Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization

Rajesh Sharma1,8, Kyoung-Jae Choi1,8, My Diem Quan1, Sonum Sharma1, Banumathi Sankaran2, Hyekyung Park4, Anel LaGrone4, Jean J. Kim3,4, Kevin R. MacKenzie1,5,6✉, Allan Chris M. Ferreon1,2✉, Choel Kim1,7✉ & Josephine C. Ferreon1,7✉

Expression of a few master transcription factors can reprogram the epigenetic landscape and three-dimensional chromatin topology of differentiated cells and achieve pluripotency. During reprogramming, thousands of long-range chromatin contacts are altered, and changes in promoter association with enhancers dramatically influence transcription. Molecular participants at these sites have been identified, but how this re-organization might be orchestrated is not known. Biomolecular condensation is implicated in subcellular organization, including the recruitment of RNA polymerase in transcriptional activation. Here, we show that reprogramming factor KLF4 undergoes biomolecular condensation even in the absence of its intrinsically disordered region. Liquid-liquid condensation of the isolated KLF4 DNA binding domain with a DNA fragment from the NANOG proximal promoter is enhanced by CpG methylation of a KLF4 cognate binding site. We propose KLF4-mediated condensation as one mechanism for selectively organizing and re-organizing the genome based on the local sequence and epigenetic state.
Kruppel like factor 4 (KLF4) is a key component of reprogramming cocktails that transform fibroblasts to induced pluripotent stem cells (iPSCs). KLF4 cooperates with transcription factors (TFs) OCT4 and SOX2 in reprogramming to silence somatic enhancers and activate enhancers of pluripotency genes, including the gene to pluri-potency' gene NANOG, which is highly expressed in embryonic stem cell (ESCs). In iPSCs, KLF4 is enriched at the NANOG and OCT4 loci, which interact through space with many other pluripotency-related genomic sites. KLF4 is enriched at ESC super-enhancers and at iPSC genomic anchors that make more than four contacts, further implicating KLF4 in chromatin organization. How KLF4 or other TFs might initiate chromatin reorganizations that determine cell fate is of intense interest.

KLF4 contacts the DNA major groove with three tandem C2H2 zinc fingers (ZnFs) that make specific interactions. The first 400 KLF4 residues are likely to be disorder because they have low sequence complexity, and intrinsically disordered regions (IDRs) of other TFs help to silence or activate gene expression. In current models for transcriptional activation, TFs bound to their cognate sites cooperate with co-localized co-activators to recruit Mediator complex and RNA polymerase II through IDR:IDR mediated biomolecular condensations. KLF4 DNA binding domain and IDR might participate in such processes in open chromatin, and the KLF4 preference for CpG methylated over unmethylated cognate sites combined with its ability to bind 6 bp partial sites in nucleosomal DNA could help target it to silenced chromatin. The ability of KLF4 to undergo biomolecular condensation could facilitate pioneer interactions with closed chromatin and, as others have speculated, might stabilize long-range contacts between genomic loci.

Here, we show that KLF4 forms a liquid-like biomolecular condensate with DNA that recruits OCT4 and SOX2. Surprisingly, the intrinsically disordered region is not essential for KLF4 condensation in cells, and a KLF4 fragment comprising the isolated DNA binding domain (DBD) condenses with DNA in vitro. KLF4 DBD condensation with a NANOG promoter duplex is strongly enhanced by CpG methylation of a KLF4 cognate site, and ZnF point mutations that weaken interactions with DNA cognate sites decrease condensation in cells and in vitro. Single molecule methods show that KLF4 tandem zinc fingers bring together short DNA duplexes in dilute solution by a bridging interaction. We propose that bridging and/or condensation with DNA in a sequence- and CpG methylation-dependent manner underlie KLF4 function as a key chromatin organizer and pioneer transcription factor in somatic cell reprogramming.

**Results**

**KLF4 forms nuclear condensates at modest expression levels.** We used expression tags to monitor the distribution of KLF4 by fluorescence microscopy in HEK 293T cells or BJ fibroblasts, the somatic cells most widely used for reprogramming. KLF4 fused to mTurquoise2 (KLF4-mTurq) localizes to the nucleus and forms small puncta or round droplets, whereas mTurquoise2 alone is diffusely distributed throughout the nucleus and the cytoplasm. Transfection produces cells with various expression levels of tagged protein; KLF4-mTurq distribution is diffuse at the lowest expression levels, but most cells that express detectable KLF4-mTurq show punctate expression or droplets. Because round droplets are hallmarks of liquid–liquid phase separation (LLPS), we monitored KLF4-mTurq fluorescence after photobleaching: fluorescence recovers rapidly in both large droplets and small puncta in BJ fibroblasts, indicating that KLF4-mTurq in the condensate diffuses rapidly and is therefore liquid-like. Time courses using 3D z-stack fluorescence imaging reveal fusion of small droplets in BJ fibroblasts, indicating a liquid-like KLF4-mTurq condensate. Treatment with 1,6-hexanediol largely dissolves the KLF4-mTurq puncta and round droplets in HEK 293T cells, consistent with a liquid-like condensate.

A KLF4-mCherry fusion expresses at lower average levels than KLF4-mTurq but also forms puncta and droplets (Supplementary Fig. 1), indicating that the identity of the expression tag is not critical to condensation. Endogenous KLF4 levels have not been reported, but the KLF4-mTurq levels quantified by brightness (0.7 μM average for cells with puncta; 2.5 or 4.0 μM average for cells with small or large droplets, respectively, see Supplementary Fig. 2) are similar to those reported for TFs SOX2 or OCT4. We expect that KLF4 expression driven by vectors in reprogramming cocktails would result in robust biomolecular condensation.

The intrinsically disordered region is dispensable for KLF4 condensation. To identify domains that contribute to KLF4-mTurq biomolecular condensation, we expressed constructs lacking either the IDR (residues 1–417) or the DNA binding domain (DBD; residues 418–513). The intrinsically disordered region is dispensable for KLF4 condensation. Endogenous KLF4 levels have not been reported, but the KLF4-mTurq levels quantified by brightness (0.7 μM average for cells with puncta; 2.5 or 4.0 μM average for cells with small or large droplets, respectively, see Supplementary Fig. 2) are similar to those reported for TFs SOX2 or OCT4. We expect that KLF4 expression driven by vectors in reprogramming cocktails would result in robust biomolecular condensation.

The KLF4 DNA binding domain phase separates with cognate DNA. Purified KLF4 DBD is readily soluble and does not condense or precipitate at physiological salt or upon addition of 10% PEG 8000, a crowding agent used to enhance the weak interactions that drive biomolecular condensation (Fig. 2d). Because proteins that bind RNA can undergo RNA-induced phase separation, we tested the ability of NANK, a 30 bp NANK co-localize, we labeled DBD with Alexa Fluor 594 (AF594) and mixed it with NANK in the presence of the dye. Two-channel fluorescence images reveal fusion of small droplets in BJ fibroblasts, indicating that the phase is liquid-like. The liquid nature of large droplets is confirmed by rapid recovery of fluorescence after photobleaching localized regions (Fig. 2g). We conclude that DBD undergoes liquid–liquid phase separation (LLPS) with NANK at physiological salt without the need for crowding agents.
DBD-NANK LLPS depends in complex ways on the component concentrations. At 6 µM DBD, 0.25 µM NANK induces readily detectable LLPS, and increasing NANK up to 1 µM increases the amount of condensate, but further increases in NANK actually produce less condensate, and at 3 µM NANK and above LLPS is no longer detected (Fig. 2h). The lack of LLPS at high NANK suggests that phase separation requires a minimum DBD:NANK ratio, perhaps to saturate NANK cognate KLF4 sites. At any given DBD concentration, LLPS is not observed below a threshold NANK concentration; this NANK threshold is lower at KLF4-mTurq (HEK 293T cells) than at KLF4-mTurq (BJ fibroblasts).
high DBD levels (Fig. 2i, Supplementary Fig. 3). These data describe a phase diagram for DBD:NANK LLPS (Fig. 2i) and challenge us to understand the nature of DBD:NANK interaction.

