Genome-wide analysis of nucleosome regulation by histone modification.

Ryu Ogawa

Abstract

In eukaryotic cells, histones are essential for chromatin conformation and transcriptional regulation. The histone octamer is composed of two copies each of four histone proteins (H2A, H2B, H3, and H4). N-terminal tails of histone proteins are targeted to many types of modifications such as methylation and acetylation. Histone modifications are implicated in regulating the chromatin structure and transcription in eukaryotic cells. Recent studies have provided genome-wide histone modification data in human by using ChIP-seq method. Here we developed a novel method to detect histone modification patterns from ChIP-seq data. We used O/E (Observation / Expectation) value based method and P-value based method. We selected top 100 modification patterns that had high value of each method. And we used clustering method to selected modification patterns. Our methods detected candidates of functional histone modification patterns. And our result showed a tendency that DNA of “Full modified” nucleosome are hyper methylated.

Keywords

Epigenetics, Histone modification, DNA methylation, Computational analysis, Genome-wide analysis

1. Introduction

In eukaryotic cells, histones are essential for chromatin conformation and transcriptional regulation. The nucleosome consists of 147 bp of DNA wound around a histone octamer in 1.65 turns. The histone octamer is composed of 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). N-terminal tails of histone proteins are targeted to many types of modifications such as methylation, acetylation, ubiquitination and phoporylation (Fig. 1.).

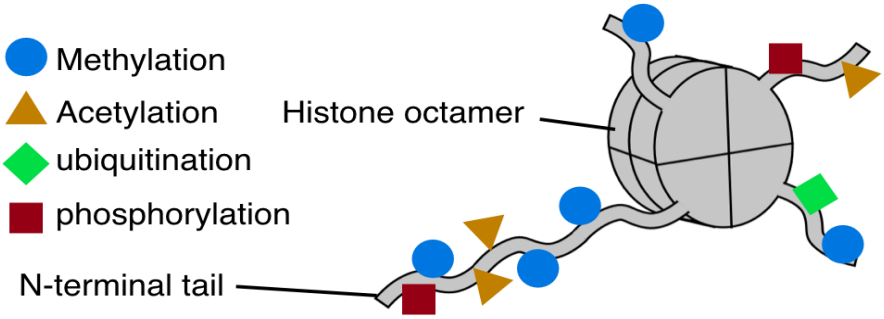

Fig. 1. A structure of histone octamer and N-terminal tails. Histone octamer has 8 N-terminaltail, and each histone tail can be target of many modifications. There is a huge variety of histone modification patterns.

Histone modifications are implicated in regulating the chromatin structure and transcription in eukaryotic cells. Importance of histone modification patterns were suggested by "histone code hypothesis" 10 years ago (Strahi *et al,* 2000). Recent studies ascertain function of each histone modification (Table. 1.). Methylated H3K4, H3K36 and H3K79 is known to characteristics of euchromatin. Heterchromatin is also characterized by methylated H3K9, H3K27, H4K20 (Li *et al.*, 2007). And histone modification levels are predictive for gene reguration. H3K4me3 and H4K20me1 in the promoter is characteristics of transactive gene. H3K79me3 and H4K20me1 in the gene body is also mark of active transcribe genes (Karlic *et al*., 2010).

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

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)

Histone modification data is accumulated, but functions of modification patterns are still unknown. Here, we analyze the function that histone modifications have combination pattern and the position that histone modifications locateon genome by using histone modifications data taken by Human T cell CD4+. We developed two method to detect functional hisutone modification patterns. DNA of “full modified” pattern nucleosome was highly methylated. This study will make a contribution to understand a role of histone modification patterns.

2. Materials and methods

2.1. Genome-wide histone modification data and DNA methylation data.

We used the ChIP-seq data of Barski *et al*. (2007) for the experimental nucleosome positioning and histone modification data of human CD4 T-cell. This data consists of 20 histone modifications and 1 variant. And we also used methylC-seq data of Lister *et al*. (2009) for the experimental DNA methylation data of human h1 Embryo Stem cell.

