Title: Capillary forces drive pioneer transcription factor-mediated DNA condensation

Thomas Quail[1,2,3,4], Stefan Golfier[1,2,3,4], Maria Elsner[1,2,3,4], Keisuke Ishihara[1,2,3,4], 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, 01062 Dresden, Germany
*correspondence: brugues@mpi-cbg.de

Summary: Capillary forces are driven by interactions between liquids and surfaces[1]. These forces are crucial for many processes ranging from biology and physics to engineering, such as the motion of aquatic insects on the surface of water[2], modulation of the material properties of spider silk[3,4], and self-assembly of small objects and microstructures[5]. Recent studies have shown that cells assemble biomolecular condensates in a manner similar to phase separation[6]. In the nucleus, these condensates are thought to drive transcription[7-10], heterochromatin formation[11,12], nucleolus assembly[13,14], and DNA repair[15,16]. Here, we test if the interaction between liquid-like condensates and chromatin generates capillary forces, which 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, and theory to show that the pioneer transcription factor FoxA1 mediates the condensation of a DNA-protein phase via a mesoscopic first-order phase transition. Surprisingly, after nucleation, capillary 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 using capillary forces. Our work suggests that the physics of DNA and protein condensation driven by capillary forces provides a general regulatory principle for chromatin organization.

Main text: To investigate how transcription factors physically organize DNA, we attached linearized λ-phage DNA to a functionalized 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 factor that regulates tissue differentiation across a range of organisms[17,18] (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 suggests that DNA mediates protein condensation of a thin layer of FoxA1 on DNA.

In our assay, DNA molecules displayed a broad distribution of end-to-end distances (L), which are determined by the locations of the biotin-streptavidin attachment points of the λ-phage DNA molecules to the coverslip (Fig. 1c, d). This end-to-end distance governs the tension on the DNA molecule, as the contour length of each individual λ-phage DNA molecule is the same[19]. For DNA strands with end-to-end distances greater than approximately 10 μm, FoxA1 generated protein condensates on DNA (Fig. 1c). In these conditions, however, FoxA1 condensation did not influence the DNA molecule (Fig. 1c). 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). To quantify FoxA1-mediated DNA condensation, we measured the cross-correlation of FoxA1-DNA intensities as a function of end-to-end distance (Fig. 1d, e; Extended Data Fig. 1b). This measure ranges from roughly zero to one, where a value of zero corresponds to FoxA1-only condensation and a value of one corresponds to DNA-FoxA1 condensation. Consistent with the ability of FoxA1 to form DNA-FoxA1 condensates at low tensions (corresponding to short end-to-end distances), the cross-correlation decayed from one to zero with increasing end-to-end distances (Fig. 1e). Thus, FoxA1 can induce the formation of a DNA–protein-rich microphase in a tension-dependent manner.

The observation that FoxA1 drives DNA condensation suggests that it can overcome the entropic tension of the DNA molecule set by the end-to-end distance[19]. Similar to tightening an elastic band, incorporating DNA into the condensates will increase the tension on the non-condensed DNA by reducing the amount of available slack. This increased tension should reduce transverse DNA fluctuations of the non-condensed DNA, consistent with a reduction of slack in the molecule. To quantify these fluctuations, we first generated time-averaged projections from movies of DNA molecules with and without FoxA1. These projections are single images that display the area explored by the molecule due to its motion in a given time period (Fig. 1d, Extended Data Fig. 1c). From these projections, we extracted line profiles of the intensities of non-condensed DNA in the direction perpendicular to the long axis of the strand. Finally, we used Gaussian fits to these line profiles to define the DNA envelope width (Extended Data Fig. 1c). 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[19] (Fig. 1f). However, in the presence of FoxA1, the 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 significantly lower than in buffer conditions for all end-to-end distances (Fig. 1f). This result 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. 1d). 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 independent strands zipped together, generating a y-shaped morphology as the condensation of FoxA1 increased over time (Fig. 1g, Extended Data Fig. 1d). 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[20] or (ii) weak protein-protein interactions driven by disordered regions of FoxA1. Structurally, FoxA1 consists of a winged helix-turn-helix DNA-binding domain and two mostly disordered domains corresponding to the N and C termini[20,21]. 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. Point mutations in the sequence-specific DNA-binding region (NH-FoxA1[20]) had virtually no influence on DNA condensation activity (Fig. 2a). Although the presence of two point mutations in one of the non-sequence-specific DNA-binding regions (RR-FoxA1[20]) 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 inhibited DNA condensation activity (Fig. 2d). Thus, non-sequence-specific binding drives FoxA1 localization to DNA, and the disordered C terminus of FoxA1 drives DNA condensation.

Our results show that the disordered C terminus of FoxA1 is functionally required for FoxA1-mediated DNA condensation. Disordered regions of proteins influence condensate formation via weak protein-protein interactions[22]. Thus, our results support the hypothesis that FoxA1 condenses onto DNA to generate a DNA–protein-rich condensate via weak protein-protein interactions. This condensate phase nucleates only if the capillary forces it generates are larger than the entropic tension on the strand. To explore this possibility, we developed

a theoretical description based on the thermodynamics of a semi-flexible polymer partially forming a liquid-like condensate. In the present case, the semi-flexible polymer is DNA and the condensation is mediated by the transcription factor. The free energy of this process contains volume and surface contributions, and the capillary work to pull DNA into the condensate (Fig. 3a). The volume contribution, $\mu\frac{4}{3}\pi R^3$, is proportional to the volume of the condensate, where μ is the condensation free energy per volume and R is the radius of the condensate. We assume that DNA is fully collapsed inside the condensate and thus its volume is proportional to the adsorbed DNA contour length, $V = \alpha\, L_d$. Here, α describes the packing efficiency given as DNA strand length per cross-sectional area. The surface contribution to the free energy, $\gamma 4\pi R^2$, is characterized by the surface tension γ of the condensate. The work $W = \int_0^{L_d} f\, dl$ is the work that must be performed by capillary forces f to bring an amount L_d of DNA into the condensate. The free energy for fixed end-to-end distance L depends on the contour length of DNA in the condensate, L_d, as

$$F(L, L_d) = -\mu\alpha L_d + \gamma 4\pi \left(\frac{3\alpha}{4\pi}\right)^{\frac{2}{3}} {L_d}^{\frac{2}{3}} + \frac{\kappa L}{4}\left(\frac{l l_d}{(l-1)(l+l_d-1)} + \log\left(\frac{(l-1)^2}{(1-l_d)^4(l+l_d-1)^2}\right)\right)$$

where $l = \frac{L}{L_t}$, $l_d = \frac{L_d}{L_t}$, $\kappa = \frac{k_B T}{P}$, k_B is the Boltzmann constant, T is the temperature, and P is the persistence length of DNA (see Supplemental information). This free energy predicts a stochastic first-order phase transition for the formation of DNA–protein microphases (Fig. 3b) with a distribution of condensate sizes $P(L_d) \sim e^{-\beta F(L, L_d)}$ for fixed L (Fig. 3c). Furthermore, it predicts a sharp transition for DNA condensation controlled by the end-to-end distance and thus the tension on the DNA molecule. The first-order nature of this behavior implies regimes of hysteresis and bistability. Our theory also predicts that the capillary forces on the DNA are buffered by the condensate.

To test this theory, we first measured DNA condensate volumes and found that they increase linearly with the amount of condensed DNA (L_d), with α=0.04 ±

164 0.005 μm^2 (Fig. 3d, Extended Data Fig. 2a, Methods). This confirms that DNA is in
165 a collapsed conformation inside the condensates. Next, we simultaneously fit the
166 predictions to the average amount of DNA contained in the condensates (L_d), and
167 the probability of nucleating a DNA condensate (P_{cond}) as a function of end-to-
168 end distance (see Methods). We calculated L_d (Fig. 3e, Extended Data Fig. 2b,
169 Extended Data Fig. 3) and P_{cond} (Fig. 3f, Extended Data Fig. 2c) using the
170 Boltzmann probability distributions (Fig. 3c) from the free energy. Our fits agree
171 quantitatively with the data and show that the amount of DNA in the
172 condensates decreases with L until a critical end-to-end distance beyond which
173 DNA condensates do not form. Below this critical length, we observed that the
174 capillary force exerted by the condensate is buffered at 0.22 pN (0.18 – 0.27 pN
175 CI), consistent with the theory (Fig. 3f). Finally, P_{cond} exhibits a sharp transition
176 at $L = 10.7$ μm (9.2 – 10.8 μm CI), in agreement with a stochastic first-order
177 phase transition (Fig. 3f). Close to the transition point FoxA1-mediated DNA
178 condensation displayed bistability. This bistability was observed in strands that
179 contained multiple FoxA1 condensates, but where only some of them condensed
180 DNA (Extended Data Fig. 4a). Our fits allowed us to extract the physical
181 parameters associated with condensate formation, namely the condensation free
182 energy per volume μ = 5.9 pN/μm^2 (4.9 – 8.9 pN/μm^2 CI) and the surface tension
183 γ=0.05 pN/μm (0.04 – 0.38 pN/μm CI), see Methods section. These parameters
184 are consistent with previous measurements for *in vitro* and *in vivo*
185 condensates[23,24].
186
187 Our theory and experiments show that two key parameters govern DNA–protein
188 condensation behavior, namely the condensation free energy per volume (μ) of
189 the DNA–protein phase and the surface tension of the condensate (γ). We
190 reasoned that different DNA-binding proteins may exhibit a range of behaviors
191 depending on the values of these parameters. First we investigated the
192 sequence-specific DNA-binding region mutant (NH-FoxA1), which also
193 condensed DNA but to a lesser extent (Fig. 2a). Quantitatively, we found that the
194 surface tension of condensates formed with this mutant was roughly unchanged
195 compared to WT-FoxA1, γ=0.03 pN/μm (0.02 – 0.09 pN/μm CI), but the free
196 energy per volume of condensation was reduced consistent with reduced DNA

