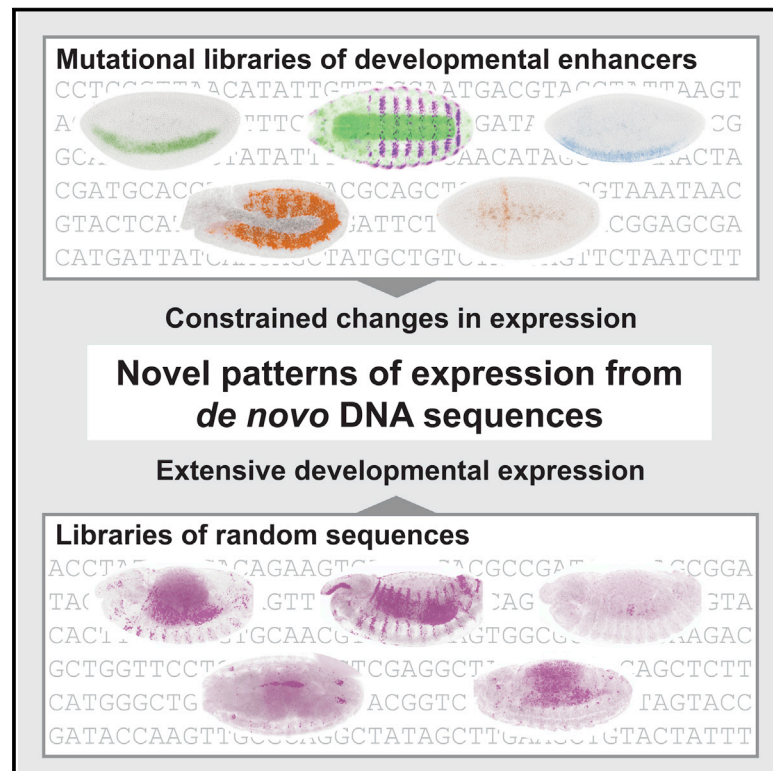


Developmental Cell

Enhancer architecture and chromatin accessibility constrain phenotypic space during *Drosophila* development

Graphical abstract



Authors

Rafael Galupa,
Gilberto Alvarez-Canales,
Noa Otilie Borst, ..., Tin Kocijan,
Christine A. Rushlow, Justin Crocker

Correspondence

rafael.galupa@embl.de (R.G.),
justin.crocker@embl.de (J.C.)

In brief

Galupa et al. compared the potential of random sequences versus point mutations of existing enhancers to generate new patterns of expression during *Drosophila melanogaster* development. Their findings point to the activity of *de novo* elements as a primary source for phenotypic novelty in developmental enhancers' evolution.

Highlights

- Mutation libraries of four developmental enhancers and three sets of random sequences
- Most point mutations rarely result in expression outside of the native pattern
- Random sequences often act as developmental enhancers
- Genomic architecture and accessibility limit developmental enhancers' evolution



Article

Enhancer architecture and chromatin accessibility constrain phenotypic space during *Drosophila* development

Rafael Galupa,^{1,*} Gilberto Alvarez-Canales,^{1,3} Noa Otilie Borst,^{1,3} Timothy Fuqua,^{1,3} Lautaro Gandara,^{1,3} Natalia Misunou,^{1,3} Kerstin Richter,^{1,3} Mariana R.P. Alves,¹ Esther Karumbi,¹ Melinda Liu Perkins,¹ Tin Kocijan,¹ Christine A. Rushlow,² and Justin Crocker^{1,4,*}

¹European Molecular Biology Laboratory, 69117 Heidelberg, Germany

²Department of Biology, New York University, New York, NY 10003, USA

³These authors contributed equally

⁴Lead contact

*Correspondence: rafael.galupa@embl.de (R.G.), justin.crocker@embl.de (J.C.)

<https://doi.org/10.1016/j.devcel.2022.12.003>

SUMMARY

Developmental enhancers bind transcription factors and dictate patterns of gene expression during development. Their molecular evolution can underlie phenotypical evolution, but the contributions of the evolutionary pathways involved remain little understood. Here, using mutation libraries in *Drosophila melanogaster* embryos, we observed that most point mutations in developmental enhancers led to changes in gene expression levels but rarely resulted in novel expression outside of the native pattern. In contrast, random sequences, often acting as developmental enhancers, drove expression across a range of cell types; random sequences including motifs for transcription factors with pioneer activity acted as enhancers even more frequently. Our findings suggest that the phenotypic landscapes of developmental enhancers are constrained by enhancer architecture and chromatin accessibility. We propose that the evolution of existing enhancers is limited in its capacity to generate novel phenotypes, whereas the activity of *de novo* elements is a primary source of phenotypic novelty.

INTRODUCTION

Morphological changes generally result from changes in the spatiotemporal regulation of gene expression during development, and thus a major theory in evolutionary developmental biology proposes anatomical evolution to be based on the genetic and molecular mechanisms underlying the evolution of spatial gene regulation.¹ In line with this, the evolution of *cis*-regulatory elements, such as developmental enhancers,² has been proposed to be a major component of phenotypical evolution across animals.^{1,3–7} The so-called “*cis*-regulatory hypothesis” proposes that mutations in enhancers are a common and continuous source of morphological variation, and a means to escape the pleiotropic effects of mutations to protein coding regions.^{1,4} For instance, the evolution of wing pigmentation “spots” in *Drosophila* involved the gain of binding sites for different transcription factors (TFs) in an enhancer controlling a pigmentation gene,⁸ whereas the loss of pelvic structures in stickleback fish occurred via mutations that abrogate the activity of an enhancer controlling the homeobox gene *Pitx1*.⁹ Molecular mechanisms of *cis*-regulatory evolution have also been proposed to include duplications of existing enhancers, *de novo* emergence from existing nonregulatory DNA and co-option or

exaptation of transposable elements or enhancers with unrelated activities.^{10–18}

Despite elegant case studies, the extent to which these mechanisms contribute to the regulatory evolution of developmental enhancers remains an open question.^{19,20} It is still unknown which changes in enhancer function are evolutionarily accessible, or how the distribution of TF binding sites might constrain the evolutionary potential of enhancers.²¹ As such, there is a lack of clarity on the molecular genetic pathways for evolutionary change in animal development based on what is functionally possible versus what is probable and permissible from the standpoint of mutational events and natural selection.¹

Here, we explored how molecular evolution of existing enhancers versus *de novo* sequences contributes to producing novel patterns of gene expression across *Drosophila melanogaster* embryos. We generated and characterized a panel of unbiased mutation libraries for both classical developmental enhancers and randomly generated sequences; this approach allows to distinguish constraints that emerge from the prior function or evolutionary histories of existing enhancers from constraints that arise from properties of the sequence or locus unrelated to selection processes.



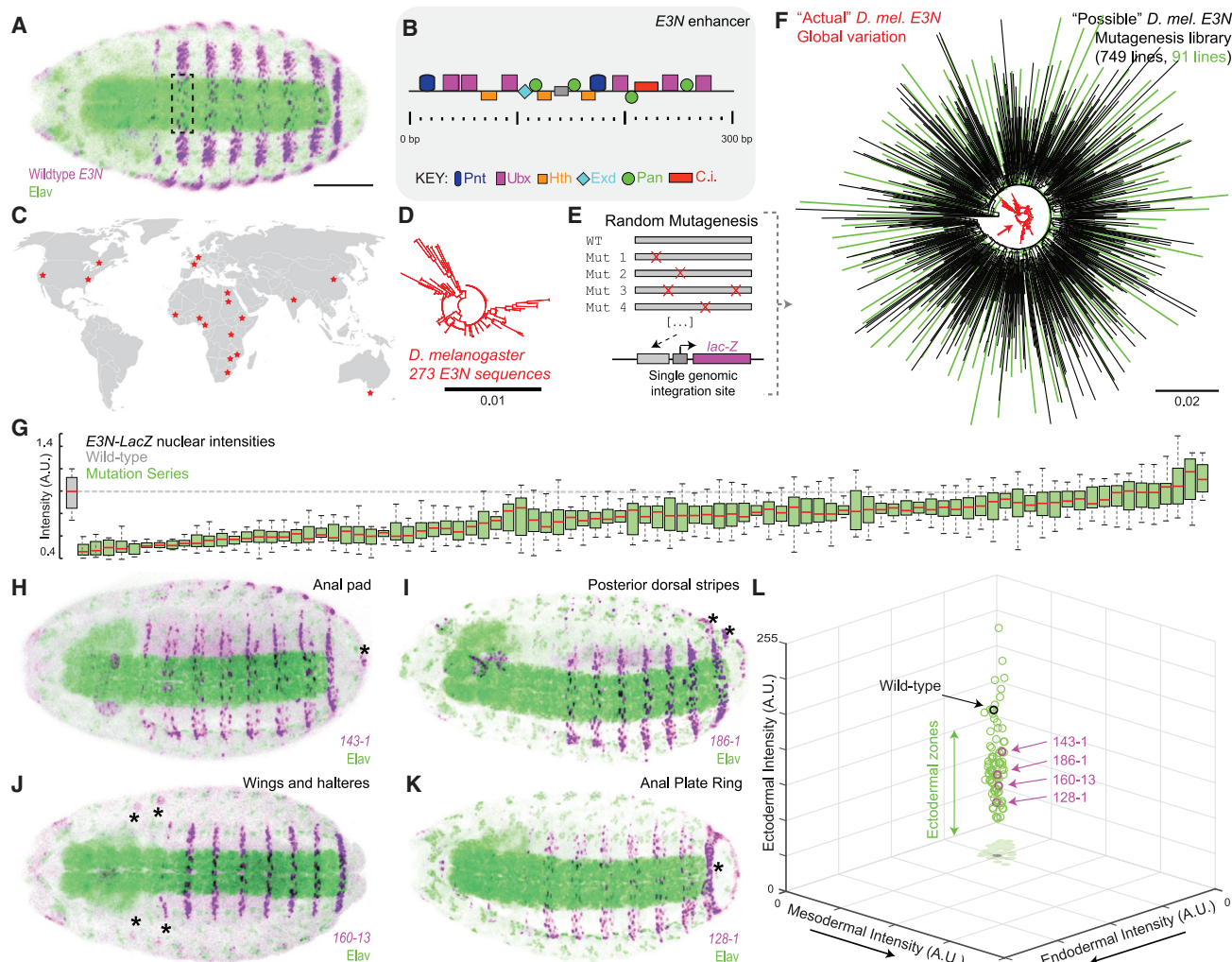


Figure 1. Mutant variants of the *E3N* enhancer have a limited capacity for expression outside native tissues and cell types

(A) Pattern of expression driven by wild-type *E3N* at stage 15 (β -galactosidase protein staining). Dashed box demarcates region quantified in (G). Scale bars, 100 μ m.

(B) Mapped binding site architecture for *E3N*.