2.2. Histone modification pattern detection from genome-wide area.

2.2.1. Overview of detecting functional histone modification patterns.

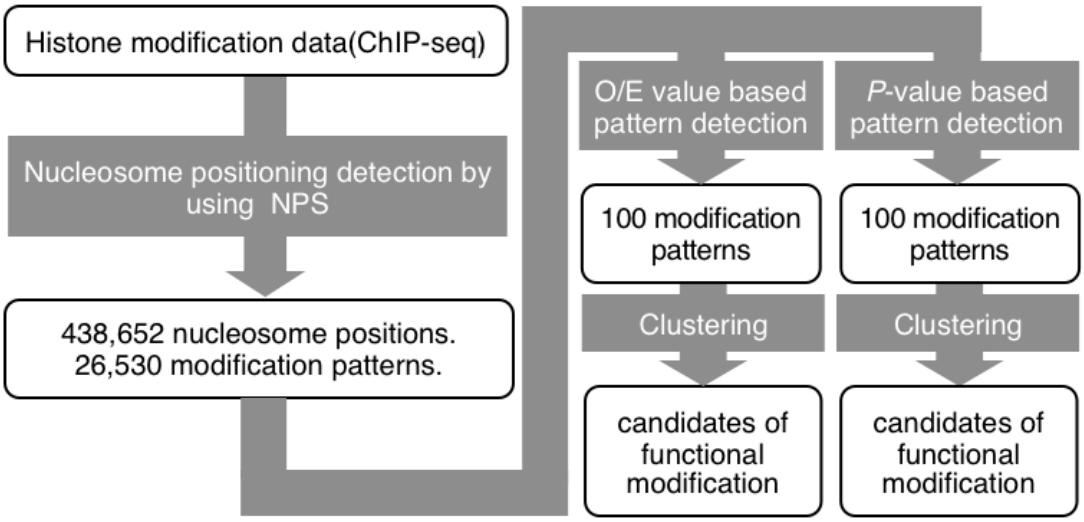

Fig. 2. Flowchart of modification pattern detection. This figure shows our modification pattern detecting methods. We developed O/E value based method and *P*-value based method.

We extract nucleosome positioning and its histone modification from ChIP-seq data by using NPS (Nucleosome Positioning from Sequencing) that is famous software to detect nucleosome positioning (Zhang *et al*., 2008). 438,652 nucleosome positions and 26,530 modification patterns were extracted by this software. And we used two algorithms to detect modification patterns. One is O/E value based method, and the other is P-value based method. We selected top 100 modification patterns which had high value of each method. And we used furthest neighbor method to clustering our candidates.

2.2.2. Hamming distance of modification.

Histone modification pattern was detected as 21 bit binary information. To consider fluctuations of histone modification pattern, we calculated the Hamming distance. The Hamming distance is the number of positions at that the corresponding bit were different. If Hamming distance between two modification patterns was lower than 1, we treated those modification patterns as same thing in this study.

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

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 $O(i)$.

2.3. Histone modification pattern detection from exon and intron.

2.3.1. Splicing dataset of human.

We used splicing data of H-DBAS. H-DBAS is a database of alternative splicing based on H-Invitational full-length cDNA (Takada *et al*. 2010). 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.

2.3.2. Comparison of modification patterns between exon and intron.

To detect characteristics of genome structure specific modification patterns, we calculated Kullback-Leibler information in each modification. Kullback-Leibler information indicates a quantity of probability distribution bias.

3. Results and discussions

3.1. Histone modification pattern detection from genome-wide area.

3.1.1. Detection of modified nucleosome positionings.

We detected 438,652 nucleosome positionings that consisted 26,530 histone modification patterns (Table. 2., Fig. 2.). Modifications that has high rate ware transactivation modifications (H3K4me1, H3K4me1, H4K20me1). The rate of nucleosomes that gain no modifications was only 0.03%. This result shows most of nucleosomes are modified. And sum of modification occurrence rate is more than 400%, so multiple histone modification is not minor event.

