Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 345 (2006) 318–323
BBRC
www.elsevier.com/locate/ybbrc

Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity

Sung Mi Kim [a], Hae Jin Kee [b,c], Gwang Hyeon Eom [b,c], Nak Won Choe [b,c], Ji Young Kim [a], Young Soo Kim [a], Seong Ki Kim [a], Hoon Kook [d], Hyun Kook [b,c], Sang Beom Seo [a,*]

[a] *Department of Life Science, College of Natural Sciences, Chung-Ang University, Seoul 156-756, South Korea*
[b] *Medical Research Center for Gene Regulation, Chonnam National University Medical School, Gwangju 501-746, South Korea*
[c] *Department of Pharmacology, Chonnam National University Medical School, Gwangju 501-746, South Korea*
[d] *Department of Pediatrics, Chonnam National University Medical School, Gwangju 501-746, South Korea*

Received 17 April 2006
Available online 27 April 2006

Abstract

Evolutionary conserved SET domains were originally identified in three *Drosophila* proteins: suppressor of variegation (Su (var) 3–9), enhancer of zeste (E(z)), and the trithorax. Some of the SET-domain containing proteins have been known to elicit methylation of histone lysine residues. Based on a search for SET-domain containing proteins using bioinformatic tools, we identified and subsequently named a novel SET domain as WHISTLE, that has histone methyltransferase (HMTase) activity. To characterize WHISTLE, we performed an HMTase assay, mass spectrometric analysis, lysine specificity, and transfection assays. Mass spectrometric and immunoblot analysis revealed that WHISTLE di-methylates H3K4 and di-, and tri-methylates H3K27 of histones. Overexpression of WHISTLE repressed transcription of the SV40 promoter. Our results suggest that WHISTLE is a novel SET domain containing a protein with specific H3K4 and H3K27 HMTase activity.

Keywords: SET domain; HMTase; WHISTLE; H3K4; H3K27

Covalent modifications of the histones have been demonstrated to play important roles in the regulation of chromatin structure and transcriptional activities [1]. Among them, the methylation of specific lysines of histone proteins modifies target chromatin structure to activate or repress the transcription of the specific genes associated with chromatin [2]. H3K4, H3K36, and H3K79 have been correlated with transcriptional activation [3,4], while that of H3K9, H3K27, and H4K20 are marks of a repressed chromatin state [5,6].

Histone methyltransferases (HMTases) are the enzymes responsible for these modifications and are specific for either lysine or arginine residues. While arginine methylation is catalyzed by the PRMT/CARM family of HMTase, most lysine methylations are carried out by a SET-domain containing HMTase. Besides intensively investigated Suv39h1, H3 methyltransferase with testis-specific expression, Suv39h2, has been reported [7]. In addition, SET1, SET7, and SET9 were characterized as H3K4-specific HMTases and E(z)/EZH2 was identified for the H3K27 specific involved in the regulation of gene transcription [8,9].

Evolutionary conserved SET domains were originally identified in three *Drosophila* proteins: suppressor of variegation (Su (var) 3–9), enhancer of zeste (E(z)), and the trithorax. In mammalian, *Drosophila* homolog of trithorax, MLl/HRX, is frequently translocated in leukemia, resulting in oncogenic fusion proteins that lack the C-terminal SET domain [10]. Another SET-domain containing protein, NSD1, has been associated with the Wolf–Hirschhorn syndrome candidate (WHSC), suggesting that some members of the SET family might be linked to human pathologies [11].

* Corresponding author. Fax: +82 2 822 3059.
E-mail address: sangbs@cau.ac.kr (S.B. Seo).

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.04.095

Recent pioneering works of whole-genome sequencing and prediction of gene structure based on numerous bioinformatic tools have enabled us to identify genes with specific domains by a homology search approach. Applying these modern methods to find SET-domain containing proteins, we were able to collect more than 200 nucleotide sequences to predict their possession of a 'SET-domain' structure, from such entries as viruses, bacteria, plants, yeast, and the human.

