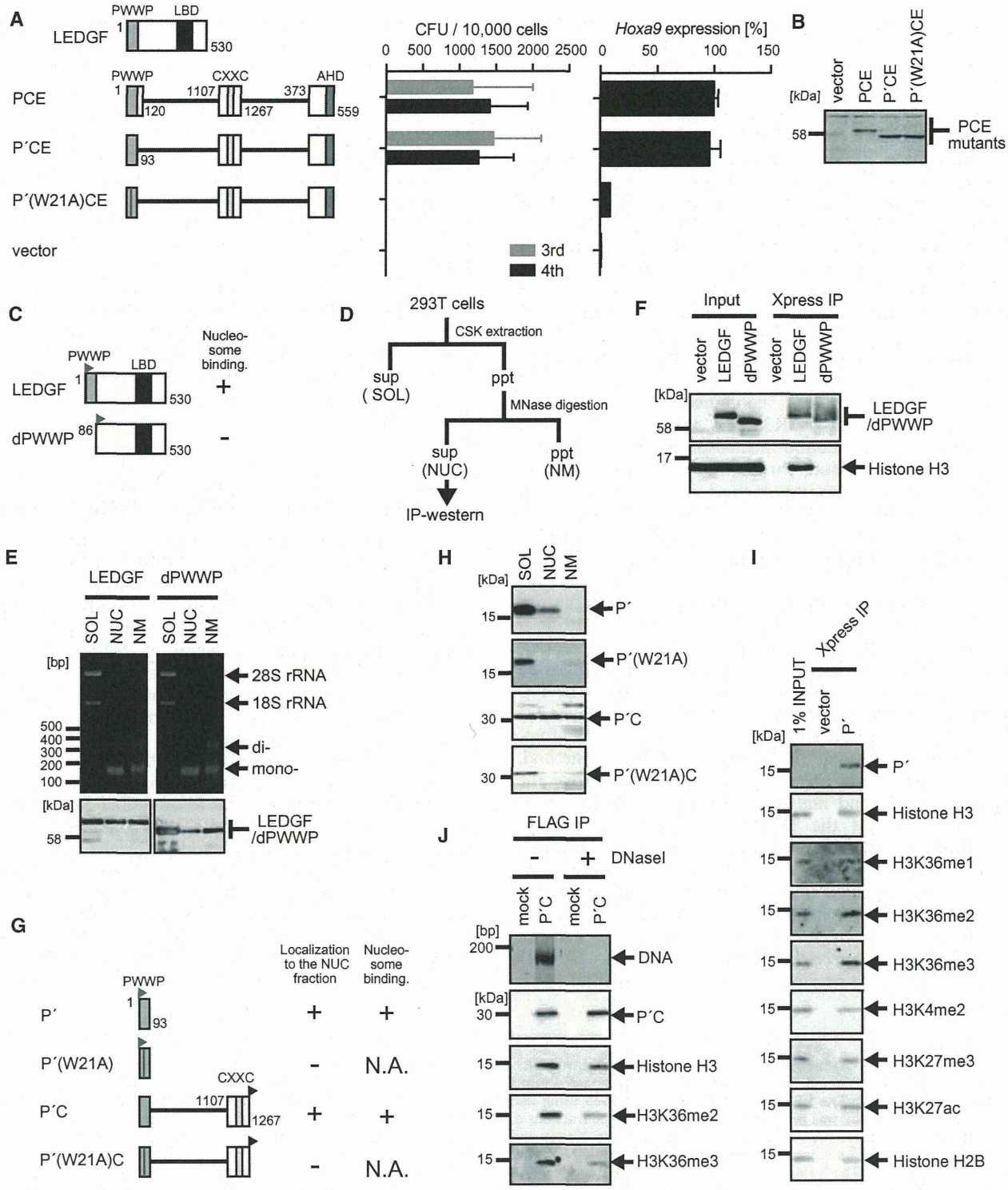

Figure 2. Nucleosome binding by the PWWP domain is essential for MLL-ENL-dependent transformation. (**A**) Transforming ability of PCE mutants containing various mutations within the PWWP domain. P′: the minimum structure of the PWWP domain. The schematic structures of LEDGF and various PCE mutants are shown (left). The CFUs at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression in the first-round colonies is expressed relative to that of PCE (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**B**) Protein expression of the PCE mutants in the packaging cells. The PCE mutant proteins were visualized using the anti-ENL antibody. (**C**) The schematic structures of LEDGF and its mutant lacking the PWWP domain (dPWWP). An Xpress tag (gray flag) is fused to the N-terminal end of the LEDGF proteins. Their ability to associate with nucleosomes is summarized. (**D**) Experimental

(continued)

factor of the MOZ histone acetyltransferase (27) and has a less conserved PWWP domain, which has been recently shown to directly bind to histone H3 peptide containing H3K36me2 or H3K36me3 *in vitro* (28). To investigate the functional differences and similarities of the PWWP domains of LEDGF, HRP2 and BRPF1, we tested whether these PWWP domains are interchangeable in the myeloid progenitor transformation assay. The P′CE mutants whose PWWP domain is replaced by those of HRP2 and LEDGF successfully activated *Hoxa9* expression and immortalized myeloid progenitors (Figure 3A and B). Nucleosome co-IP assay with the P′C mutants containing the PWWP domains of LEDGF and BRPF1 specifically enriched nucleosomes containing H3K36me2 more efficiently than those containing H3K36me3 (Figure 3C and D). Consistent with these results, mass spectrometry analysis indicated that two-thirds of P′C-bound nucleosomes contained the H3K36me2 marker, while the H3K36me3 marker was minor compared with P′-bound nucleosomes (Supplementary Figure S3). In addition, chromatin modifications characteristic to active promoters such as di-methylated histone H3 lysine 4 (H3K4me2) (Figures 2I and 3D), acetylated histone H3 lysine 9 (H3K9ac) and acetylated histone H3 lysine 27 (H3K27ac) (Supplementary Figure S3) were enriched in P′C-bound nucleosomes, suggesting that the CXXC domain confers affinity toward active promoters. To determine the targeting ability of these artificial proteins to known MLL target genes *in vivo*, we performed chromatin immunoprecipitation (ChIP) followed by qPCR on 293T cells transiently expressing those P′C mutants (Figure 3C and E). To this end, we devised a highly sensitive native ChIP protocol by altering the aforementioned nucleosome IP protocol. In this ChIP method, we resuspended the cells in CSK buffer to remove the chromatin-unbound materials and treated the cells with MNase to digest the DNAs to achieve the average length of 500–1000 bp. The highly solubilized chromatin fraction was extracted with lysis buffer and then subjected to IP. We designated this protocol fractionation-assisted native ChIP, but hereafter it is referred to as ChIP. *Ecotropic viral integration site 1* (*EVI1*), *paired-like homeodomain transcription factor 2* (*PITX2*), *HOXA7* and *HOXA9* were previously reported as potential MLL target genes (14,28–30). Although the chromatin environment in 293T cells might be different from that in leukemia cells, the P′C mutants, but not P′(W21A)C, were localized at the transcription start sites (TSSs) of *PITX2*, *HOXA7* and *HOXA9*, but not *EVI1* (Figure 3E). These genes showed enrichment of H3K4me2 near the TSS, consistent with the observation that P′C-bound nucleosomes contain high levels of H3K4me2 modification (Figure 3D–F). H3K36me2 was present in all of the P′C mutant-occupied loci, while H3K36me3 was only found at the *HOXA9* locus (Figure 3F), supporting the notion that the PWWP domain binds either H3K36me2 or H3K36me3 *in vivo*. The presence of RNA polymerase II (RNAPII) indicated that *PITX2*, *HOXA7* and *HOXA9*, but not *EVI1*, were transcriptionally active in 293T cells (Figure 3F). Thus, the PWWP domain and the CXXC domain collaboratively target the promoter-proximal regions of transcriptionally active genes containing either H3K36me2 or H3K36me3.

