

## MOLECULAR BIOLOGY

# Quantitative modulation of a spatial enhancer through the biophysical properties of a transcription factor binding site

Conrad Fallon<sup>1,2</sup>, Liucong Ling<sup>1,2</sup>, Brendon H. Cooper<sup>3</sup>, Stefano Ceolin<sup>2</sup>, Tsu-Pei Chiu<sup>3</sup>, Raktim Mitra<sup>3</sup>, Bettina Mühling<sup>2</sup>, Vani Srinivasan<sup>2</sup>, Ayse Damla Durmaz<sup>2</sup>, Daniel Veselinovic<sup>2</sup>, Mahek Kothari<sup>2</sup>, Athanasios Panagiotis Fylaktakis<sup>4</sup>, William J. Glassford<sup>5,6,7</sup>, Judith F. Kribelbauer-Swietek<sup>5,6,7</sup>, Dimitrios K. Papadopoulos<sup>4\*</sup>, Richard S. Mann<sup>5,6,7\*</sup>, Remo Rohs<sup>3,8\*</sup>, Nicolas Gompel<sup>1,2\*</sup>

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Fine-tuning of gene expression by transcription factors (TFs) is essential for normal development, but how TF binding is converted into precise expression levels is poorly understood. We used a *Drosophila* enhancer responding to the TF Distal-less (Dll) to investigate this relationship. By combining computational structure analysis and quantifying relative TFBS affinity and enhancer activity, we describe how TF concentration is converted into precise expression across the developing *Drosophila* wing. We show that, although the regulatory function of the enhancer generally follows the classical Hill equation, several additional terms are needed to describe the effects of affinity, orientation, sequence, and wing region on the maximum achievable expression and position of transcriptional activity in a tissue. We also found that Dll relies on two distinct sets of amino acid–DNA contacts for binding to higher affinity sites, but contacts are more evenly distributed at lower affinities. Our work integrates information at molecular and tissue scales to describe how the modulation of a single TFBS determines the regulatory function of an enhancer in a complex in vivo environment.

## INTRODUCTION

Cells at different positions in a developing embryo acquire their identity by expressing selected genes at precise times and levels. This controlled expression results from enhancers that contain specific transcription factor (TF) binding sites (TFBSs), which integrate input information in the form of TF concentration. The nature of the TF-TFBS interaction (binding affinity and orientation) and the combination of specific TFBSs in an enhancer (1) determine the enhancer's output quantitatively and spatially. The resulting precise expression of their target genes influences in turn the acquisition of cell identity. Deviations from normal patterning and expression levels can result in phenotypic variation within a species, diversity among species, or abnormal development and disease. Simplistically, this can be viewed as a process of TF molecules binding to their specific target sites and TF-TFBS interaction being converted into precise gene expression levels. Understanding these processes is key to understanding how specific expression is fine-tuned during

development. Here, we describe both processes using quantitative spatial expression data and computational structure analysis.

The configuration and properties of TFBSs are key determinants of enhancer function, in that their order (2, 3), orientation (4–6), spacing (7, 8), and binding affinity (9–11) all determine how TF-TFBS interactions influence gene expression (12–18). Of these, affinity changes are of key interest as they are more common—they can be caused by single point mutations—and result in notable phenotypic changes such as heart development defects or polydactyly (11, 19–21). How affinity changes affect the expression of a gene in a tissue is described by its gene-regulatory function (GRF), the relationship between TF concentration and gene expression at steady state (22, 23). Gene expression typically increases in a nonlinear manner and can be modeled using the Hill equation, which describes sigmoidal, “S”-shaped responses (23, 24) and has been used for many biological phenomena including the effect of TF and cofactor concentration and TFBS affinity and spacing, on GRFs (22, 23, 25, 26). The Hill equation therefore offers a framework to study how TF-TFBS binding is converted to gene expression and represents an essential step to predicting an enhancer's activity based on its sequence and regulatory context. In the context of gene expression, the Hill equation can be written as

$$E = \frac{E_{\max}^S}{1 + \left(\frac{K_a}{[T]}\right)^{n_H}} + c \quad (1)$$

where  $E$  is expression;  $E_{\max}^S$  is the maximum expression achievable by a given enhancer sequence when it is saturated with bound TFs (and defines the height of the curve);  $[T]$  is the TF concentration at equilibrium;  $K_a$  is the in vivo affinity or association constant of all binding sites (which are normally assumed to be equal in the model), determines the position where the enhancer is most sensitive to changes in TF concentration, and defines the inflection point of the curve (25);

<sup>1</sup>University of Bonn, Bonn Institute for Organismic Biology, Bonn, Germany. <sup>2</sup>Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Planegg-Martinsried, Germany. <sup>3</sup>Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA. <sup>4</sup>Department of Biology, University of Crete, Voutes University Campus, Heraklion, Crete 70013, Greece. <sup>5</sup>Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA. <sup>6</sup>Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA. <sup>7</sup>Department of Neuroscience, Columbia University, New York, NY 10027, USA. <sup>8</sup>Department of Chemistry, Department of Physics & Astronomy, Thomas Lord Department of Computer Science, Division for Medical Oncology in the Department of Medicine, and Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA. \*Corresponding author. Email: dpapadopoulos@uoc.gr (D.K.P.); rsm10@columbia.edu (R.S.M.); rohs@usc.edu (R.R.); ngompel@uni-bonn.de (N.G.)

$n_H$  is the Hill coefficient, which classically reflects the number of TF molecules binding cooperatively to an enhancer, and describes the steepness of the response to increases in TF concentration (and defines whether the curve resembles a gentle hill or a steep cliff) (25, 27); and  $c$  is the basal promoter function when the enhancer is inactive (24, 28). The Hill equation describes the equilibrium average occupancy of a series of TFBSs with cooperative interactions in an all-or-nothing scenario, in which the cooperative interaction is much stronger than the sum of individual interactions with each binding site. Using the equation to fit to expression data assumes that TF occupancy follows this simplified relationship and that expression is linearly proportional to site occupancy (28–32). However, aside from rarely being tested in a developing tissue, the equation limits the GRF of an enhancer to a measure of site occupancy, leaving other features of TF-TFBS interactions such as binding orientation unaccounted for, and a connection to the mechanism of TF-TFBS interaction is rarely made.

How a TF recognizes a cognate TFBS has only recently started to be understood. TFBS recognition is more complex in tissue environments, as closely related TFs may compete for similar sites; high-affinity TFBSs can often be bound by multiple members of a TF family (33). This confounds predictions of expression patterns in tissues, as ectopic expression can be caused by binding of multiple TFs, including in regions outside of where the TF attributed to the TFBS is found. Producing a precise expression pattern, therefore, requires TFs to bind to their TFBSs using specific mechanisms. Understanding how TFs acquire specificity is essential to predicting expression patterns. Only once that both the mechanism of a TF binding to its target sites and the relationship between TF-TFBS binding and gene expression are known will a clear understanding of how expression patterns emerge be possible.

To first explore the relationship between TF specificity, orientation, TFBS affinity and enhancer activity, we modulated the sequence of a known binding site for the homeodomain TF Distal-less (Dll) in a well-characterized enhancer of the *yellow* gene in *Drosophila*, *spot*<sup>196</sup>, and measured the spatial effect on expression levels alongside structural predictions of how Dll binding specificity was achieved. *spot*<sup>196</sup> is active in pupal wings (34–36). We measured the spatial activity of *spot*<sup>196</sup> variants differing in the affinity of a single Dll binding site as predicted using systematic evolution of ligands by exponential enrichment combined with next-generation sequencing (SELEX-seq) (37). In parallel, we estimated the endogenous Dll concentration using fluorescence correlation spectroscopy (FCS) and compared the shape of the GRF for each Dll site variant to investigate how the relationship between absolute TF concentration and enhancer activity corresponded to different TFBS affinities, orientations, and sequences. Last, we used a deep learning-based framework to generate predictions of the binding mechanism of Dll on different site variants, revealing how specific binding to cognate sites is initially achieved by Dll. Together, our study provides a complete view of the enhancer response to variation in the sequence of a single TFBS in a complex tissue, from input TF concentration to the spatial and quantitative distribution of enhancer activity and down to the biophysical properties of the TF-DNA interactions at the atomic level.

## RESULTS

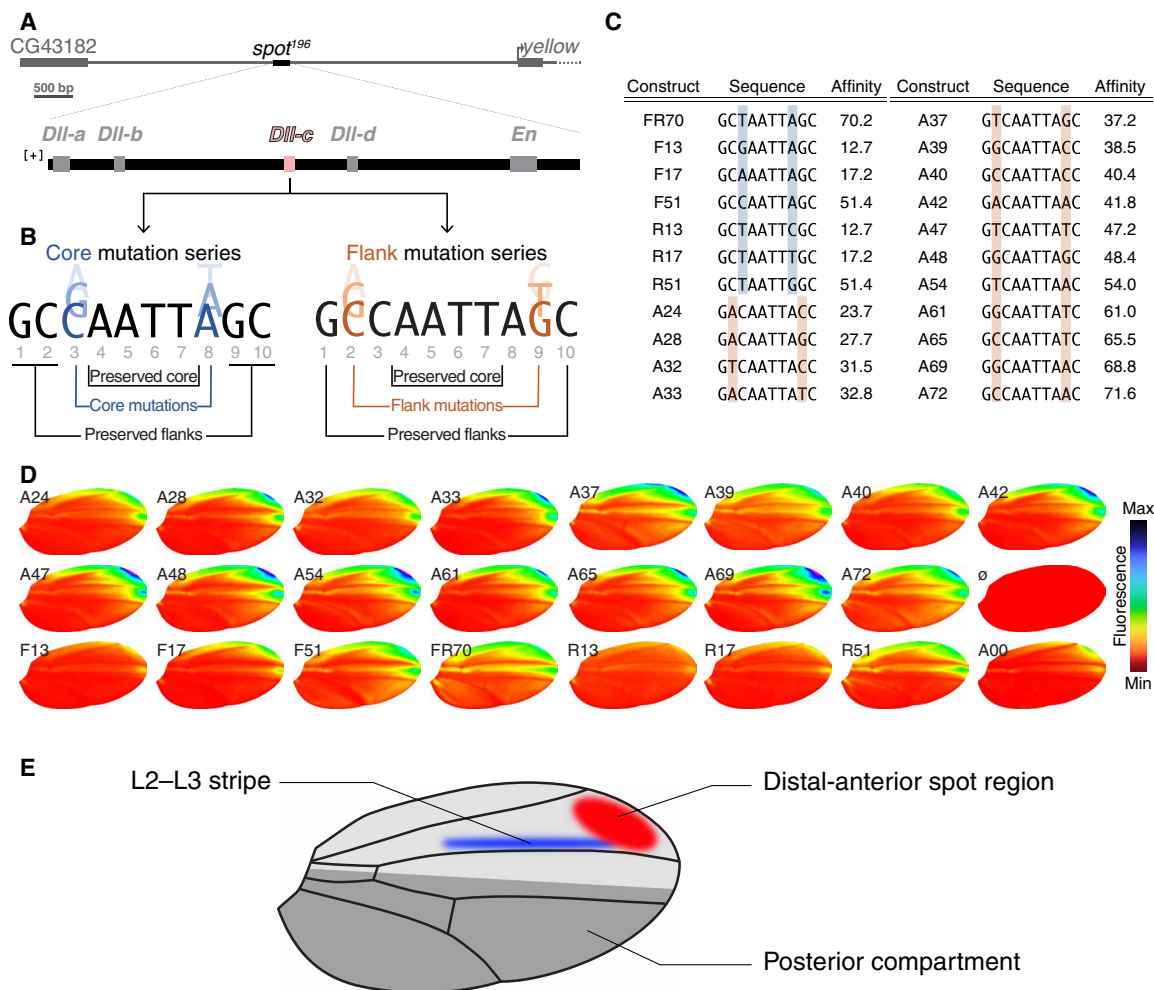
### Evaluating Dll relative binding affinity for different sequences with SELEX-seq

To design *spot*<sup>196</sup> variants with Dll sites of precisely known relative binding affinity, we characterized the in vitro monomer Dll binding

preferences using SELEX-seq. We performed SELEX-seq with full-length Dll and a SELEX-seq library containing a randomized 16-base pair (bp) region (37) (see Materials and Methods). Using the sequence data for the 16-mers, we performed motif-free alignment of 8-mer sequences using the Top-Down Crawl method (38). This distinguishes the 8-bp sequence in the 16-mers, for which Dll shows a real preference from the nonspecific sequences that appear within the same 16-mers. From this, every 8-bp motif was assigned a normalized score of 0 to 100%, which represented its in vitro (SELEX-seq-measured) relative binding affinity ( $A_{ff}$ ) (table S1). This score is related to, but distinct from,  $K_a$ , as it is fixed for each sequence, whereas  $K_a$  values will differ for each sequence depending on their chemical, in vivo context.

