Inventory of Supplemental Materials

7 supplemental figures, 7 supplemental tables, supplemental experimental procedures, and supplemental references.

Supplemental Figures

Figure S1. Quantification of the number of NICD nuclear bodies in individual cells over time and co-localization screen of NICD nuclear bodies with endogenous nuclear bodies (related to Figure 1)

Figure S2. Size-dependent effects of the point spread function on nuclear body intensity, additional snapshots of pairs of NICD molecules whose association is mediated by Arg_6 peptides, and influence of free parameters and variables (K_D, C_{sat}, B) on complex coacervation (related to Figure 2)

Figure S3. Isolated human NICD is intrinsically disordered and monomeric but phase separates with positively charged partners (related to Figure 3)

Figure S4. Expression levels of wild type NICD and its mutants are similar and do not affect qualitative conclusions of nuclear body quantification studies (related to Figure 5)

Figure S5. Charge patterning of NICD and net charge of partners affects phase separation in vitro (related to Figure 5)

Figure S6. Identification of residue types that promote nuclear body formation (related to Figure 6)

Figure S7. Distribution of amino acids for groups of sequences (related to Figure 6)

Supplemental Tables

Table S1. Mean percentage of nuclear bodies observed normalized to wild type NICD, and the number of residues in the [Y/D/R] critical residue set (related to Figure 6)

Table S2. Examples of the Monte Carlo subsampling procedure, where each iteration (SS^1, SS^2, etc.) represents a random Monte Carlo selection. Specific values are sampled from the "Raw data" set to generate 2000 randomly constructed datasets (related to Figure 6)

Table S3. NICD-like proteins (n=443) in the human proteome, identified based on similar charge properties and charge interaction elements (related to Figure 6 and Discussion)

Table S4. Enriched molecular functions and cellular components (GO) of NICD-like proteins (n=443) in the human proteome, identified based on similar charge properties and charge interaction elements (related to Figure 6 and Discussion)

Table S5. Enriched biological processes and cellular components (GO) of NICD-like proteins (n=260) in the human proteome, identified based on similar charge properties and charge interaction elements and in which tyrosine and/or leucines represent >=6% of residues in the region (related to Figure 6 and Discussion)

Table S6. Sequences of constructs used in this study (related to Experimental Procedures)

Supplemental Experimental Procedures

Supplemental References

Supplemental Figures

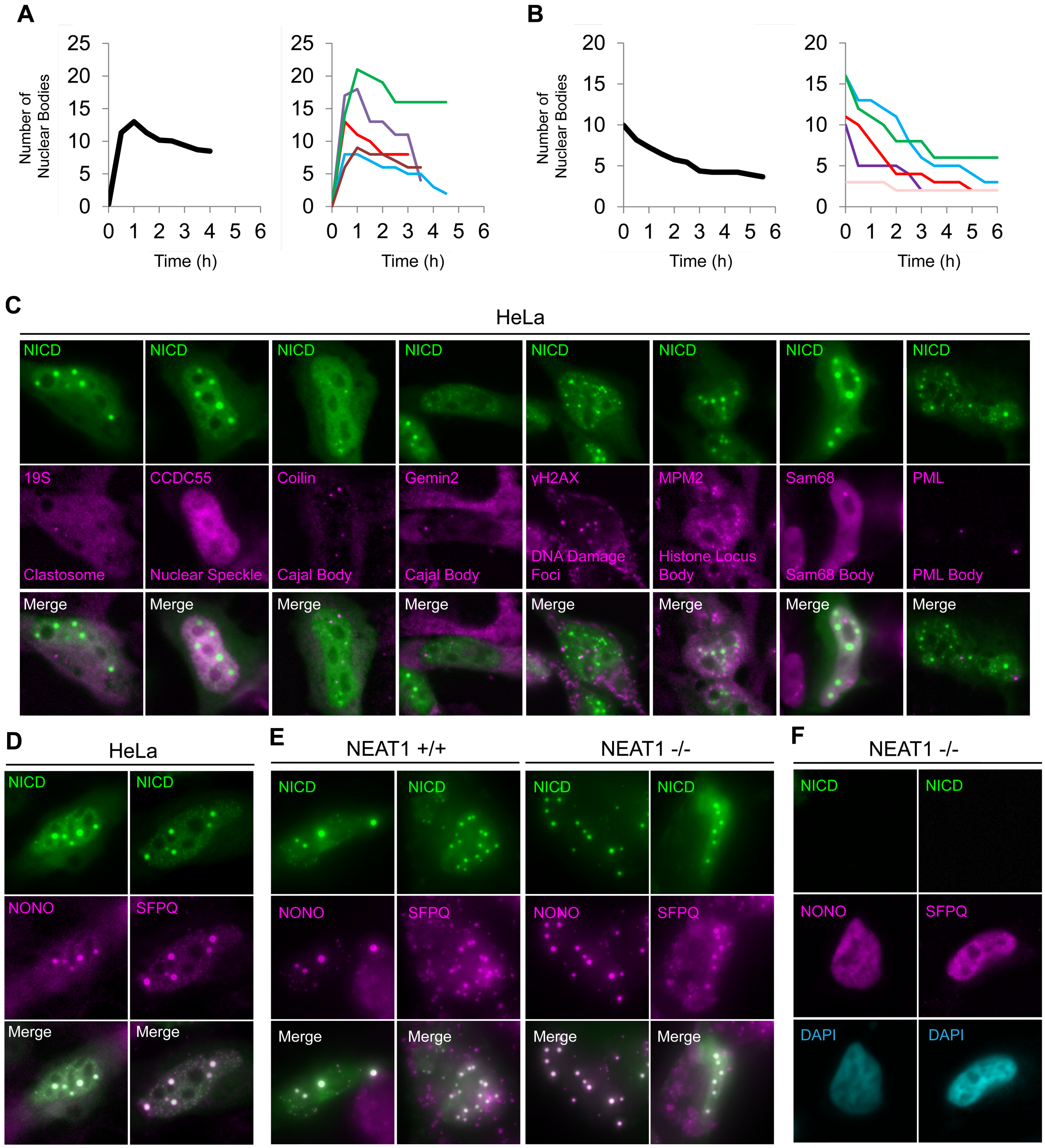

Figure S1. Quantification of the number of NICD nuclear bodies in individual cells over time and co-localization screen of NICD nuclear bodies with endogenous nuclear bodies (related to Figure 1)

(A,B) Quantification of the number of NICD nuclear bodies in individual HeLa cells transiently transfected with NICD-YPet and imaged over a 6 h period. Cells were classified based on whether they will nuclear bodies only after the start of imaging (panel A) or had already formed nuclear bodies before the start of imaging (panel B). In A, the time point preceding the first appearance of nuclear bodies was set to t_0. On average, nuclear bodies form near-coincidentally within a short time period, presumably as NICD expression increases past a threshold (A, left panel). The number of nuclear bodies in representative individual cells are shown in the right panel. Over time and despite increasing NICD expression, the average number of NICD nuclear bodies decreases (B, left panel; t_0 = start of imaging 4 h after transfection). The number of nuclear bodies in representative individual cells are shown in the right panel. (C) HeLa cells expressing NICD-YPet (green) were stained with antibodies (magenta) to visualize marker proteins (in parenthesis) characteristic of the indicated endogenous nuclear bodies: clastosome (19S), nuclear speckles (CCDC55), Cajal bodies (coilin/gemin2), DNA damage foci (γH2AX), histone locus bodies (MPM2), Sam68 bodies (Sam68), PML bodies (PML), and paraspeckles (NONO/SFPQ; shown in B). (D) NICD (green) nuclear bodies co-localized with known paraspeckle proteins, NONO and SFPQ (magenta), in HeLa cells. (E) NICD (green) nuclear bodies form in NEAT1 -/- mouse embryonic fibroblasts (MEFs), which lack paraspeckles, and in parental NEAT1 +/+ MEFs. NICD co-localizes with NONO and SFPQ (magenta) in NEAT1 -/- and +/+ cells. (F) NONO and SFPQ (magenta) are diffuse in nuclei of NEAT1 -/- MEFs that have not formed NICD nuclear bodies.

