Divergent Chromatin Accessibility Dynamics Promote Differentiated Enhancer Functions and Morphological Trait Evolution

Dissertation

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

Liucong Ling

München 2024

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

Zweitgutachter: Prof. Dr. Peter Becker

Tag der Einreichung: den 13.03.2024

Tag der mündlichen Prüfung: 21.06.2024

Declaration

Eidesstattliche Erklärung

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

München, den .....27.06.2024..... .....

Liucong Ling

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.

München, den....27.06.2024. .....

Liucong Ling

List of Publications

1. **Ling, L.**, Mühling, B., Jaenichen, R., & Gompel, N. (2023). Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*. *Science Advances*, *9*(7).

2. Xin, Y., Poul, Y. Le, **Ling, L.**, Museridze, M., Mühling, B., Jaenichen, R., … Gompel, N. (2020). Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. *Proceedings of the National Academy of Sciences*, *117*(34), 20636–20644.

3. Poul, Y. Le, Xin, Y., **Ling, L.**, Mühling, B., Jaenichen, R., Hörl, D., … Gompel, N. (2020). Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Science Advances*, *6*(49), eabe2955.

Declaration of Contributions

1. **Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*.**

   Liucong Ling, Bettina Mühling, Rita Jaenichen, Nicolas Gompel.

   SCIENCE ADVANCES **|** Vol 9, Issue 7 | 7 Feb 2023

   DOI: 10.1126/sciadv.ade6529

   Contributions: Liucong Ling conceptualized and designed the project and experiments with Prof. Dr. Nicolas Gompel. Liucong Ling performed most experiments, including all RNA-related experiments, ATAC-seq, CUT&RUN, greenCUT&RUN, fly crosses and RNAi, and imaging experiments. Liucong Ling also performed all data analysis, including qPCR analysis, genomic sequencing processing and analysis, transcription factor-DNA binding motif analysis, and imaging processing and analysis.

   Other authors' contributions:

   Nicolas Gompel: Conceptualized the project and interpreted the data with Liucong Ling.; wrote the manuscript, which Liucong Ling edited; supervised the project and acquired funding.

   Bettina Mühling: Produced all transgenic lines.

   Rita Jaenichen: Performed all other molecular biology experiments, including molecular cloning and genotyping.

   Supervisor's signature:

   Nicolas Gompel

2. 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: Liucong Ling designed the experiments to measure chromatin accessibility with Yaqun Xin and Prof. Dr. Nicolas Gompel. Liucong Ling set up the ATAC-seq experiment and performed ATAC-seq with help from Yaqun Xin. Liucong Ling processed and analyzed the ATAC-seq data.

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 Prof. Dr. Nicolas Gompel.

Mariam Museridze and 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

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: Liucong Ling performed some of the cloning and imaging experiments. Liucong Ling contributed to the interpretation of the data.

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.

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.

Mariam Museridze: investigation and formal analysis.

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—original draft, and funding acquisition.

Nicolas Gompel: conceptualization, validation, writing—original draft, visualization, supervision, project administration, and funding acquisition.

Supervisor's signature:

Nicolas Gompel

Table of Contents

List of Abbreviations

4C	circular chromosome conformation capture
A-P	anterior-posterior
Abd-B	Abdominal-B
AML	acute myeloid leukemia
APF	after puparium formation
ATAC-seq	the assay for transposase-accessible chromatin using sequencing
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
BAF	BRG1/BRM-associated factor
BAP	Brahma Associated Protein
bp	base pair
BRD4	bromodomain-containing protein 4
c-Myc	cellular Myc
C/EBP	CCAAT/enhancer binding proteins
CBP	CREB(cAMP response element-binding protein)-binding protein
CD2	cluster of differentiation 2
CD4	cluster of differentiation 4
CHD	Chromodomain-Helicase DNA-binding
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation combined with microarray chip technology
ChIP-seq	chromatin immunoprecipitation with sequencing
CNN	Convolutional Neural Network
COMPASS	Complex Proteins Associated with Set1
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference
CtBP	C-terminal-binding protein 1
CUT&RUN	cleavage under targets and release using nuclease
CUT&Tag	cleavage under targets and tagmentation
DFREs	dual-function regulatory elements
Dll	Distal-less
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease

DNase-seq	DNase I hypersensitive sites sequencing
dSfmbt	Drosophila Scm-related gene containing four mbt domains
DV	dorsoventral
E93	Ecdysone-induced protein 93F; Eip93F
EcR	ecdysone receptor
En	Engrailed
ENCODE	the Encyclopedia of DNA Elements
FAIRE-seq	Formaldehyde-Assisted Isolation of Regulatory Elements followed by sequencing
FGFR2	fibroblast growth factor receptor 2
FoxA	forkhead box A
FOXA2	forkhead box A2
GDF6	growth differentiation factor 6
GFP	green fluorescent protein
gRNA	guide RNA
GWAS	genome-wide association study
H1	histone H1
H2A	histone H2A
H2B	histone H2B
H3	histone H3
H3K27ac	histone H3 lysine 27 acetylation
H3K27me3	histone H3 lysine 27 trimethylation
H3K4	histone H3 lysine 4
H3K4me1	histone H3 lysine 4 monomethylation
H3K4me3	histone H3 lysine 4 trimethylation
H4	histone H4
HARs	Human Accelerated Regions
HAT	histone acetyltransferases
Hb	Hunchback
HDAC	histone deacetylases
hESCs	human embryonic stem cells
HNF3	hepatocyte nuclear factor 3
IgH	immunoglobulin heavy chain
iPS	induced pluripotent stem cells
ISWI	Imitation SWItch

Kr	Krüppel
LCR	locus control region
LDL	low-density lipoprotein
LDL-C	low-density lipoprotein cholesterol
Mb	megabase pair
MNase	micrococcal nuclease
MNase-seq	micrococcal nuclease digestion with deep sequencing
modENCODE	the model organism Encyclopedia of DNA Elements
MPRA	Massively Parallel Reporter Assays
mRNA	messenger ribonucleic acid
NGS	next-generation sequencing
NPCs	Neural Progenitor Cells
Nr5a2	nuclear receptor subfamily 5 group A member 2
NURF	nucleosome remodeling factor
Oct4	octamer-binding transcription factor 4
P300	E1A(adenovirus early region 1A)-associated protein p300
PBAF	polybromo-associated BAF
PBAP	Polybromo-associated BAP
PcG	Polycomb group complexes
PhoRC	Pho-repressive complex
Pitx1	Paired-like homeodomain 1
Pol II	RNA polymerase II
Pou5f1	POU domain, class 5, transcription factor 1
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PREs	Polycomb response elements
qPCR	real-time polymerase chain reaction
QTL	quantitative trait locus
REST	RE1-Silencing Transcription factor
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNAi	ribonucleic acid interference
RUNX 1/3	runt-related transcription factor 1/3
SELEX-seq	Systematic Evolution of Ligands by Exponential Enrichment with massively parallel sequencing

Shh	sonic hedgehog protein
SNPs	single nucleotide polymorphisms
SORT1	sortilin 1
Sox2	SRY(sex determining region Y)-box transcription factor 2
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAD	topologically associating domain
UAS	upstream activating sequence
Ubx	Ultrabithorax

Summary

Gene *cis*-regulatory regions regulate specific temporal and spatial gene transcription. Sequence variations in the *cis*-regulatory areas alter gene transcription and are, therefore, pivotal for the evolution of morphological traits, as well as the development of diseases. Understanding how the regulatory information is encoded in *cis*-regulatory sequences is essential for deepening our understanding of gene transcriptional regulation. In the group of species akin to the model organism *Drosophila melanogaster*, many have varied wing pigmentation patterns in males. Previous studies have shown that the expression pattern of a pigmentation-related gene, *yellow*, in pupal wings prefigured the adult pigmentation pattern. A specific enhancer of *yellow* regulates its wing expression and underlies the dark pigmentation pattern. The sequence of this enhancer varies among species, explaining changes in *yellow* expression. Several species, including *Drosophila melanogaster*, lost the wing pigmentation pattern secondarily during evolution. By which mechanisms this trait was lost remains unclear. In the main project of the present dissertation, by comparing chromatin accessibility throughout pupal wing development between *D. melanogaster* and a closely related species with wing pigmentation, I uncovered a region of increased chromatin accessibility in *D. melanogaster* corresponding to a newly evolved silencer. This silencer specifically represses the nearby enhancer regulating the pigmentation pattern. I also found that the transcription factor Eip93F is involved in regulating the increased accessibility of this silencer. Furthermore, enhancer-reporter assays suggested a gain of a repressor site in the region with increased chromatin accessibility.

In summary, this project proposes a novel model explaining the evolutionary loss of a trait: the sequence variations gained during evolution resulted in the increased chromatin accessibility of the *cis*-regulatory region, leading to the recruitment of repressors silencing a nearby enhancer. Moreover, by studying the molecular mechanism of silencing an evolutionary gained enhancer, this dissertation offers new insights into the interplay between enhancers and silencers.

Additionally, as part of the present dissertation, I collaborated with my colleagues and contributed to two other projects that centered on the *yellow* enhancer regulating the wing spot pigmentation pattern. The first project examined the boundary between this recently evolved *yellow* enhancer and its ancestral counterpart, as well as how they regulate one another. The second project involved analyzing the regulatory syntax of this enhancer to understand how it regulates the expression of *yellow* in specific spatial patterns.

Zusammenfassung

Die *cis*-regulatorischen Bereiche von Genen kontrollieren die zeitliche und räumliche Gentranskription. Sequenzvariationen in *cis*-regulatorischen Bereichen können somit Gentranskription verändern und sind dementsprechend von zentraler Bedeutung für die Evolution und Diversifizierung morphologischer Merkmale sowie für das Auftreten von Krankheiten. Die Prinzipien aufzudecken wie regulatorische Informationen in den *cis*-regulatorischen Sequenzen kodiert sind ist für ein tieferes Verständnis der Gen-Transkriptionsregulation von wesentlicher Bedeutung. Bei nahverwandten Arten vom Modellorganismus *Drosophila melanogaster* treten bei Männchen diverse Flügelpigmentierungsmuster auf. Frühere Studien belegen, dass das adulte Pigmentierungsmuster mit dem pupalen Expressionsmuster des Pigmentierungsgens, *yellow*, übereinstimmt. Ferner wurde gezeigt, dass ein spezifischer *yellow* Enhancer die Genexpression im Flügel reguliert und das dunklen Pigmentierungsmuster bestimmt. Die Sequenz dieses Enhancers ist artspezifisch unterschiedlich, was die Diversität des Pigmentierungsmusters sowie die Veränderung der *yellow* Expression zwischen Arten erklärt. Mehrere Arten, darunter *Drosophila melanogaster*, haben die Flügelpigmentierung im Laufe der Evolution unabhängig voneinander sekundär verloren. Welche Mechanismen dem Merkmalsverlust verursachen ist unklar. Das Hauptprojekts dieser Dissertation ist der Vergleich der Chromatin-Zugänglichkeit von *D. melanogaster* mit einer eng verwandten Art mit Flügelpigmentierung während der Entwicklung der Puppenflügel. Hierbei wurde eine Region mit erhöhter Chromatin-Zugänglichkeit in *D. melanogaster* aufgedeckt, die einem neu evolvierten Silencer entspricht. Dieser Silencer unterdrückt spezifisch den nahe gelegenen Enhancer, der das Pigmentierungsmuster reguliert. Ich fand auch heraus, dass der Transkriptionsfaktor Eip93F an der Regulierung der erhöhten Zugänglichkeit dieses Silencers beteiligt ist. Darüber hinaus deuten Enhancer-Reporter-Assays darauf hin, dass in der Region mit erhöhter Chromatin-Zugänglichkeit eine Repressorstelle hinzugekommen ist.

Zusammenfassend, lässt sich durch diese neuen Erkenntnisse ein neuartiges Modell aufstellen, welches den evolutionären Verlust eines Merkmals erklärt: Die im Laufe der Evolution erworbenen Sequenzvariationen führten zu einer erhöhten Chromatin-Zugänglichkeit der *cis*-regulatorischen Region. Dieses führt gleichzeitig zur Rekrutierung von Repressoren welche nahe gelegene Enhancer beeinflussen. Durch die Untersuchung des molekularen Mechanismus des Silencing eines evolutionär gewonnenen Enhancers bietet diese Dissertation neue Einblicke in das Zusammenspiel von Enhancern und Silencern.

Darüber hinaus habe ich im Rahmen dieser Dissertation an zwei weiteren Projekten mitgewirkt, die sich mit dem *yellow* Enhancer beschäftigten. Das erste Projekt untersuchte die Grenzen zwischen neu evolvierten Enhancerelementen sowie die Art und Weise, wie sie sich gegenseitig regulieren. Das zweite

Projekt befasste sich mit der Analyse der regulatorischen Syntax dieses Enhancers, um zu verstehen, wie die Expression von *yellow* in bestimmten räumlichen Mustern reguliert wird.

Introduction

1. Enhancers: essential elements in regulating gene expression in eukaryotes

1.1. Overview of gene transcriptional regulation in eukaryotes

Gene transcriptional regulation is a fundamental process in biological systems. The discovery of DNA as genetic material and its double helix structure[1] unveiled a new era of molecular biology. This discovery, followed by Jacob and Monod's work on the regulation of the *lacZ* expression in *Escherichia coli (E. coli)*[2,3], triggered tremendous research in elucidating the mechanisms governing gene expression. Although gene transcriptional regulation is more complicated in eukaryotes, the fact that *cis*-regulatory elements control gene expression has been proven true in eukaryotes for decades[4].

While Jacob and Monod's work illustrated how DNA sequence regulates gene expression without knowing the exact sequence, we now have access to the whole genome sequence, which can be probed with molecular and biochemistry experiments. The advancement of techniques helps us understand gene regulation from a genome-wide perspective rather than a single site. For instance, the Human Genome Project sequenced the entire human DNA sequence, first released in 2001[5], showing that the human genome comprises 3.2 billion base pairs, while only 1-2% of this sequence encodes proteins[6]. The instructions for gene expression are hidden in the remaining regions of the genome.

The Encyclopedia of DNA Elements (ENCODE), another project launched in 2003, integrated data from multiple newly developed sequencing methods[6–8]. These methods are coupled with specific biochemistry assays, separating *cis*-regulatory regions from the rest of the genome before sequencing. The results unraveled a systematic picture of gene expression regulated by the *cis*-regulatory regions, presenting the correlation between gene transcription and histone modification, transcription factor binding, and chromatin accessibility in these *cis*-regulatory regions[7].

Jacob and Monod's research on *E. coli* inspired further studies on gene transcription in the evolution of mammals. A. C. Wilson, Mary-Claire King, and their colleagues proposed genetic variation in regulatory regions, rather than the amino acid changes, underlies evolutionary diversity[9–11]. In his 1977 paper "Evolution and Tinkering", Jacob elaborated on how variation in *cis*-regulatory sequences may shape animal diversity[12]. This idea was supported by numerous studies focusing on the morphological variations in development and evolution, including the study of skeletal morphology in stickleback[13] and the study of pigmentation in fruit flies[14], which will be further discussed in this Introduction. Carried out alongside ECODE, the model organism Encyclopedia of DNA Elements (modENCODE) further supported Jacob's

idea from a genome-wide perspective, allowing the comparison of sequences involved in gene transcriptional regulation among species[15].

Despite these efforts, accurately predicting a given *cis*-regulatory region's function from its sequence remains challenging. One reason is that gene transcriptional regulation is highly dynamic and varies significantly during development and cellular processes; consequently, certain *cis*-regulatory regions are known to function with temporal-spatial specificity. However, the experiments exploring gene transcription capture signals only from a particular stage of cells. As such, predicting the precise function of *cis*-regulatory regions from their sequence remains difficult.

Apart from identifying *cis*-regulatory regions in the raw sequence, another focus in transcriptional regulation is understanding their mechanism of action. In eukaryotes, *cis*-regulatory regions can be thousands to millions of base pairs away from their target genes[16–18]. Also, these regions are usually wrapped into the nucleosome and even buried in a compact chromatin structure when inactive[19]. How are they activated and connected to their target genes? Here is an approximate answer: when the cellular signaling cascades are activated so that certain groups of transcription factors and proteins function together in the nucleus, the chromatin is then unpacked, the DNA is locally depleted from nucleosomes, and the regulatory regions become accessible to DNA-binding proteins, thus priming them for the assembly of the transcriptional machinery.

Studies using biochemistry and molecular biology methods have resulted in a relatively detailed model of the assembly[20–25], as briefly illustrated in Figure 1, which typically involves several components, including DNA elements for protein binding, such as enhancers and promoters, and proteins required for transcription initiation, including the Mediator, sequence-specific transcription factors, co-factors, and the RNA polymerase II (Pol II). Enhancers are looped or linked to their corresponding promoters through transcriptional machinery assembly[26,27]. This model allows enhancers to contact their targeted promoters from a distance and then recruit Pol II and the elongation complex[28]. However, this model focusing on the initiating complex of gene transcription does not present the specificity of transcriptional regulation in time and space. In other words, it does not provide a clear explanation as to why certain genes are expressed in specific cell types while others are not. To illustrate this, another layer of regulatory mechanism is required.

A part of that layer of regulation is chromatin accessibility. It intertwines with the other part of the regulation, DNA-binding transcription factors conferring spatiotemporal specificity of gene expression. The earliest proposal of accessible chromatin regions involved in regulating gene expression was brought up by several studies around 1980[29,30]. It implied that chromatin accessibility at the 5’ end of a gene

controlled transcriptional specificity. For example, Carl Wu and colleagues discovered in 1979[31,32] that light digestion of nuclei with DNase I could cleave certain regions into fragments of discrete lengths. A follow-up study mapped some of these sites to the 5' end of two heat-shock genes[29]. Studies from the Weintraub group focusing on chicken globin genes also found that DNase I-hypersensitive regions were distributed around the globin genes[33,34], and these results could only be observed in specific cell types expressing them. These studies suggested that DNase I-hypersensitive regions might have a role in regulating specific gene expression[35].

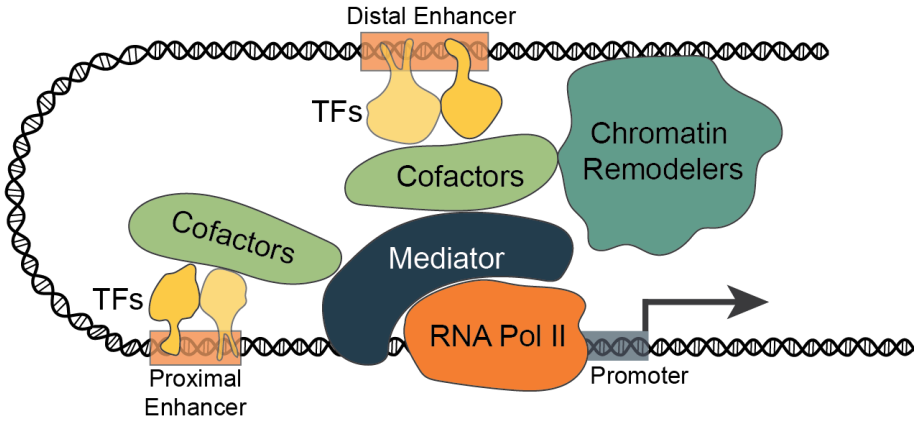

Figure 1. A sketch illustrates the transcriptional machinery at the transcriptional start site. *Enhancers recruit transcription factors (TFs), which subsequently recruit other proteins such as cofactors and chromatin remodelers. Together with these proteins, the Mediators and RNA Pol II assemble the final transcriptional machinery between enhancers and their targeted promoters and initiate the transcription.*

In 1981, a study from the Schaffner lab using SV40 discovered enhancers as a part of *cis*-regulatory components[36]. The enhancer was then defined as a DNA fragment located outside the coding region, capable of increasing gene transcription by several folds, regardless of its location or the gene it regulated. Shortly after, the first tissue-specific endogenous enhancer, functioning specifically in myeloma cells, was discovered in the intron of *IgH*[37–39]. The link between enhancer chromatin accessibility and gene expression was uncovered shortly after the discovery of the enhancer[40]. It became clear that activated enhancers were hypersensitive to DNase I, indicating that the functioning enhancers were accessible.

Due to limitations in sequencing techniques and computation power, it was only in this century that scientists uncovered genome-wide changes in chromatin accessibility and their relationship with gene transcription regulation. The ENCODE project found that about 4% of the genome sequence was accessible in at least one cell type used in the study[7]. This also means that genome-wide, the chromatin accessibility

landscape in each cell type differs, echoing the hypothesis proposed decades ago based on the discovery from several loci in isolation. The development and application of new techniques offer opportunities to study the mechanism underlying chromatin accessibility changes and how these changes affect gene transcription from a genomic perspective.

Furthermore, technological development promotes the investigation of how sequence variation in enhancers alters gene transcription. Transcription factors and chromatin proteins control chromatin accessibility by binding to the nucleosome-bound DNA and remodeling the chromatin structure or evicting the nucleosome so that the DNA is unwrapped from the nucleosome structure[41,42]. The nucleosome-free DNA motifs can then be recognized by specific transcriptional factors, and this binding further stabilizes the nucleosome-free state[41]. Any sequence variation can potentially alter the binding between these proteins and DNA, affecting the regulation of chromatin accessibility and gene transcription.

Yet, unlike reading the sequence in the protein-coding region with the codon table, we do not have an instruction book to decode the sequence in the *cis*-regulatory areas. Despite knowing that sequence variation in *cis*-regulatory regions is associated with trait evolution and the development of diseases[43], what variations and how these changes alter the regulatory function of enhancers remain unknown. Therefore, investigating the relationship between sequence variation and chromatin accessibility might offer insights into decoding the instruction book of the enhancer regulation. Eventually, these findings pave the way to understanding how enhancers evolved to regulate diverse gene expression among species and what mutations in enhancers can cause dysregulation of disease-associated gene expression.

1.2. Decoding enhancer regulatory mechanisms: understanding how sequence variation in enhancers shapes gene regulation

Despite an overall picture of gene transcriptional regulation, how the sequence of enhancers encodes spatiotemporal instructions remains unclear. For example, it is known that enhancers have binding motifs to recruit transcription factors[20]. However, locating these transcription factor binding sites is challenging because transcription factors sharing the DNA binding domain from the same family often bind to similar motifs enriched in *in vitro* experiments[44–46]. Also, multiple enhancers can regulate the same gene expression at a particular time and location during development or any cellular process[47]. Sometimes, a single enhancer is sufficient to drive gene expression, while others are only required in response to environmental stress[48]. Sequence variation can happen in these functional elements and alter the enhancer function. However, finding how the variation affects the enhancer function requires a comprehensive understanding of the code in the enhancer in the first place.

Besides the fact that mutations in enhancers may be related to specific diseases, as mentioned above, it was shown two decades ago that enhancers also harbor mutations responsible for evolutionary changes in gene transcription across species[14], as Jacob expected in his Evolution and Tinkering[12] paper. In summary, it is fundamental to understand how the enhancer sequence variation alters gene transcriptional regulation. This introduction will summarize the current literature in this direction, focusing on chromatin accessibility and the functional organization of enhancers. The discussion will be presented in six sections:

Section 2 – 3: "How" – How mutations may affect transcription factor binding sites, pioneer transcription factor binding sites, and histone modifications, changing enhancer functions.

Section 4: "When" – Mutations may affect the enhancer function during development and cellular processes. They may affect the enhancer functions at a particular developmental stage when specific transcription factors are expressed or presented.

Section 5 – 6: "Where" – Where the mutations may occur. Modifications may occur in the primary enhancers, shadow enhancers, or as part of the super-enhancer hub. The mutation effect may be difficult to observe due to these types of enhancer redundancy.

Section 7 will introduce the system I used to study how sequence variation changes enhancer regulation: *Drosophila* wing pigmentation.

2. Enhancer activation I: chromatin structure remodeling

2.1. Regulatory mechanisms of chromatin accessibility in enhancers

2.1.1. Enhancer chromatin accessibility linked to the specificity of enhancer activity

As briefly discussed in Section 1, it has been known for a long time that chromatin accessibility is a feature of active enhancers. First proposed in 1974 by R. D. Kornberg[49], the model of DNA packaging into chromatin through wrapping around histone oligomers with 200 bp as a unit was supported by following studies using nucleases to digest native chromatin, including DNase I and MNase[50,51]. Both enzymes could digest native chromatin into fragments around 200 bp, reflecting the nucleosome unit[50,51]. Around the same time, Weintraub and Groudine applied DNase I to chromatin extracted from different tissues, showing that active genes were sensitive to this nuclease[52].

Subsequently, studies from Wu *et al*. using DNase I to digest nuclei lightly showed interesting results[29,31,32]: the recombined DNA with heat shock genes showed discrete bands after the digestion, and the specific sites that account for the discrete bands could be mapped to the 5’ end of the heat shock genes. In addition, these

bands were maintained with some changes once these genes were activated. Furthermore, repeating the same experiment in the *Drosophila* embryo showed some loci had different DNase I sensitivity from that in the culture cells. The results indicated that DNase I-hypersensitive sites were unpacked from more compacted chromatin structures but were still occupied by nucleosomes or other proteins, allowing DNase I to bind and function. The study concluded that these sites might be specific for regulating gene expression. Shortly after this study, more work in *Drosophila*[53], chicken[30,33], and mammalian tissues[54] was published, showing that most, but not all, DNase I-hypersensitive sites were located close to active genes with tissue specificity.

The presence of DNase I-hypersensitive chromatin in *cis* regions of active genes led to the proposal of a model of gene transcription[29,35]: the open chromatin is required to recruit RNA Pol II to initiate gene transcription and thus confer the gene transcription specificity. This simple model has been refined by discoveries over four decades. One more component, the enhancer, was introduced in the same volume of *Cell*[36], in which the open chromatin model was presented. Following the discovery of the enhancer in 1981, a connection between DNase I-hypersensitivity and functioning enhancers was also established[40,55–57]. Also, many studies would perform a DNase I-hypersensitivity assay to examine the chromatin status of the newly discovered enhancers, reinforcing the link between DNase I-hypersensitivity and active enhancers. The chromatin region that is hypersensitive to DNase I digestion is now known as accessible or open chromatin[58].

However, are DNase I-hypersensitivity assays sufficient to screen enhancers from a given chromatin DNA region? Or can DNase I-hypersensitivity results predict enhancers? New techniques allowed the test of these hypotheses at a genome-wide scale rather than with single experiments to test each locus. The early results emphasized the link between chromatin accessibility and gene transcription[59,60]. For instance, the development of chromatin immunoprecipitation (ChIP) and DNA microarray technologies[61,62] allowed genome-wide studies of nucleosome positioning. ChIPs of H4 in budding yeast displayed reduced H4 occupancy at the 5’ regions of actively transcribed genes, meaning that these regions had a lower level of histone occupancy[60]. Similarly, a DNase-chip experiment in six human cell types showed that the cell type-specific DNase I-hypersensitive sites were mainly enhancers[63]. More recently, measuring chromatin accessibility by DNase I digestion or similar experiments coupled with next-generation sequencing (NGS) has been widely used for investigating *cis*-regulatory mechanisms[64].

The NGS-based techniques[65], including DNase-seq[64], MNase-seq[66], FAIRE-seq[67], and ATAC-seq[68], promoted the study of developmental enhancers and their accessibility changes during development and

differentiation. The results from these studies supported that chromatin accessibility is connected to dynamic gene transcriptional regulation during development and will be discussed further in Section 3.

2.1.2. Nucleosome structure and chromatin remodeling

The DNase I-hypersensitivity discussed above indicated that the chromatin structure, where DNA was wrapped around histones and was relatively insensitive to DNase I, could sometimes be converted into DNase I-hypersensitivity sites, allowing the nuclease to function. As more studies were conducted, it became increasingly evident that there was a correlation between these DNase I-hypersensitivity sites and enhancer activity. As a result, researchers started to focus on understanding the dynamics of chromatin structural organization. The nucleosome is the basic unit of the chromatin structure, so researchers directed their efforts toward understanding the nucleosome structure and how it is unpacked to free the DNA[69].

Each nucleosome comprises two copies of each histone (H2A, H2B, H3, and H4) that form an octamer structure[70] (Figure 2). This histone octamer is wrapped around a 147 bp DNA fragment and maintains contact with the DNA *via* hydrogen bonds between its arginine and the DNA minor groove[71]. The tails of histones are rich in lysine and extend from the nucleosome passing through the DNA[71]. This extension structure further stabilizes the nucleosome and is essential because bending is not favorable for DNA molecules alone[71]. In addition, linker histone H1 can add another layer of structural complexity to the nucleosome by binding to one end of the nucleosome DNA and the DNA at the dyad position, further stabilizing its structure[71].

The regulation of chromatin structure remodeling, which involves the displacement or disassembly of nucleosomes, is facilitated by chromatin remodelers[72]. The recruitment of these remodelers is typically linked to specific modifications on histone tails, referred to as histone marks[72]. These marks are mostly located on lysine residues and are subject to modifications[72]. Certain histone marks are closely related to active *cis*-regulatory regions, including promoters and enhancers, which will be discussed in the upcoming section. Although the connection between histone marks and chromatin remodeling is well-established, it remains uncertain whether these modifications are the cause or consequence of chromatin remodeling[73,74].

One of the histone modifications associated with the displacement of nucleosomes is the methylation of the histone H3 lysine 4 (H3K4) residue. This modification is regulated by the COMPASS complex containing methyltransferases[75]. The methylated H3K4 sites are associated with the binding of chromatin remodelers, such as NURF[76,77] and CHD[78] for H3K4 trimethylation (H3K4me3) and BAF for H3K4 monomethylation (H3K4me1)[79]. Another key histone modification involved in displacing nucleosomes is the acetylation of histone H3 lysine 27 acetylation (H3K27ac), regulated by histone acetyltransferases (HAT) domain

proteins[80,81]. The existence of H3K27ac is also associated with the recruitment of chromatin remodelers in the same locus, mainly the SWI/SNF family[82,83]. However, the mechanism regulating the recruitment of these chromatin remodelers remains to be uncovered.

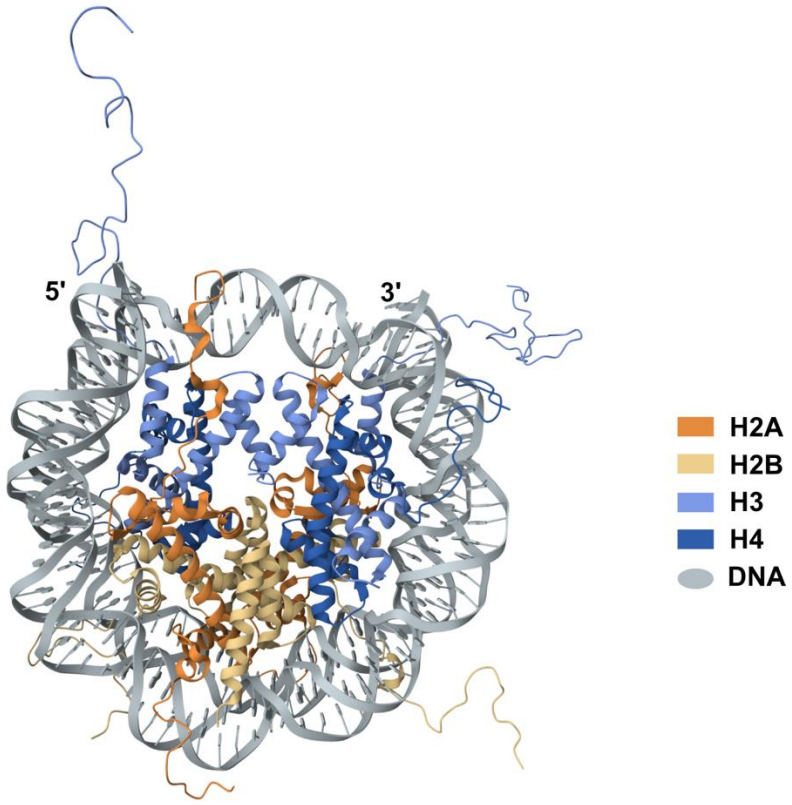

Figure 2. The crystal structure of the nucleosome core particle (PDB ID: 1KX5). *The nucleosome core particle comprises histones derived from Xenopus laevis, and a DNA fragment obtained from humans*[84]*. The histones are organized in an octamer structure with two copies of H2A, H2B, H3, and H4. Each component is marked by a distinct color, as indicated in the figure. This figure was created using Mol* on the PDB website* [84–86].

Nevertheless, chromatin remodelers are essential for displacing nucleosomes. They are protein complexes remodeling chromatin in an ATP-dependent manner, but the ATPase domains vary among different families[19]. The chromatin remodelers can be grouped into SWI/SNF, ISWI, CHD, and INO80 families based on the differences in the ATPase domain[19]. The SWI/SNF family is the major regulator of chromatin accessibility and can slide and evict nucleosomes[42,87]. First discovered in yeast as swi/snf protein, the orthologue complexes of SWI/SNF complex in flies and mammals are known as BAP and PBAP, and BAF and PBAF, respectively[19,88]. A recent study focusing on the regulatory mechanisms of SWI/SNF has suggested that the SWI/SNF complex could translocate DNA fast enough to disrupt multiple interactions

between DNA and the nucleosome, leading to the ejection of histones[42,89]. Further studies have revealed that SWI/SNF is connected to enhancers, indicating its role in regulating chromatin structure for enhancer function[83,90].

Conversely, certain chromatin remodelers, such as those from the ISWI family, act as repressive factors, hindering DNA accessibility by reassembling the nucleosome structure[19]. These remodelers form an integral part of chromatin accessibility regulation, and their recruitment at specific genomic locations shapes the accessibility landscape of chromatin.

2.1.3. Specific histone modification marks associated with enhancer activity state

As we have seen above, the control of accessibility happens primarily at the level of histone tails. Early studies associated histone modifications in *cis*-regulatory regions with gene transcription[80,91]. This led to the question of whether a particular combination of histone modifications only marks enhancers but not other *cis*-regulatory elements such as promoters, the so-called "histone code[92]." With the advent of genomics, scientists have sought to seek a universal histone modification code for each regulatory element genome-wide. In a study by Heintzman *et al.*, H3K4me1 was characterized as an enhancer-specific histone mark by ChIP-chip (ChIP followed by DNA microarray), while H3K4me3 was characterized as a promoter-specific mark. This was the first time that promoters and enhancers were distinguished globally based on histone modifications[93]. The study also found that most selected enhancers with H3K4me1 were inactive in the reporter assay.

At the same time, another histone modification, H3K27ac, was also found to be involved in regulating enhancer activity. It has long been known that histone acetylation is correlated with the accessibility and activity of regulatory elements[81,94–96]. The study from Heintzman *et al.* also found that H3K27ac was enriched in active enhancers and promoters[93]. The P300/CBP coactivator can deposit this histone mark through the HAT domain[81]. Recent research has shown that H3K27ac can recruit SWI/SNF complex through p300 to regulate nucleosome positioning, facilitating the DNA becomes accessible[83]. These studies help to explain earlier observations that histone acetylation is associated with regulatory elements and increased chromatin accessibility[93,94,97,98].

Furthermore, another study performed ChIP-seq (ChIP followed by sequencing) experiments targeting H3K4me1 and H3K27ac[98]. It measured the activity of all enhancers with both marks using corresponding gene expression levels from RNA-seq data. The results suggested that a combination of H3K4me1 and H3K27ac is the histone code for active enhancers. The inactive enhancers were mostly marked by

H3K4me1 alone. This finding was further explained by the following studies, which suggested that these enhancers were poised and had the potential to be activated upon stimulus or during development[99,100].

Researchers have widely used the presence of H3K4me1, H3K27ac, and chromatin accessibility in *cis*-regulatory regions to identify active enhancers genome-wide. As discussed previously, rather than initiating the regulation, the histone modifications at the enhancers may result from recruiting proteins to regulate enhancer activity independent of their enzyme activity but through other protein-protein interactions[73,74]. These findings suggested that the histone modifications could be the byproduct of the recruitment. As such, the detailed mechanism of the interplay between chromatin regulators and enhancers still needs to be discovered.

2.2. Pioneer transcription factors in initiating chromatin remodeling

We have discussed the regulatory mechanisms of chromatin accessibility at enhancers, but how does this process begin? It is likely that the recruitment of chromatin remodelers to specific chromosome positions is associated with transcription factors binding to the nucleosomes[101]. Most transcription factors only bind to specific DNA motifs not wrapped around nucleosomes, but some transcription factors can bind to nucleosome-bound DNA. A high-throughput *in vitro* study published in 2018 examined the binding preferences between transcription factors and nucleosome-bound DNA by using over 200 purified DNA binding domains of transcription factors or full-length transcription factors and a nucleosome-bound DNA library[102]. The study then sequenced the transcription factors-bound DNA fragments and found a small subset of transcription factors, including pioneer factors, that recognized specific nucleosome-bound DNA motifs and bound to them *in vitro*[102].

Pioneer factors are known for their ability to bind to nucleosome-bound DNA[103]. They were initially discovered by their ability to bind to nucleosomes *in vitro* and to closed chromatin *in vivo*[103,104]. This process, followed by the recruitment of cofactors or chromatin remodelers, establishes chromatin accessibility[105,106], allowing other transcription factors to bind to DNA as well. One of the classic pioneer factors is FoxA (also known as HNF3), a critical regulator in liver development that is always bound to DNA earlier than any other transcription factors during this developmental process[107]. It was the first transcription factor proven to bind to nucleosome-bound DNA, displacing H1 from the nucleosome structure and making the linker DNA more accessible[107]. The pivotal role of pioneer factors in development and disease has been further identified[108]. During the activation of zygotic genome expression, it was shown that pioneer factors govern accessibility in three distinct species: Zelda for *Drosophila* [109,110], Nanog, Pou5f1, and SoxB1 for zebrafish[111,112], and Nr5a2 for mice[113]. These recent studies combined with genome-

wide methods to profile chromatin accessibility and transcription factor-DNA binding also indicate that the binding of these pioneer factors at silent enhancers during earlier developmental stages predicts open chromatin at these enhancers at later stages, followed by corresponding gene expression.

While the pioneer factors are master regulators of multiple cellular reprogramming processes, they regulate chromatin accessibility through different mechanisms. FoxA, for instance, frees the linker DNA by displacing H1[107,114]. Some pioneer factors bind to nucleosomes and release the adjacent gyre of DNA to make it accessible for other protein-DNA binding[114,115]. Pioneer factors can also recruit chromatin remodelers upon binding to further stabilize the accessible status, although the structural details of how pioneer factors recruit chromatin remodelers are still unclear[105,106].

In summary, the binding of pioneer factors to nucleosome-bound DNA is typically the first step in opening the enhancer chromatin structure, a crucial step in enhancer activation. As discussed, pioneer factor-binding is always accompanied or followed by histone modifications and remodeling to open the enhancer-associated chromatin. Moreover, cooperating with other transcription factors[116] allows pioneer factors to establish a more stable transcription factor-nucleosome binding complex while facilitating specificity, which will be further discussed in Section 3.

3. Enhancer activation II: transcription factors

3.1. Transcription factors in enhancers shaping specific gene expression

The enhancer function is regulated by the transcription factors binding to it. The genome's general accessibility landscape, established by pioneer factors, is necessary but insufficient to activate enhancers alone. To further stabilize chromatin accessibility, pioneer factors often cooperate with other transcription factors[116–118]. For instance, during induced pluripotent stem (iPS) cell induction, Oct4, Sox2, and Nanog are pioneer factors responsible for opening necessary enhancers[119,120]. However, the induction also requires another transcription factor, c-Myc[119]. Studies showed that c-Myc only binds to accessible regions and stabilizes the binding of the three pioneer factors mentioned earlier[120]. While most transcription factors cannot bind to nucleosome-bound DNA, they are crucial for gene transcriptional regulation. Humans have around 1600 transcription factors[121], and approximately 700 are found in *Drosophila*[122]. This abundance of transcription factors allows them to function collectively to yield high specificity in gene transcriptional regulation[21].

The varied spatial distribution of transcription factors also contributes to specific gene transcriptional regulation[123]. Pioneer factors initiate development at particular stages, and cells undergo further

differentiation even if they share the same pioneer factor. This specificity stems from different subsets of transcription factors expressed in different cells, often resulting from morphogen diffusions[124]. In some cases, multiple layers of gene transcriptional regulations operate concurrently to specify cell types and final morphological patterns of the organism[125].

Many of these principles emerged from studying animal embryogenesis, such as *Drosophila melanogaster* (*D. melanogaster*). *Drosophila* embryonic development is a model system for studying gene transcriptional regulation in patterning. After fertilization, the embryo undergoes 14 mitotic cycles to form a blastoderm with cells[126]. During these cycles, the embryo is a syncytium, where transcription factors diffuse freely without passing through any cell membrane. Zygotic gene transcription, initiated at cycle 8, is activated by the pioneer factor Zelda, homogenously distributed in the embryo during this stage[110]. Other transcription factors distributed in different spatial patterns subsequently control the detailed body plan.

