Published in final edited form as:

Nature. 2020 November ; 587(7833): 235–239. doi:10.1038/s41586-020-2816-5.

Dense and pleiotropic regulatory information in a developmental enhancer

Timothy Fuqua[1,2], **Jeff Jordan**[3], **Maria Elize van Breugel**[1], **Aliaksandr Halavatyi**[1], **Christian Tischer**[1], **Peter Polidoro**[3], **Namiko Abe**[4], **Albert Tsai**[1], **Richard S. Mann**[4], **David L. Stern**[3], **Justin Crocker**[1]

[1]European Molecular Biology Laboratory, Heidelberg, Germany.

[2]Joint PhD Collaboration, EMBL and Faculty of Biosciences Heidelberg University, Heidelberg, Germany.

[3]Janelia Research Campus, Ashburn, VA, USA.

[4]Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA.

Abstract

Changes in gene regulation underlie much of phenotypic evolution[1]. However, our understanding of the potential for regulatory evolution is biased, because most evidence comes from either natural variation or limited experimental perturbations[2]. Using an automated robotics pipeline, we surveyed an unbiased mutation library for a developmental enhancer in *Drosophila melanogaster.* We found that almost all mutations altered gene expression and that parameters of gene expression–levels, location, and state–were convolved. The widespread pleiotropic effects of most mutations may constrain the evolvability of developmental enhancers. Consistent with these observations, comparisons of diverse *Drosophila* larvae revealed apparent biases in the phenotypes influenced by the enhancer. Developmental enhancers may encode a higher density of regulatory information than has been appreciated previously, imposing constraints on regulatory evolution.

Transcription factor binding sites are normally identified genetically, biochemically, or through phylogenetic footprinting[1], which has generated a biased understanding of the distribution and function of regulatory information in enhancers[2]. It has also limited our knowledge of the contributions of enhancer function to evolution. Despite many examples of individual nucleotide changes contributing to regulatory evolution, we have almost no

Correspondence and requests for materials should be addressed to D.L.S. or J.C. sternd@janelia.hhmi.org; justin.crocker@embl.de.

Author contributions T.F., M.E.v.B. and J.C. performed the measurements; R.S.M., D.L.S. and J.C. were involved in planning and supervised the work; T.F. and J.C. processed the experimental data, performed the analysis, drafted the manuscript and designed the figures with input from all the authors; J.J. and P.P. designed and built the robotics with input from J.C. A.H. and C.T. wrote the automated microscopy code and analysis tools. N.A. carried out the biochemistry experiments. T.F., A.T. and J.C. analysed the data.

Competing interests The authors declare no competing interests.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2816-5.

Peer review information *Nature* thanks Robb Krumlauf and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

understanding of what kinds of changes in enhancer function are evolutionarily accessible, or how the distribution of transcription factor binding sites might constrain the evolvability of enhancers.

Mutational scanning and saturation mutagenesis experiments can provide unbiased surveys of regulatory inputs[3-5], can detect weak regulatory interactions[6], and can identify the effects of mutations on phenotypic plasticity[7]. However, such scans are difficult to implement on developmental enhancers, because the precise expression patterns across embryos must also be examined. To facilitate these experiments on *Drosophila* developmental enhancers, we developed an automated robotics pipeline that allows the quantitative measurement of expression levels and patterns at multiple embryonic stages, across hundreds of transgenic fly lines.

Mutational scanning of the *E3N* enhancer

We selected an enhancer of the *D. melanogaster shavenbaby* (*svb*; also known as *ovo*) gene, *E3N*, which drives expression in ventral stripes of the embryo and is required to drive cell differentiation in the ventral denticle belts[8]. We chose *E3N* because the function of this enhancer is largely conserved across species, but its primary sequence has diverged. Also, it is only 292 base pairs (bp) long, but it integrates information from multiple patterning networks[9].

We generated mutagenized libraries of the *E3N* enhancer with an error-rate of about 2% per molecule, mimicking the divergence in this enhancer between *D. melanogaster* and *D. simulans* (Fig. 1a, Extended Data Fig. 1). We identified mutations in 272 of 292 bps, and the number of mutations per enhancer was approximately Poisson-distributed. Variants were combined with a heterologous *hsp70* promoter, with the caveat that this design may have discounted regulatory information that acts only upon the native *svb* promoter. Reporter constructs were integrated into the *Drosophila* genome at the same genomic location. We isolated 749 lines, which contained an average of seven mutations per enhancer, with a range between one and nineteen (Extended Data Fig. 1c).

To automate embryo handling and staining, we engineered tools for egg collections that interface with a chemical handling robot (Fig. 1b, Extended Data Fig. 2). The robot automates critical steps such as embryo fixation, vitelline membrane removal, and antibody or chemical staining of embryos, and this allows classical *Drosophila* fixation and immunohistochemistry protocols on 24 pools of embryos per experiment–providing increased throughput and reproducibility.

We first examined enhancer activity across a subset of 274 lines by staining embryos carrying enhancer variants driving a *lacZ* gene using a β-galactosidase assay. Although this assay is less sensitive than antibody staining, we found that the proportion of lines that showed no detectable *E3N* expression increased monotonically with the number of mutations in *E3N* (Fig. 1c), suggesting that the *E3N* enhancer contains a high density of regulatory information.

To identify relevant regions within the enhancer, we performed computational 'footprinting'[10]. Each mutated base in a line was scored either 0 for no change or 1 if either expression was lost or the pattern qualitatively did not resemble that of *E3N* (Fig. 1d). The total score (sum of all lines) for each base was normalized to the total number of mutations per base, smoothed, and plotted (Fig. 1e; see Methods and Extended Data Fig. 1e). We identified regions across the entire enhancer sequence that had effects on expression. Many of these regions overlapped with previously identified binding sites[8] or with consensus motifs for transcription factors, suggesting that regulatory information is distributed broadly across the enhancer.

We next calculated odds ratios across the screened lines to examine the extent to which each variant was associated with the loss of expression. This approach, which we term enhancer-wide association catalogue (EWAC), is derived from genome-wide association studies. This allowed us to test the contribution of each individual variant to expression–with the caveat that the limited sensitivity of the β-galactosidase assay may underestimate the role of some sites. We found that 22.9% (67/292) of mutations strongly altered expression ($q = 0.25$)[11] (Fig. 1f). These results identified key regions in the *E3N* enhancer that strongly alter expression when mutated, and suggested that many sites in the *E3N* enhancer are required to generate a wild-type expression pattern.

Most mutations alter gene expression

We developed an imaging pipeline for embryos stained with fluorescent antibodies to quantify phenotypic differences missed by the β-galactosidase assay (Extended Data Fig. 2). Cleared embryos were imaged using an adaptive feedback confocal microscope pipeline (see Methods). Images were subsequently compiled for analysis as montages for 117 lines (Extended Data Fig. 1) or registered using internal fiduciaries (Extended Data Fig. 3).

To examine the phenotypic effects of individual mutations on *E3N*, we screened the 18 lines from our collection that carried single mutations. All lines with single mutations showed a significant decrease in expression (Fig. 1g-i). A subset of embryos with single mutations is shown in Fig. 1g, demonstrating the substantial effects of most mutations (see also Extended Data Fig. 4). We next examined possible modes of transcriptional outputs for lines with single mutations: changes in levels, state, and location (Fig. 1h). We found that in 11 of 18 lines (about 61%), expression covaried with changes in transcriptional states or locations (Fig. 1i).

We next tested for a correlation between changes in nuclear intensity of expression for the single-base mutations and their respective conservation levels. We found no clear correlation using phyloP[12] estimates of evolutionary conservation among 27 insect species ($R^2 = 0.25$, two-tailed $P > 0.2$) or 124 insect species ($R^2 = 0.01$, two-tailed $P > 0.9$) (Fig. 1j, k, Extended Data Fig. 4t, u). These results suggest that sequence conservation is not an accurate predictor of the quantitative roles of individual sites in the *E3N* enhancer.

Mutational scanning maps binding sites

To validate these results, we selected eight lines from the library that carried mutations in a predicted Homothorax (Hth) binding motif with high footprint and EWAC scores (*Hth2*) and fewer than two mutations elsewhere (Fig. 2a, b). Although each line carried different mutations in *Hth2,* most lines drove expression in just a single row of cells. To confirm this result, we generated targeted knockouts of *Hth2* and other predicted *Hth* binding motifs in the *E3N* enhancer (Extended Data Fig. 5a-p). The *Hth2* targeted knockout exhibited low expression levels and expression in fewer cells than the wild type, similar to the mutant lines in our library (Extended Data Fig. 5g, h). We then performed Hth electromobility gel-shift assays (EMSAs) across the *E3N* enhancer (Extended Data Fig. 5q-y). We found that Hth, together with its cofactor Extradenticle (Exd), binds to four Hth motifs in vitro, including *Hth2,* while a variant Hth lacking a homeodomain (Hth^{HM}/Exd) does not (Extended Data Fig. 5s). Finally, binding of the Hox protein Ultrabithorax (Ubx) at a site adjacent to *Hth2* was enhanced in the presence of full-length Hth (Extended Data Fig. 5t), suggesting possible cooperativity between Hth and Ubx that requires the *Hth2* site (Fig. 2c).

Notably, the *Hth2* binding site in the *D. melanogaster E3N* enhancer is not present in *D. virilis* (Fig. 2c). By contrast, other sites, such as *Hth1* and *Exd*, are conserved (Fig. 2c, Extended Data Fig. 6). The *D. virilis E3N* enhancer (*E3N D. virilis*) in *D. melanogaster* drove an almost identical expression pattern to the *D. melanogaster E3N* with *Hth2* mutated (*E3N Hth2*), and both enhancers drove lower expression than the wild-type *mel-E3N* enhancer (*E3N D. mel*; Fig. 2d-f). To test whether the loss of the *Hth2* site in *D. virilis* contributes to the weaker expression of the *vir-E3N* enhancer, we ‘resurrected’ the *Hth2* motif in the *vir-E3N* enhancer (*E3N D. virilis + Hth2*; Fig. 2g). The addition of this binding site increased the expression driven by *vir-E3N* to nearly the level of the *mel-E3N* enhancer (Fig. 2h). Finally, we found that *D. virilis* larvae exhibit fewer trichomes in the domain in which *mel-E3N* is active (Fig. 2i, k), suggesting that the loss of the *Hth2* site in the *D. virilis E3N* enhancer may have contributed to the loss of ventral trichomes in *D. virilis.*

Mutations drive pleiotropic expression

Our findings demonstrate that the *E3N* enhancer contains regulatory information in most nucleotide positions. One consequence of this high density of information is that it may limit the expression patterns generated through stepwise mutations, and may bias change in some directions[13,14]. We investigated this possibility first by analysing the manifold consequences of mutations in Ubx binding sites. In *E3N*, Ubx acts as a transcriptional activator, and this enhancer contains multiple low-affinity Ubx binding sites[8,15].