197 binding, μ=1.7 pN/μm^2 (1.6 – 2.1 pN/μm^2 CI), Extended Data Fig. 5b, Fig. 4a. This
198 was also reflected in an increase in the value of the packing efficiency parameter,
199 with α = 0.09 ± 0.02 μm^2, meaning that these condensates packed less DNA
200 compared to WT-FoxA1 given the same volume of DNA-protein condensate
201 (Extended Data Fig. 5a). We also observed that NH-FoxA1-mediated condensates
202 generated a force of 0.14 pN (0.09 – 0.15 pN, CI), lower than that for WT-FoxA1.
203 This moderate decrease in force was also reflected in a shift of the transition
204 point of condensate nucleation towards a lower end-to-end distance L_c = 8.9 μm
205 (8.8 – 9.0 μm, CI), Fig. 4a. In addition, NH-FoxA1 displayed bistable DNA-protein
206 condensation activity in the neighborhood of the transition point (Extended Data
207 Fig. 4b). Next, we examined the interactions of a different transcription factor,
208 Tata-Box-binding protein (TBP) with DNA. We found that TBP also formed small
209 condensates on DNA, but did not condense DNA even at the lowest imposed DNA
210 tensions (Fig. 4b). Instead, TBP performed a diffusive motion along the DNA
211 strand (Extended Data Fig. 6). This suggests that the free energy per volume of
212 condensation is positive, and DNA-protein condensation is not
213 thermodynamically favored. Lastly, we analyzed somatic linker histone H1, a
214 protein that is structurally similar to FoxA1. However, in contrast to FoxA1, one
215 of the known functions of H1 is to compact chromatin[25], so we expected H1 to
216 strongly condense DNA. Consistent with this, we found that H1 displayed a
217 dramatically increased DNA condensation activity compared to FoxA1,
218 condensing DNA for all measured end-to-end distances (Fig. 4c). Thus, we
219 propose that the competition between condensation free energy per volume of
220 the DNA–protein phase and surface tension regulate a spectrum of DNA
221 condensation activities, which may be tuned by the structure of transcription
222 factors.
223
224 Here, we show that FoxA1 can condense DNA to form a DNA–protein-rich
225 microphase. The DNA–protein condensates form through a force-dependent
226 first-order transition for forces below a critical value. This critical force, which is
227 on the order of 0.2 pN, is set by capillary forces that the DNA–protein condensate
228 exerts on the non-condensed DNA strand. These forces are similar in magnitude
229 to those recently measured for DNA loop extrusion of the order of 0.2-1 pN[26-30].

We speculate that these weak forces are of key relevance to the mechanics of chromatin organization. Thus, our work suggests that capillary forces can act as a chromatin remodeler 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[31]. In this context, DNA–protein condensates could act as scaffolds, recruiting transcriptional co-factors. Furthermore, capillary forces generated by the condensate could also pull 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. We find that, near the transition point, assembly and disassembly of DNA–protein condensates becomes highly stochastic. This is reminiscent of the rapid dynamics associated with the initiation and cessation of transcriptional bursts observed in vivo[32], suggesting stochastic condensate dynamics as a candidate for mechanisms that underlie transcriptional bursting.

DNA–protein condensation not only provides mechanisms to facilitate enhancer–promoter contacts, but could also play a more general role in DNA compaction such as in mitotic chromatid compaction[33], and the formation of chromatin compartments[11,12]. Our work highlights the role of capillary forces in the interaction between condensates and surfaces. Such forces may also be relevant beyond chromatin in other biological contexts, including membranes and the cytoskeleton. Owing to the tension-dependent nature of transcription-factor-mediated DNA–protein condensation, our work suggests that mechanics and capillary forces from liquid-like condensates could play a key and, as yet, underappreciated role in the spatial organization of the genome and its 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. He, D. L. & Bush, J. W. M. The hydrodynamics of water-walking arthropods. *J. Fluid Mech.* **644,** 5–33 (2010).
3. 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).
4. Vollrath, F. & Edmonds, D. T. Modulation of the mechanical properties of spider silk. *Nature* **340,** 305–307 (1989).
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. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).
9. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361,** eaar2555–11 (2018).
10. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361,** 412–415 (2018).
11. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547,** 241–245 (2017).
12. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547,** 236–240 (2017).
13. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. USA* **108,** 4334–4339 (2011).
14. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165,** 1686–1697 (2016).
15. Kilic, S. *et al.* Phase separation of 53BP1 determines liquid-like behavior of DNArepair compartments. *EMBO J.* **38,** 7–17 (2019).
16. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).
17. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37,** 76–81 (2016).
18. Kaestner, K. H. The FoxA factors in organogenesis and differentiation. *Curr. Opin. Genet. Dev.* **20,** 527–532 (2010).
19. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265,** 1599–1600 (1994).
20. 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).
21. Iwafuchi, M. *et al.* Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* **52,** 418–427 (2020).
22. Zhou, H.-X., Nguemaha, V., Mazarakos, K. & Qin, S. Why Do Disordered and Structured Proteins Behave Differently in Phase Separation? *Trends Biochem. Sci.* **43,** 499–516 (2018).
23. Jawerth, L. M. *et al.* Salt-Dependent Rheology and Surface Tension of Protein Condensates Using Optical Traps. *Phys. Rev. Lett.* **121,** 258101 (2018).
24. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324,** 1729–1732 (2009).
25. Botao, X. *et* al. Histone H1 compacts DNA under force and during chromatin assembly. **23,** 4864–4871 (2012).
26. Golfier, S., Quail, T., Kimura, H. & Brugués, J. Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner. *Elife* **9,** (2020).
27. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360,** 102–105 (2018).
28. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366,** 1338–1345 (2019).
29. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366,** 1345–1349 (2019).
30. Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments. *Curr. Opin. Cell Biol.* **64,** 124–138 (2020).
31. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
32. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).
33. Yoshimura, S. H. & Hirano, T. HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? *J. Cell. Sci.* jcs.185710–8 (2016).

Acknowledgements: We thank Pavel Tomancak, Anthony Hyman, Stephan Grill, Iain Patten, Martin Loose, David Oriola, Benjamin Dalton, Patrick McCall, Vasanthnarayan Murugesan, and Claudia Meyer for helpful feedback and stimulating discussions. We would like to 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) and a Volkswagen "Life" grant number 94820.

Author contributions: T.Q. and J.B. conceived the project. T.Q. and S.G. performed imaging experiments. 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 construct and purified the protein. 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.

Author information: The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.B. (brugues@mpi-cbg.de).

Table of contents:

Figures 1-4
Methods
Extended Data Figures 1-6
Supplementary information

Figure 1

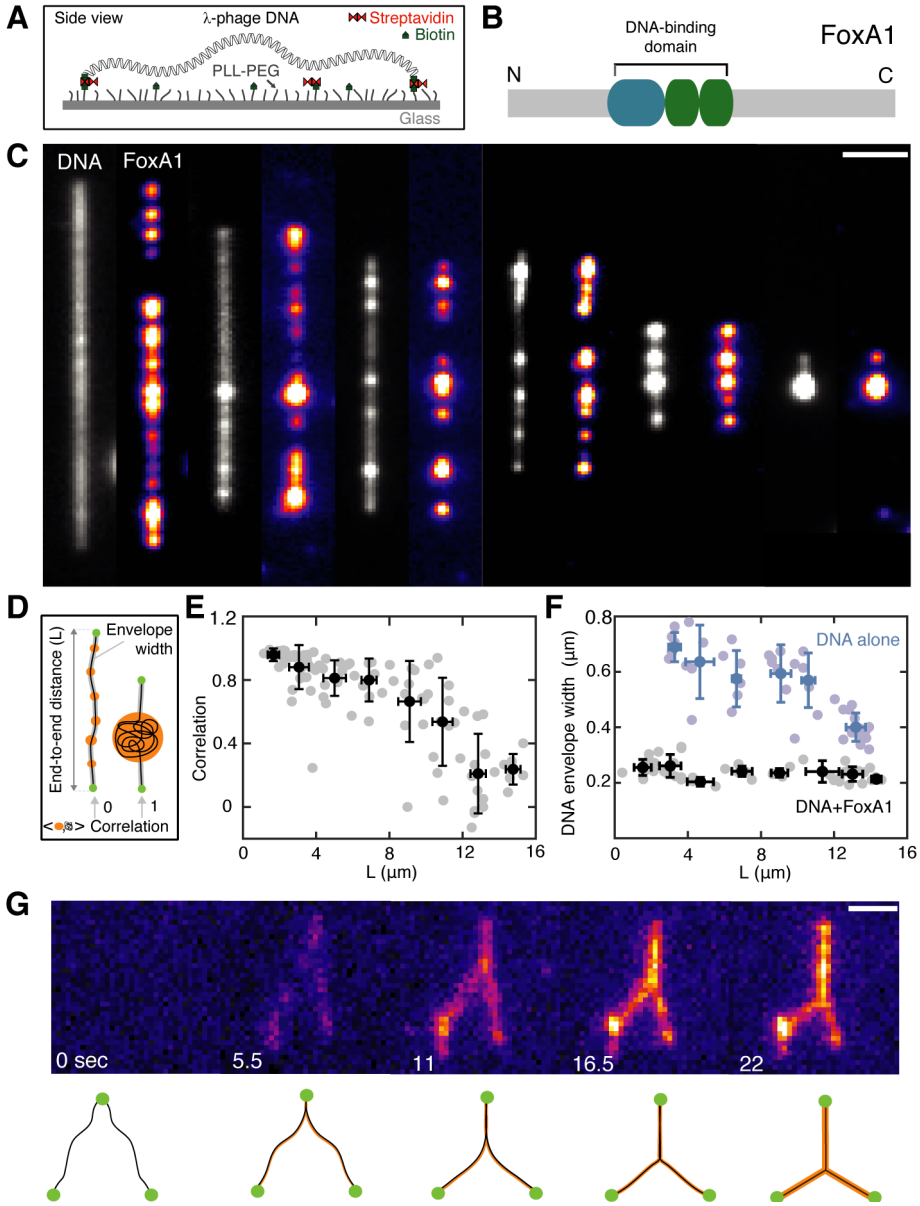


Fig. 1: FoxA1 forms DNA-FoxA1 condensates in a tension-dependent manner. (A) Schematic of single λ-phage DNA molecule assay. (B) Structure of FoxA1, consisting of a structured DNA-binding domain flanked by mostly disordered N and C termini. The DNA-binding domain has a sequence-specific binding region (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. 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). (F)

DNA envelope width measurements reveal that FoxA1-DNA condensation buffers DNA tension (blue and black dots correspond to control and DNA+FoxA1 conditions, n=46 and n=50 respectively). The data is binned every 2-μm (mean ± SD). (G) Representative images of FoxA1 zipping two independent DNA strands over time. Scale bar=1 μm.

Figure 2

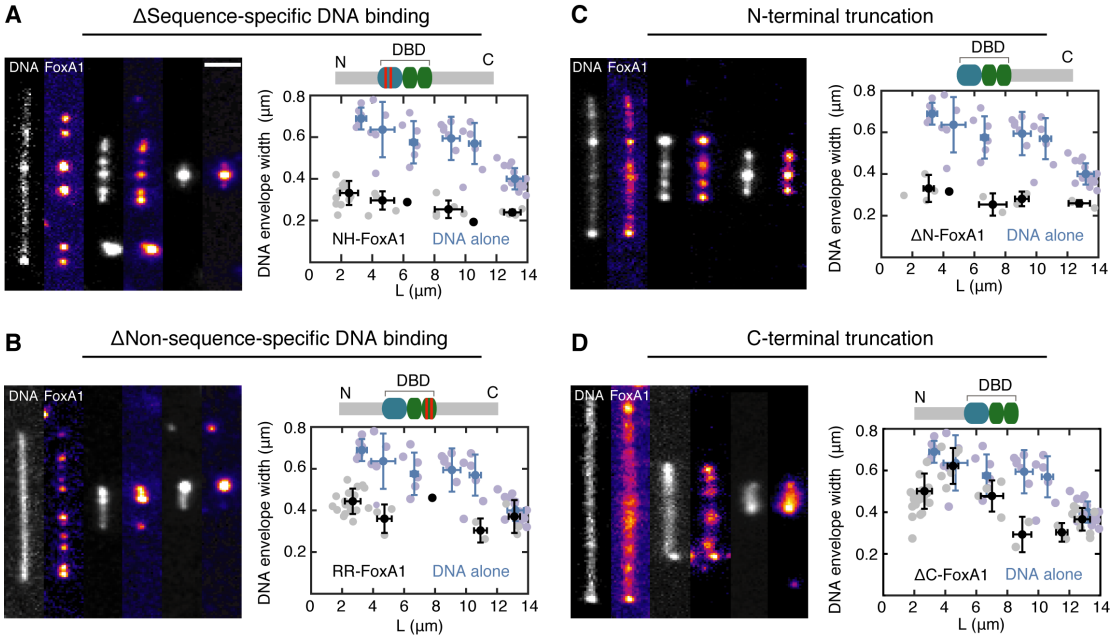
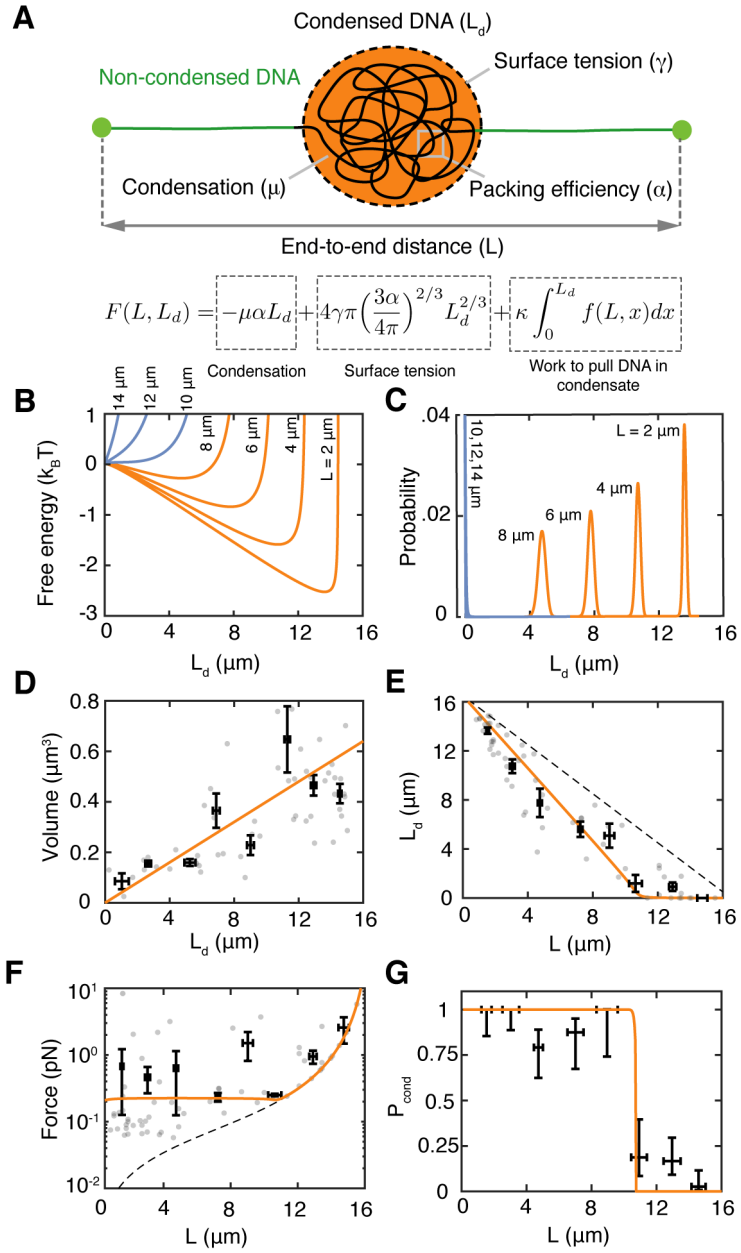


Fig. 2: Mutant analysis reveals that the C terminus of FoxA1 drives DNA condensation. Representative images of FoxA1 mutants and corresponding DNA envelope width measurements. The data is binned every 2-μm and the mean ± SD are shown in black for each mutant and in blue for control (n=46). Scale bar=2 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=15). (D) C-terminal truncation of FoxA1 ΔC-FoxA1 inhibits DNA condensation (n=34).

493 **Figure 3**
494



495
496

497 **Fig. 3: Thermodynamic description of a liquid phase condensing onto a**
498 **semi-flexible polymer explains FoxA1-mediated DNA condensation.** (A)
499 Schematic representing DNA-FoxA1 condensation (orange). DNA can be in a
500 condensed state (black) or a non-condensed state (green). DNA condensation
501 depends on the condensate surface tension (γ), condensation free energy per
502 volume (μ), and DNA packing efficiency (α). (B) Free energy profiles as a function
503 of condensed DNA (L_d) for different L reveal a first-order phase transition for
504 DNA–protein condensation (orange and blue correspond to favorable and
505 unfavorable condensation, respectively). (C) Boltzmann distributions
506 corresponding to the free energy profiles in (B). (D) Condensate volume linearly
507 increases with L_d. The orange curve represents a linear fit to individual strands
508 (n=53). For (D), (E), and (F), individual strands are represented as gray dots and
509 binned mean±SEM is in black. (E) Amount of condensed DNA as a function of L

(n=63 strands) 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) Capillary forces that DNA-protein condensates exert on non-condensed DNA are buffered (n=63 strands). Orange curve is the theoretical prediction. The gray dashed line 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 SEM, and the P_{cond} error bars are the 95% confidence intervals from a Beta distribution.

