bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Force generation by protein-DNA co-condensation

Author list: Thomas Quail[1,2,3,4], Stefan Golfier[1,2,3,4], Maria Elsner[1,2,3,4], Keisuke Ishihara[1,2,3,4], Vasanthanarayan Murugesan[1,2,3,4], Roman Renger[1], Frank Jülicher*[2,3,4], Jan Brugués*[1,2,3,4]

Affiliations:

[1]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[2]Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany
[3]Center for Systems Biology Dresden, 01307 Dresden, Germany
[4]Cluster of Excellence Physics of Life, TU Dresden, 01307 Dresden, Germany
*correspondence: julicher@pks.mpg.de, brugues@mpi-cbg.de

Abstract: Interactions between liquids and surfaces generate forces[1,2] that are crucial for many processes in biology, physics, and engineering, including the motion of insects on the surface of water[3], modulation of the material properties of spider silk[4], and self-assembly of microstructures[5]. Recent studies have shown that cells assemble biomolecular condensates via phase separation[6]. In the nucleus, these condensates are thought to drive transcription[7], heterochromatin formation[8], nucleolus assembly[9], and DNA repair[10]. Here, we show that the interaction between liquid-like condensates and DNA generates forces that might play a role in bringing distant regulatory elements of DNA together, a key step in transcriptional regulation. We combine quantitative microscopy, *in vitro* reconstitution, optical tweezers, and theory to show that the transcription factor FoxA1 mediates the condensation of a DNA-protein phase via a mesoscopic first-order phase transition. After nucleation, co-condensation forces drive growth of this phase by pulling non-condensed DNA. Altering the tension on the DNA strand enlarges or dissolves the condensates, revealing their mechanosensitive nature. These findings show that DNA condensation mediated by transcription factors could bring distant regions of DNA in close proximity, suggesting that this physical mechanism is a possible general regulatory principle for chromatin organization that may be relevant *in vivo*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Main text: Compartmentalization is key to organizing cellular biochemistry. Biomolecular condensate formation underlies the compartmentalization of many cellular functions[6]. Considerable progress has been made towards understanding the biophysical properties of condensates in bulk. However, how these condensates interact with other cellular components such as polymers, membranes, and chromatin remains unclear. Transcriptional hubs are an example of compartments in the nucleus. These hubs involve the coalescence of transcription factors, biochemical regulators of transcription, and DNA[11]. The physical nature of these transcription hubs is under debate, though recent studies have proposed that transcriptional hubs can be understood as examples of biomolecular condensates[12]. In theory, the interactions between transcriptional machinery condensates and the DNA polymer could deform DNA, potentially bridging distal regulatory elements, a critical step in gene regulation. However, we still lack a physical picture of how transcriptional regulators interact with each other and with the surface of the DNA polymer.

To investigate how transcription factors physically organize DNA, we attached linearized λ-phage DNA to a coverslip via biotin-streptavidin linkers (Fig. 1a). We used TIRF microscopy to image the interactions between DNA and Forkhead Box Protein A1 (FoxA1), a pioneer transcription factor that regulates tissue differentiation across a range of organisms[13] (Fig. 1b). Upon addition of 10 nM FoxA1-mCherry (FoxA1) to the flow chamber in the presence of DNA, FoxA1 formed protein condensates that decorated the strand (Fig. 1c). In the absence of DNA, FoxA1 did not nucleate condensates in solution at concentrations ranging from 10 to 500 nM (Extended Data Fig. 1a). The requirement for DNA in condensate formation at low concentrations suggests that DNA mediates the condensation of a thin layer of FoxA1 on DNA.

In our assay, DNA molecules displayed a broad distribution of end-to-end distances (L), determined by the DNA-coverslip attachment points (Fig. 1c, d). This end-to-end distance tunes the tension of the DNA[14]. For DNA strands with end-to-end distances greater than approximately 10 μm, FoxA1 generated protein condensates on DNA (Fig. 1c). However, FoxA1 condensation did not influence the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA molecule (Fig. 1c, leftmost pair of images). Strikingly, for DNA molecules with end-to-end distances below 10 μm, FoxA1 pulled DNA into highly enriched condensates of FoxA1 and DNA (Fig. 1c, Extended Data Fig. 1b-e) with a density of roughly 750 molecules/μm^3 (see Methods, Extended Data Fig. 2a-d). To quantify FoxA1-mediated DNA condensation, we measured the cross-correlation of FoxA1-DNA intensities as a function of end-to-end distance (see Methods, Fig. 1d,e, Extended Data Fig. 3a). Consistent with the ability of FoxA1 to form FoxA1-DNA condensates at low tensions, the cross-correlation decayed from one to zero with increasing end-to-end distance (Fig. 1e). Thus, FoxA1 mediates the formation of a DNA–protein-rich phase in a tension-dependent manner.

The observation that FoxA1 drives DNA condensation suggests that it can overcome the DNA molecule's entropic tension set by the end-to-end distance[14]. Incorporating DNA into the condensates increases the tension on the strand, thereby reducing the transverse DNA fluctuations of the non-condensed DNA. To quantify this, we measured the DNA envelope width of the non-condensed DNA fluctuations (see Methods, Extended Data Fig. 3b). In buffer, the DNA envelope width decreased as a function of end-to-end distance, consistent with the corresponding increase of DNA strand tension for increasing end-to-end distances[14] (Fig. 1f). However, in the presence of FoxA1, the DNA envelope width remained constant for all end-to-end distances as FoxA1 pulled DNA into one or more condensates. The magnitude of the DNA envelope width was lower in the presence of FoxA1 than in buffer conditions for all end-to-end distances (Fig. 1f). Taken together, this suggests that FoxA1-DNA condensates generate forces that can overcome the entropic tension of the non-condensed DNA and buffer its tension.

The observation that FoxA1 can mediate DNA condensation suggests that it could bridge distant DNA strands. To investigate this possibility, we examined DNA molecules that were bound to the same streptavidin molecule at one end (Fig. 1g, Extended Data Fig. 3c). In the absence of FoxA1, these DNA molecules form a v-shaped morphology and fluctuate independently of one another. Upon addition of FoxA1, however, we observed that the two strands zipped together, generating a

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

y-shaped morphology as the condensation of FoxA1 increased over time (Fig. 1g, Extended Data Fig. 3c). Taken together, these data demonstrate that FoxA1 can physically bridge DNA strands in both *cis* and *trans*.

Two mechanisms can be postulated to explain FoxA1-mediated DNA condensation in our experiments: (i) direct cross-linking via the multiple DNA-binding activities of FoxA1[15] or (ii) weak protein-protein interactions driven by disordered regions of FoxA1. FoxA1 consists of a winged helix-turn-helix DNA-binding domain and two N and C termini domains that are mostly disordered[15]. The DNA-binding domain contains a sequence-specific binding region composed of three alpha helices and a non-sequence-specific binding region composed of two wings. Two point mutations known to affect sequence-specific DNA binding (NH-FoxA1[15]) had virtually no influence on DNA condensation activity (Fig. 2a). Although the presence of two point mutations known to affect non-sequence-specific DNA binding (RR-FoxA1[15]) partially inhibited FoxA1 localization to the strand (Fig. 2b), this mutant still condensed DNA. In this case, condensation occurred on a time scale of minutes rather than seconds (as in WT-FoxA1), which can be explained by the delay in condensing sufficient RR-FoxA1 to the strand. These data suggest that non-sequence-specific binding drives the localization of FoxA1 to DNA but does not mediate DNA condensation through cross-linking. Furthermore, the sequence-specific binding domain of FoxA1 is dispensable for its localization to DNA *in vitro.* To probe whether FoxA1 protein-protein interactions through disordered domains mediate DNA condensation, we truncated both the N and C termini of FoxA1. Although ΔN-FoxA1 retained DNA condensation activity (Fig. 2c), truncating the disordered C terminus of FoxA1 largely inhibited DNA condensation activity (Fig. 2d). Additionally, we found that, at high FoxA1 concentrations in bulk (50 μM), 3% PEG (30K) nucleated highly-enriched spherical FoxA1 condensates (Extended Data Fig. 4a), further suggesting the existence of weak FoxA1-FoxA1 interactions. Thus, non-sequence-specific binding drives FoxA1 localization to DNA, and the disordered C terminus of FoxA1 promotes DNA condensation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our results support the hypothesis that FoxA1 condenses onto DNA to generate a DNA–protein-rich condensate via weak protein-protein interactions that exerts a pulling force on the non-condensed strand (see the section Thermodynamic description of DNA-protein condensation in the Supplementary information). To explore the thermodynamics of condensation, we developed a theoretical description based on a semi-flexible polymer partially condensing into a liquid-like condensate. Here, the semi-flexible polymer is DNA and the condensation is mediated by the transcription factor. The free energy of this process contains volume, ($\upsilon\frac{4}{3}\pi R^3$), and surface contributions, ($\gamma 4\pi R^2$), as well as a term representing the free energy of the non-condensed DNA (Fig. 3a), where υ is the condensation free energy per volume, R is the condensate radius, and γ is the surface tension of the condensate. We assume that DNA is fully collapsed inside the condensate and thus its volume is proportional to the condensed DNA contour length, $V = \alpha\, L_d$, where $1/\alpha$ describes the packing density given as DNA length per condensate volume. The free energy of the polymer, $F_p(L, L_p) = \int_0^L f(L, L_p)\, dl$, can be obtained from the force-extension curve of the polymer $f(L, L_p)$, where L_p is the contour length of the non-condensed polymer. Using $L_p = L_c - L_d$ where L_c is the contour length of λ-phage DNA (16.5 μm), the free energy is as follows,

$$F(L, L_d) = -\upsilon\alpha L_d + \gamma 4\pi \left(\frac{3\alpha}{4\pi}\right)^{\frac{2}{3}} {L_d}^{\frac{2}{3}} + \kappa\left(\frac{(L_c - L_d)^2}{4(L_c - L_d - L)} - \frac{L}{4} + \frac{L^2}{2(L_c - L_d)} - \frac{(L_c - L_d)}{4}\right)$$

where $\kappa = \frac{k_B T}{P}$, k_B is the Boltzmann constant, T is the temperature, and P is the persistence length of DNA (see the section Thermodynamic description of DNA-protein condensation in the Supplementary information). For fixed L, the minimum of $F(L, L_d)$ determines the preferred size of the condensate. This free energy predicts upon variation of L a stochastic first-order phase transition for the formation of DNA–protein condensates (Fig. 3b). The distribution of condensate sizes is then given by $P(L_d) \sim e^{-\beta F(L, L_d)}$ for fixed L (Fig. 3c). This accounts for a sharp transition of DNA condensation controlled by the end-to-end distance and thus the tension of the DNA molecule. The first-order nature of this behavior implies regimes of hysteresis and bistability. Our theory also predicts that the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

condensation forces exerted on the non-condensed DNA are kept roughly constant.

To test this theory, we first measured DNA condensate volumes and found that they increase linearly with the length of condensed DNA (L_d), with α=0.04 ± 0.01 μm^2 (Fig. 3d, Extended Data Fig. 4d, Methods). This confirms that DNA is in a collapsed conformation inside the condensates. Next, we simultaneously fit the predictions to the average amount of DNA contained in the condensates (L_d), and the probability of nucleating a DNA condensate (P_{cond}) as a function of end-to-end distance (see Methods). We calculated L_d (Fig. 3e, Extended Data Fig. 4e, Extended Data Fig. 5) and P_{cond} (Fig. 3g, Extended Data Fig. 4f) using the Boltzmann probability distributions (Fig. 3c) from the free energy. Our fits agree quantitatively with the data and show that L_d decreases with L until a critical end-to-end distance beyond which DNA condensates do not form. Below this critical length, we observed that the force exerted by the condensate is buffered at 0.21 pN (0.18 – 0.30 pN CI), consistent with the theory (Fig. 3f). To complement our force measurements, we performed optical tweezer measurements of FoxA1-mediated DNA condensation. Incubating a single λ-phage DNA molecule at either L = 6 or 8 μm in the presence of 150 nM FoxA1 generated forces on the order of 0.4-0.6 pN, consistent with the force measurements using fluorescence microscopy (Methods, Extended Data Figs. 6,7). Finally, P_{cond} exhibits a sharp transition at L = 10.5 μm (9.4 – 10.9 μm CI), in agreement with a stochastic first-order phase transition (Fig. 3g). We also observed a sudden force jump during the onset of condensate formation (as measured by the individual temporal force trajectories in the optical tweezer experiments), consistent with a first order phase transition (Extended Data Figs. 6c,7). Close to the transition point FoxA1-mediated DNA condensation displayed bistability. This bistability was observed in strands that contained multiple FoxA1 condensates, but where only some of them condensed DNA (Extended Data Fig. 8a). Our fits allowed us to extract the physical parameters associated with condensate formation, namely the condensation free energy per volume υ = 2.6 pN/μm^2 (2.3 – 5.2 pN/μm^2 CI) and the surface tension γ=0.04 pN/μm (0.04 – 0.28 pN/μm CI), see Methods section. These parameters are consistent with previous measurements for *in vitro* and *in vivo* condensates[16,17].