**KLF4 DBD forms a 3:1 complex with a NANOG promoter duplex.** KLF4 binds 9 bp cognate sites containing methylated GGCG or ‘intrinsically methylated’ GGTG17 and activates expression from the NANOG promoter through two GGTG elements36 (sites KLFA and KLFC in Fig. 3a). A GGCG element in the NANOG promoter reverse strand (KLFB in Fig. 3a) is methylated and silenced in germ cells37, and NANOG promoter hypermethylation must be reversed to achieve pluripotency38. Because the nature of KLF4 DBD binding to this DNA might be important for LLPS, we tested binding at these sites in vitro. Each of three 12 bp duplexes excerpted from the 30 bp NANOG promoter fragment NANK contains a central GG(C/T)G and binds DBD in electromobility shifts assays (EMSA) (Supplementary Fig. 4). EMSA titration of the 30 bp fragment NANK (or its CpG methylated variant, NANKm) with DBD gives complexes of three different mobilities, consistent with DBD forming 1:1, 2:1, and 3:1 complexes with these DNAs (Fig. 3a). At the same DNA concentrations, DBD forms 1:1 complexes with a 12 bp duplex from the 30 bp fragment NANKm (Supplementary Fig. 4).
concentrations, 3 DBD equivalents form a detectable 3:1 complex with NANKm, but 6 equivalents are needed to form such a complex with NANK, consistent with the KLF4 preference for GGNCWCG over GGGCG. We conclude that our KLF4 DBD preparations are well folded and show target selectivity, and that DBD can form a 3:1 complex with NANK.

**KLF4 DBD forms a 1:1 complex with a cognate NANOG dodecamer.** We determined the crystal structure of the 1:1 complex of DBD bound to a dodecamer containing the NANOG proximal promoter KLF4 site (Fig. 3a, Table 1). As with previous KLF4 crystal structures with DNA17,18, each ZnF contacts three or four base pairs in the DNA major groove (Fig. 3b). The overall structure of the DBD:NKA complex is similar to previous KLF4:decamer complexes18 and each ZnF makes base-specific contacts mediated by one or more ‘specificity residues’ at positions −1, 2, 3, and 6 of the canonical C2H2 recognition code39. Residues R473 and R479 of ZnF2 (in positions −1 and 6) and R501 of ZnF3 (in position −1) hydrogen bond with the N7s and O6s of bases G5, G6, and G8, respectively; the arginine side chains also make polar interactions with water, ions, and/or aspartate side chains (Fig. 3c). As in previously solved structures, ZnF1 makes fewer base-specific contacts than ZnF2 or ZnF3. The only base-specific contact made by ZnF1 in this structure is H446 in the decamer (354 atoms superimpose with a root-mean-square deviation of 0.97 Å). When bound to a DNA decamer containing methylated-C (PDB ID: 4M9E), KLF4 DBD binds to the GGTG of our dodecamer with the same conformation as it does to the GGGCG of a decamer (87 Ca-atoms superimpose with a root-mean-square deviation of 0.97 Å). When bound to a DNA heptamer (PDB ID: 2WBS), the ZnF2 and ZnF3 domains contact their target bases as in the decamer (354 atoms superimpose with a root-mean-square deviation of 0.48 Å), but ZnF1 adopts a new orientation18 (Fig. 3d) that has been invoked to explain a 6 bp consensus KLF4 binding site on nucleosomes16.

**KLF4 site overlap may drive non-canonical binding.** The three NANK KLF4 sites match the 9 base KLF4 consensus site from JASPAR40 at 6 or 8 positions, but because the KLF4 and KLBFX sites overlap (Fig. 4a), how a 3:1 DBD:NANK complex forms (Fig. 3a, right) is not clear. We sought to determine the structure of three DBDs bound to NANK (or NANKm), but preparing 3:1 complexes invariably gave condensation, not crystalization. We, therefore, used superpositioning to determine if the canonical interactions seen in the DBD:NKA complex can be accommodated at sites KLF4A, KLF4B, and KLF4C in a B-DNA model of the NANOG proximal promoter. Individual superpositions at each site suggest favorable protein:DNA contacts, and simultaneously placing DBD at KLF4 and either KLF4A or KLBFX generates no clashes (Fig. 4b). However, canonical DBD occupation of both KLF4A and KLF4B causes intermonomer clashes of the two modeled ZnF1 domains (Fig. 4c). We conclude that in the observed 3:1 complexes (Fig. 3a), at least one ZnF1 projects away from the DNA. Using the pose for DBD bound to a DNA heptamer (Fig. 3d, black) at KLF4 or KLBFX with a canonical DBD posed at the other site relieves the clash (Fig. 4c). The four residue ZnF1–ZnF2

---

### Table 1 Data and refinement statistics.

| KLF4:NKA |  |
|----------|---|
| **Data collection** |  |
| Space group | P 1 2 1 |
| Cell dimensions | a, b, c (Å) |
| a, b, c (Å) | 38.96, 46.69, 45.20 |
| α, β, γ (°) | 90.0, 113.5, 90.0 |
| Resolution (Å) | 22.17-2.14 (2.22-2.14) |
| Completeness (%) | 96.14 (94.20) |
| Redundancy | 3.9 (3.8) |
| R<sub>merge</sub> | 0.2289 (2.774) |
| CC<sub>1/2</sub> | 0.975 (0.304) |
| **Refinement** |  |
| Resolution (Å) | 22.17-2.14 (2.22-2.14) |
| No. reflections | 8016 (780) |
| R<sub>merge</sub>/R<sub>free</sub> | 0.1975/0.2474 |
| No. atoms | Proteins 1142 |
| Bond lengths (Å) | 5 |
| Water | 93 |
| B-factors | Protein 84 |
| Ligand/ion | 45.64 |
| R.m.s. deviations | Bond lengths (Å) 0.009 |
| Bond angles (°) | 1.03 |

---

<sup>a</sup>Values in parentheses are for the highest resolution shell.

<sup>b</sup>50.0% of the observed reflections were excluded from refinement for cross validation purposes.
linker should allow ZnF1 to adopt different orientations, consistent with tandem ZnFs acting as independently folded “beads on a string”. KLF4 would contact NANK with ZnF2 and ZnF3, using the 6 bp KLF4 binding site detected on nucleosomal DNA and retaining most of the basis for specificity inferred from structural analysis. If overlapping KLF4 sites have functional importance because they obligately expose a ZnF, then such arrangements should be conserved. The mouse NANOG promoter (Supplementary Fig. 5) contains four overlapping KLF4 cognate sites in the −90 to −65 region: two sites on the top strand contain GGTG (and match the 9 bp consensus at 7 or 8 positions), and one site on each strand...
contains GGGCG (and matches the consensus at 6 or 8 positions). DBD binding schematics for these sites suggest that the mouse NANOG promoter would also exhibit a tethered, continuously solvent-exposed ZnF1 when saturated with KLF4, and that binding could be modulated by CpG methylation. Since tight DNA binding can be achieved by two ZnFs or one ZnF flanked by a basic region, we hypothesized that the exposed ZnF1 might recruit a second DNA partner, and that such DNA bridging might drive biomolecular condensation.

KLF4 DBD can bridge two DNA duplexes. We tested this hypothesis using single molecule Förster resonance energy transfer (smFRET)44-46. In DBD:NANK models where ZnF1 is excluded from either KLFA or KLFB (Fig. 4d), the 5′ end of the NANK coding strand is 40–43 Å from the Cα of H446, a ZnF residue that canonically contacts the major groove. We therefore 5′ end labeled the NANK coding strand with fluorescent donors or acceptors, reasoning that close proximity of the excluded ZnF1 might enable it to bring another labeled DNA close enough for FRET (Fig. 4e). With two labeled DNAs co-dissolved at low concentrations (100 pM AF488-labeled NANK, 500 pM AF594-labeled NANK), donor emission is observed but FRET is not (Fig. 4f, left). FRET events induced by 1 µM unlabeled KLF4 DBD (Fig. 4f, right) show that DBD can bring two DNAs together (closer than 55 Å, the R₀ for this label pair), providing a mechanism for DBD biomolecular condensation with DNA. Although detected at dilute concentrations that do not support mesoscale LLPS (Fig. 2i), these events are likely those that drive condensation at higher concentrations. The smFRET data do not define the stoichiometry of the complex(es), but they are consistent with our non-canonical model for DBD:NANK interaction (Fig. 4e).