Figure 4

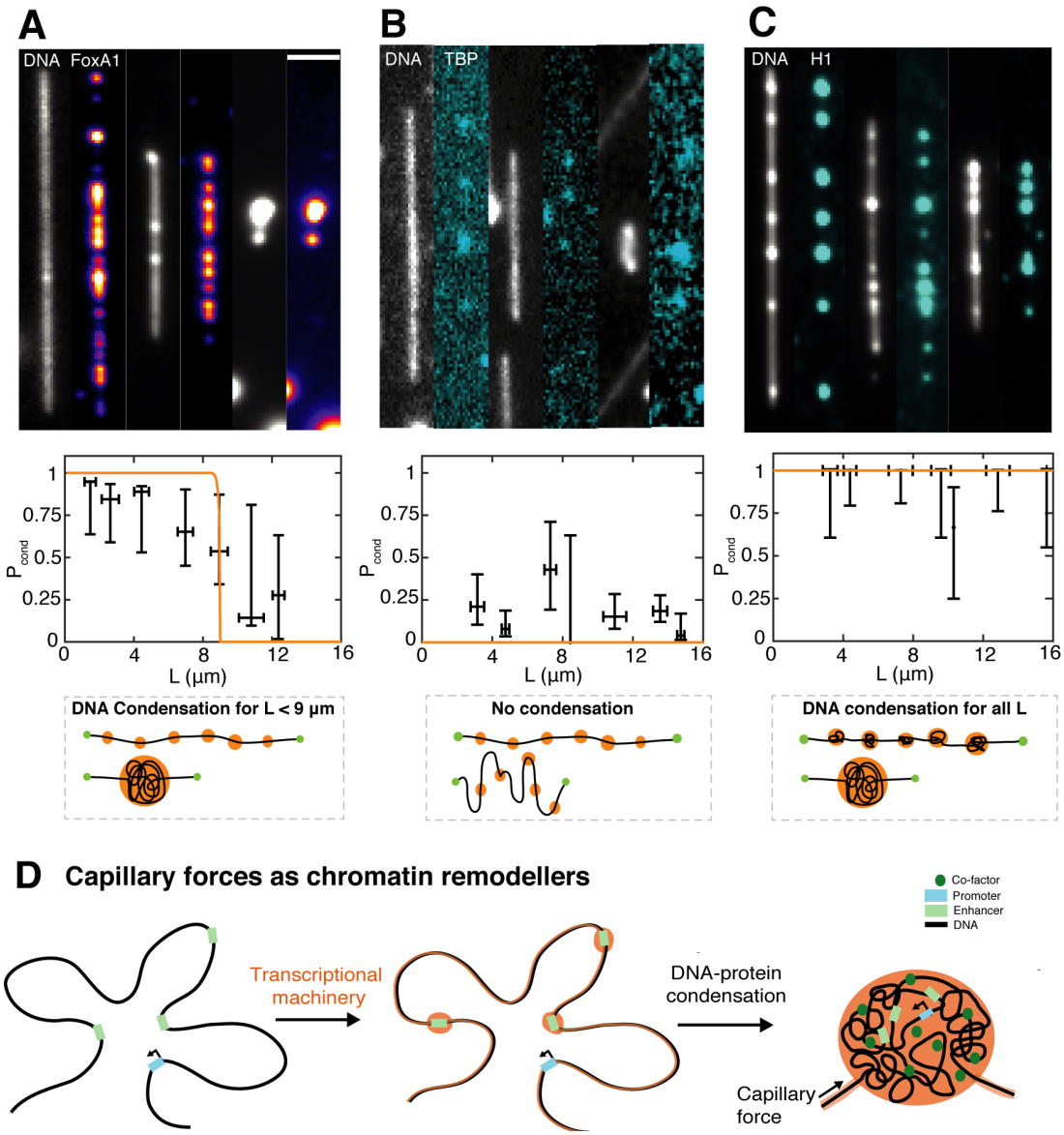


Fig. 4: Universality of protein-mediated DNA-protein condensation. Representative images of sequence-specific DNA-binding mutant NH-FoxA1 (A), Tata-box-binding protein (B), and Somatic linker histone H1 (C), with corresponding P_{cond} data below. DNA is stained with 10 nM Sytox Green for NH-FoxA1 and 25 nM Sytox Orange for both Tata-box-binding protein and Somatic linker histone H1. P_{cond} is computed from local correlation data with n=174 condensates for NH-FoxA1 (A), n=154 condensates for Tata-box-binding protein (B), and n=61 for H1 (C). Scale bar=2 μm. We found that NH-FoxA1 condensed DNA less efficiently 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 capillary forces that could serve to recruit transcriptional regulators, and potentially remodel chromatin at physiologically relevant force scales in order to properly regulate transcription.

578 **Methods**
579
580 Cloning and protein purification
581 FoxA1 was first cloned into a middle-entry Destination vector with a fluorescent
582 tag on the C terminus and a glycine-rich flexible linker region between FoxA1
583 and mCherry. A LR reaction was performed with pDEST17, a destination vector
584 with a N-terminal His_6 tag using Gateway Cloning to introduce FoxA1-mCherry
585 into this bacterial expression vector: pDEST17-FoxA1-mCherry. This vector was
586 transformed into T7 express bacterial cells (enhanced BL21 derivative, NEB
587 C2566I), and grown to OD~0.4-0.8, whereupon we added 1 mM IPTG and
588 expressed His_6-FoxA1-mCherry for 3-4 hours at 37 °C. We snap-froze the pellets
589 in liquid nitrogen and stored them at -80 °C. The pellets were redissolved in
590 binding buffer (1xBB) that contained 20 mM Tris-HCl (pH=7.9), 500 mM NaCl, 20
591 mM Imidazole, 1 mM $MgCl_2$, supplemented with a single cOmplete Protease
592 Inhibitor Cocktail tablet and Benzonase. The redissolved pellets were lysed using
593 a pressure-based homogenizer, and clarified via centrifugation. The supernatant
594 was discarded and the pellets were resuspended in 1xBB + 6 M Urea. The lysate
595 was spun again, and the supernatant was poured over an IMAC column, washed
596 with 10 column volumes (5 ml per column volume) of 1xBB + 6 M Urea, and then
597 eluted with 1xBB+6 M Urea+250 mM Imidazole. Protein-rich fractions were
598 detected using a Bradford assay, pooled, and then dialyzed overnight into
599 storage buffer (1xSB), 20 mM HEPES (pH=6.5), 100 mM KCl, 1 mM $MgCl_2$, 3 mM
600 DTT, and 5 M Urea. Next, the proteins were dialyzed in 1xSB+4 M Urea for two
601 hours, then 1xSB+3 M Urea for two hours, and finally 1xSB+2 M Urea for two
602 hours. The protein fraction was spun-concentrated to 4-5 mg/ml (~50 μM), and
603 then snap-frozen in liquid nitrogen and stored at -80 °C. Dialyzing the protein
604 from 2 M urea to lower amounts of urea led to protein precipitation. NH-FoxA1-
605 mCherry and RR-FoxA1-mCherry were obtained by modifying the pDEST17-
606 FoxA1-mCherry vector using the Q5 Site-Directed Mutagenesis Kit (NEB), and
607 purified using the same protocol given above. Specifically, for NH-FoxA1, we
608 mutated residue number 216 (where residue number 1 is start codon of the
609 open-reading frame of FoxA1) from an asparagine to an alanine, and residue
610 number 222 from a histidine to an alanine[20]. For RR-FoxA1, we changed residue
611 numbers 262 and 265 from arginines to alanines[20]. For the FoxA1 N-terminal

612 truncation, we first digested pDEST17-FoxA1-mCherry with NdeI and BamHI—
613 these digestion sites are unique and located right before the N-terminal His_6 tag
614 and directly following the C-terminal mCherry. Then we amplified the target
615 using a forward primer that contained the NdeI cut site, a His_6 tag, and the start
616 of the DNA-binding domain (residue number 144) and a reverse primer that had
617 the BamHI cut site and the end of the mCherry sequence. For the C-terminal
618 truncation, we again digested pDEST17-FoxA1-mCherry with NdeI and BamHI,
619 and amplified our target using two sets of primers: first we amplified the
620 purification tag, the N terminus, and the DNA-binding domain using a reverse
621 primer that targeted the end of the DNA-binding domain and had a unique SpeI
622 cut site. Second, we simply amplified the mCherry fluorescent tag where the
623 forward primer contained a SpeI cut site. For both constructs, we then cut the
624 template, performed a ligation, transformed into bacteria, and selected for clones
625 using ampicillin. We used somatic linker histone purified from calf thymus,
626 labeled with Alexa Fluor 488 dye, kept at 2 mg/ml (~70 μM) in 1xPBS, stored
627 snap-frozen at -80 °C (H-13188, ThermoFisher). To purify Tata-box binding
628 protein, we cloned a template of a zebrafish Tata-box-binding protein (zTBP)—a
629 gift from Nadine Vastenhouw—into a bacterial expression vector where a His_6
630 tag, a Maltose-binding protein purification tag, and an eGFP fluorescent tag were
631 on the N terminus (His_6-MBP-eGFP-zTBP). This vector was transformed into T7
632 express bacterial cells (enhanced BL21 derivative, NEB C2566I) and grown to
633 OD~0.4-0.8, whereupon we added 0.2 mM IPTG and expressed His_6-MBP-eGFP-
634 zTBP overnight at 18 °C. We lysed the cells into lysis buffer A containing 50 mM
635 Tris-HCl (pH=8.0), 1 M NaCl, 10% glycerol, 1 mM DTT, 1 mM $MgCl_2$
636 supplemented with a single cOmplete Protease Inhibitor Cocktail tablet. After
637 lysis, we added NP40 to 0.1% and clarified via centrifugation. We performed a
638 polyethylenimine precipitation to precipitate the remaining DNA in the soluble
639 fraction and then an ammonium sulfate precipitation to recover the protein
640 fraction, resuspending the precipitated proteins in lysis buffer B containing 50
641 mM Tris-HCl (pH=8.0), 1 M NaCl, 10% glycerol, 1 mM DTT, 0.1% NP40, and 20
642 mM imidazole and clarified the soluble fraction via centrifugation. We poured the
643 lysate over an IMAC column and eluted using 2xPBS, 250 mM imidazole, 10%
644 glycerol, and 1 mM DTT. We performed a Bradford assay to detect protein-rich

