

## DEVELOPMENTAL BIOLOGY

Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer

Yann Le Poul<sup>1\*</sup>, Yaqun Xin<sup>1\*</sup>, Liucong Ling<sup>1</sup>, Bettina Mühling<sup>1</sup>, Rita Jaenichen<sup>1</sup>, David Hörl<sup>2</sup>, David Bunk<sup>2</sup>, Hartmann Harz<sup>2</sup>, Heinrich Leonhardt<sup>2</sup>, Yingfei Wang<sup>3</sup>, Elena Osipova<sup>1</sup>, Mariam Museridze<sup>1</sup>, Deepak Dharmadhikari<sup>1</sup>, Eamonn Murphy<sup>1</sup>, Remo Rohs<sup>3</sup>, Stephan Preibisch<sup>4,5</sup>, Benjamin Prud'homme<sup>6†</sup>, Nicolas Gompel<sup>1†</sup>

Developmental enhancers control the expression of genes prefiguring morphological patterns. The activity of an enhancer varies among cells of a tissue, but collectively, expression levels in individual cells constitute a spatial pattern of gene expression. How the spatial and quantitative regulatory information is encoded in an enhancer sequence is elusive. To link spatial pattern and activity levels of an enhancer, we used systematic mutations of the *yellow spot* enhancer, active in developing *Drosophila* wings, and tested their effect in a reporter assay. Moreover, we developed an analytic framework based on the comprehensive quantification of spatial reporter activity. We show that the quantitative enhancer activity results from densely packed regulatory information along the sequence, and that a complex interplay between activators and multiple tiers of repressors carves the spatial pattern. Our results shed light on how an enhancer reads and integrates trans-regulatory landscape information to encode a spatial quantitative pattern.

## INTRODUCTION

Enhancers constitute a particular class of cis-regulatory elements that control in which cells a gene is transcribed, when, and at which rate (1, 2). Notably, enhancers play a central role during development in plants and animals (3), generating patterns of gene expression that delineate embryonic territories and prefigure future forms (4). How the information determining these patterns is encoded in a developmental enhancer has therefore been at the center of attention for several decades. Enhancers integrate spatial information from transcription factors (TFs) bound to them, and the number, affinity, and arrangement of TF binding sites (TFBSs) in the enhancer sequence are relevant to the enhancer spatial activity [reviewed in (5)]. However, the logic of TFBS organization that determines a spatial pattern is not sufficiently understood to reliably design a functional synthetic enhancer driving correct expression levels (6, 7).

The study of developmental enhancers has been polarized by two conceptions of gene expression patterns. Until recently, most studies have referred to enhancer activities in qualitative terms exclusively, where the notion of spatial pattern evokes discrete and relatively homogeneous domains of gene expression (8). With the rise of genomics from the early 2000s, it has become possible to precisely measure gene expression and, by extension, enhancer activity. How-

ever, whether it is measured in a given tissue or in single cells, this quantification of gene expression is done at the expense of losing spatial information [e.g., (9–11)], with few exceptions [e.g., (12, 13)]. It is nevertheless critical to appreciate that the overall levels and the spatial pattern of activity in a given tissue are intrinsically linked. Therefore, to understand how a spatial pattern of gene expression is encoded in the sequence of an enhancer, it is necessary to measure quantitative variation of gene expression in space in the tissue where the enhancer is active. Leading this endeavor, recent studies have quantified spatial enhancer activity but without considering the pattern itself as a quantitative object (13–18).

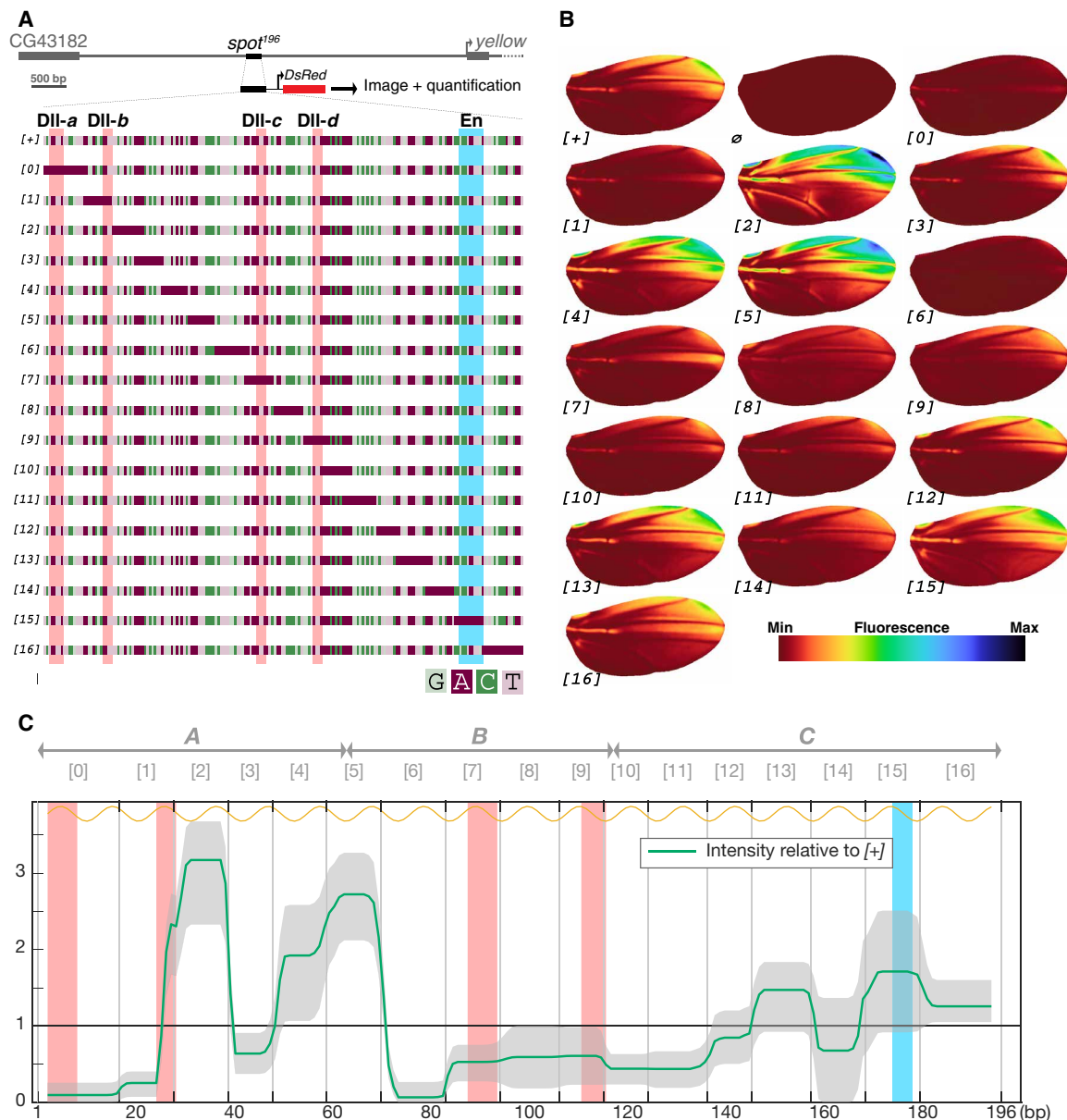
To pursue this effort of measuring quantitative variation in spatial gene expression, we have analyzed the structure and the functional logic of a compact *Drosophila* enhancer sequence with quantitative measurements of its spatial activity in fly wings. The so-called *spot*<sup>196</sup> enhancer, from the *yellow* gene of the fruit fly *Drosophila biarmipes*, drives a patterned gene expression in pupal wings with heterogeneous expression levels among cells (19–21). The *spot*<sup>196</sup> enhancer sequence contains at least four TFBSs for the activator Distal-less (Dll) and at least one TFBS for the repressor Engrailed (En) (Fig. 1A) (19, 20). Together, these inputs were considered to be sufficient to explain the spatial activity of *spot*<sup>196</sup> in the wing, with activation in the distal region and repression in the posterior wing compartment (19, 20). Grafting TFBSs for these factors on a naïve sequence in their native configuration, however, proved insufficient to produce regulatory activity in wings. This prompted us to dissect the *spot*<sup>196</sup> element further to identify what determines its regulatory activity, considering simultaneously spatial pattern and activity levels.