### Design of enhancer variants with a single modified TFBS

We designed variants of the *spot*<sup>196</sup> enhancer that differed in their predicted (SELEX-seq) affinity at a single TFBS. The enhancer contained at least four Dll binding sites and at least one for the repressor Engrailed (En) (Fig. 1A) (34, 35). One of the Dll binding sites, *Dll-c*, was previously shown to contribute moderately to enhancer activity (39) and was, therefore, chosen as the subject of this study, so that expression levels could be more finely adjusted across a wider range of values. *Dll-c* lies a few nucleotides away from a site governing accessibility but not overlapping with it. Therefore, the modulation of *Dll-c* is not expected to affect chromatin accessibility and was initially assumed to be equal in all our constructs. We then designed variants of *spot*<sup>196</sup> with alternate *Dll-c* sites using 8-mers selected from the SELEX-seq results, bearing additional random nucleotides at both ends of the site that did not significantly change the relative affinity of the sequences. Initially, we replaced the *Dll-c* site with 10-mers with relative affinities evenly spanning a range of 14 to 87% (fig. S1A). However, the overall enhancer activity was poorly correlated with *Dll-c* affinity—several constructs demonstrated artifacts that were likely caused by non-Dll TF binding (fig. S1, B and C), and in some cases, we detected alternate Dll binding sites on either side of *Dll-c* (fig. S1G). We, therefore, redesigned the enhancers to ensure Dll bound at the core *Dll-c* site, insulating it with flanking GC dinucleotides at each end (Fig. 1, B and C). In addition, we prevented the occurrence of non-Dll TF binding by minimizing the total number of base pair changes. The variants were modified from a binding site of the first series, *F51*, which produced relatively high expression. From this, we derived a series of mutants containing a single substitution in the 6-bp core of *Dll-c* at position 3 or 8 (mutating both nucleotides always produced extremely low-affinity sites) or up to two substitutions in the flanking regions at positions 2 and/or 9 (Fig. 1B). This resulted in a set of 22 site variants with affinities ranging evenly from 13 to 72% (Fig. 1C). Among the seven core mutant variants, three pairs were reverse complements of each other with the same relative affinity (*F13*, *R13*; *F17*, *R17*; and *F51*, *R51*), and one was a palindromic sequence (*FR70*). We refer to those that retain the original orientation of the wild-type TFBS core (5'-NAATTA-3') as the forward-orientation variants (*F13*, *F17*, and *F51*) and those with the opposite sequence (5'-TAATTN-3') as the reverse-orientation variants (*R13*, *R17*, and *R51*). All constructs with flanking mutations contained the same core sequence as that of *F51* (5'-CAATTA-3'). We also included a *Dll-c* null construct from a previous study on *spot*<sup>196</sup> (39), in which the 10-bp *Dll-c* site is replaced by a polyadenylate tract, here referred to as *A00*. Together, the *Dll-c* variants spanned an affinity range of 0 to 72%. We then assessed the activity of each variant using quantitative fluorescent



**Fig. 1. Design of enhancer variants with a single modified TFBS.** Reporter-construct design and phenotypes. **(A)** Map of *spot*<sup>196</sup> enhancer in relation to the *yellow* gene and location of binding sites for Dll and En. **(B)** Sequence of edited *Dll-c* binding site showing position of core and flanking substitutions. **(C)** Full set of *Dll-c* variants showing their sequence and relative binding affinity according to SELEX-seq. Construct names indicate orientation of binding site (F and A: forward; R: reverse; FR: palindromic), position of substitutions (F, R, and FR: core; A: flank), and predicted relative binding affinity as a percentage. **(D)** Average wing phenotypes showing the average reporter signal from tens of wings expressing each construct (table S3). Notably, the number in each construct name corresponds to the predicted relative binding affinity. **(E)** Map of wing showing locations of features referred to in the Results.

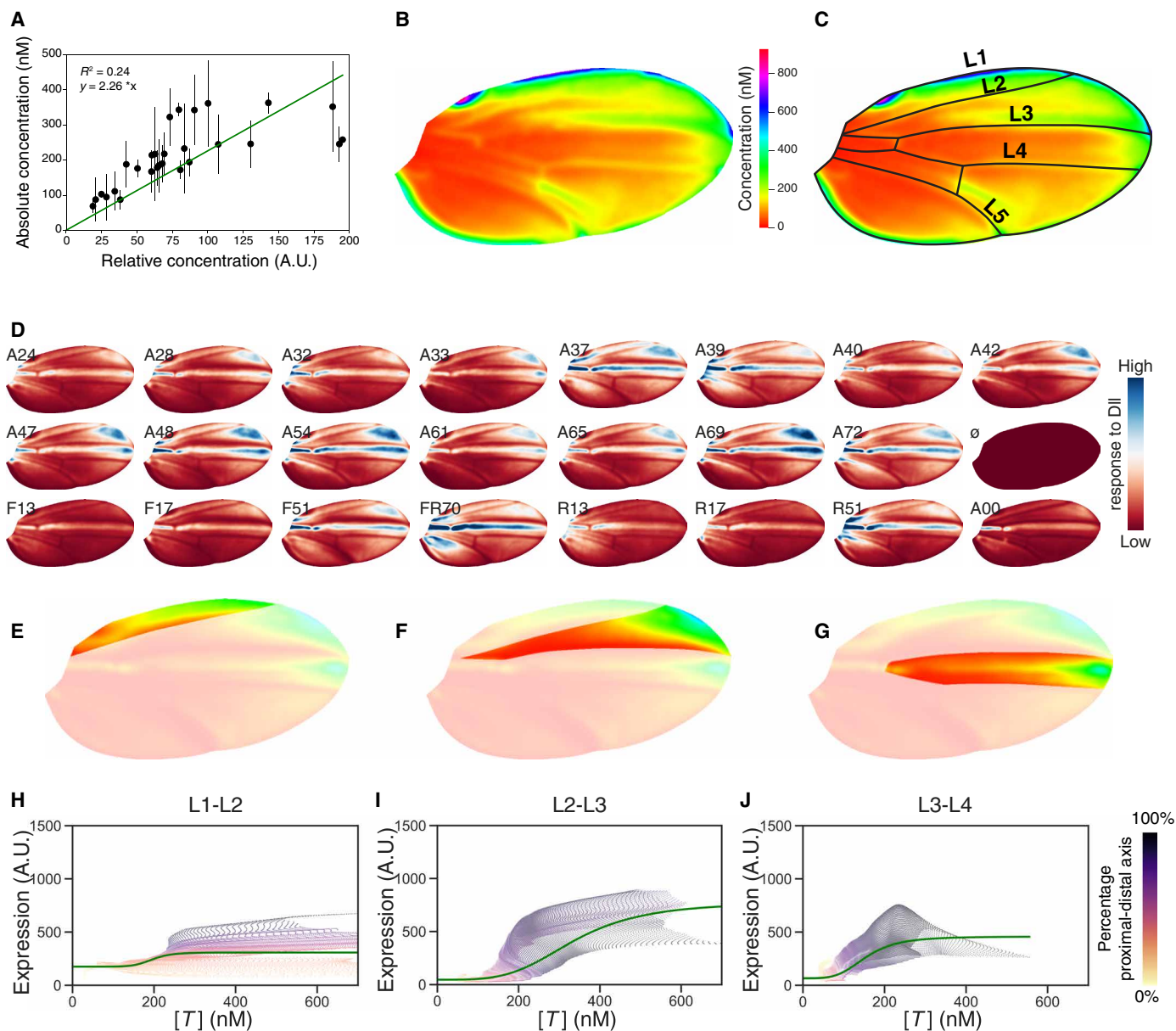
reporter assays in the adult wings (36) of recently eclosed transgenic *Drosophila melanogaster* flies (Fig. 1D and fig. S1B).

### Enhancer activity as a function of TF concentration

The transcriptional effect of TF-TFBS binding is modulated by TFBS features and by the concentration of the input TF. To determine the relative contribution of the latter, we created a map of nanomolar Dll concentration in developing wings as we have done in previous studies for other developmental TFs (40–42). We first produced a map of relative Dll concentrations using an available direct transcriptional reporter for endogenous *Dll* expression (43). Endogenous *Dll* mRNAs produced from this allele also encode a *tdTomato* reporter sequence, and the coding sequences of the *tdTomato* and Dll peptides are separated by a T2A cleavage signal. Assuming that the degradation rates of *tdTomato* and Dll are similar, the concentration of the two proteins will also be similar. We then imaged nine pupal wings at 75%

of development, registered the images onto a reference wing image (36), and derived an average image of the Dll relative concentration. In a second step, we used FCS to measure the absolute concentrations of the *tdTomato* reporter construct in combination with a histone–green fluorescent protein transgene. We acquired FCS measurements from across the pupal wing and interpolated the discrete nanomolar concentrations obtained for *tdTomato* with our map of relative Dll concentrations to generate a complete map of Dll TF concentration in the pupal wing (Fig. 2, A and B, and see Materials and Methods).

Activity maps that investigated enhancer activity relative to Dll concentration were generated, which revealed low responsiveness to Dll concentration in the posterior wing compartment (Fig. 1E) and along the veins (Fig. 2D). This is likely caused by the presence of the repressor En in the posterior half of the wing and of an unknown regulator along the veins (35, 39, 44). To generate GRFs,



**Fig. 2. Enhancer activity as a function of TF concentration.** Map of Dll concentration and reporter signal as proportion of Dll concentration. **(A)** Relative Dll concentration obtained from *tdTomato-Dll* reporter images against FCS values from *tdTomato-Dll* reporter construct in wings (see Materials and Methods for details). **(B)** Map of nanomolar Dll concentration. **(C)** Map of veins in *Drosophila* wings and nomenclature of longitudinal veins used to delineate regions for analysis. **(D)** Average signal of each reporter construct as a proportion of Dll concentration. **(E to G)** Regions considered for generating GRFs in later analysis. **(E)** L1–L2, **(F)** L2–L3, and **(G)** L3–L4. **(H to J)** Plots of Dll concentration ( $[T]$ ) against average reporter signal in each region to which a Hill equation is fitted (green line). The plots show the trend for the representative line F51 and each point on the scatter plot represents the value of  $[T]$  and expression at an individual pixel on the average phenotype map. **(H)** L1–L2, **(I)** L2–L3, and **(J)** L3–L4. The color scale shows relative distance of each point along the proximal-distal axis within each region. A.U., arbitrary unit.

we restricted our analysis to anterior wing regions with a broad range of Dll concentrations, including the L1–L2, L2–L3, and L3–L4 intervein regions (Fig. 2, C and E to G). Reporter expression was then plotted against Dll concentration across each region, and a curve based on the Hill equation was fitted to the data (Fig. 2, H to J). Among these, the expression in each construct appeared to follow the sigmoidal shape expected from the Hill function most closely in the L2–L3 region (Fig. 2I). By contrast, the L1–L2 region provided a poor fit, likely due to the difficulty in cropping the outer

wing veins from the pupal wing images, which resulted in extremely high expression levels that skewed the plots toward high levels of  $[T]$ . Similarly, the L3–L4 region also provided a poor fit for the Hill function, as it did not reach high Dll levels, except in a few cells at the distalmost extremity. Therefore, expression did not reach a plateau for  $E_{\max}^S$  to be reliably measured. In addition, the partial overlap of this region with the *En* expression domain did not allow us to keep the transregulatory environment relatively homogenous. Therefore, we obtained measurements of enhancer response to Dll

