Research on Non-linear Relationship among Histone Modifications in Yeast Genome

Panfeng Chen [a], JiHua Feng [b,*], Zenhui Shan [c], Henhen Wei [d], Huan Hu [e]

School of Electrical and Information Technology, Yunnan University of Nationalities, Kunming 650500, China

[a]973831955@qq.com, [b]feng_jihua@126.com, [c]519579098@qq.com, [d]842576200@qq.com [e]312 324981@qq.com

Keywords: Histone modification; Pearson Correlation; MINE.

Abstract. In the process of we studied chromosome modifications, we found the previous studies only focused on linear relationship between chemical modifications. So we interfered the data sets of the Yeast histone Methylation modification and Acetylation which are measured by Pothook et al. We got 16 kinds of protein modification date sets in the whole genome. Then we processed the each of data sets by aligning, averaging and normalizing data around the TSS (about 1000bp from upstream or downstream).After we grouped the whole genome histone modification data sets according to the Yeast genome transcription level data : one modified groups promote transcription, and the other modified groups inhibit transcription or are independent of transcription. In order to reveal the complex relationships between different transcriptional activity of genetically modified, we calculated the correlation among several modifications with two by algorithms–Pearson correlation analysis and MINE proposed by David N. Reshelf et al. We analyzed the results and got some non-linear relationship between histone modifications. We also found that the association between histone modifications stronger in the same modified group.

Introduction

Histone is a major element of the nucleosome which is the basic structure of the chromosome. Nucleosome consists of four kinds of histones which are H2A, H2B, H3 and H4, each of two molecules to form a group of histone octamer, its outside coiled by the DNA molecules whose length is about 146 bp. Each Core Histone's N - terminal amino acid residues can react with the effect of acetylation, methylation, phosphorylation, ubiquitin and many kinds of covalent modification [3]. Histone modification plays a very important role in life process, and the occurrence of many diseases and Histone modifications have a close relation [4, 5].

Different histone modification plays a different role in the gene expression, including histone code, signal network and charge neutralization model and other different models which have been proposed to illustrate the function of histone modifications [6, 7, 8, 9], but a single histone modification cannot independently work [6], and this trait show that there is relevance among the Histone modification, and have an influence on the gene expression. Therefore, there is biological significance to research the correlation of the Histone modification. But previous research of chromatin modification only focused on the linear relation between the chemical modification, but nonlinear relationship was not fully focused. However, there are complex relationships between the modify, and often cannot simplified only with linear, so there are a more important meaning to research the nonlinear relationship of modify.

Data and Methods

The Resources of the data. We get these research data mainly from the Dmitry K. Pokholok et al who has experimented the data of Yeast histone methylation modification and the acetylation of histone [1]. And we also get the data from NCBI databases of yeast 16 chromosomes DNA coding

sequence, David and others given in literature of 4792 high confidence yeast genome experimental data [10].

Pre-processing Data. The Interpolation of Data. Dmitry K. Pokholok use experiment to measure Yeast Histone Methylation Modification and Acetylated Modification and get the data which cannot cover the whole genome, so we will have to conduct these data with the method of interpolation processing. This paper chooses the spline interpolation method to interpolation to the original data.

The Alignment of Data Processing. According to the different location of data which are used to defy the patterns of gene, we choose the range of 1000bp between the upstream and downstream around TSS of histone modifications data, using the means of aligning (And the data corresponding to C type gene will go into reverse processing), then solve those data with stacking, averaging, smoothing, and normalizing, and thus we got 16 kinds of the average spectrum of histone modification which are around all the genes and different gene expression patterns of TSS.
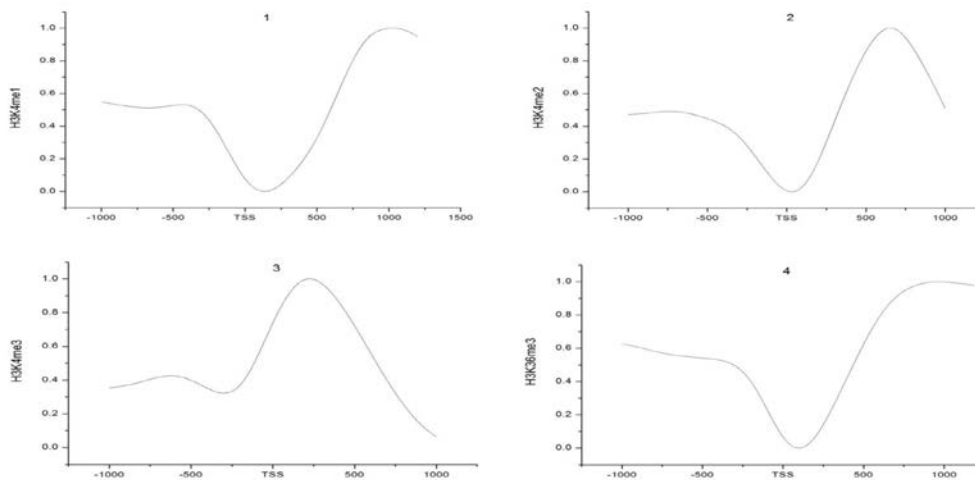