Initially, transcription factors translated from maternally deposited mRNA, such as *bicoid* and *nanos* in flies, diffuse from different parts of the embryo to form a gradient distribution of transcription factors[126]. The combination of activators and repressors with varying concentrations along the anterior-posterior (A-P) axis activates the next tier of regulators, the so-called "gap" genes at different positions along the embryo[126]. These gap genes, which are transcription factors, work together with maternal transcription factors to further activate the “pair-rule” genes, such as *even-skipped*[126]. The unique combination of transcription factors at different positions along the A-P axis regulates *even-skipped* expression in seven sharp stripes by cycle 14[126]. Moreover, ChIP-seq against Zelda showed that before the expression of *even-skipped* starts, Zelda is already bound to its enhancer region to prime the accessibility at later stages[110]. This classic example illustrates how transcription factors' sequential and combinatory binding regulates chromatin accessibility and specific gene expression.

Moreover, enhancers usually recruit multiple transcription factors to bind, compounding the complexity of enhancer regulation. The instructions for such recruitment are encoded in the enhancer sequence, known as transcription factor binding motifs. Unlike the three base-pair codes guiding amino acid assembly, these motifs vary in length and complexity.

Transcription factors bind to preferred DNA sequences as monomers, dimers[127], or tetramers[128], similar to protein-protein interactions. While transcription factors from the same family sometimes prefer similar motifs in *in vitro* experiments, they regulate distinct gene sets *in vivo*[129]. For instance, homeodomain transcription factors usually favor binding to "TAAT" motifs *in vitro*[44,129]. However, some homeodomain transcription factors resolve this discrepancy by cooperatively binding to DNA with cofactors *in vivo*,

changing their binding motifs and increasing specificity[130]. This mechanism raises the question: How prevalent are secondary binding motifs for transcription factors via cooperative binding *in vivo*?

Transcription factors not only bind to high-affinity sites but also to low-affinity sites. Recent studies have revealed that increasing the binding affinity of low-affinity sites through mutation can lead to ectopic gene expression during development[131–135]. Interestingly, the sequences of these low-affinity binding motifs sometimes differ from corresponding high-affinity motifs[131]. Therefore, one transcription factor may bind to various DNA sequences *in vivo*, further complicating the enhancer syntax.

Additionally, the arrangement of transcription factor binding sites varies. There are three proposed models: enhanceosome, billboard, and collective (See Figure 3)[21]. According to the enhanceosome model, all binding sites have fixed positions and directions[21,136]. On the other hand, the billboard model suggests flexible alignment within enhancers, though the combination of the binding motifs is fixed[21,137]. Lastly, the collective model proposes that the same group of transcription factors can bind to different enhancers in distinct arrangements[21]. These proteins sometimes can be recruited by other transcription factors without binding to DNA.

Transcription factors decode enhancer sequences and act as sophisticated regulators within open or accessible enhancers. The complex interplay between transcription factors and enhancers results in highly specific gene transcriptional regulation[138]. Consequently, enhancers are only activated when the necessary group of transcription factors with appropriate concentrations is present in the nucleus.

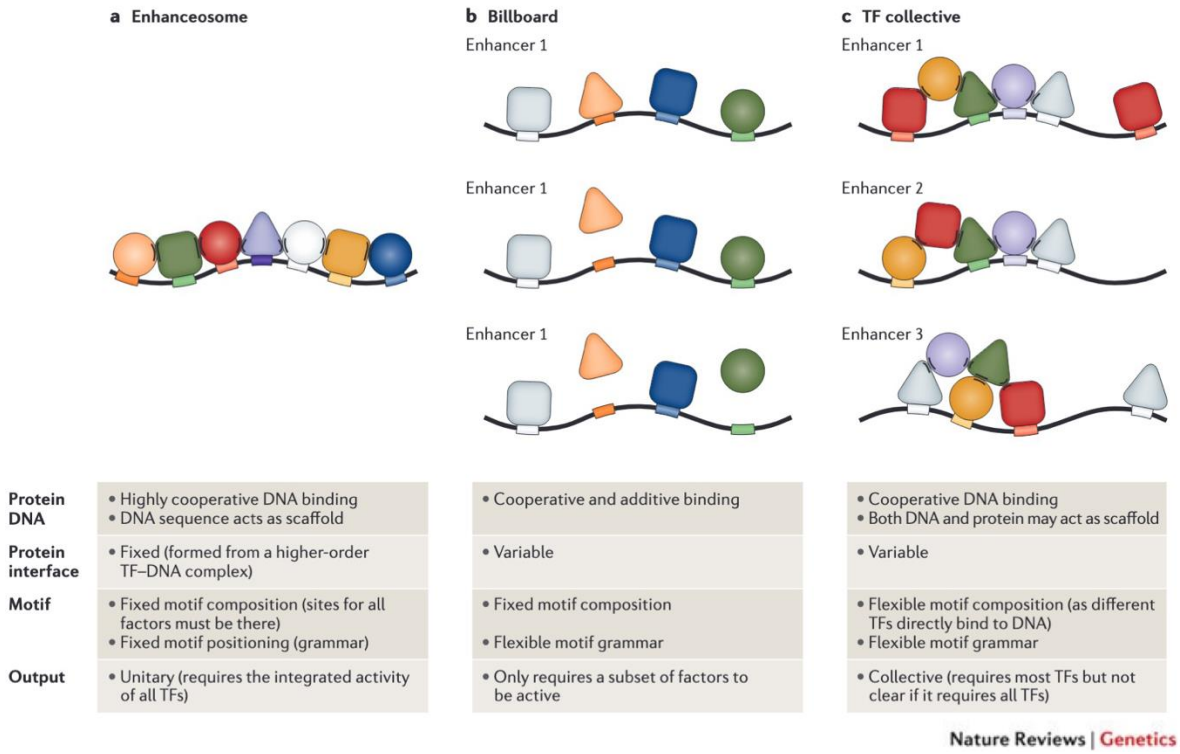

***Figure 3. Models of the arrangement of transcription factor binding to enhancer*[21].** *(a.) The enhanceosome model suggested that multiple transcription factors bind to an enhancer with a strict arrangement. (b.) The billboard model suggested that not all transcription factor binding sites need to be occupied to activate an enhancer. (c.) The collective model proposed that the same group of transcription factors can be recruited by multiple enhancers. These enhancers may not have the binding sites for all transcription factors presented here, as some transcription factors may be recruited by others without binding to enhancers directly. This figure is cited from Spitz et al. with permission from Springer Nature*[21].

3.2. Enhancer variation modulating enhancer function and shaping phenotypic evolution

The earlier part of this section summarized the role of proteins in regulating enhancer function, from modulating accessibility to initiating activity. However, pioneer factors and spatial transcription factors do not consistently bind to their preferred DNA motifs *in vivo*. Fully understanding the organization of the information within enhancers and how it guides transcription factors and other regulatory proteins remains challenging. Fortunately, naturally occurring enhancer variation and their associated phenotypical diversity offer a valuable system for investigating enhancer regulation[43].

One type of natural variation in enhancers researchers have been focusing on is those related to diseases[139]. One study from the ENCODE project was the first to combine genome-wide association study (GWAS) data with the DNase-seq[140]. This study applied DNase-seq to profile DNase I-hypersensitive sites from different cell types and tissues and compared these sites with disease-associated single nucleotide polymorphisms (SNPs). The results indicated that these SNPs tend to be located in disease-related tissue-specific DNase I-hypersensitive sites. In a separate, more in-depth study[141] focused on low-density lipoprotein cholesterol (LDL-C)-associated diseases, it was uncovered that among the LDL-C-related variants on chromosome 1p13, one variant mutated a binding motif for a liver-enriched transcription factor, C/EBP. This mutation decreased the expression level of SORT1, a gene related to LDL regulation. These studies highlighted the significance of enhancer variation, but evidence for direct causation[142] between enhancer variation and diseases remains elusive.

The idea that natural variation in enhancers could directly cause diseases is intriguing yet challenging to investigate, given that diseases often involve multiple cell types and tissues without a simple readout to evaluate the effects of mutations in enhancers. Alternatively, examining how enhancer variation contributes to phenotypic diversity in simpler systems may be more feasible. Morphological evolution results from gains and losses of discrete traits among closely related species[43]. Nearly two decades of evolutionary developmental biology have unraveled the genetic principles of these changes.

First, changes in regulatory regions of developmental genes likely represent the main driver of these gains and losses[43]. One such classic model is the pelvic apparatus in sticklebacks. Marine sticklebacks have this bony structure, homologous to the hindlimbs of terrestrial vertebrates, extending downward and backward from their ventral regions[13]. Studies have shown that more than ten fresh-water stickleback populations worldwide have lost or reduced their hindlimbs, while most marine sticklebacks have retained them[13]. This trait is directly related to the expression of the *Pitx1* gene[13]. A 2.5 kb enhancer is sufficient to drive *Pitx1*

expression in sticklebacks, which was identified and termed *pel*[143]. Multiple-sequence alignment of the loci around *Pitx1* among sticklebacks suggested that the partial deletion of *pel* in fresh-water sticklebacks led to the loss of the *Pitx1* expression[143]. Recently, a study[144] on the stickleback fish skeletal trait identified an enhancer, regulating a bone morphological protein GDF6, that was gained during evolution and underlies armor-plate size changes. Interestingly, an enhancer driving GDF6 expression in hindlimb in chimpanzees was deleted in humans. This study indicated the gain and loss of enhancers from homologous loci are shared mechanisms resulting in parallel morphological evolution across species, from fish to humans.

Additionally, gains or losses of enhancers were also characterized in other species, resulting in phenotypical changes in evolution, such as limb loss in snake[145]. However, previous studies fell short of understanding the complete regulatory mechanisms in those contexts. Most studies did not identify the transcription factors or other regulators that control these regulatory regions. To study enhancer variation that drives morphological evolution, fruit fly species in the genus *Drosophila* is a classic model organism. Closely related *Drosophila* species have undergone changes in various morphological traits, such as trichome patterns in larvae[146,147], eye size [148], and pigmentation patterns on wings[14] and abdomens[149]. Further studies have indeed discovered the gain or loss of transcription factor binding sites during enhancer evolution, leading to morphological trait evolution. Section 6 will discuss an example in which researchers investigated enhancer evolution in regulating trichome patterns on larvae. Section 7 will summarize another example focusing on enhancer evolution associated with wing pigmentation patterns, a trait used as the model for this dissertation.

So far, enhancer variation discussed in this section focuses primarily on those that directly alter phenotypes, regardless of when they function during development. As discussed in Section 2, the chromatin state of enhancers changes dynamically, allowing them to control specific gene expression during development or in response to environmental stimuli. The details of the dynamic regulation in cells and tissues will be further discussed in the next section.

4. Dynamic enhancer chromatin accessibility in regulating development

4.1. The dynamics of chromatin accessibility instruct development and differentiation

Previous studies have highlighted the variation in chromatin accessibility in different cell types. However, there is still a gap in our understanding of how changes in chromatin accessibility are linked to gene expression during development. Specifically, how is enhancer chromatin accessibility modulated during embryogenesis, when multiple organs and tissues develop from a single cell? The detailed questions include: first, whether the chromatin accessibility landscape can represent the developmental stages;

second, whether pioneer factors establish chromatin accessibility before cell differentiation or whether they function together with other transcription factors at the same developmental stages.

In Section 2, we discuss how NGS-based techniques such as DNase-seq, ATAC-seq, and ChIP-seq can help us understand enhancer accessibility regulation during different developmental stages. However, these bulk sequencing methods are limited in distinguishing between different cell types within a tissue or during embryogenesis because they rely on the amount of input cells.

Recently, single-cell RNA-seq (scRNA-seq)[150,151] and single-cell ATAC-seq (scATAC-seq)[152] techniques have revolutionized our ability to understand enhancer accessibility dynamics at a single-cell resolution during embryonic development in model organisms such as *Drosophila*[153], mice[154], and humans[155]. By analyzing scATAC-seq clusters, studies have discovered previously unknown tissue- and temporal-specific enhancers and have gained new insights into the intricacies of gene regulatory dynamics and developmental gene expression cascades.

While bulk-based sequencing methods hold statistical reliability for many studies, particularly those examining chromatin regulation at specific developmental stages within defined cell types[156], single-cell sequencing has significantly improved our ability to understand the finer details of chromatin accessibility dynamics and gene expression changes at an unprecedented resolution. In this section, I will provide a brief overview of research that examines the interplay between chromatin accessibility and gene expression during development, utilizing either bulk or single-cell sequencing-based methods.

In the ENCODE project mentioned in Section 1, one study published in 2012 conducted extensive DNase-seq on various human cells and tissues to investigate differences in chromatin accessibility[157]. Following this research, another study explored data from 38 types of primary cells to create a chromatin accessibility map that covers the development of major cell lineages post-gastrulation, human embryonic stem cells (hESCs), cells from hematopoietic lineage, and keratinocytes[158]. The study found that cells clustered based on chromatin accessibility better reflected their lineage branching than transcription levels did. This result suggests that the chromatin accessibility profile is not merely representative of the transcription in the cells but has unique patterns. Additionally, the number of accessible regions decreased as cell differentiation progressed, with different lineages having distinct sets of closed regions and additional sets of opened regions for specificity regulation - however, the details of how these chromatin accessibility changes related to cell differentiation remained unknown.

Over the last decade, multiple studies profiled changes in chromatin accessibility using stem cells or other tissues. These studies have generally agreed with the conclusions drawn from the initial chromatin

accessibility profiling across different cell types. Furthermore, several studies have unraveled detailed regulatory mechanisms governing chromatin accessibility dynamics, including the role of transcription factors in this process.

A study from McKay *et al.* focusing on *Drosophila* development aimed to determine how chromatin accessibility changes in enhancers affect the development of distinct thoracic appendages, such as forewing vs. hindwing (also called haltere in flies), as these tissues carry many similarities[159]. The researchers measured chromatin accessibility changes at three embryonic stages to determine how cell fate is decided. They also measured the chromatin accessibility differences in three types of appendage primordia (imaginal discs) during the third instar larvae stage, when the patterning phase is ending. The study showed that chromatin accessibility was similar among cells of the same developmental stage. The only enhancers with differentiated accessibility among the appendage discs were those that controlled the expression of master regulators that determine appendage types. For instance, the transcription factor Ubx, which specifies the haltere fate by repressing the forewing developmental program, showed increased chromatin accessibility in haltere discs compared to wing discs. Furthermore, accessibility differences between appendage discs and eye discs remained prominent, indicating that accessibility landscapes still differed in cells that developed into different organs.

The process that regulates chromatin accessibility during tissue differentiation may be more complex in embryonic development. Delás *et al.* conducted a recent study applying a modified ATAC-seq on a mouse stem cell model, which mimicked neural tube development in response to the Shh morphogen gradient[160]. The modified ATAC-seq allowed the application of ATAC-seq on different subtypes of cells during neural tube development. The results showed that during the development of neuron tube progenitors in response to the Shh morphogen, cells employed two strategies to modulate cell fate. One strategy used differentiated sets of transcription factors in different cells while all the cells shared common accessibility. The other strategy involved the expression of the transcription factor FOXA2, acting as a pioneer factor. The expression of FOXA2 in a specific subtype of neural progenitor cells remodeled the chromatin accessibility landscape so that only a distinct subset of enhancers would be accessible compared to other cell types.

Further, *in vivo* experiments performed by Delás *et al.* showed that, in mouse embryos, FOXA2 established the chromatin accessibility landscape before cells differentiate into neural progenitors. This finding is in line with other research indicating that master regulators facilitate chromatin accessibility before relevant gene expression occurs. A single-cell-based study on *Drosophila* embryo development revealed that the chromatin accessibility changes were mostly consistent with RNA expression[153]. The study also analyzed the loci of genes whose expression depends on Zelda, the pioneer factor in *D. melanogaster* embryogenesis.

In contrast to the general analysis, they found the enhancers of the Zelda-dependent genes opened much earlier than the expression of these genes. Another study by Bozek *et al.* assessed the chromatin accessibility from different patterning compartments at a particular stage during *Drosophila* embryo development[161]. The results from this study suggested that the enhancers regulating axis patterning were accessible before zygotic genome activation. The level of accessibility was later regulated by different subsets of transcription factors during embryo development[161].

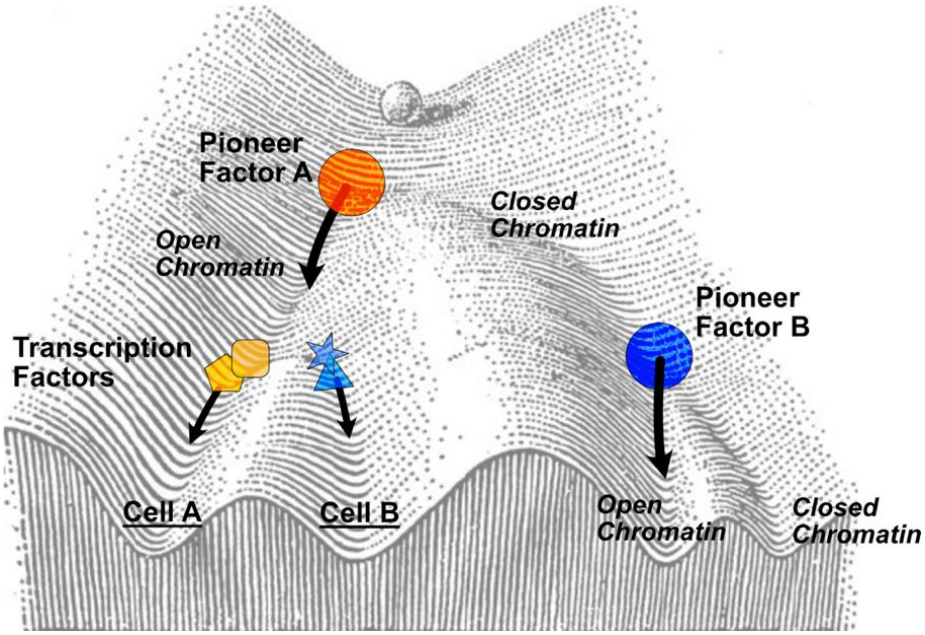

Figure 4. The Waddington epigenetic landscape and the chromatin accessibility dynamics. *The background of this figure displays the epigenetic landscape drawn by Conrad Hal Waddington[162]. Pioneer factor A selectively binds to specific enhancers, rendering their chromatin status accessible. Enhancers that are not bound by pioneer factor A remain in a closed chromatin status until their targeted pioneer factor B is activated. As cells differentiate, distinct sets of transcription factors further regulate the accessible enhancers.*

In 1957, Conrad Hal Waddington illustrated his famous model of epigenetics in his book, "The Strategy of the Gene[162]." In the epigenetic landscape he drew, he analogized the paths of cell differentiation to valleys emerging from a common starting point (Figure 4). Over the past several decades, his model has added more details. One such detail is that the dynamic of enhancer accessibility plays a crucial role in guiding cells to differentiate into subtypes (Figure 4). Cells from the early embryo or progenitor cells have more accessible regions than later stages[158]. As development progresses, some regions undergo chromatin closure. Meanwhile, a more limited set of cell-type-specific *cis*-regulatory regions has open chromatin. The

regulation of these accessible regions is orchestrated by master transcription factors or specific groups of transcription factors, each employing different mechanisms, as previously discussed. While the chromatin accessibility landscape can sometimes represent tissue identity, the regulation of enhancer accessibility dynamics is context-dependent and not a one-size-fits-all answer.

4.2. Evolutionary variation affecting enhancer function at specific developmental stages

Studying the impact of evolutionary mutations in enhancers on gene transcriptional regulation is challenging due to several factors. These include the varying developmental rates of different species, difficulty in accurately determining enhancer boundaries, which may change during development and across species, and variation in the expression profiles of transcription factors, leading to different gene expression outcomes. Recent research in the evolutionary development of the human brain tackled this question. To illustrate their findings, this section focuses on studies that explore the role of Human Accelerated Regions (HARs) and SNPs in regulating gene transcription during neuronal development.

HARs are genomic regions conserved across species that have accumulated unique mutations in humans[163,164]. Previous investigations[164,165], often relying on GWAS and transcriptomics data analysis, have suggested that HARs are crucial in regulating gene expression during human brain development and play a role in the evolution of the human brain. Because these findings remained inconclusive, some researchers sought to examine the mechanisms that govern HARs in gene transcriptional regulation more closely. These studies aim to understand how HARs function differently from their orthologous counterparts in other species.

Newly developed techniques, including NGS-based methods and modified Massively Parallel Reporter Assays (MPRA), offer scalable functional validations for HARs, providing insights into their functionality. In a recent study, 3142 out of 3171 known HARs were tested by MPRA in human and mouse neuronal cells[166]. The study revealed that many HARs act as active enhancers in neuronal cells, and the enhancer activity was traced back to sequence variation rather than trans-regulatory landscapes. These results aligned with Whalen *et al.*'s study[167], which employed modified MPRA to assess HAR enhancer activity in human and chimpanzee neuronal progenitor cells. They used DNase-seq profiles from the Roadmap Epigenomics Consortium database and *in vivo* epigenetic features from human embryo cortex samples to identify HARs with specific accessibility patterns during neuronal development stages. Subsequent DNase-seq and ChIP-seq analyses on human fetal brain Neural Progenitor Cells (NPCs) and neurons pinpointed 210 HARs with enhancer activity in neuronal development. Despite these discoveries, direct evidence proving that these HARs act as enhancers driving human-specific gene expression remained elusive.

One challenge to studying the HARs is investigating them in the context of *in vivo* development, as cell lines cannot capture the entire developmental process. With great respect to donors, researchers had the chance to conduct studies using human samples. One study investigated human cortical neurogenesis by dissecting neural progenitor-enriched and cortical plate regions from human samples at 15-17 weeks post-conception[168]. The study used a comprehensive approach that integrated ATAC-seq, RNA-seq, and Hi-C sequencing[169] and revealed a positive correlation between heightened chromatin accessibility, chromatin interactions, and concurrent upregulated gene expression levels. One significant finding was the identification of a human-specific enhancer, compared to mouse and monkey, that regulates the expression of the transcription factor FGFR2. FGFR2 is essential for neural progenitor cell proliferation. The results also showed that mutating this human-specific enhancer in human neural progenitor cells impeded their self-renewal capacity. This finding showed how these mutations in enhancers alter gene regulation and thus affect cellular processes.

A follow-up study derived neural progenitor cells from tens of human fetus samples[170]. The results also characterized the correlation between enhancer sequence variation and accessibility changes at specific development stages. Further, the results found one mutation in a human quantitative trait locus (QTL) that disrupted a transcription factor binding site located in an accessible region in both progenitor cells and neurons. This led to a decrease in the degree of accessibility of the region in progenitor cells but not in neurons. Finally, both studies compared disease-related SNPs with chromatin accessibility data and found that certain types of neuropsychiatric illness risk variants were located within the differentially accessible regions.

Together, these investigations reveal that mutations in enhancers can modify their own accessibility during specific developmental stages and spatial contexts. This, in turn, can significantly impact precise gene transcription and expression. In human neuronal development, such mutations may contribute to certain diseases by disrupting enhancer accessibility during specific developmental stages. These findings underscore the complexities of studying how mutations can rewire the gene transcriptional network, emphasizing the importance of conducting experiments with related cells or tissues at relevant developmental stages in different organisms. Additionally, it should be noted that the function of enhancers can change over time or in response to environmental stress, further adding to their complexity. The subsequent two sections will discuss how enhancers interplay with each other in detail. These studies emphasized the importance of understanding enhancer regulatory mechanisms to comprehend better or even predict human diseases.

5. The regulatory mechanisms of enhancer clusters

5.1. The classification of enhancer clusters

The introduction has highlighted the importance of enhancer accessibility in regulating gene transcription, and the notion that mutations in enhancers can alter gene transcription by changing chromatin accessibility. However, it is common in the genome that one gene is regulated by multiple enhancers, which was also discovered in the 1980s[47,171,172]. While chromatin accessibility is not the primary focus of this section, it is a feature widely used to identify enhancers for further study. This section and Section 6 will delve into the functional organization of enhancer clusters and bifunctional enhancers.

The first well-characterized enhancer cluster was the locus control region (LCR) that drives the expression of ß-globin[47,171]. LCR contains five independent DNase I-hypersensitive sites or five independent enhancers. The full expression of transgenic ß-globin in mice requires the presence of all five enhancers. Other examples of enhancer clusters include tens of LCRs (summarized in Li *et al.*[173]) in mammalian cells and a well-characterized enhancer cluster driving the expression of the pair-rule gene *even-skipped* in *Drosophila*[174,175]. Similar to the LCR of ß-globin, the *even-skipped* enhancer cluster has five independent enhancers, each driving one or two stripes during the blastoderm embryonic stage[174,176]. Similarly, the stripe expression of several other pair-rule genes, including *hairy*, *run*, *fushi tarazu*, and *odd*, are also regulated by enhancer clusters, in which each enhancer functions independently on one or two stripes[177].

As more research is conducted to understand the regulation of enhancer clusters, the list of classifications of enhancer clusters is expanding. Over the years, enhancer clusters have been identified based on biochemical markers or perturbations, leading to the classification based on different properties, such as super-enhancers[178], shadow enhancers[48], cryptic enhancers[179], and latent enhancers[180]. Among these, the super-enhancers and shadow enhancers have been extensively studied. Therefore, in this section, I will mainly focus on these two types of enhancers.

The concept of super-enhancer was first proposed in 2013 by Richard Young and his colleagues, referring to extended regions of enhancer clusters enriched with Mediator and master transcription factors[178]. A similar concept was also published around the same time in which the enhancer cluster was named stretch enhancer[181]. Super-enhancers are transcription hubs that drive critical and specific gene expression during cellular differentiation, as well as oncogene expression[182]. Unlike previous genetic methods used to identify enhancer clusters, super-enhancers can be identified genome-wide by performing and analyzing ChIP-seq against master transcription factors, Mediator, histone marks, and DNase-seq or ATAC-seq[178,182]. However, the differences between super-enhancer and previously known enhancer clusters were debated due to the similarity in their function[47,183,184]. A recent review suggested that the unique property of super-enhancers

is their sensitivity to chemical perturbation, disrupting the function of cofactor and chromatin regulator BRD4, which co-occupies enhancers with Mediator[47]. Therefore, more experimental data is needed to investigate whether super-enhancers regulation functionally differs from traditional enhancer clusters.

Shadow enhancers were first proposed by Michael Levine's group to describe enhancers located at the distal side to the promoter sharing redundant functions with the primary enhancers located at the proximal side[48,185,186]. The term is now more commonly used to refer to enhancers that share functional redundancy with other enhancers to regulate the same gene expression[187]. Before the discovery of shadow enhancers, the redundant enhancers had been identified several times, such as the enhancers of *Krüppel* (*Kr*)[172] and the enhancers activated during dorsoventral (DV) patterning in *Drosophila* embryogenesis[188], discovered by ChIP-chip against key transcription factors regulating DV patterning. Over the past decade, more genome-wide experiments have been conducted, and it has been established that shadow enhancers are prevalent in *Drosophila* and mice[189]. However, further research is needed to understand the evolution of shadow enhancers and the regulatory mechanisms involved in the cooperation within shadow enhancer clusters.

5.2. The regulatory models for enhancer cooperation

As introduced, the enhancer clusters have been named using various classifications, but it is still unclear how multiple enhancers function together despite being located thousands of base pairs apart in the genome. The classification mentioned above only provides simplified explanations of how enhancers interact and cooperate and does not predict the impact on gene transcription when one or two enhancers among the enhancer clusters are repressed. Researchers have developed methods such as Hi-C[169] to explore *in vivo* interactions among enhancers and CRISPRi[190] to repress enhancers *in vivo*. These methods have led to the discovery of more detailed regulatory models that describe the mechanisms of enhancers functioning in clusters. These regulatory models include additive, synergistic, competitive, and redundant regulation[191–194]. The synergistic model also covers both exponential and linear-logistic models by definition[192].

5.2.1. Additive model and synergistic model

Several studies have supported the additive model, including the ß-globin LCR mentioned earlier[47,171]. Another study focusing on α-globin super-enhancer dissected this super-enhancer using knock-out mouse models[195]. The study concluded that the α-globin super-enhancer also functioned in an additive manner. This model was later argued by Dukler *et al.* using biophysical modeling based on Boltzmann distribution[191]. By fitting the raw data from this study with three different model equations, it was shown that the activation models of α-globin super-enhancers prefer the biophysical-based linear-logistic model - a type of synergistic model.

Biophysical modeling may not always provide an accurate depiction of biological processes. Two recent experimental studies provided clearer perspectives of the regulatory models used by enhancer clusters and super-enhancers. These studies suggested that the enhancer clusters can regulate gene transcription in both additive and synergistic ways. For instance, one study by Lin *et al.* revisited the super-enhancer responsible for regulating Myc[196]. This super-enhancer is comprised of seven enhancers spanning a 1.9 Mb region. CRISPRi was employed in this study to repress enhancers without introducing deletion. The authors designed gRNA pools to cover all seven enhancers and perturbed these enhancers with two gRNA pairs simultaneously, targeting one or two enhancers in each sample. They then built a machine learning model based on the CRISPRi screening data and current knowledge about chromatin features in the cell lines used in this study. They found that the enhancers functioning synergistically were located far apart, while additive enhancers were within proximity. Further analysis also suggested that enhancer interactions and BRD4 occupancy determine the functioning of synergistic enhancer pairs. Applying JQ1, a BRD4 inhibitor, greatly reduced the synergistic cooperation within the MYC super-enhancer, further supporting their machine learning analysis predictions.

Rather than repressing enhancers by CRISPRi and evaluating the resulting gene expression changes, a study by Choi J. *et al.* focused on the dynamic gene expression changes regulated by enhancers[197]. The study used the transdifferentiation process as a model system and investigated the dynamic regulation of enhancer cooperation during human leukemia B-cells being transdifferentiated into macrophages. Throughout the process of transdifferentiation, multi-omics data was analyzed and profiled. The findings indicated that most paired enhancers functioned additively and target common genes, whereas a subset of enhancers targeting cell type-specific genes functioned synergistically. These enhancers were not categorized as super-enhancers, unlike MYC super-enhancers. The study by Lin *et al.* discussed above suggested that BRD4 might also associated with synergistic enhancers other than those found in MYC super-enhancers. This study confirmed that BRD4 occupied identified enhancer clusters. However, there were no significant differences in BRD4 occupancy between synergistic enhancers and additive enhancers genome-wide, indicating that the presence of BRD4 was not a feature of synergistic cooperation on its own, at least in certain contexts. Consistent with the definition of the super-enhancer, transcription factor binding sites enriched in synergistic enhancers were classified as cell-type specific transcription factors. These results suggested combining BRD4 presence and cell-specific transcription factor binding sites may be a feature for super-enhancers and other synergistic enhancer clusters.

The two examples employed distinct cellular systems and methods and proposed an intricate regulatory and cooperative model for enhancer clusters and super-enhancers. This model encompassed both additive and synergistic manners, and the regulatory mechanisms underlying synergistic mode were further explored. It

is worth noting that both studies used cancer cell lines, while BRD4 is a known drug target for many types of tumors[198]. Recently, it has been shown that in the context of individual enhancers, they have varied dependencies on cofactors[199]. It is unknown yet whether any other cofactors are involved in regulating enhancer synergistic cooperation.

Moreover, as discussed in Section 4, enhancers function dynamically during development and cellular processes, and enhancer clusters may also cooperate dynamically during development. For instance, a study that focused on enhancer clusters during *Drosophila* embryogenesis found that the enhancers of *Knirps* followed a synergistic model at earlier stages while switching to an additive model at later stages[193]. Therefore, each enhancer cluster may have unique ways of functioning together and employing different regulatory models over time.

5.2.2. Redundancy model and enhancer competition

The enhancer redundancy model is often discussed in the context of shadow enhancers. One study by Perry *et al.* found that deleting the shadow enhancer or the primary enhancer did not lead to any obvious developmental defects under normal experimental conditions[186]. A similar effect was also observed during the development of limb morphology in mice[200]. However, in both cases, a single enhancer proved insufficient when there was environmental stress or the reduced expression of development-associated genes. Since early studies on shadow enhancers mainly focused on a few key developmental enhancers, the questions remaining need to be further studied, including, first, how frequently the enhancer redundancy occurs, and second, the regulatory mechanism governs redundant enhancers.

Studies conducted on *Drosophila* and mice genomes have shed light on the frequency of shadow enhancers in the genome. In *Drosophila*, Kvon *et al.* characterized the spatial and temporal activity of enhancers that function during embryogenesis using *in situ* hybridization against reporters driven by these enhancers[201]. Subsequently, a study re-examined the data to identify redundant enhancers[189]. The conclusions drawn from this study include the following: 1. Redundant enhancers exist extensively in the genome; 2. Most redundant enhancers are partial redundancy, and they may also have activity in different spatial positions compared to the primary enhancers; as such, they are kept during evolution; 3. It is common for the same gene to have more than one redundant enhancer that regulates it. Similar conclusions were drawn from the study conducted on mice. After the discovery of shadow enhancers during limb morphology, Osterwalder *et al.* analyzed ChIP-seq against H3K27ac, RNA-seq from 29 different mouse tissues, as well as topologically associating domain (TAD) information[200]. The study found that redundant enhancers are common for genes that express dynamically during development but not for housekeeping genes. Thus, enhancer redundancy

is a common feature in both insect and mammalian development, particularly for developmental-related genes.

The regulatory mechanisms underlying redundant enhancers or shadow enhancers are not fully understood. Nevertheless, certain studies have provided some indications in this direction. Two studies[193,202] used an MS2-reporter system coupled with several known paired enhancers to quantify the enhancer activity changes during *Drosophila* embryogenesis. Both studies demonstrated that, for some enhancer pairs, the enhancer activity remained almost unchanged in time and space when the shadow enhancers were inactivated. Botham *et al.*[193] suggested that the activity of the *hunchback* primary enhancer is similar to the activity of both enhancers since they compete with one another when the concentration of Bicoid is high, with the more robust enhancer having a greater chance of interacting with the promoter. According to their proposed model, partial redundancy happens when the concentration of Bicoid drops. At low concentrations, the enhancer competition model shifts to the enhancer addition model, requiring both enhancers to achieve full expression. Another study by Scholes *et al.*[202] examined enhancer cooperation using *Kr* enhancers and also supported the competition model. However, the details of the relationship between enhancer strength and transcription factor concentration, such as at what transcription factor concentration the weak enhancer can share the regulatory function with the strong enhancer, are still unknown.

Activation of shadow enhancers can be affected by changes in the concentration of transcription factors, as well as increased environmental stress. In the context of environmental stress, such as increased temperature, when it comes to maintaining the same level of gene transcription regulation from one enhancer to two or more, how does this work? One possible answer may be that the enhancers are able to maintain a similar concentration of transcription factors in the transcription hub under different conditions. A study conducted by Tsai *et al.*[203] suggested that redundant enhancers are essential for maintaining the local transcription factor concentration at higher temperatures together with the primary enhancers, thus preserving the nuclear microenvironment. Recruiting different sets of transcription factors through the redundant enhancers may be another way to maintain the local transcription factor concentration to buffer the environmental stress. More studies are needed to illustrate the mechanism for regulating redundant enhancer cooperation.

5.2.3. Summary

Recent studies on shadow enhancers have revealed that shadow enhancers also employ the additive model in cases where there is only partial redundancy. This implies that the regulatory models for enhancer clusters, super-enhancers, and shadow enhancers may be mutually inclusive. In fact, one study found that

super-enhancers also act in a partial redundancy manner in embryonic stem cells[184]. Whether it is necessary to categorize enhancer clusters into different categories remains unresolved. Nevertheless, understanding how enhancers cooperate is essential for illustrating the gene transcriptional regulation network.

The mechanisms governing the cooperation among enhancers still need to be thoroughly discussed. Other mechanisms regulating enhancer cooperation, such as the formation of condensates among transcription factors, coactivators, and Mediator, may bring all the components from an enhancer cluster together. However, these mechanisms are yet to be reviewed. To understand enhancer regulatory mechanisms, it is crucial to understand how enhancers are regulated by chromatin remodelers, transcription factors, Mediator, and cofactors in time and space, and how they cooperate over time.

6. Bifunctional enhancers functioning as transcriptional silencers

6.1. The definition and characteristics of transcriptional silencers

In the previous section, we discussed enhancer redundancy and competition, which led to the emergence of another regulatory model suggesting that two paired enhancers may repress each other through unidentified mechanisms. For example, in *Drosophila*, the *Kr* enhancers, the enhancer pair CD1 and CD2 are redundant to drive a stripe expression pattern in the central embryo[184]. However, enhancer CD2 represses enhancer CD1 in the posterior embryo, independent of their relative positions, resulting in the absence of enhancer activity. This raises two questions: Does enhancer CD2 directly silence enhancer CD1 posteriorly? Is enhancer CD2 a bifunctional element with both activation and repression functions? When a *cis*-regulatory element, like enhancer CD2 in the posterior, silences another enhancer's activity, it is referred to as a silencer.

Silencers were discovered in yeast in 1985. They are similar to enhancers but with the opposite effect[204]. Silencers can repress gene expression, and their deletion can restore gene expression. The position or direction of the DNA element does not impact its silencing function, and flipping the element or switching the promoter to another one also does not affect its silencing effect. Silencers were defined as DNA elements repressing gene transcription regardless of their position and direction[204]. It is equally important to maintain gene silencing when a gene should not be expressed, as it is to turn on gene transcription at the right place and time. However, studies on silencers are far behind those on enhancers and promoters.

The information on silencers is limited, likely due to the confusing terminology used over the years. In a review published in 1998, silencers were classified as “transcriptional silencers” and Polycomb group complexes (PcG)-regulated ones [205]. However, this classification may need to be updated since Polycomb repressive complex 2 (PRC2) deposited H3K27me3 can be associated with “poised” enhancers and bivalent

primers rather than packing nucleosomes into heterochromatin[100]. In addition, repressors or corepressors can mediate the recruitment of PRC2[206,207], indicating that PRC2 is probably also involved in "transcriptional silencing."

Transcriptional silencing is a transcriptional repression process that prevents a gene from responding to environmental signals and stimuli[208]. This type of repression is achieved by recruiting nucleosome and DNA modifiers instead of through RNAi-mediated or local silencing via direct binding to promoters[208]. The 7th edition of the Molecular Biology of the Gene defines transcriptional silencing in this way to encompass all silencers, regardless of their specific regulatory mechanisms[208]. This includes silencers uncovered by most studies published recently using different criteria for screening. This section will discuss literature studying transcriptional silencers that follow this definition.

Additionally, silencers have yet to be studied genome-wide due to the difficulty of identifying them in NGS-based multi-omics data. This is because they are associated with diverse histone marks, making it challenging to find signature histone marks associated to them[209]. So far, the most commonly used histone mark for silencers is H3K27me3[210–213], though its specific association with silencers is debatable since it is also a mark for poised enhancers. However, a new approach uses deep learning methods for predicting silencers. A recent study conducted training on a Convolutional Neural Network (CNN) model, using the DNA sequences and certain chromatin marks across various cell types[214]. This model was validated with published MAPR data. The researchers applied this CNN model and successfully predicted silencers that were previously discovered by two independent screening experiments in K652 cells. This was achieved despite the limited overlap between the two studies [211,213]. Interestingly, the most enriched histone mark in these studies was not H3K27me3, although it contributed significantly to the dataset used for CNN training. In the future, combining deep learning techniques with NGS experiments could help construct interpretable models, enhancing our understanding of silencer features.

Despite the varied signatures among silencers, at least two chromatin or sequence features are shown to be shared among silencers, including chromatin accessibility and repressor binding sites[211,212,214,215]. While heterochromatin is often associated with preventing accessibility to transcription factors and dominating gene silencing, this is not the case with transcriptional silencing. Transcription factors and chromatin modifiers are involved in regulating the process, and accessible DNA elements are necessary for transcription factors to bind. As a result, most silencers identified so far were discovered mainly through measuring chromatin accessibility using DNase I-hypersensitivity assays or ATAC-seq. Some silencers identified through other methods were also shown to have accessibility in subsequent experiments[216,217]. While the chromatin features of silencers vary depending on specific screening criteria, accessibility

remains a consistent feature among all identified silencers[209]. Therefore, accessibility is a common feature of both silencers and enhancers.

Other than the chromatin accessibility, another shared feature among silencers based on genome-wide analysis is the presence of motifs for repressors[214]. Early studies on silencers led to the discovery of several repressors, which are now widely discussed in literature. These silencers were identified by examining *cis*-regulatory elements that repressed gene expression in different cellular processes. For example, REST-regulated silencer RE1 repressed gene expression in specific neuronal cells[218–223], while RUNX 1/3-regulated silencer S4 repressed CD4 expression during certain stages of T-cell lineage differentiation[224,225]. This feature also holds for bifunctional silencers, which act as enhancers in other tissues or developmental stages. The study of *Drosophila* embryogenesis in the early 1990s led to the discovery of silencers that functioned as enhancers in different tissues and were targeted by the factors Krüppel (Kr) and Hunchback (Hb) when they were silencers[226]. The recruitment of repressors is thus a key step for silencers to function.

