

Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger

Gang G. Wang¹, Jikui Song², Zhanxin Wang², Holger L. Dormann¹, Fabio Casadio¹, Haitao Li², Jun-Li Luo³, Dinshaw J. Patel² & C. David Allis¹

Histone H3 lysine 4 methylation (H3K4me) has been proposed as a critical component in regulating gene expression, epigenetic states, and cellular identities¹. The biological meaning of H3K4me is interpreted by conserved modules including plant homeodomain (PHD) fingers that recognize varied H3K4me states^{1,2}. The dysregulation of PHD fingers has been implicated in several human diseases, including cancers and immune or neurological disorders³. Here we report that fusing an H3K4-trimethylation (H3K4me3)-binding PHD finger, such as the carboxy-terminal PHD finger of PHF23 or JARID1A (also known as KDM5A or RBBP2), to a common fusion partner nucleoporin-98 (NUP98) as identified in human leukaemias^{4,5}, generated potent oncoproteins that arrested haematopoietic differentiation and induced acute myeloid leukaemia in murine models. In these processes, a PHD finger that specifically recognizes H3K4me3/2 marks was essential for leukaemogenesis. Mutations in PHD fingers that abrogated H3K4me3 binding also abolished leukaemic transformation. NUP98–PHD fusion prevented the differentiation-associated removal of H3K4me3 at many loci encoding lineage-specific transcription factors (*Hox(s)*, *Gata3*, *Meis1*, *Eya1* and *Pbx1*), and enforced their active gene transcription in murine haematopoietic stem/progenitor cells. Mechanistically, NUP98–PHD fusions act as ‘chromatin boundary factors’, dominating over polycomb-mediated gene silencing to ‘lock’ developmentally critical loci into an active chromatin state (H3K4me3 with induced histone acetylation), a state that defined leukaemia stem cells. Collectively, our studies represent, to our knowledge, the first report that deregulation of the PHD finger, an ‘effector’ of specific histone modification, perturbs the epigenetic dynamics on developmentally critical loci, catastrophizes cellular fate decision-making, and even causes oncogenesis during mammalian development.

Recent studies have shown that an H3K4me3-binding PHD finger in the human BPTF (also known as NURF301), ING2 or TFIID (also known as TBP) complex helps to recruit and/or stabilize these effectors and associated factors onto appropriate target promoters during transcriptional regulation^{1,6–10}. A non-methylated H3K4 (H3K4me0)-engaging PHD finger in the DNMT3L or LSD1 (also known as AOF2) complex connects the activities of DNA methylation or H3K4 demethylation to repressive chromatin^{11,12}. Notably, germ-line mutation in the PHD finger of RAG2 abrogates its recognition of H3K4me3 and causes immunodeficiency¹³. Mutations in the PHD finger of ING1 have been implicated in cancers^{3,8,14}. However, evidence supporting a causal role for PHD finger mutation and inappropriate interpretation of histone modification in oncogenesis is still elusive.

In clinically reported acute myeloid leukaemia (AML) patients^{4,5}, chromosomal translocation fuses the C-terminal PHD finger of JARID1A or PHF23 (JARID1A–PHD3 or PHF23–PHD), together with

nuclear localization signals, to NUP98, a common leukaemia fusion partner that contains transactivation activities^{15–17} (Supplementary Fig. 1). Notably, the JARID1A–PHD3 motif is excluded from an alternatively spliced isoform of JARID1A and the corresponding NUP98–JARID1A fusion (hereafter referred to as NJS), whereas it is retained in the longer fusion isoform (hereafter referred to as NJL; Fig. 1a). We asked whether JARID1A–PHD3 as a putative chromatin-‘reading’ module is involved in haematopoietic malignancies. To test this, we examined the leukaemogenic potential of both fusion isoforms using a haematopoietic progenitor transformation assay¹⁸ (Supplementary Fig. 2a). Murine bone-marrow-derived haematopoietic stem/progenitor cells transduced with an empty retrovirus or a retrovirus encoding NJS proliferated transiently and differentiated into mature cells, whereas cells transduced with NJL proliferated indefinitely as undifferentiated progenitors (Fig. 1b, c). The NJL-transduced marrow cells proliferated in a cell-autonomous manner, had typical myeloblast morphology (Fig. 1d) and expressed early myeloid progenitor antigens (c-Kit⁺ Cd11b⁺ Cd34⁺ Gr-1[–] Cd19[–] B220^{–/low}; Fig. 1e and Supplementary Fig. 2b). The arrest of myeloid differentiation by NJL indicated that it would induce leukaemia *in vivo*. Indeed, all of the 12 mice transplanted with murine marrow progenitors transduced with NJL died of AML in an average of 69 days, whereas those reconstituted with empty vector- or NJS-transduced progenitors remained healthy after 1 year (Fig. 1f). NJL-induced leukaemia exhibited a myeloid phenotype (Supplementary Fig. 2c, d), and typically presented with an enlarged spleen, packed progenitors in the bone marrow, and a massive increase in peripheral white blood cells (Supplementary Table 1 and Fig. 1g, h). Taken together, NJL represents a potent leukaemia oncogene both in cellular and animal models.

The fact that NJS failed to induce leukaemia indicated that the PHD finger is required for leukaemogenesis. Indeed, deletion of JARID1A–PHD3, but not of JARID1A sequences before or after it, abolished NJL-mediated transformation of haematopoietic cells (Supplementary Fig. 2f–h). We next addressed whether JARID1A–PHD3 recognizes histone methylation. First, only histone H3 associated with recombinant JARID1A–PHD3 using total histone extracts (Supplementary Fig. 3a). When a mini-library of H3 peptides containing unmodified, mono-, di- or tri-methylated Lys 4, Lys 9, Lys 27, Lys 36 or Lys 79 was screened in a biotinylated peptide pull-down assay, JARID1A–PHD3 only interacted with those containing H3K4me3/2 (Fig. 2a and Supplementary Fig. 3b). Such specificity was further confirmed by immunostaining and co-immunoprecipitation using Flag–NJL stable expression cells: NJL exhibited a speckled nuclear staining pattern, and significantly co-localized with H3K4me3 but not with H3K9me3 (Supplementary Fig. 4). Most of the NJL was bound to mononucleosomes containing H3K4me3 but not H3K27me3 (Supplementary Fig. 3c and Fig. 2d). Calorimetry-based measurements

¹Laboratory of Chromatin Biology & Epigenetics, The Rockefeller University, New York, New York 10065, USA. ²Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³Department of Cancer Biology, The Scripps Research Institute, Scripps Florida, Jupiter, Florida 33458, USA.

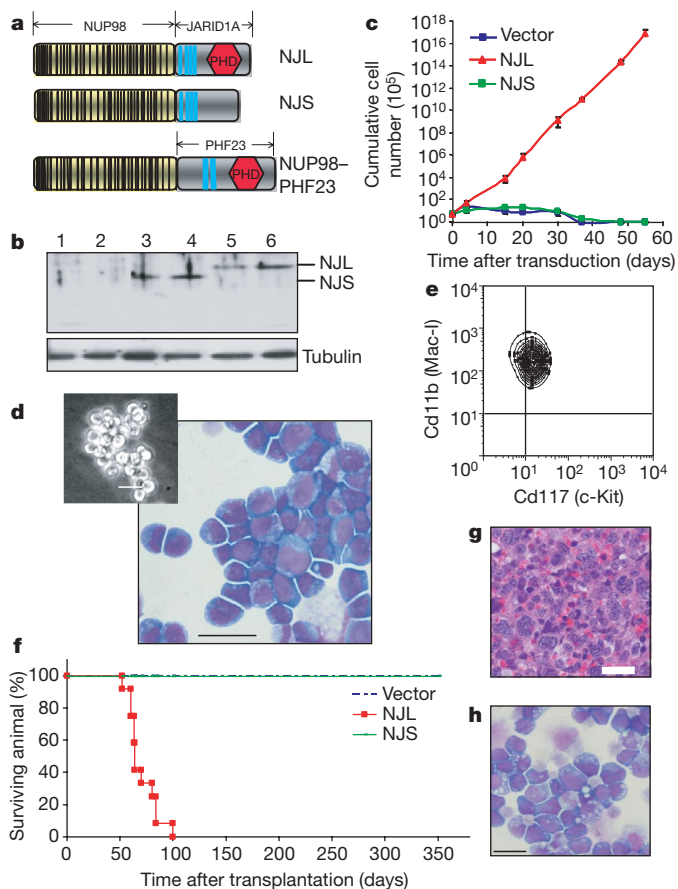


Figure 1 | The PHD finger-containing NUP98-JARID1A fusion isoform (NJL), but not that lacking the PHD finger (NJS), confers leukaemogenic potentials to haematopoietic stem/progenitor cells. **a**, The NUP98-JARID1A and NUP98-PHF23 structures are shown (see Supplementary Fig. 1 for details). **b**, Immunoblot of haematopoietic cells transduced with empty vector (lanes 1–2) or that encoding Flag-tagged NJS (lanes 3–4) or NJL (lanes 5–6). **c**, Proliferation kinetics of lineage-negative haematopoietic cells after transduction of empty vector, NJL or NJS. Data are presented as mean $\pm$ s.d. of six experiments. **d**, **e**, Wright-Giemsa staining (**d**, insert, microscopy image) and fluorescence-activated cell sorting analysis (FACS, **e**) of NJL-transformed cells. **f**, Leukaemia kinetics in mice (12 each group) after transplantation of bone marrow transduced with vector, NJL or NJS. **g**, Haematoxylin and eosin staining of spleen section and **h**, Wright-Giemsa staining of bone marrow from NJL-induced AML mice. Scale bars, 20 μ m.

revealed a dissociation constant (K_d) of $\sim 0.75 \mu$ M for JARID1A-PHD3 binding to H3K4me3, with reduced affinities for binding to H3K4me2/1/0 (Supplementary Fig. 3d).

We determined the structure of JARID1A-PHD3–H3K4me3 complexes using X-ray crystallographic and NMR spectroscopic techniques. Both analyses showed that the JARID1A-PHD3–H3K4me3 interaction was established by (1) anti-parallel β -sheet pairing between the H3 backbone and a β -sheet of JARID1A-PHD3; (2) a hydrophobic cleft formed by two Trp residues (Trp 1625 and Trp 1635) that anchor the H3K4me3 side chain; and (3) the positioning of H3R2 in an acidic pocket (Glu 1627/Asp 1629/Asp 1633) (Fig. 2b and Supplementary Figs 5b and 6c). H3K4me3 is stacked between the indole rings of two orthogonally aligned Trp residues, with intermolecular contacts shown in Fig. 2b and Supplementary Figs 5b and 6d. The X-ray (a domain-swapped dimer of one molecule and a crystallographic symmetry-related molecule) and solution NMR (monomer) analyses are summarized in Supplementary Fig. 5 (statistics in Supplementary Table 2) and Supplementary Figs 6 and 7 (statistics in Supplementary Table 3), respectively. Comparison between JARID1A-PHD3 structures in the free and H3K4me3-bound

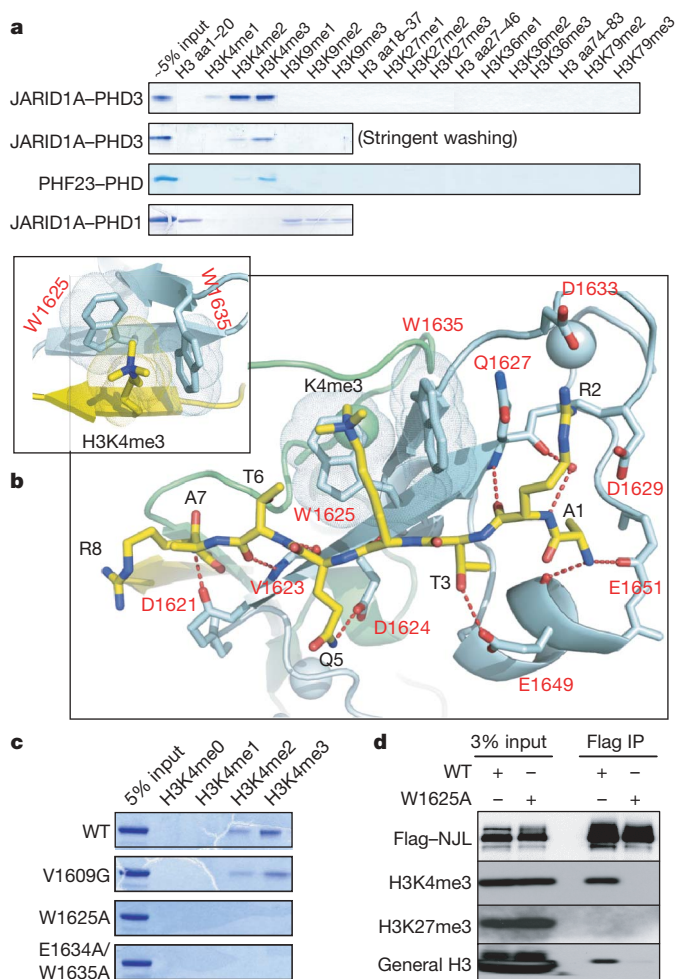


Figure 2 | JARID1A-PHD3, an essential motif for NJL-mediated leukaemia, specifically recognizes H3K4me3/2 marks. **a**, Capability of JARID1A-PHD3, PHF23-PHD and JARID1A-PHD1 (the first PHD finger of JARID1A; Supplementary Fig. 1) to interact with H3 peptides containing different states of Lys methylation, in a peptide pull-down assay. JARID1A-PHD1 interacted with H3K4me0 as BHC80-PHD¹¹. aa, amino acids. **b**, The crystal structure of JARID1A-PHD3 (cyan) complexed with H3K4me3 peptide (yellow), and a close-up view of the H3K4me3-binding channel (inset) formed by two orthogonally aligned Trp residues. The numeration of JARID1A-PHD3 and H3 residues is shown in red and black, respectively. Protein Data Bank accession number, 3GL6. **c**, Capability of wild-type (WT) or mutant JARID1A-PHD3 to bind to H3K4me3/2. **d**, Co-immunoprecipitation showing that NJL containing the wild-type, but not mutant (W1625A) PHD finger, associated with H3K4me3 or H3 in transiently transfected 293 cells.

state (Supplementary Fig. 6a, b) showed no overall conformational changes. Residues Trp 1625 and Trp 1635 are evolutionarily conserved among JARID1 homologues (Supplementary Fig. 8a). Mutations targeting these Trp residues disrupted the H3K4me3-binding both *in vitro* (Fig. 2c) and in cells (Fig. 2d). Such a two-sided H3K4me3-binding tryptophan channel is a variant form of the H3K4me3-engaging pocket involving 3–4 hydrophobic residues found in the PHD finger of BPTF⁷, ING2 (ref. 8), *Saccharomyces cerevisiae* Yng1 (ref. 19) or RAG2 (ref. 13) (Supplementary Fig. 8b–d). Yet, it exhibited a stronger H3K4me3-binding affinity ($K_d = 0.75 \mu$ M). Collectively, the PHD finger, an essential motif of NUP98-JARID1A, uniquely recognizes H3K4me3/2 using an aromatic engaging channel.

To gain insight into the mechanisms of NJL-induced AML, we used microarray analyses to compare the transcriptome of NJL-transformed marrow progenitors and that of control murine cells, committed myeloid progenitors generated as described before¹⁸. Notably, a considerable portion of genes upregulated in the NJL-transformed

progenitors were those either targeted by polycomb proteins^{20,21} or exhibiting a 'bivalent domain pattern'²² in stem cells, many of which encode developmentally critical transcription factors (*Hoxa5/a7/a9/a10*, *Gata3*, *Meis1*, *Eya1* and *Pbx1*; Supplementary Table 4). Such upregulation was further confirmed by RT-PCR using vector- versus NJL-transduced marrow cells (Supplementary Fig. 9a–c). Other *Hoxa* genes (*a1*, *a2*, *a11* and *a13*) were not expressed in NJL-transformed progenitors. We detected a similar target specificity for *Hoxa* genes using chromatin immunoprecipitation (ChIP)—NJL directly bound to the promoters of *Hoxa6–a10*, but not to distal *Hoxa1–a3* or *Hoxa11–a13* (Fig. 3a and Supplementary Fig. 9d, green). NJL-binding specificity among *Hox* clusters was correlated to H3K4me3, as H3K4me3 was abundant in *Hoxa6–a10*, but was low/absent in *Hoxa1–a4* or *Hoxa11–a13* (Fig. 3b). Enforced expression of *Hox* and *Meis1* has been shown sufficient to induce AML²³. This indicated that NJL blocks haematopoietic differentiation and induces AML by enforcing the transcription of these genes.

