

Evolution and regulatory logic of an enhancer underlying a novel pigmentation pattern in *Drosophila* wings

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München



Yaqun Xin

München, 2020

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

Erstgutachter: Prof. Dr. Nicolas Gompel

Zweitgutachter: Prof. Dr. John Parsch

Tag der Abgabe: den 19.10.2020

Tag der mündlichen Prüfung: den 11.12.2020

Declaration

Eidesstattliche Versicherung

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

München, den 10. 2020

Yaqun Xin

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist. Ich habe nicht versucht, anderweitig eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

München, den 10. 2020

Yaqun Xin

List of publications

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

Yaqun Xin, Yann Le Poul, Liucong Ling, Mariam Museridze, Bettina Mühling, Rita Jaenichen, Elena Osipova, Nicolas Gompel
PNAS August 25, 2020 117 (34) 20636-20644.

2. 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
Accepted by *Science Advances* on 09 October 2020.

3. Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele

Hélène Hinaux, Katharina Bachem, Margherita Battistara, Matteo Rossi, **Yaqun Xin**, Rita Jaenichen, Yann Le Poul, Laurent Arnoult, Johanna M. Kobler, Ilona C. Grunwald Kadow, Lisa Rodermund, Benjamin Prud'homme, Nicolas Gompel
Developmental Biology 15 June 2018.

Declaration of contribution

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

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

PNAS August 25, 2020 117 (34) 20636-20644.

Yaqun Xin performed most of the experiments, analyzed them together with other authors, and contributed to the design of the project. Yaqun Xin and Yann Le Poul are co-first authors.

Other contributions:

Yann Le Poul: designed the pipeline for wing quantification, analyzed and interpreted image quantification;

Liucong Ling: carried out the ATACseq experiments with help from Yaqun Xin;

Rita Jaenichen and Mariam Museridze: cloned some of the construct;

Bettina Mühling: transformed the reporter constructs;

Elena Osipova: contributed to the preparation of samples for imaging;

Nicolas Gompel and Yann Le Poul: conceived the project and wrote the manuscript with feedback from Yaqun Xin, Liucong Ling and Mariam Museridze;

Nicolas Gompel: analyzed and interpreted the data, and supervised the project.

Signature of the Supervisor:

Prof. Dr. Nicolas Gompel
10. 2020

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

Yann Le Poul, **Yaquun 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

Accepted by *Science Advances* on 09 October 2020.

Yaquun Xin performed most of the experiments, analyzed them together with other authors, and contributed to the design of the project. Yaquun Xin and Yann Le Poul are co-first authors.

Other contributions:

Yann Le Poul: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing—original draft, Visualization;

Liucong Ling, Mariam Museridze, Deepak Dharmadhikari, Eamonn Murphy: Investigation, Formal analysis;

Bettina Mühling, Rita Jaenichen and Elena Osipova: Investigation;

David Hörl and David Bunk: Software, Data curation;

Hartmann Harz: Methodology, Supervision;

Heinrich Leonhardt: Supervision;

Yingfei Wang: Methodology, Software, Formal analysis;

Remo Rohs: Methodology, Supervision, Funding acquisition;

Stephan Preibisch: Software, Supervision, Funding acquisition;

Benjamin Prud'homme: Conceptualization, Writing—original draft;

Nicolas Gompel: Conceptualization, Validation, Writing—original draft, Visualization, Supervision, Project administration, Funding acquisition.

Signature of the Supervisor:

Prof. Dr. Nicolas Gompel

10. 2020

3. Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele

Hélène Hinaux, Katharina Bachem, Margherita Battistara, Matteo Rossi, **Yaquun Xin**,
Rita Jaenichen, Yann Le Poul, Laurent Arnoult, Johanna M. Kobler, Ilona C. Grunwald Kadow,
Lisa Rodermund, Benjamin Prud'homme, Nicolas Gompel
Developmental Biology 15 June 2018.

Yaquun Xin contributed to the dissection and imaging of fluorescent pupae.

Signature of the Supervisor:

Prof. Dr. Nicolas Gompel
10. 2020

Declaration of contribution

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

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

PNAS August 25, 2020 117 (34) 20636-20644.

Yaqun Xin performed most of the experiments, analyzed them together with other authors, and contributed to the design of the project. Yaqun Xin and Yann Le Poul are co-first authors.

Contributions of Yann Le Poul:

Yann Le Poul conceived the project, designed the pipeline for wing quantification, analyzed and interpreted image quantification. Yann Le Poul and Nicolas Gompel wrote the manuscript with feedback from Yaqun Xin, Liucong Ling and Mariam Museridze.

Signature of Yann Le Poul:

10. 2020

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

Yann Le Poul, **Yaquun 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

Accepted by *Science Advances* on 09 October 2020.

Yaquun Xin performed most of the experiments, analyzed them together with other authors, and contributed to the design of the project. Yaquun Xin and Yann Le Poul are co-first authors.

Contributions of Yann Le Poul:

Yann Le Poul: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing—original draft, Visualization.

Signature of Yann Le Poul:

10. 2020

Table of Contents

Declaration.....	III
List of publications.....	IV
Declaration of contribution.....	V
List of abbreviations	1
Summary.....	3
Zusammenfassung	5
Introduction.....	7
1. Morphological diversity and gene regulation	7
1.1. From morphological diversities to changes in gene expression patterns.....	7
1.2. From changes in enhancers to morphological evolution	10
2. Transcriptional regulation.....	14
2.1. Transcription	14
2.2. Enhancers: general information	14
2.3. Enhancers and TF interactions.....	16
3. Enhancer evolution and morphological novelties.....	20
4. Enhancer identification	24
4.1. Enhancer prediction	24
4.2. Enhancer validation	26
4.3. Minimal enhancers.....	27
4.4. Quantifying the enhancer activities in the tissue	28
5. <i>Drosophila</i> pigmentation model	30
5.1. <i>Drosophila</i> pigmentation synthesis.....	30
5.2. <i>Drosophila</i> pigmentation regulation	31
6. A pipeline for the quantification of gene expression in <i>Drosophila</i> wing.....	35
Aim of the thesis	39

Results	41
Paper One: Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic site affecting chromatin accessibility.....	41
Manuscript Two: Regulatory encoding of quantitative variation in spatial activity of a <i>Drosophila</i> enhancer	73
Paper Three: Revisiting the developmental and cellular role of the pigmentation gene <i>yellow</i> in <i>Drosophila</i> using a tagged allele	117
General Discussion.....	139
Summary of findings.....	139
How to define an enhancer?.....	140
Pleiotropy in enhancer function	141
Evolution of novel enhancers from preexisting regulatory sequences	143
The regulatory logic of enhancer activities.....	145
References.....	151
Acknowledgements	169

List of abbreviations

ATAC	Assay for Transposase-Accessible Chromatin
aaNAT	arylalkylamine-<i>N</i>-acetyltransferase
ATP	adenosine triphosphate
CBP	CREB-binding protein
ChIP	Chromatin immunoprecipitation
CNS	central nervous system
COMPASS	Complex of Proteins Associated with Set1
CRE	<i>Cis</i>-regulatory element
DNA	deoxyribonucleic acid
DOPA	dihydroxyphenylalanine
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements
GTF	general transcription factors
MPRA	massively parallel reporter assays
NADA	N-acetyldopamine
NBAD	N-β-alanyldopamine
PCA	Principle component analysis
PCR	polymerase chain reaction
PIC	preinitiation complex
PRC	Polycomb repressive complex

RNAi	RNA interference
RNA	ribonucleic acid
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SWI2/SNF2	SWItch 2/Sucrose Non-Fermentable 2
STAT	Signal Transducers and Activators of Transcription
STARR-seq	Self-transcribing active regulatory region sequencing
TE	transposable element
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcription start site
UAS	upstream activation sequence

Summary

Novel morphological traits originate largely from the novel expression patterns of genes controlled by enhancers during development. Enhancers bind and integrate the spatial-temporal activity of transcription factors, and their combinatorial interplay determines the time, location and levels of transcriptional output. New enhancers can arise through enhancer co-option by reusing some of the regulatory information from a preexisting enhancer. While enhancer co-option is thought to be a fast and likely way to evolve new enhancers, its genetic and molecular mechanisms still remain elusive.

In this context, this thesis investigates the genetic origin of a novel enhancer and the regulatory logic underlying its function. I used the *spot* enhancer of the gene *yellow* as a model, which underlies the evolution of a morphological trait, the wing spot in *Drosophila biarmipes*. I sought to understand how the novel *spot* enhancer has evolved and what regulatory logic governs its function.

Specifically:

In the first chapter, I examined the evolutionary mechanism of *spot* enhancer in the context of the preexisting *wing blade* enhancer. By revisiting the entire *D. biarmipes yellow* 5' region with a comprehensive and quantitative method, I mapped the full activities of the novel *spot* and preexisting *wing blade* enhancers to a much larger region (3.5 kb) than previously described (1.1 kb together). Within the region, the regulatory information necessary and sufficient for the *spot* activity was inseparable from, and extensively overlapping with the *wing blade* activity. Further dissection of the shared core region revealed a pleiotropic binding site that contributed to both activities by regulating the local chromatin accessibility. I therefore confirmed that the novel *spot* activity originates from the co-option of the preexisting *wing blade* activity. The pleiotropic site for chromatin accessibility suggests a possible model where a new enhancer could evolve by co-option of chromatin accessibility input from the ancestral element, and that might facilitate the emergence and diversification of morphological traits.

In the second chapter, I investigated how the various aspects of regulatory information encoded in the *spot* enhancer sequences influenced its activity. Through introducing systematic mutations along the enhancer sequences and implementing a quantitative framework, the spatial activities on the wing of all the mutant enhancers were measured. The analysis showed an unexpected density of regulatory information within the *spot* enhancer. Moreover, it reveals an unanticipated regulatory logic underlying the activity of this enhancer and how it reads the wing trans-regulatory landscape to encode a spatial pattern.

The gene *yellow* is required for black pigment production and its expression in late pupal stage prefigures the adult wing spot pigmentation pattern. Therefore, understanding the dynamics of *yellow* expression is essential to elucidate the process of *yellow* enhancer regulation as well as pigment formation during development. Chapter three investigates the process of pigment formation in space and time using the pigment gene *yellow* in *D. melanogaster*. Firstly, a fluorescent protein-tagged *yellow* allele was generated, then the dynamics of *yellow* expression and cellular localization in relationship to the process of pigment formation was examined during development. It was found that *yellow* is expressed in a few neurons in the brain and the ventral nerve chord from the second larval instar to adult stage, indicating a neuro-developmental function of *yellow*. In addition, the results mainly showed how *yellow* expression in the adult cuticle is determined by regulated developmental processes affecting the body color, and suggested a structural role of Yellow in the establishment of pigmentation patterns.

Zusammenfassung

Der Ursprung von neu evolvierten morphologischen Merkmalen ist oftmals eine neuartiges Genexpressionsmuster, welches während der Entwicklung durch Enhancer kontrolliert wird. Transkriptionsfaktoren binden an diese Enhancer und deren räumlich-zeitliche Aktivität wird integriert. Das kombinatorische Zusammenspiel verschiedener Transkriptionsfaktoren und dem Enhancer bestimmt Zeit, Ort und Stärke der Gentranskription. Neue Enhancer können durch Enhancer-Kooptation entstehen, indem einige der regulatorischen Informationen von einem bereits vorhandenen Enhancer wiederverwendet werden. Obwohl man davon ausgeht, dass die Enhancer-Kooptation ein schneller und daher wahrscheinlicher Weg ist, neue Enhancer zu entwickeln, sind die genetischen und molekularen Mechanismen dieser Kooptation noch unbekannt.

Diese Arbeit untersucht den genetischen Ursprung eines neuartigen Enhancers und die seiner Funktion zugrunde liegende regulatorische Logik. Als Modell dient der *spot* Enhancer des Gens *yellow*, welches der Evolution eines morphologischen Merkmals zugrunde liegt, dem Flügelfleck bei *Drosophila biarmipes*. Ziel war es zu verstehen, wie dieser neuartige *spot* Enhancer evolviert ist und welche regulatorische Logik seiner Funktion zugrunde liegt.

Im Besonderen:

Im ersten Kapitel wird der evolutionäre Mechanismus untersucht welcher dem *spot* Enhancer zugrunde liegt, im Zusammenhang mit dem bereits existierenden *wing blade* Enhancer welcher eine generelle Flügelpigmentierung treibt. Indem wir die gesamte *yellow* 5'-Region von *D. biarmipes* mit einer umfassenden und quantitativen Methode überprüften, konnten Ich die gesamten Aktivitäten des neuartigen *spot* Enhancer und des bereits existierenden *wing blade* Enhancer auf eine viel größere Region (3,5 kb) abbilden als zuvor beschrieben (1,1 kb). Innerhalb dieser Region waren die regulatorischen Informationen, welche für die *spot* Aktivität notwendig und ausreichend waren, nicht von der *wing blade* Aktivität zu trennen und überlappten sich weitgehend mit ihnen. Die weitere Zergliederung der gemeinsamen Kernregion ergab eine pleiotrope Bindungsstelle, die zu beiden Aktivitäten beitrug, indem sie die lokale Chromatinzugänglichkeit regulierte. Ich konnten somit bestätigen, dass der neuartige *spot* Enhancer mittels der Kooptation des bereits vorhandenen *wing blade* Enhancers

evolviert ist. Diese pleiotrope Stelle für die Chromatinzugänglichkeit deutet auf ein mögliches Modell hin, bei dem sich ein neuer Enhancer durch Kooptation der Chromatinzugänglichkeitseingabe aus dem ursprünglichen Element entwickeln könnte, welches die Entstehung und Diversifizierung morphologischer Merkmale erleichtern könnte.

Im zweiten Kapitel untersuchten Ich, wie die verschiedenen Aspekte der in den *spot* Enhancer-Sequenzen kodierten regulatorischen Informationen deren Aktivität beeinflussten. Durch die Einführung systematischer Mutationen entlang der Enhancer-Sequenzen und die Implementierung eines quantitativen Rahmens wurden die räumlichen Aktivitäten auf dem Flügel aller mutierten Enhancer. Die Analyse zeigte eine unerwartete Dichte an regulatorischer Information innerhalb des *spot* Enhancers. Darüber hinaus enthüllt sie eine unerwartete regulatorische Logik, die der Aktivität dieses Enhancers zugrunde liegt und wie er die transregulatorische Landschaft auf dem Flügel liest, um ein räumliches Muster zu kodieren.

yellow ist für die Produktion von schwarzem Pigment erforderlich, und seine Ausprägung im späten Puppenstadium deutet auf das Pigmentmuster der erwachsenen Flügelflecken hin. Daher ist das Verständnis der Dynamik der *yellow*-Expression von wesentlicher Bedeutung, um den Prozess der *yellow* Enhancer-Regulation sowie die Pigmentbildung während der Entwicklung aufzuklären. Das dritte Kapitel untersucht den Prozess der Pigmentbildung in Raum und Zeit anhand des Pigmentgens *yellow* in *D. melanogaster*. Es wurde ein fluoreszenzmarkiertes *yellow* Allel erzeugt, und die Dynamik der *yellow* Expression und der zellulären Lokalisation im Zusammenhang mit dem Prozess der Pigmentbildung während der Entwicklung wurde untersucht. Es zeigte sich, dass Yellow in einigen wenigen Neuronen im Gehirn und in der ventralen Nervensehne vom zweiten Larvenstadium bis zum Erwachsenenstadium exprimiert wird, was auf eine neurologische Entwicklungsfunktion von Yellow hinweist. Darüber hinaus zeigten die Ergebnisse hauptsächlich, wie die *yellow* Expression in der adulten Kutikula durch regulierte Entwicklungsprozesse bestimmt wird, die die Körperfarbe beeinflussen, und legten eine strukturelle Rolle von Yellow bei der Etablierung von Pigmentierungsmustern nahe.

Introduction

1. Morphological diversity and gene regulation

1.1. From morphological diversities to changes in gene expression patterns

Today, we share the planet with millions of different species, with astonishing levels of diversity in behavior, physiology, habitats and morphology (Hickman 2018). Morphological diversity, which refers to the different characteristics in the shape, size, color, texture or organization of body structures, has been the primary criterion for species identification and classification (Foote 1997).

In animals, body structures are composed of modules, and each structure serves a unique role that is relatively independent from that of other structures. Modular organization allows morphological components to respond to natural selection individually without necessarily changing other parts. Therefore, the evolution of the genetic mechanism controlling the development of one body part can be dissociated from that of another body part. This allows further modular modifications of an individualized morphological trait and promotes morphological diversification (Gilbert, Opitz, and Raff 1996; Bedau 2009; Wagner 1996).

For example (Figure 1), some higher Diptera share a common wing plan regarding the shape and the venation pattern, but have evolved a variety of wing pigmentation patterns with different colors, dots and lines on the wing (Prud'homme, Gompel, and Carroll 2007). These diverse wing pigmentation patterns represent a starting place to study the genetic changes underlying morphological differences.

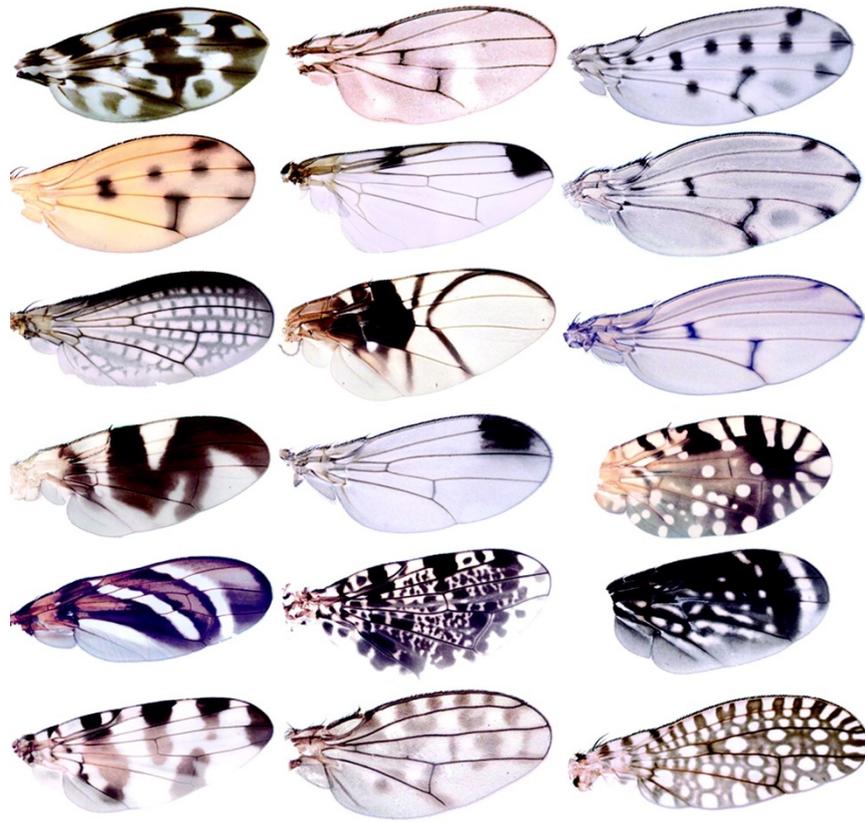


Figure 1: The various pigmentation patterns on the wings of some higher Diptera. Adapted with permission from (Prud'homme, Gompel, and Carroll 2007). Copyright (2007) National Academy of Sciences.

A main goal in biology is to identify the genetic mechanisms underlying morphological evolution. In this respect, one first step is to understand the developmental process producing these forms, and how genes control it. This involves the field of evolutionary developmental biology, referred to as “Evo Devo” (Gilbert 2003; Hall 2012).

Numerous genes that control the specification and segmentation of body structures have been identified by isolating mutants with morphological abnormalities in *Drosophila*, (Wakimoto and Kaufman 1981; E. B. Lewis 1978). These developmental genes are called toolkit genes. Characterizations of the toolkit genes revealed that many of them encoded transcription factors (TFs) or signaling molecules that regulate the expression of other genes to control tissue-specific functions during development (Zaraiskii 2001; Krumlauf 1994; Hueber and Lohmann 2008). Later, with more toolkit genes being identified in *Drosophila* as well as in vertebrates

and other animals, comparisons of the structure and biological function of the toolkit genes revealed high conservation among distantly related species (Tayas et al. 2006; Schneider and Amemiya 2016; Sanetra et al. 2005; Carroll 1995).

These discoveries, however, lead to an apparent paradox: if different animals, from flies to humans, use a similar set of toolkit genes in development, how has all the morphological diversity arisen from a common ancestor? The answer lies not in what the genes are but in how the genes are used during development, including their dynamic expression in space, time and levels and their interactions with other genes during the formation of diverse morphological structures. Molecular techniques such as in situ hybridization (Gall and Pardue 1969) or antibody staining (Coons, Creech, and Jones 1941) allow us to directly visualize the distribution of gene products, RNAs or proteins, respectively (Akam 1983; Coons, Creech, and Jones 1941; Ransick et al. 1993), especially during development. It turns out that the adult body plan is pre-patterned by expression patterns of the developmental genes in the embryo. The precise spatial-temporal expression of these genes during development prefigures and determines the final morphology (Pechmann et al. 2011; Niwa et al. 1997; Schaefer, Oliver, and Henry 1999; Jang et al. 2003; Bandyopadhyay et al. 2006; Pandur et al. 2013).

This finding has related the action of the invisible genes during development to the visible phenotypic trait in the adult. The dynamic expression pattern of the key regulatory genes are thus snapshots of the unfolding process of morphology formation during development.

Over the past three decades or so, a vast body of comparative studies across all taxonomic levels has proven that variation in gene expression patterns correlated with the evolution of morphological traits (Rawls and Kumar 2002; Carroll 2005). One iconic example was found in the Darwin's finches (Figure 2), which revealed a correlation between the changes in developmental gene expression pattern and the differences in morphological traits between closely related species. Different closely related species have evolved different shapes of the beak associated with the exploitation of various food types. Several key genes are involved in the formation of the beak. The comparative analysis of expression patterns of these genes revealed two genes whose expression correlated with the specific shape of the beak. The expression pattern of the gene *bone morphogenetic protein 4* (*Bmp4*) in the beak was shown to correlate with the depth and width of beaks among ground finches, while the expression pattern

of another target *calmodulin* (*CaM*) was shown to correlate with the long and pointed beak morphology of the cactus finches (Abzhanov et al. 2004; 2006).

In summary, changes in gene expression of developmental genes can produce morphological variation. The question of morphological diversity then can be switched to the question of how changes in gene expression patterns occur.

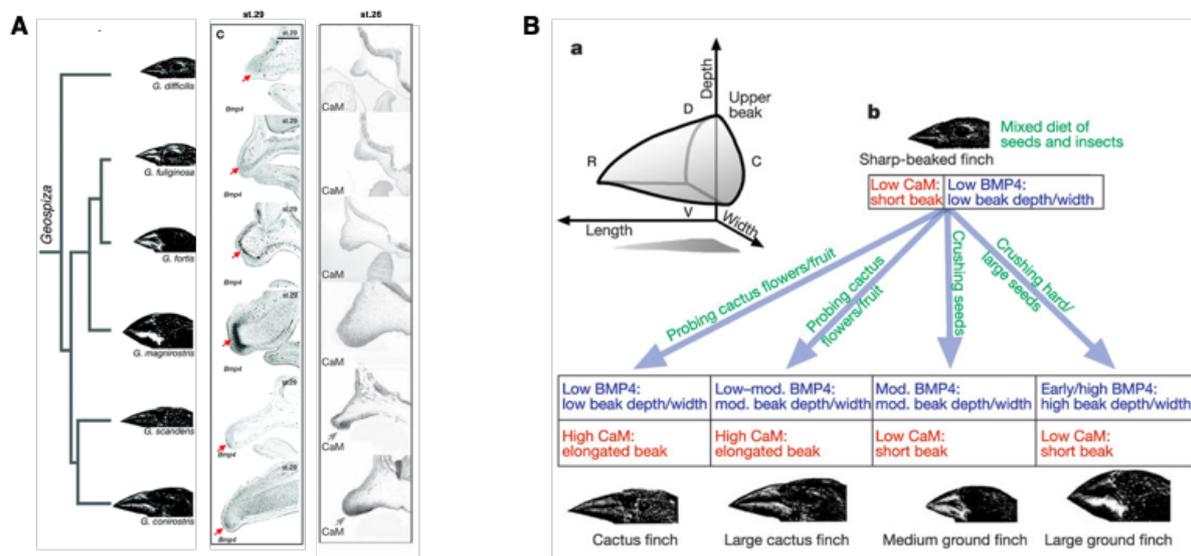


Figure 2: Changes in gene expression correlate with the diverse morphology of beaks in Darwin finches. (A) The finches display distinct beak morphologies between different species. At stage 29 (st.29), *Bmp4* is expressed in the distal-dorsal domain in the mesenchyme of the beak in species with wide beaks. *CaM* is expressed in the distal-ventral domain in the mesenchyme in species with long beaks. (B) a. The 3D structure of the developing beak. b. The model of how the expression of *Bmp4* and *CaM* regulate the development of the length and width of the beaks. Modified with permission from (Abzhanov et al. 2004; 2006).

1.2. From changes in enhancers to morphological evolution

Changes in gene expression patterns can result from the alterations of TFs that regulate the expression of these genes, and/or the *cis*-regulatory elements (CREs) mostly called enhancers which are non-coding DNA containing binding sites for TFs (C.-T. Ong and Corces 2011).

As TF coding sequences and function tend to be conserved across species, and as they are deployed in multiple tissues and developmental stages, mutations in their protein coding sequence are likely to be pleiotropic (Carroll 2008). By contrast, the relative low pleiotropy and modularity of enhancers poise them to accommodate mutations resulting in gene expression changes.

Evidence from whole-genome comparisons and the deeper understanding of the structure of enhancers has led to propose that changes in enhancers are a major source of morphological variation (Carroll 2000; Wittkopp and Kalay 2012; Kalay 2012; Nagy et al. 2018; Camino et al. 2015; Madgwick et al. 2019; Wittkopp, Vaccaro, and Carroll 2002a; Gompel et al. 2005; Jeong, Rokas, and Carroll 2006; Indjeian et al. 2016; Attanasio et al. 2013; M.-C. King and Wilson 1975; Britten and Davidson 1969; Carroll 2005). The whole-genome comparisons from chimpanzee and human show that chimpanzee shares about 99% of the human DNA, while the 1% variation is found in non-coding DNA sequences (Ebersberger et al. 2002; M.-C. King and Wilson 1975).

Protein-coding sequences are embedded in the vast non-coding sequences, which include enhancers. Enhancers function as genetic switches to determine where, when and how much a gene is expressed in the tissue during development (Melamed et al. 2016; Mike Levine 2010; Long, Prescott, and Wysocka 2016). The non-coding region of many genes encoding regulatory proteins was found to contain multiple arrayed enhancers. Each enhancer independently regulates gene expression in a specific tissue (Koshikawa 2015; Martin 2014; Gaunt and Paul 2012; Serfling, Jasin, and Schaffner 1985; Prud'homme et al. 2006; Wenick and Hobert 2004; Melamed et al. 2016; Li, Notani, and Rosenfeld 2016). Gene regulation by these individual enhancers is therefore modular, and mutational changes in one enhancer are predicted to only affect gene expression in the tissue where the enhancer drives expression. Therefore, the modular gene regulation facilitates the evolution of morphological traits of one body part independently of other parts, minimizing the deleterious penalty on fitness cost (Rebeiz et al. 2009; Carroll 2008; Adachi et al. 2003; Gomez-Skarmeta et al. 1995).

A great number of studies has demonstrated that changes in enhancers underlies the diversity of morphological traits (Wittkopp, Vaccaro, and Carroll 2002b; Gompel et al. 2005; Shapiro et al. 2004; E. Sucena et al. 2003; É. Sucena and Stern 2000; Belting, Shashikant, and Ruddle 1998). A typical example of changes in enhancers contributing to morphological variation is the reduced pelvic fin in stickleback fish (Figure 3). Compared to the marine stickleback fish, the freshwater sticklebacks have reduced pelvic fins. The changes in the size of the pelvic fin has been shown to be associated with the gene expression pattern of *pitx1*, which encodes a TF functioning in multiple tissues (Shapiro et al. 2004). The expression of *pitx1* is regulated by multiple discrete regulatory elements, each of which governs the gene function in a certain tissue. The pelvic loss in freshwater stickleback fish was shown to result from the specific loss of activity of the hindlimb element, whereas other elements regulating *pitx1* in other structures remained unaffected (Chan et al. 2010).

To conclude, this chapter has pointed the paradox that different species use a common set of toolkit genes, and how the conserved toolkit genes find their different use in various gene expression patterns, which are controlled by enhancers. In the next chapter, I will describe the molecular bases underlying enhancer function in generating precise spatial-temporal gene expression patterns.

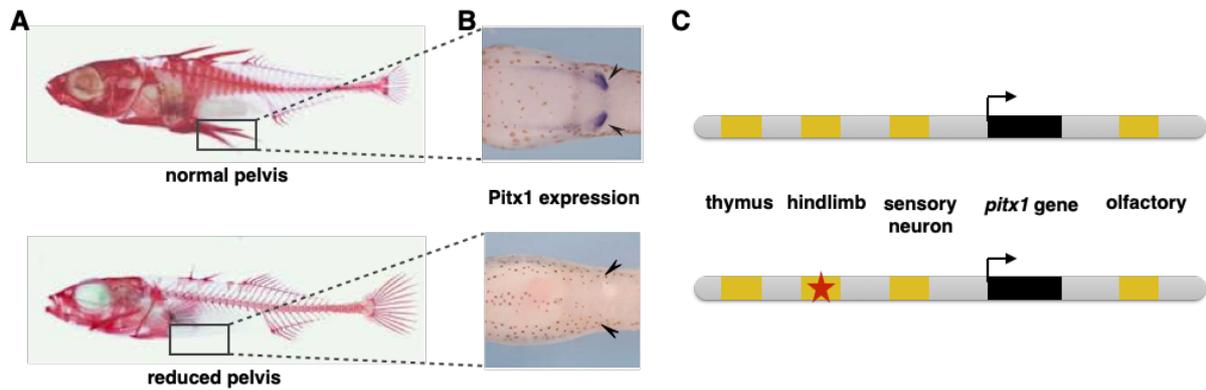


Figure 3: The evolution of the reduced pelvis in the freshwater sticklebacks. (A) The marine sticklebacks (top) have normal pelvic fins compared to the freshwater sticklebacks (bottom) which evolved reduced pelvic fins (both are indicated by the empty squares). (B) Gene *pitx1* is expressed in pelvic buds of the marine stickleback larvae, but not of freshwater sticklebacks (highlighted by the arrowheads). (C) The absence of *pitx1* gene expression is due to the disruption (red star) of the hindlimb enhancer of the *pitx1* gene, whose expression is controlled by multiple independent modular enhancers (yellow boxes). Figure 3B is modified with permission from Nature Springer (Shapiro et al. 2004), copyright 2004. Photos of Figure 3A were taken by Mike Shapiro.

2. Transcriptional regulation

2.1. Transcription

The regulation of gene expression occurs at different levels: transcription, post-transcriptional events, translation and post-translational modifications (Reményi, Schöler, and Wilmanns 2004).

During development, the regulation of transcription in space and time, which is the first step of gene expression, plays a major role (Michael Levine and Tjian 2003). Transcription is the process where genetic information contained in DNA is copied to a complementary RNA molecule that exits the nucleus. It involves large protein complexes, including, for most eukaryotic genes, the RNA polymerase II. In eukaryotes, gene transcription takes place in nucleus with three sequential steps: initiation, elongation and termination. Among these three steps, transcription initiation is the most controlled step and involves, on the DNA side, promoters and enhancers (Kadonaga 1998; Lee and Young 2000; Cramer 2019).

Promoters locate near the transcription start site (TSS) of the target gene. Promoters can be classified as core promoters which are within 100 bp around the TSS and proximal promoters which locate several hundred base pairs upstream of the TSS. Transcription typically initiates at the core promoter region. General transcription factors (GTFs) bind to the core promoter and recruit RNA polymerase II to form transcription pre-initiation complex (PIC) (Juven-Gershon et al. 2008). The proximal promoters bind to specific transcription factors to determine the tissue specificity of gene expression (Smith et al. 2006).

2.2. Enhancers: general information

Core promoters are sufficient to initiate gene transcription, but at a low basal level. To increase the transcription rate, more distantly located regulatory elements called enhancers are needed. The first enhancer was discovered in 1981 as a 72-bp sequence from the Simian Virus 40 late gene region, which was found to increase the transcription of a reporter gene by several hundred fold regardless of its relative position and orientation to the gene TSS (Moreau et al. 1981; Banerji, Rusconi, and Schaffner 1981).

Since then, enhancers were identified as short DNA sequences, usually a few hundred base pairs in length. Enhancers can be located upstream, downstream, or even within the intron of the gene and increase transcription independently of their orientation, position or distance relative to the TSS (W. Li, Notani, and Rosenfeld 2016; Melamed et al. 2016).

Functional enhancers contain clusters of 6-12 bp transcription factor binding sites (TFBSs) which can each be recognized by a specific TF (Shlyueva, Stampfel, and Stark 2014; C. T. Ong and Corces 2011; Li, Notani, and Rosenfeld 2016; Khoury and Gruss 1983; Christina Ione Swanson 2010). Once bound by TFs, enhancers can act at a distance by looping onto the core promoter, bringing activators at enhancers and general transcription factors and RNA polymerase II at the promoter together (Tolhuis et al. 2002; Krivega and Dean 2012) to increase the initiation rate of transcription (Figure 4). As a result, the distance between the TFBSs and the promoter can affect the loop formation and thus affect the efficiency of transcription (Nolis et al. 2009).

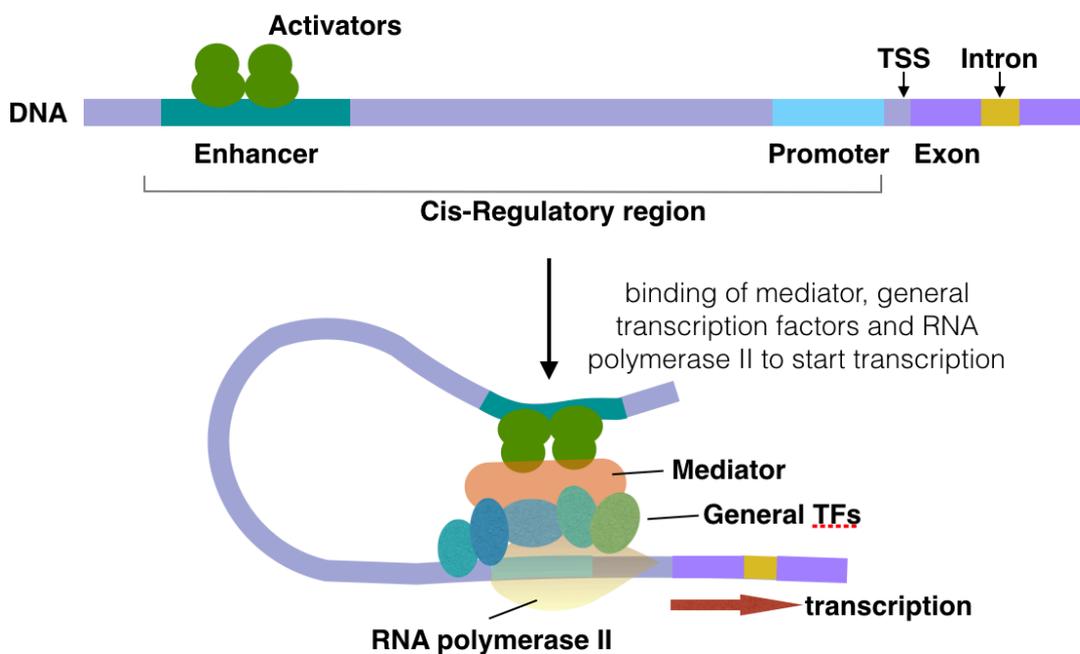


Figure 4: Overview of eukaryotic transcriptional control

2.3. Enhancers and TF interactions

Understanding the rules governing transcriptional regulation is important to predict what changes in the DNA sequences or TFs may affect gene expression, and subsequently morphological evolution.

The combinatorial interactions between enhancers and bound TFs determine the precise temporal and spatial gene expression (Reményi, Schöler, and Wilmanns 2004). Overall, the TF-enhancer interactions can be influenced by several features *in vivo*.

The TFs:

TFs are expressed spatially and temporally in a tissue-specific manner during development. Different TFs recognize and bind distinct DNA motifs with variations in the length and composition of the sequence, and even with different sensitivities and strengths to the same motif (Slattery et al. 2014; Khoueiry et al. 2017; Levy and Hannenhalli 2002; Nitta et al. 2015; Spitz and Furlong 2012). Thus, regulation of transcription is correlated with the TF binding specificity and affinity. *In vivo*, TFs often cooperatively interact with each other and their cofactors to increase their binding specificity and affinity. How the activity of TFs and cofactors is tuned is therefore essential for the precise control of transcription (Slattery et al. 2011; Siggers et al. 2011).

Chromatin accessibility:

Even when the TFs are present with proper activity in a tissue, transcription might also not happen. This is because in eukaryote the chromatin fiber can block the access of TFs and RNA polymerases to enhancers (Kornberg 1977).

Decompaction of the chromatin requires a specific class of chromatin modifying factors including histone modifying enzymes and ATP-dependent chromatin remodelers (Workman and Kingston 1998; Kingston and Narlikar 1999; Kadonaga 1998). Post-translational modifications of histone tails by histone-modifying enzymes, such as methylation and acetylation of specific residues, can affect the interplay between TFs and enhancers. For example, it has shown that active enhancers are marked with mono-methylation on lysine 4 (H3K4me1) and acetylation on lysine 27 of histone H3 (H3K27ac), while poised enhancers are marked with H3K4me1 and tri-methylation on lysine 27 of histone H3 (H3K27me3)

(Heintzman et al. 2009; Ernst et al. 2011; Heintzman and Ren 2009; C.-T. Ong and Corces 2011).

ATP-dependent chromatin remodelers, such as the SWItch 2/Sucrose Non-Fermentable 2 (SWI2/SNF2) group can reposition or evict nucleosomes so that inaccessible TFBSs can be exposed to TFs and RNA polymerases (Becker and Workman 2013).

Another special class of factors, the pioneer factors, have the intrinsic ability to bind nucleosomal DNA in condensed chromatin early in development, resulting in the opening of chromatin to create a permissive environment for transcription. Unlike other TFs, pioneer factors have the ability to bind directly to closed chromatin can evict histones from the chromatin (Cirillo L.A. et al. 2002). However, the binding of pioneer factor itself is insufficient to trigger the enhancer activity. Pioneer factors can recruit other TFs and work cooperatively to initiate gene transcription (Iwafuchi-Doi 2019; Iwafuchi-Doi and Zaret 2014; Soufi et al. 2015; Zaret and Carroll 2011).

Enhancer grammar:

The organization of TFBSs in the enhancer sequence, sometimes referred to as “enhancer grammar”, can also affect TF interactions. “Enhancer grammar” is defined as the number, order, orientation and spacing of the TFBSs and greatly affect the TF-enhancer interactions (Long, Prescott, and Wysocka 2016).

Mainly three models have been proposed regarding the grammar of the enhancer (Spitz and Furlong 2012) (Figure 5). The “enhanceosome model” represents a rigid enhancer architecture, where the enhancer activity depends on a precise array of TFBSs, which form a platform for a TF scaffold. Any changes in the TFBS organization can disrupt the protein-protein interactions and thus the enhancer activity (Merika and Thanos 2001; Thanos and Maniatis 1995). On the contrary, the “billboard model” posits a more flexible binding site grammar. Enhancer activity is maintained by the presence of the TFBSs but with great flexibility as to the order, orientation or spacing of these binding sites (Kulkarni and Arnosti 2003; Arnosti and Kulkarni 2005). This model can help explain the rapid motif turnover across species (Farley et al. 2015). The “TF collective” model represents a “no grammar” enhancer function model. In this model, TFs, which are either recruited directly by TFBSs or indirectly by other TFs, act collectively to regulate the enhancer activity. The recruitment of the collective binding is mediated by the

high-affinity TFBSs for a subset of TFs and thus not any specific TFBS is required for the enhancer function (Erceg et al. 2014; Junion et al. 2012).

Until today, the regulatory logic of “enhancer grammar” is still not fully understood. The above models to investigate the TFBS organization are tested based on the output of the enhancer activity through changing the relative order, number or distance of the known TFBSs within an enhancer. However, we are far from interpreting the enhancer activity through its sequences. The effect of mutated TFBSs on transcriptional output might come from epistatic interactions between TFBSs, which makes the results complicated to interpret (Michael Z Ludwig et al. 2000; Doniger and Fay 2007). Besides, there is a large gap in our understanding of the function of sequences between the known TFBSs. For most enhancers, the known TFBSs are not sufficient to generate the enhancer activity (Vincent, Estrada, and DePace 2016), suggesting that other uncharacterized TFBSs, as well as the remaining sequences between the TFBSs play a role too. Therefore, uncovering more of the sequence determinants of enhancer activity requires to run more systematic scans along enhancer sequences.

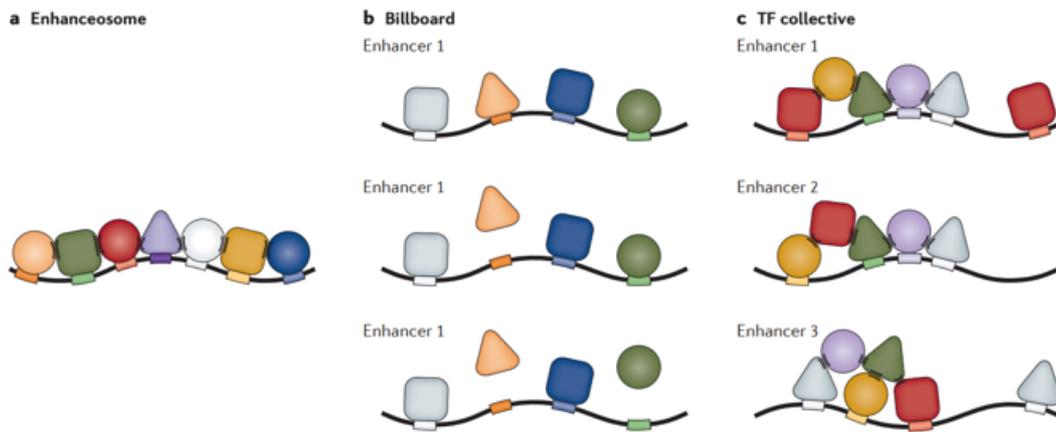


Figure 5: Models of enhancer grammar. a. The enhanceosome model, which represents a highly rigid enhancer architecture. Enhancer activity requires the presence and precise position of all the TFBSs. b. The billboard model, which represents a grammar that enhancer activity is maintained by more flexible organizations of the TFBSs. c. The TF collective model, in which TFs, either recruited directly or indirectly, act collectively to regulate the enhancer activity. From (Spitz and Furlong 2012) with permission.

3. Enhancer evolution and morphological novelties

Morphological novelties have been defined as new body structures or patterns, derived from an ancestor, and described as a qualitative change rather than just a quantitative modification in size or shape of a pre-existing trait (Wagner and Lynch 2010). Morphological novelties arise from the novel execution of organ-specific gene regulatory program during development, which is mainly controlled by enhancers (Monteiro and Podlaha 2009).

Multiple possible mechanisms can account for the origin of new enhancers, which are mainly summarized into four possible scenarios (Rebeiz and Tsiantis 2017) (Figure 6).

Transposable elements (TEs) have been shown to be a rich source for the generation of new enhancers, as TEs account for a large proportion of eukaryotic genomes (de Koning et al. 2011; Feschotte 2008). TEs carrying regulatory elements or functional binding sites can be inserted into different locations of the genome, and thus provide raw materials to derive novel enhancer activities (Feschotte 2008). The role of TEs in the evolution of novel gene expression patterns was shown by a number of studies (Emera and Wagner 2012; Lynch et al. 2011; Ting et al. 1992). For instance, TEs contributed to the evolution of a novel gene regulatory network during the pregnancy of placental mammals. It is found that TEs can directly bind TFs required for pregnancy and regulate related gene expression (Lynch et al. 2011).

Another scenario for the origin of a new expression pattern is promoter-switching. An enhancer often interacts with a specific target promoter (Burgess-Beusse et al. 2002; Butler and Kadonaga 2001). Altered enhancer-promoter specificity can therefore switch the enhancer to another promoter and result in the control a different gene. In practice, chromosome inversions can lead to novel enhancer-promoter interactions by replacing an enhancer nearby a distinct gene or by removing insulators between an enhancer and a promoter (Cande, Chopra, and Levine 2009; Imsland et al. 2012). In the beetle *Tribolium castaneum*, a chromosomal inversion allows a conserved ladybird enhancer to redirect the expression of a neighboring gene *C15* to generate a novel pattern of gene expression. In *Drosophila*, due to an insulator element at the ladybird promoter, this enhancer fails to activate *C15* expression. Such promoter-switching via genome rearrangements might be common in the evolution of diversification of the arthropods (Cande, Chopra, and Levine 2009).

New enhancers can also simply emerge *de novo* through the accumulation of random mutations from neutral DNA sequences. However, to form a cluster of functional TFBSs to generate a new expression pattern takes more evolutionary steps (Rebeiz and Tsiantis 2017; Gompel and Prud'homme 2009) than in other evolutionary scenarios.

Enhancer co-option is another mechanism for the origin of a new expression pattern. In this case, a novel activity can be derived by reusing some of the regulatory sequences or even TFBSs from the preexisting enhancers (Rebeiz and Tsiantis 2017), while keeping the preexisting enhancer activity unchanged. Therefore, sequence overlap between a new activity and an old activity, is a signature of co-option. It can also result in sequence dependency of the new and old activities.

Several studies have revealed that enhancer co-option can facilitate the evolution of novel enhancer activities and suggested that enhancer co-option might be common to derive morphological novelties (Koshikawa et al. 2015; Rebeiz et al. 2011; Gompel et al. 2005; Glassford et al. 2015). It is likely that within a region already containing an activity, bound by TFs with diverse expression patterns and with a favorable chromatin environment, fewer mutations are required to generate a novel activity. Compared to *de novo* emergence of an enhancer, enhancer co-option therefore takes fewer evolutionary steps for the origin of new enhancers (Rebeiz and Tsiantis 2017; Gompel and Prud'homme 2009).

In *Drosophila santomea*, the *Nepriylisin-1* (*Nep1*) gene has evolved a novel expression pattern in the optic lobe, which is governed by the *Nep1* optic lobe enhancer. It is found that the optic lobe enhancer is localized in a region significantly overlapping with preexisting enhancers driving expression in multiple tissues such as the retinal field, larval central nervous system (CNS), leg, and wing hinge. This indicates that the novel *Nep1* optic lobe enhancer has evolved by co-option of preexisting regulatory sequences (Rebeiz et al. 2011).

In another species *Drosophila guttifera*, a novel domain of *wingless* (*wg*) expression at the wing vein tips was found associated with an evolved vein-tip enhancer. Analysis of this vein-tip enhancer revealed that it overlapped with an ancestral enhancer that was active in the wing crossveins, indicating that the novel vein-tip enhancer of *wg* might emerge by enhancer co-option (Koshikawa et al. 2015).

The molecular correlates of enhancer co-option are the co-opted TFBSs and their functions in the novel and ancestral developmental contexts (Rebeiz and Tsiantis 2017). However, finding the binding sites used by the new and the ancestral enhancers is challenging since it requires detailed dissection within the overlapping region, and until today few studies have explored enhancer co-option in depth at the site level. One representative example is the origin of the male-specific posterior lobe of the male genitalia in *D. melanogaster*. The authors showed that the developmental program of the derived posterior lobe during pupal stage has recruited an ancestral *Hox*-regulated network required for the development of an ancestral structure, the posterior spiracle of the embryo. Binding sites for STAT and Abdominal-B (Abd-B) of an enhancer of *poxn* gene in this network were found to be used by both structures (Glassford et al. 2015). This example demonstrates how TFBSs from ancestral developmental contexts can be co-opted to a novel developmental context.

When a common set of binding sites are used in more than one developmental context, it results in TFBS pleiotropy (Preger-Ben Noon et al. 2018). The existence of pleiotropy in enhancers may impose more constraints on the enhancer sequence evolution, since a mutation affecting one expression pattern will also affect others (Andersson et al. 2014; Huang, Gulko, and Siepel 2017; Sabarís et al. 2019; Infante et al. 2015a), potentially resulting in deleterious effects.

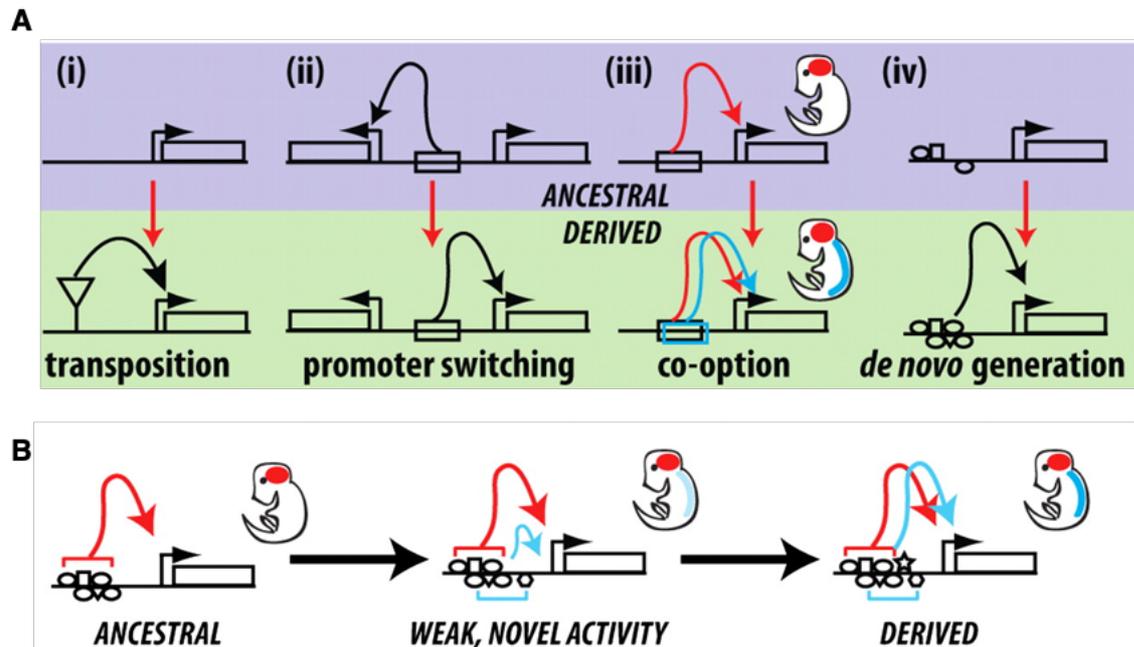


Figure 6: Scenarios of enhancer evolution. (A) The 4 different ways for the origin of new gene expression patterns. (B) The steps of enhancer co-option event. From (Rebeiz et al. 2011) with permission.

4. Enhancer identification

Given their importance for development and evolution, methods to identify or predict enhancers have been explored since the discovery of enhancers (Shlyueva, Stampfel, and Stark 2014). However, unlike the well-defined protein coding sequences (Lee and Young 2000), enhancers lack general sequence code. Also, a gene can have multiple enhancers that can locate anywhere from the TSS of the target gene (Melamed et al. 2016). All these have made it challenging to locate enhancers in the genome.

4.1. Enhancer prediction

Classically, enhancers have been identified using reporter gene assays *in vivo*. In these assays, DNA sequences located upstream or downstream of the TSS are isolated, fused with a minimal promoter and a reporter gene, and transformed to the germline of an organism for permanent integration into its genome (Shlyueva, Stampfel, and Stark 2014; O'Kane and Gehring 1987; Chiocchetti et al. 1997)

In the recent years, the development of genomic technologies and next-generation sequencing technologies have revolutionized enhancer identification on a genome-wide scale.

Initially, comparing genomic sequences between different species was used to predict enhancers. Functional regulatory sequences are under increased evolutionary constraint compared to non-functional sequences and are thus more conserved between different species (Visel, Bristow, and Pennacchio 2007). While this method is easy to compute, it is insufficient to predict the stage- and tissue-specific activity of enhancers. Besides, it may fail to predict functionally conserved enhancers with divergent sequences. Finally, it hinders the discovery of newly evolved regulatory elements (Kheradpour et al. 2007; Yáñez-Cuna et al. 2012; Blow et al. 2010).

Some of these limitations of sequence comparison can be overcome with a more recent method: chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) (Solomon, Larsen, and Varshavsky 1988). ChIP using antibodies against active enhancer histone marks (e.g., H3K4me1 and H3K27ac) and enhancer cofactors (e.g., CBP/P300) can reveal the dynamic

patterns of enhancer activity from specific stages or cell types (Yáñez-Cuna et al. 2012). However, because none of the known marks is shown to perfectly correlate with enhancer activity, this method can generate a number of false positive targets that are not functional *in vivo* (Kvon et al. 2012; Bonn et al. 2012).

Chromatin accessibility represents another complementary approach for genome-wide enhancer predictions as open chromatin is necessary for enhancer activity. Several methods have been developed to provide a genome-wide map of accessible chromatin regions, like DNase-seq (Hesselberth et al. 2009), MNase-seq (Yuan et al. 2005), Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq) (Giresi et al. 2007), and more recently Assay for Transposase-Accessible Chromatin (ATAC-seq) (Buenrostro et al. 2013). Compared to other methods, ATAC-seq requires fewer cells and a simpler protocol and gives base pair resolution of the chromatin accessible regions. ATAC-seq is performed via two-steps that involves chromatin tagmentation with sequencing adapters using the Tn5 transposase (Goryshin and Reznikoff 1998) followed by PCR amplification using barcoded primers. DNA regions identified based on chromatin accessibility are not only enhancers but also other active CREs. Furthermore, not all open regions correspond to active and functional enhancers as activation of enhancer also requires other factors during development (Arnold et al. 2013; Thurman et al. 2012).

STARR-seq (self-transcribing active regulatory region sequencing) is a technique to identify enhancers directly and quantitatively in a genome-wide manner (Arnold et al. 2013). In this experiment, enhancer fragments are placed downstream of a minimal promoter, so that active enhancers can transcribe themselves. The enhancer library is then transfected into cells and the resulted enhancer activity is measured by RNA sequencing. STARR-seq provides functional identification of enhancers. It can also measure enhancer activities from “closed” chromatin. However, it measures the enhancer activity outside the endogenous chromatin environment (Muerdter, Boryń, and Arnold 2015).

Combinations of these predictive approaches can increase the success rate of enhancer identification. However, it is important to note that the evidence provided by the genome-wide mapping of genomic features is just an indication of potential enhancer activity and should not be taken as equivalent to demonstrating regulatory function. Therefore, direct functional validation of enhancers is required.

4.2. Enhancer validation

The enhancer-reporter assays are widely used to study the function of enhancers. One way to use reporter assays to validate enhancers is by massively parallel reporter assays (MPRAs) (Inoue and Ahituv 2015). Typically, libraries of thousands of barcoded enhancer-reporter vectors are introduced into cell lines or tissues via transient transfection or electroporation (Uchikawa 2008; Corbo, Levine, and Zeller 1997). The reporter expression is then quantified by RNA-seq (Mortazavi et al. 2008; Emrich et al. 2007).

MPRAs can validate enhancers in a high throughput with a quantitative readout, and have identified numerous functional enhancers (Kwasnieski et al. 2012; Kheradpour et al. 2013; Wenick and Hobert 2004; Inoue and Ahituv 2015). However, since most of the MPRAs are conducted in an episomal manner and the enhancer-reporter constructs do not incorporate into the genome, these systems lack chromatin environment for gene expression (Inoue and Ahituv 2015). Besides, in this case the spatial-temporal expression pattern of reporter gene is lost.

Another way to validate enhancers is through transgenic assays. In *Drosophila* and other model organisms where transgenesis is possible, one primary validation method is through functional transgenic reporter assays *in vivo*, which provides a direct read-out of enhancer activity in different tissues and time points during development (Kvon 2015) (Figure 7).

In *Drosophila*, reporter assays have heavily relied on P element-mediated transgenesis. P elements are transposable elements (Castro and Carareto 2004), and their introduction has brought a breakthrough to *Drosophila* transgenesis (Rubin and Spradling 1982). However, P element can only integrate DNA fragments of relatively small size (up to a few tens of kb). Moreover, the insertion location of P element is random, increasing expression variability between lines (Venken and Bellen 2007). To be able to compare any quantitative changes between different enhancer activities, it is desirable to have all constructs inserted into the same genomic location. This can be achieved by site-specific integration using the bacteriophage ϕ C31 integrase. The ϕ C31 integrase catalyzes the irreversible recombination between the phage attachment (*attP*) site in the bacteriophage genome and a bacterial attachment (*attB*) site in the host genome (Groth et al. 2004). Thus, the ϕ C31-mediated transgenesis can result in the

site-specific integration of a vector containing attB sites into an organism genome containing attP “docking” sites (Groth et al. 2000).