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our theory and experiments show that two key parameters govern DNA–protein co-condensation, namely the condensation free energy per volume (υ) and the surface tension (γ). We reasoned that different DNA-binding proteins may exhibit a range of behaviors depending on these parameters. First, we investigated the sequence-specific DNA-binding region mutant (NH-FoxA1), which also condensed DNA but to a lesser extent (Fig. 2a). Quantitatively, we found that the surface tension of condensates formed with this mutant was roughly unchanged compared to WT-FoxA1, γ=0.065 pN/μm (0.05 – 0.07 pN/μm CI), but the free energy per volume of condensation was reduced consistent with reduced DNA binding, υ=1.05 pN/μm^2 (0.9 – 1.1 pN/μm^2 CI), Extended Data Fig. 9, Fig. 4a. This was also reflected in a decrease in the extent of DNA packing with α = 0.09 ± 0.02 μm^2 (Extended Data Fig. 9a). We also observed that NH-FoxA1-mediated condensates generated a force of 0.17 pN (0.16 – 0.19 pN, CI), lower than that for WT-FoxA1. In addition, NH-FoxA1 displayed bistable DNA–protein condensation activity in the neighborhood of the transition point (Extended Data Fig. 8b). Next, we examined the interactions of a different transcription factor Tata-Box-binding protein (TBP) with DNA. We found that TBP also formed small condensates on DNA, but did not condense DNA even at the lowest imposed DNA tensions (Fig. 4b). Instead, TBP performed a diffusive motion along the DNA strand (Extended Data Fig. 10c), suggesting that DNA-protein condensation is not thermodynamically favored. Another transcription factor, Gal4-VP16, formed condensates on DNA and condensed DNA in a tension-dependent manner consistent with FoxA1 (Extended Data Fig. 10e). Lastly, we analyzed somatic linker histone H1, a protein that is structurally similar to FoxA1. However, in contrast to FoxA1, one of the known functions of H1 is to compact chromatin[18], so we expected H1 to strongly condense DNA. Consistent with this, we found that H1 displayed a stronger DNA condensation activity compared to FoxA1, condensing DNA for all measured end-to-end distances (Fig. 4c). Interestingly, the *Xenopus* embryonic linker histone B4 condensed DNA in a tension-dependent manner but not to the same extent as H1 (Extended Data Fig. 10f). Thus, we propose that the competition between condensation free energy per volume of the DNA–protein

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phase and surface tension regulate a spectrum of DNA condensation activities, which may be tuned by the structure of transcription factors.

Here, we show that FoxA1 can condense DNA under tension to form a DNA–protein-rich phase that nucleates through a force-dependent first-order transition for forces below a critical value. This critical force, which is on the order of 0.2-0.6 pN for FoxA1, is set by co-condensation forces that the DNA–protein phase exerts on the non-condensed DNA. These forces are similar in magnitude to those recently measured for DNA loop extrusion on the order of 0.2-1 pN[19,20] and those estimated in intact nuclei from nuclear condensate fusion[21]. Thus, we speculate that these weak forces we find *in vitro* may be of relevance to the mechanics of chromatin organization, though future studies are necessary to show this. Taken together, our work suggests that co-condensation forces may act as an additional mechanism to remodel chromatin in addition to molecular motors that extrude loops and complexes that remove or displace nucleosomes (Fig. 4d).

Transcription-factor-mediated DNA–protein condensation represents a possible mechanism by which transcription factors coordinate enhancer-promoter contacts in transcriptional hubs[12]. In this context, DNA–protein condensates could act as scaffolds, pulling co-factors into the droplet (Fig. 4d). Our theoretical description reveals that these DNA–protein condensates are formed via a first-order phase transition, suggesting that they can be assembled and disassembled rapidly by changing mechanical conditions. Near the transition point, assembly and disassembly of these *in vitro* DNA–protein condensates becomes highly stochastic, reminiscent of the rapid dynamics associated with the initiation and cessation of transcriptional bursts observed in vivo[22].

We have demonstrated that protein-DNA co-condensation is associated with a difference in chemical potential between the condensed and non-condensed DNA. This difference in chemical potential is transduced by the condensate to perform mechanical work on the non-condensed DNA strand. Capillary forces represent another example of forces that involve liquid-surface interactions[1,2,23]. With both co-condensation and capillary forces, attractive interactions give rise to the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transduction of free energy into work. Such forces may also be relevant beyond chromatin in other biological contexts, including membranes and the cytoskeleton.

DNA–protein co-condensation not only provides mechanisms to facilitate enhancer–promoter contacts, but could also play a more general role in DNA compaction and maintenance of bulk chromatin rigidity in processes such as mitotic chromatid compaction[24], and the formation of chromatin compartments[8,25,26]. Owing to the tension-dependent nature of DNA–protein co-condensation, our work suggests that these forces could play a key and, as yet, underappreciated role in genome organization and transcriptional initiation. It is appealing to imagine that transcriptional outputs not only respond to concentrations of transcription factors in the nucleus, but also to mechanical cues from chromatin.

References:

1. De Gennes, P.-G., Brochard-Wyart, F. & Quere, D. *Capillarity and Wetting Phenomena*. Springer (2013).
2. Bico, J., Reyssat, É. & Roman, B. Elastocapillarity: When Surface Tension Deforms Elastic Solids. *Ann. Rev. Fluid Rev.* **50,** 629-659 (2017).
3. Hu, D.L. & Bush, J. W. M. The hydrodynamics of water-walking arthropods. *J. Fluid Mech.* **664,** 5-33 (2010).
4. Elettro, H., Neukirch, S., Vollrath, F. & Antkowiak, A. In-drop capillary spooling of spider capture thread inspires hybrid fibers with mixed solid–liquid mechanical properties. *Proc. Natl. Acad. Sci. USA* **113,** 6143–6147 (2016).
5. Mastrangeli, M. *et al.* Self-assembly from milli- to nanoscales: methods and applications. *J. Micromech. Microeng.* **19,** 083001–38 (2009).
6. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30,** 39–58 (2014).
7. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361,** eaar3958–13 (2018).
8. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547,** 236–240 (2017).
9. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165,** 1686–1697 (2016).
10. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).
11. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361,** 1341–1345 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45-54, (2019).
13. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37,** 76–81 (2016).
14. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265,** 1599–1600 (1994).
15. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23,** 804–809 (2009).
16. Jawerth, L. M. *et al.* Salt-Dependent Rheology and Surface Tension of Protein Condensates Using Optical Traps. *Phys. Rev. Lett.* **121,** 258101 (2018).
17. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324,** 1729–1732 (2009).
18. Botao, X. *et al.* Histone H1 compacts DNA under force and during chromatin assembly. *Mol. Biol. Cell.* **23,** 4864-4871 (2012).
19. Golfier, S., Quail, T., Kimura, H. & Brugués, J. Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner. *Elife* **9,** (2020).
20. Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments. *Curr. Opin. Cell Biol.* **64,** 124–138 (2020).
21. Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175,** 1481–1491 (2018).
22. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).
23. Jang, J., Schatz, G. C., & Ratner, M. A. Capillary force in atomic force microscopy. *J. Chem. Phys.* **120,** 1157-1160 (2004).
24. Yoshimura, S. H. & Hirano, T. HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? *J. Cell. Sci.* jcs.185710–8 (2016).
25. Strom, A. R. *et al.* HP1α is a chromatin crosslinker that controls nuclear and mitotic chromosome mechanics. *bioRxiv* **29,** 717–742 (2020).
26. Williams, J. F. *et al.* Phase separation enables heterochromatin domains to do mechanical work. *bioRxiv* **78,** 236–245 (2020).
27. McCall, P. M. *et al.* Quantitative phase microscopy enables precise and efficient determination of biomolecular condensate composition. *bioRxiv* **16,** 2885–19 (2020).
28. Shakya, A., Park, S., Rana, N. & King, J. T. Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization. *Biophys. J.* **118,** 753–764 (2020).
29. Turner, A. L. *et al.* Highly disordered histone H1−DNA model complexes and their condensates. 1–6 (2018). doi:10.1073/pnas.1805943115

Acknowledgements: We thank Pavel Tomancak, Anthony Hyman, Stephan Grill, Iain Patten, Martin Loose, David Oriola, Benjamin Dalton, Patrick McCall, and Claudia Meyer for helpful feedback and stimulating discussions. We would like to

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

thank Aliona Bogdanova for discussions and help with cloning as well as both the Protein Expression and Purification Facility and the Light Microscopy Facility at the Max Planck Institute of Molecular Cell Biology and Genetics. We would like to acknowledge and thank Nadine Vastenhouw for discussions and the construct containing the Tata-box-binding protein template, and Christoph Zechner for help with statistical analyses. Lastly, we would like to thank Kenneth Zaret for sending constructs, information on mutant FoxA1 proteins, and advice on FoxA1 protein purification. This work was supported by an EMBO long-term fellowship (ALTF-1456-2015) (TQ), DFG project BR 5411/1-1 (JB,VN), and a Volkswagen "Life" grant number 96827 (JB,TQ).

Author contributions: T.Q. and J.B. conceived the project. T.Q. and S.G. performed imaging experiments. S.G. established the single-strand DNA assay. T.Q. purified proteins, made constructs, and performed data analysis. T.Q., J.B., and F.J. performed theoretical calculations. M.E. made the TBP and Gal4-VP16 constructs and purified the proteins. V.M. purified B4. T.Q. and R.R. performed optical tweezer measurements. S.G. and R.R. performed data analysis and contributed to methods writing. K.I. made the initial FoxA1 construct and provided key biochemical support. J.B. and F.J. supervised the work. T.Q., J.B., and F.J. wrote the manuscript and all authors contributed ideas and reviewed the manuscript.

Competing interest statement: The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1

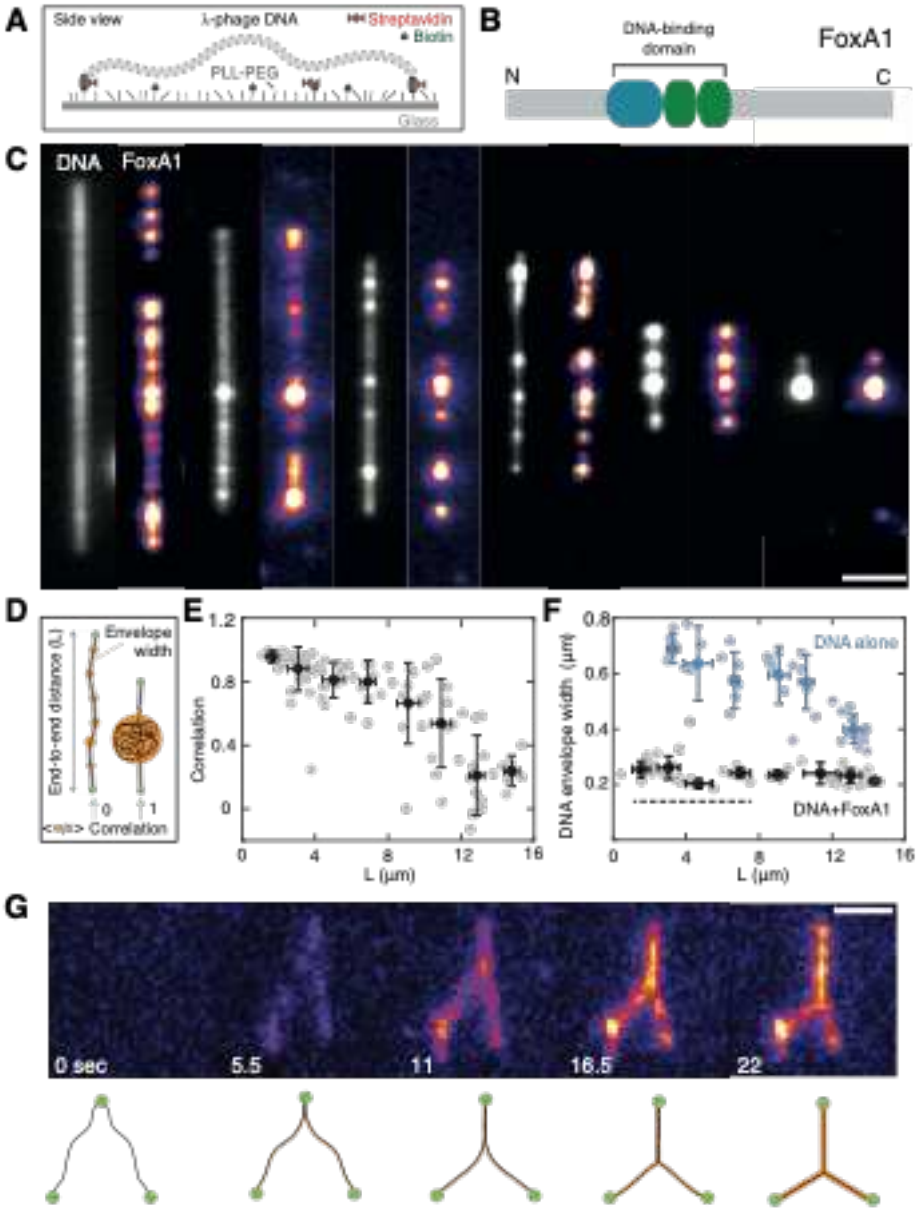


Figure 1: FoxA1 forms DNA-FoxA1 condensates in a tension-dependent manner. (A) Schematic of single λ-phage DNA molecule assay. (B) Structure of FoxA1, consisting of a structured DNA-binding domain flanked by mostly disordered N and C termini. The DNA-binding domain has a sequence-specific binding region (blue) and two non-sequence-specific binding regions (green). (C) The extent of FoxA1-mediated DNA condensation depends on the end-to-end distance of the strand. Representative time-averaged projections of FoxA1 and DNA. Note that the total amount of DNA is the same in each example. The DNA was imaged using 10 nM Sytox Green. Scale bar=2 μm. (D) Schematic displaying three main quantities used to characterize DNA-FoxA1 condensation: the end-to-end distance L; Cross-correlation of DNA and FoxA1 intensities; and DNA envelope width, a measure of transverse DNA fluctuations. (E) Cross-correlation of FoxA1 and DNA signals shows that FoxA1 condenses DNA below a critical end-to-end distance. The gray dots represent individual strands, n=107. The data is binned every 2-μm (black, mean ± SD for both correlations and strand lengths). (F) DNA

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

envelope width measurements (see Methods) reveal that FoxA1-DNA condensation buffers DNA tension (blue and black dots correspond to control and DNA+FoxA1 conditions, n=45 and n=50 respectively). The data is binned every 2-μm (mean ± SD for both the envelope widths and strand lengths). The dashed black line represents the theoretical diffraction limit. (G) Representative images of FoxA1 zipping two independent DNA strands over time. Scale bar=2 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2

