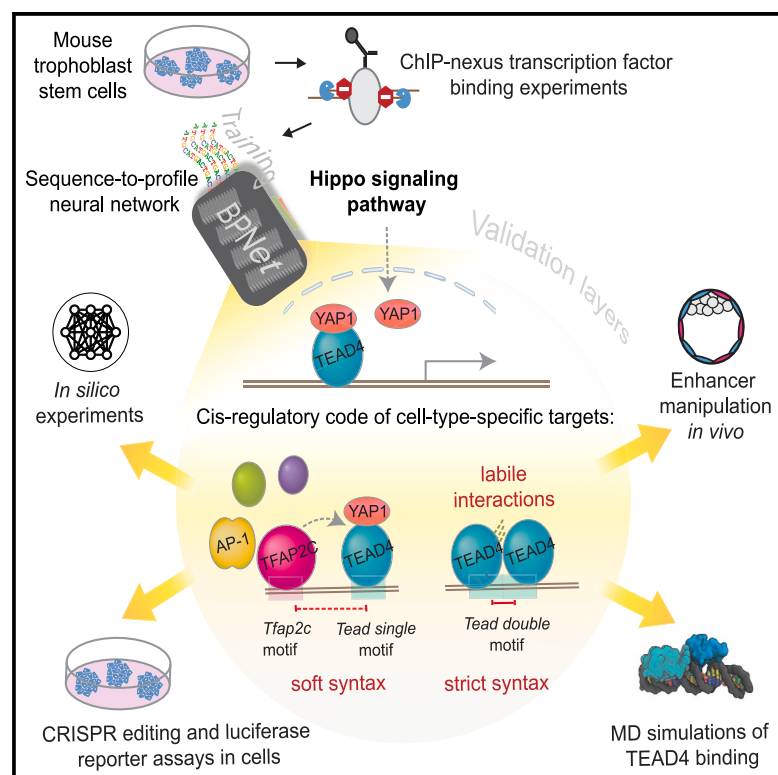


Interpreting regulatory mechanisms of Hippo signaling through a deep learning sequence model

Graphical abstract



Authors

Khyati Dalal, Charles McAnany, Melanie Weilert, Mary Cathleen McKinney, Sabrina Krueger, Julia Zeitlinger

Correspondence

jbz@stowers.org

In brief

Dalal et al. show how AI can be used to systematically uncover the sequence rules that guide how signaling pathways interact with DNA inside cells. This work provides a new lens for discovering the cooperative mechanisms that shape cell-type-specific gene expression in response to cellular signaling.

Highlights

- Interpretable deep learning uncovers *cis*-regulatory rules downstream of Hippo signaling
- YAP1 binding serves as a proxy for predicting cell-type-specific enhancer activity
- TFAP2C enhances TEAD4/YAP1 binding within nucleosome range, dependent on distance
- Two TEAD4s cooperate on the *Tead double* motif through transient, labile interactions



Article

Interpreting regulatory mechanisms of Hippo signaling through a deep learning sequence model

Khyati Dalal,^{1,2} Charles McAnany,¹ Melanie Weilert,¹ Mary Cathleen McKinney,¹ Sabrina Krueger,¹ and Julia Zeitlinger^{1,2,3,*}¹Stowers Institute for Medical Research, Kansas City, MO, USA²Department of Pathology & Laboratory Medicine, The University of Kansas Medical Center, Kansas City, KS, USA³Lead contact*Correspondence: jbz@stowers.org<https://doi.org/10.1016/j.xgen.2025.100821>

SUMMARY

Signaling pathway components are well studied, but how they mediate cell-type-specific transcription responses is an unresolved problem. Using the Hippo pathway in mouse trophoblast stem cells as a model, we show that the DNA binding of signaling effectors is driven by cell-type-specific sequence rules that can be learned genome wide by deep learning models. Through model interpretation and experimental validation, we show that motifs for the cell-type-specific transcription factor TFAP2C enhance TEAD4/YAP1 binding in a nucleosome-range and distance-dependent manner, driving synergistic enhancer activation. We also discovered that *Tead double* motifs are widespread, highly active canonical response elements. Molecular dynamics simulations suggest that TEAD4 binds them cooperatively through surprisingly labile protein-protein interactions that depend on the DNA template. These results show that the response to signaling pathways is encoded in the *cis*-regulatory sequences and that interpreting the rules reveals insights into the mechanisms by which signaling effectors influence cell-type-specific enhancer activity.

INTRODUCTION

Signaling pathways are critical for cell fate decisions during development, the generation of cell types *in vitro*, and therapeutic interventions, which often target-specific signaling pathways.¹ The signaling components and transcription factors (TFs) that function as canonical effectors of a pathway are typically well studied. Such signaling effectors recognize specific DNA sequence motifs either directly or find obligate partner TFs with DNA binding specificity. However, signaling pathways are reiteratively used during development, and which *cis*-regulatory sequences are bound by these TFs and become active enhancers that regulate target genes is highly complex and poorly understood.^{2,3} Thus, signaling pathways are critical for gene regulation, but their target specificity is one of the least understood areas of enhancer biology, making it difficult to predict the activity of enhancers or modify their function during development through targeted mutations.^{3–7}

Here, we hypothesized that the binding of signaling effectors is encoded in the *cis*-regulatory sequences. This hypothesis is supported by studies on individual enhancers showing that the target gene specificity of a signaling pathway depends on cell-type-specific TFs.^{2,8} Furthermore, genetic experiments showed that cell-type-specific TFs help determine where signaling effectors bind.^{9–13} However, there is no systematic approach to capture the genome-wide *cis*-regulatory sequence rules, identify relevant cell-type-specific TFs, and characterize the molecular mechanisms by which they influence enhancer activation.¹⁴

Capturing genome-wide *cis*-regulatory sequence rules in a cell-type-specific manner is an ideal task for a deep learning model. By learning to predict TF binding data from DNA sequence, convolutional neural networks learn complex sequence rules in an unbiased way, and the rules can be extracted from the trained models using interpretation tools.^{15–21} A key advantage is the inherent predictive accuracy and the genome-wide applicability of the learned sequence rules, something that cannot be achieved by studying motifs on individual enhancers experimentally or by analyzing motif compositions using traditional methods.

Predictive accuracy is the first step during training, where the model learns to predict TF binding profiles from sequence. The genome-wide applicability is ensured by testing the performance on withheld data that the model has not seen during training. Another advantage is that the sequence rules are learned *de novo* in an inherently combinatorial way on large amounts of data. Only upon achieving accurate predictions can the model be interpreted to extract the learned rules. Such models not only learn TF motifs, but also measurements of their relative affinities and the syntax rules, and thus the distance relationships by which motifs cooperate with each other.^{15,17,20,22,23}

We reasoned that interpretable deep learning might identify which cell-type-specific TF motifs contribute to the response to signaling, and that the motif syntax rules could pinpoint to potential molecular mechanisms. TF cooperativity is often studied with the assumption that TFs interact with each other through protein-protein interactions, exemplified by the pioneering



work of the interferon-beta enhanceosome.^{24–26} For such protein-protein interactions to occur, two motifs have to be well positioned at close distances, which should be reflected in the syntax rules. On the other hand, such strictly spaced motifs are not frequently observed in the genome, raising the question of whether TF binding cooperativity often occurs through more flexible motif syntax.^{27–29} Indeed, deep learning models suggest the existence of a soft motif syntax, which occurs at variable motif distances within ~150 bp, with stronger TF cooperativity predicted at closer distances.^{15,20}

Whether such syntax rules exist for signaling effectors has previously been difficult to decipher. ChIP-seq binding data tend to be of low resolution and display low levels of signal when the TF binds indirectly through a partner TF.^{10,11,13} Likewise, individually manipulating enhancer sequences *in vivo* limits throughput, and the effects can be difficult to interpret since they may be enhancer-specific or caused by the inadvertent disruption of other important sequences.^{30,31} Large-scale reporter assays, on the other hand, have produced conflicting results on whether motif syntax is important and, if so, their effects have only been measured for short motif distances and not been mechanistically analyzed.^{16,27,32–37} For these reasons, TF binding cooperativity downstream of signaling pathways has not been systematically studied from a sequence perspective.

To test whether the binding of signaling effectors is sequence encoded and follows syntax rules, we performed TF binding experiments at the highest resolution and leveraged our previously developed deep learning model BPNet to predict the data at base resolution from genomic sequences of 1 kb.^{15,20,22,38,39} This approach has high predictive accuracy and optimally resolves sequence rules between closely spaced motifs within enhancers.¹⁵ Since the model does not predict enhancer activity or target genes, we evaluated and validated these downstream aspects using traditional methods.

As a model system for our approach, we studied the Hippo signaling pathway in mouse trophoblast stem cells (TSCs). Hippo signaling is critical for specifying trophoblast versus inner cell mass cell fate in the early mouse embryo.^{40–45} When cells of the embryo sense that they are facing the outside, i.e., less cell density, they polarize and inactivate the Hippo pathway. This causes YAP1/TAZ to translocate to the nucleus and bind to TEAD4, which, like all TEAD family members, binds to a consensus *Tea*d motif.^{41,43,45,46} Other TFs known to be important for TSC identity include CDX2, TFAP2C, and GATA3.^{47–54} TSCs are therefore an ideal system to dissect the interactions between Hippo signaling effectors and cell-type-specific TFs in enhancer activation.

Here, we show that the binding of the signaling effectors YAP1/TEAD4 is specified by *cis*-regulatory sequence rules that apply genome-wide. We identified thousands of novel active enhancers in TSCs and show that their activity is driven by YAP1, which binds DNA with the help of cell-type-specific TFs such as TFAP2C. While the *Tfap2c*-*Tea*d motif synergy follows a soft motif syntax, we also identified the *Tea*d *double* motif as mediating strong cooperativity through strict syntax. Based on our molecular dynamics (MD) simulations, the strict distance between the two *Tea*d motifs is required for two TEAD4 to engage in labile protein-protein interactions, suggesting that the TEAD4

effector complex is assembled on DNA with the help of Hippo signaling. This demonstrates how deep learning models can uncover precise sequence rules and potential mechanisms by which signaling effectors bind in the genome to produce cell-type-specific effects.

RESULTS

A deep learning model reveals combinatorial binding motifs for Hippo TFs

We generated genome-wide, high-resolution binding data for two Hippo signaling effectors (TEAD4 and YAP1) and for potential TSC-specific partner TFs (CDX2, TFAP2C, and GATA3) by using a ChIP-exo technique called ChIP-nexus,³⁸ in which an exonuclease step generates narrow and sharp binding footprints (Figures 1A and S1A). We used TSCs derived from mouse blastocysts⁵⁵ and confirmed that they retain features of endogenous trophoblast (TE) cells by reintegrating them into the TE layer of blastocyst embryos in an aggregation assay (Figure S1B). The ChIP-nexus binding data revealed that YAP1 and TEAD4 were more correlated with each other than any other TF pair (Figure 1B), consistent with YAP1 binding to DNA through TEAD4.^{56–58}

We then trained the deep learning model BPNet to predict the base-resolution binding profiles of all TFs from 187,775 reproducibly bound genomic regions (Figure 1C) by separating chromosomes into training, validation, and test groups to confirm model accuracy. We also performed cross-validation on different chromosome combinations to ensure model stability (Figure S1F). For all TFs, we obtained high prediction accuracy for the read counts, as well as profiles and footprint positions on a par with the similarity between replicate experiments (Figures 1D and S1C–S1E). Examples include the putative enhancers of important trophoblast genes such as *Fgfr1*^{43,59} (Figure 1D), *Amotl2*, *Pard3b*, and *Krt8/18*^{60–62} (Figures S1G and S1I). These results show that the model learned general rules for this cell type to predict TF binding anywhere in the genome from sequence alone.

To understand which motifs contribute to the binding of the signaling effectors TEAD4 and YAP1, we extracted *de novo* learned motifs. Using an attribution method,⁶³ we assigned contribution scores to all bases in the input sequences and then summarized learned motifs as a contribution weight matrix (CWM).⁶⁴ The CWM motifs were then used to map “contributing” motif instances in each genomic region (Figure 1C). These mapped motifs were highly congruent with experimentally observed and predicted TF footprints (Figures 1D and S1G–S1I).

Among the discovered motifs were the known consensus motifs of the profiled TFs (Figure 1E) and two unexpected motifs: a strictly spaced *Tea*d *double* motif and a strictly spaced *Gata3* *double* motif (Figure 1E). These motifs are directly bound by TEAD4 and GATA3, respectively, as confirmed by the sharpness of the ChIP-nexus footprints (Figure 1F). YAP1 also showed sharp binding footprints on the *Tea*d motifs, suggesting a tight physical association between YAP1 and TEAD4 on DNA. Given YAP1’s stronger dependency on Hippo signaling, we focused on understanding how YAP1 binding is influenced by other TFs. Notably, the model learned motifs for TFs that we did not profile, including JUN-FOS (AP-1), CTCF, and ELF5 (Figure S1F).

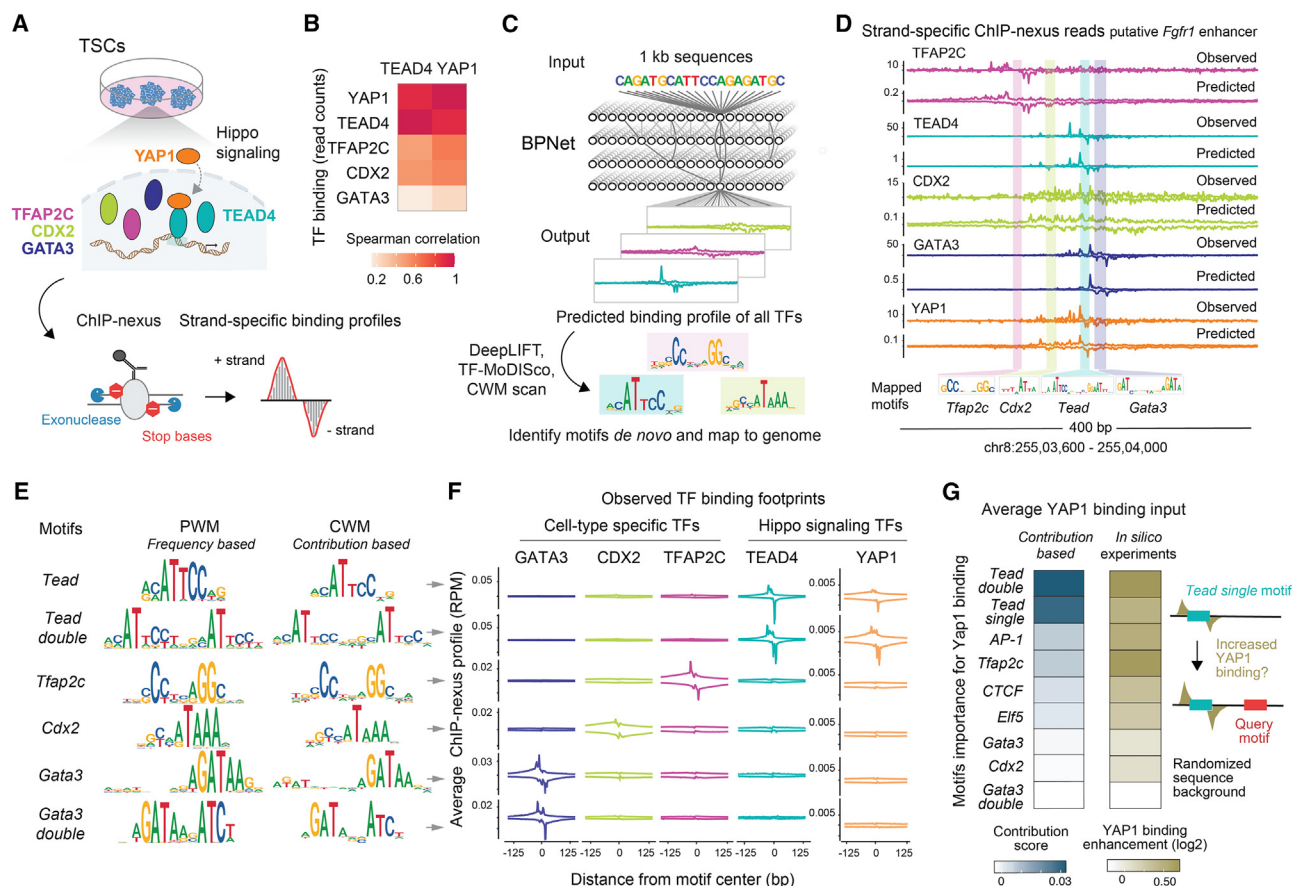


Figure 1. BPNet suggests combinatorial binding motifs for Hippo TFs

(A) Experimental design to map high-resolution binding of signaling and cell-type-specific TFs in mouse TSCs.

(B) Spearman correlations of the ChIP-nexus read counts between TFs at non-promoter binding regions show that YAP1 and TEAD4 binding are highly correlated.

(C) Schematic of the multi-task BPNet model, trained to predict ChIP-nexus TF binding from DNA sequence, and the interpretation tools that identify and map contributing motifs for each TF.

(D) At the *Fgf1* enhancer (mm10-chr8:25503600-2550400) not seen during model training, observed and BPNet-predicted base-resolution binding are visibly similar for each TF (+, strand on top; -, strand below). Quantifications of this similarity by Jensen-Shannon distance (0 = perfect concordance, 1 = no similarity, shown globally in Figure S1D) give values of 0.31 (TFAP2C), 0.32 (TEAD4), 0.32 (CDX2), 0.30 (GATA3), and 0.43 (YAP1). BPNet-mapped motifs are shown below.

(E) Learned motifs are shown as frequency-based position weight matrix (PWM) and contribution weight matrix (CWM), where the base height reflects the contribution to the TF binding predictions.

(F) Average ChIP-nexus binding footprints of all TFs at BPNet-mapped motifs, shown as reads per million (RPM), with values on + strand on top and - strand below. Sharp footprints typically indicate direct binding of the TF to the motif. YAP1 has sharp footprints on the *Tead* motifs despite binding indirectly.

(G) The importance of motifs toward YAP1 binding was assessed by two interpretation methods, one based on the contribution scores from genomic instances (blue on the left) and one based on rules derived *in silico* without genomic context (olive on the right). Contribution scores are derived by DeepLIFT relative to a dinucleotide shuffled control sequence on a log scale, thus have no unit. In the second method (right plot), motifs are injected many times into a randomized background that contains a *Tead single* motif (ACATTCTG) within 150 bp. The average predicted YAP1 binding enhancement over no added query motif is calculated.

We then used two complementary model interpretation methods¹⁵ to measure whether and how much, on average, each motif influences YAP1 binding (Figure 1G). The first method uses the contribution scores of mapped motifs from genomic regions (Figure 1G, left) and measures each motif's influence across all observed instances *in vivo*. The second method uses *in silico* experiments to measure each motif's influence in a randomized sequence background (Figure 1G, right), and thus tests the learned rules in isolation without the complexity of genomic

sequences. We injected a *Tead single* motif with or without another motif and let the model predict how much any given motif enhanced the binding of YAP1 to the *Tead single* motif. Both methods revealed similar results, providing internal validation of our model interpretation.