4.3. Minimal enhancers

Such functional validation of the predicted enhancers allows selecting a representative DNA segment driving the reporter expression in a pattern that is similar or identical to that of the endogenous gene. This DNA segment can be further dissected to a minimal enhancer, which is the smallest contiguous DNA segment that recapitulates largely the spatial-temporal expression pattern of the endogenous gene (Arnone and Davidson 1997) (Figure 7).

Over many years, while studies on minimal enhancers have shed light on our understanding of the logic of enhancer function and evolution (Istrail and Davidson 2005), they also showed some limits to our understanding of enhancers.

How minimal enhancers are identified is dependent on the assumption that enhancers are collections of compact, modular units (Shlyueva, Stampfel, and Stark 2014). As mentioned above, minimal enhancers have been identified qualitatively, with a focus on the relative spatial distribution of enhancer activity rather than on its quantitative levels. Moreover, minimal enhancers are identified by arbitrarily chosen fragments, only assessing the sequence sufficiency rather than necessity for the enhancer activity (Muller and Basler 2000; Milewski et al. 2004) (Figure 7).

However, it is widely recognized that minimal enhancers often fail to recapitulate the precise expression pattern boundaries of the native gene, and that most minimal enhancers drive significantly lower levels of gene expression compared to that of the native gene (Summerbell et al. 2000; Irvine et al. 2008; Barrière, Gordon, and Ruvinsky 2011; S. Small, Blair, and Levine 1992; Chao, Wang, and Yuh 2010; Frankel et al. 2011). Many minimal enhancers have been shown to drive expression in ectopic cell types or at ectopic developmental stages, and flanking sequences are necessary for correcting the enhancer activity (Chao, Wang, and Yuh 2010).

There is no reason to take for granted that enhancers should be short, contiguous segments with clear physical limits (Milewski et al. 2004). Different enhancers can have diverse architectures,

that some might contain complex spatial distributions of regulatory information. Actually, some enhancers have been found to spread over a larger region of several kilobases, making the boundaries of enhancers difficult to justify (Klingler et al. 1996; Stephen Small, Blair, and Levine 1996; Davis et al. 2007). For example, the regulation of the gene *runt* in *Drosophila* embryo is controlled by TFBSs dispersed over a 5 kb segment region rather than clustered into a compact element (Klingler et al. 1996). A comprehensive assay of the *Hox* gene *Ultrabithorax* (*Ubx*) locus failed to identify the *cis*-regulatory regions of *Ubx* that drive expression on the posterior second femur in *Drosophila*, suggesting this enhancer structure is complex and that the TFBSs might be scattered across a larger region outside of the region surveyed (Davis et al. 2007). Such distributed enhancers have been implicated in fine-tuning gene expression and maintaining robustness of gene expression against genetic and environmental perturbation (Michael Z. Ludwig et al. 2011; Frankel et al. 2010; Swami 2010), as well as conferring the precision of gene expression (Dunipace, Ozdemir, and Stathopoulos 2011). In addition, evolutionary conserved TFBSs have also been found outside the annotated minimal enhancers (Hare, Peterson, and Eisen 2008; M Z Ludwig, Patel, and Kreitman 1998).

4.4. Quantifying the enhancer activities in the tissue

Although the minimal enhancers are clearly insufficient to understand enhancer function and evolution, very few studies have comprehensively investigated the regulatory inputs necessary and sufficient for the full activity of an enhancer.

In this thesis, we applied a more systematic and high-resolution quantitative approach to directly measure the spatial expression pattern in a tissue. By systematically dissecting a larger DNA region, we are able to map regulatory sequences both necessary and sufficient to produce the full enhancer activity. Our high-resolution quantitative method therefore makes it possible to attribute the measured reporter expression differences to any single sequence change, which can be a powerful tool to deepen our understanding of the grammar of enhancer function.

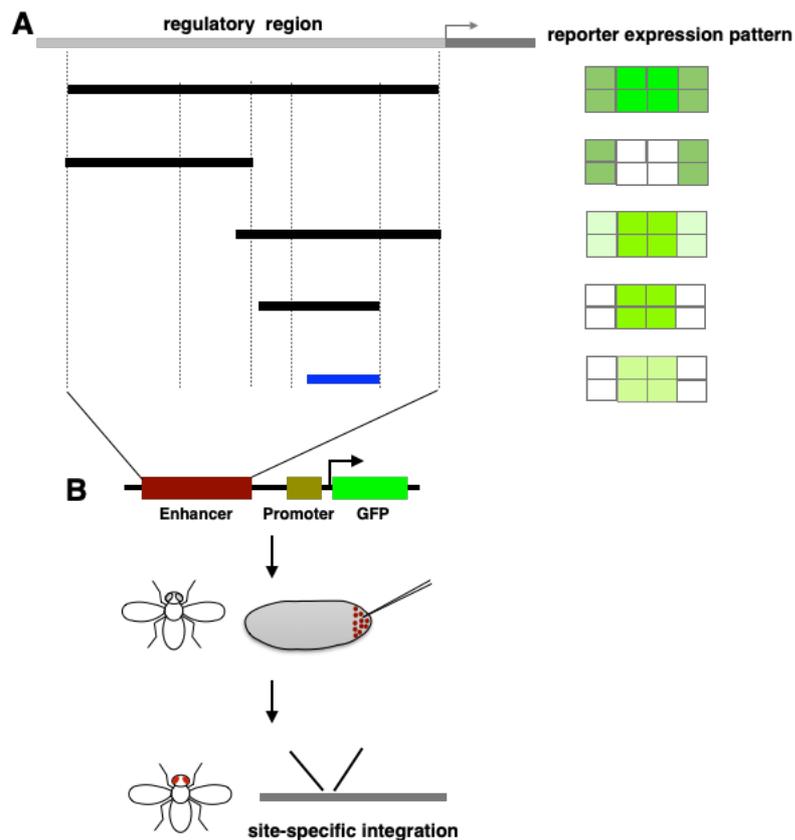


Figure 7: The identification of minimal enhancers using transgenesis in *Drosophila*. A. The arbitrary dissection of regulatory region to identify minimal enhancers (left) and the reporter gene expression patterns driven by the corresponding regulatory fragments (right). The blue fragment indicates the identified minimal enhancer. The intensity of green color indicates the intensity of reporter gene expression. B. The classical reporter assay in *Drosophila* for enhancer validation.

5. *Drosophila* pigmentation model

5.1. *Drosophila* pigmentation synthesis

Drosophila pigmentation is an ideal model to study the molecular mechanisms that underlie the evolution of morphological diversities. Pigmentation is one of the most rapidly evolving and variable traits within and among species in *Drosophila* (Prud'homme, Gompel, and Carroll 2007; Wittkopp et al. 2009). The diverse pigmentation patterns decorating fruit fly bodies as well as other animal bodies, have played an important role in organismal fitness such as mating choice, thermoregulation, ultraviolet light resistance and mimicry (COOK 1998). In addition, the availability of multiple genetic tools in *Drosophila* has offered an opportunity to elucidate the genetic and developmental mechanisms of pigmentation patterning and evolution (Hales et al. 2015; Wittkopp, Carroll, and Kopp 2003)

Understanding the process of pigmentation formation in *Drosophila*, for example, what genes are required, how do these genes interact to paint different pigmentation patterns, has laid a foundation to investigate the regulation underlying the divergence of pigmentation patterns.

In *Drosophila*, body color results from a combination of black, brown and yellowish pigments which are synthesized through a complex biochemical pathway (Figure 8) (Wittkopp, Carroll, and Kopp 2003; True 2003; Wright 1987). Tyrosine from the diet is first converted into DOPA (L-3,4-dihydroxyphenylalanine) by tyrosine hydroxylase (encoded by *pale*). DOPA is then converted into dopamine by the Dopa decarboxylase enzyme (encoded by *Ddc*). Dopamine can then be converted into different pigments. It can be converted into black pigments by Yellow (encoded by *yellow* or *y*); It can also be converted into brown pigments by phenol oxidases (PO), or be converted into *N*- β -alanyl dopamine (NBAD) through the activity of Ebony followed by being polymerized into yellow-tan pigments by PO. The conversion of dopamine into NBAD is reversible, and this reverse reaction is catalyzed by Tan (encoded by *tan* or *t*); it can also be converted into *N*-acetyl dopamine (NADA) by arylalkylamine *N*-acetyl transferases (aaNATs) and then lead to colorless pigments by PO (Stern, Road, and Cb 1998; Wittkopp, Carroll, and Kopp 2003; Massey and Wittkopp 2016).

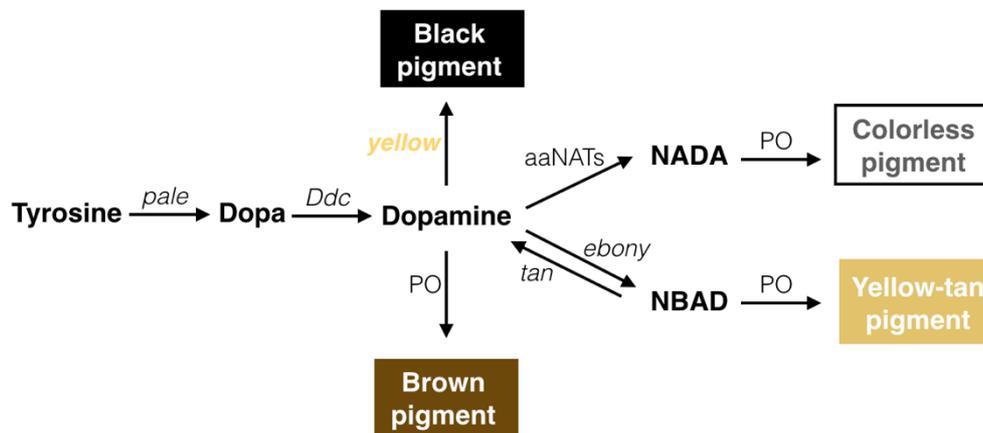


Figure 8: A simplified scheme of pigment biosynthesis pathway in *Drosophila*. The gene *yellow* is highlight in yellow color. Modified with permission from (Massey and Wittkopp 2016).

5.2. *Drosophila* pigmentation regulation

In the pigmentation synthesis pathway, one of the most studied gene is *yellow*. The gene *yellow* is required for the production of black pigment as *yellow* mutants display a light yellowish body color (Biessmann 1985; Morgan and Bridges 1916). In *Drosophila*, *Yellow* expression during pupal life prefigures the pigmentation patterns in the adult wing and the abdomen (Prud'homme et al. 2006; Arnoult et al. 2013; Camino et al. 2015; Jeong, Rokas, and Carroll 2006; Wittkopp, Vaccaro, and Carroll 2002) (Figure 9A). The different expression patterns of *yellow* on these locations are controlled by distinct modular enhancers.

In *D. melanogaster*, the male abdomen is darkened in the A5 and A6 tergites, which is an evolved trait from an ancestor that lacked this trait (Jeong, Rokas, and Carroll 2006). *Yellow* is expressed in a pattern that prefigures the abdominal pigmentation pattern, and this specific pattern is controlled within a body enhancer (Wittkopp, Vaccaro, and Carroll 2002) that contains binding sites for the Abdominal-B (Jeong, Rokas, and Carroll 2006) and Bric-a-brac (Roeske et al. 2018) transcription factors.

In the wing, Yellow is controlled by the *wing* enhancer (Figure 10). In *D. melanogaster*, Yellow is expressed at a low level throughout the wing during late pupal development, resulting in a uniform wing pigmentation. By contrast, in *D. biarmipes*, where the males have evolved a dark spot at the wing tip from an unspotted ancestor (Prud'homme et al. 2006), Yellow is also expressed at a much higher level in the spot region (Figure 9B). The novel spot-specific expression pattern was shown to be regulated by a *spot* enhancer, which evolved in the vicinity of the preexisting *wing* enhancer (Gompel et al. 2005). Further dissection of the *wing* enhancer identified a sub-fragment, the *right* element, driving *yellow* expression exclusively in the wing blade (Figure 9C). Analyses of the enhancer sequence indicate that spatial *spot* activity in *D. biarmipes* has evolved TFBSs for the transcriptional activator Distal-less (Dll) (Arnoult et al. 2013) and the repressor Engrailed (En) (Gompel et al. 2005).

The evolved *spot* element thus represents a good model to investigate the emergence of a morphological novelty. These studies, however, were unable to resolve the evolutionary origin of the *spot* element: it is possible that the *spot* element has evolved through co-option of the preexisting *wing* enhancer although the two enhancers appear to not overlap (Prud'homme et al. 2006). Alternatively, it could have evolved *de novo* in a region next to the *wing* enhancer.

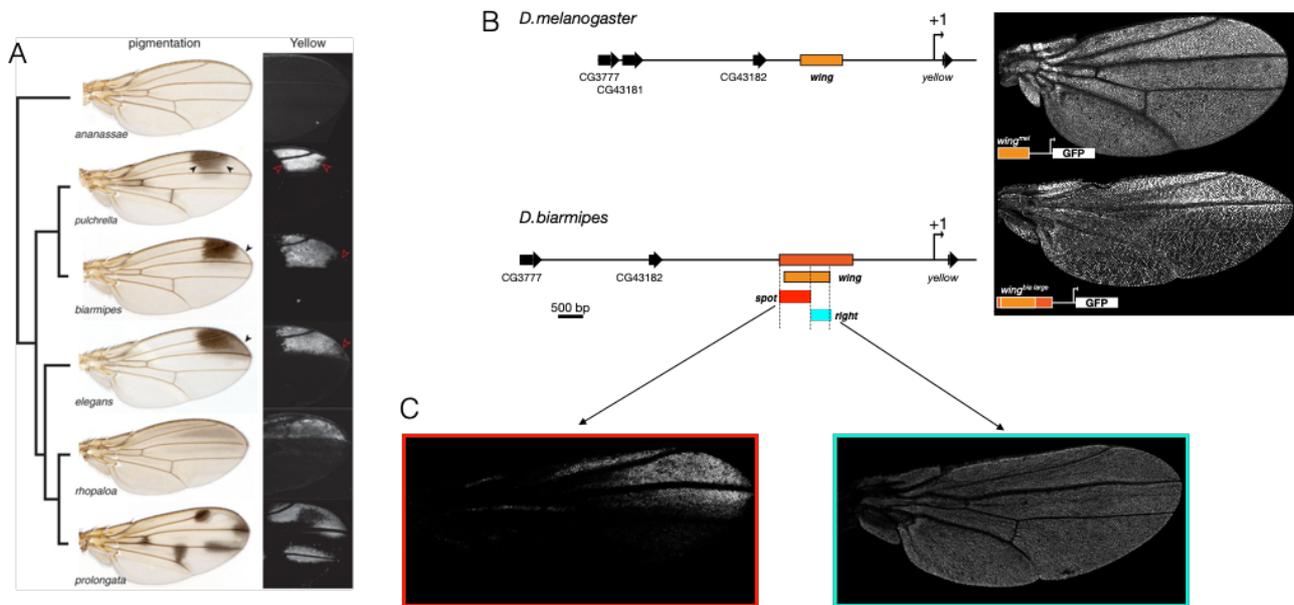


Figure 9: The novel *spot* expression pattern in *D. biarmipes* is controlled by the new *spot* enhancer of *yellow*. A. The wing spot is a morphological novelty derived from an unspotted ancestor. Expression of *yellow* prefigures the diverse spot pigmentation patterns on the wing of different *Drosophila* species. B. The *wing* enhancer from *D. melanogaster* drives a low level of activity across the wing, while the *wing* enhancer from *D. biarmipes* also drives a high level of activity in the spot region. C. Two separable elements, the *spot* and the *right* enhancer drives a spatial reporter expression pattern specifically in the spot region and the wing blade region, respectively. Panel A is adapted from (Arnoult et al. 2013) with permission. Figure B is kindly provided by Prof. Nicolas Gompel.

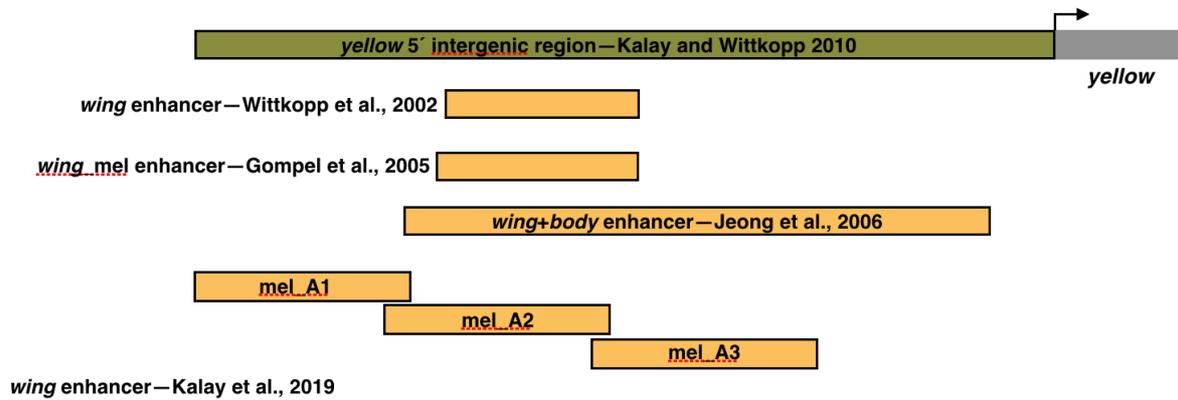


Figure 10: Summary of prior work for the mapping of the *wing* enhancer of *yellow* in *D. melanogaster*. Each of the fragment *mel_A1*, *mel_A2* and *mel_A3* drives reporter expression in the wing blade. Modified with permission from (Kalay et al. 2019).

6. A pipeline for the quantification of gene expression in *Drosophila* wing

The work described in this thesis relies on a technical pipeline developed in our lab to precisely quantify the enhancer activities in the wing. The workflow is depicted in Figure 11. The idea of this pipeline is to assess phenotypes (gene expression), not from a single wing for each given line, but from dozens of wings in order to take non-genetic variation into account in our analyses.

The pipeline comprises different steps, each optimized to reduce experimental noise. For instance, animal staging is very precise (in a 15 min time window after emergence from the pupa). The protocol for sample dissection and fixation is highly stereotyped, and so are the imaging conditions.

An important aspect of this pipeline is the alignment of wing images. Although the architecture of *Drosophila* wings is very stable, wings from two different animals do not perfectly overlap, even after scaling. This represents a limit for the quantitative comparison of spatial gene expression among individuals. To overcome this limit, the lab uses a methodology known as image registration or image alignment, where individual wing images are deformed (warped) to match a reference image. The landmarks used to align images are the wing veins. The resulting images are wings with the same shape and size, irrespective of the gene expression pattern being considered. The deformation is computed on bright-field images, and then applied to the fluorescent channel containing the gene expression information.

Wing alignment opens the possibility to perform quantitative analyses, as all the wing in a dataset share the same system of spatial coordinates, and this system can be used to describe quantitative gene expression in two dimensions.

Multiple quantitative information become available from a dataset of registered wings which are imaged under the same conditions. These include:

the average phenotype. This is a wing image representing the average value of each pixel among images of the same genotype (the same line). These images are smoothed and colored with a heat map to represent the signal intensities.

the overall variation in a dataset. This is visualized in a Principal Component Analysis (PCA). In our experiments, the first two components always captured most of the variation and corresponded to overall changes in intensity and changes in pattern.

additional tools to measure changes between lines. The difference between groups (*Diff* image) is an image representing the subtraction of two average phenotypes. Similarly, the local fold change between two genotypes is represented by the *logRatio* image. Colormaps for *Diff* and *logRatio* images represent the absolute value of differences between the compared phenotypes. Grey color means no change between the compared phenotypes.

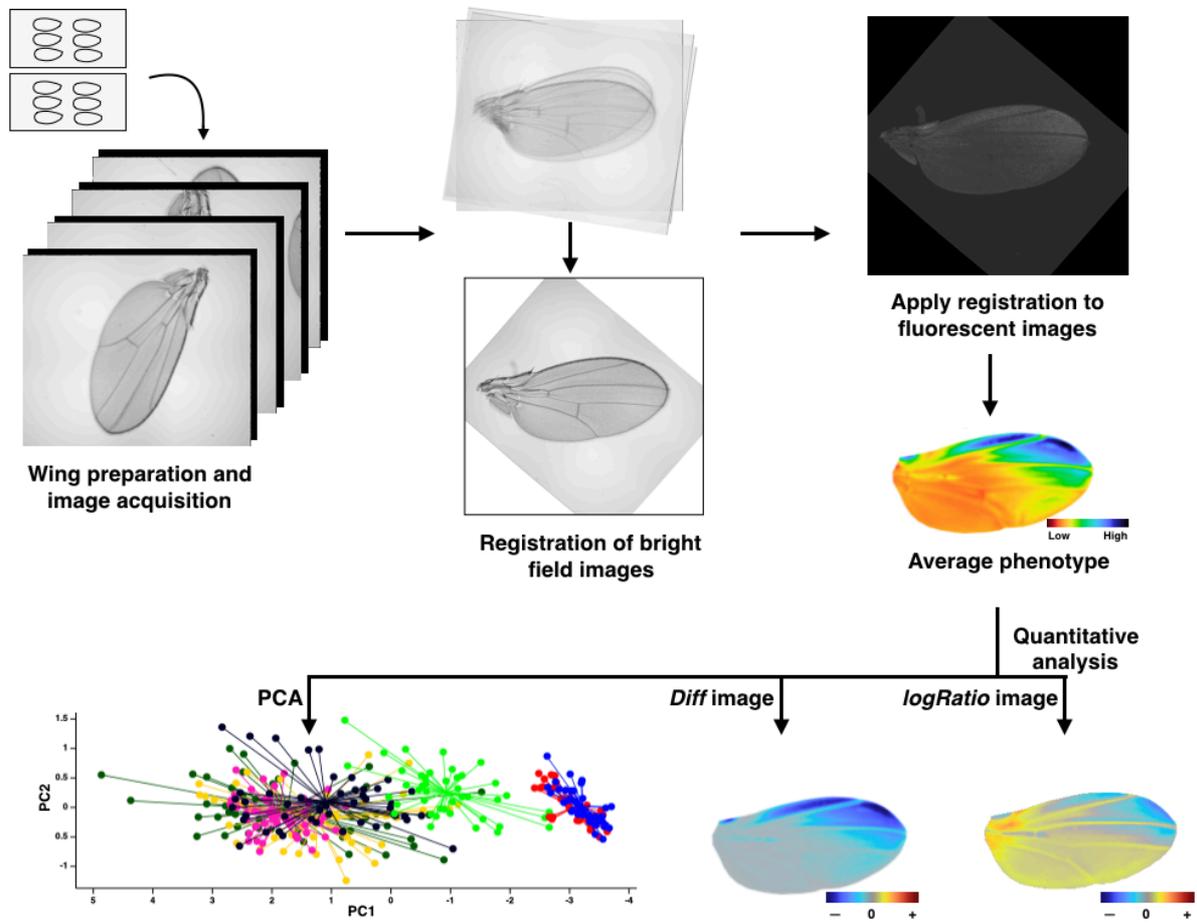


Figure 11: Pipeline for wing image quantification.

Aim of the thesis

In this thesis, I aim to investigate the evolutionary path and the regulatory logic underlying the emergence of a novel enhancer.

Firstly, I asked how a novel enhancer has evolved from a preexisting regulatory context. While enhancer co-option has provided evidence to support this phenomenon, its genetic and molecular mechanisms still remain elusive. This breaks down into several questions: What is the positional and functional relationship of the novel and the preexisting activities? Furthermore, do the novel and the preexisting activities share co-opted TFBSs? How have the shared sites, if any, contributed to both enhancer activities?

As a model system to tackle these questions, I used the regulation of the pigmentation gene *yellow* in *Drosophila*. I focused in particular on the *spot* enhancer, a derived activity found in *D. biarmipes* in the vicinity of the preexisting *wing blade* enhancer. This *spot*_enhancer underlies the evolutionary origin of a novel aspect of *yellow* expression in fly wings, and a novel pigmentation pattern. With the collaboration of Dr. Yann Le Poul, we used a quantitative and systematic approach to precisely map the sequence boundaries of the *spot* and *wing blade* enhancer activities. To assess their functional relationship, we dissected sequences upstream of *yellow* in a reporter assay.

Secondly, I investigated how an enhancer is built to generate a spatial expression pattern, trying to decipher the underlying regulatory logic. Using a similar experimental set-up, again in collaboration with Dr. Yann Le Poul, I undertook the fine dissection of the core of the *spot* enhancer, *spot*¹⁹⁶. We introduced systematic mutations along the enhancer sequence and precisely quantified the spatial pattern from all the enhancer variants.

Results

Paper One: 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

PNAS August 25, 2020 117 (34) 20636-20644.

In this paper, we investigated the evolutionary origin of the novel *spot* enhancer of *yellow* in *D. biarmipes*, which underlies the evolutionary origin of a novel wing pigmentation pattern. We used precise quantitative analysis and a systematic dissection of *yellow* regulatory regions to revisit the positional and functional relationship of the novel *spot* and the preexisting *wing blade* activities. We found that the novel *spot* activity has evolved by co-opting most of the sequences from the preexisting *wing blade* activity. We further demonstrate that a pleiotropic site within the overlapping region is required for the local chromatin accessibility, suggesting that chromatin accessibility might be a component seeding evolutionary co-option.



Enhancer evolutionary co-option through shared chromatin accessibility input

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}

^aFakultät 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 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. 1A). 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,

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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

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

²To 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.

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. 1B) 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. 1B). 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 an 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).

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. 1C). We summarized the variation in activity across the

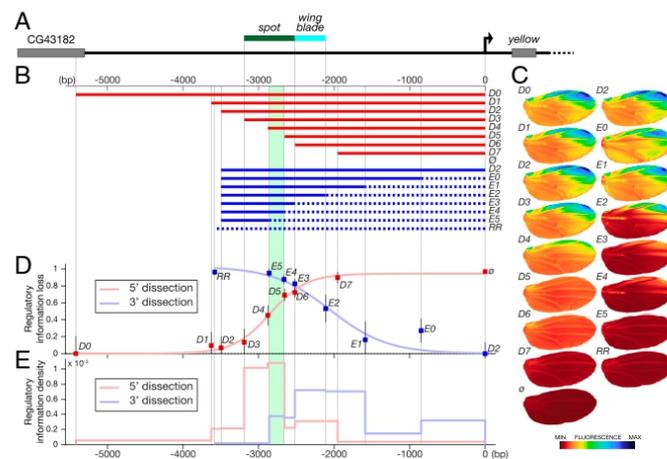


Fig. 1. Quantitative mapping of wing regulatory activities at the *yellow* locus. (A) The top line represents the 5' region of the *yellow* locus from *D. biarmipes*; the green and blue bars indicate the respective locations of *spot* and *wing blade*, respectively, as originally mapped (6). (B) Two 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); \emptyset 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 $\frac{d(P_x, P_{ref})}{d(P_x, P_{ref})}$, where P_x , P_{ref} , and P_\emptyset 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 \emptyset 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 $d(P_x, P_{ref})$. (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.

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. S1A). 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. 1E 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. 1D 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. 1D 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. 2B 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. 2B and C and SI Appendix, Fig. S1A). 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. 2A) 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. 2A, and SI Appendix, Fig. S1A). 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. 2B 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*⁹⁶ 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. 2B and C and SI Appendix, Fig. S1C 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: $F(1, 55) = 22.57$, $P = 1.4954e-05$] (horizontal double arrow in Fig. 2A and 3' curve in Fig. 2B). Reciprocally, we asked whether 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: $F(1, 68) = 56.35$, $P = 1.7205e-10$] (vertical double arrow in Fig. 2A and 5'

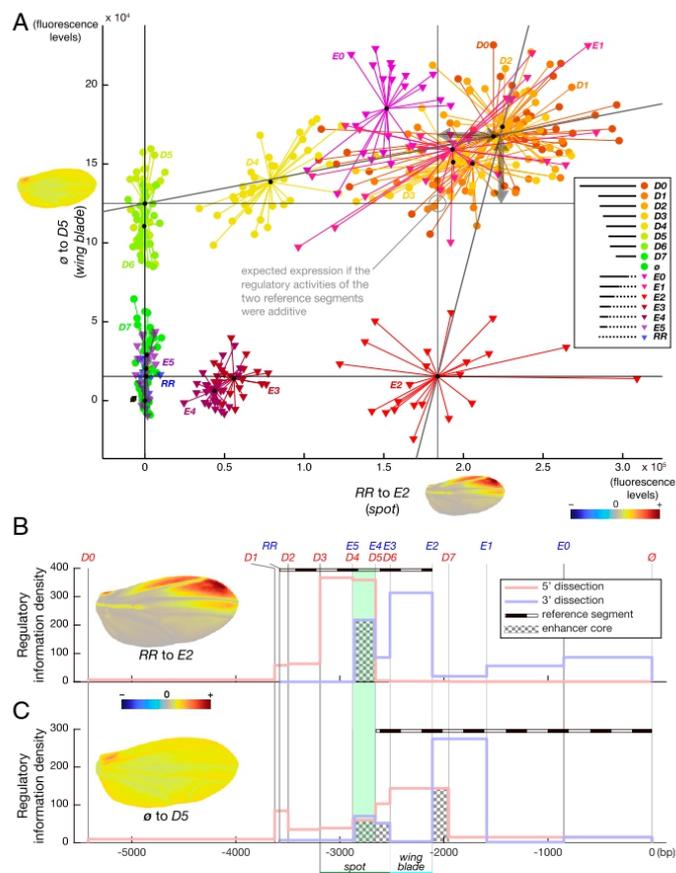


Fig. 2. *wing blade* and the *spot* activities map to overlapping sequences. (A) Representation of *wing blade* activity as a function of *spot* activity. Independent estimates were produced by projecting the PCA phenotypic space (PCA in *SI Appendix*, Fig. S1A) on a two-vector basis defined by two independent directions identified in *SI Appendix*, Fig. S1A (phenotypic directions with color map near each axis) and corresponding to *wing blade* (constructs ϕ 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.

curve in Fig. 2C). 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. 2A would form, conservatively, a perfect rectangle (indicated by four lines in the graph). Additivity would translate geometrically into the addition of the two vectors ϕ 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. 1B) 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. 1B and C and 2B, intersection between the 5' and 3' dissection curves). In addition, it contributes quantitative information both to the *spot* and the *wing 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

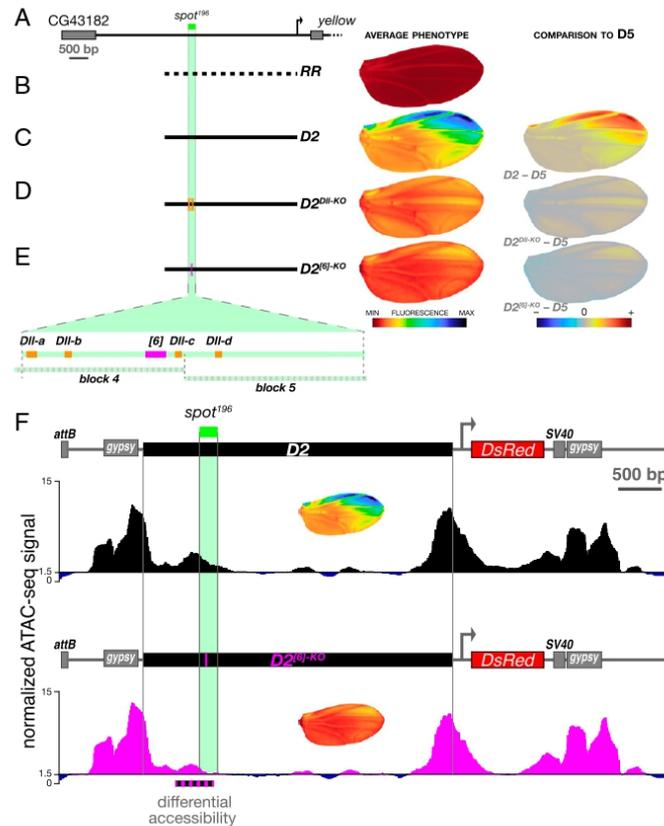


Fig. 3. Shared regulatory inputs of the *wing blade* and the *spot* activities. (A) A map of the *yellow* locus 5' region highlighting the position of the *spot*¹⁹⁶ core. (B–E) The *wing blade* and the *spot* activities are strongly affected by discrete mutations in *D2*. (Left) Construct schematics. (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. 1B and C. (C) *D2* is identical to Fig. 1B and C. (D) Mutating all four characterized Dll binding sites (16) of *spot*¹⁹⁶ in the context of *D2* (*D2*^{DII-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*^{G1-KO}, 12 bp mutated) in the context of *D2* (*D2*^{G1-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*^{G1-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 (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.

core element (SI Appendix, Figs. S1B and S2, $D2^{block5}$) reduces the *spot* activity by 61% [ANOVA $D2$ vs. $D2^{block5}$: $F(1, 44) = 516.84$, $P = 5.9730e-26$] without affecting the average levels of *wing blade* activity [ANOVA $D2$ vs. $D2^{block5}$: $F(1, 44) = 0.58$, $P = 0.452$]. By contrast, the randomization of the distal half of this core element (SI Appendix, Figs. S1B and S2, $D2^{block4}$) abolishes the *spot* activity completely and suppresses the nonadditive effects on *wing blade* activity described above [ANOVA $D2^{block4}$ vs. $D5$, $F(1, 45) = 0.025$, $P = 0.876$] (SI Appendix, Fig. S2D). 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). $D2^{Dll-KO}$ and $D2^{[6]-KO}$ resulted in strong effects on the *spot* (Fig. 3A–E and SI Appendix, Fig. S1B), and both abolished the nonadditive *wing blade* activity, bringing it to the levels of $D5$ (SI Appendix, Fig. S1B). 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. S1B). As a comparison, $D2^{[6]-KO}$ has a stronger effect on *wing blade* than $D5$, from which the whole *spot*¹⁹⁶ segment was removed (Fig. 3E and SI Appendix, Fig. S1B). 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 $D2^{[6]-KO}$ 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. 3F 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 $D2^{[6]-KO}$ (Fig. 3F). 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. 4A). 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. 4B). 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 Arnould et al. (16). We used ϕ C31-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 ϕ C31-mediated transgenesis (22) cloned at the *AvrII* site] digested with BamHI and EcoRI using In-Fusion HD Cloning Kits (Takara; catalog no. 121416). The fragment encompassing the four *Dll* sites in construct $D2^{Dll-KO}$ was synthesized in vitro by Integrated DNA Technologies. The mutations in construct $D2^{[6]-KO}$ were introduced by PCR through site-directed mutagenesis.

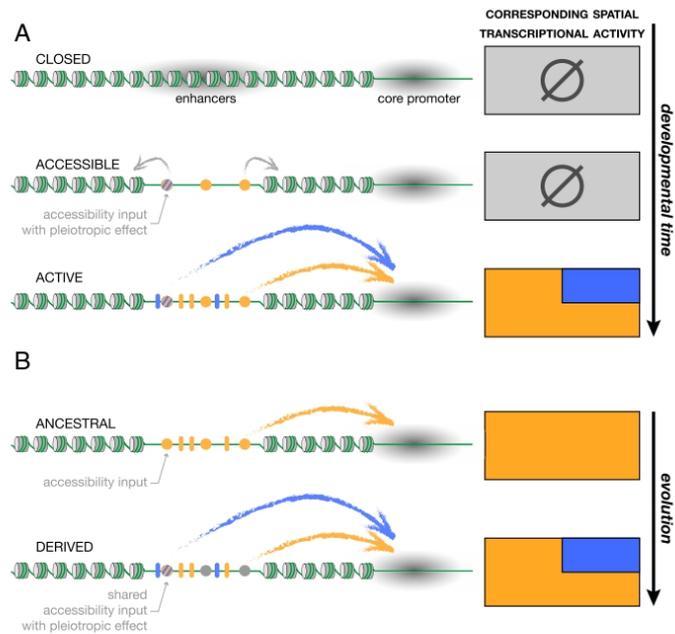


Fig. 4. Developmental enhancer pleiotropy and evolutionary enhancer co-option. (A) The developmental progression toward the activation of two interdependent enhancers inferred from our results. High nucleosome occupancy prevents access of transcription factors to the enhancers sequence (closed state; 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)], poised 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.

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. 1A) 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 synthesized 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 $-20\text{ }^{\circ}\text{C}$ until we had reached a sufficient number of flies. Staged flies were processed after a maximum of 48 h at $-20\text{ }^{\circ}\text{C}$. We dissected a single wing per male. Upon dissection, wings were immediately mounted

onto a microscope slide coated with transparent glue (see below) and fixed for 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 $4\text{ }^{\circ}\text{C}$ 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 $4\text{ }^{\circ}\text{C}$ until image acquisition.

The glue-coated slides were prepared immediately before wing mounting by dissolving adhesive tape (Tesa brand; tesafilm, reference S7912) 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 \times 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 = $4\text{ }\mu\text{m}$) 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 semiautomated 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 (\emptyset) was used as a proxy to measure noise and tissue autofluorescence. The median raw fluorescent image was computed across all \emptyset images and used to remove autofluorescence, subtracted from all raw images before the following steps. All variation of fluorescence below the median \emptyset value was discarded. The DsRed (red fluorescent protein from *Drosophila*) 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 subsequent 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 ($n_{\text{individual}} \times n_{\text{pixels}}$ 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 (\emptyset) 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 [\emptyset]). A sigmoid curve of equation $\frac{1}{1 + e^{-\alpha(t - t_{50\%})}}$, 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 \emptyset 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: 1x nuclei permeabilize buffer (NPB) buffer: 15 mM Tris-HCl, pH 7.5, 3 mM MgCl₂, 1x protease inhibitor mixture (Roche; cOmplete catalog no. 04693132001), ultrapure water (Invitrogen); 1x 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 1x 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 μ L cold 1x 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 μ L 1x 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- μ m strainer (Corning) and centrifuged at 4 °C at 1,000 \times g for 10 min.

Tagmentation. Pelleted nuclei were gently resuspended in 45 μ L ultrapure water and counted using a hemocytometer; 50,000 nuclei were then centrifuged at 4 °C at 1,000 \times g for 10 min and resuspended in 8 μ L 2x Tagment DNA (TD) buffer (Illumina; catalog no. 15027866). The tagmentation reaction followed the previous ATAC-seq protocol (29) with minor modifications: 10 μ L 2x TD buffer with nuclei, 2 μ L TD Enzyme (Illumina; catalog no. 15027865), 8 μ L ultrapure water. The reaction was terminated by the addition of 5x 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 2x PCR Master Mix (NEB; catalog no. M05415) for 9 to 11 PCR cycles and purified by Agencourt AMPure XP beads (Beckman Coulter) with double size selection (0.5x and 2.0x). Bio-analyzer 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, Fix-Mate 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 -5PMR; -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 bFullLibrarySize = 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. [GSM422134](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM422134)–[GSM422139](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM422139)).

ACKNOWLEDGMENTS. We thank Benjamin Prud'homme, Ilona Grunwald Kadow, Miltos Tsiantis, and Marta Božek 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.

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. Kwon *et al.*, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
11. A. Visei, 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. Arnould *et al.*, Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* **339**, 1423–1426 (2013).
17. P. K. Geyer, V. G. Corces, Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in *Drosophila melanogaster*. *Genes Dev.* **1**, 996–1004 (1987).
18. M. F. Walter *et al.*, Temporal and spatial expression of the yellow gene in correlation with cuticle formation and 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. Lachowicz, 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 ϕ C31. *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. Buenostro, 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. Božek, 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. Dunder, 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).



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

Xin, Le Poul, et al. Figure S1

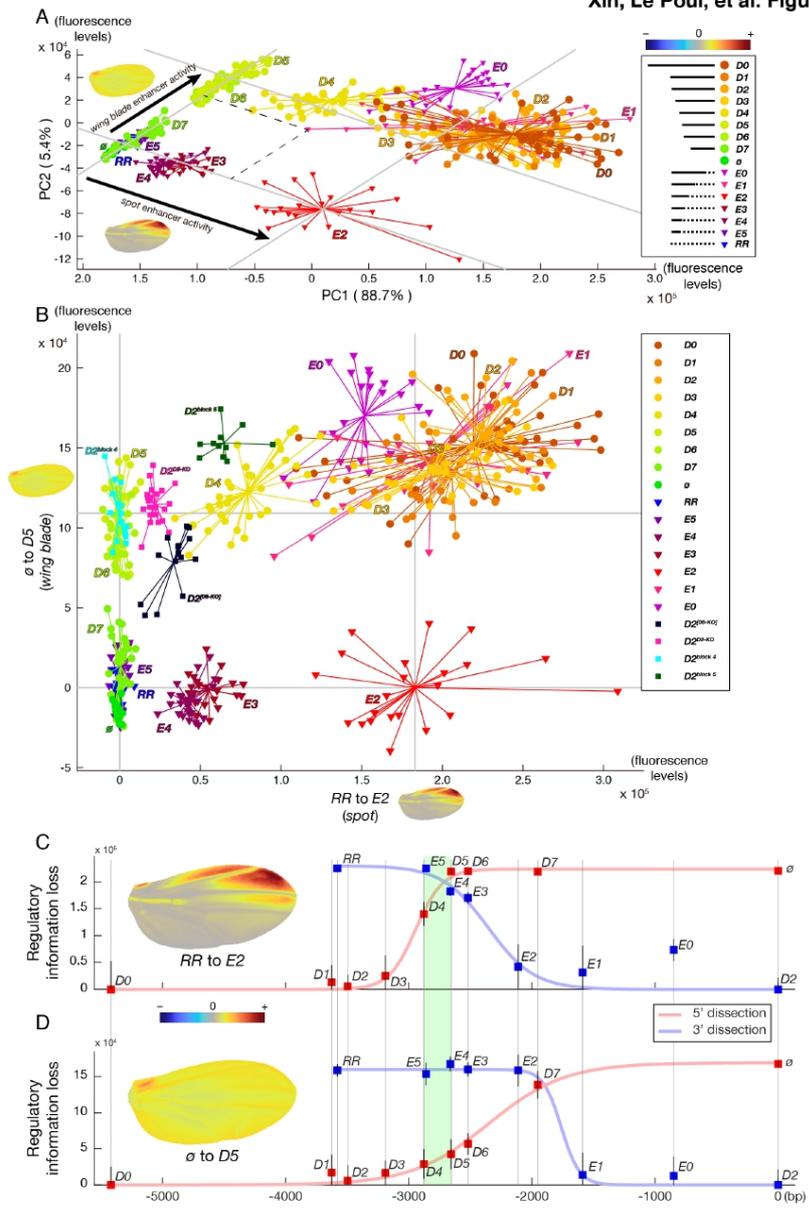


Fig. S1. Variation in reporter expression across all transgenic lines. (A) PCA of activity variation in the phenotypic space that correspond to the *wing blade* and the *spot* activities, respectively. Wings with colormap (average phenotype differences between *D5* and \emptyset , and *E2* and *RR*, 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: $D2^{block\ 4}$, $D2^{block\ 5}$, $D2^{DII-KO}$ and $D2^{[6]-KO}$. (C, D) Loss of regulatory information along *yellow* 5' region (fluorescence levels, as in Fig. 1D) for each direction defined in panel (A).

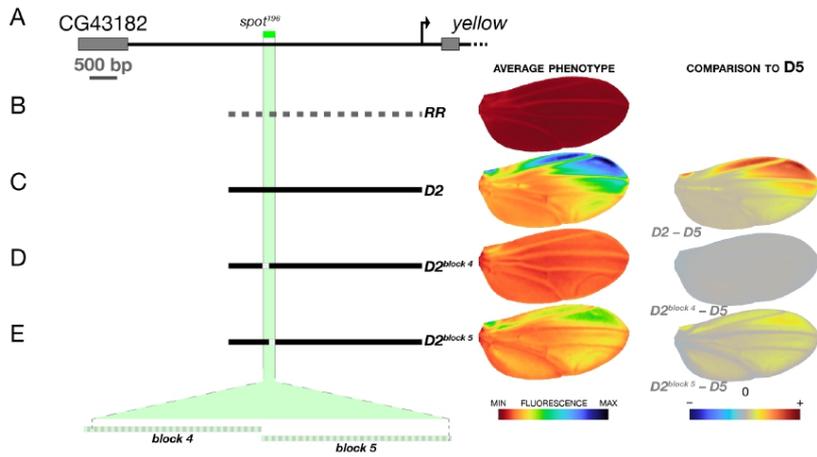


Fig. S2. Reporter activity in the wing for constructs *D2*^{block 4} and *D2*^{block 5}. (A) A map of the *yellow* 5' region highlighting the position of the *spot*¹⁹⁶ core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 4) and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

Xin, Le Poul, et al. Figure S3

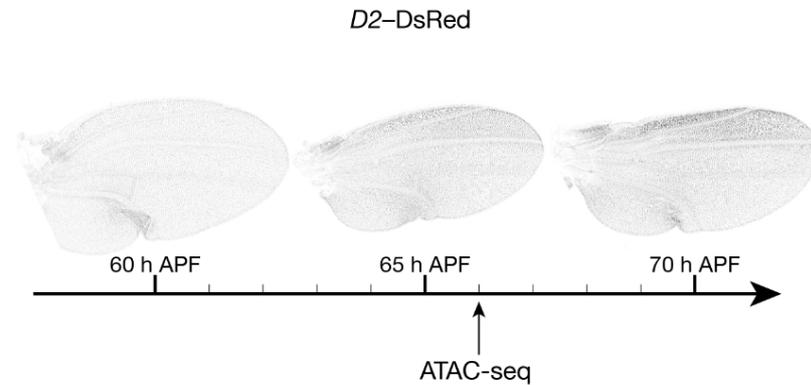


Fig. S3. Reporter activity in the wing for constructs $D2^{block4}$ and $D2^{block5}$. (A) A map of the *yellow 5'* region highlighting the position of the *spot¹⁹⁶* core. (B-E) The *wing blade* and the *spot* activities are strongly affected by sequence randomization of the distal part (block 4) and the proximal part (block 5) of the *spot¹⁹⁶* core in the context of *D2*. The portions of randomized sequence are depicted on a blown-up schematics of the *spot¹⁹⁶* 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) $D2^{block4}$ abolishes the *spot* activity and strongly reduces the *wing blade* activity. (E) $D2^{block5}$ reduces the *spot* activity and has a milder effect on the *wing blade* activity. (F) The differential effects of $D2^{block4}$, $D2^{block5}$, $D2^{DII-KO}$ and $D2^{Gj-KO}$ 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.

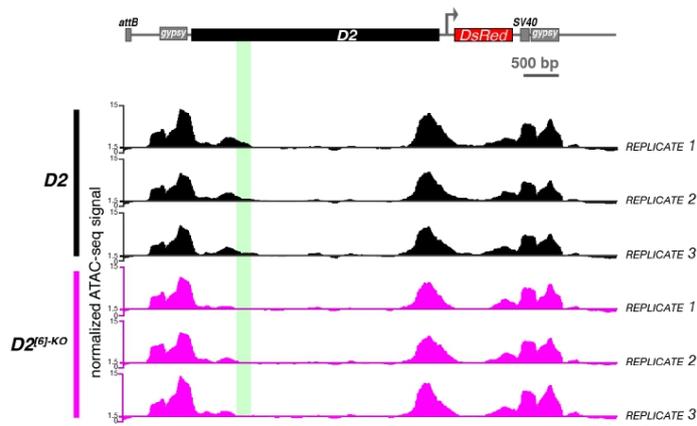


Fig. S4. ATAC-seq replicates shown separately. Chromatin accessibility at the *D2* and *D2^{[6]-KO}* transgenes at the onset of *spot* activity. The 6 ATAC-seq traces represent 3 technical replicates for each transgene that were pooled for each genotype in Fig. 3.

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, e)	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, e)	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
(e, RR)	10	0.829743	12.1836	10	25	2.57409e-07

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

Primer name	Primer sequence	Note
D0-Forward	gagcccgggcgaattaaacagcccaagctcggctgc	small letters denote adapters for in-Fusion cloning
D1-Forward	gagcccgggcgaattgccacgccagtaattctaaaagctgc	Capital letters denote D. biarmipes genomic sequence
D2-Forward	gagcccgggcgaattccttagttcagaagggcctgc	
D3-Forward	gagcccgggcgaattcgataatgcggattaccgc	
D4-Forward	gagcccgggcgaattccagggcctgtattggcattc	
D5-Forward	gagcccgggcgaattgtgccaatcatttttagtacacc	
D6-Forward	gagcccgggcgaattgaaatctctttctcggcctgt	
D7-Forward	gagcccgggcgaattgccgcaagctgcgcaaa	
D-series-Reverse	gtctctgagggatcttctgtggaccgtggcc	
E-series-Forward	gagcccgggcgaattccttagttcagaagggcctgc	
E0-Reverse	gtctctgagggatctgcacaatacaggcgctggc	
E1-Reverse	gtctctgagggatctctaccacagaaatrttag	
E2-Reverse	gtctctgagggatctcattatggccacagattag	
E3-Reverse	gtctctgagggatctctcggcaggcgccact	
E4-Reverse	gtctctgagggatcttaccatggccacatc	
E5-Reverse	gtctctgagggatctgcacaatacaggcgctggc	
Randomized Space F-Forward	gagcccgggcgaattacaaaacggcgcaaatcct	
Randomized Space F-Reverse	gtctctgagggatctgtcggcctgtctaac	
D2m10		
fragment 1-F (D2-Forward)	gcttagagccggggcgaattccttagttcagaagggcctgc	
fragment 1-R	ccttaaacgggaatcgagatgccaatcagaagggcctgc	
a fragment encompassing the 4 DII site mutations was synthesized from IDT-D2m10	TCTGcGATTCGGTTAAAGGACGccgTCTGAGCTAAACTCGGTTATG GAGAGATCTAAATTTCCCGCTTTTGGCTGAAATBAAGcCTTGAATTCC CCCGTGGCTGcTcgaAAGcAACAAGGGCGcCTGTGCTCTTTCAATGTA AATTGCAAAATGCTCAATCCGCC	
fragment 3-F	TGCAAAATGCTCAATCCGCCCT	
fragment 3-R (D2-Reverse)	gtctctgagggatcttctgtggaccgtggcc	
D2m11		
fragment 1-F (D2-Forward)	gagcccgggcgaattccttagttcagaagggcctgc	
fragment 1-R with mutation adapter	TTATTTAGTgtsaagatGGGGAAATTTAGATCTTCAT	
fragment 2-F with mutation adapter	ATCTTgctctcAATAAATTAATGAATCCCGCTCG	
fragment 2-R (D2-Reverse)	gtctctgagggatcttctgtggaccgtggcc	
D2m12		
fragment 1-F	gcttagagccggggcgaattccttagttcagaagggcctgc	
fragment 1-R with mutation adapter	TAAcGACACCGAAGGAAAGCGCTACTGGGAAAACTAAGC	
fragment 2-F with mutation adapter	CCTTCCCTGGTGCTGTGTA	
fragment 2-R	TTTAATAGCCAGCGGGAAATTCAGATGATAACGAAGTGGAG	
fragment 3-F	ATTCGCCGCTGCATTAAACAAC	
fragment 3-R	gtctctgagggatcttctgtggaccgtggcc	
D2m13		
fragment 1-F (D2-Forward)	gcttagagccggggcgaattccttagttcagaagggcctgc	
fragment 1-R with mutation adapter	GATCGCAACATGACTCATGTTGATTAATTTCAAGCCAAAAG	
fragment 2-F with mutation adapter	CATGAGTCATGTCGSAAT	
fragment 2-R	GTAGTAAAAATGATTGGCACGTGTATCAATATTATTTCCAAGTGG	
fragment 3-F	GTCGCAATCATTTTGTACACCC	
fragment 3-R (D2-Reverse)	gtctctgagggatcttctgtggaccgtggcc	
98955 sequencing primer		
921-F	gcttggctcctcagctaat	
922-R	gtctctgaggtcctcagc	

Table S3. Construct sequences.