Non-cognate DNA can drive KLF4 DBD phase separation. If bridging between NANK molecules by a single continuously excluded ZnF can drive LLPS, then DBD bound to non-cognate sites might make similar bridging interactions when one ZnF transiently leaves the major groove. To test this, we mixed DBD with four DNA duplexes of 12 to 40 bp that lack a GG(C/T)G. None induce LLPS at 3.0 µM DBD and 1 µM DNA, conditions at which NANK readily drives LLPS, but at 10 µM DBD and 3 µM duplex, all four non-cognate DNAs produce phase separated droplets (Fig. 5a). We infer that DBD bound to non-cognate DNA samples binding modes that transiently expose ZnFs to interact with another duplex.

DBD phase separation with non-cognate DNAs at these modest concentrations reinforces the idea that the failure of 6 µM DBD to support LLPS at 3 µM NANK despite robust LLPS at 1 or 2 µM NANK (Fig. 2i) results from the sequestering of DBD into canonical complexes that depopulate states in which ZnF1 is exposed. At 2:1 DBD:NANK, energetically favored canonical modes can be adopted without steric clashes, as in Fig. 4a, so few DBD will adopt either obligately or transiently exposed binding modes. We reasoned that providing cognate sites in trans might therefore dissolve pre-formed condensate by sequestering DBD. We tested this by preparing 10 µM DBD with 3 µM NANK and allowing the mixture to undergo LLPS. Adding 0.5 equivalents of NANK causes rapid, total loss of the condensate (Fig. 5b, top). The initial and final states are consistent with the phase diagram (Fig. 2i), while the rapid dissolution shows that material readily exchanges between the aqueous phase and the enriched phase. This behavior and our stoichiometry-based explanation are similar to observations and the rationale for phase separation of tandem SH3 domains with a tandem substrate47. For a 3:1 DBD:FGF4 mixture, adding 0.5 equivalents of FGF4 DNA only modestly decreases the amount of condensate (Fig. 5b, bottom) because FGF4 has no high affinity KLF4 cognate sites.

DNA sequence strongly influences LLPS threshold concentration. To determine how the DNA sequence might influence the lowest concentration at which LLPS is observed (the threshold DNA concentration), we performed LLPS assays for the non-cognate 17-mer DNA SBE, for NANK, and for the CpG methylated substrate NANKm at a range of DBD concentrations (Fig. 5c). Even at 3 µM SBE, no LLPS is seen with 1.5 or 3.0 µM DBD; with 6.0 µM DBD, the threshold SBE concentration is 1.5 µM. For NANK, the thresholds for LLPS at 1.5, 3.0, and 6.0 µM DBD are 250, 125, and 125 nM (respectively); the 6.0 µM DBD threshold for NANK is more than 10-fold lower than that of SBE. For NANKm, the LLPS thresholds at 1.5, 3.0 and 6.0 µM DBD are 63, 31, and 16 nM (respectively); these thresholds are 4–8 fold lower than those for NANK, and the threshold for NANKm at 6.0 µM DBD is at least 90-fold lower than that of SBE. The degree of DNA interaction with DBD by EMSA correlates with DBD:DNA LLPS potential (Supplementary Fig. 6), and the very low threshold concentrations for NANKm indicate that condensation can be directed to high affinity KLF4 binding sites. We conclude that the DNA sequence can dramatically alter the propensity for DBD:DNA biomolecular condensation, and that CpG methylation of the NANOG promoter KLF4 site (converting NANK to NANKm) strongly potentiates condensation.

Zinc finger domain mutations attenuate DBD:DNA condensation. If the lower threshold concentrations for NANKm compared to NANK are caused by tight DBD binding that accompanies CpG methylation (Fig. 3a), then residues that participate in KLF4 base-specific recognition (see Fig. 3c) should be important to LLPS. On the other hand, if the observed condensation depends on an unfolded DBD fraction, a different DBD surface, or a trace contaminant, then large-to-small mutations at the “specificity residues” of the C2H2 recognition code should have no effect on LLPS. We, therefore, prepared DBD carrying a ZnF2 mutation (E476D, position 3) that weakens affinity for cognate KLF4 sites and a ZnF3
mutation (R501A, position −1) that weakens affinity by EMSA (Supplementary Fig. 6). The double mutant domain (DBDE476D/R501A) shows decreased LLPS compared to wild type for all three of the tested DNAs (Fig. 5d). Even at 3.0 µM SBE, no LLPS is detected with 6.0 µM DBDE476D/R501A. For NANK, no LLPS is detected for DBDE476D/R501A at 1.5 or 3.0 µM, though the threshold concentration at 6.0 µM is unaltered from wild type. For NANKm, no LLPS is detected at 1.5 µM DBDE476D/R501A, and the thresholds for LLPS at 3.0 and 6.0 µM are 8 fold higher for DBDE476D/R501A than for DBD. Two of the four non-cognate DNAs that condense robustly at 3 µM with 10 µM DBD (Fig. 5a) show no condensation at 3 µM with 10 µM DBDE476D/R501A (Supplementary Fig. 7). We conclude that the ZnF2 and ZnF3 surfaces that contact bases in cognate DNA are important for LLPS with both non-cognate and cognate DNAs.
**Fig. 4 Models and solution data for DBD interactions with NANK.** a Consensus KLF4 binding site from JASPAR\(^{40}\) and schematic DBD contacts (left); mapping these sites and contacts onto the NANK sequence (right); in each 9 bp site (yellow), bases that match the consensus are in bold. b Posing DBD:NKA at cognate sites KLFB/KLFC (left) or KLF4/KLFC (right) on B-DNA NANK model generates no clashes. c B-DNA NANK model with canonical DBDs at KLF4 and KLFC shown as surfaces and two poses for DBD at KLFB in cartoons. ZnF1 of DBD:NKA (yellow) clashes with DBD modeled at KLF4 (top right) but an alternate ZnF1 pose (2wbs, red) does not. Positions of key residue:base contacts indicated in cyan. d Distances between modeled NANK 5′ ends and H446 Cα of ZnF1 sterically excluded from KLF4 (left) or KLFB (right) favor 5′ labeling the coding strand. e Superimposing a second duplex on the excluded ZnFs suggests that 5′ labels could be used to detect DBD-mediated DNA bridging in solution by smFRET. f FRET efficiency (Em/F) histograms of mixtures of 100 pM Alexa 488-labeled NANK and 500 pM Alexa 594-labeled NANK in the absence (left) or presence (right) of 1 μM KLF4 DBD were fit to Gaussian functions (donor emission, green; acceptor/FRET emission, red). Bottom schematic emphasizes that the atomic details of DNA:DBD:DNA bridging are not defined experimentally.

We then transfected cells with constructs carrying wild type, R501A mutant, or E476D/R501A double mutant KLF4-mTurq and assessed their distributions by microscopy. More puncta are seen for wild type than the mutants (Fig. 5e, top), but mutant expression levels are lower than wild type. Automating the identification of “punctate” cells (≥5 puncta detected) and plotting cells by their average fluorescence values reveals that both mutant proteins can be expressed at higher levels than wild type without conferring a “punctate” phenotype (Fig. 5f). At levels between 5.0 and 7.5 × 10^3 arbitrary units, all cells expressing wild type fusions are classified as “punctate” but fewer than half of those expressing mutant proteins are so classified (Fig. 5f). Visual comparison of HEK 293T cells (Fig. 5e, bottom) or BJ fibroblasts (Supplementary Fig. 8) with equivalent average fluorescence confirms that the wild type construct supports a more punctate distribution than the point mutants. We conclude that at equivalent expression levels, both KLF4R501A-mTurq and KLF4E476D/R501A-mTurq undergo biomolecular condensation less readily than KLF4-mTurq. We infer that the DNA-contacting surfaces of ZnF2 and ZnF3 are therefore important to condensation mediated by full-length KLF4, and that the observed condensation is likely mediated by KLF4 molecules whose DNA binding domain is properly folded.

