

 Open access • Posted Content • DOI:10.1101/2021.03.01.429915

Butterfly eyespots evolved via co-option of the antennal gene-regulatory network

— [Source link](#) 

Murugesan Sn, Heidi Connahs, Yuji Matsuoka, Gupta ...+8 more authors

Institutions: National University of Singapore, Miami University, Michigan Technological University, Yale-NUS College

Published on: 01 Mar 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Antennapedia

Related papers:

- [Mutants highlight the modular control of butterfly eyespot patterns.](#)
- [Spalt expression and the development of melanic color patterns in pierid butterflies](#)
- [Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait](#)
- [Comparative insights into questions of lepidopteran wing pattern homology.](#)
- [Nymphalid eyespots are co-opted to novel wing locations following a similar pattern in independent lineages](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/butterfly-eyesspots-evolved-via-co-option-of-the-antennal-4h64txbmb3>

1 **Title: Butterfly eyespots evolved via co-option of the antennal gene-regulatory network**

2

3 **Authors:** Suriya Narayanan Murugesan^{1, *, †}, Heidi Connahs^{1, *, †}, Yuji Matsuoka^{1, †}, Mainak
4 das Gupta¹, Manizah Huq¹, V Gowri¹, Sarah Monroe¹, Kevin D. Deem², Thomas Werner³,
5 Yoshinori Tomoyasu², Antónia Monteiro^{1, 4, *}

6 **Affiliations:**

7 1- Department of Biological Sciences, National University of Singapore

8 2- Department of Biology, Miami University, USA

9 3- Department of Biological Sciences, Michigan Technological University, USA

10 4- Science Division, Yale-NUS College, Singapore

11 †- Equal contributions

12 *- Authors of correspondence

13 **Email:** suriya_nm@u.nus.edu, dbshc@nus.edu.sg, antonia.monteiro@nus.edu.sg

14

15 **Abstract:** Butterfly eyespots are beautiful novel traits with an unknown developmental origin.
16 Here we show that eyespots likely originated via co-option of the antennal gene-regulatory
17 network (GRN) to novel locations on the wing. Using comparative transcriptome analysis, we
18 show that eyespots cluster with antennae relative to multiple other tissues. Furthermore, three
19 genes essential for eyespot development (*Distal-less (Dll)*, *spalt (sal)*, and *Antennapedia*
20 (*Antp*)) share similar regulatory connections as those observed in the antennal GRN. CRISPR
21 knockout of *cis*-regulatory elements (CREs) for *Dll* and *sal* led to the loss of eyespots and
22 antennae, and also legs and wings, demonstrating that these CREs are highly pleiotropic. We
23 conclude that eyespots likely re-used the ancient antennal GRN, a network previously
24 implicated also in the development of legs and wings.

25

26 **Main text:**

27 Although the hypothesis of GRN co-option is a plausible model to explain the origin of
28 morphological novelties (1), there has been limited empirical evidence to show that this
29 mechanism led to the origin of any novel trait. Several hypotheses have been proposed for the
30 origin of butterfly eyespots, a novel morphological trait. These include GRN co-option from
31 the leg (2), embryo segmentation (3), wing margin (4) and wound healing (5). These
32 hypotheses for eyespot GRN origins all rely on similarities of expression of just a few candidate
33 genes observed in eyespots and in the proposed ancestral gene network. To test whether co-
34 option of any of these networks underlies eyespot origins, we focused on the nymphalid
35 butterfly *Bicyclus anynana*, which has served as a model for studying eyespot development (6).
36 Using RNA-sequencing (RNA-seq), we examined and compared the larger collection of genes
37 expressed in a forewing eyespot of *B. anynana* with those expressed in these proposed
38 candidate ancestral traits. Additionally, we examined a few other traits, including larval head
39 horns and prolegs, and also pupal eyes and antennae (Fig. 1A).

40

41 **The transcriptome profile of eyespots and antennae cluster together**

42 We first examined which of the sampled tissues shared the most similar gene expression profile
43 to eyespot tissue, as these should cluster closer together (7). Pairwise differential expression
44 (DE) analysis using DESeq2 (8) identified 10,281 DE genes ($\log_{2}FC \geq |2|$ and $padj \leq 0.001$)
45 among all tissues sampled. Hierarchical clustering of tissues, using DE genes, resulted in
46 eyespots clustering with antennae (Fig. 1B), but tissues were also clustering according to
47 developmental stage (Fig. 1B, 1C). To circumvent the strong developmental stage signal, we
48 reanalyzed DE genes solely from 3-h-old pupae, when the eyespot tissue was dissected. We
49 found 3,839 DE genes between the tissues, with eyespots clustering with antennae, and both

50 forming an outgroup to the remaining tissues with a high approximately unbiased (AU) P -value
51 (9) (Fig. 1D).

52 To more narrowly identify the subset of genes associated with eyespot development and to
53 examine similarities in their expression profile with our candidate tissues, we next compared
54 the transcriptome of dissected eyespot tissue with adjoining control tissue in the same wing
55 sector (Fig. 1A), as done by a previous study (10). This previous study identified 183 genes
56 differentially expressed in eyespots relative to sectors of the wing without eyespots. Our new
57 DE analysis between eyespot and control wing tissues identified 652 eyespot-specific DE genes
58 with 370 being up-regulated, which included *sal*, and 282 down-regulated in eyespots (Fig. S1,
59 S2, Spreadsheet S1). We mapped the published 183 eyespot DE genes, which included *Dll* and
60 *Antp*, to the current assembled transcriptome. After removing multi-mapped genes, we retained
61 144 genes from the published study for further analysis (Spreadsheet S1). When hierarchical
62 clustering was performed, using either the newly identified 652 genes, the 144 genes previously
63 identified, or both datasets combined, we found that the eyespot transcriptome always clustered
64 with antennae with strong support AU P -value for the clade. This clustering persisted with just
65 the 370 up-regulated genes (Fig. S3A, E, F).

66

67 Given the importance of transcription factors (TFs) in development and in establishing GRNs,
68 we used 336 genes annotated as having “DNA-binding transcription factor activity
69 (GO:0003700)” and “transcription factor binding (GO:0008134)” in a separate analysis, which
70 showed eyespots again clustering with antennae (Fig. S3B). Annotation and gene enrichment
71 for the DE genes (3,839) between the 3-h-pupal stage tissues showed a strong enrichment in
72 animal organ morphogenesis (GO:0009887) and anatomical structure formation (GO:2000026)
73 (Fig. S4). Performing the clustering analysis using genes from these two groups (GO:0009887
74 and GO:2000026), in two separate analyses, reproduced the same results as the full gene set,

75 indicating that these morphogenesis genes show similar expression profiles in both eyespots
76 and antennae (Fig. S3C, 3D).

77

78 These analyses showed that eyespots and antennae form an outgroup to the other tissues,
79 including legs, which are considered serial homologs to antennae. However, eyespots express
80 a key selector gene, *Antp* which is known to give legs their unique identity and differentiate
81 them from antennae. Antp protein is known to positively regulate *Dll* and repress *sal* in the leg
82 disc of *Drosophila* (11, 12), whereas in the antennae, in the absence of Antp, Dll activates *sal*
83 (13). Comparative data across 23 butterfly species suggested that eyespots originated without
84 Antp protein expression, and that *Antp* was recruited later to the eyespot GRN in at least two
85 separate lineages, including in the ancestors of *B. anynana* (14). We therefore reasoned that if
86 eyespots are co-opted antennae, rather than co-opted legs, the regulatory interactions between
87 *Dll*, *Sal*, and *Antp* in eyespots should resemble those in insect antennae but not those in legs,
88 and that the regulatory interactions between *Antp* and the other two genes should be novel and
89 not homologous.

90

91 **Function of *sal* and regulatory interactions between *Dll*, *sal*, and *Antp* in eyespots**

92 Before establishing regulatory interactions between the three genes, we first obtained missing
93 functional data for one of these genes, *sal*, lacking for *B. anynana*. Mutations for *Dll* and *Antp*
94 were previously shown to remove eyespots, pointing to these genes as necessary for eyespot
95 development (6, 15). We disrupted the function of *sal*, using CRISPR with a single guide RNA
96 (sgRNA) targeting exon 2 (Fig. S5). *sal* crispants (mosaic mutants) showed a range of
97 phenotypes, from missing eyespots (Fig. 2B and 2D, Fig. S6) to altered chevron patterns on
98 the wing margin and the central symmetry system bands running the length of each wing (Fig.
99 2B), all mapping to patterns of *sal* expression in larval and pupal wings (Fig. 2H, 2K) (5, 16).

100 Our data confirmed phenotypes previously shown in *J. coenia* (17). However, two novel and
101 striking phenotypes were the splitting of eyespot centers into two smaller centers (Fig. 2D, Fig.
102 S6) and the partial loss of black scales in the eyespot and their replacement with orange scales
103 (Fig. 2D), resembling the “goldeneye” phenotype (18). Taken together, these results confirm
104 that *sal* is necessary for the development of eyespots and also for the development of black
105 scales.

106 To test the regulatory hierarchy between these three eyespot-essential genes, we knocked out
107 each gene in turn, using CRISPR-Cas9, and reared the mosaic individuals until the late 5th
108 instar for larval wing dissections. We performed immunohistochemistry on these wings with
109 antibodies against the protein of the targeted gene and against the other two proteins. We first
110 examined the interaction of *Dll* with *Antp*. In wild-type (wt) wings, Dll protein is expressed
111 along the wing margin and in finger-like patterns, spreading from the wing margin to the future
112 eyespot centers (Fig. 2E), whereas Antp protein is initially expressed in the center of four
113 putative eyespots (from M1 to Cu1) (19). In a *Dll* crispant forewing, Antp protein expression
114 was affected in *Dll* null cells (Fig. 2F, Fig. S7), whereas Dll protein expression was not affected
115 in *Antp* null cells in an *Antp* crispant (Fig. 2G, Fig. S8). These results suggest that *Dll* is
116 upstream of *Antp* in eyespot development. We next examined the interaction of *Dll* with *sal*.
117 In wt wings, Sal protein is broadly expressed along several wing sectors, connected to its role
118 in vein patterning (16), and also expressed in nine potential eyespot centers (Fig. 2H and 2K).
119 In *Dll* crispants, Sal expression was lost in *Dll* null clones in the eyespot centers (Fig. 2I, Fig.
120 S9), but Dll protein expression was not affected in *sal* null clones in *sal* crispants (Fig. 2J, Fig.
121 S10). These results suggest that *Dll* is also upstream of *sal* in eyespots. Finally, we examined
122 the interaction between *Antp* and *sal*. In *Antp* crispants, Sal protein expression is missing from
123 *Antp* null cells (Fig. 2L, Fig. S11). Furthermore, Antp protein expression is missing from *sal*
124 null cells in *sal* crispants (Fig. 2M, Fig. S12). Taken together, *Dll* is up-regulating both *Antp*

125 and *sal*, and *Antp* and *sal* are up-regulating each other's expression in forewing eyespots (Fig.
126 2N).

127

128 **Regulatory connections between *Dll* and *sal* in eyespot development are similar to those** 129 **in the antennae of flies**

130 We next examined whether the appendage expression and regulatory connections between
131 these three genes of *B. anynana* matched those known in fly leg and antennal development. In
132 flies, Dll protein is expressed in both appendages (20), whereas Sal is only expressed in
133 antennae and Antp only in legs of flies (13). In *B. anynana* we observed similar expression
134 profiles in antennae and thoracic legs of pupae (Fig. S13-S14). Dll is necessary for *sal*
135 expression in antennae of flies (13), as also observed in *B. anynana* eyespots (Fig. 2I). Antp,
136 however, negatively regulates *sal* expression in fly legs (12), which differs from the regulation
137 observed in eyespots, where *Antp* and *sal* up-regulate each other (Fig. 2N). The genetic
138 interaction of *Antp* and *Dll* during leg development in *Drosophila* is stage-dependent. At the
139 stage when leg primordia are formed, Antp positively regulates *Dll* expression in the thoracic
140 leg bud (11), but when leg segments are being formed, Dll negatively regulates *Antp* in the
141 distal leg elements (21). These regulatory interactions between *Dll* and *Antp* in leg development
142 are distinct from the regulatory interaction observed in eyespots (Fig. 2N). Taken together,
143 these data suggest that the regulatory interactions between *Dll* and *sal* in eyespots are likely
144 homologous to those in the insect antenna GRN. *Antp* established a novel regulatory interaction
145 to these two genes in eyespots, distinct from those found in the leg GRN of *Drosophila*. This
146 supports the later and independent addition of *Antp* to the eyespot GRN in two separate lineages
147 of butterflies, as proposed by Oliver et al. (2012) (14).

148

149 **Two pleiotropic CREs reveal a shared network between eyespots, antennae, and other**
150 **traits**

151 Evidence of GRN co-option is bolstered by the identification of shared *cis*-regulatory elements
152 (CREs) driving the expression of genes common to both the ancestral and the novel trait
153 (eyespots). To identify putative CREs specific to wing tissue with eyespots, we used
154 Formaldehyde-Assisted Isolation of Regulatory Elements using sequencing (FAIRE-seq) to
155 identify the open chromatin profile around *Dll* in forewing and hindwing pupal tissues of *B.*
156 *anymana*. We produced separate libraries from the proximal and distal regions of the wing.
157 Mapping of FAIRE-seq reads from each wing region to a previously published *Dll* BAC
158 (scaffold length of 230 kb) revealed 18 regions of open chromatin across this scaffold,
159 representing candidate CREs (Fig. 3A). A BLAST search of each candidate CRE against the
160 *B. anymana* genome revealed that most of these regions contained repetitive elements.
161 However, one candidate CRE that was open in the distal forewing at scaffold position 150 kb
162 (Fig. 3B) (*Dll319* CRE), returned a unique BLAST hit to the genome. As this region did not
163 contain any repetitive elements, we used CRISPR-Cas9 to disrupt its function. We designed
164 four guide RNAs along its 319 bp length to maximize the likelihood of its disruption (Fig. 3A,
165 Fig. S15). We obtained a variety of different phenotypes that were also observed when
166 targeting exons of the *Dll* gene using CRISPR (6): several caterpillars showed a missing or
167 necrotic thoracic leg (Fig. 3C, Fig. S16), adults were missing legs and even a hindwing (Fig.
168 3D-E), adults lacked eyespots (Fig. 3F-H), adults showed truncated antennae, pigmentation
169 defects, and loss of wing scales (Fig. 3I-J and Fig. S16-S19, Table S1), all having deletions
170 within the CRE of various sizes (Fig. S16). These findings confirm that the *Dll319* CRE is
171 pleiotropic and further suggest that eyespots use the same GRN as antennae in addition to legs
172 and wings.

173

174 In order to confirm that the *Dll319* contains a functional and pleiotropic CRE, we cloned a 917
175 bp region containing this CRE into a *piggyBac*-based reporter construct (22) and evaluated its
176 CRE activity in transgenic butterflies. We observed that embryos expressed the reporter gene
177 (EGFP) in antennae, mouthparts, as well as thoracic limbs, indicating that this CRE is sufficient
178 to drive gene expression both in antennae and legs (Fig. 3K and S20). Unfortunately, the loss
179 of this line precluded us from visualizing EGFP expression in eyespots. Using this same cloned
180 region containing the *Dll319* CRE, we also observed pleiotropic CRE activity in antennae,
181 mouthparts, legs, and genitalia, when tested in a cross-species setting with *Drosophila*
182 *melanogaster* (Fig. S21), suggesting that this region contains an ancestral and pleiotropic CRE
183 present in the ancestors of flies and butterflies.

184

185 In order to investigate the extent to which other genes of the eyespot GRN share the same
186 open-chromatin profiles as genes expressed in antennae and in other tissues, we performed an
187 Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) with the same
188 tissues used for the transcriptome analysis. A differential accessibility analysis for the open-
189 chromatin regions associated with the eyespot DE genes showed that eyespots shared the
190 greatest number of open-chromatin regions with antennae, as compared to other tissues at the
191 3-h-pupal stage (Fig. 4G, 4H). The ATAC-seq data also showed that the *Dll319* CRE is open
192 across all different stages and tissues, irrespective of the expression of *Dll* (Fig. 4A),
193 suggesting that pleiotropic CREs may always be open throughout development. To test this,
194 we further targeted a genomic region of *sal* (*sal740*) that had open-chromatin across most
195 developmental stages using CRISPR-Cas9 (Fig. 4B). We obtained aberrations in caterpillar
196 horns, adult antennae, leg and chevron patterns, as well as missing eyespots and a missing
197 wing (Fig. 4C-4F, Fig. S22), again confirming the presence of a pleiotropic CRE for a gene
198 common to both eyespots and antennae.

199

200 To further confirm that the two CREs (*Dll319* & *sal740*) drive *Dll* and *sal* in an endogenous
201 context, we reanalyzed Hi-C data from the wandering larval stage, when Dll and Sal proteins
202 are expressed in eyespot centers (Fig. 2). Using the *Dll319* and *sal740* CREs as a bait, we
203 observed that these two sequences physically interact with the *Dll* promotor and *sal* promoter,
204 respectively (Fig. S23).

205

206 By exploring the gene expression profile and functional regulatory connections of elements of
207 the eyespot GRN, we showed that eyespots, a morphological novelty in nymphalid butterflies,
208 likely evolved via co-option of the antennal GRN, the oldest urbilaterian appendage. This
209 network, initially deployed in primitive sensory systems, has been subsequently recruited and
210 modified to produce legs (23) and perhaps even wings (24, 25). We show that the transcriptome
211 profile of eyespots more closely resembles that of antennae compared to any other tested
212 appendage or butterfly tissue. Furthermore, genes known to be critical for eyespot development
213 share the same functional connections as observed in *Drosophila* antennae. Previous studies in
214 *Drosophila* had demonstrated the same CRE driving reporter gene expression in separate traits,
215 and CRE disruptions leading to pleiotropic effects on patterns of CRE activity (26). However,
216 here we show, for the first time, that disruptions to two pleiotropic CREs result in the loss of
217 both ancestral and derived traits, which provides uncontroversial evidence for GRN co-option.

218

219 The *cis*-regulatory paradigm (27) suggests that, when a gene is expressed in a different
220 developmental context, it uses a different CRE for its activation. Here we show that this does
221 not apply to traits that emerge through gene-network co-option, as the recruited network genes
222 are most likely sharing pre-existent regulatory connections (26, 28) (Fig. 4I). The origin of
223 novelties has remained an important unanswered question in biology; and here we show that

224 novelties can arise from GRN co-option, which provides a mechanism for complex traits to
225 evolve rapidly from pre-existing traits.

226
227
228
229

230 **References**

- 231 1. W. J. Glassford, W. C. Johnson, N. R. Dall, S. J. Smith, Y. Liu, W. Boll, M. Noll, M.
232 Rebeiz, Co-option of an Ancestral Hox-Regulated Network Underlies a Recently
233 Evolved Morphological Novelty. *Dev. Cell.* **34**, 520–531 (2015).
- 234 2. S. B. Carroll, J. Gates, D. N. Keys, S. W. Paddock, G. E. Panganiban, J. E. Selegue, J.
235 a Williams, Pattern formation and eyespot determination in butterfly wings. *Science*.
236 **265**, 109–14 (1994).
- 237 3. S. V Saenko, V. French, P. M. Brakefield, P. Beldade, Conserved developmental
238 processes and the formation of evolutionary novelties: examples from butterfly wings.
239 *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 1549–55 (2008).
- 240 4. L. I. Held, Rethinking Butterfly Eyespots. *Evol. Biol.* **40**, 158–168 (2013).
- 241 5. A. Monteiro, G. Glaser, S. Stockslager, N. Glansdorp, D. Ramos, Comparative insights
242 into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* **6**, 1–13 (2006).
- 243 6. H. Connahs, S. Tlili, J. van Creijl, T. Y. J. Loo, T. Das Banerjee, T. E. Saunders, A.
244 Monteiro, Activation of butterfly eyespots by *Distal-less* is consistent with a reaction-
245 diffusion process. *Dev.* **146**, 1–12 (2019).
- 246 7. J. M. Musser, G. P. Wagner, Character trees from transcriptome data: Origin and
247 individuation of morphological characters and the so-called “species signal.” *J. Exp.*
248 *Zool. Part B Mol. Dev. Evol.* **324**, 588–604 (2015).
- 249 8. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion
250 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

- 251 9. R. Suzuki, H. Shimodaira, Pvcust: An R package for assessing the uncertainty in
252 hierarchical clustering. *Bioinformatics*. **22**, 1540–1542 (2006).
- 253 10. N. Özsu, A. Monteiro, Wound healing, calcium signaling, and other novel pathways
254 are associated with the formation of butterfly eyespots. *BMC Genomics*. **18** (2017),
255 doi:10.1186/s12864-017-4175-7.
- 256 11. J. D. Uhl, A. Zandvakili, B. Gebelein, A Hox Transcription Factor Collective Binds a
257 Highly Conserved Distal-less cis-Regulatory Module to Generate Robust
258 Transcriptional Outcomes. *PLoS Genet*. **12**, 1–26 (2016).
- 259 12. J. T. Wagner-Bernholz, O. Wilson, G. Gibson, R. Schuh, W. J. Gehring, Identification
260 of target genes of the homeotic gene Antennapedia by enhancer detection (Genes and
261 Development 5 (2467-2480)). *Genes Dev*. **6**, 328 (1992).
- 262 13. P. D. Si Dong, J. Chu, G. Panganiban, Coexpression of the homeobox genes Distal-
263 less and homothorax determines Drosophila antennal identity. *Development*. **127**, 209–
264 216 (2000).
- 265 14. J. C. Oliver, X. L. Tong, L. F. Gall, W. H. Piel, A. Monteiro, A Single Origin for
266 Nymphalid Butterfly Eyespots Followed by Widespread Loss of Associated Gene
267 Expression. *PLoS Genet*. **8** (2012), doi:10.1371/journal.pgen.1002893.
- 268 15. Y. Matsuoka, A. Monteiro, Hox genes are essential for the development of novel serial
269 homologous eyespots on the wings of *Bicyclus anynana* butterflies. *Gentics*, iyaa005
270 (2020).
- 271 16. T. Das Banerjee, A. Monteiro, Molecular mechanisms underlying simplification of
272 venation patterns in holometabolous insects. *Dev*. **dev.196394** (2020),
273 doi:10.1242/dev.196394.
- 274 17. L. Zhang, R. D. Reed, Genome editing in butterflies reveals that spalt promotes and
275 Distal-less represses eyespot colour patterns. *Nat. Commun*. **7** (2016),

- 276 doi:10.1038/ncomms11769.
- 277 18. C. R. Brunetti, J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield, S. B. Carroll,
278 The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**,
279 1578–1585 (2001).
- 280 19. S. V. Saenko, M. S. P. Marialva, P. Beldade, Involvement of the conserved Hox gene
281 Antennapedia in the development and evolution of a novel trait. *Evodevo.* **2**, 9 (2011).
- 282 20. G. Panganiban, Distal-less Function During Drosophila Appendage and Sense Organ
283 Development. *Dev. Dyn.* **562**, 554–562 (2000).
- 284 21. B. S. Emerald, S. M. Cohen, Spatial and temporal regulation of the homeotic selector
285 gene Antennapedia is required for the establishment of leg identity in Drosophila. *Dev.*
286 *Biol.* **267**, 462–472 (2004).
- 287 22. Y. T. Lai, K. D. Deem, F. Borràs-Castells, N. Sambrani, H. Rudolf, K. Suryamohan, E.
288 El-Sherif, M. S. Halfon, D. J. McKay, Y. Tomoyasu, Enhancer identification and
289 activity evaluation in the red flour beetle, *Tribolium castaneum*. *Dev.* **145** (2018),
290 doi:10.1242/dev.160663.
- 291 23. G. Panganiban, J. L. R. Rubenstein, Developmental functions of the Distal-less/Dlx
292 homeobox genes. *Development.* **129**, 4371–86 (2002).
- 293 24. H. S. Bruce, N. H. Patel, *Knockout of crustacean leg patterning genes suggests that*
294 *insect wings and body walls evolved from ancient leg segments* (2020), vol. 4.
- 295 25. C. M. Clark-Hachtel, Y. Tomoyasu, Two sets of candidate crustacean wing
296 homologues and their implication for the origin of insect wings. *Nat. Ecol. Evol.* **4**,
297 1694–1702 (2020).
- 298 26. G. Sabarís, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with Multiple Roles:
299 Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**,
300 423–433 (2019).

