Collective unstructured interactions drive chromatin binding of transcription factors

Abrar A Abidi[1], Gina M Dailey[1,2], Thomas G W Graham[1,2,3], Robert Tjian[1,2]
1) Department of Molecular and Cell Biology, University of California, Berkeley, United States; 2) Howard Hughes Medical Institute, University of California, Berkeley, United States 3) Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA

Abstract

Eukaryotic transcription factors (TFs) contain both structured DNA-binding domains (DBDs) and intrinsically disordered regions (IDRs). While the structures and sequence preferences of DBDs have been extensively characterized, the role of IDR-mediated interactions in chromatin binding and nuclear organization remains poorly understood, in part because these interactions have been difficult to measure in living cells. Here, we use a recently developed single-molecule technique, proximity-assisted photoactivation (PAPA), to investigate how IDRs influence TF associations with each other and with chromatin, focusing on the factors Sp1 and Klf1. We find that the number and patterning of aromatic and basic residues within IDRs govern both TF self-association and chromatin binding. Remarkably, the isolated DBD of Sp1 binds chromatin weakly and non-specifically. By contrast, the isolated IDR interacts transiently with chromatin-bound full-length Sp1, and these interactions become more stable when even minimal DNA-binding capacity is restored. Replacing Sp1's native DBD with domains from heterologous TFs with differing sequence preferences—or even with nonspecific bacterial DBDs—preserves IDR-mediated interactions and chromatin association. PAPA measurements also reveal extensive heterotypic interactions between Sp1 and other TFs. Together, these results establish PAPA as a powerful method for studying IDR interactions in their native context and suggest that IDRs participate in widespread cooperative associations scaffolded around chromatin. In contrast to classical models, we propose that TF specificity *in vivo* emerges from ensembles of weak, dynamic, and diverse interactions—not from the intrinsic sequence preferences of DBDs alone.

Introduction

Eukaryotic cells regulate gene expression through the activity of transcription factors (TFs). For decades, the prevailing model has held that TF specificity is encoded by structured DNA-binding domains (DBDs), which locate genomic targets by recognizing short DNA sequence motifs (Brent & Ptashne, 1985; Stormo, 2013; Lambert et al., 2018). This framework—rooted in structural biology, *in vitro* biochemistry, and genomics—has come under increasing strain (Biggin, 2011). Although eukaryotic genomes are saturated with potential binding sites for any given TF, only a small fraction are occupied *in vivo* (Lambert et al., 2018; Horton et al., 2023). Moreover, TFs with near-identical DBDs that recognize the same motif often bind to distinct genomic loci within the same cell type (Hollenhorst et al., 2009; Brodsky et al., 2020; Gera et al., 2022), and many occupied sites lack recognizable motifs altogether (Slattery et al., 2014). These contradictions call into question the sufficiency of motif recognition alone in determining TF localization. Given the gap between model and observation, we aim in this study to understand how transcription factors reliably bind chromatin in a nuclear environment where potential binding sites are both overabundant and underspecified.

Unlike their prokaryotic counterparts, eukaryotic transcription factors contain extensive intrinsically disordered regions (IDRs)—polypeptide segments that lack stable three-dimensional structure (Sigler, 1988). Recent studies suggest that IDRs can influence genomic specificity (Brodsky et al., 2020; Chen et al., 2022), but the basis of their contribution remains poorly understood. The peculiar properties of IDRs may hold the key to resolving this puzzle, yet their behavior in living cells has remained largely inaccessible. The overriding obstacle has been methodological: until recently, there were no tools capable of measuring weak, transient protein interactions in their native context. Most prior studies have relied on overexpression, artificial multimerization, or tethering to synthetic gene arrays to amplify and visualize bulk co-localization events (Boija et al., 2018; Wei et al., 2020; Chong et al., 2018). These approaches operate at the scale of large, synthetic nuclear foci rather than individual molecular interactions, and in doing so, often distort the very behaviors they aim to capture (McSwiggen et al., 2019). As a result, current models of IDR function in the nucleus are largely derived from observations in engineered systems, rather than from direct measurements of native behavior in the cell.

To overcome this technical barrier and observe how unstructured proteins interact under physiological conditions, we used proximity-assisted photoactivation (PAPA), a fluorescence imaging approach that we recently developed to detect association between proteins at single-molecule resolution in live cells. Unlike approaches involving artificial tethering, multimerization, or overexpression, PAPA can detect interactions between proteins in their native context at physiologically relevant expression levels (Graham et al., 2022; Graham et al., 2024; Dahal et al., 2025). To increase the throughput of these measurements, we built an oblique line-scan (OLS) microscope, which sweeps a light sheet across the specimen in synchrony with the high-speed readout of an sCMOS camera (Driouchi, 2025).

Using PAPA to study the transcription factors Sp1 and Klf1, we found that IDRs drive multivalent homotypic interactions and contribute substantially to chromatin association. These properties depend on both the number and patterning of aromatic residues within the IDRs, and are critically stabilized by basic residues flanking the DNA-binding domain, which likely promote weak electrostatic contacts with nucleic acids. Strikingly, the isolated Sp1 DBD was nearly incapable of binding chromatin in cells and did not co-localize with the full-length protein, suggesting that the DBD does not efficiently bind the same genomic regions as full-length Sp1. In contrast, the Sp1 IDRs retained weak chromatin association even in the absence of a DBD—an association that increased markedly when the protein could bind DNA, even minimally.

Substituting Sp1's DBD with those of heterologous transcription factors—irrespective of their sequence preferences—was sufficient to stabilize chromatin association and retain interactions with the full-length protein. This suggests that while DBD-DNA interactions are insufficient on their own, they can act as generic scaffolds that promote IDR-mediated binding. Conversely, fusing Sp1's DBD to IDRs from unrelated RNA-binding proteins enhanced chromatin association and enabled interactions with wild-type Sp1. Finally, we observed widespread heterotypic interactions between Sp1 and other transcription factors, supporting the view that TFs

participate in ensembles of weak, preferential interactions that help shape where and how they localize in the genome.

Taken together, these findings suggest that individual DBDs are not the sole determinant of TF binding specificity, but one component of a higher-order regulatory architecture in which disordered regions play a central role. The insufficiency and relative nonspecificity of eukaryotic DBDs may reflect a deeper evolutionary shift: as regulatory complexity increased, rigid and deterministic sequence recognition may have been supplanted by a more flexible, integrative mode of control. Synthesizing our observations, we propose a model of “emergent specificity”, in which chromatin binding arises from the dynamic participation of both DBDs and disordered regions in ensembles of preferential, weak interactions.

Results

PAPA Detects Weak Interactions Between Disordered Regions in Living Cells

Proximity-assisted photoactivation (PAPA) is a photoswitching phenomenon observed with Janelia Fluor dyes that we have used to detect endogenous protein-protein interactions at single-molecule resolution in live cells. Two proteins are fused to the self-labeling Halo and SNAPf tags, which are labeled with the “sender” and “receiver” fluorophores Janelia Fluor 549 (JF549) and Janelia Fluor X 650 (JFX650), respectively (Fig. 1A). The JFX650 receiver fluorophore is first placed in a nonfluorescent dark state by illumination with intense red (642 nm) light. Excitation of the JF549 sender fluorophore with a green (560 nm) light pulse then reactivates nearby JFX650 receiver fluorophores. In this way, receiver-labeled proteins associated with the sender-labeled protein are preferentially switched back to a fluorescent state, permitting their subsequent detection in single-molecule imaging. As an internal control, excitation with violet (405 nm) light is used to induce direct reactivation (DR) of receiver fluorophores, independent of their proximity with the sender, providing a sample of the overall population of molecules.

To increase the throughput of these measurements, we built an OLS microscope, which sweeps a light sheet over the specimen in synchrony with the readout of an sCMOS camera (Driouchi, 2025). This provides a much wider field of view than standard HILO microscopy (Fig. 1B) while maintaining a fast frame rate and high signal-to-noise. Because our previous applications of PAPA involved structured interactions, it was unknown whether PAPA could resolve interactions mediated by IDRs. To test this, we applied PAPA to the transcription factor Sp1 (Fig. 1C).

We used CRISPR-Cas9 to insert HaloTag at the endogenous *Sp1* locus in U2OS cells and expressed a SNAPf-Sp1 fusion protein from a doxycycline-inducible transgene integrated at the AAVS1 safe harbor locus (Fig. S1A). This strategy allowed precise control over copy number and expression level. Flow cytometry confirmed that transgene expression remained below endogenous levels, minimizing artifacts associated with overexpression (Fig. S1B).

For each condition, we alternated imaging cycles with green (PAPA) and violet (DR) excitation and counted the number of reactivated receiver molecules (Fig. 1D, E). These counts were compared to cells expressing nuclear SNAPf alone as a non-interacting control. We define a

“proximity index” as the ratio of PAPA to DR reactivation events, normalized to this control (Fig. 1E). A proximity index greater than one indicates that the proteins are in proximity more often than expected by chance.