**KLF4 biomolecular condensates recruit SOX2 and OCT4.** We then tested whether SOX2 and OCT4, TFs that cooperate with KLF4 at promoters and enhancers\(^{5,6,14,19}\), would co-localize to KLF4-mediated condensates by co-expressing KLF4-mTurq with OCT4-mCherry or SOX2-mCherry. OCT4-mCherry expressed alone shows a uniform nuclear distribution at low levels, with some tiny puncta at higher expression levels (Fig. 6a). SOX2-mCherry expressed alone shows distributions consistent with SOX2 acting as a bookmark for mitosis\(^{49}\); although usually uniform or showing tiny puncta, in some cells it highlights mitotic chromosomes (Fig. 6a). After co-transfection of vectors for TF-mCherry and KLF4-mTurq, only a fraction of cells express both tagged proteins. OCT4-mCherry co-localizes to KLF4-mTurq droplets and puncta (Fig. 6b, top) in all cells where both proteins are detected (n = 73, 2 biological replicates); OCT4-mCherry droplets are never seen in the absence of KLF4-mTurq. SOX2-mCherry co-localizes to KLF4 puncta and droplets in 74% of cells where both KLF4-mTurq and SOX2-mCherry fluorescence are detected (n = 39, 2 biological replicates); when SOX2-mCherry does not co-localize with KLF4-mTurq, its distribution resembles mitotic bookmarking (Fig. 6c, bottom). We conclude that the cellular KLF4 condensate can recruit OCT4 or SOX2.

To determine if the in vitro DBD:DNA condensed phase can recruit TFs, we labeled purified full-length OCT4 and SOX2 proteins with Alexa Fluor 647, mixed them with NANK (which lacks OCT4 or SOX2 cognate binding sites) with or without DBD, and monitored the mixtures by fluorescence microscopy. OCT4-AF647 or SOX2-AF647 mixed with NANK give homogeneous mixtures, but addition of DBD drives NANK into droplets that co-localize with OCT4-AF647 or SOX2-AF647 (Fig. 6d).

We conclude that the DBD-mediated biomolecular condensate can recruit TFs, perhaps through non-specific TF:NANK interactions.

We then assessed DBD behavior with a polynucleosome substrate consisting of a 5 kbp plasmid DNA with sites for 11 nucleosomes and at least 9 GGTG motifs (Active Motif, Inc.). DBD colocalizes to droplets with polynucleosomes (Fig. 6e), and droplets induced by mixing DBD with polynucleosomes recruit labeled OCT4 or SOX2 (Fig. 6f). DBD condenses with this substrate at low concentrations: 250 nM DBD induces droplets with 210 pM plasmid/nucleosome complex (0.4 ng DNA/μl, Fig. 6g, left). This might reflect DBD enhancing the intrinsic ability of polynucleosomes to phase separate\(^{50}\). The binding of KLF4 to DNA in nucleosomes\(^{16}\) might mediate this enhancement or independently support condensation, but exposed plasmid DNA in this substrate might also drive condensation by recruiting many DBDs, giving it increased valency\(^{31,52}\).

**KLF4 DBD condenses readily with long DNAs.** To see if longer DNAs containing NANK could condense readily without nucleosomes, we examined NP, a 404 bp NANOG promoter fragment (−379 to +25) that includes NANK and 6 additional GGTG sites. 250 nM DBD readily condenses with 2.5 nM NP, but not with NANK at the same DNA weight concentration (0.6 ng/μl, 32 nM) (Fig. 6g, center panels). The threshold concentration for NANK at 6 μM DBD is 125 nM (Fig. 5b), so NP condenses at 24-fold lower DBD levels and 4-fold lower DNA weight concentration (50-fold lower mole concentration) than NANK. 250 nM DBD condenses robustly with 0.6 ng/μl (130 pM) NPE, a 7.4 kbp linear DNA containing portions of the NANOG promoter and its −5 enhancer\(^{39}\) and 93 G(C/T)G sites (Fig. 6g, right). We conclude that long DNAs condense much more readily than short DNAs.

**Discussion**

We propose that KLF4 organizes chromatin by forming condensates at genomic loci to which it is recruited in high numbers and then stabilizing the colocalization of such genomic sites when their KLF4:DNA condensates fuse during random diffusive collisions (Fig. 7). For the initial condensation, we expect that KLF4 would bind tightly to 6 bp on one DNA through ZnF2/ZnF3 but more weakly to another DNA through ZnF1, in a bridging mode, and that several KLF4 bound to one stretch of DNA would provide the valency needed to drive biomolecular condensation\(^{31,52}\). Cognate KLF4 sites\(^{19,23}\) and 93 G(C/T)G sites (Fig. 4a, c), and partial 6 bp sites\(^{16}\) that might direct KLF4:DNA condensation at particular genomic loci will have their affinities modulated by CpG methylation and their accessibility influenced by nucleosomes and by other DNA binding proteins. KLF4:DNA condensation in vitro does not require IDR:IDR interactions, but in cells the KLF4 IDR may contribute to condensation (through homotypic interactions) or to recruitment of other factors.
Chromatin modifying machinery would be able to reinforce or reverse KLF4:DNA condensation by altering the accessibility or methylation states of KLF4 binding sites.

KLF4 is found at both repressive and activating loops in PSCs, indicating that contacts mediated by KLF4:DNA condensates are not sufficient to drive transcriptional activation.

Spatial colocalization of genomic elements by KLF4:DNA condensates combined with IDR-centric models for transcriptional activation can explain many observed chromatin features. Promoters and enhancers associated with pluripotency are known to recruit KLF4 when they make long-range contacts; we propose that KLF4 is condensed with DNA at these loci, helping to stabilize the observed long-range contacts.

KLF4 mutations in cells

KLF4 DBD mutations in cells

Spatial colocalization of genomic elements by KLF4:DNA condensates combined with IDR-centric models for transcriptional activation can explain many observed chromatin features. Promoters and enhancers associated with pluripotency are known to recruit KLF4 when they make long-range contacts; we propose that KLF4 is condensed with DNA at these loci, helping to stabilize the observed long-range contacts.
Super-enhancers are larger than typical ESC enhancers, more enriched in KLF4, and able to recruit much higher levels of Mediator. These properties can be explained by extensive KLF4-DNA condensates that bring together several enhancers, whose abilities to recruit transcription machinery through IDR-IDR interactions would be increased by their mutual proximity and by recruitment of TFs to the KLF4-DNA condensate. The KLF4-mediated recruitment of histone demethylase JMJD3 or DNA demethylase TET2 may be influenced by KLF4-DNA condensation, and the KLF4-mediated recruitment of cohesin may help to topologically link remote DNA segments held together by KLF4-DNA condensation.

KLF4 is functionally implicated at the NANOG promoter in somatic cell reprogramming. We propose KLF4 binding and condensation as the first mechanistic steps in accessing the closed, highly methylated NANOG promoter during reprogramming. Silenced chromatin is compact, but KLF4 should diffuse into it readily because its folded ZnFs are small and its IDR is deformable. The human NANOG promoter has KLF4 cognate sites spaced by 15, 7, and 11 bp, so one of these sites must be partially exposed in nucleosomes, and KLF4 is known to bind to 6 bp partial sites in nucleosomal DNA. KLF4 that binds to the CpG-methylated, nucleosome-wrapped NANOG promoter can recruit more KLF4 through condensation driven by its other three ZnFs, such contacts could be long-lived. ZnFs with unidentified functional roles are also common in proteins with repressive effects in chromatin: the repressor ZFP57 binds a methylated 6 bp motif in closed chromatin with two of its seven ZnFs, and the N-terminal ZnF of the mouse repressor protein ZFP568 does not contact target DNA. The architectural protein CTCF, which interacts through its N-terminal domain with cohesin and whose binding site polarity on DNA controls chromatin looping, makes sequence-specific contacts with different target DNAs but its terminal ZnFs (ZnF1, ZnF10, and ZnF11) do not contribute to binding target DNA. Our demonstration that the KLF4 ZnF tandem array makes DNA-bridging contacts that mediate condensation suggests that other C2H2 tandem ZnF proteins may bridge DNA making transient or long-lived contacts that contribute to biological function.

**Methods**

**Bacterial strains.** The *E. coli* strain DH5α (Thermo Fisher Scientific) was used for plasmid cloning and large-scale preparations of plasmid DNAs. The *E. coli* strain BL21 Star (DE3) (Thermo Fisher Scientific) was used for large-scale protein production.

**Mammalian cell lines.** The HEK 293T cell line (from ATCC, CRL-3216), Lenti-X 293T (from Takara Bio USA, TaKaRa Bio # 632180), and BJ fibroblasts (from ATCC, CRL-2522) were cultivated in Dulbecco’s Modified Eagle Medium (DMEM, Corning) with 10% (v/v) fetal bovine serum (FBS, Corning) and 1X antibiotic-antimycotic solution (Corning). All cells used in this study tested negative for mycoplasma contamination.