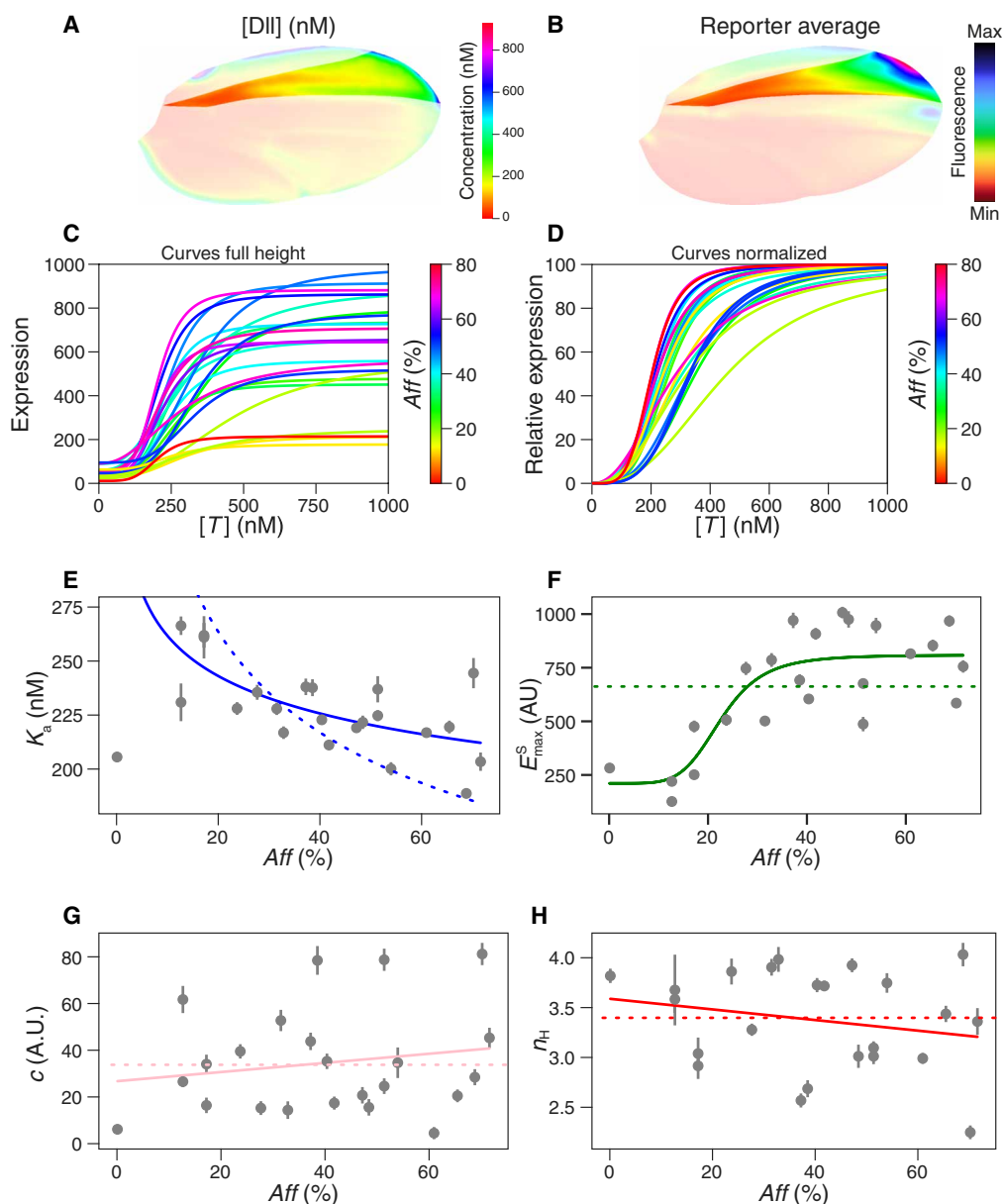
concentration from the L2–L3 intervein region (Fig. 3, A and B). Within the L2–L3 region, the Hill equation provided a good fit for the response of enhancer activity to Dll concentration ( $[T]$ ) for each variant. We, therefore, used these data to determine the GRF for each enhancer variant (Fig. 3, C and D, and fig. S2A).

We then compared the value of each parameter we measured by modulating the values for Dll concentration ( $[T]$ ) and relative affinity ( $Aff$ ) against what would be expected from the assumptions of the Hill equation. As shown in Fig. 3E

$$\text{Expected: } K_a = \left(\frac{1}{Aff}\right)^{1/3.37} * 647, \text{ Fitted: } K_a = \left(\frac{1}{Aff}\right)^{1/9.43} * 334$$

As shown in Fig. 3F

$$\text{Expected: } E_{\max}^S = 662, \text{ Fitted: } E_{\max}^S = \frac{600}{\left(\frac{22.6}{Aff}\right)^{5.27}} + 209$$



**Fig. 3. Enhancer activity as a function of TF concentration.** Relation between Hill curve parameters and binding preference to *Dll-c* variants. (A) Distribution of Dll concentration within the L2–L3 intervein region. (B) Average reporter expression in the L2–L3 region from all variants. (C) Curves generated by fitting Hill regressions to Dll concentration against the reporter signal. (D) Curves generated in C normalized by setting  $E_{\max}^S$  to 1 for all variants. (E to H) Hill parameters for each *Dll-c* variant. Error bars show the SD of parameter values determined from 100 iterations of generating average phenotypes from a random sample ( $n = 30$ ) of wings from each variant. Dotted lines indicate expected relationships between relative affinity, and each parameter and solid lines indicate the best-fitted relationship between the relative affinity and each parameter as described in the Results.

As shown in Fig. 3G

$$\text{Expected: } c = 35, \text{ Fitted: } c = 0.2 * Aff + 27$$

As shown in Fig. 3H

$$\text{Expected: } n_H = 3.37, \text{ Fitted: } n_H = -0.005 * Aff + 3.58$$

One departure we made from the conventional Hill equation was the value of  $K_a$ , which we considered to represent the association constant of all Dll binding sites in the *spot<sup>196</sup>* enhancer. Since we know that the affinity of each Dll binding site is different and that we are only modulating one of them, the term representing the rate of Dll binding to the enhancer was replaced by the product of the association constants of all Dll sites such that

$$K_a = \left( \frac{1}{Aff} \right)^{1/n_H} * p \quad (2)$$

where *Aff* is the in vitro relative affinity of the *Dll-c* site predicted from SELEX-seq data and *p* is a proportionality constant incorporating the product of the association constants of all other Dll binding sites and the scaling constant relating relative and absolute affinity. The inverse relationship between  $K_a$  and *Aff* reflects the fact that as the relative affinity of the *Dll-c* site is increased, a lower concentration of Dll is required to achieve half maximum expression level and, eventually, to saturate the site. Conversely,  $E_{max}^S$ , *c*, and  $n_H$  are expected to be constants with respect to [*T*] and *Aff*. The expected relationships between each parameter and *Aff* are shown as dotted lines in Fig. 3 (E to H). The average for each experimentally determined parameter was used wherever a value was expected to be constant, e.g., the value for  $n_H$  in the relationship between  $K_a$  and *Aff*.

To fit the model from Eq. 2 to the data for  $K_a$  and *Aff*, we fixed the experimentally determined average value of  $n_H$  (3.4) (Fig. 3H, dashed line), used *p* as a free parameter, and excluded the value for  $K_a$  where *Aff* = 0, as this would require an infinite value of  $K_a$ . This did not appear to fit the data well [Normalized Root Mean Square Error (NRMSE) = 0.12] compared with fitting the exponent in Eq. 2 as a free parameter (Fig. 3H; NRMSE = 0.06, solid line). However, fitting the exponent freely would require an  $n_H$  of 9.4, which would exceed the assumptions of the Hill model, where the number of binding sites determines the maximum possible value of  $n_H$  without additional energy expenditure (26). The poor fit to the model could also reflect the large spread in the expression data (Fig. 2I and fig. S2), which is likely caused by differing regulatory regimes in different regions across the L2–L3 region.

The most evident deviation from the predictions of the classical Hill models that can be observed in our data is the variation of  $E_{max}^S$  at saturating concentrations of the input TF. Unfortunately, the biophysical mechanisms of how enhancers and promoters engage to achieve a certain level of expression are not understood in enough detail to postulate a biophysical model relating enhancer occupancy and expression. However, our data (Fig. 3F) suggest that a sigmoidal relation between  $E_{max}^S$  and the binding site's affinity provides a better fit [Fig. 3F; constant coefficient of determination ( $R^2$ ) = 0, NRMSE = 0.39, sigmoid  $R^2$  = 0.69, NRMSE = 0.22]. We, therefore, followed a phenomenological approach and extended the original model of our GRF by assuming that

$$E = \frac{E_{max}}{1 + \left( \frac{Aff_{50}}{Aff} \right)^a} \cdot \frac{1}{1 + \left( \frac{K_a}{[T]} \right)^{n_H}} + c \quad (3)$$

where  $E_{max}$  is the general maximum expression achievable by an enhancer when site affinity and TF concentration are in excess,  $Aff_{50}$  is the affinity level required to achieve half maximum expression level, and *a* is a nondimensional exponent describing steepness. One possible interpretation of this extension is that while expression is normally assumed to relate to the average occupancy of a TF at the enhancer, independently from this being achieved by a higher frequency of binding events (from higher TF concentration) or longer residency time (from higher affinity), one could envision a direct impact of residency time on transcription. The increased residency time of TFs at the site may be more effective in triggering transcription and may increase the rate of mRNA transcription nonlinearly, for example, by producing instances of binding to the site that allow more time for transcriptional complex to become active and remain in an active state for longer. In addition, the change in  $E_{max}^S$  may reflect changes in accessibility caused by increased Dll binding resulting in competition with nucleosomes (45). This may also account for the deviation from the Hill model.

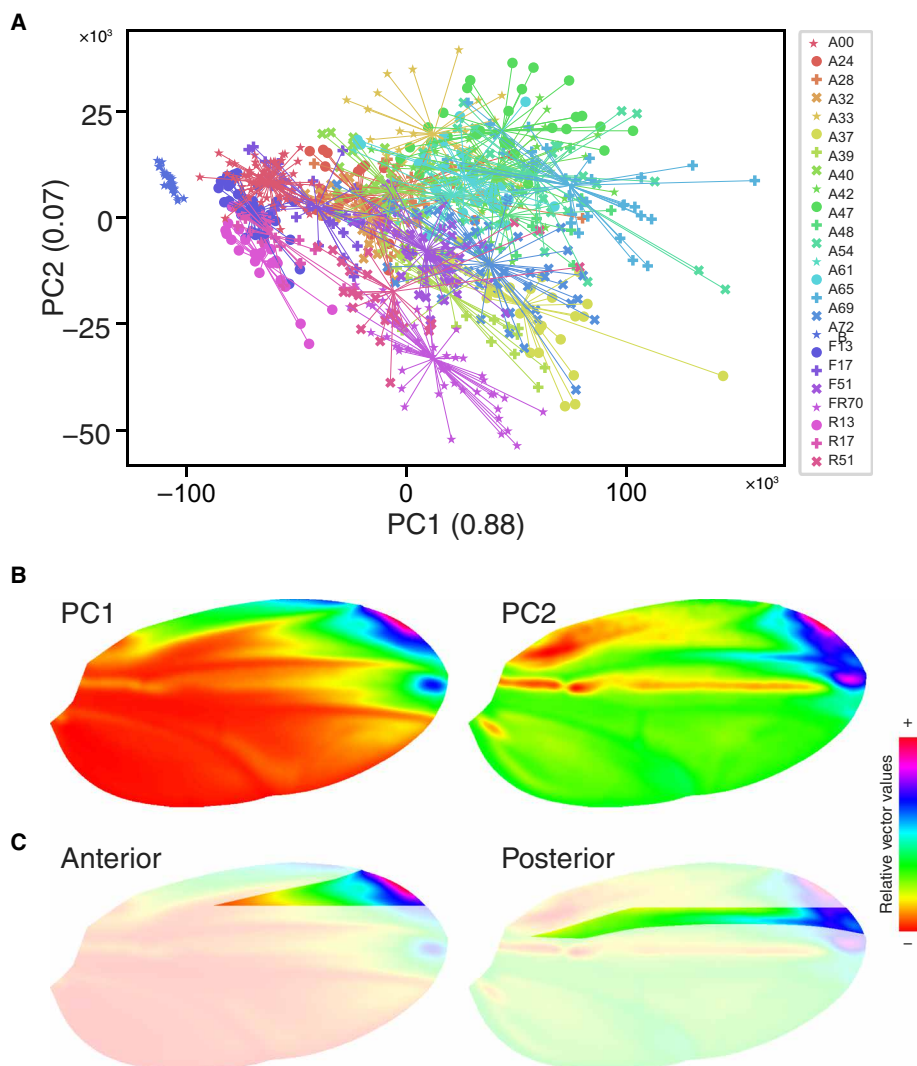
As would be expected from the model assumptions, both *c* and  $n_H$  were unaffected by *Aff*, in that both remained around a fixed value—with some variation—irrespective of *Dll-c* affinity (Fig. 3, G and H; *c*:  $R^2$  = 0.03, NRMSE = 0.63;  $n_H$ :  $R^2$  = 0.05, NRMSE = 0.14) in line with the model assumptions. Baseline expression should occur when Dll concentration is close to 0 nM. Because [*T*] is close to 0 nM, the first term on the right side of Eq. 1 tends to 0, leaving only the constant value *c* unaffected by *Dll-c* affinity. Similarly,  $n_H$  is unaffected by *Dll-c* affinity, since in the model it is assumed to be a separate quantity indicating cooperativity among binding sites. However, the physical interpretation of  $n_H$  is not well established and whether  $n_H$  changes with affinity even less so.

Subsequently, we sought to investigate how differences in the regulatory context across regions within the L2–L3 region may determine the relationship between *Aff* and each parameter, particularly the effect on the value of  $K_a$ .