PAPA revealed self-association for Sp1 (Fig. 1E). State array analysis showed that PAPA enriched molecules with the slow mobility characteristic of chromatin-bound proteins (Fig. 1E), suggesting that Sp1 self-association correlates with chromatin localization. We refer to this population as “chromatin-bound” for brevity, although we discuss caveats to this interpretation in the Discussion.

We applied the same approach to Klf1, a related zinc-finger TF that is not natively expressed in U2OS cells. Using a bicistronic transgene inserted at AAVS1, we co-expressed mixtures of either Halo-Klf1 and SNAPf-Klf1 or Halo-Klf1 and a nuclear SNAPf control. As with Sp1, PAPA detected robust self-association of Klf1 relative to the non-interacting control (Fig. S2A).

Aromatic Residues Promote IDR Self-Association and Chromatin Engagement

We next investigated the role of IDRs in Sp1 self-interaction. Sp1’s N-terminal IDR contains two aromatic-rich segments, labeled A and B (Fig. 1C). Using the inducible transgene system described above, we assayed interaction of SNAPf-tagged deletion mutants with endogenous Halo-tagged Sp1 (Fig 2A). Both A and B deletion mutants showed less PAPA signal with full-length Sp1 and reduced chromatin interaction (Fig. 2B-C). Nonetheless, PAPA enriched chromatin-bound molecules, indicating that association of the mutants with full-length Sp1 correlated with chromatin binding (Fig. 2C). The ∆A+B double mutant was more severely impaired in full-length Sp1 interaction and chromatin association, suggesting that the A and B regions contribute additively to self-association (Fig. 2B-C).

To assess the role of aromatic residues and rule out indirect effects from changes in protein length, we generated variants in which either all tyrosines or all aromatic residues outside the DBD were individually mutated to non-aromatic residues (Fig. 2A). Both variants showed substantially reduced association with full-length Sp1 and with chromatin (Fig. 2B-C). The tyrosine-only mutant produced a milder phenotype than the all-aromatic mutant, suggesting that self-association correlates with total aromatic content.

To assess whether the amino acid sequence, not just the total aromatic content, influenced self-association, we scrambled the amino acid sequences of segments A and B while preserving their overall composition (Fig. 2D). This redistribution dispersed the aromatic residues, reducing their local clustering—an alteration that, based on prior simulations and *in vitro* studies, was expected to weaken self-association (Martin et al., 2020). Surprisingly, the scrambled variant exhibited a higher proximity index than the wild-type protein and showed a slightly increased fraction of chromatin-associated molecules (Fig. 2D), indicating that the ability of the Sp1 IDR to drive self-association is remarkably tolerant of sequence rearrangement. The elevated fraction of slow-diffusing molecules suggests that altering residue order may expose latent interaction surfaces or stabilize weak, transient contacts, enhancing chromatin engagement.

The effect of residue arrangement on Sp1's self-association prompted us to ask whether naturally occurring sequence variation might have similar effects. While Sp1 is highly conserved across mammals—showing 96% sequence identity between human and mouse—Klf1 exhibits greater divergence, with only 71% identity, largely due to differences in the N-terminal disordered region. Although the total number of aromatic residues is preserved, their identities and spacing differ between species. Strikingly, mouse Klf1 exhibited a higher proximity index with Sp1 than human Klf1 and showed a slightly larger chromatin-associated fraction, suggesting that evolutionary shifts in residue order have influenced the interactions of the Klf1 IDR (Fig. S2B).

Taken together, these results indicate that Sp1 and Klf1 interactions in live cells are strongly dependent on their IDRs and influenced by the number and positioning of aromatic residues.

Basic Residues Flanking the DBD Stabilize Chromatin Binding

Sp1 contains two positively charged unstructured regions flanking its DBD: the N-terminal C region and the C-terminal D region (Fig. 1C). Such basic stretches are common among eukaryotic TFs (Oksuz et al., 2023), yet their role in chromatin engagement remains poorly understood. To investigate how these basic regions influence Sp1 interactions, we generated single and double deletion mutants lacking the C and D regions (Fig. 3A). Both Sp1ΔC and Sp1ΔD exhibited reduced interaction with endogenous Sp1, and the double mutant Sp1ΔC+D showed a further decrease (Fig. 3B). These interaction defects were milder than those observed for deletions of the aromatic-rich A and B segments. In contrast, chromatin association was more strongly impaired in the C and D deletion variants than in the A and B deletions (Fig. 3C). These findings suggest that the basic residues flanking the DBD stabilize chromatin binding—likely through electrostatic interactions with negatively charged nucleic acids. Destabilizing chromatin association, in turn, appears to weaken self-association, underscoring that these two types of interaction are interdependent.

To rule out confounding effects from changes in protein length and to neutralize positive charge beyond the C and D regions, we generated a variant in which all lysine, arginine, and histidine residues were mutated to glutamine (Fig. 3A). Strikingly, this variant exhibited minimal interaction with wild-type Sp1 and showed almost no chromatin association (Fig. 3B-C).
We initially reasoned that the precise positioning of basic residues might not be critical, given that flexible, unstructured regions can adopt a wide range of conformations. However, scrambling the C region—increasing the average distance between basic residues and the DBD—substantially reduced both self-association and chromatin engagement (Fig. 3D). These results suggest that the stabilizing effect of basic residues depends not only on their presence but on their proximity to the DBD.

The core Sp1 DBD is insufficient for chromatin binding and self-association

The basic-rich regions flanking Sp1's DBD lack the defined structure typically associated with sequence-specific nucleic acid binding. Yet in their absence, Sp1 loses its ability to stably bind chromatin, despite retaining an intact DBD and the full complement of aromatic-rich disordered

regions. To test whether this loss reflects an indirect consequence of IDR truncation, we examined the isolated DBD.

Previous DNase footprinting experiments showed that the isolated DBD of Sp1 readily binds GC box sequences *in vitro* (Kadonaga et al., 1988). Given that clustered GC box sequences are abundant in the human genome, one might expect that the isolated DBD would show both substantial chromatin binding and measurable PAPA signal with Sp1 molecules bound to adjacent GC boxes.

To the contrary, we found that the isolated DBD of Sp1 (Fig. 4A) exhibited low chromatin association and negligible PAPA signal with endogenous Sp1 (Fig. 4B, S3). Similar proximity indices were observed between endogenous Sp1 and histone H2B, a core constituent of chromatin that is not known to interact directly with Sp1 (Fig. 4B).

To rule out the possibility that the failure to detect interactions stemmed from steric hindrance by the SNAPf tag or the effect of changes in inter-dye distance on PAPA efficiency, we replaced the native IDR with a synthetic flexible linker of equivalent length. This variant also failed to interact with wild-type Sp1 and showed weak chromatin association, confirming that the DBD alone is insufficient for self-interaction or stable chromatin binding, irrespective of tag placement (Fig. 4B, S3). Supporting this conclusion, FRAP analysis revealed a much shallower bleach depth and more rapid fluorescence recovery for the isolated DBD than for full-length Sp1 (Fig. 4C).

These results show, surprisingly, that the structured DBD of Sp1 binds poorly to chromatin and fails to co-localize with full-length Sp1 in the absence of its IDRs.

Weak DNA Binding Scaffolds IDR-IDR Interactions and Stabilizes Genomic Localization

Given that the N-terminal IDR of Sp1 is necessary for chromatin binding and interaction with full-length Sp1 in cells, we next asked if it is sufficient. An N-terminal IDR fragment showed a weak though measurable PAPA signal with full-length Sp1 (Fig. 4E, middle bar). A weak but measurable interaction was likewise detected between full-length Klf1 and the Klf1 IDR (Fig. S2C).

We next examined a disease-associated Sp1 mutant containing a heterozygous frameshift that preserves the N-terminal IDR but retains only a single zinc finger in the DBD (Tummala et al., 2020). DNA binding by this truncated DBD, if any, is predicted to be weak and nonspecific, and we thus expected that its behavior would resemble that of the IDR alone. Remarkably, this mutant showed stronger interaction with wild-type Sp1 than the isolated IDR, suggesting that even weak, nonspecific DNA binding can stabilize IDR-mediated interactions between TFs.

To test the generality of this phenomenon, we examined a Klf1 mutation associated with human disease: E325K, which causes a dominant form of congenital dyserythropoietic anemia type IV (Arnaud et al., 2010). E325K alters a glutamate within the second zinc finger and is believed to confer novel DNA-binding specificity. Strikingly, it retained substantial interaction with wild-type Klf1 (Fig S2D), exceeding that of the isolated IDR.