We found that a previously characterized high-affinity variant of the Ubx motif[8] within the *E3N* enhancer activated *E3N* prematurely in stage 14 embryos (Fig. 3a-d). In addition, the expression intensities in anterior, early stripe, and naked (inter-stripe) regions driven by the Ubx high-affinity *E3N* enhancer were higher at stages 14 and 15 than for wild-type *E3N* (Fig. 3e). Furthermore, flies carrying the Ubx high-affinity *E3N* enhancer driving a *svb*-cDNA exhibited an increased number of trichomes in stripe regions and ectopic trichomes between segments (Extended Data Fig. 6c-e).

We next investigated whether these trends extended across a range of Ubx affinities. Using No Read Left Behind (NRLB)[15] to calculate the total Ubx affinity across the library, we selected mutated *E3N* enhancers with an estimated range of total Ubx affinities, with the number of mutations outside the Ubx binding sites ranging from 0 to 3. We analysed the anterior, stripe, and naked (inter-stripe) region intensities for each of the lines (Fig. 3f). We found that phenotypes in these regions were correlated and expression levels within all domains of expression increased with increased Ubx affinities (Fig. 3f). The increased expression levels we observed may be associated with precocious expression of these enhancers. Enhancers with increased Ubx affinity drove expression one embryonic stage earlier than normal and drove ectopic expression in naked regions and anterior domains. These enhancers may have higher sensitivity both to Ubx and to other homeodomain factors[8].

To investigate whether pleiotropic effects were observed in other mutated enhancers, we examined enhancer activity across a larger number of lines. The proportion of lines with ectopic *E3N* expression increased with the number of mutations (Extended Data Fig. 7), and 32.5% (38/117) of all lines generated obvious ectopic expression at stage 15, where about 34% (13/38) drove ectopic expression between the stripes. We observed pleiotropic expression in wing and haltere imaginal discs, developing mouth hooks, and other regions outside of the normal *E3N* expression domains (Extended Data Fig. 7). We next used the 274 lines and data from the β-galactosidase assay to calculate footprinting scores for pleiotropic hotspots. Unlike our loss-of-function results (Fig. 1), there was no clear signal (Extended Data Fig. 7k), possibly owing to the many mechanisms available to generate ectopic expression (Extended Data Fig. 7c-j). However, an EWAC analysis identified multiple regions (47/292 bases, about 16%) that were significantly associated with ectopic expression ($q = 0.25$, Extended Data Fig. 7l).

From the EWAC analysis, we noted that each of the Pangolin (Pan) binding motifs[16] was associated with pleotropic expression. Pan is a transcription factor that responds to Wingless/Wnt signalling and represses *svb* expression in naked regions on the ventral larval cuticle[9] (Fig. 4a, b). Disruption of *wingless* (*wg*) function leads to overexpression of Svb, resulting in a lawn of trichomes on the larval cuticle (Fig. 4a). Furthermore, 46% (20/43) of lines with mutations in Pan sites from our β-galactosidase assay showed ectopic patterns of expression (Extended Data Fig. 7m), often coupled with reduced expression in the stripe domains (18/43 lines, 42%). Thus, loss of Pan binding sites may lead to ectopic expression and trichome development.

To validate our findings, we selected lines from our library that carried mutations in Pan motifs (Fig. 4c), including a line with a single mutation near the *Hth2* site (Fig. 4b-d). Across the 13 lines carrying mutations in this Pan motif, we found both lower levels of expression in the denticle stripe domain and ectopic expression in the naked region (Fig. 4e, Extended Data Fig. 7n). Increased ectopic expression was negatively correlated with overall expression levels (Fig. 4f). These observations demonstrate that different aspects of enhancer expression may be correlated because of the pleiotropic effects of single nucleotide sites[17].

E3N architecture may constrain evolution

The widespread pleiotropic effects of *E3N* mutations suggest that the distribution of regulatory information may constrain enhancer evolution[18]. We therefore examined the larval cuticles of 60 *Drosophila* species (Fig. 4g, Extended Data Fig. 8) to see whether the ventral trichomes showed patterns of constraint. Although certain genetic perturbations can lead to the development of trichomes in the inter-stripe domain (Fig. 4a), we did not observe trichomes in this domain in any species (Fig. 4g-j). Furthermore, we detected at least twelve evolutionary losses of trichomes within ventral denticle bands. These results are consistent with the hypothesis that *E3N* activity is constrained across *Drosophila,* and that changes in *E3N* will more often lead to loss than to gain of trichomes.

Discussion

We used an unbiased method to explore regulatory information in a developmental enhancer. Strikingly, most mutations in *E3N* led to changes in transcriptional outputs, suggesting that regulatory information is distributed densely within this enhancer. The density and pleiotropic effects of mutations appear to constrain enhancer evolution. These observations were unexpected, as the DNA sequence of *E3N*, like most developmental enhancers, evolves rapidly, although its function is largely conserved. The high density of pleiotropic sites in this enhancer may help to explain why it is difficult in general to predict enhancer function from sequence and to build synthetic enhancers[2,19].

One explanation for the observed fragility of *E3N* and its rapid sequence evolution is that *E3N* has evolved along constrained mutational paths[14]. However, there are several potential caveats to this conclusion. First, removing an enhancer from its native location and placing it in a different chromatin environment may facilitate the binding of transcription factors that are not used in its native context[20]. Second, changes in binding sites may be buffered by regulatory information adjacent to the enhancer in its native context[21]. Third, our construct employed the *hsp70* promoter, and the activity of *E3N* may be modulated by the native *svb* promoter[22]. Finally, additional buffering can result from regulatory information encoded by partially redundant enhancers[23,24]. Future work could attempt similar experiments within native genomic contexts.

All mutations that generated ectopic expression in naked regions exhibited pleiotropic expression in other domains. Thus, it appears that there are no mutations in the lines we tested that enable *E3N* to escape this pleiotropic constraint. Consistent with this observation, we did not observe any *Drosophila* species with trichomes in these domains. Although it is possible that natural selection has not favoured the presence of trichomes in these domains, this potential regulatory constraint may help to explain the absence of trichomes in inter-stripe regions in the genus *Drosophila.* Other studies have shown that mutational changes in other *svb* enhancers tend to act pleiotropically across most segments of the larval body[25], suggesting that multiple *svb* enhancers are constrained.

There are additional sources of enhancer constraint. For example, *E3N* encodes low-affinity binding sites that confer specificity for a subset of Hox proteins[8]. In the case of *E3N,*

cooperativity between Hth and Ubx may enable the enhancer to drive precise patterns of expression, but increase its fragility to mutations. Finally, spatial precision can be encoded through binding site competition between transcriptional activators and repressors[26], consistent with the overlapping Pan and Hth motifs we observed in *E3N.* It may be challenging to separate the regulatory inputs of either factor independently through sequence turnover during evolution. This co-dependency may constrain evolvability because modification of overlapping sites leads to pleiotropic effects.

Our findings are consistent with the view that enhancer pleiotropy may be extensive in animal genomes, challenging the view of enhancer modularity[25]. We demonstrate an approach for exploring modes of regulatory evolution that considers both 'the possible and the actual' changes operating on regulatory regions[27], and may, therefore, allow us to predict changes in an evolutionary context.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2816-5.

Methods

Fly strains and constructs

All reporter constructs and the *E3N* mutant library were synthesized and cloned (GenScript) into the *placZattB* reporter construct. All mutant genetic sequences are listed in Supplementary Table 1.

Embryo fixation and robotics

Drosophila were loaded into egg collection chambers (Extended Data Fig. 2). Embryos were collected overnight and washed in a saline solution (0.1 M NaCl and 0.04% Triton X-100), dechorionated in 50% bleach for 90 s, and rinsed with water. For manual fixation, embryos were transferred to scintillation vials containing fixative solution (700 μl 16% PFA, 1.7 ml PBS/EGTA, 3.0 ml 100% heptane) and fixed for 25 min, shaking at 250 rpm. The lower phase was separated, and embryos were shocked isotonically using 100% methanol and rapid vortexing for 30 s. The interphase and upper phase were removed and the embryos were washed in fresh methanol. Fixation and antibody staining were tested on a series of control wells and showed no significant difference in sample fluorescence (Extended Data Fig. 2)

For automated fixation, dechorionated embryos were transferred to fixation plates (Extended Data Fig. 2) and loaded into the robot.

Antibody staining and cuticle preparations

Fixed embryos were stained using standard procedures with a chicken ant-βGal (1:500, abcam ab9361) and mouse anti-ELAV supernatant (1:25, DSHB), and conjugated with

AlexaFluor 488 and 633 secondary antibodies, respectively (1:500, Invitrogen). *Drosophila* cuticles were prepared using standard protocols[28]. The anti-ELAV antibody was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at the Department of Biology, University of Iowa.

Protein purification and EMSAs

Ubx (isoform IVa), HthHM-Exd, and HthFL-Exd constructs, protein purification, and EMSA conditions were as described previously[8].

Embryo mounting

Fixed and antibody-stained embryos were either mounted on Prolong Gold (Thermo Fischer Scientific) or mounted in BABB (benzyl alcohol/ benzyl benzoate). For BABB mounting, embryos were serially washed into 100% ethanol, 2× in BABB, and incubated overnight.

A Grace Silicone 2 × 4-well Isolator 2 × 4 (JTR8R-A2-0.5, Grace Bio-Labs) was cut in half and applied with another Isolator to a 75 × 50-mm microscope slide (Corning). Using a micropipette, 100 μl of the embryos in BABB solution were transferred to each well. The embryos were allowed to sink, and wells were connected with a thin layer of BABB and covered with a coverslip. The coverslip was sealed with three coats of clear nail polish.

Microscopy and data analysis

Cuticle preparations were imaged on a phase-contrast microscope (Zeiss). Confocal images were acquired on a Zeiss LSM 880 confocal microscope (Zeiss) either manually or using developed adaptive feedback microscopy pipeline[29] (https://git.embl.de/grp-almf/feedback-fly-embryo-crocker). The pipeline processes acquired images during the experiment and guides the microscope via the MyPic macro[30] to automatically acquire high-zoom images in the identified positions. Low-resolution overview 3D tile scans were acquired using a 5×/0.16 NA air objective and used to detect the lateral positions of embryos. For each selected position, low-resolution 3D stacks were acquired with a 20×/0.8 NA air objective lens. The low-resolution stacks were automatically analysed to obtain the embryo's bounding box. Within the bounding box a multichannel 3D stack was acquired with the same 20×/0.8 NA air objective, but now at high resolution for quantification. Imaging parameters such as lasers, emission filters, step-sizes, etc. can be found in the .lsm files at the URL provided above.