### $E_{max}$ and $K_a$ are region specific

Since we observed an unexpectedly poor correlation between  $K_a$  and *Aff*, we investigated what features might contribute to the divergence from the expected relationship within the dataset. Although the relationship between Dll concentration and expression in the L2–L3 region generally fits the pattern expected from the Hill equation, a relatively high amount of spread is observed in the data (Fig. 2I and fig. S2), reflecting differing responses to Dll concentration across space. To identify the regions contributing to this effect, we performed principal components analysis (PCA) of all wing images (Fig. 4, A and B) and found that the regions of highest variability across variants were in the distal anterior wing region (PC1) and the posterior half of the L2–L3 segment (PC2). Accordingly, variants could be grouped according to the presence or absence of a stripe of expression in L2–L3 region (Fig. 4C). This represented a notable region-specific difference in how each variant responded to Dll concentration. We, therefore, repeated the analysis in the anterior and posterior halves of the L2–L3 intervein region separately (fig. S2, B and C).

By fixing the value of  $n_H$  to the average experimentally determined value in each region (Fig. 5, G and H; anterior: 4.1, posterior: 3.7) and keeping *p* as a free parameter, the relationship between *Aff* and  $K_a$  was highly divergent between the two regions (Fig. 5, A and B). In the anterior region, *Aff* and  $K_a$  showed a better fit to the



**Fig. 4. PCA detects region-specific sources of variation.** PCA of reporter signal of all *Dll-c* variants. (A) PCA showing the contribution of the two most significant components for all variants. B is the empty construct used for background subtraction. (B) Visual representation of PC1 and PC2, where each captures the largest degree of variation nondimensionally, showing where in the wing the greatest difference in reporter activity is captured. Positive PC values correspond to higher expression levels. PC2 shows the location of the stripe of expression in the L2–L3 region. (C) Division of the L2–L3 region into anterior and posterior halves used for later analysis to either include or exclude the stripe observed in PC2.

expected model ( $R^2 = 0$ , NMRSE = 0.08; Fig. 5A) compared with that in the posterior region ( $R^2 = 0$ , NMRSE = 0.15; Fig. 5B), demonstrating that a simplistic Hill-based model of GRFs can account for enhancer activity in some regions but not in others. In addition, the values for  $K_a$  also differed in each region, with higher  $K_a$  values in the posterior region suggesting a higher threshold of Dll concentration being required to activate the enhancer.

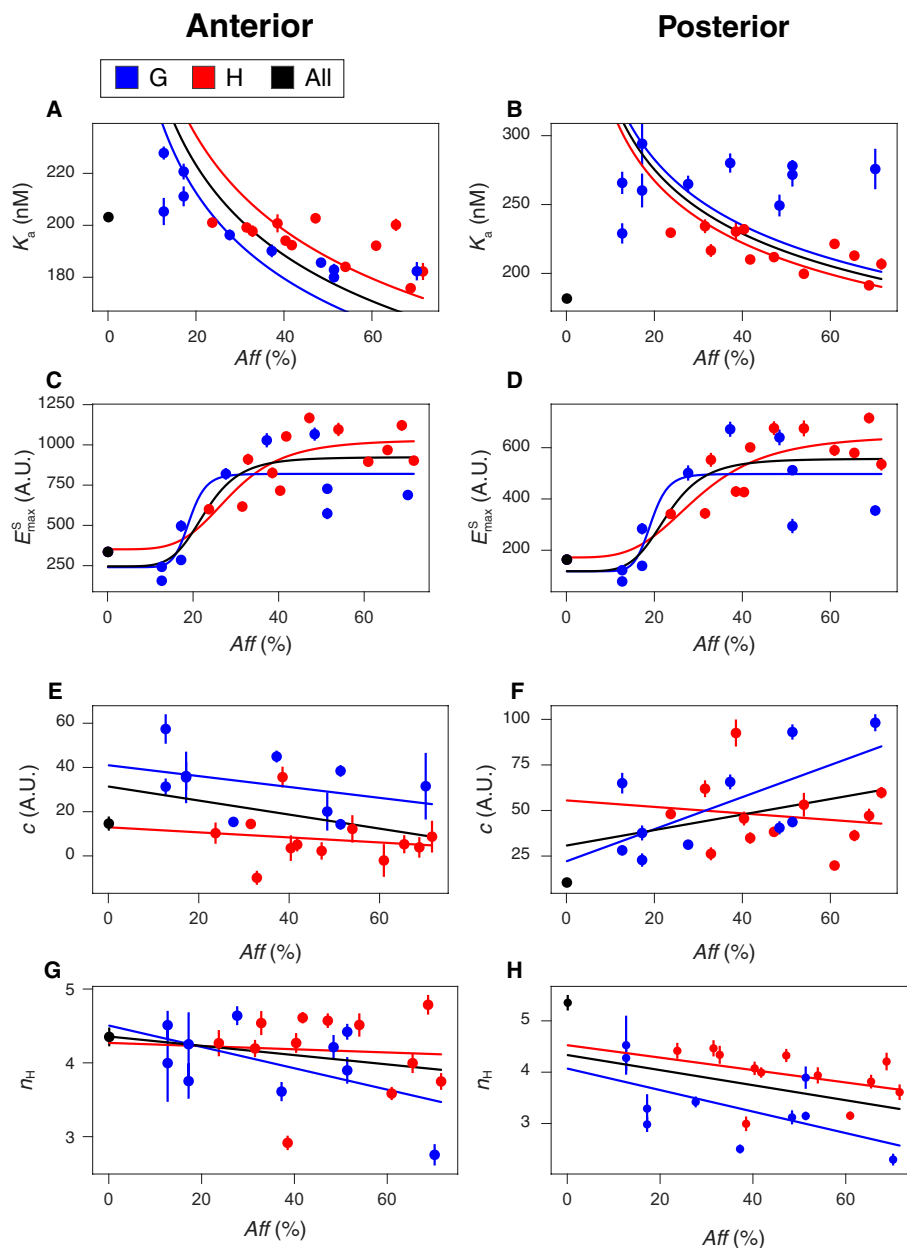
In both the anterior and posterior domain of the L2–L3 intervein region, the relationship between  $E_{\max}^S$  and  $Aff$  followed the same sigmoidal regression observed in the entire segment (anterior:  $R^2 = 0.72$ , NMRSE = 0.21; posterior,  $R^2 = 0.69$ , NMRSE = 0.24; Fig. 5, C and D) but with a reduction of  $E_{\max}^S$  by about 40% in the posterior region. As expected, baseline expression and  $n_H$  both remained constant in each half of the L2–L3 region (Fig. 5, E and F, black lines). Notably, the  $n_H$  values for the same constructs can vary across regions and may reflect

distinct modes of enhancer activation depending on the local environment. For example, if  $n_H$  relates to the number of ligands involved in cooperative binding, then the difference in  $n_H$  could reflect wing-region specific cofactors enabling binding at additional sites in the enhancer or assisting in cooperative binding between Dll molecules.

With this, the effect of region on  $E_{\max}$ ,  $n_H$ , and  $c$  can be factored into the equation, which can be rewritten as

$$E = \frac{E_{\max,R}}{1 + \left(\frac{Aff_{50}}{Aff}\right)^a} \cdot \frac{1}{1 + \left(\frac{K_a}{[T]}\right)^{n_{H,R}}} + c_R \quad (4)$$

where  $R$  is a region-specific constant modulating  $E_{\max}$ ,  $n_H$ , and  $c$ , likely reflecting the differences in TFs, cofactors, and accessibility across regions. Meanwhile, the affinity appears to have only minor—if any—effects on baseline expression and  $n_H$ . However, the region



**Fig. 5.  $E_{\max}$  and  $K_a$  are region specific.** Relation between Hill curve parameters and predicted binding preference in the anterior (A, C, E, and G) and posterior (B, D, F, and H) L2–L3 region in *Dll-c* variants with (blue) and without (red) a G at position 9. Error bars show the SD of parameter values determined from 100 iterations of generating average phenotypes from a random sample of wings from each variant.

does not appear to have a straightforward effect on  $K_a$ , whereas the relationship between  $Aff$  and  $K_a$  appears as expected in the anterior region and is lost in the posterior region. To examine how this discrepancy emerges between regions, we next divided our enhancer variants according to their sequence features.

### $K_a$ is sequence and region specific

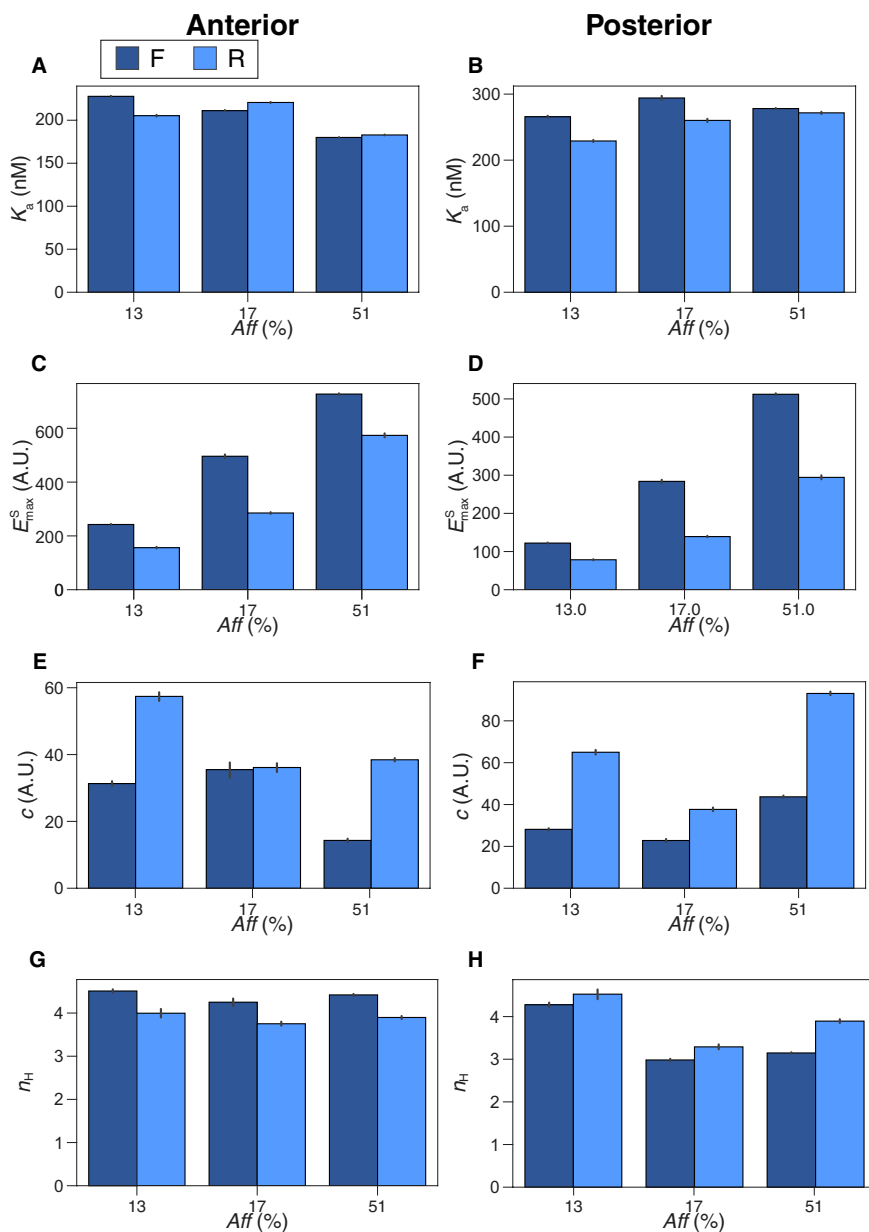
Our observations indicated that the uncoupling of  $K_a$  from SELEX-seq-measured affinity in the posterior half of the L2–L3 region corresponded to the presence or absence of a stripe of activity for each enhancer variant, as shown in PC2 (Fig. 4, B and C). A close

inspection of the variant sequences showed that all variants with a guanine (G) at position 9 of the *Dll-c* TFBS lacked the stripe, and those with any other base did have the stripe (Fig. 1D). We, therefore, separated our analyses based on the presence of a guanine (G) or any other base (H) at that position. In the anterior region, however, we observed no clear difference in the relationship between  $K_a$  and  $Aff$  in each group. Grouping the variants did not consistently change the quality of the fit (Fig. 5A; G: NRMSE = 0.06, H: NRMSE = 0.08, all: NRMSE = 0.09), suggesting that the relationship between relative affinity and  $K_a$  is not strongly affected by sequence in the anterior region. In the posterior region, fitting the expected model to each

group of variants separately produced a better fit for the H group than for the G group (Fig. 5B; G: NRMSE = 0.21; H: NRMSE = 0.06; all: NRMSE = 0.15). This indicates that the relationship between SELEX-seq-measured affinity and  $K_a$  is modulated differently by sequence depending on where in the tissue the measurement is obtained. By contrast, the sequence does not have a clear effect on  $E_{\max}^S$ . In both the anterior and posterior regions, the sigmoid fit is not significantly improved when the two groups are separated (Fig. 5C, anterior; separated: G,  $R^2 = 0.77$ , NRMSE = 0.19; H,  $R^2 = 0.73$ , NRMSE = 0.16 versus grouped:  $R^2 = 0.72$ , NRMSE = 0.21; Fig. 5D, posterior; separated: G,  $R^2 = 0.75$ , NRMSE = 0.23; H,  $R^2 = 0.72$ , NRMSE = 0.17 versus grouped:  $R^2 = 0.68$ , NRMSE = 0.24). This

suggests that the relationship between  $Aff$  and  $E_{\max}^S$  is modulated in a region-specific manner. This is unexpected as it suggests the site's capacity to induce expression (measured by  $E_{\max}^S$  in Fig. 5 (C and D)) can be uncoupled from the propensity of Dll to bind there (measured by  $K_a$  in Fig. 5, A and B), although both are correlated with SELEX-seq-measured affinity. This possibly reflects different interactions of Dll and other TFs in each region.