**Construction of mammalian plasmids.** All generated constructs and mutations were confirmed by DNA sequencing (Eurofins Genomics). The pHRT-GFP-ΔH lentiviral transfer vector was generated from pHR-CMV-TetO2_3C-Avi-His6 (Addgene #113887) by replacing the DNA fragment corresponding to the 5'Chicken RPTPs signal sequence-HRV 3C site-3' with a DNA fragment corresponding to 5'-BsmBI-H-KpnI-TEV cleavage site-eGFP-3’. The insert fragment was amplified from the plasmid encoding TEV-eGFP using the primers eGFP-F/eGFP-R (see Supplementary Table 1 for all primers). The vector fragment was amplified

---

**Fig. 5 DNA affinity influences KLF4-mediated liquid condensation.** a Fluorescence microscopy images of 3 μM duplex DNA in TS buffer and 100 nM YOYO-1 with (bottom) or without (top) 10 μM DBD. Similar results were obtained for 2 replicates. b Fluorescence microscopy images of DBD-DNA mixtures at 3:1 and 3:1.5 ratios show that adding cognate DNA can reverse LLPS. Similar results were obtained for 2 replicates. c Mixing DBD and DNAs at different concentrations produces condensates (black circles) or homogeneous solutions (open circles) that define LLPS phase diagrams. Images were scored as condensate if the coefficient of variation was >0.2 and the mean fluorescence intensity was >0.4 arbitrary units. Values were determined using ImageJ from 2-3 independent replicates. Fluorescence images at right are taken at 3 μM DBD and 250 nM DNA, conditions boxed in red in the phase diagrams. d LLPS phase diagrams determined as in (c) but for double mutant DBD/R501A with DNAs. e Fluorescence images of HEK 293T cells expressing KLF4-mTurq (left), KLF4R501A-mTurq (middle) and KLF4E476D/R501A-mTurq (right). Bottom row shows close-ups of single cells with the same mean fluorescence (-5000 AU). Similar results were obtained for 2 biological replicates. f HEK 293T cells expressing KLF4-mTurq variants classified as diffuse (gray, left) or punctate (red) are plotted by mean fluorescence intensity; the long horizontal bars mark the median, and the top and bottom bars on the vertical lines denote the 90-10 percentiles. Wildtype “diffuse” cells show lower fluorescence than those cells expressing KLF4R501A-mTurq or KLF4E476D/R501A-mTurq (n = 112 diffuse cells for each fusion across 2 biological replicates). Cells with at least 5 distinct puncta (0.5 μm spots with intensity center >1500 arbitrary units, determined using Imaris software) were classified as punctate. Statistical tests were performed using two-sided Student’s paired t-test. Y-axis limit was set to 30,000 for visualization purposes. Fields that yielded >112 diffuse cells also gave punctate cell counts of 209 (WT), 39 (KLF4R501A-mTurq) and 64 (KLF4E476D/R501A-mTurq) across 2 biological replicates.
from pH-CMV-TetO2-3C-Avi-His6 using the primers pHRT-1F/pHRT-1R. The insert and vector fragments were ligated together using Gibson Assembly Master Mix (NEB) according to the manufacturer’s protocol.

Lentiviral vectors pHRT-mTurquoise2-AH and pHRT-mCherry-AH were generated by replacing the eGFP gene in pHRT-mTurquoise2-AH with mTurquoise2 (mTu) and mCherry (mCh), respectively. The mTu insert was amplified from pmTurquoise2-Tubulin (Addgene #36202) using the primers mTu-1F/mTu-1R. The mTu gene has the A206K mutation to ensure obligate-monomer state71. The mCh insert was amplified from pBRY-nuclear mCherry-IRES-PURO (Addgene #52409) using the primers mTu-1F/mTu-1R. The vector fragment was amplified from pHRT-GFP-AH using the primers pHRT-1F/pHRT-2R. The insert and vector fragments were ligated together using Gibson Assembly Master Mix.

Vector pHRT-KLF4-mTurquoise2-AH was constructed to express KLF4 with a C-terminal TEV cleavage site (mTurquoise2-Avi-His6; mTu-AH) in a lentiviral expression system. The KLF4 insert was amplified from the plasmid encoding KLF4 (GeneArt) using primers KLF4-1F/KLF4-1. The insert fragment was digested with BamHI/KpnI and ligated into BamHI/KpnI-digested pHRT-mTurquoise2-AH lentiviral transfer vector using T4 DNA ligase (Promega).

pHRT-KLF4(2–417)-mTu-AH and pHRT-KLF4(418–513)-mTu-AH were constructed to express labeled KLF4 deletion constructs. The KLF4 coding regions were amplified and digested with BamHI/KpnI and ligated into pHRT-mTurquoise2-AH lentiviral transfer vector using T4 DNA ligase (Promega).

Fig. 6 KLF4 condensates recruit TFs and form at low concentrations with long DNAs. a Fluorescence microscopy images of HEK 293T cells expressing OCT4-mCherry (left) or SOX2-mCherry (right). Some SOX2-mCherry distributions suggest mitotic bookmarking49 (white arrow). b Fluorescence microscopy images of HEK 293T cells. Screening ~1000 cells from 2 transfection replicates identified 73 cells that co-express OCT4-mCherry and KLF4-mTurquoise2; all 73 show tags colocalized to droplets (examples top and bottom). c Fluorescence microscopy images of HEK 293T cells. Screening ~1000 cells from 2 transfection replicates identified 39 cells that co-express SOX2-mCherry and KLF4-mTurquoise2; 29 cells show tags colocalized to droplets (example, top row) and 10 cells do not (example, bottom row). d Fluorescence microscopy images of 1.5 μM NANK DNA with 100 nM YOYO-1 and 50 nM OCT4-AF647 (top) or 70 nM SOX2-AF647 (bottom) without (left column) or with (right three columns) 9 μM DBD. Similar results were obtained for 2 replicates.

e Fluorescence microscopy images of polynucleosomes (green; 20 ng DNA/μl, Active Motif, Inc., visualized with 100 nM YOYO-1) with 10 μM DBD (red; trace labeled 1:100 with DBD-AF594). Similar results were obtained for 2 replicates.

f Fluorescence microscopy images of polynucleosomes (green; 11 ng DNA/μl, visualized with 100 nM YOYO-1) with OCT4-AF647 (50 nM; purple, top row) or SOX2-AF647 (70 nM; purple, bottom row) alone (left) or with 1 μM DBD (three right panels). Similar results were obtained for 2 replicates. g Longer DNAs undergo LLPS with DBD at low concentrations. Fluorescence microscopy images of DBD with 30 bp NANK, 404 bp NP (NANOG promoter), 7.4 kbp NPE (NANOG promoter enhancer) and 5 kbp plasmid DNA in nucleosomes (left to right panels, respectively). Conditions: DBD (250 nM) mixed with different DNA (mass equivalent of 0.6 ng/μl; 32 nM NANK or 2.5 nM NP or 140 pM NPE or 210 pM plasmid DNA concentration in nucleosomes) in T5 buffer with 100 nM YOYO-1. Similar results were obtained for 2 replicates.
Corresponding to the intrinsically disordered region (IDR, residues 2-417) of the DNA binding domain (DBD, residues 418-513), two fragments were amplified by the plasmid encoding human KLF4 gene (GeneArt) using primer sets KLF4-2F/KLF4-2R or KLF4-3F/KLF4-3R, respectively, and ligated into BamHI/Kpn1-digested pHR-T-mTu-AH lentiviral transfer vector using Gibson Assembly Master Mix. To construct the KLF4 single mutant (R501A), two DNA fragments were amplified from pHRT-KLF4-mTu-AH using the primer sets (KLF4-4F/KLF4-4R and KLF4-5F/KLF4-5R, respectively) and ligated together using Gibson Assembly Master Mix.

To construct the KLF4 double mutant (E476D/R501A) expression vector pHRT-KLF4(418 R501A)-mTu-AH, two DNA fragments were amplified from pHRT-KLF4 R501A-mTu-AH using the primer sets KLF4-6F/KLF4-10R according to the manufacturer optimized protocols. Lentiviruses were harvested 3 days post-transfection, filtered through 0.45 μm-pore size PES filters, and concentrated 100 times using Lenti Concentrator (OriGene). The lentiviral titer concentration, determined by Lenti Virus Titer Kit HIV-1 p23 Elisa Assay (OriGene), was ~2 × 10^8 TU/mL.