>RR
 ACAAACCGGGCGGACAAATCCTACAACGGTCACCTTGCAATCCCTTTACTCCCGGCAATACTGCACGCGTG
 ACTAATTTCTCATACTCGTTGCCAATAATCCTCCACACTCCGCACCCTCGATCTCAGTAAACCGGACATGGGG
 CCTATTTCAATAGTTCCGAGCGTGACCAGTTTTTTGTCAATCTTGAGCGAATTTAAGGGACGGCTACCGCA
 AGTACCGCGTCTTTATCAGTAGGGATTTCAAGTCGTTTGCCTACGCTATGTGGAGTAGTGGACCCCGGTGCTC
 TCAGTACGTGTAATAACCTGGTACGGTGATGAATTTGTTATTTACCTCAGTGGAAATACTCAGTGCAGTA
 AGAGCTGTCGTAGTGGTCTTCGGATGTGGTTCGCAGAGTTATGTCTGTGTACAGCGTAGATCGTTAATAAC
 GCGAACGGTGGCCACCCACAATAATGTGGGAGGCCACATACATGCTCACCATCGAACTCAGGATTGCTGT
 GGTAACGTTTAGGGCGACCGAACGGAGTGTAAATGGTTATCCACC CGGTGCCACATGATTTCAGTGCAGT
 GCGTACC CGAAGATATTCCTGGGAACCCAAGTAACTTATCGACATTTAATGGCGGTTCCCTTGGCAATGC
 CTGGTTTATAATAACGACATATATTTATGTGGGCTTTCTATACTCTGCGTAAACAAGTAAATCCCTTCCCT
 CGGTGCTGTTAAAGTGGAGAGGATCCTGGTTCCTGTGGCCAGCATTCCTTAGGATGTATAATAACGGCCG
 ACAGAGCTTCCAACTCCACTTCGTTATCATCTGACATGAGTCACTGTTGCGGATCTTCAAGGTAACCTCA
 TACACATCATTTCAATAATGACTTTGATGACCTCATCGCTTTTAGTCGCCACTTGGAAATAATTTGATAGC
 AGTGCATTATTTGAGCCTTATCCGAGAGCCTCATAGGTGGCGTTTATATCTTGAGTACATATAAGTGCAGG
 CGTAGGGTTCACCACTTAAGTGTGAAGCTAGGGGTTTCGCCGCCGGGACGAGGATGGGAGTCTCGGG
 GCCAGGTGCATTAAGCTCGGTAACCTCGGATGAACTCGGTACAGTGAAGTCCAGCTCAGGCACGGGAGTGC
 ATATGGGGTATGAAGGCTACAGAGACGCTCAATATGATTTCTTCCCAATCAAGTTTAGAGAAATAAAGAT
 CGATATTTGTGGGACCGGTGTATACAGGATCTAGAGGACCGAATAAATAATCGCTTTGTGCGTATGCCGA
 TCGGTACCGCGCAACAACCTACCTGAGAAAGCTTGTGGGACGCCTAAGTGAAGTACTAGCGGTCAAAAAGC
 CAGACGGGTGACCGCGGAAAAGGACCCGCTGTTAGAGCACTTATCTTCTGTCTGTGACATTTCAAGG
 GCTGGAATTCCTGTGGGACTCGGATCCCTCTATGCCCGCATACGCTCGTTGATGTATCCTGATTAAGCT
 AGTCTCTTTGAACCTGTGCAGTCCACGGGATAGCCGACCTTTCGGAGTTTGTGTCTTTCTTCCATA
 TGCTTCGTGTAATACATTTATTCACAACAATAAATAAGAGGGACCTGTCTAAAGAACACACATGGC
 AAAGTGGGAAATACAACCGAAAAGTGGTCCATAAACAAGAACCTGAATCACTCAGGAATGAGAACCATC
 GTGAGCCTTACGACAAATTTACCCATGGCATCTAAATGGCGAGTACTTTACAACGCCGACAAAGTAGCT
 TACGAATCATGTGACCGGAGTATCAATAATTTGTATGAGTCTCACCCAGATTTGATCCGCGTTAAGCT
 CACCCGTTAGGCAAACTCTTTGGCATCGATGGTAGTTAGCTCCATGTAAACAATTTCTACTAGAGGTAGC
 CCAGCGTGGCGCGCTTACCTATTTAGGGTTTGTATCGCCCTTTAGTAGAGTGGGGTCCGGCTTCCAGTAT
 CGAATAGATGATCTGCTCTTTGGATAGTGGCTGACAAAAGTACTAAGGGAAATCTTTATCTTATACTAAGT
 CCGAGGACAGTGGGTATAGACATGGGAAGCACTCACTCAGATATGTATAGACACAGCAAAATCGTGTATTTA
 AAATCAATTCCTTAGATTTATACGAAAAGTAAGAAACAGGGCAACACGACATTTGAAACAGTTACGTAAT
 GCGTTTTGCGGTATGCAAAAGTTTCAATTTGCTTAGGAGTGTGGCCATATGAGATTCATCCGCTCTTCTC
 GACGTGAGAGGGACCCAGCTTTGCTTCTTAAACAATAATCCTGTCGCTTAGGCAATTTTATGGTGTCTG
 GTTAAATTTGTTCAATACATCACATCCACGCTTATTCCTGTGCTCCGCTCCATTTTATGCTCGAATCCAGC
 GTAGGCGGTCATATGTGCTCTTTATTTAGCAGTTGGGAGGTTCTGTTCTCGGGAGATGTCGTTTGGACC
 ATTTGCCGATCTCGCGCCAGAGCGTGGAAATACCCGACATCTGGCTTTCAACTTGGCAAACTCAGGGA
 GTCGTCATCACTTGACCTAGAACC CGGTGGGGGACCAGTTATCTGCTTAGCCATTTCCGGTTATGGCG
 GAGCATATTAACAACCGGTGAAAATTTCTGGCCAGCTTACAAAAGTGTCAAAATGGGAGCCCGAACCCAC
 CGAGGCATACGATGCTGGCTAGCTCGTACCATTTATGAGATGAACATACTGAACCTCACTACTGGGTAA
 TACGAATAAGTACATATACTaAAAATCTTTTCGGTAGCGAATAAGCTATATATACATAAAATAAATGAGCT
 CAACAACCAAGTGAAGTGCACCAGGGGATTTGGTAAAAGCTCGACCGACTGCTTTTTTAAATCGATGGAATA
 TACTATCTTGAATCTTACATAAATATTTATGACAAATGACTTGAATTTAATGTAAAAATTTATTTACGTTA
 AAAAAATATAATGATGTC CAATGCAACCTAAGTTCGCAAGACTCAGAATCAACCGCATCGGGGTGAAAAGTT
 ACAGTTTTACAGATACCGAGATCAATTTTCGTTTAATAAGTTGACTTCTGGTACACTAAACAGTTAATG
 TTATAACAAAGTATGAAAGTTTGTAGACAAATAGCCTAGAACCAGGACTGAGTACCTAATAATGTCAACGCG
 CGAGCAATCAAGATCTCGAAGACGAGGCGTACGCATTTGCCCGAGAGGCCATCTCAGGCGCCACCCATGG
 TCCGAGAGCAACCCAGATTCAGAGTCAACCAAGAAAGACGAGCTAGCGATCAATTAAGTAAGGTAACATA
 GATGAAACAAATGGTATTTGGAATGCAAGCCAAACAACGGATGATCTGACTGCACCCGAATAGAAGTTT
 GTATGGACACAATGCGCTATCACCGCTGGCTGACGATCGCACGCCCTAGCCCTCAAAAGAGCAAGCT
 GCGGTGATAACTGTACTGATGGCACTTAAACCTGAACCCGTTTTCGCTGTAACGAATGTATCAAAAACA
 CGGCAACGTTCCGTTAGACGGACCGGACCA

>D0
 AAGAGCCCAAGGTCGGTCTTTAGCTTGGCCAAAACCTACCCATCCAGCTGGCACTTCCACCAACGGCACC
 AAGACACGAGCGGAAAATAAAAAGCCACACCACCCACTTAGAATCTCCGTTTAGCAGCAGTTGTTCAACA

GAATTGGCTGGCTTCGGCCGGTT CAGCCTCAGTTGATAAATTATATAATATCTATGTCTTGGCTATCGC
TGGGCCCTAATTGGCCAGACAAAAGCACCGT TTTTATGCATAACTGGAGGCTTACAATTTGGCCTTCGAC
ACCGGGTCTTACATTTGCCAAAAAGAACGAGAACTCgGCAAGCCAATTACACTCGAAGAAGCAGCGGG
ATCGTTTCGATGCC TAGCCCTGGGCCAGTTCAATCACTCCCGCCGATAAATAGCCGGCTCTGCAATGGGA
ACTTTTCCAAGACGAGATCGATCTTTGGAAAACACCACCTCAGTTTCCATTTCTGT TTTTGGTCCGGA
AGTGGCATCGTGTCTTCCAGAAAGCGCTCCAAATGGTGCCACCATTAGCCAGGGGGAAGCCGGCGGCAG
TCACTTAGCTGTGCAATTTAAATAC TTTTAAATGATTAATAATTGCGGCGAGGCAAGTCCAAAACAAG
ACGACGAGGACGACTTGGCTGTGACGTTTTCGATGCAACCCGACCGGGGACTGCCACTCTTTAGCCAGTTA
ATTGGCAGCAAAAGCGACAGTGCAGCGGCAGCAACTGCTTTTACCAGGAAATCAATAAACGCTCGTCCA
GCGCAAAAGTAATCGCAACACGCACATCTCAATTTCCGTTGGCAGAAAAAACCCCTCACCAGCTCAGTTT
CCCGTGCCAAATTAACCAGAGCAACATAGCCAGTTTCTTCTCCTGCGGCATGTGAAAAGGCAACAGTG
CTGGCCGAAAAATACCAGCAAAACCTGAGTTCTAGTTGCGATTTCGGAATTTGGACTATAAAAAGCGG
CCGTGCGGTAGCTTCTTCACTACAACAGTCCAAAAGCATCTCCAACATGAAGTTCTTCCAGCAATCG
TCCTGGGTCTGATGCTCGTCTGGCCATCATGGGTTCCGTGGCTAGCGCCAAGCCAGGAGGCGGAGGAA
CCGAGGAGTCTGGTCCGGGACTCAGAGTCCGGGACAGTCCGTGCCGAGGAGCCAGCAGGACTACCT
CAACCTGGCGGACCTCACACTGCCGCTCCACATGCTGGTGAAC TAGAGCCCGGGAATTTAGAGGATTT
TTAACAATCTTTGTTTTTTTTTGTCAATCAATGTACTATAATTGCAAGAATTTACGGTGTTCATATTT
CAATAAACCTATTTAAAGCTGAATACATAATTTACGTTAATAAATGTTCTTGATACGATAAAATTTACTTAA
TTAAAATACATTACATTTCACTTAAATATTTGTAAAATAAAATATATTTAAAAAATATTTAATCCTAG
TTGTGGGATTCATAGACTTATTTATTTGTTTTTATTAATTTGTAATTTGTATCAAAAAGTTATTTTGCCA
AACAGTGAATCTTAAATATATATCAAGTTCAATGCACAAATTAACCTATATAAATGTGACCAAAAATTTAG
AAATCAACCTATGTAATAAATTAACAACAACCAATCATATCTGAATTTAAATATATAAAAAGAGGAGATAA
ACCATTTATAAAATGGTCTCACCTTTTTTTAGTTTATTTGATGCATGTTTTAATTTTGTAAAAATCATATTT
CTGATGCTATTTCAATTTTCCAGCCAGTAACTCAAAAAGTCGACCTATCACTCTCCCCCTCTTATATTTG
ACCTATAAATACCCACCGCAAAATGCCCAACCAACCTAACCCACAGAGCCAGTTCCGGTGT TTTAATGGAC
AATTTATCCTTTAGTTT CAGAAGCCCTCCCTTTATGCGTATTTCCCGCTTGGCTGCGAATACGCCCTAACGA
AATTAATCGAGCCCTGAAACCCAGTTTCCGGTAAGTGCTTCTTATGAATATTTTCCATTTACTTTAATTTGA
AGGCTGCCAATTTGGTGGTCCCGAGTGTGTGACTGCTGGCAATGAGGCGGTAATACGTTAAGTCGGAGCT
CGGAAACGGGATGATGGACCAAGTTGAGGCGAAGTACATCAATCTCAATTTGCCGCACTTATCGAACGGTTG
CCTTGGGAGATCGCTGCGATTGTTTATCGATAAATCGCCGATTACCGCGCTGAGCGGCTTTAAAGACCCA
TAAGAAATGCCGGCATGGCGGCTTTAGATAAGTAAGTCTGCGGGCGCTCATAAATTTGAGCGCGATCCA
CGATTTATGCACTCGCTGAAAAGCTATTACCATTAGGCTTTTCCGACACACGGATTTTCCGCTTGCCT
GAGACAAGTGCAGCGCGGAGTTGCAGGCAAAATTTATGCTGAGGCAATGCGCGGCAATGCTTACCCGAA
ATCAAAATACGGCAACCTCTATTCACCTATTTGCTTAGTTTTTGGCAGTGAGCGGCGCAGCGCTTGTAT
TGGCATCTAAATTTATTCGTTTAAAGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTT
CCCCGCTTTTGGCTTGAATAAATTAATCGAATTC CCGCTGGCTATTAACAACACAAAAGGCGCTCTCGT
CTGTTTCAATGTAATTTGCAAAATGCTCAATCGCCATAATGATGTGCGCCATGCAATAGTTTGTGCGCA
ATCATTTTATGACACCCCTA ACTGGTGT TTTCTACGCATAAATATGTCATGGCTTAGGGCTTTTGGTG
GACTTACCAACTGAAGAAGACGATTGTGGGGTGGCTTTGGCGCAGTGCGCGCTTGGAGCAGGAAATCTC
TTTCTCGGCTGTCTGATTTTGGCCAAAGCAAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGG
TCAGGGCCGTGGACATTGAACTTGA AAAACGCAAGCCAGCGCGCAAAAACATTTGATTTCAACGAACGGCAAGTG
CTGCGGGCATGGGTGCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTTCCCGGCCAACACTGG
GGGAGAAAAATAAAAGGAAAAATGTTT CAGGCTGCCATAAGTGGGAAAAAGGAAAAACAAAATGAAAAC
GGCCGGCAATGTCACTCGGCATTCGCTTGATTTTCCGCTAACCTCGCAGCGGCTCTGTGTGTAATAAT
GTCTAATGTTGCATGCCGTTGCATAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTATTTATTTTFA
CTTCTGCCATGTTCCGCTGCCACCGTATTTCTTTTCCGCACTTAGTGCCTCGCTTGATAATGATGTT
TTGTTTTTCCGCGGACAAAACCTCGTTTCGATTATTTGGAAAAAGCGGTATAAATCATCGCCCGCAAGTCT
GGCAAAACAGCAAAATGAAAAC TGCAAGCTGAAAAC TGA AAACTGAAAAC TGA AACCCAAACAAACACAGC
ATCCACACGACGAGGTGAAAATGAAAATAAATA CGGACTGAGCAGCTGAAAACGAGTCAATTCGATTCAA
ATTCGAGGTTCAACGGCTCCCGCATCGCATCATTAAGTGCCTTCGCTGATACGCGGCTCTTATGCA
ACGAGCACACAAATTAATTAATAAGCGTCTGGTTGTTTCCGCTTGGCTTTTGGGACCTGCGCATCGCAA
TAAATTAAGGACGATTAGTCGCAATATGTCACATAGTTGGGCTGCTTACTTTCTGTGGGTGAGCCG
AGCGCAGAAATGCGGCAAGGATCGAGTTAAACCGCTTTTCCGAGGCAAGAGTTTTTCCGATTTTGCAT
AAAATCGCAACGCATAAGTGGCAAGCATTGATGAAAAC TCGGGAAAAAGAAATAAAAAATTTAAAAAT
AAATATAAATTTATGGCAGAACTTAAGAAAAC TAAATTTGAAAATACTTCTTCTTAGGAAACTGTCCTAGGAA
TATTTGTTTTTCCAGCATTCATAATTTCTCCATCTTTTGTCTATTTGCCAGACATTTCTTGGC
CGAAGTGTAGCTGGTGGGCTCCAGATTAATGCAAAACCTTCGTCAGCGGAGGTCGTAACGATCTTTTG
CCATTTGGCTCGTTCAATATGCGTGTGGTATAGCTTTATTTTGGCATTTCCTCTTTTTTGCACACG

TGCACTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGTTCGCACTCGCTTGGCCATGGC
 CATTAGAGCATTACCCGTTAGGGGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAAC
 TTACGGCCAGCTGAGTGGAGTGTGGAAACGCACTTCTTAATTTTCGGCGGTATGTAACTCGAGCTGAGTG
 TCGGATACATATGCCAAAATCACCTGCTCATAATAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTG
 AATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTCTG
 TTCGCAAGACCTGTCCAGATACTGTTTATAGGTATAATTTAAGTGCATATCAGGTTTATTTACAT
 TTATATCGTATTTATATGGTAACTGCAGCAGATGCTGTGCTACAAAATTTAGAATCATTAAAAACAACATA
 TTTGCCACAGAAAATGTGTGAAATAAATAAACTAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCATA
 AATAATATTAAGAAATAAAGAAAATCAGTAGATGTAAGTACTTCTGACCTACGTTGCATGGTATTC
 AATAAAGACTCGAAAATACTCTCACTCACTGAAGTGAACCCAGTGTTTTGTAAATGGCTAGCACATAAAT
 CAGCTGAATCTTAAACGTATCTGAAGCCAGGAGTGTGGAGAATTCGGTGTGCCAAAAACCAAGACC
 AGACCAACCCCTTTCAAACCTTATGAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGCTCAGGGGAT
 CGGGGCGCTGATACTCGCACTTAATAAACATGCGTGAATAAATCAATCAGCGAAGACAAAAGCCACGCA
 AGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCAGGGTGCATATAGACCATAAAGTCCGCATGGTGG
 CCACCAACCGAGCCACCGAAAGCAGCGAATGGCCGAACCCGAAAGTGGCGCTTCTGTTTCGCTTCCA
 TTGGCTGCTTTCGCTTCGGGAAAAAACCTCATATAAACCTGGCCGACATATTGAGTCCAACAGTCC
 TAAGCGGCCACGGTCCACAGAA

>D1

GCCAGCCAGTAATCTAAAAAGTCGACCTATCACTCTCCCTCTTATATTTTCGACCTATAAATACCCACCG
 CAAATGCCGCAACCAACCTAACCCACAGAGCCAGTTCGGGTTGTTTAAATGACCAATATTCCTTAGTTCAGA
 AGGCGCTGCCTTTATGCGTATTTCCCGCTTGGCTGCGAATACGCCAATACGAAATTAATCGAGCCCGTAA
 ACCCAGTTTCGGTAAGTGTCTTTATGAAATATTTCCATTTACTTTAATFGAAGGCTGCCAATGTGTGTG
 CCGGAGTGTGACTGCTGGCCAAATGAGGCGGTAAATACGTTAAGTCGGAGCTGCGGAACGGGATGATGGA
 CCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAAGCGTTCCTTGGGGAGATCGCTGC
 GATTTGTTTATCGATAAATCGCCGATTAACCGCTGAGCGGTCTTAAAGACCCATAAGAAAATCGCCGATGG
 CGGCTTTAGATAAGTAAGTCGTGGGGCGCTCATAAATTTGAGCGCGATCCACGGATTTATGCACTCGCT
 GAAAAGCTATTACCAATTAGGCTTTTTCGGCACCGGATTTTTCGCTTGGCTGAGACAAGTGCAGCGCGG
 CAGTTGCAGGCAAAATATGTGTGAGGCAATGCCGCGGCGATGCTACACCGAAATCAAATACGGCAACCT
 CTATCACTTATTTGCTTAGTTTTTCGGCAGTGCAGCGCCAGCGCTTGTATTGGCATCTAATTTATTCGG
 TTTAAGGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGCTTTTGGCTTGAA
 TAAATTAATCGAATTCCTCGCTGGCTATTAACACACAAAAGCGCTCTCGCTGTTTCAATGTAAATG
 CAAATTCGCTCAATCGCCTAATTGATGTGCCATGCATAGTTTTGTGCCAATCATTTTATAGTACACCC
 CTAACCTGGTGTCTTCTACGCATAAATATGTGCCATGGCTTAGGGCTTTTGGTGGACTTACCAACTGAAGAA
 ACGATTTGGGGGGTGGCTTTGGCGCAGTGCAGCGCTGCGAGCAGGAAATCTTTCTCGGCTGTCTGAT
 TTTGGCCAAAGCAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGCTCAGGGCCGTGGACATG
 AACTTGAAGACGACCCAGCGCCGAAAACATTTGATTAACGAAACGGCAAGTGTGCGCGGATGGGTGTC
 TCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTCCCGCGCCAAACACTGGGGGGAGAAAATAAAGG
 AAAATGTTTCAGGCTGCCATAAGTGGGAAAAAGGAAAACAAAACATGAAACACGGGCCGGCAATGTCACT
 CGGCATTCGCTTGAATTTCCGCCTAACCTCGCAGCGGCTCTGTGTGTAATAATGTCTAATGTGCAATGCGG
 GTTGCATAAATCGTGTGGCAATTTATGCCAGAGAGATTGCTTATTTATTTTACTTTCTGCCATGTTCCG
 TGCCACCGTATTTCTTTTCGGCCACTTAGTCCGCTCCGCTTGATAAATGATGTTTGTTTTTCGCCGGGACA
 AACTCGTTTCGATTTATGGGAAAAGCGCTATAAATCATCGCCGCGAAGTCTGGCAAAACAGCAAAATGA
 AACTGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAACAAACACAGCATCCACACGACGAGGTG
 AAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTTCAAATTCAGGTTCAACGGCT
 CCGCGGATCGCATCATTAAGTGCCTTCGCTGGATACGCGGCTTATGCAACGAGCACACACAATTA
 TTAATAAGCGTCTGGTGTTCGGCCGCTTTTCGGGACTGCGCGATCGCAATAAATTAAGGCAGCATT
 GTCGCAATATGTGCGCACATAGTTGGGCTGCTTACTTTCTGTGGGTGAGCCGAGCGCAGAAATGCCGCCAA
 GGGATCGAGTTAAACCGCTTTTCGGCAGCCAGAGTTCGCAATTTGCAATAAATCGGCAACGCATAA
 GTGGCAAGCATTTGATGAACTCGGGAAAAGAACTAAAAATATTTAAAAATAAATAAATTTATGGCA
 GAACTTAAGAACTAATTTGAAATACCTTCTTTAGGAAACTGTCCTAGGAAATATTTGTTTCCCGCAGCA
 TTGCTCAATATTTCTCCATCTTTTCTTATGTCAGACATTTCTTGGCCGAAGTGTAGCTGGTGGG
 TCTCCAGATTAATGCAACCACCTTCGTCAGCGGAGTCTGTAACCTATCTTTGCCAATTTGGCTCGTTCAT
 TATGCGTGTGGTATAGCTTTATTTTTCGCATTTTCCCTCTTTTTCGACCGAGTGCAGTTGGGCAAGAGA
 GTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTGCTTTCGGCCATGGCCATTAGAGCATTACCCGC
 TTAGGGCGCCCTAAAGTCCAGGTGGTCCCGAGGACCACAAGAGTATGCAACTTACGGCCAGCTGAGTGG
 AGTGTGGAAACGCACTTCTTAATTTTCGGCGTTATGTAACCTCGAGCTGAGTGTGCGATACATATGCCAAA
 ATCACCTGCTCATAATTAGCGGAAACCACTGTTTGGCCCTCGCCGACTGTGAATCATCaGAGCTGCCCA

11

ATCGAAATCAAAGCCAAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGCTGTTTCGCAAAGACCTGTCC
AGATACTCTGTTTATAGGTATAATTAATTAAGTGCATATCAGGTTTATTTTACATTTATATCGTATTTATATTG
GTAACCTGCAGCAGATGCTGTGTACAAATTTAGAAATCATTTAAAAACAAACATATTTGCCACAGAAAATGTG
TGAATAAATTAAC TAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCTAAAAATAATATTATGAATAAT
CAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCATAAAGACTCGAAAATA
CTCTCACTCACTGTAAGTGAACCCAGTGTGTTTGTAAATGGCTAGCACATAAATCAGCTGAATCC TAAACGT
ATCTGAAGGCCAGGAGTGTGGAGAAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCATACCC TTTCAA
ACCTTATGAAAAATGGCAAGCCCGCCGAAAGGTGTTGGCCGGTCCAGGGGATTTCGGGGCCCGTGATACTC
GCACTTAATAAAACATCGCGTGAATAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGAAGCCAAAAGTGT
CGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCACCCGAGCCACCG
AAAGCAGCCGAATGGCCGAAACCCCGAAGTGGCGCCTTCGTTTTCGCTTCCATTGGCCTGCCTTCGTCTT
CGGAGAAAAAACCTCATATAAAACGTGGCCGACATATTTAGTCCAAAGTCCGTAAGCCGCGCCACGGTCCA
CAGAA

>D2

CCTTAGTTTCAGAAGCGCCTGCCTTTATGCTATTTCCCGCTTGCCTGCGAATACGCCCTAACGAAATTA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTTTTATGAATATTTTCCATTTACTTTAATTGAAGGCTG
CCAATTTGTTGGTCCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTCGGAA
CGGATGATGGACAGTGTGAGGCGAAGTACATCAATCTCATTTCGCCCACTTATCGAAGCGTGTGCTTGG
GGAGATCGTCCGATTTGTTTATCGATAATCGCCGATTTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATCGGGCGATGGCGCTTTAGATAAGTAAGTCTCGGGCGCTCATAAATTTTCGAGCGCATCCACGGAT
TATGCACTCGCTGAAAAAGCTATTACATTTAGGCTTTTCGCGACCACGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGCACTTGCAGGCAAAATTTATGTGTGAGGCAATGCCCGGGCATGTCTACACCGAAATCAAA
TTAGGGCAACCTCTATTCTACTTATTTGCTTAGTTTTTCGCGCAGTGTGAGCGCCAGCGCTTGTATTGGCAT
CTAATTTATCCGTTTAAAGGACGCAATTTTCGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
TTTTGGCTTGAATAAATTAATCGAATTCGCCGCTGCTTATTAACACACAAAAGGCGCTCTCGCTGTTT
CAATGTAATTTGCAATTTGCTCAATCGCC TAAATGATGTGCGCCATGCAATAGTTTTGTGCCAATCAT
TTTAGTACACCCCTAACTGGTGTGTTTCTACGCATAATATGTGCCATGGCTTAGGGCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTTGTGGGGTGCCTTTGGCGCAGTGCGCCCTGCGAGCAGGAAATCTCTTTCTC
GGCTGTCTGATTTTGGCCAAAGCAAAATAATCCGGCTGGCAGATAGGCAGAGGGACCCGGCGGTGAGGG
CCGTGGACATGAACTGAAAAACGACGCGAGCGCCGAAAAACATTTGTATTC AACGAACGGCAAGTGTGCGC
GGCATGGGTGTCTCTGGCTAAGGTACGGCGGTTGGGCAACAGGTTTTCGCCGCAACACTGGGGGAG
AAAAATAAAGGAAATGTTTCAGGCTGCCATAAGTGGGAAAAAGGAAAAACAAACATGAAACACGGGGCG
GGCAATGTCACTCGGCATTCGCTTGTATTTTCCGCTAACTCGCAGCGGTCTGTGTGTAATAATGTCTAA
TGTGTGCATGGCGTTGCATAATCGTGTGGCAATTTATGCCAGAGAGATTGCTTATTTATTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCCTCCGCTTGATAATGATGTTTGTGTT
TTCGCCGGGCAAACTCGTTTCGATTTATGGAAAAAGCGCTATAAATCATCGCCGCCGAAAGTCTGGCAAA
ACAGCAAAATGAAAACTGCAAGCTGAAAACTGAAAACTGAAAACTGTAACCCAAACAAACACAGCATCCCA
CACGACGAGGTGAAAAATAAATAACGACTGAGCAGTAAAAACGAGTCAATTCGATTCAAATTCGA
GGTTCAACGGCTGCCGCGATCGCATTTAAGTGCCTTTCGCTGGATACGGCGCTCTTATGCAACGAGC
ACACACAAATTAATAAAGCGTCTGGTGTGTTTCGGCTGGCTTTTTCGGGACTGCGCGATCGCAATAAAT
AAGGCAGCATTAGTTCGCAATTAATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGCCAAAGGATCGAGTTAAACCGCTTTTCCGCGAGCCAAAGGTTTTCGCAATTTGCAATAAATC
GGCAACGCATAAAGTGGCGAAGCATTTGATGAAACTCGGGGAAAGAAAGTAAAAATAATTTAAAAATAAAT
AAATTTATGGCAGAACTTAAAGAACTAATTTGAAATACTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
TTTTCCCGCAGATTGCTCAATATTTCTCCATCTTTTGGCTTATGCCCAGACATTTTCTTGGCCGAGT
GTAGCTGTTGGTCTCCAGATTAATGCAAAACACTTCGTCAGCGGAGGTGTAACGTATCTTTGCCATTT
TGGCTCGTTCATTAATGCGTGTGGTATAGCTTTATTTTGGCATTTCCTCTTTTTCACCAGCTGCAAGT
TGGGCCAAGAGGATTTATGCGAATCGGTGCGATTTTCGGGTTTTCGCACTCGCTTGGCGCCATGGCCATTAG
AGCATTTACCCGCTTAGGGCGCC TAAAGTCCAGGTGTTCCAGGACCAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGTGGAACGCACTTCTAATTTTCGGCGTTATGTAACCTCGAGCTGAGTGTGCGAT
ACATATGCCAAAATCACC TGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGACTGTGAATCAT
CaGAGCTGCCAATCGAATCAAAGC CAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGCTGTTCGCA
AAGACCTGTCCAGATAC TCTGTTTATAGGTATAATTAATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTAATTTGGTAACTGCAGCAGATGCTGTGTACAAATTTAGAAATCATTTAAAAACAAACATATTTGCC
ACAGAAAAATGTTGTAATAATTAAC TAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCTAAAAATAAT
ATTTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTTCAATAAA
GACTCGAAAAACTCTCACTCACTGTAAGTGAACCCAGTGTGTTTGTAAATTTGCCCTAGCACATAAATCAGCTG

AATCCTAAACGTATCTGAAGGCCAGGAGTGTCCGGAGAATTCCGGTGTGCCAAAAACCAAAGACCAAAGACCA
 TACCCTTTCAAACCTTATGAAAATGGCAAGCCCGGCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
 CCGGTGATACCTGCCTTAAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCACGGCCTAGAGA
 AGCCAAAGTGTCCGAAGTGGCCGATCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
 CCCGAGCCACCGAAAGCAGCGAATGGCCGAAACCCGAAAGTGGCGCCTTCGTTTTCGCTTCCATTTGGCC
 TGCTTTCGCTTTCGGAGAAAAAACCTCATATAAACCTGGCCGACATATGAGTCCAACAGTCTGTAAGCG
 CGCCACGGTCCACAGAA

>D3

CGATAATCGCCGATFACCGCGTGTAGCGGCTTTAAAGACCATAAGAAAATGCGCGATGGCGGCTTTAGA
 TAAGTAAGTCTCGGGGCGCTCATAAATTCGAGCGCGATCCACGGATTTATGCACCTCGCTGGAAAAGCTA
 TTACCAATTAGGCTTTTCGGGACACGGATTTTCGCTGCTGAGACAAGTGCAGCGCGGAGTTCGAGG
 CAAATTTAGTGTGAGGCAATGCCGCGGGCATGTCACACC GAAATCAAATFACGGCAACCTTATTCACCT
 ATTTGCTTAGTTTTTTGGCAGTGTAGCGGCCAGCGCCTTGTATTGGCATCTAATTTATTCGTTTAAAGGACG
 CAATTTCTGAGCTAAAACCTCGCTTATGAGAGATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATC
 GAATTCGCCGCTGGCTATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAATTTGCAAAATGGCTC
 AATCCGCCAATATGATGTGCGCCATGCAATAGTTTTGTGCCAATCATTTTTAGTACACCCCTAATCGGTG
 TTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGTGGACTTACCAACTGAAAGAAGACGATTTGTG
 GGGTGGCTTTGGCGCAGTGGCGCCGCGAGCAGGAAATCTCTTTCGGCTGTCTGATTTTGGCCAAG
 ACAATAAATCCGGCTGGCAGATAGGAGAGGGGACCCGGCGGTCAAGGGCTGGACATTTGAACTTGA
 CGAGCCAGCCGAAAAACATTGATTTCAACGAACGGCAAGTGTGCGCGCATGGGTGTCTCTGGCTAAG
 GTTACGGCGTTGGGCAACAGGTTTTCCCGCGCCAACACTGGGGGGAGAAAAATAAAGGAAAAATGTTCA
 TAAATACGGAATGGGAAAAAGAAAAACAAACATGAAACACGGCCCGGCAATGTCACTCGGCATTCGC
 TTGATTTTCGCCCTAACTCGCAGCGGCTCTGTGTAAATAATGTCATAATGTCATGCCGTTGCATAAT
 CGTGTGGCAATTATGCCAGAGATTCGCTTATTTATTTTTACTTTCTGCCATGTTCCGCTGCCACCGTA
 TTTCTTTTCGGCCACTTGTGCGCTCGCTTGATAATGATGTTTTGTTTTTCGCGGGACAACTCGTTTTC
 GATTTATGGGAAAAGCGCTATAAATCATCGCCCGGAGGCTGGCAAAAACAGCAAAATGAAAACGTCAAG
 CTGAAAACGTAAAACGTAAAACGTAAACCAAAACAAACAGCATCCACACGACGAGGTGAAAATGAAAA
 TAAATACGGAATGAGCGACTGAAAACGAGTCAATTCGATTTCAAATGCAAGTTCAACGGCTCCCGCGATC
 GCATCAATTAAGTGCCTTCGCTGGATACGGCGCTTATGCAACGAGCACACAAATTAATTAATTAAGCG
 TCTGGTTGTTTTGGCCTGGCTTTTGGCGACTGCGCATCGCAATAAATTAAGGCAGCATTAGTGCCAATTA
 TGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCAGAATGCGGCAAGGGATCGAGT
 TAAACCGCTTTTCCGAGGCAAGAGTTTTTCGCATTTTGCATAAATCGGCAACGCATAAGTGGCGAAGC
 ATTTGATGAACTGCGGAAAAGAAAGTAAAAAATTTAAAAATAAATAAATTTATGGCAGAACTTAAGA
 AACFAATTTGAAATACTTCTCTTAGGAAACTGTCCCTAGGAATATTTGTTTTTCCCGCATGTCTCAATA
 TTTCTCCATCTTTTGTCTTATGCCCAGACATTTTCTTGGCCGAAGTGTAGCTGGTGGTCTCCAGATT
 AATGCAAACTTTCGCTAGCGGAGTGTGTAACGATCTTTTGGCCATTTGGCTCGTTTATGCGGTGTG
 GTATAGCTTTATTTTTGCAATTTCCCTCTTTTTTGCACCAGCTGCAGTTGGGCAAGAGAGTTATGCGAA
 TCGGTGCGATTTTCGGGTTTTTCGCACTCGCTTGGCGCCATGGCCATTAGAGCATTACCCTGTTAGGGGCC
 CTAAGTCCAGGTGCTCCCGAGGACCAAGAGATTTGCAACTTACGGCCAGCTGAGTGGAGTGTGGAA
 CGCACTTCTTAATTTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGATACATATGCCAAAAACCTGCT
 CATAATTAGCGGAAACCACTGTTTGGCCCTCGCCGACTGTGAATCATCAGAGCTGCCCAATCGAAATCA
 AAGCCAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTGTTTCGCAAAAGACTGTCCAGATACTCTG
 TTTATAGGTATAATTAATTAAGTGCATATCAGGTTTATTTACATTTATATCGTATTAATTTGGTAACGCA
 CAGATGCTGTGTACAAAATTTAGAATCATTTAAAACAAACATATTTGCCACAGAAAAATGTGTGAATAAT
 AACTAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCTAAAAATAATTTATGAATAATCAAGAAAAAT
 CAGTAGATGGTAAAGTACTTTCGTACCTACGTTGCATGGTATTTCAATAAAGACTGAAAAATACTCTCACTCA
 CTGTAAGTGAACCCAGTGTTTTGTAAATTGCTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGC
 CAGGAGTGTCCGAGAAATTCGCTGTCCAAAACCAAGACCAAAGACCATACCTTTTCAAACCTTATGAA
 AAATGGCAAGCCCGGCAAGGTTTGGCCGTCAGGGGATTCGGGGCCCGTGATACTCGCACCTTAATA
 AACATGCGTGAATAATCAATCAGCGAAGACAAAAGCCACGCATGAAAGAAGCCAAAGTGTCCGAAGTGGCC
 GATCCACGGGTGACCATAAGACCATAAAGTCCGATGGTGGACACCACCCGAGCCACCGAAAGCAGCCG
 AATGGCCGAAACCCGAAAGTGGCGCCTTCGTTTTTCGCTTCCATTGGCTGCCCTTCGTTCTCGGAGAAAA
 AACCTCATATAAAGCTGGCCGACATATTGAGTCCAACAGTCTGAAGCGCCACCGTCCACAGAA

>D4

TCATAATATTCGGTTAAGGACGCAATTTCTGAGCTAAAACCTGCTTATGGAGAGATCTAAATTTCCCGG
 CTTTTGGCTTGAATAAATTAATCGAATTCGCCGCTGGCTATTTAAAACACACAAAAGGCGCTCTCGTCTGTT

TCAAATGTAAATGCAAATGCTCAATCCGCCTAATTGATGTGCGCCATGCAATAGTTTTGTGCCAATCAT
TTTTAGTACACCCCTAACGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTT
ACCAACTGAAAGAACGATTTGTGGGGTGCCTTTGGCGCAGTGGCGCCCTGCGAGCAGGAAATCTCTTTCT
CGCCTGTCTGATTTTGGCCAAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTGAGG
GCGGTGGACATTGAACCTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGGCAAGTGTGCG
CGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTGGGCAACAGGTTTTCCCGCCGCAACACTGGGGGGA
GAAAATAAAAAGGAAAATGTTCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAAACCGGGCC
GGCAATGTCTACCTCGCATTTGGCTTGTATTTCCGCCTAACTCGCAGCGCTCTGTGTGTAATAATGTCTA
ATGTTGCATGCGGTTGCATAAATCTGTGGCAATATGCGCAGAGAGATTCGCTTATTTATTTTACTTTCT
TGCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGGCTCCGCTTGATAATGATGTTTTGT
TTTTCCCGGGACAACTCGTTTTGATTTTGGGAAAAGCGGTATAAATCATCGCCGCGAAGTCTGGCAA
AACAGCAAAATGAAAACGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAAACAAACACAGCATCC
ACGGCAGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTCG
AGGTTCAACGGCTGCGCGGATCGCATCATTAAGTGGCGCTTCGCTGGATACGGCGCTCTTATGCAACGAG
CACACAAATTAATTAAGCGCTCTGGTTTTCGGCTGGCTTTTGGCGGACTGCGCATCGCAATAAAT
TAAGCAGCATTAGTCGCAATTTATGTCCACATAGTTGGCTGCTTACTTTCTGTGGGTGAGCGGCGG
AGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCGCGAGGCCAAGAGTTTTTCGCAATTTGCATAAAAT
CGCAACGCATAAGTGGCGAAGCATTGATGAAACGCGGAAAAGAGTAAAAAATATTTAAAAATAAATA
TAAATTTATGGCAGAACTAAGAACTAATTTGAAATACTTCTTCTTAGGAACTGTCCCTAGGAATATTT
GTTTTCCCGCAGCATGCTCAATATTTCCCTCATCTTTTGGCTTATTGCGCAGACATTTCTCTGGCCGAA
GTAGCTGGTGGGTCTCCAGATTAATGCAACCACTTCGTGAGCGGAGGTCGTAACGTATCTTTGCCCAT
TTGGCTCGTTTATGCTGTGGTATAGCTTTATTTTGGCATTTCCTCTTTTTTGCACCAGCTGCAG
TTGGCCCAAGAGAGTTATGCGAATCGGTGCCAATTTTCGGGTTTTCGCACTCGCTTGGCGCATGGCCATTA
GAGCATTACCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCAGGGACCAAGAGTATTGCAACTTAGC
GCCAGCTGAGTGGAGTGTGGAACGCCTTCTAATTTTCGGCGGTTATGTAACCTCGAGCTGAGTGTGCGA
TACATAAGCCAAAACACCTGCTCATAAATAGCGGAAACCAACTGTTGGCCCTCGCCGACTGTGAATCA
TCAGAGCTGCCAAATGCAAAATCAAGCCAAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTCTTCCG
AAAGACCTGTCCAGATACTCTGTTATAGGTATAAATTAAGTGCATATCAGGTTTATTTACATTTATA
TCGTATTTATAATGGTAACGTCAGCAGATGCTGTGCTACAAAATTTAGAATCATTTAAAAACAAACATTTTC
CACAGAAAATGTGTGAAATAAATAAAGCTAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCTAAAATA
TATTATGAATAATCAAGAAAATCAGTAGATGGTAAAGTACTTTCGTACCTACGTTGCATGGTATTCAATAA
AGACTCGAAAATACCTCCTACTCTAAGTGAACCCAGTGTTTTGAATTCGCTAGCACATAAATCAGT
GAATCCFAAACGTACTGTAAGCCAGGAGTCTCGGAGAAATTCGGTGTGCCAAAACCAAGACCAAGACC
ATACCTTTTCAAAACCTTATGAAAAATGGCAAGCCGGCGAAAAGGTTGGCCGGTCCAGGGGATTCGGGG
GCCGTGATACTCGCACTTAATAAACATCGGTGAAAATCAATCAGCGAAGACAAAAGCCACGCACCTAGAAG
AAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACATAAAGTCCGCATGGTGGACCACC
ACCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCGAAAGTGGCGCCTTCGTTTTTCGCTTCCATTGGC
CTGCCTTCGTCTTCGGAGAAAAAACCTCATATAAAACGTGGCCGACATATTGAGTCCAACAGTCGTAAGC
GCGCCACGGTCCACAGAA

>D5

GTGCCAATCATTTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAAATATGTGCCATGGCTTAGGGCCCTT
TTGGTGGACTTACCAACTGAAGAACGATTTGTGGGGTGCCTTTGGCGCAGTGGCGCCCTGCGAGCAGGA
AATCTCTTTCTCGGCTGTCTGATTTTGGCCAAAGACAATAAATCCGGCTGGCAGATAGGCAGAGGGGACC
CGCGGCTCAGGGCCGTGGACATTGAACCTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAACGG
CAAGTGTGCGCGGATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCGCCGCAAC
CACTGGGGGAGAAAATAAAAAGGAAAATGTTTCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATG
AAACACGGCCGGGCAATGTCACTCGGCATTGCTTTGTTTTCCGCCTAACTCGCAGCGGCTCTGTGTGTA
AATAATGTCTAATGTGCATGCCGTTGCATAAATCGTGTGGCAATTTATGCGAGAGATTGCTTATTTAT
TTTTTACTTTCTGCAATGTTCCGCTGCCACGTAATTTCTTTTCGGCCACTTAGTGGCGCTCCGCTTGATAAT
GATGTTTTGTTTTTCGCCGGGACAAACTCGTTTCGATTATTGGGAAAAGCGGTATAAATCATCGCCGCGC
AAGTCTGGCAAAACAGCAAAATGAAAACGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAAACAAA
CACAGCATCCACACGAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTA
ATTCAAATGTCAGGTTCAACGGCTGCCGCGATCGCATCATTAAGTGGCGCTTCGCTGGATACGGCGCT
TATGCAACGAGCACACAAATTAATTAAGCGTCTGGTGTTCGGCCGCTTTTGGCGGACTGCGCGA
TCGCAATAAATTAAGGAGCATTAGTGCATTAATGTCACATAGTTGGGCTGCTTACTTTCTGTGGGT
GAGCCGAGCGAGAATGCGCCAAAGGATCGAGTAAACCGCTTTTCCGAGGCCAAGAGTTTTTCGCAAT
TTGCATAAAATCGGCAACGCATAAGTGGCGAAGCATTGATGAACTGCGGAAAAGAGTAAAAAATTTTT

AAAAATAAATAAAATTTATGGCAGAACCTTAAGAACTAATTTGAAATACCTTCTTCTTAGGAAACTGTCCC
TAGGAATATTTGTTTTCCCGCAGCATTTGCTCAATATTTCTCCATCTTTTTGCTTATTGCCAGACATTTTC
CTTGGCCGAAGTGTAGCTGGTGGTCTCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAACCGTA
TCTTTGCCCATTTGGCTCGTTCATATGCGGTGGTATAGCTTTATTTTTGCCATTTCCCTCTTTTTTGC
ACCAGCTGCAGTTGGGCCAAGAGAGTTATGCCAATCGGTGCGATTTTCGGGTTTTTCGCACTCGCTTGGGC
CATGGCCATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCAGGGACCAAGAGTAT
TGCAACTTACGGCCAGCTGAGTGGAGTGTGGAACGCACTTCTTAATTTGGCGGTTATGTAACCTCGAGC
TGAGTGTGGATACATATGCCAAAATCACCCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGG
ACTGTGAATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAAGTCAATCGAAGCCAGGTAATTCATATTTA
GTGCTGTTTCGCAAAGACCTGTCCAGATACCTCTGTTTATAGGTATAATTTAAGTGCATATCAGGTTTAT
TTACATTTATATCGTATTATATTTGGTAACTGCAGCAGATGCTGTGCTACAAAATTTAGAATCATTTAAAACA
AACATAATTTGCCACAGAAAATGTGTGAAATAAATTAACATAAAGCTTTGGATGAAGTAAAAAGCCATAAAA
GCCTAAAAATAATATTAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATG
GTATTCATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAAATGCCTAGCAC
ATAAATCAGCTGAATCTTAACGTATCTGAAGCCAGGAGTGTCCGAGAAATTCGGTGTGCCAAAAACAAA
GACCAAAGACCATACCTTTCAAACCTTATGAAAATGGCAAGCCCGGAAAGGTTTGGCCGGTCCAG
GGGATTCGGGGGCCCGTGATCTCGCACTTAATAAACATGCGTGAAAATCAATCAGCGAAGACAAAAGCCA
CGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCA
GTTGGACCACCACCCGAGCCACGAAAGCAGCCGAATGGCCGAAACCCGAAGTTGGCCCTTCGTTTTCC
CTTCCATTTGGCCCTGCTTTCGTAAGAAAAAACCTCATATAAAACGTGGCCGACATATTTAGTCCAA
CAGTCGTAAGCGCCACGGTCCACAGAA

>D6

AGGAAATCTCTTTCGCGCTGTCTGATTTTGGCCAAGACAATAAATCCGGCTGGCAGATAGGCAGAGGG
GACCCGGCGGTCCAGGGCCGTGGACATGAACTTGAACCGCAAGCCAGCCGCGGAAACATTTGATTTCAACGA
ACGGCAAGTGTCCCGGCATGGGTGCTCTGCGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCG
CCAACTGCGGGGAGAAAATAAAAAGGAAAATGTTCCAGGCTGCATAAGTGGGAAAAGGAAAACAAA
CATGAAACACGGCCCGGCAATGCACTCGGCATTCGCTTGTATTTCCGCTAACTCGCAGCGGTCCTGTG
TGTAATAAATGCTAATGTTGCTGCGGTGCAATAAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTAT
TTATTTTTTACTTTCTGCGCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGGCTCCGCTTGA
TAATGATGTTTTGTTTTTCGCCGGGCAAACTCGTTTTGATTATTTGGAAAAGCGCGTATAAATCATCGCC
CGCGAACTCTGGCAAACAGCAAATGAAAATGCAAGCTGAAAATGAAAATGAAAATGAAAATGAAAATGAAA
CAAACACAGCATCCACACGACGAGGTGAAAATGAAAATAAATACCGACTGAGCGACTGAAAACGAGTCAA
TTGATTCAAAATGCAAGGTTCAACGGTGC CGCCGATCGCATTAAGTGGCCCTTCGCTGGATACGCGG
CTCTTATGCAACGACACACCAATTAATTAAGCGTCTGTTGTTTTCGCCGCGGCTTTTCGGGACCTG
CCGATCGCAATAAATAAGGAGCATTAAGTCCAAATTAATGTCACATAGTTGGGCTGCTTACTTTCTGT
GGGTGAGCCGAGCGAGAATGCGGCCAAGGATCGAGTAAACCGCTTTTCCGAGGCCAAGAGTTTTTCG
CATTTTTGCATAAAAATCGGCAACGCATAAAGTGGCGAAGCATTTGATGAACTGCGGAAAAGAAAGTAAAAA
ATTTAAAAATAAATAAATTTATGGCAGAATTAAGAACTAATTTGAAATACCTTCTTCTTAGGAAACTG
TCCCTAGGAAATTTGTTTTCCCGCAGCATGCTCAATATTTCTCCATCTTTTTGCTTATTGCCAGACAT
TTTTCTTGGCCGAAGTGTAGCTGGTGGGCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTA
CGTATCTTTGCCATTTGGCTCGTTCATTAATGCGGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTT
TTGCACCAGCTGCAGTTGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTTCGCACTCGCTT
CGCCATGGCCATTAGAGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCGAGGGACCAAGA
GTATGCAACTTACGGCCAGCTGAGTGGAGTGTGGAACGCACTTCTTAATTTTCGGCGGTTATGTAACCTC
GAGCTGAGTGTCCGATACATATGCCAAAATCACCCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTC
CCGACTGTGAATCATCaGAGCTGCCCAATCGAAATCAAAGCCAAAGTCAATCGAAGCCAGGTAATTCATA
TTTATGCTGTTTCGCAAAGACCTGTCCAGATACCTGTTTATAGGTATAATTTAAGTGCATATCAGGT
TTATTTACATTTATATCGTATTATATTTGTTAACTGCAGCAGATGCTGTGCTACAAAATTTAGAATCATTTAA
AACAAACATATTTGCCACAGAAAATGTGTGAAATAAATTAACATAAAGCTTTGGATGAAGTAAAAAGCCA
TAAAGCCATAAATAATATTAATAAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTG
CATGGTATTTCAATAAAGACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAAATGCTA
GCACATAAATCAGCTGAATCCTAAACGTATCTGAAGCCAGGAGTGTCCGAGAAATTCGGTGTGCCAAAA
CAAAGCCAAAAGACCATACCTTTCAAACCTTATGAAAATGGCAAGCCCGGCGAAAAGGTTTGGCCGGT
CCAGGGGATTCGGGGCCCGTGATACFCGCCTTAATAAACATGCGTGAAAATCAATCAGCGAAGCAAAA
GCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCC
GCATGGTGGACCAACCCGAGCCACGAAAGCAGCCGAATGGCCGAAACCCGAAGTTGGCCCTTCGTT

15

TTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAACCTCATATAAAAACGTGGCCGACATATTGAGT
CCAAACAGTCGTAAGCGGCCACGGTCCACAGAA

>D7

GCCGCCGAAGCTGGCAAAAACAGCAAATGAAAACCTGCAAGCTGAAAACCTGAAAACCTGAAAACCTGTAACCC
AAACAAACACAGCATCCACACGACGAGGTGAAAATGAAAATAAAATACGGACTGAGCGACTGAAAACGAGT
CAATTCGATTCAAATGCGAGTTCAACGGCTGCCGGCGATCGCATATTAAGTGCAGCTTCGCTGGATACG
CGGCTCTTATGCAACGAGCACACACAATTAATTAATAAGCGTCTGGTTGTTTCGGCCCTGGCTTTTCGGGAC
CTGCCGATCGCAATAAATAAGCAGCATTAGTCGCAATTAATGTGCCACATAGTTGGGCTGCTTACTTTTC
TGTGGGTGAGCCGAGCGCAGAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCGGCAGGCCAAGAGTTT
TCGCATTTTGCATAAAAACCGCAACGCATAAGTGGCGAAGCATTGATGAAAACCTGCGGAAAAGAAGTAAAA
AATATTTAAAAATAAATAAATAAATTAATGCGAGAATTAAGAAAACATAATTTGAAAATACTTCTTCTTAGGAAA
CTGTCCCTAGGAATATTTGTTTCCCGCAGCATGCTCAATATTTCTCCATCTTTTGTCTTATTGCCGAGA
CATTTTCTTGGCCGAAGTGTAGCTGGTGGGTCTCCAGATTAATGCAAAACCACTTCGTCAGCGGAGGTCGT
AAAGCTATCTTTGCCCATTTGGCTCGTTCATTATGCGTGTGGTATAGCTTTATTTTTCGCAATTTTCCCTCT
TTTTTGCACCAAGCTGCAGTTGGGCCAAGAGATTATGCGAATCGTGCAGTTTTCGGGTTTTTCGCACTCCG
TTGGCCCATGGCCATTAGAGCATTACCGCTTAGGGCCCTAAAGTCCAGGTGGTCCCAGGGACCACA
AGAGTATTCGCAACTTACGGCCAGCTGAGTGGAGTCTGGAACGCACCTTCTAATTTTCGGCGTTATGTAAC
CTCGAGCTGAGTGTGCGATACATATGCCAAAATCACCTGCTCATAATTAGCGAAAACCACTGTTTGGCCC
TCGCCGATGTGAATCATCAGAGCTGCCAATCGAAAATCAAGCCAAAGTCAATCGAAGCCAGGTAAATTT
ATATTTAGTGTGTTGCGCAAGACCTGTCCAGATCTCTGTTATAGGTATAAATTAAGTGCATATCA
GGTTATTTACATTTATATCGTATTAATTTGGTAACGCGAGATGCTGTGTACAAATTTAGAATCATTT
TAAACAAACATATTTGCCACAGAAAATGTGTGAAATAATTAACCTAAAAGCTTTGGATGAAGTAAAAAAG
CCATAAAGCCATAAATAAATTTATGAATAATCAAGAAAATCAGTAGATGTTAAAGTACTTTCGTACCTACG
TTGCATGGTATTCATAAAGACTCGAAAATCTCTCACCTGTAAGTGAACCCAGTGTTTTGTAAATTCG
CTAGCACATAAATCAGCTGAATCCTAAACGTATCTGAAGGCCAGGAGTGTGCGGAGAATTCGGTGTGCCAAA
AACCAAAGACCAGAACCATACCTTTCAAACCTTATGAAAATGGCAAGCCCGGCCGAAAGGTGTGGCC
GGTCCAGGGGATTCGGGGCCGTGACTCGCACTTAATAAACATGCGTGAATAAATCAATCAGCGAAGACA
AAGCCACGCACTAGAAGAAGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACATAAAG
TCCGATGTTGGACCAACCCGAGCCAGCAAAGCAGCCGATGGCCGAAACCCCGAAGTTGGCCCTTC
GTTTTCGCTTCCATTGGCCTGCCTTCGTCTTCGGAGAAAAAACCTCATATAAAAACGTGGCCGACATATTG
AGTCCAACAGTCGTAAGCCGCCACGGTCCACAGAA

>E0

CCTTAGTTTCAGAAGGCGCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCCTAACGAAATTA
TCGAGCCGTAACCCAGTTTCGGTAAGTGTCTTTATGAATATTTCCATTTACTTTAATTAAGGCTG
CCAAATTTGGTGGCCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCCGAGCTCGGAA
CGGATGATGGACCAAGTTGAGGCGAAGTACATCAATCTCATTTCGCCGCACTTATCGAACGGTTGCTTGG
GGAGATCGTGCATTTGTTTATCGATAATCGCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGCGATGGCGCTTTAGATAAAGTAAGTCTCGGGCGCTCATAAATTTTCGAGCGGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACCATTAGGCTTTTCGCGACACGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGGAGTTCAGGCAAAATATGTTGTGAGGCAATGCCGCGGATGCTTACACCGAAATCAAA
TTACGGCAACCTCTATCTACTTTTGTCTTGTGTTTTCGCGAGTGTGAGCGCCAGCGCTTGTATTGGCAT
CTAATTTATTCGTTTAAAGGACGCAATTTTCGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
TTTTGGCTTGAATAAATAAATCGAATTTCCCGCTGGCTATTAACACACAAAAGGCGCTCTCGTCTGTTT
CAATGTAATTTGCAAAATGCTCAATCCGCCTAATGATGTGCGCCATGCAATAGTTTGTGCCAATCATTT
TTTGTACACCCCTAAGTGTGTTTCTACGCATAATAATGTCATGCGCTTAGGCTTTTGTGGACTTTC
CCAACTGAAGAAGACGATTTGTTGGGGTGGCTTTGGCGCAGTGCAGCGCTGCGAGCAGGAAATCTCTTCTC
GGCTGTCTGATTTTGGCCAAAGCAAAATAATCCGGCTGGCAGATAGGACAGAGGGGACCCGGCGTCCAGG
CCGTGGACATGAACTTGAACACGCGCCAGCCGCGAAAACATTGATTCACGAAACGCAAGTGTGCGC
GGCATGGGTGCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCGCCGCAACACTGGGGGAG
AAAATAAAAAGGAAAATGTTTCCAGCTGCATAAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGCG
GGCAATGCTACTCGCATTCGCTTGAATTTCCCGCTAAGTGCAGCGGTCCTGTGTGTAATAATGCTTAA
TGTTCGATGCCGTTGCATAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCCTCGCTTGATAATGATGTTTGTGTT
TTCGCCGGGCAAACTCGTTTCGATTTATGGGAAAAGCGCTATAAATCATCGCCGCGAAGTCTGGCAAA
ACAGCAAAATGAAAACGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAAACAAACACAGCATCCCA
CACGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTCGA

GGTTCACCGGCTGCCGGCATCGCATCATTAAAGTGCCTTCGCTGGATACGCGGCTTTATGCAACGAGC
 ACACACAATTAATAAAGCGTCTGGTTGTTTCGGCTGGCTTTTGGCGACCTGCCGATCGCAATAAAT
 AAGCAGCATAGTCGCAATATAGTGCCACATAGTTGGGCTGCTACTTTTCTGTGGGTGAGCCGAGCGCA
 GAATGCCGCAAGGATCGAGTTAAACCGCTTTCCGCGAGCCAAAGATTTTCGCATTTTCATATAAATC
 GGCAACGCATAAAGTGGCGAAGCATTGATGAACTCGGGAAAAGAAGTAAAAATATTTAAAAATAAATAT
 AAATTTATGGCAGAACTTAAAGAACTAATTTGAAATACCTTCTTCTTAGGAAACTGTCCCTAGGAATATTTG
 TTTCCCGCAGCATTGCTCAATATTTCCCTCCATCTTTTGGCTTATTGCCAGACATTTTCCCTGGCCGAAGT
 GTAGCTGGTGGGCTCCAGATTAATGCAACCACCTTCGTCAGCGGAGGTCGTAACGTATCTTTGCCATT
 TGGCTCGTTCATTAATGCGTGTGTATAGCTTATTTTGGCATTTCCTCTTTTTCGACCAGCTGCAGT
 TGGCCAAAGAGATTTATGCGAATCGGTGCGATTTTCGGGTTTCGCACTCGCTTGGCCATGGCCATTAG
 AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCCAGGGACCACAAGAGTATTGCAACTTACGG
 CCAGCTGAGTGGAGTGTGGAAACGCACTTCTAAATTTCCGGCGTTATGTAACCTCGAGCTGAGTGTGCGAT
 ACATATGCCAAAATCACCCTGCTCATAATTAGCGGAAACCACTGTTTGGCCCTCGCCGACTGTGAATCAT
 CaGAGCTGCCAATCGAAATCA

>E1

CCTTAGTTCAGAAGGCGCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCCTAACGAAATTA
 TCAGACCCGTAACCCAGTTTCGGTAAGTGTCTCTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
 CCAATTTGGTGGCCGAGTGTGTGACTGCTGGCCAATGAGGCGTAATACGTTAAGTCGGAGCTGCGGAA
 CCGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTTCGCCGCACTTATCGAACGGTTGCCCTGG
 GGAGATCGCTCGCATTTTATCGATAATCGCCGATTACCGGCTGAGCGGCTTAAAGACCCATAAGAA
 ATGCGGCGATGGCGGCTTAGATAAGTAAGTCTCGGGGCGCTCATAAATTTGAGCGCGATCCACGGATT
 TATGCACTCGCTGGAAAAGCTATTACCATTTAGGCTTTTCGCGACCCGATTTTCGCTGCGCTGAGACA
 AGTGCAGCGCGCAGTTGCAAGGCAAAATATGTGTGAGGCAATGCGCGGCGATGCTACACCGAAATCAA
 TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTGGCAGTGAGCGGCCAGCGCTTGTATTGGCAT
 CTAATTTATTCGTTTAAAGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
 TTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTATTAACACACAAAAGGCGCTCTCGCTGT
 CAATGTAATTTGCAAAATGCTCAATCCGCCTAATGATGTGCGCCATGCAATAGTTTTGTGCCAATCATT
 TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATAATGTCATGGCTTAGGCTTTTGGTGGACTTA
 CCAACTGAAGAGACGATTTGTTGGGGTGGCTTTGGCGCAGTGGCGCCTGCGAGCAGGAAATCTCTTTC
 GGCTGTCTGATTTTGGCCAAAGACAATAAATCCGGCTGGCAGATAGGACAGAGGGGACCCGGCGGTGAGG
 CCGTGGACATGAACTTGAACCGCAAGCAGCGCAGCGCCGAAAACATTTGATTCACGAACGGCAAGTGTGCG
 GGATGGGTGCTCTGGCTAAGTTAGCGCGTGGGCAACAGGTTTTCCCGGCCAACACTGGGGGAG
 AAAATAAAAGGAAAATGTTTCCGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGCG
 GGAATGTCACTCGCATTCGCTTGAATTTCCGCTAACTCGCAGCGGCTCTGTGTGTAATAATGCTAA
 TGTTCATGCGGTTGCAATAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT
 GCCATGTTCCGCTGCCACCGTATTTCTTTCCGCGCACTTAGTGCCTCGCTTGATAATGATGTTTGT
 TTCGCGGGGACAAACTCGTTTCGATTATGGGAAAAGCGCTATAAATCATCGCCGCGAAGTCTGGCAAA
 ACAGCAAAATGAAAACCTGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAAACACAGCATCCCA
 CACGACGAGGTGAAAATGAAAATAAATACGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTCGA
 GGTTCACCGGCTGCCGGCATCGCATCATTAAAGTGCCTTCGCTGGATACGCGGCTCTTATGCAACGAGC
 ACACACAATTAATAAAGCGTCTGGTTGTTTCGGCTGGCTTTTGGCGACCTGCCGATCGCAATAAAT
 AAGCAGCATAGTCGCAATTAATGTCACATAGTTGGGCTGCTACTTTTCTGTGGGTGAGC