We first introduced systematic small-scale mutations along the 196 base pairs (bp) of the enhancer sequence to test the necessity of the mutated positions; we then randomized large blocks of the enhancer sequence to test the sufficiency of the remaining intact sequence to drive activity. To assess the activity of each mutant enhancer, we devised a pipeline that uses comprehensive descriptors to quantify variations in reporter activity levels across the wing of *Drosophila melanogaster* transgenic lines. Our quantitative analysis revealed a

<sup>1</sup>Evolutionary Ecology, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. <sup>2</sup>Human Biology and Bioimaging, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. <sup>3</sup>Quantitative and Computational Biology, Departments of Biological Sciences, Chemistry, Physics and Astronomy, and Computer Science, University of Southern California, Los Angeles, CA 90089, USA. <sup>4</sup>Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany. <sup>5</sup>Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. <sup>6</sup>Aix-Marseille Université, CNRS, IBDM, Institut de Biologie du Développement de Marseille, Campus de Luminy Case 907, 13288 Marseille Cedex 9, France.

\*These authors contributed equally to this work.

†Corresponding author. Email: benjamin.prudhomme@univ-amu.fr (B.P.); gompel@bio.lmu.de (N.G.)



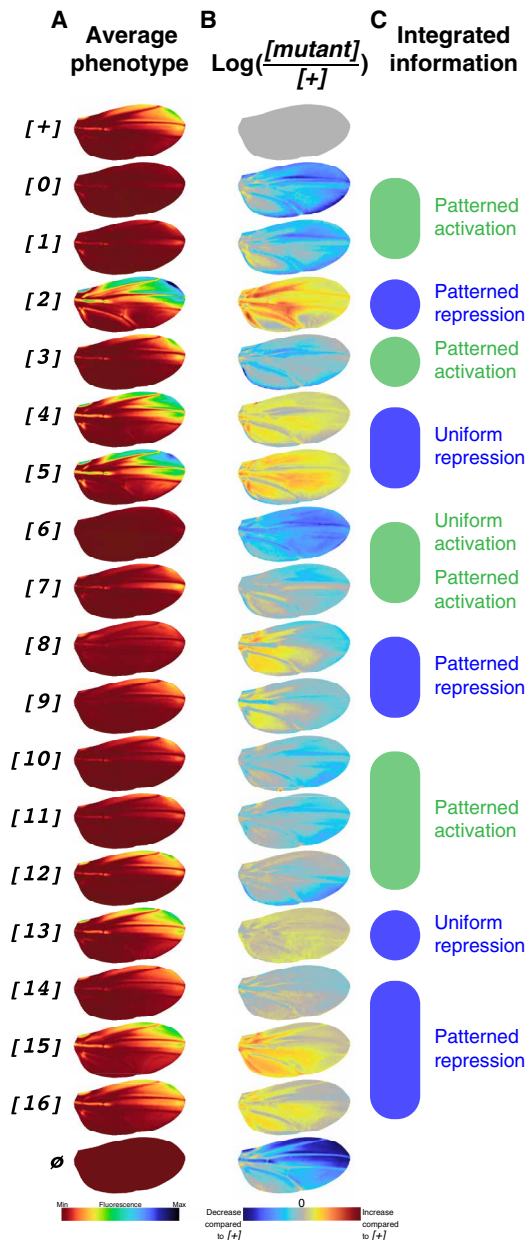
**Fig. 1. A mutational scan of the *D. biarmipes*  $spot^{196}$  enhancer with a quantitative reporter assay.** (A) Wild-type ([+]) and mutant ([0] to [16]) versions of the  $spot^{196}$  enhancer from the *D. biarmipes* *yellow* locus (depicted at the top) were cloned upstream of a DsRed reporter to assay their respective activities in transgenic *D. melanogaster*. Each mutant targets a position of the enhancer, where the native sequence was replaced by an A-tract (color code: light green, guanine; purple, adenine; dark green, cytosine; pink, thymine). Four characterized binding sites for the TF Distal-less (DII-a, DII-b, DII-c, and DII-d) (19) are highlighted in red, and a single binding site for the TF Engrailed (20) is highlighted in blue across all constructs. (B) Average wing reporter expression for each construct depicted in (A) and an empty reporter vector ( $\emptyset$ ). Each wing image is produced from 11 to 77 individual wing images (38 on average; data file S2), aligned onto a unique wing model. The average image is smoothed, and intensity levels are indicated by a colormap. (C) Mutational effect on intensity of activity along the  $spot^{196}$  sequence. The phenotypic effect of each mutation described in (A) along the  $spot^{196}$  sequence (x axis) is plotted as the average level of expression across the wing relative to the wild-type average levels. Shaded gray areas around the curve represent the 95% confidence interval of the average levels per position. “1” on the y axis represents the mean wild-type intensity of reporter expression. The graph shows how each construct departs from the wild-type activity (see Materials and Methods). Mutation positions in constructs [0] to [16] are indicated above the graph. The locations of blocks A, B, and C, analyzed in Fig. 3, are also indicated above the graph. The yellow curve above the graph indicates the helical phasing.

high density of regulatory information, with all mutated positions along the  $spot^{196}$  enhancer sequence contributing significantly to the activity levels. It also outlined an unanticipated regulatory logic for this enhancer, where the spatial pattern in the wing results from a complex interplay between activators and multiple tiers of repressors carving a spatial pattern.

## RESULTS

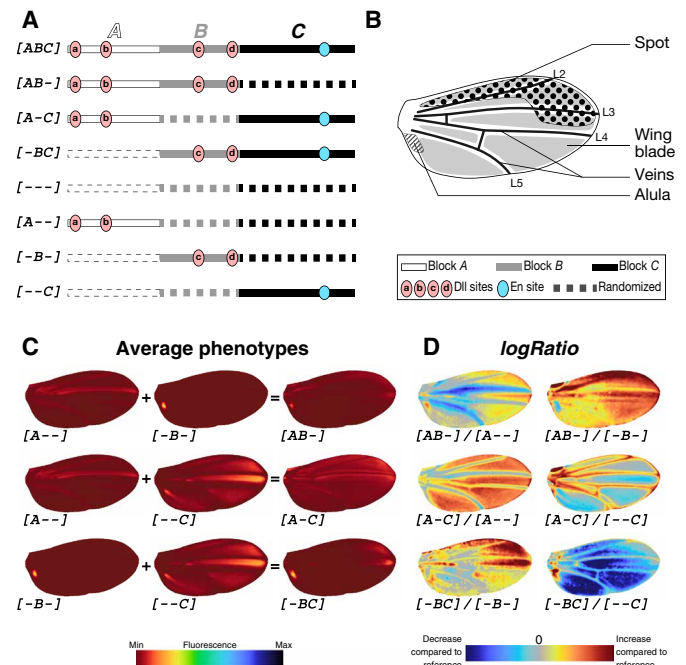
### Regulatory information distributed along the entire $spot^{196}$ enhancer contributes to its quantitative spatial activity in the wing

We first systematically evaluated the potential role of all positions along the  $spot^{196}$  enhancer sequence to produce an activity pattern



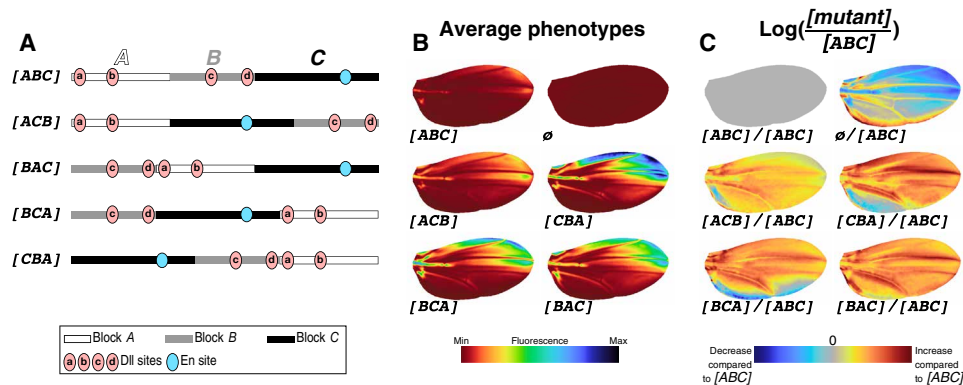
**Fig. 2. Trans-regulatory integration along the *spot*<sup>196</sup> sequence.** (A) Average phenotypes reproduced from Fig. 1B. (B) *logRatio* images [ $\log([mutant]/[+])$ ] for intensity values of each pixel of registered wing images reveal what spatial information is integrated by each position along the enhancer sequence. For instance, a blue region on an image indicates that the enhancer position contains information for activation in this region. When mutated, this enhancer position results in lower activity than *[+]* in this region of the wing. Note that *logRatio* illustrates local changes between *[+]* and mutants far better than image differences (fig. S3) in regions of relatively low activity. (C) Summary of spatial information integrated along the enhancer sequence.