Using Fiji (https://imagej.net/Fiji)[31], images for each embryo were max-projected and compiled into montages using the MontageMaker plugin (Extended Data Fig. 3a). Embryos could also be registered (Extended Data Fig. 3b-d). Embryos were rotated in 3D space using a fiducial stain (ELAV, DSHB). Maximum projections of the ventral half of the embryo were calculated, and multiple images elastically transformed to one another[32], creating composite expression patterns.

Phenotypes were analysed using a sliding window method (Extended Data Fig. 3e), where a box was drawn between the T1 and T2 segments and centred over T2, and the average intensity for the T2 stripe was taken as the ROI was dragged down the embryo. Additionally,

a state-measuring method was employed (Extended Data Fig. 3f), wherein a nucleus-sized circlular ROI measured nuclear intensities down a single column of cells. Finally, entire expression profiles in Fig. 3 were generated by drawing a single box around the expression pattern and using ProfilePlot (Extended Data Fig. 3g). Box plots were generated in Matlab using the notBoxPlot plugin[33], and violin plots for state-measurements were generated in Matlab using distribution Plot.m[34].

Individual nuclei in Fig. 1 were identified using the automated threshold algorithm on FIJI and a watershed to split large ROIs. Average intensities for each nucleus were measured and plotted. The number of nuclei and embryos from left to right are n = 375, 72, 278, 98, 142, 169, 168, 247, 169, 136, 107, 325, 177, 211, 241, 221, 272, 256 nuclei; N = 3, 4, 10, 7, 7, 8, 10, 10, 10, 8, 5, 10, 10, 8, 8, 8, 8 embryos.

β-Galactosidase staining and imaging

Embryos were washed and dechorionated as described in 'Embryo fixation and robotics', and fixed in a 1:1 solution of fixative (2% formaldehyde + 0.2% glutaraldehyde + PBS) and 100% heptane for 20 min, shaking at 200 rpm. Fixative solution and heptane were removed, and embryos were patted dry on paper towels. Embryos were washed 3× in PBT for 10 min, shaking at 150 rpm. Embryos were mixed with β-galactosidase staining solution (5-bromo-4-chloro-3-indolyl-β-D-galactosidase (20 mg/ml DMF0, 400 mM potassium ferricyanide, 400 mM potassium ferrocyanide, 200 mM magnesium chloride, H_2O), and incubated at 37 °C for 1 h. Staining was stopped with 3× PBT washes. Embryos were imaged with a Leica DFC420C digital camera using a Leica MZ16F stereomicroscope.

Calculating footprinting scores

Enhancer sequences were aligned to each other using the pairwise2 alignment function in Biopython[35]. For this assay, deletions in the sequences were treated as mismatches, and we removed inserted bases for the analysis, making all sequences 292 bp. Alignments were changed to binary values, where 0 = match, and 1 = mismatch.

For each base (i = 1…292) in the *E3N* enhancer and each mutagenized line (j = 1…274), a score $a_{i,j}$ = 0 was assigned if the base in the line was not mutated. For a mutated base, $a_{i,j}$ = 1. The mutation coverage M_i (Fig. 1h) is the sum across all 274 lines:

$$M_i = \sum_{j}^{274} a_{i,j}$$

For each base (i = 1…292) in the *E3N* enhancer and each mutagenized line (j = 1…274), a score $s_{i,j}$ = 0 was assigned if the base in the line was not mutated or was mutated but did not alter the *E3N* expression pattern. For a mutated base that affected the *E3N* expression pattern, $s_{i,j}$ = 1. The total score S_i at each base i is the sum across all lines:

$$S_i = \sum_{j}^{274} s_{i,j}$$

If a specific base was mutated fewer than five times across all lines, the score was discarded ($S_i \equiv$ NaN in the data). A normalized footprint score σ_i was calculated as $\sigma_i = s_i/M_i$.

The plot in Fig. 1j was computed in Matlab using the Matlab smooth-data function with a Gaussian-weighted moving average window of 5 bases. Excluded bases (S_i = NaN) were ignored by this script as described in its documentation. Footprinting data are available in Supplementary Table 3.

Calculating EWAC scores

For each base (i = 1…292) in the *E3N* enhancer, the total score A_i at each base i is S_i (see 'Calculating footprinting scores') subtracted from the total coverage M_i: $A_i = M_i - S_i$. If a value was not available for S_i (S_i = NaN), A_i was set to 0.5.

The total score C_i at each base (i = 1…292) is the score A_i subtracted from the total number of lines without any expression (Q = 129), subtracted from the total number of lines (J = 274): $C_i = J - Q - A_i$.

The total score D_i at each base (i = 1…292) is the score S_i (see 'Calculating footprinting scores') subtracted from the total number of lines without any expression (Q = 129): $D_i = Q - S_i$.

For each base (i = 1…292) a 2 × 2 contingency table was generated for the values previously calculated:

	No change	Change
Mut+	A_i	S_i
Mut−	C_i	D_i

P values were calculated using chi2_contingency from SciPy[36] and compiled using Pandas[37] (Extended Data Fig. 1). EWAC scores in Fig. 1 are $-\log_{10}P$. Q values were calculated as described[11]. EWAC P values are available in Supplementary Table 2.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

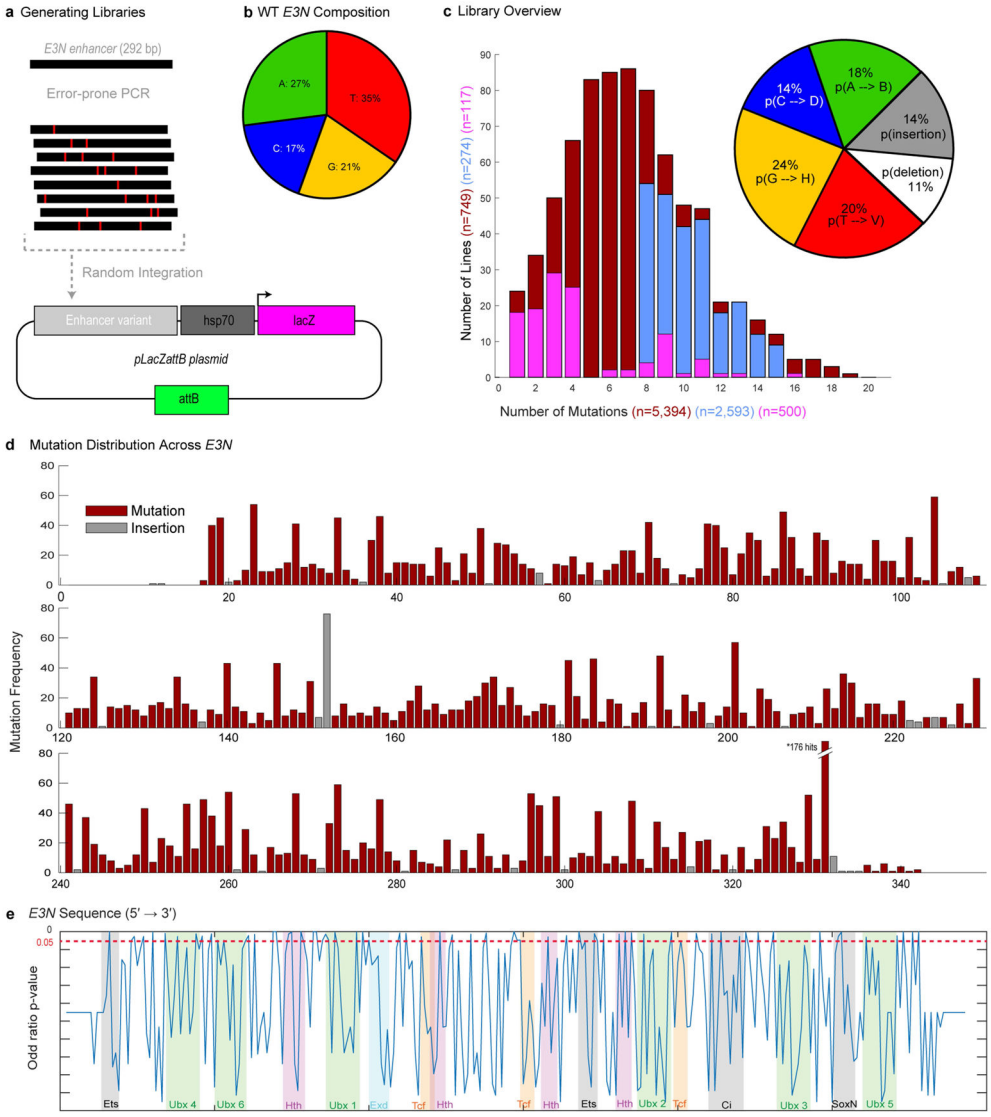
Data availability

The original images (cuticle preparations and embryo images, organized into zip files) and data files are available for download and are indexed at: https://www.embl.de/download/crocker/Dense_and_pleiotropic_regulatory_information_in_a_developmental_enhancer/index.html. All fly lines will be made available from the corresponding authors upon reasonable request.

Code availability

CAD files and links to the software can also be found at Github: https://github.com/tfuqua95/Flyspresso-CAD-files.