To sum up, the recent findings about the genome-wide features of silencers support the regulatory mechanisms characterized in the early days on a single locus; much like enhancers, silencers function by modulating chromatin accessibility and recruiting transcription factors to regulate gene expression. Moreover, recent findings also indicated that it is expected to observe silencers functioning as enhancers in other tissues or different developmental stages[214,215,227]. However, the detailed mechanism of how these bifunctional silencers or enhancers are regulated in different contexts is not fully understood. The following section will primarily focus on bifunctional silencers or enhancers.

6.2. The regulatory mechanisms of silencers and bifunctional enhancers

6.2.1. Bifunctional silencing regulated by repressors

One of the main characteristics of transcriptional silencers is the enrichment of repressor binding motifs, indicating the important role of repressors in mediating silencer activity. Studies conducted on *Drosophila* embryos have revealed the mechanisms of gene repression during embryogenesis, resulting in precise gene expression patterns. In 1993, a study showed that elements responsible for driving specific gene expression and bound by repressors could function as silencers, as they could repress the activity of enhancers and promoters targeting different genes[226]. The study further identified a silencer of *Abd-B*, which acted as an enhancer driving reporter gene expression pattern similar to that of *Abd-B* at the same stage. This enhancer activity was switched to silencing in the anterior embryo, where the repressor Kr is expressed. The study also characterized another two silencers regulating *Abd-B* expression *via* Hb and Polycomb. All three fragments tested showed enhancer activity in the region where the repressors were not expressed. These results suggested that *cis*-regulatory elements can be bifunctional in regulating gene expression during

developmental patterning. The example of the enhancer CD2 in regulating *Kr* expression discussed earlier also indicates bifunctionality in a well-characterized mode of gene regulation. Recently, the concept of bifunctional regulatory elements has been revisited, and it has been observed that switching the function between enhancer and silencer happens not only in patterning body plans but also in other contexts, which will be discussed later.

6.2.2. Bifunctional silencing regulated by Polycomb-group complexes

PcG has been long thought to be involved in gene silencing by depositing H3K27me3[228–230]. In *Drosophila*, PcG is recruited to a type of DNA element called Polycomb response elements (PREs) by recognizing DNA motifs by one of its complexes, the Pho-repressive complex (PhoRC)[231]. PhoRC contains transcription factors Pho and dSfmbt, which bind to PREs and recruit PRC1 there[230,231]. To better understand the relationship between PREs and gene transcription during *Drosophila* embryogenesis, a study by Ereg *et al.* profiled PhoRC-bound elements genome-wide and assessed their function in gene transcription[227]. They first expanded the pool of PREs by overlapping ChIP-seq profiles against Pho and dSfmbt, then analyzed these PREs with known developmental enhancers and profiles of H3K27me3, H3K27ac, and RNA-seq. Their findings showed that 25% of these PREs are known developmental enhancers that function in other developmental contexts. These regions were marked with H3K27me3 and chromatin accessibility while depleting H3K27ac and RNA transcripts. The chromatin features of these bifunctional PREs, chromatin accessibility, H3K27me3 mark, and depletion of H3K27ac were further supported by the following studies on silencers. Interestingly, these marks also feature poised enhancers, referring to the inactive enhancers that will be activated later during development. These results suggest that PREs-regulated transcriptional silencing elements also exhibit dynamic changes between enhancer and silencer, similar to other bifunctional silencers characterized by repressor binding.

6.2.3. The regulation of the switch between bifunctional silencers and enhancers

The mechanisms that govern the transition between silencers and enhancers remained elusive until recent research focused on genome-wide screening of silencers. About ten studies have used different screening methods to search for silencers in the past few years[210–217,232,233]. These studies have revealed that some of the silencers identified from screening can also act as enhancers in other contexts. However, the proportion of these bifunctional silencers among all silencers identified in the same study is inconsistent. Besides the study above on PRE, two other studies investigated these bifunctional silencers.

In one study, Gisselbrecht *et al.* aimed to screen transcriptional silencers in mesoderm during *Drosophila* embryogenesis[215]. They marked the cells from mesoderm with a CD2+ marker, which is normally not expressed in *Drosophila*, and designed constructs specifically for the silencer screen. These constructs were

fused with candidate silencers 5' of a universal enhancer and followed by a GFP tag. They tested various candidate elements, including known enhancers not functioning in the mesoderm but in other tissues, regions with both active and repressive histone marks, regions bound by co-repressors, and known silencers. However, only the known enhancers yielded some silencing effects for further analysis. Among the tested enhancers, only 10% showed a silencing effect. They also found that the repressor Snail regulated 41% of the silencers tested. A 4C analysis and Hi-C data supported the previous finding that the binding of Snail prevented the adjacent enhancers from contacting their promoters. The study concluded that anti-looping by Snail and looping to the promoter by silencers were the two mechanisms that silenced gene transcription, not silencers looping to enhancers. In total, the study tested 29 silencers.

In recent years, combining machine learning and deep learning techniques with multi-omics data has provided new insights into understanding transcription factor-DNA binding and enhancer syntax. In the study briefly discussed above, Huang *et al.* conducted a study on bifunctional silencers in the context of cell lineage differentiation, revealing the potential to uncover these silencers' regulatory syntax[214]. Using a CNN model, the researchers predicted silencers during T-cell lineage differentiation. They then focused on a specific subset of silencers in T cells that were enhancers in H1 hESCs, accounting for 6% of all silencers in T cells. They also verified that these silencers, termed dual-function regulatory elements (DFREs), targeted the same genes in H1 hESC, where they functioned as enhancers. Subsequent analysis revealed that the binding preference of transcription factors changed during the transition from enhancers to silencers, indicating that these DFREs could recruit different sets of transcription factors to regulate their functions. Furthermore, the study showed that the high-affinity binding sites of these transcription factors and their approximal regions did not overlap. This suggested that the transcription factor binding sites formed clusters for their enhancer or silencer functions, respectively, and that these two clusters may not overlap.

Taken together, the bifunctional silencers likely harbor both activator and repressor binding sites and switch functions by recruiting distinct sets of transcription factors. The expression of these transcription factors can be regulated through developmental and other cellular signals. The recruited transcription factors trigger changes in chromatin marks, modified by cofactors and chromatin modifiers, further affecting the enhancer-promoter looping. The cooperative process can be modulated through binding between PRE and PcG or other repressor binding sites and corresponding cofactors. Although the DNA sequence instructs these dynamic processes, it remains unclear how enhancers switch to silencer mode, or *vice versa*, by cooperating with transcription factors, cofactors, and chromatin remodelers and modifiers. Analyzing multi-omics data from different stages of cell lineage differentiation using deep learning models is promising for offering more insights into the regulation of bifunctional enhancers.

6.2.4. Summary

Silencers, discovered and defined nearly four decades ago, have recently caught attention again. Bifunctional silencers are not uncommon and have been identified through large-scale silencer screen studies. These elements can function either as enhancers or as silencers, depending on the developmental context. Moreover, the idea that silencers are accessible when functioning is supported by large-scale silencer screens. Presumably, in this way, enhancers can be turned into silencers once different sets of transcription factors bind them during development and differentiation. Such a switch may allow for transient silencing of targeted gene transcription without remodeling euchromatin into heterochromatin. More questions remain to be answered to understand silencers and bifunctional silencers further. For instance, it is unclear whether bifunctional elements function in the context of enhancer clusters or super-enhancers. Additionally, it is unclear whether we can identify a unique combination of chromatin features that specifically mark silencers in the same way promoters and enhancers are specifically marked.

6.3. The effect of sequence variation on bifunctional enhancers

Recent studies on silencers have shown a correlation between silencers and disease-related SNPs[209], similar to studies on the effect of sequence variation on chromatin accessibility dynamics. However, the specific mechanisms by which these variants may alter the regulatory functions of elements in silencers or bifunctional silencers remain unclear.

A recent study investigated the changes in chromatin 3D structure in acute myeloid leukemia (AML)[234], revealing that newly formed silencer-promoter loops in patients' samples also play a role in the disease in addition to the aberrantly formed enhancer-promoter loops. CRISPR-based tests further confirmed the effect on gene expression from these silencer-promoter loops in a cell line derived from an AML patient. Although this study did not investigate further how the mutations in AML patients regulate the genomic structural changes to form new silencer-promoter loops, it suggested that their counterparts, enhancers, should not overshadow the importance of silencers.

Studies focusing on evolutionary sequence variation in silencers are limited. However, some studies have characterized the gain of repressor binding sites contributing to trait loss. For instance, the fly *D. sechellia* is a closely related species to *D. melanogaster*, whose larvae have lost specific patches of cuticular hairs called trichomes. This loss is the result of an evolutionary change, the loss of expression of *shavenbaby*, a transcription factor governing trichome development, in so-called quaternary cells[235]. This loss led to the naked cuticle formation without quaternary trichomes in the first-instar larvae[147,235]. The reduction in *E6* enhancer activity, which regulates *shavenbaby* expression in these cells, plays a significant role in silencing *shavenbaby* expression, specifically in *D. sechellia*[147].

A study conducted by Preger-Ben Noon *et al.* revealed that the loss of activator binding sites in *E6* and the gain of a repressor binding site located 100-bp upstream of *E6B*, the core enhancer of *E6*, led to repression of the *D. sechellia*'s *E6B* activity[236]. Placing this repressive element upstream of another *shavenbaby* enhancer, *E3N*, also suppressed the enhancer function. Although the original work did not discuss it, this evolved repressive element is likely an evolved silencer that silences the *E6B* activity in *D. sechellia*, contributing to the naked cuticle in this species. The study further supported the insight on the conversion between enhancers and silencers: the sequences encoded regulatory potential for both enhancers and silencers; and the switch from enhancer to silencer function depends on the expression of transcription factors present in each cell.

Identifying potential repressor binding sites by perturbing the sequence of an enhancer can be difficult. This approach may also create new transcription factor binding sites, making it more challenging to interpret the results accurately. However, recently developed methods such as ATAC-seq, CUT&RUN[237], and CUT&Tag[238] enable chromatin accessibility, histone marks, and transcription factor-DNA binding to be measured using minimal experimental material. By combining these methods with machine learning and deep learning analysis pipelines, it may soon be possible to characterize sequence-based features of silencers and, as such, facilitate the discovery of evolutionary silencers and disease-related silencers.

7. *Drosophila* wing pigmentation – A model system used in this study

7.1. The evolution of wing pigmentation pattern in *Drosophila*

As discussed in previous sections, changes in DNA sequences can affect the function of enhancers by altering the binding sites of transcription factors. This, in turn, can change the accessibility and activity of enhancers and influence their interactions with other enhancers and their ability to silence genes. However, unlike genes that code for proteins, we do not yet fully understand the organization of enhancer sequences, also known as enhancer grammar or syntax[138]. This makes it challenging to predict enhancer functions based solely on their sequences.

Understanding the determinants of enhancer functions can be facilitated by exploring natural variation in evolutionary morphological traits. These enhancers share functional elements regulating gene expression during development, but some evolved to achieve spatial specificity. For instance, in the case of trichome pattern evolution in fruit flies, conserved binding sites for activators were functioning, while a new binding site for a repressor was gained in one species[236]. As a result, trichomes were lost in the area where the repressor was expressed.

The wing pigmentation pattern of fruit flies is another well-studied morphological trait in evolutionary biology[14,239,240]. The gene *yellow* is responsible for the dark pigmentation in adult fruit flies.[241] It encodes Yellow, a protein involved in melanin synthesis[241]. However, it is still unknown whether it functions as an enzyme to promote melanin synthesis or as a ligand protein to activate related signaling pathways[242]. Nonetheless, previous studies revealed that *yellow* expression patterns in pupal wings prefigure the corresponding black pigmentation patterns in adult wings [14]. Enhancer-reporter assays have also shown that the 5' sequence upstream of the *yellow* locus in spotted species drives reporter gene expression in a spot-like pattern, while non-spotted species do not exhibit this pattern[14,240]. These results indicate that the variation in *yellow* enhancers underlies the evolution of the pigmentation pattern in fruit fly wings.

Previous studies have shown that *yellow* enhancers from the *D. melanogaster* group can be divided into two fragments: the "*spot*" enhancer, which drives specific expression at the anterior distal tip of the wing, and the "*wing large*" enhancer, which drives uniform expression across pupal wings[14,240]. In a spotted species, *D. elegans*, the *spot* enhancer sequence is similar to that of its closely related species, *D. gunungcola*, which lacks a dark pigmentation pattern on male wings. By comparing the *spot* enhancer sequences between these two species, Prud'homme *et al.* found that replacing several base pairs in *D. gunungcola* with the homologous sequence from *D. elegans* resulted in the modified *D. gunungcola* enhancer being able to drive a spot-like pattern[240]. This result indicated that the mutations that evolved in the gained *spot* enhancer contributed to the secondary loss of the spot trait in this group.

The *yellow* enhancer from a species closely related to *D. melanogaster*, *D. biarmipes* (Figure 5a), has been studied to understand how this enhancer controls the specific expression of *yellow* in space and how sequence variation affects its function[14,243–245]. The dissection of enhancer *spot*^{bia} showed that a minimal enhancer with 196 bp could drive reporter gene expression in a spot-like pattern in the pupal wing[14]. At least two transcription factors, Distal-less (Dll) and Engrailed (En) regulate *spot*^{bia196} enhancer activity[14]. Dll is a homeodomain transcription factor activating *spot* enhancer in the distal region of the wing (Figure 5d)[243,246,247]. En is a homeodomain repressor repressing *spot* enhancer activity in the posterior part of the wing (Figure 5d)[14,248]. While some of the binding sites for Dll and En are known[14,243,245], the evolution of the *spot* enhancer is still unclear. For example, a comparison of the *yellow* 5' sequence from spotted and unspotted species in the *D. melanogaster* group showed that a key Dll binding site[243] located in *spot*^{bia196} was conserved in spotted species and even in *D. melanogaster*, which has no spot pattern (Figure 5b). Given that all enhancer-reporter assays discussed so far were performed in *D. melanogaster*, these findings raised the question of how *D. melanogaster* lost its *spot* enhancer despite preserving a highly conserved Dll binding site and having the necessary transcription factor landscape to drive the spot-like expression. Unlike the evolutionary loss of potential activator binding sites in *D. gunungcola*, resulting in the loss of the *spot*

activity, the *spot* orthologous region in *D. melanogaster* might employ a more complex path to achieve the secondary loss of *spot* activity.

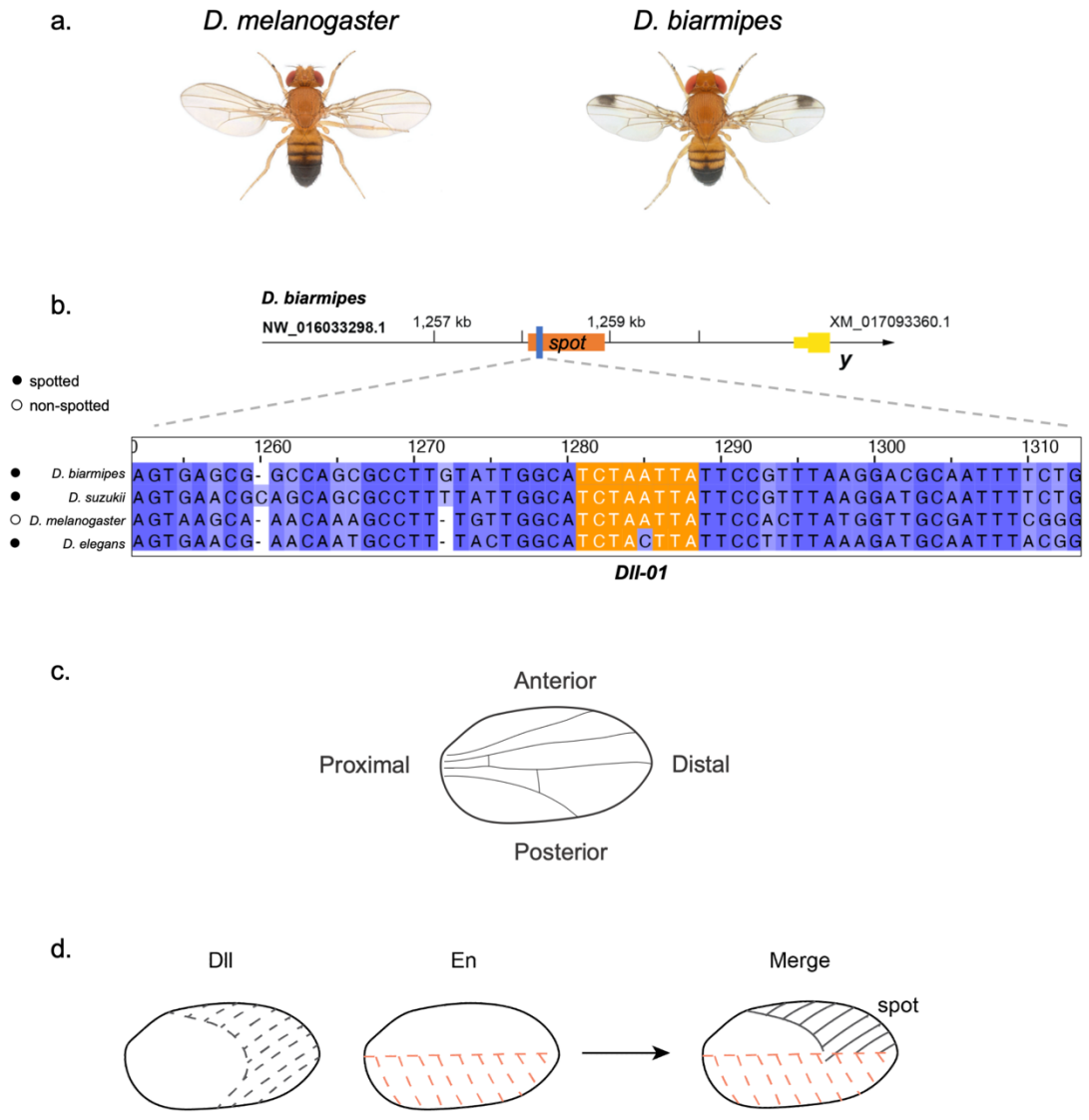

Figure 5. The wing spot pigmentation pattern in Drosophila melanogaster. *(a.) The male fly from D. biarmipes presents a dark spot pigmentation pattern in the wings, while the one from D. melanogaster does not have this pattern. (Images are from Nicolas Gompel.) (b.) The multiple sequence alignment at the yellow 5' locus shows that a strong Dll binding site, Dll-01, is conserved among several spotted species and D. melanogaster. (c.) A sketch of the Drosophila wing. (d.) The sketch illustrates the expression patterns of Dll and En in the wings.*

The enhancers harbor information for recruiting transcription factors, including activators, repressors, and others regulating chromatin accessibility. As discussed above, activating enhancers requires multi-layer regulators to function in coordination. Meanwhile, it is worth noting that Dll, which is necessary to activate *spot*^{bia}, is also expressed in wing discs in the third instar larvae. However, Dll alone cannot activate *yellow* expression in wing discs. This suggests other transcription factors regulate the *yellow* enhancer function together with Dll.

Furthermore, the reporter gene driven by the *yellow*^{bia} enhancer exhibited a whole-wing pattern first in pupal wings, followed by a spot-like pattern in late pupal wings[244]. This dynamic reporter expression pattern indicated that the *yellow* expression dynamics may differ in spotted and non-spotted species, and so do the enhancer activity dynamics. Can the chromatin accessibility reflect the *yellow* enhancer activity dynamics? What other transcription factors are involved in regulating the enhancer activity with Dll? Does the enhancer sequence variation among species alter the *yellow* enhancer accessibility dynamics by creating or mutating transcription factor binding sites? Answering these questions is challenging but advantageous for understanding enhancer syntax.

7.2. Gene expression dynamics during wing development

The expression of *yellow* is known to be dynamic during development[249], and the regulation of this expression during pupal wing development is still not fully understood. Previous studies on *yellow* enhancers in pupal wings have mainly focused on late-stage pupal wings or newly emerged adult wings. Understanding how gene transcriptional regulation is controlled during pupal wing development might provide further insight into how the sequence information of the *yellow* enhancer is organized.

During pupal development, fruit flies undergo metamorphosis to transition from larvae to adults. At 25°C, male *D. melanogaster* usually takes 96 to 100 hours to develop from white pupae to newly emerged adults. The first 12-14 hours after puparium formation (APF) are defined as the prepupal stage when head eversion takes place[250]. Pupal wings, originating from larval wing imaginal discs, undergo morphological changes throughout pupal development (Figure 6)[251]. Precise control of gene expression in space and time during the entire pupal stage is necessary for proper wing development. The tempo of this regulation is overall determined by hormonal signals.

The steroid hormone ecdysone regulates metamorphosis in fruit flies[252]. Ecdysone binds to the ecdysone receptor (EcR) and activates transcription factors that sequentially regulate the expression of more transcription factors[252]. The onset of *yellow* expression in pupal wings is likely controlled by ecdysone pulses. A study conducted by Sobala *et al.*, using RNA-seq on pupal wings from 42h APF to 96h APF,

revealed the dynamic gene expression profiles[253]. The expression level of *yellow* in *D. melanogaster* spiked in 52h APF pupal wings and then decreased to a low expression level at later stages. The onset of *yellow* expression may follow one of the ecdysone pulses, which takes place at around 30-40h APF[252,254,255]. This ecdysone pulse controls pupal development after the initial ecdysone pulse for pupation at the end of the third instar larval stage.

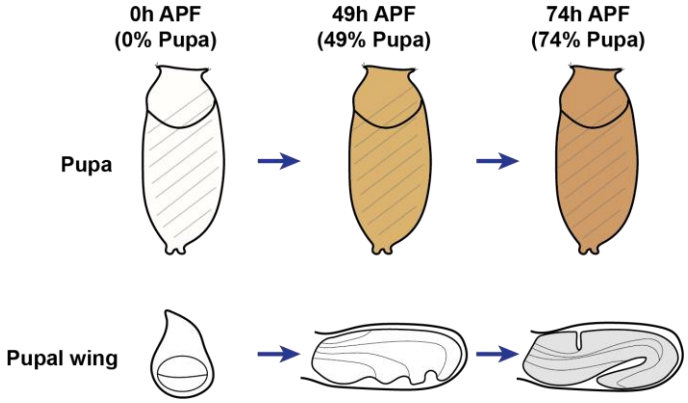

Figure 6. The morphology of the pupal wing changes during pupal development. *The schematic shows examples of pupae at different stages and their corresponding pupal wings. These stages include the prepupa, mid-stage pupa, and late-stage pupa. The descriptions of developmental stages in this dissertation are referred to in either APF or percentage of pupal development. The conversion between the two metrics has been done here to be interchangeable. The pupal wing figures were drawn based on M. C. Diaz de la et al [251].*

Eip93F (E93) is a transcription factor that responds directly to ecdysone pulses[256]. It is essential for puparium formation and adult development[257,258]. In particular, in the development of wings, research has shown that the absence of E93 significantly reduces the size of pupal wings[259]. Uyehara *et al.* conducted a study investigating how E93 functions in gene regulation by performing FAIRE-seq in wing discs, prepupal, and early pupal wings from normal or E93-mutated strains[260]. The results showed that depletion of E93 resulted in either gain or loss of accessibility from hundreds to several thousands of enhancers during wing development. Another study focusing on salivary glands supported the dual-functional regulation of E93[258]: Lam *et al.* generated E93 knock-out fly strains and performed RNA-seq in prepupae, illustrating that E93 can upregulate and downregulate gene expression in salivary glands.

In another study by Nystrom *et al.*, E93 was overexpressed in wing discs, a stage when E93 is not normally expressed, to determine if it could promote ectopic accessibility of enhancers[261]. The results showed that while E93 was needed, it was not sufficient to promote the accessibility of all enhancers. The 5' locus of

yellow was one of the loci where accessibility was significantly decreased in the E93 mutant strain during early pupal development. Still, overexpression of E93 in wing discs could not promote accessibility. These findings suggest that E93 works in coordination with other transcription factors to regulate the accessibility of the *yellow* enhancer during wing development. According to the RNA-seq data[253], E93 expression decreased after the ecdysone pulses, but increased in late pupal wings from 62h to 96h APF.

Interestingly, the reporter gene driven by the $spot^{bia}$ starts to show a spot-like pattern in 65-70h APF pupae. Therefore, I reasoned that the increased expression of E93 in late pupal wings may be linked to the activation of $spot^{bia}$ enhancers. Further study on the relationship between E93 and accessibility changes in *yellow* enhancers in late pupal wings may offer new insights into *yellow* enhancer dynamics.

Moreover, it is worth noting that all the studies on gene transcription dynamics during wing development were conducted in *D. melanogaster* species. However, whether other spotted species within the *D. melanogaster* group share similar gene expression profiles remains unclear. It is reasonable to assume that changes in enhancer sequences have led to the gain or loss of transcription factor binding sites involved in regulating chromatin accessibility, altering the enhancer dynamics and function. To investigate the evolution of wing pigmentation patterns, it is necessary to understand the regulation of *yellow* enhancer dynamics across different species. This includes understanding the changes in the accessibility of *yellow* enhancers during development and the transcription factors responsible for regulating such dynamic changes.

7.3. Methods to study enhancer regulatory mechanisms in pupal wings

In recent years, NGS-based techniques have revolutionized genome-wide sequencing, making it more affordable and accessible[65]. With NGS, it is now possible to measure chromatin properties throughout the entire genome with a single experiment, whereas previously, only a single locus could be studied. Several techniques can be used to examine chromatin dynamic changes in enhancers. These include DNase-seq, MNase-seq, and FAIRE-seq to measure chromatin accessibility[262] and ChIP-seq to study histone marks and transcription factor-DNA binding. These cutting-edge techniques and robust data analysis tools present new opportunities for studying gene transcriptional regulation. However, these NGS-based techniques require significant input cells, equivalent to tens or even over a hundred pupal wings, making them challenging to apply to developmental tissues.

More recent techniques based on the transposes Tn5 allow a much lower amount of input cells (500-50000) and are time-efficient[68]. These techniques provide opportunities to study chromatin changes in tissues that are difficult to dissect and have fewer cells. Two such methods are ATAC-seq, which profiles chromatin

accessibility[68], and CUT&Tag, which profiles chromatin marks deposition and transcription factor binding[238]. Another similar method is CUT&RUN, which uses MNase instead of Tn5 and requires fewer cells and less time to perform the experiment than ChIP-seq[263].

While these NGS-based techniques discussed in this introduction have proven successful in cell lines and tissues, they have yet to be as commonly used in the case of pupal wings, especially in later stages when the *spot* enhancer is activated. This is mainly because the cuticle structure of late pupal wings protects the cells. Also, there is an expansion of cell volume compared to those in the prepupal stage. There are also limited cells, numbering around 20,000 per wing (Gompel lab unpublished data). As a result, collecting enough intact wing cells from individuals at the same stage and processing them using NGS-based techniques is challenging. However, with the development of ATAC-seq and CUT&RUN techniques, it is now possible to explore the chromatin dynamics in these late pupal wings.

In this dissertation, I applied ATAC-seq, CUT&RUN, and ChIP-seq to pupal wings by modifying protocols from published resources[68,264–266]. I also profiled chromatin accessibility changes in pupal wings dissected from 0h APF pupae to 78h APF pupae. The results offer insights into the chromatin dynamic changes in a highly differentiated tissue. Moreover, comparing the chromatin accessibility changes between *D. melanogaster* and *D. biarmipes* has led to new assumptions about how the morphological trait was lost during evolution.

Aim of the study

This dissertation aims to understand how enhancer sequence variation affects targeted gene transcription and how it influenced the evolution of morphological traits. To study this, I employed the wing pigmentation pattern in *Drosophila* as a model system. In the *D. melanogaster* group, the formation of this trait depends, among others, on the pigmentation-related gene *yellow*, necessary for the formation of black pigments. Therefore, in this dissertation, I mainly focused on understanding how the *yellow* enhancer variation regulates its expression pattern in the wings.

In a main project, I investigated through which mechanisms *yellow spot* activity was lost from *D. melanogaster* during evolution. For this purpose, I profiled and analyzed the chromatin accessibility during pupal wing development and carefully dissected *yellow* enhancer sequences from spotted species, *D. biarmipes*, and non-spotted species, *D. melanogaster*. In addition, I have also investigated the transcription factors that may be involved in regulating differentiated chromatin accessibility and repressing the *yellow* enhancer activity in *D. melanogaster*. The ultimate goal of this study is to identify the determinants of the *yellow* enhancer in *D. melanogaster* that led to the loss of the spot pattern.

Using the same experimental system, I contributed to two additional projects related to enhancer function. The first project aimed to investigate the potential boundary between the recently evolved *spot* enhancer and the ancestral part of the *yellow* enhancer in *D. biarmipes*. Additionally, the project sought to explore the interplay between these two enhancers, given their proximity. In my doctoral project, I examined the role of a particular element in the *spot* enhancer involved in regulating the chromatin accessibility of both enhancers.

The second project aimed to understand the regulatory syntax underlying the *spot* enhancer activity in *D. biarmipes*. Previous studies have identified two transcription factors that regulate the *spot* enhancer activity. The other transcription factors responsible for regulating the enhancer activity beyond these two remain unidentified. The mechanisms through which these factors modulate enhancer activity are also elusive. As part of this dissertation, I worked with my colleagues to introduce mutations into the $spot^{bia196}$ enhancer and examine the consequences of these mutations.

In conclusion, the goal of this study is to further our understanding of how complex organisms evolve by examining how transcriptional enhancers function. A particular emphasis was given to the interplay between the accessibility of the enhancer and its regulatory function.

Results

Paper I. Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*.

Liucong Ling, Bettina Mühling, Rita Jaenichen, Nicolas Gompel.

SCIENCE ADVANCES | Vol 9, Issue 7 | 7 Feb 2023

DOI: 10.1126/sciadv.ade6529

Citation: Ling, L., Mühling, B., Jaenichen, R. & Gompel, N. Increased chromatin accessibility promotes the evolution of a transcriptional silencer in Drosophila. *Sci Adv* **9**, (2023).

SCIENCE ADVANCES | RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*

Liucong Ling, Bettina Mühling, Rita Jaenichen, Nicolas Gompel*

The loss of discrete morphological traits, the most common evolutionary transition, is typically driven by changes in developmental gene expression. Mutations accumulating in regulatory elements of these genes can disrupt DNA binding sites for transcription factors patterning their spatial expression, or delete entire enhancers. Regulatory elements, however, may be silenced through changes in chromatin accessibility or the emergence of repressive elements. Here, we show that increased chromatin accessibility at the gene *yellow*, combined with the gain of a repressor site, underlies the loss of a wing spot pigmentation pattern in a *Drosophila* species. The gain of accessibility of this repressive element is regulated by E93, a transcription factor governing the progress of metamorphosis. This convoluted evolutionary scenario contrasts with the parsimonious mutational paths generally envisioned and often documented for morphological losses. It illustrates how evolutionary changes in chromatin accessibility may directly contribute to morphological diversification.

INTRODUCTION

The diversity of morphological characters in plants and animals breaks down into gains, losses, and modulation of traits. From a developmental standpoint, morphological characters of adult organisms are typically foreshadowed by discrete patterns of gene expression, driven by specific transcriptional enhancers during embryogenesis or postembryonic development (*1*–*4*). From an evolutionary standpoint, changes in these *cis*-regulatory elements drive changes in gene expression and thereby produce morphological variation (*5*). Consequently, the loss of a discrete trait, the most common evolutionary transition (*6*), often stems from the loss of function or deletion of enhancers underlying its formation (*3*, *7*–*9*). The literature on enhancers and their evolution emphasizes their spatial activity and how transcription factor binding sites (TFBSs) for spatial regulators encode patterns (*10*–*15*). Recently, another aspect of enhancer biology has come under scrutiny, their accessibility (*16*, *17*), reflecting the degree of local nucleosome occupancy and chromatin compaction. DNA accessibility of an enhancer is a prerequisite to the action of transcription factors (TFs) acting as spatial regulators (*18*). It is controlled during development by different parameters, including the action of pioneer TFs (*19*) and transcriptional activity per se (*20*, *21*), and was shown to vary between species, along with TF binding (*22*). As such, the modulation of enhancer accessibility represents an obvious level of transcriptional regulation that could contribute to morphological divergence between species. We explored this possibility using a well-defined transcriptional enhancer associated with morphological evolution in *Drosophila*, the *spot* enhancer of the pigmentation gene *yellow* (*8*, *23*–*26*). Several *Drosophila* species have patterns of black pigmentation on their wings, as, for instance, the wing-spotted fly *Drosophila biarmipes* (Fig. 1A). This character was gained once in a common ancestor of a species group containing *D. biarmipes* and the model species *Drosophila melanogaster*. It was subsequently lost several times in this group, including in *D. melanogaster* itself (Fig. 1B) (*8*). The initial gain of a wing spot involved the co-option of *yellow*, a gene necessary for the production of black pigments, through the newly evolved *spot* regulatory activity (Fig. 1C) (*23*, *24*). In at least one species from this group, *Drosophila gunungcola*, that lost its wing spot, the *spot* activity decayed through point mutations (*8*). Focusing on another species that has secondarily lost its wing spot (*8*), the model organism *D. melanogaster* itself, we investigated the evolution of accessibility of a regulatory region that has lost its enhancer activity.

Chair of Evolutionary Ecology, Ludwig-Maximilians Universität München, Faculty of Biology, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.
*Corresponding author. Email: gompel@bio.lmu.de

RESULTS

Enhancer accessibility and enhancer activity show discrepancy at the *yellow* locus of *D. melanogaster*

To understand whether enhancer accessibility might also play a role in the loss of *spot* activity, we compared chromatin states at the *yellow* locus during wing pupal development between the spotted *D. biarmipes* and the nonspotted *D. melanogaster*. We first monitored the dynamic of *yellow* expression in developing wings of both species with quantitative polymerase chain reaction (qPCR) (fig. S1A) to facilitate the interpretation of accessibility. The two species showed marked differences in the expression dynamic, where *yellow* transcripts faded in *D. melanogaster* after 56% pupal development, while they slightly increased in *D. biarmipes*, in line with the *spot* activity (*23*). We then profiled accessibility upstream of *yellow* transcription at selected stages in the wings of both species using assay of transposase-accessible chromatin using sequencing (ATAC-seq) (Fig. 1, E and F, and fig. S1A) (*27*). At the onset of *yellow* expression (47% of pupal development), accessibility increased in both species with two adjacent peaks (Fig. 1F, peak-1 and peak-2), which tended to merge and fade at later stages. We were intrigued by two observations. First, the region of *D. melanogaster* orthologous to *D. biarmipes spot* enhancer (peak-1) (*26*) showed persistent accessibility throughout pupal development (Fig. 1F and fig. S1), although we detected no *spot* activity in *D. melanogaster* in a reporter assay ($y5'^{mel}$; Fig. 1, D and E). We initially interpreted this as residual accessibility but found this unexpected, given that the wing spot was lost at least 10 million years ago in the

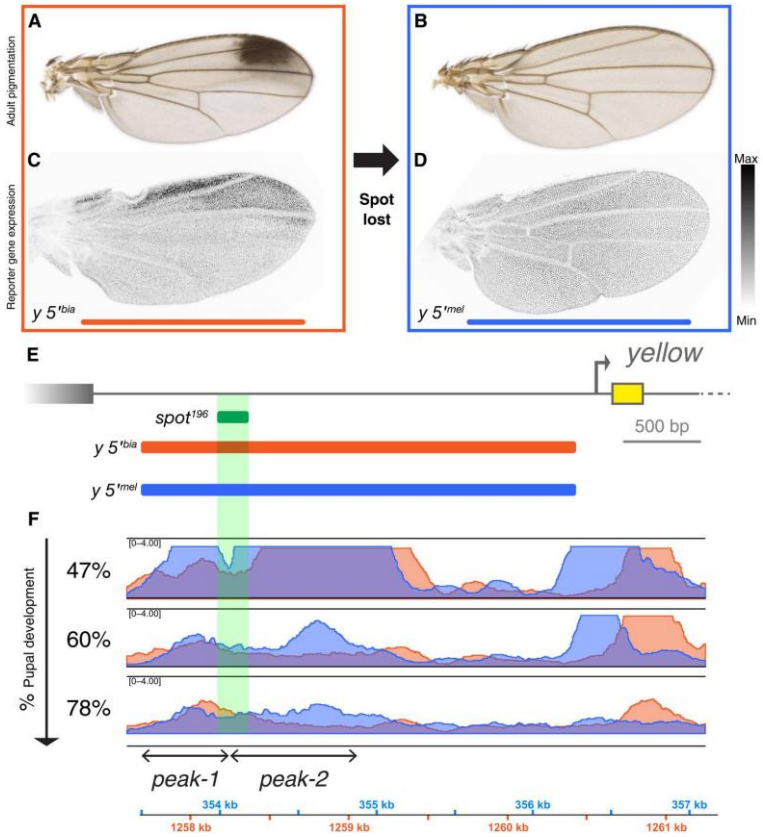

Fig. 1. Changes in accessibility in the regulatory regions of *yellow* associated with a loss in wing pigmentation. (**A** and **B**) Adult wings of *D. biarmipes* (A) and *D. melanogaster* (B). (**C** and **D**) Regulatory regions upstream of *yellow* transcription start site of *D. biarmipes* (C) and *D. melanogaster* (D) drive reporter expression in the wings of transgenic *D. melanogaster* at 80% of pupal development. The reporter activity levels foreshadow adult pigmentation of the respective species (see grayscale for fluorescence levels in all figures). (**E**) A map of the *yellow* locus indicates the relative positions of the fragments tested in (C) and (D) as well as the previously characterized *D. biarmipes* minimal enhancer *spot*^{196} (*23*). bp, base pair. (**F**) Assay of transposase accessible chromatin sequencing (ATAC-seq) tracks of the corresponding *yellow* region from pupal wings at different stages (expressed in percentage of the pupal life duration). Note the presence of two peaks of accessibility, best separated at 47% of pupal development, a stage that we used to define the sequence coordinates of these peaks by visual inspection. Here and in all subsequent figures, the blue color denotes *D. melanogaster* and the orange color denotes *D. biarmipes*.

D. melanogaster lineage (*8*), leaving enough time to this DNA segment to lose accessibility. Second, we noted that peak-2 decreased largely in *D. biarmipes* after midpupal life (Fig. 1F and fig. S1A) but persisted in *D. melanogaster*, suggesting a possible link to the evolution of wing pigmentation (*26*).

A novel silencer and a cryptic enhancer underlie the loss of a wing spot in *D. melanogaster*

To understand the functional significance of these accessible regions, we derived a series of reporter constructs to assay the regulatory activities of the corresponding *D. melanogaster* sequences (Fig. 2, A to D). To our surprise, *D. melanogaster* sequences corresponding to peak-1 alone (construct *peak-1*^{mel}; Fig. 2C) drove reporter expression in the spot region of the wing, in a spatial pattern comparable to the activity of the orthologous segment of *D. biarmipes* (*23*), albeit with weaker levels (fig. S3). We found that *peak-1*^{mel} activity depended on the TF Distal-less (Dll; fig. S2, A and B), as does *D. biarmipes* enhancer (*23*), confirming the functional orthology of these regulatory elements. This activity, however, was absent in a construct containing a fragment encompassing peak-1 and peak-2 [construct *peaks-(1+2)*^{mel}; Fig. 2B], which drove uniform wing expression resembling that of *y 5'*^{mel} (Fig. 1D). These results showed that the *spot* activity of *D. melanogaster* was not lost during evolution but instead became cryptic, masked by a repressive activity contained within peak-2 region.