Fig. 7 KLF4:DNA condensation as an organizer of chromatin. a KLF4 levels rise early in reprogramming, leading to KLF4 biomolecular condensation. b-g Proposed roles for KLF4:DNA condensates in chromatin reorganization during reprogramming. b Pluripotency-related sites in closed chromatin of somatic cells lack KLF4. c Sites will recruit KLF4 and form local KLF4:DNA condensates as expression levels rise. d Random diffusive encounters of cis-acting elements leads to fusion of their condensates and persistent colocalization. e Locis on different chromosomes can also be co-localized in KLF4:DNA condensates. f Within the condensate, KLF4 makes DNA bridging contacts at cognate 9 bp KLF4 sites (yellow), especially where such sites overlap, but perhaps also at 6 bp partial sites or non-cognate sites. The KLF4:DNA condensate enriches TFs relative to solution, helping to saturate binding sites for TFs OCT4 and SOX2 and to recruit co-activators such as CBP/p300 to activate NANOG expression and gain access to pluripotency.

Construction of bacterial plasmids. To express fluorescein-conjugated Nano tag-His6-TEV cleavage site (NH66) fused KLF4 DBD in the bacterial expression system, the KLF4 DBD gene (insert) was amplified with the plasmid encoding KLF4 (GeneArt) using primer sets (KLF4-8F/KLF4-8R). The vector fragment was amplified from pET15-Nano6HT-SMAD1 (DNASU) using primer sets (pET15-1F/pET15-1R). The insert and vector fragments were ligated together using Gibson Assembly Master Mix to give pET15-NH66-KLF4(418–513). KLF4 DBD mutations E476D and R501A were introduced into plasmid pET15-NH66-KLF4(418–513) using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent) and the primers KLF4 E476D and KLF4 R501A according to the manufacturer’s protocols. Product pET15-NH66-KLF4(418–513) E476D R501A was verified by DNA sequencing; residue numbering follows the human gene product.

Lentiviral transfection and transduction. Recombinant lentiviruses were produced by co-transfection of Lenti-X 293T cells (1 × 10^6 cells in gelatin-coated 10 cm cell culture dish) with lentiviral transfer construct (1.2 pmol), psPAX2 packaging plasmid (1.2 pmol; Addgene #12260), and pMD.2G envelope plasmid (0.7 pmol; Addgene #12259) using the Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific). Transfection was performed according to the manufacturer’s optimized protocols. Lentiviruses were harvested 3 days post-transfection, filtered through 0.45 μm-pore size PES filters, and concentrated 100 times using Lenti Concentrator (OriGene). The lentiviral titer concentration, determined by Lentivirus Titer Kit HIV-1 p23 Elisa Assay (OriGene), was ~2 × 10^8 TU/mL.

Lentiviral transduction was carried out with 1 × 10^5 host cells in 24-well cell culture plate and 10 multiplicities of infection (MOIs) of lentivirus in the presence of 8 μg/mL of polybrene. Lentivirus was removed after overnight incubation, and fresh cell culture media was added. Two days post-transduction, expression of the
fluorescent proteins (eGFP, mTurquoise2, and mCherry) were verified using an EVO fluorescence microscope (Thermo Fisher Scientific). Transduced cells were passaged in culture as bulk preparation for functional assays.

**Purification of KLF4 DBD (418-513; DBD) WT and double mutant (E476D/R501A).** Either pET15-NH6-KLF4(418–513) or pET15-NH6-KLF4(418–513) _E476D_R501A_ were transformed into E. coli BL21 Star competent cells (Novagen, Merck KGaA, Darmstadt, Germany). Transformed cells were grown at 37°C in Terrific Broth media containing 100 μg/ml carbenicillin antibiotic until optical density at 600 nm (OD600) reached 0.6. The culture was then transferred to an 18°C shaker for shaking at 250 rpm until OD600 reached 0.8–1.0. Protein expression was induced with 1 mM IPTG, followed by overnight growth at 18°C with shaking at 250 rpm. Cells from the overnight culture were harvested by centrifugation at 7900 × g. Both WT and double mutant (E476D/R501A) KLF4 DBD were purified with the same procedure. The cell culture pellets were resuspended in denaturing lysis buffer (6 M urea, 1 mM 2-mercaptoethanol, 0.5 M NaCl, 10 mM sodium phosphate, pH 8). The resuspended pellets were lysed using a cell homogenizer (Avestin, Ottawa, Canada), with the soluble fraction separated from the cell debris by centrifugation at 38,700 × g for 30 min at 4°C. The supernatant was combined with 3 volumes of 0.1% (v/v) TFA, and acidi

**Fluorescent labeling of OCT4 and SOX2.** OCT4 and SOX2 were labeled using Alexa Fluor 488 (AF488) maleimide (Thermo Fisher Scientific) using standard methods described previously. Briefly, the proteins were dissolved in 6 M GdnHCl, 20 mM Tris pH 8, mixed with 3–4 molar excess of Alexa Fluor 647 (AF647) maleimide dyes, and incubated for 1 h at RT. Samples were then mixed with 3-fold excess of 0.1% TFA/dH2O and purified by reverse phase HPLC. Purified fluorescent labeled samples were lyophilized and stored at −80°C. OCT4-AF647 and OCT4-AF488 protein samples had 45% and 135% fluorescent labeling efficiency, respectively. For colocalization experiments (Fig. 6d, f), SOX2-AF647 was replaced with 6 M GdnHCl−EDTA−200 mM sodium phosphate buffer, pH 8. OCT4-AF647 protein was dissolved in 6 M GdnHCl and then mixed with OCT4−NP−5 columns (GE Healthcare) to a final concentration of ~100 nM in 10 mM sodium phosphate buffer, pH 8. Samples were snap frozen for storage at −80°C before use.

**Fluorescent labeling of NANK.** NANK DNA oligos with 5′ amino modified C6 (DTT) were purified by ethanol precipitation and labeled with a 10-fold molar excess of dye (Alexa Fluor 488 or 594 NHS ester; Invitrogen). NANK has 30% (Alexa Fluor 488) and 50% (Alexa Fluor 594) fluorescent labeling efficiency. The labeling reactions were performed at 30°C with 30–60 min incubation. The Alexa Fluor 488− or 594-labeled NANK were then ethanol precipitated; the collected free NHS or NHS− were dissolved in 0.1 M triethanolamine acetate at pH 7. Excess unconjugated dyes were removed by passing twice over NAP-5 columns (GE Healthcare).

**Preparation of 404 bp NP (NANOG promoter) and 7.4 kbp NPE (NANOG promoter enhancer).** The human NANOG promoter and enhancer was ligated into pGL-Nanog-Luc (Addgene #25990) using the forward primer hNan-F2 (or -F1) and the reverse primer hNan-R1 (or -R2); the primers in parentheses are unlabeled. The 404 bp PCR fragments were purified using QIAquick Gel Extraction Kit (Qagen). The pGL-Nanog-P5E minus plasmid23 containing the mouse 1535 bp NANOG promoter and 1337 bp enhancer (~3 kbp from NANOG promoter) was digested with PmeI (NEB) to linearize the plasmid. The DNA fragment containing NANOG promoter and enhancer was purified using QIAquick Gel Extraction Kit (Qagen).

**Electrophoretic mobility shift assay (EMSA).** The binding reactions for the EMSA consisted of 1× EMSA buffer (0.01 mg/ml BSA, 0.1 mM DTT and 0.05 mM TCEP, 5% glycerol, 50 mM NaCl, 20 mM Tris pH 8) and unlabeled (50 mM−15 mM) protein (see figure legends for the exact protein and DNA concentration, and buffer conditions). Protein concentrations were prepared by 2-fold serial dilution. Samples were loaded onto either 10%, 12% or 4% gel in TBE or Sybr® Green for 20 min prior to imaging. The gels were then imaged using ChemiDoc with the appropriate filters and analyzed through the Image Lab software (Bio-Rad). EMSAs were performed with 2–3 independent replicates.