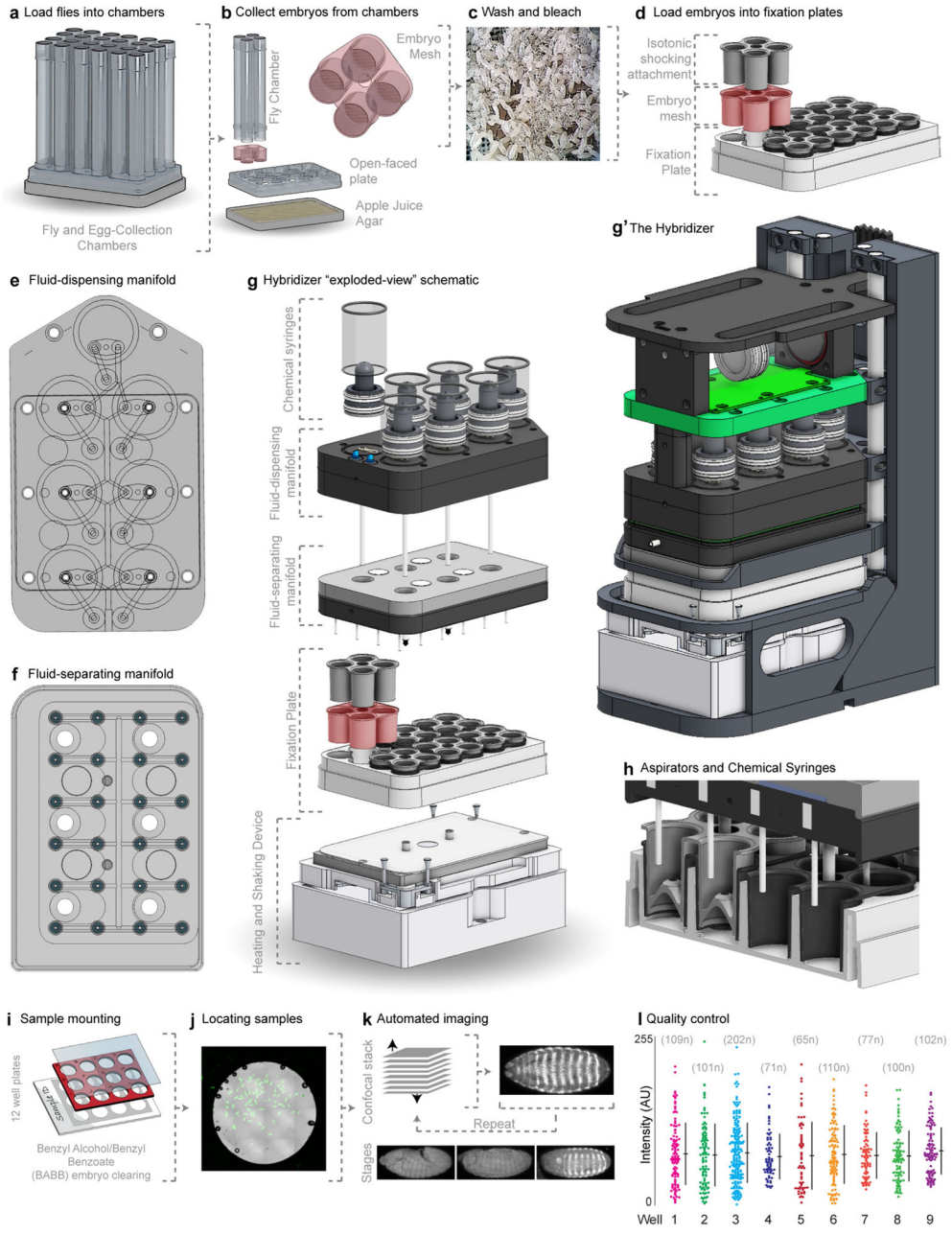
Extended Data



Extended Data Fig. 1 |. Distribution of mutations in the *E3N* enhancer library.

a, Mutant enhancer variants of *E3N* were created via degenerate PCR and integrated into the *placZattB* plasmid, which contains a minimalized core *hsp70* promoter and the *lacZ* reporter gene. Plasmids were integrated into the *Drosophila* genome at the *attP2* site. **b**, Pie chart depicting base-pair composition of the WT *E3N* enhancer. **c**, (Left) Histogram for all 749 mutants (dark red) is approximately normal with an average of 7 mutations per mutant. Magenta bars denote lines antibody stained (117 total), and blue lines indicate lines that were also Beta-Galactosidase stained (274 total). (Right) pie chart shows probability of

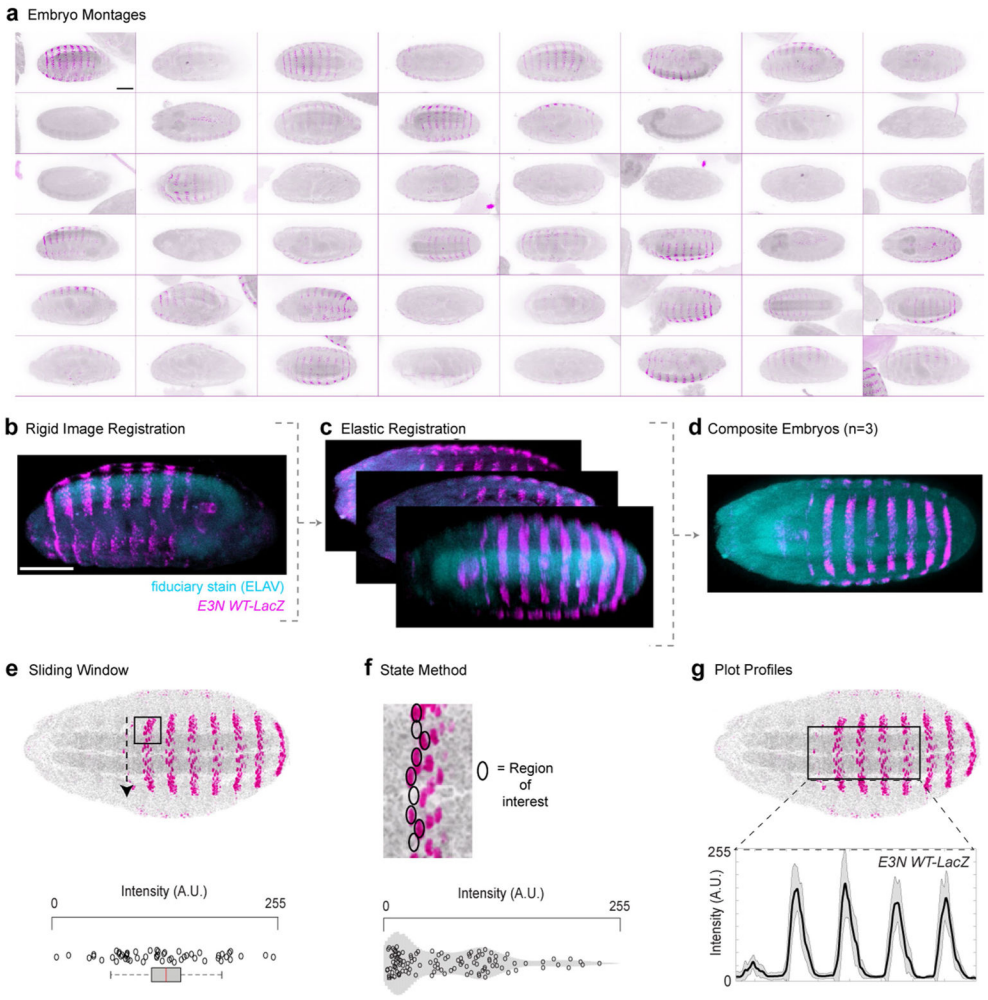
mutation normalized to ATCG composition (see **b**). **d**, Manhattan plot shows the summation of all mutations within the *E3N* library. **e**, Unsmoothened "footprinting scores" from Fig. 1h. Scores plotted linearly over transcription factor binding motifs (colored and shaded regions) across the *E3N* genomic sequence.



Extended Data Fig. 2 |. An automated platform for fixing, staining, and imaging *Drosophila* embryos.

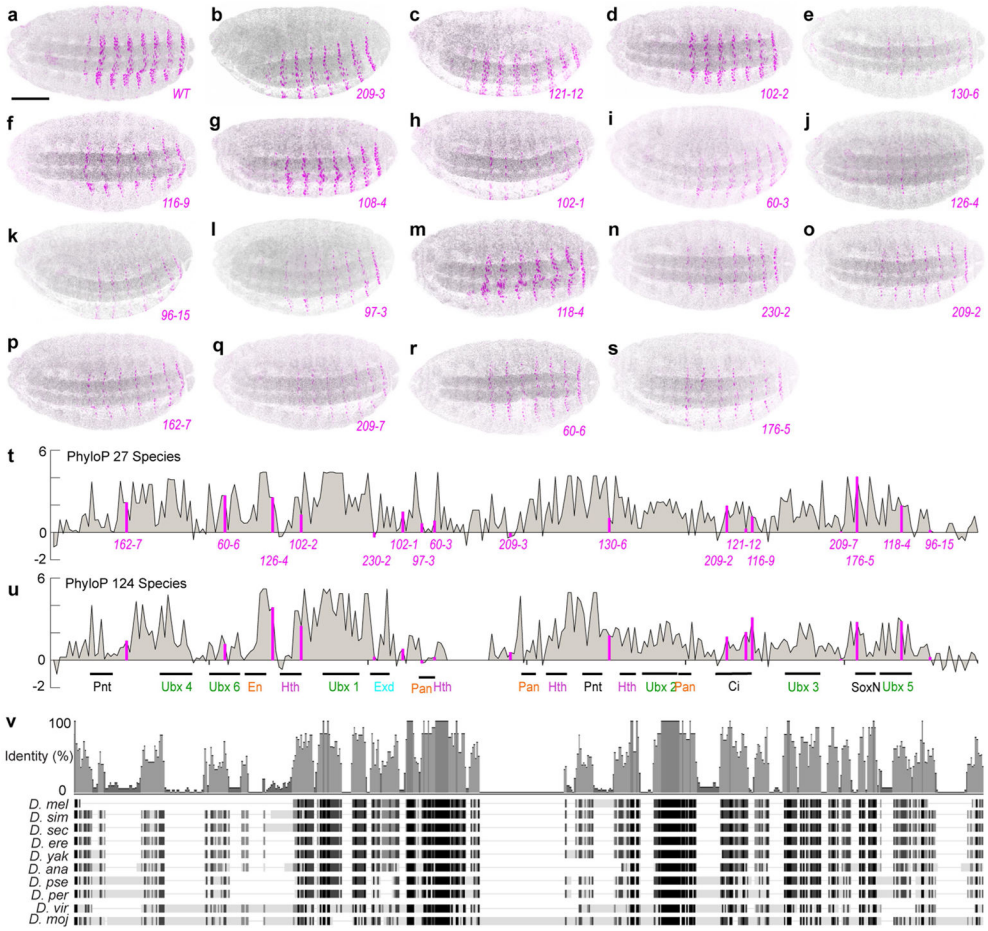
a–d, Collecting *Drosophila* embryos. (**a**) Custom fly chambers were made, holding up to 24 different strains. (**b**) An explosion-view of the fly chambers. Embryo meshes (red) can attach and detach from the fly chambers and are suspended above an apple juice-agar plate.