(C) Collection locations of sequenced *Drosophila melanogaster* strains.⁹⁷

(D) Phylogenetic tree of *E3N* sequences across *D. melanogaster* strains. Units of the scale bar are units of substitutions per site of the sequence alignment.

(E) Schematic of enhancer variants and reporter gene construct used for integration into the *D. melanogaster* genome.

(F) Phylogenetic tree of *E3N* sequences across *D. melanogaster* strains (red) and of *E3N* sequences from our mutational library (black and green; in green, 91 lines selected for further characterization). Units of the scale bar are units of substitutions per site of the sequence alignment.

(G) Nuclear intensities of the A2 segment region (see region demarcated in A) across 91 lines, normalized to wild-type *E3N* ($n = 10$ embryos per line). a.u., arbitrary units of fluorescence intensity.

(H–K) Examples of mutant variants leading to reporter expression outside the wild-type *E3N* pattern. In (H) and (I), the expression associated to esophagus is likely an artifact of the construct used, as observed in other lines unrelated to *E3N*.

(L) 3D plot showing fluorescence intensities for 91 lines across three regions of the embryo with different germ-layer origins (see Figures S1G and S1H). Each dot corresponds to the average value for one variant enhancer line.

RESULTS

Constrained capacity for enhancer-driven expression outside of native expression patterns

We first set out to investigate whether and how mutations across developmental enhancers could lead to ectopic, novel expression patterns. We have previously generated a mutation library for the *E3N* enhancer (292 bp), which regulates the expression

of *shavenbaby* (*svb*; Figures 1A and 1B).²¹ This mutation library included 749 variants and most mutations led to changes in transcriptional outputs (e.g., levels and location).²¹ This library represents a ~ 6 times larger sequence space than the natural variation found for *D. melanogaster E3N* from samples across the world (Figures 1C–1F and S1A). To investigate novel expression patterns, we selected a subset of lines harboring 1–10 point mutations for further characterization (see STAR Methods for

further details); these lines come from different regions of the sequence space covered by the total library (Figure 1F; Table S1A) and showed a spectrum of effects in terms of expression levels (Figure 1G). We found that 22% of the lines showed expression outside of the usual *E3N*-driven ventral stripes, in regions such as prospective anal pads, wing and haltere imaginal discs and other structures (Figures 1H–1K). However, these regions are ectopic regions for the enhancer but not for the target gene—they correspond to ectoderm-derived regions where *svb* is expressed.^{22,23}

To evaluate ectopic expression across regions derived from different germ layers, we quantified reporter expression intensity in the selected lines (Figures S1G and S1H) and detected no expression in regions derived from germ layers other than the ectoderm (Figure 1L), whereas variable levels of expression along the “ectoderm” axis could be seen (Figure 1L). These results suggest that evolving new patterns of expression upon point mutations of a developmental enhancer is possible but developmentally biased to specific lineages.

The emergence of ectopic expression patterns upon mutagenesis of developmental enhancers is rare

The hierarchical position of an enhancer in its gene regulatory network(s) is expected to influence the effects of its loss or redeployment;¹² enhancers at *higher* positions in the network are expected to be more robust, given that mutations affecting them would have more pleiotropic effects. We thus decided to analyze additional “classical” enhancers involved in early development (*higher* in the network) to explore whether the transcriptional constraints we observed for *E3N* mutagenesis are a general property of developmental enhancers, or linked to the fact that *E3N* regulates a terminal selector gene (*lower* in the network) in later development.²⁴ The “early” enhancers we explored include *eveS2* (484 bp), important for anterior-posterior specification (Figures 2A and 2B),^{25–27} and *rhoNEE* and *twiPE* (359 and 290 bp, respectively), both involved in dorsoventral patterning (Figures 2E–2G), in the neurogenic ectoderm and mesoderm, respectively.^{28–33} For each of these enhancers, we generated mutant libraries using the same setup as for the *E3N* library:²¹ each variant was cloned upstream of a heterologous *hsp70* promoter driving *lacZ* reporter expression and integrated into the *Drosophila* genome at a specific landing site, amenable to expression across different tissues and stages (Figures S1B–S1F). Using a PCR error-rate of ~0.5% per molecule, we isolated enhancer variants containing approximately 1–5 mutations in 12–36 independent fly lines per enhancer (Table S1).

We examined reporter activity across all lines in the early embryo (stage 5) and found similar trends for all of them. On the one hand, mutations often led to significant changes in expression levels, and on the other hand, changes in expression were restricted to the native pattern—no ectopic expression was observed. For *eveS2* (Figure 2A), each variant contained a single mutation only, almost none overlapping a known binding site (Figures 2B and 2C). Yet, 75% led to significantly reduced expression compared with control (Figures 2D and 2H), suggesting that it is relatively easy to “break” the minimal *eveS2* enhancer, consistent with unsuccessful attempts to build this enhancer *de novo*.^{34,35} In no case did we observe expression outside of the *eve* stripe 2 region. Similar results were found

for *rhoNEE* and *twiPE*: 47% and 77% of enhancer variants, respectively, showed statistically significant changes in nuclear intensities compared with control (Figure 2H); for *rhoNEE*, 18% showed higher expression and 29% showed lower expression (Figure 2J); for *twiPE*, these values were 18% and 59% respectively (Figure 2I). These effects did not seem to correlate with the number of mutations per enhancer (Figure S2A) nor with the length of the enhancer (compare Figures 2B and 2F with 2H). Again, despite clear changes in levels for most mutant variants, we noted that expression outside of the typical area of expression for each enhancer was never observed—quantification of expression in control and mutant lines across regions of the embryo that will give rise to ectoderm (lateral region of the embryo), endoderm (posterior region of the embryo), and mesoderm (ventral region of the embryo; regions highlighted in Figures 2E and 2G) revealed that mutant lines showed changed levels of expression but always within the ectoderm and “mesoderm” regions only, for *rhoNEE* and *twiPE* enhancers, respectively (Figure 2K). In summary, most mutations led to changes in expression levels within native zones of expression; thus, the results suggest that the “molecular evolution” by point mutations of developmental enhancers is not likely to result in novel expression patterns.

Considering that such pleiotropic effects could be revealed throughout development,²² we analyzed expression in embryos at later stages (stage 9 and 14) for the *rhoNEE* (Figures 3A and S2B) and *twiPE* libraries (Figure 3E and S2C), but we observed no ectopic expression in the mutant lines compared with the control (Figures 3B–3D, 3F, and 3G). We also generated an additional mutational library for *tinB*, a 411-bp enhancer that controls a mesoderm-specific gene throughout a broad developmental window (Figures 3H and 3I; Table S1).^{36,37} Similar to what we found for early enhancers, 47% of enhancer variants showed significant changes in enhancer activity (Figures 3J and S2D; 20% showed increased expression, 27% showed decreased expression), yet no ectopic expression was observed (Figure 3K).

Finally, we tested whether ectopic expression could be “forced” upon recruitment of a ubiquitously expressed synthetic TF. The *rhoNEE* enhancer has been previously engineered to contain binding sites for a transcription activator-like effector (TALE) DNA-binding protein.³⁸ We crossed fly lines harboring *rhoNEE* enhancers with one, two, or three TALE-binding sites with a line containing a TALE protein fused to the strong activation domain VP64³⁹ and expressed via an ubiquitous *nos::Gal4* driver, and quantified expression across different regions of the early embryo (Figure S2E). The higher the number of binding sites for the synthetic TF, the higher the expression within the usual regions of *rhoNEE* expression. However, it was not until there were two or more binding sites (16-bp long) that appreciable expression was generated outside of the native zones of expression (Figure S2E). Together, these results reveal that the *rhoNEE* enhancer is not “intrinsically” refractory to expression outside of its usual pattern of expression but rather requires a considerably larger recruitment of activators to the locus. The fact that we do not observe ectopic expression in the enhancer libraries analyzed suggests that evolutionary constraints are imposed on developmental enhancers.

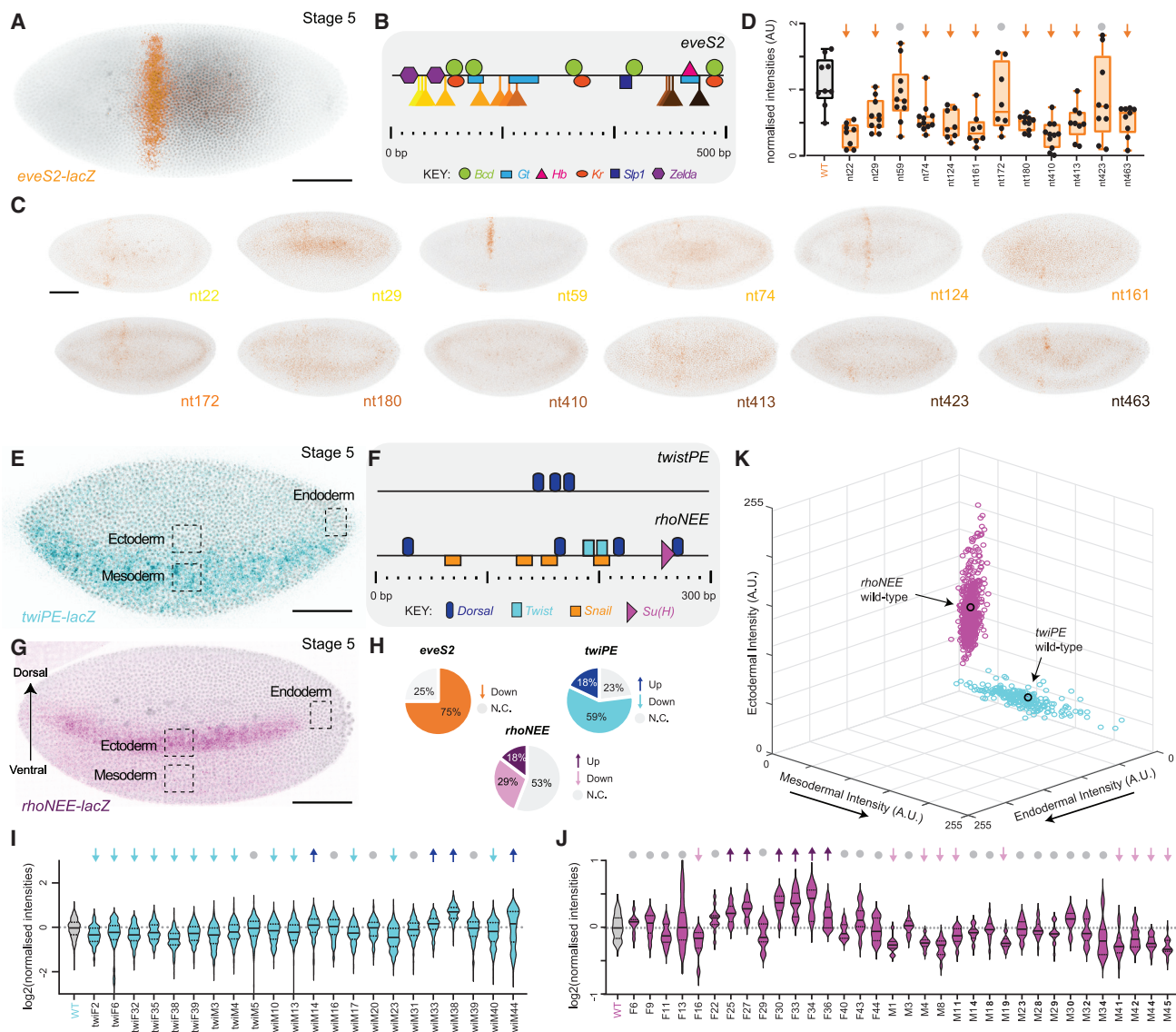


Figure 2. Mutagenesis across early developmental enhancers alters gene expression only within native patterns of expression

(A) Pattern of expression driven by wild-type *eveS2* at stage 5 (*lacZ* mRNA staining). Scale bars, 100 μ m.

