Article

Phase separation drives aberrant chromatin looping and cancer development

https://doi.org/10.1038/s41586-021-03662-5

Received: 17 January 2020

Accepted: 21 May 2021

Published online: 23 June 2021

Check for updates

Jeong Hyun Ahn[1,2], Eric S. Davis[3], Timothy A. Daugird[4], Shuai Zhao[1,2], Ivana Yoseli Quiroga[5], Hidetaka Uryu[1,2], Jie Li[1,6], Aaron J. Storey[7], Yi-Hsuan Tsai[1], Daniel P. Keeley[8], Samuel G. Mackintosh[7], Ricky D. Edmondson[7], Stephanie D. Byrum[7], Ling Cai[1,2,9], Alan J. Tackett[7], Deyou Zheng[10], Wesley R. Legant[4,11], Douglas H. Phanstiel[1,3,5,12,13] ✉ & Gang Greg Wang[1,2,4,6,13] ✉

The development of cancer is intimately associated with genetic abnormalities that target proteins with intrinsically disordered regions (IDRs). In human haematological malignancies, recurrent chromosomal translocation of nucleoporin (NUP98 or NUP214) generates an aberrant chimera that invariably retains the nucleoporin IDR—tandemly dispersed repeats of phenylalanine and glycine residues[1,2]. However, how unstructured IDRs contribute to oncogenesis remains unclear. Here we show that IDRs contained within NUP98–HOXA9, a homeodomain-containing transcription factor chimera recurrently detected in leukaemias[1,2], are essential for establishing liquid–liquid phase separation (LLPS) puncta of chimera and for inducing leukaemic transformation. Notably, LLPS of NUP98–HOXA9 not only promotes chromatin occupancy of chimera transcription factors, but also is required for the formation of a broad 'super-enhancer'-like binding pattern typically seen at leukaemogenic genes, which potentiates transcriptional activation. An artificial HOX chimera, created by replacing the phenylalanine and glycine repeats of NUP98 with an unrelated LLPS-forming IDR of the FUS protein[3,4], had similar enhancing effects on the genome-wide binding and target gene activation of the chimera. Deeply sequenced Hi-C revealed that phase-separated NUP98–HOXA9 induces CTCF-independent chromatin loops that are enriched at proto-oncogenes. Together, this report describes a proof-of-principle example in which cancer acquires mutation to establish oncogenic transcription factor condensates via phase separation, which simultaneously enhances their genomic targeting and induces organization of aberrant three-dimensional chromatin structure during tumourous transformation. As LLPS-competent molecules are frequently implicated in diseases[1,2,4–7], this mechanism can potentially be generalized to many malignant and pathological settings.

IDRs within various proteins—including transcription factors, chromatin modulators and RNA-binding proteins—form liquid droplets via phase separation, which affects myriad biological processes ranging from organelle formation and stress tolerance to gene transcription[4,5,8–10]. Notably, many cancers are characterized by recurrent fusions between genes encoding IDR-containing and chromatin-binding proteins. For instance, a subset of leukaemias that display poor prognosis carry a characteristic chromosomal translocation that produces a gene fusion between an IDR-containing segment of nucleoporin and a chromatin/DNA-binding factor[1,2,11,12]. Similarly, in Ewing's sarcoma, aberrant fusion occurs between transcription factors and the IDR of RNA-binding proteins[7]. Both chromatin-binding and IDR-containing domains were previously shown to be essential for tumorigenicity, which supports chromatin deregulation as a general mechanism[1,11,12]. However, how IDRs contribute to gene mis-regulation and oncogenesis is unclear.

IDRs induce transcription factor phase separation

We aimed to define the role for IDR and potentially phase separation in tumorigenicity by characterizing the NUP98–HOXA9 protein fusion, which

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [2]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [3]Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [4]Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [5]Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [6]Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [7]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA. [8]UNC Neuroscience Center and Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. [9]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [10]Departments of Genetics, Neurology, and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA. [11]Joint Department of Biomedical Engineering, University of North Carolina, Chapel Hill, and North Carolina State University, Raleigh, NC, USA. [12]Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [13]These authors jointly supervised this work: Douglas H. Phanstiel, Gang Greg Wang. ✉e-mail: douglas_phanstiel@med.unc.edu; greg_wang@med.unc.edu

shares similarity with other NUP98–transcription factor chimeras identified from various leukaemia subtypes[1,2]. NUP98–HOXA9 contains two protein motifs from NUP98—dispersed phenylalaine and glycine (FG) repeats and a GLE2-binding sequence (GLEBS) (Extended Data Fig. 1a). Deletion of GLEBS did not interfere with NUP98–HOXA9-mediated transformation of primary haematopoietic stem and progenitor cells (HSPCs) (Extended Data Fig. 1b, c). Normally, NUP98 is mainly localized at the nuclear periphery. Live-cell imaging showed that full-length and GLEBS-deleted NUP98–HOXA9 displayed a pattern of nucleoplasmic puncta (Extended Data Fig. 1d). Immunoblotting showed that the levels of NUP98 and NUP98–HOXA9 were comparable (Extended Data Fig. 1d). Thus, NUP98–HOXA9-mediated HSPC transformation and condensate formation are GLEBS-independent. To investigate the role for the NUP98 IDR in leukaemogenesis, we mainly used GLEBS-deleted NUP98–HOXA9 (hereafter referred to as N-IDR$_{WT}$/A9) (Fig. 1a, b).

To determine whether N-IDR$_{WT}$/A9 puncta are established via LLPS, we used several approaches[8–10,13]. First, we found that N-IDR$_{WT}$/A9 puncta were sensitive to treatment with 1,6-hexanediol, a chemical used to disrupt phase-separated condensates[8–10,13] (Fig. 1c). Second, the purified NUP98 IDR (N-IDR) proteins formed liquid condensates in vitro (38× FG) (Fig. 1d). To further assess concentration dependency and importance of multivalency conferred by FG-repeats for condensate formation, we generated recombinant N-IDR proteins that contained a varying number of FG repeats (Extended Data Fig. 1e, f). While N-IDR containing 38× or 36× FG repeats formed liquid droplets in a concentration-dependent fashion (Fig. 1d), those with 27× or 11× FG repeats were unable to phase separate under the same conditions (not shown). Only with the assistance of a crowding agent and at higher concentrations was the 27× FG-repeat-containing N-IDR able to establish condensates in vitro (Fig. 1d). However, when mixed with N-IDR proteins that contained 38× FG repeats, those with 11× or 27× FG repeats were readily incorporated into formed condensates in vitro (Fig. 1e). Imaging of cells expressing N-IDR/A9 with the varying FG-repeat number corroborated in vitro findings—compared with chimeras with 38× or 36× FG repeats, those with fewer FG repeats formed fewer condensates in cells (27×) or could not at all (11×), which is similar to that seen with the HOXA9 fusion segment alone (Fig. 1f, Extended Data Fig. 1g). In addition, DNA binding is dispensable for forming LLPS-like NUP98–HOXA9 puncta. Relative to N-IDR$_{WT}$/A9, its DNA-binding-defective form (carrying an N51S homeodomain mutation[14,15]) formed considerably fewer but larger puncta (Extended Data Fig. 1a, d, h), which were also readily detected as droplet-like nuclear structures even under the phase-contrast microscope (Fig. 1g). This indicates that chromatin binding of NUP98–HOXA9 may spatially restrict condensates from further coalescence, which occurs more readily with the N51S-mutant puncta. Condensates of NUP98–HOXA9(N51S) were also sensitive to 1,6-hexanediol treatment (Extended Data Fig. 1h). Notably, live-cell imaging after induction of GFP–NUP98-HOXA9(N51S) showed events of coalescence in which several small condensates collided producing a larger one (Fig. 1h, Supplementary Video 1), which is a characteristic of liquid condensates[13]. Together, IDR within NUP98–HOXA9 establishes LLPS in a valency-dependent and concentration-dependent manner.

IDRs in transcription factors drive oncogenesis

To investigate the roles for IDR and LLPS in leukaemogenesis, we mutated phenylalanine in the FG repeats of chimeras to serine (Fig. 1a)—a mutation previously shown to disable hydrogel formation by FG repeats in vitro[16]. Such Phe-to-Ser mutations did not affect the protein stability but abolished the nucleoplasmic droplet formation by N-IDR$_{WT}$/A9 carrying either wild-type or N51S-mutated homeodomain, which supports a crucial requirement of FG repeats for LLPS in cells (Fig. 1b, c, Extended Data Figs. 1h, 2a, b). NUP98–HOXA9 was reported to interact, either directly or indirectly, with coactivators such as CBP–p300[17] and MLL–NSL complexes[18]. We next queried whether

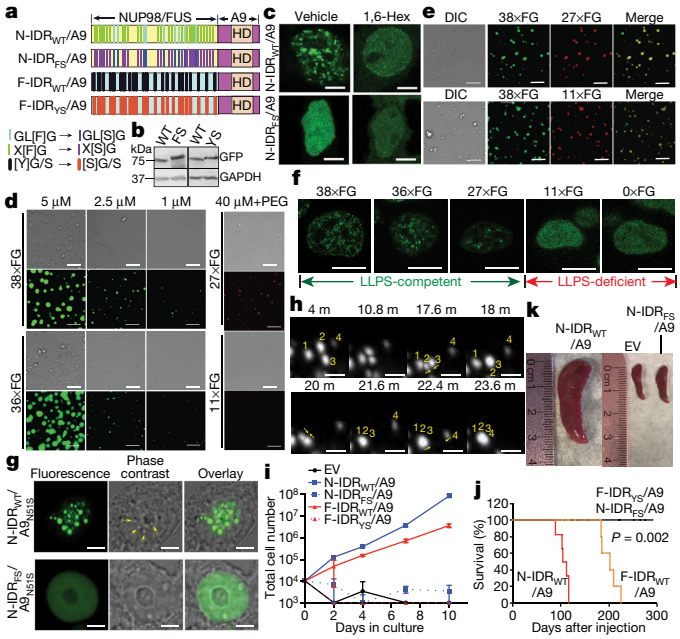


Fig. 1 | IDRs within chimeric transcription factor oncoproteins establish phase-separated assemblies, inducing leukaemogenesis. a, Scheme for N-IDR/A9 and F-IDR/A9 chimera, with the Phe-to-Ser and Tyr-to-Ser mutations introduced to the NUP98 and FUS IDRs, respectively. HD, homeodomain. **b**, **c**, Immunoblotting (**b**; GADPH was used as a loading control) and live-cell fluorescence (**c**) for GFP-tagged chimera carrying the wild-type (WT) or mutant IDR in 293FT cells. 1,6-Hex, 1,6-hexanediol. Scale bars, 10 μm. For gel source data, see Supplementary Fig. 1. **d**, **e**, Differential interference contrast (DIC) and concurrent fluorescence imaging (bottom) of N-IDR recombinant proteins that contain varying number of FG repeats, prepared at the indicated concentration with either single protein species (**d**) or a mixture of the two (**e**). PEG, polyethylene glycol-3350. Scale bars, 10 μm. **f**, Live-cell imaging of GFP-tagged N-IDR/A9 with the indicated number of FG repeats. Scale bars, 10 μm. **g**, Live-cell imaging (GFP) and concurrent phase-contrast imaging for N51S-mutated GFP–NUP98–HOXA9 with either wild-type (top) or Phe-to-Ser-mutated IDR (bottom). Arrows indicate droplet-like structures. Scale bars, 10 μm. **h**, Coalescence of GFP–NUP98–HOXA9 condensates (N51S-mutated). Scale bars, 2 μm. **i**, Proliferation of mouse HSPCs transduced with empty vector (EV) or the indicated chimera ($n = 3$ independent biological replicates; data are mean ± s.d.). **j**, Kaplan–Meier survival plot of mice after transplantation of HSPCs transduced with the indicated chimera ($n = 5$ mice per group). *P* values were calculated by two-sided log-rank test. **k**, Splenomegaly associated with N-IDR$_{WT}$/A9-induced leukaemias, three months after transplantation of infected HSPCs into mice.

Phe-to-Ser mutations perturbed such interaction networks by using BioID and found that most N-IDR$_{WT}$/A9- and N-IDR$_{FS}$/A9-interacting proteins were shared, including all reported interactors and many general transcriptional machinery proteins (Extended Data Fig. 2c, Supplementary Table 1). To examine the relationship between IDR-mediated LLPS and leukaemogenesis further, we performed the retrovirus-mediated oncogene transduction and transformation assays with mouse HSPCs, and found that, unlike N-IDR$_{WT}$/A9 that efficiently formed nuclear condensates and had a potent HSPC-transforming capacity as previously described[19], the Phe-to-Ser mutant was unable to establish puncta in HSPCs, did not transform HSPCs in vitro, and was unable to induce leukaemia in vivo (Fig. 1i–k, Extended Data Fig. 2d–g). We further assessed the involvement of IDR and LLPS in leukaemogenesis with an artificial chimera termed F-IDR$_{WT}$/A9 by fusing the homeodomain of HOXA9 to an unrelated IDR of the FUS protein that can phase separate[20,21] (Fig. 1a, b). As expected, F-IDR$_{WT}$/A9 formed puncta in cells, a process that was suppressed by treatment with 1,6-hexanediol or a condensate-disrupting mutation[20] (F-IDR$_{YS}$/A9) (Fig. 1a, b, Extended

Data Fig. 2h). Consistent with NUP98–HOXA9, only the IDR-intact and not the Tyr-to-Ser mutant form of F-IDR/A9 caused leukaemic transformation in vitro and in vivo (Fig. 1i, j, Extended Data Fig. 2g, i). Altogether, LLPS-forming IDRs retained within chimeric transcription factors are essential for cancerous transformation.

IDRs enhance genomic binding of chimeras

NUP98–HOXA9 binds DNA via the homeodomain, causing gene deregulation during leukaemogenesis. Next, we assessed the effect of IDR-mediated phase separation on chromatin targeting of NUP98–HOXA9 by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq) to map genome-wide binding of LLPS-competent N-IDR$_{WT}$/A9 versus LLPS-incompetent N-IDR$_{FS}$/A9 in their corresponding stable expression cells. Here, 293FT cells provide a system for assessing direct gene-regulatory effects of NUP98–HOXA9, because its cellular state is relatively stable and not apparently altered after transduction of the chimera, in contrast to what was observed in HSPCs such as differentiation arrest[18,19] (Extended Data Fig. 2e, f). ChIP–seq using antibodies of different tags attached to N-IDR/A9 produced robust, highly correlated signals, whereas ChIP–seq with non-tagged cells generated almost no binding (Extended Data Fig. 3a–c). Both N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 showed preferential binding to intergenic and intronic enhancers, with binding most enriched in expected motifs of HOX-related transcription factors (Extended Data Fig. 3d–g). Despite shared features seen for their targeting, N-IDR$_{WT}$/A9 had a notably enhanced genomic occupancy relative to N-IDR$_{FS}$/A9, irrespective of peak subclasses defined by unsupervised clustering (Fig. 2a). Also, the broad and dense super-enhancer-like peaks are unique to N-IDR$_{WT}$/A9 (Supplementary Table 2) and enriched at development- and leukaemia-associated genes (Extended Data Fig. 3h), exemplified by *HOX*, *PBX3* and *MEIS1* (Fig. 2b, c, Extended Data Fig. 4a–e). Super-enhancer calling by N-IDR$_{WT}$/A9 or H3K27ac verified their dense binding at proto-oncogenes (Extended Data Fig. 5a–c).

To further assess the role for IDR-induced LLPS in chromatin targeting of chimeras, we used several additional strategies. First, the treatment of 1,6-hexanediol markedly decreased chromatin occupancy of N-IDR$_{WT}$/A9, whereas it had minimal effects on the overall binding of N-IDR$_{FS}$/A9 (Fig. 2d, Extended Data Fig. 5d, e). Treatment with 1,6-hexanediol also suppressed the formation of a vast majority of broad N-IDR$_{WT}$/A9 peaks (Extended Data Fig. 5f, Supplementary Table 2). As a result, overall binding of N-IDR$_{WT}$/A9 after treatment with 1,6-hexanediol more closely resembled that of LLPS-incompetent N-IDR$_{FS}$/A9, compared with N-IDR$_{WT}$/A9 without treatment (Extended Data Fig. 5g). Second, we turned to F-IDR/A9 and tested whether the FUS IDR is sufficient to enhance genomic binding of the chimera. ChIP–seq analysis revealed that these two chimeras carrying unrelated LLPS-competent IDRs showed similar binding patterns—F-IDR$_{WT}$/A9 shows significantly enhanced genomic targeting and broad binding at AML-related oncogenes, in contrast to F-IDR$_{YS}$/A9 (Fig. 3a, b, Extended Data Figs. 4, 6a, b, Supplementary Table 3). ChIP–seq for N-IDR$_{WT}$/A9 in mouse leukaemias uncovered similar super-enhancer-like peaks at oncogenes, which overlapped those found in 293FT cells (Extended Data Fig. 6c–e). ChIP combined with quantitative PCR (ChIP–qPCR) verified the enhanced enrichment of N-IDR$_{WT}$/A9 and F-IDR$_{WT}$/A9, relative to their corresponding IDR mutant, and suppressive effect by 1,6-hexanediol on binding of N-IDR$_{WT}$/A9, but not its LLPS-defective mutant, to the tested loci (Extended Data Fig. 6f, g). Third, we used cells that expressed NUP98–HOXA9 with varied numbers of FG repeats, which were either LLPS-competent or LLPS-incompetent, and ChIP–qPCR detected significantly enhanced enrichment of LLPS-competent and not LLPS-incompetent fusions at loci that show broad N-IDR$_{WT}$/A9 binding (Fig. 3c), which indicates a crucial FG-repeat number required for establishing LLPS and intensified binding of chimeras. Lastly, we conducted single-molecule imaging studies to evaluate chromatin occupancy of N-IDR$_{WT}$/A9 relative to N-IDR$_{FS}$/A9.

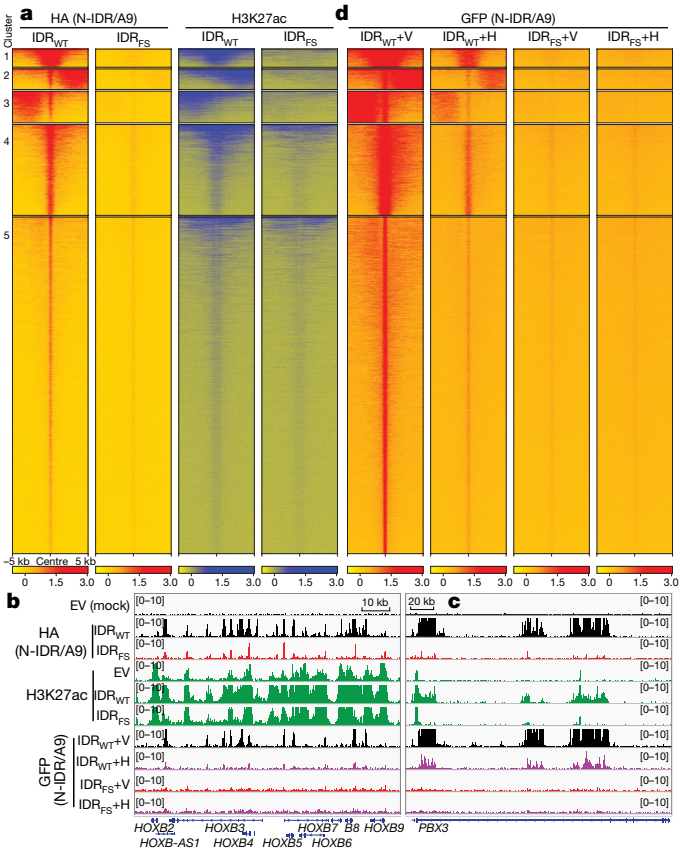


Fig. 2 | Phase separation markedly enhances chromatin binding of NUP98–HOXA9, featured with broad, super-enhancer-like genomic occupancy. **a**, **d**, Heat maps for *k*-means clustering of ChIP–seq signals in 293FT cells that express haemagglutinin (HA)-tagged (**a**; input-normalized) or GFP-tagged (**a**; spike-in control normalized) N-IDR/A9 with either wild-type or Phe-to-Ser-mutated IDRs. Cells in **d** were treated with 10% of 1,6-hexanediol (+H), compared with vehicle (+V), for 1 min. Each row represents a peak called for wild-type samples (first column) ± 5 kb from peak centre). **b**, **c**, IGV tracks of the indicated ChIP–seq signals at *HOXB* (**b**) and *PBX3* (**c**) in 293FT cells. EV-transduced cells act as a ChIP control.