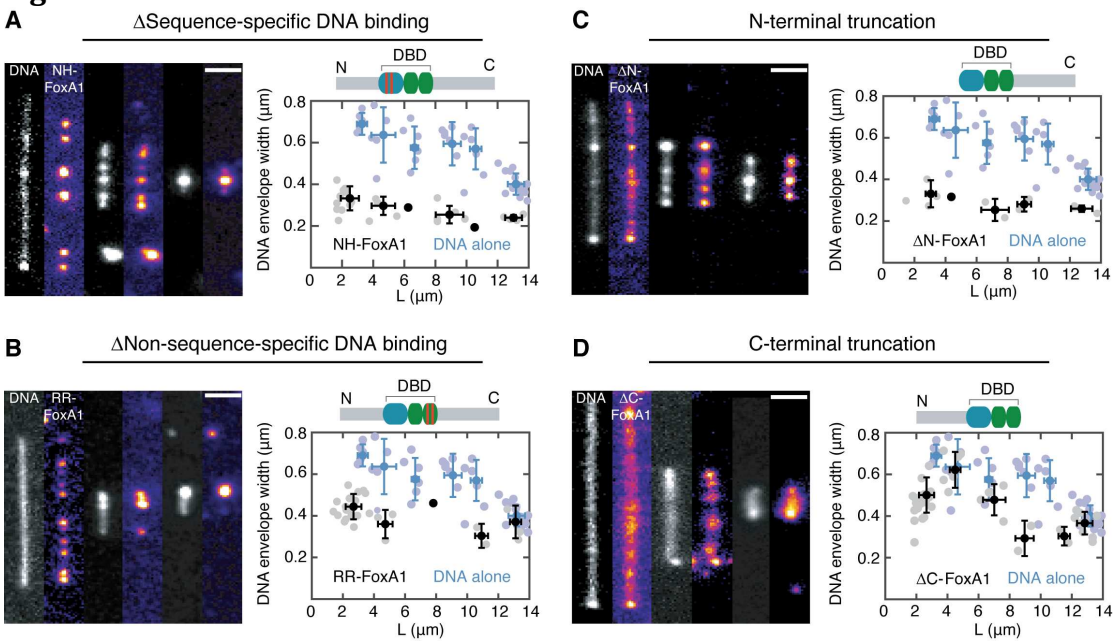


Figure 2: Mutant analysis reveals that the C terminus of FoxA1 drives DNA condensation. Representative images and DNA envelope width measurements for FoxA1 mutants. The data is binned every 2-μm and the mean ± SD (for both the envelope width and the strand length) are shown in black for each mutant and in blue for control (n=45). Scale bars=2 μm. (A) Sequence-specific DNA binding mutant NH-FoxA1 condenses DNA (n=30). (B) Non-sequence-specific DNA-binding mutant RR-FoxA1 condenses DNA (n=28). (C) N-terminal truncation of FoxA1 ΔN-FoxA1 condenses DNA (n=13). (D) C-terminal truncation of FoxA1 ΔC-FoxA1 inhibits DNA condensation (n=44). In all conditions, the protein concentration was 10 nM.

Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

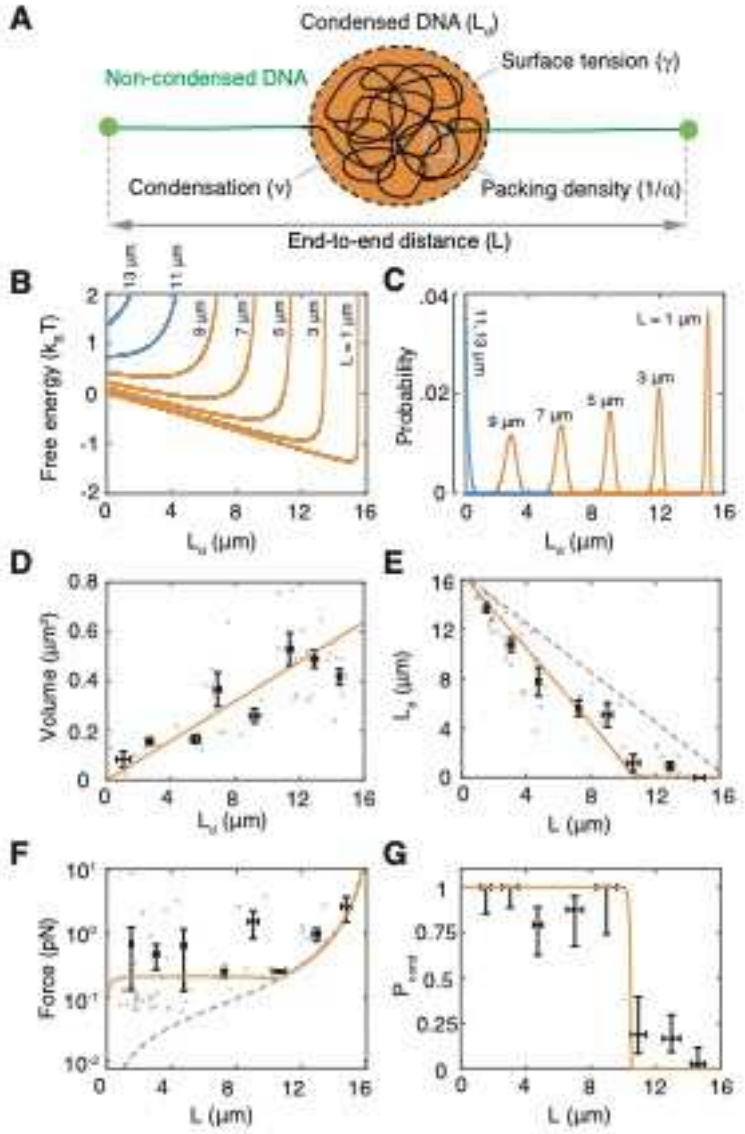


Figure 3: Thermodynamic description of a liquid phase condensing onto a semi-flexible polymer explains FoxA1-mediated DNA condensation. (A) Schematic representing DNA-FoxA1 condensation (orange). DNA can be in a condensed state (black) or a non-condensed state (green). DNA condensation depends on the condensate surface tension (γ), condensation free energy per volume (υ), and DNA packing efficiency (α). (B) Free energy profiles as a function of condensed DNA (L_d) for different L reveal a first-order phase transition for DNA–protein condensation (orange and blue correspond to favorable and unfavorable condensation, respectively). (C) Boltzmann distributions corresponding to the free energy profiles in (B). (D) Condensate volume linearly increases with L_d. The orange curve represents a linear fit to individual strands (n=47). For (D), (E), and (F), individual strands are represented as gray dots and binned mean±SEM is in black. (E) Amount of condensed DNA as a function of L (n=63) reveals sharp transition. Orange curve represents optimal theoretical fit. The gray dashed-line corresponds to the limit of maximum condensation where L_d is equal to the contour length of DNA (16.5 μm) minus L. (F) Condensation forces that DNA-protein condensates exert on non-condensed DNA are buffered (n=62). Orange curve is the theoretical prediction. The gray dashed line

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

represents the force when L_d=0. (G) Probability to nucleate a DNA-FoxA1 condensate (P_{cond}) reveals a sharp transition at a critical end-to-end distance. P_{cond} is computed from binned local correlation data (n=181 condensates). The end-to-end distance error bars are the SD and the P_{cond} error bars are the 95% confidence intervals from a Beta distribution.

Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

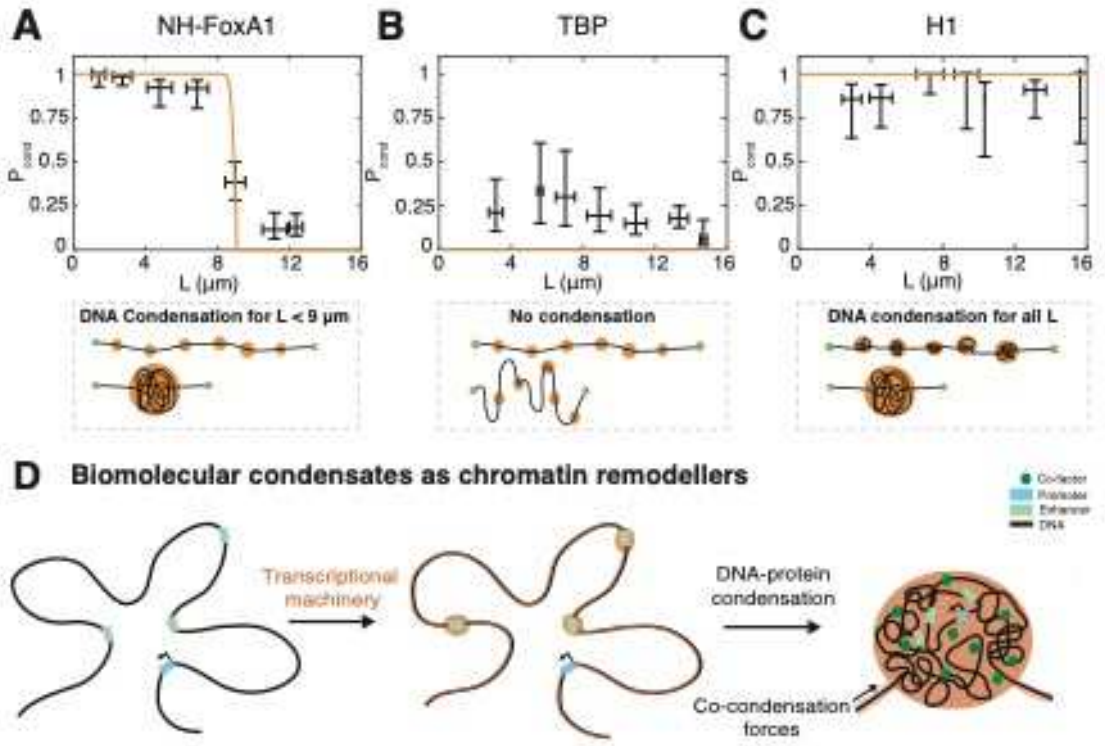


Figure 4: Universality of protein-DNA co-condensation. Probability to form a protein-DNA co-condensate for NH-FoxA1 (A), Tata-box-binding protein (B), and Somatic linker histone H1 (C). P_{cond} is computed from local correlation data with n=361 condensates for NH-FoxA1 (A), n=247 condensates for Tata-box-binding protein (B), and n=101 for H1 (C). Scale bar=2 μm. The error bars for the end-to-end distance are SD and the P_{cond} error bars are the 95% confidence intervals from a Beta distribution. We found that NH-FoxA1 condensed DNA less strongly than WT-FoxA1, TBP could not condense DNA for any end-to-end distance, and H1 condensed DNA for all measured end-to-end distances. (D) Biomolecular condensates generate condensation forces that could serve to recruit transcriptional regulators, and potentially remodel chromatin at physiologically relevant force scales in order to properly regulate transcription. See Figure 2 in the Supplementary Information for representative protein-DNA images of NH-FoxA1, TBP, and H1.

Methods

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

584
585 Cloning and protein purification

586 FoxA1-mCherry was introduced into a bacterial expression vector with an N-
587 terminal His_6 tag using Gateway cloning. Unlabeled FoxA1 was cloned and purified
588 the same way. This vector was transformed into T7 express cells (enhanced BL21
589 derivative, NEB C2566I), grown to OD~0.4-0.8, whereupon we added 1 mM IPTG
590 and expressed His_6-FoxA1-mCherry for 3-4 hours at 37°C. We thawed frozen
591 pellets in binding buffer (1xBB) that contained 20 mM Tris-HCl (pH=7.9), 500 mM
592 NaCl, 20 mM Imidazole, 1 mM $MgCl_2$, supplemented with protease inhibitors and
593 Benzonase. The redissolved pellets were lysed and clarified via centrifugation.
594 Discarding the supernatant, we resuspended the pellets in 1xBB + 6 M Urea, spun,
595 collected the supernatant and poured it over an IMAC column, eluting the protein
596 with 1xBB+6 M Urea+250 mM Imidazole. We dialyzed overnight into storage
597 buffer (1xSB), 20 mM HEPES (pH=6.5), 100 mM KCl, 1 mM $MgCl_2$, 3 mM DTT, and
598 5 M Urea. Multiple dialysis rounds reduced the concentration of urea. Finally, the
599 protein was dialyzed into 1xSB+2 M Urea, spun-concentrated to 4-5 mg/ml (~50
600 μM), and then snap-frozen nitrogen and stored at -80°C. NH-FoxA1-mCherry and
601 RR-FoxA1-mCherry were obtained following[15] using the Q5 Site-Directed
602 Mutagenesis Kit. The truncation constructs were generated using restriction
603 digestion-ligation approaches coupled with PCR. We used Alexa-488-labeled
604 somatic linker histone H1 purified from calf thymus (H-13188, ThermoFisher). To
605 purify mCherry-B4, the gene (Genscript) was cloned into a bacterial expression
606 vector with N-terminal His_6 and mCherry tags, transformed into T7 express cells,
607 grown to OD~0.7 and supplemented with 0.8 mM IPTG and expressed at 37°C for
608 four hours. Resuspending the pellets in lysis buffer, 1xPBS with 500 mM NaCl, 1
609 mM DTT plus protease inhibitors and Benzonase, we then lysed the cells, collected
610 the supernatant, ran the supernatant over an IMAC column, and eluted the protein
611 with lysis buffer+250 mM Imidazole. The protein was dialyzed into 1xPBS+500
612 mM NaCl overnight, spun-concentrated, snap-frozen, and stored at -80 °C. We
613 purified labeled versions of Tata-box binding protein and Gal4-VP16 using similar
614 purification strategies. Both vectors—His_6-MBP-eGFP-zTBP and His_6-Gal4-GFP-
615 VP16—were transformed into T7 express cells, grown to OD~0.6, whereupon we
616 added 0.2 mM IPTG and expressed overnight at 18 °C. We lysed the cells into

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

buffer containing 50 mM Tris-HCl (pH=8.0), 1 M NaCl, 10% glycerol, 1 mM DTT, 1 mM $MgCl_2$ supplemented with protease inhibitors. For subsequent steps, 10 μM $ZnSO_4$ was added to buffers for the Gal4-VP16 purification. After lysis, we added NP40 to 0.1% and clarified via centrifugation. We performed a polyethylenimine precipitation to precipitate DNA and then an ammonium sulfate precipitation to recover the protein, resuspending the precipitated proteins in buffer containing 50 mM Tris-HCl (pH=8.0), 1 M NaCl, 10% glycerol, 1 mM DTT, 0.1% NP40, and 20 mM imidazole and clarified the soluble fraction via centrifugation. We poured the lysate over an IMAC column and eluted the protein using 2xPBS, 250 mM imidazole, 10% glycerol, and 1 mM DTT. We pooled protein fractions and dialyzed TBP overnight into 20 mM HEPES pH=7.7, 150 mM KCl, 10% glycerol, and 1 mM DTT and Gal4-VP16 into 20 mM Hepes (pH=7.7), 100 mM KCl, 50 mM Sucrose, 0.1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM DTT, and 10 μM $ZnSO_4$. We then spun-concentrated the proteins, snap-froze using liquid nitrogen, and stored at -80°C.