(B) Known binding site architecture for *eveS2*. Location of point mutations is indicated.

(C) Examples of stained embryos from different *eveS2* single-nucleotide mutant variants. The name of each line corresponds to the location of the point mutation (compare with B).

(D) Fluorescence intensities of the region where the wild-type *eveS2* shows a stripe across 12 single-nucleotide *eveS2* variants ($n = 8-11$ embryos per line). Lines marked with an arrow are statistically significantly different from wild type ($p < 0.05$; two-tailed t test). a.u., arbitrary units of fluorescence intensity.

(E) Pattern of expression driven by wild-type *twiPE* at stage 5 (*lacZ* mRNA staining).

(F) Known binding site architecture for *twiPE* and *rhoNEE*.

(G) Pattern of expression driven by wild-type *rhoNEE* at stage 5 (*lacZ* mRNA staining).

(H) Summary of changes in expression levels for the *eveS2*, *twiPE*, and *rhoNEE* lines.

(I and J) Nuclear intensities across *twiPE* (I) and *rhoNEE* (J) variants ($n = 6-27$ embryos per line). Lines marked with an arrow (up or down) are statistically significant from wild type ($p < 0.05$; two-tailed t test).

(K) 3D plot showing fluorescence intensities for *twiPE* (blue) and *rhoNEE* (purple) lines across three regions of the embryo illustrated in (I) and (J). Each dot corresponds to one embryo; three embryos per line were quantified.

Random sequences lead to extensive expression across developmental time and space

We interrogated the extent to which *de novo* sequences, devoid of evolutionary constraints, could act as enhancers and drive expression across the embryo and across development. We

synthesized random sequences (~ 180 bp), inserted them upstream of *hsp70* promoter driving *lacZ* (similarly to the enhancer libraries) and integrated them into the fly genome at the same genomic location (Figures 4A and S3A). These sequences included a motif (UAS) for the yeast Gal4 TF,^{40,41} which

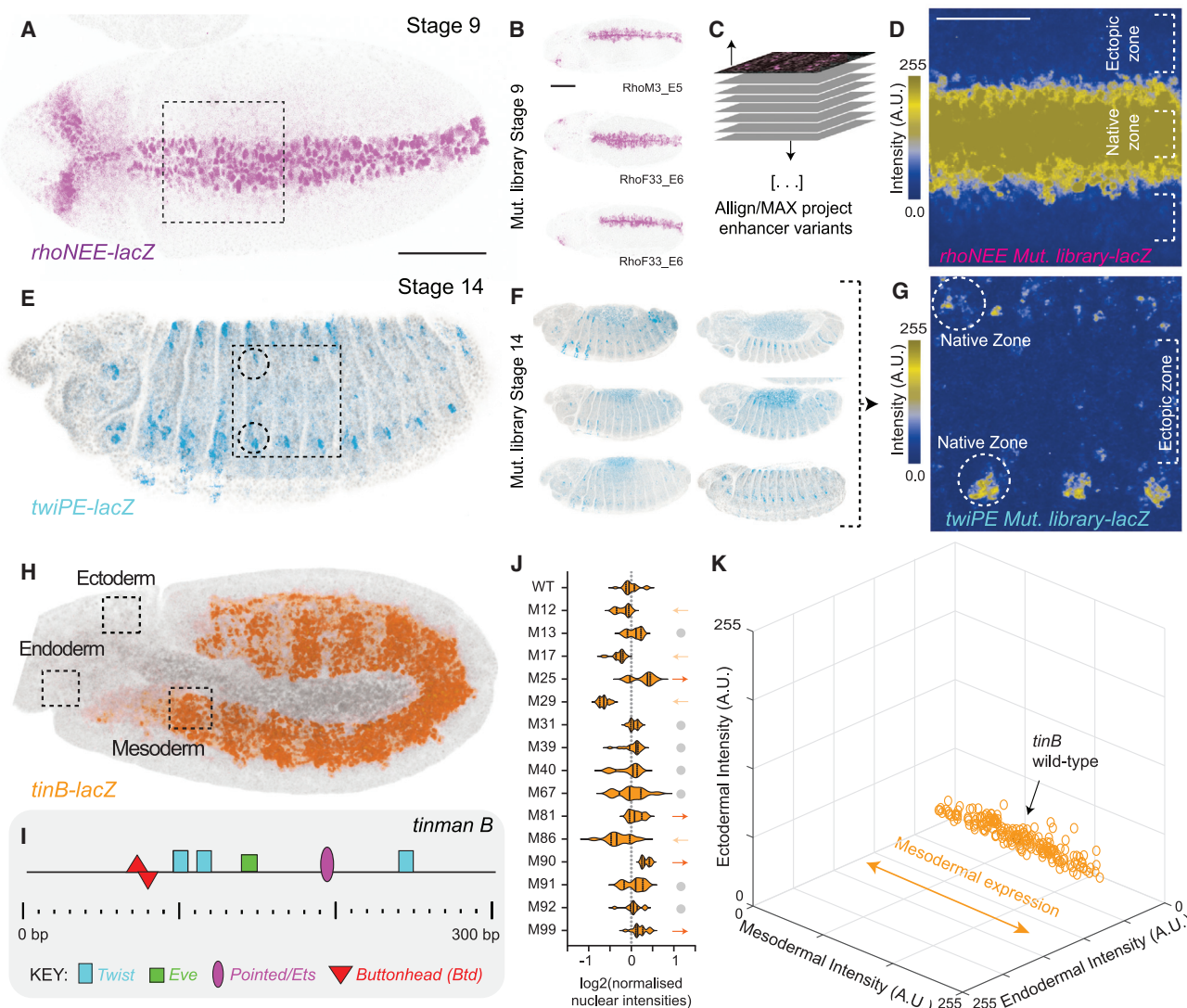


Figure 3. Mutagenesis across late developmental enhancers alters gene expression only within native patterns of expression

(A) Pattern of expression driven by wild-type *rhoNEE* at stage 9 (β -galactosidase protein staining). Scale bars, 100 μ m.
 (B) Examples of stained embryos from different *rhoNEE* mutant variants. Scale bars, 100 μ m.
 (C) Schematic of alignment and overlaying of individual Z projections of maximum intensity for *rhoNEE* mutant variants.
 (D) Heatmap of aggregated Z projections. Scale bars, 50 μ m.
 (E) Pattern of expression driven by wild-type *twiPE* at stage 14 (β -galactosidase protein staining).
 (F) Examples of stained embryos from different *twiPE* mutant variants.
 (G) Heatmap of aggregated Z projections upon alignment of individual Z projections of maximum intensity for *twiPE* mutant variants.
 (H) Pattern of expression driven by wild-type *tinB* at stage 10 (β -galactosidase protein staining).
 (I) Known binding site architecture for *tinB*.
 (J) Nuclear intensities across *tinB* variants (n = 10–18 embryos per line). Lines marked with an arrow (up or down) are statistically significant from wild type (p < 0.05; two-tailed t test).
 (K) 3D plot showing fluorescence intensities for *tinB* lines across three regions of the embryo as illustrated in (H). Each dot corresponds to one embryo; at least ten embryos per line were quantified.

is not present in the fly and thus, this motif should be “neutral”; this design was chosen so that these sequences have a comparable architecture to libraries containing other motifs (see later). We isolated 56 fly lines harboring unique sequences (Table S1), for which we stained embryos at different stages to determine reporter gene’s expression pattern(s). Surprisingly, 86% of sequences led to changes in reporter expression at least

in some cells and/or at some developmental stage, compared with expression of the reporter with no sequence cloned upstream (Figures 4B–4D and S3B). The other surprising observation was that despite such pervasive expression, we never observed expression in the early embryo (Figure 4C). Given the variable consensus sites found in multicellular systems, such libraries are expected to have a range of motifs with variable

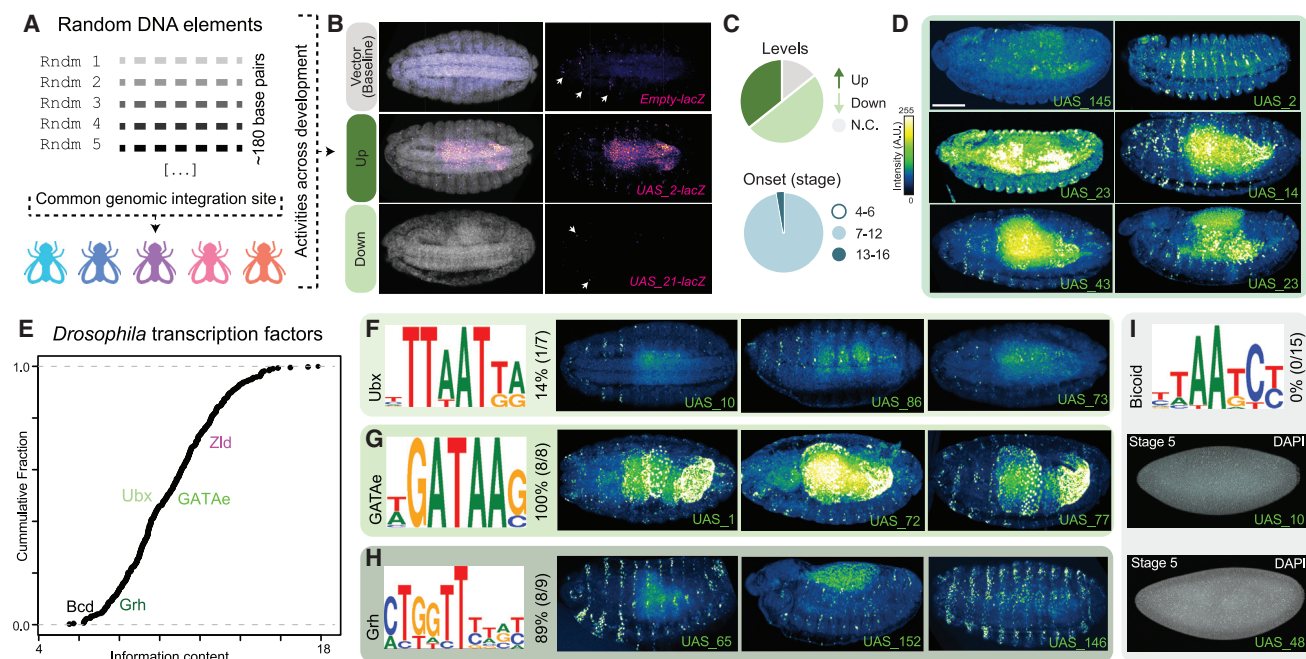


Figure 4. Random DNA sequences often drive reporter expression during development