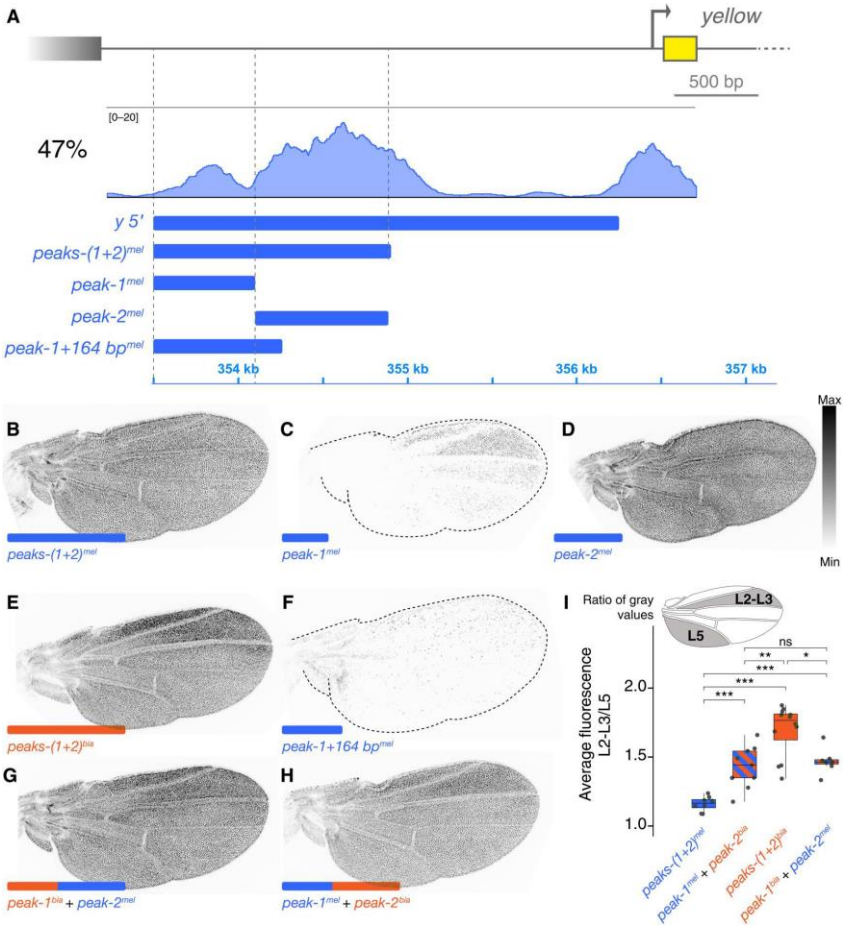

Fig. 2. A novel silencer represses the *spot* activity in *D. melanogaster*. (**A**) Design of reporter constructs to map the regulatory activities of accessible regions at the *yellow* locus of *D. melanogaster*. *peak-1* and *peak-2* segments were defined based on visual inspection of the accessibility profiles, rather than based on peak calling. (**B** to **H**) Reporter activity in transgenic *D. melanogaster* wings at 80% of pupal development for the following constructs: *peaks-(1+2)*^{mel} (B), *peak-1*^{mel} (C), *peak-2*^{mel} (D), *peaks-(1+2)*^{bia} (E), *peak-1+164 bp*^{mel} (F), *peak-1*^{bia}*+peak-2*^{mel} (G), and *peak-1*^{mel}*+peak-2*^{bia} (H). (**I**) Quantification of reporter activities in (B) to (H). Statistical differences were examined using the Wilcoxon test. All box plots show the median and first and third quartiles of the data, overlaid with individual data points. ns, not significant ($P > 0.05$); *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. The contour of wings with weak reporter activity is identified with a dotted line. All wings in this figure are homozygous for the respective transgenes, and images were enhanced using the same settings, which differ from those used in other figures.

To test whether the observed repressive activity was specific to *D. melanogaster* or ancestral to both species, we used chimeric constructs. Specifically, we assayed the repressive effect of *D. melanogaster peak-2* (*peak-2*^{mel}) on *D. biarmipes spot* activity (*peak-1*^{bia}) and reciprocally (Fig. 2, E to H). A *peak-1*^{bia}*+peak-2*^{mel} construct resulted in reduced *spot* activity compared to a *peaks-(1+2)*^{bia} construct (Fig. 2, E, H, and I), indicating that the novel *cis*-repressing activity of *D. melanogaster* can target a heterologous *spot* enhancer. Conversely, while *D. melanogaster spot* enhancer was silenced in its original context [construct *peaks-(1+2)*^{mel}; Fig. 2B], it became active when paired to *D. biarmipes peak-2* orthologous sequence (construct *peak-1*^{mel}*+peak-2*^{bia}; Fig. 2, H and I), confirming the absence of the repressive activity in *D. biarmipes peak-2*.

The DNA segment corresponding to peak-2 overlaps with another enhancer, driving lower levels of *yellow* expression uniformly across the wing (Fig. 2D and fig. S3C), the *wing blade* enhancer (*26*, *28*). In an attempt to disentangle the *wing blade* activity from the *spot*-repressing activity, we mapped the latter further with reporter constructs. We found that a 164–base pair (bp) segment adjacent to peak-1 and devoid of *wing blade* activity

was sufficient to abolish most of the *spot* activity (Fig. 2F). This segment overlapped with a short stretch of differentially accessible DNA between *D. biarmipes* and *D. melanogaster* (fig. S1B, inset).

Together, these results suggest two evolutionary steps that lead to the emergence of a silencer (*29*), suppressing the *spot* activity in *D. melanogaster*: a gain or increase of accessibility in the region immediately adjacent to the core of the *spot* activity, $spot^{196}$ (Fig. 1, E and F) (*26*), and a gain of repressive activity in the newly accessible region. We sought to characterize these events.

Increased accessibility of the *D. melanogaster spot* silencer is promoted by the TF E93, an effector of the ecdysone pathway

We first examined the origin of accessibility of this 164-bp repressive segment. Genome-wide accessibility of regulatory elements in developing wings is governed by the ecdysone pathway (*30*, *31*). Notably, the TF E93, uniformly expressed across the pupal wing (fig. S2E) (*31*), directly regulates chromatin accessibility at thousands of sites during *D. melanogaster* pupal wing development (*30*). To test a possible control of E93 on the novel silencer, we depleted E93 from pupal wings using RNA interference (RNAi) (UAS-shE93 and NP3537-Gal4; two independent RNAi lines) and monitored the consequences on *peak-1+164* bp^{mel} activity. We found that the depletion of E93 was sufficient to restore a *spot* activity with this construct (Fig. 3, B and D; and fig. S2, C and D). Wondering whether this effect was direct, we scanned *peak-1+164* bp^{mel} with PWMEnrich (*32*), which uses a similar position weight matrix to that published by Uyehara *et al.* (*30*) for E93. This scan identified several putative binding sites, including four sites within the 164-bp repressive segment, three of which are not conserved in the orthologous region of spotted species (Fig. 3A and fig. S4). Using a modified version of CUT&RUN (*33*), we then showed that a tagged version of E93 ($E93^{GFSTF}$) (*30*) bound upstream of *yellow* transcription start site while the *spot* activity was being repressed. In particular, a small peak of E93 binding centered on the 164-bp repressive segment in phase with a peak of accessibility and a peak of H3K27ac, a mark for active regulatory elements (Fig. 3D and fig. S1D) (*34*). This was consistent with chromatin immunoprecipitation sequencing data showing that E93 binds to the region of peak-2 in 24-hour after puparium formation (APF) pupal wings (fig. S1D) (*30*). Last, to clarify the exact role of E93 in the repression of the *spot* activity, we examined previously published E93 Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)-seq data on pupal wings from (*30*) (at 44% of pupal development, somewhat earlier than the stages we have studied with ATAC-seq). We found that the region of peak-2 at the *yellow* locus of *D. melanogaster* had a markedly reduced accessibility in *E93* mutants compared to wild type (fig. S1D). We concluded from these experiments that the accessibility to the *spot*-repressive segment was granted by the ecdysone pathway effector E93, presumably directly.

Because the *spot* enhancer, both the cryptic version of *D. melanogaster* (*peak-1*^{mel}) and the active version of *D. biarmipes*, is regulated by the spatial regulator Dll, we were intrigued by previous work showing a genetic interaction between these two factors (*35*). In the context of touch organ development on fly legs, E93 has a permissive role, enabling the expression and action of Dll and thereby the formation of bracts, cuticular productions at the base of sensory bristles. Although this work did not examine the molecular mode of action of E93 in bract cells, its conclusion is perfectly consistent with the permissive role of E93 that we uncovered at the *yellow* locus. It also hints at a possible direct partnership between E93 and Dll.

Repression of the *spot* activity likely evolved through splicing within the newly accessible region

Turning to the repressive activity proper, we noticed that the *spot*-repressive segment of *D. melanogaster* had sequence gaps in the vicinity of putative E93 TFBSs, compared to the orthologous segments of the spotted species *D. biarmipes* and *Drosophila suzukii* (Fig. 4A). While the exact position of these indels depended on sequence alignment parameters, they nevertheless resulted in sequence splicing in *D. melanogaster* at two positions (site 1 and site 2; Fig. 4A) in comparison to other species. We reasoned that such splicing might have created TFBSs for repressors and tested this possibility with point mutations at the spliced sites. A mutation of the first candidate repressor site (*peak-1+164* $bp^{mel\ KO\ site\ 1}$) resulted in a weak but clear derepression of the cryptic *spot* activity (Fig. 4, B, C, and E), while a mutation of the second site (*peak-1+164* $bp^{mel\ KO\ site\ 2}$) did not lead to derepression (Fig. 4, B, D, and E). In summary, our data suggest that the *spot* activity of *D. melanogaster* is constitutively silenced through sites for an unidentified repressor, located within a 164-bp DNA segment adjacent to the core $spot^{196}$, and accessible in pupal wings.

DISCUSSION

The results presented here show a direct involvement of chromatin accessibility changes to regulatory and morphological evolution (Fig. 4F). We propose that, along with the gains and losses of TFBS for spatial regulators, modulation of enhancer accessibility constitutes a level of phenotypic diversification that may have been overlooked. Genome-wide analyses among ecotypes of the plant *Arabidopsis thaliana* suggested, for instance, that most divergences in accessibility of regulatory regions were not associated with changes in gene expression (*36*); likewise, a comparative survey of accessibility changes between two yeast species concluded that changes in accessibility might have modest consequences on gene regulation (*37*). It is conceivable that the contribution of accessibility changes to regulatory evolution is difficult to assess, because the functional significance of accessibility peaks is not necessarily proportional to peak size and might be biased by peak-calling procedures.

Rather than an enhancer decay or deletion (*8*, *38*), the evolutionary mechanism that led to the loss of activity of an enhancer of the pigmentation gene *yellow* is convoluted and implies several evolutionary steps (Fig. 4F). The silenced *spot* activity represents cryptic variation (*39*) potentially available for the reemergence of a pigmentation spot or perhaps expressed under certain environmental conditions or genomic background. Epigenetic changes have been shown to affect pigmentation through the modification of accessibility of specific regulatory regions in response to temperature changes (*40*).

In terms of mutational path and likelihood (*41*), the activity of an enhancer may be lost through direct mutations in its TFBSs, rather than repressed by the emergence of a new *cis*-regulatory element. The former situation has been abundantly documented in insects and in vertebrates (*7*–*9*, *42*–*44*). A rare and interesting exception involves the gain of a repressor TFBS leading to enhancer silencing

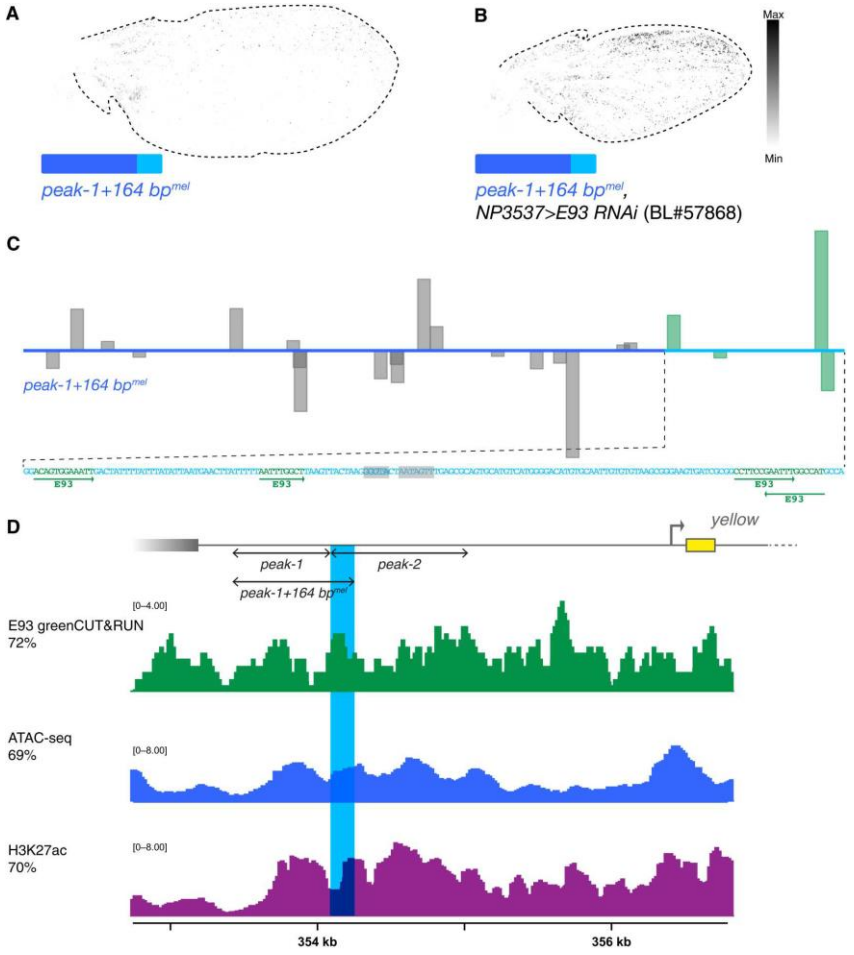

Fig. 3. The ecdysone effector E93 regulates accessibility of a novel silencer at the *D. melanogaster yellow* locus. (**A** and **B**) The *spot*-repressing activity (A) is reduced when depleting E93 from pupal wings by means of RNA interference (RNAi) (B). The knockdown of E93 function also affects general wing development in this line, resulting in deformed wings. (**C**) Predicted E93 sites and their relative binding affinities (proportional to bar lengths) along *peak-1+164 bp*^{mel} [based on the binding motif defined in (*30*)]. Note that the predicted repressive segment mapped in Fig. 2 contains four putative E93 binding sites, highlighted in green. (**D**) E93 binds to the 164-bp repressive segment. A greenCUT&RUN experiment (*33*) shows a peak of E93 binding phased with a small peak of accessibility and marks for active regulatory elements (H3K27ac) centered on the 164-bp repressive segment (turquoise blue). Wings in (A) and (B) are heterozygous for the transgene, and both images were enhanced using the same settings, which differ from those used in other figures.

(*45*). These different studies, however, did not examine whether accessibility had changed along with the reduced or lost enhancer activity. It is conceivable that when an enhancer accumulates mutations in TFBSs for activators, it also accumulates changes in sites regulating accessibility, but these accessibility sites may be mutated first, corresponding to the onset of enhancer decay.

The situation that we observed in *D. melanogaster*, however, requires a more complex explanation. One possibility is that the silencing of the *spot* evolved by serendipity and was rapidly fixed, for instance, under sexual selection from females preferring non-spotted males. Alternatively, the entangled regulatory structure of *yellow* in spotted species (*26*) may have constrained the mutational path to losing the *spot* activity through silencing. For reasons that remain elusive, the direct mutation of *spot* activator sites may not have been permissible in terms of natural selection, for instance, because of pleiotropic effects on the *wing blade* activity. The emergence of a silencer may have offered an alternative route,

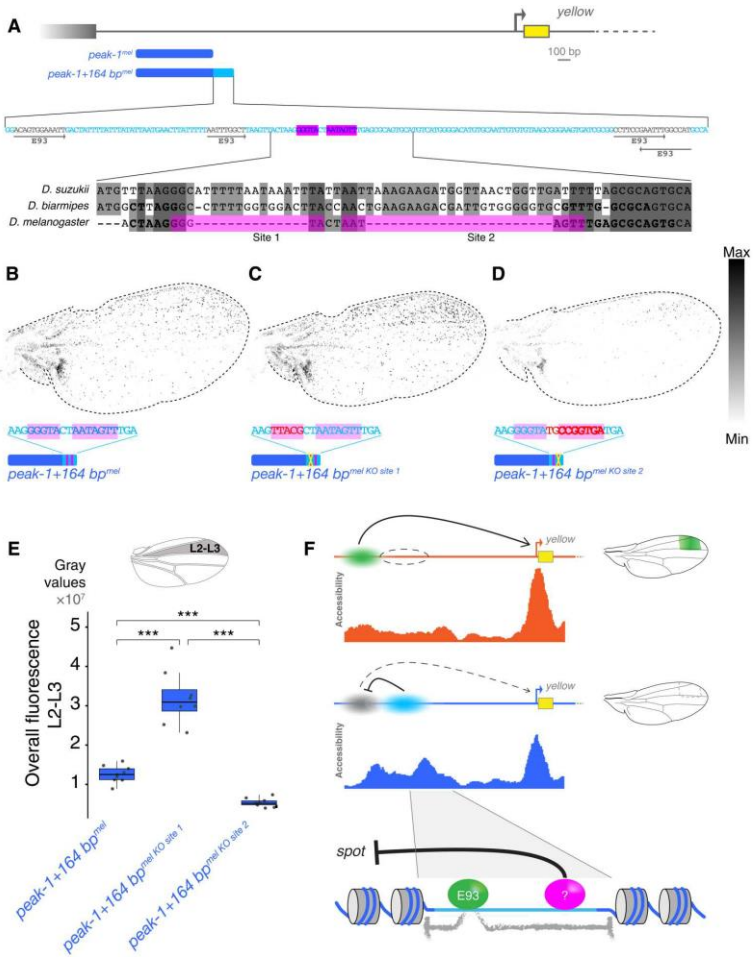

Fig. 4. The evolution of repressive sites in a newly accessible region of *yellow*. (**A**) The 164-bp *D. melanogaster*–repressive segment misses short stretches of sequence compared to spotted species (indels), creating two potential binding sites for repressors (site 1 and site 2, highlighted in magenta). (**B** to **D**) Function of sites 1 and 2, defined in (A). Mutation of site 1 (C) in the context of *peak-1+164 bp*^{mel} results in a derepression of the *spot* activity compared to the activity of the wild-type sequence (B). Mutation of site 2 (D) in the context of *peak-1+164 bp*^{mel} resulted in even less activity than the wild-type construct. (**E**) Quantification of reporter activities shown in (B) to (D) (overall fluorescent signal between veins L2 and L3). All wings in this figure are homozygous for the respective transgenes, and images were enhanced using the same settings, which differ from those used in other figures. Statistical differences were examined using the Wilcoxon test. All box plots show the median and first and third quartiles of the data, overlaid with individual data points. ***$P < 0.001$. (**F**) A model for the evolutionary emergence and function of novel silencer at the *yellow* locus. The top schematic depicts an active *spot* enhancer (green) in a spotted species such as *D. biarmipes*, devoid of silencer. The bottom schematic shows a repressed *spot* enhancer (gray) under the control of a novel silencer (turquoise) in *D. melanogaster*. This silencer is accessible through the local action of E93 (green) and antagonizes the *spot* activity through an unidentified repressor (magenta, "?").

circumventing pleiotropic deleterious effects on other aspects of *yellow* regulation.

MATERIALS AND METHODS

Fly husbandry, fly stocks, and dissection

Our *Drosophila* stocks were maintained on standard cornmeal medium at 25°C (unless otherwise specified) with a 12-hour day/12-hour night light cycle. The stocks used in this study are as follows: *D. biarmipes* genome line (cultured at 20°C) (*46*); *D.*

melanogaster Canton S (cultured at 20°C); *D. melanogaster* UAS-E93 RNAi lines P{TRiP.HMC04773}attP40 (Bloomington, no. 57868) and KK108140 (Vienna *Drosophila* Resource Center, no. V104390) (*47*); *D. melanogaster* UAS-Dll RNAi line (Bloomington, no. 29337); *D. melanogaster* protein trap line $E93^{GFSTF}$ (Bloomington, no. 43675) (*30*); NP3537, tub-GAL80^{ts} [wing-specific Gal4 driver combined to a thermosensitive Gal80 repressor (*48*)]; and *D. melanogaster* line *D2* (*26*). All UAS RNAi lines were driven by the wing-specific Gal4 driver NP3537 (*48*) and cultured at 25°C, where Gal4 is repressed by a tub-GAL80^{ts} transgene. Selected wandering third instar larvae or white pre-pupae were transferred to 29°C and imaged 72 to 75 hours later. Pupal stages are expressed in percentage of pupal development, where white pre-pupae correspond to 0% and flies about to emerge from the pupal case correspond to 100%. This establishes a unique scale for staging pupae from different species, with pupal life of slightly different durations or pupae raised at different temperatures. Wings for genomic experiments were dissected either on microscope slides (for pupae older than 60%) or in cold phosphate-buffered saline (PBS) (for pupae younger than 60%) and rinsed in cold PBS following previous descriptions (*49*, *50*).

Transgenesis

All reporter constructs were injected as described in Arnoult *et al.* (*23*). We used ϕC31-mediated transgenesis (*51*) and integrated all constructs at the genomic *attP* site *VK00016* (attP line; Bloomington, no. 9735) on chromosome 2 (*52*). The enhancer sequence of all transgenic stocks was genotyped before imaging. The sequences of all fragments that we tested are provided in table S1. All primers are listed in table S2.

Molecular biology

Constructs for enhancer-reporter assay were cloned as described before (*26*) with some modifications. PCR-amplified fragments derived from *D. biarmipes* and *D. melanogaster* genomic strains used for ATAC-seq were cloned into pRedSA digested with Eco RI and Bam HI, using the T4 ligase (New England Biolabs, USA). Chimeric constructs were stitched using type IIS restriction enzyme cloning. The 164-bp fragments with mutations were synthesized by Integrated DNA Technologies (USA) with adaptors for type IIS restriction enzymes and then cloned into pRedSA.

ATAC-seq

ATAC-seq was performed as described in Xin *et al.* (*26*) with modifications. For nuclei preparation: Dissected wings were immediately moved into cold 1× lysis buffer after rinsing. Twenty-four 0% pupal development wing discs, 14 to 17 wings from at least 11 individuals at 47% pupal development, and 24 wings from later stages were used for the following steps, respectively. Only the pupal wings older than 60% of pupal development were cut before disruption. Samples were incubated on ice for 20 to 30 min before and after disruption. For tagmentation: The reaction was set up with 18 μl of Tagment DNA Buffer (Illumina, no. 15027866) with nuclei plus 2 μl of Tagment DNA Enzyme (Illumina, no. 15027865). ATAC-seq library preparation and sequencing were performed as described in Xin *et al.* (*26*), with two replicates per stage (biological replicates). ATAC-seq libraries were then processed as described in in Xin *et al.* (*26*) until peak calling. Normalized bedGraph files were then generated from two merged replicates using MACS2 with the following settings: --keep-dup all; -q 0.01; --nomodel; --shift -100; --extsize 200; -B --SPMR. Peak calling was performed using HOMER (*53*) with the following settings: -style histone -size 100 -minDist 100 -gsize 1.2e+8 -o auto.

CUT&RUN

We adapted the manufacturer's protocol (EpiCypher, Chapel Hill, USA) in the following subsections.

Buffers

1× PBS with 6 mM $MgCl_2$; ATAC-seq 1× lysis buffer; cross-link buffer: 10 mM Hepes (pH 8.0), 100 mM NaCl, 1 mM EDTA (pH 8.0), 0.5 mM EGTA (pH 8.0), and 0.1% formaldehyde (Sigma-Aldrich, F1635); 2.5 M glycine in 1× PBS; quench solution: 1× PBS, 125 mM glycine, and 0.1% Triton X-100. Other buffers follow EpiCypher protocol, including bead activation buffer, cross-link (XL) prewash buffer, XL wash buffer (WB), digitonin buffer with 0.01% digitonin, antibody buffer, and stop buffer.

Cell preparation

Thirty-five to 45 wings per biological replicate at 67% pupal development from *D. melanogaster* line *D2* (*26*) were transferred into a 5-ml Eppendorf tube with cold 1× PBS with 6 mM $MgCl_2$, rocked at 4°C with 1× cold lysis buffer for 5 min, and then lightly cross-linked (1 min) with the cross-link buffer. The wings were then rocked at room temperature for 1 min and immediately quenched with glycine, transferred into a glass well with 50 μl of cold XL WB buffer, and cut coarsely (three pieces per wing). Homogenization was performed with a 2-ml dounce using pestle A for 12 strokes and then pestle B for 30 strokes. The homogenate was incubated on ice for 40 to 50 min and then centrifuged at 1000*g* for 10 min at 4°C with an addition of 500 μl of XL WB. The nuclei pellet was resuspended with 100 μl of XL WB and processed for CUT&RUN.

CUT&RUN against H3K27ac

We followed instructions from EpiCypher, including bead activation, binding cells to activated beads, binding of antibodies, binding of Protein A and G fused to Micrococcal Nuclease (pAG-MNase), targeted chromatin digestion and release, and reverse cross-linking. DNA was then purified with QIAGEN MinElute kit and subsequently processed for library preparation. One microliter of antibody against H3K27ac (Active Motif, no. 39034) was used. Two biological replicates were done.

greenCUT&RUN against E93

Cells were prepared as described above for canonical CUT&RUN with minor modifications. Seventy to 90 pupal wings (72% pupal development) per replicate from E93 protein trap (Bloomington, no. 43675) were used for the experiment (two biological replicates were done). The wings were rocked at room temperature for 4 min for cross-link, instead of a 1-min light cross-link. We then followed the protocol from Koidl *et al.* (*33*) until the DNA isolation step. The concentration of digitonin in the digitonin buffer was 0.05%. Reverse cross-link and DNA isolation were done as described above for CUT&RUN against H3K27ac. The DNA was then processed for library preparation. MNase-coupled green fluorescent protein nanobodies were from M. Timmers.

Library preparation, sequencing, and data analysis

For both CUT&RUN and greenCUT&RUN, we used the NEBNext Ultra II DNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 2) for library construction, following the protocol from Liu (*54*). Sequencing was done as in Xin *et al.* (*26*). The sequencing depth is 5 to 10 million reads per

library. Demultiplexed and trimmed libraries were aligned to the reference genome UCSC (University of California, Santa Cruz) dm6 using Bowtie2 (*55*, *56*) with parameters from Meers *et al.* (*57*): --end-to-end --very-sensitive --no-mixed --no-discordant -q --phred33 -I 10 -X 700. The aligned reads were filtered and cleaned as described for ATAC-seq in (*26*). Peak calling was done by MACS2 (*58*) with the following settings: -f BAMPE --keep-dup all -q 0.01 -g 1.2e+8 -B --SPMR.

yellow RNA quantification

RNA extraction

Ten to 12 wings from at least 7 individuals (earlier than 60% of pupal development) and 15 to 20 wings from at least 10 individuals (older than 60% of pupal development) were dissected, rinsed twice in cold PBS, and then transferred into ~150 μl of RLT buffer (QIAGEN) with β-mercaptoethanol and stored at −80°C until further processing. Before performing RNA extraction, each sample was brought up to 350 μl with RLT buffer and transferred into tubes with 2.38-mm metal beads (QIAGEN PowerBead). The wings were homogenized using a microtube homogenizer (BeadBug) for 50 s, power at 400, repeated once. The following steps were performed on the basis of the QIAGEN RNeasy Kit protocol. The RNA was then treated with deoxyribonuclease I (M0303S, New England Biolabs, USA) and then cleaned up using the Monarch RNA Cleanup kit (New England Biolabs, USA). The following stages were selected for RNA extraction: 30, 56, 61, 70, 77, and 86% of pupal development.

RT-PCR

cDNA was synthesized using ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs, USA) following the manufacturer's instructions. Six microliters of RNA was used for cDNA synthesis. Reverse transcription PCR (RT-PCR) was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, USA), and cDNA was synthesized as above. Primers used for RT-PCR are described in Table S2. RT-PCR products were checked on 1.3% agarose gel.

Quantitative real-time PCR

Quantitative real-time PCR was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, USA) and SYBR Green I Nucleic Acid Gel Stain (Invitrogen), and cDNA was generated as described above. The expression levels of *Actin 42A* at each time point was used as an internal control. The primers were the same as the ones used for RT-PCR. The *C*t values were measured by the Bio-Rad CFX Connect Real-Time PCR Detection System, and the *C*t values from technical duplicates and biological duplicates were averaged for relative expression calculation.

Imaging

Sample preparation and microscopy

All transgenic wings imaged in this study were homozygous for the reporter construct, unless otherwise stated. Male white pre-pupae were left to develop for 90 to 92 hours at 25°C. Pupal wings were dissected as described above, transferred into water to unfold them, then mounted in PBS on a microscope slide with a coverslip, and immediately imaged with a Ti2-Eclipse Nikon microscope. All images were acquired as in Xin *et al.* (*26*).

Fluorescence quantification

z stacks were projected using Fiji (*59*) with maximum intensity. Projected images were further processed in MATLAB to generate masks segmenting nuclei with fluorescent signal. Specifically, a Gaussian high-pass filter was used to detect and segment nuclei before generating nuclei binary masks. These masks were then used to measure image intensity in Fiji. At least seven individual wings were measured for each construct. In Fig. 2I, the degree of repression of *spot* activity was measured as the ratio of average fluorescence (overall intensity/area) between two areas of the wing: The region comprised between veins L2 and L3 and the region posterior to vein L5. The former is the region where *spot* activity is modulated, and the latter is a reference for uniform *wing blade* activity. For each wing image, the region between L2 and L3 as well as the region posterior to L5 were selected manually, and the mean intensity of each was measured using the masks described above. In Fig. 4E, the overall intensity of the region between L2 and L3 was measured after applying a mask segmenting the nuclei but was not compared to the region posterior to L5 as above, as these constructs are devoid of *wing blade* activity. The number of nuclei per wing, counted using different reporter lines with expression across the wing, does not vary much in this genetic background (12,392 ± 889; $n = 36$) and is not likely to bias the overall intensity signal.

Image enhancement

Images in the figures were enhanced for visual clarity in agreement with the journal image integrity standards, exclusively with linear adjustments.

E93 motif search

Software from Bioconductor, PWMEnrich (*32*), was used for motif scan. A built-in motif database for *Drosophila* was used for scanning motifs. "Eip93F_SANGER_10_FBgn0013948" was used for plotting E93 binding sites on *yellow* sequences (*peak-1+164 bp*^{mel}; Fig. 3C).

Multisequence alignment

Multisequence alignment was performed with MUltiple Sequence Comparison by Log-Expectation (MUSCLE) from European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) online tools (*60*) using default settings. All sequences from *yellow* upstream regulatory regions are from the reference genomes of *D. suzukii*, *D. melanogaster*, and *D. biarmipes* and listed in table S1 as *y 5′ full*^{suz}, *y 5′ full*^{mel}, and *y 5′ full*^{bia}, respectively.

Supplementary Materials

This PDF file includes:
Figs. S1 to S4
Legends for tables S1 and S2

Other Supplementary Material for this manuscript includes the following:
Tables S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. S. B. Carroll, From pattern to gene, from gene to pattern. *Int. J. Dev. Biol.* **42**, 305–309 (1998).
2. P. Piazza, C. D. Bailey, M. Cartolano, J. Krieger, J. Cao, S. Ossowski, K. Schneeberger, F. He, J. de Meaux, N. Hall, N. Macleod, D. Filatov, A. Hay, M. Tsiantis, *Arabidopsis thaliana* leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep. *Curr. Biol.* **20**, 2223–2228 (2010).

3. 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).
4. E. Sucena, I. Delon, I. Jones, F. Payre, D. L. Stern, Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938 (2003).
5. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
6. S. J. Adamowicz, A. Purvis, From more to fewer? Testing an allegedly pervasive trend in the evolution of morphological structure. *Evolution* **60**, 1402–1416 (2006).
7. E. Z. Kvon, O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, B. J. Mannion, V. Tissieres, C. S. Pickle, I. Plajzer-Frick, E. A. Lee, M. Kato, T. H. Garvin, J. A. Akiyama, V. Afzal, J. Lopez-Rios, E. M. Rubin, D. E. Dickel, L. A. Pennacchio, A. Visel, Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642.e11 (2016).
8. B. Prud'homme, N. Gompel, A. Rokas, V. A. Kassner, T. M. Williams, S. D. Yeh, J. R. True, S. B. Carroll, Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).
9. N. Frankel, D. F. Erezyilmaz, A. P. McGregor, S. Wang, F. Payre, D. L. Stern, Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598–603 (2011).
10. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
11. M. Z. Ludwig, A. Palsson, E. Alekseeva, C. M. Bergman, J. Nathan, M. Kreitman, Functional evolution of a cis-regulatory module. *PLoS Biol.* **3**, e93 (2005).
12. M. Levine, Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
13. E. E. Hare, B. K. Peterson, M. B. Eisen, A careful look at binding site reorganization in the even-skipped enhancers of *Drosophila* and sepsids. *PLOS Genet.* **4**, e1000268 (2008).
14. D. Papatsenko, Y. Goltsev, M. Levine, Organization of developmental enhancers in the *Drosophila* embryo. *Nucleic Acids Res.* **37**, 5665–5677 (2009).
15. J. Pu, Z. Wang, H. Cong, J. S. R. Chin, J. Justen, C. Finet, J. Y. Yew, H. Chung, Repression precedes independent evolutionary gains of a highly specific gene expression pattern. *Cell Rep.* **37**, 109896 (2021).
16. D. J. McKay, J. D. Lieb, A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev. Cell* **27**, 306–318 (2013).
17. D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: From properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
18. M. Bozek, N. Gompel, Developmental transcriptional enhancers: A subtle interplay between accessibility and activity. *Bioessays* **42**, e1900188 (2020).
19. L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K. S. Zaret, Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
20. J. Jacobs, M. Atkins, K. Davie, H. Imrichova, L. Romanelli, V. Christiaens, G. Hulselmans, D. Potier, J. Wouters, I. I. Taskiran, G. Paciello, C. B. González-Blas, D. Koldere, S. Aibar, G. Halder, S. Aerts, The transcription factor *Grainy head* primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
21. Y. Sun, C.-Y. Nien, K. Chen, H.-Y. Liu, J. Johnston, J. Zeitlinger, C. Rushlow, Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
22. P. C. Peng, P. Khoueiry, C. Girardot, J. P. Reddington, D. A. Garfield, E. E. M. Furlong, S. Sinha, The role of chromatin accessibility in cis-regulatory evolution. *Genome Biol. Evol.* **11**, 1813–1828 (2019).
23. 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).
24. 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).
25. Y. Le Poul, Y. Xin, L. Ling, B. Muhling, R. Jaenichen, D. Horl, D. Bunk, H. Harz, H. Leonhardt, Y. Wang, E. Osipova, M. Museridze, D. Dharmadhikari, E. Murphy, R. Rohs, S. Preibisch, B. Prud'homme, N. Gompel, Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci. Adv.* **6**, eabe2955 (2020).
26. 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).
27. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.21–221.29.9 (2015).
28. P. J. Wittkopp, K. Vaccaro, S. B. Carroll, Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547–1556 (2002).
29. S. Ogbourne, T. M. Antalis, Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* **331**, 1–14 (1998).
30. 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 Devel.* **31**, 862–875 (2017).
31. S. L. Nystrom, M. J. Niederhuber, D. J. McKay, Expression of E93 provides an instructive cue to control dynamic enhancer activity and chromatin accessibility during development. *Development* **147**, dev181909 (2020).
32. R. Stojnic, D. Diez, PWMEnrich: PWM enrichment analysis. *R package*, (2021).
33. S. Koidl, H. T. M. Timmers, greenCUT&RUN: Efficient genomic profiling of GFP-tagged transcription factors and chromatin regulators. *Curr. Protoc.* **1**, e266 (2021).
34. 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).
35. X. Mou, D. M. Duncan, E. H. Baehrecke, I. Duncan, Control of target gene specificity during metamorphosis by the steroid response gene E93. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2949–2954 (2012).
36. C. M. Alexandre, J. R. Urton, K. Jean-Baptiste, J. Huddleston, M. W. Dorrity, J. T. Cuperus, A. M. Sullivan, F. Bemm, D. Jolic, A. A. Arsovski, A. Thompson, J. L. Nemhauser, S. Fields, D. Weigel, K. L. Bubb, C. Queitsch, Complex relationships between chromatin accessibility, sequence divergence, and gene expression in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **35**, 837–854 (2018).
37. C. F. Connelly, J. Wakefield, J. M. Akey, Evolution and genetic architecture of chromatin accessibility and function in yeast. *PLOS Genet.* **10**, e1004427 (2014).
38. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal Jr., M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jonsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305 (2010).
39. P. J. Polaczyk, R. Gasperini, G. Gibson, Naturally occurring genetic variation affects *Drosophila* photoreceptor determination. *Dev. Genes Evol.* **207**, 462–470 (1998).
40. J. M. Gibert, E. Mouchel-Vielh, S. De Castro, F. Peronnet, Phenotypic plasticity through transcriptional regulation of the evolutionary hotspot gene tan in *Drosophila melanogaster*. *PLOS Genet.* **12**, e1006218 (2016).
41. N. Gompel, B. Prud'homme, The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47 (2009).
42. 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).
43. F. Leal, M. J. Cohn, Loss and re-emergence of legs in snakes by modular evolution of sonic hedgehog and HOXD enhancers. *Curr. Biol.* **26**, 2966–2973 (2016).
44. C. A. Guenther, B. Tasic, L. Luo, M. A. Bedell, D. M. Kingsley, A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* **46**, 748–752 (2014).
45. E. Preger-Ben Noon, F. P. Davis, D. L. Stern, Evolved repression overcomes enhancer robustness. *Dev. Cell* **39**, 572–584 (2016).
46. Z. X. Chen, D. Sturgill, J. Qu, H. Jiang, S. Park, N. Boley, A. M. Suzuki, A. R. Fletcher, D. C. Plachetzki, P. C. FitzGerald, C. G. Artieri, J. Atallah, O. Barmina, J. B. Brown, K. P. Blankenburg, E. Clough, A. Dasgupta, S. Gubbala, Y. Han, J. C. Jayaseelan, D. Kalra, Y. A. Kim, C. L. Kovar, S. L. Lee, M. Li, J. D. Malley, J. H. Malone, T. Mathew, N. R. Mattiuzzo, M. Munidasa, D. M. Muzny, F. Ongeri, L. Perales, T. M. Przytycka, L. L. Pu, G. Robinson, R. L. Thornton, N. Saada, S. E. Scherer, H. E. Smith, C. Vinson, C. B. Warner, K. C. Worley, Y. Q. Wu, X. Zou, P. Cherbas, M. Kellis, M. B. Eisen, F. Piano, K. Kionte, D. H. Fitch, P. W. Sternberg, A. D. Cutter, M. O. Duff, R. A. Hoskins, B. R. Graveley, R. A. Gibbs, P. J. Bickel, A. Kopp, P. Carninci, S. E. Celniker, B. Oliver, S. Richards, Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* **24**, 1209–1223 (2014).
47. M. C. Pahl, S. E. Doyle, S. E. Siegrist, E93 integrates neuroblast intrinsic state with developmental time to terminate MB neurogenesis via autophagy. *Curr. Biol.* **29**, 750–762.e3 (2019).
48. 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).
49. C. Bolatto, C. Parada, V. Colmenares, A rapid and efficient method to dissect pupal wings of *Drosophila* suitable for immunodetections or PCR assays. *J. Vis. Exp.* , 55854 (2017).
50. K. Flegel, D. Sun, O. Grushko, Y. Ma, L. Buttitta, Live cell cycle analysis of drosophila tissues using the attune acoustic focusing cytometer and vybrant dyecycle violet DNA stain. *J. Vis. Exp.* , e50239 (2013).

51. A. Groth, M. Fish, R. Nusse, M. Calos, Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775–1782 (2004).
52. 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).
53. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
54. N. Liu, *Library Prep for CUT&RUN with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (E7645)* (2021); www.protocols.io.
55. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. B. Langmead, C. Wilks, V. Antonescu, R. Charles, Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
57. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314 (2019).
58. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
59. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
60. F. Madeira, M. Pearce, A. R. N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, R. Lopez, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).

Acknowledgments: We thank A. Maugarny-Calès and M. Timmers for sharing stocks and reagents. We are grateful to B. Prud'homme and S. Ceolin for constructive comments on the manuscript, to V. Courtier for advice on the literature, to L. Osipova for assistance with fly maintenance, to Y. Xin for help with sample collection, and to S. Feng for technical advice. **Funding:** This work was supported by the Graduate School of Quantitative Biosciences Munich (QBM; to L.L.), the Human Frontiers Science Program (program grant RGP0021/2018 to N.G.), and the Deutsche Forschungsgemeinschaft (grants INST 86/1783-1 LAGG, GO 2495/5-1, and GO 2495/11-1 to N.G.). **Author contributions:** L.L. and N.G. conceptualized the project. L.L. designed all experiments and performed all RNA-related, genomic, and imaging experiments and all data analysis. B.M. produced all transgenic lines. R.J. performed all other molecular biology experiments. L.L. and N.G. analyzed and interpreted the data. N.G. wrote the manuscript, which L.L. edited. N.G. supervised the project and acquired funding. **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. All raw and processed ATAC and CUT&RUN sequencing data are available on NCBI Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) through the accession number GSE222717.