Table. 2. Detail of detected nucleosome positions.

modification	number of Positions	Rate (%)	modification	number of positions	rate (%)
H3K4me1	254,748	58.1	H3K36me1	41,080	9.37
H3K9me1	251,219	57.3	H3K79me1	26,267	5.99
H4K20me1	209,660	47.8	H3K27me2	24,619	5.61
H2BK5me1	205,864	46.9	H3R2me2	22,468	5.12
H3K4me2	183,363	41.8	H4K20me3	22,291	5.08
H3K4me3	153,220	34.9	H4R3me2	19,234	4.38
H3K27me1	140,633	32.1	H3K27me3	18,405	4.20
H3K36me3	120,050	27.4	H3K9me2	11,969	2.73
H2AZ	108,198	24.7	H3K9me3	10,726	2.45
H3R2me1	61,340	14.0	H3K79me2	6,295	1.44
H3K79me3	53,177	12.1	no modifications	147	0.034

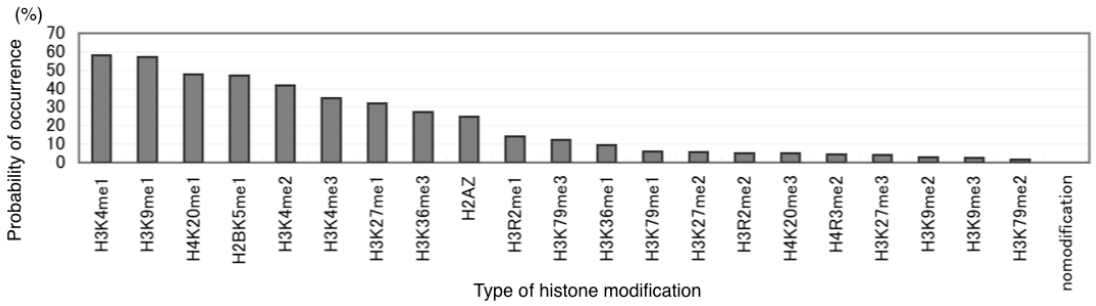

Fig. 3. Probability of modification occurrence. Y-axis indicates the name of (%) modification, and X-axis indicates the probability of modification occurrence (%).