>E2

CCTTAGTTCAGAAGGCGCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCCTAACGAAATTA
 TCAGACCCGTAACCCAGTTTCGGTAAGTGTCTCTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
 CCAATTTGGTGGCCGAGTGTGTGACTGCTGGCCAATGAGGCGTAATACGTTAAGTCGGAGCTGCGGAA
 CCGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTTCGCCGCACTTATCGAACGGTTGCCCTGG
 GGAGATCGCTCGCATTTTATCGATAATCGCCGATTACCGGCTGAGCGGCTTAAAGACCCATAAGAA
 ATGCGGCGATGGCGGCTTAGATAAGTAAGTCTCGGGGCGCTCATAAATTTGAGCGCGATCCACGGATT
 TATGCACTCGCTGGAAAAGCTATTACCATTTAGGCTTTTCGCGACCCGATTTTTCGCTTGCCTGAGACA
 AGTGCAGCGCGCAGTTGCAAGCAATTTATGTGTGAGGCAATGCGCGGCGATGCTACACCGAAATCAA
 TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTGGCAGTGAGCGGCCAGCGCTTGTATTGGCAT
 CTAATTTATTCGTTTAAAGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
 TTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTATTAACACACAAAAGGCGCTCTCGCTGT
 CAATGTAATTTGCAAAATGCTCAATCCGCCTAATGATGTGCGCCATGCAATAGTTTTGTGCCAATCATT
 TTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATAATGTCATGGCTTAGGCTTTTGGTGGACTTA

17

CCAACTGAAGAAGACGATTGTGGGGTGCCTTTGGCGCAGTGC CGCCTGCGAGCAGGAAATCTCTTTCTC
GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGGTGAGGG
CCGTGGACATTGAACCTGAAAACGCAGCCAGCGCCGAAAACATTGTATTCAACGAAACGGCAAGTGTGCGC
GGCATGGGTGTCCTGCGCTAAGGTACGCGCGTTGGCAACAGGTTTTCCCGCCCAACACTGGGGGGAG
AAAAATAAAGGAAATGTTTCCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGGCCG
GGCAATGTCACTCGGCATTCGCTTGTATTTCCCGCTAACTCGCAGCGGTCTGTGTGTAATAATGCTTAA
TGTTCATGCGCGTTGCATAAATCGTGTGGCAATTATG

>E3

CCTTAGTTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTCTTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
CCAAATTTGGTGCCTCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTGCGCGACTTATCGAACGGTTGCCTTGG
GGAGATCGTGCATTTGTTATCGATAAATCGCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGGCATGGCGCTTAGATAAAGTAAGTCTGCGGGCGCTCATAAAATTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACATTAGGCTTTTCGCGACACCGGATTTTTCGCTTGCCTTGG
AGTGCAGCGCGGAGTTGAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCTACTTATTGCTTAGTTTTTTGCGCAGTGCAGCGCCAGCGCTTGTATTGGCAT
CTAATTTATCCGTTTAAAGGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
TTTTGGCTTGAATAAATAATCAATCGAATTCGCCGCTGGCTATTAACACACAAAAGGCGCTCTGCTGTGTT
CAATGTAATTTGCAAAATGCTCAATCGCCTAATGTATGTGCGCCATGCAATAGTTTTGTGCCAATCAT
TTTTAGTACACCCCTAACTGGTGTCTTCTACGCATAATATGTGCCATGGCTTAGGGCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTGTGGGGTGCCTTTGGCGCAGTGC CGCCTGCGAGC

>E4

CCTTAGTTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTCTTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
CCAAATTTGGTGCCTCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTGCGCGACTTATCGAACGGTTGCCTTGG
GGAGATCGTGCATTTGTTATCGATAAATCGCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGGCATGGCGCTTAGATAAAGTAAGTCTGCGGGCGCTCATAAAATTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACATTAGGCTTTTCGCGACACCGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGGAGTTGAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCTACTTATTGCTTAGTTTTTTGCGCAGTGCAGCGCCAGCGCTTGTATTGGCAT
CTAATTTATCCGTTTAAAGGACGCAATTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
TTTTGGCTTGAATAAATAATCAATCGAATTCGCCGCTGGCTATTAACACACAAAAGGCGCTCTGCTGTGTT
CAATGTAATTTGCAAAATGCTCAATCGCCTAATGTATGTGCGCCATGCAAT

>E5

CCTTAGTTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTCTTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
CCAAATTTGGTGCCTCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTGCGCGACTTATCGAACGGTTGCCTTGG
GGAGATCGTGCATTTGTTATCGATAAATCGCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGGCATGGCGCTTAGATAAAGTAAGTCTGCGGGCGCTCATAAAATTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACATTAGGCTTTTCGCGACACCGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGGAGTTGAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACCTCTATTCTACTTATTGCTTAGTTTTTTGCGCAGTGCAGCGCCAGCGCTTGTATTGGCA

>D2-D11K0

CCTTAGTTTCAGAAGGCGCCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTAA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTCTTTATGAATATTTCCATTTACTTTAATGAAAGGCTG
CCAAATTTGGTGCCTCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTGCGGAA
CGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTGCGCGACTTATCGAACGGTTGCCTTGG
GGAGATCGTGCATTTGTTATCGATAAATCGCCGATTACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATGCGGGCATGGCGCTTAGATAAAGTAAGTCTGCGGGCGCTCATAAAATTCGAGCGCGATCCACGGATT
TATGCACTCGCTGGAAAAGCTATTACATTAGGCTTTTCGCGACACCGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGGAGTTGAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA
AGTGCAGCGCGGAGTTGAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA

TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTGGCGAGTGAGCGGCCAGCGCCTTGATTTGGCAT
 CTcgcgATTCCGTTTAAAGGACGcgcgTTCAGACTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
 TTTTGGCTTGAATAAAcgcgTCGAATTCGCCGCTGGCTcgcgAAACACACAAAAGCGCTCTCGTCTGTTT
 CAATGTAATTTGCAAAATGCTCAATCCGCCTAATTGATGTGCGCCATGCAATAGTTTTGTGCCAATCATT
 TTTAGTACACCCCTAACTGGTGTTCCTACGCATAATATGTGCCATGGCTTAGGGCCTTTTGGTGGACTTA
 CCAACTGAAGAGACGATTTGTGGGGTGGCTTTGGCGCAGTGCAGCGCTTCCGAGCAGGAAATCTCTTTCTC
 GGCTGCTGTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGTCCAGG
 CCGTGGACATTTGAACCTGAAAACGCAGCCAGCGCCGAAAACATTTGTATTCAACGAACGGCAAGTGTGCGC
 GGATGGGTGTCTCGGCTAAGTTACGGCGGTTGGGCAACAGGTTTTTCCCGGCCAACACTGGGGGGAG
 AAAATAAAAAGGAAAATGTTTCCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGCCG
 GGCAATGTCACTCGGCATTCGCTTGATTTTTCCGCCTAACCTCGCAGCGGCTCTGTGTGTAATAATGTCTAA
 TGTGTGATGCGGTTGCAATAATCGTGTGGCAATTATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT
 GCCATGTTCCGCTGCCACCGTATTTCTTTTCCGCCACTTAGTGCCTCCGCTTGATAATGATGTTTTGTTTT
 TTCGCCGGGACAAAACCTCGTTTCGATTTATGGGAAAAGCGGTATAAATCATCGCCGCCGAAGTCTGGCAAA
 ACAGCAAAATGAAAACCTGAAGCTGAAAACCTGAAAACCTGAAACCAAAACAAACAGCATCCCA
 CACGACGAGGTGAAAATGAAAATAAATACGACTGAGCAGTGAACGAGTCAATTCGATTCAAATTCGA
 GGTTCACCGGCTGCCGGCAGTGCATCATTAAAGTGGCCTTCGCTGGATACGGGCTCTTATGCAACGAGC
 ACACACAAATTAATAAAGCGTCTGTGTGTTCCGGCTGGCTTTTGGCGACTGCGGATCGCAATAAAT
 AAGGCAGCATTAGTCCGCAATATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
 GAATGCCGGCCAAAGGATCGAATTAACCCGCTTTTCCGAGGCCAAGAGTTTTTCGCATTTGCATAAAATC
 GGCACCGCATAAAGTGGCGAAGCATTGATGAAAACCTCGGGAAAAGAGTAAAAAATATTTAAAAATAAAT
 AAATTTATGGCAGAATTAAGAAAATAATTTGAATACTTCTTCTTAGGAAAACCTGCTCCAGGAATATTTG
 TTTTCCCGCAGCATTTGCTCAATAATTTCCCTCCATCTTTTGGCTTATTGCCAGACATTTTCTTGGCCGAGT
 GTAGCTGGTGGGCTCCAGATTAATGCAAAACCTTCGTCAGCGGAGGTGTAACAGTATCTTTGCCCAT
 TGGCTCGTTCATATGCGTGTGGTATAGCTTTATTTTTGCCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
 TGGCCAAAGAGTATGCGAAATCGGTCGATTTTCGGGTTTTTCGCACCTGCTTCCGGCCATGGCCATTAG
 AGCATTACCCTTGGGGCGCCCTAAAGTCCAGGTGGTCCCGAGGACCAAGAGTATTGCAACTTACGG
 CCAGCTGAGTGGAGTGTGGAACGCACCTCTTAATTTCCGGCGTATGTAACCTCGAGCTGAGTGTGGAT
 ACATATGCCAAAATCACCTGCTCATAATTAGCGGAAAACCAACTGTTTGGCCCTCGCCGACTGTGAATCAT
 CaGAGCTGCCAATCGAAATCAAGCAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTGTTCCGA
 AAGACCTGTCCAGATACCTGTTTATAGGTATAATTTAAGTGCATATCAGGTTTATTTACATTTATAT
 CGTATTTATTTGGTAACTGCAGCAGATGCTGTGCTACAAAATTTAGAATCATTTAAAAACAAACATATTTGCC
 ACAGAAAATGTTGAAATTAATTAACATAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCTAAAATAAT
 ATTTATGAATAATCAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
 GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAAATGCTTAGCACATAAATCAGCTG
 AATCCTAAACGTATCTGAAGCCAGGAGTGTCCGAGAATTCGGTGTGCCAAAACCAAGACCAAGACCA
 TACCCTTTCAAACCTTATGAAAATGGCAAGCCCGCGAAAAGGTGTGGCCGGTCCAGGGGATTCGGGGG
 CCGGTGATACTCGCACCTTAATAAACATGCGTGAATAATCAATCAGCGAAGCAAAAAGCCACGCACGAGA
 AGCCAAAGTGTCCGAAGTGGCCGATCCAGGGTGACCATATAGACCATAAAGTCCGATGGTGGACCACCA
 CCGAGCCACCGAAAGCAGCCGAAATGCGCGAAACCCGAAAGTGGCGCTTCGTTTTCCGCTTCCATTTGGCC
 TGCTTTCGCTTTCCGAGAAAACCTCATATAAAAACGTTGGCCGACATATTGAGTCCAACAGTCCGTAAGCG
 CGCCACGGTCCACAGAA

>D2- [6]
 CCTTAGTTTCAAGGCGCCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCCTAACGAAATTA
 TCGAGCCCGTAAACCCAGTTTCGGTAAGTGCTCTTTATGAAATTTTCCATTTACTTTAATGAAAGGCTG
 CCAATTTGGTGGCCGAGTGTGTGACTGCTGGCCAATGAGGCGTAATACGTTAAGTCCGAGCTCGGAA
 CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTGCGCCGACTTATCGAACGGTTGCCTTGG
 GGAGATCGCTGCCGATTTTATCGATAATCGCCGATTACCGCGCTGAGCGGCTTAAAGACCCATAAGAA
 ATCGCGGATGGCGCTTTAGATAAGTAAAGTCTGTCGGGCGCTCATAAATTTGAGCGCGATCCACGGATT
 TATGCACTCGCTGGAAAAGCTATTACATTAGGCTTTTCCGCGACACGGATTTTCCGCTTGCCTGAGACA
 AGTGCAGCGCGCAGTTGCAAGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAA
 TTACGGCAACCTCTATTCACTTATTTGCTTAGTTTTTGGCGAGTGAGCGGCCAGCGCCTTGATTTGGCAT
 CTAATTTATTCGTTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGAGATCTAAATTTCCCGC
 ctttactaATAAATTAATCGAATTCGCCGCTGGCTATTAACACACAAAAGCGCTCTCGTCTGTTTT
 CAATGTAATTTGCAAAATGCTCAATCCGCCTAATTGATGTGCGCCATGCAATAGTTTTGTGCCAATCATT
 TTTAGTACACCCCTAACTGGTGTTTTCTACGATAAATAATGTCATGGCTTAGGGCCTTTTGGTGGACTTA
 CCAACTGAAGAGACGATTTGTGGGGTGGCTTTGGCGCAGTGCAGCGCTTCCGAGCAGGAAATCTCTTCTC

GGCCTGTCTGATTTTGGCCAAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGACCCGGCGTTCAGGG
CCGTGGACATTTGAACCTGAAAACGACGCGAGCCGAAAACATTGTATTCACGAACGGCAAGTGTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTTCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACCTCGGCATTCGCTTGATTTTCCGCCCTAACTCGCAGCGGCTCTGTGTGTAATAATGTCTAA
TGTGTCATGCCGTTGCAATAATCGTGTGGCAATATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCCTCCGCTTGATAATGATGTTTTGTTT
TTCCGCCGGGACAACTCGTTTCGATTTATGGGAAAAGCGGTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAAATGAAAACCTGCAAGCTGAAAACGAAAACGAAAACGTAACCCAAAACAAAACACAGCATCCCA
CAGGACGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAATTTGCA
GGTTCAACGGCTGCCGGGATCGCATCATTAAGTGGCCTTTCGCTGGATACGGGCTCTTATGCAACGAGC
ACACACAATTAATAAAGCGTCTGTGTTTTCGGCTGGCTTTTTCGGGACTGCCGATCGCAATAAATTA
AAGGCAGCATTAGTGCATAATTAAGTGCACATAGTTGGCTGCTTACTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCGAGTTAAACCGCTTTTCCGCGAGCCAAAGGTTTTTTCGCATTTTGCATAAAAATC
GGCAACGCATAAGTGGCGAAGCATTTGATGAACTGCGGAAAAGAAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAACTAATTTGAAATACTTCTTCTTAGGAACTGTCCTTAGGAAATTTG
TTTTCCCCAGCATTTGCTCAATAATTTCTCCATCTTTTTCGCTTATTGCCAGACATTTTCTTGGCCGAAGT
GTAGCTGGTGGGTCCTCAGATTAATGCAACCACTTCGTCAGCGGAGGTCGTAACAGTATCTTTGCCCAT
TGGCTCGTCTATTAAGCGTGTGATAGCTTTATTTTGGCATTTCCTCTTTTTCACCAGCTGCAGT
TGGCCAAAGAGTTTATGCGAATCGGTGGATTTTCGGGTTTTTCGCATCGCTTGGCCATGGCCATAG
AGCATTACCCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCAGGACCACAAGAGTATTGCAACTACGG
CCAGCTGAGTGGAGTGTGGAACGCACTCTTAATTTTCGGCGTTTATGTAACCTCGAGCTGAGTGTGGAT
ACATATGCCAAAATCACCTGCTCATAATTAAGCGAAAACCAACTGTTTGGCCCTCGCCGACTGTGAATCAT
CaGAGCTGCCAATCGAAATCAAAGCAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTCTGTCCGA
AAGACCTGTCCAGATATCTGTTTATAGGTATAATTTAAGTGCATATCAGGTTTATTTACATTTATAT
GGATTTATTTGGTAACTGCAGCAGATGCTGTGTACAAATTTAGAAATCATTTAAAACAAAACATATTTGCC
ACAGAAAATGTTGAAAATAAATAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCTAAAATAAT
ATTATGAATAATCAAAGAAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTCAATAAA
GACTCGAAAATACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAAATTTGCCCTAGCACATAAATCAGCTG
AATCCTAAACCTATCTGAAGCCAGGAGTGTCCGAGAAATTCGGTGTGCCAAAACCAAAGACCAAAGCCCA
TACCCTTTCAAACCTTATGAAAATGGCAAGCCCGGCCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACCTCGCACTTAATAAACAATGCGTGAATAACAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGTTGGACCCCA
CCCGAGCCACCGAAAGCAGCCGAAATGCGGAAACCCGAAAGTTGGCGCTTTCGTTTTCTGCTTCCATTTGGCC
TGCTTCTGCTTTTCGGAGAAAACCCATATAAAAACGTTGGCCGACATATGAGTCCAAACAGTCCGTAAGCG
CGCCACGGTCCACAGAA

>D2-block4

CCCTAGTTCAGAAGGCGCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTA
TCGAGCCCGTAAACCCAGTTCGGTAAGTGTCTCTTTATGAATATTTCCATTTACTTTAATTTGAAGGCTG
CCAAATTTGGTGCCTCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTTCGAGCTGCGGAA
CGGGATGATGGACAGTTGAGGCGAAGTACATCAATCTCATTTCGCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCATTTGTTTATCGATAATCGCCGATACCGCGCTGAGCGGCTTAAAGACCCATAAGAA
ATCGGGCGATGGCGCTTTAGATAAGTAAAGTCTCGGGCGCTCATAAATTTTCGAGCCGATCCACGGATTT
TATGCACTCGCTGAAAAGCTATTACCATTAGGCTTTTCGGGACCACGGATTTTTCGCTTGCCTGAGACA
AGTGCAGCGCGGCACTTGCAGGCAAAATATGTGTGAGGCAATGCGCGGGCATGTCTACACCGAAATCAAA
TTACGGCAACTCTATCTACTTATTTGCTTAGTTTTTTCGCGAGTGCCTTTCCCTCGGTGCTGTGTTAA
AGTGGAGAGGATCCGTTTCTGTGGCCAGCATTCGTTAGGATGTATAAATACGGCGCACAGAGCCTTCCA
ACCTCCCACTTCGTTATCATCTGAATTCGCCGCTGGCTATTAACACACAAAAGGCGCTCTCGTCTGTTTT
CAATGTAATTTGCAAAATGCTCAATCGCCTAATTTGATGTGCGCCATGCAATAGTTTTGTGCCAATCATT
TTTAGTACACCCCTAACTGGTGTCTTACGCATAAATATGTCATGGCTTAGGGCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTTGGGGGTGCGTTTGGCGCAGTGCAGCCCTGCGAGCAGGAAATCTCTTTC
GGCTGTCTGATTTTGGCAGACAAATAAATCCGGCTGGCAGATAGGCAGAGGGACCCGGCGTTCAGGG
CCGTGGACATTTGAACTGAAAACGACGCGAGCGCCGAAAACATTTGATTTCAACGAAACGGCAAGTGTGCGC
GGCATGGGTGTCTCTGGCTAAGGTTACGGCGGTTGGGCAACAGGTTTTCCCCCGGCCAACACTGGGGGGAG
AAAATAAAAAGGAAAATGTTTCAGGCTGCCATAAGTGGGAAAAGGAAAACAAAACATGAAACACGGGCCG
GGCAATGTCACTCGCATTCGCTTGTATTTTCCGCCCTAACTCGCAGCGGCTCTGTGTGTAATAATATGCTAA
TGTGTCATGCCGTTGCAATAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT

GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAAACCTCGTTTCGATTTATGGGAAAAGCGCTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATGAAAACCTGCAAGCTGAAAACCTGAAAACCTGAAAACCTGTAACCCAAAACAAACACAGCATCCCA
CAGCAGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAAATTGCA
GGTTCAACGGCTGCCGGCAGTCGCATCATTAAGTGGCCCTTCGCTGGATACGGGGCTTTATGCAACGAGC
ACACACAATTAATAAAGCGTCTGGTTGTTTCGGCCCTGGCTTTTGGCGACCTGCCGATCGCAATAAAT
AAGCAGCATTAGTCGCAATATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGATCGAGTTAAACCGCTTTTCCGAGGCCAAGAGTTTTTCGCATTTTGCATAAAAATC
GGCAACGCATAAAGTGGCGAAGCATTGATGAAAACCTGCGGAAAAGAGTAAAAAATATTTAAAAATAAAT
AAATTTATGGCAGAACTTAAGAAAATAATTTGAAATACCTTCTTCTTAGGAACTGTCCCTAGGAATATTTG
TTTTCCCGCAGCATTTGCTCAATAATTTCTCCATCTTTTGGCTTATTTGCCAGACATTTTCTTGGCCGAAGT
GTAGCTGGTGGGCTCCAGATTAAAGTCAACCCACTTCGTCAGCGGAGGTCGTAACCGTATCTTTGCCCAT
TGGCTCTTCTATATGCGTGTGTATAGCTTATTTTGCATTTTCCCTCTTTTGTCCAGCTGCGAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTGGGTTTTGCGACTCGCTTGGGCCATGGCCATTAG
AGCATTACCCTTAAAGGCGCCATAAGTCCAGTGGTCCCGAGGACCAAGAGATTGCAACTTACGG
CCAGCTGAGTGGAGTGTGGAAACGCACTTCTTAAATTTGGCGGTTATGTAACCTCGAGCTGAGTGGCGAT
ACATATGCCAAAATCACCTGCTCATAATAGCGGAAACCAACTGTTTGGCCCTCGCCGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCAGTCAATCGAAGCCAGGTAATTCATATTTAGTCTGTTCCGA
AAGACTGTCCAGATACCTCTTATAGGTATAATTAAGTGCATATCAGGTTTTATTTACATTTATAT
CGTATTATTTGGTAACTGCAGCAGATGCTGTGCTACAAAATTTAGAATCATTTAAAAACAAACATATTTGCC
ACAGAAAATGTTGTAATAAATTAACATAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCTAAAATAAT
ATTTGAAATAATCAAAGAAAATCAGTAGATGTAAGTACTTTCGTACCTACGTTGCATGGTATTTCAATAAA
AAGCTCGAAAACTACTCTCACTCACTGTAAGTGAACCCAGTGTTTTGTAAATTTGCCCTAGCACATAAATCAGCTG
AATCCTAAACCTATCTGAAGGCCAGGAGTGTCCGAGAAATTCGGTGTGCCAAAAACCAAGACCAAGACCA
TACCCTTTCAAAACCTTATGAAAATGGCAAGCCCGGCCGAAAGGTGTTGGCCGGTCCAGGGGATTCGGGG
CCCGTGAATCTGCCACTTAAATAAACATGCGTGAATAAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACATAAAGTCCGATGTGGACCAACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCGAAAGTTGGCCGCTTCCGTTTTCCGTTCCATTTGGCC
TGCCCTTCTGCTTTCCGAGAAAAAACCCTCATATAAAACCTGGCCGACATATTTAGTCCAACAGTCCGTAAGCG
CGCCACGGTCCACAGAA

>D2-block5

CCTTAGTTTCAGAAGGCGCTGCCTTTATGCGTATTTCCCGCTTGCCTGCGAATACGCCTAACGAAATTA
TCGAGCCCGTAAACCCAGTTTCGGTAAGTGTCTTTTATGAATATTTCCATTTACTTTAATTTGAAGGCTG
CAAATTTGGTGGCCGAGTGTGTGACTGCTGGCCAAATGAGGCGGTAATACGTTAAGTCGGAGCTCGGAA
CGGGATGATGGACCAGTTGAGGCGAAGTACATCAATCTCATTTGCCCGCACTTATCGAACGGTTGCCTTGG
GGAGATCGCTGCGATTTGTTTATCGATAATCGCCGATPACCGCGCTGAGCGGTCTTAAAGACCCATAAGAA
ATCGGGCGATGGCGCTTTAGATAAAGTAAAGTCTCGGGCGCTCATAAAATTTGAGCGCGATCCACGGAT
TATGCACTCGCTGAAAAGCTATTACATTTAGGCTTTTCCGACCACGGATTTTCCGCTTGCCTGAGACA
AGTGCAGCGCGCAGTTGAGGCAAAATATGTTGTGAGGCAATGCCCGGGCATGTCTACCCGAAATCAAA
TTACGGCAACCTCTATTTCACTTATTTGCTTAGTTTTTGGCGAGTGAAGCGCCAGCGCTTGTATTGGCAT
CTAATTTATCCGTTTAAAGGACGCAATTTCTGAGCTAAAACCTGCTTATGGAGAGATCTAAATTTCCCGC
TTTTGGCTTGAATAAATTAATCGACATGAGTCATGTTGCGGATCTTCAAGGTAACCTATACATCATTTCC
ATAATGACTTTGATGACCTCATCGCTTTTAGTCGCCACTTGGAAATAATTTGATAGCAGTGCCAAATCAT
TTTTAGTACACCCCTAACTGGTGTTTTCTACGCATAATATGTGCCATGGCTTAGGGCTTTTGGTGGACTTA
CCAACTGAAGAAGACGATTTGTTGGGGTGGCTTTGGCGCAGTGCAGCGCTTCCGAGCAGGAAATCTCTTTCTC
GGCTGTCTGATTTTGGCCAAGACAAAATAAATCCGGCTGGCAGATAGGCAGAGGGGACCCGGCGTCAAGG
CCGTGGACATTTGAACCTGAAAACGACGCGCAGCGCCGAAAACATTTGATTTCAACGAAACGGCAAGTGTGCGC
GGCATGGGTGTCTCTGGCTAAGGTACGGCGGTTGGCAACAGGTTTTTCCCGCCAACTGGGGGAG
AAAAATAAAAGAAAATGTTTCAAGCTCCATAAGTGGGAAAAGAAAACAAAACATGAAAACACGGGCCG
GGCAATGTCACTCGGCATTTGCTTATTTCCGCTAACTCGCAGCGGCTCTGTGTGTAATAATGTCTAA
TGTGTGATGCCGTTGCAATAATCGTGTGGCAATTTATGCCAGAGAGATTCGCTTATTTATTTTACTTTCT
GCCATGTTCCGCTGCCACCGTATTTCTTTTCGGCCACTTAGTGCGCTCCGCTTGATAATGATGTTTTGTTT
TTCGCCGGGACAAAACCTGTTTCGATTTATGGAAAAGCGCTATAAATCATCGCCGCCGAAGTCTGGCAAA
ACAGCAAATGAAAACCTGCAAGCTGAAAACCTGAAAACCTGTAACCCAAAACAAACACAGCATCCCA
CAGCAGAGGTGAAAATGAAAATAAATACGGACTGAGCGACTGAAAACGAGTCAATTCGATTCAAAATTGCA
GGTTCAACGGCTGCCGGCATCGCATCATTAAGTGGCCCTTCGCTGGATACGGCGCTTTATGCAACGAGC
ACACACAATTAATAAAGCGTCTGGTTGTTTCGGCCCTGGCTTTTGGCGACCTGCCGATCGCAATAAAT

AAGGCAGCATTAGTCGCAATTATGTGCCACATAGTTGGGCTGCTTACTTTTCTGTGGGTGAGCCGAGCGCA
GAATGCGGCCAAGGGATCAGTTAAACCGCTTTTCCGCAGGCCAAGAGTTTTTCGCATTTTGCATAAAAATC
GGCAACGCATAAGTGGCGAAGCATTGATGAAACTGCGGAAAAGAAGTAAAAAATATTTAAAAATAAATAT
AAATTTATGGCAGAACTTAAGAAACTAATTTGAAATACCTTCTTCTTAGGAACTGTCCCTAGGAATATTTG
TTTTCCCGAGCATTTGCTCAATATTTCTCCATCTTTTTGCTTATTGCCAGACATTTTCTTGGCCGAAGT
GTAGCTGGTGGGCTCCAGATTAATGCAAACCACTTCGTCAGCGGAGGTCGTAACGTATCTTTGCCCATTT
TGGCTCGTTCATATGCGTGTGGTATAGCTTTATTTTGCATTTTCCCTCTTTTTTGCACCAGCTGCAGT
TGGGCCAAGAGAGTTATGCGAATCGGTGCGATTTTCGGGTTTTTCGCACTCGCTTGGGCCATGGCCATTAG
AGCATTACCCGCTTAGGGCGCCCTAAAGTCCAGGTGGTCCCAGGGACCACAAGAGTATTGCAACTTACGG
CCAGCTGAGTGGAGTGCTGGAACGCACTCTTAATTTGCGCGTTATGTAACCTCGAGCTGAGTGTCCGAT
ACATATGCCAAAATCACCCTGCTCATAATTAGCGGAAACCAACTGTTTGGCCCTCGCCGGACTGTGAATCAT
CaGAGCTGCCCAATCGAAATCAAAGCCAAGTCAATCGAAGCCAGGTAATTCATATTTAGTGTCTCGCA
AAGACCTGTCCAGATACCTGTTTATAGGTATAATTAAGTGCATATCAGGTTTATTTACATTTATAT
CGTATTTATTTGGTAACTGCAGCAGATGCTGTGCTACAAATTTAGAATCATTTAAAACAAACATATTTGCC
ACAGAAAATGTGTGAAATAATTAACATAAAAGCTTTGGATGAAGTAAAAAGCCATAAAGCCATAAATAAT
ATTATGAATAATCAAAGAAATCAGTAGATGGTAAAGTACTTCGTACCTACGTTGCATGGTATTTCAATAAA
GACTCGAAAATACTCTACTCTACTGTAAGTGAACCCAGTGTTTTGTAAATGCTTAGCACATAAATCAGCTG
AATCCTAAACGTATCTGAAGGCCAGGAGTTCGGAGAATTCGGTGTGCCAAAAACCAAAGACCAAAGACCA
TACCCTTTCAAACCTTATGAAAAATGGCAAGCCCGCGCAAAGGTGTTGGCCGGTCCAGGGGATTCGGGGG
CCCGTGATACTCGCACTTAATAAACATGCGTGAATAATCAATCAGCGAAGACAAAAGCCACGCACTAGAAGA
AGCCAAAGTGTCCGAAGTGGCCGATCCACGGGTGACCATATAGACCATAAAGTCCGCATGGTGGACCACCA
CCCGAGCCACCGAAAGCAGCCGAATGGCCGAAACCCCGAAGTTGGCGCCTTCGTTTTTCGCTTCCATTGGCC
TGCTTTCGTCCTTCGGAGAAAAAACCTCATATAAAACGTGGCCGACATATGAGTCCAACAGTCCGTAAGGC
CGCCACGGTCCACAGAA

Manuscript Two: 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

Accepted by *Science Advances* on 09 October 2020.

This manuscript deciphers the regulatory logic of enhancer structure and function, trying to elucidate how the spatial enhancer activity is encoded in the enhancer sequence. We used the minimal version of the *spot* enhancer, which is 196 bp, as a model. We introduced systematic mutations along the *spot* enhancer sequences, and precisely quantified their effects on the spatial activity of the *spot* enhancer on the wing. Our results reveal a highly density of regulatory information distributed along the *spot* enhancer sequence, and deepen our understanding on how enhancer reads the wing trans-regulatory environment to encode a spatial pattern.

|

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

Authors

Yann Le Poul^{1,†}, Yaqun Xin^{1,†}, Liucong Ling¹, Bettina Mühling¹, Rita Jaenichen¹, David Hörl³, David Bunk³, Hartmann Harz³, Heinrich Leonhardt³, Yingfei Wang⁵, Elena Osipova¹, Mariam Museridze¹, Deepak Dharmadhikari¹, Eamonn Murphy¹, Remo Rohs⁵, Stephan Preibisch^{2,6}, Benjamin Prud'homme^{4*} and Nicolas Gompel^{1*}

Affiliations

¹ Chair of Evolutionary Ecology, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

² Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany.

³ Chair of Human Biology and Bioimaging, Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

⁴ Aix-Marseille Université, CNRS, IBDM, Institut de Biologie du Développement de Marseille, Campus de Luminy Case 907, 13288 Marseille Cedex 9, France.

⁵ Quantitative and Computational Biology, Departments of Biological Sciences, Chemistry, Physics & Astronomy, and Computer Science, University of Southern California, Los Angeles, California 90089, USA.

⁶ Janelia Research Campus, Howard Hughes Medical Institute, VA, Ashburn, USA.

† These authors contributed equally to this work

* correspondence: benjamin.prudhomme@univ-amu.fr; gompel@bio.lmu.de

Abstract

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 carve 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, the affinity and the arrangement of TF binding sites (TFBSs) in the enhancer sequence are relevant to the enhancer spatial activity (reviewed in (5)). Yet, the logic of TFBS organization that determines a spatial pattern is not sufficiently understood to reliably design 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 pattern elements of enhancer activity, but with limited spatial or quantitative resolution (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 4 TFBSs for the activator Distal-less (Dll) and at least one TFBS for the repressor Engrailed (En) (19, 20) (Figure 1A). 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 naive sequence in their native configuration, however, proved insufficient to produce regulatory activity in wings (B. Prud'homme and N. Gompel unpublished results). 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 enhancer sequence to test 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 *D. melanogaster* transgenic lines. Our quantitative analysis revealed a 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 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. Of note, 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 10-18 bp (11.5 bp on average) of poly(dA:dT), also known as A-tracts (22) at adjacent positions along the sequence (Figure 1A). Thus, the sequence of each of the 17 constructs (*spot*¹⁹⁶ [0] to *spot*¹⁹⁶ [16], or [0] to [16] in short, Figure 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 (one 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 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 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 or not the activity driven by a given enhancer sequence significantly differs from any other, wild type or mutant, we used the scores produced by Principal Component 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 (Figure 1B) differs significantly from that of [+], as measured in the PCA space (Figure S1 and Data file S1). This means that the activity of each mutant had some features, more or less pronounced, that significantly differentiates 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 (Figure 1C). This distribution of regulatory information, the magnitude and the direction of the effects, including several successions of over-expressing and under-expressing 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 non-exclusive 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. Such changes in rigidity, which we have evaluated for our mutant series (Figure S2A), may affect TF-TF interactions (Figure 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

To visualize the changes in spatial activity caused by each mutation, we computed the log of the pixel-wise ratio between two average phenotypes (single mutants over *[+]*) at every pixel (30), hereafter noted *logRatio*. As detailed in the supplementary methods, *logRatio* images reveal in which proportion a mutant affects the enhancer activity across the wing. Therefore, *logRatio* images highlight local effects of low activity that would be eluded by stronger activity levels in other areas of the wing. In this respect, *logRatio* images further support our previous conclusion that all the sequences we have mutated affect the activity pattern, one way or another (Figure 2), and therefore that regulatory information is densely packed in the *spot*¹⁹⁶ sequence.

logRatio images also reflect, to some extent, 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 the *spot¹⁹⁶* (19, 20). The disruption of an En binding site (Figure 1A,B, construct [15]) resulted in a proportional increase of activity in the posterior wing compartment (75%, $F(1,124) = 77.8$, $p=8.8818e-15$). The $\log([15]/[+])$ image (Figure 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. Interestingly, contrary to what the average phenotypes suggested (Figure 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 variation in the overall fluorescence signal (Figure 1C), illustrating the power of the *logRatio* analysis to detect local effects of low activity. Mutations that disrupted characterized Dll binding sites (Figure 1A,B, constructs [0], [1], [7] and [9]) resulted in strong reduction in reporter expression (90% $F(1,74) = 143.3$, $p=0$; 75%, $F(1,78) = 109.3$, $p=2.2204e-16$; 47%, $F(1,107) = 75.4$, $p=4.8073e-14$ and 39%, $F(1,74) = 23.2$, $p=7.6363e-06$, 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 (Figure 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 over-expression, is discussed below. Mutations of Dll sites, however, clearly have non-additive 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 non-additivity could be explained by 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 (Figure 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 homeodomain 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 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 of Figure 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% ($F(1,69) = 17.6$, $p=7.8336e-05$) 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 folds ($F(1,72) = 511.5$, $p=0$), 1.9 folds ($F(1,85) = 103.2$, $p=2.2204e-16$) and 2.7 folds ($F(1,82) = 426.5$, $p=0$) compared to [+], respectively). Construct [3] proportionally decreases the expression mostly around wing veins (Figure 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, *D. pseudoobscura*, where no other wing blade activity concealed it (20). Interestingly, 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] vs. [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], [5]) with a global effect on the enhancer activity across the wing is also unexpected.

The analysis revealed another activating stretch of sequence, between 116-137 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 anterior distal wing quadrant. Finally, segments [13], [14], and [15] showed a succession of repressing and activating sites, as we have seen for segments [2] - [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 non-additivity of effects at Dll binding sites, three repressing and four activating novel segments distributed in alternation along the enhancer, and the variety of their effects pointed to a complex regulatory logic, involving more (possibly 6 to 8) 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. Yet, 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 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, 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, as 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 (Figure 3A), and produced constructs where selected blocks were replaced by 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 Figure 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*] (Figure S4A-C). Reciprocally, randomizing block *C* (construct [*AB-*]) results in a uniform collapse of the activity (Figure S4A-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 the *spot*¹⁹⁶, in the context of these reporter constructs.

Block *A* alone ([*A*--]) produces high levels of expression in the veins (Figure S4A-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*-], Figure 3B-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 in fact 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 (Figure 3). Such interactions are best visualized with *logRatios*, comparing this time double-block constructs to single-block constructs used as references (Figure 3D and Figure S4D-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 ($\log([-BC]/[--C])$, Figure 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 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 in fact result from its repressive effect on direct repressors of activity, mostly at the wing tip. This indirect activation

model would explain the non-additivity 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 *spot¹⁹⁶* 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* (Figure 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 (4 out of 5 possible permutations) drive significantly higher levels of expression than the wild type [*ABC*] (*[ACB]*: 2.9 folds (F(1,98) = 191.8, p=0); *[BAC]*: 6 folds (F(1,93) = 589.1, p=0); *[BCA]*: 5.8 fold (F(1,93) = 589.1, p=0); *[CBA]*: 8.4 folds (F(1,93) = 1664.2, p=0); Figure 4B), yet with minor effects on the activity distribution proportionally to the wild type (Figure 4C). We concluded from these experiments that, in terms of pattern, the regulatory output is generally resilient to large-scale rearrangements. As long as all inputs are present in the sequence, the spatial activity is deployed in a similar pattern, yet its quantitative activity is strongly modulated. Because they have little influence on the activity pattern, the rearrangements may not change the nature of the interactions within the enhancer or with the core promoter. Although we would need to challenge this conclusion with additional constructs and blocks with different breakpoints, we speculate that, molecularly, the block randomization perturbs the action of some of the uniformly repressing elements. It highlights the robustness of the enhancer logic to produce a given patterned activity.

Discussion

With this work, we have set to decipher the regulatory logic of an enhancer, *spot¹⁹⁶*. The view point 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 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)). Yet, our study clearly shows that to decipher regulatory logic, and eventually design synthetic enhancers, understanding which regulatory components are sufficient to build an enhancer activity is key.

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 Figure 2 and Figure S3). Indeed, 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 major effect of TF-TF interactions

A-tracts are known to change local conformational properties of DNA. As such, 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. Yet, 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 (Figure 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 *spot¹⁹⁶*. Instead, we think that the effects of A-tract mutations are mostly due to disrupted TFBSs along the enhancer sequence.

The regulatory logic underlying *spot¹⁹⁶* enhancer activity

The main finding of our study is that the *spot¹⁹⁶* enhancer likely integrates 6 to 8 distinct regulatory inputs, with multiple layers of cross-interactions (Figure 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 (Figures 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 *spot¹⁹⁶*. The main levels of *spot¹⁹⁶* 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 (Figure 5A). This activation is in turn globally repressed throughout the wing by an unknown repressor whose action masks that of the global activator (Figure 5B). Upon this 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 (Figure 5C) but the spatial range of this repression is limited to the anterior wing compartment by another repressor acting across the posterior wing compartment (Figure 5D). The former local repression could be mediated by Dll itself, a hypothesis compatible with the non-additive effects of Dll TFBS mutations, while 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 find surprising that shuffling blocks of sequence resulted in dramatic 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, impacting 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 evolutionary steps of the emergence of *spot*¹⁹⁶ perhaps reflect in the regulatory logic of this enhancer. The *spot*¹⁹⁶ element evolved from the co-option of a pre-existing *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 impact 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 *Drosophila 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 ϕ C31-mediated transgenesis (40) and integrated all constructs at the genomic *attP* site *VK00016* (41) on chromosome 2. All transgenic lines were genotyped to ascertain that the enhancer sequence was correct.

Molecular biology. All 196 bp constructs derived from the *D. biarmipes spot*¹⁹⁶ sequence were synthesized *in vitro* by a Biotech company (Integrated DNA Technologies, Coralville, United States, Cat. #121416). Table S1 provides a list of all constructs and their sequences. Each construct was cloned by In-Fusion (Takara, Mountain View, United States) in our pRedSA vector (a custom version of the transformation vector pRed H-Stinger (42) with a 284 bp *attB* site for ϕ C31-mediated transgenesis (40) cloned at the *AvrII* site of pRed H-Stinger). All constructs in Figure 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 Figure 3 were cloned by cutting pRedSA with *BamH I* and *EcoR I*, 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 minutes 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 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 covered with Vectashield (Vector Laboratories, Burlingame, United States). 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 (2 rolls in 100 ml 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 Ti2-Eclipse Nikon microscope equipped with a Nikon 10x plan apochromatic lens (N.A. 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 (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 have 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 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. 14 landmarks placed on vein intersections and end points, and 26 sliding landmarks equally spaced along the veins were placed on bright field images using a semi-automatized pipeline. Landmark coordinates on the image were then used to warp with a deformable model (thin plate spline) bright field and fluorescent images to match the landmarks of an arbitrarily chosen reference wing by the thin plate spline interpolation (43). All wings were then in the same coordinate system, defined by their venation.

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

then subsampled by a factor of 2. We used the 89735 pixels inside the wings as descriptors of the phenotype for all subsequence 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 (Figure 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, grey colors meaning that the two compared phenotypes are equal.

Mutation effect direction and intensity. We proposed to represent the necessity of a stretch of sequence along the enhancer with the activity levels of mutants of this stretch relatively to wild-type (*[+]*) activity. To summarize the overall effect of mutants (overexpression or underexpression), we measured the average level of activity across each wing relatively to that of a 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, as 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 *[+]* were 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 relatively to *[+]* sequence. DNA shape descriptors were computed by the R package DNASHapeR (46). Of note, 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.

Principal component analysis (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 mutant significantly differ from [+], we used comprehensive and unbiased phenotype descriptors provided by principal component analysis (PCA), which removes correlation between pixel intensities and describe the variation in reporter gene expression. PCA was calculated on the matrix regrouping intensities of all pixels for every individual, of dimensions ($n_{\text{individuals}} \times n_{\text{pixels}}$ on the wing). The significance of the difference between two constructs considers the multivariate variation of the phenotypes, and is tested using MANOVA on all 5 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 relatively to the wild-type levels.

DNA rigidity scores. A-tracts are runs of consecutive A/T bp without a TpA step. Stacking interactions and inter-bp 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 sub-sequence. 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.

References

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**, R754-763 (2010).
3. I. S. Peter, E. H. Davidson, *Genomic Control Process: Development and Evolution* (Academic Press, San Diego, United States, ed. 1st, 2015), pp. 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. 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**, (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 encoding of developmental regulatory information may constrain evolvability. *bioRxiv*, 2020.2004.2017.046052 (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**, (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. 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).
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. Mühlhling, R. Jaenichen, E. Osipova, N. Gompel, Enhancer evolutionary co-option through shared chromatin accessibility input. *Proceedings of the National Academy of Sciences*, 202004003 (2020).
22. S. Neidle, *Principles of Nucleic Acid Structure* (Academic Press, 2010), pp. 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-Marcais, 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 e1215 (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**, D111-117 (2011).
32. A. J. Stewart, S. Hannehalli, J. B. Plotkin, Why transcription factor binding sites are ten nucleotides long. *Genetics* **192**, 973-985 (2012).
33. D. N. Arnosti, S. Barolo, M. Levine, S. Small, The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205-214 (1996).
34. E. K. Farley, K. M. Olson, W. Zhang, D. S. Rokhsar, M. S. Levine, Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc Natl Acad Sci U S A* **113**, 6508-6513 (2016).
35. V. Bertrand, C. Hudson, D. Caillol, C. Popovici, P. Lemaire, Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* **115**, 615-627 (2003).
36. D. Thanos, T. Maniatis, Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091-1100 (1995).
37. C. I. Swanson, D. B. Schwimmer, S. Barolo, Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr Biol* **21**, 1186-1196 (2011).
38. C. C. Fowlkes, C. L. Hendriks, S. V. Keranen, G. H. Weber, O. Rubel, M. Y. Huang, S. Chatoor, A. H. DePace, L. Simirenko, C. Henriquez, A. Beaton, R. Weizmann, S.

- Celniker, B. Hamann, D. W. Knowles, M. D. Biggin, M. B. Eisen, J. Malik, A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* **133**, 364-374 (2008).
39. N. Martinez-Abadias, 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. 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).
 41. 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).
 42. 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).
 43. M. F. Hutchinson, Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems* **9**, 385-403 (1995).
 44. C. I. d. I. E. c. (CIE), *Colorimetry* (CIE Central Bureau, Vienna, Austria, ed. 4th edition, 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**, W56-62 (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 NG, SP and RR), the Deutsche Forschungsgemeinschaft (grants INST 86/1783-1 LAGG and GO 2495/5-1 to NG), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013 / ERC Grant Agreement n° 615789 to BP) and the National Institutes of Health (grant R35GM130376 to RR). YX was supported by a fellowship from the China Scholarship Council (fellowship 201506990003). LL was supported by a DFG fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). MM and DD are recipients of fellowships from the German Academic Exchange Service (DAAD). EM was supported by the Amgen Scholar program of the LMU.

Author contributions: YLP: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing—original draft, Visualization; YX: Validation, Investigation, Formal analysis, Data curation; LL: Investigation, Formal analysis; BM: Investigation; RJ: Investigation; DH: Software, Data curation; DB: Software, Data curation; HH: Methodology, Supervision; HL: Supervision; YW: Methodology, Software, Formal analysis; EO: Investigation; MM: Investigation, Formal analysis; DD: Investigation, Formal analysis; EM: Investigation, Formal analysis; RR: Methodology, Supervision, Funding acquisition; SP: Software, Supervision, Funding acquisition; BP: Conceptualization, Writing—original draft, Funding acquisition; NG: Conceptualization, Validation, Writing—original draft, Visualization, Supervision, Project administration, Funding acquisition.

Competing interests: The authors declare no competing interests.

Figures

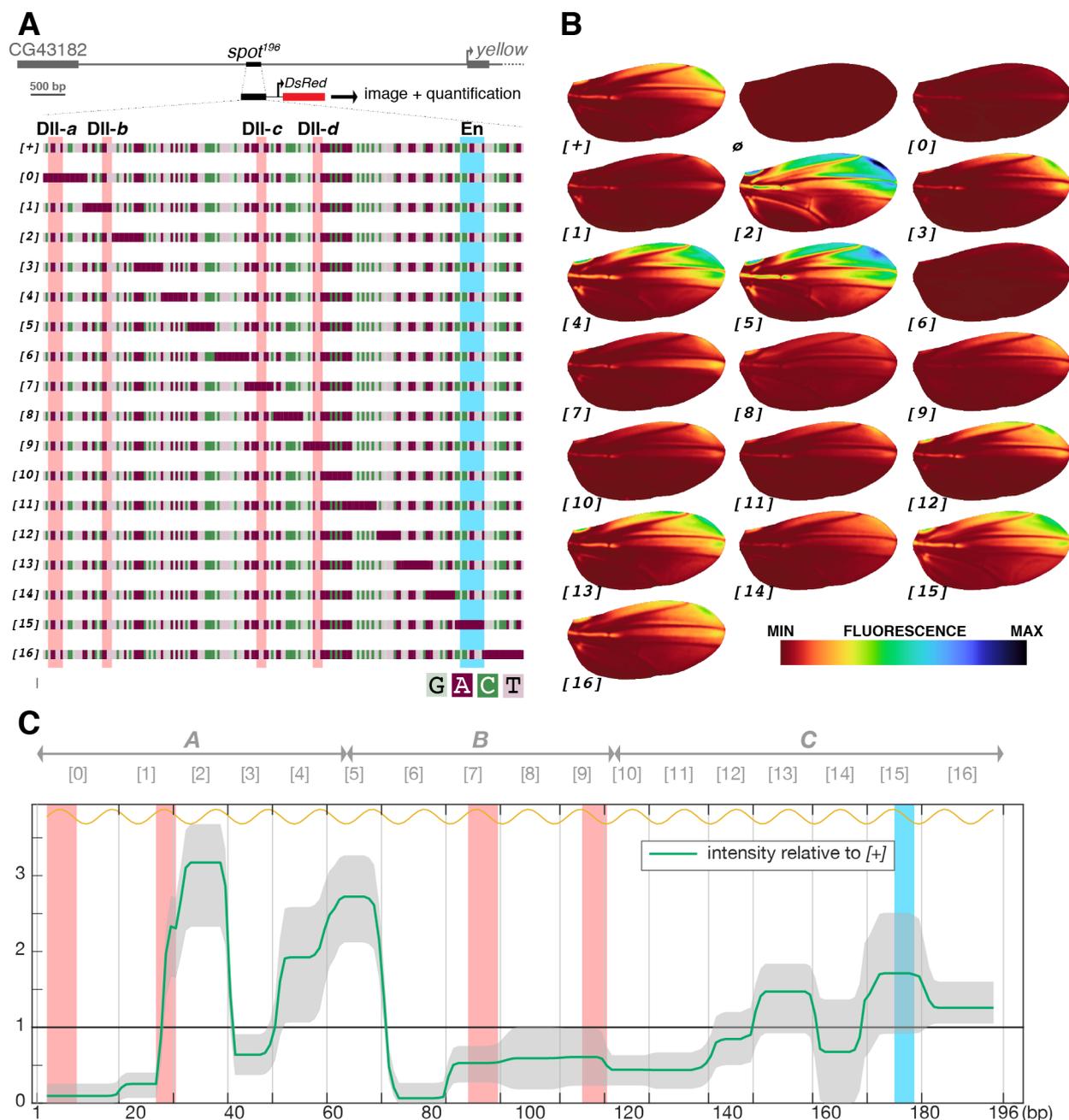


Figure 1. A mutational scan of the *Drosophila biarmipes spot¹⁹⁶* enhancer with a quantitative reporter assay. (A) Wild-type ([+]) and mutant ([0] to [16]) versions of the *spot¹⁹⁶* enhancer from the *D. biarmipes yellow* locus (depicted at the panel top) were cloned upstream of a DsRed reporter to assay their respective activities in transgenic *D. melanogaster*. Each mutant targets a position of the enhancer where the native sequence was replaced by an A-tract (color code: light green=guanine, purple=adenine, dark green=cytosine, pink=thymine). Four characterized binding sites for the TF Distal-less (DII-a, DII-b, DII-c and DII-d) (19) are

highlighted in red and a single binding site for the TF Engrailed (20) is highlighted in blue across all constructs. **(B)** Average wing reporter expression for each construct depicted in **(A)** and an empty reporter vector (\emptyset). Each wing image is produced from 11 to 77 individual wing images (38 on average; Data file S2), aligned onto a unique wing model. The average image is smoothed and intensity levels are indicated by a colormap. **(C)** Mutational effect on intensity of activity along the *spot¹⁹⁶* sequence. The phenotypic effect of each mutation described in **(A)** along the *spot¹⁹⁶* sequence (x-axis) is plotted as the average level of expression across the wing relatively to the wild-type average levels. Shaded grey 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 methods). Mutation positions in constructs [0]-[16] are indicated above the graph. The locations of blocs *A*, *B* and *C*, analyzed in Figure 3 are also indicated above the graph. The yellow curve above the graph indicates the helical phasing.

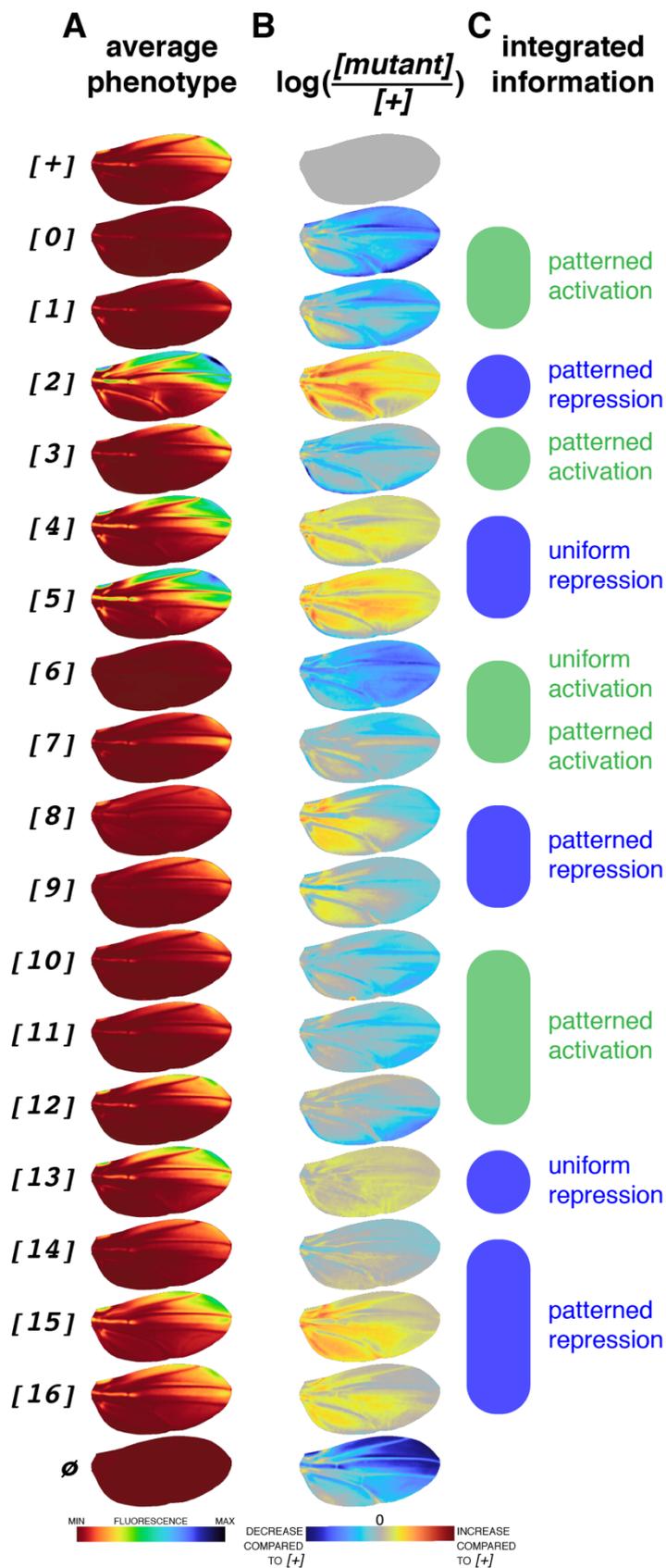


Figure 2. *trans*-regulatory integration along the *spot*¹⁹⁶ sequence. (A) Average phenotypes reproduced from Figure 1B. (B) *logRatio* images ($\log([mutant]/[+])$ for intensity values of each pixel of registered wing images) reveal what spatial information is integrated by each position

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 (Figure S3) in regions of relatively low activity. (C) Summary of spatial information integrated along the enhancer sequence.

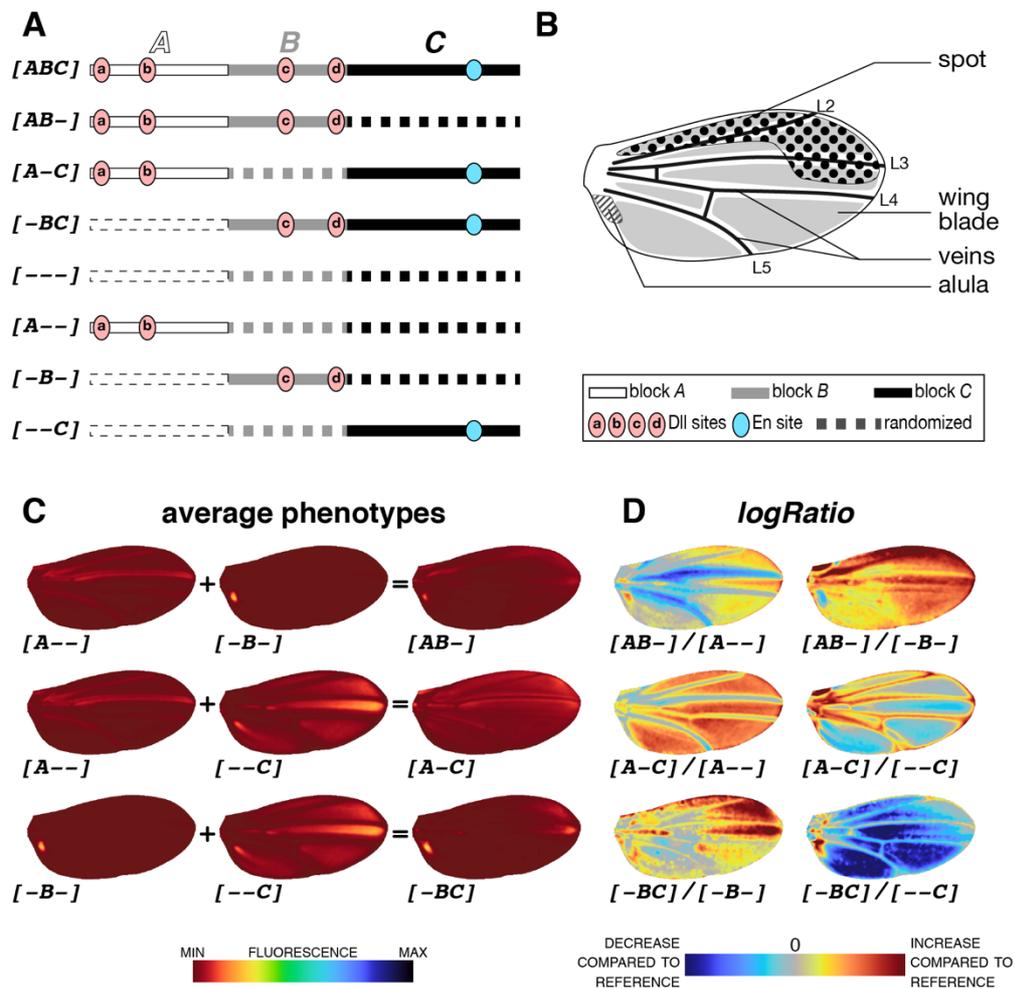


Figure 3. Regulatory interactions in the *spot*¹⁹⁶ sequence. (A) Schematics of constructs with block randomizations. *spot*¹⁹⁶ sequence was arbitrarily divided into 3 blocks (A: 63 bp; B: 54 bp; C: 79 bp). In each construct, the sequence of one, two or all 3 blocks was randomized. (B) Terminology for 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 non-randomized 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 Figure 4B. (D) Block interactions is best visualized with *logRatio* images of constructs 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 Figure S4C-F.