- 301 27. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory
302 evolution. *Light Evol.* **1**, 109–127 (2007).
- 303 28. A. Monteiro, O. Podlaha, Wings, horns, and butterfly eyespots: how do complex traits
304 evolve? *PLoS Biol.* **7**, e37 (2009).
- 305 29. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler
306 transform. *Bioinformatics.* **25**, 1754–1760 (2009).
- 307 30. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G.
308 Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools.
309 *Bioinformatics.* **25**, 2078–2079 (2009).
- 310 31. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing
311 genomic features. *Bioinformatics.* **26**, 841–842 (2010).
- 312 32. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C.
313 Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Shirley, Model-based analysis of
314 ChIP-Seq (MACS). *Genome Biol.* **9** (2008), doi:10.1186/gb-2008-9-9-r137.
- 315 33. Y. Naito, K. Hino, H. Bono, K. Ui-Tei, CRISPRdirect: Software for designing
316 CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics.* **31**, 1120–1123
317 (2015).
- 318 34. K. N. Eckermann, H. M. M. Ahmed, M. KaramiNejadRanjbar, S. Dippel, C. E.
319 Ogaugwu, P. Kitzmann, M. D. Isah, E. A. Wimmer, Hyperactive piggyBac transposase
320 improves transformation efficiency in diverse insect species. *Insect Biochem. Mol.*
321 *Biol.* **98**, 16–24 (2018).
- 322 35. C. J. Evans, J. M. Olson, K. T. Ngo, E. Kim, N. E. Lee, E. Kuoy, A. N. Patananan, D.
323 Sitz, P. T. Tran, M. T. Do, K. Yackle, A. Cespedes, V. Hartenstein, G. B. Call, U.
324 Banerjee, G-TRACE: Rapid Gal4-based cell lineage analysis in *Drosophila*. *Nat.*
325 *Methods.* **6**, 603–605 (2009).

- 326 36. B. Bushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner. *United States N. p*
327 (2014), (available at <https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice->
328 aware-aligner).
- 329 37. E. Kopylova, L. Noé, H. Touzet, SortMeRNA: Fast and accurate filtering of ribosomal
330 RNAs in metatranscriptomic data. *Bioinformatics*. **28**, 3211–3217 (2012).
- 331 38. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level
332 expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.
333 *Nat. Protoc.* **11**, 1650–1667 (2016).
- 334 39. M. S. Campbel, C. Holt, B. Moore, M. Yandell, *Genome Annotation and Curation*
335 *Using MAKER and MAKER-P* (2008), vol. 48.
- 336 40. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, D. Philip, J. Bowden, M. B.
337 Couger, D. Eccles, B. Li, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N.
338 Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N.
339 Friedman, A. Regev, *De novo transcript sequence reconstruction from RNA-Seq:*
340 *reference generation and analysis with Trinity* (2013), vol. 8.
- 341 41. C. R. Fisher, J. L. Wegrzyn, E. L. Jockusch, Co-option of wing-patterning genes
342 underlies the evolution of the treehopper helmet. *Nat. Ecol. Evol.* **4**, 250–260 (2020).
- 343 42. R Core Team, R: A language and environment for statistical computing (2020).
- 344 43. M. R. Corces, A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-
345 Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria,
346 S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A.
347 Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf, H. Y. Chang, An improved
348 ATAC-seq protocol reduces background and enables interrogation of frozen tissues.
349 *Nat. Methods*. **14**, 959–962 (2017).
- 350 44. A. P. Boyle, J. Guinney, G. E. Crawford, T. S. Furey, F-Seq: A feature density

- 351 estimator for high-throughput sequence tags. *Bioinformatics*. **24**, 2537–2538 (2008).
- 352 45. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program
353 for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).
- 354 46. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, DeepTools: A flexible
355 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).
- 356 47. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E.
357 L. Aiden, A. Mathematics, Juicer provides a one-click system for analyzing loop-
358 resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2018).
- 359 48. J. Ray, P. R. Munn, A. Vihervaara, J. J. Lewis, A. Ozer, C. G. Danko, J. T. Lis,
360 Chromatin conformation remains stable upon extensive transcriptional changes driven
361 by heat shock. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19431–19439 (2019).
- 362 49. R. W. Nowell, B. Elsworth, V. Oostra, B. J. Zwaan, C. W. Wheat, M. Saastamoinen, I.
363 J. Saccheri, A. E. van't Hof, B. R. Wasik, H. Connahs, M. L. Aslam, S. Kumar, R. J.
364 Challis, A. Monteiro, P. M. Brakefield, M. Blaxter, A high-coverage draft genome of
365 the mycalesine butterfly *Bicyclus anynana*. *Gigascience*. **6**, 1–7 (2017).
- 366 50. P. Beldade, S. V. Saenko, N. Pul, A. D. Long, A gene-based linkage map for *Bicyclus*
367 *anynana* butterflies allows for a comprehensive analysis of synteny with the
368 lepidopteran reference genome. *PLoS Genet.* **5** (2009),
369 doi:10.1371/journal.pgen.1000366.
- 370 51. J. Catchen, A. Amores, S. Bassham, *G3: Genes|Genomes|Genetics*, in
371 press, doi:10.1534/g3.120.401485.
- 372 52. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov,
373 BUSCO: Assessing genome assembly and annotation completeness with single-copy
374 orthologs. *Bioinformatics*. **31**, 3210–3212 (2015).
- 375 53. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J.

- 376 Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M.
377 Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale
378 protein function classification. *Bioinformatics*. **30**, 1236–1240 (2014).
- 379 54. J. Dainat, AGAT: Another Gff Analysis Toolkit to handle annotations in any
380 GTF/GFF format.(Version v0.4.0) (2020), doi:10.5281/ZENODO.4205393.
- 381 55. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using
382 DIAMOND. *Nat. Methods*. **12**, 59–60 (2014).
- 383 56. S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M.
384 Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and
385 data mining with the Blast2GO suite. *Nucleic Acids Res*. **36**, 3420–3435 (2008).

386

387

388 **Acknowledgement:**

389 We would like to thank NUS HPC for allowing access to perform the bioinformatic analysis.

390 Tirtha Das Banerjee for his comments on manuscript.

391

392 **Funding:**

393 This work was supported by the Ministry of Education, Singapore award MOE2015-T2-2-159

394 and the National Research Foundation, Singapore NRF Investigatorship award NRF-NRFI05-

395 2019-0006 and NRF-CRP20-2017-0001 to AM. SNM was supported by a Yale-NUS

396 Scholarship. YT is thankful to the National Science Foundation (NSF) (grant IOS1557936) and

397 the U.S. Department of Agriculture (USDA) (grant 2018- 08230).

398

399 **Authors contribution:**

400 SNM, HC, YM, and AM designed and conceived the project. SNM, HC, YM, MDG, MH,
401 GV, SM, KDD performed the experiments. TW and YT critically revised and, SNM, HC,
402 YM, AM analyzed the data and wrote the manuscript.

403

404 **Competing interests:**

405 Authors declare that they have no competing interests

406

407 **Data Availability:**

408 All raw Illumina reads of RNA-seq, ATAC-seq and Hi-C are available under NCBI

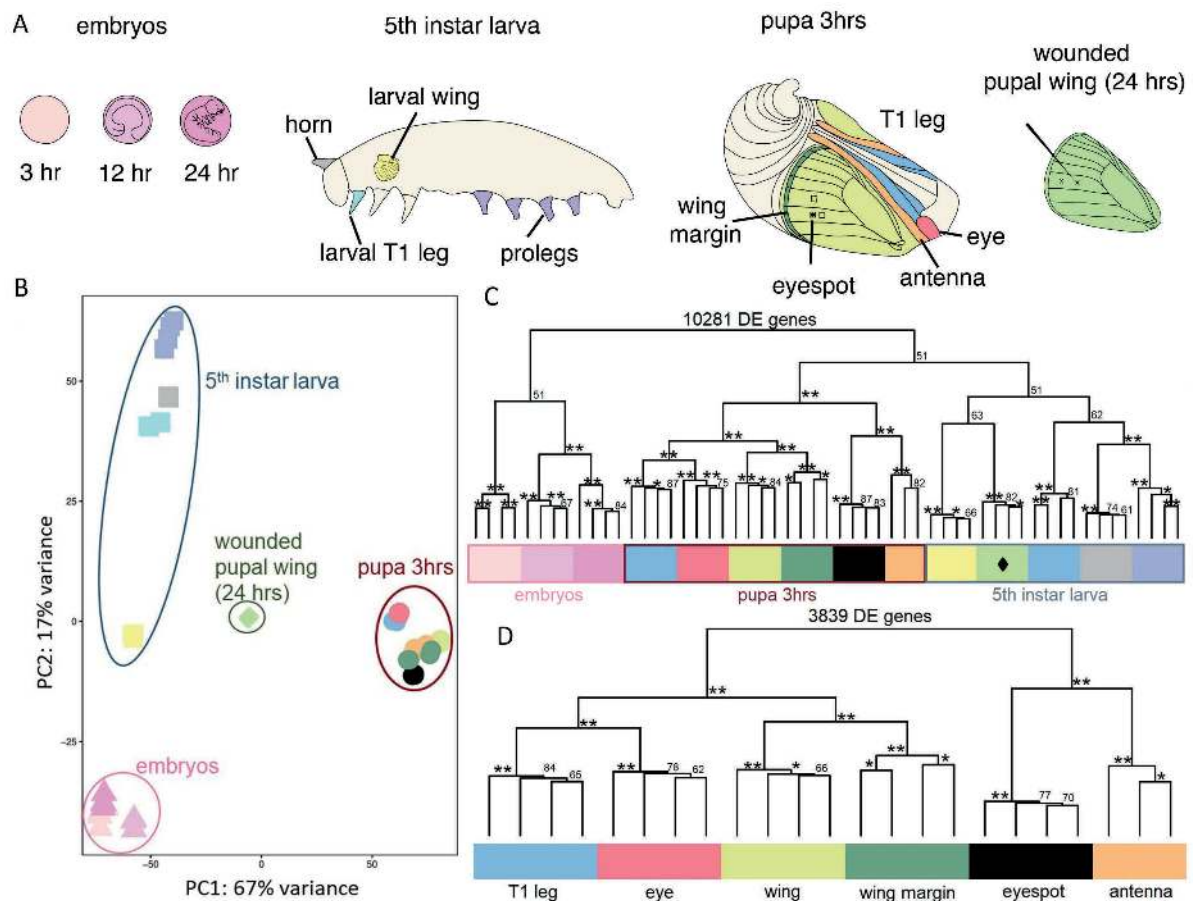
409 Bioproject (PRJNA685019)

410

411

412

413



414

415

416 **Fig. 1: Tissues used for RNA-seq analysis and character tree constructed using the**

417 **differentially expressed (DE) genes.** (A). We used 16 tissue groups from three separate

418 developmental stages of *B anynana* for RNA extractions. Embryos at 3 hrs, 12 hrs and 24 hrs

419 after egg laying. Larval forewings, T1-legs, horns and prolegs. Pupal antenna, T1-leg,

420 forewing, eye, wing margin, eyespot, and two eyespot control tissues all dissected at 3hrs after

421 pupation, and a wounded wing dissected at 24hrs after pupation. (B). PCA using 10281 DE

422 genes obtained from pairwise comparisons between different tissues. Tissues are clustering

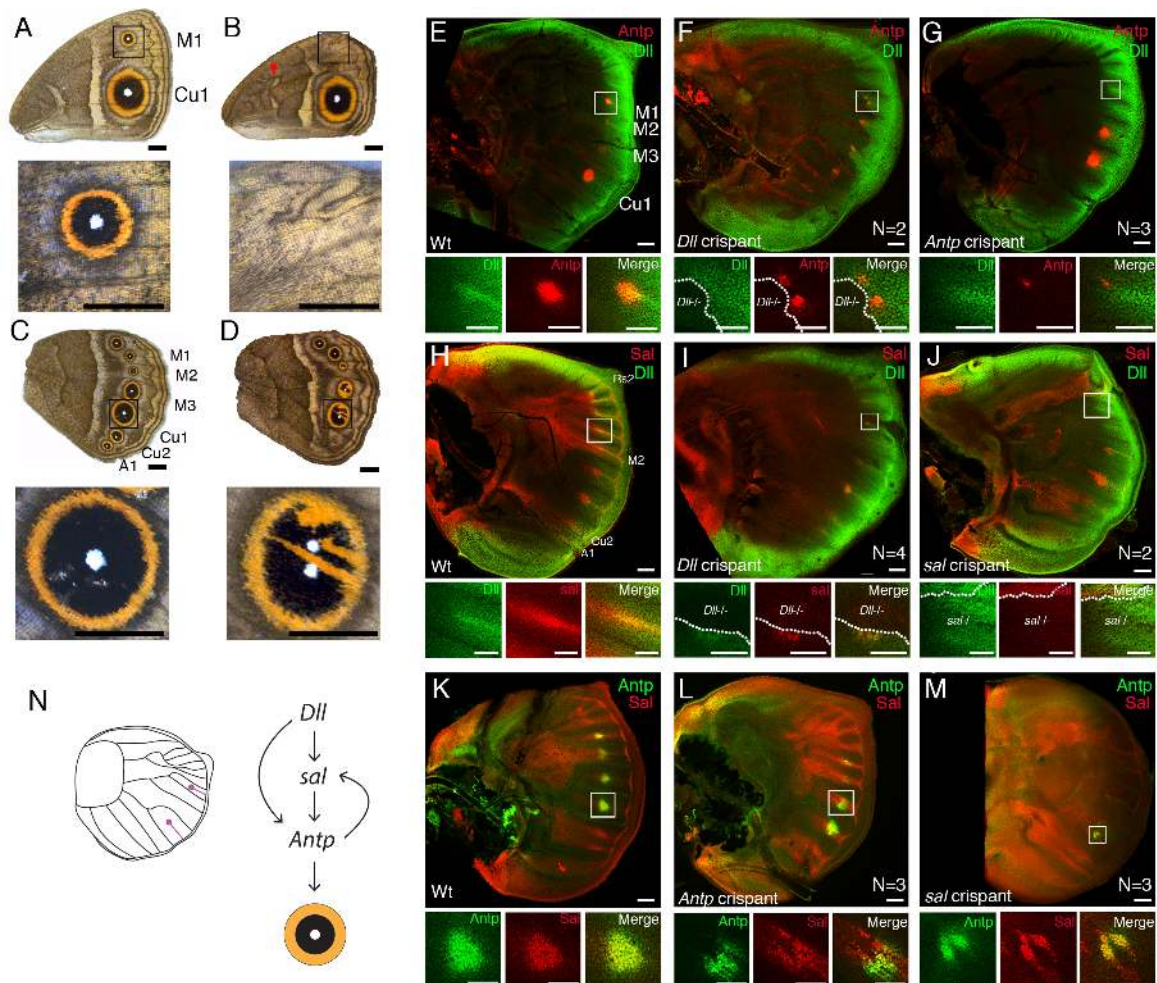
423 according to their developmental stages. (C). Character tree constructed using 10281 DE genes

424 showing eyespot tissue clustering with antenna tissue first, and next with tissues from same

425 developmental stage, except for a 24 hrs wounded wing (◆), which clustered with larval wing

426 tissue (D). Character tree constructed using 3839 DE genes from 3hr pupal stage showing

427 eyespot tissue clusters with antenna tissue, which together form an outgroup to the rest of the
428 samples. ** - 100 unbiased (AU) p-value; * - 90-99 unbiased (AU) p-value; ◆ - wounded pupal
429 wing (24 hrs)
430
431
432
433
434
435
436
437
438
439
440



441

442

443

444

445 **Fig. 2. Function of *sal* and regulatory interactions between *Dll*, *sal*, and *Antp* inferred**

446 **with CRISPR and immunohistochemistry (A). Wt female forewing. (B) *sal* crisant female**

447 **forewing (C) Wt female hindwing. (D) *sal* crisant female hindwing (E) Expression pattern of**

448 **Dll and Antp proteins in Wt forewing. (F) Expression pattern of Dll and Antp proteins in *Dll***

449 **crisant forewing. (G) Expression pattern of Dll and Antp proteins in an *Antp* crisant forewing.**

450 **(H) Expression pattern of Dll and Sal proteins in Wt forewing. (I) Expression pattern of Dll and**

451 **Sal proteins in *Dll* crisant forewing. (J) Expression pattern of Dll and Sal proteins in *sal***

452 **crisant forewing. (K) Expression pattern of Sal and Antp proteins in Wt forewing. (L)**

453 Expression pattern of Sal and Antp proteins in *Antp* crispant forewing. (M) Expression pattern
454 of Sal and Antp proteins in *sal* crispant forewing. White square regions were highly magnified.
455 (N) Schematic diagram of genetic interaction among *Dll*, *sal*, and *Antp* in the eyespot region of
456 a developing forewing. Scale bars in A-D: 5 mm for whole wings and wing details. Scale bars
457 in E-M: 100 μ m in low and 50 μ m, in high magnification

458

459

460

461

462

463

464

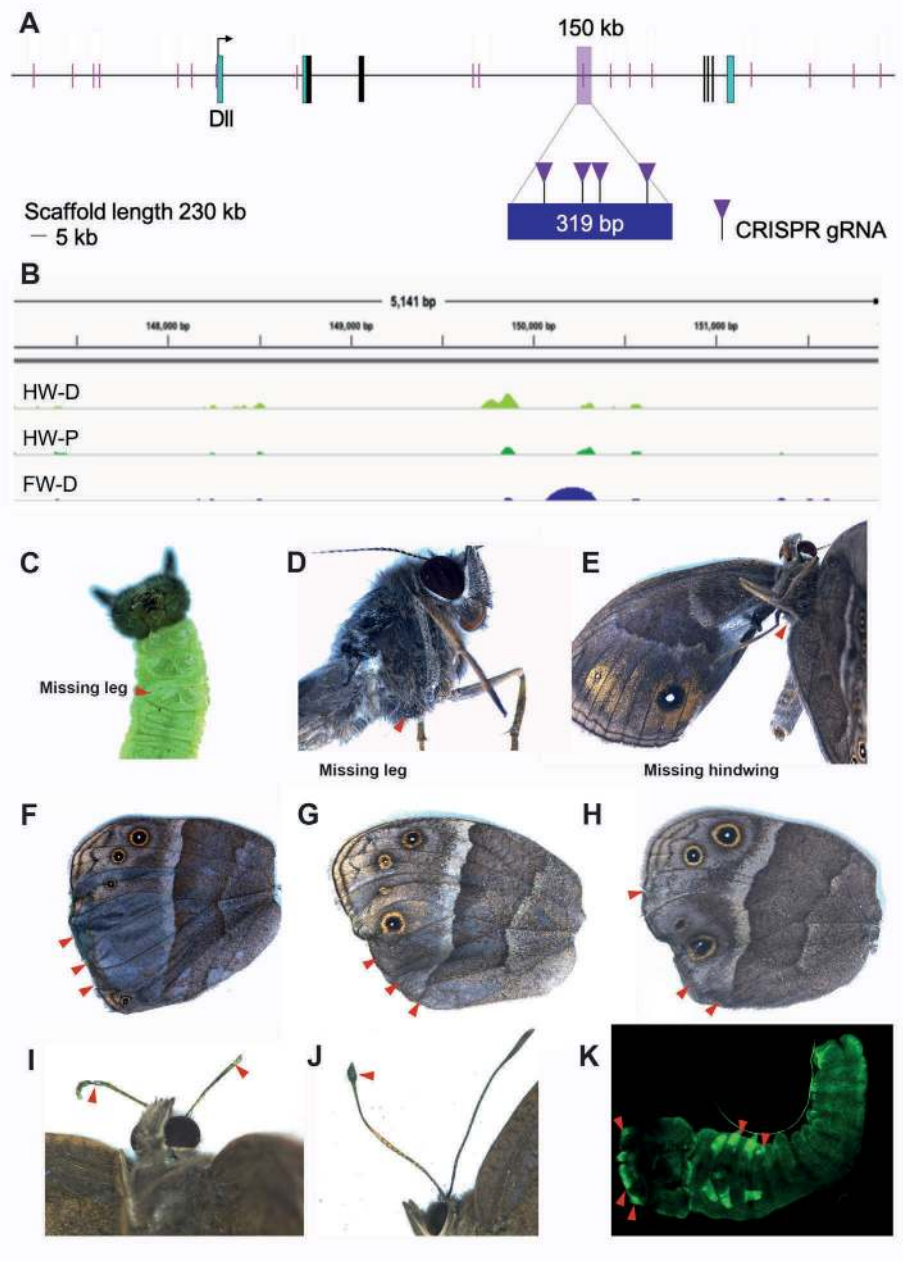
465

466

467

468

469



470

471

472

473 **Fig. 3. Multiple traits are affected by disruptions of a single *Distal-less* CRE.** (A) *B.*