(**c**) Embryos are collected onto the embryo meshes and washed with saline solution and bleached. (**d**) Embryos are loaded into a fixation plate. **e–h**, Components of the robot. (**e**) The fluid-dispensing manifold. Seven pneumo-hydraulic syringe pumps are coupled to the fluid-dispensing manifold; one pump for priming chemicals into the fluid-dispensing manifold, and six pumps for dispensing chemicals into the fixation plate. (**f**) The fluid-separating manifold uses 24 small syringes to aspirate fluid from the isotonic shocking attachments. (**g** and **g'**) Different components of the robot. (**h**) Cross-section of the fixation plate and aspiration tips and syringes. 24 small aspiration tips draw fluid from the top of each well within the isotonic shocking attachment and six main dual-purpose tips dispense and aspirate fluid into and out of the bottom of the wells. **i–k**, The adaptive feedback imaging pipeline. (**i**) Samples are mounted on multi-well slides. (**j**) An overview tile-scan of each well is taken and x,y coordinates for embryos (green) are identified either manually or computationally. (**k**) For each coordinate, a fast, low-resolution confocal stack is automatically acquired. An algorithm determines the embryo's z position and rotation, yielding a bounding box within which a high-resolution, 3D stack of the entire embryo is acquired. See also Methods. **l**, Control *E3N* WT embryos were fixed and stained on the robot. A single embryo in the same orientation and age from each well was selected and the individual nuclear fluorescence intensities were measured in AU, arbitrary units of fluorescence intensity. In plots, centre line is mean, upper and lower limits are standard deviation.



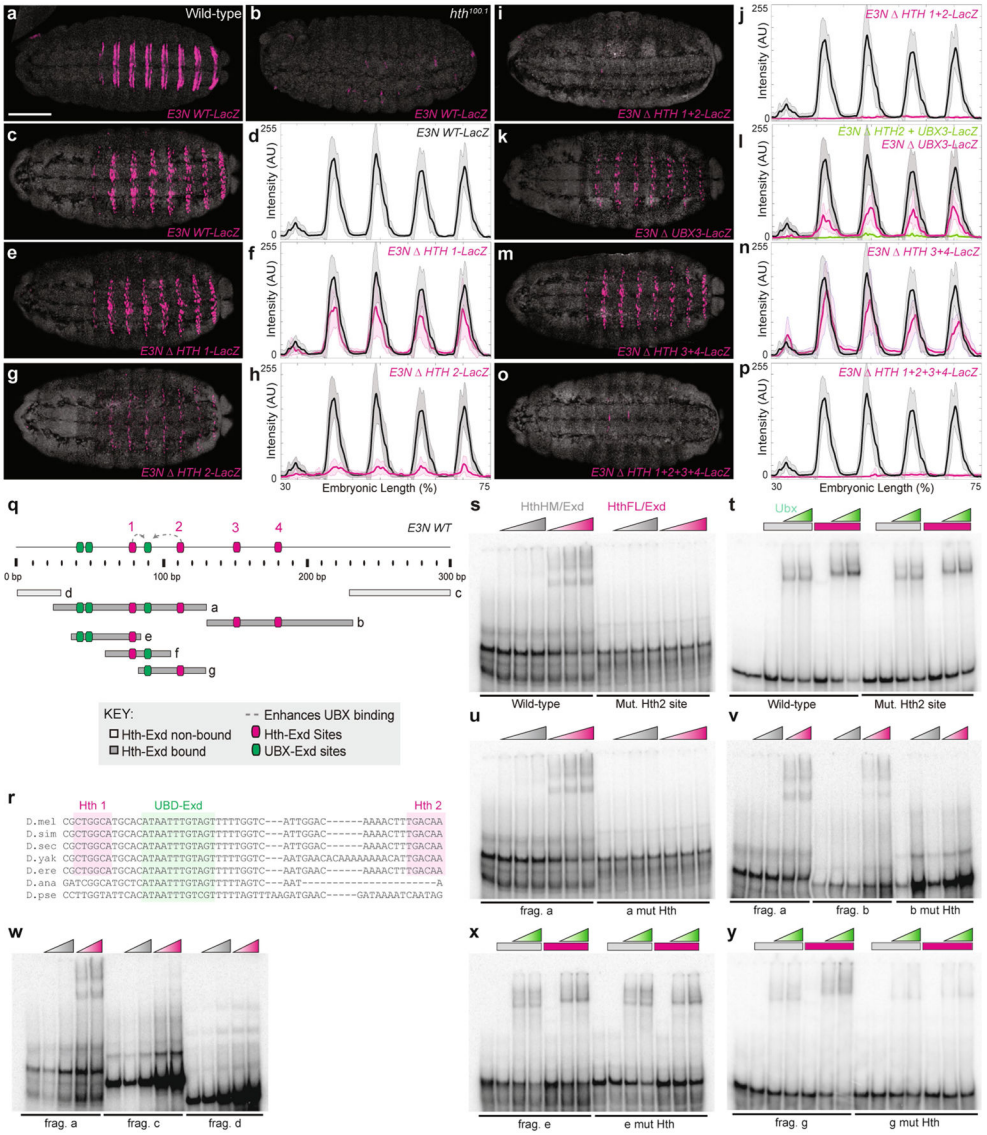
Extended Data Fig. 3 |. Methods of image and data analysis.

a, Images acquired from automated imaging are compiled into a large montage image. **b–d**, Registering multiple images using fiducials. An embryo acquired during automated imaging (**b**) can be automatically rotated in 3D space using ELAV (teal) as a fiducial. Once properly rotated, maximum projections of the ventral half can be computed (**c**). Finally, the 2-D projections can be elastically registered – or deformed – to align multiple samples (**d**). **e–g**, Methods of measuring expression patterns. (**e**) Sliding window analysis. A box is drawn between A2 and A3 and centred within A2. Multiple measurements are taken, sliding the box across the stripe. Each point on the boxplot represents one measurement within the box. In box plots, centre line is mean, upper and lower limits are standard deviation and whiskers show 95% CIs. (**f**) State method analysis. A row of cell-sized regions of interest are dragged down across the A2 stripe. Each point on the boxplot represents a single cell. (**g**) Plot profile analysis. A box is drawn from the A1 to A5 and the mean intensity is taken for each column of pixels and plotted (N = 10 embryos). Shaded areas indicate ± 1 SD, solid line is the mean expression. Scale bars, 100 μm. Embryos are matched to scale respectively in (**a**) and in (**b-e**), and (**g**).



Extended Data Fig. 4 |. Single base pair mutations and *E3N* conservation.

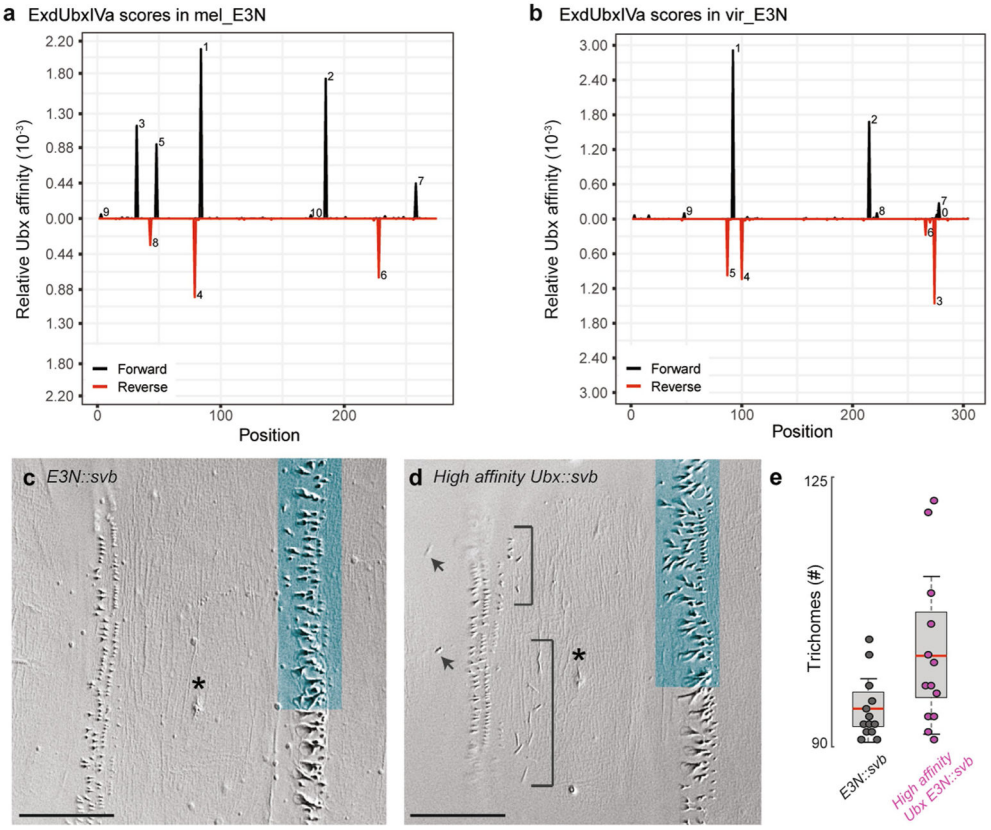
a–s, Example embryos carrying individual *E3N::lacZ* variants with single mutations. Constructs are ordered from smallest to largest effect sizes. **t**, **u**, PhyloP scores across the *E3N* enhancer sequence. Locations of the single mutations and their PhyloP scores are highlighted as magenta bars. **v**, *E3N* sequence alignment between 10 *Drosophila* species. Scale bars, 100 μm (**a**). Embryos are matched to scale respectively (**a – s**).



Extended Data Fig. 5 –. Testing additional Hth-Exd motifs in *E3N*.