DNA functionalization, cover slip PEGylation, and DNA micro-channel preparation

To biotinylate DNA purified from λ-phage (λ-phage DNA), we followed the protocol given in[19]. Each end of the biotinylated λ-phage DNA had two biotin molecules. To PEGylate the cover slips and prepare the DNA microchannels we followed the protocol given in[19].

DNA and protein imaging

We fluorescently stained immobilized DNA strands with 10 nM Sytox Green in Cirillo buffer (20 mM HEPES, pH=7.8, 50 mM KCl, 2 or 3 mM DTT, 5% glycerol, 100 μg/ml BSA). For experiments with H1 and TBP, we imaged DNA using 25 nM Sytox Orange. We used protein concentrations of 10 nM. We used a Nikon Eclipse microscope with a Nikon 100x/NA 1.49 oil SR Apo TIRF and an Andor iXon3 EMCCD camera using a frame-rate of 100 – 300ms. A highly inclined and laminated optical sheet (HILO) was established using a Nikon Ti-TIRF-E unit mounted onto the microscope stand.

Optical tweezer measurements

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We performed optical tweezer experiments using a C-Trap G2 system (Lumicks) in a microfluidics flowcell (Lumicks), providing separate laminar flow channels. For each experiment, we trapped two streptavidin-coated polystyrene beads (Spherotec SVP-40-5). Once trapped, we moved these beads to a channel containing biotinylated λ-phage DNA (Lumicks) at a concentration of 0.5 μg/ml, whereupon we used an automated "tether-finder" routine to capture a single molecule between the two beads. Once a single λ-phage DNA molecule was attached to the two beads, we moved the trapped beads to a buffer-only channel (containing Cirillo buffer: 20 mM HEPES, pH=7.8, 50 mM KCl, 3 mM DTT, 5% glycerol, 100 μg/ml BSA). In the buffer-only channel, we fixed the molecule's end-to-end distance at either L=6 or 8 μm. We then moved the tethered DNA to a channel containing 150 nM FoxA1 in Cirillo buffer or another buffer-only channel (as a control) and tracked the force and imaged the FoxA1-mCherry fluorescence for 100 seconds.

Bulk phase separation assays

We performed bulk phase separation assays with FoxA1-mCherry, NH-FoxA1-mCherry, and somatic linker histone H1. The storage buffer for FoxA1 and NH-FoxA1 was 20 mM HEPES (pH=6.5), 100 mM KCl, 1 mM $MgCl_2$, 3 mM DTT, and 2 M Urea. The storage buffer for H1 was 1xPBS. For FoxA1, we combined 6 μl of FoxA1 (at 50 μM) and 1 μl of 20% 30K poly-ethylene glycol (PEG). For NH-FoxA1, we combined 9 μl and 1 μl of 20% 30K PEG. For H1, we combined 9 μl H1 and 1 μl 100 μM 32-base pair ssDNA. We prepared flow channels with double-sided tape on the cover slide and attached a PEGylated cover slip to the tape. We imaged the condensates using spinning disk microscopy and a 60x objective.

FoxA1 molecule number estimation

To estimate the number of FoxA1-mCherry molecules per condensate, we quantified the intensity of single FoxA1-mCherry molecules bound non-specifically to the slide. Around each segmented spot of DNA-independent FoxA1 intensity, we cropped an area of 10x10 pixels, performed a background subtraction and summed the remaining intensity in the cropped area. To determine the contribution of the background, the same method was applied to

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

682 10x10 pixel areas void of FoxA1 signal intensity. The resulting distribution of
683 these integrated signal intensities reveals consecutive peaks that are evenly
684 spaced by an average intensity of about 400 a.u., allowing us to calculate the
685 number of molecules. This approach should interpreted as a lower bound
686 estimate of the number of FoxA1-mCherry molecules per condensate, as it
687 neglects effects such as fluorescent quenching[27].
688
689 Hydrodynamic stretching of DNA
690 DNA molecules bound at only one end to the slide were hydrodynamically
691 stretched using a constant flow rate of 100 µl/min of 0.5 nM FoxA1-mCherry in
692 Cirillo buffer with 10 nM Sytox Orange. The flow rate was sustained for tens of
693 seconds using a programmable syringe pump (Pro Sense B.V., NE-501).
694
695 Strand length calculation
696 To calculate the end-to-end distance, we generated time-averaged projections of
697 FoxA1 and DNA and integrated these projections along the strand's orthogonal
698 axis. To find the profile's "left" edge, we computed the gradient of the signal and
699 determined the position where the gradient went through a threshold (defined as
700 0.2). We then took all the points from the start of the signal to this position,
701 performed a background subtraction, and fit an exponential to these points. To
702 ensure that we included the entire DNA signal, we defined the fitted threshold for
703 both the left and the right edges as three-quarters of the value of the fitted
704 exponential value at the point when the gradient had gone through the intensity
705 threshold. Using this fitted threshold, we computed the position values for the left
706 and the right sides, and computed the end-to-end distance as the difference
707 between these two positions.
708
709 Global cross-correlation analysis
710 We generated time-averaged projections from movies of both FoxA1 and DNA, and
711 then summed the intensities in the orthogonal axis to the strand, generating line
712 profiles. We then calculated the strand length and cropped both the FoxA1 and
713 DNA line profiles from the edges of the strand. We then subtracted the mean value
714 from these cropped line profiles, normalized the amplitudes of the signals by their

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Euclidean distances, and computed the zero-lag cross-correlation coefficient of the normalized signals, which we defined as "Correlation": $R(\tau = 0) = \sum_{n=1}^{N} \overline{x_n y_n}$, where τ is the number of lags, N is the number of points in the normalized FoxA1 and DNA signals, $\overline{x_n}$ is the nth entry of the normalized FoxA1 signal, and $\overline{y_n}$ is the nth entry of the normalized FoxA1 signal. In general, Correlation values range from -1 to 1, but in our experimental data the values range from roughly 0 to 1, where 1 represents the formation of DNA-FoxA1 condensates and 0 represents the formation of only FoxA1 condensates (no DNA condensation).

DNA envelope width calculation

To compute the DNA envelope width, we first generated time-averaged projections from movies of FoxA1 and DNA. We then selected segments of the strand that did not contain FoxA1—regions of non-condensed DNA. Using these segments, we extracted a line profile of the DNA signal orthogonal to the strand that gave the maximum width. We then subtracted off the background of the DNA profile, normalized the signal's amplitude using the Euclidean distance, and fit a Gaussian. We defined the DNA envelope width as $\sqrt{2}\sigma$, which represents the square root of two times the standard deviation of the fitted Gaussian. The theoretical diffraction limit is calculated using the Rayleigh criterion, a measure of the minimal resolvable distance between two point sources in close proximity for a given set of imaging conditions: $d = \frac{0.61\,\lambda}{NA}$, where λ represents the imaging wavelength and NA is the numerical aperture. For our imaging setup, d = 0.2 μm, which is approximately 2σ of the fluorescent source from the DNA. As the DNA envelope width is defined as $\sqrt{2}\sigma$, our "diffraction limit" as given by the dashed line in Fig. 1f is given as 0.14 μm.

Condensate volume analysis

To calculate condensate volumes, we generated time-averaged DNA-FoxA1 projections and then localized the peaks of the DNA condensates. Using the peak locations, we extracted background-subtracted one-dimensional profiles of the DNA condensates in the orthogonal axis to the strand—these profiles went through the peak location. We fit Gaussians to these profiles without normalizing

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the amplitude. To define the radii of the condensates, we computed the gradient of the fitted Gaussians and defined the condensate "edges" as when the absolute value of the gradient of the Gaussian function gradient went through a threshold value (defined as one, and determined by comparing with fluorescence). Assuming condensates are spherical, we computed the condensate volume as $V = \frac{4}{3}\pi R^3$, where R is the condensate's radius. To compute a condensate volume for strands with multiple condensates, we simply added up the volumes for each condensate.

Condensed DNA amount analysis

To compute the amount of condensed DNA, L_d, we generated time-averaged projections of DNA and FoxA1 signals, integrating the DNA signal in the orthogonal direction to the strand. We then defined condensed vs non-condensed DNA with $\text{Threshold}_{\text{drop}}$: the median value of the profile plus a tolerance. Intensity values below $\text{Threshold}_{\text{drop}}$ were defined as pixels of non-condensed DNA, and intensity values above $\text{Threshold}_{\text{drop}}$ were defined as pixels of condensed DNA. This assumption was also consistent with the measured FoxA1 signal, where FoxA1 signals clearly localized to regions of condensed DNA, as defined by the $\text{Threshold}_{\text{drop}}$. The tolerance value was used to suppress artefactual fluctuations of the non-condensed DNA signal in the neighborhood of the median. To optimize the tolerance value, we assume that L_d as a function of L is linear for lower values of L (<5 μm) with a y-intercept equal to the contour length of the DNA molecule (16.5 μm), as this is consistent with our theoretical description. We plotted the y-intercepts of the linear fits as a function of tolerance and found that tolerance=500 gives a y-intercept equal to 16.5 and generates DNA-FoxA1 condensates up to 10 μm consistent with our data and analysis (Extended Data Fig. 6). To calculate the DNA length contained within the droplet, we integrated the intensities from pixels above $\text{Threshold}_{\text{drop}}$, divided this value by the sum of the total intensity of the profile, and then multiplied this ratio by the contour length of λ-phage DNA, 16.5 μm. The non-condensed DNA length was calculated as simply the contour length minus L_d. We used the same tolerance = 500 for the NH-FoxA1 mutant analysis.

Force analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To calculate the force that the condensate exerts on the non-condensed DNA, we used the worm-like chain model, which relates λ-phage DNA's extension and force. Upon addition of FoxA1, the amount of non-condensed DNA reduces, and the extension changes as follows, $E = \frac{L}{L_c - L_d}$, where L_d is the amount of condensed DNA, L is the end-to-end distance, and L_c is the total contour length of the molecule. We then directly compute the force using the worm-like-chain model,

$$F = \kappa\left(\frac{1}{4}(1-E)^{-2} - \frac{1}{4} + E\right)$$

Condensate nucleation probability analysis

To calculate the probability of the formation of a DNA-protein condensate as a function of end-to-end distance, we localized the peaks of the FoxA1 condensates from time-averaged projections of FoxA1 and DNA. We then extracted 0.9 μm x 0.5 μm windows centered around the localized FoxA1 peaks of both the FoxA1 and DNA signals—with the window's long axis going with the strand and the short axis as orthogonal to the strand. We then computed the zero-lag normalized cross-correlation coefficient as follows:

$$C_{loc} = \frac{\sum\sum(f(x,y) - \mu_f)(g(x,y) - \mu_g)}{\sqrt{\sum f(x,y)^2 - {\mu_f}^2}\sqrt{\sum g(x,y)^2 - {\mu_g}^2}}$$

where $f(x,y)$ is the DNA, $g(x,y)$ is FoxA1, μ_f is the mean of the DNA image, and μ_g is the mean of the FoxA1 image. This generates values from -1 to 1. For FoxA1-mediated DNA condensation, the values for particular condensates are close to 1. When FoxA1 fails to condense DNA, owing to the morphology of the underlying DNA strand and the small number of pixels, we obtain values that range from -1 to roughly 0.5. To obtain a value for P_{cond} as a function of end-to-end distance, we selected a threshold of 0.75—C_{loc} values above the threshold are considered as "condensed" and values below would be considered "non-condensed". We binned the C_{loc} data in 2-μm increments as a function of end-to-end distance, and calculated P_{cond} by taking the number of "condensed" condensates and dividing it by the total number of condensates within the bin. The confidence intervals for P_{cond} in each respective bin are computed by computing the 95% confidence interval of a beta-distribution, which represents the probability distribution for a

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bernoulli process that takes into account the total number of successes with respect to the total number of attempts.