DBD specificity is not essential for IDR interactions between TFs

To further investigate how the sequence-specificity of DBDs influences IDR interactions, we replaced Sp1's DBD with two structurally and functionally distinct alternatives: the DBD of human CTCF and the bacterial LacI protein (Fig. 5A). The former recognizes highly specific endogenous sequences, while the latter lacks specific binding sites in the human genome. Strikingly, both chimeras retained substantial interaction with wild-type Sp1 (Fig. 5B), while neither DBD alone exhibited detectable interaction (Fig. 5B). These results reinforce the idea that DNA binding—whether sequence-specific or not—can stabilize IDR interactions.

Chromatin Association Enables Heterotypic IDR Interactions Among Diverse Transcription Factors

To assess the generality of IDR-mediated TF interactions, we first asked whether IDRs from functionally unrelated proteins could associate with the IDR of Sp1. A previous study using a synthetic gene array found no significant co-localization between Sp1's IDR and those of EWS, FUS, or TAF15 (Chong et al., 2018). Revisiting these interactions using PAPA, we similarly observed little to no proximity between the freely diffusing IDRs and wild-type Sp1 on native chromatin (Fig. 5E).

However, when these same IDRs were fused to DBDs (Fig. 5D), they exhibited increased interactions with Sp1 (Fig. 5E) and a higher proportion of chromatin-bound, slow-diffusing molecules (Fig. 5F). These results indicate that chromatin binding can unmask latent interaction potential between IDRs that otherwise interact poorly.

To test whether such heterotypic interactions represent a broader feature of TF behavior, we next measured PAPA signals between endogenous Sp1 and six other TFs, each bearing substantial IDRs (Fig. 6A). In every case, we observed interactions above background, with PAPA selectively enriching chromatin-bound molecules (Fig. 6B, C). These findings suggest that Sp1 engages in a spectrum of weak, chromatin-localized interactions with other disordered TFs.

Together, these results reveal that promiscuous IDR–IDR interactions—often invisible in free solution—are potentiated by chromatin association. Chromatin therefore acts as a scaffold that concentrates disordered regions, allowing TFs to participate in dynamic ensembles held together by weak, unstructured interactions.

Discussion

Transcription factor binding has long been understood through the lens of structured DNA-binding domains recognizing specific DNA sequence motifs. However, our findings challenge this modular view. Using PAPA, a single-molecule imaging method uniquely suited to probing weak, transient interactions in live cells, we reveal an alternative picture in which transcription factor co-localization along chromatin emerges through ensembles of low-affinity interactions mediated by intrinsically disordered regions and scaffolded by DNA binding. Rather than reflecting pre-programmed sequence recognition, TF occupancy reflects the dynamic behavior of multivalent IDR ensembles along chromatin.

Given that fluorophore reactivation occurs rapidly, PAPA is uniquely sensitive to weak, transient interactions that escape detection by traditional biochemical methods. By tagging proteins with spectrally distinct self-labeling dyes, we detect associations at the single-molecule level in their native nuclear environment. To scale these measurements across many interaction pairs, we built an OLS microscope, which synchronizes light-sheet excitation with fast camera readout to dramatically increase acquisition throughput. This platform enabled systematic profiling of interaction strengths and chromatin association for dozens of TF-IDR variants and heterotypic combinations across hundreds to thousands of cells per experiment. No prior method allows for this combination of resolution, sensitivity, and throughput in live-cell IDR interaction measurements.

These live-cell measurements uncovered several surprising patterns. The isolated Sp1 DBD, despite binding GC box motifs *in vitro*, failed to co-localize with full-length Sp1 in cells. Conversely, TF variants with heterologous DBDs—even DBDs lacking sequence-specificity—retained robust interactions with wild-type Sp1. Heterotypic TF-IDR pairs that showed negligible association in isolation exhibited strong interactions when directed to chromatin by DBDs. These results suggest that specificity cannot be attributed to DBD motif recognition alone. Instead, DNA association locally concentrates TFs along a chromatin scaffold, facilitating interactions that are improbable in solution.

This mode of assembly differs fundamentally from classical models of cooperative binding, which assume discrete, stereospecific interactions among folded domains at well-defined genomic loci. In contrast, we propose a model of *emergent specificity*, in which specificity arises from the collective behavior of weak, flexible, and promiscuous interactions that are dynamically scaffolded by chromatin. This ensemble view could explain how transcription factors with near-identical DBDs bind distinct genomic regions, why many *in vivo* binding sites lack canonical motifs, and how IDR composition influences occupancy patterns without encoding direct sequence recognition.

This model may also help to explain why evolution has selected for eukaryotic TFs which, unlike their prokaryotic counterparts, are enriched for IDRs and typically recognize shorter, more degenerate motifs. This may reflect a transition toward a mode of regulation in which specificity emerges not from rigid recognition of individual TF motifs but from highly distributed TF interactions mediated by disordered regions.

Our findings may also help explain the functional capacity of oncogenic fusion proteins such as EWS-FLI1 and FUS-CHOP. While EWS and FUS are not DNA-binding transcription factors, fusion of their IDRs to DBDs may allow them to integrate aberrantly into transcriptional complexes. Our data suggest that these proteins do not act as rogue factors with novel specificity, but rather as opportunistic participants in pre-existing IDR interaction networks scaffolded around chromatin.

While PAPA offers unparalleled resolution for detecting weak interactions in living cells, a key limitation is that it does not reveal at which specific genomic loci interactions occur. In addition, SMT cannot distinguish direct DNA binding from indirect association via large protein complexes or RNAs. Integration with genomic-resolution methods like ChIP and CUT&RUN will be critical—though care must be taken, as such techniques may disrupt the very interactions we aim to study. Although our results establish a strong link between IDR interactions and chromatin binding, further work is needed to connect these interactions to transcriptional outcomes.

Our study reframes transcription factor binding to chromatin as a property not of individual structured domains, but of distributed ensembles of TFs organized through IDR-IDR associations. These dynamic, low-affinity interactions—directly visualized in living cells—very likely influence where TFs bind along the genome to regulate gene expression.