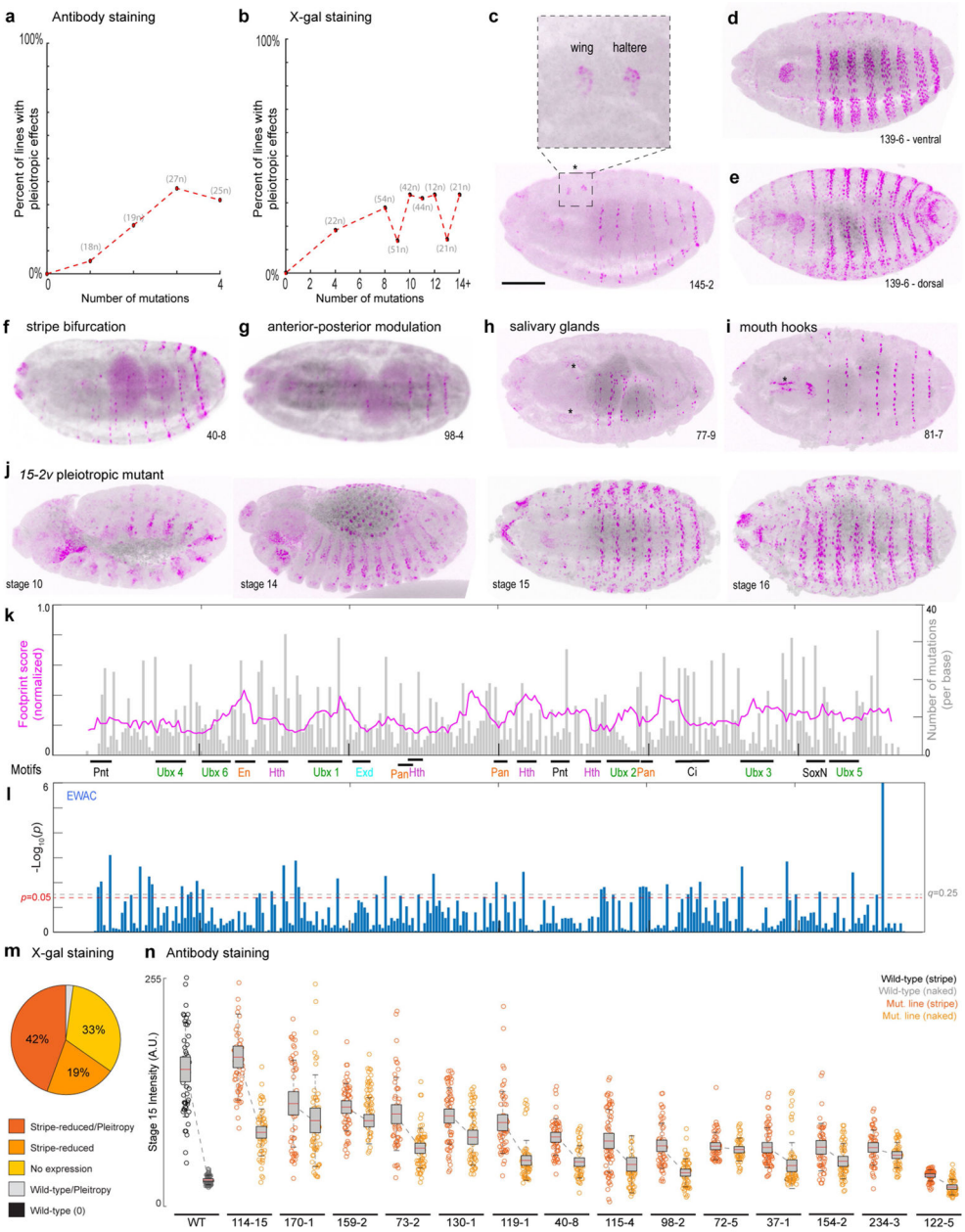
a, **b**, Embryos carrying *E3N::lacZ* reporter constructs in a WT w^{1118} background (**a**) and *hth* homeodomain-less (HthHM) $hth^{100.1}$ background stained with anti-β-Galactosidase (**b**). **c–p**, Embryos carrying *E3N::lacZ* reporter constructs stained with anti-β-Galactosidase adjacent to their respective expression plot profiles. Constructs contain mutations in Hth1 (CTGGCA → CCCCCC), Hth2 (TGACAA → CCCCCC), Hth3 (TTGTCG → CCCCCC), and Hth4 (TGAGAG → CCCCCC). (**c** and **d**) *E3N* WT. (**e** and **f**) *E3N* with Hth1 site changed. (**g** and **h**) *E3N* with Hth2 site changed. (**i** and **j**) *E3N* with Hth1 and Hth2 sites changed. (**k** and **l**) *E3N* with Ubx3 site changed (CATAATTTGT → CAGGGTTTGT). (**m** and **n**) *E3N* with Hth3 and Hth4 sites changed. (**o** and **p**) *E3N* with Hth1, Hth2, Hth3, and Hth4 sites changed. In all plots, the black and magenta lines denote the average expression driven by the wild-type and modified enhancers, respectively (n = 10 for each genotype). Shaded areas indicate ± 1 s.d. AU, arbitrary units of fluorescence intensity. **q**, top, Schematic for the *E3N* enhancer, denoting binding sites and possible protein-to-protein interactions. **q**,

bottom, Schematic for different *E3N* fragments tested. **r**, Multiple-species alignment of Hth1, 2 and the UBX-Exd site. **s–y**, Electromobility shift assays (EMSA) for different fragments of *E3N* denoted in (**q**). All EMSAs were run on native (non-denatured) gels. HthHM/Exd is the homeodomain-less (HthHM) isoform of Hth incubated with Exd. HthFL/Exd is the Hth isoform with a homeodomain, incubated with Exd. Fragments tested with the WT Hth binding site and a mutated form. (**s**) EMSA for *fragment-f* with Hth2 mutated (**t**) additionally with increasing Ubx concentrations. (**u)** EMSA for *fragment-a* with Hth1 and 2 mutated. (**v**) EMSA for *fragment-a* and *fragment-b* with Hth3 and Hth4 mutated. (**w**) EMSA for *fragment-a, fragment-c*, and *fragment-d*. (**x**) EMSA for *fragment-e* with Hth1 mutated. (**y**) EMSA for *fragment-g* with Hth2 mutated. Scale bars, 100 μm. Embryos are matched to scale.



Extended Data Fig. 6 | The effects of Ubx affinity on morphology.
a, **b**, Schematic output from NRLB[15] shows predicted binding affinity for Exd::Ubx heterodimers across the *E3N* sequence, where black peaks are on the 5′ strand and red peaks respectively on the 3′ strand. Affinity plots are shown for *Drosophila melanogaster* (**a**) and *Drosophila virilis* (**b**). **c**, **d**, *Drosophila* cuticle preps for flies with WT *E3N* driving *svb* cDNA (**c**), or *E3N* with increased Ubx binding affinity driving *svb* cDNA (**d**). Trichomes were counted within a region of interest (teal box) defined by anatomic epithelial sensory cells (*). Arrows and brackets demarcate ectopic trichomes. **e**, Boxplots comparing trichome numbers in the A1 segment in the region of interest from panels (**c**) and (**d**) ($n = 13$, $P <$

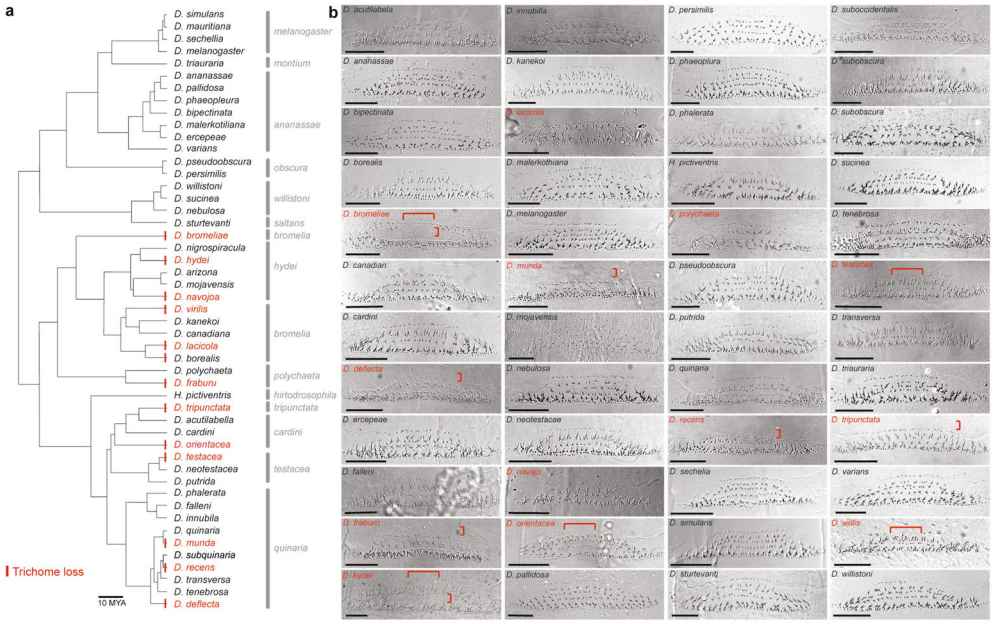
0.02), see also Tsai et al., 2018[24]. In box plots, centre line is mean, upper and lower limits are standard deviation and whiskers show 95% CIs. Scale bars, 25 μm each.



Extended Data Fig. 7 | Extensive pleiotropic effects across the *E3N* enhancer.

a, **b**, Plot comparing the percent of lines with pleiotropic or ectopic expression versus the number of mutations based on antibody staining (**a**) and Beta-Galactosidase staining (**b**). **c–j**, A subset of mutants with pleiotropic effects. (**c**) Line *145-2* drives ectopic expression in the developing wing and haltere discs (7/7 embryos). (**d**) Line *139-6* drives wider stripes and increased expression, as well as ectopic expression between the stripes and (**e**) on the dorsal side, (5/5 embryos). (**f**) Line *40-8* drives a split stripe pattern, where the middle row of

nuclei within the stripes is not active (6/6 embryos). (**g**) Line *93-4* expression varies along the anterior-posterior axis (5/5 embryos). (**h**) Line *77-9* drives ectopic expression in the salivary glands (5/5 embryos). (**i**) Line *81-7* drives expression in the developing mouth hooks (5/5 embryos.) (**j**) Line *15-2v* activates expression at stage 10 and drives ectopic expression throughout the embryo in multiple developmental stages (14/14 embryos). **k**, Plot of footprinting scores versus *E3N* sequence. Magenta is the footprinting score (σ_j, see methods). The higher the peak, the higher probability that a mutation will change expression. Gray plots are the mutation coverage for the number of lines screened per base (M_j, see methods). **l**, EWAC scores represent p-values from a log of odds ratio test for the association of a mutation changing expression. Dashed lines denote p- and q-values[11], respectively. See Materials and Methods. **m**, Plot comparing the percent of lines with changed expression for mutations in the overlapping Pan/Hth site. **n**, Quantification of the staining intensities in the stripe and naked domains with the indicated reporter construct using the "sliding window" technique (Extended Data Fig. 3e). N = the number of embryos per line measured, from left to right: 10, 10, 7, 9, 10, 7, 10, 6, 10, 8, 4, 8, 10, 10, 10. In box plots, centre line is mean, upper and lower limits are standard deviation and whiskers show 95% CIs. Scale bars, 100 μm. Embryos are matched to scale (**c – j**).