Parameter fitting of the thermodynamic description and confidence intervals

To fit α, we used a linear fit of the condensate volumes for individual strands as a function of L_d. The confidence intervals are the 95 per cent CI generated from directly fitting the points. To fit the surface tension γ and condensation free energy per volume υ, we minimized the error of the average $\overline{L_d}(L)$ and $P_{cond}(L)$ with respect to the data to optimize the parameter values. We used the normalized Boltzmann distribution $P(L_d) = \frac{e^{-\beta F(L,L_d)}}{\int_0^{L_c-L} e^{-\beta F(l)} dl}$ to calculate $\overline{L_d} = \int_0^{L_c-L} lP(l)dl$. To compute $P_{cond}(L)$, we localized the position of the local maximum in the free energy, L_d^{max} for a given L and then computed the probability to "not" nucleate a droplet from the Boltzmann distribution $\int_0^{L_d^{max}} P(l)dl$, which gives $P_{cond} = 1 - \int_0^{L_d^{max}} P(l)dl$. To minimize the error, we binned the data in 2-μm-width bins. For each "binned" mean for both condensed DNA and condensation probability, we computed the squared residual of the mean value with respect to the theoretical expression. For residuals calculated from $\overline{L_d}(L)$, we normalized each residual by the squared standard error of the mean, and then summed the normalized residuals to obtain the error. For residuals calculated from $P_{nuc}(L)$, we normalized each residual by the variance of the beta distribution, $P_{nuc}{}_{cond}^{\sigma^2} = \frac{(1+k)(1-k+N)}{(2N^2(3+N))}$ and then summed the normalized residuals to obtain the error. For the global error, we simply added the error from both deviations in $\overline{L_d}(L)$ and $P_{cond}(L)$. We then iterated through a range of values for (γ, υ) and computed the total error associated with each set of parameter values, exponentiated the negative values of the total error matrix, and computed the largest combined value to select the parameter values. To calculate the parameters' confidence intervals, we obtained one-dimensional profiles of the integrated exponentiated total error for υ as a function of γ and γ as a function of υ. The peaks of these profiles represented the values that we selected for our best-fit parameters. We assumed that these profiles represented probability distributions for parameter selection, and then calculated the left and right bounds where the area under the curve between these bounds

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

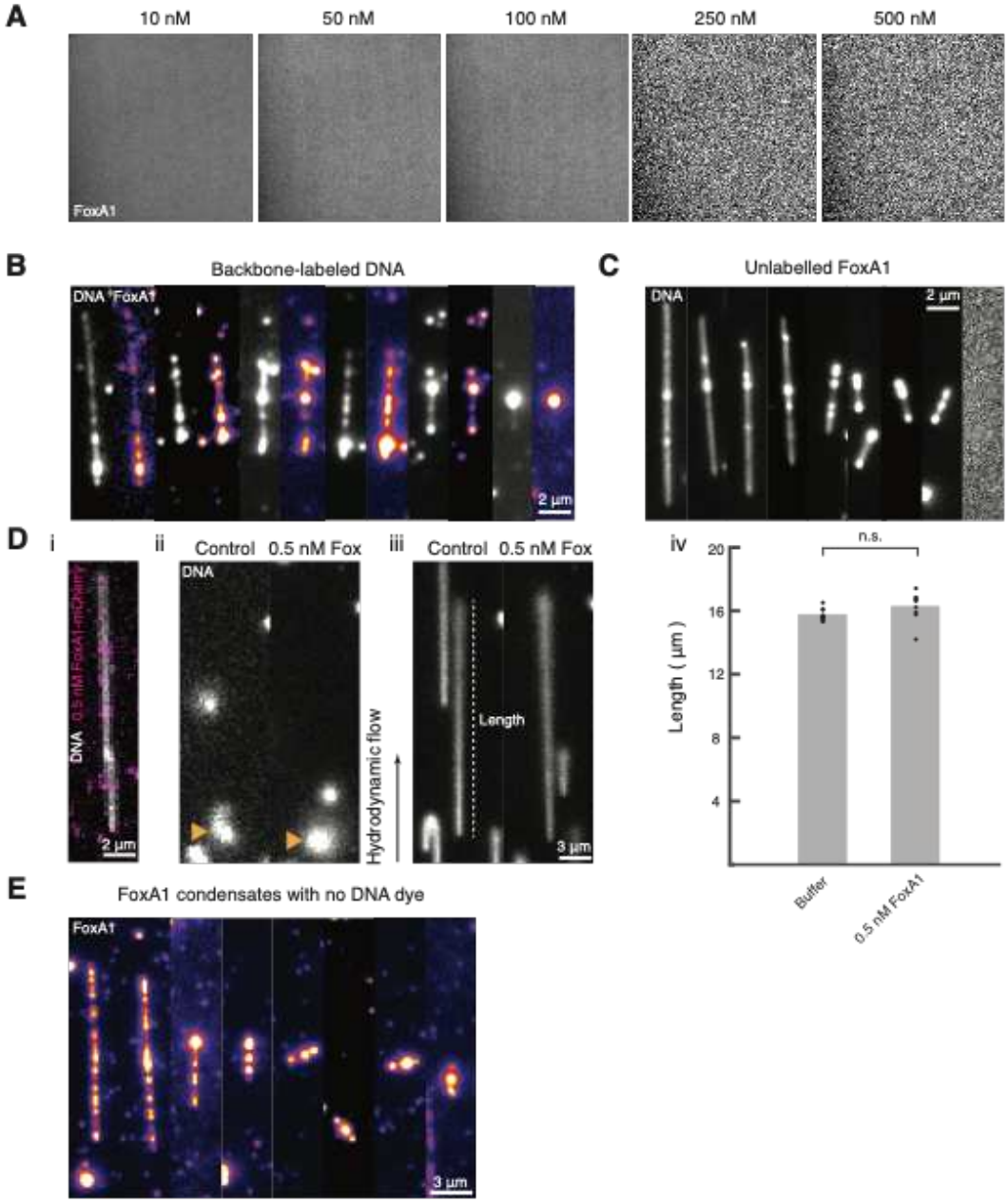
represented 95 per cent of the area. These left and right bounds represent the lower and upper values of our confidence intervals. To compute the 95 per cent confidence intervals for the force for each respective end-to-end distance value, we scanned through (γ, υ) parameter space and computed the value of L_d for each set of parameters. We then plotted these values against the probability that these parameter values were the "true" values—simply the probability from the exponentiated error matrix. Integrating the points under the Probability vs. L_d curve and dividing this by the total area under this curve, we generated a probability distribution function from which we could compute the 95% confidence intervals for L_d. Because the force was constant, to compute the confidence intervals for the force, we calculated the force using the worm-like chain model using corresponding L_d values for an end-to-end distance that retained FoxA1-mediated DNA condensation. To compute the confidence intervals for L_{crit}, we scanned through (γ, υ) parameter space and computed L_{crit} for each set of parameters. We then plotted L_{crit} values with the corresponding values from the probability that these parameter values were true (again, the exponentiated error matrix). Integrating the points under the Probability vs L_{crit} curve and dividing this by the total area under this curve, we generated a probability distribution function from which we could compute the 95% confidence intervals for L_{crit}.

Data availability statement: Source data files are made available for this paper. Data generated and analysed supporting the findings of this manuscript will be made available upon reasonable request.

Code availability statement: Code generated supporting the findings of this manuscript will be made available upon reasonable request.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Figure 1



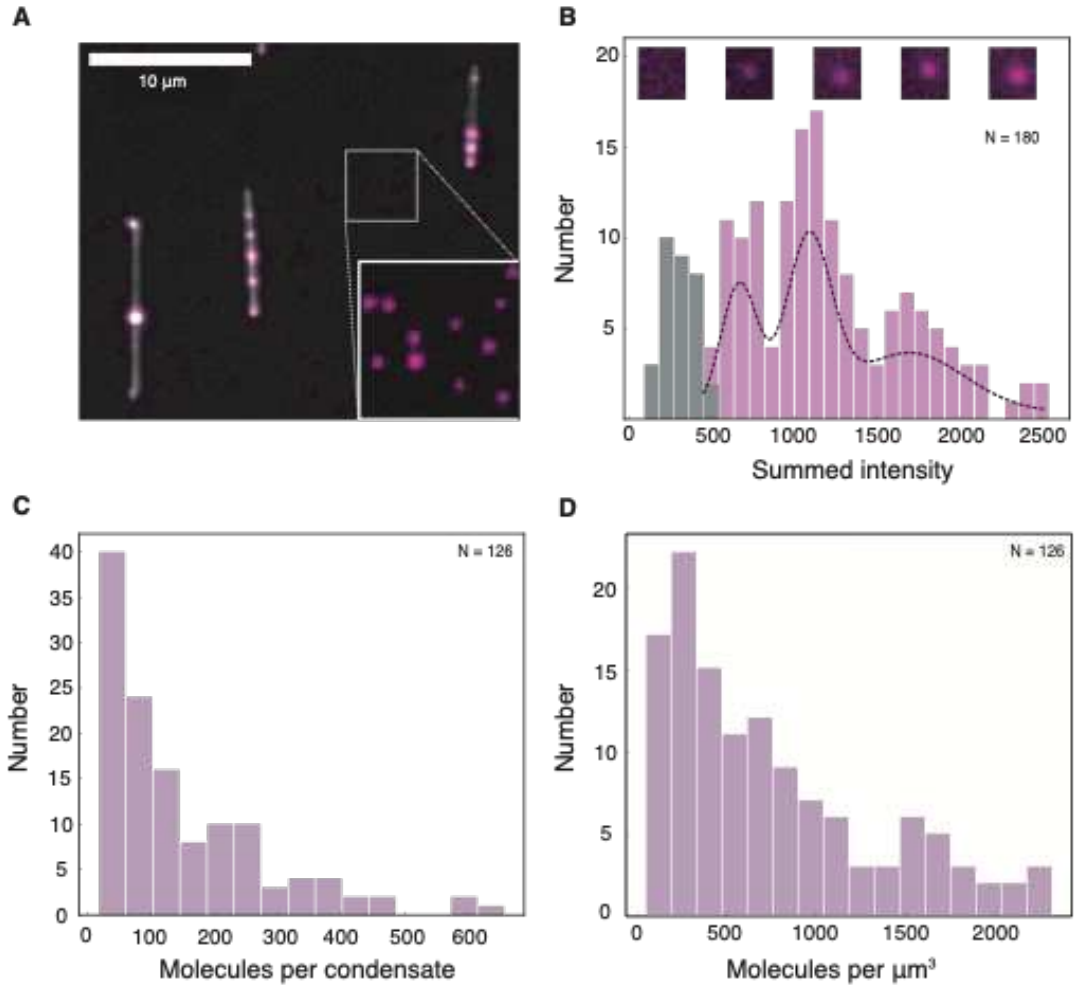
Extended Data Figure 1: Experimental controls for FoxA1-mediated DNA condensation. (A) Representative fluorescent images of FoxA1-mCherry in buffer (20 mM HEPES, pH=7.8, 50 mM KCl, 2 mM DTT, 5% glycerol, 100 μg/ml BSA) at different concentrations, 10-500 nM, in the absence of DNA reveals that FoxA1 does not form condensates in bulk at these concentrations. Using spinning disk microscopy and a 60x objective, we acquired images 70 μm x 70 μm in size with an exposure time of 250 msec and a time stamp of 500 msec to generate movies 30 seconds in duration. For all measured concentrations we generated n=3 movies and did not observe any FoxA1 condensation. (B) FoxA1-mCherry condenses λ-phage DNA molecules with Cy5 dye covalently attached to the phosphate backbone of DNA (Label-IT Nucleic Acid Labeling Kit, Cy5, Mirus). (C) Unlabeled FoxA1 condenses DNA (visualized with 10 nM Sytox Green). The rightmost panel is a representative image of the mCherry 561 nM imaging channel, revealing that

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the FoxA1 molecule does not have a mCherry fluorophore. (D) Sparse labeling of FoxA1 (0.5 nM) does not influence the persistence length and contour length of λ-phage DNA, as determined by hydrodynamic stretching (see Methods). (i) FoxA1 (purple) is sparsely bound to DNA (in grey), visualized with 10 nM Sytox Green. (ii) Snapshots of unstretched DNA molecules bound at only one end to the coverslip before hydrodynamic stretching in both control and 0.5 nM FoxA1 conditions. The yellow arrows point to the DNA molecules. (iii) Snapshots of stretched DNA molecules bound at one end to the coverslip during hydrodynamic stretching in both control and 0.5 nM FoxA1 conditions. (iv) Quantification of stretched DNA lengths in both control (n=10) and 0.5 nM FoxA1 (n=9) conditions reveals that there is no significant difference in the length under hydrodynamic stretching (unpaired t-test, p=0.11). (E) FoxA1 condensates imaged in the absence of DNA dye are consistent in size with that of FoxA1 condensates formed in the presence of DNA dye.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Figure 2



