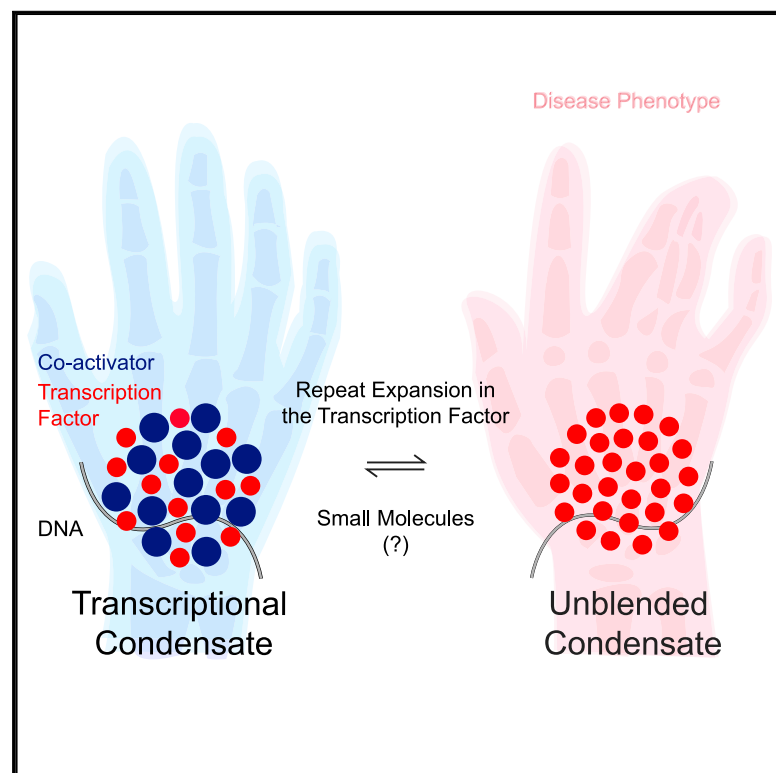


Unblending of Transcriptional Condensates in Human Repeat Expansion Disease

Graphical Abstract



Authors

Shaon Basu, Sebastian D. Mackowiak, Henri Niskanen, ..., Alexander Meissner, Daniel M. Ibrahim, Denes Hnisz

Correspondence

hnisz@molgen.mpg.de

In Brief

Disease-associated repeat expansions within transcription factors alter their capacity to co-condense with transcriptional co-activators.

Highlights

- Repeat expansions in transcription factors (TFs) alter their phase separation capacity
- Repeat expansions in TFs perturb the composition of TF-containing condensates
- Hoxd13 repeat expansions alter the transcriptional program in a synpolydactyly model
- Features of intrinsically disordered regions in TFs are linked to condensation behavior

Article

Unblending of Transcriptional Condensates in Human Repeat Expansion Disease

Shaon Basu,^{1,10} Sebastian D. Mackowiak,^{1,10} Henri Niskanen,^{1,9} Dora Knezevic,^{1,9} Vahid Asimi,¹ Stefanie Grosswendt,¹ Hylkje Geertsema,² Salaheddine Ali,^{3,4,5} Ivana Jerković,^{3,8} Helge Ewers,² Stefan Mundlos,^{3,4,5} Alexander Meissner,^{1,6,7} Daniel M. Ibrahim,^{3,4,5,11} and Denes Hnisz^{1,11,12,*}

¹Department of Genome Regulation, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

²Institute for Chemistry and Biochemistry, Free University Berlin, 14195 Berlin, Germany

³RG Development and Disease, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

⁴Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany

⁵Charité-Universitätsmedizin Berlin, BCRT-Berlin Institute of Health Center for Regenerative Therapies, 10178 Berlin, Germany

⁶Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁷Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

⁸Present address: CNRS-Institute of Human Genetics, Montpellier, France

⁹These authors contributed equally

¹⁰These authors contributed equally

¹¹Senior author

¹²Lead Contact

*Correspondence: hnisz@molgen.mpg.de

<https://doi.org/10.1016/j.cell.2020.04.018>

SUMMARY

Expansions of amino acid repeats occur in >20 inherited human disorders, and many occur in intrinsically disordered regions (IDRs) of transcription factors (TFs). Such diseases are associated with protein aggregation, but the contribution of aggregates to pathology has been controversial. Here, we report that alanine repeat expansions in the HOXD13 TF, which cause hereditary synpolydactyly in humans, alter its phase separation capacity and its capacity to co-condense with transcriptional co-activators. HOXD13 repeat expansions perturb the composition of HOXD13-containing condensates *in vitro* and *in vivo* and alter the transcriptional program in a cell-specific manner in a mouse model of synpolydactyly. Disease-associated repeat expansions in other TFs (HOXA13, RUNX2, and TBP) were similarly found to alter their phase separation. These results suggest that unblending of transcriptional condensates may underlie human pathologies. We present a molecular classification of TF IDRs, which provides a framework to dissect TF function in diseases associated with transcriptional dysregulation.

INTRODUCTION

More than 30 inherited human disorders are caused by an abnormal expansion of short, repetitive DNA sequence elements (Albrecht and Mundlos, 2005; Darling and Uversky, 2017; La Spada and Taylor, 2010; Orr and Zoghbi, 2007). The majority (>20) of such repeat expansions occur in protein coding genes and lead to expansions of homopolymeric alanine or glutamine repeats in cellular proteins. Glutamine repeat expansions are typically associated with devastating neurodegenerative diseases (e.g., Huntington's disease and spinocerebellar ataxia) (La Spada and Taylor, 2010; Orr and Zoghbi, 2007), and alanine repeat expansions are typically associated with severe developmental disorders (e.g., X-linked mental retardation, congenital ventral hypoventilation, and synpolydactyly) (Albrecht and Mundlos, 2005).

Virtually all investigations to date have focused on three features of proteins that contain disease-associated repeat expan-

sions, as cause of their pathology: their proclivity to form solid aggregates, alteration in their subcellular localization, and alteration of their proteolytic processing (Darling and Uversky, 2017; Orr and Zoghbi, 2007; Ross and Poirier, 2004). For example, many studies have established a correlation between aggregate formation by the mutant HTT protein and degeneration of specific neurons in Huntington's disease (Davies et al., 1997; Ross and Poirier, 2004; Zoghbi and Orr, 2000). However, several lines of evidence suggest that perturbed function of the soluble, appropriately localized fraction of the repeat-expanded protein may be responsible for its pathological effect, rather than aggregates (Ross, 2002; Saudou et al., 1998; Truant et al., 2008). Improved understanding of the function of the affected proteins and how the repeat expansions interfere with those functions would facilitate the development of therapeutics for this family of disease.

The majority (15/20) of disease-associated repeat expansions occur in nuclear proteins, most of which are sequence-specific

transcription factors (TFs) (Darling and Uversky, 2017). For example, expansions of an alanine repeat in the homeobox TF HOXD13 cause synpolydactyly, a hereditary limb malformation disorder (Muragaki et al., 1996), and expansions of a glutamine repeat in the highly conserved TF TATA-box binding protein (TBP) causes spinocerebellar ataxia type 17, a progressive neurodegenerative disease (Nakamura et al., 2001). New insights into how TFs interact with components of the transcription machinery to control gene expression programs would thus likely afford significant advance in our understanding of how repeat expansions in transcriptional regulators lead to pathology.

Control of gene transcription in eukaryotes involves the recruitment of RNA polymerase II (RNAPII) to genomic sites by sequence-specific TFs assisted by diverse transcriptional co-activators (Levine et al., 2014). Separation of liquids into a dense and dilute phase underlies the formation of several subcellular membraneless organelles (Alberti et al., 2019; Banani et al., 2017; Shin and Brangwynne, 2017), and recent evidence suggests that the assembly of the transcription machinery at genomic sites occurs through liquid-liquid phase separation, leading to the formation of transcriptional condensates (Boehning et al., 2018; Chong et al., 2018; Hnisz et al., 2017; Li et al., 2019; Sabari et al., 2018). For example, several TFs (e.g., FET family TFs, OCT4, and SP1), co-activators (e.g., Mediator and BRD4), and RNAPII contain intrinsically disordered regions (IDRs) that drive their phase separation, and these factors form discrete nuclear puncta in mammalian cells (Boehning et al., 2018; Cho et al., 2018; Chong et al., 2018; Kwon et al., 2013; Lu et al., 2018; Sabari et al., 2018). While TF- and co-activator-containing condensates are sensitive to short-chain aliphatic alcohols that dissolve various intracellular membraneless organelles (Boehning et al., 2018; Chong et al., 2018; Sabari et al., 2018), the functional importance of phase separation in transcriptional control has been unclear.

Here, we investigate the hypothesis that repeat expansions in IDRs of TFs alter the phase separation capacity of those TFs, and their capacity to form transcriptional condensates with components of the transcription machinery. We found that the IDR of the HOXD13 TF drives phase separation and co-condensation with the Mediator co-activator. Synpolydactyly-associated alanine repeat expansions facilitate homotypic HOXD13 phase separation, and the mutant HOXD13 IDRs co-condense less readily with Mediator than the wild-type (WT) HOXD13 IDR, a phenomenon we term “condensate unblending.” In a mouse synpolydactyly model, HOXD13 repeat expansion led to cell-type-specific transcriptional changes of HOXD13 targets in disease-relevant cells. We propose that disease-associated mutations in TF IDRs alter the features that drive the TF’s condensation behavior and present a molecular classification of TF IDRs as a framework to dissect the pathomechanism of diseases associated with TF dysfunction.

RESULTS

The HOXD13 IDR Drives Phase Separation

Disease-associated repeat expansions are significantly enriched in TFs ($p < 10^{-4}$, Fisher’s test; Figure 1A). To investigate

the hypothesis that disease-associated repeat expansions alter the phase separation capacity of the TFs in which they occur, we first focused on the homeobox TF HOXD13 as a proof of concept, because the genetics of *HOXD13* in disease is well characterized (Albrecht et al., 2004; Kuss et al., 2009; Muragaki et al., 1996; Villavicencio-Lorini et al., 2010). During mammalian embryogenesis, *Hoxd13* is expressed in the limb bud (Figure 1B) and controls skeletal morphogenesis (Villavicencio-Lorini et al., 2010). *HOXD13* mutations cause hereditary limb malformations (e.g., synpolydactyly), a syndrome characterized by extra digits and digit fusions (Muragaki et al., 1996). High-resolution confocal imaging of the HOXD13 protein in mouse limb bud cells revealed that HOXD13 forms discrete nuclear puncta (Figure 1B). HOXD13 puncta were detected using different HOXD13 antibodies (Figures S1A and S1B), and in several human cell lines that express HOXD13 (Figures S1B and S1C). Stochastic optical reconstruction microscopy (STORM) revealed that HOXD13 puncta in limb bud cells were ~ 100 nm in size and occurred in less DNA dense parts of the nucleus (Figure 1C), similar to previously described co-activator and RNAPII puncta (Cho et al., 2018). These results are consistent with the notion that HOXD13 may be a component of transcriptional condensates.

IDRs in proteins are known to drive phase separation (Banani et al., 2017; Shin and Brangwynne, 2017), and the HOXD13 N terminus has sequence features predictive of an IDR (Figure 1D). Therefore, we used an optogenetic platform to investigate whether the HOXD13 IDR can drive phase separation *in vivo*. In brief, the optoDroplet assay involves expression of a fusion protein consisting of the IDR of interest, mCherry, and the photolyase domain of the *Arabidopsis thaliana* CRY2 protein. Excitation of CRY2 with blue light stimulates its self-association, which leads to an increase of local concentration of the fused IDR (Shin et al., 2017). IDRs that drive phase separation subsequently facilitate the formation of liquid-like droplets (Figure 1E), which tend not to form in the absence of the IDR (Sabari et al., 2018; Shin et al., 2017). The HOXD13 IDR fused to mCherry and CRY2 facilitated the formation of droplets upon blue-light stimulation in HEK293T cells (Figures 1F, 1G, and S1D). As expected for phase-separated condensates, the extent of HOXD13 IDR droplet formation correlated with the expression level of the fusion protein (Figure S1E). Fluorescence recovery after photobleaching (FRAP) revealed rapid, liquid-like recovery rate of the HOXD13 IDR-mCherry-CRY2 droplets (Figures 1H and S1F), and the droplets were sometimes observed to undergo fusion (Figure 1I), which are characteristics of phase-separated condensates (Alberti et al., 2019). These results suggest that the HOXD13 IDR has the capacity to drive phase separation *in vivo*.

If the HOXD13 IDR can drive phase separation, the purified IDR should form liquid-like droplets *in vitro*. Purified recombinant HOXD13 IDR-mCherry fusion protein indeed formed droplets in the presence of 10% PEG-8000, while an mCherry control did not (Figure 1J). As expected for phase-separated condensates (Alberti et al., 2019), the HOXD13 IDR droplets had spherical shape, and their size scaled with the concentration of the protein (Figure 1J). These results indicate that the HOXD13 IDR can form condensates *in vitro*.

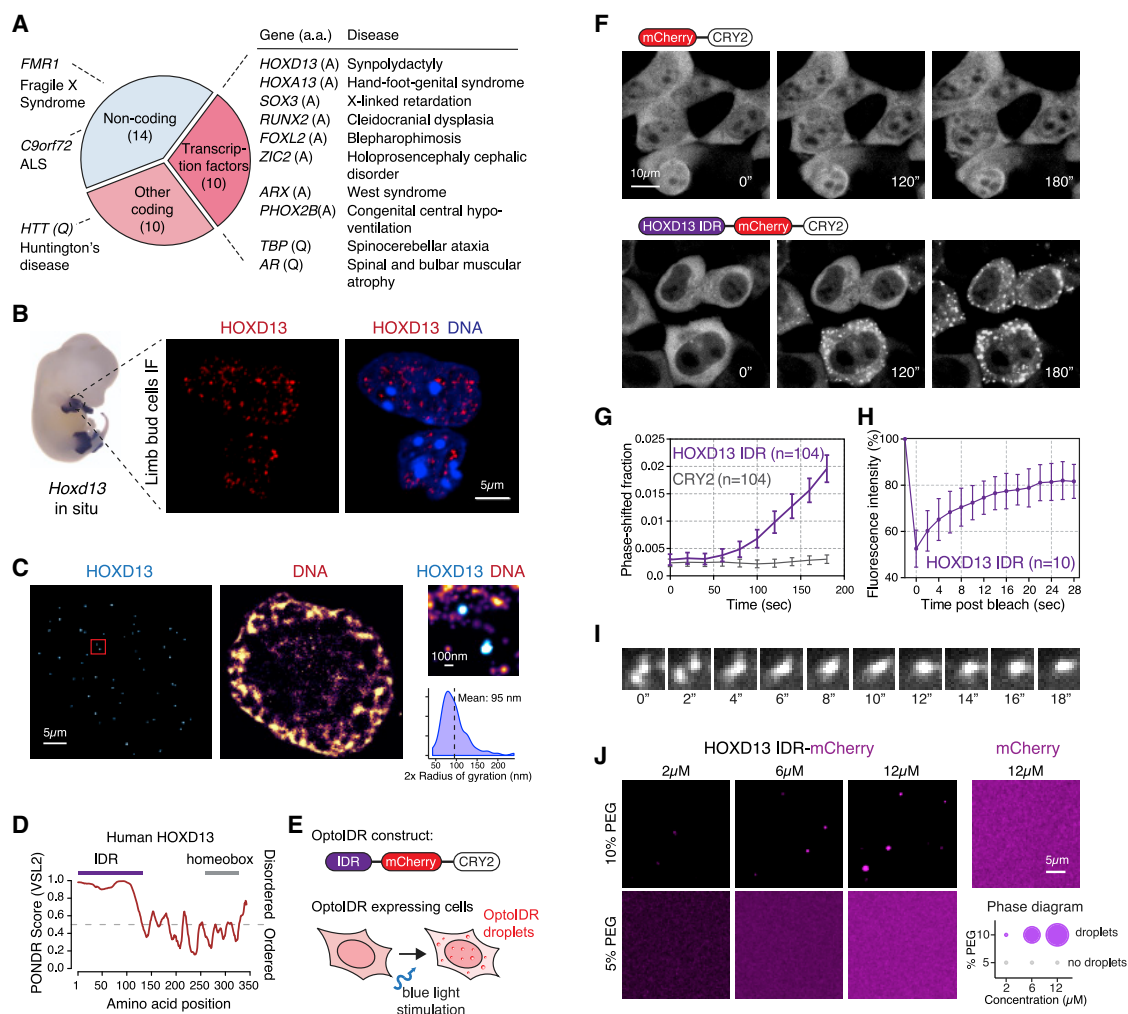


Figure 1. The HOXD13 IDR Drives Phase Separation

(A) Disease-associated repeat expansions in humans. (aa, amino acid)

(B) Left: *Hoxd13* whole mount *in situ* hybridization in an E12.5 mouse embryo. Right: HOXD13 Immunofluorescence (IF) in E12.5 mouse limb bud cells.

(C) Stochastic optical reconstruction microscopy (STORM) images of E12.5 mouse limb bud cells. The zoomed-in area on the right is highlighted with a red box on the left.

(D) Graph plotting intrinsic disorder for human HOXD13. The IDR cloned for subsequent experiments is highlighted with a purple bar.

(E) Scheme of the optoDroplet assay. The optoIDR construct consists of the HOXD13 IDR fused to mCherry and the *A. thaliana* CRY2 PHR domain.

(F) Representative images of live HEK293T cells expressing mCherry-CRY2 (top) and HOXD13 IDR-mCherry-CRY2 (bottom) fusion proteins. Cells were stimulated with 488 nm laser every 20 s for 3 min.

(G) Quantification of the fraction of the cytoplasmic area occupied by HOXD13 IDR-mCherry-CRY2 and mCherry-CRY2 droplets in HEK293T cells over time. Data displayed as mean \pm SEM.

(H) Fluorescence intensity of HOXD13 IDR-mCherry-CRY2 droplets before, during, and after photobleaching. Data displayed as mean \pm SD.

(I) Time-lapse images of a droplet fusion event in HEK293T cells expressing HOXD13 IDR- mCherry-CRY2 fusion protein.

(J) Left: representative images of droplet formation by purified HOXD13-mCherry and mCherry at the indicated concentrations. Right: phase diagram of HOXD13-mCherry in the presence of different concentrations of PEG-8000. The size of the circles is proportional to the size of droplets detected in the respective buffer conditions.

See also Figure S1.

Synpolydactyly-Associated Repeat Expansions Enhance HOXD13 IDR Phase Separation

Expansions of an alanine repeat in the IDR of HOXD13 cause synpolydactyly, a congenital limb malformation (Figures 2A and 2B) (Kuss et al., 2009; Muragaki et al., 1996). Of note, HOXD13 mutants that contain short (e.g., +7A) synpolydactyly-

associated expansions have not been described to form aggregates, suggesting that protein aggregation does not explain the pathology of short repeat expansions (Albrecht et al., 2004; Villavicencio-Lorini et al., 2010). Because the repeat expansions occur within the HOXD13 IDR, which promotes phase separation, we hypothesized that the repeat expansions may alter the

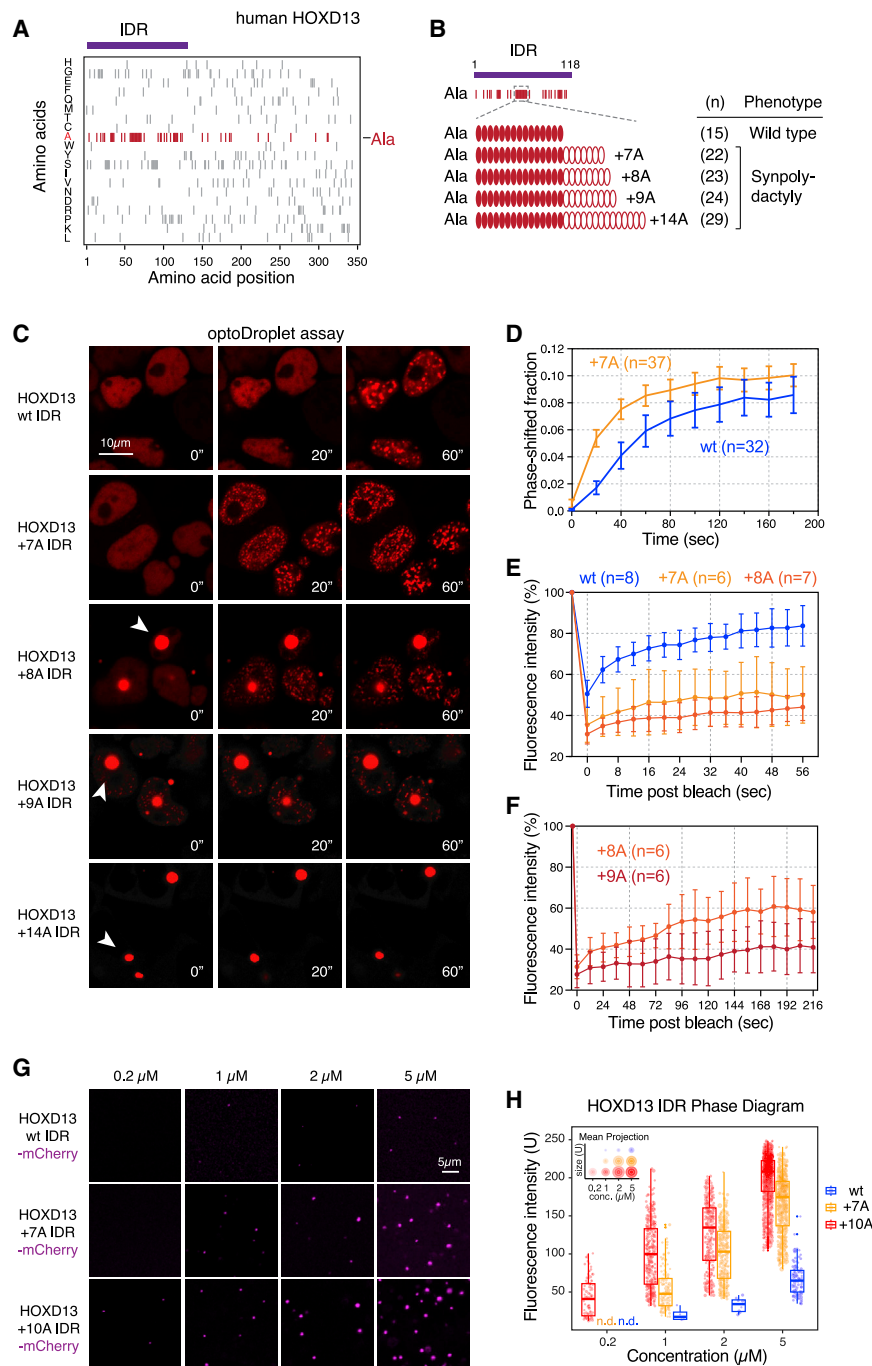


Figure 2. Synpolydactyly-Associated Repeat Expansions Enhance HOXD13 IDR Phase Separation

(A) Amino acid composition of human HOXD13. Ticks represent amino acids indicated on the y axis at the positions indicated on the x axis. The IDR cloned for subsequent experiments is highlighted with a purple bar.

(B) Alanines within the HOXD13 IDR sequence are indicated as red ticks. The central alanine repeat consists of 15As in the WT protein.

(C) Representative images of live HEK293T nuclei expressing WT and repeat-expanded HOXD13 IDR-mCherry-CRY2 fusion proteins. Cells were stimulated with 488 nm laser every 20 s for 3 min. Arrowheads highlight spontaneously forming IDR condensates present without 488 nm laser stimulation.

(D) Quantification of the fraction of the nuclear area occupied by HOXD13 WT IDR-mCherry-CRY2 and HOXD13 +7A IDR-mCherry-CRY2 droplets in HEK293T cells over time. Data displayed as mean \pm SEM.

(E) Fluorescence intensity of light-induced WT, +7A, and +8A HOXD13 IDR droplets before, during, and after photobleaching. Data displayed as mean \pm SD.

(F) Fluorescence intensity of +8A and +9A spontaneously formed HOXD13 IDR condensates before, during, and after photobleaching. Data displayed as mean \pm SD.

(G) Representative images of droplet formation by purified HOXD13 IDR-mCherry fusion proteins in droplet formation buffer.

(H) Phase diagram of HOXD13 IDR-mCherry fusion proteins. Every dot represents a detected droplet. Inset: projected average size of the droplets as mean \pm SD (middle circle, mean; inner and outer circle, SD). n.d.: not detected.

See also Figure S2.

phase separation capacity of the HOXD13 IDR. HOXD13 IDRs including the WT alanine repeat and several synpolydactyly-associated expansions (+7A, +8A, +9A, and +14A) were thus tested in the optoDroplet system. To ensure investigation of the phase separation capacity of the HOXD13 IDR in its nuclear context, an SV40 nuclear localization sequence (NLS) was included in all subsequent fusion constructs. Short expansions enhanced the rate of light-induced droplet formation in live cell nuclei (Figures 2C, 2D, and S2A). In addition, spontaneously

formed HOXD13 IDR condensates were observed in cells expressing the +8A, +9A, and +14A alleles, and the ratio of the protein in the spontaneous condensates and the soluble (diffuse) fraction correlated with the length of the repeat expansion (Figure 2C). DNA staining confirmed that the spontaneous condensates formed by the +8A, +9A alleles were nuclear, whereas the condensates formed by the +14A allele were cytosolic (Figure S2B). These results suggest that alanine repeat expansions enhance the phase-separation capacity of the HOXD13 IDR.

Phase-separated condensates deep beyond the phase boundary can transition to a gel-like state characterized by arrested, yet reversible dynamics (Shin et al., 2017). Consistent with this notion, the light-induced droplets formed by repeat-expanded HOXD13 IDRs exhibited considerably slower FRAP rates than droplets formed by the wild-type IDR (Figure 2E).

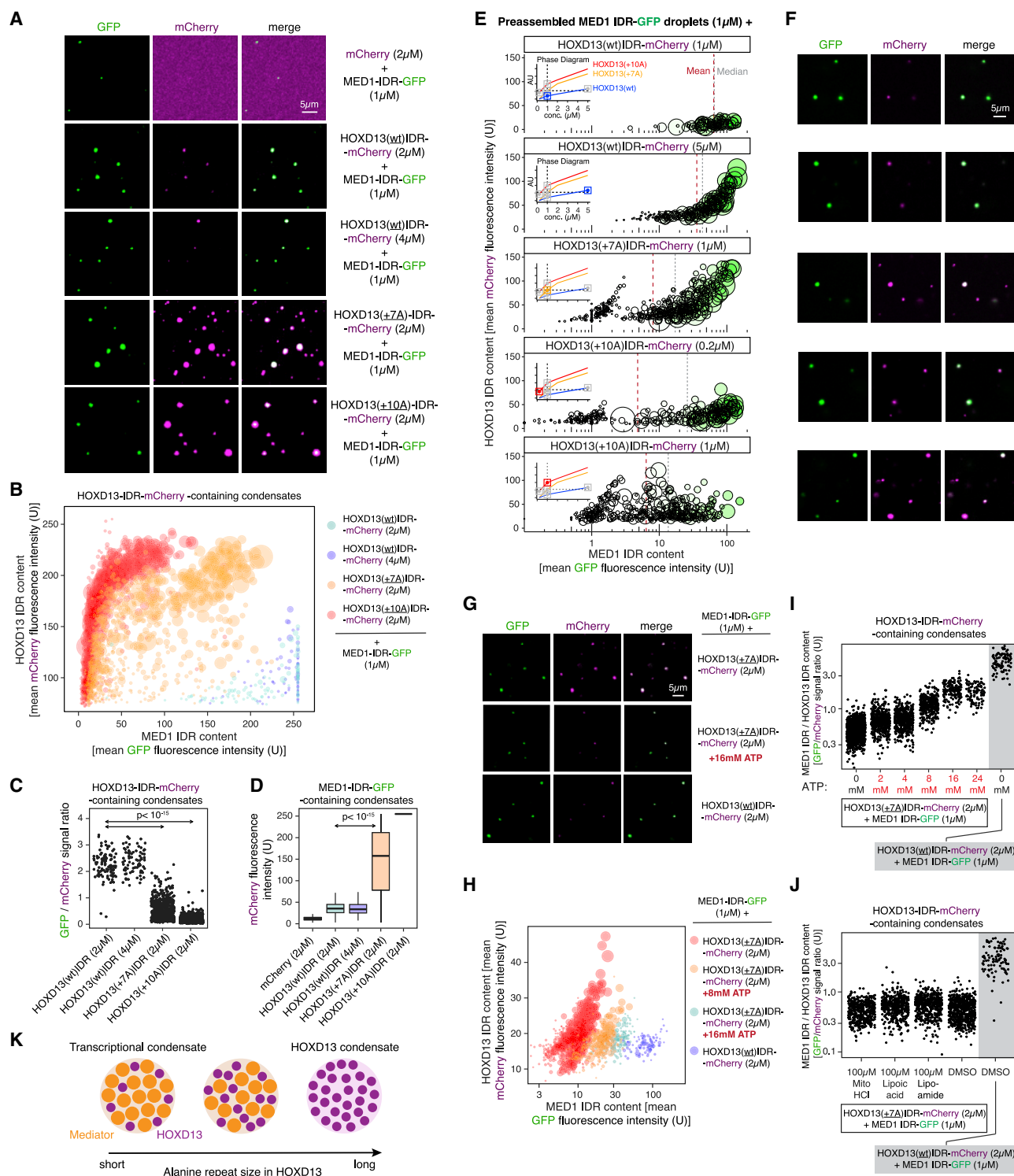


Figure 3. Synpolydactyly-Associated Repeat Expansions Alter the Composition of Hoxd13-Containing Condensates *In Vitro*

(A) Representative images of droplet formation by purified MED1-IDR-GFP and HOXD13 IDR-mCherry fusion proteins in droplet formation buffer. (B) Quantification of GFP and mCherry fluorescence intensity in HOXD13 IDR-mCherry containing droplets in the indicated MED1 IDR-GFP mixing experiments. Each dot represents one droplet, and the size of the dot is proportional to the size of the droplet. (C) Quantification of the ratio of GFP and mCherry fluorescence intensity in HOXD13 IDR-mCherry containing droplets in the indicated MED1 IDR-GFP mixing experiments. p value is from a Welch's t test.