Measurements of single-molecule speed and track displacement showed N-IDR$_{WT}$/A9 to be significantly less mobile than N-IDR$_{FS}$/A9 (Extended Data Fig. 7). Two-state kinetic modelling of single-molecule trajectories[22] showed that, compared with N-IDR$_{FS}$/A9, N-IDR$_{WT}$/A9 had a greater fraction of molecules in the low-diffusion bound state and had slower diffusion coefficients (Fig. 3d, Extended Data Fig. 7f, g), which suggests that assemblies of transcription factors, confined within phase-separated puncta, engage target DNA sequences more tightly and generally display slower diffusion, compared with LLPS-defective transcription factors. Collectively, using both genetic and pharmacological approaches, we have demonstrated a causal role for IDR-mediated LLPS in establishing enhanced targeting of chimeric transcription factors, particularly those seen at super-enhancer-like peaks.

IDRs potentiate target gene activation

To assess the relationship between NUP98–HOXA9 binding and gene activation, we conducted histone 3 Lys27 acetylation (H3K27ac) ChIP–seq and observed that increased chimera transcription factor binding is correlated with increased H3K27ac (Fig. 2a–c, Extended Data Fig. 4). Immunofluorescence also revealed co-localization of N-IDR$_{WT}$/A9 'dots' with H3K27ac, in comparison to H3K9me3 (Extended Data Fig. 8a, b). To define the role for IDR in target gene regulation further, we performed

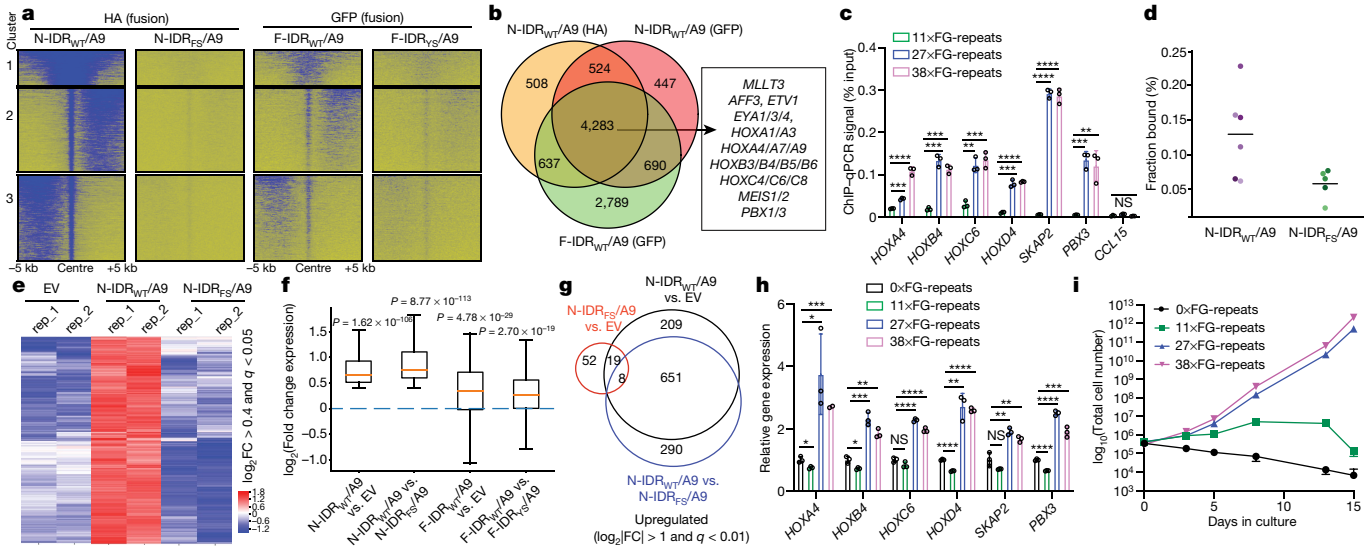


Fig. 3 | Creation of an artificial F-IDR/A9 chimera and alteration of the FG-repeat valency in NUP98–HOXA9 demonstrate a role for IDR and LLPS in promoting target oncogene activation and cancerous transformation. **a**, ChIP–seq signal heat maps showing N-IDR/A9 (HA-tagged; left) and F-IDR/A9 (GFP-tagged; right), either wild-type or IDR-mutated (FS or YS), in 293FT cells. See also Extended Data Fig. 6a. **b**, Venn diagram using direct targets of N-IDR$_{WT}$/A9 or F-IDR$_{WT}$/A9 in 293FT cells, with a battery of leukaemia-related oncogenes highlighted. *MLLT3* is also known as *AF9*. **c**, ChIP–qPCR for binding of GFP-tagged N-IDR/A9 with the indicated number of FG repeats at examined loci in 293FT cells ($n = 3$ independent samples; data are mean ± s.d.). CCL15 acts as a negative control for ChIP. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, two-sided *t*-test. **d**, Single-molecule imaging estimated the fraction of chromatin-bound N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 in 293FT stable cells. Presented are values based on two-state kinetic modelling (individual s.d. < 0.0003). Black bars denote mean values. **e**, Heat map of 303 genes upregulated in 293FT cells after transduction of N-IDR$_{WT}$/A9, compared with empty vector and N-IDR$_{FS}$/A9. **f**, Box plots showing relative expression of 303 N-IDR$_{WT}$/A9-activated genes in **e** among the indicated pairwise comparison of 293FT cells. Boxes extend from the first to the third quartile values of the dataset; line denotes median value; whiskers show the data range. *P* values were determined by two-sided *t*-test. **g**, Venn diagram using genes upregulated in mouse HPSCs after transduction of the indicated construct. **h**, RT–qPCR for oncogenes in 293FT cells expressing chimera with the indicated number of FG-repeats ($n = 3$ independent samples; data are mean ± s.d.). Expression was normalized to the 0 × FG-repeat sample. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, two-sided *t*-test. NS, not significant. **i**, Proliferation of mouse HSPCs transduced with N-IDR fusion carrying the indicated number of FG repeats ($n = 3$ independent replicates; data are mean ± s.d.).

RNA sequencing (RNA-seq) analysis in 293FT cells with stable chimera expression and identified 303 differentially expressed genes that were significantly upregulated by N-IDR$_{WT}$/A9, compared with mock treatment and N-IDR$_{FS}$/A9 (Fig. 3e, Supplementary Table 4), the effect confirmed by quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 8c). IDR-dependent gene activation was also observed in 293FT cells with expression of F-IDR$_{WT}$/A9 versus F-IDR$_{YS}$/A9 (Extended Data Fig. 8d, Supplementary Table 5), albeit gene activation of F-IDR$_{WT}$/A9 is less than that of N-IDR$_{WT}$/A9 (Fig. 3f), in agreement with a relatively less oncogenic potency by the former in vivo (Fig. 1i, j). In addition, RNA-seq of mouse HSPCs transduced with fusion relative to mock control corroborated that N-IDR$_{WT}$/A9, but not N-IDR$_{FS}$/A9, sustains oncogenic gene-expression programs, which again include *HOX*, *MEIS* and *PBX* family genes and other signatures related to leukaemia and HSPCs (Fig. 3g, Extended Data Fig. 8e, f, Supplementary Table 6); as expected, differentiation-related gene sets were suppressed in the N-IDR$_{WT}$/A9 sample (Extended Data Fig. 8f). Gene-regulatory effects of the artificial chimera F-IDR$_{WT}$/A9 were similar to those of N-IDR$_{WT}$/A9 in HSPCs (Extended Data Fig. 8g, Supplementary Table 7). Furthermore, a reduction in the FG-repeat number, which decreased LLPS competence, also significantly decreased the effects of the chimera on oncogene transcription and HSPC transformation (Fig. 3h, i). Thus, genomic profiling of independent models strongly supports a crucial role for IDRs in activating proto-oncogenes, many of which carry super-enhancer-like elements bound by chimeric transcription factors and H3K27ac.

IDRs and LLPS induce chromatin looping

Increasing evidence suggests that the phase separation of chromatin-associated factors can modulate gene transcription via alterations to three-dimensional chromatin structure[9,23–26]. However, so far there is little direct evidence that phase separation can form DNA loops similar to those created by CTCF and cohesin, nor that such phase separation-driven loops have a causal role in human disease. To test the ability of NUP98–HOXA9 to form chromatin loops via LLPS, we generated Hi-C profiles of 293FT cells that expressed either N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9, which revealed 6,615 DNA loops (Fig. 4a) and high correlation between replicates (Extended Data Fig. 9a, b). To determine the effect of N-IDR$_{WT}$/A9 on Hi-C contact frequency, we aggregated the interaction counts between the 500 most strongly N-IDR$_{WT}$/A9-occupied sites for both N-IDR$_{WT}$/A9- and N-IDR$_{FS}$/A9-expressing cells. Regions with high occupancy of N-IDR$_{WT}$/A9 exhibited increased interaction frequencies, even between binding sites separated by great distances (greater than 2 Mb) or on different chromosomes entirely (Fig. 4b). Increased interaction frequencies were not observed between the same loci in cells expressing N-IDR$_{FS}$/A9 (Fig. 4b). Differential analysis revealed 232 loops specific to N-IDR$_{WT}$/A9 and 52 specific to N-IDR$_{FS}$/A9 (DESeq2, $P < 0.01$) (Fig. 4a, c–e). Most (91%) N-IDR$_{WT}$/A9-specific-loop anchors overlapped N-IDR$_{WT}$/A9 binding, whereas only 31% overlapped a CTCF-binding site (Fig. 4f, Extended Data Fig. 9c). Thus, N-IDR$_{WT}$/A9 loops form in a largely CTCF-independent manner, consistent with a phase-separation-driven mechanism. Chromatin conformation capture (3C) followed by qPCR (3C–qPCR) after treatment with 1,6-hexanediol showed that the N-IDR$_{WT}$/A9-specific loop at *PBX3*, but not an unrelated CTCF loop, was significantly disrupted (Extended Data Fig. 9d–g). The vast majority (82%) of N-IDR$_{WT}$/A9-specific-loop anchors overlapped H3K27ac, in contrast to only 31% observed for non-differential loop anchors (Fig. 4f), which suggests that N-IDR$_{WT}$/A9-specific loops rewire connections between enhancers and target genes. Indeed, genes with promoters that overlapped N-IDR$_{WT}$/A9-specific-loop anchors exhibited

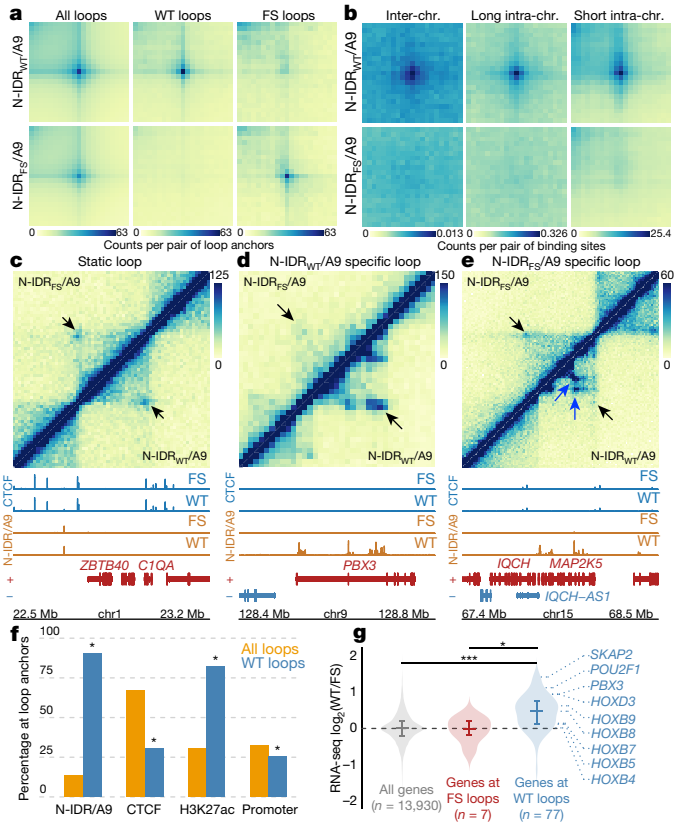


Fig. 4 | Phase-separation-competent IDRs within NUP98–HOXA9 induce CTCF-independent looping at oncogenes. a, Aggregate peak analysis (APA) for all loops ($n = 6{,}615$), Wild-type-specific ($n = 232$) and FS-specific ($n = 52$) loops defined by Hi-C in 293FT cells expressing N-IDR$_{WT}$/A9 (top) or N-IDR$_{FS}$/A9 (bottom). Pixel colour represents the mean interaction counts per loop, plotted on a common scale. **b**, APA plots at 10-kb resolution for interactions between the 500 strongest N-IDR/A9 binding sites in cells with N-IDR$_{WT}$/A9 (top) or N-IDR$_{FS}$/A9 (bottom). Paired interactions were categorized as inter-chromosomal ($n = 95{,}959$), long (≥2 Mb) intra-chromosomal ($n = 6{,}298$), or short (<2 Mb) intra-chromosomal ($n = 574$). Pixel colour represents the mean interaction counts per pair of loci interrogated. Colour scale in each plot is adjusted to the maximum value. **c**–**e**, Non-differential static (**c**), N-IDR$_{WT}$/A9-specific (**d**; 'Gained in WT' at PBX3) and N-IDR$_{FS}$/A9-specific loop (**e**; 'Lost in WT') detected by Hi-C (arrowheads in top panel) with 293FT cells expressing N-IDR$_{WT}$/A9 (below diagonal) or N-IDR$_{FS}$/A9 (above diagonal). Bottom panels show CTCF (blue) and N-IDR/A9 (orange) ChIP–seq signals (gene tracks shown below) in same cells. **f**, Percentage of the indicated feature present at either all loops or WT-specific loops. *$P < 0.001$, permutation test (Methods). **g**, Relative expression of genes associated with wild-type-specific ($n = 77$) and FS-specific loops ($n = 7$) in 293FT cells expressing N-IDR$_{WT}$/A9 versus N-IDR$_{FS}$/A9. *$P < 0.05$; ***$P < 0.00001$, Benjamini–Hochberg-adjustment method.

increased expression in N-IDR$_{WT}$/A9-expressing cells, compared to those with N-IDR$_{FS}$/A9, which further supports a regulatory role of these loops (Fig. 4g). The upregulated genes at N-IDR$_{WT}$/A9-specific-loop anchors include proto-oncogenes such as *HOX* and *PBX3* (Fig. 4d, g, Extended Data Fig. 10a–c). These results support the idea that IDRs of chimeric transcription factors induce DNA looping between super-enhancer-like targeting sites and oncogenes via phase separation.

Discussion

In summary, we show that the LLPS-competent IDR contained within NUP98–HOXA9 is crucial for leukaemogenesis and activation of the oncogenic gene-expression program. These effects are mediated by the ability of the IDR to (1) enhance transcription factor binding to genomic targets, and/or (2) promote long-distance looping between enhancers and oncogene promoters (Extended Data Fig. 11). We demonstrated these effects by both genetic (IDR mutagenesis or replacement with an unrelated one and changing the FG-repeats valency) and pharmacological methods. This study provides a proof-of-principle example of an oncogenic mutation that promotes LLPS-driven transcription factor binding and 3D chromatin reorganization during transformation of tumours. As a wide range of IDR-containing LLPS-competent molecules are implicated in diseases[1,2,4–7], this mechanism can potentially be generalized to many pathological settings.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03662-5.

1. Gough, S. M., Slape, C. I. & Aplan, P. D. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. *Blood* **118**, 6247–6257 (2011).
2. Mendes, A. & Fahrenkrog, B. NUP214 in leukemia: it's more than transport. *Cells* **8**, E76 (2019).
3. Murray, D. T. et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017).
4. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 196–213 (2021).
5. Boija, A., Klein, I. A. & Young, R. A. Biomolecular condensates and cancer. *Cancer Cell* **39**, 174–192 (2021).
6. Wan, L. et al. Impaired cell fate through gain-of-function mutations in a chromatin reader. *Nature* **577**, 121–126 (2020).
7. Kovar, H. Dr. Jekyll and Mr. Hyde: the two faces of the FUS/EWS/TAF15 protein family. *Sarcoma* **2011**, 837474 (2011).
8. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
9. Nair, S. J. et al. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* **26**, 193–203 (2019).
10. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
11. Wang, G. G. et al. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847–851 (2009).
12. Jankovic, D. et al. Leukemogenic mechanisms and targets of a NUP98/HHEX fusion in acute myeloid leukemia. *Blood* **111**, 5672–5682 (2008).
13. Pak, C. W. et al. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).
14. LaRonde-LeBlanc, N. A. & Wolberger, C. Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev.* **17**, 2060–2072 (2003).
15. 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).
16. Frey, S., Richter, R. P. & Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006).
17. 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).
18. Xu, H. et al. NUP98 fusion proteins interact with the NSL and MLL1 complexes to drive leukemogenesis. *Cancer Cell* **30**, 863–878 (2016).
19. Kroon, E., Thorsteinsdottir, U., Mayotte, N., Nakamura, T. & Sauvageau, G. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. *EMBO J.* **20**, 350–361 (2001).
20. Wang, J. et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16 (2018).
21. Qamar, S. et al. FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* **173**, 720–734.e15 (2018).
22. Hansen, A. S. et al. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7**, e33125 (2018).
23. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
24. Wang, L. et al. Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Mol. Cell* **76**, 646–659.e6 (2019).
25. Gibson, B. A. et al. Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484.e21 (2019).
26. Shin, Y. et al. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* **175**, 1481–1491.e13 (2018).

Methods

Plasmid construction

The MSCV-based retroviral vector for expression of NUP98–HOXA9 fusion has previously been described[27] and the mammalian expression constructs containing various tagged NUP98–HOXA9 (such as GFP–NUP98–HOXA9 in an inducible expression vector[28]) were gifts from M. Kamps, B. Fahrenkrog and J. Schwaller. The IDR (amino acids 1–215) of FUS can phase separate and is used for creating an artificial fusion of F-IDR/A9. To generate various chimera constructs of N-IDR/A9 or F-IDR/A9 fusions, we synthesized the gBlocks (IDT) that contain cDNA segments of both fusion partners fused in-frame, with a 3×HA-3×Flag tag added at the C terminus. Each gBlock fragment was cloned into the MSCV retroviral vector with a drug selection marker (Puro or Neo). For live-cell imaging studies, we replaced the 3×HA-3×Flag tag in fusion constructs with EGFP by subcloning. For generating a series of constructs with a varying number of NUP98 FG repeats, we used the following NUP98 portion as its fusion segment in the expression vector: amino acids 1–468 as 38 × FG repeats, 1–468(Δ132–224) as 36 × FG repeats, 65–468 (Δ132–224) as 27 × FG repeats and 357–468 as 11 × FG repeats. For bacterial expression of IDR, the same fragments with varying number of FG repeats were cloned into the pRSFDuet-1 vector (a gift from J. Song). For single-molecule tracking studies, we synthesized gBlocks (IDT) that contain cDNA segments of a HaloTag with flanking enzymatic sites of MluI and XhoI to replace the 3×HA-3×Flag tag described in the above expression vectors. All plasmids used were confirmed by sequencing before use and are listed in Supplementary Table 8.

