viral infection (Fig. 1e). These results indicate that NES1 is not sufficient, but its flanking regions including leucine-540 are necessary for cytoplasmic localization of CALM-AF10. Thus, we concluded that the NES2 region is the minimal NES that mediates cytoplasmic localization of CALM-AF10.

The nuclear export signal within CALM is necessary for CALM-AF10-induced immortalization of cells *in vitro*. We next investigated whether the NES within CALM-AF10 is required for leukemogenesis. To this end, primary murine bone-marrow stem/progenitor cells (HSPC) were infected with retrovirus encoding CALM-AF10, CALM^{NES4A}-AF10, NES2-AF10 and mAF10. Serial-replating assays revealed that both CALM-AF10 and NES2-AF10 immortalized HSPC, and that the cells formed colonies for at least five rounds of replating (Fig. 2a). By contrast, neither mAF10 nor CALM^{NES4A}-AF10, which lacks a functional CALM NES, could immortalize cells. Transduced cells with elevated colony-forming abilities also exhibited upregulation of the *Hoxa* cluster (*Hoxa5*, *Hoxa7*, *Hoxa9* and *Hoxa10*) and *Meis1* genes (Fig. 2b)[7,9]. These results indicated that the CALM NES is necessary for CALM-AF10 to immortalize hematopoietic stem/progenitor cells.

The nuclear export signal within CALM-AF10 is necessary to induce leukemia *in vivo*. To determine whether CALM-AF10 and NES2-AF10 can induce leukemia in mice, we injected bone-marrow progenitor cells transduced with CALM-AF10 and NES2-AF10 into lethally irradiated mice. Seven out of eight mice transplanted with cells expressing CALM-AF10 developed leukemia within 6 months after transplantation (Fig. 3a), and all mice transplanted with cells expressing NES2-AF10 developed leukemia within 3 months after transplantation. When cells prepared from bone marrow of these leukemic mice were transplanted into secondary recipient mice, all recipients promptly developed leukemia (medians: CALM-AF10 donors, 21 days [n = 4]; NES2-AF10 donors,

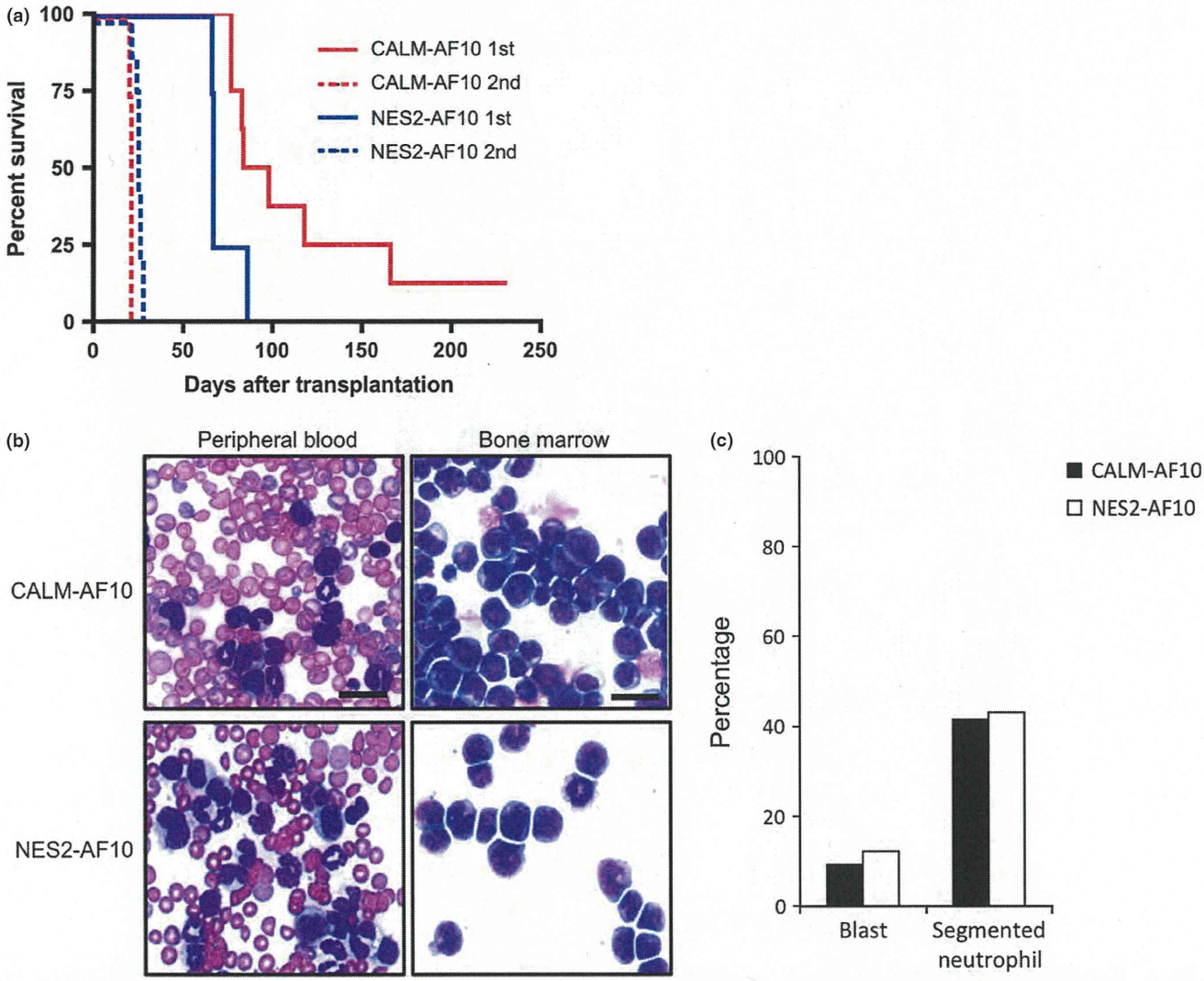

Fig. 3. The nuclear export signal within clathrin assembly lymphoid myeloid leukemia protein (CALM) is sufficient for leukemic transformation by CALM-AF10. (a) Survival of mice injected with murine bone-marrow cells transduced with FLAG-CALM-AF10 or FLAG-NES2-AF10. The leukemia-free survivals of the mice were analyzed. CALM-AF10 primary transplantation, n = 8; CALM-AF10 secondary transplantation, n = 4; NES2-AF10 primary transplantation, n = 4; NES2-AF10 secondary transplantation, n = 9. (b) Peripheral blood smears and bone-marrow cytospins were stained with May-Giemsa from CALM-AF10-transduced or NES2-AF10-transduced bone-marrow cells. Original magnification is 400×. (c) Population of blasts and segmented neutrophils in bone-marrow cells shown in (b). The scale bars represent 20 μm.