The chromatin context comprising H3K36me2/3 and non-methylated CpGs is targeted by MLL fusion proteins

Next, we examined the structural requirements of the CXXC domain. A deletion mutant (P′C′E) that lacks the sequences flanking the CXXC domain successfully transformed myeloid progenitors (Figure 4A and B). Therefore, the CXXC domain, but not its flanking sequences, is required for the proper targeting of MLL-ENL. The CXXC domain of MLL directly binds to non-methylated CpGs (31–34). Point mutations that have been proved to abolish the binding ability to non-methylated CpGs (C1155A and K1186A) (31) resulted in a loss of transforming activity (Figure 4A and B), while the Q1162A substitution, which has no effect on non-methylated CpG-specific binding, did not compromise transformation. Mutations in the KFGG motif (K1178A and K1179A) that severely attenuate the binding ability to non-methylated CpGs (31) also abolished transforming activity. The artificial protein (P′C′) composed of the minimum structures of the PWWP domain and the CXXC domain localized in the nucleosome fraction (Figure 4C and D) and associated with nucleosomes (Figure 4E). The P′C′ mutants carrying CpG binding-deficient mutations retained the nucleosome

Figure 2. Continued
scheme of subcellular fractionation. (**E**) Distribution of DNAs/RNAs (top) and LEDGF proteins (bottom) in the three subfractions. 293T cells transiently expressing the indicated proteins were subfractionated into the soluble fraction (SOL), the nucleosome fraction (NUC) and the nuclear matrix fraction (NM). The LEDGF proteins were visualized using anti-Xpress antibody. (**F**) Nucleosomes associated with LEDGF and dPWWP. Nucleosome co-IP assay was performed with the nucleosome fractions of 293T cells transiently expressing the LEDGF proteins. The LEDGF mutant proteins and histone H3 proteins were visualized using anti-Xpress and histone H3 antibodies, respectively. (**G**) The schematic structures of the P′ and the P′C mutants. An Xpress tag (gray flag) is fused to the N-terminal end of P′ mutants. A FLAG tag (black flag) is fused to the C-terminal end of the P′C mutants. The localization to the nucleosome fraction of each mutant and its ability to bind nucleosomes are summarized. N.A., not applicable. (**H**) Subcellular distribution of the P′ and P′C mutants. Experiments were performed as in (E). The P′ mutants were visualized by using the anti-Xpress antibody. The P′C mutants were visualized by using the anti-FLAG antibody. (**I**) Nucleosomes associated with P′. P′ and its associating nucleosomes were co-purified by co-IP from the NUC fraction of 293T cells transiently expressing P′ as in (F). Nucleosomes were visualized by antibodies specific for mono-methylated histone H3 lysine 36 (H3K36me1), H3K36me2 and H3K36me3 or by anti-histone H2B and -H3 antibodies. Di-methylated histone H3 lysine 4 (H3K4me2), tri-methylated histone H3 lysine 27 (H3K27me3) and acetylated histone H3 lysine 27 (H3K27ac) were also analyzed for comparison. (**J**) The role of DNAs in P′C–nucleosome interaction. P′C and its associating nucleosomes were co-purified by co-IP using the anti-FLAG antibody from the NUC fraction of 293T cells expressing P′C. The precipitates were treated with or without DNase I (300 kU/ml) for 10 min and washed to remove unbound material. DNAs co-precipitated with P′C were visualized by SYBR Green staining. Nucleosomes were visualized by antibodies specific for H3K36me2 and H3K36me3 modification or by the anti-histone-H3 antibody.

