Quantitative analysis of regulatory information distribution in a developmental locus of *Drosophila*

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)
an der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Mariam Museridze

München 2025

Diese Dissertation wurde angefertigt unter der Leitung von Prof. Dr. Nicolas Gompel im Bereich von Fakultät für Biologie an der Ludwig-Maximilians-Universität München.

Erstgutachter: Prof. Dr. Nicolas Gompel

Zweitgutachterin: Dr. Tamara Mikeladze-Dvali

Tag der Einreichung: 09.04.2025

Tag der mündlichen Prüfung: 21.07.2025

Declaration

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

Basel, den 01.11.2025 .....

Mariam Museridze

Erklärung

Hiermit erkläre ich,

☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

Basel, den 01.11.2025 .....

Mariam Museridze

Table of Contents

List of Publications

1. Xin, Yaqun, Yann Le Poul, Liucong Ling, **Mariam Museridze**, Bettina Mühling, Rita Jaenichen, Elena Osipova, and Nicolas Gompel. 2020. “Enhancer Evolutionary Co-Option through Shared Chromatin Accessibility Input.” Proceedings of the National Academy of Sciences of the United States of America 117 (34): 20636–44

2. **Mariam Museridze**, Stefano Ceolin, Bettina Mühling, Srishti Ramanathan, Olga Barmina, Pallavi Santhi Sekhar, and Nicolas Gompel. 2024. “Entangled and Non-Modular Enhancer Sequences Producing Independent Spatial Activities.” Science Advances 10 (47): eadr9856.

3. Yann Le Poul, Yann, 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. 2020. “Regulatory Encoding of Quantitative Variation in Spatial Activity of a Drosophila Enhancer.” Science Advances 6 (49): eabe2955.

Declaration of Contributions

1. Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility

Yaqun Xin*, Yann Le Poul*, Liucong Ling, **Mariam Museridze**, Bettina Mühling, Rita Jaenichen, Elena Osipova, Nicolas Gompel

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES | Vol. 117, No. 34 | 25 August 2020

DOI: 10.1073/pnas.2004003117

Contributions:

Mariam Museridze: Contributed to experiments and formal analysis; designed and cloned several constructs; assisted with imaging.

Other authors' contributions:

Yaqun Xin (co-first author*): Contributed to the design of the research; performed all imaging experiments and most cloning experiment; analyzed the data with other authors.

Yann Le Poul (co-first author*): Developed the pipeline for wing imaging quantification; analyzed the data, wrote the paper with Nicolas Gompel.

Liucong Ling: Designed, performed and analyzed ATAC-seq experiments.

Rita Jaenichen: Performed some of the cloning experiments.

Bettina Mühling: Produced all transgenic lines.

Elena Osipova: Contributed to the sample preparation for imaging experiments.

Nicolas Gompel: Designed and supervised the research, analyzed the data with other authors, wrote the paper with Yann Le Poul, and acquired funding.

Supervisor's signature:

Nicolas Gompel 01.11.2025

2. Entangled and non-modular enhancer sequences producing independent spatial activities

Mariam Museridze*, Stefano Ceolin*, Bettina Mühling, Srishti Ramanathan, Olga Barmina, Pallavi Santhi Sekhar, and Nicolas Gompel

SCIENCE ADVANCES | Vol 10, Issue 47 | 20 Nov 2024
DOI: 10.1126/sciadv.adr9856

Contributions:

Mariam Museridze (co-first author*): conceptualized and designed the project and experiments together with Nicolas Gompel and Stefano Ceolin. Mariam Museridze performed most of the experiments and imaging, and carried out part of the analysis. Mariam Museridze wrote the manuscript together with Nicolas Gompel.

Stefano Ceolin (co-first author*): contributed to the project and experimental design, developed image registration and analysis pipelines. Stefano Ceolin performed part of the imaging together with other authors. He also provided invaluable feedback on the manuscript.

Other authors' contributions:

Bettina Mühling: Produced all transgenic lines.

Srishti Ramanathan: Performed some of the cloning experiments and did imaging.

Olga Barmina: Performed abdominal transcriptome experiments.

Pallavi Santhi Sekhar: Contributed to the sample preparation and imaging.

Nicolas Gompel: Designed and supervised the research, analyzed the data with other authors, wrote the paper with Mariam Museridze, and acquired funding.

Co-first author's signature

Mariam Museridze 01.11.2025

Co-first author's signature:

Stefano Ceolin 01.11.2025

Supervisor's signature:

Nicolas Gompel 01.11.2025

3. 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, Nicolas Gompel

SCIENCE ADVANCES | Vol 6, Issue 49 | 2 Dec 2020
DOI: 10.1126/sciadv.abe2955

Contributions:
Mariam Museridze: Contributed to experiments, formal analysis and assisted with imaging.

Other authors' contributions:
Yann Le Poul (co-first author*): conceptualization, methodology, software, validation, formal analysis, data curation, writing—original draft, and visualization.
Yaqun Xin (co-first author*): validation, investigation, formal analysis, and data curation.
Liucong Ling: performed some of the cloning and imaging experiments and contributed to the interpretation of the data.
Bettina Mühling: investigation.
Rita Jaenichen: investigation.
David Hörl: software and data curation.
David Bunk: software and data curation.
Hartmann Harz: methodology and supervision.
Heinrich Leonhardt: supervision.
Yingfei Wang: methodology, software, and formal analysis.
Elena Osipova: investigation.
Deepak Dharmadhikari: investigation and formal analysis.
Eamonn Murphy: investigation and formal analysis.
Remo Rohs: methodology, supervision, and funding acquisition.
Stephan Preibisch: software, supervision, and funding acquisition.
Benjamin Prud'homme: conceptualization, writing of the original draft, and funding acquisition.
Nicolas Gompel: conceptualization, validation, writing of the manuscript, visualization, supervision, project administration, and funding acquisition.

Supervisor's signature:
Nicolas Gompel 01.11.2025

List of Abbreviations

3C	chromosome conformation capture
3D	three-dimensional
AP-1	activator protein 1
ATAC-seq	assay for transposase-accessible chromatin using sequencing
ChIP-seq	chromatin immunoprecipitation sequencing
DNA	deoxyribonucleic acid
E-P	enhancer-promoter
ENCODE	encyclopedia of DNA elements
Eve	even-skipped
eRNA	enhancer RNA
Fgf8	fibroblast growth factor 8
GRN	gene regulatory network
H3K27ac	histone H3 lysine 27 acetylation
H3K4me1	histone H3 lysine 4 monomethylation
IFN-β	interferon beta
IgH	immunoglobulin heavy chain
Med1	mediator complex subunit 1
MPRA	massively parallel reporter assay
Oct4	octamer-binding transcription factor 4
OCT4	octamer-binding transcription factor 4
Pax6	paired box gene 6
Pol II	RNA polymerase II
RNA	ribonucleic acid
Rho	rhomboid
Shh	sonic hedgehog
Sox2	SRY-box transcription factor 2
SV40	simian virus 40
Svb	shavenbaby
TAD	topologically associating domain
TF	transcription factor
TFBS	transcription factor binding site

Twi	twist
UAS	upstream activating sequence
Ubx	Ultrabithorax
Y	yellow
ZRS	zone of polarizing activity regulatory sequence

Abstract

The development of a multicellular organism relies on the precise spatio-temporal expression of regulatory genes, which assign cell identities and define spatial domains for future traits. This expression is controlled by enhancers—regulatory DNA sequences that determine when, where, and how much of a gene is expressed. Enhancers are traditionally described as small, modular elements that activate transcription regardless of their orientation or distance from the target promoter. This definition is largely based on qualitative assessments that focus on an enhancer's ability to drive a specific spatial expression pattern in reporter assays. However, such an approach often overlooks DNA fragments that do not alter the spatial pattern but may still influence transcript abundance, leading to a disproportionate focus on minimal enhancer elements necessary for spatial specificity. Studying regulatory elements in a broader context and using quantitative approaches is important, as our perception of enhancers and their structure strongly influences our understanding of the mechanisms that shape the regulatory landscape and drive the evolution of morphological phenotypes.

Using the regulatory locus of the *yellow* gene in *Drosophila*, this dissertation investigates how regulatory information is distributed and how spatial and quantitative aspects of gene expression are encoded in the DNA sequence. By systematically dissecting the regulatory region and employing an image registration pipeline, this study quantitatively analyzes enhancer activities responsible for gene expression in the wings and body of the fly. The findings reveal that *yellow* enhancers span a large genomic region, extensively overlap, and are densely packed with regulatory information. Moreover, the tested enhancers exhibit high levels of pleiotropy, challenging the conventional view of enhancer modularity. These observations have significant implications for both enhancer function and evolution.

Zusammenfassung

Die Entwicklung eines multizelligen Organismus beruht auf der präzisen räumlich-zeitlichen Expression regulatorischer Gene, die Zellidentitäten zuweisen und räumliche Domänen für zukünftige Merkmale definieren. Diese Expression wird durch Enhancer kontrolliert – regulatorische DNA-Sequenzen, die bestimmen, wann, wo und in welchem Ausmaß ein Gen exprimiert wird. Enhancer werden traditionell als kleine, modulare Elemente beschrieben, die die Transkription unabhängig von ihrer Orientierung oder Entfernung zum Zielpromotor aktivieren. Diese Definition basiert größtenteils auf qualitativen Bewertungen, die sich auf die Fähigkeit eines Enhancers konzentrieren, in Reporter-Assays ein spezifisches räumliches Expressionsmuster zu erzeugen. Ein solcher Ansatz übersieht jedoch häufig DNA-Fragmente, die das räumliche Muster nicht verändern, aber dennoch die Transkriptmenge beeinflussen können, was zu einer unverhältnismäßigen Fokussierung auf minimale Enhancer-Elemente führt, die für die räumliche Spezifität notwendig sind. Die Untersuchung regulatorischer Elemente in einem breiteren Kontext und unter Anwendung quantitativer Methoden ist wichtig, da unsere Wahrnehmung von Enhancern und ihrer Struktur unser Verständnis der Mechanismen, die die regulatorische Landschaft formen und die Evolution morphologischer Merkmale antreiben, maßgeblich beeinflusst.

Anhand des regulatorischen Lokus des *yellow*-Gens in *Drosophila* untersucht diese Dissertation, wie regulatorische Information verteilt ist und wie räumliche sowie quantitative Aspekte der Genexpression in der DNA-Sequenz kodiert sind. Durch die systematische Zerlegung der regulatorischen Region und den Einsatz einer Bildregistrierungs-Pipeline analysiert diese Studie quantitativ die Enhancer-Aktivitäten, die für die Genexpression in den Flügeln und dem Körper der Fliege verantwortlich sind. Die Ergebnisse zeigen, dass *yellow*-Enhancer eine große genomische Region umfassen, sich stark überlappen und dicht mit regulatorischer Information gepackt sind. Darüber hinaus weisen die getesteten Enhancer ein hohes Maß an Pleiotropie auf, was die konventionelle Vorstellung von Enhancer-Modularität infrage stellt. Diese Beobachtungen haben weitreichende Implikationen sowohl für die Funktion als auch für die Evolution von Enhancern.

Introduction

1. Mechanisms of Enhancer Function

1.1 Enhancers modulate gene expression.

Development depends on precise expression of regulatory genes

Multicellular organisms start from a single cell, and through developmental process, grow and shape into their adult form. After fertilization, the zygote begins to divide, differentiating into distinct cell types and increasing the spatial complexity of the organism. In each species this process is highly reproducible and follows a conserved scheme. Development is orchestrated by a set of regulatory genes that define the identity, fate, and function of each cell. In the early stages of development, these genes first outline the general body plan of the organism, and with each subsequent step, define more spatial domains, creating a blueprint for future traits. For example, in *Drosophila* embryos, the presence of factors like Bicoid and Hunchback in one part of the embryo defines the anterior-posterior axis, while the presence of Dorsal and Toll outlines dorso-ventral regions [1], [2]. Later *Hox* genes specify different body segments such as the head, mouthparts, thoracic, and abdominal segments [3]–[5]. Often, the expression patterns of these regulatory genes prefigure the shape of the trait they regulate. The expression patterns of segmentation genes such as *Krueppel*, *hairy*, and *giant* in the blastoderm of a *Drosophila* embryo resemble the segmented pattern observed on the larval cuticle 24 hours later. Disruptions in the expression of patterning genes can result in developmental disorders and diseases. For example, a *Krueppel* mutation is lethal, and embryos homozygous for this mutation lack thoracic and anterior abdominal segments (Fig. 1), [1], [6]. The effects of mutations in patterning genes have also been observed in humans. The *Hox* genes encode transcription factors that play a fundamental role in embryonic morphogenesis and are functionally conserved throughout the animal kingdom [7], [8]. Hox proteins are essential for determining primary and secondary body axes. Mutations in *Hox* genes have been linked to various human disorders, including skeletal and limb malformations, neurodevelopmental disorders, and craniofacial abnormalities. Additionally, because *Hox* genes regulate cellular proliferation and growth control, their mutations have also been associated with an increased risk of developing cancers [9].

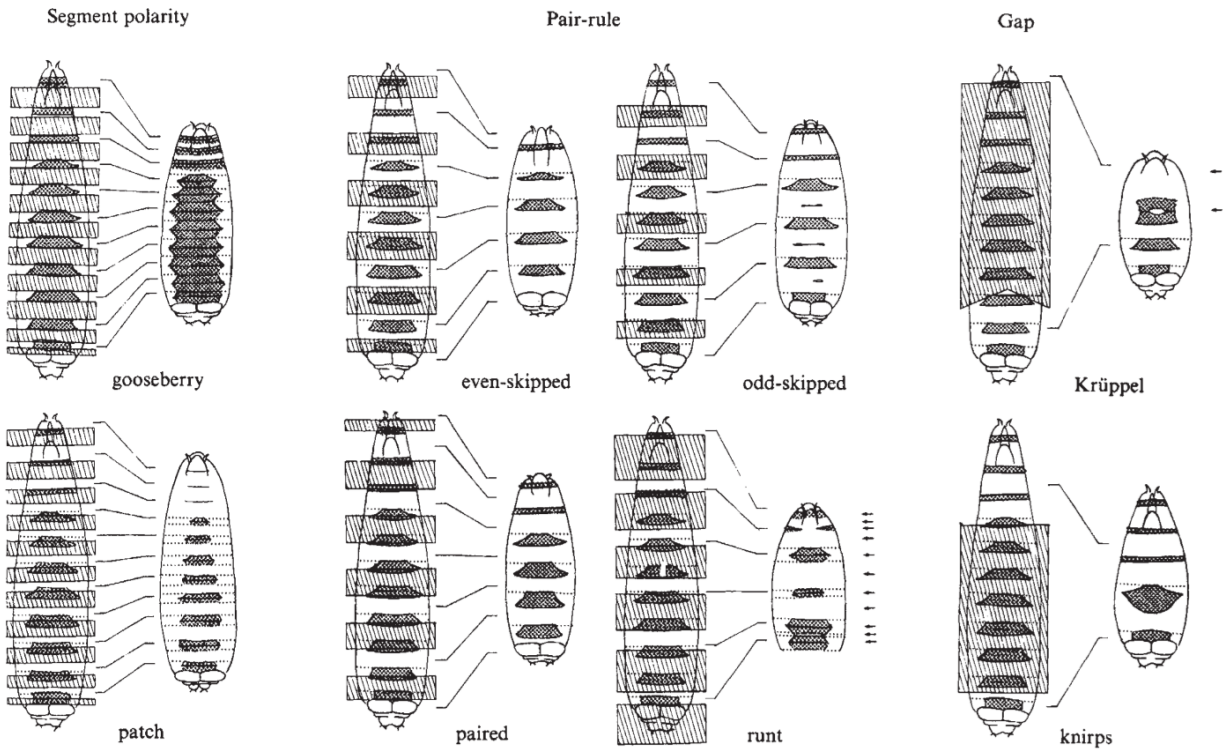

Figure 1. Phenotypes of *Drosophila* mutant larvae lacking different patterning genes. The expression domains of each gene are indicated by rectangles in the larva on the left. Mutant larvae on the right lack specific patterning genes, resulting in the loss of corresponding segments. This figure is reused from C. Nüsslein-Volhard & E. Wieschaus, 1980, with permission from the publisher.

Enhancers are regulatory sequences modulating gene expression

In 1953 the uncovering of DNA's structure marked the beginning of a new era in biology. The genetic code was deciphered, technologies advanced, and within the next two decades, scientists were able to amplify, clone, and sequence genomic DNA. It soon became clear that the complexity of an organism was not determined by the number of genes, but rather by the variety of gene regulatory mechanisms. The first steps in understanding gene regulation were taken with the discovery of the classic bacterial operon, which included a repressor and an activator located upstream of the transcription start site [10], [11]. However, the transcription regulation of eukaryotic genes mostly remained a mystery for another two decades. Then, in 1981, Banerji et al. demonstrated that a piece of DNA from the genome of the simian virus 40 (SV40) had the capacity to alter gene expression [12]. When cloned together with the hemoglobin β-chain gene from the rabbit and tested in HeLa cells, this 72 bp long fragment enhanced the transcription of the gene by 200 times. Similar elements were soon discovered in the genomes of the polyoma, bovine papilloma, and Moloney sarcoma viruses, and later in the DNA of the mouse immunoglobulin constant region locus [13]. These genomic elements, termed enhancers, were shown to activate gene transcription irrespective of their orientation and location relative to the transcription start site (Fig. 2), [14].

Enhancers regulate quantitative and spatio-temporal aspects of gene expression

Not only were enhancers modulating gene expression, but they were doing so in a tissue-specific manner. Polyoma virus showed higher efficiency in mouse cells and SV40 virus induced stronger expression in primate cells [15]. Similarly research done on the immunoglobulin heavy chain (*IgH*) enhancer identified sequence motifs essential for the cell type-dependent activation and determined that binding of the lymphoid-specific factors were major determinants of such activity [16]. During development, enhancer specificity allows precise activation of regulatory gene expression in both space and time.

Enhancers regulate transcription levels, that is the rate of transcription initiation from a core promoter by RNA Pol II. The production of the correct amount of transcript for a given gene is crucial for the proper functioning of an organism. Changes in transcript abundance can lead to developmental disorders. For example, mutations in the *ZRS* enhancer, which regulates the *Sonic hedgehog* (*Shh*) gene, can result in various limb defects, including preaxial polydactyly type 2,

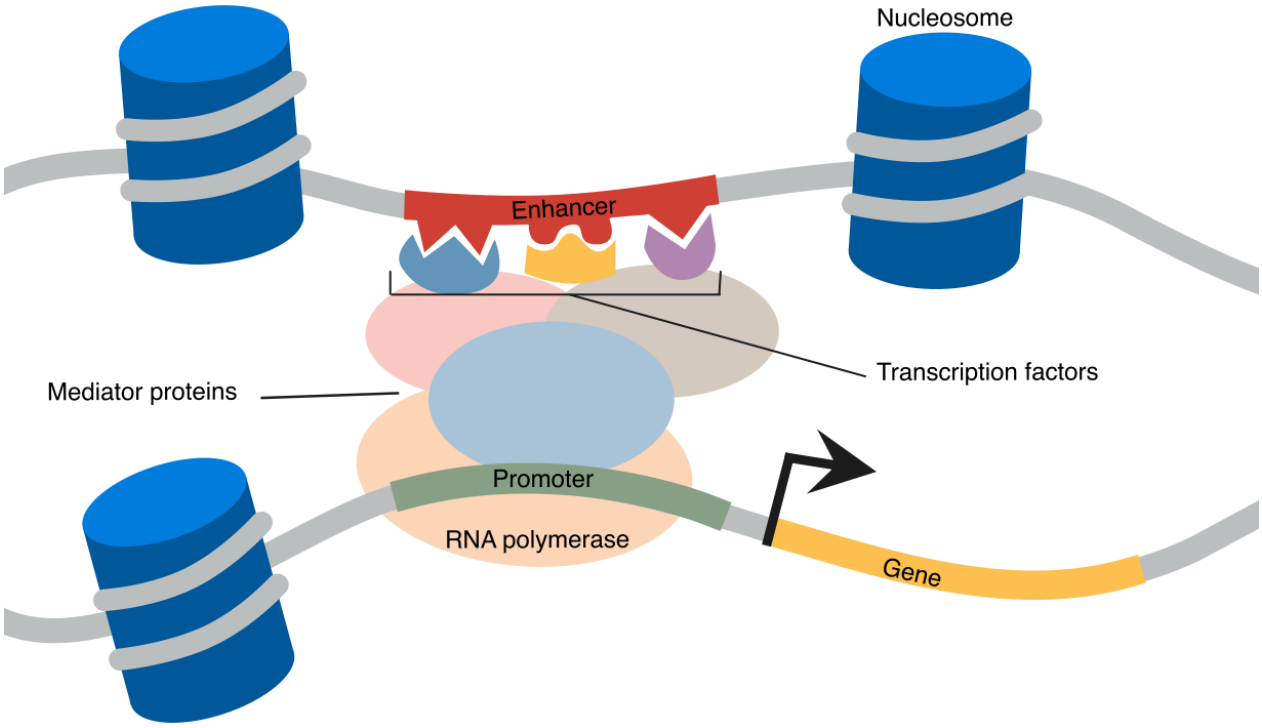

Figure 2. Schematic representation of a regulatory locus. Regulatory elements are located in accessible chromatin regions, allowing transcription factors to bind to DNA. This binding leads to the recruitment of mediator proteins and RNA polymerase, initiating transcription.

triphalangeal thumb polysyndactyly, syndactyly type 4, and Werner mesomelic syndrome [17], [18].

1.2 The role of transcription factors

Enhancers act as a platform that interacts with transcription factors

Enhancer sequences contain DNA motifs recognized by transcription factors (TFs), a group of proteins involved in transcribing DNA into RNA [19]–[21]. TFs contain a DNA-binding domain that recognizes and binds to specific target sequences in the genome, as well as other domains that mediate interactions with various proteins or small molecules. The interaction between TFs and transcription factor binding sites (TFBSs) depends on shape complementarity and the ability to establish electrostatic, hydrophobic, or hydrogen bond interactions. Binding can also occur at nonspecific or low-affinity sites, although they are often nonfunctional. Functional binding, on the other hand, takes place in regulatory regions where multiple binding sites are present. This process is followed by the recruitment of coactivators and mediator complex components, resulting in enhancer-promoter looping, which ultimately leads to transcriptional activation.

Enhancers are estimated to be regulated by 5-6 TFs that can act as both activators and repressors. A combination of such tissue-specific factors will define regions of enhancer activity. A classic example is the regulation of the *even-skipped* (*eve*) gene in *Drosophila*. *eve* is a segmentation gene that plays a crucial role in the embryonic establishment of the antero-posterior axis and is expressed in a series of seven stripes, defining the body plan of a fly. The enhancer driving the expression of its second stripe is one of the most studied enhancers and is regulated by four different TFs that bind to 12 sites distributed over the length of the enhancer. Two activators, Bicoid and Hunchback, drive the expression of *eve* in the entire anterior half of the embryo; however, localized repressors, Giant and Krueppel, restrict its expression within the stripe 2 domain [22].

Transcription factors work synergistically

TFs often do not act in isolation but exhibit cooperativity, significantly enhancing their regulatory impact on gene expression. This cooperativity can manifest through several mechanisms, one of which is the simultaneous binding of multiple TFs to adjacent DNA sites within an enhancer region [23], [24]. Such collective binding can lead to a synergistic effect on transcriptional activation by facilitating the recruitment and binding of additional cofactors or

by altering the chromatin structure to make it more accessible for transcriptional machinery [25].

In addition to enhancing activation levels, cooperativity can modify the binding preferences of TFs. Sequence motifs to which TFs bind are 8-10 bp long and in long DNA sequences would occur often. Cooperative binding can increase specificity in a cell-type-specific manner. For example, during early *Drosophila* embryogenesis, Twist has a preference for sites co-bound by Zelda, Snail, and Dorsal, but at later stages, its binding correlates with binding sites for Tinman and Chorion factor 2 [26]. Direct protein-protein interactions also play an important role and can affect specificity, depending on whether the factor is acting alone or in a pair [27]. For example, the transcription factor Fos can bind to DNA as a homodimer toward the minor groove or as a heterodimer with Jun toward the major groove [28]. The differential binding of these dimers is driven by DNA bending in opposite directions and is believed to lead to opposing transcriptional effects.

TFs provide cell-type-specific instructions for enhancers, allowing them to act at the right time and place and to correctly respond to environmental changes. However, not only the type and combination of TFs can modulate gene expression. Aspects like binding site affinity, orientation, or distribution can also affect transcription initiation. The role of these parameters in gene expression is called enhancer grammar and will be discussed in a later chapter.

1.3 DNA accessibility

The role of chromatin in gene regulation

Inside the nucleus of eukaryotic cells, DNA is not just floating freely. Instead, it is carefully organized and packaged to fit within the limited space. This organization is achieved by compacting the genome into chromatin, a dynamic complex of DNA, proteins, and RNA that plays a crucial role in regulating gene expression, replication, and repair. The structural unit of chromatin is called a nucleosome, which consists of a DNA molecule wrapped around histone proteins, forming a bead-like structure (Fig. 2). The occupancy and topological organization of nucleosomes determine how easily other proteins can interact with chromatinized DNA [29]–[31]. The topological organization of nucleosomes is not uniform throughout the genome. In heterochromatin, a transcriptionally inactive and densely structured form of chromatin, nucleosomes are tightly packed, restricting access to regulatory proteins. In contrast, euchromatin is more loosely packed and actively participates in transcription [32], [33]. Chemical modification of histones, which involves the addition or removal of chemical groups

to histone proteins, can alter how DNA is packaged and accessed. These modifications, such as methylation, acetylation, and phosphorylation, can either loosen or tighten chromatin structure, thereby influencing whether genes are actively expressed or kept silent [34].

Enhancer activity requires accessible DNA

While enhancers function through the binding of TFs, they alone cannot guarantee gene expression. DNA accessibility is a fundamental requirement for enhancer function, as it allows TFs and other regulatory proteins to access specific DNA sequences necessary for initiating transcription. The chromatin structure at enhancers is less compact, with reduced nucleosome occupancy and specific histone modifications that promote an open chromatin state, thus facilitating TF binding [35]. Only 2-3% of the total DNA sequence in the genome is accessible, yet it captures more than 90% of regions bound by TFs [33]. The accessibility landscape is not static and responds to regulatory inputs. Nucleosome occupancy can range from closed to accessible. Between these two states is permissive chromatin, which is sufficiently dynamic to allow TFs to establish a locally open chromatin conformation [36], [37]. This regulation is bidirectional: TFs compete with histones to modulate local DNA access, and in turn, the cell-type-specific accessibility landscape, regulated by other factors, such as pioneer factors, promotes TF binding [33], [38]–[41].

Enhancers self-regulate accessibility using pioneer factors

Pioneer factors are a unique class of TFs that play a pivotal role in chromatin dynamics. They can bind specific sites on nucleosomal DNA. Thereby, they facilitate binding of other TFs by promoting the transition between closed and open states of chromatin —a crucial step for lineage-specific gene expression during development [42], [43]. For example, the pioneer factor Zelda is essential for zygotic gene activation in *Drosophila*. Embryos lacking Zelda are defective and fail to activate genes essential for blastoderm formation [44]. Pioneer factors have a high affinity for closed chromatin and can bind their target sites on nucleosomes. After binding, they induce chromatin decompaction and partial DNA unwrapping, such as the binding of OCT4, which has been shown to result in the unwrapping of 25 base pairs [45]. They not only contribute to chromatin remodeling but also play a role in the demethylation of mammalian enhancers, increasing their activity and gene expression [46]. The local regulation of DNA accessibility is an essential feature of enhancers that allows precise cell type-specific regulation of gene expression, and the evolution of binding sites for pioneer factors has been shown to contribute to the evolution of novel traits, as developed in the following section.

Modification of accessibility plays an important role in regulatory evolution

Evolution of regulatory regions can occur through the acquisition or loss of binding sites de novo or within existing enhancers. These changes include not only elements encoding spatial information, like tissue-specific TFs, but also binding sites for factors that regulate DNA accessibility. Comparing regulatory sequences between *Drosophila melanogaster* and *Drosophila erecta* has shown differences in the distribution of motifs for both patterning factors like Bicoid and the pioneer factor Zelda, which induces open chromatin [47]. Changes in the accessibility profile can be a step leading to the evolution of novel morphological traits. For example, some *Drosophila* species have a pigmentation pattern on the wing that has been lost and gained multiple times during evolution [48]. It has been demonstrated that a regulatory site identified in *melanogaster*, a spotless species, ensures local opening of chromatin that uncovers a repressor binding site, suppressing expression in the spot region [49]. Mutations can occur anywhere in the DNA, but the tissue-specific accessibility landscape of the genome determines which binding sites can be functionally active and which will remain silent.

1.4 Genome three-dimensional architecture

Enhancers loop to the promoter to transfer regulatory information

Various aspects of enhancer function have been described over the years, including the binding of TFs, nucleosome depletion, the presence of histone post-translational modifications, and even the transcription of enhancers. Nevertheless, we still do not fully understand the precise mechanism by which enhancers interact with their target genes. It has been determined that before an enhancer can activate gene expression, it needs to establish a connection with its cognate promoter. This connection provides the means for transferring information necessary for transcription initiation. Numerous studies have explored a variety of possible mechanisms, which have been summarized and discussed in detail [50]–[52]. These mechanisms may include the transfer of protein complexes from enhancers to promoters, the movement of TFs that help recruit co-activators to the promoter, the establishment of local concentrations of crucial proteins near the promoter through clustering or phase separation, and enhancer-mediated post-translational modification of promoter-bound histones.

Genome's three-dimensional conformation can facilitate enhancer-promoter interactions

Enhancers can activate different promoters over large distances, skipping other regulatory elements that might reside in between. The precise mechanism of the search for a cognate promoter is not known, but there are multiple factors that might contribute to enhancer-

promoter (E-P) specificity. The presence of specific and compatible sequence motifs in enhancers and promoters might be necessary for establishing the connection. It has been proposed that housekeeping and developmental genes have distinct sequence motifs that only allow interactions within these groups [53]. Another mechanism for ensuring specificity might come from promoter repression through reduced accessibility, which prevents unwanted activation. The 3D genome structure has also been proposed as another mechanism regulating E-P interactions. The spatial organization of the genome brings certain regions into close proximity, through topologically associating domains (TADs) that enable enhancers to interact more frequently with their target promoters or other regulatory sequences [54], [55]. Additionally, TADs can act as insulators, restricting interactions with other genomic regions. The three-dimensional organization of DNA provides an additional mechanism for regulating enhancer activity and shaping the genome's regulatory landscape by determining which elements can interact. For example, a study by Keough et al. analyzed genome conformation across 241 mammalian genomes and identified conserved loci that evolved at an accelerated rate in the human lineage. They demonstrated that changes in 3D genome organization can drive species-specific traits by altering regulatory interactions between genes and enhancers [56].

1.5 Enhancers and Evolution

Changes in cis drive morphological evolution

After the discovery of DNA structure and advancement in molecular biology techniques, scientists sought to understand what kind of changes in genes are responsible for the evolution of morphological diversity. In 1975, King and Wilson investigated genetic differences between species that had been thoroughly compared at the organismal level and substantially differed in anatomy and way of life, namely humans and chimpanzees [57]. They compared the amino acid sequences of 44 proteins and found only a 1% divergence between the two species. This high similarity contrasted sharply with the significant phenotypic differences between humans and chimpanzees. Further studies also showed high conservation of developmental genes across animal taxa. For example, the *Pax6* gene, a master regulator for eye morphogenesis, is highly conserved in coding sequence and function among humans, mice, and even *Drosophila* [58], [59]. It was becoming clear that substantial differences in the morphology of various species had to stem from regulatory changes. Morphological evolution seems to be driven by changes in regulatory sequences rather than protein sequences; i.e., through cis-regulatory elements rather than trans-acting factors.

Gene regulatory networks are modified through changes in cis

A gene regulatory network (GRN) is a collection of molecular regulators that interact with each other to govern gene expression and control a variety of cellular functions such as cell growth, tissue patterning, cell migration, morphogenesis, differentiation, or cell death. The fundamental unit of such networks is a link between a TF and a *cis*-regulatory element. A single factor can bind to and activate a high number of genes. The androgen receptor, for example, has been shown to bind to *cis*-regulatory elements of 172 genes [60], and the transcription factor Twist has been demonstrated to target 500 genes [61]. Global alterations in such vast networks can disrupt crucial developmental processes. Deleterious mutations in the protein-coding sequences of factors like Twist will affect the expression of hundreds of genes and may not be tolerated by the organism. Changes in the *cis*-regulatory elements, on the other hand, can modify the network without affecting a vast number of its members. Mutations in enhancers can create or disrupt binding sites, thus creating or removing links between different transcription factors, a process often referred to as network rewiring. In this scenario, it is possible to reshape a regulatory network without jeopardizing the survival of the organism, but yet potentially contributing to the evolution of novel traits.

Enhancer modularity lowers the fitness cost of mutations

One of the main reasons why enhancers are thought to be the biggest contributors to morphological evolution is their modular nature [62]. One of the distinct features that has been demonstrated since the early days of their discovery is the ability to drive specific and localized gene expression. For a gene that is active in different tissues or at different time points, each activity can be regulated by a distinct modular enhancer. For example, the pigmentation gene *yellow* is expressed in various tissues in *Drosophila* and is regulated by multiple enhancers that drive expression in the body (head, thorax, abdomen), wings, wing veins, and bristles. Modularity has strong evolutionary implications, as it provides regulatory flexibility, which may be essential for the evolution of novel traits. Deleterious mutations occurring within a modular enhancer will affect only a limited population of cells, while activities in other tissues, regulated by different enhancers, remain unaffected. Such compartmentalization reduces the fitness cost of mutation by having a limited effect on gene activity and giving the organism higher chances of survival. By contrast, a mutation in a gene's protein-coding sequence that alters its function will impact all cells in which the gene is expressed. Depending on the functional importance of the gene, such changes might not be tolerated. For example, the gene

wingless plays a crucial role in signaling during multiple developmental processes, including segment polarity and wing development; mutations in its coding region are lethal. However, a small deletion in its *cis*-regulatory region results in a viable adult fly that lacks wings [63]. This illustrates a broader principle in developmental biology: genes involved in fundamental processes often exhibit complex spatio-temporal expression patterns, regulated by multiple enhancers for tissue- and stage-specific control. Such regulatory complexity enables mutations in individual enhancers to have localized effects without disrupting essential gene functions across the entire organism.

2. Enhancer Architecture

The structure of regulatory sequences reflects their function

The way we define enhancers is not just a technical matter - it fundamentally shapes how we understand their function and evolution. Their definition is heavily influenced by the methods used to study them. The types of DNA fragments tested for reporter activity, the experimental setup, and the parameters used for computational analysis shape how we represent enhancers and, as a result, how we think about their evolutionary potential. In this section, I will discuss the current representation of enhancers and their architecture at different scales, from how regulatory information is distributed in the genome to specific loci and clusters of binding sites. I will explore how various methods influence our perception, what parameters should be considered when investigating enhancer function, and what assumptions we can make regarding the evolutionary mechanisms that shape the regulatory landscape.

2.1 Genomic Location

The difficulty to predict the localization of regulatory sequence

Genome size varies greatly among organisms and does not directly correlate with the number of genes or organismal complexity. However, as genomes increase in size, the proportion of coding sequence to non-coding DNA changes drastically. In viruses, prokaryotes, and smaller eukaryotes, non-coding DNA occupies 15%-20% of the genome. By contrast, in genomes larger than 100 Mb, around 90%-98% of the genome consists of non-coding sequences [64]. In human and mouse genomes, only 1.5% encodes proteins. Interestingly, however, 5% of these genomes are under purifying selection [65], suggesting the importance of additional features that might include untranslated regions, non-protein-coding genes, chromosomal structural elements, and regulatory sequences such as enhancers. Multiple studies have shown an association between

large intergenic sequences and the complexity of gene expression [66], [67]. Analyses of non-coding sequences in *C. elegans* and *D. melanogaster* revealed that developmental genes with complex functions are flanked by longer intergenic sequences compared to housekeeping genes [68]. Another study found that neural genes expressed in multiple tissues have longer intergenic regions compared to neural genes expressed only in a few tissues [67]. It is hard to estimate the proportion of the genome with regulatory function, as enhancer prediction is associated with numerous difficulties. While protein-coding genes have well-defined structures and conserved features, such as start and stop codons, enhancers lack a general language for their sequences. Regulatory elements can be located upstream or downstream of genes, or within introns, and can act on their target genes either proximally or from great distances. All these factors make it difficult to identify enhancers and their target genes.

Features associated with enhancers are used for their identification

Various characteristics are used for the genome-wide identification of regulatory elements. Enhancers are associated with accessible chromatin, specific epigenetic marks, and transcription factor binding sites. Advances in DNA sequencing technologies have facilitated the development of genome-wide methods for identifying putative enhancers. Specific histone modifications, such as H3K27ac and H3K4me1, are hallmarks of active and primed enhancers, and can be used to identify regulatory sequences in the genome. The Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is a technique used in genomics to assess chromatin accessibility. It helps identify open regions in the DNA where transcription factors and other regulatory proteins can bind, indicating active areas of gene regulation. This method utilizes a transposase enzyme that inserts sequencing adapters into open chromatin regions. The DNA is then sequenced, allowing researchers to map these accessible areas across the genome. Another way to localize putative enhancers is with Chromatin Immunoprecipitation sequencing (ChIP-seq). It is a powerful method used to analyze protein interactions with DNA and can be used to globally map binding sites for any protein of interest.

Additionally, Chromosome Conformation Capture (3C) assays facilitate analysis of the genome's three-dimensional architecture and identification of interacting sequences. This method involves crosslinking, ligation, and subsequent sequencing of fragments that are proximate in three-dimensional space. Because enhancers interact with their target promoters and other enhancers, a genomic map of these interactions can help identify functional elements.

In addition to sequencing-based methods, other high-throughput approaches like Massively Parallel Reporter Assays (MPRA) have been developed to experimentally evaluate the

regulatory potential of genomic sequences. This method allows researchers to test thousands of fragments of DNA to see how each element affects gene expression, helping to identify and characterize functional regulatory elements.

While methods for enhancer prediction continue to improve, it's crucial to recognize that these approaches are highly sensitive to the experimental design and arbitrarily set parameters during analysis [50]. A study comparing nine different MPRA methods highlighted that design differences, such as integrated versus episomal reporters, the 5′ versus 3′ location of the enhancer, and the size of the tested fragment, all influence the results [69]. 3C methods vary in their capture radius, significantly influencing the reported contact probability between interacting sequences. If the enhancer-promoter interaction radius is significantly smaller than the capture radius, it will lead to an overestimation of interaction probabilities. If it is much larger, some enhancer-promoter interactions will be missed. Analysis pipelines can introduce biases in enhancer prediction from sequencing-based methods due to sensitivity to various parameters. Algorithms need to filter out noise and determine active peaks. Setting high thresholds during peak calling might miss weaker signals, while low thresholds might include too much noise. Machine learning approaches for predicting enhancers depend on the training dataset and reflect biases in the results. Most putative enhancers remain experimentally untested, making it difficult to predict their activity or functional relevance in the genome.

The genomic location of regulatory regions is often conserved

Although enhancer sequences are not always conserved across species, their genomic location often is [48], [70]–[72]. A comparison of the regulatory loci of the *shavenbaby* gene between *D. melanogaster* and *D. virilis*, species that diverged 40 million years ago, revealed no clear orthologs and very little sequence conservation. However, functional analysis of the *D. virilis* regulatory region showed that five out of six enhancers were located in the same relative positions and drove similar expression patterns, despite their lack of sequence conservation [72]. Similarly, a study of the *yellow* regulatory region across six *Drosophila* species found that while the local arrangement of enhancers varied significantly, they were all mapped within two regulatory regions adjacent to the *yellow* transcription start site [73]. This conservation suggests that enhancer function may require specific distance to the promoter. Genomic location may also be influenced by the global cell type-specific chromatin architecture. As enhancers have been shown to loop to their target promoters and interact with other *cis*-regulatory elements, the three-dimensional conformation of DNA likely plays a crucial role in facilitating these interactions [74]. It has been demonstrated that the characteristics of inter-chromosomal

contacts are largely conserved across vertebrate 3D genomes. This includes conserved TAD boundaries and similar gene-gene contact profiles among related species. Regions that differ in 3D organization have been associated with species-specific gene expression [75], [76]. This suggests that genome architecture can define the regions where regulatory information resides and can constrain the evolution of regulatory elements outside of permissible loci.