The *Tead* motifs, which we will refer to as *Tead single* and *Tead double* motifs, were both important as expected, but the *AP-1* and *Tfap2c* motifs also had a sizable contribution to YAP1 binding (Figure 1G). *AP-1* is present in many cell types and has

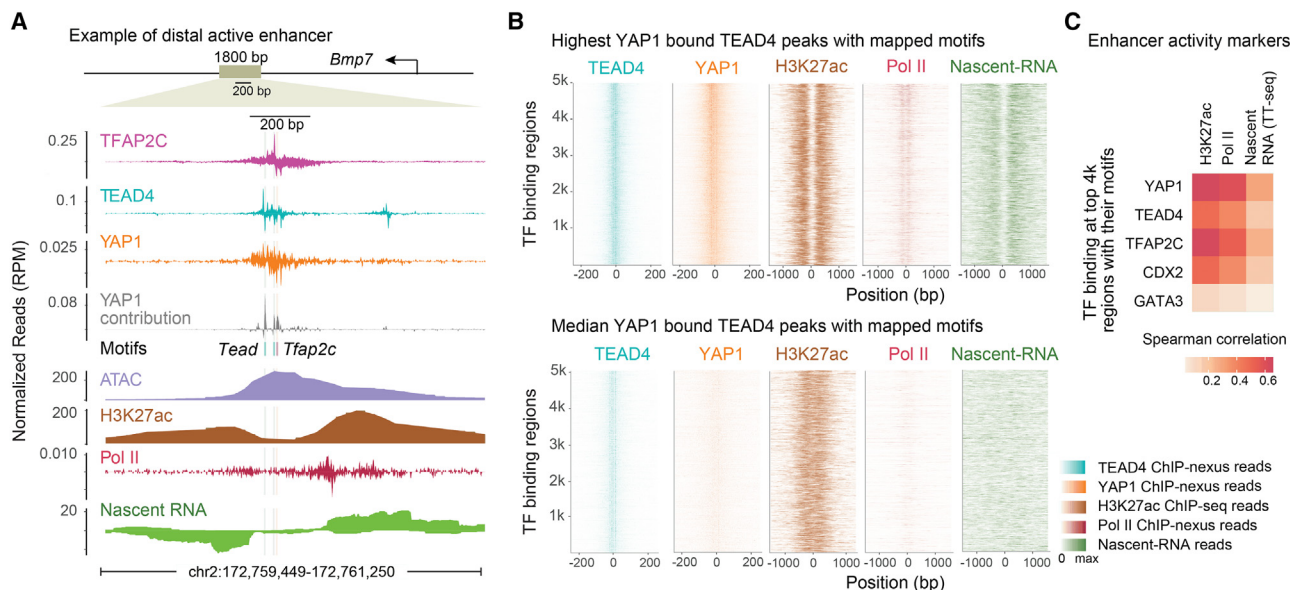


Figure 2. YAP1 binding correlates with enhancer activity markers

(A) An example of an active enhancer ~100 kb downstream of the *Bmp7* gene, showing ChIP-nexus TF binding for TFAP2C, TEAD4, and YAP1 alongside BPNet-mapped motifs *Tead* and *Tfap2c* and predicted YAP1 binding contribution. Additional tracks are the fragment coverage for ATAC-seq, H3K27ac ChIP-seq, Pol II ChIP-nexus data, and nascent RNA-seq derived from TT-seq.

(B) Profile heatmaps of TEAD4 and YAP1 ChIP-nexus data at the 5,000 TEAD4 peaks with the highest YAP1 binding (top) and 5,000 peaks with median YAP1 binding (bottom). Regions with highest YAP1 have an active enhancer signature of H3K27ac ChIP-seq, Pol II ChIP-nexus, and Nascent-RNA reads (top).

(C) A heatmap depicting Spearman correlations between ChIP-nexus TF binding and the enhancer activity markers at the top 4,000 non-promoter peaks containing their motif. YAP1 correlates best, followed by TFAP2C.

previously been shown to cooperate with TEAD and YAP in cancer cell lines,^{65–68} confirming the model's predictions. However, we also identified a strong contribution from TFAP2C, which is critical for specifying TE.^{47,48,54,69–71}

The strong contribution of TFAP2C suggests that Hippo signaling is influenced by cell-type-specific TFs as hypothesized. Interestingly though, the model did not assign all cell-type-specific TF motifs the same importance (Figure 1G). For example, CDX2 and GATA3 were not predicted to help YAP1 bind, although they are critical for trophoblast identity.^{50–52} This suggests that the rules by which TFs boost the binding of signaling effectors are not obvious, but that, with the help of TF binding data, these rules can be learned with a deep learning model.

YAP1 binding correlates with markers of enhancer activity

Having analyzed TSC-specific YAP1 binding, we investigated whether high YAP1 binding levels are indicative of enhancer activation. We expect YAP1 to be a strong activator based on previous molecular evidence,^{72–74} but many other TFs have transactivation domains, and thus it is unclear how much YAP1 contributes to enhancer activation at a genome-wide level.

Since no individual assay unambiguously measures enhancer activity,⁷⁵ we performed experiments in TSCs to profile multiple markers of active enhancers: ChIP-nexus for RNA polymerase II (Pol II), TT-seq to capture enhancer transcription, ATAC-seq to measure chromatin accessibility, and ChIP-seq for H3K27ac

found on nucleosomes flanking active enhancers^{76–78} (Figure S2A). We identified thousands of enhancers that showed enhancer transcription, chromatin accessibility, and H3K27ac, thus meeting our definition of being active enhancers (Table S4). We selected a few enhancers near important trophoblast genes for further characterization. Named after the nearest gene, these include a *Bmp7*, *Rin3*, *Ezr*, *Cited2*, *Amotl2*, *Bmp7*, *Dst*, and *Tjp1* enhancer. They were validated by cloning the minimal central region into a luciferase reporter assay and measuring their activity in TSCs (Figure S2B).

These data revealed that active enhancers are indeed associated with high levels of YAP1 binding (Figures 2A and S2E–S2H). As an example, strong binding footprints of TFAP2C, TEAD4, and YAP1 are found at the putative downstream *Bmp7* enhancer, and the contribution scores show that BPNet used the *Tfap2c* and *Tead* single motifs to predict YAP1 binding (Figure 2A). This region possesses all the characteristic features of active enhancers, with central ATAC-seq accessibility, flanking H3K27ac signal, Pol II occupancy, and bidirectional nascent RNA transcription (Figure 2A).

To examine the global correlation between YAP1 binding and the markers of enhancer activity, we selected all TEAD4 peaks with a *Tead* motif; we then compared the 5,000 regions with the highest YAP1 binding to 5,000 regions with median levels of YAP1 binding (Figure 2B). The top YAP1 bound regions showed strong H3K27ac signal, Pol II binding, and nascent transcription adjacent to the central region, while no strong evidence of enhancer activity was observed for the more lowly bound set

(Figure 2B). To quantify each TF's effect, we calculated the pairwise correlation between each TF's binding and enhancer activity markers (H3K27ac, Pol II, and nascent RNA) (Figure 2C). Among all TFs, YAP1 binding correlated best with the enhancer activity markers. Given the strong transactivation potential of YAP1,^{72–74,79} we conclude that YAP1 binding is an important determinant for enhancer activation in TSCs, and hence predicting YAP1 binding should serve as a proxy for predicting enhancer activity.

Enhancer activation involves DNA-distance-dependent TF cooperativity

If YAP1 binding occurs cooperatively and promotes enhancer activation, the cooperating motifs might activate transcription synergistically. Synergistic activation by two motifs has been documented,^{16,80–82} but the mechanisms are not clear and could vary. We focused on the *Tead single* and *Tfap2c* motifs since this motif pair is frequently found at active enhancers (Figure 2). In addition, genes near these active enhancers are enriched for cell fate commitment and GTPase regulation (Figure S2C), consistent with previous studies.^{43,51,54,69} Synergistic activation could occur if the *Tead single* and *Tfap2c* motif not only contribute to activation on their own but cause additional activation by promoting YAP1 binding. If so, the activity of such an enhancer would depend on BPNNet-learned syntax rules for YAP1 binding and thus may depend on the distance between the motifs.

To test for synergistic activation, we performed luciferase assays using the 200 bp minimal *Bmp7* enhancer, which has a *Tead single* and *Tfap2c* motif (shown in Figure 2A). To perturb combinations of these motifs in a controlled way while reducing the chance of introducing unknown variables, we used BPNNet: for each motif, we mutated the two bases that contributed most to the predictions and tested whether this led to decreased YAP1 binding (Figure S2D).

Luciferase assays showed that mutating either motif alone was sufficient to strongly reduce the activity while mutating both almost completely abolished the activity (Figure 3A). Thus, in the presence of one motif, the putative *Bmp7* enhancer produced only moderate activity, while together, they resulted in activity that exceeded the sum of each motif's effect. These results show that the *Tead single* and *Tfap2c* motifs mediate activation synergistically, presumably at least in part by increasing YAP1 binding.

To test whether the *Tfap2c* motif boosts YAP1 binding in a distance-dependent manner, we examined the contribution scores of all *Tfap2c* motifs found near a *Tead single* motif. This showed that *Tfap2c* motifs had significantly higher scores in regions where the motif was within nucleosome distance (<150 bp) and showed even higher scores when closer (Wilcoxon test $p < 2e-16$, Figures 3B and S3A). When we plotted these contribution scores in the genomic regions directly, they visually decreased with further distances between the *Tead single* motif and the *Tfap2c* motif (Figures 3C, S3C, and S3D). This suggests that TFAP2C enhances YAP1 binding in a distance-dependent manner through soft motif syntax.

We next investigated whether TFAP2C directly helps the recruitment of YAP1 or whether the effect on YAP1 binding is mediated through increased TEAD4 binding (Figure 3D). To

distinguish between these possibilities, we performed *in silico* experiments with randomized sequences and analyzed the binding enhancement of TEAD4, YAP1, and TFAP2C with different motif distances. This revealed that YAP1 and TEAD4 binding both depend on the distance of the nearby *Tfap2c* motif, causing an over 2.5-fold increase in binding of both when the *Tfap2c* motif is close (Figures 3D and S3B). Notably, the reverse was not necessarily true: TFAP2C binding was not substantially increased (<1.5-fold) in the presence of a nearby *Tead single* motif (Figure 3D), but showed some increase in the presence of a *Tead double* motif (Figures S3B and S3D). Given that binding cooperativity is usually assumed to be mutual,⁸⁴ this directionality is surprising but consistent with previous observations of soft motif syntax.¹⁵

To validate the *Tead single-Tfap2c* soft motif syntax, we performed luciferase reporter experiments on the putative *Rin3*, *Dst*, and *Adcy7* enhancers. Using BPNNet predictions as a guide for designing experiments, we changed the distances between the *Tead single* and *Tfap2c* motifs at three independent regions, by deleting the *Tead single* motif through minimal mutations and introducing the same motif at a different location (Figures 3F and S3E). In all three cases, the reporter activity of the enhancer changed in the expected direction. For example, when we moved the *Tead single* motif in the minimal *Rin3* enhancer further away from the *Tfap2c* motif (from 20 bp away to 60 bp away), BPNNet predicted lower TEAD4 binding (Figure 3E). This lower binding mirrored the lower activity measured in the luciferase assay (Figure 3F). Moving the two motifs closer to each other increased the luciferase reporter activity of the minimal *Dst* and *Adcy7* enhancers (Figure S3E).

To confirm that these distance effects are also observed in the genomic context, we performed CRISPR-Cas9-induced mutations in TSCs using homologous recombination on the endogenous *Rin3* enhancer (Figures S3F and S3G). ChIP experiments on this edited cell line confirmed the reduced TEAD4 binding, H3K27ac, and Pol II levels at the *Rin3* enhancer (Figure 3G), while other enhancers remained unchanged (Figure S3H). This demonstrates that changing the distance between motifs through controlled minimal mutations measurably affects enhancer activity markers in an *in vivo* endogenous context.

Taken together, we have identified genome-wide rules by which cell-type-specific TF motifs can enhance the binding of the Hippo signaling effectors. We validated that *Tfap2c* motifs enhance TEAD4/YAP1 through soft motif syntax in a distance-dependent manner, resulting in synergistic enhancer activation. This could represent a general mechanism for how cell-type-specific TFs boost the activity of signaling pathways.

The *Tead double* motif is a canonical element of the Hippo pathway

So far, we have focused on the genome-wide rules and mechanisms by which cell-type-specific TFs influence Hippo pathway activity. To push the boundaries of what can be learned with our approach, we next asked whether we could also discover novel molecular details of the Hippo pathway effectors themselves, which are mechanistically well studied.

Notably, BPNNet discovered the strictly spaced *Tead double* motif (Figure 1E), which is not considered a canonical regulatory

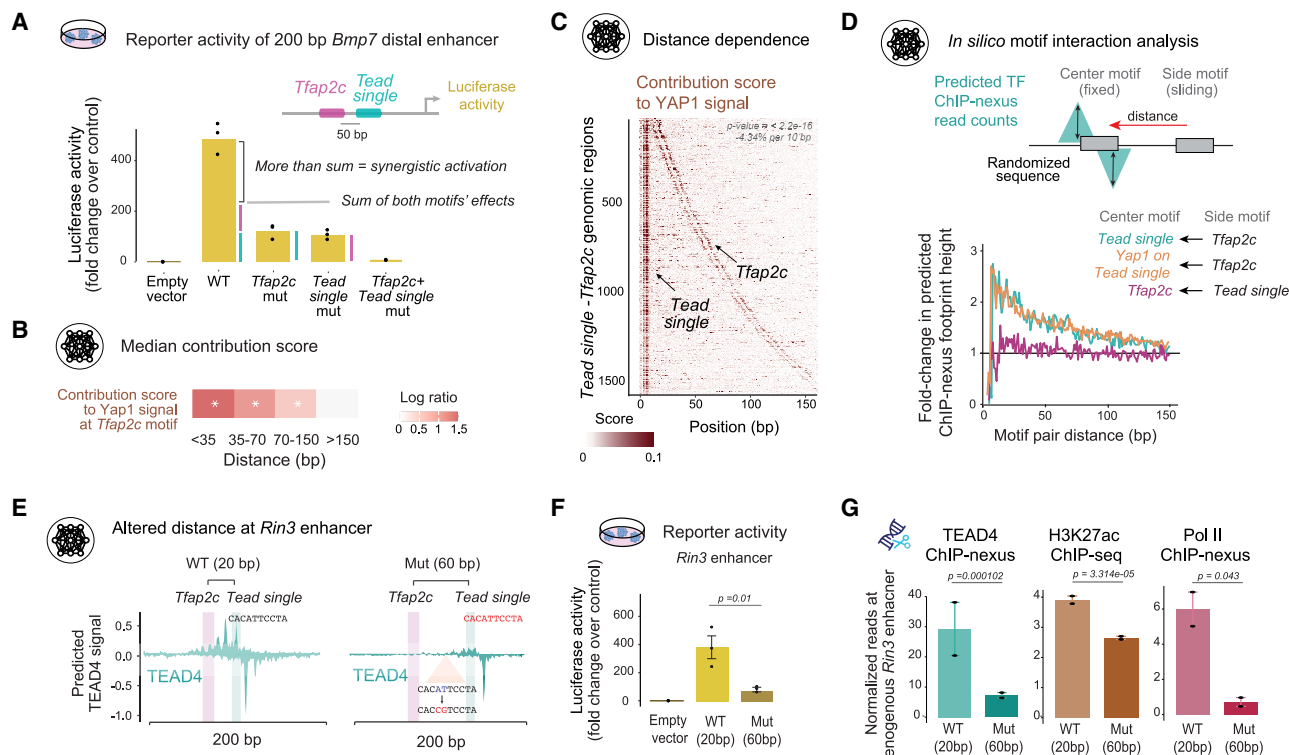


Figure 3. Enhancer activation involves DNA-distance-dependent cooperativity

(A) Luciferase assay of a 200 bp *Bmp7* enhancer (mm10-chr2:172,760,183-172,760,382) in TSCs shows that the *Tead single* and *Tfp2c* motifs function synergistically, producing reporter activity greater than the sum of each individual motif. Black dots show the three biological replicates per construct.

(B) The BPNNet contribution scores of *Tfp2c* motifs toward YAP1 binding are significantly higher when they are in close distance to a *Tead single* motif. For each distance interval, the median log ratio of the contribution scores over the baseline scores where *Tfp2c* motif >150 bp away is shown ($p < 2e-16$, Wilcoxon test).

(C) Heatmap of YAP1 binding contribution scores at genomic regions ordered by *Tead single*-*Tfp2c* motif distance. The contribution of both motifs decreases with larger distances. Linear regression gives a -4.34% slope per 10 bp and $p < 2.2e-16$ (Table S5 and STAR Methods).

(D) By injecting motifs *in silico* into randomized sequences, the average enhancement of TF binding to the center motif in the presence of a side motif is predicted by BPNNet.¹⁵ The results show a distance-dependent enhancement of TEAD4 and YAP1 binding in the presence of a *Tfp2c* motif.

(E) Predicted TEAD4 binding at the *Rin3* enhancer where the *Tfp2c* and *Tead single* motifs are 20 bp apart (left) and after the distance was increased to 60 bp between motifs (right). The motif was moved by inserting an identical new motif further away and mutating the most important bases within the original *Tead single* motif.

(F) Luciferase assays of the wild-type (WT) and mutated (Mut) 200 bp minimal *Rin3* enhancer in TSCs, all in three biological replicates and normalized to the empty vector control, show a significant decrease ($p < 0.05$, Student's *t* test).

(G) After mutating the endogenous *Rin3* enhancer through sequential CRISPR (Mut), TEAD4 ChIP-nexus binding (left), H3K27ac ChIP-seq enrichment (center), and Pol II ChIP-nexus occupancy (right) were all reduced compared with WT. The scale for ChIP-nexus is reads per million (RPM), for H3K27ac levels $\log_2(\text{ChIP-seq}/\text{WCE reads})$, in a 1 kb window. Two biological replicate values (black dots); error bars show standard deviation (SD). The p values were derived by DESeq2 (v.1.34).⁸³

element of the Hippo pathway,^{43,85,86} although it has been discovered multiple times.^{74,87,88} The first characterization occurred on the SV40 enhancer, but its identity remained unclear since it did not resemble the *Tead single* motif.⁸⁹⁻⁹³ Even after the *Tead double* motif had been discovered in *Drosophila* and cancer cells,^{74,87,94,95} it often remained unreported in genomics studies.^{43,96}

To test whether the *Tead double* motif has simply been overlooked or whether the BPNNet approach is particularly suitable to learn this motif, we analyzed our data using traditional genomics approaches (Figure 4A). We found that it is easy to miss the *Tead double* motif by MEME or HOMER (Figure S4A), but with prior knowledge of suitable settings, a position weight matrix (PWM) similar to that discovered by BPNNet/TF-MoDISco can

be identified (Figure 4A). However, when FIMO was used to map motif instances in the bound genomic regions with this PWM, the mapped motifs disagreed with the *in vivo* TEAD4 ChIP-nexus data. While CWM-mapped *Tead double* motifs consistently show strong TEAD4 footprints, confirming their correct mapping, the PWM-mapped motifs show far fewer footprints (Figure 4B). This problem was less pronounced with *Tead single* motifs, which mapped more accurately with PWM scanning (Figure S4B). This shows that BPNNet outperforms traditional methods in mapping functional motif instances,¹⁵ and that *Tead double* motifs are particularly challenging to map.