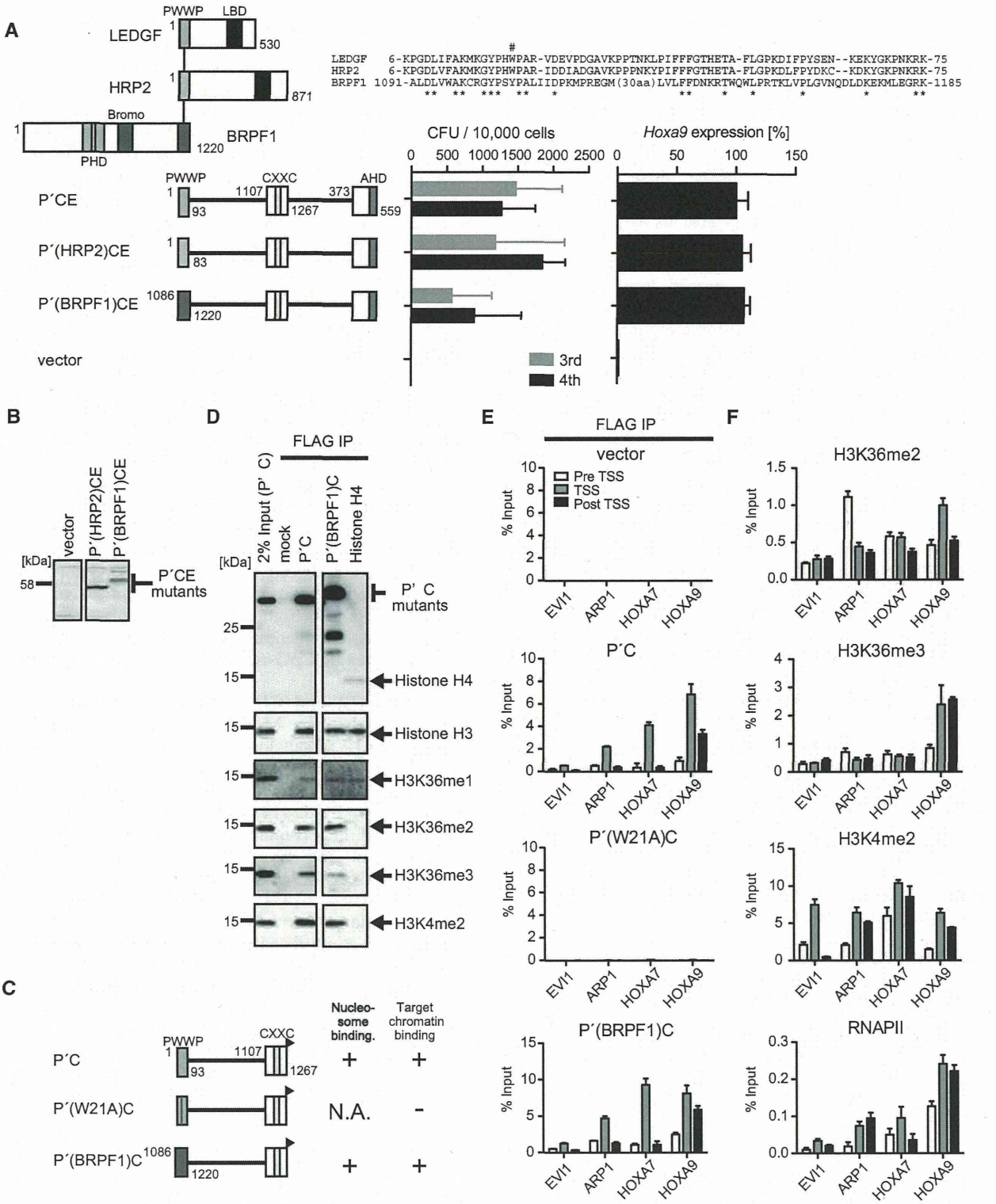

Figure 3. The PWWP domain and the CXXC domain target active promoters containing H3K36me2/3 *in vivo*. (**A**) Transforming ability of P′CE mutants containing three different PWWP domains. The schematic structures of LEDGF, HRP2, BRPF1 and various P′CE mutants are shown (left). Sequence alignment of the three PWWP domains is shown. Asterisk, conserved residue. The tryptophan residue mutated in Figure 2A is indicated by #. The CFUs at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression in the first-round colonies is expressed relative to that of P′CE (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**B**) Protein expression of the P′CE mutants carrying non-LEDGF PWWP domains in the packaging cells. The P′CE mutant proteins were visualized using the

(continued)

binding activity because they had an intact PWWP domain (Figure 4C–E). To investigate the CpG methylation status of the MLL target genes, we performed CIRA using the recombinant CXXC domain protein followed by qPCR, which showed that *PITX2*, *HOXA7* and *HOXA9* loci contained ample amounts of non-methylated CpGs near the TSSs in 293T cells (Figure 4F). ChIP-qPCR analysis of 293T cells transiently expressing these P′C′ mutants showed that only the wild-type control, not the CpG binding-deficient mutants, was able to associate with the promoters of MLL target genes (Figure 4G). Thus, binding to non-methylated CpGs through the CXXC domain is essential for the proper targeting of MLL fusion proteins.

There are dozens of CXXC domain-containing proteins in mammals. To examine the functional similarity of different CXXC domains, we generated several domain swap mutants of P′C′E whose CXXC domain is replaced by other CXXC domains (Figure 5A and B). The CXXC domains of MLL2 (also known as HRX2, MLL4 and KMT2B) (35), DNA (cytosine-5)-methyltransferase 1 (DNMT1) (36) and F-box and leucine-rich repeat protein 11 (FBXL11; also known as KDM2A) (37) were previously shown to bind to non-methylated CpGs. P′C′E mutants carrying these three non-MLL CXXC domains were competent for leukemic transformation (Figure 5A and B). Furthermore, the P′C′ mutant with the CXXC domain of FBXL11 (Figure 5C and Supplementary Figure S4) was able to associate with MLL target genes in 293T cells (Figure 5D). These results indicate that the three different CXXC domains are functionally equivalent in terms of MLL-dependent gene activation, confirming the notion that the binding ability to non-methylated CpGs, a common function of those three CXXC domains, is required for targeting of MLL fusion proteins.

Taken together, the minimum functional modules required for the targeting of MLL fusion protein are the PWWP domain and the CXXC domain, which bind to nucleosomes with H3K36me2/3 and non-methylated CpGs, respectively. Therefore, the MLL fusion complex targets a chromatin context in which H3K36me2/3 and non-methylated CpGs coexist.

The MLL-AF6 complex localizes at transcriptionally active promoters containing H3K36me2/3 and non-methylated CpGs *in vivo*

To determine the genomic landscape of the MLL target genes under physiological conditions, we performed a genome-wide ChIP analysis of cells that express an MLL fusion protein endogenously. ML-2 cells express MLL-AF6, but not wild-type MLL (38,39); therefore, the ChIP signal obtained from ML-2 cells using the anti-MLL antibody can be ascribed to the presence of MLL-AF6. Our ChIP analysis followed by deep sequencing (ChIP-seq) using the anti-MLL antibody identified 154 TSSs of 131 MLL-AF6 target genes (Supplementary Table S1) by virtue of enriched ChIP signals of MLL-AF6 near the TSSs. These regions included previously identified MLL target genes such as *HOXA9*, *HOXA10*, *cyclin-dependent kinase inhibitor 1B* (*CDKN1B*), *CDKN2C*, *transcription factor 4* (*TCF4*) and *runt-related transcription factor 2* (*RUNX2*) (14,40), and some newly identified genes such as *special AT-rich sequence-binding protein-1* (*SATB1*), *ribosomal protein L31* (*RPL31*) and *RPL10A* (Supplementary Figure S5A and Supplementary Table S1). Gene ontology analysis of these MLL-AF6 target genes indicates that MLL-AF6 mainly associates with genes involved in gene expression and translation (Supplementary Figure S5B). The average distribution of MLL-AF6 at the MLL target loci indicates that MLL-AF6 localizes near the TSSs (Figure 6A). MLL target loci contain a relatively large amount of CpG sequence near the TSSs. CIRA and MIRA followed by deep sequencing (CIRA-seq and MIRA-seq respectively) showed that MLL target loci are rich with non-methylated CpGs but are devoid of methylated CpGs. MLL-AF6-occupied loci also exhibited an enrichment of RNAPII and H3K36me3 (Figure 6A and Supplementary Figure S5C). Thus, endogenously expressed MLL-AF6 specifically localizes at CpG-rich promoters of transcriptionally active genes.