Tissue culture and stable cell line generation

293FT (Thermo Fisher R70007), a fast-growing variant of the HEK293T cell line, and HeLa (ATCC CCL-2) cells were obtained from commercial vendors and maintained using recommended culture conditions. Authentication of cell identities, including those parental and derived lines, was ensured by the Tissue Culture Facility affiliated to UNC Lineberger Comprehensive Cancer Center with genetic signature profiling and fingerprinting analyses[29]. A routine examination for any possible mycoplasma contamination was performed every month with kits (Lonza). Cells in a passage of less than 10 were used. Retrovirus or lentivirus was packaged and produced in 293FT cells, and the stable cell lines were generated by viral infection followed by drug selection as previously performed[30,31]. The 293FT cell lines with stable expression of chimera carrying either wild-type or mutant IDRs were first examined by western blotting and immunofluorescence of the transgene, and the same sets of cells then used throughout this study for various assays such as live-cell imaging and genomic profiling (RNA-seq, ChIP–seq and Hi-C).

Antibodies and western blotting

Immunoblotting was performed as previously described[30,31]. Affinity-purified antibodies against endogenous NUP98 (raised in rabbits against NUP98 amino acids 51–223 covering GLEBS) was a gift from J. M. van Deursen and used as previously described[17,32]. The antibodies used (including the antibody source and dilution) are listed in Supplementary Table 8.

Fixed cell immunofluorescence

293FT cells were grown on polylysine-coated coverslips (Corning, 354085) for 24 h at a 37 °C incubator. For non-adherent mouse HPSCs, 0.1 million cells were added on top of polylysine-coated coverslips and centrifuged for 30 min at 1,600*g*. The cover slips were briefly washed with PBS and then fixed in 4% formaldehyde (Thermo Scientific, 28908) for 10 min at room temperature. Fixed cell samples were washed with cold PBS three times and incubated in PBS plus 0.1% Triton X-100 for 10 min, followed by washing with PBS for three times and incubation in blocking buffer (1% BSA in PBS plus 0.1% Tween-20) for 30 min. After discarding the blocking buffer, the fixed samples were incubated with a primary antibody diluted in the blocking buffer for 2 h at room temperature or overnight at 4 °C in a humidified chamber, and then washed with PBS plus 0.1% Tween-20 for three times (3 min each time). Lastly, the samples were incubated with the secondary antibody conjugated to appropriate fluorophores for 2 h at room temperature and washed three times with PBST before adding the mounting medium (Thermo Scientific, P36935). The slides were then dried overnight at dark before imaging on the Olympus FV1000 confocal microscope with a 100×/1.4NA Plan Apochromat oil immersion objective. DAPI was imaged with an excitation of 405nm and emission from 430–470nm, Alexa Fluor 488 was imaged with an excitation of 488nm and emission from 505–540nm, and Alexa Fluor 594 was imaged with an excitation of 559 nm and emission from 575–675nm.

Live-cell imaging

For live-cell imaging, cells were grown on 35-mm dish with 20-mm glass bottom well (Cellvis, D35-20-1.5-N) for 24 h before imaging. Live-cell imaging was conducted on Olympus FV1000 confocal microscope using 60× and 100× oil objectives. Three-dimensional lattice light sheet microscopy movies of fusion events were acquired on the lattice light sheet system as previously described[33] using a square lattice excitation with numerical apertures of 0.5 (outer) and 0.42 (inner). Three-dimensional volumes of cells, acquired every 24 seconds, were imaged by scanning the coverslip along the sample-plane axis and consisted of 140 planes spaced 360 nm apart. Raw data was de-skewed and deconvolved via 10 iterations of Richardson Lucy deconvolution, using an experimentally measured point spread function prior to quantification. To capture the events of coalescence in which multiple small liquid condensates of chimera are fused into a single one, we used Hela cells with stable expression of doxycycline-inducible GFP-tagged NUP98-HOXA9^{N51S} for live-cell imaging upon chimera expression induction.

Chemical treatment

To test the sensitivity of protein aggregates to 1,6-hexanediol treatment, 10% of 1,6-hexanediol (Sigma-Aldrich, 240117) were prepared in PBS. Throughout this study, the 1,6-hexanediol treatment condition was 10% for 1 min. Such 1,6-hexanediol-treated cells, together with the vehicle-treated control cells, were used for various experiments such as immediate imaging or fixation with 1% formaldehyde for subsequent ChIP–seq experiments.

Recombinant protein purification

For bacterial expression of IDR proteins, the His6× tag-containing pRSFDuet-1 vector that contains NUP98 segment covering FG repeats was transformed into Rosetta 2 (DE3) competent cells (Sigma, 71397). Three litres of bacterial cultures were grown at 37 °C for 12 h and then added with a final concentration of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for overnight induction at 16 °C. Bacterial cells were spun down at 6,340*g* for 15 min, resuspended and lysed in 6 M guanidine hydrochloride added with 20 mM imidazole. After brief sonication, lysates were centrifuged 30,600*g* for 1 h at 4 °C, and supernatants were collected. Supernatants were run through Ni-column (Qiagen, 30250) and washed sequentially with the following buffers: 2 M guanidine hydrochloride with 20 mM imidazole, 2 M guanidine hydrochloride with 1 M NaCl, and 2 M guanidine hydrochloride with 20 mM imidazole. The His6×-tagged target proteins were eluted in 2 M guanidium hydrochloride with 500 mM imidazole, with 50 µl of elution assessed by SDS–PAGE after ethanol precipitation. Then, protein samples were further purified on size exclusion column 10/300 SD75 (GE healthcare) using the AKTA purifier (GE Healthcare, AKTA pure 25) in SEC buffer (2 M guanidine hydrochloride). Fractions with purified target proteins were combined and concentrated using microcon-10 filter (Millipore, MRCPRT010) to reach the sample concentration ranging from 27 µM to 255 µM and kept at −80 °C for storage.

In vitro phase separation assay

We first carried out the labelling of recombinant protein with the Alexa Flour 488 and 594 protein labelling kit (ThermoFisher, A30006 and A3008) according to manufacturer's protocols. To set up the in vitro phase separation assays, the labelled proteins were mixed with unlabelled ones at a ratio of 1:20, and such a mixture further diluted to a desired concentration in the Eppendorf tubes with either TBS buffer alone (50mM Tris-HCl pH 7.5, 150 mM NaCl) or TBS plus a crowding agent such as 20% of polyethylene glycol (PEG) 3350 (ThermoFisher, NC0620958). Imaging was carried out immediately with samples transferred to a 35-mm dish with 20-mm glass bottom well (Cellvis, D35-20-1.5-N) using Olympus FV3000RS Confocal microscope with 100× oil objective. For fluorescence imaging studies with a mixture of two species of N-IDR recombinant proteins containing FG-repeats in different numbers, we used those with 38 × FG-repeats in the final concentration of 2.5 μM in the TBS buffer (labelled with Alexa Flour 488), which was mixed with those labelled with Alexa Flour 594, either carrying 27 × FG-repeats (a final concentration of 2.5 μM) or 11 × FG repeats (a final concentration of 6 μM).

Colocalization analysis

Colocalization analysis between fusion and H3K27ac or H3K9me3 was performed using the EzColocalization plugin in FIJI version 1.53[34]. Colocalization was measured using the Pearson's correlation coefficient (PCC). An a-priori power analysis of pilot data was performed in G*Power (*z*-tests, two independent Pearson *r* values) and showed that a sample size of at least 388 cells would be required to determine significance at $P > 0.05$ given an effect size of 0.24. For analysis, nuclei were manually segmented by hand tracing with the polygon selection tool, then converted into binary masks used in the EzColocalization plugin to restrict colocalization analysis to the nuclei. PCC values for each cell were averaged and the calculated means were compared with an independent two-tailed Student's *t*-test.

Purification, transduction, and cultivation of primary mouse HSPCs

Primary bone marrow cells were obtained from femur and tibia of 10-week-old female Balb/C mice and then subject to a lineage-negative (Lin^-) enrichment protocol to remove differentiated cell populations as previously described[32,35]. Lin^- enriched HSPCs were first stimulated in the base medium (OptiMEM, Invitrogen, 31985) supplemented with 15% of FBS (Invitrogen, 16000-044), 1% of antibiotics, 50 μM of β-mercaptoethanol and a cytokine cocktail that contains 10 ng ml^{-1} each of mouse SCF (Peprotech), FLT3 ligand (FLT3L; Sigma), IL-3 (Peprotech) and IL-6 (Peprotech) for 4 days as previously described[11,32,35]. Two days after infection with retrovirus, mouse HSPCs were subject to drug selection and then plated for assaying proliferation and differentiation in the same liquid base medium with SCF alone as previously described[11,32,35]. These in vitro cultured HSPC cells were routinely monitored under microscopy and cellular morphology examined by Wright–Giemsa staining as previously described[11,32,35]. For HSPCs transduced with a bicistronic GFP-containing retroviral construct, we also scored relative proliferation of GFP-positive HSPCs by FACS every 2–3 days after infection.

Flow cytometry (FACS) analysis

Cells were washed once in the cold FACS buffer (PBS with 1% of FBS added) and then resuspended and incubated in the FACS buffer added with the respective antibodies (1:100 dilution) for 30 min on ice. The cell pellets were washed with FACS buffer and the stained cells were subject to analysis with the FACS machine (Attune Nxt, Thermo Fisher; available in UNC Flow Cytometry Core Facility). Data were analysed using FlowJo software.

In vivo leukaemogenic assay

All animal experiments were approved by and performed in accord with the guidelines of Institutional Animal Care and Use Committee (IACUC) at the University of North Carolina (UNC) at Chapel Hill. Mice were purchased from the Jackson Laboratory and maintained by the Animal Studies Core, UNC Lineberger Comprehensive Cancer Center. Determination of potential leukaemogenic properties of the oncogene was carried out as previously described[32,36], and no statistical method was used to determine size of cohorts, with investigators blinded to allocation during assays. In brief, 0.5 million of freshly infected and selected murine HSPCs were transplanted to syngeneic 10-week-old female Balb/C mice (JAX lab, 000651) via tail vein injection (carried out by Animal Studies Core of UNC Cancer Center). Mice were regularly monitored with complete blood counting with the collected peripheral blood and abdomen palpation for early signs of leukaemia such as lethargy, increased white blood cell counts and enlarged spleen[30]. Mice exhibiting leukaemic phenotypes were euthanized followed by pathological and histological analyses as described[32,36]. Haematoxylin and eosin (H&E) staining of spleen sections was carried by UNC Pathology Core as previously described[37].

BioID

A BirA cDNA sequence (a gift from B. Strahl) was inserted into N terminus of target protein in the MSCV based retroviral vector, followed by viral production and establishment of 293FT stable expression cells. Proximity-dependent labelling of interacting proteins or BioID was conducted as previously described[38–40]. In brief, 293FT stable cells were collected from five 15-cm plates after treatment with 50 μM of biotin for 24 h, and then washed twice with cold PBS. The cell pellets were resuspended in 1 ml of RIPA lysis buffer (10% glycerol, 25mM Tris-HCl pH 8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% NP-40, 0.2% sodium deoxycholate), and lysates were added with 1 μl of benzonase (Sigma-Aldrich, E1014) followed by incubation on ice for 1 h. After centrifugation at maximum speed for 30 min at 4 °C, the supernatant was collected and incubated with Neutravidin beads (Thermo Fisher, 29204) overnight at 4 °C. The Neutravidin beads were then washed twice with the RIPA buffer and TAP lysis buffer (10% glycerol, 350mM NaCl, 2 mM EDTA, 0.1% NP-40, 50 mM HEPES, pH 8) sequentially. Finally, the beads were washed three times with the ABC buffer (50 mM ammonium bicarbonate, pH 8) and subjected to mass spectrometry-based analysis.

Mass spectrometry-based protein identification

Proteins were eluted from beads by adding 50 μl 2 × Laemmli buffer (Boston Bioproducts) and heating at 95 °C for 5 min. A total of 5 0μl of each sample was resolved by SDS–PAGE using a 4–20% Tris-glycine wedge well gel (Invitrogen) and visualized by Coomassie staining. Each SDS–PAGE gel lane was sectioned into 12 segments of equal volume. Each segment was subjected to in-gel trypsin digestion as follows. Gel slices were destained in 50% methanol (Fisher), 50 mM ammonium bicarbonate (Sigma-Aldrich), followed by reduction in 10 mM Tris [2-carboxyethyl] phosphine (Pierce) and alkylation in 50 mM iodoacetamide (Sigma-Aldrich). Gel slices were then dehydrated in acetonitrile (Fisher), followed by addition of 100 ng porcine sequencing grade modified trypsin (Promega) in 50 mM ammonium bicarbonate (Sigma-Aldrich) and incubation at 37 °C for 12–16 h. Peptide products were then acidified in 0.1% formic acid (Pierce). Tryptic peptides were separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 150 × 0.075 mm column using a nanoAcquity UPLC system (Waters). Peptides were eluted using a 30 min gradient from 97:3 to 67:33 buffer A:B ratio (buffer A: 0.1% formic acid, 0.5% acetonitrile; buffer B: 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.15 kV) followed by MS/MS analysis using higher-energy collisional dissociation (HCD) on an Orbitrap Fusion Tribrid mass spectrometer (Thermo) in top-speed data-dependent

mode. MS data were acquired using the FTMS analyser in profile mode at a resolution of 240,000 over a range of 375 to 1,500 *m/z*. Following HCD activation, MS/MS data were acquired using the ion trap analyser in centroid mode and normal mass range with precursor mass-dependent normalized collision energy between 28.0 and 31.0. Proteins were identified by searching the UniProtKB database restricted to *Homo Sapiens* using Mascot (Matrix Science) with a parent ion tolerance of 3 ppm and a fragment ion tolerance of 0.5 Da, fixed modifications for carbamidomethyl of cysteine, and variable modifications for oxidation on methionine and acetyl on N terminus. Scaffold (Proteome Software) was used to verify MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established with less than 1.0% false discovery by the Scaffold Local false discovery rate algorithm. Protein identifications were accepted if they could be established with less than 1.0% false discovery and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm[41]. Proteins were filtered out if they had a spectral count <8 in all sample groups and the counts were normalized to $\log_2$-normalized spectral abundance factor (NSAF) values. Significant interacting proteins were defined with a cut-off of a $\log_2$-transformed fold change above 2 in the experimental versus control samples.

ChIP–seq

ChIP–seq was carried out as previously described[30,42]. In brief, cells were fixed in 1% formaldehyde (Thermo Scientific, 28908) for 10 min, followed by quenching with 125 mM glycine for 5 min. Cells were then washed twice with cold PBS added with protease inhibitors (Sigma-Aldrich, 4693132001), and then subjected to resuspension and incubation in LB1 buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100), LB2 buffer (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), and LB3 buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine). The cell nuclei were collected for sonication using Bioruptor sonicator (Diagenode, B01020001; at high-energy setting for 45 cycles with 30 s on and 30 s off). After treatment with Triton X-100 (1% as a final concentration), the supernatant was collected after centrifugation (20,000*g* for 10 min at 4 °C) for incubation with the dynabeads (Invitrogen, 11204D) that are pre-bound with antibodies for around 8 h at 4 °C. After a series of wash, the chromatin–protein complexes bound to beads were eluted, subject to reverse crosslink overnight at 65 °C, and treated with RNase (Roche, 11119915001; 1 h at 37 °C) and then protease K (Roche, 03115828001; 2 h at 55 °C). The final DNA sample, as well as 1% of input chromatin, was recovered using PCR purification kit (Qiagen, 28106). The ChIP–seq library was prepared using NEBNext Ultra II kit (NEB, E7645L) following the manufacturer's instructions. ChIP–seq libraries were sequenced on the Nextseq 550 system using Nextseq 550 High Output Kit v2.5 (Illumina, 20024906). For ChIP–seq of HA-tagged N-IDR/A9 (with either wild-type or mutated IDRs), we used the matched input signals for signal normalization; for ChIP–seq of GFP-tagged N-IDR/A9 (with either wild-type or mutated IDRs), we used signals of *Drosophila* spike-in chromatin for normalization as previously described[43] (Active Motif spike-in ChIP–seq reagents, 53083 and 61686).

ChIP–seq data analysis

ChIP–seq data alignment, filtration, peak calling and assignment, and cross-sample comparison were performed as previously described[30,42] with slight modifications. In brief, ChIP–seq reads were aligned to human genome build GRCh37/hg19 or to mouse genome build GRCm38/mm10 using STAR version 2.7.1a[44]. The MACS2 software was used for peak identification with data from input as controls and default parameters[45]. Homer (ver 4.10.0) 'annotatePeaks' and 'findMotifsGenome' functions were used to annotate the called peaks and to find enriched motifs in these called peaks. Alignment files in the bam format were also transformed into read coverage files (bigWig format) using DeepTools[46]. Genomic binding profiles were generated using the deepTools 'bamCompare' functions with options [–operation ratio–pseudocount 1 -binSize 10–extendReads 250] and normalized to the matched input. The resulting bigWig files were visualized in the Integrative Genome Viewer (IGV). Heat maps for ChIP–seq signals were generated using the deepTools 'computeMatrix' and 'plotHeatmap' functions. ROSE were used for defining super-enhancers[47], with input signals used as control for normalization and peaks at ±2.5 kb from the transcriptional start site excluded. Homer mergePeaks was used to determine overlap of ChIP–seq peaks with default settings.

RNA-seq and data analysis

RNA-seq was performed as previously described[42,48]. For 293FT cells, the same stable expression lines used for ChIP–seq were used. For mouse HPSCs, cells were collected for RNA isolation 7 days after viral transduction and drug selection in the OptiMEM medium supplemented with the HSPC-supporting cytokines. In brief, total RNAs were purified using RNeasy Plus kit (Qiagen, 74136) and further processed with Turbo DNA-free kit (Thermo Fisher, AM1907) to ensure the purity of RNA sample. For RNA-seq, the RNA samples were either sent to Novogene or processed using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) and NEBNext Ultra II RNA library Prep kit (NEB, E7770) as per the manufacturer's instructions. The multiplexed RNA-seq libraries were subjected for deep sequencing using the Illumina NextSeq500 platform (available in the UNC Sequencing Facility) with the Nextseq 550 High Output Kit v2.5 (Illumina, 20024906). For data analysis, RNA-seq reads were mapped to the reference genome followed by differential gene expression analysis as previously described[42,48]. In brief, RNA-seq reads were mapped using MapSplice[49] and quantified using RSEM[50]. Read counts were upper-quantile normalized and $\log_2$-transformed. Raw read counts were used for differential gene expression analysis by DESeq[51]. Gene Ontology analysis was done using the C5 gene set of Molecular Signature Database (MsigDB) collections available in GSEA website[52].

ChIP–qPCR or RT–qPCR

ChIP–qPCR or RT–qPCR was performed as previously described[37,42]. ChIP DNA was prepared as described above for ChIP–seq, whereas total RNA was used to generate cDNA with the iScript cDNA Synthesis kit (Biorad, 1708890) for qPCR.