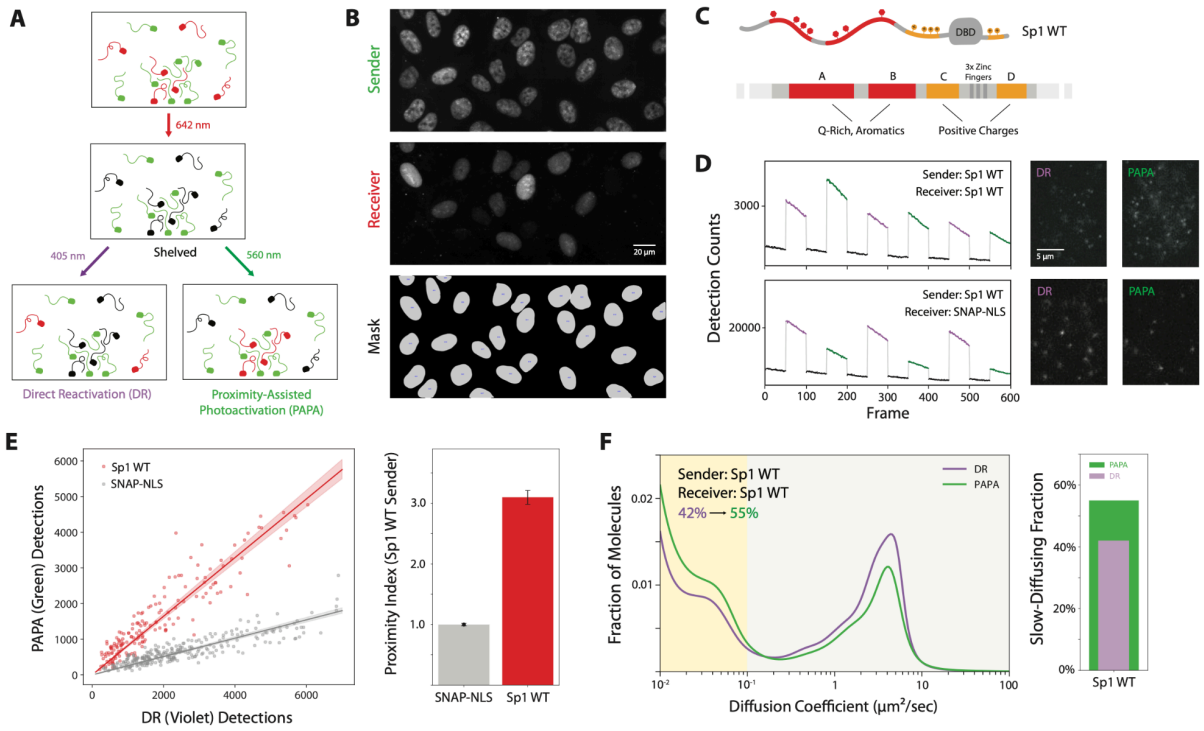

Figure 1. PAPA detects Sp1 self-association in living cells.
(A) Schematic of proximity-assisted photoactivation (PAPA). Receiver-labeled proteins (red) are shelved into a nonfluorescent dark state using 642 nm light. Subsequent excitation of nearby sender-labeled proteins (green) with 560 nm light selectively reactivates nearby receivers. As a control, 405 nm illumination induces direct reactivation (DR) of receivers independent of their proximity to the sender.
(B) High-throughput single-molecule imaging using an OLS microscope. Representative widefield images show sender (JF549-Halo-Sp1), receiver (JFX650-SNAPf-Sp1), and nuclear segmentation masks used for quantification.
(C) Domain diagram of Sp1 showing the locations of Halo and SNAPf tags. The N-terminal A and B regions are enriched in glutamine and aromatic residues, while the C and D regions flanking the DBD are enriched in basic residues.
(D) Detection counts over time under alternating violet (DR) and green (PAPA) illumination for cells co-expressing Sp1 as both sender and receiver (top) or Sp1 with a non-interacting nuclear SNAPf control (bottom). Insets show representative single-molecule frames for each condition and illumination mode.
(E) Left: Scatterplot of DR versus PAPA detection counts per nucleus. Right: Bar plot showing proximity index (PAPA/DR), normalized to non-interacting SNAPf-NLS control.
(F) Left: Diffusion coefficient distributions of PAPA- and DR-reactivated molecules. Right: Slow-moving (< 0.1 µm²/s) fraction of molecules reactivated by PAPA (green) and DR (violet).

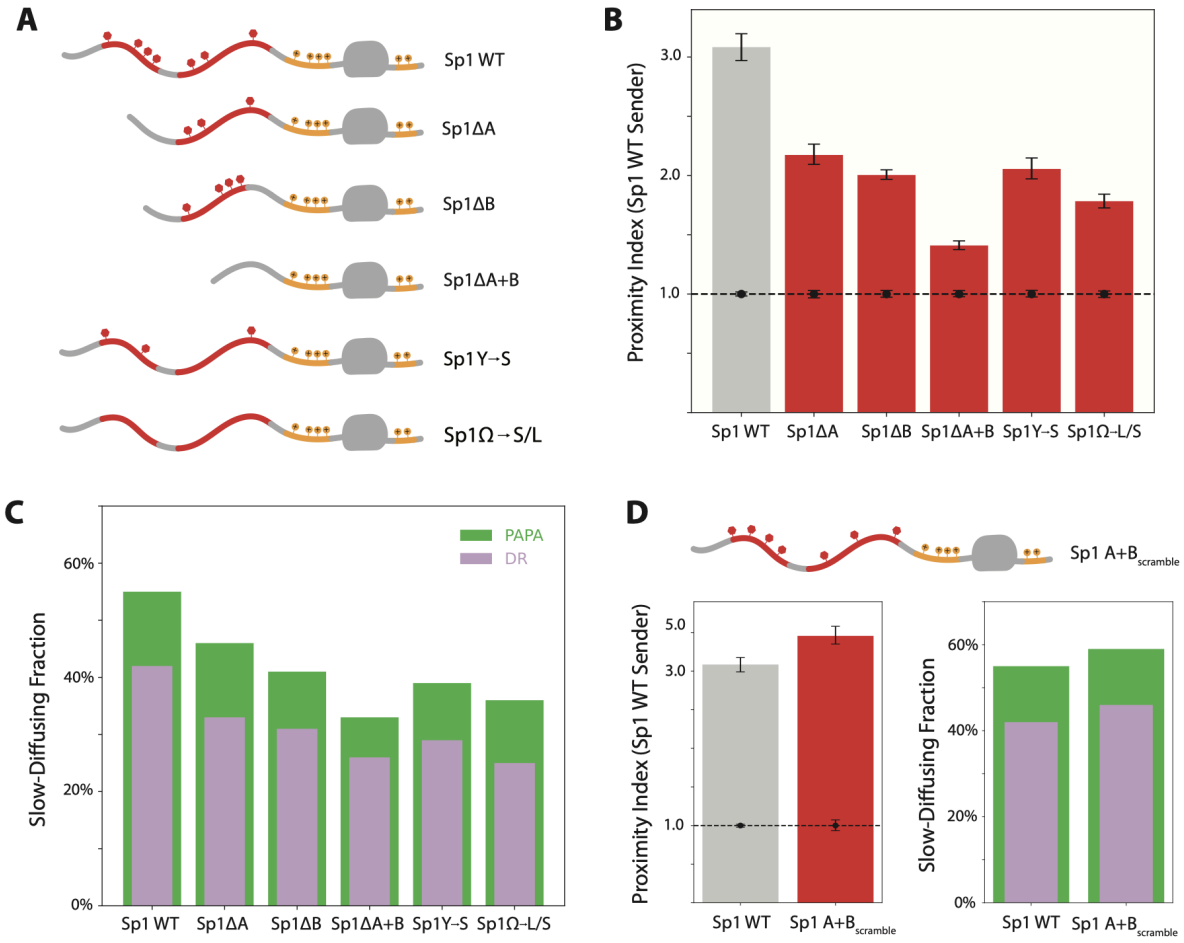

Figure 2. Aromatic residues in Sp1 IDRs drive self-association and chromatin binding.
(A) Domain diagrams of A and B domain mutants.
(B) PAPA signal between full-length Sp1 and each variant in (A).
(C) Slow-moving (< 0.1 μm²/s) fraction of molecules reactivated by PAPA (green) and DR (violet).
(D) Left: Effect of the A+$B_{scramble}$ variant on PAPA signal with full-length Sp1. Right: Corresponding slow-diffusing fractions.

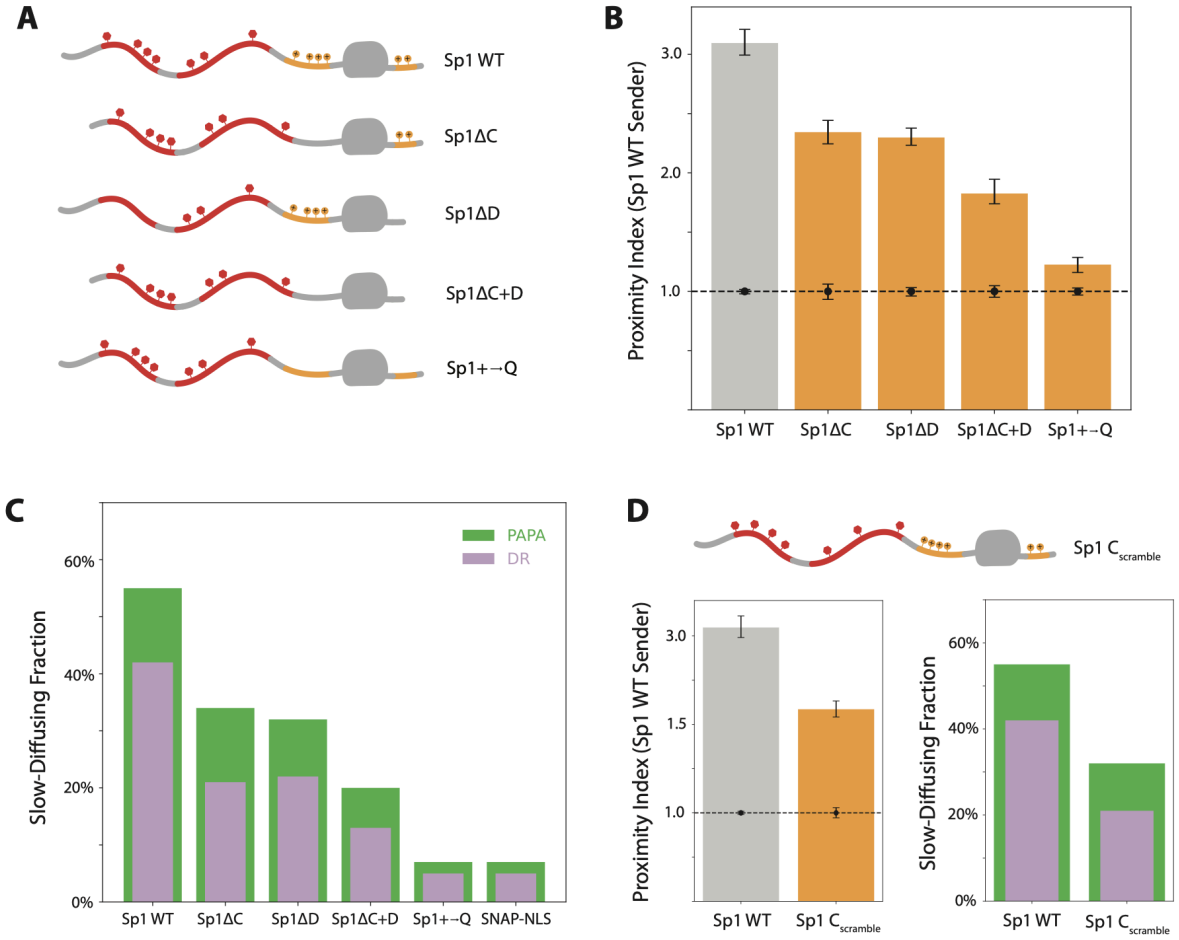

Figure 3. Basic residues flanking the Sp1 DBD stabilize self-association and are necessary for chromatin binding.