We next performed *in silico* experiments to test whether the *Tead double* motif itself was learned more accurately by BPNNet or whether the genomic context provided the necessary context

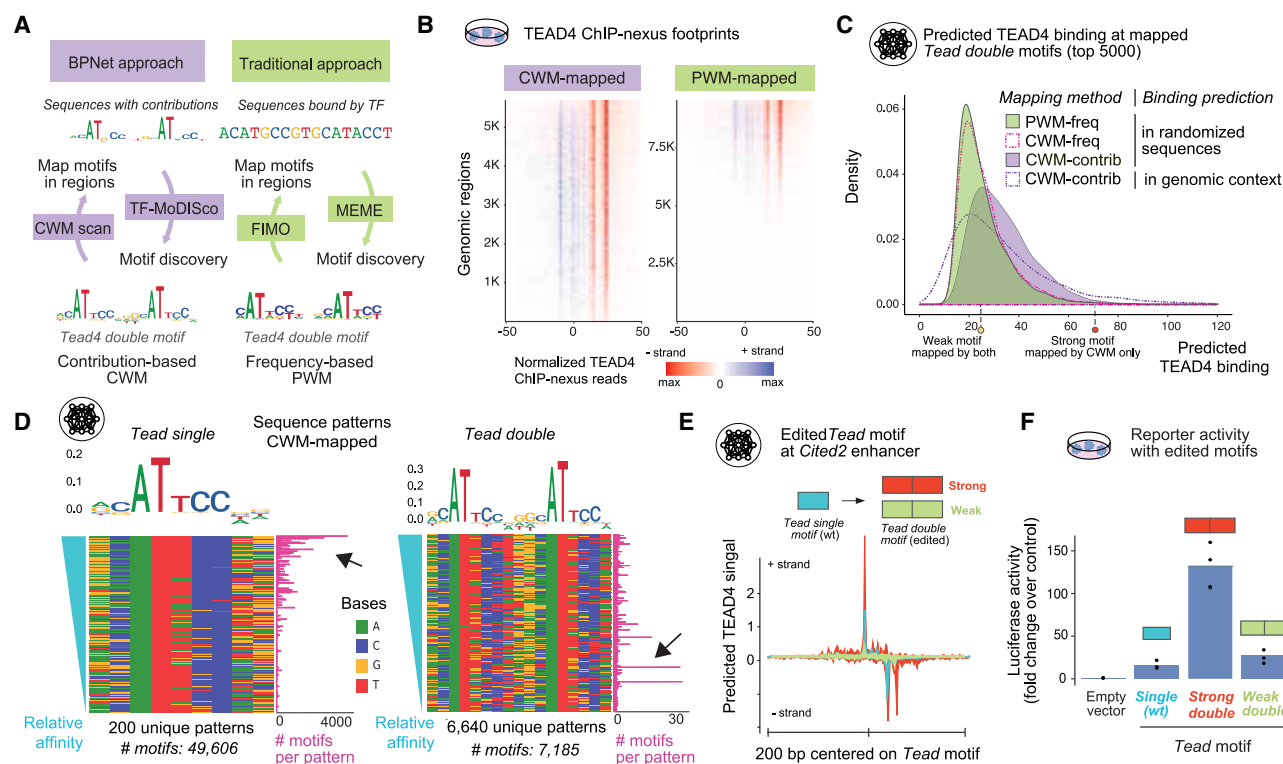


Figure 4. The *Tead double* motif is widespread, highly variable, and active

(A) Comparison of the BPNNet and traditional approach for *de novo* motif discovery and mapping the *Tead double* motif in genome sequences. The BPNNet approach uses TEAD4 contribution scores and CWM motifs, while MEME was used on the central 101 bp of TEAD4 peaks to derive frequency-based PWMs and map motifs by match scores. With correct settings, both methods identified the *Tead double* motif.

(B) However, *Tead double* motifs mapped by CWM scanning consistently show TEAD4 *in vivo* ChIP-nexus binding footprints throughout, while motifs mapped by PWM using FIMO show much fewer footprints; + strand (blue) and – strand (red). Regions were centered on the motif's left side and sorted by ChIP-nexus signal.

(C) Predicted TEAD4 binding at mapped *Tead double* motifs (top 5000). The density plot shows that the frequency-based PWM- or CWM-mapped motifs injected into randomized sequences perform similarly, while contribution-based CWM-mapped motifs have higher predicted values. The same CWM-mapped motifs predicted in their native genomic context show a wider distribution.

(D) All unique sequence patterns for the *Tead single* and *double* motifs, sorted by relative motif affinities (teal, left) and CWM logo on top. The motif frequency (purple, right) shows that high-affinity *Tead single* motifs occur most frequently, while *Tead double* motifs do not (black arrow).

(E) BPNNet-predicted TEAD4 binding profile at the *Cited2* enhancer when the *Tead single* motif was replaced with a strong or a weak *Tead double* motif.

(F) Luciferase assay of the 200 bp minimal *Cited2* enhancer (mm10-chr10:17,579,590–17,579,789), normalized over the empty vector control, for the three motifs, each in three biological replicates.

to make better predictions (Figure 4C). The key to CWM scanning is that it considers BPNNet model-derived contribution scores in each individual genomic region (Figure 4A). These contribution scores depend on the binding strength of the motif itself and the genomic context, such as neighboring motifs. To distinguish which one was key, we *in silico* injected the PWM-mapped and CWM-mapped motifs into a randomized sequence background and predicted TEAD4 binding (Figure 4C, methods). This revealed that the CWM-mapped motifs themselves had higher predicted binding than the PWM-mapped motifs, and that this was due to the contribution scores. Genomic context modulated the binding predictions, but did not explain the better performance of CWM scanning (Figures 4C and S4C). This shows that the CWM-mapped *Tead double* motifs are more accurate because they incorporate BPNNet's learned binding strength, and that the binding strength of this motif is not well modeled by a PWM.

The challenging aspect of the *Tead double* motif is that it is a long sequence pattern where many bases do not strongly contribute and are highly variable among the mapped motif sequences (Figure 4D). When we analyzed the sequence patterns of the more regular *Tead single* motifs, the vast majority are commonly occurring patterns, with high-affinity motifs being the most frequent (Figure 4D, left). In contrast, only a tiny fraction of the mapped *Tead double* motifs have recurring sequence patterns, and the most frequent patterns are not those with the highest predicted affinity (Figure 4D, right). This can explain why a frequency-based PWM representation does not accurately reflect the binding strength and why it has been difficult to identify the *Tead double* motif on the SV40 enhancer, while the BPNNet model learned this motif (Figure S4D).

Having established the widespread occurrence of the *Tead double* motif, we asked how much the *Tead double* motif promotes enhancer activity compared with a *Tead single* motif.

Using our validated putative *Cited2* enhancer, we replaced the high-affinity *Tead single* motif with either a weak *Tead double* motif (mapped by both PWM and CWM scanning), and a strong *Tead double* motif (mapped by CWM scanning only, highlighted in Figure 4C). BPNet predicted that replacement with the strong *Tead double* motif caused a large increase in TEAD4 binding, while the weaker one caused a reduction in binding (Figure 4E). When assayed in a luciferase assay, the strong *Tead double* motif caused an over 8-fold increase in activity compared with the wild-type *Tead single* motif. Interestingly, even the weak *Tead double* motif showed increased activity (~1.9-fold) over the *Tead single* motif despite lower TEAD4 binding (Figure 4F). This shows that even a weaker *Tead double* motif is highly active, explaining why it can be functional in so many different sequence patterns.

These results suggest that the *Tead double* motif is an important element of the Hippo pathway in TSCs. To test whether this can be generalized to other TEAD family members and cell types, we analyzed BPNet models trained on TEAD1-4 ChIP-seq data from the ENCODE portal^{97,98} (<https://www.encodeproject.org/>). BPNet discovered the *Tead double* motif in diverse human cell types for different TEAD family members (Figure S4E), showing that the *Tead double* motif is generally a widespread canonical motif of the Hippo pathway.

Genome-wide TEAD4 cooperativity through labile protein-protein interactions

The nature of the *Tead double* motif, its strong activity, and strict spacing, suggest that it is bound cooperatively by two TEAD4 molecules. Such cooperativity is also supported by previous gel shift analyses,^{87,94,95,90,99} but it is not known how strong the two TEAD4 molecules interact with each other. They could already come as a dimer, as known for JUN-FOS,¹⁰⁰ or they could only interact on DNA when a corresponding DNA sequence brings them into contact.^{24,101} We, therefore, explored the mechanism of TEAD4 cooperativity using BPNet and all-atom MD simulations (Figure 5).

The MD simulations were performed by placing two TEAD4 DNA binding domains (PDB: 5GZB)¹⁰² on the high-affinity functional *Tead double* motif from the putative *Tjp1* enhancer (Figures 5A, 5B, S5A, and S5B). This complex was highly similar to one predicted by AlphaFold3¹⁰⁴ (Figure S5M) and remained stable during the simulations over 500 ns, allowing us to quantify the protein-DNA and protein-protein interactions over time (measured as buried surface area Å²) (Figures 5C and S5G, left). We also confirmed that the results were similar using the *Tead double* motif from the putative *Amotl2* enhancer (Figures S5A and S5B, bottom, and S5E) or when using a different force field (CHARMM27 instead of FF19SB)¹⁰⁵ (Figure S5F).

Our MD simulations revealed that the intermolecular TEAD4 protein-protein interactions are sensitive to the DNA template, consistent with BPNet's predicted changes in binding levels (Figures S5H and S5N, left). The protein-protein interactions were strongest in the presence of the correctly spaced *Tead double* motif, and weaker when simulating with a 1-bp deletion, a 1-bp insertion, or mutations in one-half of the DNA sequence (Figures 5B, 5C, S5C, S5D, and S5G, right). The protein-protein

contacts occur through amino acid residues within a region that is important for TEAD4's cooperative binding to the *Tead double* motif⁹⁹ and is highly conserved between TEAD family members and across evolution (Figures 5D and 5E), consistent with the widespread role of the *Tead double* motif.

At the same time, these protein-protein interactions were surprisingly dynamic and labile (Figure 5C), arguing against TEAD4 forming dimers before binding to DNA. They changed on the scale of ~100 ns, which is much shorter than the tens of microseconds seen for stable protein-protein complexes.¹⁰⁶ The contacts made by the involved residues also varied in their molecular details over time. Contacts such as hydrogen bonds (Figure 5D, inset) formed and then dissociated, with no single interaction persisting for the entire trajectory (Video S1). This suggests that the initial DNA-templated protein-protein interactions between two TEAD4 molecules are labile and dynamic in nature, although binding of this complex *in vivo* may be further stabilized by the Hippo effector partners YAP1 or TAZ.¹⁰⁷

To analyze TEAD4 binding cooperativity *in vivo* and test its dependence on Hippo signaling, we leveraged BPNet (Figures 5F and 5G). Since BPNet accurately predicts TEAD4 binding on *Tead single* and *Tead double* motifs, we measured the cooperativity on the *Tead double* motif as the observed increase in binding over the expected additive binding on each *Tead* motif contained in the *Tead double* motif (Figure 5F). This showed consistently strong cooperativity across the genome, with an average of ~4-fold higher TEAD4 binding than expected from an additive model (Figure 5F).

To test whether this cooperativity was dependent on Hippo signaling, we performed TEAD4 ChIP-nexus in mouse embryonic stem cells (ESCs) where YAP1/TAZ is not nuclear,⁴¹ and trained an independent BPNet model on these data (Figure S5I). In these cells, TEAD4 binding was overall lower than in TSCs, resulting in much fewer bound instances (Figures 5G, S5J, and S5K). This is consistent with TEAD4 binding being stabilized at the *Tead double* motif by partners in TSCs.¹⁰⁷ However, we observed cooperativity even in ESCs, with a >2-fold increase over the additive signal (Figure 5G). This is consistent with the MD model, where some degree of TEAD4 cooperativity on the *Tead double* motif is observed in the absence of YAP1/TAZ.

Finally, we investigated whether the labile interactions we observe for TEAD4-TEAD4 are within range of what has been observed for other TFs that interact on DNA (Figure 5H). Since MD simulations are rarely performed, we quantified and compared the protein-protein buried surface area between TFs from crystal structures,²⁴ preferably on TFs studied here. Our high-affinity simulation gave an average buried surface area of 112 Å² between the two TEAD4 molecules. For comparison, the JUN-FOS dimer is held together through much stronger protein-protein interactions (2,466 Å²), consistent with dimerization before binding DNA.¹⁰⁰ GATA-GATA¹⁰⁸ formed intermediate interactions (530 Å²), suggesting that the complex is DNA templated but not as labile as the TEAD4-TEAD4 complex. Finally, we found interaction strengths akin to those of TEAD4-TEAD4 (Figure 5H), most notably those of IRF with RelA (105 Å²), p50 (94 Å²), and ATF-2 (180 Å²) in the enhanceosome model (Figure S5L). This example is particularly relevant

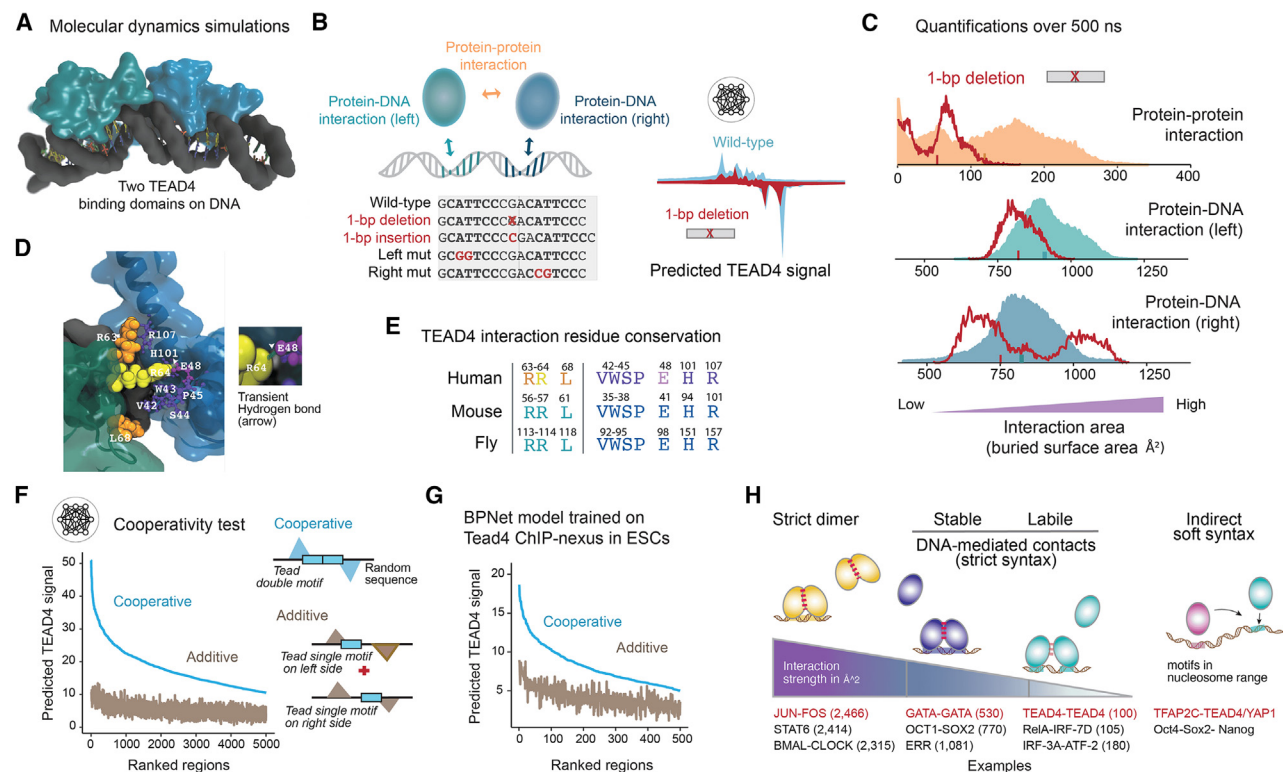


Figure 5. BPNet and MD simulations reveal insights into TEAD4 cooperative binding at double motifs

(A) Using the known structure of human TEAD4 bound to a single motif,¹⁰² we constructed a model of two TEAD4 DNA binding domains simultaneously bound to a high-affinity *Tead double* motif.

(B) Graphic showing which interactions were measured to quantify binding strength and which sequences were used as controls. The 1-bp deletion was chosen because BPNet predicts it to cause lower TEAD4 footprints (right), suggesting that cooperative binding depends on optimal spacing.

(C) Buried surface area distributions from the MD simulations with the high-affinity *Tead double* motif (GCATTCCCGACATTCCC) shown as solid areas, and the 1-bp deletion (GCATTCCCGACATTCCC) shown as a red line. Lower protein-protein interaction values for the 1-bp deletion and insertion suggest lower cooperativity at suboptimal spacing (see also Figure S5D). The mean protein-protein interaction is not a reliable measurement for sequence affinity but the difference between the two sides is a sequence-specific feature that we experimentally validated (Figure S5N, right).

(D) TEAD4 residues involved in interprotein interactions (defined by $\leq 4 \text{\AA}$ in $>20\%$ of simulation frames) are shown as space-filling spheres on the left TEAD4 and as ball-and-stick atoms on the right TEAD4. In a representative frame, R64 and E48 form a hydrogen bond (green dashed line, inset), but it forms/dissociates throughout the trajectory. A simulation video shows the entire protein-protein interaction dynamics (Video S1).

(E) Interactions residues are conserved across *Drosophila* Scalloped, mouse TEAD4, and human TEAD4 based on Clustal Omega multiple sequence alignment (UniProt IDs: P30052, Q62296, Q15561).

(F) *In silico* analysis of the TEAD4 cooperativity in mouse TSCs on all mapped *Tead double* motif sequences, injected into random sequences either as a whole (cooperative) or each half separately and then added (additive). The predicted signal was summed in a 50 bp window and averaged across all random sequences. Motifs were ordered by the predicted TEAD4 signal on the whole motif.

(G) Same as in (F), but mapped *Tead double* motifs and predictions were in mouse ESCs.

(H) Quantification of protein-protein buried surface area ($\text{\AA}^2$) from structures of TFs that form dimers in solution (PDB: 1JNM [AP-1]; 4Y5W [STAT6]; 4H10 [BMAL:CLOCK]), TFs that have stable DNA-mediated interactions (PDB: 3DFV [GATA-GATA]; 104X [OCT1-SOX2]; 8CEF [ERR]), TFs with weak interactions (our simulations of TEAD4, PDB: 5GZB; RelA-IRF-7D and IRF-3A-ATF-2¹⁰³), and TFs that cooperate with soft syntax and may not directly interact; examples include TFAP2C and TEAD4, and OCT4-SOX2 and NANOG.¹⁵ Highlighted in red are TFs for which motifs were identified here with BPNet.

as the TFs have been reported to cooperate with a measurable effect on transcription.^{25,82,109} These results point toward labile interactions being a plausible mechanism for DNA-templated cooperativity.

Taken together, our results suggest that two TEAD4 molecules cooperate with each other on the *Tead double* motif through transient and labile interactions (Figure 5H). We propose that labile interactions make this type of cooperativity highly dependent on the sequence template and allow additional stabilization by partners to further fine-tune the readout.