(legend continued on next page)

Transient light stimulation revealed that formation of these droplets was reversible (Figure S2C). In addition, the spontaneous condensates formed by the +8A and +9A HOXD13 IDRs exhibited slow, but detectable FRAP rates (Figure 2F). Co-expression of an mCherry-tagged HOXD13 +8A IDR, which forms spontaneous condensates, with YFP-tagged HOXD13 WT and +7A IDRs confirmed that recruitment of IDR fusion proteins to IDR condensates is dependent on the IDR and correlates with the length of the alanine repeat (Figures S2D and S2E). These results suggest that the alanine repeat expansion alters the material properties of HOXD13 IDR optoDroplets.

To further probe the effect of alanine repeat expansion on HOXD13 phase separation, we purified various recombinant HOXD13 IDR-mCherry fusion proteins and investigated their phase separation capacity in droplet formation buffer *in vitro*. The +7A and +10A IDR mutants formed more, and more concentrated droplets compared to the WT IDR at similar concentrations (Figures 2G and 2H). Furthermore, the concentrations at which droplets appeared (i.e., the saturation concentration C_{sat}) were lower for +7A and +10A IDRs compared to the WT IDR (Figures 2G and 2H). Taken together, these results suggest that the alanine repeat expansion enhances phase separation of the HOXD13 IDR and are consistent with the previously described correlation between the length of the alanine repeat expansion and disease severity (Kuss et al., 2009; Muragaki et al., 1996).

Synpolydactyly-Associated Repeat Expansions Alter the Composition of HOXD13-Containing Condensates

Recent studies indicate that TFs can form heterotypic condensates with the IDR of the MED1 subunit of the Mediator co-activator *in vitro*, and co-condensation of TFs with the MED1 IDR requires the TFs' IDR (Boija et al., 2018). We thus hypothesized that the alanine repeat expansion of the HOXD13 IDR alters its ability to co-condense with transcriptional co-activators. To test this model, we purified recombinant MED1 IDR-GFP fusion protein and mixed it with various purified HOXD13 IDR-mCherry fusion proteins. Of note, the condensation behavior of the MED1 IDR is similar to that of purified partial Mediator complex, and the MED1 IDR has thus been a useful surrogate for Mediator condensates *in vitro* (Boija et al., 2018; Guo et al., 2019; Sabari et al., 2018). We found that heterotypic MED1 IDR-HOXD13 IDR co-condensates had dramatically altered composition when HOXD13 IDRs containing synpolydactyly-associated repeat expansions (+7A and +10A) were used in the mixing experiments. The MED1 IDR droplets incorporated the WT

HOXD13 IDR at several concentrations, and the WT HOXD13 IDR exclusively occurred in co-condensates with the MED1 IDR (Figure 3A). On the other hand, +7A and +10A HOXD13 IDRs were more enriched in MED1 IDR-containing droplets ($p < 10^{-15}$ Welch's t test), but the heterotypic droplets contained on average significantly less MED1 IDR than WT HOXD13 IDR co-condensates ($p < 10^{-15}$ Welch's t test) (Figures 3A–3D). These results suggest that the repeat-expanded HOXD13 IDR co-condenses with less MED1 IDR than the WT HOXD13 IDR, a phenomenon we term “condensate unblending.”

We next tested whether various HOXD13 IDRs unblend from the MED1 IDR even at concentrations at which they form comparable homotypic droplets alone. To this end, we preassembled MED1 IDR-GFP droplets for 30 min and mixed the preassembled MED1 IDR-GFP droplets with various concentrations of HOXD13 IDR-mCherry fusion proteins. These included concentrations where the HOXD13 proteins alone formed droplets similar in number, size, and protein content (WT, 5 μ M; +7A, 1 μ M; +10A, 0.2 μ M; see Figures 2H and 3E insets). The WT HOXD13 IDR at 1 μ M was incorporated in MED1 IDR droplets, and its enrichment was concentration-dependent (Figures 3E and 3F, top two panels). On the other hand, small condensates that consisted of almost exclusively HOXD13 +7A IDR were apparent in mixtures containing 1 μ M HOXD13 +7A IDR, and the median MED1 IDR content of the condensates was substantially lower relative to the MED1 IDR content of the HOXD13 WT IDR-MED1 IDR co-condensates (Figures 3E and 3F, compare first and third panels). A similar effect was observed when 0.2 μ M HOXD13 +10A IDR was used in the mixing experiment (Figures 3E and 3F). These results suggest that condensate unblending occurs at various concentrations *in vitro*.

The unblending of repeat-expanded HOXD13 IDRs from MED1 IDR-containing co-condensates indicates that the alanine repeat expansion increases the preference for homotypic HOXD13 IDR-IDR interactions over heterotypic interactions with the MED1 IDR. If this model is true, disrupting weak hydrophobic interactions could revert the composition of repeat-expanded HOXD13 IDR-containing MED1 co-condensates to be more similar to the composition of HOXD13 WT IDR-MED1 IDR co-condensates (i.e., “reblend” them). To test this notion, we treated HOXD13 IDR-MED1 IDR co-condensates with ATP, a small hydrophilic molecule known to solubilize hydrophobic molecules at mM concentrations (Patel et al., 2017). ATP reblended HOXD13 +7A IDR-MED1 IDR co-condensates in a dose-dependent manner (Figures 3G–3I). The reblending effect appeared specific to ATP, as lipoic acid, lipoamide, and

(D) Quantification of mCherry fluorescence intensity in MED1 IDR-GFP containing droplets in the indicated MED1 IDR-GFP mixing experiments. p value is from a Welch's t test.

(E) Quantification of GFP and mCherry fluorescence intensity in HOXD13 IDR-mCherry containing droplets. Each dot represents one droplet. The size of the dot is proportional to the size of the droplet, and the color of the dot is scaled to the MED1 signal in the droplet. Inset: a simplified phase diagram of HOXD13 IDRs based on the data displayed in Figure 2H. x axis is in \log_{10} scale.

(F) Representative images of the mixtures in (E).

(G) Representative images of droplets formed by purified MED1-IDR-GFP and HOXD13 IDR-mCherry fusion proteins.

(H) Quantification of GFP and mCherry fluorescence intensity in HOXD13 IDR-mCherry containing droplets in the indicated MED1 IDR-GFP mixing experiments. Each dot represents one droplet, and the size of the dot is proportional to the size of the droplet. x axis is in \log_{10} scale.

(I and J) Quantification of the ratio of GFP and mCherry fluorescence in HOXD13 IDR-mCherry containing droplets in the indicated mixing experiments. The y axes are in \log_{10} scale.

(K) Condensate unblending model of the impact of HOXD13 alanine repeat expansions.

mitoxantrone—compounds that were recently reported to dissolve condensates formed by stress granule proteins (Wheeler et al., 2019)—failed to reblend HOXD13 +7A IDR-MED1 IDR co-condensates (Figure 3J). Taken together, these results suggest that HOXD13 repeat expansion leads to unblending of HOXD13 IDR-MED1 IDR co-condensates (Figure 3K).

Altered Composition and Properties of Repeat-Expanded HOXD13-Condensates *In Vivo*

Next, we sought to investigate the effect of disease-associated repeat expansions on endogenous HOXD13-containing condensates *in vivo*. To this end, we isolated limb bud cells from homozygous *spdh* mouse embryos and WT littermates. *Spdh* mice carry +7A repeat-expanded *Hoxd13* alleles, and homozygous *spdh* mice exhibit synpolydactyly (Bruneau et al., 2001). We first tested whether transcriptional co-activators are associated with HOXD13-containing condensates in limb bud cells using antibody staining and STORM imaging. Because no high quality Mediator antibodies were available to co-stain with HOXD13 in STORM, BRD4 was visualized instead of Mediator. BRD4 is a ubiquitous co-activator that co-purifies with Mediator (Jiang et al., 1998), co-condenses with Mediator (Sabari et al., 2018), and BRD4 chemical inhibition dissolves Mediator condensates *in vivo* (Cho et al., 2018). We found that BRD4 more frequently overlapped HOXD13 puncta in WT limb bud cells than in *spdh* limb bud cells harboring HOXD13 +7A alleles (1.5-fold difference in Manders coefficient, p value < 0.001, two-tailed t test) (Figures 4A and 4B). As a control, HP1 α was co-visualized with HOXD13, and the two proteins showed negligible overlap (Figure 4B). These results suggest that HOXD13-condensates have altered composition *in vivo*.

To investigate the biophysical properties of HOXD13-containing condensates *in vivo*, WT and *spdh* limb bud cells were treated with 1,6-hexanediol (1,6-HD), a short-chain aliphatic alcohol that dissolves various intracellular membraneless organelles (Boehning et al., 2018; Chong et al., 2018; Sabari et al., 2018). HOXD13 was subsequently visualized with immunofluorescence. Both WT and +7A HOXD13 localized within discrete nuclear puncta (Figure 4C), but the puncta detected in limb bud cells expressing +7A HOXD13 displayed considerably reduced sensitivity to 1,6-HD (Figures 4C, 4D, S3A, and S3B). These results suggest that a short alanine repeat expansion (+7A) alters the biophysical properties of HOXD13-containing puncta in limb bud cells.

To test whether repeat-expansion impairs the ability of HOXD13 condensates to recruit Mediator *in vivo*, various HOXD13 IDRs were tethered to a LacO array in U2OS cells expressing an ectopic MED1 IDR-YFP fusion protein (Figure 4E) (Janicki et al., 2004). MED1 IDR-YFP was found mildly enriched at the LacO array occupied by the HOXD13 WT IDR tether, and its incorporation was significantly reduced in the +7A and +10A HOXD13 IDR tethers ($p < 10^{-2}$, Welch's t test) (Figures 4E and 4F). Consistent with the notion that HOXD13 repeat expansion impairs Mediator recruitment to HOXD13 condensates, spontaneously forming HOXD13 +8A IDR condensates observed in HEK293T cells excluded the MED1 IDR (Figures S2D and S2E).

The Mediator co-activator plays key roles in recruiting RNAPII to TF-bound genes (Levine et al., 2014), so a reduction of MED1 content in mutant HOXD13-containing condensates would be expected to reduce the transcriptional activity of repeat-

expanded HOXD13. Consistent with this idea, repeat-expanded HOXD13 alleles displayed reduced activity in luciferase reporter assays (Figure 4G).

To rule out that the repeat expansion affects DNA binding of HOXD13, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) on FLAG-tagged HOXD13 proteins expressed in a chicken transgenic cell system (Ibrahim et al., 2013). The genome-wide binding of +7A HOXD13 was virtually indistinguishable from the binding of WT HOXD13 (Figures S3C–S3E). As a control, a HOXD13 allele containing a Q317R mutation was included in the analyses. The Q317R mutation occurs in the DNA-binding domain in HOXD13 and substantially alters HOXD13 binding genome-wide (Figures S3C–S3E) (Ibrahim et al., 2013).

Taken together, these results suggest that repeat expansions alter the composition and biophysical properties of HOXD13-containing condensates in disease-relevant limb bud cells and transgenic cell systems and reduce HOXD13-dependent transcriptional activity without affecting DNA binding.

HOXD13 Repeat Expansion Alters the Transcriptional Program of Several Cell Types in a Cell-Specific Manner

The condensate unblending model predicts that repeat expansions alter the composition and properties of HOXD13-containing transcriptional condensates, which leads to deregulated gene expression programs. To comprehensively assess the impact of HOXD13 repeat expansion on gene expression in disease-relevant cells, we performed single-cell RNA-seq (scRNA-seq) on limb buds from mouse embryos that carry either WT or +7A repeat-expanded *Hoxd13* alleles (*spdh* mice). In total, 9,655 single-cell transcriptomes were captured, with on average >60,000 transcripts and ~3,500 genes detected per cell. The single cell transcriptomes of 4,464 WT cells were clustered to generate a reference map of cell states in the limb bud (Figures 5A–5C and S4A–S4C), and 4,147 *spdh* limb bud cells were assigned to the 11 WT cell states (Figures 5A–5C and S4A–S4C). These analyses revealed that +7A HOXD13 affected the abundance primarily of two cell types: interdigital mesenchymal cells were depleted, and proximal chondrocytes were enriched in the *spdh* limb bud (Figures 5C and S5A–S5C), consistent with previous observations (Kuss et al., 2009; Villavicencio-Lorini et al., 2010).

Examination of dysregulated genes within individual cell states revealed cell-type-specific changes in the transcriptional program of several cell types in the *spdh* limb. For example, genes associated with mesenchyme differentiation (e.g., *Msx1*, *Msx2*, *Tgfb2*) and digit morphogenesis (e.g., *Hoxd12*, *Hoxd13*) were downregulated in the *spdh* interdigital mesenchymal cells (cluster 4; whose abundance was lower in the *spdh* limbs) (Figures 5D and 5E). On the other hand, the transcriptome profile of proximal chondrocytes (cluster 6), did not substantially change (Figure 5D). HOXD13 ChIP-seq in WT limbs confirmed that the topologically associating domains (TADs) containing genes dysregulated in the interdigital mesenchyme cells were significantly enriched for HOXD13 binding ($p < 10^{-5}$ Wilcoxon test) (Figures 5F–5H and S5D–S5G). Capture-C (chromosome conformation) data in WT limb bud cells confirmed interactions between the HOXD13-bound elements and those genes (Figures 5F, 5H, and S5D). We also performed ChIP with reference exogenous genome (ChIP-Rx) for the transcription-associated H3K27Ac modification in

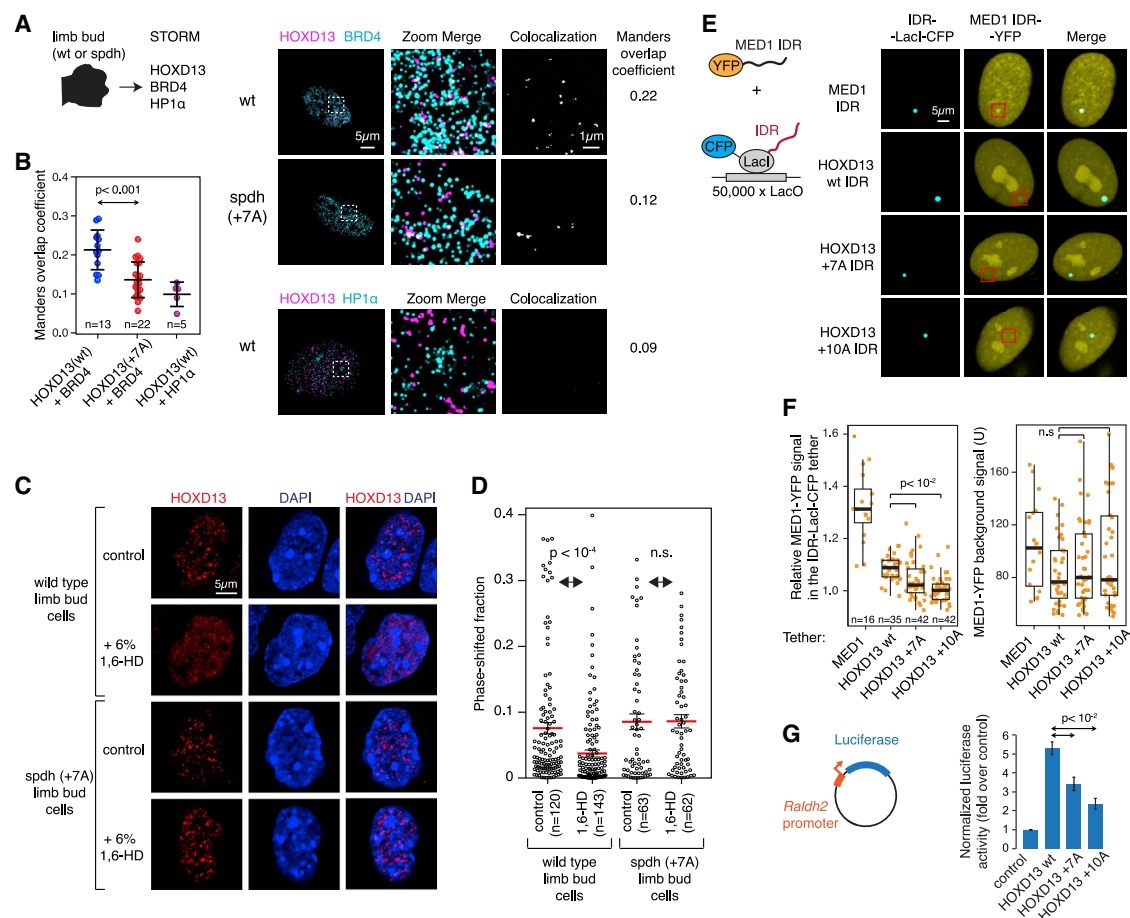


Figure 4. Altered Composition and Properties of Repeat-Expanded HOXD13-Condensates In Vivo

(A) Left: experimental scheme. Right: stochastic optical reconstruction microscopy (STORM) images of WT and spdh E12.5 mouse limb bud cells. Nuclei were cropped in images. The zoomed-in area on the right is highlighted with a white box in the middle.

(B) Manders overlap coefficients (MOC) of the STORM co-localizations. Each dot represents the MOC calculated from localization data from a single nucleus. p value is from a Student's t test.

(C) Left: experimental scheme. Right: representative images of WT and spdh E12.5 mouse limb bud cells with or without treatment with 6% 1,6-hexanediol for 1 min.

(D) Quantification of the fraction of the nuclear area occupied by HOXD13 puncta in mouse limb bud cells (displayed in C) with or without treatment with 6% 1,6-hexanediol for 1 min. p values are from a Student's t-test.

(E) Fluorescence images of ectopically expressed MED1 IDR-YFP in U2OS cells co-transfected with the indicated HOXD13 IDR-LacI-CFP fusion constructs.

(F) Quantification of the relative MED1 IDR-YFP signal intensity in the tether foci. p values are from a Welch's t test.

(G) Luciferase reporter assays of HOXD13 WT, +7A, and +10A mutants co-expressed with a *Raldh2*-luciferase reporter construct. p value is from a Student's t test.

See also Figure S3.

WT and spdh limb buds, and found a significant reduction of H3K27Ac at the nearest HOXD13-bound sites around the genes downregulated in interdigital mesenchymal cells ($p = 0.03$, Wilcoxon test) (Figures 5I and 5J). These results suggest that HOXD13 repeat expansion leads to reduced transcription of key differentiation HOXD13-target genes in interdigital mesenchymal cells associated with the synpolydactyly phenotype.

Disease-Associated Repeat Expansions Alter the Phase Separation Capacity of Other TFs

Amino acid repeat expansions in TFs occur in various diseases (Figure 1A). We therefore investigated whether the phase separation capacity of other TFs may be altered by disease-associated repeat expansions.

HOXA13 is a homeobox TF involved in vertebrate limb and urogenital tract development (Albrecht and Mundlos, 2005), and alanine repeat expansions in the HOXA13 IDR cause hand-foot genital syndrome (HFGS) (Figure 6A) (Goodman et al., 2000).

The HOXA13 IDR facilitated phase separation in the optoDroplet system (Figures 6B, 6C, and S6A–S6C), and HOXA13 IDR droplets exhibited liquid-like FRAP rate (Figure 6D). The HOXA13 IDR containing a short (+7A) HFGS-linked expansion tended to form aggregates with negligible FRAP rate (Figures 6B and 6D). However, a portion of the repeat-expanded

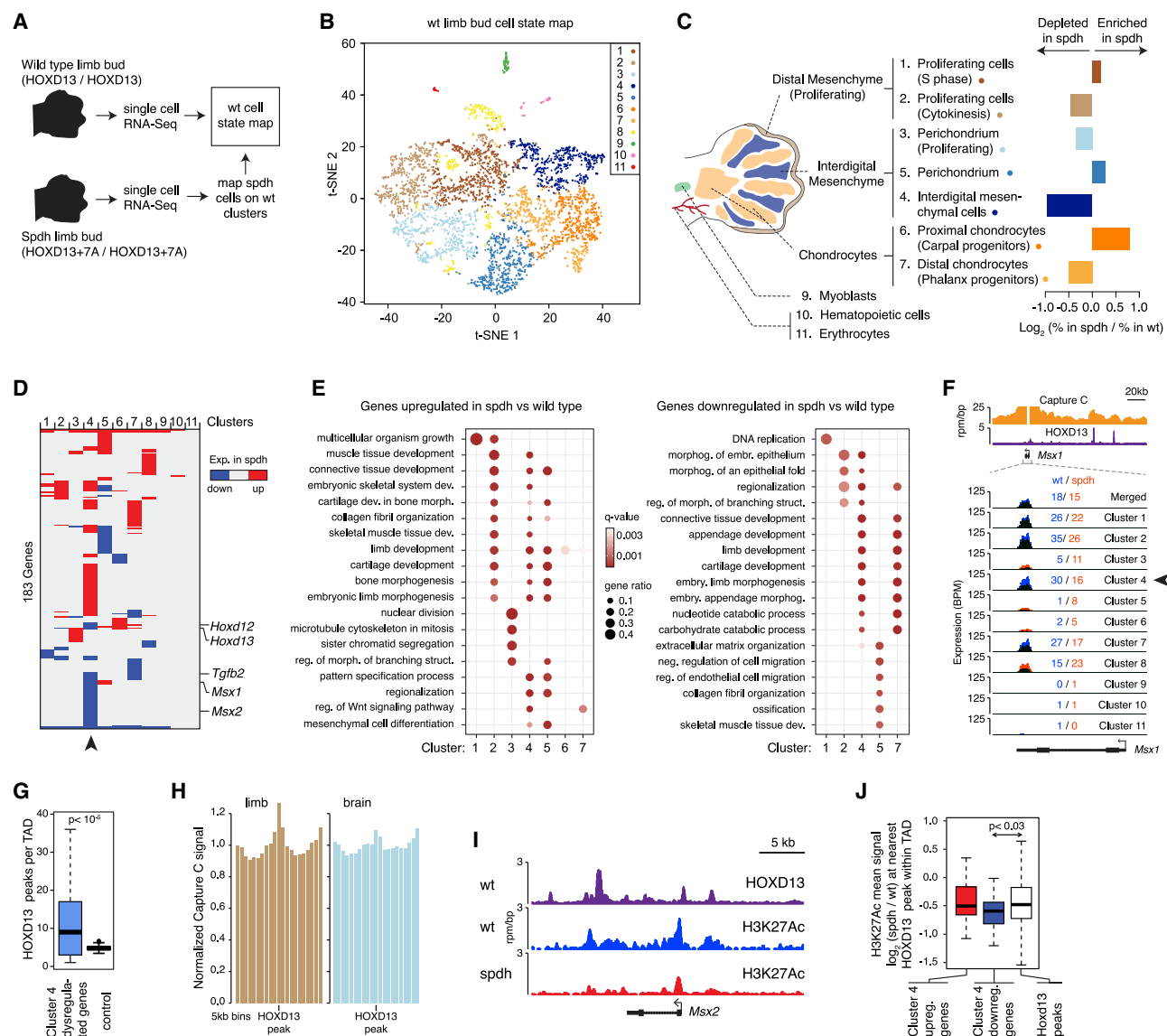
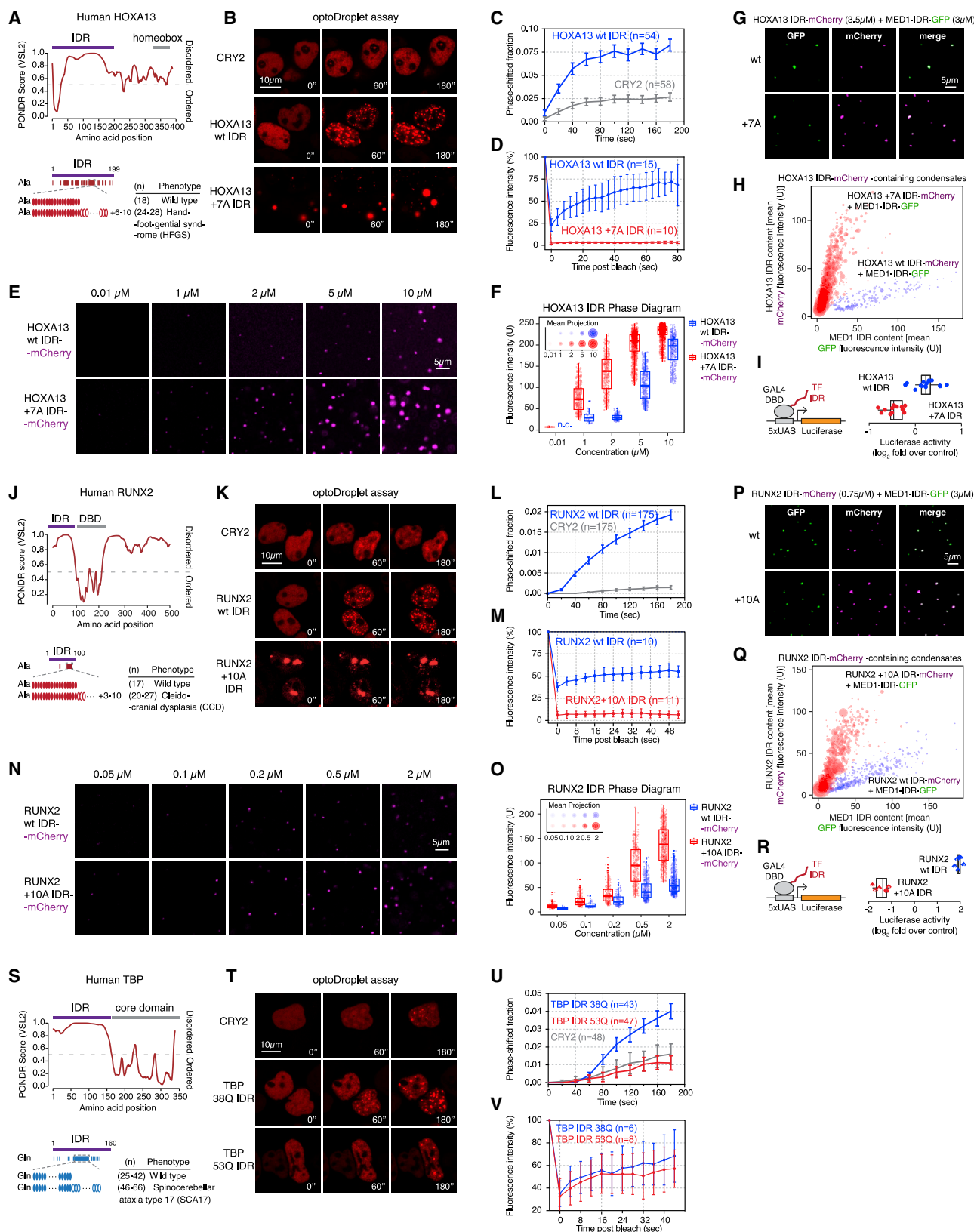


Figure 5. HOXD13 Repeat Expansion Alters the Transcriptional Program of Several Cell Types in a Cell-Specific Manner

(A) Scheme of the scRNA-seq experiment strategy.
 (B) Visualization of the WT scRNA-seq data using t-distributed stochastic neighbor embedding (t-SNE).
 (C) Changes in cell type composition in spdh limb buds. Displayed are the relative changes in the proportions of cells that belong to the designated clusters (i.e., cell states) between WT and spdh E12.5 limb buds.
 (D) Heatmap of differentially up- or downregulated genes in the spdh limb bud relative to WT within the 11 cell clusters. Arrowhead highlights the interdigital mesenchymal cells.
 (E) Gene Ontology (GO) term enrichment analysis of differentially up- or downregulated genes in the spdh limb bud relative to WT within individual cell clusters.
 (F) Profiles of Capture C, HOXD13 ChIP-seq, and scRNA-seq data at the *Msx1* locus. The mean expression value in spdh (red) and WT cells (blue) within each cluster are also displayed. Arrowhead highlights the expression level in interdigital mesenchymal cells, where the expression difference is the most profound.
 (G) Number of HOXD13 peaks in topologically associating domains (TADs) that contain a gene dysregulated in cluster 4. p value is from a Wilcoxon rank sum test.
 (H) Mean Capture C signal around HOXD13 peaks within topologically associating domains (TADs) that contain a gene dysregulated in cluster 4.
 (I) ChIP-Seq binding profiles around the *Msx2* locus.
 (J) Quantification of the mean H3K27Ac signal at the nearest HOXD13 binding sites around the indicated genes within the same TAD. p value is from a Wilcoxon rank-sum test.

See also [Figures S4](#) and [S5](#).



(legend on next page)

HOXA13 IDR remained soluble within nuclei, and formed light-induced droplets at concentrations where the HOXA13 WT IDR did not (Figure S6D). Furthermore, the +7A expansion enhanced droplet formation of mCherry-tagged, purified HOXA13 IDR *in vitro* (Figures 6E and 6F), and lead to unblending of HOXA13 IDR from MED1-IDR co-condensates (Figures 6G and 6H). Last, the +7A expansion significantly reduced transcriptional activity of the HOXA13 IDR fused to a GAL4 DNA-binding domain (DBD) in a luciferase reporter system ($p < 10^{-3}$, two-tailed t test) (Figure 6I). These results suggest that the HOXA13 IDR can drive phase separation, and a pathological alanine repeat expansion alters its phase separation capacity, co-condensation with the MED1 IDR, and transcriptional activity.