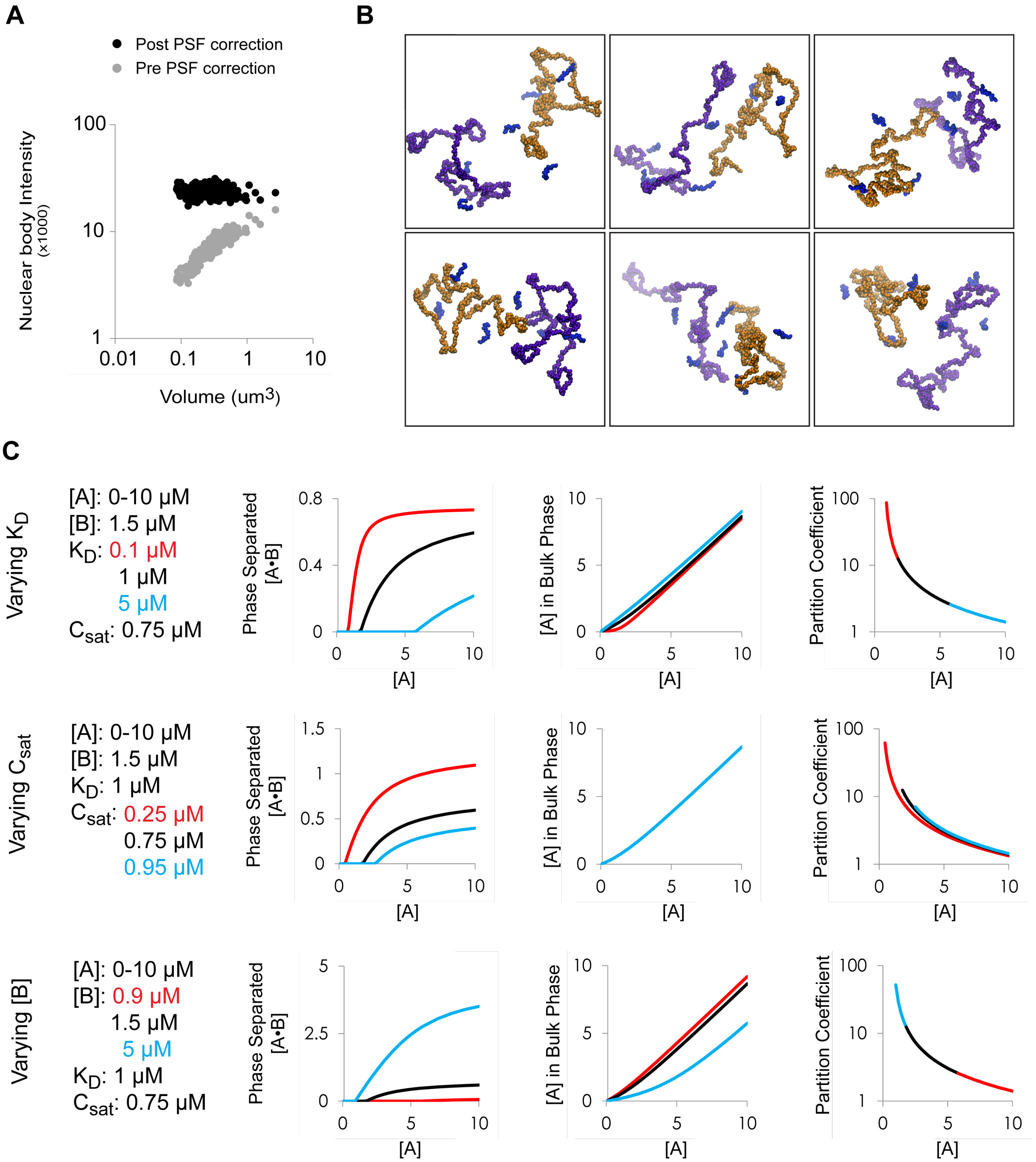
A
Post PSF correction
Pre PSF correction
Nuclear body Intensity (x1000)
100
10
1
0.01
0.1
1
10
Volume (um3)
B
C
Varying K_D
[A]: 0-10 μM
[B]: 1.5 μM
K_D: 0.1 μM
1 μM
5 μM
C_sat: 0.75 μM
Phase Separated [A•B]
0.8
0.4
0
5
10
[A]
[A] in Bulk Phase
Partition Coefficient
Varying C_sat
[A]: 0-10 μM
[B]: 1.5 μM
K_D: 1 μM
C_sat: 0.25 μM
0.75 μM
0.95 μM
1.5
1
0.5
Varying [B]
[A]: 0-10 μM
[B]: 0.9 μM
1.5 μM
5 μM
K_D: 1 μM
C_sat: 0.75 μM
2.5

Figure S2. Size-dependent effects of the point spread function on nuclear body intensity, additional snapshots of pairs of NICD molecules whose association is mediated by Arg_6 peptides, and influence of free parameters and variables (K_D, saturation concentration (C_{sat}), B) on complex coacervation (related to Figure 2)

(A) Quantification of nuclear body intensity (n=239 nuclear bodies from 30 cells) before (gray) and after (black) correcting for the effect of the point spread function, as described in Supplemental Methods. Before correcting for the point spread function, nuclear body intensity scales with volume. After correcting for the point spread function, nuclear body intensity is independent of volume. (B) Representative snapshots of pairs of NICD molecules from atomistic Monte Carlo simulations in the presence of Arg_6 peptides. Atomistic Monte Carlo simulations were performed using the ABSINTH implicit solvation model and forcefield paradigm. Backbone-only representations are shown, with the two NICDs shown in magenta and orange and Arg_6 peptides shown in blue. (C) Influence of free parameters and variables, K_D, C_{sat}, and [B], on complex coacervation of [A•B] when 0-10 μM of A is added to B. The amount of phase separated [A•B] increases with lower K_D and C_{sat} and with more B (left panels). The amount of A in the bulk phase does not significantly change except when more B is present (middle panels). The maximum partition coefficient increases with lower K_D and C_{sat} and with more B (right panels). In general, the qualitative behaviors of the system do not change over a substantial range of parameter values.

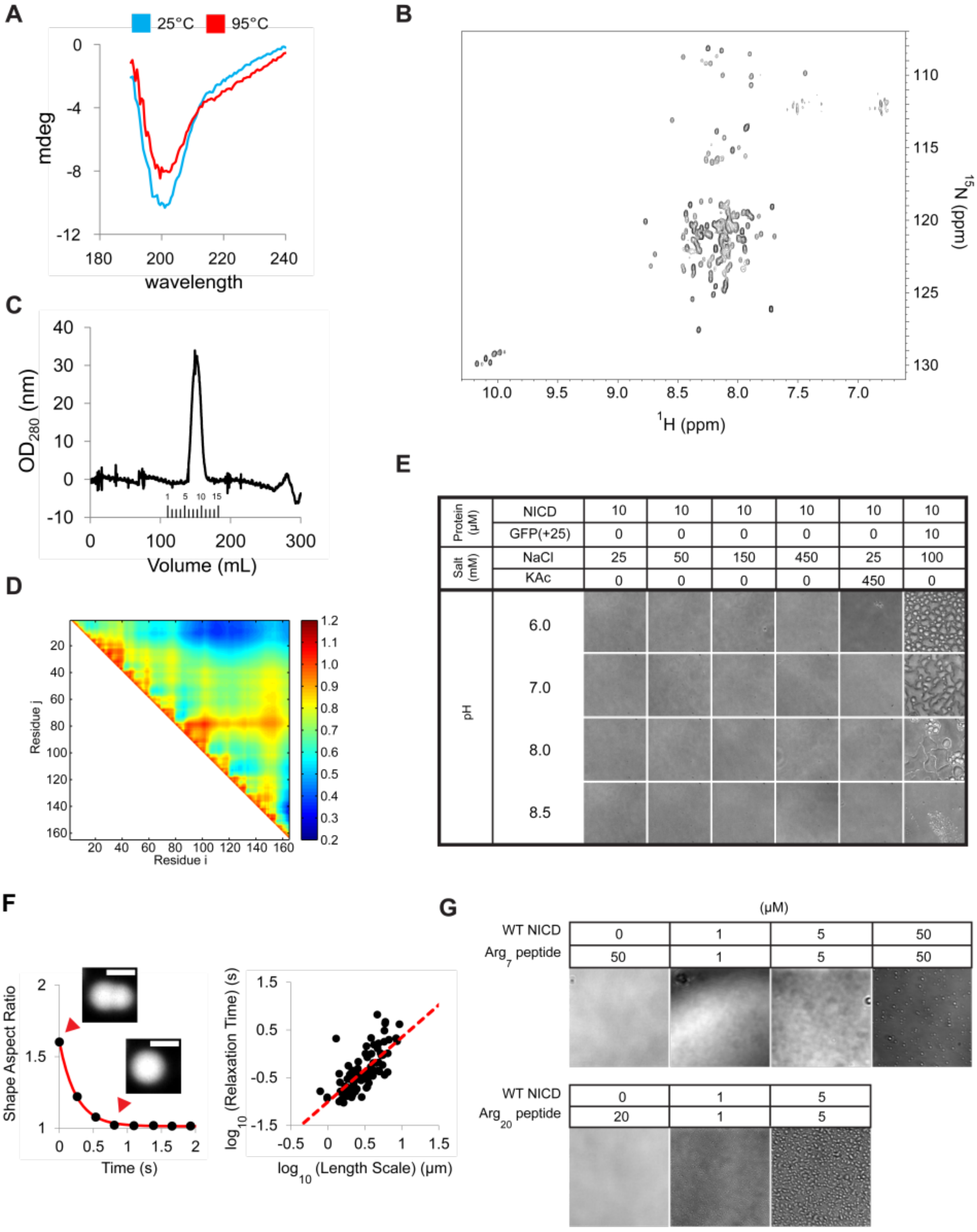
A
25°C
95°C
mdeg
0
-4
-8
-12
180
200
220
240
wavelength
B
110
115
120
125
130
^{15}N (ppm)
10.0
9.5
9.0
8.5
8.0
7.5
7.0
^{1}H (ppm)
C
OD_{280} (nm)
40
30
20
10
0
-10
1 5 10 15
0
100
200
300
Volume (mL)
D
Residue j
Residue i
20 40 60 80 100 120 140 160
1.2 1.1 1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2
E
Protein (μM)
NICD 10 10 10 10 10 10
GFP(+25) 0 0 0 0 0 10
Salt (mM)
NaCl 25 50 150 450 25 100
KAc 0 0 0 0 450 0
pH
6.0
7.0
8.0
8.5
F
Shape Aspect Ratio
2
1.5
1
0 0.5 1 1.5 2
Time (s)
$\log_{10}$ (Relaxation Time) (s)
1.5
0.5
-0.5
-1.5
-0.5 0 0.5 1 1.5
$\log_{10}$ (Length Scale) (μm)
G
(μM)
WT NICD 0 1 5 50
Arg_7 peptide 50 1 5 50
WT NICD 0 1 5
Arg_{20} peptide 20 1 5

Figure S3. Isolated human NICD is intrinsically disordered and monomeric but phase separates with positively charged partners (related to Figure 3)

(A) CD spectra of NICD at 25°C (cyan) and 95°C (red). Only minor changes in the spectra are observed after melting indicating low secondary structure content. (B) $^{1}H/^{15}N$ HSQC spectrum of 87 μM ^{15}N-labeled NICD recorded at 25°C on an 800 MHz NMR spectrometer. The poor chemical shift dispersion suggests a lack of persistent secondary structure. (C) NICD migrates as a single monodisperse peak on a Superdex 75 pg 26/600 gel filtration column, column void volume is approximately 110 mL, near the beginning of the first fraction. (D) Average distance between i^{th} and j^{th} residues of NICD were determined from atomistic Monte Carlo simulations and normalized by the expected distance from a sequence-matched self-avoiding random chain. The normalized distance is indicated by heat map colors. The ensemble does not show well-defined close-range interactions, consistent with a disordered structure. (E) Solutions of isolated recombinant NICD were prepared in different buffers, varying pH (6.0-8.5) and salt conditions (NaCl: 25-450 mM; KAc: 450 mM), and imaged by light microscopy. The solution remained homogenous under all conditions, including at 4°C and 25°C. In 150 mM NaCl and at pH 6.0-8.5, phase separated droplets were observed when supercharged GFP (+25) was added to NICD (rightmost column). (F) Left: representative example of the time course of fusion of phase separated droplets *in vitro* (WT NICD with scGFP(+20)). Right: time constants of relaxation are plotted versus length scale of droplets. The inverse capillary velocity is 0.25 s/μm. (G) Solutions containing equimolar concentrations of NICD (1, 5, and 50 μM) and Arg_7 peptide (1, 5, and 50 μM) (A) or Arg_{20} peptide (1 and 5 μM) (B) were imaged for evidence of phase separation by light microscopy. Both peptides were homogeneously distributed in isolation (50 and 20 μM for Arg_7 and Arg_{20}, respectively), but phase separated with NICD. In both cases, higher concentrations of the two species resulted in more robust phase separation (rightmost panel).