25 days [$n = 9$]). Morphological analysis revealed large populations of blast cells in leukemic mice receiving cells transduced with either CALM-AF10 or NES2-AF10 (Fig. 3b, c). Flow cytometry analysis showed that cells expressing CALM-AF10 and NES2-AF10 in the bone marrow cells of primary recipient mice were Mac1$^+$, CSF1R$^+$ and c-kit$^+$ (Fig. 4a). Moreover, as shown in Figure 4b, *Hoxa5*, *Hoxa7*, *Hoxa9*, *Hoxa10* and *Meis1* expression levels were upregulated in cells expressing CALM-AF10 and NES2-AF10 compared with normal bone marrow cells, although upregulation of *Hoxa5* and *Meis1* in primary recipient mice harboring NES2-AF10 was not significant ($P = 0.084$ and $P = 0.093$, respectively). These

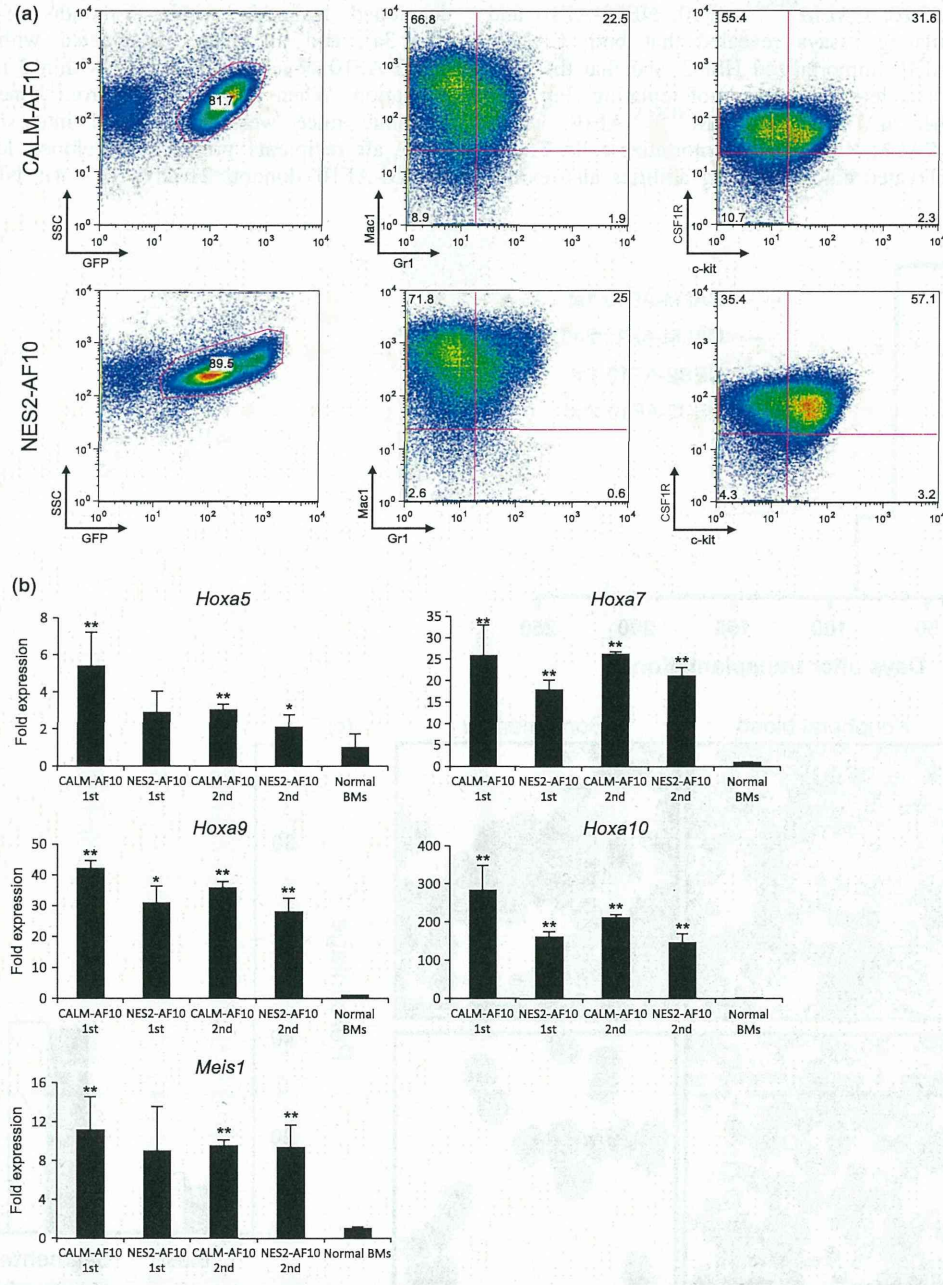

Fig. 4. Characterization of leukemic cells *in vivo*. (a) Flow cytometric analysis of leukemic cells. Murine bone-marrow cells were prepared from mice that developed leukemia after receiving transplantation of tumor cells transduced with CALM-AF10 or NES2-AF10, and were co-stained for Gr-1, Mac-1, colony stimulating factor 1 receptor (CSF1R) and c-kit; data are representative of CALM-AF10 primary transplantation ($n = 3$) and NES2-AF10 primary transplantation ($n = 3$). (b) *Hoxa* cluster and *Meis1* expression in mice receiving cells transduced with wild-type and mutant CALM-AF10. RNA transcripts were analyzed by real-time PCR of bone-marrow cells in mice that developed leukemia after CALM-AF10 and NES2-AF10 bone-marrow transplantation. Expression levels of *Hoxa5*, *Hoxa7*, *Hoxa9*, *Hoxa10* and *Meis1* were normalized against *Actb* and compared with wild-type whole bone marrow. Data are shown as means ± SEM from three independent leukemic mice. $*P < 0.05$; $**P < 0.01$ (*vs* normal bone-marrow cells).