Some genomic regions accumulate regulatory elements

Is regulatory information distributed evenly across the genome? Some studies point to genomic hotspots densely packed with regulatory elements. A series of genome-wide sequencing-based studies introduced the concept of "super-enhancers." These are classes of regulatory regions characterized by high levels of activator binding and chromatin modifications. Initially identified in mouse embryonic stem cells, super-enhancers were marked by strong binding of the master regulators Oct4, Sox2, and Nanog, clustering within 12.5 kb of each other, and showing enrichment of Med1 [77]. Subsequently, they have been identified in a variety of cell types based on the enrichment of binding sites for master regulators near genes that determine cell fate. Super-enhancers are associated with pluripotency genes, exhibit strong activity in luciferase-based reporter assays, are located close to the transcription start site (TSS), and are often found in conserved regulatory regions [77]–[79]. However, whether super-enhancers represent a novel regulatory paradigm distinct from previously defined enhancers has been debated [80]. Many features of super-enhancers may be a direct result of a biased definition. The use of specific parameters to define super-enhancers does not always have a strong biological rationale. For example, setting high thresholds for peak calling during the analysis of ChIP-seq data might explain the particularly strong activator binding observed in super-enhancers, and selecting regions with a high number of binding sites could account for their high expression levels in reporter assays. Additionally, the choice of transcription factors analyzed can influence the perceived association of super-enhancers with specific biological functions.

Enhancers are more likely to evolve in regions that already harbor regulatory activity. Accessible chromatin is essential for the development of functional enhancers. Regions with open chromatin are more likely to accumulate regulatory sequences than areas with closed chromatin, which require an additional evolutionary step to modify chromatin conformation. Co-opting the activity of existing enhancers may provide a quicker path to developing novel functions [81]. For example, a *yellow* enhancer driving a spotted expression pattern in *Drosophila* wings evolved independently twice, in the *D. melanogaster* and *D. obscura* groups.

Despite evolving in different parts of the *yellow* locus, both instances of evolution occurred within the context of preexisting wing enhancers [48]. For the *D. melanogaster* group, the *spot* enhancer evolved in the vicinity of the enhancer driving a background activity in the wing while in the *D. obscura* group, it evolved next to an ancestral element activating gene expression in the wing veins.

2.2 Enhancer size and enhancer cores

Enhancers are classically defined as short DNA fragments

Enhancers are usually described as small regulatory elements spanning a few hundred base pairs. Small clusters of binding sites activated by a few transcription factors are sufficient for transcription initiation. The first described enhancer, *SV40*, consisted of only 72 bp, and even a single binding site was shown to have activity in synthetic enhancers [12], [82]. While some studies suggest that regulatory information in the genome can be distributed over larger fragments, the vast majority of research focuses on sequences under 800 bp. This is mainly due to the focus on identifying fragments necessary for driving an expression pattern that recapitulates the expression pattern of the endogenous gene, sometimes referred to as minimal or core enhancers [22], [83]. Another reason why enhancers are described as short is that most studies use qualitative approaches focusing on spatial expression dismissing qualitative aspect and as a result not taking into consideration fragments that modulate the amount of transcripts. While minimal enhancers are essential for transcription initiation, they can be insufficient to drive endogenous levels of gene expression.

Regulatory information can be distributed over longer fragment

Qualitative approaches commonly used in enhancer research often overlook the quantitative effects on gene expression, neglecting the contribution of larger enhancer fragments. Enhancer activity is tested using reporter constructs and assessed by the presence or absence of the correct expression pattern. Increasing the size of the tested enhancer often does not alter the expression pattern and, as a result, is not considered to significantly contribute to gene expression or contain essential regulatory elements. However, evidence suggests that regions flanking minimal enhancers contribute to transcriptional output. For example, in the *even-skipped* locus, adding 320 bp to the minimal *stripe 2* enhancer increased transcription levels fivefold [84]. Such quantitative differences can be crucial for the successful outcome of a developmental process, or even for survival, particularly under unfavorable environmental conditions, where robust gene expression is essential. Transgenic fly lines containing only the minimal *stripe 2*

enhancer (with flanking regions removed) produce viable flies under normal conditions. However, environmental perturbations like temperature changes negatively impact viability, suggesting that these flanking regions play a role in temperature compensation, likely by increasing levels from the minimal enhancer expression [85].

Sequences outside of minimal enhancers also contribute to evolutionary changes. A study dissecting the *even-skipped* locus in *Drosophila erecta* showed that the *stripe 2* enhancer ortholog failed to drive significant expression in *D. melanogaster*. However, an extended enhancer region containing multiple binding motifs for the pioneer transcription factor Zelda was able to drive normal expression [47]. This finding highlights the importance of studying enhancer function and evolution at the level of entire loci rather than focusing solely on minimal enhancers.

Sequences outside of the minimal enhancer can have distinct functions

Minimal enhancers contain information that regulates transcription in both space and time; however, additional aspects of gene regulation might be encoded in *cis*. Analysis of sequences beyond minimal enhancers suggests that longer fragments may carry diverse types of information. A study described facilitator elements adjacent to the classical enhancer of the human adenosine deaminase gene. These elements do not exhibit independent activity but enhance reporter expression, possibly by influencing chromatin structure [86]. Another study investigating facilitator elements in the α-globin super-enhancer suggested their role in enhancer RNA (eRNA) production, cofactor recruitment, and enhancer-promoter interactions [87]. A study comparing MPRA results across different fragment lengths found significant differences in enhancer activity between short (192 bp) and long (678 bp) sequences. Features associated with enhancer activity in longer fragments included components of RNA Polymerase III, the AP-1 complex, core histone-modifying enzymes, and an increased number of transcription factor binding sites [69]. Differences in the structure of enhancers have also been reported in a study that described “complex enhancers” composed of fragments from different evolutionary origins, with older “core” and younger “derived” sequences. The report indicates that derived sequences have distinct constraint profiles, transcription factor-binding preferences, and tolerance to variation compared to cores [88]. These studies suggest that information isn’t distributed evenly in enhancers and might contribute to different aspects of transcription regulation.

Methodologies can bias our perception of enhancer size

Methods used for studying and analyzing enhancers can introduce biases regarding enhancer size and distribution of regulatory information. The focus on shorter sequences in experimental approaches also extends to computational approaches. Massively parallel reporter assays predominantly use short DNA fragments under 200 bp. Training machine-learning models on such data biases enhancer predictions toward short sequences, reinforcing the assumption that regulatory information is confined to small DNA fragments. Although longer fragments are more challenging to test experimentally, they contain a substantial amount of biological information [69]. Enhancer prediction using sequencing-based methods such as ATAC-seq and ChIP-seq also strongly depends on arbitrarily chosen parameters, which can bias predictions toward short regulatory fragments. During analysis of sequencing data, focusing only on regions with very high peaks and binning them into small discrete intervals may lead to the selective identification of short, highly active enhancers while overlooking larger regulatory regions that play diverse roles in gene expression.

Some studies predict longer enhancers

Despite the prevailing focus on short enhancers, exceptions exist. A study using ENCODE data to predict enhancers in human pancreatic islets coined the concept of stretch enhancers, which span large genomic regions [89]. While most predicted enhancers were under 800 bp, a significant fraction contained regulatory elements spread over 3,000 bp. Stretch enhancers were more cell-type-specific and were associated with robust gene expression. Within this analysis, enhancer length correlated with cell specificity, suggesting that longer enhancers play crucial roles in tissue-specific gene regulation. Stretch enhancers appear to differ from super-enhancers in several ways; they are located further away from the transcription start site, and are less conserved than super-enhancers, and contain more markers of poised enhancers, that are partially repressed and activate gene expression after reception of the appropriate signal [79]. Stretch enhancers were shown to have reproducible activity in cell culture and mouse embryo reporter assays. However, it remains unclear whether they consist of evenly distributed regulatory elements or contain localized clusters of transcription factor binding sites. Further studies comparing stretch enhancers with classical enhancers that include extended flanking regions could provide insights into their mechanism of action.

Research on minimal enhancers has significantly advanced our understanding of gene regulation and led to indispensable tools for genomic research such as the GAL4/UAS system. Studying shorter sequences is often more convenient, both experimentally and computationally.

However, to develop a nuanced understanding of enhancer function and the evolutionary mechanisms shaping genomes, it is crucial to analyze larger DNA loci and consider how experimental and analytical designs influence enhancer characterization.

Enhancers are defined as modular fragments

Modular activity of enhancers is one of their fundamental features and is reflected in their structure. A textbook image of a regulatory region is a series of regulatory blocks sequentially distributed on DNA and driving independent, distinct activities. A classical example of a modular enhancer is the *even-skipped (eve)* locus. *Eve* encodes a homeodomain protein that is crucial for segmentation in *Drosophila* and is expressed in seven distinct stripes during embryogenesis. This activity is driven by multiple enhancers, with each enhancer regulating the expression of one or two stripes. These enhancers were initially described as being ~500 bp long, separated by kilobase-long spacers that ensure proper gene function [90]. Such structure might be shaped evolutionarily, by constraining enhancers to narrow fragments that allow tissue-specific activity or provide spacing that might be important for interactions among regulatory sequences. Another mechanism creating distinct modules might involve insertion of transposable elements that acquired regulatory function.

Pleiotropic nature of enhancers

One regulatory element can activate multiple expressions. This ability of a DNA fragment to encode regulatory information for driving multiple expressions in different contexts is referred to as enhancer pleiotropy. An increasing number of studies describe pleiotropic enhancers, which regulate multiple genes or function in multiple tissues. Different activities can be regulated by distinct sets of TFBSs located in the same region, or by pleiotropic binding sites that are active in multiple contexts. For example, a single base pair mutation in the enhancer of the *scute* gene has been shown to contribute to bristle loss in developing genitalia and bristle gain in developing legs [91]. Similarly, pleiotropic sites have been identified in the *E6* enhancer of the *shavenbaby* gene, regulating gene activity simultaneously in embryonic and pupal epidermis [92]. Deciphering whether an enhancer contains independent sets of binding sites that are intertwined or pleiotropic binding sites is challenging because it requires detailed analysis of separate motifs and their contributions to different activities.

The strict modular view of enhancers may be partially reinforced by experimental design. If enhancer activity is tested only in specific tissues or at specific time points, additional

regulatory roles may be overlooked. Furthermore, qualitative analyses may fail to detect subtle changes in expression patterns across different tissues. For example, the *shavenbaby* gene is regulated by seven enhancers, originally described as functioning exclusively in embryos. However, thorough and systematic dissection revealed that these enhancers are active in multiple developmental contexts, driving expression in various tissues during larval and pupal development [92]. Even in classic models like *eve*, later studies suggested potential overlap in regulatory information between the *stripe 2* and *stripe 3+7* enhancers, challenging the strict modular structure [93]. Another pleiotropic enhancer of *eve* has been shown to drive expression in ganglion mother cells, neurons, and the anal plate ring [94].

Understanding the structure of enhancers and the level of their modularity is important as it uncovers evolutionary constraints of regulatory sequences. While modular enhancers have a limited effect on gene function, deleterious mutations in pleiotropic sequences will affect larger cell populations and might not be permitted. A study demonstrated enhancers that show activity in both the developing limbs and genitalia of mice. Similar enhancers were also identified in Anolis lizards and surprisingly snakes. The conservation of limb enhancer sequences in snakes can be explained by their utilization during genitalia development [95]. Another study compared liver enhancers identified from genome-wide histone modification profiles of ten mammalian species and showed that enhancers exhibiting higher conservation are active across more cellular contexts, regulate more genes, and are less tolerant to loss-of-function mutations than species-specific enhancers [96]. These studies highlight how the pleiotropic nature can result in stronger constraints on regulatory sequences.

2.3 Enhancer grammar

Enhancer sequence encodes rules of gene transcription

For an enhancer to function effectively, multiple transcription factors must bind to TFBSs leading to the recruitment of transcription machinery and the activation of gene expression. This process involves various DNA-protein and protein-protein interactions that can be influenced by features such as the number, spacing, orientation, and binding affinities of TFBSs. The set of rules determining how an enhancer sequence translates into expression activity is known as enhancer grammar. One of the earliest studies on enhancer grammar focused on the human interferon-beta (*IFN-β*) gene, demonstrating that the proper function of the gene relies on a precise helical relationship between individual TFBSs [97]. Another study showed the interplay between the affinity and organization of transcription factor binding sites, where

organization can compensate for low affinity and vice versa [98]. At the same time, some enhancers show flexibility regarding their sequence composition. Many enhancers in closely related species have very low similarity in the enhancer sequences while expressing genes in a conserved pattern. A comparison of the *even-skipped* locus in six species of scavenger flies and *Drosophila melanogaster* revealed that while their expression patterns were nearly identical, there was little sequence similarity between species. Only a small number of short (20-30 bp) conserved sequences enriched for pairs of overlapping or adjacent binding sites were identified [71]. This indicates a certain flexibility in the regulatory sequence that allows different versions of enhancers to produce similar expression patterns. There is a debate over which aspects of transcription factor binding site distribution are crucial for enhancer function, how much flexibility there is, and whether there are general rules that apply to all enhancers.

Three main models describe rules of the enhancer grammar

There are several models that try to summarize rules of the enhancer grammar. The Enhanceosome Model suggests a very rigid organization where the exact arrangement of TFBSs is crucial for the function of the enhancer. This model suggests that a precise spatial configuration of transcription factors bound to the enhancer is required to initiate transcription. The most iconic example of the enhanceosome is the *IFN-β* gene where single point mutations that move or remove the binding site of individual proteins disable its enhancer, suggesting that an overall protein-DNA superstructure is crucial [97].

The Billboard Model represents the opposite scenario, where there are no strict rules on the arrangement of TFBSs. In this model, the presence of necessary TFBSs within the enhancer is sufficient, regardless of their order or orientation. A study testing a compact synthetic enhancer in *Drosophila* embryos showed that closely apposed transcription factor binding sites can be interpreted independently by transcriptional machinery [99]. While the Billboard model suggests complete flexibility regarding site arrangement, enhancers that function in any combination and arrangement of binding sites have not been demonstrated.

The TF-Collective Model suggests a more flexible interaction where multiple transcription factors collectively occupy the enhancer without a strict motif arrangement, emphasizing the importance of transcription factor interactions over the specific sequence or spatial configuration. Using ChIP-seq and reporter assays, researchers studied heart enhancers in flies and discovered that active enhancers need a certain number of TFBSs to attract necessary proteins through direct DNA interactions. Once enough TFs bind to these sites, they can help

recruit additional TFs through interactions among themselves [100]. Such co-occupancy occurs in the absence of specific motif organization.

In reality, these models most likely represent the range of possibilities in transcription regulation, with billboard and enhanceosome models representing opposite ends of a spectrum. A single enhancer can incorporate aspects of these models to various degrees. The degree to which enhancers incorporate these features is shaped by molecular mechanisms driving the activity of separate enhancers and evolutionary constraints that differ based on the rigidity of the grammar. Billboard and TF-collective models are more flexible and operate under low sequence conservation as they can easily adapt to TFBS changes. The enhanceosome model, on the other hand, will be more conserved, as any alterations in the sequence, order, or spacing of TFBSs will disrupt enhancer function. Understanding enhancer grammar is crucial because it enables us to decipher the complex instructions for development encoded in our genomes, helping us design synthetic, tissue-specific enhancers and predict how genetic variations within enhancers influence diseases and evolutionary adaptations.

3. Yellow as a model for studying enhancer function and evolution.

yellow is developmental gene regulating pigmentation

yellow (*y*) is a developmental gene that plays a crucial role in insect pigmentation. As a member of the gene family that regulates melanin synthesis pathways, it contributes to the production of dark pigments in the cuticle, forming the characteristic black coloration patterns observed on the body, wings, larval mouthparts and bristles of the fruit fly. Mutations in the *y* gene result in lighter, yellowish body pigmentation, hence its name.

The precise molecular function of *y* remains unknown, but it is thought to play a direct role in melanin synthesis. One crucial step in melanin production in insects is the conversion of dopachrome to 5,6-dihydroxyindole, and some members of the *y* gene family have been shown to act in this process as dopachrome conversion enzymes (DCE) [101]. Additionally, *yellow* has been shown to act in an alternative melanin production pathway that uses dopamine as a melanin precursor [102]. Yellow-h has been identified as an enzyme in this biosynthetic pathway [103]. Other hypotheses include Yellow acting as a hormone or a growth factor [104] or serving as an anchoring pigment in the cuticle layer [105], [106].

yellow is regulated by multiple enhancers

yellow is expressed in different tissues, and its expression pattern has been shown to prefigure pigmentation in adult flies across many species. This pleiotropic activity is regulated by multiple modular, tissue-specific enhancers that control expression in the body (head, thorax, abdomen), wings, wing veins, and bristles. These enhancers are distributed in two regions: in the intergenic region directly upstream of the *y* transcription start site and in the intronic region. Kalay and Wittkopp analyzed the *y* regulatory regions of six species, including members of the *Drosophila* (*mojavensis*, *virilis*, *grimshawi*) and *Sophophora* (*melanogaster*, *pseudoobscura*, *willistoni*) subgenera [107]. This study revealed that while the bristle enhancer was consistently mapped to the intron, enhancers driving expression in the thorax, abdomen, and wings, on the other hand, map to different regions in the locus. Depending on the species, these activities may be driven by intergenic, intronic, or both regions.

Changes in yellow regulatory sequences contribute to the morphological evolution

The *yellow* gene's regulatory region has been a model for studying the evolution of morphological features. Pigmentation is a rapidly evolving trait, and changes in the regulatory sequences of *y* have been shown to contribute to this evolution. A dark pigmentation spot on the wing of some *Drosophila* species is one of the best-studied examples of evolution in *cis*. The spot has been gained and lost multiple times in different species [48]. Within the *D. melanogaster* group, there have been at least five independent losses of *spot* enhancer activity. For example, in *D. mimetica* and *D. gunungcola*, mutations in a small number of binding sites were sufficient to abolish *y* expression in the spot region. Interestingly, wing spots have evolved independently at least twice in the *D. melanogaster* and *D. obscura* groups. In *D. tristis* (*D. obscura* group), the *intron-spot* enhancer evolved within an intron, unlike in the *D. melanogaster* group, where the *spot* enhancer, a distinct element with a similar activity, was mapped in the upstream intergenic region.

Yellow enhancers demonstrate pleiotropic activities

Most studies depict the *y* regulatory region as a collection of modular enhancers, each driving distinct tissue-specific expression. However, more detailed analyses suggest that this view may oversimplify the structure of regulatory sequences. Kalay et al. (2019) demonstrated that fragments from intergenic and intronic regions in *D. melanogaster*, *D. pseudoobscura*, and *D. willistoni* could drive multiple activities in different spatial elements. Various enhancer fragments were driving overlapping activities, demonstrating robustness and expressing cryptic

patterns that might reveal the evolutionary history of the region. Such analysis demonstrated the complex organization of the region and highlights the need for detailed and systematic analysis.

Aim of the Study

This dissertation investigates how regulatory information is distributed in a locus containing multiple enhancers. I present studies that test enhancer modularity and explore the evolutionary mechanisms that shape regulatory regions.

As a model, I used the *yellow* locus in *Drosophila biarmipes*. *Yellow* is a developmental gene that prefigures pigmentation in the body and wings of a fly and is regulated by multiple enhancers located directly upstream of its transcription start site (5' regulatory region). We created a series of reporter lines containing various fragments of the 5' region and imaged fly wings and the abdomen to investigate how changes in the enhancer sequences affect the expression of the reporter. We used an image registration pipeline to overlap images and statistically analyze changes in fluorescence levels and expression patterns.

In the first chapter, I helped investigate the positional relationship between two enhancers that are thought to have evolved through co-option. The wing blade enhancer drives a background expression pattern in the fly wing and is thought to serve as the foundation for the evolution of the novel spot enhancer. The goal was to map these enhancers and examine whether there is a functional dependence between them.

In the second chapter, I expand the investigation to the body enhancer by evaluating its activity in the fly abdomen. To understand how multiple activities are encoded in the same region, I analyzed the types of expression patterns regulated in the abdomen and how this information is distributed throughout the entire 5' regulatory locus. The goal was to assess the sequence and functional independence of multiple enhancers driving the activity of the *yellow* gene.

In the last chapter, I contributed to a project focusing on the local arrangement of binding sites in the spot enhancer. Using systematic mutations, we tested how spatial and quantitative aspects of gene expression are encoded in the regulatory region.

Results

Paper I: Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility

Yaqun Xin, Yann Le Poul, Liucong Ling, **Mariam Museridze**, Bettina Mühling, Rita Jaenichen, Elena Osipova, Nicolas Gompel

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES | Vol. 117, No. 34 | 25 August 2020

DOI: 10.1073/pnas.2004003117

From: Xin, Y. et al. Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. Proc. Natl. Acad. Sci. 117, 20636–20644 (2020).

Correction

DEVELOPMENTAL BIOLOGY

Correction for "Enhancer evolutionary co-option through shared chromatin accessibility input," by Yaqun Xin, Yann Le Poul, Liucong Ling, Mariam Museridze, Bettina Mühling, Rita Jaenichen, Elena Osipova, and Nicolas Gompel, which was first published August 10, 2020; 10.1073/pnas.2004003117 (*Proc. Natl. Acad. Sci. U.S.A.* **117**, 20636–20644).

The authors note that, due to a printer's error, the article title "Enhancer evolutionary co-option through shared chromatin accessibility input" appeared incorrectly. The correct title should be "Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility." The online version has been corrected.

First published September 28, 2020.

www.pnas.org/cgi/doi/10.1073/pnas.2018481117

Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility

Yaqun Xin[a,1], Yann Le Poul[a,1], Liucong Ling[a], Mariam Museridze[a], Bettina Mühling[a], Rita Jaenichen[a], Elena Osipova[a], and Nicolas Gompel[a,2]

[a]Fakultät für Biologie, Biozentrum, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

Edited by William J. McGinnis, University of California San Diego, La Jolla, CA, and approved July 13, 2020 (received for review March 2, 2020)

The diversity of forms in multicellular organisms originates largely from the spatial redeployment of developmental genes [S. B. Carroll, *Cell* 134, 25–36 (2008)]. Several scenarios can explain the emergence of *cis*-regulatory elements that govern novel aspects of a gene expression pattern [M. Rebeiz, M. Tsiantis, *Curr. Opin. Genet. Dev.* 45, 115–123 (2017)]. One scenario, enhancer co-option, holds that a DNA sequence producing an ancestral regulatory activity also becomes the template for a new regulatory activity, sharing regulatory information. While enhancer co-option might fuel morphological diversification, it has rarely been documented [W. J. Glassford et al., *Dev. Cell* 34, 520–531 (2015)]. Moreover, if two regulatory activities are borne from the same sequence, their modularity, considered a defining feature of enhancers [J. Banerji, L. Olson, W. Schaffner, *Cell* 33, 729–740 (1983)], might be affected by pleiotropy. Sequence overlap may thereby play a determinant role in enhancer function and evolution. Here, we investigated this problem with two regulatory activities of the *Drosophila* gene *yellow*, the novel *spot* enhancer and the ancestral *wing blade* enhancer. We used precise and comprehensive quantification of each activity in *Drosophila* wings to systematically map their sequences along the locus. We show that the *spot* enhancer has co-opted the sequences of the *wing blade* enhancer. We also identified a pleiotropic site necessary for DNA accessibility of a shared regulatory region. While the evolutionary steps leading to the derived activity are still unknown, such pleiotropy suggests that enhancer accessibility could be one of the molecular mechanisms seeding evolutionary co-option.

transcriptional regulation | regulatory evolution | pattern formation | chromatin | enhancer

Evolutionary co-option happens when an ancestral biological object is recycled to a new function while maintaining its ancestral role. Novel *cis*-regulatory elements (transcriptional enhancers), for instance, may emerge through co-option of a preexisting element. In this case, the ancestral and the derived regulatory functions map to overlapping DNA segments, which we define as structural co-option. They may share ancestral components such as ancestral transcription factor binding sites (TFBSs), bringing co-option to a functional level but resulting in a functional dependency or pleiotropy (1–5). Because the boundaries of transcriptional enhancers are difficult to define precisely, it is most often challenging to assess sequence overlap and regulatory pleiotropy when a new regulatory activity emerges in the vicinity of an ancestral activity (6–8). An enhancer is typically defined on the basis of its activity, notably in a transgenic context, using reporter assays as a segment of sequence sufficient to direct a spatiotemporal transcriptional activity resembling that of their original target gene (9–12). In developmental biology, enhancer boundaries are defined from a DNA sequence sufficient to recapitulate specific elements of the endogenous expression pattern of the corresponding gene. This definition has several limits. One limit, not addressed in this study, is that the biological context in which enhancer activity is assessed differs from the native genomic and transcriptional context. Another limit is that it focuses on the relative spatial distribution of the regulatory activity, the pattern, rather than on its quantitative aspects and is therefore likely to reveal only partial enhancer sequences and to miss pleiotropic effects. Moreover, fragments are often chosen either arbitrarily or based on sequence conservation or genomic marks to limit the risk of disrupting functional features. These fragments can pinpoint minimal enhancers but fail to determine whether the same sequences at their locus of origin are necessary and sufficient to recapitulate the transcriptional activity of their cognate target gene (13–15). Finally, the representation of enhancers as rectangular boxes or stretches of sequence eludes the actual distribution of regulatory information along the enhancer sequence with different segments contributing different inputs (activation, repression, permissivity) and different activity levels. In an attempt to overcome most of these limits, we examine here the molecular relationship that a new regulatory activity entertains with a nearby ancestral activity.

While the wings of *Drosophila* are uniformly shaded with light gray pigment, some species, including *Drosophila biarmipes*, have gained a pattern of dark pigmentation, a spot, at the wing tip (7). The expression of the gene *yellow* (*y*) in the wings during pupal life is necessary both to the wing blade shading and to the spot pattern (16). These two components of *yellow* wing expression result from two distinct regulatory regions, the ancestral *wing*

Significance

Form diversity is fueled by changes in the expression of genes that build organisms. New expression often results from the emergence of new DNA switches, known as transcriptional enhancers. Many enhancers are thought to appear through the recycling of older enhancers, a process called evolutionary co-option. Enhancer co-option is difficult to assess, and the molecular mechanisms explaining its prevalence are elusive. Using state-of-the-art quantification and analyses, we reveal that the sequences of an ancestral and a derived enhancer overlap extensively. They contain specific binding sites for regulators imparting spatial activities. We found that the two enhancers also share a site facilitating access to chromatin in a region where they overlap.

Author contributions: Y.X., Y.L.P., L.L., and N.G. designed research; Y.X., Y.L.P., L.L., M.M., B.M., R.J., and E.O. performed research; Y.X., Y.L.P., L.L., M.M., and N.G. analyzed data; and Y.L.P. and N.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Y.X. and Y.L.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: gompel@bio.lmu.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2004003117/-/DCSupplemental.

First published August 10, 2020.

blade enhancer (referred to as "*wing*" in other publications) and the recently evolved *spot* enhancer (6, 7, 17–21). In *D. biarmipes*, both activities map within 6 kb upstream of *y* transcription start site (6) (*y* 5′ region) (Fig. 1*A*). Two short adjacent regulatory fragments (~1.1 kb together) within this *y* 5′ region drive distinct spatial expression in the spot and uniformly in the wing blade, respectively (6, 16). It is, however, unclear to what extent sequences surrounding these fragments at their locus of origin also contribute to each transcriptional activity. It is equally unclear whether or not the contributing sequences of the two enhancers overlap. Because both activities are driven in the same tissue and developmental stage, it is technically and conceptually challenging to evaluate the distribution of regulatory information quantitatively and assess possible pleiotropic effects.

Testing the hypothesis of enhancer structural co-option in our system required us to link regulatory information distributed in DNA to activities measured with quantitative spatial reporter expression. Using classical reporter assays in transgenic *Drosophila*, we mapped regulatory information with two series of nested fragments, depleting sequence information from the 3′ end or the 5′ end. This approach reveals the contribution of DNA segments along the sequence, including sequences that cannot drive activity alone and whose activity depends on nearby sequences. A simple qualitative assessment of the reporter activity resulting from each construct is, however, insufficient to produce a precise regulatory map. Moreover, qualitative or semiquantitative approaches would not allow us to separately measure each regulatory activity because of the spatial and temporal overlap with the other activity. This prompted us to develop a generic quantification pipeline to comprehensively describe variation in reporter expression levels across the wing. Finally, with an appropriate analytical framework, we have mathematically separated the two activities, although they drive in the same tissue and developmental stage. Our results indicate that the regulatory information spans a much wider region than previously described and that, unexpectedly, the ancestral *wing blade* and the derived *spot* activities overlap extensively. Further, the molecular dissection of the overlapping region led us to uncover a site with pleiotropic effects in the core of the derived enhancer, which proved to regulate chromatin accessibility.

Results

To evaluate how the *wing blade* and the *spot* activities are distributed along *y* 5′ sequences of *D. biarmipes* and to test whether they are intertwined, we derived two series of reporter constructs from the *y* 5′ region (Fig. 1*B*) and tested them in *Drosophila melanogaster*. The first series (*D*) consists of distal (5′) truncations, while in the second series (*E*), we randomized increasingly longer segments of wild-type proximal (3′) sequence, keeping the total fragment size constant (identical to that of construct *D2*). In each series, the largest intact fragment is a reference for the complete regulatory information (*D0* in the 5′ dissection and *D2* in the 3′ dissection) (Fig. 1*B*). These two series allow us to measure how a segment modulates regulatory information, when the information in 3′ (*D* series) or in 5′ (*E* series) of this segment is preserved. We define as enhancer core any segment that, in its local genomic context (including the distance to the core promoter), is necessary and sufficient to drive significant levels of a given activity (see below).

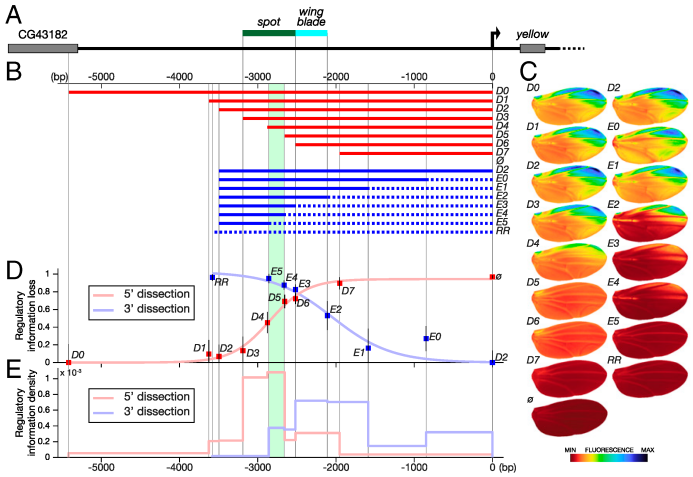

Fig. 1. Quantitative mapping of wing regulatory activities at the *yellow* locus. (*A*) The top line represents the 5′ region of the *yellow* locus from *D. biarmipes*; the green and blue bars indicate the respective locations of *spot* and *wing blade*, respectively, as originally mapped (6). (*B*) Two series of fragments derived from *y* 5′ region (*D* series, red; *E* series, blue) were tested in reporter constructs in *D. melanogaster*. The dotted lines in the *E* series represent randomized sequences (*Materials and Methods*); *ø* and *RR* stand for an empty reporter vector and a vector containing a completely randomized fragment, respectively. The area shaded in green in *B*, *D*, and *E* identifies a previously studied regulatory component (16), *spot*^{196}. (*C*) Images of average reporter expression of all individuals for each construct in the wing at emergence from the pupa according to the color map below. Note that *spot*^{196} appears strictly necessary to any activity in the spot region (compare *D4* with *D5* and compare *E4* with *E5*). (*D*) Overall loss of regulatory information (fluorescence levels) along the sequence (base pairs). The loss of phenotypic information measures how much truncating or randomizing a fragment affects the whole activity relatively. It is estimated by the ratio $\frac{d(P_x, P_{ref})}{d(P_{\varnothing}, P_{ref})}$, where P_x, P_{ref}, and $P_{\varnothing}$ are the phenotypes of construct *x*, construct *D0* or *D2* (the largest constructs of each series as a reference for that series), and the empty construct *ø* in the PCA space, respectively, plotted as a function of the distance to the starting point of the randomization (series *E*) or truncation (series *D*). Error bars represent the SD of the phenotype of each construct in PCA space normalized by the distance $d(P_{\varnothing}, P_{ref})$. (*E*) Density of regulatory information along the *y* 5′ region (fluorescence levels per base pair). It is technically the first derivative of the regulatory information loss shown in *D*. For each series, it represents the phenotypic distance (in PCA space) between two consecutive constructs divided by the number of base pairs that changed between those two constructs. It indicates the regulatory contribution per base pair of each DNA segment measured in each series.

We imaged 27 wings on average (minimum 22; maximum 39) for each construct and used them to precisely quantify spatial reporter expression (referred to as phenotype) driven by each construct in the wings of transgenic *D. melanogaster*, used here as an experimental recipient with site-specific transgenesis (22) (Fig. 1*C*). We summarized the variation in activity across the wing (both pattern and levels) from each series of constructs with principal component analysis (PCA), producing a comprehensive description of the phenotypic variation (*SI Appendix*, Fig. S1*A*). We define the overall loss of regulatory information for each construct as the amount of change in activity compared with the activity of a reference construct. To estimate this loss, we use the distance between the average phenotypes, as described in PCA space. This distance takes any change of activity into account. As this measure is more informative when represented relatively, we normalized the loss of regulatory information to the total amount of regulatory information brought by the enhancer, as estimated by the distance between the reference activity and the empty construct. The relative loss is therefore given by the following formula:

$$\frac{d(P_x, P_{ref})}{d(P_\emptyset, P_{ref})},$$

where P_x, P_{ref}, and $P_\emptyset$ are the average phenotypes of construct x, the reference construct (*D0* or *D2*, the largest constructs of each series), and the empty construct $\emptyset$, respectively, and $d(P_x,P_{ref})$ is the distance between these average phenotypes. Hence, this ratio estimates the loss of regulatory output of each construct compared with the largest construct of the series. In contrast to classical reporter assays testing the sole sufficiency of candidate regulatory fragments to produce a spatial pattern, the combined series reveal a surprisingly large stretch of the regulatory activities along *y* 5′ sequences (the regulatory activity of each construct is significantly different from that of the largest construct of the series) (*SI Appendix*, Table S1). Further, Fig. 1*E* establishes the contribution of each segment to these activity differences (intensity effect/base pair). Consistent with previous work (6), the 5′ series (*D*) shows that most of the regulatory activity maps within ∼1.7 kb (−3.6 to −2 kb) (Fig. 1 *D* and *E*). The 3′ dissection, however, reveals additional regulatory information contributing to the activity, located proximally to this 1.7-kb segment and extending to *y* promoter region (Fig. 1 *D* and *E*). These results demonstrate that *y* regulatory activities in the wing extend over 3 kb (conservative) to 4 kb upstream of *y* promoter, a much broader region than previously assessed (6, 7).

To specifically address the question of regulatory co-option, we then examined the sequence relationship between *spot* and *wing blade* activities. It was first necessary, however, to mathematically separate the *wing blade* and the *spot* activities to then evaluate to what extent they map to distinct segments. In the PCA of all constructs, we found that both the *D* and the *E* series varied mostly along a combination of two additive directions in the phenotype space, explaining a large part (69%) of the phenotype variance resulting from the two dissection series. We noticed that these two directions correspond to a near-uniform increase in expression across the wing and an increased expression mostly at the anterior distal wing tip, respectively. These two directions map to overlapping sequence segments: −2,656 to 0 bp ($\emptyset$ to *D5*) and −3,496 to −2,519 bp (*RR* to *E2*, where RR is a segment of randomized sequence; see *Materials and Methods*), respectively (reference segments in Fig. 2 *B* and *C*). The segment driving a uniform pattern of activity fully includes the originally defined *wing blade* enhancer (6) but not the full original *spot* enhancer. Surprisingly, the segment driving a spotted pattern of activity includes both the originally defined *spot* and *wing blade* enhancers (6), despite its very low activity in the wing blade.

Hence, guided by the structure of the phenotypic space, we extracted representations of the actual patterns of activity driven by the *wing blade* and the *spot* enhancers, where *D5* and *E2* are representative segments of each direction, respectively (Fig. 2 *B* and *C* and *SI Appendix*, Fig. S1*A*). The segments defining the two activities (−3,496 to −2,519 bp for the *spot* activity and −2,656 to 0 bp for the *wing blade* activity) share regulatory information, indicating that our estimate of the structural co-option is conservative as it tends to minimize the measured sequence overlap between the two activities. It is important to note that the definition of those two directions (independently representing the *spot* and *wing blade* activities) (axes of Fig. 2*A*) is not linked to prior knowledge on these enhancers, neither from the phenotypic nor the sequence point of view. The fact that those data-driven directions correspond to uniform and spotted activities confirms that the two activities map mainly, when the two series are considered separately, to different segments. It also shows that the full 5′ region of *y* drives mainly two different activities, apparently relatively independently. Structural co-option implies that at least some segments of *y* 5′ contribute to the *wing blade* and *spot* activities simultaneously. Because the two activities overlap in space in the wing, they cannot be distinguished by simply measuring the separate reporter expression in their respective domains. To independently evaluate the uniform activity and the patterned, spotted activity, we projected the phenotype of each individual wing in the two-dimensional basis defined by these two phenotypic directions using a mathematical operation called change of basis (*Materials and Methods*, Fig. 2*A*, and *SI Appendix*, Fig. S1*A*). With the possibility to evaluate *wing blade* and *spot* activities independently, we quantified the contribution of each DNA segment to the respective activities.

We first tested whether, in the case of the *wing blade* and *spot* enhancers, the enhancer cores, as defined above, mapped to the same region. In our experimental system, the core of an enhancer is a segment sufficient to contribute a uniform or a spotted activity in the wing when either flanking 5′ or 3′ regions are missing. Because of the particular enhancer configuration in our system, each dissection series is simultaneously testing the sufficiency of a segment for one activity and its necessity for the other activity. This definition takes the preserved distance of regulatory information to the core promoter into account as well as the local genomic context at the *yellow* locus. We submit that this approach is more informative than testing the sole sufficiency of an isolated segment, as is classically done (21). These cores can logically be visualized in Fig. 2 *B* and *C* as the intersection between the 5′ and 3′ dissection curves. The core of the *spot* activity as revealed here coincides exactly with the $spot^{196}$ enhancer, defined in previous work (6, 16). For the *wing blade* enhancer, interestingly, there are two cores (from −2,111 to −1,953 bp and from −2,877 to −2,518 bp) flanking what was previously defined as the *wing blade* enhancer (6). Thus, there are two regions sufficient to drive a significant amount of *wing blade* activity when either 5′ or 3′ regulatory information is missing. Moreover, the overlap between the core of the *spot* enhancer and one of the cores of the *wing blade* enhancer reveals that a region inside the *spot* enhancer is sufficient to drive a substantial amount of expression in the wing blade.

