Nucleic Acids Research, 2025, **53**, gkaf056
https://doi.org/10.1093/nar/gkaf056
Advance access publication date: 12 February 2025
NAR Breakthrough Article

NAR Breakthrough Article

Biomolecular condensation of human IDRs initiates endogenous transcription via intrachromosomal looping or high-density promoter localization

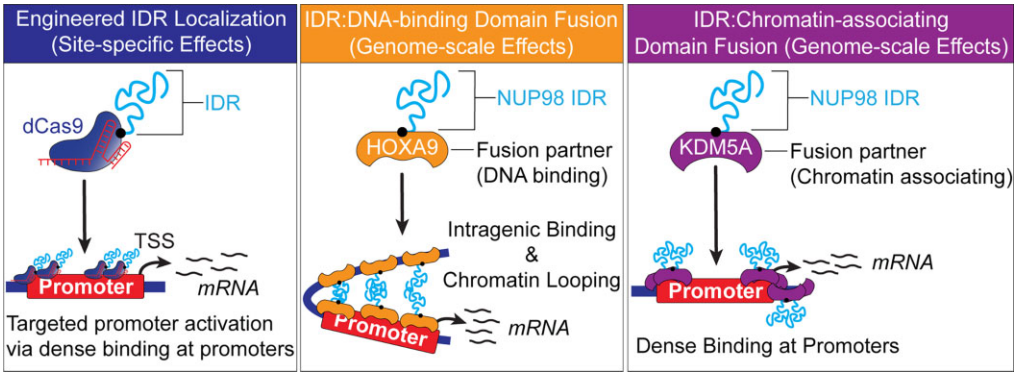
Jing Li[1], Shizhe Liu[2], Sunghwan Kim[1], Jacob Goell[1], Zachary Allen Drum[3], John Patrick Flores[3], Alex J. Ma[1], Barun Mahata[1], Mario Escobar[1], Alex Raterink[4], Jeong Hyun Ahn[5], Erik R. Terán[2], Rosa Selenia Guerra-Resendez[4], Yuhao Zhou[1], Bo Yu[6], Michael R. Diehl[1,7], Gang Greg Wang[5,8], Anna-Karin Gustavsson[2,4,7], Douglas H. Phanstiel[3], Isaac B. Hilton[1,2,4,*]

[1]Department of Bioengineering, Rice University, Houston, TX, 77030, United States
[2]Department of BioSciences, Rice University, Houston, TX, 77030, United States
[3]Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, 27599, United States
[4]Systems, Synthetic, and Physical Biology Graduate Program, Rice University, Houston, TX, 77030, United States
[5]Lineberger Comprehensive Cancer Center and Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, 27599, United States
[6]Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200080, China
[7]Department of Chemistry, Rice University, Houston, TX, 77030, United States
[8]Department of Pharmacology and Cancer Biology and Duke Cancer Institute, Duke University, Durham, NC, 27710, United States
*To whom correspondence should be addressed. Email: isaac.hilton@rice.edu

Abstract

Protein intrinsically disordered regions (IDRs) are critical gene-regulatory components and aberrant fusions between IDRs and DNA-binding/chromatin-associating domains cause diverse human cancers. Despite this importance, how IDRs influence gene expression, and how aberrant IDR fusion proteins provoke oncogenesis, remains incompletely understood. Here we develop a series of synthetic dCas9-IDR fusions to establish that locus-specific recruitment of IDRs can be sufficient to stimulate endogenous gene expression. Using dCas9 fused to the paradigmatic leukemogenic NUP98 IDR, we also demonstrate that IDRs can activate transcription via localized biomolecular condensation and in a manner that is dependent upon overall IDR concentration, local binding density, and amino acid composition. To better clarify the oncogenic role of IDRs, we construct clinically observed NUP98 IDR fusions and show that, while generally non-overlapping, oncogenic NUP98-IDR fusions convergently drive a core leukemogenic gene expression program in donor-derived human hematopoietic stem cells. Interestingly, we find that this leukemic program arises through differing mechanistic routes based upon IDR fusion partner; either distributed intragenic binding and intrachromosomal looping, or dense binding at promoters. Altogether, our studies clarify the gene-regulatory roles of IDRs and, for the NUP98 IDR, connect this capacity to pathological cellular programs, creating potential opportunities for generalized and mechanistically tailored therapies.

Graphical abstract



Received: August 24, 2024. **Revised:** January 2, 2025. **Editorial Decision:** January 7, 2025. **Accepted:** January 22, 2025

Introduction

More than 40% of the human proteome contains intrinsically disordered regions (IDRs) that promote cooperative interactions that in turn facilitate a variety of critical cellular processes [1, 2]. Transcriptional regulators are highly enriched in IDRs, which is thought to enable interactions with, and/or recruitment of, other cofactors to precisely control chromatin dynamics and gene expression [3, 4]. Despite this prevalence and fundamental importance, the functions of IDRs remain incompletely defined. In addition, the IDRs from many different classes of human proteins, ranging from RNA/DNA binding (e.g. EWSR1 and NONO), chromatin remodeling (e.g. SS18), and nuclear membrane-associated (e.g. NUP98), can inappropriately fuse with diverse DNA-binding or chromatin-associating domains [5–9]. These fusion events often drive aberrant cellular processes, including protein mislocalization [10], dysregulated gene expression [11, 12], inappropriate chromatin dynamics [13], and oncogenesis [7, 11]. Therefore, deciphering how IDRs function holds tremendous promise for both understanding cellular processes and uncovering new therapeutic opportunities for diseases driven by pathological IDR fusion proteins, particularly IDR-driven human cancers.

For example, the NUP98 IDR, which is made up of aromatic phenylalanine-glycine (FG) repeats, is associated with hematopoietic malignancies when aberrantly fused to protein partners that engage the human genome [5]. At least 30 unique NUP98 IDR fusion partners have been identified in cancer patients [14, 15], and these fusion partners encompass a wide array of direct DNA binding domains (i.e. HOX and PMX) as well as epigenetic regulatory domains such as homeodomains, PHD, and SET domains [14, 15]. Additionally, NUP98 IDRs of different lengths, harboring as few as 19 to as many as 39 FG repeats, can fuse with this array of different binding partners and have been observed in human cancers [14, 16–18]. This multilayered diversity among oncogenic NUP98 fusions has complicated our understanding of how the NUP98 IDR impacts gene expression and raised the question as to whether heterogenous NUP98 IDR fusions mediate discrete pathological gene expression profiles or drive oncogenesis through conserved mechanistic routes.

Here, we engineered a set of CRISPR/Cas-based fusion proteins which demonstrate that human IDRs can be sufficient to initiate endogenous human gene expression in a phase separation-associated manner. This transactivation capability is dependent upon IDR concentration, binding site frequency, and IDR composition. We also generate genome-wide evidence that representatives from two major classes of oncogenic NUP98 IDR fusion proteins; DNA-binding (NUP98-HOXA9) and chromatin-associating (NUP98-KDM5A), activate a conserved core set of leukemogenic factors in primary donor-derived human hemopoietic stem cells (HSCs), however, these fusion proteins exhibit non-overlapping genomic binding landscapes and differing modes of transcriptional activation.

Collectively, our site-specific data gathered using synthetic dCas9-IDRs and genome-scale datasets generated using NUP98-KDM5A or NUP98-HOXA9 fusions connect the transcriptional activity of IDRs to the promotion of oncogenesis in donor derived HSCs. Given that NUP98 IDR-associated cancers are clinically intractable and linked with poor prognoses, our findings may provide new opportunities for therapeutic intervention. Further, since numerous human cancers can arise from IDR fusions with DNA-binding/chromatin-associating domains [7, 19], our findings here using the paradigmatic NUP98 IDR are likely extensible to other cancers driven by IDR fusions. Finally, since IDRs are a prevalent molecular component within human gene-regulatory factors [3, 20], understanding their functions could lead to enhanced synthetic control of phase-separation, chromatin dynamics, and gene expression [21–23].

Materials and methods

Cell culture

HEK293T cells (ATCC, CRL-11268) and U2OS P21-MS2 cells [24] were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, 31-053-028) supplemented with 10% FBS (Sigma, F2442) and 1% penicillin/streptomycin (Gibco, 15140). CD34^{+} human HSCs (StemCell, 70060) were cultured in StemSpan Leukemic Cell Culture Kit (StemCell, 09720) and Complete MethoCult Optimum media (StemCell, H4034). All cell lines were maintained at 37 °C with 5% CO2.

Plasmid construction

For dCas9-encoding vectors, the cloning backbone was modified from pLV-dCas9-p300-P2A-Puro (Addgene, 83889) by inserting an SV40 NLS peptide at the N-terminal of dCas9-p300). EWSR1, SS18, NONO1, and NONO2 are amplified from complementary DNA (cDNA) of HEK293T cells. NUP98a, NUP98b, eNUP98, and various FG repeats were amplified from the pPEP-TEV-cNup98 (amino acids 1–498; Addgene, 38037). HOXA9 (amino acids 164–272) was amplified from pFUW-tetO-HOXA9 (Addgene, 139827) and KDM5A (amino acids 1485–1690) was partially (amino acids 1485–1622) amplified from pBabe-puro/HA-FLAG-RBP2 (Addgene, 14802) and partially (amino acids 1623–1690) synthesized as a gBlock by IDT. The NUP98 F/S mutant was partly amplified from the IDRFS vector [13] and partially synthesized as a gBlock by IDT. EGFP was amplified from miniCMV-EGFP-hPGK-BSD (Addgene,188519). The dCas9-NUP98a, dCas9-NUP98b, dCas9-eNUP98, dCas9-FG, dCas9-F/S, dCas9-NUP98-HOXA9, and dCas9-NUP98-KDM5A plasmids (and associated EGFP harboring variants) were created by subcloning the corresponding amplified fragments into a BamHI-digested plasmid (modified from Addgene, 83889) backbone via NEBuilder HiFi DNA Assembly (NEB, E2621). The doxycycline inducible TRE-dCas9-NUP98a was created by subcloning the NUP98a fragments into a BamHI digested TRE-dCas9-p300 plasmid (unpublished). For constructs without dCas9, the cloning backbone was modified from pLV-dCas9-p300-P2A-Puro (Addgene, 83889) by replacing the dCas9-p300-p2a-puro cassette with a hPGK-puro cassette using NEBuilder HiFi DNA Assembly (NEB, E2621). NUP98a-EGFP, NUP98b-EGFP, various FG domains-EGFP, NUP98a/b-HOXA9-EGFP, and NUP98a/b-KDM5A-EGFP were subcloned into XbaI/BamHI digested backbone by NEBuilder HiFi DNA Assembly (NEB, E2621). All amplified PCR products or gBlocks were cloned onto the N-terminus of dCas9 and dCas9-eNUP98 vectors by NEBuilder HiFi DNA Assembly (NEB, E2621). Protein sequences of all dCas9 constructs are shown in Supplementary Note 1. All gRNAs were cloned into the pSPgRNA backbone (Addgene, 47108) and all gRNA protospacer targets are listed in Supplementary Table S3.

Live cell imaging

4.0e5 HEK293T cells or 2e5 U2OS P21-MS2 cells were seeded in 35 mm glass-bottom petri dishes with 10 mm microwells (MatTek, P35G-1.5–10-C). In experiments with only one fluorophore and no dCas9/gRNA, 500 ng of indicated EGFP/mCherry containing vectors were used. In experiments using dCas9 and gRNAs, 375 ng of dCas9 and EGFP fusion encoding vector and 125 ng of gRNA vector were used. In experiments using two fluorophores (i.e. p300-mCherry and FG-EGFP), 250 ng of EGFP containing vector, and 250 ng of mCherry containing vector were used. Cells were transiently transfected using Lipofectamine 3000 (Invitrogen, L3000015). HEK293T cells were stained using NucBlue Live ReadyProbes Reagent (Hoechst 33342; Invitrogen, R37605) ∼20 or ∼6–20 h post-transfection. U2OS P21-MS2 cells were stained using NucBlue Live ReadyProbes Reagent (Hoechst 33342; Invitrogen, R37605) ∼10 h post-transfection. After incubation, cells were imaged using a Nikon ECLIPSE Ti2-E (541094). Fluorescence quantification was performed using NIS-Elements AR (v5.10.01).

Flow cytometry

EGFP fusions were transfected into HEK293T cells. Cells were harvested ∼20 h post-transfection and then washed with phosphate-buffered saline (PBS, Fisher, BP3994) and then resuspended in 1% bovine serum albumin (BSA, Fisher, BP9706-100). Cells were then analyzed using a Sony SA3800 flow cytometer. Data were analyzed using FlowJo software (v.10).