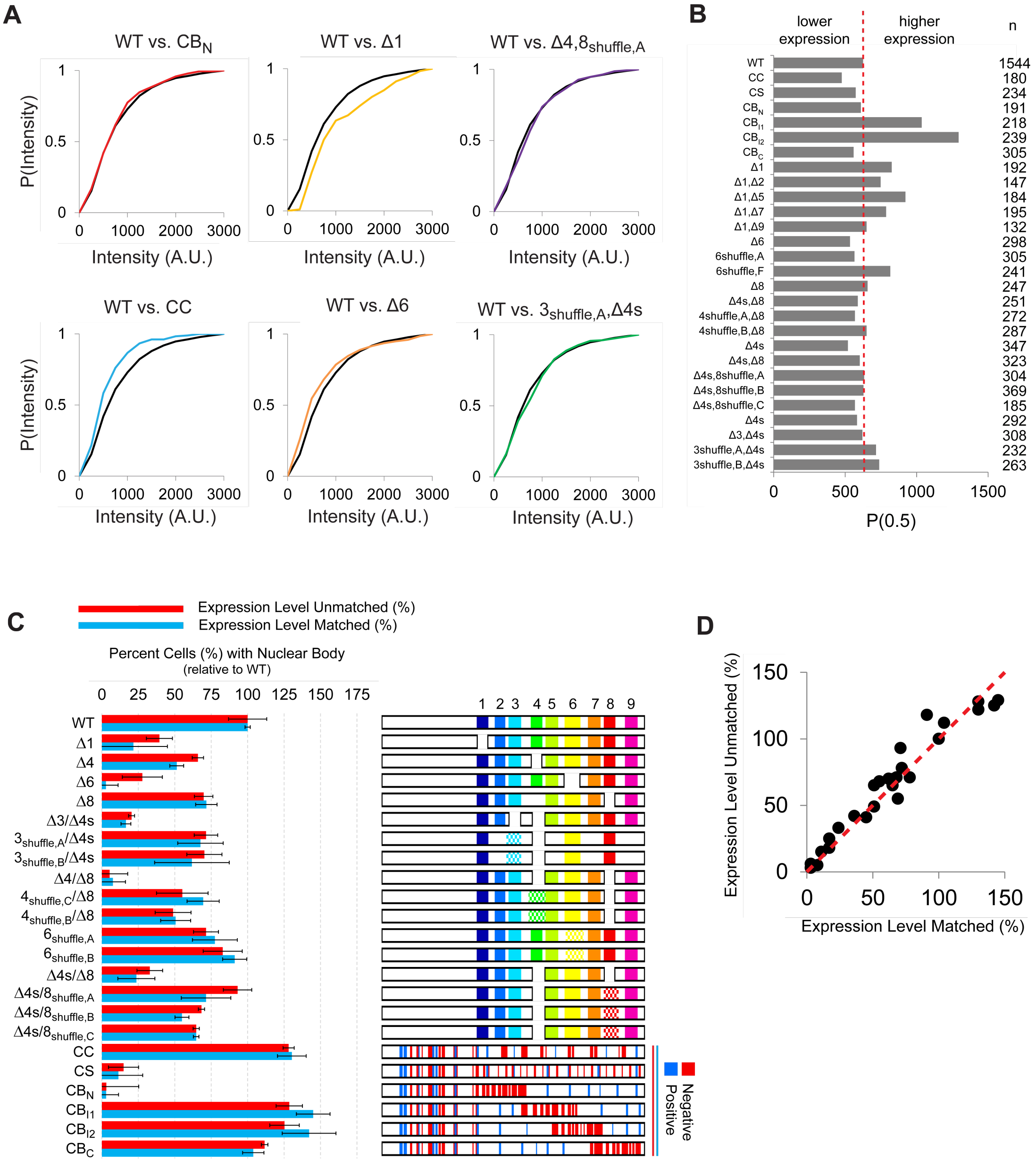

Figure S4. Expression levels of wild type NICD and its mutants are similar and do not affect qualitative conclusions of nuclear body quantification studies (related to Figure 5)

(A) Cumulative frequency distributions (CFDs) of nucleoplasm intensities in HeLa cells expressing representative NICD charge mutants (CC (red) and CB$_N$ (cyan)), deletion mutants (Δ1 (yellow) and Δ6 (orange)), and shuffle mutants (Δ4,8$_{shuffle,A}$ (magenta) and 3$_{shuffle,A}$,Δ4s (green)) or wild type NICD (black) (B) Median nucleoplasm intensity (P(0.5)) for the indicated mutants representing different mutation classes: charge, deletion, and shuffle mutants. The red dashed line indicates the median nucleoplasm intensity of wild type NICD. The total number of cells (n) analyzed for each CFD is listed. (C) Quantification of nuclear body formation in cells expressing NICD mutants unmatched (red) or matched (blue) for expression level. Data are represented as mean ± SEM. (D) Comparison of percentage of cells containing nuclear bodies when expressing different NICD mutants, relative to those expressing wild type NICD. When quantifying cell populations using matched (X-axis) or unmatched (Y-axis) cellular expression levels, no qualitative differences are seen. Red dashed line indicates perfect agreement between the two values.

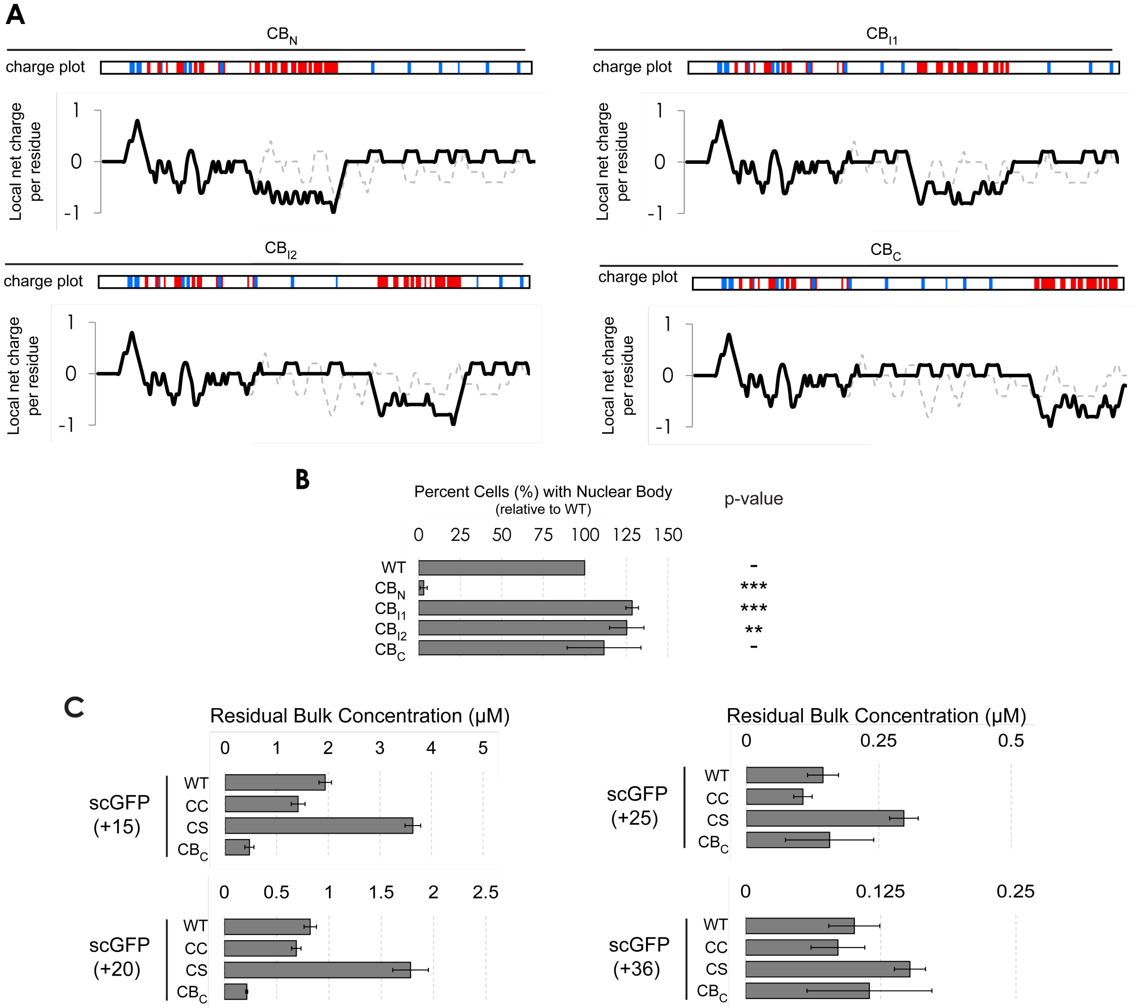

Figure S5. Charge patterning of NICD and net charge of partners affects phase separation in vitro (related to Figure 5)