Single molecule tracking, lattice light sheet microscopy, and data analysis

Three-dimensional lattice light sheet microscopy movies of cells were acquired on a modified version of the lattice light sheet system as previously described[33] using a square lattice excitation with numerical apertures of 0.4 (outer) and 0.3 (inner). Time intervals and imaging duration are specified in the legends for each dataset presented. Single-molecule tracking was performed on the same system by focusing on a single plane within the nucleus of cells expressing Halo-tag protein fusions. Before imaging, cells were incubated win 1 nM of Halo Tag-Janelia Fluor 549 ligand for 20 min and then washed in PBS[53]. After transferring to the microscope, single planes within the nucleus of each cells were imaged under the same lattice illumination parameters above for a total of 20,000–40,000 frames with 20 ms exposures. Before tracking, images were pre-processed with a rolling ball background subtraction and histogram equalization contrast enhancement using ImageJ. Single molecules were then tracked using the TrackMate plugin for ImageJ[54]. To account for variation in protein expression levels between cells and avoid potential tracking artefacts due to different densities of fluorescent molecules, Pandas software library for python[55] was used to register single particle tracking datasets such that the number particles within a rolling 100 window was consistent both within and between conditions. We controlled for photobleaching and phototoxicity by confirming that mean molecular speeds within a single cell did not

vary substantially throughout the course of the imaging experiment. Finally, molecular trajectories were fit to a two-state kinetic model using Spot-On[22] to estimate the mean diffusion coefficients and fraction of molecular populations for both the slow-diffusing/bound state and rapidly diffusing/free state.

In situ Hi-C

In situ Hi-C was performed as previously described[56]. Five million cells were crosslinked in 1% formaldehyde for 10 min with stirring and quenched by adding 2.5 M glycine to a final concentration of 0.2 M for 5 min with rocking. Cells were pelleted by spinning at 300*g* for 5 min at 4 °C. The pellet was washed with cold PBS and spun again before freezing in liquid nitrogen. Cells were lysed with 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630 and protease inhibitors (Sigma, P8340) for 15 min on ice. Cells were pelleted and washed once more using the same buffer. Pellets were resuspended in 50 µl of 0.5% SDS and incubated for 7 min at 62 °C. Next, reactions were quenched with 145 µl of water and 25µl of 10% Triton X-100 (Sigma, 93443) at 37 °C for 15 min. Chromatin was digested overnight with 25 µl of 10X NEBuffer2 and 100 U of MboI at 37 °C with rotation. Reactions were incubated at 62 °C for 20 min to inactivate MboI and then cooled to room temperature. Fragment overhangs were repaired by adding 37.5 µl of 0.4 mM biotin-14-dATP, 1.5 µl of 10mM dCTP, 1.5 µl of 10mM dGTP, 1.5 µl of 10mM dTTP, and 8 µl of 5 U µl^{-1} DNA polymerase I, large (Klenow) fragment and incubating at 37 °C for 1.5 h. Ligation was performed by adding 667 µl of water, 120 µl of 10X NEB T4 DNA ligase buffer, 100 µl of 10% Triton X-100, 12 µl of 10 mg ml^{-1} BSA, and 1 µl of 2,000 U µl^{-1} T4 DNA ligase and incubating at room temperature for 4 h with slow rotation. Samples were pelleted at 2,500*g* and resuspended in 432 µl of water, 18 µl of 20 mg ml^{-1} proteinase K, 50 µl of 10% SDS, 46 µl of 5M NaCl and incubated for 30 min at 55 °C. The temperature was raised to 68 °C and incubated overnight. Samples were cooled to room temperature. Then, 874 µl of pure ethanol and 55 µl of 3 M sodium acetate pH 5.2 were added to each tube which were subsequently incubated for 15 min at −80 °C. Tubes were spun at maximum speed at 2 °C for 15 min and washed twice with 70% ethanol. The resulting pellet was resuspended in 130 µl of 10 mM Tris-HCl, pH 8, and incubated at 37 °C for 15 min. DNA was sheared using an LE220 Covaris Focused-ultrasonicator to a fragment size of 300–500 bp. Sheared DNA was size selected using AMPure XP beads. One hundred and ten µl of beads were added to each reaction and incubated for 5 min. Using a magnetic stand, supernatant was removed and added to a fresh tube. Then, 30 µl of fresh AMPure XP beads were added and incubated for 5 min. Beads were separated on a magnet and washed twice with 700 µl of 70% ethanol without mixing. Beads were left to dry and then sample was eluted using 300 µl of 10 mM Tris-HCl, pH 8. 150 of 10 mg ml^{-1} Dynabeads MyOne Streptavidin T1 beads were washed resuspended in 300 µl of 10 mM Tris HCl, pH 7.5. This solution was added to the samples and incubated for 15 min at room temperature. Beads were washed twice with 600 µl Tween Washing Buffer (TWB; 250 µl Tris-HCl, pH 7.5, 50 µl 0.5 M EDTA, 10 ml 5 M NaCl, 25 µl Tween-20, and 39.675 ml water) at 55 °C for 2 min with shaking. Sheared ends were repaired by adding 88 µl 1× NEB T4 DNA ligase buffer with 1 mM ATP, 2 µl of 25 mM dNTP mix, 5 µl of 10 U µl^{-1} NEB T4 PNK, 4 µl of 3 U µl^{-1} NEB T4 DNA polymerase I, 1 µl of 5 U µl^{-1} NEB DNA polymerase I, large (Klenow) fragment and incubating at room temperature for 30 min. Beads were washed two more times with TWB for 2 min at 55 °C with shaking. Beads were washed once with 100 µl of 1× NEBuffer 2 and resuspended in 90 µl of 1× NEBuffer 2, 5 µl of 10 mM dATP, 5 µl of 5 U µl^{-1} NEB Klenow exo minus, and incubated at 37 °C for 30 min. Beads were washed two more times with TWB for 2 min at 55 °C with shaking. Beads were washed once in 50 µl of 1× Quick Ligation reaction buffer and resuspended in 50 µl of 1× Quick Ligation reaction buffer. Then, 2 µl of NEB DNA Quick ligase and 3 µl of an Illumina-indexed adaptor were added and the solution was incubated for 15 min at room temperature. Beads were reclaimed using the magnet and washed two more times with TWB for 2 min at 55 °C with shaking. Beads were washed once in 100 µl of 10 mM Tris-HCl, pH 8, and resuspended in 50 µl of 10 mM Tris-HCl, pH 8. Hi-C libraries were amplified for 7–12 cycles in 5 µl of PCR primer cocktail, 20 µl of Enhanced PCR mix, and 25 µl of DNA on beads. The PCR settings included 3 min of 95 °C followed by 7–12 cycles of 20 s at 98 °C, 15 s at 60 °C, and 30 s at 72 °C. Samples were then held at 72 °C for 5 min before lowering to 4 °C until samples were collected. Amplified samples were brought to 250 µl with 10 mM Tris-HCl, pH 8. Samples were separated on a magnet and supernatant was transferred to a new tube. One hundred and seventy-five µl of AMPure XP beads were added to each sample and incubated for 5 min. Beads were separated on a magnet and washed once with 700 µl of 70% ethanol. Supernatant was discarded. One hundred µl of 10 mM Tris-HCl and 70 µl of fresh AMPure XP beads were added and the solution was incubated for 5 min at room temperature. Beads were separated with a magnet and washed twice with 700 µl 70% ethanol. Beads were left to dry and DNA was eluted in 25 µl of Tris HCl, pH 8.0. The resulting libraries were next quantified by Qubit and Bioanalyzer. A low depth sequencing was performed first using the MiniSeq sequencer system (Illumina) and analysed using the Juicer pipeline[57] to assess quality control before deep sequencing (NovaSeq S4). Each Hi-C library was assessed in biological and technical duplicate achieving a total of 3 billion reads per cell line.

Hi-C data processing and analysis

In situ Hi-C datasets were processed using the Juicer Hi-C pipeline with default parameters as previously described[57]. MboI was used as the restriction enzyme, and reads were aligned to the hg19 human reference genome with bwa (version 0.7.17). Data were processed for 3,058,370,530 Hi-C read pairs in N-IDR$_{WT}$/A9 cells, yielding 1,791,818,927 Hi-C contacts (58.59%) and 2,914,343,903 Hi-C read pairs in N-IDR$_{FS}$/A9 cells, yielding 1,708,441,327 Hi-C contacts (58.62%). Hi-C matrices were constructed for each individual replicate for downstream analysis. A Hi-C mega map was constructed by combining all replicates for each condition (that is, N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9). For visualization, the resulting Hi-C contact matrices were normalized with a matrix balancing algorithm as previously described[58] ('KR') to adjust for regional background differences in chromatin accessibility.

Loops were detected using HiCCUPS from the Juicer tools software (version 1.11.09) as previously described[56] via the following command: 'hiccups -m 2048 -c 2 -r 5000,10000,25000 -k KR -f 0.1,0.1,0.1 -p 4,2,1 -i 8,6,4 -t 0.2,1.5,1.5,1.75 -d 30000,30000,60000'. A total of 4,788 loops were identified in N-IDR$_{WT}$/A9 and 2,826 loops were identified in N-IDR$_{FS}$/A9 for a total of 7,616 loops at 10-kb resolution. After filtering out redundant loops, 6,615 combined loops remained. Unnormalized loop counts were extracted using the straw api[57] for all loops in each replicate (8 total). Differential loops between N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 were determined using DESeq2[59], including biological replicate and condition as covariates in the model. 232 N-IDR$_{WT}$/A9-specific loops and 52 N-IDR$_{FS}$/A9-specific loops were considered significantly differential at a Benjamini–Hochberg adjusted *P* value ≤ 0.01.

APA of N-IDR/A9 binding site interactions was conducted in R using straw. All unique, paired interactions between the 500 strongest N-IDR$_{WT}$/A9 ChIP–seq binding sites were categorized into (1) inter-chromosomal (n = 95,959), (2) long (≥ 2 Mb) intra-chromosomal (n = 6,298), or (3) short (< 2 Mb) intra-chromosomal (n = 574) interactions. Short interactions were filtered out such that the corner of the APA plot would not intersect the diagonal, reducing them from n = 574 to n = 309. Unnormalized pixel values ±10 surrounding pixels were extracted from N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 Hi-C files at 10-kb resolution for each interaction pair. Resulting 21×21, 10-kb pixel matrices were aggregated and normalized to the number of binding site pairs.

APA of differential loop calls was conducted in R using straw. APA was run for all loops (n = 6,615), N-IDR$_{WT}$/A9-specific loops (n = 232), and N-IDR$_{FS}$/A9-specific loops (n = 52) using both N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 Hi-C. Short interactions were filtered out as described above, reducing the

number of interactions to $n = 3{,}427$, $n = 121$, and $n = 24$ for all, N-IDR$_{WT}$/A9-specific and N-IDR$_{FS}$/A9-specific loops, respectively. Unnormalized pixels were extracted with straw producing a 21×21 pixel matrix at 10-kb resolution that was aggregated and normalized by the number of loops per group.

All loops were partitioned as either N-IDR$_{WT}$/A9-specific loops (WT loops) or N-IDR$_{FS}$/A9-specific loops (FS loops) based on differential loop calling (as described above) and then split into separate loop anchors. Loop anchors were then intersected (bedtoolsr) with several features including ChIP–seq peaks for NUP98–HOXA9, CTCF, or H3K27Ac in both cell types (N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9) and with promoter regions (defined as 1,000 bp upstream of transcription start sites). Permutation testing was used to calculate *P* values for each feature's intersection with loop anchors. In short, the observed percentage of each feature present at wild-type or FS loop anchors was calculated. The expected percentage was determined by randomly sampling an equivalent number of loop anchors from all loop anchors called, then calculating the percentage overlap with each feature. This procedure was repeated 1,000 times to create a distribution of expected values. *P* values were determined by summing the number of expected values greater than (or less than if the observed value was less than the mean) the observed value for that feature.

All loops were partitioned as either N-IDR$_{WT}$/A9-specific loops (WT loops) or N-IDR$_{FS}$/A9-specific loops (FS loops) based on differential loop calling (as described above). Each loop was then intersected with 5-kb windows around the transcription start sites of genes using the bedtoolsr 'pairtobed' function with either end of the loop constituting an overlap. The $\log_2$-transformed fold change in expression value (WT/FS) of genes overlapping either end of a wild-type or FS differential loop were plotted along with the expression of all genes. A Dunn's multiple comparison test following a Kruskal–Wallis test showed a statistically significant difference in expression between wild-type-specific gene-loops and either FS-specific gene-loops ($P = 0.015$) or all genes ($P < 0.001$), after *P* value correction with the Benjamini–Hochberg procedure. In this study, wild-type-specific loops were present in the N-IDR$_{WT}$/A9-expressing cells and absent in N-IDR$_{FS}$/A9 cells whereas mutant-specific loops were absent in N-IDR$_{WT}$/A9 cells and present in N-IDR$_{FS}$/A9 cells, supporting accurate calling of differential loops.

3C–qPCR

Cell samples were processed and analysed as previously described with slight modifications[37]. In brief, 10 million of cells were fixed in 1% formaldehyde at room temperature for 10 min, followed by quenching in 0.125 M glycine for 5 min. Fixed cells were washed in cold PBS and lysed in ice-cold lysis buffer (10 mM Tris-Cl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitor cocktail) for 1 h at 4 °C. Nuclei were collected by centrifugation at 2,400*g* for 5 min and digested with 800 U of Bgl-II enzyme, added with 0.3% of SDS and 1.8% of Triton X-100 in the molecular-grade water with respective enzyme digestion buffer (1.2×) for overnight at 37 °C. After inactivation at 65 °C for 20 min with 1.6% of SDS, digested chromatin was subjected to ligation by T4 ligase (NEB) with 1% Triton X-100 for overnight at 16 °C, followed by 30 min incubation at room temperature. Ligated chromatin was treated with protease K for overnight at 65 °C and then treated with RNase for 2 h at 37 °C, followed by DNA purification with the phenol–chloroform extraction protocol. For qPCR, the obtained DNA was diluted 50-fold and used as a template. Primers were designed for the respective genomic loci with chromatin loop as detected by Hi-C mapping experiment. All PCR products were sequenced to confirm that they are indeed correctly ligated products from two distant genomic loci where chromatin loop is expected to form between them. All the primers used for 3C–qPCR are listed in Supplementary Table 8.

Statistics and reproducibility

Experimental data are presented as the mean ± s.d. of three independent experiments unless otherwise noted. Statistical analysis was carried out with two-sided Student's *t*-test for comparing the two sets of data with assumed normal distribution. We used a log-rank test for the Kaplan–Meier survival curve to define statistical significance. A *P* value of less than 0.05 was considered to be significant. Statistical significance levels are denoted as follows: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$.

Sample numbers are indicated in the figure legends. Results of images or staining (shown in Figs. 1c–h, k and Extended Data Figs. 1d, f, h, 2b, d, e, h, 7a, 8a) and western blotting (Fig. 1b and Extended Data Figs. 1b, d, f, g, 2a) were reproducible with at least three independent experiments or prepared samples, with the representative ones shown in the figures.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Next-generation sequencing datasets including those of ChIP–seq, RNA-seq and Hi-C used in this current study are deposited in the NCBI GEO under the accession number GSE144643. The mass spectrometry-based proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023548 and 10.6019/PXD023548. Source data are provided with this paper.

Code availability

The scripts for genomic data analyses and all other data are available from the corresponding author upon request.

27. Calvo, K. R., Sykes, D. B., Pasillas, M. P. & Kamps, M. P. Nup98-HoxA9 immortalizes myeloid progenitors, enforces expression of Hoxa9, Hoxa7 and Meis1, and alters cytokine-specific responses in a manner similar to that induced by retroviral co-expression of Hoxa9 and Meis1. *Oncogene* **21**, 4247–4256 (2002).
28. Fahrenkrog, B. et al. Expression of leukemia-associated Nup98 fusion proteins generates an aberrant nuclear envelope phenotype. *PLoS One* **11**, e0152321 (2016).
29. Yu, M. et al. A resource for cell line authentication, annotation and quality control. *Nature* **520**, 307–311 (2015).
30. Xu, B. et al. Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. *Blood* **125**, 346–357 (2015).
31. Cai, L. et al. An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell* **49**, 571–582 (2013).
32. Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* **9**, 804–812 (2007).
33. Chen, B.-C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
34. Stauffer, W., Sheng, H. & Lim, H. N. EzColocalization: An ImageJ plugin for visualizing and measuring colocalization in cells and organisms. *Sci. Rep.* **8**, 15764 (2018).
35. Wang, G. G. et al. Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. *Nat. Methods* **3**, 287–293 (2006).
36. 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).
37. Lu, R. et al. Epigenetic perturbations by Arg882-mutated dnmt3a potentiate aberrant stem cell gene-expression program and acute leukemia development. *Cancer Cell* **30**, 92–107 (2016).
38. Roux, K. J., Kim, D. I. & Burke, B. BioID: a screen for protein–protein interactions. *Curr. Protoc. Protein Sci.* **74**, 19.23.11–19.23.14 (2013).
39. Roux, K. J., Kim, D. I., Burke, B. & May, D. G. BioID: a screen for protein–protein interactions. *Curr. Protoc Protein Sci.* **91**, 19.23.11–19.23.15 (2018).
40. Li, J. et al. ZMYND11-MBTD1 induces leukemogenesis through hijacking NuA4/TIP60 acetyltransferase complex and a PWWP-mediated chromatin association mechanism. *Nat. Commun.* **12**, 1045 (2021).
41. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
42. Cai, L. et al. ZFX Mediates non-canonical oncogenic functions of the androgen receptor splice variant 7 in castrate-resistant prostate cancer. *Mol. Cell* **72**, 341–354 (2018).
43. Egan, B. et al. An alternative approach to ChIP-Seq normalization enables detection of genome-wide changes in histone H3 lysine 27 trimethylation upon ezh2 inhibition. *PLoS One* **11**, e0166438 (2016).
44. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
45. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
46. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016).
47. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).

48. Ren, Z. et al. PHF19 promotes multiple myeloma tumorigenicity through PRC2 activation and broad H3K27me3 domain formation. *Blood* **134**, 1176–1189 (2019).
49. Wang, K. et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*. **38**, e178 (2010).
50. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
51. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol*. **11**, R106 (2010).
52. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
53. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250, 3, 250 (2015).
54. Tinevez, J.-Y. et al. TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
55. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
56. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
57. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. **3**, 95–98 (2016).
58. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal*. **33**, 1029–1047 (2012).
59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
60. Ren, Y., Seo, H.-S., Blobel, G. & Hoelz, A. Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. *Proc. Natl Acad. Sci. USA* **107**, 10406–10411 (2010).
61. Yung, E. et al. Delineating domains and functions of NUP98 contributing to the leukemogenic activity of NUP98-HOX fusions. *Leuk. Res*. **35**, 545–550 (2011).

Acknowledgements We thank M. Kamps, B. Strahl, B. Fahrenkrog, J. Schwaller, J. van Deursen, J. Song and J. Hao for providing reagents used in the study and the Wang laboratory members and J. Bear for discussion and technical support. We thank J. Lippincott-Schwartz for help with lattice light sheet microscopy and J. Rowley and A. Gladfelter for discussion and input. We thank UNC for facilities, including Imaging Core, High-Throughput Sequencing Facility (HTSF), Bioinformatics Core, Flow Cytometry Core, Tissue Culture Facility and Animal Studies Core, for their professional assistance of this work. We thank S. Pattenden for use of the Covaris LE220 instrument which was provided by the North Carolina Biotechnology Center Institute Development Program grant 2017-IDG-1005. The cores affiliated to UNC Cancer Center are supported in part by the UNC Lineberger Comprehensive Cancer Center Core Support Grant P30-CA016086 and UNC Neuroscience Microscopy Core supported, in part, by funding from the NIH-NINDS Neuroscience Center Support Grant P30 NS045892 and the NIH-NICHD Intellectual and Developmental Disabilities Research Center Support Grant U54 HD079124. This work was supported by NIH grants (R01-CA215284 and R01-CA218600 to G.G.W.; R35-GM128645 to D.H.P.; DP2GM136653 to W.R.L.; P20GM121293, R24GM137786, R01CA236209, S10OD018445, and TL1TR003109 to A.J.T; R01HL148128 and R01HL153920 to D.Z.), a Kimmel Scholar Award (to G.G.W.), Gabrielle's Angel Foundation for Cancer Research (to G.G.W.), Gilead Sciences Research Scholars Program in haematology/oncology (to G.G.W.), When Everyone Survives (WES) Leukemia Research Foundation (to G.G.W.) and UNC Lineberger Stimulus Awards (to D.H.P. and to L.C.). E.S.D. was supported by the NIH-NIGMS training grant T32-GM067553. W.R.L. is a Searle Scholar, a Beckman Foundation Young Investigator, and a Packard Fellow for Science and Engineering. G.G.W. is an American Cancer Society (ACS) Research Scholar, an American Society of Hematology (ASH) Scholar in basic science, and a Leukemia and Lymphoma Society (LLS) Scholar.