Reverse-transcription quantitative PCR

HEK293T cells were transiently transfected with respective dCas9 expression vector (375 ng) and single gRNA vectors or equimolar pooled gRNA expression vectors (125 ng) in 24-well plates using Lipofectamine 3000 (Invitrogen, L3000015). RNA was extracted from HEK293T cells ∼72 h post-transfection using the Qiagen RNeasy Plus mini kit (Qiagen, 74136). One microgram of RNA was used for cDNA synthesis using the iScript advanced cDNA synthesis kit (Bio-Rad, 1725038). Reverse-transcription quantitative PCR (RT-qPCR) was performed using Luna Universal qPCR Master Mix (NEB, M3003E) on a CFX96 Real-Time PCR Detection System with a C1000 Thermal Cycler (Bio-Rad, 1855195). Baselines were subtracted using the baseline subtraction curve fit analysis mode and thresholds were automatically calculated using the Bio-Rad CFX Manager software (version 2.1). Results are expressed as fold change above mock-transfected cells after normalization to *GAPDH* expression using the ΔΔCt method. The messenger RNA (mRNA) nonlinear curves were performed using the log (agonist) versus response function of GraphPad Prism version 10.0.0. All qPCR primers and conditions are listed in Supplementary Table S4.

Lentivirus production

HEK293T cells were seeded at ∼40% confluency in 10 cm plates 1 day prior to transfection. Cells were transfected with lentiviral production plasmids at ∼80%–90% confluency on the next day. For each transfection, 10 μg of plasmid containing the vector of interest, 10 μg of pMD2.G (Addgene, 12259), and 15 μg of psPAX2 (Addgene, 12260) were transfected using Lipofectamine 3000 (Invitrogen, L3000015). Six hours post-transfection the media was changed. The supernatant was harvested at ∼24 and ∼48 h post-transfection and filtered through a 0.45 μm PVDF filter (Millipore, SLGVM33RS). Viruses were concentrated using Lenti-X Concentrator (Takara, 631232), aliquoted, and stored at −80 °C. Lentiviral titers were measured by the Lenti-X qRT-PCR Titration Kit (Takara, 631232).

Lentiviral transduction

Twenty-four-well plates were treated with 20 μg/ml RetroNectin (Takara Bio, T100B) as per manufacturer's instruction. 2e6 $CD34^+$ human HSCs (StemCell, 70060) were then mixed with 2.0e9 lentiviral particles and Lentiboost-P (1:100 dilution, Siron Biotech), and were then seeded onto the RetroNectin-treated plates. After centrifugation at 800 g at 4°C for 90 min, plates were placed at 37°C overnight. The cell medium was changed, and cells were cultured for another ∼48 h. The transduced cells were imaged using a fluorescent microscope (Nikon ECLIPSE Ti2-E; 541094) and EGFP-positive cells were isolated using FACS cell sorting (Sony Biotechnology, MA900) and cultured in Complete MethoCult Optimum media (StemCell, H4034) for 2 weeks.

HEK293T cells were transduced with lentiviruses encoding EGFP, NUP98a-HOXA9-EGFP, NUP98a-KDM5A-EGFP, or TRE-dCas9-NUP98a in 6-well plates at an MOI of 1. Briefly, 1e6 cells in 2 ml of media supplemented with 8 μg/ml polybrene (Sigma, TR-1003-G) were added to each well. Then, 48 h post-transduction, cells were passaged and EGFP positive cells were isolated by FACS cell sorting (Sony Biotechnology, MA900) and cultured in complete DMEM media for 2 weeks with passaging every 3 days. TRE-dCas9-NUP98a transduced HEK293T cells were selected with DMEM medium supplemented with 1 μg/ml puromycin for 1 week. The selected cells were transfected with indicated gRNA pools and rtTA plasmids. Doxycycline (0–100 nM) was added ∼6 h post-transfection. The cells were visualized at ∼20 h post-transfection. Cells were collected at 72 h post-transfection for RT-qPCR and Western blotting analysis.

Western blotting

Twenty microgram of protein was loaded for SDS PAGE and transferred onto a PVDF membrane for Western blots. Primary antibodies (α-FLAG; Sigma–Aldrich, F1804) were used at a 1:1000 dilution in 1× Tris Buffered Saline with 1% Casein (Bio-Rad, 1610782EDU). Secondary α-mouse HRP (Cell Signaling, 7076) was used at a 1:3000 dilution in 1× Tris Buffered Saline with 1% Casein (Bio-Rad, 1610782EDU). Membranes were exposed after addition of ECL (Bio-Rad, 170-5060). Tubulin was detected with hFAB™ Rhodamine Anti-Tubulin Primary Antibody (Bio-Rad, 12004166).

CUT&RUN assays

CUT&RUN assays were carried out using the Epicypher CUTANA ChIP/CUT&RUN Kit (Epicypher 14-1048). Briefly, 5.0e5 transfected cells were harvested, immobilized on concanavalin A beads, and permeabilized in 0.01% digitonin cell permeabilization buffer. The cell-bead conjugate mixture was then divided equally into 2 aliquots and incubated in 50 μl antibody buffer with either 0.5 μg of anti-H3K27ac antibody (Abcam, 4729) or 0.5 μg of control rabbit IgG

antibody (Abcam, 37415) overnight at 4°C. After washing the beads, pAG-MNase was added to the immobilized cells and the solution was incubated for 2 h at 4°C to digest and release chromatin DNA. For CUT&RUN-qPCR assays, 1 μl of purified DNA from both H3K27ac antibody treatment and rabbit IgG control treatment was then assayed using qPCR. Relative enrichment of H3K27ac is expressed as fold change relative to the corresponding rabbit IgG control samples using the ΔΔCt method. All qPCR primers and conditions are listed in Supplementary Table S4.

RNA-seq and analysis

RNA sequencing (RNA-seq) was performed in duplicate for each experimental condition. RNA was isolated from transduced CD34+ human hematopoietic stem cells (HSCs) using the RNeasy Plus mini kit (Qiagen, 74136). Total RNA samples were normalized to 10 ng each, based on picogreen quantitation. RNA-seq libraries were constructed using the Takara SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian (Takara, 634485). The Switching Mechanism at 5′ end of RNA template (SMARTer) cDNA synthesis technology generates directionally identifiable ds-cDNA, incorporated with Illumina sequencing adapters and sample barcodes. Following cDNA synthesis and purification, ribosomal cDNA reduction is achieved by using probes (R-Probes v3) that target mammalian ribosomal cDNA and human mitochondrial ribosomal cDNA and ZapR v3 cleaves these fragments. Following amplification of the library using universal primers and a final purification using Takara NucleoMag NGS Clean-up and Size Select beads, the final library is ready for quantification and sequencing. Read 1 sequenced reads map to the antisense strand of the original RNA. The resulting libraries were quantitated by picogreen and fragment size assessed with the Agilent 2100 Bioanalyzer.

All samples were pooled equimolarly and re-quantitated by qPCR using the Applied Biosystems ViiA7 Quantitative PCR instrument and a KAPA Library Quant Kit (p/n KK4824), and also re-assessed on the Bioanalyzer. Cluster generation by exclusion amplification (ExAMP): using the concentration from the ViiA7 TM qPCR machine above, 150 pM of equimolarly pooled library is loaded onto one lane of the NovaSeq S1 flowcell (Illumina p/n 20028317) following the XP Workflow protocol (Illumina kit p/n 20021664) and amplified by ExAMP onto a nanowell-designed, patterned flowcell using the Illumina NovaSeq 6000 sequencing instrument. A combination of PhiX Control v3 adapter-ligated library (Illumina p/n FC-110-3001) and RNA library are spiked-in at 10% by weight to ensure balanced diversity and to monitor clustering and sequencing performance. A paired-end 150 bp cycle run was used to sequence the flowcell on a NovaSeq 6000 Sequencing System. An average of 60 million read pairs per sample was sequenced. Fastq file generation was executed using Illumina's cloud-based informatics platform, BaseSpace Sequencing Hub. Reads were aligned to the GRCh38 transcriptome using HISAT2 (2.2.1) [25]. Transcript abundance was calculated using feature Counts from the subread package (v2.0.3) [26], and differential expression was determined in R studio (v2022.07.2, R 4.2.2) using the DESeq2 (v1.38.3) [27] analysis package with default parameters. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed using clusterProfiler (v4.6.0) [28] and sequence read coverage was visualized by the IGV (2.16.0) [29].

Colony formation assay

One thousand CD34+ human HSCs transduced with EGFP, NUP98a-HOXA9-EGFP, or NUP98a-KDM5A-EGFP. Cells were sorted for EGFP positivity 3 days post-transduction and then plated in Complete MethoCult Optimum media (StemCell, H4034). After 2 weeks of growth, colonies were counted, and cells were washed with PBS before being replated in Complete MethoCult Optimum media. After another 2 weeks of growth, colonies were counted again.

ChIP-sequencing and analysis

HEK293T cells were transfected with indicated NUP98 fusions (10 μg) or cotransfected with indicated dCas9 fusion (7.5 μg) expression vectors and gRNA (2.5 μg) constructs in 10 cm plates in biological duplicates for each condition tested. Approximately 72 h post-transfection, cells were cross-linked for 10 min at RT using 1% formaldehyde (Sigma F8775-25ML) and then the reaction was stopped by the addition of glycine to a final concentration of 125 mM. Cells were harvested and washed with ice cold 1× PBS and suspended in Farnham lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, 0.5% NP-40) supplemented with protease inhibitor (Thermo Scientific, A32965). Cells were then pelleted and resuspended in RIPA buffer (1× PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitor. Approximately 2.5e7 cells were used for each ChIP experiment. Chromatin in RIPA buffer was sheared to a median fragment size of around 250 bp using a ioruptor XL (Diagenode). Five microgram of α-FLAG antibody (Sigma–Aldrich, F1804) was incubated with 50 μl Mouse IgG magnetic beads (Life Technologies, 11202D) for ∼16 h at 4°C, respectively. Antibody-linked magnetic beads were washed three times with PBS/BSA buffer (1× PBS and 5 mg/ml BSA) and sheared chromatin was incubated with corresponding antibody-linked magnetic beads at 4°C overnight and then washed five times with LiCl IP wash buffer (100 mM Tris, pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate). Cross-links were then reversed via overnight incubation at 65°C and DNA was purified using QIAquick PCR purification kit (Qiagen, 28106) for ChIP-seq. Input DNA was prepared from ∼1.0e6 cells.

ChIP DNA samples were quantified using Qubit 2.0 Fluorometer (ThermoFisher Scientific) and the DNA integrity was checked with 4200 TapeStation (Agilent Technologies). ChIP-seq libraries were prepared using the NEB NextUltra DNA Library Preparation kit (NEB, 7645). Briefly, the ChIP DNA was end-repaired and adapters were ligated after adenylation of the 3′ ends. Adapter-ligated DNA was size-selected, followed by clean up, and limited cycle PCR enrichment. The ChIP library was validated using Agilent TapeStation (Agilent) and quantified using Qubit 2.0 Fluorometer as well as real time PCR (KAPA Biosystems). The sequencing libraries were multiplexed and clustered onto one lane of a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq 4000 equivalent instrument. Sequencing was performed using a 2 × 150 paired end conFiguration. Image analysis and base calling were conducted by the HiSeq Control Software. Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into Fastq files and de-multiplexed using the bcl2fastq (v2.17). One mismatch was allowed for index sequence identification. For ChIP-seq analysis, reads were aligned to the GRCh38 reference genome using Bowtie2 (v2.4.5) [30] and peaks were called using MACS2 (v2.2.7.1)

[31] with a q-value ≤ 0.01. Resulting peaks were assessed and analyzed with deepTools (v3.5.4) [32], DiffBind (v3.8.4) [33], and ChIPSeeker (v1.34.1) [34].