(A) Domain diagrams of C and D domain variants.

(B) PAPA signal between full-length Sp1 and each variant in (A).

(C) Slow-moving (< 0.1 μm²/s) fraction of molecules reactivated by PAPA (green) and DR (violet).

(D) Left: Effect of the $C_{scramble}$ variant on PAPA with full-length Sp1. Right: Corresponding slow-diffusing fractions.

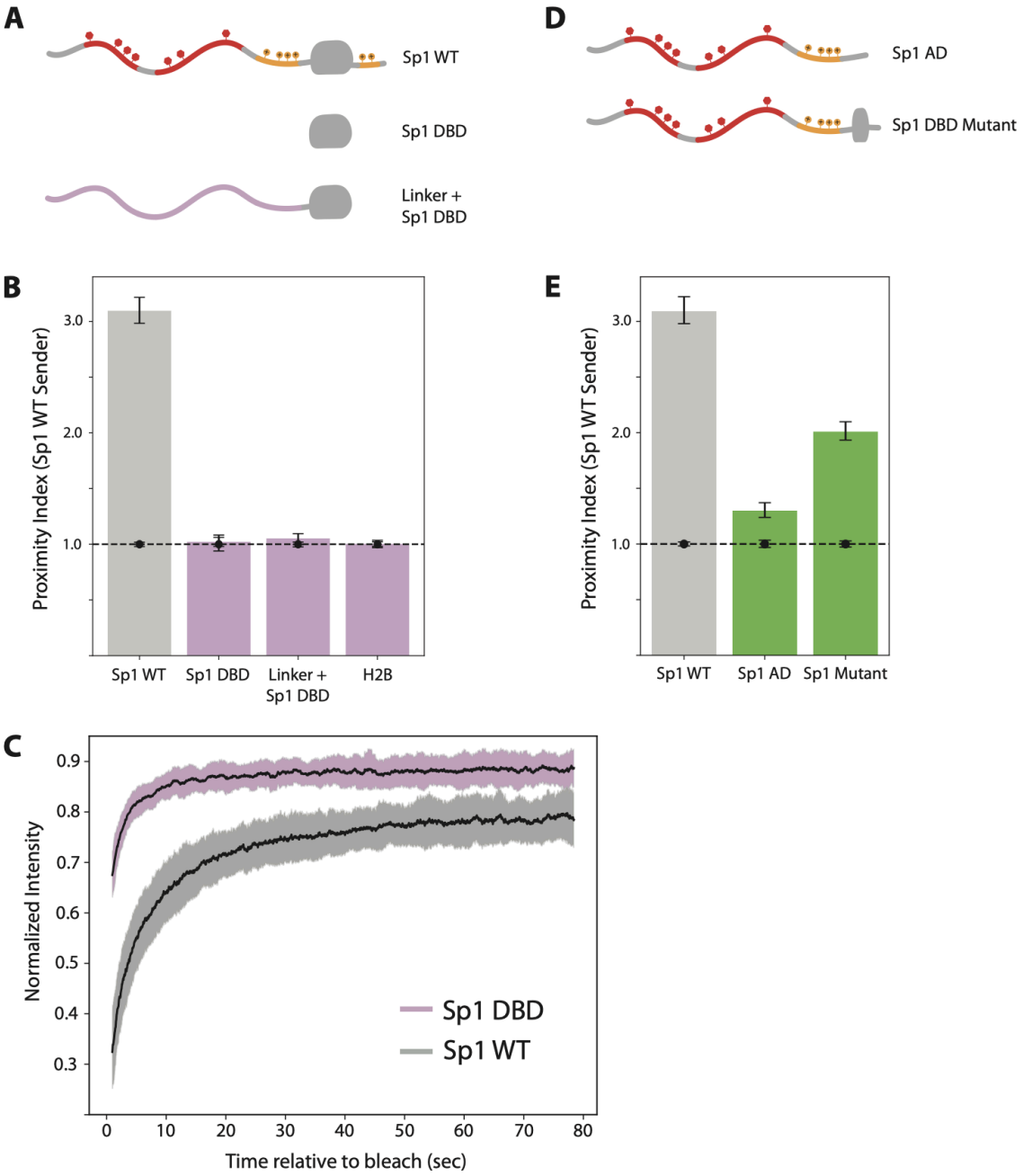

Figure 4. The Sp1 DBD does not co-localize with wild-type Sp1 and binds chromatin poorly.

(A) Domain diagrams of IDR deletion and linker replacement mutants.

(B) PAPA signal between full-length Sp1 and each variant in (A), with histone H2B-SNAPf as an additional non-interacting negative control.

(C) Fluorescence recovery after photobleaching (FRAP) curves comparing full-length Sp1 and the isolated DBD.

(D) Cartoons of the isolated IDR (Sp1 AD) and single-zinc finger mutant (Sp1 DBD Mutant).

(E) PAPA signal between full-length Sp1 and the variants in (D).

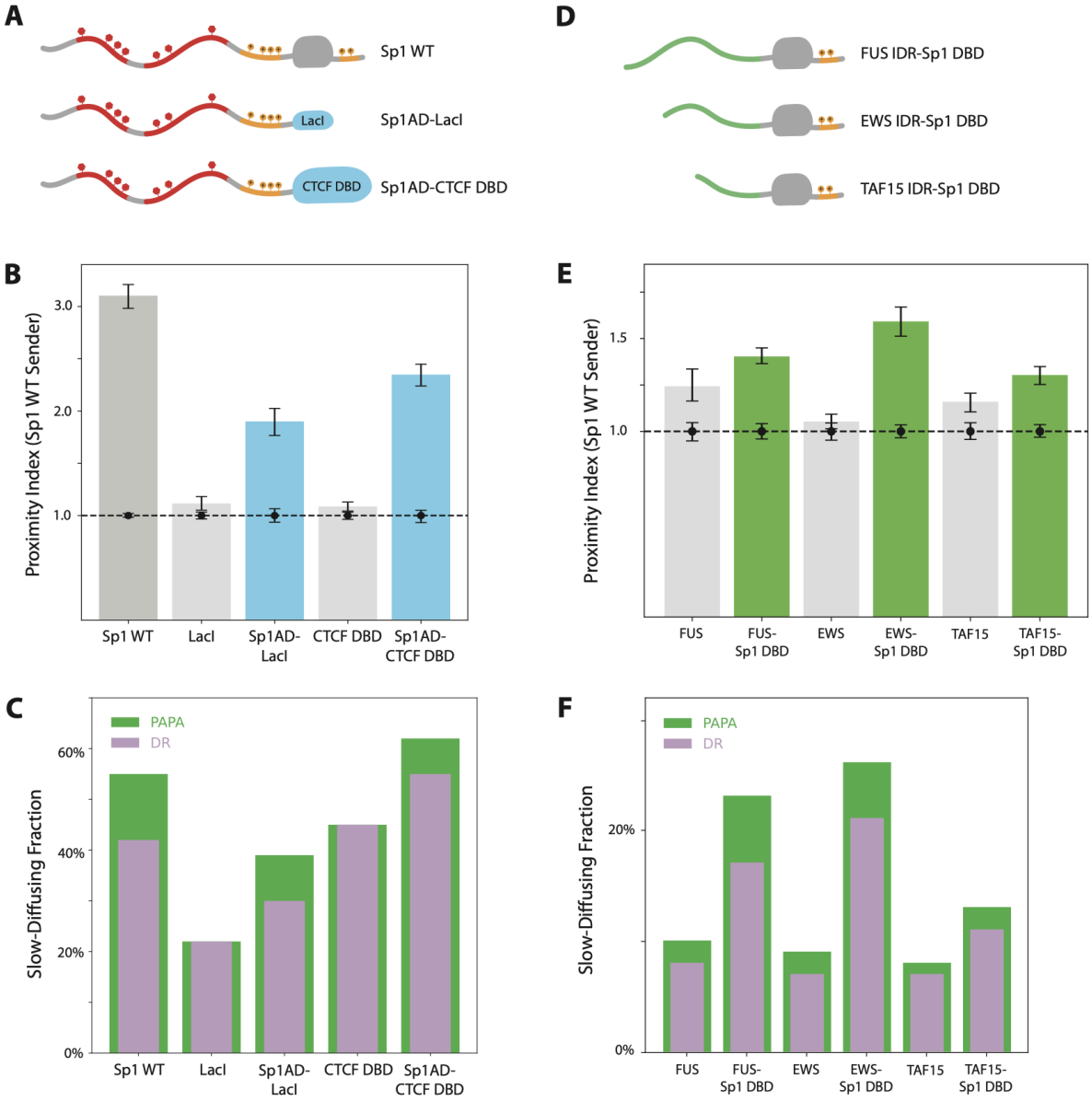

Figure 5. Sp1 variants with heterologous DBD or IDR swaps retain association with wild-type Sp1.