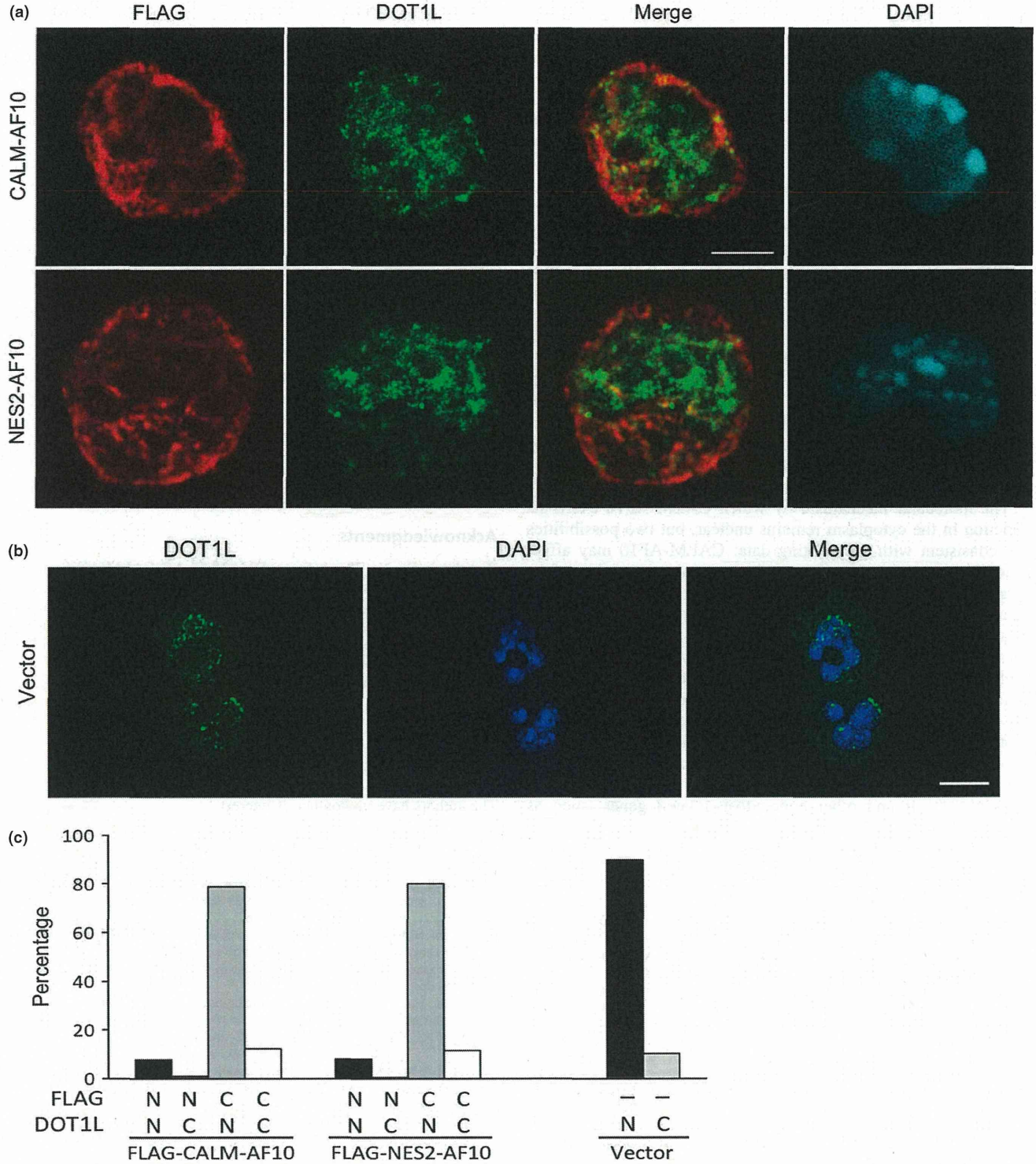

Fig. 5. Dot1L mainly localize in the nucleus in CALM-AF10-induced or NES2-AF10-induced leukemic cells. (a) Subcellular distribution of endogenous Dot1L in CALM-AF10-induced or NES2-AF10-induced leukemic cells. Cytospins of the cells were stained with anti-FLAG antibody (red), anti-DOT1L antibody (green) and DAPI (blue) and observed by confocal laser scanning microscopy. Note that GFP expression was not detected in the condition. (b) Subcellular distribution of endogenous Dot1L in the control vector-infected murine using fluorescence microscopy. (c) Population of leukemia cells expressing DOT1L and CALM-AF10 or FLAG-NES2-AF10 in the nucleus and the cytoplasm shown in (a) and (b). The scale bar represents 5 μm in (a) and 10 μm in (b).

data demonstrate that the NES within CALM-AF10 is a critical element for induction of leukemia.

It has been reported that CALM-AF10 interacts with the histone methyltransferase DOT1L to mediate H3K79 hypermethylation at the *Hoxa5* locus.[9] To determine whether Dot1L colocalizes with CALM-AF10 and NES2-AF10 in the leukemia cells, we performed immunofluorescence analysis. Dot1L mainly localized in the nucleus while CALM-AF10 and NES2-AF10 mainly localized in the cytoplasm (Fig. 5a,b). Dot1L partially colocalized with both CALM-AF10 and NES2-AF10, but neither CALM-AF10 nor NES2-AF10 altered the localization of Dot1L (Fig. 5a).

Discussion

AF10 and CALM localize diffusely in the nucleus and cytoplasm, respectively, whereas CALM-AF10 primarily localizes in cytoplasmic speckle domains (see Fig. 1c,d). The fact that CALM-AF10 regulates histone methylation at the *Hoxa5* locus suggests that CALM-AF10 is likely to function in the nucleus.[9] However, the results described here indicate that the CALM NES is essential for CALM-AF10-induced leukemogenesis, suggesting that cytoplasmic localization (or shuttling between nucleus and cytoplasm) is critical for the function of CALM-AF10. During the preparation of the manuscript, another group reported similar findings,[19] indicating that the results are reproducible and the conclusions can be validated using alternative experimental systems.

The molecular mechanism by which CALM-AF10 exerts its function in the cytoplasm remains unclear, but two possibilities are consistent with the existing data: CALM-AF10 may affect cytoplasmic signaling pathways that regulate expression of its target genes, including *HoxA* cluster genes; alternatively, CALM-AF10 may affect the functions of transcriptional regulators by changing their localization from the nucleus to the cytoplasm. DOT1L, a candidate mediator of CALM-AF10-induced leukemia, interacts with AF10 and induces H3K79 hypermethylation at *Hoxa5*.[9] However, our present data suggest that CALM-AF10 and NES2-AF10 did not affect the localization of Dot1L (see Fig. 5a). It is possible that CALM-AF10 squelches DOT1L inhibitors by exporting them to the cytoplasm.

CALM plays an important role in clathrin-mediated endocytosis.[17,20,21] It and other endocytosis-related genes, such as EPS15, EEN, CLTC and HIP1, are involved in multiple types of leukemia-associated chromosomal translocations (e.g. MLL-CALM, MLL-EPS15, MLL-EEN, CLTC-ALK and HIP1-PDGFBR),[22–26] suggesting that these leukemia-associated fusions might affect endocytosis in a manner that contributes to leukemogenesis. However, recent reports have shown that the clathrin-binding domain of CALM is not essential for CALM-AF10-mediated leukemogenesis.[27,28] Here, we show that nuclear export of CAL-AF10 is critical for the leukemogenesis. Because the endocytosis-related proteins mentioned above are also exported from the nucleus to the cytoplasm, as in the case of CALM,[11,23,29] it is possible that changes in the localization of fusions involving endocytosis-related proteins have some shared consequence that is important for leukemogenesis.