3C-qPCR

1.0e7 transduced HEK293T cells were crosslinked in 9.5 ml of 1% formaldehyde in PBS with 10% FBS (Sigma–Aldrich, F8775) and then incubated at room temperature for 10 min. 1.425 ml of ice-cold 1M glycine was added to a final concentration of 130 mM glycine to quench crosslinking similar to previous methods [35]. Briefly, crosslinked cells were then pelleted and then lysed in 5 ml of lysis buffer (10 mM Tris–HCl, pH 7.5; 10mM NaCl; 5 mM $MgCl_2$; 0.1 mM EGTA; 1 × complete protease inhibitor; Fisher, A32965) on ice for 10 min, and then centrifuged for 5 min at 400 g at 4°C for nuclei extraction. The extracted nuclei were transferred to 1 × CutSmart digestion buffer (NEB, B7204) with 0.3% SDS (Invitrogen 24730020) and then shaken at 900 rpm for 1 h at 37°C, after which, 2% Triton X-100 (Sigma–Aldrich, T9284) was added, and another 1 h incubation with shaking at 900 rpm at 37°C was performed. 400 U of HindIII (NEB R0104) was then added, and the nuclear DNA was digested at 37°C while shaking at 900 rpm overnight. The solution was then brought to 7 ml total volume in T4 ligation buffer (NEB B0202) with a final concentration of 1.6% SDS and 1% Triton X-100. The mixture was then gently shaken at 37°C for 1 h. One hundred Weiss units of T4 DNA ligase (NEB M0202) was then added to the nuclear DNA, and the solution was ligated for 4 h at 16°C and then for 30 min at room temperature. The reaction was then incubated with 300 μg of proteinase K (Qiagen 19131) at 65°C overnight. The next day 300 μg of RNase (Qiagen 9101) was added and the solution was incubated at 37°C for 45 min. The nuclear DNA was then purified using phenol–chloroform (Fisher BP1752I400) extraction. One thousand nanogram of extracted DNA was used for qPCR. Results are expressed as relative contact normalized to the non-transfected HEK293T cells. All qPCR primers and conditions are listed in Supplementary Table S4.

Results

IDRs from diverse human proteins are sufficient to stimulate endogenous gene expression when localized to endogenous promoters

The IDRs from EWSR1, SS18, NONO, and NUP98 can aberrantly fuse with DNA-binding/chromatin-associating domains, which disrupts transcription and promotes oncogenesis [6, 7, 11, 12]. To clarify the phase separation capacity and potential oncogenic role of these IDRs, we first fused EGFP to the C-terminus of the EWSR1, SS18, NONO1 (NONO variant 1), NONO2 (NONO variant 2), NUP98a (longer IDR isoform of NUP98), and NUP98b (shorter IDR isoform of NUP98) IDRs (Supplementary Fig. S1A and B) and then monitored the expression levels and cellular distributions of these fusions when transfected into human HEK293T cells. IDR-EGFP fusions exhibited varied expression levels, and each IDR tested facilitated increases in the nuclear fraction of EGFP compared to EGFP controls (Fig. 1A, top and Supplementary Fig. S1C and D). In addition, each IDR-EGFP fusion formed biomolecular condensates in HEK293T cells (Fig. 1A, top and Fig. 1B). Although the expression level of the NUP98a-EGFP fusion was lower than other IDR fusion proteins, NUP98a-EGFP nevertheless exhibited particularly sharp, clustered EGFP puncta in transfected cell nuclei (Supplementary Fig. S1C and E, Fig. 1A, top and Fig. 1B). Notably, adding a nuclear localization sequence (NLS) to each IDR fusion further increased nuclear import of all IDR fusions and phase separation behaviors of NONO1, NONO2, and NUP98b (Fig. 1A, bottom; Fig. 1B and Supplementary Fig. S1B–E), suggesting that higher nuclear protein concentrations of IDRs bolster these activities.

To quantify the gene-regulatory role of IDRs at endogenous human loci, we next fused each of these IDRs to the C-terminus of dCas9 (Fig. 1C) and targeted these fusion proteins to the *IL1RN* promoter using four specific gRNAs. All dCas9-IDR fusions significantly activated *IL1RN* transcription (Fig. 1D), albeit this activation was relatively modest compared to other conventional CRISPRa tools {e.g. dCas9-p300 [36] or dCas9-VPR [37]}. Importantly, transfection of these IDRs alone had no effect upon *IL1RN* expression (Supplementary Fig. S1F), demonstrating that dCas9-mediated targeting of these IDRs is required for transcriptional activation of *IL1RN*. Altogether, these data demonstrate that human IDRs can undergo different degrees of nuclear import and spontaneous condensation in a concentration-related (i.e. enhanced via NLS) manner in human cells, and further, that human IDRs can be sufficient to initiate gene expression upon targeted binding within endogenous human promoters.

Biomolecular condensation of the NUP98 IDR at human promoters drives gene expression

Given the condensation behavior of IDRs (Fig. 1), we hypothesized that IDRs functioned via the formation of so called transcriptional condensates [38, 39], in which protein concentration and binding density tightly govern gene regulatory control [40–42]. To test this hypothesis, we constructed a doxycycline (dox) inducible dCas9-NUP98a fusion and delivered this fusion protein into HEK293T cells using lentiviral transduction (MOI = 1.0). We measured the fraction of dCas9-NUP98a-EGFP positive cells and dCas9-NUP98a-EGFP protein levels using a range of dox concentrations (0–100 nM) using flow cytometry and observed that even though the number of dCas9-NUP98a-EGFP positive cells saturated at 20 nM dox, the dCas9-NUP98a-EGFP protein levels continued to increase with higher dox concentrations (>20 nM; Supplementary Fig. S2A–C). Further, we found that dCas9-NUP98a-EGFP condensed into puncta and activated gRNA-targeted promoters at very low dox concentrations (5 nM), which became more pronounced at higher dox concentrations (Fig. 2A–C). We also observed that increasing the binding density/local dCas9-NUP98a concentration by modulating the gRNA target site frequency at promoters resulted in dose-responsive increases in gene expression (Fig. 2D) and that there was a nonlinear relationship between the concentration of dCas9-NUP98a and transcriptional activation (Fig. 2C and D, Supplementary Fig. S2D and E). Interestingly, this activation profile was similar to dCas9 fused to the acidic transactivation domain VP64 (Supplementary Fig. S2F and G) but distinct from other enzymatic epigenome editing tools (i.e. dCas9-p300), which have been shown to activate gene expression more linearly [36, 43]. For instance, the dCas9-p300 epigenome editor, which is capable of direct histone

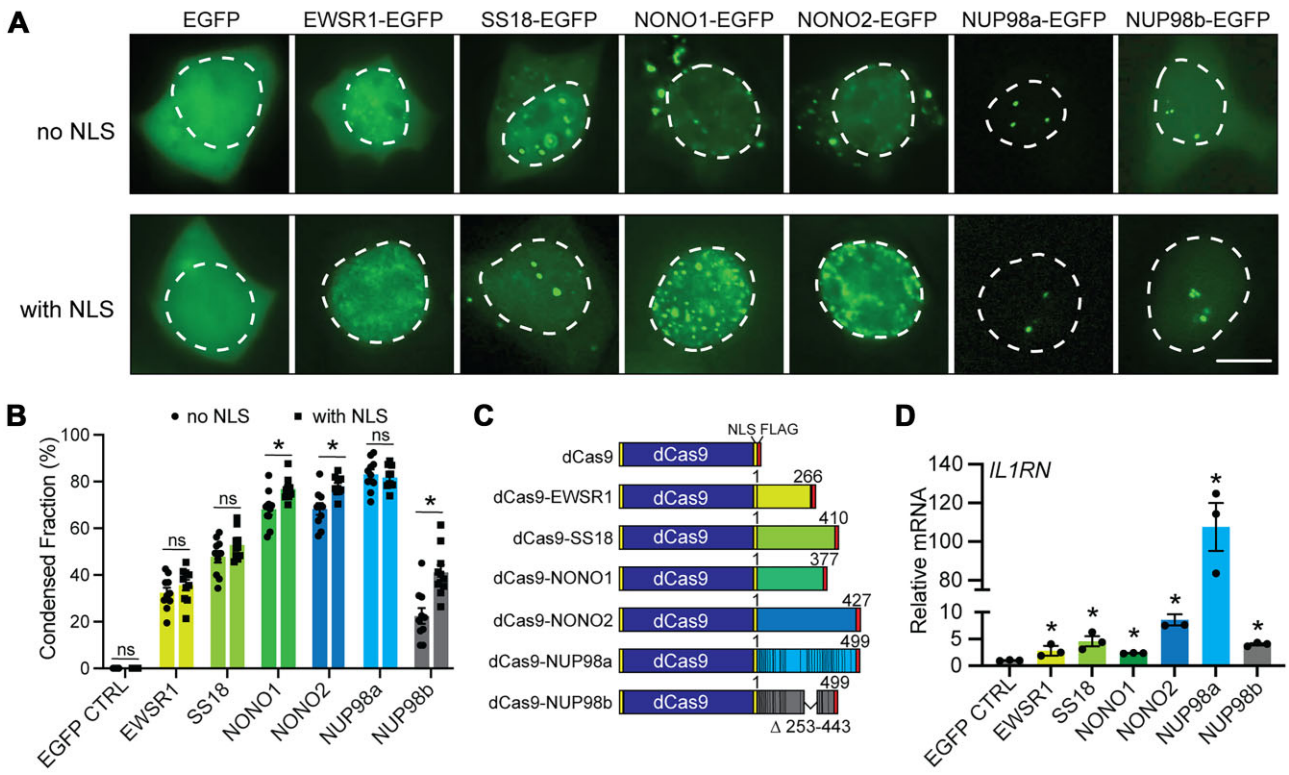


Figure 1. IDR domains are sufficient to stimulate human gene promoters. (**A**) Expression and cellular localization of EGFP-tagged IDRs from EWSR1, SS18, NONO1, NONO2, NUP98a, and NUP98b in HEK293T cells ∼20 h post-transfection. Top: IDR fusions without NLS and bottom: IDR fusions with NLS. Dotted lines encircle nuclei and scale bar = 10 μm. (**B**) Quantification of condensed EGFP fractions in HEK293T cells ∼20 h post-transfection of EGFP tagged IDR domains from EWSR1, SS18, NONO1, NONO2, NUP98a, and NUP98b with or without NLS. $n = 10$ randomly selected cells; error bars, s.e.m. (**C**) Schematics of dCas9, dCas9-EWSR1, dCas9-SS18, dCas9-NONO1, dCas9-NONO2, dCas9-NUP98a, and dCas9-NUP98b. (**D**) RT-qPCR for *IL1RN* mRNA levels ∼72 h post-transfection of dCas9 or indicated dCas9-IDR fusions and 4 corresponding *IL1RN* promoter-targeting gRNAs. mRNA levels were normalized to non-transfected control cells. Asterisks (*) indicate $P < 0.05$ via two-tailed student's *t*-test. $n = 3$ independent experiments; error bars; s.e.m.

acetylation (and lacks natural p300 IDRs found in full length p300), exhibited a linear relationship between the number of gRNA target sites and gene expression (Supplementary Fig. S2H and I).

We also observed that constitutively expressed dCas9-NUP98a-EGFP formed clear nuclear puncta when transiently transfected into HEK293T cells (Fig. 2E, top; Supplementary Fig. S2J, top). In contrast, dCas9-p300-EGFP, did not form puncta under the same conditions. Moreover, puncta catalyzed by dCas9-NUP98a-EGFP were highly sensitive to 10% 1,6-hexanediol (Fig. 2E, bottom; Supplementary Fig. S2J, bottom), which is a commonly used reagent to interrogate phase separation behavior [44]. To further investigate the link between phase separation behavior and dCas9-NUP98a-mediated gene activation, we treated cells with low dose (0.4%) 1,6-hexanediol at 6 h post-transfection. Even at this relatively low concentration, the ability of dCas9-NUP98a to activate gene expression from human promoters was significantly diminished (Fig. 2F). However, similar treatment did not significantly alter the ability of dCas9-p300 to activate gene expression (Fig. 2G). To next determine whether dCas9-NUP98a condensates were physically localized with gRNA target sites and linked with transcription, we transfected U2OS p21-MS2 + MCP-EGFP cells [24] with dCas9-NUP98a-mCherry and *p21* promoter-targeting gRNAs and observed that indeed, dCas9-NUP98a-mCherry condensates were co-localized with *p21* transcripts (Fig. 2H and Supplementary Fig. S2K). Altogether, these data demonstrate that biomolecular condensation is an intrinsic property of NUP98a in human cells that is required for stimulating gene expression from promoters.