RUNX2 is a RUNT family TF that controls bone morphogenesis, and expansions of a short alanine and glutamine repeat in the RUNX2 IDR are associated with cleidocranial dysplasia (CCD), a disorder of severe skeletal defects (Figure 6J) (Mastushita et al., 2015; Shibata et al., 2016). The RUNX2 IDR facilitated phase separation in the optoDroplet system (Figures 6K, 6L, and S6E–S6G), and RUNX2 IDR droplets exhibited liquid-like FRAP rate (Figure 6M). The RUNX2 IDR containing a CCD-associated alanine expansion a (+10A) tended to form solid aggregates with negligible FRAP rate (Figures 6K and 6M). However, a portion of the repeat-expanded RUNX2 IDR remained soluble within nuclei and formed light-induced droplets at concentrations where the RUNX2 WT IDR did not (Figure S6H). Furthermore, the +10A expansion enhanced droplet formation of mCherry-tagged, purified RUNX2 IDR *in vitro* (Figures 6N and 6O) and led to unblending of RUNX2 IDR from MED1-IDR co-condensates (Figures 6P and 6Q). Last, the +10A expansion significantly reduced transcriptional activity of the RUNX2 IDR fused to a GAL4 DBD in a luciferase reporter system ($p < 10^{-3}$, two-tailed t test) (Figure 6R). These results suggest that the RUNX2 IDR can drive phase separation, and a pathological alanine repeat expansion alters its phase separation capacity, co-condensation with the MED1 IDR, and transcriptional activity.

The TATA-box binding factor TBP is a highly conserved general TF that plays a key role in transcription initiation. The TBP

N terminus contains a polymorphic polyglutamine (Q) repeat, whose typical size ranges between 25–42 glutamines, and repeats consisting of >46 glutamines are linked to spinocerebellar ataxia type 17 (SCA17), a progressive neurodegenerative disease (Figure 6S) (Nakamura et al., 2001). Fixed cell immunofluorescence revealed that TBP forms puncta in murine cells (Figure S6I). Purified recombinant TBP IDR-mCherry fusion protein formed spherical droplets whose size scaled with the concentration of the protein (Figures S6J and S6K). Furthermore, a TBP IDR with WT glutamine repeat length (38Q) facilitated phase separation over the CRY2 control in the optoDroplet system (Figures 6T, 6U, and S6L–S6N), and TBP IDR droplets exhibited liquid-like FRAP rate (Figure 6V). In contrast, an SCA17-associated polyglutamine repeat expansion (53Q) inhibited TBP IDR droplet formation (Figures 6T, 6U, and S6M). These results suggest that the TBP IDR can drive phase separation, and its phase separation capacity is inhibited by pathological glutamine repeat expansion.

A Catalog of Human TF IDRs

TFs typically consist of highly structured DBDs and intrinsically disordered activation domains (ADs/IDRs) (Lambert et al., 2018; Mitchell and Tjian, 1989; Staby et al., 2017). TF IDRs are low complexity protein sequences, several of which are known to drive phase separation (Boija et al., 2018; Chong et al., 2018). TF IDRs frequently contain homopolymeric repeats (proline, serine, alanine, glycine, glutamine, and histidine), but only expansions of alanine and glutamine repeats have been linked to human pathologies to date. We thus reasoned that various molecular features of TF IDRs dictate their phase separation capacity. To gain insights into those features, we created a catalog of IDRs in ~1,500 human TFs. We first identified IDRs and DBDs in human TFs and clustered them based on amino acid composition, hydrophobicity, aliphatic index, stability, isoelectric point, and disorder. The clustering algorithm separated DBDs and IDRs with 96% efficiency (Figure S7A) and could even separate DBD families previously annotated based on structural homology (Figure S7B) (Lambert et al., 2018). The TF IDRs were

Figure 6. Disease-Associated Repeat Expansions Alter the Phase Separation Capacity of Other TF IDRs

(A, J, and S) Graphs plotting intrinsic disorder for (A) HOXA13, (J) RUNX2, and (S) TBP. The IDRs cloned for subsequent experiments are highlighted with a purple bar.
(B, K, and T) Representative images of HEK293T nuclei expressing the indicated TF IDR-mCherry-CRY2 fusion proteins. Cells were stimulated with 488 nm laser every 20 s for 3 min.
(C, L, and U) Quantification of the fraction of the nuclear area occupied by droplets of the indicated TF IDR-mCherry-CRY2 fusion proteins in HEK293T nuclei over time. Data displayed as mean \pm SEM.
(D, M, and V) Fluorescence intensity of droplets of the indicated TF IDR-mCherry-CRY2 fusion proteins before, during, and after photobleaching. For the HOXA13 +7A IDR and the RUNX2 +10A IDR, the spontaneously formed droplets were bleached, for all other fusion proteins the light-induced droplets were bleached. Data displayed as mean \pm SD.
(E and N) Representative images of droplet formation by purified TF IDR-mCherry fusion proteins in droplet formation buffer.
(F and O) Phase diagram of TF IDR-mCherry fusion proteins. Every dot represents a detected droplet. Inset: projected average size of the droplets as mean \pm SD (middle circle, mean; inner and outer circle, SD). n.d., not detected
(G and P) Representative images of droplet formation by purified MED1-IDR-GFP and TF IDR-mCherry fusion proteins in droplet formation buffer with 10% PEG-8000.
(H and Q) Quantification of GFP and mCherry fluorescence intensity in TF IDR-mCherry containing droplets in the indicated MED1 IDR-GFP mixing experiments. Each dot represents one droplet, and the size of the dot is proportional to the size of the droplet.
(I and R) Left: GAL4 activation assay schematic. The luciferase reporter plasmid and the expression vector for the GAL4 DBD-TF IDR fusion proteins were transfected into HEK293T cells. Right: Luciferase reporter activity of the indicated TF IDRs fused to GAL4-DBD. $p < 10^{-3}$ for both WT/mutant comparisons (Student's t test).
See also Figure S6.

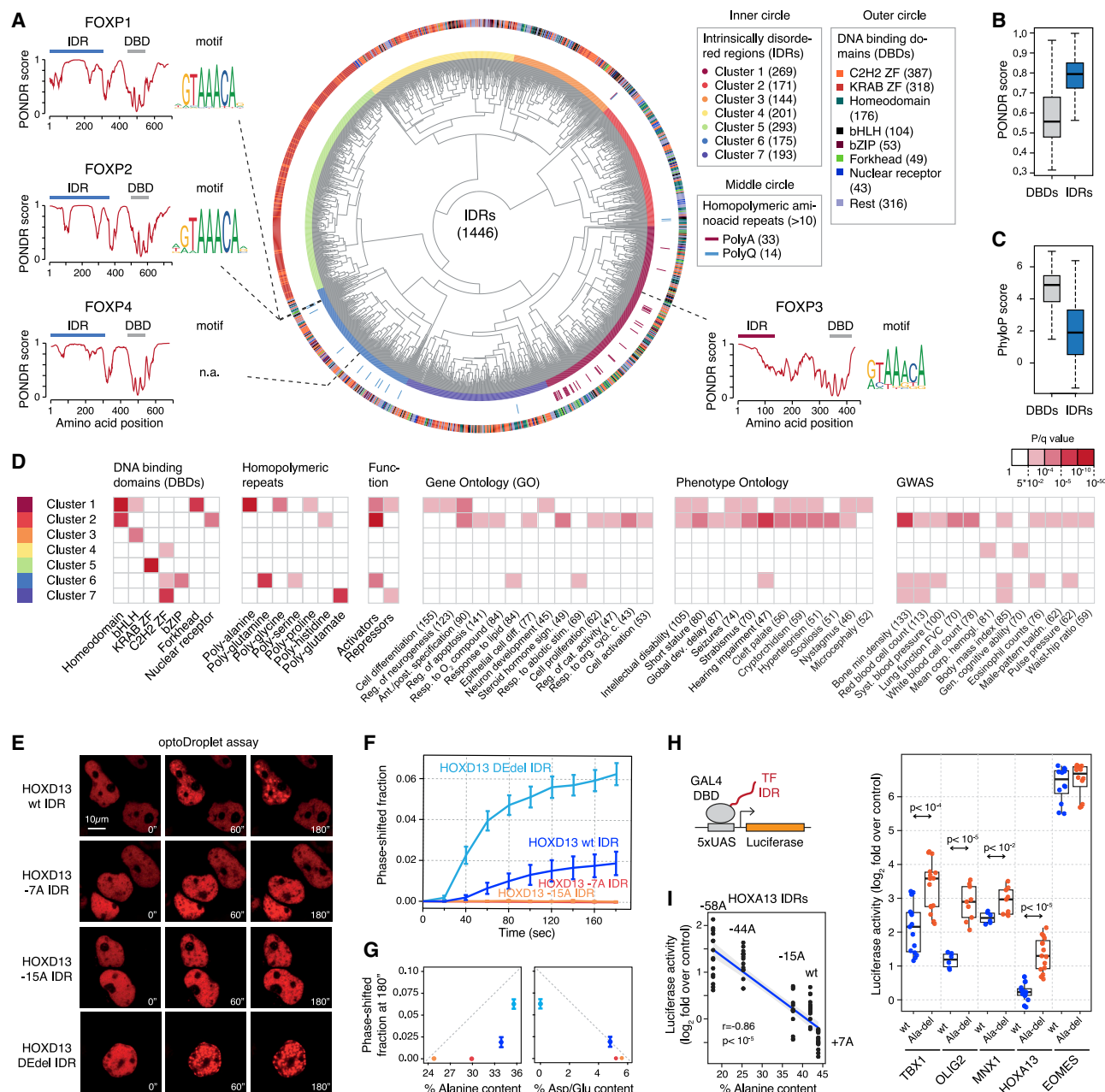


Figure 7. A Catalog of Human TF IDRs

(A) Classification of human TF IDRs. The inner circle depicts the clusters of TF IDRs. The outer circle includes the annotation of the DBDs of the TFs whose IDRs were classified in the inner circle.

(B) Boxplot of POND scores (disorder) of human TF DBDs and IDRs.

(C) Boxplot of phyloP scores (conservation) of human TF DBDs and IDRs

(D) Enrichment of TFs whose IDRs belong to the seven IDR clusters for the indicated sequence features, functional and phenotypic categories. Red box highlights significant enrichment ($q < 0.05$).

(E) Representative images of HEK293T cells expressing the indicated HOXD13 IDR-mCherry-CRY2 fusion proteins. Cells were stimulated with 488 nm laser every 20 s for 3 min.

(F) Quantification of the fraction of the nuclear area occupied by HOXD13 IDR-mCherry-CRY2 droplets in HEK293T cells over time. Data displayed as mean \pm SEM.

(G) Plot of the nuclear area occupied by HOXD13 IDR-mCherry-CRY2 droplets versus the alanine content and Asp/Glu content of the IDR constructs. Data displayed as mean \pm SEM.

(legend continued on next page)

broadly distributed in seven major clusters based on various features (Figures 7A and S7C–S7E), and as expected, were on average more disordered and less conserved than DBDs (Figure 7B–C). The IDRs in individual clusters belong to TFs from diverse DBD families, and TFs that have similar DBDs contain IDRs that belong to diverse IDR clusters. For example, 3/4 FOXP family members have a cluster 6 IDR, whereas FOXP3 has a cluster 1 IDR, and the DNA-binding specificity of these TFs is virtually identical (Figure 7A). A notable, but expected, exception was IDR cluster 5, which predominantly consisted of KRAB-zinc finger TFs (Figure 7A). Overall, TFs in diverse IDR clusters were enriched for various homopolymeric repeats and were associated with a spectrum of biological processes, human phenotypes, and disease-associated genetic variants (Figures 7A and 7D).

One key feature of cluster 1 IDRs was high alanine content (Figure S7E), and this cluster included 32 of the 33 poly-alanine-containing IDRs ($p < 10^{-16}$, Fisher's exact test) (Figure 7A). We thus hypothesized that the hydrophobic alanine residues may drive the phase separation capacity of cluster 1 IDRs. To test this model, we engineered various repeat-deletion mutant (–7A and –15A) HOXD13 IDRs, an IDR in which the negatively charged residues were deleted (DEdel), and tested their phase separation capacity in the optoDroplet system (Figure 7E). Deletion of 7 and 15 alanines, which decreased hydrophobicity, inhibited droplet formation, while deletion of the negatively charged residues enhanced droplet formation by the HOXD13 IDR (Figures 7E–7G, S7F, and S7G). We then selected 10 cluster 1 IDRs and tested the contribution of the poly-alanines to transcriptional activity in the GAL4 DBD-luciferase system. 5/10 IDRs functioned as activators in this minimal system, and for 4/5 of the activators, deletion of alanine repeats enhanced reporter activity (Figures 7H and S7H). Analysis of further alanine mutants revealed that the reporter activity of HOXA13 IDR negatively correlated with the number of alanines (Figure 7I). These results suggest that the alanine repeat drives homotypic phase separation of cluster 1 IDRs, and that poly-alanine-driven TF phase separation can inhibit transactivation.

DISCUSSION

The results presented here support a model that disease-associated amino acid repeat expansions in TF IDRs alter the TFs' phase separation capacity and their ability to co-condense with transcriptional coactivators. For the HOXD13 TF, synpolydactyly-associated alanine repeat expansions enhanced the phase separation capacity of the HOXD13 IDR, and the mutant HOXD13 IDR was able to co-condense with dramatically less Mediator than the WT IDR, a phenomenon we term “condensate unblending” (Figure 3K). In a mouse model of synpolydactyly, HOXD13 repeat expansion altered the co-activator content and biophysical properties of HOXD13-containing condensates

and led to cell-type-specific transcriptional changes of HOXD13 target genes in disease-relevant cells. For HOXA13 and RUNX2, disease-associated alanine repeat expansions similarly enhanced homotypic phase separation and led to unblending from Mediator, while for TBP, a disease-associated glutamine repeat expansion reduced the phase separation capacity of its IDR. We propose that aberrant phase separation of TFs and unblending of transcriptional condensates may underlie human pathologies associated with mutations in the IDRs of transcriptional regulators.

Several lines of evidence indicate that altered phase separation underpins the effect of HOXD13 repeat expansions. The saturation concentration C_{sat} (at which condensates are observed) is lower for purified mutant HOXD13 IDRs than the WT IDR (Figures 2G and 2H), and the mutations enhance condensate formation in the optoDroplet system (Figures 2C and 2D). The lower C_{sat} of mutant HOXD13 IDRs is in turn associated with an increase in TF IDR content and reduced MED1 IDR content of heterotypic co-condensates *in vitro* (Figures 3A–3F), and reduced co-activator-HOXD13 association *in vivo* (Figures 4A and 4B). This effect is consistent with recent reports that heterotypic interactions dominate phase separation of endogenous condensates, and C_{sat} of heterotypic condensates can be modulated by physico-chemical properties of their components (Choi et al., 2019; Riback et al., 2019). Furthermore, the condensate unblending model may help explain why the (+7A) repeat expanded *Hoxd13* allele is genetically a dominant negative allele (Albrecht and Mundlos, 2005; Villavicencio-Lorini et al., 2010), why the phenotype of repeat expansions is distinct from the phenotype of HOXD13 deactivating mutations (Bruneau et al., 2001; Dollé et al., 1993), and why the length of the repeat expansion correlates with disease severity (Goodman et al., 1997).

Repeat expansion diseases include severe, incurable neurodegenerative and developmental diseases typically associated with the presence of large protein aggregates (Albrecht and Mundlos, 2005; Orr and Zoghbi, 2007; Zoghbi and Orr, 2000). The condensate model presented here may explain several observations of the pathology of repeat expansion diseases that are less readily explained by toxicity of aggregates. For example, for poly-alanine expansions that occur in TFs, aggregates, or nuclear inclusions were only described in overexpression systems but not in primary tissue to date (Albrecht et al., 2004; Villavicencio-Lorini et al., 2010). Furthermore, the short +7A repeat expansion of HOXD13 recapitulates the human synpolydactyly phenotype in mice, yet without any evidence of protein aggregation in the limb bud (Albrecht et al., 2004; Villavicencio-Lorini et al., 2010). For polyglutamine expansion diseases (e.g., Huntington's disease), previous studies have established a correlation between aggregate formation by the mutant huntingtin (HTT) protein and degeneration of specific neurons (Davies et al., 1997; Ross and Poirier, 2004). However, manipulation of the ubiquitin-proteasome pathway was reported to decrease aggregate

(H) Left: GAL4 activation assay schematic. The luciferase reporter plasmid and the expression vector for the GAL4 DBD-TF IDR fusion proteins were transfected into HEK293T cells. Right: Luciferase reporter activity of the indicated TF IDRs fused to GAL4-DBD. p values are from a Welch's t test.

(I) Normalized luciferase activity of the indicated HOXA13 IDRs fused to GAL4 DBD. The blue line is a linear regression line, and the gray zones denote the 95% confidence interval. r denotes a Pearson's correlation coefficient. p value is from a t test.

See also Figure S7.

formation without decreasing toxicity of the repeat-expanded HTT protein in primary cell models, suggesting that the soluble fraction of the protein may be responsible for its cytotoxic effects (Saudou et al., 1998; Truant et al., 2008). Changes in the phase separation capacity and miscibility of the repeat expanded proteins in heterotypic condensates are consistent with these diverse observations.

Disease-associated repeat expansions occur not just in TFs, but in other nuclear proteins (e.g., ataxin genes), several of which have been linked to chromatin regulation and transcription (Darling and Uversky, 2017; La Spada and Taylor, 2010; Orr and Zoghbi, 2007). It is thus plausible that disease-associated changes may alter the phase separation capacity and miscibility of various nuclear proteins within heterotypic condensates, which in turn perturbs cellular gene expression programs. Mutations in the IDR of cellular proteins and repeat expansions in RNA have indeed been recently linked to altered phase separation (Jain and Vale, 2017; Meyer et al., 2018; Molliex et al., 2015; Patel et al., 2015). Dysregulated phase separation may thus underlie a wide spectrum of diseases.

The results presented here also provide insights into the function of TF activation domains (ADs). TF ADs are typically intrinsically disordered sequences that interact with components of the transcription machinery, and recent studies suggest that various TF ADs (i.e., IDRs) have the capacity to phase separate and to co-condense with the Mediator co-activator *in vitro*, although the functional importance of these observations has been unclear (Boija et al., 2018; Chong et al., 2018). The results presented here provide evidence for the importance of biochemical characteristics that drive TF phase separation for normal TF function *in vivo*.

Eukaryotic genomes tend to encode hundreds of TFs, whose ADs have limited sequence-level similarity, but many ADs are nevertheless functionally interchangeable (Lambert et al., 2018; Mitchell and Tjian, 1989; Staby et al., 2017; Stampfel et al., 2015). A key outstanding question is whether the interchangeability of TF IDRs is determined by their phase separation capacity and whether the phase separation capacity of interchangeable TF IDRs has a shared molecular basis. The phase separation of the IDR of the FUS RNA-binding protein, for example, is predominantly driven by cation- π interactions between tyrosine and arginine residues (Wang et al., 2018), while the phase separation capacity of the OCT4 and GCN4 TFs' IDRs is predominantly driven by negatively charged residues (Boija et al., 2018). Our draft catalog of human TF IDRs indicates that at least for a subset of TFs, their phase separation may be driven by alanine residues, and different TF IDR clusters correlate with various biological functions and phenotypic effects. Future work into the molecular basis of TF phase separation, and how developmental signaling, post-translational modifications, and genetic variants impact the miscibility of TFs in heterotypic condensates with components of the transcriptional machinery may thus open new condensate-directed therapeutic avenues for diseases associated with transcriptional dysregulation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Cell culture
 - Animal model
- **METHOD DETAILS**
 - Experimental design
 - Construct generation
 - Isolation of limb bud cells
 - Mouse genotyping
 - Whole mount *in situ* hybridization (WISH)
 - Cell treatments
 - Western blot
 - Immunofluorescence and confocal microscopy
 - Stochastic optical reconstruction microscopy (STORM)
 - Live cell imaging
 - Protein purification and *in vitro* droplet formation assays
 - Luciferase reporter assays
 - H3K27Ac ChIP-seq with S2 spike-in control (ChIP-Rx)
 - Flag-Hoxd13 ChIP-Seq from chicken micromass (chMM) cells
 - Single-cell RNA-seq data generation
 - Repeat expansions in HOXD13, HOXA13, RUNX2 and TBP
 - Protein disorder analysis of HOXD13, HOXA13, RUNX2, TBP and FOXP1-4
 - IDR classification: data retrieval and preprocessing
 - IDR/DBD clustering and annotation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Confocal/Fluorescence image analysis
 - STORM image analysis
 - OptoDroplet assay analysis
 - LacO-LacI tethering assay analysis
 - *In vitro* condensate assay analysis
 - Statistical analysis
 - Single-cell RNA-seq data pre-processing
 - Single-cell RNA-seq analysis: filtering and normalization
 - Cluster identification
 - Assignment of cell types to clusters
 - Cluster 1. Proliferating cells (S phase) of the distal proliferating mesenchyme
 - Cluster 2. Proliferating cells (Cytokinesis) of the distal proliferating mesenchyme
 - Cluster 3. Perichondrium (Proliferating)
 - Cluster 4. Interdigital mesenchyme
 - Cluster 5. Perichondrium
 - Cluster 6. Proximal chondrocytes (carpal progenitors)
 - Cluster 7. Distal chondrocytes (phalanx progenitors)
 - Cluster 8. Proliferating cells
 - Cluster 9. Myoblasts
 - Cluster 10. Hematopoietic cells
 - Cluster 11. Erythrocytes

- Quantification of cell state proportions
- Differential expression analysis
- Combined wild-type and *spdh* scRNA-seq data analysis
- Gene ontology (GO) term enrichment analysis
- ChIP-Seq data (previously published)
- ChIP-Seq data processing
- H3K27Ac ChIP-Seq analysis
- Capture C data
- Topologically Associating Domains (TADs)
- Enrichment of HOXD13 ChIP-Seq peaks around Cluster 4 dysregulated genes
- Enrichment of Capture C signal at HOXD13 peaks
- Mean expression value of individual genes

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.04.018>.

ACKNOWLEDGMENTS

We thank members of the Meissner laboratory, Dario Lupianez, Phillip Sharp, and Martin Vingron for discussions. We are grateful to Charles Haggerty for S2 cells; Abishek Sampath Kumar for help with scRNA-seq libraries; Sigmar Stricker for help with scRNA-seq cluster-to-cell type assignment; Asita Stiege for help with the luciferase assays; Helene Kretzmer for discussions on scRNA-seq; Alicia Zamudio and Jurian Schuijers for help with LacI-tethering experiments; Rene Buschow, Thorsten Mielke, and the MPIMG Imaging Facility for help with microscopy and image analysis; Celine Hillgardt, Christin Franke, and the MPIMG transgenic facility for help with mouse work; and the MPIMG Sequencing Core for help with sequencing. All data are available in the supplementary materials, and the sequencing data were deposited at GEO: GSE128818. This work was funded by the Max Planck Society and partially supported by grants from the NIH (1P50HG006193 to A.M.) and the Deutsche Forschungsgemeinschaft (DFG), including DFG project 278001972-TRR 186 to H.E., DFG grant KR3985/7 to S.M., and SPP2202 Priority Program grants from the DFG: MU880/16-1 (to S.M.), IB 139/1-1 (to D.M.I.), and HN 4/1-1 (to D.H.). H.N. is supported by a fellowship of the Emil Aaltonen Foundation.

AUTHOR CONTRIBUTIONS

S.B., S.D.M., D.M.I., and D.H. designed and conceived the study. S.B., S.D.M., V.A., D.K., and D.H. prepared the manuscript with the assistance of the other authors. S.B. performed protein purification and *in vitro* condensation experiments. S.B. and D.K. performed optoDroplet experiments. S.B. and V.A. performed IF experiments. S.B. and H.G. performed super-resolution microscopy. S.B. and D.K. performed optoDroplet image analysis. S.A., D.M.I., and S.B. performed limb bud cell isolation. S.G. performed scRNA-seq. H.N. performed LacI-tethering and luciferase experiments. S.B. performed LacI-tethering, STORM, and *in vitro* condensate image analysis. S.B., V.A., D.K., and H.N. generated DNA constructs. S.B. and S.A. performed H3K27Ac ChIP-seq. I.J. performed Hoxd13-FLAG ChIP-seq. S.D.M. performed scRNA-seq, ChIP-seq, and Capture C data analyses, as well as identification and molecular classification of TF IDRs. H.E., A.M., S.M., D.M.I., and D.H. provided resources, funding, and expertise. D.H. supervised the work.

DECLARATION OF INTERESTS

The Max Planck Society has filed a patent application based on this paper.

Received: August 1, 2019
Revised: January 16, 2020

Accepted: April 13, 2020

Published: May 7, 2020

REFERENCES

- Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A., and de Crombrughe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 16, 2813–2828.
- Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 176, 419–434.
- Albrecht, A., and Mundlos, S. (2005). The other trinucleotide repeat: polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* 15, 285–293.
- Albrecht, A.N., Kornak, U., Böddrich, A., Süring, K., Robinson, P.N., Stiege, A.C., Lurz, R., Stricker, S., Wanker, E.E., and Mundlos, S. (2004). A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. *Hum. Mol. Genet.* 13, 2351–2359.
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–603.
- Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298.
- Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691–1692.
- Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrughe, B. (1999). Sox9 is required for cartilage formation. *Nat. Genet.* 22, 85–89.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008 <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* 25, 833–840.
- Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842–1855.
- Boulet, A.M., and Capecchi, M.R. (2004). Multiple roles of Hoxa11 and Hoxd11 in the formation of the mammalian forelimb zeugopod. *Development* 131, 299–309.
- Bruneau, S., Johnson, K.R., Yamamoto, M., Kuroiwa, A., and Duboule, D. (2001). The mouse Hoxd13(*spdh*) mutation, a polyalanine expansion similar to human type II synpolydactyly (SPD), disrupts the function but not the expression of other Hoxd genes. *Dev. Biol.* 237, 345–353.
- Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malan-gone, C., McMahon, A., Morales, J., Mountjoy, E., Solis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415.
- Choi, J.M., Dar, F., and Pappu, R.V. (2019). LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLoS Comput. Biol.* 15, e1007028.
- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging

dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361, eaar2555.

Darling, A.L., and Uversky, V.N. (2017). Intrinsic Disorder in Proteins with Pathogenic Repeat Expansions. *Molecules* 22, E2027.

Davidson, D.R., Crawley, A., Hill, R.E., and Tickle, C. (1991). Position-dependent expression of two related homeobox genes in developing vertebrate limbs. *Nature* 352, 429–431.

Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L., and Bates, G.P. (1997). Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90, 537–548.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Dollé, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbauer, B., Chambon, P., and Duboule, D. (1993). Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neonatal limbs. *Cell* 75, 431–441.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48.

Endesfelder, U., Malkusch, S., Fricke, F., and Heilemann, M. (2014). A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* 141, 629–638.

Fabrizius, V., Lefebvre, J., Geertsema, H., Marino, S.F., and Ewers, H. (2018). Rapid and efficient C-terminal labeling of nanobodies for DNA-PAINT. *J. Phys. D Appl. Phys.* 51.

Gallili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720.

Goodman, F.R., Mundlos, S., Muragaki, Y., Donnai, D., Giovannucci-Uzielli, M.L., Lapi, E., Majewski, F., McGaughan, J., McKeown, C., Reardon, W., et al. (1997). Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proc. Natl. Acad. Sci. USA* 94, 7458–7463.

Goodman, F.R., Bacchelli, C., Brady, A.F., Brueton, L.A., Fryns, J.P., Mortlock, D.P., Innis, J.W., Holmes, L.B., Donnay, A.E., Feingold, M., et al. (2000). Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am. J. Hum. Genet.* 67, 197–202.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.

Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548.

Hasty, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N., and Klein, W.H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* 364, 501–506.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23.

Ibrahim, D.M., Hansen, P., Rödelberger, C., Stiege, A.C., Doelken, S.C., Horn, D., Jäger, M., Janetzki, C., Krawitz, P., Leschik, G., et al. (2013). Distinct global shifts in genomic binding profiles of limb malformation-associated HOXD13 mutations. *Genome Res.* 23, 2091–2102.

Innis, J.W., Mortlock, D., Chen, Z., Ludwig, M., Williams, M.E., Williams, T.M., Doyle, C.D., Shao, Z., Glynn, M., Mikulic, D., et al. (2004). Polyalanine expansion in HOXA13: three new affected families and the molecular consequences in a mouse model. *Hum. Mol. Genet.* 13, 2841–2851.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. *Nature* 546, 243–247.

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., and Spec-

tor, D.L. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell* 116, 683–698.

Jerković, I., Ibrahim, D.M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., González Navarrete, I., Robinson, P.N., Hecht, J., and Mundlos, S. (2017). Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF. *PLoS Genet.* 13, e1006567.

Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl. Acad. Sci. USA* 95, 8538–8543.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496.

Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 42, D966–D974.

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpfli, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despain, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat. Cell Biol.* 21, 305–310.

Kuss, P., Villavicencio-Lorini, P., Witte, F., Klose, J., Albrecht, A.N., Seemann, P., Hecht, J., and Mundlos, S. (2009). Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *J. Clin. Invest.* 119, 146–156.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155, 1049–1060.

La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* 11, 247–258.

Lagache, T., Grassart, A., Dallongeville, S., Faklaris, O., Sauvonnnet, N., Dufour, A., Danglot, L., and Olivo-Marin, J.C. (2018). Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat. Commun.* 9, 698.

Lallemand, Y., Bensoussan, V., Clément, C.S., and Robert, B. (2009). Mxg genes are important apoptosis effectors downstream of the Shh/Gli3 pathway in the limb. *Dev. Biol.* 331, 189–198.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* 175, 598–599.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* 157, 13–25.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, J., Dong, A., Saydaminova, K., Chang, H., Wang, G., Ochiai, H., Yamamoto, T., and Pertsinidis, A. (2019). Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* 178, 491–506.