(A) Schematic of the UAS library.
(B) Expression patterns at stage 15 were compared with the reporter with no sequence cloned upstream (top) and classified as “up” (middle) or “down” (bottom), depending on whether expression was increased or decreased, respectively.
(C) Summary of changes in expression levels at stage 15 (top) based on panel (B), and of developmental period in which expression is first observed (bottom).
(D) Examples of stained embryos from different random DNA sequences.
(E) Cumulative distribution function of the expected frequency of *Drosophila* TF motifs in random DNA.
(F) Ubx motif, percentage of lines showing expression among random DNA lines with a Ubx motif and examples of corresponding embryos.
(G) GATA motif, percentage of lines showing expression among random DNA lines with a GATA motif and examples of corresponding embryos.
(H) Grh motif, percentage of lines showing expression among random DNA lines with a Grh motif and examples of corresponding embryos.
(I) Bicoid motif, percentage of lines showing expression among random DNA lines with a Bcd motif and examples of corresponding embryos.

information content^{42,43} (Figure 4E). To explore the expression patterns observed, we conducted motif searches across all random sequences for *Drosophila* developmental TFs (Figure 4E; STAR Methods). Motifs found included Ultrabithorax (Ubx), GATA, Grainyhead (Grh), and Bicoid (Bcd) motifs (Figures 4F–4I). Interestingly, 100% or 80% of the random DNA elements containing a GATA or Grh motif, respectively, showed expression (Figures 4G and 4H), consistent with their previously reported predictive power^{44,45} and with the expression patterns of the respective TFs. In contrast, only 14% of elements with a Ubx motif showed expression (Figure 4F), and none of the elements containing a Bcd motif showed expression (Figure 4I), consistent with the absence of expression in the early embryo for all random sequences. We calculated whether our random sequences were biased for motifs of late-development TFs, but this did not explain the absence of early expression (average per sequence: ~3.9 hits per early-specific motif versus ~3.4 hits per late-specific motif; see STAR Methods).

Specific motifs can potentiate emergence of enhancer activity

Completely random sequences thus seem to have a high potential of driving expression, and this can be associated to particular motifs. Given the association between chromatin accessibility

and transcriptional permissiveness,⁴⁶ as well as studies suggesting that chromatin accessibility might underlie enhancer evolution,^{47,48} we generated “biased” random libraries in which we included a Grh motif (Figure 5A; 7 lines, Table S1) or a Zelda motif (Figure 5E; 41 lines; Table S1) approximately at the center of random sequences. Grh and Zelda are TFs in the fly reported to have “pioneer activity”^{49,50}—their binding is associated with “opening” chromatin, rendering enhancers more accessible to binding by other TFs.^{51–58} Though Zelda is usually associated with early fly development, it is expressed throughout development (Figure S4A), and its late embryonic knockout has phenotypic consequences (Figure S4B). Consistent with the idea of pioneer activity, an even higher proportion of random sequences from the Grh and Zelda-biased libraries drove expression compared with the UAS library (Figures 5B, 5C, 5F, 5G, and S4C). Not only a higher number of lines was associated with expression for the biased libraries, but also expression levels were higher when compared with the UAS library, regardless of the region of the embryo (Figures 5D, 5H, and 5I). To further test the potential of these motifs, we added one or two Zelda motifs to the developmental enhancers we tested initially (*eveS2*, *rhoNEE*, *twiPE*, and *tinB*) and found a significant increase in reporter expression levels for all enhancers within their native patterns of expression (Figures S5A–S5H). For the

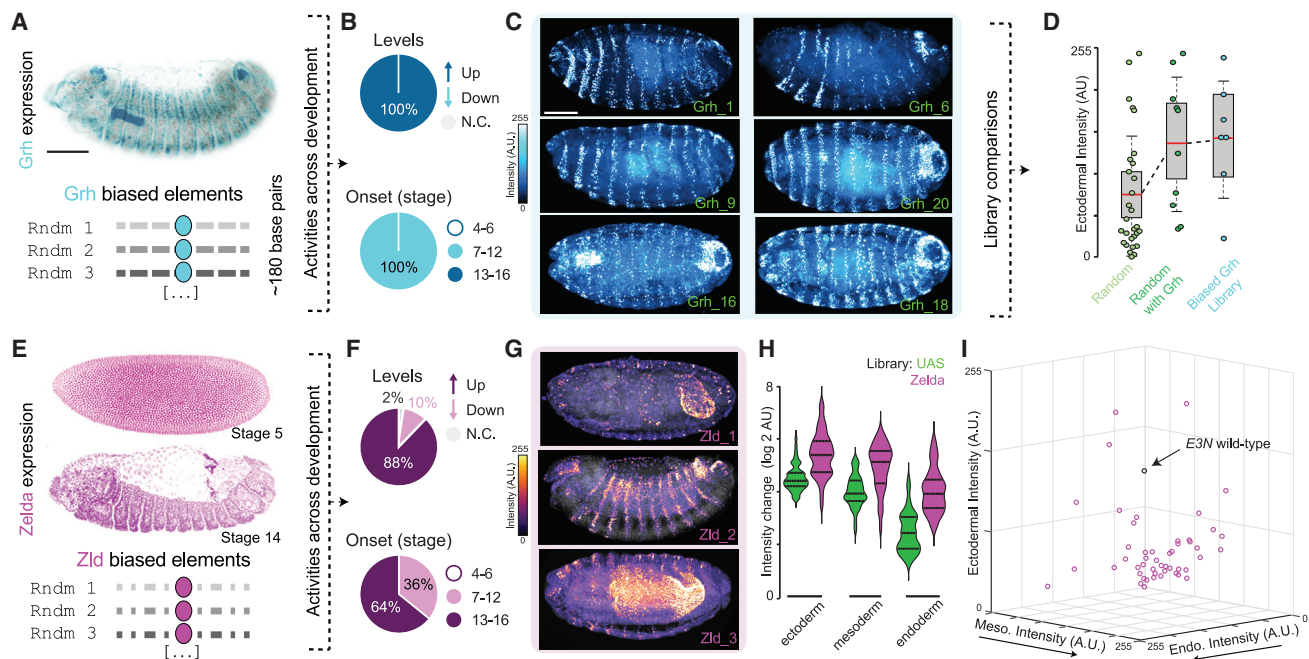


Figure 5. Specific DNA motifs enhance likelihood of reporter expression during development

(A) Staining for Grh transcription factor (top) and schematic of the Grh-library (bottom).
 (B) Summary of changes in expression levels (top) compared with the reporter with no sequence cloned upstream (Figure 4B) and of developmental period in which expression is first observed (bottom).
 (C) Examples of stained embryos from different Grh-biased sequences.
 (D) Quantification of fluorescent intensities in ectoderm-associated region for all random DNA sequences, for random DNA sequences with Grh motifs (subset of all random DNA sequences) and for Grh-biased sequences.
 (E) Staining for Zld transcription factor (top) and schematic of the Zld-library (bottom).
 (F) Summary of changes in expression levels at stage 15 (top) compared with the reporter with no sequence cloned upstream (Figure 4B) and of developmental period in which expression is first observed (bottom).
 (G) Examples of stained embryos from different Zld-biased sequences.
 (H) Quantification of fluorescent intensities for Zld-biased lines across three regions of the embryo (see Figures S1G and S1H).
 (I) 3D plot showing fluorescence intensities for Zld-biased lines, based on (H). Each dot corresponds to one line. For reference, fluorescence intensity for the wild-type *E3N* sequence is shown (from Figure 1L).

eveS2 lines, we additionally observed novel, ectopic expression (Figures S5A–S5H), suggesting that the *Zelda* motifs might “unlock” cryptic sites contained in *eveS2*. We tested whether *eveS2* contained more predicted motifs than the other enhancers, but we did not find any significant differences in the number of hits (0.07 for *eveS2* versus 0.10, 0.12, and 0.05 for *rhoNEE*, *tinB*, and *twiPE*, respectively; normalized per enhancer length).

To explore the possibility that the occurrence of specific motifs throughout the genome might contribute to the emergence of (*de novo*) enhancers, we selected genomic sequences containing high-affinity Ubx/Hth motifs (ATGATTATGAC)⁵⁹ present in *D. melanogaster* but not in other *Drosophila* species (Figures S5I–S5L). Such motifs have been demonstrated to augment chromatin accessibility⁶⁰ and are broadly used across development, providing a counterpoint to our synthetic libraries. Strikingly, when we tested their enhancer potential with the *lacZ* reporter assay, all sequences showed enhancer activity (Figures S5I–S5L). Mutating the Ubx/Hth motif in each of those sequences led to a dramatic reduction in expression for six out of seven of those sequences (Figures S5I–S5L), indicating that these motifs clearly have the capacity to drive expression across development. These results support the idea that specific

sequence motifs might prime genomic sequences to act and/or evolve as enhancers.