(A) The positions of negatively (D, E; red) and positively (K, R; blue) charged residues (charge plots) and linear charge density scores are shown for wild type (WT) NICD (gray dashed) and charge mutants (CB$_{N,\ I1,\ I2,\ C}$; black solid). All proteins have the same overall amino acid composition, and thus the same net charge. In CB$_{N,\ I1,\ I2,\ C}$, negatively charged amino acids in residues 61-165 were aggregated into a single block located at different positions in the protein. (B) Quantification of nuclear body formation for NICD charge mutants. Three of the four CB mutants produced an equal or greater percentage of HeLa cells with nuclear bodies than did WT NICD, indicating a single charge block is sufficient. Data are represented as mean ± SEM and p-values for comparison to WT NICD are: * < 0.05, ** < 0.01, *** < 0.001. (C) NICD charge mutants (CC, CS, CB$_C$; 5 μM), with the same net charge but with different charge patterning, were mixed with supercharged GFPs of increasing positive charge (5 μM) and the residual bulk concentrations of scGFP were quantified. Increasing net positive charge of scGFP promotes phase separation. Residual bulk concentration for WT NICD decreases ~20-fold over the range of positive charges tested (+15 to +36). With the same supercharged GFP species, residual bulk concentrations for NICD mutants with higher local charge density (CC and CB) were always lower than or equal to those for WT NICD. The NICD mutant with lower local charge density (CS) behaved oppositely, with higher bulk concentration than WT NICD. Data are represented as mean ± SEM.

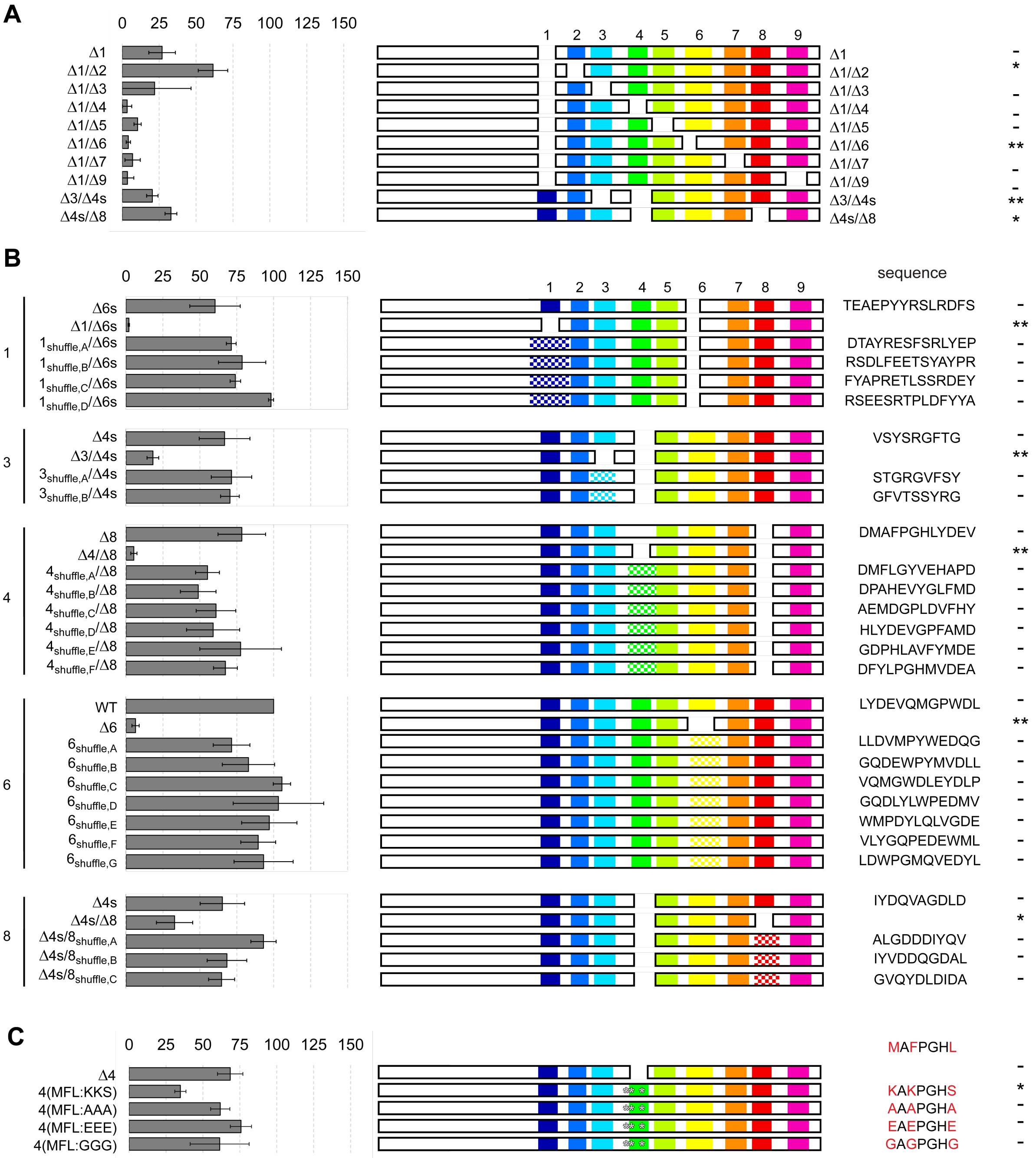

Figure S6. Identification of residue types that promote nuclear body formation (related to Figure 6)

(A) Normalized (to WT) percent of HeLa cells containing nuclear puncta (left) when expressing constructs deleted for multiple sequence elements (schematically illustrated at right). (B) Quantification of nuclear body formation for additional locally shuffled sequences. The positions of shuffled sequences are indicated (checkered region), and the native sequence (top of each group) and shuffled sequence are shown. (C) Quantification of nuclear body formation for additional NICD point mutants, wherein the same residues were mutated to different amino acids. The positions of the mutated residues are indicated (white asterisks), and the native residues and residues they were mutated to are shown. In panels A-C, data are represented as mean ± SEM and p-values for comparison to Δ1 mutant, respective single deletion mutants and WT NICD, and Δ4 mutant respectively are: * < 0.05, ** < 0.01, *** < 0.001.

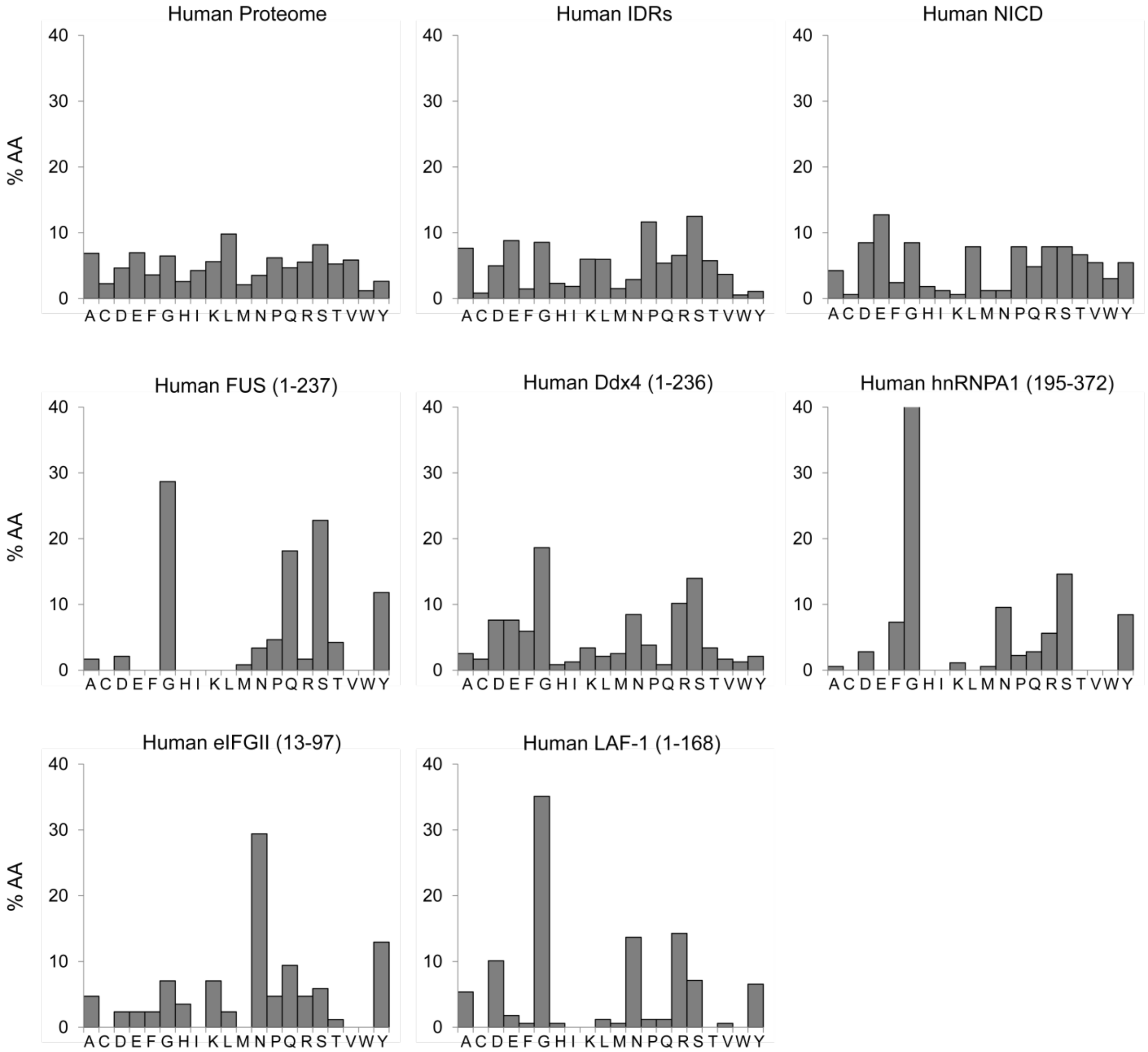

Figure S7. Distribution of amino acids for groups of sequences (related to Figure 6)

The distribution of amino acids (AAs) for various groups of sequences were analyzed, including for the entire human proteome (UniProt: UP000005640_9606), IDRs in the human proteome predicted by the MobiDB consensus prediction, and sundry IDRs (NICD, FUS (1-237), Ddx4 (1-236), hnRNPA1 (195-372), eIFGII (13-97), Laf-1 (1-168)) shown previously to liquid-liquid phase separate. The sequence of NICD is not low complexity and its AA distribution is more typical of the average human IDR.