FG repeat numbers within the NUP98 IDR dictate phase separation and transactivation intensities

The NUP98 IDR is made up of FG repeats, and different numbers of FG repeats (from 19 to 39 FG repeats) are detected in NUP98-associated cancers [14, 17, 18]. This motivated us to next test how different FG repeat numbers control NUP98 fusion protein properties. We constructed a series of EGFP-tagged NUP98 FG repeats (spanning 0, 11, 21, 30, and 39 repeats) and associated EGFP-tagged dCas9-FG repeat fusions and measured their expression levels and the ability of these fusion proteins to form puncta in transfected cell nuclei (Supplementary Fig. S3A and B). Interestingly, expression levels were inversely correlated with FG repeat numbers, however nuclear localization and condensation behaviors increased in proportion to increasing FG repeats (Fig. 3A, top; Fig. 3B, left; Supplementary Fig. S3C and E, top; Supplementary Fig. S3D). Although puncta were observed in HEK293T cell nuclei harboring as few as 21 NUP98 FG repeats, a marked increase in numbers of puncta was observed above 30 NUP98 FG repeats (Fig. 3A, top; Supplementary Fig. S3F, top). Similar results were obtained when this series of FG

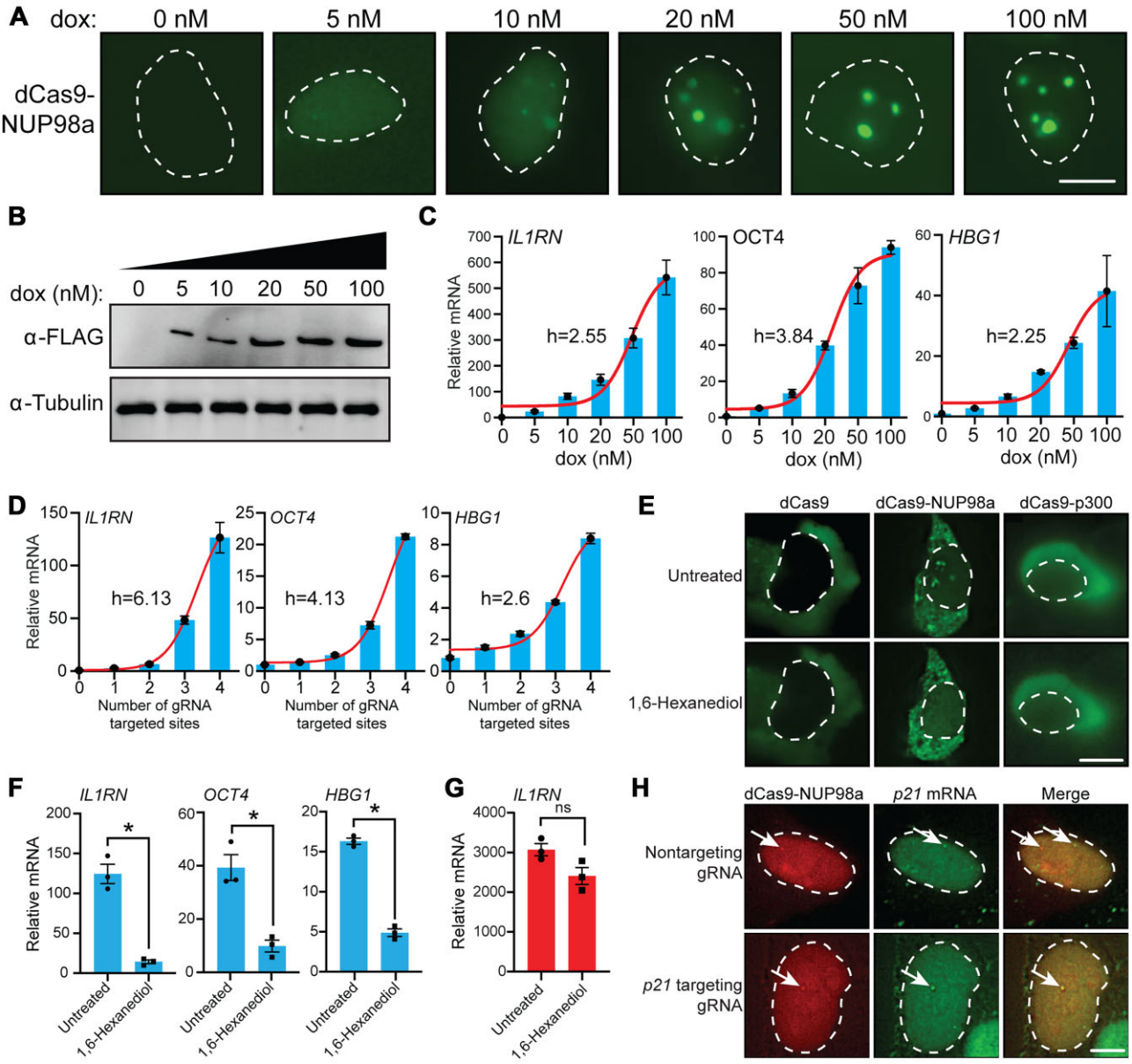


Figure 2. Biomolecular condensation of the NUP98 IDR at human promoters stimulates gene expression. (**A**) Expression and localization of dCas9-NUP98a in HEK293T cells transduced with TRE-dCas9-NUP98a lentivirus over indicated ranges of doxycycline (dox) for ∼20 h. Dotted lines encircle nuclei and scale bar = 10 μm. (**B**) Expression of dCas9-NUP98a in HEK293T cells transduced with TRE-dCas9-NUP98a lentivirus over indicated ranges of dox for ∼48 h. (**C**) RT-qPCR for *IL1RN* (left), *OCT4* (middle), or *HBG1* (right) in HEK293T cells transduced with TRE-dCas9-NUP98a lentivirus over indicated ranges of dox ∼72 h post-transfection of corresponding gRNAs. Dox was added to cells 6 h post-transfection of gRNAs. The dose-response curve across dox concentrations is indicated. mRNA levels were normalized to cells transfected with a non-targeting gRNA. (**D**) mRNA levels of *IL1RN*, *OCT4*, or *HBG1* were measured by RT-qPCR ∼72 h after co-transfection of dCas9-NUP98a and increasing numbers of promoter-targeting gRNA (s). The dose-response curve across gRNA-targeted sites is indicated. (**E**) Expression and localization of dCas9-EGFP, dCas9-NUP98a-EGFP, or dCas9-p300-EGFP ∼8 h post-transfection in HEK293T cells before and 1 min after treatment with 10% 1,6-hexanediol. Scale bar = 10 μm. (**F**) RT-qPCR of *IL1RN* (left), *OCT4* (middle), or *HBG1* (right) mRNA ∼72 h post-transfection of dCas9-NUP98a fusions and corresponding gRNAs in the absence or presence of 0.4% 1,6-hexanediol. (**G**) RT-qPCR of *IL1RN* mRNA ∼72 h post-transfection of dCas9-p300 and 4 corresponding gRNAs in the absence or presence of 0.4% 1,6-hexanediol. (**H**) Expression and localization of dCas9-NUP98a-mCherry in U2OS p21-MS2 + MCP-EGFP cells ∼10 h post-transfection of *p21* targeting or nontargeting control gRNAs. Dotted lines encircle nuclei, arrows indicate NUP98 condensation (mCherry), *p21* mRNA (EGFP), or colocalization and scale bar = 10 μm. For panels (**C**) and (**D**), Hill coefficients (H) were determined using nonlinear regression in GraphPad Prism 10. For panels (D), (F), and (G), mRNA levels were normalized to non-transfected control cells. Asterisks (*) indicate $P < 0.05$ via two-tailed student's *t*-test. $n = 3$ independent experiments; error bars; s.e.m. ns; not significant.

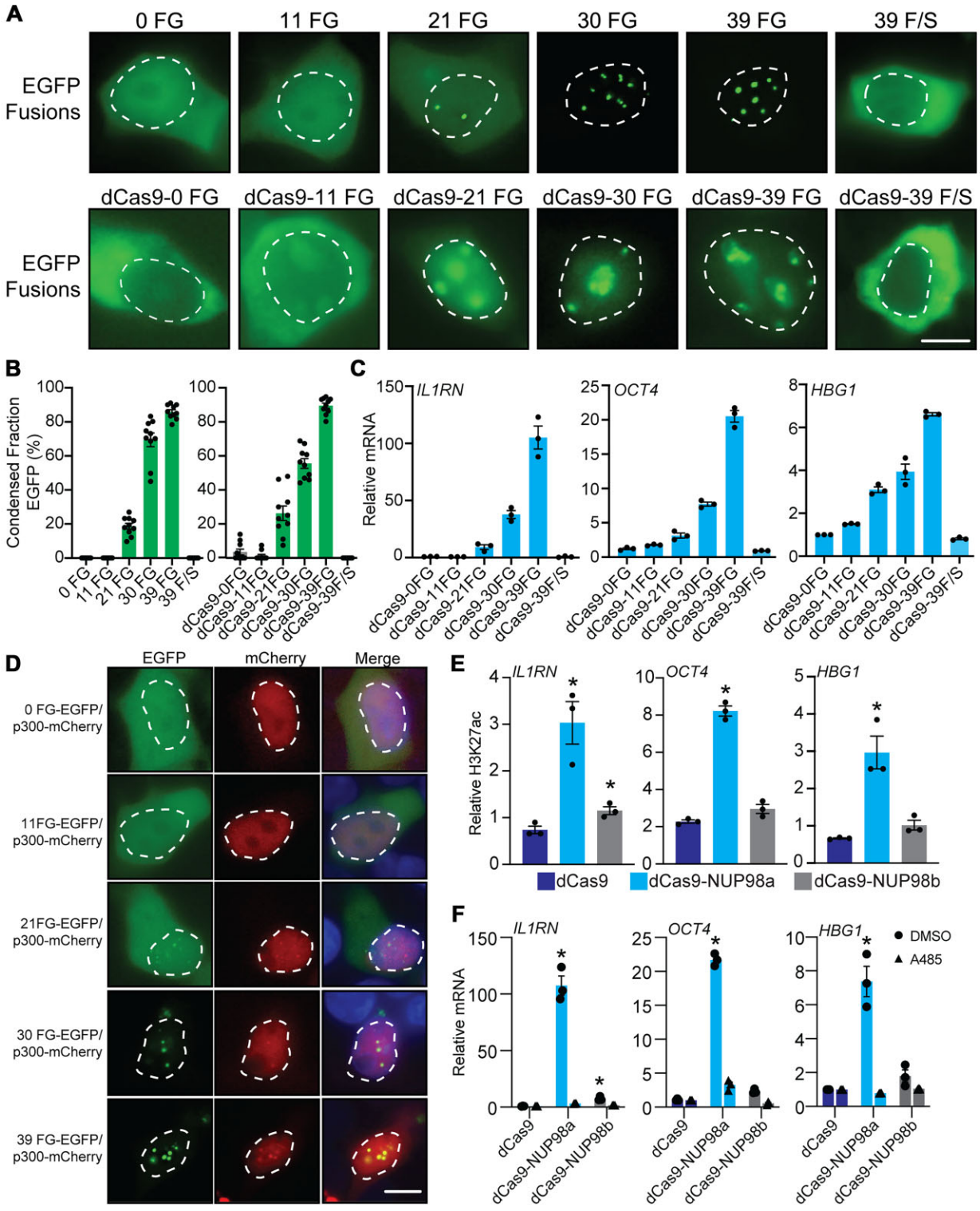


Figure 3. NUP98 IDR FG repeat numbers dictate phase separation and transactivation intensities. (**A**) Expression and cellular localization of EGFP-tagged NUP98 FG or phenylalanine to serine (F/S) mutant and dCas9-NUP98 FG (or F/S mutant) repeat variants (top and bottom, respectively) in HEK293T cells ∼20 h post-transfection. Dotted lines encircle nuclei and scale bar = 10 μm. (**B**) Quantification of condensed EGFP fractions in HEK293T cells ∼20 h post-transfection of EGFP tagged NUP98 FG (or F/S mutant, left), and dCas9-NUP98 FG (or F/S mutant, right) repeat variants. $n = 10$ randomly selected cells; error bars, s.e.m. (**C**) RT-qPCR for *IL1RN* (left), *OCT4* (middle), or *HBG1* (right) mRNA level ∼72 h post-transfection of dCas9-EGFP or indicated dCas9-FG repeat variants and corresponding gRNAs. mRNA levels were normalized to non-transfected control cells. (**D**) Co-expression of NUP98 FG-EGFP fusion variants and p300-mCherry. Dotted lines encircle nuclei and scale bar = 10 μm. (**E**) Relative enrichment of H3K27ac at *IL1RN* (left), *OCT4* (middle), and *HBG1* (right) promoters ∼72 h post-transfection with dCas9 or indicated dCas9-NUP98 fusions and corresponding gRNAs. H3K27ac levels were normalized to IgG control samples. (**F**) RT-qPCR for *IL1RN* (left), *OCT4* (middle), and *HBG1* (right) mRNA levels ∼72 h post-transfection of dCas9 or indicated dCas9-NUP98 fusion proteins and corresponding gRNAs in the presence of DMSO or the p300 inhibitor A485. 20 μM of A485 or an equal volume of DMSO was added to cells ∼6 h post-transfection. mRNA levels were normalized to non-transfected control cells. For panels C, E, and F, asterisks (*) indicate $P < 0.05$ via two-tailed student's *t*-test. $n = 3$ independent experiments; error bars; s.e.m.

repeats and EGFP was fused to dCas9, despite puncta being more diffuse in the context of dCas9-FG repeat fusions. (Fig. 3A, bottom; Fig. 3B, right; Supplementary Fig. S3C and E–F, bottom). These effects had no significant impact on cell nuclei sizes (Supplementary Fig. S3G).