It has been reported that A-cluster *Hox* gene expression is high in haematopoietic stem cells and early progenitors, followed by down-regulation and shut-off during terminal differentiation²⁴. Our *ex vivo* murine haematopoietic stem/progenitor cell system recapitulated such dynamics: coincident to the silencing of a haematopoietic stem cell marker and the activation of a differentiation marker (Supplementary Fig. 9f), *Hoxa9* and *Hoxa10* were downregulated >10- or 60-fold, respectively, in 8 days of culture (Fig. 3c). The concurrent loss of *Hoxa9*- and *Hoxa10*-associated H3K4me3 was observed in these cells (Fig. 3e). Notably, NJL persistently enforced high levels of *Hoxa9/a10* expression and *Hoxa9/a10*-associated H3K4me3 in marrow cells, whereas *Hoxa9/a10* expression were silenced 10 days after transduction of vector or NJS in similarly maintained cells (Fig. 3c–e). To rigorously test the role of H3K4me3 recognition during leukaemogenesis, we mutated the H3K4me3-engaging residues. NJL containing a mutation on Trp 1625 or Trp 1635 failed to bind to H3K4me3 or H3 (Fig. 2d), failed to bind to the *Hoxa9* promoter that exhibited high H3K4me3 in 293 cells (Fig. 4a and

Supplementary Fig. 9i), failed to enforce the *Hoxa9* expression (Fig. 4b) or *Hoxa9*-associated H3K4me3 in haematopoietic progenitors (Fig. 4c), and failed to transform the haematopoietic cells (Fig. 4d), whereas the irrelevant mutation Val1609Gly did not affect these activities (Supplementary Fig. 10e). To assess whether the NJL-induced phenotype was unique to JARID1A-PHD3, we investigated another similar *de novo* translocation, NUP98–PHF23 (Fig. 1a)⁵, and also swapped JARID1A-PHD3 with other PHD fingers reported before. PHF23-PHD specifically engaged H3K4me3/2 as predicted (Fig. 2a)¹; NUP98–PHF23 robustly enforced *Hoxa9*-associated H3K4me3 and transformed haematopoietic progenitors (Fig. 4c, e and Supplementary Fig. 10). Notably, swapping JARID1A-PHD3 with another H3K4me3/2-binding PHD finger from ING2 (ref. 8) or even *S. cerevisiae* Yng1 (ref. 19) also succeeded in the transformation, whereas replacing it with an H3K4me0-binding PHD finger, either BHC80-PHD¹¹ or JARID1A-PHD1 (Fig. 2a), abolished the transformation (Fig. 4c, e). Therefore, engaging H3K4me3/2 by NUP98–PHD fusion causes leukaemia by enforcing an active state on developmentally critical loci.

Because the H3K4me3 recognition cannot provide DNA sequence specificity and yet NJL-upregulated genes were enriched with polycomb-targeted^{20,21} or 'bivalent domain' genes²² in stem cells (for example, *Hox* genes, *Gata3* and *Meis1*; Supplementary Table 4), we asked whether such specificity is due to their dynamically regulated characteristics. Towards this end, we examined the effect of NJL on two distinct gene classes: developmentally critical genes, and house-keeping genes that exhibit constitutive H3K4me3 (Supplementary Fig. 11a, top panel). Interestingly, although NJL bound to housekeeping genes, it had little effect on their expression during cell differentiation (Supplementary Fig. 11a, middle and bottom panels). Thus, NJL tends to affect the developmentally critical loci specifically during haematopoiesis. We next pursued the possibility that NJL interferes with the activities of polycomb proteins at these developmentally critical loci. Using ChIP, we found that whereas Ezh2 or Suz12 was spread throughout *Hoxa* clusters in vehicle-infected marrow progeni-

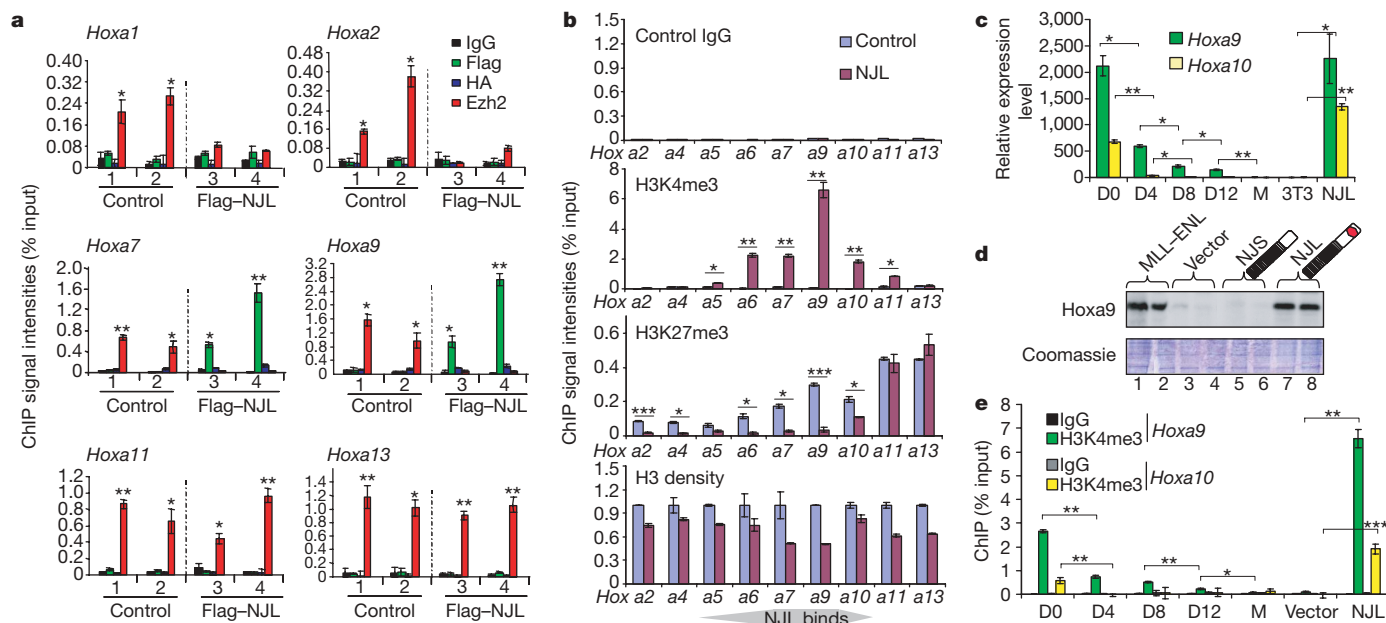


Figure 3 | NUP98–JARID1A enforced high H3K4me3 and active transcription associated with developmentally critical loci such as *Hox*. **a**, ChIP for NJL- or Ezh2-binding to *Hoxa* gene promoters in a committed myeloid progenitor line¹⁸ (1), or in haematopoietic stem/progenitor cells 3 weeks after transduction of control vector (2), or 3×Flag-tagged NJL (3–4). HA, haemagglutinin. **b**, ChIP of H3K4me3, H3K27me3 and general H3 among the *Hoxa* gene cluster in haematopoietic progenitors 3 weeks after transduction of vector or NJL. **c**, *Hoxa9* and *Hoxa10* expression in

haematopoietic stem/progenitor cells after *in vitro* cultivation (0, 4, 8 and 12 days (D)), and in macrophages (M), NIH-3T3 fibroblasts or NJL-transformed progenitors. **d**, Anti-Hoxa9 blot in marrow progenitors 10 days after transduction of MLL-ENL, empty vector, NJS or NJL. **e**, ChIP for *Hoxa9* or *Hoxa10* promoter-associated H3K4me3 in haematopoietic stem/progenitor cells after days of *in vitro* culture, and in macrophages and marrow progenitors 20 days after transduction of vector or NJL. *n* = 3, error bars indicate s.d.; **P* < 0.01, ***P* < 0.001 and ****P* < 10^{−4}.

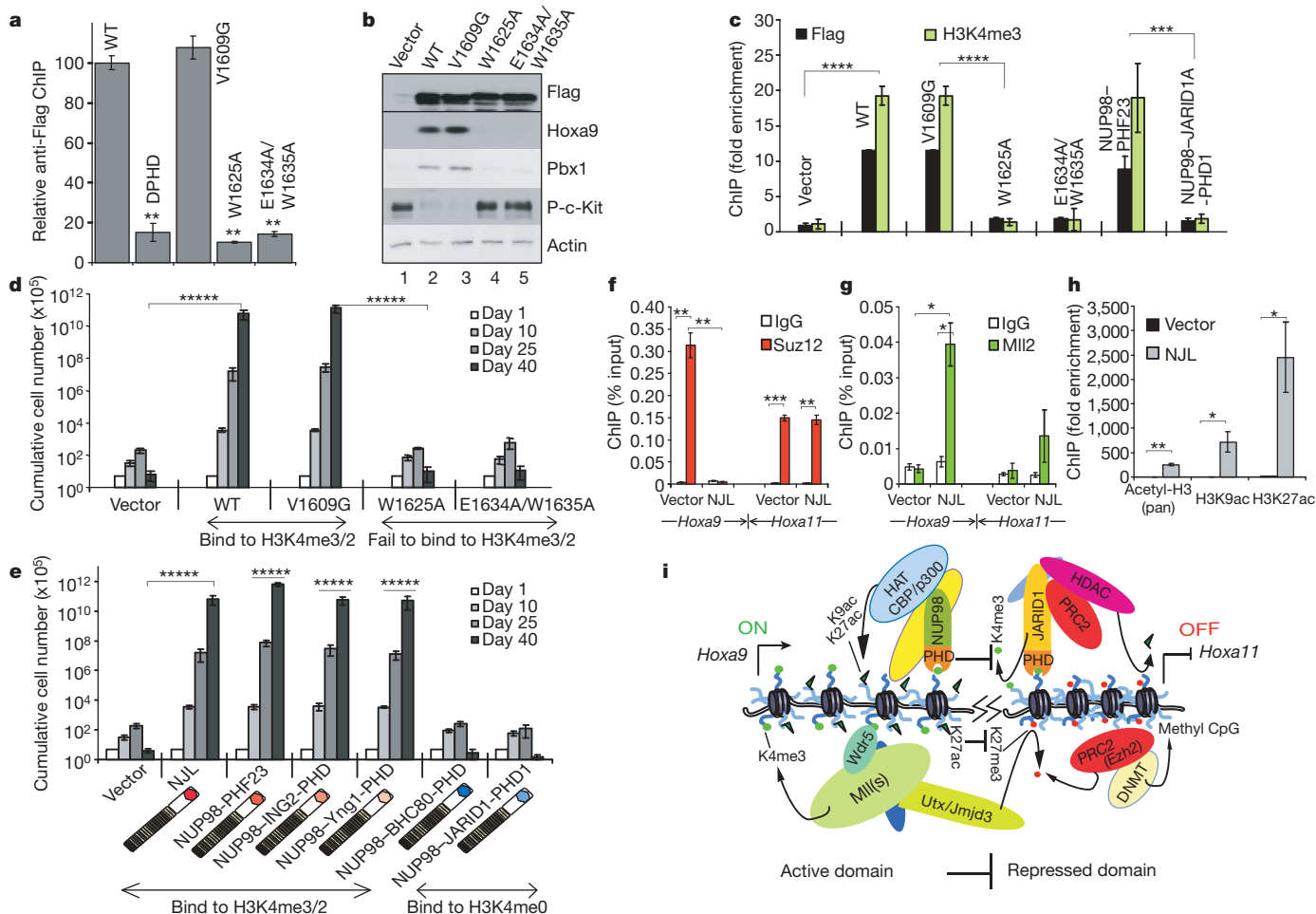


Figure 4 | The H3K4me3/2 engagement by NUP98-JARID1A perturbs the epigenetic state of developmentally critical loci during haematopoiesis.

a, The impact of mutations on Flag–NJL binding to *HOXA9* in 293 cells. WT, wild type. **b**, Immunoblot of haematopoietic progenitors 10 days after transduction of vector, wild-type or mutant NJL. Phosphorylated c-Kit (P-c-Kit) is a marker of mast cells. Actin is shown as a loading control. **c**, ChIP for *Hoxa9* promoter-associated NUP98-fusion proteins (3×Flag-tagged) and H3K4me3 in marrow progenitors 10 days after transduction. **d**, **e**, Transforming capacities after introducing mutation to NJL (**d**) or those by NUP98–PHF23 (**e**) or after replacing JARID1A–PHD3 with another PHD

tors that underwent differentiation, these polycomb factors were restricted within *Hoxa11–a13* in NJL-infected progenitors (Figs 3a and 4f, red). In the NJL-transduced cells, H3K27me3 was also only detected at *Hoxa13–a11*; the differentiation-associated spreading of H3K27me3 was inhibited at a region from *Hoxa10* to *Hoxa1* (Fig. 3b). The spreading of polycomb factors from distal *Hox* loci (*a13–a11*) seemed to be blocked at *Hoxa10–a9* by NJL that was bound there (Fig. 3a and Supplementary Fig. 9d). A similar result was also found at *Meis1* (Supplementary Fig. 9e). Consistent with previous reports^{15,16}, the recruitment of p300 (also known as EP300) and a marked increase in H3 acetylation (>2,000-fold for H3K27 acetylation) were observed on *Hoxa9* in NJL-transduced cells (Fig. 4h and Supplementary Fig. 11b). Collectively, NUP98–PHD fusion dominated over the spreading of polycomb and enforced an H3K4me3/acetylated histone state at developmentally critical loci, an epigenetic state that defines leukaemia stem cells.

We have demonstrated for the first time, to our knowledge, that fusing an H3K4me3-engaging PHD finger (plus the nuclear localization signal) to a common partner such as NUP98 is sufficient to induce leukaemia. We showed that NUP98–PHD fusion prevented the silencing of critical loci encoding master transcription factors (*Hox* genes, *Gata3*, *Mes1* and *Pbx1*) during haematopoietic differentiation.

finger that engages either H3K4me3/2 or H3K4me0. The total progenitor number was counted at day 1, 10, 25 and 40. **f–h** ChIP for Suz12 (**f**), Mll2-binding to *Hoxa9/a11* (**g**), and *Hoxa9*-associated H3 acetylation (**h**) in marrow progenitors 15 days after transduction of vector or NJL. Error bar indicates s.d.; $n = 3$; * $P < 0.05$, ** $P < 0.005$, *** $P < 10^{-4}$ and **** $P < 10^{-6}$. **i**, A scheme showing that NUP98–PHD fusion acts as a boundary factor and prevents the spreading of polycomb factors from *Hoxa13/a11* to *Hoxa9*, thus inhibiting H3K4me3 removal and H3K27me3 addition during haematopoiesis.

NUP98 fusion partners can be placed into two groups: DNA-binding homeodomain and chromatin-associated factors including PHD fingers (JARID1A and PHF23)¹⁷. Although the existence of an additional unknown ligand is possible for PHD fingers in the latter group (as the H3K4 site cannot be mutated in mammals), the most straightforward interpretation of our findings is that binding H3K4me3/2 marks are responsible for leukaemia described here. In support, a genetic interaction was demonstrated in yeast between H3K4 and the Yng1 PHD finger²⁵, a module that imparted similar oncogenic properties when swapped into our assays (Fig. 4e). Several PHD fingers exist in NSD1, another NUP98-fusion partner¹⁶, however, none contains critical H3K4me3-engaging residues¹. Thus, our report demonstrates that inappropriate interpretation of histone modification can actively induce a deregulation of developmentally critical loci, perturb cellular/epigenetic identities, and even induce oncogenesis. NUP98–PHD fusion coordinates acts of H3K4me3/2 and histone acetylation, mimicking mechanisms used by evolutionarily conserved ING(s)-complexes for robust gene activation^{19,26} (Supplementary Fig. 12). H3K4me3 bound by NUP98–PHD may serve as a ‘seed’ of propagation mediated by Wdr5–Mll2/3 complexes^{1,27} that is also coupled with Utx/Jmjd3-mediated H3K27 demethylation^{28,29}, as we detected high levels of Wdr5, Rbbp5 and Mll2 on *Hoxa9* in

NJL-transduced marrow cells (Fig. 4g and Supplementary Fig. 11c, d). We suggest that NUP98-PHD acts as a 'boundary factor', using the PHD finger to protect H3K4me3 from Jarid1-mediated demethylation²⁹ and also inducing H3K27ac to block H3K27me addition (Fig. 4i). In support, we observed a bivalent domain feature²² at *Hoxa11-a10*, the junction region of two antagonizing mechanisms (Fig. 3b). A loss-of-function mutation of RAG2-PHD in immunodeficiency, and a gain-of-function mutation involving PHD fingers in malignancies described here, indicate new types of disease that arise from 'misinterpreting' the 'histone code'^{33,30}. With ~200 PHD fingers in the human genome and some intimately associated with diseases³, we expect similar 'mis-reading' mechanisms may be responsible for some unstudied diseases. These pathologies, together with those caused by 'mis-writing' or 'mis-erasing'²⁹ histone modifications, underscore the importance in investigating the biological readout of histone marks.

METHODS SUMMARY

Haematopoietic cell transformation assays. Protocols for the culture of primary murine haematopoietic stem/progenitor cells were previously described¹⁸. In brief, 100,000 lineage-negative bone marrow stem/progenitor cells were subjected to retroviral infection, followed by kinetics analyses of proliferation versus differentiation in *ex vivo* culture system as described before¹⁸.