Supplemental Tables

Table S1. Mean percentage of nuclear bodies observed normalized to wild type NICD, and the number of residues in the [Y/D/R] critical residue set (related to Figure 6)

The sequence of fourteen single and double deletions are shown, highlighting P (magenta), aromatics (orange), R (blue), and D/E (red). Note the critical residue set (CRS) shown is purely to illustrate the methods of the analysis.

Single/Double deletion	**Construct**	**Sequence of deleted residues**	**% of cells with nuclear body**	**# of residues from [Y/D/R] CRS**
Single deletion	Δ1	PYYRSLR	35	4
Single deletion	Δ3	SYSRGFTG	66	2
Single deletion	Δ4	FPGHLYD	68	2
Single deletion	Δ6	LYDEVQMGPWDL	28	3
Single deletion	Δ8	YDQVAGD	70	2
Double deletion	Δ1,Δ2	PYYRSLR, QLPPTQE	61	4
Double deletion	Δ1,Δ3	PYYRSLR, SYSRGFTG	22	6
Double deletion	Δ1,Δ4	PYYRSLR, FPGHLYD	3	6
Double deletion	Δ1,Δ5	PYYRSLR, ERTYPPSG	9	6
Double deletion	Δ1,Δ6	PYYRSLR, QMGPW	4	6
Double deletion	Δ1,Δ7	PYYRSLR, EDTYQDPR	6	8
Double deletion	Δ1,Δ9	PYYRSLR, DSLPFELR	3	6
Double deletion	Δ3,Δ4s	VSYSRGFTG, PGHLYDEV	20	4
Double deletion	Δ4s,Δ8	PGHLYDEV, YDQVAGD	28	4

Table S2. Examples of the Monte Carlo subsampling procedure, where each iteration (SS^1, SS^2, etc.) represents a random Monte Carlo selection (related to Figure 6)

Specific values for each construct are sampled from the "Raw data" set to generate 2000 randomly constructed datasets.

Construct	**Raw data**	**SS^1**	**SS^2**	**SS^3**	**…**	**SS^{2000}**
Δ1	[A,B,C,D,E,F]	C	A	E		D
Δ3	[G,H,J]	H	J	J		G
Δ4	[K,L]	L	L	K		L
Δ6	[M,N,P]	M	P	N		P
Δ8	[Q,R,S,T,U]	S	Q	R		S

Table S3. NICD-like proteins (n=443) in the human proteome, identified based on similar charge properties and charge interaction elements (related to Figure 6 and Discussion)

See accompanying spreadsheet.

Table S4. Enriched molecular functions and cellular component (GO) of NICD-like proteins (n=443) in the human proteome, identified based on similar charge properties and charge interaction elements (related to Figure 6 and Discussion)

GO molecular functions analysis implicates NICD-like proteins in nucleic acid binding. GO cellular components analysis annotates NICD-like proteins as being components of nuclear complexes and localized in nuclear body structures, including the nucleolus.

	Annotation	**Fold Enrichment**	**P value**
GO (molecular function)	Histone binding (*GO:0042393*)	> 5	1.40E-04
	Chromatin binding (*GO: 0003682*)	4.28	2.26E-02
	Poly(A) RNA binding (*GO: 0044822*)	2.30	2.80E-05
	RNA binding (*GO: 0003723*)	2.19	8.19E-05
	Nucleic acid binding (*GO: 0003676*)	2.04	3.30E-06
GO (cellular component)	Histone methyltransferase complex (*GO:0035097*)	>5	9.00E-03
	Methyltransferase complex (*GO:0034708*)	>5	5.76E-03
	Nucleolus (*GO:0005730*)	3.05	9.65 E-07
	Transferase complex (*GO:1990234*)	2.83	2.37E-03
	Nucleoplasm part (*GO:0044451*)	2.79	4.65E-03

Table S5. Enriched biological processes (GO) of NICD-like proteins (n=260) in the human proteome, identified based on similar charge properties and charge interaction elements and in which tyrosine and/or leucines represent >=6% of residues in the region (related to Figure 6 and Discussion)

GO molecular functions analysis implicates NICD-like proteins in regulation of nucleic acid biosynthesis. GO cellular components analysis annotates NICD-like proteins as being components of nuclear complexes and localized in nuclear body structures, including the nucleolus.

	Annotation	**Fold Enrichment**	**P value**
GO (Biological process)	Negative regulation of transcription from RNA polymerase II promoter (*GO:0000122*)	3.98	3.03E-02
	Negative regulation of transcription, DNA-templated (*GO:0045892*)	3.06	2.72E-02
	Negative regulation of nucleic acid-templated transcription (*GO:1903507*)	3.01	3.51E-02
	Regulation of RNA biosynthetic process (*GO:2001141*)	2.30	1.14E-02
	Regulation of cellular macromolecule biosynthetic process (*GO:2000112*)	2.22	1.33E-02
GO (cellular component)	Histone methyltransferase complex (GO:0035097)	> 5	1.15E-02
	Methyltransferase complex (*GO:0034708*)	>5	3.63E-02
	Nucleolus (*GO:0005730*)	2.76	3.97E-02
	Nucleoplasm (*GO:0005654*)	2.04	8.98E-04
	Nuclear lumen (*GO:0031981*)	2.04	3.28E-05

Table S6. Sequences of constructs used in this study (related to Experimental Procedures)

See accompanying spreadsheet.

Supplemental Experimental Procedures

Molecular cloning

Human wild type and mutant (Y3F) NICD sequences were amplified from pEBB plasmids encoding CD16/7-Nephrin(WT)-GFP or CD16/7-Nephrin(Y3F)-GFP (kind gift from Nina Jones and Tony Pawson) using KOD polymerase (EMD Millipore, Billerica, MA, USA). NICD PCR fragments were inserted into pEGFP-N1 using XhoI and KpnI (New England Biolabs (NEB), Ipswich, MA, USA). NICD single and double deletion mutants were generated by amplifying WT or single-deletion plasmids, respectively, minus the region-of-interest using KOD Xtreme polymerase (EMD Millipore). Ends of PCR products were phosphorylated using T4 PNK (3' phosphatase minus; NEB), and self-ligated using T4 DNA ligase (NEB). NICD point or shuffle mutants were generated by the same method described above except primers containing mutations or the shuffled sequence, respectively, were used.

For recombinant NICD, codon-optimized human wild type NICD sequence with a C-terminal His8x tag was synthesized by Genscript (Genscript, Piscataway NJ, USA), and inserted into a modified pMAL-C2 vector (2MBP-Tev-NdeI-BamHI) using NdeI and BamHI (NEB). The modified pMAL-C2 contains two N-terminal tandem MBP domains, which can be cleaved from the fused protein using Tev protease. A Tev cleavage site was also added after the NICD sequence and before the His_8 tag, permitting the removal of both tags with Tev protease. Mutant NICD sequences (charge: CC, CS, CB_C) were inserted into the same modified pMAL-C2 vector using NdeI and BamHI (NEB). Restriction sites were added by PCR. Plasmids encoding supercharged GFPs were kind gifts from David Liu.

Cell culture and transfection

HeLa cells were cultured in DMEM (Invitrogen, Carlsbad, CA, USA) supplemented with 10% FBS (Invitrogen), 1X GlutaMAX (Thermo Fisher, Waltham, MA, USA), and 100 U/g pen/strep (Thermo Fisher). NEAT1 -/- and +/+ MEFs were cultured in 1:1 mixture of DMEM (Invitrogen) and Ham's F-12 (Invitrogen) supplemented with 10% FBS, 1X GlutaMAX, and 100 U/g pen/strep. For all imaging experiments, cells were plated onto glass bottom 35 mm dishes (MatTek, Ashland, MA, USA). Transfections were performed using Lipofectamine 2000 (Thermo Fisher) and 1.6 µg DNA per dish 12-16 h after plating.

Cellular FRAP imaging and analysis

10-12 h after transfection with wild type NICD, HeLa cells were switched to live cell imaging medium [Leibovitz L-15 medium (Thermo Fisher) supplemented with 10% FBS (Invitrogen), 1X GlutaMAX (Thermo Fisher), and 100 U/g pen/strep (Thermo Fisher)] and equilibrated on the heated (37°C) stage for 10-15 min. Nuclear bodies were imaged using a 63x objective (1.4 NA) on a Zeiss LSM 510 (Carl Zeiss, Jena, Germany) laser scanning confocal microscope. Imaging settings were set as follows: 488 nm laser power = 25%, excitation transmission = 1%, 12-bit image depth, 5x zoom (55.8×55.8 nm pixel size), 100×100 pixel scan region, scan speed = 7, 500 ms interval. Photobleach settings were set as follows: excitation transmission = 100%, 4 iterations, 12×12 circular region. Custom ImageJ/FIJI scripts were used to analyze fluorescence recovery after photobleaching. The photobleached region was empirically determined by image subtraction ($Image_{pre} - Image_{post}$ photobleaching). The intensity within the photobleached region was measured at each time point and normalized to the initial intensity ($I_{t=0}$). Data are represented as mean ± SEM.