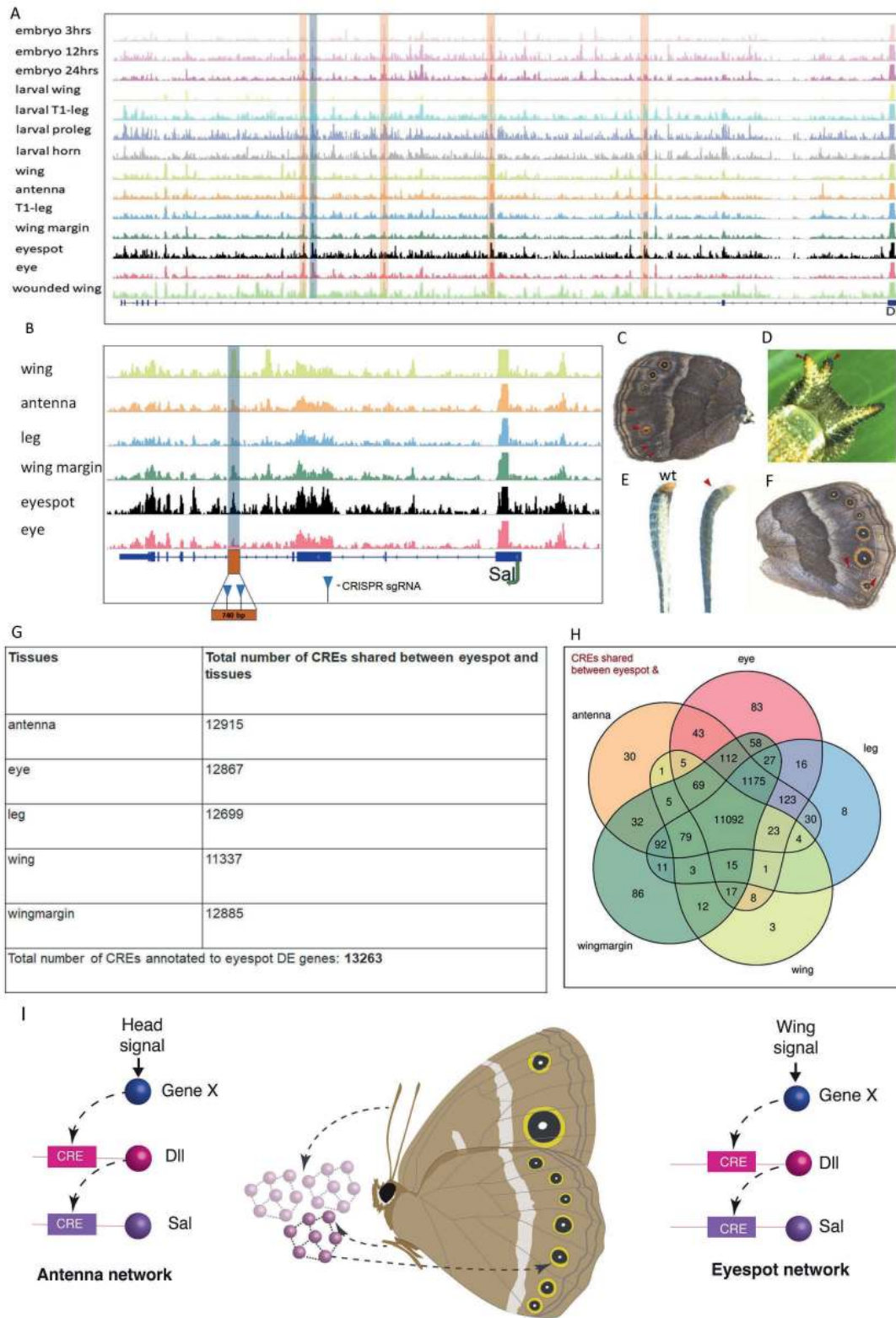
474 *anynana Distal-less* BAC visualized using IGV showing all 18 FAIRE-seq open chromatin

475 regions at 24 hours post-pupation (short red lines). First exon (UTR in blue) shows open

476 chromatin region (highlighted by a short red line) at position 54 kb at the transcriptional start

477 site of *Distal-less*. The CRE at position 150 kb (*DII319*; highlighted with a pink bar) is open in

478 *B. anynana* forewing and was targeted with CRISPR. Four RNA guides were used
479 simultaneously to target this region. (B) FAIRE-seq results showing an open region of
480 chromatin in the distal forewing (FW-D) at position 150 kb on the *Distal-less* BAC (blue peak).
481 (C-E) crispant phenotypes from the same individual: with a missing thoracic leg as a caterpillar,
482 and same missing thoracic leg and also missing hindwing as an adult. (F-H) Crispant wing
483 phenotypes showing loss of eyespots and pigmentation defects. (I-J) Crispants showing
484 antennal defects (K) Transgenic embryo showing EGFP expression driven by the *Dll319* CRE
485 in mouthparts, antennae, legs, and pleuropodia (red arrows from left to right).
486
487



488

489

490

491 **Fig. 4. Visualization of open chromatin around *Dll* and *sal* genomic regions for different**
492 **tissues and identification of a *sal* pleiotropic CRE.** (A). ATAC-seq reads around the *Dll*
493 genomic region with highlights in the open regions shared across different tissues (orange) and
494 the targeted *Dll319* (blue). (B). ATAC peak regions from 3hr pupal tissues around the *sal*
495 genomic region with the CRE (*sal740*) targeted highlighted in blue. (C-F). *sal740* crispant
496 phenotypes: Missing and reduced eyespots (C), split horn (D), thinner and discolored antenna
497 compared to wild type (E), lost chevrons in the wing margin and ectopic vein in the Cu2 sector
498 (F). (G). Table with the total number of open peaks associated with eyespot DE genes and
499 number of peaks shared between eyespots and different tissues. (H). Venn diagram showing
500 the number of open chromatin regions shared between different tissue groups. (I). Schematic
501 illustrating the hypothesis that eyespots evolved via co-option of an antennal GRN with genes
502 (*Dll* and *sal*) in the GRN reusing the same CREs in both antennae and eyespot development.

503

504

505

506

507

508

509

510

511

512

513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554

Supplementary Materials for

Butterfly eyespots evolved via co-option of the antennal gene-regulatory network

Suriya Narayanan Murugesan^{1,*, †}, Heidi Connahs^{1,*, †}, Yuji Matsuoka^{1, †}, Mainak das Gupta¹,
Manizah Huq¹, V Gowri¹, Sarah Monroe¹, Kevin D. Deem², Thomas Werner³, Yoshinori
Tomoyasu², Antónia Monteiro^{1,4,*}

Correspondence to : suriya_nm@u.nus.edu, dbshc@nus.edu.sg, antonia.monteiro@nus.edu.sg

555 **Materials and Methods**

556

557 **Butterfly husbandry**

558 *Bicyclus anynana* were maintained in lab populations and reared at 27°C and 60% humidity
559 inside a climate room with 12:12 h light:dark cycle. All larvae were supplied with young corn
560 leaves to complete their development until pupation. Following pupation, the pupae were
561 collected and placed in a separate cage until they emerged. The butterflies were fed every other
562 day with banana on moist cotton in Petri dishes.

563

564 **Wing library preparation and FAIRE-seq analysis**

565 Wings were dissected from *B. anynana* at ~22-26 hours post-pupation. For control input
566 libraries (non-enriched), 2 whole forewings and 2 whole hindwings were pooled. Three
567 FAIRE-enriched libraries were prepared in total, including a forewing distal library (the pupal
568 wing was cut in half and the distal region was used for the library) and 2 hindwing libraries,
569 using both the proximal and distal regions of the wing. All FAIRE-enriched libraries were
570 prepared from 7-8 pooled wing tissues. Libraries were prepared by Genotypic Technology
571 (India) as 75bp pair-end reads and sequenced, using Illumina NextSeq. Raw reads were quality-
572 checked and reads with phred scores >30 were retained for downstream analyses. Following
573 the removal of adapters and low-quality bases, the reads were aligned to a *B. anynana* BAC
574 sequence containing *Dll*, with BWA (0.7.13)(29), using the following parameters: -k INT, -w
575 INT, -A INT, -B INT, -O INT, -E INT, -L INT, -U INT. The resulting SAM files were
576 converted to BAM files, using SAMtools-0.1.7a(30). The BAM files were converted to sorted
577 BAM, followed by removal of PCR duplicates. The final BAM files were converted to
578 BEDgraph files, using BEDtools-2.14.3(31). Peaks were called with MACS2 software(32),
579 using the aligned enriched and input (control) files with the q-value (minimum FDR) cutoff to
580 call significant regions. The command `bdgcmp` script was used on the enriched and input
581 BEDgraph files to generate fold-enrichment and log likelihood scores. This command also
582 removed noise from the enriched sample relative to the control. The BEDgraph files were
583 converted to BigWig files for visualization in Integrative genomic viewer (IGV).

584

585 **Identifying cis-regulatory elements (CREs) for CRISPR-Cas9 experiments**

586 The FAIRE-seq data were visualized using IGV. All 18 candidate CREs identified around the
587 *Dll* locus were blasted against the *B. anynana* genome in LepBase to verify whether they were
588 unique in the genome. Most of the candidate CREs were not unique and had multiple hits
589 throughout the genome. One of the unique regions, the CRE *Dll319*, was selected as a suitable
590 target for CRISPR knock-out.

591

592 **Single guide RNA design and production**

593 Single guide RNA (sgRNA) target sequences for *sal* were selected based on their GC content
594 (around 60%) and the number of mismatch sequences relative to other sequences in the genome
595 (> 3 sites). In addition, we selected target sequences that started with a guanidine for subsequent
596 *in vitro* transcription by T7 RNA polymerase. sgRNA for the *Dll319* CRE were designed using
597 CRISPR Direct(33), corresponding to GGN20NGG. We designed 4 guides spanning the length
598 of the CRE (Fig. 3B, Fig. S15, and Table S2). Two guides were designed targeting the *sal740*

599 region (Fig. S24, and Table S2). The sgRNA templates were created by a PCR reaction with
600 overlapping primers, using Q5 polymerase (New England Biolabs). Constructs were
601 transcribed using T7 polymerase and (10X) transcription buffer (New England Biolabs),
602 RNase inhibitor (Ribolock), NTPs (10 mM) and 600 ng of the PCR template. The final sample
603 volume was 40 μ L. Samples were incubated for 16 h at 37°C and then treated with 2 μ L of
604 DNase 1 at 37°C for 15 minutes. Samples were purified by ethanol precipitation, and RNA
605 size and integrity were confirmed by gel electrophoresis.

606

607 **Cas9 mRNA production**

608 The plasmid pT3TS-nCas9n (Addgene) was linearized with XbaI (NEB) and purified by
609 phenol/chloroform purification and ethanol precipitation. pT3TS-nCas9n was a gift from
610 Wenbiao Chen (Addgene plasmid # 46757; <http://n2t.net/addgene:46757>;
611 RRID:Addgene_46757). *In vitro* transcription of mRNA was performed using the
612 mMESSAGE mMACHINE T3 kit (Ambion). One microgram of linearized plasmid was used
613 as a template, and a poly(A) tail was added to the synthesized mRNA by using the Poly(A)
614 Tailing Kit (Thermo Fisher). The A-tailed RNA was purified by lithium-chloride precipitation
615 and then dissolved in RNase-free water and stored at -80°C. The *Cas9* transcript was used for
616 producing *sal* crispants, and for the analysis of regulatory interactions among *Dll*, *Antp*, and
617 *sal*.

618

619 ***In vitro* cleavage assay for the *Dll319* CRE**

620 The sgRNAs were tested using an *in vitro* cleavage assay. Wild-type genomic DNA was
621 amplified using primers that were at least 200 bp from the sgRNA sites. sgRNA (200 ng/ μ L
622 per guide), Cas9 protein (800 ng/ μ L) (stored in a buffer containing 300 mM NaCl, 0.1 mM
623 EDTA, 1 mM DTT, 10 mM Tris-HCl, 50% glycerol pH 7.4 at 25°C), NEB buffer 3 (1 μ L) and
624 BSA (1 μ L) were brought to a final volume of 10 μ L with nuclease-free water and incubated
625 at 37°C. After 15 minutes of incubation, the purified amplicon (100 ng) was added to the
626 sample, which was then incubated for an additional 1-2 h at 37°C. The entire reaction volume
627 was analyzed on a 1%-agarose gel. Cas9 protein was purchased from NEB EnGen Cas9 NLS.
628 The cleavage assay confirmed that each guide successfully cleaved the PCR amplicon.

629

630 **Embryo injections**

631 Wild-type lab populations of *B. anynana* adults were provided with corn plants to lay eggs.
632 The eggs were collected within 1.5 h of oviposition and placed onto 1-mm-wide strips of
633 double-sided tape in plastic Petri dishes (90 mm). Cas9 protein (final concentration 800 ng/ μ L)
634 and sgRNA (final concentration 200 ng/ μ L per guide) for all 4 guides were prepared in a total
635 volume of 10 μ L and incubated for 15 min at 37°C prior to injection along with 0.5 μ L of food
636 dye to improve visualization of the injected sample into the embryos. For *sal* crispants, Cas9
637 mRNA (500 μ g/ μ L final concentration) and sgRNA (500 μ g/ μ L final concentration) were
638 injected along with one tenth of the volume of food dye. For *sal740* CRE, eggs were injected
639 with the mix of Cas9 protein (final concentration 800 ng/ μ L) and sgRNA (final concentration
640 400 ng/ μ L per guide). The injection mixture was kept on ice after the incubation and prior to
641 injection. Embryo injections were carried out by nitrogen-driven injections through glass
642 capillary needles. Injected eggs were stored in closed Petri dishes containing cotton balls that

643 were dampened daily to maintain humidity. The hatched larvae were reared in small paper cups
644 for 1 week and then moved to corn plants to complete their development. Tables S1, S3 and
645 S4 summarize the injection results.

646

647 ***In vivo* cleavage assay and genotyping of *sal* crispants**

648 Genomic DNA was extracted with a SDS-based method from a pool of 5 injected embryos that
649 did not hatch. About 250 bp of sequence spanning the target sequence was amplified with
650 PCR BIO Taq Mix Red (PCRBIOSYSTEMS), and PCR conditions were optimized until there
651 were no smears, primer dimers, or extra bands. Primers are listed in Table S2. The PCR
652 products were purified with the Gene JET PCR purification kit (Thermo Fisher). Two hundred
653 nanograms of PCR product were denatured and re-annealed in 10x NEB2 buffer. One
654 microliter of T7 endonuclease I (NEB) was added to the sample, while 1 μ L of MQ water was
655 added to a negative control. Immediately after the incubation for 15 min at 37°C, all the
656 reactions were analyzed on a 3% agarose gel. Amplicons that showed positive cleavage from
657 the T7 endonuclease I assay were subcloned into the pGEM-Teasy vector (Promega) through
658 TA cloning. For each target, we picked 8 colonies, extracted the plasmid with a traditional
659 alkali-SDS method, and performed a Polyethylene glycol (PEG) precipitation. Sequence
660 analysis was performed with the BIGDYE terminator kit and a 3730xl DNA Analyzer (Thermo
661 Fisher).

662

663 **Screening and genotyping *Dll319* crispants**

664 Newly emerged caterpillars were screened under a microscope to look for developmental
665 defects affecting any regions where *Dll* is expressed, such as the thoracic legs, mouthparts, and
666 prolegs. Any caterpillars exhibiting defects were imaged and reared individually in paper cups
667 until the butterflies eclosed. Caterpillars that died were immediately frozen for DNA isolation
668 and genotyping. All other surviving caterpillars with no apparent developmental abnormalities
669 were reared in groups on corn plants and fed *ad-libitum* every 2 days until pupation. The
670 eclosed butterflies were frozen individually at -20°C. Each butterfly was carefully screened
671 under a microscope and examined for asymmetric crispant phenotypes, focusing particularly
672 on phenotypes expected for a *Dll* knock-out, such as appendage, eyespot, or pigmentation
673 defects.

674

675 **Colony PCR to identify CRE deletions**

676 For selected crispants, genomic DNA was extracted from the thorax (E.Z.N.A tissue DNA kit)
677 and used for PCR to prepare samples for genotyping. The samples were visualized on a gel to
678 confirm the correct size band and the PCR product was purified using a Thermo Scientific PCR
679 purification kit. The DNA was cloned into a pGEM T-Easy Vector (Promega) and the plasmid
680 was transformed into DH5 alpha *E. coli*. White colony selection was used for colony PCR. The
681 bands were visualized on a 1%-agarose gel to look for bands with shifts relative to the WT
682 band. PCR products from colonies showing evidence of a deletion were submitted for Sanger
683 sequencing PCR (Axil Scientific, Singapore), including a sample that was amplified from *B.*
684 *anymana* wild-type genomic DNA.

685

686 **Butterfly enhancer reporter assay**

687 A 917 bp region containing the *Dll319* CRE was cloned into the piggyGUE vector via the
688 GATEWAY technology (Thermo Fisher). piggyGUE is the EGFP version of piggyGUM, the
689 piggyBac-based reporter construct that was previously published(22). The details of piggyGUE
690 will be published elsewhere (Deem and Tomoyasu, unpublished). The 917 bp region was
691 amplified from *B. anynana* wild-type genomic DNA using a primer containing CACC at the
692 5' end for directional cloning. The PCR product was cloned into the pENTR vector and further
693 cloned into the piggyGUE vector via a LR reaction, as described by Lai et al., 2018(22). Four
694 microliters of the LR reaction mix were used for bacterial transformation. After sequence
695 analysis to confirm the presence of SNPs in the *Dll319* CRE, plasmid DNA was amplified,
696 using a Midiprep kit (Qiagen). The piggyGUE *Dll319* CRE plasmid was diluted to 1µg/µL
697 and mixed in a 1:1 ratio with a hyperactive *piggyBac* transposase plasmid(34). Embryos
698 (n=550) were collected from *B. anynana* butterflies reared at 27°C and were injected ~1-hour
699 after egg laying with the plasmid solution and a small amount of food dye, using a glass
700 injection needle and nitrogen gas pressure. Eggs were transferred in a Petri dish to a chamber
701 and kept moist to prevent dehydration. From this batch of eggs, 40 caterpillars hatched and
702 were reared in paper cups during the first week and then transferred to cages with corn plants
703 to complete their development. At all stages, caterpillars were fed corn *ad-libitum*. From this
704 batch of caterpillars, 19 reached adulthood (10 females and 9 males). These butterflies were
705 evenly distributed into 4 cages (~5/cage) and placed with respective wild-type males and
706 females for breeding. We were unable to observe any *dsRed* signal (the positive marker of
707 transgenesis driven by the 3xP3 promoter) in the eyes of the caterpillars from the F1 or F2
708 generation, despite ubiquitous *dsRed* signal in some 1st-instar larvae (only) of the F1 generation,
709 which were used later for outcrossing to wild-type individuals. This ubiquitous signal was not
710 observed again in the offspring of these larvae. We collected eggs from the F3 generation and
711 dissected some embryos for EGFP antibody staining. Two out of the four dissected embryos
712 did show expression of EGFP driven by the *Dll319* CRE in the embryonic antennae,
713 mouthparts, thoracic legs and pleuropodia (Fig. 3K, Fig. S10). Subsequent hemolymph PCR
714 genetic screening in individuals of the 4th generation failed to identify additional positive
715 individuals and the line was lost.

716

717 ***Drosophila* enhancer reporter assay**

718 The same 917 bp sequence that contained The *Dll319* CRE was directionally cloned into
719 pENTR-D, then GATEWAY cloned into the piggyPhiGUGd, the Gal4-delta version of the
720 previously reported piggyBac-based reporter construct(22). piggyPhiGUGd also has an attB
721 site, allowing phiC31 transgenesis. The detail of piggyPhiGUE will be published elsewhere
722 (Deem and Tomoyasu, unpublished). For *Drosophila* transgenesis, the piggyPhiGUGd *Dll319*
723 CRE construct was transformed into the attP2 site (68A4) through phiC31 integrase-mediated
724 transgenesis system with EGFP as a visible marker (BestGene *Drosophila* transgenic service).
725 Established transgenic flies were crossed with G-TRACE(35) to visualize the tissues with CRE
726 activities.