ChIP-qPCR and CIRA-qPCR analyses on each gene confirmed that the MLL-AF6 complex localized at the major MLL target loci including *HOXA7*, *CDKN2C*, *CDKN1B* and *HOXA9* but was completely absent at the gene loci of *PITX2* and *MEIS1*, which are potentially regulated by MLL (14) but are transcriptionally inactive in this cell line (Figure 6B and C and Supplementary Figure S5D and E). MLL-AF6, menin and LEDGF co-localized near the TSSs of MLL target genes where non-methylated CpGs are abundant (Figure 6B and Supplementary Figure S5D). H3K36me2 was ubiquitously present and relatively enriched at the transcriptionally active loci. The MLL-AF6 complex localized at the promoter-proximal regions of *HOXA7* and *CDKN2C*, which were H3K36me2-rich but H3K36me3-poor (Figure 6B and C). On the other hand, the MLL-AF6 complex was highly enriched near the TSS of *HOXA9*, where H3K36me2 was scarce but H3K36me3 was abundant (Figure 6B and C and Supplementary Figure S5D and E). These results confirm

Figure 3. Continued
anti-ENL antibody. (**C**) The schematic structures of various P′C mutants. A FLAG tag (black flag) is fused to the C-terminal end of the P′C mutants. Their abilities to bind nucleosomes and target promoters are summarized. (**D**) Nucleosomes associated with the P′C mutants. The P′C mutants and their associating nucleosomes were co-purified by co-IP using the anti-FLAG antibody from the NUC fraction of 293T cells transiently expressing the P′C mutants. Nucleosomes were visualized by antibodies specific for histone H3 or the indicated modifications. Nucleosomes co-purified with FLAG-tagged histone H4 were also analyzed for comparison. (**E**) Chromatin targeting ability of the P′C mutants. Genomic localization of the P′C mutants was analyzed by ChIP using the anti-FLAG antibody in 293T cells transiently expressing the P′C mutants. The precipitated DNAs were analyzed by qPCR using specific probes for pre-TSS (−1.0 to −0.5 kb of TSS), TSS (0 to +0.5 kb of TSS) and post-TSS (+1.0 to 1.5 kb of TSS) of the indicated genes. (**F**) Epigenetic status of MLL target genes in 293T cells. ChIP analysis was performed on 293T cells using antibodies specific for H3K36me2, H3K36me3, H3K4me2 and RNAPII. The precipitated DNAs were analyzed by qPCR as in (E).

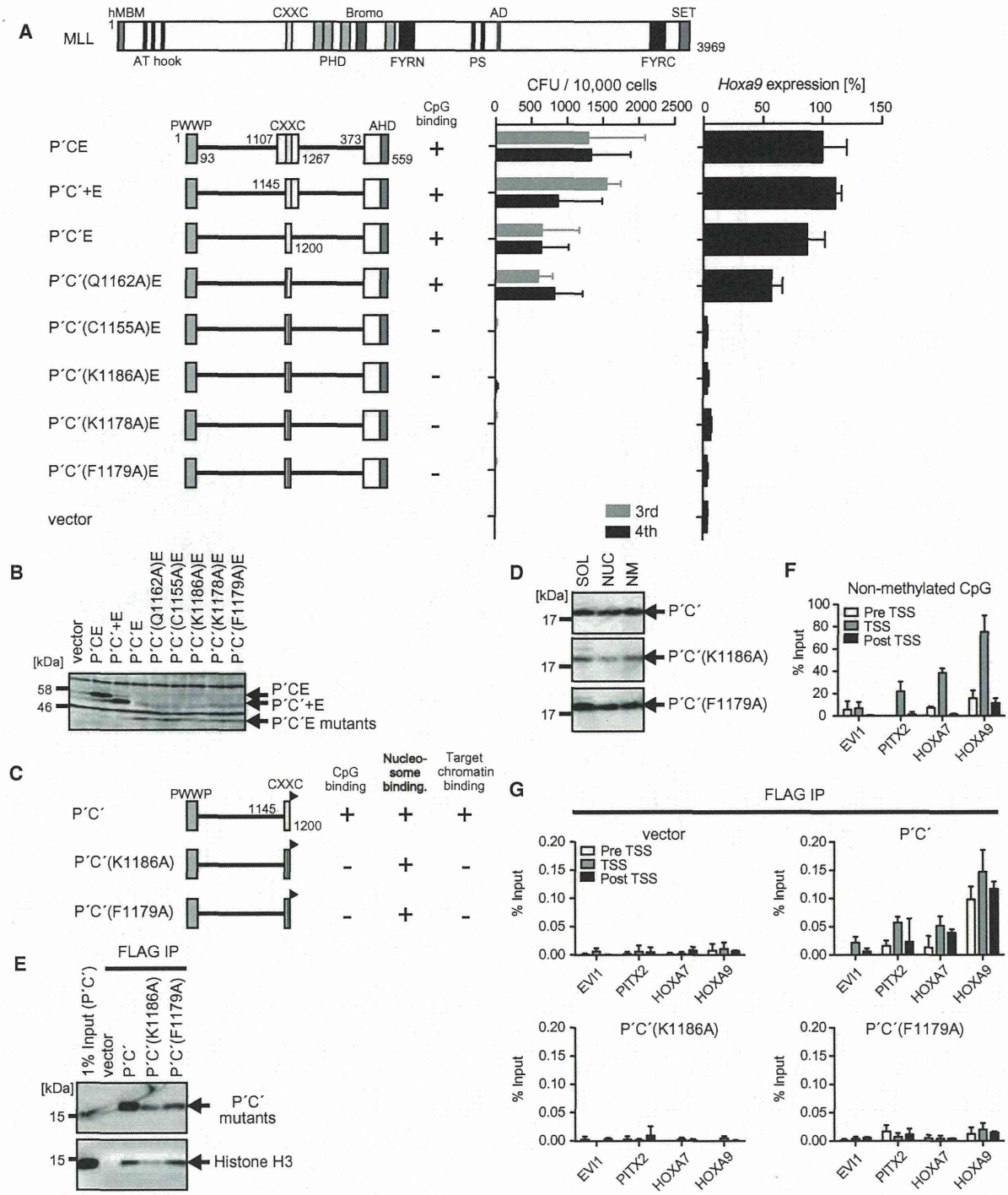

Figure 4. MLL-ENL recognizes the chromatin context with H3K36me2/3 and non-methylated CpGs. (**A**) Transforming ability of P′CE mutants with various mutations within the CXXC domain. The schematic structures of MLL and various P′CE mutants are shown (left). C′: the minimum structure of the CXXC domain; C′+, the minimum structure of the CXXC domain and its C-terminal flanking region. Their ability to bind to non-methylated CpGs is summarized. The CFUs at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression in the first-round colonies is expressed relative to the value of P′CE (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**B**) Protein expression of the P′CE mutants in the packaging cells. The P′CE mutant proteins were visualized by the anti-ENL antibody. (**C**) The schematic structures of various P′C′ mutants. A FLAG tag (black flag) is fused to the C-terminal end of the P′C′ mutants. Their abilities to bind non-methylated CpGs, nucleosomes and target promoters are summarized. (**D**) Subcellular distribution of the P′C′ mutants. Experiments were performed as in Figure 2E. The P′C′ mutants were visualized by the anti-FLAG antibody. (**E**) Nucleosomes associated with the P′C′ mutants. The P′C mutants and their associating nucleosomes were analyzed as in Figure 2F. (**F**) Distribution of non-methylated CpGs at the MLL target genes. Genomic DNAs of 293T cells were analyzed by CIRA. The precipitated DNAs were analyzed by qPCR as in Figure 3E. (**G**) Chromatin targeting ability of the P′C mutants. Genomic localization of the P′C′ mutants was analyzed by ChIP as in Figure 3E.