Lu, M.F., Cheng, H.T., Kern, M.J., Potter, S.S., Tran, B., Diekwisch, T.G., and Martin, J.F. (1999a). prx-1 functions cooperatively with another paired-related homeobox gene, prx-2, to maintain cell fates within the craniofacial mesenchyme. *Development* 126, 495–504.

Lu, M.F., Cheng, H.T., Lacy, A.R., Kern, M.J., Argao, E.A., Potter, S.S., Olson, E.N., and Martin, J.F. (1999b). Paired-related homeobox genes cooperate in handplate and hindlimb zeugopod morphogenesis. *Dev. Biol.* 205, 145–157.

- Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 558, 318–323.
- Maltecca, F., Filla, A., Castaldo, I., Coppola, G., Fragassi, N.A., Carella, M., Bruni, A., Cocozza, S., Casari, G., Servadio, A., and De Michele, G. (2003). Intergenerational instability and marked anticipation in SCA-17. *Neurology* 61, 1441–1443.
- Manders, E.M.M., Verbeek, F.J., and Aten, J.A. (1993). Measurement of colocalization of objects in dual-colour confocal images. *J. Microsc.* 169, 375–382.
- Mastushita, M., Kitoh, H., Subasioglu, A., Kurt Colak, F., Dundar, M., Mishima, K., Nishida, Y., and Ishiguro, N. (2015). A Glutamine Repeat Variant of the RUNX2 Gene Causes Cleidocranial Dysplasia. *Mol. Syndromol.* 6, 50–53.
- Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, D142–D147.
- Meyer, K., Kirchner, M., Uyar, B., Cheng, J.Y., Russo, G., Hernandez-Miranda, L.R., Szymborska, A., Zaubner, H., Rudolph, I.M., Willnow, T.E., et al. (2018). Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. *Cell* 175, 239–253.
- Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245, 371–378.
- Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163, 123–133.
- Mundlos, S., Otto, F., Mundlos, C., Mulliken, J.B., Aylsworth, A.S., Albright, S., Lindhout, D., Cole, W.G., Henn, W., Knoll, J.H., et al. (1997). Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* 89, 773–779.
- Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B.R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272, 548–551.
- Nakamura, K., Jeong, S.Y., Uchihara, T., Anno, M., Nagashima, K., Nagashima, T., Ikeda, S., Tsuji, S., and Kanazawa, I. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum. Mol. Genet.* 10, 1441–1448.
- Neubüser, A., Koseki, H., and Balling, R. (1995). Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1. *Dev. Biol.* 170, 701–716.
- Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., and Sato, M. (2015). CRISPR-Cas9-based photoactivatable transcription system. *Chem. Biol.* 22, 169–174.
- O'Rourke, M.P., Soo, K., Behringer, R.R., Hui, C.C., and Tam, P.P. (2002). Twist plays an essential role in FGF and SHH signal transduction during mouse limb development. *Dev. Biol.* 248, 143–156.
- Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* 9, 1163–1170.
- Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575–621.
- Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoyanov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077.
- Patel, A., Malinowska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., and Hyman, A.A. (2017). ATP as a biological hydrotrope. *Science* 356, 753–756.
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7, 208.
- R Development Core Team (2018). R: A language and environment for statistical computing (R Foundation for Statistical Computing). <http://www.R-project.org/>.
- Ramirez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44 (W1), W160–5.
- Riback, J.A., Zhu, L., Ferrolino, M.C., Tolbert, M., Mitrea, D.M., Sanders, D.W., Wei, M.T., Kriwacki, R.W., and Brangwynne, C.P. (2019). Composition dependent phase separation underlies directional flux through the nucleolus. *bioRxiv*. <https://doi.org/10.1101/809210>.
- Ross, C.A. (2002). Polyglutamine pathogenesis: emergence of unifying mechanisms for Huntington's disease and related disorders. *Neuron* 35, 819–822.
- Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med.* 10 (Suppl), S10–S17.
- Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Co-activator condensation at super-enhancers links phase separation and gene control. *Science* 361, eaar3958.
- Saudou, F., Finkbeiner, S., Devys, D., and Greenberg, M.E. (1998). Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95, 55–66.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.
- Sheth, R., Barozzi, I., Langlais, D., Osterwalder, M., Nemec, S., Carlson, H.L., Stadler, H.S., Visel, A., Drouin, J., and Kmita, M. (2016). Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. *Cell Rep.* 17, 2913–2926.
- Shibata, A., Machida, J., Yamaguchi, S., Kimura, M., Tatematsu, T., Miyachi, H., Matsushita, M., Kitoh, H., Ishiguro, N., Nakayama, A., et al. (2016). Characterisation of novel RUNX2 mutation with alanine tract expansion from Japanese cleidocranial dysplasia patient. *Mutagenesis* 31, 61–67.
- Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. *Science* 357, eaaf4382.
- Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E., and Brangwynne, C.P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* 168, 159–171.
- Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., and Skriver, K. (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.* 474, 2509–2532.
- Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* 528, 147–151.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234.
- Swinehart, I.T., Schlientz, A.J., Quintanilla, C.A., Mortlock, D.P., and Wellik, D.M. (2013). Hox11 genes are required for regional patterning and integration of muscle, tendon and bone. *Development* 140, 4574–4582.
- Truant, R., Atwal, R.S., Desmond, C., Munsie, L., and Tran, T. (2008). Huntington's disease: revisiting the aggregation hypothesis in polyglutamine neurodegenerative diseases. *FEBS J.* 275, 4252–4262.
- Villavicencio-Lorini, P., Kuss, P., Friedrich, J., Haupt, J., Farooq, M., Türkmen, S., Duboule, D., Hecht, J., and Mundlos, S. (2010). Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation. *J. Clin. Invest.* 120, 1994–2004.
- Wang, J., Choi, J.M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaître, R., Pozniakovskiy, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174, 688–699.

- Wheeler, J.R., Lee, H.O., Poser, I., Pal, A., Doeleman, T., Kishigami, S., Kour, S., Anderson, E.N., Marrone, L., Murthy, A.C., et al. (2019). Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv*. <https://doi.org/10.1101/721001>.
- Yoshida, M., Hata, K., Takashima, R., Ono, K., Nakamura, E., Takahata, Y., Murakami, T., Iseki, S., Takano-Yamamoto, T., Nishimura, R., and Yoneda, T. (2015). The transcription factor Foxc1 is necessary for Ihh-Gli2-regulated endochondral ossification. *Nat. Commun.* 6, 6653.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
- Zamudio, A.V., Dall'Agnese, A., Henninger, J.E., Manteiga, J.C., Afeyan, L.K., Hannett, N.M., Coffey, E.L., Li, C.H., Oksuz, O., Sabari, B.R., et al. (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell* 76, 753–766.
- Zhang, P., Jimenez, S.A., and Stokes, D.G. (2003). Regulation of human COL9A1 gene expression. Activation of the proximal promoter region by SOX9. *J. Biol. Chem.* 278, 117–123.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
- Zhao, Q., Eberspaecher, H., Lefebvre, V., and De Crombrughe, B. (1997). Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. *Dev. Dyn.* 209, 377–386.
- Zoghbi, H.Y., and Orr, H.T. (2000). Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23, 217–247.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
HOXD13	ThermoFisher	PA5-66661; RRID: AB_2662801
HOXD13	Abcam	ab229234
HOXD13	Abcam	ab19866; RRID: AB_733004
HSP90	BD Biosciences	610419; RRID: AB_397799
TBP	ThermoFisher	MA 3080; RRID: AB_2633325
BRD4	Santa Cruz	Sc-518021
HP1 α	Millipore	05-689; RRID: AB_2290973
H3K27Ac	Diagenode	C15410174; RRID: AB_2716835
Goat anti-Rabbit IgG Cy3	Invitrogen	A10520; RRID: AB_2534029
Goat anti-Mouse IgG Alexa Fluor 647	Invitrogen	A21236; RRID: AB_141725
Goat anti-Rabbit IgG Alexa Fluor 647	Invitrogen	A32733; RRID: AB_2633282
Donkey anti-Rabbit IgG Alexa Fluor 568	Invitrogen	A10042; RRID: AB_2534017
Digoxigenin-AP	Roche	11093274910; RRID: AB_514497
Bacterial and Virus Strains		
BL21(DE3)	NEB	C2527H
Chemicals, Peptides, and Recombinant Proteins		
1,6-Hexanediol	Sigma	240117
ATP	Jena Bioscience	NU-1010
Mitoxantrone HCl	Sigma	M6545
(\pm)- α -Lipoic acid	Sigma	62320
Lipoamide	Medchemexpress	HY-B1142
Urea	Sigma	U5378
Critical Commercial Assays		
NEBuilder HiFi DNA Assembly Master Mix	NEB	E2521
Q5 Polymerase	NEB	M0491
DNA Midi / Mini Prep Kit	QIAGEN	51183 / 51304
Dual-Glo Luciferase Assay System	Promega	E2920
FuGENE HD Transfection Reagent	Promega	E2311
LipoD293 <i>In Vitro</i> DNA Transfection Reagent	Signagen	SL100668
Deposited Data		
scRNaseq WT_E12.5	This Paper	GEO: GSM3685917
scRNaseq SPDH_E12.5	This Paper	GEO: GSM3685918
WT_H3K27ac_ChIPSeq_spikein_Rep1	This Paper	GEO: GSM4230262
WT_H3K27ac_ChIPSeq_spikein_Rep2	This Paper	GEO: GSM4230263
SPDH_H3K27ac_ChIPSeq_spikein_Rep1	This Paper	GEO: GSM4230264
SPDH_H3K27ac_ChIPSeq_spikein_Rep2	This Paper	GEO: GSM4230265
WT_H3K27ac_ChIPSeq_spikein_INPUT_Rep1	This Paper	GEO: GSM4230266
WT_H3K27ac_ChIPSeq_spikein_INPUT_Rep2	This Paper	GEO: GSM4230267
SPDH_H3K27ac_ChIPSeq_spikein_INPUT_Rep1	This Paper	GEO: GSM4230268
SPDH_H3K27ac_ChIPSeq_spikein_INPUT_Rep2	This Paper	GEO: GSM4230269
WT_HOXD13_transfected_chicken_Rep1	This Paper	GEO: GSM4230270
WT_HOXD13_transfected_chicken_Rep2	This Paper	GEO: GSM4230271

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SPDH_transfected_chicken	This Paper	GEO: GSM4230272
WT_E11.5_HOXD13_ChIP-seq	Sheth et al., 2016	GEO: GSM2151014
Experimental Models: Cell Lines		
HEK293T cells	ATCC	N/A
Kelly Cells	DSMZ	ACC-355
SH-SY5Y cells	DSMZ	ACC-209
U2OS-2-6-3 cells	Zamudio et al., 2019	N/A
Cos-7 cells	ATCC	CRL-1651
Experimental Models: Organisms/Strains		
WT mouse embryos E12.5	Mundlos laboratory (MPIMG) littermates to <i>spdh</i>	N/A
SPDH mouse embryos E12.5	Mundlos laboratory (MPIMG) B6C3Fe <i>a/a-Hoxd13^{spdh}/J</i>	JAX #002875
Recombinant DNA		
pHR-mCherry-CRY2	Shin et al., 2017	Addgene Plasmid #101221
pHR-HOXD13-IDR-mCherry-CRY2	This Paper	Deposited to Addgene Cat# 145258
pHR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145259
pHR-HOXD13-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145260
pHR-HOXD13(+7A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145261
pHR-HOXD13(+8A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145262
pHR-HOXD13(+9A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145263
pHR-HOXD13(+14A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145264
pHR-HOXA13-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145265
pHR-HOXA13(+7A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145266
pHR-RUNX2-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145267
pHR-RUNX2(+10A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145268
pHR-TBP(38Q)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145269
pHR-TBP(53Q)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145270
pHR-HOXD13(−7A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145271
pHR-HOXD13(−15A)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145272
pHR-HOXD13(ΔEdel)-IDR-mCherry-CRY2-NLS	This Paper	Deposited to Addgene Cat# 145273
pHR-MCS-YFP-NLS	This Paper	Deposited to Addgene Cat# 145274
pHR-MCS-MED1long-IDR-YFP-NLS	This Paper	Deposited to Addgene Cat# 145276
pHR-MED1trunc-IDR-YFP-NLS	This Paper	Deposited to Addgene Cat# 145275
pHR-HOXD13-IDR-YFP-NLS	This Paper	Deposited to Addgene Cat# 145277
pHR-HOXD13(+7A)-IDR-YFP-NLS	This Paper	Deposited to Addgene Cat# 145278
pET-45b(+)	Novagen	Novagen Cat# 71327
pET-45b-mCherry	This paper	Deposited to Addgene Cat# 145279
pET-45b-HOXD13-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145280
pET-45b-HOXD13(+7A)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145281
pET-45b-HOXD13(+10A)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145282
pET-45b-TBP(38Q)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145283
pET-45b-HOXA13(WT)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145284
pET-45b-HOXA13(+7A)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145285
pET-45b-RUNX2(WT)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145286
pET-45b-RUNX2(+10A)-IDR-mCherry	This Paper	Deposited to Addgene Cat# 145287
pET-28a-mEGFP-MED1-IDR	Sabari et al., 2018	N/A
GAL4-UAS-Luciferase reporter	Nihongaki et al., 2015	Addgene Plasmid #64125

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRL-CMV	Promega	E2261
pGL2-Promoter-Aldh1a2	Kuss et al., 2009	N/A
pcDNA3.1-Hoxd13wt	Kuss et al., 2009	N/A
pcDNA3.1-Hoxd13(+7A)	Kuss et al., 2009	N/A
pcDNA3.1-Hoxd13(+10A)	Kuss et al., 2009	N/A
pCFP-LacI-MED1	Zamudio et al., 2019	N/A
pCFP-LacI-HOXD13wt-IDR	This Paper	Deposited to Addgene Cat# 145255
pCFP-LacI-HOXD13(+7A)-IDR	This Paper	Deposited to Addgene Cat# 145256
pCFP-LacI-HOXD13(+10A)-IDR	This Paper	Deposited to Addgene Cat# 145257
pGAL4-DBD-MCS	This Paper	Deposited to Addgene Cat# 145245
pGAL4-DBD-MCS-long	This Paper	Deposited to Addgene Cat# 145246
pGAL4-DBD-HOXA13wt-IDR	This Paper	Deposited to Addgene Cat# 145240
pGAL4-DBD-HOXA13+7A-IDR	This Paper	Deposited to Addgene Cat# 145239
pGAL4-DBD-HOXA13-15A-IDR	This Paper	Deposited to Addgene Cat# 145236
pGAL4-DBD-HOXA13-44A-IDR	This Paper	Deposited to Addgene Cat# 145237
pGAL4-DBD-HOXA13-58A-IDR	This Paper	Deposited to Addgene Cat# 145238
pGAL4-DBD-HOXD13wt-IDR	This Paper	Deposited to Addgene Cat# 145244
pGAL4-DBD-HOXD13-Adel-15A-IDR	This Paper	Deposited to Addgene Cat# 145243
pGAL4-DBD-RUNX2wt-IDR	This Paper	Deposited to Addgene Cat# 145252
pGAL4-DBD-RUNX2+10A-IDR	This Paper	Deposited to Addgene Cat# 145251
pGAL4-DBD-EOMESwt-IDR	This Paper	Deposited to Addgene Cat# 145231
pGAL4-DBD-EOMES-Adel-IDR	This Paper	Deposited to Addgene Cat# 145230
pGAL4-DBD-TBX1wt-IDR	This Paper	Deposited to Addgene Cat# 145254
pGAL4-DBD-TBX1-Adel-IDR	This Paper	Deposited to Addgene Cat# 145253
pGAL4-DBD-OLIG2wt-IDR	This Paper	Deposited to Addgene Cat# 145250
pGAL4-DBD-OLIG2-Adel-IDR	This Paper	Deposited to Addgene Cat# 145249
pGAL4-DBD-MNX1wt-IDR	This Paper	Deposited to Addgene Cat# 145248
pGAL4-DBD-MNX1-Adel-IDR	This Paper	Deposited to Addgene Cat# 145247
pGAL4-DBD-BHLHE41wt-IDR	This Paper	Deposited to Addgene Cat# 145229
pGAL4-DBD-BHLHE41-Adel-IDR	This Paper	Deposited to Addgene Cat# 145228
pGAL4-DBD-GSX2wt-IDR	This Paper	Deposited to Addgene Cat# 145233
pGAL4-DBD-GSX2-Adel-IDR	This Paper	Deposited to Addgene Cat# 145232
pGAL4-DBD-HEY2wt-IDR	This Paper	Deposited to Addgene Cat# 145235
pGAL4-DBD-HEY2-Adel-IDR	This Paper	Deposited to Addgene Cat# 145234
pGAL4-DBD-HOXD11wt-IDR	This Paper	Deposited to Addgene Cat# 145242
pGAL4-DBD-HOXD11-Adel-IDR	This Paper	Deposited to Addgene Cat# 145241
RCASBP(A)-3xFlag-mHoxd13	Ibrahim et al., 2013	Deposited to Addgene Cat# 145288
RCASBP(A)-3xFlag-mHoxd13+7Ala	This Paper	Deposited to Addgene Cat# 145289
Software and Algorithms		
arivis Vision4D 3.1	arivis	https://www.arivis.com/en/imaging-science/arivis-vision4d
PONDR	Peng et al., 2006	http://www.pondr.com/
ProtParam	Artimo et al., 2012	https://web.expasy.org/protparam/
R-Studio	R-Studio	https://www.rstudio.com/
R	R Development Core Team, 2018	https://www.R-project.org/
bamtools	Barnett et al., 2011	https://github.com/pezmaster31/bamtools
samtools	Li et al., 2009	http://samtools.sourceforge.net
bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
bwa	Li and Durbin, 2009	http://bio-bwa.sourceforge.net
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
Cell Ranger	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
Seurat 2.4.3	Butler et al., 2018	https://satijalab.org/seurat/
Macs14	Zhang et al., 2008	https://github.com/taoliu/MACS
Macs2	Zhang et al., 2008	https://github.com/taoliu/MACS
bamToGff_turbo.py	BradnerLab	https://github.com/BradnerLab/pipeline
Zen 3.0 (blue edition)	Zeiss	https://www.zeiss.com/microscopy/int/products/microscope-software/zen.html
FIJI	Schindelin et al., 2012	http://fiji.sc/
Vutara SRX 6.04.14	Bruker Corporation	https://www.bruker.com/
Code for image analyses, IDR analyses and sequence analyses	This Paper	https://github.com/hnislzlab/hoxd13 https://github.com/BasuShaon/ChemicalBiology
Other		
Non-Inducing Medium (MDG)	Studier, 2005	N/A
Automatic Induction Medium (ZYM-5052)	Studier, 2005	N/A
His GraviTrap Columns	GE-Healthcare	11003399
Ibittreat μ -Slide 8 Well Chambered Coverslips	Ibidi	80826

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Denes Hnislz (hnislz@molgen.mpg.de).

Materials Availability

All unique reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement. Plasmids generated in the study were deposited at Addgene, and are available at <http://www.addgene.org>.

Data and Code Availability

All next generation sequencing data generated in the study were deposited at the Gene Expression Omnibus (GEO) under the accession number GSE128818.

Original data including all raw microscopy images were deposited at Mendeley Data (<https://data.mendeley.com/>) under <https://doi.org/10.17632/ztd6wzcv7h.13>.

Code supporting the study were deposited at Github (<https://github.com/hnislzlab/hoxd13>). Code for droplet visualization is available at <https://github.com/BasuShaon/ChemicalBiology>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

HEK293T, Kelly, SH-SY5Y, U2OS-2-6-3, and Cos7 cells were cultured in DMEM with GlutaMAX (ThermoFisher Scientific, 10566-016) supplemented with 10% FBS (Sigma Aldrich, F4135), and 100 U/mL penicillin-streptomycin (GIBCO, 15140), at 37°C with 5% CO₂ in a humidified incubator. Cells were negative for mycoplasma.

Animal model

All animal procedures were conducted in the animal facility of the MPIMG, and were approved by the local authorities (LAGeSo Berlin) under the license number #ZH-120. Mice were monitored daily for food and water intake, and animals were inspected by the Chair of the Animal Welfare and Research Committee and the head of the Animal Facility. For staged matings, females were controlled for

CTGCCACGCCAGCTTCGGAGAGTTCTGGGggtaccGGAGGCGGGATGGTTTCCAAAGGAGAGGAGGATAATATGGCTATAATTAA
AGAGTTTATGCGGTTCAAGgtgcatatggaggatctg

RUNX2 wt gene fragment:

gcaaggcaggACTAGTGCCACCATGGCATCAAACAGCCTCTTCAGCACAGTGACACCATGTCAGCAAACTTCTTTGGGATCCGA
GCACCAGCCGGCGCTTCAGCCCCCTCCAGCAGCCTGCAGCCCGGCAAAATGAGCGACGTGAGCCCGGTGGTGGCTGCGCA
ACAGCAGCAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGGAGGCGGCGGCGGC
CGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGAGCTGCAGTGCCCCGTTACGGCCGCCCCACGACAACCGCACCATGG
TGGAGACGTACGaagccaggc

RUNX2 +10A gene fragment:

gcaaggcaggACTAGTGCCACCATGGCATCAAACAGCCTCTTCAGCACAGTGACACCATGTCAGCAAACTTCTTTGGGATCCGA
GCACCAGCCGGCGCTTCAGCCCCCTCCAGCAGCCTGCAGCCCGGCAAAATGAGCGACGTGAGCCCGGTGGTGGCTGCGCA
ACAGCAGCAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGGAGGCGGCGGCGCAGC
CGCGGCTGCCGCGAGCGGCTGCAGCGGCGGCGGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGAGCTGCAGTGCCCCG
TTGCGGCCGCCCCACGACAACCGCACCATGGTGGAGACGTACGaagccaggc

Multiple cloning site (MCS) gene fragment:

tcaagtacgGGCGCGCCAGTCTCCGACAGACTGAGTCGCCCGGgaagagagggccACTAGTcgcgaccacgCGCCGGCGggagcagagac
GCGATCGCggagcgcaacGTACGgggtctggtctgctggttccATGGTTTCCAAAGGAGAGGAGGATAATATGGCTATAATTAAAGAGTTTAT
GCGGTTCAAGGTGCATATGcggaggcagg

For the generation of IDR-YFP constructs, a YFP fragment was first PCR amplified from an mCitrine vector, and cloned into the pHR-mCherry-CRY2-NLS plasmid using *BlpI* and *NdeI* restriction sites, to generate pHR-YFP-NLS. HOXD13(wt)-IDR and HOXD13(+7A)-IDRs were subcloned from optoIDR vectors into the pHR-YFP-NLS plasmid, to generate pHR-HOXD13-IDR-YFP-NLS and pHR-HOXD13(+7A)-IDR-YFP-NLS using *AscI* and *NdeI* restriction sites. For the generation of pHR-MED1trunc-IDR-YFP-NLS, the MED1-IDR was PCR amplified from a MED1 IDR expression vector (Boija et al., 2018; Sabari et al., 2018), and subcloned into pHR-YFP-NLS using *AscI* and *NdeI* restriction sites. To generate pHR-MCS-MED1long-IDR-YFP-NLS, a multiple cloning site was introduced into pHR-YFP-NLS via *AscI* and *NdeI* sites to generate pHR-MCS-YFP-NLS. The MED1-IDR was then PCR amplified from a MED1 IDR expression vector (Boija et al., 2018; Sabari et al., 2018), and ligated into pHR-MCS-YFP-NLS using the *AsiI* and *BsiWI* restriction sites. pHR-MED1trunc-IDR-YFP-NLS was used for exclusion microscopy with OptoIDR constructs (Figures S2D and S2E), and pHR-MCS-MED1long-IDR-YFP-NLS was used for *Lacl* tethering experiments (Figures 4E and 4F).

mEYFP forward primer:

tgaggttcaaggtgcatatgGTGAGCAAGGGCGAGGAGC

mEYFP reverse primer:

aggtgggagttgcggccgctcatatcttcttcttttggaccgctccCTGTACAGCTCGTCCATGC

MED1trunc-IDR forward primer:

cacaaccctcactcgccgagcagctcccgacagactgagtcgccccggggggatctgGAGCATCACAGTGGTAGTCAGGG

MED1trunc-IDR reverse primer:

gcgaggaggataacaATTAATtctccgctacctccAAGATCATCATCTTCCTCCCCAATC

Multiple cloning site (MCS) for YFP gene fragment:

tcaagtacgGGCGCGCCAGTCTCCGACAGACTGAGTCGCCCGGgaagagagggccACTAGTcgcgaccacgCGCCGGCGggagcagagacG
CGATCGCggagcgcaacGTACGgggtctggtctgctggttccCATATGtcacgtcagt

MCS-MED1long-IDR forward primer:

gaggccGCGATCGCgagctaccatgGAGCATCACAGTGGTAGTCAGGG

MCS-MED1long-IDR reverse primer:

gcgaggaggataaacacgtacgtccgctacctccAAGATCATCATCTTCCTCCCCAATC

For the generation of plasmids for the expression of recombinant proteins, pET-45b(+) was ordered from Novagen (Sigma Cat No. 71327). Gene fragments encoding IDRs tagged with mCherry were PCR amplified from pHR-IDR-mCherry-CRY2 vectors (described above) using Q5 polymerase (NEB M0494S), and ligated into pET-45b(+) using *AgeI* and *HindIII* restriction sites. This pipeline was used to generate pET-45b-HOXD13-IDR-mCherry, pET-45b-HOXD13(+7A)-IDR-mCherry, pET-45b-HOXD13(+10A)-IDR-mCherry, pET-45b-HOXA13(WT)-IDR-mCherry, pET-45b-RUNX2(WT)-IDR-mCherry, pET-45b-RUNX2(+10A)-IDR-mCherry, pET-45b-HOXA13(+7A)-IDR-mCherry, and pET-45b-TBP38Q-IDR-mCherry. For PCR amplification of RUNX2, a modified forward primer was used to anneal downstream of the MCS in the pHR-RUNX2-IDR-mCherry-CRY2 vectors (see below). To create pET-45b-mCherry, the IDR region of pET-45b-HOXD13-IDR-mCherry was excised using *KpnI* and the backbone (containing mCherry) was religated.

IDR-mCherry forward primer:

taccggtgacagtggatccggagctac

RUNX2 IDR-mCherry forward primer:

taccggtGgaagagagggccACTAGTGC

IDR-mCherry reverse primer:

cgcaagcttctgtacaattcatcatgc

In the GAL4-DBD luciferase system, GAL4 DBD-IDR fusion constructs were expressed from a backbone of GAL4-DBD expression vector [GI-GAL4-DBD (Addgene 42500)], from which GI-GAL4 sequence was removed with EcoRI and XbaI digestion. GAL4-DBD was PCR amplified from pCAG-GAL4-DBD-GBP2 vector (Addgene 49439) while introducing a short multiple cloning site C-terminal to GAL4-DBD allowing IDR cloning. Synthetic DNA fragments for wt and alanine repeat-deletion IDRs of GSX2, HOXD11, TBX1, EOMES, BHLHE41, and MNX1 sequences were flanked by AsiSI and BsiWI sequences for cloning, and purchased from Genewiz. For HEY2 and OLIG2, wt IDR sequences were PCR amplified from human iPSC gDNA and alanine repeat-deletion sequences were ordered from Genewiz. For HOXA13 and HOXD13, IDR sequences were PCR amplified from previously described vectors, except HOXA13 –44A and –58A deletion sequences were ordered from Genewiz. The HOXA13 –15A construct lacks alanines 88-103. In case of RUNX2, sequences were cloned from optoIDR constructs with SpeI and BsiWI restriction enzymes and inserted into the GAL4-DBD vector with a longer multiple cloning site with NheI and BsiWI restriction sites.