727

728 **Antibody staining of *B. anynana* embryos and wings**

729 Two-day-old embryos, as well as fifth-instar larval and pupal wing tissues were dissected in
730 PBS buffer under the microscope. The samples were fixed in 4% formaldehyde/Fix buffer (0.1

731 M PIPES pH 6.9, 1 mM EGTA pH 6.9, 1.0% Triton x-100, 2 mM MgSO₄) for 30 min on ice.
732 The samples were washed with 0.02% PBSTx (PBS + Triton x-100) 3 times every 10 min, and
733 then blocked in 5% BSA/PBSTx for 1 h. The samples were then incubated in 5% BSA/PBSTx
734 with the primary antibody, and incubated at 4°C overnight. As primary antibodies, we used a
735 rabbit polyclonal anti-Dll antibody (at 1:200, a gift from Grace Boekhoff-Falk), a mouse
736 monoclonal anti-Antp 4C3 antibody (at 1:200; Developmental Studies Hybridoma Bank), a
737 rabbit anti-Sal antibody (at 1:20,000 for wings and pupal tissues, and 1:2,000 for embryos; de
738 Celis et al., 1999), and a rabbit anti-EGFP antibody (at 1:200; Abcam ab290) for the transgenic
739 embryos at 24h (n=4) and wt controls. For double staining, we added two primary antibodies
740 to the same tube. The wings were washed with PBSTx 3 times every 10 min. Then, we replaced
741 the PBSTx with 5% BSA/PBSTx to block for 1 hour, followed by the incubation with the
742 secondary antibody (1:200) in 5% BSA/PBSTx at 4°C for 2 h. The wings were washed with
743 PBSTx 3 times every 10 min, followed by mounting the wings in ProLong Gold Antifade
744 Mountant (Thermo Fisher). The images were taken under an Olympus FV3000 Confocal Laser
745 Scanning Microscope.

746

747 **Sample collection and library preparation for RNA sequencing**

748 In order to identify gene expression patterns specific to eyespot formation on the developing
749 wings, we extracted RNA from sixteen different tissue types at four developmental time points:
750 3-4-hour-old, 12-13-hour-old, and 24-25-hour-old embryos; T1 legs, prolegs, forewings, and
751 horns from wandering caterpillars; T1 legs, antennae, forewings, forewing margins, eyes,
752 eyespots, and two control tissues adjacent to eyespot centers from 3-h-old pupae (**Fig. 1A**). For
753 wing wounding experiments, we poked one wing between 17 to 18 h after pupation in two
754 different places in the M3 sector, using a fine tungsten needle with a diameter of 0.25 mm and
755 0.001 mm at the tip (FST- 10130-10). We collected the wings 6 hours later, which corresponds
756 to 23-24 h after pupation (Monteiro et al 2006). We performed the experiments with four
757 biological replicates for each tissue type with 10 to 25 female individuals in each replicate
758 (both left and right tissues were used, except for the wounded pupal wings, where a single wing
759 was used) (Table S5). Total RNA was extracted in 70 µL of nuclease-free water, using Qiagen
760 RNA Plus Mini Kit. RNA quantity and integrity were measured using a Nanodrop and an RNA
761 Bleach gel (Aranda et al 2013). RNA libraries were prepared, using the Truseq stranded mRNA
762 kit from Illumina. Forty million reads were sequenced for each replicate, using Novoseq 6000
763 with 150bp paired-end and an average insert size of 250-300 bp. Library preparation and
764 sequencing were carried out at AIT Novogene, Singapore. In order to avoid batch effects, we
765 randomized the sample extraction and RNA isolation, such that two replicates of the same
766 group were never extracted at same time.

767

768 **RNA-seq analysis**

769 The raw RNA-seq data were quality-controlled and filtered. Adapter sequences and reads with
770 low quality (less than Q30) were trimmed, using bbduk scripts (ktrim=r, k=23, mink=11,
771 hdist=1, tpe, tbo, qtrim=rl, trimq=30, minlen=40). In order to remove any bacterial
772 contamination in the samples, we used the bbsplit script, which is a part of the bmap tools
773 (36). All bacterial genomes were downloaded from NCBI (last downloaded in June 2018), and
774 the reads were mapped to the bacterial genomes, using bmap. Only reads whose pairs also

775 passed through the filter were further analyzed. To remove any ribosomal RNA sequences from
776 the RNA-seq data, the reads were aligned to the eukaryotic rRNA database available in
777 sortmeRNA (37). The processed reads from different samples were then mapped to the BaGv2
778 genome, using hisat2 (38) (mapping statistics in Table S6), resulting in bam files that were
779 sorted by genomic positions, using samtools (30). They were used as inputs in StringTie (38)
780 to create the initial transcriptome assembly with 71,042 transcripts, which was used to annotate
781 the genome using Maker v.3 (39), resulting in 18,196 genes with 29,389 transcripts.

782

783 **RNA-seq differentially expressed (DE) gene analysis**

784 A read count matrix of the annotated genes was obtained for the samples using StringTie (38).
785 We used the GO terms to filter out any ribosomal genes before obtaining the read counts. This
786 approach led to the removal of 496 genes to a final set of 17,700 genes, which was used
787 throughout the analysis. Correlations between the replicate samples was analyzed using
788 DESeq2 (8) with a sample distance matrix. One of the antennal samples was removed due to
789 its poor correlation with its other biological replicates. The remaining samples were used for
790 the downstream analyses (Fig. S25).

791

792 **Identifying eyespot-specific DE genes**

793 To identify eyespot-specific genes, a pairwise DE analysis was performed between eyespot and
794 control adjacent tissues, Nes1 and Nes2, using DESeq2 (Fig. 1A, Fig. S2). Common genes
795 upregulated and downregulated between eyespot vs. Nes1 and eyespot vs. Nes2 with an
796 adjusted P-value (padj) of 0.05 were chosen as eyespot-specific DE genes (Spreadsheet S1).

797

798 **RNA hierarchical sample clustering**

799 In order to identify the tissue with the closest gene expression profile to eyespot, we used all
800 tissue samples except the eyespot control tissue samples. DE analysis between the multiple
801 tissues was performed, using `run_DE_analysis.pl` script provided in Trinity tool, using
802 DESeq2 as the method of choice for this analysis (40). Hierarchical clustering was performed
803 for the different tissues, using genes that showed a log2fold change of $|2|$ and padj value of
804 0.001, as in Fisher et al (2020) (41). Clustering was performed using an Euclidean distance
805 matrix derived, using the DE genes for the tissues with the `hclust` function in R(42). The `pvclust`
806 package(9) in R was used to calculate the uncertainty in the hierarchical clustering with a 1000
807 bootstrap value.

808

809 **ATAC-seq library preparation**

810 We prepared ATAC libraries for the same set of tissues as we did for the RNA-seq experiment,
811 except for the eyespot control tissues (Table S7). Library preparation failed for a few groups
812 leading to 2 to 4 biological replicates per group. Tissues were collected, flash-frozen in liquid
813 nitrogen and stored in -80°C , before we extracted nuclei and prepared the libraries. We used 10
814 to 25 individuals and approximately 80,000 nuclei per replicate. Libraries were prepared as
815 described in the Omni-ATAC protocol (43) with slight modifications. Individual tissues
816 extracted at different time periods during the process were randomized and pooled into each
817 replicate before extracting the nuclei. The tissues were thawed and homogenized in 2 mL of
818 ice cold 1X homogenization buffer (HB) in a 2-mL-glass douncer. Homogenization was

819 performed by 10 strokes with pestle A, followed by 15 strokes with pestle B. The homogenized
820 mixture was left on ice for 2 min before filtering it through a 100- μ m- nylon mesh filter into a
821 DNA “low bind” 2-mL Eppendorf tube (Z666556-250EA). The filtered mixture was
822 centrifuged at 2500 rpm, and the pellet (the nuclei) was collected along with 50 μ L of the
823 solution at the bottom, keeping unwanted cytoplasmic RNAs in the top layers. The filtered
824 nuclei were diluted in ATAC-Resuspension Buffer (RSB buffer), and 10 μ L of the solution
825 were used to count the nuclei, using a hemocytometer. Approximately 80,000 cells were used
826 for each replicate to prepare the libraries. The tagmentation enzyme (TDE1) was obtained from
827 Illumina (Illumina tagment dna tde1 enzyme and buffer smaller kits – 20034197). As the
828 concentration of the TDE1 and cell number greatly affect the identification of open chromatin
829 regions, we estimated the amount of enzyme needed for each reaction, using the formula:
830 volume of enzyme = genome size of *B anynana* [475MB] * number of cells [80,000] *2.5/
831 (genome size of humans [3200MB] *50,000). We used 0.65 μ L (final concentration of 10.4
832 nM) of enzyme for each reaction. The Omni-ATAC transposition reaction was carried out as
833 follows: 80,000 cells suspended in ATAC-RSB buffer were centrifuged at 2500 rpm for 10
834 min at 4° C. The supernatant was removed, and the nuclei-containing pellet was kept. To
835 perform the cell lysis, 50 μ L of ice-cold ATAC-RSB were added to the pellet, along with 0.1%
836 NP40, 0.1% Tween 20, and 0.01% Digitonin. The mixture was incubated for 3 min on ice.
837 Subsequently, 1 mL of ATAC-RSB buffer containing only 0.1% Tween and no NP40 nor
838 digitonin was added, and the mixture was centrifuged at 2500 rpm. The supernatant was
839 discarded, and the pellet was retained, to which 50 μ L of transposition mixture (6.5 μ L 2x TD
840 buffer, 0.65 μ L transposase (10.4 nM final concentration), 16.5 μ L PBS, 0.5 μ L 1% digitonin,
841 0.5 μ L 10% Tween-20, 25.35 μ L H₂O) were added. The reaction was incubated for 25 minutes
842 at 37° C at 1000 rpm in a thermomixer. After the transpositions and tagmentation occurred, the
843 samples were prepared for sequencing by adding Illumina/Nextera adapters with dual indexing
844 and further PCR amplified for 10 cycles. The PCR products were purified, using a Zymo-DNA
845 Clean & Concentrator-5 kit, and the DNA fragments were size-selected between 50 – 1500 bp,
846 using the ProNex Size-Selective Purification System (NG2002) from Promega. The samples
847 were sequenced, using Novoseq 6000 with an average read depth of 30 million and 2x50 bp
848 paired end reads by AIT Novogene, Singapore.

849

850

851 **ATAC-seq peak calling**

852 ATAC reads were processed, using bbdut scripts from bbmap tools, to remove any adapters.
853 The reads were mapped to the BaGv2 genome, using bowtie with the -x 1500 and -m1
854 parameters. Only reads with insert sizes of 1500 bp or less and only those mapping to a unique
855 region of the genome were mapped. Reads mapped to the mitochondrial genome were
856 removed, using samtools idxstats, and reads marked for PCR duplicates were also removed,
857 using GATK Markduplicates. We kept only paired-end mapped reads with a phred quality
858 score of Q20 and above for downstream analysis. Because the Tn5 transposase binds to DNA
859 as a dimer and inserts adapters of 9 bp in length at the insertion sites, the start sites of the
860 mapped reads were adjusted to an offset of +4 bp in the forward strand and -5 bp in the reverse
861 strand. The bam files were converted to bed files, using Bedtools (31), and we used F-Seq (44)
862 to call peaks for each sample. Bedtools intersect was used to identify the common set of peaks

863 for each tissue type. Peaks from all samples were merged if they were separated by 50 bp, using
864 Bedtools merge to create 313,425 consensus peaks used for the downstream analyses.
865 Featurecount from the Subread package(45) was used to extract a read count matrix
866 corresponding to the consensus peaks for all samples. The FRiP score, which is defined as the
867 fraction of all reads that are mapped to peaks across the entire genome, was used to measure
868 the quality of the ATAC-seq data. Our ATAC-seq data showed a median FRiP score of 0.846,
869 which is higher than the ENCODE standard (>0.3) for the fraction of reads falling into peaks
870 (Table S8). And deepTools (46) was used to access the sample correlation between the
871 replicates and quality of the libraries (Fig. S26).

872

873 **ATAC-seq differential peak analysis:**

874 Differential peak analysis was performed using DESeq2. Peaks were considered differentially
875 accessible with a padj value of 0.05. We also mapped the *Dll319* peak identified from the
876 FAIRE data to the BaGv2 genome, using blastn, to identify its position in the new genome
877 assembly and test whether the ATAC-seq analysis was also able to identify it. To identify
878 potential CREs for *sal*, ATAC peaks from 3hr pupal tissues were visualized using IGV and we
879 targeted one potential candidate region (*sal740*) within the intronic region of *sal* gene loci
880 which is open across almost of the tissues.

881

882 **Hi-C analysis and Virtual 4C**

883 The Hi-C library used for scaffolding the *B. anynana* genome was reanalyzed, using the *Dll319*
884 and *sal740* region as bait, to verify whether these regions interacted with the promoter of *Dll*
885 and *sal* respectively. Libraries were mapped to the BaGv2 assembly, using Juicer (47). We
886 used the contact map obtained from Juicer to construct a virtual 4C plot for the window around
887 the *Dll319* and *sal740* regions by placing reads in a 3-kb bin, using the script from Ray et al.,
888 2019 (48).

889

890 **Screening and genotyping *sal740* crispants**

891 Caterpillars that emerged were carefully screened under the microscope for any defects in their
892 body, especially in the head region where *sal* expression is observed. Individuals showing any
893 abnormalities were imaged and grown separately in a cup whereas all others were grown in a
894 separate cage. Adults were immediately frozen at -20° C after emergence and screened later
895 under a microscope for any defects in eyespots, wings, legs and in antenna.

896

897 **Hi-C Genome Assembly**

898 Eggs were collected from a single pair of mated *B. anynana* butterflies and reared. Eighteen
899 female siblings were harvested at the wandering stage for DNA extraction. Guts were removed,
900 and the samples were immediately flash-frozen in liquid nitrogen and stored at -80 °C before
901 the samples were sent to Dovetail Genomics to perform Chicago and Hi-C library preparation
902 and analysis. The Chicago library preparation uses *in vitro* chromatin fixation, digestion, and
903 crosslinking of regions in the genome that are close to each other in terms of 3D chromatin
904 architecture. In order to sort and scaffold the genome, 233 million reads (2x150bp) were
905 sequenced from the Chicago library and mapped to the previously published *B. anynana*
906 genome (v1.2) with 10,800 scaffolds(49). The HiRise pipeline was used to identify mis-

907 assemblies, to break the scaffolds, and to sort the scaffolds. Only scaffolds greater than 1kb in
908 length (n=5027) were used because scaffolds needed to be long enough for the read pairs to
909 align and be scaffolded in accordance with the likelihood model used by HiRise. Next, 153
910 million reads (2*150) sequenced from the Hi-C library were mapped to the genome assembly
911 output generated from the Chicago-HiRise pipeline to identify any mis-assemblies from the
912 Chicago pipeline and correct them to produce a final genome assembly of high contiguity.
913 The genome assembly obtained from the HiC pipeline was ordered, using the available linkage
914 information from Beldade et al., (2009)(50), using Chromonomer(51). Two hundred eighty-
915 nine SNP FASTA sequences were mapped to the Hi-C assembly, using blastn to identify their
916 corresponding positions in the Hi-C genome. Using the SNP position obtained from blastn, a
917 list describing the genetic map was manually created, which later passed through
918 Chromonomer to sort the Hi-C assembly resulting in the final assembly (**BaGv2**) that was used
919 for the current study. The BUSCO score(52) was used to check for the completeness of the
920 gene sets in the assembly.

921

922 *Genome Annotation*

923 The genome was repeat-masked for transposable elements, small repeats, and tandem repeats
924 before annotation as described in Nowell et al., 2017 (49) The soft repeat-masked genome was
925 annotated, using four rounds of Maker v.3 (39). The transcriptome assembled from the RNA-
926 Seq data and gene sequences annotated from the previous version of the genome were
927 combined and used as transcripts for the species, with transcriptome and protein sequences
928 from *Pieris rapi*, *Junonia coenia*, and *Bombyx mori* as relative transcripts and protein
929 homology evidences for the first round of gene predictions. Output gene predictions from each
930 round were used as input for the next round. Snap and Augustus were used for the second round
931 of gene predictions, followed by Genemark for the third round of gene modelling. Then we
932 performed one final round of Snap and Augustus predictions. The minimum length of 35 amino
933 acids was set for gene predictions. The predicted gene models were kept for genes that had an
934 Annotation Edit Distance (AED) score of < 1 and/or had a gene ontology obtained from
935 Interproscan (53). This resulted in 18,189 genes with 29,490 transcripts. In order to correct the
936 annotations and produce a standardized gff3 file, the gff file obtained from Maker was run
937 through agat_convert_sp_gxf2gxf.pl script, which is a part of AGAT tools (54). This step
938 resulted in the removal of 82 identical isoforms and added the missing gene features, leading
939 to a total of 18,196 genes with 29,389 transcripts. Functional annotation was performed by
940 locally blasting the transcripts against a non-redundant (nr) protein database, using diamond
941 blast (55), and a gene ontology analysis was performed using Interproscan in Blast2Go(56).
942 Finally, the blast results were merged with the interproscan results in Blast2Go to produce a
943 final functional annotation for the genome.

944

945 **Supplementary Results**

946

947 ***Bicyclus anynana* Hi-C Genome Assembly**

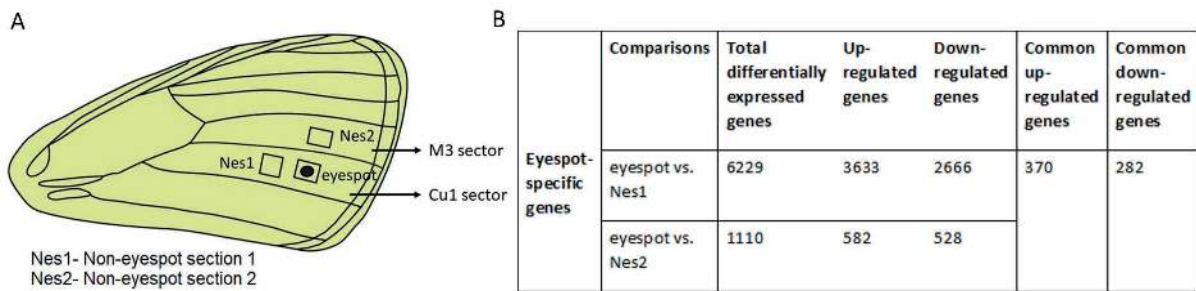
948 The published version of the *B. anynana* genome assembly (475 MB) contains 10,800 scaffolds
949 with an N90 value of 99.3 kB (49). To improve the current assembly, we performed
950 scaffolding, using a two-step approach, one with Chicago-HiRise followed by Hi-C-based

951 scaffolding. Chicago-HiRise scaffolding performed on the published version resulted in 3512
952 new joins with 634 breaks remaining in the genome, raising the N90 value to 840 kB. The Hi-
953 C scaffolding that followed corrected mis-joints from the Chicago-HiRise scaffolding by
954 creating four new breaks but made 512 new joints, improving the N90 value further to 12.073
955 MB and placing 98% of the bases (467.62 MB) into 28 scaffolds, achieving a near-
956 chromosomal level assembly. Following the Hi-C assembly and using the linkage map from
957 Beldade et al. (2009)(50) obtained for the 28 chromosomes of *B. anynana*, we produced one
958 manual break and one joint to achieve congruence of the two data sets. We were able to map
959 171 markers out of the 289 markers from the 28 linkage groups (50) in the genome, resulting
960 in an ordered chromosomal level assembly of 475.8 MB

961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994

995 **Supplementary Figures**

996



997

998

999

1000

1001

1002

1003

1004

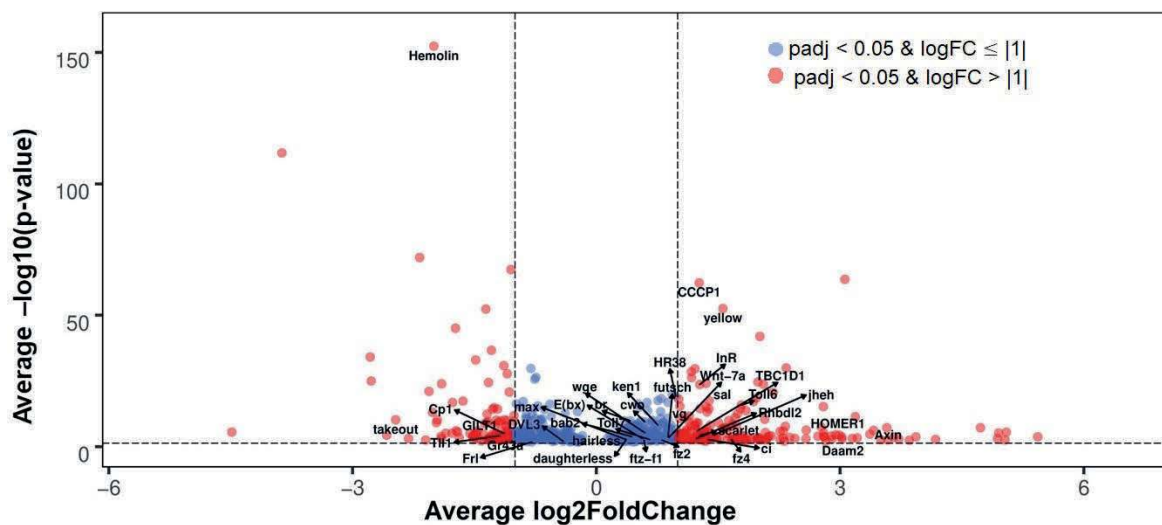
1005

1006

1007

1008

Fig. S1. Tissue selection to identify eyespot-specific DE genes. (A) Forewing image highlighting the regions chosen to perform DE analysis to identify eyespot genes. Nes1 and Nes2 are two control tissues representing tissue from the same and from a more anterior wing sector, where no eyespots develop. (B) Table Showing the number of genes differentially expressed (DE) in each of the comparisons, leading to 652 eyespot-specific DE genes that were in common across both comparisons.