Author contributions J.H.A. designed the research, performed experiments, interpreted data and wrote the manuscript. J.H.A., Y-H. T., H.U., J. L., L. C., D.Z. and G.G.W. performed genomic data analysis. J.H.A. and S.Z. performed in vitro phase separation assays. D.P.K. conducted imaging quantification analysis. A.J.S., S.G.M., R.D.E. and S.D.B. performed proteomic analysis under the supervision of A.J.T. T.A.D. and J.H.A. performed single molecule tracking studies under the supervision of W.R.L. J.H.A. and J.L. performed murine leukaemia assays. E.S.D., I.Y.Q. and J.H.A. performed Hi-C mapping, data analysis and interpretation under the supervision of D.H.P. G.G.W. conceived the idea, supervised and designed the research, interpreted data, and wrote the manuscript with the inputs from all authors.

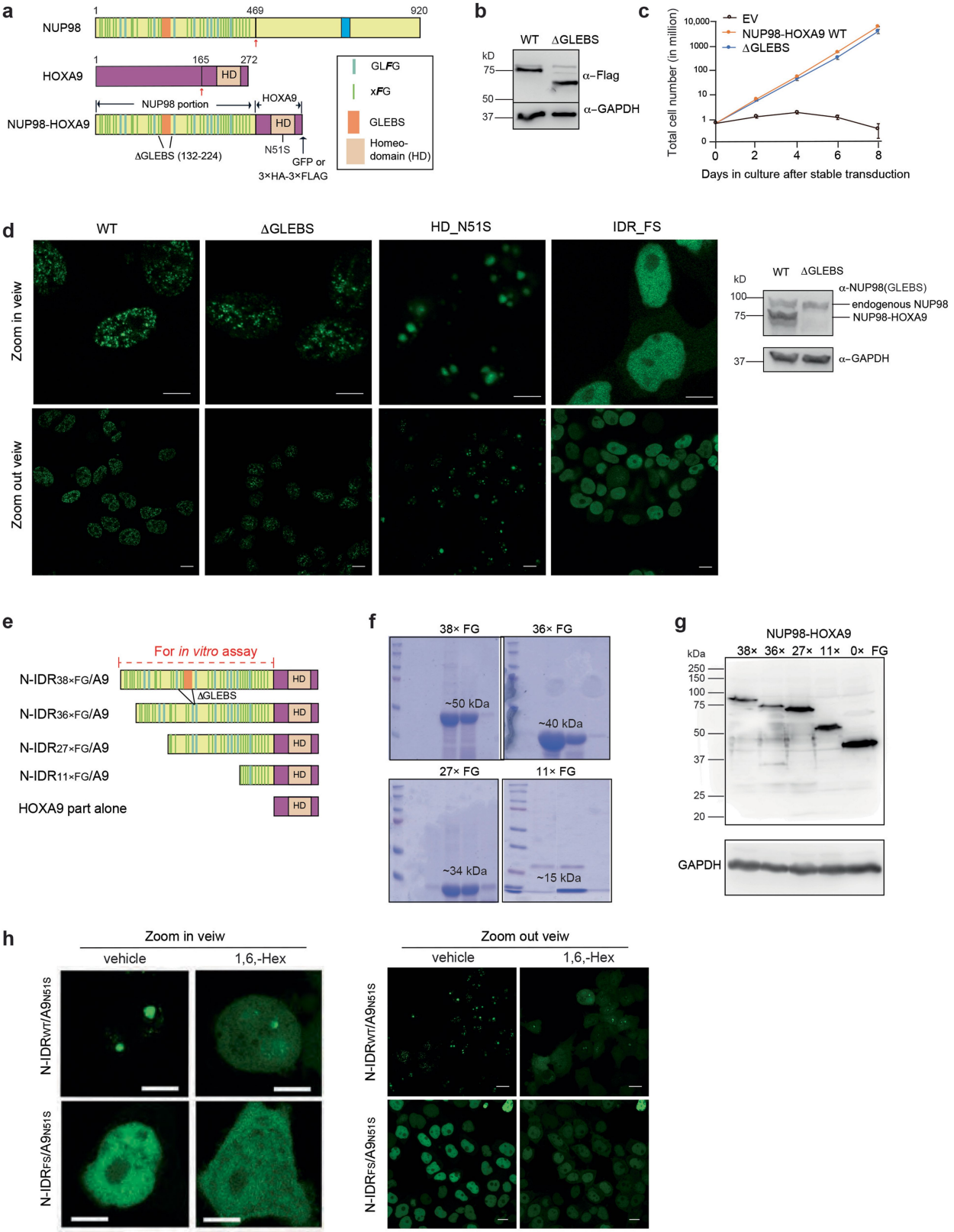
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03662-5.

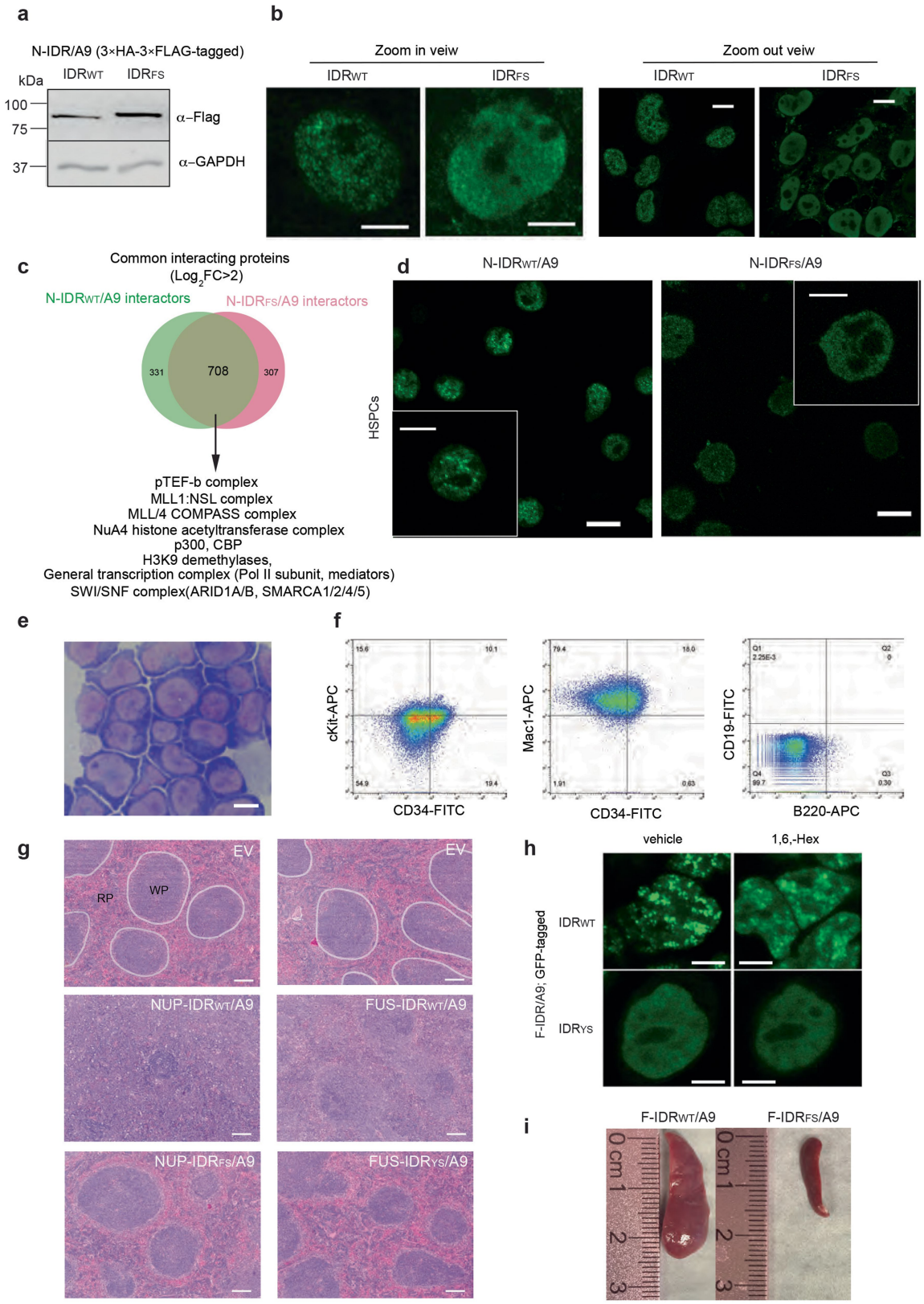
Correspondence and requests for materials should be addressed to D.H.P. or G.G.W.

Peer review information *Nature* thanks Ari Melnick, Tanja Mittag and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | IDR retained within the leukaemia-related NUP98–HOXA9 chimera forms phase-separated condensates in vitro and is essential for establishing phase-separated chimeric transcription factor assemblies in the nucleus. a, Schematic showing the domain architecture of normal NUP98 (top), normal HOXA9 (middle) and leukaemic NUP98–HOXA9 chimera (bottom; with either GFP or 3×HA–3×Flag tag fused to C terminus). The GLFG or non-GLFG (xFG) motif contents, which make up IDR, and other important domains are shown. GLEBS represents the GLE2-binding sequence, which directs the NUP98 interaction with GLE2 (also known as RAE1) for mRNA export when NUP98 acts as component of nuclear pore complex[60]. Red arrows indicate the common breakage point of NUP98 and HOXA9. **b**, Immunoblotting of NUP98–HOXA9, either full-length (WT) or with GLEBS deleted (Δ132-224; see **a**), as detected by the indicated antibodies after stable transduction into primary mouse HSPCs. For gel source data, see Supplementary Fig. 1. **c**, Mouse HSPCs stably transduced with wild-type or GLEBS-deleted NUP98–HOXA9 showed similar proliferation in liquid cultures ($n = 3$ independent cell cultures per group), in agreement to previous reports[17,61]. Empty vector (EV)-infected HSPCs served as a control. Data are mean ± s.d. **d**, Live-cell fluorescence imaging (GFP; with zoomed-in and zoomed-out views shown in the top and bottom panels, respectively) of 293FT cells with stable transduction of GFP-tagged NUP98–HOXA9, wild-type, GLEBS-deleted (also referred to as N-IDR$_{WT}$/A9; see Fig. 1a) or carrying a DNA-binding-defective mutation in homeodomain (HD$_{N51S}$) or a Phe-to-Ser mutation that substitutes Phe residues within all FG repeats to Ser (IDR$_{FS}$, also referred to as N-IDR$_{FS}$/A9; see Fig. 1a). The right panel shows immunoblotting of endogenous normal NUP98 in 293FT cells, as well as the stably transduced exogenous NUP98–HOXA9, either wild-type (lane 1) or GLEBS-deleted (lane 2), as detected by antibodies against GLEBS of NUP98[17]. For gel source data, see Supplementary Fig. 1. Scale bars, 10 μm. **e**, Schematic of the indicated N-IDR fusion domains with a varying number of FG repeats. The IDR portion used for in vitro assay in main Fig. 1d is indicated by a red dotted line. **f**, SDS–PAGE images showing recombinant N-IDR domain protein with the indicated varying number of FG repeats (His6×-tagged; see **e**), purified with Ni-column and an additional size exclusion column purification step. The protein size is labelled above the recombinant protein. **g**, Anti-GFP immunoblotting for GFP-tagged NUP98–HOXA9 chimera with the indicated varying number of FG repeats described in **e** after stable transduction in 293FT cells. For gel source data, see Supplementary Fig. 1. **h**, Live-cell fluorescence imaging for the N51S-mutated N-IDR/A9 (GFP-tagged) with either wild-type (top) or the Phe-to-Ser mutated IDR (bottom) in 293FT stable expression lines before (left) and after (right) treatment with 10% 1,6-hexanediol for 1 min. The left panels show zoomed-in images of a representative cell from the right panels of zoomed-out cell images. Scale bar, 10 μm.

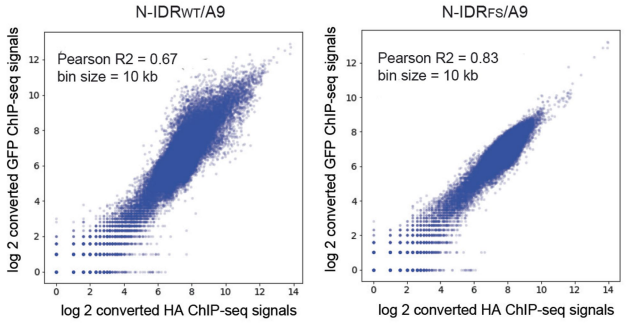


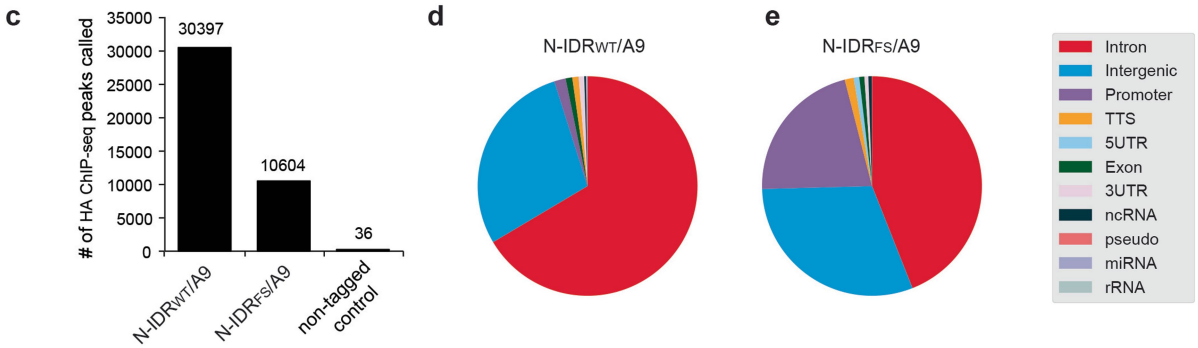
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | IDR contained within chimeric transcription factor is required for leukaemic transformation of primary mouse HSPCs. **a**, **b**, Immunoblotting (**a**) and fixed cell immunostaining (**b**; anti-Flag) of the LLPS-competent N-IDR$_{WT}$/A9 and LLPS-incompetent N-IDR$_{FS}$/A9 after stable transduction in 293FT cells. The left panel of **b** shows a zoomed-in view on the right panel. Scale bars, 10 μm. For gel source data, see Supplementary Fig. 1. **c**, Venn diagram shows significant overlap between the N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 interactomes as detected by BioID, with the cut-off value set as the $\log_2$-transformed fold change value above 2 compared with control. Examples of the detected interacting proteins are shown below. **d**–**f**, Immunostaining (**d**; anti-GFP), Wright–Giemsa staining (**e**) and FACS with the indicated surface markers (**f**) using mouse HSPCs 1 month after transduction of N-IDR$_{WT}$/A9 (GFP or 3×HA-3×Flag-tagged), which revealed a typical acute myeloid leukaemia phenotype (cKit$^+$, CD34+, Mac1^{high}, CD19$^-$, B220$^-$). The insert in **d** shows a zoomed-in view of the representative cell. Scale bars, 5 μm. For FACS gating strategy, see Supplementary Fig. 1. **g**, H&E-stained spleen section images for the indicated cohort at 10 × magnification. White pulp (WP) is outlined with white line for the sample from mice transplanted with empty vector-infected HSPCs (top). Note that clear demarcation between white pulp and red pulp (RP), as observed in cohorts receiving either empty vector or the mutant forms of fusion (bottom), is lost in those with N-IDR$_{WT}$/A9 and F-IDR$_{WT}$/A9 (middle) due to an excessive expansion of transformed leukaemia cells that infiltrated into spleen, leading to splenomegaly observed in **i** and Fig. 1k. **h**, Live-cell fluorescence (GFP) imaging of 293FT cells with stable expression of an artificial HOXA9 chimera created by replacing the NUP98 FG repeats with an unrelated IDR of the RNA-binding protein FUS, either wild-type or Tyr-to-Ser mutated (hereafter referred to as the F-IDR$_{WT}$/A9 and F-IDR$_{YS}$/A9 fusion, respectively; see Fig. 1a), before and after treatment with 10% 1,6-hexanediol for 1 min. Scale bar, 10 μm. **i**, Representative image of spleen from mice 7 months after transplantation of mouse HPSCs stably transduced with either F-IDR$_{WT}$/A9 (left) or F-IDR$_{YS}$/A9 (right).

a

Cell sample	Treatment	Epitope for ChIP	Aligned read # (duplicate removed)
N-IDR(WT)/A9, HA-tagged	none	input	61,503,386
		HA ChIP	50,816,405
N-IDR(FS)/A9, HA-tagged	none	input	73,139,298
		HA ChIP	66,113,889
N-IDR(WT)/A9, GFP-tagged	mock	GFP ChIP	27,942,764
		H2Av ChIP (fly spikein)	38,240
N-IDR(FS)/A9, GFP-tagged	mock	GFP ChIP	38,070,512
		H2Av ChIP (fly spikein)	47,506
N-IDR(WT)/A9, GFP-tagged	1,6-hexanediol	GFP ChIP	33,251,112
		H2Av ChIP (fly spikein)	45,119
N-IDR(FS)/A9, GFP-tagged	1,6-hexanediol	GFP ChIP	34,402,931
		H2Av ChIP (fly spikein)	40,713