Through homology searching and by following functional analyses, we herein report WHISTLE (*WHSC1-like 1 isoform 9 with methyltransferase activity to lysine*), to be a novel SET-domain containing protein with apparent specific HMTase activity. WHISTLE is a short alternative splice version of the Wolf–Hirschhorn Syndrome Candidate 1-like 1 (WHSC1L1), which was previously described to have the structure similar to Wolf-Hirschhorn Syndrome Candidate 1 (WHSC1) gene, but its functional characteristics have yet to be reported. In this study, we describe that a novel SET-domain containing protein WHISTLE methylates histone H3K4 and H3K27 residues and facilitates transcriptional repression.

Materials and methods

Bioinformatics. To identify proteins of the SET domain EnsMart, a data mining tool in Sanger web site (www.ensembl.org) was utilized along with the filter of pfam ID (PF00856). The homology of the proteins in human and mouse genomes was investigated and the novel proteins were selected. For those SET-containing proteins, possible alternative splice forms were collected and compared by a genome browser in Ensemble (www.ensembl.org). After retrieving the amino acid sequences, such proteins were blasted to Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to evaluate the probability to fit a SET domain and to find other structures. The candidate genes of interest were amplified by suitable primers with appropriate mouse cDNA libraries.

Plasmid construction. The full-length open reading frame of WHISTLE and SET domain-deleted WHISTLE were amplified by PCR from a mouse cDNA library (Clontech). The PCR products were then inserted into a bacterial expression vector (pGEX-4T1, Amersham Pharmacia Biotech) or into pcDNA3.1-HisTOPO (Invitrogen).

Peptide synthesis and LC–MS. Peptides (H3N1, H3N2, H3N3, and H4N1) were synthesized based on their N-terminal amino acid sequences of H3 and H4 histones (Peptron). A small quantity (100 μM) of synthetic peptides were used as substrate in the HMTase assay with WHISTLE; the reaction was stopped by TCA precipitation. After removing the precipitates by centrifugation, the supernatants were retrieved and methylated peptides in the supernatants analyzed by liquid chromatography–mass spectrometry (LC–MS) at the Korea Basic Science Institute.

Purification of recombinant proteins. The expression plasmids were transformed to *Escherichia coli* BL21 (DE3, Invitrogen) and the GST-WHISTLE protein was induced by IPTG; the cells were lysed in NET buffer (50 mM Tris–HCl (pH 7.8), 0.5 M NaCl, 1 mM EDTA, 0.01% Triton X-100, 1× protease inhibitor cocktail, and 0.5 g/L PMSF). Bacterial lysates were applied to glutathione–Sepharose beads and bead bound WHISTLE and part of the proteins were eluted.

HMTase assay. The HMTase assays were carried out according to a previous report and with some modifications [12]. The methylation reaction was set up as a 40 μL reaction mixture containing 1 g/L histone (Sigma), 150 nCi of *S*-adenosyl-[methyl-^{14}C]-L-methionine [^{14}C-SAM] (15 Ci/mmol, Amersham Pharmacia Biotech), and the enzyme (bead bound WHISTLE) in the HMTase assay buffer (50 mM Tris–HCl (pH 8.5), 20 mM KCl, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 1.25 M sucrose). Following overnight incubation at 28 °C, samples were loaded onto SDS–PAGE gels.

Cell culture and transient transfection. NIH3T3 cells were grown in DMEM with 10% fetal bovine serum (Gibco) and 0.05% antibiotics in a CO_2 incubator at 37 °C. In a luciferase assay (2.5×10^4 cells/well/48-well plate at their 50–60% confluence stage) cells were transfected with 100–300 ng of the expression plasmid and 200 ng of the reporter plasmid by Lipofectamine 2000 reagents as directed. After 48 h, the cells were harvested and subjected to the luciferase assay (Promega). The amount of DNA in each transfection was kept constant by adding a fixed amount of pcDNA3.0 vector.

Patients' samples. Bone marrow aspirates were obtained from seven patients with pediatric leukemias at the time of diagnosis. Control marrow samples were obtained from staging neuroblastoma patients subsequently shown to have no marrow involvement ($n = 2$), and a normal bone marrow donor ($n = 1$). Informed consent for the use of the marrow cells for research was obtained from parents in all cases, and study approval was obtained from the Institutional Review Board of the Chonnam National University Hospital.