1009

1010

1011

1012

1013

1014

1015

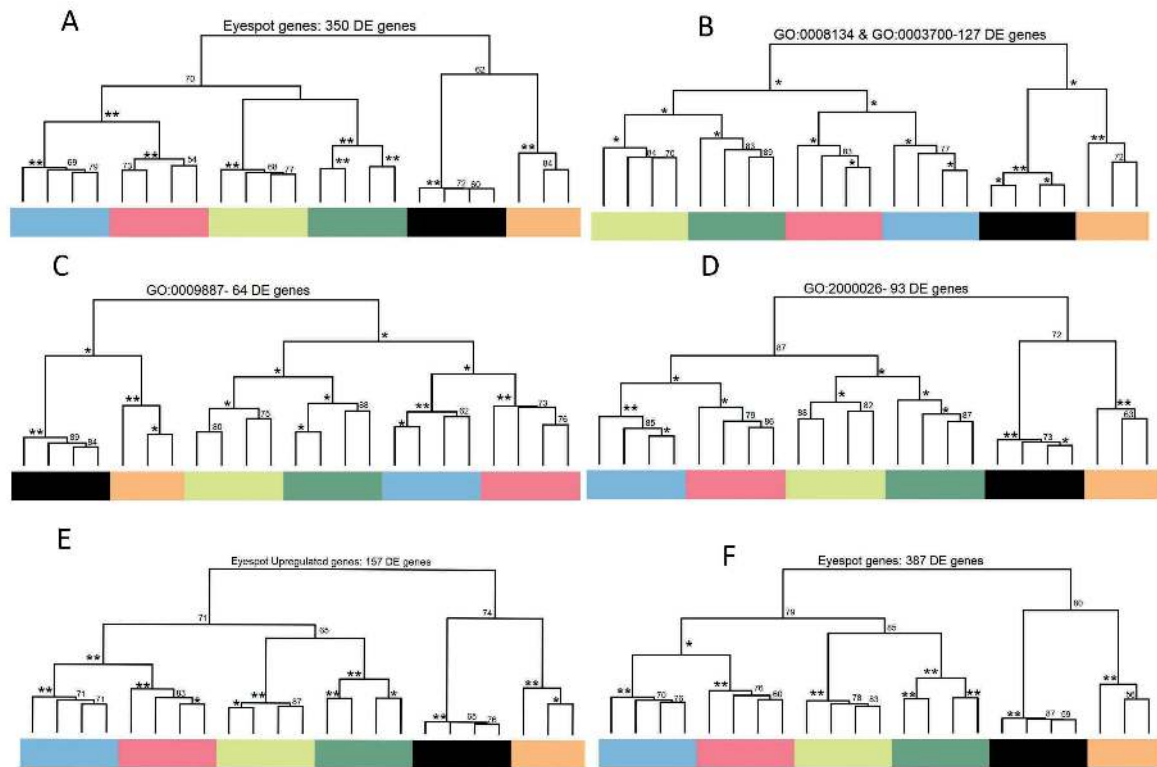
1016

1017

1018

1019

Fig. S2. Set of DE genes between the “eyespot” and “control” wing tissues (with adjusted p -values < 0.05). Up-regulated genes show a positive X-axis value, while down-regulated genes show a negative X-axis value. The \log_2FC and p -values were averaged among the two control tissue comparisons.



1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

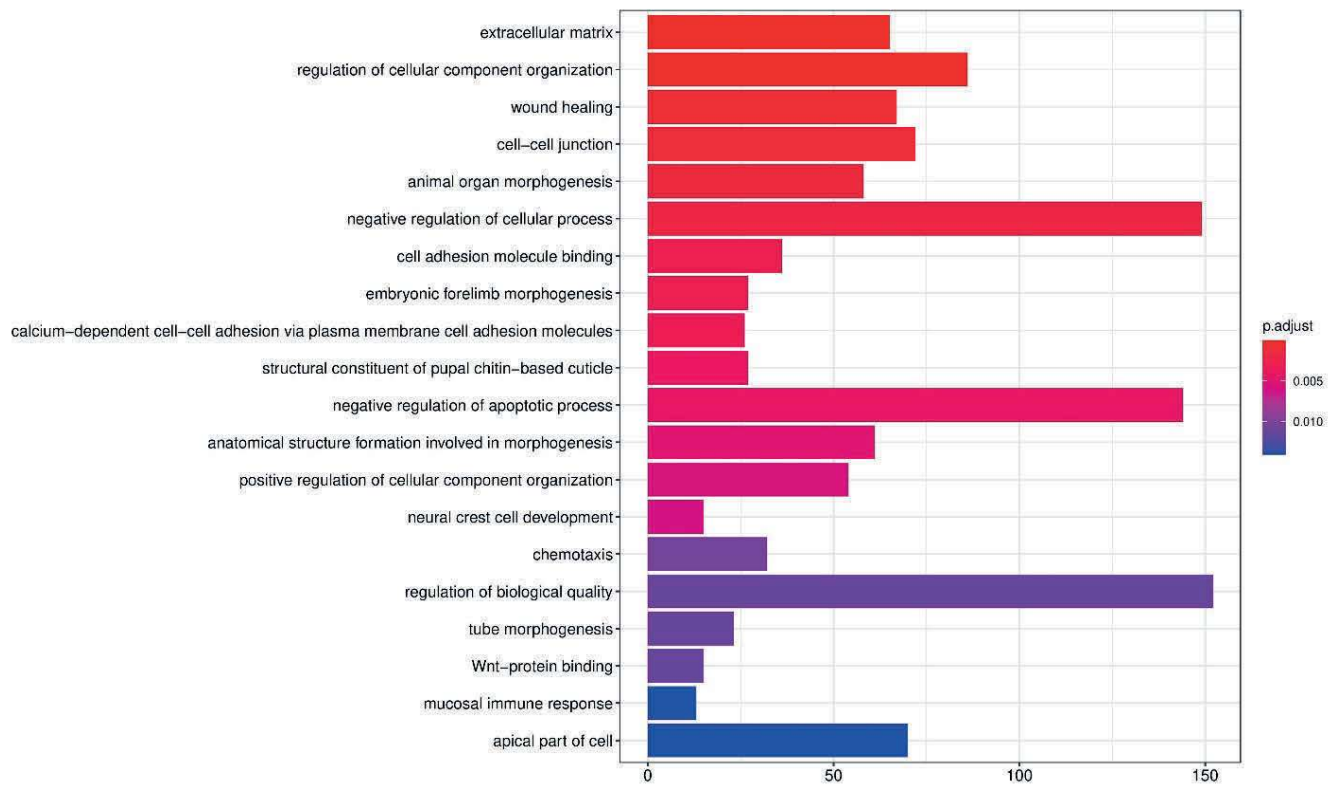
1043

1044

1045

Fig. S3. Character trees for 3-h pupal tissues, using various gene subsets revealing that eyespot gene expression patterns cluster with those of antennae. “DE genes” represents the number differentially expressed genes ($p_{adj} < 0.001$ and $\log_2FC \geq |2|$) from the initial subset of genes. Trees constructed using: A. eyespot-specific genes (652 genes), B. Transcription factors and co-factors (336 genes), C. genes enriched for GO terms associated with animal organ morphogenesis (108 genes), D. genes enriched for GO terms corresponding to anatomical structure formation involved in morphogenesis (165 genes), E. genes up-regulated in eyespots (370 genes), F. combined eyespot-specific DE genes predicted in this study as well as those published in Ozsu and Monteiro 2017 (10) (a total of 775 genes). In all six trees, eyespots gene expression patterns cluster with those of antennae and form an outgroup in the tree with another clade where eyes cluster with legs, and wings with wing margins.

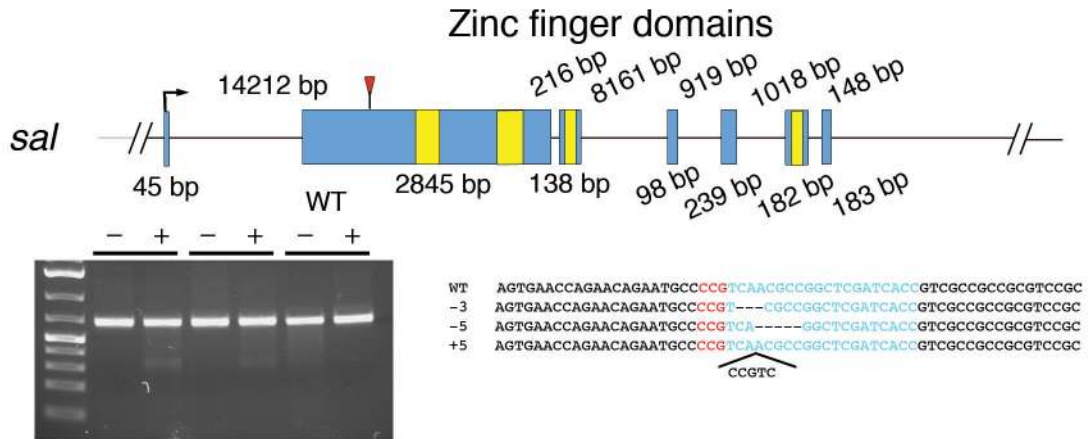
1046
1047
1048
1049
1050



1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070

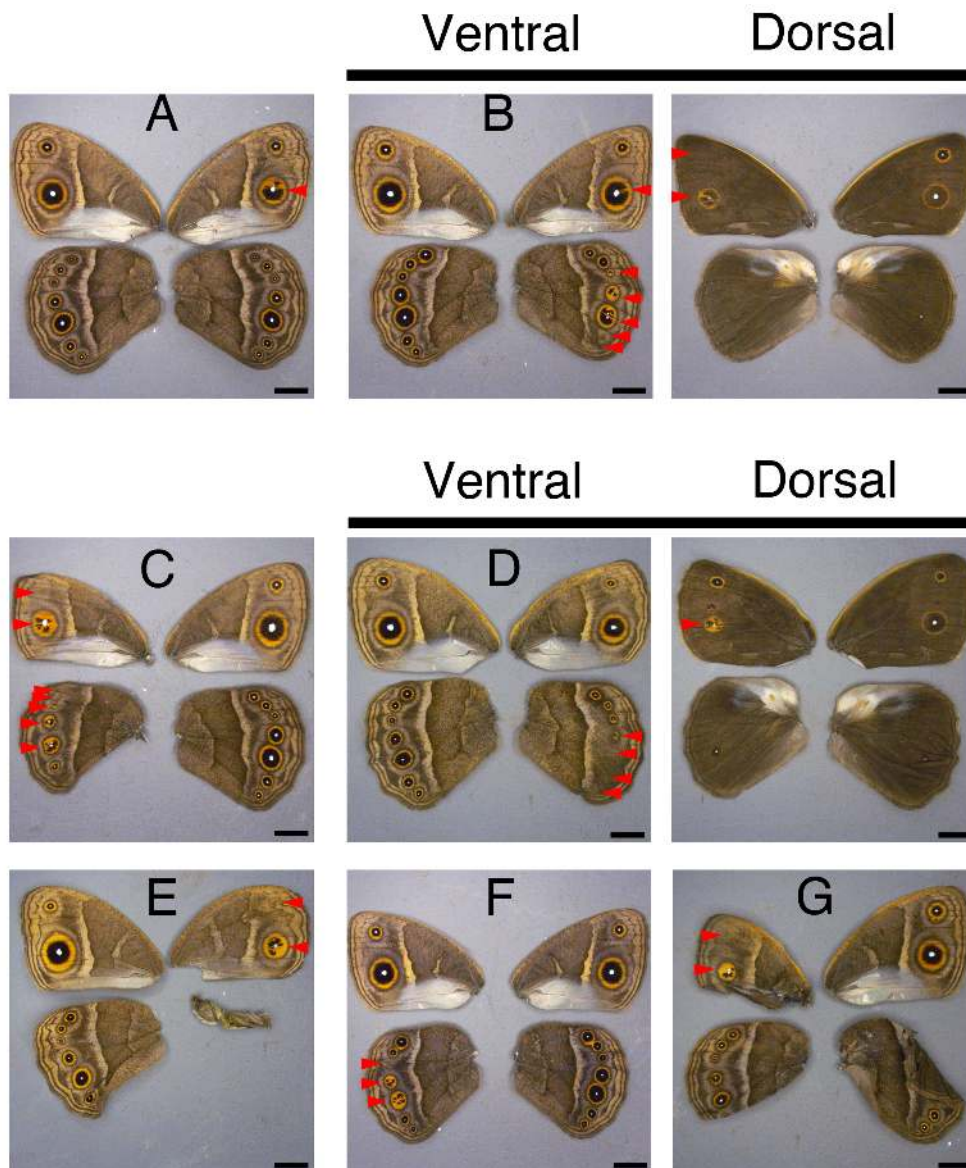
Fig. S4. Gene-set-enrichment plot for the 3839 DE genes from 3-h pupal tissues. The gene-enrichment analysis shows the various functions DE genes are involved.

1071
1072



1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086

Fig. S5. T7 endonuclease I assay and sequence analysis for CRISPR-Cas9 mutations in *sal*. Schematic representation of the *sal* gene structure. Blue boxes indicate exons, and yellow-coloured regions inside exons indicate functional domains. Each functional domain was annotated using a conserved-domain search at NCBI. Red arrowheads indicate the CRISPR target region. The gel shows the result of a T7 endonuclease assay performed on embryos after injection of sgRNA and Cas9 mRNA or protein or performed on Wt embryos (last two lanes). We performed the assay with two different samples for technical replicates. “Minus” lanes indicate a negative control, where T7 endonuclease was not added to the reaction. “Plus” lanes indicate the presence of T7 endonuclease. The expected sizes of digested bands were observed only from the lanes containing the T7 enzyme. Sanger sequence results indicate that an indel mutation was generated around the target site. Blue-coloured sequences indicate the sgRNA target sequence, and red-coloured sequences indicate the PAM sequence.



1087

1088

Fig. S6. *sal* crispant phenotypes

1089

(A) The Cu1 eyespot on the right forewing showed a transformation of orange scales into black

1090

scales. (B) The Cu1 eyespot on the right forewing and the M1, M2, M3, and Cu1 eyespots on

1091

the right hindwing showed transformation of black into orange scales. The Cu2 eyespot got

1092

reduced in size, and the A1 eyespot on the right hindwing disappeared. On the dorsal side, the

1093

M1 eyespot on the right forewing disappeared, but the Cu1 eyespot on the right forewing

1094

showed a transformation of black into orange scales. (C) The M1 eyespot on the left forewing

1095

was reduced in size, and the Cu1 eyespot showed a transformation of black into orange scales.

1096

On the hindwing, the Rs eyespot disappeared, and the M1, M2, M3, and Cu1 eyespots showed

1097

transformation of black into orange scales. (D) The M3 eyespot was reduced in size, and the

1098

Cu1, Cu2, and A1 eyespots disappeared. On the dorsal forewing, the Cu1 eyespot showed

1099

transformation of black into orange scales. (E) The M1 eyespot on the right forewing was

1100

reduced in size, and the Cu1 eyespot showed transformation of black into orange scales. The

1101

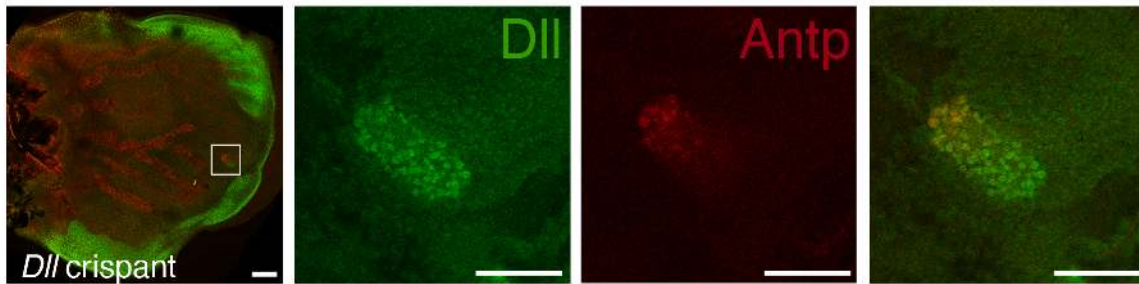
chevron pattern on the wing margin and the central symmetry system bands running the length

1102

of each wing were distorted. (F) The M2, M3, and Cu1 eyespots on the left hindwing showed

1103 transformation of black into orange scales. (G) The M1 eyespot on the left forewing
1104 disappeared, and the Cu1 eyespot showed transformation of orange into black scales. Mutated
1105 eyespots are marked with red arrowheads. Scale; 5 mm.

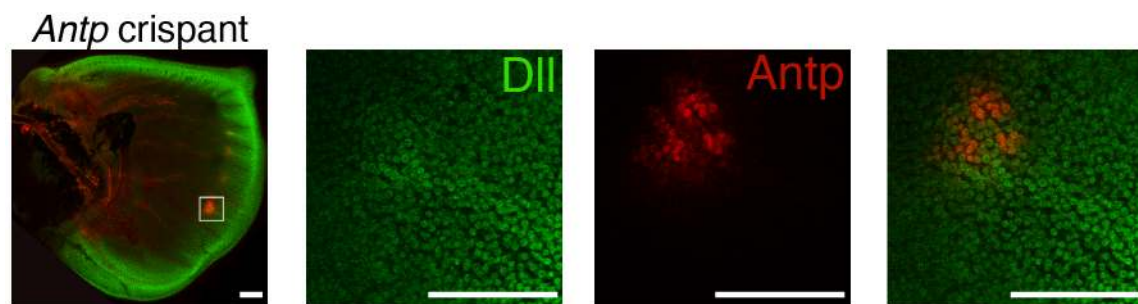
1106
1107
1108
1109
1110



1111
1112
1113
1114
1115

Fig. S7. Expression pattern of Dll and Antp proteins in a *Dll* crispant

Antp expression was only observed within the *Dll*-positive cells. Scale; 100 μ m for low magnification and 50 μ m for high magnification.



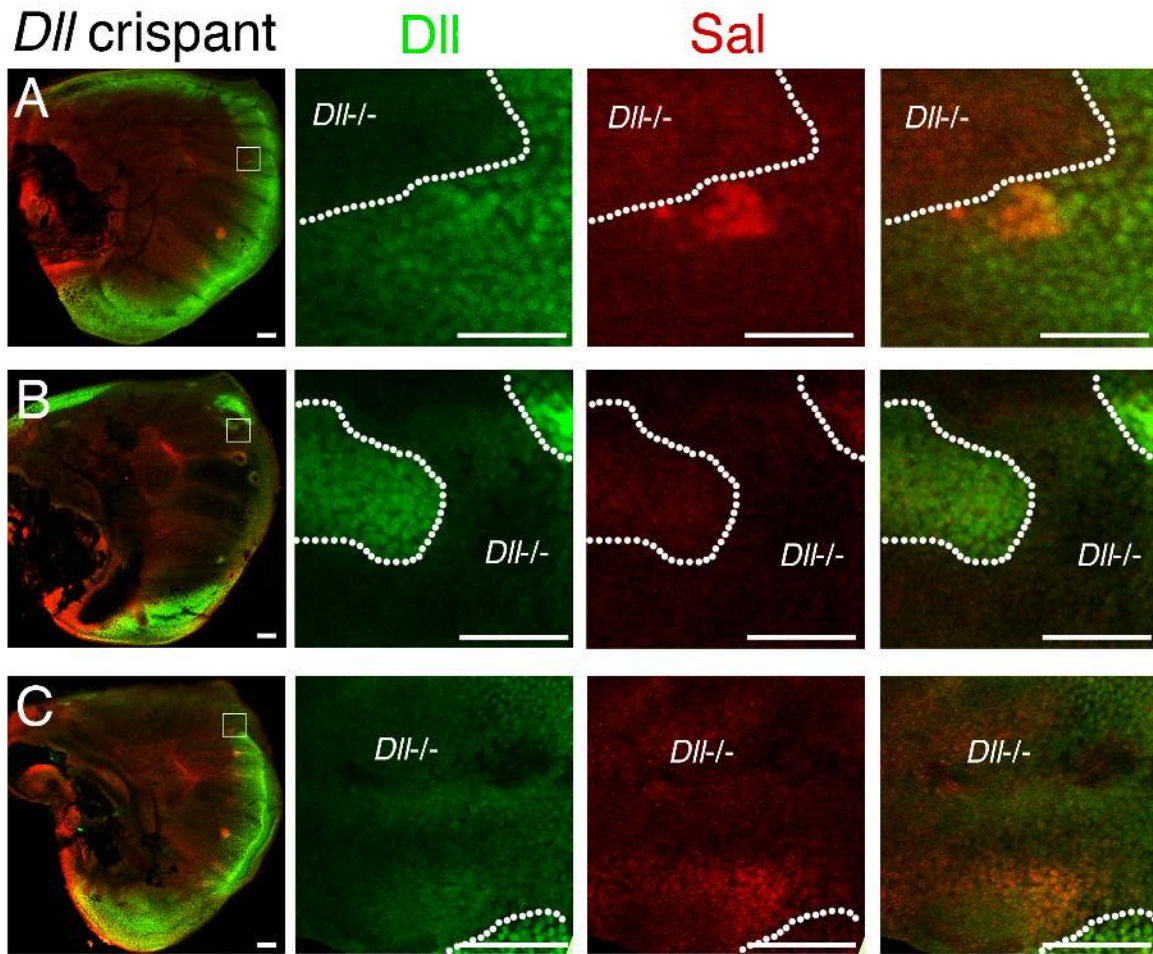
1116

1117 **Fig. S8. Expression pattern of Dll and Antp proteins in an *Antp* crispant**

1118 (A) *Antp* expression in Cu1 eyespot was partially lost, but Dll expression was not affected.