Peptide pull-down assay. Pull-down using biotinylated histone peptide and recombinant protein was performed as described^{6,11}. After binding, peptide-avidin beads were washed extensively in solution containing 50 mM Tris, pH 7.5, 150 mM NaCl (250 mM as stringent washing), 0.05% NP-40, 0.3 mg ml⁻¹ BSA and 1 mM dithiothreitol (DTT).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2008; accepted 2 April 2009.

Published online 10 May 2009; corrected 11 June 2009 (see full-text HTML version for details).

- Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15–30 (2007).
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
- Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **647**, 3–12 (2008).
- van Zutven, L. J. *et al.* Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes Chromosom. Cancer* **45**, 437–446 (2006).
- Reader, J. C., Meekins, J. S., Gojo, I. & Ning, Y. A novel NUP98-PHF23 fusion resulting from a cryptic translocation t(11;17)(p15;p13) in acute myeloid leukemia. *Leukemia* **21**, 842–844 (2007).
- Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
- Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
- Pena, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).
- Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
- Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
- Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718–722 (2007).
- Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**, 714–717 (2007).
- Matthews, A. G. *et al.* RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106–1110 (2007).
- Gong, W., Suzuki, K., Russell, M. & Riabowol, K. Function of the ING family of PHD proteins in cancer. *Int. J. Biochem. Cell Biol.* **37**, 1054–1065 (2005).

- Kasper, L. H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell. Biol.* **19**, 764–776 (1999).
- Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to *Hox-A* gene activation and leukaemogenesis. *Nature Cell Biol.* **9**, 804–812 (2007).
- Moore, M. A. *et al.* NUP98 dysregulation in myeloid leukemogenesis. *Ann. NY Acad. Sci.* **1106**, 114–142 (2007).
- Wang, G. G., Pasillas, M. P. & Kamps, M. P. Meis1 programs transcription of *FLT3* and cancer stem cell character, using a mechanism that requires interaction with Pbx and a novel function of the Meis1 C-terminus. *Blood* **106**, 254–264 (2005).
- Taverna, S. D. *et al.* Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell* **24**, 785–796 (2006).
- Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
- Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Kroon, E. *et al.* *Hoxa9* transforms primary bone marrow cells through specific collaboration with *Meis1a* but not *Pbx1b*. *EMBO J.* **17**, 3714–3725 (1998).
- Pineault, N., Helgason, C. D., Lawrence, H. J. & Humphries, R. K. Differential expression of *Hox*, *Meis1*, and *Pbx1* genes in primitive cells throughout murine hematopoietic ontogeny. *Exp. Hematol.* **30**, 49–57 (2002).
- Martin, D. G. *et al.* The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol. Cell. Biol.* **26**, 7871–7879 (2006).
- Doyon, Y. *et al.* ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–64 (2006).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).
- Cloos, P. A., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140 (2008).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Baker, P. Chi, A. Ruthenberg and A. Xiao for critical reading and help in manuscript preparation, C. Hughes for antibodies, and other Allis laboratory members for their advice. We are extremely grateful to M. Kamps and S. Rafii for shared plasmids and expert advice. Thanks to E. Coutavas, W. Chen, E. Ezhkova and the R. Roeder laboratory for help on cell and animal work. G.G.W. is supported by a Leukemia & Lymphoma Society Fellow award and a Choh-Hao Li Memorial Fund Scholar award; H.L.D. is supported by a predoctoral Boehringer Ingelheim Foundation fellowship; This research was supported by the National Institutes of Health (NIH) grant and funds of Rockefeller University to C.D.A., funds of Abby Rockefeller Mauze Trust and the Dewitt Wallace and Maloris Foundations to D.J.P., US Department of Defense CDMRP grant to J.-L.L., and a joint Starr Foundation Cancer Consortium grant.

Author Contributions G.G.W. and C.D.A. designed the study. G.G.W. performed most of the cellular and molecular experiments, and wrote the paper; J.S. performed protein preparation, NMR structure determination and crystallization; Z.W. performed crystallographic analyses; H.L.D. and G.G.W. did the immunostaining; F.C. participated in plasmid/protein preparation; H.L. performed isothermal titration calorimetry measurement; G.G.W. and J.-L.L. performed animal studies; D.J.P. and C.D.A. supervised the structural and functional aspects of the project, respectively, and helped with manuscript preparation.

Author Information The structural coordinates of JARID1A-PHD3 in the H3-bound or free state have been submitted to the Protein Data Bank under accessions 3GL6, 2KGG and 2KGI, and the chemical shift assignment of NMR structures to BioMagResBank under accessions 16209 and 16210. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.D.A. (alliscd@rockefeller.edu).

METHODS

Plasmid construction and retroviral expression system. The human *NUP98-JARID1A* fusion complementary DNA⁴ was generated by ligating *NUP98* sequences encoding amino acids 1–514 to those encoding amino acids 1489–1690 of *JARID1A* transcript variant 1 (NCBI accession number NM_001042603) or amino acids 1489–1641 of *JARID1A* transcript variant 2 (NCBI accession number NM_005056), producing two fusion isoforms (NJL or NJS), respectively. The same method was used to generate *NUP98-PHF23* (ref. 5). The fusion cDNA with an amino-terminal 3×FLAG was cloned into MSCV retroviral expression vector (Clontech). *JARID1A*, *PHF23* and *BHC80* cDNAs were purchased from Open Biosystems. *NUP98* plasmids were provided by J. M. van Deursen, *Hoxa9* by M. P. Kamps, *MLL-ENL* by R. K. Slany, *Yng1* by S. D. Taverna, *CBX7* by E. Bernstein, and *ING2* by Z. Tang. Mutations were generated by site-directed mutagenesis, and all used plasmids were confirmed by sequencing.

Purification and culture of primary haematopoietic cells. Bone marrow cells collected from the femur and tibia of balb/c or b/6 mice were subject to lineage-negative (Lin^-) enrichment using Haematopoietic Progenitor Enrichment Kit (StemCell Technologies or Miltenyi Biotec) to remove cells expressing differentiation antigens as described before¹⁶. Approximately 400,000 Lin^- -enriched haematopoietic progenitors were obtained per mouse with ~10% c-Kit⁺ Lin^- Sca1⁺ haematopoietic stem cells. Before retroviral infection, Lin^- -enriched haematopoietic progenitors were stimulated in OptiMEM base medium (Invitrogen) complemented with 10% FBS (Invitrogen), 1% antibiotics, 50 μM β -mercaptoethanol and a cytokine cocktail containing stem-cell factor (SCF; supernatant of SCF-producer cells), 5 ng ml⁻¹ FLT3 ligand (Sigma), 5 ng ml⁻² IL3 and IL6 (Miltenyi) for 2–3 days as described^{18,31}. After retroviral infection and selection (1 $\mu\text{g ml}^{-1}$ puromycin), marrow cells were plated in the same medium with SCF as the sole cytokine. Cell splitting and replating to fresh medium were performed every 3–4 days to keep the cell number <2 million per well (6- or 12-well plates). Cell morphology was examined by Wright–Giemsa staining. Macrophages were obtained by culture of marrow cells in M-CSF (Miltenyi) for 1–2 weeks as described³². Immortalized cell lines that mimic committed neutrophil-macrophage progenitors were generated as described previously^{18,31,33}.

Murine bone marrow transplantation leukaemogenic assay. The leukaemogenic potentials of oncogenes were evaluated in sublethally irradiated syngeneic mice, after tail vein injection with 100,000 bone-marrow-derived Lin^- cells that were infected with retrovirus encoding the fusion gene as described¹⁸. Mice exhibiting leukaemic phenotype were subjected to pathological analyses.

Recombinant protein production and glutathione S-transferase pull-down. *JARID1A-PHD3* (amino acids 1601–1660) glutathione S-transferase (GST)-fusion proteins were produced using a previously described protocol¹⁹. GST pull-down using total histone extracts was performed as described with modification⁹. In brief, ~2 μg GST-fusion protein bound to glutathione beads (Amersham) were incubated with 10 μg of calf thymus histone extracts (Worthington) in a binding buffer containing 50 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 0.5% NP-40, 0.2 mM EDTA, 1 mM DTT and protease inhibitor cocktail (Roche) at 4 °C for 4 h.

Native co-immunoprecipitation. Mononucleosome-containing fractions were prepared as described before⁶. In brief, intact nuclei were subject to limited micrococcal nuclease (MNase) digestion, so that the major form of released chromatin is mononucleosome. After the removal of the insoluble fraction by centrifugation, supernatant containing mononucleosomes was then incubated with Flag or HA-agarose beads (Sigma), or with Dynal magnetic beads (Invitrogen) coupled with anti-H3K4me3 (Abcam) or control antibodies. After extensive washing, precipitated proteins were subject to immunoblot.

Isothermal titration calorimetry measurements. Calorimetric experiments were conducted at 25 °C with a MicroCal iTC200 instrument as described⁷. Recombinant *JARID1A-PHD3* protein and H3_{1–15}K4me (H3 amino acids 1–15, with Lys4 methylated) peptides were dialysed overnight against 25 mM Tris-HCl, pH 7.5, 50 mM KCl and 2 mM β -mercaptoethanol. Protein concentration was determined by absorbance spectroscopy (Tyr $\epsilon_{280} = 1,420 \text{ M}^{-1} \text{ cm}^{-1}$; Trp $\epsilon_{280} = 5,600 \text{ M}^{-1} \text{ cm}^{-1}$; Cys $\epsilon_{280} = 125 \text{ M}^{-1} \text{ cm}^{-1}$). H3_{1–15}K4me peptides were quantified by the absorbance of an added C-terminal Tyr with $\epsilon_{280} = 1,280 \text{ M}^{-1} \text{ cm}^{-1}$ for peptide. Acquired calorimetric titration data were analysed using software Origin7.0 (MicroCal, iTC200) on the basis of a 1:1 binding stoichiometry.

Antibodies and immunoblot. Antibodies used were anti-Flag (Sigma; M2), anti-HA (Covance, MMS101), anti-Hoxa9 (Upstate, 07-178), anti-Pbx1 (Santa Cruz, sc889), anti-phospho-c-Kit (Cell Signaling) and anti-Tubulin (Sigma).

ChIP analysis. ChIP analysis was performed using an Upstate ChIP kit and a protocol described before³⁴. One-to-two-million cells per ChIP were used for histones, and 2–3 million for others. Antibodies and the amount used were anti-Flag (Sigma M2, 1–3 μg), anti-HA (Covance MMS101, 1–3 μg), anti-H3K4me3 (Upstate 07-473, 1 μl ; Abcam 8580, 0.5 μg), anti-H3K27me3 (Upstate 07-449, 0.5 μg), anti-acetyl-H3 (Upstate 06-599, 1 μg), anti-general H3 (Abcam

1791, 0.5 μg), anti-acetyl-H3K9 (Upstate 06-942, 1 μg), anti-acetyl-H3K27 (Abcam 4729, 1 μg), anti-Ezh2 (Cell Signaling 4905, 4–5 μl), anti-Suz12 (Upstate 07-379, 2 μl), anti-Mll2 (Bethyl A300-113A, 4 μg), anti-WDR5 (Upstate 07-706, 2 μg), anti-RBBP5 (Bethyl A300-109A, 3 μg ; a gift from C. Hughes) and anti-p300 (Santa Cruz, N15/C20, 10 μg). The same amount of nonspecific IgG (Upstate) was used as antibody control, and a silenced intragenic locus, Chr8Int, was used as a locus control for H3K4me3 or activator binding as described²¹. The promoter sequence was acquired from the UCSC genomic browser (<http://genome.ucsc.edu>). ChIP primers are shown in Supplementary Table 5. ChIP signals were represented as the percentage of signals from total chromatin used, and the fold of enrichment was calculated by normalizing against signals of nonspecific IgG.

Microarray analysis. Total RNA was extracted and the transcript expression was quantified using Affymetrix GeneChip Mouse arrays as described¹⁸. RNA hybridization, scanning and signal quantification were performed by the Rockefeller University Genomic Resources Center. Hybridization signals were retrieved and normalized, followed by differential expression analysis and statistical analysis using GeneSpring Analysis Platform GX 7.0 (Agilent Technologies).

RT-PCR analysis. Reverse transcription of RNA was performed using the random hexamer and Invitrogen Superscript III kit. Usually the PCR amplicon (~90–200 bp) is designed to span over large intron regions. Exon–intron information was obtained from the UCSC genomic browser. Quantitative PCR was performed in triplicate using SYBR green master mix reagent (Applied Biosystem) on a Stratagene Mx3005P QPCR system. Primer information is shown in Supplementary Table 5.

Flow cytometry. Cells were blocked with BD FcBlock (2.4G2) and stained on ice with fluoro-conjugated antibodies (1:1,000 dilution of Cd117-FITC, Sca-1-PE-CY7, Cd34-APC, Cd34-FITC, Cd11b-APC, Gr-1-PE, Cd19-PE or B220-PE, BD Biosciences), and analysed on a BD FACS Calibur cytometer. Data were collected and analysed using CellQuestPro and FlowJo software.

Immunofluorescence microscopy. Suspension-cultured haematopoietic cells were attached to coverslips treated with 0.01% (w/v) poly-lysine, followed by 15-min fixation in 4% paraformaldehyde and 10-min solubilization in PBS, 0.2% Triton-X100 and 0.2% NP-40. After a 30-min block in PBS, 2.5% BSA and 10% normal goat serum, cells were stained with primary antibodies (M2 anti-Flag (1:1,000–2,000 dilution of 1 mg ml⁻¹), rabbit anti-H3K4me3 (Upstate 07-473 or Abcam 8580, 1:2,000) or rabbit anti-H3K9me3 antibodies (Upstate 07-442, 1:1,000)) followed by washing and staining with fluorescent-labelled secondary antibodies. After washing, fluorescent signal was visualized and analysed with a DeltaVision Image Restoration Microscope (Applied Precision) and a Confocal Microscope (Olympus). Deconvolution microscopy image analysis was performed to reassign the out-of-focus blurred light to its origin³⁵, and subcellular co-localization analysis was carried out from stacks of deconvolved images using ImageJ (W. S. Rasband, <http://rsb.info.nih.gov/ij/>) and the plugin JACO³⁶. Confocal microscopy analysis was performed as previously described³⁷. Co-immunostaining statistics was analysed using Pearson's Coefficient of Correlation method. Image acquisition, processing and analyses were performed with help from Rockefeller University Bio-Imaging Center, and detailed protocols are available on request.

Statistics. All results are presented as the mean and s.d. Statistical analyses were performed using Student's *t*-test.

Protein preparation for structural studies. The gene fragment encoding *JARID1A-PHD3* was fused to the C terminus of a His(6×)-SUMO tag in a modified pRSFDuet-1 vector (Novagen), with a ubiquitin-like-protease (ULP) cleavage site located at the linker region. The bacterial expressed protein was purified using a Ni-NTA affinity column, followed by ULP cleavage, separation of *JARID1A-PHD3* from His(6×)-SUMO via a second Ni-NTA chromatography step, and gel filtration. The *JARID1A-PHD3*–H3K4me3 complex was obtained by mixing *JARID1A-PHD3* protein with an equal molar amount of H3_{1–9}K4me3 peptides (H3 amino acids 1–9, with Lys4 trimethylated), then purified by gel filtration, and concentrated by ultrafiltration.

Crystal growth. The crystals of *JARID1A-PHD3*–H3K4me3 complexes were obtained by equilibrating a reservoir consisting of 20% (w/v) poly(ethylene glycol) monomethyl ether 2000, 10 mM nickel (II) chloride hexahydrate, and 0.1 M Tris, pH 8.5, with a hanging drop consisting of 1 μl of the reservoir solution and 1 μl of a 27 mg ml⁻¹ protein solution in 10 mM Tris, pH 8.0, 0.1 mM ZnCl₂, 5 mM DTT and 50 mM NaCl (Crystal Screen 2 kit, Hampton Research). A mixture of the well solution with 10% (v/v) glycerol was used as a cryoprotectant.

Data collection and structure determination. An anomalous diffraction data set for the *JARID1A-PHD3*–H3K4me3 complex was collected at the zinc anomalous peak wavelength (1.28215 Å) at beamline NE-CAT 24ID-C, Advanced Photon Source, Chicago. The data set was indexed, integrated and merged to 2.2 Å using the program HKL2000. The crystal belongs to *I*₄ space group and contains one molecule per asymmetric unit. Heavy-atom search, single-wavelength anomalous dispersion (SAD) phasing and model building

were performed with the PHENIX³⁸ software package. Three zinc atoms were unambiguously identified for SAD phasing, and ~90% residues of the protein–peptide complex were successfully built into the initial model. The PHENIX–model was further manually rebuilt using COOT³⁹ and refined using REFMAC5⁴⁰ in successive cycles. The final refined structure has R_{work} and R_{free} values of 0.200 and 0.234, respectively (Supplementary Table 2). One molecule forms a domain-swapped dimer with a crystallographic symmetry-related molecule (Supplementary Fig. 5a). The swapped segment spans the first 14 residues from the N terminus.