Submitted 30 August 2022
Accepted 17 January 2023
Published 17 February 2023
10.1126/sciadv.ade6529

Science Advances
AAAS

Supplementary Materials for

Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*

Liucong Ling *et al.*

Corresponding author: Nicolas Gompel, gompel@bio.lmu.de

Sci. Adv. **9**, eade6529 (2023)
DOI: 10.1126/sciadv.ade6529

The PDF file includes:

Figs. S1 to S4
Legends for tables S1 and S2

Other Supplementary Material for this manuscript includes the following:

Tables S1 and S2

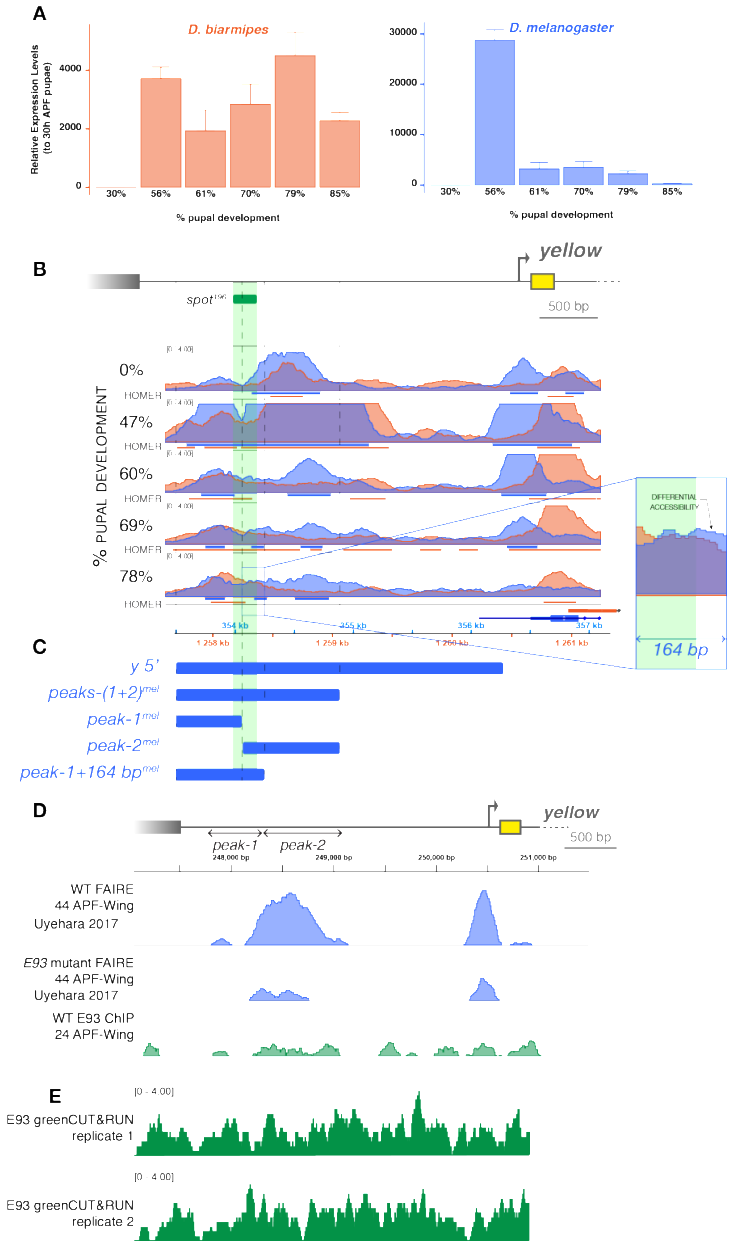
A
D. biarmipes
D. melanogaster
Relative Expression Levels (to 30h APF pupae)
% pupal development
30%
56%
61%
70%
79%
85%
B
yellow
500 bp
% PUPAL DEVELOPMENT
0%
47%
60%
69%
78%
HOMER
DIFFERENTIAL ACCESSIBILITY
164 bp
C
y 5'
peaks-(1+2)mel
peak-1mel
peak-2mel
peak-1+164 bpmel
D
peak-1
peak-2
WT FAIRE
44 APF-Wing
Uyehara 2017
E93 mutant FAIRE
44 APF-Wing
Uyehara 2017
WT E93 ChIP
24 APF-Wing
E
E93 greenCUT&RUN
replicate 1
E93 greenCUT&RUN
replicate 2

Fig. S1. Genomics at the *yellow* locus.
(A) dynamics of *yellow* relative expression during wing development between *D. biarmipes* (orange) and *D. melanogaster* (blue) measured by qPCR. Note that elevated expression persists until late metamorphosis in *D. biarmipes* but not in *D. melanogaster*. **(B)** comparative ATAC-seq tracks at the *yellow* locus during pupal development of *D. biarmipes* (orange) and *D. melanogaster* (blue). **(C)** reporter construct design already showed in Fig. 2, for reference. **(D)** FAIRE-seq and ChIP-seq tracks at the *yellow* locus replotted from published data (*8*). The accessibility indicated by FAIRE-seq peaks at 44 hours after puparium formation (APF) is consistent with peak-1 and peak-2 in our ATAC-seq data, both in terms of position and quantity. Remarkably, the peaks are considerably reduced when E93 is mutated, indicating the requirement of this factor to grant accessibility of *yellow* wing regulatory elements. Moreover, ChIP-seq against E93 (green) at 24 h APF indicates that this factor is bound in the regions corresponding to peak-1 and peak-2. **(E)** a greenCUT&RUN experiment against a tagged E93 ($E93^{GFSTF}$ (*8*)) shows that this factor continues to bind *yellow* regulatory regions when the *spot* activity is being silenced (replicate 1 is already shown in Fig. 3D).

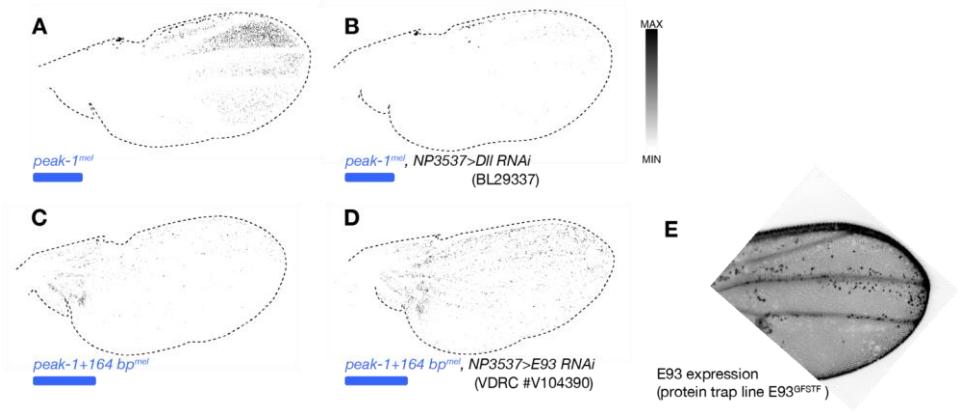

Fig. S2. Down-regulation and expression of *yellow* regulators in *D. melanogaster*.
(A-B) Distal-less (Dll) regulates the cryptic *spot* activity in *D. melanogaster*. The depletion of Dll by RNAi in developing wings (88% development) results in a loss of activity of *peak-1*, consistent with the role of Dll in spotted species (*26, 28*). **(C-D)** E93 is necessary for the repressive activity of the 164-bp silencer. The depletion of E93 by RNAi in developing wing (85% of pupal development) results in an increase of *spot* activity of *peak-1+164 bp^{mel}*. The effect is weaker but consistent with what we found with another RNAi line (Fig. 3b-d) **(E)** In line with previous work (*8*), E93, here monitored with the protein trap line *E93^{GFSTF}*, is distributed homogenously in pupal wings at 73% of pupal development, a stage when the 164-bp silencer is active. Wings in panels (A-D) are heterozygous for the respective transgenes and these four images were enhanced using the same settings, which differ from those used in other figures.

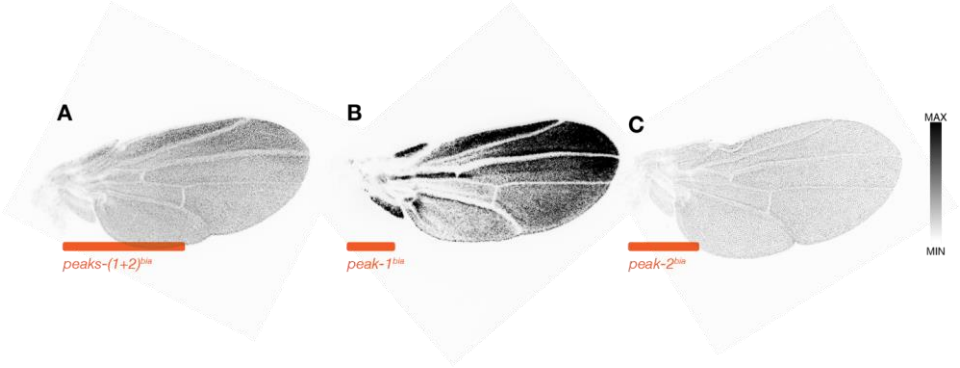

Fig. S3. Regulatory activities of *yellow* accessible regions from *D. biarmipes*.
Reporter activity in transgenic *D. melanogaster* wings at 80% of pupal development for the following constructs: *peaks-(1+2)*^{bia} **(A)**, *peak-1*^{bia} **(B)**, *peak-2*^{bia} **(C).** Note that the very strong reporter expression of *peak1*^{bia} is typically what we observe when this segment is placed adjacent to the core promoter. All wings in this figure are homozygous for the respective transgenes and images were enhanced using the same settings as those used in Fig. 2B-H.

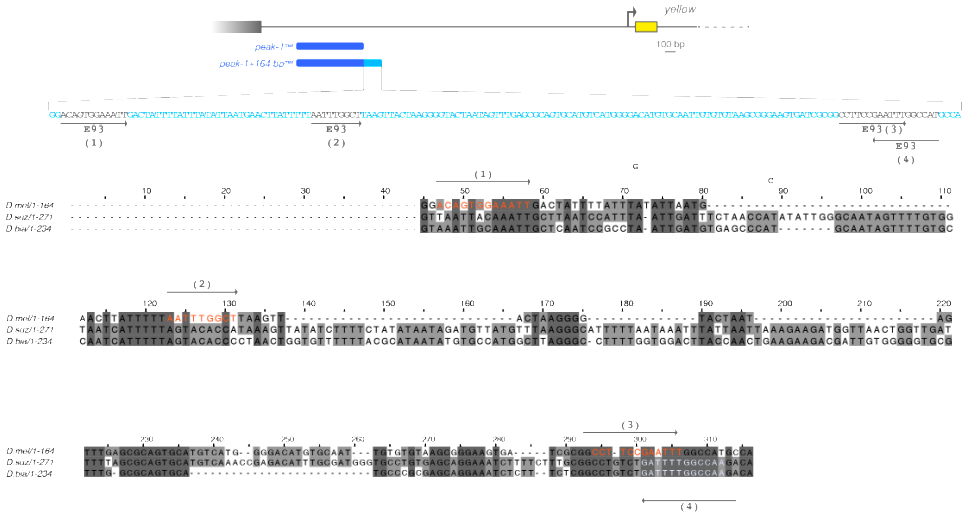

Fig. S4. Evolution of E93 putative binding sites in the 164-bp silencer.

An alignment of *D. melanogaster* 164-bp silencer sequence to the orthologous regions of the spotted species *D. biarmipes* and *D. suzukii* shows that three out of four putative binding sites are not conserved in the spotted species.

Table S1. Insert sequences of reporter constructs.

Table S2. Primers for PCR

>y 5'mel

TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTGGAAATTGACTATTTTATTTATATT
AATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGGGTACTAATAGTTTGAGCGCAGTGCATGTCATG
GGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATTTGGCCATGCCAAATAATCC
CAGCTCGAAAGGAGGGGACCCGGCGGTCAGGGCCATGGACATTGAACTTGAAAAAAAAAAAAAACACAAAA
ATATATAACACAAAACGGAAAATGCTGTGTACCGCTTATGTTAGAGAAGTTGAGCAACGGGTTTTTCGTT
TTGCAGTCACGATGGATTTCCAAATTAGTGTAGGAGGGGGGAGGGGAGGGAGGGAGATAATGTCCAGGCT
GCCATAAGTGGGGAATAAGGAAAATAAAACATGAAACACGGGTCGGGCAATGTCATGCGGTATTCGGCTT
TGCTTTCCGCCCAAGTTGAAGTGATCCTGTGTGTAAATAATGTCGAATGTTGCCGGTCGGTTGCATAAGC
GTTAGTCAATTATGCCAAAGAGATCTGATTTGTGGAAGCTTTTTTTTGACCACTTAGCGCGCTCCGCTTGA
TGTTGTTTTGTTTTGTGCTGGGGCAGAAAACTTGTTTCAATTATTGGGAAAAGTGCGTATAAATCATTGC
CGCAAGCTCTGAAAAGCGAAAAAGAAAAACAGTAACCAAACAGACAAACGCAGCATCCCCCCACACAATT
AAGCAAAAACTTGAAAACAAGTCAATTCGAAAAAAATTATAGGTTCAACGGCTGCAGCGATCGCATCATT
AGTTGCGTTTTTAGTAAATACACCATTTCATTACACAACACACACAATTAATTAATAAAACTGTACTGGT
TATTTCAAGTGTGTCTTTTAATAAGCCTGCCGATCGCAATAAATTCGAGCAGCATTGCCGGTAATTTTGT
GCAACATATTTTTCGATTGCCACACCGTGTTTGTTTATTTTTCTGTGGGTGCAATGATTTAGAATGCGGG
CAAGGGATCAAGTTGAACCACTTCTAAGAAAAAATAGCATTGCATAAATGATATAGAGTCCAAAAACTAC
ACAAATTCAATAGCAGTAATGGTTACATTAGCTTTGAAATTGTTTTTAGACATCCGAAGAAATAAGATTA
AATTTAAACGGCATTCTTTAATTTGTATTTTAATATTTTGAGAGGTTTTCCTTATTTAAAGTGTAGATTA
TTGAGGATTAATGCAAACCACTTTATCTGCGGAGGTCGTAAAACGTATTTTTACCCATTTGCATGTTTAT
TATGCGTGTGGCTGGTTGTATTACTTTACTTAAGTTTTGCAATTTTTTCTTTAGCAAGCAGGTGCATTTG
GGCCAAGAGATATATGCGATCGCTTTCGGTTCGAATTTTTAACATTTACTTGCGGCGATGGTCATTAGAG
CATTACCCACTTAGGGCACCCCCAACATCCAGTTGATTTTCAGGGACCACAATATTTTAAATAACAGCTA
GTGGAATTACCTAAAAGCGCTTTCGTCCCTTTTGAAATTTTATGTAACACTCAATTATATTTATGTATAT
GTATGCTCAAAATCACCTGCCAATAACTAGCGGAAACCAAATATTTGACCCTCAGTGAATTGTGAATCAT
CGGTGACGCCCAATCGAAATCCAATCCTAAGCAATTGAAACGAGCACGAGTTCCAATTTAATAGTATACA
AGGAAACACCTGCTTTAAATACTCTACATAGTACACGTTATAATAACGATTTATTTGATATTTCTGGATT
TTTGTCTGCATGTATTTCATATAATATTGATTTGATTTTTTTAATGAATTGAACTAAAAAATCATATTAG
AACATTTTTTGCAGTCGCCGATAAAGATGAACACTGTTCTCAGAACACAACTGTCATGTATTAAGCTTTC
AGATTTTCAGAAATTTGGAGAGCAATGCATTCTATGCACGAGCCTCCTGGCCTTACAATTTACTTGTTTG
AAATTAGATCGTCAAATAAAGTCCCTAAAATTAAATAAATAGTAGTCACAACTTTAAAATAGGTCTTAAT
CTTTTAGGGTACCGAAAGGTATTTCGGCACAAATCAGCGCAGTTTTAAATGTCGATGAAGGCCAAAAATC
ATACCAAACCCAG

>peaks-(1+2)mel

TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT

TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTGGAAATTGACTATTTTATTTATATT
AATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGGGTACTAATAGTTTGAGCGCAGTGCATGTCATG
GGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATTTGGCCATGCCAAATAATCC
CAGCTCGAAAGGAGGGGACCCGGCGGTCAGGGCCATGGACATTGAACTTGAAAAAAAAAAAAAACACAAAA
ATATATAACACAAAACGGAAAATGCTGTGTACCGCTTATGTTAGAGAAGTTGAGCAACGGGTTTTTCGTT
TTGCAGTCACGATGGATTTCCAAATTAGTGTAGGAGGGGGGAGGGGAGGGAGGGAGATAATGTCCAGGCT
GCCATAAGTGGGGAATAAGGAAAATAAAACATGAAACACGGGTCGGGCAATGTCATGCGGTATTCGGCTT
TGCTTTCCGCCCAAGTTGAAGTGATCCTGTGTGTAAATAATGTCGAATGTTGCCGGTCGGTTGCATAAGC
GTTAGTCAATTATGCCAAAGAGATCTGATTTGTGGAAGCTTTTTTTGACCACTTAGCGCGCTCCGCTTGA
TGTTGTTTTGTTTTGTGCTGGGGCAGAAAACTTGTTTCAATTATTGGGAAAAGTGCGTATAAATCATTGC
CGCAAGCTCTGAAAAGCGAAAAAGAAAAACAGTAACCAAACAGACAAACGCAGCATCCCCCCACACAATT
AAGCAAAAACTTGAAAACAAGTCAATTCGAAAAAAATTATAGGTTCAACGGCTGCAGCGATCGCATCATT
AG

>*peak-1*^{mel}
TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCG

>*peak-2*^{mel}
GGACAGTGGAAATTGACTATTTTATTTATATTAATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGG
GTACTAATAGTTTGAGCGCAGTGCATGTCATGGGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGC
GGCCTTCCGAATTTGGCCATGCCAAATAATCCCAGCTCGAAAGGAGGGGACCCGGCGGTCAGGGCCATGG
ACATTGAACTTGAAAAAAAAAAAAAACACAAAAATATATAACACAAAACGGAAAATGCTGTGTACCGCTTA
TGTTAGAGAAGTTGAGCAACGGGTTTTTCGTTTTGCAGTCACGATGGATTTCCAAATTAGTGTAGGAGGG
GGGAGGGGAGGGAGGGAGATAATGTCCAGGCTGCCATAAGTGGGGAATAAGGAAAATAAAACATGAAACA
CGGGTCGGGCAATGTCATGCGGTATTCGGCTTTGCTTTCCGCCCAAGTTGAAGTGATCCTGTGTGTAAAT
AATGTCGAATGTTGCCGGTCGGTTGCATAAGCGTTAGTCAATTATGCCAAAGAGATCTGATTTGTGGAAG
CTTTTTTTGACCACTTAGCGCGCTCCGCTTGATGTTGTTTTGTTTTGTGCTGGGGCAGAAAACTTGTTTC
AATTATTGGGAAAAGTGCGTATAAATCATTGCCGCAAGCTCTGAAAAGCGAAAAAGAAAAACAGTAACCA
AACAGACAAACGCAGCATCCCCCCACACAATTAAGCAAAAACTTGAAAACAAGTCAATTCGAAAAAAATT
ATAGGTTCAACGGCTGCAGCGATCGCATCATTAG

>*peak-1+164 bp*^{mel}
TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT

CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTGGAAATTGACTATTTTATTTATATT
AATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGGGTACTAATAGTTTGAGCGCAGTGCATGTCATG
GGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATTTGGCCATGCCA

>y 5'bia
TGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAACGGGAT
GAGGGACCAGTTGAGGCGAAGTATATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGA
TCGCTGCGATTGTTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGC
GGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTAT
GCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAG
TGCGGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAAT
TACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCG
CTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGT
TTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGAGCCCATGCAATAGTTTTGTGCCAATC
ATTTTTAGTACACCCCTAACTGGTGTTTTTTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGA
CTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCATGCCCGCGAGCAGGAAATCTCT
TTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGG
TCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGT
GCTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCTGGCCAACACT
GGGGGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAA
CACGGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTCCGGCTAACTCGCAGCGGTCCTGTGTGTAAA
TAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATT
TTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCACTTGATAAT
GATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCGATTATGGGGAAAAGCGCGTATAAATCATCGCCGCC
GAAGTCTGGCAAAACAGCAAATTGAAAACTGCAAACTGCAAACTGAAAACTGTAACCCAAACAAACACAG
CATCCCACACGACGAGATGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTC
AAATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTAT
GCAACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATC
GCAATAAATTAAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTG
AGCCGAGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCGGGCCAAGAGTTTTTCGCATT
TTGCATAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATT
TAAAAATATATATAAATTTATGGCATGTTCTGTACAGAACTTATGAAGCTAATTTGAAATACTTCTTCTT
AGGAAACTGTCCCTAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATT
TCCCAGACATTTTCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGG
AGGTCGTAAACGTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATT
TTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTT
CGCACTCGCTTGCGGCCATGGCCATTAGAGCATTACCCACTTAGGGCGCCCTAAAGTCCAGGTGGTCCCC
AGGGACCACAAGAGTATTGCAACTTACGGCCAGCTGAGTGGAGTGCCGGAACGCACTTCTTTAATTTCGG
CGGTTATGTAACCTCGAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACC
AACTGTTTGGCCCTCGCCGGACTGTGAATCATCGGAGCTGCCCAATCGAAATCAAAGCCAAGCCAATCGA
AGCCCAGGTAATTCATATTTAATGCTGTTAGCAAAGACCTGTCCTAGATACTCTGTATAAAAGTATAATT
ATTAAGTGCATATCAGGTTTATTTACATTTATATCGTATAATATTGGTAAGTGCAGCAGATGCTGTGATC
TAAATTTAGAGTCATTTAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTT
TGGATGAAGTAAGAAAGCCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTA
AAGTATTTCGTACCTACGTTGCATGGTATTCAATAAAGATTCAAAAATACTCTCACCCACTGTAAGTGAA
CCCAGTGTTTCGTAATTGCCTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTC

GGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAA
GCCCGG

>*peaks-(1+2)*^{bia}
TGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAACGGGAT
GATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGA
TCGCTGCGATTGCTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGC
GGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTAT
GCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAG
TGCGGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAAT
TACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCG
CTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGT
TTCAATGTAAATTGCAAATTGCTCAATCCGCCTAATTGATGTGCGCCCATGCAATAGTTTTGTGCCAATC
ATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGA
CTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCACGCCTGCGAGCAGGAAATCTCT
TTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGG
TCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGT
GCTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACT
GGGAGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAA
CACGGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCCTAACTCGCAGCGGTCCTGTGTGTAAA
TAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATT
TTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAAT
GATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCAATTATTGGGAAAAGCGCGTATAAATCATCGCCGCC
AAAGTCTGGCAAAACAGCAAATTGAAAACTGCAATCTGAAAACTGAAAACTGAAAACTGTAACCCAAACA
AACACAGCATCCCACACGACGAGATGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACAAGTCAAT
TCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGATCGCATCATTAAG

>*peak-1*^{mel} + *peak-2*^{bia}
TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGAAATTGCAAATTGCTCAATCCGCCTAATTGAT
GTGCGCCCATGCAATAGTTTTGTGCCAATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAAT
ATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAAGAAGACGATTGTGGGGGTGCGTTTGGC
GCAGTGCACGCCTGCGAGCAGGAAATCTCTTTCTCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCG
GCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGACATTGAACTTGAAAACGCAGCCAGCGC
CGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATGGGTGTCTCTGGCTAAGGTTACGGCGGT
TGGGCAACAGGTTTTCCCCCGGCCAACACTGGGAGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATA
AGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCGGGCAATGTCACTCGGCATTCGCTTGATTTTC
CGCCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGC
AATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCTGCCATGTTCCGCTGCCACCGTATTTCTTT
TCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCAATTAT
TGGGAAAAGCGCGTATAAATCATCGCCGCCAAAGTCTGGCAAAACAGCAAATTGAAAACTGCAATCTGAA

AACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCACACGACGAGATGAAAATGAAAATAAA
TACGGACTGAGCGACTGAAAACAAGTCAATTCGATTCAAATTGCAGGTTCAACGGCTGCCGGCGATCGCA
TCATTAAG

>*peak-1*^{bia} + *peak-2*^{mel}
TGGTGCCCGAGTGTTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAACGGGAT
GATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGA
TCGCTGCGATTGCTTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGC
GGCGATGGCGGCTTTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTAT
GCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAG
TGCGGCGCGGCAGTTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAAT
TACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCAT
CTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCG
CTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGT
TTCAATGTGGACAGTGGAAATTGACTATTTTATTTATATTAATGAACTTATTTTTAATTTGGCTTAAGTT
ACTAAGGGGTACTAATAGTTTGAGCGCAGTGCATGTCATGGGGACATGTGCAATTGTGTGTAAGCGGGAA
GTGATCGCGGCCTTCCGAATTTGGCCATGCCAAATAATCCCAGCTCGAAAGGAGGGGACCCGGCGGTCAG
GGCCATGGACATTGAACTTGAAAAAAAAAAAAACACAAAAATATATAACACAAAACGGAAAATGCTGTGT
ACCGCTTATGTTAGAGAAGTTGAGCAACGGGTTTTTCGTTTTGCAGTCACGATGGATTTCCAAATTAGTG
TAGGAGGGGGGAGGGGAGGGAGGGAGATAATGTCCAGGCTGCCATAAGTGGGGAATAAGGAAAATAAAAC
ATGAAACACGGGTCGGGCAATGTCATGCGGTATTCGGCTTTGCTTTCCGCCCAAGTTGAAGTGATCCTGT
GTGTAAATAATGTCGAATGTTGCCGGTCGGTTGCATAAGCGTTAGTCAATTATGCCAAAGAGATCTGATT
TGTGGAAGCTTTTTTTGACCACTTAGCGCGCTCCGCTTGATGTTGTTTTGTTTTGTGCTGGGGCAGAAAA
CTTGTTTCAATTATTGGGAAAAGTGCGTATAAATCATTGCCGCAAGCTCTGAAAAGCGAAAAAGAAAAAC
AGTAACCAAACAGACAAACGCAGCATCCCCCCACACAATTAAGCAAAAACTTGAAAACAAGTCAATTCGA
AAAAAATTATAGGTTCAACGGCTGCAGCGATCGCATCATTAG

>*peak-1 + 164 bp*$^{mel\ KO\ site\ 1}$
TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA
CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTGGAAATTGACTATTTTATTTATATT
AATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGTTACGCTAATAGTTTGAGCGCAGTGCATGTCATG
GGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATTTGGCCATGCCA

>*peak-1 + 164 bp*$^{mel\ KO\ site\ 2}$
TGGTAAATCAATGATTCGACTGTTGTTGTATAAAAATGGCGAAACTGGAAACACTAAACTAGTAAATGCG
AAATGTATCATATGGATTTTCATATTTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCG
ACTATTAAATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAATAGCCATAAAATATGCAACTGACG
ATGGCTTAAGATAAATACGTCGCAGAGTCACTCATAAATTTCGAACGCAGCCCGCTGATTTACCTACCCC
TCTAAACGATTCATAGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGC
ACGAGACAAGTGCACCACCGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGAT
TACACAGAAATCAAATTACGACAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATT
CCACTTATGGTTGCGATTTCGGGAGCTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGA

CATTTAAAATTTACAAATAAAGTTTCTCTATTCAATCGGGACAGTGGAAATTGACTATTTTATTTATATT
AATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGGGGTATGCCGGTGATGAGCGCAGTGCATGTCATG
GGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATTTGGCCATGCCA

>*y 5′ full*^{suz} [*D. suzukii* reference sequence]
ATCTTACTTTTATTAAATTTAAATTTATATGTCAAGTTTATTTTGCCACCCAGTAAATCTATAAAATATA
CCTATAATTTTCTATCCGTAAAATTTCGACCTATAAATACCCACCGCAAATTTCGAAACCAACTTATTCA
TAACCCACAAAGCCAGTTCGATTTGTTTAATGCTCAATTATCCTTAGTTCAGAAGGTATTTGCTTTTTAT
GCATATTTTCCCGCTTACCTATAAATACGCCTAACGAAATTTATTTAAACCTATAAACCCAGTTTCAGTA
AGTGCTTCTTTATGAATATTTTCCATCTAGTTATTTTAATTGAAAGCTACAAATTGTGGTCCATAAATGT
TTTGACTGCTGGCAAATGTGGTGGTTGTACTTTTGATAAATCAATGCAGCGAAACGGGATGATGAACAGT
GAACCAGTTAATGCGAAATATATATCATATATCAATCTCATTTACTTACTTATAGTTTTCTCGGACCGCC
ACTAATGGAAATGTTTGCCTTGAGGAGATCGGGCGATTATTAATCGATAATTGCCCGATTACCACGCTGA
GTAAACTTAAAGACCCATAAGAAATGCAACCGATGGTGGCTTAAGATAAGTAAGTCGTCGAGTTGCTCAT
AAATTTCGAATGCAGTCCGCAGATTTATGCAGACCACGGAAAAGCTATTACTATTAAGCTTTTCTCGACC
ACGGATTTTTCCGCTTGCATGAGACAAGTGCACCGCAGCAATTGCAGGCAAATTATGTCTGAGGCAATGT
TTTGGGGCATGTCTACACCGAAATCAAATTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTTGCGC
AGTGAACGCAGCAGCGCCTTTTATTGGCATCTAATTATTCCGTTTAAGGATGCAATTTTCTGAGCTAAAA
CACACTTATGGAGAGATCTAAATTTCTCTGCATTTGGCTTGAATAAATTAATCGAATTCCCCGCTGACTA
TTAAAATAGACAAAAAGGCGATCTCTTGTATTTTAATGTTAATTACAAATTGCTTAATCCATTTAATTGA
TTTCTAACCATATATTGGGCAATAGTTTTGTGGTAATCATTTTTAGTACACCATAAAGTTATATCTTTTC
TATATAATAGATGTTATGTTTAAGGGCATTTTTAATAAATTTATTAATTAAAGAAGATGGTTAACTGGTT
GATTTTTAGCGCAGTGCATGTCAAACCGAGACATTTGCGATGGGTGCCTGTGAGCAGGAAATCTTTTCTT
TGCGGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGGAGATAAGAAGAGGGGATCCGGCGGTC
AGGGCCATGGACATTGAACTTGAAAACGCAGAAAGCGCTTAAAACATTGTGTTTAACGAACGACAAATGC
TGTGCGGTGGTGTGTCTTTGGCTAAGGTTAGGGGAGTTTGGCAACAGGTTTTCCACTTGCCAATGATAGA
TGGAAATCCAAACTGGGGGGGAGAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGAAAAAAGGAA
AACAAAACATGAAACACGGGCCGGGCAATGTCATGCGGCATTTCGCTTGGTTTTCGCTTCATTTTCCGCC
TAAGTCGCAGTGATCCTGTGTGTAAGTAATGTCTAATGTTGCATGCCGGTTGCATAATCGTGTGGCAATT
ATGCCAAAGAGATTCGTTTATTTATTTTTAATTTTTTTTTTATCATGTTCCGCTGCTACTATATTTTTTTG
GCCACTTAGTGCGCCGCGCTTGATAATGATGTTTTGTTTTTCGCGTGCTGGCGCAAAGACAAACCCGTTT
CAATTATTGGGAAAAGCGCGTATAAATCATCGCTGCTGAAATCTGCGAAAACTGAAAAATGAAAACTGAA
AAGTGTAACCCAAAACAAACGCAAACAAACATAGCATCCCACACGACAAAATCAAAATCAAAATGGATAT
GAAAAACAGAAAAACGAGTCAATTCGATTTAAATTATAGGTTCAACGGCTGCTGGCGATCGCATCATTAA
TTGCGCTTTCAGTGGATACTCGGCTGCGTTATGCAACAAACACACACAATTAATTAATAAAAGCGTACTG
GTTATTTTAGCCGTGGCTTTTTATGGACCTGCCGATCGCAATAAATTTAAGCAGCATTAGTCGTAATTAT
GTGCAACATAGTTTCGGCTGCCGCGTGTCTCGGTTCATTTTTCTGTGGGTGAGCCGAAGTCCAGAATGCG
GCCAAGGGATCAAGTTAAACCACTTTTCCGCACACCACACAGCTTTTTCGCATTTTGCATAATATTACCA
TAGCATAAACAGCGAAAACAAATGCCAAGAATGAATAAAACGAGGAAAACAATTCAAAAATATTTTCAGA
AGTATCATTCATAATGTTTACCAGGTTTTTTTTATTAAGTGTTCTAAGTAGAATTAAAAAAAAAAGAAAA
TGAACGATATTTATTAGGCAAATCGAATTCTTTGTCACCATTCTTGTTACGAAATTGTCTCCAAATATAT
TTGTTTTTCCTAGACTTACTCAGACTTTAGTTGCACCTATCAAATAATAATACTTGTTTTGCGATTTTTC
TGAAAGATTTTCCTTAGCCAAAGTGTAGATTGTGGGACTCCAGATTAATGCAAACCACTTCATCAGCTGA
GGTCGTAAAACGTATCTTTACCTATTTGGTTCGTTCATTATGCGTGTGGTATAACGTTATTTTTTTAGAT
TTTTTCCTAACCAGCTGCAGTTGGGCCAAAGAGTTATGCGAACGCCTTCGGATTTTTTTCGAGTTTTCAC
ATTCACTTGCGGCGATGGCCATTAGAGCAAAACCCACTTAGGGCGCCCCTAAAATCCAGGTGGTCCACAG
GGACCACAACAGCATTGCAAATTACAGTCAGTTGAGTGGAGTTACCGAAACGCACTTCTTTAATTTTGGC
GGCCTTATGTAACTCTCGAATTGTATGTGCGATAAATATGCTAAAAATCACCTGCTCATAATTAGCGGAA
ACCAATAGTTTGACCCTCGCTGAATTGTGAATCATCGGAGCTGCCCAATCGAAATCAAAGCTAAACCATT

CGAAGCCAGCCCAAGTAATTCATATTTAATGTTGTACAAGAAATACCCGCTCTGAATACTCTGTATCCAT
AAGCATTATACTAATATATAATTCATAACCATATCTGGTTTACTTACATACAAATCATACATTATAGTTA
AGTGCACTAGCTATTTTTTGTTGCCCTAGATTTAATTTATATAAACTTTGCATATTTAAAATCCATTTAA
AAATGAGTCGCCTTTTAAAAATTCTTTTGAAGGGAATTTAGATGAACGCATATATTAACCTTAAATTAAC
CCTTAATAATCCATAATTTGTTACCTTAGTAAAAAAGTAAAATAAAAATGCCCCCACCCATTGTAAGTGA
ACCTAATGTTTTGTAGTTGATTTGCACATAAATCAGCAAAATCCTAAATGTAGATGAAGGCCAGGAATTT
CAAAGAATTCGGTGTGCCAAAAACCATACCCTTTCAAAATTATATGAAAAATGGCAAGTCCGGCGAAAGG
TGTTGGCTGATTCAGTGAATTCGGAGTTTTGAGATACTTGTATTTATTAAACACGCGTGAAAATCAATCA
CTGAAGACAAAAGCCACGCACTAAAAAAAAAAACATAAGTATCCGAAGTTGGCCGATCCATGGGTGACCAA
ATAGATCATATAGTGGACCACCGCCACCCGAACCGTTGAAAACAACCGAATCACTGAAACCACCGAAGTT
GGCGCCTTCGTTTTCGTTTTCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAAAACCTCATATAAAACGTGG
CCGACATATTGTGGCCAACAGTCGTAAGCGCGCCA

>*y 5′ full*^{mel} [*D. melanogaster* reference sequence]
TAAAATTATAGATCGACATTACATTGTAATAAATTGTTTTAATAATAATATTATTATTATTTGATATTGT
TTTTTTTTAAATATAAAAATAAAAATGAAATTTTACGATGTATTTTTTGATTAAAGGGTTGGATTGATAA
CTAAATTGAGGGTTTTTTTTTTTTTTACTAAATAAATTGGTAAATCAATGATTCGACTGTTGTTGTATAAA
AATGGCGAAACTGGAAACACTAAACTAGTAAATGCGAAATGTATCATATGGATTTTCATATTTTTTCCAC
TGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCGACTATTAAATGATTATCGCCCGATTACCACATTG
AGTGGTTTAAAATAGCCATAAAATATGCAACTGACGATGGCTTAAGATAAATACGTCGCAGAGTCACTCA
TAAATTTCGAACGCAGCCCGCTGATTTACCTACCCCTCTAAACGATTCATAGTATATGTACGAGTATATC
CACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGCACGAGACAAGTGCACCACCGCAATTGCAGGCAAA
TTATGTCTGAGGTAATGATTCCGTTTCGTGCAAGATTACACAGAAATCAAATTACGACAACCTTTATTCA
GTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATTCCACTTATGGTTGCGATTTCGGGAGCTACAATCG
GTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGACATTTAAAATTTACAAATAAAGTTTCTCTATTCA
ATCGGGACAGTGGAAATTGACTATTTTATTTATATTAATGAACTTATTTTTAATTTGGCTTAAGTTACTA
AGGGGTACTAATAGTTTGAGCGCAGTGCATGTCATGGGGACATGTGCAATTGTGTGTAAGCGGGAAGTGA
TCGCGGCCTTCCGAATTTGGCCATGCCAAATAATCCCAGCTCGAAAGGAGGGGACCCGGCGGTCAGGGCC
ATGGACATTGAACTTGAAAAAAAAAAAAACACAAAAATATATAACACAAAACGGAAAATGCTGTGTACCG
CTTATGTTAGAGAAGTTGAGCAACGGGTTTTTCGTTTTGCAGTCACGATGGATTTCCAAATTAGTGTAGG
AGGGGGGAGGGGAGGGAGGGAGATAATGTCCAGGCTGCCATAAGTGGGGAATAAGGAAAATAAAACATGA
AACACGGGTCGGGCAATGTCATGCGGTATTCGGCTTTGCTTTCCGCCCAAGTTGAAGTGATCCTGTGTGT
AAATAATGTCGAATGTTGCCGGTCGGTTGCATAAGCGTTAGTCAATTATGCCAAAGAGATCTGATTTGTG
GAAGCTTTTTTTGACCACTTAGCGCGCTCCGCTTGATGTTGTTTTGTTTTGTGCTGGGGCAGAAAACTTG
TTTCAATTATTGGGAAAAGTGCGTATAAATCATTGCCGCAAGCTCTGAAAAGCGAAAAAGAAAAACAGTA
ACCAAACAGACAAACGCAGCATCCCCCCACACAATTAAGCAAAAACTTGAAAACAAGTCAATTCGAAAAA
AATTATAGGTTCAACGGCTGCAGCGATCGCATCATTAGTTGCGTTTTTAGTAAATACACCATTTCATTAC
ACAACACACACAATTAATTAATAAAACTGTACTGGTTATTTCAAGTGTGTCTTTTAATAAGCCTGCCGAT
CGCAATAAATTCGAGCAGCATTGCCGGTAATTTTGTGCAACATATTTTTCGATTGCCACACCGTGTTTGT
TTATTTTTCTGTGGGTGCAATGATTTAGAATGCGGGCAAGGGATCAAGTTGAACCACTTCTAAGAAAAAA
TAGCATTGCATAAATGATATAGAGTCCAAAAACTACACAAATTCAATAGCAGTAATGGTTACATTAGCTT
TGAAATTGTTTTTAGACATCCGAAGAAATAAGATTAAATTTAAACGGCATTCTTTAATTTGTATTTTAAT
ATTTTGAGAGGTTTTCCTTATTTAAAGTGTAGATTATTGAGGATTAATGCAAACCACTTTATCTGCGGAG
GTCGTAAAACGTATTTTTACCCATTTGCATGTTTATTATGCGTGTGGCTGGTTGTATTACTTTACTTAAG
TTTTGCAATTTTTTCTTTAGCAAGCAGGTGCATTTGGGCCAAGAGATATATGCGATCGCTTTCGGTTCGA
ATTTTTAACATTTACTTGCGGCGATGGTCATTAGAGCATTACCCACTTAGGGCACCCCCAACATCCAGTT
GATTTTCAGGGACCACAATATTTTAAATAACAGCTAGTGGAATTACCTAAAAGCGCTTTCGTCCCTTTTG
AAATTTTATGTAACACTCAATTATATTTATGTATATGTATGCTCAAAATCACCTGCCAATAACTAGCGGA
AACCAAATATTTGACCCTCAGTGAATTGTGAATCATCGGTGACGCCCAATCGAAATCCAATCCTAAGCAA

TTGAAACGAGCACGAGTTCCAATTTAATAGTATACAAGGAAACACCTGCTTTAAATACTCTACATAGTAC
ACGTTATAATAACGATTTATTTGATATTTCTGGATTTTTGTCTGCATGTATTTCATATAATATTGATTTG
ATTTTTTTAATGAATTGAACTAAAAAATCATATTAGAACATTTTTTGCAGTCGCCGATAAAGATGAACAC
TGTTCTCAGAACACAACTGTCATGTATTAAGCTTTCAGATTTTCAGAAATTTGGAGAGCAATGCATTCTA
TGCACGAGCCTCCTGGCCTTACAATTTACTTGTTTGAAATTAGATCGTCAAATAAAGTCCCTAAAATTAA
ATAAATAGTAGTCACAACTTTAAAATAGGTCTTAATCTTTTAGGGTACCGAAAGGTATTTCGGCACAAAT
CAGCGCAGTTTTAAATGTCGATGAAGGCCAAAAATCATACCAAACCCAGCGAAAGGTGATGTCTGACTCA
TTAAATTGGGGGATTCGAGTGTATTTATTAAACATGCGTGAAAATCAATCATGGAAGACAAAACGCAAAG
TTGGCCGATCTATGGGAACAGCATAAGCCACCTGATTACCCGAACACTGAACCACCCGAATCACTAAAAC
CACCGAAGTTGGCGCGCGCCTTCGTTTTCATTTTCATTGGCCTGTCTTCGTCTTCGGAGAAAAAAAACTT
CATATAAAACGCGGCCGACATATTATGGCCACCAGTCGTTACCG