DISCUSSION

We used transgenesis-based mutagenesis and *de novo* gene synthesis during fly embryogenesis to investigate evolutionary pathways for enhancer activity. We used fly development to explore how novel patterns of gene expression might appear from either molecular evolution of developmental enhancers or random sequences. Notably, while reporter gene assays and minimal enhancers may not reflect the full regulatory activities of native loci,^{61–63} such an approach allows us to evaluate a broad range of “possible” enhancer variation in a controlled experimental setup, without associated fitness costs and allowing a broader exploration of evolution and development without the complexities and historical contingencies found in nature. Furthermore, using such an assay in a developmental model system, which generates an embryo in 24 h, we can assay regulatory activities across ~100,000 cells of different lineage origins.⁶⁴

Using this approach, we found that most mutations in enhancers led to changes in levels of reporter gene expression,

but almost entirely within their native zones of expression (Figures 1, 2, and 3), similar to previous studies using transgenic mutagenesis of the *Shh* enhancer in murine embryos,⁶⁵ or the *E3N* enhancer²¹ and the wing *spot*¹⁹⁶ enhancer⁶⁶ in fly embryos. Consistent with our results, known phenotypic evolution through nucleotide mutations of standing regulatory elements seems to appear either through changes in the levels or timings of expression within native zones or the loss of regulatory activities. For example, the evolution of pigmentation spots in fly wings occurred via a specific spatial increase in the melanin protein Yellow, which is uniformly expressed at low levels throughout the developing wings of fruit flies;⁸ see Frankel et al.⁶⁷ and Rebeiz et al.⁶⁸ for other examples of evolution within native patterns of expression. Evolution of other traits such as thoracic ribs in vertebrates,⁶⁹ limbs in snakes,⁷⁰ pelvic structures in sticklebacks,⁹ and seed shattering in rice⁷¹ are all associated with loss of enhancer activity due to internal enhancer mutations. Additionally, mutations have been found to occur less often in functionally constrained regions of the genome, suggesting that mutation bias may reduce the occurrence of deleterious mutations in regulatory regions.⁷²

Consistent with these results, phenotypic novelties underlain by enhancer-associated ectopic gains of expression are reportedly due to transposon mobilization,^{73–76} rearrangements in chromosome topology^{77–79} or *de novo* evolution of enhancers from DNA sequences with unrelated or nonregulatory activities.^{13,17,19,80–83} Previous studies have explored the potential of random DNA sequences to lead to reporter gene expression, either as enhancers or promoters, especially in cell lines of prokaryotic or eukaryotic origin.^{45,84,85} These have shown that there is a short (or sometimes null) mutational distance between random sequences and active *cis*-regulatory elements,⁸⁵ which may improve evolvability. In our study, we tested random sequences in a developmental context and found that most showed enhancer activity across several types of tissues and developmental stages (Figure 4). These results are consistent with a study that tested enhancer activity of all 6-mers in developing zebrafish embryos and found a diverse range of expression for ~38% of the sequences at two developmental stages.²⁰ We observed expression driven by random sequences even in the absence of motifs within their sequence for TFs with pioneering activity (Figure 4). Yet, when such motifs were included, nearly all sequences acted as “strong” enhancers (leading to high levels of expression) (Figure 5), consistent with the “evolutionary barrier” to the formation of a novel enhancer being lower in regions that already contain motifs for DNA-binding factors, which can “act cooperatively with newly emerging sites.”¹¹

It is interesting to note that, despite the high potential of random sequences to be expressed during development and across cell types, we never observed expression prior to gastrulation; this was not evaluated in the zebrafish study or in other studies. This may be due to the rapid rates of early fruit fly development, in which gene expression patterns are highly dynamic, and cell-fate specifications occur within minutes.⁸⁶ As such, there may be extensive regulatory demands placed on transcriptional enhancers, reflected in the clusters of high-affinity binding sites common across early embryonic developmental enhancers⁸⁷ as well as their extensive conservation in function⁸⁸ and location.⁸⁹ In the future, it will be interesting to explore how regulatory de-

mands that change across development—such as nuclear differentiation, network cross-talk, and metabolic changes—are reflected in regulatory architectures and their evolvability.

The observation that most random sequences led to expression suggests that the potential of any sequence within the genome to drive expression is enormous and thus “an important playground for creating new regulatory variability and evolutionary innovation.”⁸⁰ This was further supported by the regulatory potential of the genomic sequences we tested, containing Ubx/Hth motifs; indeed, the results from our work imply that enhancers would more likely evolve from sequences that contain or are biased toward specific motifs (e.g., GATA and Zelda). Perhaps the challenge from an evolutionary perspective has not been what allows expression, but what prevents expression; thus, mechanisms that repress “spurious” expression might have evolved across genomes. This is in line with propositions that nucleosomal DNA in eukaryotes has evolved to repress transcription,^{90,91} along with transcriptional repressors and other mechanisms such as DNA methylation, as a response (at least partially) to “the unbearable ease of expression” present in prokaryotes.⁹² The action of such repressive mechanisms could also explain why mutagenesis of developmental enhancers, which are subject to evolutionary selection, does not easily lead to expression outside their native patterns of expression. In sum, our findings raise exciting questions about the evolution of enhancers and the emergence of novel patterns of expression that may underlie new phenotypes, suggesting an underappreciated role for *de novo* evolution of enhancers by happenstance. Genetic theories of morphological evolution will benefit from comparing controlled, multi-dimensional laboratory experiments with standing variation;⁹³ such an integrative approach could provide the frameworks that will enable us to make both transcriptional and evolutionary predictions.

Limitations of the study

One limitation of our study lies on the numbers—we have tested a significant number of enhancer variants, but it is still possible that we would have captured ectopic expression more frequently had we tested a larger set of enhancer variants. Also, in principle, a higher number of mutations per enhancer could have enhanced the likelihood of ectopic expression. Previous work from our lab with the *E3N* enhancer reported that indeed the proportion of lines with ectopic expression increased with the number of mutations.²¹ However, this increase plateaued around 20%–30% for lines with ~3+ mutations per enhancer and in this study, the number of mutations in the enhancer variants for *twiPE*, *rhoNEE*, and *tinB* ranges from 1 to 7 mutations, so we could have expected to capture a number of lines with ectopic expression. Importantly, our assay captures millions of years of variation in a controlled setting decoupled from fitness costs. It is also possible that ectopic expression might be present in developmental stages that we have not analyzed. Finally, would the results be different if we had used a different promoter? We have not tested this formally, but based on published literature, we believe that using a different promoter would not have major implications in the results observed. Testing a total of 27 enhancer-promoter combinations in human cells, efficiency of enhancers has been shown to be approximately the same irrespective of the type of

promoter used,⁹⁴ and a recent combinatorial analysis of 1,000 human promoters and 1,000 human enhancers confirmed that most enhancers activate all promoters by similar amounts.⁹⁵ These studies, in cell lines, could only address levels of expression, not spatial patterns—but very recently published results from the lab⁹⁶ show that developmental promoters in fly embryos can drive a range of outputs but do not affect spatial aspects of expression, only levels.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
 - Fly strains and constructs
 - Embryos collection and fixation
 - Reporter gene expression analysis
 - Motif prediction analysis of random sequences
 - Information content
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.devcel.2022.12.003>.

ACKNOWLEDGMENTS

We thank GenScript for gifting us the random DNA libraries, Hsiao-Yun Liu for making the protein for the Zelda antibody prep, Garth Ilsley for discussions about pattern quantifications, Claire Standley for title suggestions and Denis Krndija for critical feedback on the manuscript. We are grateful to other members of the Crocker lab for helpful suggestions and discussions during the course of the project, in particular Xueying Li. We also thank the ALMF imaging platform and Alessandra Reversi for some of the fly injections. Stocks obtained from the Bloomington Drosophila Stock Center (NIH P40D018537) were used in this study. R.G. and L.G. were supported by fellowships from the European Molecular Biology Laboratory Interdisciplinary Postdoc Programme (EIPOD) under Marie Skłodowska-Curie Actions COFUND (664726 and 847543, respectively). Research in the Crocker lab is supported by the European Molecular Biology Laboratory (EMBL).

AUTHOR CONTRIBUTIONS

Conceptualization: R.G., T.F., and J.C. Investigation: R.G., G.A.-C., M.R.P.A., N.O.B., T.F., L.G., N.M., K.R., E.K., T.K., C.A.R., and J.C. Methodology: R.G., G.A.-C., N.O.B., T.F., L.G., K.R., and J.C. Formal analysis: R.G., G.A.-C., N.O.B., T.F., L.G., N.M., and J.C. Data curation: R.G., N.M., and J.C. Visualization: R.G., T.F., K.R., N.M., and J.C. Software: G.A.-C., M.L.P., and J.C. Resources: C.A.R. Supervision: R.G., T.F., and J.C. Project administration: R.G. and J.C. Funding acquisition: J.C. Writing – original draft: R.G. and J.C. Writing – review & editing: R.G., G.A.-C., N.O.B., T.F., L.G., N.M., M.R.P.A., M.L.P., and J.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: August 4, 2022

Revised: October 18, 2022

Accepted: December 7, 2022

Published: January 9, 2023

REFERENCES

1. Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. <https://doi.org/10.1016/J.CELL.2008.06.030>.
2. Jindal, G.A., and Farley, E.K. (2021). Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev. Cell* 56, 575–587. <https://doi.org/10.1016/J.DEVCEL.2021.02.016>.
3. Monteiro, A., and Gupta, M.D. (2016). Identifying coopted networks and causative mutations in the origin of novel complex traits. *Curr. Top. Dev. Biol.* 119, 205–226. <https://doi.org/10.1016/BS.CTDB.2016.03.003>.
4. Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* 62, 2155–2177. <https://doi.org/10.1111/J.1558-5646.2008.00450.X>.
5. Majic, P., and Payne, J.L. (2020). Enhancers facilitate the birth of de novo genes and gene integration into regulatory networks. *Mol. Biol. Evol.* 37, 1165–1178. <https://doi.org/10.1093/MOLBEV/MSZ300>.
6. Nghe, P., de Vos, M.G.J., Kingma, E., Kogenaru, M., Poelwijk, F.J., Laan, L., and Tans, S.J. (2020). Predicting evolution using regulatory architecture. *Annu. Rev. Biophys.* 49, 181–197. <https://doi.org/10.1146/annurev-biophys-070317-032939>.
7. Koshikawa, S. (2015). Enhancer modularity and the evolution of new traits. *Fly (Austin)* 9, 155–159. <https://doi.org/10.1080/19336934.2016.1151129>.
8. Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487. <https://doi.org/10.1038/nature03235>.
9. Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305. <https://doi.org/10.1126/science.1182213>.
10. Kvon, E.Z., Waymack, R., Gad, M., and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. *Nat. Rev. Genet.* 22, 324–336. <https://doi.org/10.1038/s41576-020-00311-x>.
11. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167, 1170–1187. <https://doi.org/10.1016/J.CELL.2016.09.018>.
12. Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* 10, 141–148. <https://doi.org/10.1038/nrg2499>.
13. Emera, D., Yin, J., Reilly, S.K., Gockley, J., and Noonan, J.P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. USA* 113, E2617–E2626. <https://doi.org/10.1073/PNAS.1603718113>.
14. Indjeian, V.B., Kingman, G.A., Jones, F.C., Guenther, C.A., Grimwood, J., Schmutz, J., Myers, R.M., and Kingsley, D.M. (2016). Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell* 164, 45–56. <https://doi.org/10.1016/J.CELL.2015.12.007>.
15. Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* 43, 1154–1159. <https://doi.org/10.1038/ng.917>.