Results

Bioinformatics screening and cloning

WHSC1L1 is predicted to be present in the 8th chromosome of both the mouse (gene ID: ENSMUSG00000054823) and the human (gene ID: ENSG00000147548). Data from blasting WHSC1L1 to the Conserved Domain Database suggested that the SET domain is localized in the carboxy terminal of the protein. The WHSC1L1 gene is expected to have several transcript variants and among them, the WHSC1L1 transcript variant 1 (XM_911776) is predicted to have the longest transcript of 9635 bp. Transcript variants 1 and 9 are expected to have an intact SET domain at the carboxy terminal. The carboxy terminal (400 bp) was amplified to prepare probes for Northern tissue blotting. As shown in Fig. 1A, the longer transcript at 10 kb was detected ubiquitously, while a strong signal at 18S was detected in testis tissue, suggesting the existence of shorter transcript variants of the WHSC1L1. Since the main purpose of this study was to assess the HMTase activity of the novel SET-containing protein, we continued to investigate the shortest transcript variant with an intact SET domain. We tried to amplify the transcript (variant 9) with specific primer sets from the mouse testis cDNA library (Clontech) and named it WHISTLE.

To investigate the pathological significance of WHISTLE in leukemia patients, RT-PCR-based WHISTLE expression analysis was performed for acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL) patients. By PCR analysis with specific primers for WHISTLE, it has been demonstrated that WHISTLE expressions were increased in both AML and ALL patients. Amino acid sequence analysis revealed that WHISTLE consists of 506 amino acids with three apparent distinct domains: PWWP, SET, and PostSET (Fig. 1C). The SET domain is located in the 215–337 amino acid sequence and shortly thereafter, the PostSET domain with a small cysteine-rich region was located at amino acids 340–356.

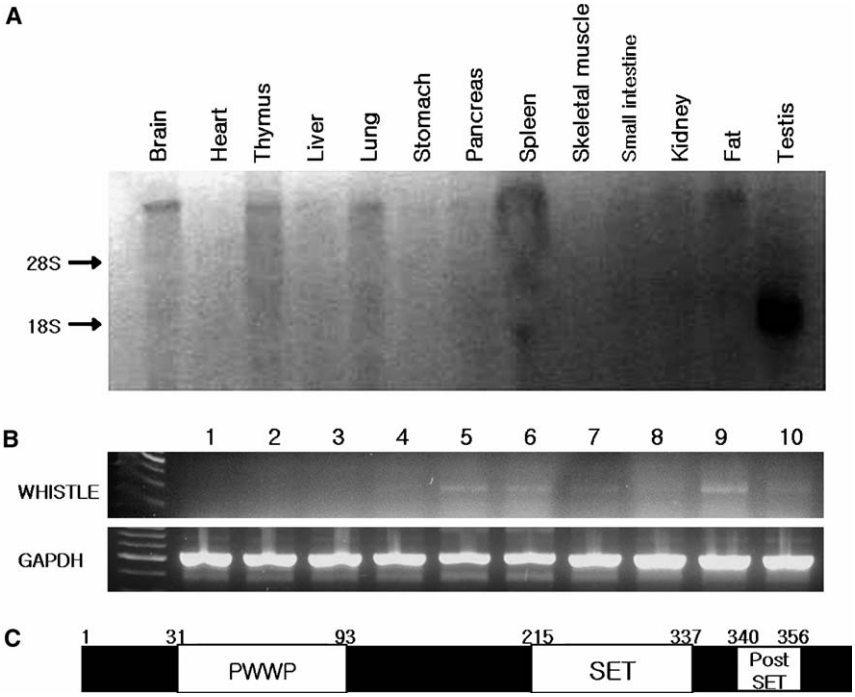

Fig. 1. Cloning of WHISTLE. (A) Northern blot results for WHISTLE. A mouse adult tissue blot was purchased and hybridized to a ^{32}P-labeled WHISTLE cDNA probe. (B) WHISTLE expression analysis by semi-quantitative RT-PCR from mononuclear cells of AML and ALL patients: 1–3, normal; 4–7, AML patients; 8–10, ALL patients. (C) Schematic view of WHISTLE domains: amino acids 31–93 were predicted to be the PWWP domain; amino acids 215–337 were predicted to be the SET domain; amino acids 340–356 were predicted to be the PostSET domain.