and wild-type levels of gene expression. We generated a series of mutants scanning the element and thereby testing the necessity of short adjacent segments to the enhancer function. Notably, we made no prior assumption (e.g., predicted TFBSs) on the function of the mutated nucleotides. We maximized the disruption of se-



**Fig. 3. Regulatory interactions in the *spot*<sup>196</sup> sequence.** (A) Schematics of constructs with block randomizations. The *spot*<sup>196</sup> sequence was arbitrarily divided into three blocks (A, 63 bp; B, 54 bp; C, 79 bp). In each construct, the sequence of one, two, or all three blocks was randomized. (B) Terminology for parts of the wing where constructs from (A) drive reporter expression. (C) Average phenotypes resulting from constructs in (A). Constructs where single blocks remain indicate the sufficiency of these blocks to promote wing activity: A in the veins, B in the alula, and C at high levels across the wing blade. Constructs with two nonrandomized blocks show the effect of one block on the other. For instance, B is sufficient to suppress the wing blade activation promoted by C, as seen by comparing [-B-], [-C], and [-BC]. Colormap of average phenotypes normalized for all constructs of the block series, including block permutations of Fig. 4B. (D) Block interactions are best visualized with *logRatio* images of construct phenotypes shown in (C). For each *logRatio*, the denominator is the reference construct, and the image shows on a logarithmic scale how much the construct in the numerator changes compared to this reference. For instance,  $\log([-BC]/[-C])$  shows the effect of B on C, a global repression, except in the spot region. Colormap indicates an increase or a decrease of activity compared to the reference (denominator). For an overview of all comparisons, particularly the relative contribution of each block to the entire enhancer activity, see fig. S4 (C to F).

quence information by introducing stretches of 10 to 18 bp (11.5 bp on average) of poly(dA:dT), also known as A-tracts (22), at adjacent positions along the sequence (Fig. 1A). Thus, the sequence of each of the 17 constructs (*spot*<sup>196</sup>[0] to *spot*<sup>196</sup>[16], or [0] to [16] in short; Fig. 1A) is identical to the wild-type *spot*<sup>196</sup> (*[+]* in short), except for one segment where the sequence was replaced by the corresponding number of adenines. These mutations affect the local sequence composition, without changing distances or helical phasing in the rest of the enhancer. We measured activities of each mutant enhancer in the wing of the corresponding reporter construct line of *D. melanogaster*, here used as an experimental recipient for site-specific integration. In brief, for each reporter construct line, we imaged individually around 30 male wings (1 wing per fly) under bright-field and fluorescent light. We detected the venation on the bright-field images of all wings and used it to compare reporter activity across wings. For this, we applied a deformable model to



**Fig. 4. Block permutations scale the activity of the *spot*<sup>196</sup> enhancer.** (A) Schematics of constructs with block permutations. In this series, the same blocks of sequences as in Fig. 3A were permuted. (B) Average phenotypes resulting from constructs in (A). Colormap of average phenotypes normalized for all constructs of the block series, including block randomizations of Fig. 3C and fig. S4B. (C) Average phenotypes in (B) compared to the average phenotype of the wild-type [ABC] (*logRatio*). Note that, in contrast to constructs with randomized blocks (Fig. 3), constructs with block permutations result in near-uniform changes of activity across the wing. Colormap indicates an increase or a decrease of activity compared to the wild-type enhancer [ABC].

warp the fluorescent image of each wing, using landmarks placed along the veins of the corresponding bright-field image and aligning them to a reference venation (see Materials and Methods for details). The resulting dataset is a collection of fluorescence images for which the venation of all specimens is perfectly aligned. These images, represented as the list of fluorescence intensity of all pixels, constitute the basis of all our quantitative dissection. To assess whether the activity driven by a given enhancer sequence significantly differs from any other, wild type or mutant, we used the scores produced by principal components analysis (PCA) that comprehensively summarizes the variation of the pixel intensities across wings. To visualize the reporter activity per line, we used images representing the average activity per pixel (hereafter average phenotype).

The activity of each mutant (Fig. 1B) differs significantly from that of [+], as measured in the PCA space (fig. S1 and data file S1). This means that the activity of each mutant had some features, more or less pronounced, that significantly differentiate its activity from [+], revealing the high density of regulatory information distributed along the sequence of *spot*<sup>196</sup>. The magnitude and direction of the effects, however, vary widely among mutants, ranging from activity levels well above those of [+] to a near-complete loss of activity.

The average activity levels of each mutant construct in the wing relative to the average activity levels of [+] show how effect directions and intensities are distributed along the enhancer sequence (Fig. 1C). This distribution of regulatory information and the magnitude and direction of the effects, including several successions of overexpressing and underexpressing mutants, suggest a more complex enhancer structure than previously thought (20). The density of regulatory information is also reminiscent of what has been found for other enhancers (13, 23, 24).

In principle, the localized mutations we introduced can affect the *spot*<sup>196</sup> enhancer function through nonexclusive molecular mechanisms. Mutations may affect TF-DNA interactions by disrupting TFBS cores or by influencing TF binding at neighboring TFBSs [for instance, by altering DNA shape properties (25, 26)]. A-tract mutations may also influence nucleosome positioning and

thereby the binding of TFs at adjacent sites (27). Not exclusively, because of stacking interactions between adjacent As and Ts, they increase local DNA rigidity (22, 28, 29) and may thereby hinder or modulate TF interactions. These changes in rigidity, which we have evaluated for our mutant series (fig. S2A), may affect TF-TF interactions (fig. S2B). Regardless of the precise molecular mechanisms underlying the mutations we introduced in the *spot*<sup>196</sup> sequence, we wanted to assess how they affect the integration of spatial information along the enhancer sequence.

### An enhancer's view on the wing trans-regulatory landscape revealed by *logRatio* images

We have introduced a spatial visualization of the intensity of effect of a mutation on the enhancer activity. We computed the pixel-wise log of the ratio between two average phenotypes (single mutants over [+]) at every pixel (30), hereafter noted *logRatio*. The advantages of using *logRatio* are detailed in the Supplementary Materials and briefly summarized here. *logRatio* images show visually how much a mutant affects the enhancer activity across the wing proportionally to the local activity level. By contrast, the absolute difference in expression is generally locally linked to the level of expression. Therefore, effects in areas of high activity tend to be much more visible than those in areas of low activity (compare Fig. 2 and fig. S3). *logRatio* images instead represent the local proportional effects and are therefore suitable to reveal the variety of spatial effects of mutations, irrespective of the expression pattern itself.

Depending on how TF integration is modified by a mutant, *logRatio* images can also reflect the distribution of the individual spatial inputs received and integrated along the *spot*<sup>196</sup> sequence. They can be particularly informative when both a TFBS and the spatial distribution of the cognate TF are known, as they shed light on how directly the TF information is integrated. This is the case for En and Dll, for which TFBSs have been previously characterized in *spot*<sup>196</sup> (19, 20). The disruption of an En binding site (Fig. 1, A and B, construct [15]) resulted in a proportional increase of activity in the posterior wing compartment (75%,  $F_{1,124} = 77.8$ ,  $P = 8.8818 \times 10^{-15}$ ). The  $\log([15]/[+])$  image (Fig. 2) shows that mutant [15] proportionally affects the activity mostly in the posterior wing. The effect

correlates with En distribution (20) and is consistent with the repressive effect of its TF. Contrary to what the average phenotypes suggested (Fig. 1C), mutant [16] shows a very similar *logRatio* to that of [15], albeit with only 25% increase in activity. The effect of mutant [16] was barely discernible when considering the variation in the overall fluorescence signal (Fig. 1C), illustrating the power of the *logRatio* analysis to detect local effects in areas of low activity. Mutations that disrupted characterized Dll binding sites (Fig. 1, A and B, constructs [0], [1], [7], and [9]) resulted in strong reduction in reporter expression (90%,  $F_{1,74} = 143.3$ ,  $P = 0$ ; 75%,  $F_{1,78} = 109.3$ ,  $P = 2.2204 \times 10^{-16}$ ; 47%,  $F_{1,107} = 75.4$ ,  $P = 4.8073 \times 10^{-14}$ ; and 39%,  $F_{1,74} = 23.2$ ,  $P = 7.6363 \times 10^{-6}$ , respectively; data file S1). The *logRatio* images for mutants [0], [1], and, to a lesser extent, [7] show a patterned decrease of activity in line with Dll distribution in the wing (Fig. 2) (19), with a proportionally stronger loss of activity toward the distal wing margin. This corroborates previous evidence that Dll binds to these sites. The respective *logRatio* images for segments [0] and [1] correlate with levels of Dll across the wing. This suggests that these sites individually integrate mostly Dll information and do so in a near-linear fashion. Site [9], which produces a relatively different picture with areas showing overexpression, is discussed below. Mutations of Dll sites, however, have nonadditive effects, as mutants [0], [1], [7], and [9] result in a decrease of activity levels by 90, 75, 47, and 39% compared to [+], respectively. This nonadditivity could be explained by a strong cooperative binding of Dll at these sites or, alternatively, by considering that these Dll TFBS are interacting with other sites in the sequence.