1119 Scale; 100 μ m for low magnification and 50 μ m for high magnification.

1120



1121

1122

Fig. S9. Expression pattern of Dll and Sal proteins in *Dll* crispants

1123

(A) Sal expression was lost in *Dll* null mutant cells of the M1 eyespot. (B) Cells from the

1124

middle of the wing were broadly mutated and lost Sal activity. (C) Cells anterior to the M1

1125

eyespot were mutated, which resulted in the loss of Sal expression in the eyespot centres. (D)

1126

Cells from the middle of the wing were broadly mutated, and in some wing sectors, Sal was

1127

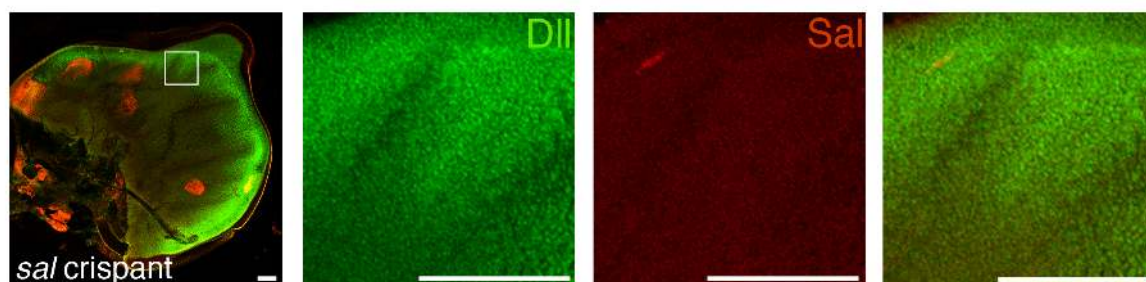
ectopically expressed in the distal wing region. Scale; 100 μm for low magnification and 50

1128

μm for high magnification.

1129

1130

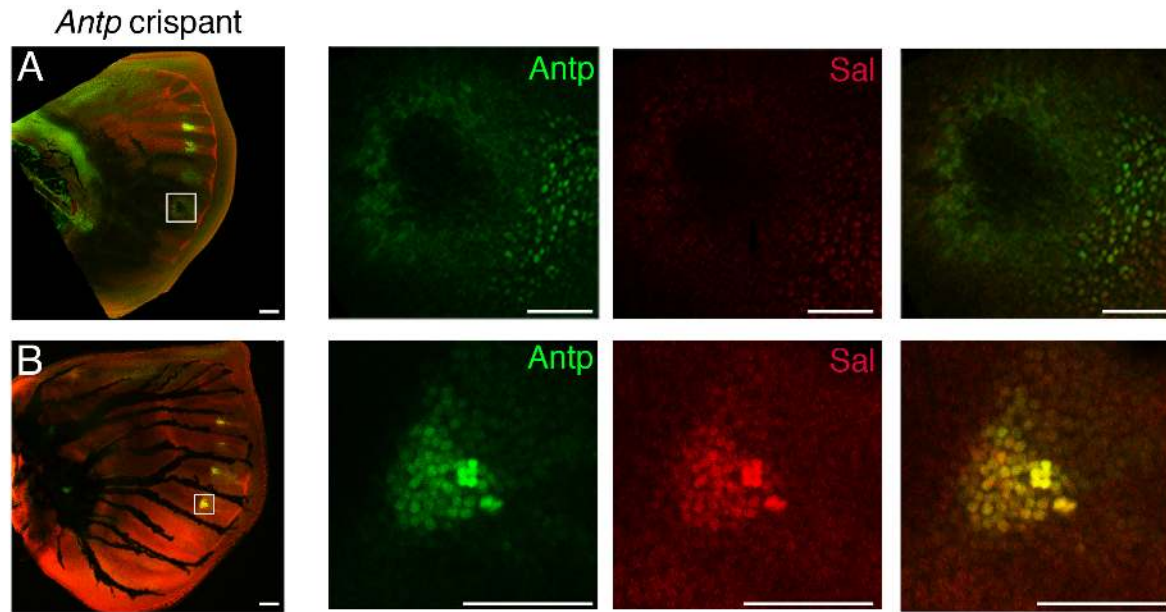


1131

1132 **Fig. S10. Expression pattern of Dll and Sal proteins in a *sal* crispant**

1133 The distal wing region was broadly mutated for Sal activity, but Dll expression was not
1134 affected. Scale 50 μ m.

1135



1136

1137

Fig. S11. Expression pattern of Antp and Sal proteins in *Antp* crispants

1138

(A) Sal expression was lost in *Antp* null mutant cells in the Cu1 eyespot. (B) Some Cu1 eyespot

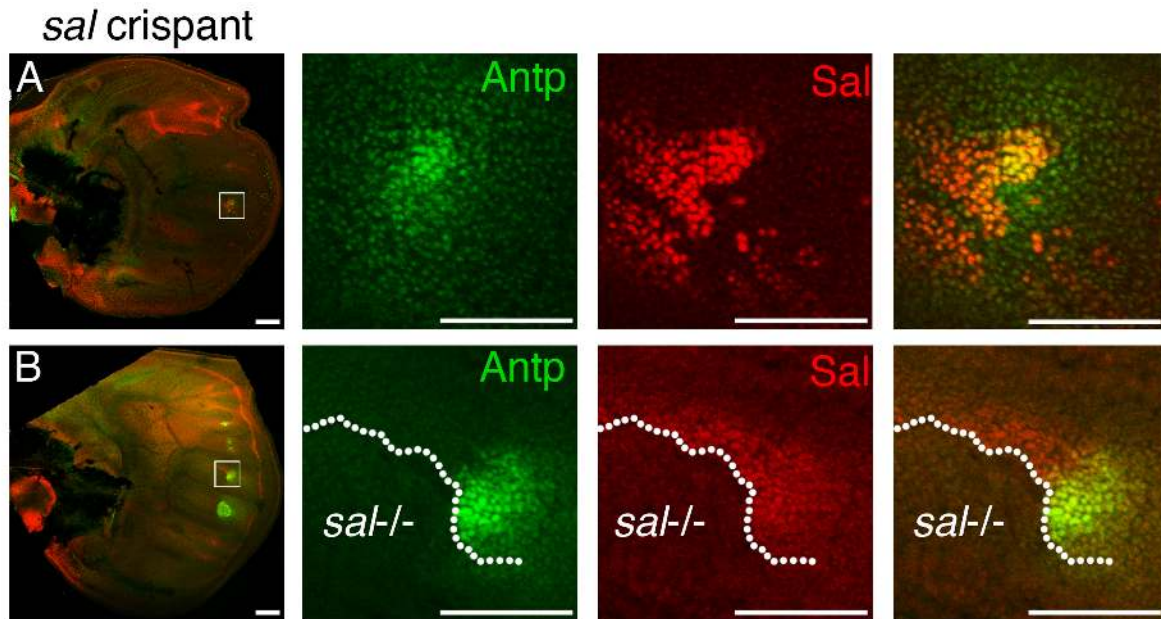
1139

cells lost Antp activity, and Sal expression was only detected in the Antp-positive cells. Scale;

1140

100 μm for low magnification and 50 μm for high magnification.

1141



1142

1143

Fig. S12. Expression pattern of Antp and Sal proteins in *sal* crispants

1144

(A) The wing is broadly mutated for Sal activity. Some Cu1 eyespot cells lost Sal expression,

1145

and Antp expression was only detected in the Sal-positive cells. (B) Cells around the M3

1146

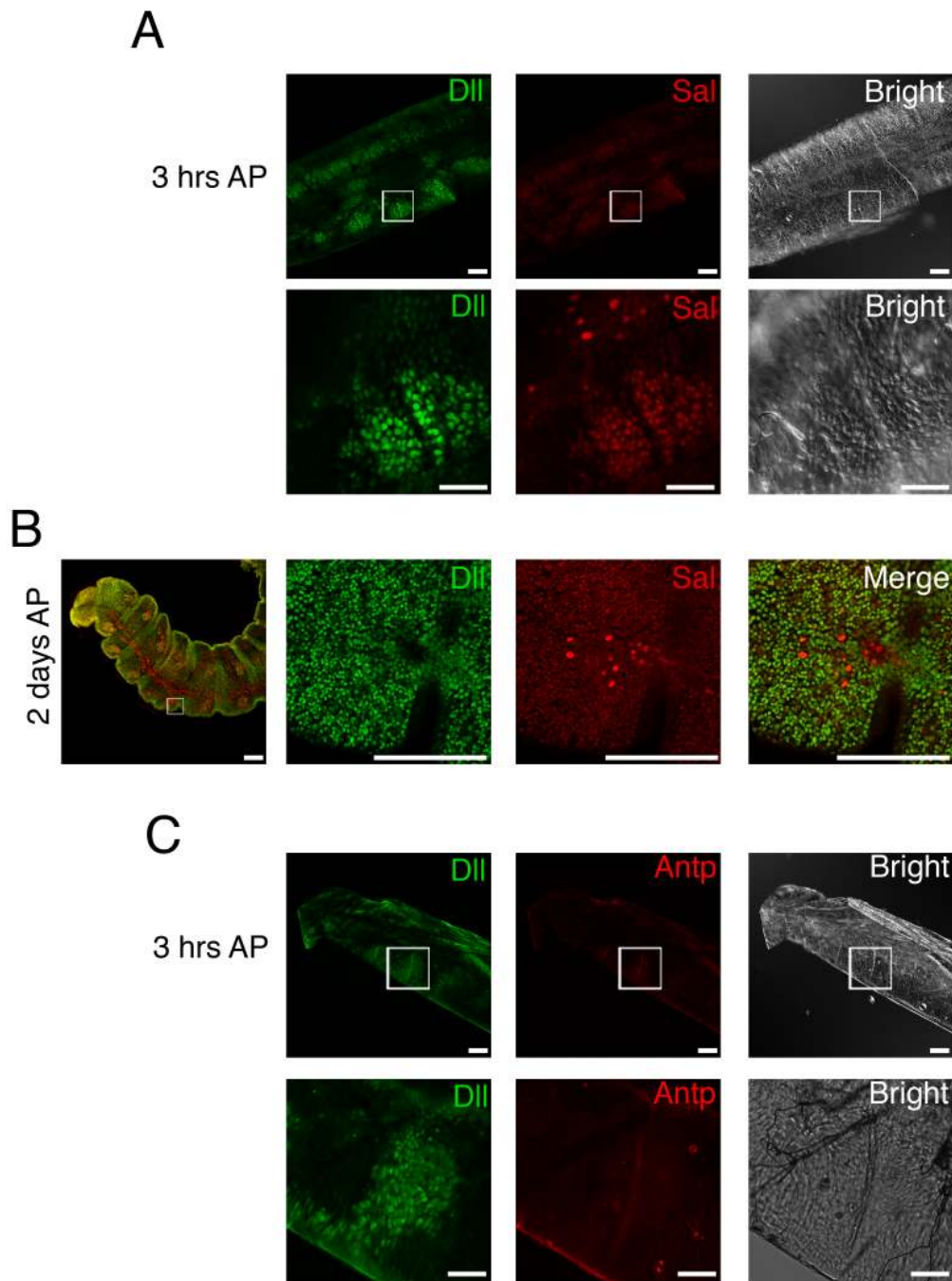
eyespot were mutated, and Antp expression was detected only in the Sal-positive cells. Scale;

1147

100 μ m for low magnification and 50 μ m for high magnification.

1148

1149



1150

1151 **Fig. S13. Expression of Dll, Sal, and Antp proteins in pupal antennae**

1152 (A) Dll and Sal were co-expressed in the segments of the developing antenna of a 3-h-old pupa.

1153 (B) Expression patterns of Dll and Sal in the pupal antenna, 2 days after pupation (AP). Dll

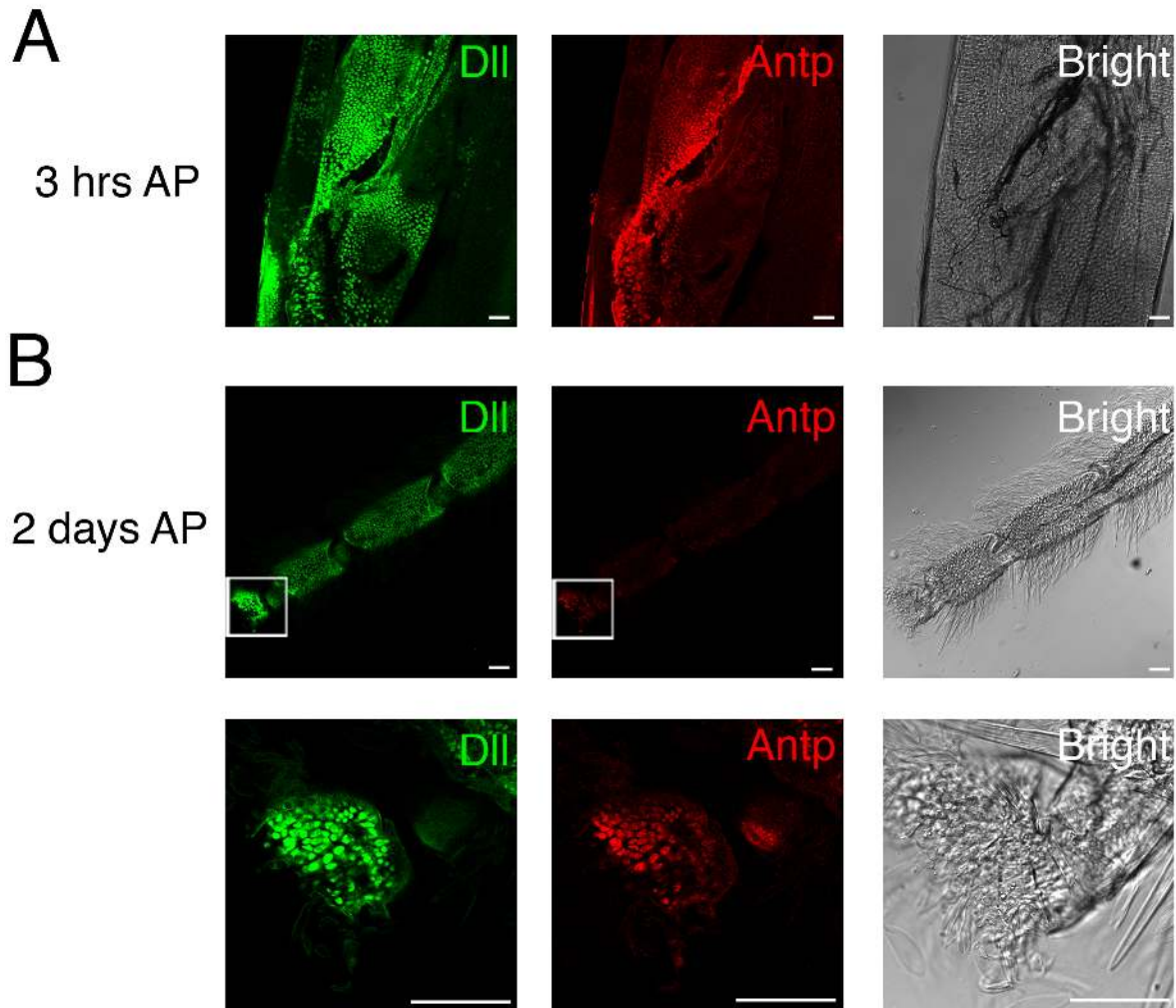
1154 was ubiquitously expressed in the antenna, but Sal expression is observed in the neurons. (C)

1155 Expression patterns of Dll and Antp in the antenna of a 3-h-old pupa. Antp expression was

1156 not observed in the antenna. The regions within the white squares are shown at higher

1157 magnification. Scale; 100 μ m for low magnification and 50 μ m for high magnification.

1158

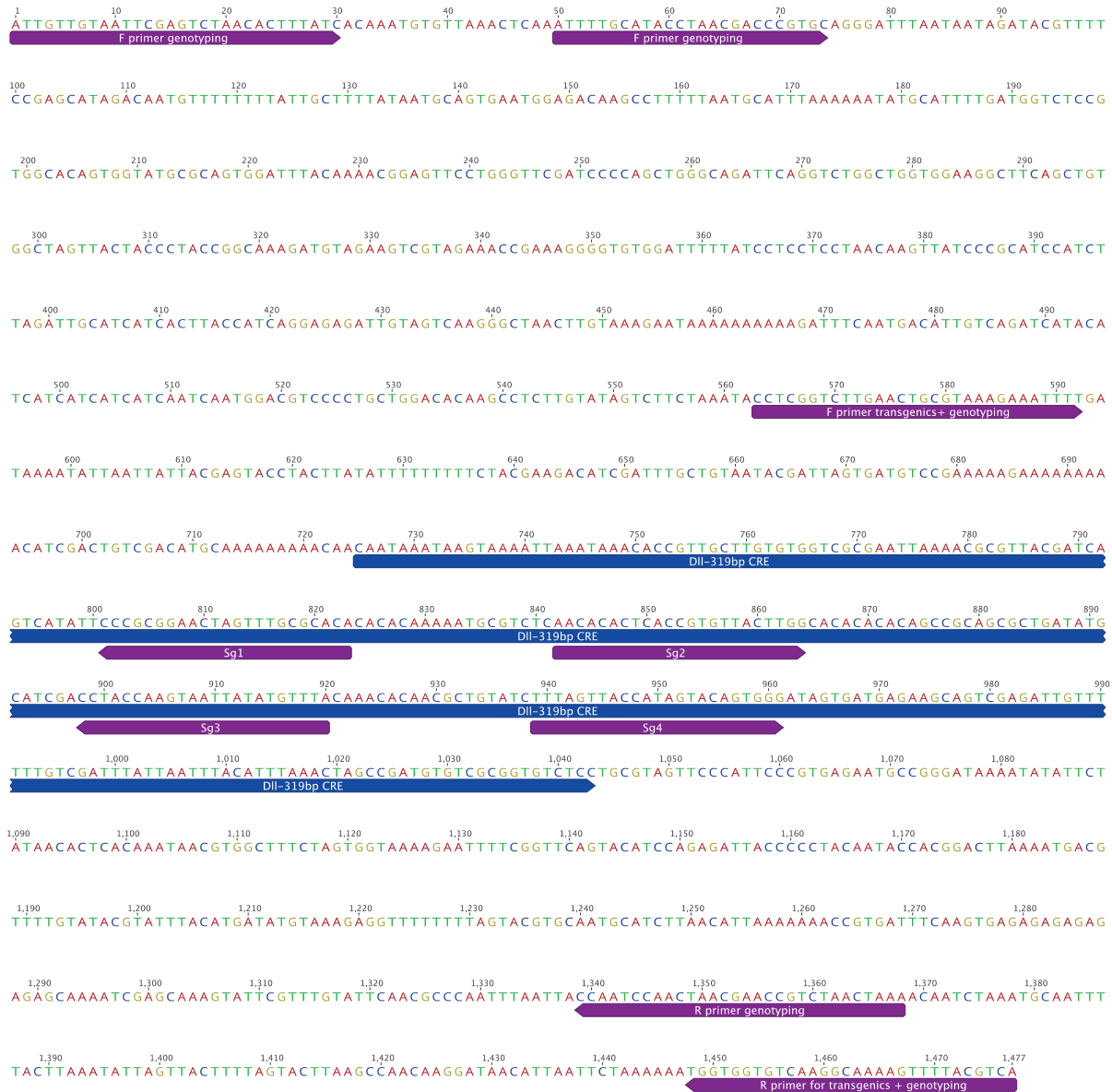


1159

1160 **Fig. S14. Expression patterns of Dll, and Antp proteins in pupal legs**

1161 (A) Expression pattern of Dll and Antp in the developing leg of a 3 -h-old pupa. Dll and Antp
1162 are co-localized in the developing pupal leg. (B) Expression patterns of Dll and Antp in the leg
1163 of a 3-days-old pupa. Dll is ubiquitously expressed in the distal region of a developing pupal
1164 leg. Antp is also expressed in the developing pupal leg. The regions within the white squares
1165 are shown at higher magnification. Scale; 100 μ m.

1166



1167

1168 **Fig. S15.** *Dll3* CRE annotated with CRISPR guides and primers used for the transgenics and

1169 genotyping crispants.