**Cristallization and X-ray data collection.** The KL44 DBD:NKA complex was crystallized by hanging drop vapor diffusion method at 20°C. Purified and refolded human KL44 (418–513) in 0.5 mM DTT, 20 mM Tris-Cl, pH 8.0, and 3.3 mM HgCl2 was mixed with 1.2 molar excess of dodecameric DNA (12 mer: 5′-AGG GGG TGC TGC-3′). Crystals of KL44 DBD:NKA were obtained by mixing equal volumes of KL44 DBD:NKA complex (40 mg/ml total protein) with different concentrations of precipitants (60% PEG 3350, 0.1 M sodium acetate, pH 6 and 0.1 M sodium acetate, pH 8) and stored at 2-3°C for 1-2 weeks.
macromolecule) with 0.2 M sodium iodide (pH 7.0) and 20% w/v polyethylene glycol 3,330 reservoir solution. Single crystal X-ray diffraction data were collected at 100 K on the Beam Line 5.0.2 Advanced Light Source (UC Berkeley, USA) at wavelength (λ) = 1.00 Å, using an ADSQ Q170 CCD detector. The collected data were integrated and scaled using iMOSl and SCALA, respectively.7,47

Structure solution and refinement. The crystal structure of KLF4 DBD:NKA was determined by molecular replacement method using Phaser.48 A prior crystal structure of KLF4 DNA binding domain (PDB ID: 2WBS) was used as search model. The unique solution was obtained for the molecule in the asymmetric unit. The dodecameric DNA was traced and fitted manually into electron density. The final model was obtained by iterative cycles of manual rebuilding using Coot49 and refinement using phenix.refine50. PyMOL visualization program (https://pymol.org) was used for all the structural analyses and preparation of figures. The statistical data on data collection and refinement are summarized in Supplementary Table 1. Residue numbering follows the human gene product; previous structures with identical ZnF sequences have been numbered according to the mouse gene product.

In vitro LLPS microscopy imaging. Monitoring for the presence/absence of LLPS droplets was performed at room temperature using EVOS fluorescence imaging system (Thermo Fisher Scientific) with bright field and/or necessary filters (CFP (mTurquoise2), GFP (YOYO-1, AF488), Texas Red (AF594), Cy5 (AF647)). For a set of experiments, the same light power and exposure time was used. Conditions for each set of experiments are detailed in the figure legends. To construct LLPS diagrams, various concentrations of KLF4 DBD and DNAs (cognate and non-cognate) were prepared with either of the following buffers: TS buffer (70 mM NaCl, 12.5 mM Tris, pH 7.4; Figs. 2i, 5) or TS2 Buffer (140 mM NaCl, 25 mM Tris, pH 7.4; Supplementary Fig. 3). 100 nM YOYO-1 was added to samples in which the DNA was to be imaged by fluorescence microscopy. Specific conditions for the experiments are in the figure legends. LLPS diagrams were based on images obtained after 30 min of incubation. To assess colocalization of KLF4 DBD-NANK droplets and full length OCT4 or SOX2, KLF4 DBD (9 µM) was mixed with NANK DNA (1.5 µM) and either OCT4-AF647 (95 nM) or SOX2-AF647 (1:40 nM) in TS buffer. Samples were incubated for 30 min to 1 h prior to imaging. Experiments were performed in 2–3 independent replicates. To assess colocalization of KLF4 DBD with recombinant polycomissures purchased from Active Motif, as in Fig. 6e, the commercial polycomissures (H31; 20 µg protein + 24 kg/lbs plasmid DNA; 0.55 µg/µl) in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 2 mM DTT, 20% glycerol were diluted (final concentration is 20 ng/µl) in TS buffer and mixed with KLF4 DBD (10 µM). To assess colocalization of KLF4 DBD, polycomissures, and OCT4 or SOX2, KLF4 DBD (1 µM) was mixed with commercial polycomissures (11 ng/µl and OCT4-AF647 (50 nM) or SOX2-AF647 (70 nM)) in TS buffer. Samples were incubated for 30 min to 1 h prior to imaging. Experiments were performed in 2–3 independent replicates.

Fluorescence live cell confocal imaging. Fluorescence imaging of live cells (HEK 293 T cells and BJ fibroblasts plated on poly-L-lysine coated 35 mm ibidi µdish transwell with complete µdishes medium) was performed 2–3 days after transient transduction using EVOS fluorescence imaging system (Thermo Fisher Scientific) or LSM780 and LSM880 laser-scanning confocal microscope system (Zeiss, Oberkochen, Germany) at 37 °C and 5% CO2 with a ×60 oil objective. Images were analyzed using Fiji (ImageJ 1.52c), Zen 2.3 (Zeiss, Oberkochen, Germany) and Imaris v9.2 (Zurich, Switzerland) microscopy image analysis software. Images were taken at 3–5 different field locations for each biological replicate.

Fluorescence recovery after photo-bleaching (FRAP) imaging in cells and in vitro. FRAP imaging of KLF4-mTurq droplets and puncta/cluster in HEK 293T or BJ fibroblasts cells (2–3 days after lentiviral transduction) were performed using a Zeiss LSM780 and LSM880 laser-scanning confocal microscope system at 37 °C and 5% CO2. Different nuclear region of interest (ROI) spots (~0.5–2 µm diameter) were selected. ROIs were drawn in the ImageJ software. Following 2–3 baseline images, ROIs were bleached for 50-200 iterations at 100% laser power (458 nm and 488 nm), and were imaged for up to 2–4 min post-bleaching for fluorescence recovery. FRAP recovery curves were corrected for background photobleaching (reference ROI in a separate droplet) and normalized against pre-bleach intensity values. FRAP data are fitted with an exponential function in the software Origin (Fig. 1c). FRAP imaging of LLPS droplets in vitro was achieved for LLPS droplets prepared by mixing trace-labeled KLF4 DBD (9 µM unlabeled DBD, 50 nM DBD-AF594) with trace-labeled NANK DNA (1.5 µM unlabeled NANK, 180 nM NANK-AF488). After ~1 h sample incubation, FRAP imaging was performed on the droplets that had formed and settled close to the部ment surface. Using Zeiss LSM780 (with ×60 objective), different regions (~1 µm diameter ROI) were bleached with 100% power (488 nm) and 90% power (594 nm) for 100 iterations. Pre- and post-bleaching images (simultaneous 488 and 594 nm channels) were collected for ~15 min with 5 s intervals. After background subtraction (reference ROI in separate droplet) and normalization, the FRAP recovery curves (means and standard deviations) were plotted in the software Origin (Fig. 2g).

Single-molecule Förster resonance energy transfer (smFRET). KLF4 DBD and NANK DNA binding interactions were monitored by single-molecule spectroscopy using a custom-built confocal laser microscopy system (ISS, Champaign, Illinois). smFRET measurements were conducted in TS buffer (70 mM NaCl, 12.5 mM Tris, pH 7.4) at room temperature (21 ± 1 °C) by mixing 100 pM NANK 5'-labeled with Alexa Fluor 488 (FRET donor; Thermo Fisher Scientific) and 500 pM NANK 5'-labeled with Alexa Fluor 594 (FRET acceptor; Thermo Fisher Scientific), with or without 1 µM KLF4 DBD. Measurements were performed with 2 independent replicates. Freely diffusing FRET samples were excited with a 488-nm laser (ISS; ~115 µW). Fluorescence emission was split into donor-acceptor fluorescence by a 605-nm long pass beam splitter dichroic, and donor and acceptor signals were further filtered using 535/50-nm and 641/75-nm bandpass emission filters, respectively. Emission was detected using SPCM-ARQH-16 Avalanche photodiode detectors (Excitellas Technologies Corp., Waltham, MA). Data acquisition and FRET efficiency analysis were performed using Vistavision (64) 4.2.220.0 (ISS), correcting for acceptor emission due to direct excitation (1%) and fluorescence bleed-through of donor emission into the acceptor channel (5%), applying a binning time of 500 µs. There were 40,335 and 33,160 events collected for DNA samples without DBD and with DBD, respectively (Fig. 4). smFRET histograms were fitted to Gaussian functions using OriginPro 2020 (OriginLab, Northampton, MA, USA). FRET efficiencies (E_FRET) were calculated (using a value of unity for γ from the corrected donor (I_D) and acceptor (I_A) fluorescence intensities as given by:

\[
E_{\text{FRET}} = \frac{I_D}{N + I_D} + \frac{N}{I_A}
\]

LLPS quantification and statistical analysis. To construct phase diagrams, a matrix of different nucleic acid and protein concentrations were mixed and incubated for 30 min. Images (fixed size of 153 × 114.7 µm) were collected at the same focal plane using EVOS microscopy system. The mean fluorescent intensities and standard deviation of the *t* images were determined by the ImageJ software. Data from 2–3 independent replicates were averaged; the coefficient of variation (CV) is determined by the standard deviation divided by the mean. Positive phase separation for a particular condition is determined by CV > 0.2 and mean fluorescent intensity > 4 arb. units (Figs. 2 and 5).