Molecular exchange between the nucleus and cytoplasm takes place through nuclear pore complexes (NPC). Fusion proteins containing NUP98 and NUP214, which are components of the NPC, have been found in AML and T-ALL;[30,31] as in cells expressing CALM-AF10, a set of *Hoxa* and *Meis1* genes are upregulated in leukemia cells expressing these NUP98 fusions and NUP214 fusions.[32,33] In addition, the NPC-component fusions interact with CRM1, the major receptor for the nuclear export of protein.[33,34] These observations suggest that alteration of the localization of certain factors by NUP98 fusions and NUP214 fusions might be important for leukemogenesis, and that a common mechanism may underlie leukemias induced by CALM and NUP fusions.

Acknowledgments

We thank Dr Y. Zhang for pcDNA3β-FLAG-CALM-AF10 plasmid. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Health, Labor, and Welfare and from the Ministry of Education, Culture, Sports, Science, and Technology; the National Cancer Center Research and Development Fund; the Naito Foundation; Cosmetology Research Foundation; and Nara Women's University Intramural Grant for Project Research. MS is a Research Fellow for Yong Scientist of Japan Society for the Promotion of Science.

Disclosure Statement

The authors have no conflict of interest.

References

1 Dreyling MH, MartinezCliment JA, Zheng M, Mao J, Rowley JD, Bohlander SK. The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc Natl Acad Sci USA* 1996; **93**: 4804–9.
2 Bohlander SK, Muschinsky V, Schrader K *et al.* Molecular analysis of the CALM/AF10 fusion: identical rearrangements in acute myeloid leukemia, acute lymphoblastic leukemia and malignant lymphoma patients. *Leukemia* 2000; **14**: 93–9.
3 Narita M, Shimizu K, Hayashi Y *et al.* Consistent detection of CALM-AF10 chimaeric transcripts in haematological malignancies with t(10;11)(p13;q14) and identification of novel transcripts. *Br J Haematol* 1999; **105**: 928–37.
4 Linder B, Newman R, Jones LK *et al.* Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *J Mol Biol* 2000; **299**: 369–78.
5 Wysocka J, Swigut T, Xiao H *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 2006; **442**: 86–90.
6 Greif PA, Tizazu B, Krause A, Kremmer E, Bohlander SK. The leukemogenic CALM/AF10 fusion protein alters the subcellular localization of the lymphoid regulator Ikaros. *Oncogene* 2008; **27**: 2886–96.
7 Caudell D, Zhang Z, Chung YJ, Aplan PD. Expression of a CALM-AF10 fusion gene leads to Hoxa cluster overexpression and acute leukemia in transgenic mice. *Cancer Res* 2007; **67**: 8022–31.
8 Mulaw MA, Krause A, Deshpande AJ *et al.* CALM/AF10-positive leukemias show upregulation of genes involved in chromatin assembly and DNA repair processes and of genes adjacent to the breakpoint at 10p12. *Leukemia* 2012; **26**: 1012–9.
9 Okada Y, Jiang Q, Lemieux M, Jeannotte L, Su L, Zhang Y. Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. *Nat Cell Biol* 2006; **8**: 1017–24.
10 Chen L, Deshpande AJ, Banka D *et al.* Abrogation of MLL-AF10 and CALM-AF10-mediated transformation through genetic inactivation or pharmacological inhibition of the H3K79 methyltransferase Dot1l. *Leukemia* 2013; **27**: 813–22.
11 Vecchi M, Polo S, Poupon V, van de Loo JW, Benmerah A, Di Fiore PP. Nucleocytoplasmic shuttling of endocytic proteins. *J Cell Biol* 2001; **153**: 1511–7.
12 Borkhardt A, Haas OA, Strobl W *et al.* A novel type of Mll/Af10 fusion transcript in a child with acute megakaryocytic leukemia (Aml-M7). *Leukemia* 1995; **9**: 1796–7.
13 Brandimarte L, Pierini V, Di Giacomo D *et al.* New MLLT10 gene recombinations in pediatric T-acute lymphoblastic leukemia. *Blood* 2013; **121**: 5064–7.

14 Lee JW, Liao PC, Young KC *et al.* Identification of hnRNPH1, NF45, and C14orf166 as novel host interacting partners of the mature hepatitis C virus core protein. *J Proteome Res* 2011; **10**: 4522–34.
15 Ennas MG, Sorio C, Greim R *et al.* The human ALL-1/MLL/HRX antigen is predominantly localized in the nucleus of resting and proliferating peripheral blood mononuclear cells. *Cancer Res* 1997; **57**: 2035–41.
16 Yedavalli VSRK, Neuveut C, Chi YH, Kleiman L, Jeang KT. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* 2004; **119**: 381–92.
17 Tebar F, Bohlander SK, Sorkin A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol Biol Cell* 1999; **10**: 2687–702.
18 Takagi M, Nishioka M, Kakihara H *et al.* Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl Environ Microbiol* 1997; **63**: 4504–10.
19 Conway AE, Scotland PB, Lavau CP, Wechsler DS. A CALM-derived nuclear export signal is essential for CALM-AF10-mediated leukemogenesis. *Blood* 2013; **121**: 4758–68.
20 Ford MGJ, Pearse BMF, Higgins MK *et al.* Simultaneous binding of PtdIns (4,5)P-2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* 2001; **291**: 1051–5.
21 Meyerholz A, Hinrichsen L, Groos S, Esk PC, Brandes G, Ungewickell EJ. Effect of clathrin assembly lymphoid myeloid leukemia protein depletion on clathrin coat formation. *Traffic* 2005; **6**: 1225–34.
22 Bernard OA, Mauchauffe M, Mecucci C, Vandenberghe H, Berger R. A novel gene, Af-1p, fused to Hrx in T(1,11)(P32, Q23), is not related to Af-4, Af-9 nor Enl. *Oncogene* 1994; **9**: 1039–45.
23 So CW, So CKC, Cheung N, Chew SL, Sham MH, Chan LC. The interaction between EEN and Abi-1, two MLL fusion partners, and synaptojanin and dynamin: implications for leukaemogenesis. *Leukemia* 2000; **14**: 594–601.
24 Bridge JA, Kanamori M, Ma ZG *et al.* Fusion of the ALK gene to the clathrin heavy chain gene, CLTC, in inflammatory myofibroblastic tumor. *Am J Pathol* 2001; **159**: 411–5.
25 Wechsler DS, Engstrom LD, Alexander BM, Motto DG, Roulston D. A novel chromosomal inversion at 11q23 in infant acute myeloid leukemia fuses MLL to CALM, a gene that encodes a clathrin assembly protein. *Genes Chromosom Cancer* 2003; **36**: 26–36.
26 Ross TS, Bernard OA, Berger R, Gilliland DG. Fusion of Huntingtin interacting protein 1 to platelet-derived growth factor beta receptor (PDGF beta R) in chronic myelomonocytic leukemia with t(5;7)(q33;q11.2). *Blood* 1998; **91**: 4419–26.
27 Deshpande AJ, Rouhi A, Lin Y *et al.* The clathrin-binding domain of CALM and the OM-LZ domain of AF10 are sufficient to induce acute myeloid leukemia in mice. *Leukemia* 2011; **25**: 1718–27.
28 Stoddart A, Tennant TR, Fernald AA, Anastasi J, Brodsky FM, Le Beau MM. The clathrin-binding domain of CALM-AF10 alters the phenotype of myeloid neoplasms in mice. *Oncogene* 2012; **31**: 494–506.
29 Engqvist-Goldstein AEY, Kessels MM, Chopra VS, Hayden MR, Drubin DG. An actin-binding protein of the Sla2/Huntingtin interacting protein 1 family is a novel component of clathrin-coated pits and vesicles. *J Cell Biol* 1999; **147**: 1503–18.
30 von Lindern M, Breems D, van Baal S, Adriaansen H, Grosveld G. Characterization of the translocation breakpoint sequences of two DEK-CAN fusion genes present in t(6;9) acute myeloid leukemia and a SET-CAN fusion gene found in a case of acute undifferentiated leukemia. *Genes Chromosom Cancer* 1992; **5**: 227–34.
31 Nakamura T, Largaespada DA, Lee MP *et al.* Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat Genet* 1996; **12**: 154–8.
32 Nakamura T. NUP98 fusion in human leukemia: dysregulation of the nuclear pore and homeodomain proteins. *Int J Hematol* 2005; **82**: 21–7.
33 Van Vlierberghe P, van Grotel M, Tchinda J *et al.* The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* 2008; **111**: 4668–80.
34 Takeda A, Sarma NJ, Abdul-Nabi AM, Yaseen NR. Inhibition of CRM1-mediated nuclear export of transcription factors by leukemogenic NUP98 fusion proteins. *J Biol Chem* 2010; **285**: 16248–57.