Extended Data Fig. 8 l. Cuticle preps from 60 *Drosophila* species across approximately 100 million years of evolution.

a, Phylogenetic tree of *Drosophila* species studied here, spanning approximately 150 million years of evolution. Red indicates a loss of trichomes. **b**, Representative cuticle preps for *Drosophila* species. See also Fig. 4. Scale bars, 25 μm each.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

A.T. is supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, TS 428/1-1) and EMBL. R.S.M. is supported by funding from R35GM118336. We thank J. Zaugg and J. Wirbel for statistical advice; C. J. Standley, X. Li, R. M. Galupa, G. A. Canales, M. Perkins and M. R. P. Alves for comments on the manuscript; C. Rastogi and H. Bussemaker for the NRLB algorithms; K. Richter and J. Sager for technical support; I. Jones for mounting the *Drosophila* species larvae; T. Laverty for *Drosophila* assistance; A. Milberger for help with CAD design; and J. Payne for discussions.

References

1. Wittkopp PJ & Kalay G Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet 13, 59–69 (2011). [PubMed: 22143240]
2. Crocker J & Ilsley GR Using synthetic biology to study gene regulatory evolution. Curr. Opin. Genet. Dev 47, 91–101 (2017). [PubMed: 28968519]
3. Mogno I, Kwasnieski JC & Cohen BA Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. Genome Res. 23, 1908–1915 (2013). [PubMed: 23921661]
4. Patwardhan RP et al. Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol 30, 265–270 (2012). [PubMed: 22371081]
5. Weingarten-Gabbay S et al. Systematic interrogation of human promoters. Genome Res. 29, 171–183 (2019). [PubMed: 30622120]
6. de Boer CG et al. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nat. Biotechnol 38, 56–65 (2020). [PubMed: 31792407]
7. Duveau F, Yuan DC, Metzger BPH, Hodgins-Davis A & Wittkopp PJ Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. Proc. Natl Acad. Sci. USA 114, E11218–E11227 (2017). [PubMed: 29259117]
8. Crocker J et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell 160, 191–203 (2015). [PubMed: 25557079]
9. Payre F Genetic control of epidermis differentiation in *Drosophila*. Int. J. Dev. Biol 48, 207–215 (2004). [PubMed: 15272387]
10. Belliveau NM et al. Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. Proc. Natl Acad. Sci. USA 115, E4796–E4805 (2018). [PubMed: 29728462]
11. Storey JD, Tibshirani R Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003). [PubMed: 12883005]
12. Pollard KS, Hubisz MJ, Rosenbloom KR & Siepel A Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121 (2010). [PubMed: 19858363]
13. Smith JM et al. Developmental constraints and evolution: a perspective from the Mountain Lake Conference on Development and Evolution. Q. Rev. Biol 60, 265–287 (1985).
14. Uller T, Moczek AP, Watson RA, Brakefield PM & Laland KN Developmental bias and evolution: a regulatory network perspective. Genetics 209, 949–966 (2018). [PubMed: 30049818]
15. Rastogi C et al. Accurate and sensitive quantification of protein-DNA binding affinity. Proc. Natl Acad. Sci. USA 115, E3692–E3701 (2018). [PubMed: 29610332]
16. Chang MV, Chang JL, Gangopadhyay A, Shearer A & Cadigan KM Activation of wingless targets requires bipartite recognition of DNA by TCF. Curr. Biol 18, 1877–1881 (2008). [PubMed: 19062282]
17. Nagy O et al. Correlated Evolution of Two Copulatory Organs via a Single cis-Regulatory Nucleotide Change. Curr. Biol 28, 3450–3457.e13 (2018). [PubMed: 30344115]
18. Sabarís G, Laiker I, Preger-Ben Noon E & Frankel N Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. Trends Genet. 35, 423–433 (2019). [PubMed: 31005339]
19. Vincent BJ, Estrada J & DePace AH The appeasement of Doug: a synthetic approach to enhancer biology. Integr. Biol 8, 475–484 (2016).

20. Dey SS, Foley JE, Limsirichai P, Schaffer DV & Arkin AP Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol 11, 806 (2015). [PubMed: 25943345]
21. Stern DL et al. Genetic and transgenic reagents for *Drosophila simulans, D. mauritiana, D. yakuba, D. santomea, and D. virilis*. G3 7, 1339–1347 (2017). [PubMed: 28280212]
22. Zabidi MA et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature 518, 556–559 (2015). [PubMed: 25517091]
23. Frankel N et al. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466, 490–493 (2010). [PubMed: 20512118]
24. Tsai A, Alves MR & Crocker J Multi-enhancer transcriptional hubs confer phenotypic robustness. eLife 8, e45325 (2019). [PubMed: 31294690]
25. Preger-Ben Noon E et al. Comprehensive analysis of a *cis*-regulatory region reveals pleiotropy in enhancer function. Cell Rep. 22, 3021–3031 (2018). [PubMed: 29539428]
26. Crocker J, Tsai A & Stern DL A fully synthetic transcriptional platform for a multicellular eukaryote. Cell Rep. 18, 287–296 (2017). [PubMed: 28052257]
27. Jacob F The Possible and the Actual (Univ. Washington Press, 1994).
28. Stern DL & Sucena E Preparation of cuticles from unhatched first-instar *Drosophila* larvae. Cold Spring Harb. Protoc 2011, 065532 (2011).
29. Tischer C, Hilsenstein V, Hanson K & Pepperkok R Adaptive fluorescence microscopy by online feedback image analysis. Methods Cell Biol 123, 489–503 (2014). [PubMed: 24974044]
30. Politi AZ et al. Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. Nat. Protoc 13, 1445–1464 (2018). [PubMed: 29844523]
31. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012). [PubMed: 22743772]
32. Arganda-Carreras I et al. in Computer Vision Approaches to Medical Image Analysis. CVAMIA 2006. Lecture Notes in Computer Science Vol. 4241 (eds Beichel RR & Sonka M) (Springer, 2006).
33. Campbell R notBoxPlot https://github.com/raacampbell/notBoxPlot (2020).
34. Jonas. Violin Plots for Plotting Multiple Distributions (distributionPlot.m) https://uk.mathworks.com/matlabcentral/fileexchange/23661-violin-plots-for-plotting-multiple-distributions-distributionplot-m (2020).
35. Cock PJA et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423 (2009). [PubMed: 19304878]
36. Virtanen P et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272 (2020). [PubMed: 32015543]
37. McKinney W Data structures for statistical computing in Python. Proc. 9th Python Sci. Conf (2010).

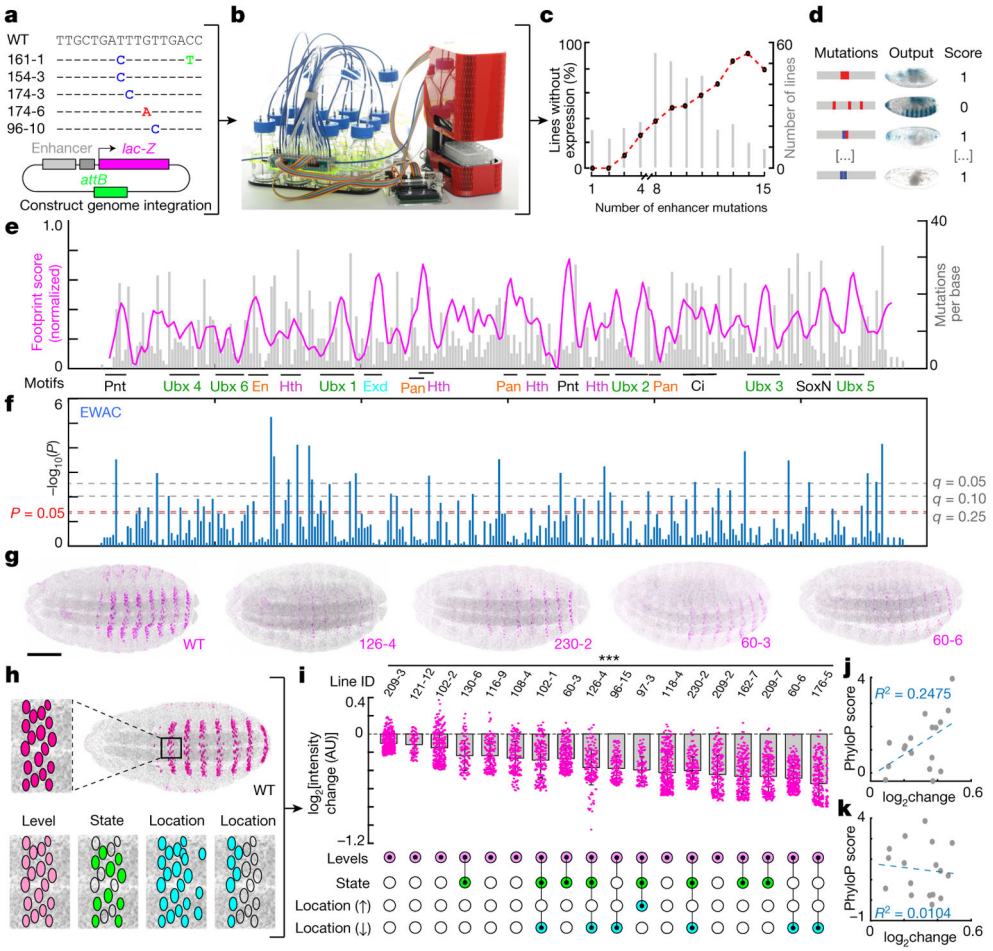


Fig. 1 |. Most nucleotide mutations in *E3N* alter gene expression.

a, Subset of *E3N* mutagenesis library and schematic of the reporter construct used for integration of enhancers into the *Drosophila* genome (Extended Data Fig. 1). **b**, The liquid-handling robot (Extended Data Fig. 2). **c**, Percentage of lines with no expression plotted against number of mutations per line. **d**, Lines were scored as 1 (positive) or 0 (no visible expression defects; see Methods). **e**, Footprinting scores plotted along the *E3N* sequence. Magenta line, footprinting score (σ_j, see Methods). Higher peaks show a higher probability that a mutation will change expression. Grey histogram, number of mutations per base for the screened lines (M_j, see Methods). **f**, EWAC scores represent *P* values (*t*-test, two-tailed) from a log of odds ratio test for the association of a mutation with changing expression (see Methods). **g**, Examples of fly embryos with single-mutant *E3N::lacZ* reporter constructs. WT, wild-type; numbers are line ID numbers. **h**, Schematic of possible changes to expression outputs. **i**, Top, nuclear intensity changes associated with single mutations compared to wild-type *E3N* ($n = 212$ nuclei; 8 embryos). Mean ± s.d.; ***$P < 0.01$, two-tailed *t*-tests. See Methods for sample sizes. AU, arbitrary units of fluorescence intensity. Bottom, Changes in expression output for the single mutations (Extended Data Fig. 3). **j, k**, Pearson correlations between mutation effect sizes and PhyloP scores for 27 (**j**) and 124 species (**k**) with least squares linear regression, and R^2 values. **g**, Scale bar, 100 μm; embryo in **h** matched for scale.