Further investigating the interweaving of the two activities, we found, strikingly, that the sequences contributing to them largely overlap (Fig. 2 *B* and *C* and *SI Appendix*, Fig. S1 *C* and *D*). We asked whether sequences 3′ to the *spot* reference segment also contributed significant regulatory information to the *spot* activity. To this end, we compared *D2* (the largest fragment of the *E* series) with *E2*, in which these 3′ sequences are randomized (−2,111 to 0 bp) and found that this region contributes a substantial and unexpected amount of *spot* activity [22%, ANOVA: $F(1, 55) = 22.57$, $P = 1.4954\text{e-}05$] (horizontal double arrow in Fig. 2*A* and 3′ curve in Fig. 2*B*). Reciprocally, we asked whether

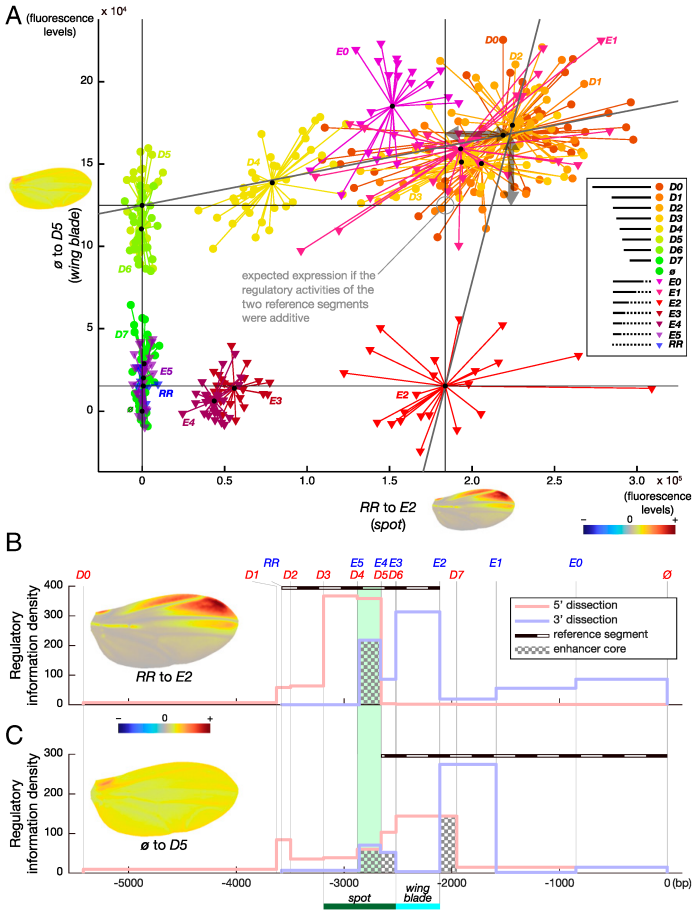

Fig. 2. *wing blade* and the *spot* activities map to overlapping sequences. (*A*) Representation of *wing blade* activity as a function of *spot* activity. Independent estimates were produced by projecting the PCA phenotypic space (PCA in *SI Appendix*, Fig. S1*A*) on a two-vector basis defined by two independent directions identified in *SI Appendix*, Fig. S1*A* (phenotypic directions with color map near each axis) and corresponding to *wing blade* (constructs *ø* to *D5*; dotted line in *B*) and *spot* (constructs *RR* to *E2*; dotted line in *C*) activities. The mathematical change from the PCA coordinate basis to this two-vector coordinate basis affords the separation and independent measurements of both activities, although they occur in the same tissue. This graph shows for each individual wing (dots and triangles) of each reporter line the contribution to the *wing blade* and *spot* activities. Small black dots mark the center of a cluster for each construct. Note that constructs driving both activities (*D0* to *D4*, *E0* to *E1*) produce more expression than expected if the activities were strictly additive (i.e., they lie above the point of strict additivity of the activities driven by the two reference segments of the *wing blade* and the *spot* activities; the resulting nonadditive effects are shown with double arrows). (*B* and *C*) Density of regulatory information along the *y* 5′ region (fluorescence levels per base pair) as measured specifically (*Materials and Methods*) for the *spot* activity (*B*) and the *wing blade* activity (*C*). Construct boundaries are delineated with vertical gray lines labeled with the construct name on top in *B* and *C*. The original *spot* and *wing blade* boundaries (6) are indicated by a green bar and a blue bar, respectively, for comparison. Color scheme is the same as in Fig. 1. Enhancer cores, defined in the results as the intersection between the 5′ and 3′ dissection curves, are highlighted with a checkerboard pattern in *B* and *C*.

sequences 5′ to the *wing blade* reference segment also contributed significant regulatory information to the *wing blade* activity. When comparing *D0* (the largest construct of the *D* series) with *D5*, in which these 5′ sequences are truncated, we observed an increase of *wing blade* activity of 34% [ANOVA: $F(1, 68) = 56.35$, $P = 1.7205\text{e-}10$] (vertical double arrow in Fig. 2*A* and 5′ curve in Fig. 2*C*). If activities driven by the truncated segment in *D5* (−5,419 to −2,656 bp) and the randomized segment in *E2* (−2,111 to 0 bp) were strictly additive, the phenotypes in Fig. 2*A* would form, conservatively, a perfect rectangle (indicated by four lines in the graph). Additivity would translate geometrically into the addition of the two vectors *ø* to *D5* and *RR* to *E2*, placing the maximum of each activity measured along each direction at the top right corner of this rectangle. Yet, this is not the case, indicating that the sequences contributing to the *spot* activity between −2.8 kb and the core promoter and those contributing to the *wing blade* activity between −5,419 and −2,656 bp are not sufficient to drive the maximum activity. Their effects require the presence of sequences in 5′ for the *spot* activity and sequence in 3′ for the *wing blade* activity, respectively. This is confirmed by the fact that those same sequences show very little to no effect in 5′ dissection for the *spot* activity and in the 3′ dissection for the

wing blade activity. We concluded from this analysis that, although their cores are partially distinct, the derived *spot* activity is largely intertwined in the DNA segment driving the ancestral *wing blade* activity. This strongly suggests that the *spot* enhancer evolved by co-opting the ancestral regulatory segment and raises the possibility that the two enhancer regions share pleiotropic inputs. The notion of enhancer pleiotropy is suggested or discussed as such by several other studies (23–26). In two cases, enhancer pleiotropy was shown to directly result from shared TFBSs in enhancers active in different tissues and at different times of development (3, 27). Although it is unclear whether the *wing blade* and *spot* activities share regulatory information that would result in enhancer pleiotropy, our observations prompted us to explore the modalities of these regulatory interactions further.

In principle, the *spot* and the *wing blade* enhancer, although intertwined, may be functionally independent, with separate sets of intermingled TFBSs. They may on the contrary share TFBSs. In our quantitative mapping (Fig. 1), we noticed that the overlap between the *spot* and *wing blade* activities encompasses a 196-bp fragment (the segment between *D4* and *D5*) (Fig. 1*B*) with interesting regulatory properties. It is indeed necessary for the overall *spot* activity (i.e., any construct missing this fragment displays no spot pattern) (Figs. 1 *B* and *C* and 2*B*, intersection between the 5′ and 3′ dissection curves). In addition, it contributes quantitative information both to the *spot* and the *wing*

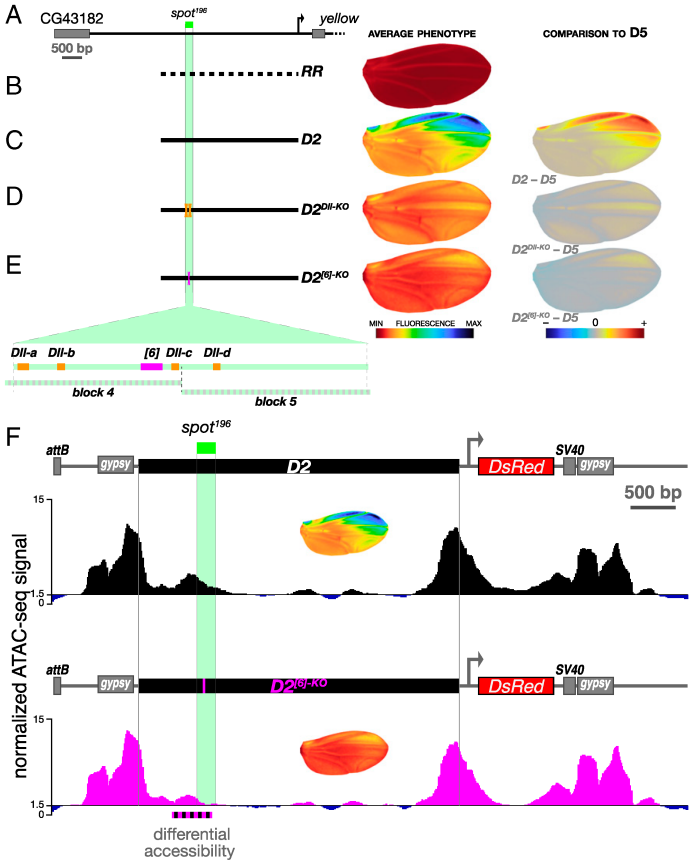

Fig. 3. Shared regulatory inputs of the *wing blade* and the *spot* activities. (*A*) A map of the *yellow* locus 5′ region highlighting the position of the $spot^{196}$ core. (*B–E*) The *wing blade* and the *spot* activities are strongly affected by discrete mutations in *D2*. (*Left*) Construct schematics. (*Center*) Average phenotype. (*Right*) Comparison with construct *D5* phenotype (difference). The positions of mutated sites as well as those of blocks 4 and 5 (*SI Appendix*, Fig. S2) are depicted on blown-up schematics of the $spot^{196}$ core in *E*. (*B*) *RR* is a negative control, the same randomized fragment as in Fig. 1 *B* and *C*. (*C*) *D2* is identical to Fig. 1 *B* and *C*. (*D*) Mutating all four characterized Dll binding sites (16) of $spot^{196}$ in the context of *D2* ($D2^{Dll\text{-}KO}$) reduces the *spot* activity strongly and the *wing blade* activity moderately, as seen when comparing this mutant construct with *D5*. (*E*) Mutating a newly identified activator site (28) of $spot^{196}$ ($spot^{196\ [6]}$, 12 bp mutated) in the context of *D2* ($D2^{[6]\text{-}KO}$) reduces both *spot* and *wing blade* activities strongly, as seen when comparing this mutant construct with *D5*. (*F*) Chromatin accessibility measured with ATAC-seq at the *D2* and $D2^{[6]\text{-}KO}$ transgenes at the onset of *spot* activity (*SI Appendix*, Fig. S3) (66-h pupal wings) differs significantly in a 500-bp region overlapping $spot^{196\ [6]}$ (dotted black and magenta line). This is the only region in the entire locus identified as a differentiated site using diffBind (50, 51) and DEseq2 (52) analyses (*Materials and Methods*) (adjusted *P* value from the DESeq2 analysis: 7.21E-08). ATAC-seq traces represent the pooled signal of three replicates for each transgenic line (*SI Appendix*, Fig. S4). The discrepancy between the enhancer boundaries defined in Fig. 1 and the accessible region of *F* may stem in part from the different stages at which these properties were assessed. Average activity phenotypes of each construct also shown in *C* and *E* are indicated in *Insets* under each construct diagram as a reminder.

blade activities, as we have seen above (Figs. 1 and 2), and is a second enhancer core of the *wing blade* activity. We confirmed this core function of the *spot* activity when we randomized small blocks of sequence (100 bp) overlapping the 196-bp fragment in the context of *D2*. The randomization of the proximal half of this core element (*SI Appendix*, Figs. S1*B* and S2, $D2^{block5}$) reduces the *spot* activity by 61% [ANOVA *D2* vs. $D2^{block\ 5}$: $F(1, 44) = 516.84$, $P = 5.9730e\text{-}26$] without affecting the average levels of *wing blade* activity [ANOVA *D2* vs. $D2^{block\ 5}$: $F(1, 44) = 0.58$, $P = 0.452$]. By contrast, the randomization of the distal half of this core element (*SI Appendix*, Figs. S1*B* and S2, $D2^{block4}$) abolishes the *spot* activity completely and suppresses the nonadditive effects on *wing blade* activity described above [ANOVA $D2^{block\ 4}$ vs. *D5*, $F(1, 45) = 0.025$, $P = 0.876$] (*SI Appendix*, Fig. S2*D*). In previous studies (6, 16), we had analyzed these 196 bp (called $spot^{196}$) because they represented a minimal enhancer to understand the evolution of a spatial expression pattern (not the transcription levels). In particular, we found that this fragment was activated by the transcription factor (TF) Distal-less (Dll) through at least four TFBSs (16), three of which map to the region randomized in $D2^{block4}$ (Fig. 3). In a recent and independent dissection of $spot^{196}$, we identified a potential site for one or more unknown transcription factor(s), $spot^{196\ [6]}$, whose mutation (12 bp) nearly abolishes $spot^{196}$ activity completely (28). It is conceivable that these sites necessary for the *spot* activity also influence the *wing blade* activity, thereby producing pleiotropic effects. We mutated them in the context of *D2* to measure their relative contribution to the *spot* and the *wing blade* activities (Fig. 3). $D2^{Dll\text{-}KO}$ and $D2^{[6]\text{-}KO}$ resulted in strong effects on the *spot* (Fig. 3 *A*–*E* and *SI Appendix*, Fig. S1*B*), and both abolished the nonadditive *wing blade* activity, bringing it to the levels of *D5* (*SI Appendix*, Fig. S1*B*). Mutating the sole site $spot^{196\ [6]}$ in *D2*, along with abolishing 85% of the *spot* activity, also reduced the *wing blade* activity by 44% compared with *D2* (*SI Appendix*, Fig. S1*B*). As a comparison, $D2^{[6]\text{-}KO}$ has a stronger effect on *wing blade* than *D5*, from which the whole $spot^{196}$ segment was removed (Fig. 3*E* and *SI Appendix*, Fig. S1*B*). We were intrigued by these results, as the mutation $spot^{196\ [6]}$ had an effect on the *wing blade* activity only when the rest of the $spot^{196}$ was intact. This suggested that site $spot^{196\ [6]}$ could act indirectly on the *wing blade* activity by preventing, for example, the action of repressors regulating both activities. As the effect on the *wing blade* activity is not observed in $D2^{block4}$, which also randomizes site $spot^{196\ [6]}$, it is likely that sites for repressors acting on both activities are located within the 100 bp randomized in $D2^{block4}$. In our separate dissection of $spot^{196}$ (28), we reached a similar conclusion for the role of $spot^{196\ [6]}$. Even without knowing the molecular mechanism at work, our results suggest that $spot^{196\ [6]}$ could be the target site of a global, permissive activator of both activities in the context of segment $spot^{196}$. They demonstrate that *spot* and *wing blade* enhance transcription from shared, pleiotropic DNA sites. Because $spot^{196\ [6]}$ shows an effect on the *wing blade* activity not observed when mutating Dll TFBSs, we reasoned that the TFBSs for Dll and site $spot^{196\ [6]}$ may convey different information. We have previously shown that Dll primarily instructs the spatial pattern of the *spot* enhancer (16). The global spatial effect of site $spot^{196\ [6]}$, by contrast, suggests a permissive role such as the control of DNA accessibility in this regulatory region. To test this hypothesis, we compared the DNA accessibility of constructs *D2* and $D2^{[6]\text{-}KO}$ using ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) (29) in pupal wings at the onset of activation of the *wing blade* and the *spot* (Fig. 3*F* and *SI Appendix*, Fig. S3). While the genome-wide accessibility profiles of the two transgenic lines were similar, we observed a striking and specific disappearance of the accessibility peak overlapping the two activities in $D2^{[6]\text{-}KO}$ (Fig. 3*F*). These results suggest that the effect of site $spot^{196\ [6]}$ for the *wing blade* and the *spot* activities could stem from its effect on accessibility of a shared segment. We speculate that it could prime *yellow* regulatory activities in the wing by responding to a pioneer transcription factor (30–32), although its sequence does not resemble known motifs (33) of TFs expressed in pupal wings (16).

Discussion

Our results give a molecular snapshot of the evolutionary situation of two enhancers that today are entangled. In the 15 My since the emergence of the *spot* activity (7), the turnover of TFBSs in this region has likely been important, and there is no indication that the very inputs at work today are those involved in the original events of regulatory co-option. Our results, nevertheless, show that the sequences contributing the two activities largely overlap and that at least one site, $spot^{196\ [6]}$, influences both *wing blade* and *spot* activities in the wing. This is, therefore, a characterized case of enhancer pleiotropy. One molecular function associated with this site, as we have shown, is the regulation of chromatin accessibility. We envision the following sequence of events in this regulatory region during development. The regulatory region inaccessible to TFs at earlier developmental stages produces no activity in the wing (Fig. 4*A*). Site $spot^{196\ [6]}$ and probably several other sites, possibly through the interaction with a pioneer factor binding nucleosomal DNA, contribute to loosen local chromatin, resulting in enhancers poised for transcriptional activity (34). After the access to the enhancer sequences is granted, activator and repressor TFs bind to their cognate sites, and the respective enhancer activities start. This general developmental time line (silenced, poised, active enhancer) is supported by numerous recent publications (30, 35). In line with our results, the notion that enhancers control and fine tune their own accessibility is gaining rapid ground (30, 34). The pleiotropic effect of $spot^{196\ [6]}$ and its effect on chromatin opening suggest that, in contrast to the instructive role of Dll (this work and ref. 16) or Engrailed TFBS (6), it may be a site targeted by a pioneer transcription factor (32). As removing this site shows a pleiotropic effect only in the context of an intact $spot^{196}$, we suppose that its role on chromatin opening may give way to TFs preventing global repressors in the $spot^{196}$ acting pleiotropically on both activities.

The question of the evolutionary history of this pleiotropic site is still unclear, and to understand whether or not it is ancestral will require further work. The extensive interweaving that we observed between the *spot* and the *wing blade* enhancers, however, suggests that the evolution of the *spot* activity is tightly linked to the ancestral *wing blade* activity. TFBSs for spatial regulators of an enhancer emerge through random mutations. Mutations in an accessible region resulting in a TFBS for a spatial regulator, unlike mutations trapped in compacted chromatin, have the potential to contribute to a new spatial activity (Fig. 4*B*). In evolutionary terms, this means a shorter mutational path to gaining a regulatory activity (36) and therefore, an increased likelihood (37). Such shortcuts to the emergence of new regulatory activities may explain the apparent prevalence of enhancer co-option.

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 Arnoult et al. (16). We used ϕC31-mediated transgenesis (22) and integrated all constructs at the genomic *attP* site *VK00016* on chromosome 2 (38). The enhancer sequence of all transgenic stocks was genotyped before imaging.

Molecular Biology. Fragments of the *D* series were amplified by PCR from *D. biarmipes* [genome strain (39)] with Phusion polymerase (NEB) and cloned into our transformation vector pRedSA [a custom version of the transformation vector pRed H-Stinger (40) with a 284-bp *attB* site for ϕC31-mediated transgenesis (22) cloned at the AvrII site] digested with BamHI and EcoRI

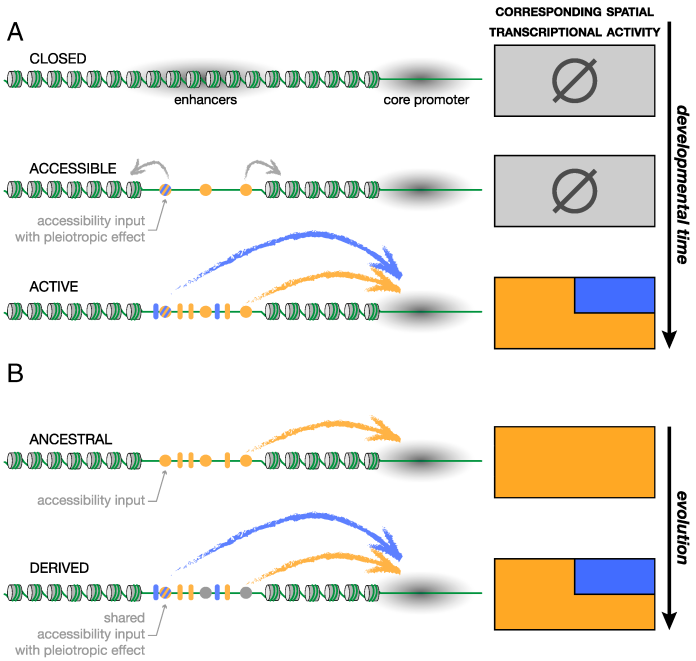

Fig. 4. Developmental enhancer pleiotropy and evolutionary enhancer co-option. (*A*) The developmental progression toward the activation of two interdependent enhancers inferred from our results. High nucleosome occupancy prevents access of transcription factors to the enhancers sequence (closed state; gray shading). Later during development, one or more specific sites in the regulatory sequence (pleiotropic accessibility input; colored circles) determine accessibility across a tissue [for instance, upon priming by a pioneer factor (34)], poising the region for transcriptional activity. Upon binding of specific regulators to their cognate TFBSs (blue and orange ovals), the enhancers become transcriptionally active, producing specific spatial activity patterns. (*B*) A speculative model of the emergence of a new enhancer by co-option. Some of the accessibility sites may be ancestral sites controlling the local accessibility of the regulatory region. During evolution, new TFBSs for spatial regulators, gained in the already accessible region, have the potential to promote the derived activity (blue), unlike TFBSs emerging from mutations in inaccessible regions. In this scenario, the derived activity co-opts an otherwise accessible ancestral activity, creating de facto pleiotropic regulatory information.

using In-Fusion HD Cloning Kits (Takara; catalog no. 121416). The fragment encompassing the four *Dll* sites in construct $D2^{Dll\text{-}KO}$ was synthetized in vitro by Integrated DNA Technologies. The mutations in construct $D2^{[06]\text{-}KO}$ were introduced by PCR through site-directed mutagenesis.

Constructs from the *E* series were produced similarly, but the fragments were made of two components stitched by PCR: a distal part amplified from *D. biarmipes* genome, as above, and a proximal part (dotted line in Fig. 1*A*) amplified from a unique randomized fragment (see below). Likewise, the randomized parts in constructs $D2^{block\ 4}$ and $D2^{block\ 5}$ were amplified from the same randomized fragment and stitched to *D. biarmipes* amplicons.

A randomized sequence was derived from the distal 4 kb of *D0* by randomizing 100-bp segments separately to preserve the local guanine–cytosine content and used for all constructs with randomized sequence. We generated it with an online DNA sequence randomizer (https://faculty.ucr.edu/~mmaduro/random.htm). The 4-kb fragment was synthetized in vitro by Integrated DNA Technologies and used as PCR template to amplify randomized spacers in *E*-series constructs as well as constructs $D2^{block\ 4}$, $D2^{block\ 5}$, and *RR*.

All primers are listed in *SI Appendix*, Table S2. The sequences of all fragments we tested are provided in *SI Appendix*, Table S3. Both *D* and *E* series keep the distance to the core promoter unaffected.

Imaging.

Sample preparation. All transgenic wings imaged in this study were heterozygous for the reporter construct. Males were selected minutes after emergence from pupa, a stage that we call "postemergence," 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 °C until we had reached a sufficient number of flies. Staged flies were processed after a maximum of 48 h at −20 °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 30 min at room temperature in 4% paraformaldehyde diluted in phosphate buffer saline 1% Triton X-100 (PBST). Slides with mounted wings were then rinsed in PBST and kept in a PBST bath at 4 °C until the next day. Slides were then removed from PBST, and the wings were covered with Vectashield (Vector Laboratories). The samples were then covered with a coverslip. Preparations were stored for a maximum of 48 h at 4 °C until image acquisition.

The glue-coated slides were prepared immediately before wing mounting by dissolving adhesive tape (Tesa brand; tesafilm, reference 57912) in heptane (two rolls in 100 mL heptane) and spreading a thin layer of this solution onto a clean microscope slide. After the heptane had evaporated (under a fume hood), the slide was ready for wing mounting.

Microscopy. All wing images were acquired as 16-bit images on a Ti2-Eclipse Nikon microscope equipped with a 10× plan apochromatic lens (numerical aperture 0.45) and a 5.5-M scientific complementary metal oxide semiconductor camera (PCO). Each wing was imaged as a tile of several *z* stacks (*z* step = 4 μm) with 50% overlap between tiles. Each image comprises a fluorescent (TRITC-B filter cube) and a bright-field channel, the latter being used for later image alignment.

z Projection. Stitched three-dimensional stacks were projected to two-dimensional (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 pixel). Both bright-field and fluorescent 2D images were reconstituted by taking the value of the sharpest slice for each pixel.

Image Quantification and Analysis.

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 semiautomatized pipeline. Landmark coordinates on the image were then used to warp bright-field and fluorescent images to match the landmarks of an arbitrarily chosen reference wing by the thin plate spline interpolation (41). All wings were then in the same coordinate system, defined by their venation.

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

Average phenotype images and differences, color maps, and normalization. Average reporter expression images were computed as the average smoothed fluorescence intensity at every pixel among all individuals in a given group (27 individuals per transgenic line on average). The difference between groups was computed as the difference between the average of the groups. Averages and difference images were represented using colors equally spaced in CIELAB perceptual color space (42). With these color maps, the perceived difference in colors corresponds to the actual difference in signal. Color maps were spread between the minimal and maximal signals across all averages for average phenotypes and between minus and plus the absolute value of all difference for the phenotype differences.

PCA. 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*_individual × *n* pixels on the wing). The average phenotype of a construct was described as the average score in the PCA space among all wings of the construct, taking all components into account. Of note, in our calculations, working in the PCA space is equivalent to working directly in the image space. The variance of multidimensional phenotypes in PCA space was measured as the trace of the covariance matrix within each construct. SD was calculated as the square root of this variance.

Overall regulatory information loss. The overall amount of regulatory information lost or modified in successive fragments for each reporter construct series was approximated to the phenotypic distance to the respective largest fragment (*D0* for the *D* series, *D2* for the *E* series) in PCA space divided by the phenotypic distance between the largest construct of the series and the empty construct (*ø*) for normalization purpose. Consequently, while this phenotypic distance is zero for the largest construct, it increases as regulatory information is removed from the enhancer sequence as a result of truncation or randomization. The overall regulatory information loss reaches one when no regulatory information is left (i.e., when a construct has an average phenotype similar to that of the empty construct [*ø*]). A sigmoid curve of equation $\frac{Q}{1+e^{-\alpha*(t-t_{half})}}$, where t is the position along the enhancer sequence, was fitted to the measurements. The amount of regulatory information for each activity was calculated similarly but using *wing blade* and *spot* enhancer-independent measurement (see below) instead of the phenotypic distance described above.

Density of regulatory information per base. The amount of regulatory information brought by a segment of DNA was calculated as the absolute value of the difference between two consecutive fragments, of either the phenotypic distance to the full enhancer for the overall density or the *wing blade* and *spot* enhancer-independent measurements (see below) for the activity specific densities, divided by the differential fragment length. It represents the average amount of information (in terms of fluorescence intensity) per base pair, assuming that it is spread evenly across the modified sequence. To represent regulatory information, be it activating or repressing information, we used the absolute value of the change in the measure of activity, resulting in a similar representation of repression and activation.

Wing blade and spot enhancer-independent measurements. To measure independently the signal brought by the two enhancers, all individuals were projected from the PCA space onto a new two-vector basis, defined by the direction between *ø* and *D5* and the direction between *RR* and *E2*, both normalized to unit length. The coordinates in this two-vector basis represent directly reconstructed values for each activity as two independent measurements. These directions were chosen following the two independent directions of variations observed in the PCA space. Because *D5* and *E2* share 546 common nonmodified nucleotides, this is a conservative estimate of the independent effects in the context of measuring overlapping effect. The difference of expression of either activity between two groups was measured as the difference between the group average of the *wing blade* activity or *spot* activity coordinates described above.

Wing blade and spot regulatory information loss and density. The amount of regulatory information estimated specifically for each activity was calculated similarly to the overall regulatory information loss but using *wing blade* and *spot* enhancer-independent measurements (see above) instead of the phenotypic distance. The density of regulatory information specifically for the two activities was computed the same way as the overall regulatory information.

ATAC-Seq.

Buffers. Buffers for the purification of nuclei from pupal wings were prepared according to the omni–ATAC-seq protocol (43) with some modifications: 1× nuclei permeabilize buffer (NPB) buffer: 15 mM Tris·HCl, pH 7.5, 3 mM $MgCl_2$, 1× protease inhibitor mixture (Roche; cOmplete catalog no. 04693132001), ultrapure water (Invitrogen); 1× lysis buffer: NPB, 1% (vol/vol) Nonidet P-40 (Sigma), 1% (vol/vol) TWEEN 20 (Sigma), 0.1% (vol/vol) Digitonin (Promega), 1 mM dithiothreitol; and 1× wash buffer: NPB, 2% (vol/vol) Nonidet P-40, 10 mM NaCl.

Nuclei preparation. Male white pupa (0 to 1 h after puparium formation) were left to develop for 66 h at 25 °C. Twenty-four pupal wings were then dissected, rinsed twice in cold phosphate-buffered saline, and transferred into 100 μL cold 1× lysis buffer. The wings were cut coarsely into three to four pieces, transferred into a 2-mL Dounce homogenizer (Kimble), and further disrupted by 12 strokes using pestle A. The homogenate was let to rest on ice for 5 min and then further processed with 20 strokes using pestle B. After an additional 10 min of incubation on ice, 900 μL 1× wash buffer was added. A 20-mL syringe and a 20 1/2-gauge needle (Becton Dickinson) were employed to separate cells from the wing cuticle. The mixture was then filtered with a 40-μM strainer (Corning) and centrifuged at 4 °C at 1,000 × *g* for 10 min.

Tagmentation. Pelleted nuclei were gently resuspended in 45 μL ultrapure water and counted using a hemocytometer; 50,000 nuclei were then centrifuged at 4 °C at 1,000 × *g* for 10 min and resuspended in 8 μL 2× Tagment DNA (TD) buffer (Illumina; catalog no. 15027866). The tagmentation reaction followed the previous ATAC-seq protocol (29) with minor modifications: 10 μL 2× TD buffer with nuclei, 2 μL TD Enzyme (Illumina; catalog no. 15027865), 8 μL ultrapure water. The reaction was terminated by the addition of 5× volume PB buffer from the Qiagen MinElute kit, and the library was then purified following the kit's instruction. ATAC-seq libraries were amplified by NEBNext High-Fidelity 2× PCR Master Mix (NEB; catalog no. M0541S) for 9 to 11 PCR cycles and purified by Agencourt AMPure XP beads (Beckman Coulter) with double size selection (0.5× and 2.0×). Bioanalyzer with HS-DNA chip (Agilent) was used to determine the library quality and the final concentration for sequencing.

Sequencing and data processing. The sequencing was carried on an Illumina HiSeq1500 at LAFUGA (Laboratory for Functional Genome Analysis), Gene Center, Ludwig-Maximilians-Universität München, with pair-end settings. The reads for each library were around 50 to 70 million. The sequenced libraries were then demultiplexed, trimmed, and aligned to the reference genome UCSC (University of California, Santa Cruz) dm6 using Bowtie2 (44, 45) with following settings: –X 2000;–fr;–very-sensitive. The aligned reads were then filtered by Picard (46) with the following steps: clean sam, FixMate information, MarkDuplicate. The PCR duplicates were subsequently removed by SAMtools (47). Deeptools (48) was used to obtain the correlation among replicates. Peak calling was performed on three replicates together using MACS2 (49) with the following settings: –keep-dup all; –q 0.01;–nomodel;–shift −100;–extsize 200; –B –SPMR;–call-summits. The differentiated peak analysis was done with diffBind (50, 51) using DESeq2 (52) settings. Three replicates were used for each line. All counts were normalized with the setting bFullLinrarySize = TRUE. All raw and processed ATAC sequencing data have been submitted to the National Center for Biotechnology Information Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under the following accession numbers: pupal wing, D2_66 hAPF_rep1 (GSM4222134); pupal wing, D2_66hAPF_rep2 (GSM4222135); pupal

wing, D2_66hAPF_rep3 (GSM4222136); pupal wing, D206KO_66hAPF_rep1 (GSM4222137); pupal wing, D206KO_66hAPF_rep2 (GSM4222138); and pupal wing, D206KO_66hAPF_rep3 (GSM4222139).

Data Availability. ATAC-seq data have been deposited in GEO (accession nos. GSM4222134–GSM4222139).

ACKNOWLEDGMENTS. We thank Benjamin Prud'homme, Ilona Grunwald Kadow, Miltos Tsiantis, and Marta Bożek for insightful comments on the manuscript. We also thank S. Krebs and H. Blum (LAFUGA at Gene Center, Ludwig-Maximilians-Universität München) for support with sequencing. This work was supported by funds from the Ludwig Maximilians Universität München, The Graduate School of Quantitative Biosciences Munich, Human Frontiers Science Program Grant RGP0021/2018, and Deutsche Forschungsgemeinschaft Grants INST 86/1783-1 LAGG (to N.G.) and GO 2495/5-1 (to N.G.). Y.X. was supported by China Scholarship Council Fellowship 201506990003. M.M. was the recipient of a fellowship from the German Academic Exchange Service.

1. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
2. M. Rebeiz, M. Tsiantis, Enhancer evolution and the origins of morphological novelty. *Curr. Opin. Genet. Dev.* **45**, 115–123 (2017).
3. W. J. Glassford *et al.*, Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. *Dev. Cell* **34**, 520–531 (2015).
4. J. Banerji, L. Olson, W. Schaffner, A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729–740 (1983).
5. G. Sabaris, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with multiple roles: Pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet.* **35**, 423–433 (2019).
6. 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).
7. B. Prud'homme *et al.*, Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).
8. A. Monteiro, O. Podlaha, Wings, horns, and butterfly eyespots: How do complex traits evolve? *PLoS Biol.* **7**, e37 (2009).
9. B. D. Pfeiffer *et al.*, Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9715–9720 (2008).
10. E. Z. Kvon *et al.*, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
11. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser–A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
12. B. P. Berman *et al.*, Computational identification of developmental enhancers: Conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5**, R61 (2004).
13. J. Crocker, D. L. Stern, Functional regulatory evolution outside of the minimal *even-skipped* stripe 2 enhancer. *Development* **144**, 3095–3101 (2017).
14. N. Frankel, Multiple layers of complexity in cis-regulatory regions of developmental genes. *Dev. Dyn.* **241**, 1857–1866 (2012).
15. M. Marinić, T. Aktas, S. Ruf, F. Spitz, An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
16. L. Arnoult *et al.*, Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* **339**, 1423–1426 (2013).
17. P. K. Geyer, V. G. Corces, Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in *Drosophila melanogaster*. *Genes Dev.* **1**, 996–1004 (1987).
18. M. F. Walter *et al.*, Temporal and spatial expression of the *yellow* gene in correlation with cuticle formation and dopa decarboxylase activity in *Drosophila* development. *Dev. Biol.* **147**, 32–45 (1991).
19. P. J. Wittkopp, J. R. True, S. B. Carroll, Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* **129**, 1849–1858 (2002).
20. P. J. Wittkopp, K. Vaccaro, S. B. Carroll, Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547–1556 (2002).
21. G. Kalay, J. Lachowiec, U. Rosas, M. R. Dome, P. Wittkopp, Redundant and cryptic enhancer activities of the *Drosophila yellow* gene. *Genetics* **212**, 343–360 (2019).
22. 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).
23. E. Preger-Ben Noon *et al.*, Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* **22**, 3021–3031 (2018).
24. R. Barrio, J. F. de Celis, S. Bolshakov, F. C. Kafatos, Identification of regulatory regions driving the expression of the *Drosophila spalt* complex at different developmental stages. *Dev. Biol.* **215**, 33–47 (1999).
25. R. B. Emmons, D. Duncan, I. Duncan, Regulation of the *Drosophila* distal antennal determinant spineless. *Dev. Biol.* **302**, 412–426 (2007).
26. J. T. Wagner-Bernholz, C. Wilson, G. Gibson, R. Schuh, W. J. Gehring, Identification of target genes of the homeotic gene *Antennapedia* by enhancer detection. *Genes Dev.* **5**, 2467–2480 (1991).
27. O. Nagy *et al.*, Correlated evolution of two copulatory organs via a single cis-regulatory nucleotide change. *Curr. Biol.* **28**, 3450–3457.e13 (2018).
28. Y. Le Poul *et al.*, Deciphering the regulatory logic of a *Drosophila* enhancer through systematic sequence mutagenesis and quantitative image analysis. https://doi.org/10.1101/2020.06.24.169748(25 June 2020).
29. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
30. J. Jacobs *et al.*, The transcription factor *Grainy head* primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
31. Y. Sun *et al.*, *Zelda* overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
32. K. S. Zaret, S. E. Mango, Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
33. T. L. Bailey *et al.*, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
34. 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).
35. J. Crocker, A. Tsai, D. L. Stern, A fully synthetic transcriptional platform for a multicellular eukaryote. *Cell Rep.* **18**, 287–296 (2017).
36. N. Gompel, B. Prud'homme, The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47 (2009).
37. I. Maeso, J. J. Tena, Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework. *Semin. Cell Dev. Biol.* **57**, 2–10 (2016).
38. K. J. 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).
39. Z. X. Chen *et al.*, Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* **24**, 1209–1223 (2014).
40. S. Barolo, B. Castro, J. W. Posakony, New *Drosophila* transgenic reporters: Insulated P-element vectors expressing fast-maturing RFP. *Biotechniques* **36**, 436–440, 442 (2004).
41. M. F. Hutchinson, Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.* **9**, 385–403 (1995).
42. CIE, *Colorimetry*, (CIE Central Bureau, Vienna, Austria, ed. 4, 2018).
43. M. R. Corces *et al.*, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
44. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. B. Langmead, C. Wilks, V. Antonescu, R. Charles, Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
46. Broad Institute, Picard (2019). https://broadinstitute.github.io/picard/. Accessed 3 September 2019.
47. H. Li *et al.*; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
49. Y. Zhang *et al.*, Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
50. R. Stark, G. Brown, DiffBind: Differential binding analysis of ChIP-Seq peak data. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf. Accessed 6 September 2019.
51. C. S. Ross-Innes *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
52. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

PNAS
www.pnas.org

Supplementary Information for

An ancestral and a derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting of chromatin accessibility

Yaqun Xin, Yann Le Poul, Liucong Ling, Mariam Museridze, Bettina Mühling, Rita Jaenichen, Elena Osipova, Nicolas Gompel

Paste corresponding author: Nicolas Gompel
Email: gompel@bio.lmu.de

This PDF file includes:

Figs. S1 to S4
Tables S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.2004003117

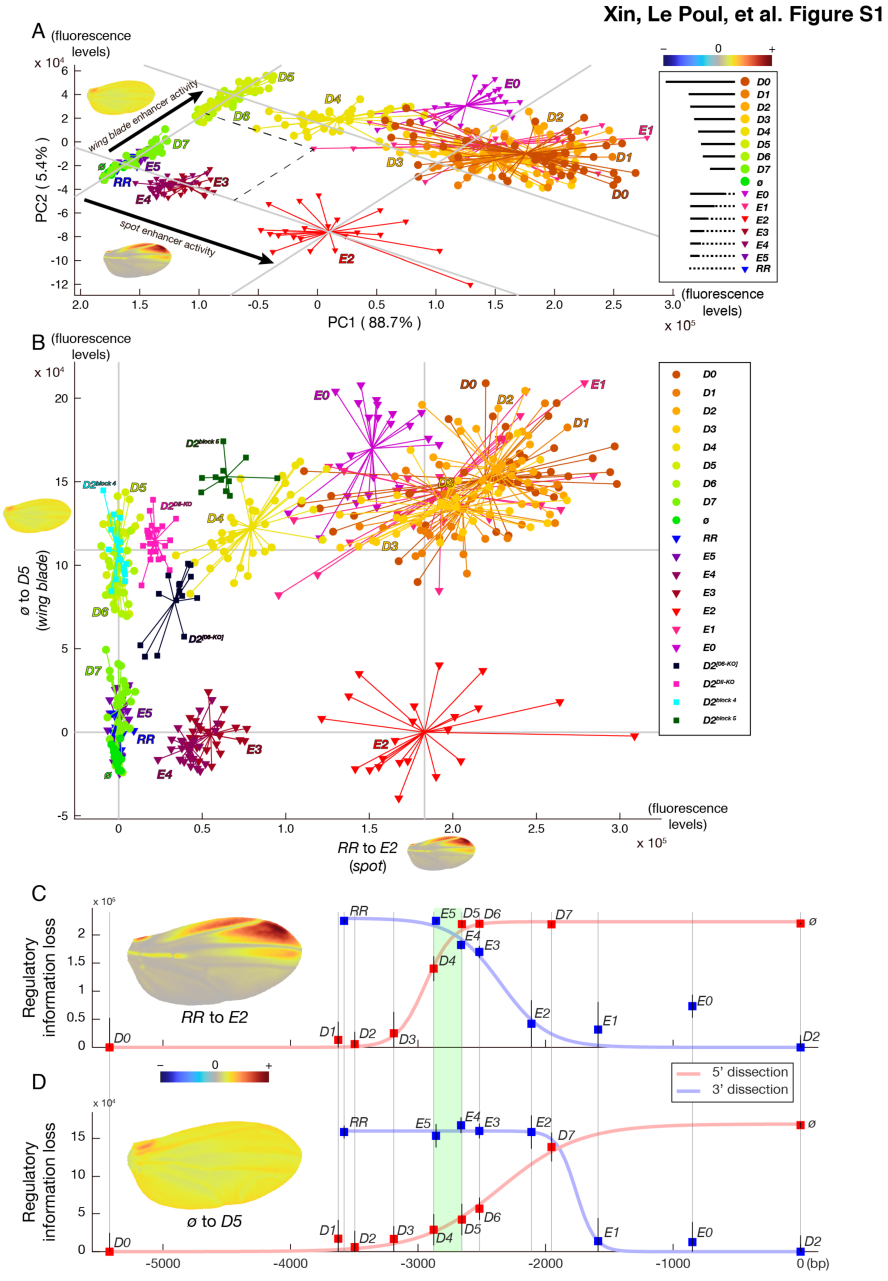
Xin, Le Poul, et al. Figure S1
A
(fluorescence levels)
x 10^4
PC2 (5.4%)
PC1 (88.7%)
x 10^5
wing blade enhancer activity
spot enhancer activity
B
ø to D5 (wing blade)
RR to E2 (spot)
C
Regulatory information loss
RR to E2
D
ø to D5
5' dissection
3' dissection
-5000
-4000
-3000
-2000
-1000
0 (bp)

Fig. S1. Variation in reporter expression across all transgenic lines. (A) PCA of activity variation for constructs of the D and E series (Fig. 1). Black arrows identify 2 directions of variation in the phenotypic space that correspond to the *wing blade* and the *spot* activities, respectively. Wings with colormap (average phenotype differences between *D5* and *ø*, and *E2* and *RR*, respectively) illustrate the corresponding phenotypic variation. We defined a 2-vector basis with these two independent directions, in which we projected each individual wing phenotype (black dotted lines indicate the projections) to produce panel (B) (below) and Fig. 2A. (B) Projection of PC1 and PC2 from (A) in the new 2-vector basis showing in addition to all D and E series constructs the following mutants: $D2^{block\ 4}$, $D2^{block\ 5}$, $D2^{Dll\text{-}KO}$ and $D2^{[6]\text{-}KO}$. (C, D) Loss of regulatory information along *yellow* 5' region (fluorescence levels, as in Fig. 1D) for each direction defined in panel (A).