Consistent with this continuum of phase separation behavior, although 21 NUP98 FG repeats activated gene expression when localized to human promoters, potency was enhanced upon localization of 30 or 39 NUP98 FG repeats (Fig. 3C and Supplementary Fig. S3H). Thus, these data connect the critical threshold of NUP98 FG repeats observed in oncogenic NUP98 fusions to gene regulatory properties. Importantly, mutation of phenylalanine to serine (F–S) completely abolished nuclear import, puncta formation, and gene activation capabilities (Fig. 3A–C and Supplementary Fig. S3A) and restoration of nuclear localization to the 39 F/S repeat mutant through the addition of NLSs failed to rescue condensation capacity, demonstrating that NUP98 IDR composition (i.e. the FG repeats) are key factors driving phase separation and gene activation (Fig. 3A–C and Supplementary Fig. S3I).

Immunoprecipitation [45] and Bio-ID [13] have also shown that NUP98 FG repeats can interact with the transcriptional cofactor p300. Therefore, we hypothesized that NUP98 FG repeats might activate gene expression, at least in part, though interaction with p300 and changes to acetylation status at occupied sites. To test this hypothesis, we co-expressed p300 fused to mCherry and various FG repeats in HEK293T cells and observed that p300 and NUP98 FG-mediated condensates co-localized at ≥21 FG repeats (Fig. 3D and Supplementary Fig. S3J). Notably, this is also the minimum FG repeat number that showed biomolecular condensation and gene activation (Fig. 3C). Consistent with this co-condensation, we also observed that targeting dCas9-NUP98a, or the shorter NUP98b isoform (21 FG repeats) to human promoters resulted in increased H3K27ac levels (Fig. 3E) and further, that treatment of cells with the p300 inhibitor A485, abolished the ability of dCas9-NUP98 fusions to activate gene expression from targeted promoters (Fig. 3F). Altogether these data demonstrate that co-condensation of NUP98 FG repeats and other transcriptional co-factors, such as p300, and associated epigenetic modifications (such as H3K27ac) play an essential role in NUP98 FG repeat-mediated gene activation, and potentially oncogenesis.

Different NUP98 fusions provoke a shared leukemogenic gene expression program in primary human HSCs

Having demonstrated that IDRs can stimulate endogenous human gene expression upon site-specific binding to promoters using dCas9 (Figs 1–3), we next sought to determine how IDR fusions observed in human pathologies disrupt transcription in clinically proximal cell types. Therefore, we selected paradigms from two major classes of NUP98 IDR fusions that have been frequently observed in leukemia patients NUP98-HOXA9 (DNA-binding) and NUP98-KDM5A (chromatin-associating) [14, 46–48] for genome-scale analysis in primary human HSCs. We first constructed NUP98-HOXA9 and KDM5A fusions harboring the most common oncogenic NUP98 IDR fragment (NUP98a; 39 FG repeats). We also fused EGFP to the C-terminus of each fusion for simple and robust analysis. NUP98a-HOXA9-EGFP and NUP98a-KDM5A-EGFP (hereafter NUP98a-HOXA9 and NUP98a-KDM5A) strongly localized to the nuclei of primary mobilized donor-derived human peripheral blood CD34 + HSCs and formed clear nuclear puncta after lentiviral transduction (Fig. 4A and Supplementary Fig. S4A). RNA-seq analysis on EGFP positive HSCs showed that NUP98a-HOXA9 and NUP98a-KDM5A displayed distinct transcriptomes relative to one another and to EGFP control transduced cells (Fig. 4B). In fact, only ∼17% (410 out of 2434) of all differentially expressed genes (DEGs) overlapped between HSCs transduced with NUP98a-HOXA9 and NUP98a-KDM5A (Fig. 4C and Supplementary Table S1).

Further comparative RNA-seq analyses revealed that despite this largely non-overlapping transcriptomic dysregulation, both NUP98-HOXA9 and NUP98-KDM5A caused substantial perturbations in the KEGG [49] transcriptional misregulation in cancer pathway (Fig. 4D and E). In particular, each fusion markedly upregulated the expression of the HOXA9/MEIS1 transcriptional program [50] as well as several other leukemia-associated transcription factors {e.g. *RUNX2* [51], *WT1* [52], *ZBTB16* [53], and *MLLT3* [54]} that together stimulate the proliferation and aberrant transcription of leukemia cells. A key cytokine in HSC lineage commitment, *CSF1* [55], was also downregulated in HSCs transduced with either NUP98a-HOXA9 and NUP98a-KDM5A (Fig. 4D and F), consistent with emerging evidence that NUP98 IDR fusions can disrupt human HSC differentiation programs [56]. Notably, we observed important exceptions to this shared core leukemogenic factors between NUP98a-HOXA9 and NUP98a-KDM5A, in that only NUP98a-HOXA9 triggered the expression of additional oncogenes {e.g. *PBX3* [57], *SUPT3H* [58], and *FUT8* [59], Fig. 4D}, which likely act synergistically to facilitate leukemic transformation in NUP98-HOXA9-associated cancers. Importantly, we also observed that NUP98a-HOXA9 and NUP98a-KDM5A also activated leukemogenic factors (*MEIS1*, *MLLT3*, *PBX3*, *SUPT3H*, and *FUT8*) in HEK293T cells (Supplementary Fig. S4I).

In rare cases, shorter NUP98 FG repeats (<30 FG repeats) have also been detected in NUP98-associated leukemias [17, 18]. Therefore, we also tested how shorter isoforms of NUP98 (i.e. NUP98b; 21 FG repeats) affected transcriptional programs in human HSCs (Supplementary Fig. S4). We found that NUP98b fusions also potently localized to HSC nuclei, stimulated the transcriptional misregulation in cancer KEGG pathway and promoted the proliferation of human HSCs; albeit to a lesser extent than longer NUP98a (39 FG repeats) IDR fusions (Supplementary Fig. S4C–H and J). These data demonstrate that NUP98 IDR fusions with different C-terminal fusion partners and FG repeat lengths have the capacity to promote leukemogenic gene expression in human cells, including primary human HSCs.

NUP98a-HOXA9 and NUP98a-KDM5A display vastly different patterns of engagement with the human genome

To further clarify the global effects of the NUP98 IDR in human cells, we investigated how oncogenic fusion partners altered the cellular localization and genomic binding profiles of the NUP98 IDR. To do so, we transfected NUP98a-EGFP, NUP98a-KDM5A, or NUP98a-HOXA9 into HEK293T cells, and then monitored fusion protein expression and measured

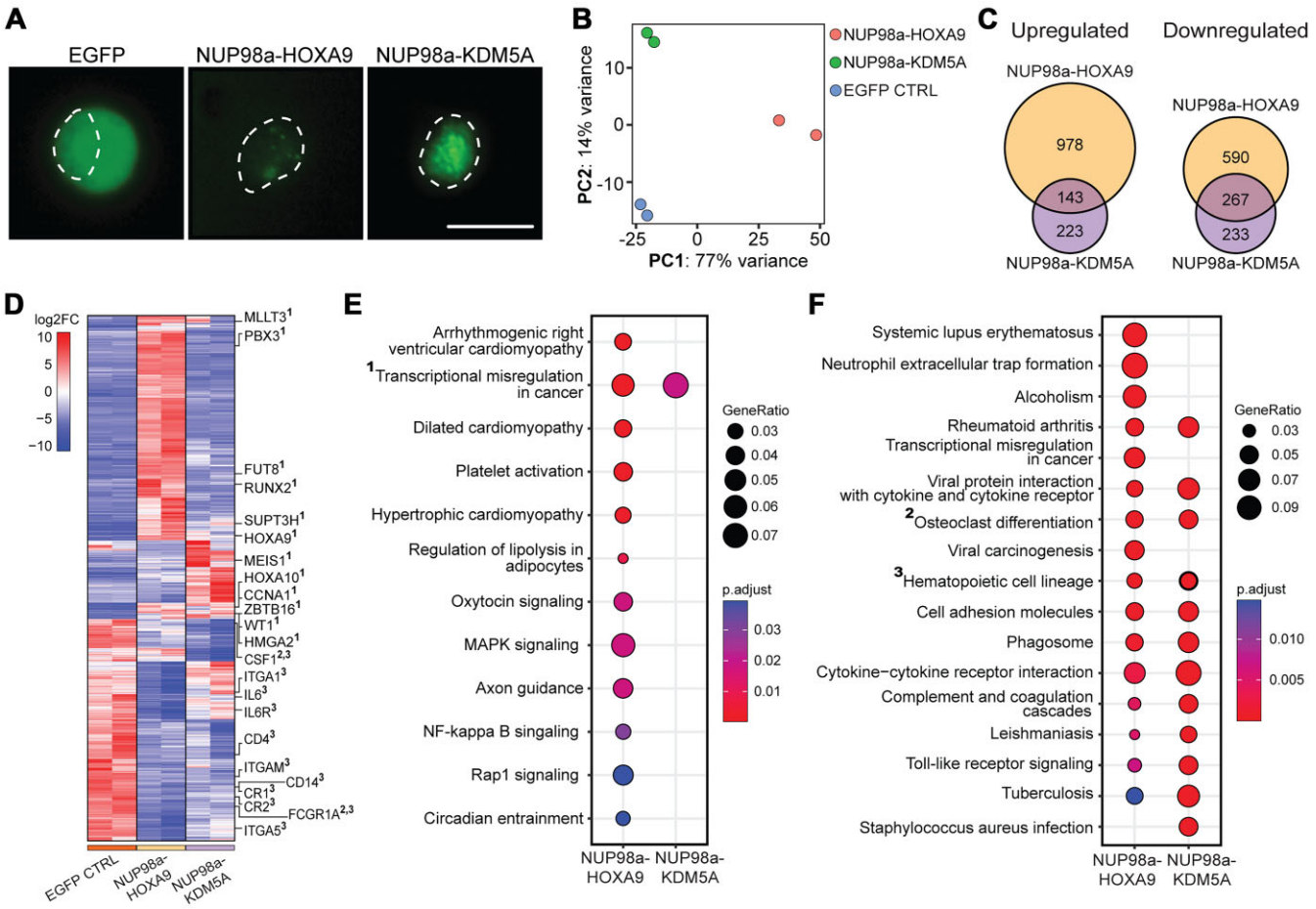


Figure 4. NUP98-HOXA9 and NUP98-KDM5A drive a shared leukemogenic transcriptional program in primary human HSCs. (**A**) Expression of EGFP control, NUP98a-HOXA9-EGFP, or NUP98a-KDM5A-EGFP ∼72 h post lentiviral transduction in primary human HSCs. Dotted lines encircle nuclei and scale bar = 10 μm. (**B**) PCA analysis of HSC transcriptomes after transduction with EGFP, NUP98a-HOXA9-EGFP, or NUP98a-KDM5A-EGFP. (**C**) Venn diagram showing the number and overlap of significantly upregulated (left) and downregulated (right) genes ∼14 days post lentiviral transduction. (**D**) Heatmap showing DEGs in human HSCs transduced with EGFP, NUP98a-HOXA9-EGFP, or NUP98a-KDM5A-EGFP. (**E**) KEGG analysis of all upregulated genes in human HSCs transduced with NUP98a-HOXA9-EGFP or NUP98a-KDM5A-EGFP. (**F**) KEGG analysis of all downregulated genes in human HSCs transduced with NUP98a-HOXA9-EGFP or NUP98a-KDM5A-EGFP.

genomic binding profiles using ChIP-seq. HEK293T cells were used due to technical difficulties associated with performing ChIP-seq in primary human HSCs. NUP98a-EGFP displayed potent nuclear import and condensed into sharp puncta in HEK293T cell nuclei but did not display significant levels of interaction with the human genome relative to mock transfected cells (Fig. 5A). However, fusion of either HOXA9 or KDM5A C-terminal domains to NUP98a resulted in a far more diffuse pattern of puncta formation in HEK293T cell nuclei and led to significant occupancy at thousands [∼18 000 and ∼2000, respectively, using DiffBind with a threshold of false discovery rate (FDR) < 0.05] of sites in the human genome (Fig. 5B and C). Interestingly, although the shorter NUP98b fragment formed fewer puncta and was more diffusely distributed compared to NUP98a, HOXA9, and KDM5A nevertheless drastically increased the nuclear localization and condensation of NUP98b, again with no effects on nuclei sizes (Supplementary Fig. S5A–C). These data demonstrate that fusion partners orchestrate the genomic binding of NUP98 FG repeats and likely also promote or reinforce their nuclear import, which we suspect is related to the oncogenic potential of NUP98 fusions, particularly those harboring fewer than ∼30 FG repeats.