WHISTLE has methyltransferase activity when incubated with NIH3T3 cell extracts

The presence of a cysteine-rich region in WHISTLE raises the possibility that this protein has intact HMTase activity, since the domain is a characteristic feature of the Suv39h family of proteins with HMTase activity. To investigate whether the SET domain of WHISTLE indeed has enzyme activity, we performed an in vitro methylation assay with purified GST-WHISTLE protein with histone as substrate. We could not observe any HMTase activity in our experimental condition (Fig. 2), however, even when we used several combinations of buffers. The failure of the assay raised the possibility that WHISTLE had HMTase activity but only in an in vivo environment, thus requiring other cellular components for enzymatic activity.

Therefore, we designed an HMTase assay that would consider the presence of cellular lysates, and thereby mimic an in vivo environment for the protein. Purified WHISTLE protein was immunoprecipitated with NIH3T3 cell extracts and used in an in vitro HMTase assay. When the GST-WHISTLE protein beads were incubated with NIH3T3 cell extracts, as shown in Fig. 2, HMTase activity was clearly observed. As described above, no enzyme activity was seen when purified GST-WHISTLE was assayed while in contrast, PRMT1 (a positive control to methylate Arg3 on histone H4) elicited HMTase activity even in the absence of cell lysates, suggesting that their methylation mechanisms are different. HMTase activity was absent when histones and cell extracts alone were assayed as negative controls.

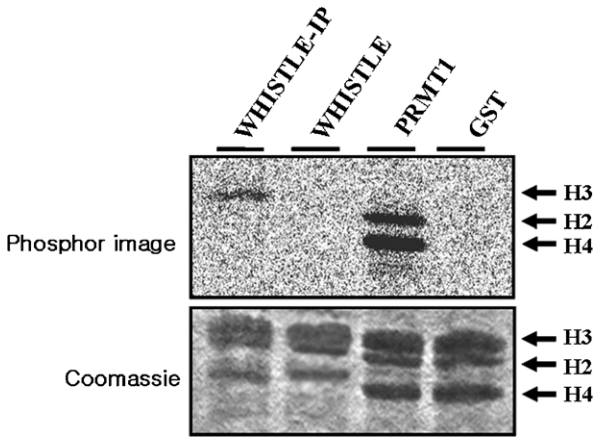

Fig. 2. HMTase activity of WHISTLE. Cell lysates incubated with WHISTLE beads and PRMT1 were detected with ^{14}C-labeled histones. Cell lysates incubated with WHISTLE beads were immunoprecipitated with NIH3T3 cell extracts overnight at 4 °C. PRMT1 is a positive control and both histone (no enzyme) and cell extract are negative controls. The upper panel is the result for phosphorimager; the lower panel is the result for the Coomassie staining.

H3K4 and H3K27 methylation and transcriptional repression by WHISTLE

To determine what lysine residue is methylated by WHISTLE, lysine specificity for methylation was investigated. We synthesized the four peptides based on the N-terminus of H3 and H4 of histones: H3N1, H3N2, H3N3, and H4N1. Each peptide consists of seven amino acids and has only one lysine as a target for methylation (Fig. 3). After