Xin, Le Poul, et al. Figure S2

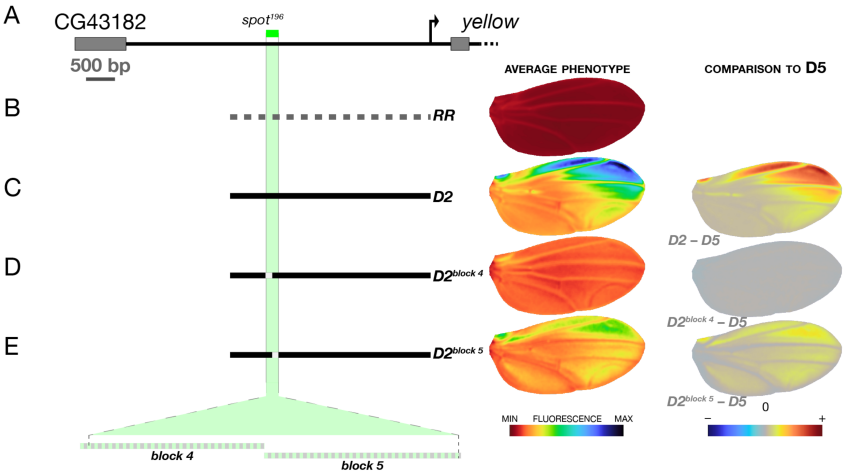

Fig. S2. Reporter activity in the wing for constructs *D2*$^{block\ 4}$ and *D2*$^{block\ 5}$. (A) A map of the *yellow* 5' region highlighting the position of the *spot*^{196} core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 4) and the proximal part (block 5) of the *spot*^{196} core in in *D2*. Left: construct schematics; middle; average phenotype; right: comparison (difference) to construct *D5* phenotype, which drives partial, uniform *wing blade* activity. The portions of randomized sequence are depicted on a blown-up schematics of the *spot*^{196} core under panel (E) with dashed green lines. (B) *RR* is the same negative control, a randomized fragment, as in Fig. 1. (C) *D2* is identical to Fig. 1. (D) *D2*$^{block\ 4}$ abolishes the *spot* activity and strongly reduces the *wing blade* activity. (E) *D2*$^{block\ 5}$ reduces the *spot* activity and has a milder effect on the *wing blade* activity.

Xin, Le Poul, et al. Figure S3

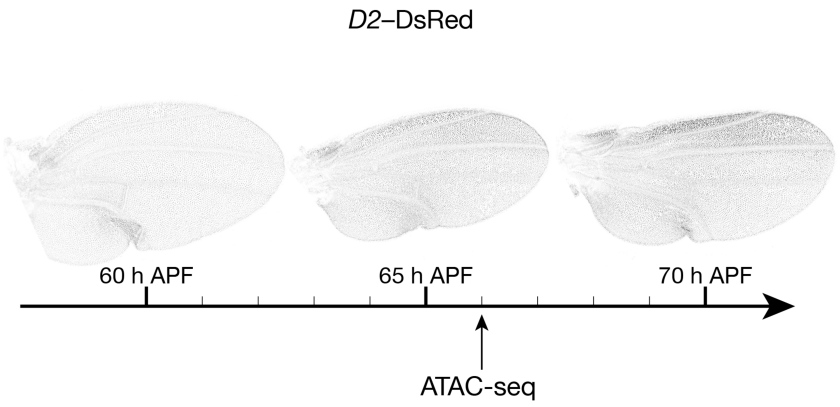

Fig. S3. Reporter activity in the wing for constructs *D2*$^{block\ 4}$ and *D2*$^{block\ 5}$. (A) A map of the *yellow* 5' region highlighting the position of the *spot*^{196} core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 4) and the proximal part (block 5) of the *spot*^{196} core in the context of *D2*. The portions of randomized sequence are depicted on a blown-up schematics of the *spot*^{196} core under panel e with dashed green lines. (B) *RR* is the same negative control, a randomized fragment, as in Fig. 1. (C) *D2* is identical to Fig. 1. (D) *D2*$^{block\ 4}$ abolishes the *spot* activity and strongly reduces the *wing blade* activity. (E) *D2*$^{block\ 5}$ reduces the *spot* activity and has a milder effect on the *wing blade* activity. (F) The differential effects of *D2*$^{block\ 4}$, *D2*$^{block\ 5}$, *D2*$^{Dll\text{-}KO}$ and *D2*$^{[6]\text{-}KO}$ and are best seen when subtracting the uniform *wing blade* activity of *D5*.Type or paste caption here. Create a page break and paste in the Figure above the caption.

Xin, Le Poul, et al. Figure S4

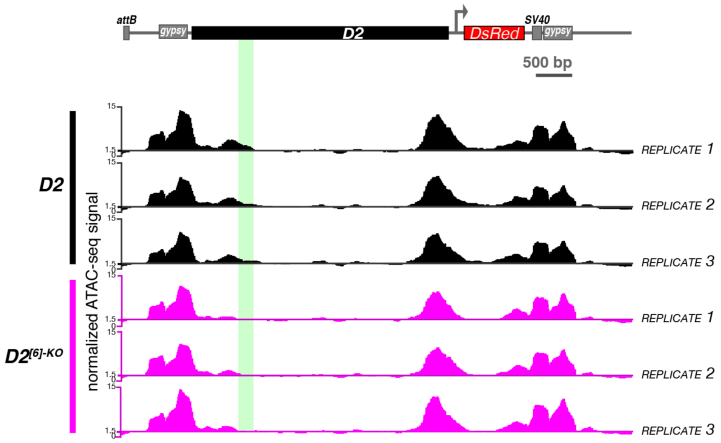

Fig. S4. ATAC-seq replicates shown separately. Chromatin accessibility at the *D2* and *D2*$^{[6]\text{-}KO}$ transgenes at the onset of *spot* activity. The 6 ATAC-seq traces represent 3 technical replicates for each transgene that were pooled for each genotype in Fig. 3.

Table S1. Results of MANOVA between selected pairs of constructs on the 10 first PC explaining 97.4% of variance.

Genotypes	Df	Pillai	Approx. F	Num df	Den df	P
(D0, D1)	10	0.704254	13.3352	10	56	1.43856e-11
(D0, D2)	10	0.648637	11.6302	10	63	5.16992e-11
(D0, D3)	10	0.821631	26.2562	10	57	0
(D0, D4)	10	0.932406	88283	10	64	0
(D0, D5)	10	0.959188	138664	10	59	0
(D0, D6)	10	0.955799	125418	10	58	0
(D0, D7)	10	0.962682	144462	10	56	0
(D0, ø)	10	0.975699	212.8	10	53	0
(D0, D1)	10	0.704254	13.3352	10	56	1.43856e-11
(D1, D2)	10	0.485949	4.91573	10	52	5.50417e-05
(D2, D3)	10	0.795837	20.6596	10	53	6.32827e-15
(D3, D4)	10	0.928855	70.5011	10	54	0
(D4, D5)	10	0.934419	79.7903	10	56	0
(D5, D6)	10	0.357372	2.78055	10	50	0.0082051
(D6, D7)	10	0.912014	48.7176	10	47	0
(D7, ø)	10	0.689185	9.31285	10	42	6.8268e-08
(D2, E0)	10	0.970208	153062	10	47	0
(D2, E1)	10	0.887467	41.7975	10	53	0
(D2, E2)	10	0.966521	132801	10	46	0
(D2, E3)	10	0.98647	342686	10	47	0
(D2, E4)	10	0.987002	379671	10	50	0
(D2, E5)	10	0.98802	404126	10	49	0
(D2, RR)	10	0.982305	194301	10	35	0
(D2, E0)	10	0.970208	153062	10	47	0
(E0, E1)	10	0.946754	72.9004	10	41	0
(E1, E2)	10	0.935671	58.1807	10	40	0
(E2, E3)	10	0.970607	112275	10	34	0
(E3, E4)	10	0.818479	17.1342	10	38	3.45354e-11
(E4, E5)	10	0.949004	74.4369	10	40	0
(E5, RR)	10	0.45029	2.04785	10	25	0.0710936
(ø, RR)	10	0.829743	12.1836	10	25	2.57409e-07

Table S2. Primers used in this study. Small letters in the sequences denote adapters for in-Fusion cloning, sequence in red denote mutations introduced in the wild-type sequence.

Primer name	Primer sequence	note
D0-Forward	gagcccgggcgaattAAGAGCCCAAGGTCGGTCG	small letters denote adapters for In-Fusion cloning
D1-Forward	gagcccgggcgaattGCCAGCCAGTAATCTAAAAAGTCG	Capital letters denote *D. biarmipes* genomic sequence
D2-Forward	gagcccgggcgaattCCTTAGTTCAGAAGGCGCCTGC	
D3-Forward	gagcccgggcgaattCGATAATCGCCCGATTACCGC	
D4-Forward	gagcccgggcgaattCCAGCGCCTTGTATTGGCATC	
D5-Forward	gagcccgggcgaattGTGCCAATCATTTTTAGTACACCC	
D6-Forward	gagcccgggcgaattAGGAAATCTCTTTCTCGGCCTGT	
D7-Forward	gagcccgggcgaattGCCGCCGAAGTCTGGCAAA	
D-series-Reverse	gctcctcgagggatcTTCTGTGGACCGTGGCGC	
E-series-Forward	gagcccgggcgaattCCTTAGTTCAGAAGGCGCCTGC	
E0-Reverse	gctcctcgagggatcTGCCAATACAAGGCGCTGGC	
E1-Reverse	gctcctcgagggatcGCTCACCCACAGAAAAGTAAG	
E2-Reverse	gctcctcgagggatcCATAATTGCCACACGATTATG	
E3-Reverse	gctcctcgagggatcTGCTCGCAGGCGCGCACT	
E4-Reverse	gctcctcgagggatcATTGCATGGGCGCACATC	
E5-Reverse	gctcctcgagggatcTGCCAATACAAGGCGCTGGC	
Randomized Spacer-Forward	gagcccgggcgaattACAAAACCGGCGGACAAATCCT	
Randomized Spacer-Reverse	gctcctcgagggatcTGGTCCGGTCCGTCTAACC	
$D2^{Dll\text{-}KO}$		
fragment1-F (D2-Forward)	gtctagagcccgggcgaatt CCTTAGTTCAGAAGGCGCCTGC	
fragment1-R	CCTTAAACGGAATcgcgAGATGCCAATACAAGGCGCTGGC	
a fragment encompassing the 4 Dll site mutations was synthesized from IDT :$D2^{Dll\text{-}KO}$	TCTcgcgATTCCGTTTAAGGACGCgcgTTCTGAGCTAAAACTCGCTTATG GAGAGATCTAAATTTCCCCGCTTTTGGCTTGAATAAAcgcgTCGAATTCC CCGCTGGCTcgcgAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTA AATTGCAAATTGCTCAATCCGCC	
fragment3-F	TGCAAATTGCTCAATCCGCCT	
fragment3-R (D2-Reverse)	gctcctcgagggatcTTCTGTGGACCGTGGCGC	
$D2^{Dl\text{-}KO}$		
fragment1-F (D2-Forward)	gagcccgggcgaattCCTTAGTTCAGAAGGCGCCTGC	
fragment1-R with mutation adapter	TTATTtagtgtaaagatGGGGAAATTTAGATCTCTCCAT	
fragment2-F with mutation adapter	atctttacactaAATAAATTAATCGAATTCCCCGCTGG	
fragment2-R (D2-Reverse)	gctcctcgagggatcTTCTGTGGACCGTGGCGC	
$D2^{knock\text{-}4}$		
fragment1-F	gtctagagcccgggcgaatt CCTTAGTTCAGAAGGCGCCTGC	
fragment1-R with mutation adapter	TAACAGACACCGAGGGAAAG CGCTCACTGCGCAAAAAACTAAGC	
fragment2-F with mutation adapter	CTTTCCCTCGGTGTCTGTTAA	
fragment2-R	TTTAATAGCCAGCGGGGAAT TCAGATGATAACGAAGTGGGAG	
fragment3-F	ATTCCCCGCTGGCTATTAAAACAC	
fragment3-R	gctcctcgagggatc TTCTGTGGACCGTGGCGC	
$D2^{knock\text{-}5}$		
fragment1-F (D2-Forward)	gtctagagcccgggcgaatt CCTTAGTTCAGAAGGCGCCTGC	
fragment1-R with mutation adapter	GATCCGCAACATGACTCATG TCGATTAATTTATTCAAGCCAAAAG	
fragment2-F with mutation adapter	CATGAGTCATGTTGCGGAT	
fragment2-R	GTACTAAAAATGATTGGCAC TGCTATCAATATTATTCCAAGTGG	
fragment3-F	GTGCCAATCATTTTTAGTACACCC	
fragment3-R (D2-Reverse)	gctcctcgagggatc TTCTGTGGACCGTGGCGC	
pRedSA sequencing primer		
B21-F	gcttggctgcatcacgtaat	
B22-R	gtcctcgaagttcatcacgc	

Table S3. Construct sequences.

>RR
ACAAAACCGGCGGACAAATCCTACAAACGGTCACTTGCAATCCCTTTACTCCCGGCAATACTGCACGCGTG
ACTAATTCTCATACTCGTTGCCAATATCCTCCACACTCCGCACCCTCGATCTCAGTAAACCGGACATGGGG
CCTATTTCAATAGTTCCGAGCGTGACCAGTTTTTTGTCATTCTTGAGCGAATTTAAGGGACGGCTACCGCA
AGTACGCGTCTTTATCAGTAGGGATTTCAGTCGTTTGCGTACGCTATGTGGAGTAGTGGACCCCCGTGCTC
TCAGTAGCTGTAAATAACCTGGTACGGTGATGAATTTGTTATTTACCTCAGTGGAAATACTCAGTGCAGTA
AGAGCTGTCGTAGTGGTCTTCGGATGTGGTTCGCAGAGTTATGTCCTGTGTCAGCGTAGATCGTTAATAAC
GCGAACGGTGGCCCACCCACAATAATGTGGGAGGCCACATACATGCTCACCATCGAACTCACGATTGCTGT
GGTATACGTTTAGGGCGACCGAACGGAGTGTTAATGGTTATCCACCCGGCTGCCACATGATTCAGTGCGAT
GCGTACCCGAAGATATTCTCGGGAACCCAAGTAACTTATCGACATTTAATTGGCGGTTCCCCTTGCAATGC
CTGGGTTATATAAACGACATATATTTATGTTGGGTCTTCTATACTCTGCGTAACAAGTAAATCCTTTCCCT
CGGTGTCTGTTAAAGTGGAGAGGATCCTGGTTCCTGTGGCCAGCATTCCGTAGGATGTATAATAACGGCGC
ACAGAGCCTTCCAACCTCCCACTTCGTTATCATCTGACATGAGTCATGTTGCGGATCTTCAAGGTAACTCA
TACACATCATTCCATAATGACTTTGATGACCTCATCGCTTTTAGTCGCCCACTTGGAATAATATTGATAGC
AGTGCATTATTGAGCCTTATCCGAGAGCCTCATAGGTGGCGTTTATATCTTGAGTACATATAAGTGCGCGG
CGTAGGGTTCACCATCTAAGTGCTGAAGCTAGGGGGTTCGCCGCGCCGGGCACGAGGATGGCAGTCTCGGG
GCCACGTGCATTAAGCTCGGTAACTCGGATGAAACTCGGTACCAGTGAGTCCACGTCAGGCACGGGAGTGC
ATATGGGGTATGAAGGTCTACAGAGACGCTCAATATGATTCTTCACCAATCAAGTTTAGAGAAATAAAGAT
CGATATTTGTGGGGCACGGGTGTATACAGGATCTAGAGGACCGAATAAAATCCGCTTTGTGCGTATGCCGA
TGCGTACCCGGCGAACAAACTACCTGAGAAAGCTTGTGGGACGCCTAAGTGAGTACTAGCGGTCAAAAAGC
CAGACGGGGTGACGCCGGGAAAAGGACCCGCCTGTTAGAGCACTTATCTTCTGTTCTGTGACATTTCAAGG
GCTGGAATTCTCTGTTGGGACTCGGATCCTCTATGCCCGCATACGCCTCGTTGATGTATCCTGATTAAGCT
AGGTCTCTTTGAACTTGTGCAGCTCCACGGGATAGCCGAACGTTTCGGAGTTTGTGTGTCTTTCTTCCATA
TGCTTCGTGTAATTACATTTATTCCACAAACAATAAAATAGAGGGGACCTGTCTAAAGAACAACACATGGC
AAAGTGGGAATACAACCAGAAAAGTGGTCCAATAAACAAAGAACGTGAATCACTCAGGAATGAGAACCATC
GTGAGCCTTCAGCAACAATTACCCATGGCATCTAAATGGCGAGTACTTTACAACGCCTGACAAAGATAGCT
TACGAATCATGTGACGCGAGTATCAATAATTTTGTATGAGTCTCACCCAGATTCTGATCCGCCGTTAAGCT
CACCCGTTAGGCAAACTCTTTGGCATCGATGGTAGTTAGCTCCATGTAAACAATTCTTACTAGAGGTAGGC
CCAGCGTGCGCGCGCTTACCTATTGAGGGTTTGATCGCCCTTTAGTAGAGTCGGGGTCCGGCTTCAGGTAT
CGAATAGATGATCTGCTCTTGGATAGTGGCTGACAAAAGTACTAAGGGAAATCTTTATCCTTATACTAAGT
CCGAGGACAGTGGGTATAGACATGGGAAGCACTCACTCAGATATGTATAGACACAGCAAATCGTGTATTTA
AAATCAATTCCTTAGATTATTACGAAAAGATAAGAAACAGGGCAACACGACATTGGAACAGTTACGTAATT
GCGTTTTGCGGTTATGCAAAGTTTCATTTCGCTAGGAGTGTGGCCCATATGAGATTCATCCGCTTCTTCTC
GACGTGAGAGGGACCCCAGCTTTGCTTCTCTTAACAATAATCCTGTCGCTTAGGCAATTTTATGGTGTCGT
GTTTAATTGTTCAATACATCACATCCACGCTTATTCCTGTCGCTCCGCTCCATTTTATGCTCGAATCCAGC
GTAGGCGGGTCATATGTGCCTCTTTATTGAGCAGTTGGGAGGTTCGTTTCTCGGGAGATGTCGTTTTGACC
ATTGCCGATCCTCGGCGCCAGAGCGTGGAAATACCCCGACATCTGGCTTTCAACTTGGCAAACCTCAGGGA
GTCGTCATCACTTGACCTAGAACCCGGTCGGGGGCACCGATTATCTGCCTTAGCCATTTCCGGTTATGGCG
GAGCATATTATCAACGCGTGAAAATTCTGGCCACGTCTTACAAAGTGTCAAATGGGAGCCCGAACGCCCAC
CGAGGCATACGATGCTGGCTAGCTCGTACCATTTATGAGATGAACTAATCTGAACCTCACCTACTGGGTAA
TACGAATAAGTCACTATACTaAAATACTTTTCGGTAGCGAATAAGCTATATATTACATAAATAAATGAGCT
CAACAACCAGTGAGATGCACCAGGGGATTGGTTAAAAGCTCGACCGACTGCTTTTTTAAATCGATGGAATA
TACTATCTTGAATCTTACATAAATTATTATGACAAATGACTTGAATTTAATGTAAAAATTATTTTACGTTA
AAAAATTATAATGATGTCCAATGCAACCTAAGTCGCGAAGACTCAGAATCAACGGCATCGGGGTGAAAGTT
ACAGTTTTACAGATACCGAGATCAATTTTCCGTTTAATAAGTTGACTTCTGGTACACTAAACCAGTTAATG
TTATAACAAAGTATGAAAGTTTGTAGACAAATAGCCTAGAACCGGACTGAGTACCTAATAATGTCAACGCG
CGAGCAATCAAGATCTCGAAGACGAGGCGTACGCATTGCCCCGAGAGGCCATCTCAGGCGCGCACCCATGG
TCCGAGAGCAACCCAGATTCAGAGTCAACCAAGAAGACGAGCCTAGCGATCAATTAAGTAAGGTAACTATA
GATGAAACCAAATGGTATTGGAATGCAAAGCCAAACAACGGATGATCCTGACTGCACCCGAATAGAAGTTC
GTATGGACACAATGCGCTATCACCCGCCTGGCTGACGATCGCACGCCCCTAGCCCCCTCAAAGAGCAAGCT
GCGGTGATAACTGTACTGATGGCACTCTAAACCTGAACCCCGTTTCCGTCGTAACGAATGTATCAAAAACA
CGGCAACGTTCGGTTAGACGGACCGGACCA

>D0
AAGAGCCCAAGGTCGGTCGTTTAGCTTGGCCAAAACCTACCCATCCAGCTGGCACTTCCACCAACGGCACC
AAGACACGAGCGGAAAATAAAAAGCCACACCACCCCACTTAGAACTCCGTTTAGCAGCAGTTGTTCAAACA

GAAATTGGCTGGCTTCGGCCGGTTCAGCcTCAGTTGATAATTATTATAATATCTATGTTCTTGCCTATCGC
TGGGCCCTAATTGGCCCAGACAAAGGCACCGTTTTTATGCATAACTGGAGGCTTACAATTTGGCcTTCGAC
ACGCGGTCCTTCACATTGCCAAAAAAGAACGAGAACTCgGCAAGCCAATTACACTCGAAGAAGCAGCGGGG
ATCGTTTCGATGCCTAGCCCTGGGCCAGTTCAATCACTCCCGCCGATAATTAGCCGGCCTCTGCAATGGGA
ACTTTTCCAAGACGAGATCGATTCTTGGGAAAACACCACCTCAGTTTCCATTTCTGTTTTTTGGGTCCGGA
AGTGGCATCGTGTCTTCCCAGAAGCGCCTCCAAATGGTGCCACCATTAGCCAGGGGGAAGCCGGGCGGCAG
TCACTTAGCTGCTGCAATTTAAATACTTTTTAATTGATTACTAATTGCGGCGAGGCAAGTGCCAAAACAAG
ACGACGAGGACGACTTGGCTGTGACGTTTTCGATGCAACCCGACCGGGGACTGCCACTCTTTAGCCAGTTA
ATTGGCAGCAAAAGCGACAGTGGCAGCGGCAGCAACTGCTTTTCACCAGGAAATCAATAAACGCTCGTCCA
GCGGCAAAAGTAATCGCAACACGCACATCTCAATTTCGGTGGCAGAAAAAAAACCCTCACCAGCTCAGTTC
CCCGTGCCAAATTAACCAGAGCAAACATAGCCCAGTTTCTTCTCCTGCGGCATGTGAAAAGGCAAACAGTG
CTGGCCGGAAAATACCCAGCAAAACACCTGAGTTCTAGTTGCGATTTTCGGAATTGGACTATAAAAGGCGG
CCGTCGGGTAGCTTTCTTCACTCACAACCAGTCCAAAAGCATCTCCAACATGAAGTTCTTCCAGCAAATCG
TCCTGGGTCTGATGCTCGTCCTGGCCATCATGGGTTCGCTGGCTAGCGCCAAGCCCCAGGAGGCCGAGGAA
CCGGAGGAGTCGCTGGTCGGGGACTCAGAGTCGGGACAGTCCGTGCCCGAGGACGCCCAGCAGGACTACCT
CAACGTGGCGGACCTCACCACTGCCGCTCCTACATGGTGGTGGAACTAGAGCCCGGGAATTCTAGAGGATT
TTAACAATCTTTTGTTTTTTTTTGCTAATCTAATGTACTATAATTGCAAAGAATTTACGGTGTTCCATATT
CAATAAACCTATTTAAAGCTGAATACAATATTTACGTTAATAAATGTTCTTGATACGATAAATTTACTTAA
TTAAAATACATTACATTTCAGTTAAATATTTGTAAAATAAAATATATTTAAAAAAATATTTAATTCACTAG
TTGTGGGAGTTCATTAGACTTATTATTTGTTTTTATTAAATTGTAATTTGTATCAAAAGTTTATTTTGCCA
AACAGTGAATCTTAAAATATATATCAAGTTCATTGCACAAATTAACTTATAAATTGTCACCAAAAATTTAG
AAATCAACCTATGTAAATAAATTTAACAAACCAATCATATCTTGAATTTAAATATATAAAAGAGGAGATAA
ACCATTTATAAAATGGTCTCACCTTTTTTTAGTTTATTTGATGCATGTTTTAATTTTGCTAAAATCATATT
CTGATGTCTATTCATTTTGCCAGCCAGTAATCTAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTCG
ACCTATAAATACCCACCGCAAATGCCGCAACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAATGGAC
AATTATCCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGA
AATTAATCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGA
AGGCTGCCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCT
GCGGAACGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTG
CCTTGGGGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCA
TAAGAAATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCA
CGGATTTATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCT
GAGACAAGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAA
ATCAAATTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTAT
TGGCATCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTT
CCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGT
CTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCA
ATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTG
GACTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTC
TTTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGG
TCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTG
CTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGG
GGGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACAC
GGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAAT
GTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTA
CTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTT
TTGTTTTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCT
GGCAAAACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGC
ATCCCACACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAA
ATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCA
ACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAA
TAAATTAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCG
AGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCAT
AAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAAT
AAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAA
TATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGC
CGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTG
CCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGC

10

TGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGC
CATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAAC
TTACGGCCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTG
TGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTG
AATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTG
TTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACAT
TTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATA
TTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAA
AATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTC
AATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAAT
CAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAA
AGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATT
CGGGGGCCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACT
AGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGA
CCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCA
TTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCG
TAAGCGCGCCACGGTCCACAGAA

>D1
GCCAGCCAGTAATCTAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTCGACCTATAAATACCCACCG
CAAATGCCGCAACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAATGGACAATTATCCTTAGTTCAGA
AGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAATCGAGCCCGTAA
ACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTGCCAATTGTGGTG
CCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAACGGGATGATGGA
CCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGATCGCTGC
GATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGCGGCGATGG
CGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTATGCACTCGCT
GGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAGTGCAGCGCGG
CAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAATTACGGCAACCT
CTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCATCTAATTATTCCG
TTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGCTTTTGGCTTGAA
TAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTG
CAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATTTTTAGTACACCC
CTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAAGAA
GACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTCGGCCTGTCTGAT
TTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGACATTG
AACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATGGGTGTC
TCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAGAAAATAAAAAGG
AAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCGGGCAATGTCACT
CGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAATGTTGCATGCCG
GTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCTGCCATGTTCCGC
TGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTTTTCGCCGGGACA
AACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAAACAGCAAATTGA
AAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCACACGACGAGGTG
AAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCAGGTTCAACGGCT
GCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGCACACACAATTAA
TTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATTAAGGCAGCATTA
GTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCAGAATGCGGCCAA
GGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATCGGCAACGCATAA
GTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATATAAATTTATGGCA
GAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTGTTTTCCCCAGCA
TTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGTGTAGCTGGTGGG
TCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATTTGGCTCGTTCAT
TATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGA
GTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAGAGCATTACCCGC
TTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGGCCAGCTGAGTGG
AGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGATACATATGCCAAA
ATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCATCaGAGCTGCCCA

ATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCAAAGACCTGTCCC
AGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATATCGTATTATATTG
GTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCCACAGAAAATGTG
TGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAATATTATGAATAAT
CAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAAGACTCGAAAATA
CTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTGAATCCTAAACGT
ATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCCTTTCAAA
ACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGGCCCGTGATACTC
GCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGAAGCCAAAGTGTC
CGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCACCCGAGCCACCG
AAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCCTGCCTTCGTCTT
CGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCGCGCCACGGTCCA
CAGAA

>D2
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTG

AATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCC
TGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCG
CGCCACGGTCCACAGAA

>D3
CGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGCGGCGATGGCGGCTTTAGA
TAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTATGCACTCGCTGGAAAAGCTA
TTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAGTGCAGCGCGGCAGTTGCAGG
CAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAATTACGGCAACCTCTATTCACTT
ATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCATCTAATTATTCCGTTTAAGGACG
CAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATC
GAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATTTTTAGTACACCCCTAACTGGTG
TTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAAGAAGACGATTGTG
GGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAG
ACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGACATTGAACTTGAAAA
CGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATGGGTGTCTCTGGCTAAG
GTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAGAAAATAAAAAGGAAAATGTTCA
GGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCGGGCAATGTCACTCGGCATTCGC
TTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAATGTTGCATGCCGGTTGCATAAT
CGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCTGCCATGTTCCGCTGCCACCGTA
TTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTTTTCGCCGGGACAAACTCGTTTC
GATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAAACAGCAAATTGAAAACTGCAAG
CTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCACACGACGAGGTGAAAATGAAAA
TAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGATC
GCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGCACACACAATTAATTAATAAGCG
TCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATTAAGGCAGCATTAGTCGCAATTA
TGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCAGAATGCGGCCAAGGGATCGAGT
TAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATCGGCAACGCATAAGTGGCGAAGC
ATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATATAAATTTATGGCAGAACTTAAGA
AACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATA
TTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATT
AATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTG
GTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAA
TCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAGAGCATTACCCGCTTAGGGCGCC
CTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGGCCAGCTGAGTGGAGTGCTGGAA
CGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCT
CATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCATCaGAGCTGCCCAATCGAAATCA
AAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCAAAGACCTGTCCCAGATACTCTG
TTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATATCGTATTATATTGGTAACTGCAG
CAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATT
AAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAAT
CAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAAGACTCGAAAATACTCTCACTCA
CTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGC
CAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAA
AAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGGCCCGTGATACTCGCACTTAATA
AACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCC
GATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCACCCGAGCCACCGAAAGCAGCCG
AATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAA
AACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCGCGCCACGGTCCACAGAA

>D4
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCG
CTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTT

TCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCAT
TTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTT
ACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCT
CGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGG
GCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCG
CGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGA
GAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCC
GGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTA
ATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTC
TGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTT
TTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAA
AACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCC
ACACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGC
AGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAG
CACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAAT
TAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGC
AGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAAT
CGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATA
TAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTT
GTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAG
TGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCAT
TTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAG
TTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTA
GAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACG
GCCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGA
TACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCA
TCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGC
AAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATA
TCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGC
CACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAA
TATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAA
AGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCT
GAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACC
ATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGG
GCCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAG
AAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACC
ACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGC
CTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGC
GCGCCACGGTCCACAGAA

>D5
GTGCCAATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTT
TTGGTGGACTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGA
AATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACC
CGGCGGTCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGG
CAAGTGCTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAA
CACTGGGGGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATG
AAACACGGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTA
AATAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTAT
TTTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAAT
GATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCG
AAGTCTGGCAAAACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAA
CACAGCATCCCACACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCG
ATTCAAATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCT
TATGCAACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGA
TCGCAATAAATTAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGT
GAGCCGAGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATT
TTGCATAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTT

AAAAATAAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCC
TAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTC
CTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTA
TCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGC
ACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGC
CATGGCCATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTAT
TGCAACTTACGGCCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGC
TGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGG
ACTGTGAATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTA
GTGCTGTTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTAT
TTACATTTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACA
AACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAA
GCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATG
GTATTCAATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCAC
ATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAA
GACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAG
GGGATTCGGGGGCCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCA
CGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCAT
GGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCG
CTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAA
CAGTCGTAAGCGCGCCACGGTCCACAGAA

>D6
AGGAAATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGG
GACCCGGCGGTCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGA
ACGGCAAGTGCTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGG
CCAACACTGGGGGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAA
CATGAAACACGGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTG
TGTAAATAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTAT
TTATTTTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGA
TAATGATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCC
GCCGAAGTCTGGCAAAACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAA
CAAACACAGCATCCCACACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAA
TTCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGG
CTCTTATGCAACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTG
CCGATCGCAATAAATTAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGT
GGGTGAGCCGAGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCG
CATTTTGCATAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAAT
ATTTAAAAATAAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTG
TCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACAT
TTTCCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAA
CGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTT
TTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTG
CGGCCATGGCCATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGA
GTATTGCAACTTACGGCCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTC
GAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCG
CCGGACTGTGAATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATA
TTTAGTGCTGTTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGT
TTATTTACATTTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAA
AACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCA
TAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTG
CATGGTATTCAATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTA
GCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAAC
CAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGT
CCAGGGGATTCGGGGGCCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAA
GCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCC
GCATGGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTT

TTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGT
CCAACAGTCGTAAGCGCGCCACGGTCCACAGAA

>D7
GCCGCCGAAGTCTGGCAAAACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCC
AAACAAACACAGCATCCCACACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGT
CAATTCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACG
CGGCTCTTATGCAACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGAC
CTGCCGATCGCAATAAATTAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTC
TGTGGGTGAGCCGAGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTT
TCGCATTTTGCATAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAA
AATATTTAAAAATAAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAA
CTGTCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGA
CATTTTCCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGT
AAACGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCT
TTTTTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGC
TTGCGGCCATGGCCATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACA
AGAGTATTGCAACTTACGGCCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAAC
CTCGAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCC
TCGCCGGACTGTGAATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTC
ATATTTAGTGCTGTTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCA
GGTTTATTTACATTTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATT
TAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAG
CCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACG
TTGCATGGTATTCAATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGC
CTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAA
AACCAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCC
GGTCCAGGGGATTCGGGGGCCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACA
AAAGCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAG
TCCGCATGGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTC
GTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTG
AGTCCAACAGTCGTAAGCGCGCCACGGTCCACAGAA

>E0
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA

GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCA

>E1
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGC

>E2
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA

CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATG

>E3
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGC

>E4
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAAT

>E5
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCA

>D2-DllKO
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA

TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTcgcgATTCCGTTTAAGGACGCcgcgTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAAcgcgTCGAATTCCCCGCTGGCTcgcgAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTG
AATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCC
TGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCG
CGCCACGGTCCACAGAA

>D2-[6]
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCat
ctttacactaAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC

GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTG
AATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCC
TGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCG
CGCCACGGTCCACAGAA

>D2-block4
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGCTTTCCCTCGGTGTCTGTTAA
AGTGGAGAGGATCCTGGTTCCTGTGGCCAGCATTCCGTAGGATGTATAATAACGGCGCACAGAGCCTTCCA
ACCTCCCACTTCGTTATCATCTGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT

GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT
AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTG
AATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCC
TGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCG
CGCCACGGTCCACAGAA

>D2-block5
CCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGC
TTTTGGCTTGAATAAATTAATCGACATGAGTCATGTTGCGGATCTTCAAGGTAACTCATACACATCATTCC
ATAATGACTTTGATGACCTCATCGCTTTTAGTCGCCCACTTGGAATAATATTGATAGCAGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGG
CCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAA
TGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCA
GGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
ACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATT

AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTGCCCAGACATTTTCCTTGGCCGAAGT
GTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATT
TGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTG
AATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCC
TGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCG
CGCCACGGTCCACAGAA

Paper II: Entangled and non-modular enhancer sequences producing independent spatial activities

Mariam Museridze, Stefano Ceolin, Bettina Mühling, Srishti Ramanathan, Olga Barmina, Pallavi Santhi Sekhar, and Nicolas Gompel

SCIENCE ADVANCES | Vol 10, Issue 47 | 20 Nov 2024
DOI: 10.1126/sciadv.adr9856

From: Mariam Museridze, Stefano Ceolin, Bettina Mühling, Srishti Ramanathan, Olga Barmina, Pallavi Santhi Sekhar, and Nicolas Gompel. 2024. “Entangled and Non-Modular Enhancer Sequences Producing Independent Spatial Activities.” Science Advances 10 (47): eadr9856.

SCIENCE ADVANCES | RESEARCH ARTICLE

GENETICS

Entangled and non-modular enhancer sequences producing independent spatial activities

Mariam Museridze[1,2]†, Stefano Ceolin[1]†, Bettina Mühling[1], Srishti Ramanathan[1], Olga Barmina[3], Pallavi Santhi Sekhar[1], Nicolas Gompel[1,2]*

The modularity of transcriptional enhancers is central to our understanding of morphological evolution, allowing specific changes to a gene expression pattern component, without affecting others. Enhancer modularity refers to physically separated stretches of regulatory sequence producing discrete spatiotemporal transcriptional activity. This concept stems from assays that test the sufficiency of a DNA segment to drive spatial reporter expression resembling that of the corresponding gene. Focusing on spatial patterns, it overlooks quantitative aspects of gene expression, underestimating the regulatory sequence actually required to reach full endogenous expression levels. Here, we show that five regulatory activities of the gene *yellow* in *Drosophila*, classically described as modular, result from extensively overlapping sequences, with broadly distributed regulatory information. Nevertheless, the independent regulatory activities of these entangled enhancers appear to be nucleated by specific segments that we called enhancer cores. Our work calls for a reappraisal of enhancer definition and properties, as well as of the consequences on regulatory evolution.

INTRODUCTION

The embryonic blueprint of gene expression prefiguring morphology results from the activity of separate cis-regulatory elements (*1*–*5*). It was hypothesized (*6*, *7*) and then demonstrated (*8*–*11*) [for review, see (*12*)] that mutations in enhancers are a primary driver for the evolution of discrete morphological traits, such as gains and losses of limbs in vertebrates or decorative elements in insects. Owing to their reduced pleiotropy, mutations affecting single enhancers facilitate changes in specific aspects of morphology with no deleterious effect on other traits (*12*, *13*). Hence, a modular representation of enhancers bears strongly on our understanding of regulatory evolution.

The original notion of enhancer modularity (*14*, *15*) was quickly generalized in transgenic animals and plants with reporter assays testing the capacity of arbitrarily chosen DNA segments to recapitulate elements of the gene's spatial expression (*16*). This notion is based on the sufficiency of DNA segments to drive specific expression patterns, leaving out completely how much transcript these sequences yield compared to the endogenous levels of transcript. Yet, the amount of gene transcript is key to normal development (*17*, *18*). Moreover, a minimal enhancer producing a correct spatiotemporal pattern may be insufficient for a robust expression under unfavorable environmental or genetic conditions (*19*–*21*). It has been repeatedly shown that mutations occurring outside minimal enhancers can contribute to phenotypic evolution (*21*–*23*). Last, several papers have recently challenged the lack of pleiotropy of enhancers (*24*–*26*).

The modular definition of enhancers devolved from functional assays has been reinforced by genomic approaches. Discrete regions of open chromatin or peaks of specific epigenetic marks for active enhancers have been used to map enhancers genome-wide (*27*–*32*). Yet, the extent of sequence that these peaks span is influenced by the choice of cutoff values or other methodological biases. For instance, the processing of datasets assessing genome-wide chromatin accessibility starts with the binning of raw data into putative regulatory elements (*33*), leading to a circular argument. As a result, of these different approaches, transgenic assays, or genomic methods, enhancers are deemed to span 100 to 1000 base pairs, although the sequence necessary for a full regulatory activity, pattern and levels, may be larger (*26*).

Additional considerations undermine the notion of enhancer modularity or their compactness. While short regulatory segments usually drive spatially restricted expression, they are often found to have broad activity, targeting multiple tissues (*2*, *24*, *25*, *34*–*36*). By contrast, single enhancers are sometimes insufficient to produce robust gene expression in a given tissue: The additive activity of redundant or shadow enhancers was shown to enable robust gene expression (*19*, *20*, *37*). Other sequences, referred to as facilitator elements, do not drive activity on their own but enhance the function of neighboring enhancers (*38*). They are part of so-called super-enhancers, which are themselves described as large clusters of enhancers sufficient to assign cell identity, a concept that was criticized for its lack of functional support (*39*). These different concepts assume some level of sequence modularity, but a systematic analysis of intervening regions between these regulatory elements has been lacking.

Does the picture of discrete and modular enhancers hold when one considers the quantitative dimension of transcription? Using a quantitative reporter assay, we addressed this question by precisely mapping regulatory information at the *yellow* (*y*) locus in *Drosophila*, a typical locus with several independent enhancers. Their activities, classically mapped with reporter assays, prefigure a gray background or a black spot on the wing, stripes, a longitudinal band, a sexually dimorphic block of pigmentation on the abdomen, or black sensory bristles over the body, respectively (*9*, *40*–*42*). A recent systematic dissection of *y* regulatory

[1]Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Planegg-Martinsried, Germany. [2]University of Bonn, Bonn Institute for Organismic Biology, Bonn, Germany. [3]Department of Evolution and Ecology, University of California, Davis, Davis, CA, USA.
*Corresponding author. Email: ngompel@uni-bonn.de
†These authors contributed equally to this work.

regions in *Drosophila melanogaster* (*34*), however, suggested more distributed activities than the textbook picture of separate and independent enhancers would predict. Along the same lines, focusing on the wing-spotted species *Drosophila biarmipes*, we showed that the *wing blade* and *spot* enhancers, respectively prefiguring the background gray pigmentation of the wing and a dark distal wing spot, were actually broader, extensively overlapping, and shared regulatory information (*43*). The sequence overlap of these two enhancers may, however, simply result from a recent cooption process, and the regulatory activities may resolve with enough evolutionary time.

To investigate the relationship between distinct regulatory activities, we compared the wing enhancers to an ancient regulatory sequence, the *body* enhancer. The *body* enhancer, found in various *Drosophila* species (*42*, *44*), is active in the head, thoracic, and abdominal epidermis during metamorphosis (*45*, *46*), defining a complex spatial pattern. We mapped regulatory information of the *body* enhancer and compared it to our previous map of the wing enhancers (*43*). The *body* enhancer spans the entire sequence of the two wing enhancers, refuting their modularity. The sequences of these activities are entangled with, however, different distributions of regulatory information.