3.1.2. Clustering of functional modification pattern candidates.

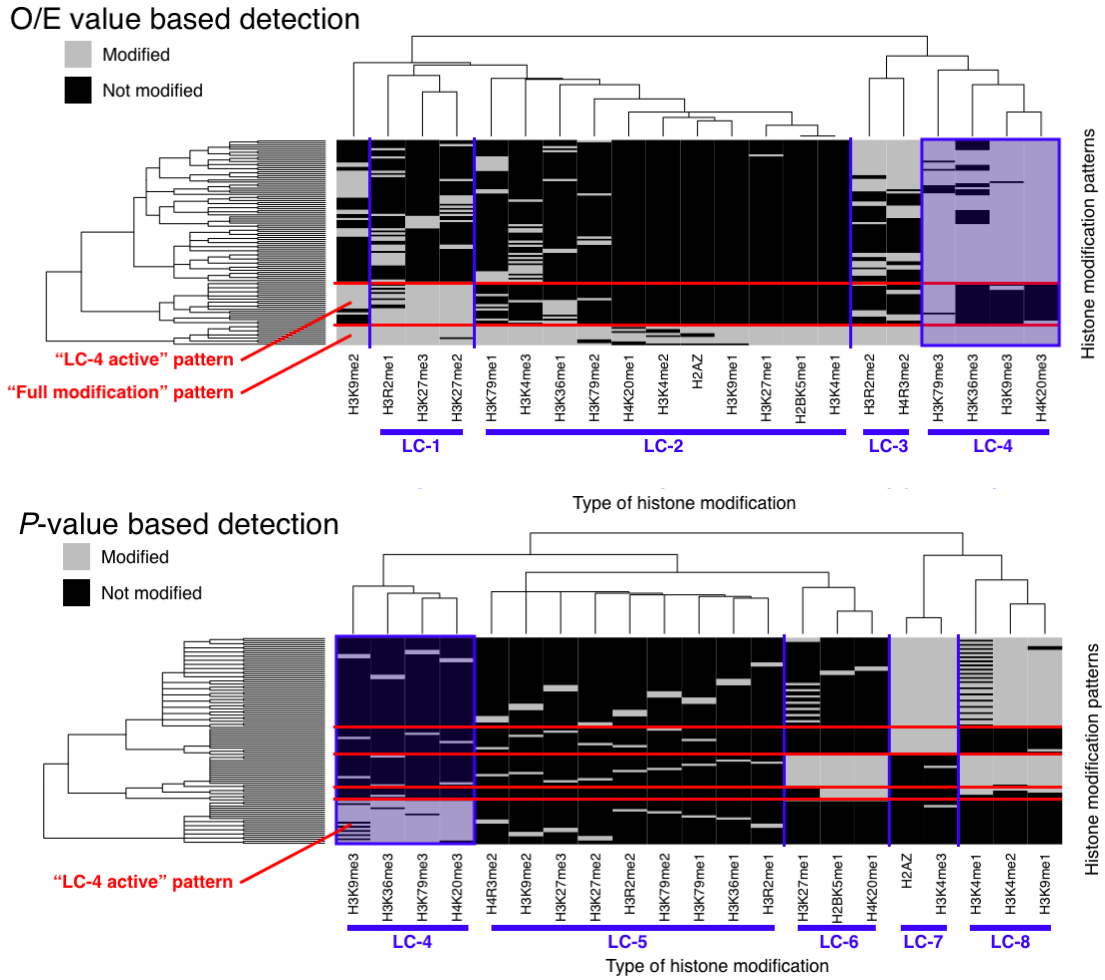

Fig. 4. Clustering of functional modification pattern candidates. Y-axis indicates each modification patterns, and X-axis indicates the name of modification. Gray color means "modified" and black color means "not modified". Red line represents the border of modification pattern and slue line indicates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

We detected 8 candidates of functional modification patterns. We also detected local combinations, LC-1 to 7. LC-4 is commonly detected by O/E value method and *P*-value method. LC-4 consists of H3K9me3, H3K36me3, H3K79me3, H4K20me3. "LC-4 active" modification pattern was observed in the clustering result of both methods. "Full modification" patterns had a highest O/E value.

3.1.3. DNA of "full modified" nucleosomes were highly methylated.

(a) (b)

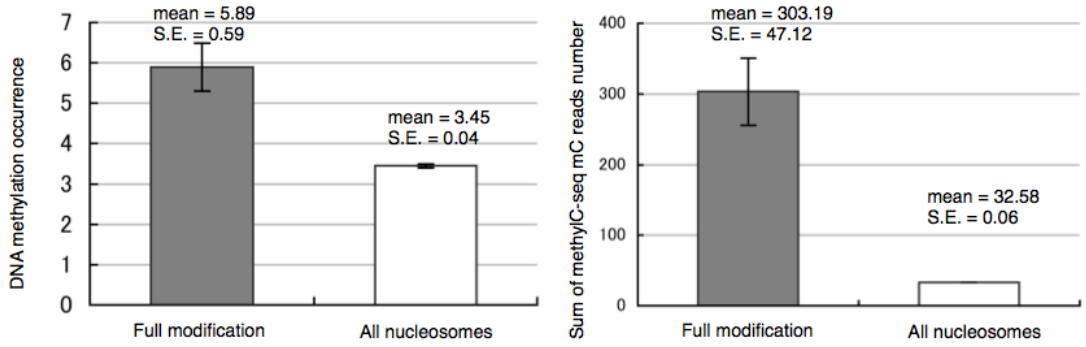

Fig. 5. Comparison of DNA methylation level between "full modified" nucleosomes and all nucleosomes. (a) Comparison of DNA methylation occourance. Y-axis indicates the number of DNA methylation, and X-axis indicates a modification pattern. (b) Comparison of methylC-seq mC reads number. Y-axis indicates the sum of methylC-seq mC reads number, and X-axis indicates a modification pattern.