A redesigned enhancer shows that the *Tead double* motif increases gene activation in mouse embryos

Having shown that the *Tead double* motif mediates TEAD4 cooperative binding genome-wide, we wanted to confirm that it mediates the response to the Hippo pathway *in vivo*. We added the *Tead double* motif to an endogenous enhancer in TSCs and re-integrated these cells into mouse embryos. We selected the putative *Ezr* enhancer because it contains a *Tead single* motif, has all the hallmarks of being an active TSC enhancer (Figure S2F), and its putative target gene *Ezr* encodes an actin-associated

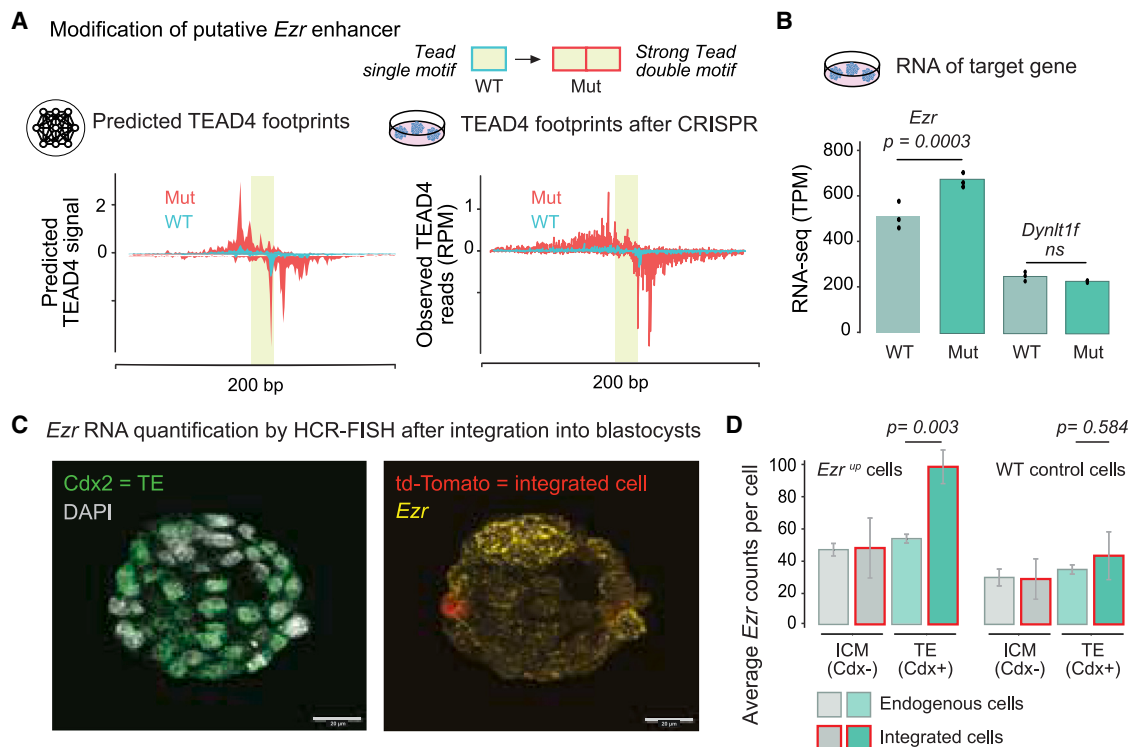


Figure 6. CRISPR-Cas9 enhancer design validated by BPNet increases target gene activity in mouse embryos

(A) At the putative *Ezr* enhancer in mouse TSCs (mm10-chr17:6,827,705-6,827,905), the *Tead* single motif in the wild-type (WT) sequence was mutated (Mut) into a strong *Tead* double motif through CRISPR-Cas9-mediated homologous recombination. BPNet predicts increased TEAD4 binding (left); + strand on top, – strand on the same scale below, lime box: motif width. ChIP-nexus experiments confirm this change (right); scale is reads per million (RPM).

(B) RNA-seq data in transcripts per million (TPM) in WT and Mut cells. Differential expression p values were derived using edgeR on three biological replicates. (C) HCR-FISH was performed on aggregated mouse blastocyst embryos with incorporated WT or Mut (edited *Ezr*^{up}) cells, using probes for *Cdx2*, *Ezr*, and *td-Tomato*.

(D) Quantification of average *Ezr* counts (average *Cdx2* counts shown in Figure S5E). *Ezr* transcripts were significantly increased for edited cells but not for WT cells among *Cdx2*+ TE lineage cells (Student's t test $p < 0.05$). Error bars show standard error of the mean (SEM).

protein that is highly expressed in TSCs.^{54,110,111} We hypothesized that replacing the *Tead* single motif with a high-affinity *Tead* double motif would increase the enhancer's response to Hippo signaling *in vivo*.

While it is typically easy to destroy the activity of an endogenous enhancer by mutating relevant TF motifs,³⁴ it is more challenging to engineer mutations that increase the activity since sequence changes can have unexpected side effects.^{30,31} However, deep learning models are ideal for exploring and evaluating possible mutations, as shown in *Drosophila*.^{16,112} We, therefore, replaced the *Tead* single motif with a *Tead* double motif in such a way that BPNet predicts an increase in TEAD4 binding (Figure 6A, left).

We then used CRISPR-Cas9-induced homologous recombination to edit the endogenous enhancer in our TSCs (Figure S6A). We performed TEAD4 ChIP-nexus experiments on the edited cells and found that the TEAD4 binding footprint indeed matched the one predicted by BPNet (Figures 6A, right, and S6B). We also tested the modified enhancer sequence in the luciferase assay and found it to increase enhancer activity (Figure S6C). The increase was moderate (~1.5-fold), likely because the wild-type enhancer activity was already high to begin with. To test whether

this change affects the expression of *Ezr*, we performed RNA-seq on the edited cells with the *Tead* double motif and the wild-type cells with the *Tead* single motif. This revealed a moderate, statistically significant increase in *Ezr* transcript levels in the edited cells (Figure 6B).

We next tested whether this edit increases *Ezr* expression in mouse embryos in a cell-type-specific way. We marked the edited TSCs (and wild-type TSCs as control) with td-Tomato, aggregated these cells with early mouse embryos at the 4–8 cell stage, and analyzed the embryos when they reached the blastocyst stage, where the outer TE cells are clearly distinguishable from the inner cell mass (ICM) by their expression of *Cdx2*. We performed HCR-FISH to precisely quantify the expression of *Ezr* and *Cdx2* in these embryos (Figures 6C and S6D–S6F).

Ezr transcripts were specifically increased in edited cells but not wild-type cells, and only when they became TE cells with nuclear YAP1 (Figure 6C, right). Not all added TSCs maintained their TE identity but occasionally lost *Cdx2* expression and acquired ICM identity (Figure S6E), consistent with cell fate plasticity at this stage.⁴⁴ Notably, when the edited cells lost TE identity, *Ezr* transcripts were no longer increased. These findings show that the increased activity of the *Tead* double motif is

specific to the cell type with nuclear YAP1. They also demonstrate that, with the help of BPNet's predictive framework, enhancers can be manipulated to respond more strongly to a cell-type-specific signaling pathway *in vivo*.

DISCUSSION

Here, we show as a proof of principle that the cell-type specificity of the Hippo signaling pathway in TSCs is encoded in the *cis*-regulatory sequences and that the sequence rules reflect the mechanisms by which these sequences are read out. Canonical effectors such as TEAD4 and YAP1 of the Hippo pathway are well studied, but how these effectors mediate different transcriptional responses in different cell types has been a challenging problem, often attributed to the combinatorial complexity of the signaling components or effects of chromatin. The fact that we can train a deep learning model to accurately predict the binding of the signaling effectors in TSCs from sequence alone, and that the binding is predictive of enhancer activation, shows that the cell-type-specific response to Hippo signaling is sequence encoded. Moreover, we found that the learned sequence rules are precise enough to reveal mechanistic insights into how signaling pathway effectors function and interact with cell-type-specific partner TFs.

The power of this approach comes from training a highly predictive base-resolution sequence model. Since the model can accurately predict test sequences it has never seen during training, it has learned general TF binding rules that apply genome-wide for the cell type it was trained on. These rules include the motif's strength (as shown here for the *Tead double* motif) and incorporate input from neighboring motifs (e.g., the contribution of *Tfap2c* toward YAP1 binding). As a result, the tens of thousands of motifs that we mapped inside active enhancers are more accurate than motifs mapped by traditional methods. For example, a frequency-based PWM could not accurately map the long and variable *Tead double* motif in the genome, explaining why studies in the past did not identify this motif as a widespread canonical element of the Hippo pathway. Furthermore, we used the predictive accuracy of the model for designing follow-up experiments, e.g., performing minimal mutations to achieve a certain outcome.

A key focus of this study was, however, to interpret the trained model to derive mechanistic insights. We found that the model learned precise rules by which motifs combinatorially predict YAP1 binding, including syntax rules that determine how the distances between motifs affect the binding cooperativity. These syntax rules should depend on the molecular mechanisms by which the corresponding TFs interact. Since we re-discovered expected motifs and their interactions, and validated novel findings, we conclude that the sequence rules indeed mirror molecular mechanisms downstream of the Hippo signaling pathway.

The motif syntax rules suggested that cell-type-specific TFAP2C boosts YAP1 binding and enhancer activation at motif distances of up to ~150 bp, with stronger effects at closer distances. This soft syntax, which we experimentally validated, is unlikely to involve specific protein-protein interactions and instead points to a nucleosome-mediated mechanism.^{15,20,26,113–117} Such a mechanism, while molecularly unclear, could explain how signaling TFs can receive input from a wide variety of TFs

in different cell types without having evolved specific protein-protein interactions. Nevertheless, we found that some cell-type-specific TFs had a stronger contribution to YAP1 binding than others, suggesting that specific properties make them better partner TFs. TFAP2C is likely a strong partner because it is highly expressed, can pioneer chromatin,^{118,119} and interacts with co-activators,¹²⁰ which could help YAP1 form condensates.^{73,79} An important challenge in the future will be to measure how TFs interact with each other in different cell types and which properties drive these interactions.

A different type of syntax was found for the *Tead double* motif, where two strictly spaced *Tead* motifs mediate cooperative TEAD4 binding. Using MD simulations, we found that this motif allows two TEAD4 molecules to directly interact with each other, but the protein-protein interactions were surprisingly labile, transient, and dependent on the DNA sequence. Such weak interactions between TFs have been observed in the enhanceosome crystal structure, but their dynamic nature and significance were not known at the time. We propose that such transient interactions are strong enough to stabilize the complex but weak enough that they depend on matching DNA sequences and can be further stabilized through signaling pathway activities.

In summary, our sequence-driven interpretable deep learning approach reveals motif dependencies and syntax rules that correspond to distinct molecular mechanisms and suggest novel hypotheses that can be further studied experimentally. For example, we also discovered a palindromic *Gata double* motif. Since previous studies have shown that GATA zinc fingers can bind DNA cooperatively,^{108,121} it should be possible to identify the structural basis for GATA3 binding to this specific motif. Importantly, our approach identifies TF cooperativities that are in many cases cell-type specific and thus open the door to systematically study how signaling pathways target different enhancers in different cell types.

Limitations of the study

A limitation of our approach is that it depends on high-quality binding data in the cell type of interest. While BPNet can model lower-resolution ChIP-seq data,^{15,122} the cell-type specificity of signaling effectors makes obtaining comprehensive data across cell types beyond ENCODE challenging. Model training and interpretation also have limitations. For example, we noticed that YAP1 binding correlates better with enhancer activity markers than TEAD4 binding, yet the model interrogation suggested that YAP1 binding is mostly boosted through increased binding of TEAD4. Finally, the precision of MD simulations can give a deceptive sense of accuracy if the simulation does not adequately sample conformation space. Since our simulated system is large, we cannot capture a complete equilibrium ensemble, making MD results a snapshot of the possible behavior of the TEAD4-DNA complex.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Julia Zeitlinger (jbz@stowers.org).

Materials availability

CRISPR-Cas9 cell lines and complete hydrated MD trajectories generated in this study are available upon request.

Data and code availability

- The raw and processed data for ChIP-nexus, ChIP-seq, ATAC-seq, TT-seq, and RNA-seq experiments have been deposited in GEO under series accession number GSE252463. The genomic datasets used in the paper can be viewed on the UCSC Genome Browser: [Link](https://research.stowers.org/zeitlingerlab/protocols.html)
- The ChIP-nexus protocol description can be found at <https://research.stowers.org/zeitlingerlab/protocols.html>. The trained BPNet model is available at Zenodo <https://zenodo.org/records/14894986>. Original data, including MD simulation trajectories and microscopy images, can be accessed from the Stowers Original Data Repository at <http://www.stowers.org/research/publications/libpb-2440>. All code used to process and analyze the data in this paper can be accessed at https://github.com/zeitlingerlab/Dala_hippo_signaling_2024.

ACKNOWLEDGMENTS

We thank Žiga Avsec, Robb Krumlauf, Helen McNeill, Anshul Kundaje, and Zeitlinger lab members for their helpful comments and suggestions on the manuscript. We thank the following Stowers Institute core facilities for their support: Sequencing and Discovery Genomics (Anoja Perera, Michael Peterson, and Amanda Lawlor), Histology (Dai Tsuchiya, Yongfu Wang, and Seth Malloy), Transgenic and Reproductive Technologies team (Michael Durnin and Andrea Moran), Cells Tissues and Organoids Center (Yan Wang, Naresh Kumar Rajendran, Sonia Ghosh, Maria Katt, Olga Kenzior, Shilpa Waduware, and Chongbei Zhao), Lab Services (Stacey Walker), Cytometry (Kevin Ferro, Jose Javier, KyeongMin Bae, and Jeff Haug), and Computational Biology (Hua Li, Madelaine Gogol, and Hassan Huzaifa). The research reported in this publication was supported by the Stowers Institute for Medical Research, United States and NIH grant R01HG010211 to J.Z.

AUTHOR CONTRIBUTIONS

K.D. and J.Z. conceived the project as part of K.D.'s thesis research to fulfill the University of Kansas Medical Center requirements. K.D. and J.Z. designed the genomics and other experiments that K.D. performed. Deep learning model training, computational analysis, and *in silico* experiments were performed by K.D. and M.W. Additional genomics data analyses were done by K.D. ATAC-seq experiments were performed by S.K. MD experiments were conceived and designed by C.M., K.D., and J.Z.; performed by C.M.; and analyzed by C.M., K.D., and J.Z. Embryo aggregation, imaging, and analysis were conceived by K.D. and J.Z. Imaging and analysis was performed by M.C.M. The manuscript was prepared by K.D. and J.Z. with input from all authors.

DECLARATION OF INTERESTS

J.Z. owns a patent on ChIP-nexus (no. 10287628).

STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **METHOD DETAILS**
 - Mouse stem cell culture
 - ChIP-nexus, PAtCh-Cap, and ChIP-seq experiments
 - Luciferase assays
 - TT-seq experiments
 - ATAC-seq experiments
 - RNA-seq experiments
 - CRISPR-Cas9 experiments
 - Mice strains and superovulation

- Lentivirus transduction of fluorescent td-tomato in TSCs
- Aggregation assays to obtain chimeric embryos
- Immunofluorescence stainings of chimeric embryos
- HCR-FISH on chimeric embryos
- Molecular dynamics simulations
- ChIP-nexus data processing
- ChIP-seq data processing
- TT-seq data processing
- ATAC-seq data processing
- RNA-seq data processing
- HCR-FISH image analysis
- Molecular dynamics analysis and visualization
- BPNet model training
- Motif extraction, curation, and island generation
- Visualization of YAP1 binding and enhancer activity markers
- Motif pair interaction analysis
- Enhancer regions selection for reporter assay
- Extracting regions with different *Tead double* motif spacings
- *Tead* motifs variant analysis
- Genome-wide TEAD4 binding cooperativity on *Tead double* motifs
- Benchmarking motif discovery and analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2025.100821>.

Received: February 22, 2024

Revised: September 23, 2024

Accepted: March 5, 2025

Published: April 1, 2025

REFERENCES

1. Hamidi, S., and Alev, C. (2022). In vitro models of pre- and post-gastrulation embryonic development. *Curr. Opin. Genet. Dev.* 77, 101985. <https://doi.org/10.1016/j.gde.2022.101985>.
2. Barolo, S., and Posakony, J.W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* 16, 1167–1181. <https://doi.org/10.1101/gad.976502>.
3. Weidemüller, P., Kholmatov, M., Petsalaki, E., and Zaugg, J.B. (2021). Transcription factors: Bridge between cell signaling and gene regulation. *Proteomics* 21, e2000034. <https://doi.org/10.1002/pmic.202000034>.
4. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117. <https://doi.org/10.1016/j.cell.2008.04.043>.
5. Zeitlinger, J. (2020). Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Struct. Biol.* 23, 22–31. <https://doi.org/10.1016/j.coisb.2020.08.002>.
6. Jindal, G.A., and Farley, E.K. (2021). Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev. Cell* 56, 575–587. <https://doi.org/10.1016/j.devcel.2021.02.016>.
7. Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., and Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* 58, 362–370. <https://doi.org/10.1016/j.molcel.2015.02.014>.
8. Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. *Cell* 103, 75–85. [https://doi.org/10.1016/s0092-8674\(00\)00106-9](https://doi.org/10.1016/s0092-8674(00)00106-9).

9. Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R., and Young, R.A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* 113, 395–404. [https://doi.org/10.1016/s0092-8674\(03\)00301-5](https://doi.org/10.1016/s0092-8674(03)00301-5).
10. Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.-C., DiBiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L., et al. (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* 147, 577–589. <https://doi.org/10.1016/j.cell.2011.09.044>.
11. Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilo-deau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell* 147, 565–576. <https://doi.org/10.1016/j.cell.2011.08.050>.
12. Xie, L., Torigoe, S.E., Xiao, J., Mai, D.H., Li, L., Davis, F.P., Dong, P., Marie-Nelly, H., Grimm, J., Lavis, L., et al. (2017). A dynamic interplay of enhancer elements regulates Klf4 expression in naïve pluripotency. *Genes Dev.* 31, 1795–1808. <https://doi.org/10.1101/gad.303321.117>.
13. Blassberg, R., Patel, H., Watson, T., Gouti, M., Metzis, V., Delás, M.J., and Briscoe, J. (2022). Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat. Cell Biol.* 24, 633–644. <https://doi.org/10.1038/s41556-022-00910-2>.
14. Siggers, T., and Gordán, R. (2014). Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res.* 42, 2099–2111. <https://doi.org/10.1093/nar/gkt1112>.
15. Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Frof, R., McAnany, C., Gagneur, J., Kundaje, A., and Zeitlinger, J. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* 53, 354–366. <https://doi.org/10.1038/s41588-021-00782-6>.
16. de Almeida, B.P., Reiter, F., Pagani, M., and Stark, A. (2022). Deep-STARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* 54, 613–624. <https://doi.org/10.1038/s41588-022-01048-5>.
17. Kim, D.S., Risca, V.I., Reynolds, D.L., Chappell, J., Rubin, A.J., Jung, N., Donohue, L.K.H., Lopez-Pajares, V., Kathiria, A., Shi, M., et al. (2021). The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. *Nat. Genet.* 53, 1564–1576. <https://doi.org/10.1038/s41588-021-00947-3>.
18. Kelley, D.R., Snoek, J., and Rinn, J.L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* 26, 990–999. <https://doi.org/10.1101/gr.200535.115>.
19. Greenside, P., Shimko, T., Fordyce, P., and Kundaje, A. (2018). Discovering epistatic feature interactions from neural network models of regulatory DNA sequences. *Bioinformatics* 34, i629–i637. <https://doi.org/10.1093/bioinformatics/bty575>.
20. Brennan, K.J., Weilert, M., Krueger, S., Pampari, A., Liu, H.-Y., Yang, A.W.H., Morrison, J.A., Hughes, T.R., Rushlow, C.A., Kundaje, A., and Zeitlinger, J. (2023). Chromatin accessibility in the Drosophila embryo is determined by transcription factor pioneering and enhancer activation. *Dev. Cell* 58, 1898–1916. <https://doi.org/10.1016/j.devcel.2023.07.007>.
21. Xu, C., Kleinschmidt, H., Yang, J., Leith, E., Johnson, J., Tan, S., Mahony, S., and Bai, L. (2023). Systematic Dissection of Sequence Features Affecting the Binding Specificity of a Pioneer Factor Reveals Binding Synergy Between FOXA1 and AP-1. Preprint at bioRxiv, 2023.11.08.566246. <https://doi.org/10.1101/2023.11.08.566246>.
22. Horton, C.A., Alexandari, A.M., Hayes, M.G.B., Marklund, E., Schaepe, J.M., Aditham, A.K., Shah, N., Shrikumar, A., Afek, A., Greenleaf, W.J., et al. (2022). Short tandem repeats bind transcription factors to tune eukaryotic gene expression. Preprint at bioRxiv. <https://doi.org/10.1101/2022.05.24.493321>.
23. Alexandari, A.M., Horton, C.A., Shrikumar, A., Shah, N., Li, E., Weilert, M., Pufall, M.A., Zeitlinger, J., Fordyce, P.M., and Kundaje, A. (2023). De novo distillation of thermodynamic affinity from deep learning regulatory sequence models of in vivo protein-DNA binding. Preprint at bioRxiv, 2023.05.11.540401. <https://doi.org/10.1101/2023.05.11.540401>.
24. Morgunova, E., and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* 47, 1–8. <https://doi.org/10.1016/j.sbi.2017.03.006>.
25. Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. *Cell* 129, 1111–1123. <https://doi.org/10.1016/j.cell.2007.05.019>.
26. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187. <https://doi.org/10.1016/j.cell.2016.09.018>.
27. Singh, G., Mullany, S., Moorthy, S.D., Zhang, R., Mehdi, T., Tian, R., Duncan, A.G., Moses, A.M., and Mitchell, J.A. (2021). A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. *Genome Res.* 31, 564–575. <https://doi.org/10.1101/gr.272468.120>.
28. Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* 94, 890–898. <https://doi.org/10.1002/jcb.20352>.
29. Martin, V., Zhuang, F., Zhang, Y., Pinheiro, K., and Gordán, R. (2023). High-throughput data and modeling reveal insights into the mechanisms of cooperative DNA-binding by transcription factor proteins. *Nucleic Acids Res.* 51, 11600–11612. <https://doi.org/10.1093/nar/gkad872>.
30. Crocker, J., and Ilesley, G.R. (2017). Using synthetic biology to study gene regulatory evolution. *Curr. Opin. Genet. Dev.* 47, 91–101. <https://doi.org/10.1016/j.gde.2017.09.001>.
31. Vincent, B.J., Estrada, J., and DePace, A.H. (2016). The appeasement of Doug: a synthetic approach to enhancer biology. *Integr. Biol.* 8, 475–484. <https://doi.org/10.1039/c5ib00321k>.
32. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., et al. (2022). Sequence determinants of human gene regulatory elements. *Nat. Genet.* 54, 283–294. <https://doi.org/10.1038/s41588-021-01009-4>.
33. Fiore, C., and Cohen, B.A. (2016). Interactions between pluripotency factors specify cis-regulation in embryonic stem cells. *Genome Res.* 26, 778–786. <https://doi.org/10.1101/gr.200733.115>.
34. Reiter, F., de Almeida, B.P., and Stark, A. (2023). Enhancers display constrained sequence flexibility and context-specific modulation of motif function. *Genome Res.* 33, 346–358. <https://doi.org/10.1101/gr.277246.122>.
35. King, D.M., Hong, C.K.Y., Shepherdson, J.L., Granas, D.M., Maricque, B.B., and Cohen, B.A. (2020). Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. *Elife* 9, e41279. <https://doi.org/10.7554/eLife.41279>.
36. Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. USA* 113, 6508–6513. <https://doi.org/10.1073/pnas.1605085113>.
37. Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. *Science* 350, 325–328. <https://doi.org/10.1126/science.aac6948>.
38. He, Q., Johnston, J., and Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* 33, 395–401. <https://doi.org/10.1038/nbt.3121>.
39. Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419. <https://doi.org/10.1016/j.cell.2011.11.013>.
40. Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of

trophoblast in pre-implantation mouse embryos. *Mech. Dev.* 125, 270–283. <https://doi.org/10.1016/j.mod.2007.11.002>.

41. Nishioka, N., Inoue, K.I., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., et al. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophoblast from inner cell mass. *Dev. Cell* 16, 398–410. <https://doi.org/10.1016/j.devcel.2009.02.003>.
42. Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophoblast lineage at the beginning of mammalian development. *Development* 134, 3827–3836. <https://doi.org/10.1242/dev.010223>.
43. Home, P., Saha, B., Ray, S., Dutta, D., Gunewardena, S., Yoo, B., Pal, A., Vivian, J.L., Larson, M., Petroff, M., et al. (2012). Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proc. Natl. Acad. Sci. USA* 109, 7362–7367. <https://doi.org/10.1073/pnas.1201595109>.
44. Posfai, E., Petropoulos, S., de Barros, F.R.O., Schell, J.P., Jurisica, I., Sandberg, R., Lanner, F., and Rossant, J. (2017). Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *Elife* 6, e22906. <https://doi.org/10.7554/eLife.22906>.
45. Karasek, C., Ashry, M., Driscoll, C.S., and Knott, J.G. (2020). A tale of two cell-fates: role of the Hippo signaling pathway and transcription factors in early lineage formation in mouse preimplantation embryos. *Mol. Hum. Reprod.* 26, 653–664. <https://doi.org/10.1093/molehr/gaaa052>.
46. Vassilev, A., Kaneko, K.J., Shu, H., Zhao, Y., and DePamphilis, M.L. (2001). TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. *Genes Dev.* 15, 1229–1241. <https://doi.org/10.1101/gad.888601>.
47. Knott, J.G., and Paul, S. (2014). Transcriptional regulators of the trophoblast lineage in mammals with hemochorial placentation. *Reproduction* 148, R121–R136. <https://doi.org/10.1530/REP-14-0072>.
48. Kuckenberg, P., Buhl, S., Woynecki, T., van Fürden, B., Tolkunova, E., Seiffe, F., Moser, M., Tomilin, A., Winterhager, E., and Schorle, H. (2010). The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophoblast formation. *Mol. Cell Biol.* 30, 3310–3320. <https://doi.org/10.1128/MCB.01215-09>.
49. Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophoblast differentiation. *Cell* 123, 917–929. <https://doi.org/10.1016/j.cell.2005.08.040>.
50. Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophoblast of peri-implantation embryo and directly regulates Cdx2 gene expression. *J. Biol. Chem.* 284, 28729–28737. <https://doi.org/10.1074/jbc.M109.016840>.
51. Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* 137, 395–403. <https://doi.org/10.1242/dev.038828>.
52. Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsakso-phak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophoblast in the mouse blastocyst. *Development* 132, 2093–2102. <https://doi.org/10.1242/dev.01801>.
53. Chi, F., Sharpley, M.S., Nagaraj, R., Roy, S.S., and Banerjee, U. (2020). Glycolysis-Independent Glucose Metabolism Distinguishes TE from ICM Fate during Mammalian Embryogenesis. *Dev. Cell* 53, 9–26. <https://doi.org/10.1016/j.devcel.2020.02.015>.
54. Zhu, M., Cornwall-Scoones, J., Wang, P., Handford, C.E., Na, J., Thomson, M., and Zernicka-Goetz, M. (2020). Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* 370, eabd2703. <https://doi.org/10.1126/science.abd2703>.
55. Singh, V.P., and Gerton, J.L. (2021). Protocol for mouse trophoblast stem cell isolation, differentiation, and cytokine detection. *STAR Protoc.* 2, 100242. <https://doi.org/10.1016/j.xpro.2020.100242>.
56. Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J.D., Wang, C.-Y., Chinnaiyan, A.M., et al. (2008). TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* 22, 1962–1971. <https://doi.org/10.1101/gad.1664408>.
57. Li, Z., Zhao, B., Wang, P., Chen, F., Dong, Z., Yang, H., Guan, K.-L., and Xu, Y. (2010). Structural insights into the YAP and TEAD complex. *Genes Dev.* 24, 235–240. <https://doi.org/10.1101/gad.1865810>.
58. Chen, L., Chan, S.W., Zhang, X., Walsh, M., Lim, C.J., Hong, W., and Song, H. (2010). Structural basis of YAP recognition by TEAD4 in the hippo pathway. *Genes Dev.* 24, 290–300. <https://doi.org/10.1101/gad.1865310>.
59. Kurowski, A., Molotkov, A., and Soriano, P. (2019). FGFR1 regulates trophoblast development and facilitates blastocyst implantation. *Dev. Biol.* 446, 94–101. <https://doi.org/10.1016/j.ydbio.2018.12.008>.
60. Leung, C.Y., and Zernicka-Goetz, M. (2013). Angiogenin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat. Commun.* 4, 2251. <https://doi.org/10.1038/ncomms3251>.
61. Plusa, B., Frankenberg, S., Chalmers, A., Hadjantonakis, A.-K., Moore, C.A., Papalopulu, N., Papaioannou, V.E., Glover, D.M., and Zernicka-Goetz, M. (2005). Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J. Cell Sci.* 118, 505–515. <https://doi.org/10.1242/jcs.01666>.
62. Lim, H.Y.G., Alvarez, Y.D., Gasnier, M., Wang, Y., Tetlak, P., Bissiere, S., Wang, H., Biro, M., and Plachta, N. (2020). Keratins are asymmetrically inherited fate determinants in the mammalian embryo. *Nature* 585, 404–409. <https://doi.org/10.1038/s41586-020-2647-4>.
63. Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. Preprint at arXiv. <https://doi.org/10.48550/arxiv.1704.02685>.
64. Shrikumar, A., Tian, K., Shcherbina, A., Avsec, Z., Banerjee, A., Sharmin, M., Nair, S., and Kundaje, A. (2018). TF-MoDISco V0.4.2.2-alpha: Technical Note. Preprint at arXiv. <https://doi.org/10.48550/arxiv.1811.00416>.
65. He, L., Pratt, H., Gao, M., Wei, F., Weng, Z., and Struhl, K. (2021). YAP and TAZ are transcriptional co-activators of AP-1 proteins and STAT3 during breast cellular transformation. *Elife* 10, e67312. <https://doi.org/10.7554/eLife.67312>.
66. Zancanato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Biccato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* 17, 1218–1227. <https://doi.org/10.1038/ncb3216>.
67. Liu, X., Li, H., Rajurkar, M., Li, Q., Cotton, J.L., Ou, J., Zhu, L.J., Goel, H.L., Mercurio, A.M., Park, J.-S., et al. (2016). Tead and AP1 coordinate transcription and motility. *Cell Rep.* 14, 1169–1180. <https://doi.org/10.1016/j.celrep.2015.12.104>.
68. Pascual, J., Jacobs, J., Sansores-Garcia, L., Natarajan, M., Zeitlinger, J., Aerts, S., Halder, G., and Hamaratoglu, F. (2017). Hippo reprograms the transcriptional response to ras signaling. *Dev. Cell* 42, 667–680. <https://doi.org/10.1016/j.devcel.2017.08.013>.
69. Cao, Z., Carey, T.S., Ganguly, A., Wilson, C.A., Paul, S., and Knott, J.G. (2015). Transcription factor AP-2γ induces early Cdx2 expression and represses HIPPO signaling to specify the trophoblast lineage. *Development* 142, 1606–1615. <https://doi.org/10.1242/dev.120238>.
70. Kuckenberg, P., Kubaczka, C., and Schorle, H. (2012). The role of transcription factor Tcfap2c/TFAP2C in trophoblast development. *Reprod. Biomed. Online* 25, 12–20. <https://doi.org/10.1016/j.rbmo.2012.02.015>.
71. Winger, Q., Huang, J., Auman, H.J., Lewandoski, M., and Williams, T. (2006). Analysis of transcription factor AP-2 expression and function

- p>
during mouse preimplantation development.
- Biol. Reprod.*
- 75, 324–333.
- <https://doi.org/10.1095/biolreprod.106.052407>
- .
72. Galli, G.G., Carrara, M., Yuan, W.-C., Valdes-Quezada, C., Gurung, B., Pepe-Mooney, B., Zhang, T., Geeven, G., Gray, N.S., de Laat, W., et al. (2015). YAP Drives Growth by Controlling Transcriptional Pause Release from Dynamic Enhancers. *Mol. Cell* 60, 328–337. <https://doi.org/10.1016/j.molcel.2015.09.001>.
 73. Hao, S., Fuehrer, H., Flores, E., Demmerle, J., Lippincott-Schwartz, J., Liu, Z., Sukenik, S., and Cai, D. (2022). YAP condensates are highly organized hubs for YAP/TEAD transcription. Preprint at bioRxiv. <https://doi.org/10.1101/2022.10.24.513621>.
 74. Stein, C., Bardet, A.F., Roma, G., Bergling, S., Clay, I., Ruchti, A., Agarinis, C., Schmelzle, T., Bouwmeester, T., Schübeler, D., and Bauer, A. (2015). YAP1 Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers. *PLoS Genet.* 11, e1005465. <https://doi.org/10.1371/journal.pgen.1005465>.
 75. Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* 32, 202–223. <https://doi.org/10.1101/gad.310367.117>.
 76. Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M., and Zeitlinger, J. (2016). Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.* 17, 196. <https://doi.org/10.1186/s13059-016-1057-2>.
 77. Creighton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936. <https://doi.org/10.1073/pnas.1016071107>.
 78. Kristjánsdóttir, K., Dziubek, A., Kang, H.M., and Kwak, H. (2020). Population-scale study of eRNA transcription reveals bipartite functional enhancer architecture. *Nat. Commun.* 11, 5963. <https://doi.org/10.1038/s41467-020-19829-z>.
 79. Cai, D., Feliciano, D., Dong, P., Flores, E., Gruebele, M., Porat-Shliom, N., Sukenik, S., Liu, Z., and Lippincott-Schwartz, J. (2019). Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. *Nat. Cell Biol.* 21, 1578–1589. <https://doi.org/10.1038/s41556-019-0433-z>.
 80. Giniger, E., and Ptashne, M. (1988). Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Natl. Acad. Sci. USA* 85, 382–386. <https://doi.org/10.1073/pnas.85.2.382>.
 81. Fromental, C., Kanno, M., Nomiya, H., and Chambon, P. (1988). Cooperativity and hierarchical levels of functional organization in the SV40 enhancer. *Cell* 54, 943–953. [https://doi.org/10.1016/0092-8674\(88\)90109-2](https://doi.org/10.1016/0092-8674(88)90109-2).
 82. Kim, T.K., and Maniatis, T. (1997). The mechanism of transcriptional synergy of an in vitro assembled interferon-beta enhanceosome. *Mol. Cell* 1, 119–129. [https://doi.org/10.1016/S1097-2765\(00\)80013-1](https://doi.org/10.1016/S1097-2765(00)80013-1).
 83. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
 84. Geisel, N., and Gerland, U. (2011). Physical limits on cooperative protein-DNA binding and the kinetics of combinatorial transcription regulation. *Biophys. J.* 101, 1569–1579. <https://doi.org/10.1016/j.bpj.2011.08.041>.
 85. Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B., and Jiang, J. (2008). The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. *Dev. Cell* 14, 377–387. <https://doi.org/10.1016/j.devcel.2008.01.006>.
 86. Lin, K.C., Park, H.W., and Guan, K.-L. (2017). Regulation of the hippo pathway transcription factor TEAD. *Trends Biochem. Sci.* 42, 862–872. <https://doi.org/10.1016/j.tibs.2017.09.003>.
 87. Anbanandam, A., Albarado, D.C., Nguyen, C.T., Halder, G., Gao, X., and Veeraraghavan, S. (2006). Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proc. Natl. Acad. Sci. USA* 103, 17225–17230. <https://doi.org/10.1073/pnas.0607171103>.
 88. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009>.
 89. Xiao, J.H., Davidson, I., Matthes, H., Garnier, J.M., and Chambon, P. (1991). Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* 65, 551–568. [https://doi.org/10.1016/0092-8674\(91\)90088-g](https://doi.org/10.1016/0092-8674(91)90088-g).
 90. Davidson, I., Xiao, J.H., Rosales, R., Staub, A., and Chambon, P. (1988). The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. *Cell* 54, 931–942. [https://doi.org/10.1016/0092-8674\(88\)90108-0](https://doi.org/10.1016/0092-8674(88)90108-0).
 91. Herr, W. (1993). The SV40 enhancer: Transcriptional regulation through a hierarchy of combinatorial interactions. *Semin. Virol.* 4, 3–13. [https://doi.org/10.1016/1044-5773\(93\)80003-7](https://doi.org/10.1016/1044-5773(93)80003-7).
 92. Jacquemin, P., Hwang, J.J., Martial, J.A., Dollé, P., and Davidson, I. (1996). A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. *J. Biol. Chem.* 271, 21775–21785. <https://doi.org/10.1074/jbc.271.36.21775>.
 93. Jacquemin, P., and Davidson, I. (1997). The role of the TEF transcription factors in cardiogenesis and other developmental processes. *Trends Cardiovasc. Med.* 7, 192–197. [https://doi.org/10.1016/S1050-1738\(97\)00052-2](https://doi.org/10.1016/S1050-1738(97)00052-2).
 94. Jiang, S.W., Desai, D., Khan, S., and Eberhardt, N.L. (2000). Cooperative binding of TEF-1 to repeated GGAATG-related consensus elements with restricted spatial separation and orientation. *DNA Cell Biol.* 19, 507–514. <https://doi.org/10.1089/10445490050128430>.
 95. Halder, G., and Carroll, S.B. (2001). Binding of the Vestigial co-factor switches the DNA-target selectivity of the Scalloped selector protein. *Development* 128, 3295–3305. <https://doi.org/10.1242/dev.128.17.3295>.
 96. Lee, B.-K., Jang, Y.J., Kim, M., LeBlanc, L., Rhee, C., Lee, J., Beck, S., Shen, W., and Kim, J. (2019). Super-enhancer-guided mapping of regulatory networks controlling mouse trophoblast stem cells. *Nat. Commun.* 10, 4749. <https://doi.org/10.1038/s41467-019-12720-6>.
 97. Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Res.* 44, D726–D732. <https://doi.org/10.1093/nar/gkv1160>.
 98. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* 48, D882–D889. <https://doi.org/10.1093/nar/gkz1062>.
 99. Lee, D.-S., Vonrhein, C., Albarado, D., Raman, C.S., and Veeraraghavan, S. (2016). A Potential Structural Switch for Regulating DNA-Binding by TEAD Transcription Factors. *J. Mol. Biol.* 428, 2557–2568. <https://doi.org/10.1016/j.jmb.2016.03.008>.
 100. Patel, L.R., Curran, T., and Kerppola, T.K. (1994). Energy transfer analysis of Fos-Jun dimerization and DNA binding. *Proc. Natl. Acad. Sci. USA* 91, 7360–7364. <https://doi.org/10.1073/pnas.91.15.7360>.
 101. Chen, L. (1999). Combinatorial gene regulation by eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* 9, 48–55. [https://doi.org/10.1016/S0959-440X\(99\)80007-4](https://doi.org/10.1016/S0959-440X(99)80007-4).
 102. Shi, Z., He, F., Chen, M., Hua, L., Wang, W., Jiao, S., and Zhou, Z. (2017). DNA-binding mechanism of the Hippo pathway transcription factor TEAD4. *Oncogene* 36, 4362–4369. <https://doi.org/10.1038/onc.2017.24>.
 103. Panne, D. (2008). The enhanceosome. *Curr. Opin. Struct. Biol.* 18, 236–242. <https://doi.org/10.1016/j.sbi.2007.12.002>.

104. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
105. MacKerell, A.D., Banavali, N., and Foloppe, N. (2000). Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56, 257–265.
106. Pan, A.C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T.M., and Shaw, D.E. (2019). Atomic-level characterization of protein-protein association. *Proc. Natl. Acad. Sci. USA* 116, 4244–4249. <https://doi.org/10.1073/pnas.1815431116>.
107. Kaan, H.Y.K., Chan, S.W., Tan, S.K.J., Guo, F., Lim, C.J., Hong, W., and Song, H. (2017). Crystal structure of TAZ-TEAD complex reveals a distinct interaction mode from that of YAP-TEAD complex. *Sci. Rep.* 7, 2035. <https://doi.org/10.1038/s41598-017-02219-9>.
108. Chen, Y., Bates, D.L., Dey, R., Chen, P.-H., Machado, A.C.D., Laird-Offringa, I.A., Rohs, R., and Chen, L. (2012). DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.* 2, 1197–1206. <https://doi.org/10.1016/j.celrep.2012.10.012>.
109. Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100. [https://doi.org/10.1016/0092-8674\(95\)90136-1](https://doi.org/10.1016/0092-8674(95)90136-1).
110. Louvet, S., Aghion, J., Santa-Maria, A., Mangeat, P., and Maro, B. (1996). Ezrin becomes restricted to outer cells following asymmetrical division in the preimplantation mouse embryo. *Dev. Biol.* 177, 568–579. <https://doi.org/10.1006/dbio.1996.0186>.
111. Zenker, J., White, M.D., Gasnier, M., Alvarez, Y.D., Lim, H.Y.G., Bissiere, S., Biro, M., and Plachta, N. (2018). Expanding actin rings zipper the mouse embryo for blastocyst formation. *Cell* 173, 776–791. <https://doi.org/10.1016/j.cell.2018.02.035>.
112. Janssens, J., Aibar, S., Taskiran, I.I., Ismail, J.N., Gomez, A.E., Aughey, G., Spanier, K.I., De Rop, F.V., González-Blas, C.B., Dionne, M., et al. (2022). Decoding gene regulation in the fly brain. *Nature* 607, 630–636. <https://doi.org/10.1038/s41586-021-04262-z>.
113. Sönmez, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., and Krebs, A.R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. *Mol. Cell* 81, 255–267. <https://doi.org/10.1016/j.molcel.2020.11.015>.
114. Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., and Greenberg, M.E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol. Cell* 68, 1067–1082. <https://doi.org/10.1016/j.molcel.2017.11.026>.
115. Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. USA* 107, 22534–22539. <https://doi.org/10.1073/pnas.0913805107>.
116. Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell Biol.* 15, 1405–1421. <https://doi.org/10.1128/MCB.15.3.1405>.
117. Miller, J.A., and Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. *Mol. Cell Biol.* 23, 1623–1632. <https://doi.org/10.1128/MCB.23.5.1623-1632.2003>.
118. Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol. Cell* 75, 921–932. <https://doi.org/10.1016/j.molcel.2019.06.009>.
119. Kenny, C., Dilshat, R., Seberg, H.E., Van Otterloo, E., Bonde, G., Helverson, A., Franke, C.M., Steingrimsson, E., and Cornell, R.A. (2022). TFAP2 paralogs facilitate chromatin access for MITF at pigmentation and cell proliferation genes. *PLoS Genet.* 18, e1010207. <https://doi.org/10.1371/journal.pgen.1010207>.
120. Bragança, J., Eloranta, J.J., Bamforth, S.D., Ibbitt, J.C., Hurst, H.C., and Bhattacharya, S. (2003). Physical and functional interactions among AP-2 transcription factors, p300/CREB-binding protein, and CITED2. *J. Biol. Chem.* 278, 16021–16029. <https://doi.org/10.1074/jbc.M208144200>.
121. Trainor, C.D., Omichinski, J.G., Vandergon, T.L., Gronenborn, A.M., Clore, G.M., and Felsenfeld, G. (1996). A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol. Cell Biol.* 16, 2238–2247. <https://doi.org/10.1128/MCB.16.5.2238>.
122. Maven, B.E.J., Gifford, C.A., Weiler, M., Gonzalez-Teran, B., Hüttenhain, R., Pelonero, A., Ivey, K.N., Samse-Knapp, K., Kwong, W., Gordon, D., et al. (2023). The multi-lineage transcription factor ISL1 controls cardiomyocyte cell fate through interaction with NKX2.5. *Stem Cell Rep.* 18, 2138–2153. <https://doi.org/10.1016/j.stemcr.2023.09.014>.
123. Gressel, S., Lidschreiber, K., and Cramer, P. (2019). Transient transcriptome sequencing: experimental protocol to monitor genome-wide RNA synthesis including enhancer transcription v1. <https://doi.org/10.17504/protocols.io.3vzgn76>.
124. Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.* 8, 2022–2032. <https://doi.org/10.1038/nprot.2013.118>.
125. Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., et al. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* 153, 044130. <https://doi.org/10.1063/5.0014475>.
126. Fiorin, G., Klein, M.L., and Hénin, J. (2013). Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* 111, 3345–3362. <https://doi.org/10.1080/00268976.2013.813594>.
127. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. j.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.
128. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
129. 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. <https://doi.org/10.1093/bioinformatics/bts635>.
130. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, 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. <https://doi.org/10.1186/gb-2008-9-9-r137>.
131. Chollet. (2015). Francois and others (Keras). <https://keras.io>.
132. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1603.04467>.
133. Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1412.6980>.
134. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis | SpringerLink. https://link.springer.com/book/10.1007/978-3-319-24277-4?trk=public_post_comment-text.
135. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
136. Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., III, Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., et al. (2020). AMBER 2020 (University of California).
137. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>.

138. Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* 23, 566–579. <https://doi.org/10.1002/prot.340230412>.
139. Ahlers, J., Althviz Moré, D., Amsalem, O., Anderson, A., Bokota, G., and Boone, P. (2023). napari: a multi-dimensional image viewer for Python. Zenodo. <https://zenodo.org/records/15029515>.
140. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>.
141. Terootea, T.W., Pozner, A., and Buck-Koehntop, B.A. (2016). PatCh-Cap: input strategy for improving analysis of ChIP-exo data sets and beyond. *Nucleic Acids Res.* 44, e159. <https://doi.org/10.1093/nar/gkw741>.
142. Koenecke, N., Johnston, J., He, Q., Meier, S., and Zeitlinger, J. (2017). Drosophila poised enhancers are generated during tissue patterning with the help of repression. *Genome Res.* 27, 64–74. <https://doi.org/10.1101/gr.209486.116>.
143. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14, 959–962. <https://doi.org/10.1038/nmeth.4396>.
144. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218. <https://doi.org/10.1038/nmeth.2688>.
145. Hennig, B.P., Velten, L., Racke, I., Tu, C.S., Thoms, M., Rybin, V., Besir, H., Remans, K., and Steinmetz, L.M. (2018). Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3 (Bethesda)* 8, 79–89. <https://doi.org/10.1534/g3.117.300257>.
146. Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033–2040. <https://doi.org/10.1101/gr.177881.114>.
147. Connelly, J.P., and Pruett-Miller, S.M. (2019). CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. *Sci. Rep.* 9, 4194. <https://doi.org/10.1038/s41598-019-40896-w>.
148. 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. <https://doi.org/10.1038/nmeth.2019>.
149. Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., and Hutchison, G.R. (2012). Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* 4, 17. <https://doi.org/10.1186/1758-2946-4-17>.
150. Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Miguez, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theor. Comput.* 16, 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
151. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., et al. (2016). Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* 13, 55–58. <https://doi.org/10.1038/nmeth.3658>.
152. Izadi, S., Anandakrishnan, R., and Onufriev, A.V. (2014). Building water models: A different approach. *J. Phys. Chem. Lett.* 5, 3863–3871. <https://doi.org/10.1021/jz501780a>.
153. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 5, 1752–1779. <https://doi.org/10.1214/11-AOAS466>.
154. Ramírez, 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, W160–W165. <https://doi.org/10.1093/nar/gkw257>.
155. 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. <https://doi.org/10.1093/bioinformatics/btp352>.
156. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. <https://doi.org/10.1093/nar/gku365>.
157. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18, 100–106. <https://doi.org/10.1038/s41592-020-01018-x>.
158. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
159. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
160. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
161. McAnany, C.E., and Mura, C. (2016). Claws, Disorder, and Conformational Dynamics of the C-Terminal Region of Human Desmoplakin. *J. Phys. Chem. B* 120, 8654–8667. <https://doi.org/10.1021/acs.jpcc.6b03261>.
162. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>.
163. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-TEAD4	Abcam	ab58310
Goat polyclonal anti-TFAP2C	R&Dsystems	AF5059
Rabbit polyclonal anti-CDX2	Bethyl Laboratories	A300-692A
Rabbit monoclonal anti-GATA3	Cell Signaling	5852T
Rabbit polyclonal anti-YAP1	Cell Signaling	14074S
Rabbit monoclonal anti-Pol II	Cell Signaling	D8L4Y
Rabbit polyclonal anti-H3K27ac	Active Motif	39135
Mouse monoclonal anti-CDX2 for immunofluorescence	BioGenex	MU392A-5UC
Rabbit monoclonal anti-Nanog for immunofluorescence	Cell Signaling	8822
Anti-rabbit IgG Alexa Fluor 488 secondary antibody	Biotium	20015
Anti-rabbit IgG Alexa Fluor 647 secondary antibody	Biotium	20047
Chemicals, peptides, and recombinant proteins		
37% formaldehyde solution	VWR	50-00-0
Dynabeads Protein A	ThermoFisher	10008D
phi29 DNA polymerase	New England Biolabs	M0269S
Lambda exonuclease	New England Biolabs	M0262S
Q5 High-Fidelity 2x Master Mix	New England Biolabs	M0492S
dNTP solution mix	New England Biolabs	N0447S
RNase A	ThermoFisher	EN0531
Phenol:chloroform:isoamyl alcohol (25:24:1) (v/v/v)	VWR	136112-00-0
Proteinase K	ThermoFisher	25530049
DAPI	BioLegend	422801
crRNA and ssODN sequences	IDT	Table S2
Alt-R HiFi Cas9 Nuclease V3 protein and tracerRNA, ATTO-550	IDT	1081059 and 1075928
rLV.EF1.tdTomato-9	Takara	0036VCT
Critical commercial assays		
End Repair Module	New England Biolabs	E6050S
dA-Tailing Module	New England Biolabs	E6053S
Quick Ligation Kit	New England Biolabs	M2200S
Monarch DNA Gel Extraction Kit	New England Biolabs	T1020
Monarch PCR & DNA Cleanup Kit	New England Biolabs	T1030
Hybridization Chain Reaction (HCR) v3.0	Molecular Instruments	N/A
Infusion cloning	Takara	638947
Dual-Glo luciferase assay system	Promega	N1521
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold Set	Illumina	RS-122-2303
TruSeq poly-A Stranded mRNA Library Prep Kit	Illumina	20020595
NEBNext Ultra II DNA library prep kit	NEB	E7645
Deposited data		
Raw and analyzed NGS and PBM data	This paper	GEO: GSE252463
Trained deep learning model	This paper	https://zenodo.org/records/14894986

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Raw images and MD trajectories	This paper	https://collaboration.stowers.org/d1e17b38-9a3d-4900-a623-ac32836b32fc/
Experimental models: Organisms/strains		
Mouse Trophoblast Stem Cells (TSCs)	Singh et al. ⁵⁵	https://pubmed.ncbi.nlm.nih.gov/33458704/
Mouse Embryonic Stem Cells (ESCs)	Avsec et al. ¹⁵	https://www.nature.com/articles/s41588-021-00782-6#Sec10
First CRISPR TSC line: CRISPR-Cas9 at <i>Tead motif</i> within locus chr17:6,827,802-6,827,811 in mouse Trophoblast Stem Cells (TSC)	This paper	N/A
Second CRISPR TSC line: CRISPR-Cas9 at <i>Tead motif</i> within locus chr12:102,262,024-102,262,033 in mouse Trophoblast Stem Cells (TSC)	This paper	N/A
Oligonucleotides		
Oligonucleotides for ChIP-nexus, see Table S2	IDT	https://research.stowers.org/zeitlingerlab/protocols.html
Illumina Index primer 1: 5'-CAAGCAGAAGA CGGCATACGAGAT[i7]GTCTCGTGGGCTCGG-3'	IDT	https://support-docs.illumina.com/SHARE/AdapterSequences/Content/SHARE/AdapterSeq/Nextera/SequencesNextera_Illumina.htm
Illumina Index primer 2: 5'-AATGATACGGC GACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC-3'	IDT	https://support-docs.illumina.com/SHARE/AdapterSequences/Content/SHARE/AdapterSeq/Nextera/SequencesNextera_Illumina.htm
Illumina Transposase adapter read 1 (Nextera A): 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'	IDT	https://support-docs.illumina.com/SHARE/AdapterSequences/Content/SHARE/AdapterSeq/Nextera/SequencesNextera_Illumina.htm
Illumina Transposase adapter read 2 (Nextera B): 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'	IDT	https://support-docs.illumina.com/SHARE/AdapterSequences/Content/SHARE/AdapterSeq/Nextera/SequencesNextera_Illumina.htm
Mosaic end primer: 5Phos/CTGTCTCTTATACA/3ddC/	IDT	Tn5mC1.1-A1block
Recombinant DNA		
pETM11-Sumo3-Tn5 plasmid	Hennig et al. ¹²³	E54K,L372P
His6-tagged SenP2 protease plasmid	Hennig et al. ¹²³	N/A
Software and algorithms		
FIJI	Schindelin et al. ¹²⁴	https://fiji.sc/
Cutadapt v.2.5	Martin 2011 ¹²⁵	https://cutadapt.readthedocs.io/en/v2.5/
Bowtie (v.1.1.2) and Bowtie2 (v.2.4.2)	Langmead and Salzberg ¹²⁶	https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
MACS2 v.2.2.6	Zhang et al. ¹²⁷	https://github.com/macs3-project/MACS
Irreproducible Discovery Rate framework v.2.0.3	Li et al. ¹²⁸	https://github.com/nboley/idr
STAR v.2.7.3	Dobin et al. ¹²⁹	https://code.google.com/archive/p/rna-star/
deepTools2 v.3.1.3	Ramírez et al. ¹³⁰	https://deeptools.readthedocs.io/en/latest/
BPNet software v.0.0.23	Avsec et al. ¹⁵	https://github.com/kundajelab/bpnet/
Keras v.2.2.4	Chollet F ¹³¹	https://pypi.org/project/keras/
TensorFlow1 backend v.1.7	Abadi et al. ¹³²	https://www.tensorflow.org/install/pip
Adam optimizer	Kingma and Ba ¹³³	N/A
DeepLIFT v.0.6.9.0	Shrikumar et al. ⁶³	https://github.com/kundajelab/DeepExplain
TF-MoDISco v.0.4.2.2	Shrikumar et al. ⁶⁴	https://github.com/kundajelab/tfmodisco

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DESeq2 v.1.34.0	Love et al. ⁸³	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
R v.4.1.1 and v.4.2.3	R core team	https://www.r-project.org/
Rstudio	RStudio	https://rstudio.com
ggplot2 v.3.5.1	Wickham ¹³⁴	https://ggplot2.tidyverse.org/
VMD 1.9.3	Humphrey et al. ¹³⁵	http://www.ks.uiuc.edu/Research/vmd/
AmberTools20 and CHARMM27	Case et al. ¹³⁶ and Mackerell et al. ¹⁰⁴	https://ambermd.org/AmberTools.php https://mackerell.umaryland.edu/charmm_ff.shtml#gromacs
MEME v.5.3 and v.5.5.3	Bailey et al. ¹³⁷	https://meme-suite.org/meme/tools/meme
FIMO v.5.5.3	Grant et al. ¹³⁸	https://meme-suite.org/meme/tools/fimo
AlphaFold3	Abramson et al. ¹⁰²	https://alphafoldserver.com/
Napari v0.4.0	Ahlers et al. ¹³⁹	https://doi.org/10.5281/zenodo.3555620
Cellpose deep learning	Stringer et al. ¹⁴⁰	www.cellpose.org
Other		
All code and analyses that contributed to this work	This paper	https://github.com/zeitlingerlab/Dalal_hippo_signaling_2024
Bioruptor Pico sonication device	Diagenode	https://www.diagenode.com/en/p/bioruptor-pico-sonication-device
Confocal scanning microscope	Zeiss	800
Spinning disk microscope	Nikon	Eclipse Ti2
Quantification and Statistical Analysis		
Linear regression model was implemented in R v.4.1.1 using the lm function. From the fitted model, key parameters (slope, intercept, <i>p</i> -value) were extracted. (Table S5)	R core team	https://www.r-project.org/

METHOD DETAILS

Mouse stem cell culture

Mouse trophoblast stem cells (TSCs) were a gift from Vijay Pratap Singh and were maintained in a feeder-free culture as described.⁵⁵ Briefly, feeder conditioned medium (Feeder-CM) was prepared by culturing γ -irradiated MEFs (mouse embryonic fibroblasts) in TS medium (RPMI 1640 medium, FBS 20%, 50 μ g/mL of Penicillin and streptomycin (100 $\times$), 1 mM of Sodium pyruvate (100 mM), 0.1 mM of β -Mercaptoethanol (20 mM), 2 mM GlutaMAX (200 mM) for 72 h and then filtered with a 0.45 μ m filter. 70% Feeder-CM plus 30% TS medium (70cond) supplemented with growth factor FGF4 (R&D System) and heparin (Sigma) (70cond +1.5 $\times$ F4H Medium) was used to maintain TSCs in feeder-free conditions. Mouse embryonic stem cells (ESCs) were cultured and maintained as previously described.¹⁵

ChIP-nexus, PAtCh-Cap, and ChIP-seq experiments

For each ChIP-nexus experiment, 10⁶ TSCs were used. Cells were washed with PBS and cross-linked with 1% formaldehyde (Fisher Scientific) in PBS for 10 min at RT. The reaction was quenched with 125 mM glycine. Fixed cells were washed twice with cold PBS and resuspended in cold lysis buffer (15 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5% *N*-lauroylsarcosine, 0.1% sodium deoxycholate and 0.1% SDS), incubated for 10 min on ice and sonicated with a Bioruptor Pico (Diagenode) for five cycles of 30 s on and 30 s off. The ChIP-nexus procedure and data processing were performed as previously described,³⁸ except that the ChIP-nexus adapter mix contained four fixed barcodes (ACTG, CTGA, GACT, and TGAC), and PCR library amplification was performed directly after circularization of the purified DNA fragments (without the addition of the oligo and BamHI digestion). PAtCh-Cap was performed as previously described¹⁴¹ with 10% of sheared chromatin from 10⁶ TSCs. ChIP-seq experiments, including a whole cell extract (WCE) control, were performed as described¹⁴² with 10⁶ TSCs per ChIP. For each ChIP, 5–10 μ g of antibody was coupled to 50–100 μ L of Protein A or Protein G Dynabeads (Invitrogen). The following antibodies were used: anti-TFAP2C (R & D systems, no. AF5059), anti-TEAD4 (Abcam, no. ab58310), anti-CDX2 (Bethyl Laboratories, no. A300-692A), anti-GATA3 (Cell Signaling, no. 5852T), anti-YAP1 (Cell Signaling, no. 14074S), anti-Pol II (Cell Signaling, no. D8L4Y), anti-H3K27ac (ChIP-seq) (Active motif, no. 39135). For all experiments, at least two biological replicates were prepared—that is, the

experiments were performed on different days, starting with cells from a different passage number. Single-end sequencing was performed on an Illumina NextSeq 500 instrument (75 cycles). The full ChIP-nexus protocol can be found on the Zeitlinger lab website at <https://research.stowers.org/zeitlingerlab/protocols.html>.