645 fractions, pooled these fractions, and then dialyzed the proteins overnight into
646 20 mM HEPES pH=7.7, 150 mM KCl, 10% glycerol, and 1 mM DTT. We spun-
647 concentrated the proteins to 10 mg/ml (~100 μM), snap-froze using liquid
648 nitrogen, and stored at -80 °C.
649
650 DNA functionalization
651 To biotinylate DNA purified from λ-phage (λ-phage DNA), we combined 10 μg of
652 λ-phage DNA (NEB, N3011S) and 5 μl of a 10X polymerase buffer (50 mM Tris-
653 HCl, pH=7.2, 10 mM $MgSO_4$, 100 μM DTT) to a total reaction volume of 50 μl. We
654 then heated the mixture up to 65 °C for 7 minutes to break apart the λ-phage
655 DNA's sticky ends. After heat treatment, we added 100x molar excess of
656 biotinylated dATP, biotinylated dUTP, and dGTP, and dCTP. We then added 1 unit
657 (~1 μl) of Klenow enzyme, mixed well, and incubated overnight at room
658 temperature. We purified the biotinylated λ-phage DNA using ethanol
659 precipitation and stored at -20 °C.
660
661 PEGylation of cover slips and DNA micro-channel preparation
662 We functionalized glass cover slips with mPEG and PEG-Biotin. Briefly, we
663 sonicated coverslips first in acetone for 15 minutes followed by 5 rinses with
664 MilliQ water, and then another sonication step in 5 M KOH for 40 minutes. After
665 rinsing the coverslips 3 times with water and then 3 times with methanol, we
666 dried the coverslips with N_2. We silanized the coverslips combining 250 ml
667 methanol, 12.5 ml acetic acid, and 2.5 ml (3-aminopropyl)-trimethoxysilane,
668 incubating the coverslips in this mixture for 10 minutes at room temperature,
669 sonicating for 1 minute, and then incubating the coverslips for an additional 10
670 minutes. Next we rinsed the coverslips once with methanol, once with water, and
671 once again methanol, and dried with N_2. Then we mixed 100 mg mPEG and ~1.5
672 mg Biotin-PEG with 450 μl PEGylation buffer (0.1M Sodium Bicarbonate,
673 pH=8.5), and spun the reaction at 10000 RPM for 1 minute. We pipetted 25 μl of
674 the PEG mixture onto a dried, silanized coverslip and put another coverslip on
675 top, generating a coverslip sandwich. We incubated these sandwiches over night
676 in distilled water-filled pipette tip-boxes in the dark. After incubation, we
677 carefully disassembled the coverslips, rinsed with MilliQ water, and dried with

678 N_2. To generate a channel for imaging, we first drilled holes through a cleaned
679 cover slide—these holes acted as channel inlets and outlets. We placed custom-
680 designed, laser-cut double-sided tape onto the coverslip, defining the channel
681 geometry. We then placed a functionalized PEG-biotinylated coverslip on top of
682 the double-sided tape, sealing the channel on either end with Valap. We filled the
683 channel with ~10-15 µl of 0.1 mg/ml free streptavidin, incubating the channel
684 with streptavidin for 1 minute. To remove the free, unbound streptavidin, we
685 flushed ~100 µl channel washing buffer (40 mM Tris-HCl, pH=8.0, 20 mM NaCl,
686 0.4 mM EDTA) through the channel, using the drilled holes as channel inlets and
687 outlets. We added 20 µl of 1:1000 biotinylated λ-phage DNA (~5 pM), incubating
688 it for ~1 min and then washed the channel with 3x100 µl of channel washing
689 buffer.
690
691 DNA imaging
692 For live imaging, we fluorescently stained immobilized DNA strands with 10-20
693 nM Sytox Green in imaging buffer (20 mM HEPES, pH=7.8, 50 mM KCl, 2 mM
694 DTT, 5% glycerol, 100 µg/ml BSA). For experiments with H1-Alexa-488 and
695 MBP-TBP-eGFP, we imaged DNA with 25 nM Sytox Orange. All experiments were
696 performed with this buffer and the concentration we used for all proteins was 10
697 nM. In general, protein concentrations above 10 nM led to an increase in the
698 number of strands that stuck to the coverslip. Before performing an experiment,
699 we pre-diluted our stock concentration 1:1000 and then performed a hard-spin
700 at 13K at 4°C for 1 minute. We then further diluted the proteins to 10 nM and
701 performed another hard-spin at 13K at 4°C for 1 minute. We excited Sytox
702 Green- or Sytox Orange-labelled DNA using a 488 nm and 561 nm laser
703 respectively, imaging the strands using a Nikon Eclipse microscope stand with a
704 Nikon 100x/NA 1.49 oil SR Apo TIRF and an Andor iXon3 EMCCD camera using a
705 frame-rate of 100 – 300ms. A highly inclined and laminated optical sheet (HILO)
706 microscopy mode was established using a Nikon Ti-TIRF-E unit mounted onto
707 the microscope stand to improve signal-to-noise ratio by cutting off background
708 fluorescence signal from unbound DNA dye in the buffer.
709
710 Strand length calculation

711 To calculate the end-to-end distance, we generated time-averaged projections of
712 FoxA1 and DNA and integrated these projections along the orthogonal axis to the
713 strand. To find the profiles' left edge, we computed the gradient of the DNA
714 signal's one-dimensional integrated intensity and computed the position when
715 the gradient signal went through a threshold (defined as 0.2). We then took all
716 the points from the start of the signal to this position, background-subtracted
717 this cropped signal, and fit an exponential to these points. To ensure that we
718 were including all of the DNA signal, we defined the threshold for both the left
719 and the right edges as three-quarters of the value of the fitted exponential value
720 at the point when the gradient had gone through the threshold. This threshold
721 value gave rise to end-to-end distances consistent from what we could observe
722 from the fluorescent images and from qualitative measurements that we
723 performed in FIJI. Using this threshold, we background-subtracted our signal and
724 simply computed the position values at which the signal went through the
725 threshold on the left and the right sides, and computed the end-to-end distance
726 as the difference between these two positions.
727
728 Global cross-correlation analysis
729 We generated time-averaged projections of both the FoxA1 and DNA signal from
730 movies roughly 30-60 sec in duration, and then summed the intensities in the
731 orthogonal axis to the strand, generating line profiles for both the FoxA1 and
732 DNA signals. We then normalized the amplitudes of the signals by their
733 Euclidean distances, and computed the zero-lag cross-correlation coefficient of
734 the normalized signals, which we defined as "Correlation". In general, Correlation
735 values range from -1 to 1, but, in our experimental data, the values range from
736 roughly 0 to 1, where 1 represents the formation of DNA-FoxA1 condensates and
737 0 represents the formation of only FoxA1 condensates (no DNA condensation).
738
739 DNA envelope width calculation
740 To compute the DNA envelope width, we first generated time-averaged
741 projections from movies of FoxA1 and DNA. We then selected segments of the
742 strand that did not contain FoxA1—regions of non-condensed DNA. Using these
743 segments, we then 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, scaled the signal's amplitude using the Euclidean distance, and fit a Gaussian. We defined the DNA envelope width as 2σ, which represents two times the standard deviation of the fitted Gaussian.

Condensate volume analysis

To calculate the volume, we first generated time-averaged DNA-FoxA1 projections, and then localized the peaks of the DNA condensates. Using the location of the peaks, 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 then fit Gaussians to these profiles without normalizing 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 contour length of DNA contained within a DNA-protein condensate—defined hereafter as condensed DNA amount (L_d)—we used a fluorescence-based approach. We first generated time-averaged projections of DNA and FoxA1 signals, and integrated the DNA signal in the orthogonal direction to the DNA 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 associated with non-condensed DNA, and intensity values above $\text{Threshold}_{\text{drop}}$ were defined as pixels associated with 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 (Fig. S3). In order 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.