Using a crystal of the complex following pH optimization of crystallization conditions, we were able to collect one 1.9 Å data set at wavelength 0.97949 Å at the same beamline. The crystal belongs to the same crystal form as the previous one. We solved the high-resolution structure by molecular replacement using PHASER⁴¹ with the above 2.2 Å model after removing all the water molecules and some flexible residues. Structure refinement was done using CNSsolve⁴², cycled with manual model building in COOT. Before the refinement, the same R_{free} set of reflections were transferred from the low-resolution data using the program Freerflag in CCP4 suite⁴³ for effective cross validation. For both data sets, ~10% reflections were selected in a ‘random’ mode throughout the resolution range. After resetting the overall B-factor to 20 Å² and rigid body refinement, simulated annealing starting at 5,000K was performed to reduce model bias before extensive B-factor and positional refinement. The final model contains full-length JARID1A-PHD3 (1609–1659) with one extra serine at the N terminus from the expression vector, histone H3_{1–9}K4me3, three zinc ions and 32 water molecules. The JARID1A-PHD3–H3K4me3 complex in the crystal shows that one molecule forms a domain-swapped dimer with a crystallographically symmetry-related molecule (Supplementary Fig. 5a). Two zinc ions are integral to the folding of the PHD finger, whereas the third zinc ion locates at the interface between two domain-swapped dimers, thereby mediating crystal packing (Supplementary Fig. 5c). The R_{work} and R_{free} of the final structure are 0.208 and 0.225, respectively (Supplementary Table 2).

Isotopic labelling, NMR data collection and structure determination. Samples used for NMR chemical shift assignments, ¹⁵N relaxation measurements, and structure determination contained 0.2–0.5 mM uniformly-¹⁵N- or [¹³C, ¹⁵N]-labelled JARID1A-PHD3 in the free state and in complex with unlabelled H3_{1–9}K4me3 peptide dissolved in NMR buffer (20 mM Na-phosphate, 1 mM ZnCl₂, 5 mM DTT, 90% H₂O/10% D₂O) at pH 7.0. The sample used for measurements of ¹⁵N-¹H residual dipolar couplings (RDCs) contained 0.2 mM JARID1A-PHD3 aligned in 12 mg ml^{−1} of bacteriophage Pf1 (Alsa), 10 mM MOPS, 200 mM NaCl, pH 7.0.

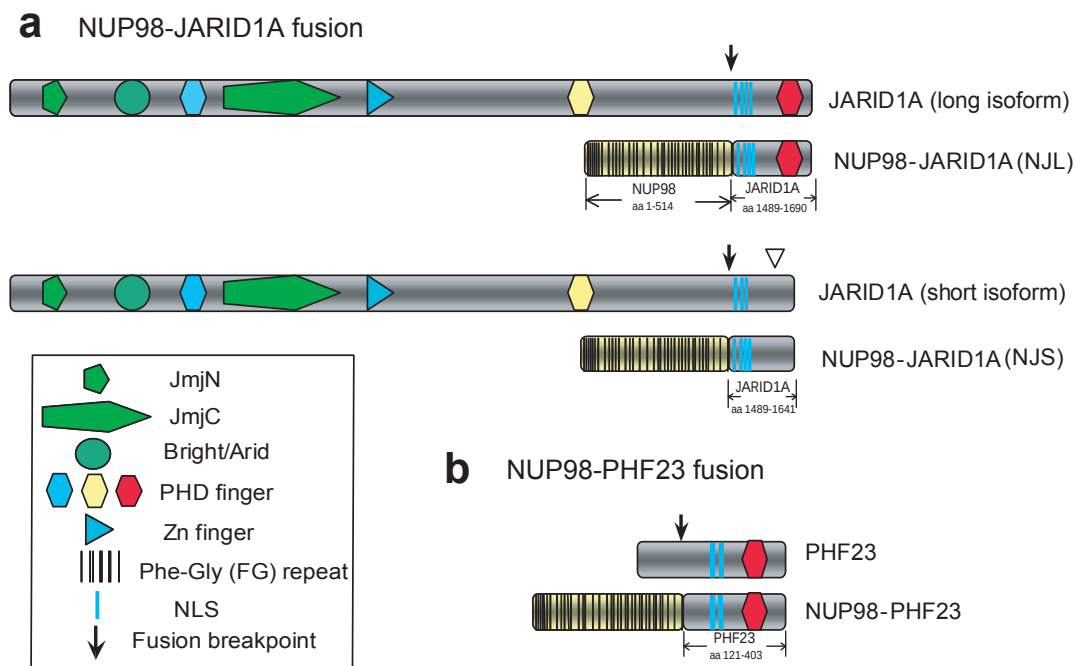
All NMR spectra were collected at the New York Structural Biology Center using 800 MHz Bruker NMR spectrometers equipped with ¹H, ¹⁵N, ¹³C triple-resonance cryogenic probes. Unless indicated otherwise, the sample temperature was controlled at 20 °C. A suite of three-dimensional (3D) heteronuclear NMR experiments, including HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH and HCCH-TOCSY were acquired for sequential backbone and non-aromatic side-chain assignments of JARID1A-PHD3 both in the free state and in complex with H3_{1–9}K4me3 peptide in solution. Two-dimensional (2D) nuclear Overhauser enhancement spectroscopy (NOESY) ($\tau_{\text{mix}} = 100$ ms), 3D ¹⁵N-edited NOESY-HSQC (heteronuclear single quantum correlation) ($\tau_{\text{mix}} = 100$ ms), 3D aromatic ¹³C-edited NOESY-HSQC ($\tau_{\text{mix}} = 100$ ms) and 3D aliphatic ¹³C-edited NOESY-HSQC ($\tau_{\text{mix}} = 100$ ms) data sets were acquired and used for additional assignments (side-chain amide and aromatic groups) and distance constraints. To selectively observe the nuclear Overhauser effect (NOEs) between JARID1A-PHD3 and H3_{1–9}K4me3 peptide, a [¹³C, ¹⁵N]-filtered, ¹³C-edited NOESY ($\tau_{\text{mix}} = 120$ ms) spectrum⁴⁴ of uniformly [¹⁵N, ¹³C]-labelled JARID1A-PHD3 bound to unlabelled H3_{1–9}K4me3 peptide was recorded. One-bond N-H RDCs were determined by using the in-phase anti-phase (IPAP) ¹⁵N-HSQC pulse sequence at 25 °C⁴⁵. Standard pulse sequences⁴⁶ were used for measurements of the ¹⁵N relaxation rates (R_1 , R_2) of JARID1A-PHD3 at 25 °C. The spectra were processed and analysed, respectively, with the NMRPipe⁴⁷ and Sparky (http://www.cgl.ucsf.edu/home/sparky) software. The solution structures of JARID1A-PHD3 both in the free state and in complex with H3_{1–9}K4me3 peptide were first calculated using the CYANA program⁴⁸. Interproton distance constraints were derived from 2D NOESY, 3D ¹⁵N-edited NOESY-HSQC and 3D ¹³C-edited NOESY-HSQC spectra. Backbone ϕ and ψ angles were derived from TALOS-based analysis of backbone chemical shifts⁴⁹. Several hydrogen bonds derived from chemical shift analysis and from observed NOEs characteristic for α -helices and β -sheets, were added in the final rounds of structure refinement. Of the 100 final structures calculated by CYANA, 20 structures with the lowest target functions were chosen for further refinement using the Xplor-NIH program⁵⁰, in which ¹D_{NH} RDC restraints, physical force field terms and explicit solvent terms⁵¹ were added to the calculation. The final structures were validated by Procheck-NMR⁵², and the statistics for the 20 final structures are listed in Supplementary Table 3.

Monomeric state of JARID1A-PHD3 in free and H3_{1–9}K4me3-bound states in solution. The oligomeric states of JARID1A-PHD3 (molecular weight 5.8 kDa) and JARID1A-PHD3–H3_{1–9}K4me3 complex (molecular weight 6.8 kDa) were first evaluated by comparing their elution volumes on a Superdex G75 16/60 column, with the calibration line derived from several molecular standards. As shown in Supplementary Fig. 7a, the elution volumes of both free JARID1A-PHD3 and the JARID1A-PHD3–H3_{1–9}K4me3 complex are comparable with those expected for their monomeric states, but considerably larger than those expected for their dimeric states. This suggests that JARID1A-PHD3 is monomeric in solution, for both the free and H3_{1–9}K4me3-bound states. Furthermore, the rotational correlation times of JARID1A-PHD3 in the free and H3_{1–9}K4me3-bound states were estimated as 3.5 ns and 4.5 ns, respectively, on the basis of an analysis of ¹⁵N R_2/R_1 relaxation time ratios (Supplementary Fig. 7b) using the quadratic diffusion program⁵³. These values are consistent with isotropic tumbling values of a monomeric protein of their respective sizes, providing further support that both free and complexed JARID1A-PHD3 exist as monomers in solution.

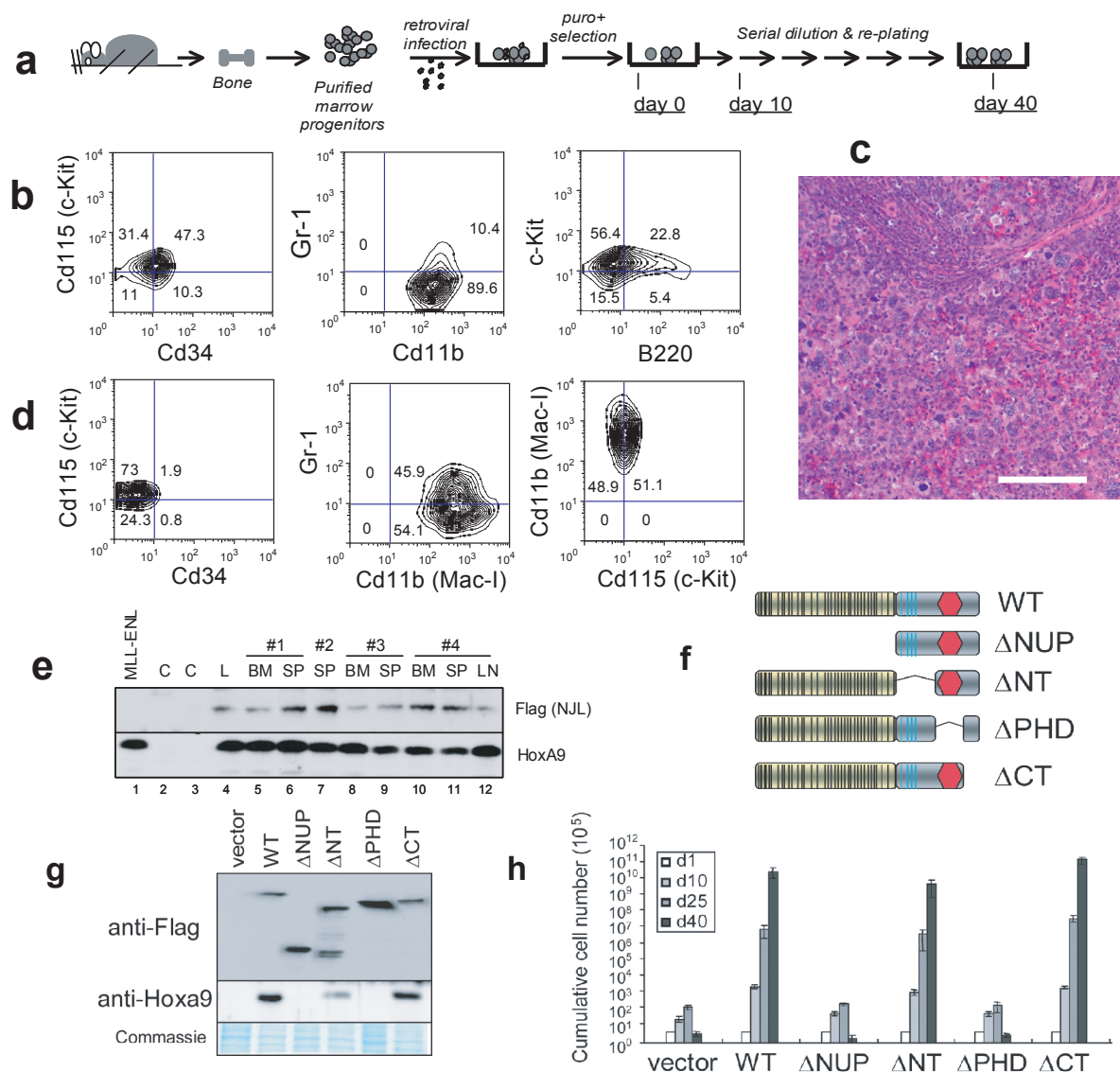
Thus, although the JARID1A-PHD3–H3K4me3 complex exhibits a domain-swapped dimer in the crystal, gel filtration and NMR relaxation measurements (Supplementary Fig. 7) clearly showed such a complex to be monomeric in solution. Hence the domain-swapped dimerization observed in the crystal is probably a characteristic feature of the crystalline state, originating perhaps in packing interactions.

31. Wang, G. G. *et al.* Quantitative production of macrophages or neutrophils *ex vivo* using conditional Hoxb8. *Nature Methods* **3**, 287–293 (2006).
32. Sawka-Verhelle, D. *et al.* PE-1/METS, an antiproliferative Ets repressor factor, is induced by CREB-1/CREM-1 during macrophage differentiation. *J. Biol. Chem.* **279**, 17772–17784 (2004).
33. Calvo, K. R., Sykes, D. B., Pasillas, M. & Kamps, M. P. Hoxa9 immortalizes a granulocyte-macrophage colony-stimulating factor-dependent promyelocyte capable of biphenotypic differentiation to neutrophils or macrophages, independent of enforced meis expression. *Mol. Cell. Biol.* **20**, 3274–3285 (2000).
34. Okada, Y. *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* **121**, 167–178 (2005).
35. Wallace, W., Schaefer, L. H. & Swedlow, J. R. A workingperson's guide to deconvolution in light microscopy. *Biotechniques* **31**, 1076–1078 (2001).
36. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
37. Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* **D 58**, 1948–1954 (2002).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **D 60**, 2126–2132 (2004).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D 53**, 240–255 (1997).
41. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr.* **D 63**, 32–41 (2007).
42. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
43. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D 50**, 760–763 (1994).
44. Zwahlen, C. *et al.* Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721 (1997).
45. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
46. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
47. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
48. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
49. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
50. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
51. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **50**, 496–506 (2003).
52. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
53. Lee, L. K., Rance, M., Chazin, W. J. & Palmer, A. G. III. Rotational diffusion anisotropy of proteins from simultaneous analysis of ¹⁵N and ¹³C alpha nuclear spin relaxation. *J. Biomol. NMR* **9**, 287–298 (1997).

