Molecular Cell, Volume *75*

Supplemental Information

Enhancer Features that Drive Formation of Transcriptional Condensates

Krishna Shrinivas, Benjamin R. Sabari, Eliot L. Coffey, Isaac A. Klein, Ann Boija, Alicia V. Zamudio, Jurian Schuijers, Nancy M. Hannett, Phillip A. Sharp, Richard A. Young, and Arup K. Chakraborty

Figure S1

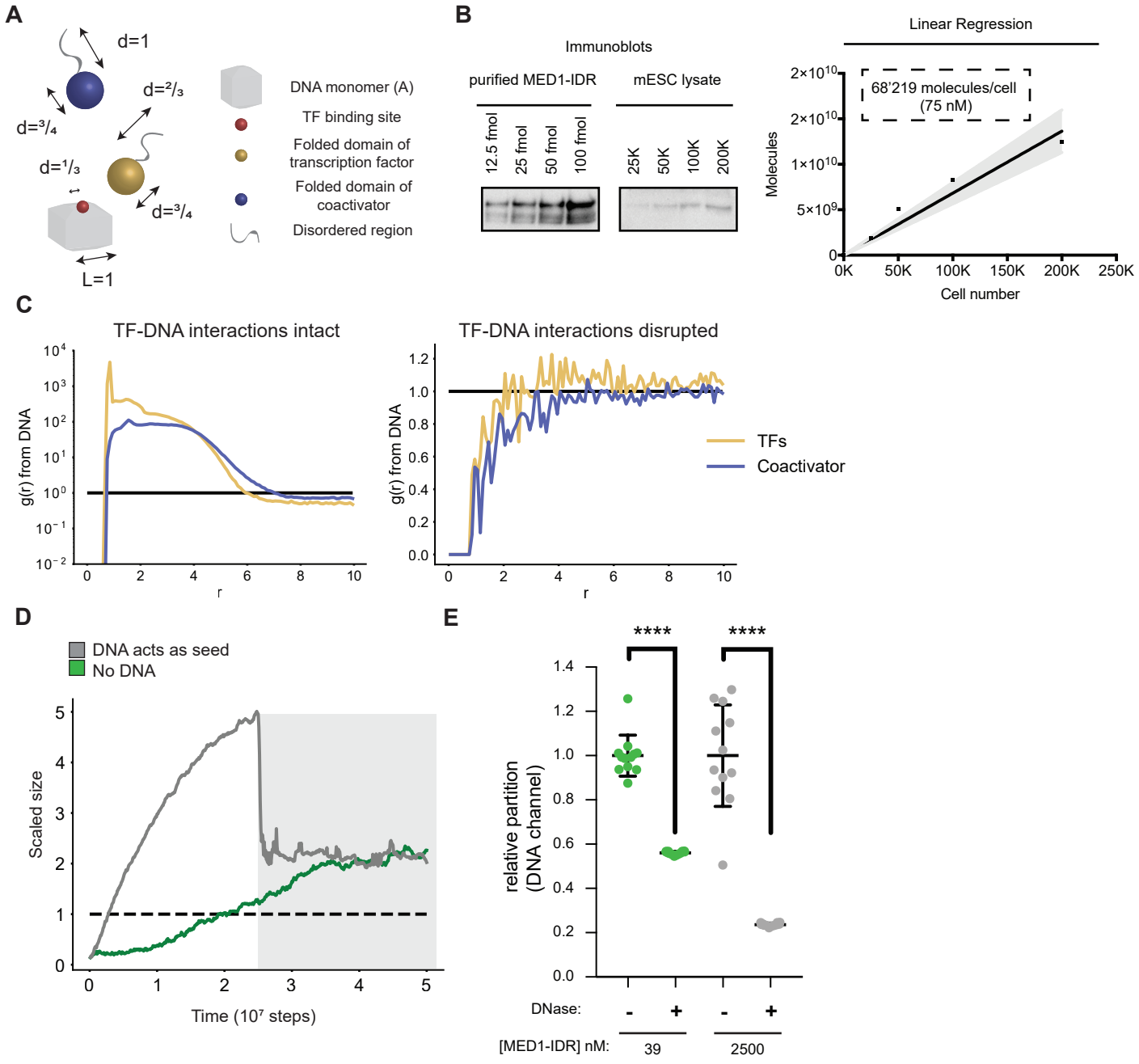

Figure S1: Multivalent TF-DNA interactions promote phase separation of TFs and coactivators, Related to Figures 1,2