ChIP-seq analyses also revealed that NUP98a-HOXA9 and NUP98a-KDM5A exhibited non-overlapping genomic binding profiles (Supplementary Fig. S5D and E), particularly with respect to human TSSs (Fig. 5D). In fact, ∼94% of NUP98a-KDM5A binding sites were located within promoter regions (<1 kb from TSSs), whereas only ∼5% of NUP98a-HOXA9 bound within 1 kb of annotated TSSs (Fig. 5E and F, Supplementary Fig. S5F). Instead, most NUP98a-HOXA9 binding sites (∼80%) were located within introns and distal intergenic regions >10 kb from TSSs (only ∼2% of NUP98a-KDM5A bound introns or distal intergenic regions). Using ChIPseeker [34] we also found that NUP98a-HOXA9 and NUP98a-KDM5A bound within annotated promoters, UTRs, exons, introns, downstream, or distal intergenic regions of ∼5475 and ∼2325 genes, respectively. However, only ∼738 of these genes were shared between fusions (Supplementary Fig. S5G and Supplementary Table S2). KEGG analyses also showed that while the binding of NUP98a-HOXA9 and NUP98a-KDM5A was associated with different biological pathways, one important exception was a shared association

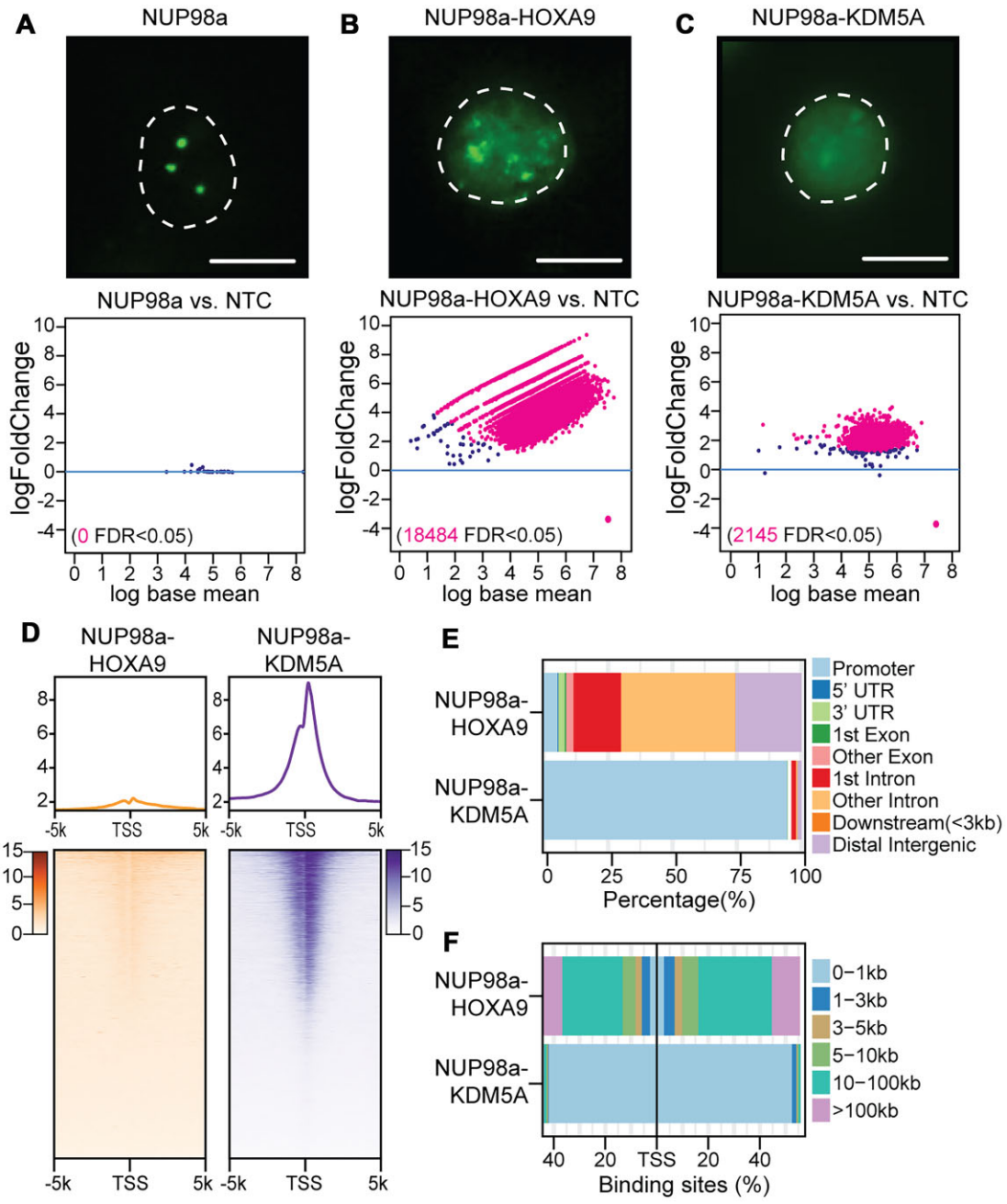


Figure 5. NUP98a-HOXA9 and NUP98a-KDM5A exhibit non-overlapping binding landscapes in human genome. (A–C) (upper panels) Fluorescence microscopy of NUP98a-EGFP control (panel **A**), NUP98a-HOXA9 (panel **B**), and NUP98a-KDM5A (panel **C**) in HEK293T cells ∼20 h post-transfection. All plasmid constructs harbor C-terminal EGFP fusions. Dotted lines encircle nuclei and scale bar = 10 μm. (lower panels) Genome-wide binding analysis of NUP98a control (panel A), NUP98a-HOXA9 (panel B), and NUP98a-KDM5A (panel C) ∼72 h post-transfection in HEK293T cells. Significant binding (FDR < 0.05) is indicated. (**D**) Genomic distribution of NUP98a-HOXA9 and NUP98a-KDM5A relative to all human transcription start sites (TSSs) ∼72 h post-transfection in HEK293T cells. (**E** and **F**) The distribution of NUP98a-HOXA9 and NUP98a-KDM5A binding with respect to genomic annotations (panel E) and distance to TSSs (panel F) using ChIPseeker.

with genes involved in oncogenic pathways (Supplementary Fig. S5H). Collectively, these data demonstrate that although the genomic binding landscapes of NUP98a-HOXA9 and NUP98a-KDM5A are largely non-overlapping in terms of annotated genomic sites (i.e. intergenic regions versus promoters) and in enriched genes and associated signaling pathways, these fusion proteins nevertheless share a propensity for binding key sites in the human genome associated with the onset and maintenance of oncogenesis.

NUP98-HOXA9 and NUP98-KDM5A activate human genes via distinct mechanisms

To better understand how NUP98a-HOXA9 and NUP98a-KDM5A stimulated the expression of leukemogenic genes {*MLLT3*, *MEIS1*, *FUT8*, *HMGA2*, *PBX3*, *SUPT3H*, *RUNX2*, and *HOXA9*, *ZBTB16*; Fig. 4, Supplementary Fig. S4, and reference [60]}, we more deeply interrogated our ChIP-seq data at these key loci in HEK293T cells (Fig. 6A and Supplementary Fig. S6A). These loci were

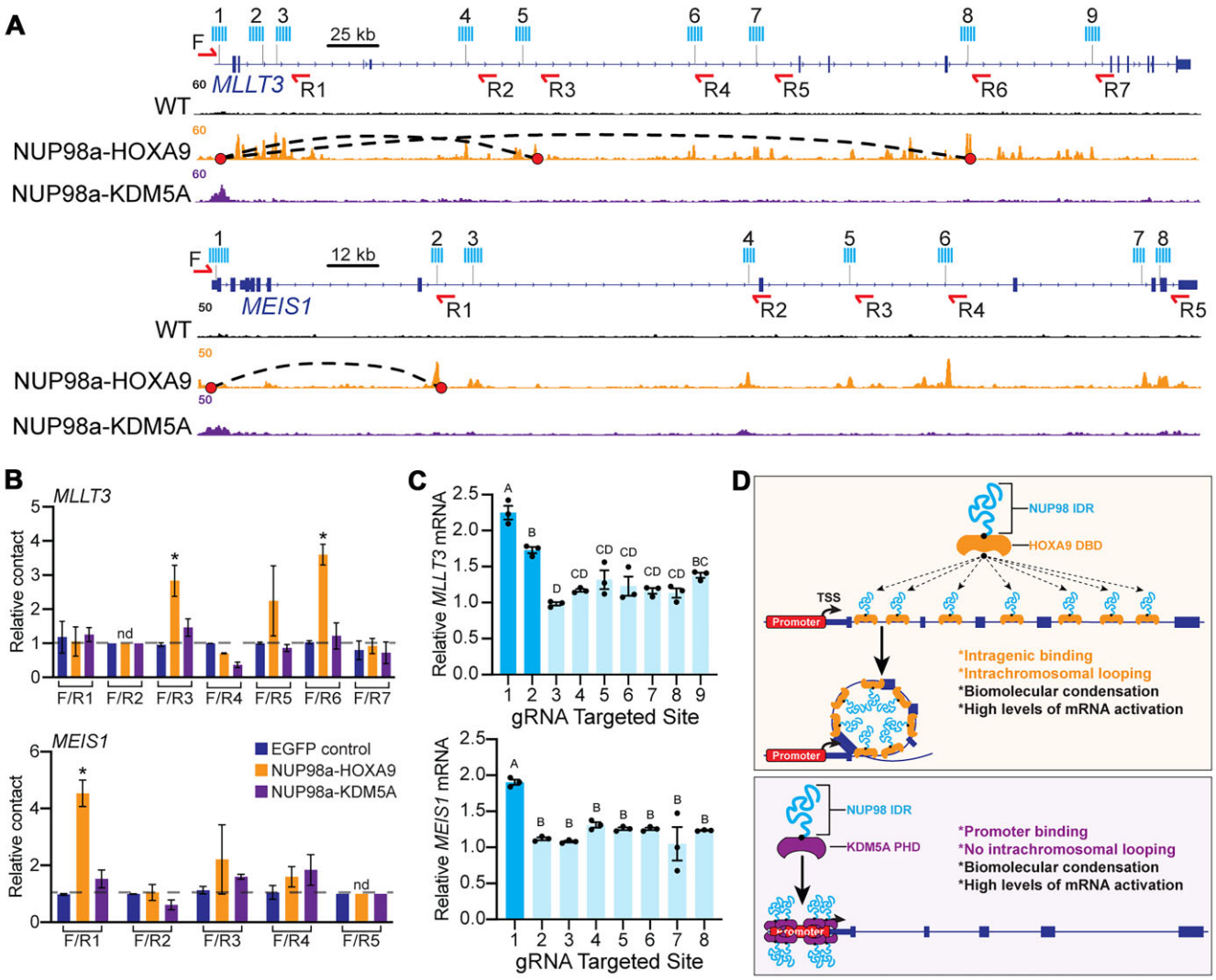


Figure 6. NUP98-HOXA9 and NUP98-KDM5A activate human genes via distinct mechanisms. (**A**) The genomic regions encompassing the *MLLT3* and *MEIS1* genes are shown along with ChIP-seq signals from wildtype HEK293T cells and HEK293T cells transfected with either NUP98a-HOXA9 or NUP98a-KDM5A. Genomic coordinates are based upon GRCh38/hg38. Bars indicate gRNA locations and arrows denote 3C-qPCR primers. Chromatin contacts detected via 3C-qPCR are labelled as arced lines. (**B**) Relative contact (using 3C-qPCR) between the *MLLT3* or *MEIS1* promoters and NUP98a-HOXA9 bound introns 2-weeks post-transduction with EGFP, NUP98a-HOXA9-EGFP, and NUP98a-KDM5A-EGFP in HEK293T. nd; not detected. Relative contact was normalized to non-transfected control cells. Asterisks (*) indicate $P < 0.05$ via two-tailed student's *t*-test. $n = 3$ independent experiments; error bars; s.e.m. (**C**) RT-qPCR for *MLLT3* (top) and *MEIS1* (bottom) mRNA ~72 h post-transfection of dCas9-NUP98a and corresponding gRNAs in HEK293T cells. mRNA levels were normalized to non-transfected control cells. Different letters above error bars indicate significant difference between different groups ($P < 0.05$) using one-way analysis of variance followed by Tukey's post-hoc test. (**D**) Illustration of gene activation models for NUP98-HOXA9 and NUP98-KDM5A.

indeed bound by NUP98a-HOXA9 and NUP98a-KDM5A, albeit with strikingly different binding patterns. For instance, NUP98a-HOXA9 was highly enriched across multiple binding sites spanning entire gene bodies, while NUP98a-KDM5A showed highly localized enrichment near promoters. Binding of NUP98 IDR fusions has been shown to modulate chromosomal contacts [13], therefore we hypothesized that the binding of NUP98a-HOXA9 and NUP98a-KDM5A at these genes was also changing chromatin conformation. To test this hypothesis, we performed 3C-qPCR in HEK293T cells transduced with EGFP, NUP98a-HOXA9, or NUP98a-KDM5A. We observed that NUP98a-HOXA9 catalyzed increased *de novo* chromatin contacts between NUP98a-HOXA9 sites near the *MLLT3* and *MEIS1* promoters and other binding sites within the introns of these genes (Fig. 6B). In contrast, no increases in chromosomal contacts were observed in NUP98a-KDM5A transfected cells (Fig. 6B). These data suggest NUP98a-KDM5A directly activates human genes by localization of the NUP98 IDR at promoters, while NUP98a-HOXA9 likely controls transcription through the creation of chromatin contacts between promoters and other bound sites across gene bodies and/or intergenic regions.