Timelapse imaging and analysis of nuclear body fusion and formation

To image nuclear body fusion, HeLa cells were switched to live cell imaging medium 10-12 h after transfection with wild type NICD and equilibrated on a heated (37°C) stage for 15-30 min. Cells were imaged using a 63x objective (1.4 NA) on a Zeiss LSM 510 (Carl Zeiss) laser scanning confocal microscope. Imaging settings were set as follows: 488 nm laser power = 25%, excitation transmission = 1%, 8-bit image depth, 5x zoom (55.8×55.8 nm pixel size), 150×150 pixel scan region, scan speed = 7, 10 s image interval. Volumes of nuclear bodies were estimated by measuring their diameter (FWHM) before and after fusion. The predicted volume was estimated by summing the volumes of the two individual nuclear bodies that fuse.

To image nuclear body formation, HeLa cells were switched to live cell imaging medium 6 h after transfection with wild type NICD and equilibrated on a heated (37°C) stage for 30 min. To adequately track cells, Z-stacks (36 μm, 1 μm interval) were collected every 30 min (to minimize photobleaching) using a 63x objective (1.42 NA) on a Deltavision RT (GE Healthcare, Little Chalfont, United Kingdom) widefield microscope equipped with a Coolsnap HQ camera (Photometrics, Tucson, AZ, USA). The time point preceding the first instance of a nuclear body was set to t=0, and maximum intensity projections of Z-stacks are shown.

High-resolution 3D imaging of nuclear bodies and image analysis

10 h after transfection, HeLa cells were fixed for 15 min/RT using 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA) in the following fixation buffer: 10 mM MES (pH 6.1), 138 mM KCl, 10 mM EGTA (pH 8.0), 3 mM $MgCl_2$. Fixed cells were washed and maintained in PBS until imaging. Z-stacks of the entire cell (0.1 μm step) were collected using a 60x objective (1.4 NA) on an Andor (Andor, Belfast, Northern Ireland) spinning disc confocal microscope (Nikon Ti Eclipse microscope base; Nikon, Tokyo, Japan) equipped with a Yokogawa CSU X (Yokogawa Electric, Tokyo, Japan) and an Andor NEO sCMOS camera (Andor). All images were background subtracted (using an image of a solution of phosphate buffered saline (PBS)), and flatfield corrected (using an image of recombinant enhanced green fluorescent protein (EGFP)). Cellular autofluorescence was measured by imaging untransfected HeLa cells using the same imaging parameters as above, and determined to be negligible. A custom semi-automated ImageJ script was used to detect nuclear bodies, determine their centroid in 3D, and measure their full-width half maximum (FWHM) diameter and maximum intensity.

Volume fraction, nuclear body intensity, molar fraction, and partition coefficient

The volumes of nuclear bodies were calculated by assuming a spherical shape. For smaller nuclear bodies, which were below the transverse resolution limit (i.e., diffraction-limited; ~550 nm), their diameter could not be directly measured. Instead, we indirectly estimated their diameter relying on a calibration that relates the effect of the empirical point spread function (PSF) on the apparent intensity of computer-generated spheres (Fink et al, 1998). For the calibration, ideal spheres of the same intensity but different diameters were generated in ImageJ and convolved using an empirically measured point spread function. The point spread function was measured using sub-diffraction sized beads (200 nm) resuspended in PBS using the same spinning disc confocal microscope (Andor) and imaging parameters as described above. The total nuclear volume for each cell was also measured. Volume fraction was calculated as $(V_{total,NBs}) / (V_{total,nucleus})$.

Due to diffraction, the apparent size and intensity of an object can be inaccurate below or near the diffraction-limit. This is particularly true for apparent intensity since axial resolution is considerably worse than transverse resolution. To illustrate, even for a computer-generated sphere as large as 2 μm in diameter, which is larger than any NICD body measured, its apparent intensity is still reduced by the PSF. To account for the effect of the PSF, we used the same calibration as above to apply an intensity correction factor that depends on nuclear body size. Therefore, for this analysis, we limited our dataset to nuclear bodies slightly larger than the transverse resolution limit (diameter > 550 nm). The mean intensity of nuclear bodies ($I_{mean,NBs}$) was estimated as the average value of nuclear body intensities after PSF correction.

Molar fraction was calculated as $[(V_{total,NBs}) * (I_{mean,NBs})] / [(V_{total,nucleus}) * (I_{mean,nucleoplasm})]$. Aggregate volume of nuclear bodies ($V_{total,NBs}$), PSF-corrected mean intensity of nuclear bodies ($I_{mean,NBs}$), total nuclear volume

($V_{total,nucleus}$), and mean intensity of the nucleoplasm ($I_{mean,nucleoplasm}$) were measured as described above. Partition coefficients for each nuclear body were calculated as ($I_{mean,NBs}$) / ($I_{mean,nucleoplasm}$, in cell where nuclear body originated from).

Atomistic simulations of NICD

All simulations were performed using the CAMPARI Monte Carlo (MC) modeling suite (http://campari.sourceforge.net), which uses the ABSINTH implicit solvent model and force field paradigm (Vitalis and Pappu, 2009). CAMPARI implements a range of MC moves that explore the proteins' degrees of freedom, rotation and translation of full protein chains and ions, and explicit consideration of the conformational degrees of freedom coupled to proline ring systems (Radhakrishnan et al., 2012). Each simulation was initiated from a random, non-overlapping starting conformation. Under ABSINTH, protein atoms and mobile solution ions are modeled with atomic detail, whereas the solvent is treated using a mean field representation. Simulations utilize parameters from the abs_3.2_opls.prm parameter file for the interactions of polypeptide atoms and the parameters of Mao and Pappu (Mao and Pappu, 2012) for solution Na^+ and Cl^- ions. The dielectric constant for water was set to 78.2. Spherical cutoffs of 10 and 14 Å were used to truncate and taper the Lennard-Jones and electrostatic interactions, respectively.

Single protein simulations used temperature replica exchange Monte Carlo (T-REMC) to enhance the quality of simulation sampling. T-REMC used the following temperature schedule: 280 K, 285 K, 289 K, 294 K, 298 K, 310 K, 320 K, 330 K, 340 K, 355 K, 370 K, 385 K, 400 K. Temperature swaps were attempted between adjacent temperatures every 50,000 MC steps. All analysis was performed on conformations from the 310 K ensemble. T-REMC simulations consisted of 61 million MC steps per temperature, with the first million discarded as equilibration. Each independent T-REMC simulations took around three weeks of wall-clock time. Dimer simulations were run at a single temperature (310 K), and consisted of 70 million MC steps per replica, with the first ten million discarded as equilibration, and took around five weeks of wall-clock time per independent replica.

Simulation analysis was performed using in-house analysis tools, MDTraj (McGibbon et al., 2015) and routines built into CAMPARI. Figures were generated using MATLAB.

Quantification of nuclear body formation

HeLa cells were switched to live cell imaging medium 10-12 h after transfecting with indicated NICD expression plasmids and equilibrated on a heated (37°C) stage for 15-30 min. Transfection-positive cells within a grid (700×700 µm) were imaged. Z-stacks (25 µm, 1 µm interval) were collected using a 63x objective (1.42 NA) on a Deltavision RT (GE Healthcare) widefield microscope equipped with a Coolsnap HQ camera (Photometrics). Images were background subtracted (rolling ball algorithm; 50 µm). Transfected cells above a minimal threshold of intensity (A.U. > 250) were evaluated for the presence of nuclear bodies. For some constructs, the nucleoplasm intensity (in a region adjacent to the nuclear body) was also measured to compare nuclear body formation at matched expression levels. Differences between groups were determined using an unpaired Student's *t* test and *p* values are shown. Data are represented as mean ± SEM of 2-12 independent experiments.

Statistical analysis of critical residues

The contributions of specific residue types to nuclear body formation were determined by assessing how well the loss of unique combinations of residues (critical residue sets, CRS) correlated with changes in nuclear body formation. For each CRS, the percentage of cells that formed nuclear bodies was plotted against the number of CRS residues lost with each deletion construct, and Pearson's correlation coefficient were determined. An example of values using the CRS, [Y/D/R], is shown in Table S1. To account for variance in the experimental data in a non-parametric manner, subsampled datasets (n=2000) were constructed through Monte Carlo subsampling of experimental values (i.e., percentage of cells that formed nuclear bodies) from individual experiments (each deletion construct was assayed between 2 and 12 times). A simplified example of the Monte Carlo subsampling process is shown in Table S2. A distribution of correlation values (n=2000) were determined using subsampled datasets and significant differences between different CRSs were evaluated using unpaired Student's *t* test. Highly correlated CRSs were further evaluated to determine which residues, if any, were consistently enriched relative to a normalized background.

Definition of charge interaction elements (CIEs)

Sequence analysis for identifying CIEs was performed using localCIDER (Holehouse et al., 2015). The parameters used to define charge interaction elements were based on previous work in polymer and polyampholyte physics. A five residue window corresponds to the number of residues beyond which the combined balance of chain-chain and chain-solvent interaction energy is on the order of thermal fluctuations (kT) (Dobrynin and Rubinstein, 1995; Pappu et al., 2008). This length-scale is also referred to as a blob. A charge threshold of -0.35 corresponds to the net-charge per residue limit at which a polymer enters the strong (negative) polyelectrolyte regime on the Das plot (Das and Pappu, 2013). Finally, given a sliding window of 5 and a sliding step-size of 1, a length cutoff of four residues or longer corresponds to the length-scale at which a region will have a net charge between -0.2 and -0.4 and remain equal to or greater than the blob length scale, given the appearance of flanking residues around the element, which by definition must have a net charge of -0.2. A length cutoff of three would be too lenient, while a length cutoff of five would allow for negatively charged blob-sized regions to remain unidentified as CIEs. A cutoff of 4 residues offers an ideal compromise.