RESULTS

The *y body* enhancer encompasses the entire *y* 5′ region

y is expressed in body pupal epidermis (*42*, *47*) under the control of regulatory sequences mapping 5′ of *y* transcription start site (TSS) (*41*, *45*, *46*). The dissection of these sequences from *D. melanogaster* with transgenic reporter assays identified a 1-kb segment, the *body* enhancer, sufficient to recapitulate *y* spatial expression (*42*). Whether this segment is sufficient to drive endogenous transcription levels is, however, unknown. We therefore first mapped the entire regulatory sequence necessary and sufficient to drive *y* transcription in the fly body, relying on a quantitative reporter assay. For the sake of comparison with our previous mapping, we used *D. biarmipes* sequences, and, for simplicity, we focused on abdominal expression. Starting from a 5.4-kb segment upstream of *y* TSS, we tested the regulatory activity of segments with deletions from the 5′ end (*D* series), as well as progressive sequence randomization from the 3′ end in male pupae [*E* series; Fig. 1, A and C, and (*43*)]. To map the boundaries of regulatory activities in the abdomen, we calculated how much reporter expression was lost or gained for each construct compared to the activity of the largest construct of the corresponding series (Fig. 1B; see Materials and Methods). For instance, removing a distal segment (line *D1*), about 4 kb upstream of the original *body* enhancer (*42*), already had a significant effect on the abdominal expression (Fig. 1B). Generally, we found that all constructs differed in activity from the largest construct of each series (table S1). We concluded that the regulatory information determining abdominal expression spans up to 5.4 kb and overlaps largely with the previously mapped regulatory activities of the wing (*43*), challenging the notion of short and modular enhancers. Average phenotypes (Fig. 1C) revealed changes not only in the levels of regulatory activity among constructs but also in spatial distribution. We therefore sought to understand how the distributed regulatory information upstream of *y* TSS was organized to control different spatial pattern elements.

PCA highlights the composite nature of *body* activities

To this end, we relied on a principal components analysis (PCA) of phenotypic variation across all constructs (Fig. 2, A and B, and fig. S1), as it gave us access to the fine relationship between regulatory content and enhancer activity. The first component [principal component 1 (PC1), 72% of the total variation; Fig. 2, A and B], mostly captured broad uniform expression changes across the abdomen. The second component (PC2, 18% of the total variation) captured changes in two distinct spatial activities simultaneously, in the upper four segments and the last two segments, respectively corresponding to broad uniform expression (*34*) and the sexually dimorphic male expression (Fig. 2, A and B). The third component showed an identified artifact caused by gaps between the last two abdominal segments, due to ventral bending (fig. S1). We did not consider PC3 for further analysis. The fourth component (PC4, 1% of the total variation) shows changes in the basal part of each segment, identifying the banded pattern of the fly abdomen (Fig. 2A). In summary, PC1, PC2, and PC4 appear to jointly capture variation in the three main pattern components of the global *body* activity, although these PCs are not necessarily independent.

stripes, *broad*, and *dimorphic*, three independent enhancers defining *yellow* abdominal expression

To test the possible regulatory independence of these activities, we sought to identify sequence segments that affect single spatial pattern components. To this end, we first calculated differences between average phenotypes of consecutive lines (fig. S2). We observed that the average phenotype differences between certain lines corresponded precisely to pure individual activities (Fig. 2C): *D3-D4* reflects the stripes component, *E1-E2* reflects the broad uniform component, and *D2-E0* reflects the dimorphic component, suggesting that the corresponding sequence segments may affect the respective activities independently. To directly test these independent regulatory contributions, we examined reporter constructs with randomized sequence in place of these segments (Fig. 2D): Δ *stripes* (segment *D3-D4* randomized in the context of *D0*), Δ *broad* (segment *E1-E2* randomized in the context of *D0*), and Δ *dimorphic* (segment *D2-E0* randomized in the context of *D2* = construct *E0* of Fig. 1). Compared to the activity of the largest construct *D0*, which contains all three pattern components—stripes, broad expression, and dimorphic expression, we observed that only a single component was missing in each randomized line (Fig. 2E): Δ *stripes* had no visible stripes, Δ *broad* missed broad expression in the upper segments, and Δ *dimorphic* was devoid of strong expression in the posterior segments. These results showed that the three *body* activities are, to some extent, functionally independent of each other. Further, they suggested that each of these segments could be an enhancer core, meaning a stretch of sequence necessary to seed the spatial regulatory activity but insufficient to account for the endogenous levels of expression (*43*). To evaluate the extent of independence of the three activities—*broad*, *dimorphic*, and *stripes*—and to lastly examine the distribution of regulatory information upstream of the *y* TSS, we went on to decompose the expression patterns as a sum of these three activities in the PC space. The decomposition is achieved through a simple operation, a change of basis, whereby the PC space is re-projected in a new coordinate system made of three vectors corresponding to the pure activities. The *stripes* activity is defined by the vector (*D0* to Δ

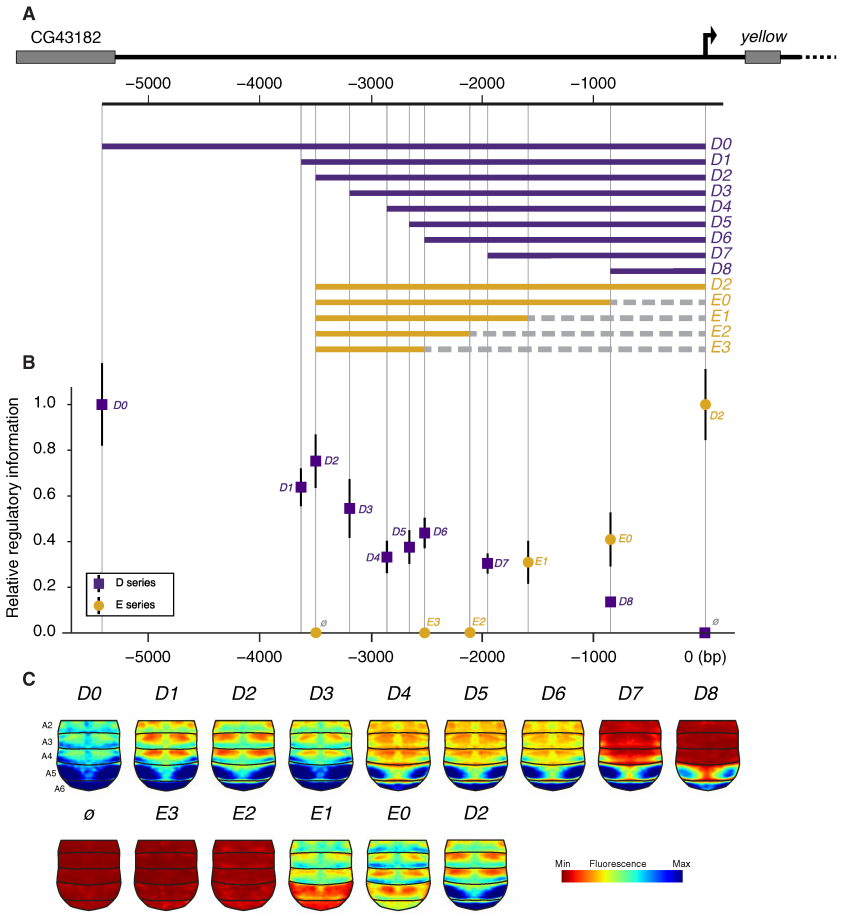

Fig. 1. Mapping regulatory activities upstream of the *yellow* gene. (**A**) Reporter construct series used to evaluate the regulatory content of regions located 5′ of *y*. The first series, D (purple), is progressively trimmed from the 5′ end, while the proximal end of constructs of the second series, E (ocher), is increasingly randomized, without changing the distance to the TSS. Note that construct *D2* appears in the *E* series, too, as the reference construct from which all E constructs were derived. (**B**) Relative regulatory information (fluorescence levels) for each reporter line compared to the reference line of the series. bp, base pairs. (**C**) Average reporter expression (average phenotype) in the abdominal epidermis for each line. ø denotes a line with an empty reporter vector. In all figures, only male abdomens were examined.

stripes) in the original PC space, the *broad* activity by the vector (*D0* to Δ *broad*), and the *dimorphic* activity by the vector (*D2* to Δ *dimorphic*). The resulting mixed components of the re-projected space show changes only in stripes, broad, or dimorphic pattern elements, respectively (Fig. 3).

Entangled wing and body enhancers upstream of *yellow*

In this re-projected space, we could then measure the independent contribution of each fragment in the dissection series to each pattern component and compare these distributions of regulatory information to those of the wing enhancers that we calculated in (*43*) (Fig. 3). The results show that, in spite of their extensive overlap, the distribution of regulatory information along the sequence is very different among activities. Together, our results outline an alternative architecture of *y* regulatory regions, whereby enhancer cores necessary to initiate a regulatory activity are surrounded by regulatory sequences contributing to the full endogenous levels of transcription. It is at odds with the current thread-like image of modular regulatory elements at *y* or, generally, at developmental loci. This could be a peculiarity of the *y* locus. Yet, the methodological biases of ascertainment caused by classical reporter assays and genomic data processing (*33*) speak for an entangled architecture of regulatory regions in general. This is only comforted by circumstantial evidence that regulatory regions have

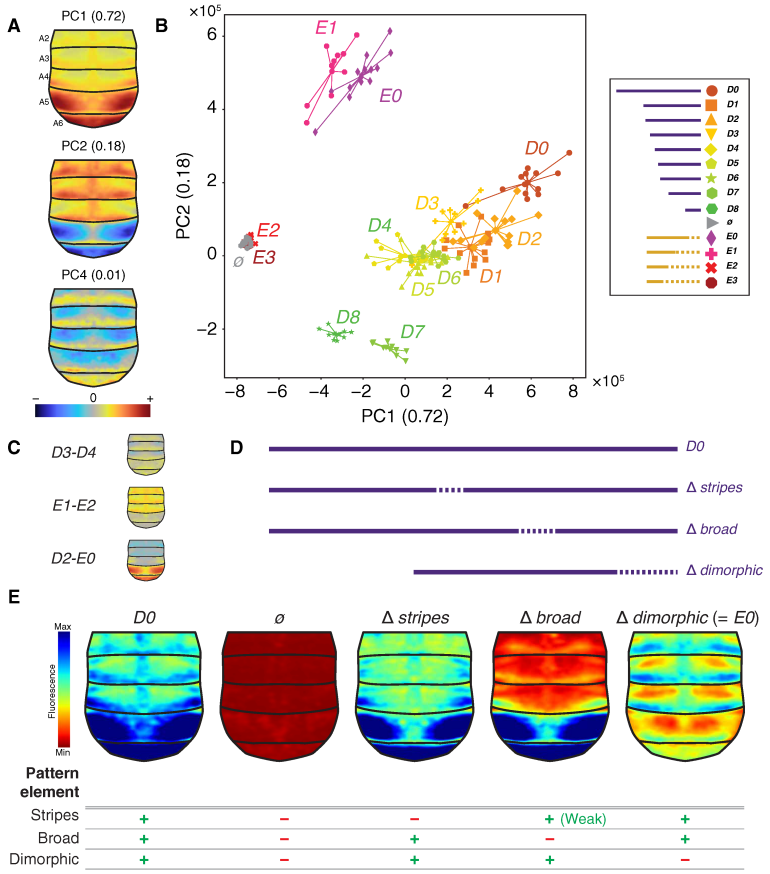

Fig. 2. The *body* enhancer pattern results from three distinct activities. (**A**) Phenotypic directions captured by three principal components (PCs) in the space of variation for all reporter lines shown in Fig. 1. These directions correspond to broad uniform expression (PC1), a combination of broad uniform expression and sexually dimorphic male expression (PC2), and stripes (PC4). PC3 identifies an artifact caused by stretching between consecutive segments and is shown in fig. S1. (**B**) Principal components analysis (PCA) of phenotypic space showing the first two PCs. Each dot represents a single abdomen. All samples from the same construct are connected by lines to their average. (**C**) Subtraction of average phenotypes of selected lines identifies specific pattern elements: *D3-D4*, stripes; *E1-E2*, broad uniform expression; and *D2-E0*, sexually dimorphic element. (**D**) Constructs with randomized sequence segments according to (C). (**E**) Resulting average phenotypes of lines with randomized segments with an annotation of pattern elements.

been underestimated (*21*–*23*, *26*). This emerging possibility raises a fundamental question.

If enhancers are not modular stretches of sequence, then how do they independently produce regulatory activities in distinct subsets of cells? At the *y* locus, each enhancer responds to a subset of transcription factors (TFs) governing its activity, but TFs present in the wing may also be present in the abdomen and could activate wing enhancers in abdominal cells. We compared the TFs expressed in pupal wings (*48*, *49*) to those expressed in abdominal epidermis at the same stage (using a previously unpublished microarray dataset, see Materials and Methods). Of 302 wing TFs and 240 abdominal TFs, we found that 78 are expressed in both tissues during pupal life (table S2), potentially exposing their respective enhancers to bleed-through activities.

DISCUSSION

Our reappraisal of *yellow* regulatory architecture with a quantitative method challenges the textbook picture of enhancers as small and discrete boxes surrounding the coding sequence of a gene (*50*). The difference arises, we reckon, from taking expression levels into consideration. We would therefore not be surprised if other loci with complex expression also had large and overlapping enhancers. The notion of distributed regulatory information has previously been proposed for the *y* locus (*34*) and may generally relate to the concepts of redundant enhancers (*51*) as well as super-enhancers (*38*, *39*). It is conceivable that intervening regions between shadow enhancers contribute to increase regulatory activity but do not have the ability to drive expression alone, reminiscent of facilitator elements (*38*).

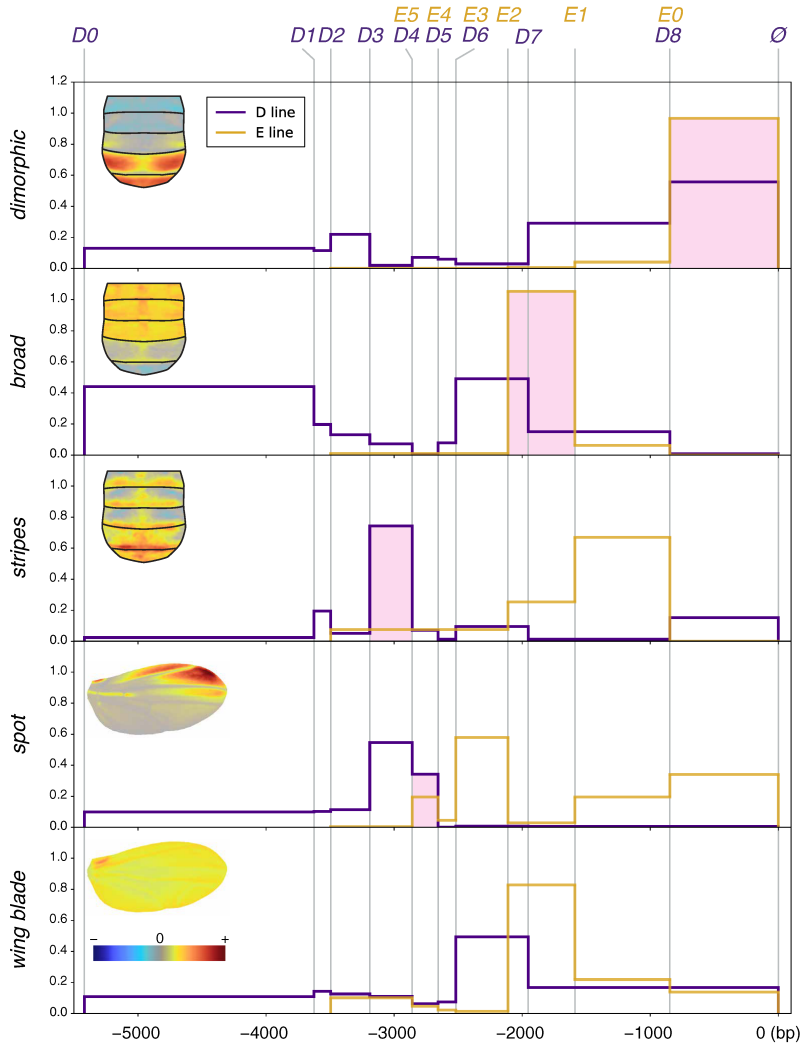

Fig. 3. Regulatory architecture upstream of *yellow*. The abdomen and wing images with a colormap correspond to the mixed components, representing the variation of different regulatory activities in the PC space, after a change of basis. While PCs of the original PCA captured multiple activities, the mixed components resulting from a change of basis capture variation in single activities (i.e., phenotypic directions). The contribution of each sequence segment along *y* 5′ region to each mixed PC is a map of the regulatory information for each enhancer. While all activities span most of the 5.4-kb region, we note that enhancer cores (pink shading) identified in Fig. 2 and (*43*) are not overlapping.

We mapped regulatory information at the *y* locus in a transgenic and heterologous context. This could, in principle, bias our assessment. We have shown, however, that the same *spot* enhancer reporter construct drove similar expression in *D. melanogaster* and in *D. biarmipes* transgenic flies (*48*). The only significant difference was the shape of the spot pattern, extending proximally toward the wing hinge in *D. melanogaster*, but confined distally in *D. biarmipes*. This difference, we found, was due to a change in the spatial distribution of the activator Distal-less (*48*). Apart from this, we have also established that this region and other *Drosophila* enhancers contain the information determining their own accessibility (*43*, *52*). Hence, studying enhancers in a different genomic context than their endogenous locus should not bias how their regulatory content is read by TFs.

Entangled regulatory information at the *y* locus resolves into functionally independent enhancers. We submit that the regulatory independence may reside in the function of enhancer cores and hypothesize that each core subordinates all flanking regulatory information through an unknown mechanism to selectively produce the full activity of a given enhancer. This could involve selective accessibility or selective use of TF binding sites, or both, as we have shown in the case of the *spot* enhancer (*43*). With the advent of methodologies describing the dynamics of the three-dimensional (3D) genome, the representation of gene regulation is changing from selective and transient looping of a remote enhancer onto the core promoter region to more stable multi-enhancer hubs gathered around the TSS of a gene (*53*, *54*). It is conceivable that a complex architecture of entangled enhancers forms a relatively stable hub around a core promoter, where core enhancer segments control the selective involvement of each activity in the respective cell types.

Last, the notion of enhancer modularity is the cornerstone of Evolutionary Developmental Biology, predicting that regulatory changes in developmental genes drive morphological diversification because these changes circumvent pleiotropy and foster evolvability (*12*). If enhancers share extensive portions of sequence, then how are their independent changes possible? One possibility is that binding sites for TFs involved in distinct activities are intermingled rather than shared. This would still afford changes in selected activities with minimal effects on other enhancers. On the other hand, if the 3D folding of a regulatory region with entangled enhancers is relevant for its function, then the evolvability of regulatory sequences may be more limited than previously assumed.

MATERIALS AND 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 (*48*). We used ϕC31-mediated transgenesis (*55*) and integrated all constructs at the genomic attP site VK00016 on chromosome 2 (*56*). The enhancer sequence of all transgenic stocks was genotyped before imaging. Of note, all reporter constructs contain regulatory regions from *D. biarmipes* and are tested in *D. melanogaster* transgenic flies.

Molecular biology

Lines from the D and E series and the randomized sequence for the Δ *stripes* and Δ *broad* were taken from (*43*). Fragments of *y* 5′ sequences for lines *D8*, Δ *stripes*, and Δ *broad* were amplified with Phusion polymerase (New England Biolabs) and cloned into our transformation vector pRedSA (*43*) digested with Eco RI and Bam HI using In-Fusion HD Cloning Kits (Takara, catalog no. 121416). Randomized sequences were amplified from the fragment used in (*43*). Primers are listed in table S3. The sequences are provided in table S4.

Imaging

Sample preparation

All transgenic abdomens imaged in this study were from male flies heterozygous for the reporter construct, obtained from a cross between homozygous reporter construct flies and a marker line [;*en*-Gal4, UAS-GFP/CyO; *pnr*-Gal4/TM6B, (*57*), FlyBaseID FBrf0098595]. White pupae were left to age 90 hours at 25°C. Pupal case was removed, and pupae were mounted in halocarbon oil (Sigma-Aldrich, CAS no. 9002-83-9) on a microscope slide with cover slips and appropriate spacers and immediately imaged.

Microscopy

All abdomen images were acquired as 12-bit images on a Leica SP5-2 confocal microscope using an HC PL APO CS 10×/0.40 IMM lens with a HyD SP Hybrid-Detector. Each image comprises two fluorescent channels (DsRed to report *y* regulatory activities and *en*-Gal4, *pnr*-Gal4, UAS-GFP as positional landmark for image registration). Image stacks were acquired in 4.99-µm steps.

Image quantification and analysis

We used a comprehensive pipeline for the registration, segmentation, and analysis of multichannel 3D fluorescence microscopy images. This pipeline, now supporting images with three color channels, facilitates the analysis of gene expression patterns in the epithelial cells of the *Drosophila* abdomen by producing a registered 2D map of expression on the abdomen surface starting from a 3D image stack (fig. S3A). The pipeline begins with an automated preprocessing step (fig. S3B), where segmentation of the image stack is performed using an automatically determined threshold. This threshold is set to ensure that the segmented volume covers a specified range of the entire image volume. After segmentation, the object is refined using morphological transformations and mesh fitting to fill any holes. Following preprocessing, the registration step involves rendering the segmented volume in an interactive interface. Corresponding points on the source and reference objects are selected, and the image stack is rigidly rotated and rescaled to minimize the distances between these points (fig. S3C). In the projection phase, the surface of the registered objects is projected into 2D images using a modified sinusoidal projection. The object is sliced, and, for each slice, the profile of the bright object is interpolated with a spline curve. Image brightness is then read out along the curve by taking the local maxima along the local normal direction. The resulting 1D brightness profiles from each slice form the rows of the 2D projected image, which are aligned at a predefined meridian plane (fig. S3D). The final step involves labeling and elastic warping. A graphical interface built with the PYSimpleGUI library enables manual labeling and registration of the 2D images. User-selected points are used to elastically warp the images onto a reference model using thin-plate spline registration (fig. S3E). This methodology ensures precise registration and analysis of multichannel 3D images, thereby facilitating the detailed study of gene expression patterns in the *Drosophila* abdomen. The Python-based code, together with an example of Jupyter Notebooks, installation instructions, and a minimal test dataset, is available at https://github.com/UniBonn-GompelLab/3D-DrosophilaRegistration.

Relative regulatory information and statistical analysis

We showed how much regulatory information (fluorescence levels) is contained in each line and compared it to the construct containing the largest regulatory fragment. First, to measure the amount of fluorescence, we calculate a squared Euclidean distance from each sample to the average phenotype of the line with an empty vector (ø). Practically, we subtracted from each projected abdomen image the average image of the empty line (ø) and then squared and sum all the values. Next, we calculated an average value of these distances for each line. Last, to show the relative amount of regulatory information, we divided all the line averages by the average of the biggest construct in the series (*D0* line for the *D* series and *D2* line for the E

series). The statistical significance of differences among lines was assessed with Kolmogorov-Smirnov tests (table S1).

Density of regulatory information

We calculated the PCA from the matrix of dimensions (n_individuals × n_pixels on the abdomen), regrouping intensities of all pixels for every individual. Mixed components representing the *stripes*, *broad*, and *dimorphic* activities were obtained by a change of basis of the original PC space. We calculated vectors defining a new coordinate system as differences between the control lines and randomized lines without enchancer cores: *D0* and Δ *stripes*, *D0* and Δ *broad*, and *D2* and Δ *dimorphic*. The amount of regulatory information brought by a segment of DNA for a given activity was calculated as the absolute value of the difference between the phenotypic distances of two consecutive fragments, relative to the phenotypic distance from the empty line (ø) to the largest construct in the series, using independent measurements for the *broad*, *stripes*, *dimorphic*, *wing blade*, and *spot* activities. To represent regulatory information, we used the absolute value of the change in the measure of activity, resulting in a similar depiction of repression and activation.

Abdominal transcriptome

RNA isolation

RNA isolation was performed on wild type flies using standard TRIzol (Invitrogen) protocol from abdominal dorsal epithelium of pupae 72 hours after puparium formation. The resulting RNA was further purified using "RNeasy Mini Protocol" for RNA cleanup (QIAGEN). RNA concentration was measured by spectrophotometer, and RNA quantity, quality, and size distribution were checked on a Bioanalyzer (Agilent Technologies).

Microarray experiment

In microarray experiments, each tissue was analyzed in three replicates consisting of 25 individuals each. RNA amplification was done according to Affymetrix "One cycle cDNA synthesis" and "Synthesis of biotin-labeled cRNA" protocols. Microarray hybridizations were conducted by Affymetrix facility.

Supplementary Materials

The PDF file includes:

Figs. S1 to S3
Table S4
Legends for tables S1 to S3
References

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S3

REFERENCES AND NOTES

1. J. L. Gomez-Skarmeta, I. Rodriguez, C. Martinez, J. Culi, D. Ferres-Marco, D. Beamonte, J. Modolell, Cis-regulation of achaete and scute: Shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev.* **9**, 1869–1882 (1995).
2. E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yanez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, A. Stark, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
3. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
4. M. L. Howard, E. H. Davidson, *cis*-Regulatory control circuits in development. *Dev. Biol.* **271**, 109–118 (2004).
5. C. H. Yuh, E. H. Davidson, Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069–1082 (1996).
6. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
7. S. B. Carroll, Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).
8. X. Wang, H. M. Chamberlin, Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48* ovo gene. *Genes Dev.* **16**, 2345–2349 (2002).
9. 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).
10. A. P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, D. L. Stern, Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
11. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jonsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
12. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
13. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U.S.A.* **104**(Suppl. 1), 8605–8612 (2007).
14. J. Banerji, S. Rusconi, W. Schaffner, Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
15. W. Schaffner, Enhancers, enhancers – From their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).
16. K. Harding, T. Hoey, R. Warrior, M. Levine, Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *EMBO J.* **8**, 1205–1212 (1989).
17. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
18. A. Pavlopoulos, M. Akam, Hox gene *Ultrabithorax* regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2855–2860 (2011).
19. M. W. Perry, A. N. Boettiger, J. P. Bothma, M. Levine, Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
20. N. Frankel, G. K. Davis, D. Vargas, S. Wang, F. Payre, D. L. Stern, Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).
21. M. Z. Ludwig, Manu, R. Kittler, K. P. White, M. Kreitman, Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. *PLOS Genet.* **7**, e1002364 (2011).
22. J. Crocker, D. L. Stern, Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. *Development* **144**, 3095–3101 (2017).
23. R. D. Bickel, A. Kopp, S. V. Nuzhdin, Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. *PLOS Genet.* **7**, e1001275 (2011).
24. E. Preger-Ben Noon, G. Sabaris, D. M. Ortiz, J. Sager, A. Liebowitz, D. L. Stern, N. Frankel, Comprehensive analysis of a *cis*-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* **22**, 3021–3031 (2018).
25. G. Sabaris, I. Laiker, E. P.-B. Noon, N. Frankel, Actors with multiple roles: Pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet.* **35**, 423–433 (2019).
26. G. Sabaris, D. M. Ortiz, I. Laiker, I. Mayansky, S. Naik, G. Cavalli, D. L. Stern, E. Preger-Ben Noon, N. Frankel, The density of regulatory information is a major determinant of evolutionary constraint on noncoding DNA in *Drosophila*. *Mol. Biol. Evol.* **41**, msae004 (2024).
27. D. J. McKay, J. D. Lieb, A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev. Cell* **27**, 306–318 (2013).
28. J. C. Pearson, D. J. McKay, J. D. Lieb, S. T. Crews, Chromatin profiling of *Drosophila* CNS subpopulations identifies active transcriptional enhancers. *Development* **143**, 3723–3732 (2016).
29. C. M. Uyehara, S. L. Nystrom, M. J. Niederhuber, M. Leatham-Jensen, Y. Ma, L. A. Buttitta, D. J. McKay, Hormone-dependent control of developmental timing through regulation of chromatin accessibility. *Genes Dev.* **31**, 862–875 (2017).
30. M. Bozek, R. Cortini, A. E. Storti, U. Unnerstall, U. Gaul, N. Gompel, ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Res.* **29**, 771–783 (2019).
31. M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
32. S. Fu, Q. Wang, J. E. Moore, M. J. Purcaro, H. E. Pratt, K. Fan, C. Gu, C. Jiang, R. Zhu, A. Kundaje, A. Lu, Z. Weng, Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers. *Nucleic Acids Res.* **46**, 11184–11201 (2018).
33. W. Saelens, O. Pushkarev, B. Deplancke, ChromatinHD connects single-cell DNA accessibility and conformation to gene expression through scale-adaptive machine learning. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.07.21.549899.

34. G. Kalay, J. Lachowiec, U. Rosas, M. R. Dome, P. J. Wittkopp, Redundant and cryptic enhancer activities of the *Drosophila yellow* gene. *Genetics* **212**, 343–360 (2019).
35. I. Laiker, N. Frankel, Pleiotropic enhancers are ubiquitous regulatory elements in the human genome. *Genome Biol. Evol.* **14**, evac071 (2022).
36. O. Nagy, I. Nuez, R. Savisaar, A. E. Peluffo, A. Yassin, M. Lang, D. L. Stern, D. R. Matute, J. R. David, V. Courtier-Orgogozo, Correlated evolution of two copulatory organs via a single *cis*-regulatory nucleotide change. *Curr. Biol.* **28**, 3450–3457.e13 (2018).
37. J. W. Hong, D. A. Hendrix, M. S. Levine, Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
38. J. W. Blayney, H. Francis, A. Rampasekova, B. Camellato, L. Mitchell, R. Stolper, L. Cornell, C. Babbs, J. D. Boeke, D. R. Higgs, M. Kassouf, Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* **186**, 5826–5839.e18 (2023).
39. S. Pott, J. D. Lieb, What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
40. S. Jeong, A. Rokas, S. B. Carroll, Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* **125**, 1387–1399 (2006).
41. P. K. Geyer, V. G. Corces, Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. *Genes Dev.* **1**, 996–1004 (1987).
42. P. J. Wittkopp, K. Vaccaro, S. B. Carroll, Evolution of *yellow* gene regulation and pigmentation in Drosophila. *Curr. Biol.* **12**, 1547–1556 (2002).
43. Y. Xin, Y. Le Poul, L. Ling, M. Museridze, B. Muhling, R. Jaenichen, E. Osipova, N. Gompel, Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20636–20644 (2020).
44. G. Kalay, P. J. Wittkopp, Nomadic enhancers: Tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among *Drosophila* species. *PLOS Genet.* **6**, e1001222 (2010).
45. H. Biessmann, J. M. Mason, Progressive loss of DNA sequences from terminal chromosome deficiencies in Drosophila melanogaster. *EMBO J.* **7**, 1081–1086 (1988).
46. M. Martin, Y. B. Meng, W. Chia, Regulatory elements involved in the tissue-specific expression of the *yellow* gene of *Drosophila*. *Mol. Gen. Genet.* **218**, 118–126 (1989).
47. H. Hinaux, K. Bachem, M. Battistara, M. Rossi, Y. Xin, R. Jaenichen, Y. Le Poul, L. Arnoult, J. M. Kobler, I. C. Grunwald Kadow, L. Rodermund, B. Prud'homme, N. Gompel, Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele. *Dev. Biol.* **438**, 111–123 (2018).
48. 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).
49. L. F. Sobala, P. N. Adler, The gene expression program for the formation of wing cuticle in Drosophila. *PLOS Genet.* **12**, e1006100 (2016).
50. J. D. Watson, T. A. Baker, P. B. Bell, A. Gann, M. Levine, S. C. Harrison, *Molecular Biology of the Gene* (Pearson, ed. 7, 2013).
51. E. Z. Kvon, R. Waymack, M. Gad, Z. Wunderlich, Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).
52. 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).
53. B. Lim, M. S. Levine, Enhancer-promoter communication: Hubs or loops? *Curr. Opin. Genet. Dev.* **67**, 5–9 (2021).
54. A. Panigrahi, B. W. O'Malley, Mechanisms of enhancer action: The known and the unknown. *Genome Biol.* **22**, 108 (2021).
55. 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).
56. K. J. 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).
57. M. Calleja, H. Herranz, C. Estella, J. Casal, P. Lawrence, P. Simpson, G. Morata, Generation of medial and lateral dorsal body domains by the *pannier* gene of *Drosophila*. *Development* **127**, 3971–3980 (2000).
58. G. Struhl, D. A. Barbash, P. A. Lawrence, Hedgehog organises the pattern and polarity of epidermal cells in the *Drosophila* abdomen. *Development* **124**, 2143–2154 (1997).

Acknowledgments: We are grateful to B. Prud'homme for insightful comments on the manuscript. **Funding:** This work was supported by the Graduate School of Quantitative Biosciences Munich (QBM, to M.M.) and the Deutsche Forschungsgemeinschaft (GO2495/16-1, to N.G.). **Author contributions:** Conceptualization: M.M., N.G., and S.C. Methodology: M.M., N.G., and S.C. Investigation: M.M., S.C., B.M., S.R., P.S.S., and O.B. Visualization: M.M., N.G., and S.C. Funding acquisition: N.G. Project administration: N.G. Supervision: N.G. Resources: O.B. and B.M. Software: M.M. and S.C. Validation: S.R., M.M., N.G., and S.C. Writing—original draft: N.G. and M.M. Writing—review and editing: M.M., N.G., and S.C. **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.

Submitted 24 July 2024
Accepted 17 October 2024
Published 20 November 2024
10.1126/sciadv.adr9856

Science Advances
AAAS

Supplementary Materials for

Entangled and non-modular enhancer sequences producing independent spatial activities

Mariam Museridze *et al.*

Corresponding author: Nicolas Gompel, ngompel@uni-bonn.de

Sci. Adv. **10**, eadr9856 (2024)
DOI: 10.1126/sciadv.adr9856

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

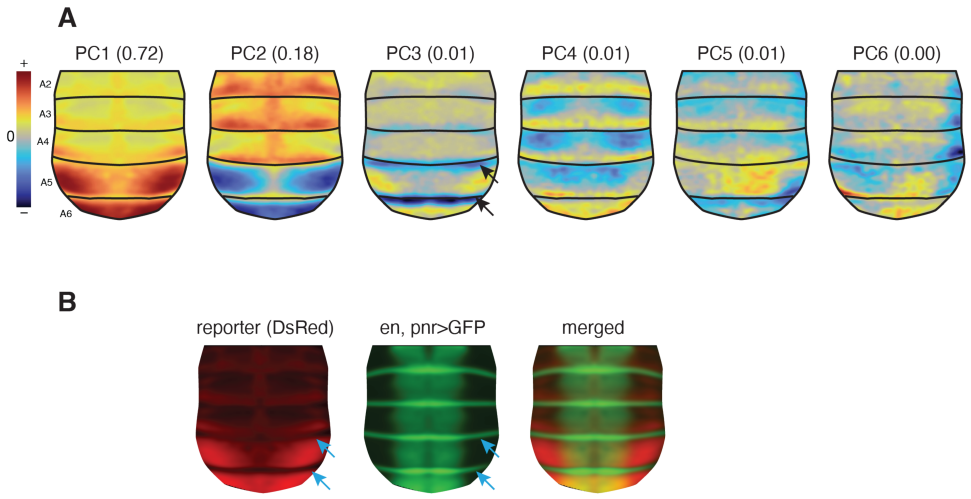

Fig. S1. Directions in the principal component analysis of reporter lines. (**A**) Phenotypic directions of the first 6 principal components. PC1, PC2 and PC4 appear to capture biological information directly related to the regulatory dissection (see main text), PC3 captures mostly the gap between posterior segments (arrows) caused by their ventral bending and stretching, and PC5 and PC6 capture experimental noise (note the asymmetrical patterns). (**B**) The intersegmental gaps between segments A4-A5 and A5-A6 (blue arrows pointing to identical positions as in (A)) are visible on an average intensity projection of all reporter line images used in the PCA. For each reporter line image (DsRed), we also acquired a second image of fluorescent landmarks used for image registration and delineated by the combined expression of *engrailed*-Gal4 (*en*, segmental stripes) and *pannier*-Gal4 (*pnr*, longitudinal band), driving a UAS-GFP transgene. The merged image shows that the gaps captured by PC3 correspond to the posterior compartment of each segment defined by the expression of *engrailed*, an unpigmented domain of soft cuticle (*58*).

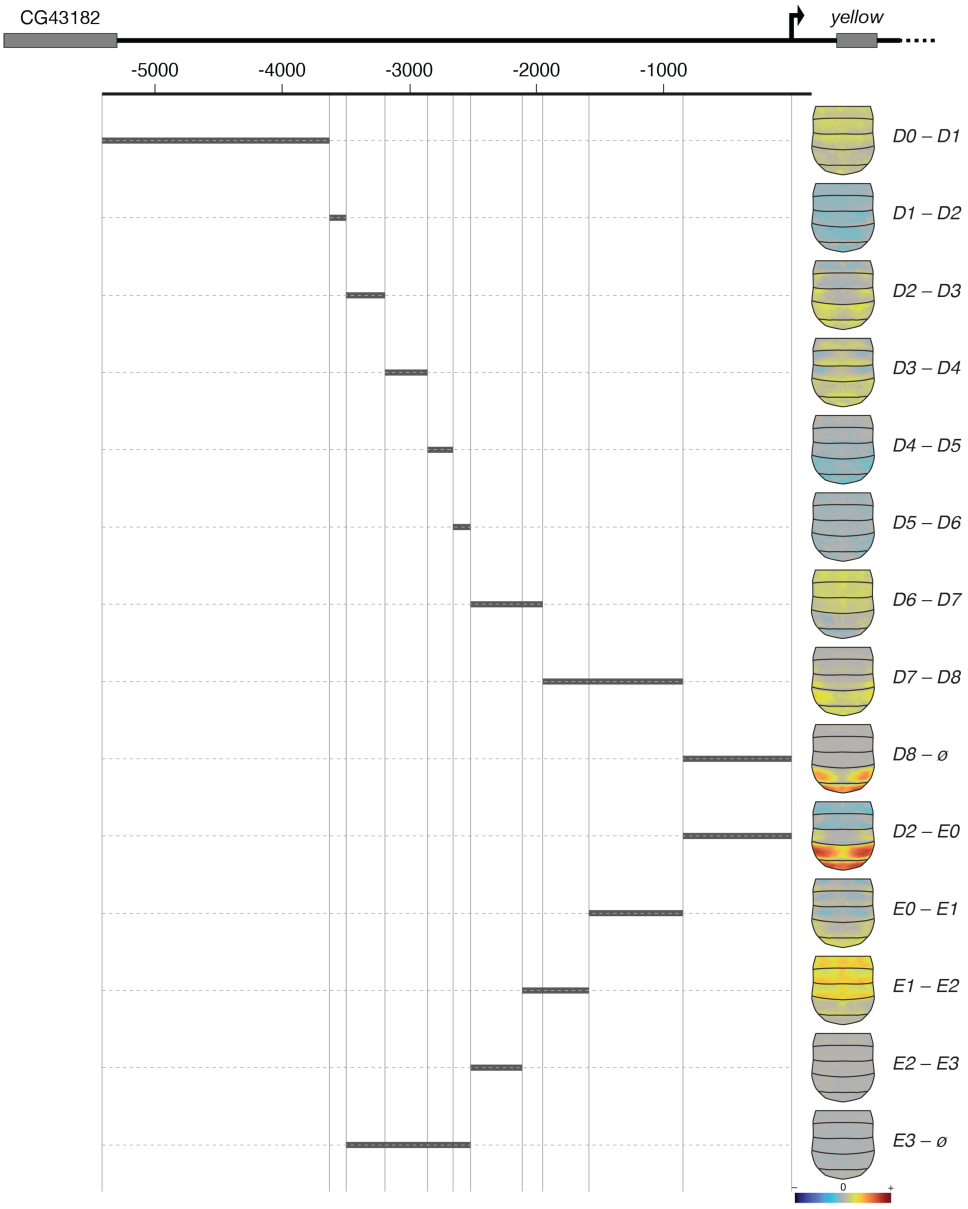

Fig. S2. Subtraction of average phenotypes between consecutive lines highlights potential enhancer cores. Each horizontal gray segment represents the difference between two consecutive constructs. For instance, the difference *D0 – D1* result in a distal-most segment. The

abdomens on the right represent the calculated phenotypic difference between the average phenotypes of the corresponding lines, where differential fluorescent levels are indicated by a color map.