GAL4-DBD forward primer:

GATCGAATTCATGAAGCTACTGTCTTCTATCGAACAAGCA

GAL4-DBD (short MCS) reverse primer:

GATCTCTAGACGTACGACCGGTGCTAGCGATCGCCGATACAGTCAACTGTCTTTGACCTTTGTT

GAL4-DBD (long MCS) reverse primer:

GATCTCTAGACTTAAGCGTACGACCGGTGCGGCCGCTAGCGATCGCCGATACAGTCAACTGTCTTTGACCTTTGTT

OLIG2 forward primer:

GATCGCGATCGCTTTCCACCCGTCGGCCTGCG

OLIG2 reverse primer:

GATCCGTACGTCACTTGGCGTCGGAGGTGAGG

HEY2 forward primer:

GATCGCGATCGCTCATCTCAGCACTTGCGCCACCC

HEY2 reverse primer:

GATCCGTACGTTAAAAAGCTCCAACCTCTGTCCCCAG

GSX2 wt sequence:

GCGATCGCTTCGCGCTCCTTCTATGTGCGACTCGCTCATCATAAAGGATACATCCCGCCCCGGCCCCGTCTCTCCCCGAACCTCA
CCCTGGTCCCCGATTTTTTATACCCTTGGGAATGCCTCCGCCCTTGGTCATGTCCGTGTCTGGTCCCGGGTGCCCCGAGTAGGAAG
AGTGGTGCCTTTTGTGTTTGCCCGTTGTGTGTACATCCCATCTGCATAGTTCTAGAGGCAGTGAGGTGCGGGTTCAGGCGGTG
CAGGTGCAGGTGTAAGTGGAGCAGGGGGATCAGGCGTGCGGGGGCAGCCGGGGCTCTGCCGTTGCTCAAGGGTCAGTTCAG
TAGCGCTCCGGGTGATGCCAATTTTGTCCGAGAGTAAATCAGCGCACCATCATCACCATCCCCGCAACACCACCACCA
TCATCAACCACAACAGCCTGGATCAGCGGCCGCCGAGCGGCAGCTGCGGCCGAGCGGCTGCTGCTGCCGCGCTTGACAT
CCACAACATCAGCTCCGGTATGCACCGCTACCTGACGTACG

GSX2 Ala-del sequence:

GCGATCGCTTCGCGCTCCTTCTATGTGCGACTCGCTCATCATCAAGGACACCTCACGGCCTGCGCCCTCGCTGCCTGAACCTCA
CCCCGGACAGACTTCTTTATACCACTCGGGATGCCACCACCTCTCGTTATGTCTGTGTGTCAGGCCCGGGTGCCTGCCCCGCAAG
AGCGGCGCGTTCTGCGTGTGCCCTCTCTGCGTCACTTCGCATCTGCACTCCTCTAGGGGGAGTGTTGGTGTGATCTGGTGGAG
CTGGTGCCGGTGTACTGGAGCTGGGGGTTACGGGGTAGCGGGAGCAGCGGGTGTCTTCTCTCTTAAAGGTCAAGTCTCTTC
GGCTCCTGGGGACGCGCAGTTTTGCCCGCGGGTGAACCATGCGCATCATCATCACCATCCTCCGAGCACCACCATCACCATCA
TCAGCCACAACAACAGGTTTATTGGGACATCCACAACACCACGCACCCGTATGTACAGCTACCTGACGTACG

HOXD11 wt sequence:

GCGATCGCTCTGGAACGCGCTAAGTGGCCCTACAGAGGAGGCGGAGGTGGGGGATCAGCTGGCGGAGGGAGCTCAGGAGG
CGGACCAGGTGGCGGTGGTGGTGGCGCGGGTGGGTACGCACCCTATTACGAAGAGGCAGCTATGCAGCGGGAGTTGCTTCCAC
CAGCTGGGCGCCGACCCGATGTTCTGTTTAAAGGCTCCAGAACCCGTATGCGCTGCGCCGGGACCCCGCACGGTCCCGCCGG
GGCGGCTAGTAACCTTCTACTCTGCAGTGGGCCGCAACGGTATTTTGGCGCAAGGTTTGTATCAGTTTACGAAGCAGCGCCCGGA
CCTCCATTTGCGGGCCCCGCAACCGCCACCACCTCCTGCTCCTCCACAACCAGAAGGTGCCGCTGACAAGGGTGACCCTAGGAC
AGGTGCTGGTGGCGGCGGTGGCTCACCTTGTACGAAGGCAACGCCCGGTTCTGAGCCCAAAGGTGCAGCGGAAGGTAGCGGC
GGAGACGGTGAAGGGCCACCTGGTGGAGCGGGAGCCGAGAAGAGTTCTTACGCGGTAGCTCCACAACGGTCTCGCAAGAAGCG
GTGCCCATATACCAAGTACTGACGTACG

HOXD11 Ala-del sequence:

GCGATCGCTCTGGAACGCGCTAAGTGGCCCTACAGAGGAGGCGGAGGTGGGGGATCAGCTGGCGGAGGGAGCTCAGGAGGC
GGACCAGGTGGCGGTGGTGGTGGCGCGGGTGGGTACGCACCCTATTACGAAGAGGCAGCTATGCAGCGGGAGTTGCTTCCACC
AGCTGGGCGCCGACCCGATGTTCTGTTTAAAGGCTCCAGAACCCGTATGCGCTGCGCCGGGACCCCGCACGGTCCCGCCGGG
CGGCTAGTAACCTTCTACTCTGCAGTGGGCCGCAACGGTATTTTCCGCAAGGTTTGTATCAGTTTACGAAGCAGCGCCCGGACCT
CCATTTGCGGGCCCCGCAACCGCCACCACCTCCTGCTCCTCCACAACCAGAAGGTGCCGCTGACAAGGGTGACCCTAGGACAGGT
GCTGGTGGCGGCGGTGGCTCACCTTGTACGAAGGCAACGCCCGGTTCTGAGCCCAAAGGTGCAGCGGAAGGTAGCGGCGGAGA
CGGTGAAGGGCCACCTGGTGGAGCGGGAGCCGAGAAGAGTTCTTACGCGGTAGCTCCACAACGGTCTCGCAAGAAGCGGTGCC
ATATACCAAGTACTGACGTACG

TBX1 wt sequence:

GCGATCGCTGACCCGGAAGACTGGCCCTCGAAACACCGGCCAGGAGCACTCCCGCTGATGTCCGCTTTTCGCGCGGTCTAGGA
ATCCCGTAGCTTCCCCTACGCAACCTTCCGGCACCGAGAAAGATGCAGCAGAGGCACGCCGGGAGTTCCAGCGGGACGCAGGA
GGGCCCCGCTGTTCTCGGAGATCCGGCTCATCCGCCACAACCTTCTGCTAGGGTATTGTCTCCTAGTTTGCCTGGGGCAGGAGGTG
CTGGCGGCTGGTGCCTCTCCAGGAGCCCCAGGTGGGAGACCATCCCCCCAAATCCAGAGCTGCGATTGGAAGCGCCCGG
GCTTCTGAACCTTTGCACCACCATCCATACAAATACCCAGCTGCTGCATATGATCACTACCTTGGGGCAAAAAGTAGACCGGCACC
GTACCCCTGCCCCGACTGCGGGGGCACGGATACCATCCTCACGCGACCCACATCATCATCACCATCCCGTTAGTCCGGCAGC
GGCCGCTGCAGCTGCTGCGGCTGCCGACGCCGCTGCTGCAACATGTACAGCAGCGCTGGAGCAGCTCCTCCCGGTAGTTATGA
TTACTGCCCCAGATGACGTACG

TBX1 Ala-del sequence:

GCGATCGCTGACCCTGAAGACTGGCCGAGGAACCATAGGCCAGGCGCACTCCCCCTTATGAGTGCCTTCGCTCGATCCAGAA
ATCCGTCGCAAGTCCAACACAACCATCTGGAACCGAGAAAGATGCAGCTGAAGCACGACGGGAGTTCCAGCGAGATGCCGGG
GGACCTGCAGTACTGGGTGATCCCGCACATCCTCCCCAACTTTTGGCAGAGTATTGAGTCCCTCCCTCTGCTGGTGGAGGAG
CTGGCGGTCTCGTACCCTTCCAGGAGCGCCGGGTGGTCCGCCATCCCCGCCTAACCCAGAATCCGACTTGAGGCTCCAGGC
GCTTCCGAACCGCTGCATCATCATCCGTACAAGTACCTTACGATCATTACCTGGGCGCTAAGAGCAGACCGGCTCCGTACCCTC
TGCCGGGCTCAGAGGACATGGGTATCACCTCACGCGACCCGACACCACATCACCATCCTGTGTACCCGAACATGTATTCCA
GCGCGGGGCGGCTCCTCCTGGTAGCTACGATTACTGCTTAGATGACGTACG

EOMES wt sequence:

GCGATCGCTAACCTGCCTGGCGCGCACTTCTACCCGCTGGAGAGCGCTCGCGGAGGTAGTGGGGGCTCTGCGGGCCACCTT
CCCTCCGCTGCGCCATCACCGCAAAAACCTTGACCTCGATAAGGCTAGTAAAAATTTTCAGGCTCTCTGTATGTGAAGCGGTGAG
CGGGGAACCGGCGGCGAGCCTCCGCTGGTGCACACGCGCGATGTTGTAGATACGGATGCTGGCGACGCTTTTGCAGTGTCTG
CTGCTGTAGCAAAACCTGGCCCTCCTGATGGGAGAAAAGGCTCTCCATGCGGAGAGGAGGAACCTTCCAAGTGCCGAGCTGCAG
CAGCTGCAGCGGCGAGCAGCAGCGGCTGCCACGGCACGGTATTCTATGGATTCACTTTCCTCTGAAAGATACTACCTCCAGTCTCC
TGGCCACAAAGGTCTGAGCTCGCCGCCCTTGTTCATTGTTCCCTACCAGGCAGCCGAGGCGCTCCACACGGCCCCGGTTTA
TCCTGCTCCCAATGGTGCAAGATACCCATATGGGAGTATGTTGCCACCTGGAGGCTTCCCGGCGGCGCTGTGCCACAGGGCG
GGCTCAGTTTGGACCTGGAGCAGGTGCTGGTAGTGGGGCTGGTGGTTCTCAGGAGGGGGTGGCGGTCTGCTGACTTACCAATA
CAGCCAAGGAGCCCCGCTATGCGCCGTACCTGGAGCCGACGCGCGGGATCTTGCTGACGTACG

EOMES Ala-del sequence:

GCGATCGCTAACCTGCCCGGTGCACATTTTATCCTCTTGAATCTGCACGGGGGGGCGAGCGGTGGCTCTGCTGGCCACTTGCC
GTCTGCGGCTCCTAGTCCCCAAAGCTCGATCTTGACAAAGCTAGTAAGAAGTTTTCAGGAAGTCTCAGTTGCGAAGCTGTTCCG
GGGAGCCTTCAGCTGGTGCCCCGCGGCTATGTTGAGTGACACCGATGCAGGTGACGCTTTCGCTAGCGTGGCAAAGCCGGGC
CCCCCGATGGGCGGAAGGGGAGCCCTTGGCGGTGAGGAGGAACCTCCGTCTACGGCTCGGTACAGTATGGATTCCCTGAGCAG
CGAGAGATACTACCTTCAGAGTCCGGGTCCGCAGGGGAGTGAGTTGGCAGCCCCCTGTTCTTGTTCATATCAAGGAGCGCC
CCATGGTCTGTGTACCCTGCCCCCAACGGAGCCCGGTACCCCTACGGGTCAATGCTTCCGCCTGGCGGATTTCTGCGCGGT
ATGTCTCCTGCTGAGAGCCCAATTTCGGCCCTGGCGCTGGGGCGGGTAGTGGGGCGGGCGGATCTTCAGGTGGCGGAGGGGGT
CAGGAACCTATCAATATTCCCAAGGGGCGCCACTGTACGGTCCCTACCCTGGGGCTGCCGCGAGCTGGTAGTTGCTGACGTACG

BHLHE41 wt sequence:

GCGATCGCTTGCGTGCCCGTCATCCAGCGGACTCAGCCCTCCGCAGAGCTGGCCGCTGAAAATGATACTGATACCGACTCCG
GCTACGGAGGTGAAGCTGAGGCAAGGCCGACCGGGAGAAAGGAAAAGGCGCGGGTGTTCCTCCGCGTCACAATAAAACAGGAG
CCCCCAGGAGAAGACTCACCTGCACCTAAGAGGATGAAGCTGGATTCCAGGGGAGGAGGTTCTGGAGGCGGACCCGAGGCG
GCCTGCTGGGCCAGACCTTTGCTCAGGCCTGATGCGGCGCTGTTGTCCAGTCTTGTGCGATTTGGAGGCGGGGGAGGAGCA
CCATTCCCACAGCCACCGTTCTGTCTCCCTTTTTGTTTTCTGTACCTAGTTATGTTCAACCTTTTCTTGATAAGAGCGGACTCGAA
AAGTATCTGTATCCACCGTTCCCACTTCTGTATCCTGGAATCCCCGCGCCCTTCCATGCCTGAGCAGTGTATTGTCTCCGCCCC
CGAGAAAGCTGGCACCTGTTGCCGATGAGGTGGCTCCGCTGGGAGCACCTCACCCACAACATCCGCACGGTAGGACTCACT
TGCTTTTCGCTGGCCCCCGGGAACCTGGGAACCCCGAGTCTAGCGCTCAAGAGGATCCAGCCAACCCGGCAAGGAGGCTCCC
TGACGTACG

BHLHE41 Ala-del sequence:

GCGATCGCTTGCGTGCCCGTCATCCAGCGGACTCAGCCCTCCGCAGAGCTGGCCGCTGAAAATGATACTGATACCGACTCCG
GCTACGGAGGTGAAGCTGAGGCAAGGCCGACCGGGAGAAAGGAAAAGGCGCGGGTGTTCCTCCGCGTCACAATAAAACAGGAG
CCCCCAGGAGAAGACTCACCTGCACCTAAGAGGATGAAGCTGGATTCCAGGGGAGGAGGTTCTGGAGGCGGACCCGAGGCG
GCCTGCTGGGCCAGACCTTTGCTCAGGCCTGATGCGGCGCTGTTGTCCAGTCTTGTGCGATTTGGAGGCGGGGGAGGAGCA
CCATTCCCACAGCCACCGTTCTGTCTCCCTTTTTGTTTTCTGTACCTAGTTATGTTCAACCTTTTCTTGATAAGAGCGGACTCGAA
AAGTATCTGTATCCACCGTTCCCACTTCTGTATCCTGGAATCCCCGCGCCCTTCCATGCCTGAGCAGTGTATTGTCTCCGCCCC
CGAGAAAGCTGGCACCTGTTGCCGATGAGGTGGCTCCGCTGGGAGCACCTCACCCACAACATCCGCACGGTAGGACTCACT
TGCTTTTCGCTGGCCCCCGGGAACCTGGGAACCCCGAGTCTAGCGCTCAAGAGGATCCAGCCAACCCGGCAAGGAGGCTCCC
TGACGTACG

HOXA13 –58A (Ala-del) sequence:

GCGATCGCTGAGTTGAACAAAAACATGGAGGGGGGAGCCGGTGGAGGAGGTTTTCCCATCCGGGGGGGAATTTAGTGTTAA
CCAGTGCCGCAACCTGATGGCGCATCCAGCGCCTCTTGCGCCGGGGGCCGCATCAGCGTACAGTTCTGCTCCTGGGGAAGCAC
CCCCGTCCAGCAGTTCTGGCGGACCAGGGCCCGCAGGCCCTGCTGGTGCTGAGGCGGCGAAACAGTGCAGCCCCCTGTTCTCAG
AGCTCATCCGGTCCCGCCGCCCTCCCATACGGATACTTTGGCTCAGGGTACTACCCGTGCGCGAGAATGGGCCCTCACCCCAAC
GCCATAAAGTCATGTGCCAACCCGCGTCATTTGCCGATAAGTACATGGACACTGCTGCCCCCGCGCGGAGGAGTTAGTAGCA
GGGCGAAGGAGTGACGTACG

HOXA13 –44A sequence:

GCGATCGCTGAGTTGAACAAAAACATGGAGGGGGGGGCTGGAGGAGGTGGCTTCCCGCATCCTGCTGCTGCGGCAGCTGGGG
GGAATTTTTAGTCAATCAGTGCAGAAACCTGATGGCGCACCCCTGCCCACTTGACCCGGGAGCTGCATCAGCGTACAGCTCAGC
GCCGGGAGAAGCTCCACCGAGCTCATCTAGTGGTGGACCCGGACAGCGGTCCCGCTGGGGCCGAGGCTGCGAAACAATGTT
CCCCCTGTTCTGCAGCCGCACAGTCTCTTCAGGCCCTGCGGCATTGCCATATGGATACTTTGGGTGAGGCTATTATCCCTGCGCC
CGCATGGGTCTCATCAAATGCGATCAAGAGCTGCGCGCAGCCTGCTAGTGCCGCGGCAGCGGCAGCCTTTGCTGATAAATACA
TGATACCGCGGGGCCCGCAGCAGAAGAATTTAGCAGCCGAGCTAAGGAGTGACGTACG

MXN1 wt sequence:

GCGATCGCTGCGGTTGATCCTCCTAGAGCTGCGAGTGCTCAGAGTGCCCCCTTGGCCTTGTTACGAGTCTGGCAGCTGCTGC
TAGCGGCACGGGTGGGGGAGGCGGCGGTGGGGGAGCTAGTGGGGGACATCAGGCTCATGCTCTCCAGCTTCTCCGAGCCG
CCTGCCGCTCCCGCCGATAGGCTCCGAGCAGAGAGTCCATCACCGCCTCGGCTGTTGGCAGCCCCACTGTGCTTTGCTTCCAAA
CCAGGGTTTTTGGGCGCTGGTGGCGGAGGGGGAGGAACGGGAGGTGGTCACGGCGGCCCTCATCACCATGCTCATCCGGGTGC
CGCTGCGGCTGCTGCGGCAGCAGCTGCGGCGGCAGCCGAGGTGGTCTTGCACTCGGTCTTACCCGGGCGGAGCACAAGGT
GGCGCGGACTTCTGCACAAGCAGCTCTGTATGGGCACCCAGTCTATGGGTATAGTGCGGCGGCTGCGGCAGCCGCCCTTGC
AGGGCAACACCCAGCCTTGCTTATTCTTATCCTCAGGTACAGGGGGCGCATCCAGCACACCCGGCGGACCCGATAAAGTGACG
TACG

MXN1 Ala-del sequence:

GCGATCGCTGCGGTGGACCCCCACGAGCCGCTCAGCGCAATCTGCGCCATTGGCTCTGGTCACTTCATTGTCAGGAACAG
GAGGTGGGGGTGGAGGAGGAGTGCGTCAGGCGGAACGAGCGGGAGCTGTTACCCGGCCTCATCCGAACCTCCGGCGGCACC
CGCCGACAGACTTCGGGCTGAAAGCCCGTCCCCCCCCACGGCTCCTGGCTGCGCATTGTGCTCTGTTGCCTAAACCGGGATTCT
GGGTGCAGGTGGTGGGGGAGGGGTACAGGCGGAGGTACGGCGGCCACACCATCATGCTCACCCTGGAGGCGGGTTGGCG
CTCGGCCTCCATCCCGGGGTGCCAAGGTGGGGCTGGATTGCCAGCACAAGCGGCTCTGTATGGACATCCGGTATATGGCTAT
TCTCTTGCTGGACAGCACCTGCTCTGTCTATAGTTACCTCAGGTACAGGGGGCCCATCCAGCGCATCCCGCAGACCCTATCA
AG TGACGTACG

HEY2 Ala-del sequence:

GCGATCGCTCATCTCAGCACTTGCGCCACCCAGCGGGAGATGACATCCTCCATGGCCCACCACCATCATCCGCTCCACCCGCA
TCACTGGTTCCACCACCTGCCCGCAGCCCTGCTCCAGCCCAACGGCCTCCATGCCTCAGAGTCAACCCCTTGTGCGCTCTCCAC
AACTTCAGAAAGTGCTCCTGCCACGGCTCTGCTCTCCTCACGGCCACGTTTGCCCATGCGGATTACGCCCTCCGAATGCCATC
CACGGGCAGCGTCGCCCCCTGCGTGCCACCTCTCTCCACCTCTCTTGTCCCTCTCTGCCACCGTCCACACCCACAGCTTCCC
TCTGTCTTCCGCGGGGCGATTCCCCATGCTTCCCCCAAACGTGACAGCCATCAGCCCGCCCTTGTGAGTATCAGCCACGTCCAGT
CCTCAGCAGACCAGCAGTGGAACAAACAATAAACCTTACCGACCTGGGGGACAGAAGTTGGAGCTTTTAAACGTACG

OLIG2 Ala-del sequence:

GCGATCGCTTTCCATCCCTCTGCTTGGCGGAGGTCTGGCGCACAGCGCACCTTTGCCCGCTGCCACGGCACATCCACATGCGG
CACATCATCCGGCCGTGCACCATCCTATACTCCCGCCCGTATCTAGCGCATCATTGCCGGGGAGTGGCCTCCCTAGTGAGGTTT
CATCAGACCACCACATGGACTTCTTAAATCCCCGTCCCCCTTGGTGGTGGTGGGGGTGGATCTGGAGCGTCCGGAGGCTTTCAA
CATTGGGGCGGAATGCCATGCCGTGTTCCATGTGCCAAGTCCCCCTCCACATCACCACGTCTCTGCAATGGGTGCTGGGTAC
TCCCTAGGCTCACTAGCGACGCCAAGTAGCGTACG

For LacO-LacI tethering experiments (Figures 4E and 4F), a vector containing CFP-LacI followed by a multiple cloning site (MCS) was prepared from CFP-LacI-MED1-IDR (Zamudio et al., 2019) by removing MED1-IDR with BamHI + EcoRI digestion. Next, a multiple cloning site (MCS) was introduced with annealed, BamHI + EcoRI digested oligonucleotides (below), and AsiSI/BsiWI sites were used to subclone HOXD13 IDR sequences from previously described vectors.

MCS forward primer:

GATCGGATCCGCGATCGCTAGCACCGGTCTGACGtctagaggggccGAATTCGATC

MCS reverse primer:

GATCGAATTCGGGCCCTCTAGACGTACGACCGGTGCTAGCGATCGCGGATCCGATC

All constructs were sequence verified. The plasmids used in this study are listed in Table S1.

Isolation of limb bud cells

Limb buds from E12.5 wild-type and *spdh* homozygous embryos were micro-dissected individually and digested with Trypsin-EDTA 0.05% (GIBCO) for 15 minutes at 37°C and gently dissociated by pipetting after 5, 10, and 15 minutes. Cells were mixed in cell culture media (DMEM, 10%FBS, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin) and a single-cell suspension was obtained using a

40µm cell strainer (Falcon). About 150,000 limb bud cells from each embryo were seeded in 1000µl medium onto fibronectin-coated glass coverslips in 12 well plates. After 30-60 minutes, additional cell culture medium was added, and cells were grown for 24 hours. After 24 hours, the cells were rinsed twice with PBS and fixed 15 minutes at room temperature with 4% PFA/PBS.

Mouse genotyping

For genotyping, DNA was extracted from embryos and animals for strain maintenance using QuickExtract (Lucigen, QE09050), and PCR amplified using SPDH_F (ACTTGGGACATGGATGGGCT) and SPDH_R (CGCTCAGAGGTCCCTGGGTA) primers. PCR mix contained 1µl sample and 4% DMSO in 25µl reaction volume [95°C 3min, (95°C 20sec, 51°C 20sec 72°C 30sec)x35, 72°C 7min].

Whole mount *in situ* hybridization (WISH)

WISH was performed as previously described (Kuss et al., 2009; Villavicencio-Lorini et al., 2010). In brief, WISH was performed using a digoxigenin-labeled antisense riboprobe for *Hoxd13* transcribed from cloned gene-specific probes with PCR DIG Probe Synthesis Kit (Roche). Embryos were fixed overnight in 4% paraformaldehyde (PFA) in PBS, washed with 0.1% Tween in PBS, and dehydrated stepwise in 25%, 50% and 75% methanol/PBST and stored at –20°C in 100% methanol. For WISH, embryos were rehydrated on ice in reverse methanol/PBST steps, washed in PBS-Tween, bleached in 6% H₂O₂ in PBST for 1 h and washed in PBS-Tween. Embryos were then proteinase K-treated (10 µg/ml in PBS-Tween) for 3 min and the reaction was quenched with glycine in PBS-Tween. Embryos were washed in PBS-Tween and finally re-fixed for 20 min with 4% PFA in PBS, 0.2% glutaraldehyde and 0.1% Tween. After further washing steps with PBS-Tween, embryos were incubated at 68°C in L1 buffer [50% deionized formamide, 5X SSC, 1% SDS, 0.1% Tween-20 in diethyl pyrocarbonate (DEPC); pH 4.5] for 10 min.

For pre-hybridization, embryos were incubated for 2h at 68°C in hybridization buffer 1 (L1 with 0.1% tRNA and 0.05% heparin). For subsequent probe hybridization embryos were incubated overnight at 68°C in hybridization buffer 2 (hybridization buffer 1 with 0.1% tRNA, 0.05% heparin and 1:500 digoxigenin (DIG) probe). Unbound probe was removed through repeated washing steps; 3 × 30 min at 68°C with L1, L2 (50% deionized formamide, 2 × SSC pH 4.5, 0.1% Tween 20 in DEPC, pH 4.5) and L3 (2 × SSC pH 4.5, 0.1% Tween-20 in DEPC, pH 4.5).

For signal detection, embryos were treated for 1 h with RNase solution (0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween 20, 100 µg/ml RNase A in H₂O), followed by washing in TBST 1 (140mM NaCl, 2.7mM KCl, 25mM Tris-HCl, 1% Tween 20; pH 7.5), and blocking (2h at room temperature in blocking solution [TBST 1 with 2% calf-serum and 0.2% bovine serum albumin (BSA)]). Overnight incubation with Anti-Dig antibody conjugated to alkaline phosphatase (1:5,000) at 4°C (Roche, catalog no. 11093274910) was followed by 8x30 min washing steps at room temperature with TBST 2 (TBST with 0.1% Tween 20, and 0.05% levamisole/tetramisole), and samples were left overnight at 4°C. Finally, embryos were stained after equilibration in alkaline phosphate buffer (0.02 M NaCl, 0.05 M MgCl₂, 0.1% Tween 20, 0.1 M Tris-HCl, and 0.05% levamisole/tetramisole in H₂O) 3X20 min, followed by staining with BM Purple AP Substrate (Roche). The Stained embryos were imaged using a Zeiss Discovery V12 microscope and Leica DFC420 digital camera.

Cell treatments

Transfection: For transient transfection and live cell imaging (Figures 1F, 1G, 2C–2F, 4E, 4F, 6B–6D, 6K–6M, 6T–6V, 7E–7G, S1F, S2A–S2E, S6B–S6D, S6F–S6H, S6M, S6N, S7F, and S7G) HEK293T cells or U2OS-2-6-3 cells were seeded onto chambered coverslips (Ibidi, 80826-90), and transiently transfected 20-24 h later using lipofectamine 2000 (Signagen, SL100668) or FuGENE HD (Promega) according to the manufacturer's protocols. Each transfection series was repeated at least twice.

For 1,6-hexanediol (1,6-HD) treatment, isolated limb bud cells attached to fibronectin coated glass coverslips were treated with 6% or 0% 1,6-hexanediol (Sigma, 240117) in 1 mL cell culture media for 1 minute at 37°C. After treatment, cells were washed with PBS, fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) for 10 minutes, and stored at 4°C until processing for immunofluorescence and microscopy. Limb bud cells isolated from two replicates (i.e., two pups per genotype) were used for 1,6-HD treatment.

Western blot

Cells were washed once with ice-cold PBS, and lysed in RIPA buffer (Thermo, 88900) supplemented with protease inhibitors (Thermo, 87786). Protein concentration was determined using bicinchoninic acid assay (Thermo, 23225), according to the manufacturer's protocol. Equal protein amounts were brought to same volumes with lysis buffer, reducing agent and LDS loading buffer. Lysates were then heated to 95°C for 10 minutes, and separated on 4%–12% Tris-acetate gels (Invitrogen, NP0322BOX). Protein was transferred onto a nitrocellulose membrane in an iBlot 2 apparatus. After transfer, the membrane was blocked with Li-Cor blocking buffer (Li-Cor, 927-500000) for 1 hour at room temperature with shaking. The membranes were incubated with 1:750 anti-HOXD13 (abcam ab229234) and 1:3000 anti-HSP90 (BD Biosciences, 610419) antibody diluted in 5% non-fat milk in TBST overnight at 4°C with shaking. The next day, membranes were washed three times with TBST for 5 minutes at room temperature, incubated with fluorescent anti-mouse (IRDye 800CW Donkey anti-Mouse, Li-Cor P/N 925-32212) and anti-rabbit (IRDye 680LT Donkey anti-Rabbit, Li-Cor P/N 925-68023) secondary antibodies at 1:10000 dilution according to the manufacturer's protocol, and washed three times in TBS-T for 5 minutes in the dark. Membranes were imaged on a LICOR Odyssey CLx imager.

Immunofluorescence and confocal microscopy

Isolated limb bud cells, Kelly cells, and SH-SY5Y cells attached to coated glass coverslips were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) in PBS for 10 min and stored at 4°C in PBS or processed immediately. After two washes in PBS, cells were permeabilized in 0.1% Triton X-100 (Thermo Scientific, 85111) in PBS for 10 min at room temperature (RT). Following three washes with PBS, cells were blocked with the blocking solution (10% fetal bovine serum (FBS), 0.1% Triton X-100 in PBS) for 30 min. Cells were then incubated with the primary antibody [anti-HOXD13 Invitrogen PA5-66661 1:250 dilution (Figure 4C) or anti-HOXD13 abcam ab19866 1:150 dilution (Figure 1B)] in blocking solution at 4°C overnight. After three 10 min washes in PBS, cells were incubated with the secondary antibody (donkey anti-Rabbit IgG Alexa Fluor 568 Invitrogen A10042 1:1000 dilution) in blocking solution for 1 h at room temperature in the dark. Cells were washed three times in PBS and nuclei were counterstained with 0.24mg/mL 4',6-diamidino-2-phenylindole (DAPI) in PBS for 3 min at RT in the dark. Following five washes in PBS, coverslips were mounted onto slides with Vectashield (Vector, H-1000) and sealed using transparent nail polish. Images were acquired on a confocal laser-scanning microscope (Zeiss LSM880, 63x oil objective, NA 1.4, 1 Airy Unit). Raw images (.czi files) were processed in FIJI (Schindelin et al., 2012). Expression of HOXD13 protein in nuclei of limb bud cells, Kelly cells, and SH-SY5Y cells lines was examined with three HOXD13 antibodies, as displayed in Figures 1B and S1A–S1C.