Published online 23 January 2014

Nucleic Acids Research, 2014, Vol. 42, No. 7 ***4241–4256***
doi:10.1093/nar/gkt1394

MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters

Hiroshi Okuda[1,†], Marie Kawaguchi[2,†], Akinori Kanai[3,†], Hirotaka Matsui[3,†], Takeshi Kawamura[4], Toshiya Inaba[3], Issay Kitabayashi[2] and Akihiko Yokoyama[1,*]

[1]Laboratory for Malignancy Control Research, Medical Innovation Center, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan, [2]Division of Hematological Malignancy, National Cancer Center Research Institute, Tokyo 104-0045, Japan, [3]Department of Molecular Oncology and Leukemia Program Project, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima 734-8553, Japan and [4]Department of Molecular Biology and Medicine, Laboratory for System Biology and Medicine (LSBM), Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Tokyo 153-8904, Japan

Received July 19, 2013; Revised December 17, 2013; Accepted December 19, 2013

ABSTRACT

Mixed-lineage leukemia (MLL) maintains the expression of cellular memory genes during development, while leukemic MLL fusion proteins aberrantly maintain expression of hematopoietic stem cell program genes such as *HOXA9* to cause leukemia. However, the molecular mechanism of gene activation is unclear. Here we show that only two functional modules are necessary and sufficient for target recognition: those that bind to non-methylated CpGs and di-/tri-methylated histone H3 lysine 36 (H3K36me2/3). An artificial protein composed of the two targeting modules and an interaction domain for AF4-family coactivators can functionally substitute for MLL fusion proteins. Because H3K36me2/3 markers are indicative of active transcription, MLL fusion proteins target previously active CpG-rich genes and activate transcription by recruiting coactivators thereto. Our results indicate that such chromatin context-dependent gene activation is the fundamental mechanism by which MLL fusion proteins maintain the expression of the cellular memory/hematopoietic stem cell program genes.

INTRODUCTION

The *MLL* gene encodes an epigenetic regulator that maintains *HOX* gene expression during embryogenesis (1). *HOX* genes are so-called cellular memory genes because their expression is maintained throughout the development. In the hematopoietic lineage, the MLL protein (also known as HRX, MLL1 and KMT2A) activates the transcription of posterior *HOXA* genes (e.g. *HOXA7-A10*) (2,3). Posterior *HOXA* genes are hematopoietic stem cell (HSC) program genes (4) that promote the self-renewal of HSCs/immature progenitors (5). In normal hematopoiesis, their expression is maintained by MLL in the HSC/immature progenitor compartments, which diminishes as cells differentiate. Chromosomal translocation generates *MLL* fusion genes, whose products constitutively activate the posterior *HOXA* genes, which results in aberrant self-renewal of hematopoietic progenitors, leading to leukemia (6). However, the precise molecular mechanism by which MLL and MLL fusion proteins activate their target genes remains unclear. MLL fusion proteins exert their oncogenic functions as a complex with the lens epithelium-derived growth factor (LEDGF) (also known as PSIP1) (7). Disruption of *Psip1* in mice causes homeotic skeletal transformation, a characteristic phenotype caused by dysregulation of *Hox* gene expression (8). LEDGF also facilitates the specific integration of the HIV genome into transcriptionally active regions, presumably by tethering the HIV genome/integrase complex with transcriptionally active chromatin (9,10). In leukemia, *MLL* frequently fuses with the *ALL1-fused gene from chromosome 4* (*AF4*)- and *eleven-nineteen leukemia* (*ENL*)-family genes, whose protein products are the components of a transcriptional coactivator, termed AEP (the AF4 family/ENL family/PTEFb complex) (11–14). MLL fusion proteins constitutively recruit AEP components to activate transcription through direct interaction (e.g. MLL-ENL) or by an as-yet-uncharacterized indirect mechanism (e.g. MLL-AF6) (14). These previous studies postulate that MLL fusion proteins recognize their target genes through the MLL portion and constitutively recruit the AEP coactivator through the fusion partner portion.

*To whom correspondence should be addressed. Tel: +81 75 366 7441; Fax: +81 75 752 9132; Email: yokoyama@dsk.med.kyoto-u.ac.jp

†These authors contributed equally to the paper as first authors.

Here, we report that MLL fusion proteins recognize a specific chromatin context to activate cellular memory genes. Our structure/function analyses demonstrate that MLL fusion proteins recognize their target genes through two functional domains: the PWWP domain of LEDGF and the CXXC domain of MLL that specifically bind to H3K36me2/3 and non-methylated CpGs, respectively. Because H3K36me2 and H3K36me3 are generally linked to gene activation (15,16), a previously transcribed gene containing non-methylated CpGs in its promoter is subjected to transcriptional activation by MLL fusion proteins. These studies provide a novel chromatin context-dependent gene activation mechanism by which MLL fusion proteins maintain cellular memory.

MATERIALS AND METHODS

Cell culture

The human leukemia cell line ML-2 was cultured in RPMI 1640 medium supplemented with 10% fetal calf serum and non-essential amino acids. 293T, plat-E and HeLa cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and non-essential amino acids.