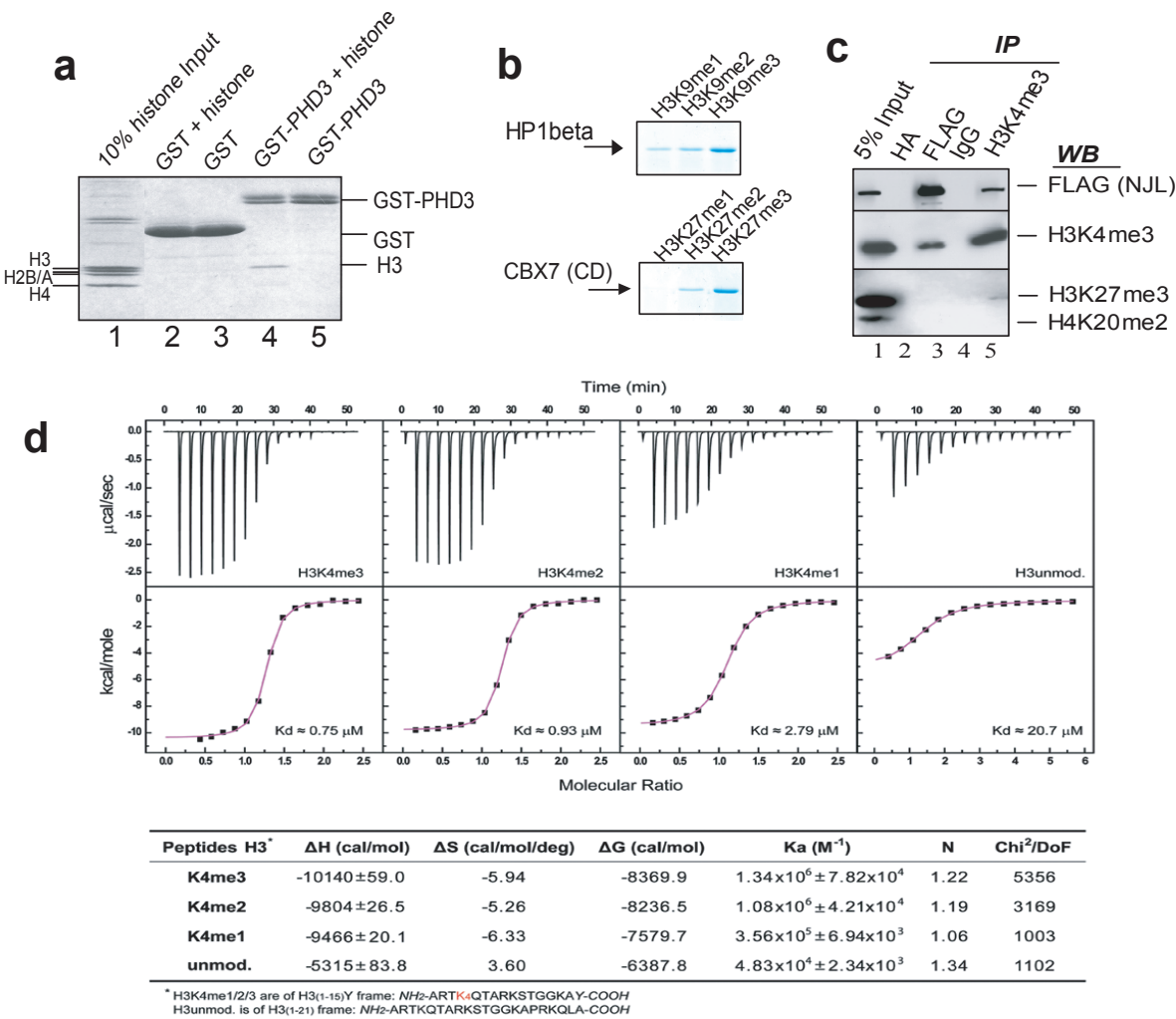
SUPPLEMENTARY INFORMATION



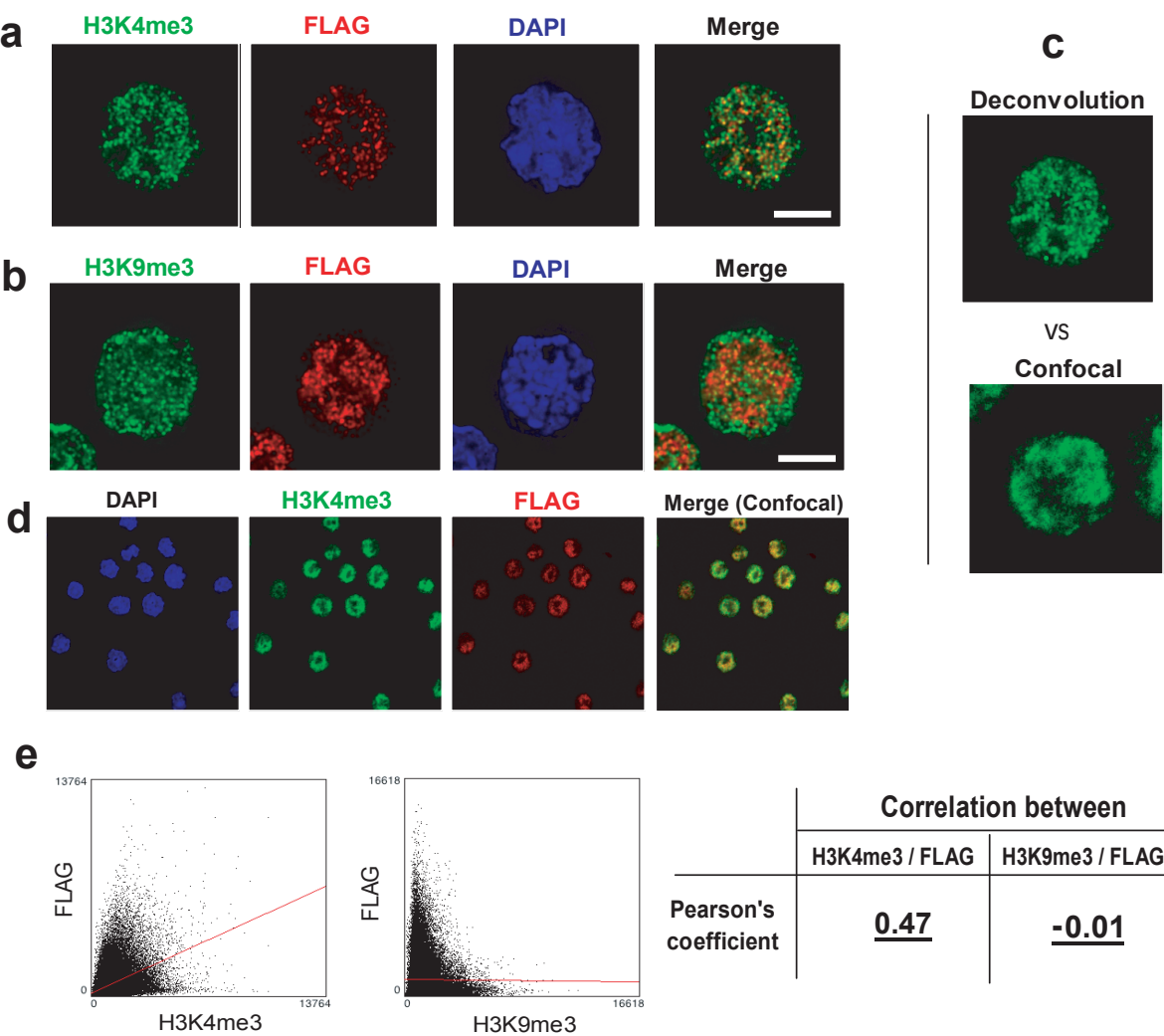
Supplementary Figure 1| Abnormal chromosomal translocations as reported in human leukemia patients fuse the N-terminal part of NUP98 to the C-terminus of JARID1A or PHF23— a region harboring a single PHD finger and adjacent nuclear localization signals (NLS)^{1,2}. **a**, Structure of wildtype JARID1A (also known as RBP2 and KDM5A) protein and NUP98-JARID1A fusion. Due to an alternative splicing that incorporates a cryptic exon (marked with an inverted triangle), the third PHD finger at C-terminus (JARID1A_{PHD3}) is excluded from the shorter isoform of JARID1A. As a result, the JARID1A_{PHD3} domain is excluded from the corresponding shorter NUP98-JARID1A fusion isoform (hereafter referred to as NJS, bottom panel), while it is retained in the longer fusion isoform (hereafter referred to as NJL, top panel). The designation of each motif is shown in the legend box. **b**, Structure of PHF23 and NUP98-PHF23. NUP98-PHF23 exhibits a high similarity to NJL.



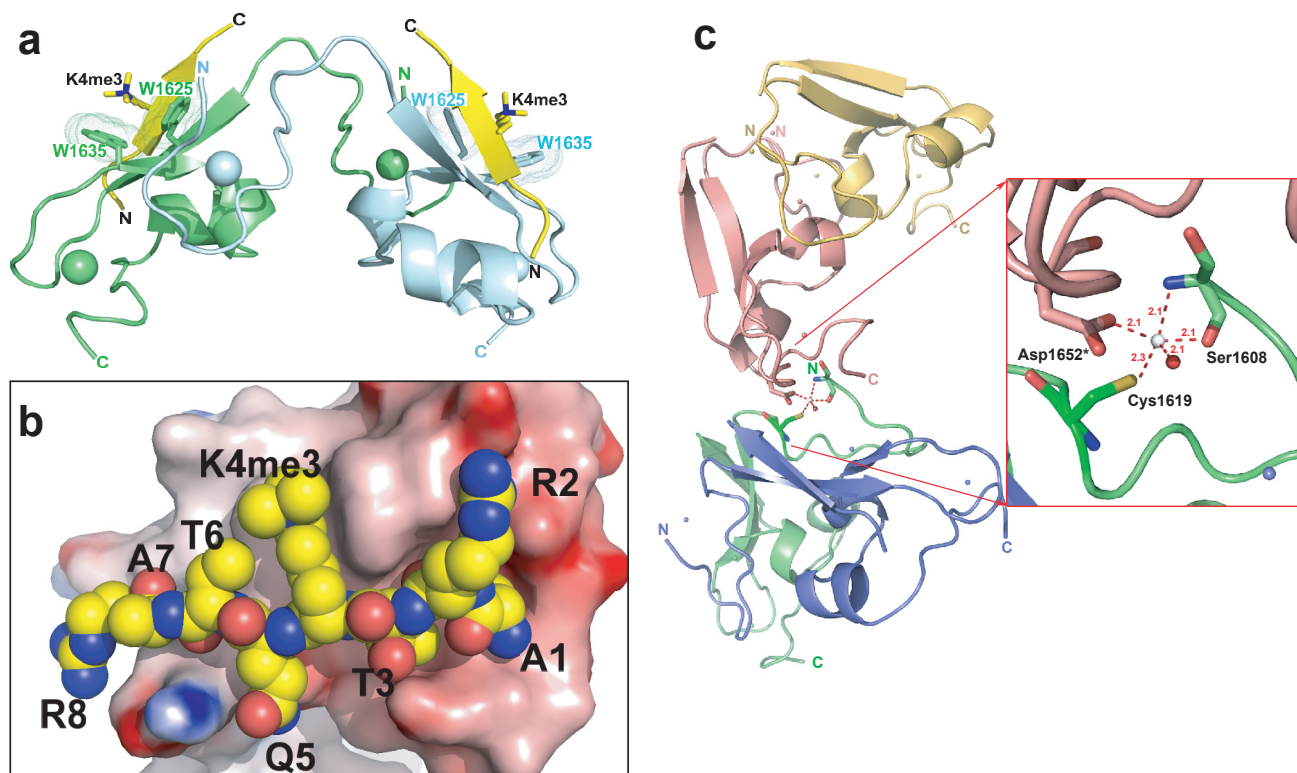
Supplementary Figure 2| NUP98-JARID1A confers leukomogenic potential to lineage-negative hematopoietic cells in murine bone marrow transplantation leukemic models. **a**, Scheme of *in vitro* hematopoietic cell serial re-plating assay to evaluate the effect of oncogenes on hematopoietic differentiation versus proliferation as described before^{3,4}. **b**, FACS analysis revealing the expression of surface markers specific for both early progenitors (cKit⁺ Cd34⁺) and committed myeloid lineages (Mac1⁺ Gr1⁺ B220^{low} Cd19⁻) in NJL-transduced progenitors four weeks after retroviral transduction. **c**, Hematoxylin-eosin staining of spleen sections demonstrated massive infiltration and growth of leukemic cells in mice transplanted with NJL-transduced hematopoietic cells. Scale bar, 200 micrometer. **d**, FACS analysis of leukemic cells extracted from bone marrow of NJL-induced AML exhibited a myeloid phenotype (Cd34⁺ cKit⁺ Mac1⁺ Gr1⁺). **e**, Immunoblot of NJL (FLAG-tagged) and Hoxa9 in marrow cells transduced with a known Hoxa9 activator MLL-ENL (lane 1), empty vector (C, lanes 2-3) or NJL (L, lane 4), and in leukemic cells extracted from bone marrow (BM), spleen (SP) or lymph nodes (LN) from AML mice. The number above lanes indicates mouse identification number. **f**, Diagram showing NJL deletion constructs. **g**, Immunoblot of NJL (Flag tagged) and Hoxa9 in hematopoietic cells 10 days after retroviral infection. **h**, Capability of NJL deletion mutants to transform lineage-negative hematopoietic cells. Data are presented as mean $\pm$ s.d. of the progenitor number among four experiments counted at day 1 (open bars), 10 (light gray), 25 (dark gray) and 40 (black).



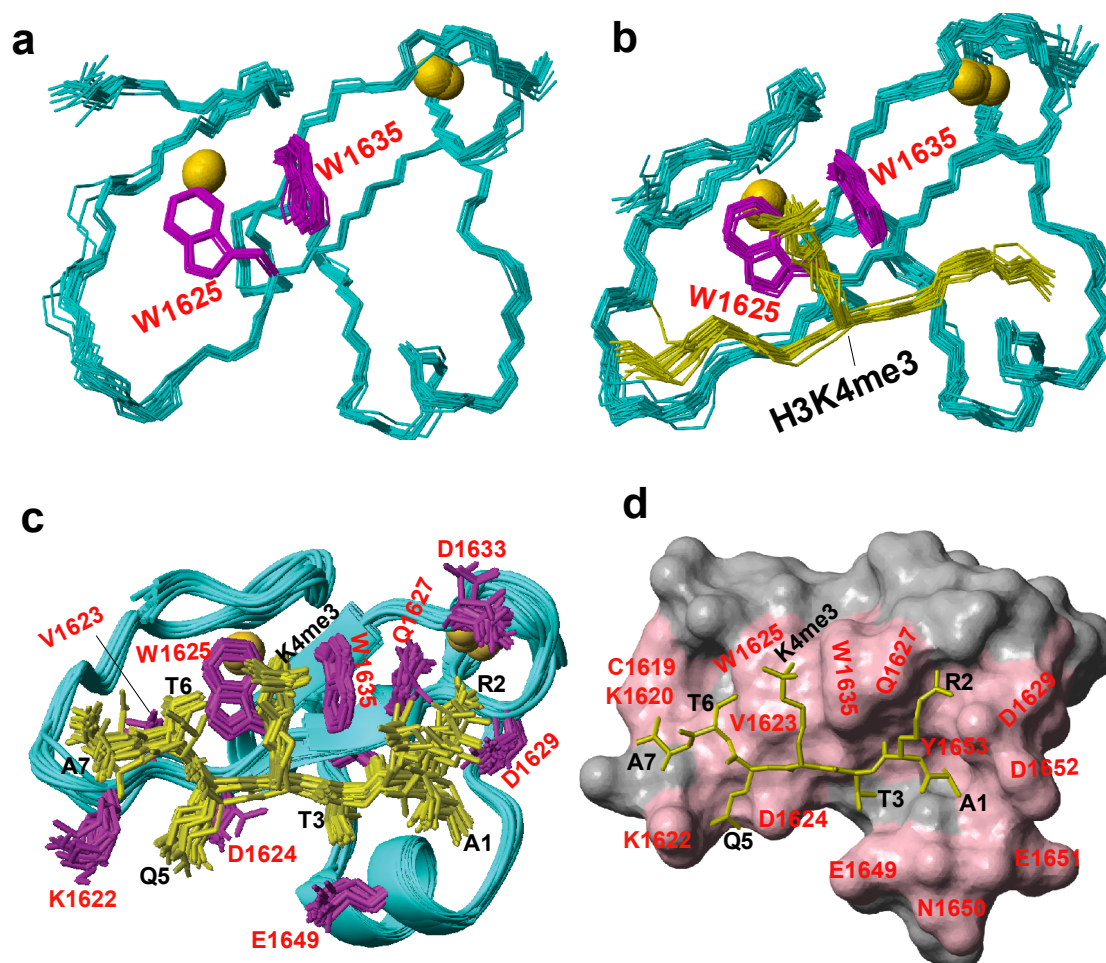
Supplementary Figure 3| Interaction between H3K4me3/2 and JARID1A_{PHD3}. a, Recombinant GST-JARID1A_{PHD3} fusion protein, but not GST alone, interacted with histone H3 and not other histone species (lane 4 versus 2). The 17% SDS-PAGE gel was loaded with 10% input of histone extracts used (lane 1) and the sample of GST pull-down using either GST alone (lanes 2-3) or GST-JARID1A_{PHD3} (lanes 4-5) in the presence or absence of histones, followed by visualization by coomassie blue. b, The chromodomain (CD) of HP1-beta⁵ and CBX7⁶ serves as positive control for H3K9me- and H3K27me3/2-binding effectors respectively in biotinylated peptide pull down shown in Fig 2a. c, CoIP of mononucleosomes and associated factors after limited micrococcal nuclease (MNase)-digestion of nuclei from Flag-NJL stable expression cells. Loaded were IP samples using anti-HA or M2 anti-FLAG agarose beads, or beads coupled with IgG or anti-H3K4me3 antibodies, followed by western blot (WB) with antibodies suggested on the right. Anti-H3K4me3 beads precipitated ~5% of total H3K4me3 together with ~5% of total FLAG-NJL, indicating that the vast majority of NJL bound to H3K4me3-containing mononucleosomes that also showed little H3K27me3 or H4K20me2; anti-FLAG beads precipitated nearly all of NJL together with only ~0.5% of total H3K4me3, indicating a ligand excess (H3K4me3); Similar but stronger interaction for H3K4me3 was observed after overexpression of NJL in 293 cells (Fig. 2d). d, Isothermal titration calorimetry (ITC) assays of JARID1A_{PHD3} bound to H3 peptides as a function of H3K4 methylation state. ITC titration experimental curves are shown in top panel and the corresponding enthalpy fitting plots are shown in the middle panel. Titrations were performed using the same protein batch with a concentration of 0.125 mM. Peptide concentrations are 1.5 mM for H3₁₋₁₅K4me1/2/3, and 3.0mM for unmodified H3₁₋₁₅ peptide. A summary of thermodynamic and curve fitting parameters are listed in the bottom table.