Regarding baseline expression and  $n_H$  (Fig. 5, E to H), only one notable region-specific difference was observed between groups of variants. The  $n_H$  value in the posterior region of L2–L3 was higher in the H group than that in the G group (Fig. 6H). This could indicate the presence of an additional parameter involved in expression in



**Fig. 6. Orientation determines  $E_{\max}^S$  but does not affect  $K_a$ .** Relation between Hill curve parameters and binding orientation in the anterior (A, C, E, and G) and posterior (B, D, F, and H) L2–L3 regions in *Dll-c* core mutants with forward (dark blue) and reverse (light blue) orientations. Error bars show the SD of parameter values determined from 100 iterations of generating average phenotypes from random sample of wings from each variant.

the H group or a missing one in the G group—possibly the effect of a cofactor assisting Dll binding at the site.

Together, these results suggest that *Dll-c* sequence has a region-specific effect on  $K_a$  values, which may be reflected in region-specific changes to  $n_H$ , whereas it does not appear to influence  $E_{\max}$  or  $Aff$ . Region-specific sequence effects on  $K_a$  can, therefore, be represented in a modification of Eq. 2 as ( $S,R$ ), such that

$$K_a = \left( \frac{1}{Aff} \right)^{1/n_H} * p_{S,R} \quad (5)$$

where  $p_{S,R}$  likely represents differences in the ability of a TF to bind to certain sequences in the presence of region-specific cofactors or a competing TF binding to an overlapping site.

### Orientation determines $E_{\max}^T$ but does not affect $K_a$

In our design, we also controlled for TFBS orientation to describe its effect on the GRF parameters. Among the core variants, three pairs of sequences had equal affinity and opposite orientation (Fig. 1C). Since all these sequences share a G at position 9, the value of  $K_a$  and  $n_H$  should not be influenced by their sequence per se.

No systematic difference between the values of  $K_a$  or  $n_H$  between forward and reverse orientation variants in either the anterior or posterior region was observed (Fig. 6, A, B, and E to H). This is consistent with the principle that  $K_a$  is primarily determined by relative affinity and sequence. Furthermore,  $n_H$  was not significantly affected by orientation, suggesting the same number of TF and cofactor molecules likely bind to the site in both orientations and both regions. While we did detect a correlation between  $c$  and orientation—some reverse sites seemed to have higher values of  $c$ —these values were obtained from regions of low Dll concentration, where Dll occupancy is likely to be very low, and were not enough to ascribe a mechanical justification for linking orientation to baseline expression. Conversely,  $E_{\max}^S$  was consistently reduced in reverse orientation variants by 40 to 50% posteriorly and 20 to 50% anteriorly (Fig. 6, C and D). This is consistent with other studies on the effect of binding orientation, which suggest that reversing orientation can significantly reduce expression (4–6, 12, 13, 16). Our results, however, further show that the reduction specifically concerns the parameter  $E_{\max}^S$ . We can, therefore, represent the effect of TFBS orientation as  $O$ , an additional coefficient of  $E_{\max}^S$

$$E = \frac{E_{\max,O,R}}{1 + \left( \frac{Aff_{50}}{Aff} \right)^a} \cdot \frac{1}{1 + \left( \frac{K_a}{[T]} \right)^{n_{H,R}}} + c_R \quad (6)$$

where  $Aff_{50}$  and  $K_a$  are described in Eqs. 4 and 5, respectively. This likely reflects the impact of orientation, not on a TF's ability to bind to a TFBS (since the affinity and sequence are the same) but on its ability to cooperate with its binding partners once bound (since the TF is presumably physically orientated in the opposite direction). Together, these results provided a deeper understanding of how expression patterning is achieved through TF-TFBS binding and the potential effects on transcriptional complex assembly. This finally led us to investigate how Dll achieves specificity for its binding sites in the first place.

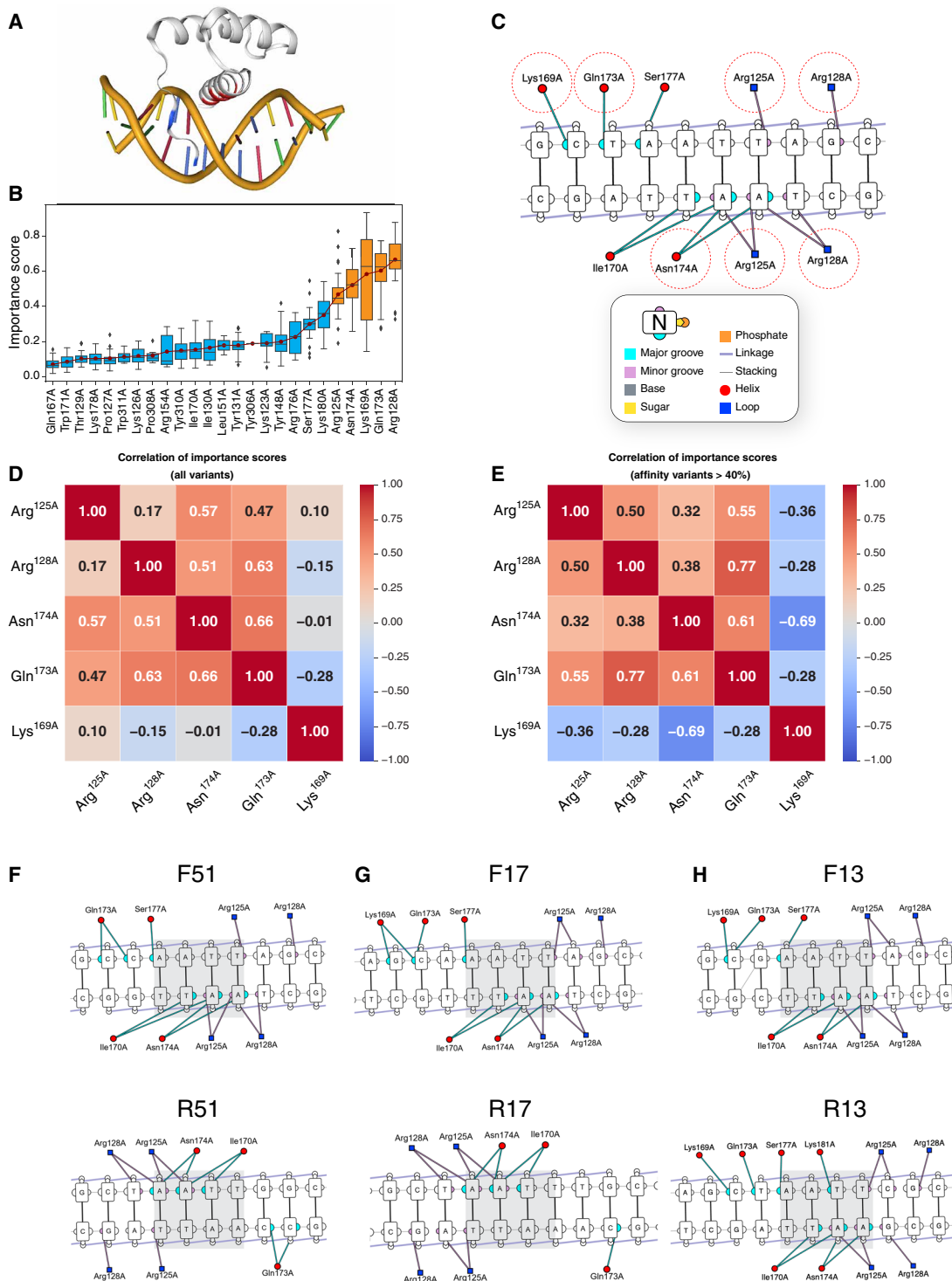
### Structural analysis of Dll binding site interactions using DeepPBS

In eukaryotic nuclei of developing organisms, which are densely packed with chromatin, TFs undergo three-dimensional (3D) diffusion

and multiple, transient, electrostatic interactions with nonspecific DNA until they find their target TFBSs to which they bind specifically to achieve precise gene control. Specific binding depends both on the TF and the target-DNA sequence and structure. TF specificity is a measure of the range of binding sites TFs will bind to with increased affinity, relative to the base level of affinity (often several orders of magnitude lower) at which nonspecific and transient TF binding occurs (46–48). Simply put, a highly specific TF will have a smaller number of TFBSs to which it stably binds [for an in-depth explanation, see (1, 49)]. Generally, different TF families will rely on certain interactions between conserved residues, such as Asn<sup>51</sup> of the recognition  $\alpha$  helix of homeodomain TFs interacting with core TAAT sequences, whereas additional amino acid–DNA interactions flanking the core can be distinct to individual homeodomain family members (50). These additional interactions increase the affinity of binding to different subsets of DNA sequences. Together, all these interactions determine the range of sites to which a TF will bind, and the participating specific residue–DNA interactions can diverge between TFs to ensure that different sets of target genes are expressed. As a homeodomain TF, Dll binds both with the third helix and the N-terminal arm of the homeodomain, its DNA binding domain (DBD), to the major and minor grooves of the DNA, respectively (Fig. 7A). To understand how Dll specificity is achieved, we explored how the Dll binding mechanism differed structurally among sites spanning a broad range of affinities.

Recently developed computational approaches can reveal which interactions a TF relies on for binding and determine the binding specificity of the TF–DNA complex. To investigate the binding mechanism of Dll to each of our *Dll-c* site variants, we used DeepPBS (51), a deep learning–based tool for predicting TF–DNA binding specificity, which can analyze both experimentally determined and computationally predicted structures. We first used AlphaFold 3 (52) to model each Dll–*Dll-c* complex and then applied DeepPBS to assign importance scores at the residue and atomic level, identifying key interactions responsible for achieving binding specificity. DeepPBS computes relative importance (RI) scores, which indicate the RI of protein–DNA contacts for the predicted position weight matrices (PWMs), where higher importance scores indicate interactions crucial for TF specificity (51). The results reveal residues that restrict the TF to a specific subset of sequences (Fig. 7B).

The most important Dll residues for conferring binding specificity (importance score of >0.4) were identified to be Lys<sup>169</sup>, Asn<sup>174</sup>, and Gln<sup>173</sup>, which bind to the major groove and Arg<sup>128</sup> and Arg<sup>125</sup>, which bind to the minor groove (Fig. 7, B and C, and table S2). While the sequences used in this study were overall similar, two distinct modes of binding specificity emerged. These binding modes were distinguished by the importance scores, primarily of Lys<sup>169</sup> alone or of Gln<sup>173</sup>, Arg<sup>128</sup>, Asn<sup>174</sup>, and Arg<sup>125</sup> together. Across all variants, the importance score of Lys<sup>169</sup> was found to be strongly anticorrelated with the importance scores for the other residues, whereas the importance scores of Gln<sup>173</sup>, Arg<sup>128</sup>, Asn<sup>174</sup>, and Arg<sup>125</sup> were positively correlated with each other (Fig. 7D and table S2), suggesting that Dll tends to prefer one mechanism or another. Furthermore, the relation was stronger among sequences with higher affinity (>40%) (Fig. 7E). This suggests that at high-affinity levels, Dll binding specificity is more likely to be achieved primarily through Lys<sup>169</sup>–linked interactions or a combination of the other four residues, whereas at lower affinities, specificity may be achieved through a combination of interactions of all residues with DNA. It



**Fig. 7. DII relies on two binding mechanisms and binds in opposite orientation on reverse complementary sites.** Importance scores of DII residues for conferring specificity to different *Dil-c* site variants. **(A)** Structure of DII DBD and its interaction with DNA. **(B)** Importance scores for each residue, sorted by their average value. The red dots indicate the mean importance score for each residue, with the residues with mean importance scores >0.4 highlighted in orange (Arg<sup>125</sup>A, Arg<sup>128</sup>A, Asn<sup>174</sup>A, Gln<sup>173</sup>A and Lys<sup>169</sup>A, with A indicating the chain). **(C)** Overview of interactions between DII residues and the *Dil-c* site (shown in the case of *FR70*). Red solid circles indicate residues in a helix, whereas blue squares represent residues in a loop. Color key indicates contacts between residues and DNA: Blue connectors from residues to DNA indicate contacts in the major groove, whereas magenta connectors indicate contacts in the minor groove—no contacts are shown between residues and sugars (yellow) or phosphates (orange). **(D)** Correlation of importance scores highlighting the relationships among Arg<sup>125</sup>A, Arg<sup>128</sup>A, Asn<sup>174</sup>A, Gln<sup>173</sup>A, and Lys<sup>169</sup>A contacts for higher affinity *Dil-c* variants (>40%). **(E)** Correlation between importance scores highlighting the relationships among Arg<sup>125</sup>A, Arg<sup>128</sup>A, Asn<sup>174</sup>A, Gln<sup>173</sup>A, and Lys<sup>169</sup>A contacts for all *Dil-c* variants. **(F)** Interactions between DII and variants *F51* and *R51*. **(G)** Interactions between DII and variants *F17* and *R17*. **(H)** Interactions between DII and variants *F13* and *R13*.

has previously been suggested that proteins achieve specificity for lower affinity sites by inducing DNA bending (50, 53, 54). The results shown here may reflect that DNA bending of lower affinity sequences is associated with a more even distribution of interaction importance between TF residues and DNA, as the interaction relies on a higher number of residues.