Stochastic optical reconstruction microscopy (STORM)

Isolated wild-type and homozygous *spdh* E12.5 limb bud cells attached to fibronectin-coated glass coverslips were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) for 20 min at room temperature. After 2 x washes in PBS, fixed samples were treated with permeabilization solution (PBS supplemented with 0.1% Triton X-100) for 10 minutes at room temperature. Samples were then treated with blocking solution (permeabilization solution supplemented with 10% fetal bovine serum) for 30 min. After blocking, samples were incubated with primary anti-HOXD13 antibody (Thermo, PA5-66661, 1:250 in blocking solution), BRD4 antibody (Clone A-7, Santa Cruz, sc-518021, 1:250 in blocking solution), or HP1 α antibody (clone15.19s2, Millipore/Merck, 1:250 in blocking solution) overnight at 4°C, and then washed 3x in PBS. Stained and washed samples were then incubated with secondary antibody (goat anti-Rabbit IgG Alexa Fluor 647 Invitrogen A32733 / goat anti-Rabbit IgG Cy3 Invitrogen A21235, 1:1000 in blocking buffer for HOXD13 primary and goat anti-Mouse IgG Alexa Fluor 647 Invitrogen A10520, 1:1000 in blocking buffer for BRD4 / HP1 α primary) for 45 minutes, after which samples were washed 3 x in PBS. For imaging, samples were placed in a one-well magnetic chamber, covered in switching buffer consisting of 0.15 M 2-mercaptoethanol/0.2 M Tris, pH 8.0 with 4% (w/v) glucose, 0.25 mg/ml glucose-oxidase, 20 μ g/ml catalase. For nuclei visualization (Figure 1C), switching buffer was supplemented with 5 nM Sytox Orange (Thermo, S11368). The BRD4 antibody was validated in-house using dBET6-induced degradation of endogenous BRD proteins. Images were acquired as described in (Fabricius et al., 2018) with a Vutara 352 super resolution microscope (Bruker) equipped with a Hamamatsu ORCA Flash4.0 sCMOS for super resolution imaging and a 60x oil immersion TIRF objective with numerical aperture 1.49 (Olympus). Data were acquired with TIRF/HILO-illumination at a laser-power density of 62.5 kW/cm² using an 532 and 639 nm laser. Images for co-IF STORM analyses were acquired from three biological replicates (three limb bud samples per genotype, harvested from three independent mouse matings).

Live cell imaging

Live cell imaging experiments were performed on an LSM880 confocal microscope (Zeiss) equipped with an incubation chamber and a heated stage at 37°C. Images were acquired with either a Plan-Apochromat 40x0.95 Korr M27 or a 63x1.40 oil DIC objective.

OptoDroplet assay (Figures 1F, 1G, 2C, 2D, 6B, 6C, 6K, 6L, 6T, 6U, 7E, 7F, S6D, and S6H): The optoIDR assay was adapted from Shin et al. (2017). Approximately 20,000 cells were seeded per well in chambered coverslips (Ibidi, 80826-90) one day before transfection. The following day, cells were transfected with 200ng of plasmid encoding respective optoIDR constructs. Cell culture medium was refreshed 48 hours post-transfection, and cells were imaged on a Zeiss LSM 880 confocal microscope. Droplet formation was induced using scans with the 488 nm laser every 20 s for the duration of imaging (Figures 1F, 1G, 2C, 2D, 6B, 6C, 6K, 6L, 6T, 6U, 7E, 7F, S6D, and S6H). For image acquisition, mCherry fluorescence was stimulated using 561nm laser scans every 20 s. The constructs used for the optoDroplet experiments described in Figures 1F–1I and S1D–S1F did not include the SV40 NLS sequence, and the fusion protein displayed cytosolic localization. All other data was generated using constructs that included the SV40 NLS, and the fusion proteins displayed nuclear localization (verified e.g., in Figure S2B).

For FRAP experiments of light induced droplets (Figures 1H, 2E, 6D, 6M, 6V, and S1F) droplet formation was induced with continuous 488 nm light for 90 s. Droplets were then bleached with 561 nm light, and fluorescence recovery was imaged every 4 s in the presence of simultaneous 488 nm light stimulation. Bleaching was performed on a region enclosing a single droplet using 2 iterations of 100% laser power. For RUNX2 and RUNX2+10A (Figure 6M), bleaching was on a region enclosing a single droplet using a single iteration of 100% laser power. For FRAP experiments of spontaneous HOXD13 IDR condensates (Figure 2F), condensates were bleached with 561 nm light and recovery was imaged every 12 s.

Nuclei in live cells were visualized with the addition of Hoechst 33342 for 5min at 37°C in the imaging medium (Figure S2B).

To test the reversibility of optoIDR droplet formation (Figure S2C) cells were stimulated with continuous 488 nm laser scanning for 30 s. mCherry fluorescence was detected in the absence of 488 nm stimulation for 15 minutes at the time intervals displayed in Figure S2C.

For exclusion microscopy (Figures S2D and S2E), approximately 20,000 cells were seeded per well on chambered coverslips one day before co-transfection with optoIDR and IDR-YFP constructs. For all co-transfections, the total amount of transfected DNA was kept at a constant 300 ng/well. 48 hours post-transfection, culture media was refreshed, and cells were imaged on a Zeiss LSM 880 confocal microscope. For dual color image acquisition, mCherry fluorescence was stimulated at 561 nm, and YFP fluorescence was stimulated at 513 nm. Images for exclusion microscopy were acquired across two experimental replicates.

All experiments using optoIDR constructs were repeated at least two times.

LacO-LacI assays (Figures 4E and 4F): Tethering experiments were adapted from Chong et al. (2018) and Zamudio et al. (2019). Briefly, 20,000 U2OS cells were seeded on chambered coverslips one day before co-transfecting the cells with 100 ng of CFP-LacI-HOXD13-IDR plasmid and 100 ng of MED1-IDR-YFP-NLS plasmid with FuGENE HD reagent (Promega). Imaging was performed on live cells 48–80h after transfections. For the HOXD13 WT, +7A, and +10A tether series, images were acquired across three experimental replicates, and for the control (MED1 tether with MED1-YFP) images were acquired across two experimental replicates.

Protein purification and *in vitro* droplet formation assays

For protein expression, plasmids were transformed into BL21(DE3) (NEB M0491S) cells, and grown in home-made automatic-induction medium (AIM), as described in Studier (2005). In brief, 10 ml of overnight culture in minimal media (MDG) was used to seed 100 ml of AIM (ZYM-5052) supplemented with ampicillin and trace metals. This culture was incubated at 37°C with vigorous shaking for 2 hours, after which the temperature was brought down to 16°C. Cells were harvested when the cultures became dark magenta or bright green (24–48 hours later), after which pellets were frozen at –80°C for at least 16 hours. For protein purification, pellets were resuspended in 30 mL of Buffer A (50 mM Tris pH 7.5, 300 mM NaCl) supplemented with 8 M urea (Sigma, U5378) and cComplete protease inhibitors (Sigma, 11697498001). This denatured suspension was sonicated and clarified by centrifugation at 20,000 g for 20 minutes at 4°C. Supernatants containing fusion proteins were loaded onto a His GraviTrap column (GE HealthCare, 11003399) pre-equilibrated in Buffer A supplemented with 8 M urea. The loaded column was washed with 30 column volumes (CV) of 6% Buffer B (50 mM Tris pH 7.5, 300 mM NaCl, 500 mM Imidazole) in Buffer A supplemented with 8 M urea, and another 20 CV of 6% Buffer B to remove denaturant. Proteins were eluted in 3 CV of 50% Buffer B, immediately diluted 1:2 in storage buffer (50 mM Tris pH 7.5, 125 mM NaCl, 1 mM DTT, 10% Glycerol), and then concentrated using 3000 MWCO centrifugal filters (Merck, UFC803008) by spinning at 7500 g for 30 minutes at 4°C. The resulting fraction was then diluted 1:100 in storage buffer, re-concentrated, and then stored at –80°C. All fusion proteins were purified in this same manner. For *in vitro* droplet formation experiments (Figures 1J, 2G, 2H, 3A–3J, 6E–6H, 6N–6Q, and S6K), recombinant mCherry or mEGFP fusion proteins were measured for concentration, and then diluted or mixed to desired concentrations in storage buffer. For microscopy, these solutions were further diluted 1:2 in either 20% or 10% PEG-8000 in de-ionized water (w/v). 10 μ l of this suspension was pipetted onto chambered coverslips (Ibidi, 80826-90), and immediately imaged using a LSM880 confocal microscope equipped with a 63x1.40 oil DIC objective and AiryScan detector. All images were acquired from within the solution interface, and performed before droplets settled onto the bottom of the coverslip, as described in Sabari et al. (2018) and Boija et al. (2018). For droplet assays using preformed GFP-MED1 IDR condensates (Figures 3E and 3F), GFP-MED1 IDR droplets were allowed to form for 30 minutes at room temperature in the presence of 10% PEG-8000 in protein Lo-bind tubes (Eppendorf, 30108094), before proceeding with co-condensation assays. For compound treatments of co-condensates (Figures 3G–3J), small molecules [ATP (Jena Bioscience, NU-1010), Mitoxantrone dihydrochloride (Sigma M6545), (\pm)- α -Lipoic acid (Sigma 62320), and Lipoamide (Medchemexpress HY-B1142)] were directly diluted into droplet mixtures from appropriate stocks with or without vehicle (DMSO, Sigma D2650) to the desired final concentrations, and immediately imaged using confocal microscopy. Droplet formation experiments (Figures 1J, 2G, 2H, 3I, 3J, 6E–6H, 6N–6Q, and S6K) were repeated at least two times. Experiments for Figures 3A–3C were repeated seven times.

Luciferase reporter assays

The murine proximal enhancer of *Aldh1a2* (Kuss et al., 2009) was cloned into the pGL2-enhancer luciferase reporter vector and co-transfected into Cos7 cells with pcDNA3.1 (wild-type or mutant *Hoxd13*) expression plasmids and pRL-CMV (Promega). For transfection, 8×10^4 Cos7 cells were cultured per well in 24-well plates and transfection of 455ng DNA per well (250ng pGL2-*Aldh1a2*, 200ng pcDNA3.1-*Hoxd13*, 5ng pRL-CMV) was carried out using FuGENE HD (Promega) according per the manufacturer's instructions. 24 hours post transfection, reporter activity in 2.5 μ l of cell lysate was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's instructions (Figure 4G).

Measurements were performed similarly in GAL4-DBD luciferase assays (Figures 6I, 6R, 7H, 7I, and S7H), except GAL4UAS-Luciferase reporter [Addgene 64125, described in Nihongaki et al. (2015)] was co-transfected with GAL4-IDR fusion constructs and with pRL-CMV into HEK293T cells, and 10 μ l of cell lysate was used in measurements. All experiments were performed in triplicates and repeated 2–5 times. The number of measurements was 15 for HOXA13, TBX1, GSX2 and HEY2 series, 12 for EOMES, HOXD13 and BHLHE41-Adel, 9 for MNX1-Adel, OLIG2-Adel and HOXD11-Adel, and 6 for RUNX2, MNX1wt, OLIG2wt, BHLHE41wt and HOXD11wt.

H3K27Ac ChIP-seq with S2 spike-in control (ChIP-Rx)

Isolated E12.5 limb bud cells (hand plate) from 2 wild-type, 2 homozygous *spdh*, and 2 heterozygous *spdh* mice (approximately 3.5 million cells each) across 2 separate het x het mouse crosses were fixed in 10 minutes in 1% formaldehyde on ice and quenched with

2.5M glycine. Each pool of 3.5 million cells (wt, homozygous *spdh*, and heterozygous *spdh*) was spiked in with 1 million fixed fly S2 cells (Orlando et al., 2014), after which cell suspensions were lysed in lysis buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% Glycerin, 0.5% NP-40, 0.25 Triton X-100, pH 7.5 supplemented with protease inhibitors and Na-Butyrate). After assessing for optimal shearing conditions (47 cycles of 30 s of sonication at high settings using Diagenode Bioruptor) and total chromatin amount, 10–15 μ g of sheared chromatin was incubated with 4 μ l of H3K27Ac antibody (Diagenode C15410174) in a total of 1.2 mL of buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-DOC, 0.5% N-Laroylsarcosine, pH 8.0 supplemented with protease inhibitors and Na-Butyrate) overnight with gentle rotation at 4°C. After incubation, 30 μ l of Protein G beads were added to the chromatin and antibody suspensions, and allowed to incubate overnight with gentle rotation at 4°C. Samples were then washed 7 times with 1 mL of RIPA buffer (to 50 mM HEPES-KOH, 1 mM EDTA, 1% NP-40, 0.7% Na-DOC, 500 mM Li-Cl, pH 7.55 supplemented with protease inhibitors and Na-Butyrate). Beads were then washed with TE buffer (supplemented with protease inhibitors and Na-Butyrate) and then centrifuged down to remove TE buffer. Chromatin was then eluted using 210 μ l of elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS, pH 8.0) and heated to 65°C with vigorous shaking. Eluates were then treated with 5 μ l of Proteinase K overnight at 65°C. The next day, 4 μ l of RNase A was added to the samples, vortexed, spun down and incubated at 37°C for 30 minutes. Chromatin was then extracted using phenol-chloroform, and precipitated using 70% ice cold EtOH, with centrifugation at max speed for 10 minutes. Supernatant was removed, and the resulting pellet was dissolved in 30–50 μ l of de-ionized water. Total yield was assessed by Qubit, and then sent for sequencing. Respective ChIP samples and input controls for wild-type, homozygous *spdh*, and heterozygous *spdh* were paired-end sequenced using Illumina high-throughput sequencing, to a depth of around 25 million reads.

Flag-Hoxd13 ChIP-Seq from chicken micromass (chMM) cells

ChIP-Seq for wild-type and mutant (+7A) *Hoxd13* in chicken micromass cells (Figures S3C–S3E) were performed as described (Jerković et al., 2017) using 3xFlag tagged versions of murine *Hoxd13*. To create FLAG-tagged fusion protein for ChIP-Seq, coding sequence of mHoxd13+7A was cloned with ClaI and SpeI sites into modified RCASPB(A) vector, as described for mHoxd13wt (Ibrahim et al., 2013). RCASPB(A)-3xFLAGHoxd13 (wt or mutant) virus was mixed with mesenchymal cells from HH22-24 embryonic limb buds at the time of seeding [10 μ l drops à 2x10⁷ cells in chMM medium (DMEM:HAM's-F12 with 10% FBS, 10% CS, 1% L-glutamine and 1% Penicillin-Streptomycin)]. *Hoxd13* expressing chMM cultures were harvested after 6 days of culture by adding digestion solution (0.1% collagenase (Sigma, #C9891) and 0.1% Trypsin in 1x PBS). Cells were resuspended in 10 mL cold chMM medium and fixed for 10 min on ice with 1% formaldehyde. Extraction of nuclear lysate was performed as described for the H3K27Ac ChIP above. Chromatin was sonicated with a Diagenode Bioruptor (45 cycles—30 s pulse, 30 s pause, HI power). For ChIP, 25–35 μ g of chromatin was incubated with 6–8 μ g of mouse anti-FLAG (Sigma M2, F1804) antibody overnight. The next day 30 μ l of Protein G beads were added to the suspension and incubated overnight at 4°C. Washes and DNA extraction was performed as described above. Eluted DNA was used for library generation using NEBNext ChIP-seq Library Prep Master Mix for Illumina (New England Biolabs), fragments between 300–450bp were selected and sequenced single-end 50bp on a Illumina HiSeq 1500 using Illumina TruSeq v3 chemistry.

Single-cell RNA-seq data generation

E12.5 littermate embryos (wt and homozygous *spdh*) were obtained from staged matings of heterozygous *spdh* mice (Bruneau et al., 2001). Individual littermate embryos were separated and kept at 4°C in PBS during genotyping. DNA was extracted from embryonic forebrain tissue of each embryo using QuickExtract (Epicenter, QE09050) and used for immediate PCR genotyping. Distal limb buds (hand plate) from 1 wild-type and 1 homozygous *spdh* embryo were then dissected and trypsinized for a total of 30 minutes in 500 μ l pre-warmed 3T-trypsin and gently dissociated by pipetting after 10, 15 and 25 minutes. Trypsinization was stopped by addition of cell culture medium (DMEM, 10% FBS, 2 mM L-Glutamine, 50 U/ml Penicillin/Streptomycin). Cell dissociation was verified microscopically and cell viability was assessed with Trypan-blue staining.

The cell suspension was passed through a Flowmi strainer (40 μ m, Scienceware H13680-0040), centrifuged at 1,200 rpm at 4°C and washed with 1 mL PBS containing 0.4% FBS for a total of 3 washes. For single-cell RNA sequencing with a target cell recovery of about 5,000 cells, 8,000 cells were used as input for the generation of a droplet emulsion using the Chromium Controller. Single-cell libraries were generated following the manual instructions of 10x Genomics (Chromium Single Cell 3' v2), with the exception of fewer PCR cycles in cDNA amplification (11 cycles) and sample index PCR (10 cycles) than recommended in order to increase complexity of libraries. Libraries were sequenced with a minimum of 300 million paired end reads and parameters as described in the manual. Further details are listed in Figure S4A.

Repeat expansions in HOXD13, HOXA13, RUNX2 and TBP

The repeat expansion information displayed in Figures 2B, 6A, 6J, and 6S were curated from multiple studies, some of which were summarized in previous reviews (Albrecht and Mundlos, 2005; Darling and Uversky, 2017). The first +7A, +8 and +10 polyA repeat expansions associated with synpolydactyly were described in Muragaki et al. (1996). Additional expansions, including pedigrees with +9A and +14A mutations were described in Goodman et al. (1997). HOXA13 contains three short alanine repeats (14, 12 and 18 alanines, respectively), and hand-foot-genital syndrome (HFGS)—associated mutations in all three repeats have been described (Goodman et al., 2000; Innis et al., 2004). TBP polyQ repeat expansions were described in several early reports (Maltecca et al., 2003; Nakamura et al., 2001). The RUNX2 alanine expansions were described in (Mundlos et al., 1997).

Protein disorder analysis of HOXD13, HOXA13, RUNX2, TBP and FOXP1-4

Intrinsically disordered regions (IDRs) in HOXD13, HOXA13, RUNX2, TBP and FOXP1-4 (Figures 1D, 6A, 6J, 6S, and 7A) were predicted using the publicly available Predictor of Naturally Disordered Regions (PONDR) algorithm (VSL2) (Peng et al., 2006) as previously described (Boija et al., 2018; Sabari et al., 2018). In brief, the algorithm utilizes neural network predictors, trained on non-redundant sets of ordered and disordered sequences, to predict disordered regions within queried sequences. An amino acid in a protein was considered disordered if the VSL2 algorithm score was > 0.5 (<http://www.pondr.com/>) (Peng et al., 2006).

IDR classification: data retrieval and preprocessing

Protein sequences for all human RefSeq gene models were downloaded from UCSC (Karolchik et al., 2004) on November 9th 2018. For each protein, the PONDR (disorder) scores (VSL2) were obtained from PONDR (<http://www.pondr.com>) (Peng et al., 2006). An intrinsically disordered region (IDR) was defined as a sequence of at least 50 consecutive amino acids with a PONDR (VSL2) score above 0.5. Two IDRs in the same protein were fused if less than 40 consecutive amino acids with a score below 0.5 separated them. For the classification of TF IDRs (Figures 7A and S7A) the longest IDR from all protein isoforms of the same gene was used. The list of transcription factors and their DNA binding domain (DBD) sequences used in Figures 7A, S7A, and S7B were retrieved from a previous study (Lambert et al., 2018) (<http://humantfs.ccb.utoronto.ca/>). In particular, the DBD sequences of TFs with a homeobox, bHLH, bZIP and forkhead domain were used in Figure S7B. The amino acid composition, GRAVY (hydrophobicity) score, instability score, aliphatic index and isoelectric point (pI) of each DBD and IDR was extracted using the ProtParam webservice (Artimo et al., 2012). PhyloP conservation scores based on the 100way multiple species alignment with hg19 as the reference species were obtained from UCSC. The list of DBD and IDR sequences are listed in Table S7.

IDR/DBD clustering and annotation

The amino acid composition, aliphatic index, isoelectric point (pI), PONDR, GRAVY and instability scores for each IDR of all TFs was used as input. Principal component analysis (PCA) was performed to identify the most variable features, and the first ten PCs (that explain 80% of the variability of the input data) were subsequently used (Figure S7C). The transformed data were then subject to K-means clustering. We determined the Bayesian information criterion (BIC) for various values of “k” and identified k = 7 as an optimal cluster number as the inflection point on the BIC plot (Figure S7D) (Schwarz, 1978). In particular, we performed k-means clustering 1,000 times with different cluster starting points by setting different seed values in R and calculated BIC scores for each run. The inflection point (the optimal k for the K-means clustering) of each repetition was identified as the one with the maximum distance to a fit distribution. The majority of the 1,000 repetitions had the inflection point at k = 7. For each TF, the DBD annotation of a previous study was used (Lambert et al., 2018). The presence of a homopolymeric repeat was defined using a minimum repeat length of ten amino acids in the IDR (Figure 7A).

In Figure S7E, the features that distinguish each cluster from the other six clusters are shown. For this analysis the mean value of each feature was calculated for each cluster. The statistical significance between the distribution of the values for a feature in a cluster and the distribution of the values for a feature in the other six clusters was determined using Kolmogorov–Smirnov test. Bonferroni corrected p values were log₁₀ transformed. Red color indicates enrichment of a feature in the corresponding cluster (i.e., the mean value of the feature in the cluster is higher than the mean value of the feature in the other six clusters). Blue color indicates depletion of a feature in the corresponding cluster (i.e., the mean value of the feature in the cluster is lower than the mean value of the feature in the other six clusters). The depth of the color is scaled to the Bonferroni corrected p values.

Figures 7A, S7A, and S7B used the R packages *circize* (Gu et al., 2014) and *dendextend* (Galili, 2015) to visualize the identified IDR clusters, DBD clusters and IDR and DBD clusters. The inner circle in Figure 7A represents the identified 7 IDR clusters. The middle circle highlights the TF IDRs with a homopolymeric alanine or glutamine repeat in the IDR, and the outer circle highlights the transcription factor family a TF IDR belongs to. The FOXP1/2/3 DNA binding motifs in the outsets were obtained from the Jaspas database (Mathelier et al., 2014).

In Figure S7A we used the classification tool on the 1,446 IDR sequences and 365 DBD sequences from the homeodomain, bHLH, bZIP and forkhead TFs for which the DBD amino acid sequences were available in Lambert et al. (2018). The efficiency of the classification tool in discriminating DBDs and IDRs is calculated as the percentage of IDRs and DBDs being present in clusters 1 and 2 of this circular representation (Figure S7A). Cluster 1 and 2 contain 95% of all DBDs while 96% of IDRs are in the other clusters.

For the DBD classification in Figure S7B we used a k of 4 for the k-means clustering since DBDs that belong to 4 TF families (homeodomain, bHLH, bZIP and forkhead) were used as input. The inner circle represents the 4 clusters identified by our classifier, and the outer circle shows the TF family annotation. The transcription factor families have been recovered with 85% average efficiency. For each of the four DBD clusters, we determined the most abundant TF family in a cluster and the percentage of all of the family members being recovered in this particular cluster. Cluster 1 consists of 88% homeobox TFs, cluster 2 consists of 76% bHLH TFs, cluster 3 consists of 82% bZIP TFs and cluster 4 consists of 97% forkhead TFs.

The enrichment analyses of TFs whose IDRs comprise the 7 IDR clusters for categories displayed in Figure 7D was done as follows. The enrichment of Gene Ontology (GO) terms for biological processes was carried out by using the GOrilla webservice (Eden et al., 2009) with two unranked lists. The TFs found in a cluster were used as the target gene set and all TFs were used as the background set. A p value of 10^{−3} was kept as threshold. We curated then the most specific GO terms for the target set which correspond to the leafs of the resulting GO term tree. Only GO terms that had at least 20 genes and not more than 1,000 genes in the background gene

set were kept. In Figure 7D, only GO terms that had at least 15 genes overlap with at least one of the seven IDR clusters, and an FDR of maximum 5% were displayed. Phenotype ontology terms were retrieved from <https://hpo.jax.org> (Köhler et al., 2014) (ALL_SOURCES_ALL_FREQUENCIES_genes_to_phenotype.txt). Fisher's exact test in R was carried out for each of the TFs in the seven clusters by setting the parameter `alternative = 'greater'` and a p value cutoff of 0.05. The significant phenotype terms were then filtered in the same way as the GO terms. The GWAS associations were obtained from <https://www.ebi.ac.uk/gwas/> (Buniello et al., 2019) (gwas_catalog_v1.0-associations_e96_r2019-06-20.tsv) and we used the exact same processing as for the phenotype ontology terms. The complete set of terms and enrichment scores are listed in Table S7. The TF annotation of activators and/or repressors (Figure 7D) was obtained by using the R bioconductor package, and the GO.db was queried for terms 'activator' or 'repressor' and intersected with the term 'transcript'. This resulted in 430 activators and 231 repressors.

QUANTIFICATION AND STATISTICAL ANALYSIS

Confocal/Fluorescence image analysis

HOXD13 puncta detected in fixed cell immunofluorescence (Figures 4C and 4D) were analyzed in Zen Blue 3.0. Nuclear regions were first detected by Otsu thresholding on the DAPI counterstain signal. HOXD13 puncta within nuclear regions were detected using a fixed minimal pixel intensity threshold, segmented on morphology, and then filtered based on fluorescence intensity and circularity. After adjusting parameters on 4–5 images, the established pipeline was used to analyze all images from cell-types and treatment conditions (7 images for untreated wild-type limb bud, 7 images for 6% 1–6 hexanediol (1–6 HD) treated wild-type limb bud, 18 images for untreated spdh limb bud, 12 images 6% 1–6 HD treated spdh limb bud). Values for nuclear area, HOXD13 puncta area, HOXD13 puncta mean fluorescence intensity, and HOXD13 puncta count were measured for 120 wild-type cells, 143 1–6 HD treated wild-type cells, 63 spdh cells, and 62 1–6 HD treated spdh cells. The numbers of cells are displayed as “n” in Figures 4D, S3A, and S3B. For data visualization, phase-shift ratios, puncta count, and mean fluorescent intensities of HOXD13 puncta in each sample were plotted as indicated in figure legends (Figures 4D, S3A, and S3B). Detailed description of the phase-shift ratio calculation is found in the OptoDroplet assay analysis section below.

STORM image analysis

STORM images were localized, visualized, and analyzed using Vutara SRX software, version 6.04.14 (Bruker). For HOXD13 cluster analysis (Figure 1C), images were collected with a 50 ms acquisition time and 5,000 images were used to reconstruct the super-resolution composite with minimal background thresholding. The HOXD13 density map was generated by point splatting localized particles with radial precision. The histogram in Figure 1C shows the frequency in size distribution of HOXD13 clusters within the projected cell nucleus of an E12.5 mouse limb bud cell. To estimate HOXD13 condensate size, HOXD13 localizations were clustered using density-based spatial clustering of applications with noise (DBSCAN). DBSCAN parameters (specific distance of 100 nanometers) were set using the spatial distribution tool embedded in the Vutara SRX analysis interface and the radius of gyration for clustered HOXD13 localizations (15–50 particle clusters) was measured. The Nearest Neighbors (NeNa) localization precision (Endesfelder et al., 2014) was 9 nm for HOXD13 localizations. NeNa reports the average localization precision of a single molecule localization microscopy (SMLM) experiment. In brief, the routine is dependent on the distance distribution (NN_{adj}) of nearest neighbors in adjacent frames, and is closely related to the pairwise displacement distribution of Gaussian-distributed localizations of a single, frequently localized molecule (Endesfelder et al., 2014).

For Co-IF STORM analysis (Figures 4A and 4B), 4,000 images (2,000 for each probe channel) were collected with a 50 ms acquisition time for each cell (25 cells from wt samples, 22 cells from spdh samples, and 5 cells in negative control sample). HOXD13, BRD4 and HP1 α particles were localized with fixed minimal background thresholds on both channels to create super-resolution composites. Single cells (seen in Figure 4A) were then cropped from field of views using the selection tool embedded in the Vutara SRX visualization interface. Cells were then filtered for similar localization densities using a maximum localization density cut off, and assessed for probe overlap using the Manders overlap coefficient (MOC). The numbers of cells is indicated as “n” in Figure 4B (n = 13 for HOXD13/BRD4 co-localization in wt cells, n = 22 for HOXD13/BRD4 co-localization in spdh cells, and n = 5 for HOXD13/HP1 α co-localization in wt cells). Since the MOC calculates pixel-based overlap between 2 channel images acquired from diffraction-limited microscopy systems (Manders et al., 1993; Lagache et al., 2018), STORM composites were first converted into pixel-based images to mimic a diffraction-limited system, and then these images were used to calculate the MOC. Each cell was analyzed for probe overlap using the same conversion parameters (particle size = 150 nanometers, max radius = 500 nanometers) using the co-localization tool embedded in the Vutara SRX analysis interface. Density maps in Figure 4A were generated by point splatting localizations with size constants to better reflect the conversion parameters used for overlap analysis. For visualization of probe overlap (right most column in Figure 4A), merged density maps were processed in FIJI to isolate the pixel intersection between the two probe channels.