A. Schematic illustrating geometries of the simulated molecules (in arbitrary units), including relative sizes of the folded and disordered domains. In typical simulations, the total length of the DNA chain is 10 DNA monomers.
B. Immunoblot of recombinant MED1-IDR at indicated concentrations or lysates from the indicated number of cells is shown on the top panel. Linear regression (bottom panel) is carried out to estimate number and concentration of MED1-IDR per cell (dashed box, bottom panel) (see methods for details).
C. The radial density function (g(r)) is computed around DNA at low concentrations for TFs (yellow) and coactivators (blue), before (left panel) and after (right panel) disruption of TF-DNA interactions. TFs and coactivators form a largely uniform dense phase incorporating DNA (high values and overlap of g(r)), which is lost upon disruption of TF-DNA interactions and condensate dissolution.
D. Dynamics of condensate assembly at conditions with (grey) and without DNA (green line) is represented by average scaled size on y-axis, and time (in simulation steps after initialization) on the x-axis. DNA promotes rate of assembly at high concentrations. However, DNA is not required for condensate stability, as evidenced by high values of scaled size after disruption of TF-DNA interactions (shown by a dark grey background).
E. Scatter-plot depiction of ODNA_20 partition ratio between condensate and background, at high (gray) and low MED1-IDR concentrations (orange) in conditions without DNase I addition (-) or with DNase I addition (+). The partition ratio is normalized to the (-) condition, showing that addition of DNase I degrades DNA.