From the structural analysis, we also obtained evidence for the change in Dll orientation relative to DNA in the reverse-complement sites F/R51 and F/R17, whereby TF-DNA interactions with the major groove occur at the 5' end in one and 3' end in the other (Fig 7, F to H). As noted previously, this likely reflects the effect of reversing the orientation of the TFBS sequence on expression, where Dll binds in the opposite orientation and may be less able to interact with its binding partners. However, for the F/R13 pair, major groove interactions occur at the 5' end in both sequences (Fig. 7H)—this is true for all five of the most likely binding configurations predicted by AlphaFold 3. This may be a function of low affinity, where the precise sequence of the site—and the interaction between residues and specific nucleotides—becomes less important than external features such as the structure of the extended DNA region and promotes Dll to favor the forward orientation. Because the effect on expression is unlikely to act through changes in residence time at the site, it may reflect that this configuration is suboptimal for the protein to interact with its binding partners before transcription initiation, even with the protein binding in the forward orientation. Alternatively, there may still be minor binding species that bind in the reverse orientation, which are not among the most likely predicted by AlphaFold 3.

Our structural analysis showed that Dll binding to its target sites relies on two distinct mechanisms at higher affinity sites and a more evenly distributed set of interactions at lower affinity sites. Furthermore, we have provided structural evidence for Dll's binding in opposite orientation on complementary sites, which likely explains the mechanism by which  $E_{\max}^S$  is decreased in oppositely orientated binding sites.

## DISCUSSION

Our study illustrates how biophysical properties of a TFBS in an enhancer influence its transcriptional output, integrating TF concentration and local regulatory environment to determine expression levels. Starting from a simple model where TFBS affinity changes the value of  $K_a$ , we show that affinity has a positive sigmoidal effect on maximum expression (potentially through changes to the duration of TF binding at the *Dll-c* site resulting in increased transcriptional burst frequency or in increased local accessibility). However, we detected locations and sequence-specific factors that can disrupt the relationship between affinity,  $K_a$ , and maximum expression and affect the baseline expression and  $n_H$ , highlighting the complexity of gene regulation and the prediction of gene expression from sequence changes in complex tissues. We also found that reversing the orientation of a TFBS produces a change specifically in the value of maximum expression and provide structural evidence that this may result from Dll binding in an opposite orientation on the binding site and/or binding with suboptimal configuration for driving expression. Last, we showed that the Dll interactions with its binding site tended to act through two different mechanisms—one mostly dependent on interactions between Lys<sup>169</sup> and another with Gln<sup>173</sup>, Arg<sup>128</sup>, Asn<sup>174</sup>, and Arg<sup>125</sup>—whereas the distinction between the two mechanisms became stronger at high affinities.

Our study aimed to bridge the gap between data-intensive studies of enhancer regulation, such as those using massively parallel reporter assays often using single-cell data [e.g., (55, 56)], and studies on the general principles of phenotypic divergence in developing organisms. The fact that high-affinity binding and expression are relative terms and often lack quantitative values is generally overlooked in developmental studies. For example, in a recent study on digit formation in mice, affinity-optimizing mutations resulted in ectopic expression of *Shh* and caused polydactyly (21); however, the levels at which affinity changes result in a new phenotype, and the affinity values that span transitional states between phenotypes are rarely shown. The wild-type *Dll-c* site has an affinity that would place it in the transitional range of the *Aff* versus  $E_{\max}^S$  curve (Fig. 3D and fig. S1). Given a sigmoidal relationship between *Aff* versus  $E_{\max}^S$ , selection on TFBS affinities may drive TFBSs toward this transitional range explaining why the highest affinity levels are rarely seen in the wild (11).

Other studies have suggested a trade-off among TFBS affinity, specificity of binding, and subsequent spatial expression (33, 57). Affinity-modulating mutations in TFBSs can modify the expression domain of an enhancer by enabling interactions between TFBSs and new TFs because related TFs are likely to bind to similar high-affinity sites (57–59). We showed that specific sequence features can disrupt the expected relationship between *Aff* and  $K_a$ , correlating with a change in the spatial activity of the enhancer (the posterior L2–L3 stripe). Because this shift appears to be based on a sequence feature at a broad range of affinities, it may not reflect two related TFs binding to the same high-affinity site. Instead, it may be due to the creation of an alternate binding site for a different TF. This highlights the possibility that, in addition to the trade-off between affinity and specificity, sequence- and region-specific modifiers that entirely interrupt the relationship between TFBS features and expression may exist. What we refer to as the “sequence and region-specific modifiers” likely represent the combination of TF, cofactors, and accessibility differences across regions, whose effects need to be determined to assess the effect of mutations.

Similarly, TFBSs are often wrongly assumed to be more strictly defined than they are. Advances in the study of low-affinity binding clusters suggests that the boundaries of individual TF binding sites are often more loosely specified (60). Even in this study, the first round of *Dll-c* site variants showed that the relationship between TFBS features and expression are only as useful as the definition of the TFBS. In other words, TFs can find other favorable sites if their “usual” binding site is altered. One of the reasons we could identify a relationship between affinity and expression in our study is that we deliberately designed our *Dll-c* variants to favor Dll binding directly at the center of *Dll-c* and disrupt potential sites on either side of it. A difficulty in predicting the effect of TFBS mutations is to determine whether a TF will continue to bind at that site when a mutation has been introduced. However, there is no reason why the enhancer would not be governed by the same set of rules concerning enhancer activity and affinity once bound.

To predict the activity of an enhancer from its sequence requires studies into how expression levels are achieved. Numerous studies using machine learning have shown that this is within reach (14, 56, 61) and that models that predict enhancer activity from a sequence are available. However, current models are generally trained for specific cell types or time points and are, therefore, generally not interpretable as they do not show how this expression is achieved. Our work has aimed to complement these studies using a living animal organ to

integrate developmental and positional information into unpacking the mechanisms of how enhancer sequences mediate the conversion of TF signals into expression. High-throughput binding data are available for an increasing number of TFs, and computational tools can predict binding mechanisms for TF-TFBS interactions. However, even in our relatively controlled environment, we see that nearly identical sequences can behave very differently across distances, equivalent of only a handful of cells. For a machine learning model to predict those differences, a major limiting factor is the inability to determine nongenetic inputs (including the transenvironment of TFs and chromatin organization) experimentally. Similarly, the enhancers that can be designed to achieve precise expression in tissue cultures by machine learning models are often based on higher-affinity TFBSs and are unlikely to reproduce the complex patterns observed in developing tissues, which results from multiple overlapping sites of varying affinities. Also, these synthetic enhancers are designed without factoring in quantitative expression levels. Here, we have carefully dissected the relationship between each parameter across a complex tissue based on qualitatively richer data than that typically used to train machine learning models. If these models were trained on data taking into account precise concentrations of TFs and transcripts across different regions, they would say much more about gene expression than what is currently possible.

Last, while using the Hill equation as a model of gene regulation is not meant to provide a comprehensive, mechanistic view of transcription, identifying the gaps between the assumptions of these models and the data collected is essential for building more complex models of gene regulation. Equation 2 is drawn from a model of transcription in which all binding sites are cooperatively bound, and expression is linearly proportional to site occupancy. To go beyond this model, one would have to check if other models incorporating additional enhancer-protein interactions, energy expenditure (25, 62), molecular condensates, and changes in chromatin accessibility (63–65) could better fit the data (ensuring the experiments provide enough resolution/statistical power to make this claim). In addition, the study would need to be extended experimentally by identifying these additional factors and perturb their expression or check if they are expressed in a gradient and include them in the GRF model, testing indirectly different hypotheses for their interaction.

Together, our study indicates that even a relatively simple model of gene regulation may be used to make predictions of gene activity. While the complexity of developing tissues requires adjusting the model in different contexts, we show precisely where some of these adjustments should be made.

## MATERIALS AND METHODS

### STAR methods

#### Fly husbandry

Our *D. melanogaster* stocks were maintained on standard cornmeal medium at 25°C with a 12-hour:12-hour day:night light cycle.

#### Transgenesis

All reporter constructs were injected as in (34). We used  $\phi$ C31-mediated transgenesis (66) and integrated all constructs at the genomic *attP* site VK00016 (67) on chromosome 2. All transgenic lines were genotyped to ascertain that the enhancer sequence was correct.

#### SELEX-seq

Dll was produced and purified from a pET-11a expression vector containing full-length *Dll* coding sequence and a His tag. SELEX-seq

was performed as previously described (37, 68, 69). Briefly, a SELEX-seq library with a 16-bp randomized region was prepared. The library was then incubated with 20 nM Dll (~10-fold less than DNA) for 30 min at room temperature before loading onto an electrophoretic mobility shift assay gel. Following gel purification of the band corresponding to Dll monomers bound to DNA, both round 0 (R0) and round 1 (R1) libraries were amplified using Q5 polymerase and NEBnext Multiple Oligos for Illumina (New England Biolabs). The final library was purified using a native TBE-gel, sequenced with an Illumina NextSeq desktop sequencer and analyzed using the Top-Down Crawl method (38) to assign values of relative binding affinity to each sequence.

#### Construct design

In the initial round, constructs were chosen sequence agnostically to obtain a range of relative affinities from 14 to 87%. Each construct was analyzed to ensure that no additional Dll binding sites were formed. In the second round, the sequences were modified forms of the line I-51 from the initial round. In both cases, an additional reporter construct, devoid of *spot*<sup>196</sup> insert ( $\emptyset$ ), was used.

#### Molecular biology

All 196-bp constructs derived from the *Drosophila biarmipes spot*<sup>196</sup> sequence were synthesized in vitro by Integrated DNA Technologies, Coralville, IA, USA (catalog no. 121416). Table S1 provides a list of all constructs and their sequences. Each construct was cloned by In-Fusion (Takara, San Jose, CA, USA) into our pRedSA vector as described in (39).

#### Wing preparation and imaging

All transgenic wings imaged in this study were homozygous for the reporter construct. For each reporter line, we imaged 15 to 30 wings of freshly eclosed flies. All samples were prepared, imaged, and processed as in (39).

#### Image processing and quantification

Wing images were aligned using wing veins as landmarks, as described in detail in (39). Our image registration pipeline is accessible at <https://github.com/UniBonn-GompelLab/Image-registration-gui>. Procedures to describe fluorescent signal, derive average phenotypes or overall intensity values, apply color maps, and produce PCAs are also presented in (39). The affinity modulation of a single site produced a range of activities where the highest level (A69) was six times greater than the lowest (A00). To generate a GRF of enhancer activity with respect to TFBS affinity, the average signal intensity from the entire wing was used as a measure of enhancer activity.

#### Fluorescence correlation spectroscopy

Nanomolar Dll concentrations were obtained using FCS (70). As a first step, we determined the relative concentration of Dll across pupal wings at a stage when the *spot*<sup>196</sup> enhancer is active, i.e., at 75% of pupal development (36). To estimate relative concentrations, we used a transcriptional reporter, *tdTomato-Dll*, in which the coding sequence of a *tdTomato* reporter was inserted at the *Dll* locus, upstream of the *Dll* coding sequence, with a T2A cleavage signal between both open reading frames (43). We imaged nine stage-matched pupal wings (75%; 82 hours of pupal development) using a Ti2 Eclipse Nikon microscope as described above. The images were aligned, and we processed the resulting images as we did for *spot*<sup>196</sup> variants (see above). From this, we derived an average image, which represented a map of the relative Dll concentration across the wing. Assuming that the protein lifetime of *tdTomato* did not differ significantly from that of Dll, we considered that *tdTomato* fluorescence intensity was a reasonable proxy of Dll concentration, with

one molecule of Dll protein being produced for each tdTomato protein. Therefore, we used this *tdTomato-Dll* allele to measure the absolute concentration of tdTomato in 75% pupal wings using FCS.