1170

1171

1172

1173

1174

1175

1176

1177

1178

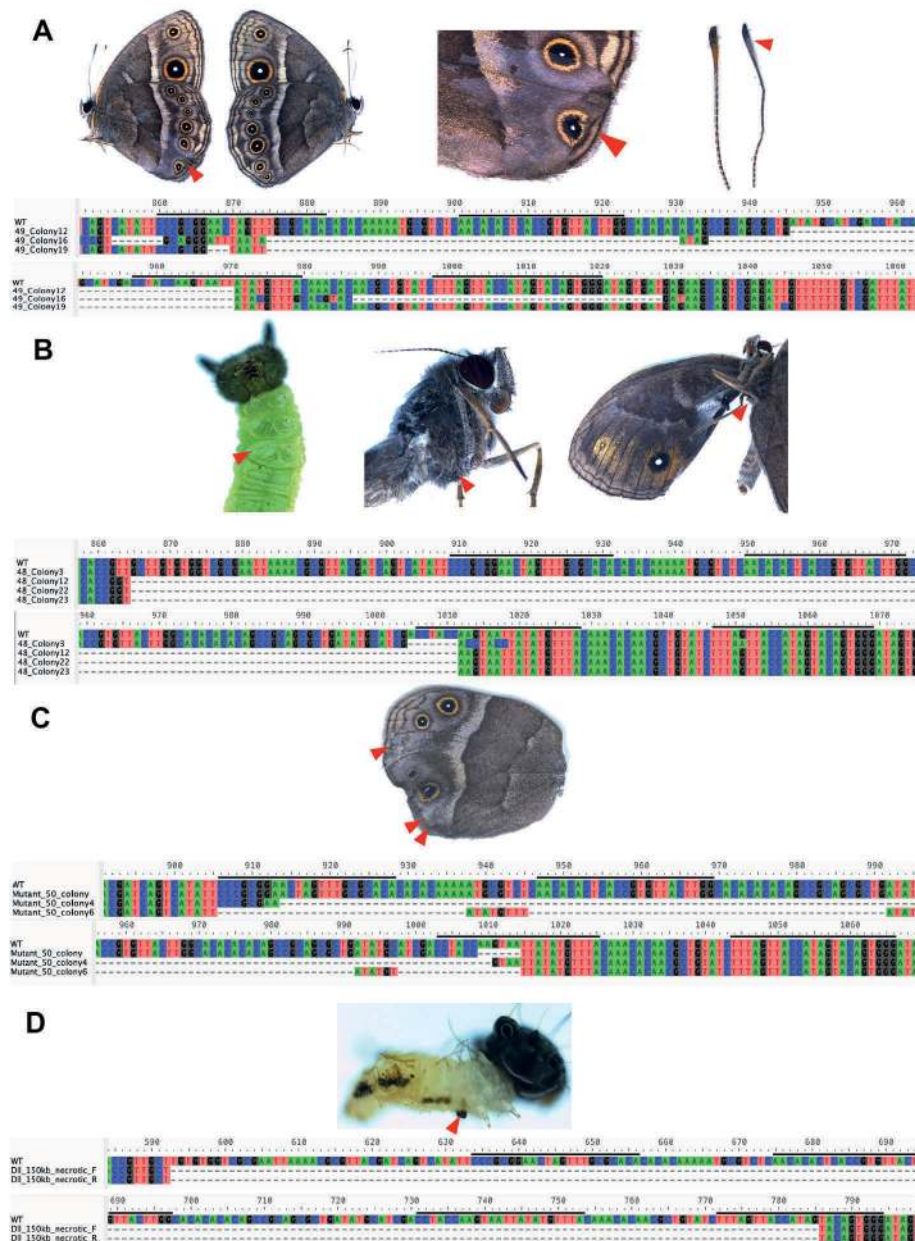
1179

1180

1181

1182

1183



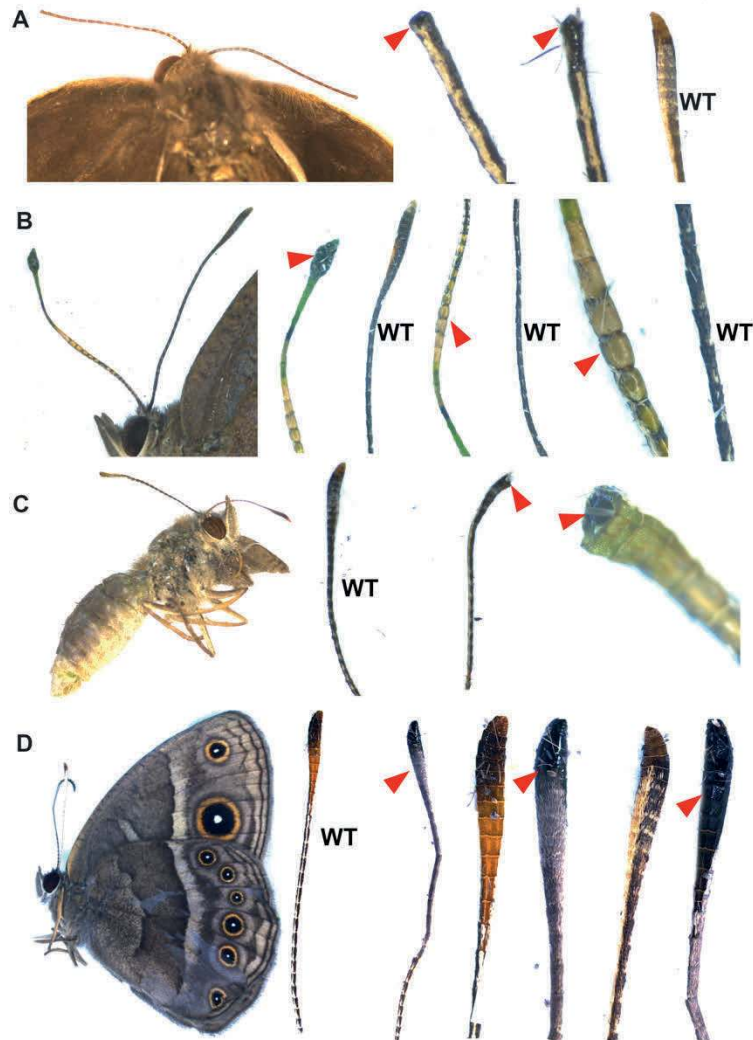
1184

1185

1186 **Fig. S16.** Sanger sequencing results following CRISPR-targeting of the *Dll319* CRE in four
1187 crisprant individuals, WT refers to the wild-type sequence. Positions of the CRISPR guides are
1188 shown as horizontal black lines above each sequence. (A) Colony PCR sequencing results with
1189 variable sized deletions for a crisprant with a missing eyespot; pigmentation defects are visible
1190 on the wing and antenna. The left image shows the wing defects. The mirror image on the right
1191 shows wild-type wing phenotype. (B) The same individual from Fig. 3C is shown with a
1192 missing T3 leg in the larva and adult, as well as a missing wing. The sequence of a colony PCR
1193 product revealed a 147 bp deletion. (C) Colony PCR sequencing result showing a 108 bp
1194 deletion for an individual with three missing eyespots. (D) Sanger sequencing showing a 193

1195 bp deletion from a whole larva that showed areas of necrosis and a missing distal tip of a T3
1196 leg.

1197
1198
1199

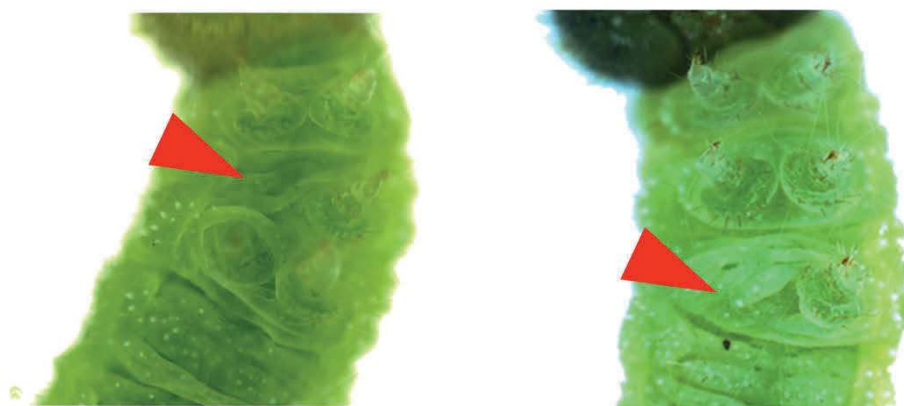


1200
1201

1202 **Fig. S17.** Four individual crispants showing antennal deformities following disruption of the
1203 *Dll319* CRE. (A) Both antennae of this butterfly were missing the distal tip. (B) One antenna
1204 showed a developmental abnormality, changing the shape of the distal tip. Additionally, the
1205 stem of the antenna had a different morphology to the wild-type and also showed loss of scales.
1206 (C) The very distal tip of one antenna in this individual was missing. (D) One antenna was
1207 crooked and showed a change in pigmentation from brown to grey. Furthermore, we noted a
1208 loss of scales on one side of the antenna that has been replaced by a shiny black cuticle, as
1209 compared to wild-type. Red arrows point to the antennae with developmental defects.

1210
1211
1212
1213
1214

1215
1216



1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245

Fig. S18. Targeting of the *Dll319* CRE led to losses of either T2 or T3 legs.



1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271

Fig. S19. A *Dll319* crispant shows white wing patches due to a complete loss of wing scales.



1272

1273

1274 **Fig. S20.** Second replicate of an EGFP-expressing embryo from the F3 transgenic generation,
1275 where the *EGFP* gene was driven by the *Dll3/19* CRE. EGFP was visualized through the use of
1276 anti-EGFP antibodies.

1277

1278

1279

1280

1281

1282

1283

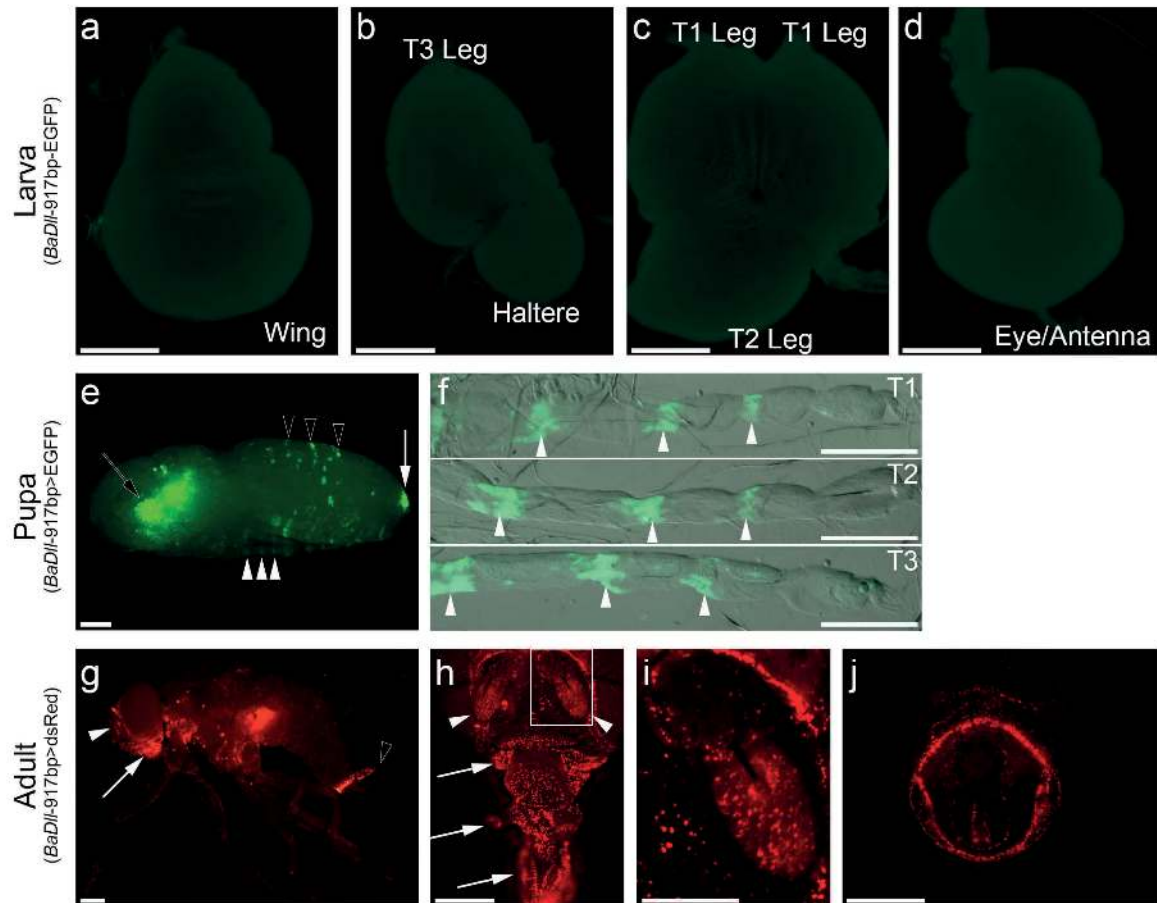
1284

1285

1286

1287

1288



1289

1290

1291

1292

1293

1294

1295

1296

1297

1298

1299

1300

1301

1302

1303

1304

1305

1306

1307

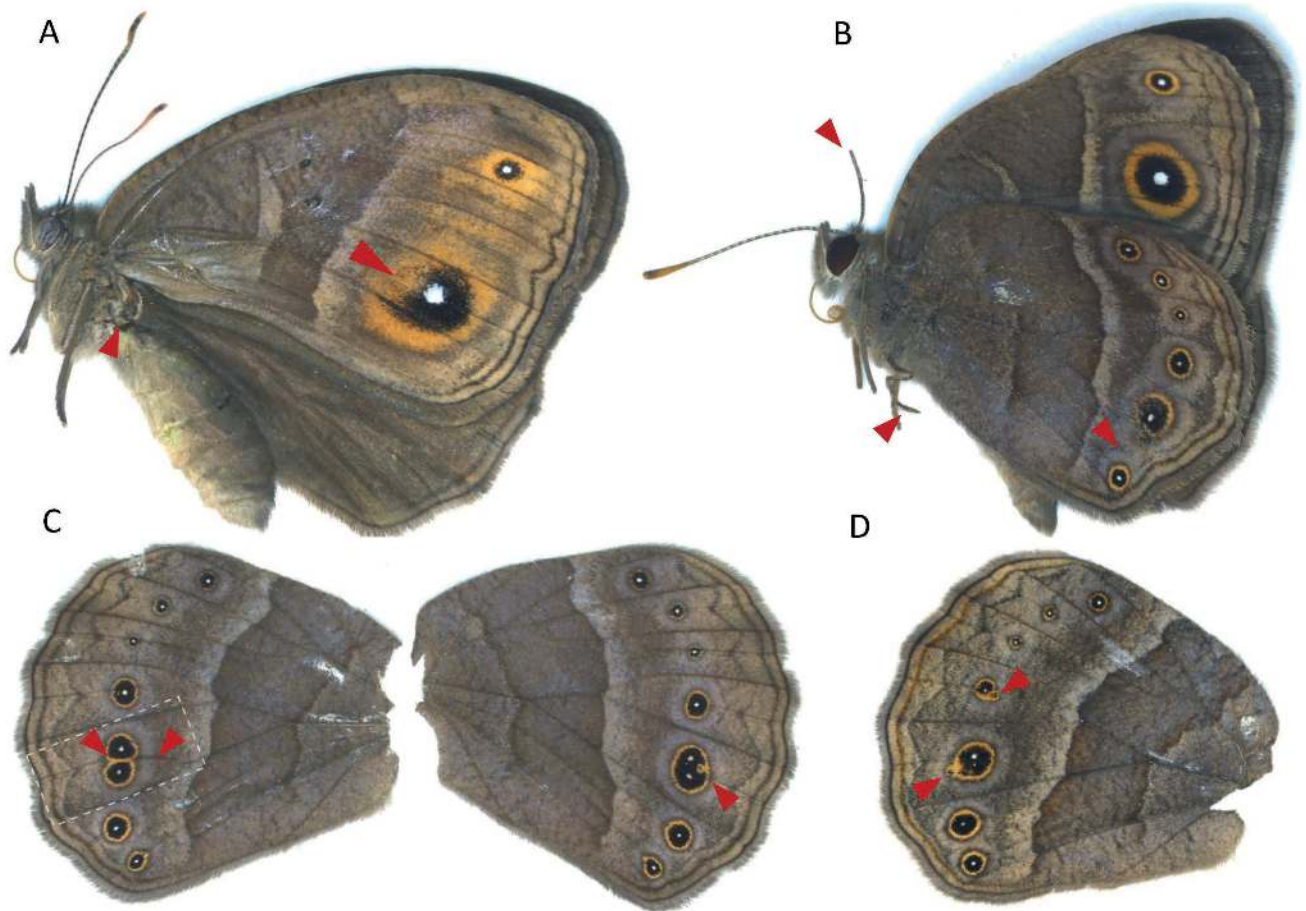
1308

1309

1310

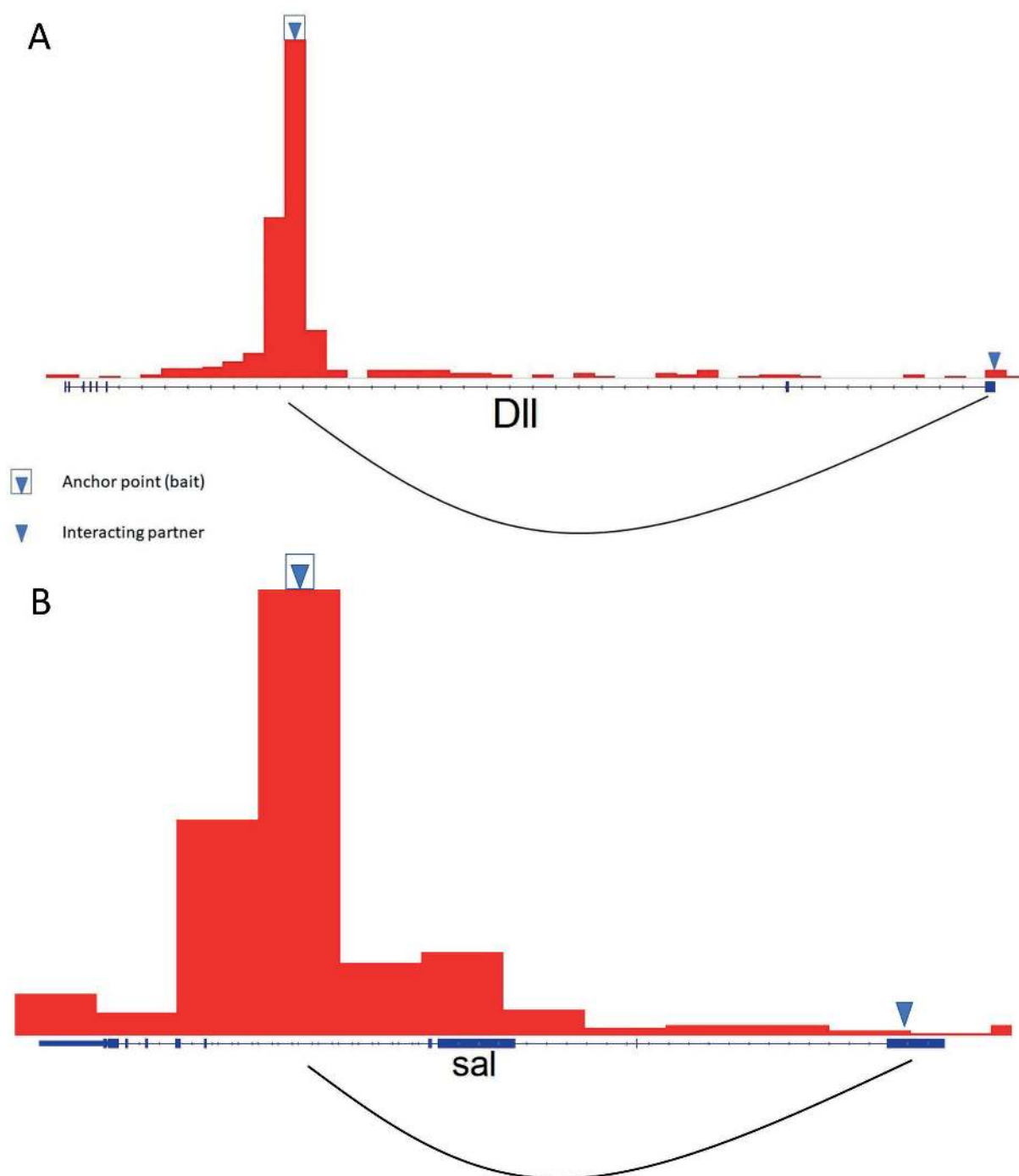
Fig. S21: Cross-species reporter assay of *Dll319* CRE in *D. melanogaster*. **a-d.** Imaginal discs of *Dll319*-EGFP flies. No noticeable enhancer activity is detected at this stage. **e,f.** *Dll319* CRE activity during the pupal stage as visualized by *Dll319*-Gal4>UAS-EGFP. **e.** Enhancer activity is detected in the pupal legs (white arrowheads), genitalia (white arrow), and in the abdomen (black arrowheads). EGFP in the eye (black arrow) is transgenic marker expression. **f.** Close-up images of the pupal legs, showing the enhancer activity in the tarsal segments (white arrowheads). **g-j.** *Dll319* CRE activity at the adult stage visualized by *Dll319*-Gal4>UAS-dsRed. **g.** *Dll319* CRE drives low ubiquitous ectodermal expression, with increased expression in the adult antennae (white arrowhead), mouthparts (white arrow), and genitalia (black arrowhead). **h.** Close-up of adult head, showing the increased activity in antennae (white arrowheads) and mouthparts (white arrows). **i.** Close-up of antenna enhancer activity within the area indicated by the white box in **h.** **j.** Close-up of enhancer activity in the adult genitalia. A lack of enhancer activity during the last larval stage may be due to the nature of this enhancer or a limitation of testing the activity of this enhancer in a cross-species setting (or a combination of both). Nonetheless, the presence of increased enhancer activity in multiple tissues, especially with these that are homologous to antennae, suggests that *Dll319* contains a pleiotropic CRE. Scale bars indicate 100 μ m (**a-d, f, i**) and 200 μ m (**e, g, h, j**).