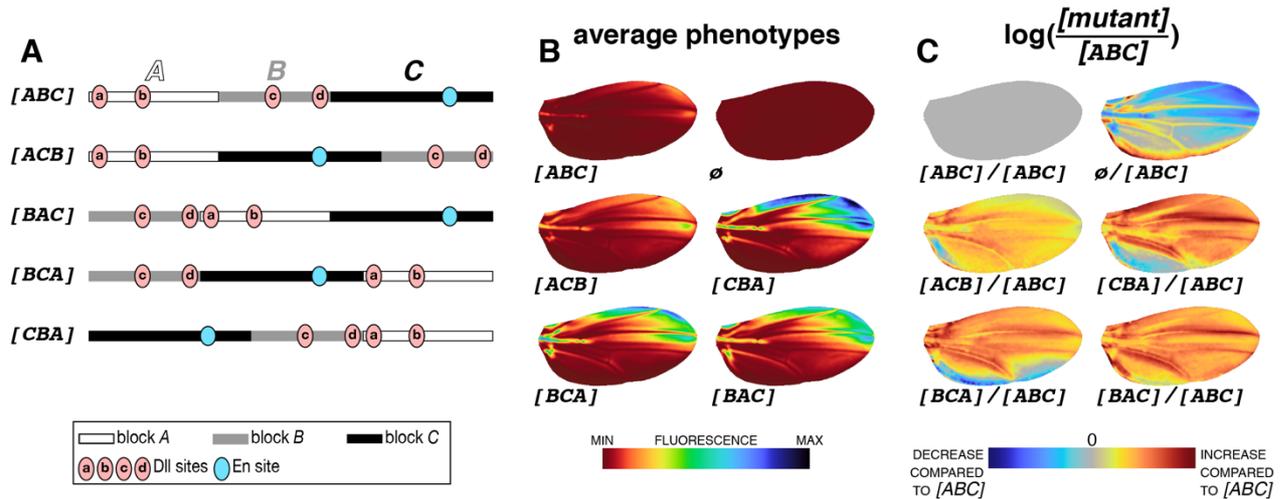


Figure 4. Block permutations scale the activity of the *spot*¹⁹⁶ enhancer. (A) Schematics of constructs with block permutations. In this series, the same blocks of sequences as in Figure 3A were permuted. (B) Average phenotypes resulting from constructs in (A). Colormap of average phenotypes normalized for all constructs of the block series, including block randomizations of Figure 3C and Figure 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 (Figure 3), constructs with block permutations results 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].

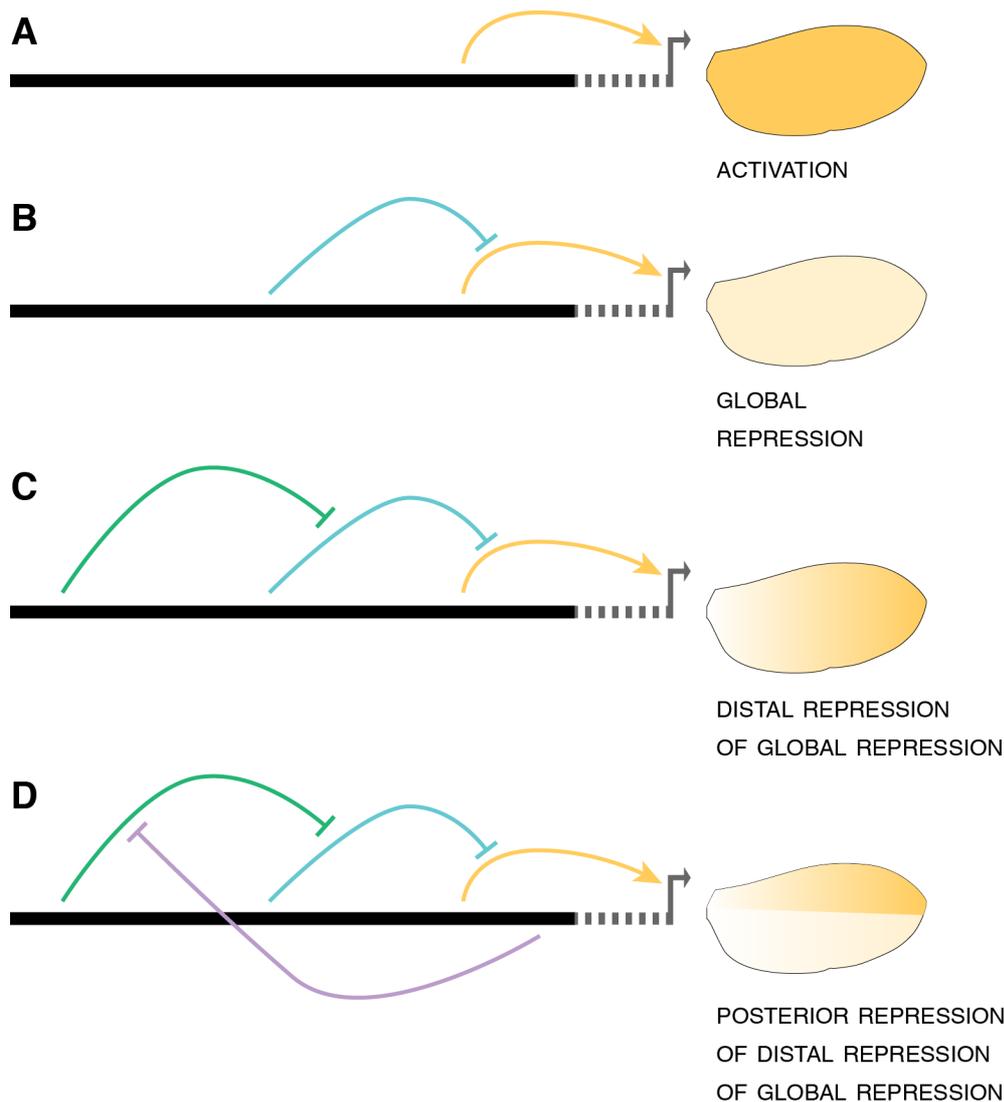


Figure 5. A model of the regulatory logic governing the *spot*¹⁹⁶ enhancer. (A)-(D) The schematics shows step by step how regulatory information and interactions integrated along the enhancer sequence produce a spatial pattern of activity. (A) three independent inputs, respectively in blocks *A*, *B* 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 suppress activity in the wing blade (stemming from blocks *A* and *B*) and the veins (stemming from blocks *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.

H2: Supplementary Materials

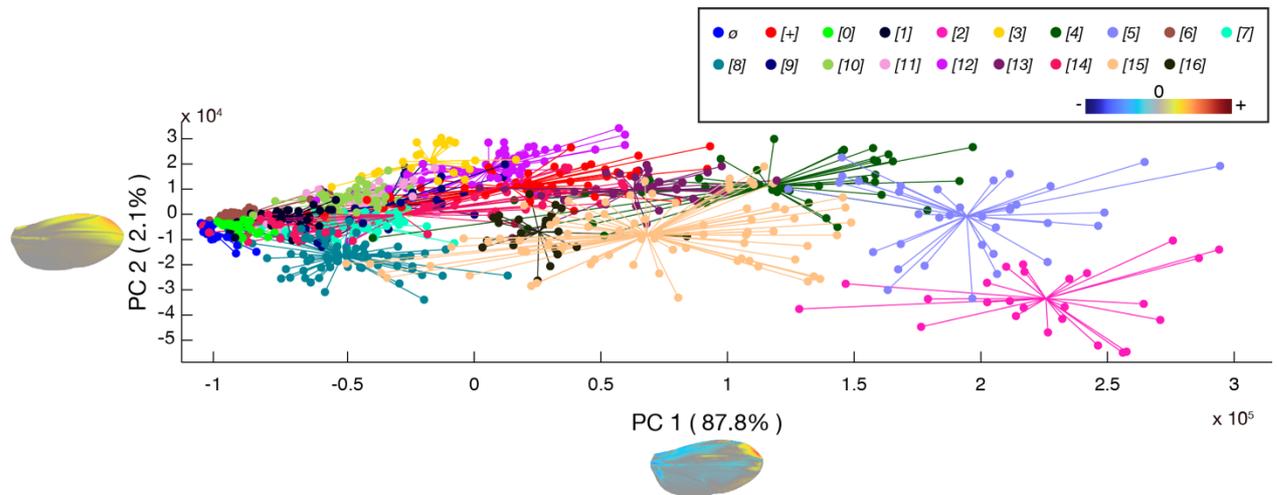


Figure S1. First two axes of variation in a principal component analysis of all individual wings used to generate the average reporter expression of Figure 1. Each wing is depicted by a colored dot, and each construct by a color. PC1 captures 87.8% of the variation and corresponds to overall changes in the activity of the *spot¹⁹⁶* 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.

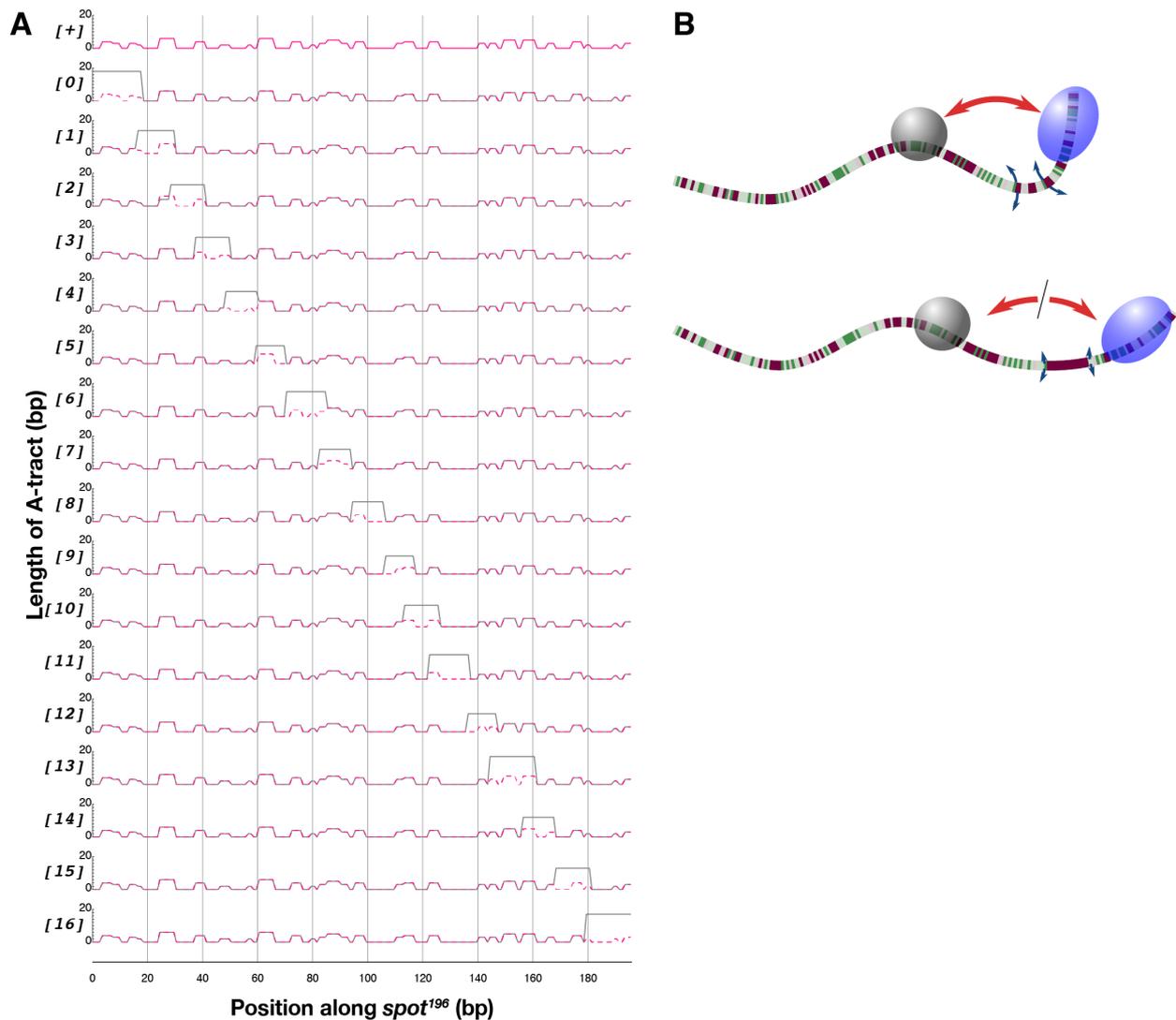


Figure S2. Local rigidity along the wild-type and mutant *spot*¹⁹⁶. (A) Each graph is a plot of the length of the longest consecutive A_nT_n sequence that a base pair participates in, a proxy for sequence rigidity at this position. The first graph on top is the wild type (*[+]*) alone. The remaining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

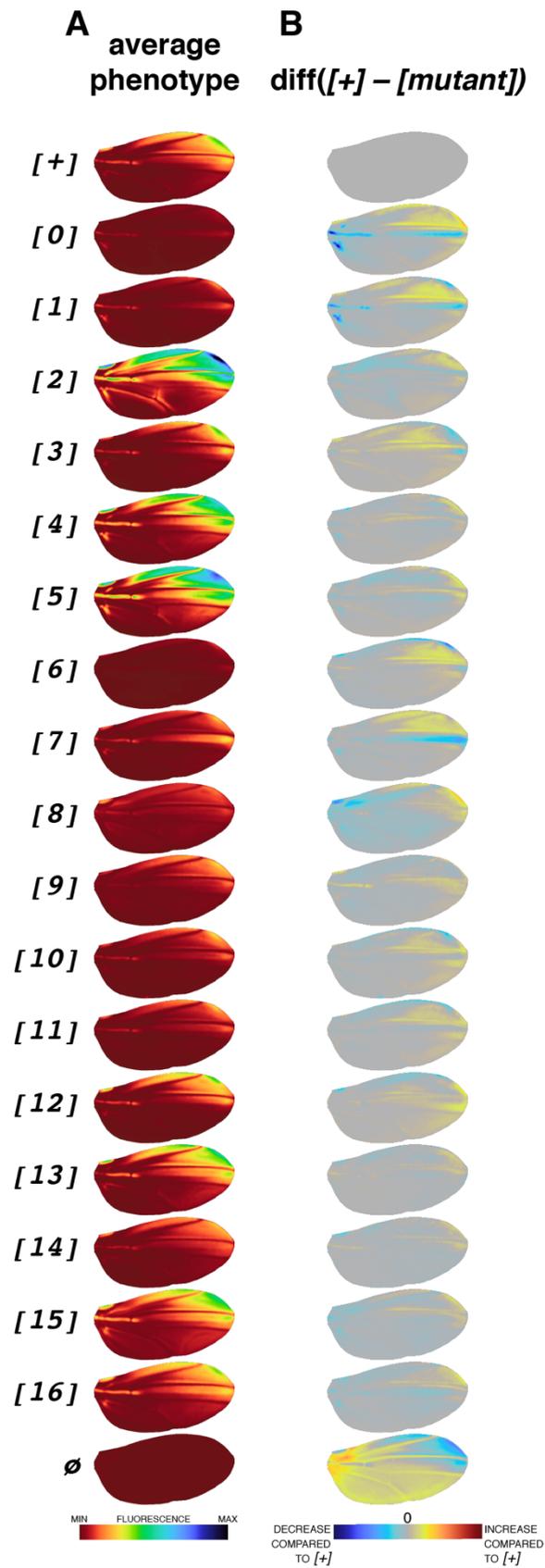


Figure S3. Pattern changes between wild-type and mutant *spot¹⁹⁶* constructs. (A) Average phenotypes reproduced from Figure 1B. (B) difference images ($[+] - [mutant]$) for intensity values of each pixel of registered wing images) highlight changes in the distribution of the enhancer activity across the wing. Note that this operation introduces a visual bias towards changes in region of high expression, contrasting with *logRatio* images of Figure 2.

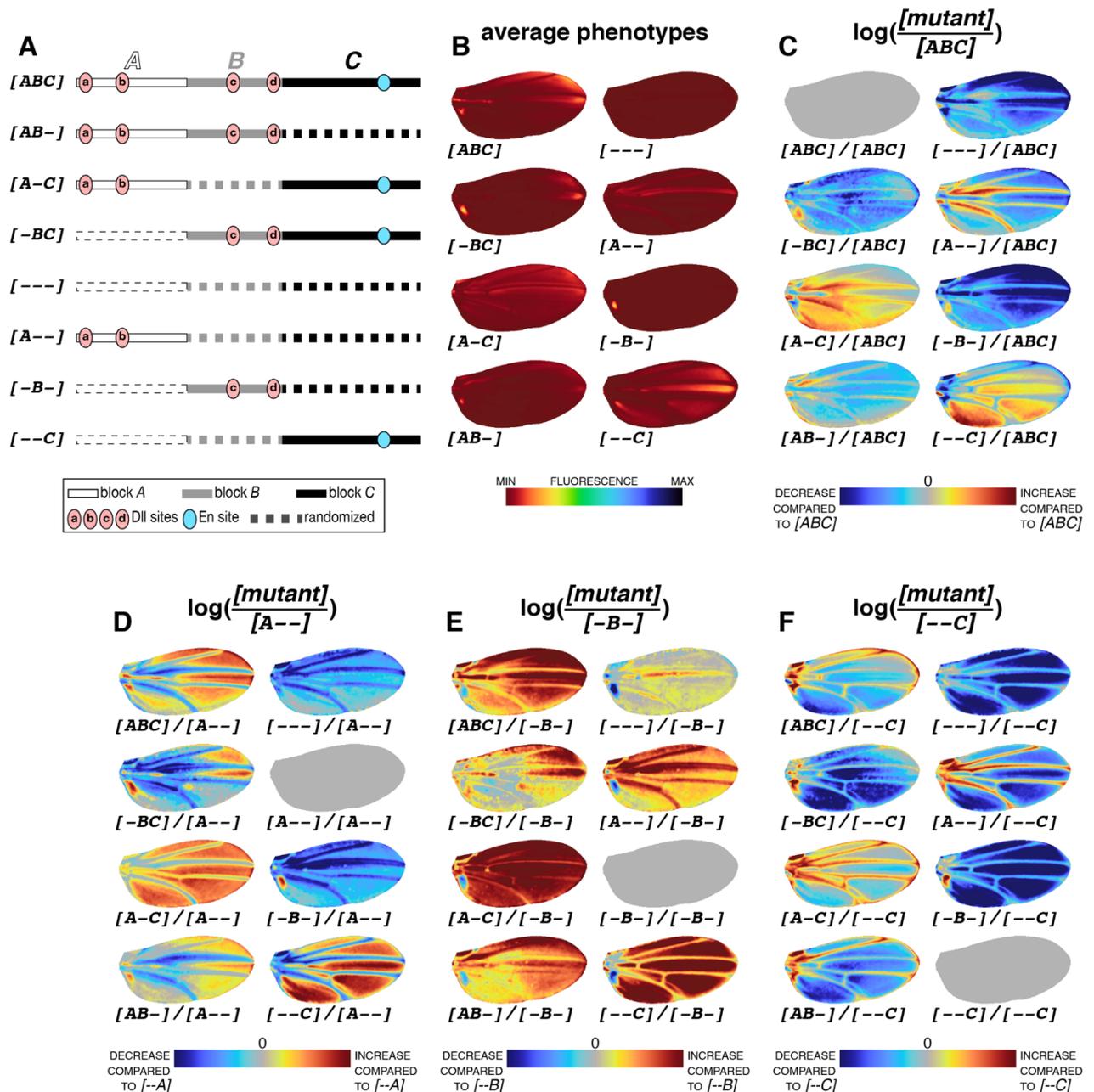


Figure S4. *logRatio* of all block constructs. (A) Schematics of block constructs repeated from Figure 3A for legibility. (B) Average phenotypes of constructs shown in (A), repeated from Figure 3B for legibility. Colormap of average phenotypes normalized for all constructs of the block series, including 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 [+]
TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT
- single mutants [0] to [16]
>*spot*¹⁹⁶ [0]
AAAAAAAAAAAAAAAAAAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAG
AGATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCT
ATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [1]
TCTAATTATTCCGTTTAAAAAAAAAAAAAATTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [2]
TCTAATTATTCCGTTTAAAGGACGCAATTAAAAAAAAAAAAAACTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [3]
TCTAATTATTCCGTTTAAAGGACGCAATTTTCTGAGCTAAAAAAAAAAAAAATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [4]
TCTAATTATTCCGTTTAAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTAAAAAA
AAAATAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCT
ATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [5]
TCTAATTATTCCGTTTAAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCAAAAAAAAAAAAAGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCT
ATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT
- >*spot*¹⁹⁶ [6]
TCTAATTATTCCGTTTAAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCCAAAAAAAAAAAAAAATAAATTAATCGAATTCCCCGCTGGC
TATTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC

CAATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [7]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAAAAAAAAAAAAAGAATTTCCCGCTGGCT
ATTAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [8]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCAAAAAAAAAAAAAGGCT
ATTAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [9]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTAAAA
AAAAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [10]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTA
TTAAAAAAAAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [11]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTA
TTAAACACACAAAAAAAAAAAAAAAATCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [12]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTA
TTAAACACACAAAAGGCGCTCTCGAAAAAAAAAAAAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [13]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTA
TTAAACACACAAAAGGCGCTCTCGTCTGTTTCAAAAAAAAAAAAAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [14]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTTCCCGCTGGCTA
TTAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAAAAAAA
AACCGCCTAATTGATGTGCGCCCATGCAAT

>spot¹⁹⁶ [15]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATAAAAAAAAAAAAAATGTGCGCCCATGCAAT

>spot¹⁹⁶ [16]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGAAAAAAAAAAAAAAAAAAAA

- Permutations of blocks

>spot¹⁹⁶ [ACB]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAACACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCT
CAATCCGCCTAATTGATGTGCGCCCATGCAATTTCCCGCTTTTGGCTTGAATAA
ATTAATCGAATTCCCCGCTGGCTATTA

>spot¹⁹⁶ [BAC]

TTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTA

>spot¹⁹⁶ [BCA]

TTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTATTA

>spot¹⁹⁶ [CBA]

CACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCAATCCGC
CTAATTGATGTGCGCCCATGCAATTTCCCGCTTTTGGCTTGAATAAATTAATCG
AATTCCCCGCTGGCTATTA

- Randomized blocks

>spot¹⁹⁶ [A--]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAATCCGAATTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACC
ACATGTTGTCGACTCAGAAACATTATCCCATTACGCGTAAGCAAAAAATGCGT
CCTTATCGAACTTACACTCGCCTGCGTTGGT

>spot¹⁹⁶ [-B-]

ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAA
ACTGTTGCATGTTTCCCGCTTTTGGCTTGAATAAATTAATCGAATTCCCCGCTGGCTA
TTAAAAGTTCGACTCAGAAACATTATCCCATTACGCGTAAGCAAAAAATGCGTC
CTTATCGAACTTACACTCGCCTGCGTTGGT

>*spot*¹⁹⁶ [-C]

ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAA
ACTGTTGCATGTCCGAATTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACC
ACATGTT**CACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT**

>*spot*¹⁹⁶ [AB-]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAA**TTCCCGCTTTTGGCTTGAATAAATTAATCGAATTC**CCCGCTGGCTA
TTAAAAGT**CGACTCAGAAACATTAT**TCCATTTACGCGTAAGCAAAAAATGCGTC
CTTATCGAACTTACACTCGCCTGCGTTGGT

>*spot*¹⁹⁶ [A-C]

TCTAATTATTCCGTTTAAGGACGCAATTTTCTGAGCTAAAACCTCGCTTATGGAGA
GATCTAAA**TCCGAATTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACC
ACATGTT**CACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTC
AATCCGCCTAATTGATGTGCGCCCATGCAAT

>*spot*¹⁹⁶ [-BC]

ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAA
ACTGTTGCATGTT**TTCCCGCTTTTGGCTTGAATAAATTAATCGAATTC**CCCGCTGGCTA
TTAAA**ACACAAAAGGCGCTCTCGTCTGTTTCAATGTAAATTGCAAATTGCTCA
ATCCGCCTAATTGATGTGCGCCCATGCAAT**

>*spot*¹⁹⁶ [---]

ATAATATTGCATCTCATTGTGGTGCTAGATAATCATCTAGGCTAAATCCAAA
ACTGTTGCATGTCCGAATTTTTCTTGTCCGACTAGAAACGACTAATTTAGCCGTACC
ACATGTTGT**CGACTCAGAAACATTAT**TCCATTTACGCGTAAGCAAAAAATGCGT
CCTTATCGAACTTACACTCGCCTGCGTTGGT

<u>genotype</u>	<u>number of individuals</u>
\emptyset	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
<i>WT</i> -[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 (30), 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.

Paper Three: Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele

Hélène Hinaux, Katharina Bachem, Margherita Battistara, Matteo Rossi, **Yaqun Xin**,
Rita Jaenichen, Yann Le Poul, Laurent Arnoult, Johanna M. Kobler, Ilona C. Grunwald Kadow,
Lisa Rodermund, Benjamin Prud'homme, Nicolas Gompel

Developmental Biology 15 June 2018.

This paper investigates the temporal and developmental process of pigment formation using the pigment gene *yellow* in *Drosophila*. We generated a fluorescent protein-tagged *yellow* allele, and then examined the dynamics of Yellow distribution and cellular targeting in relationship to the process of pigment formation during development. Our analysis resolves the relationship between Yellow expression in space and time, its cellular distribution in the epidermis during development, and its function in pigment formation. In addition, the results showed that Yellow is expressed in a few neurons in the brain and the ventral nerve chord from the second larval instar to adult stage, indicating a neuro-developmental function of *yellow*. Finally, the results suggested a structural role of Yellow in the establishment of pigmentation patterns.



Original research article

Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele



Hélène Hinaux^a, Katharina Bachem^a, Margherita Battistara^a, Matteo Rossi^a, Yaqun Xin^a, Rita Jaenichen^a, Yann Le Poul^a, Laurent Arnoult^b, Johanna M. Kobler^{c,d}, Ilona C. Grunwald Kadow^c, Lisa Rodermund^a, Benjamin Prud'homme^b, Nicolas Gompel^{a,*}

^a Ludwig-Maximilians Universität München, Fakultät für Biologie, Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany

^b Aix-Marseille Université, CNRS, IBDM, Institut de Biologie du Développement de Marseille, Campus de Luminy Case 907, 13288 Marseille Cedex 9, France

^c Technical University of Munich, School of Life Sciences, ZIEL – Institute for Food And Health, Liesel-Beckmann-Str. 4, 85354 Freising, Germany

^d Chemosensory Coding, Max-Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Planegg-Martinsried, Germany

ARTICLE INFO

Keywords:

Pigmentation
Insect
Melanin
Cell trafficking
Live imaging
Pattern boundary

ABSTRACT

Pigmentation is a diverse and ecologically relevant trait in insects. Pigment formation has been studied extensively at the genetic and biochemical levels. The temporality of pigment formation during animal development, however, is more elusive. Here, we examine this temporality, focusing on *yellow*, a gene involved in the formation of black melanin. We generated a protein-tagged *yellow* allele in the fruit fly *Drosophila melanogaster*, which allowed us to precisely describe Yellow expression pattern at the tissue and cellular levels throughout development. We found Yellow expressed in the pupal epidermis in patterns prefiguring black pigmentation. We also found Yellow expressed in a few central neurons from the second larval instar to adult stages, including a subset of neurons adjacent to the clock neurons marked by the gene *Pdf*. We then specifically examined the dynamics of Yellow expression domain and subcellular localization in relationship to pigment formation. In particular, we showed how a late step of re-internalization is regulated by the large low-density lipoprotein receptor-related protein Megalin. Finally we suggest a new function for Yellow in the establishment of sharp pigmentation pattern boundaries, whereby this protein may assume a structural role, anchoring pigment deposits or pigmentation enzymes in the cuticle.

1. Introduction

Closely related animal species with a shared body plan often look strikingly dissimilar because of diverging coloration patterns. In insects, the diversification of pigmentation patterns among closely related species reaches heights, for instance in butterflies or beetles, which exploit the riches of their colorful motifs under various selection regimes (sexual selection, crypsis, predator intimidation) (Edwards et al., 2007; Kronforst and Papa, 2015; Wilson et al., 2015).

Probably because of this prevalent role of pigmentation in species morphological diversification, researchers have sought to understand how pigmentation patterns are physically built during animal development. In insects, the research on pigmentation has taken different routes over the last century. On one hand, geneticists have isolated plethora of mutants with pigmentation defects in *Drosophila* (*yellow* (Brehme, 1941; Morgan and Bridges, 1916); *ebony* (Bridges and Morgan, 1923); *black* (Bridges and Morgan, 1919; Lindsley and

Grell, 1968) *ple* (Budnik and White, 1987; Jurgens et al., 1984) or *tan* (Brehme, 1941). These are particularly well represented in genetic screens, as they are easily seen and often viable under laboratory conditions. On the other hand, biochemists have deciphered the enzymatic pathways leading to pigment deposits and their intermediate metabolites. Pigments are precipitates embedded in the insect cuticle, an extra-cellular matrix composed of lipids, proteins, chitin and catecholamines, and their formation results from a complex biochemical conversion (Locke, 2001; Massey and Wittkopp, 2016; Moussian, 2010; Sugumaran and Barek, 2016; Wright, 1987). Attempts to superimpose these two layers, a biochemical pathway and a genetic network, have reached mixed results, and the function of several genes with specific pigmentation phenotypes remains unknown.

While a more complete picture of the correspondence between genes and intermediate metabolites would help understand better how pigments are made, at least two other dimensions await documentation. First, the production of pigments is a cellular process and it is

* Corresponding author.

E-mail address: gompel@bio.lmu.de (N. Gompel).

<https://doi.org/10.1016/j.ydbio.2018.04.003>

Received 7 February 2018; Received in revised form 28 March 2018; Accepted 6 April 2018

Available online 07 April 2018

0012-1606/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

necessary to understand where the genetic and the biochemical networks are active in a cell. Precursors circulate in the insect hemolymph, are internalized into cells, partially processed in their cytoplasm, and secreted into the forming cuticle where they are converted into pigment deposit (True et al., 1999). Where, in this general framework, are the different gene products at work? Second, the production of pigments is a developmental process, and the temporal dynamic of this process has been largely overlooked at the expense of the spatial determinants of pigment distribution (Gompel and Carroll, 2003; Wittkopp et al., 2002a). This has started to change with the developmental survey of gene expression (Sobala and Adler, 2016). RNA-seq from *D. melanogaster* pupal wings highlights temporal differences in pigmentation gene expression, shedding a new light on the dynamic of the biochemical pathways: *ebony* and *black* are expressed at high levels at the end of pupal life (96 h after puparium formation, h APF) while *yellow* mRNA levels peak at around 52 h APF (Sobala and Adler, 2016). Pigmentation itself appears in the wing blade only around 80 h APF, in the deep layers of the procuticle (Sobala and Adler, 2016) and its formation is thus intricately linked to cuticle deposition.

In an attempt to integrate different dimensions of pigment formation, we are here revisiting the developmental role of the gene *yellow*, a gene necessary – but not sufficient – for the production of black pigments in *Drosophila*. The molecular function of Yellow is unknown (Drapeau et al., 2003; Li and Christensen, 2011). It is a secreted protein (Komezos and Chia, 1992; Wittkopp et al., 2002a). In pupal wings, it is apparently deposited in the distal procuticle, and internalized when the proximal procuticle is secreted (Riedel et al., 2011). Its expression correlates with black melanin patterns (Drapeau et al., 2003; Riedel et al., 2011; Walter et al., 1991; Wittkopp et al., 2002a), and its function is necessary for the production of black melanin (Lindsley and Grell, 1968; Nash, 1976) although it does not appear to function as an enzyme (Wright, 1987).

To survey the dynamic of Yellow protein in time and at the subcellular level in living animals, we have created a fluorescently tagged allele to produce a Yellow::mCherry fusion protein. We have used this functional allele to follow the subcellular localization of Yellow in genetic experiments aimed at interfering with trafficking at the cellular membrane. Our results indicate that Yellow is expressed in a very precise spatio-temporal pattern during *Drosophila* pupal life, shortly preceding the onset of black pigment accumulation. The protein is targeted to the extracellular compartment from the onset of its production and during most of the pupal life. However, during the last few hours of pupal life, some amount of the Yellow protein is internalized and accumulates in the cytoplasm.

2. Results

2.1. Revisiting Yellow expression pattern using a *D. melanogaster* $y^{mCherry}$ line

To track Yellow expression and localization in all tissues of living flies, we first generated a *D. melanogaster* *yellow* allele tagged with the mCherry fluorescent protein gene (Shaner et al., 2004), using the CRISPR-Cas9 technology. In brief, we created the *yellow::mCherry* fusion allele (later referred to as $y^{mCherry}$) by repairing a CRISPR-mediated double-stranded break in *yellow* exon 2 with a template containing the custom in-frame fusion (Fig. 1A, Fig. S1A, Text S1, Text S2, details in Material and Methods). Flies with a modified *yellow* locus were initially screened by PCR (Fig. S1D). Using Sanger sequencing of a portion of the modified *yellow* locus, encompassing the fragment inserted by CRISPR-mediated homologous recombination, we determined that the allele was in line with our design. We later confirmed *mCherry* integration at the *yellow* locus by genomic DNA sequencing of the *yellow::mCherry* line at a 11 × coverage (Fig. S1F–G). The sequence coverage was however insufficient and the size of the library

too small to univocally prove that the insertion was unique.

To evaluate the fidelity of the $y^{mCherry}$ reporter line, we first compared its expression in pupal wings throughout development by Western blot to that of Yellow in wild-type flies. In Canton-S, Yellow was detected in pupal wings from 54 h APF onwards as a 60 kDa band using a polyclonal antibody. A faint band could sometimes be seen at 46 h APF, suggesting that the onset of Yellow expression might be around 46 h APF. Likewise, we revealed a 100 kDa band in $y^{mCherry}$ wing extracts using the same anti-Yellow antibody. Its expression follows a similar temporal dynamic as wild-type Yellow (Fig. 1B). In both wild-type and $y^{mCherry}$ extracts, the highest protein levels were reached at 62 or 70 h APF, while Yellow was barely detectable at 90 h APF. These results were consistent with our expectations of size, and with published work (Walter et al., 1991; Wittkopp et al., 2002a). We also confirmed this expression dynamic in $y^{mCherry}$ pupal wings using an mCherry antibody (not shown). We concluded that our $y^{mCherry}$ allele reports Yellow expression dynamic accurately.

We then used the $y^{mCherry}$ allele to survey Yellow expression in wholemount flies. The fluorescence pattern proved to be consistent with Yellow expression described in fixed tissue (Drapeau et al., 2003; Riedel et al., 2011; Walter et al., 1991; Wittkopp et al., 2002a). It was also consistent with reporter construct expression under the control of *yellow* regulatory regions (Gompel et al., 2005; Jeong et al., 2006; Wittkopp et al., 2002b). In pupae (Fig. 1C–F), we detected fluorescence in transversal stripes prefiguring the abdominal banding pigmentation pattern, on the dorsal thorax in 3 longitudinal stripes (trident) and at the basis of each bristle, including in the male sex combs (inset in Fig. 1E), as well as in the mouthparts. While the fluorescence is eluded by the accumulation of pigments (see below), we noticed that it persists, at least in the wings, until adulthood (Fig. 1G,H), in line with timing of pigment precursor conversion.

In addition to the epidermal expression, we surveyed Yellow::mCherry expression in the brain at different stages, as a few reports invoke its function in neurons (Bastock, 1956; Drapeau et al., 2003; Radovic et al., 2002). We identified a small number of cells that express Yellow::mCherry in the larval brain and ventral nerve chord, at least from the second larval instar (Fig. 2A–E). A similar expression in L3 brains was already reported (Drapeau et al., 2006). We also found a similar scattered pattern in the adult brain, where the expression appears to be confined to a few cells adjacent to the optic lobe lamina, and another few ventral to the suboesophageal ganglion. We did not find positive cells in the ventral nerve chord. The expression in the adult brain was reminiscent of Pigment-dispersing factor (Pdf) expression in clock neurons (Helfrich-Forster and Homberg, 1993). We re-evaluated this expression in flies combining $y^{mCherry}$, *Pdf-Gal4* and *UAS-GFP* (Park et al., 2000), and found that the clock neurons are directly neighboring the Yellow::mCherry expressing cells (Fig. 2F, G). At all stages, in the central nervous system, Yellow::mCherry expression was strong in the soma, but was also occasionally visible in cellular extensions resembling neurites. This discrete spatio-temporal pattern of Yellow expression in the brain contrasts sharply with earlier reports of a widespread expression of cytoplasmic Yellow across the 3rd instar larval brain, with upregulation in cells expressing the male forms of Fruitless proteins (*Fru^M*) (Drapeau et al., 2003; Radovic et al., 2002).

2.2. $y^{mCherry}$ functions normally to produce wild-type pigmentation

A fusion protein may alter the normal function of a gene, for instance by destabilizing the tertiary protein structure. To evaluate the consequences of the tagged allele on *yellow*'s function, we have quantified its effect on pigmentation. $y^{mCherry}$ adult flies are superficially undistinguished from the wild-type parental line used to generate this allele (Fig. 1I). Yet, pigmentation is a quantitative trait, and subtle differences in pigmentation levels may pass unnoticed. To evaluate the functionality of our $y^{mCherry}$ allele for black pigmentation, we compared wings of age-matched adult flies with different genotypes (Fig. 1J), and measured the

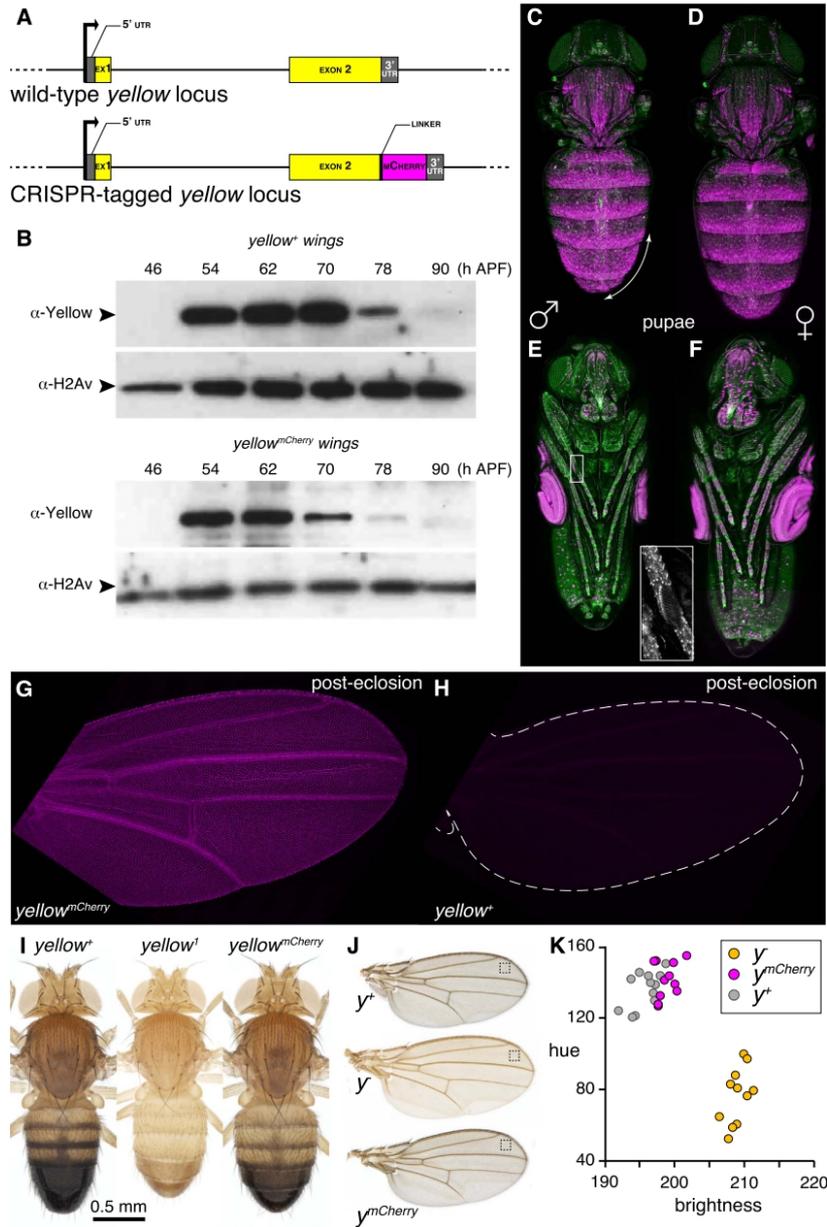


Fig. 1. A fluorescently tagged *yellow* allele. (A) Map of the *yellow::mCherry* locus compared to the wild-type *yellow* locus. *mCherry* was inserted by CRISPR/Cas9-mediated homologous recombination in frame between *yellow* exon 2 CDS and 3' UTR. A short linker sequence (Waldo et al., 1999) was added in hope to preserve the Yellow function. (B) Western blots of pupal wing protein extracts, probed with a Yellow antibody. Top: extracts from Canton-S (*y*⁺) flies; bottom: extracts from *y*^{mCherry} flies. Loading control: H2Av antibody (full gels on Fig. S1E). Of note, we detected higher molecular weight products with the Yellow antibody in pupal wings older than 62 h APF, both in Canton-S and in *y*^{mCherry} flies (Fig. S1E). These may represent crosslinked Yellow to proteins of the maturing cuticle. (C,D,E,F) Confocal images of whole 72–74 h APF *y*^{mCherry}, utrophin-GFP male (C,E) and female (D,F) pupae, mounted dorsally (C,D) or ventrally (E,F). Anterior is up. The arrow highlights the male specific expression of Yellow throughout abdominal epidermis in segments A5–A6 (C). Note the expression in male sex combs (boxed region in (E) and higher magnification in inset). (G,H) Post-eclosion wings of *y*^{mCherry} (I) or *y*⁺ (J) females, imaged under identical

confocal settings (4 wings were examined for each genotype and had identical signal to those shown). (I) $y^{mCherry}$ flies are normally pigmented. Dorsal views of 5 day-old males of the genotypes y^+ (left), y^1 (middle) and $y^{mCherry}$ (right) showing no difference in pigmentation color or intensity between $yellow^+$ and $y^{mCherry}$. (J) Representative wings of 5-day old males and (K) quantification of pigmentation in similar samples of the genotypes y^+ (gray circles, $n = 12$), y^1 (yellow circles, $n = 11$) and $y^{mCherry}$ (red circles, $n = 11$). Pigmentation differences were analyzed by plotting brightness against hue measured in the distal part of each wing, between the veins L2 and L3 (squares on the left panel indicate the region that was analyzed) (see methods).

levels of pigmentation (I J,K). To this end, we have used wing color images acquired under identical conditions, encoded in an HSB color space (Joblove and Greenberg, 1978), and have plotted the hue against the brightness for each wing. This analysis confirmed that pigmentation levels and tone are identical in the wild type and in $y^{mCherry}$ flies, but strikingly different from that of $yellow$ null mutants (Fig. 1K). Similarly, our quantitative analysis of abdominal pigmentation between wild type and $y^{mCherry}$ flies (not shown) revealed no difference. These results demonstrate that the Yellow::mCherry fusion protein is functional to produce normal pigmentation.

In conclusion, we have generated a functional allele of $yellow$ that reports with accuracy the localization in time and space of the endogenous gene product in live animals.

2.3. Live dynamics of Yellow expression in the developing pupal wings

Although its biochemical function remains elusive, Yellow was shown to be secreted by epidermal cells (Kornezos and Chia, 1992; Riedel et al., 2011; Walter et al., 1991; Wittkopp et al., 2002a) and embedded in the cuticle. Adult cuticle is a complex layered structure, produced through the tightly regulated expression of many genes through development (Sobala and Adler, 2016). As $yellow$ confers the specific dark color to the cuticle, we wondered how its developmental expression is coordinated with cuticle deposition during development. We first examined the distribution of Yellow::mCherry during development at the tissue level (this section), and then at the

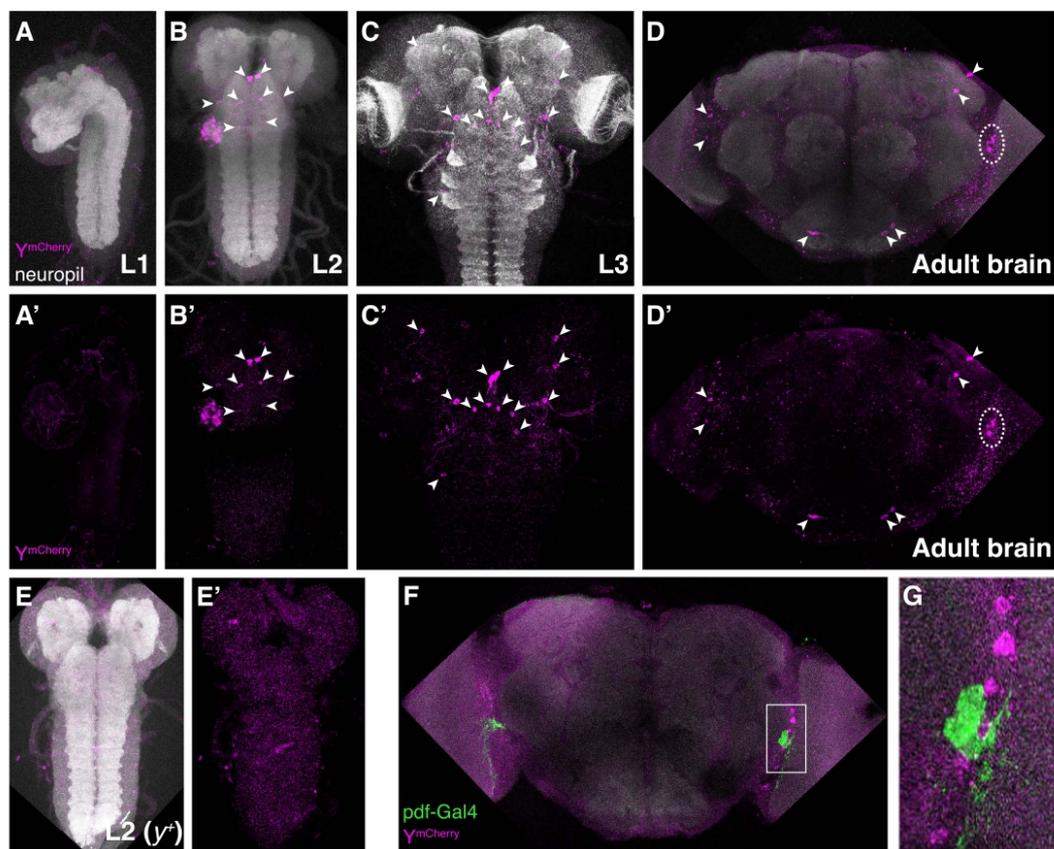


Fig. 2. Yellow is expressed in the central nervous system of *Drosophila*. Confocal projection images of larval L1 (A, A'), L2 (B, B', E, E'), L3 (C, C') and adult (D, D', F, G) brain stacks. (A–D, A'–D') Brains of $y^{mCherry}$ flies stained with an anti-N-Cadherin antibody (labeling the neuropil; shown in white) and anti-DsRed (labeling Yellow::mCherry; shown in magenta) antibody. The top line shows the merged images, the bottom line the DsRed channel alone. Arrowheads and dotted lines point to areas of Yellow::mCherry expression. (E, E') Brain of a wild-type L2 larva stained as in (B, B') confirming the specificity of the stainings in (A–D, A'–D'). (F, G) a partial brain stack projection showing clusters of Yellow::mCherry cells in the vicinity of Pdf-expressing cells (marked by Pdf-Gal4 > UAS-GFP). (G) is an high-resolution view of the region boxed in (F).

subcellular level (next section). For this and subsequent analyses, we concentrated on wings. In this tissue, it is relatively simple to follow the developmental process, from the establishment of a pigmentation gene's blueprint, to the differentiation of the actual pigments in the acellular adult wing.

We initially recorded male *y^{mCherry}* pupae with live time-lapse imaging from 44 to 73 h APF (Fig. 3A). We then quantified the fluorescent signal in the wings (Fig. 3B–C). For all individuals (n = 3), fluorescence appeared around 50 h APF, increased rapidly until 56–59 h APF, and plateaued until 66–68 h APF. This is in agreement with the expression dynamics deduced from Western blots (Fig. 1B) and transcriptomic analysis indicating that the onset of *yellow* transcription in the wing is between 42 and 52 h APF (Sobala and Adler, 2016). Fluorescence then decreased abruptly until it was no longer visible under our imaging conditions, around 72–73 h APF. Our Western blots experiments, though, detected Yellow in the wings, albeit at lower levels, at 78 and 90 h APF (Fig. 1B). Additionally, we had detected fluorescence in *y^{mCherry}* wings after eclosion using confocal imaging (Fig. 1G). Therefore, the fading of fluorescence from pupae was unexpected. We noted that in the time-lapse movies, the decrease in fluorescence occurred simultaneously in all pupal tissues (Movie S1), and correlated precisely with the accumulation of pigmentation (Fig. 3D–E). This suggested that Yellow::mCherry protein was still present at later pupal stages, but that the fluorescent signal was masked by pigmentation.

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.ydbio.2018.04.003>.

In the abdominal epidermis, Yellow expression prefigures the adult banding pattern of pigmentation. From a broad, fuzzy domain in each segment, the expression refines over time to a sharp transversal band. This refinement is not synchronous across segments, but instead follows a temporal sequence from anterior to posterior segments completed after 75 h APF (Wittkopp et al., 2002a). We examined the spatial dynamic of expression in the pupal abdomen of *y^{mCherry}* animals at 65, 70 and 75 h APF, focusing segments A3 and A4. The expression was already sharp at 65 h APF (Fig. S2A). The quantification of fluorescence intensity profiles along the segments, however, did not reveal any changes between the different stages for the A3 segment (Fig. S2C, E). For segment A4, we tentatively observed a very subtle refinement of the anterior boundary (Fig. S2B, D).

2.4. Developmental dynamics of Yellow subcellular localization

To understand how Yellow is produced in relationship to cuticle deposition, we compared the distribution of Yellow::mCherry to that of other markers in developing pupal wings: an mCD8::GFP fusion protein (Lee and Luo, 1999) to outline the cytoplasm, and an indicator of chitin production (Fig. 4). It is indeed possible to directly monitor cuticle localization in transgenic flies expressing the chitin reporter *ChitVis-Tomato* (Sobala et al., 2015, 2016). We could, however, not directly compare *ChitVis-Tomato* to Yellow::mCherry distribution in the same cells, as these fluorescent reporters have overlapping emission spectra. Instead, we compared their respective distributions to that of mCD8::GFP (*mCD8::GFP*, *ChitVis-Tomato* on one hand, and *mCD8::GFP*, *y^{mCherry}* on the other hand) to infer where Yellow localizes during cuticle deposition (Fig. 4, Fig. S3) in wing cells. Yellow is not detected at 46 h APF (Fig. 4D), a stage at which chitin is already present in wing hairs (trichomes) (Fig. 4A). Yellow then accumulates in the wing blade trichomes shortly after the onset of its expression (54 h APF, Fig. 4E, Fig. S3H). Also at 54 h APF, we detect very faint Yellow::mCherry signal at the apical outline of the cells (Fig. S3S, T). This is the stage at which the envelope and the epicuticle have been deposited on the pupal wing (Sobala and Adler, 2016). At this stage, chitin still appears limited to the trichomes (Fig. 4B). Yellow::mCherry signal decreases in the trichomes between 70 and 78 h APF (Fig. 4E–F, Fig. S3J–K), possibly again because of the

accumulating dark pigments (see Fig. 3D–E). Chitin is also visible, lining-up the cell contours, in addition to its presence in trichomes from 62 h APF onwards (Fig. 4C, Fig. S3C–F). At the same stage, Yellow::mCherry signal distinctly outlines cells at the apex (Fig. S3U, V), suggesting that Yellow is incorporated in the cuticle shortly after, or together with chitin. Nevertheless, at 78 h APF and even more so at 90 h APF (Fig. 5F, H, Fig. S3K, L), Yellow::mCherry accumulates into the cytoplasm. This cytoplasmic signal could in principle result either from newly expressed Yellow::mCherry that is not exported to the cuticle, or from cuticular Yellow::mCherry that is re-internalized (Riedel et al., 2011).

We concluded from these experiments that Yellow production is tightly correlated in space and time with the process of cuticle formation.

2.5. Regulation of Yellow subcellular localization

The tight timing of Yellow expression and cellular dynamics in the epidermis may reflect a structural role in the cuticle as much as its requirement for pigment production. We next investigated how Yellow subcellular localization influences pigmentation, by knocking down a gene that could control Yellow trafficking. This study, together with previous reports (Kornezos and Chia, 1992; Riedel et al., 2011; Walter et al., 1991) indicates that Yellow is secreted and later possibly reinternalized. To examine the control of this phenomenon, we sought to impair endocytosis, and examined the consequences on Yellow::mCherry and on pigment formation in the wing. Megalin, a large low-density lipoprotein receptor-related protein involved in endocytosis, is thought to control Yellow endocytosis at the end of pupal life (Riedel et al., 2011). Yet, Yellow endocytosis was studied in an ectopic context, in third instar wing discs, a tissue that normally does not express Yellow. We first confirmed that *megalin* (*mgl*) RNAi knockdown resulted in darker pigmentation and in a more fragile cuticle in the wing (Fig. 5A–B) (Riedel et al., 2011). We have quantified this overall phenotype by comparing *mgl* RNAi wings to control wings imaged under identical conditions (Fig. 5C). Because these *mgl* RNAi wings show other overall morphological differences (in particular they tend to be smaller than wild-type wings), we wondered whether their darker appearance was really the result of additional pigment deposits. We examine the coloration difference at a higher magnification and found that the density of trichomes is slightly increased in *mgl* RNAi wings (Fig. S4), and could contribute to the darker phenotype. We therefore measured the pigmentation between trichomes in the wing blade of the RNAi and control samples. We found that the cuticle is distinctly darker in the wing blade between trichomes (Fig. 5D–E). Moreover, we found that the pigmentation changes are mainly due to dark specks at the base of many trichomes. These specks are occasionally present, but smaller and fainter, in control wings. Interestingly, this phenotype is mirrored in the cellular accumulation of Yellow::mCherry in *mgl* RNAi wings (see below).