Force analysis

In order to calculate the force that the condensate exerts against the non-condensed DNA strand, we used the worm-like chain model. The worm-like chain model gives a relationship between the extension of λ-phage DNA —$E = \frac{L}{L_t}$, where L is the end-to-end distance and L_t is the total contour length of the molecule (16.5 μm, for λ-phage DNA)—and the force. Quantitatively, upon addition of FoxA1, the amount of non-condensed DNA reduces, and the extension changes as follows, $E = \frac{L}{L_t - L_d}$, where L_d is the amount of condensed DNA. With this expression, we 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 that a FoxA1 condensate would generate a DNA condensate as a function of end-to-end distance, we first localized the peaks of the FoxA1 condensates from time-averaged projections taken from movies 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. This cut the image up into smaller boxes. We then computed the zero-lag normalized cross-correlation coefficient as follows:

809 $$C_{loc} = \frac{\sum\sum(f(x,y)-\mu_f)(g(x,y)-\mu_g)}{\sqrt{\sum f(x,y)^2 - {\mu_f}^2}\sqrt{\sum g(x,y)^2 - {\mu_g}^2}}$$

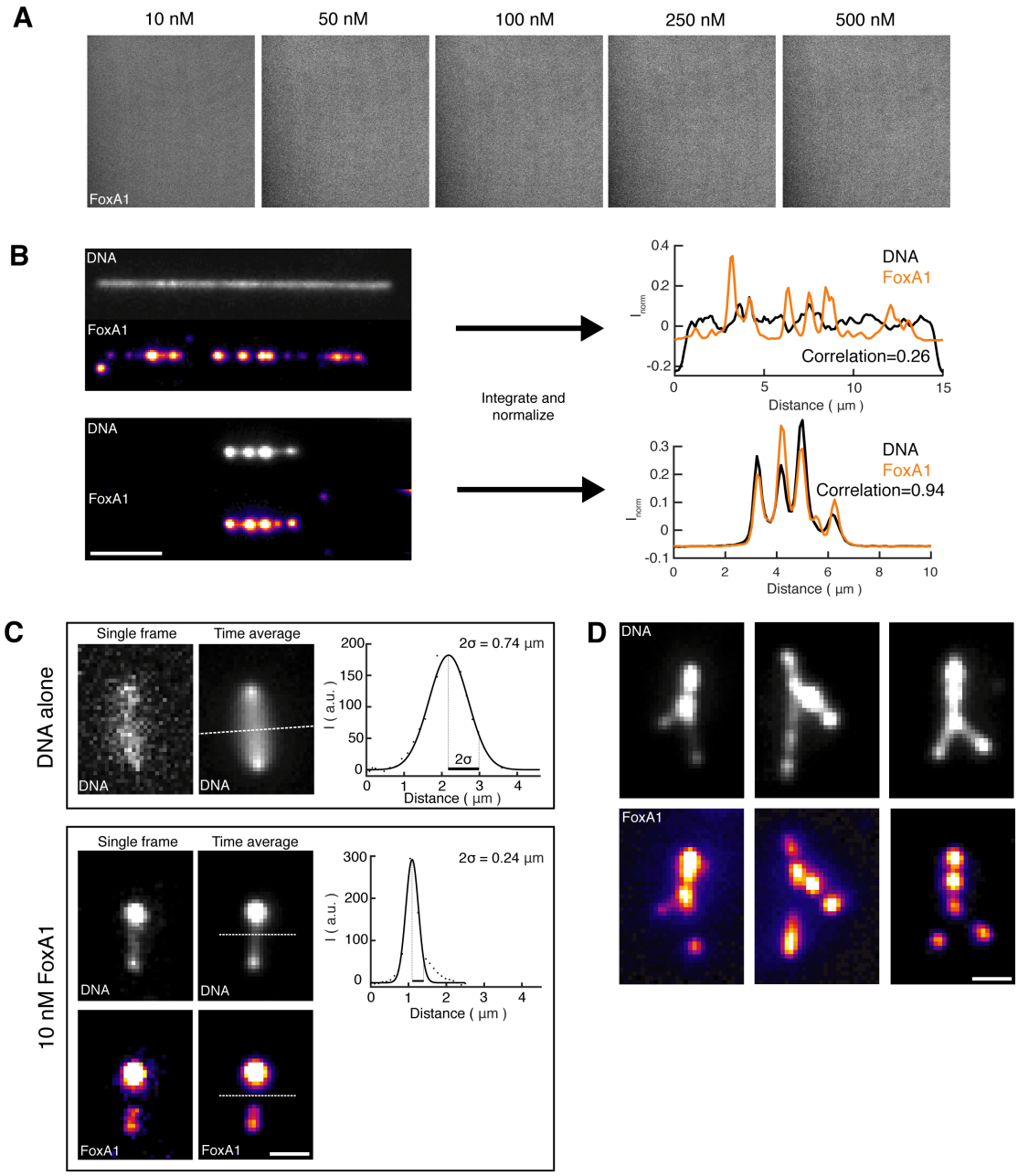
810 where $f(x,y)$ is the DNA, $g(x,y)$ is FoxA1, μ_f is the mean of the DNA image, and
811 μ_g is the mean of the FoxA1 image. This generates values from -1 to 1. For FoxA1-
812 mediated DNA condensation, the values for particular condensates are clearly
813 close to 1. When FoxA1 fails to mediate DNA condensation, owing to the
814 morphology of the underlying DNA strand and the small number of pixels, we
815 obtain values that range from -1 to roughly 0.5. To obtain a value for P_{nuc} as a
816 function of end-to-end distance, we selected a threshold of 0.75—C_{loc} values
817 above the threshold are considered as "condensed" and values below would be
818 considered "non-condensed". Hence, we binned the C_{loc} data in 2-μm increments
819 as a function of end-to-end distance, and calculated P_{cond} by simply taking the
820 number of "condensed" condensates and dividing it by the total number of
821 condensates within the bin. The confidence intervals for P_{cond} in each respective
822 bin are computed by computing the 95% confidence interval of a beta-
823 distribution, which represents the probability distribution for a Bernoulli
824 process that takes into account the total number of successes with respect to the
825 total number of attempts.
826
827 <u>Parameter fitting and confidence intervals</u>
828 Our thermodynamic description contains three parameters: α, the DNA packing
829 efficiency given as DNA strand length as a function of cross-sectional area; γ, the
830 surface tension of the condensate; and μ, the condensation free energy per
831 volume. To fit α, we computed condensate volumes for individual strands and
832 plotted these volumes as a function of condensed DNA (L_d) and fit a linear
833 function to the individual points, giving 0.04 μm^2 for WT-FoxA1 and 0.09 μm^2 for
834 NH-FoxA1. The confidence intervals presented in the main text represent the 95
835 per cent confidence intervals that were generated from fitting the individual
836 points directly. To fit γ and μ, we generated theoretical expressions for both L_d
837 and P_{cond} as a function of L, and minimized the error of these theoretical
838 expressions with respect to the experimental data to optimize the parameter
839 values. First, to compute a theoretical expression of $L_d(L)$, we generated free
840 energy profiles as a function of condensed DNA (L_d) for values of L from 0 to L_t,

where L_t is λ-phage DNA's contour length, 16.5 μm. For each individual free energy profile, we computed the normalized Boltzmann distribution $P(L_d) = \frac{e^{-\beta F(L,L_d)}}{\int_0^{L_t-L} e^{-\beta F(l)} dl}$ and then calculated the mean of the Boltzmann distribution $\overline{L_d} = \int_0^{L_t-L} lP(l)dl$, giving rise to $\overline{L_d}(L)$. To compute theoretical expressions of $P_{cond}(L)$, we again generated free energy profiles for a range of L from $[0, L_t]$ plotted as a function of L_d. We then localized the position of the local maximum in the free energy, L_d^{max}, and then computed the probability to "not" nucleate a droplet from the Boltzmann distribution $\int_0^{L_d^{max}} P(l)dl$. Following from this, we then computed $P_{cond} = 1 - \int_0^{L_d^{max}} P(l)dl$, giving rise to $P_{cond}(L)$. To optimally fit γ and μ, we computed weighted errors from the theoretical expressions as compared to the experimental data. For the following error minimization analysis, we compared the theoretical expressions against the mean of data that was binned in 2- μm-width bins. Briefly, for a given set of (γ, μ), we computed theoretical expressions for both $\overline{L_d}(L)$ and $P_{cond}(L)$. Then, for each "binned" mean for both condensed DNA and condensation probability, we computed the squared residual of the mean value with 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,

$P_{nuc}{}^{\sigma^2}_{cond} = \frac{(1+k)(1-k+N)}{(2N^2(3+N))}$ 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, μ) 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 confidence intervals for the fitted parameters from the error minimization, we first obtained one-dimensional profiles of the integrated exponentiated total error for μ as a function of γ and γ as a function of μ. 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 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, μ) 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, μ) 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}.

905 **Extended Data Figure 1:**



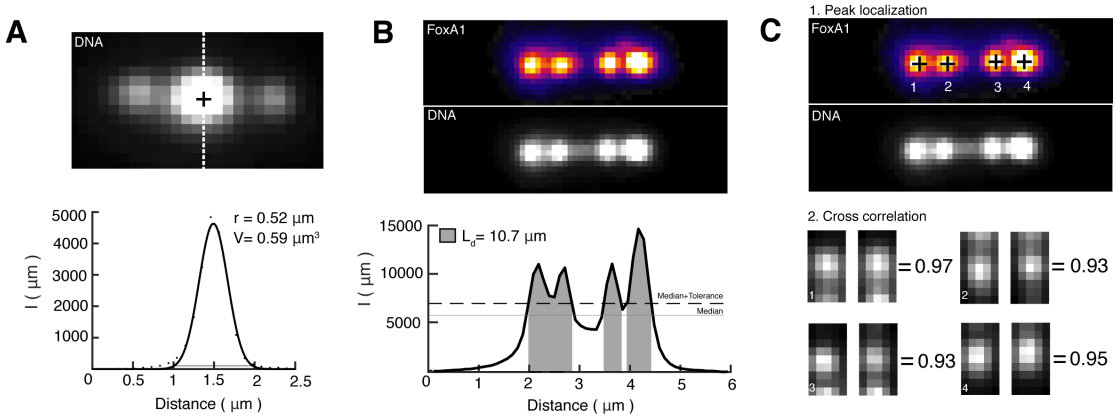
906

907

908 **Extended Data Fig. 1: Quantification of FoxA1-mediated DNA condensation.**
909 (A) Representative fluorescent images of FoxA1-mCherry in buffer (20 mM
910 HEPES, pH=7.8, 50 mM KCl, 2 mM DTT, 5% glycerol, 100 µg/ml BSA) at different
911 concentrations, 10-500 nM, in the absence of DNA reveals that FoxA1 does not
912 form condensates in bulk. Using spinning disk microscopy and a 60x objective,
913 we acquired images 70 μm x 70 μm in size with an exposure time of 250 msec
914 and a time stamp of 500 msec to generate movies 30 seconds in duration. For all
915 measured concentrations we generated n=3 movies and did not observe any
916 FoxA1 condensation. (B) Global cross-correlation between FoxA1 and DNA
917 reveals FoxA1-mediatd DNA condensation. Left, representative fluorescent time-
918 averaged projections of DNA and FoxA1 at two different end-to-end distances.