FCS experiments were performed as previously described in (42) with the following modifications. Measurements were performed on a Zeiss LSM780 inverted setup, comprising diode of 405 nm; Ar multiline of 458 nm, 488 nm, and 514 nm; diode-pumped solid-state laser (DPSS) of 561 nm; and 633-nm HeNe lasers. The instrument was modified to enable detection using silicon avalanche photo detectors (APDs) (SPCM-AQR-1X; PerkinElmer, USA) for imaging and FCS. Before performing FCS measurements, images were recorded at a 512 × 512 pixel resolution. Compared with photomultiplier tubes (PMTs), which are normally used as detectors in conventional confocal laser scanning microscopy, APDs are characterized by higher quantum yield and collection efficiency—about 70% in APDs as compared to 15 to 25% in PMTs, higher gain, negligible dark current, and better efficiency in detecting red fluorescence. Enhanced fluorescence detection efficiency enabled image collection using fast scanning (1 to 5 μs/pixel). This further enhances the signal-to-noise ratio by preventing fluorescence loss due to triplet state formation, thus enabling fluorescence imaging with single-molecule sensitivity. In addition, low laser intensities (150 to 750 μW) could be applied for imaging, significantly reducing phototoxicity (71). A C-Apochromat 40×/1.2 W ultraviolet-visible-infrared objective was used throughout. Fluorescence intensity fluctuations were recorded in arrays of 10 consecutive measurements, with each measurement lasting 10 s.

FCS measurements were performed by recording fluorescence intensity fluctuations in a very small, approximately ellipsoidal observation volume element (OVE) (about 0.2-μm wide and 1-μm long), which was generated in pupal wing cells by focusing the laser light through the microscope objective and collecting the fluorescence light through the same objective using a pinhole in front of the detector to block out-of-focus light. The fluorescence intensity fluctuations, caused by fluorescently labeled molecules passing through the OVE, are analyzed using temporal autocorrelation analysis.

In temporal autocorrelation analysis, we first derive the autocorrelation function

$$G(\tau) = 1 + \frac{\langle \delta I(t) \cdot \delta I(t + \tau) \rangle}{\langle I(t) \rangle^2}$$

where  $\delta I(t) = I(t) - \langle I(t) \rangle$  is the deviation from the mean intensity at time  $t$  and  $\delta I(t + \tau) = I(t + \tau) - \langle I(t) \rangle$  is the deviation from the mean intensity at time  $t + \tau$ . For further analysis, an autocorrelation curve is derived by plotting  $G(\tau)$  as a function of the lag time, i.e., the autocorrelation time  $\tau$ .

To derive information about molecular numbers and their corresponding diffusion time, the experimentally obtained autocorrelation curves were fitted with a model describing free 3D diffusion and triplet formation. Since fitting tdTomato autocorrelation curves with a one-component diffusion model did not result in satisfactory fitting (as observed by visual inspection and residuals analysis), it is likely that a subpopulation of the tdTomato molecules diffuses at slower speeds, presumably reflecting incomplete self-cleavage of the T2A peptide. Fitting the autocorrelation curves with a two-component model was, therefore, identified as simplest and best

suitable for fitting the experimentally derived autocorrelation curves and used throughout the analysis

$$G(\tau) = 1 + \frac{1}{N} \left[ \frac{1-y}{\left(1 + \frac{\tau}{\tau_{D_1}}\right) \sqrt{1 + \frac{w_{xy}^2 \tau}{w_z^2 \tau_{D_1}}}} + \frac{y}{\left(1 + \frac{\tau}{\tau_{D_2}}\right) \sqrt{1 + \frac{w_{xy}^2 \tau}{w_z^2 \tau_{D_2}}}} \right] \left(1 + \frac{\tau}{1-T} e^{-\frac{\tau}{\tau_T}}\right)$$

In the above equation,  $N$  is the average number of molecules in the OVE;  $y$  is the fraction of the slowly moving tdTomato molecules;  $\tau_{D_1}$  is the diffusion time of the free tdTomato molecules;  $\tau_{D_2}$  is the diffusion time of slowly-moving tdTomato molecules;  $w_{xy}$  and  $w_z$  are radial and axial parameters, respectively, related to spatial properties of the OVE;  $T$  is the average equilibrium fraction of molecules in the triplet state; and  $\tau_T$  the triplet correlation time related to rate constants for intersystem crossing and the triplet decay. Spatial properties of the detection volume, represented by the square of the

ratio of the axial and radial parameters  $\left[\left(\frac{w_z}{w_{xy}}\right)^2\right]$ , were determined in calibration measurements performed using a solution of Rhodamine 6G for which the diffusion coefficient ( $D$ ) is known to be  $D_{Rh6G} = 4.1 \cdot 10^{-10} \text{ m}^2 \text{ s}^{-1}$  (72). The diffusion time,  $\tau_D$ , measured by FCS, is related to the translation diffusion coefficient  $D$  by

$$\tau_D = \frac{w_{xy}^2}{4D}$$

Averaged curves were analyzed using the software for online data analysis, and the nonlinear least square fitting of the autocorrelation curve was performed using the Levenberg-Marquardt algorithm. The quality of the fitting was evaluated by visual inspection and residuals analysis.

We carried out one to four FCS measurements at different positions on 16 individual pupal wings. To translate the average number of molecules in the OVE ( $N$ ) into molar concentration, the size of the OVE, i.e., the axial and radial parameters ( $w_z$  and  $w_{xy}$ , respectively) were determined in calibration experiments with Alexa488 or Rhodamine 6G dyes. The volume of the OVE, approximated by a prolate ellipsoid, was determined as follows

$$\begin{aligned} V_{\text{OVE}} &= \pi^{\frac{3}{2}} \cdot w_{xy}^2 \cdot z_0 = 5.57 \cdot 0.15873^2 \cdot 1 \\ &= 0.14034 \cdot 10^{-18} \text{ m}^3 = 0.14 \cdot 10^{-15} \text{ L} \end{aligned}$$

Thereafter, the average number of molecules in the OVE ( $N$ ) was converted into molar concentration ( $C$ ) using the relationship

$$C = \frac{N}{N_A \cdot V_{\text{OVE}}}$$

where  $N_A$  is the Avogadro number ( $6.022 \cdot 10^{23} \text{ mol}^{-1}$ ), which indicates that one molecule in the OVE corresponds, on average, to a concentration of 11.84 nM concentration of tdTomato molecules in the cell.

Our data covered a broad range of concentrations across the wings, spanning the entire Dll gradient, and correlated well with the Dll relative concentration at each position ( $[\text{Dll}]_{\text{absolute}} = 2.26 \cdot [\text{Dll}]_{\text{relative}}$ ), where  $[\text{Dll}]_{\text{absolute}}$  is the absolute concentration of Dll in nanomolar

measured by FCS and  $[DII]_{\text{relative}}$  is the relative concentration of DII in arbitrary units (Fig. 3A). Last, we interpolated the discrete measurements to our map of relative DII concentrations to obtain a map of absolute concentrations (Fig. 3B).

### Data analysis

To obtain estimates for  $K_a$ ,  $E_{\text{max}}$ ,  $c$  and  $n_H$  for each construct and to have a measure of the uncertainty of those estimates, bootstrapping was used to generate multiple average expression profiles based on a random selection of images ( $n = 30$ ) from each construct, and curves fitted to the resulting profile; this process was repeated 100 times. As in (36), PCA was used to remove correlation between pixel intensities, to concentrate the variance on few variables, and, therefore, to describe the variation in intensity and pattern of reporter gene expression in a comprehensive and unbiased way with few dimensions. PCA was calculated on the matrix of dimensions ( $n_{\text{individual}} \times n$  pixels on the wing).

### Structure predictions

Structure predictions for DII bound to each *DII-c* site variants were generated using AlphaFold 3 (52), which produces five top-ranked predictions per complex, with the highest-ranked model selected for analysis based on confidence scores. Each site variant contained a 10-bp binding site flanked by 10 bp on both sides, totaling 30 bp. To access TF-DNA interactions, specificity importance scores were assigned to each DII residue using DeepPBS (51), a deep-learning tool that quantifies the contribution of individual residues to DNA binding specificity. DeepPBS residue-level importance scores measure the impact of amino acid atoms on binding specificity by perturbing interactions within the TF-DNA complex, with nodes representing the protein on one side and DNA on the other side. Systematic inactivation of each amino acid was performed, and changes in PWM were measured, with large differences indicating stronger interactions. The importance score is a nonunit value as it reflects relative changes rather than absolute measurements. Visualization of DII-residue interactions with the *DII-c* site was performed using DNAProDB (73) based on the AlphaFold 3–predicted structures of protein-DNA complexes.

## Supplementary Materials

### The PDF file includes:

Figs. S1 and S2  
Legends for tables S1 to S3

### Other Supplementary Material for this manuscript includes the following:

Tables S1 to S3

## REFERENCES

- G. D. Stormo, *Introduction to Protein-DNA Interactions: Structure, Thermodynamics, and Bioinformatics* (Cold Spring Harbor Laboratory Press, 2013).
- F. Liu, J. W. Posakony, Role of architecture in the function and specificity of two Notch-regulated transcriptional enhancer modules. *PLOS Genet.* **8**, e1002796 (2012).
- D. M. King, C. K. Y. Hong, J. L. Shepherdson, D. M. Granas, B. B. Maricque, B. A. Cohen, Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. *eLife* **9**, e41279 (2020).
- J. W. Cave, F. Loh, J. W. Surpris, L. Xia, M. A. Caudy, A DNA transcription code for cell-specific gene activation by notch signaling. *Curr. Biol.* **15**, 94–104 (2005).
- K. Senger, G. W. Armstrong, W. J. Rowell, J. M. Kwan, M. Markstein, M. Levine, Immunity regulatory DNAs share common organizational features in *Drosophila*. *Mol. Cell* **13**, 19–32 (2004).
- I. Georgakopoulos-Soares, C. Deng, V. Agarwal, C. S. Y. Chan, J. Zhao, F. Inoue, N. Ahituv, Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nat. Commun.* **14**, 2333 (2023).
- S. D. Hanes, G. Riddihough, D. Ish-Horowitz, R. Brent, Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. *Mol. Cell. Biol.* **14**, 3364–3375 (1994).
- H. N. Cai, D. N. Arnosti, M. Levine, Long-range repression in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9309–9314 (1996).
- E. D. Flynn, A. L. Tsu, S. Kasela, S. Kim-Hellmuth, F. Aguet, K. G. Ardlie, H. J. Bussemaker, P. Mohammadi, T. Lappalainen, Transcription factor regulation of eQTL activity across individuals and tissues. *PLOS Genet.* **18**, e1009719 (2022).
- S. S. Nishizaki, N. Ng, S. Dong, R. S. Porter, C. Morterud, C. Williams, C. Asman, J. A. Switzenberg, A. P. Boyle, Predicting the effects of SNPs on transcription factor binding affinity. *Bioinformatics* **36**, 364–372 (2020).
- E. K. Farley, K. M. Olson, W. Zhang, A. J. Brandt, D. S. Rokhsar, M. S. Levine, Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).
- C. D. Arnold, D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath, A. Stark, Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
- R. R. Catarino, A. Stark, Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* **32**, 202–223 (2018).
- Z. Avsec, M. Weilert, A. Shrikumar, S. Krueger, A. Alexandari, K. Dalal, R. Profp, C. McAnany, J. Gagneur, A. Kundaje, J. Zeitlinger, Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).
- G. A. Jindal, E. K. Farley, Enhancer grammar in development, evolution, and disease: Dependencies and interplay. *Dev. Cell* **56**, 575–587 (2021).
- D. Panne, The enhanceosome. *Curr. Opin. Struct. Biol.* **18**, 236–242 (2008).
- D. Thanos, T. Maniatis, Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).
- D. N. Arnosti, M. M. Kulkarni, Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* **94**, 890–898 (2005).
- E. K. Farley, K. M. Olson, M. S. Levine, Regulatory principles governing tissue specificity of developmental enhancers. *Cold Spring Harb. Symp. Quant. Biol.* **80**, 27–32 (2015).
- G. A. Jindal, A. T. Bantle, J. J. Solvason, J. L. Grudzien, A. D'Antonio-Chronowska, F. Lim, S. H. Le, B. P. Song, M. F. Ragsac, A. Klie, R. O. Larsen, K. A. Frazer, E. K. Farley, Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. *Dev. Cell* **58**, 2206–2216.e5 (2023).
- F. Lim, J. J. Solvason, G. E. Ryan, S. H. Le, G. A. Jindal, P. Steffen, S. K. Jandu, E. K. Farley, Affinity-optimizing enhancer variants disrupt development. *Nature* **626**, 151–159 (2024).
- H. D. Kim, E. K. O'Shea, A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* **15**, 1192–1198 (2008).
- J. Park, J. Estrada, G. Johnson, B. J. Vincent, C. Ricci-Tam, M. D. Bragdon, Y. Shulgina, A. Cha, Z. Wunderlich, J. Gunawardena, A. H. DePace, Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. *eLife* **8**, e41266 (2019).
- S. A. Frank, Input-output relations in biological systems: Measurement, information and the Hill equation. *Biol. Direct* **8**, 31 (2013).
- J. Estrada, F. Wong, A. DePace, J. Gunawardena, Information integration and energy expenditure in gene regulation. *Cell* **166**, 234–244 (2016).
- R. Martinez-Corral, K. M. Nam, A. H. DePace, J. Gunawardena, The Hill function is the universal Hopfield barrier for sharpness of input-output responses. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2318329121 (2024).
- T. Gregor, D. W. Tank, E. F. Wieschaus, W. Bialek, Probing the limits to positional information. *Cell* **130**, 153–164 (2007).
- M. S. Sherman, B. A. Cohen, Thermodynamic state ensemble models of cis-regulation. *PLOS Comput. Biol.* **8**, e1002407 (2012).
- J. Goutsias, S. Kim, A nonlinear discrete dynamical model for transcriptional regulation: Construction and properties. *Biophys. J.* **86**, 1922–1945 (2004).
- H. G. Garcia, A. Sanchez, T. Kuhlman, J. Kondev, R. Phillips, Transcription by the numbers redux: Experiments and calculations that surprise. *Trends Cell Biol.* **20**, 723–733 (2010).
- R. Phillips, J. Kondev, J. Theriot, H. Garcia, *Physical biology of the cell* (Garland Science, 2012).
- J. Gunawardena, Models in biology: Accurate descriptions of our pathetic thinking. *BMC Biol.* **12**, 29 (2014).
- J. Crocker, N. Abe, L. Rinaldi, A. P. McGregor, N. Frankel, S. Wang, A. Alsawadi, P. Valenti, S. Plaza, F. Payre, R. S. Mann, D. L. Stern, Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
- L. Arnoult, K. F. Su, D. Manoel, C. Minervino, J. Magrina, N. Gompel, B. Prud'homme, Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* **339**, 1423–1426 (2013).
- N. Gompel, B. Prud'homme, P. J. Wittkopp, V. A. Kassner, S. B. Carroll, Chance caught on the wing: Cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481–487 (2005).