To study the transcriptional effects of the NUP98 IDR at specific ChIP-seq enriched sites within these loci, we next targeted the engineered dCas9-NUP98a to genomic segments matching the sites occupied by NUP98a-HOXA9 and NUP98a-KDM5A within *MTTL3* and *MEIS1* (Fig. 6A). We found that targeting of dCas9-NUP98a to the *MTTL3* or *MEIS1* promoters significantly stimulated the transcription of these genes (Fig. 6C). However, we were unable to stimulate

MLLT3 or *MEIS1* expression by targeting dCas9-NUP98a to any distal genomic segments occupied by NUP98a-HOXA9 (Fig. 6C). Interestingly, dCas9-NUP98a was also unable to stimulate gene expression when targeted to intergenic *OCT4* or *HBG1* enhancers, or to augment gene expression when targeted to enhancers and cognate promoters [36, 61, 62] (Supplementary Fig. S6B and C). Collectively, we conclude that the transcriptional effects of the NUP98 IDR are exerted via high-density localization or proximity to human promoters, which is sufficient to drive pathological gene expression in NUP98 IDR-associated cancers. Our data also suggest that the targeting of the NUP98 IDR to short segments (<500 bp) in introns or distal enhancers using dCas9-NUP98a is insufficient to activate human genes, however, the high-density binding across longer genomic segments (>5 kb) as observed in NUP98a-HOXA9 binding, likely catalyzes increased *de novo* chromatin contacts/looping, which ultimately results in recruiting high levels of NUP98 FG repeats to promoters and in turn pathological gene activation (Fig. 6D).

Discussion

Here, we demonstrate that human IDRs can be sufficient to initiate transcription at endogenous human promoters. Although we show that several human IDRs (EWSR1, SS18, NONO, and NUP98) harbor this potential (Fig. 1), we focus here on the NUP98 IDR due to (i) its transcriptional potency, (ii) its high degree of condensation, and (iii) its role in treatment resistant cancers. In particular, we dissect the regulatory role of the NUP98 IDR and associated FG repeats at specific loci using synthetic dCas9-NUP98 fusions. We also study the genome-wide effects of these biomolecules using paradigmatic NUP98 IDR (FG repeat)-harboring oncogenic fusions (NUP98-HOXA9 and NUP98-KDM5A). Altogether, our studies and prior reports [11, 13] suggest that the inappropriate localization and condensation of the NUP98 IDR at key genomic sites is likely an important factor in NUP98 IDR-driven leukemias. In primary human HSCs this aberrant activity provokes a leukemic gene expression program in which genes encoding differentiation trajectories are suppressed and those stimulating proliferation are activated. These findings open new therapeutic opportunities with respect to NUP98 IDR-associated leukemias, for instance, through disrupting the ability of NUP98 to import into cell nuclei, phase separate/condense, and/or to modulate the key target genes identified here.

IDR-containing proteins can phase separate after exceeding a critical concentration [63], and aromatic amino acid resides can contribute to liquid–liquid phase separation through pi-pi interactions [64]. The NUP98 IDR, which consists of aromatic FG repeats, has been shown to spontaneously and rapidly phase separate *in vitro* [65]. Here, we observed that the EGFP-tagged NUP98 IDR and NUP98 IDR-harboring fusion proteins spontaneously phase separate in human cells (Figs 1–4). We note that we cannot exclude the possibility that the EGFP tag may slightly affect the phase separation properties of NUP98 FG repeats. We also showed that the aromatic phenylalanine (F) residue within the NUP98 IDR determines phase separation and gene-regulatory capacities at endogenous sites within human cells (Fig. 3). These findings align with recent thermodynamic models suggesting a log-linear relationship between NUP98 FG repeat numbers and partition coefficients *in vitro* [66]. We also directly connected the condensation behavior of the NUP98 IDR to gene activation at endogenous human promoters and interestingly, our data show that the NUP98 IDR, displays a nonlinear dose-responsive gene regulatory curve at human promoters (Fig. 2A–D), which is considered a defining feature of transcriptional condensates [40–42]. However, we note that dox administration is not without limitations with respect to quantification [67]. We further showed that engineered, gRNA-targeted dCas9-NUP98 condensates are localized to sites of targeted transactivation in U2OS p21-MS2 + MCP-EGFP cells [24] (Fig. 2H), further supporting a model in which NUP98 FG repeats largely exert their effects as founding members of transcriptional condensates.

IDR-driven phase separation behavior has also been linked to transcriptional control and chromatin looping at super enhancers in mammalian cells [42, 68 , 69]. The well-defined phase separation property of the NUP98 IDR [65] and the binding pattern and chromatin looping catalyzed by NUP98a-HOXA9 identified here and previously [13], suggest that NUP98a-HOXA9 adopts super enhancer-like transcriptional control during leukemic transformation. Moreover, our studies suggest that the super enhancer-like transcriptional control of NUP98a-HOXA9 fusions is intensified by the high-density of FG repeats bound across long stretches (>5 kb) of genomic DNA (Fig. 6A and Supplementary Fig. S6A). For instance, our RNA-seq and ChIP-seq data show that NUP98a-HOXA9 strongly activates the *PBX3* [57], *SUPT3H* [58], and *FUT8* [59] proto-oncogenes through dense clusters of binding across respective gene bodies, whereas NUP98a-KDM5A is unable to activate these genes through localized promoter binding (Supplementary Table S1). This strict dependence on binding site number and density over a long genomic window is a hallmark of super enhancers {average window ∼8.7 kb [70]} and may explain why targeting dCas9-NUP98a to short (∼500–1000 bp) 'typical' enhancers or to short segments of NUP98a-HOXA9 bound introns is insufficient to stimulate transcription (Supplementary Fig. S6B). In addition, our data demonstrate that dCas9-NUP98a functions similarly to dCas9-KDM5A, in which high levels of focused enrichment at target promoters leads to phase separation and increased gene expression (Fig. 6D).

The genomic binding profiles and transcriptional regulatory mechanisms of NUP98-HOXA9 and NUP98a-KDM5A likely exemplify the oncogenic processes of the myriad other homeodomain-containing and chromatin-modifying NUP98-IDR fusion proteins observed in cancers. For instance, homeodomain fusions occur in numerous other NUP98 IDR-associated leukemias (e.g. HOXA11/13, HOXC11/13, HOXD11/13, PMX1/2, and HHEX) [14]. Furthermore, another chromatin-modifying fusion partner, NSD1 has also recently been shown to display substantial binding around promoter regions [71], suggesting that NUP98-NSD1 fusions act similarly to NUP98-KDM5A and trigger leukemic transformation through aberrant focal condensation at promoters. Nevertheless, our studies identify a shared set of stimulated genes between these two major groups of NUP98 fusions. Therefore, we speculate that NUP98-based malignancies may have common therapeutic dependencies that could be exploited for effective and ubiquitous treatment options. For instance, our data and previous studies [46, 50] establish that HOXA9-based regulatory components, such as HOXA9 and MEIS1 are commonly triggered by NUP98-HOXA9 and NUP98-KDM5A. Indeed, although NUP98-HOXA9 and

NUP98a-KDM5A are associated with different subtypes of leukemia [72], unfortunately patients with either fusion exhibit similarly low survival rates [46]. However, we also note that NUP98-HOXA9 and NUP98-KDM5A have many non-overlapping effects, and thus unique therapeutic strategies could also be tailored for specific NUP98 fusions [46].

Altogether, our studies here demonstrate that the NUP98 IDR can drive biomolecular condensation and activate human genes, which in turn clarifies the biological role of the NUP98 IDR fusions in oncogenesis. Further interrogation of these processes could be useful for the treatment of NUP98 IDR-driven human cancers, as well as other cancers caused by inappropriate fusions between IDRs and proteins that bind the human genome [19]. Moreover, recent efforts have shown that the important biomolecular properties of IDRs can be harnessed to enhance cellular engineering applications [21–23], suggesting that NUP98 IDR and other IDRs could be further developed to create new opportunities for bioengineering and synthetic biology.

Acknowledgements

The authors thank all members of the Hilton lab for helpful discussions and insights. J.L. and I.B.H. conceived the project and designed experiments. J.L. performed most experiments with the assistance of S.L., S.K., J.G., Y.Z., E.R.T., and A.R. J.L. analyzed the data with the assistance of Z.A.D., J.P.F., A.J.M., B.M., B.Y., and M.E. J.L. and I.B.H. wrote the manuscript with input from all authors.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

J.L., S.L., J.G., B.M, M.E., and I.B.H. are inventors on patents related to this work and other genome and epigenome editing technologies. The remaining authors declare no competing interests.

Funding

This work was supported by a Cancer Prevention & Research Institute of Texas (CPRIT) Award (RR170030) and National Institutes of Health (NIH) Awards (R35GM143532 and R56HG012206) to I.B.H. M.D.E. was supported by an American Heart Association predoctoral fellowship program (917025).

Data availability

RNA and ChIP sequencing data will be released from the NCBI Gene Expression Omnibus repository (GSE248061) upon publication. Key plasmids are available on Addgene, and all plasmids and reagents are available from the authors upon request.