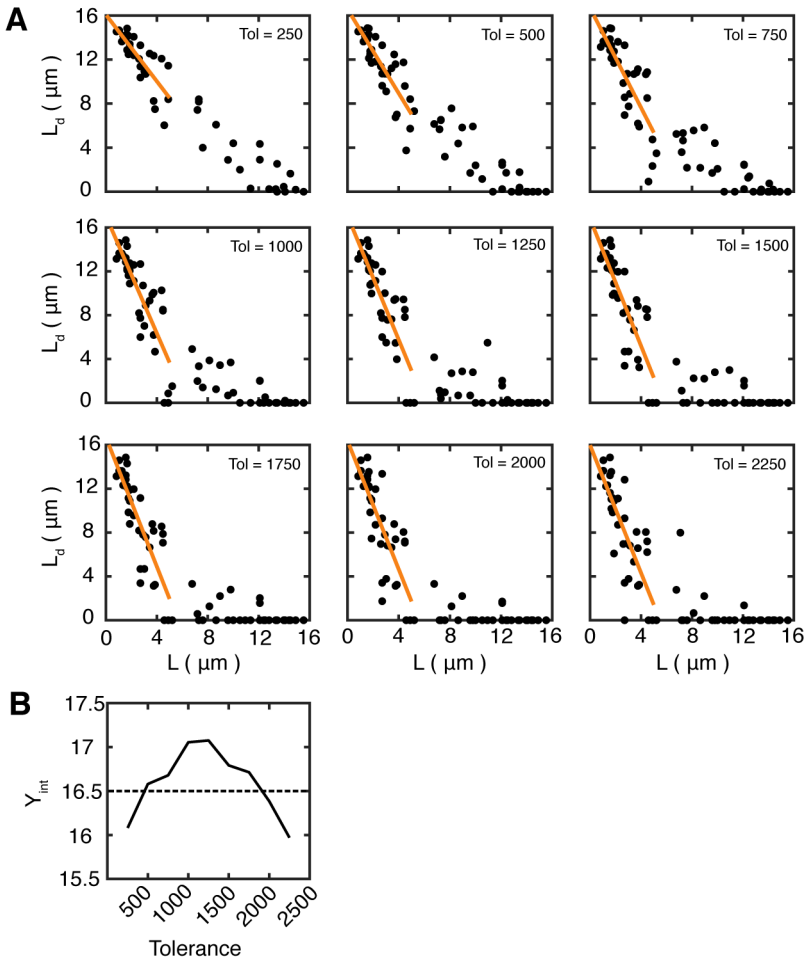
Integrating both the DNA and FoxA1 signals along the axis orthogonal to the long axis of the strand gave rise to line profiles, which we normalized, and then plotted as a function of distance (DNA in black and FoxA1 in orange). We then computed the zero-lag cross-correlation coefficient (C_{global}). (C) 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 2 times the standard deviation of the Gaussian fit. (D) Three representative examples of FoxA1-mediated zipping. These images are time-averaged projections of both FoxA1 and DNA.

Extended Data Figure 2:



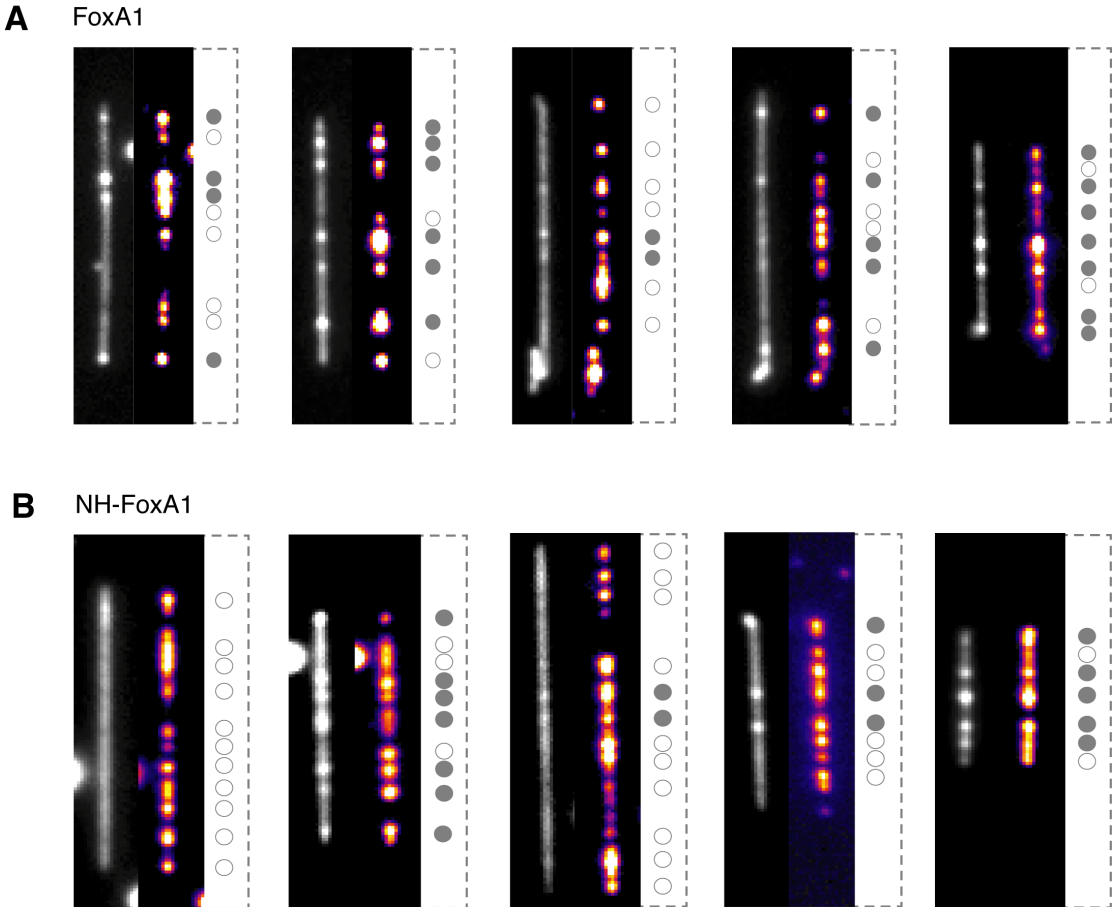
Extended Data Fig. 2: Quantification of condensate volume, force and condensation probability. (A) Condensate volume calculation. Top, a representative time-averaged projection of a single DNA condensate. The black dashed cross represents the peak location. The white dashed line is the line profile position that we extract from the image governed by the peak location. Below, the black dots are the background-subtracted values of the line profile and the solid black line is a fit to a Gaussian function. The gray line represents the threshold value computed from the gradient of the Gaussian function that defines the edges of the condensate (see Methods). These positions define a condensate diameter that, assuming the condensates are spherical, allows to compute the volume of the condensate. (B) Condensed DNA calculation. Top, representative time-averaged projections of DNA and FoxA1. Below, a line profile (black) was computed from integrating the DNA signal along the orthogonal axis to the strand. To define condensed versus non-condensed regions of DNA, we defined a threshold given by the median of the signal (gray) plus a tolerance (black dashed). Above this threshold, pixels count towards condensed DNA, and, below this threshold to non-condensed DNA. (C) Local correlation calculation. In order to compute P_{cond} as a function of end-to-end distance, we computed the correlations between individual condensates. We first localized the peaks of FoxA1 condensates and then cropped 1 μm x 0.5 μm boxes, centered around these peaks, of both the FoxA1 and corresponding DNA signals. The black dashed crosses give the location of the peaks and the images below are the boxes for both the FoxA1 and the DNA signal (FoxA1 to the left, DNA to the right). We then computed the normalized correlation of these images.