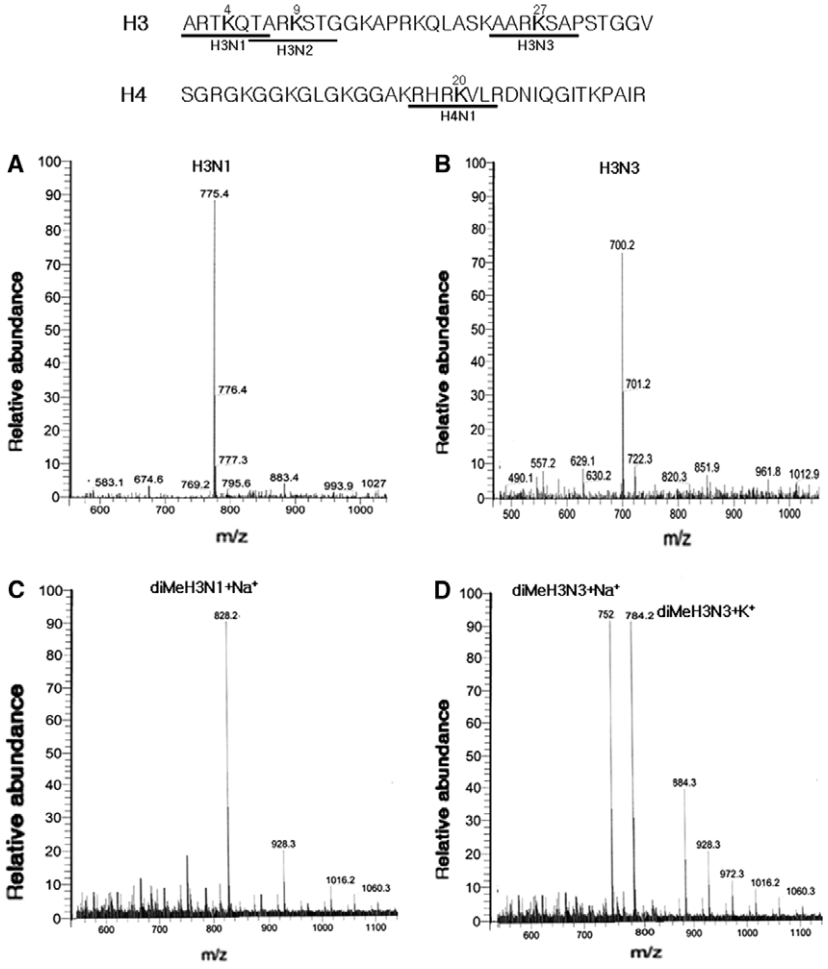

Fig. 3. Lysine specificity of WHISTLE HMTase activity. LC–MS results show that H3N1 peptides were dimethylated (A,C); H3N3 peptides were dimethylated and trimethylated (B,D). The mass of nonmethylated H3N1 peptides was 775.4 and dimethylated H3N1 was 828.2. Nonmethylated H3N3 peptide was 700.2, dimethylated H3N3 was 752.4, and trimethylated H3N3 was 784.2.

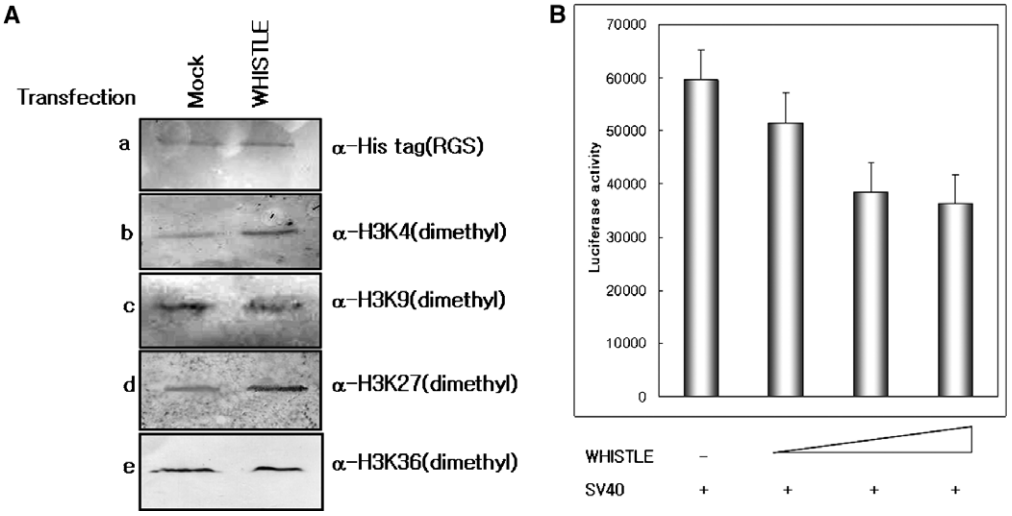

Fig. 4. Overexpression of WHISTLE increased the H3K4 and H3K27 dimethylation in NIH3T3 cells and induced transcriptional repression. (A) Transient transfected cell extracts were immunoblotted against anti-dimethyl H3K4, anti-dimethyl H3K9, anti-dimethyl H3K27, and anti-dimethyl H3K36. Anti-His antibodies were used as a control in detecting the expression proteins. (B) Overexpression of WHISTLE inhibits SV40-dependent transcriptional activation. NIH3T3 cells were transfected with SV40 and increasing concentrations of WHISTLE plasmids as indicated. Following transfection, cells were grown for 48 h, and cell extracts were prepared and assayed for luciferase activity.