Fig.1 The picture of Histone modification after interpolation and align

The Classification of Histone Modification. Through interpolation, we get the data of whole genome which was modified by the histone, according to different levels of yeast gene transcription, we divided those levels into four grades, and from high to low is that the mRNA/h is equal or greater than50 mRNA/h; mRNA/h is greater than or equal to 20 less than 50; mRNA/h is greater than 1 and less than 20; mRNA/h is less than or equal to 1. In this way, we can according to the level of transcription of each gene be classified, so that all the genes will be divided into four categories. After that, we class for all of the genes in its TSS peripheral selection, the length of 1000bp downstream, to perform the above alignment treatment (in this process normalization is not required), to thereby obtain a different level of transcription TSS around histone modification data. So we can get in different transcription conditions of 16 kinds of histone modifications on downstream 1000bp modified TSS at around the level. Below are some histone modifications Figure aligned up under different transcriptional level later.
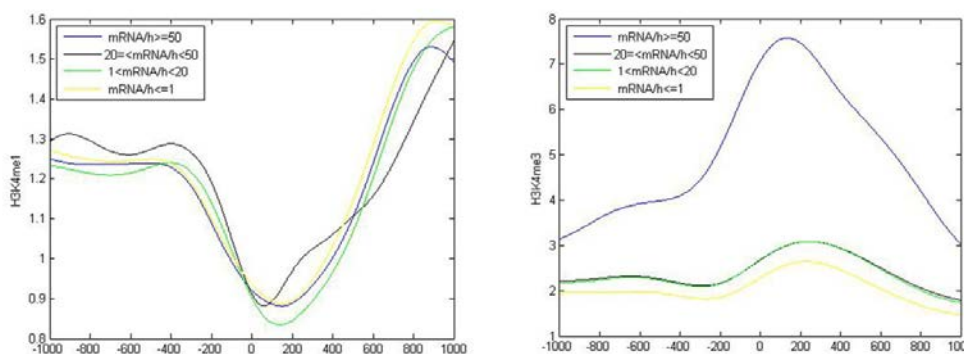
We observed at different levels of transcription of each histone modification 1000bp downstream around the modified diagram TSS and found: in different transcription conditions, the level of histone modification and somewhat increased with the increasing of the level of transcription, and some did

not substantially change, some with slightly reduced transcript levels increased. We take all of the histone modifications divided into two types: one is a transcriptional enhancer modification groups(H3K9ac,H3K14ac,H4ac,H3K4me3,H3K36me3,H3K14acvsWCE.YPD,ESA1.YPD,GCN5. YPD,GCN4.AA), another class is a transcriptional repressor of transcription modification or unrelated groups (H3K4me1, H3K4me2, H3K79me3, H3.YPD, H4.YPD, IgG. YPD, noAB. YPD).

Pearson Co-relational Analysis Method. Pearson linear correlation coefficient can indicate the degree of correlation between two variables, which is calculated as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

Where n is the sample size , $\bar{x}$ is the average value of the variable x , $\bar{y}$ is the average value of the variable y , r value for the variable y in [-1 , + 1] varies between , if r> 0, then x is related to y, if r < 0 , then x and y negative correlation , the greater the absolute value of r , the linear correlation are greater .