Figure S2

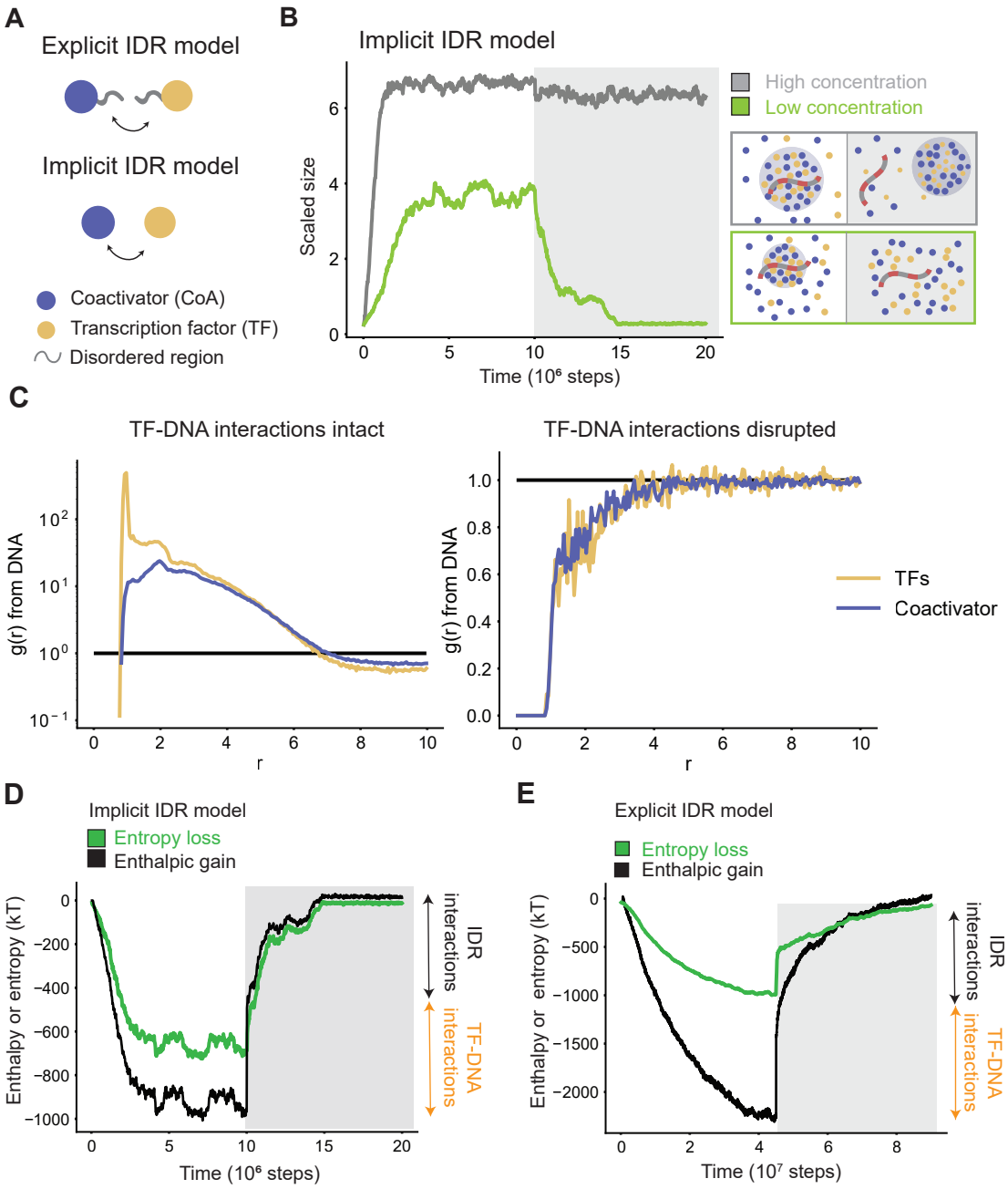

Figure S2: Simplified computational model recapitulates all features of explicit-IDR model, Related to Figure 2

A. Schematic cartoon of difference between explicit IDR model and implicit IDR model.
B. Dynamics of condensate assembly/disassembly at three different protein concentrations (gray = high concentration, orange = low concentration, black = lower concentration) is represented by average scaled size on the y-axis, and time (in simulation steps after initialization) on the x-axis. TF-DNA interactions are disrupted after steady state is reached (shown by a dark gray background). Schematic of phase behavior is presented next to simulation data, enclosed in boxes whose colors match the respective lines.
C. The radial density function (g(r)) is computed around DNA at low concentrations for TFs (yellow) and coactivators (blue), before (left panel) and after (right panel) disruption of TF-DNA interactions. TFs and coactivators form a largely uniform dense phase incorporating DNA (high values and overlap of g(r)), which is lost upon disruption of TF-DNA interactions and condensate dissolution.
D. Energetic attractions (black line) compensate entropic loss (green line) during condensate assembly, but disruption of TF-DNA interactions (magnitude =orange double arrow) causes dissolution at low concentrations.
E. Explicit-IDR simulations show a compensation of entropic loss (green line) by energetic attractions (black line) during condensate assembly, and disruption of TF-DNA interactions causes dissolution. However, the estimate of entropy loss from simulations is an under-count to the total loss of entropy, missing effects of configurational entropy.