(A) Domain diagrams of wild-type Sp1 and chimeric variants in which the Sp1 DBD is replaced with LacI or the CTCF DBD.

(B) PAPA signal between full-length Sp1 and each variant in (A).

(C) Slow-moving (< 0.1 μm²/s) fraction of molecules reactivated by PAPA (green) and DR (violet) for each variant in (A).

(D) Domain diagrams of constructs in which IDRs from FUS, EWS, or TAF15 replace the N-terminal IDR of Sp1.

(E) PAPA signal between full-length Sp1 and each variant in (D).

(F) Slow-moving fraction of PAPA/DR-reactivated molecules for each variant in (D).

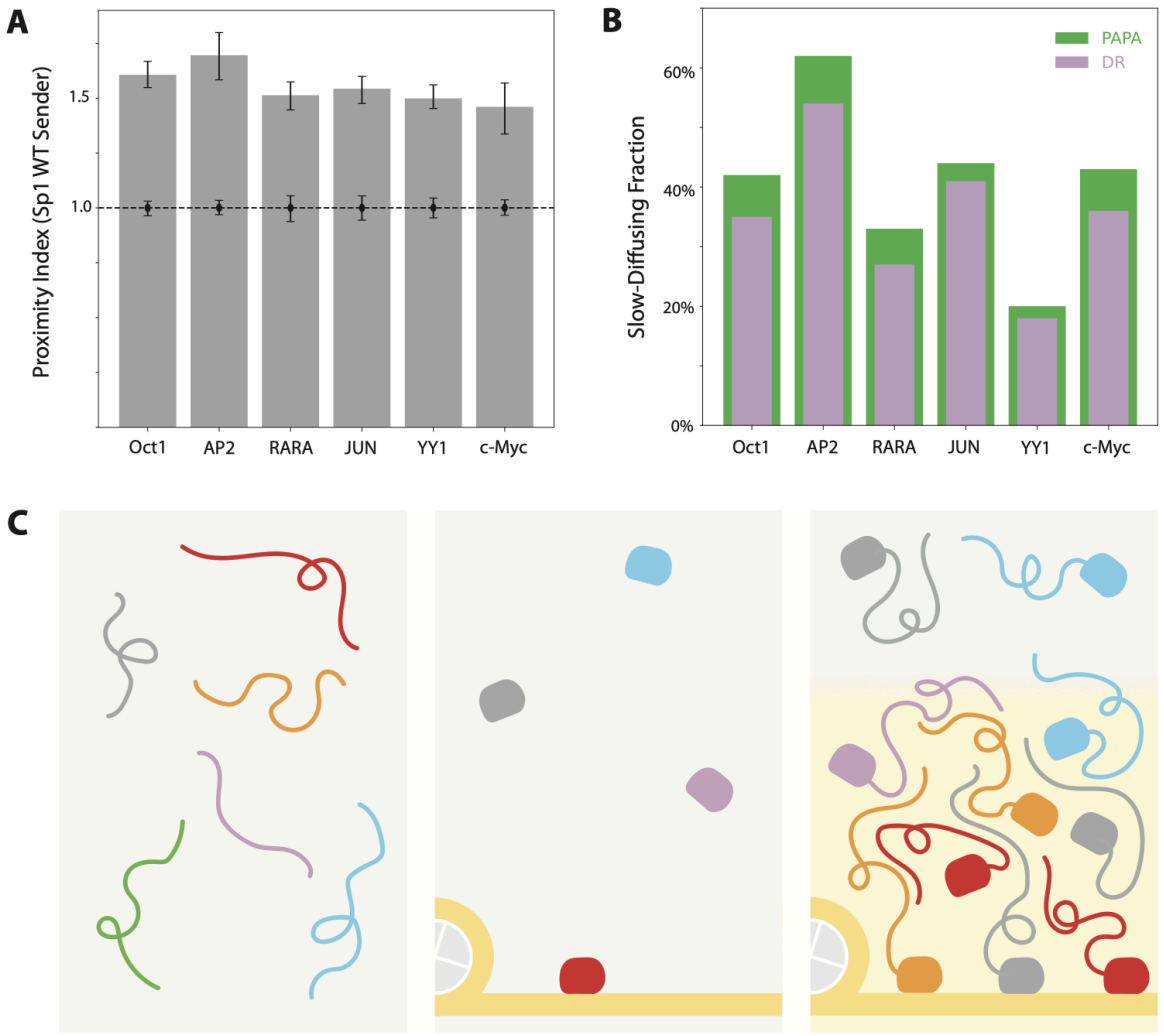

Figure 6. Sp1 shows extensive heterotypic interactions with other transcription factors.
(A) PAPA signal between endogenous Sp1 and SNAPf-tagged TFs expressed from transgenes.
(B) Slow-diffusing (< 0.1 μm²/s) fraction of PAPA/DR-reactivated molecules in (A).
(C) Model of TF binding to chromatin. Left: Freely diffusing intrinsically disordered regions interact with low probability at endogenous concentrations. Middle: Isolated DNA-binding domains interact weakly and nonspecifically with chromatin. Right: IDR-mediated interactions between TFs, scaffolded by DBD-DNA interactions, promote assembly of TF ensembles. Chromatin binding by a given TF is thus influenced both by direct DBD-DNA interactions and by IDR-IDR interactions with neighboring TFs.