In addition, we noted that, despite mutating a Dll TFBS, mutant [9] showed a substantially different *logRatio* than [0] and [1] but similar to [8], with a repressing activity in the posterior wing compartment, proximally, and a distal activation (Fig. 2B). This dual effect could be explained by the disruption of the Dll site along with a distinct TFBS for a posterior repressor. Alternatively, a single TFBS could be used by different TFs with opposite activities. In this regard, we note that the homeodomains of Dll and En have similar binding motifs (31) and could both bind the Dll TFBS disrupted by [9] (and possibly [8]). The posterior repression of En and the distal activation of Dll seem compatible with this hypothesis.

### Unraveling trans-regulatory integration along the *spot*<sup>196</sup> sequence

Following the same approach, we next analyzed the information integrated in other segments. Apart from the known Dll and En TFBSs, the enhancer scan in Fig. 1C identified several segments with strong quantitative effects on the regulatory activity. Between the two pairs of Dll TFBSs, we found an alternation of activating sites [[3] and [6], reducing overall levels by 36% ( $F_{1,69} = 17.6$ ,  $P = 7.8336 \times 10^{-5}$ ) and 93% ( $F_{1,98} = 284.9$ ,  $P = 0$ ) compared to [+], respectively] and strong repressing sites [[2], [4], and [5], with an overall level increase of 3.2-fold ( $F_{1,72} = 511.5$ ,  $P = 0$ ), 1.9-fold ( $F_{1,85} = 103.2$ ,  $P = 2.2204 \times 10^{-16}$ ), and 2.7-fold ( $F_{1,82} = 426.5$ ,  $P = 0$ ) compared to [+], respectively]. Construct [3] proportionally decreases the expression mostly around the wing veins (Fig. 2B), suggesting that this segment integrates information from an activator of the vein regions. We had found a similar activity for this region of *yellow* from another species, *Drosophila pseudoobscura*, where no other wing blade activity concealed it (20). The *logRatio* of mutant [6], with a stronger, more uniform effect than for the other mutants

that repress the activity, suggests a different trans-regulatory integration than Dll sites. We have recently shown that this site regulates the chromatin state of the enhancer (21). Regarding segments with a repressive effect, mutants [4] and [5] result in a fairly uniform relative increase in expression, different from the activity of [2], indicating that the information integrated by these two regions ([2] versus [4] and [5]) likely involves different TFs. Three segments, [6], [0], and [1] (the last two containing previously known Dll binding sites), each decrease the activity levels by 75% or more. Finding additional strong repressive sites ([2], [4], and [5]) with a global effect on the enhancer activity across the wing is also unexpected.

The analysis revealed another activating stretch of the sequence, between 116 and 137 bp, as mutated segments [10] and [11] decreased activity by 56% relative to [+], and showed very similar *logRatios*. Mutant [12] showed a mixed effect, with practically, in absolute terms, no effect in the anterior distal wing quadrant. Last, segments [13], [14], and [15] showed a succession of repressing and activating sites, as we have seen for segments [2] to [6], although with a lower amplitude. Mutant [13] caused an overall increase in activity (1.4-fold relative to [+]) with, proportionally, a uniform effect across the wing (*logRatio*). By contrast, mutant [14] decreased the overall activity by 36%, with a *logRatio* indicating an activating effect in the spot region and a repressive effect in the proximal part of the posterior wing compartment, similarly to mutants [8] and [9] but with lesser effects.

Together, this first dissection, focusing on the necessity of segments for the enhancer activity at the scale of a TFBS, which is typically 10 bp long (32), suggested a much higher density of regulatory information in the *spot*<sup>196</sup> enhancer than previously described (19, 20). The nonadditivity of effects at Dll binding sites, three repressing and four activating and previously unidentified segments distributed in alternation along the enhancer, and the variety of their effects pointed to a complex regulatory logic, involving more (possibly six to eight) factors than just Dll and En. We resorted to a different approach to further probe the regulatory logic of *spot*<sup>196</sup>.

### An interplay of activating and repressing inputs produces a spatial pattern of enhancer activity

The first series of mutations informed us on the contribution of the different elementary components of the *spot*<sup>196</sup> enhancer sequence to its regulatory activity. However, it failed to explain how these components integrated by each segment interact to produce the enhancer activity. To unravel the regulatory logic of this enhancer, it is required to understand not only which segments are sufficient to drive expression but also how elementary components underlying the regulatory logic influence each other. To evaluate the sufficiency of, and interactions between, different segments, we would require to test all possible combinations of mutated segments, namely, a combinatorial dissection. Doing this at the same segment resolution as above is unrealistic, because the number of constructs grows with each permutation. Instead, we used three sequence blocks of comparable sizes in the *spot*<sup>196</sup> enhancer—A, B, and C, defined arbitrarily (Fig. 3A)—and produced constructs where selected blocks were replaced by a randomized sequence (noted “.”). This second series, therefore, consists of eight constructs, including all combinations of one, two, or three randomized blocks, a wild-type [ABC] (which has strictly the same sequence as [+], from the first series), and a fully randomized sequence, [---].

With these constructs, we can track which segments, identified in the first series as necessary for activation in the context of the whole *spot*<sup>196</sup>, are also sufficient to drive activity (table S3; see Fig. 1C for the correspondence between the two series of mutations). Of the three blocks (constructs [A-], [-B-], and [--C]), only block C is sufficient to produce activity levels comparable to those of the wild-type *spot*<sup>196</sup> in the wing blade, although with a different pattern from [ABC] (fig. S4, A to C). Reciprocally, randomizing block C (construct [AB-]) results in a uniform collapse of the activity (fig. S4, A to C). We concluded that the sequence of block C contains information necessary and sufficient to drive high levels of activity in the wing in the context of our experiment. This is particularly interesting because C does not contain previously identified Dll TFBSs or strong activating segments. By contrast, blocks A and B, although they each contain two Dll sites, do not drive wing blade expression. The activating segments in block C revealed in the first dissection, particularly segments [10] and [11], are therefore candidates to drive the main activity of *spot*<sup>196</sup> in the context of these reporter constructs.

Block A alone ([A-]) produces high levels of expression in the veins (fig. S4, A to C). Combined with block C (construct [A-C]), it also increases the vein expression compared to C alone. We concluded that A is sufficient to drive expression in the veins. Segment [3], which proportionally decreased the activity mostly in the veins, could therefore be the necessary counterpart for this activation.

Block B alone drives expression only near the wing hinge, in a region called the alula ([-B-]; Fig. 3, B to D). The first dissection series, however, did not identify a mutated segment within block B that affected specifically the alula.

The necessity of Dll binding sites (in segments [0], [1], [7], and [9]) and of segment [6], and their insufficiency to drive activity in the wing blade in the context of block A alone, block B alone, or blocks A and B combined, suggest that these sites with a strong activation effect function as permissive sites. We next focused on understanding the interplay between repressing and activating sites to shed light on how the *spot*<sup>196</sup> patterning information is built. In the first series of constructs, we identified several strong repressing segments in block A ([2] and [4]) and block B ([5]). Using sufficiency reasoning with the second series of constructs, we further investigated how these inputs interacted with other parts of the enhancer (Fig. 3). These interactions are best visualized with *logRatios*, comparing this time double-block constructs to single-block constructs used as references (Fig. 3D and fig. S4, D to F). Block B has a strong repressive effect on block C throughout the wing, except at the anterior distal tip, where C activity is nearly unchanged [ $\log([-BC]/[-C])$ ; Fig. 3D]. Likewise,  $\log([AB-]/[A-])$  shows that B also represses the vein expression driven by A. Similarly, block A represses the C activity across the wing blade, except in the spot region  $\log([A-C]/[-C])$ . We have seen above that blocks A and B both contain not only strong repressing segments but also known Dll TFBSs. Because both A and B show a repressive effect on block C, except in the spot region, we submit that the apparent patterned activation by Dll may result from its repressive effect on direct repressors of activity, mostly at the wing tip. This indirect activation model would explain the nonadditivity of the individual Dll binding sites observed in the first construct series and why grafting Dll TFBSs on a naïve DNA sequence is not sufficient to create a wing spot pattern. Together, these results outline an unexpectedly complex

regulatory logic that contrasts with the simple model we had initially proposed (19, 20) and involves multiple activators and several tiers of repressors.