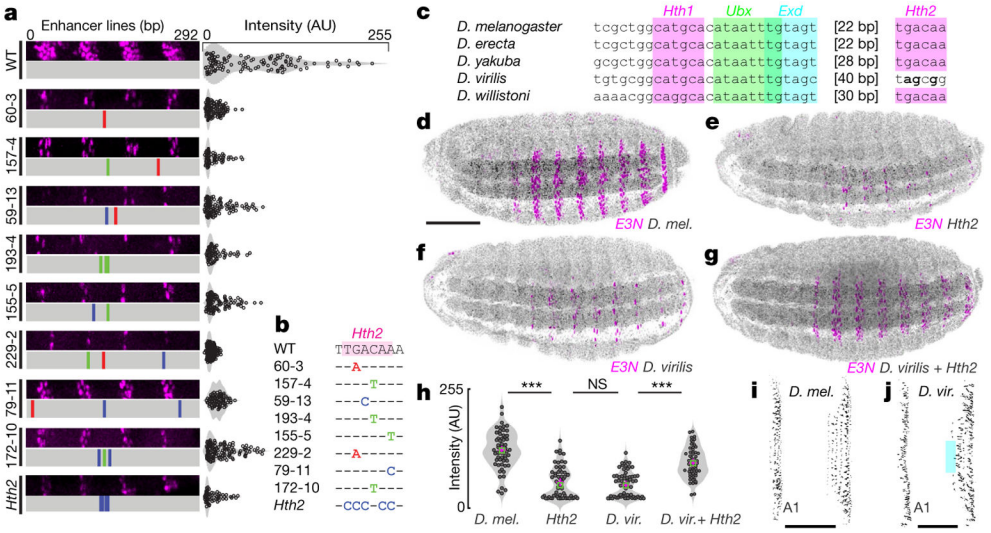


Fig. 2 |. Mutational scanning identifies a Hth binding site associated with a changed evolutionary phenotype.

a, Ventral abdominal segments from lines with mutations in an Hth2 binding site (left) and individual cell-staining intensities (right). **b**, Sequences of Hth2 binding site in lines tested. **c**, Conservation of *E3N* in *Drosophila*, highlighting conserved binding sites (coloured areas). **d–g**, Embryos bearing *E3N::lacZ* constructs in a *D. melanogaster* background, stained with anti-β-galactosidase. **d**, *D. melanogaster* (*D. mel.*) wild-type *E3N::lacZ* reporter construct. **e**, *D. melanogaster E3N::lacZ* reporter construct with a mutated *Hth2* site. **f**, *Drosophila virilis E3N::lacZ* reporter construct. **g**, *D. virilis E3N::lacZ* reporter construct with rescued *Hth2* site (5′-TGACAA). **h**, Single-cell quantification of staining intensities in embryos bearing indicated constructs. *D. vir., D. virilis*. Magenta crosses, mean; green squares, median. Two-tailed *t*-test; ***$P < 0.01$, NS not significant ($n = 50$, 10 embryos). Left to right: $P < 0.01$, $P = 0.34$, $P < 0.01$. **i**, **j**, Cuticle preparations for *D. melanogaster* (**k**) and *D. virilis* (**l**) showing ventral regions. Scale bars, 50 μm. Teal rectangle highlights trichomes that are absent in *D. virilis*. **f**, Scale bar, 100 μm; embryos in **d–g** matched for scale.

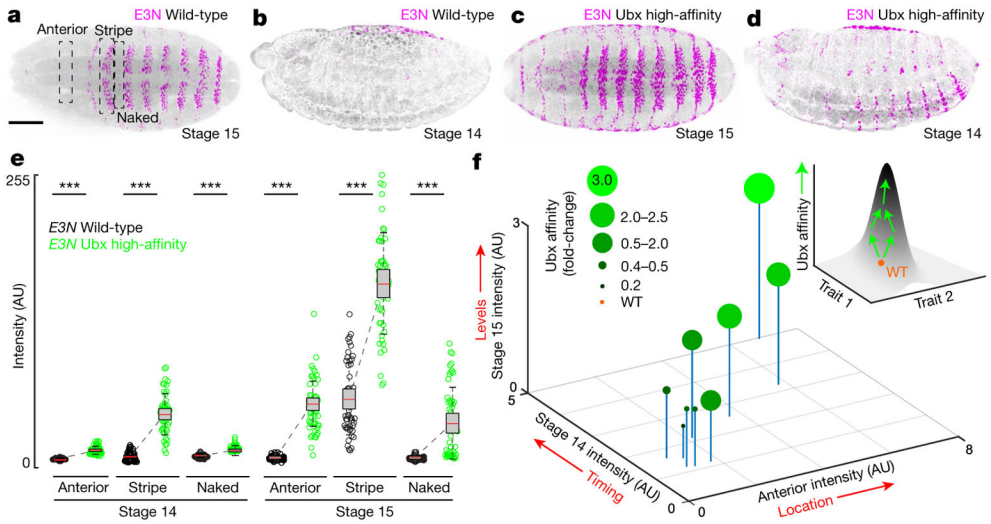


Fig. 3 |. Ubx mutations often simultaneously change levels, timing, and locations of *E3N* expression.

a–d, Embryos carrying *E3N::lacZ* reporter constructs stained with anti-β-galactosidase. **a**, **b**, Wild-type *E3N::lacZ* at stages 15 and 14. **c**, **d**, *E3N::lacZ* construct with increased Ubx affinity at stages 15 and 14. **e**, Measurements of stripe (intra), naked (inter-stripe), and anterior (ectopic) nuclear expression by wild-type *E3N::lacZ* and Ubx high-affinity *E3N::lacZ*. Centre line, mean; upper and lower limits, s.d.; whiskers, 95% confidence intervals (CIs). Two-tailed *t*-test, ***$P < 0.01$ ($n = 50$, 10 embryos). **f**, Change in expression between mutant lines and wild-type, comparing total Ubx affinity (shades of green and circle size), anterior intensity, stage 14 intensity, and stage 15 intensity. Inset, model for Ubx affinity and linked traits. **a**, Scale bar, 100 μm; embryos in **a–d** are matched for scale.

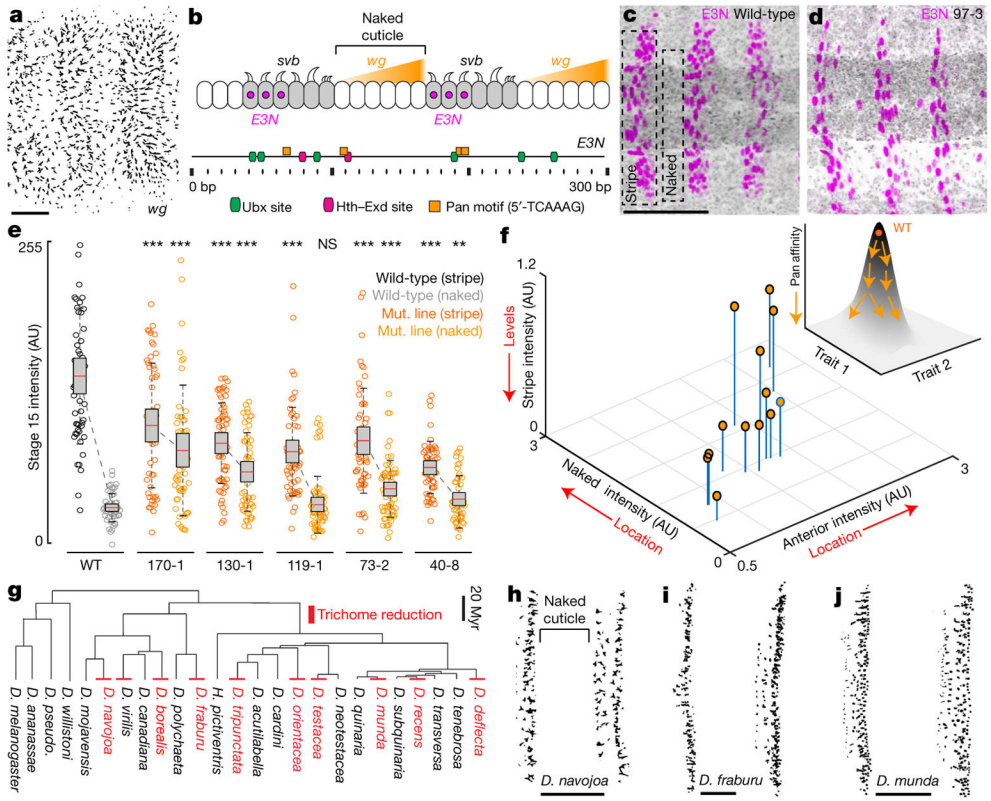


Fig. 4 | *E3N* enhancer architecture may constrain evolution.

a, Ventral views of *wg*-deficient (wg^{CX4}) cuticle preparations. **b**, Top, model of *wg* expression in naked cuticle regions that represses *svb* expression through Pan (Tcf) activity. Bottom, schematic of the *E3N* enhancer denoting binding motifs for Ubx, Hth, and Pan. **c**, **d**, Embryos carrying *E3N::lacZ* reporter constructs stained with anti-β-galactosidase; **c**, wild-type; **d**, *97-3::lacZ* line (Extended Data Fig. 4). **e**, Nuclear expression in naked and stripe regions for a subset of Pan mutation lines. Centre line, mean; upper and lower limits, s.d.; whiskers, 95% CIs. Two-tailed *t*-test; ***$P < 0.01$, NS not significant ($n = 50$, 10 embryos) (Extended Data Fig. 7). **f**, Changes in intensities compared to wild-type in inter- and intra-stripe regions with anterior expression. Inset, model of the relationship between Pan affinities and effects on multiple traits. **g**, Phylogenetic tree of *Drosophila* species (including *Hirtodrosophila pictiventris*) examined here noting losses (red) of trichomes in the naked region (there were no examples of gains). **h–j**, Cuticle preparations from *D. navojoa* (**h**), *D. fraburu* (**i**), and *D. munda* (**j**) (Extended Data Fig. 8). Scale bars, 100 μm (**c**, **d**); 25 μm (**a**, **h-j**).