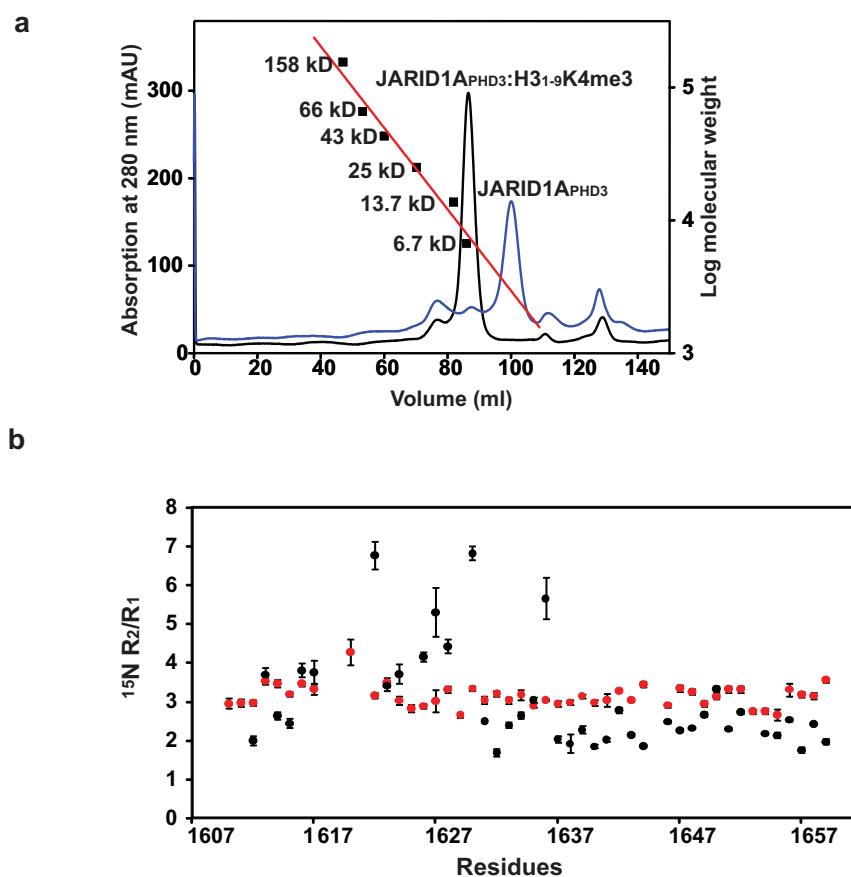
Supplementary Figure 4| Microscopy analyses of localization of NJL fusion protein and histone modification mark. a-b, Deconvolved microscopy images revealing that NJL (anti-FLAG, red) significantly colocalized with H3K4me3 (green, panel a), but not H3K9me3 (green, panel b), in NJL-transformed hematopoietic progenitors. Scale bar, 5 micrometer. c, Comparison between deconvolved microscopy and confocal microscopy images. d, Confocal microscopy images revealing that the nuclear immunostaining pattern of NJL (anti-FLAG, red) largely co-localized with H3K4me3 staining (green) in marrow progenitors. e, Statistical analysis of the co-immunostaining correlation by using Pearson's Coefficient method. Plotted were relative intensities of anti-Flag signals and those of anti-H3K4me3 (left) or anti-H3K9me3 (middle) staining in a scatter plot. The table (right) shows a calculated Pearson's coefficient of 0.47 for anti-H3K4me3/anti-FLAG correlation that revealed a positive correlation, as well as a Pearson's coefficient of -0.01 for anti-H3K9me3/anti-FLAG correlation that proved lack of correlation. It should be noted that H3K4me3:FLAG correlation in immunostaining may be underestimated, as the PHD finger binding can interfere with the efficient antigen presentation of H3K4me3.



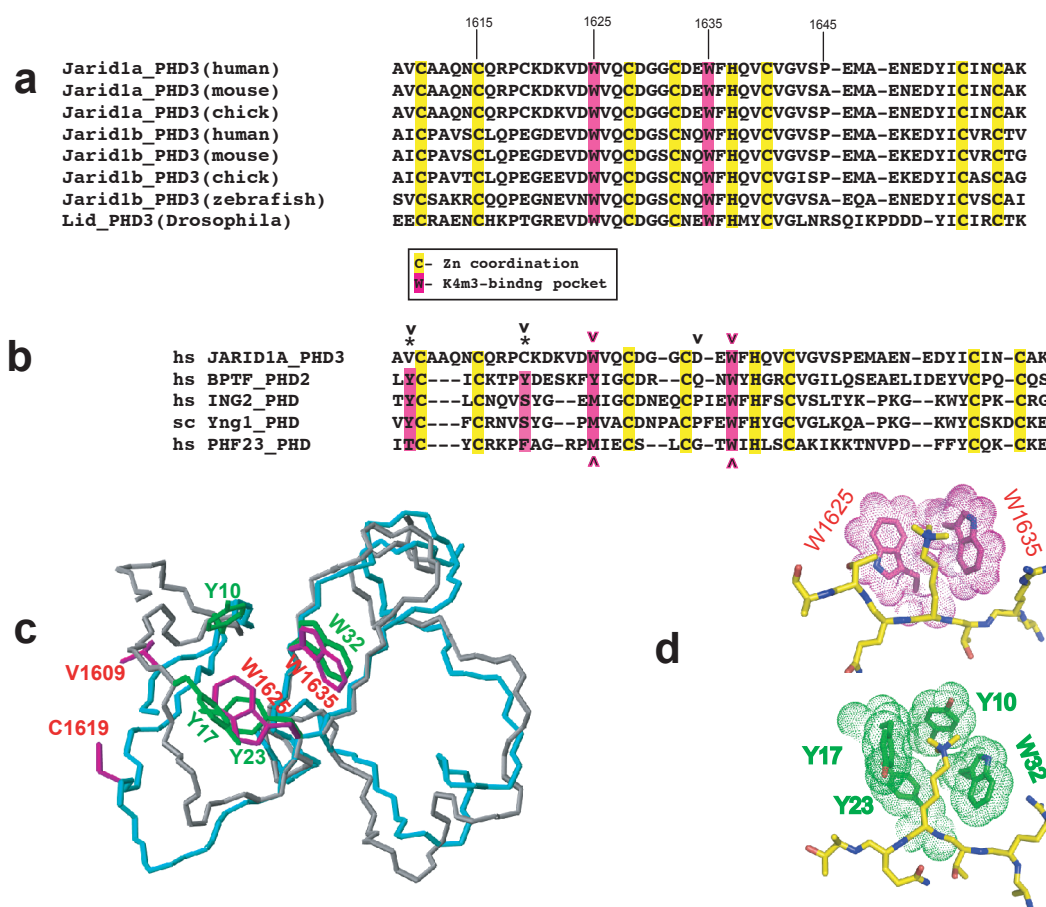
Supplementary Figure 5| The X-ray structure of JARID1A_{PHD3} in complex with H3K4me3 peptide at 1.90 Å resolution. a, The entire structure of the complex in the crystal showed that one molecule forms a domain-swapped dimer with a crystallographic symmetry-related molecule. The individual domains are colored green and cyan in a ribbon representation, together with the side chains of W1625 and W1635. The two Zn ions bound to each PHD finger are shown as colored balls. The H₃₁₋₉K4me3 peptide is colored in yellow, with its beta-strand in a ribbon representation, together with the side chain of H3K4me3. b, A view of the intermolecular contacts in the complex. The PHD finger is shown in an electrostatic surface representation (with red and blue representing acidic and basic patches) while the bound H₃₁₋₉K4me3 peptide is shown in a space-filling representation. This view is similar to the ribbon view of the protein in the complex shown in Fig. 2b. Note that the walls of the H3K4me3-binding surface channel involve Trp1635 (right wall) and Trp1625 (back wall), both from the protein, as well as the side chain of H3T6 (left wall) of the bound peptide. c, Zinc-induced crystal packing between domain-swapped dimers. Left, one domain-swapped dimer is colored in blue and green, and the other colored in yellow and salmon. Each color represents one JARID1A_{PHD3}-H3K4me3 complex. The zinc is located at the interface between the two domain-swapped dimers. Right, a close-up view of the zinc coordination. The zinc ion is chelated by residues S1608 and C1619 from one monomer, D1652* from an adjacent monomer and one water molecule. Note that the N-terminal S1608 (coming from the expression vector) engages zinc via a five-membered chelate.



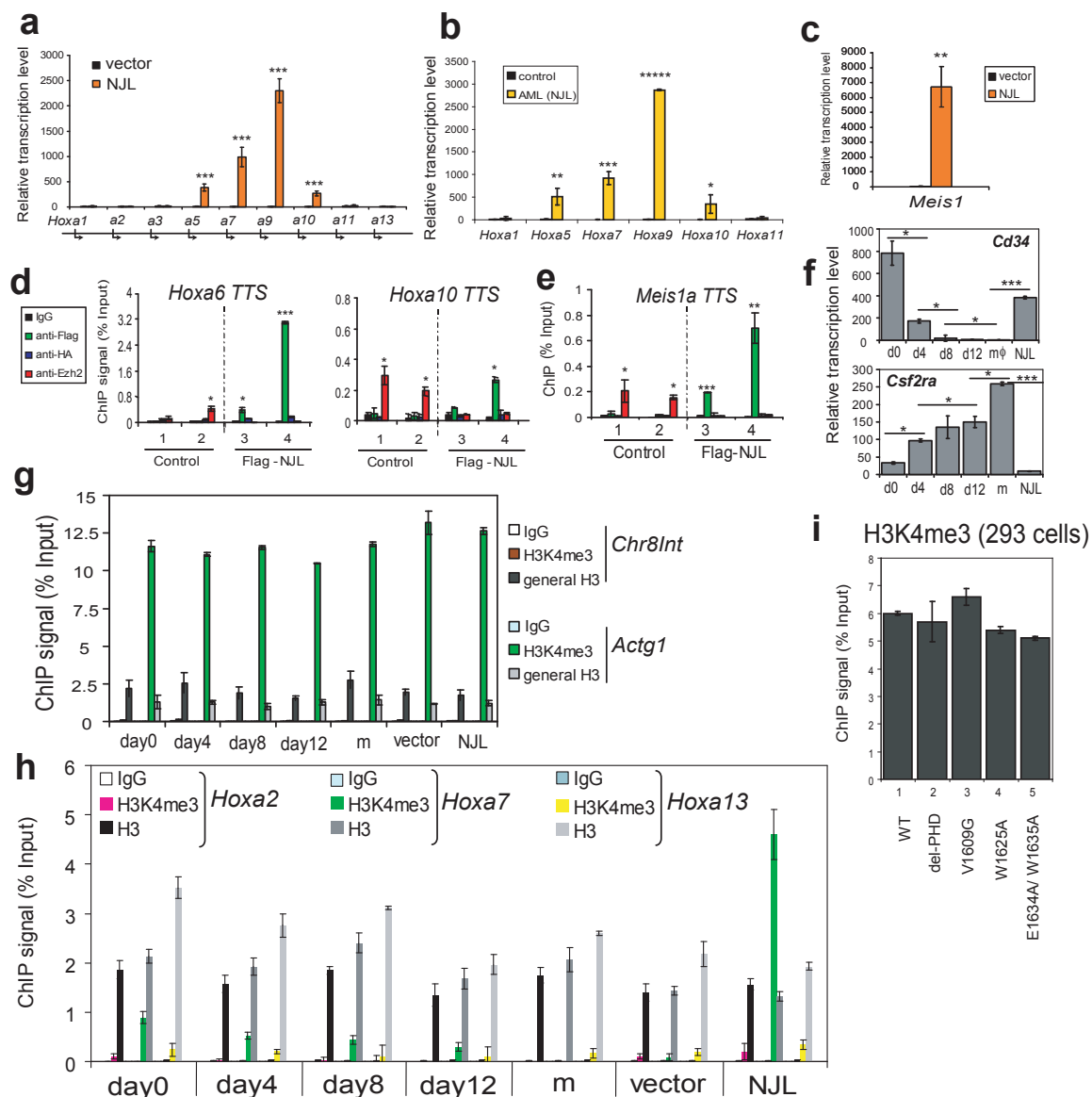
Supplementary Figure 6| NMR spectroscopic analysis of the JARID1A_{PHD3} structure in the free state and when bound to H3K4me3 peptide. a, Backbone superposition of 20 energy-minimized conformers representing the NMR structure of JARID1A_{PHD3} in the free state. The PHD finger is colored in green, aromatic side chains of H3K4me3-engaging Trp1625 and Trp1635 in purple, and zinc ion in gold sphere. b, Backbone superposition of 20 energy-minimized conformers representing the NMR structure of JARID1A_{PHD3} in complex with H3₁₋₉K4me3 peptide (H3 amino acids 1 to 9, containing H3K4me3). The side chains of H3K4me3 and H3 peptide backbone are colored in yellow. c, Ensemble of 20 NMR structures of JARID1A_{PHD3} in complex with H3₁₋₉K4me3 peptide, with side chains involved in complex formation with H3₁₋₉K4me3 peptide. Side chains from the PHD finger are colored in purple and the H3 peptide in yellow. d, Surface representation of JARID1A_{PHD3} in complex with H3₁₋₉K4me3. Surface residues that undergo the largest chemical shift upon binding are highlighted in pink.



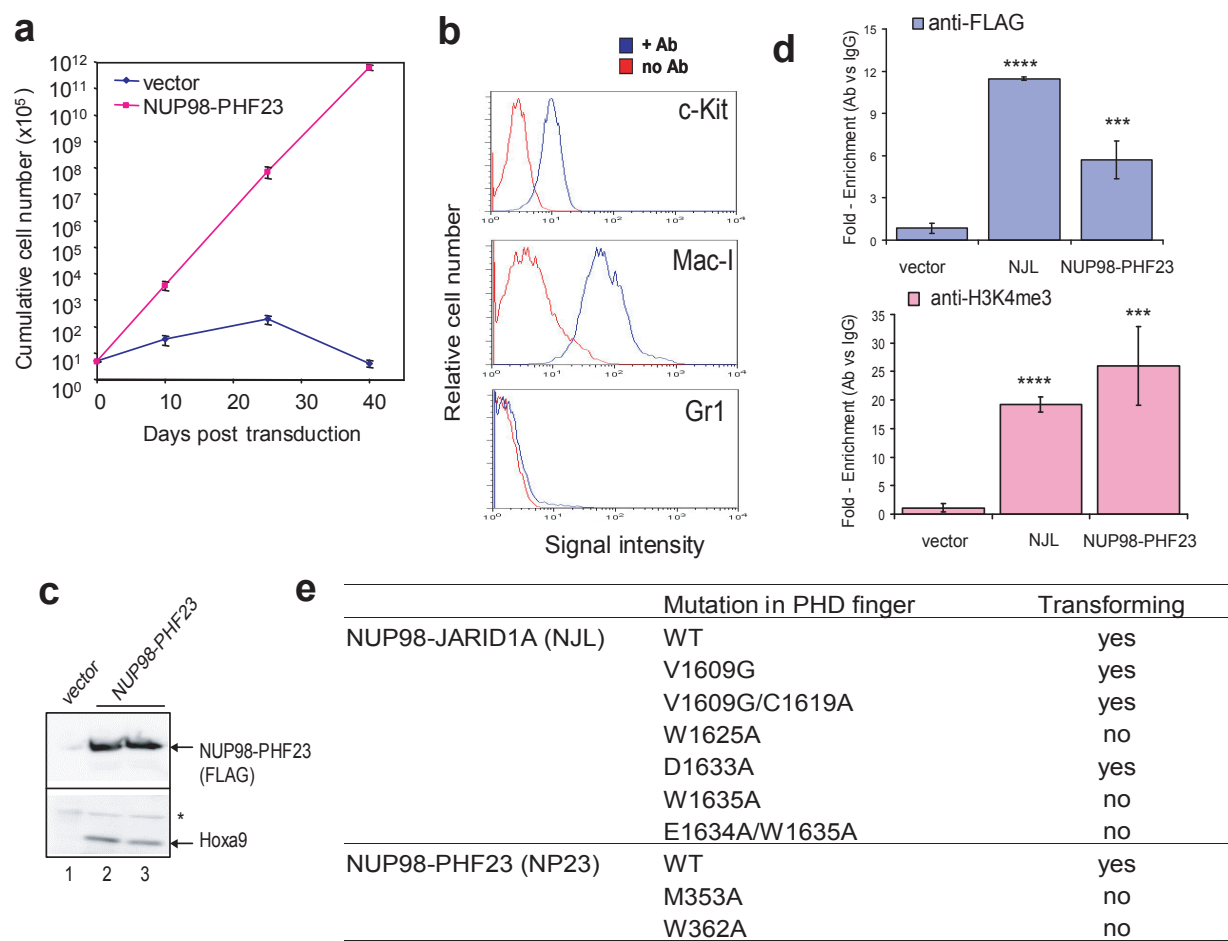
Supplementary Figure 7| Gel filtration and NMR relaxation studies established a monomeric state of JARID1A_{PHD3} in both the free and H3K4me3-bound states. a, The gel filtration profiles of free JARID1A_{PHD3} (blue) and JARID1A_{PHD3}-H3₍₁₋₉₎K4me3 complex (black), with the calibration line obtained from various molecular standards shown in red. **b,** Sequence dependent backbone ¹⁵N R₂/R₁ ratios of free (black) and H3₁₋₉K4me3-bound JARID1A_{PHD3} complex (red) at 600 MHz ¹H resonance frequency. The domain-swapped dimerization observed in the crystal is likely to be a characteristic feature of the crystalline state, originating perhaps in packing interactions.