### Sequence reorganization affects activity levels of the *spot*<sup>196</sup> enhancer, not its spatial output

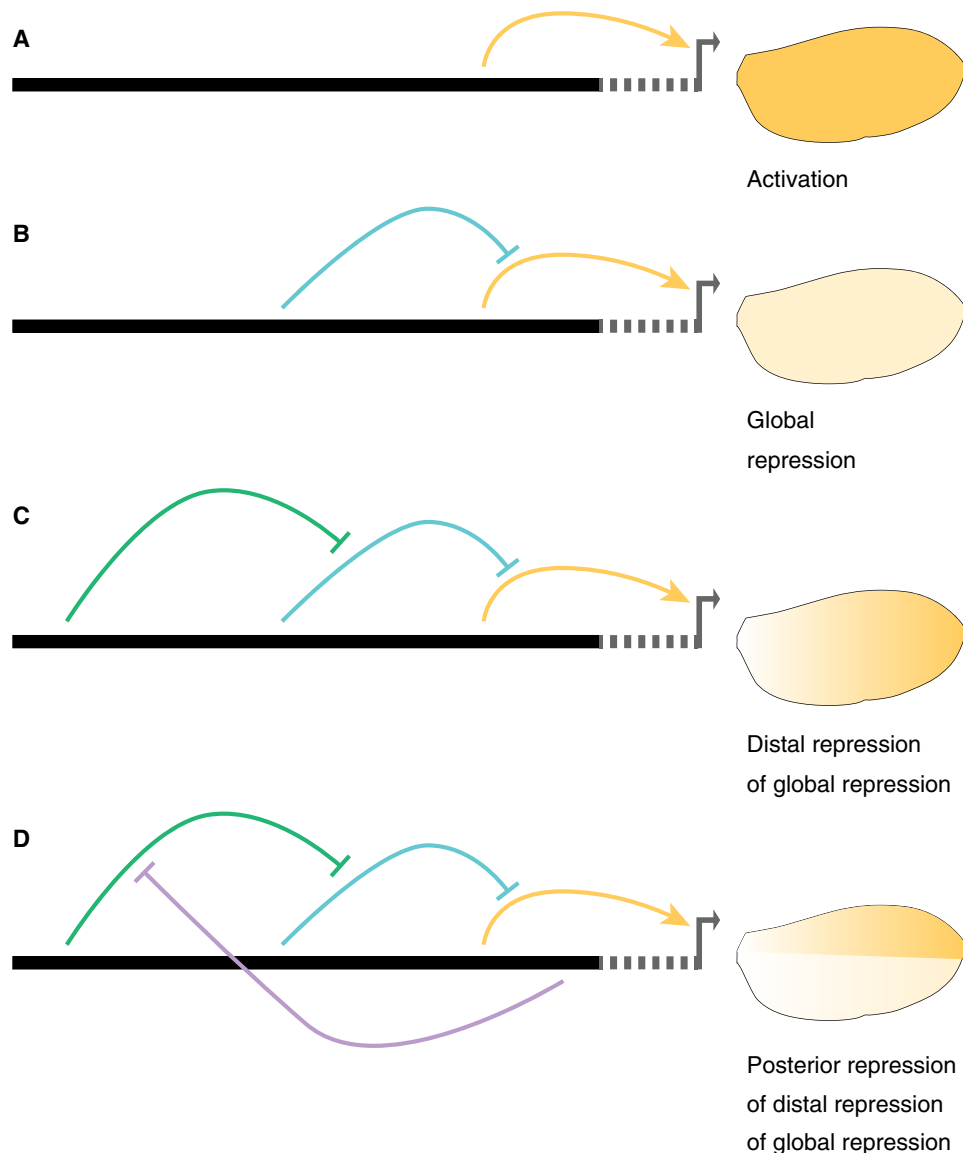
In a final series of experiments, we wondered whether the complex regulatory architecture uncovered by the first two mutant series was sensitive to the organization of the inputs. To test the effect of changes in the organization of enhancer logical elements, we introduced new constructs with permutations of blocks A, B, and C (Fig. 4A). These permutations preserve the entire regulatory content of the enhancer, except at the junction of adjacent blocks where regulatory information may be lost or created. All permutations that we have tested (four of five possible permutations) drive significantly higher levels of expression than the wild-type [ABC] [[ACB]: 2.9-fold ( $F_{1,98} = 191.8, P = 0$ ); [BAC]: 6-fold ( $F_{1,93} = 589.1, P = 0$ ); [BCA]: 5.8-fold ( $F_{1,93} = 589.1, P = 0$ ); [CBA]: 8.4-fold ( $F_{1,93} = 1664.2, P = 0$ ); Fig. 4B] yet with minor effects on the activity distribution proportionally to the wild type (Fig. 4C). We concluded from these experiments that, in terms of pattern, the regulatory output is generally resilient to large-scale rearrangements. As long as all inputs are present in the sequence, the spatial activity is deployed in a similar pattern, yet its quantitative activity is strongly modulated. Because they have little influence on the activity pattern, the rearrangements may not change the nature of the interactions within the enhancer or with the core promoter. Although we would need to challenge this conclusion with additional constructs and blocks with different breakpoints, we speculate that, molecularly, the block randomization perturbs the action of some of the uniformly repressing elements. It highlights the robustness of the enhancer logic to produce a given patterned activity.

## DISCUSSION

With this work, we have set to decipher the regulatory logic of an enhancer, *spot*<sup>196</sup>. The viewpoint presented here is the information that the enhancer integrates along its sequence. Combined with the quantitative measurement of enhancer activity in a tissue, the wing, this information reveals the enhancer regulatory logic and how it reads the wing trans-regulatory environment to encode a spatial pattern. The strength of our arguments stems from the introduction of two complementary aspects of the method (discussed in the following sections): one to combine the assessment of necessity and sufficiency of regulatory information in our analysis and another to compare the spatial activity of enhancer variants (*logRatio*).

### Regulatory necessity and regulatory sufficiency

When dissecting a regulatory element, it is straightforward to assess the necessity of a TFBS or any stretch of the sequence to the activity, by introducing mutations. It is generally more difficult to assess whether the same sequence is sufficient to promote regulatory activity at all, and most enhancer dissections are focusing on necessity analysis [see, for instance, (12, 17, 19, 20, 23, 33–37)]. However, our study shows that, to decipher regulatory logic and eventually design synthetic enhancers, understanding which regulatory components are sufficient to build an enhancer activity is key.



**Fig. 5. A model of the regulatory logic governing the *spot*<sup>196</sup> enhancer.** (A to D) The schematics show step by step how regulatory information and interactions integrated along the enhancer sequence produce a spatial pattern of activity. (A) Three independent inputs, respectively, in blocks A, B, and C promote activity (arrows) in the wing veins, the alula, and the wing blade, as illustrated with average phenotypes of constructs [A-], [B-], and [C-], respectively. Note that activity levels in the wing blade, stemming from block C, match the final levels of the *spot*<sup>196</sup> enhancer activity in the spot region. (B) A first set of repressive inputs suppresses activity in the wing blade (stemming from blocks A and B) and the veins (stemming from block B). The overall combined output of the initial activation and the global repressive inputs is a near-complete loss of activity, except in the alula. (C) A second set of repressive inputs, whose action is localized in the distal wing region, counters the global repression, thereby carving a pattern of distal activity promoted by block C. (D) The distal activity is repressed in the posterior wing compartment, likely through the repressive action of Engrailed, resulting in a final pattern of activity in the spot region.