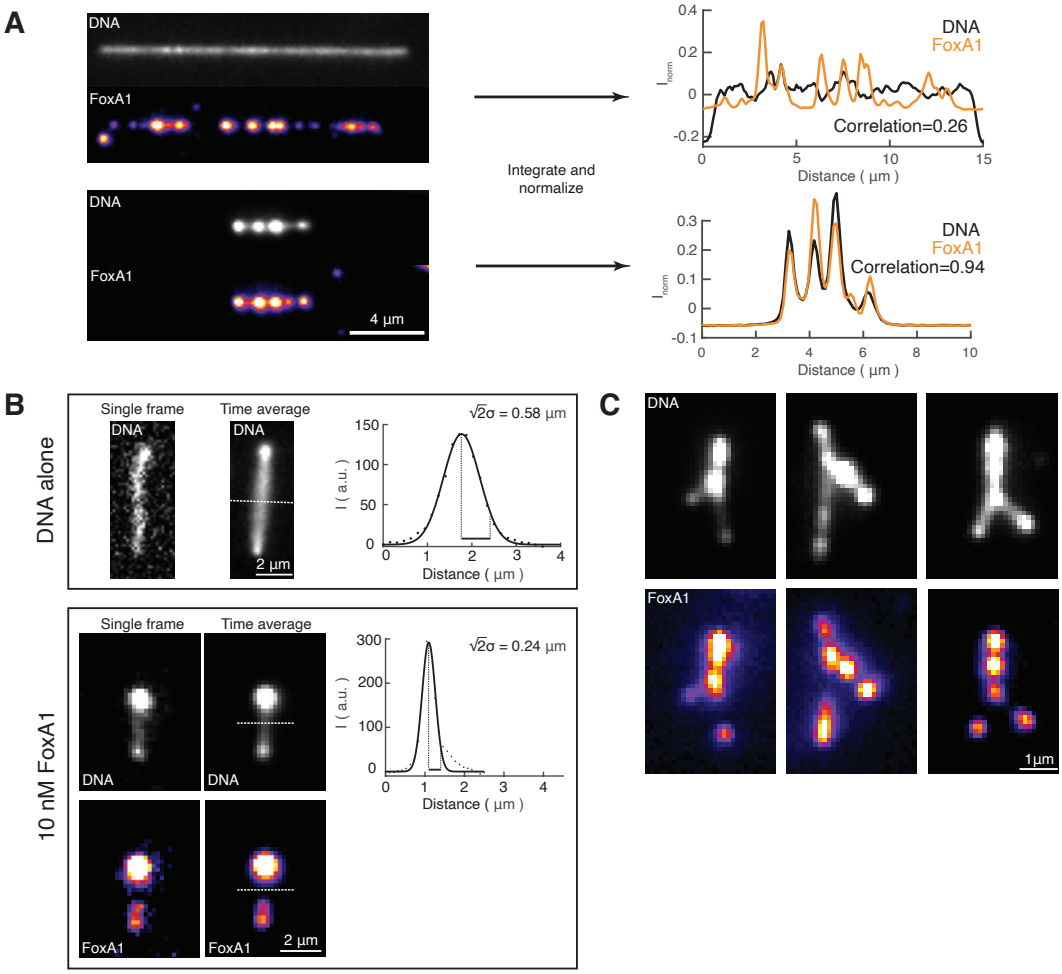
Extended Data Figure 2: Counting FoxA1 molecules in condensates. (A) Representative image of three DNA strands with FoxA1 condensates. The inset shows an area of the PEGylated glass slide void of DNA. Increased contrast reveals the presence of individual spots of FoxA1 non-specifically bound to the coverslip. (B) Histogram of integrated intensities of these DNA-independent FoxA1 to calibrate the amount of fluorescence per molecule. The grey bars represent the integrated background intensity of areas where no FoxA1 signal could be detected (maximum at 289 a.u.). Pink bars represent the integrated intensity of individual spots of DNA-independent FoxA1 signal. Black dotted line is a multi-Gaussian fit to the pink histogram, indicating consecutive peaks in the histogram at intensities of 683, 1096 and 1706 (a.u.), suggesting an integrated intensity of 400 a.u. per FoxA1 molecule. Representative images (10x10 pixels) of background (left) and individual DNA-independent FoxA1 spots used in this analysis are placed above the histogram according to their integrated signal intensity. (C) Histogram of the number of FoxA1 molecules in FoxA1 condensates on DNA, calculated based on an integrated intensity of 400 a.u. per FoxA1 molecule, determined in (B). The mean number of molecules is 150 per condensate. (D) Histogram of the density of FoxA1 molecules in the FoxA1-DNA condensates analyzed in (C). The mean value is 750 molecules per μm^3. These estimates represent lower bounds as previous studies

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

have demonstrated that fluorescent-based methods for estimating the number of molecules neglect effects such as quenching and can underestimate the number of molecules by as much as 50 fold[27].

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

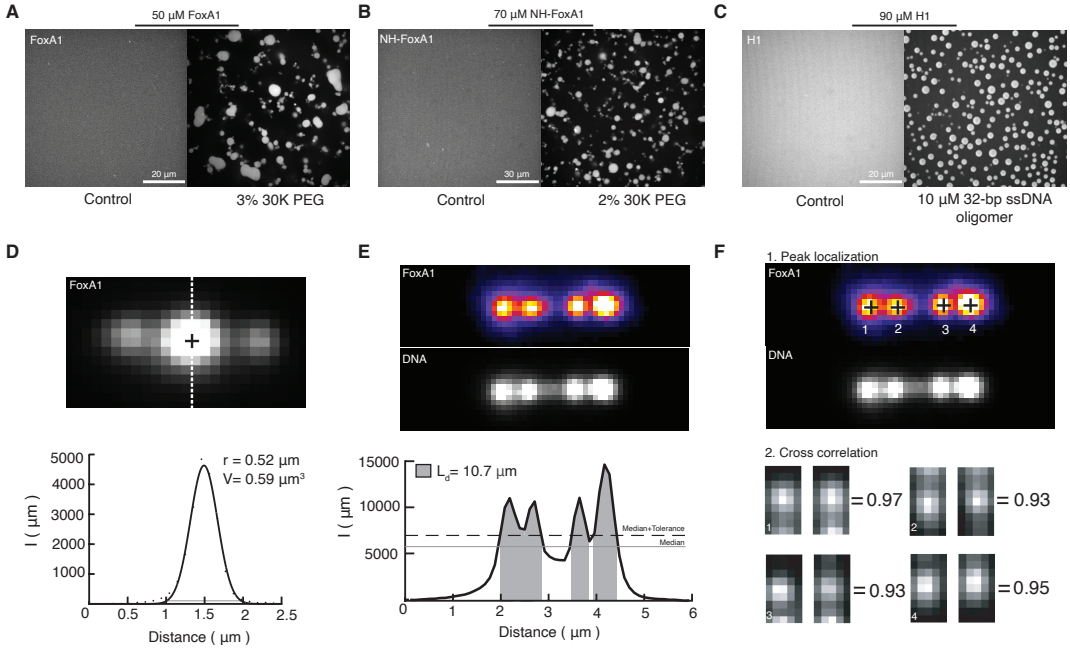
Extended Data Figure 3



Extended Data Figure 3: Quantification of FoxA1-mediated DNA condensation. (A) Global cross-correlation between FoxA1 and DNA reveals FoxA1-mediatd DNA condensation. Left, representative fluorescent time-averaged projections of DNA and FoxA1 at two different end-to-end distances. Integrating both the DNA and FoxA1 signals along the axis orthogonal to the long axis of the strand gave rise to line profiles, which we normalized, and then plotted as a function of distance (DNA in black and FoxA1 in orange). We then computed the zero-lag cross-correlation coefficient defined as "Correlation" (see Methods). (B) DNA envelope width measure measures the transverse fluctuation of non-condensed DNA. Top box: DNA alone condition. Bottom box: DNA+FoxA1 condition. For both conditions, we display representative fluorescent images of single frames and time-averaged projections of the DNA and FoxA1 signals. The white dashed line represents the maximum width of the DNA signal along the orthogonal axis of the non-condensed DNA. The black dots in the profile represent the background-subtracted points from the white dashed line, and the black line represents a Gaussian fit. The DNA envelope width was defined as $\sqrt{2}\sigma$, where σ is the standard deviation of the Gaussian fit. (C) Three representative examples of FoxA1-mediated zipping. These images are time-averaged projections of both FoxA1 and DNA.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

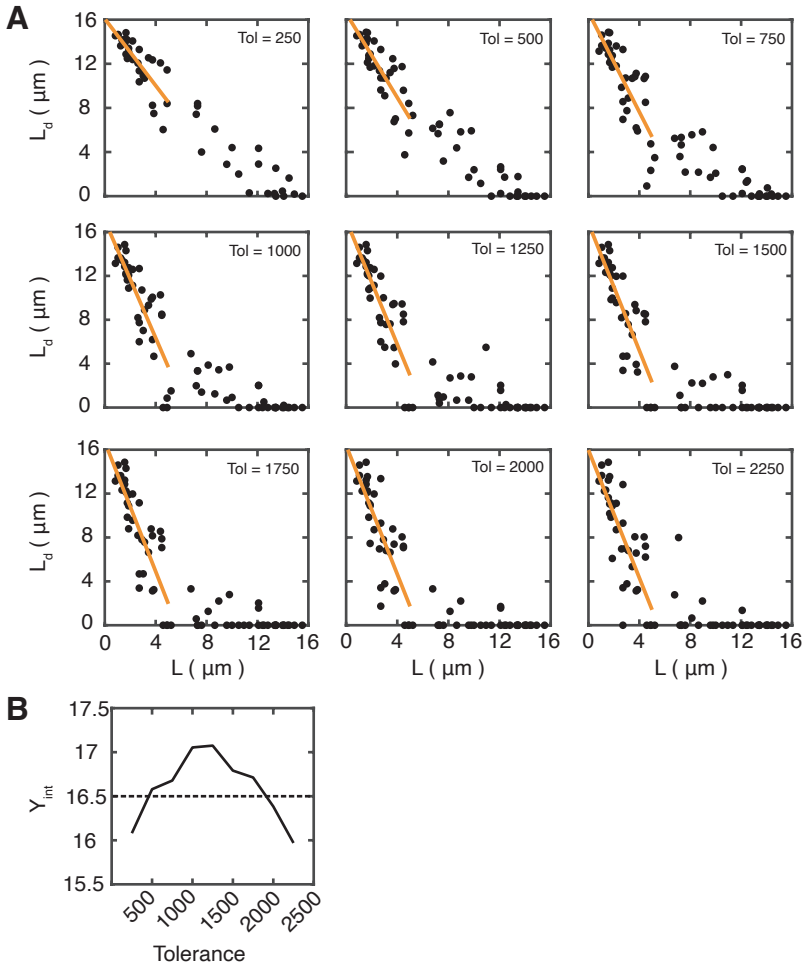
Extended Data Figure 4



Extended Data Figure 4: Bulk biomolecular condensate formation and quantification of condensate volume, condensed DNA length, and condensation probability. (A) Three per cent 30K PEG triggers FoxA1 condensate formation in bulk at 50 μM in storage buffer: 20 mM HEPES (pH=6.5), 100 mM KCl, 1 mM $MgCl_2$, 3 mM DTT, and 2 M Urea. (B) Two per cent 30K PEG triggers NH-FoxA1 condensate formation in bulk at 70 μM in storage buffer. (C) The addition of 10 μM 32-BP ssDNA oligomers nucleated droplets of H1 in bulk at 90 μM that exhibited features of liquid-like droplets consistent with literature[28,29]. These data demonstrate that H1-DNA form liquid-like condensates, which could be driven via transient cross-linking of H1 and DNA or H1-H1 interactions. Both mechanisms are accounted for in our free energy description. (D) Condensate volume quantification of a representative time-averaged projection of a FoxA1-DNA condensate, where the black cross is the condensate peak location and the white dashed line is the intersecting profile to measure the volume. Lower panel: the black dots are the profile's background-subtracted values and the solid black line is a Gaussian fit. The gray line represents the threshold value computed from the gradient of the Gaussian function that defines the edges of the condensate (see Methods). (E) Condensed DNA length quantification of a representative time-averaged projection of FoxA1 and DNA. Below: the integrated one-dimensional DNA profile is defined into condensed versus non-condensed regions using the median of the profile's median (gray) plus a tolerance (black dashed). (F) Local correlation quantification of a representative time-averaged projection of FoxA1 and DNA. The condensates were localized (black crosses) and then 0.9 μm x 0.5 μm boxes centered around these peaks were cropped. The correlations between the cropped regions of FoxA1 (left) and DNA (right) were then computed.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

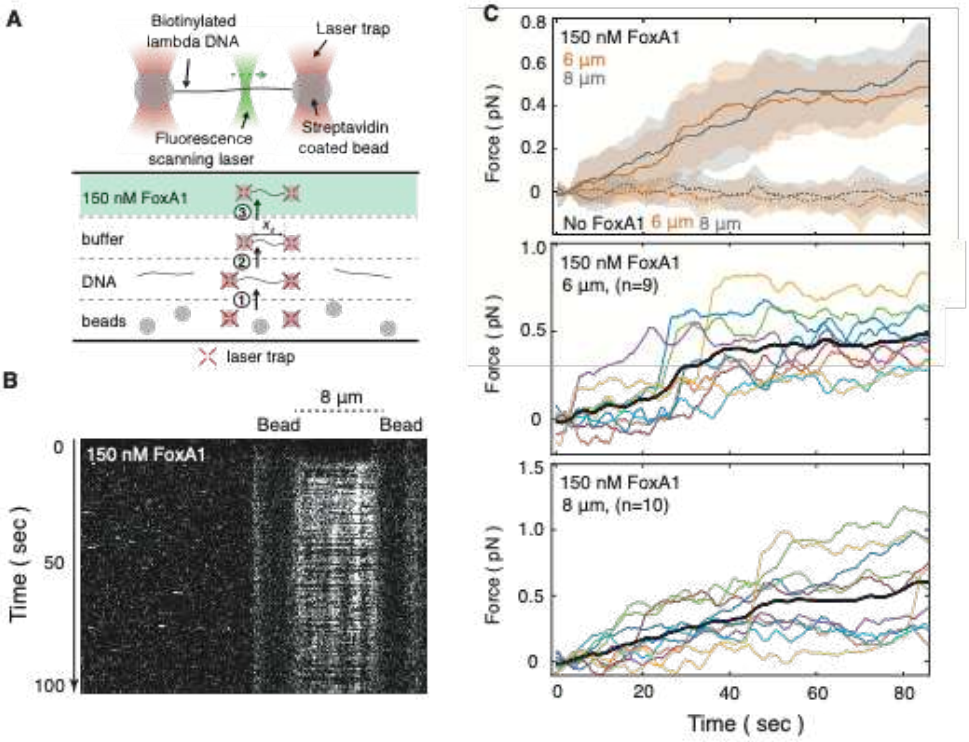
Extended Data Figure 5



Extended Data Figure 5: Tolerance value calculation. Quantification of the condensed DNA length as a function of end-to-end distance for a range of tolerance values. Condensed DNA length is computed by defining regions of condensed versus non-condensed DNA using a threshold composed of the signal's median value plus a tolerance. (A) Condensed DNA length is plotted as a function of end-to-end distance L for tolerance values from 250 to 2250 where the black dots represent the condensed DNA length for individual strands and the orange curve represent linear fits to these points for end-to-end distance below 5 µm. (B) Y intercept of the fitted linear curves. A tolerance=500 was selected as the y intercept was equal to the contour length of λ-phage DNA (16.5 µm) and gave FoxA1-DNA condensate formation up to approximately 10 µm, consistent with experimental observations (see Methods).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