Vector construction

The pMSCV-neo MLL-ENL, PME (7) and Bcl2-ires-E2A-HLF (17) vectors are described elsewhere. Various MLL fusion constructs were generated by restriction enzyme digestion or polymerase chain reaction (PCR)-based mutagenesis. The cDNAs were cloned into the pMSCV neo vector (for virus production) (Clontech, Mountain View, CA) or the pCMV5 vector (for transient expression). The expression vectors for FLAG-tagged GAL4 fusion proteins were constructed by PCR using pM (Clontech) as template and cloned into the pCMV5 vector. Various LEDGF constructs were generated by PCR-based mutagenesis and cloned into the pcDNA4 HisMax vector (Invitrogen, Carlsbad, CA).

Western blotting

Western blotting was performed as described elsewhere (14). The antibodies used in this study have been listed in Supplementary Table S2.

Reverse transcription-PCR

RNAs were prepared using the RNeasy kit (Qiagen, Valencia, CA) and reverse-transcribed with the Superscript III first strand cDNA synthesis kit with oligo(dT) primers (Invitrogen). Expression of PCE was confirmed by regular PCR using the primer pair 5′-AGA ATTCGATATCGGAAACATGACTCGCGATTTCAA ACC-3′/5′-TCACGTGTTCGCGATGCGACGGGCTTT CGTGGAGGAG-3′. Quantitative PCR (qPCR) for other genes was performed as described previously (14) using TaqMan probes [*Gapdh* (Mm99999915_g1), *Hoxa7* (Mm00657963_m1), *Hoxa9* (Mm00439364_m1), *Hoxa10* (Mm00433966_m1), *Cbx5* (Mm00483092_m1), *Myb* (Mm00501741_m1) and *Hmgb3* (Mm01377544_gH) (Applied Biosystems)]. The expression levels, normalized to *Gapdh*, were estimated using a standard curve and the relative quantification method, as described in ABI User Bulletin #2.

Virus production

The ecotropic retrovirus was produced using plat-E packaging cells (18). Supernatant medium containing the virus was harvested 24–48 h after transfection and used for viral transduction.

Myeloid progenitor transformation assay

The myeloid progenitor transformation assay was performed using cells harvested from the femurs and tibiae of C57BL/6 mice. C-kit-positive cells were enriched using magnetic beads conjugated with anti-c-kit antibody (Miltenyi Biotech, Bergisch Gladbach, Germany), transduced with recombinant retrovirus by spinoculation (19) and plated in methylcellulose medium (Iscove's modified Dulbecco's medium, 20% fetal bovine serum, 1.6% methylcellulose, 100 μM β-mercaptoethanol) containing stem cell factor, interleukin-3 and granulocyte macrophage colony-stimulating factor (10 ng/ml). The colony-forming units per 10^4 plated cells at the third and fourth rounds were quantified after 5–7 days of culture.

Subcellular fractionation and nucleosome co-immunoprecipitation analysis

Subcellular fractions of 293T cells were obtained by CSK buffer extraction (20) and MNase treatment. 293T cells cultured in a 10-cm dish were resuspended in 1 ml of CSK buffer [100 mM NaCl, 10 mM PIPES (pH 6.8), 3 mM $MgCl_2$, 1 mM EGTA (pH 7.6), 0.3 M sucrose, 0.5% Triton X-100, 5 mM sodium butyrate, 0.5 mM DTT, EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) and 2 mM vanadyl ribonucleoside complexes (Sigma, St. Louis, MO)], incubated on ice for 5 min and then centrifuged (400*g*, 4°C, 4 min). The supernatant (the soluble fraction) was transferred to a new tube, and the pellet was resuspended in 1 ml of MNase buffer [50 mM Tris-HCl (pH 7.5), 4 mM $MgCl_2$, 1 mM $CaCl_2$, 0.3 M sucrose, 5 mM sodium butyrate, 0.5 mM DTT and protease inhibitor cocktail]. One unit of MNase (Sigma) was added to the suspension, and the mixture was incubated at 37°C for 10–12 min to obtain mononucleosomes. The MNase reaction was stopped by adding EDTA (pH 8.0) at a final concentration of 20 mM. The reaction mixture was centrifuged (13 000 rpm, 4°C, 5 min) to separate the supernatant (the nucleosome fraction) and the pellet. The pellet was resuspended in elution buffer (1% sodium dodecyl sulfate, 50 mM $NaHCO_3$). The nucleosome fractions were subjected to immunoprecipitation (IP) using specific antibodies; washed five times with MNase buffer with 20 mM EDTA and analyzed by western blotting, Oriole staining (Bio-Rad, Hercules, CA), SYBR Green staining (Lonza, Basel, Switzerland) and mass spectrometry. Optionally, the precipitates were washed with MNase buffer three times, treated with DNase I (Qiagen) for 10 min at 37°C and washed five times with MNase buffer with 20 mM

EDTA. The precipitates were analyzed by western blotting and SYBR Green staining.

Fractionation-assisted native chromatin immunoprecipitation

293T cells cultured in a 10-cm dish or ML-2 cells (2×10^7) were suspended in 1 ml of CSK buffer to remove chromatin-unbound materials. The pellet was then resuspended in 1 ml of MNase buffer and mixed with 0.2–1 U of MNase for 5–10 min at 37°C to obtain chromatin with DNA lengths of 150–3000 base pairs (bp). The MNase reaction was stopped by adding EDTA (pH 8.0) at a final concentration of 20 mM, and then the mixture was placed on ice. Lysis buffer (1 ml) [10% glycerol, 20 mM sodium phosphate (pH 7.0), 250 mM NaCl, 30 mM sodium pyrophosphate, 0.1% Nonidet P-40, 5 mM EDTA, 10 mM NaF and protease inhibitor cocktail] (21) was added to solubilize the proteins, and the mixture was spun down to remove insoluble material. The supernatant was then subjected to IP. Approximately 1 μg of the primary antibodies listed in Supplementary Table S2 was added to 400 μl of chromatin suspension, and the mixture was incubated for 4–6 h at 4°C. Ten microliters of protein-G magnetic beads (Invitrogen) was added to each sample, and the mixture was incubated for 2 h with rotation. The beads were washed five times with 500 μl of a 1:1 mixture of MNase buffer containing 20 mM EDTA and lysis buffer. For histone modifications, incubation and washing were performed with higher NaCl concentrations (plus 400 mM) to strip chromatin-binding proteins. DNAs were harvested from the precipitates by dissolving in elution buffer followed by phenol/chloroform/iso-amyl alcohol extraction and ethanol precipitation. Deep sequencing of the precipitated DNAs was performed using the Illumina genome analyzer IIx at the Joint Usage/Research Center (RIRBM), Hiroshima University. qPCR was performed on the precipitated DNAs using the custom-made primer sets described in Supplementary Table S3. The value relative to the input was determined using a standard curve and the relative quantification method as described in the ABI User Bulletin #2. Optionally, co-precipitated proteins were harvested by dissolving the immunoprecipitates in elution buffer, mixed with an equal volume of 2× sodium dodecyl sulfate-polyacrylamide gel electrophoresis sample buffer and subjected to western blotting.