Recombinant protein expression and purification

Recombinant proteins were expressed in *E. coli* BL21(DE3)T1^R. Bacteria were grown in Luria Broth (LB; Thermo Fisher) at 37°C until culture density reached optical density at 600 nm of 0.6-0.8. For recombinant NICD and its mutants, cultures were incubated with 4 g/L dextrose and antibiotic (100 mg/L Amp; Thermo Fisher) and induced with 1 mM IPTG (Gold Biotechnology, Olivette, MO, USA) at 37°C for 3-4 h. For supercharged GFP mutants, cultures were incubated with antibiotic (100 mg/L Amp; Thermo Fisher) and induced with 1 mM IPTG (Gold Biotechnology) at 30°C for 4 h. Bacteria were lysed by homogenization (EmulsiFlex-C5; Avestin, Ottawa, Canada) and then centrifuged at ~30,000 x *g* for 30-40 min at 4°C.

For recombinant NICD and its mutants, pelleted cells were resuspended in 20 mM Tris/HCl (pH 8.0, at 22°C), 20 mM NaCl, 2 mM EDTA (pH 8.0) (Sigma-Aldrich, St. Louis, MO, USA), 2 mM DTT (Gold Biotechnology), and supplemented with PMSF to 1 mM before snap freezing in liquid N_2. Cells were thawed in cool H_2O and lysed as described above. Soluble NICD was purified using Amylose High Flow (NEB) affinity chromatography, followed by Nickel Sepharose Fast Flow (GE Healthcare) affinity chromatography, and cleaved with Tev protease overnight at 4°C. Proteins were then further purified using SOURCE 15Q anion exchange chromatography (GE Healthcare) packed in a ~4 mL, 10 mm × 50 mm (width × height) column. The Tev reaction was diluted bound to the SOURCE 15Q column, washed, and eluted with a resolving gradient of 20 mM imidazole (pH 7.0) (Sigma-Aldrich), 1 mM DTT (Gold Biotechnology), and NaCl varying from 150 mM to 350 mM, developed over 25 column volumes (CVs). For NICD mutants, optimized SOURCE 15Q anion exchange gradients were used and developed over 30 CVs: 150-600 mM NaCl for CC, 30-300 mM NaCl for CS, and 200-1600 mM NaCl for CB_C. Fractions were analyzed by SDS-PAGE, pooled, and passed over a gel filtration column (Superdex 75pg) into a final buffer of 10 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 1 mM DTT. Protein identity was confirmed by mass spectrometry (expected mass = 20,174.78, observed mass = 20,174.95 Da), concentrated to ca. 100 μM using ultrafiltration (Amicon Ultra-3; EMD Millipore). Concentration was determined from the absorbance at 280 nm using a Nanodrop spectrophotometer and extinction coefficient of 42,400 M^{-1} cm^{-1}.

Supercharged GFP mutants were lysed (as above) in high salt buffer: 2.0 M NaCl, 20 mM $NaPO_4$ (pH 7.4), 20 mM imidazole (pH 8.0) (Sigma-Aldrich), 5 mM BME (Sigma-Aldrich), and protease inhibitor cocktail. They were then purified with Ni-NTA resin and eluted in high salt buffer (2.0 M NaCl, 20 mM $NaPO_4$ (pH 7.4), 500 mM imidazole pH 8.0 (Sigma-Aldrich), and 5 mM BME) for scGFP mutants >= +20 and in medium salt buffer (0.5 M NaCl, 20 mM $NaPO_4$ (pH 7.4), 500 mM imidazole pH 8.0, and 5 mM BME) for scGFP mutants < +20. Eluate was diluted with no salt buffer (20 mM NaPO4 (pH 6.0) and 1 mM DTT) to reduce the salt concentration to 100-250 mM, and purified with a Source 15S column using 25 mM $NaPO_4$ pH 6.0, 1 mM EDTA pH 8.0, 1 mM DTT (Buffer

A) and plus 1 M NaCl (Buffer B). Relevant protein fractions were further purified with a Superdex 75 column equilibrated in 10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM EDTA pH 8.0, and 1 mM DTT. Proteins were concentrated via ultrafiltration (Amicon Ultra-10; EMD Millpore). The concentrations of all proteins were measured from the absorbance at 280 nm using a Nanodrop spectrophotometer (Thermo Fisher) and extinction coefficients were calculated using ExPASy ProtParam (Artimo et al., 2012).

Labeling of NICD with Alexa Fluor 568

NICD, which contains a sole cysteine residue, was labeled with Alexa Fluor 568 maleimide (Thermo Fisher) following the purification protocol described above but with the following modifications. Following Tev cleavage, 5mM DTT was added to NICD, which was then diluted and purified using SOURCE 15Q anion exchange chromatography using with the same gradient described above but using buffers without DTT. Pooled peak fractions were quantified by UV absorption (as above) and labeled with a four-fold excess of fresh Alexa Fluor 568 maleimide dye (DMSO) for 10 min at room temperature. The labeling reaction was quenched with a five-fold excess of DTT. High efficiency labeling was confirmed by a subtle, but complete, shift in SDS-PAGE mobility, and further purified using Superdex 75 gel filtration chromatography using buffers described above. The concentration of Alexa Fluor 568 labeled NICD was determined using extinction coefficient at 578 nm of 88,000 M^{-1} cm^{-1}.

Preparation of Arg peptides

Two custom synthetic peptides, CR7 (sequence = CRRRRRRR), and CR20 (sequence = CRRRRRRRRRRRRRRRRRRRR), were synthesized by GenScript (NJ, USA). After solid phase synthesis, peptides were HPLC purified to >98% purity. Known amounts of peptide (by mass) were lyophilized in separate containers. Dry 1 mg aliquots of peptide were dissolved in 10 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 1 mM DTT and used the same day as required.

Imaging of isolated recombinant proteins

To prevent non-specific binding of proteins to well surfaces, polystyrene 384-well plates (Sigma-Aldrich) were pretreated with 1% Tween-20 (Sigma-Aldrich) for 30 min to 1 h and then washed before adding protein samples. NICD (WT, 10 μM) or supercharged GFP (+25, 10 μM), prepared in different buffers varying salt (NaCl: 25, 50, 150, and 450 mM; KAc: 450 mM) and pH (6.0, 7.0, 8.0, and 8.5) conditions, were then directly added to each well. To prevent evaporation, wells were sealed with clear sealing tape (Thermo Fisher) and the solution was incubated overnight in the dark. After overnight incubation, Z-stacks (50 μm at 3 μm intervals) of brightfield images were collected using a 20x objective (0.4 NA) on a custom spinning disc confocal microscope (Biovision Technologies, Chester Springs, PA, USA; Leica DMI6000B microscope base, Leica, Wetzlar, Germany) equipped with a Yokogawa CSU XI (Yokogawa Electronics) and a PCO Edge 4.2 sCMOS camera (PCO, Kelheim, Germany). Phase separation was determined by the presence of droplets.

2D phase diagram imaging and analysis

Polystyrene 384-well plates (Sigma-Aldrich) were prepared as described above with slight modifications. Before pretreating with 1% Tween-20, wells were washed with ultrapure H_2O and EtOH, and then dried with argon gas. Different concentrations of NICD and scGFP were added to a final volume of 25 μL in 10 mM HEPES (pH 7.5) and 150 mM NaCl, and solutions were incubated overnight as described above. After overnight incubation, Z-stacks (50 μm at 3 μm intervals) of fluorescent images at four different positions were collected using the custom spinning disc confocal microscope described above.

Maximum projection images of Z-stacks were prepared before background subtracting using a rolling ball algorithm with a radius ~25-fold the PSF. Images were segmented using the Triangle threshold method, and particles were identified using "Analyze Particles…" (circularity of 0.4-1.0 and area of particles >5-fold the FWHM of the PSF) in FIJI. Images were scored automatically for the presence of phase separated droplets using custom Python scripts, although images with obvious inclusions were manually rejected. The latter set of images represented 2-5% of each dataset. A well was scored as positive for phase separation if at least 2 of the 4 images acquired from the well satisfied the following criteria: the most intense pixel value was larger than a threshold value (250 A.U. with ImagEM X2 EMCCD camera, 1000 A.U. with PCO Edge 4.2 sCMOS camera), and either one particle was 5-fold larger than a diffraction-limited object or at least two particles were 4-fold larger than a diffraction-limited object. Images were also scored positive if the average intensity was significantly greater than the background, indicating that a large fraction of the image area was covered by the second phase. Scores from automatic scoring methods were compared to blind scores (i.e., image names were reassigned randomly) by an individual, which agreed for >95% of images. In general, greater sensitivity to small and/or dim signals in blind scoring resulted in disagreements between the two scoring methods. At least two data sets were collected for each pair of proteins and data shown are representative.

Residual bulk concentration analysis

To prevent non-specific binding of proteins to the sidewall of Eppendorf tubes, 1.5 mL Eppendorf tubes were pretreated with 1% Tween-20 (30 µL) (Sigma-Aldrich) for 15 min and washed before adding protein samples. NICD (WT or mutants, 5 µM) and supercharged GFPs (5 µM) were then added to a final volume of 25 µL and the solution was incubated for 5 min/RT. In experiments where the residual bulk concentration of NICD was measured, Alexa 568-labeled NICD (0.5 µM, 10% of total) was added. The condensed phase was pelleted by centrifugation at 10,000 x *g* for 1 min at 4°C. The supernatant (bulk phase, ~20-25 µL) was transferred to polystyrene 384-well plates (Thermo Fisher) and Z-stacks (75 µm at 10 µm intervals) were collected using a 20x objective (0.4 NA) on a custom spinning disc confocal microscope (Biovision Technologies; Leica DMI6000B microscope base) equipped with a Yokogawa CSU XI (Yokogawa Electronics) and a PCO Edge 4.2 sCMOS camera (PCO). For each experiment, calibration standards of known concentrations of either Alexa 568-labeled NICD or supercharged GFPs were also prepared and imaged. All images were background subtracted (using an image of buffer only) and flatfield corrected (using an image of WT EGFP) before preparing maximum projection images of Z-stacks. The average intensities of maximum projection images were measured and residual bulk concentrations were calculated using calibration curves. Data are represented as mean ± SEM of three independent experiments.

