2011 Autumn Progress Report

Genome-wide analysis of histone modification patterns in human

Ryu Ogawa[1,2] and Masaru Tomita[1,2]

[1]Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan

[2]Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-8520, Japan

Abstract

Histones are essential for chromatin conformation and transcriptional regulation in eukaryotic cells. N-terminal tails of histone proteins are targeted to many types of modifications such as methylation, acetylation, phosphorylation and ubiquitination. Recent studies have provided genome-wide histone modification data in human by using ChIP-seq method. However, functions of histone modification patterns are largely unknown. In this study, we developed a novel method to detect modification patterns. To detect bivalent histone modification, we used relative O/E value that donated the difference between O/E values of bivalent modifications in target region and those in the whole genome region. We detected 5 bivalent histone modification candidates. Network analysis showed candidates were important node of modification network. This results support our candidate of bivalent histone modification pairs.

Keywords

Epigenetics, Histone modification, Computational analysis, Genome-wide analysis, Bivalent chromatin

1. Introduction

Histones are essential for chromatin conformation and transcriptional regulation in eukaryotic cells. Around a histone octamer in 1.65 turns, the nucleosome is consists of 147 bp of DNA wound. The histone octamer makes up two copies each of four histone proteins that are called H2A, H2B, H3, and H4 (Kornberg 1974; Drew and Travers 1985; Luger *et al*., 1997; Davey *et al*., 2002). Such as methylation, acetylation, ubiquitination and phoporylation, N-terminal tails of histone proteins are targeted to a lot of types of modifications (Fig. 1-a.).

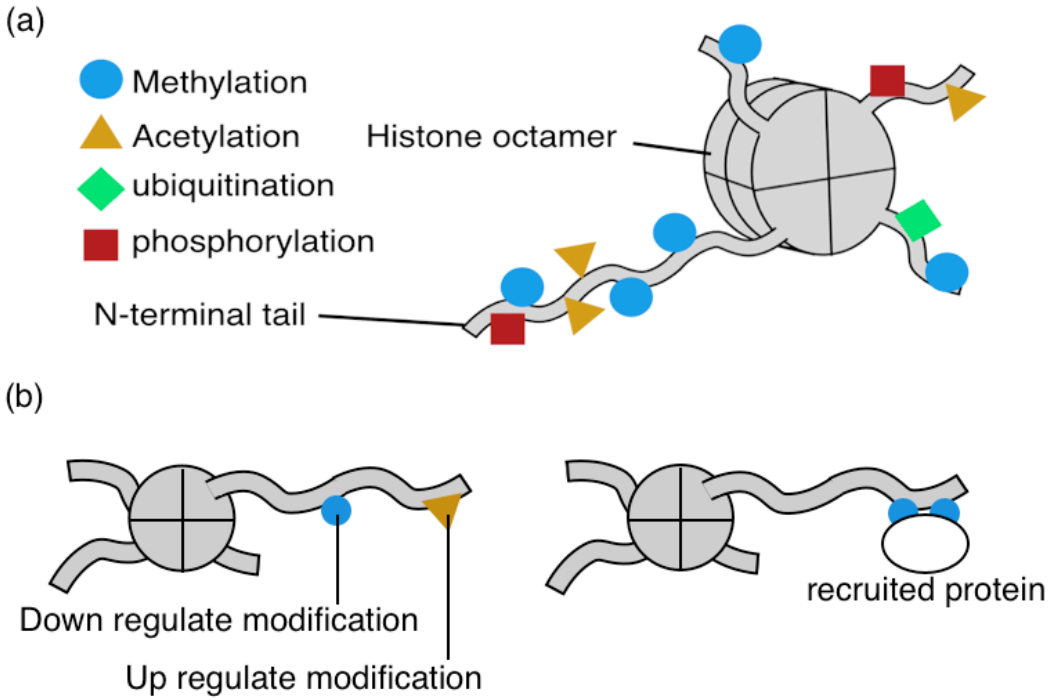

Fig. 1. A structure of histone octamer and N-terminal tails. (a) A structure of modified histone. Histone octamer has 8 N-terminaltail, and each histone tail can be target of many modifications. There are huge varieties of histone modification patterns. (b) A structure of bivalent modification. Functions of histone modification patterns are largely unknown