Liquid chromatography-tandem mass spectrometry analysis

The method for trypsin digestion of protein has been described previously (22). Tandem mass spectrometry analysis was performed using an LTQ Orbitrap ELITE ETD mass spectrometer (Thermo Fisher Scientific). The methods used for liquid chromatography-tandem mass spectrometry (LC-MS/MS) were slightly modified from those described previously (23). The mass spectrometer was operated in data-dependent acquisition mode in which MS acquisition with a mass range of m/z 400–1000 was automatically switched to MS/MS acquisition under the control of Xcalibur software. The top four precursor ions in the MS scan were selected by Orbitrap, with resolution R = 240 000, and those in the subsequent MS/MS scans, with an ion trap in automated gain control (AGC) mode, where AGC values were 1×10^6 and 1.00×10^4 for full MS and MS/MS, respectively. For fragmentation, electron transfer dissociation was used.

In vivo leukemogenesis assay

C-kit-positive cells (2×10^5) prepared from mouse femurs and tibiae were transduced with retrovirus by spinoculation and intravenously transplanted into sublethally irradiated (two doses of 500 rad in 2 days) C57BL/6 mice. Moribund mice were sacrificed, and the spleen cells were either subjected to cytospin preparation followed by May-Grunwald/Giemsa staining or temporarily cultured in methylcellulose medium in the presence of G418 (1 mg/ml) to remove untransformed cells, and then subjected to secondary transplantation or reverse transcription-PCR (RT-PCR) analysis.

CpG island recovery assay

CpG island recovery assays for non-methylated CpGs (CIRA) and methylated CpGs (MIRA) were performed using the Unmethyl Collector kit and Methyl Collector Ultra kit, respectively (Active Motif, Carlsbad, CA). Deep sequencing after CIRA and MIRA was carried out at the Joint Usage/Research Center (RIRBM), Hiroshima University.

Transactivation assay

Transactivation assays using the pFR-luc reporter (Clontech) were performed as described elsewhere (14). Relative luciferase activities were normalized to *Renilla* luciferase activity and expressed in terms of the average values and standard deviations (SDs) of triplicate determinations relative to the GAL4 DNA binding domain controls.

RESULTS

Murine leukemia models define the major functional modules required for leukemic transformation by MLL-ENL

MLL fusion proteins form a trimeric complex with menin and LEDGF through the MLL portion (7). Because MLL fusion proteins associate with LEDGF through menin as a mediator, an MLL-ENL mutant (MEdNter) lacking the high-affinity menin-binding motif failed to transform hematopoietic progenitors in a myeloid progenitor transformation assay (Figure 1A), in which successful transformation is represented by vigorous colony-forming activity in the third and fourth rounds of plating, and elevated expression of *Hoxa9* in first-round colonies (Figure 1B). An artificial MLL-ENL fusion protein (PME), in which high-affinity menin-binding motif is replaced by the PWWP domain of LEDGF, transformed myeloid progenitors despite its inability to form the trimeric complex (7). Therefore, the PWWP domain is

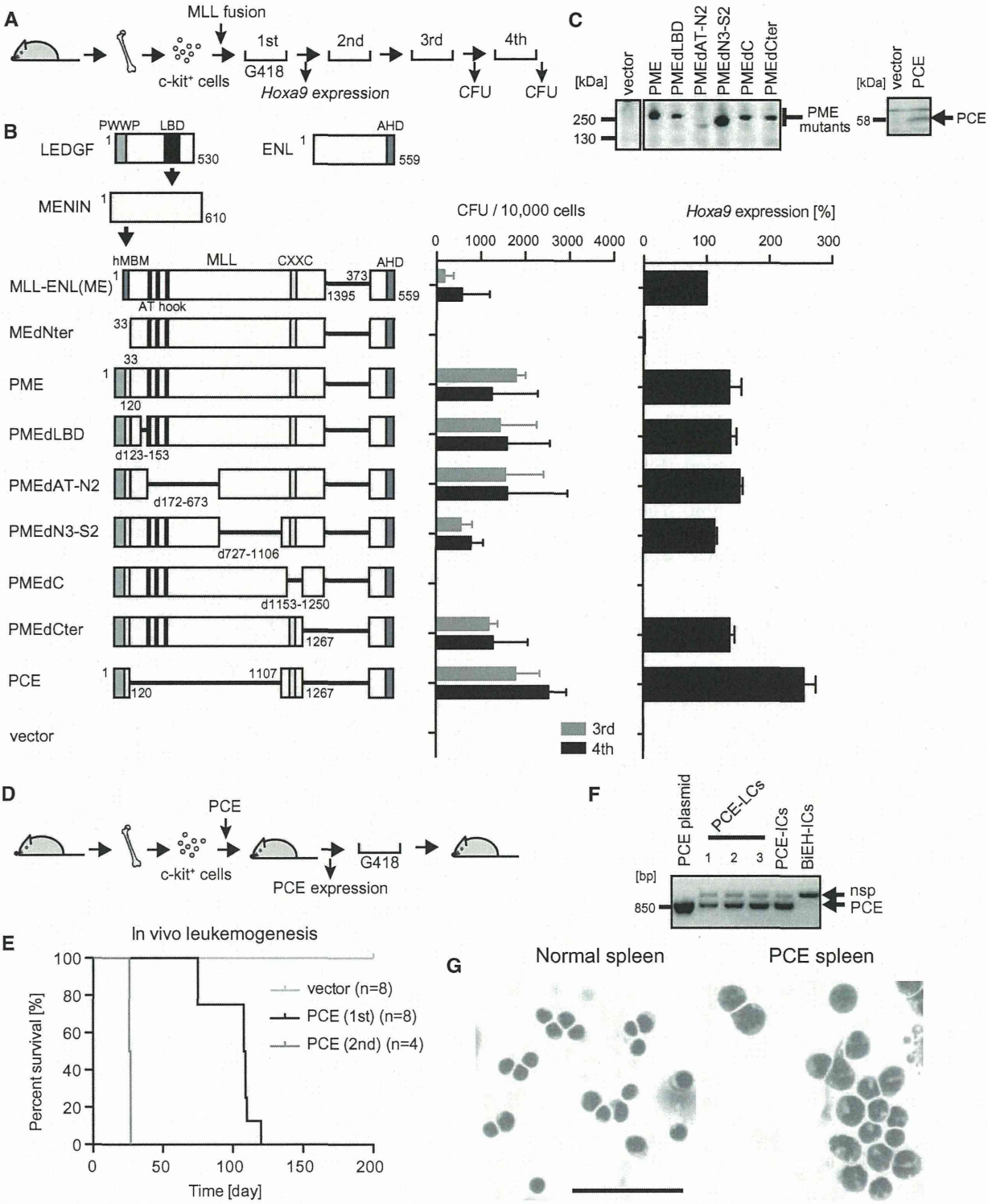

Figure 1. Major functional domains required for leukemogenesis by MLL-ENL. (**A**) Experimental scheme for the myeloid progenitor transformation assay. (**B**) Transforming ability of various MLL-ENL (ME) mutants. The schematic structures of LEDGF, MENIN, ENL and various ME mutants are shown, with the key functional modules indicated (left). P: the PWWP domain. C: the CXXC domain. E: the ENL portion included in MLL-ENL. The numbers of colony-forming units (CFUs) at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression is presented relative to the value of ME (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**C**) Protein expression of the PWWP-MLL-ENL (PME) and PCE mutants in the packaging cells. The PME and PCE proteins were visualized using the anti-ENL antibody. Molecular standards are shown on the left. (**D**) Experimental scheme for the *in vivo* leukemogenesis assay. (**E**) Survival of the transplanted