Quantification of fluorescence intensities and puncta in cells. Statistical tests (student’s paired *t*-test) performed on experimental data and their representations are performed using Origin and noted in the figure legends. Puncta/droplet identification was determined through the Spots Algorithm in Imaris software v9.2. Only spots that are localized in the nucleus, >500 nm in diameter and >1500 arb. units center intensity were chosen. The mean fluorescent intensities were determined by the Imaris software for the HEK 293T cells (Figs. 2 and 5) and Imagel software for BJ fibroblasts (Supplementary Fig. 8).

Nuclear concentration determination of KLF4-mTurq. Nuclear concentrations of transiently expressed KLF4-mTurq were determined using a calibration plot of the fluorescence intensity/exposure time versus concentration of purified mTurquoise2 protein (Supplementary Fig. 2b). Using an EVOS fluorescence microscope, ×60 objective and CFP filter (Thermo Fisher Scientific), HEK 293T cells expressing KLF4-mTurq were imaged on 35 mm ibidi µ-dishes were imaged using 30% power, 15 ms exposure time. The nuclear boundaries were manually drawn in ImageJ and the mean fluorescent intensities quantified. The calibration plot was linearly fitted using Origin.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability Source for plots, raw data for counts and intensity measurements, and uncropped gel images generated in this study are provided in a Source data file. The structure factors and coordinates for the KLF4 DBD:KLFA structure have been deposited in the Protein Data Bank under the accession number 6vtx. Source data are provided with this paper.

Received: 8 July 2020; Accepted: 31 August 2021; Published online: 22 September 2021

References
1. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872 (2007).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006).
14. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers. *Cell* **133**, 2151–2164 (2011).

5. Chronis, C. et al. Cooperative binding of transcription factors orchestrates developmental patterning. *Science* **360**, 918–921 (2018).

30. Zhang, H. et al. RNA controls PolyQ protein phase transitions. *Science* **360**, 918–921 (2018).

3. Schmidt, R. & Plath, K. The roles of the reprogramming factors Oct4, Sox2 and KLF4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol.* **13**, 251 (2012).

4. Beers, J. et al. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 13119 (2015).

22. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

6. Wei, Z. et al. Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells** **27**, 2969–2978 (2009).

7. Zhang, P., Andrianakos, R., Yang, Y., Liu, C. & Lu, W. Kruppel-like factor 4 (KLF4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285**, 9180–9189 (2010).

8. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell** **122**, 947–956 (2005).

11. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **451**, 1127–1131 (2008).

21. Larson, A. G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

24. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and macrophase differentiation. *Cell Mol. Life Sci.* **68**, 3211–3231 (2011).

56. Z. et al. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Stem Cells** **33**, 36–47 (2015).

16. Souza, M. C. et al. DNA conformation induces adaptable binding by tandem zinc fingers. *Mol. Cell** **75**, 543–559 (2019).

212. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

5. Chronis, C. et al. Cooperative binding of transcription factors orchestrates developmental patterning. *Science* **360**, 918–921 (2018).

4. Beers, J. et al. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 13119 (2015).

22. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

6. Wei, Z. et al. Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells** **27**, 2969–2978 (2009).

7. Zhang, P., Andrianakos, R., Yang, Y., Liu, C. & Lu, W. Kruppel-like factor 4 (KLF4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285**, 9180–9189 (2010).

8. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell** **122**, 947–956 (2005).

11. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **451**, 1127–1131 (2008).

21. Larson, A. G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

24. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and macrophase differentiation. *Cell Mol. Life Sci.* **68**, 3211–3231 (2011).

56. Z. et al. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Stem Cells** **33**, 36–47 (2015).

16. Souza, M. C. et al. DNA conformation induces adaptable binding by tandem zinc fingers. *Mol. Cell** **75**, 543–559 (2019).

212. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

5. Chronis, C. et al. Cooperative binding of transcription factors orchestrates developmental patterning. *Science* **360**, 918–921 (2018).

4. Beers, J. et al. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 13119 (2015).

22. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

6. Wei, Z. et al. Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells** **27**, 2969–2978 (2009).

7. Zhang, P., Andrianakos, R., Yang, Y., Liu, C. & Lu, W. Kruppel-like factor 4 (KLF4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285**, 9180–9189 (2010).

8. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell** **122**, 947–956 (2005).

11. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **451**, 1127–1131 (2008).

21. Larson, A. G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

24. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and macrophase differentiation. *Cell Mol. Life Sci.* **68**, 3211–3231 (2011).

56. Z. et al. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Stem Cells** **33**, 36–47 (2015).

16. Souza, M. C. et al. DNA conformation induces adaptable binding by tandem zinc fingers. *Mol. Cell** **75**, 543–559 (2019).

212. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

5. Chronis, C. et al. Cooperative binding of transcription factors orchestrates developmental patterning. *Science* **360**, 918–921 (2018).

4. Beers, J. et al. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 13119 (2015).

22. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

6. Wei, Z. et al. Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells** **27**, 2969–2978 (2009).

7. Zhang, P., Andrianakos, R., Yang, Y., Liu, C. & Lu, W. Kruppel-like factor 4 (KLF4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285**, 9180–9189 (2010).

8. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell** **122**, 947–956 (2005).

11. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **451**, 1127–1131 (2008).

21. Larson, A. G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

24. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and macrophase differentiation. *Cell Mol. Life Sci.* **68**, 3211–3231 (2011).

56. Z. et al. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Stem Cells** **33**, 36–47 (2015).

16. Souza, M. C. et al. DNA conformation induces adaptable binding by tandem zinc fingers. *Mol. Cell** **75**, 543–559 (2019).

212. Price, J. A. et al. Nuclear phase separation and domain formation in the chromatin fiber. *Cell Rep.* **26**, 1413–1423 (2018).

5. Chronis, C. et al. Cooperative binding of transcription factors orchestrates developmental patterning. *Science* **360**, 918–921 (2018).

4. Beers, J. et al. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci. Rep.* **5**, 13119 (2015).
69. de Wit, E. et al. CTCF binding polarity determines chromatin looping. Mol. Cell 60, 676–684 (2015).

70. Hashimoto, H. et al. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. Mol. Cell 66, 711–720 e713 (2017).

71. von Stetten, D., Noirclerc-Savoye, M., Goedhart, J., Gadella, T. W. Jr. & Royant, A. Structure of a fluorescent protein from Aequorea victoria bearing the obligate-monomer mutation A206K. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 68, 871–882 (2012).

72. Edelboch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochem. J. 6, 1948–1954 (1967).

73. Tatsukawa, A. V., You, Y. & Owczarzy, R. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. Biophys. Chem. 133, 66–70 (2008).

74. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D. Biol. Crystallogr. 67, 282–292 (2011).

75. Battye, T. G., Kontoglouinnis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D. Biol. Crystallogr. 67, 271–281 (2011).

76. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

77. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. 60, 2126–2132 (2004).

78. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D. Biol. Crystallogr. 68, 352–367 (2012).

Acknowledgements
We thank the Baylor College of Medicine Integrated Microscopy Core for the use of the confocal microscopes. We thank Phoebe S. Tsoi for help with DBD protein expression and for reading the manuscript. This work was supported by NIH grants R01 GM122763 to J.C.F. and R21 NS107792 to A.C.M.F. Additional funding was provided by R01 NS105874 and R21 NS109678 to A.C.M.F., by a Core Facility Support Award from the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01 and the Howard Hughes Medical Cancer Prevention and Research Institute of Texas (grant RP160805) to Dr. Martin Matzuk, and by R01 DK121970 and R61 HD099995 to Dr. Feng Li. The ALS-ENABLE facilities are supported in part by the National Institutes of Health, National Institute of Biomedical Imaging and Bioengineering, grant 5P41EB015906-33 to Dr. Kevin MacKenzie, by R01 DK121970 and R61 HD099995 to Dr. Feng Li. The ALS-ENABLE facilities are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01 and the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Of

Material in this article are included in the article Supplementary information available at https://doi.org/10.1038/s41467-021-25761-7.

Additional information
Competing interests
The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021