the HMTase assay for each peptide, the specificity of methylation was analyzed by LC–MS. As shown by Fig. 3, dimethylated H3N1, dimethylated H3N3, and trimethylated H3N2 were detected. The majority of the H3N1 peptides were dimethylated (MW 828.2), which yielded larger peptides than the control nonmethylated peptides (MW 775.4). H3N3 peptides (MW 700.2 in nonmethylated form) were also observed in both dimethylated (MW 752.4) and trimethylated (MW 784.2) states. Likewise, H4N1 peptides turned out to be trimethylated (data not shown). In contrast, however, the H3N2 peptides did not show any peak changes. The LC–MS results indicated that WHISTLE has HMTase activity specific for the H3K4 and H3K27 residues.

We attempted to demonstrate if the methylations could take place in vivo. If so, this would implicate greater significance to the functional characteristics of WHISTLE. We preformed immunoblotting with specific antibodies in a WHISTLE-overexpressed cell condition. After transient transfection with WHISTLE, the cell lysates were probed with H3K4-dimethyl, H3K9-dimethyl, H3K27-dimethyl, and H3K36-dimethyl antibodies. In agreement with LC–MS analysis, H3K4 and H3K27 dimethylation were notably increased in the WHISTLE-transfected cells, when compared to mock-transfected controls (Fig. 4A). In contrast, H3K4 and H3K36 dimethylations remained unchanged.

To investigate the in vivo effects of H3K9, H3K27 methylation on transcription, NIH3T3 cells were co-transfected with the pcDNA3.1-WHISTLE and SV40 reporter construct and the luciferase activity was measured. As the concentration of WHISTLE increased, the luciferase activity decreased in a dose-dependent manner (Fig. 4B). These results suggest WHISTLE exerts repressive activities toward basal SV40 transcription.

Discussion

The Wolf–Hirschhorn syndrome (WHS) is a multiple malformation syndrome caused by the loss of one copy of a distal segment of chromosome 4p [13]. WHS is a rare anomaly with an incidence of 1/50,000 in births, that is characterized by various degrees of mental and growth retardation. WHS candidate 1 (WHSC1, also known as MMSET) was identified initially by its unique location in the translocation region [14]. In addition to its genomic location, the expression profiles of WHSC1 and its deduced function reveal the gene to be an excellent candidate for causative genes involved in cancer or other abnormalities in association with WHS. The WHSC1-like gene 1 (WHSC1L1) was also known as NSD3 (nuclear receptor-binding, su(var), enhancer-of-zeste, and trithorax domain-containing protein 3). WHSC1L1 has been identified by a database search using the WHSC1 sequence against the human genome sequence. By this homology search, it has been suggested that WHSC1L1 might have a role in embryonic development, although its functions remain to be identified [15].

We applied bioinformatic tools in this study to find the SET-domain containing proteins with unknown HMTase functions. Among them, we have identified the WHSC1L1 isoform 9 as possessing HMTase activity and named it WHISTLE. It is intriguing and quite unique that WHISTLE requires certain cellular components for its HMTase activity. SETDB1, however, could be an alternate example of WHISTLE. It has been reported that SETDB1 failed to show any appreciable methylation of core histones when the purified SETDB1 protein was assayed, however, HMTase activity has been detected in vivo [16]. Based on the SETDB1 case, it is plausible that WHISTLE also requires cofactors from the cell lysates for expression of full enzymatic activity, which implies that identifying those cofactors could be an important next step for further investigations.

The present result that WHISTLE was involved in transcriptional repression led us to conclude that WHISTLE can function as a transcriptional repressor through H3K4 and H3K27-specific dimethylation activities. In our study, we also observed that the protein is dysregulated in leukemia, again suggesting that it has a functional role in the development of certain type of cancers. The increase in the transcript amount is quite uniform despite the leukemia type, implying that WHISTLE might have a common role in the progression of leukemia. Thus, the functional role in hematologic malignancies as shown in the present study remains to be clarified.