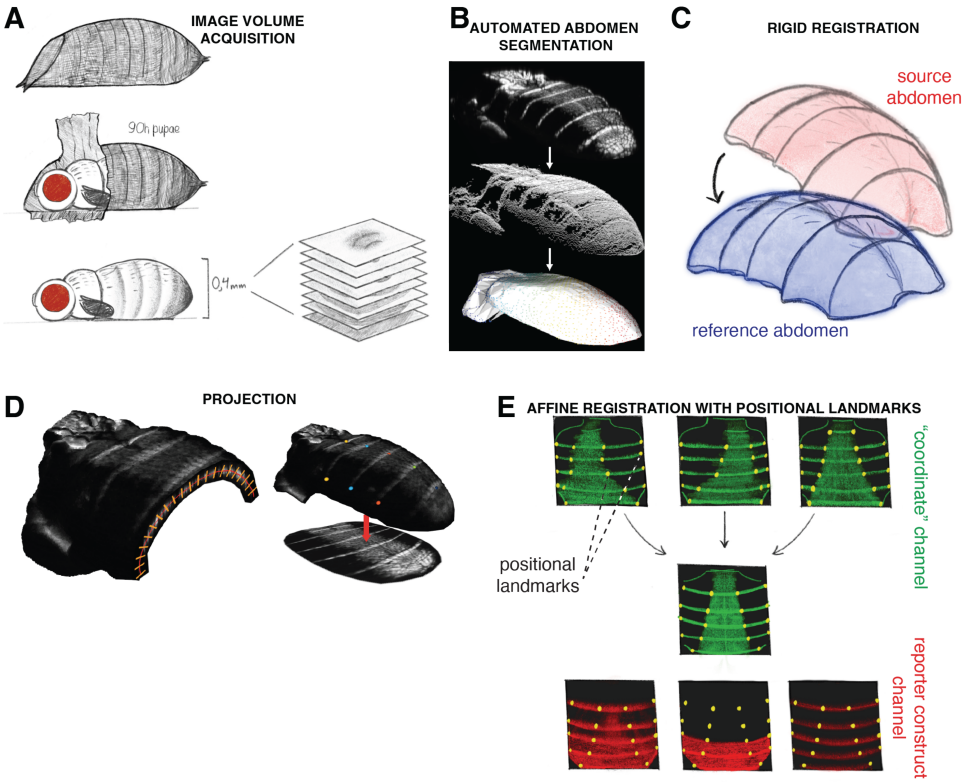

Fig. S3. Pipeline for the registration, segmentation, and analysis of multi-channel 3D fluorescence microscopy images. (**A**) Acquisition of a 3D image volume of a *Drosophila* pupal abdomen (**B**) Segmentation of the abdomen from the 3D volume. (**C**) Rotation and rescaling of the source and the reference objects (rigid registration). (**D**) 2D projection of the 3D object using a modified sinusoidal projection. (**E**) Registration of the 2D images with user-placed positional landmarks (yellow dots).

Table S1. Statistics for among-line expression comparisons. Comparison of expression changes (fluorescence levels) between the reference line for each of series and the trimmed/randomized lines, and comparison between adjacent lines. P-values were obtained using the Kolmogorov-Smirnov test. The table is provided as a separate file.

Table S2. Wing vs. abdominal epidermis TF comparison. The list of TFs expressed in pupal wings is the intersection of two published datasets: the micro-array dataset from the wings collected at 80 h after puparium formation (*48*) and RNA sequencing results from 72 h pupae (*49*). This list was compared to the TFs expressed in 72 h after puparium formation abdominal epidermis (previously unpublished micro-array dataset, see methods). TFs shared between the two tissues are highlighted (overlap wing/abdomen). The table is provided as a separate file.

Table S3. Primer sequences. Primers for cloning of the *D8*, *Δ stripes* and *Δ broad* lines. Primers include additional sequence, creating overhangs for In-fusion cloning. The table is provided as a separate file.

Table S4. Construct sequences.

>D8

AAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTTAGTGCTGTTCGCAAAGACCTGTCCCAGA
TACTCTGTTTATAGGTATAATTATTAAGTGCATATCAGGTTTATTTACATTTATATCGTATTAT
ATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCCA
CAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAAAAAGCCATAAAGCCTAA
AATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCAT
GGTATTCAATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAATTG
CCTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGT
GTGCCAAAAACCAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGG
CGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGGCCCGTGATACTCGCACTTAATAAACATGCG
TGAAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCG
ATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCACCCGAGCCACCGAAAG
CAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTCGCTTCCATTGGCCTGCCTTCGT
CTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGCGCG
CCACGGTCCACAGAA

>*Δ stripes*

AAGAGCCCAAGGTCGGTCGTTTAGCTTGGCCAAAACCTACCCATCCAGCTGGCACTTCCACCAA
CGGCACCAAGACACGAGCGGAAAATAAAAAGCCACACCACCCCACTTAGAACTCCGTTTAGCAG
CAGTTGTTCAAACAGAAATTGGCTGGCTTCGGCCGGTTCAGCcTCAGTTGATAATTATTATAAT
ATCTATGTTCTTGCCTATCGCTGGGCCCTAATTGGCCCAGACAAAGGCACCGTTTTTATGCATA
ACTGGAGGCTTACAATTTGGCcTTCGACACGCGGTCCTTCACATTGCCAAAAAAGAACGAGAAC
TCgGCAAGCCAATTACACTCGAAGAAGCAGCGGGGATCGTTTCGATGCCTAGCCCTGGGCCAGT
TCAATCACTCCCGCCGATAATTAGCCGGCCTCTGCAATGGGAACTTTTCCAAGACGAGATCGAT
TCTTGGGAAAACACCACCTCAGTTTCCATTTCTGTTTTTTGGGTCCGGAAGTGGCATCGTGTCT
TCCCAGAAGCGCCTCCAAATGGTGCCACCATTAGCCAGGGGGAAGCCGGGCGGCAGTCACTTAG
CTGCTGCAATTTAAATACTTTTTAATTGATTACTAATTGCGGCGAGGCAAGTGCCAAAACAAGA
CGACGAGGACGACTTGGCTGTGACGTTTTCGATGCAACCCGACCGGGGACTGCCACTCTTTAGC
CAGTTAATTGGCAGCAAAAGCGACAGTGGCAGCGGCAGCAACTGCTTTTCACCAGGAAATCAAT
AAACGCTCGTCCAGCGGCAAAAGTAATCGCAACACGCACATCTCAATTTCGGTGGCAGAAAAAA
AACCCTCACCAGCTCAGTTCCCCGTGCCAAATTAACCAGAGCAAACATAGCCCAGTTTCTTCTC
CTGCGGCATGTGAAAAGGCAAACAGTGCTGGCCGGAAAATACCCAGCAAAACACCTGAGTTCTA
GTTGCGATTTTCGGAATTGGACTATAAAAGGCGGCCGTCGGGTAGCTTTCTTCACTCACAACCA
GTCCAAAAGCATCTCCAACATGAAGTTCTTCCAGCAAATCGTCCTGGGTCTGATGCTCGTCCTG
GCCATCATGGGTTCGCTGGCTAGCGCCAAGCCCCAGGAGGCCGAGGAACCGGAGGAGTCGCTGG
TCGGGGACTCAGAGTCGGGACAGTCCGTGCCCGAGGACGCCCAGCAGGACTACCTCAACGTGGC
GGACCTCACCACTGCCGCTCCTACATGGTGGTGGAACTAGAGCCCGGGAATTCTAGAGGATTTT
AACAATCTTTTGTTTTTTTTTGCTAATCTAATGTACTATAATTGCAAAGAATTTACGGTGTTCC
ATATTCAATAAACCTATTTAAAGCTGAATACAATATTTACGTTAATAAATGTTCTTGATACGAT
AAATTTACTTAATTAAAATACATTACATTTCAGTTAAATATTTGTAAAATAAAATATATTTAAA
AAAATATTTAATTCACTAGTTGTGGGAGTTCATTAGACTTATTATTTGTTTTTATTAAATTGTA
ATTTGTATCAAAAGTTTATTTTGCCAAACAGTGAATCTTAAAATATATATCAAGTTCATTGCAC
AAATTAACTTATAAATTGTCACCAAAAATTTAGAAATCAACCTATGTAAATAAATTTAACAAAC
CAATCATATCTTGAATTTAAATATATAAAAGAGGAGATAAACCATTTATAAAATGGTCTCACCT
TTTTTTAGTTTATTTGATGCATGTTTTAATTTTGCTAAAATCATATTCTGATGTCTATTCATTT
TGCCAGCCAGTAATCTAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTCGACCTATAAAT

ACCCACCGCAAATGCCGCAACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAATGGACAAT
TATCCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTA
ACGAAATTAATCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTT
ACTTTAATTGAAGGCTGCCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTA
ATACGTTAAGTCGGAGCTGCGGAACGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCA
TTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGATCGCTGCGATTGTTTATGAGTTATGTCC
TGTGTCAGCGTAGATCGTTAATAACGCGAACGGTGGCCCACCCACAATAATGTGGGAGGCCACA
TACATGCTCACCATCGAACTCACGATTGCTGTGGTATACGTTTAGGGCGACCGAACGGAGTGTT
AATGGTTATCCACCCGGCTGCCACATGATTCAGTGCGATGCGTACCCGAAGATATTCTCGGGAA
CCCAAGTAACTTATCGACATTTAATTGGCGGTTCCCCTTGCAATGCCTGGGTTATATAAACGAC
ATATATTTATGTTGGGTCTTCTATACTCTGCGTAACAAGTAAATCCTTTCCCTCGGTGTCTGTT
ATCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAA
TTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAG
GCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCAT
GCAATAGTTTTGTGCCAATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGT
GCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTG
GCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAGACAAA
TAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGACATTGAACTTGAA
AACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATGGGTGTCT
CTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAGAAAATA
AAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCC
GGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATA
ATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATT
TATTTTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCC
GCTTGATAATGATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGCGT
ATAAATCATCGCCGCCGAAGTCTGGCAAAACAGCAAATTGAAAACTGCAAGCTGAAAACTGAAA
ACTGAAAACTGTAACCCAAACAAACACAGCATCCCACACGACGAGGTGAAAATGAAAATAAATA
CGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGAT
CGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGCACACACAATTAATT
AATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATTAAGGCAG
CATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCA
TAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATAT
TTAAAAATAAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGA
AACTGTCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTA
TTGCCCAGACATTTTCCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTT
CGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGC
TTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAA
TCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAGAGCATTACCCGCTTA
GGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGGCCAGCTGA
GTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGATA
CATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTG
AATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTT
AGTGCTGTTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATC
AGGTTTATTTACATTTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTA
GAATCATTTAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTG
GATGAAGTAAAAAAGCCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGAT

GGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAAGACTCGAAAATACTCTCACTCAC
TGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTGAATCCTAAACGTATCT
GAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCCTTTC
AAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGGCC
CGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTA
GAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCAT
GGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTC
GTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGA
CATATTGAGTCCAACAGTCGTAAGCGCGCCACGGTCCACAGAA

>*Δ broad*

AAGAGCCCAAGGTCGGTCGTTTAGCTTGGCCAAAACCTACCCATCCAGCTGGCACTTCCACCAA
CGGCACCAAGACACGAGCGGAAAATAAAAAGCCACACCACCCCACTTAGAACTCCGTTTAGCAG
CAGTTGTTCAAACAGAAATTGGCTGGCTTCGGCCGGTTCAGCcTCAGTTGATAATTATTATAAT
ATCTATGTTCTTGCCTATCGCTGGGCCCTAATTGGCCCAGACAAAGGCACCGTTTTTATGCATA
ACTGGAGGCTTACAATTTGGCcTTCGACACGCGGTCCTTCACATTGCCAAAAAAGAACGAGAAC
TCgGCAAGCCAATTACACTCGAAGAAGCAGCGGGGATCGTTTCGATGCCTAGCCCTGGGCCAGT
TCAATCACTCCCGCCGATAATTAGCCGGCCTCTGCAATGGGAACTTTTCCAAGACGAGATCGAT
TCTTGGGAAAACACCACCTCAGTTTCCATTTCTGTTTTTTGGGTCCGGAAGTGGCATCGTGTCT
TCCCAGAAGCGCCTCCAAATGGTGCCACCATTAGCCAGGGGGAAGCCGGGCGGCAGTCACTTAG
CTGCTGCAATTTAAATACTTTTTAATTGATTACTAATTGCGGCGAGGCAAGTGCCAAAACAAGA
CGACGAGGACGACTTGGCTGTGACGTTTTCGATGCAACCCGACCGGGGACTGCCACTCTTTAGC
CAGTTAATTGGCAGCAAAAGCGACAGTGGCAGCGGCAGCAACTGCTTTTCACCAGGAAATCAAT
AAACGCTCGTCCAGCGGCAAAAGTAATCGCAACACGCACATCTCAATTTCGGTGGCAGAAAAAA
AACCCTCACCAGCTCAGTTCCCCGTGCCAAATTAACCAGAGCAAACATAGCCCAGTTTCTTCTC
CTGCGGCATGTGAAAAGGCAAACAGTGCTGGCCGGAAAATACCCAGCAAAACACCTGAGTTCTA
GTTGCGATTTTCGGAATTGGACTATAAAAGGCGGCCGTCGGGTAGCTTTCTTCACTCACAACCA
GTCCAAAAGCATCTCCAACATGAAGTTCTTCCAGCAAATCGTCCTGGGTCTGATGCTCGTCCTG
GCCATCATGGGTTCGCTGGCTAGCGCCAAGCCCCAGGAGGCCGAGGAACCGGAGGAGTCGCTGG
TCGGGGACTCAGAGTCGGGACAGTCCGTGCCCGAGGACGCCCAGCAGGACTACCTCAACGTGGC
GGACCTCACCACTGCCGCTCCTACATGGTGGTGGAACTAGAGCCCGGGAATTCTAGAGGATTTT
AACAATCTTTTGTTTTTTTTTGCTAATCTAATGTACTATAATTGCAAAGAATTTACGGTGTTCC
ATATTCAATAAACCTATTTAAAGCTGAATACAATATTTACGTTAATAAATGTTCTTGATACGAT
AAATTTACTTAATTAAAATACATTACATTTCAGTTAAATATTTGTAAAATAAAATATATTTAAA
AAAATATTTAATTCACTAGTTGTGGGAGTTCATTAGACTTATTATTTGTTTTTATTAAATTGTA
ATTTGTATCAAAAGTTTATTTTGCCAAACAGTGAATCTTAAAATATATATCAAGTTCATTGCAC
AAATTAACTTATAAATTGTCACCAAAAATTTAGAAATCAACCTATGTAAATAAATTTAACAAAC
CAATCATATCTTGAATTTAAATATATAAAAGAGGAGATAAACCATTTATAAAATGGTCTCACCT
TTTTTTAGTTTATTTGATGCATGTTTTAATTTTGCTAAAATCATATTCTGATGTCTATTCATTT
TGCCAGCCAGTAATCTAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTCGACCTATAAAT
ACCCACCGCAAATGCCGCAACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAATGGACAAT
TATCCTTAGTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTA
ACGAAATTAATCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTTCTTTATGAATATTTTCCATTT
ACTTTAATTGAAGGCTGCCAATTGTGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTA
ATACGTTAAGTCGGAGCTGCGGAACGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCA
TTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGATCGCTGCGATTGTTTATCGATAATCGCC
CGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGCGGCGATGGCGGCTTTAGATAAG
TAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTATGCACTCGCTGGAAAAG

CTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAGTGCAGCGCGG
CAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAATTACG
GCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGC
ATCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAA
TTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAG
GCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCAT
GCAATAGTTTTGTGCCAATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGT
GCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTG
GCGCAGTGCGCGCCTGCGAGCAGGAAATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAGACAAA
TAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGACATTGAACTTGAA
AACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATGGGTGTCT
CTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAGAAAATA
AAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCC
GGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAATA
ATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGTTGATGTATCCTGATTAAG
CTAGGTCTCTTTGAACTTGTGCAGCTCCACGGGATAGCCGAACGTTTCGGAGTTTGTGTGTCTT
TCTTCCATATGCTTCGTGTAATTACATTTATTCCACAAACAATAAAATAGAGGGGACCTGTCTA
AAGAACAACACATGGCAAAGTGGGAATACAACCAGAAAAGTGGTCCAATAAACAAAGAACGTGA
ATCACTCAGGAATGAGAACCATCGTGAGCCTTCAGCAACAATTACCCATGGCATCTAAATGGCG
AGTACTTTACAACGCCTGACAAAGATAGCTTACGAATCATGTGACGCGAGTATCAATAATTTTG
TATGAGTCTCACCCAGATTCTGATCCGCCGTTAAGCTCACCCGTTAGGCAAACTCTTTGGCATC
GATGGTAGTTAGCTCCATGTAAACAATTCTTACTAGAGGTAGGCCCAGCGTGCGCGCGCTTACC
TATTGAGGGTTTGATCGCCCTTTAGTAGAGTCGGGGTCCGGCTTCAGGTATCGAATCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCA
TAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATAT
TTAAAAATAAATATAAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACTTCTTCTTAGGA
AACTGTCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTA
TTGCCCAGACATTTTCCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTT
CGTCAGCGGAGGTCGTAAACGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGC
TTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAA
TCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGCGGCCATGGCCATTAGAGCATTACCCGCTTA
GGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGGCCAGCTGA
GTGGAGTGCTGGAACGCACTTCTTAATTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGATA
CATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTG
AATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCCAGGTAATTCATATTT
AGTGCTGTTCGCAAAGACCTGTCCCAGATACTCTGTTTATAGGTATAATTATTAAGTGCATATC
AGGTTTATTTACATTTATATCGTATTATATTGGTAACTGCAGCAGATGCTGTGCTACAAATTTA
GAATCATTTAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTG
GATGAAGTAAAAAAGCCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGAT
GGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAAGACTCGAAAATACTCTCACTCAC
TGTAAGTGAACCCAGTGTTTTGTAATTGCCTAGCACATAAATCAGCTGAATCCTAAACGTATCT
GAAGGCCAGGAGTGTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCCTTTC
AAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGGCC
CGTGATACTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACTA
GAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCAT
GGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTC

GTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAACCTCATATAAAACGTGGCCGA
CATATTGAGTCCAACAGTCGTAAGCGCGCCACGGTCCACAGAA

REFERENCES AND NOTES

1. J. L. Gomez-Skarmeta, I. Rodriguez, C. Martinez, J. Culi, D. Ferres-Marco, D. Beamonte, J. Modolell, Cis-regulation of achaete and scute: Shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev.* **9**, 1869–1882 (1995).

2. E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yanez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, A. Stark, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).

3. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).

4. M. L. Howard, E. H. Davidson, *cis*-Regulatory control circuits in development. *Dev. Biol.* **271**, 109–118 (2004).

5. C. H. Yuh, E. H. Davidson, Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069–1082 (1996).

6. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).

7. S. B. Carroll, Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).

8. X. Wang, H. M. Chamberlin, Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48* ovo gene. *Genes Dev.* **16**, 2345–2349 (2002).

9. 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).

10. A. P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, D. L. Stern, Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).

11. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jonsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).

12. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).

13. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U.S.A.* **104**(Suppl 1), 8605–8612 (2007).

14. J. Banerji, S. Rusconi, W. Schaffner, Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).

15. W. Schaffner, Enhancers, enhancers – From their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).

16. K. Harding, T. Hoey, R. Warrior, M. Levine, Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *EMBO J.* **8**, 1205–1212 (1989).

17. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).

18. A. Pavlopoulos, M. Akam, Hox gene *Ultrabithorax* regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2855–2860 (2011).

19. M. W. Perry, A. N. Boettiger, J. P. Bothma, M. Levine, Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).

20. N. Frankel, G. K. Davis, D. Vargas, S. Wang, F. Payre, D. L. Stern, Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).

21. M. Z. Ludwig, Manu, R. Kittler, K. P. White, M. Kreitman, Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. *PLOS Genet.* **7**, e1002364 (2011).

22. J. Crocker, D. L. Stern, Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. *Development* **144**, 3095–3101 (2017).

23. R. D. Bickel, A. Kopp, S. V. Nuzhdin, Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. *PLOS Genet.* **7**, e1001275 (2011).

24. E. Preger-Ben Noon, G. Sabaris, D. M. Ortiz, J. Sager, A. Liebowitz, D. L. Stern, N. Frankel, Comprehensive analysis of a *cis*-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* **22**, 3021–3031 (2018).

25. G. Sabaris, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with multiple roles: Pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet.* **35**, 423–433 (2019).

26. G. Sabaris, D. M. Ortiz, I. Laiker, I. Mayansky, S. Naik, G. Cavalli, D. L. Stern, E. Preger-Ben Noon, N. Frankel, The density of regulatory information is a major determinant of evolutionary constraint on noncoding DNA in *Drosophila*. *Mol. Biol. Evol.* **41**, msae004 (2024).

27. D. J. McKay, J. D. Lieb, A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev. Cell* **27**, 306–318 (2013).

28. J. C. Pearson, D. J. McKay, J. D. Lieb, S. T. Crews, Chromatin profiling of *Drosophila* CNS subpopulations identifies active transcriptional enhancers. *Development* **143**, 3723–3732 (2016).

29. C. M. Uyehara, S. L. Nystrom, M. J. Niederhuber, M. Leatham-Jensen, Y. Ma, L. A. Buttitta, D. J. McKay, Hormone-dependent control of developmental timing through regulation of chromatin accessibility. *Genes Dev.* **31**, 862–875 (2017).

30. M. Bozek, R. Cortini, A. E. Storti, U. Unnerstall, U. Gaul, N. Gompel, ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Res.* **29**, 771–783 (2019).

31. M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).

32. S. Fu, Q. Wang, J. E. Moore, M. J. Purcaro, H. E. Pratt, K. Fan, C. Gu, C. Jiang, R. Zhu, A. Kundaje, A. Lu, Z. Weng, Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers. *Nucleic Acids Res.* **46**, 11184–11201 (2018).

33. W. Saelens, O. Pushkarev, B. Deplancke, ChromatinHD connects single-cell DNA accessibility and conformation to gene expression through scale-adaptive machine learning. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.07.21.549899.

34. G. Kalay, J. Lachowiec, U. Rosas, M. R. Dome, P. J. Wittkopp, Redundant and cryptic enhancer activities of the *Drosophila yellow* gene. *Genetics* **212**, 343–360 (2019).

35. I. Laiker, N. Frankel, Pleiotropic enhancers are ubiquitous regulatory elements in the human genome. *Genome Biol. Evol.* **14**, evac071 (2022).

36. O. Nagy, I. Nuez, R. Savisaar, A. E. Peluffo, A. Yassin, M. Lang, D. L. Stern, D. R. Matute, J. R. David, V. Courtier-Orgogozo, Correlated evolution of two copulatory organs via a single *cis*-regulatory nucleotide change. *Curr. Biol.* **28**, 3450–3457.e13 (2018).

37. J. W. Hong, D. A. Hendrix, M. S. Levine, Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).

38. J. W. Blayney, H. Francis, A. Rampasekova, B. Camellato, L. Mitchell, R. Stolper, L. Cornell, C. Babbs, J. D. Boeke, D. R. Higgs, M. Kassouf, Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* **186**, 5826–5839.e18 (2023).

39. S. Pott, J. D. Lieb, What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).

40. S. Jeong, A. Rokas, S. B. Carroll, Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* **125**, 1387–1399 (2006).

41. P. K. Geyer, V. G. Corces, Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. *Genes Dev.* **1**, 996–1004 (1987).

42. P. J. Wittkopp, K. Vaccaro, S. B. Carroll, Evolution of *yellow* gene regulation and pigmentation in Drosophila. *Curr. Biol.* **12**, 1547–1556 (2002).

43. Y. Xin, Y. Le Poul, L. Ling, M. Museridze, B. Muhling, R. Jaenichen, E. Osipova, N. Gompel, Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20636–20644 (2020).

44. G. Kalay, P. J. Wittkopp, Nomadic enhancers: Tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among *Drosophila* species. *PLOS Genet.* **6**, e1001222 (2010).

45. H. Biessmann, J. M. Mason, Progressive loss of DNA sequences from terminal chromosome deficiencies in Drosophila melanogaster. *EMBO J.* **7**, 1081–1086 (1988).

46. M. Martin, Y. B. Meng, W. Chia, Regulatory elements involved in the tissue-specific expression of the *yellow* gene of *Drosophila*. *Mol. Gen. Genet.* **218**, 118–126 (1989).

47. H. Hinaux, K. Bachem, M. Battistara, M. Rossi, Y. Xin, R. Jaenichen, Y. Le Poul, L. Arnoult, J. M. Kobler, I. C. Grunwald Kadow, L. Rodermund, B. Prud'homme, N. Gompel, Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele. *Dev. Biol.* **438**, 111–123 (2018).

48. 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).

49. L. F. Sobala, P. N. Adler, The gene expression program for the formation of wing cuticle in Drosophila. *PLOS Genet.* **12**, e1006100 (2016).

50. J. D. Watson, T. A. Baker, P. B. Bell, A. Gann, M. Levine, S. C. Harrison, *Molecular Biology of the Gene* (Pearson, ed. 7, 2013).

51. E. Z. Kvon, R. Waymack, M. Gad, Z. Wunderlich, Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).

52. 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).

53. B. Lim, M. S. Levine, Enhancer-promoter communication: Hubs or loops? *Curr. Opin. Genet. Dev.* **67**, 5–9 (2021).

54. A. Panigrahi, B. W. O'Malley, Mechanisms of enhancer action: The known and the unknown. *Genome Biol.* **22**, 108 (2021).

55. 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).

56. K. J. 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).

57. M. Calleja, H. Herranz, C. Estella, J. Casal, P. Lawrence, P. Simpson, G. Morata, Generation of medial and lateral dorsal body domains by the *pannier* gene of *Drosophila*. *Development* **127**, 3971–3980 (2000).

58. G. Struhl, D. A. Barbash, P. A. Lawrence, Hedgehog organises the pattern and polarity of epidermal cells in the *Drosophila* abdomen. *Development* **124**, 2143–2154 (1997).

D Lines			
Line 1	Line 2	Kolmogorov-Smirnov statistic	P-value
D0	D1	0.916666667	5.384E-06
D0	D2	0.833333333	0.0001154
D0	D3	0.916666667	0.0007541
D0	D4	1	3.846E-07
D0	D5	1	2.071E-07
D0	D6	1	1.479E-06
D0	D7	1	1.479E-06
D0	D8	1	6.804E-06
D0	empty	1	3.846E-07
D1	D2	0.538461538	0.029515
D2	D3	0.666666667	0.0340557
D3	D4	0.833333333	0.0028748
D4	D5	0.395604396	0.198732
D5	D6	0.493506494	0.0683769
D6	D7	0.818181818	0.0006549
D7	D8	1	1.191E-05
D8	empty	1	4.021E-06

E lines			
Line 1	Line 2	Kolmogorov-Smirnov statistic	P-value
D2	E0	1	3.846E-07
D2	E1	1	4.021E-06
D2	E2	1	1.748E-06
D2	E3	1	1.748E-06
D2	empty	1	1.923E-07
E0	E1	0.638888889	0.0179567
E1	E2	1	2.165E-05
E2	E3	0.7	0.0123406
E3	empty	0.438461538	0.171838

pupal abdominal epidermis genes	pupal wing genes
Adf1	Atf3
Alh	abo
Antp	ash2
aop	achi
Atf3	Ada2b
Atf6	Adf1
bi	adp
bip2	Aef1
bon	Alh
Brf	Antp
bs	aop
bun	apt
Camta	ap
caz	ara
cg	pnt
CG10462	az2
CG10949	bi
CG11247	Bgb
CG11617	bigmax
CG11710	bs
CG11902	bbx
CG12299	bon
CG13185	BEAF-32
CG13796	bowl
CG14711	BtbVII
CG15073	bun
CG15439	cbt
CG17068	cic
CG17359	cnc
CG17385	caup
CG17612	CG10209
CG18262	Zif
CG2202	CG10274
CG2712	CG10321
CG2926	CG10366
CG31126	CG10543
CG31460	CG10565
CG31709	CG10631
CG32847	CG10949
CG3407	CG10979
CG3909	Kdm2
CG42724	CG11122
CG44002	CG11247
CG45071	CG11414
CG4730	CG11456
CG4914	pdm3
CG4935	CG11696
CG5708	CG11723
CG6654	CG11906

CG6686	E(var)3-9
CG7099	CG12054
CG7839	CG12071
CG8301	mxc
CG8319	CG12219
CG8578	CG12236
CG9650	CG12299
CG9705	CG12391
CG9895	MEP-1
Chi	CG13204
Clamp	mld
Cnot4	CG14117
CrebA	CG14667
CREG	CG14710
CSN5	CG14711
ct	Asciz
cwo	CG15011
D1	CG15073
Dad	CG1529
dl	CG15514
dlt	CG15715
dmrt93B	CG1603
Dp	CG1620
DpseGA22108	Coop
Dr	CG1647
Dsp1	cwo
dsx	CG17186
Dya	CG17385
Dyak	CG17806
dysf	CG17829
E(bx)	Xrp1
e(r)	CG1792
E(spl)mβ-HLH	Clamp
E(var)3-9	Br140
e(y)3	CG2116
ea	CG2202
EcR	CG2790
Eip74EF	CG2889
Eip75B	CG30020
Eip93F	Blos1
EloB	CG3065
EloC	Nf-YC
en	CG31224
escl	Cenp-C
Etl1	CG31365
Ets97D	CG31388
exd	CG31510
FoxK	CG32006
ft	CG32767
fu2	CG32772

GABA-B-R3	CG3281
Gas41	Glut4EF
gcm	CG3407
Gdi	Jarid2
gem	CG3815
grau	CG3838
grh	CG3847
gsb-n	Nf-YA
H	CG3995
Hand	CG4282
HDAC6	CG4360
HmgD	CG4424
HmgZ	CG4936
Hr3	CG6254
Hr38	CG6276
Hsf	CG6654
hth	CG6686
inv	CG6689
jim	CG6701
Kr-h1	CG6769
l(1)10Bb	CG6791
lid	CG45071
luna	Unr
Mad	CG7372
maf-S	CG7518
Max	CG7987
MBD-R2	row
mbf1	CG8108
Med	Vps45
MED10	ADD1
MED11	CG8301
MED18	CG8319
MED23	CG8388
MED30	Rbsn-5
Meics	CG8765
Mes4	CG8944
Mitf	Rabex-5
mof	CG9215
Mondo	Mtp
MP1	CG9705
mub	CG9776
Nelf-E	M1BP
net	CG9932
Neu2	chn
NFAT	CHES-1-like
Nfi	cg
Nf-YB	corto
NK7.1	crol
nom	crp
nonA	crc

nonA-l	CTCF
odd	ci
osa	cyc
p53	CrebA
Paip2	D19A
pan	D19B
Pcl	d4
Pdp1	da
pnr	Deaf1
pnt	dwg
Pop2	Dek
Poxn	dm
Qtzl	disco-r
Rbf	Dref
reb	dom
Ref1	dl
Ref2	Dsp1
REPTOR	Doc1
REPTOR-BP	Doc3
rn	Dp
RnpS1	Dr
row	E2f2
Rpb10	eg
Rpb11	EcR
Rpb4	Eip74EF
Rpb5	Eip75B
Rpb7	ecd
Rpb8	Eip93F
Rpll15	elB
Rpll18	en
Rpll33	E(bx)
Rplll128	ewg
RpL7-like	esg
Rtf1	ERR
schlank	Ets97D
Scm	Ets98B
SF1	emc
sima	exd
simj	foxo
Sirt1	fru
Sirt4	ftz-f1
skd	GATAd
slbo	gem
Smox	Gnf1
Smr	grn
Snr1	grh
Sox21b	grau
SoxN	Gug
Spec	hang
Spp	Hsf

spt4	HLH106
Spt5	Hnf4
sqz	her
SREBP	HmgZ
srp	hth
Sry-delta	Hr38
Ssb-c31a	Hr39
Ssdp	Hr46
su(Hw)	Hr96
Su(var)205	Hr78
Taf10	HP1b
Taf10b	HP1c
Taf11	inv
Taf4	jim
TBPH	Jra
TFAM	kay
Tfb1	ken
Tfb4	kin17
TfIIA-L	klu
TfIIA-S	knrl
TfIIB	kn
TfIIFalpha	lilli
TH1	Lmpt
tHMG1	schlank
Tif-IA	lola
toe	maf-S
toy	Max
Trf2	MBD-like
Trl	MBD-R2
trr	Med
Tudor-SN	Mi-2
tx	TFAM
Ubx	Mio
Utx	Mnt
vis	mod
Vps45	Mad
vri	MTA1-like
Vsx2	MTF-1
vvl	mbf1
woc	Myb
Zif	mip120
	Mrtf
	Mef2
	nej
	NFAT
	noc
	nub
	Optix
	ovo
	p53

pan
Pdp1
pita
pho
phol
Pcl
ph-p
Psc
pdm2
Pep
Rel
rn
sd
shn
sbb
Scm
sv
stc
Stat92E
sima
sba
slbo
Smox
Sox14
Sox15
salm
ss
sqz
Su(z)12
Su(H)
Su(var)205
Su(z)2
tai
tgo
Trf
Tip60
E2f1
trh
ttk
trx
Trl
trr
usp
Usf
vvl
vis
wdn
woc
Xbp1
yem

z
zf30C
zfh1

name	sequence
D0 for	GTCTAGAGCCCGGGCGAATTAAGAGCCCAAGGTCGGTCG
D0 Rev	CCGGCGCTCCTCGAGGGATCTTCTGTGGACCGTGGCGC
D3-4 A Reverse	GACATAACTCATAAACAATCGCAGCGATCTCCC
Spacer D3-4 for	GATTGTTTATGAGTTATGTCCTGTGTCAGCG
Spacer D3-4 Rev	AATAATTAGATAACAGACACCGAGGGAAAGG
D3-4 B Forward	GTGTCTGTTATCTAATTATTCCGTTTAAGGACGC
E2-1 A Reverse	GATACATCAACATAATTGCCACACGATTATGC
Spacer E2-1 for	GGCAATTATGTTGATGTATCCTGATTAAGCTAGGT
Spacer E2-1 rev	TCTGCGCTCGATTCGATACCTGAAGCCGGAC
E2-1 B Forward	GGTATCGAATCGAGCGCAGAATGCGGCC

Paper III. 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, Nicolas Gompel

SCIENCE ADVANCES | Vol 6, Issue 49 | 2 Dec 2020

DOI: 10.1126/sciadv.abe2955

From: Poul, Y. L. et al. Regulatory encoding of quantitative variation in spatial activity of a Drosophila enhancer. Sci Adv 6, eabe2955 (2020).

Supplementary data files S1 – S4 in Excel format:

https://www.science.org/doi/10.1126/sciadv.abe2955

SCIENCE ADVANCES | RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

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

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

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. However, 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 $spot^{196}$ 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 $spot^{196}$ 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 $spot^{196}$ 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 $spot^{196}$ 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

[1]Evolutionary Ecology, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. [2]Human Biology and Bioimaging, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. [3]Quantitative and Computational Biology, Departments of Biological Sciences, Chemistry, Physics and Astronomy, and Computer Science, University of Southern California, Los Angeles, CA 90089, USA. [4]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany. [5]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. [6]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.)

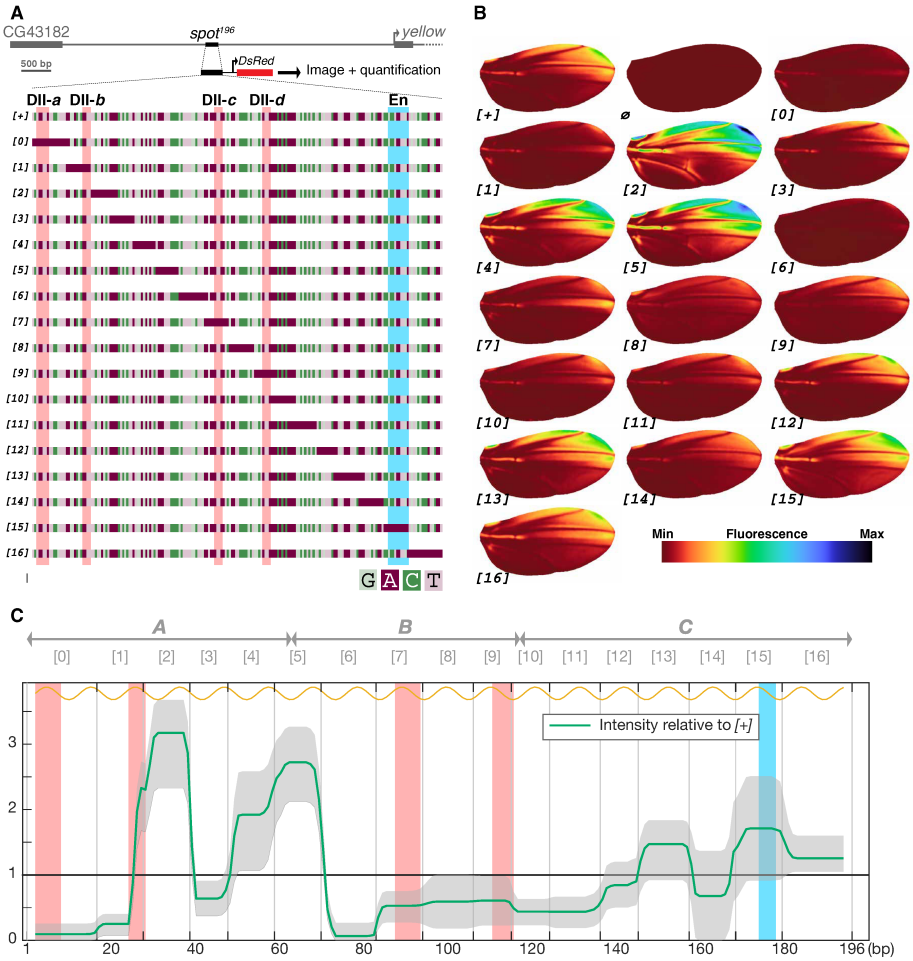

Fig. 1. A mutational scan of the *D. biarmipes spot*^{196} enhancer with a quantitative reporter assay. (**A**) Wild-type (*[+]*) and mutant (*[0]* to *[16]*) versions of the *spot*^{196} enhancer from the *D. biarmipes yellow* locus (depicted at the top) were cloned upstream of a DsRed reporter to assay 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 (Dll-a, Dll-b, Dll-c, and Dll-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 (ø). 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 smoothened, 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

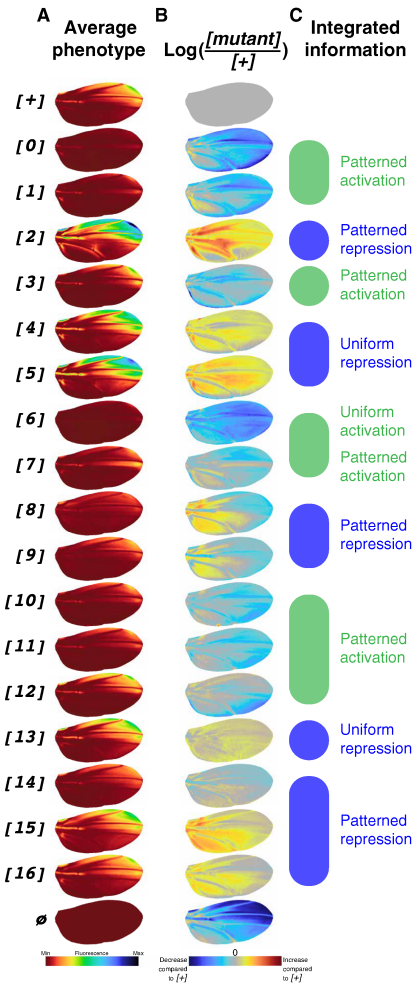

Fig. 2. Trans-regulatory integration along the *spot*^{196} sequence. (**A**) Average phenotypes reproduced from Fig. 1B. (**B**) *logRatio* images [log([*mutant*]/[+]) for intensity values of each pixel of registered wing images] reveal what spatial information is integrated by each position along the 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.

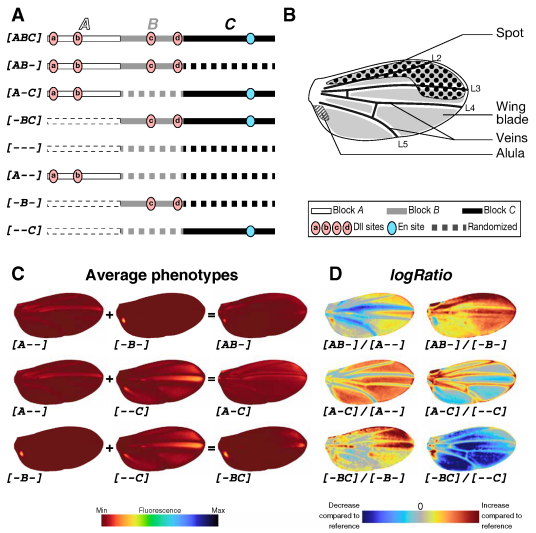

Fig. 3. Regulatory interactions in the *spot*^{196} sequence. (**A**) Schematics of constructs with block randomizations. The *spot*^{196} sequence was arbitrarily divided into three blocks (A, 63 bp; B, 54 bp; C, 79 bp). In each construct, the sequence of one, two, or all three blocks was randomized. (**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).