OptoDroplet assay analysis

Image analysis for optoDroplet assays (Figures 1F, 1G, 2C, 2D, 6B, 6C, 6K, 6L, 6T, 6U, 7E, and 7F) was performed in Zen Blue 3.0 (Zeiss), Arivis Vision4D or FIJI. For optoDroplet analysis in Zen Blue 3.0, single cells were cropped out of image field recordings of cells receiving green light stimulation. Nuclear and cytosolic mCherry signals within cell recordings were then detected by Otsu

thresholding and size filtering to define primary regions of interest (ROIs). OptoDroplets were detected within the primary ROIs using a second fixed pixel intensity threshold. All detected optoDroplets were automatically segmented using the watershed algorithm. For optoDroplet analysis in Arivis Vision4D, images were first converted to an appropriate file format in the Arivis SIS converter. Nuclear mCherry signals were automatically thresholded and size-filtered to define primary ROIs. OptoDroplets were detected as secondary ROIs that exceeded the threshold limit of the primary ROI. All parameters for both software approaches were determined empirically by adjusting the analysis pipeline on 4–5 raw image field recordings for each experimental set. The pipeline was then used to analyze at least 50–200 cells per genotype (indicated as “n” in each figure panel) with identical parameters. In Figures 1G, 2D, 6C, 6L, 6U, and 7F the phase-shifted fraction was calculated as the total area of detected optoDroplets within a primary ROI, divided by the area of the corresponding primary ROI. This fraction was then averaged over all detected cells (indicated as “n” in each figure panel) for each genotype, and plotted over time.

$$\text{phaseshifted fraction at time point } t = \frac{\sum_{i=1}^n \left\{ \frac{\text{area of optoDroplets within ROI at time } t_i}{\text{area of ROI}_i \text{ at time } t_i} \right\}}{n}$$

To control for differences in optoIDR expression levels, minimum and maximum thresholds on mean mCherry fluorescence at time point 0” were used. Mean fluorescence intensities of ROIs (cells or nuclei) at time point 0” used for phase-shift fraction calculations are displayed in Figures S1D, S2A, S6B, S6F, S6M, and S7F.

For phase-shifted fraction comparison in the HOXD13 deletion series in Figure 7E–F, 30 cells in the mCherry signal intensity range of 75 to 100 at the initial time point (i.e., pre-light induction) were selected at random for each genotype using Arivis Vision4D processing. The phase-shifted score was averaged over the 30 cells per genotype, and plotted over time. In Figure 7G, IDR amino acid composition is plotted against the mean phase-shift score at 180” of green light stimulation for each genotype (Figure 7G).

For FRAP analysis of OptoIDR constructs, mean pixel intensity of regions of interests were measured using FIJI, and normalized to pre-bleaching intensity. Captured intensity traces were averaged over multiple replicates for each genotype (indicated as “n” in figure panels), and values were plotted against time.

For exclusion microscopy analysis (Figures S2D and S2E) primary regions of interest corresponding to spontaneously formed HOXD13(+8A)-mCherry-CRY2-NLS condensates were detected based on a fixed intensity threshold. Mean intensities of these primary regions were measured on both mCherry and YFP channels, and a 4-pixel wide ring with a 1-pixel gap from the primary region was used as a reference (background) region. The mean intensity value of this reference region was used to normalize the mean intensity of the enclosed primary region. To quantify exclusion of IDR-YFP constructs, normalized signals of primary regions, in both channels, were plotted against the corresponding genotype. Each “n” in Figure S2E indicates a unique spontaneous +8A optoIDR condensate.

$$\text{Relative Signal within optoIDR condensate} = \frac{\text{mean YFP intensity within optoIDR condensate}}{\text{mean YFP intensity of background ring}}$$

LacO-LacI tethering assay analysis

For LacO-LacI image analysis (Figures 4E and 4F) primary regions of interest corresponding to IDR-LacI-CFP tethers were detected based on a fixed intensity threshold on the cyan channel using ZEN Blue 3.0. Mean intensities of these primary regions were measured on both YFP and cyan channels, and a 3-pixel wide ring with a 1-pixel gap from the primary region was used as a reference (background) region. The mean intensity value of this reference region was used to normalize the mean intensity of the enclosed primary region. To quantify partitioning of MED1 IDR-YFP constructs, normalized YFP signals at the primary regions were plotted for each co-transfection (Figure 4F). For quantification, only tethers within cells of similar expression levels of YFP were used, by gating on mean intensities of reference regions as seen in the right panel in Figure 4F. Additionally, only tethers exhibiting at least 5 fold CFP signal over background were used for enrichment analysis. Each “n” in Figure 4F indicates a unique LacI-CFP tether that abides by these gating conditions (n = 35 for HOXD13 WT + MED1, n = 42 for HOXD13 +7A + MED1, n = 42 for HOXD13 +10A + MED1, n = 16 for MED1 + MED1). For data visualization (Figure 4F), values were plotted as indicated using R-studio.

$$\text{Relative Signal within LacI tether} = \frac{\text{mean YFP intensity within LacI tether}}{\text{mean YFP intensity of background ring}}$$

In vitro condensate assay analysis

Data for in-vitro condensation experiments were acquired from 4–8 image fields for each condition. For data analyses (Figures 1J, 3A–3D, 3H–3J, 6G, 6H, 6P, 6Q, S6J, and S6K) droplet regions were first detected using a fixed minimal intensity threshold on either the mEGFP or mCherry channel using Zen Blue 3.0. For preassembled MED1 IDR-GFP scaffold experiments (Figures 3E and 3F) and TF IDR mutant phase diagrams (Figures 2G, 2H, 6E, 6F, 6N, and 6O), droplet regions were first detected by three sigma thresholding on the mCherry channel in Zen Blue. To ensure robust detection of droplets, detected regions were then gated with minimum and maximum filters on size and pixel deviation. Mean fluorescence intensity, area, and diameter of droplet regions were measured on both channels, and plotted as described in Figures 1J, 2H, 3B–3E, 3H–3J, 6F, 6H, 6O, 6Q, and S6K using R-Studio. The flushed

insets in Figure 3E were generated using the Directlabels package. Detailed breakdown of visualization approaches and the utilized code can be found at <https://github.com/BasuShaon/ChemicalBiology>.

Statistical analysis

Statistical analyses were carried out in Graphpad Prism 7 and R Studio (Figures 1G, 1H, 1J, 2D–2F, 2G, 3B–3D, 3E, 3H–3J, 4B, 4D, 4F, 6C, 6D, 6F, 6H, 6L, 6M, 6O, 6Q, 6U, 6V, 6G, S1D, S1E, S2A, S2D, S2E, S3A, S3B, S6B, S6C, S6F, S6G, S6K, S6M, S6N, and S6F). Cluster analysis and overlap analysis of STORM localizations (Figures 1C and 4B) was carried out using Vutara SRX v 6.04.14 analysis software. In Figures 1G, 2D, 4D, 6C, 6L, 6U, and 7F–7G, values are displayed as mean \pm SEM, and this information is included in the corresponding figure legends. In Figures 1H, 2E, 2F, 4B, 4G, 6D, 6M, and 6V, data are displayed as mean \pm SD, and this information is included in the corresponding figure legends. In Figures 2H, 3D, 4F, 6F, 6I, 6O, 6R, 7H, S3A, S3B, and S7H, the centerline corresponds to the median, and the upper and lower bounding box indicates the interquartile range of the displayed data. Dashed vertical lines in Figures 1C and 3E indicate the mean or median value on the x axis, as described in figure legends. In Figures S1E, S6C, S6G, and S6N, the “r” represents the Pearson’s correlation coefficient between phase-shift score and cellular expression levels of optoIDR constructs. Pairwise statistical testing in Figures 3C, 3D, 4B, 4D, 4F, 7I, S1D, S2A, S2E, S3A, S3B, S6B, S6F, S6M, and S7F was done using Student’s t test, K-S test, or Welch’s Two sample t test, as described in the respective figure legends. The sample size “n” in Figures 1G, 1H, 2D–2F, 4B, 4D, 4F, 6C, 6D, 6L, 6M, 6U, 6V, S1D, S2A, S2E, S6B, S6F, S6M, and S7G can be found in the figure panels. In Figure S7G, Pearson’s correlation coefficients and associated p values were calculated with the cor.test function in R. The –15A and –7A HOXD13-IDR constructs had less than 4 out of 30 data points being non-zero. No methods were used to determine if the data met the assumptions of the statistical approach.

Single-cell RNA-seq data pre-processing

The single-cell RNA-seq datasets were processed with 10x Genomics’ Cell Ranger pipeline version 3.0.0 and mapped to the mouse genome version mm10. The mapped reads from the Cell Ranger output in identified clusters 1 to 11 for the wild-type and SPDH samples were remapped to mouse genome mm9 in separate batches (one per cluster and condition) with the STAR aligner version 2.5.3a (Dobin et al., 2013) and subsequently transformed to bigwig files for UCSC genome browser visualization (Figures 5F and S5D) by using bamCoverage from deeptools version 3.1.2 (Ramírez et al., 2016) with normalization method BPM.

Single-cell RNA-seq analysis: filtering and normalization

The Cell Ranger expression profiles for wild-type and spdH cells were loaded into R version 3.5.1 (R Core Team, 2018) and were further processed by the R Seurat package version 2.3.4 (Butler et al., 2018). In particular, only cells that had between 200 and 7,000 genes detected as expressed, and less than 5% of mitochondrial reads were kept for further analysis. All remaining cells were then log-normalized and scaled up to 10,000 total reads using Seurat’s default settings.

Cluster identification

To identify cell populations in the wild-type scRNA-Seq data the following Seurat functions were used: FindVariableGenes, RunPCA, RunTSNE and FindClusters (Butler et al., 2018). First, the genes with the highest variation across all WT cells were determined. Second, those genes were standard normalized across all cells. Third, a PCA was run on those genes to identify the top ten directions with the highest variances across all wild-type cells and those were subject to t-SNE to get a better visual separation of the cells from the PCA input. Last, clusters were identified by first building a shared nearest neighbor graph and then running the Louvain algorithm on it (Blondel et al., 2008). The number of clusters was determined by the optimum of the modularity function from the Louvain algorithm.

Assignment of cell types to clusters

The marker genes in the 11 clusters of the wild-type reference map were identified by running the FindAllMarkers functions of the Seurat package (Butler et al., 2018), and the marker genes are listed in Table S2. The clusters were then assigned to cell types using a combination of approaches including GO term enrichment analysis, inspection of localized gene expression data from whole mount *in situ* hybridization [e.g., using the Gene eXpression Database (GXD)], and literature data on marker genes of cell types in the developing limb, as described below.

Cluster 1. Proliferating cells (S phase) of the distal proliferating mesenchyme

The marker genes in this cluster are enriched for GO terms associated with DNA replication (e.g., *Pcna*) (Figure S4B). The marker genes in the cluster also include previously described distal limb mesenchyme markers *Msx1* (Davidson et al., 1991; Lallemand et al., 2009), *Prrx2* (Lu et al., 1999a, 1999b), and *Twist1* (O’Rourke et al., 2002) (Figure S4B).

Cluster 2. Proliferating cells (Cytokinesis) of the distal proliferating mesenchyme

The marker genes in this cluster are enriched for GO terms associated with nuclear division and cytokinesis. The marker genes in the cluster also include previously described distal limb mesenchyme markers *Msx1* (Davidson et al., 1991; Lallemand et al., 2009), *Msx2* (Davidson et al., 1991; Lallemand et al., 2009), *Prrx2* (Lu et al., 1999a, 1999b), and *Twist1* (O’Rourke et al., 2002) (Figure S4B).

Cluster 3. Perichondrium (Proliferating)

The marker genes in this cluster are enriched for GO terms associated with cell cycle. The marker genes include perichondrium marker genes *Col1a1* (Villavicencio-Lorini et al., 2010), *Crabp1* (Villavicencio-Lorini et al., 2010) and *Hoxa11* (Boulet and Capecchi, 2004; Swinehart et al., 2013) (Figure S4B).

Cluster 4. Interdigital mesenchyme

The marker genes in this cluster include *Aldh1a2* (*Raldh*), a gene exclusively expressed in interdigital mesenchymal cells (Kuss et al., 2009). The marker genes in the cluster also include previously described distal limb mesenchyme markers *Msx1* (Davidson et al., 1991; Lallemand et al., 2009) and *Msx2* (Davidson et al., 1991; Lallemand et al., 2009) (Figure S4B).

Cluster 5. Perichondrium

The marker genes in the cluster include perichondrium marker genes *Col1a1* (Villavicencio-Lorini et al., 2010), *Crabp1* (Villavicencio-Lorini et al., 2010) and *Hoxa11* (Boulet and Capecchi, 2004; Swinehart et al., 2013) (Figure S4B).

Cluster 6. Proximal chondrocytes (carpal progenitors)

The marker genes in this cluster are enriched for GO terms associated with chondrocyte differentiation and skeletal system morphogenesis. The marker genes include *Sox6* (Akiyama et al., 2002), *Sox9* (Bi et al., 1999), *Foxc1* (Yoshida et al., 2015), and *Col2a1* (Zhao et al., 1997) (Figure S4B).

Cluster 7. Distal chondrocytes (phalanx progenitors)

The marker genes in this cluster are enriched for GO terms associated with chondrocyte differentiation and cartilage condensation. The marker genes include *Sox9* (Bi et al., 1999), *Sox5* (Akiyama et al., 2002), *Col2a1* (Zhao et al., 1997), *Col9a1* (Zhang et al., 2003), and *Pax9* (Neubüser et al., 1995) (Figure S4B).

Cluster 8. Proliferating cells

The marker genes in this cluster are enriched for GO terms associated with cell division, and housekeeping cellular functions, and lack classic cell type markers that would enable clear identification of their origin. The cells likely represent proliferating undifferentiated cells.

Cluster 9. Myoblasts

The marker genes in this cluster are enriched for GO terms associated with skeletal muscle cell differentiation. The marker genes in the cluster include *Myod1* (Davis et al., 1987) and *Myog* (Hasty et al., 1993) (Figure S4B).

Cluster 10. Hematopoietic cells

The marker genes in this cluster are enriched for GO terms associated with immune system function. The marker genes in the cluster include cell surface antigens *Fcgr1* and *CD48* (Figure S4B).

Cluster 11. Erythrocytes

The marker genes in this cluster are enriched for GO terms associated with erythrocyte maturation. The marker genes in the cluster include several hemoglobin genes (*Hba-x*, *Hba-a1*, *Hbb-bs*, *Hbb-bh1*) (Figure S4B).

Quantification of cell state proportions

Spdh cells were assigned to the identified 11 clusters of the wild-type reference cell state map using the nearest neighbor method. Spdh cells were assigned to the wild-type cluster that had the most similar mean expression profile. For each cluster and condition we normalized the cell counts by dividing by the total number of wild-type or spdh cells used. Last, we determined the log₂ ratio of these normalized cell state percentages in the spdh and wild-type samples per cluster (Figures 5C and S4A). The expression values in Table S2 are the normalized values output by the “AverageExpression” function of Seurat. Expression values shown in Figure S4A are log1p-transformed values.

Differential expression analysis

Differential expression analysis was performed within clusters between wild-type and assigned spdh cells by using the FindMarkers function of the Seurat package. The differentially expressed genes within the clusters are listed in Table S3. To display differentially expressed genes in Figure 5D, a q value cutoff of 0.05 was used.

Combined wild-type and spdh scRNA-seq data analysis

To determine whether the abundance of specific cell types was altered in the spdh limb with an alternative clustering approach, the following analysis was performed (Figures S5A–S5C). Wild-type and spdh cells were processed together, and only cells that had at least a single *Hoxd13* read were considered for further processing. Preprocessing steps were then performed as described above,

and PCA, t-SNE and cluster identification were run on the combined pool (wild-type and *spdh*) of *Hoxd13*-expressing cells together. The clusters identified thus comprised of both wild-type and *spdh* cells. Marker genes were identified by the FindAllMarkers function from the Seurat package. Last, the overlap of marker genes of the clusters from this approach with the marker genes from the clusters of the wild-type cells only (above) was determined (Figures S5A–S5C).

Gene ontology (GO) term enrichment analysis

Gene Ontology (GO) term enrichment was performed on differentially expressed genes per individual cluster using the *enrichGO* function of the *clusterProfiler* R package (version 3.10.1) (Yu et al., 2012). The cut-off value for significance was set at $q < 0.05$. As a control, cluster gene assignment was resampled both with and without replacement and the GO term enrichment repeated. Redundant terms were removed from the graphical display of the results, and terms displayed in Figure 5E were further filtered for q -value. The complete list of GO terms is listed in Table S5.

ChIP-Seq data (previously published)

The HOXD13 ChIP-Seq data was downloaded from GEO (accession number GSE81358), and was described in a previous study (Sheth et al., 2016). Previously published murine FLAG-HOXD13 and FLAG-HOXD13 Q317R ChIP-Seq data in chicken micromass were obtained from GEO (accession number GSE44799).

ChIP-Seq data processing

The HOXD13 ChIP-Seq data was aligned to the mm9 genome assembly using bowtie version 1.0.0 (Langmead et al., 2009) with parameters ‘-n 2 -e 70 -m 1 -k 1 -best -l 200’. If paired-end data was available we used only mate 1 from these as MACS version 1.4.2 (Zhang et al., 2008) uses only the first mate of a paired end mapping file. Peak calling in the HOXD13 ChIP-Seq data was done with MACS version 1.4.2 and parameters ‘-w -S -space = 50 -keep-dup = auto -pvalue = 1e-10’. The HOXD13 MACS peak co-ordinates are listed in Table S4A.

The murine FLAG-Hoxd13 transfected into chicken micromass ChIP-Seq datasets were aligned to the chicken genome (assembly gg3) using bowtie with parameters ‘-n 2 -e 70 -m 1 -k 1 -best -l 200’. If a library was sequenced on more than one lane then the mapping files of the same samples were merged with *bamtools merge*. Peak calling was done with MACS version 1.4.2 and parameters ‘-w -S -space = 50 -keep-dup = auto -pvalue = 1e-5’.

The paired-end sequencing reads of the murine H3K27Ac ChIP-Seq data with *D. melanogaster* spike-ins were first adaptor- and quality trimmed by using *cutadapt* version 1.18 with options -q 20 -m 25 -a AGATCGGAAGAGC -A AGATCGGAAGAGC -O 5. Trimmed reads were aligned to the mouse genome mm9 or fly (*D. melanogaster*) genome dm6 with *bwa* version 0.7.17-r1188 using the ‘mem’ command and default options. A sorted bam file and corresponding index was created with *samtools* version 1.9. Duplicate reads have been identified and filtered by using *MarkDuplicates* from the *gatk* suite version 4.1.4.0. Deduplicated reads were kept if their MAPQ value was at least 15 by using *samtools* with option -q 15. Peaks for the mapped mouse reads/fly reads were called using the *callpeak* program from *macs2* version 2.1.2, using Input ChIP as control and options ‘-narrow’. Technical replicates were merged using the *merge* program from *bamtools* version 2.5.1 and global normalization using reads per million mapped reads per base pair (rpm/bp) was performed.

Genome wide rpm/bp normalized coverage profiles were created by binning the mouse genome into 3mb regions and subsequently using *bamToGFF_turbo.py* from <https://github.com/BradnerLab/pipeline> (downloaded Oct. 2018) with options -s both -e 200 -m 60000 resulting in 50bp sized bins. H3K27Ac rpm/bp values for each bin were finally normalized by the spike-in factor obtained from the *D. melanogaster* spike-in ChIP-Seq reads (Orlando et al., 2014).

H3K27Ac ChIP-Seq analysis

In Figure 5J, the fold change in spike-in normalized H3K27Ac ChIP-Seq signal between *spdh* and wt samples was calculated as follows. First, the nearest HOXD13 peak within the same TAD around to the TSS of genes that were either downregulated or upregulated in cluster 4 was identified. The mean ChIP-Seq read density in the wt and *spdh* samples were calculated in a ± 1 kb window around the HOXD13 peak. As control set, the nearest HOXD13 peaks to genes within the same TADs were included. Only regions that had a mean K27Ac signal of at least 0.01 rpm/bp in both samples were kept.

In Figure S3D the union of ChIP-Seq peaks in the respective samples was identified, and the signal in $\log_{10}(\text{rpm/bp})$ in each sample at each region is plotted against the value of the same region in the other sample. Spearman’s rho was used as a correlation measure. The number of peaks in each panel is found on the y axis in Figure S3E.

The heatmaps in Figure S3E were created by taking the union of the HOXD13 peaks in the wt and mutant datasets, and extending each region by 1kb. Each region was divided into 50 bins and rpm/bp values were calculated for each bin with the *bamToGFF_turbo.py*. The same procedure was performed for 3kb flanking regions to the left and right of each peak region.

Capture C data

Capture C profiles for hindlimb and midbrain tissues from mouse embryos at the developmental stage E11.5 were obtained from GEO (accession number GSE84795). In particular the following files were used:

GSM2251426_CC-HL-E115-Wt-Mm_Merged-Smoothed-5kb-Norm.tar.gz
GSM2251422_CC-MB-E105-Wt-Mm_Merged-Smoothed-5kb-Norm.tar.gz

Topologically Associating Domains (TADs)

Topologically associating domains (TADs) in mouse E12.5 limb bud cells were described in a previous study (Kraft et al., 2019). The genomic co-ordinates of TADs are listed in Table S4B.

Enrichment of HOXD13 ChIP-Seq peaks around Cluster 4 dysregulated genes

The number of HOXD13 ChIP-Seq peaks within TADs that contain a Cluster 4 dysregulated gene was calculated as follows. First, TADs that contain the transcription start site (TSS) of at least one gene dysregulated in Cluster 4 (Figure 5G; Table S3) were identified, and the number of HOXD13 peaks within those 82 TADs were calculated (Figure 5G). Refseq annotation catalog 89 (24th Sep. 2018) was used to define TSSs. As a control, 82 TADs of all TADs that contain at least one gene that is not a Cluster 4 dysregulated gene, were randomly selected, and the average number of HOXD13 peaks within those TADs was calculated. This process was iterated 1,000 times, and the mean number of peaks of the 1,000 iterations is plotted in Figure 5G.

The analysis described in Figure S5E was performed as follows. First, TADs that contain at least one gene dysregulated in Cluster 4 in *spdh* versus wild-type cells (Table S3) were identified. Plotted is the percentage of these TADs that contain at least one HOXD13 ChIP-Seq peak (80/82 TADs, red bar). As a control, all other TADs that contain at least one gene, that is not a cluster 4 dysregulated gene, were identified, and the percentage of these TADs that contain at least one HOXD13 ChIP-Seq peak is plotted (2,760/3,567 TADs blue bar). To assess the statistical significance of the difference (Figure S5F), we randomly selected 82 TADs from the control TAD set 1,000 times, and plotted the distribution of the percentage of those TADs that contain at least one HOXD13 ChIP-Seq peak. The vertical red line represents the percentage of TADs with a Cluster 4 dysregulated gene that contains at least one HOXD13 peak (Figure S5E, red bar). The empirical p value obtained from this procedure was 0. Further information on the genes and HOXD13 peaks is listed in Table S6.

In Figure S5G, the distance between the transcription start site (TSS) of Cluster 4 dysregulated genes and the nearest HOXD13 peak is plotted in the left. For this analysis, Cluster 4 dysregulated genes that are contained within a TAD that also contains a HOXD13 peak were used. As a control (right) the following analysis was performed: first the TADs that contain at least one HOXD13 peak and at least one gene that is not a Cluster 4 dysregulated genes were identified. 84 of these TADs were randomly selected 1,000 times, and the distances between the TSS of the genes to the nearest HOXD13 peak are plotted. The vertical lines represent the median values. The x axis is in \log_2 scale.

Enrichment of Capture C signal at HOXD13 peaks

The mean contact frequency between Cluster 4 dysregulated genes and HOXD13 ChIP-Seq peaks (Figure 5H) was calculated as follows. First, the genes dysregulated in Cluster 4 were identified. 25 of these genes were used as Capture C viewpoints in the Capture C data described above, and these genes were kept for further analysis. The HOXD13 peaks within the TADs that contain these 25 genes were identified, and only the HOXD13 peaks that were separated by at least 50kb from the viewpoint gene were kept for further analysis (to reduce the elevated background signal close to the viewpoints). The summit positions of the HOXD13 peaks were then identified, and the mean Capture C signal around the summit position were plotted in 5kb bins around the summit (Figure 5H left). As a control, the same genomic co-ordinates of the peaks and bins were used to calculate the mean Capture C signal in embryonic midbrain tissue (Figure 5H right). The signal on the plot within the two samples is normalized against the highest mean bin signal in the respective sample.

Mean expression value of individual genes

For genes depicted in Figures 5F and S5D the coverage signal from the bigWig files of the 11 Clusters was exported. For each gene (*Msx1*, *Msx2*, *Hoxd12*, *Hoxd13* and *Sall1*), the mean coverage of the 3'UTR plus last coding exon was calculated as bins per million mapped reads. The mean coverage values in the wild-type cells are highlighted in blue, and the mean coverage values in the *spdh* cells are highlighted in red. Values are rounded to the nearest integer (Figures 5F and S5D).

Supplemental Figures

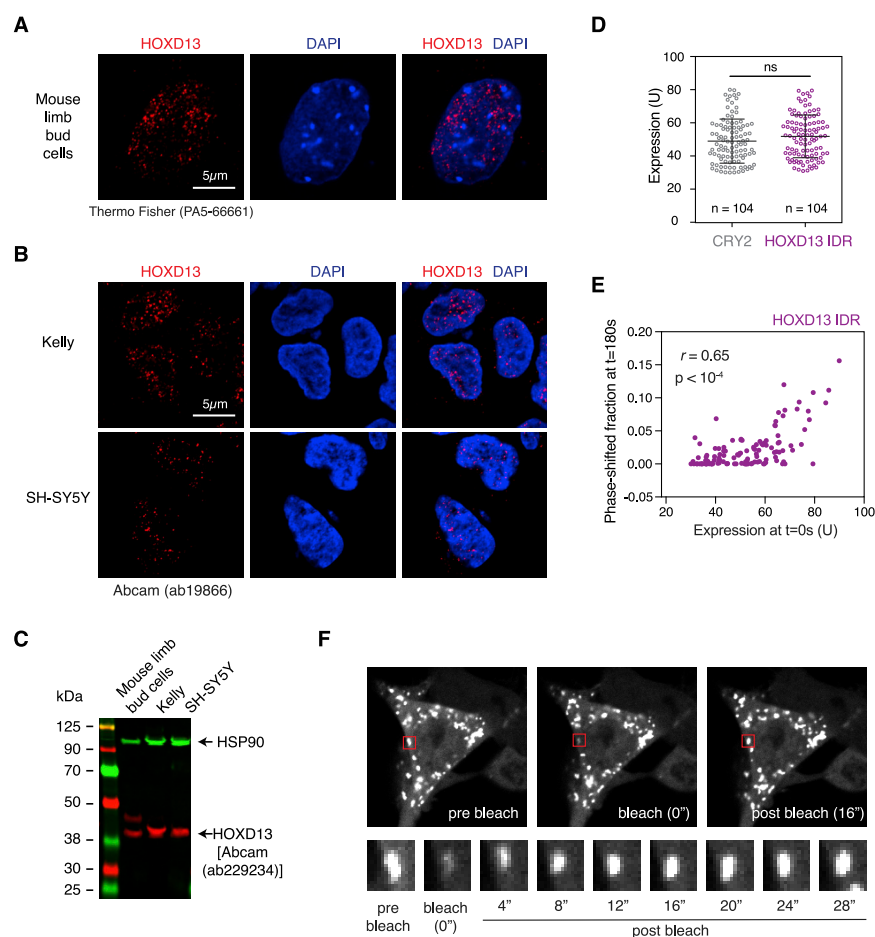


Figure S1. HOXD13 Forms Condensates, Related to Figure 1

(A) HOXD13 Immunofluorescence (IF) using the indicated antibody in E12.5 mouse limb bud cells.

(B) HOXD13 Immunofluorescence using the indicated antibody in human neuroblastoma cell lines Kelly and SH-SY5Y.

(C) western blot analysis of HOXD13 protein levels in mouse limb bud, Kelly and SH-SY5Y cells. HSP90 was used as a loading control.

(D) Expression level of the mCherry-CRY2 and HOXD13 IDR-mCherry-CRY2 fusion proteins in cells used in the optoDroplet experiments displayed in Figures 1F and 1G. The expression level was quantified as the mean fluorescence intensity of mCherry in the cells expressing the fusion proteins at time point 0 (i.e., before the 488nm laser stimulation). r denotes a Pearson's correlation coefficient.

(E) The expression level of the HOXD13 IDR-mCherry-CRY2 fusion protein correlates with droplet formation in HEK293T cells. Plotted is the fraction of the cytoplasmic area occupied by HOXD13 IDR-mCherry-CRY2 droplets after 3 minutes of 488nm laser excitation at every 20 s versus the mean expression of the fusion protein at time point 0 (i.e., before the 488nm laser stimulation). r denotes a Pearson's correlation coefficient.

(F) Time lapse images of live HEK293T cells expressing HOXD13 IDR-mCherry-CRY2 fusion protein before, during and after photobleaching. The bleached droplet is highlighted with a red box, and a zoom-in version of the box is displayed below at higher temporal resolution.