Acknowledgments

This work was supported by grants from the Basic Research Program of the KOSEF (R01-2004-000-10121-0) and Cellular Bio Discovery Research Program Grant from MOST (2004M1AN01000125000) in South Korea. H.K. was supported in part by the Korea Science and Engineering Foundation through the Medical Research Center for Gene Regulation (R13-2002-013-03002-0) at Chonnam National University. H.J.K. was supported by a Korea Research Foundation grant (KRF-2004-C00141).

References

[1] B.D. Strahl, C.D. Allis, Gene silencing: trans-histone regulatory pathway in chromatin, Nature 418 (2002) 498.
[2] T. Kouzarides, Histone methylation in transcriptional control, Curr. Opin. Genet. Dev. 12 (2002) 198–209.
[3] H. Santos-Rosa, R. Schneider, A.J. Bannister, J. Sherriff, B.E. Bernstein, N.C. Emre, S.L. Schreiber, J. Mellor, T. Kouzarides, Active genes are tri-methylated at K4 of histone H3, Nature 419 (2002) 407–411.
[4] H.H. Ng, D.N. Ciccone, K.B. Morshead, M.A. Oettinger, K. Struhl, Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation, Proc. Natl. Acad. Sci. USA 100 (2003) 1820–1825.
[5] A.W. Snowden, P.D. Gregory, C.C. Case, C.O. Pabo, Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo, Curr. Biol. 12 (2002) 2159–2166.
[6] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R.S. Jones, Y. Zhang, Role of histone H3 lysine 27

methylation in Polycomb-group silencing, Science 298 (2002) 1039–1043.

[7] D. O'Carroll, H. Scherthan, A.H. Peters, S. Opravil, A.R. Haynes, G. Laible, S. Rea, M. Schmid, A. Lebersorger, M. Jerratsch, L. Sattler, M.G. Mattei, P. Denny, S.D. Brown, D. Schweizer, T. Jenuwein, Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression, Mol. Cell. Biol. 20 (2000) 9423–9433.

[8] A. Roguev, D. Schaft, A. Shevchenko, W.W. Pijnappel, M. Wilm, R. Aasland, A.F. Stewart, The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4, EMBO J. 20 (2001) 7137–7148.

[9] K. Nishioka, S. Chuikov, K. Sarma, H. Erdjument-Bromage, C.D. Allis, P. Tempst, D. Reinberg, Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation, Genes Dev. 16 (2002) 479–489.

[10] P.M. Waring, M.L. Cleary, Disruption of a homolog of trithorax by 11q23 translocations: leukemogenic and transcriptional implications, Curr. Top. Microbiol. Immunol. 220 (1997) 1–23.

[11] N. Kurotaki, N. Harada, K. Yoshiura, S. Sugano, N. Niikawa, N. Matsumoto, Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene, Gene 279 (2001) 197–204.

[12] S. Rea, F. Eisenhaber, D. O'Carroll, B.D. Strahl, Z.W. Sun, M. Schmid, S. Opravil, K. Mechtler, C.P. Ponting, C.D. Allis, T. Jenuwein, Regulation of chromatin structure by site-specific histone H3 methyltransferases, Nature 406 (2000) 593–599.

[13] K.Y. Gandelman, L. Gibson, M.S. Meyn, T.L. Yang-Feng, Molecular definition of the smallest region of deletion overlap in the Wolf–Hirschhorn syndrome, Am. J. Hum. Genet. 51 (1992) 571–578.

[14] I. Stec, T.J. Wright, G.B. Ommen, P.A.J. de Boer, A. Haeringen, A.F.M. Moorman, M.R. Altherr, J.T. Dunnen, WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphy gene maps in the Wolf–Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma, Hum. Mol. Genet. 7 (1998) 1527.

[15] I. Stec, G.B. Ommen, J.T. Dunnen, WHSC1L1, on Human Chromosome 8p11.2 WHSC1L1, on human chromosome 8p11.2, closely resembles WHSC1 and maps to a duplicated region shared with 4p16.3, Genomics 76 (2001) 5–8.

[16] D.C. Schultz, K. Ayyanathan, D. Negorev, G.G. Maul, F.J. Rauscher, 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 (2002) 919–932.