16. Fong, S.L., and Capra, J.A. (2022). Function and constraint in enhancers with multiple evolutionary origins. <https://doi.org/10.1101/2022.01.05.475150>.
17. Rebeiz, M., Jikomes, N., Kassner, V.A., and Carroll, S.B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc. Natl. Acad. Sci. USA* 108, 10036–10043. <https://doi.org/10.1073/pnas.1105937108>.
18. Koshikawa, S., Giorgianni, M.W., Vaccaro, K., Kassner, V.A., Yoder, J.H., Werner, T., and Carroll, S.B. (2015). Gain of cis-regulatory activities underlies novel domains of wingless gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 112, 7524–7529. <https://doi.org/10.1073/PNAS.1509022112>.
19. Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* 46, 685–692. <https://doi.org/10.1038/ng.3009>.
20. Smith, R.P., Riesenfeld, S.J., Holloway, A.K., Li, Q., Murphy, K.K., Feliciano, N.M., Orecchia, L., Oksenberg, N., Pollard, K.S., and Ahituv, N. (2013). A compact, in vivo screen of all 6-mers reveals drivers of tissue-specific expression and guides synthetic regulatory element design. *Genome Biol.* 14, R72. <https://doi.org/10.1186/GB-2013-14-7-R72/COMMENTS>.
21. Fuqua, T., Jordan, J., van Breugel, M.E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai, A., Mann, R.S., Stern, D.L., and Crocker, J. (2020). Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* 587, 235–239. <https://doi.org/10.1038/s41586-020-2816-5>.
22. Preger-Ben Noon, E., Sabarís, G., Ortiz, D.M., Sager, J., Liebowitz, A., Stern, D.L., and Frankel, N. (2018). Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* 22, 3021–3031. <https://doi.org/10.1016/j.celrep.2018.02.073>.
23. Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 7305, 490–493. <https://doi.org/10.1038/nature09158>.
24. Allan, D.W., and Thor, S. (2015). Transcriptional selectors, masters, and combinatorial codes: regulatory principles of neural subtype specification. *Wiley Interdiscip. Rev. Dev. Biol.* 4, 505–528. <https://doi.org/10.1002/WDEV.191>.
25. Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Science* 254, 1385–1387. <https://doi.org/10.1126/SCIENCE.1683715>.
26. Small, S., Kraut, R., Hoey, T., Warrior, R., and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* 5, 827–839. <https://doi.org/10.1101/GAD.5.5.827>.
27. Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* 11, 4047–4057. <https://doi.org/10.1002/J.1460-2075.1992.TB05498.X>.
28. Bier, E., Jan, L.Y., and Jan, Y.N. (1990). rhomboid, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*. *Genes Dev.* 4, 190–203. <https://doi.org/10.1101/GAD.4.2.190>.
29. Ip, Y.T., Park, R.E., Kosman, D., Bier, E., and Levine, M. (1992). The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev.* 6, 1728–1739. <https://doi.org/10.1101/GAD.6.9.1728>.
30. Markstein, M., Zinzen, R., Markstein, P., Yee, K.-P., Erives, A., Stathopoulos, A., and Levine, M. (2004). A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development* 131, 2387–2394. <https://doi.org/10.1242/DEV.01124>.
31. Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev.* 5, 1881–1891. <https://doi.org/10.1101/GAD.5.10.1881>.
32. Pan, D.J., Huang, J.D., and Courey, A.J. (1991). Functional analysis of the *Drosophila* twist promoter reveals a dorsal-binding ventral activator region. *Genes Dev.* 5, 1892–1901. <https://doi.org/10.1101/GAD.5.10.1892>.
33. Thisse, C., Perrin-Schmitt, F., Stetzel, C., and Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila* twist gene by the dorsal gene product. *Cell* 65, 1191–1201. [https://doi.org/10.1016/0092-8674\(91\)90014-P](https://doi.org/10.1016/0092-8674(91)90014-P).
34. Crocker, J., and Ilsley, G.R. (2017). Using synthetic biology to study gene regulatory evolution. *Curr. Opin. Genet. Dev.* 47, 91–101. <https://doi.org/10.1016/J.GDE.2017.09.001>.
35. Vincent, B.J., Estrada, J., and DePace, A.H. (2016). The appeasement of Doug: a synthetic approach to enhancer biology. *Integr. Biol. (Camb)* 8, 475–484. <https://doi.org/10.1039/c5ib00321k>.
36. Zaffran, S., Reim, I., Qian, L., Lo, P.C., Bodmer, R., and Frasch, M. (2006). Cardioblast-intrinsic Tinman activity controls proper diversification and differentiation of myocardial cells in *Drosophila*. *Development* 133, 4073–4083. <https://doi.org/10.1242/DEV.02586>.
37. Yin, Z., Xu, X.L., and Frasch, M. (1997). Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* 124, 4971–4982. <https://doi.org/10.1242/DEV.124.24.4971>.
38. Crocker, J., Ilsley, G.R., and Stern, D.L. (2016). Quantitatively predictable control of *Drosophila* transcriptional enhancers in vivo with engineered transcription factors. *Nat. Genet.* 48, 292–298. <https://doi.org/10.1038/ng.3509>.
39. Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. USA* 95, 14628–14633. <https://doi.org/10.1073/pnas.95.25.14628>.
40. Webster, N., Jin, J.R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast UASG is a transcriptional enhancer in human hela cells in the presence of the GAL4 trans-activator. *Cell* 52, 169–178. [https://doi.org/10.1016/0092-8674\(88\)90505-3](https://doi.org/10.1016/0092-8674(88)90505-3).
41. Kakidani, H., and Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. *Cell* 52, 161–167. [https://doi.org/10.1016/0092-8674\(88\)90504-1](https://doi.org/10.1016/0092-8674(88)90504-1).
42. de Boer, C.G., Vaishnav, E.D., Sadeh, R., Abeyta, E.L., Friedman, N., and Regev, A. (2019). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* 38, 56–65. <https://doi.org/10.1038/s41587-019-0315-8>.
43. Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* 25, 434–440. <https://doi.org/10.1016/J.TIG.2009.08.003>.
44. Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* 512, 91–95. <https://doi.org/10.1038/nature13395>.
45. de Almeida, B.P., Reiter, F., Pagani, M., and Stark, A. (2022). DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* 54, 613–624. <https://doi.org/10.1038/S41588-022-01048-5>.
46. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 4, 207–220. <https://doi.org/10.1038/s41576-018-0089-8>.
47. Peng, P.C., Khoeiry, P., Girardot, C., Reddington, J.P., Garfield, D.A., Furlong, E.E.M., and Sinha, S. (2019). The role of chromatin accessibility in cis-regulatory evolution. *Genome Biol. Evol.* 11, 1813–1828. <https://doi.org/10.1093/GBE/EVZ103>.
48. Xin, Y., le Poul, Y., Ling, L., Museridze, M., Mühling, B., Jaenichen, R., Osipova, E., and Gompel, N. (2020). Enhancer evolutionary co-option

- through shared chromatin accessibility input. *Proc. Natl. Acad. Sci. USA* 117, 20636–20644. <https://doi.org/10.1073/PNAS.2004003117>.
49. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25, 2227–2241. <https://doi.org/10.1101/GAD.176826.111>.
 50. Hansen, J.L., Loell, K.J., and Cohen, B.A. (2022). The pioneer factor hypothesis is not necessary to explain ectopic liver gene activation. *eLife* 11, e73358. <https://doi.org/10.7554/eLife.73358>.
 51. Sun, Y., Nien, C.Y., Chen, K., Liu, H.Y., Johnston, J., Zeitlinger, J., and Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* 25, 1703–1714. <https://doi.org/10.1101/GR.192542.115>.
 52. Schulz, K.N., Bondra, E.R., Moshe, A., Villalta, J.E., Lieb, J.D., Kaplan, T., McKay, D.J., and Harrison, M.M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Res.* 25, 1715–1726. <https://doi.org/10.1101/GR.192682.115>.
 53. Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I.I., et al. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* 50, 1011–1020. <https://doi.org/10.1038/s41588-018-0140-x>.
 54. Foo, S.M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.Y., Kirov, N., Shvartsman, S.Y., and Rushlow, C.A. (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr. Biol.* 24, 1341–1346. <https://doi.org/10.1016/j.cub.2014.04.032>.
 55. Nevil, M., Gibson, T.J., Bartolutti, C., Iyengar, A., and Harrison, M.M. (2020). Establishment of chromatin accessibility by the conserved transcription factor Grainy head is developmentally regulated. *Development* 147, dev185009. <https://doi.org/10.1242/dev.185009>.
 56. Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda Binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* 7, e1002266. <https://doi.org/10.1371/JOURNAL.PGEN.1002266>.
 57. Iwafuchi-Doi, M. (2019). The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 11, e1427.
 58. Larson, E.D., Komori, H., Gibson, T.J., Ostgaard, C.M., Hamm, D.C., Schnell, J.M., Lee, C.Y., and Harrison, M.M. (2021). Cell-type-specific chromatin occupancy by the pioneer factor Zelda drives key developmental transitions in *Drosophila*. *Nat. Commun.* 12, 7153. <https://doi.org/10.1038/s41467-021-27506-y>.
 59. Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., and Mann, R.S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell* 147, 1270–1282. <https://doi.org/10.1016/j.cell.2011.10.053>.
 60. Loker, R., Sanner, J.E., and Mann, R.S. (2021). Cell-type-specific Hox regulatory strategies orchestrate tissue identity. *Curr. Biol.* 31, 4246–4255.e4. <https://doi.org/10.1016/j.cub.2021.07.030>.
 61. López-Rivera, F., Foster Rhoades, O.K., Vincent, B.J., Pym, E.C.G., Bragdon, M.D.J., Estrada, J., DePace, A.H., and Wunderlich, Z. (2020). A mutation in the *Drosophila melanogaster* eve stripe 2 minimal enhancer is buffered by flanking sequences. *G3 (Bethesda)* 10, 4473–4482. <https://doi.org/10.1534/G3.120.401777>.
 62. Halfon, M.S. (2019). Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases. *Trends Genet.* 35, 93–103. <https://doi.org/10.1016/j.tig.2018.11.004>.
 63. Lindhorst, D., and Halfon, M.S. (2022). Reporter gene assays and chromatin-level assays define substantially non-overlapping sets of enhancer sequences. <https://doi.org/10.1101/2022.04.21.489091>.
 64. Song, Y., Park, J.O., Tanner, L., Nagano, Y., Rabinowitz, J.D., and Shvartsman, S.Y. (2019). Energy budget of *Drosophila* embryogenesis. *Curr. Biol.* 29, R566–R567. <https://doi.org/10.1016/J.CUB.2019.05.025>.
 65. Kvon, E.Z., Zhu, Y., Kelman, G., Novak, C.S., Plajzer-Frick, I., Kato, M., Garvin, T.H., Pham, Q., Harrington, A.N., Hunter, R.D., et al. (2020). Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell* 180, 1262–1271.e15. <https://doi.org/10.1016/J.CELL.2020.02.031>.
 66. le Poul, Y., Xin, Y., Ling, L., Mühling, B., Jaenichen, R., Hörl, D., Bunk, D., Harz, H., Leonhardt, H., Wang, Y., et al. (2020). Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci. Adv.* 6, 1262–1271.e15. <https://doi.org/10.1126/sciadv.abe2955>.
 67. Frankel, N.S., Erezilmaz, D.F., McGregor, A.P., Wang, S., Payre, F., and Stern, D.L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474, 598–603. <https://doi.org/10.1038/nature10200>.
 68. Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., and Carroll, S.B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326, 1663–1667. <https://doi.org/10.1126/science.1178357>.
 69. Guerreiro, I., Nunes, A., Woltering, J.M., Casaca, A., Nóvoa, A., Vinagre, T., Hunter, M.E., Duboule, D., and Mallo, M. (2013). Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc. Natl. Acad. Sci. USA* 110, 10682–10686. <https://doi.org/10.1073/pnas.1300592110>.
 70. Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167, 633–642.e11. <https://doi.org/10.1016/J.CELL.2016.09.028/ATTACHMENT/137E2323-216B-446E-869C-32653F4064C5/MC1.PDF>.
 71. Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. <https://doi.org/10.1126/science.1126410>.
 72. Monroe, J.G., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., Klein, M., Hildebrandt, J., Neumann, M., Kliebenstein, D., et al. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature* 602, 101–105. <https://doi.org/10.1038/s41586-021-04269-6>.
 73. Emera, D., and Wagner, G.P. (2012). Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Brief. Funct. Genomics* 11, 267–276. <https://doi.org/10.1093/BFGP/ELS013>.
 74. Oliver, K.R., and Greene, W.K. (2009). Transposable elements: powerful facilitators of evolution. *BioEssays* 31, 703–714. <https://doi.org/10.1002/BIES.200800219>.
 75. Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. <https://doi.org/10.1038/nrg2337>.
 76. Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762. <https://doi.org/10.1101/GR.080663.108>.
 77. Gilbertson, S.E., Walter, H.C., Gardner, K., Wren, S.N., Vahedi, G., and Weinmann, A.S. (2022). Topologically associating domains are disrupted by evolutionary genome rearrangements forming species-specific enhancer connections in mice and humans. *Cell Rep.* 39, 110769. <https://doi.org/10.1016/J.CELREP.2022.110769>.
 78. Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* 32, 225–237. <https://doi.org/10.1016/j.tig.2016.01.003>.
 79. Galupa, R., and Heard, E. (2017). Topologically associating domains in chromosome architecture and gene regulatory landscapes during