The Arithmetic of MINE. The Brief Introduction of the Arithmetic of MINE. MINE algorithm is the degree of correlation between the two variables algorithm David N. Reshelf [1], who proposed a calculation, is the Maximal Information-based Nonparametric Exploration abbreviations. Has its limitations because of the correlation between the two variables using the Pearson correlation coefficient is a measure , that is, the method can find the degree of linear correlation between two variables , while for the non-linear dependence of the case, the method cannot accurately determine their degree of correlation . The core idea of the MINE algorithm is to calculate the MIC (the Maximal Information Coefficient) between two variables values, basic idea is that if there is an association between two variables, then you can draw on a network on both the scatterplot trellis, and the trellis can be divided scatter and parcel to the grid. Thus, in order to calculate the value of a variable data on MIC, we will explore a combination of grid map of all ranks (from 2 to multiply by 2 column values based on the maximum number of samples of data rows), and then calculates the ranks can be achieved in each portfolio the maximum possible mutual information (mutual information). (Note: So, here row by row should get the maximum value of a mutual information) Then, we will normalize the mutual information value, make it into a value between 0-1 so you can compare. We define a characteristic matrix M, so that M= {m_{xy}}, which is the maximum value for each row of interaction information obtained at the previous combination, and the MIC value is the maximum value of the mutual information values.

The research of relativity

After we group histone modifications, we use the data of histone modifications around the TSS data which is got by pre-processing, we calculate the Pearson Correlation Coefficient between them.
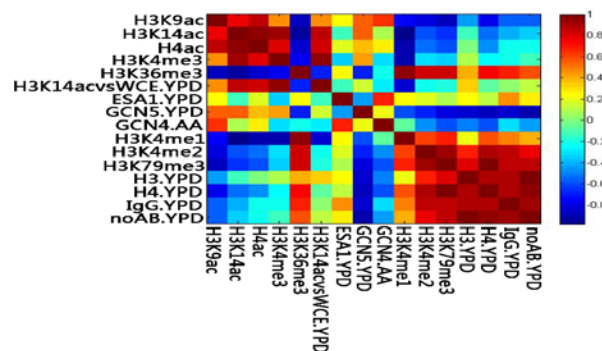

Fig.3 Pearson Correlation between Histone modifications

From Figure (3) we can find: except individual histone modifications (eg H3K36me3), most of Histone modification showing a strong positive correlation between the same group , or little

correlation between different group, or showing a strong negative correlation . This shows the same group of histone modification make same function in transcription.

Pearson coefficient can only reflect correlation of liner, but there is not a simple linear relationship between histone modification, the relationship between them are complex, only use Pearson coefficients to reflect the relationship between them there may be error. So we took advantage of MINE algorithm to calculate the values of MIC between them.
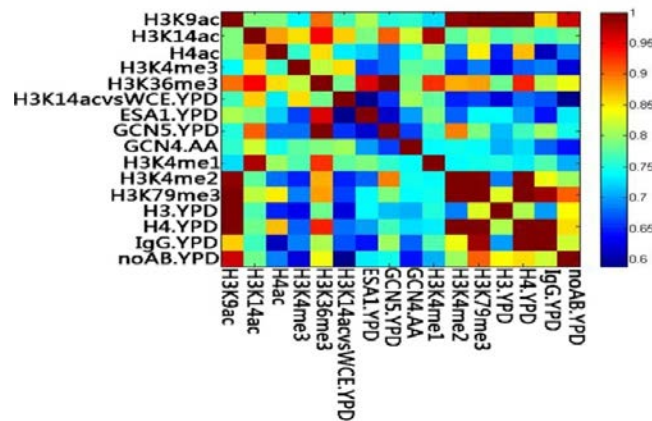

Fig.4 MIC between Histone modifications

Compare fig (3) and fig(4) , we can find : when Pearson coefficient between histone modifications are big ,their values of MIC values also large , indicating that there is a linear relationship between these histone modifications ; and when Pearson coefficient between histone modifications are small, some MIC value was large, indicating that there are some non-linear relationship between these histone modifications . As H3K9ac and H3.YPD, Pearson coefficient between them is about -0.3, and their MIC values are close to 1. To study the relationship between histone modifications which value of Pearson coefficient is small but has a great value of MIC , we find them, and then draw a two-dimensional scatter plot between them and draw a three-dimensional map which plus genetic locus.
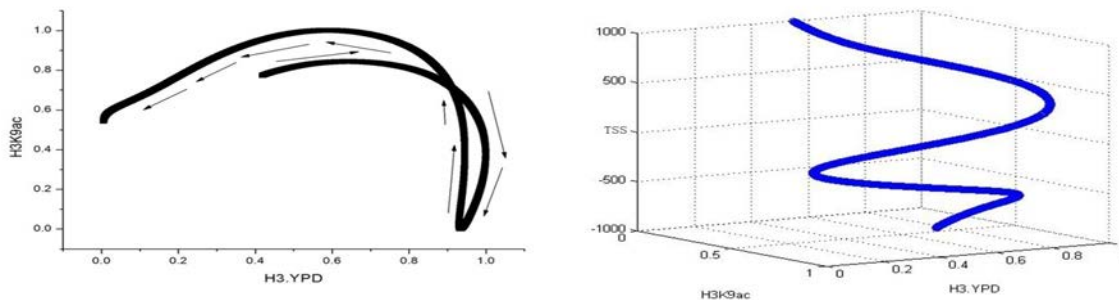