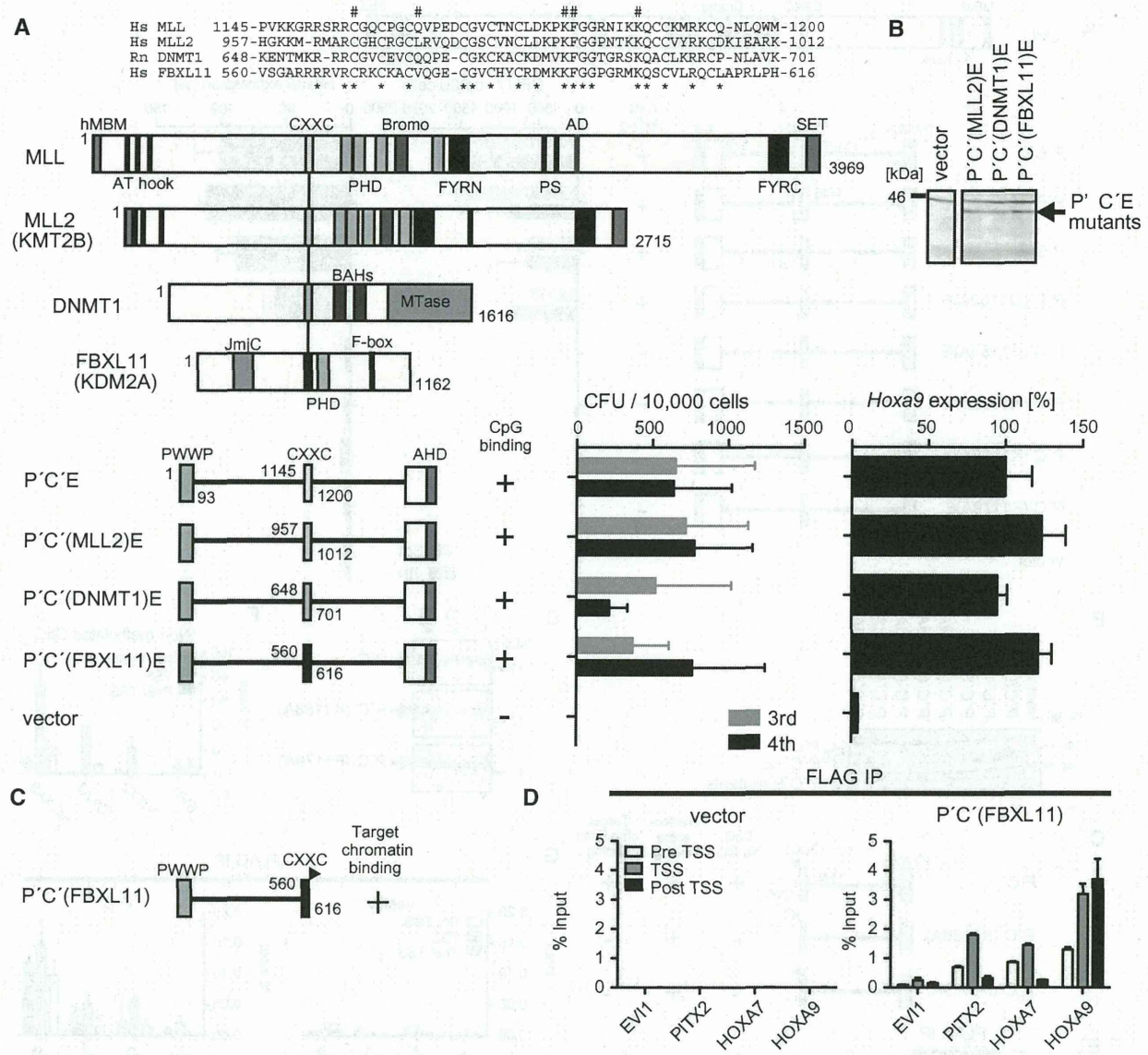

Figure 5. Non-MLL CXXC domains can functionally substitute for the MLL CXXC domain. (**A**) Transforming ability of P′C′E mutants with various non-MLL CXXC domains. The schematic structures of MLL, MLL2, DNMT1, FBXL11 and various P′C′E mutants are shown (left). Sequence alignment of the CXXC domains of MLL, MLL2, DNMT1 and FBXL11. The mutated residues in Figure 4A are indicated by #. Asterisk, conserved residue. The CFUs at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression in the first-round colonies is expressed relative to the value of P′C′E (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**B**) Protein expression of the P′C′E mutants carrying non-MLL CXXC domains in the packaging cells. The P′C′E mutant proteins were visualized by the anti-ENL antibody. (**C**) The schematic structure of the P′C′ mutant carrying the CXXC domain of FBXL11 [P′C′(FBXL11)]. A FLAG tag is fused to the C-terminal end of the CXXC domain (black flag). Its ability to associate with the target promoters is indicated. (**D**) Chromatin targeting ability of P′C′(FBXL11). Genomic localization of P′C′(FBXL11) was analyzed by ChIP as in Figure 3E.

the notion that the presence of either H3K36me2 or H3K36me3 is required for targeting of the MLL fusion complex. Hence, our data strongly indicate that the MLL-fusion complex targets the promoter-proximal regions containing H3K36me2/3 and non-methylated CpGs under physiological conditions.

It should be noted that the MLL fusion complex was also found by ChIP-qPCR at a housekeeping gene, *glyceraldehyde phosphate dehydrogenase* (*GAPDH*), whose promoter-proximal region also contained H3K36me2/3 and non-methylated CpGs near the TSS (Figure 6B and Supplementary Figure S5D). However, the localization of the MLL-AF6 complex was much weaker than that at well-known MLL target genes. These results suggest that the extent of occupation by the MLL fusion complex is influenced not only by the