Crossing a *mgl* RNAi transgenic line to our *y^{mCherry}* line allowed us to directly visualize the effect of *mgl* knockdown on Yellow localization. At 70 h APF, in control wings, Yellow is mainly located in trichomes (Fig. 5F). In *mgl* RNAi wings, Yellow signal is present in trichomes, but also in spots at the cell surface – either in the cuticle between the trichomes or close to the membrane (Fig. 5G). At 90 h APF, when Yellow signal is barely detectable in wing hairs in control wings, it is still strong in *mgl* knockdown (Fig. 5H–I). It is also still present in aggregates at the apical cell surface. These Yellow aggregates (Fig. 5G, I) are reminiscent of the dark pigmentation specks seen in adult *mgl* RNAi wings (Fig. 5D). We concluded that in the absence of Mgl, high amounts of Yellow accumulate at the apical outline of the cells, not just in trichomes, leading to the production of dark specks. In line with this increased accumulation of Yellow in the cuticle of *mgl* RNAi wings, we also noted that at 90 h APF, Yellow::mCherry signal in the cytoplasm is stronger in control wings than in *mgl* RNAi wings, although we didn't

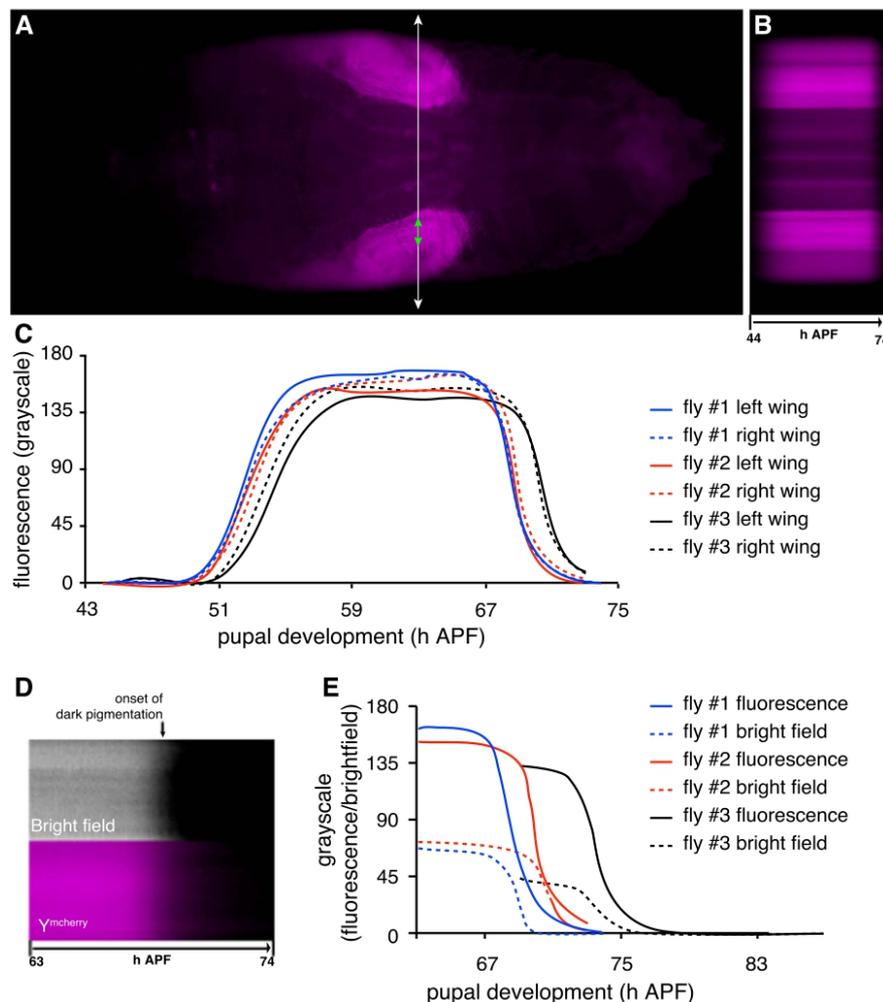


Fig. 3. Live temporal dynamics of Yellow expression in the developing pupal wings. (A) Still frame of a live male $y^{mCherry}$ pupa imaged in time-lapse from 44 to 74 h APF. (B) Chromogram showing the variation of fluorescence intensity during development for one particular position along the antero-posterior axis at the level of the wings as shown by a white line on panel (A). (C) Quantification of fluorescence intensity in the wings of 3 $y^{mCherry}$ male pupae imaged with the same time-lapse settings. (D) Chromograms of the same $y^{mCherry}$ male pupa as in A at the level of the wing (see green arrow in A) for brightfield (top) and fluorescence (bottom) between 63 and 74 h APF. Note the correlation of pigmentation appearance and fluorescence decline (arrow). (E) Quantification of fluorescence and brightfield intensities in 3 $y^{mCherry}$ male pupae, showing some variability in pigmentation appearance. Of note, brightfield settings could not be adjusted finely to be identical, which explains the differences in maximum brightfield intensity between pupae.

find this difference to be statistically significant (Fig. 5H–J). These results reinforce the notion that Mgl is involved in the internalization of Yellow from the cuticle at the end of pupal development, thereby regulating its amount and cuticular embedding for dark pigment production.

2.6. A structural role of Yellow in the cuticle?

The expression and cellular trafficking of Yellow, its regulated embedding in the cuticle produced by a given cell, result in a variety of pigmentation phenotypes at the level of an entire animal. These include light homogeneous gray dusking of certain body parts, but also

specific pigmentation patterns characterized, in the fly, by a steep transition from the background to the pigmented area. In addition to imparting dark pigmentation to these patterns, Yellow has been proposed to act as a cuticular anchor for pigmentation, around which phenol oxidases, as well as other enzymes, would irreversibly cross link catecholamines to cuticular compounds (Walter et al., 1991). If this anchor model is correct, catecholamines should diffuse more in Yellow's absence, and one would expect *yellow* mutants to display fuzzier pattern boundaries. To test this hypothesis, we examined another *Drosophila* species, *D. biarmipes*, whose males harbor a solid spot of dark pigments (Fig. 6A). We have isolated a *yellow* null mutant from this species (Arnoult et al., 2013). A wing spot remains visible in

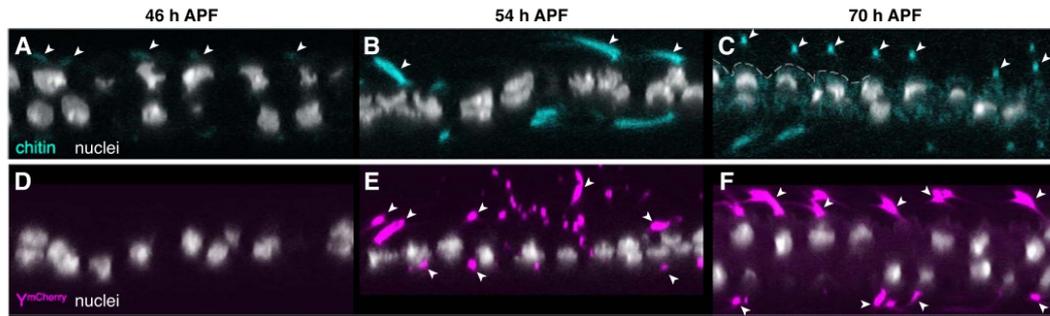


Fig. 4. Temporal dynamics of Yellow subcellular localization in the developing wing. Optical confocal sections through *ChitVis-Tomato* pupal wings (A-C) and *y^{mCherry}* pupal wings (D-F) at 46 (A, D), 54 (B, E) and 70 h APF (C, F). *ChitVis-Tomato*, which labels chitin, was imaged with the same settings at all stages and is shown in blue. Yellow::mCherry was imaged with the same settings at all stages and is shown in magenta (n = 3 for each stage and genotype). Nuclei are marked with DAPI (white). Arrowheads point to expression in the trichomes. The dotted line in C highlights the enrichment of *ChitVis-Tomato* at the apical outline of the cells.

y⁻ males (Fig. 6B), yet it is faint and no longer dark. The background pigmentation is also reduced throughout the wing. The overall location of the pigmentation pattern is, however, similar. We compared the profile of average pigmentation intensity at the transition from the wing background to the spot between *y⁻* and wild-type males (Fig. 6E, see methods). This analysis first showed that the profiles were more

variable inside the spot for the *y⁻* wings, both within and across individuals. Because the wing background pigmentation is more than 3 times less intense in the *y⁻* compared to wild type, we could not compare the global slope of the profile without introducing major biases. Nevertheless, at the outer edge of the more pigmented area in *y⁻* flies, the pigmentation levels were consistently higher compared to wild

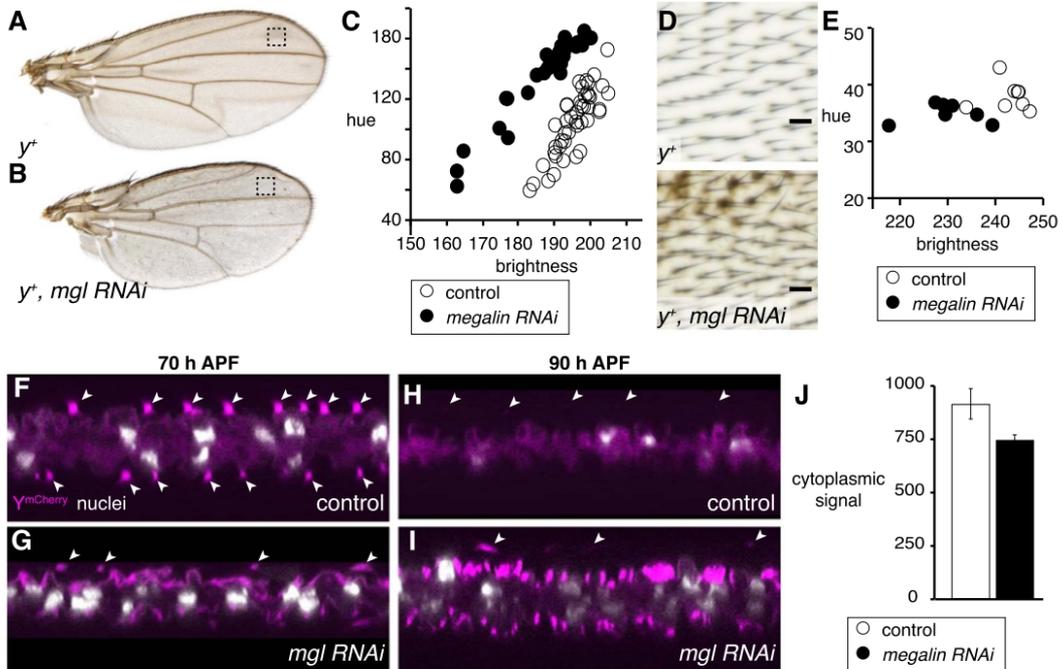


Fig. 5. Megalin controls pigmentation levels in the wing through reinternalization of Yellow. (A-B) Representative wings of 5-day old males of the genotypes *y⁺;UAS-mgl-shRNA* (A) and *y⁺; NP3537 > UAS-mgl-shRNA* (B). (C) Quantification of pigmentation in samples of the genotypes *y⁺* (control, white circles, n = 44) and *y⁺; NP3537 > UAS-mgl-shRNA* (black circles, n = 31). Pigmentation differences were analyzed by plotting brightness against hue measured in the distal part of each wing, between the veins L2 and L3 (box in A, B) (see methods). (D) Details of representative *y⁺;UAS-mgl-shRNA* control (top) and *y⁺; NP3537 > UAS-mgl-shRNA* wings (bottom) showing pigmentation specks at the basis of some trichomes (scale bar: 10 μm). (E) Quantification of pigmentation between trichomes in such pictures (n = 8 wings of each genotype, each wing sampled with 4 20-pixel squares). (F-I) Optical confocal sections through pupal wings of the genotype *y^{mCherry}* (F, H) and *y^{mCherry}; NP3537 > UAS-mgl-shRNA* (G, I) at 70 (F, G) and 90 h APF (H, I). Yellow::mCherry was imaged with the same settings at all stages and is shown in magenta (n = 5 for each stage and genotype). Nuclei are marked with DAPI (white). Arrowheads point to expression in the trichomes. (J) Quantification of cytoplasmic Yellow::mCherry signal (n = 5 wings of each genotype, each wing sliced 3 times and each slice sampled with 20 30-pixel squares).

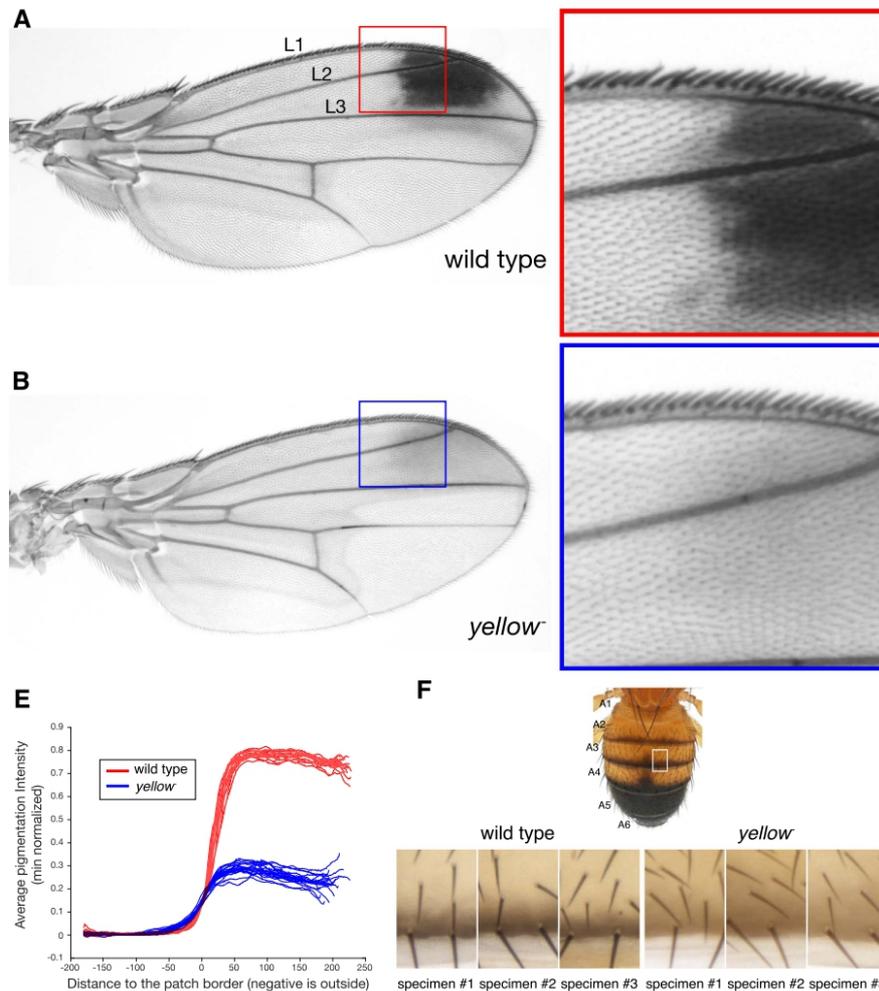


Fig. 6. Pigmentation intensity profiling at wing spot boundary between wild-type and *yellow* *D. biarmipes* males. (A–B) Representative wings of 5-day old *D. biarmipes* males of the genotypes y^+ (A) and y (B). L1–L3 indicate veins referred to in methods. (C–D) Higher resolution views of the regions boxed in (A, B). (E) Profiling of pigmentation intensity at the transition between wing background and spot area after alignment of 16 wild-type (red) and 11 y - (blue) wings, and normalization of the lower values (see methods). (F) Picture of a male *D. biarmipes* abdomen (top), showing the portion of A3 segment that was compared between y^+ and y samples (bottom).

type, suggesting that the pattern was more diffuse in *yellow* (Fig. 6E, Fig. S5). We then used the same flies and examined the boundaries of the banding pigmentation pattern on their abdomen (Fig. 6F). Comparing the anterior boundary of a band on segment A3 for instance between wild-type flies and y mutants also shows a striking difference in the sharpness of the pattern boundary. While the boundary is sharp and parallel to the anterior border of the segment in the wild type, it is fuzzy and irregular in the mutant, resulting in a broader band.

We concluded from these results that, in addition to contributing a dark hue to pigmentation, Yellow may also play a role in anchoring pigment deposits into the cuticle. Alternatively, we cannot exclude that the pigments produced in a *yellow* mutant are more mobile by nature than the black pigment deposits characteristic of the wild type, and explain the apparent dispersion that we describe above.

3. Discussion

3.1. Understanding pigment formation in time and space, and from cell resolution to tissue level phenotype

The literature on pigment formation in insects focuses largely on the biochemistry of this process (Czapla et al., 1990; Sugumaran and Barek, 2016; Wright, 1987), as well as on the spatial control of pigment deposition at the level of tissues (Kronforst et al., 2012; Wittkopp et al., 2003). The temporal and developmental dimension of pigment formation, as well as the cellular processes leading a structure to be colored are generally not well understood. But perhaps more limiting to understand the process of pigment formation, very few studies attempt to connect these different dimensions: development, cellular processes, tissue

patterning and molecular function. In an attempt to connect these dimensions, our motivation with the present work was to focus on a key determinant of black pigment formation in *Drosophila*, *yellow*, and analyze its contribution from cell to tissue across development. The generation of a functional, fluorescent protein-fusion *yellow* allele allowed us indeed to examine the process of pigment formation over developmental time. By doing so, we collected direct information on the dynamics of Yellow expression and cellular targeting in relationship to the process of cuticle deposition. We found that in the wing Yellow is exported to the apical membrane at the level of trichomes soon after the cuticle envelope and the trichomes have formed, at the onset of epicuticle secretion (52 h APF) (Sobala and Adler, 2016). Following endogenous Yellow trafficking over development in the wing, we also confirmed that the excess of protein is reinternalized by the epidermal cells (Riedel et al., 2011). In the wing, these cells later disappear (Kiger et al., 2007), and the balance of Yellow levels is therefore set during pupal development. If this process does not take place properly, we found that Yellow accumulates in patches, presumably in the cuticle, and that this accumulation correlates with specks of pigmentation at the base of trichomes of the adult wing, explaining the overall darker color of the wings. In this way, our results do integrate the description of a cellular process over developmental time, and its overall phenotypic consequences on the animal body.

In line with previous work, our study showed how Yellow protein levels in the adult cuticle are determined by regulated developmental processes impacting the final color. The embedding of Yellow in the cuticle may also assume a structural role in the establishment of pigmentation patterns (Drapeau, 2003; Li and Christensen, 2011; Walter et al., 1991). In *Drosophila*, as in many other insects, the animal emerging from a pupa is pale, or completely unpigmented. Pigment pattern develop in the hours to days following eclosion. They result from diffusing pigment precursors being locally converted into colored precipitates (True et al., 1999). These pigment deposits are not thought to be diffusible, but it is conceivable that they nevertheless spread over a few cell diameters over time. In that respect, using mosaic gynandromorphs, Hannah (Hannah, 1953) showed that cuticle and bristles of *y⁻* genotype could have wild-type pigmentation when located in the vicinity of wild-type, Yellow-expressing, clones. She interpreted this result as the diffusion of pigment-producing substances. It may indicate that wild-type Yellow protein diffuses in the cuticle after being produced from wild-type clones, or that black melanin diffuses in the cuticle after its conversion in the cuticle of wild-type clones. Her experiments however were not designed to answer the question of whether pigments and/or pigment-producing enzymes diffused faster in the cuticle in the absence of Yellow. Pigment patterns, in particular in insects, are characterized by sharp boundaries. The proteins involved in pigment precursor conversion may also stabilize pigment deposits, and thereby contribute to sharper pattern boundaries. We explored this possibility by examining the edge of a pigmentation pattern element in *Drosophila* wings. Our results show that the boundary becomes fuzzy in the absence of Yellow, consistent with an anchoring, structural role of this protein in pigment deposition. Yellow, through its cysteine and methionine residues, could cross-link 5,6-indole quinones in the cuticle (Geyer et al., 1986). Melanins are usually found associated with proteins (Mason, 1955), and Yellow could be associated to dopamine-melanin (Gibert et al., 2017). In its absence, either an altered form of melanin would be produced, or indole-5,6-quinones would self-polymerize in the presence of beta-alanine, resulting in the formation of a tan instead of black pigment (Sherald, 1980). However, no structural protein interacting with 5,6-indole quinones has been identified to date in any species (Sugumaran and Berek, 2016).

3.2. *Yellow*, a pleiotropic gene with a neuro-developmental function

The diversification of pigmentation pattern in *Drosophila* is intimately and recurrently related to changes in the transcriptional

regulation of *yellow*, rather than in its coding sequence (Arnoult et al., 2013; Gompel et al., 2005; Jeong et al., 2006; Prud'homme et al., 2006; Rebeiz et al., 2009; Werner et al., 2010; Williams et al., 2008). This mode of evolution is presumably imposed by the pleiotropic effects of mutations in its coding sequence, not tolerable by natural selection. While a change in the regulation of *yellow* expression may affect this or that pattern element, it is immediately clear that a *yellow* protein mutant is globally changing color (Fig. 11). Adding to the pleiotropy hypothesis, several studies have invoked a behavioral function for this gene, in particular during male courtship (Drapeau et al., 2006). Yet, the neuronal correlate of this behavior remains elusive. Indirect evidence show that a 300 bp regulatory element may control *yellow* expression in two neurons of the larval brain (Drapeau et al., 2006), but it remains unclear whether the lack of *yellow* expression in two neurons of the larval brain compromises male courtship circuitry, or whether *yellow* is expressed later on in the adult brain, in neurons affecting aspects of the male courtship, including wing extension (Bastock, 1956). Using our tagged allele, we explored Yellow expression in the nervous system across the fly life cycle. We confirmed the L3 larval expression in a small subset of cells, not just 2, scattered in the brain and the ventral nerve chord (Drapeau et al., 2006). We also found a similar expression pattern earlier on during larval life, and later on, in the adult. *yellow* has been implicated in the control of male courtship, but the neuronal correlate of this function remains elusive. One could expect Yellow expression in the brain to be dimorphic, or to overlap with Fru^M expression (Stockinger et al., 2005) or both. Our results do not indicate that Yellow distribution is dimorphic in the brain, and its comparison to the published expression of Fru^M does not suggest overlap. At the cellular level, it is also strikingly different from that of epidermal cells: Yellow is confined to the cell cytoplasm in the brain, suggesting a differential mode of production or cellular addressing. Our results deepen the mystery of *yellow* function in the brain, but open the door to the survey and identification of specific neuronal drivers to analyze the role of *yellow* in behavior.

4. Material and methods

4.1. Fly cultures

All stocks were grown on standard cornmeal medium. *M{Act5C-Cas9, 3XP3-RFP, w⁺}ZH-2A, w¹¹¹⁸* was a gift from Frank Schnorrer's lab (Port et al., 2014). *w⁺; nab-Gal4^{NP3537}, tub-Gal80^{ts} / TM6, Sb, Tb* is a wing-specific driver active throughout development (Arnoult et al., 2013; Pavlopoulos and Akam, 2011). *w⁺; UAS-ChTVis-Tomato* line was a gift from Paul Adler's lab (Sobala et al., 2015). *w⁺; sqh-utrophin::GFP* (Rauzi et al., 2010) was a gift from Anne Classen. Other lines were obtained from the Bloomington *Drosophila* Stock Center (BDSC), the Vienna *Drosophila* Resource Center (VDRC), or derived from these stocks with the following references:

- (1) Canton-S
- (2) *w⁺*
- (3) *y¹ w⁺; UAS-mCD8::GFP* (Lee and Luo, 1999).
- (4) *UAS- mgl-shRNA* (VDRC #27242).
- (5) *P{w[+mC]=Pdf-GAL4. P2.4}X, y[1] w[*];* (Park et al., 2000)

4.2. Molecular biology

4.2.1. Repair construct

Part of the *yellow* locus (intron, exon2, 3' UTR and 3' intergenic sequence) was amplified from wild-type *D. melanogaster* genomic DNA using the primers: yFE (CAA TGC TGG GCT CAA TTG GA) and yRI (GCC TGC TCT TTG TTC CTC TG). The resulting amplicon was cloned into a *pJet1.2* plasmid (ThermoFisher Scientific). The *pJet-yellow* vector was digested with *HpaI* (NEB) and used for an InFusion reaction (CloneTech) with an amplicon consisting in the mCherry

sequence with a 5' linker (Waldo et al., 1999). This amplicon was generated by PCR on the pTV-Cherry vector (Baena-Lopez et al., 2013) (a gift from Jean-Paul Vincent) using the primers InFusion-F2 (ATC ATC AGC ATC AAG GTT CCG CTG GCT CCG CTG CTG GTT CTG GC) and InFusion-mCherry-Rev (CCGTGTGTAGGATTATGTTACTTGTACA GCTCGTCCATGCC). The *pJet-yellow-mCherry* vector was then mutated at the target site of sgRNA F4 (see below) to minimize risks of cuts (synonymous mutations). Two amplicons were generated by PCR on the *pJet-yellow-mCherry* vector (using primers *mel_y_PacI_Fw* (GGA ATT TAG GCA GAA ATT CCA G) / *mel_y_mut_Rv* (TCG AAT CCT CGT ATC CGT GGT CAA) and *mel_y_mut_Fw* (AGG ATT CGA AGA TAC GAG CTA CCT G) / *mel_y_StuI_Rv* (GTG CTG GTT GAA AAT ATA GGC C)). The 2 PCR products were combined by overlap extension PCR, to generate a mutated *PacI-StuI* fragment. In parallel, the *pJet-yellow-mCherry* vector was digested by *PacI* and *StuI* and used for an InFusion reaction (CloneTech) with the mutated *PacI-StuI* fragment. This resulted in a *pJet-yellow_F4mut-mCherry* vector, hereafter called the repair construct (sequence in Text S1).

4.2.2. sgRNA

The sgRNA y2 (GGA TGA GTG TGG TCG GCT GTG TTT TAG AGC TAG AAA TAG CAA GTT AAA ATA AGG CTA GTC CGT TAT CAA CTT GAA AAA GTG GCA CCG AGT CCG TGC TTT T) was described by Bassett and colleagues (Bassett et al., 2013). The sgRNA F4 (GGT GAC CAC GGA TAC GCG AAT TGT TTT AGA GCT AGA AAT AGC AAG TTA AAA TAA GGC TAG TCC GTT ATC AAC TTG AAA AAG TGG CAC CGA GTC GGT GCT TTT) was designed using www.flyrnai.org/crispr2 (Housden et al., 2015). Both sgRNAs were produced as described (Bassett and Liu, 2014).

4.3. Embryo injections

505 embryos of the *y**, *M{Act5C-Cas9, 3XP3-RFP, w*}ZH-2A, w¹¹¹⁸* line were injected with water solution of sgRNA y2 at 40 ng/ μ L. 105 G0 adults were screened, 92 showed mosaic *yellow* clones. 4 independent *yellow* mutant lines (*y^{CRISPR y2}*) were recovered (Fig. S1C). 1164 embryos *y^{CRISPR y2}, M{Act5C-Cas9, 3XP3-RFP, w*}ZH-2A, w¹¹¹⁸* line were injected with a dilution of the sgRNA F4 (80 ng/ μ L) and the repair construct *pJet-yellow_F4mut-mCherry* (150 ng/ μ L). 71 G0 adults were screened, 5 showed mosaic wild-type clones. Among them, two had wild-type pigmented flies in their progeny, from which lines were established. We extracted genomic DNA from these lines and ran a diagnostic PCR (*y-mel-ex1-Fw* (AAG CCA CCT GAT TAC CCG AA)/*y-insert-Rv2* (CAC GAT GAC TGA TGT GTG GT)) to confirm the repair (Fig. S1D). A portion of the *yellow* locus was then amplified (*y-intron-Fw* (AGC AAA TCG GTA GTG GCA AC)/*y-insert-Rv2* (CAC GAT GAC TGA TGT GTG GT)) and cloned into a pCR™8/GW/TOPO™ vector (ThermoFisher Scientific) which was sequenced with Sanger technology at Eurofins genomics. This showed that the whole repair sequence was introduced, with no mutation, at the endogenous *yellow* locus, as the fragment cloned was larger on the 5' end than the homology arm.

4.4. Genome sequencing

The genomic DNA of 40 females from the *y^{mCherry}* line was purified using the Blood and Cell Culture DNA Midi kit (Qiagen). The library was prepared using the 1S Plus Kit (Swift Biosciences) with a mean library size of 400 bp, and 11 million reads were sequenced paired end (2 * 50 bp) on a HiSeq. 1500 by the LAFUGA sequencing facility of the Ludwig-Maximilians University Gene Center in Munich. Analyses were performed on the LAFUGA Galaxy web server. Briefly all reads were BLASTed against the sequence of the repair construct with a cutoff of 0.0001. 390 mate reads both BLASTed on the repair construct, and BLAST results were used to infer the length of each insert (Fig. S1F). In 27 cases, only one read of the pair BLASTed on the repair construct.

The 27 mate reads were retrieved and BLASTed against the *Drosophila melanogaster* genome (release r6.17) downloaded from Flybase (Gramates et al., 2017) (Fig. S1G).

4.5. Immunocytochemistry

4.5.1. Antibody production

A polyclonal anti-Yellow antibody was produced at the Ludwig-Maximilians University Veterinary school by immunizing 2 rabbits with a purified Yellow-GST protein produced from the expression vector Dmel-Yellow-GST in pGEX-5 × 1 (a gift from Trisha Wittkopp; (Wittkopp et al., 2002a)). Sera were collected 2 months after immunization and affinity purified.

4.5.2. Western blot

Pupal wings were ground in 2x Laemmli buffer (0.125 M Tris-Cl, pH 6.8, 4.1% SDS, 3.1% DTT, 20% glycerol, bromophenol blue), and boiled at 95 °C for 5 min. Protein samples were then spun down and the supernatants were run on a 10% acrylamide SDS-PAGE gel (~6 wings were used per lane). Western blots were performed as in (Wittkopp et al., 2002a), using the rabbit anti-Yellow antibody described above (1:200), a rabbit anti-mCherry antibody (Novus NBP2-25157) (1:2000) and a rabbit H2Av antibody (1:2000) (a gift from Carla Margulies, produced as in (Leach et al., 2000)). The secondary antibody, a goat anti-rabbit IgG-HRP conjugate (Bio-Rad 170-6515) was used at 1:10000. Detection was performed using the Immobilon Western Chemiluminescent HRP substrate (Merck).

4.6. Pigmentation quantification

4.6.1. *D. melanogaster* wings

Male flies raised at 20 °C were collected upon hatching and left to mature 5 days at 20 °C for pigmentation analysis of the *y^{mCherry}*, *w** line, compared to *w** and *y¹ w** flies. For the experiments with *UAS-mgl-shRNA* line, flies were collected after hatching and left to mature 7 days at 20 °C, then stored in 80% ethanol at -20 °C until dissection. A single wing per individual was dissected and mounted in Hoyer's medium (Ashburner, 1989). Wings were imaged under a Leica Macroscope equipped with a Manta G-609B/C camera (GigE camera with Sony ICX694, Allied Vision, Exton, PA) driven by nVision software (Impuls Imaging GmbH, Türkheim) using a diffuse back lighting table (DBL-2020-WT, MBJ Imaging, Hamburg) for illumination. The resulting color images were converted to HSB coordinates (Joblove and Greenberg, 1978). Brightness and hue were averaged from a 50-pixel square in the distal part of the wing between veins L2 and L3, using Fiji (Schindelin et al., 2012)).

4.6.2. *D. biarmipes* wings

Wings of 11 *y¹* and 16 wild-type 5 day-old adult males were prepared, imaged and registered on a reference wing as in (Arnoult et al., 2013). Average intensity profiles represent the average pigmentation intensity (= 255 - gray level) in the compartments between L1 and L2 veins and between L2 and L3 veins (boxed region in Fig. 6A,B) relative to the distance to an arbitrary proximal limit between more pigmented area and wing background. For each individual, the limit was the proximal border of the area defined by a constant threshold above the background average intensity (see examples on Fig. S5). The threshold was defined as the half of the difference between the average intensity inside and outside the spot, averaged across all *y¹* mutant wings. The threshold allows to center the profiles on a given intensity reference. For each point inside the two compartments, the distance to the proximal spot boundary was calculated using the distance transform (Borgefors, 1986). The averaging of pigmentation intensity for each distance bin resulted in a profile of average pigment intensity, the distance coordinate being relative to a comparable reference. This approach is robust to variation in pattern boundary. It allows to

compute an average profile for each individual that integrates the information all along the spot, not just on a single line. The minimum value near the spot boundary of each profile, i. e., the average intensity of the background has been normalized to 0, to allow the comparison of the slope of the increase in pigmentation intensity. Although it does not reflect the perceived spot border for the wild type (Fig. S5A₃), this constant threshold allows to compare, on common reference, the slope of the transition from the background to the pigmented area, by aligning the profiles (Fig. S5A₂–B₂, A₄–B₄).

4.6.3. Wholemout flies

Specimens from Fig. 11 are 5-day old males raised at 20 °C, anaesthetized and imaged as in (Chyb and Gompel, 2013).

4.7. Fluorescent sample preparation and imaging

4.7.1. Wholemout pupae

$y^{mCherry}$; *sqh-utrophin::GFP* pupae were dissected out of their pupal case, mounted live in Voltalef oil to permit gas exchanges, and imaged in tiles with a Leica TCS SP5 II confocal microscope with the 10× objective. Tiled image stacks were stitched in Fiji using the Stitching plugin (Preibisch et al., 2009).

4.7.2. Pupal abdomen

65 h, 70 h and 75 h APF $y^{mCherry}$; *sqh-utrophin::GFP* male pupae were dissected from their pupal case, mounted live in Voltalef oil to permit gas exchanges, and imaged with a Leica TCS SP5 II confocal microscope with the 20× objective, focusing on the A3 and A4 abdominal segments. Image stacks were projected (maximum intensity) and analyzed using Fiji. For each sample, two 150-pixels wide areas, one on each side of the midline, encompassing the whole antero-posterior length of the segment A3 or A4, were blurred using Gaussian blur (sigma 4) and analyzed using the Plot Profile tool of Fiji. The average of all profiles is shown (Fig. S2B–C). The blurred areas of age-matched samples were also projected (average intensity) to get an average of the expression pattern for each segment and at each stage.

4.7.3. Central nervous system

Brains from unsexed L1 (3 $y^{mCherry}$, 3 y^+ controls), L2 (5 $y^{mCherry}$, 1 y^+ control) and L3 (6 $y^{mCherry}$, 5 y^+ controls) larvae as well as 5-day old adult female flies (7 $y^{mCherry}$, 5 $y^{mCherry}$, *pdf > GFP*; 2 y^+ controls) were dissected in cold PBS and collected in 1% paraformaldehyde on ice. Following fixation with 4% paraformaldehyde at room temperature (20 min for L1, 30 min for L2 and L3, 45 min for adult brains), brains were washed 3 × 10 min in PBS, 0.1% Triton X-100 and then incubated in blocking solution for 1 h (3% normal goat serum, 0.1% Triton X-100 in PBS). Primary antibodies were incubated overnight at 4 °C in blocking solution. After washing 3 × 20 min in PBS, 0.1% Triton X-100, brains were incubated with secondary antibodies for 4 h at room temperature in blocking solution and again washed in PBS, 0.1% Triton X-100 for 3 × 20 min and 1 × 1 h. All tissues were mounted with Vectashield mounting medium. All microscopic observations were made at a Leica SP8 confocal microscope. Image stacks were analyzed and projected (maximum intensity) using Fiji (Schindelin et al., 2012).

The primary antibodies used were rat anti-N-cadherin (anti-N-cad DN-Ex #8, Developmental Studies Hybridoma Bank, 1:200), Living Colors® rabbit anti-DsRed (Clontech, 1:200) and 75–132 anti-GFP primary antibody (specific to full GFP, monoclonal, NeuroMab, clone N86/38, 1:200). The following secondary antibodies were used: goat anti-rat Alexa 488 (invitrogen, 1:200), goat anti-rabbit Cy3 (Jackson ImmunoResearch, 1:200) and anti-mouse Alexa488 (molecular probes, 1:250), respectively.

4.7.4. Time-lapse of pupae

Male white prepupae ($y^{mCherry}$, w^*) were selected and left to develop at 25 °C for 44 h, then placed individually, ventral side up, in

a humid chamber. The chamber was a 55 mm Petri dish coated with humid tissue, covered with parafilm, and with a hatch at the center of the parafilm lid, for imaging with the 10x objective of a Zeiss Imager M2 wide field microscope. The whole pupa was scanned every 450 s under bright field and fluorescent light with a pco sensicam camera driven with a custom camera software (Lim et al., 2016). Imaging was performed for approximately 48 h at 20 °C. Development at 20 °C was assumed to be 1.5 times slower than at 25 °C (Ludwig and Cable, 1933; Powsner, 1935), and developmental times indicated on the figures are equivalent to development at 25 °C. Time-lapse stacks were analyzed using Fiji. For each wing, 6 squares of 12*12 pixels were quantified on the best focused section of the stack and the means of these measures were plotted against time.

4.7.5. Pupal wings

Male white prepupae of 4 genotypes (1. $y^{mCherry}$, w^* ; *UAS-mCD8::GFP/+*; *NP3537/+* called " $y^{mCherry}$, *mCD8::GFP*" – 2. y^+ , w^* ; *UAS-ChtVis-Tomato/ UAS-mCD8::GFP*; *NP3537/+* called "*ChtVis-Tomato*, *mCD8::GFP*" – 3. y^+ , w^* ; *UAS-mCD8::GFP/+*; *NP3537/+* called " y^+ *mCD8::GFP*" – 4. $y^{mCherry}$, w^* ; *UAS-mCD8::GFP/UAS-mgl-shRNA*; *NP3537/+* called " $y^{mCherry}$, *mgl RNAi*") were selected and aged at 25 °C until the appropriate developmental point. They were then dissected out of the pupal case, wings were removed from their envelope and allowed to unfold in distilled water. They were then fixed in PBS 4% PFA for 30 min at room temperature, washed in PBS and mounted in Vectashield with DAPI (Vector Laboratories). They were then imaged with a Leica TCS SP5 II confocal microscope, using the 63× objective. Stacks were resliced along the z-axis in Fiji, accounting for the chromatic shift. The imaging of y^+ , *mCD8::GFP* wings revealed low levels of autofluorescence, much weaker than the signal produced by Yellow::mCherry (Fig. S3).

All images were processed with Adobe Photoshop, using linear enhancement as well as gamma correction.

Acknowledgements

We thank Haris Khan and Carla Margulies for help with Western blots and for sharing reagents, Nadin Memar for help with time-lapse imaging, Karin Merk and Klaus Förstemann for test of the sgRNAs in cell culture, Helmut Blum and Stefan Krebs of the Gene Center Genomics Unit (LMU, Munich) for library preparation and Illumina sequencing, Yan Jaszczyszyn and Cloelia Dard-Dascal of the I2BC High Throughput Sequencing Core Facility for help with the analysis of the sequencing data. We thank Frank Schnorrer, Paul Adler Anne Classen, Jean-Paul Vincent, Aaron Voigt and Barry Thompson for sharing fly stocks and reagents. We are grateful to Patricia Wittkopp and Artyom Kopp for constructive comments on the manuscript. We also thank Irina Hein for discussion in the early phase of the project and Sabine Radetzki for technical support with the molecular biology. Stocks obtained from the Bloomington Drosophila Stock Center (NIH P400D018537) were used in this study.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 701691, and from the Ludwig-Maximilians University of Munich. LR was supported by the Amgen Scholar program of the LMU; JK is supported by the Max Planck Society; YX is supported by a China Scholarship Council Ph.D. fellowship.

Appendix A. Supporting information

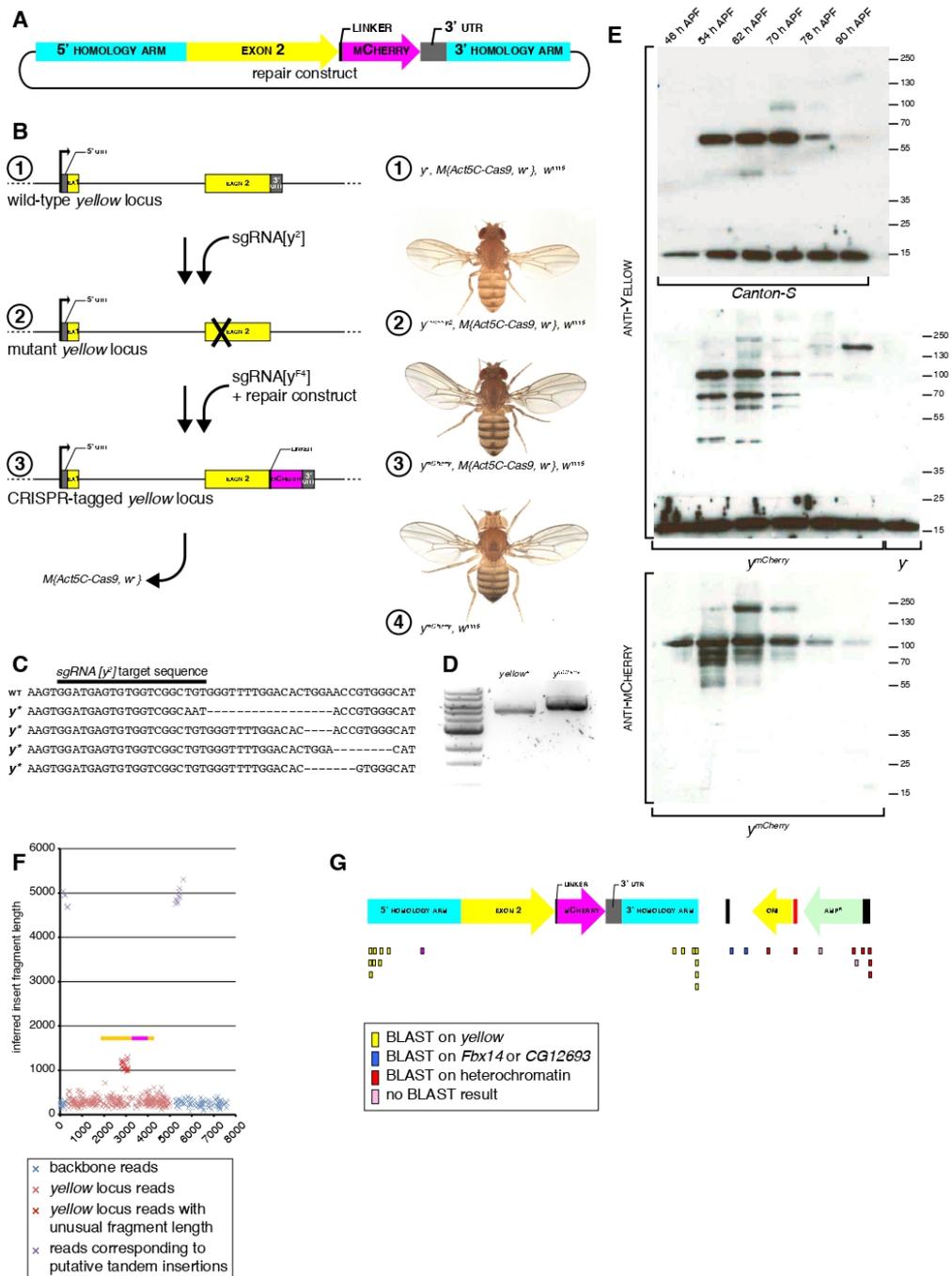
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2018.04.003.

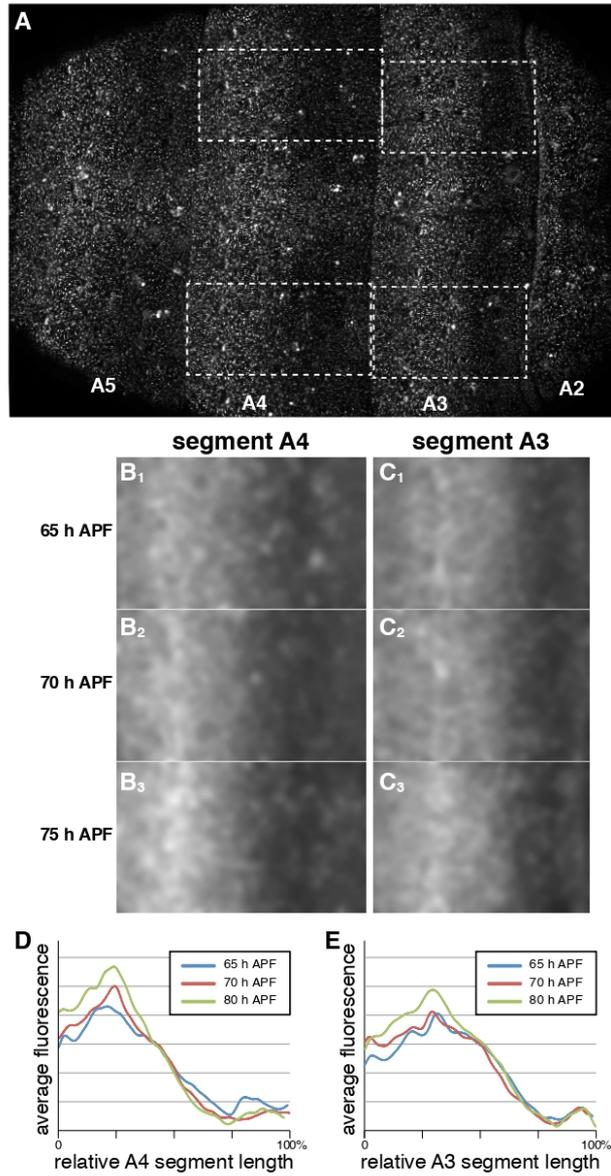
References

- Arnoult, L., Su, K.F., Manoel, D., Minervino, C., Magrina, J., Gompel, N., Prud'homme, B., 2013. Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* 339, 1423–1426.
- Ashburner, M., 1989. *Drosophila: A Laboratory Handbook and Manual*. Two volumes.
- Baena-Lopez, L.A., Alexandre, C., Mitchell, A., Pasakarnis, L., Vincent, J.P., 2013. Accelerated homologous recombination and subsequent genome modification in *Drosophila*. *Development* 140, 4818–4825.
- Bassett, A., Liu, J.L., 2014. CRISPR/Cas9 mediated genome engineering in *Drosophila*. *Methods* 69, 128–136.
- Bassett, A.R., Tibbit, C., Ponting, C.P., Liu, J.L., 2013. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep.* 4, 220–228.
- Bastock, M., 1956. A gene mutation which changes behavior pattern. *Evolution* 10, 421–439.
- Borgefors, G., 1986. Distance transformations in digital images. *Comput. Vision., Graph., Image Process.* 34, 344–371.
- Brehme, K.S., 1941. The effect of adult body color mutations upon the Larva of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 27, 254–261.
- Bridges, C.B., Morgan, T.H., 1919. Contributions to the genetics of *Drosophila melanogaster*. *Carne. Inst. Wash. Publ.* 278, 123–304.
- Bridges, C.B., Morgan, T.H., 1923. The third-chromosome group of mutant characters of *Drosophila melanogaster*. *Carne. Inst. Wash. Publ.* 327, 1–251.
- Budnik, V., White, K., 1987. Genetic dissection of dopamine and serotonin synthesis in the nervous system of *Drosophila melanogaster*. *J. Neurogenet.* 4, 309–314.
- Chyb, S., Gompel, N., 2013. Atlas of *Drosophila* morphology: Wild-type and Classical Mutants. Academic Press, London; Waltham, MA.
- Czapla, T.H., Hopkins, L.T., Kramer, K.J., 1990. Catecholamines and related o-diphenols in cockroach hemolymph and cuticle during sclerotization and melanization: comparative studies on the order Diptera. *J. Comp. Physiol. B, Biochem. Syst. Environ. Physiol.* 160, 175–181.
- Drapeau, M.D., 2003. A novel hypothesis on the biochemical role of the *Drosophila* Yellow protein. *Biochem. Biophys. Res. Commun.* 311, 1–3.
- Drapeau, M.D., Cyran, S.A., Viering, M.M., Geyer, P.K., Long, A.D., 2006. A cis-regulatory sequence within the yellow locus of *Drosophila melanogaster* required for normal male mating success. *Genetics* 172, 1009–1030.
- Drapeau, M.D., Radovic, A., Wittkopp, P.J., Long, A.D., 2003. A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the *Drosophila melanogaster* larval brain. *J. Neurobiol.* 55, 53–72.
- Edwards, K.A., Doescher, L.T., Kaneshiro, K.Y., Yamamoto, D., 2007. A database of wing diversity in the Hawaiian *Drosophila*. *PLoS One* 2, e487.
- Geyer, P.K., Spana, C., Cores, V.G., 1986. On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *EMBO J.* 5, 2657–2662.
- Gibert, J.M., Mouchel-Viehl, E., Peromet, F., 2017. Modulation of yellow expression contributes to thermal plasticity of female abdominal pigmentation in *Drosophila melanogaster*. *Sci. Rep.* 7, 43370.
- Gompel, N., Carroll, S.B., 2003. Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* 424, 931–935.
- Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., Carroll, S.B., 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487.
- Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., Falls, K., Goodman, J.L., Hu, Y., Ponting, L., Schroeder, A.J., Strelets, V.B., Thurmond, J., Zhou, P., the FlyBase, C., 2017. FlyBase at 25: looking to the future. *Nucleic Acids Res.* 45, D663–D671.
- Hannah, A.M., 1953. Non-autonomy of yellow in gynandromorphs of *Drosophila melanogaster*. *J. Exp. Zool.* 123, 523–560.
- Helfrich-Forster, C., Homborg, U., 1993. Pigment-dispersing hormone-immunoreactive neurons in the nervous system of wild-type *Drosophila melanogaster* and of several mutants with altered circadian rhythmicity. *J. Comp. Neurol.* 337, 177–190.
- Housden, B.E., Valvezan, A.J., Kelley, C., Sopko, R., Hu, Y., Roedel, C., Lin, S., Buckner, M., Tao, R., Yilmazel, B., Mohr, S.E., Manning, B.D., Perrimon, N., 2015. Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci. Signal.* 8, rs9.
- Jeong, S., Rokas, A., Carroll, S.B., 2006. Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* 125, 1387–1399.
- Joblove, G., Greenberg, D.P., 1978. Color spaces for computer graphics. *Comput. Graph.* 12, 20–25.
- Jurgens, G., Wieschaus, E., Nusslein-Volhard, C., Kluding, H., 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: II. Zygotic loci on the third chromosome. *Wilhelm. Roux's Arch. Dev. Biol.* 193, 283–295.
- Kiger, J.A., Jr., Natzle, J.E., Kimbrell, D.A., Paddy, M.R., Kleinhesselink, K., Green, M.M., 2007. Tissue remodeling during maturation of the *Drosophila* wing. *Dev. Biol.* 301, 178–191.
- Kornzeos, A., Chia, W., 1992. Apical secretion and association of the *Drosophila* yellow gene product with developing larval cuticle structures during embryogenesis. *Mol. General. Genet.* MGG 235, 397–405.
- Kronforst, M.R., Barsh, G.S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S.P., Protas, M., Rosenblum, E.B., Schneider, C.J., Hoekstra, H.E., 2012. Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell Melanoma Res.* 25, 411–433.
- Kronforst, M.R., Papa, R., 2015. The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. *Genetics* 200, 1–19.
- Leach, T.J., Mazzeo, M., Chotkowski, H.L., Madigan, J.P., Wotring, M.G., Glaser, R.L., 2000. Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster*. *J. Biol. Chem.* 275, 23267–23272.
- Lee, T., Luo, L., 1999. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* 22, 451–461.
- Li, J., Christensen, B.M., 2011. Biological function of insect yellow gene family. In: Liu, T., Kang, L. (Eds.), *Recent Advances in Entomological Research*. Springer, Berlin, Heidelberg.
- Lim, M.A., Chitturi, J., Laskova, V., Meng, J., Findeis, D., Wiekenberg, A., Mulcahy, B., Luo, L., Li, Y., Lu, Y., Hung, W., Qu, Y., Ho, C.Y., Holmyard, D., Ji, N., McWhirter, R., Samuel, A.D., Miller, D.M., Schnabel, R., Calarco, J.A., Zhen, M., 2016. Neuroendocrine modulation sustains the *C. elegans* forward motor state. *eLife* 5.
- Lindsley, D.L., Grell, E.H., 1968. *Genetic variations of Drosophila melanogaster*.
- Locke, M., 2001. The Wigglesworth lecture: insects for studying fundamental problems in biology. *J. Insect Physiol.* 47, 495–507.
- Ludwig, D., Cable, R.M., 1933. The effect of alternating temperatures on the pupal development of *Drosophila melanogaster* Meigen. *Physiol. Zool.* 6, 493–508.
- Mason, H.S., 1955. Comparative biochemistry of the phenolase complex. *Adv. Enzymol. Relat. Subj. Biochem.* 16, 105–184.
- Massey, J.H., Wittkopp, P.J., 2016. The genetic basis of pigmentation differences within and between *Drosophila* species. *Curr. Top. Dev. Biol.* 119, 27–61.
- Morgan, T.H., Bridges, C.B., 1916. Sex-linked inheritance in *Drosophila* 237. Carnegie Institute of Washington Publication, 1–88.
- Moussian, B., 2010. Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochem. Mol. Biol.* 40, 363–375.
- Nash, W.G., 1976. Patterns of pigmentation color states regulated by the *y* locus in *Drosophila melanogaster*. *Dev. Biol.* 48, 336–343.
- Park, J.H., Helfrich-Forster, C., Lee, G., Liu, L., Rosbash, M., Hall, J.C., 2000. Differential regulation of circadian pacemaker output by separate clock genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 97, 3608–3613.
- Pavlopoulos, A., Akam, M., 2011. Hox gene Ultrabithorax regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. *Proc. Natl. Acad. Sci. USA* 108, 2855–2860.
- Port, F., Chen, H.M., Lee, T., Bullock, S.L., 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 111, E2967–E2976.
- Powson, L., 1935. The effects of temperature on the durations of the developmental stages of *Drosophila melanogaster*. *Physiol. Zool.* 8, 474–520.
- Preibisch, S., Saalfeld, S., Tomancak, P., 2009. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* 25, 1463–1465.
- Prud'homme, B., Gompel, N., Rokas, A., Kassner, V.A., Williams, T.M., Yeh, S.D., True, J.R., Carroll, S.B., 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* 440, 1050–1053.
- Radovic, A., Wittkopp, P.J., Long, A.D., Drapeau, M.D., 2002. Immunohistochemical colocalization of Yellow and male-specific Fruitless in *Drosophila melanogaster* neuroblasts. *Biochem. Biophys. Res. Commun.* 293, 1262–1264.
- Rauzi, M., Lenne, P.F., Lecuit, T., 2010. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature* 468, 1110–1114.
- Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., Carroll, S.B., 2009. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326, 1663–1667.
- Riedel, F., Vorkel, D., Eaton, S., 2011. Megalin-dependent yellow endocytosis restricts melanization in the *Drosophila* cuticle. *Development* 138, 149–158.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Gopman, B.N., Palmer, A.E., Tsien, R.Y., 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Drosophila* Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* 22, 1567–1572.
- Sherald, A.F., 1980. Sclerotization and coloration of the insect cuticle. *Experientia* 36, 143–146.
- Sobala, L.F., Adler, P.N., 2016. The gene expression program for the formation of wing cuticle in *Drosophila*. *PLoS Genet.* 12, e1006100.
- Sobala, L.F., Wang, Y., Adler, P.N., 2015. ChVis-Tomato, a genetic reporter for in vivo visualization of chitin deposition in *Drosophila*. *Development* 142, 3974–3981.
- Sobala, L.F., Wang, Y., Adler, P.N., 2016. Correction: ChVis-Tomato, a genetic reporter for in vivo visualization of chitin deposition in *Drosophila*. *Development* 143, 3638.
- Stockinger, P., Kvisianski, D., Rotkopf, S., Tirian, L., Dickson, B.J., 2005. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* 121, 795–807.
- Sugumar, M., Berek, H., 2016. Critical analysis of the melanogenic pathway in insects and higher animals. *Int. J. Mol. Sci.* 17.
- True, J.R., Edwards, K.A., Yamamoto, D., Carroll, S.B., 1999. *Drosophila* wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterning. *Curr. Biol.* CB 9, 1382–1391.
- Waldo, G.S., Standish, B.M., Berendzen, J., Terwilliger, T.C., 1999. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695.
- Walter, M.F., Black, B.C., Afshar, G., Kermahon, A.Y., Wright, T.R., Biessmann, H., 1991. Temporal and spatial expression of the yellow gene in correlation with cuticle formation and dopa decarboxylase activity in *Drosophila* development. *Dev. Biol.* 147, 32–45.
- Werner, T., Koshikawa, S., Williams, T.M., Carroll, S.B., 2010. Generation of a novel wing colour pattern by the Wingless morphogen. *Nature* 464, 1143–1148.
- Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., Carroll, S.B., 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in

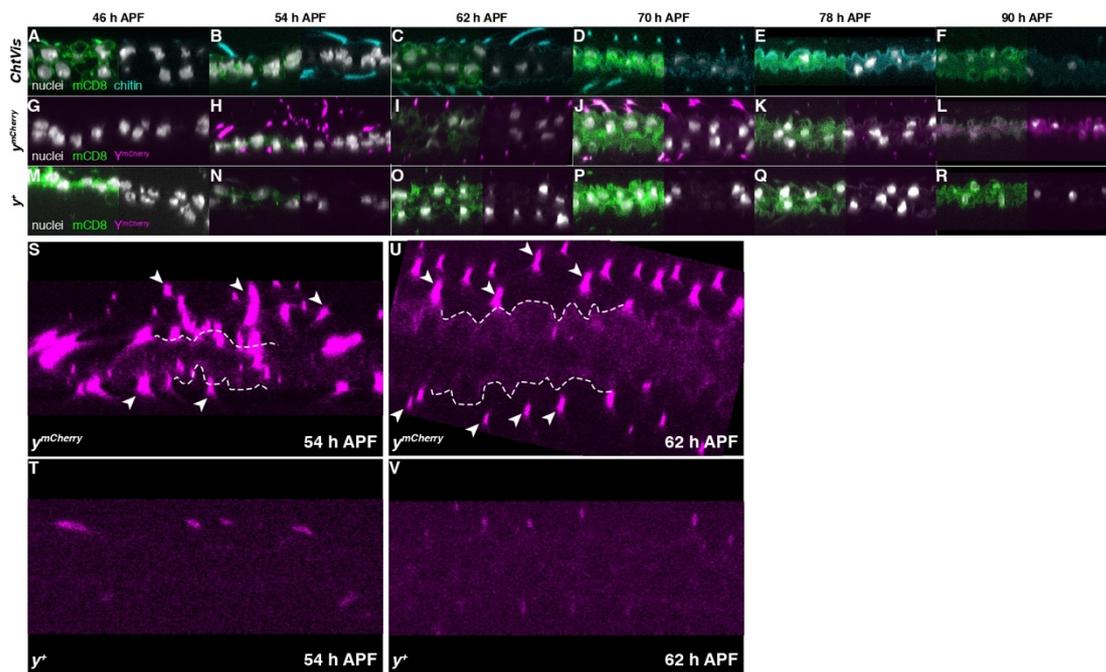
- Drosophila*. Cell 134, 610–623.
- Wilson, J.S., Jahner, J.P., Forister, M.L., Sheehan, E.S., Williams, K.A., Pitts, J.P., 2015. North American velvet ants form one of the world's largest known Mullerian mimicry complexes. *Curr. Biol.*: CB 25, R704–R706.
- Wittkopp, P.J., Carroll, S.B., Kopp, A., 2003. Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet.*: TIG 19, 495–504.
- Wittkopp, P.J., True, J.R., Carroll, S.B., 2002a. Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129, 1849–1858.
- Wittkopp, P.J., Vaccaro, K., Carroll, S.B., 2002b. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.*: CB 12, 1547–1556.
- Wright, T.R., 1987. The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. *Adv. Genet.* 24, 127–222.