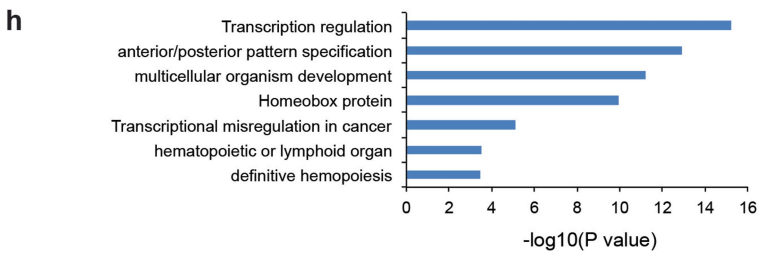


f

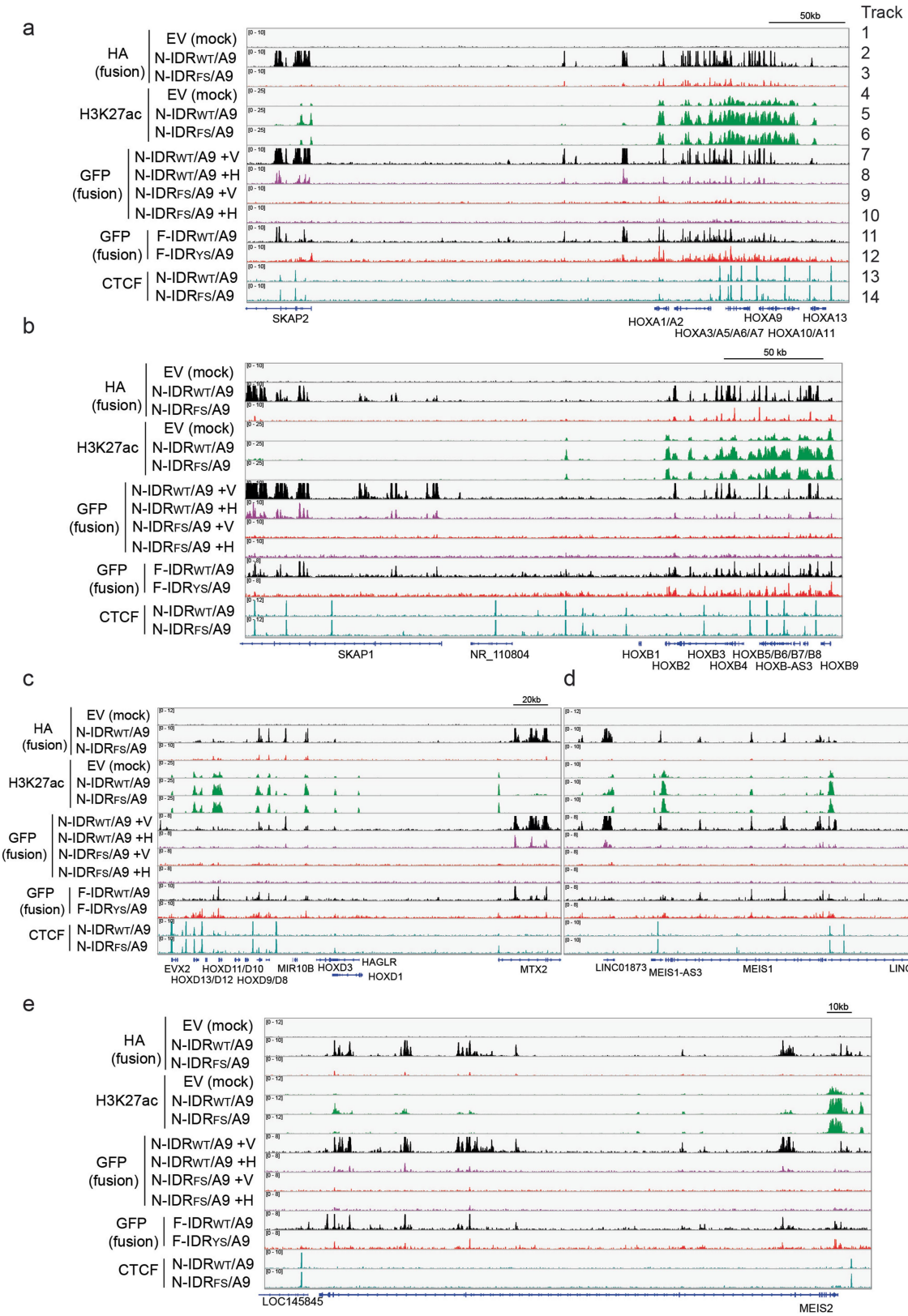
Rank	Motif	Name	P-value	% Targets with Motif
1		CDX4(Homeobox)/ Zebrafish Embryos-Cdx4. Myc-ChIP-Seq (GSE48254)	1e-3152	48.36%
2		Hoxa11(Homeobox) /ChickenMSG- Hoxa11.Flag-ChIP-Seq (GSE86088)	1e-3094	77.41%
3		Hoxd11(Homeobox)/ ChickenMSG- Hoxd11.Flag-ChIP-Seq (GSE86088)	1e-2981	75.44%
4		HOXB13(Homeobox)/ ProstateTumor- HOXB13-ChIP-Seq (GSE56288)	1e-2512	47.37%
5		Unknown(Homeobox)/ Limb-p300-ChIP-Seq	1e-2419	36.52%
6		Hoxd13(Homeobox)/ ChickenMSG-Hoxd13.Flag-ChIP-Seq (GSE86088)	1e-2361	62.50%
7		Cdx2(Homeobox)/ mES-Cdx2-ChIP-Seq (GSE14586)	1e-2150	41.41%
8		Hoxa9(Homeobox)/ ChickenMSG-Hoxa9.Flag-ChIP-Seq (GSE86088)	1e-2071	71.99%
9		Hoxa13(Homeobox)/ ChickenMSG- Hoxa13.Flag-ChIP-Seq (GSE86088)	1e-1963	75.80%
10		Hoxd10 (Homeobox)/ ChickenMSG-Hoxd10.Flag-ChIP-Seq (GSE86088)	1e-1930	42.83%

g

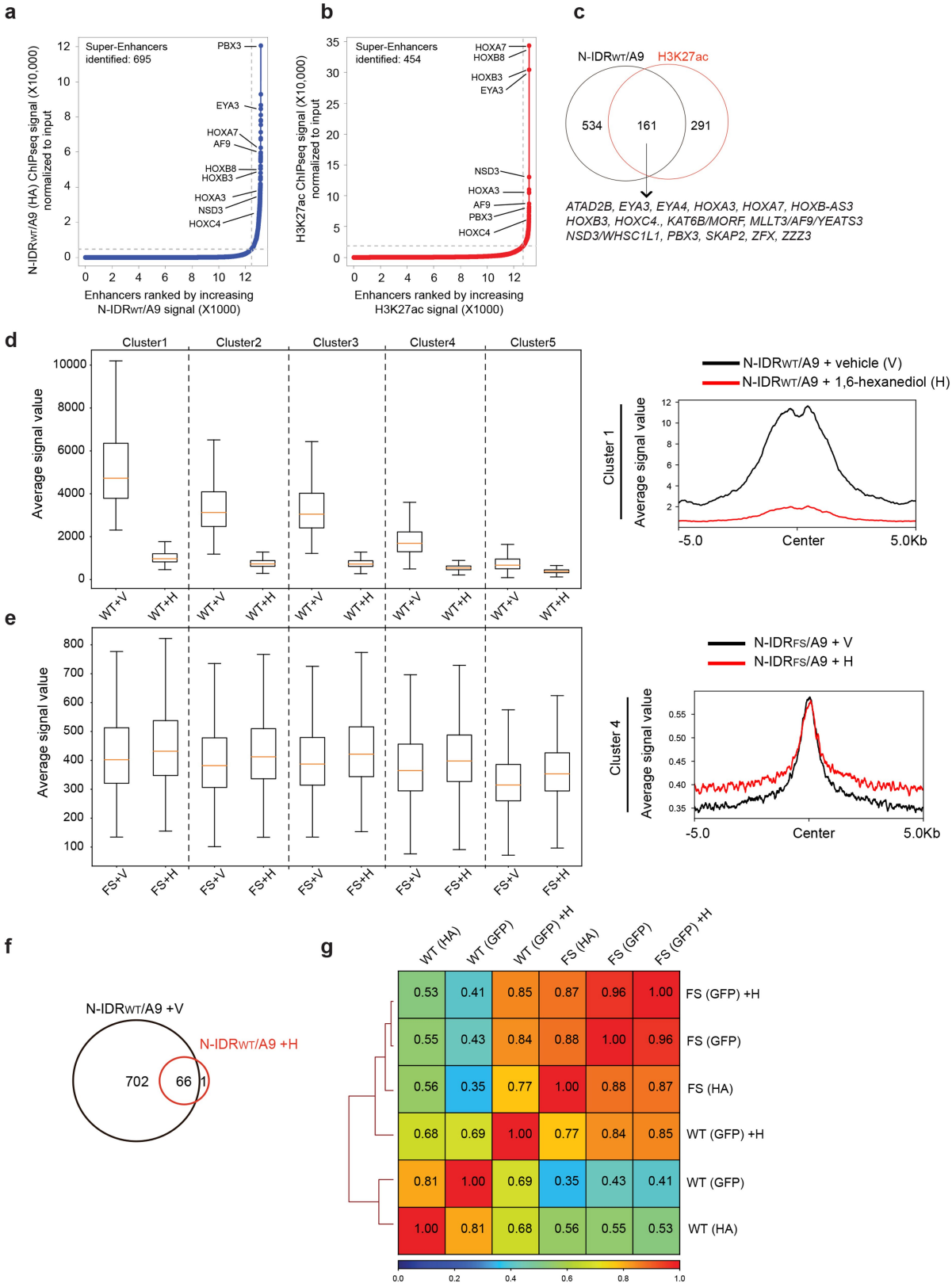
Rank	Motif	Name	P-value	% Targets with Motif
1		CDX4(Homeobox)/ Zebrafish Embryos-Cdx4. Myc-ChIP-Seq (GSE48254)	1e-1417	42.01%
2		HOXB13(Homeobox)/ ProstateTumor- HOXB13-ChIP-Seq (GSE56288)	1e-1327	41.01%
3		Hoxd11(Homeobox)/ ChickenMSG- Hoxd11.Flag-ChIP-Seq (GSE86088)	1e-1198	65.69%
4		Hoxa11 (Homeobox) ChickenMSG- Hoxa11.Flag-ChIP-Seq (GSE86088)	1e-1117	65.76%
5		Unknown(Homeobox)/ Limb-p300-ChIP-Seq	1e-1028	32.24%
6		Hoxd13(Homeobox)/ ChickenMSG-Hoxd13.Flag-ChIP-Seq (GSE86088)	1e-1009	50.72%
7		Hoxa13(Homeobox)/ ChickenMSG- Hoxa13.Flag-ChIP-Seq (GSE86088)	1e-902	62.18%
8		Cdx2(Homeobox)/ mES-Cdx2-ChIP-Seq (GSE14586)	1e-899	31.97%
9		Hoxa9(Homeobox)/ ChickenMSG-Hoxa9.Flag-ChIP-Seq (GSE86088)	1e-800	62.07%
10		Hoxc9(Homeobox)/ Ainv15-Hoxc9-ChIP-Seq (GSE21812)	1e-777	24.43%



Extended Data Fig. 3 | ChIP–seq reveals binding patterns of NUP98–HOXA9 that carries either wild-type or an Phe-to-Ser mutated IDR. a, Summary of the counts of ChIP–seq read tags for the indicated samples. **b**, Scatterplots showing correlation of global N-IDR$_{WT}$/A9 (left) or N-IDR$_{FS}$/A9 (right) ChIP–seq signals using either HA (*x* axis) or GFP (*y* axis) antibodies in two biological replicates of 293FT stable cells. Coefficient of determination (R^2) is determined by Pearson correlation. **c**, Total number of the called HA ChIP–seq peaks in stable 293FT cell lines expressing HA-tagged N-IDR$_{WT}$/A9 (left) or N-IDR$_{FS}$/A9 (middle) or empty vector control (right). **d**, **e**, Pie chart showing distribution of the indicated annotation feature among the called N-IDR$_{WT}$/A9 (**d**) or N-IDR$_{FS}$/A9 (**e**) ChIP–seq peaks in 293FT stable expression cells. **f**, **g**, Summary of the most enriched motifs identified within the called N-IDR$_{WT}$/A9 (**f**) or N-IDR$_{FS}$/A9 (**g**) ChIP–seq peaks in 293FT stable expression cells. Motif enrichment was statistically determined by ZOOPS scoring (zero or one occurrence per sequence) coupled with the hypergeometric enrichment calculations. **h**, Gene Ontology analysis of genes associated with broad super-enhancer-like peaks of N-IDR$_{WT}$/A9 as identified in 293FT stable cells. *P* values were determined by Fisher's exact test.

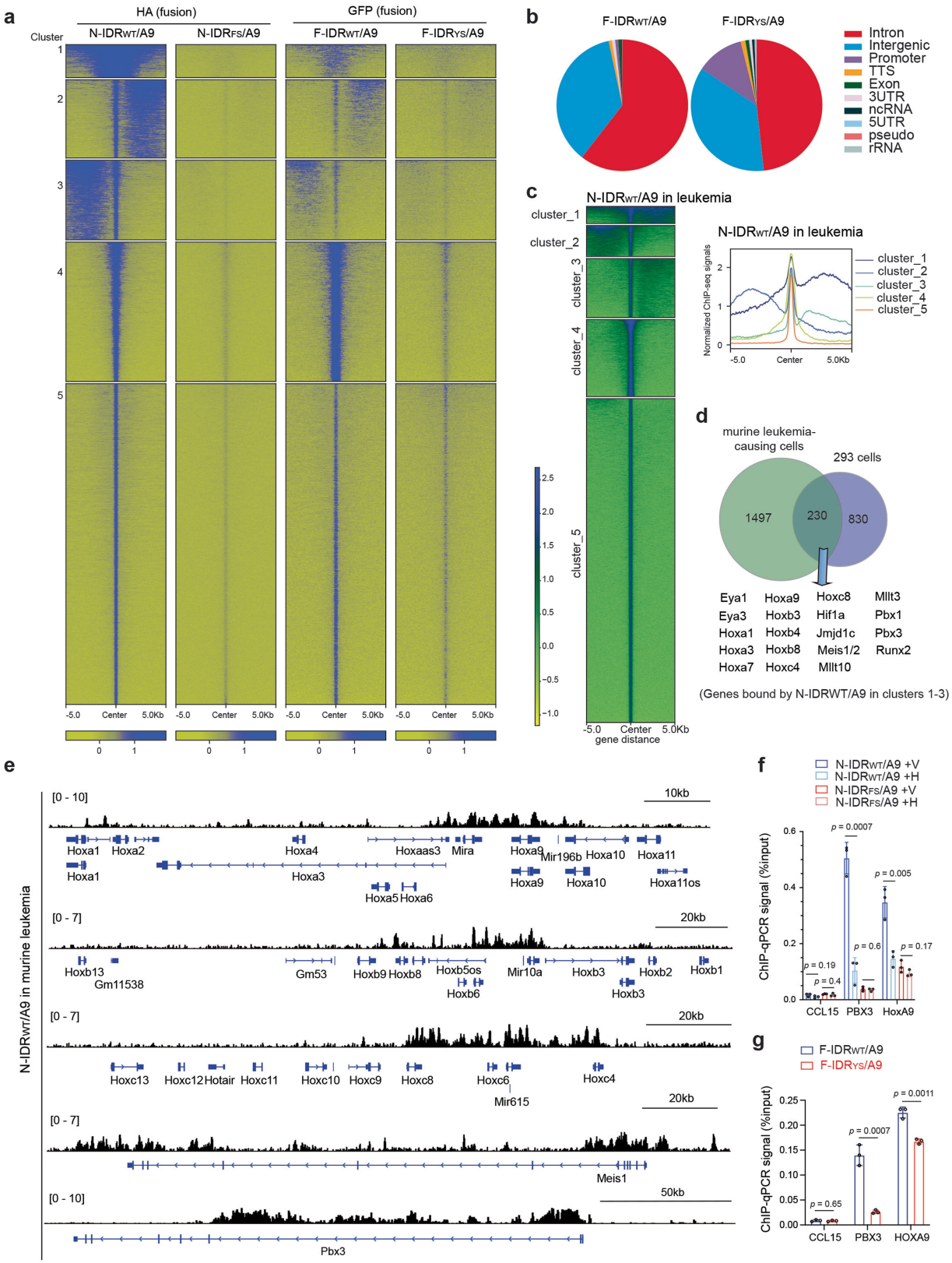


Extended Data Fig. 4 | Enhanced chromatin occupancy, as well as a broad super-enhancer-like binding pattern typically seen at leukaemia-related genomic loci, is characteristic for the LLPS-competent NUP98–HOXA9 (N-IDR$_{WT}$/A9) and not its LLPS-incompetent IDR mutant (N-IDR$_{FS}$/A9). **a**–**e**, Integrative genomics viewer (IGV) views for the indicated ChIP–seq signal at the well-known leukaemia-associated loci such as the *HOXA* (**a**), *HOXB* (**b**) and *HOXD* (**c**) gene clusters, *MEIS1* (**d**) and *MEIS2* (**e**). Samples from top to bottom are HA (tracks 1–3) and H3K27ac (tracks 4–6) ChIP–seq signals in the 293FT cells stably expressed with either empty vector (tracks 1 and 4; EV in track 1 acts as a negative control for HA ChIP) or the HA-tagged N-IDR$_{WT}$/A9 (tracks 2 and 5) or N-IDR$_{FS}$/A9 (tracks 3 and 6), GFP ChIP–seq signals (tracks 7–12) in the 293FT cells stably expressed with GFP-tagged N-IDR$_{WT}$/A9 (tracks 7–8 represent samples after treatment with vehicle or 10% 1,6-hexanediol, respectively, for 1 min), N-IDR$_{FS}$/A9 (tracks 9–10 represent samples after treatment with vehicle or 1,6-hexanediol, respectively), F-IDR$_{WT}$/A9 (track 11) or F-IDR$_{YS}$/A9 (track 12), as well as CTCF ChIP–seq in 293FT cells with N-IDR$_{WT}$/A9 (track 13) or N-IDR$_{FS}$/A9 (track 14). HA and CTCF ChIP–seq signals were normalized to input signals, whereas GFP ChIP–seq, conducted in the spike-in controlled experiments, normalized to the spike-in *Drosophila* chromatin signals (those from antibody of a *Drosophila*-specific histone, H2Av).