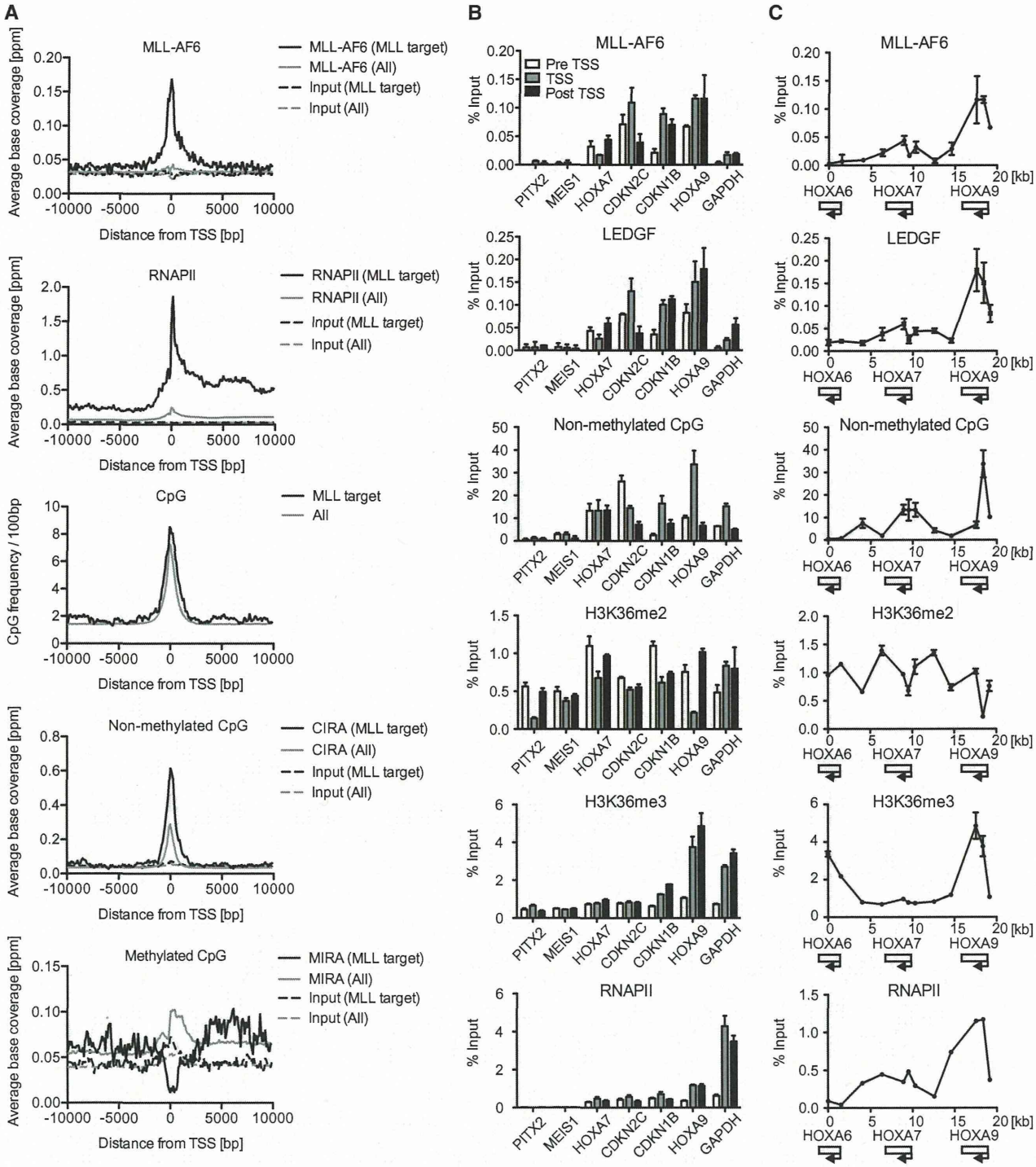

Figure 6. Genomic landscape of MLL-AF6 target genes in ML-2 cells. (**A**) The average distribution of MLL-AF6, RNAPII, CpG dinucleotide, non-methylated CpGs and methylated CpGs at the 154 MLL-AF6-occupied TSSs. The input data and all TSS data are included for comparison. ChIP-seq or CIRA/MIRA-seq tags were clustered into 100-bp bins. On the *x* axis, the position of the associated TSS is designated as 0 bp. The *y* axis indicates the normalized ChIP-seq tag count (ppm). The average number of CpG dinucleotide in each bin was also plotted. (**B**) Genomic localization of MLL-AF6, LEDGF, non-methylated CpGs, H3K36me2, H3K36me3 and RNAPII in ML-2 cells at various gene loci. Genomic localization was determined by ChIP-qPCR or CIRA-qPCR. The precipitated DNAs were analyzed using the specific probes for pre-TSS (−1.0 to −0.5 kb of TSS), TSS (0 to +0.5 kb of TSS) and post-TSS (+1.0 to 1.5 kb of TSS) of the indicated genes. (**C**) Genomic localization of MLL-AF6, LEDGF, non-methylated CpGs, H3K36me2, H3K36me3 and RNAPII in ML-2 cells at the posterior *HOXA* loci. Relative locations of the *HOXA6*, *HOXA7* and *HOXA9* loci are shown at the bottom.