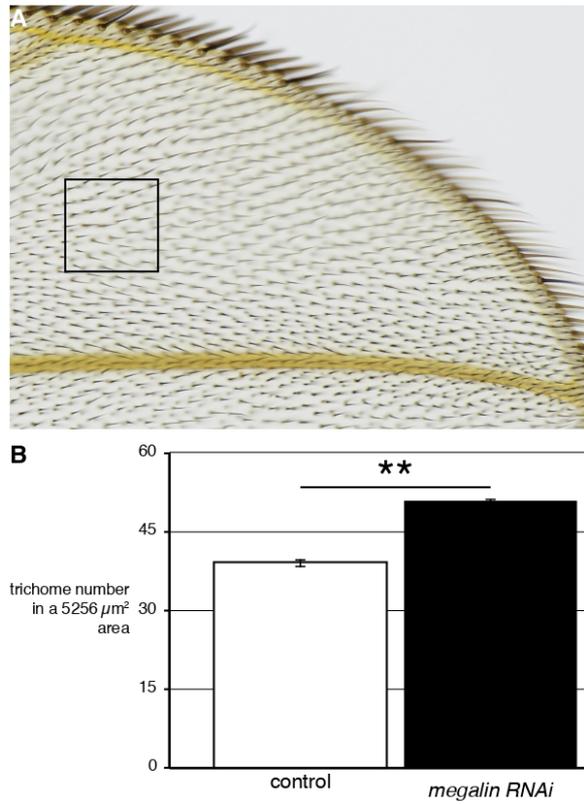
Hinaux et al. Figure S1



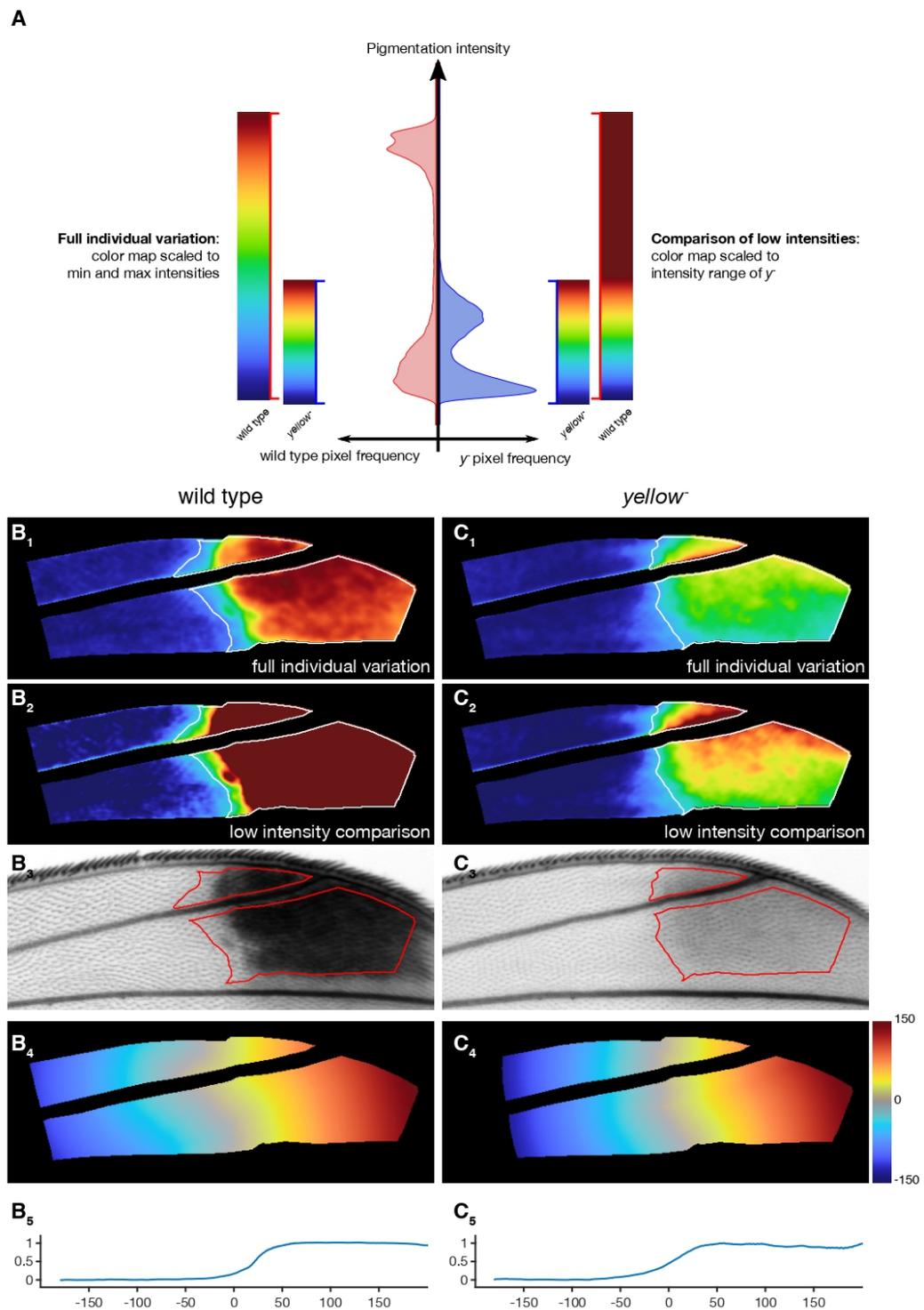


Hinaux et al. Figure S3





Hinaux et al. Figure S5



>pJet1.2_ymCherry_F4_mutated

GCCCCTGCAGCCGAATTATATTATTTTTGCCCCAAATAATTTTTAACAAAAGCTCTGAAGCTTCTTCATTTAAATCTTTAG
ATGATAC TTCATCTGAAAATGTCCCAATTAGTAGCATCACGCTGTGAGTAAGTCTAAACCATTTTTTTATGTTGTA
TTATCTC TAATCTTACTACTCGATGAGTTTTCCGGTATTATCTCTATTTTTAACTTGGAGCAGGTCCATTCATGTTTTT
TTCATCATAGTGAATAAAATCAACTGCTTTAACACTTGTGCCTGAACACCATATCCATCCGGCGTAATACGACTACTAT
AGGGAGAGCGGGCCGAGATCTCCGGATGGCTCGAGTTTTTTCAGCAAGATCAATGCTGGGCTCAATTGGAAAAACTATT
CTATGAAGATTTGAGTAAATAAATTTTTGATTTAAAAAGCCCA TGGTTATCGCGACA ACTAGTACGGGACAAGATTACT
GTTTAAATCAAGTGTGAAATATCAAAATCAAAATCGGATTCCGATCGGGAAAGTTGTA TCCGATTCTGAAACTAAAACAC
AGAAATGCAACATTTTCCGATATCGACTGAGCTGAGCTGAGGCTATTTTATACAGATTCATTAGGCCACCAGCCATTGAATA
TACCCAGTCAATTGAGCTACTCGATAGTTGATCAACTTAGCTTTTGTCAACGAGTGAACGCATAAACTACTACATCAAC
GATATTTGCGGCCCATTTCAAGCTAAAAGTTTCACTTAAATACAAATAAGATTAGAAAAAATATCTGAATGAAAAAATG
TTGAGACATATTTCTTGGAAAAAGGAGAACCTCAAGACAGTCGAAAAAATGTTTACAATGAAAAATGTTGAAAAATCATGA
AGCAGATAAATCGTCAAGTTCGAGGTTTTAGGACTGAAAGAGCACATGTCAAAATATAAATTTGTTCAAATACTTTATA
TTTACTGAAATGAGATTGTTATTTTAAAGTTATGAATTAATAAGATTGAAAGGTGCATTATGCTCAAATGATATTT
ATCGCAACCCCGGTTACTTTGTAAAGCAAAAACGCCCTGGTTTTGATTTTTAAGAAGATGGGTCCGTAATCGATAAAAGC
TATATTTCTGGTCTGTTGAGTCTCACTCGCTGTATAAAAAACATTAAGTTCCAGAAAACAATAATGTC TTTAAAT
TCAATTAACGAAGAAATAAAGAAAGGAAAAAAGAACTGGAGCGGAAATCGGTCGAAATACTGCCAATGGCCCATATCAAT
AACAGCATATATGGTATACATATTGATAATGATGTGACAGCAATTTGCTTACAGCGGTAATGACATCGCAAATGCGAC
GCAACTTGCAATAGTGCCAAATATGACTGAAGTACATATAGCCGGGATCTTTTAACTAAACTTCCAGTAGATGTACAA
GCAGAAAAAGAGCCATTAGCACGGCAGTTACCATTGCTTATGATTCTTGTGTTCCAAAATAATGACAAATAGGTATATA
AATAATTAATCAAGCAAAACATAAGCGATTCTAAATTTACCTTTACATCTGTATGCATTTACATATATCCAGAAAACAGACA
GCATAAAGCTTGCACATTTGCTTAGTATAATAATCCAAAGAAAGGAAATTTAGGCAGAAATCCAGTTAATTAATATTCAAA
ACAAACTTTATTTAGTGCCCAATAATAGTTTGGCCCTGCTAATCTCTCTATTTTATTTTTTAGGGATTCCGGCCACTCT
GACCTATATAAACATGGACCGCAGTTTTGACGGGTTACCCGGAGCTAATTCGGTATCCAGATTGGCGCTCAAATACAGCTG
GAGATTGCGCCAACAGTATTACCACTGCC TACCGCATTAAGTGGATGAGTGTGGTCCGGCTGTGGTTTTGGACACTGGA
ACCGTGGGCATCGGCAATACCACCCTAATCCGTGCCCTATCGGGTAAATGTCTTTGACTTGACCACGGATACGaGgAT
TCGaAGATACGAGCTACCTGGCGTGGACAAAATCCAAATACCTTTTATAGCTAACATTTGCCGTGGATATAGGCAAAAAT
CGCATGATGCAATGCTATTTTTGCCGATGAAATGGGATACGGCTTGTATGCTTACTCTGGGAACTGAACAAGTCCCTGG
AGATTCTCGGCACATTCGTATTTTTTCCCAGATCCATTGAGGGGCGATTTCAATGTCGCTGGTATTACTTCCAAATGGGG
CGAGGAGGATATTTGGTATGTCCCTTTCGCCCATTCGATCGGATGGTTATCGTACCCTGTACTTTAGTCCGTTAGCAA
GTCATCGACAATTTGCCGATCCACGAGGATTTGAGGGATGAAACCAGGACGGAAGATAGCTATCATGACTTTGTTGCC
TTAGATGAACGGGTCCAAACCTCCATAACCACTTCACTGTGATGAGCGATGATGGAAATGAGCTGTTCAATTTAATAGA
TCAAATGCAGTGGGTTGCTGGCACTCATCAATGCCGTACTCACCGCAATTTTCATGGCATTGTGGATCCGGATGACGTTG
GCTTAGTTTTTCCGGCCGATGTGAAAAATGATGAGAACAACAAACGTTTGGGTTCTATCCGATAGGATGCCCGTTTTCTTG
CTGCTGACTTGGATTATTTTCAGATACTAATTTCCGAATTTACACGGCTCCTTGGCCACTTTAATGAGAATCTGTGTG
TGATTTGAGGAATAACGCCTATGGGCCCAAATACCGTTTCAATACCAAACAAGCCGTTTTTGC CAATGGGTCCACCGT
TATATACGAAACAATATCGTCTGTCTTGGCCACAGAAACCTCAGACCAGCTGGGCTTCTCGCCGCTCCTCCAGTCCG
ACTTATTTGCCCGCAATTCAGGCAATGTAGTCTCCAGTATTAGTGTCTCTACAAATCTGTGGGTCCTGCAGGAGTGGGA
GGTGCCAAAGGCTATATTTTTCAACAGCACAACGCCATAAATACGAGACAAGTGGTCCCCATCTATTTCCCAACTC
AACCCGCCAACCCGGTGGCCAGGATGGTGGGTTAAAAACTTA TGTGAATGCCCGCAATCTGGGTGGTGGCATCATCAG
CATCAAGGTTCCGCTGGCTCCGCTGCTGGTTTGGCGAATTCATGGTGAAGGCGGAGGAGGATAACATGGCCATCAT
CAAGGAGTTATCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGG
GCCGCCCTACGAGGGCACCCAGACCGCAAGCTGAAGGTGACCAAGGTTGGCCCTGCCCTTCCGCTGGGACATCCCTG
TCCCTCAGTTTATGTACGGCTCCAAGGCTACGTGAAGCACC CGCCGACATCCCGACTACTTGAAGCTGTCTTCCC
CGAGGGCTTCAAGTGGGAGCGGTGATGAACCTCGAGGACGGCGGCGTGGTGAACCGTGAACCGGACTCCTCCTGCAGG
ACGGCGAGTTTATCTACAAGGTGAAGCTGCGCGGCACCACTTCCCCTCCGACGGCCCGTAATGACAGAAAGACCATG
GGCTGGGAGGCTCTCCGAGCGGATGTACCCGAGGACGGCCCTGAAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAA
GGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCCCTACAACG
TCAACATCAAGTTGGACATCACCTCCCAACAGGACTACACCATCGTGAACAGTACGAACCGCCGAGGGCCGCCAC
TCCACCGCGGCATGGACGAGCTGTACAAGTAACATAATCCTACACACGGTACTTGGGTATATTTCTCACACACTCGATTG
ATGTAAGAATATTTAAAGACAACAACATAGGGCAACAGCGGTTAAAAAACCACATGACGTATGAGCAAGTGGCAAATC
AATACTTTATCTAGTTATGTTAAGCAAAAAATAACAATAAATCAACTTTTTTTGAAGGTTAAGAGTTTACGC AATTTTC
TTGAGCGGAAAAAGCGGAAAAATGTAAGTATGCATAAATTTCAATATATCAACAACGTACATTTTCTGGAGTACTAC
TACCAGGCAAGAAAGTAGGTTGATAAAGCTATGCACAAGATCTGTATATGAGTGGGTTTCGCATCTTCTGGATCATAGG
ACTTGAGAACATACAAAACAACATAAATCAGAAAAGTTGCCAGCAATCACTTTAGTTTTCTGCAAGCTAGATCCAC
CTGCTACTATCCCTACTTTCTTCTGTTGGGTTGTGAGAGCTCTACCTACAGTTGATTAACCCACACATCAGTCACTG
TTTTTTTACAATTTCCTAATTTCTTACTTTGATTTGCCACTTTTTTTGGCAGGGACTTCTATTTGCCAGTATATATTA
TGTGTTGCATACCTTGGTCA CAGCAAAAATAAAAACAAAAATACGCATGGTAAAAGAAATGCTTCTTTTTTACGCACAA
CGGTAGACGAAGCATATAATTTTACATATCGGAAATGTTCTTGAACAGAGCTTAGTAAATTTATACGCAATATTAGTT
AGTGTATTTAATTTTCCCTTTTAGTTTACCTTGTGTGTAATCAGGGAAAGTGATAATGTATTTAATAATAATTAAT

AGTTTTAATGATACAAATCTATTGCGCAGAAATCCGATGGGTAATTACAACATTATGATTCCTTACATTATAGATACAT
GCAAAATGTATATCTCGACTATAGCGTTCCTCTCTGTTTATGCTACAAATGTAATGTACTAAAGAAAATATTTTATAAA
GTATAAAATTAATATATTATGACATTTGATATATGTCATTTATAACGTGGAAAGAATAATTTATATACATTTGAAAGTAT
CAAATTAGTAAGCACTTTTTTAAAAATCATAATTAACAATTAACAATTCCAATAAAAAATGTTTTTGGGTGCATATTAGT
TTGAAGAATCAAACTTTTAAAAATAGAGTTTGAATTCGAGCATACATACATACATAAGGCAAAATTTAGAGCTTTTTTT
TTCCGGTGCATCGGCAATTTATTTTACTACTACCAGGCGACTGTTGATATTTCCAGTCTCGGGCGGGGCAATTTGTCAA
CTGTGCAAAAAACAACAGAGGAACAAGAGCAGGCATCTTTCTAGAAGATCTCCTACAAATATTTCTCAGCTGCCATGGAAA
ATCGATGTTCTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAACCTTATATTAAGAACTATGCTAACACCTCA
TCAGGAACCGTTGTAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAACCTACGAGCTAAATATTTCAATATGTTT
CTCTTGACCAACTTTATTTCTGCATTTTTTTTGAACGAGGTTTAGAGCAAGCTTCAGGAACTGAGACAGGAATTTATTA
AAAATTTAAATTTTGAAGAAAGTTCAGGGTTAATAGCATCCATTTTTTGTCTTTCGCAAGTTCCTCAGCATTTTAAACAAA
GACGCTCTTTTTGACATGTTTAAAGTTTAACTCCTGTGTGAAATTTGTTATCCGCTCACAAATTCACACATTTATACGAG
CCGGAAGCATAAAGTGTAAAGCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATGCGTTCGCGTTCACATGCAATTT
GCTTTCAGTTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAAATCGGCAACCGCGGGGAGAGGCGGTTTGCATATTGG
GGCTCTTCCGCTTCTCCTCGCTCACTGACTCGCTGCGCTCGGCTGTTCCGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGAGGAAAGAACATGTTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
CGTAAAAAGGCCCGTGTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
GAGGTGGCGAAACCAGCAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCCGA
CCCTGCCGCTTACCGGATACCTGTCCGCTTCTCCTTCCGGGAAAGCGTGGCGCTTTCATAGCTCAGCGTGTAGGTAT
CTCAGTTTCGGTGTAGGTGCTTCCGCTCCAGCTGGGCTGTGTGCACGAACCCCGCTCAGCCCGACCGCTGCGCTTATC
CGGTAACATATCGTCTTGAGTCCAAACCGGTAAGACAGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
GAGCGAGGTATGTAGGCGGTGTACAGAGTTCCTGAAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGT
ATCTGGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAACCCGCTGGTAT
CGGTGGTTTTTTTTGTTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTTTACGG
GGTCTGACGCTCAGTGAACGAAACTCACGTTAAGGGATTTTGGTCAATGAGATTATCAAAAAGGATCTTACCTAGATC
CTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTGCTGACTCCCGTGTGATATAACTACGA
TAGGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCAGACCCAGCTCACCGGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTTGTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGGCGAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTAC
GCTCGTCTGTTTGGTATGGCTTCATTCAGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCTTCCGCTCCTCCGATCGTTGTGCAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGC
ACTGCATAAATCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGTGACTCAACCAAGTCAATCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCTCAATACGGGATAATACCGGCCACATAGCAGAACTTTTAAA
GTGCTCATCATTGGAACCGTTCTTCCGGGCGAAAACTCTCAAGGATCTTACCGCTGTGAGATCCAGTTTCGATGTAACC
CACTCGTGCACCAACTGATCTTCAGCATCTTTACTTTTACCAGGCTTTCTGGGTGAGCAAAAAAGGAAAGGCAAAATG
CCGCAAAAAGGGAATAAGGGCGACACGAAATGTTGAATACTCATACTCTTCTTTTTCAATATATTTGAAGCATTTTAT
CAGGGTTATTGCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAAAATAGGGGTTCCGCGCACATTTCC
CCGAAAAGTCCACCTGACGTCTAAGAAAACCATTTATTTATCATGACATTAACCTATAAAAAATAGGCGTATCAGGAGCC

#name	#location	#size	#directionality	#type
T7 promoter		305..323	19 =>	promoter
5' Homology Arm		372..1744	1373 ==	misc_feature
yellow exon 2 - mCherry fusion		1744..3873	2130 =>	CDS
exon2		1745..3129	1385 ==	misc_feature
Linker (Waldo, 1995)		3130..3162	33 ==	misc_feature
mCherry		3163..3873	711 =>	CDS
3'UTR		3874..4114	241 ==	5'UTR
3' Homology Arm		4115..5235	1121 ==	misc_feature
lac operator		5646..5662	17 ==	protein_bind
lac UV5 promoter		5670..5700	31 <=	promoter
CAP binding site		5715..5736	22 ==	protein_bind
ori		6027..6615	589 <=	rep_origin
AmpR		6786..7646	861 <=	CDS
AmpR promoter		7647..7751	105 <=	promoter

General Discussion

Summary of findings

Morphological diversity results largely from the recycling of the conserved developmental genes. This recycling happens mainly at the level of CREs such as enhancers. Their emergence and their modification during evolution result in new gene expression patterns, in particular for developmental genes. Enhancers are known as short modular DNA segments controlling gene expression. They contain TFBSs bound by TFs in a sequence-specific and tissue-specific manner, and their combinatorial interplay determines the activity of an enhancer.

One of the mechanisms underlying enhancer emergence is the co-option of preexisting regulatory information. While enhancer co-option represents a likely evolutionary path for the birth of new enhancers, its molecular mechanism still remains elusive.

In this thesis, using the *spot* enhancer underlying the evolutionary origin of a novel *spot* expression pattern, I answered two main questions:

1. How has the novel *spot* enhancer evolved in the context of the preexisting *wing blade* enhancer?

For the first question, starting with two qualitatively and arbitrarily defined enhancers, the *spot* and the *wing blade* enhancers, I used a quantitative and systematic approach to map the sequence boundaries of segments driving each full enhancer activity. The results showed that the full *spot* and *wing blade* activities were located in a much larger region (3.5 kb) than previously described (1.1 kb together). The regulatory information both necessary and sufficient for the respective *spot* and *wing blade* activities was extensively overlapping. The results further revealed that a particular site contributed to both activities and was required for the local chromatin accessibility. The results thus demonstrated a case of novel enhancer

evolution by co-option of an ancestral enhancer, and suggested that chromatin accessibility might be one of the molecular components seeding evolutionary co-option.

2. What is the regulatory logic underlying the activity of the minimal *spot*¹⁹⁶ enhancer?

Having identified that the *spot* enhancer evolved through co-option, we sought to understand how the core of this regulatory activity, the short segment named *spot*¹⁹⁶, was built. In this part, we used systematic and quantitative analysis to deconstruct the regulatory logic of the *spot*¹⁹⁶ element and understand how it reads the wing *trans*-regulatory environment to encode a spatial pattern. The results revealed an unexpected density of regulatory information, and also uncovered multiple tiers of repression to produce the spatial pattern of activity.

I will hereafter discuss some aspects of these results.

How to define an enhancer?

By precisely quantifying the spatial enhancer activity, I found that a ~3.5-kb region from the *yellow* 5' drove full activities of both the novel *spot* and preexisting *wing blade* enhancers in *D. biarmipes*. Regulatory sequences required for the *spot* activity also contribute to the *wing blade* activity. Their extensive positional and functional overlap therefore highlight the pleiotropic nature of these enhancers and undermines the conception of enhancer modularity.

Over the 40 years since their discovery, enhancers have been depicted as discrete, arrayed boxes with clearly cut boundaries (Shlyueva, Stampfel, and Stark 2014). Enhancers have been identified by testing the sufficiency, rather than the necessity, of short DNA fragments to drive a similar or identical expression pattern (in a reporter assay) as that of the gene they normally control. The criterion to assess the similar expression, however, is based on qualitative measurements, ignoring the expression levels (Milewski et al. 2004; Corbo, Levine, and Zeller 1997). Therefore, most of the well-studied enhancers likely drive lower levels and imprecise expression patterns compared to those of the endogenous gene.

Our precise quantitative measurement, which takes both levels and pattern into account, enabled us to map all the regulatory sequences necessary and sufficient to recapitulate the faithful enhancer activities.

It is found that compared to the minimal *spot*¹⁹⁶ enhancer (196 bp) identified previously (Gompel et al. 2005), which drove a spot spatial expression pattern but at weak levels, the full *spot* enhancer element (~3.5 kb) drove a much higher expression levels as well as a more precise expression pattern with sharper boundaries on the wing. The sequences flanking minimal enhancers have been implicated in fine-tuning gene expression, conferring the precision of gene expression (Dunipace, Ozdemir, and Stathopoulos 2011). Our findings of the contribution of the flanking sequences thus provided evidence for the above implications.

In addition, flanking sequences were also implicated in the robustness of gene expression against genetic and environmental perturbation (Michael Z. Ludwig et al. 2011; Frankel et al. 2010; Swami 2010). A minimal “stripe element” of the embryonic segmentation gene *even-skipped* in *Drosophila* was shown to be sufficient for the biological function under normal conditions. However, under genetic and environmental perturbation, the flanking sequences are required for robustness of gene expression and fly viability. I speculate that the full *spot* enhancer, which spans ~3.5 kb of the *y* 5' region, might be required for a more robust expression of *yellow*, which therefore can increase the robustness of the wing pigmentation spot.

Our results also indicate that regulatory information required for the full enhancer activities is continuously distributed over a much larger region, rather than clustered into short modular DNA elements. This is consistent with other findings. For example, the regulation of *runt* gene in *Drosophila* embryos is controlled by TFBSs dispersed over a 5 kb segment region (Klingler et al. 1996). A systematic survey of the *Ultrabithorax* (*Ubx*) locus failed to identify the *cis*-regulatory regions of *Ubx* that drive expression on the posterior second femur in *Drosophila*, suggesting that this enhancer structure is complex and regulatory information might be spread over a larger region (Davis et al. 2007). These results call for a reappraisal of our definition of enhancers and a possible change in our methodology when attempting to isolate enhancers.

Pleiotropy in enhancer function

Our precise quantitative analysis of the *yellow* 5' region in *D. biarmipes* provided a revised understanding of the *cis*-regulatory architecture of gene *yellow*. It is surprising to find that the

yellow 5' regulatory region is less modular than previously thought. Rather, the *cis*-regulatory architecture of *yellow 5'* region displays pleiotropy as the same segment and TFBSs affect enhancer activities in both the spot and the wing blade. Actually, I also found that the same segments also affect the enhancer activities in the abdomen (data not shown).

Similar with our results, a recent study found that 5 characterized CREs of *optix* underlying wing color pattern variation in *Heliconius* butterflies are pleiotropic and functionally interdependent. 4 of them are shown to be required for the red color pattern and wing vein development (J. J. Lewis et al. 2019).

Recently, several genome-wide studies have also confirmed that thousands of enhancers function in multiple developmental contexts (Infante et al. 2015; Lonfat et al. 2014; Preger-Ben Noon et al. 2018; McKay and Lieb 2013; Fish, Chen, and Capra 2017; J. J. Lewis et al. 2016; Schep et al. 2016; Vizcaya-Molina et al. 2018). For example, Infante et al. showed that in mouse and Anolis Lizards, many enhancers were used during the limb and phallus development. A specific pleiotropic enhancer *HLEB* of *Tbx4* contributed to the development of both hindlimb and genitalia in mice (Infante et al. 2015b). Lonfat et al. also showed several enhancers were shared by the digits and the genitalia structure (Lonfat et al. 2014). These cases indicate that enhancer pleiotropy is gaining rapid attention and might be pervasive.

When two distinct regulatory activities share a DNA segment, within the shared region, the pleiotropy can simply result from this overlap, although both activities rely on distinct sets of TFBSs. Alternatively, the two activity may share same TFBSs, which deepens the level of enhancer pleiotropy and lead to “site pleiotropy” (Preger-Ben Noon et al. 2018). The further mutational analysis of the *spot*¹⁹⁶ core that I presented here revealed that at least two sets of TFBSs, Dll sites and site [6], whose mutations significantly reduced both the *spot* and the *wing blade* activities, functioned in these two contexts.

Since enhancer pleiotropy is not frequently investigated, and TFBS pleiotropy even less so, few studies have dissected individual pleiotropic enhancer to this resolution. One example concerns enhancers from the *Drosophila shavenbaby* (*svb*) gene. In a comprehensive study of enhancers of *D. melanogaster svb* gene, the same TFBSs were found to be used in the epidermis in embryonic and pupal development in one enhancer, while in another enhancer, distinct sets of TFBSs were used to regulate different expression patterns (Preger-Ben Noon et al. 2018).

In another study, this time on an enhancer of the *Drosophila* gene *poxn*, the same TFBSs for Abd-B and STAT were shown to be required for the development of the adult male genitalia posterior lobe and for the development of the embryonic posterior spiracle (Glassford et al. 2015).

These two examples of TFBS pleiotropy have been shown with enhancers active in distinct tissues and at different developmental stages, where the mutational effect of the same TFBS can be examined by tissue dissection followed by imaging. However, if the two activities to be examined are overlapping in the same tissue and at the same developmental timing, such as the *spot* and the *wing blade* activities in our studies, the investigation of site pleiotropy could be hindered because it is impossible to separate these two activities. Our image quantification analysis can separate the two overlapping expression patterns and attribute each activity to respective DNA sequences. It therefore provides insight into the mechanism of enhancer pleiotropy and site pleiotropy when two activities occur simultaneously in the same tissue. Our approach could inspire other scientists to explore the consequences of pleiotropy on phenotypic variation in similar situations.

Even if a same TFBS is used in distinct regulatory instances, it may be used differently. Because TFs are expressed differentially in a tissue- and time-specific manner (Spitz and Furlong 2012), the same TFBSs can be bound by distinct TFs in different developmental contexts. This is the case for a site in the *Drosophila scute* enhancer. In the genitalia, this site is bound by Abd-B where Abd-B is expressed, while in the leg a single mutation alters its binding for another factor expressed in the leg (Nagy et al. 2018). On the contrary, if the same TFBS is used at the same space and time, where the same TFs are available to the site, the pleiotropic site might respond to a common TF. Therefore, it is tempting to speculate that the pleiotropic site identified in our study might control the *spot* and the *wing blade* activities by responding to a pioneer factor which opens the local chromatin accessibility (Cirillo et al. 2002). However, this speculation needs to be confirmed.

Evolution of novel enhancers from preexisting regulatory sequences

In *D. biarmipes*, the novel *spot* enhancer, which underlies the evolution of a morphological novelty, was found in the vicinity of the preexisting *wing blade* enhancer (Gompel et al. 2005). While the positional proximity suggested a case of enhancer co-option, the scenario of *de novo* emergence could not be ruled out. In this thesis, I revisited the positional relationship of the novel *spot* and the preexisting *wing blade* activities using a quantitative measurement instead, and found that the regulatory sequences required for these two activities overlapped extensively, and the sequences required for the *spot* activity also contributed to the *wing blade* activity. A detailed analysis of the shared core region revealed that at the timing when the *spot* enhancer became active, the chromatin was accessible around the core region, and a shared site was required for this local chromatin accessibility. This has confirmed that the novel *spot* activity has evolved by co-option of sequences from the ancestral *wing blade* element.

Consistent with my results, several examples of new enhancer origin have involved the co-option of preexisting regulatory sequences, and this phenomenon might be common.

A novel expression pattern of the *Nep1* gene in the optic lobe in *D. santomea* has evolved by reusing preexisting regulatory sequences that are active in multiple tissues (Rebeiz et al. 2011). In another species *D. guttifera*, a novel vein-tip expression pattern of the gene *wingless* evolved by co-opting an ancestral enhancer (Koshikawa et al. 2015). The TFBSs for STAT and Abd-B of the *poxn* enhancer active in the ancestral posterior spiracle are reused to evolve a novel expression pattern in the derived posterior lobe structure of the male *Drosophila* genitalia (Glassford et al. 2015). It is conceivable that enhancer co-option is widely spread across the whole genome. Reinforcing this idea, a genome-wide analysis of enhancers in mammalian liver showed a great number of enhancers have recently emerged by exaptation from ancestral DNA sequences (Villar et al. 2015). Compared to *de novo* emergence, the emergence of new enhancers through co-option requires fewer mutational steps and is therefore more likely (Gompel and Prud'homme 2009).

Indeed, enhancer co-option represents a shorter evolutionary path. The emergence of a new enhancer activity requires a correct (functional) combination of TFBSs. Generating such TFBSs by random mutation is a game of chance. By reusing some preexisting regulatory information, in the form of TFBSs already bound by TFs and functional, the new activity can emerge with fewer mutations. Especially when the new and the preexisting activities are in the

same tissue and at the same developmental stage, multiple spatially expressed TFs are available for any new generated TFBS. This could shorten the evolutionary path to build a new activity.

The other important biological reason proposed to explain enhancer co-option is the favorable chromatin environment. From all possible mutations generating new TFBSs, only those occurring in accessible chromatin regions have the chance to be caught by TFs. The increased chromatin accessibility of the preexisting enhancer therefore provides a favorable chromatin environment for the emergence of a new activity. Within a chromatin region that is already accessible, unlike in a compacted chromatin region, any mutation resulting in the formation of new TFBSs could be immediately exposed to the prepatterned spatial inputs and more efficiently be incorporated to the novel regulatory context.

The importance of chromatin accessibility in enhancer evolution has been extensively studied genome-wide using bioinformatics in recent years (Peng et al. 2019; Maeso and Tena 2016). However, no studies have confirmed the role of chromatin accessibility at the molecular level of individual enhancer. The pleiotropic site contributing to both the novel *spot* and the preexisting *wing blade* activities by maintaining the local chromatin accessibility therefore opens a door to investigate the proposed role of chromatin accessibility in enhancer co-option.

Of course, it is still not known whether this open chromatin profile observed is ancestral or derived. Sequence comparison of this pleiotropic site reveals its conservation between different closely related species (not shown), which suggests a conserved function to contribute to the chromatin accessibility. On the other hand, it cannot be ruled out that other sites nearby, such as the sites for Dll, are also necessary for the local chromatin accessibility as no other sites have been tested by ATAC-seq. It is also possible that the pleiotropic site [6] is necessary but not sufficient for the chromatin opening. To solve these open questions, ATAC-seq on the spotted and outgroup species as well as on constructs with additional mutations are needed.

The regulatory logic of enhancer activities

Previous work has shown that the *spot*¹⁹⁶ element contains at least 4 TFBSs for the activator Dll and one TFBS for the repressor En, and these two inputs appear sufficient to explain the spatial *spot* activity in the wing (Gompel et al. 2005; Arnoult et al. 2013). However, combining

TFBSs for these factors on a naive sequence failed to produce the spatial spot activity (B. Prud'homme and N. Gompel unpublished results). This suggests that other sequences are required for the spot activity. Yet the function of the remaining sequences between the known TFBSs has not been explored.

We used systematic mutation of the *yellow spot¹⁹⁶* enhancer and quantified the effect of each mutated position on the enhancer activity. Our results demonstrate an unexpected density of regulatory information along the *spot¹⁹⁶* enhancer. Other than the characterized Dll and En TFBSs, all other mutated positions are shown to significantly contribute to the enhancer activity. Moreover, the results demonstrate that the spatial spot pattern is achieved by a complex interplay between activators and multiple tiers of repressors.

An enhancer contains clusters of TFBSs as well as the remaining sequences between the TFBSs. Deciphering the regulatory logic of an enhancer therefore requires understanding how all these sequences within the enhancer encode the gene expression output. Our results here shed light on several aspects of how a spatial quantitative pattern is encoded.

A high density of regulatory information along the *spot¹⁹⁶* sequence

Our quantitative analysis of the activities from the enhancer variants have demonstrated that there is a lot more information in the *spot¹⁹⁶* sequence than just the characterized TFBSs, and the regulatory logic of this enhancer *in vivo* is much more complex.

Most studies to decipher the enhancer “grammar” have focused on the necessity of the characterized TFBSs for activators and repressors (Weingarten-Gabbay and Segal 2014; Spitz and Furlong 2012; Barolo 2016). The remaining sequences between the TFBSs are less examined. These remaining sequences turn out to be necessary for the enhancer function, as TFBSs alone are not sufficient to generate the enhancer activity. Most synthetic enhancers built by combining characterized TFBSs in their original arrangement have failed to recapitulate the native gene expression pattern in transgenic animals (Johnson et al. 2008; Vincent, Estrada, and DePace 2016).

Similar conclusions have been drawn by other studies. For example, Swanson et al. performed an exhaustive dissection of the *sparkling* (*spa*) enhancer of the *Drosophila pax2* gene, and discovered that other than the twelve identified TFBSs, *spa* enhancer is densely packed with previously unknown regulatory sequences which are necessary for normal enhancer activity *in vivo* (Christina I Swanson, Evans, and Barolo 2010). Another similar study found that randomizing the spacer sequences separating the characterized ETS and GATA sites of the *Ciona Otx* gene enhancer affected the level of the enhancer activity (Guérault-Bellone et al. 2017).

Compared to the above studies, as well as others that also reveal the function of the spacing sequence on enhancer activity including those through massively parallel reporter assays in cell lines or embryos (Verfaillie et al. 2016; Inoue and Ahituv 2015; Grossman et al. 2017; Farley et al. 2016; D. M. King et al. 2020; Farley et al. 2015), our study here has revealed several different aspects, which will deepen the current understanding on enhancer regulatory logic. First, unlike most *in vivo* studies that keep the fixed TFBSs, we introduced systematic mutations at all the positions along the enhancer without prior assumption on the role of sequences at each position. Second, instead of using randomized sequences which could potentially introduce unexpected TFBSs, we maximized the disruption of sequence information by introducing A-tracts. Thus, each enhancer mutant is identical to the wild-type except for the replaced segment. Third, our quantitative approach enables us to directly quantify the spatial enhancer activity in the tissue, which can better discriminate the distinct sequence functions on enhancer activity (*e.g.*, pattern or levels. More discussion below). In this respect, our method can be adapted to other systems to better understand the regulatory logic of different sets of enhancers.

Of course, it cannot be ruled out that the replaced A-tracts might influence TF-TF interactions and might also create unknown TFBSs that could affect the enhancer activity.

Regarding the function of the remaining uncharacterized sequences, it is possible that these sequences might contain functional TFBSs for Dll or other unknown TFs, as ChIP-seq has revealed that a TF can bind multiple different motifs with different intensities *in vivo* (Wang et al. 2012; Yang et al. 2006). Besides, these sequences might also affect the transcriptional rate by affecting DNA looping, chromatin structure or TF-TF interactions (Amit et al. 2011).

The *trans*-regulatory integration along the *spot*¹⁹⁶ enhancer sequence

We used *logRatio* images ($\log ([\text{mutant}] / [+])$) to decipher the possible *trans*-regulatory landscape integrated into *spot*¹⁹⁶ enhancer sequence. *logRatio* images can visually reveal how much a mutant affects the enhancer activity across the wing proportionally to the local activity levels, and can reflect logically the spatial distribution of the inputs received and integrated along the *spot*¹⁹⁶ sequence.

The *logRatio* images reveal two classes of spatial information integrated by each position along the *spot*¹⁹⁶: patterned activation/repression and uniform activation/repression.

For the 2 known patterned TFs Dll and En (Gompel et al. 2005; Arnoult et al. 2013), our *logRatio* images have successfully shown that the mutants of Dll and En TFBSs proportionally affect the activity in the wing that correlate with their distribution and are consistent with their effect.

Interestingly, the *logRatio* of mutant [6] reveals a stronger, more uniform decrease in enhancer activity across the wing. In another study we have shown that this site is required for the local chromatin accessibility (Xin et al. 2020). This suggests that site [6] most probably integrates chromatin regulators such as pioneer factors (Zaret and Carroll 2011a; Iwafuchi-Doi 2019), which uniformly express in the wing, to activate the *spot*¹⁹⁶ activity. It might also be the same case for site [5] whose *logRatio* reveals a stronger, more uniformly increased effect on the wing. It is tempting to speculate that site [5] might integrate chromatin regulators with uniform expression in the wing to repress the *spot*¹⁹⁶ activity. If this is the case, then the DNA sequence encompassing site [5] and site [6] will be required for a proper chromatin environment to maintain a correct activity of *spot*¹⁹⁶. To find out, whether chromatin regulators bind to this region will require further experiments. Previous work (Arnoult et al. 2013) with RNAi screens on wing TFs has identified the chromatin regulator Trithorax (TRX) (Kassis, Kennison, and Tamkun 2017) as an activator for the *spot*¹⁹⁶ activity. To test whether TRX could function through interactions with site [6] or other sites nearby, further work such as CHIP-seq on TRX or histone marks (*e.g.*, H3K4me1) (Tie et al. 2014) will be required.

Since the activation of enhancers requires an open chromatin, we speculate that the reason why most synthetic enhancers are not functional might be because they lack such accessibility sites. Consistent with this, Crocker et al. constructed a fully synthetic enhancer platform in *Drosophila* embryos. They confirmed that activation of synthetic enhancers requires the pioneer factor Zelda to establish an open chromatin state (Crocker, Tsai, and Stern 2017). In the *Drosophila* embryo, Zelda is expressed ubiquitously and appears to increase enhancer accessibility (Foo et al. 2014; Harrison et al. 2011; Schulz et al. 2015; Xu et al. 2014).

Overall, the analysis of the mutational effects by *logRatio* images suggest that more (possibly 6 to 8) factors might regulate the *spot¹⁹⁶* enhancer. These factors include activators for the uniform wing blade expression and for expression in the wing vein. These factors possibly work in concert for the construction of the *spot¹⁹⁶* activity. Some factors such as Dll and En can contribute to the patterning of the spatial spot pattern, while some others might be required for the correct TF-TFBS binding, chromatin state or the 3D DNA structure which could affect enhancer-promoter interactions.

Evolution of the spatial spot expression pattern by multiple tiers of repression

Our results from the systematic and combinatorial dissections of the *spot¹⁹⁶* enhancer demonstrate that the spatial spot expression pattern mainly results from multiple tiers of repression rather than from local activation. The *spot¹⁹⁶* enhancer mainly integrates two unknown activators, one promoting a uniform expression in the wing blade, and the other activating activity in the wing veins. This activation is in turn repressed by an unknown repressor throughout the wing. Upon this first two regulatory layers, the spatial spot expression pattern is established by two spatially localized repressors, resulting in a spatially limited repression to the anterior distal region of the wing whose boundary is confined by another repressor in the posterior compartment of the wing.

This result is consistent with the idea that spatial enhancer activity is often determined by the localized activity of transcriptional repressors (Mannervik et al. 1999; Spitz and Furlong 2012), and this is confirmed by multiple studies from the dissection of regulatory enhancers (E. H. Davidson 2006; Ferrándiz, Liljegren, and Yanofsky 2000; Roeder, Ferrándiz, and Yanofsky 2003; Stanojevic, Small, and Levine 1991). For example, the spatial activity of the *Drosophila*

even-skipped (eve) stripe 2 enhancer is activated by broadly expressed Bicoid and Hunchback, and repressed by Giant on the anterior and Krüppel on the posterior region of the embryo (Stanojevic, Small, and Levine 1991).

The spatial spot expression pattern has evolved from the preexisting uniform activity in the wing blade. These two activities share at least one pleiotropic binding site which integrates uniformly expressed activator(s) (Gompel et al. 2005; Xin et al. 2020). To avoid pleiotropic effects, evolution of binding sites for spatially and/or temporally restricted repressors might serve as a possible mechanism for a novel spot expression pattern to emerge from the uniform activity in the wing blade.

Given that the evolutionary rate of gene repression is significantly higher than the rate of gene activation (Oakley, Østman, and Wilson 2006), recruitment of regulatory repression might be a general rule for enhancer and phenotypic evolution.

Several studies have provided evidence that the evolution of differential repression often involves changes in gene expression patterns and morphological diversities (Ronshaugen, McGinnis, and McGinnis 2002; Carroll, Weatherbee, and Langeland 1995; Ochi et al. 2012; Preger-Ben Noon, Davis, and Stern 2016). In *D. sechellia*, the evolution of a spatially restricted repressor Abrupt for a *shavenbaby (svb)* gene enhancer results in the complete loss of the gene expression in the embryo and thus contributes to morphological evolution (Preger-Ben Noon, Davis, and Stern 2016). Similarly, the positioning of insect wings on two thoracic segment was shown to have evolved through the repression of the wing developmental program in different segments by different homeotic genes (Carroll, Weatherbee, and Langeland 1995).

References

- Abzhanov, Arhat, Winston P Kuo, Christine Hartmann, B Rosemary Grant, Peter R Grant, and Clifford J Tabin. 2006. “The Calmodulin Pathway and Evolution of Elongated Beak Morphology in Darwin’s Finches.” *Nature* 442 (7102): 563–67. <https://doi.org/10.1038/nature04843>.
- Abzhanov, Arhat, Meredith Protas, B Rosemary Grant, Peter R Grant, and Clifford J Tabin. 2004. “Bmp4 and Morphological Variation of Beaks in Darwin’s Finches.” *Science* 305 (5689): 1462 LP – 1465. <https://doi.org/10.1126/science.1098095>.
- Adachi, Yoshitsugu, Bernd Hauck, Jason Clements, Hiroshi Kawauchi, Mitsuhiko Kurusu, Yoko Totani, Yuan Yuan Kang, et al. 2003. “Conserved Cis-Regulatory Modules Mediate Complex Neural Expression Patterns of the Eyeless Gene in the Drosophila Brain.” *Mechanisms of Development* 120 (10): 1113–26. <https://doi.org/https://doi.org/10.1016/j.mod.2003.08.007>.
- Akam, M E. 1983. “The Location of Ultrabithorax Transcripts in Drosophila Tissue Sections.” *The EMBO Journal* 2 (11): 2075–84. <https://pubmed.ncbi.nlm.nih.gov/6416829>.
- Amit, Roei, Hernan G Garcia, Rob Phillips, and Scott E Fraser. 2011. “Building Enhancers from the Ground up: A Synthetic Biology Approach.” *Cell* 146 (1): 105–18. <https://doi.org/10.1016/j.cell.2011.06.024>.
- Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. “An Atlas of Active Enhancers across Human Cell Types and Tissues.” *Nature* 507 (7493): 455–61. <https://doi.org/10.1038/nature12787>.
- Arnold, Cosmas D, Daniel Gerlach, Christoph Stelzer, Łukasz M Boryń, Martina Rath, and Alexander Stark. 2013. “Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-Seq.” *Science* 339 (6123): 1074 LP – 1077. <https://doi.org/10.1126/science.1232542>.
- Arnold, M I, and E H Davidson. 1997. “The Hardwiring of Development: Organization and Function of Genomic Regulatory Systems.” *Development* 124 (10): 1851 LP – 1864. <http://dev.biologists.org/content/124/10/1851.abstract>.
- Arnosti, David N, and Meghana M Kulkarni. 2005. “Transcriptional Enhancers: Intelligent Enhanceosomes or Flexible Billboards?” *Journal of Cellular Biochemistry* 94 (5): 890–98. <https://doi.org/10.1002/jcb.20352>.
- Arnoult, Laurent, Kathy F.Y. Su, Diogo Manoel, Caroline Minervino, Justine Magriña, Nicolas Gompel, and Benjamin Prud’homme. 2013. “Emergence and Diversification of Fly Pigmentation through Evolution of a Gene Regulatory Module.” *Science* 339 (6126): 1423–26. <https://doi.org/10.1126/science.1233749>.
- Attanasio, Catia, Alex S Nord, Yiwen Zhu, Matthew J Blow, Zirong Li, Denise K Liberton, Harris Morrison, et al. 2013. “Fine Tuning of Craniofacial Morphology by Distant-Acting Enhancers.” *Science* 342 (6157): 1241006. <https://doi.org/10.1126/science.1241006>.
- Bandyopadhyay, Amitabha, Kunikazu Tsuji, Karen Cox, Brian D Harfe, Vicki Rosen, and Clifford J Tabin. 2006. “Genetic Analysis of the Roles of BMP2, BMP4, and BMP7 in

- Limb Patterning and Skeletogenesis.” *PLOS Genetics* 2 (12): e216. <https://doi.org/10.1371/journal.pgen.0020216>.
- Banerji, Julian, Sandro Rusconi, and Walter Schaffner. 1981. “Expression of a β -Globin Gene Is Enhanced by Remote SV40 DNA Sequences.” *Cell* 27 (2 PART 1): 299–308. [https://doi.org/10.1016/0092-8674\(81\)90413-X](https://doi.org/10.1016/0092-8674(81)90413-X).
- Barolo, Scott. 2016. “How to Tune an Enhancer.” *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. <https://doi.org/10.1073/pnas.1606109113>.
- Barrière, Antoine, Kacy L Gordon, and Ilya Ruvinsky. 2011. “Distinct Functional Constraints Partition Sequence Conservation in a Cis-Regulatory Element.” *PLOS Genetics* 7 (6): e1002095. <https://doi.org/10.1371/journal.pgen.1002095>.
- Becker, Peter B, and Jerry L Workman. 2013. “Nucleosome Remodeling and Epigenetics.” *Cold Spring Harbor Perspectives in Biology* 5 (9): a017905. <https://doi.org/10.1101/cshperspect.a017905>.
- Bedau, Mark A. 2009. “The Evolution of Complexity BT - Mapping the Future of Biology: Evolving Concepts and Theories.” In , edited by Anouk Barberousse, Michel Morange, and Thomas Pradeu, 111–30. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-9636-5_8.
- Belting, Heinz-Georg, Cooduvalli S Shashikant, and Frank H Ruddle. 1998. “Modification of Expression and Cis -Regulation of Hoxc8 in the Evolution of Diverged Axial Morphology.” *Proceedings of the National Academy of Sciences* 95 (5): 2355 LP – 2360. <https://doi.org/10.1073/pnas.95.5.2355>.
- Biessmann, H. 1985. “Molecular Analysis of the Yellow Gene (y) Region of *Drosophila Melanogaster*.” *Proceedings of the National Academy of Sciences* 82 (21): 7369 LP – 7373. <https://doi.org/10.1073/pnas.82.21.7369>.
- Blow, Matthew J, David J McCulley, Zirong Li, Tao Zhang, Jennifer A Akiyama, Amy Holt, Ingrid Plajzer-Frick, et al. 2010. “ChIP-Seq Identification of Weakly Conserved Heart Enhancers.” *Nature Genetics* 42 (9): 806–10. <https://doi.org/10.1038/ng.650>.
- Bonn, Stefan, Robert P Zinzen, Charles Girardot, E Hilary Gustafson, Alexis Perez-Gonzalez, Nicolas Delhomme, Yad Ghavi-Helm, Bartek Wilczyński, Andrew Riddell, and Eileen E M Furlong. 2012. “Tissue-Specific Analysis of Chromatin State Identifies Temporal Signatures of Enhancer Activity during Embryonic Development.” *Nature Genetics* 44 (2): 148–56. <https://doi.org/10.1038/ng.1064>.
- Britten, Roy J, and Eric H Davidson. 1969. “Gene Regulation for Higher Cells: A Theory.” *Science* 165 (3891): 349 LP – 357. <https://doi.org/10.1126/science.165.3891.349>.
- Buenrostro, Jason D, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. 2013. “Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position.” *Nature Methods* 10 (12): 1213–18. <https://doi.org/10.1038/nmeth.2688>.
- Burgess-Beusse, Bonnie, Catherine Farrell, Miklos Gaszner, Michael Litt, Vesco Mutskov, Felix Recillas-Targa, Melanie Simpson, Adam West, and Gary Felsenfeld. 2002. “The Insulation of Genes from External Enhancers and Silencing Chromatin.” *Proceedings of the National Academy of Sciences of the United States of America* 99 Suppl 4 (Suppl 4): 16433–37. <https://doi.org/10.1073/pnas.162342499>.
- Butler, J E, and J T Kadonaga. 2001. “Enhancer-Promoter Specificity Mediated by DPE or TATA Core Promoter Motifs.” *Genes & Development* 15 (19): 2515–19. <https://doi.org/10.1101/gad.924301>.
- Camino, Eric M, John C Butts, Alison Ordway, Jordan E Vellky, Mark Rebeiz, and Thomas M Williams. 2015. “The Evolutionary Origination and Diversification of a Dimorphic

- Gene Regulatory Network through Parallel Innovations in Cis and Trans.” *PLoS Genetics* 11 (4): e1005136–e1005136. <https://doi.org/10.1371/journal.pgen.1005136>.
- Cande, Jessica Doran, Vivek S Chopra, and Michael Levine. 2009. “Evolving Enhancer-Promoter Interactions within the Tinman Complex of the Flour Beetle, *Tribolium Castaneum*.” *Development (Cambridge, England)* 136 (18): 3153–60. <https://doi.org/10.1242/dev.038034>.
- Carroll, Sean B. 2008. “Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution.” *Cell* 134 (1): 25–36. <https://doi.org/10.1016/j.cell.2008.06.030>.
- Carroll, Sean B. 1995. “Homeotic Genes and the Evolution of Arthropods and Chordates.” *Nature* 376 (6540): 479–85. <https://doi.org/10.1038/376479a0>.
- . 2000. “Endless Forms: The Evolution Minireview of Gene Regulation and Morphological Diversity The Similarities in Gene Content among Long-Diverged Phyla Did Not Meet Initial Expectations That the Expansion of Gene Families Would Track the Evolution of Later, M.” *Cell* 101: 577–80. [http://www.cell.com/cell/pdf/S0092-8674\(00\)80868-5.pdf](http://www.cell.com/cell/pdf/S0092-8674(00)80868-5.pdf).
- . 2005. “Evolution at Two Levels: On Genes and Form.” *PLOS Biology* 3 (7): e245. <https://doi.org/10.1371/journal.pbio.0030245>.
- Carroll, Sean B, Scott D Weatherbee, and James A Langeland. 1995. “Homeotic Genes and the Regulation and Evolution of Insect Wing Number.” *Nature* 375 (6526): 58–61. <https://doi.org/10.1038/375058a0>.
- Castro, Juliana P, and Claudia M A Carareto. 2004. “Drosophila Melanogaster P Transposable Elements: Mechanisms of Transposition and Regulation.” *Genetica* 121 (2): 107–18. <https://doi.org/10.1023/B:GENE.0000040382.48039.a2>.
- Chan, Yingguang Frank, Melissa E Marks, Felicity C Jones, Guadalupe Villarreal Jr, Michael D Shapiro, Shannon D Brady, Audrey M Southwick, et al. 2010. “Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer.” *Science (New York, N.Y.)* 327 (5963): 302–5. <https://doi.org/10.1126/science.1182213>.
- Chao, Chung-Hao, Horng-Dar Wang, and Chiou-Hwa Yuh. 2010. “Complexity of Cis-Regulatory Organization of Six3a during Forebrain and Eye Development in Zebrafish.” *BMC Developmental Biology* 10 (March): 35. <https://doi.org/10.1186/1471-213X-10-35>.
- Chiocchetti, Annalisa, Emanuela Tolosano, Emilio Hirsch, Lorenzo Silengo, and Fiorella Altruda. 1997. “Green Fluorescent Protein as a Reporter of Gene Expression in Transgenic Mice.” *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1352 (2): 193–202. [https://doi.org/https://doi.org/10.1016/S0167-4781\(97\)00010-9](https://doi.org/https://doi.org/10.1016/S0167-4781(97)00010-9).
- Cirillo L.A., Lin F R Cuesta I Friedman D Jarnik M Zaret K S, Lisa Ann Cirillo, Frank Robert Lin, Isabel Cuesta, Dara Friedman, Michal Jarnik, and Kenneth S Zaret. 2002. “Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4.” *Molecular Cell* 9 (2): 279–89. <http://www.ncbi.nlm.nih.gov/pubmed/11864602>.
- Cirillo, Lisa Ann, Frank Robert Lin, Isabel Cuesta, Dara Friedman, Michal Jarnik, and Kenneth S Zaret. 2002. “Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4.” *Molecular Cell* 9 (2): 279–89. [https://doi.org/https://doi.org/10.1016/S1097-2765\(02\)00459-8](https://doi.org/https://doi.org/10.1016/S1097-2765(02)00459-8).
- COOK, LAURENCE. 1998. “Melanism. Evolution in Action. By M. E. N. MAJERUS. Oxford University Press, Oxford 1998. 338+xiii Pages. Price £23.95 (Paperback). ISBN 0 19 854982 2.” *Genetical Research* 72 (1): 73–75. <https://doi.org/DOI:10.1017/S0016672398229739>.
- Coons, Albert H, Hugh J Creech, and R Norman Jones. 1941. “Immunological Properties of an Antibody Containing a Fluorescent Group.” *Proceedings of the Society for Experimental*