The mean of DNA methylation occurrence was 5.89 for "Full modification" pattern (S.E. = 0.592) and 3.45 for all nucleosomes (S.E. = 0.045). DNA of "Full modification" pattern nucleosome had a significantly higher value than that of all nucleosomes (Wilcoxon test; $P < 4.1$ e-7). The mean of sum of methylC-seq mC reads number was 303.19 for "Full modification" pattern (S.E. = 47.116) and 32.58 for all nucleosomes (S.E. = 0.056). DNA of "Full modification" pattern nucleosome had a significantly higher value than that of all nucleosomes (Wilcoxon test; $P < 2.2$ e-16).

This result shows a relationship between "full modified" pattern and DNA methylation. Our methods could detect candidates of functional histone modification patterns. "Full modification" pattern was one of our candidates. And DNA of "full modified" nucleosome were hyper methylated. Those results suggested our method could detect functional histone modification.

3.2. Histone modification pattern detection from exons and introns.

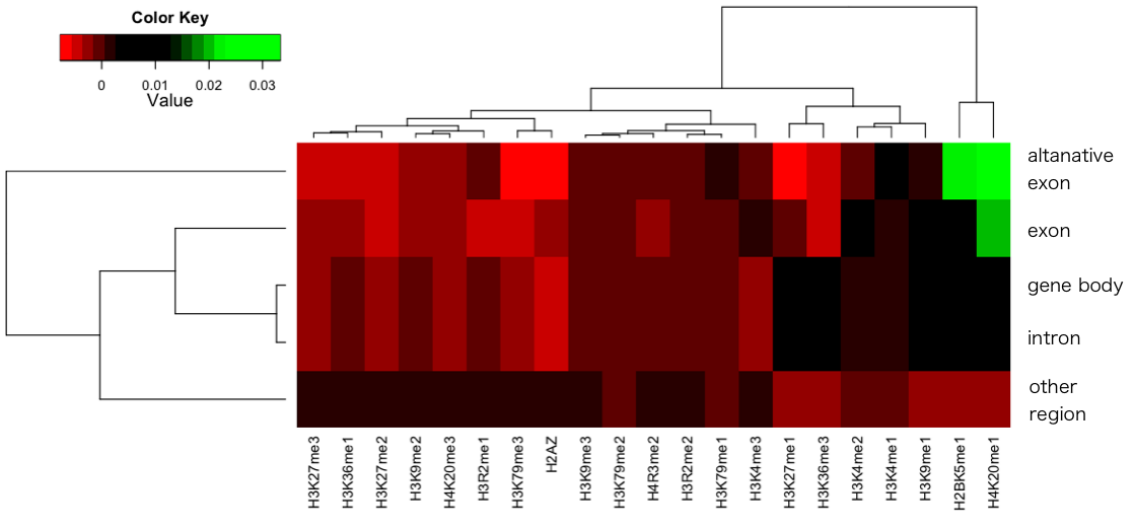

Fig. 6. Comparison of modification patterns between exon and intron. Y-axis indicates position of histones, and X-axis indicates the name of modification. Color key is Kullback-Leibler information for hisotone modification probability of genome and specific region.

Modification pattern of gene body and intron had similar tendency. The cause of this similarity is that introns account for a substantial fraction of gene body. Exon and altanative exons had also similar patterns. H4K20me1 showed high Kullback-Leibler information in exon and altanative exon. H2BK5me1 was altanative exon specific modification. H4K20me1 is known as transactive modification and silencing modification (Beisel, C. *et al*. 2002; Nishioka, K et al. 2002). There is a possibility that H4K20me1 is signal of altanative exon.

4. Future works

Our future work is finding more function of those candidates and to research relationships between the function and the modification patterns. To find functional modification patterns, we will analyze localization of histone modification patterns, such as around transcription start sites, enhancer sites, and promoter sites.

5. 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.