Figure S3

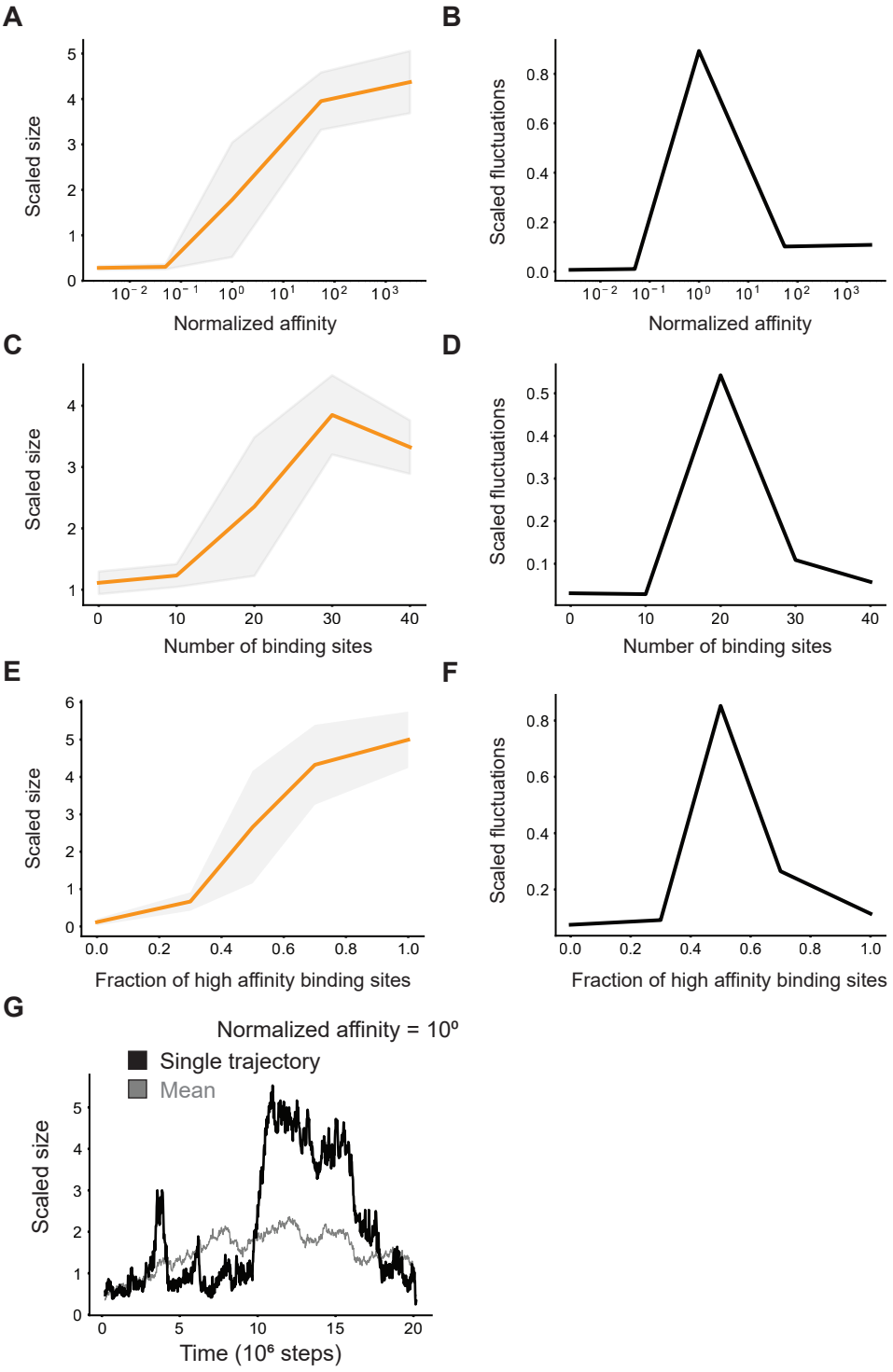

Figure S3: Normalized fluctuations exhibit a sharp peak across the transition point in scaled-size, characteristic of phase transitions, Related to Figures 3,4

Simulations predict a shift in scaled size from stoichiometric binding (≈1) to phase separation (>1) with increasing affinity for TF binding sites on DNA (**A**), valency of TF binding sites (**C**), or fraction of high-affinity binding sites (**E**).

Normalized fluctuations in scaled size (variance over mean) shows a peak near threshold affinity (**B**), valency (**D**), or fraction (**F**); affinity normalized to threshold affinity of E=12kT, fraction of binding sites normalized to total of 30.

G. Typical simulation trajectories show dynamic formation and disassembly of clusters (Transition between low and high scaled size) at threshold affinities.

Figure S4

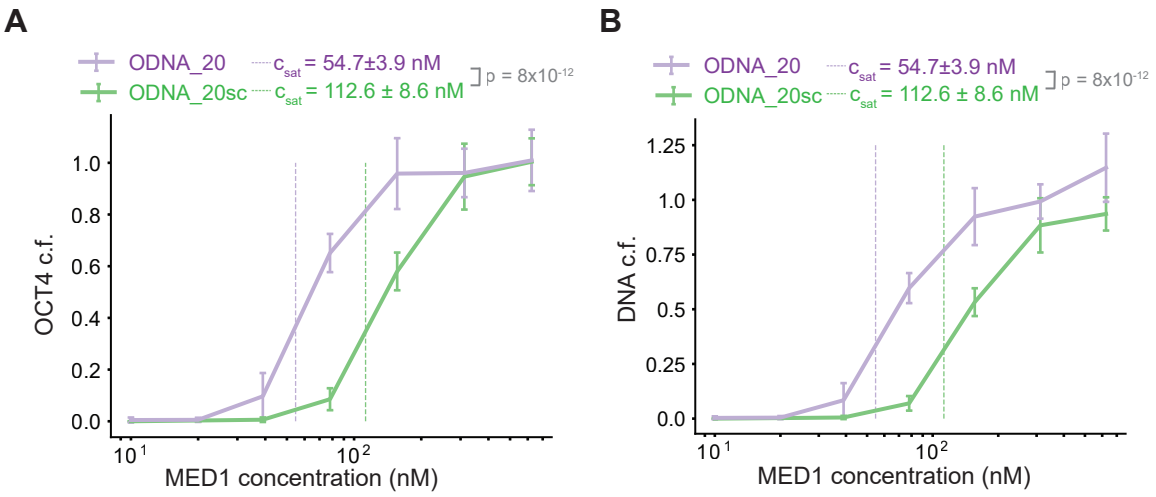

Figure S4: Phase separation of all components is promoted at lower coactivator concentrations by multivalent DNA with high density of TF binding sites, Related to Figures 3,4

A. Condensed fraction of OCT4 (in units of percentage) for ODNA_20 (purple) and ODNA_20sc (green) across a range of MED1-IDR concentrations.
B. Condensed fraction (in units of percentage) of ODNA_20 (purple) and ODNA_20sc (green) across a range of MED1-IDR concentrations.