>*y 5′ full*^{bia} [*D. biarmipes* reference sequence]
CGTATATAAATGTTCTTGTTATGATACATTTTACTAAATTAAAAGTTAAATAATTAAAATATTTAATTTA
CTCATTGTGGGAGTTCATTAGACTTATTGTTTGTTTTTATTAAATTGTAATTTGTATCAAAAGTTTAAGT
TAAAGTCTTAAAATATATATCAAGTTCATTGCACAAATTAACTTAAAAATATGTACCAAAAATGTAGAAA
TCAACCTATTTAATCTAAATGTAAATAAATGTAACAAACCAATCATGTCTTAAATTAAATATGTTTTTAA
GCCTAACACTTATATTTAAAAAATAATTATAATTCAGACTTAAATTTTAAAATCGTCTCACCTTTTTTTA
GTTTATTTGATGCATGTTTTAATTTTGCTAAAATCGTATTTTGATGTCTATTCATTTTGCCAGCCAGTAA
TCTAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTCGACCTATAAATACCCACCGCAAATGCCGCA
ACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAATGGACAATTATCCTTAGTTCAGAAAGCGCCTGC
CTTTATGCGTATTTCCCCGCTTGCCTGCGAATACGCCTAACGAAATTAATCGAGCCCGTAAACCCAGTTT
CGATAAGTGCTTCTTTATGAATATTTTCCATTTACTTTAATTGAGGGCTGCAAATTGTGGTGCCCGAGTG
TTGTGACTGCTGGCCAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAACGGGATGAGGGACCAGTTG
AGGCGAAGTATATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGGGGAGATCGCTGCGATTGT
TTATCGATAATCGCCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAAATGCGGCGATGGCGGCT
TTAGATAAGTAAGTCGTCGGGGCGCTCATAAATTTCGAGCGCGATCCACGGATTTATGCACTCGCTGGAA
AAGCTATTACCATTAGGCTTTTCGCGACCACGGATTTTTCCGCTTGCCTGAGACAAGTGCGGCGCGGCAG
TTGCAGGCAAATTATGTGTGAGGCAATGCCGCGGGCATGTCTACACCGAAATCAAATTACGGCAACCTCT
ATTCACTTATTTGCTTAGTTTTTTGCGCAGTGAGCGGCCAGCGCCTTGTATTGGCATCTAATTATTCCGT
TTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGATCTAAATTTCCCCGCTTTTGGCTTGAA
TAAATTAATCGAATTCCCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATT
GCAAATTGCTCAATCCGCCTAATTGATGTGAGCCCATGCAATAGTTTTGTGCCAATCATTTTTAGTACAC
CCCTAACTGGTGTTTTTTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTACCAACTGAA
GAAGACGATTGTGGGGGTGCGTTTGGCGCAGTGCATGCCCGCGAGCAGGAAATCTCTTTCTCGGCCTGTC
TGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTCAGGGCCGTGGA
CATTGAACTTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGCTGCGCGGCATG
GGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCTGGCCAACACTGGGGGGAGAAAAT
AAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGGAAAAAGGAAAACAAAACATGAAACACGGGCCGGGCA
ATGTCACTCGGCATTCGCTTGATTTTCCGGCTAACTCGCAGCGGTCCTGTGTGTAAATAATGTCTAATGT
TGCATGCCGGTTGCATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTTTACTTTCTGC
CATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCACTTGATAATGATGTTTTGTTTT
TCGCCGGGACAAACTCGTTTCGATTATGGGGAAAAGCGCGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATTGAAAACTGCAAACTGCAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCACACGAC
GAGATGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTGCAGGTTC
AACGGCTGCCGGCGATCGCATCATTAAGTGCGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGCACAC
ACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCTGGCTTTTGCGGACCTGCCGATCGCAATAAATTAAG
GCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCAGA
ATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCGGGCCAAGAGTTTTTCGCATTTTGCATAAAATCG

GCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGGAAAAGAAGTAAAAAATATTTAAAAATATATAT
AAATTTATGGCATGTTCTGTACAGAACTTATGAAGCTAATTTGAAATACTTCTTCTTAGGAAACTGTCCC
TAGGAATATTTGTTTTCCCCAGCATTGCTCAATATTTCCTCCATCTTTTTGCTTATTTCCCAGACATTTT
CTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAAACGT
ATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTT
GCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGC
GGCCATGGCCATTAGAGCATTACCCACTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGA
GTATTGCAACTTACGGCCAGCTGAGTGGAGTGCCGGAACGCACTTCTTTAATTTCGGCGGTTATGTAACC
TCGAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCC
TCGCCGGACTGTGAATCATCGGAGCTGCCCAATCGAAATCAAAGCCAAGCCAATCGAAGCCCAGGTAATT
CATATTTAATGCTGTTAGCAAAGACCTGTCCTAGATACTCTGTATAAAAGTATAATTATTAAGTGCATAT
CAGGTTTATTTACATTTATATCGTATAATATTGGTAAGTGCAGCAGATGCTGTGATCTAAATTTAGAGTC
ATTTAAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAAACTAAAAGCTTTGGATGAAGTAAG
AAAGCCATAAAGCCTAAAATAATATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTATTTCGTAC
CTACGTTGCATGGTATTCAATAAAGATTCAAAAATACTCTCACCCACTGTAAGTGAACCCAGTGTTTCGT
AATTGCCTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTCGGAGAATTCGGTG
TGCCAAAAACCAAAGACCAAAGACCATACCCTTTCAAAACCTTATGAAAAATGGCAAGCCCGGCGAAAGG
TGTTGGCCGGTCCAGGGGATTCGGGGGCCCGAGATACTCGCACTTAATAAACATGCGTGAAAATCAATCA
GCGAAGACAAAAGCCACGCACTAGAAGAAACCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATA
GACCATAAAGTCCGCATGGTGGACCACCACCCGAGCCACCGAAAGCAGCCGAATGGCTGAAACCCCGAAG
TTGGCGCCTTCGTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAAAACCTCATATAAAACGT
GGCCGACATATTGTGGCCAACAGTC

Primers	PrimerID	Sequence
Actin42_FWD	PID55-actin42A_fwd:	GCGTCGGTCAATTCAATCTT
Actin42_REV	PID56-actin42A_rev:	AAGCTGCAACCTCTTCGTCA
y-CDS-D.melanogaster_FWD	RT_007-ymelCDS_Fwd_1:	TGGAGCCAGCTGGACTTTG
y-CDS-D.melanogaster_REV	RT_008-ymelCDS_Rev_1:	GTCAAACTGCGGTCCATGTT
y-CDS-D.biarmipes_FWD	RT_011-ybiaCDS_Fwd_1:	TGGAACCAGCTGGACTTCG
y-CDS-D.biarmipes_REV	RT_012-ybiaCDS_Rev_1:	CAGGGTGCGGTCCATGTT

Primers for reporter constructs			
1	*y 5'*^{mel}	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	TCCGGATCCCTGGGTTTGGTATGATTTTTG
2	peaks-(1+2)^{mel}	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	TCCGGATCCCTAATGATGCGATCGCTG
3	peak-1^{mel}	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	TCCGGATCCCGATTGAATAGAGAAACTTTATT
4	peak-2^{mel}	F	ACCTGGAATTCGGACAGTGGAAATTGACT
		R	TCCGGATCCCTAATGATGCGATCGCTG
5	peak-1+164 bp^{mel}	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	TCCGGATCCTGGCATGGCCAAATTCGG
6	*y 5'*^{bia}	F	TTCCGTCTCTAATTCTGGTGCCCGAGTGTTGTGACTG
		R	TCCGGATCCCCGGGCTTGCCATTTTTCATAAGGTTTTGA
7	peaks-(1+2)^{bia}	F	TTCCGTCTCTAATTCTGGTGCCCGAGTGTTGTGACTG
		R	TCCGGATCCCTTAATGATGCGATCGCCGGC
8	peak-1^{mel}+peak-2^{bia}	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	CTTGAAGACCTATTTCGATTGAATAGAGAAACTTTATTT
		F	TTCCGTCTCTAAATTGCAAATTGCTCAATCCGCC
		R	TCCGGATCCCTTAATGATGCGATCGCCGGC
9	peak-1^{bia}+peak-2^{mel}	F	TTCCGTCTCTAATTCTGGTGCCCGAGTGTTGTGACTG
		R	CTTGAAGACCTGTCCACATTGAAACAGACGAGA
		F	CTTGAAGACCTGGACAGTGGAAATTGACTATT
		R	TCCGGATCCCTAATGATGCGATCGCTG
10	peak-1+164 bp$^{mel\ KO\ sit}$	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	CTTGAAGACCTCGATTGAATAGAGAAACTTTATT
11	peak-1+164 bp$^{mel\ KO\ sit}$	F	ACCTGGAATTCTGGTAAATCAATGATTCGACT
		R	CTTGAAGACCTCGATTGAATAGAGAAACTTTATT

Primers for RT-PCR			
1	Actin42_FWD		GCGTCGGTCAATTCAATCTT
2	Actin42_REV		AAGCTGCAACCTCTTCGTCA
3	y-CDS-D.melanogaster_F		TGGAGCCAGCTGGACTTTG
4	y-CDS-D.melanogaster_R		GTCAAACTGCGGTCCATGTT
5	y-CDS-D.biarmipes_FWD		TGGAACCAGCTGGACTTCG
6	y-CDS-D.biarmipes_REV		CAGGGTGCGGTCCATGTT

Synthesized fragments

1	Fragment_164bp$^{mel\ KO\ site\ 1}$	CTTGAAGACCTATCGGGACAGTGGAAATTGACTATTTTATTT ATATTAATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGT TACGCTAATAGTTTGAGCGCAGTGCATGTCATGGGGACATGT GCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATT TGGCCATGCCAGGATCCGGA
2	Fragment_164bp$^{mel\ KO\ site\ 2}$	CTTGAAGACCTATCGGGACAGTGGAAATTGACTATTTTATTT ATATTAATGAACTTATTTTTAATTTGGCTTAAGTTACTAAGG GGTATGCCGGTGATGAGCGCAGTGCATGTCATGGGGACATGT GCAATTGTGTGTAAGCGGGAAGTGATCGCGGCCTTCCGAATT TGGCCATGCCAGGATCCGGA

Paper II. 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

Citation: 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).

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).

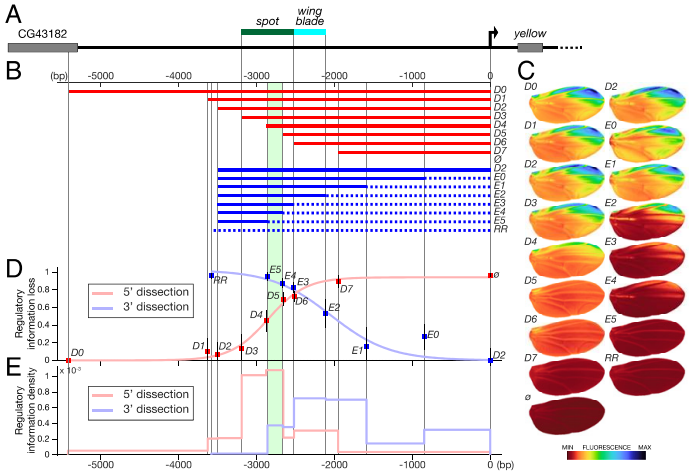

Fig. 1. Quantitative mapping of wing regulatory activities at the *yellow* locus. (*A*) The top line represents the 5′ region of the *yellow* locus from *D. biarmipes*; the green and blue bars indicate the respective locations of *spot* and *wing blade*, respectively, as originally mapped (6). (*B*) 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

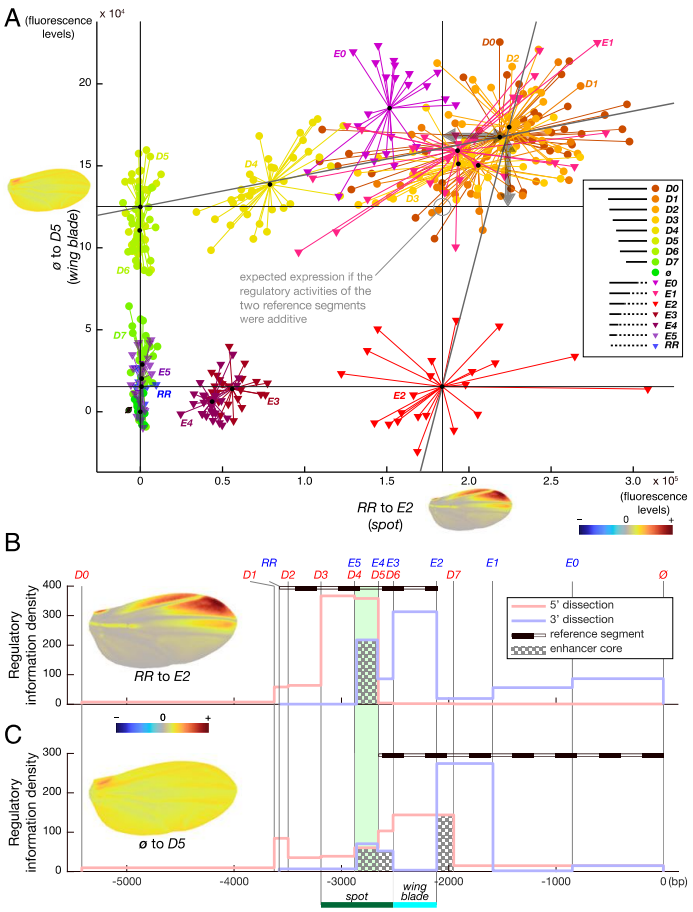

Fig. 2. *wing blade* and the *spot* activities map to overlapping sequences. (*A*) Representation of *wing blade* activity as a function of *spot* activity. Independent estimates were produced by projecting the PCA phenotypic space (PCA in *SI Appendix*, Fig. S1*A*) on a two-vector basis defined by two 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*

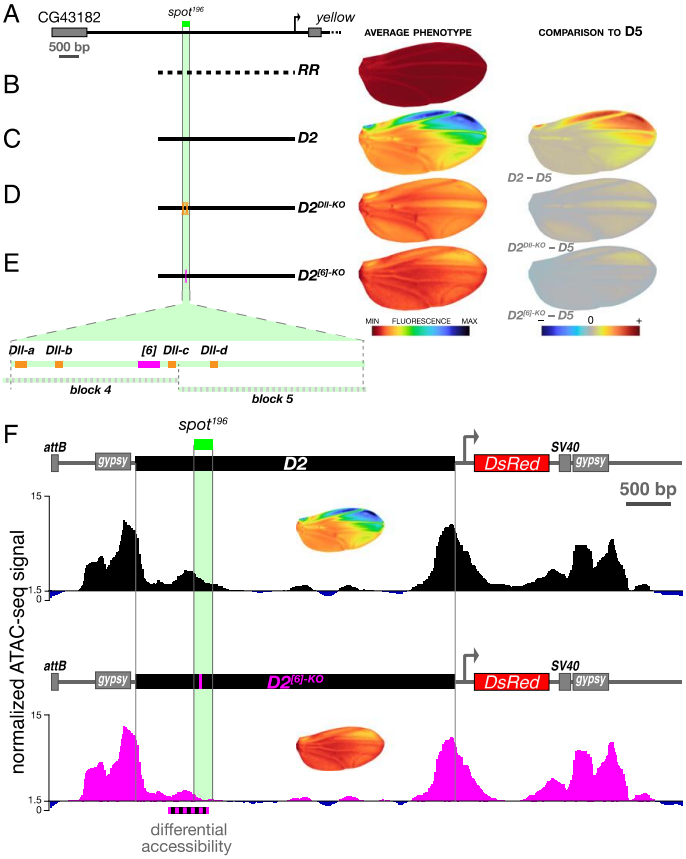

Fig. 3. Shared regulatory inputs of the *wing blade* and the *spot* activities. (*A*) A map of the *yellow* locus 5′ region highlighting the position of the $spot^{196}$ core. (*B–E*) The *wing blade* and the *spot* activities are strongly affected by discrete mutations in *D2*. (*Left*) Construct schematics. (*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

DEVELOPMENTAL BIOLOGY

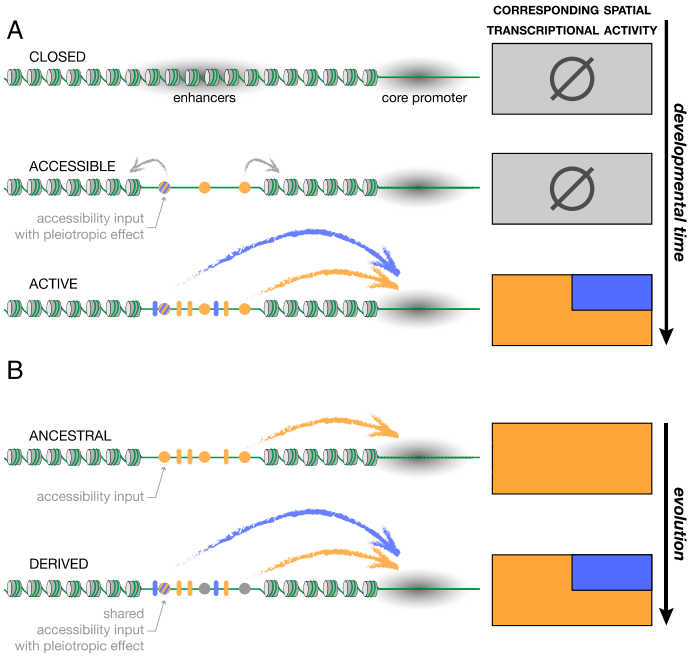

Fig. 4. Developmental enhancer pleiotropy and evolutionary enhancer co-option. (*A*) The developmental progression toward the activation of two interdependent enhancers inferred from our results. High nucleosome occupancy prevents access of transcription factors to the enhancers sequence (closed state; 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 $\frac{Q}{1+e^{-s*(t-t_{half})}}$, 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 cutide 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).

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

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

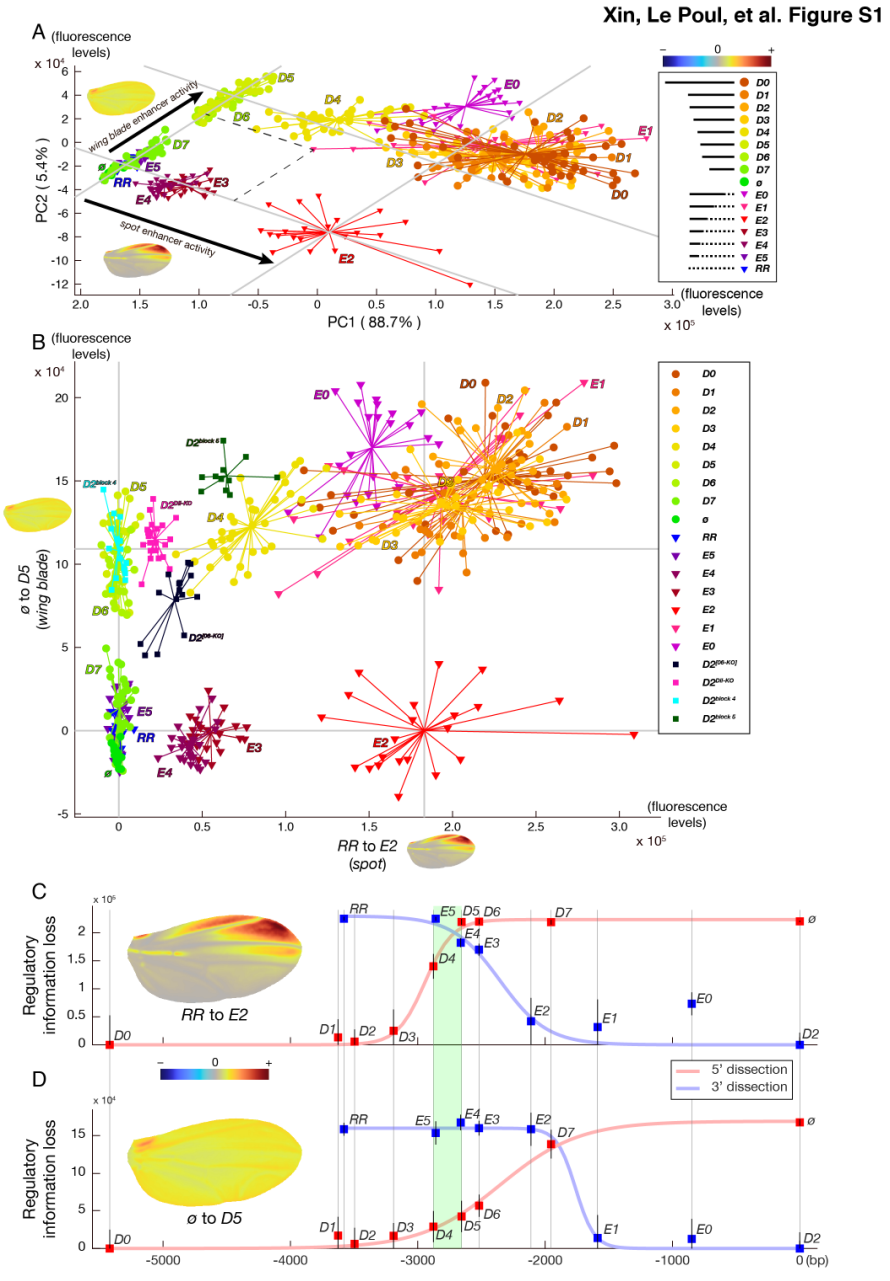
Xin, Le Poul, et al. Figure S1
A (fluorescence levels)
PC2 (5.4%)
PC1 (88.7%)
wing blade enhancer activity
spot enhancer activity
(fluorescence levels)
B (fluorescence levels)
ø to D5 (wing blade)
RR to E2 (spot)
(fluorescence levels)
C
Regulatory information loss
RR to E2
D
Regulatory information loss
ø to D5
5' dissection
3' dissection
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

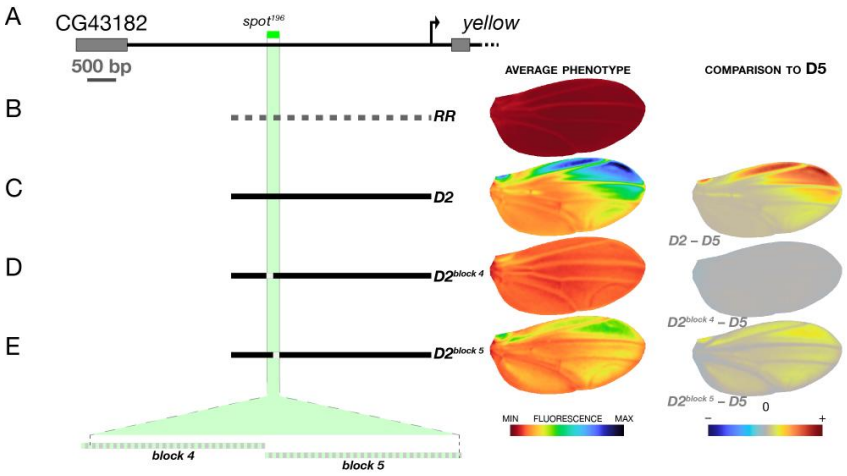

Fig. S2. Reporter activity in the wing for constructs *D2*$^{block\ 4}$ and *D2*$^{block\ 5}$. (A) A map of the *yellow* 5' region highlighting the position of the *spot*^{196} core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 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.

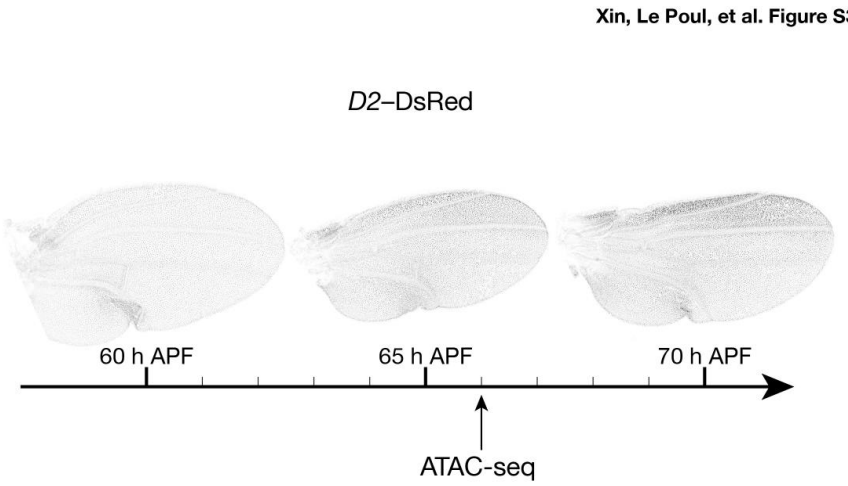

Fig. S3. Reporter activity in the wing for constructs *D2*$^{block\ 4}$ and *D2*$^{block\ 5}$. (A) A map of the *yellow* 5' region highlighting the position of the *spot*^{196} core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 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

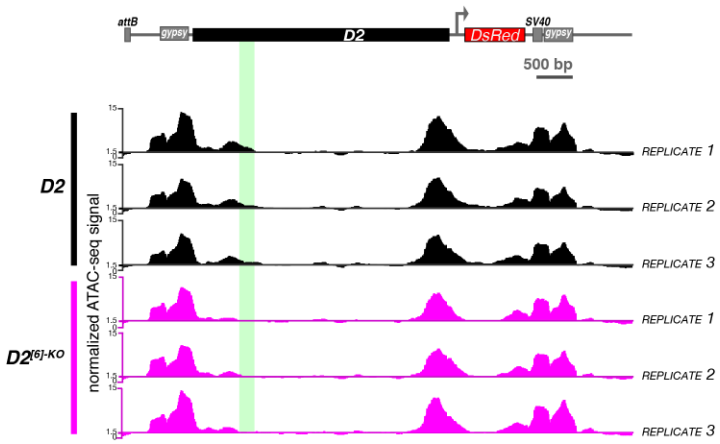

Fig. S4. ATAC-seq replicates shown separately. Chromatin accessibility at the *D2* and *D2*$^{[6]\text{-}KO}$ transgenes at the onset of *spot* activity. The 6 ATAC-seq traces represent 3 technical replicates for each transgene that were pooled for each genotype in Fig. 3.

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 clining, 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	
$D2^{DII-KO}$		
fragment1-F (D2-Forward)	gtctagagcccgggcgaatt CCTTAGTTCAGAAGGCGCCTGC	
fragment1-R	CCTTAAACGGAATcgcgAGATGCCAATACAAGGCGCTGGC	
a fragment encompassing the 4 DII site mutations was synthesized from IDT ;$D2^{DII-KO}$	TCTcgcgATTCCGTTTAAGGACGCcgcgTTCTGAGCTAAAACTCGCTTATG GAGAGATCTAAATTTCCCCGCTTTTGGCTTGAATAAAcgcgTCGAATTCC CCGCTGGCTcgcgAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTA AATTGCAAATTGCTCAATCCGCC	
fragment3-F	TGCAAATTGCTCAATCCGCCT	
fragment3-R (D2-Reverse)	gctcctcgagggatcTTCTGTGGACCGTGGCGC	
$D2^{HG-KO}$		
fragment1-F (D2-Forward)	gagcccgggcgaattCCTTAGTTCAGAAGGCGCCTGC	
fragment1-R with mutation adapter	TTATTtagtgtaaagatGGGGAAATTTAGATCTCTCCAT	
fragment2-F with mutation adapter	atctttacactaAATAAATTAATCGAATTCCCCGCTGG	
fragment2-R (D2-Reverse)	gctcctcgagggatcTTCTGTGGACCGTGGCGC	
$D2^{block\ 4}$		
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	
$D2^{block\ 5}$		
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

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
TATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTTGCACCAGCTGCAGTTGGGCCAAGAGA
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
TTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGGTTTTCGCACTCGCTTG
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-D1lKO
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 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

Citation: 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.)

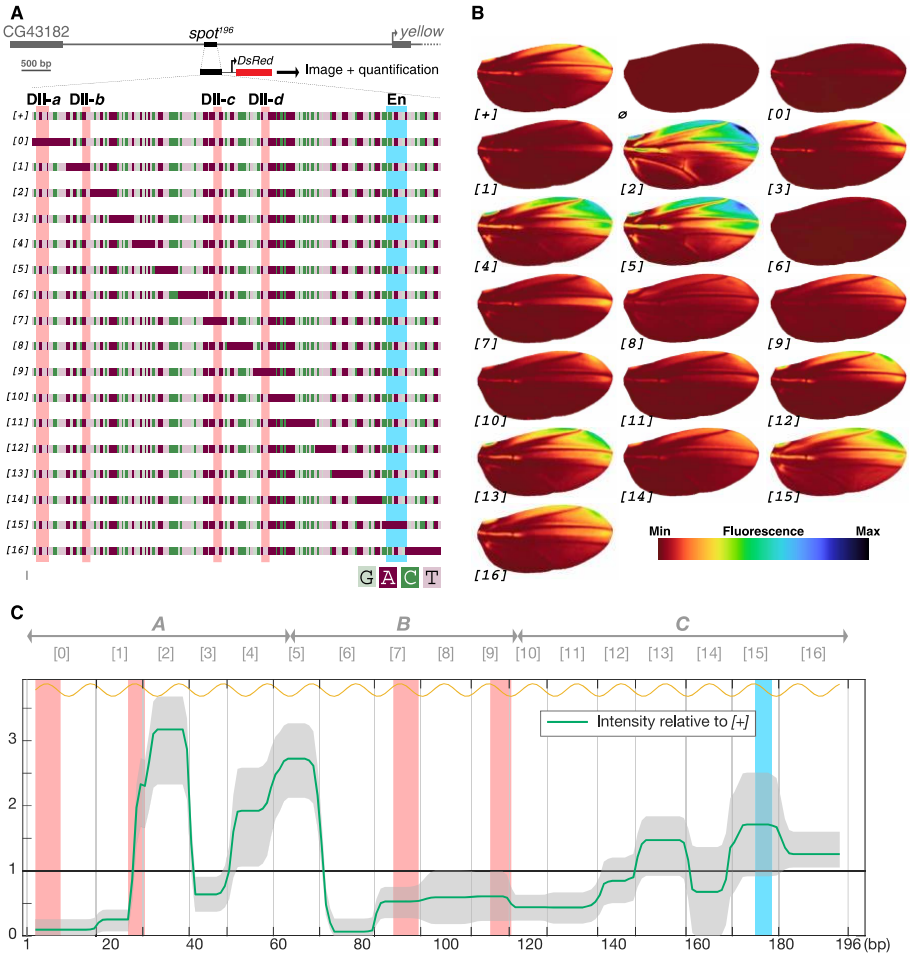

Fig. 1. A mutational scan of the *D. biarmipes spot*^{196} enhancer with a quantitative reporter assay. (**A**) Wild-type (*[+]*) and mutant (*[0]* to *[16]*) versions of the *spot*^{196} enhancer from the *D. biarmipes yellow* locus (depicted at the top) were cloned upstream of a DsRed reporter to assay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

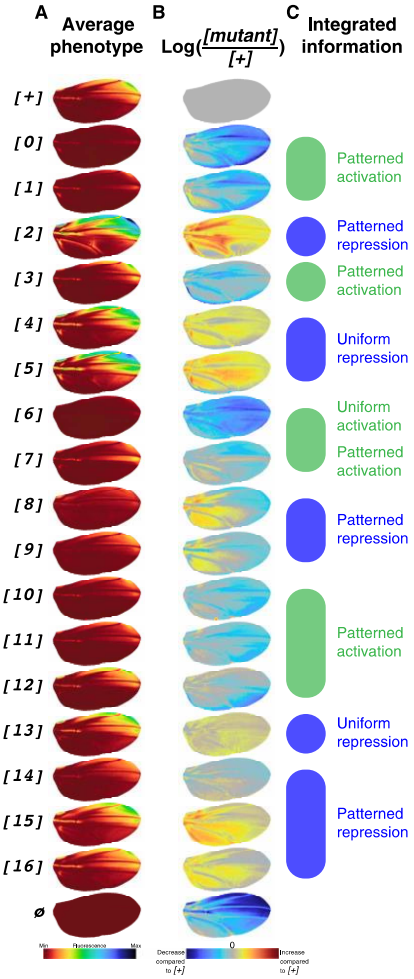

Fig. 2. Trans-regulatory integration along the *spot*^{196} sequence. (**A**) Average phenotypes reproduced from Fig. 1B. (**B**) *logRatio* images [log([*mutant*]/[*+*]) for intensity values of each pixel of registered wing images] reveal what spatial information is integrated by each position along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

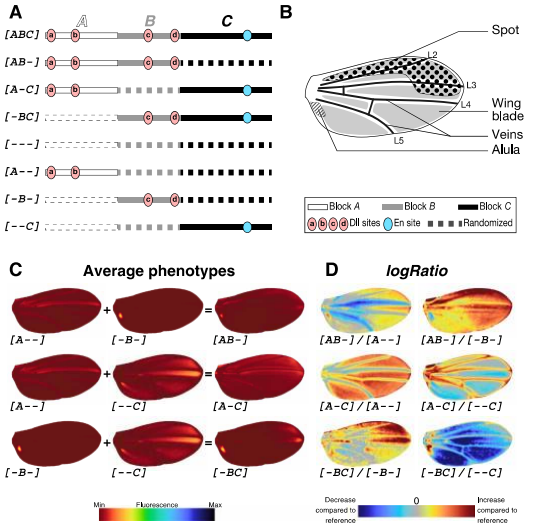

Fig. 3. Regulatory interactions in the *spot*^{196} sequence. (**A**) Schematics of constructs with block randomizations. The *spot*^{196} sequence was arbitrarily divided into three blocks (A, 63 bp; B, 54 bp; C, 79 bp). In each construct, the sequence of one, two, or all three blocks was randomized.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

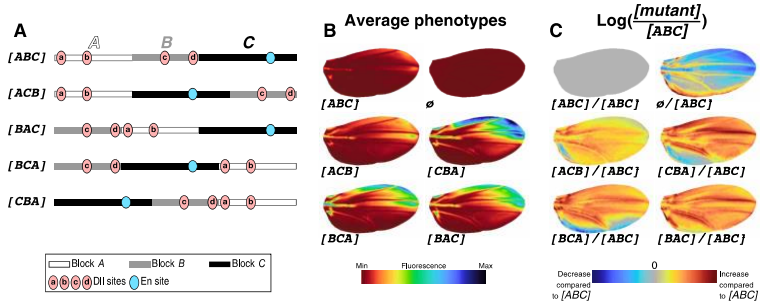

Fig. 4. Block permutations scale the activity of the *spot*^{196} enhancer. (**A**) Schematics of constructs with block permutations. In this series, the same blocks of sequences as in Fig. 3A were permutated. (**B**) Average phenotypes resulting from constructs in (A). Colormap of average phenotypes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

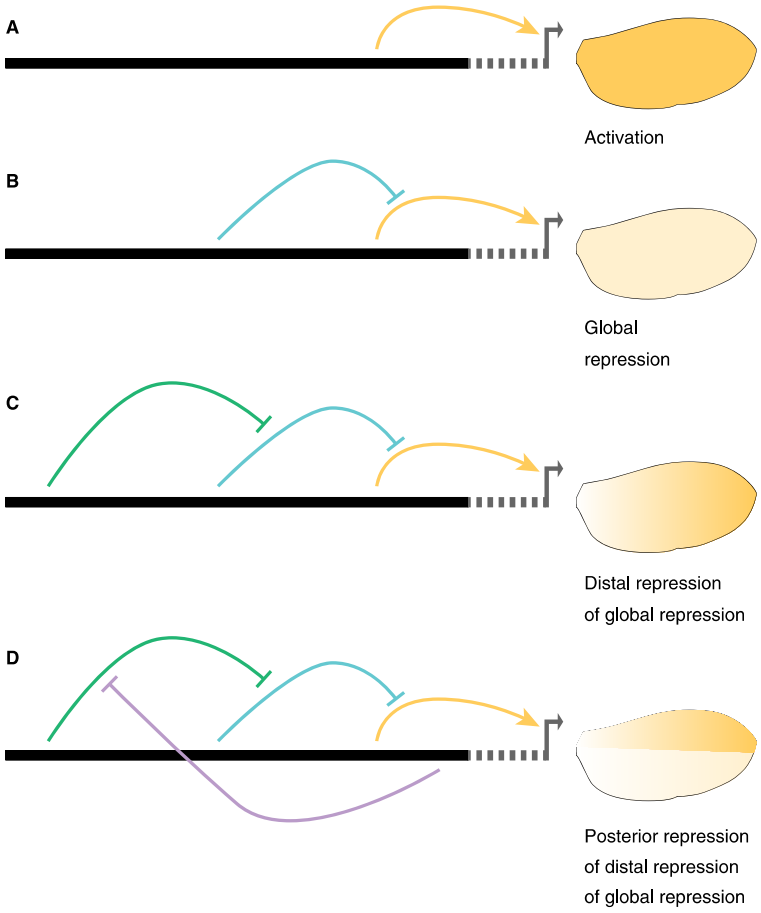

Fig. 5. A model of the regulatory logic governing the *spot*^{196} enhancer. (**A** to **D**) The schematics show step by step how regulatory information and interactions integrated along the enhancer sequence produce a spatial pattern of activity. (A) Three independent inputs, respectively, in blocks *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).

ScienceAdvances

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

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

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

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

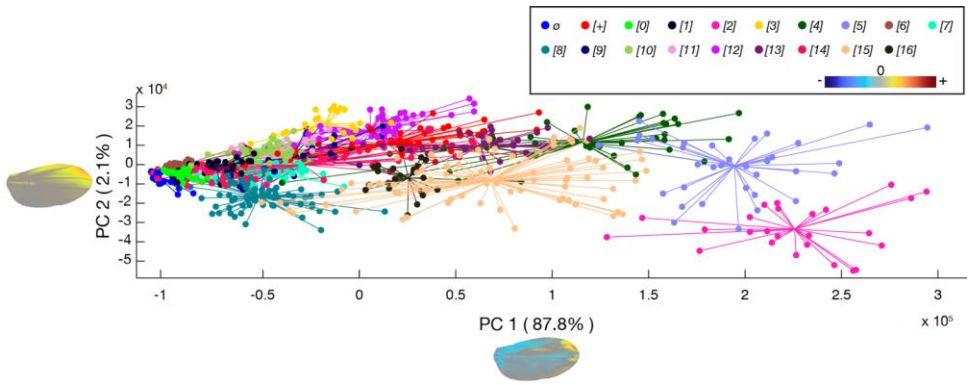

Figure S1. **First two axes of variation in a principal component analysis of all individual wings used to generate the average reporter expression of Figure 1**. Each wing is depicted by a colored dot, and each construct by a color. PC1 captures 87.8% of the variation and corresponds to overall changes 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.

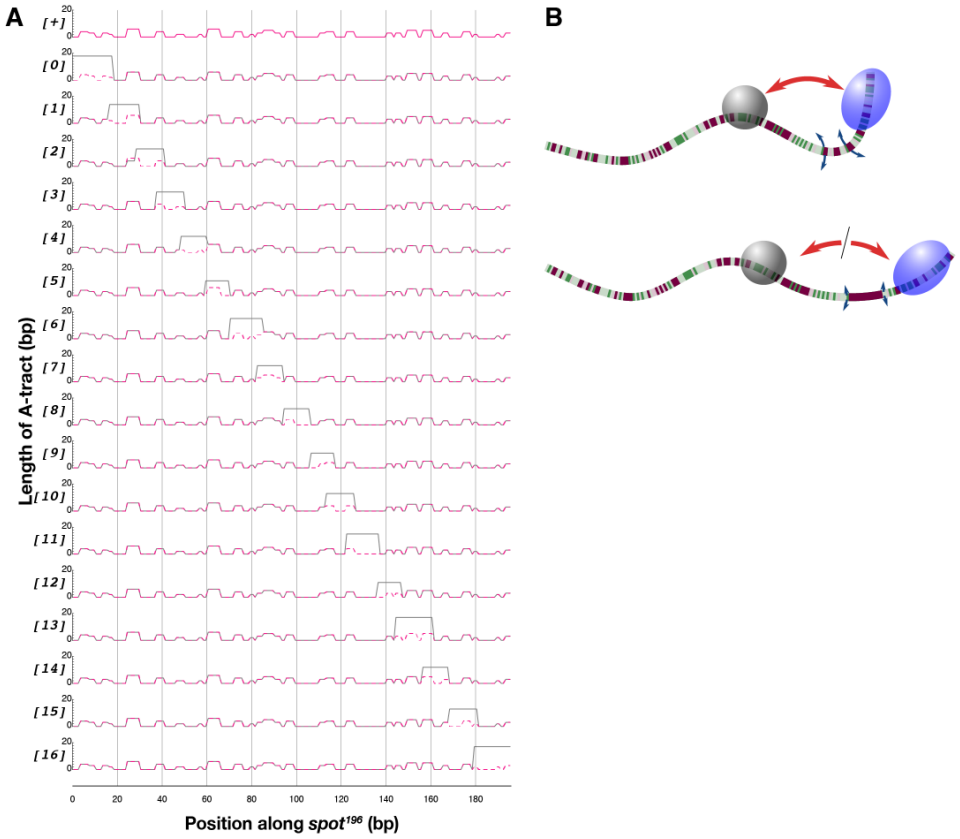

Figure S2. Local rigidity along the wild-type and mutant *spot*^{196}. (**A**) Each graph is a plot of the length of the longest consecutive A_nT_n sequence that a base pair participates in, a proxy for sequence rigidity at this position. The first graph on top is the wild type (*[+]*) alone. The remaining 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.