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 sequence 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*$^{196\,[0]}$ to *spot*$^{196\,[16]}$, or *[0]* to *[16]* in short; Fig. 1A) is identical to the wild-type *spot*^{196} (*[+]* 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

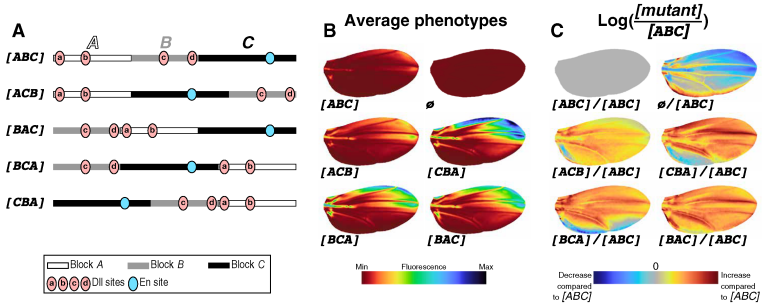

Fig. 4. Block permutations scale the activity of the *spot*^{196} enhancer. (**A**) Schematics of constructs with block permutations. In this series, the same blocks of sequences as in Fig. 3A were permutated. (**B**) Average phenotypes resulting from constructs in (A). Colormap of average phenotypes 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*^{196}. 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*^{196} 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*^{196} 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*^{196} 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*^{196} (*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*^{196} 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*^{196} 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*^{196}.

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*^{196} 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*^{196} 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*^{196}, 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*^{196} 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*^{196} 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*^{196} 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*^{196} 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 perturbates 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*^{196}. 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.

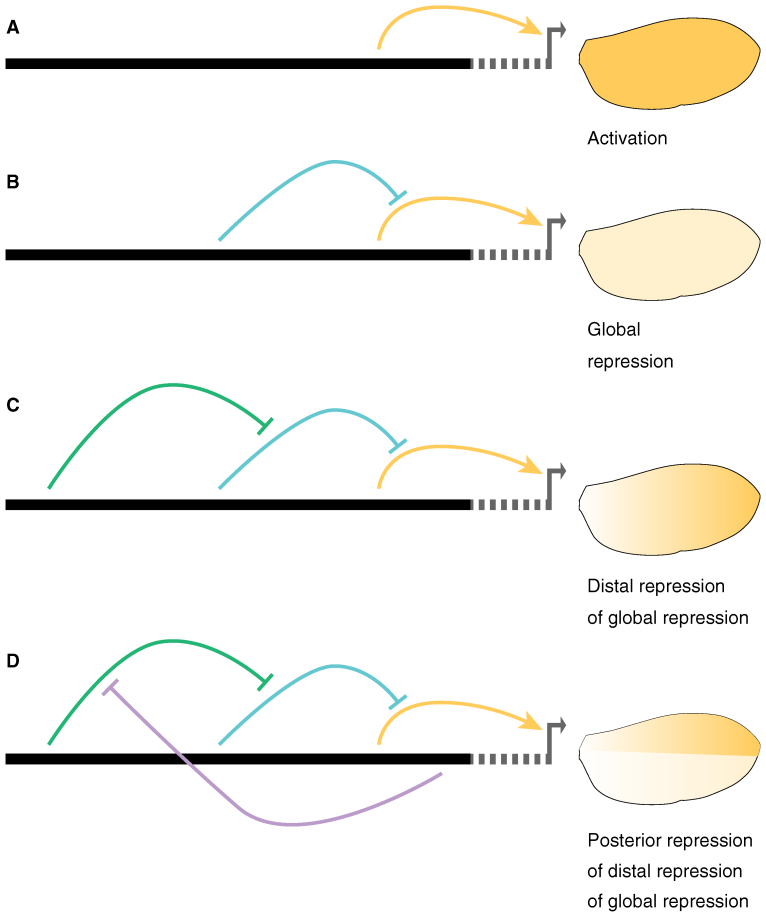

Fig. 5. A model of the regulatory logic governing the *spot*^{196} enhancer. (**A** to **D**) The schematics show step by step how regulatory information and interactions integrated along the enhancer sequence produce a spatial pattern of activity. (A) Three independent inputs, respectively, in blocks *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*^{196} 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 activities 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 $spot^{196}$. Instead, we think that the effects of A-tract mutations are mostly due to disrupted TFBSs along the enhancer sequence.

The regulatory logic underlying $spot^{196}$ enhancer activity

The main finding of our study is that the $spot^{196}$ 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 $spot^{196}$. The main levels of $spot^{196}$ 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 $spot^{196}$. The $spot^{196}$ 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 $spot^{196}$ 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 $spot^{196}$ 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 ϕC31-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* $spot^{196}$ 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 ϕC31-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′-GAGCCCGGGCGAATT-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′-GAGCCCGGGCGAATT-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°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°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°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°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× 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 µm) 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 (ø) was used as a proxy to measure noise and tissue autofluorescence. The median raw fluorescent image was computed across all ø images and used to remove autofluorescence, subtracted from all raw images before the following steps. All variation of fluorescence below the median ø 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 [~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_individuals × n_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

1. D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: From properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
2. M. Levine, Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
3. I. S. Peter, E. H. Davidson, *Genomic Control Process: Development and Evolution* (Academic Press, ed. 1, 2015), p. 460.
4. S. B. Carroll, From pattern to gene, from gene to pattern. *Int. J. Dev. Biol.* **42**, 305–309 (1998).
5. F. Spitz, E. E. M. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
6. J. Crocker, A. Tsai, D. L. Stern, A fully synthetic transcriptional platform for a multicellular eukaryote. *Cell Rep.* **18**, 287–296 (2017).
7. 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).
8. E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Elsevier, 2010).
9. 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).
10. 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).
11. 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).
12. 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).
13. T. Fuqua, J. Jordan, M. E. van Breugel, A. Halavatyi, 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).
14. 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).
15. J. Crocker, D. L. Stern, Functional regulatory evolution outside of the minimal *even-skipped* stripe 2 enhancer. *Development* **144**, 3095–3101 (2017).
16. 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).
17. 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).
18. 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).
19. 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).
20. 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).
21. 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).
22. S. Neidle, *Principles of Nucleic Acid Structure* (Academic Press, 2010), p. 302.
23. 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).
24. E. Z. Kvon, Y. Zhu, G. Kelman, C. S. Novak, I. Plajzer-Frick, M. Kato, T. H. Garvin, Q. Pham, A. 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).
25. 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).
26. 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).
27. 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).
28. 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).
29. 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).
30. 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).
31. L. J. Zhu, R. G. Christensen, M. Kazemian, C. J. Hull, M. S. Enuameh, 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. Hannenhalli, 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. Rübel, M.-Y. Huang, S. Chatoor, A. H. De Pace, L. Simirenko, C. Henriquez, A. Beaton, R. Weiszmann, 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 φC31. *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. Mühling, 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).

Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/49/eabe2955/DC1

Supplementary Materials for

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*, Nicolas Gompel*

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

Published 2 December 2020, *Sci. Adv.* **6**, eabe2955 (2020)
DOI: 10.1126/sciadv.abe2955

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/49/eabe2955/DC1)

Supplementary Materials

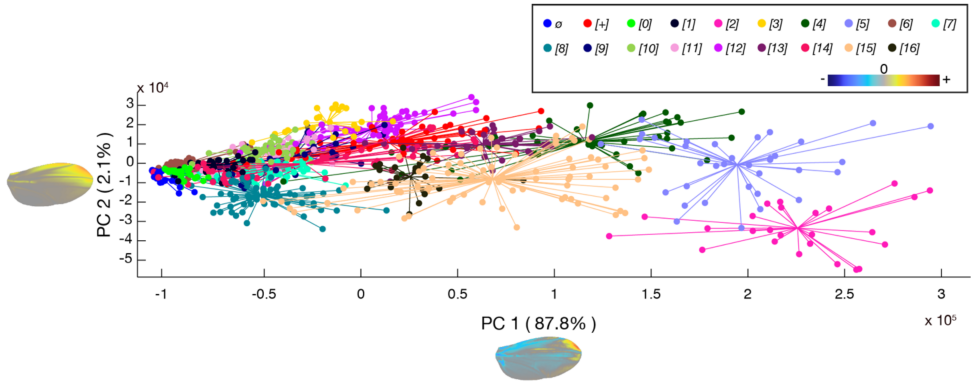

Figure S1. **First two axes of variation in a principal component analysis of all individual wings used to generate the average reporter expression of Figure 1**. Each wing is depicted by a colored dot, and each construct by a color. PC1 captures 87.8% of the variation and corresponds to overall changes in the activity of the *spot*^{196} CRE. PC2 captures 2.1% of the variation and appears to represent spatial difference in CRE activity between lines. The direction of variation along each principal component is represented on a wing with a colormap next to each axis.

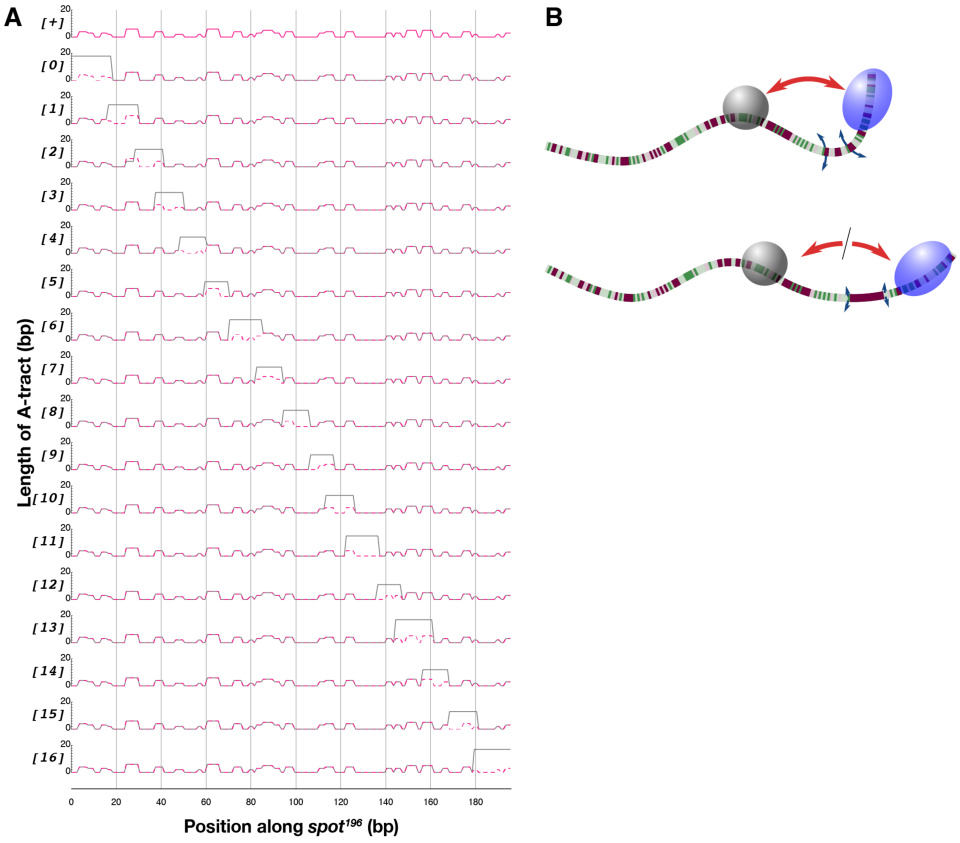

Figure S2. Local rigidity along the wild-type and mutant *spot*^{196}. (**A**) Each graph is a plot of the length of the longest consecutive A_nT_n sequence that a base pair participates in, a proxy for sequence rigidity at this position. The first graph on top is the wild type (*[+]*) alone. The remaining graphs show plots for each mutant (*[0]*, …, *[16]*) with a solid black line, compared to the wild type represented with a dotted magenta line. (**B**) Schematics illustrating the hypothetical consequence of local DNA rigidity (caused by an A-tract) on TF interactions. A flexible linker between two TFBSs would favor interactions between 2 bound TFs, while a stiffer linker of the same length would limit, or prevent these interactions.

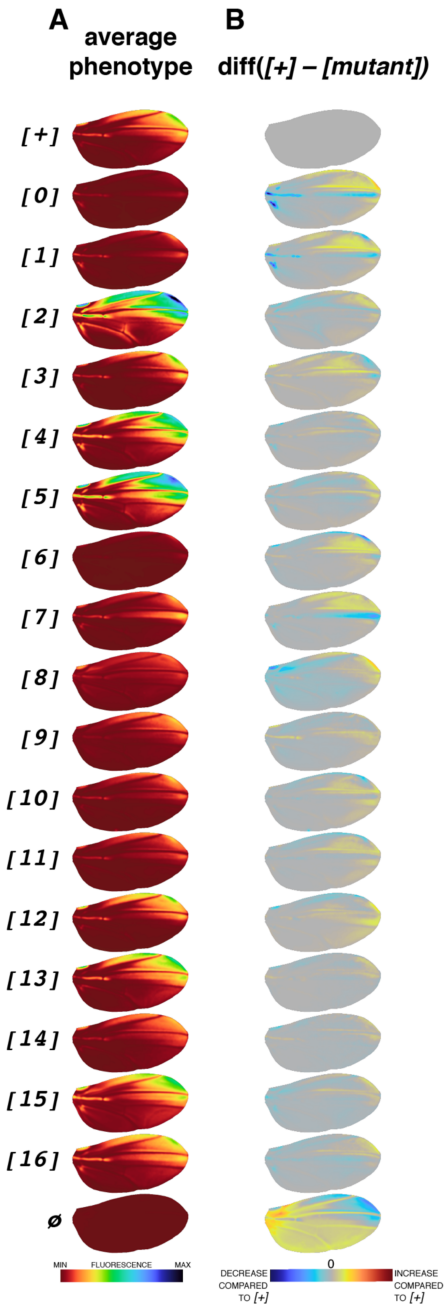
A
average phenotype
B
diff([+] – [mutant])
[+]
[0]
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
ø
MIN
FLUORESCENCE
MAX
0
DECREASE COMPARED TO [+]
INCREASE COMPARED TO [+]

Figure S3. **Pattern changes between wild-type and mutant *spot*^{196} constructs**. (**A**) Average phenotypes reproduced from Figure 1B. (**B**) difference images (*[+] – [mutant]*) for intensity values of each pixel of registered wing images) highlight changes in the distribution of the enhancer activity across the wing. Note that this operation introduces a visual bias towards changes in region of high expression, contrasting with *logRatio* images of Figure 2.

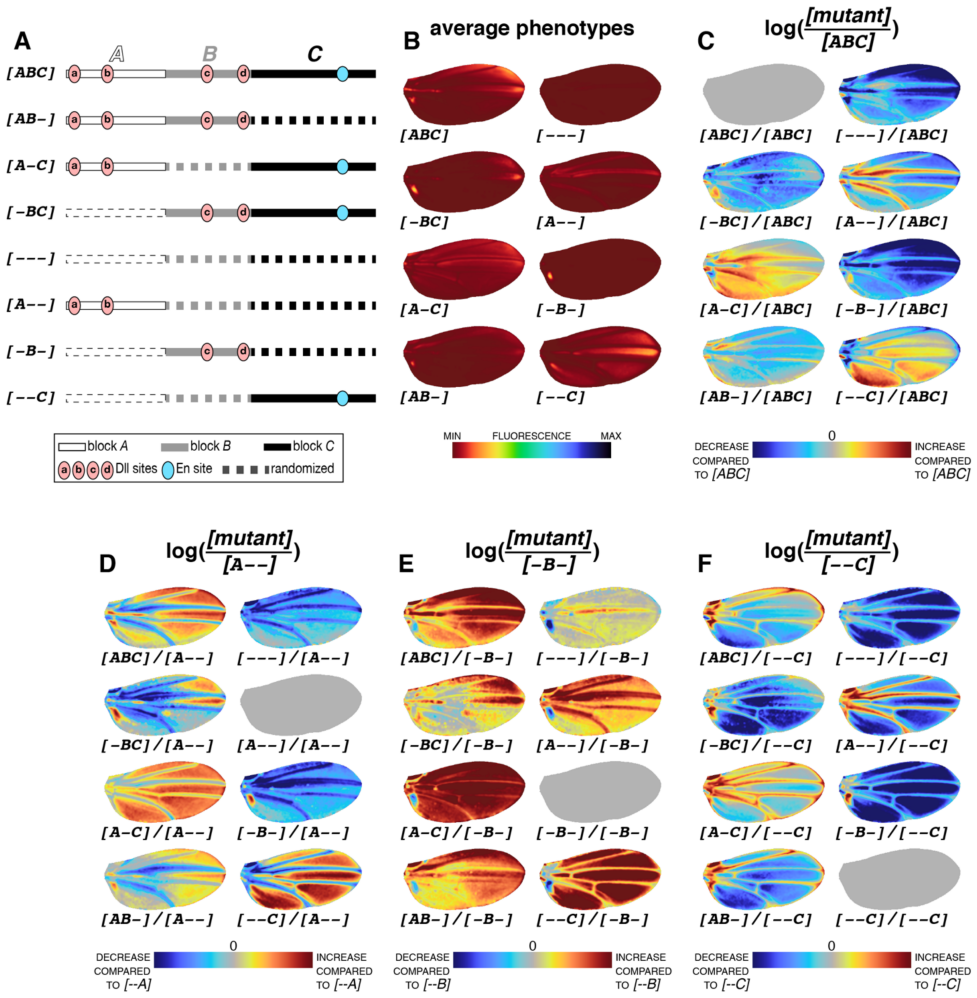

Figure S4. ***logRatio* of all block constructs.** (**A**) Schematics of block constructs repeated from Figure 3A for legibility. (**B**) Average phenotypes of constructs shown in (**A**), repeated from Figure 3B for legibility. Colormap of average phenotypes normalized for all constructs of the block series, including block permutations of Figure 4B. (**C**) Average phenotypes in (**B**) compared to the average phenotype of the wild type *[ABC]* (*logRatio*). (**D**) Average phenotypes in (**B**) compared to the average phenotype of *[A--]* (*logRatio*). (**E**) Average phenotypes in (**B**) compared to the average phenotype of *[-B-]* (*logRatio*). (**F**) Average phenotypes in (**B**) compared to the average phenotype of *[--C]* (*logRatio*). Colormaps in (**C**)-(**F**) indicate an increase or a decrease of activity compared to the reference (denominator).

Table S1. Sequences of *spot*^{196} enhancer variants.

- wild type *[+]* or *[ABC]*

>*spot*$^{196\ [+]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

- single mutants *[0]* to *[16]*

>*spot*$^{196\ [0]}$
AAAAAAAAAAAAAAAAAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAG
ATCTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTA
AAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCG
CCTAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [1]}$
TCTAATTATTCCGTTTAAAAAAAAAAAATTCTGAGCTAAAACTCGCTTATGGAGAGA
TCTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [2]}$
TCTAATTATTCCGTTTAAGGACGCAATTAAAAAAAAAAAAACTCGCTTATGGAGAGA
TCTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [3]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAAAAAAAAATGGAGAGA
TCTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [4]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTAAAAAAAAAA
ATAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [5]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CAAAAAAAAAAAGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [6]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCAAAAAAAAAAAAAATAAATTAATCGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC

TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [7]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAAAAAAAAAAAAGAATTCCCCGCTGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [8]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCAAAAAAAAAAAAAGGCTATTAA
AACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [9]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTAAAAAAAAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*[196 [10]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
AAAAAAAAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [11]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAAAAAAAAAAATCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [12]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGAAAAAAAAAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [13]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAAAAAAAAAAAAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*[196 [14]]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAAAAAAAAAACCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [15]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATAAAAA
AAAAAAATGTGCGCCCATGCAAT

>*spot*$^{196\ [16]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGAAAAAAAAAAAAAAAA

- Permutations of blocks

> *spot*$^{196\ [ACB]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCC
GCCTAATTGATGTGCGCCCATGCAATTTTCCCCGCTTTTGGCTTGAATAAATTAATCG
AATTCCCCGCTGGCTATTAAAA

>*spot*$^{196\ [BAC]}$
TTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAATCTA
ATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*$^{196\ [BCA]}$
TTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACA
CAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTG
ATGTGCGCCCATGCAATTCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAA
CTCGCTTATGGAGAGATCTAAA

>*spot*$^{196\ [CBA]}$
CACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCTA
ATTGATGTGCGCCCATGCAATTTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTC
CCCGCTGGCTATTAAAATCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAA
CTCGCTTATGGAGAGATCTAAA

- Randomized blocks

>*spot*$^{196\ [A--]}$
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATCCGAATTTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACCACATGT
TGTCGACTCAGAAACATTATTCCCATTTACGCGTAAGCAAAAAATGCGTCCTTATCGA
ACTTACACTCGCCTGCGTTGGT

>*spot*$^{196\ [-B-]}$
ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAAACTGTT
GCATGTTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
AGTCGACTCAGAAACATTATTCCCATTTACGCGTAAGCAAAAAATGCGTCCTTATCG
AACTTACACTCGCCTGCGTTGGT

>*spot*^{196} *[--C]*
ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAAACTGTT
GCATGTCCGAATTTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACCACATGT
TCACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[AB-]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
AGTCGACTCAGAAACATTATTCCCATTTACGCGTAAGCAAAAAATGCGTCCTTATCG
AACTTACACTCGCCTGCGTTGGT

>*spot*^{196} *[A-C]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATCCGAATTTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACCACATGT
TCACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[-BC]*
ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAAACTGTT
GCATGTTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCCT
AATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[---]*
ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAAACTGTT
GCATGTCCGAATTTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACCACATGT
TGTCGACTCAGAAACATTATTCCCATTTACGCGTAAGCAAAAAATGCGTCCTTATCGA
ACTTACACTCGCCTGCGTTGGT

genotype	number of individuals
ø	38
[+]	49
[0]	27
[1]	31
[2]	25
[3]	22
[4]	38
[5]	35
[6]	51
[7]	60
[8]	67
[9]	27
[10]	46
[11]	33
[12]	61
[13]	39
[14]	44
[15]	77
[16]	23
WT-[ABC]	61
[-BC]	32
[A-C]	49
[AB-]	24
[A--]	33
[-B-]	35
[--C]	32
[---]	37
[ACB]	39
[BAC]	34
[BCA]	37
[CBA]	34

Table S2. Number of individuals analyzed for each construct in this study.

	regulatory potential (sufficiency)	necessity
[A--]	A is sufficient for vein expression	
[-B-]	B is sufficient for alula expression	
[--C]	C is sufficient for wing blade expression	
[AB-]		C is necessary for high levels in the spot
[A-C]	A is sufficient to repress wing blade expression (outside of spot region)	B is necessary for alula expression B is necessary for full spot levels
[-BC]	B is sufficient to repress wing blade expression (outside of spot region)	A is necessary for full spot levels

Table S3. Analysis of necessity and sufficiency of each block.

Data file S1. Scores for the PCA shown in Figure S1.

Data file S2. **Significance of difference in activity between pairs of groups, using the first 6 principal components.**

Data file S3. Significance of the difference in average expression levels among constructs of the first mutant series (*[0]-[16]*).

Data file S4. Significance of difference in average expression levels among constructs of the second mutant series (blocks).

Additional notes on *logRatios*.

Using average phenotypes to evaluate the effect of the mutations we introduced is useful but limited. Indeed, the differences we observe are visually driven by changes in regions of the wing with elevated enhancer activity. It is then difficult to appreciate whether a mutation affects enhancer activity locally or uniformly across the wing. Differential gene expression is generally represented using log ratios (see reference *(30)* in main text), which measure the fold changes in expression level of a gene relative to a reference (*e.g.*, the expression of the same gene under different conditions). We applied this principle to our image data to visually compare the activity of different constructs across the wing. Classical log ratio translates here to the log of the pixel-wise ratio between two average phenotypes at every pixel (hereafter noted *logRatio*). *logRatio* images of mutants *vs.* wild type are of particular interest to decipher the regulatory logic, because they reveal in which proportion a mutant affects the enhancer activity across the wing. Compared to absolute difference, *logRatio* are not driven by regions with high levels of expression, but by regions with a large fold change, irrespective of the wild-type activity pattern. In a theoretical case where the enhancer activity depends directly and linearly on a given TF concentration, the *logRatio* image reflects logically the spatial distribution of this particular TF. This is also the case if this integration of this TF information is only modulated by uniformly distributed TFs. The underlying logic is straightforward: in this theoretical case, a sequence mutation breaking the interaction between the DNA and the TF will have a significant effect on the phenotype. The intensity of the local phenotypic effect (relatively to the wild-type levels) will depend on the local intensity of the TF-DNA interaction across the wing, and therefore on the

local concentration of the TF. Logically, this interaction is not happening where the TF is absent, with no effect on the phenotype. For any situation departing from these ideal conditions, the resemblance between the *logRatio* and the TF distribution is compromised. For instance, when a TF is locally repressed by another, *logRatio* will correspond to the net loss of spatial information integration, including the loss of this repression. The *logRatio* of a mutant affecting a known TFBS for which the corresponding TF distribution is known therefore informs us on its contribution in the regulatory logic of the enhancer, and how linearly this integration happens. Moreover, even without additional knowledge on the regulatory logic and TF spatial variation, the variety of *logRatio* patterns suggests the action of different spatial inputs integrated by the enhancer.

Discussion

Short Summary

The aim of this dissertation was to explore how multiple enhancers are encoded within the same locus, using the developmental gene *yellow* as a model system. In the *Drosophila biarmipes* genome, the region directly upstream of the *yellow* transcription start site encodes regulatory elements driving multiple activities in the body and wings of a fly. We dissected the region, created a series of transgenic flies containing reporter constructs, and used an image registration pipeline to systematically and quantitatively analyze the distribution of regulatory information. The *yellow 5'* region was shown to be densely packed with overlapping enhancers stretching over large distances. The presented studies challenge the classical notion of enhancers as small, modular elements and highlight how experimental biases can skew our perception of the regulatory landscape in the genome.

In the first chapter, we compared the activities of two enhancers driving transcription in fly wings. We demonstrated sequence overlap between the *wing blade* and *spot* enhancers and identified shared transcription factor binding sites. These pleiotropic sites have been shown to contribute to both wing activities, with one site shown to increase local DNA accessibility. These binding sites indicate a co-option event, where a novel enhancer evolved in the context of preexisting regulatory elements. In this case, the *spot* enhancer likely evolved based on the *wing blade* enhancer, which had sequences regulating local accessibility, thereby creating a favorable environment for the evolution of novel regulatory activities. This study highlights the role of accessibility in the evolution of novel morphological traits.

Complementing this, the second study extended the focus to the *body* enhancer that drives activities in the head, thorax, and abdomen of a fly, and maps to the same locus. While the sequence overlap between the two wing enhancers might have been a result of recent evolution through co-option, the *body* enhancer is much older and should have had enough time to resolve from the wing enhancer sequences. Analysis of the reporter lines demonstrated that elements regulating *yellow* expression in the abdomen are distributed throughout the whole upstream regulatory region and overlap with previously described enhancers. Additionally, using the image registration pipeline and quantitative analysis, we demonstrated that the *body* enhancer consists of at least three independent activities that drive background, dimorphic, and stripe expression patterns in the abdomen. The region, previously thought to consist of discrete and

modular enhancers, was shown instead to have a complex architecture with overlapping regulatory sequences.

The first two chapters of this dissertation explore the distribution of regulatory information over a large region. The third chapter investigates how spatial pattern and activity levels are encoded on a smaller scale within the 196 base pairs of the spot enhancer core. Despite the regulators and their binding sites having been previously identified, this enhancer was found to be much more densely packed with regulatory information. Each of the 16 reporter lines, which collectively mutate every single position in the enhancer, showed significant expression differences compared to the control. This study highlights that the spatial regulation of spot expression results from the interplay of multiple tiers of transcription factors, including at least one activator and three repressors. Interestingly, rearranging these sites affects only the transcription levels of the reporter gene, not the spatial distribution of expression, highlighting the flexibility of the regulatory grammar.

How do we define enhancers?

Mapping enhancer boundaries

The standard image of enhancers as presented in textbooks stems from the assays used to define them. Most often, these are reporter assays in which small DNA fragments are tested for their ability to drive spatio-temporal activity resembling the expression pattern of the endogenous gene. When such a fragment drives reporter expression in transgenic organisms or transfected cells, it is deemed to contain an enhancer. These analyses of regulatory activity are often qualitative, focusing on the presence of an expression pattern without considering whether the tested fragment is sufficient to produce endogenous levels of transcription. As a result, additional DNA sequences necessary to confer transcriptional robustness are often overlooked or dismissed.

Although work on minimal enhancers has been incredibly valuable in uncovering mechanisms of enhancer function, it has also created a biased image of enhancers as short, modular DNA fragments. This image has prompted researchers to focus on the identification and analysis of small fragments, resulting in a circular argument. While some studies have demonstrated that sequences outside of minimal enhancers contribute to expression activity, the field still lacks a nuanced understanding of how regulatory information is encoded on a larger scale.

The mapping of the enhancers in the *yellow* 5' regulatory region showed that the *wing blade* and *spot* enhancers spread over 3.5kb, and the *body* enhancer covers all 5.4 kb of the tested region. Removing fragments within these windows significantly affects the expression of the given enhancers. Such broad distribution is surprising and does not align with the canonical image of enhancers that cover only a few hundred base pairs. Other studies, such as Parker et al., have described long enhancers through computational analysis of the chromatin and transcription profiles of different cell types [89]. This study shows that enhancer length increases with cell specificity and refers to all enhancers longer than 3 kb as stretch enhancers. A wider distribution of regulatory sequences than initially thought bears on our understanding of transcriptional regulation both at the functional and evolutionary levels. Functionally, stretch enhancers are associated with cell type-specific and robust expression and often fall into locus control regions that regulate the expression of important cell type-specific genes. It has been hypothesized that such regulatory hotspots might act as beacons that attract tissue-specific transcription factors. *yellow* is a developmental gene with cell-specific expression, but whether it resides in a locus control region or in the vicinity of other epidermis-specific genes is not known.

Is there a functional or selective need for enhancers to be long? Extended sequences might provide robustness in expression, buffering the effects of individual mutations. Research on minimal enhancers has frequently demonstrated their sensitivity to sequence perturbations. Systematic mutation of the $spot^{196}$ sequence resulted in significant changes in reporter expression. Similar results were observed with minimal enhancers of other developmental genes. Enhancers of *svb*, *eve*, *rho*, and *twi*, measuring 292, 484, 359, and 290 bp respectively, were tested for mutation sensitivity [108], [109]. It was found that these enhancers were strongly affected by most changes in their sequence. For example, 75% of single nucleotide variants of the *ES2* enhancer showed a significant decrease in reporter expression. These genes are crucial for embryonic patterning during *Drosophila* development, and changes in their expression may not be tolerated. Sensitivity of enhancers to mutations would be expected to result in high sequence conservation during evolution; however, this is not what we observe. While the expression patterns of various developmental genes are highly conserved between species, their enhancer sequences often differ significantly. For instance, when *eve* enhancer sequences were compared among six species of *Sepsidae*—scavenger flies that diverged from *Drosophila* over 100 million years ago—very little sequence similarity was found. Nevertheless, they drove the same pattern of gene expression as those of the model species *D. melanogaster* [71]. Similar observations were made when comparing enhancer sequences of

svb among *D. virilis* and *D. ezoana*. Although there was almost no sequence conservation, homologous genomic loci recapitulated similar expression patterns in *Drosophila* larvae [72]. Such findings indicate a high degree of flexibility in enhancers for binding site turnover while maintaining their patterns of activity.

The vulnerability of minimal enhancers to mutations, as tested in reporter constructs, does not align with the idea that regulatory sequences are flexible to binding site turnover—a concept supported by the observation of divergent enhancer sequences among related species. The most likely explanation is the lack of expression robustness provided by minimal enhancers. Flanking sequences often contain additional regulatory elements, and their removal renders minimal enhancers more sensitive to sequence perturbations. It is plausible to speculate that the evolution of large enhancers is driven by the need to stabilize gene expression and buffer against deleterious mutations and might be a common trait of enhancers rather than an exception.

Defining regulatory activities

The studies presented in this dissertation highlight the importance of quantitative analysis. By measuring the fluorescence driven by various fragments, we identified elements that, while not essential for transcription initiation, significantly contribute to enhancer activities and are necessary to reach the full levels of gene expression. Simultaneously, our image registration pipeline allowed us to measure changes not only in expression levels but also patterns. Although the *body* enhancer is typically mapped as a single element, it has been suggested that it consists of multiple activities [107]. PCA identified multiple directions of phenotypic change observed in the abdomen, including general changes in fluorescence levels and changes of expression patterns in the dimorphic region, background expression in the upper four segments, and alterations in the striped pattern. We experimentally tested multiple regions that appeared to contribute to these activities, confirming their ability to drive distinct expression patterns.

Enhancer modularity

Modularity is one of the fundamental features of enhancers, allowing them to discretely modify distinct activities of genes with complex expression patterns. Developmental genes are often active in multiple contexts and are necessary for the formation of multiple traits in different populations of cells and different body parts. The ability to regulate each of these activities separately provides greater flexibility for modifying gene function without disrupting all of its roles. For example, the gene *eyeless* is required for the development of the fly eye, as well as the patterning of different brain regions and the central nervous system. Its activity is regulated

by six distinct enhancers, each approximately 1 kb in length, with each enhancer driving a specific expression pattern: in the eye, different brain lobes at various developmental stages, and the central nervous system [110]. Modifications to any of these enhancers affect only a single aspect of *eyeless* activity without disturbing its other functions. In terms of natural selection, this architecture reduces the likelihood of deleterious effects when spontaneous mutations arise in *eyeless* enhancers in natural fly populations. This modular organization is thought to be a key reason why the evolution of morphological traits occurs primarily through *cis*-regulatory changes rather than *trans*-regulatory modifications. Mutations do arise in *cis* and *trans*, but the mutations in cis are more likely to be retained.

One of the main discoveries of this dissertation is the lack of sequence modularity at the *y* locus (Fig. 3). In the first two papers, we identified and mapped five distinct regulatory activities controlling *y* gene expression. Systematic dissection of the region revealed extensive pleiotropy, with all investigated enhancers extensively overlapping. Such pleiotropy can arise through two possible scenarios: it may result from pleiotropic TFBSs that contribute to the regulation of multiple activities, or from two independent sets of TFBSs that are interwoven (Fig. 4). Distinguishing between these cases is often challenging, as it requires detailed dissection of the region. We observe binding site pleiotropy in the wing blade and spot enhancers, where five TFBSs have been shown to regulate both activities. Mutating these sites restricts *y* expression in the wings. Interestingly, however, these mutations do not affect reporter gene expression in the abdomen (data not shown).

We also observe a complex relationship between abdominal regulatory activities. On one hand, randomizing smaller sections within the context of the entire region demonstrated a certain level of independence. We were able to remove each of the three elements—stripes, dimorphic, and broad—without disrupting the others, suggesting some degree of functional independence. On the other hand, randomizing regulatory information in a 2 kb fragment at the 3′ end, including the stripe element, which is located 1 kb upstream of the core promoter, resulted in the complete loss of abdominal expression. This suggests two possible explanations: either a minimal set of regulatory elements at the 3′ end must be activated simultaneously to enable expression in the abdomen, or

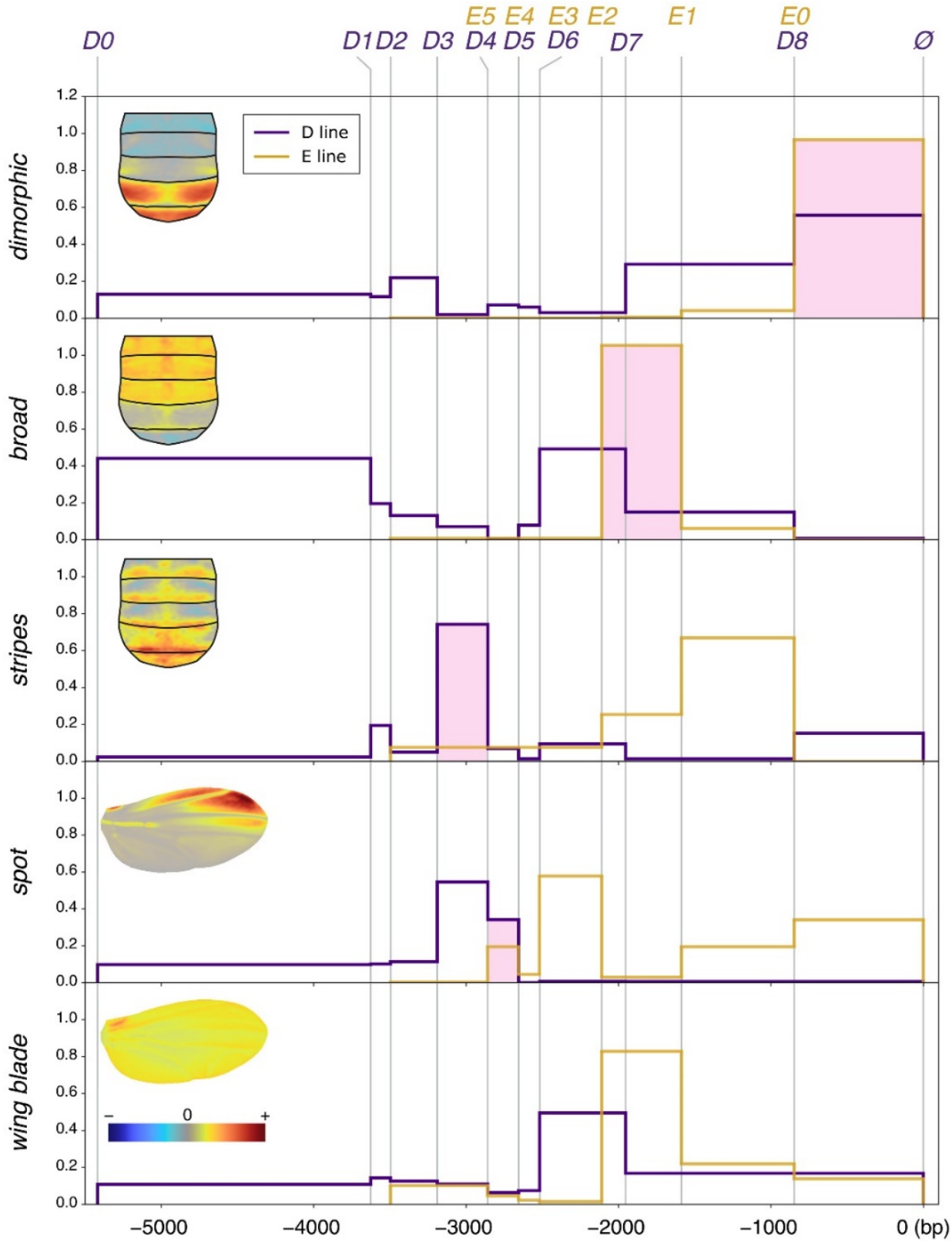
D0
D1
D2
D3
E5
D4
E4
D5
E3
D6
E2
D7
E1
E0
D8
Ø
D line
E line
dimorphic
broad
stripes
spot
wing blade
1.2
1.0
0.8
0.6
0.4
0.2
0.0
–
0
+
−5000
−4000
−3000
−2000
−1000
0 (bp)

Figure 3. Distribution of regulatory information at the *y* locus. Each panel shows the distribution of regulatory information driving the expression of each of the five identified enhancers, which regulate activities in the dimorphic region of the abdomen, background abdominal expression, stripes, wing spot, and background expression in the fly wings. The X-axis indicates the distance to the *yellow* transcription start site, and the Y-axis of each panel shows the relative amount of regulatory information encoded by each of the dissected fragments. The distributions are calculated separately for the D- and E-series fragment deletions. Vertical gray lines indicate the cut sites, which correspond to the outermost boundaries of the transgenic lines (5′ end for the D series, 3′ end for the E series). The labels above the figure show the names of the respective lines. The figure shows that enhancers spread over broad regions and overlap. The regulatory elements of each enhancer are distributed unevenly and contain an enhancer core, a fragment that is necessary for transcription activation and sufficient to drive activity independently of the flanking sequences.

From Mariam Museridze, Stefano Ceolin, Bettina Mühling, Srishti Ramanathan, Olga Barmina, Pallavi Santhi Sekhar, and Nicolas Gompel. 2024. “Entangled and Non-Modular Enhancer Sequences Producing Independent Spatial Activities.” Science Advances 10 (47): eadr9856.