36. Y. Xin, Y. Le Poul, L. Ling, M. Museridze, B. Muhling, R. Jaenichen, E. Osipova, N. Gompel, Enhancer evolutionary co-option through shared chromatin accessibility input. *Cur. Opin. Chem. Biol.* **117**, 20636–20644 (2020).
37. T. R. Riley, M. Slattery, N. Abe, C. Rastogi, D. Liu, R. S. Mann, H. J. Bussemaker, SELEX-seq: A method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods Mol. Biol.* **1196**, 255–278 (2014).
38. B. H. Cooper, T. P. Chiu, R. Rohs, Top-down crawl: A method for the ultra-rapid and motif-free alignment of sequences with associated binding metrics. *Bioinformatics* **38**, 5121–5123 (2022).
39. Y. Le Poul, Y. Xin, L. Ling, B. Muhling, R. Jaenichen, D. Horl, D. Bunk, H. Harz, H. Leonhardt, Y. Wang, E. Osipova, M. Museridze, D. Dharmadhikari, E. Murphy, R. Rohs, S. Preibisch, B. Prud'homme, N. Gompel, Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci. Adv.* **6**, eabe2955 (2020).
40. R. Giri, S. Brady, D. K. Papadopoulos, R. W. Carthew, Single-cell Senseless protein analysis reveals metastable states during the transition to a sensory organ fate. *iScience* **25**, 105097 (2022).
41. R. Giri, D. K. Papadopoulos, D. M. Posadas, H. K. Potluri, P. Tomancak, M. Mani, R. W. Carthew, Ordered patterning of the sensory system is susceptible to stochastic features of gene expression. *eLife* **9**, e53638 (2020).
42. D. K. Papadopoulos, K. Skouloudaki, Y. Engstrom, L. Terenius, R. Rigler, C. Zechner, V. Vukojevic, P. Tomancak, Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch. *Development* **146**, dev168179 (2019).
43. V. Chaudhary, S. Hingole, J. Frei, F. Port, D. Strutt, M. Boutros, Robust Wnt signaling is maintained by a Wg protein gradient and Fz2 receptor activity in the developing *Drosophila* wing. *Development* **146**, dev174789 (2019).
44. G. Morata, P. A. Lawrence, Control of compartment development by the engrailed gene in *Drosophila*. *Nature* **255**, 614–617 (1975).
45. M. Bozek, N. Gompel, Developmental transcriptional enhancers: A subtle interplay between accessibility and activity: Considering quantitative accessibility changes between different regulatory states of an enhancer deconvolutes the complex relationship between accessibility and activity. *Bioessays* **42**, e1900188 (2020).
46. V. Vukojevic, D. K. Papadopoulos, L. Terenius, W. J. Gehring, R. Rigler, Quantitative study of synthetic Hox transcription factor-DNA interactions in live cells. *Int. J. Mol. Sci.* **107**, 4093–4098 (2010).
47. J. Chen, Z. Zhang, L. Li, B. C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, Z. Liu, Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
48. M. Raccaud, E. T. Friman, A. B. Alber, H. Agarwal, C. Deluz, T. Kuhn, J. C. M. Gebhardt, D. M. Suter, Mitotic chromosome binding predicts transcription factor properties in interphase. *Nat. Commun.* **10**, 487 (2019).
49. G. D. Stormo, Y. Zhao, Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.* **11**, 751–760 (2010).
50. R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, R. S. Mann, Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
51. R. Mitra, J. Li, J. M. Sagendorf, Y. Jiang, A. S. Cohen, T. P. Chiu, C. J. Glasscock, R. Rohs, Geometric deep learning of protein-DNA binding specificity. *Nat. Methods* **21**, 1674–1683 (2024).
52. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Zengulyte, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Zidek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
53. R. S. Hegde, The papillomavirus E2 proteins: Structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 343–360 (2002).
54. S. S. Kim, J. K. Tam, A. F. Wang, R. S. Hegde, The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**, 31245–31254 (2000).
55. C. Bravo Gonzalez-Blas, I. Matetovici, H. Hillen, I. I. Taskiran, R. Vandepoel, V. Christiaens, L. Sansores-Garcia, E. Verboven, G. Hulselmans, S. Poovathingal, J. Demeulemeester, N. Psatha, D. Mauduit, G. Halder, S. Aerts, Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. *Nat. Cell Biol.* **26**, 153–167 (2024).
56. B. P. de Almeida, F. Reiter, M. Pagani, A. Stark, DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* **54**, 613–624 (2022).
57. J. F. Kribelbauer, C. Rastogi, H. J. Bussemaker, R. S. Mann, Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* **35**, 357–379 (2019).
58. A. Jolma, J. Yan, T. Whittington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
59. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J. C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F. Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
60. A. Shahein, M. Lopez-Malo, I. Istomin, E. J. Olson, S. Cheng, S. J. Maerkl, Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. *Nat. Commun.* **13**, 5273 (2022).
61. I. I. Taskiran, K. I. Spanier, H. Dickmanken, N. Kempynck, A. Pancikova, E. C. Eksi, G. Hulselmans, J. N. Ismail, K. Theunis, R. Vandepoel, V. Christiaens, D. Mauduit, S. Aerts, Cell-type-directed design of synthetic enhancers. *Nature* **626**, 212–220 (2024).
62. J. Trojanowski, K. Rippe, Transcription factor binding and activity on chromatin. *Curr. Opin. Syst. Biol.* **31**, 100438 (2022).
63. M. Mir, A. Reimer, J. E. Haines, X. Y. Li, M. Stadler, H. Garcia, M. B. Eisen, X. Darzacq, Dense Bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev.* **31**, 1784–1794 (2017).
64. G. Pei, H. Lyons, P. Li, B. R. Sabari, Transcription regulation by biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **26**, 213–236 (2025).
65. M. Woringer, X. Darzacq, I. Izeddin, Geometry of the nucleus: A perspective on gene expression regulation. *Curr. Opin. Chem. Biol.* **20**, 112–119 (2014).
66. A. C. Groth, M. Fish, R. Nusse, M. P. Calos, Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775–1782 (2004).
67. K. J. Venken, Y. He, R. A. Hoskins, H. J. Bellen, Placman: A BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**, 1747–1751 (2006).
68. J. F. Kribelbauer, R. E. Loker, S. Feng, C. Rastogi, N. Abe, H. T. Rube, H. J. Bussemaker, R. S. Mann, Context-dependent gene regulation by homeodomain transcription factor complexes revealed by shape-readout deficient proteins. *Mol. Cell* **78**, 152–167.e11 (2020).
69. M. Slattery, T. Riley, P. Liu, N. Abe, P. Gomez-Alcala, I. Dror, T. Zhou, R. Rohs, B. Honig, H. J. Bussemaker, R. S. Mann, Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
70. D. Magde, E. L. Elson, W. W. Webb, Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers* **13**, 29–61 (1974).
71. V. Vukojevic, M. Heidkamp, Y. Ming, B. Johansson, L. Terenius, R. Rigler, Quantitative single-molecule imaging by confocal laser scanning microscopy. *Int. J. Mol. Sci.* **105**, 18176–18181 (2008).
72. C. B. Müller, A. Loman, V. Pacheco, F. Koberling, D. Willbold, W. Richterling, J. Enderlein, Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *Europhys. Lett.* **83**, 46001 (2008).
73. R. Mitra, A. S. Cohen, J. M. Sagendorf, H. M. Berman, R. Rohs, DNAProDB: An updated database for the automated and interactive analysis of protein-DNA complexes. *Nucleic Acids Res.* **53**, D396–D402 (2025).

**Acknowledgments:** We are grateful to G. Álvarez Canales for comments on the manuscript, M. Boutros and F. Port for sharing the dTomato-Dll fly line, and P. Tomancak for facilitating the FCS experiments. **Funding:** This work was supported by the Graduate School of Quantitative Biosciences Munich (QBM to C.F. and L.L.) and the Deutsche Forschungsgemeinschaft (GO2495/11-1 to N.G.), the National Institutes of Health (R35GM130376 to R.R. and R35GM118336 to R.S.M.), a USC SBIR/STTR Planning Award (to R.R.), and the Light Microscopy Facility (LMF) of the Center for Molecular and Cellular Bioengineering of the Technical University Dresden. The LMF is supported by a European Fund for Regional Development. A.P.F. and D.K.P. were supported by an EMBO Advanced Collaboration Grant (#10377) for the implementation of FCS experiments at the Max Planck Institute of Molecular Cell Biology and the Center for Molecular and Cellular Bioengineering of the Technical University Dresden. **Author contributions:** Conceptualization: C.F., L.L., B.H.C., S.C., D.K.P., R.S.M., R.R., and N.G. Methodology: C.F., L.L., B.H.C., S.C., T.-P.C., D.K.P., and R.M. Software: S.C. and C.F. Validation: C.F., L.L., A.D.D., D.V., and M.K. Formal analysis: C.F., L.L., T.-P.C., R.M., B.H.C., and S.C. Investigation: C.F., L.L., T.-P.C., R.M., B.H.C., B.M., V.S., A.D.D., D.V., M.K., A.P.F., W.J.G., J.F.K.-S., and D.K.P. Resources: Data curation: N.G. and C.F. Writing—original draft: C.F. and N.G. Writing—review and editing: C.F., L.L., S.C., T.-P.C., D.K.P., R.S.M., R.R., and N.G. Visualization: C.F., B.H.C., S.C., T.-P.C., and N.G. Supervision: N.G. Project administration: N.G. Funding acquisition: N.G., R.S.M., R.R., and D.K.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. Materials availability: Reagents used in this study, including reporter constructs and fly lines, are available from one of the corresponding authors, N.G. ngompel@uni-bonn.de. Data and code availability: Image datasets and code used for image processing and analysis in this paper are available at <https://doi.org/10.5061/dryad.mgqkn99c8>. The wing registration software is available at <https://doi.org/10.5281/zenodo.17702271>.

Submitted 7 June 2025

Accepted 4 December 2025

Published 7 January 2026

10.1126/sciadv.adz5902

## Quantitative modulation of a spatial enhancer through the biophysical properties of a transcription factor binding site

Conrad Fallon, Liucong Ling, Brendon H. Cooper, Stefano Ceolin, Tsu-Pei Chiu, Raktim Mitra, Bettina Mühling, Vani Srinivasan, Ayse Damla Durmaz, Daniel Veselinovic, Mahek Kothari, Athanasios Panagiotis Fylaktakis, William J. Glassford, Judith F. Kribelbauer-Swietek, Dimitrios K. Papadopoulos, Richard S. Mann, Remo Rohs, and Nicolas Gompel

*Sci. Adv.* **12** (2), eadz5902. DOI: 10.1126/sciadv.adz5902

### View the article online

<https://www.science.org/doi/10.1126/sciadv.adz5902>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).