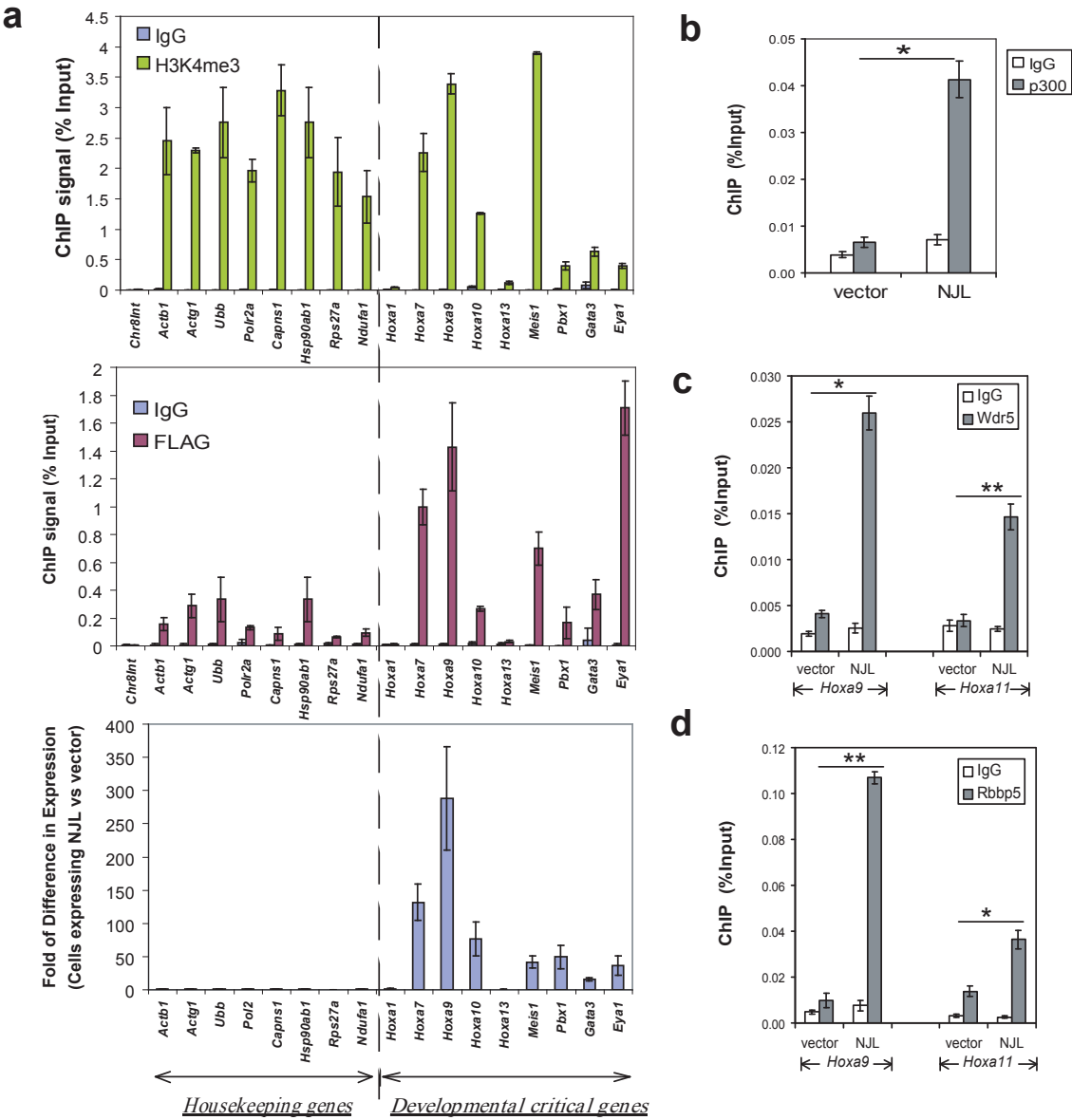
Supplementary Figure 8| Comparison of JARID1A_{PHD3} and other H3K4me3/2-binding PHD fingers. a, Sequence alignment of the C-terminal PHD finger among JARID1 homologues of human, chicken and fly revealed an evolutionary conservation of the two H3K4me3-engaging Trp residues (in magenta). Amino acid numeration of human JARID1A_{PHD3} was shown above the sequence. b, Sequence alignment of H3K4me3-binding PHD fingers of JARID1A, BPTF, ING2, Yng1 and PHF23. Stars indicate the position of V1609 and C1619, two JARID1A_{PHD3} residues that are not involved in H3K4me3 binding; Arrowhead indicates the position of mutation that interferes with (magenta ones) or does not affect (black) leukemic transformation shown in Supplementary Fig. 10e. c, Structural alignment of the backbone and H3K4me3-engaging residues of JARID1A_{PHD3} (in cyan) and BPTF_{PHD2} (in gray) revealed that V1609 and C1619 of JARID1A_{PHD3} do not participate in engaging H3K4me3, a feature different from PHD finger of BPTF, ING2 and Yng1. Structure and numeration of BPTF_{PHD} was adapted from previous study⁷. d, Comparison of positioning of the H3K4me3-engaging aromatic residues in the PHD fingers of JARID1A (in purple, top) and BPTF (in green, bottom). W1625 and W1635 of JARID1A_{PHD3} align well with Y17, Y23 and W32 of BPTF_{PHD2}.



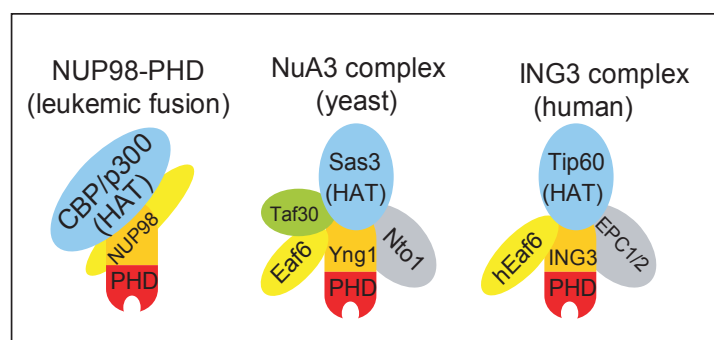
Supplementary Figure 9 | NUP98-JARID1A enforced the transcription of A-cluster *Hox* genes and *Meis1* and the H3K4me3 associated with their promoter. a, Quantitative RT-PCR of *Hox* A-clusters in hematopoietic progenitors 15 days after transduction of empty vector (black) or NJL (orange). *, $P < 0.01$; **, $P < 0.001$; ***, $P < 10^{-4}$; ****, $P < 10^{-5}$. b, RT-PCR of A-cluster *Hox* genes in leukemic cells derived from NJL-induced AML (orange) or in vector-infected control marrow cells (black). c, Quantitative RT-PCR of *Meis1* in hematopoietic progenitors 21 days after transduction of empty vector (black) or NJL (orange). d & e, ChIP of the recruitment of NJL (FLAG-tagged, green bar) or Ezh2-polycomb complex (red bar) to the transcriptional start site (TTS) of *Hoxa6*, *Hoxa10* and *Meis1* in marrow cells three weeks post transduction of empty vector (line 1), control neutrophil/macrophage progenitors³ (line 2), or two independent lines transduced with NJL (two and three weeks post transduction as for line 3 and 4, respectively). f, The expression level of *Cd34* (HSC marker) and GM-CSF receptor α (*Csf2ra*) in marrow lineage-negative progenitors during *in vitro* culture (day 0, 4, 8 and 12), macrophage (m) and NJL-transformed progenitors. g, ChIP of a silenced intragenic locus *Chr8Int* or a housekeeping gene *Actg1* promoter as negative or positive control for H3K4me3 in presented in Fig. 3e and Supplementary Fig. 9h. h, ChIP of H3K4me3 and general H3 at *Hoxa2*, *a7* and *Hoxa13* promoters among lineage-negative hematopoietic progenitors after days of *in vitro* cultivation, macrophage, and marrow cells 20 days post transduction of empty vector or NJL. i, Similar H3K4me3 level on *HOXA9* promoter in HEK293 cells after transient transfection of wildtype or mutant NJL.



Supplementary Figure 10| The clinically reported NUP98-PHF23 fusion² harbors the same leukomogenic potentials as NUP98-JARID1A. a, Altered proliferation kinetics of lineage-negative hematopoietic stem/progenitor cells in SCF-containing medium after retroviral transduction of either empty vector (blue) or NUP98-PHF23 (red). Error bar, s.d. of four different cultures. b, FACS revealed that NUP98-PHF23 transformed progenitors expressed the surface marker of myeloid progenitors (c-Kit+ MacI+) but not that of differentiated neutrophils (Gr1-). c, Immunoblot analysis of the NUP98-PHF23 fusion protein (Flag-tagged) and Hoxa9 in hematopoietic cells ten days post transduction of empty retroviral vector (lane 1) or that encoding NUP98-PHF23 (lanes 2-3). d, ChIP analysis revealed the recruitment of Flag-tagged NUP98-fusion proteins (left) and high H3K4me3 levels (right) on *Hoxa9* promoter in hematopoietic cells ten days post retroviral transduction. Cells transduced with empty vector were used as negative control. Y-axis represents fold of enrichment in specific signals versus those from non-specific antibodies. ***, $P < 10^{-4}$; ****, $P < 10^{-5}$. e, Summary of the effect of PHD finger mutations on the transforming ability of NJL or NUP98-PHF23. The numeration of residues is based on wildtype JARID1A or PHF23 (NCBI accession #117957305).



Supplementary Figure 11| NUP98-JARID1A perturbs epigenetic states of developmentally critical loci during hematopoiesis. a, ChIP analyses of NJL-recruitment (middle panel, anti-Flag) to the promoter of housekeeping genes and development regulator genes, both exhibiting high H3K4me3 (top) in NJL-transduced marrow progenitors. NJL specifically prevented the repression of developmental regulator genes (bottom) as measured by comparing gene transcript levels among marrow cells 20 days post transduction of NJL versus empty vector. Chr8Int, a silenced control locus⁸. Gene information is provided in Supplementary Table 5. b, ChIP analysis of the recruitment of p300 histone acetyltransferase onto the *Hoxa9* promoter in cultured bone marrow cells 15 days after transduction of empty vector (black) or NJL (grey). c-d, ChIP of the recruitment of MLL2-complex core components, WDR5 and RBBP5, on the *Hoxa9* and *Hoxa11* promoters in marrow cells 15 days after transduction of empty vector or NJL.



Supplementary Figure 12| A scheme of NJL-mediated leukemogenesis. Leukemic fusion oncoprotein NJL connects H3K4me3 to histone acetylation, mimicking mechanisms utilized by evolutionarily conserved NuA3/NuA4 (yeast Yng1⁹ and Yng2¹⁰) and ING-containing complexes (human ING3, ING4 and ING5)¹⁰ for achieving robust transcriptional activation.

References in Supplementary Figures

1. van Zutven, L.J. et al. Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes Chromosomes Cancer* 45, 437-46 (2006).
2. Reader, J.C., Meekins, J.S., Gojo, I. & Ning, Y. A novel NUP98-PHF23 fusion resulting from a cryptic translocation t(11;17)(p15;p13) in acute myeloid leukemia. *Leukemia* 21, 842-4 (2007).
3. Wang, G.G., Pasillas, M.P. & Kamps, M.P. Meis1 programs transcription of FLT3 and cancer stem cell character, using a mechanism that requires interaction with Pbx and a novel function of the Meis1 C-terminus. *Blood* 106, 254-64 (2005).
4. Wang, G.G. et al. Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. *Nat Methods* 3, 287-93 (2006).
5. Fischle, W. et al. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438, 1116-22 (2005).
6. Bernstein, E. et al. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* 26, 2560-9 (2006).
7. Li, H. et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442, 91-5 (2006).
8. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349-53 (2006).
9. Taverna, S.D. et al. Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol Cell* 24, 785-96 (2006).
10. Doyon, Y. et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* 21, 51-64 (2006).

Supplementary Table 1. Phenotypes of leukemia induced by NJL in murine bone marrow transplantation model.

<i>Mice transduced with</i>	<i>Latency (days)</i>	<i>Organ size (mg)</i>		<i>WBC count (1,000/μ)^b</i>	<i>FACS of leukemic cells^a</i>		
		<i>Spleen</i>	<i>Lymph Node</i>		<i>Mac-I+</i>	<i>B220+</i>	<i>Cd19+</i>
<i>NJL</i>	69 $\pm$ 12 ^c	737 $\pm$ 220	69 $\pm$ 46	113 $\pm$ 57	89 $\pm$ 7	13 $\pm$ 10	5 $\pm$ 4
<i>NJS</i>	N.A.	105 $\pm$ 8	<2	3 $\pm$ 0.5	N.A.	N.A.	N.A.
<i>empty vector</i>	N.A.	107 $\pm$ 6	<2	N.A.	N.A.	N.A.	N.A.

^a FACS showing the percentage of staining-positive leukemia cells from either spleen and bone marrow of AML mice;

^b WBC, white blood cells in peripheral blood;

^c Values presented in form of averages $\pm$ standard deviation (SD);

N.A., not applicable.

Supplementary Table 2. Data collection and refinement statistics

Crystal	JARID1A with H3 ₁₋₉ K4me3	JARID1A with H3 ₁₋₉ K4me3*
Beam line	APS-24ID-C	APS-24ID-C
Wavelength	0.97949	1.28215
Space group	I4 ₁	I4 ₁
Unit cell		
a, b, c (Å)	49.95, 49.95, 86.45	50.58, 50.58, 86.14
A, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50-1.90 (1.97-1.90)	50-2.20 (2.28-2.20)
R _{sym}	0.065 (0.180)	0.105 (0.242)
I/σ (I)	55.5 (6.46)	51.1 (7.0)
Completeness (%)	97.8 (85.1)	96.0 (76.5)
Redundancy	8.0 (4.4)	6.1 (5.6)
Number of unique reflections	8182	5333
R _{work} /R _{free} (%)	20.8 (22.5)	20.0(23.4)
Number of non-H atoms		
Protein	467	454
Water	32	20
Other ligands	3	3
Average B factors (Å ²)		
Protein	37.7	29.8
Water	42.2	39.9
Other ligands	31.9	32.3
R.m.s. deviations		
Bond lengths (Å)	0.006	0.02
Bond angles (°)	1.43	1.72

* Data set used to build the initial model using zinc anomalous signals.

Supplementary Table 3.**Statistics for twenty energy-minimized NMR-based solution conformers of JARID1A_{PHD3} in the free state and in complex with H3₁₋₉K4me3 peptide^a.**

Distance constraints	JARID1A _{PHD3} (free state)	JARID1A _{PHD3} -H3 ₁₋₉ K4me3 complex
JARID1A _{PHD3}		
Long [(i-j)>5] (intramolecular)	375	368
Medium [1<(i-j)≤5]	159	178
Sequential [(i-j)=1]	91	69
Intraresidue [i=j]	154	175
Intermolecular		84
H3 ₁₋₉ K4Me3		21
Intraresidue [i=j]		37
Sequential [(i-j)=1]		11
Dihedral angle constraints (φ and ψ)	57	59
Hydrogen bond constraints	6	11
Residual dipolar coupling constraints (¹ D _{NH})	44	47
Average pairwise R.M.S.D. to the mean structure (Å)		
Backbone (C ^α , C', N, O)	0.29 ± 0.09	0.41 ± 0.11
Heavy atoms	0.61 ± 0.10	0.80 ± 0.08
Deviations from idealized covalent geometry		
Bond (Å)	0.018 ± 0.001	0.018 ± 0.001
Angles (°)	1.398 ± 0.029	1.773 ± 0.264
Impropers (°)	1.523 ± 0.071	2.044 ± 0.213
R.M.S.D. from experimental distance restraints (Å)	0.020 ± 0.001	0.007 ± 0.001
R.M.S.D. from experimental dihedral restraints (Å)	0.489 ± 0.084	0.199 ± 0.024
Ramachandran statistics (% of all residues)		
Most favored	84.0	85.8
Additionally allowed	15.9	13.4
Generously allowed	0.1	0.9
Disallowed	0.0	0.0

^aStatistics are given for JARID1A_{PHD3} and residues 1-7 of H3₁₋₉K4Me3 peptide.

Supplementary Table 4. Genes up regulated in NJL-transformed hematopoietic stem/progenitor cells.