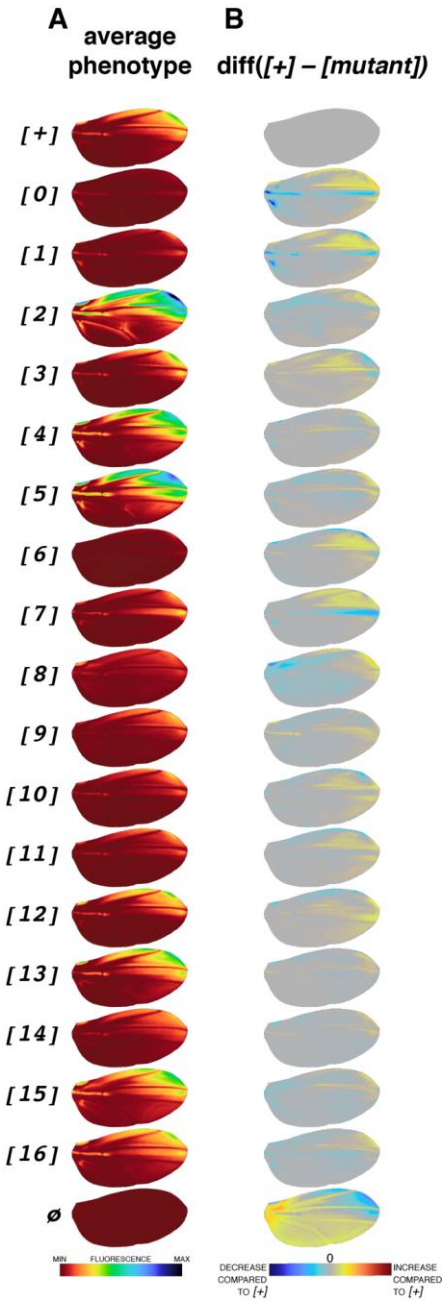
A
average phenotype
B
diff([+] – [mutant])
[+]
[0]
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
ø
MIN
FLUORESCENCE
MAX
0
DECREASE COMPARED TO [+]
INCREASE COMPARED TO [+]

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.

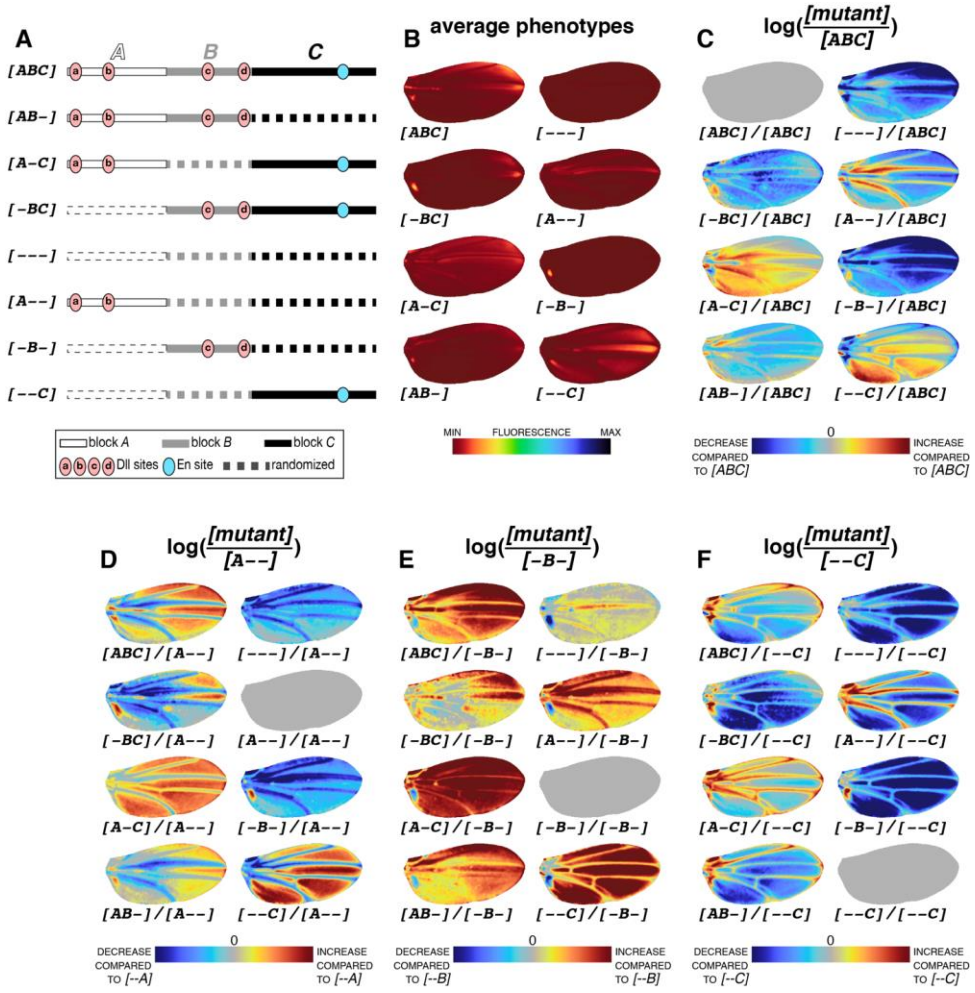

Figure S4. ***logRatio* of all block constructs.** (**A**) Schematics of block constructs repeated from Figure 3A for legibility. (**B**) Average phenotypes of constructs shown in (**A**), repeated from Figure 3B for legibility. Colormap of average phenotypes normalized for all constructs of the block 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
CTAAATTTCCCCAAAAAAAAAAAAAAATAAATTAATCGAATTCCCCGCTGGCTATTAA
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
AAAAAAAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[11]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAAAAAAAAAAATCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[12]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGAAAAAAAAAAATGTAAATTGCAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[13]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAAAAAAAAAAAAAAATTGCTCAATCCGCC
TAATTGATGTGCGCCCATGCAAT

>*spot*^{196} *[14]*
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACTCGCTTATGGAGAGAT
CTAAATTTCCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTAAA
ACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAAAAAAAAAAACCGCC
TAATTGATGTGCGCCCATGCAAT

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

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

- 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

1. Summary of results

This dissertation explored the mechanisms governing the loss of the wing pigmentation pattern in *D. melanogaster*, focusing on the dynamic changes occurring in chromatin during wing development. Over the past two decades, studies have used wing pigmentation patterns as a model system to illustrate how enhancer variation resulted in the evolution of morphological traits. At the same time, the development of new NGS-based techniques allowed studies to investigate the gene transcriptional regulation genome-wide. The resolution of these techniques, such as ATAC-seq, can reach a very high level, up to several hundred base pairs in this study. Our understanding of gene transcription regulation has been growing fast. However, few of these techniques have been applied to late-stage pupal wings, where the pigmentation-related enhancers are activated. In this dissertation, I modified the published protocols for several NGS-based methods and successfully applied them on pupal wings from 0h APF pupae to 78h APF pupae.

In Paper I of Results, by conducting an extensive analysis of chromatin accessibility changes in pupal wings from both spotted and non-spotted species, I have then unveiled a previously unknown process. Specifically, this study has revealed a novel evolutional pathway where the evolutionarily gained *spot* enhancer in *D. melanogaster* becomes silenced by a nearby enhancer during the late stage of pupal development. What sets this discovery apart is its departure from the previously documented pathways associated with the loss of activator binding sites and the gains of repressor binding sites. Instead, the silencing was achieved by increasing the accessibility of the nearby enhancer, allowing for the recruitment of repressors and repressing the *spot* enhancer activity.

Additionally, in Paper II of Results, I examined the function of an element in *spot* enhancer from *D. biarmipes* by measuring the chromatin accessibility changes before and after mutating this element. The results showed that this element is involved in regulating the chromatin accessibility of both the newly evolved *spot* enhancer and the ancestral *wing blade* enhancer. These two enhancers regulate *yellow* expression in the wing with distinct patterns, indicating this element functions with a pleiotropic effect.

In Paper III of Results, I worked with my colleagues to further investigate the arrangement of transcription factor binding sites in the minimum *spot* enhancer from *D. biarmipes*, $spot^{bia196}$. We introduced a series of mutations across the enhancer and analyzed their effect on enhancer activity by quantifying the reporter expression pattern driven by $spot^{bia196}$. The results showed that most mutations altered the reporter expression pattern quantitatively if they did not abolish the enhancer activity. This finding suggested that

contrary to the previously proposed flexible models, the information instructing the enhancer quantitative activity level was densely distributed in *spot*^{bia196}.

The main focus of this dissertation is to investigate the complex mechanism of transcriptional silencing that resulted in the loss of the pigmentation trait in *D. melanogaster*, as summarized earlier. Unraveling the precise mechanism behind the loss of pigmentation can be challenging, and the findings presented in this dissertation only provide a partial understanding of this process. In this discussion, I will delve into the intricacies of this silencing process from several perspectives of gene regulation, including dynamic gene expression, transcription factor regulation, the dual role of enhancers as silencers, and the potential interactions between enhancers.

2. The dynamic gene expression during pupal wing development

The gene expression is precisely regulated to switch on and off during development and cellular processes. This dynamic regulation is governed by *cis*-regulatory elements that work in concert with specific transcription factors and chromatin proteins. As discussed in the Introduction (Section 4) of this dissertation, the chromatin accessibility landscape is a distinct characteristic that helps to categorize cells based on their developmental or differentiated stages. Furthermore, alterations in chromatin accessibility in enhancer regions are essential in guiding the dynamic expression of genes[153,158].

Understanding how changes in chromatin accessibility occur over time and space is a fundamental question in molecular biology. Equally significant is the understanding of enhancer sequence variations that influence chromatin accessibility, resulting in diverse gene expression patterns and phenotypes in evolution. While studies focusing on the former question have mainly looked at embryogenesis and cell differentiation[153,158], to address the latter, we need to use appropriate evolutionary developmental model systems.

Several studies have been conducted on HARs during human brain development[166,167]. The facial development of chimpanzees and humans has also been utilized as an evolutionary developmental model[267]. However, these developmental processes are complex and challenging to study. This dissertation presents a *Drosophila* wing development model that uses various wing pigmentation patterns found in *Drosophila* species as a phenotypical readout to study how sequence variation modifies enhancer evolution.

Investigating gene dynamic expression during pupal wing development among closely related species has offered valuable insights. A previous study found that the *wing blade* enhancer of the gene *yellow* in *D. melanogaster* could initiate reporter gene expression in a homogenous wing pattern rather than a spot-like

pattern. Results from this dissertation show that the *D. biarmipes* orthologous enhancer initially induces a uniform reporter expression pattern, followed by a spot-like expression pattern in the 80h to 87h APF wings (Figure S1 in Appendix, also see Paper II of the Results). These findings strongly suggest that distinct elements govern enhancer activities at different stages of development. Additionally, they support the *spot* evolution model proposed in previous studies, which entails the conservation of *wing* enhancers across various species and the acquisition of *spot* enhancers specifically within the *D. melanogaster* group[14,240]. To further validate the observed *yellow* enhancer activity in the enhancer-reporter assay and its relevance to the *yellow* enhancers in their original loci, qPCR analysis was conducted to measure *yellow* expression dynamics during pupal development in both *D. melanogaster* and *D. biarmipes*. The results illustrate the differential *yellow* expression dynamics between these two species, underscoring the intricate and varied regulatory mechanisms governing wing pigmentation patterns, which will be further discussed below.

First, the dynamic expression of *yellow* is consistent with the dynamic changes of the chromatin accessibility at the *yellow* 5' region in both species. As shown in the first paper in the Results [268], the chromatin accessibility in *yellow* enhancers was low in 0h APF pupal wings, implying that *yellow* was not expressed at that stage. Although the 30h APF pupal wing was used in qPCR for a similar expression level of *Actin* among all pupal wings, the low expression level of *yellow* in 0h APF pupal wings was shown in the pilot experiments. The chromatin accessibility in 48h APF pupal wings was much higher than that in the wings from all other stages, also reflected by the observation from qPCR. Of note, the chromatin accessibility at many different loci was the same or even increased during pupal wing development, indicating that the strong signal in the *yellow* 5' region in 48h APF wings was not from an experimental batch effect. The overlap of chromatin accessibility dynamics and *yellow* expression changes supported the hypothesis of DNase I-hypersensitivity sites involved in gene regulation proposed four decades ago, as discussed in the Introduction. However, the ATAC-seq results from 60h to 78h APF wings also show the limitation of bulk ATAC-seq when the gene expression is not uniform but has a spatially-restricted pattern. Without the enhancer-reporter assay, it would be difficult to tell how the different accessibility patterns in late pupal wings would affect the gene expression. A combination of scATAC-seq and scRNA-seq may further help to identify the transcription factors involved in regulating *yellow* enhancer accessibility in late pupal wings from different species.

Additionally, understanding the dynamic expression of *yellow* may provide valuable insights into the activation of the *spot* enhancer and other enhancers in late-stage pupal wings. We can look for other genes with similar expression dynamics by studying gene expression changes across the entire genome. Besides, the transcription factors that govern the activation of the *spot* enhancer may also regulate other enhancers during this developmental stage. A thorough analysis of the enhancers that control developmental genes,

for instance, analyzing transcription factor binding motifs enriched from the ATAC-seq data, can be performed. Further investigations could then concentrate on determining the roles of transcription factors, cofactors, and chromatin proteins in controlling enhancer accessibility and activity throughout the wing development. These efforts could potentially improve our understanding of the mechanisms that regulate dynamic gene expression.

For instance, one question worth studying is how transcription factors cooperate with cofactors and chromatin proteins to activate regulatory elements precisely, such as the *spot* enhancer. The findings of such studies could also help us understand how gene expression dynamics are affected by evolutionary variation among different species. Such comparative analyses also provide a framework for understanding how the sequence variation contributes to the enhancer evolution and thus leads to divergent developmental processes.

3. The role of E93 in regulating the enhancer activity during wing development

Cis-regulatory elements intricately govern gene expression dynamics. These elements have dynamic accessibility changes to control the recruitment of transcription factors, thus regulating the temporal and spatial expression of genes. A group of proteins regulates the activity of these elements, mainly enhancers in this dissertation, ranging from transcription factors to chromatin remodelers. However, transcription factors are primarily responsible for deciphering the instructions encoded in enhancer sequences. Pioneer factors can recognize their binding motifs even when DNA is wrapped around nucleosomes[103]. This interaction with nucleosome-bound DNA promotes local accessibility of targeted enhancers, facilitating subsequent binding of other transcription factors[103]. However, pioneer factors may not act alone in modulating enhancer chromatin accessibility[116]. Transcription factors recruited by specific enhancers can also contribute to regulating chromatin accessibility, in addition to their function in activating or repressing enhancers[116,120]. The arrangement of these transcription factor binding sites with different functions within enhancers is still a longstanding question in the field of gene transcriptional regulation[138].

In *Drosophila*, transcription factors such as Dorsal and E93 are known to regulate chromatin accessibility[118,260]. Interestingly, both Dorsal and E93 present dual functions, simultaneously activating and repressing different subsets of enhancers within the same nuclei. A recent study by Brennan *et al.* elucidated how Dorsal decodes enhancer sequences to carry out these opposing functions[118]. When acting as an activator, Dorsal modulates enhancer chromatin accessibility. However, in enhancers containing a repressor motif adjacent to the Dorsal motif, its influence on chromatin accessibility is negated despite binding to these enhancers.

Similarly, previous research on E93 revealed its ability to function as both an activator and a repressor within the same nuclei[260]. Yet, the specific sequence features that determine E93's function remain elusive. Furthermore, it remains uncertain whether the decoding mechanisms identified during embryogenesis are applicable to tissue differentiation processes later in life.

The transcription factor E93 plays a crucial role in the developmental stages of the *Drosophila* wing, particularly during the prepupal and mid-pupal stages[260]. It orchestrates changes in chromatin accessibility, and its expression responds to the EcR pulses, a steroid hormone process necessary for metamorphosis[256]. The published data, including E93 ChIP-seq and related FAIRE-seq from wing imaginal discs and pupal wings, shows that E93 binds to the *yellow* 5' region in early pupal wings[260]. When E93 is mutated, accessibility is lost in 44h APF wings[260]. In addition, there is a correlation between the appearance of prominent accessibility in 47h APF wings at the *yellow* 5' locus and the initiation of *yellow* expression (Results of this dissertation), supporting E93's role as a major transcription factor in regulating *yellow* 5' accessibility in pupal wings.

In the context of *peak-1+164-bp*^{mel}, the activity of *spot* in *peak-1*^{mel} is suppressed by a 164-bp element (Figure 7). Knocking down E93 in this strain resulted in an increased reporter gene expression, which showed a spot-like pattern. However, some questions remain to be answered regarding the role of E93 in regulating the *yellow* enhancer. It is yet to be determined whether E93 binds directly to the 164-bp element or influences the enhancer activity by regulating upstream transcription factors. Further investigation is required to identify the E93 binding motifs within this 164-bp element and to gain insights into the evolutionary acquisition of these binding sites.

Following a methodology similar to that used by Uyehara *et al.*[260], I used an E93 protein-trap strain available in the Bloomington *Drosophila* Stock Center from the *Drosophila* Gene Disruption Project[269]. This stock expresses E93 fused with GFP at the endogenous site[269], allowing the profiling of E93-DNA binding using an anti-GFP antibody. Specifically, it allowed me to employ a modified version of the CUT&RUN, greenCUT&RUN[270,271]. The greenCUT&RUN involves using a nanobody against GFP coupled with MNase, which differs from the conventional approach of using an antibody and MNase separately in standard CUT&RUN. Commercial GFP antibodies were found to introduce significant noise in the ChIP-seq experiment, demanding a higher amount of input sample to overcome the noise.

While successfully implementing CUT&RUN against H3K27ac and Dll in 72h APF wings, I encountered certain challenges with greenCUT&RUN. I reasoned that the size of the conjugated protein is larger than the antibody or the MNase alone, making it more difficult to pass through the nuclear envelop. This may

lead to the low amount of DNA harvested from the experiment. An increased digitonin concentration in the experimental setup likely improved the results by enhancing membrane permeability. The yield of the DNA then allowed the following preparation of libraries for sequencing. However, the observed noise persisted throughout the genome. The findings presented in this dissertation on greenCUT&RUN indicated a mere enrichment of E93 in the *yellow* 5' region. It is worth noting that the greenCUT&RUN used in this dissertation included an extra light cross-linking step, which was not part of the original protocol. Further improvement in data quality is essential to establish a more precise E93 binding profile with enhanced resolution in the late pupal wings.

Throughout my dissertation, I attempted to locate E93 binding sites within the 164-bp element. I mutated three predicted E93 binding sites in the enhancer-reporter assay with *peak-1+164-bp*^{mel}, but I found that neither single nor triple mutations increased *spot*^{mel} activity. One possibility is that E93 binds to an unconventional or low-affinity site that was not predicted using the motif from JASPAR. It has been reported that low-affinity binding sites can dominantly regulate developmental gene transcription. In a parallel study, I found that a cluster of low-affinity sites in *spot*^{196bia} is crucial for the enhancer activity (manuscript in preparation). Therefore, it is possible that E93 may bind to a cluster of low-affinity binding sites in the 164-bp element, surpassing the predictive capacity of its known binding matrices.

However, we cannot rule out the possibility that E93 indirectly regulates the *yellow* enhancer accessibility. As a responder of EcR pulses, E93 governs downstream genes, including transcription factors. A strain used in this dissertation (Paper II of the Results) carried a 3 kb fragment from *yellow*^{bia} 5' enhancer fused with DsRed termed D2, and the DsRed in this strain expressed in a spot-like pattern with wing blade signal. In this context, reducing the E93 expression level led to an overall increase in expression across the wing blade. Confusing enough, the signal only decreased in the region between L1 and L2 (unpublished data). The recent study by Brennan *et al.* showed that transcription factor Dorsal can be an activator when it regulates accessibility while being a repressor when the cofactor binds nearby[118]. Likely, the role of E93 is also context-dependent. These findings and observations from the *peak-1+164-bp*^{mel} strain indicate the complex regulatory role of E93 at the *yellow* 5' loci, potentially modulating the enhancer activity through both direct and indirect pathways.

To conclude, further investigation is required to understand the role of E93 in silencing *spot*^{mel}. The improvement to the greenCUT&RUN protocol for E93-GFP could provide high-resolution profiles and help in revealing the *in vivo* binding preferences of E93. Additionally, an *in vitro* E93 binding assay, such as SELEX-seq[272], might help identify low-affinity binding sites. Furthermore, evaluating the changes in accessibility before and after the depletion of E93 in late-stage pupal wings could offer important insights.

An RNAi screen targeting transcription factors that exhibit upregulation upon E93 downregulation in pupal wings may also help identify downstream transcription factors that potentially regulate *yellow*^{mel} 5' accessibility in late pupal wings.

As discussed in Section 1, elucidating the impact of evolutionary sequence variation on enhancer function represents a future direction of this dissertation. Investigating how enhancer sequences modulate the function of E93 contributes to our understanding of enhancer syntax in the context of wing development. This, in turn, may facilitate the study of how evolutionary variations influence the dynamics of differentiated enhancers, leading to diverse gene expression patterns among species.

4. The silencing of the *spot* enhancer in *D. melanogaster*

This dissertation uncovered an unconventional pathway contributing to losing a morphological trait during evolution. In contrast to another non-spotted species, *D. gunungcola*, which lost its the spot through direct mutations, presumably of binding sites within the *spot* enhancer[240], *D. melanogaster* retains a *spot* enhancer capable of driving reporter gene expression in a spot-like pattern. However, in a previous study, confusion arose due to the proximity of the element silencing *spot*^{mel} activity to the 3' end of *spot*^{mel}, leading to the assumption that *D. melanogaster* had lost the *spot* enhancer[14]. Nevertheless, careful analysis of *yellow* expression dynamics and chromatin accessibility changes in this dissertation successfully distinguished the *spot*^{mel} enhancer from the *yellow*^{mel} 5' locus, thereby preserving the *spot* activity. Additionally, a 164-bp element located at the 3' end of *spot*^{mel} was identified, indicating its potential to repress the *spot* activity. However, subsequent experiments revealed that, unlike *peak-2*^{mel}, this 164-bp element could not silence the activity driven by *peak-1*^{bia} (unpublished data). In this section, the potential silencing mechanisms of *peak-2*^{mel} will be discussed.

4.1. The function of the 164-bp element

The first question of the function of the 164-bp element is whether it also represses *yellow*^{mel} *wing blade* activity. The Results of this dissertation (briefly shown in Figure 7) indicate that either *peak-2*^{mel} or *peak-2*^{bia} could initiate reporter gene expression in the wing blade. Given that the 164-bp element is part of *peak-2*^{mel}, the fragment devoid of this element is denoted as *peak-2-164-bp*^{mel} (Figure 7). Preliminary data suggested that *peak-2-164-bp*^{mel} maintains *wing blade* activity, indicating that the onset of *wing blade* activity is independent of the initiation or repression of *spot*^{mel}. This finding aligns with the observations in the *yellow*^{bia} enhancer, as discussed in Paper II of the Results.

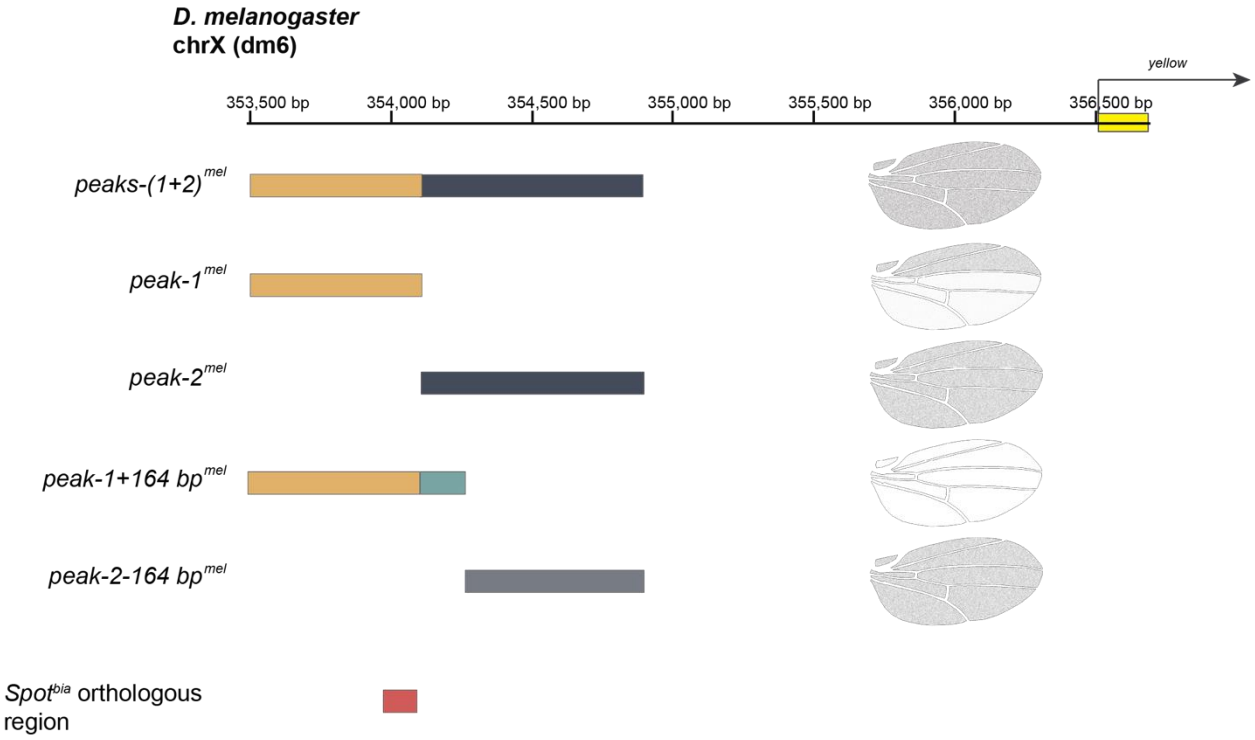

Figure 7. A brief recapture of the results from Paper I. *The map illustrates some constructs designed in this dissertation and their corresponding reporter expression pattern. The peak-(1+2)^{bia} is the orthologous region of the peak-(1+2)^{mel} in D. biarmipes, and so are the other peak^{bia} discussed here. Both peak-1^{bia} and peak-1^{mel} contain elements with spot activity.*

However, the duration of *wing blade* activity in *D. melanogaster* remains unknown. According to qPCR results, the *yellow* gene's expression levels seem to diminish significantly during later developmental stages compared to earlier ones. To facilitate imaging of pupal wings with thick cuticles and slight pigmentation, this dissertation employs DsRed as a reporter gene reflecting enhancer activity[273,274]. DsRed is a relatively stable fluorescent protein that can remain in cells for several days[275]. Consequently, it is still uncertain whether the wing blade signal observed in late pupal wings originates from 40-50h APF wings or subsequent stages. To address this issue, one potential approach would be to use a UAS-TransTimer strain and cross it with Gal-4 lines containing different enhancer fragments for testing. The UAS-TransTimer strain uses a version of destabilized fluorescent proteins that can be degraded within two hours in cells[276]. Therefore, no fluorescent signal should be observed in late pupal wings if the enhancers only function earlier.

Nonetheless, whether active or not, the accessibility of *peak-2-164-bp^{mel}* persisted until at least 78h APF in *Drosophila* wings. This raises the question of whether *peak-2-164-bp^{mel}* continues functioning during the

late stages of wing development. Since the *peak-2* functions in mid-stage pupal wings, as it is the stage when the initiation of *yellow* expression occurs across the wing blade, the 164-bp is presumably not a repression element to the *wing blade* activity during this stage. Suppose *peak-2-164-bp*^{mel} maintains its *wing blade* activity during the late stages of pupal development when the *spot*^{mel} is repressed by the 164-bp element (Figure 8b). In that case, it suggests that the repressors recruited by the 164-bp element may not be effective in repressing *peak-2-164-bp*^{mel} activity. This could be due to either the relative weakness of the repressor or the distance between the repressor and the *wing blade* activator, which may hinder the function of the repressor on the *wing blade* activator. Additionally, the observed reduction in *spot*^{bia} activity by *peak-2*^{mel}, but not the 164-bp element, implies that an activated *peak-2*^{mel} during the late stage could support the enhancer competition model in which *peak-2*^{mel} competes with the promoter with spot enhancers in late-stage pupal wings.

Alternatively, *peak-2-164-bp*^{mel} may only activate the wing blade signal during the mid-stage of pupal wings (shown in Figure 8c). The qPCR results indicate that *peak-2-164-bp*^{mel} may no longer have enhancer activity in late-stage pupal wings. As discussed in the introduction, the key feature of silencers is their chromatin accessibility during functional phases. Considering its accessibility, repressor binding sites, and capacity to suppress enhancer activity across different species would reinforce the idea that *peak-2*^{mel} is a silencer in late-pupal wings. The enhancer competition and silencing models will be further discussed in this section.

Another remaining question for the 164-bp element is its role in maintaining accessibility to *peak-2-164-bp*^{bia} during late pupal wing development. While it is likely that the silencing function is due to a gain of E93 binding sites that facilitate accessibility and a gain of repressor binding sites that suppress *spot*^{mel}, it is uncertain whether this enhanced accessibility extends to the entire *peak-2*^{mel}. In other words, it is unknown whether deleting the 164-bp element would eliminate the entire *peak-2*^{mel} accessibility in late-stage pupal wings. Furthermore, the *peak-2*^{mel} region maintains conserved *wing blade* activity across the *D. melanogaster* group[14,240], raising the question of whether the 164-bp element is the only fragment in *peak-2*^{mel} responsible for losing the *spot* enhancer in *D. melanogaster* during evolution.

In this dissertation, I performed ATAC-seq in transgenic *D. melanogaster* to examine the accessibility of the *yellow* enhancer *D. biarmipes*, which was not impacted by the endogenous *yellow*^{mel} enhancer (Paper II of the Results). This approach enabled the analysis of the accessibility of *yellow* enhancers and the measurement of the activity with the reporter gene using the same strain. To begin the test of the hypothesis that if the 164-bp element regulates *peak-2*^{mel} accessibility in late-stage pupal wings, we can substitute the 164-bp homologous sequence in *peak-2*^{bia} with the 164-bp element from *D. melanogaster*. The enhancer accessibility can be profiled by ATAC-seq, and the enhancer activity can be measured by imaging the

reporter gene. It is possible that the164-bp element can also maintain the accessibility of *peak-2-164-bp*^{bia} in late-pupal wings. If that is the case, then it is likely that the 164-bp element, while not essential for initiating yellow expression in mid-pupal wings, can facilitate peak-2 accessible in late-pupal wings. Therefore, the sequence variation in this 164-bp element but not those in *peak-2-164-bp*^{mel} would be thought to contribute to the observed loss of the *spot*.

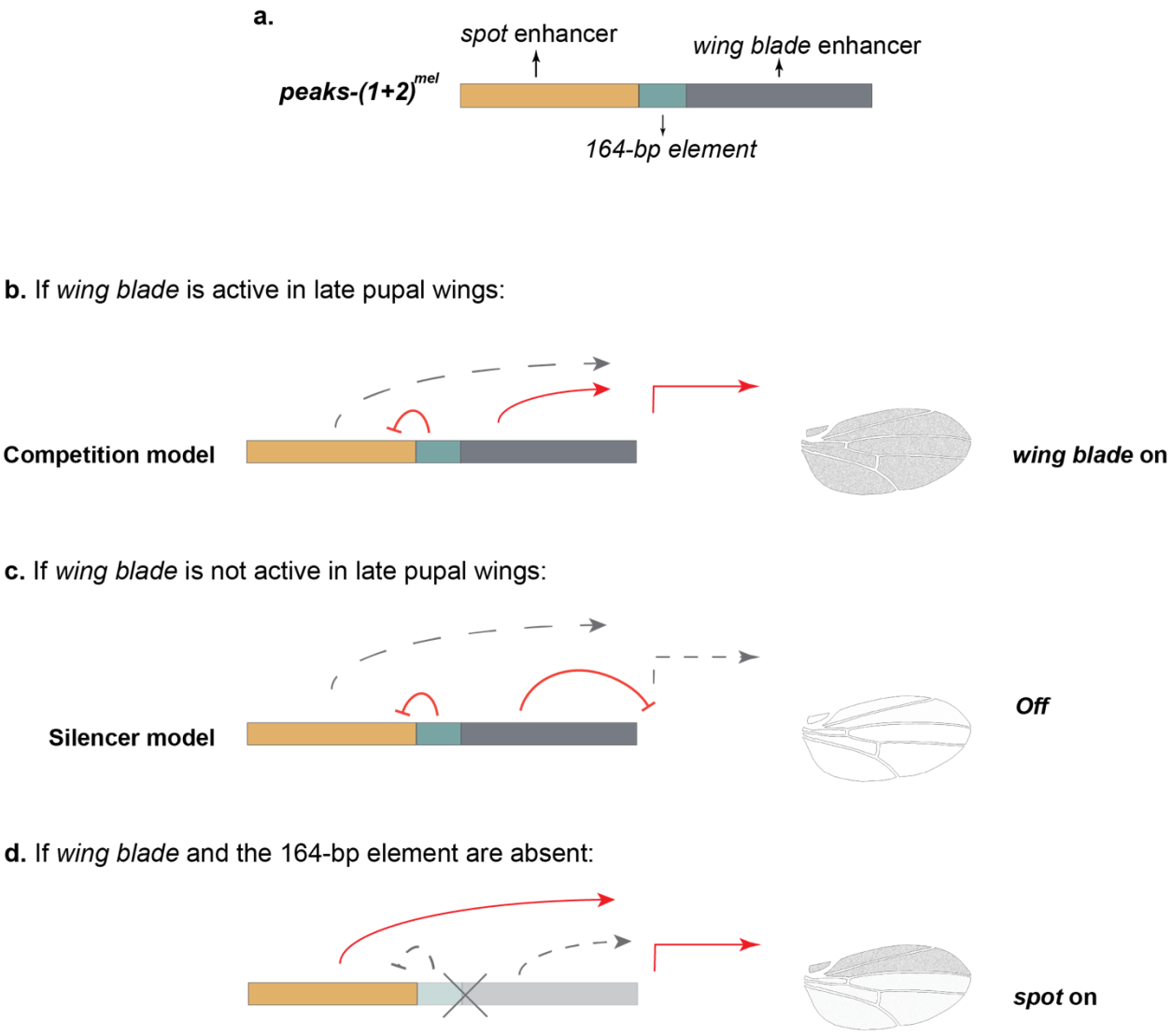

Figure 8. Two enhancer functional models may explain the findings of this dissertation. *(a.) The functional elements of the yellow^{mel} enhancer include three components: the spot enhancer, the 164-bp repressing element, and the wing blade enhancer. The illustration is the same as in Figure 7. (b. and c.) Both the competition and silencer models can explain the silencing of the spot^{mel} function. (d.) The spot^{mel} only functions when peak-2^{mel} is absent.*

4.2. Competition model and silencer model

The competition model shows that enhancers may not always work together synergistically, as evidenced by several studies[193,202]. In some cases, the co-existence of two enhancers could reduce gene transcription, particularly when redundancy or shadow enhancer function is involved. However, studying this model often requires live-cell imaging, and setting up such experiments in pupal wings is challenging.

One potential experiment could be to swap the positions of the *peak-1*^{mel} and *peak-2*^{mel}, as they have similar accessibility in late-stage pupal wings. In the context of two proximal enhancers, the enhancer closer to the promoter might have a greater chance of interacting with the promoter. As previously stated, the competition model requires the *wing blade* activity of *peak-2*^{mel} during late-stage pupal wings (Figure 8b). Based on the model, the *peak-1*^{mel} with *spot* activity might be a partially redundant enhancer regulating the same gene in a distinct pattern, partially overlapping with the primary enhancer (*peak-2*^{mel}) within the same tissue.

As the accessibility of *peak-2* decreases in late pupal wings, as is the case with *peak-2*^{bia}, the redundant *spot* enhancer may become more active than the wing blade enhancer, resulting in a spot-like expression pattern. Furthermore, the competition model proposes that redundant enhancers, like shadow enhancers, may drive phenotypic evolution. The conservation of the *wing blade* enhancer in both spotted and non-spotted species underscores the importance of acquiring a second enhancer that regulates *yellow* expression, primarily impacting morphological trait evolution rather than initiating expression in pupal wings.

According to the competition model, the evolution of the *peak-2*^{mel} or *wing blade* enhancer in *D. melanogaster* was facilitated by acquiring transcription factor binding sites. These binding sites recruited different transcription factors compared to those in *D. biarmipes*. As a result, the *wing blade* enhancer in *D. melanogaster* was still accessible and active during the late pupal stages. Additionally, it also recruited transcription factors responsible for repressing the *peak-1*^{mel}, or the *spot* enhancer. This intricate evolutionary trajectory prompts further questions about the frequency of such complex enhancer evolution pathways.

The findings presented in this dissertation also support the role of *peak-2*^{mel} as an evolutionary silencer in late-stage pupal wings (Figure 8c). There are two facets to be discussed in this context: the evolutionary development of the silencer and its functional mechanism. To elucidate the evolutionary development of the silencer, one question needs to be further studied: whether the 164-bp element maintains the accessibility of the *peak-2-164-bp*^{mel} accessible in late-stage pupal wings. If the evidence supports the 164-bp element as responsible for this, it proves its essential role in the evolutionary acquisition of the silencer.

Alternatively, other variants in *peak-2-164-bp*^{mel} should be considered if they maintain accessibility in late-stage pupal wings. Such an evolutionary path mirrors the complexity questioned in the evolution of the competitive enhancer model discussed above.

5. The silencing mechanism: enhancer competition or enhancer silencing?

Nearly four decades have passed since the publication of the first study on silencers, and recent research has revisited this topic with a genome-wide perspective. These studies underscore the significance of silencers in developmental processes, cell differentiation, and disease pathogenesis[209,211,277]. Section 6 of the Introduction has summarized the general mechanism regulating silencer function. Briefly, silencers recruit repressors to directly inhibit the activity of enhancers or regulate enhancers to be compacted back into the chromatin structure by recruiting corepressors and chromatin modification proteins.

Typically, silencers exhibit an enrichment of repressor binding motifs. However, despite attempts at motif analysis and RNAi to deplete potential repressors, no repressor binding to the evolved repressor binding site in the 164-bp element was identified in this dissertation. A comprehensive approach to studying the silencing mechanism might involve meticulously analyzing transcriptomics and accessibility, employing scRNA-seq and scATAC-seq techniques. These methods can help identify the specific repressors expressed at the appropriate temporal and spatial points during silencing, thereby refining the candidate list of repressors for further investigation.

Another way to study the silencing mechanism is to investigate the cofactors and chromatin regulators involved in silencing. While previous studies on screening silencers have yielded varying conclusions about shared chromatin features, the study by Huang *et al.*, summarized in Section 6 of the Introduction, successfully predicted silencers through cell lineage data and deep learning[214]. This study pointed out that chromatin accessibility, repressor motifs, and the H3K27me3 marks appear to be commonly shared by silencers. Also, the binding of repressors initiates the recruitment of their corepressors, which, in turn, further recruits chromatin modifiers such as HDAC family proteins[278]. A recent study by Jacobs *et al.* demonstrated the specificity of these corepressors in regulating gene transcription[279]. Therefore, integrating the current ATAC-seq dataset with H3K27me3 and ChIP-seq against corepressors such as CtBP could offer insights into the mechanisms underlying the silencing of *peak-2*^{mel}.

However, the mechanisms governing gene transcriptional silencing may be more intricate than the general model suggests. Based on the discussion above, another possible approach to silencing an enhancer is by competing with another enhancer for the same promoter. In this scenario, although the gene expression pattern may change, it may not be completely silenced if the two enhancers recruit different subsets of

activators but not repressors. Section 5.2.2 in the Introduction has discussed the enhancer competition model, which is particularly observable in developmental systems offering a spatial and temporal resolution of transcription factor activity. This model has thus been studied in *Drosophila* embryogenesis, where well-studied enhancers drive gene expression in diverse patterns throughout embryogenesis[193,202].

The enhancer silencing process uncovered in this dissertation presents an alternative system for studying silencing models. While the embryogenesis process is conserved across *Drosophila* species, the expression patterns of the *yellow* gene in wings differ. In addition to investigating silencing mechanisms, this system provides an opportunity to answer the question: How does sequence variation switch enhancers into silencers during evolution?