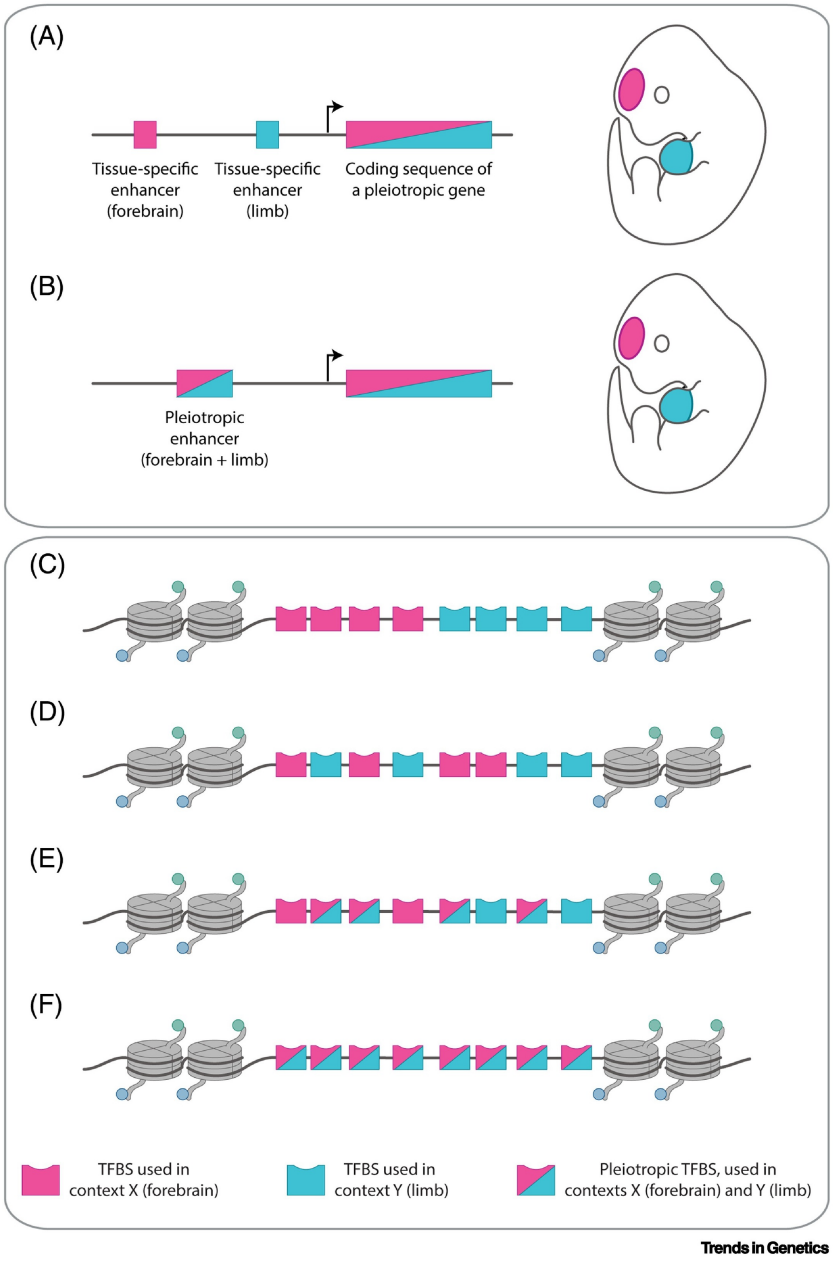

Figure 4. Enhancer Pleiotropy and Site Pleiotropy. (A) Two modular enhancers can drive distinct activities of one gene (in forebrain and limbs). **(B)** A pleiotropic enhancer can activate expression of the gene in both contexts. Distinct sets of transcription factor binding sites can be spatially separated, creating distinct modules **(C)** or be intertwined **(D)**. The two activities of the gene can also be driven by pleiotropic binding sites **(E, F)**. The figure is reused from Sabaris et al. (2019) with permission from the publisher [111].

interactions between different enhancer elements are required for proper regulatory activity. To distinguish between the two possibilities, additional experiments are required. Combined randomization of the dimorphic and broad enhancer cores may reveal their relationship with the stripe elements. Comparing accessibility profiles between the wild type and dissected fragments could show whether the 3′ end plays a role in providing accessibility in the abdominal epithelium. Additionally, chromatin conformation capture assays could highlight potential 3D interactions among *y* enhancers.

Mapping regulatory information density also revealed fragments that appear to strongly contribute to activities in both the wing and abdomen. A stretch of regulatory DNA located between the dissection sites of E3 and E1 (−2518 bp and −1587 bp from the *y* transcription start site) seems to be necessary for reporter expression driven by both the *wing blade* and *broad* enhancers. Interestingly, both enhancers drive background activity in their respective tissues. This suggests that this region contains a single enhancer driving background expression in the epithelia of different body parts. It may be activated by general transcription factors present in epithelial tissues. To better assess the role of this fragment, it will be important to test its sufficiency in driving expression in both tissues and its necessity within the context of the entire regulatory region.

How does the same region drive independent activities?

The projects presented in this thesis describe a complex regulatory region densely packed with overlapping enhancers. Nevertheless, we still observe a certain level of independence in enhancer activities. How is functional independence achieved? There are multiple mechanisms through which the same DNA fragment can be interpreted differently in distinct cell types.

One possible explanation relies on the presence of completely independent sets of transcription factors. Different cell types contain distinct transcription factors that define cell identity and activate genes necessary for cell type-specific functions. Within the same regulatory region, binding sites for entirely separate sets of transcription factors can evolve and function independently. However, existing data indicate that the wing and abdominal epithelium contain largely overlapping transcription factor repertoires [112]. Additionally, many transcription factors do not bind exclusively to a single motif but instead recognize a range of motifs that can be occupied by multiple factors. Therefore, complete independence of regulatory activities is more likely to depend on a few transcription factors that are differentially expressed across tissues and have distinct binding preferences. It may also involve cooperative interactions

between factors that can modulate their binding specificity. Such constraints could limit the number of proteins that contribute to the evolution of novel traits in a given cell type. To determine whether a pleiotropic regulatory region contains independent binding sites, it is essential to identify the transcription factors active in *yellow*-expressing cells across different tissues and pinpoint potential candidates. Additionally, sequence analysis can be used to identify putative binding sites and experimental approaches such as ChIP-seq could further illuminate functional binding sites within the region.

Why would a single locus accumulate such a high number of activities with independent regulatory elements? This scenario might result from a limited number of regions with regulatory potential. Various factors influence the accumulation of regulatory activities. The location of these loci may be constrained by global DNA accessibility in a given tissue and the 3D conformation of the genome, which permits specific enhancer-promoter interactions. As a result, all regulatory elements for a specific gene may evolve within these restricted regions. As long as mutations are not deleterious to the organism, novel binding sites can continue to accumulate, creating pleiotropic loci and spreading across the entire region with regulatory potential. The lack of constraint on the number of enhancer elements for a given gene is also suggested by the abundance of shadow enhancers—redundant enhancers that regulate similar or overlapping expression patterns of the same gene [113]. The ubiquity of these redundant enhancers has been demonstrated in both invertebrates and vertebrates, including mammals [114]–[116].

Another mechanism for separating regulatory activities across tissues is differential chromatin accessibility. It has been shown that accessibility can be modulated locally at the scale of a few hundred base pairs. Performing ATAC-seq would enable the comparison of regulatory landscapes and help identify differences that may contribute to the independent activation of different enhancer elements. It has been demonstrated that differential chromatin accessibility in similar cell types can lead to differential expression of master regulators. When genome-wide open chromatin profiles were compared between *Drosophila* wing and haltere imaginal discs, the same set of enhancers was found to be accessible in both tissues. Strikingly, out of 3,525 peaks, only five were specifically open in halteres. Four of these were located at the *Ubx* locus, a known master regulator of haltere development. These results indicate that, between similar cell types, small differences in chromatin accessibility can determine distinct gene activities [117]. It is possible to speculate that, within the *yellow* regulatory region, local differences in accessibility could result in distinct usage of the same locus across different cell types.

An additional layer of regulation may arise from the three-dimensional organization of DNA. Cooperation among multiple enhancers may be necessary for cell type-specific activation of *yellow*. Interactions among multiple regulatory elements have been shown to contribute to gene regulation. During mammalian development, proper function of the *Fgf8* gene—a key signaling factor—requires the collective activation of multiple enhancers that form an interaction network [118]. Similarly, direct contacts between multiple enhancers regulating *HoxD* genes have been demonstrated in mice. Montavon and colleagues describe a "regulatory archipelago" composed of multiple regulatory elements that has distinct 3D architectures in active versus inactive cells. The collective activation and physical interactions among several enhancer-like sequences were shown to contribute to gene activation. How different enhancers interact with the *y* promoter and with each other remains unknown. It is possible that these interactions are regulated differently across cell types, leading to distinct patterns of gene expression.

Can we generalize these findings?

This dissertation describes a pleiotropic locus densely packed with regulatory information. Can these findings be generalized to other genes, or is this phenomenon specific to the *yellow* locus? *Yellow* is one of the best-studied models for enhancer evolution, and for years, its regulatory region has been depicted as a locus containing multiple discrete modules driving distinct expression patterns. This view was largely shaped by methodological biases that focused on identifying minimal fragments capable of driving specific activities. However, systematic and quantitative dissection has revealed a more complex architecture. The approaches used to study *yellow* are standard in enhancer research and have been applied to the majority of cis-regulatory regions. Therefore, it is plausible to speculate that the widespread depiction of enhancers as modular elements may result more from methodological bias than from their actual structural organization.

An increasing number of studies is identifying pleiotropic enhancers, with their findings summarized in detail [111]. For example, a genome-wide screening for the H3K27ac chromatin mark—a common predictor of active enhancers—identified nearly 90,000 enhancers across different mouse tissues at multiple developmental stages. Remarkably, 52% of these enhancers were active in more than one tissue [119]. Pleiotropic enhancers have also been experimentally identified in various species. A study on *Hox* loci in mice identified an enhancer active in both the mammary bud and limb bud [120]. A different study showed pleiotropic binding site in the *Bmp6* gene enhancer in stickleback fish that regulates expression in both developing fins and tooth epithelia [121]. Even enhancers previously thought to be strictly modular, such as those

regulating *eve* and *svb*, have been shown to exhibit pleiotropic activity. Detailed analysis of *svb* enhancers, for example, revealed that elements driving its expression in embryos are also active in various tissues during larval and pupal development [92]. These findings emphasize that methodological approaches play a crucial role in shaping our perception of regulatory landscapes—what may initially appear modular can often turn out to be pleiotropic when analyzed with different techniques. In order to develop deeper understanding of enhancer function and evolution we should not study and analyze them as small regulatory boxes but understand that they can have uneven, entangled and pleiotropic distribution of regulatory information (Fig. 5).

During genome-wide analyses, genes and their regulatory sequences are often categorized into distinct groups, such as developmental and housekeeping genes, tissue-specific genes, regulators, and effectors. These classifications may be reflected in the structural organization of their regulatory sequences. Some studies suggest that differences in length, location, and binding site motifs distinguish the regulatory sequences of developmental and housekeeping genes [53], [122], [123]. Developmental genes tend to be located farther from their target promoters but are often confined within TADs. Certain binding sites help determine enhancer-promoter specificity between developmental and housekeeping genes. Additionally, genes involved in cell fate determination have been shown to cluster together, forming regulatory hotspots known as super-enhancers. The specific functions of genes may impose constraints that shape regulatory landscape in the genome, leading to distinct structural features. It remains unclear whether *yellow* fits neatly into one of these categories, but it is reasonable to consider it as a representative example of a regulatory locus rather than an exception.

The role of enhancers in evolution

But, if modularity is not a common feature of most enhancers, how do we reconcile this with the observed prevalence of *cis*-regulatory evolution? The answer may lie in the type of information that enhancers encode. Morphological diversity in living organisms arises from tightly orchestrated developmental processes. The shapes and colors of various traits are dictated by the expression patterns of regulatory genes, which determine the developmental fate of the cells in which they are active. Changes in the spatial distribution of regulatory gene expression can directly alter morphological traits. Unlike protein-coding sequences, which do not encode spatial information, enhancers regulate the spatiotemporal aspects of gene expression. By defining the domains in which specific traits will develop, enhancers can drive morphological changes through modifications in gene expression patterns. The role of

enhancers in evolution might rely on the type of information they encode rather than modularity. Nevertheless, the level of independence will still influence the evolutionary processes. Rather than representing distinct categories, modularity and pleiotropy likely exist on a spectrum, with different *cis*-regulatory elements falling at various points along this scale. The degree of modularity influences evolutionary constraints, with pleiotropic enhancers shown to be more conserved than tissue-specific enhancers [96], [124]. Within our model system, we have identified regulatory elements with varying levels of independence, reflecting their evolutionary history.

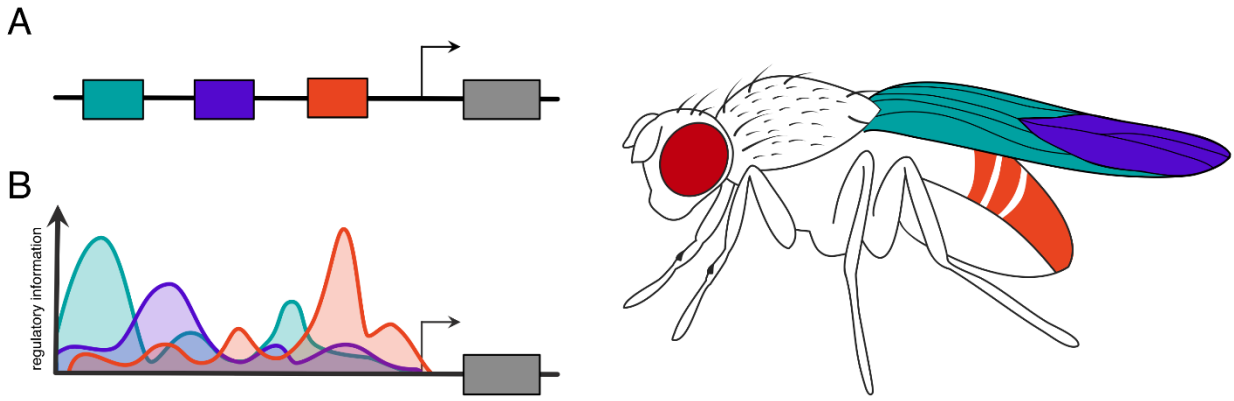

Figure 5. Enhancer structure. (**A**) Textbook representation of enhancers as short modular boxes driving distinct regulatory activities. Studies presented in this dissertation show that the distribution of regulatory information in enhancers can be uneven, entangled, and pleiotropic (**B**).

References

[1] C. Nüsslein-Volhard and E. Wieschaus, “Mutations affecting segment number and polarity in Drosophila,” *Nature*, vol. 287, no. 5785, pp. 795–801, Oct. 1980.

[2] C. Nüsslein-Volhard, H. G. Frohnhöfer, and R. Lehmann, “Determination of anteroposterior polarity in Drosophila,” *Science*, vol. 238, no. 4834, pp. 1675–1681, Dec. 1987.

[3] P. A. Lawrence and G. Morata, “Homeobox genes: their function in Drosophila segmentation and pattern formation,” *Cell*, vol. 78, no. 2, pp. 181–189, Jul. 1994.

[4] R. S. Mann and G. Morata, “The developmental and molecular biology of genes that subdivide the body of Drosophila,” *Annu. Rev. Cell Dev. Biol.*, vol. 16, no. 1, pp. 243–271, 2000.

[5] E. B. Lewis, “Genes and developmental pathways,” *Am. Zool.*, vol. 3, no. 1, pp. 33–56, Feb. 1963.

[6] H. Gloor, “Schädigungsmuster eines Letalfaktors (Kr) von Drosophila melanogaster,” *Arch. Jul. Klaus Stiftung*, vol. 29, pp. 277–287, 1950.

[7] R. Krumlauf, “Hox genes in vertebrate development,” *Cell*, vol. 78, no. 2, pp. 191–201, Jul. 1994.

[8] M. Mark, F. M. Rijli, and P. Chambon, “Homeobox genes in embryogenesis and pathogenesis,” *Pediatr. Res.*, vol. 42, no. 4, pp. 421–429, Oct. 1997.

[9] S. C. Quinonez and J. W. Innis, “Human HOX gene disorders,” *Mol. Genet. Metab.*, vol. 111, no. 1, pp. 4–15, Jan. 2014.

[10] F. Jacob and J. Monod, “Genetic regulatory mechanisms in the synthesis of proteins,” *J. Mol. Biol.*, vol. 3, no. 3, pp. 318–356, Jun. 1961.

[11] D. H. Alpers and G. M. Tomkins, “The order of induction and deinduction of the enzymes of the lactose operon in e. Coli,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 53, no. 4, pp. 797–802, Apr. 1965.

[12] J. Banerji, S. Rusconi, and W. Schaffner, “Expression of a β-globin gene is enhanced by remote SV40 DNA sequences,” *Cell*, vol. 27, no. 2, pp. 299–308, Dec. 1981.

[13] J. Banerji, L. Olson, and W. Schaffner, “A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes,” *Cell*, vol. 33, no. 3, pp. 729–740, Jul. 1983.

[14] W. Schaffner, “Enhancers, enhancers - from their discovery to today’s universe of transcription enhancers,” *Biol. Chem.*, vol. 396, no. 4, pp. 311–327, Apr. 2015.

[15] J. de Villiers, L. Olson, C. Tyndall, and W. Schaffner, “Transcriptional ‘enhancers’ from SV40 and polyoma virus show a cell type preference,” *Nucleic Acids Res.*, vol. 10, no. 24, pp. 7965–7976, 1982.

[16] T. Gerster, P. Matthias, M. Thali, J. Jiricny, and W. Schaffner, “Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer,” *EMBO J.*, vol. 6, no. 5, pp. 1323–1330, May 1987.

[17] D. Wieczorek *et al.*, “A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb,” *Hum. Mutat.*, vol. 31, no. 1, pp. 81–89, Jan. 2010.

[18] L. A. Lettice, P. Devenney, C. De Angelis, and R. E. Hill, “The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression,” *Cell Rep.*, vol. 20, no. 6, pp. 1396–1408, Aug. 2017.

[19] F. Spitz and E. E. M. Furlong, “Transcription factors: from enhancer binding to developmental control,” *Nat. Rev. Genet.*, vol. 13, no. 9, pp. 613–626, Sep. 2012.

[20] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, “A census of human transcription factors: function, expression and evolution,” *Nat. Rev. Genet.*, vol. 10, no. 4, pp. 252–263, Apr. 2009.

[21] P. J. Mitchell and R. Tjian, “Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins,” *Science*, vol. 245, no. 4916, pp. 371–378, Jul. 1989.

[22] S. Small, A. Blair, and M. Levine, “Regulation of even-skipped stripe 2 in the Drosophila embryo,” *EMBO J.*, vol. 11, no. 11, pp. 4047–4057, Nov. 1992.

[23] M. S. Halfon *et al.*, “Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors,” *Cell*, vol. 103, no. 1, pp. 63–74, Sep. 2000.

[24] C.-H. Yuh, A. Ransick, P. Martinez, R. J. Britten, and E. H. Davidson, “Complexity and organization of DNA-protein interactions in the 5′-regulatory region of an endoderm-specific marker gene in the sea urchin embryo,” *Mech. Dev.*, vol. 47, no. 2, pp. 165–186, Aug. 1994.

[25] I. Barozzi, P. Bora, and M. J. Morelli, “Corrigendum: Comparative evaluation of DNase-seq footprint identification strategies,” *Front. Genet.*, vol. 5, Sep. 2014.

[26] J. O. Yáñez-Cuna, H. Q. Dinh, E. Z. Kvon, D. Shlyueva, and A. Stark, "Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding," *Genome Res.*, vol. 22, no. 10, pp. 2018–2030, Oct. 2012.

[27] A. Jolma *et al.*, "DNA-dependent formation of transcription factor pairs alters their binding specificity," *Nature*, vol. 527, no. 7578, pp. 384–388, Nov. 2015.

[28] T. K. Kerppola and T. Curran, "Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity," *Cell*, vol. 66, no. 2, pp. 317–326, Jul. 1991.

[29] A. Kossel, "Ueber einen peptonartigen Bestandtheil des Zellkerns," *Biol. Chem.*, vol. 8, no. 6, pp. 511–515, Jan. 1884.

[30] R. D. Kornberg, "Chromatin structure: a repeating unit of histones and DNA," *Science*, vol. 184, no. 4139, pp. 868–871, May 1974.

[31] N. Kaplan *et al.*, "The DNA-encoded nucleosome organization of a eukaryotic genome," *Nature*, vol. 458, no. 7236, pp. 362–366, Mar. 2009.

[32] C.-K. Lee, Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb, "Evidence for nucleosome depletion at active regulatory regions genome-wide," *Nat. Genet.*, vol. 36, no. 8, pp. 900–905, Aug. 2004.

[33] R. E. Thurman *et al.*, "The accessible chromatin landscape of the human genome," *Nature*, vol. 489, no. 7414, pp. 75–82, Sep. 2012.

[34] C. D. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control," *Nat. Rev. Genet.*, vol. 17, no. 8, pp. 487–500, Aug. 2016.

[35] M. J. Guertin and J. T. Lis, "Mechanisms by which transcription factors gain access to target sequence elements in chromatin," *Curr. Opin. Genet. Dev.*, vol. 23, no. 2, pp. 116–123, Apr. 2013.

[36] J. Svaren, E. Klebanow, L. Sealy, and R. Chalkley, "Analysis of the competition between nucleosome formation and transcription factor binding," *J. Biol. Chem.*, vol. 269, no. 12, pp. 9335–9344, Mar. 1994.

[37] E. E. Swinstead, V. Paakinaho, D. M. Presman, and G. L. Hager, "Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective," *Bioessays*, vol. 38, no. 11, pp. 1150–1157, Nov. 2016.

[38] G. Felsenfeld, J. Boyes, J. Chung, D. Clark, and V. Studitsky, "Chromatin structure and gene expression," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 18, pp. 9384–9388, Sep. 1996.

[39] S. John *et al.*, “Chromatin accessibility pre-determines glucocorticoid receptor binding patterns,” *Nat. Genet.*, vol. 43, no. 3, pp. 264–268, Mar. 2011.

[40] B. Di Stefano *et al.*, “C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4,” *Nat. Cell Biol.*, vol. 18, no. 4, pp. 371–381, Apr. 2016.

[41] L. Grøntved *et al.*, “C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements,” *EMBO J.*, vol. 32, no. 11, pp. 1568–1583, May 2013.

[42] J. Ungerbäck *et al.*, “Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by PU.1,” *bioRxiv*, bioRxiv, 25-May-2018.

[43] J. Gassler *et al.*, “Zygotic genome activation by the totipotency pioneer factor Nr5a2,” *Science*, vol. 378, no. 6626, pp. 1305–1315, Dec. 2022.

[44] H.-L. Liang, C.-Y. Nien, H.-Y. Liu, M. M. Metzstein, N. Kirov, and C. Rushlow, “The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila,” *Nature*, vol. 456, no. 7220, pp. 400–403, Nov. 2008.

[45] R. Guan, T. Lian, B.-R. Zhou, D. Wheeler, and Y. Bai, “Structural mechanism of LIN28B nucleosome targeting by OCT4,” *Mol. Cell*, vol. 83, no. 12, pp. 1970-1982.e6, Jun. 2023.

[46] L. Vanzan *et al.*, “High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites,” *Nat. Commun.*, vol. 12, no. 1, p. 3337, Jun. 2021.

[47] J. Crocker and D. L. Stern, “Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer,” *Development*, vol. 144, no. 17, pp. 3095–3101, Sep. 2017.

[48] B. Prud’homme *et al.*, “Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene,” *Nature*, vol. 440, no. 7087, pp. 1050–1053, Apr. 2006.

[49] L. Ling, B. Mühling, R. Jaenichen, and N. Gompel, “Increased chromatin accessibility promotes the evolution of a transcriptional silencer in Drosophila,” *Sci. Adv.*, vol. 9, no. 7, p. eade6529, Feb. 2023.

[50] J. H. Yang and A. S. Hansen, “Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation,” *Nat. Rev. Mol. Cell Biol.*, vol. 25, no. 7, pp. 574–591, Jul. 2024.

[51] K. Hamamoto and T. Fukaya, “Molecular architecture of enhancer-promoter interaction,” *Curr. Opin. Cell Biol.*, vol. 74, pp. 62–70, Feb. 2022.

[52] S. Schoenfelder and P. Fraser, “Long-range enhancer-promoter contacts in gene expression control,” *Nat. Rev. Genet.*, vol. 20, no. 8, pp. 437–455, Aug. 2019.

[53] M. A. Zabidi *et al.*, “Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation,” *Nature*, vol. 518, no. 7540, pp. 556–559, Feb. 2015.

[54] C. Hou, R. Dale, and A. Dean, “Cell type specificity of chromatin organization mediated by CTCF and cohesin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 8, pp. 3651–3656, Feb. 2010.

[55] T. Sexton *et al.*, “Three-dimensional folding and functional organization principles of the Drosophila genome,” *Cell*, vol. 148, no. 3, pp. 458–472, Feb. 2012.

[56] K. C. Keough *et al.*, “Three-dimensional genome rewiring in loci with human accelerated regions,” *Science*, vol. 380, no. 6643, p. eabm1696, Apr. 2023.

[57] M. C. King and A. C. Wilson, “Evolution at two levels in humans and chimpanzees: Their macromolecules are so alike that regulatory mutations may account for their biological differences,” *Science*, vol. 188, no. 4184, pp. 107–116, Apr. 1975.

[58] R. Quiring, U. Walldorf, U. Kloter, and W. J. Gehring, “Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans,” *Science*, vol. 265, no. 5173, pp. 785–789, Aug. 1994.

[59] G. Halder, P. Callaerts, and W. J. Gehring, “Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila,” *Science*, vol. 267, no. 5205, pp. 1788–1792, Mar. 1995.

[60] E. C. Bolton, A. Y. So, C. Chaivorapol, C. M. Haqq, H. Li, and K. R. Yamamoto, “Cell- and gene-specific regulation of primary target genes by the androgen receptor,” *Genes Dev.*, vol. 21, no. 16, pp. 2005–2017, Aug. 2007.

[61] T. Sandmann, C. Girardot, M. Brehme, W. Tongprasit, V. Stolc, and E. E. M. Furlong, “A core transcriptional network for early mesoderm development in Drosophila melanogaster,” *Genes Dev.*, vol. 21, no. 4, pp. 436–449, Feb. 2007.

[62] S. B. Carroll, “Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution,” *Cell*, vol. 134, no. 1, pp. 25–36, Jul. 2008.

[63] R. P. Sharma and V. L. Chopra, “Effect of the Wingless (wg1) mutation on wing and haltere development in Drosophila melanogaster,” *Dev. Biol.*, vol. 48, no. 2, pp. 461–465, Feb. 1976.

[64] M. Lynch and B. Walsh, *The origins of genome architecture*. New York, NY: Oxford University Press, 2007.

[65] R. H. Waterston, E. S. Lander, and J. E. Sulston, “On the sequencing of the human genome,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 6, pp. 3712–3716, Mar. 2002.

[66] S. A. Shabalina, A. Y. Ogurtsov, V. A. Kondrashov, and A. S. Kondrashov, “Selective constraint in intergenic regions of human and mouse genomes,” *Trends Genet.*, vol. 17, no. 7, pp. 373–376, Jul. 2001.

[67] R. Jaura *et al.*, “Extended intergenic DNA contributes to neuron-specific expression of neighboring genes in the mammalian nervous system,” *Nat. Commun.*, vol. 13, no. 1, p. 2733, May 2022.

[68] C. E. Nelson, B. M. Hersh, and S. B. Carroll, “The regulatory content of intergenic DNA shapes genome architecture,” *Genome Biol.*, vol. 5, no. 4, p. R25, Mar. 2004.

[69] J. C. Klein *et al.*, “A systematic evaluation of the design and context dependencies of massively parallel reporter assays,” *Nat. Methods*, vol. 17, no. 11, pp. 1083–1091, Nov. 2020.

[70] J. Cande, Y. Goltsev, and M. S. Levine, “Conservation of enhancer location in divergent insects,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 34, pp. 14414–14419, Aug. 2009.

[71] E. E. Hare, B. K. Peterson, and M. B. Eisen, “A careful look at binding site reorganization in the even-skipped enhancers of Drosophila and sepsids,” *PLoS genetics*, vol. 4, no. 11. journals.plos.org, p. e1000268, Nov-2008.

[72] D. L. Stern and N. Frankel, “The structure and evolution of cis-regulatory regions: the shavenbaby story,” *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 368, no. 1632, p. 20130028, Dec. 2013.

[73] G. Kalay and P. J. Wittkopp, “Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species,” *PLoS Genet.*, vol. 6, no. 11, p. e1001222, Nov. 2010.

[74] O. Symmons *et al.*, “Functional and topological characteristics of mammalian regulatory domains,” *Genome Res.*, vol. 24, no. 3, pp. 390–400, Mar. 2014.

[75] D. Li *et al.*, “Comparative 3D genome architecture in vertebrates,” *BMC Biol.*, vol. 20, no. 1, p. 99, May 2022.

[76] S. Fanucchi, Y. Shibayama, S. Burd, M. S. Weinberg, and M. M. Mhlanga, “Chromosomal contact permits transcription between coregulated genes,” *Cell*, vol. 155, no. 3, pp. 606–620, Oct. 2013.

[77] W. A. Whyte *et al.*, “Master transcription factors and mediator establish super-enhancers at key cell identity genes,” *Cell*, vol. 153, no. 2, pp. 307–319, Apr. 2013.

[78] J. Lovén *et al.*, “Selective inhibition of tumor oncogenes by disruption of super-enhancers,” *Cell*, vol. 153, no. 2, pp. 320–334, Apr. 2013.

[79] A. Khan, A. Mathelier, and X. Zhang, “Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers,” *Epigenetics*, vol. 13, no. 9, pp. 910–922, Oct. 2018.

[80] S. Pott and J. D. Lieb, “What are super-enhancers?,” *Nat. Genet.*, vol. 47, no. 1, pp. 8–12, Jan. 2015.

[81] N. Gompel and B. Prud’homme, “The causes of repeated genetic evolution,” *Dev. Biol.*, vol. 332, no. 1, pp. 36–47, Aug. 2009.

[82] S. Alamos *et al.*, “Minimal synthetic enhancers reveal control of the probability of transcriptional engagement and its timing by a morphogen gradient,” *Cell Syst.*, vol. 14, no. 3, pp. 220-236.e3, Mar. 2023.

[83] Y. Yan, C. A. Jones, C. D. Sigmund, K. W. Gross, and D. F. Catanzaro, “Conserved enhancer elements in human and mouse renin genes have different transcriptional effects in As4.1 cells,” *Circ. Res.*, vol. 81, no. 4, pp. 558–566, Oct. 1997.

[84] K. A. Barr and J. Reinitz, “A sequence level model of an intact locus predicts the location and function of nonadditive enhancers,” *PLoS One*, vol. 12, no. 7, p. e0180861, Jul. 2017.

[85] M. Z. Ludwig, Manu, R. Kittler, K. P. White, and M. Kreitman, “Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness,” *PLoS Genet.*, vol. 7, no. 11, p. e1002364, Nov. 2011.

[86] B. J. Aronow *et al.*, “Dissecting a locus control region: Facilitation of enhancer function by extended enhancer-flanking sequences,” *Mol. Cell. Biol.*, vol. 15, no. 2, pp. 1123–1135, Feb. 1995.

[87] J. W. Blayney *et al.*, “Super-enhancers include classical enhancers and facilitators to fully activate gene expression,” *Cell*, vol. 186, no. 26, pp. 5826-5839.e18, Dec. 2023.

[88] S. L. Fong and J. A. Capra, “Function and constraint in enhancer sequences with multiple evolutionary origins,” *Genome Biol. Evol.*, vol. 14, no. 11, Nov. 2022.

[89] S. C. J. Parker *et al.*, “Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 44, pp. 17921–17926, Oct. 2013.

[90] S. Small, D. N. Arnosti, and M. Levine, “Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter,” *Development*, vol. 119, no. 3, pp. 762–772, Nov. 1993.

[91] O. Nagy *et al.*, “Correlated evolution of two copulatory organs via a single cis-regulatory nucleotide change,” *Curr. Biol.*, vol. 28, no. 21, pp. 3450-3457.e13, Nov. 2018.

[92] E. Preger-Ben Noon *et al.*, “Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function,” *Cell Rep.*, vol. 22, no. 11, pp. 3021–3031, Mar. 2018.

[93] S. Small, A. Blair, and M. Levine, “Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo,” *Dev. Biol.*, vol. 175, no. 2, pp. 314–324, May 1996.

[94] M. Fujioka, Y. Emi-Sarker, G. L. Yusibova, T. Goto, and J. B. Jaynes, “Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients,” *Development*, vol. 126, no. 11, pp. 2527–2538, Jun. 1999.

[95] C. R. Infante *et al.*, “Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes,” *Dev. Cell*, vol. 35, no. 1, pp. 107–119, Oct. 2015.

[96] A. Fish, L. Chen, and J. A. Capra, “Gene regulatory enhancers with evolutionarily conserved activity are more pleiotropic than those with species-specific activity,” *Genome Biol. Evol.*, vol. 9, no. 10, pp. 2615–2625, Oct. 2017.

[97] D. Thanos and T. Maniatis, “Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome,” *Cell*, vol. 83, no. 7, pp. 1091–1100, Dec. 1995.

[98] B. P. Song *et al.*, “Diverse logics and grammar encode notochord enhancers,” *Cell Rep.*, vol. 42, no. 2, p. 112052, Feb. 2023.

[99] M. M. Kulkarni and D. N. Arnosti, “Information display by transcriptional enhancers,” *Development*, vol. 130, no. 26, pp. 6569–6575, Dec. 2003.

[100] G. Junion *et al.*, “A transcription factor collective defines cardiac cell fate and reflects lineage history,” *Cell*, vol. 148, no. 3, pp. 473–486, Feb. 2012.

[101] Q. Han, J. Fang, H. Ding, J. K. Johnson, B. M. Christensen, and J. Li, “Identification of Drosophila melanogaster yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes,” *Biochem. J.*, vol. 368, no. Pt 1, pp. 333–340, Nov. 2002.

[102] P. J. Wittkopp, J. R. True, and S. B. Carroll, “Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns,” *Development*, vol. 129, no. 8, pp. 1849–1858, Apr. 2002.

[103] H. Barek, H. Zhao, K. Heath, A. Veraksa, and M. Sugumaran, “Drosophila yellow-h encodes dopaminechrome tautomerase: A new enzyme in the eumelanin biosynthetic pathway,” *Pigment Cell Melanoma Res.*, vol. 35, no. 1, pp. 26–37, Jan. 2022.

[104] M. D. Drapeau, “A novel hypothesis on the biochemical role of the Drosophila Yellow protein,” *Biochem. Biophys. Res. Commun.*, vol. 311, no. 1, pp. 1–3, Nov. 2003.

[105] P. K. Geyer, C. Spana, and V. G. Corces, “On the molecular mechanism of gypsy-induced mutations at the yellow locus of Drosophila melanogaster,” *EMBO J.*, vol. 5, no. 10, pp. 2657–2662, Oct. 1986.

[106] H. Hinaux *et al.*, “Revisiting the developmental and cellular role of the pigmentation gene yellow in Drosophila using a tagged allele,” *Dev. Biol.*, vol. 438, no. 2, pp. 111–123, Jun. 2018.

[107] G. Kalay, J. Lachowiec, U. Rosas, M. R. Dome, and P. Wittkopp, “Redundant and Cryptic Enhancer Activities of the Drosophila yellow Gene,” *Genetics*, vol. 212, no. 1, pp. 343–360, May 2019.

[108] T. Fuqua *et al.*, “Dense and pleiotropic regulatory information in a developmental enhancer,” *Nature*, vol. 587, no. 7833, pp. 235–239, Nov. 2020.

[109] R. Galupa *et al.*, “Enhancer architecture and chromatin accessibility constrain phenotypic space during Drosophila development,” *Dev. Cell*, vol. 58, no. 1, pp. 51-62.e4, Jan. 2023.

[110] Y. Adachi *et al.*, “Conserved cis-regulatory modules mediate complex neural expression patterns of the eyeless gene in the Drosophila brain,” *Mech. Dev.*, vol. 120, no. 10, pp. 1113–1126, Oct. 2003.

[111] G. Sabarís, I. Laiker, E. Preger-Ben Noon, and N. Frankel, “Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity,” *Trends Genet.*, vol. 35, no. 6, pp. 423–433, Jun. 2019.

[112] M. Museridze *et al.*, “Entangled and non-modular enhancer sequences producing independent spatial activities,” *Sci. Adv.*, vol. 10, no. 47, p. eadr9856, Nov. 2024.

[113] E. Cannavò *et al.*, “Shadow enhancers are pervasive features of developmental regulatory networks,” *Curr. Biol.*, vol. 26, no. 1, pp. 38–51, Jan. 2016.

[114] J.-W. Hong, D. A. Hendrix, and M. S. Levine, “Shadow enhancers as a source of evolutionary novelty,” *Science*, vol. 321, no. 5894, p. 1314, Sep. 2008.

[115] O. Hobert, “Gene regulation: enhancers stepping out of the shadow,” *Curr. Biol.*, vol. 20, no. 17, pp. R697-9, Sep. 2010.

[116] B. Antosova *et al.*, “The gene regulatory network of lens induction is wired through Meis-dependent shadow enhancers of Pax6,” *PLoS Genet.*, vol. 12, no. 12, p. e1006441, Dec. 2016.

[117] D. J. McKay and J. D. Lieb, “A common set of DNA regulatory elements shapes Drosophila appendages,” *Dev. Cell*, vol. 27, no. 3, pp. 306–318, Nov. 2013.

[118] M. Marinić, T. Aktas, S. Ruf, and F. Spitz, “An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape,” *Dev. Cell*, vol. 24, no. 5, pp. 530–542, Mar. 2013.

[119] A. S. Nord *et al.*, “Rapid and pervasive changes in genome-wide enhancer usage during mammalian development,” *Cell*, vol. 155, no. 7, pp. 1521–1531, Dec. 2013.

[120] R. Schep *et al.*, “Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 48, pp. E7720–E7729, Nov. 2016.

[121] P. A. Erickson, P. A. Cleves, N. A. Ellis, K. T. Schwalbach, J. C. Hart, and C. T. Miller, “A 190 base pair, TGF-β responsive tooth and fin enhancer is required for stickleback Bmp6 expression,” *Dev. Biol.*, vol. 401, no. 2, pp. 310–323, May 2015.

[122] I. Abnizova, C. Stapel, R. T. Boekhorst, J. T. H. Lee, and M. Hemberg, “Integrative analysis of transcriptomic and epigenomic data reveals distinct patterns for developmental and housekeeping gene regulation,” *BMC Biol.*, vol. 22, no. 1, p. 78, Apr. 2024.

[123] I. Zhimulev, T. Vatolina, V. Levitsky, and A. Tsukanov, “Developmental and housekeeping genes: Two types of genetic organization in the Drosophila genome,” *Int. J. Mol. Sci.*, vol. 25, no. 7, Apr. 2024.

[124] R. Andersson *et al.*, “An atlas of active enhancers across human cell types and tissues,” *Nature*, vol. 507, no. 7493, pp. 455–461, Mar. 2014.

Acknowledgments

My PhD journey, like many others, has had its share of ups and downs, but I've always been incredibly lucky to be surrounded by amazing people, whom I would like to thank.

I would like to thank my supervisor, Nicolas, for his guidance and encouragement during the years I spent in his group. His door was always open—for both scientific discussions and regular chats. Nicolas was just as supportive when projects worked well as he was during the first two years when everything was failing. I learned a lot from him and couldn't have asked for a better mentor.

I would like to thank Yaqun for introducing me to the lab, and Stefano, who was a central part of the project. I deeply appreciate not only their expertise and mentorship, but also their positivity and friendliness, which made working on these projects an incredibly enjoyable experience. Thanks to Rita and Bettina for being magicians in the lab and always figuring out how to solve problems. I'm grateful to all the present and past members of the Gompel group—not only are they inspiring scientists, but also amazing chefs, musicians, artists, climbers, and, most importantly, great friends. I want to thank Liucong, Yunhan, Xinyi, Conrad, Vijayaditya, Natascha, Bilal, Gilberto, Alex, Saskia, Bärbel, Emelie, Lara, Deepak, Yann, and many others for making every working day enjoyable.

I'd like to thank my thesis committee members, Tamuna and Axel, for being so encouraging during my PhD and for giving great advice about both projects and life in academia.

I'm grateful to the graduate school QMB, which provided scientific and financial support. I'm very thankful for the amazing support from its coordinators Markus, Christophe and Beate. The QMB retreats and summit clubs have been some of my favorite experiences during my time in Germany.

I've been surrounded by so many amazing people at the Biozentrum who enriched my life. I'm grateful for all the breakfasts in the department kitchen, coffees at Chicco, Russian lunches at Mensa, and happy hours in the garden.

Most importantly, the Biozentrum is where I met Tom—an incredible scientist, friend, and partner. The past few years would have been impossible without his care, encouragement, and inspiration. I'm so thankful to him for helping me with science, encouraging me to reach for more, taking me on adventures, and making me eat fruit.

When I moved to Germany, I hoped to find friends, but I ended up finding a family. I'm so lucky to have met Flo, Hanna, Jelena, Max, and Morgan. I want to thank them for always hyping me up, taking care of me, helping me write every single email and being the most amazing friends anyone could ask for.

Being in a foreign country can be hard, but I've been lucky to have friends who always made me feel like I was still home. I want to thank Lika, Tata, Annie, Revi, Kvantri, Mindia, Anna and Diana for inspiring and supporting me, even from afar.

Finally, I want to thank my family—especially my mom and dad, Lena and Temo—for always being there for me. They showed so much support and pride in everything I did, even if I failed to explain why I'd been taking fly butt pictures for five years. They always made me feel like the smartest and most successful person in the world, and encouraged me to keep going.