Luciferase assays

Selected genomic regions of range 175–200 bp were synthesized using GeneArt Strings DNA fragments along with restriction enzyme sites for with Kpn1 (NEB, R0142) and Xho1 (NEB, R0146) to clone in pNL3.2 vector. pNL3.2 vectors (Promega) were digested and cloned using an Infusion master mix (Takara) upstream of the luciferase gene. Stellar competent cells (Takara) were used for transformation and downstream miniprep (Qiagen), following the manufacturer's protocol. The cloned sequences were confirmed using the Sanger sequencing method. 2.5–5⁵ TSCs were used to transfect a total of 500 ng DNA with lipofectamine2000 in a ratio of 1:2 (DNA to lipofectamine2000), following the manufacturer's protocol. Cells were co-transfected with 1:100 ratio of control (pGL4.54[luc2/TK]) and reporter construct (pNL3.2[NlucP/minP]). Cells were transfected in suspension for 15–20 min and resuspended with media to grow in each well of the 24-well plate.

Luciferase assays were performed using a Dual-Glo luciferase assay system (Promega). After 24 h, cells were harvested, and the NanoDLR assay protocol was followed per the manufacturer's instructions to take luminescence measurements with SpectraMax iD3 Plate Reader. Reporter luminescence signals were normalized according to their corresponding control luminescence signals, resulting in relative luciferase activity. Replicate luciferase assay experiments were performed independently three times (Table S3).

TT-seq experiments

TT-seq experiments were performed in three biological replicates on TSCs across three biological replicates, as described in¹²³ <https://www.protocols.io/view/transient-transcriptome-sequencing-experimental-pr-3byl42y22vo5/v1>. Libraries were prepared using TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold Set to degrade rRNA. Approximately 10⁶ cells were seeded in a 100 cm dish (~80% confluency) and incubated with 500 μ M of 4sU (Sigma) at 37°C, 5% CO₂ for 15 min. The cells were then collected by adding 4.5 mL of TRIzol lysis reagent (Life Technologies Corp), incubated for 5 min on ice, flash-frozen, and stored at –80°C. In the biotinylation of the 4sU labeled RNA step, acid-phenol-chloroform (ThermoFisher) was used instead of chloroform.

ATAC-seq experiments

For each ATAC-seq experiment, 1–5⁵ or 2⁵ TSCs were harvested, washed with PBS, and resuspended in ATAC Resuspension Buffer (RSB, 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 3mM MgCl₂) with 0.1% IGEPAL CA-630. Tn5 transposition was performed as previously described.^{143,144} Briefly, the cells were incubated for 3 min on ice in ATAC RSB supplemented with 0.1% IGEPAL CA-630, 0.1% Tween 20, and 0.01% Digitonin (Promega, G9441). The reaction was quenched with ATAC RSB with 0.1% Tween 20 and centrifugation. Tagmentation took place at 37°C and 1000 rpm for 30 min in a 50 μ L reaction volume containing 10 μ L of 5x Tagmentation Buffer (50 mM Tris-HCl pH 7.5, 25 mM MgCl₂, 50% DMF), 0.5 μ L 10% Tween 20, 0.5 μ L 1% Digitonin, 1–2 μ M assembled transposome and water. Tn5 transposase was purified in-house as previously described.¹⁴⁵ Tn5 was loaded with previously reported oligonucleotides Tn5ME-A/Tn5mC1.1-A1block and Tn5ME-B/Tn5mC1.1-A1block^{124,146} by mixing equal amounts of purified Tn5 protein and annealed oligonucleotides for 30 min at RT. After tagmentation, the DNA fragments were purified using the Monarch PCR & DNA Cleanup Kit (NEB). Libraries were constructed using Illumina Nextera Dual Indexing, and qPCR was used to prevent over-amplification as described.¹⁴³ At least three biological replicates were generated, and paired-end sequencing was performed on an Illumina NextSeq 500 instrument (2 x 75 bp).

RNA-seq experiments

TSCs wild-type and CRISPR-Cas9 edited cells were grown separately in wells of a 6-well plate and harvested at 80% confluency (~2⁶ cells) using 500 μ L of TRIzol reagent (Life Technologies Corp). The lysate was incubated for 5 min on ice, flash-frozen, and stored at –80°C. For RNA extraction, the lysate was quickly thawed at 65°C, cooled on ice for 5 min, and vortexed. Then, 100 μ L of chloroform was added per 0.5 mL of TRIzol lysis reagent, shaken vigorously for 15 s, and incubated for 3 min. The sample was centrifuged at 4°C and 7000 x g for 25 min, and the upper colorless aqueous phase was transferred to a new tube. 250 μ L of isopropanol was then added, incubated for 10 min at 4°C, and centrifuged at 4°C and 12,000 x g for 10 min. The total RNA precipitate formed a white gel-like pellet at the bottom of the tube, which was washed with 75% ethanol, air-dried for 5–10 min, and resuspended in 20 μ L of RNase-free water. DNase treatment was performed using the TURBO DNase kit per the manufacturer's instructions: adding 1 μ L of TURBO DNase and 2 μ L of DNase buffer to the dissolved RNA and incubating at 37°C for 30 min. To inactivate TURBO DNase, the RNA samples were extracted with phenol/chloroform (Sigma). The sample was centrifuged at 4°C and 7000 x g for 25 min, and the upper colorless aqueous phase was transferred to a new tube. 250 μ L of isopropanol was added, incubated for 10 min at 4°C, and centrifuged at 4°C and 12,000 x g for 10 min. The total RNA precipitate formed a white gel-like pellet at the bottom of the tube, which was washed with 75% ethanol, air-dried for 5–10 min, and resuspended in 20 μ L of RNase-free water. The samples were incubated in a water bath or heat block set at 55°C–60°C for 10–15 min. The RNA concentration was determined using a NanoDrop spectrophotometer, and the RNA integrity was checked on a 2100 Bioanalyzer using an Agilent RNA 6000 Nano Kit. mRNA-stranded libraries were prepared using a TruSeq poly-A Stranded mRNA Library Prep Kit and sequenced on an Illumina NextSeq 2000 P2 platform with 2 x 100 bp single-end reads. Three biological replicates were performed for wild-type and CRISPR-Cas9 edited cells.

CRISPR-Cas9 experiments

In the first CRISPR TSC line, the *Tead double* motif on chr17:6,827,802-6,827,811 (mm10) was mutated from ACATTCCAGA (wild-type) to GCATTCCAGGAATTCCA (mutant). In a second CRISPR TSC line, the *Tead single* motif (CACATTCCTA) on chr12:102,262,024–102,262,033 (mm10) was first inserted at 60 bp downstream of the *TFAP2C* motif (GGGCCCCAGGGCC) and then in a second round of CRISPR-Cas9 editing *Tead single* motif was mutated from CACATTCCTA (wild-type) to CACCGTCCTA (mutant) at its original position. crRNA target sites were designed using the IDT target predictor tool by evaluating the predicted on-target efficiency score and off-target potential. Alt-R CRISPR-Cas9 crRNA was designed to contain ~40 bases of homology from the targeted cut site (gRNA and ssODN sequences are shown in Table S2). Equimolar amounts (stock of 100 μ M) of Alt-R crRNA and tracrRNA-ATTO550 were mixed to form gRNA at a final concentration of 50 μ M. The mixture was heated at 95°C for 5 min and cooled at RT. The single-stranded donor oligonucleotides (ssODN) were designed to contain ~40 bases of homology from the targeted cut site (crRNA and ssODN sequences were designed using the IDT software tool). A ribonucleoprotein (RNP) complex was formed by combining 150 pmol of gRNA (crRNA+tracrRNA) and 125 pmol of Cas9 HiFi v3 protein (IDT) with hybridization for 20 min at RT. The RNP was combined with 100 pmol of ssODN donor and 100 pmol of electroporation enhancer v2 and delivered to 1.5×10^5 cells by Neon electroporation (1,400 V, 10 ms, 3 pulses; Neon Transfection System, MPK5000, Life Technologies). Immediately after electroporation, cells were cultured in 0.5 μ M Alt-R HDR enhancer V2 of 0.69 mM. After 24 h, cells were washed with PBS before FACS sorting on S6 FACSsymphony. Single cells were directly sorted into 96-well plates. Cells were screened for the expected mutations through paired-end sequencing on an Illumina MiSeq instrument (250 cycles). On-target indel frequency and expected mutations were analyzed using CRIS.py.¹⁴⁷ Clones with the intended homozygous mutation and sequence alignments >90% were chosen for further experiments, except for the 2nd CRISPR line, where we found one Indel and SNP within 500 bp of the original *Tead single* motif position, but these changes were predicted to be neutral by BPNNet.

Mice strains and superovulation

C57BL/6J (B6) strain of mice were used from the Stowers Institute for Medical Research (SIMR) core production colony. Three to four-week-old females were superovulated using 5IU of PMSG (Genway Biotech), followed by 5IU HCG (Sigma Aldrich) 46–48 h later. Following HCG, females were paired with B6 males and checked for a copulatory plug the next morning, indicating successful mating. Fertilized embryos were collected from the plugged females at 1.5 dpc (2-cell stage) by flushing M2 (Millipore) through the infundibulum and out the uterine horn using a blunt needle. Embryos were then incubated overnight at 37°C under 5% CO₂ in humidified air in 4-well culture dishes containing KSOM media (Millipore). Experiments were approved by the SIMR IACUC and were performed following the committees' guiding principles.

Lentivirus transduction of fluorescent td-tomato in TSCs

Two days (48 h) before the transduction of wild-type or CRISPR-Cas9 edited (putative *Ezr* region edits) TSCs, cells were seeded at 1×10^5 cells per well in triplicate with 3 mL of media per well of 6-well plate. On the day of transduction, cells were small-sized colonies of about 30–40% confluence; the old media were removed and washed once with PBS and replaced with 2 mL of media. The cells were infected with prepackaged lentiviral particles (constitutive reporter vector expressing tdTomato fluorescent protein gene driven by EF1a promoter (Takara) at MOI of 20 (Stock: 3.5×10^{-10} TU/ml) along with polybrene (4 μ g/ml) (Sigma) for 24 h before being replaced with a fresh medium. Four days after transduction, the td-Tomato-positive cells were selected using puromycin antibiotic selection (1 μ g/ml) (InvivoGen) and were kept under selection until the positive colonies reached 60–80% confluence. Once cells reached 80% confluence, the positive cells were FACS sorted on S6 FACSsymphony, expanded, and used for embryo aggregation experiments.

Aggregation assays to obtain chimeric embryos

To prepare the aggregation plates, six indentations on the bottom of the 35 $\times$ 10 mm plates were made using an aggregation needle (BLS) sterilized with 70% alcohol and added a drop of KSOM. All drops of KSOM were covered with mineral oil (Sigma). On the morning of the aggregation, the embryos (8–16 cell stage) were washed through M2 and then placed in drops of Tyrode's solution (Sigma). After about 30 s, the zona pellucida began to dissolve. Once the zona was dissolved, the embryos were picked up and rinsed through a drop of M2 to neutralize the Tyrode's solution, then placed in a drop of KSOM. Embryos were moved from this dish to the aggregate plates, placing an embryo into each indentation. Clumps of td-Tomato transduced TSCs were then picked up with a mouth pipette and moved onto each embryo in the aggregate plate. Once settled and in contact with the embryo, the aggregation plates were cultured in an incubator at 37°C under 5% CO₂ in humidified air for 46–48 h until the embryos reached the blastocyst stage. Chimeric blastocysts were fixed with 4% paraformaldehyde (ThermoFisher) for 20 min and washed three times with PBS before mounting them on a glass bottom plate (Cellvis) coated with poly-L-lysine (Sigma). Embryos were imaged with a Zeiss LSM800, an upright confocal laser scanning microscope.

Immunofluorescence stainings of chimeric embryos

A few fixed chimeric blastocysts were used for immunofluorescence stainings. The embryos were washed three times with PBS-T (PBS with 0.1% Triton X-100) and incubated in PBS-T for 1 h or longer at 4°C. Embryos were then washed two times with PBS for 10 min each and incubated with 300 μ L of superbloc solution (ThermoFisher) for 90 min at RT, before adding the primary antibodies: CDX2 (BioGenex-MU392A-5UC) and Nanog (Cell Signaling, 8822S) with 10 μ g/mL of DAPI from BioLegend:422801. The CDX2

antibody came with a signal enhancing reagent, which was used to replace 75% of the superblock solution while incubating with the primary antibody. The embryos were incubated overnight at 4°C, covered from light. The next day, the samples were washed for 10 min each three times with PBST (0.1% Triton X-100) at RT and once with PBS at RT. Secondary antibodies were added (biotin:20015,20047) in special PBS (ThermoFisher) at a 1:300 dilution with DAPI 2 μ L in 1 mL of 10 μ g/mL (BioLegend) and kept on light rotation for 2 h at RT. Samples were then washed three times with PBS for 10 min. Samples were imaged immediately or kept for up to a week at 4°C before imaging. Imaging was performed with an upright confocal laser scanning microscope (Zeiss LSM800) with 40x magnification. Maximum intensity Z projections and adjustments to the brightness and contrast were performed in ImageJ/FIJI.¹⁴⁸ Samples larger than the field of view were taken as tiled images and stitched with the Grid/Collection Stitching plugin in ImageJ.

HCR-FISH on chimeric embryos

Embryos were fixed in 4% paraformaldehyde for 20 min and washed three times in PBS +0.2% Triton X- for 10 min each. RNA FISH experiments were performed using HCR v3.0 (Molecular Instrument Inc.). The RNA sequences that were used to design probes are listed along with the chosen amplifiers and probe set size: Ezr (NM_009510.2, B4,32), CDX2 (NM_007673.3, B1,29), tdTomato (B5,16). The following amplifiers with Alexa fluorophore were used: 488, 546, and 647. The fixed embryos were serially dehydrated into methanol and stored at -20°C until use. To rehydrate the embryos, they were washed in PBS +0.1% Triton X-(PBST). Embryos were incubated in the hybridization buffer for 30 min at 37°C, then in the hybridization buffer containing the probes at 37°C for 16 h. Embryos were washed 5 times with the wash buffer for 5 min each, then 2 times in 5x SSCT (5x SSC +0.1% Tween 20). Amplifiers were snap-cooled by heating at 95°C for 90 s and cooled to RT for 30 min under dark conditions. Embryos were incubated in an amplification buffer for 30 min at RT before adding the amplifiers and incubating the embryos for 80 min at RT in a humid chamber under dark conditions. Embryos were washed 4 times in 5x SSCT for 5 min each, stained with DAPI (10 μ g/ml) from BioLegend in 5x SSCT for 30 min, then washed two times in 5x SSCT. Embryos were stored in 5x SSC at 4°C until imaging. Images of labeled chimeric blastocysts were acquired with an Orca Flash 4.0 sCMOS at full resolution on a Nikon Eclipse Ti2 microscope equipped with a Yokagawa CSU W1 Spinning Disk Confocal with 50 μ m pinholes. A Nikon 40x long working distance water immersion objective, NA 1.15, was used to acquire all channels with exposure times of DAPI: 20ms, Alexa 488: 200ms, Alexa 546: 250ms, and Alexa 647: 250ms.

Molecular dynamics simulations

System preparation and simulation procedure: Canonical B-form DNA was created for each simulated sequence using Avogadro 1.2.0.¹⁴⁹ The TEAD4 structure was taken from PDB:5GZB¹⁰², and selenomethionine residues were replaced with regular methionine by simply renaming the selenium atom to sulfur. In order to align the protein structure from 5GZB onto the created DNA structures, we aligned the phosphorus atoms from the 4th to 10th residue on chain B of the PDB (which correspond to the bases CATTCT) to the corresponding atoms on the created DNA. Since we simulated TEAD4 dimers, we performed this alignment twice, once for each binding site. This alignment was accomplished using VMD 1.9.3.¹³⁵ We combined the two translated copies of the TEAD4 protein and the synthetic DNA sequence into one system using AmberTools20.¹³⁶ We used the ff19SB force field for protein atoms,¹⁵⁰ the bsc1 force field for DNA,¹⁵¹ and the OPC for water and ions.¹⁵² (For a control simulation using CHARMM27, the same procedure was used to create the initial structure, but the system was built in VMD and TIP3P was used as the water model.) Systems were solvated in truncated octahedra of water with a 12 Å padding between the solute and cell edge. Systems were charge-neutralized with K⁺ ions, and additional K⁺ and Cl⁻ ions were added to bring the system to a concentration of approximately 150 mM KCl. Systems were minimized using NAMD 2.13.¹²⁵ During minimization, a cutoff distance of 9 Å was used (12 Å for CHARMM27), and solvent bonds were held rigid, though all solute bonds were unrestrained. A timestep of 2 fs was used, and PME electrostatics was applied with a grid spacing of 1 Å. Ten thousand steps of minimization were performed. For thermalization and production, we used a GPU-enabled build of NAMD 2.14.¹²⁵ The same parameters were used as in the minimization, except for the introduction of a Langevin piston to maintain the system pressure at 1 atm and a harmonic collective variable restraint¹²⁶ to prevent the ends of the DNA from fraying during the simulation. This restraint was applied between H1 from the terminal guanine and N3 of the terminal cytosine. (The DNA ends with a GC pair on each end, and both ends of the DNA were restrained in the same way.) A force constant of 1 kcal/mol/Å² was applied to maintain this distance at 2 Å. During thermalization, all velocities were started from zero and gradually warmed by applying a Langevin thermostat to raise the system temperature to 310 K during a ten ps simulation. The thermalized systems were equilibrated for ten ns, the only difference in configuration from the thermalization simulation being the timestep (increased from 1 fs to 2 fs) and the use of rigid bonds (all bonds, including hydrogen, was made rigid during equilibration and production. Coordinates were saved for every ps for both equilibration and production runs. Simulation stability was verified by plotting protein and DNA RMSD values; all simulations were stable. To verify that the simulations had reached equilibrium, we measured BSA values in the first and last fifth of the trajectory; they showed similar distributions in each case except for the deletion, where the right-hand TEAD4 detached from the DNA toward the end of the simulation. We have provided dehydrated trajectories along with all analysis scripts in Python (Python.org), D (dlang.org), and VMD¹³⁵ in one folder. Complete, hydrated trajectories, totaling approximately 5 TB of data, are available upon reasonable request. All MD simulation trajectories can be accessed at <https://collaboration.stowers.org/d1e17b38-9a3d-4900-a623-ac32836b32fc/>.