In eukaryotic cells, histone modifications has relationship between regulating transcription and the chromatin structure. It suggested that importance of histone modification patterns were “histone code hypothesis” 10 years ago (Strahi *et al,* 2000). It made sure of function of each histone modification in recent studies (Supplementary table.). Methylated H3K4, H3K36 and H3K79 are famous of characteristics of euchromatin. Also methylated H3K9, H3K27, H4K20 characterize heterchromatin (Li *et al.*, 2007). Furthermore, histone modification levels predict gene regulation. H3K4me3 and H4K20me1

in the promoter are known to characteristics of transactive gene. Also, H3K79me3 and H4K20me1 in the gene body characterize active transcribe genes (Karlic *et al*., 2010). The combination of histone modification is getting a lot more attention lately. To name a few, it is the combination of the histone modifications between different two functions and the protein that cognizes some histone modifications (fig1-b). Functions of histone modification patterns are still unknown.

In Recent studies, using chromatin immunoprecipitation (ChIP) assay cupled with next-generation sequencing (ChIP-seq), to identify genome-wide nucleosome positioning have provided large-scale nucleosomal DNA sequence data. It detected 20 histone methylations and 1 histone vatiant to use ChIP-seq in human CD4+ T cells (Barski *et al*., 2007). These data are genome-wide data and one of the largest histone modification data.

2. Materials and methods

2.1. Materials

2.1.1. Genome-wide histone modification data.

We used the ChIP-seq data human CD4 T-cell(Barski *et al*. 2007). This data was the experimental nucleosome positioning and histone modification data. This data consisted of 20 histone modifications and 1 variant.

2.1.2. Splicing dataset of human.

The RNA splicing data was downloaded form H-DBAS(Takada *et al*. 2010). H-DBAS was a database of alternative splicing based on H-Invitational full-length cDNA. This database was updated at 2010, so the dataset is not old. And also this dataset had high accuracy because it is based on full-length cDNA.