Fig.5 the 2D scatter plot and 3D scatter plot contains the gene of Histone modification

From fig (5), we can found that the relationship between H3K9ac and H3.YPD is similar to parabola, According to the three-dimensional map, we can marked trend of genetic locus to the two-dimensional scatter plot and find the relationship between H3K9ac and H3.YPD is similar to parabola.

Further, from fig (4), we found the values of MIC are larger in same group in the whole, indicating that the histone modifications of same group have a stronger correlation.

Conclusion

This paper mainly studies the correlation between histone modifications around the TSS. After we grouped the whole genome histone modification data sets according to the Yeast genome transcription level data: one modified groups promote transcription, and the other modified groups inhibit transcription or are independent of transcription. Then calculate the Pearson coefficient between histone modifications by algorithms–Pearson correlation analysis. Because of the

limitations of algorithms–Pearson correlation analysis, we also calculated values of MIC between histone modifications by MINE. Compared the values of Pearson coefficient and MIC,

When the values of Pearson coefficient and MIC are both large, indicating that there is a linear relationship between them, and those who Pearson coefficient is small and the value of MIC large, indicating that there is a nonlinear relationship between them. Finally, we find the nonlinear relationship between histone modifications by constructing a three-dimensional map. Meanwhile, we find: from the Pearson coefficient, the same group present strong positive correlation, either the correlation between different groups is very small, or present a strong negative correlation. From the value of MIC, the values of MIC are greater in the same group. So we believe that: a causal relationship may exist between the same groups or the histone modifications in the same groups play the same role in gene activity. Johnson found through research, constructed the different expression level of yeast strain which from mutation forms of Histone H3, and knock out acetylation enzyme Gcn5 and Sas3 of H3K14,will lead to the modification level decrease, this display H3K14ac is necessary for H3K4me3 [11]. Relevant research results show: H3K4me3 and H3K9ac are related to the activation of gene expression in promoter region [1]; H3K4me1 and H3K4me2 are inhibit gene expression [12]. These support our view.

Acknowledgments

References

[1] Dmitry K. Pokholok, Christopher T, et al. Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. Cell, Vol. 122, 517–527, August 26, 2005.

[2] David N. Reshef, Yakir A. Reshef, et al. Detecting Novel Associations in Large Data Sets. Science 334, 1518 (2011).

[3] Strahl B D,Allis C D, et al. The language of covalent histone modification[J]. Nature, 2000,403(6765) p.41-45

[4] Kondo Y,Shen L, et al.Critical role of histone methylation in tumor suppressor gene Silencing in colorectal cancer[J]. Mol Cell Biol,2003,23(1) p.206-215.

[5] Fraga M F,Ballestar E, et al.Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a commom hallmark of human cancer[J],Nat Genet,2005,37(4) p.391-400.

[6] Strahl BD,Allis CD. The language of covalent histone modifications. Nature,2000, 403, p.4145

[7] Turner BM. Histone acetylation and an epigenetic code. Bioessays, 2000, 22 p.836-845

[8] Wade PA, Pruss D,Wolffe AP. Histone acetylation;chromatin in action. Sci,1997, 22, p.128-132

[9] Schreiber SL, Bernstein BE. Signaling network model of chromatin. Cell, 2002, 111,p.771-778

[10] W. Lee, D. Tillo, N. Bray, R.H. Morse, R.W. Davis, T.R. Hughes, C. Nislow, A high-resolution atlas of nucleosome occupancy in yeast, Nat Genet 39 (2007) p.1235-1244.

[11] Johnson I. Investigating histone methylation in yeast: regulation of H3K4me3, and the role of the methyl-histone binding domains of Iswlb. Masters Degree Thesis, University of Calgary, Canada, 2010,p. 26-30.