- p>development, disease, and evolution. Cold Spring Harb. Symp. Quant. Biol. 82, 267–278.
- <https://doi.org/10.1101/sqb.2017.82.035030>
- .
80. Eichenlaub, M.P., and Ettwiller, L. (2011). De novo genesis of enhancers in vertebrates. PLoS Biol. 9, e1001188. <https://doi.org/10.1371/JOURNAL.PBIO.1001188>.
 81. Li, S., Hannehalli, S., and Ovcharenko, I. (2022). De novo human brain enhancers created by single nucleotide mutations. <https://doi.org/10.1101/2021.07.04.451055>.
 82. Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., et al. (2012). Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 22, 1059–1068. <https://doi.org/10.1101/GR.133546.111>.
 83. Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., FitzPatrick, D.R., Afzal, V., et al. (2008). Human-specific gain of function in a developmental enhancer. Science 321, 1346–1350. <https://doi.org/10.1126/science.1159974>.
 84. Vaishnav, E.D., de Boer, C.G., Molinet, J., Yassour, M., Fan, L., Adiconis, X., Thompson, D.A., Levin, J.Z., Cubillos, F.A., and Regev, A. (2022). The evolution, evolvability and engineering of gene regulatory DNA. Nature 7907, 455–463. <https://doi.org/10.1038/s41586-022-04506-6>.
 85. Yona, A.H., Alm, E.J., and Gore, J. (2018). Random sequences rapidly evolve into de novo promoters. Nat. Commun. 9, 1530. <https://doi.org/10.1038/s41467-018-04026-w>.
 86. Surkova, S., Golubkova, E., Mamon, L., and Samsonova, M. (2018). Dynamic maternal gradients and morphogenetic networks in *Drosophila* early embryo. Biosystems. 173, 207–213. <https://doi.org/10.1016/J.BIOSYSTEMS.2018.10.009>.
 87. Crocker, J., Noon, E.P.B., and Stern, D.L. (2016). The soft touch: low-affinity transcription factor binding sites in development and evolution. Curr. Top. Dev. Biol. 117, 455–469.
 88. Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008). Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. PLoS Genet. 4, e1000106. <https://doi.org/10.1371/JOURNAL.PGEN.1000106>.
 89. Cande, J., Goltsev, Y., and Levine, M.S. (2009). Conservation of enhancer location in divergent insects. Proc. Natl. Acad. Sci. USA 106, 14414–14419. <https://doi.org/10.1073/pnas.0905754106>.
 90. Wade, J.T., and Grainger, D.C. (2018). Spurious transcription and its impact on cell function. Transcription 9, 182–189. <https://doi.org/10.1080/21541264.2017.1381794>.
 91. Muers, M. (2013). Chromatin: evolutionary insights into nucleosomes. Nat. Rev. Genet. 14, 78. <https://doi.org/10.1038/nrg3412>.
 92. Gophna, U. (2018). The unbearable ease of expression—how avoidance of spurious transcription can shape G+C content in bacterial genomes. FEMS Microbiol. Lett. 365, 267. <https://doi.org/10.1093/FEMSLE/FNY267>.
 93. Laland, K.N., Uller, T., Feldman, M.W., Sterelny, K., Müller, G.B., Moczek, A., Jablonka, E., and Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. Proc. Biol. Sci. 282, 20151019. <https://doi.org/10.1098/RSPB.2015.1019>.
 94. Kermekchiev, M., Pettersson, M., Matthias, P., and Schaffner, W. (1991). Every enhancer works with every promoter for all the combinations tested: could new regulatory pathways evolve by enhancer shuffling? Gene Expr. 1, 71–81.
 95. Bergman, D.T., Jones, T.R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H.Y., Nasser, J., Kane, M., Rios, A., et al. (2022). Compatibility rules of human enhancer and promoter sequences. Nature 607, 176–184. <https://doi.org/10.1038/s41586-022-04877-w>.
 96. Li, X.C., Fuqua, T., van Breugel, M.E., and Crocker, J. (2022). Mutational scans reveal differential evolvability of *Drosophila* promoters and enhancers. <https://doi.org/10.1101/2022.10.17.512533>.
 97. Lack, J.B., Lange, J.D., Tang, A.D., Corbett-Detig, R.B., and Pool, J.E. (2016). A thousand fly genomes: an expanded *Drosophila* genome nexus. Mol. Biol. Evol. 33, 3308–3313. <https://doi.org/10.1093/MOLBEV/MSW195>.
 98. Fuqua, T., Jordan, J., Halavatyi, A., Tischer, C., Richter, K., and Crocker, J. (2021). An open-source semi-automated robotics pipeline for embryo immunohistochemistry. Sci. Rep. 11, 10314. <https://doi.org/10.1038/s41598-021-89676-5>.
 99. R Core Team (2021). R: a language and environment for statistical computing (R Foundation for Statistical Computing).
 100. Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living *Drosophila* embryos links polymerase activity to patterning. Curr. Biol. 23, 2140–2145. <https://doi.org/10.1016/J.CUB.2013.08.054>.
 101. Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enuameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic Acids Res. 39, D111–D117. <https://doi.org/10.1093/NAR/GKQ858>.
 102. Li, L., and Wunderlich, Z. (2017). An enhancer's length and composition are shaped by its regulatory task. Front. Genet. 8, 63. <https://doi.org/10.3389/fgene.2017.00063>.
 103. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018. <https://doi.org/10.1093/BIOINFORMATICS/BTR064>.
 104. Schneider, T.D., Stormo, G.D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. J. Mol. Biol. 188, 415–431. [https://doi.org/10.1016/0022-2836\(86\)90165-8](https://doi.org/10.1016/0022-2836(86)90165-8).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-betagalactosidase	Promega	Z378; RRID: AB_2313752
Donkey anti-mouse AlexaFluor 555	ThermoFisher	A31570; RRID: AB_2536180
Sheep anti-digoxigenin	Roche	11333089001; RRID: AB_514496
Rabbit anti-FITC	ThermoFisher	A889; RRID: AB_221561
Donkey anti-rabbit AlexaFluor 488	ThermoFisher	A21206; RRID: AB_2535792
Donkey anti-sheep AlexaFluor 555	ThermoFisher	A21436; RRID: AB_2535857
Chemicals, peptides, and recombinant proteins		
Blocking solution	Roche	11921673001
In-vitro transcription labeling kit	Roche	11175025910
Prolong Gold with DAPI	ThermoFisher	P36935
Paraformaldehyde ampules (10x 10ml)	EMS Diasum	15710
NaCl	Merck	1064041000
TritonX-100	Sigma	T9284
EGTA	Sigma	E3889
Heptane	Sigma	246654-1L
Heparin	Sigma	H3393-100KU
Methanol	Merck	1060091000
Formamide	Sigma	47671-1L-F
SSC 20x	Sigma	SRE0068-1L
Salmon Sperm DNA Solution UltraPure	ThermoFisher	15632011
Tween-20	Sigma	P9416
Ethanol	Merck	1009831000
Xylenes	VWR	ACRO422685000
Experimental models: Organisms/strains		
attP2 line	Bloomington Drosophila Stock Center	25710
VK33 line	Bloomington Drosophila Stock Center	32543
w1118 line	Bloomington Drosophila Stock Center	5905

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Justin Crocker (justin.crocker@embl.de).

Materials availability

Plasmids, DNA libraries and fly lines generated in this study will be made available from [lead contact](#) upon request. A mutual transfer agreement might have to be signed for some materials.

Data and code availability

- Data reported in this paper will be shared by the [lead contact](#) upon request.
- All original code is available in this paper's [supplemental information](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The experimental model in this study is *Drosophila melanogaster*. All transgenic lines generated in this study were based on lines attP2 (Bloomington Stock Number: 25710) and VK33 (Bloomington Stock Number: 32543). When outcrosses were needed, we used line w1118 (Bloomington Stock Number: 5905). Information on developmental stages used is described in the main text, figures and/or [method details](#) (see below). All fly lines were kept in mite-free conditions, grown at room temperature in the lab or in a room where temperature is set at 25°C, in plastic vials or plastic bottles supplemented with fly food (1.8% m/v yeast, 1% m/v soy flour, 8% m/v cornmeal, 8% m/v malt extract, 1.2% m/v agar, 2.2% m/v sugar beet syrup, 0.625% v/v propionic acid, 1.2% v/v nipagin 20% in ethanol). All fly lines were flipped to new vials/bottles at least every two weeks.