2.2. Overview of detecting functional histone modification patterns.

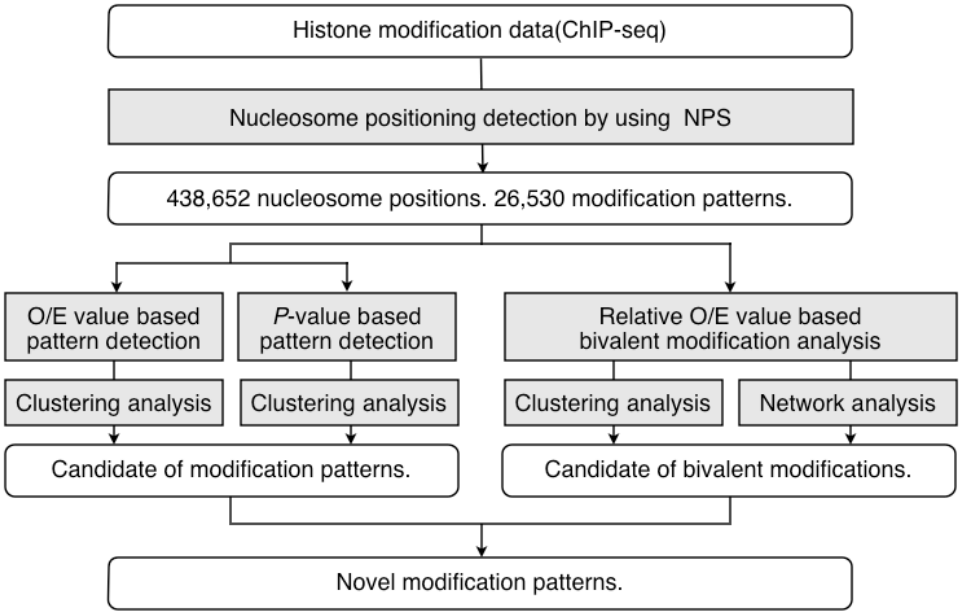

Fig. 2. Flowchart of modification pattern detection. This figure shows our modification pattern detecting methods. We developed modification pattern detecting method and bivalent modification detecting method.

By using NPS (Nucleosome Positioning from Sequencing), we extracted positions of modified histones from ChIP-seq data. NPS is famous software to detect nucleosome positioning (Zhang *et al*., 2008). We detected 438,652 nucleosome positions and 26,530 modification patterns. We developed modification pattern detecting method and bivalent modification detecting method.

We used two algorithms to detect modification patterns. Those were O/E value based method and P-value based method. We selected top 100 modification patterns that had high value of each method. And we used furthest neighbor method to clustering our candidates.

To detect bivalent histone modification, we used relative O/E value that donated the difference between O/E values of bivalent modifications in target region (e.g.: exon region, intron region) and those in the whole genome region. We analyzed those result using clustering method and network method to detect bivalent modifications.

2.3. Histone modification pattern detection

2.3.1. Vector conversion of histone modification patterns

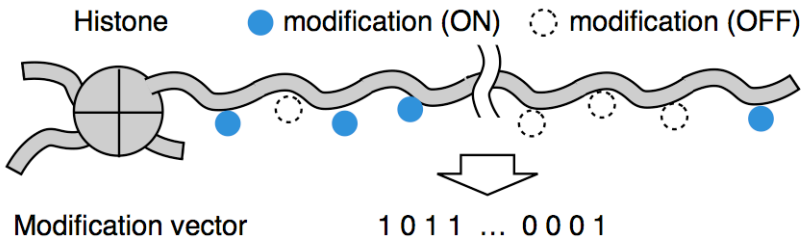

Fig. 3. Vector conversion of histone modification pattern. This figure shows schematic diagram of vector conversion of histone modification. Gray object is histone, filled blue circle indicates modified site and dashed line circle indicates no-modification site.

To analyze histone modification pattern, we conversed histone modification information to vector data. We integrated NPS outputs and created 21-dimentional vectors by translating on-midificaiton information to 1 and off-modificaiton information to 0. We detected 438,652 histone modification vectors, and there were 26,530 modification patterns contained in those vectors.

2.3.2. O/E value based pattern detection and *P*-value based pattern detection.

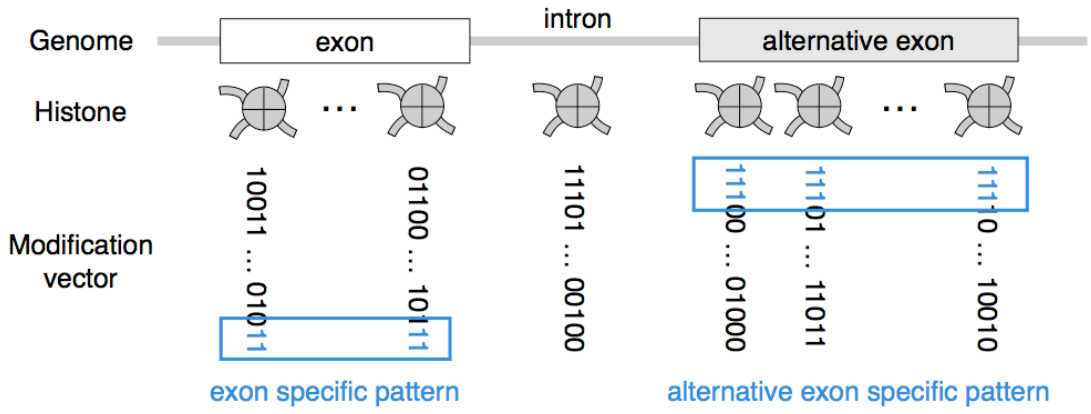

Fig. 4. Detection of region specific histone modification pattern. This figure shows schematic diagram of detection of region specific histone modification pattern. Detecting region specific patterns, we applied clustering analysis to O/E value method and *P*-value method.