(continued)

the single functional module within menin and LEDGF that is required for MLL-ENL-dependent transformation.

To identify the structural requirements of MLL fusion proteins besides the PWWP domain, we examined the transforming ability of a series of PME mutants with various internal deletions in the *MLL* gene. A mutant lacking the CXXC domain (PMEdC) failed to transform myeloid progenitors (Figure 1B and C). Conversely, a mutant (PCE) consisting only of the PWWP domain, the CXXC domain and the ENL portion activated *Hoxa9* expression and immortalized myeloid progenitors. The PCE mutant induced leukemia in recipient mice within 3–4 months in the *in vivo* leukemogenesis assay (Figure 1D–F). The PCE-induced leukemic blasts had monocytic features (Figure 1G) and elevated expression of posterior *Hoxa* genes and leukemia stem cell signature genes including *Myb*, *Hmgb3* and *Cbx5* (Supplementary Figure S1), which are major characteristics of MLL-associated myeloid leukemias (24). The leukemic blasts recapitulated the same disease in secondary recipients with a much shorter latency period (Figure 1E). These results clearly show that PCE is as functional as MLL-ENL for leukemic transformation. Thus, the PWWP domain, the CXXC domain and the ENL portion are the major structural elements required for MLL-associated leukemogenesis.

Binding to a specific nucleosome through the PWWP domain is essential for MLL fusion-dependent gene activation

To further analyze the function of the PWWP domain, we next tested a series of artificial PCE derivatives with various alterations in the PWWP domain for the transforming ability of myeloid progenitors (Figure 2A and B). A mutant (P′CE) with the minimum structure of the PWWP domain exhibited this transformation ability, whereas an alanine substitution of the evolutionarily conserved tryptophan (W21) resulted in a loss of the transformation activity. The PWWP domain of LEDGF has been implicated in chromatin binding (25,26). To examine its chromatin-binding capacity, we transiently expressed LEDGF and its mutant (dPWWP) lacking the PWWP domain in 293T cells (Figure 2C), and we divided the cellular components into three subfractions (Figure 2D and Supplementary Figure S2A). In this procedure, (i) chromatin-unbound soluble materials were extracted by cytoskeletal (CSK) buffer (the soluble fraction); (ii) chromatin and its associated factors were eluted by DNA fragmentation with micrococcal nuclease (MNase) (the nucleosome fraction); and (iii) insoluble materials after MNase treatment were solubilized by denaturing detergent (the nuclear matrix fraction). Nearly all the nucleosomes in the nucleosome fraction were mononucleosomes based on their DNA length (~150 bp) (Figure 2E, top panels). LEDGF substantially localized to the nucleosome fraction (Figure 2E, bottom panels) and was able to pull down mononucleosomes (Figure 2F), whereas dPWWP mainly localized to the soluble fraction (Figure 2E) and failed to pull down nucleosomes (Figure 2F). The PWWP domain (P′) by itself localized in the nucleosome fraction (Figure 2G and H) and pulled down endogenous nucleosomes (Figure 2I). Previously, one group showed that LEDGF had specifically associated with the histone H3 peptide containing either H3K36me2 or H3K36me3 (25), while another group showed that the recombinant PWWP domain of LEDGF directly binds to H3K36me3 but not to H3K36me2 *in vitro* (26), suggesting that the peptide pull-down assay tends to overrepresent H3K36me3–PWWP interaction, depending on the assay conditions. Our nucleosome co-immunoprecipitation (co-IP) assay, which can detect protein–nucleosome interaction under physiological conditions, showed that the nucleosomes co-precipitated with P′ contained high amounts of not only H3K36me3 but also H3K36me2 (Figure 2I), indicating that the PWWP domain of LEDGF is capable of binding to either H3K36me2 or H3K36me3 *in vivo*. In accordance with these results, an artificial protein (P′C) composed of the PWWP domain and the CXXC domain was substantially localized in the nucleosome fraction (Figures 2G and H) and co-precipitated with mononucleosomes and DNAs (Supplementary Figure S2B and C). The W21A substitution mutants of P′ and P′C exclusively localized to the soluble fraction (Figures 2G and H) and therefore were not used for the nucleosome co-IP analysis (Supplementary Figures S2D and E). These results indicate that the PWWP domain, and not the CXXC domain, is mainly responsible for stable association with chromatin. The purified P′C–nucleosome complex contained no other proteins but histones at the stoichiometric level (Supplementary Figure S2B). Complete removal of the DNAs by DNase I treatment did not abolish the P′C–nucleosome interaction (Figure 2J) or P′–nucleosome interaction (Supplementary Figure S2F). Thus, the PWWP domain directly associates with histones. Taken together, these findings show that the binding ability of the PWWP domain to specific histones is required for MLL fusion-dependent gene activation.

The PWWP domain and the CXXC domain target transcriptionally active promoters containing either H3K36me2 or H3K36me3

We next examined the functional similarities among various PWWP domains. HRP2 is a close homolog of LEDGF and has a highly homologous PWWP domain (Figure 3A). The bromodomain and PHD finger-containing 1 (BRPF1) protein is an associated

Figure 1. Continued
animals in the *in vivo* leukemogenesis assay. Light gray, vector; black, first PCE transplantation; dark gray, second PCE transplantation. *n*, number of animals analyzed. (**F**) Expression of PCE in the PCE-induced leukemia cells (PCE-LCs). Leukemic cells derived from moribund mice were cultured *ex vivo* in the presence of G418 and analyzed by RT-PCR. The PCE plasmid, PCE-immortalized cells (PCE-ICs) and Bcl2/E2A-HLF-immortalized cells (BiEH-ICs) were included for comparison. nsp, non-specific band. (**G**) Morphology of PCE-LCs in the spleen. The splenic cells harvested from a PCE-leukemic mouse were stained by the May-Grunwald/Giemsa staining method. Normal spleen is included for comparison. Scale bar, 50 μm.