- Arabidopsis Fruit Development.” *Science* 289 (5478): 436 LP – 438. <https://doi.org/10.1126/science.289.5478.436>.
- Feschotte, Cédric. 2008. “Transposable Elements and the Evolution of Regulatory Networks.” *Nature Reviews Genetics* 9 (5): 397–405. <https://doi.org/10.1038/nrg2337>.
- Fish, Alexandra, Ling Chen, and John A. Capra. 2017. “Gene Regulatory Enhancers with Evolutionarily Conserved Activity Are More Pleiotropic than Those with Species-Specific Activity.” *Genome Biology and Evolution* 9 (10): 2615–25. <https://doi.org/10.1093/gbe/evx194>.
- Foo, Sun Melody, Yujia Sun, Bomyi Lim, Ruta Ziukaite, Kevin O’Brien, Chung-Yi Nien, Nikolai Kirov, Stanislav Y. Shvartsman, and Christine A. Rushlow. 2014. “Zelda Potentiates Morphogen Activity by Increasing Chromatin Accessibility.” *Current Biology* 24 (12): 1341–46. <https://doi.org/https://doi.org/10.1016/j.cub.2014.04.032>.
- Foote, Mike. 1997. “The Evolution of Morphological Diversity.” *Annual Review of Ecology and Systematics* 28 (1): 129–52. <https://doi.org/10.1146/annurev.ecolsys.28.1.129>.
- Frankel, Nicolás, Gregory K Davis, Diego Vargas, Shu Wang, François Payre, and David L Stern. 2010. “Phenotypic Robustness Conferred by Apparently Redundant Transcriptional Enhancers.” *Nature* 466 (7305): 490–93. <https://doi.org/10.1038/nature09158>.
- Frankel, Nicolás, Deniz F Erezyilmaz, Alistair P McGregor, Shu Wang, François Payre, and David L Stern. 2011. “Morphological Evolution Caused by Many Subtle-Effect Substitutions in Regulatory DNA.” *Nature* 474 (7353): 598–603. <https://doi.org/10.1038/nature10200>.
- Gall, Joseph G, and Mary Lou Pardue. 1969. “FORMATION AND DETECTION OF RNA-DNA HYBRID MOLECULES IN CYTOLOGICAL PREPARATIONS.” *Proceedings of the National Academy of Sciences* 63 (2): 378 LP – 383. <https://doi.org/10.1073/pnas.63.2.378>.
- Gaunt, Stephen J., and Yu Lee Paul. 2012. “Changes in Cis-Regulatory Elements during Morphological Evolution.” *Biology* 1 (3): 557–74. <https://doi.org/10.3390/biology1030557>.
- Gilbert, Scott F. 2003. “The Morphogenesis of Evolutionary Developmental Biology.” *International Journal of Developmental Biology* 47 (7–8): 467–77. <https://doi.org/10.1387/ijdb.14756322>.
- Gilbert, Scott F, John M Opitz, and Rudolf A Raff. 1996. “Resynthesizing Evolutionary and Developmental Biology.” *Developmental Biology* 173 (2): 357–72. <https://doi.org/https://doi.org/10.1006/dbio.1996.0032>.
- Giresi, Paul G, Jonghwan Kim, Ryan M McDaniell, Vishwanath R Iyer, and Jason D Lieb. 2007. “FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin.” *Genome Research* 17 (6): 877–85. <https://doi.org/10.1101/gr.5533506>.
- Glassford, William J., Winslow C. Johnson, Natalie R. Dall, Sarah Jacquelyn Smith, Yang Liu, Werner Boll, Markus Noll, and Mark Rebeiz. 2015. “Co-Option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty.” *Developmental Cell* 34 (5): 520–31. <https://doi.org/10.1016/j.devcel.2015.08.005>.
- Gomez-Skarmeta, J. L., I. Rodriguez, C. Martinez, J. Culi, D. Ferres-Marco, D. Beamonte, and J. Modolell. 1995. “Cis-Regulation of Achaete and Scute: Shared Enhancer-like Elements Drive Their Coexpression in Proneural Clusters of the Imaginal Discs.” *Genes and Development* 9 (15): 1869–82. <https://doi.org/10.1101/gad.9.15.1869>.
- Gompel, Nicolas, and Benjamin Prud’homme. 2009. “The Causes of Repeated Genetic Evolution.” *Developmental Biology* 332 (1): 36–47. <https://doi.org/https://doi.org/10.1016/j.ydbio.2009.04.040>.
- Gompel, Nicolas, Benjamin Prud’homme, Patricia J Wittkopp, Victoria A Kassner, and Sean

- B Carroll. 2005. “Chance Caught on the Wing: Cis-Regulatory Evolution and the Origin of Pigment Patterns in *Drosophila*.” *Nature* 433 (7025): 481–87. <https://doi.org/10.1038/nature03235>.
- Goryshin, Igor Yu, and William S Reznikoff. 1998. “Tn5 in Vitro Transposition .” *Journal of Biological Chemistry* 273 (13): 7367–74. <https://doi.org/10.1074/jbc.273.13.7367>.
- Grossman, Sharon R, Xiaolan Zhang, Li Wang, Jesse Engreitz, Alexandre Melnikov, Peter Rogov, Ryan Tewhey, et al. 2017. “Systematic Dissection of Genomic Features Determining Transcription Factor Binding and Enhancer Function.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (7): E1291–1300. <https://doi.org/10.1073/pnas.1621150114>.
- Groth, Amy C, Matthew Fish, Roel Nusse, and Michele P Calos. 2004. “Construction of Transgenic *Drosophila* by Using the Site-Specific Integrase from Phage PhiC31.” *Genetics* 166 (4): 1775–82. <https://doi.org/10.1534/genetics.166.4.1775>.
- Groth, Amy C, Eric C Olivares, Bhaskar Thyagarajan, and Michele P Calos. 2000. “A Phage Integrase Directs Efficient Site-Specific Integration in Human Cells.” *Proceedings of the National Academy of Sciences* 97 (11): 5995 LP – 6000. <https://doi.org/10.1073/pnas.090527097>.
- Guérault-Bellone, Marion, Kazuhiro R Nitta, Willi Kari, Edwin Jacox, Rémy Beulé Dautat, Renaud Vincentelli, Carine Diarra, et al. 2017. “Spacer Sequences Separating Transcription Factor Binding Motifs Set Enhancer Quality and Strength.” *BioRxiv*, January, 98830. <https://doi.org/10.1101/098830>.
- Hales, Karen G., Christopher A. Korey, Amanda M. Larracunte, and David M. Roberts. 2015. “Genetics on the Fly: A Primer on the *Drosophila* Model System.” *Genetics* 201 (3): 815–42. <https://doi.org/10.1534/genetics.115.183392>.
- Hall, Brian K. 2012. “Evolutionary Developmental Biology (Evo-Devo): Past, Present, and Future.” *Evolution: Education and Outreach* 5 (2): 184–93. <https://doi.org/10.1007/s12052-012-0418-x>.
- Hare, Emily E, Brant K Peterson, and Michael B Eisen. 2008. “A Careful Look at Binding Site Reorganization in the Even-Skipped Enhancers of *Drosophila* and Sepsids.” *PLOS Genetics* 4 (11): e1000268. <https://doi.org/10.1371/journal.pgen.1000268>.
- Harrison, Melissa M, Xiao-Yong Li, Tommy Kaplan, Michael R Botchan, and Michael B Eisen. 2011. “Zelda Binding in the Early *Drosophila Melanogaster* Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition.” *PLOS Genetics* 7 (10): e1002266. <https://doi.org/10.1371/journal.pgen.1002266>.
- Heintzman, Nathaniel D, Gary C Hon, R David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F Harp, Zhen Ye, et al. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459 (7243): 108–12. <https://doi.org/10.1038/nature07829>.
- Heintzman, Nathaniel D, and Bing Ren. 2009. “Finding Distal Regulatory Elements in the Human Genome.” *Current Opinion in Genetics & Development* 19 (6): 541–49. <https://doi.org/10.1016/j.gde.2009.09.006>.
- Hesselberth, Jay R, Xiaoyu Chen, Zhihong Zhang, Peter J Sabo, Richard Sandstrom, Alex P Reynolds, Robert E Thurman, et al. 2009. “Global Mapping of Protein-DNA Interactions in Vivo by Digital Genomic Footprinting.” *Nature Methods* 6 (4): 283–89. <https://doi.org/10.1038/nmeth.1313>.
- Hickman, Cleveland P. 2018. *Animal Diversity*.
- Huang, Yi-Fei, Brad Gulko, and Adam Siepel. 2017. “Fast, Scalable Prediction of Deleterious Noncoding Variants from Functional and Population Genomic Data.” *Nature Genetics* 49 (4): 618–24. <https://doi.org/10.1038/ng.3810>.
- Hueber, Stefanie D, and Ingrid Lohmann. 2008. “Shaping Segments: Hox Gene Function in

- the Genomic Age.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 30 (10): 965–79. <https://doi.org/10.1002/bies.20823>.
- Imsland, Freyja, Chungang Feng, Henrik Boije, Bertrand Bed’hom, Valérie Fillon, Ben Dorshorst, Carl-Johan Rubin, et al. 2012. “The Rose-Comb Mutation in Chickens Constitutes a Structural Rearrangement Causing Both Altered Comb Morphology and Defective Sperm Motility.” *PLoS Genetics* 8 (6): e1002775–e1002775. <https://doi.org/10.1371/journal.pgen.1002775>.
- Indjeian, Vahan B, Garrett A Kingman, Felicity C Jones, Catherine A Guenther, Jane Grimwood, Jeremy Schmutz, Richard M Myers, and David M Kingsley. 2016. “Evolving New Skeletal Traits by Cis-Regulatory Changes in Bone Morphogenetic Proteins.” *Cell* 164 (1–2): 45–56. <https://doi.org/10.1016/j.cell.2015.12.007>.
- Infante, Carlos R., Alexandra G. Mihala, Sungdae Park, Jialiang S. Wang, Kenji K. Johnson, James D. Lauderdale, and Douglas B. Menke. 2015a. “Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-Genital *cis*-Regulatory Element in Snakes.” *Developmental Cell* 35 (1): 107–19. <https://doi.org/10.1016/j.devcel.2015.09.003>.
- Infante, Carlos R., Alexandra G. Mihala, Sungdae Park, Jialiang S. Wang, Kenji K. Johnson, James D. Lauderdale, and Douglas B. Menke. 2015b. “Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-Genital *Cis*-Regulatory Element in Snakes.” *Developmental Cell* 35 (1): 107–19. <https://doi.org/10.1016/j.devcel.2015.09.003>.
- Inoue, Fumitaka, and Nadav Ahituv. 2015. “Decoding Enhancers Using Massively Parallel Reporter Assays.” *Genomics* 106 (3): 159–64. <https://doi.org/https://doi.org/10.1016/j.ygeno.2015.06.005>.
- Irvine, Steven Q, Vera C Fonseca, Michael A Zompa, and Rajee Antony. 2008. “Cis-Regulatory Organization of the Pax6 Gene in the Ascidian *Ciona intestinalis*.” *Developmental Biology* 317 (2): 649–59. <https://doi.org/10.1016/j.ydbio.2008.01.036>.
- Istrail, Sorin, and Eric H Davidson. 2005. “Logic Functions of the Genomic *Cis*-Regulatory Code.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (14): 4954 LP – 4959. <https://doi.org/10.1073/pnas.0409624102>.
- Iwafuchi-Doi, Makiko. 2019. “The Mechanistic Basis for Chromatin Regulation by Pioneer Transcription Factors.” *WIREs Systems Biology and Medicine* 11 (1): e1427. <https://doi.org/10.1002/wsbm.1427>.
- Iwafuchi-Doi, Makiko, and Kenneth S Zaret. 2014. “Pioneer Transcription Factors in Cell Reprogramming.” *Genes & Development* 28 (24): 2679–92. <https://doi.org/10.1101/gad.253443.114>.
- Jang, Chuen-Chuen, Ju-Lan Chao, Nikolas Jones, Li-Chin Yao, Dmitri A Bessarab, Yien M Kuo, Susie Jun, Claude Desplan, Steven K Beckendorf, and Y Henry Sun. 2003. “Two Pax Genes, *Eye* and *Eyeless*, Act Cooperatively in Promoting *Drosophila* Eye Development.” *Development* 130 (13): 2939 LP – 2951. <https://doi.org/10.1242/dev.00522>.
- Jeong, Sangyun, Antonis Rokas, and Sean B. Carroll. 2006. “Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in *Drosophila* Evolution.” *Cell* 125 (7): 1387–99. <https://doi.org/10.1016/j.cell.2006.04.043>.
- Johnson, Lisa A, Ying Zhao, Krista Golden, and Scott Barolo. 2008. “Reverse-Engineering a Transcriptional Enhancer: A Case Study in *Drosophila*.” *Tissue Engineering. Part A* 14 (9): 1549–59. <https://doi.org/10.1089/ten.tea.2008.0074>.
- Junion, Guillaume, Mikhail Spivakov, Charles Girardot, Martina Braun, E. Hilary Gustafson, Ewan Birney, and Eileen E.M. Furlong. 2012. “A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History.” *Cell* 148 (3): 473–86.

- <https://doi.org/10.1016/j.cell.2012.01.030>.
- Juven-Gershon, Tamar, Jer-Yuan Hsu, Joshua Wm Theisen, and James T Kadonaga. 2008. "The RNA Polymerase II Core Promoter - the Gateway to Transcription." *Current Opinion in Cell Biology* 20 (3): 253–59. <https://doi.org/10.1016/j.ceb.2008.03.003>.
- Kadonaga, James T. 1998. "Eukaryotic Transcription: An Interlaced Network of Transcription Factors and Chromatin-Modifying Machines." *Cell* 92 (3): 307–13. [https://doi.org/10.1016/S0092-8674\(00\)80924-1](https://doi.org/10.1016/S0092-8674(00)80924-1).
- Kalay, Gizem. 2012. "Rapid Evolution of Cis-Regulatory Architecture and Activity in the Drosophila Yellow Gene." *ProQuest Dissertations and Theses*, 185. http://ezproxy.nottingham.ac.uk/login?url=https://search.proquest.com/docview/1027750821?accountid=8018%5Cnhttp://sfx.nottingham.ac.uk/sfx_local/?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Pro.
- Kalay, Gizem, Jennifer Lachowicz, Ulises Rosas, Mackenzie R. Dome, and Patricia Wittkopp. 2019. *Redundant and Cryptic Enhancer Activities of the Drosophila Yellow Gene. Genetics*. Vol. 212. <https://doi.org/10.1534/genetics.119.301985>.
- Kassis, Judith A., James A. Kennison, and John W. Tamkun. 2017. "Polycomb and Trithorax Group Genes in Drosophila." *Genetics* 206 (4): 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- Kheradpour, Pouya, Jason Ernst, Alexandre Melnikov, Peter Rogov, Li Wang, Xiaolan Zhang, Jessica Alston, Tarjei S Mikkelsen, and Manolis Kellis. 2013. "Systematic Dissection of Regulatory Motifs in 2000 Predicted Human Enhancers Using a Massively Parallel Reporter Assay." *Genome Research* 23 (5): 800–811. <https://doi.org/10.1101/gr.144899.112>.
- Kheradpour, Pouya, Alexander Stark, Sushmita Roy, and Manolis Kellis. 2007. "Reliable Prediction of Regulator Targets Using 12 Drosophila Genomes." *Genome Research* 17 (12): 1919–31. <https://doi.org/10.1101/gr.7090407>.
- Khoueiry, Pierre, Charles Girardot, Lucia Ciglar, Pei Chen Peng, E. Hilary Gustafson, Saurabh Sinha, and Eileen E.M. Furlong. 2017. "Uncoupling Evolutionary Changes in DNA Sequence, Transcription Factor Occupancy and Enhancer Activity." *ELife* 6: 1–29. <https://doi.org/10.7554/eLife.28440>.
- Khoury, George, and Peter Gruss. 1983. "Enhancer Elements." *Cell* 33 (2): 313–14. [https://doi.org/10.1016/0092-8674\(83\)90410-5](https://doi.org/10.1016/0092-8674(83)90410-5).
- King, Dana M, Clarice Kit Yee Hong, James L Shepherdson, David M Granas, Brett B Maricque, and Barak A Cohen. 2020. "Synthetic and Genomic Regulatory Elements Reveal Aspects of Cis-Regulatory Grammar in Mouse Embryonic Stem Cells." *ELife* 9 (February): e41279. <https://doi.org/10.7554/eLife.41279>.
- King, Mary-Claire, and A C Wilson. 1975. "Evolution at Two Levels in Humans and Chimpanzees." *Science* 188 (4184): 107–16. <http://www.jstor.org/stable/1739875>.
- Kingston, Robert E., and Geeta J. Narlikar. 1999. "ATP-Dependent Remodeling and Acetylation as Regulators of Chromatin Fluidity." *Genes and Development* 13 (18): 2339–52. <https://doi.org/10.1101/gad.13.18.2339>.
- Klingler, Martin, Johnny Soong, Barbara Butler, and J.Peter Gergen. 1996. "Disperse versus Compact Elements for the Regulation of *fruntStripes* in *Drosophila*." *Developmental Biology* 177 (1): 73–84. <https://doi.org/https://doi.org/10.1006/dbio.1996.0146>.
- Koning, A P Jason de, Wanjun Gu, Todd A Castoe, Mark A Batzer, and David D Pollock. 2011. "Repetitive Elements May Comprise Over Two-Thirds of the Human Genome." *PLOS Genetics* 7 (12): e1002384. <https://doi.org/10.1371/journal.pgen.1002384>.
- Kornberg, Roger D. 1977. "Structure of Chromatin." *Annual Review of Biochemistry* 46 (1): 931–54. <https://doi.org/10.1146/annurev.bi.46.070177.004435>.

- Koshikawa, Shigeyuki. 2015. “Enhancer Modularity and the Evolution of New Traits.” *Fly* 9 (4): 155–59. <https://doi.org/10.1080/19336934.2016.1151129>.
- Koshikawa, Shigeyuki, Matt W. Giorgianni, Kathy Vaccaro, Victoria A. Kassner, John H. Yoder, Thomas Werner, and Sean B. Carroll. 2015. “Gain of Cis-Regulatory Activities Underlies Novel Domains of Wingless Gene Expression in *Drosophila*.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (24): 7524–29. <https://doi.org/10.1073/pnas.1509022112>.
- Krivega, Ivan, and Ann Dean. 2012. “Enhancer and Promoter Interactions-Long Distance Calls.” *Current Opinion in Genetics & Development* 22 (2): 79–85. <https://doi.org/10.1016/j.gde.2011.11.001>.
- Krumlauf, R. 1994. “Hox Genes in Vertebrate Development.” *Cell* 78 (2): 191–201. [https://doi.org/10.1016/0092-8674\(94\)90290-9](https://doi.org/10.1016/0092-8674(94)90290-9).
- Kulkarni, Meghana M, and David N Arnosti. 2003. “Information Display by Transcriptional Enhancers.” *Development* 130 (26): 6569 LP – 6575. <https://doi.org/10.1242/dev.00890>.
- Kvon, Evgeny Z. 2015. “Using Transgenic Reporter Assays to Functionally Characterize Enhancers in Animals.” *Genomics* 106 (3): 185–92. <https://doi.org/https://doi.org/10.1016/j.ygeno.2015.06.007>.
- Kvon, Evgeny Z, Gerald Stampfel, J Omar Yáñez-Cuna, Barry J Dickson, and Alexander Stark. 2012. “HOT Regions Function as Patterned Developmental Enhancers and Have a Distinct Cis-Regulatory Signature.” *Genes & Development* 26 (9): 908–13. <https://doi.org/10.1101/gad.188052.112>.
- Kwasnieski, Jamie C, Ilaria Mogno, Connie A Myers, Joseph C Corbo, and Barak A Cohen. 2012. “Complex Effects of Nucleotide Variants in a Mammalian Cis-Regulatory Element.” *Proceedings of the National Academy of Sciences* 109 (47): 19498 LP – 19503. <https://doi.org/10.1073/pnas.1210678109>.
- Lee, Tong Ihn, and Richard A Young. 2000. “TRANSCRIPTION OF EUKARYOTIC PROTEIN-CODING GENES.” *Annual Review of Genetics* 34 (1): 77–137. <https://doi.org/10.1146/annurev.genet.34.1.77>.
- Levine, Michael, and Robert Tjian. 2003. “Transcription Regulation and Animal Diversity.” *Nature* 424 (6945): 147–51. <https://doi.org/10.1038/nature01763>.
- Levine, Mike. 2010. “Transcriptional Enhancers in Animal Development and Evolution.” *Current Biology : CB* 20 (17): R754–63. <https://doi.org/10.1016/j.cub.2010.06.070>.
- Levy, Samuel, and Sridhar Hannenhalli. 2002. “Identification of Transcription Factor Binding Sites in the Human Genome Sequence .” *Mammalian Genome* 13 (9): 510–14. <https://doi.org/10.1007/s00335-002-2175-6>.
- Lewis, E B. 1978. “A Gene Complex Controlling Segmentation in *Drosophila*.” *Nature* 276 (5688): 565–70. <https://doi.org/10.1038/276565a0>.
- Lewis, James J., Karin R.L. van der Burg, Anyi Mazo-Vargas, and Robert D. Reed. 2016. “ChIP-Seq-Annotated *Heliconius Erato* Genome Highlights Patterns of Cis-Regulatory Evolution in Lepidoptera.” *Cell Reports* 16 (11): 2855–63. <https://doi.org/10.1016/j.celrep.2016.08.042>.
- Lewis, James J, Rachel C Geltman, Patrick C Pollak, Kathleen E Rondem, Steven M Van Belleghem, Melissa J Hubisz, Paul R Munn, et al. 2019. “Parallel Evolution of Ancient, Pleiotropic Enhancers Underlies Butterfly Wing Pattern Mimicry.” *Proceedings of the National Academy of Sciences* 116 (48): 24174 LP – 24183. <https://doi.org/10.1073/pnas.1907068116>.
- Li, Wenbo, Dimple Notani, and Michael G. Rosenfeld. 2016. “Enhancers as Non-Coding RNA Transcription Units: Recent Insights and Future Perspectives.” *Nature Reviews Genetics* 17 (4): 207–23. <https://doi.org/10.1038/nrg.2016.4>.
- Lonfat, Nicolas, Thomas Montavon, Fabrice Darbellay, Sandra Gitto, and Denis Duboule.

2014. “Convergent Evolution of Complex Regulatory Landscapes and Pleiotropy at Hox Loci.” *Science* 346 (6212): 1004–6. <https://doi.org/10.1126/science.1257493>.
- Long, Hannah K., Sara L. Prescott, and Joanna Wysocka. 2016. “Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.” *Cell* 167 (5): 1170–87. <https://doi.org/10.1016/j.cell.2016.09.018>.
- Ludwig, M Z, N H Patel, and M Kreitman. 1998. “Functional Analysis of Eve Stripe 2 Enhancer Evolution in Drosophila: Rules Governing Conservation and Change.” *Development* 125 (5): 949 LP – 958. <http://dev.biologists.org/content/125/5/949.abstract>.
- Ludwig, Michael Z., Manu, Ralf Kittler, Kevin P. White, and Martin Kreitman. 2011. “Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness.” *PLoS Genetics* 7 (11). <https://doi.org/10.1371/journal.pgen.1002364>.
- Ludwig, Michael Z, Casey Bergman, Nipam H Patel, and Martin Kreitman. 2000. “Evidence for Stabilizing Selection in a Eukaryotic Enhancer Element.” *Nature* 403 (6769): 564–67. <https://doi.org/10.1038/35000615>.
- Lynch, Vincent J, Robert D Leclerc, Gemma May, and Günter P Wagner. 2011. “Transposon-Mediated Rewiring of Gene Regulatory Networks Contributed to the Evolution of Pregnancy in Mammals.” *Nature Genetics* 43 (11): 1154–59. <https://doi.org/10.1038/ng.917>.
- Madgwick, Alicia, Marta Silvia Magri, Christelle Dantec, Damien Gailly, Ulla Maj Fiuza, Léo Guignard, Sabrina Hettinger, Jose Luis Gomez-Skarmeta, and Patrick Lemaire. 2019. “Evolution of Embryonic Cis-Regulatory Landscapes between Divergent Phallusia and Ciona Ascidiarians.” *Developmental Biology* 448 (2): 71–87. <https://doi.org/10.1016/j.ydbio.2019.01.003>.
- Maeso, Ignacio, and Juan J Tena. 2016. “Favorable Genomic Environments for Cis-Regulatory Evolution: A Novel Theoretical Framework.” *Seminars in Cell & Developmental Biology* 57: 2–10. <https://doi.org/10.1016/j.semcdb.2015.12.003>.
- Mannervik, Mattias, Yutaka Nibu, Hailan Zhang, and Michael Levine. 1999. “Transcriptional Coregulators in Development.” *Science* 284 (5414): 606 LP – 609. <https://doi.org/10.1126/science.284.5414.606>.
- Martin, Tara Laine. 2014. “Assessing Modularity of Developmental Enhancers in Drosophila Melanogaster,” 1–93. <https://doi.org/10.2307/2373099>.
- Massey, J H, and P J Wittkopp. 2016. “Chapter Two - The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species.” In *Genes and Evolution*, edited by Virginie B T - Current Topics in Developmental Biology Orgogozo, 119:27–61. Academic Press. <https://doi.org/https://doi.org/10.1016/bs.ctdb.2016.03.004>.
- McKay, Daniel J., and Jason D. Lieb. 2013. “A Common Set of DNA Regulatory Elements Shapes Drosophila Appendages.” *Developmental Cell* 27 (3): 306–18. <https://doi.org/10.1016/j.devcel.2013.10.009>.
- Melamed, Philippa, Yahav Yosefzon, Sergei Rudnizky, and Lilach Pnueli. 2016. “Transcriptional Enhancers: Transcription, Function and Flexibility.” *Transcription* 7 (1): 26–31. <https://doi.org/10.1080/21541264.2015.1128517>.
- Merika, Menie, and Dimitris Thanos. 2001. “Enhanceosomes.” *Current Opinion in Genetics & Development* 11 (2): 205–8. [https://doi.org/https://doi.org/10.1016/S0959-437X\(00\)00180-5](https://doi.org/https://doi.org/10.1016/S0959-437X(00)00180-5).
- Milewski, Rita C., Neil C. Chi, Jun Li, Christopher Brown, Min Min Lu, and Jonathan A. Epstein. 2004a. “Identification of Minimal Enhancer Elements Sufficient for Pax3 Expression in Neural Crest and Implication of Tead2 as a Regulator of Pax3.” *Development* 131 (4): 829–37. <https://doi.org/10.1242/dev.00975>.
- Milewski, Rita C, Neil C Chi, Jun Li, Christopher Brown, Min Min Lu, and Jonathan A Epstein.

- 2004b. “Identification of Minimal Enhancer Elements Sufficient for Pax3 Expression in Neural Crest and Implication of Tead2 as a Regulator of Pax3.” *Development* 131 (4): 829 LP – 837. <https://doi.org/10.1242/dev.00975>.
- Monteiro, Antónia, and Ondrej Podlaha. 2009. “Wings, Horns, and Butterfly Eyespots: How Do Complex Traits Evolve?” *PLoS Biology* 7 (2): 0209–16. <https://doi.org/10.1371/journal.pbio.1000037>.
- Moreau, P, R Hen, B Wasylyk, R Everett, M P Gaub, and P Chambon. 1981. “The SV40 72 Base Repair Repeat Has a Striking Effect on Gene Expression Both in SV40 and Other Chimeric Recombinants.” *Nucleic Acids Research* 9 (22): 6047–68. <https://doi.org/10.1093/nar.9.22.6047>.
- Morgan, Thomas Hunt, and Calvin B Bridges. 1916. *Sex-Linked Inheritance in Drosophila*. Washington: Carnegie Institution of Washington.
- Mortazavi, Ali, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. 2008. “Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq.” *Nature Methods* 5 (7): 621–28. <https://doi.org/10.1038/nmeth.1226>.
- Muerdter, Felix, Łukasz M Boryń, and Cosmas D Arnold. 2015. “STARR-Seq — Principles and Applications.” *Genomics* 106 (3): 145–50. <https://doi.org/https://doi.org/10.1016/j.ygeno.2015.06.001>.
- Muller, B, and K Basler. 2000. “The Repressor and Activator Forms of Cubitus Interruptus Control Hedgehog Target Genes through Common Generic Gli-Binding Sites.” *Development* 127 (14): 2999 LP – 3007. <http://dev.biologists.org/content/127/14/2999.abstract>.
- Nagy, Olga, Isabelle Nuez, Rosina Savisaar, Alexandre E Peluffo, Amir Yassin, Michael Lang, David L Stern, Daniel R Matute, Jean R David, and Virginie Courtier-Orgogozo. 2018. “Correlated Evolution of Two Copulatory Organs via a Single *cis*-Regulatory Nucleotide Change.” *Current Biology* 28 (21): 3450-3457.e13. <https://doi.org/10.1016/j.cub.2018.08.047>.
- Nitta, Kazuhiro R., Arttu Jolma, Yimeng Yin, Ekaterina Morgunova, Teemu Kivioja, Junaid Akhtar, Korneel Hens, et al. 2015. “Conservation of Transcription Factor Binding Specificities across 600 Million Years of Bilateria Evolution.” *ELife* 2015 (4): 1–20. <https://doi.org/10.7554/eLife.04837>.
- Niwa, Nao, Mariko Saitoh, Hideyo Ohuchi, Hidefumi Yoshioka, and Sumihare Noji. 1997. “Correlation between *Distal-Less* Expression Patterns and Structures of Appendages in Development of the Two-Spotted Cricket, *Gryllus bimaculatus*.” *Zoological Science* 14 (1): 115–25. <https://doi.org/10.2108/zsj.14.115>.
- Nolis, Ilias K, Daniel J McKay, Eva Mantouvalou, Stavros Lomvardas, Menie Merika, and Dimitris Thanos. 2009. “Transcription Factors Mediate Long-Range Enhancer–Promoter Interactions.” *Proceedings of the National Academy of Sciences* 106 (48): 20222 LP – 20227. <https://doi.org/10.1073/pnas.0902454106>.
- O’Kane, C.J. and Gehring, W.J.. 1987. “Detection in situ of genomic regulatory elements in *Drosophila*.” *Proceedings of the National Academy of Sciences* 84(24), pp.9123-9127. <https://doi.org/10.1073/pnas.84.24.9123>.
- Oakley, Todd H, Bjørn Østman, and Asa C V Wilson. 2006. “Repression and Loss of Gene Expression Outpaces Activation and Gain in Recently Duplicated Fly Genes.” *Proceedings of the National Academy of Sciences* 103 (31): 11637 LP – 11641. <https://doi.org/10.1073/pnas.0600750103>.
- Ochi, Haruki, Tomoko Tamai, Hiroki Nagano, Akane Kawaguchi, Norihiro Sudou, and Hajime Ogino. 2012. “Evolution of a Tissue-Specific Silencer Underlies Divergence in the Expression of Pax2 and Pax8 Paralogues.” *Nature Communications* 3 (1): 848.

- <https://doi.org/10.1038/ncomms1851>.
- Ong, Chin-Tong, and Victor G Corces. 2011. “Enhancer Function: New Insights into the Regulation of Tissue-Specific Gene Expression.” *Nature Reviews Genetics* 12 (4): 283–93. <https://doi.org/10.1038/nrg2957>.
- Ong, Chin Tong, and Victor G. Corces. 2011. “Enhancer Function: New Insights into the Regulation of Tissue-Specific Gene Expression.” *Nature Reviews Genetics* 12 (4): 283–93. <https://doi.org/10.1038/nrg2957>.
- Pandur, Petra, Ioan Ovidiu Sirbu, Susanne J Köhl, Melanie Philipp, and Michael Köhl. 2013. “Islet1-Expressing Cardiac Progenitor Cells: A Comparison across Species.” *Development Genes and Evolution* 223 (1): 117–29. <https://doi.org/10.1007/s00427-012-0400-1>.
- Pechmann, Matthias, Sara Khadjeh, Natascha Turetzek, Alistair P McGregor, Wim G M Damen, and Nikola-Michael Prpic. 2011. “Novel Function of Distal-Less as a Gap Gene during Spider Segmentation.” *PLOS Genetics* 7 (10): e1002342. <https://doi.org/10.1371/journal.pgen.1002342>.
- Peng, Pei-Chen, Pierre Khoueiry, Charles Girardot, James P Reddington, David A Garfield, Eileen E M Furlong, and Saurabh Sinha. 2019. “The Role of Chromatin Accessibility in Cis-Regulatory Evolution.” *Genome Biology and Evolution* 11 (7): 1813–28. <https://doi.org/10.1093/gbe/evz103>.
- Preger-Ben Noon, Ella, Fred P. Davis, and David L. Stern. 2016. “Evolved Repression Overcomes Enhancer Robustness.” *Developmental Cell* 39 (5): 572–84. <https://doi.org/10.1016/j.devcel.2016.10.010>.
- Preger-Ben Noon, Ella, Gonzalo Sabaris, Daniela M. Ortiz, Jonathan Sager, Anna Liebowitz, David L. Stern, and Nicolás Frankel. 2018. “Comprehensive Analysis of a Cis-Regulatory Region Reveals Pleiotropy in Enhancer Function.” *Cell Reports* 22 (11): 3021–31. <https://doi.org/10.1016/j.celrep.2018.02.073>.
- Prud’homme, Benjamin, Nicolas Gompel, and Sean B Carroll. 2007. “Emerging Principles of Regulatory Evolution.” *Proceedings of the National Academy of Sciences of the United States of America* 104 Suppl (Suppl 1): 8605–12. <https://doi.org/10.1073/pnas.0700488104>.
- Prud’homme, Benjamin, Nicolas Gompel, Antonis Rokas, Victoria A. Kassner, Thomas M. Williams, Shu Dan Yeh, John R. True, and Sean B. Carroll. 2006. “Repeated Morphological Evolution through Cis-Regulatory Changes in a Pleiotropic Gene.” *Nature* 440 (7087): 1050–53. <https://doi.org/10.1038/nature04597>.
- Ransick, Andrew, Susan Ernst, Roy J Britten, and Eric H Davidson. 1993. “Whole Mount in Situ Hybridization Shows Endo 16 to Be a Marker for the Vegetal Plate Territory in Sea Urchin Embryos.” *Mechanisms of Development* 42 (3): 117–24. [https://doi.org/https://doi.org/10.1016/0925-4773\(93\)90001-E](https://doi.org/https://doi.org/10.1016/0925-4773(93)90001-E).
- Rawls, Alan, and Sudhir Kumar. 2002. “Genomic Regulatory Systems: Development and Evolution. By Eric H Davidson.” *The Quarterly Review of Biology* 77 (4): 456. <https://doi.org/10.1086/374461>.
- Rebeiz, Mark, Nick Jikomes, Victoria A. Kassner, and Sean B. Carroll. 2011. “Evolutionary Origin of a Novel Gene Expression Pattern through Co-Option of the Latent Activities of Existing Regulatory Sequences.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (25): 10036–43. <https://doi.org/10.1073/pnas.1105937108>.
- Rebeiz, Mark, John E Pool, Victoria A Kassner, Charles F Aquadro, and Sean B Carroll. 2009. “Stepwise Modification of a Modular Enhancer Underlies Adaptation in a Drosophila Population.” *Science* 326 (5960): 1663 LP – 1667. <https://doi.org/10.1126/science.1178357>.
- Rebeiz, Mark, and Miltos Tsiantis. 2017. “Enhancer Evolution and the Origins of

- Morphological Novelty.” *Current Opinion in Genetics and Development* 45: 115–23. <https://doi.org/10.1016/j.gde.2017.04.006>.
- Reményi, Attila, Hans R Schöler, and Matthias Wilmanns. 2004. “Combinatorial Control of Gene Expression.” *Nature Structural & Molecular Biology* 11 (9): 812–15. <https://doi.org/10.1038/nsmb820>.
- Roeder, Adrienne H K, Cristina Ferrándiz, and Martin F Yanofsky. 2003. “The Role of the REPLUMLESS Homeodomain Protein in Patterning the *Arabidopsis* Fruit.” *Current Biology* 13 (18): 1630–35. <https://doi.org/10.1016/j.cub.2003.08.027>.
- Roeske, Maxwell J., Eric M. Camino, Sumant Grover, Mark Rebeiz, and Thomas Michael Williams. 2018. “Cis-Regulatory Evolution Integrated the Bric-à -Brac Transcription Factors into a Novel Fruit Fly Gene Regulatory Network.” *ELife* 7: 1–28. <https://doi.org/10.7554/eLife.32273>.
- Ronshaugen, Matthew, Nadine McGinnis, and William McGinnis. 2002. “Hox Protein Mutation and Macroevolution of the Insect Body Plan.” *Nature* 415 (6874): 914–17. <https://doi.org/10.1038/nature716>.
- Rubin, G M, and A C Spradling. 1982. “Genetic Transformation of *Drosophila* with Transposable Element Vectors.” *Science* 218 (4570): 348 LP – 353. <https://doi.org/10.1126/science.6289436>.
- Sabarís, Gonzalo, Ian Laiker, Ella Preger-Ben Noon, and Nicolás Frankel. 2019. “Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity.” *Trends in Genetics* 35 (6): 423–33. <https://doi.org/10.1016/j.tig.2019.03.006>.
- Sanetra, Matthias, Gerrit Begemann, May-Britt Becker, and Axel Meyer. 2005. “Conservation and Co-Option in Developmental Programmes: The Importance of Homology Relationships.” *Frontiers in Zoology* 2 (October): 15. <https://doi.org/10.1186/1742-9994-2-15>.
- Schaefer, Jonathan J, Guillermo Oliver, and Jonathan J Henry. 1999. “Conservation of Gene Expression during Embryonic Lens Formation and Cornea-Lens Transdifferentiation in *Xenopus Laevis*.” *Developmental Dynamics* 215 (4): 308–18. [https://doi.org/10.1002/\(SICI\)1097-0177\(199908\)215:4<308::AID-AJA3>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0177(199908)215:4<308::AID-AJA3>3.0.CO;2-I).
- Schep, Ruben, Anamaria Necsulea, Eddie Rodríguez-Carballo, Isabel Guerreiro, Guillaume Andrey, Thi Hanh Nguyen Huynh, Virginie Marcet, Jozsef Zákány, Denis Duboule, and Leonardo Beccari. 2016. “Control of Hoxd Gene Transcription in the Mammary Bud by Hijacking a Preexisting Regulatory Landscape.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (48): E7720–29. <https://doi.org/10.1073/pnas.1617141113>.
- Schneider, I, and C Amemiya. 2016. “Developmental-Genetic Toolkit for Evolutionary Developmental Biology.” In , edited by Richard M B T - Encyclopedia of Evolutionary Biology Kliman, 404–8. Oxford: Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-800049-6.00128-1>.
- Schulz, Katharine N, Eliana R Bondra, Arbel Moshe, Jacqueline E Villalta, Jason D Lieb, Tommy Kaplan, Daniel J McKay, and Melissa M Harrison. 2015. “Zelda Is Differentially Required for Chromatin Accessibility, Transcription Factor Binding, and Gene Expression in the Early *Drosophila* Embryo.” *Genome Research* 25 (11): 1715–26. <https://doi.org/10.1101/gr.192682.115>.
- Serfling, Edgar, Maria Jasin, and Walter Schaffner. 1985. “Enhancers and Eukaryotic Gene Transcription.” *Trends in Genetics* 1: 224–30. [https://doi.org/https://doi.org/10.1016/0168-9525\(85\)90088-5](https://doi.org/https://doi.org/10.1016/0168-9525(85)90088-5).
- Shapiro, Michael D, Melissa E Marks, Catherine L Peichel, Benjamin K Blackman, Kirsten S Nereng, Bjarni Jónsson, Dolph Schluter, and David M Kingsley. 2004. “Genetic and Developmental Basis of Evolutionary Pelvic Reduction in Threespine Sticklebacks.”

- Nature* 428 (6984): 717–23. <https://doi.org/10.1038/nature02415>.
- Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. “Transcriptional Enhancers: From Properties to Genome-Wide Predictions.” *Nature Reviews Genetics* 15 (4): 272–86. <https://doi.org/10.1038/nrg3682>.
- Siggers, Trevor, Michael H Duyzend, Jessica Reddy, Sidra Khan, and Martha L Bulyk. 2011. “Non-DNA-Binding Cofactors Enhance DNA-Binding Specificity of a Transcriptional Regulatory Complex.” *Molecular Systems Biology* 7 (December): 555. <https://doi.org/10.1038/msb.2011.89>.
- Slattery, Matthew, Todd Riley, Peng Liu, Namiko Abe, Pilar Gomez-Alcala, Iris Dror, Tianyin Zhou, et al. 2011. “Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins.” *Cell* 147 (6): 1270–82. <https://doi.org/10.1016/j.cell.2011.10.053>.
- Slattery, Matthew, Tianyin Zhou, Lin Yang, Ana Carolina Dantas Machado, Raluca Gordân, and Remo Rohs. 2014. “Absence of a Simple Code: How Transcription Factors Read the Genome.” *Trends in Biochemical Sciences* 39 (9): 381–99. <https://doi.org/https://doi.org/10.1016/j.tibs.2014.07.002>.
- Small, S., A. Blair, and M. Levine. 1992. “Regulation of Even-Skipped Stripe 2 in the *Drosophila* Embryo.” *The EMBO Journal* 11 (11): 4047–57. <https://doi.org/10.1002/j.1460-2075.1992.tb05498.x>.
- Small, Stephen, Adrienne Blair, and Michael Levine. 1996. “Regulation of Two Pair-Rule Stripes by a Single Enhancer in The *Drosophila* Embryo.” *Developmental Biology* 175 (2): 314–24. <https://doi.org/https://doi.org/10.1006/dbio.1996.0117>.
- Smith, Andrew D, Pavel Sumazin, Zhenyu Xuan, and Michael Q Zhang. 2006. “DNA Motifs in Human and Mouse Proximal Promoters Predict Tissue-Specific Expression.” *Proceedings of the National Academy of Sciences* 103 (16): 6275 LP – 6280. <https://doi.org/10.1073/pnas.0508169103>.
- Solomon, Mark J, Pamela L Larsen, and Alexander Varshavsky. 1988. “Mapping Protein-DNA Interactions in Vivo with Formaldehyde: Evidence That Histone H4 Is Retained on a Highly Transcribed Gene.” *Cell* 53 (6): 937–47. [https://doi.org/10.1016/S0092-8674\(88\)90469-2](https://doi.org/10.1016/S0092-8674(88)90469-2).
- Soufi, Abdenour, Meilin Fernandez Garcia, Artur Jaroszewicz, Nebiyu Osman, Matteo Pellegrini, and Kenneth S. Zaret. 2015. “Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming.” *Cell* 161 (3): 555–68. <https://doi.org/https://doi.org/10.1016/j.cell.2015.03.017>.
- Spitz, François, and Eileen E M Furlong. 2012. “Transcription Factors: From Enhancer Binding to Developmental Control.” *Nature Reviews Genetics* 13 (9): 613–26. <https://doi.org/10.1038/nrg3207>.
- Stanojevic, D, S Small, and M Levine. 1991. “Regulation of a Segmentation Stripe by Overlapping Activators and Repressors in the *Drosophila* Embryo.” *Science* 254 (5036): 1385 LP – 1387. <https://doi.org/10.1126/science.1683715>.
- Stern, David L, Tennis Court Road, and Cambridge Cb. 1998. “Between *Drosophila* Species.” *Nature* 396 (December): 463–66. <https://doi.org/10.1016/bs.ctdb.2016.03.004>.The.
- Sucena, Elio, Isabelle Delon, Isaac Jones, François Payre, and David L Stern. 2003. “Regulatory Evolution of *Shavenbaby/Ovo* Underlies Multiple Cases of Morphological Parallelism.” *Nature* 424 (6951): 935–38. <https://doi.org/10.1038/nature01768>.
- Sucena, Elio, and David L Stern. 2000. “Divergence of Larval Morphology between *Drosophila Sechellia* and Its Sibling Species Caused by Cis-Regulatory Evolution of *Ovo/Shaven-Baby*,” *Proceedings of the National Academy of Sciences* 97 (9): 4530 LP – 4534. <https://doi.org/10.1073/pnas.97.9.4530>.

- Summerbell, D, P R Ashby, O Coutelle, D Cox, S Yee, and P W Rigby. 2000. “The Expression of Myf5 in the Developing Mouse Embryo Is Controlled by Discrete and Dispersed Enhancers Specific for Particular Populations of Skeletal Muscle Precursors.” *Development* 127 (17): 3745 LP – 3757. <http://dev.biologists.org/content/127/17/3745.abstract>.
- Swami, Meera. 2010. “Shadow Enhancers Confer Robustness.” *Nature Reviews Genetics* 11 (7): 455. <https://doi.org/10.1038/nrg2818>.
- Swanson, Christina I, Nicole C Evans, and Scott Barolo. 2010. “Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer.” *Developmental Cell* 18 (3): 359–70. <https://doi.org/10.1016/j.devcel.2009.12.026>.
- Swanson, Christina Ione. 2010. “Structure, Function, and Evolution of a Signal-Regulated Enhancer.”
- Thanos, Dimitris, and Tom Maniatis. 1995. “Virus Induction of Human IFN β Gene Expression Requires the Assembly of an Enhanceosome.” *Cell* 83 (7): 1091–1100. [https://doi.org/https://doi.org/10.1016/0092-8674\(95\)90136-1](https://doi.org/https://doi.org/10.1016/0092-8674(95)90136-1).
- Thurman, Robert E, Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T Maurano, Eric Haugen, Nathan C Sheffield, et al. 2012. “The Accessible Chromatin Landscape of the Human Genome.” *Nature* 489 (7414): 75–82. <https://doi.org/10.1038/nature11232>.
- Tie, Feng, Rakhee Banerjee, Alina R Saiakhova, Benny Howard, Kelsey E Monteith, Peter C Scacheri, Michael S Cosgrove, and Peter J Harte. 2014. “Trithorax Monomethylates Histone H3K4 and Interacts Directly with CBP to Promote H3K27 Acetylation and Antagonize Polycomb Silencing.” *Development (Cambridge, England)* 141 (5): 1129–39. <https://doi.org/10.1242/dev.102392>.
- Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson, and M. H. Meisler. 1992. “Endogenous Retroviral Sequences Are Required for Tissue-Specific Expression of a Human Salivary Amylase Gene.” *Genes and Development* 6 (8): 1457–65. <https://doi.org/10.1101/gad.6.8.1457>.
- Tolhuis, Bas, Robert-Jan Palstra, Erik Splinter, Frank Grosveld, and Wouter de Laat. 2002. “Looping and Interaction between Hypersensitive Sites in the Active β -Globin Locus.” *Molecular Cell* 10 (6): 1453–65. [https://doi.org/10.1016/S1097-2765\(02\)00781-5](https://doi.org/10.1016/S1097-2765(02)00781-5).
- True, John R. 2003. “Insect Melanism: The Molecules Matter.” *Trends in Ecology & Evolution* 18 (12): 640–47. <https://doi.org/https://doi.org/10.1016/j.tree.2003.09.006>.
- Tyas, David A, T Ian Simpson, Catherine B Carr, Dirk A Kleinjan, Veronica van Heyningen, John O Mason, and David J Price. 2006. “Functional Conservation of Pax6 Regulatory Elements in Humans and Mice Demonstrated with a Novel Transgenic Reporter Mouse.” *BMC Developmental Biology* 6 (May): 21. <https://doi.org/10.1186/1471-213X-6-21>.
- Uchikawa, Masanori. 2008. “Enhancer Analysis by Chicken Embryo Electroporation with Aid of Genome Comparison.” *Development, Growth & Differentiation* 50 (6): 467–74. <https://doi.org/10.1111/j.1440-169X.2008.01028.x>.
- Venken, Koen J T, and Hugo J Bellen. 2007. “Transgenesis Upgrades for *Drosophila Melanogaster*.” *Development* 134 (20): 3571 LP – 3584. <https://doi.org/10.1242/dev.005686>.
- Verfaillie, Annelien, Dmitry Svetlichnyy, Hana Imrichova, Kristofer Davie, Mark Fiers, Zeynep Kalender Atak, Gert Hulselmans, Valerie Christiaens, and Stein Aerts. 2016. “Multiplex Enhancer-Reporter Assays Uncover Unsophisticated TP53 Enhancer Logic.” *Genome Research* 26 (7): 882–95. <https://doi.org/10.1101/gr.204149.116>.
- Villar, Diego, Camille Berthelot, Sarah Aldridge, Tim F. Rayner, Margus Lukk, Miguel Pignatelli, Thomas J. Park, et al. 2015. “Enhancer Evolution across 20 Mammalian

- Species.” *Cell* 160 (3): 554–66. <https://doi.org/10.1016/j.cell.2015.01.006>.
- Vincent, Ben J, Javier Estrada, and Angela H DePace. 2016. “The Appeasement of Doug: A Synthetic Approach to Enhancer Biology.” *Integrative Biology* 8 (4): 475–84. <https://doi.org/10.1039/c5ib00321k>.
- Visel, Axel, James Bristow, and Len A Pennacchio. 2007. “Enhancer Identification through Comparative Genomics.” *Seminars in Cell & Developmental Biology* 18 (1): 140–52. <https://doi.org/10.1016/j.semcdb.2006.12.014>.
- Vizcaya-Molina, Elena, Cecilia C. Klein, Florenci Serras, Rakesh K. Mishra, Roderic Guigó, and Montserrat Corominas. 2018. “Damage-Responsive Elements in *Drosophila* Regeneration.” *Genome Research* 28 (12): 1841–51. <https://doi.org/10.1101/gr.233098.117>.
- Wagner, Günter P., and Vincent J. Lynch. 2010. “Evolutionary Novelty.” *Current Biology* 20 (2): 48–52. <https://doi.org/10.1016/j.cub.2009.11.010>.
- Wagner, Günter P. 1996. “Homologues, Natural Kinds and the Evolution of Modularity.” *American Zoologist* 36 (1): 36–43. <http://www.jstor.org/stable/3883984>.
- Wakimoto, Barbara T, and Thomas C Kaufman. 1981. “Analysis of Larval Segmentation in Lethal Genotypes Associated with the Antennapedia Gene Complex in *Drosophila Melanogaster*.” *Developmental Biology* 81 (1): 51–64. [https://doi.org/https://doi.org/10.1016/0012-1606\(81\)90347-X](https://doi.org/https://doi.org/10.1016/0012-1606(81)90347-X).
- Wang, Jie, Jiali Zhuang, Sowmya Iyer, XinYing Lin, Troy W Whitfield, Melissa C Greven, Brian G Pierce, et al. 2012. “Sequence Features and Chromatin Structure around the Genomic Regions Bound by 119 Human Transcription Factors.” *Genome Research* 22 (9): 1798–1812. <https://doi.org/10.1101/gr.139105.112>.
- Weingarten-Gabbay, Shira, and Eran Segal. 2014. “The Grammar of Transcriptional Regulation.” *Human Genetics*. Springer Verlag. <https://doi.org/10.1007/s00439-013-1413-1>.
- Wenick, Adam S, and Oliver Hobert. 2004. “Genomic Cis-Regulatory Architecture and Trans-Acting Regulators of a Single Interneuron-Specific Gene Battery in *C. Elegans*.” *Developmental Cell* 6 (6): 757–70. <https://doi.org/https://doi.org/10.1016/j.devcel.2004.05.004>.
- Wittkopp, Patricia J., Sean B. Carroll, and Artyom Kopp. 2003a. “Evolution in Black and White: Genetic Control of Pigment Patterns in *Drosophila*.” *Trends in Genetics* 19 (9): 495–504. [https://doi.org/10.1016/S0168-9525\(03\)00194-X](https://doi.org/10.1016/S0168-9525(03)00194-X).
- Wittkopp, Patricia J., and Gizem Kalay. 2012. “Cis-Regulatory Elements: Molecular Mechanisms and Evolutionary Processes Underlying Divergence.” *Nature Reviews Genetics* 13 (1): 59–69. <https://doi.org/10.1038/nrg3095>.
- Wittkopp, Patricia J., Kathy Vaccaro, and Sean B. Carroll. 2002a. “Evolution of Yellow Gene Regulation and Pigmentation in *Drosophila*.” *Current Biology* 12 (18): 1547–56. [https://doi.org/10.1016/S0960-9822\(02\)01113-2](https://doi.org/10.1016/S0960-9822(02)01113-2).
- Wittkopp, Patricia J, Sean B Carroll, and Artyom Kopp. 2003b. “Evolution in Black and White: Genetic Control of Pigment Patterns in *Drosophila*.” *Trends in Genetics* 19 (9): 495–504. [https://doi.org/10.1016/S0168-9525\(03\)00194-X](https://doi.org/10.1016/S0168-9525(03)00194-X).
- Wittkopp, Patricia J, Emma E Stewart, Lisa L Arnold, Adam H Neidert, Belinda K Haerum, Elizabeth M Thompson, Saleh Akhras, Gabriel Smith-Winberry, and Laura Shefner. 2009. “Intraspecific Polymorphism to Interspecific Divergence: Genetics of Pigmentation in *Drosophila*.” *Science* 326 (5952): 540 LP – 544. <https://doi.org/10.1126/science.1176980>.
- Wittkopp, Patricia J, Kathy Vaccaro, and Sean B Carroll. 2002b. “Evolution of *yellow* Gene Regulation and Pigmentation in *Drosophila*.” *Current Biology* 12 (18): 1547–56. [https://doi.org/10.1016/S0960-9822\(02\)01113-2](https://doi.org/10.1016/S0960-9822(02)01113-2).

- Workman, J L, and R E Kingston. 1998. "ALTERATION OF NUCLEOSOME STRUCTURE AS A MECHANISM OF TRANSCRIPTIONAL REGULATION." *Annual Review of Biochemistry* 67 (1): 545–79. <https://doi.org/10.1146/annurev.biochem.67.1.545>.
- Wright, Theodore R F. 1987. "The Genetics Of Biogenic Amine Metabolism, Sclerotization, And Melanization In *Drosophila Melanogaster***This Review Is Dedicated to Professor Ernst Caspari in Recognition of His Pioneering Research in Biochemical Genetics." In *Molecular Genetics of Development*, edited by John G Scandalios and E W B T - Advances in Genetics Caspari, 24:127–222. Academic Press. [https://doi.org/https://doi.org/10.1016/S0065-2660\(08\)60008-5](https://doi.org/https://doi.org/10.1016/S0065-2660(08)60008-5).
- Xin, Yaqu, Yann Le Poul, Liucong Ling, Mariam Museridze, Bettina Mühling, Rita Jaenichen, Elena Osipova, and Nicolas Gompel. 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 LP – 20644. <https://doi.org/10.1073/pnas.2004003117>.
- Xu, Zhe, Hongtao Chen, Jia Ling, Danyang Yu, Paolo Struffi, and Stephen Small. 2014. "Impacts of the Ubiquitous Factor Zelda on Bicoid-Dependent DNA Binding and Transcription in *Drosophila*." *Genes & Development* 28 (6): 608–21. <https://doi.org/10.1101/gad.234534.113>.
- Yáñez-Cuna, J Omar, Huy Q Dinh, Evgeny Z Kvon, Daria Shlyueva, and Alexander Stark. 2012. "Uncovering Cis-Regulatory Sequence Requirements for Context-Specific Transcription Factor Binding." *Genome Research* 22 (10): 2018–30. <https://doi.org/10.1101/gr.132811.111>.
- Yang, Annie, Zhou Zhu, Philipp Kapranov, Frank McKeon, George M Church, Thomas R Gingeras, and Kevin Struhl. 2006. "Relationships between P63 Binding, DNA Sequence, Transcription Activity, and Biological Function in Human Cells." *Molecular Cell* 24 (4): 593–602. <https://doi.org/10.1016/j.molcel.2006.10.018>.
- Yuan, Guo-Cheng, Yuen-Jong Liu, Michael F Dion, Michael D Slack, Lani F Wu, Steven J Altschuler, and Oliver J Rando. 2005. "Genome-Scale Identification of Nucleosome Positions in S. Cerevisiae." *Science* 309 (5734): 626 LP – 630. <https://doi.org/10.1126/science.1112178>.
- Zaraiskii, A G. 2001. "HOX Genes in Embryogenesis and Phylogenesis." *Russian Journal of Developmental Biology* 32 (1): 1–10. <https://doi.org/10.1023/A:1009463208459>.
- Zaret, Kenneth S, and Jason S Carroll. 2011a. "Pioneer Transcription Factors : Establishing Competence for Gene Expression," 2227–41. <https://doi.org/10.1101/gad.176826.111>. GENES.
- . 2011b. "Pioneer Transcription Factors: Establishing Competence for Gene Expression." *Genes & Development* 25 (21): 2227–41. <https://doi.org/10.1101/gad.176826.111>.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Nicolas Gompel. Especially during my writing of the thesis, his tremendous patience and understanding encouraged me going through this difficult time. As a supervisor, he spent time to guide me and train me to be a PhD: to think critically, to be independent, to learn from mistakes. As a mentor, he is always there to listen to me and support me to be better: to discover my strengths, to learn from others' strengths. As a scientist, his passion for science has inspired me to pursue my dreams.

Thanks to my TAC members, Prof. Andreas Ladurner and Dr. Tobias Straub, for their precious advice and feedback for my research. Also, thanks to the graduate school Life Science Munich for creating a scientific and communicative environment during my PhD. Thanks to my thesis committee members for their evaluable time.

Thanks to all the lab members. Thanks for creating such a caring and friendly working environment. I will never forget all the coffee and table soccer time. I would like to especially thank my lab mate and friend Liucong Ling for her encouraging and inspiring discussions on science and life in the fly room, and Yann Le Poul for his great help with the data analysis.

Thanks to my family and my dear friends. Their unconditional love and trust are always my motivation to go further.

Finally, thanks to all the people, who have been in my life and made my living in Munich unforgettable.