To detect the functional modification patterns, we calculated O/E (Observation / Expectation) value $f(x,h)$ of modification pattern x and Hamming distance h using the equation (1) and (2):

$$f(x,h) = \frac{\sum_{i \in H(x,h)} O(i)}{\sum_{i \in H(x,h)} E(i)} \qquad (1)$$

$$E(i) = 438652 \prod_{j \in M(i)} pj \qquad (2)$$

where $O(i)$ is the number of observed modification pattern i, $E(i)$ is the expected value of modification pattern i. $H(x,h)$ is modification combinations that of Hamming distance from x is h. $E(x)$ can be defined as: where pi is probability of occurrence of each histone modificaiton i. $M(i)$ is each modification that is part of modification pattern i. 438652 is the total number of detected nucleosomes. P-value of modification pattern x can be calculate by using binormal distribution, expectation function $E(i)$ and observation function $O(i)$.

After calculating O/E value and P-value, we selected top 100 modification patterns that had high value of each method. And we used clustering method to selected modification patterns.

2.4. Relative O/E value based detection of bivalent histone modification pair.

Relative O/E value was calculated the difference between O/E values of bivalent modifications in target region and O/E values in the whole genome. If relative O/E value of modification pair A-B was high, A-B is modified same histone in a high probability in target region. We analyzed modification pairs that were scored by relative O/E value using clustering method and network method to detect bivalent modifications.

3. Results and discussions

3.1. Histone modification pattern detection.

3.1.1. Detection of modified nucleosome positionings.

We found 438,652 nucleosome positionings that made up 26,530 histone modification patterns. Modifications with the rate of high were transactivation modifications (H3K4me1, H3K4me1, H4K20me1). The rate of nucleosomes without modifications was only 0.03%. This result suggests that most of nucleosomes are modified. And sum of modification occurrence rate with more than 400% means that multiple histone modification is major event.