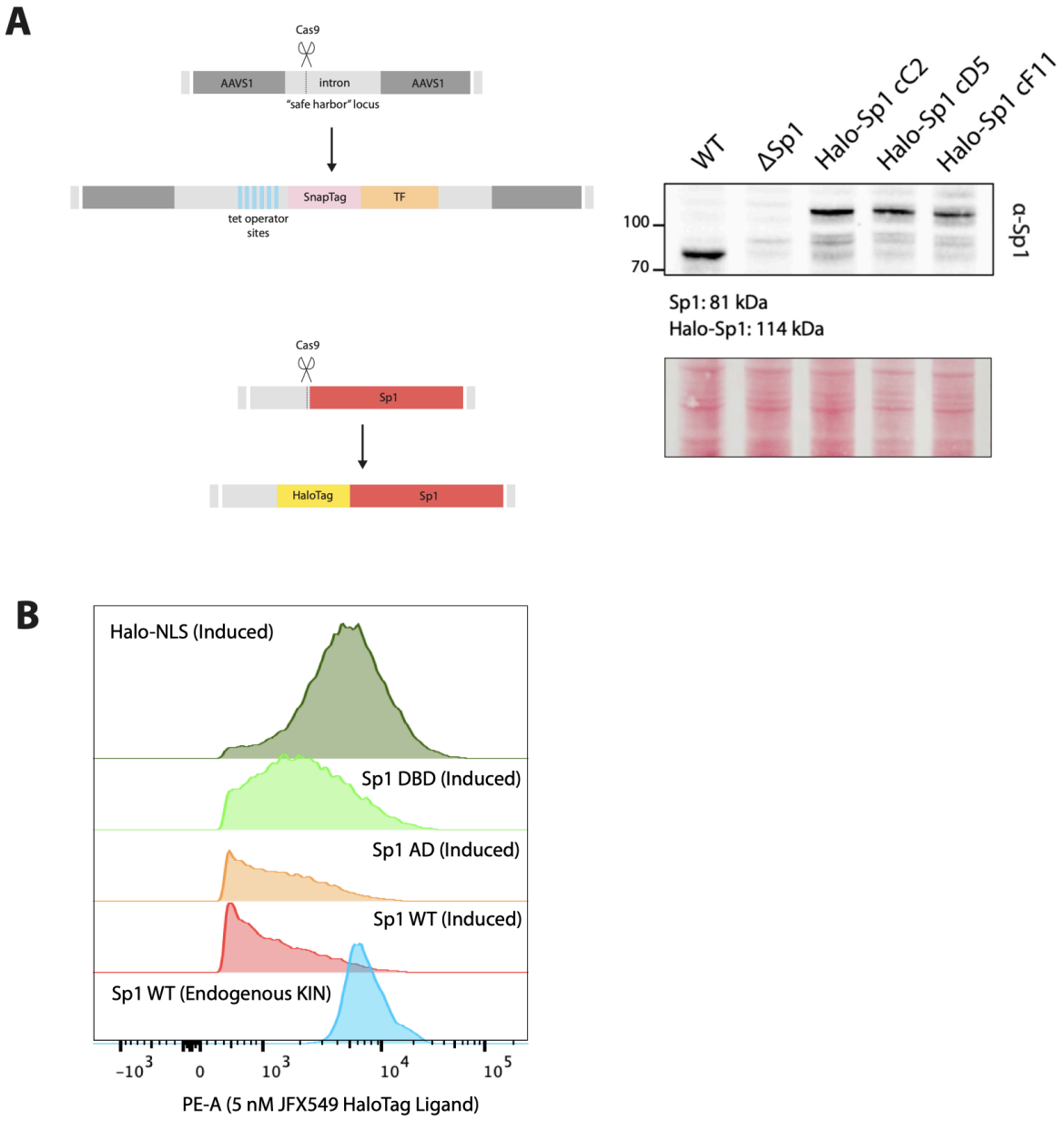

Figure S1. Engineering strategy for HaloTag and SNAPfTag fusions enables PAPA measurements at sub-physiological expression levels.
(A) CRISPR-Cas9 editing was used to engineer a HaloTag fusion at the Sp1 endogenous locus. SNAPfTag fusion proteins driven by an inducible promoter were integrated at the AAVS1 locus. Western blot shows correct band size for HaloTag knock-in as well as Sp1 knockout.
(B) Flow cytometry quantifies fluorescence for inducible transgene expression and endogenous HaloTag fusion.

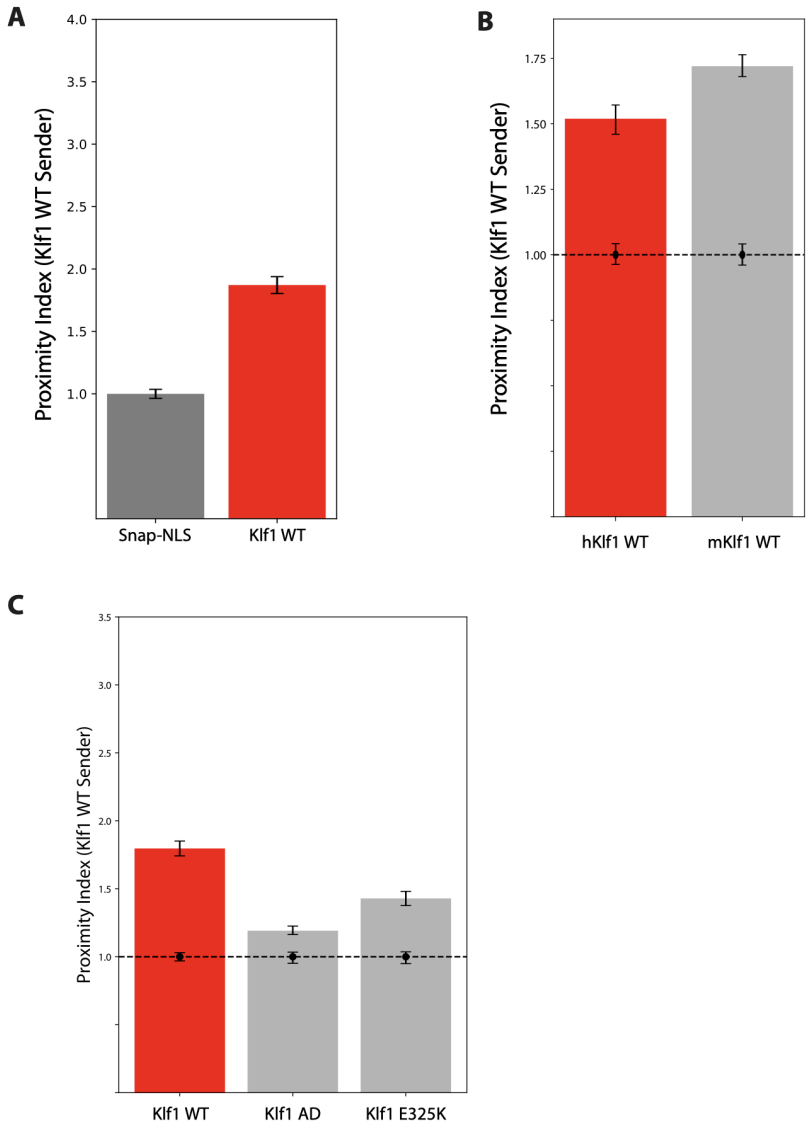

Figure S2. PAPA detects interactions between Klf1 and different Klf1 variants.
(A) PAPA signal between sender-labeled and receiver-labeled wild-type Klf1.
(B) PAPA signal between sender-labeled human Klf1 and receiver-labeled mouse Klf1.
(C) PAPA signal between sender-labeled human Klf1 and receiver-labeled N-terminal IDR (AD) of Klf1 or the E325K mutant.

Methods

Cell culture and cell line generation

U2OS cells were cultured at 37 °C with 5% CO_2 in high-glucose Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher #12800082) supplemented with 10% fetal bovine serum and penicillin/streptomycin. Cells were routinely tested for mycoplasma contamination by PCR.
To insert a HaloTag on endogenous *Sp1*, cells were co-transfected with a homology donor plasmid and a plasmid co-expressing Cas9 and an sgRNA. Cells were stained with a Janelia Fluor Halo ligand, and individual Halo-positive cells were flow-sorted into 96-well plates. Homozygous knock-in clones were validated by PCR, western blot, and sequencing.
For *Sp1* knockout, cells were co-transfected with plasmids expressing Cas9 and sgRNAs targeting sequences flanking the *Sp1* coding region. Venus-positive cells (reporting construct delivery) were flow-sorted into 96-well plates, and deletion clones were validated by PCR and western blot.

Doxycycline-inducible transgenes were inserted into the AAVS1 safe-harbor locus along with a linked puromycin resistance cassette. Following transfection, stable integrants were selected using medium supplemented with 1 µg/mL puromycin.

Oblique line-scan microscopy

We constructed an oblique line-scan microscope on a Nikon Ti2 E inverted motorized microscope stand equipped with a SR Plan Apo IR 60X NA 1.27 water immersion objective lens. Laser lines (405 nm, 488 nm, 560 nm, and 642 nm) were combined using dichroic mirrors, directed through an acousto-optic tunable filter (AOTF), and passed through a single-mode optical fiber using two achromatic FiberPorts (ThorLabs PAF2-A7A). A light sheet was generated using a Powell lens (Laserline Optics, LOCP-8.9R10-1.2) and swept over the sample using a 1D galvo mirror (Thorlabs GVS201). A field-programmable gate array (FPGA) controlled by custom LabView software was used to modulate the AOTF and sweep the light sheet in synchrony with the exposure of a Hamamatsu Orca Fusion BT camera running in lightsheet mode. Samples were kept at 37 °C with 5% CO_2 within an OKO Labs stage incubation chamber.

PAPA-SMT

Cells were plated in 35-mm MatTek glass-bottom dishes in phenol-red free DMEM and induced for 36 hours with 1 µg/mL doxycycline while also staining with 50 nM of JF549 Halo ligand and 50 nM of JFX650 SNAP ligand. Prior to imaging, cells were washed briefly in 1x phosphate-buffered saline, destained twice for 30 min with phenol-red free DMEM, and transferred to fresh phenol-red free DMEM (Gibco #31053036).

For OLS imaging, an automated script was used to acquire data across multiple fields of view per condition. Receiver fluorophores (JFX650) were shelved by illuminating the sample with 642 nm light at full power (2.2 W at the laser head). Movies were recorded under continued 642 nm illumination at a frame rate of 10 ms per frame. After acquiring 50 baseline frames, the sample was illuminated with either 405 nm light at full power (100 mW at the laser head) or 561 nm light at 50% AOTF (1.1 W at the laser head) for 50 frames, and 50 more frames were imaged to detect reactivated fluorophores. Receiver dyes were then re-shelved using a second

pulse of 642 nm light for 50 frames, and this sequence was repeated for three cycles, alternating between 405 nm (direct reactivation) and 561 nm (PAPA).

Laser output through the objective was measured prior to the experiment using a slide-mounted power sensor (Thorlabs S170C) at levels compatible with the sensor's detection range. Under these conditions, measured powers were 0.65 mW for 405 nm (full power), 22 mW for 561 nm (5% AOTF), and 30 mW for 640 nm (200 mW laser head power). These measurements served to validate the consistency of laser output across sessions. To verify measurement reproducibility, a positive control (Halo-Sp1 + SNAP-Sp1) and negative control (Halo-Sp1 + SNAP-NLS) were imaged at the beginning and end of each imaging session.
Klf1-expressing cells were imaged using highly inclined and laminated optical sheet (HILO) illumination on a custom microscope previously described by Hansen et al. (2018). Approximate laser power densities at the sample were 52 W/cm² for 405 nm (violet), 100 W/cm² for 561 nm (green), and 2.3 kW/cm² for 639 nm (red). Emission was collected through a Semrock 676/37 bandpass filter.

