were broad and more pronounced for active genes, with distinct patterns for promoters classified according to CpG-content or histone methylation marks.

Throughout the genome nucleosome occupancy was correlated with certain histone methylation or acetylation modifications. In addition, the average nucleosome-repeat length increased during differentiation by 5-7 base pairs, with local variations for specific regions. Our results revealed regulatory mechanisms of cell differentiation that involve nucleosome repositioning. Here we have used this dataset to develop and test the computational software for the analysis of this type of data (Teif *et al.*, 2013; Teif and Rippe, 2012). To demonstrate its applicability, we combined these data with DNA methylation and hydroxymethylation data, as well as the maps of histone variants and DNA-modifying enzymes and studied the interplay of nucleosome positioning DNA methylation and TF binding.

Results and Discussion

We have developed a software suite for the biophysical analysis of high-throughput sequencing experiments, and applied it to the interplay between nucleosome positioning, DNA methylation and transcription factor binding during ES cell differentiation. DNA cytosine methylation (5mC) and hydroxymethylation (5hmC) are among the most important epigenetic marks. The interplay of 5mC/5hmC marks, the arrangement of nucleosomes and transcription factors (TFs) links DNA methylation with cellular gene expression programs but the underlying mechanisms are poorly understood. Here, we analyzed nucleosome positioning, DNA methylation and TF binding in conjunction with additional dinucleosome occupancy maps. Our study provides a novel quantitative description for the relations between DNA methylation/demethylation, TF binding and nucleosome occupancy changes.

Acknowledgements

References

Teif VB, Erdel F, Beshnova DA, Vainshtein Y, Mallm JP, Rippe K (2013) Taking into account nucleosomes for predicting gene expression.

Teif VB and Rippe K (2012) Calculating transcription factor binding maps for chromatin. *Brief Bioinform* **13**, 187-201.

Teif VB, Vainstein E, Marth K, Mallm J-P., Caudron-Herger M, Höfer T, Rippe K (2012). Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* **19**, 1185-1192.