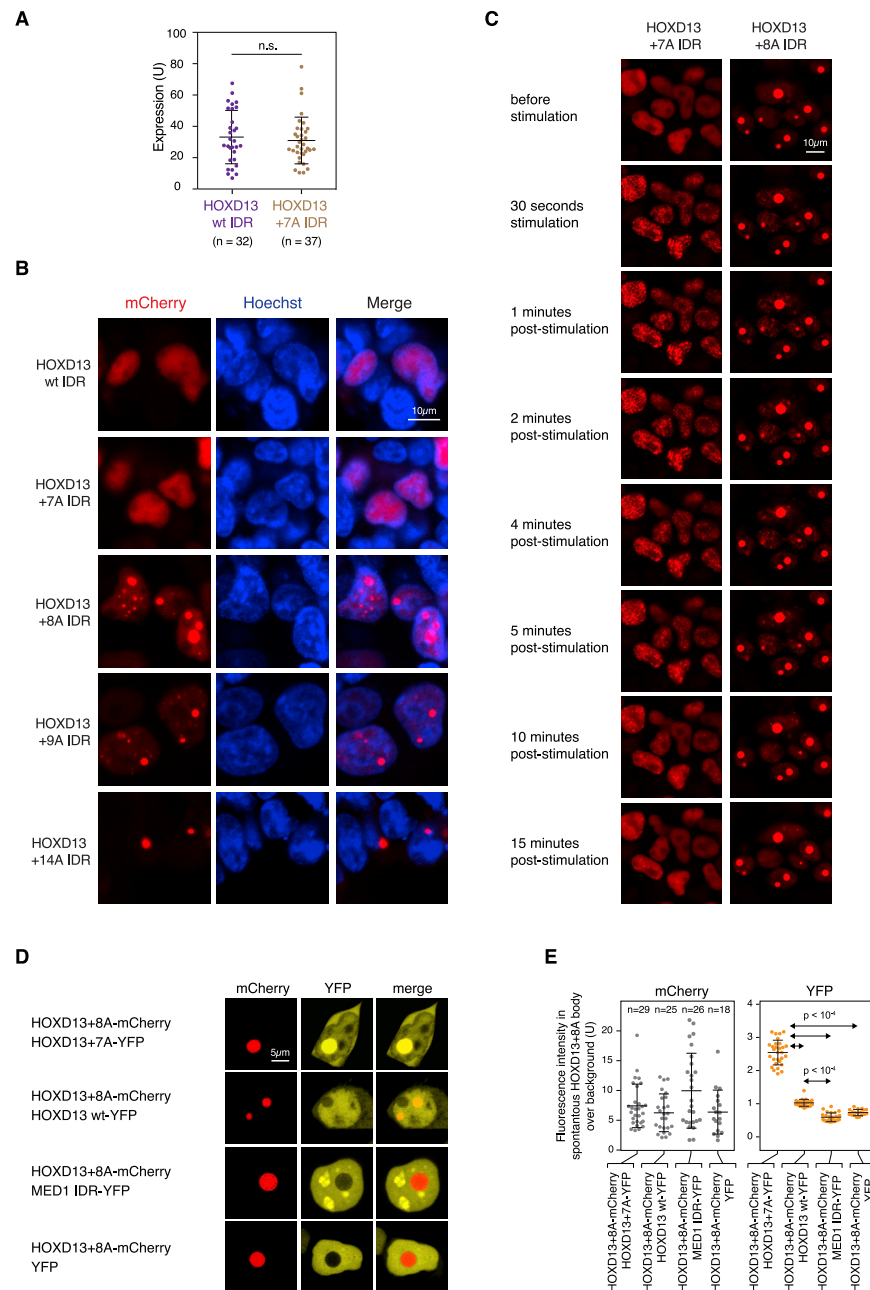


Figure S2. Synpolydactyly-Associated Repeat Expansions Enhance HOXD13 IDR Phase Separation in HEK293T Cells, Related to Figures 2 and 4

(A) Expression level of the HOXD13 wt IDR mCherry-CRY2 and HOXD13 +7A IDR mCherry-CRY2 fusion proteins in cells used in the optoDroplet experiments displayed in Figures 2C and 2D. The expression level was quantified as the mean fluorescence intensity of mCherry in the cells expressing the fusion proteins at time point 0 (i.e., before the 488nm laser stimulation).

(B) Images of nuclei of live HEK-293T cells expressing HOXD13 IDR-mCherry-CRY2 fusion proteins. The +8A and +9A IDRs form spontaneous nuclear bodies, whereas the spontaneously formed +14 IDR bodies appear cytosolic.

(C) The formation of light induced +7A and +8A HOXD13 IDR droplets is reversible. Displayed are images of live HEK-293T nuclei expressing HOXD13 +7A IDR mCherry-CRY2 and HOXD13 +8A IDR mCherry-CRY2 fusion proteins, before, at and after a 30 s pulse of 488nm laser stimulation.

(D) Spontaneously formed HOXD13 +8A IDR bodies incorporate the HOXD13 +7A IDR more readily than the HOXD13 wt IDR, and exclude MED1 IDR and YFP alone. Displayed are representative images of live HEK293T nuclei expressing the indicated fusion proteins.

(E) Quantification of fluorescence intensity of HOXD13 +7A IDR-YFP, HOXD13 wt IDR-YFP, MED1 IDR-YFP and YFP within spontaneously formed HOXD13 +8A IDR bodies.

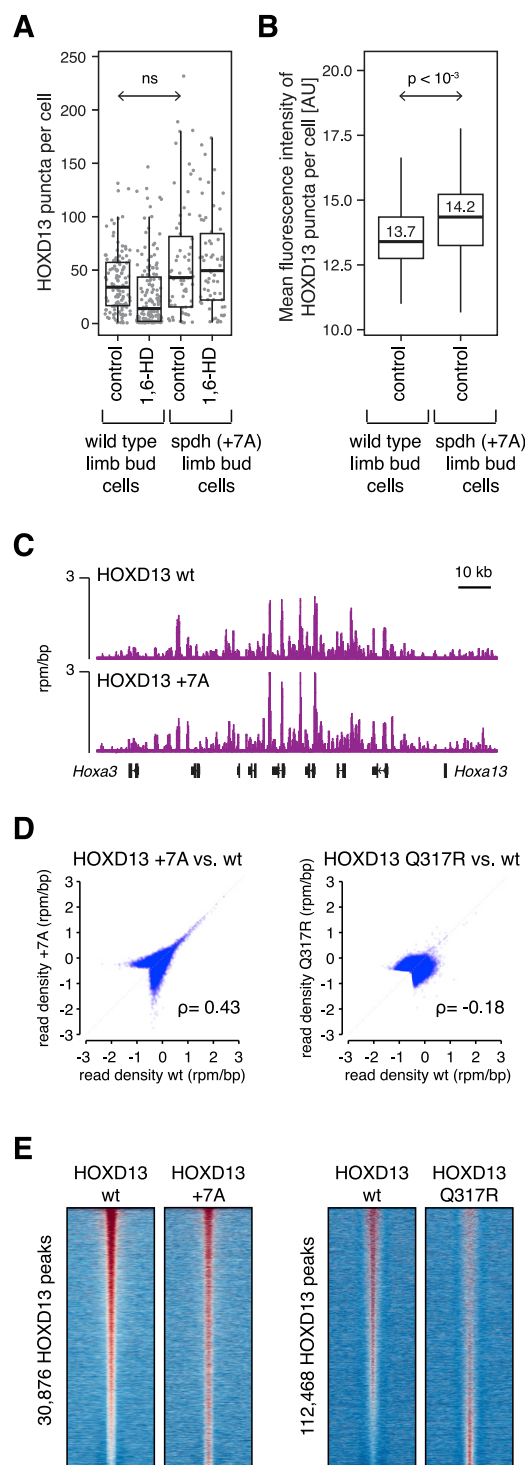


Figure S3. +7A Repeat Expansion Has Negligible Effect on HOXD13 DNA Binding, Related to Figure 4

(A) Number of HOXD13 puncta in mouse limb bud cells (displayed in Figures 4C and 4D) with or without treatment with 6% 1,6-hexanediol for 1min.

(B) Mean fluorescence signal intensity in HOXD13 puncta in mouse limb bud cells (displayed in Figures 4C and 4D). P value is from a Student's t test.

(C) ChIP-Seq binding profile of wild-type (wt) and +7A repeat-expanded HOXD13 at the *Hoxa* cluster locus in Chicken ChMM cells.

(D) Read density of wt and mutant HOXD13 binding at the union of the binding peaks detected in the wt and mutant HOXD13 ChIPs.

(E) Heatmap representation of the read density of wt and mutant HOXD13 binding at the union of the binding peaks detected in the wt and mutant HOXD13 ChIPs. The binding peak \pm 3kb is displayed, centered on the binding peak. The regions are ranked according to signal in the wt sample (top: high, bottom: low signal).

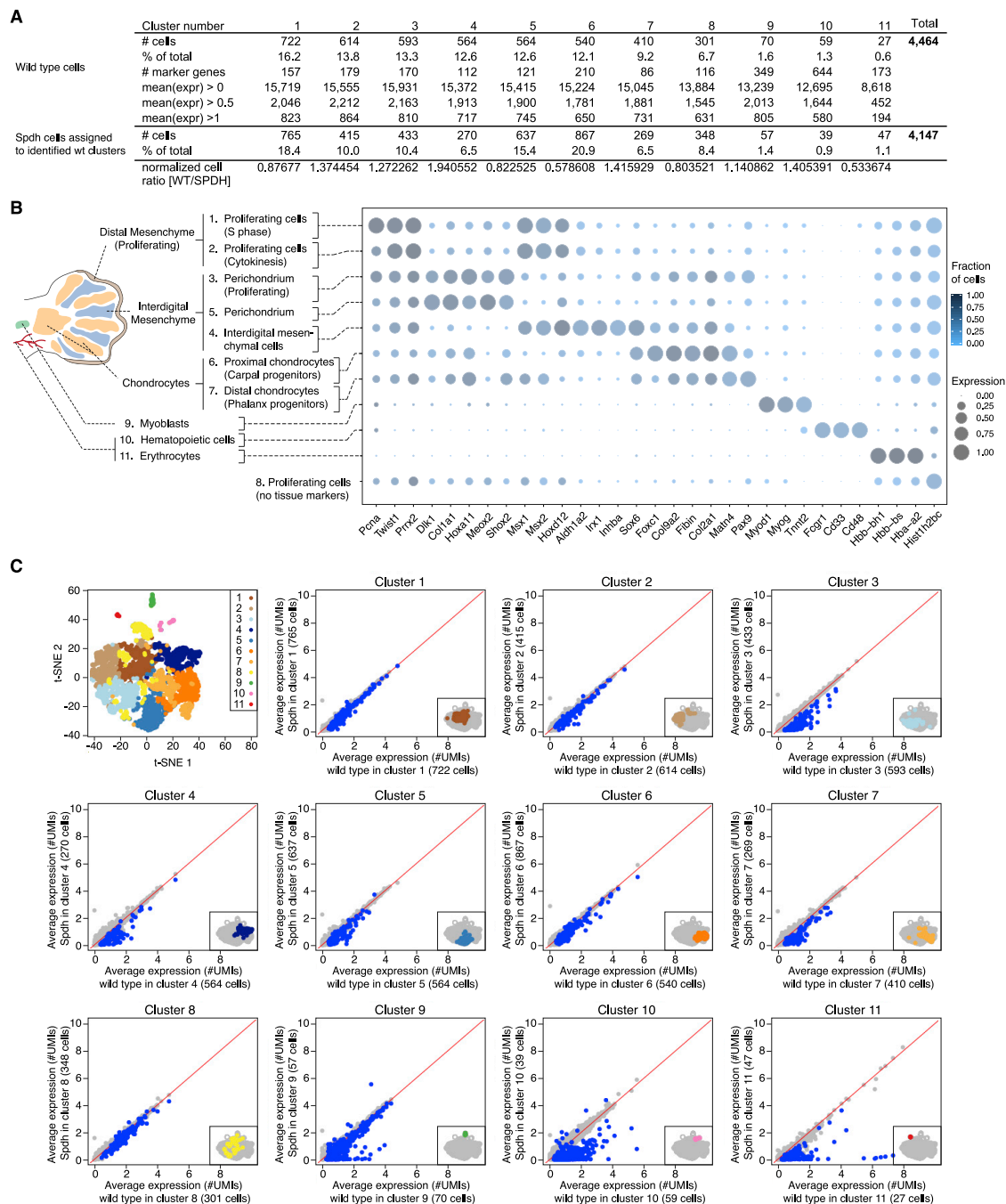


Figure S4. Summary Statistics of the scRNA-Seq Data and Cluster Characterization, Related to Figure 5

(A) Summary statistics of the scRNA-Seq data for the 11 clusters of the reference cell state map.

(B) Expression features of marker genes in the clusters of the wt reference cell state map. The 11 clusters were assigned to cell states based on marker genes described in the [STAR Methods](#) section. Displayed are the expression levels of a subset of marker genes in the individual clusters as a dot plot. The size of the dot is proportional to the total level of expression of the gene in the individual clusters, and the color of the dot is scaled to the proportion of cells within the individual clusters in which the expression of the gene is detected.

(C) (top left) Visualization of the wild-type scRNA-seq data using t-distributed Stochastic Neighbor Embedding (t-SNE). (rest) Gene expression changes in spdh cells versus wild-type cells within the individual clusters. Plotted are the average expression values of all genes detected in each cluster in spdh cells (y axis) versus wt cells (x axis). The respective marker genes in each cluster are highlighted in blue.

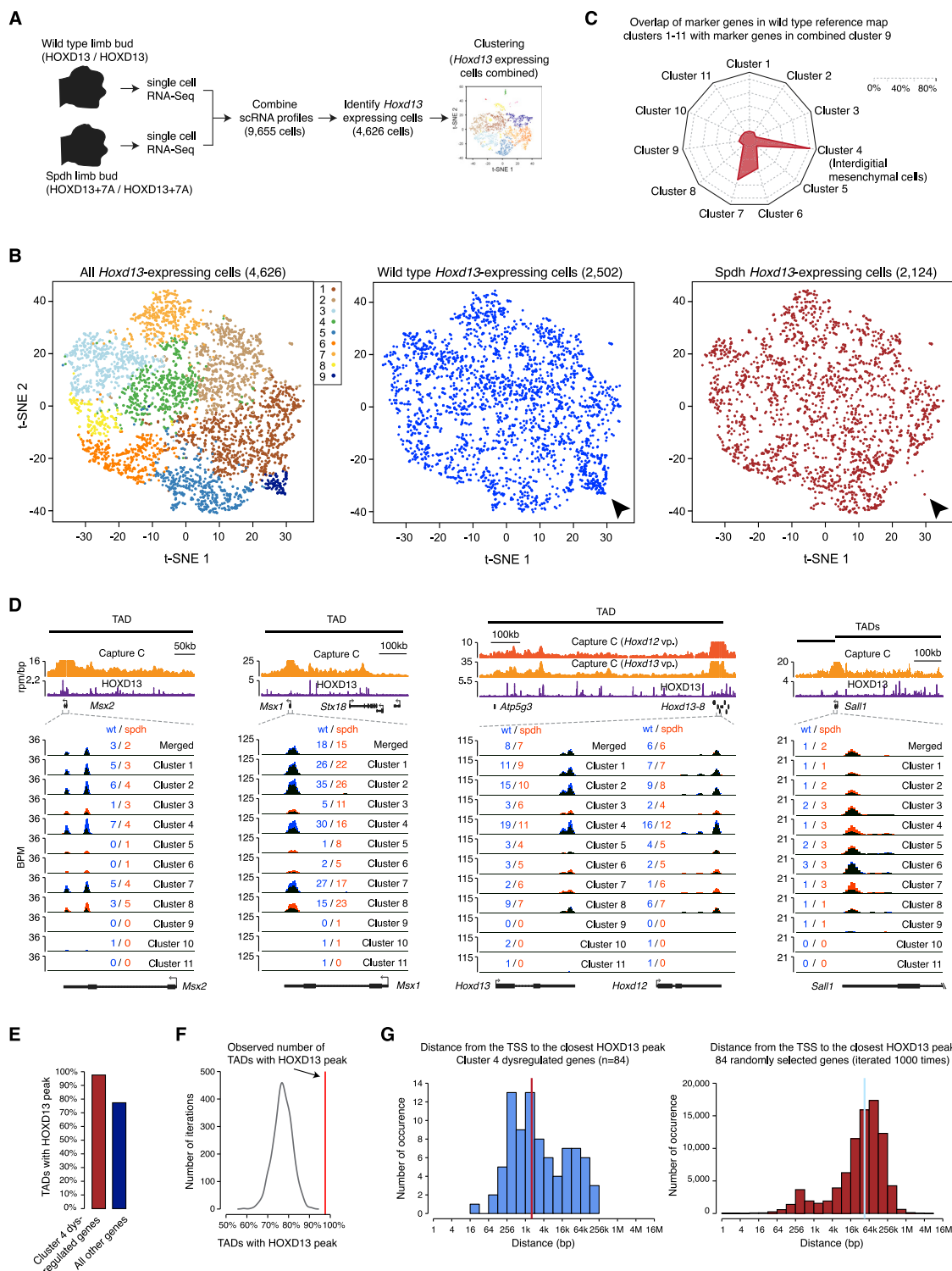


Figure S5. Characterization of Cell State and Gene Expression Changes in the spdh Limb Bud, Related to Figure 5

(A) Scheme of the workflow of an alternative analytical strategy to identify changes in the proportion of cell types in spdh limb buds. scRNA-Seq profiles of both wt and spdh cells were pooled, *Hoxd13*-expressing cells were selected, and the combined *Hoxd13*-expressing cells were clustered as described in the [STAR Methods](#).

(legend continued on next page)

(B) Visualization of the combined scRNA-seq data using t-distributed Stochastic Neighbor Embedding (t-SNE). Clusters identified in the combined data are highlighted in the left plot. Wild-type cells only (middle) and *spd* cells only (right) are highlighted in the t-SNE plot. The arrowhead highlights Cluster 9, which predominantly consists of wt cells, and is largely absent in the *spd* limb bud.

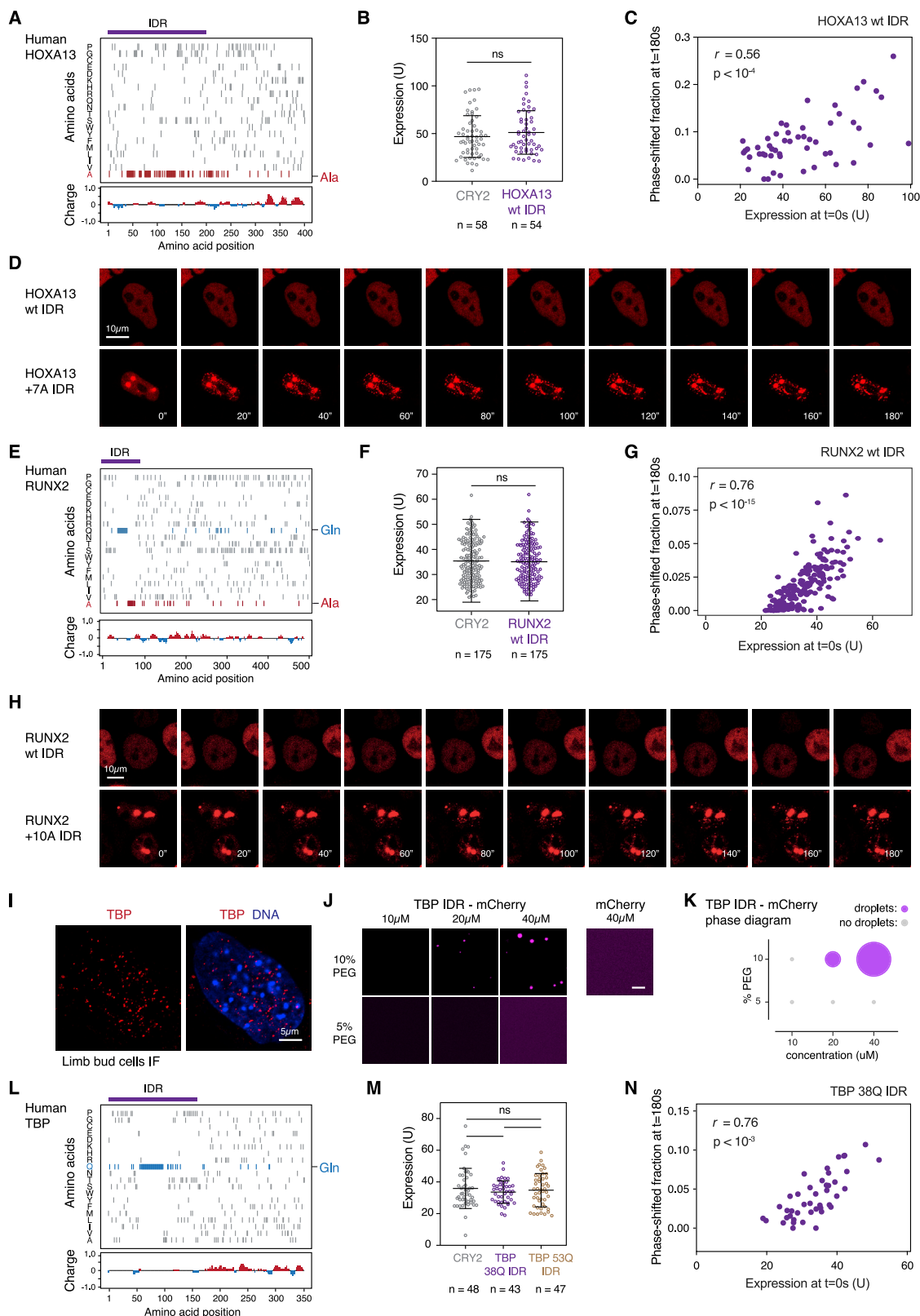
(C) The marker genes of Cluster 9 in the combined map have a substantial overlap with the marker genes in the interdigital mesenchyme cells (Cluster 4) identified in the wild-type reference cell state map. Displayed is a radar plot, where each arm in the radar represents the overlap between the marker genes of Cluster 9 in the combined map, with the marker genes in the clusters identified in the wild-type reference cell state map.

(D) Profiles of Capture C, HOXD13 ChIP-Seq and scRNA-Seq data at the *Msx2*, *Msx1*, *Hoxd13-Hoxd12* and *Sal1* loci. The topologically associating domain (TAD) at the locus is denoted by a black bar above the Capture C profile. The mean expression value in *spd* (red) and wt cells (blue) within each cluster is also displayed.

(E) Bar chart displaying the percentage of TADs that contain at least one gene dysregulated in Cluster 4 which also contain a HOXD13 binding peak (80/82 TADs), and the percentage of TADs that contain at least one gene that also contain a HOXD13 binding peak (2,755/3,562 TADs).

(F) The observed number of TADs that contain at least one gene dysregulated in Cluster 4 and a HOXD13 binding peak is significantly greater than the number of randomly permuted TADs containing a HOXD13 binding peak. 82 TADs were randomly selected 1000 times out of the 3,562 TADs, and the y axis shows the number of permutations in which the % of TADs denoted on the x axis contained at least one HOXD13 ChIP-Seq binding peak. The empirical P value is 0.

(G) (*left*) Distribution of the distance between the transcription start site (TSS) of the 84 genes dysregulated in Cluster 4 and the nearest HOXD13 ChIP-Seq peak. The median distance is highlighted by a red line. (*right*) Distribution of the distance between the transcription start site (TSS) of the 84 randomly selected genes and the nearest HOXD13 ChIP-Seq peak. The distance calculation was iterated 1000 times. The median distance is highlighted by a blue line.



(legend on next page)

Figure S6. The HOXA13 IDR, the RUNX2 IDR, and the TBP IDR Form Condensates, Related to Figure 6

- (A, E, L) Amino acid composition of human HOXA13, RUNX2 and TBP. Ticks represent amino acids indicated on the y axis at the positions indicated on the x axis. The cloned IDR is highlighted with a purple bar.
- (B, F, M) Expression level of the indicated fusion proteins in cells used in the optoDroplet experiments displayed in Figures 6B, 6K, and 6T. The expression level was quantified as the mean fluorescence intensity of mCherry in the cells expressing the fusion proteins at time point 0 (i.e., before the 488nm laser stimulation).
- (C, G, N) The expression level of the TF IDR-mCherry-CRY2 fusion proteins correlates with droplet formation in HEK293T cells. Plotted is the fraction of the nuclear area occupied by TF IDR-mCherry-CRY2 droplets after 3 minutes of 488nm laser excitation at every 20 s versus the mean expression of the fusion protein at time point 0 (i.e., before the 488nm laser stimulation). r denotes a Pearson's correlation coefficient.
- (D, H) Representative images of HEK293T nuclei expressing the indicated TF IDR-mCherry-CRY2 fusion proteins. Cells were stimulated with 488nm laser every 20 s for 3 minutes. Note the emergence of light-induced droplets in the HOXA13 +7A IDR-mCherry-CRY2 –expressing cells, and the RUNX2 +10A IDR-mCherry-CRY2 –expressing cells.
- (I) TBP Immunofluorescence (IF) in E12.5 mouse limb bud cells.
- (J) Representative images of droplet formation by purified TBP-mCherry and mCherry at the indicated concentrations in droplet formation buffer.
- (K) Phase diagram of TBP-mCherry in the presence of different concentrations of PEG-8000. The size of the circles is proportional to the size of droplets detected in the respective buffer conditions.

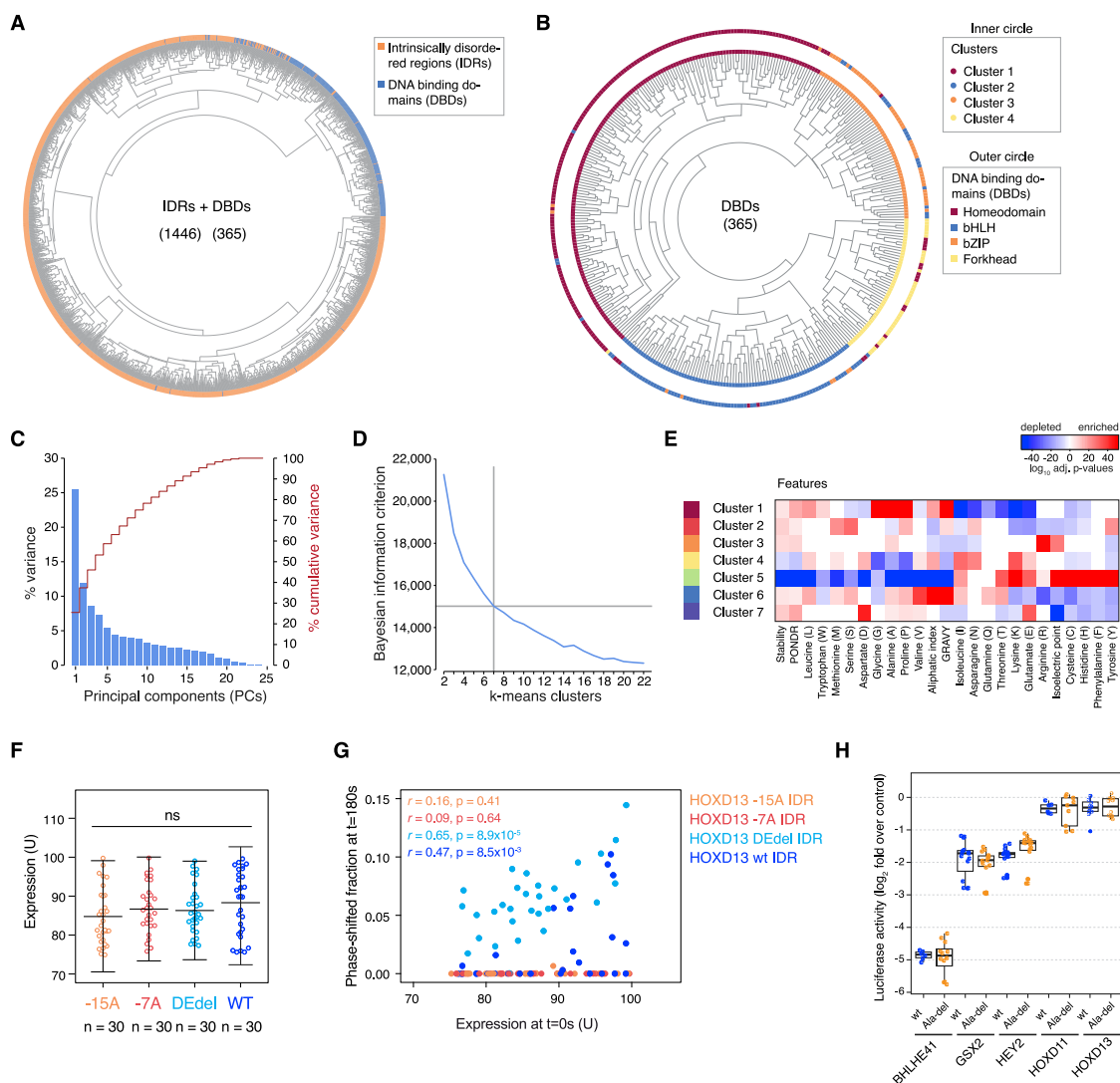


Figure S7. Molecular Characterization of Human TF IDRs, Related to Figure 7

(A) The classification tool efficiently separates human TF DBDs from IDRs. The circle depicts the annotation of the respective input protein sequence as a DBD or an IDR.

(B) The classification tool efficiently separates human TF DBDs families. The inner circle depicts the output of the classification (4 Clusters), and the outer circle depicts the annotation of DBDs of the respective input protein sequence.

(C) Screeplot showing the contribution of each principal component (PC) to the variance. The cumulative variance is depicted with a red curve.

(D) Bayesian information criterion (BIC) plot for various values of “k” tested in the IDR clustering analyses. The optimal cluster number (k = 7) was selected as the value at the inflection point of the BIC curve.

(E) Contribution of features (parameters) to the clustering of the individual TF IDR clusters. Features are highlighted in red/blue when statistically significant enrichment/depletion (Kolmogorov–Smirnov test) of a feature in a cluster compared to all other clusters was detected. The color code is proportional to $\pm \log_{10}$ (P value).

(F) Expression level of the HOXD13 IDR-mCherry-CRY2 fusion proteins in cells used in the optoDroplet experiments displayed in Figures 7E and 7F. The expression level was quantified as the mean fluorescence intensity of mCherry in the cells expressing the fusion proteins at time point 0 (i.e., before the 488nm laser stimulation).

(G) Dependence of droplet formation on expression level. The expression levels of the HOXD13 wt IDR-mCherry-CRY2, and HOXD13 DEdel IDR-mCherry-CRY2 fusion proteins correlate with droplet formation in HEK293T cells. Plotted is the fraction of the nuclear area occupied by the HOXD13 IDR-mCherry-CRY2 droplets after 3 minutes of 488nm laser excitation at every 20 s versus the mean expression of the fusion protein at time point 0 (i.e., before the 488nm laser stimulation). r denotes a Pearson’s correlation coefficient.

(H) Luciferase reporter activity of the indicated TF IDRs fused to GAL4-DBD.