Extended Data Figure 3:



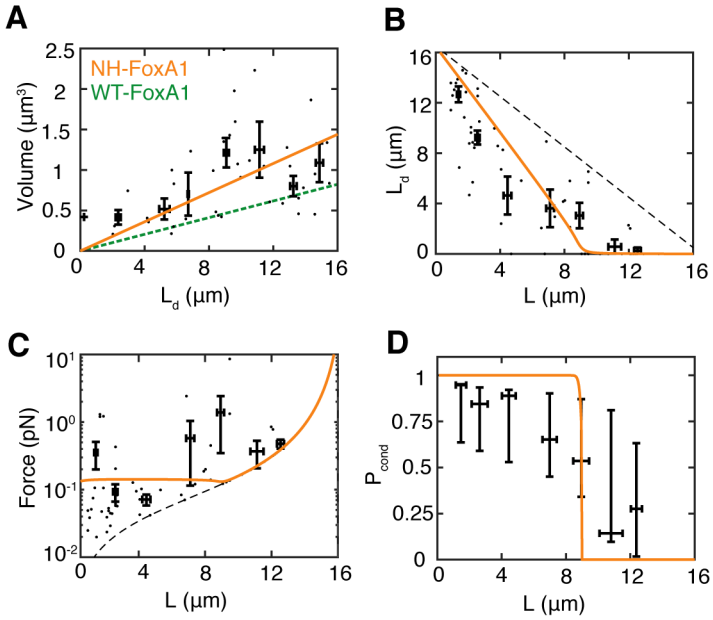
Extended Data Fig. 3: Tolerance value calculation. To calculate the amount of condensed DNA (L_d), we define a threshold for integrated DNA signal intensities that controls how we differentiate between condensed versus non-condensed DNA. The threshold is the median value of the signal plus a tolerance—a tolerance is required to suppress artefactual fluctuations of the signal close to the median value. We plotted L_d as a function of end-to-end distance L for different tolerance values. The black dots represent the amount of condensed DNA for individual strands and the orange curve represent linear fits to these points for end-to-end distance below 5 μm. (B) 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). We plotted the y-intercepts of the linear as a function of the tolerance and found two values with y-intercepts equal to 16.5. We selected a tolerance=500, as this was consistent with FoxA1-mediated DNA-FoxA1 condensate formation up to approximately 10 μm, which we experimentally observed, as opposed to a tolerance~2000, that gave DNA-FoxA1 condensate formation up to only 5 μm, inconsistent with our measurements and analysis (see Methods).

Extended Data Figure 4:



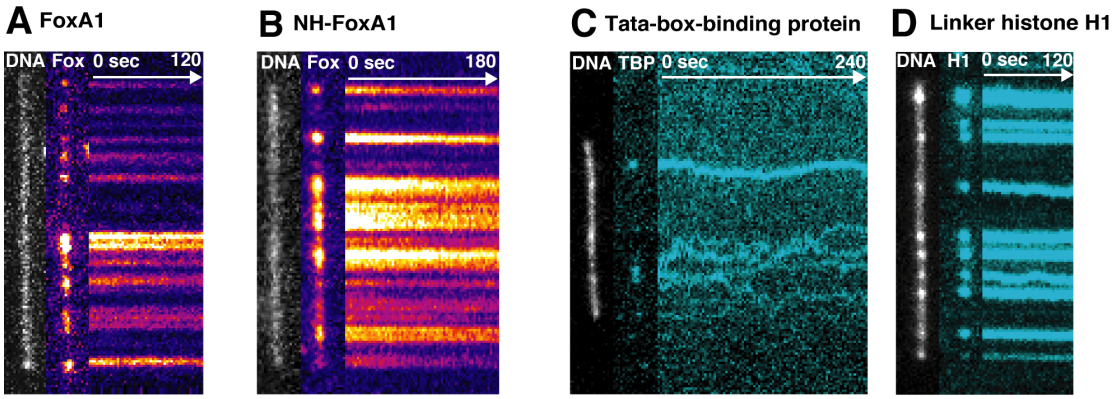
Extended Data Fig. 4: Bistability of FoxA1-mediated DNA condensation. (A) Representative time-averaged projections of DNA and FoxA1 signals show that FoxA1 condenses DNA in an all-or-nothing manner. On the right side of each pair of images, we localized the FoxA1 condensates and showed whether FoxA1 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.

Extended Data Figure 5



Extended Data Fig. 5: Quantification of NH-FoxA1-mediated DNA condensation. (A) Condensate volume as a function of condensed DNA amount (L_d). The black dots represent individual strands (n=43) and the data is binned every 2 μm (mean ± SEM). Fitting the individual points, we obtained a linear curve with slope 0.09 μm^2 given in orange. The green dashed line is a guide to demonstrate the theoretically determined fit for WT-FoxA1 (Fig. 3D) with a slope of 0.04 μm^2. (B) Condensed DNA amount as a function of end-to-end distance (L). The black dots represent individual strands (n=57) and the data is binned every 2 μm (mean ± SEM). The orange curve is the expression computed from the theoretical description parameter values determined through error minimization (see Methods). (C) The force that the condensate exerts on the non-condensed DNA as a function of end-to-end distance. The black dots represent individual strands (n=57) 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 0.14 pN. The dashed black line represents the force exerted on the non-condensed strand when L_d=0. (D) The probability for FoxA1 to form a DNA-FoxA1 condensate as a function of end-to-end distance. Local correlations of individual FoxA1 condensates with DNA (Fig. S2 C) are first calculated and binned into 2-μm-width bins. P_{cond} is then computed by taking all of the values of the local correlations that are above the threshold value 0.75 in each bin and dividing this number by the total number of condensates in the bin. There are a total number of n=174 condensates used for this analysis. The dashed lines represent the P_{cond} value as computed within the bin with ± SEM for the end-to-end distance of the strand. The interval for P_{cond} are computed by computing the 95% confidence interval of a beta-distribution, which represents the probability distribution for a Bernoulli process that takes into account the total number of successes with respect to the total number of attempts.

Extended Data Figure 6



Extended Data Fig. 6: 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.

Supplementary Information

1 Thermodynamic description of DNA-protein condensation

We consider the change of free energy associated with nucleating a condensate of radius R that contains DNA and FoxA1. The free energy of this process contains volume and surface contributions as well as the capillary work to pull DNA into the condensate,

$$F(R) = -\mu \frac{4}{3}\pi R^3 + \gamma 4\pi R^2 + W \tag{1}$$

where μ is the condensation free energy per volume, γ is the surface tension of the condensate, and W is the capillary work. We assume that the DNA in the condensate is in a collapsed state, which corresponds to the maximal packing allowed by excluded volume. With this assumption, the droplet volume and the amount of condensed DNA are linearly related, $V = \alpha L_d$, or $R^3 = \frac{3\alpha}{4\pi} L_d$, where α describes the DNA packing efficiency given as DNA strand length per cross-sectional area, and L_d is the amount of condensed DNA. We can then obtain the free energy of nucleating a condensate as a function of L_d,

$$F(L_d) = -\mu\alpha L_d + \gamma 4\pi \left(\frac{3\alpha}{4\pi}\right)^{2/3} L_d^{2/3} + W. \tag{2}$$

The work $W = \int_0^{L_d} f dx$ is the work that must be performed by capillary forces f to pull an

amount L_d into the condensate. To define f, we use a modified version of the phenomenological force-extension curve of the worm-like-chain model for λ-phage DNA that takes into account the amount of available non-condensed DNA (19). To this end, we modified the extension that the non-condensed DNA feels following the nucleation of a FoxA1-DNA condensate. Extension is a quantity that defines the ratio of the end-to-end distance of the DNA strand, L, to its contour length, L_t (for λ-phage DNA L_t = 16.5 μm). Here, FoxA1 mediates the nucleation of a DNA condensate phase, reeling in available slack into the condensate, which modifies the extension of the non-condensed strand: $\frac{L}{L_t - L_d}$. Using this extension, we can obtain an expression for the external force that the strand is subjected to, f, and define a modified version for the free energy of nucleating a FoxA1-DNA condensate:

$$F(L, L_d) = -\mu\alpha L_d + \gamma 4\pi \left(\frac{3\alpha}{4\pi}\right)^{2/3} L_d^{2/3} + \frac{k_B T}{P} \int_0^{L_d} \left(\frac{1}{4}(1 - \frac{L}{L_t - x})^{-2} - \frac{1}{4} + \frac{L}{L_t - x}\right) dx \tag{3}$$

where k_B is the Boltzmann constant, T is the temperature, and P is the persistence length of DNA. For what follows we define $\kappa = \frac{k_B T}{P}$. Solving the work integral we obtain,

$$F(L, L_d) = -\mu\alpha L_d + \gamma 4\pi \Big(\frac{3\alpha}{4\pi}\Big)^{2/3} L_d^{2/3} + \frac{\kappa L}{4} \Big(\frac{l l_d}{(l-1)(l + l_d - 1)} + \log \Big(\frac{(l-1)^2}{(1 - l_d)^4 (l + l_d - 1)^2} \Big) \Big), \tag{4}$$

where $l = \frac{L}{L_t}$ and $l_d = \frac{L_d}{L_t}$. Using this expression for the free energy, we can vary the end-to-end distance L, and obtain different profiles for the free energy as a function of L_d (see Fig. 3b). For example, at low L values—where the strand is not under much tension—we observe that there is a minimum value for L_d close to L_t. This means that, at this value of L, FoxA1 has mediated the generation of a FoxA1-DNA condensate using almost all of the strand's DNA. As L increases, however, the local minimum shifts to lower values of L_d and ultimately crosses zero, giving rise to a branch of metastable solutions. For even higher L values, the

metastable solution disappears and the global minimum is at $L_d = 0$ dominates (Fig. 3b and 3d). This sharp transition corresponds to a first-order phase transition. Notably, with optimized parameter values (see supplementary methods), the energy gained at low L is on the order of 2-3 $k_B T$, meaning 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,

$$P(L_d) = \frac{e^{-\beta F(L_d)}}{\int_0^{L_t - L} e^{-\beta F(x)} dx} \tag{5}$$

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: $\bar{L}_d(L) = \int_0^{L_t - L} x P(x) dx$ which then allows us also to calculate the magnitude of the capillary forces using the modified worm-like chain model given by f.