### A visual tool to compare spatial activities driven by enhancer variants

We introduced a new representation to compare activities between enhancer variants, typically a wild type and a mutant. Proportional effects, or local fold changes, as revealed by *logRatio* produce representations that are independent from the distribution of the reference activity. They also better reflect the distribution of factors in trans and their variations as seen by the enhancer (here, across the wing) than differential comparisons (compare Fig. 2 and fig. S3). Differential comparisons are dominated by regions of high activi-

ties and thereby focusing our attention to the regions of high variation of activity. By contrast, *logRatios* reveal strong effects in regions of low activity that would hardly be visible using differential comparisons, highlighting some cryptic components of the regulatory logic. When additional knowledge about TFBSs and TF distribution will become available, they will also inform us on the contribution of the TF in the regulatory logic. In this respect, the introduction of *logRatios* in our analysis has proven useful and could be adapted to any system where image alignment is possible, such as *Drosophila* blastoderm embryos (38) or developing mouse limbs (39).

### A-tracts did not disrupt the major effect of TF-TF interactions

A-tracts are known to change local conformational properties of DNA. Hence, our A-tract mutations could influence the regulatory logic not only by directly disrupting the information contained in the sequence they replaced but also, indirectly, by introducing more changes than wanted. As an alternative, sequence randomization, however, is more likely to create spurious TFBSs, which is difficult to control for, especially if all the determinants of the enhancer activity are not known. The possible occurrence of undesired and undetected TFBSs would have biased our interpretation of the effect of individual segments and, consequently, of the regulatory logic of the enhancer. The chance that A-tracts introduce new TFBSs in the enhancer sequence is quite low compared to sequence randomization, which is why we favored this mutational approach for the analysis of short, individual segments. However, A-tracts can modify various physical properties of the DNA molecule and, in turn, influence interactions between TFs binding the enhancer. The disruption of a TF-TF interaction due to the introduction of an A-tract between two TFBSs (fig. S2B) would be revealed if mutating a particular segment would have an effect similar to the effect of mutating immediately adjacent flanking segments. We note, however, that we do not have such situation in our dataset. This suggests that the A-tracts we introduced, if anything, only mildly altered TF-TF interactions through changes in the physical properties of *spot*<sup>196</sup>. Instead, we think that the effects of A-tract mutations are mostly due to disrupted TFBSs along the enhancer sequence.

### The regulatory logic underlying *spot*<sup>196</sup> enhancer activity

The main finding of our study is that the *spot*<sup>196</sup> enhancer likely integrates six to eight distinct regulatory inputs, with multiple layers of cross-interactions (Fig. 5). We had previously proposed that the spot pattern resulted from the integration of only two spatial regulators: the activator Dll and the repressor En (19, 20). The regulatory density that we reveal here (Figs. 1C and 2) is reminiscent of what has been found for other enhancers (13, 23, 24). A logical analysis of systematic mutations along the enhancer gives a different status to the factors controlling *spot*<sup>196</sup>. The main levels of *spot*<sup>196</sup> activity across the wing blade seem to result mostly from two unknown activators: one promoting a relatively uniform expression in the wing blade, and another along the veins (Fig. 5A). This activation is, in turn, globally repressed throughout the wing by an unknown repressor whose action masks that of the global activator (Fig. 5B). Upon these first two regulatory layers, the actual spot pattern of activity is carved by two local repressions. A distal repression counteracts the effect of the global repressor in the distal region of the wing (Fig. 5C), but the spatial range of this repression is limited to the anterior wing compartment by another repressor acting across the posterior wing compartment (Fig. 5D). The former local repression could be mediated by Dll itself, a hypothesis compatible with the nonadditive effects of Dll TFBS mutations, whereas the latter is almost certainly due to En. Thus, the pattern of activity results not so much from local activation but from multiple tiers of repressors.

One would expect this complex set of interactions between TFs that bind along the enhancer sequence to be vulnerable to sequence reorganization. We unexpectedly find that shuffling blocks of the sequence resulted in marked changes in activity levels with little effect on the activity pattern. Similarly, many of the mutations still produced a pattern of activity quite similar to the one of [+]. This

suggests that the exact organization of the different inputs and the absence of some of these inputs do not affect the TF-enhancer and TF-TF interactions required for a patterned activity, which here translates mainly to the role of Dll in repressing global repressors and the repressing role of En. The frequency of these interactions, or the interactions with the core promoter, may, however, change significantly upon sequence modifications, affecting transcription rate. In other words, the regulatory logic described above is robust to changes for the production of a spatial pattern but less so for the tuning of enhancer activity levels.

The regulatory logic of this enhancer perhaps reflects the evolutionary steps of the emergence of *spot*<sup>196</sup>. The *spot*<sup>196</sup> element evolved from the co-option of a preexisting *wing blade* enhancer (20). The sequences of this ancestral *wing blade* enhancer and the evolutionary-derived *spot*<sup>196</sup> overlap and share at least one common input (21). This perspective is consistent with the idea that a novel pattern emerged by the progressive evolution of multiple tiers of repression carving a spot pattern from a uniform regulatory activity in the wing blade. To further deconstruct the regulatory logic governing the *spot*<sup>196</sup> enhancer and its evolution, one first task will be to investigate how some of the mutations we introduced affect the activity of a broader fragment containing the entire *spot* activity (and the *wing blade* enhancer), closer to the native context of this enhancer. Another challenging step will be to identify the direct inputs integrated along its sequence. It will also be necessary to characterize their biochemical interactions with DNA and with one another. Ultimately, to fully grasp the enhancer logic will mean to be able to recreate these interactions in a functional synthetic regulatory element.

## MATERIALS AND METHODS

### Fly husbandry

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

### Transgenesis

All reporter constructs were injected as in (19). We used  $\phi$ C31-mediated transgenesis (40) and integrated all constructs at the genomic *attP* site VK00016 (41) on chromosome 2. All transgenic lines were genotyped to ascertain that the enhancer sequence was correct.

### Molecular biology

All 196-bp constructs derived from the *D. biarmipes spot*<sup>196</sup> sequence were synthesized in vitro by a biotech company (Integrated DNA Technologies, Coralville, USA; catalog no. 121416). Table S1 provides a list of all constructs and their sequences. Each construct was cloned by In-Fusion (Takara, Mountain View, USA) in our pRedSA vector [a custom version of the transformation vector pRed H-Stinger (42) with a 284-bp *attB* site for  $\phi$ C31-mediated transgenesis (40) cloned at the Avr II site of pRed H-Stinger]. All constructs in Fig. 1 were cloned by cutting pRedSA with Kpn I and Nhe I and using the following homology arms for In-Fusion cloning: 5'-GAG-CCC GGCGAATT-3' and 5'-GATCCCTCGAGGAGC-3'. Likewise, constructs in Fig. 3 were cloned by cutting pRedSA with Bam HI and Eco RI and using the following homology arms for In-Fusion cloning: 5'-GAGCCC GGCGAATT-3' and 5'-GATCCCTCGAGGAGC-3'.



### Wing preparation and imaging

All transgenic wings imaged in this study were homozygous for the reporter construct. Males were selected at emergence from pupa, a stage that we call “post-emergence,” when their wings are unfolded but still slightly curled. When flies were massively emerging from an amplified stock, we collected every 10 min and froze staged flies at  $-20^{\circ}\text{C}$  until we had reached a sufficient number of flies. In any case, staged flies were processed after a maximum of 48 hours at  $-20^{\circ}\text{C}$ . We dissected a single wing per male. Upon dissection, wings were immediately mounted onto a microscope slide coated with transparent glue (see below) and fixed for 1 hour at room temperature in 4% paraformaldehyde diluted in phosphate-buffered saline–1% Triton X-100 (PBST). Slides with mounted wings were then rinsed in PBST and kept in a PBST bath at  $4^{\circ}\text{C}$  until the next day. Slides were then removed from PBST, and the wings were covered with Vectashield (Vector Laboratories, Burlingame, USA). The samples were then covered with a coverslip. Preparations were stored for a maximum of 48 hours at  $4^{\circ}\text{C}$  until image acquisition.

