Analyzing Histone Codes by a Computational Approach

Yayoi Natsume-Kitatani\textsuperscript{1,2} \hspace{1cm} Motoki Shiga\textsuperscript{1,2}
natsume@kuicr.kyoto-u.ac.jp \hspace{1cm} shiga@kuicr.kyoto-u.ac.jp

Hiroshi Mamitsuka\textsuperscript{1,2}
mami@kuicr.kyoto-u.ac.jp

\textsuperscript{1} Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-0011, Japan.
\textsuperscript{2} Institute for Bioinformatics Research and Development of Japan Science and Technology Agency (JST-BIRD), Japan.

Keywords: histone code, histone acetylation, ChIP-chip, microarray

1 Introduction

Histone forms an octamer, which is essential for DNA packaging in nucleus. Another important aspect of histone is its N-terminal tail, which is subjected to posttranslational modifications such as methylation, phosphorylation and acetylation [1]. Recent experimental findings on histone modification lead to the so-called “histone code” (or modification patterns related with gene regulation), which can be key to elucidating epigenetic control of gene expression. The histone code has been hard to understand because of its complexity, while there has been growing demand to decipher the histone code mainly because it has attracted attention as a target for developing new therapeutics by utilizing the information of the histone code [2]. Under this background, we have developed a genome-wide computational method, which discovers patterns of histone acetylation, being tightly correlated with transcription factor (TF)-binding and gene expression.

2 Method and Results

2.1 Methods

Our method first generated two sets of ten gene clusters, one being in terms of histone acetylation and the other being in terms of TF-binding, from two ChIP-chip datasets in yeast (both in 2% glucose YPD) [3] [4], and then generated a matrix (which we call TF-HM) with 10 x 10 elements where genes in each element behave in a similar manner according to both TF-binding and histone acetylation. As a clustering method, we used spectral clustering [5], in which we estimate parameters of a mixture of von Mises Fisher distributions based on an EM (Expectation maximization) algorithm [6]. We further partitioned genes into five clusters by using microarray datasets (eg. GSE9217, glucose depletion condition in yeast). For each of the five clusters we checked the number of genes which are assigned to each element of TF-HM and, out of the total 500 (=5 x 10 x 10) elements, selected elements (which we call pattern-elements) which have the most characteristic patterns in TF-binding, histone acetylation and gene expression under some prespecified condition.
2.2 Results

From gene expression profiles in glucose depletion condition, we obtained five pattern-elements. Our result showed that four out of the five pattern-elements were dependent on H4 Lys16 acetylation, which implies that the subsequent deacetylation of H4 Lys16 plays an essential role in gene regulation. In addition, we obtained several unknown histone code patterns. Respiration-related genes, which were dependent of co-acetylation of all lysine residues in H4, were up-regulated regardless that H4 Lys16 was significantly (p<0.05) acetylated. Ribosome biogenesis-related genes, which were dependent of co-acetylation of all lysine residues in H3, H4 Lys12, H2A Lys7, were up-regulated independently of H4 Lys16. Furthermore, we detected particular patterns in the histone code from several microarray datasets.

3 Discussions

Under the glucose depletion condition, yeast cells switch the energy-supplying reaction from fermentation to respiration. This change results in elevation of cellular NAD⁺ concentration, which activates the class III histone deacetylase (HDAC) Sir2 and invokes gene inactivation by NAD⁺-dependent deacetylation of H4 Lys16 [7]. Overall the patterns we detected were well consistent with the literature. The histone acetylation pattern in respiration-related genes demonstrates that co-acetylation of all lysine residues in H4 might function to counteract gene inactivation caused by single acetylation of H4 Lys16. Our method is the first attempt to detect histone code patterns computationally. Another microarray dataset obtained under the same condition as the ChIP-chip datasets we employed would be useful to find other new histone code patterns. Similarly we may find other patterns by considering different modification, such as methylation and phosphorylation, by using a ChIP-chip dataset toward methylated or phosphorylated histones.

References