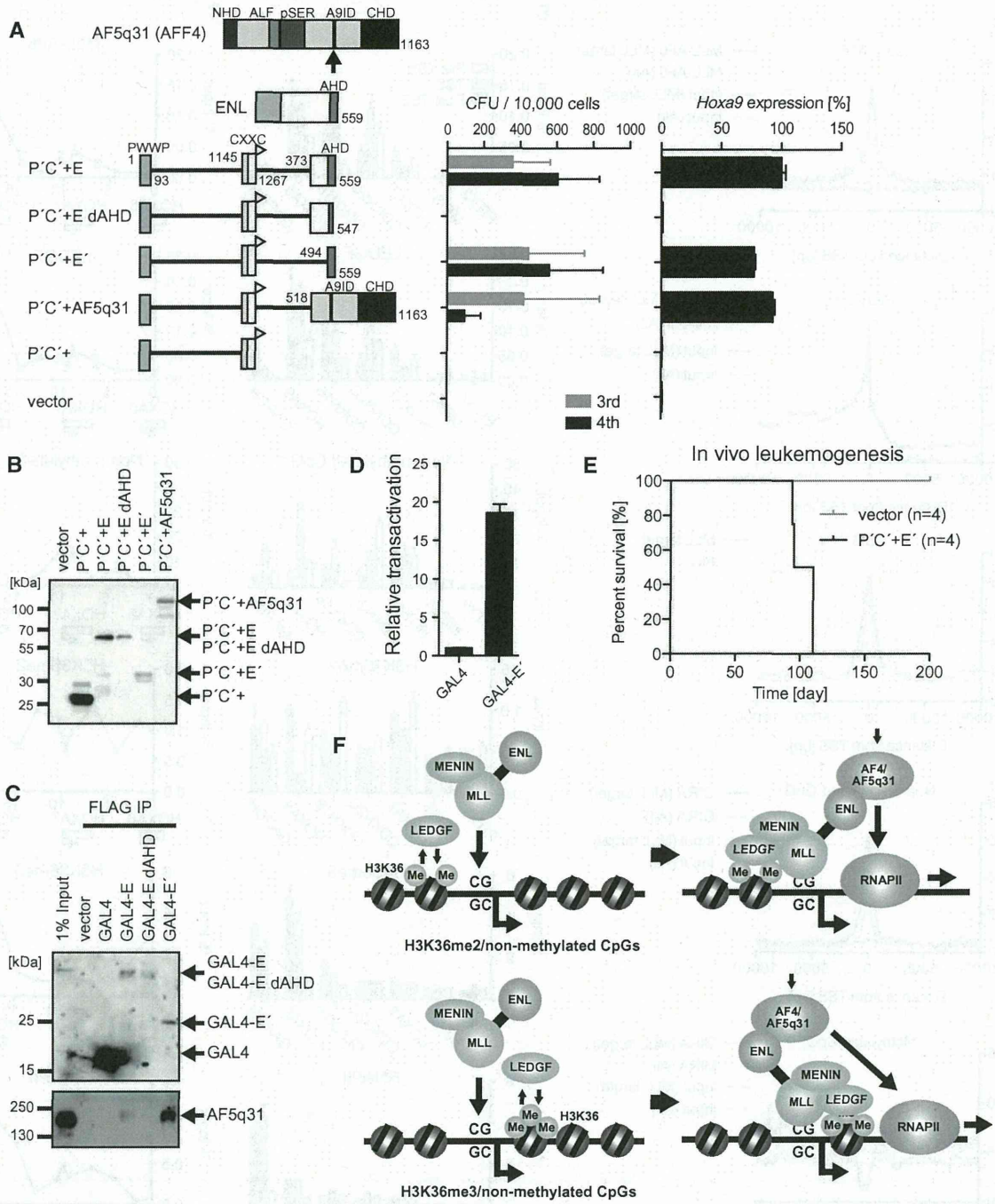

Figure 7. MLL fusion-dependent gene activation is mediated by AF4-family coactivators. (**A**) Transforming ability of P′C′+ mutants fused with ENL and AF5q31. The schematic structures of ENL, AF5q31 and various P′C′+ mutants are shown (left). An HA tag (white flag) is fused to the C-terminal end of the P′C′+ structure. E dAHD:the ENL portion lacking the ANC1 homology domain (AHD). E′: the minimum transformation domain of ENL composed of AHD. The CFUs at the third and fourth rounds of replating are shown with error bars (SD of >3 independent experiments) (middle). *Hoxa9* expression in the first-round colonies is expressed relative to the value of P′C′+E (arbitrarily set at 100%) with error bars (SD of triplicate PCRs) (right). (**B**) Protein expression of the P′C′+ mutants fused with ENL and AF5q31 in the packaging cells. The P′C′+ mutant proteins were visualized by the anti-HA antibody. (**C**) Association of AHD and the AF4-family proteins. A series of FLAG-tagged GAL4-ENL fusion proteins (Supplementary Figure S6) were transiently expressed in 293T cells and subjected to ChIP using the anti-FLAG antibody. The precipitates were analyzed by western blotting using the anti-FLAG and anti-AF5q31 antibodies. (**D**) Transcriptional activation activity of AHD. A FLAG-tagged GAL4 DNA binding domain fused with AHD (GAL4-E′) was transiently expressed in 293T cells with the pFR-luc reporter plasmid (5 GAL4 binding sequences and a TATA box are placed in front of the luciferase gene) and subjected to luciferase assay. Error bars represent the SD of triplicate analyses. (**E**) Survival of the transplanted animals in the *in vivo* leukemogenesis assay. Gray, vector; black, P′C′+E. *n*, number of animals analyzed. (**F**) Models of MLL-ENL-dependent gene activation.

amounts of H3K36me2/3 and non-methylated CpGs but also by other factors or environments that are yet undetermined.

An interaction domain for AF4-family coactivators and the targeting modules are the minimum elements required for MLL-ENL-dependent leukemogenesis

Next, we analyzed the structural requirements of the fusion partner portion. We tested the transforming ability of various fusion partners that were fused with the minimum targeting structure (P′C′+) composed of the PWWP domain and the CXXC domain (the C-terminal flanking sequence of the CXXC domain was included to increase the protein stability). An artificial protein (P′C′+E′) composed of the minimum targeting structure and the ANC1 homology domain (AHD) of ENL activated *Hoxa9* expression and transformed myeloid progenitors, whereas a P′C′+E mutant lacking a part of AHD failed to transform myeloid progenitors (Figure 7A and B). AHD is a binding platform for AF4 family transcriptional coactivators (14). Indeed, the GAL4 DNA binding domain fused with AHD (GAL4-E′) associated with AF5q31 (also known as AFF4 and MCEF) (Figure 7C and Supplementary Figure S6), which is a member of the AF4 family, and exhibited transcriptional activation activity (Figure 7D). Removing AHD from P′C′+E′ resulted in the loss of transforming ability (Figure 7A and B), whereas adding back the AF5q31 portion restored the transforming ability, indicating that association with the AF4-family coactivator is essential for MLL-ENL-dependent transformation. *In vivo* leukemogenesis assay showed that P′C′+E′ induced leukemia *in vivo* (Figure 7E). Therefore, the PWWP domain, the CXXC domain and AHD are the minimum essential structures required for leukemic transformation. Taken together, these results show that MLL fusion proteins target the promoter-proximal regions of transcriptionally active CpG-rich genes through dual recognition of epigenetic markers and activate transcription by recruiting transcriptional coactivators such as AF4-family proteins (Figure 7F).

DISCUSSION

This study demonstrates that MLL fusion proteins target the promoter-proximal regions of previously active genes by simultaneously recognizing H3K36me2/3 and non-methylated CpGs and activate transcription by recruiting transcriptional coactivators. Our structure–function analysis demonstrated that an MLL-ENL mutant composed only of the PWWP domain, the CXXC domain and AHD was able to activate *Hoxa9* expression and transform myeloid progenitors. These results imply that the PWWP domain and the CXXC domain are the only structures essentially required for target recognition by MLL fusion proteins. The PWWP domain specifically binds to either H3K36me2 or H3K36me3, both of which are indicators of active transcription. The CXXC domain binds to non-methylated CpGs, which are clustered in the active gene promoters. *HOX* loci are enriched with CpGs and are therefore preferentially subjected to MLL fusion-dependent gene activation. Our ChIP/CIRA analyses showed that the MLL fusion complex localizes at the promoter-proximal region, where non-methylated CpGs and H3K36me2/3 are present. Thus, dual recognition of the two epigenetic markers is the major determinant for MLL fusion proteins to identify their target genes. AHD is the platform to recruit AF4 family transcriptional coactivators, which confer transcriptional activation activity. The fundamental mechanism of MLL fusion-dependent gene activation is threefold: (i) LEDGF scans the genome and binds to H3K36me2/3 through the PWWP domain. (ii) The MLL fusion/menin complex binds to non-methylated CpGs in the promoter through the CXXC domain and a nearby chromatin-bound LEDGF. These two points of contact allow the MLL fusion complex to stably associate with its target chromatin. (iii) The MLL fusion complex activates transcription by recruiting transcriptional coactivators (Figure 7F). However, it should be noted that these three domains are the minimum elements, not all the elements, involved in MLL fusion-dependent leukemogenesis in the context of a full-length MLL fusion protein. The activity of the MLL fusion complex is likely regulated through intricate regulatory mechanisms. Factors that modulate its function may also play critical roles in MLL fusion-dependent leukemogenesis.

Aberrant self-renewal is an essential feature of oncogene-driven proliferation. It requires the continuous expression of a gene set that is required for self-renewal through rounds of cell divisions; however, its mechanisms are largely unclear. Because the PWWP and CXXC domains are sufficient for target recognition, the MLL fusion complex identifies posterior *HOXA* genes by a chromatin context-dependent mechanism rather than by mediating sequence-specific transcription factors. The H3K36me3 marker is generally deposited into chromatin in a transcriptional elongation-coupled manner by the SETD2 methyltransferase (also known as KMT3A, HYPB and SET2), which associates with actively elongating RNAPII (41–43). The H3K36me2 marker is deposited into chromatin by various SET domain-containing proteins, including NSD2, and is generally linked to transcriptional activation (16). Hence, previously transcribed genes containing H3K36me2/3 and non-methylated CpGs at the promoter-proximal regions are subject to gene activation by MLL fusion proteins. Consequently, the expression of previously transcribed HSC program genes is maintained at high levels through rounds of cell division to cause aberrant self-renewal.