3.1.2. Clustering of functional modification pattern candidates.

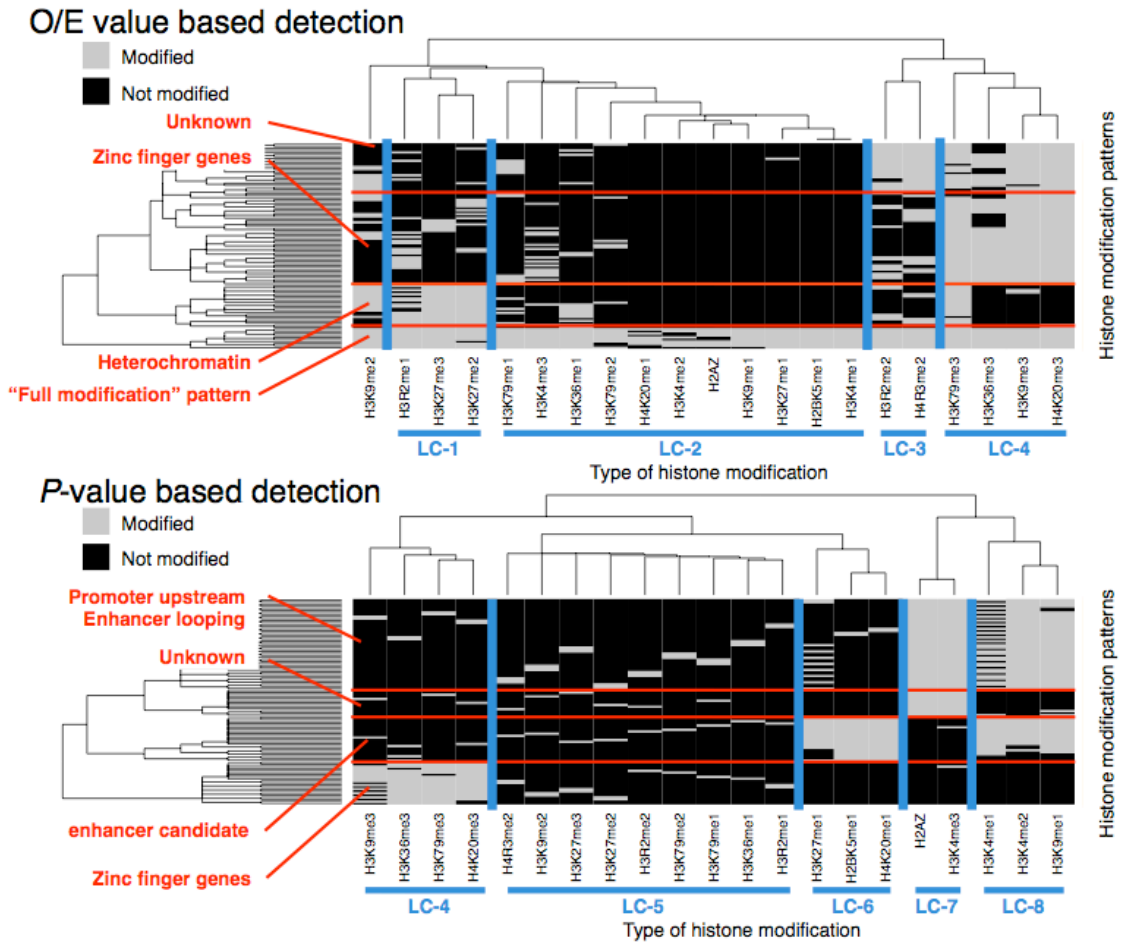

Fig. 5. Clustering of functional modification pattern candidates. Y-axis indicates each modification patterns, and X-axis indicates the name of modification. Gray color means "modified" and black color means "not modified". Red line represents the border of modification pattern and blue line indicates the border of local combination. LC means local combination. LC-4 was detected by O/E value based method and P-value based method.

Our method could detect 6 known functional modification patterns and 2 novel modification patterns. And we could detect 8 local combinations(LC). LC-4 is commonly detected by O/E value method and *P*-value method. LC-4 consists of H3K9me3, H3K36me3, H3K79me3, H4K20me3. These results suggested our method is effective to detect functional modification patterns. Our future work is finding more function of those candidates and to investigate relationships between the function and the modification patterns.