The glue-coated slides were prepared immediately before wing mounting by dissolving adhesive tape (Tesa brand, tesafilm, ref. 57912) in heptane (two rolls in 100 ml of heptane) and spreading a thin layer of this solution onto a clean microscope slide. Once the heptane had evaporated (under a fume hood), the slide was ready for wing mounting. All wing images were acquired as 16-bit images on a Ti2 Eclipse Nikon microscope equipped with a Nikon 10 $\times$  plan apochromatic lens (numerical aperture, 0.45; Nikon Corporation, Tokyo, Japan) and a pco.edge 5.5 Mpx sCMOS camera (PCO, Kelheim, Germany) under illumination from a Lumencor SOLA SE II light source (Lumencor, Beaverton, OR, USA). Each wing was imaged by tiling and stitching of several z-stacks (z-step, 4  $\mu\text{m}$ ) with 50% overlap between tiles. Each image comprises a fluorescent channel (ET-DSRed filter cube, Chroma Technology Corporation, Bellows Falls, VT, USA) and a bright-field channel (acquired using flat field correction from the Nikon NIS-Elements software throughout), the latter being used for later image alignment. To ensure that fluorescence measurements are comparable between imaging sessions, we used identical settings for the fluorescence light source (100% output), light path, and camera (20-ms exposure time, no active shutter) to achieve comparable fluorescence excitation.

### Z-projection

Stitched three-dimensional (3D) stacks were projected to 2D images for subsequent analysis. The local sharpness average of the bright-field channel was computed for each pixel position in each z-slice, and an index of the slice with the maximum sharpness was recorded and smoothed with a Gaussian kernel ( $\sigma = 5$  px). Both bright-field and fluorescent 2D images were reconstituted by taking the value of the sharpest slice for each pixel.

### Image alignment

Wing images were aligned using the veins as a reference. Fourteen landmarks placed on vein intersections and end points and 26 sliding landmarks equally spaced along the veins were placed on bright-field images using a semi-automatized pipeline. Landmark coordinates on the image were then used to warp with a deformable model (thin plate spline) bright-field and fluorescent images to match the landmarks of an arbitrarily chosen reference wing by the thin plate spline interpolation (43). All wings were then in the same coordinate system, defined by their venation.

### Fluorescent signal description

A transgenic line with an empty reporter vector ( $\emptyset$ ) was used as a proxy to measure noise and tissue autofluorescence. The median raw fluorescent image was computed across all  $\emptyset$  images and used to remove autofluorescence, subtracted from all raw images before the following steps. All variation of fluorescence below the median  $\emptyset$  value was discarded. The DsRed reporter signal was mostly localized in the cell nuclei. We measured the local average fluorescent levels by smoothing fluorescence intensity, through a Gaussian filter ( $\sigma = 8$  px) on the raw 2D fluorescent signal. The  $\sigma$  corresponded roughly to two times the distance between the adjacent nuclei. To lower the memory requirement, images were then subsampled by a factor of 2. We used the 89,735 pixels inside the wings as descriptors of the phenotype for all subsequent analyses.

### Average phenotypes, differences, *logRatio* colormaps, and normalization

Average reporter expression phenotypes were computed as the average smoothed fluorescence intensity at every pixel among all individuals in a given group (tens of individuals from the same transgenic line). The difference between groups was computed as the pixel-wise difference between the average of the groups (fig. S3). *logRatio* between two constructs represents the fold change of a phenotype relative to another and is calculated as the pixel-wise logarithm of the ratio between the two phenotypes. Averages, difference, and *logRatio* images were represented using colors equally spaced in CIELAB perceptual color space (44). With these colormaps, the perceived difference in colors corresponds to the actual difference in signal. Colormaps were spread between the minimal and maximal signals across all averages for average phenotypes. Difference and *logRatio* spread between minus and plus represent the absolute value of all difference for the phenotype differences, with gray colors indicating that the two compared phenotypes are equal.

### Mutation effect direction and intensity

We proposed to represent the necessity of a stretch of the sequence along the enhancer with the activity levels of mutants of this stretch relative to the wild-type ( $[+]$ ) activity. To summarize the overall effect of mutants (overexpression or underexpression), we measured the average level of activity across each wing relative to that of the reference. The reference level was defined as the average level of activity of all  $[+]$  individuals. The value at each position corresponds to the average of all individuals that present a sequence that have an effect on this position. The effect of a mutation is not strictly limited to the mutated bases, because they can also modify properties of DNA of flanking positions (45). To take this effect into account and produce a more realistic and conservative estimation of necessity measure at each position, we weighted the phenotypic contribution of each mutant line to the measure by the strength of the changes they introduce to the DNA shape descriptors at this position. At each position, the phenotype of constructs not affecting the DNA shape descriptors compared to  $[+]$  was not considered. When two mutants modify the DNA shape descriptors at one position, typically near the junction of two adjacent mutations, the effect at this position was computed as the weighted average of the effect of the two mutants, where the weight is the extent of the DNA shape modification relative to the  $[+]$  sequence. DNA shape descriptors were computed by the R package DNashapeR (46).

Notably, with an average of 11.5 bp, our A-tract mutations are somewhat larger than an average eukaryotic TFBS [ $\sim 10$  bp (32)], and each mutation is likely to affect up to two TFBSs. This size represents the limit of regulatory content that we can discriminate in this study.

### PCA and difference significance

The intensity measure is an average of the overall and variable expression across the wing. Hence, mutations causing a different effect on the phenotype can have the same intensity value. To test whether the mutant significantly differs from [+], we used comprehensive and unbiased phenotype descriptors provided by PCA, which removes the correlation between pixel intensities and describes the variation in reporter gene expression. PCA was calculated on the matrix regrouping intensities of all pixels for every individual, of dimensions ( $n_{\text{individuals}} \times n_{\text{pixels}}$  on the wing). The significance of the difference between two constructs considers the multivariate variation of the phenotypes and is tested using multivariate analysis of variance (MANOVA) on all five first components explaining more than 0.5% of the total variance (data file S3).

### Overall expression intensity and significance

The overall expression level was measured for each individual as the average intensity across the wing. This was used to test the significance of overall increase and decrease in expression levels relative to the wild-type levels.

### DNA rigidity scores

A-tracts are runs of consecutive A/T base pair without a TpA step. Stacking interactions and inter-base pair hydrogen bonds in ApA (TpT) or ApT steps of A-tracts lead to conformational rigidity (28). The length of an A-tract directly correlates with increased rigidity (47). To parametrize DNA rigidity at nucleotide resolution, we used A-tract length as a metric. For each position in a given DNA sequence, we find the longest consecutive run of the form  $A_nT_m$  that contains this position (with the requirement of  $n \geq 0$ ,  $m \geq 0$ , and  $n + m \geq 2$ ), and score DNA rigidity at that position using the length of this subsequence. For example, the sequence AATCGCAT will map to the scores 3,3,3,0,0,0,2,2 because AAT and AT are A-tracts of lengths 3 and 2 bp, respectively.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/49/eabe2955/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

- D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: From properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
- M. Levine, Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
- I. S. Peter, E. H. Davidson, *Genomic Control Process: Development and Evolution* (Academic Press, ed. 1, 2015), p. 460.
- S. B. Carroll, From pattern to gene, from gene to pattern. *Int. J. Dev. Biol.* **42**, 305–309 (1998).
- F. Spitz, E. E. M. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
- J. Crocker, A. Tsai, D. L. Stern, A fully synthetic transcriptional platform for a multicellular eukaryote. *Cell Rep.* **18**, 287–296 (2017).
- B. J. Vincent, J. Estrada, A. H. DePace, The appeasement of Doug: A synthetic approach to enhancer biology. *Integr. Biol. (Camb.)* **8**, 475–484 (2016).
- E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Elsevier, 2010).