METHOD DETAILS

Fly strains and constructs

E3N library

We categorized the 749 variants from the original *E3N* library²¹ by the number of mutations each variant contained. Using python's random module, we randomly selected 10 lines from the single mutation bin to the 10 mutations bin (selection without replacement, 100 lines total). If the line had been lost, we re-ran the generator to select a replacement line. Each line was fixed, stained, and imaged with a control *E3N* WT. If there were less than 10 embryos in the correct orientation and developmental stage on the microscope slide, the line was removed from the analysis. In the end, we had 10 lines with one mutation, 9 with two mutations, 9 with three mutations, 9 with four mutations, 7 with five mutations, 9 with six mutations, 9 with seven mutations, 10 with eight mutations, 8 with nine mutations, and 10 with ten mutations (please see [Table S1](#) for details). Thus, 505 total mutations were tested in 91 mutants, with the average number of mutations in this dataset ~5.5 (505/91). Sequences of the selected lines are provided in [Table S1](#).

All other libraries

All mutant and random enhancer sequences were synthesized and cloned (GenScript) into pLacZattB plasmid at HindIII/XbaI site (see more details below about the synthesis of random sequences). *eveS2*-related lines were injected into attP2 line, all other constructs were injected into VK33 line; injections done by Genetivision. Transgenic lines were homozygosed and genotyped; sequences are listed in [Table S1](#).

Synthesis of random sequences

Random sequences were synthesised by GenScript using their Precision Mutant Libraries service. Thousands of random DNA variants were synthesised, with no selection requirements other than size (chosen to be ~180bp, which was at the time the maximum size that their technology allowed). More specifically, 70-100bp random sequences were synthesised as oligo pools flanked on one side by either HindIII or XbaI sites, and on the other side by a partial motif for BsaI (a TIS restriction endonuclease). These random sequences and the backbone plasmid (pLacZattB) were digested with HindIII, XbaI and/or BsaI, and digested fragments were then mixed and assembled to the final plasmid library. The structure of the final library is [backbone-HindIII-random-TIS-random-XbaI-backbone].

Embryos collection and fixation

Flies were loaded into egg collection chambers, left to acclimatize for 3-4 days and then embryos were collected for either four or sixteen hours, for early and late stages, respectively. Embryos were dechorionated in 5% bleach for 2min, abundantly rinsed with water and washed in a saline solution (0.1 M NaCl and 0.04% Triton X-100), before transfer to scintillation vials containing fixative solution (700 μ l 16% PFA, 1.7 ml PBS/EGTA, 3.0 ml 100% heptane). Embryos were fixed for 25 min, shaking at 250 rpm. The lower phase was then removed, 4.6 mL 100% methanol added and vials vortexed at maximum speed for 1min. The interphase and upper phase were removed and the embryos were washed thrice in fresh methanol. Embryos were stored at -20 °C until processed.

Reporter gene expression analysis

In situ hybridization (probes)

probes for *lacZ* (reporter) and *snail* (internal control) were generated from PCR products using the in vitro transcription (IVT) kit from Roche (#11175025910) and following manufacturer's instructions. A list of primer sequences for each PCR product can be found in [Table S1](#). For each gene, distinct PCR products were pooled before IVT reaction. Probes were diluted in hybridization buffer (Hyb; 50% formamide, 4X SSC, 100 μ g/mL salmon DNA, 50 μ g/mL heparin, 0.1% Tween-20) at 50ng/ μ L. Prior to hybridization, a probe solution was prepared (per sample, 50 ng of each probe in 100 μ L), denatured at 80 °C for 5min, then immediately put on ice for 5min, and finally incubated at 56 °C for 10min before added to the embryos.

In situ hybridization (procedure)

embryos stored in methanol were washed in methanol/ethanol (50:50), three-times in 100% ethanol and then permeabilized in xylenes (90% in ethanol) for 1h, after which embryos were washed six times in ethanol and three times in methanol. Embryos were then washed three times in PBT (PBS + 0.1% Tween-20) before post-fixation for 25min in fixative solution (225 μ l 16% PFA, 500 μ l PBT). Embryos were then washed several times in PBT for 40 min, followed by a wash in PBT/Hyb (50:50) at room temperature and a 30min-wash in pre-warmed Hyb at 56 °C. Embryos were then incubated with probe solution at 56°C overnight. The next day,

embryos were washed in Hyb (three quick washes followed by three 30-min washes), then in Hyb/PBT (50:50), then in PBT several times for one hour before incubated for 30 min in blocking solution (Roche #11921673001; diluted 1:5 in PBT). Embryos were then incubated in blocking + primary antibodies diluted 1:500 (anti-DIG, Roche #11333089001; anti-FITC, ThermoFisher #A889) at 4 °C overnight. The next day, embryos were washed in PBT (three quick washes followed by four 15-min washes), and then incubated at room temperature in blocking solution + secondary antibodies diluted 1:500 (AlexaFluor 488 and 555, ThermoFisher #A21206 and #A21436, respectively). After 2 hours, embryos were washed in PBT (three quick washes followed by four 15-min washes), mounted on Prolong Gold with DAPI (ThermoFisher, P36935) and left to cure overnight before imaging.

Immunofluorescence

embryos stored in methanol were washed in PBT (three quick washes followed by four 15-min washes), then in blocking solution for 30 min (Roche #11921673001; diluted 1:5 in PBT), before incubated overnight at 4 °C in blocking solution + primary antibody diluted 1:500 (mouse anti-beta-galactosidase, Promega #Z378). The next day, embryos were washed in PBT (three quick washes followed by four 15-min washes), and then incubated at room temperature in blocking solution + secondary antibody (donkey anti-mouse AlexaFluor 555, ThermoFisher #A31570). After 2 hours, embryos were washed in PBT (three quick washes followed by four 15-min washes), mounted on Prolong Gold with DAPI (ThermoFisher, P36935) and left to cure overnight before imaging.

Microscopy and data analysis

embryos were imaged using a confocal microscope Zeiss LSM 880 confocal. Images were processed using a combination of automated scripts with manual curation. For 3D plots showing signal intensity across three regions of the embryo (Figures 1I, 2K, 3G, and 5I), images were analyzed in ImageJ: a circular ROI of constant size was used to measure average intensity across the different regions (selected as shown in figures); a co-stain for *snail* was used to help demarcate the presumptive germ layers. The number of lines/embryos analyzed for each case are indicated in figure legends. For analyzing *E3N* mutant lines,⁹⁸ individual nuclei were identified using the automated threshold algorithm on ImageJ and a watershed to split large ROIs; average intensities for each nucleus were measured. For analyzing *eveS2* mutant lines, we used ImageJ to perform Z-projections of max intensity, and a MATLAB (version R2018b; The MathWorks, Inc.) automated image analysis pipeline (Data S1) was developed to capture expression signal along the AP axis on stage 5 embryos. For automated rotation, an ellipse was fitted on a masked embryo, and embryos were rotated based on the maximum Feret diameter. For quantification, a section with 30% of the height of the embryo was taken at a middle position and along the AP axis of each embryo. From this image section, the intensities from all the rows in the image matrix were averaged for each pixel position along the AP axis. The integration and analysis from each of these resultant AP embryo expression profiles were done in R.⁹⁹ These expression profiles were smoothed with a Gaussian filter and then a linear interpolation was performed in order to have fixed samples number for the AP axis. Background removal and normalization were done based on the 10% and 50% quantile intensities, respectively, from the last 20% of the egg length. All embryos expression profiles per each genetic line were bootstrapped in order to see their reporter expression distribution along the AP axis. The bootstrapping was done using a confidence interval of 95% with 1000 replicates. For analyzing *twiPE* mutant lines, we used ImageJ to perform background subtraction from Z-projections of max intensity, rotate embryos to a vertical position and select a ROI at a defined position based on the intersection between 50% of the embryo long axis and the border of the *snail* RNA signal. We then used MorphoLibJ plugin in ImageJ to mask nuclei (volume higher than 3) and extracted intensities. For analyzing *rhoNEE* mutant lines, we used a custom code written in MATLAB (version R2018b; The MathWorks, Inc.) (Data S2); briefly, individual nuclei were segmented from the DAPI channel using a subroutine from the LivemRNA software package.¹⁰⁰ Stripes were then automatically identified by the following procedure: (1) bin nuclei by anterior-posterior (AP) coordinate; (2) within each bin, calculate a smoothed fluorescence profile along the dorsoventral (DV) coordinate based on the average fluorescence of each nucleus and its DV position; (3) identify peaks in the fluorescence profile for each bin; (4) align peaks across bins. Within each bin, nuclei falling within the AP coordinates for the half maximum height of a peak (on either side) were automatically considered to belong to the corresponding stripe. Manual curation was applied to fix any errors in stripe identification. Each stripe was then fitted lengthwise (AP axis) with a piecewise linear function through the middle, where for each line segment the stripe width was calculated perpendicular to the segment as the largest distance between the centers of nuclei “belonging” to the segment (i.e., nuclei with AP position falling between the AP coordinates of the two ends of the segment). Overall stripe width was calculated as the average of the widths of constituent segments. For analyzing *tinB* mutant lines, Z-projections of max intensity were generated using ImageJ and then embryos rotated and cropped to the minimum size in which the entire embryo still fitted the image. Composite images were then concatenated together and a montage was made using a scale factor of 1.0. Next, nuclear intensities were measured for each embryo in the montage. Channels were split, and in the DAPI channel the montage was smoothed twice. A threshold was manually set and applied, after which we used the “analyze particles” function based on a selection range of 100 to infinity. This threshold range was overlaid with the reporter channel, and nuclear intensities per embryo were retrieved using the ROI Manager.

Motif prediction analysis of random sequences

Position weight matrices (PWMs) for *Drosophila melanogaster* and their logos were obtained from FlyFactorSurvey.¹⁰¹ PWMs for specific stages of fly development were retrieved from Li and Wunderlich.¹⁰² Motif search analysis was done using FIMO¹⁰³ and setting a threshold p-value of 0.001. The top 30% highest PWM-scores were selected to explore putative candidates for TFs binding sites.

Information content

Information content for each of the TF motifs can be estimated using the Kullback-Leibler distance:

$$I_{\text{motif}} = \sum_{i=1}^L \sum_{n=A}^T p_{i,n} \log_2 \left(\frac{p_{i,n}}{b_n} \right)$$

where $p_{i,n}$ is the probability of observing the nucleotide “n” at position “i” and b_n is the background frequency of nucleotide “n”. These values can be an indicative of how frequent a motif hit is expected by chance where $2^{-I_{\text{motif}}}$ is an approximation of the probability for this event.^{43,104} The empirical cumulative distribution plot for the information content scores was done in R.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details of experiments can be found in the figure legends, figures and/or results, including the statistical tests used, exact value of n and what n represents.