In summary, our study identifies the minimum structural requirements of MLL fusion proteins for *HOX* gene activation and reveals the chromatin context recognized by them. A surprisingly simple combination of epigenetic markers that defines previously active gene promoters is used in MLL fusion-dependent cellular memory maintenance. Thus, our results provide significant insights into the molecular mechanisms of cellular memory maintenance and aberrant self-renewal caused by MLL fusion proteins.

ACCESSION NUMBERS

We generated ChIP-seq and CIRA/MIRA-seq data in this study. Mass data obtained in this study have been deposited under accession numbers DRA000555, DRA000557–DRA000561 in the DDBJ (DNA Data Bank of Japan) Sequence Read Archive.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Chikako Hatanaka, Yukari Jindai, Liu Kehong and Dr. Takahiro Kihara for technical assistance; Dr. Toshio Kitamura for providing plat-E cells; and Drs. Hiroshi Handa and Yuki Yamaguchi for discussion and critical reading of the manuscript.

FUNDING

Japan Society for the Promotion of Science (KAKENNHI) [23689050, 22118003, 25118511, 25670450 to A.Y., 25870373 to H.O.]. Funding for open access charge: Japan Society for the Promotion of Science (KAKENNHI) [22118003 to A.Y.].

Conflict of interest. A.Y. has research funding from Dainippon Sumitomo Pharma Co., Ltd. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1. Yu,B.D., Hanson,R.D., Hess,J.L., Horning,S.E. and Korsmeyer,S.J. (1998) MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proc. Natl Acad. Sci. USA*, **95**, 10632–10636.
2. Jude,C.D., Climer,L., Xu,D., Artinger,E., Fisher,J.K. and Ernst,P. (2007) Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell*, **1**, 324–337.
3. Yagi,H., Deguchi,K., Aono,A., Tani,Y., Kishimoto,T. and Komori,T. (1998) Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood*, **92**, 108–117.
4. Krivtsov,A.V., Twomey,D., Feng,Z., Stubbs,M.C., Wang,Y., Faber,J., Levine,J.E., Wang,J., Hahn,W.C., Gilliland,D.G. *et al.* (2006) Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature*, **442**, 818–822.
5. Thorsteinsdottir,U., Mamo,A., Kroon,E., Jerome,L., Bijl,J., Lawrence,H.J., Humphries,K. and Sauvageau,G. (2002) Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood*, **99**, 121–129.
6. Ayton,P.M. and Cleary,M.L. (2003) Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. *Genes Dev.*, **17**, 2298–2307.
7. Yokoyama,A. and Cleary,M.L. (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*, **14**, 36–46.
8. Sutherland,H.G., Newton,K., Brownstein,D.G., Holmes,M.C., Kress,C., Semple,C.A. and Bickmore,W.A. (2006) Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations. *Mol. Cell. Biol.*, **26**, 7201–7210.
9. Ciuffi,A. and Bushman,F.D. (2006) Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet.*, **22**, 388–395.
10. Shun,M.C., Raghavendra,N.K., Vandegraaff,N., Daigle,J.E., Hughes,S., Kellam,P., Cherepanov,P. and Engelman,A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.
11. Biswas,D., Milne,T.A., Basrur,V., Kim,J., Elenitoba-Johnson,K.S., Allis,C.D. and Roeder,R.G. (2011) Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proc. Natl Acad. Sci. USA*, **108**, 15751–15756.
12. Lin,C., Smith,E.R., Takahashi,H., Lai,K.C., Martin-Brown,S., Florens,L., Washburn,M.P., Conaway,J.W., Conaway,R.C. and Shilatifard,A. (2010) AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol. Cell*, **37**, 429–437.
13. Mueller,D., Bach,C., Zeisig,D., Garcia-Cuellar,M.P., Monroe,S., Sreekumar,A., Zhou,R., Nesvizhskii,A., Chinnaiyan,A., Hess,J.L. *et al.* (2007) A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, **110**, 4445–4454.
14. Yokoyama,A., Lin,M., Naresh,A., Kitabayashi,I. and Cleary,M.L. (2010) A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell*, **17**, 198–212.
15. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
16. Kuo,A.J., Cheung,P., Chen,K., Zee,B.M., Kioi,M., Lauring,J., Xi,Y., Park,B.H., Shi,X., Garcia,B.A. *et al.* (2011) NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell*, **44**, 609–620.
17. Smith,K.S., Rhee,J.W. and Cleary,M.L. (2002) Transformation of bone marrow B-cell progenitors by E2a-Hlf requires coexpression of Bcl-2. *Mol. Cell. Biol.*, **22**, 7678–7687.
18. Morita,S., Kojima,T. and Kitamura,T. (2000) Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.*, **7**, 1063–1066.
19. Lavau,C., Szilvassy,S.J., Slany,R. and Cleary,M.L. (1997) Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J.*, **16**, 4226–4237.
20. Yokoyama,A., Kawaguchi,Y., Kitabayashi,I., Ohki,M. and Hirai,K. (2001) The conserved domain CR2 of Epstein-Barr virus nuclear antigen leader protein is responsible not only for nuclear matrix association but also for nuclear localization. *Virology*, **279**, 401–413.
21. Yokoyama,A., Wang,Z., Wysocka,J., Sanyal,M., Aufiero,D.J., Kitabayashi,I., Herr,W. and Cleary,M.L. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell. Biol.*, **24**, 5639–5649.
22. Fujinoki,M., Kawamura,T., Toda,T., Ohtake,H., Ishimoda-Takagi,T., Shimizu,N., Yamaoka,S. and Okuno,M. (2003) Identification of 36-kDa flagellar phosphoproteins associated with hamster sperm motility. *J. Biochem.*, **133**, 361–369.
23. Daigo,K., Yamaguchi,N., Kawamura,T., Matsubara,K., Jiang,S., Ohashi,R., Sudou,Y., Kodama,T., Naito,M., Inoue,K. *et al.* (2012) The proteomic profile of circulating pentraxin 3 (PTX3) complex in sepsis demonstrates the interaction with azurocidin 1 and other components of neutrophil extracellular traps. *Mol. Cell. Proteomics*, **11**, M111.015073.
24. Somervaille,T.C., Matheny,C.J., Spencer,G.J., Iwasaki,M., Rinn,J.L., Witten,D.M., Chang,H.Y., Shurtleff,S.A., Downing,J.R. and Cleary,M.L. (2009) Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell*, **4**, 129–140.
25. Daugaard,M., Baude,A., Fugger,K., Povlsen,L.K., Beck,H., Sorensen,C.S., Petersen,N.H., Sorensen,P.H., Lukas,C., Bartek,J. *et al.* (2012) LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.*, **19**, 803–810.