Solid lines represent mean and error bars represent single standard deviations across replicates.

Figure S5

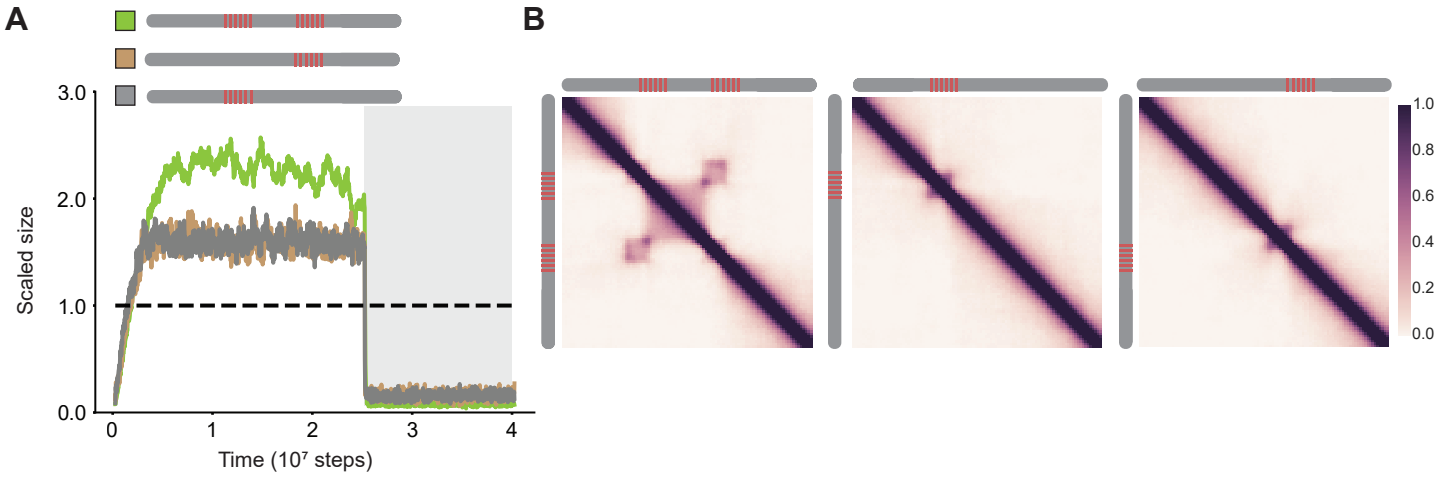

Figure S5: Patches of TF binding site interact over long distances to assemble the transcription machinery, Related to Figure 6

A. Scaled size versus simulation time steps comparing three different distribution of binding site number and distribution (as shown in the schematic legend). Dark gray background signifies disruption of TF-DNA interactions.
B. Contact frequency maps (see methods) show long-range interactions (right panel, checkerboard-like patterns) for DNA with different patch number and distribution.

Figure S6

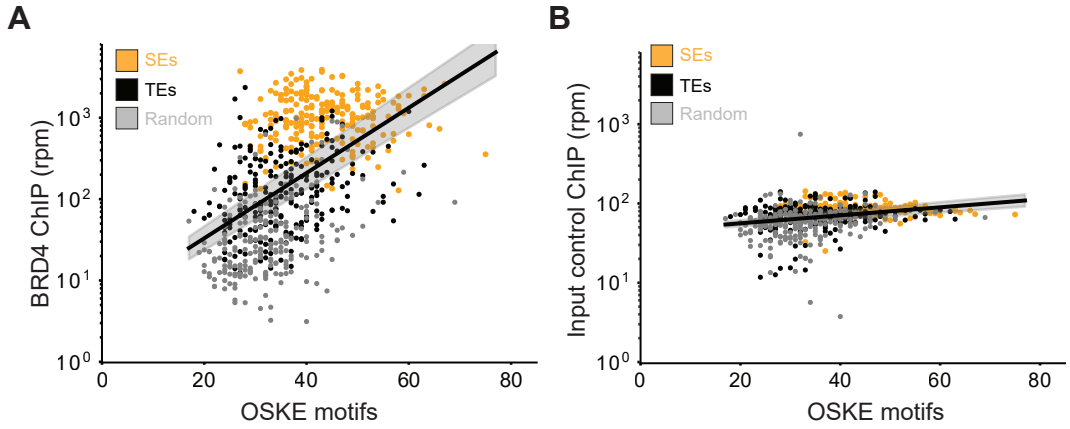

Fig S6. Mammalian genomes show correlation between high occupancy of coactivator and motif density at regulatory elements, Related to Figure 7

A. BRD4 ChIP-Seq counts (y-axis, reads-per-million) against total motifs of OCT4+SOX2+KLF4+ESRRB over 20kb regions centered on SEs (orange), TEs (black), and random loci (gray).
B. Same as **(A)** with data from sequenced input.