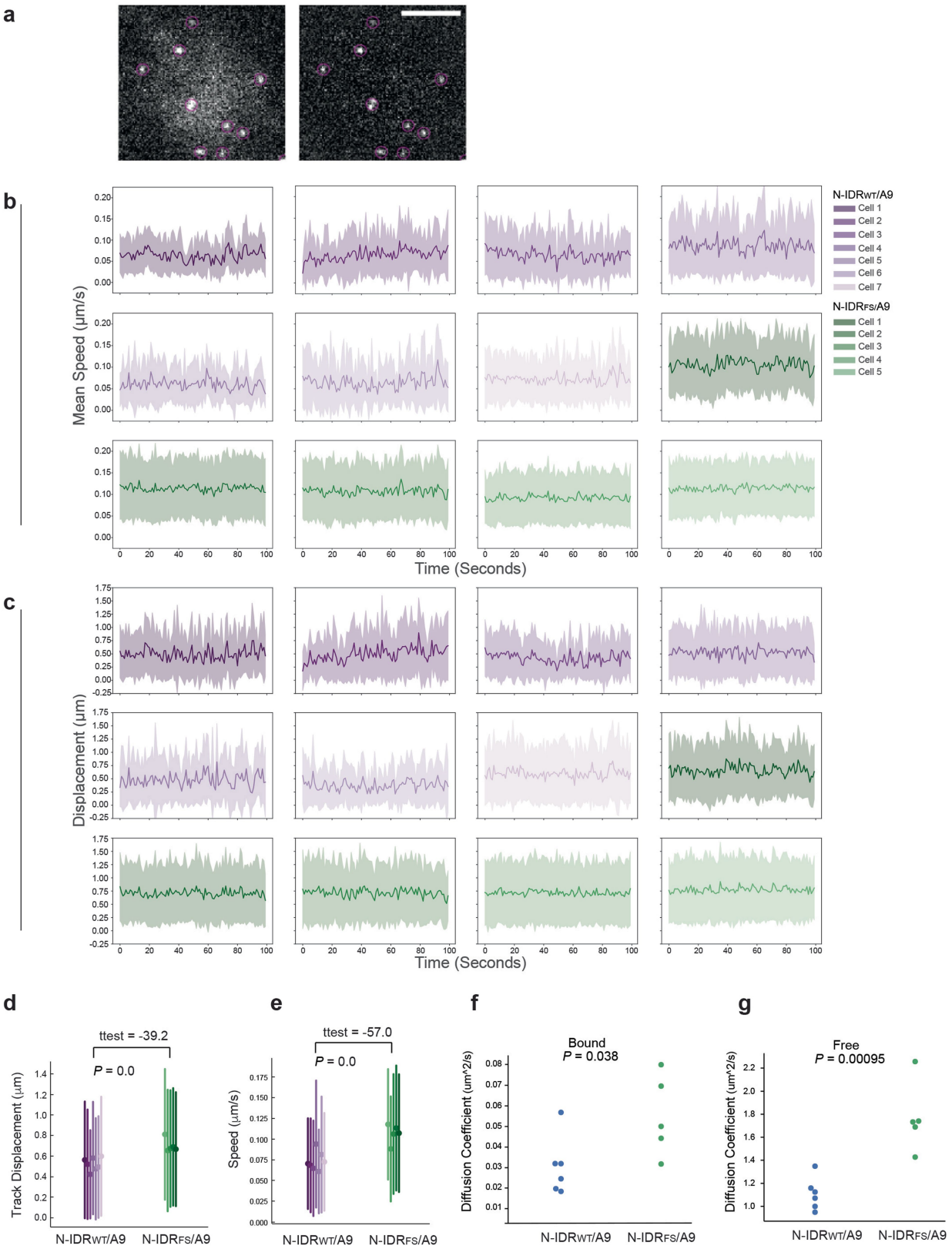
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Formation of the enhanced and broad super-enhancer-like binding patterns of leukaemia-related chimera transcription factors requires an intact phase-separation-competent IDR. a, **b**, Hockey-stick plot shows distribution of the input-normalized ChIP–seq signals of N-IDR$_{WT}$/A9 (**a**) or H3K27ac (**b**) across all enhancers annotated by H3K27ac peaks (transcriptional start site ± 2.5 kb regions were excluded) in 293FT cells. Dotted line indicates the threshold level set by the ROSE algorithm to call super-enhancers. Relative rankings of super-enhancers associated with some example genes are shown. **c**, Venn diagram illustrates overlap among super-enhancers called based on N-IDR$_{WT}$/A9 and H3K27ac ChIP–seq signals. **d**, **e**, Box plots showing averaged ChIP–seq signals for *k*-means clustered peaks (see Fig. 2b) of the LLPS-competent N-IDR$_{WT}$/A9 (WT; **d**) show a marked reduction in binding after treatment of 293FT stable cells with 1,6-hexanediol (WT+H), relative to treatment with vehicle control (WT+V); this reduction is particularly significant for peak clusters 1–3 shown in Fig. 2b. By contrast, genomic binding of N-IDR$_{FS}$/A9 (FS; **e**) shows general insensitivity to the same treatment of 1,6-hexanediol (FS+H) in comparison to mock (FS+V). Right, averaged ChIP–seq signal distribution profiles are shown for N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 over a 10-kb region in the indicated peak cluster as an example. Box plots as defined in Fig. 3f. **f**, Venn diagram to compare genes associated with the broad super-enhancer-like peaks of N-IDR$_{WT}$/A9 after treatment with 1,6-hexanediol (+H), relative to vehicle control (+V), after treatment for 1 min. **g**, Hierarchical clustered heat maps for the pairwise correlation of ChIP–seq signals between each of the indicated sample. The coefficients were determined by Pearson correlation. HA and GFP represent ChIP–seq for HA-tagged and GFP-tagged chimera transcription factors, respectively.

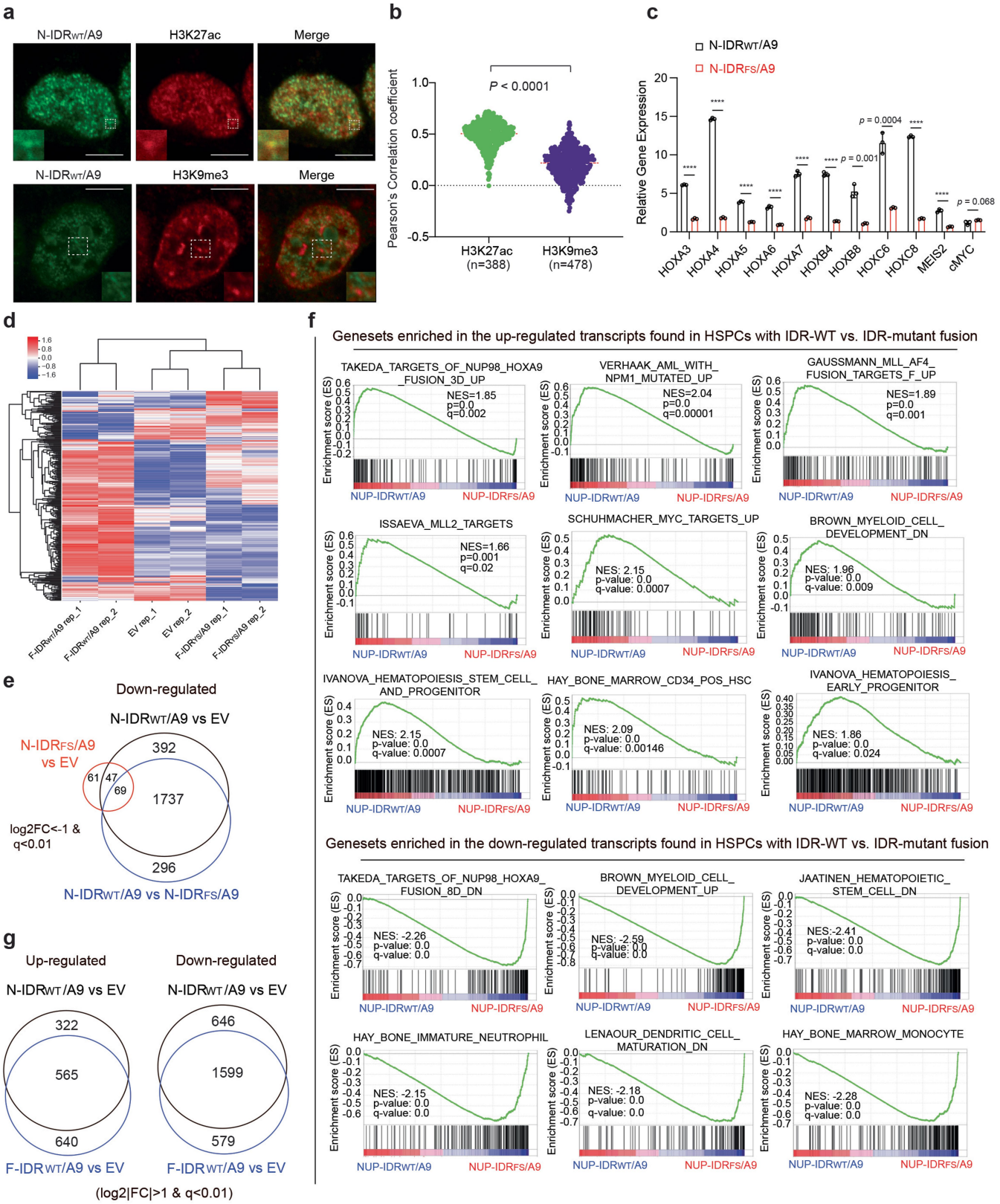


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | The phase-separation-promoting property within F-IDR is sufficient to induce the enhanced binding of the chimeric transcription factor. a, Heat maps showing the *k*-mean clustering of ChIP–seq signals for chimeric transcription factors that contain the NUP98 IDR (N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9, two panels on the left) or FUS IDR (F-IDR$_{WT}$/A9 and F-IDR$_{YS}$/A9, two panels on the right) reveal a similarly enhanced binding for the LLPS-competent chimera that carries a wild-type form of IDR, relative to its LLPS-incompetent IDR mutant, in 293FT stable expression cells. Note that, although to a lesser degree, the artificially created F-IDR$_{WT}$/A9 fusion also displays a broad, super-enhancer-like binding pattern at the same sites observed with the N-IDR$_{WT}$/A9 fusion. **b**, Pie chart showing percentage distribution of the indicated genomic annotation feature among the ChIP–seq peaks of GFP-tagged F-IDR$_{WT}$/A9 (left) or F-IDR$_{YS}$/A9 (right) in the 293FT stable expression cells. **c**, Heat maps (left) and its averaged ChIP–seq signal distribution profiles (right) for *k*-mean clustered peaks of N-IDR$_{WT}$/A9 in the transformed mouse HPSCs. **d**, Venn diagram showing overlap between the annotated genes associated with the clusters 1–3 of N-IDR$_{WT}$/A9 ChIP–seq peaks detected in the transformed mouse HPSCs (left) and the 293FT stable expression cells (right). Examples of the shared oncogenes are shown below. **e**, IGV views of N-IDR$_{WT}$/A9 ChIP–seq signals (GFP-tagged) at the indicated loci in mouse HSPCs transformed by this chimera. **f**, ChIP–qPCR to assess the binding of GFP-tagged N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9 at *CCL15* (a negative control region), *PBX3* and *HOXA9* in the 293FT stable cells after treatment with 10% 1,6-hexanediol for 1 min (+H), relative to mock (+V). ChIP signals, normalized to those of input, are presented as mean ± s.d. of three replicate experiments. **g**, ChIP–qPCR to assess the binding of GFP-tagged F-IDR$_{WT}$/A9 or F-IDR$_{YS}$/A9 at *CCL15* (a negative control region), *PBX3* and *HOXA9* in the 293FT stable cells. ChIP signals, normalized to those of input, are presented as mean ± s.d. of three replicates.

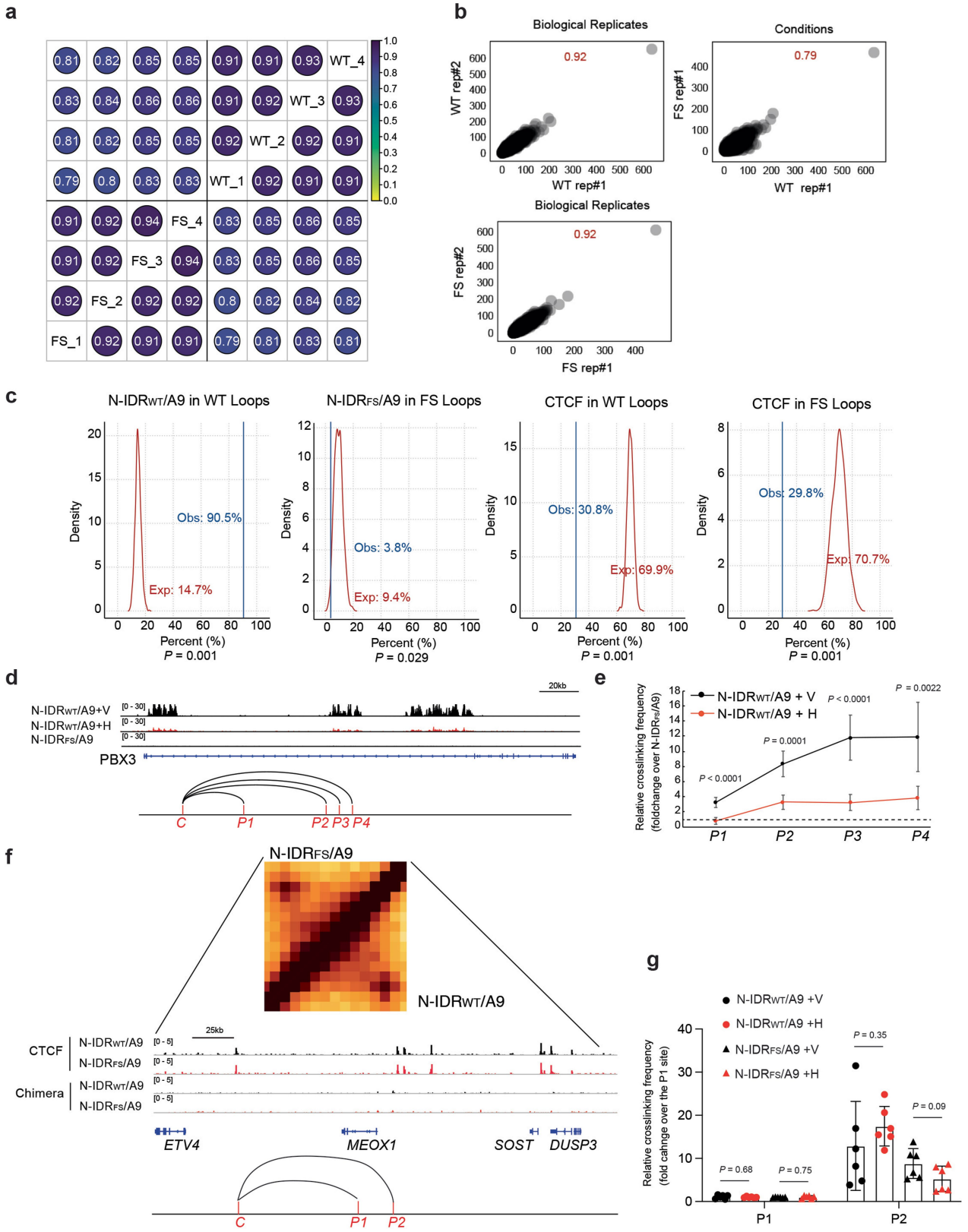


Extended Data Fig. 7 | Single-molecule tracking shows that phase-separation-competent N-IDR$_{WT}$/A9 proteins behave with less dynamic characteristics, compared with phase-separation-incompetent N-IDR$_{FS}$/A9. a, Representative images of single-molecule particles identification in an N-IDR$_{WT}$/A9-expressing cell, either the original captured image (left) or after processing to remove background (right). Scale bars, 5 µm. **b**, **c**, Single-particle tracks for mean speed (**b**) and mean displacement (**c**) of either N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9 single molecules within the temporally registered reference frame binned into 1-s intervals. **d**, **e**, Displacement (**d**) and mean velocity (**e**) of single-particle tracks indicate that N-IDR$_{WT}$/A9 with the LLPS-competent IDR (WT) is less mobile and navigates nuclear space at a slower rate than its LLPS-incompetent IDR mutant (FS). Dots indicate mean values in a single cell. Line indicates one standard deviation. *P* values determined by two-sided *t*-test. **f**, **g**, The diffusion coefficient for chromatin-bound (**f**) and freely diffusing states (**g**) of N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9, calculated based on single-molecule tracking studies of its 293FT stable expression cells. *P* values determined by two-sided *t*-test.



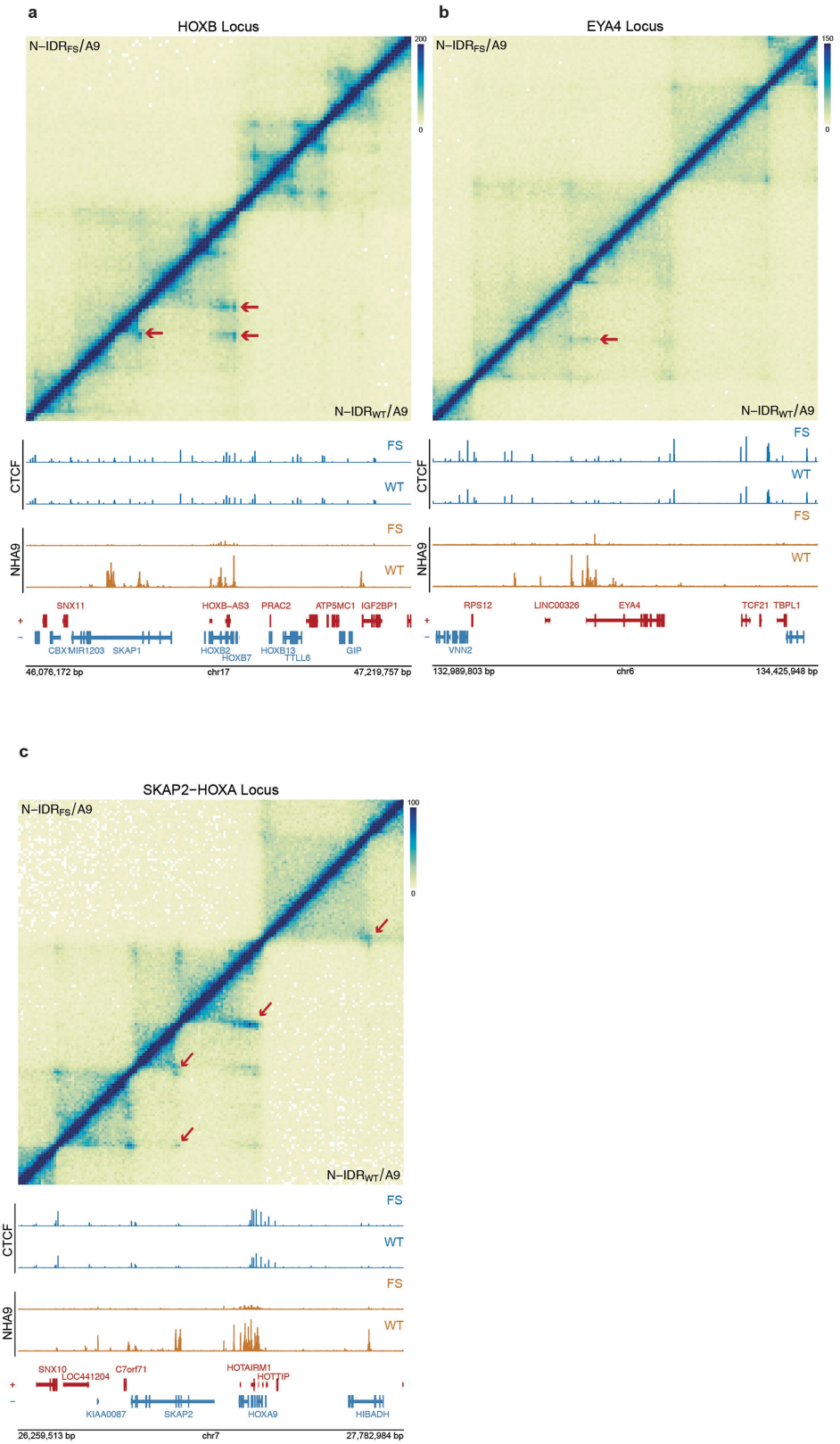
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | An LLPS-competent IDR within the leukaemia-related transcription factor chimera is essential for potentiating transcriptional activation of the downstream oncogenic gene-expression program. a, Fixed cell immunostaining for the 3×HA-3×Flag-tagged N-IDR$_{WT}$/A9 (left; anti-Flag) and the indicated histone modification (middle) in the 293FT stable expression cells. Top panels show the enlarged images of an example region within the white dotted box shown in the bottom panels, in which the transcription factor chimera is co-localized with H3K27ac (top) and not H3K9me3 (bottom). Scale bars, 10 μm. **b**, Pearson's correlation coefficient values between N-IDR$_{WT}$/A9 and the indicated histone modification. The red dotted line indicates the calculated average value of each plot. The calculated means (red dotted lines) were compared with an independent two-tailed Student's *t*-test. n, the number of cells analysed. **c**, RT–qPCR to assess the effect of phase separation in target gene expression in 293FT cells. All of the tested *HOX* and *MEIS2* genes are direct targets of both N-IDR$_{WT}$/A9 and N-IDR$_{FS}$/A9 based on ChIP–seq, whereas *MYC* is not and serves as a negative control. Note that LLPS-competent N-IDR$_{WT}$/A9 induces significantly more upregulation of target genes, relative to LLPS-incompetent N-IDR$_{FS}$/A9. PCR signals were normalized first to those of an internal control (18S RNA) and then to vector-expressing cells and presented as mean ± s.d. of three replicated experiments. ***$P < 0.001$; ****$P < 0.0001$; two-sided *t*-test. n.s., not significant. **d**, Heat map illustrating relative expression of the 374 genes that show significant upregulation post-transduction of F-IDR$_{WT}$/A9, compared to empty vector and its IDR-mutant form (F-IDR$_{YS}$/A9), in 293FT stable expression cells. **e**, Venn diagrams showing the overlap of the significantly downregulated genes identified 7 days after transduction of the indicated construct into mouse HPSCs. **f**, Gene set enrichment analysis (GSEA) shows that, compared with that of N-IDR$_{FS}$/A9, the expression N-IDR$_{WT}$/A9 in mouse HPSCs is positively correlated with the indicated leukaemia- or HSPC-related gene sets (top) and negatively correlated with the indicated differentiation-related gene sets (bottom). The *P* value was calculated by an empirical phenotype-based permutation test; the false discovery rate (*q*) is adjusted for gene set size and several hypotheses testing whereas the *P* value is not. **g**, Venn diagrams showing the overlap of the significantly upregulated (left) or downregulated (right) genes identified after transduction of the indicated construct into mouse HPSCs.