3.1.3. Histone modification pattern detection from alternative exon region.

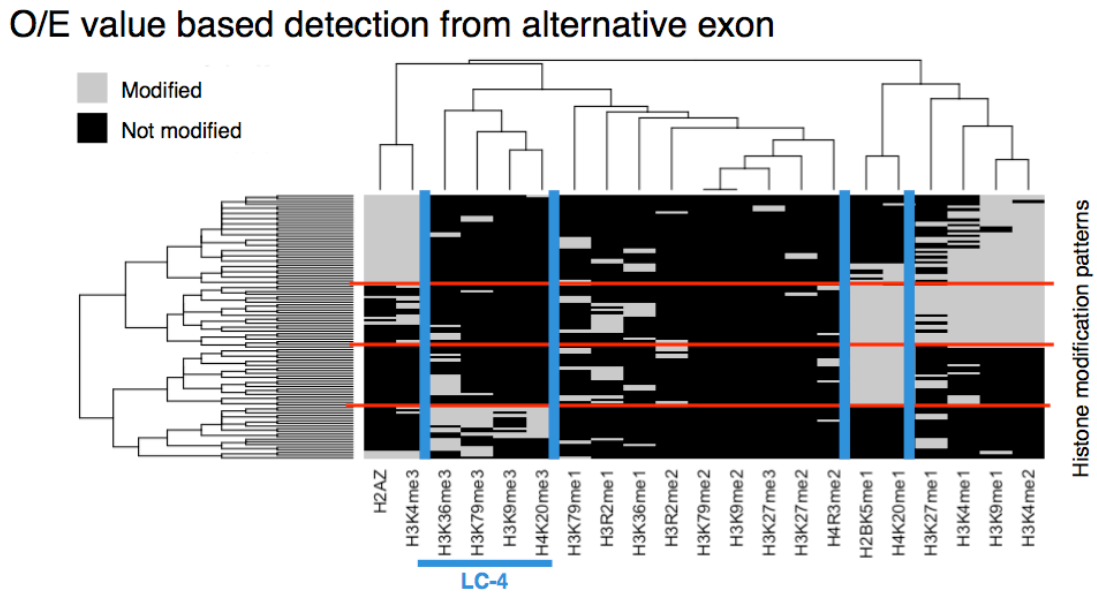

Fig. 6. Clustering of functional modification pattern candidates. Y-axis indicates each modification patterns, and X-axis indicates the name of modification. Gray color means "modified" and black color means "not modified". Red line represents the border of modification pattern and blue line indicates the border of local combination. LC-4 was detected in alternative exon region.