Droplet intensity imaging and analysis

To prevent non-specific binding of proteins to well surfaces, polystyrene 384-well plates were pretreated with 1% Tween-20 (30 µL) (Sigma-Aldrich) for 15 min and then washed before adding protein samples. NICD (WT or mutants, 5 µM) and supercharged GFPs (5 µM) were then directly added to each well to a final volume of 25 µL. To prevent evaporation, wells were sealed with clear sealing tape and the solution was incubated overnight in the dark. In experiments where the droplet intensity of NICD was measured, different amounts of Alexa 568-labeled NICD (0.01 µM, 0.005 µM, 0.001 µM, 0.0005 µM) were added. After overnight incubation, Z-stacks (50 µm at 3 µm intervals) were collected using a 20x objective (0.4 NA) on a custom spinning disc confocal microscope (Biovision Technologies; Leica DMI6000B microscope base) equipped with a Yokogawa CSU XI (Yokogawa Electronics) and a PCO Edge 4.2 sCMOS camera (PCO). All images were background subtracted (using an image of buffer only) and flatfield corrected (using an image of WT EGFP) before preparing maximum projection images. The diameters (in µm) and maximum intensities of droplets were measured. The intensities of droplets whose diameter were between 10.2 and 12.9 µm were used calculate the average droplet intensity for each combination of NICD and scGFP. Droplet intensity values for each combination of NICD and scGFP were normalized to the average droplet intensity value of WT NICD + scGFP(+15). Data are represented as mean ± SEM of three independent experiments.

In vitro FRAP imaging and analysis

Phase separated droplets were prepared in polystyrene 384-well plates (Thermo Fisher) as described above. After overnight incubation, droplets were photobleached using a 450 nm laser and imaged (1 h total at 1 min intervals) using a 20x objective (0.4 NA) on a custom spinning disc confocal microscope (Biovision Technologies; Leica DMI6000B microscope base) equipped with a Yokogawa CSU XI (Yokogawa Electronics) and a PCO Edge 4.2 sCMOS camera (PCO). The photobleached region was empirically determined by image subtraction ($\text{Image}_{\text{pre}}$ – $\text{Image}_{\text{post}}$ photobleaching). The intensity within the photobleached region was measured at each time point and normalized to the initial intensity ($I_{t=0}$). Data are represented as mean ± SEM of three independent measurements.

Bioinformatics analysis

Disordered regions in the human proteome (Uniprot: UP000005640_9606; (Consortium, 2014)) were analyzed based on the consensus prediction of MobiDB 2.0. All sequences were fetched with MobiDB's web API. MobiDB 2.0 combines ten different disorder predictors and provides a majority vote on a per-residue basis to assign a 'disordered' or 'ordered' prediction. To reduce sequence fragmentation, these data were post processed by converting any region – disordered or ordered – shorter than four residues to the identity of its surrounding regions. The disordered regions identified with MobiDB 2.0 were used to analyze the distribution of amino acids reported in Figure S7.

NICD-like proteins were identified using the following criteria to further filter disordered regions. All regions shorter than 100 residues and with an FCR < 0.25 were discarded. NICD-like protein sequences were required to have at least twice as many negatively charged residues as positively charged residues and at least twenty residues encompassed in CIEs (as defined previously). As an additional filter, any sequence with less than 6% tyrosine + leucine content was discarded.

Gene Ontology molecular function over-representation was performed on the PANTHER webserver (http://pantherdb.org/) using Release 2015-08-06. All sequence analysis was performed using the Python analysis framework, localCIDER (Holehouse et al., 2015), and the CIDER webserver. All additional sequence analysis was performed with Python.

Immunostaining and antibodies

HeLa, NEAT1 -/- MEFs, or NEAT1 +/+ MEFs were cultured and plated as described above. 12-20 h after transfection, cells were fixed for 10 min at room temperature using 4% paraformaldehyde (Electron Microscopy Sciences) in the following fixation buffer: 10 mM MES (pH 6.1), 138 mM KCl, 10 mM EGTA (pH 8.0), 3 mM $MgCl_2$. Fixed cells were washed with PBS, permeabilized with 0.2% Triton-X 100 in PBS or cold (-20°C) MeOH, washed with PBS again, and blocked with 1% bovine serum albumin (BSA; Sigma-Aldrich) in PBS. The following primary antibodies were diluted in 1% BSA in PBS and used at the indicated dilutions for 1.5 h/RT: anti-NONO (Santa Cruz, sc-376865, 1:50; Santa Cruz Biotechnology, Dallas, TX, USA), anti-coilin (Santa Cruz, sc-32860, 1:50), anti-gemin2 (Santa Cruz, sc-32806, 1:50), anti-19S (Enzo Life Science, BML-PW8965, 1:100; Enzo Life Science, Farmingdale, NY, USA), anti-gamma-H2AX (Bethyl Labs, IHC-00059, 1:100; Bethyl Laboratories, Montgomery, TX, USA), anti-MPM-2 (EMD Millipore, 05-368, 1:100), anti-CCDC55 (Sigma-Aldrich, HPA015603, 1:200), anti-PML (Santa Cruz, sc-5621, 1:50), anti-Sam68 (Santa Cruz, sc-333, 1:100), anti-SFPQ (Sigma-Aldrich, SAB4200501, 1:100). Secondary antibodies were diluted in 1% BSA in PBS and used at the indicated dilutions for one hour at room temperature : goat anti-rabbit Alexa488 (Invitrogen, 1:400) and goat anti-mouse Alexa647 (Invitrogen, 1:400). Cells were washed with PBS for 45 minutes at room temperature after incubation with the primary and secondary antibody. Cells were mounted with Prolong Gold Antifade with DAPI (Invitrogen) overnight.

Circular dichroism

NICD in 10 mM HEPES (pH 7.4) (Sigma-Aldrich), 150 mM NaCl, 1 mM EDTA pH 8.0, and 1 mM DTT (Gold Biotechnology) was exchanged into 10 mM $NaPO_4$ (pH 7.5), 100 mM NaF, and 2.5 mM BME by repeated concentration and dilution with a PVDF (0.22 µm pore size) centrifugation filtration unit (EMD Millipore). Prior to analysis, NICD sample was diluted to 0.1 mg/mL, degassed, and transferred to a 1 mm pathlength quartz cuvette.

Spectra (190-240 nm) were collected at 25°C and 95°C, using a bandwidth of 1 nm, data integration time of 1 s, data pitch of 0.1 nm, and scan speed of 50 nm/min, on a JASCO 815 spectrometer (JASCO, Easton, MD, USA). Each spectrum represents an average of five wavelength scans collected from the same samples under the same conditions. No visible precipitates were noted in samples at either temperature.

Nuclear magnetic resonance (NMR) spectroscopy

^{15}N enrichment of NICD was achieved by growing bacterial cells in M9 minimal medium supplemented with natural abundance glucose and ^{15}N-ammonium chloride as the sole source of carbon and nitrogen, respectively (Muchmore et al., 1989). Isotopically labeled NICD was purified following methods described above, except that the running buffer for gel filtration (and hence final sample buffer) was 20 mM Na_3PO_4 (pH 7.0), 100 mM NaCl, 1 mM EDTA, and 1 mM DTT. Following concentration by ultrafiltration, 0.01% w/v sodium azide was added to prolong the stability of the sample. NMR spectrum of NICD was acquired on a Varian Inova 800 MHz NMR spectrometer (Varian, Palo Alto, CA, USA) equipped with a cryogenically-cooled ^{1}H/^{13}C/^{15}N probe operating at 298 K. The concentration of NICD was (87 μM) in 90% H_2O/10% D_2O. The two-dimensional ^{1}H-^{15}N HSQC (Bodenhausen and Ruben, 1980; Mori et al., 1995) spectrum was collected with spectral widths of 11990.4 Hz and 2595.5 Hz and acquisition times of 85.4 ms and 24.6 ms in the ^{1}H and ^{15}N dimensions, respectively. An inter-scan delay of 1.2 s was employed between successive transients. Data was processed using NMRPipe (Delaglio et al., 1995). Data set was zero-filled prior to Fourier transformation. A 90° phase-shifted squared sine bell was employed for the processing of directly and indirectly detected time domains.

Supplemental References

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., *et al.* (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res *40*, W597-603.

Bodenhausen, G., and Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chemical Physics Letters *69*, 185-189.

Consortium, U. (2014). UniProt: a hub for protein information. Nucleic acids research, gku989.

Das, R.K., and Pappu, R.V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences *110*, 13392-13397.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR *6*, 277-293.

Dobrynin, A.V., and Rubinstein, M. (1995). Flory theory of a polyampholyte chain. Journal de Physique II *5*, 677-695.

Holehouse, A.S., Ahad, J., Das, R.K., and Pappu, R.V. (2015). CIDER: Classification of Intrinsically Disordered Ensemble Regions. Biophysical Journal *108*, 228a.

Mao, A.H., and Pappu, R.V. (2012). Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions. J Chem Phys *137*, 064104.

McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernandez, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J., and Pande, V.S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J *109*, 1528-1532.

Mori, S., Abeygunawardana, C., Johnson, M., and Vanzijl, P. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. Journal of Magnetic Resonance, Series B *108*, 94-98.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E., and Dahlquist, F.W. (1989). Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. Methods Enzymol *177*, 44-73.

Pappu, R.V., Wang, X., Vitalis, A., and Crick, S.L. (2008). A polymer physics perspective on driving forces and mechanisms for protein aggregation. Archives of biochemistry and biophysics *469*, 132-141.
Radhakrishnan, A., Vitalis, A., Mao, A.H., Steffen, A.T., and Pappu, R.V. (2012). Improved atomistic Monte Carlo simulations demonstrate that poly-L-proline adopts heterogeneous ensembles of conformations of semi-rigid segments interrupted by kinks. J Phys Chem B *116*, 6862-6871.
Vitalis, A., and Pappu, R.V. (2009). ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. J Comput Chem *30*, 673-699.