- 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).
- 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).
- M. Kircher, C. Xiong, B. Martin, M. Schubach, F. Inoue, R. J. A. Bell, J. F. Costello, J. Shendure, N. Ahituv, Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat. Commun.* **10**, 3583 (2019).
- A.-R. Kim, C. Martinez, J. Ionides, A. F. Ramos, M. Z. Ludwig, N. Ogawa, D. H. Sharp, J. Reinitz, Rearrangements of 2.5 kilobases of noncoding DNA from the *Drosophila even-skipped* locus define predictive rules of genomic cis-regulatory logic. *PLoS Genet.* **9**, e1003243 (2013).
- T. Fuqua, J. Jordan, M. E. van Breugel, A. Halavaty, C. Tischer, P. Polidoro, N. Abe, A. Tsai, R. S. Mann, D. L. Stern, J. Crocker, Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* **587**, 235–239 (2020).
- J. Dufourt, A. Trullo, J. Hunter, C. Fernandez, J. Lazaro, M. Dejean, L. Morales, S. Nait-Amer, K. N. Schulz, M. M. Harrison, C. Favard, O. Radulescu, M. Lagha, Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat. Commun.* **9**, 5194 (2018).
- J. Crocker, D. L. Stern, Functional regulatory evolution outside of the minimal *even-skipped* stripe 2 enhancer. *Development* **144**, 3095–3101 (2017).
- 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).
- 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).
- L. Bentovim, T. T. Harden, A. H. DePace, Transcriptional precision and accuracy in development: From measurements to models and mechanisms. *Development* **144**, 3855–3866 (2017).
- L. Arnoult, K. F. Y. Su, D. Manoel, C. Minervino, J. Magriña, 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).
- Y. Xin, Y. Le Poul, L. Ling, M. Museridze, B. Mühling, R. Jaenichen, E. Osipova, N. Gompel, Enhancer evolutionary co-option through shared chromatin accessibility input. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20636–20644 (2020).
- S. Neidle, *Principles of Nucleic Acid Structure* (Academic Press, 2010), p. 302.
- C. I. Swanson, N. C. Evans, S. Barolo, Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359–370 (2010).
- E. Z. Kvon, Y. Zhu, G. Kelman, C. S. Novak, I. Plajzer-Frick, M. Kato, T. H. Garvin, Q. Pham, N. Harrington, R. D. Hunter, J. Godoy, E. M. Meky, J. A. Akiyama, V. Afzal, S. Tran, F. Escande, B. Gilbert-Dussardier, N. Jean-Marçais, S. Hudaiberdiev, I. Ovcharenko, M. B. Dobbs, C. A. Gurnett, S. Manouvrier-Hanu, F. Petit, A. Visel, D. E. Dickel, L. A. Pennacchio, Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. *Cell* **180**, 1262–1271.e15 (2020).
- M. Slattery, T. Zhou, L. Yang, A. C. Dantas Machado, R. Gordan, R. Rohs, Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
- N. Abe, I. Dror, L. Yang, M. Slattery, T. Zhou, H. J. Bussemaker, R. Rohs, R. S. Mann, Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307–318 (2015).
- I. Barozzi, M. Simonatto, S. Bonifacio, L. Yang, R. Rohs, S. Ghisletti, G. Natoli, Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell* **54**, 844–857 (2014).
- H. C. Nelson, J. T. Finch, B. F. Luisi, A. Klug, The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221–226 (1987).
- B. Suter, G. Schnappauf, F. Thoma, Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters *in vivo*. *Nucleic Acids Res.* **28**, 4083–4089 (2000).
- M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- L. J. Zhu, R. G. Christensen, M. Kazemian, C. J. Hull, M. S. Enameh, M. D. Basciotta, J. A. Brasefield, C. Zhu, Y. Asriyan, D. S. Lapointe, S. Sinha, S. A. Wolfe, M. H. Brodsky, FlyFactorSurvey: A database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39**, D111–D117 (2011).

32. A. J. Stewart, S. Hannehalli, J. B. Plotkin, Why transcription factor binding sites are ten nucleotides long. *Genetics* **192**, 973–985 (2012).
33. D. N. Arnosti, S. Barolo, M. Levine, S. Small, The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205–214 (1996).
34. E. K. Farley, K. M. Olson, W. Zhang, D. S. Rokhsar, M. S. Levine, Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6508–6513 (2016).
35. V. Bertrand, C. Hudson, D. Caillol, C. Popovici, P. Lemaire, Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* **115**, 615–627 (2003).
36. D. Thanos, T. Maniatis, Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).
37. C. I. Swanson, D. B. Schwimmer, S. Barolo, Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196 (2011).
38. C. C. Fowlkes, C. L. L. Hendriks, S. V. E. Keränen, G. H. Weber, O. Rübél, M.-Y. Huang, S. Chatoor, A. H. De Pace, L. Simirenko, C. Henriquez, A. Beaton, R. Weizmann, S. Celniker, B. Hamann, D. W. Knowles, M. D. Biggin, M. B. Eisen, J. Malik, A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* **133**, 364–374 (2008).
39. N. Martínez-Abadías, R. Mateu, M. Niksic, L. Russo, J. Sharpe, Geometric morphometrics on gene expression patterns within phenotypes: A case example on limb development. *Syst. Biol.* **65**, 194–211 (2016).
40. A. C. Groth, M. Fish, R. Nusse, M. P. Calos, Construction of transgenic *Drosophila* by using the site-specific integrase from phage  $\phi$ C31. *Genetics* **166**, 1775–1782 (2004).
41. K. J. T. Venken, Y. He, R. A. Hoskins, H. J. Bellen, P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**, 1747–1751 (2006).
42. S. Barolo, B. Castro, J. W. Posakony, New *Drosophila* transgenic reporters: Insulated P-element vectors expressing fast-maturing RFP. *Biotechniques* **36**, 436–442 (2004).
43. M. F. Hutchinson, Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.* **9**, 385–403 (1995).
44. E. C. Carter, J. D. Schanda, R. Hirschler, S. Jost, M. R. Luo, M. Melgosa, Y. Ohno, M. R. Pointer, D. C. Rich, F. Viénot, L. Whitehead, J. H. Wold, *Colorimetry* (CIE Central Bureau, ed. 4, 2018).
45. T. Zhou, L. Yang, Y. Lu, I. Dror, A. C. Dantas Machado, T. Ghane, R. Di Felice, R. Rohs, DNASHape: A method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* **41**, W56–W62 (2013).
46. T.-P. Chiu, F. Comoglio, T. Zhou, L. Yang, R. Paro, R. Rohs, DNASHapeR: An R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics* **32**, 1211–1213 (2016).
47. R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, B. Honig, The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253 (2009).

#### Acknowledgments

**Funding:** This work was supported by funds from the Ludwig Maximilian University of Munich, the Human Frontiers Science Program (program grant RGP0021/2018 to N.G., S.P., and R.R.), the Deutsche Forschungsgemeinschaft (grants INST 86/1783-1 LAGG and GO 2495/5-1 to N.G. and SPP 2202 to H.L. and H.H.), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013/ERC Grant Agreement no. 615789 to B.P.), and the NIH (grant R35GM130376 to R.R.). Y.X. was supported by a fellowship from the China Scholarship Council (fellowship 201506990003). L.L. was supported by a DFG fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). M.M. and D.D. are recipients of fellowships from the German Academic Exchange Service (DAAD). E.M. was supported by the Amgen Scholar program of the LMU. **Author contributions:** Y.L.P.: conceptualization, methodology, software, validation, formal analysis, data curation, writing—original draft, and visualization; Y.X.: validation, investigation, formal analysis, and data curation; L.L.: investigation and formal analysis; B.M.: investigation; R.J.: investigation; D.H.: software and data curation; D.B.: software and data curation; H.H.: methodology and supervision; H.L.: supervision; Y.W.: methodology, software, and formal analysis; E.O.: investigation; M.M.: investigation and formal analysis; D.D.: investigation and formal analysis; E.M.: investigation and formal analysis; R.R.: methodology, supervision, and funding acquisition; S.P.: software, supervision, and funding acquisition; B.P.: conceptualization, writing—original draft, and funding acquisition; N.G.: conceptualization, validation, writing—original draft, visualization, supervision, project administration, and funding acquisition. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 12 August 2020

Accepted 20 October 2020

Published 2 December 2020

10.1126/sciadv.abe2955

**Citation:** Y. Le Poul, Y. Xin, L. Ling, B. Mühlhling, R. Jaenichen, D. Hörl, 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).

## Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer

Yann Le Poul, Yaqun Xin, Liucong Ling, Bettina Mühling, Rita Jaenichen, David Hörl, David Bunk, Hartmann Harz, Heinrich Leonhardt, Yingfei Wang, Elena Osipova, Mariam Museridze, Deepak Dharmadhikari, Eamonn Murphy, Remo Rohs, Stephan Preibisch, Benjamin Prud'homme and Nicolas Gompel

*Sci Adv* 6 (49), eabe2955.  
DOI: 10.1126/sciadv.abe2955

ARTICLE TOOLS	<a href="http://advances.sciencemag.org/content/6/49/eabe2955">http://advances.sciencemag.org/content/6/49/eabe2955</a>
SUPPLEMENTARY MATERIALS	<a href="http://advances.sciencemag.org/content/suppl/2020/11/30/6.49.eabe2955.DC1">http://advances.sciencemag.org/content/suppl/2020/11/30/6.49.eabe2955.DC1</a>
REFERENCES	This article cites 43 articles, 10 of which you can access for free <a href="http://advances.sciencemag.org/content/6/49/eabe2955#BIBL">http://advances.sciencemag.org/content/6/49/eabe2955#BIBL</a>
PERMISSIONS	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

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 © 2020 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).