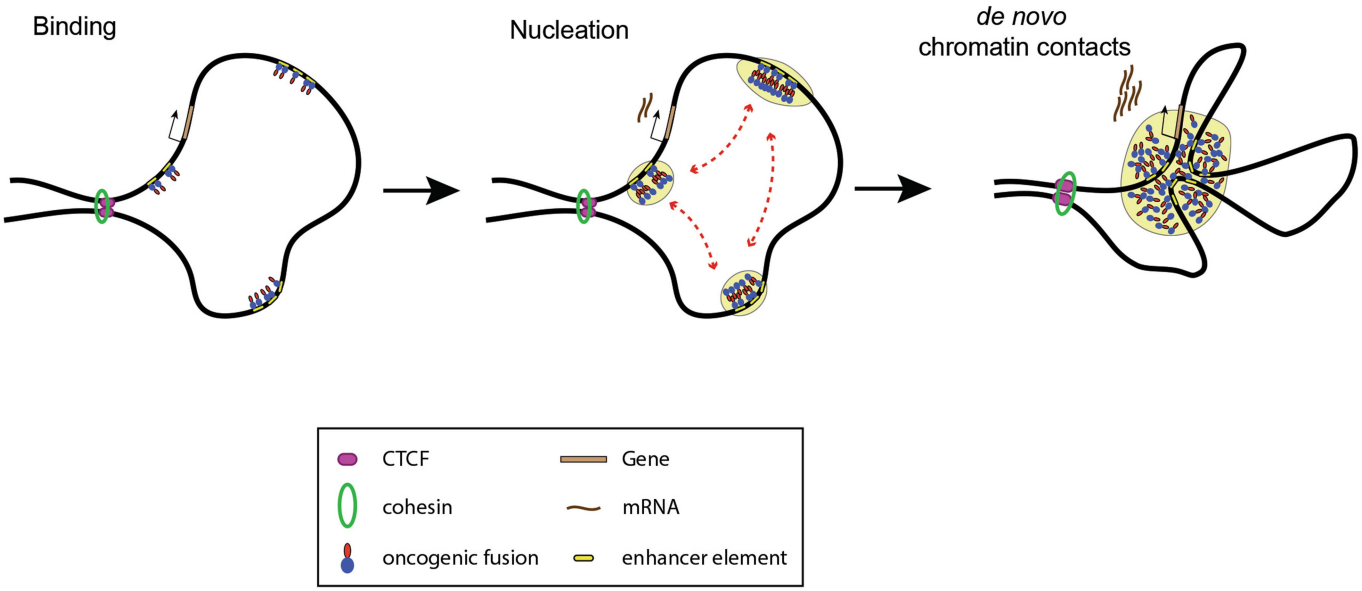


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Hi-C mapping reveals that a phase-separation-competent IDR within NUP98–HOXA9 is required to induce formation of CTCF-independent chromatin loops at the leukaemia-related genomic loci. a, Matrix of Pearson correlation coefficients of loop counts among and between biological replicates of N-IDR$_{WT}$/A9 (WT; $n = 4$ replicates) or N-IDR$_{FS}$/A9 (FS; $n = 4$ replicates) conditions. Numbers following WT or FS indicate biological replicate for that condition. **b**, Example correlation plots of loop counts between biological replicates and conditions. **c**, All loops were partitioned into either WT- or FS-specific loops and split into separate loop anchors. Loop anchors were then intersected with ChIP–seq peaks of N-IDR/A9 or CTCF. The percentage of observed (Obs.) overlaps for each feature is shown as a vertical blue line. The red line shows the expected (Exp.) distribution of overlaps as determined by randomly sampling loop anchors and calculating the overlap of each feature 1,000 times. *P* values were determined by summing the number of expected values greater than (or less than if the observed value was less than the mean) the observed value for that feature. **d**–**g**, 3C–qPCR assays measuring the change in crosslinking frequency of either an N-IDR$_{WT}$/A9-specific loop at the *PBX3* locus (**d**, **e**) or a CTCF-dependent loop (**f**, **g**; at Chr17 (41604677–41883642)) after treatment of 293FT stable cells with 10% 1,6-hexanediol for 1 min (+H), relative to mock (+V). The IGV view panels at **d** and **f** show the indicated ChIP–seq signals, with positions of the used 3C–PCR primers labelled under IGV tracks. PCR was performed using the same constant forward primer (C) paired with a differently numbered reverse primer (P1 to P4) at each locus tested. Panels **e** and **g** are plotted with signals of 3C–qPCR measuring the relative crosslinking frequency at *PBX3* (**d**, **e**) or a Chr17 locus with CTCF loop (**f**, **g**) before (V) and after (H) treatment with 1,6-hexanediol. Signals in **e** are normalized to those of the N-IDR$_{FS}$/A9-expressing cells ($n = 3$ replicated experiments). *P* values were determined by two-sided *t*-test. Data are mean ± s.d. of three or six replicates.



Extended Data Fig. 10 | Hi-C mapping reveals the chromatin loops specific to cells with the LLPS-competent NUP98–HOXA9, compared with the LLPS-competent mutant, at leukaemia-relevant gene loci. Views for Hi-C mapping, RNA-seq and ChIP–seq for CTCF, N-IDR/A9 and H3K27ac at the *HOXB* (**a**), *EYA4* (**b**), and *SKAP2-HOXA* loci (**c**) in 293FT stable cells expressing either N-IDR$_{WT}$/A9 (WT) or N-IDR$_{FS}$/A9 (FS). Hi-C mapping views (top) show results from the N-IDR$_{WT}$/A9 or N-IDR$_{FS}$/A9 expressing cells (bottom and top diagonal, respectively). Corresponding ChIP–seq and gene tracks are shown below each Hi-C plot. N-IDR$_{WT}$/A9 loops are indicated by red arrows.



Extended Data Fig. 11 | Model illustrating requirement of LLPS-competent IDR within NUP98–HOXA9 for leukaemogenesis and activation of the oncogenic gene-expression program. The LLPS-competent IDR contained with NUP98–HOXA9 is crucial for promoting long-distance chromatin looping between proto-oncogene promoter and enhancers, which thus induces an oncogenic gene-expression program and malignant development.

nature research

Corresponding author(s): Gang Greg Wang

Last updated by author(s): Apr 16, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom scripts were used to collect data for this study.

Data analysis	FlowJo X.0.7 was used for Flow cytometry analysis. ChIP-seq peak calling Trim galore version 0.6.1 was used for adapter trimming and quality filter for all reads. STAR v2.7.1a was used to align the reads to the human genome (hg19). Samtools (v1.9), Picard MarkDuplicates funtion (ver 2.20.4), and bedtools (v2.28.0) were used to remove the non-primary alignment, PCR duplicates, or blacklist regions from aligned data. MACS2(v2.1.1) was used for peak calling. Deeptools (v3.3.0) was used to make bigwig files, heatmaps, and averaged plottings of ChIP-seq signal. These bigwig files were visualized using IGV v2.5.3. Custom scripts of R(v3.5.0) or Python(v3.6) were used for some statistical analysis. Hi-C Analysis The Juicer pipeline v1.5.6 (Durand et al. 2016) was used to create Hi-C maps, using bwa v0.7.17 (Li H. and Durbin R. 2009) to align reads to the human genome (hg19). Loops were detected using HiCCUPS from the Juicer tools software v1.11.09 (Durand et al. 2016) with the following parameters: -m 2048 -c 2 -r 5000,10000,25000 -k KR -f 0.1,0.1,0.1 -p 4,2,1 -i 8,6,4 -t 0.2,1.5,1.5,1.75 -d 30000,30000,60000. Raw counts for each looping interaction were extracted with the Straw API v0.0.1 (Durand et al. 2016) and used for differential loop calling in R (v3.5.1) with DESeq2 v1.22.2 (Love, Huber, and Anders 2014). Aggregate peak analysis was conducted in R (v3.5.1) using the Straw API (v0.0.1). Overlap between loop anchors and ChIP-seq peaks (see ChIP-seq peak calling) was performed using Bedtools v2.29.2 (Quinlan et al. 2010) and bedtoolsr v2.29.0-5 (Patwardhan et al. 2020). Statistical analysis was conducted in R (v3.5.1) with the FSA package v0.8.31 (Ogle et al. 2020). RNA-seq analysis The fastq files were aligned to the GRCh38 human genome (GRCh38.d1.vd1.fa) or the mm10 mouse genome (GRCm38.p4) using STAR v2.4.2 (Dobin et al. 2013) with the following parameters: --outSAMtype BAM Unsorted --quantMode TranscriptomeSAM. Transcript abundance for each sample was estimated with salmon v0.1.19 (Patro et al. 2017) to quantify the transcriptome defined by Gencode v22. Gene level counts were summed across isoforms and genes with low counts (maximum expression < 10) were filtered for the downstream analyses. We tested genes for differential expression in DESeq2 v1.38.2 (Love, Huber, and Anders 2014) in R.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Next-generation sequencing data from ChIP-seq, RNA-seq and Hi-C experiments for this current study are deposited in the NCBI GEO under accession number GSE144643. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023548 and 10.6019/PXD023548. Gel raw images (for Fig 1b and Extended Data Figs 1b, 1d, 1g and 2a) and FACS gating control (for Extended Data Fig 2f) are provided in Supplementary Fig 1 of the paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was estimated from the preliminary experiments or from our previously published studies (Nature 459, 847-851 (2009); Nat Commun 12, 1045, doi:10.1038/s41467-021-21357-3 (2021)). No statistical method was applied to predetermine sample size. The cell counting, RT-qPCR, ChIP-qPCR and 3C-qPCR were conducted with three independent experiments or samples. ChIP-seq of fusion was conducted with two independent samples (for example, either GFP or HA ChIP-seq). RNA-seq was conducted with two independent samples. Hi-C was conducted with four independent experiments and, following confirmation of high correlation, Hi-C read tags were combined to gain depth.
Data exclusions	No
Replication	All experiments were reproduced reliably in independent setting for at least two or three times.
Randomization	Mice were randomly allocated into experimental groups by the mouse core facility at UNC-CH.
Blinding	Investigators were blinded to group allocation during experiments and not blinded to data analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: anti-HA tag antibody-ChIP Grade, Abcam, ab9110; anti-GFP, Abcam, ab290; anti-FLAG, Sigma, F1804; anti-H3K27Ac, Abcam, ab4729; anti-H3K9me3, Abcam, ab8898; anti-CTCF, Diagenode, C15410210; anti-GAPDH, Santa Cruz, sc-25778; anti-H2Av, Activemotif, 39715; anti-NUP98 GLEBS (A gift from Jan Deurson; refer to the paper: Molecular and cellular biology 19, 764-776); goat anti-mouse IgG HRP, Santa Cruz Biotechnology, sc-2005; goat anti-rabbit IgG HRP, Santa Cruz Biotechnology, sc-2004; Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594, Life technologies, A-21207; Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 488, Invitrogen, A-32723. The information of the antibody source and the used dilutions is provided in the Supplementary Table 8 of the paper.

Validation: All antibodies were validated by Western blotting or immunofluorescence using cells expressing tagged protein of interest or gene knockdown vs. the negative control sample. For histone antibodies, specificity had been extensively tested by the suppliers and independent investigators (for example, the users' generated database for histone antibodies: http://www.histoneantibodies.com/). For validation of anti-NUP98 GLBES, we have used cell line expressing GFP-tagged NUP98-HOXA9, which is either full-length or with the GLEBS deletion, and confirmed its specificity based on a strong signal for full-length NUP98-HOXA9 and a lack of signal with the GLEBS-deleted NUP98-HOXA9.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): Cell lines used in the study included HEK293T (ATCC #CRL-3216), HeLa (ATCC #CCL-2), HeLa NUP98-HOXA9 N51S (A gift from Dr. Birthe Fahrenkrog).

Authentication: Authentication of cell line identities, including parental and their derived lines, was ensured by the Tissue Culture Facility (TCF) affiliated to UNC Lineberger Comprehensive Cancer Center with the genetic signature profiling and fingerprinting analysis.

Mycoplasma contamination: Cells were negative for mycoplasm. Every month, a routine examination of cell lines in culture for any possible mycoplasma contamination was performed using commercially available detection kits (Lonza).

Commonly misidentified lines (See ICLAC register): No commonly misidentified cell lines were used in the study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals: All animal experiments were approved by and performed in accordance with the guidelines of the UNC Institutional Animal Care and Use Committee (IACUC). The balb/C mice (female 10-week-old) were purchased from Jax Lab and maintained by Animal Studies Core, the UNC Lineberger Comprehensive Cancer Center.

Wild animals: No wild animals were used in the study.

Field-collected samples: No field-collected samples were used in the study.

Ethics oversight: All animal experiments were approved by and performed in accordance with the guidelines of the UNC Institutional Animal Care and Use Committee (IACUC). Experienced staff of UNC Animal Studies Core provides support of daily care and monitoring.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144643
Files in database submission	HEK293 EV control cell Input (matched to HA-IP) HEK293 EV control cell HA-IP HEK293 WT NUP98-HOXA9 (HA-tagged) Input (matched to HA-IP) HEK293 WT NUP98-HOXA9 (HA-tagged) HA-IP HEK293 FS mutant NUP98-HOXA9 (HA-tagged) Input (matched to HA-IP) HEK293 FS mutant NUP98-HOXA9 (HA-tagged) HA-IP HEK293 WT NUP98-HOXA9 (GFP-tagged; mock treated) Input HEK293 WT NUP98-HOXA9 (GFP-tagged; mock treated) GFP-IP HEK293 WT NUP98-HOXA9 (GFP-tagged; Hex treated) Input HEK293 WT NUP98-HOXA9 (GFP-tagged; Hex treated) GFP-IP HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; mock treated) Input HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; mock treated) GFP-IP HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; Hex treated) Input HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; Hex treated) GFP-IP HEK293 EV control cell Input (matched to H3K27ac-IP) HEK293 EV control cell H3K27ac-IP HEK293 WT NUP98-HOXA9 (HA-tagged) Input (matched to H3K27ac-IP) HEK293 WT NUP98-HOXA9 (HA-tagged) H3K27ac-IP HEK293 WT NUP98-HOXA9 CTCF-IP HEK293 FS mutant NUP98-HOXA9 (HA-tagged) Input (matched to H3K27ac-IP) HEK293 FS mutant NUP98-HOXA9 (HA-tagged) H3K27ac-IP HEK293 FS mutant NUP98-HOXA9 CTCF-IP HEK293 WT FUS-HOXA9 (GFP-tagged) Input HEK293 WT FUS-HOXA9 (GFP-tagged) GFP-IP HEK293 YS mutant FUS-HOXA9 (GFP-tagged) Input HEK293 YS mutant FUS-HOXA9 (GFP-tagged) GFP-IP mouse HPSC WT NUP98-HOXA9 (GFP-tagged) Input mouse HPSC WT NUP98-HOXA9 (GFP-tagged) GFP-IP
Genome browser session (e.g. UCSC)	no longer applicable

Methodology

Replicates	We have used two different tags (GFP and HA) for NUP98 fusion.
Sequencing depth	sample name raw reads mapped reads Read type Read length mouse HPSC WT NUP98-HOXA9 (GFP-tagged) GFP-IP 40931756 38010893 single-end 50 mouse HPSC WT NUP98-HOXA9 (GFP-tagged) Input 33846318 29295519 single-end 50 HEK293 FS mutant NUP98-HOXA9 (HA-tagged) H3K27ac-IP 70506235 67860778 single-end 50 HEK293 WT NUP98-HOXA9 (HA-tagged) H3K27ac-IP 73156008 70478585 single-end 50 HEK293 EV control cell Input (matched to HA-IP) 68931398 62957130 single-end 50 HEK293 EV control cell HA-IP 75884089 69324916 single-end 50 HEK293 FS mutant NUP98-HOXA9 (HA-tagged) Input (matched to HA-IP) 80050489 74412157 single-end 50 HEK293 FS mutant NUP98-HOXA9 (HA-tagged) HA-IP 90580775 80879709 single-end 50 HEK293 WT NUP98-HOXA9 (HA-tagged) Input (matched to HA-IP) 74581179 67897568 single-end 50 HEK293 WT NUP98-HOXA9 (HA-tagged) HA-IP 61108652 56802384 single-end 50 HEK293 EV control cell H3K27ac-IP 70455472 67944972 single-end 50 HEK293 EV control cell Input (matched to H3K27ac-IP) 126621376 119865471 single-end 50 HEK293 FS mutant NUP98-HOXA9 CTCF-IP 33514784 32534621 single-end 50 HEK293 FS mutant NUP98-HOXA9 (HA-tagged) Input (matched to CTCF-IP) 23761971 22651990 single-end 50 HEK293 WT NUP98-HOXA9 CTCF-IP 34651089 33512424 single-end 50 HEK293 WT NUP98-HOXA9 (HA-tagged) Input (matched to CTCF-IP) 25039295 23814744 single-end 50 HEK293 YS mutant FUS-HOXA9 (GFP-tagged) Input 26768771 25946511 single-end 50 HEK293 YS mutant FUS-HOXA9 (GFP-tagged) GFP-IP 35721610 34874243 single-end 50 HEK293 WT FUS-HOXA9 (GFP-tagged) Input 27059808 26036690 single-end 50 HEK293 WT FUS-HOXA9 (GFP-tagged) GFP-IP 31143303 30434182 single-end 50 HEK293 WT NUP98-HOXA9 (GFP-tagged; mock treated) Input 37351884 35758008 single-end 50 HEK293 WT NUP98-HOXA9 (GFP-tagged; mock treated) GFP-IP 31911380 29855733 single-end 50 HEK293 WT NUP98-HOXA9 (GFP-tagged; Hex treated) Input 28660721 27593795 single-end 50 HEK293 WT NUP98-HOXA9 (GFP-tagged; Hex treated) GFP-IP 37312622 35303786 single-end 50 HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; mock treated) Input 39984308 38610417 single-end 50 HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; mock treated) GFP-IP 41810282 40068777 single-end 50 HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; Hex treated) Input 25485571 24688075 single-end 50 HEK293 FS mutant NUP98-HOXA9 (GFP-tagged; Hex treated) GFP-IP 37538073 36222062 single-end 50
Antibodies	Rabbit anti-HA, Abcam, ab9110; Rabbit anti-GFP, Abcam, ab290; Rabbit anti-H3K27 acetyl, Abcam, ab4729; Rabbit anti-CTCF, Diagenode, C15410210
Peak calling parameters	MACS2 with the parameter: --nomodel --pvalue 1e-05 --extsize 250

Data quality	FastQC
Software	STAR (v2.7.1a), DeepTools(v3.3.0), MACS(v2.1.1),

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were washed once in the cold FACS buffer (PBS with 5% of FBS added) to a final density of 10 million cells/ml and then resuspended and incubated in the FACS buffer added with the respective antibodies (1:100 dilution) for 30 min on ice. The cell pellets were washed with FACS buffer and the stained cells were subject to analysis with the FACS machine.
Instrument	Attune Nxt (Life Technologies)
Software	FlowJo X.0.7 was used for Flow cytometry analysis
Cell population abundance	Not applicable (all cells used).
Gating strategy	Parental cells without antibody staining and set gates for analysis. The gating control and strategy data is provided in Supplementary Figure of the paper.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.