The black line represents a linear fit inferred between the logarithmic ChIP signal and motif count, and the grey shaded regions represent the 95% confidence intervals in the inferred slope. The linear model explains a sizable fraction of the observed variance ($R^2 \approx 0.28$) for the BRD4 signal, but not for input control ($R^2 \approx 0.07$).

Figure label	Varied/key parameter
Fig 1C	$L = 48$ units
Fig S1D	$+/-$ DNA, $L = 40$
Fig 2A, Fig S2E	$L = 48{,}40$ units
Fig 2C, Fig S2D	$L = 28$ units
Fig 3A, Fig S3A-B	$E_{TF-DNA} = 6{,}9{,}12{,}16{,}20$ kT
Fig 3E	$E_{IDR-IDR} = 0{,}0.5{,}0.75{,}1{,}1.25{,}1.5$
Fig S3E-F	N_master=0,9,15,21,30
Fig 4A	$N_{binding-sites} = 0{,}10{,}20{,}30{,}40$
Fig 3B ,Fig 3F, Fig 4B	$N_{coA} = 100{,}175{,}250{,}300{,}400$
Fig 5A	$DNA = (D_{40}A_{20}D_{40}, (AD_4)_{20})$
Fig 6A-6B	$DNA = (D_{30}A_{10}D_{20}A_{10}D_{30}, (AD_4)_{20})$
Fig S5	$DNA = D_{30}A_{10}D_{20}A_{10}D_{30}, D_{30}A_{10}D_{60}, D_{60}A_{10}D_{30}$

Table S1. Key simulation parameters, Related to Figures 1,2,3,4,5

Above table highlights variables changed for different simulation plots. Additional details on simulation parameters are split as:

Explicit-IDR simulations in Fig 1C, Fig 2A,Fig S2: Key constant parameters are $DNA = A_{10}, TF = BD_4, coactivator = CE_9$, $N_{DNA} = 1$, $N_{TF} = 100$, $N_{coA} = 200$, $E_{TF-DNA} = 20kT$, $E_{D-D} = E_{D-E} = 1\ kT, E_{D-E} = 1.25\ kT$.

Implicit-IDR simulations in Fig 2C, 3A-B, 3E-F, Fig 4A-B Fig S2, Fig S3: Key constant parameters are $DNA = A_{10}, TF = B, coactivator = C$, $N_{DNA} = 1$, $N_{TF} = 100$, $N_{coA} = 300$, $E_{TF-DNA} = 16kT$, $E_{TF-TF} = 1\ kT$, $E_{TF-coA}, E_{coA-coA} = 1.5\ kT$, $L_{box} = 28\ units$.

Implicit-IDR large DNA simulations in Fig 5A,6A-B,S5: Key constant parameters are $DNA = D_{40}A_{20}D_{40}, TF = B, coactivator = C$, $N_{DNA} = 1$, $N_{TF} = 1000$, $N_{coA} = 3000$, $E_{TF-DNA} = 16kT$, $E_{TF-TF} = 1\ kT$, $E_{TF-coA}, E_{coA-coA} = 1.5\ kT$, $L_{box} = 72\ units$.