References

1. Oates ME, Romero P, Ishida T *et al.* D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* 2013;**41**:D508–6. https://doi.org/10.1093/nar/gks1226
2. Holehouse AS, Kragelund BB. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol* 2024;**25**:187–211. https://doi.org/10.1038/s41580-023-00673-0
3. Cermakova K, Hodges HC. Interaction modules that impart specificity to disordered protein. *Trends Biochem Sci* 2023;**48**:477–90. https://doi.org/10.1016/j.tibs.2023.01.004
4. Mar M, Nitsenko K, Heidarsson PO. Multifunctional intrinsically disordered regions in transcription factors. *Chemistry* 2023;**29**:e202203369. https://doi.org/10.1002/chem.202203369
5. Franks TM, Hetzer MW. The role of Nup98 in transcription regulation in healthy and diseased cells. *Trends Cell Biol* 2013;**23**:112–7. https://doi.org/10.1016/j.tcb.2012.10.013
6. Feng P, Li L, Deng T *et al.* NONO and tumorigenesis: more than splicing. *J Cell Mol Med* 2020;**24**:4368–76. https://doi.org/10.1111/jcmm.15141
7. Quiroga IY, Ahn JH, Wang GG *et al.* Oncogenic fusion proteins and their role in three-dimensional chromatin structure, phase separation, and cancer. *Curr Opin Genet Dev* 2022;**74**:101901. https://doi.org/10.1016/j.gde.2022.101901
8. Shirnekhi HK, Chandra B, Kriwacki RW. The role of phase-separated condensates in fusion oncoprotein-driven cancers. *Annu Rev Cancer Biol* 2023;**7**:73–91. https://doi.org/10.1146/annurev-cancerbio-061421-122050
9. Brien GL, Stegmaier K, Armstrong SA. Targeting chromatin complexes in fusion protein-driven malignancies. *Nat Rev Cancer* 2019;**19**:255–69. https://doi.org/10.1038/s41568-019-0132-x
10. Franks TM, McCloskey A, Shokirev MN *et al.* Nup98 recruits the Wdr82-Set1A/COMPASS complex to promoters to regulate H3K4 trimethylation in hematopoietic progenitor cells. *Genes Dev* 2017;**31**:2222–34. https://doi.org/10.1101/gad.306753.117
11. Chandra B, Michmerhuizen NL, Shirnekhi HK *et al.* Phase Separation Mediates NUP98 Fusion Oncoprotein Leukemic Transformation. *Cancer Discov* 2022;**12**:1152–69. https://doi.org/10.1158/2159-8290.CD-21-0674
12. Terlecki-Zaniewicz S, Humer T, Eder T *et al.* Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression. *Nat Struct Mol Biol* 2021;**28**:190–201. https://doi.org/10.1038/s41594-020-00550-w
13. Ahn JH, Davis ES, Daugird TA *et al.* Phase separation drives aberrant chromatin looping and cancer development. *Nature* 2021;**595**:591–5. https://doi.org/10.1038/s41586-021-03662-5
14. Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. *Blood* 2011;**118**:6247–57. https://doi.org/10.1182/blood-2011-07-328880
15. Michmerhuizen NL, Klco JM, Mullighan CG. Mechanistic insights and potential therapeutic approaches for NUP98-rearranged hematologic malignancies. *Blood* 2020;**136**:2275–89. https://doi.org/10.1182/blood.2020007093
16. Arai Y, Hosoda F, Kobayashi H *et al.* The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, NUP98, with the putative RNA helicase gene, DDX10. *Blood* 1997;**89**:3936–44. https://doi.org/10.1182/blood.V89.11.3936
17. Hussey DJ, Moore S, Nicola M *et al.* Fusion of the NUP98 gene with the LEDGF/p52 gene defines a recurrent acute myeloid leukemia translocation. *BMC Genet* 2001;**2**:20. https://doi.org/10.1186/1471-2156-2-20
18. Kasper LH, Brindle PK, Schnabel CA *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol Cell Biol* 1999;**19**:764–76. https://doi.org/10.1128/MCB.19.1.764
19. Niu X, Zhang L, Wu Y *et al.* Biomolecular condensates: formation mechanisms, biological functions, and therapeutic targets. *MedComm* 2023;**4**:e223. https://doi.org/10.1002/mco2.223
20. Naderi J, Magalhaes AP, Kibar G *et al.* An activity-specificity trade-off encoded in human transcription factors. *Nat Cell Biol* 2024;**26**:1309–21. https://doi.org/10.1038/s41556-024-01411-0

21. Liu J, Chen Y, Nong B *et al.* CRISPR-assisted transcription activation by phase separation proteins. *Protein Cell* 2023;**14**:874–87. https://doi.org/10.1093/procel/pwad013
22. Ma S, Liao K, Li M *et al.* Phase-separated DropCRISPRa platform for efficient gene activation in mammalian cells and mice. *Nucleic Acids Res* 2023;**51**:5271–84. https://doi.org/10.1093/nar/gkad301
23. Kim YJ, Lee M Jr, Lee YT *et al.* Light-activated macromolecular phase separation modulates transcription by reconfiguring chromatin interactions. *Sci Adv* 2023;**9**:eadg1123. https://doi.org/10.1126/sciadv.adg1123
24. Carvajal LA, Neriah DB, Senecal A *et al.* Dual inhibition of MDMX and MDM2 as a therapeutic strategy in leukemia. *Sci Transl Med* 2018;**10**:eaao3003. https://doi.org/10.1126/scitranslmed.aao3003
25. Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15. https://doi.org/10.1038/s41587-019-0201-4
26. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;**30**:923–30. https://doi.org/10.1093/bioinformatics/btt656
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. https://doi.org/10.1186/s13059-014-0550-8
28. Wu T, Hu E, Xu S *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;**2**:100141.
29. Robinson JT, Thorvaldsdottir H, Turner D *et al.* igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* 2023;**39**:btac830. https://doi.org/10.1093/bioinformatics/btac830
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
31. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137. https://doi.org/10.1186/gb-2008-9-9-r137
32. Ramirez F, Ryan DP, Gruning B *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;**44**:W160–5. https://doi.org/10.1093/nar/gkw257
33. Ross-Innes CS, Stark R, Teschendorff AE *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012;**481**:389–93. https://doi.org/10.1038/nature10730
34. Wang Q, Li M, Wu T *et al.* Exploring epigenomic datasets by ChIPseeker. *Curr Protoc* 2022;**2**:e585. https://doi.org/10.1002/cpz1.585
35. Hagege H, Klous P, Braem C *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2007;**2**:1722–33. https://doi.org/10.1038/nprot.2007.243
36. Hilton IB, D'Ippolito AM, Vockley CM *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015;**33**:510–7. https://doi.org/10.1038/nbt.3199
37. Wang K, Escobar M, Li J *et al.* Systematic comparison of CRISPR-based transcriptional activators uncovers gene-regulatory features of enhancer-promoter interactions. *Nucleic Acids Res* 2022;**50**:7842–55. https://doi.org/10.1093/nar/gkac582
38. Mathias KM, Liu Y, Wan L. Dysregulation of transcriptional condensates in human disease: mechanisms, biological functions, and open questions. *Curr Opin Genet Dev* 2024;**86**:102203. https://doi.org/10.1016/j.gde.2024.102203
39. Sabari BR. Biomolecular condensates and gene activation in development and disease. *Dev Cell* 2020;**55**:84–96. https://doi.org/10.1016/j.devcel.2020.09.005
40. Shrinivas K, Sabari BR, Coffey EL *et al.* Enhancer features that drive formation of transcriptional condensates. *Mol Cell* 2019;**75**:549–61. https://doi.org/10.1016/j.molcel.2019.07.009
41. Ma L, Gao Z, Wu J *et al.* Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. *Mol Cell* 2021;**81**:1682–97. https://doi.org/10.1016/j.molcel.2021.01.031
42. Hnisz D, Shrinivas K, Young RA *et al.* A phase separation model for transcriptional control. *Cell* 2017;**169**:13–23. https://doi.org/10.1016/j.cell.2017.02.007
43. Li J, Mahata B, Escobar M *et al.* Programmable human histone phosphorylation and gene activation using a CRISPR/Cas9-based chromatin kinase. *Nat Commun* 2021;**12**:896. https://doi.org/10.1038/s41467-021-21188-2
44. Kroschwald S, Maharana S, Simon AW. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* 2017; http://dx.doi.org/10.19185/matters.201702000010
45. Rio-Machin A, Gomez-Lopez G, Munoz J *et al.* The molecular pathogenesis of the NUP98-HOXA9 fusion protein in acute myeloid leukemia. *Leukemia* 2017;**31**:2000–2005. https://doi.org/10.1038/leu.2017.194
46. Bertrums EJM, Smith JL, Harmon L *et al.* Comprehensive molecular and clinical characterization of NUP98 fusions in pediatric acute myeloid leukemia. *Haematol* 2023;**108**:2044–58. https://doi.org/10.3324/haematol.2022.281653
47. Chou WC, Chen CY, Hou HA *et al.* Acute myeloid leukemia bearing t(7;11)(p15;p15) is a distinct cytogenetic entity with poor outcome and a distinct mutation profile: comparative analysis of 493 adult patients. *Leukemia* 2009;**23**:1303–10. https://doi.org/10.1038/leu.2009.25
48. Noort S, Wander P, Alonzo TA *et al.* The clinical and biological characteristics of NUP98-KDM5A in pediatric acute myeloid leukemia. *Haematol* 2021;**106**:630–4. https://doi.org/10.3324/haematol.2019.236745
49. Kanehisa M, Furumichi M, Sato Y *et al.* KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 2023;**51**:D587–92. https://doi.org/10.1093/nar/gkac963
50. Collins CT, Hess JL. Deregulation of the HOXA9/MEIS1 axis in acute leukemia. *Curr Opin Hematol* 2016;**23**:354–61. https://doi.org/10.1097/MOH.0000000000000245
51. Matthijssens F, Sharma ND, Nysus M *et al.* RUNX2 regulates leukemic cell metabolism and chemotaxis in high-risk T cell acute lymphoblastic leukemia. *J Clin Invest* 2021;**131**. e141566 https://doi.org/10.1172/JCI141566
52. Rampal R, Figueroa ME. Wilms tumor 1 mutations in the pathogenesis of acute myeloid leukemia. *Haematologica* 2016;**101**:672–9. https://doi.org/10.3324/haematol.2015.141796
53. Suliman BA, Xu D, Williams BR. The promyelocytic leukemia zinc finger protein: two decades of molecular oncology. *Front Oncol* 2012;**2**:74. https://doi.org/10.3389/fonc.2012.00074
54. Calvanese V, Nguyen AT, Bolan TJ *et al.* MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature* 2019;**576**:281–6. https://doi.org/10.1038/s41586-019-1790-2
55. Stanley ER. Lineage commitment: cytokines instruct, at last! *Cell Stem Cell* 2009;**5**:234–6. https://doi.org/10.1016/j.stem.2009.08.015
56. Takeda A, Goolsby C, Yaseen NR. NUP98-HOXA9 induces long-term proliferation and blocks differentiation of primary human CD34+ hematopoietic cells. *Cancer Res* 2006;**66**:6628–37. https://doi.org/10.1158/0008-5472.CAN-06-0458
57. Li Z, Zhang Z, Li Y *et al.* PBX3 is an important cofactor of HOXA9 in leukemogenesis. *Blood* 2013;**121**:1422–31. https://doi.org/10.1182/blood-2012-07-442004
58. Nakamura Y, Kayano H, Kakegawa E *et al.* Identification of SUPT3H as a novel 8q24/MYC partner in blastic plasmacytoid dendritic cell neoplasm with t(6;8)(p21;q24) translocation. *Blood Cancer J* 2015;**5**:e301. https://doi.org/10.1038/bcj.2015.26
59. Agrawal P, Fontanals-Cirera B, Sokolova E *et al.* A systems biology approach identifies FUT8 as a driver of melanoma metastasis. *Cancer Cell* 2017;**31**:804–19. https://doi.org/10.1016/j.ccell.2017.05.007

60. Liu Y, Luo M, Jin Z *et al.* dbLGL: an online leukemia gene and literature database for the retrospective comparison of adult and childhood leukemia genetics with literature evidence. *Database (Oxford)* 2018;**2018**;bay062. https://doi.org/10.1093/database/bay062
61. Deng W, Rupon JW, Krivega I *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 2014;**158**:849–60. https://doi.org/10.1016/j.cell.2014.05.050
62. Carter D, Chakalova L, Osborne CS *et al.* Long-range chromatin regulatory interactions *in vivo*. *Nat Genet* 2002;**32**:623–6. https://doi.org/10.1038/ng1051
63. Alberti S, Gladfelter A, Mittag T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* 2019;**176**:419–34. https://doi.org/10.1016/j.cell.2018.12.035
64. Vernon RM, Chong PA, Tsang B *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* 2018;**7**:e31486. https://doi.org/10.7554/eLife.31486
65. Schmidt HB, Gorlich D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* 2015;**4**:e04251. https://doi.org/10.7554/eLife.04251
66. Ng SC, Gorlich D. A simple thermodynamic description of phase separation of Nup98 FG domains. *Nat Commun* 2022;**13**:6172. https://doi.org/10.1038/s41467-022-33697-9
67. Kabaria SR, Bae Y, Ehmann ME *et al.* Programmable promoter editing for precise control of transgene expression. bioRxiv, https://doi.org/10.1101/2024.06.19.599813, 14 July 2024, preprint: not peer reviewed.
68. Cho WK, Spille JH, Hecht M *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 2018;**361**:412–5. https://doi.org/10.1126/science.aar4199
69. Sabari BR, Dall'Agnese A, Boija A *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 2018;**361**;eaar3958. https://doi.org/10.1126/science.aar3958
70. Hnisz D, Abraham BJ, Lee TI *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 2013;**155**:934–47. https://doi.org/10.1016/j.cell.2013.09.053
71. Ren Z, Kim A, Huang YT *et al.* A PRC2-Kdm5b axis sustains tumorigenicity of acute myeloid leukemia. *Proc Natl Acad Sci USA* 2022;**119**;e2122940119. https://doi.org/10.1073/pnas.2122940119
72. Mohanty S. NUP98 rearrangements in AML: molecular mechanisms and clinical implications. *Onco* 2023;**3**:147–64. https://doi.org/10.3390/onco3030011

Received: August 24, 2024. **Revised:** January 2, 2025. **Editorial Decision:** January 7, 2025. **Accepted:** January 22, 2025