6. Unveiling evolutionary silencers through profiling multi-omics data

Studies focusing on enhancer variation and phenotypical evolution traditionally relied on enhancer-reporter assays. These assays are known for their robustness in identifying enhancer functions and have been widely used in exploring the effect of different enhancer lengths and mutations. However, recent research has implied the limitations of using minimum enhancer constructs in these assays. Several studies, including one in this dissertation (Paper III of the Results), have indicated that the information is distributed densely across enhancers when only a minimum enhancer was used in the assay. Specifically, mutations at almost all positions across the minimal enhancers affect its activity. Such findings do not align with the evolutionary conserved enhancers controlling developmental gene expression. For instance, a study that used several different inbred lines from *D. melanogaster* showed that most enhancers share conserved features and functions during embryogenesis despite their non-identical sequences[280]. Additionally, even with sequence variation across different species, the studies of the *yellow* regulatory region evolution showed that its *wing blade* enhancer is conserved among species[14,240]. To comprehensively understand the regulatory mechanisms underlying enhancers in their native context, it is essential to consider the broader genomic contexts of the minimal enhancer.

Throughout this dissertation, I established ATAC-seq, ChIP-seq, and CUT&RUN to investigate gene transcriptional regulation in prepupal and pupal wings. I also profiled chromatin accessibility changes during the pupal wing development of two closely related species within the *D. melanogaster* group. Besides ChIP-seq against H3K27ac, discussed as part of the results in this dissertation (see Paper I of the Results), I also modified protocols for ChIP-seq and CUT&RUN against Dll in 72h APF pupal wings from *D. melanogaster* during this study. These protocols can potentially be adapted for earlier-stage pupal wings and for studying other transcription factors or cofactors if their corresponding antibodies are available. The wing pigmentation pattern is a relatively straightforward morphological trait that involves a two-

dimensional epithelial tissue with a dark pattern. This contrasts with other more complex morphological models that explore enhancer evolution, such as butterfly wings and human facial morphology. By establishing these multi-omics methods in the wings from pupal developmental stages covering from 0h APF to 78h APF, this dissertation paves the way for future research on enhancer evolution using the wing pigmentation pattern as the model system.

Through analyzing chromatin dynamic changes, I found an alternative evolutionary pathway leading to trait loss. This pathway differs from what previous research has shown. Earlier studies employing multi-omics data analysis have indicated that changes in sequence variation within *cis*-regulatory elements can affect chromatin dynamics in enhancers, leading to divergent gene expression[168,170,281,282]. Thus, enhancer evolution is crucial in differentiating tissue development among species and disease-related functional changes. However, the relationship between the sequence variation and the silencer evolution remains elusive. Since 2019, around ten studies have been published focusing on identifying silencers (see Section 6.2 of the Introduction). Some of these studies highlight *cis*-regulatory regions that can have open chromatin in different cell types during development and cellular differentiation, with divergent functions. A study by Gisselbrecht *et al.* revealed that most silencers identified in this study are enhancers in other parts of embryos at different stages[215].

Furthermore, likely to be a silencer in late-stage pupae, *peak-2*^{mel} is an example of how the acquisition of stage-specific silencers can contribute to the evolution of phenotypical traits. As previously mentioned, the study from Huang *et al.* indicates that the gain of silencers, often located proximal to enhancers, has distinct sets of repressor binding motifs from the activators binding sites in nearby enhancers[214]. This might be related to the loss of phenotypical traits due to the gain of repressor binding sites. For example, the gain of a repressor binding site next to the core enhancer led to the loss of trichome in *D. sechellia*[236], as discussed in Section 6.3 in the Introduction. Given this proximity, instead of mutating enhancers using enhancer-reporter assay, a new approach to studying the evolutionary loss of traits could involve systematically screening evolutionary silencers. It can be done by comprehensively profiling chromatin and gene expression dynamics during pupal development of closely related species.

Moreover, recent studies have shown that enhancer activation and gene transcription involve the recruitment of specific cofactors[199,279]. It would be intriguing to explore whether evolutionary silencers lead to distinct recruitment of co-factors, resulting in particular changes in chromatin landscapes. The *D. melanogaster* group exhibits considerable divergence in wing pigmentation patterns among species[239]. Examining the dynamic changes in chromatin accessibility and cofactor binding during pupal wing

development could provide new insights into how sequence variation contributes to the evolutionary conserved and diverse gene transcriptional regulation.

This dissertation has discovered a novel evolutionary pathway involved in silencing a gained enhancer during development. Nonetheless, as discussed in this section, this discovery raises further questions about the regulatory mechanisms governing enhancers and silencers and their intricate interplay in the context of developmental evolution. More comprehensive investigations are needed to broaden the understanding of the evolution of enhancers and silencers.

References

1. WATSON, J. D. & CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

2. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).

3. Lewis, M. A Tale of Two Repressors. *J. Mol. Biol.* **409**, 14–27 (2011).

4. Gann, A. Jacob and Monod: From Operons to EvoDevo. *Curr. Biol.* **20**, R718–R723 (2010).

5. Consortium, I. H. G. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

6. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

7. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).

8. Consortium, T. E. P. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).

9. King, M.-C. & Wilson, A. C. Evolution at Two Levels in Humans and Chimpanzees. *Science* **188**, 107–116 (1975).

10. Prager, E. M. & Wilson, A. C. Slow evolutionary loss of the potential for interspecific hybridization in birds: a manifestation of slow regulatory evolution. *Proc. Natl. Acad. Sci.* **72**, 200–204 (1975).

11. Wilson, A. C., Maxson, L. R. & Sarich, V. M. Two Types of Molecular Evolution. Evidence from Studies of Interspecific Hybridization. *Proc. Natl. Acad. Sci.* **71**, 2843–2847 (1974).

12. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).

13. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).

14. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**, nature03235 (2005).

15. Consortium, T. modENCODE *et al.* Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **330**, 1787–1797 (2010).

16. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & Laat, W. de. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol Cell* **10**, 1453–1465 (2002).

17. Schoenfelder, S., Clay, I. & Fraser, P. The transcriptional interactome: gene expression in 3D. *Curr. Opin. Genet. Dev.* **20**, 127–133 (2010).

18. Amano, T. *et al.* Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Dev. Cell* **16**, 47–57 (2009).

19. Clapier, C. R. & Cairns, B. R. The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem* **78**, 273–304 (2009).

20. Mitchell, P. J. & Tjian, R. Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins. *Science* **245**, 371–378 (1989).

21. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, nrg3207 (2012).

22. Levine, M., Cattoglio, C. & Tjian, R. Looping Back to Leap Forward: Transcription Enters a New Era. *Cell* **157**, 13–25 (2014).

23. Zabidi, M. A. & Stark, A. Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet.* **32**, 801–814 (2016).

24. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).

25. Soutourina, J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Bio* **19**, 262–274 (2018).

26. Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477 (1999).

27. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).

28. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).

29. Wu, C. The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854–860 (1980).

30. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. *Cell* **27**, 45–55 (1981).

31. Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* **16**, 797–806 (1979).

32. Wu, C., Wong, Y.-C. & Elgin, S. C. R. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16**, 807–814 (1979).

33. Stalder, J. *et al.* Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**, 451–460 (1980).

34. Weintraub, H., Larsen, A. & Groudine, M. α-globin-gene switching during the development of chicken embryos: Expression and chromosome structure. *Cell* **24**, 333–344 (1981).

35. Elgin, S. C. R. DNAase I-hypersensitive sites of chromatin. *Cell* **27**, 413–415 (1981).

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

37. Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729–740 (1983).

38. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**, 717–728 (1983).

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

40. Parslow, T. G. & Granner, D. K. Chromatin changes accompany immunoglobulin κ gene activation: a potential control region within the gene. *Nature* **299**, 449–451 (1982).

41. Levine, M. Transcriptional Enhancers in Animal Development and Evolution. *Curr Biol* **20**, R754–R763 (2010).

42. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).

43. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**, 206–216 (2007).

44. Bobola, N. & Merabet, S. Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity. *Curr Opin Genet Dev* **43**, 1–8 (2017).

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

46. Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor–DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* **43**, 110–119 (2017).

47. Grosveld, F., Staalduinen, J. van & Stadhouders, R. Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. *Annu. Rev. Genom. Hum. Genet.* **22**, 1–20 (2021).

48. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary Novelty. *Science* **321**, 1314–1314 (2008).

49. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868–871 (1974).

50. NOLL, M. Subunit structure of chromatin. *Nature* **251**, 249–251 (1974).

51. Noll, M. Internal structure of the chromatin subunit. *Nucleic Acids Res.* **1**, 1573–1578 (1974).

52. Weintraub, H. & Groudine, M. Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science* **193**, 848–856 (1976).

53. Shermoen, A. W. & Beckendorf, S. K. A complex of interacting DNAase I-hypersensitive sites near the drosophila glue protein gene, Sgs4. *Cell* **29**, 601–607 (1982).

54. Wu, C. & Gilbert, W. Tissue-specific exposure of chromatin structure at the 5’ terminus of the rat preproinsulin II gene. *Proc National Acad Sci* **78**, 1577–1580 (1981).

55. Boss, M. A. Gene regulation: Enhancer elements in immunoglobulin genes. *Nature* **303**, 281–282 (1983).

56. Bryan, P. N. & Folk, W. R. Enhancer sequences responsible for DNase I hypersensitivity in polyomavirus chromatin. *Mol Cell Biol* **6**, 2249–2252 (1986).

57. Akmal, M., el-Ghor, A. & Burk, R. D. DNase I hypersensitive site maps to the HBV enhancer. *Virology* **172**, 478–488 (1989).

58. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207–220 (2019).

59. He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nat Genet* **42**, 343–347 (2010).

60. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* **36**, 900–905 (2004).

61. Ren, B. *et al.* Genome-Wide Location and Function of DNA Binding Proteins. *Science* **290**, 2306–2309 (2000).

62. Shalon, D., Smith, S. J. & Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639–645 (1996).

63. Xi, H. *et al.* Identification and Characterization of Cell Type–Specific and Ubiquitous Chromatin Regulatory Structures in the Human Genome. *Plos Genet* **3**, e136 (2007).

64. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**, 311–322 (2008).

65. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-Throughput Sequencing Technologies. *Mol Cell* **58**, 586–597 (2015).

66. Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. *Genome Res.* **16**, 1505–1516 (2006).

67. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**, 877–885 (2007).

68. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. 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).

69. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Bio* **4**, 809–814 (2003).

70. McGhee, J. D. & Felsenfeld, G. Nucleosome Structure. *Annu. Rev. Biochem.* **49**, 1115–1156 (1980).

71. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).

72. Becker, P. B. & Workman, J. L. Nucleosome Remodeling and Epigenetics. *Csh Perspect Biol* **5**, a017905 (2013).

73. Morgan, M. A. J. & Shilatifard, A. Epigenetic moonlighting: Catalytic-independent functions of histone modifiers in regulating transcription. *Sci Adv* **9**, eadg6593 (2023).

74. Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet* **27**, 389–396 (2011).

75. Cenik, B. K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and disease. *Nat. Rev. Genet.* **22**, 38–58 (2021).

76. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).

77. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).

78. Sims, R. J. *et al.* Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains*. *J. Biol. Chem.* **280**, 41789–41792 (2005).

79. Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).

80. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).

81. Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131–3141 (2009).

82. Mathur, R. *et al.* ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* **49**, 296–302 (2017).

83. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648 (2017).

84. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution† † We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *J. Mol. Biol.* **319**, 1097–1113 (2002).

85. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

86. Sehnal, D. *et al.* Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).

87. Whitehouse, I. *et al.* Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature* **400**, 784–787 (1999).

88. Becker, P. B. & Hörz, W. ATP-DEPENDENT NUCLEOSOME REMODELING. *Annu. Rev. Biochem.* **71**, 247–273 (2002).

89. Clapier, C. R. *et al.* Regulation of DNA Translocation Efficiency within the Chromatin Remodeler RSC/Sth1 Potentiates Nucleosome Sliding and Ejection. *Mol. Cell* **62**, 453–461 (2016).

90. Schwabish, M. A. & Struhl, K. The Swi/Snf Complex Is Important for Histone Eviction during Transcriptional Activation and RNA Polymerase II Elongation In Vivo. *Mol. Cell. Biol.* **27**, 6987–6995 (2007).

91. Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720 (1999).

92. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

93. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–318 (2007).

94. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**, 599–606 (1998).

95. Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly Specific Antibodies Determine Histone Acetylation Site Usage in Yeast Heterochromatin and Euchromatin. *Mol. Cell* **8**, 473–479 (2001).

96. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897–903 (2008).

97. Felsenfeld, G. Chromatin Unfolds. *Cell* **86**, 13–19 (1996).

98. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc National Acad Sci* **107**, 21931–21936 (2010).

99. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).

100. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).

101. Swinstead, E. E., Paakinaho, V., Presman, D. M. & Hager, G. L. Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective. *Bioessays* **38**, 1150–1157 (2016).

102. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).

103. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Gene Dev* **25**, 2227–2241 (2011).

104. Zaret, K. S. Pioneer Transcription Factors Initiating Gene Network Changes. *Annu Rev Genet* **54**, 1–19 (2020).

105. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6**, e22631 (2017).

106. Frederick, M. A. *et al.* A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **30**, 31–37 (2023).

107. Cirillo, L. A. *et al.* Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279–289 (2002).

108. Larson, E. D., Marsh, A. J. & Harrison, M. M. Pioneering the developmental frontier. *Mol Cell* (2021) doi:10.1016/j.molcel.2021.02.020.

109. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400–403 (2008).

110. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. *Plos Genet* **7**, e1002266 (2011).

111. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).

112. Pálfy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet.* **16**, e1008546 (2020).

113. Gassler, J. *et al.* Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).

114. Luzete-Monteiro, E. & Zaret, K. S. Structures and consequences of pioneer factor binding to nucleosomes. *Curr. Opin. Struct. Biol.* **75**, 102425 (2022).

115. Tan, C. & Takada, S. Nucleosome allostery in pioneer transcription factor binding. *Proc National Acad Sci* **117**, 20586–20596 (2020).

116. Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat Rev Mol Cell Bio* **23**, 449–464 (2022).

117. Filion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224 (2010).

118. Brennan, K. J. *et al.* Chromatin accessibility in the Drosophila embryo is determined by transcription factor pioneering and enhancer activation. *Dev. Cell* (2023) doi:10.1016/j.devcel.2023.07.007.

119. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).

120. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* **151**, 994–1004 (2012).

121. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).

122. Hammonds, A. S. *et al.* Spatial expression of transcription factors in Drosophilaembryonic organ development. *Genome Biol.* **14**, R140 (2013).

123. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).

124. Foo, S. M. *et al.* Zelda Potentiates Morphogen Activity by Increasing Chromatin Accessibility. *Curr. Biol.* **24**, 1341–1346 (2014).

125. Gibson, T. J. & Harrison, M. M. Protein-intrinsic properties and epigenetic effects regulate pioneer-factor binding and function. *Biorxiv* 2023.03.18.533281 (2023) doi:10.1101/2023.03.18.533281.

126. Lawrence, P. A. *The Making of a Fly: The Genetics of Animal Design*. (Blackwell Scientific Publications, Oxford, 1992).

127. Siebenlist, U., Franzoso, G. & Brown, K. Structure, Regulation and Function of NF-kappaB. *Annu. Rev. Cell Dev. Biol.* **10**, 405–455 (1994).

128. El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nat. Genet.* **1**, 45–49 (1992).

129. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annu Rev Cell Dev Bi* **35**, 1–23 (2019).

130. Slattery, M. *et al.* Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. *Cell* **147**, 1270–1282 (2011).

131. Crocker, J. *et al.* Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**, 191–203 (2015).

132. Rastogi, C. *et al.* Accurate and sensitive quantification of protein-DNA binding affinity supplementary. *Proc National Acad Sci* **115**, 201714376 (2018).

133. Farley, E. K. *et al.* Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).

134. Jindal, G. A. *et al.* Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. *Dev. Cell* **58**, 2206-2216.e5 (2023).

135. Lim, F. *et al.* Affinity-optimizing enhancer variants disrupt development. *Nature* **626**, 151–159 (2024).

136. Merika, M. & Thanos, D. Enhanceosomes. *Curr Opin Genet Dev* **11**, 205–208 (2001).

137. Kulkarni, M. M. & Arnosti, D. N. Information display by transcriptional enhancers. *Development* **130**, 6569–6575 (2003).

138. Zeitlinger, J. Seven myths of how transcription factors read the cis-regulatory code. *Curr Opin Syst Biology* (2020) doi:10.1016/j.coisb.2020.08.002.

139. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).

140. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).

141. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).

142. Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease. *Cell* **177**, 132–145 (2019).

143. Chan, Y. F. *et al.* Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a *Pitx1* Enhancer. *Science* **327**, 302–305 (2010).

144. Indjeian, V. B. *et al.* Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell* **164**, 45–56 (2016).

145. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642.e11 (2016).

146. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).

147. Frankel, N. *et al.* Morphological evolution caused by many subtle-effect substitutions in a transcriptional enhancer. *Nature* **474**, 598–603 (2011).

148. Ramaekers, A. *et al.* Altering the Temporal Regulation of One Transcription Factor Drives Evolutionary Trade-Offs between Head Sensory Organs. *Dev. Cell* **50**, 780-792.e7 (2019).

149. Jeong, S., Rokas, A. & Carroll, S. B. Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in Drosophila Evolution. *Cell* **125**, 1387–1399 (2006).

150. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

151. Jaitin, D. A. *et al.* Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science* **343**, 776–779 (2014).

152. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

153. Calderon, D. *et al.* The continuum of Drosophila embryonic development at single-cell resolution. *Science* **377**, eabn5800–eabn5800 (2022).

154. Qiu, C. *et al.* Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* **54**, 328–341 (2022).

155. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, (2020).

156. Reddington, J. P. *et al.* Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. *Dev Cell* **55**, 648-664.e9 (2020).

157. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

158. Stergachis, A. B. *et al.* Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes. *Cell* **154**, 888–903 (2013).

159. McKay, D. J. & Lieb, J. D. A Common Set of DNA Regulatory Elements Shapes Drosophila Appendages. *Dev Cell* **27**, 306–318 (2013).

160. Delás, M. J. *et al.* Developmental cell fate choice in neural tube progenitors employs two distinct cis-regulatory strategies. *Dev Cell* **58**, 3-17.e8 (2023).

161. Bozek, M. *et al.* ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. *Genome Res* **29**, 771–783 (2019).

162. Waddington, C. H. *The Strategy of the Genes*. (Routledge, 1957). doi:10.4324/9781315765471.

163. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated Evolution of Conserved Noncoding Sequences in Humans. *Science* **314**, 786–786 (2006).

164. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).

165. Doan, R. N. *et al.* Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* **167**, 341-354.e12 (2016).

166. Girskis, K. M. *et al.* Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239-3251.e7 (2021).

167. Whalen, S. *et al.* Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* **111**, 857-873.e8 (2023).

168. Torre-Ubieta, L. de la *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304.e18 (2018).

169. Belton, J.-M. *et al.* Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).

170. Liang, D. *et al.* Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat Neurosci* 1–13 (2021) doi:10.1038/s41593-021-00858-w.

171. Grosveld, F., Assendelft, G. B. van, Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).

172. Hoch, M., Schröder, C., Seifert, E. & Jäckle, H. cis-acting control elements for Krüppel expression in the Drosophila embryo. *EMBO J.* **9**, 2587–2595 (1990).

173. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077–3086 (2002).

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

175. Goto, T., Macdonald, P. & Maniatis, T. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413–422 (1989).

176. Vincent, B. J., Estrada, J. & DePace, A. H. The appeasement of Doug: a synthetic approach to enhancer biology. *Integr Biol* **8**, 475–484 (2016).

177. Schroeder, M. D., Greer, C. & Gaul, U. How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. *Development* **138**, 3067–3078 (2011).

178. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**, 307–319 (2013).

179. Kalay, G., Lachowiec, J., Rosas, U., Dome, M. R. & Wittkopp, P. J. Redundant and Cryptic Enhancer Activities of the Drosophila yellow Gene. *Genetics* **212**, genetics.301985.2019 (2019).

180. Ostuni, R. *et al.* Latent Enhancers Activated by Stimulation in Differentiated Cells. *Cell* **152**, 157–171 (2013).

181. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci.* **110**, 17921–17926 (2013).

182. Lovén, J. *et al.* Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* **153**, 320–334 (2013).

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

184. Moorthy, S. D. *et al.* Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27**, 246–258 (2017).

185. Boettiger, A. N. & Levine, M. Synchronous and Stochastic Patterns of Gene Activation in the Drosophila Embryo. *Science* **325**, 471–473 (2009).

186. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow Enhancers Foster Robustness of Drosophila Gastrulation. *Curr Biol* **20**, 1562–1567 (2010).

187. Kvon, E. Z., Waymack, R., Elabd, M. G. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat Rev Genet* 1–13 (2021) doi:10.1038/s41576-020-00311-x.

188. Zeitlinger, J. *et al.* Whole-genome ChIP–chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Gene Dev* **21**, 385–390 (2007).

189. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* **26**, 38–51 (2016).

190. Fulco, C. P. *et al.* Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).

191. Dukler, N., Gulko, B., Huang, Y.-F. & Siepel, A. Is a super-enhancer greater than the sum of its parts? *Nat. Genet.* **49**, 2–3 (2017).

192. Carey, M. The Enhanceosome and Transcriptional Synergy. *Cell* **92**, 5–8 (1998).

193. Bothma, J. P. *et al.* Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* **4**, e07956 (2015).

194. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358–368 (2016).

195. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).

196. Lin, X. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* eabk3512 (2022) doi:10.1126/science.abk3512.

197. Choi, J. *et al.* Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. *Elife* **10**, e65381 (2021).

198. Donati, B., Lorenzini, E. & Ciarrocchi, A. BRD4 and Cancer: going beyond transcriptional regulation. *Mol. Cancer* **17**, 164 (2018).

199. Neumayr, C. *et al.* Differential cofactor dependencies define distinct types of human enhancers. *Nature* **606**, 406–413 (2022).

200. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).

201. Kvon, E. Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91 (2014).

202. Scholes, C., Biette, K. M., Harden, T. T. & DePace, A. H. Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. *Cell Reports* **26**, 2407-2418.e5 (2019).

203. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife* **6**, e28975 (2017).

204. Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. Characterization of a "silencer" in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**, 41–48 (1985).

205. OGBOURNE, S. & ANTALIS, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* **331**, 1–14 (1998).

206. Schuettengruber, B., Bourbon, H.-M., Croce, L. D. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).

207. Dietrich, N. *et al.* REST–Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. *PLoS Genet.* **8**, e1002494 (2012).

208. Watson, J. D. *et al. Molecular Biology of the Gene*. (Pearson, 2013).

209. Pang, B., Weerd, J. H. van, Hamoen, F. L. & Snyder, M. P. Identification of non-coding silencer elements and their regulation of gene expression. *Nat. Rev. Mol. Cell Biol.* **24**, 383–395 (2023).

210. Cai, Y. *et al.* H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat. Commun.* **12**, 719 (2021).

211. Pang, B. & Snyder, M. P. Systematic identification of silencers in human cells. *Nat Genet* **52**, 254–263 (2020).

212. Huang, D., Petrykowska, H. M., Miller, B. F., Elnitski, L. & Ovcharenko, I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res* **29**, 657–667 (2019).

213. Jayavelu, N. D., Jajodia, A., Mishra, A. & Hawkins, R. D. Candidate silencer elements for the human and mouse genomes. *Nat. Commun.* **11**, 1061 (2020).

214. Huang, D. & Ovcharenko, I. Enhancer-silencer transitions in the human genome. *Genome Res.* **32**, gr.275992.121 (2022).

215. Gisselbrecht, S. S. *et al.* Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Mol Cell* **77**, 324-337.e8 (2020).

216. Ngan, C. Y. *et al.* Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* **52**, 264–272 (2020).

217. Hussain, S. *et al.* Short tandem repeats are important contributors to silencer elements in T cells. *Nucleic Acids Res.* **51**, 4845–4866 (2023).

218. Maue, R. A., Kraner, S. D., Goodman, R. H. & Mandelt, G. Neuron-specific expression of the rat brain type II sodium channel gene is directed by upstream regulatory elements. *Neuron* **4**, 223–231 (1990).

219. Mori, N., Stein, R., Sigmund, O. & Anderson, D. J. A cell type-preferred silencer element that controls the neural-specific expression of the SCG10 gene. *Neuron* **4**, 583–594 (1990).

220. Mori, N., Schoenherr, C., Vandenbergh, D. J. & Anderson, D. J. A common silencer element in the SCG10 and type II Na+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells. *Neuron* **9**, 45–54 (1992).

221. Kraner, S. D., Chong, J. A., Tsay, H.-J. & Mandel, G. Silencing the type II sodium channel gene: A model for neural-specific gene regulation. *Neuron* **9**, 37–44 (1992).

222. Schoenherr, C. J. & Anderson, D. J. The Neuron-Restrictive Silencer Factor (NRSF): A Coordinate Repressor of Multiple Neuron-Specific Genes. *Science* **267**, 1360–1363 (1995).

223. Chong, J. A. *et al.* REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).

224. Sawada, S., Scarborough, J. D., Killeen, N. & Littman, D. R. A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* **77**, 917–929 (1994).

225. Taniuchi, I. *et al.* Differential Requirements for Runx Proteins in CD4 Repression and Epigenetic Silencing during T Lymphocyte Development. *Cell* **111**, 621–633 (2002).

226. Busturia, A. & Bienz, M. Silencers in abdominal-B, a homeotic Drosophila gene. *EMBO J.* **12**, 1415–1425 (1993).

227. Erceg, J. *et al.* Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. *Genes Dev.* **31**, 590–602 (2017).

228. Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570 (1978).

229. Müller, J. *et al.* Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* **111**, 197–208 (2002).

230. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49**, 808–824 (2013).

231. Klymenko, T. *et al.* A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* **20**, 1110–1122 (2006).

232. Hansen, T. J. & Hodges, E. ATAC-STARR-seq reveals transcription factor-bound activators and silencers across the chromatin accessible human genome. *Genome Res.* **32**, gr.276766.122 (2022).

233. Zhang, Y. *et al.* Super-silencers regulated by chromatin interactions control apoptotic genes. *bioRxiv* 2022.01.17.476559 (2022) doi:10.1101/2022.01.17.476559.

234. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387–398 (2022).

235. McGregor, A. P. *et al.* Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).

236. Preger-Ben Noon, E., Davis, F. P. & Stern, D. L. Evolved Repression Overcomes Enhancer Robustness. *Dev Cell* **39**, 572–584 (2016).

237. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

238. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).

239. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences* **104**, 8605–8612 (2007).

240. Prud'homme, B. *et al.* Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, nature04597 (2006).

241. True, J. R. Insect melanism: the molecules matter. *Trends Ecol. Evol.* **18**, 640–647 (2003).

242. Drapeau, M. D. A novel hypothesis on the biochemical role of the Drosophila Yellow protein. *Biochem. Biophys. Res. Commun.* **311**, 1–3 (2003).

243. Arnoult, L. *et al.* Emergence and Diversification of Fly Pigmentation Through Evolution of a Gene Regulatory Module. *Science* **339**, 1423–1426 (2013).

244. Xin, Y. *et al.* Enhancer evolutionary co-option through shared chromatin accessibility input. *Proc National Acad Sci* **117**, 20636–20644 (2020).

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

246. Campbell, G. & Tomlinson, A. The roles of the homeobox genes aristaless and Distal-less in patterning the legs and wings of Drosophila. *Dev Camb Engl* **125**, 4483–93 (1998).

247. Gorfinkiel, N., Morata, G. & Guerrero, I. The homeobox gene Distal-less induces ventral appendage development in Drosophila. *Genes Dev.* **11**, 2259–2271 (1997).

248. Blair, S. S. engrailed expression in the anterior lineage compartment of the developing wing blade of Drosophila. *Development* **115**, 21–33 (1992).

249. Hinaux, H. *et al.* Revisiting the developmental and cellular role of the pigmentation gene yellow in Drosophila using a tagged allele. *Dev Biol* **438**, 111–123 (2018).

250. Yalonetskaya, A., Mondragon, A. A., Elguero, J. & McCall, K. I Spy in the Developing Fly a Multitude of Ways to Die. *J. Dev. Biol.* **6**, 26 (2018).

251. Loza, M. C. D. de la & Thompson, B. J. Forces shaping the Drosophila wing. *Mech Develop* **144**, 23–32 (2017).

252. Thummel, C. S. Flies on steroids — Drosophila metamorphosis and the mechanisms of steroid hormone action. *Trends Genet.* **12**, 306–310 (1996).

253. Sobala, L. F. & Adler, P. N. The Gene Expression Program for the Formation of Wing Cuticle in Drosophila. *Plos Genet* **12**, e1006100 (2016).

254. Dubrovsky, E. B. Hormonal cross talk in insect development. *Trends Endocrinol. Metab.* **16**, 6–11 (2005).

255. RIDDIFORD, L. M. & TRUMAN, J. W. Hormone Receptors and the Regulation of Insect Metamorphosis1. *Am. Zoöl.* **33**, 340–347 (1993).

256. Baehrecke, E. H. & Thummel, C. S. The Drosophila E93 Gene from the 93F Early Puff Displays Stage- and Tissue-Specific Regulation by 20-Hydroxyecdysone. *Dev. Biol.* **171**, 85–97 (1995).

257. Lee, C.-Y. *et al.* E93 Directs Steroid-Triggered Programmed Cell Death in Drosophila. *Mol Cell* **6**, 433–443 (2000).

258. Lam, G., Nam, H.-J., Velentzas, P. D., Baehrecke, E. H. & Thummel, C. S. Drosophila E93 promotes adult development and suppresses larval responses to ecdysone during metamorphosis. *Dev. Biol.* **481**, 104–115 (2022).

259. Mou, X., Duncan, D. M., Baehrecke, E. H. & Duncan, I. Control of target gene specificity during metamorphosis by the steroid response gene E93. *Proc National Acad Sci* **109**, 2949–2954 (2012).

260. Uyehara, C. M. *et al.* Hormone-dependent control of developmental timing through regulation of chromatin accessibility. *Gene Dev* **31**, 862–875 (2017).

261. Nystrom, S. L., Niederhuber, M. J. & McKay, D. J. Expression of E93 provides an instructive cue to control dynamic enhancer activity and chromatin accessibility during development. *Development* **147**, dev181909 (2020).

262. Minnoye, L. *et al.* Chromatin accessibility profiling methods. *Nat Rev Methods Primers* **1**, 10 (2021).

263. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006 (2018).

264. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).

265. Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* **21**, 432–439 (2018).

266. Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* (2018) doi:10.1038/nature25981.

267. Prescott, S. L. *et al.* Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell* **163**, 68–83 (2015).

268. Ling, L., Mühling, B., Jaenichen, R. & Gompel, N. Increased chromatin accessibility promotes the evolution of a transcriptional silencer in Drosophila. *Sci Adv* **9**, (2023).

269. Nagarkar-Jaiswal, S. *et al.* A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *Elife* **4**, e05338 (2015).

270. Nizamuddin, S. *et al.* Integrating quantitative proteomics with accurate genome profiling of transcription factors by greenCUT&RUN. *Nucleic Acids Res* **49**, e49–e49 (2021).

271. Koidl, S. & Timmers, H. T. M. greenCUT&RUN: Efficient Genomic Profiling of GFP-Tagged Transcription Factors and Chromatin Regulators. *Curr Protoc* **1**, e266 (2021).

272. Riley, T. R. *et al.* SELEX-seq: A Method for Characterizing the Complete Repertoire of Binding Site Preferences for Transcription Factor Complexes. *Methods in molecular biology (Clifton, N.J.)* **1196**, 255–278 (2014).

273. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci.* **97**, 11984–11989 (2000).

274. Bevis, B. J. & Glick, B. S. Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.* **20**, 83–87 (2002).

275. Verkhusha, V. V. *et al.* High Stability of Discosoma DsRed As Compared to Aequorea EGFP. *Biochemistry* **42**, 7879–7884 (2003).

276. He, L., Binari, R., Huang, J., Falo-Sanjuan, J. & Perrimon, N. In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer. *Elife* **8**, e46181 (2019).

277. Segert, J. A., Gisselbrecht, S. S. & Bulyk, M. L. Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends Genet* **37**, 514–527 (2021).

278. Park, S.-Y. & Kim, J.-S. A short guide to histone deacetylases including recent progress on class II enzymes. *Exp. Mol. Med.* **52**, 204–212 (2020).

279. Jacobs, J., Pagani, M., Wenzl, C. & Stark, A. Widespread regulatory specificities between transcriptional co-repressors and enhancers in Drosophila. *Science* **381**, 198–204 (2023).

280. Floc'hlay, S. *et al.* Cis-acting variation is common across regulatory layers but is often buffered during embryonic development. *Genome Res* **31**, gr.266338.120 (2020).

281. Chatterjee, S. *et al.* Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. *Cell* **167**, 355-368.e10 (2016).

282. Malkmus, J. *et al.* Spatial regulation by multiple Gremlin1 enhancers provides digit development with cis-regulatory robustness and evolutionary plasticity. *Nat. Commun.* **12**, 5557 (2021).

Appendix

Supplementary Figure S1:

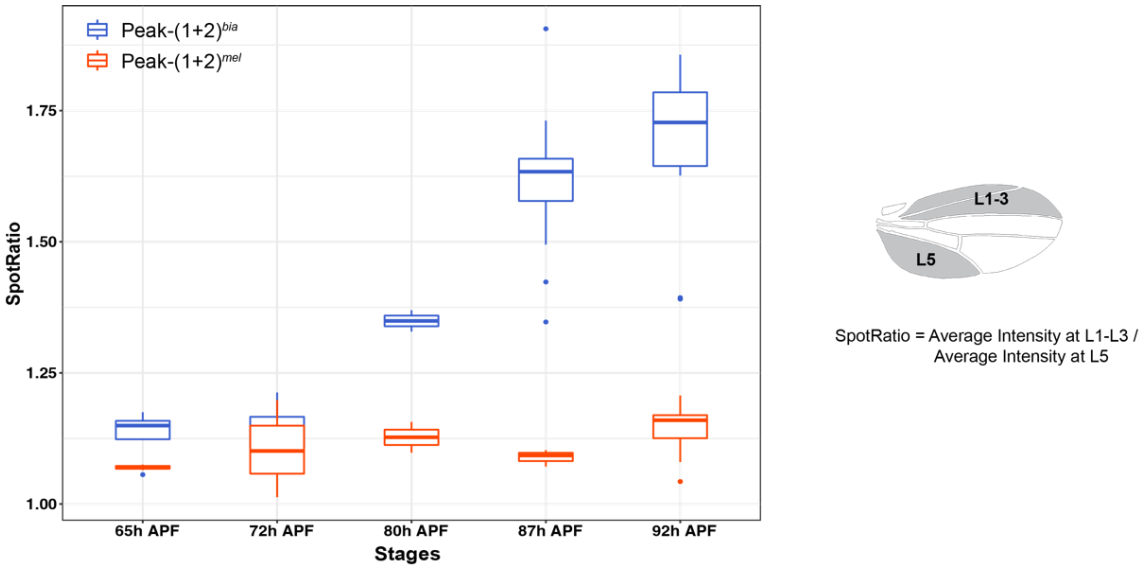

Figure S1. *The quantification of reporter expression pattern changes in late pupal wings, driven by yellow enhancers from D. biarmipes and D. melanogaster, respectively. Similar to the imaging quantification analysis presented in the Results, the left panel quantifies the reporter signal in pupal wings from different stages. The calculation of the spot pattern is indicated in the right panel. The number of individual wings collected for peak-(1+2)^{mel} from 65h APF to 92h APF is 2; 6; 2; 3; 8. The number of individual wings collected for peak-(1+2)^{bia} from 65h APF to 92h APF is 4; 3; 2; 13;11. The location of peak(1+2) is illustrated in Figure 7.*

Acknowledgments

Moving to Munich in the summer of 2017 marked the beginning of an unexpectedly long journey towards my Ph.D. Throughout these years, the support and encouragement from the people I encountered in Munich kept me going. While I cannot list all the help and memorable experiences I've had, I will certainly never forget them.

I want to express my gratitude to Prof. Nicolas Gompel for supervising my doctoral study and for his unwavering support over the years. Nicolas helped me work on projects that I was passionate about, encouraged me to learn advanced techniques and analysis, and provided opportunities to attend conferences and courses. Moreover, Nicolas encouraged me to lead collaborations with other labs, which helped me gain invaluable experience. I can hardly imagine having a doctoral study that was so enriching with diverse experiences. I am also thankful for Nicolas' patience and understanding when I was struggling with writing this dissertation.

During my time at the Gompel lab, I have received countless help and encouragement from my colleagues. Without the help of Rita and Bettina, I would not have been able to publish my paper last year. My time at the fly room would have been much less enjoyable if Yaqun had not been there with her optimistic presence. Therefore, I would like to extend my sincere appreciation to all the current and past members at Gompel lab, including Mari, Yunhan, Vijay, Xinyi, Conrad, Natascha, Stefano, Rita, Bettina, Yaqun, Deepak, Yann, and others for the help, support, encouragement, and memorable time we shared at the lab.

I feel grateful to have been a student at the QBM graduate school. The friendly and welcoming atmosphere at QBM was a great source of support to me during my first year in Munich, which was unexpectedly difficult. I want to express my appreciation to everyone I met at QBM. In particular, I am thankful to Julia for her help during my first year here. Without her encouragement, I might have quit my doctoral study a long time ago. I would also like to thank Markus for all the help and courses he offered over the years, as well as for the summit club. Also, I want to thank QBM for supporting me in attending conferences and meetings and conducting short-term research in the US.

I would like to extend my gratitude to Prof. Peter Becker and Dr. Justin Crocker for their guidance and support as members of my Thesis Advisory Committee. They offered valuable suggestions and advice throughout my doctoral study. Additionally, I extend my thanks to Prof. Richard Mann from Columbia University and Prof. Remo Rohs from USC for their support and for giving me the opportunity to collaborate with their labs. I was fortunate to conduct research at the Mann lab for a couple of months at the beginning of my doctoral study. Apart from the help I received from my colleagues at the Mann lab,

the encouragement from Judith, Bill, and Siqian and their thought-provoking discussions were instrumental in my research after that. I am also thankful for the insightful discussions with Remo and his lab, which led me to explore a new direction for my future study, although it is not included in this dissertation. Finally, I would like to thank Yingfei, Tsu-Pei, Yibei, and the other members of the Rohs lab for their warm hospitality during my visit to LA and for their contribution to the project.

I would like to thank Dr. Xiaodong Wang and all of my former colleagues from the Wang lab. I am truly grateful to Xiaodong for allowing me to work at his lab. The experiences and knowledge I gained during my several years there helped me immensely in my further studies. I am also honored to have continued support from Xiaodong and my former colleagues even after I moved to Munich.

I am grateful for all the wonderful moments I shared with my friends in Munich and China while climbing and playing badminton. Those moments we spent in the climbing gyms, in the mountains, and on the courts helped me relieve all the stress from my studies. I am also thankful to the people who helped me have a chance to enjoy those activities. Moreover, I appreciate the friendship of all my friends, near and far, for their endless support during my study. I want to thank Kan, Yi, and their friends in Munich for their help since my first visit to Munich and for all the memorable times we spent together.

Last but not least, I would like to express my deepest gratitude to my parents for their unwavering support throughout my academic and personal journey. Without their understanding and encouragement, I would not have had the opportunity to pursue my studies in the US, work in Beijing, spend years studying in Munich, or enjoy my limited part-time travel and climb. They have supported me through all the ups and downs, granting me the freedom to grow and become the person I aspire to be.

Curriculum vitae

Liucong Ling

Academic Education

Sept 2017 – Jun 2024	Ludwig-Maximilians-University Munich, Munich, Germany Graduate School of Quantitative Biosciences Munich (QBM) Faculty of Biology (Since Mar 2018) *Ph.D. Student*
Aug 2011 – Dec 2012	State University of New York at Stony Brook, Stony Brook, USA *Master of Science in Biochemistry and Cell Biology*
Sept 2007 – Jun 2011	Nanjing Forestry University, Nanjing, China *Bachelor of Science in Biotechnology*

Work Experience

Mar 2014 – Jun 2017	National Institute of Biological Sciences, Beijing, China Research Group of Dr. Xiaodong Wang *Research Assistant*

Publications

1. **Ling, L.**, Mühling, B., Jaenichen, R., & Gompel, N. (2023). Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*. *Science Advances*, *9*(7).
2. Poul, Y. Le, Xin, Y., **Ling, L.**, Mühling, B., Jaenichen, R., Hörl, D., … Gompel, N. (2020). Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Science Advances*, *6*(49), eabe2955.
3. Xin, Y., Poul, Y. Le, **Ling, L.**, Museridze, M., Mühling, B., Jaenichen, R., … Gompel, N. (2020). Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility. *Proceedings of the National Academy of Sciences*, *117*(34), 20636–20644.
4. Fan, W., Guo, J., Gao, B., Zhang, W., **Ling, L.**, Xu, T., … Wang, X. (2019). Flotillin-mediated endocytosis and ALIX–syntenin–1-mediated exocytosis protect the cell membrane from damage caused by necroptosis. *Science Signaling*, *12*(583).

Fellowships

Sept 2017 – Dec 2018	Deutschen Forschungsgemeinschaft (DFG) Ph.D. Fellowship (Through QBM)