Name	Sequence
ODNA_20	TGTAAAACGACGGCCAGTGGATCCTAGGCTTA**ATTTGCAT**TGC AGTAC**ATTTGCAT**GCATGAAT**ATTTGCAT**TAAGCTTG**ATTTGC AT**GTTTCAGA**ATTTGCAT**CGGCTAGC**ATTTGCAT**GGGCTAGA **ATTTGCAT**GCCGGATA**ATTTGCAT**GGCGATTC**ATTTGCAT**GC CAAATC**ATTTGCAT**GCATGAAC**ATTTGCAT**GGCTTACA**ATTT GCAT**GAAACATA**ATTTGCAT**CGATCGAA**ATTTGCAT**GTAGCC GA**ATTTGCAT**GTAGCTAA**ATTTGCAT**GAAATCGG**ATTTGCAT** GTAGCAAT**ATTTGCAT**CTAGCCTA**ATTTGCAT**ACCCTAGC**ATT TGCAT**TAGATTCGGCGGCCGCGTCATAGCTGTTTCCTG
ODNA_20sc	TGTAAAACGACGGCCAGTGGATCCTAGGCTTAATTGCCTCATC CCCTGAAATCGTTAGTGATCAGACCATTCTCTATTAATTTTAGG TGACTCTGAATCTAAATAAACATCTTTGAGATATGCTTACGATA TAATGATCACTTAAGTCATCATTTGTTATCTTACAGATTTGAGA TGCCAACTTTGTGGTGGCCTTAAATTGTAAGCTGAAAACCGTG AAGGAAGAGCGTTTTTGGCATATAGGTGAACTCGGTTCGTTAG CATCAGTCCGGTTCATCTGCTAGGCTGTTATCTATTATTTTATTA TTCTAAATTGTGACGACGTGATAGTGGCAATCACTGACTAGAT TCGGCGGCCGCGTCATAGCTGTTTCCTG
ODNA_5_uniform (ODNA_5)	TGTAAAACGACGGCCAGTGGATCCTAGGCTTA**ATTTGCAT**TGC AGTACATGACTCAGCATGAATAGAGTACGTAAGCTTGGTGATC ACGTTTCAGA**ATTTGCAT**CGGCTAGCAGAGTACGGGGCTAGA GACTGCTAGCCGGATAGACTGCTAGGCGATTC**ATTTGCAT**GCC AAATCATGACTCAGCATGAACATGACTCAGGCTTACAGTGATC ACGAAACATA**ATTTGCAT**CGATCGAAAGAGTACGGTAGCCGA GTGATCACGTAGCTAAGACTGCTAGAAATCGG**ATTTGCAT**GTA GCAATATGACTCACTAGCCTAAGAGTACGACCCTAGCGTGATC ACTAGATTCGGCGGCCGCGTCATAGCTGTTTCCTG
ODNA_5_middle (ODNA_5M)	TGTAAAACGACGGCCAGTGGATCCTAGGCTTAATCTTTAATGC AGTACATGACTCAGCATGAATAGAGTACGTAAGCTTGGTGATC ACGTTTCAGATCGAAATTCGGCTAGCAGAGTACGGGGCTAGAG ACTGCTAGCCGGATA**ATTTGCAT**GGCGATTC**ATTTGCAT**GCCA AATC**ATTTGCAT**GCATGAAC**ATTTGCAT**GGCTTACA**ATTTGC AT**GAAACATACCCAGTAGCGATCGAAAGAGTACGGTAGCCGA GTGATCACGTAGCTAAGACTGCTAGAAATCGGGGGTCATCGTA GCAATATGACTCACTAGCCTAAGAGTACGACCCTAGCGTGATC ACTAGATTCGGCGGCCGCGTCATAGCTGTTTCCTG

Table S2. Annotated sequence of DNAs used in droplet assays, Related to Figures 1,2,3,4
Sequence for each DNA species used in droplet assays. M13 (-21) and M13 reverse primer sequences used in PCR to fluorescently label and amplify DNA are underlined. The octamer motif sequence (ATTTGCAT) is bolded.

# of binding sites	Sequence
0	CAGTGGATCCTAGGCTTAATTGCCTCATCCCCTGAAATCGTTA GTGATCAGACCATTCTCTATTAATTTTAGGTGACTCTGAATCT AAATAAACATCTTTGAGATATGCTTACGATATAATGATCACTT AAGTCATCATTTGTTATCTTACAGATTTGAGATGCCAACTTTG TGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGAGC GTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTCCG GTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAATT GTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCGGC CGCGTCA
1	CAGTGGATCCTAGGCTTA**ATTTGCAT**ATCCCCTGAAATCGTTA GTGATCAGACCATTCTCTATTAATTTTAGGTGACTCTGAATCT AAATAAACATCTTTGAGATATGCTTACGATATAATGATCACTT AAGTCATCATTTGTTATCTTACAGATTTGAGATGCCAACTTTG TGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGAGC GTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTCCG GTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAATT GTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCGGC CGCGTCA
2	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TCTATTAATTTTAGGTGACTCTGAATC TAAATAAACATCTTTGAGATATGCTTACGATATAATGATCACT TAAGTCATCATTTGTTATCTTACAGATTTGAGATGCCAACTTT GTGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGAG CGTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTCC GGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAAT TGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCGG CCGCGTCA
3	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**AAACATCTTTGAGATATGCTTACGATATAATGATCACT TAAGTCATCATTTGTTATCTTACAGATTTGAGATGCCAACTTT GTGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGAG CGTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTCC GGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAAT TGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCGG CCGCGTCA
4	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**CAGCTAGCAGAGTACGGGGCTAGA**ATTTGCAT**ATCAC TTAAGTCATCATTTGTTATCTTACAGATTTGAGATGCCAACTTT GTGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGAG CGTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTCC GGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAAT TGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCGG CCGCGTCA
5	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**CAGCTAGCAGAGTACGGGGCTAGA**ATTTGCAT**TCCGG ATAGACTGCTAGGCGATTC**ATTTGCAT**TTTGAGATGCCAACTT