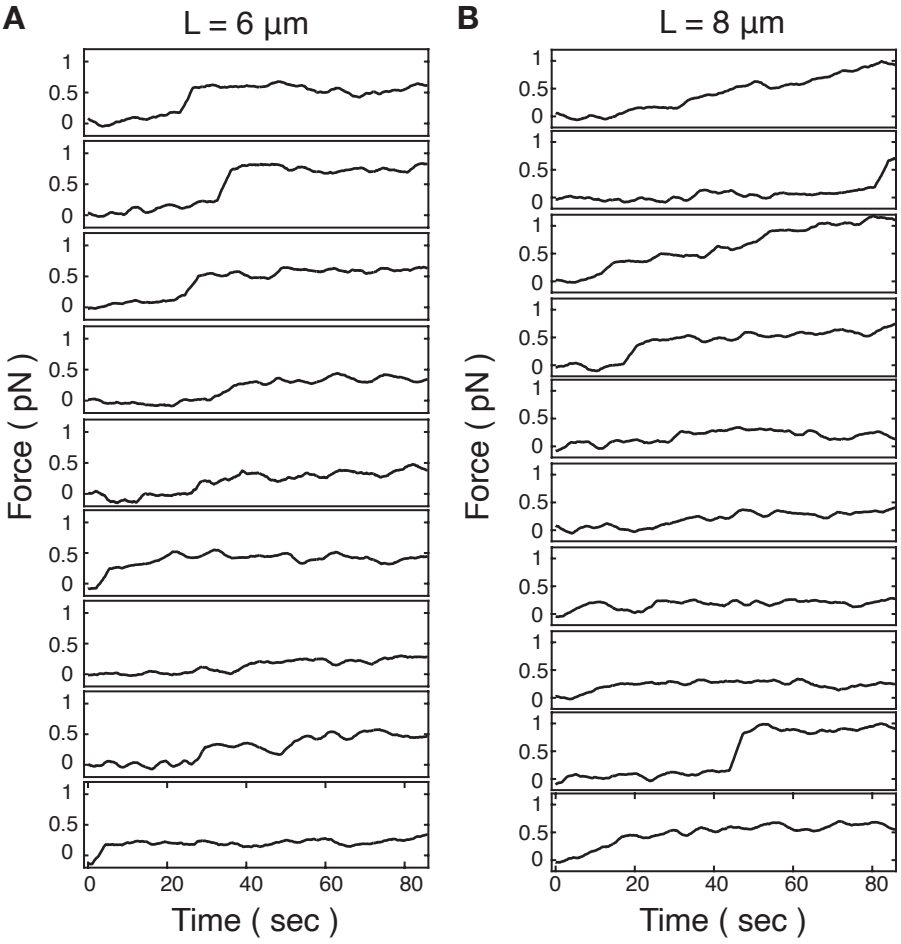
Extended Data Figure 6



Extended Data Figure 6: Optical tweezer measurements reveal that FoxA1 generates forces on the order of 0.4-0.6 pN. (A) Schematic outlining optical tweezer experimental design (see Methods). (B) Representative kymograph reveals that FoxA1 condensates co-localize with a single molecule of λ-phage DNA trapped between two beads at an end-to-end distance of 8 µm. (C) Force trajectories for single DNA molecules reveal forces on the order of 0.4-0.6 pN when in FoxA1-containing buffer. (Top panel) This panel displays the mean ± STD of force trajectories for each condition (n=9 for +FoxA1 with L=6 µm, n=10 for +FoxA1 with L=8 µm, n=10 for control with L=6 µm, and n=13 for control with L=8 µm.). This average force is slightly higher than what we measured in Fig. 3F using fluorescence, though a comparison of the relative errors reveals that both measurements give rise to comparable forces close to their respective detection limits and within the error bars. Additionally, the optical tweezer measurements were performed at a higher FoxA1 concentration—this was due to the large amount of tubing from the entry port to the flowcell in the custom-built Lumicks system, representing a considerable amount of surface for the protein to non-specifically bind to. We found that 150 nM FoxA1 was necessary to elicit a force response and to observe FoxA1 condensate formation on DNA. We conducted these measurements in the presence of 150 nM FoxA1 in Cirillo buffer 20 mM HEPES, pH=7.8, 50 mM KCl, 3 mM DTT, 5% glycerol, 100 µg/ml BSA (solid lines) and in the presence of Cirillo buffer only (hatched lines) at end-to-end distances of L=6 (orange) or 8 µm (grey). Individual force trajectories for λ-phage DNA in the presence of buffer containing 150 nM FoxA1 with an initial end-to-end distance of 6 µm (middle panel) and 8 µm (bottom panel) reveal jumps in force, consistent with a first-order phase transition. These trajectories are re-plotted for clarity in Extended Data Fig. 7.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

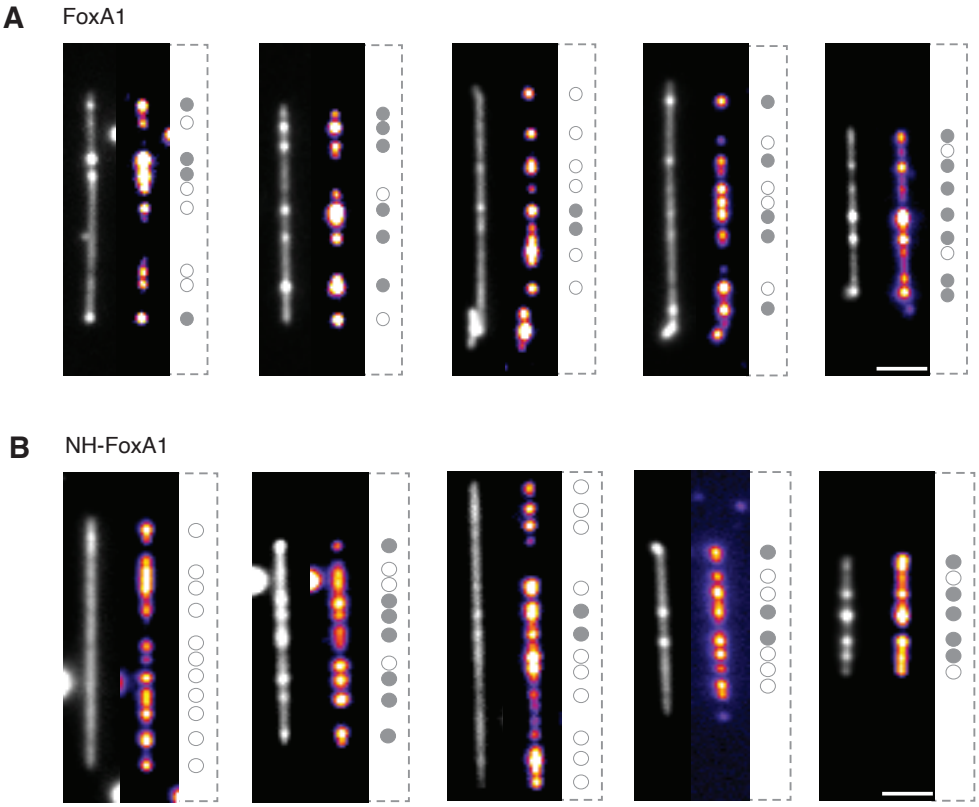
Extended Data Figure 7



Extended Data Figure 7: Individual temporal optical tweezer force measurements. Temporal force measurements from optical tweezers with an initial end-to-end distance of 6 μm (n=9 strands) (A) and 8 μm (n=10 strands) (B) in the presence of 150 nM FoxA1. These data are the same as in Extended Data Fig. 6c, and are re-plotted individually for clarity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

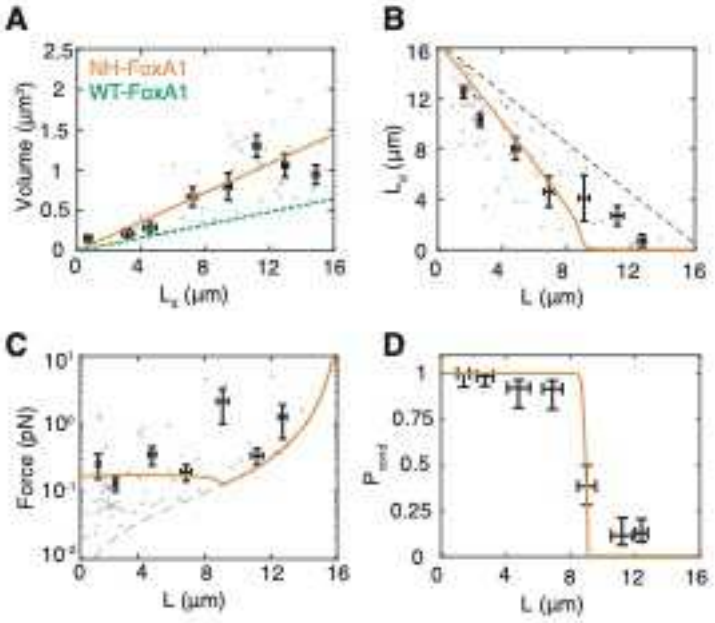
Extended Data Figure 8



Extended Data Figure 8: Bistability of FoxA1-mediated DNA condensation. (A) Representative time-averaged projections of DNA and FoxA1 signals show that FoxA1 condenses DNA in an all-or-nothing manner. On the right side of each pair of images, we localized the FoxA1 condensates and showed whether FoxA1 condenses DNA (filled-in gray circle) or not (open circle). Interestingly, there is a mixed population, revealing the bistable nature of the condensation process. (B) Representative images of condensation bistability for the sequence-specific DNA-binding mutant, NH-FoxA1. Scale bars = 2 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

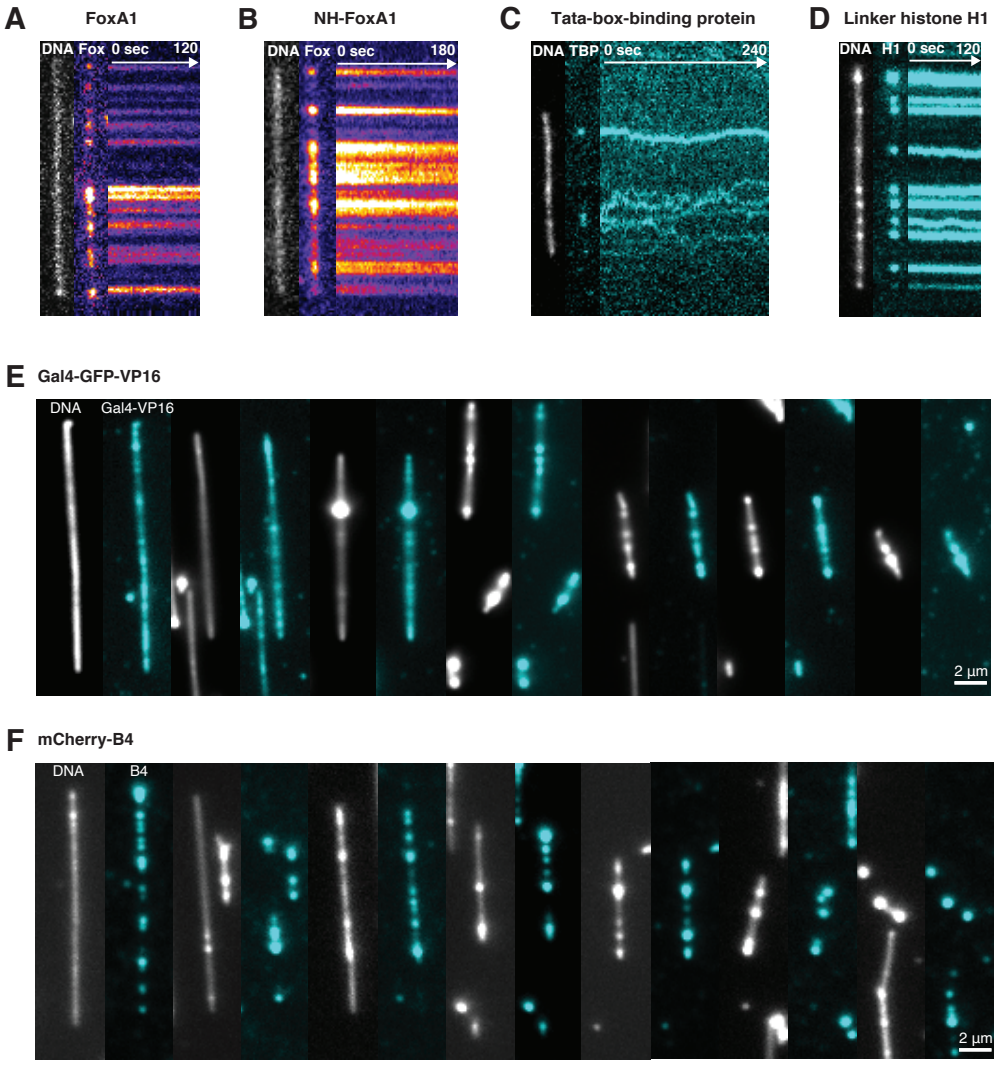
Extended Data Figure 9



Extended Data Figure 9: Quantification of NH-FoxA1-mediated DNA condensation. (A) Condensate volume as a function of condensed DNA length (L_d). The grey dots represent individual strands (n=47) and the data is binned every 2 μm (mean ± SEM). The individual data are points are fit with a linear curve with a slope of 0.09 μm² given in orange. The green dashed line is the WT-FoxA1 fit (slope=0.04 μm²). (B) Condensed DNA length as a function of end-to-end distance. The black dots represent individual strands (n=70) and the data is binned every 2 μm (mean ± SEM). The orange curve is the expression computed from the theoretical description with parameter values determined through error minimization (see Methods). The black hatched line represents the DNA's contour length (16.5 μm) minus the end-to-end distance. (C) The force that the condensate exerts on the non-condensed DNA as a function of end-to-end distance. The grey dots represent individual strands (n=68) and the data is binned every 2 μm (mean ± SEM). The orange curve is the expression computed from the theoretical expression of L_d versus L from panel B for the force. NH-FoxA1 generates forces at roughly 0.17 pN. The dashed black line represents the force exerted on the non-condensed strand when L_d=0. (D) Probability for NH-FoxA1 to form a DNA-FoxA1 condensate reveals a sharp transition at a critical end-to-end distance. Local correlations of individual FoxA1 condensates with DNA (Extended Data Fig. 4c) are calculated, binned into 2-μm-width bins, and P_{cond} is calculated (see Methods). There are a total number of n=361 condensates used for this analysis. The dashed lines represent the P_{cond} value as computed within the bin with ± SD for the strand's end-to-end distance. The confidence intervals for P_{cond} are computed by computing the 95% confidence interval of a beta-distribution (see Methods). The orange curve represents P_{cond} computed from the theoretical description with parameter values determined through error minimization.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.17.302299; this version posted July 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Figure 10



Extended Data Figure 10: Dynamics of DNA-binding proteins. (A) Representative images of FoxA1 condensates on DNA. The kymograph reveals FoxA1 condensates do not move on DNA. (B) NH-FoxA1 condensates remain stable on DNA and do not move. (C) TBP condensates exhibit diffusive-like behavior on DNA. (D) Similar to FoxA1 condensation, H1 condensates do not exhibit diffusive-like behavior on DNA. (E) Representative images of Gal4-GFP-VP16-mediated DNA condensation. DNA was imaged with 10 nM Sytox Orange. (F) Representative images of mCherry-B4-mediated DNA condensation. DNA was imaged with 10 nM Sytox Green.

