The Dynamic Epigenome - Analysis of the Distribution of Histone Modifications

Dissertation, 2013

Lydia Steiner

There is a genome in a cell, as everyone knows, but there is also an epigenome. The epigenome regulates the transcription of the underlying genome. In the last decade, it was discovered that the epigenome state and its regulation are important for differentiation and development. Diseases were identified which are caused by errors in the epigenome state and its regulation. Therefore, several studies aiming at measuring epigenome states in different organisms and cell types and thus, provide huge amount of data.

In the dissertation, a pipeline had been developed to analyze and characterize histone modifications with respect to different cell types. Application of this pipeline had been shown for a published data set of mouse consisting of data for H3K4me3, H3K27me3, and H3K9me3 measured in embryonic stem cells (ESCs), embryonic fibroblasts (MEFs) and neuronal progenitors (NPCs).

It had been shown that each cell type and each modification exhibits its own characteristic length and CpG-distribution of the regions modified. While literature reported CpG-dependency had been proven by the analysis, it had been discovered that CpG-dependency is lost in NPCs. Additionally CpG-dependencies had been more detailed described by the average CpG-density of the modified regions in the length dependent manner.

Analysis of the overlaps of different modifications and different cell types had led to the observation that regions modified are covered by another modification data set or a data set from another cell type either completely or not at all. This had led to the suggestion that genomic regions are defined which are modified or not. Length distributions had been fitted to the data assuming cooperative unis. Modification reaction is performed with a cooperative component. Thus, similar behavior as observed had been predicted by such a model.

Methods for the detection of the epigenetic patterns are presented in this dissertation. Therefore, a segmentation method is developed to segment the genome guided by the data sets. Based on this segmentation, the epigenome states as well as epigenetic variation can be studied. Different visualization methods are developed to highlight the patterns in the segmentation data. Application of the segmentation AND visualization methods to the mouse data set had resulted in not only colorful squares but also in biological conclusions! It demonstrate the power of the developed methods. Analysis of the pattern had shown that many observations reported in literature are true but are not universal. For example, housekeeping genes are marked by H3K4me3 in ESCs but not MEFs and NPCs. Bivalent regions indeed turn into H3K4me3 marked and H3K27me3 marked but more often stay bivalent. Furthermore, it had been found that increasing stability of the patterns discovered correlates with the numbers of marks.

Finally, results obtained in this thesis had shown that there are several mechanisms to set the different marks. Some of them are coupled such that different modifications are set to the same genomic loci. However, the biological and biochemical mechanisms had remained unclear. Furthermore, mechanisms exist which set a single mark and are uncoupled from the mechanism of the other marks. relevancy and customer satisfaction of a given topic.