	TGTGGTGGCCTTAAATTGTAAGCTGAAAACCGTGAAGGAAGA GCGTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGTC CGGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAAA TTGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGCG GCCGCGTCA
6	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**CAGCTAGCAGAGTACGGGGCTAGA**ATTTGCAT**TCCGG ATAGACTGCTAGGCGATTC**ATTTGCAT**GCCAAATCATGACTC AGCATGAAC**ATTTGCAT**TGTAAGCTGAAAACCGTGAAGGAAG AGCGTTTTTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGT CCGGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAA ATTGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGC GGCCGCGTCA
7	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**CAGCTAGCAGAGTACGGGGCTAGA**ATTTGCAT**TCCGG ATAGACTGCTAGGCGATTC**ATTTGCAT**GCCAAATCATGACTC AGCATGAAC**ATTTGCAT**GGCTTACAGTGATCACGAAACATA**A** **TTTGCAT**TTGGCATATAGGTGAACTCGGTTCGTTAGCATCAGT CCGGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTAA ATTGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGGC GGCCGCGTCA
8	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGAATCA TCATGAAT**ATTTGCAT**TAAGCTTGGTGATCACGTTTCAGA**ATT** **TGCAT**CAGCTAGCAGAGTACGGGGCTAGA**ATTTGCAT**TCCGG ATAGACTGCTAGGCGATTC**ATTTGCAT**GCCAAATCATGACTC AGCATGAAC**ATTTGCAT**GGCTTACAGTGATCACGAAACATA**A** **TTTGCAT**CGATCGAAAGAGTACGGTAGCCGA**ATTTGCAT**CAG TCCGGTTCATCTGCTAGGCTGTTATCTATTATTTTATTATTCTA AATTGTGACGACGTGATAGTGGCAATCACTGACTAGATTCGG CGGCCGCGTCA
5 (with lower density)	CAGTGGATCCTAGGCTTA**ATTTGCAT**TGCAGTACATGACTCA GCATGAATAGAGTACGTAAGCTTGGTGATCACGTTTCAGA**AT** **TTGCAT**CGGCTAGCAGAGTACGGGGCTAGAGACTGCTAGCCG GATAGACTGCTAGGCGATTC**ATTTGCAT**GCCAAATCATGACT CAGCATGAACATGACTCAGGCTTACAGTGATCACGAAACATA **ATTTGCAT**CGATCGAAAGAGTACGGTAGCCGAGTGATCACGT AGCTAAGACTGCTAGAAATCGG**ATTTGCAT**GTAGCAATATGA CTCACTAGCCTAAGAGTACGACCCTAGCGTGATCACTAGATTC GGCGGCCGCGTCA

Table S3. Annotated sequence of DNAs used in luciferase reporter assays, Related to Figure 4

Sequences tested in luciferase reporter assays are provided here with the octamer motif sequence (ATTTGCAT) bolded. The sequences provided were cloned into the SalI site of the previously characterized pGL3-basic OCT4 (Whyte et al 2013) as described in methods.

Movie S1. Simulation trajectory of phase separation mediated by DNA at low protein concentrations, Related to Figure 2

Typical simulation trajectory of DNA-scaffolded condensate formation and subsequent disassembly (sub-titles) at low protein concentrations (Ref green line in Fig 2A). DNA particles are in red, TFs in yellow, coactivators in blue (same as Fig 1A), and all IDRs are in grey. The movie is played at a 10x acceleration in the frame-rate.

Movie S2. Simulation trajectory of phase separation mediated by DNA at high protein concentrations, Related to Figure 2

Typical simulation trajectory of DNA-scaffolded condensate formation and ejection of DNA after disruption of TF-DNA interactions(sub-titles) at high protein concentrations (Ref grey line in Fig 2A). DNA particles are in red, TFs in yellow, coactivators in blue (same as Fig 1A), and all IDRs are in grey. The movie is played at a 10x acceleration in the frame-rate.