Supplementary Information

1 Thermodynamic description of DNA-protein condensation

We consider the free energy associated with nucleating a condensate that contains DNA and FoxA1. The free energy of this process contains volume and surface contributions of the DNA-protein condensate as well as the free energy of the DNA polymer outside the condensate,

$$F(L, L_d) = F_d(L_d) + F_p(L, L_d), \tag{1}$$

where L is the end-to-end distance of the DNA, L_d is the length of condensed DNA, F_d is the free energy of the condensate, and F_p is the free energy of the DNA polymer outside the condensate. Assuming that the DNA co-condenses with the protein to form a dense condensed phase with defined volume fraction of DNA, the droplet volume and the length of condensed DNA are linearly related, $V = \alpha L_d$, or $R^3 = \frac{3\alpha}{4\pi} L_d$, where $1/\alpha$ describes the DNA packing density given as DNA length per condensate volume. We can then obtain the condensate free energy of nucleating a condensate as a function of L_d and end-to-end distance L as

$$F_d(L_d) = -\nu\alpha L_d + \gamma 4\pi \Big(\frac{3\alpha}{4\pi}\Big)^{2/3} L_d^{2/3}, \tag{2}$$

where ν is the condensation free energy per volume, and γ is the surface tension of the condensate. The free energy of the polymer $F_p(L, L_d)$ is related to the external force applied to pin the free DNA polymer and its associated chemical potential by

$$f_{\text{text}} = \left.\frac{\partial F_p}{\partial L}\right|_{L_d}, \;\; \mu_p = -a\left.\frac{\partial F_p}{\partial L_d}\right|_{L}, \tag{3}$$

where a is the length of a base pair. The force-extension relation for λ-phage DNA has been extensively studied previously, and here we use the phenomenological force-extension curve of the worm-like-chain model for λ-phage DNA (14) with contour length, L_c (for λ-phage DNA L_c = 16.5 μm). If a length L_d of the DNA is condensed, the extension of the non-condensed strand is $x = \frac{L}{L_c - L_d}$. The force on the strand then can be expressed as

$$f_{\text{text}} = \frac{k_B T}{P}\left(\frac{1}{4}\frac{1}{(1-x)^2} - \frac{1}{4} + x\right), \tag{4}$$

where k_B is the Boltzmann constant, T is the temperature, and P is the persistence length of DNA. For what follows we define $\kappa = \frac{k_B T}{P}$. From this expression of the force and its relation to the free energy of the DNA polymer (equation 3), we can obtain the free energy of the DNA polymer outside the condensate as $F_p(L, L_d) = \int_0^L dL' f_{\text{text}}(L', L_d)$, leading to

$$F_p(L, L_d) = \kappa\left(\frac{1}{4}\frac{(L_c - L_d)^2}{L_c - L_d - L} - \frac{1}{4}L + \frac{1}{2}\frac{L^2}{L_c - L_d} - \frac{1}{4}(L_c - L_d)\right) \tag{5}$$

The total free energy associated with nucleating a FoxA1-DNA condensate on a DNA strand reads:

$$F(L, L_d) = -\nu\alpha L_d + \gamma 4\pi\left(\frac{3\alpha}{4\pi}\right)^{2/3} L_d^{2/3} + \kappa\left(\frac{1}{4}\frac{(L_c - L_d)^2}{L_c - L_d - L} - \frac{1}{4}L + \frac{1}{2}\frac{L^2}{L_c - L_d} - \frac{1}{4}(L_c - L_d)\right) \tag{6}$$

The equilibrium between condensate and polymer is given by $\frac{\partial F}{\partial L_d} = 0$, which is equivalent to equilibrating the chemical potentials of the condensate and free polymer,

$$a\frac{\partial F}{\partial L_d} = \mu_d(L_d) - \mu_p(L, L_d) = 0, \tag{7}$$

with

$$\mu_p(L, L_d) = -a\kappa\Big(\frac{1}{4} + \frac{L^2}{2(Lc - Ld)^2} + \frac{(Lc - Ld)^2}{4(Lc - L - Ld)^2} - \frac{Lc - Ld}{2(Lc - L - Ld)}\Big) \tag{8}$$

Using the expression for the total free energy, we can vary the length L_d of condensed polymer and obtain profiles for the free energy as a function of L_d, which depend on the end-to-end distance L (see Fig. 3b). For L values close to 0—where the strand is not under tension—we observe that there is a minimum of F for L_d close to L_c. This means that, at this end-to-end distance, FoxA1 has mediated the generation of a FoxA1-DNA condensate using almost all of the DNA in the strand. As L increases, however, the local minimum shifts to lower values of L_d and ultimately F at the minimum becomes higher than the free energy without condensate $F(L_d = 0)$, giving rise to a branch of metastable states. For even higher L values, the metastable state disappears and the global minimum is at $L_d = 0$ (Fig. 3b). This sharp transition corresponds to a first-order phase transition. Simple scaling arguments are useful to generate intuition for the conditions necessary for condensate formation, and for the condensate to pull DNA. Briefly, there are three energy scales associated with this problem: the energy associated to create a droplet, which is $\nu\alpha L$; the surface energy of scale $4\gamma\pi\left(\frac{3\alpha}{4\pi}\right)^{2/3}L^{2/3}$; and lastly the energy scale associated to the non-condensed polymer $L\frac{k_BT}{P}$. First, to create a droplet, $\frac{\nu(4\pi)^{2/3}(\alpha L)^{1/3}}{4\pi 3^{2/3}\gamma} > 1$. Once condensation is favorable, in order for the droplet to pull DNA, $\frac{\nu\alpha P}{k_BT} > 1$. Notably, fitting the parameter values (see Methods) demonstrated that, at low L, the free energy gained by the system is on order of 1-2 k_BT, implying that stochasticity

is relevant for the condensation process. To account for the inherent stochastic nature of the condensation, we compute the probability of nucleating a DNA-protein condensate of size L_d using Boltzmann distributions from the corresponding energy profiles,

$$P(L_d) = \frac{e^{-\beta F(L,L_d)}}{\int_0^{L_c-L} dL'_d e^{-\beta F(L,L'_d)}} \tag{9}$$

where $\beta = \frac{1}{k_B T}$. To determine the relationship between L_d and L, we compute the mean L_d value of these Boltzmann distributions: $\bar{L}_d = \int_0^{L_c-L} L_d P(L_d) dL_d$ which then allows us also to calculate the magnitude of the condensation forces using the worm-like chain model given in Eq. (4).

2 Supplementary Information Figures

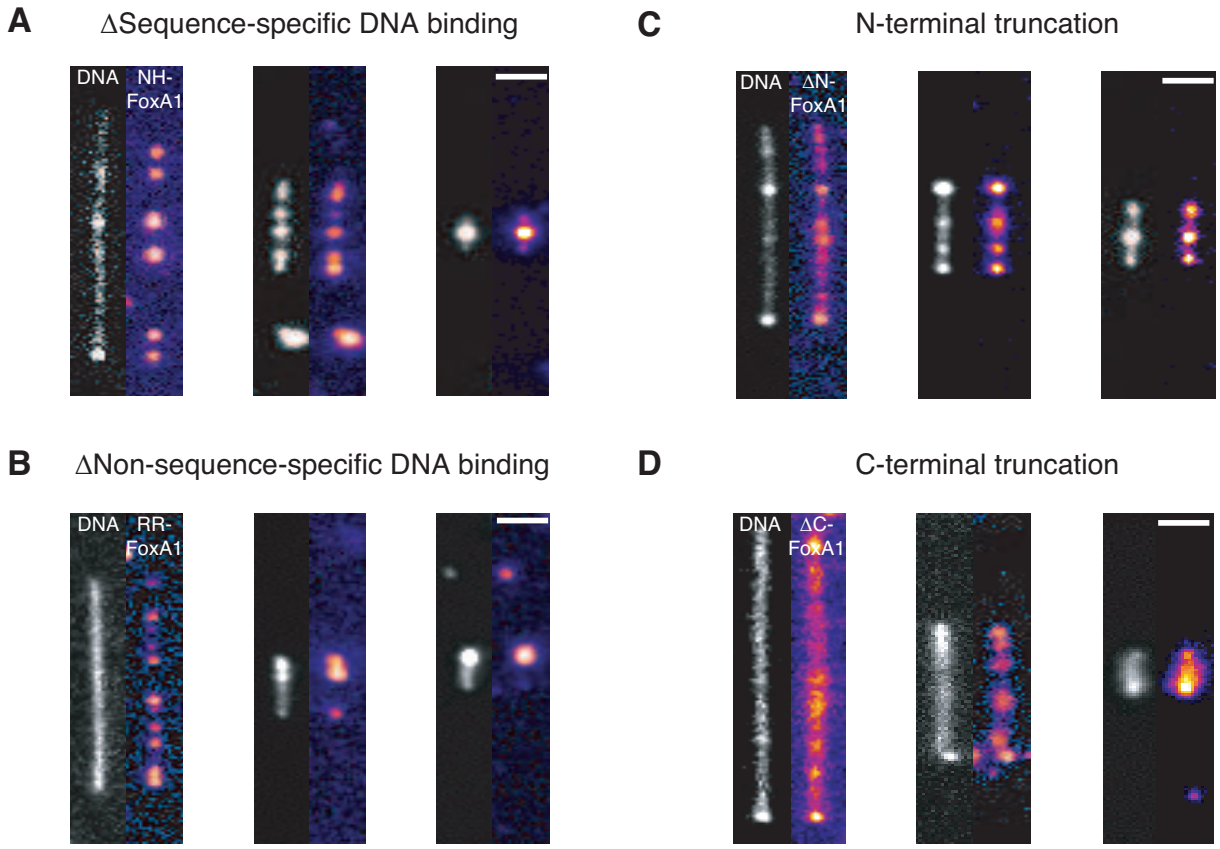


Figure 1: Representative images of sequence-specific-binding NH-FoxA1 mutant (A) non-sequence-specific-binding RR-FoxA1 mutant (B) N-terminal FoxA1 truncation (C) and C-terminal FoxA1 truncation (D). The scale bars are 2 μm. DNA is imaged with 10 nM Sytox Green. Note that the contour length of each DNA molecules is constant (16.5 μm) but the end-to-end distance is different.

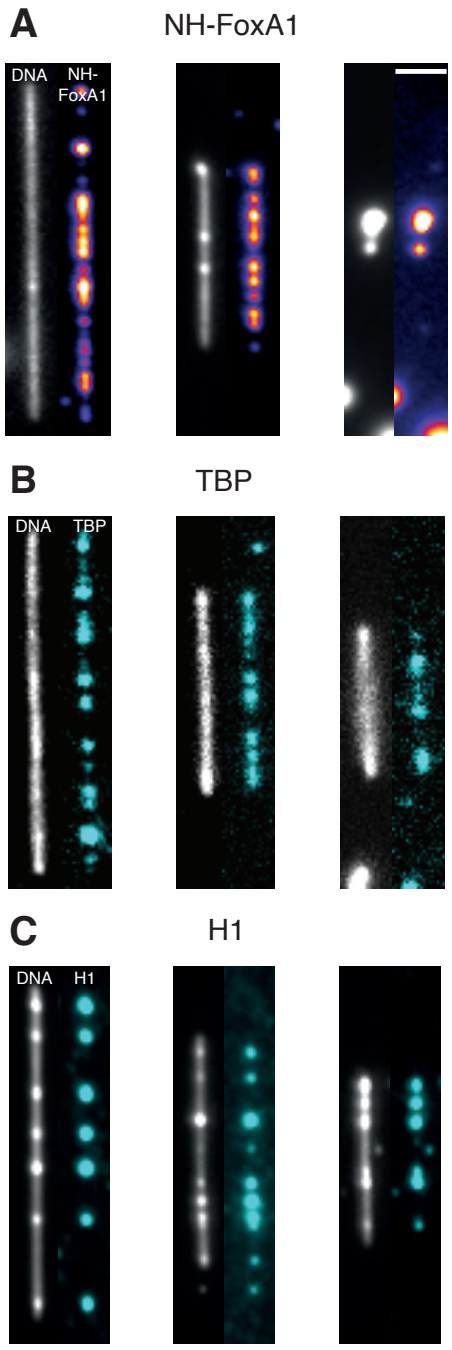


Figure 2: Representative images for NH-FoxA1 (A) Tata-box-binding protein (B) and somatic linker histone H1 (C). The images are time-averaged projections of movies for NH-FoxA1 and H1 but single images for TBP owing to TBP's diffusivity. The scale bar is 2 μm.