3.3. Detecting bivalent histone modification pairs from alternative exon region

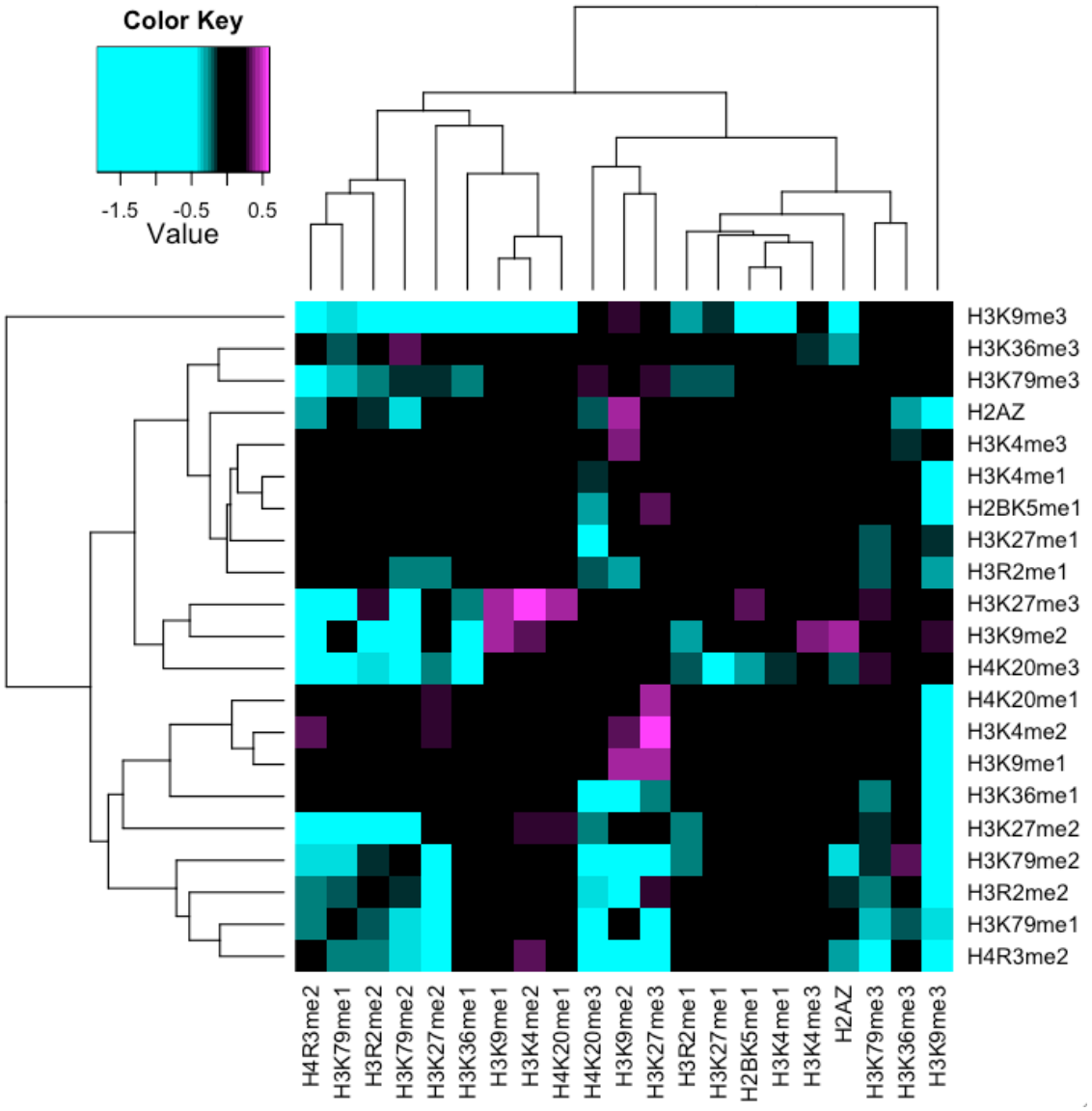

Fig. 7. Clustering of modification pair in alternative exon region. X-axis and Y-axis indicate name of histone modification. Color key is relative O/E value of bivalent histone modification in alternative exon region.

We detected candidate of bivalent modification pairs. H3K27me3–H3K9me1 pair, H3K27me3–H3K4me2 pair and H3K27me3–H3K20me1 pair were high probability. H3K9me1, H3K4me2 and H4K20me1 were contained in same cluster. H3K27me3 and H3K9me1 are known as gene repressing modification, and H3K4me2 and H4K20me1 are known as gene activating modification. H2AZ-H3K9me2 pair and H3K4me3-H3K9me2 pair had also high value. H3K4me3 is activating modification, and H3K9me2 and H2AZ are repressing modification.

3.4. Bivalent modification network.

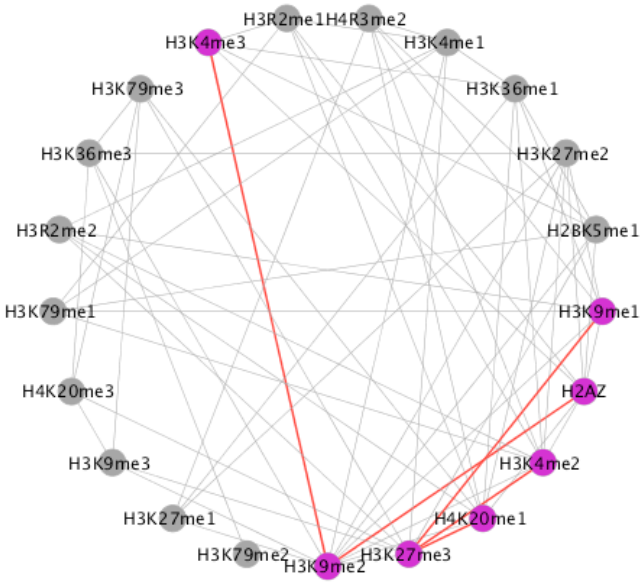

Fig. 8. Histone modification network in alternative exon region. A node means a histone and an edge shows a histone modification pair. We drew an edge if relative O/E value is above average. We arranged a node that has large degree in clockwise direction from H3K9me2. We showed a modification pair that was gained by clustering analysis as a red edge, and the modification that is consists of a pair as a pink edge.