Gene ID ¹	Genbank number	Fold Change ²	NJL-transformed progenitor ³			Control progenitors ⁴			Gene description
			Normalized	Raw	SD_Raw	Normalized	Raw	SD_Raw	
<i>Hoxa9</i> ^{†‡}	NM_010456	160.80	165.33	4971.47	580.41	1.03	31.61	2.14	homeo box A9
<i>Eya1</i> ^{†‡}	BB760085	29.79	19.70	1008.27	644.57	0.66	27.82	6.25	homolog of <i>Drosophila</i> eyes absent 1
<i>Padi4</i>	NM_011061	28.52	6.75	1240.85	915.52	0.24	38.22	10.34	peptidyl arginine deiminase, type IV
<i>Hoxa10</i> ^{†‡}	NM_008263	18.78	13.21	214.13	95.19	0.70	10.81	1.58	homeo box A10
<i>Cmya4</i>	AV220213	16.38	16.95	305.29	164.96	1.03	16.60	1.26	Mus musculus unc-45 homolog B (C. elegans); cardiomyopathy-associated 4
<i>Serpina3g</i>	BC002065	16.25	10.86	1261.04	505.59	0.67	81.06	33.26	serine (or cysteine) proteinase inhibitor, clade A, member 3G
<i>Gata3</i> ^{†‡}	NM_008091	16.16	13.50	332.29	30.50	0.84	21.22	3.49	GATA binding protein 3
<i>Hoxa5</i> ^{†‡}	BC011063	12.85	11.47	337.92	37.79	0.89	27.25	5.77	homeo box A5
<i>Agtr1</i>	BB483357	10.21	10.55	232.41	120.00	1.03	21.43	2.97	angiotensin receptor-like 1
<i>Tp120b</i>	BM238658	9.84	9.27	491.60	101.15	0.94	51.18	10.43	TBP-interacting protein 120B (TIP120B)
<i>Ms4a7</i>	BC024402	8.96	4.84	814.49	567.19	0.54	68.53	7.39	membrane-spanning 4-domains, subfamily A, member 7
<i>Chd3</i>	NM_146019	8.39	4.87	1595.47	175.27	0.58	195.16	22.12	chromodomain helicase DNA binding protein 3
<i>Gem</i> [‡]	U10551	7.31	37.81	416.89	254.29	5.17	80.86	61.44	GTP binding protein (gene overexpressed in skeletal muscle)
<i>Stat4</i>	NM_011487	7.26	11.63	1068.03	309.05	1.60	162.52	96.77	signal transducer and activator of transcription 4
<i>Pbx1</i>	L27453	6.75	3.51	148.44	80.62	0.52	20.46	7.82	pre B-cell leukemia transcription factor 1
<i>Sh3d19</i>	NM_012059	6.45	9.02	304.82	211.71	1.40	49.00	43.51	SH3 domain protein D19
<i>2210010C04Rik</i>	AK008695	6.25	9.99	322.01	330.82	1.60	32.09	14.84	RIKEN cDNA 2210010C04 gene
<i>Hoxa7</i> ^{†‡}	NM_010455	6.12	5.27	1086.06	124.14	0.86	181.43	11.81	homeo box A7
<i>Plgir</i> [‡]	NM_008967	6.08	5.48	371.01	131.12	0.90	61.16	14.90	prostaglandin I receptor (IP)
<i>1700097N02Rik</i>	AA266367	5.57	9.70	876.31	420.94	1.74	149.71	33.58	RIKEN cDNA1700122G02; unknown EST
<i>Ivl</i>	AV009441	5.26	6.32	231.66	8.27	1.20	48.48	23.50	involucrin
<i>Agpt1</i>	NM_009640	5.04	9.50	691.40	251.29	1.88	189.55	134.91	angiopoietin 1
<i>Havcr2</i>	AF450241	4.91	5.35	353.07	206.40	1.09	75.35	42.09	hepatitis A virus cellular receptor 2
<i>Mbp</i> [‡]	AI323506	4.88	2.61	232.03	154.18	0.54	43.60	13.24	myelin basic protein
<i>Arhgap6</i> ^{†‡}	AF177664	4.87	5.64	114.48	24.11	1.16	25.79	11.27	Rho GTPase activating protein 6
<i>Cacna1b</i> [‡]	AV326040	4.86	11.07	109.95	43.92	2.28	24.29	12.90	calcium channel, voltage-dependent, N type, alpha 1B subunit
<i>Ms4a4c</i>	NM_022429	4.82	27.19	474.59	271.84	5.65	144.93	111.87	membrane-spanning 4-domains, subfamily A, member 4C
<i>Era</i> [†]	AV329219	4.80	69.17	2154.15	284.20	14.43	1170.96	998.99	avian erythroblastosis virus E-26 (v-ets) oncogene related
<i>Armxc1</i>	BC021410	4.78	4.99	108.26	74.24	1.04	20.32	5.83	armadillo repeat containing, X-linked 1;
<i>F2H3</i>	NM_007975	4.76	4.66	340.05	38.96	0.98	76.27	26.09	coagulation factor II (thrombin) receptor-like 3
<i>Dnahc8</i>	AF356522	4.73	4.64	485.75	268.09	0.98	113.81	90.02	dynein, axonemal, heavy chain 8
<i>Slco3a1</i> [‡]	NM_023908	4.71	2.08	323.89	212.05	0.44	61.57	16.19	solute carrier organic anion transporter family, member 3a1
<i>0610012H03Rik</i> [‡]	AK002603	4.63	2.19	176.87	126.38	0.47	27.97	5.68	hypothetical 4-hydroxybenzoyl-CoA thioesterase active site containing protein
<i>Krt80</i>	NM_028770	4.62	3.26	501.06	100.62	0.71	109.92	7.31	keratin 80
<i>Hpse</i> [‡]	BG094050	4.62	3.66	3269.04	990.83	0.79	767.42	340.48	heparanase
<i>Meis1</i> ^{†‡}	AW547821	4.61	16.20	333.19	21.83	3.51	104.95	75.09	myeloid ecotropic viral integration site 1
<i>Cobl1</i> [‡]	AV080881	4.58	9.54	86.54	17.44	2.08	21.13	10.01	Cobl-like 1
<i>Tacstd2</i> ^{†‡}	AV241768	4.40	6.80	251.15	62.93	1.55	75.21	50.31	tumor-associated calcium signal transducer 2
<i>Gpr56</i>	NM_018882	4.39	10.48	670.50	226.87	2.39	190.59	122.98	G protein-coupled receptor 56
<i>Itgb7</i>	NM_013566	4.35	9.22	2024.26	267.53	2.12	695.80	508.11	integrin beta 7
<i>Saa3</i>	NM_011315	4.33	3.90	1426.99	1008.86	0.90	314.82	270.98	serum amyloid A 3
<i>5033430I15Rik</i>	AK017206	4.24	5.08	340.85	241.27	1.20	67.03	13.89	RIKEN cDNA 5033430I15 gene, hypothetical protein LOC76000
<i>Rai14</i>	NM_030690	4.24	4.56	786.37	290.81	1.08	196.39	84.44	retinoic acid induced 14
<i>Armxc2</i>	BB392869	4.22	4.00	91.12	109.03	0.95	12.27	1.06	armadillo repeat containing, X-linked 2;
<i>Kcnn4</i>	NM_008433	4.16	9.30	627.22	72.53	2.23	210.18	157.32	potassium intermediate/small conductance Ca-activated channel, subfamily N4
<i>Hsh2d</i>	BG091940	4.16	10.75	442.70	38.77	2.59	162.92	122.51	hematopoietic SH2 protein
<i>Pdgfrb</i>	AA499047	4.10	9.78	180.77	202.21	2.38	28.76	18.15	platelet derived growth factor receptor, beta
<i>Hdc</i>	BG072171	4.08	11.00	470.32	325.59	2.70	99.37	18.23	histidine decarboxylase
<i>6720489N17Rik</i>	BB076796	4.07	4.21	103.26	21.72	1.03	26.91	10.27	RIKEN cDNA 6720489N17 gene
<i>Mvri1</i>	U63408	4.05	4.16	146.86	121.60	1.03	28.60	0.71	MRV integration site 1
<i>Dnmt3a</i>	BB795491	4.00	3.56	951.56	165.03	0.89	242.07	24.72	DNA methyltransferase 3A

¹ Bold/Underlined are genes that have been reported or implicated to be critical for the hematopoietic development or the leukemia induction;

[†] Genes targeted by polycomb proteins in embryonic stem (ES) cells (Lee TI, et al. *Cell*. 2006,125(2):301-13; Boyer LA, et al. *Nature*. 2006,441(7091):349-53).

[‡] Loci exhibiting “chromatin bivalent domain” pattern in ES or MEF cells (Mikkelsen TS, et al. *Nature*. 2007;448(7153):553-60).

² Fold change represents fold of change in normalized signals among NJL-transformed progenitors versus those among control progenitors;

³ Normalized represents scores after initial normalization of raw hybridization signals obtained from multiple probes on Affymetrix mouse gene array 430 (for same gene) and subsequent normalization against the medium value of hybridization intensities across gene arrays by using GeneSpring CX 7.0 microarray analysis platform; Raw, average value of raw hybridization signals; SD_raw, standard deviation of raw values;

⁴ Control progenitors were committed neutrophil/macrophage progenitor cell lines generated by using exogenously expressed leukemic oncogenes that block myeloid progenitor differentiation as previously described (Reference: Calvo KR, Sykes DB, Pasillas M, Kamps MP. *Mol Cell Biol*. 2000, 20(9):3274-85; Wang GG, Pasillas MP, Kamps MP. *Blood*. 2005, 106(1):254-64; Wang GG, Calvo KR, et al. *Nat Methods*. 2006,3(4):287-93).

Supplementary Table 5. Primers used for real-time PCR analyses following the RT-PCR or ChIP experiment.

RT-PCR primer

Gene ID	Organism	Gene Name	Target	Forward primer	Reverse primer	Tm	Ref.
<i>Hoxa1</i>	Mouse	Homoe box A1	cds	TTCTCCAGCGCAGACCTTG	GCACTGCGTTGGGTTGACC	60	1
<i>Hoxa2</i>	Mouse	Homoe box A2	cds	TTCCCAGTTTCGCCTTAACC	CAGTTCTGGCCCATTTGTGAC	60	2
<i>Hoxa3</i>	Mouse	Homoe box A3	cds	CCTTCCCTTTCTCCTCTGC	ACTGACAGCCTTCCAGCAAC	60	2
<i>Hoxa5</i>	Mouse	Homoe box A5	cds	GCAAGCTGCACATTAGTCAC	GCATGAGCTATTTGATCCT	60	3
<i>Hoxa7</i>	Mouse	Homoe box A7	cds	CGGGCTTATACAATGTCAACAG	AAATGGAATTCCTCTCCAGTTC	60	3
<i>Hoxa9</i>	Mouse	Homoe box A9	cds	ACAATGCCGAGAATGAGAGC	CAGCGTCTGGTGTGTTGTGT	60	
<i>Hoxa9</i>	Mouse		5'-UTR	CCTGATGGCGTGATTAATTG	CTGGCCTTGCCCTCTGACTC	60	
<i>Hoxa9</i>	Mouse		3'-UTR	AGGACTGTCCGTCTCCCTCT	GAGATGAGGCCTGGGATTAG	60	
<i>Hoxa10</i>	Mouse	Homoe box A10	cds	CTCCAGCCCCCTCAGAAAAC	TGTAAGGGCAGCGTTTCTTC	60	
<i>Hoxa11</i>	Mouse	Homoe box A11	cds	CACACTGAGGACAAGGCCG	AAGAACTCTCGTCCAGCTCTC	60	1
<i>Hoxa13</i>	Mouse	Homoe box A13	cds	CCAAATGACTGCCCAAAG	CCTATAGGAGCTGGCGTCTG	60	
<i>Pbx1</i>	Mouse	pre B-cell leukemia factor 1	cds	GCGCCGGGAGCCCATTTCTGC	GGTCCCTCCGGCCCCATCCTG	60	2
<i>Meis1</i>	Mouse	Myeloid ecotropic viral integration site 1	cds	AAGGTGATGGCTTGGACAAC	TGTGCCAACTGCTTTTCTG	60	
<i>Cd34</i>	Mouse	Cd34 antigen	cds	CTGGAATCCGAGAAGTGAGG	GCCTCCTCCTTTTCACACAG	60	
<i>Csf2ra</i>	Mouse	GM-CSF receptor α	5'-UTR	AAGCCCCCTGTCTCAGATG	TGAGTTCCCGGAAGCAGTAG	60	
<i>Gapdh</i>	Mouse	Glyceraldehyde 3-phosphate dehydrogenase	cds	ATGACATCAAGAAGGTGGTGAAG	TCCTTGGAGGCCATGTAGG	60	3

Supplementary Table 5. Continued.

ChIP primers (real-time PCR analyses)

Gene ID	Organism	Gene Name	Target	Forward primer	Reverse primer	Tm	Ref.
<u>Developmentally critical loci</u>							
<i>Hoxa1</i>	Mouse		Promoter	TCACTGAGTGATTGGATCCTGC	GGAGGAAGTGAGAAAGTTGGCAC	59	4
<i>Hoxa2</i>	Mouse		Promoter	GACAAGGTTGAAATTGGACCG	CAAATTGTCATTGGGCAGAAGC	59	4
<i>Hoxa4</i>	Mouse		Promoter	CTCTGGAATAAAACGAAGGAGGC	GGACAAAGAATCAAAGGGCGAG	59	4
<i>Hoxa5</i>	Mouse		Promoter	atggaactcgcagggaatg	cttcgcacctcgggcttc	59	
<i>Hoxa6</i>	Mouse		Promoter	CTTTCTTTTTTGCCTTCATGG	TTGTCAGGTTTCTGTTTGGG	59	4
<i>Hoxa7</i>	Mouse		Promoter	AACCCCTCCCTAAACGCCTC	AAAAGGTCGCCAGTCTTCCAG	59	4
<i>Hoxa9</i>	Mouse		Promoter	ATCTGTATGCCTAGTCCCGCTCC	TTGATGTTGACTGGCGATTTC	60	4
<i>Hoxa10</i>	Mouse		Promoter	CTGGCTCTTGAACCTGTACCCC	CAAGGGTGCTTCCAAATAGTC	60	4
<i>Hoxa11</i>	Mouse		Promoter	GGAAGCAACAGATCGTCACTCG	TGAGTTACACCGGCGATTACG	59	4
<i>Hoxa13</i>	Mouse		Promoter	CCTGTTGGTTCAGGAGAAGTC	CCAAACTTCCAGAGAAATGTCC	60	4
<i>Meis1</i>	Mouse		TSS	gcagttgcaagaggagag	gcccgcttccctgaatc	60	
<i>Pbx1</i>	Mouse		TSS	CTCCTCCTCCTCCTCTCAGG	CTCTTTCTCCTCCGTGTCTC	59	
<i>GATA3</i>	Mouse	GATA-binding protein 3	TSS	GGAATCAGTGTGCAAGTGTGG	GAAAAGAATGAGGGCGATCC	59	
<i>Eya1</i>	Mouse	Homolog of <i>Drosophila</i> eyes absent 1	TSS	GGGATGAGATTTTGCTGTTTCT	AGGCTTCCAGTCGGAAGTG	59	
<i>HOXA9</i>	Human		Promoter	acacacctccacctggtcac	ccaagcccagaattctcac	59	
<u>Housekeeping genes</u>							
<i>Actb1</i>	Mouse	Actin, beta	TSS	TTGATAGTTCGCCATGGATGACGA	ATCGATCCCCAAGAAAACCCCA	55	#
<i>Actg1</i>	Mouse	Actin, gamma, cytoplasmic 1	TSS	cgaggacattttctgagttcg	ATTTCTTCTTCCATTGCGATctg	58	
<i>Polr2a</i>	Mouse	Polymerase (RNA) II, polypeptide A	TSS	GCACCATCAAGAGAGTGACAG	gcatcactgacctatcttc	58	
<i>Ubb</i>	Mouse	Ubiquitin B	TSS	TTTCTGTGAGGGTGTTTCGAC	aagccccctctctaatacttc	58	
<i>Capns1</i>	Mouse	Calpain, small subunit 1	TSS	cgtaagtgaacgggacactc	atgggttgaatctgagaacacc	58	
<i>Hsp90ab1</i>	Mouse	Heat shock protein 1, beta	TSS	gttcgggagctctctGAGTC	GACGGGAAGCCGAGTAGTAG	58	
<i>Rps27a</i>	Mouse	Ribosomal protein S27a	TSS	aaatgcacctctgggacaagtag	aggcttcttatacagccgatg	58	
<i>Ndufa1</i>	Mouse	NADH dehydrogenase 1 alpha	TSS	ttagagatcccaagttctgaaac	acaggattctttctgtctctac	58	
<u>Silenced control locus</u>							
<i>Chr8Int</i>	Mouse	Chr. 8 Intragenic		AAGGGGCCTCTGCTTAAAAA	AGAGCTCCATGGCAGGTAGA	55	5

The information of promoter and transcriptional start site (TSS) was acquired from UCSC genomic browser (<http://genome.ucsc.edu>).

References:

- Christensen J, et al., *Cell* 128 (6), 1063 (2007);
 - Glaser S, et al., *Development* 133 (8), 1423 (2006);
 - Wang GG, et al., *Mol. & Cell. Biol.* 26 (10), 3902 (2006);
 - Bracken AP, et al., *Genes & dev.* 20 (9), 1123 (2006).
 - Boyer LA, et al., *Nature* 441 (7091), 349 (2006).
- #, Primers for *Actb1* were provided by A. Goldberg.