PAPA-SMT data analysis
Molecules were tracked as previously described using quot (https://github.com/alecheckert/quot), a Python toolkit for high-accuracy single-molecule localization and trajectory analysis. Nuclei were segmented using the Cellpose algorithm (Stringer et al., 2021), applied to 560 nm fluorescence images after Gaussian smoothing ($\sigma = 10$), with an estimated nuclear diameter of 80 pixels. Segmented regions smaller than 5,000 pixels in area were excluded, and adjacent or touching nuclei were separated using a custom boundary-drawing routine. Nuclear masks were overlaid on 642 nm intensity images to compute a per-nucleus leakage score, defined as the mean 642 nm signal in a 4-pixel ring surrounding the nuclear mask divided by the nuclear mean intensity. Cells with high leakage scores (>0.7) were excluded to eliminate those exhibiting cytoplasmic 642 nm signal. Additional filters excluded nuclei with atypically low or high detection counts and those contacting the top or bottom edge of the image field. Trajectory data were intersected with nuclear masks to assign detections to individual cells, and only cells passing all criteria were retained for analysis. Diffusion coefficient distributions were inferred using SASPT (https://github.com/alecheckert/saspt), a Bayesian state array framework for modeling heterogeneous mobility states from single-molecule trajectories.

To calculate proximity indices, detections were counted in each frame within segmented nuclear masks and grouped into four time windows per cycle: frames 0–49 (baseline preceding violet pulse), 50–99 (violet-induced direct reactivation, DR), 100–149 (baseline preceding green pulse), and 150–199 (green-induced proximity-assisted photoactivation, PAPA). To correct for spontaneous reactivation of dark-state Janelia Fluor dyes—which occurs independently of laser input and follows an exponential decay—baseline counts were fit to a single-exponential model and extrapolated into the corresponding post-pulse window. This projected background was subtracted from the observed detections, and negative values were truncated to zero. The total number of background-corrected green detections (PAPA) was divided by the number of background-corrected violet detections (DR) to yield a green-to-violet (GV) ratio. Each

experimental GV ratio was normalized to the GV ratio of a non-interacting SNAPf-3xNLS negative control to compute the final proximity index. A proximity index of 1 indicates no increase in proximity-induced signal above background. To estimate statistical uncertainty, GV ratios were recalculated using bootstrap resampling of individual cells (100 replicates per condition). Error bars represent the 2.5th and 97.5th percentiles of the resulting bootstrap distribution. Uncertainty in the control condition is shown as a fixed error bar at a proximity index of 1.

To visualize the relationship between DR and PAPA-reactivated molecules at the single-cell level, we plotted background-corrected PAPA detections (green channel) against DR detections (violet channel) for each cell across three imaging cycles. For each pulse type, background was estimated as the number of detections in the 50 frames preceding the pulse, and subtracted from the number observed in the 50 frames following the pulse. Cells with low detection counts (<100) or negative corrected values were excluded. A linear regression was performed without intercept to quantify the proportionality between green and violet signals, and 95% confidence intervals for the fitted line were estimated by bootstrapping cell samples (500 replicates). This analysis provides an independent estimate of PAPA efficiency across experimental conditions and highlights the heterogeneity of single-cell responses.

Fluorescence recovery after photobleaching (FRAP)
FRAP experiments were performed using a Zeiss LSM 900 confocal microscope. Cells labeled with fluorescent dye were imaged using a 561 nm laser and a 150 µm pinhole. For each measurement, the region of interest was set to encompass the entire nucleus, and imaging was conducted at the maximum scan speed permitted by the frame size, with bidirectional scanning and 8-bit depth. Image sequences were acquired with a 150 ms interval between frames for a total of 600 frames (90 seconds). Photobleaching was applied on frame 16, following a 15-frame pre-bleach baseline period used for normalization. 20 movies were acquired per condition, and each condition was imaged in two independent experiments conducted on separate days.

FRAP data analysis
FRAP movies were analyzed using a custom Python pipeline (https://github.com/vinsfan368/FRAPanalysis). A maximum intensity projection of each movie was used to generate a nuclear mask, which was smoothed and thresholded to segment the nucleus. The position of the bleach spot was extracted from CZI file metadata and corrected for lateral drift using 2D image cross-correlation.
For each frame, the mean fluorescence intensity within the bleach spot was measured. Background signal was estimated as the median pixel intensity outside the nuclear mask. Recovery was calculated as the background-subtracted intensity ratio of the bleach spot to the full nucleus. This ratio was normalized to the average ratio over the pre-bleach frames (frames 1–15), such that full recovery corresponds to a normalized value of 1.

Western blotting
Cells were lysed directly in SDS sample buffer by pelleting approximately 1×10^6 cells from a six-well plate, resuspending in 100 µL of SDS buffer, and heating at 95 °C for 10 minutes.

Lysates were resolved by SDS-PAGE and transferred to nitrocellulose membranes using wet transfer at a constant current of 90 mA overnight in transfer buffer (15 mM Tris-HCl, 20 mM glycine, 20% methanol, 0.0375% SDS). Membranes were briefly stained with Ponceau S to verify transfer, then blocked for 1 hour at room temperature in TBS containing 5% bovine serum albumin (BSA) with gentle agitation. Primary antibody incubation was carried out overnight at 4 °C in TBST containing 5% BSA. To detect Sp1, membranes were incubated with anti-Sp1 mouse monoclonal antibody (Santa Cruz Biotechnology, sc-17824). After three washes in TBST, membranes were incubated for 1 hour at room temperature with an HRP-conjugated anti-mouse secondary antibody (1:5000 dilution in 5% BSA in TBST). Blots were then washed three times in TBST with brief (~2-minute) incubations under agitation prior to chemiluminescent detection.

Acknowledgments
Thanks to Eric Betzig for sharing a home-built OLS microscope design, Amir Hay, Ziyuan Chen, and Gokul Upadhyayula for help with microscope construction and troubleshooting, and Dan Milkie for assistance with software. This work was supported by early-stage investigator developmental funds from NIH grant RM1GM139738 (to T.G.). R.T. is an Investigator of the Howard Hughes Medical Institute. We would like to thank the Tjian-Darzacq group and the members of our RM1 collaboration team for helpful discussions.

References

Alberti, S., Gladfelter, A., & Mittag, T. Considerations and challenges in studying liquid–liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).

Arnaud, L. et al. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am. J. Hum. Genet.* **87**, 721–727 (2010).
Biggin, M. D. Animal transcription networks as highly connected, quantitative continua. *Dev. Cell* **21**, 611–626 (2011).

Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).

Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729–736 (1985).

Brodsky, S. et al. Intrinsically disordered regions direct transcription-factor in vivo binding specificity. *Mol. Cell* **79**, 459–471.e4 (2020).

Chen, Y. et al. Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. *eLife* **11**, e75064 (2022).

Chong, S. et al. Imaging dynamic and selective low-complexity-domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).

Dahal, L. et al. Surprising features of nuclear receptor interaction networks revealed by live-cell single-molecule imaging. *eLife* **12,** e92979 (2025).

Driouchi, A. et al. Oblique line-scan illumination enables expansive, accurate and sensitive single-protein measurements in living cells. *Nat. Methods* **22**, 559–568 (2025).

Gera, T., Jonas, F., More, R. & Barkai, N. Evolution of binding preferences among whole-genome-duplicated transcription factors. *eLife* **11**, e73225 (2022).

Graham, T. et al. Proximity-assisted photoactivation enables nanometer-scale mapping of transient protein interactions in living cells. *eLife* **11**, e76870 (2022).

Graham, T. G. W. et al. Single-molecule live imaging of subunit interactions and exchange within cellular regulatory complexes. *bioRxiv* 10.1101/2024.06.25.600644 (2024).

Hollenhorst, P. C. et al. DNA-sequence specificity determinants associate with distinct transcription-factor functions. *Genome Res.* **19**, 2090–2100 (2009).

Horton, C. A. et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science* **381**, eadd1250 (2023).

Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**, 1566–1570 (1988).

Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).

Martin, E. W. et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).

McSwiggen, D. T. et al. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).

Oksuz, O. et al. Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, 2449–2463.e13 (2023).

Sigler, P. B. Transcriptional activation: acid blobs and negative noodles. *Nature* **333**, 210–212 (1988).

Slattery, M. et al. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).

Stormo, G. D. Modeling the specificity of protein–DNA interactions. *Quant. Biol.* **1**, 115–130 (2013).

Tummala, H. et al. A frameshift variant in SP1 triggers superactivation of Sp1-mediated transcription in familial bone marrow failure. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17151–17155 (2020).

Wei, M.-T. et al. Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, 1187–1196 (2020).