We used the cytoscape to draw a modification network. This network is degree sorted circle layout. Those modifications had high degree. Those were important node of modification network. This results support our candidate of bivalent hisitone modification pairs.

4. Acknowledgements

This research was supported by research funds from the Yamagata prefectural government and Tsuruoka City, Japan.

References

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**(4), 823-837.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., Sauer, F. (2002) Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. *Nature*, **419**(6909), 857–862.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., Allis, C.D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev*, **15**(24), 3286–3295.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**(5595), 1039–1043.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W. and Richmond, T.J. (2002) Solvent-mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol*., **319**, 1097–1113.

Drew, H.R. and Travers, A.A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.*, **186**, 773–790.

Huyen, Y., Zgheib, O., Ditullio, R.A., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., Halazonetis, T.D. (2004) Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature*, **432**(7015), 406–411.

Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K., Vingron, M. (2010) Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A.*, **107**(7), 2926-2931.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868–871.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., Greenblatt, J. (2003) Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.*, **23**(12), 4207–4218.

Li, B., Carey, M., Workman, J.L. (2007) The role of chromatin during transcription. *Cell*, **128**(4), 707-719.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H.,

Thomson, J.A., Ren, B., Ecker, J.R. (2009) Human DNA methylomes at base resolution show widespread epigenetic differences. *Nature*, **462**(7271), 296-297.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, **389**, 251–260.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., Lis, J.T., Allis, C.D., Reinberg, D. (2002) PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell.*, **9**(6), 1201–13.

Sanders, S.L., Portoso, M., Mata, J., Bähler, J., Allshire, R.C., Kouzarides, T. (2004) Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell,* **119**(5), 603–614.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., Jenuwein, T. (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.*, **18**(11), 1251–1262.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., Rauscher, F.J. (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev*., **16**(8), 919–932.

Sedkov, Y., Cho, E., Petruk, S., Cherbas, L., Smith, S.T., Jones, R.S., Cherbas, P., Canaani, E., Jaynes, J.B., Mazo, A. (2003) Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. *Nature,* **426**(6962), 78–83.

Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**(6765), 41-45.

Tachibana, M., Sugimoto, K., Fukushima, T., Shinkai, Y. (2001) Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.,* **276**(27), 25309–25317.

Takeda, J., Suzuki, Y., Sakate, R., Sato, Y., Gojobori, T., Imanishi, T., Sugano, S. (2010) H-DBAS: human-transcriptome database for alternative splicing: update 2010. *Nucleic Acids Research,* **38** (Database Issue), D86-D90.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., Zhang, Y. (2001) Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol. Cell*, **8**(6), 1207–1217.

Zhang, Y., Shin, H., Song, J.S., Lei, Y., Liu, X.S. (2008) Identifying Positioned Nucleosomes with Epigenetic Marks in Human from ChIP-Seq. *BMC Genomics*, **9**, 537.

Supplementary materials

Supplementary Table. 1. Function and enzymes of histone modifications.

modification	enzyme	function	refarence
H3K4me1	MLL, ALL-1	activation	(Sedkov *et al*., 2003)
H3K4me2	Set1	permissive euchromatin	(Briggs *et al.,* 2001)
H3K4me3	Set7/9	activation	(Wang *et al*., 2001)
H3K9me1	G9a	repression (genome inprinting)	(Tachibana *et al*., 2001)
H3K9me3	SETDB1	repression	(Schultz *et al*., 2002)
H3K27me1	Ezh2	silencing	(Cao *et al*., 2002)
H3K36me1	set2	activation (elongation)	(Krogan *et al.,* 2003)
H3K79me1	Dot1	check point response, activation	(Huyen *et al*., 2004)
H4K20me1	Set9	check point response, activation	(Sanders *et al.,* 2004)
H4K20me3	Suv4-20h	heterochromatin	(Schotta *et al*., 2004)