ChIP-nexus data processing

ChIP-nexus and PATCh-Cap single-end sequencing reads were pre-processed by trimming off fixed and random barcodes and re-assigning them to FASTQ read names. ChIP-nexus adapter fragments were trimmed from the 3' end of the fragments using cutadapt (v.2.5).¹²⁷ ChIP-nexus and PATCh-Cap reads were aligned using bowtie (v.1.1.2)¹²⁸ and its bowtie to the *Mus Musculus* genome assembly mm10. Aligned ChIP-nexus and PATCh-Cap BAM files were deduplicated based on unique fragment coordinates and barcode assignments. ChIP-nexus coverage was normalized was acquired through reads per million (RPM) normalization, where the ChIP-nexus sample coverage was scaled by the total number of reads divided by 10^6 . ChIP-nexus peaks were mapped using MACS2(v.2.2.6)¹³⁰ with parameters designed to restimulate the full fragment length coverage instead of the single stop base coverage (--keep-dup = all -f = BAM --shift = -75 --extsize = 150). ChIP-nexus peaks were filtered for reproducibility in a pairwise fashion using the Irreproducible Discovery Rate framework (IDR) (v.2.0.3).¹⁵³ The IDR framework selected the peaks used for downstream analysis from the largest pairwise comparison.

ChIP-seq data processing

ChIP-seq single-end sequencing reads were aligned to the *Mus Musculus* genome assembly mm10 using bowtie2 (v.2.4.2).¹²⁸ Aligned ChIP-seq BAM files were deduplicated based on unique fragment coordinates and fragments extended based on the average experiment fragment length as determined with an Agilent 2100 Bioanalyzer. Normalized ChIP-seq coverage was acquired using the deepTools subfeature bamCompare (v.3.1.3)¹⁵⁴ using parameters to generate RPKM or log₂ fold-change scaling (--scaleFactorsMethod = None --normalizeUsingRPKM --binSize = 50) or (--scaleFactorsMethod = readCount --operation = log2 --binSize = 50). ChIP-seq peaks were mapped using MACS2 (v.2.2.6)¹³⁰ with default parameters and an applied background coverage using the associated WCE ChIP-seq control experiment. ChIP-seq peaks were filtered for pairwise reproducibility using the Irreproducible Discovery Rate framework (IDR) (v.2.0.3).¹⁵³

TT-seq data processing

TT-seq 75 bp paired-end sequencing reads were aligned using STAR(v.2.7.3)¹²⁹ to the *Mus Musculus* genome assembly mm10 with the following parameters: outFilterMismatchNmax 2, outFilterMultimapScoreRange 0. SAMtools (v.1.14)¹⁵⁵ were then used to keep alignments with mapping quality greater than 255 (-q 255), and only proper pairs (-f 2) were selected. Strand-specific BAM files for each replicate and combined were generated using the following parameters (samtools view -b -f 128 -F 16; -b -f 80; -b -f 144; -b -f 64 -F 16) and (samtools merge plus_128.bam with plus_80.bam and minus_144.bam with minus_64.bam). Normalized TT-seq coverage was generated using bamCoverage (v.3.1.3)¹⁵⁶ parameter Reads Per Kilobase per Million mapped reads (RPKM).

ATAC-seq data processing

ATAC-seq paired-end sequencing reads were aligned using bowtie2 (v.2.4.2)¹²⁸ to the *Mus Musculus* genome assembly mm10. Normalized ATAC-seq coverage was acquired through RPKM normalization along with following parameters: -bs = 50 --minFragmentLength 10 --maxFragmentLength 1000 --ignoreDuplicate --extendReads

RNA-seq data processing

RNA-seq 100 bp single-end sequencing reads were aligned to the *Mus Musculus* genome assembly mm10 using STAR (v.2.7.3)¹²⁹ with the following parameters: outSAMtype BAM SortedByCoordinate, outSAMprimaryFlag OneBestScore, outFilterMultimapNmax 20, outFilterMismatchNoverLmax 0.1, outFilterType BySJout, alignSJoverhangMin 8, alignSJDBoverhangMin 1, outFilterMismatchNmax 999, alignIntronMin 20, alignIntronMax 1000000, alignMatesGapMax 1000000, limitBAMsortRAM 10000000000, outSAMAttributes NH HI MD AS nM, quantMode TranscriptomeSAM GeneCounts. Rsem-calculate-expression (v.1.3.0)¹⁴⁰ was used to generate an expression table with the following parameters: no-bam-output, estimate-rspd, strandedness reverse.

HCR-FISH image analysis

Images were analyzed in Python 3.9. 3D masks were created with the DAPI label using the Cellpose deep learning package.¹⁵⁷ After a small Gaussian blur of width $1 \times 2 \times 2$ pixels, Cellpose segmentation was performed with the cyto model with a diameter of 60 and minimum cell size of 10000. The trophoblast cells were segmented well with this method in 3D, but the crowded inner cell mass cells were frequently corrected by hand using Napari (<https://doi.org/10.5281/zenodo.3555620>). The masked DAPI was expanded by 4 pixels in the xy direction to encompass more of the cytoplasm HCR label in each cell. The small HCR puncta were found by first performing a Gaussian blur of $2 \times 5 \times 5$ width, then a Laplace filter using Gaussian derivatives with sigma = 0.1, 0.5, 0.5. Finally, the local maximum peaks in intensity are found using the Scikit-image peak_local_max function with a threshold of 11 for the CDX2 channel and 15 for both Ezrin and tdTomato channels. The number of HCR puncta found inside each masked cell was recorded. A threshold was determined to categorize a cell as CDX2 positive (greater than 5 HCR puncta) or tdTomato positive (greater than 15 HCR puncta). The threshold for tdTomato is greater because the HCR hairpin signal is sometimes found on the outside surface of the blastocyst, forming brighter and larger puncta compared to the interior cell signal, which would cause too many cells to be categorized as tdTomato positive. All thresholds are held constant between all images of blastocysts.

Molecular dynamics analysis and visualization

Solvent-accessible surface areas were calculated using VMD,¹³⁵ with a 1.5 Å radius around all atoms. The buried surface area between the two systems was calculated by subtracting the surface area of the combined system from the sum of the surface areas of each component system. Plots were generated using Matplotlib,¹⁵⁸ Scipy,¹⁵⁹ and NumPy¹⁶⁰ with Python 3.10 (python.org). The same protocol was used to calculate the buried surface area values of crystal structures and the enhanceosome model. Figures were generated with Tachyon¹²⁵ in VMD; trajectory frames were aligned using a frame-by-frame aligner developed previously.¹⁶¹ Secondary structures were determined using STRIDE.¹³⁸

BPNet model training

BPNet (v.0.0.23) architecture and software were applied as previously described.¹⁵ Model inputs were 1000 bp genomic sequences centered on the ChIP-nexus peaks of TF of interest. Model outputs were the predicted counts (total reads across each region) and predicted profile (coverage signal across each region) for TFAP2C, TEAD4, CDX2, YAP1, and GATA3 ChIP-nexus experiments. ~150K IDR-reproducible peaks from TFAP2C, TEAD4, CDX2, YAP1, and GATA3 ChIP-nexus experiments were pooled and used as model inputs. Validation datasets were peaks across chr5,6,7,19; test datasets were peaks across chr1,8,9, and peaks across chrX and Y chromosomes were excluded from the analysis. The remaining regions were used for model training. Hyper-parameters were the default BPNet architecture. The trained model performance was assessed by comparing (1) area under the PrecisionRecall Curves (auPRC) for profiles over different bins of resolution between observed ChIP-nexus profiles and predicted BPNet profiles (Figure S1C) (2) Jensen-shannon distance for TF binding profile between observed ChIP-nexus signals to predicted BPNet signals for each TF and (3) counts correlations of observed ChIP-nexus signals to predicted BPNet signals for each TF (Figure S1E) as previously described.¹⁵ The auPRC values were benchmarked alongside replicate-replicate, observed random, and observed-average observed profile comparisons to establish an in-context understanding of predicted profile accuracy. All BPNet models were implemented and trained using Keras (v2.2.4), TensorFlow1 backend (v.1.70), and the Adam optimizer¹⁵⁰. The training used an NVIDIA TITAN RTX GPU with CUDA v9.0 and cuDNN v7.0.5 drivers. To obtain the *Tead double* motifs in ESCs for analysis Figure 5G, TEAD4 ChIP-nexus experiments were pooled and used as model inputs to train a single TF model; ~15K IDR-reproducible peaks were used. Validation peak datasets across chr 1,7,3,14, test peak datasets across chr2,8,9, and peaks across chromosomes X and Y were excluded from the analysis. Hyper-parameters, model performances, and BPNet implementation were performed as described above. PAtCh-Cap control in ESCs was from.¹⁵ We performed DeepLIFT and TF-MoDISco on the trained model to generate an ESCs-specific *Tead motif* set. For analysis in Figure 5G, we used *Tead double* motifs from fold 1. Additional models were trained with the same architecture as part of 3-fold validation (fold 2 and fold 3). Spearman counts correlation values (top right) were determined by comparing the observed ChIP-nexus counts with BPNet's predicted counts at TEAD4 ChIP-nexus peaks in ESCs (Figure S5I).

Motif extraction, curation, and island generation

DeepLIFT (v0.6.9.0, derived from the Kundaje Lab fork of DeepExplain (<https://github.com/kundajelab/DeepExplain>)⁶³ was applied to the trained BPNet model to generate the contribution of each base across a given input sequence to the predicted output counts and profile signals. Contribution scores for counts and profile outputs were generated for all 5 TF tasks. TF-MoDISco (v0.4.2.2)⁶⁴ was then applied for each TF separately. Regions of high counts contribution were identified, clustered based on within-group contribution and sequence similarity, and then consolidated into motifs. The *Tfap2c*, *Tead*, *Cdx2*, *Gata3*, and *Yap1* motifs were manually identified based on their similarity to the known motif and the sharp average ChIP-nexus binding footprint of the corresponding TF. Once motifs were characterized and confirmed, they were used to label genomic instances by CWM scanning as previously described.¹⁵ Briefly, a motif was mapped based on both Jaccardian similarity to the TF-MoDISco contribution weight matrix (CWM) and sufficient total absolute contribution across the mapped motif. Then, motifs were filtered for redundant assignment of palindromic sequences and overlapping peaks. To obtain regions of mapped motif combinations with enhancers for downstream measurement of enhancer activity to get specific mapped motif pairs, motif islands were generated as described.²⁰ Each island starts as a 500 bp (enhancer window) region centered on the motif and gets clustered and merged with another nearby motif island if they overlap. In this manner, islands get extended if a motif is within less than 500 bp. The motif islands, by their motif combinations with motif numbers, read sums of TFs binding and enhancer activity (provided in Table S4).

Visualization of YAP1 binding and enhancer activity markers

To visualize the correlation between YAP1 binding and the markers of enhancer activity, we selected regions using the following criteria: BPNet-mapped motifs that were absent of ERVs, were within TEAD4 peaks, and showed TEAD4 binding. At those regions, we calculated the total ChIP-nexus read counts for YAP1, selected regions above the median value, and sorted based on the total read counts. These regions were then divided into the top 5000 regions with high YAP1 reads and the 5000 regions with median YAP1 binding. We used this set to calculate normalized reads and generate the TEAD4, YAP1, H3K27ac, Pol II, and Nascent-RNA heatmap.

Motif pair interaction analysis

We selected mapped regions with only one motif pair from the motif islands set for the following motif-pair combinations: *Tfap2c*-*Tead*, *Cdx2*-*Tead*, *Gata3*-*Tead4*, and *Tfap2c*-*Tead double*. We then sorted the regions by the distance between the two motifs

and included distances less than 160 bp for display. YAP1 contribution scores from the binding model were used to make heatmaps in ggplot. The *in silico* motif interaction analysis and odds ratio calculations for the co-occurrence likelihood of motif pairs were performed as described.¹⁵ Statistics for [Figures 3C, S3C, and S3D](#) provided in [Table S5](#) and described in [STAR Methods](#).

Yap1 binding enhancement was calculated by measuring BPNNet model predictions across computationally generated sequences where we injected two motifs—a *Tead single* motif and another motif (y axis on 1G) into 256 randomized sequences. We next averaged predictions across these 256 trials. More precisely, this *in silico* motif interaction analysis was performed to measure “YAP1 binding enhancement” as described previously.¹⁵ When injecting two motif sequences (motif A and motif B) across motif pair distances (*d*) ranging up to 150 bp. Yap1 binding predictions were measured in these different simulation scenarios where the sum of the read counts predicted across a 50 bp window centered on motif A. We measured four different cases: (1) neither motif A nor motif B was injected into the sequence (*h*0), (2) motif A only was injected into the sequence (*h*A), (3) motif B only was injected into the sequence (*h*B), and (4) motif A and motif B were both injected into the sequence at a designated distance (*h*AB). After each case was measured across all motif combinations and distances, then averaged across trials, the *in silico* binding enhancement for each motif in a motif pair was calculated using the following equation: $\log_2((hAB - (hB - h0))/hA)$. The motif pairs considered were combinations of the highest affinity sequence representations of *Tead single* (ACATTCCTG), *Tfap2c* (CCCTCAGGC), *Cdx2* (GCCATAAA), *Gata3* (AGATAAG), *Jun-Fos* (ATGAGTCAT), *CTCF* (CCACTAGGGGGCG), *Elf5* (CCGGAAG), *Gata3 double* (AGATAAGATCT) and *Tead double* (ACATTCCTGGCATTC).

Enhancer regions selection for reporter assay

This analysis was to predict TFAP2C and TEAD4 binding on genomic regions with different motif distances and how this changes upon editing the distance between the motifs. From our islands, we selected regions with one *Tead single* and one *Tfap2c* motif within less than 200 bp and resized them to 400 bp putative enhancers, and recorded the coordinates of *Tfap2c* and *Tead4* motifs within the putative enhancers for mutations. We then identified the nearest genes using the biomaRt package. For each putative enhancer, we generated sequences for wild-type, mutated motif for each at its original position by mutating the two most contributing nucleotides to the least contributing within that motif. Then, we inserted the same motif at distances in multiples of 10 or 15 within a 400 bp window. These sequences were combined into an array to predict TF binding and contributions at a motif range of 50 bp. The resulting unique enhancer values were exported in R for plotting. The luciferase assay and CRISPR regions were selected by high binding of TEAD4, TFAP2C, and YAP1 at these putative enhancers and by context-relevant gene targets.

Extracting regions with different *Tead double* motif spacings

To map regions in the mm10 genome with different *Tead double* motif spacings, we used pattern matching (with no mismatches) to identify instances of a single *Tead* motif (RMATTCWD). Then, regions with two motifs within 23 bp were identified, and the frequencies by which each motif spacing occurred were recorded. Thus, for a motif spacing of 2, the matched sequence is RMATTCNNRMATTCNN. The predicted TEAD4 binding signal was then calculated for all motifs injected into randomized sequences and averaged over 256 iterations. The results from each spacing were then averaged.

Tead motifs variant analysis

To assess the distribution of motif variant frequency, identical sequence patterns of CWM-mapped *Tead single* and *double* motif patterns were grouped, analyzed and visualized. To obtain a robust representation, only patterns that occurred in the top 90th percentile and occurred at least 10 times were considered. After injecting each sequence pattern into 256 random sequences, BPNNet was used to predict TEAD4 binding. The average predicted signal for each pattern, along with the pattern frequency, was plotted using ggplot.

Genome-wide TEAD4 binding cooperativity on *Tead double* motifs

This analysis aimed to investigate the potential synergy between each side (each *Tead* motif) of the *Tead double* motifs. We selected regions that did not overlap with either ERVs or promoter regions, extracted the sequences of the *Tead double* motifs, and oriented them in the 5'>3' direction. We then split the motif sequences into two half-sites, each corresponding to a *Tead single* motif. We then predicted the binding of TEAD4 at the half-sites and the complete double motifs injected into random sequences. The values for the two half-sites were summed and compared to those for the complete double motifs as a measure of synergy between the two half-sites of TEAD4 double motifs.

Benchmarking motif discovery and analysis

This analysis aimed to benchmark the motif discovery and scanning approaches leveraging PWMs (frequency-based) versus CWMs (contribution-based) for analyzing *Tead* motifs. We performed traditional PWM scanning, leveraging MEME (5.3 and 5.5.3) and FIMO (5.5.3)^{137,162} We ran MEME (*-mod anr*) on the top 1,000 TEAD4 ChIP-nexus peaks within the central 101 bp regions of each peak summit, returning both a *Tead single* and *Tead double* motif. Using FIMO (*--skip-matched-sequence --parse-genomic-coord --max-strand --max-stored-scores 10000000*), we performed two rounds of motif scanning. The first round was conducted on *Tead single* and *Tead double* motif PWMs returned by MEME, described above (called “PWM-freq”). The second round was conducted on *Tead single* and *Tead double* motif PWMs reconstituted from seqlets returned by TF-MoDISco (called “CWM-freq”).

We compared each set of FIMO-mapped motifs, alongside the motifs mapped using CWM-scanning, described above (called “CWM-contrib”) (Figure 4C). This resulted in three groups of mapped motifs.

For the three groups of motif mappings (PWM-freq, CWM-freq, and CWM-contrib), we scored the sequence match of each motif from each mapping group using the CWM-freq PWM as a reference. We next performed in silico analysis of each mapping group, taking their respective top 5K *Tead* double motifs and top 20K *Tead* single motifs based on the sequence-match score described above. We injected each unique motif sequence from these top-scoring motifs into randomized DNA sequences and predicted TEAD4 binding. We evaluated genomic context influencing TEAD4 binding predictions by predicting genomic regions that mapped CWM-contrib motifs labeled as “CWM-contrib in genomic context.”

We performed additional benchmarking of CWM-versus PWM-scanning by leveraging an alternative motif mapping software called HOMER (4.9.1)¹⁶³ to identify motifs. Run under default parameters of fragment size of 200 bp and a variety of allowed motif lengths (i.e., 9, 13, 18, 20 bp) from the top 1000 TEAD4 ChIP-nexus peaks, HOMER returned a matched *Tead single* motif, not the double motif. However, when we looked at the list of *de novo* motifs similar to the *Tead single* motif, we could identify the *Tead double* motif by specifically searching for the pattern (Figure S4A). Given the long list, it is unlikely that the motif would stand out to someone not looking for it.