1311
1312
1313



1314
1315 **Fig. S22.** *sal* CRE (*sal740*) crispants phenotypes. A. Loss of hindwing and black pigment in
1316 the forewing. B. Loss of eyespot in Cu2 hindwing sector, loss of distal part of the antenna and
1317 crooked T3-leg. C. Left wing: Additional vein formation in Cu1 sector of hindwing with split
1318 of Cu1 eyespot. Right wing: No ectopic vein but also split of Cu1 eyespot showing two white
1319 centres and another eyespot pigment deformity. D. Ectopic eyespot in Cu1 and M3 sector in
1320 hindwing.
1321

1322

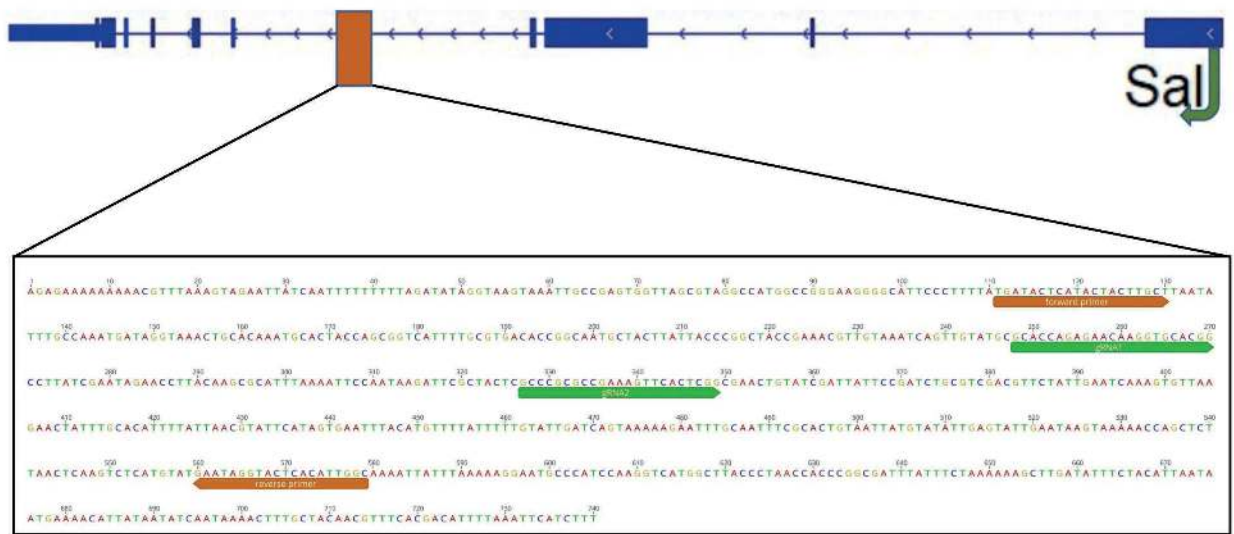


1323

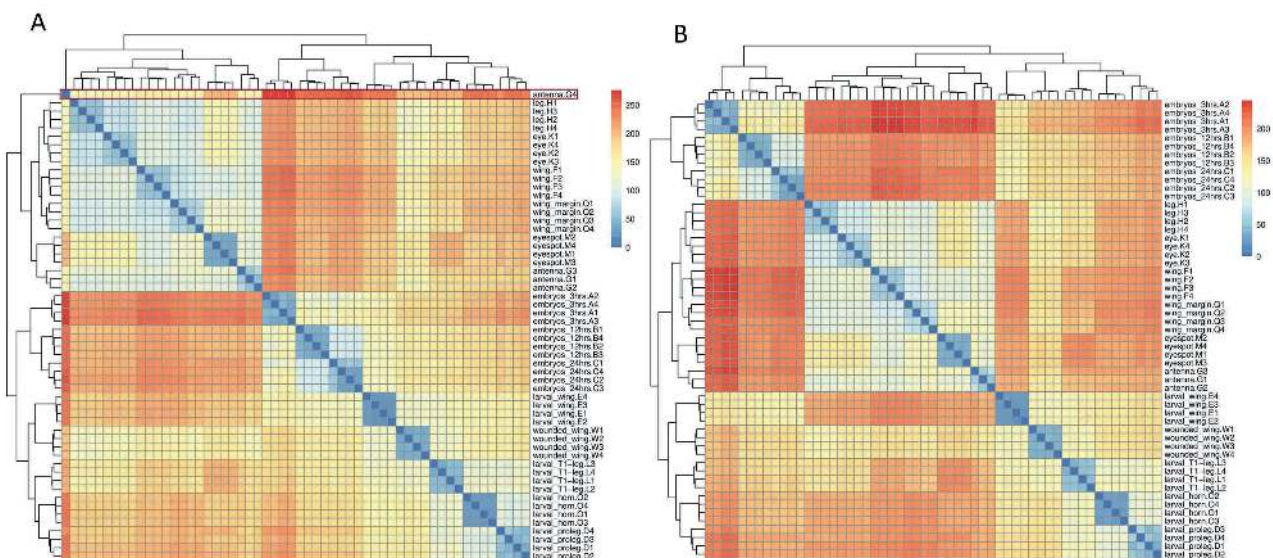
1324 **Fig. S23.** Virtual 4c plot for *Dll319* and *sal740*. (A) The graph around *Dll319* region showing

1325 its interaction with *Dll* promoter region. (B) Graph around *sal740* region shows its interaction

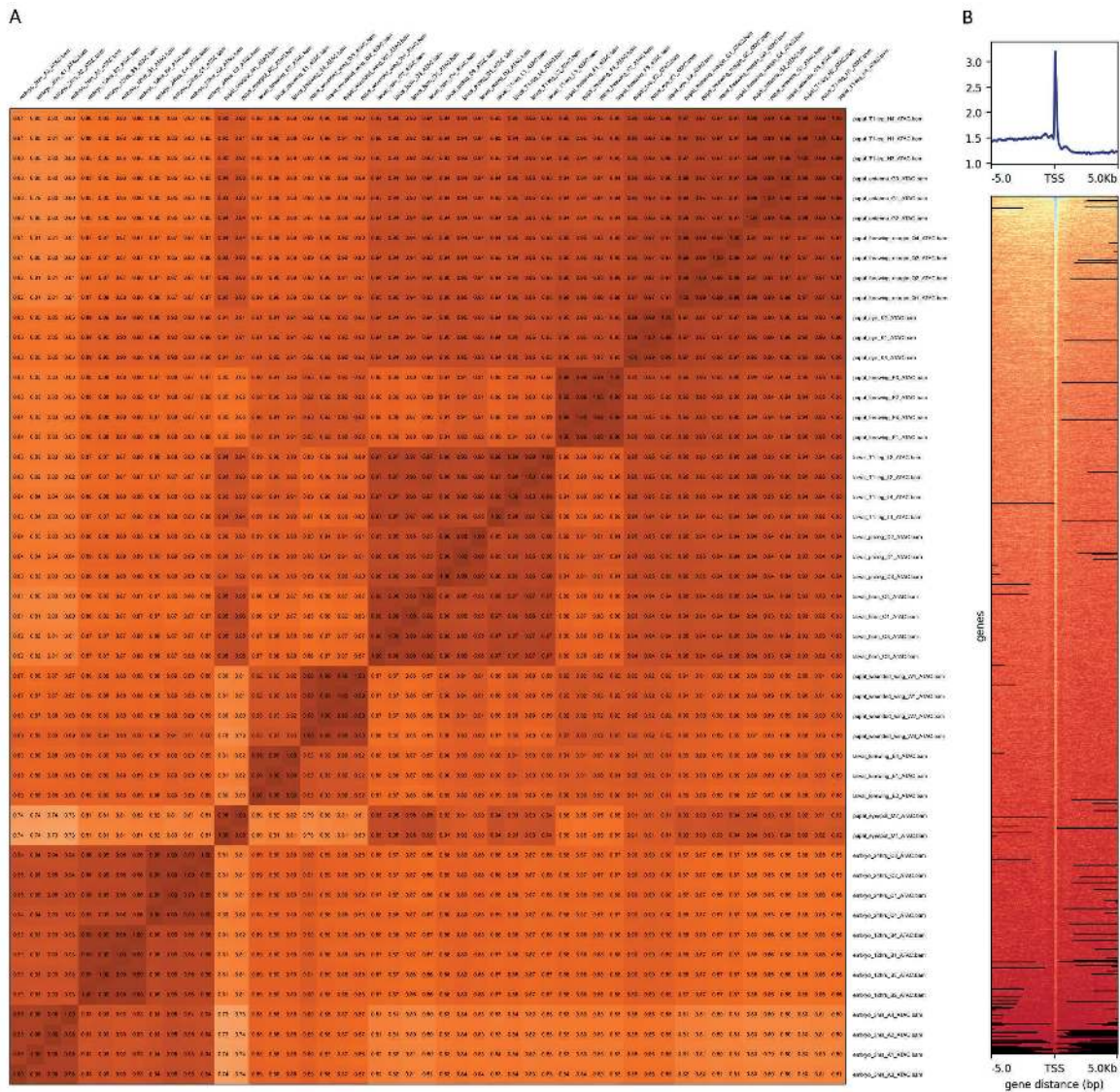
1326 with *sal* promoter region



1327
 1328 **Fig. S24.** *sal740* CRE region annotated with CRISPR guides and primers used for knockout
 1329 and genotyping
 1330



1331
 1332 **Fig. S25.** Sample clustering using an Euclidean distance matrix before and after sample
 1333 filtering. (A) Initial sample clustering shows antenna. G4 is an outlier. (B) After excluding the
 1334 outlier (antenna.G4), the replicates from each group cluster together.
 1335
 1336



1337
1338 **Fig. S26.** Quality assessment of ATAC libraries. A. Correlation matrix between the replicates
1339 in the group. B. ATAC peaks are highly enriched in Transcription start sites (TSS) and this
1340 highlights libraries of good quality.

1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353

1354

1355

1356

1357 **Supplementary Tables**

1358

1359 **Table S1.** Number of injected individuals that displayed developmental defects due to
1360 CRISPR-targeting of the *Dll319* CRE.

No. eggs injected	No. hatched	Survival 3 rd instar	Leg mutants	Missing eyespots	Deformed antennae	Missing wings	Pigment defects
366	40	35	1	0	0	0	0
646	56	51	1	2	0	1	3
422	87	76	1	2	1	0	4
797	193	186	1	0	4	0	0

1361

1362

1363 **Table S2.** Primers used for CRISPR guide RNAs and for genotyping crispants.

1364

Primer description	Sequence (5' to 3')
CRISPR guides for <i>Dll-319bp</i> CRE	
Sg1 (- strand)	TGTGCGCAAAGTTCGCGGG
Sg2	AACACACTCACCGTGTTACTTGG
Sg3 (- strand)	TAAACATATAATTACTTGGTAGG
Sg4	TTTAGTTACCATAGTACAGTGGG
Primers used for genotyping larval mutants	FP: CCTCGGTCTTGAAGTGCCTAAAGAAATTT
	RP: TTTAGTTAGACGGTTCGTTAGTTGGATTGG
Primers used for genotyping adult mutants	FP: ATTGTTGTAATTCGAGTCTAACACTTTATC
	RP: TGACGTAAACTTTTGCCTTGACACCACCA
	FP: ATTTTGCATACCTAACGACCCGTGC
	RP: CGTAAACTTTTGCCTTGACACCACC

Primers used for the transgenics	FP: CCTCGGTCTTGAAGTGGCGTAAAGAAATTTT
	RP: TGACGTAAACTTTTGCCTTGACACCACCA
CRISPR guides for <i>sal</i>	GGTGATCGAGCCGGCGTTGACGG
Primers used for genotyping <i>sal</i> crispants	FR: GCATCGACAAGA TGCTGAAA
	RP: TTCATTTAGGGACGGTGGAG
CRISPR guides for <i>sal740</i> CRE	
+ strand	GCACCAGAGAACAAGGTGCA CGG
+ strand	GCCCGCGCCGAAAGTTCACT CGG
Primers used for genotyping <i>sal740</i> CRE	FP: TGATACTCATACTACTTGCT
	RP: GCCAATGTGAGTACCTATTC

1365

1366

1367

1368

1369 **Table S3.** Number of injected individuals that displayed developmental defects due to
1370 CRISPR-targeting of *sal*.

No. eggs injected	No. hatched	Survival 3 rd instar	No. surviving until adulthood	Number of adults showing phenotype
108	N.A.	N.A.	51	11
102	N.A.	N.A.	23	5

1371

1372

1373 **Table S4.** Number of injected individuals that displayed developmental defects due to
1374 CRISPR-targeting of the *sal740* CRE
1375

No. eggs injected	No. hatched	Survival 3 rd instar	Leg mutants	Eyespot mutant	Deformed antennae	Wing mutant	Horn defect	Vein defect
174	23	17	0	2	1	0	1	1
116	27	21	1	4	2	1	0	1

1376

1377

1378 **Table S5. Tissue types and number of individuals used for each replicate in RNA-seq**

Tissue	Number of individuals
Cu1 control tissue – Nes1	20
Cu1 eyespot	20
M3 control tissue - Nes2	20
Wings (larval, pupal 3 h, wounded 24 h)	15
Embryos (3 h, 12 h, 24 h)	15
T1- legs (larval, pupal 3 h)	15
Pupal antennae (3 h)	15
Pupal eyes (3 h)	15
Pupal wing margins	15
Larval prolegs	15
Larval horns	15

1379 Dissections were performed from the right and left sides of each animal, except for wounded
 1380 wings, where only one side of each pupa was wounded.

1381

1382

1383

1384 **Table S6. RNA sequencing data. Read-depth and alignment rate**

Group	Sample	Before filtration	After filtration	Read mapped (percentage)
Embryo (4 h)	A1	113835036	98548967	84535028 (85.78)
	A2	85798982	72804246	62563452 (85.93)
	A3	84005872	71811396	61845018 (86.12)
	A4	87964122	75009490	64255225 (85.66)
Embryo (12 h)	B1	118541088	92418462	80238385 (86.82)
	B2	83149314	69051367	60021312 (86.92)
	B3	84868922	71215091	62150653 (87.27)
	B4	95915936	80838013	70347725 (87.02)
Embryo (24 h)	C1	104046272	83316043	73709053 (88.47)
	C2	87307238	75070312	65274564 (86.95)
	C3	81284220	69109463	60064998 (86.91)
	C4	94828668	79682561	69289785 (86.96)
Larval prolegs	D1	83276248	77847505	69820993 (89.69)
	D2	79714564	73485189	66512225 (90.51)

	D3	81727632	75488734	66017212 (87.45)
	D4	127697360	120323993	108186292 (89.91)
Larval forewings	E1	94157316	89083615	80816020 (90.72)
	E2	87677324	82697367	75030588 (90.73)
	E3	87813076	82573316	74758890 (90.54)
	E4	93771904	85869930	77692802 (90.48)
Pupal forewings (3 h)	F1	86993380	81635451	73605012 (90.16)
	F2	92558328	85676141	77497720 (90.45)
	F3	110561420	100513137	90731802 (90.27)
	F4	98618432	89622914	80778116 (90.13)
Pupal antennae (3 h)	G1	121763488	97360783	89616569 (92.05)
	G2	98867180	91442924	83616498 (91.44)
	G3	110612002	100686643	93855202 (93.22)
	G4	131127280	112107640	102392344 (91.33)
Pupal T1-legs (3 h)	H1	82555558	73463615	66269516 (90.21)
	H2	82664720	76726364	69433046 (90.49)
	H3	92650310	83394624	75461955 (90.49)
	H4	103884724	95341919	86365208 (90.58)
Pupal eyes (3 h)	K1	105415992	97797591	86952723 (88.91)
	K2	104493966	90166098	80474351 (89.25)
	K3	126805006	119393689	106917955 (89.55)
	K4	89001494	83262184	74215276 (89.13)
Larval T1-legs	L1	86355848	71592616	62684382 (87.56)
	L2	87998220	74129624	64661129 (87.23)
	L3	85044900	74009714	65079663 (87.93)
	L4	85073358	72756514	64015288 (87.99)
Pupal eyespots (3h)	M1	95689310	79421201	71472125 (89.99)
	M2	88548194	72248301	64941843 (89.89)
	M3	99417446	82401746	74051771 (89.87)
	M4	83145168	68910114	61942120 (89.89)
Pupal eyespot control tissue (Nes2) (3h)	N1	84383612	67629139	61068234 (90.30)
	N2	94301260	77361040	69687990 (90.08)
	N3	82445922	67265063	60665549 (90.19)
	N4	88114168	72856592	65653903 (90.11)
Larval horns	O1	99068014	85624813	75362272 (88.01)
	O2	91060940	79279017	69618673 (87.81)
	O3	82291644	71445496	62786392 (87.88)
	O4	90492250	79350706	69725988 (87.87)
Pupal eyespot control	P1	82963650	72783137	65682543 (90.24)
	P2	80007172	72558427	65495973 (90.27)
	P3	88945496	82952035	74929608 (90.33)

tissue (Nes1) (3h)	P4	87810090	82518479	73776623 (89.41)
Pupal forewing margins (3 h)	Q1	95209370	84816919	76522011 (90.22)
	Q2	92102414	83759967	75370417 (89.98)
	Q3	108138096	93532198	83304314 (89.06)
	Q4	89856780	78373499	70107705 (89.45)
Wounded pupal wings (24 h)	W1	103749366	98026397	87874345 (89.64)
	W2	103912114	90239092	81000646 (89.76)
	W3	83506698	78712471	71236861 (90.50)
	W4	93126586	87624287	79304892 (90.51)

1385

1386

1387

1388

1389

Table S7. Tissues types and numbers of individuals used for each replicate in ATAC-seq

Tissue	Number of individuals
Cu1 eyespot	25
Wings (larval, pupal 3 h, wounded 24 h)	20
Embryos (3 h, 12 h, 24 h)	15
T1- legs (larval, pupal 3 h)	20
Pupal antennae (3 h)	20
Pupal eyes (3 h)	20
Pupal wing margins	20
Larval prolegs	20
Larval horns	20

1390

Dissections were performed from the right and left sides of each animal, except for wounded wings, where only one side of each pupa was wounded.

1391

1392

1393

1394

1395

1396

Table S8: ATAC-Seq reads. Read depth and FRiP score

1397

Group	Sample	Reads mapped to genome	Reads mapped to peaks	FRiP score
Embryos (4 h)	A1	7929162	7069208	0.892
	A2	8722848	7650722	0.878
	A3	7590222	6645944	0.876

	A4	8079972	7027172	0.87
Embryos (12 h)	B1	28845464	24414192	0.847
	B2	21022574	17806982	0.848
	B3	26855374	22789314	0.849
	B4	24438812	20817352	0.852
Embryos (24 h)	C1	23947046	20470934	0.855
	C2	19317116	16720430	0.866
	C3	17459032	15060032	0.863
	C4	23499084	20482478	0.872
Larval prolegs	D1	17098004	14066208	0.823
	D2	16925178	13848684	0.819
	D3	18225832	14975016	0.822
Larval forewings	E1	22238552	19258772	0.867
	E2	22896714	20094916	0.878
	E4	22130896	19302416	0.873
Pupal forewings (3 h)	F1	32928940	28260602	0.859
	F2	25811542	22215800	0.861
	F3	30779202	26438266	0.859
	F4	30407764	26060672	0.858
Pupal antennae (3 h)	G1	39990330	33025110	0.826
	G2	26276830	21934216	0.835
	G3	35808400	29831484	0.834
Pupal T1-legs (3 h)	H1	35962956	29719496	0.827
	H3	30691458	25407180	0.828
	H4	32904694	27491442	0.836
Pupal eyes (3 h)	K1	27369070	23057554	0.843
	K2	25115380	21124918	0.842
	K3	23566544	19761176	0.839
Larval T1-legs	L1	24453864	20554604	0.841
	L2	28109100	22717706	0.809
	L3	27081264	22121078	0.817
	L4	22582662	18430048	0.817
Pupal eyespots (3 h)	M1	23591594	19145932	0.812
	M2	24294062	20029744	0.825
Larval horns	O1	21485294	17388048	0.81
	O2	20946732	16941090	0.809
	O3	19939442	16082748	0.807
	O4	27568916	21767434	0.79
Pupal forewing margins (3 h)	Q1	79547324	65788012	0.828
	Q2	31692538	26044216	0.822
	Q3	24654044	20246276	0.822
	Q4	25965642	21625304	0.833

Wounded pupal wings (24 h)	W1	30988206	26493042	0.855
	W2	36814878	31509656	0.856
	W3	49790726	43261278	0.869
	W4	27833492	23647830	0.85

1398

1399

1400

1401

1402 References

- 1403 1. W. J. Glassford, W. C. Johnson, N. R. Dall, S. J. Smith, Y. Liu, W. Boll, M. Noll, M.
1404 Rebeiz, Co-option of an Ancestral Hox-Regulated Network Underlies a Recently
1405 Evolved Morphological Novelty. *Dev. Cell.* **34**, 520–531 (2015).
- 1406 2. S. B. Carroll, J. Gates, D. N. Keys, S. W. Paddock, G. E. Panganiban, J. E. Selegue, J.
1407 a Williams, Pattern formation and eyespot determination in butterfly wings. *Science.*
1408 **265**, 109–14 (1994).
- 1409 3. S. V Saenko, V. French, P. M. Brakefield, P. Beldade, Conserved developmental
1410 processes and the formation of evolutionary novelties: examples from butterfly wings.
1411 *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 1549–55 (2008).
- 1412 4. L. I. Held, Rethinking Butterfly Eyespots. *Evol. Biol.* **40**, 158–168 (2013).
- 1413 5. A. Monteiro, G. Glaser, S. Stockslager, N. Glansdorp, D. Ramos, Comparative insights
1414 into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* **6**, 1–13 (2006).
- 1415 6. H. Connahs, S. Tlili, J. van Creijl, T. Y. J. Loo, T. Das Banerjee, T. E. Saunders, A.
1416 Monteiro, Activation of butterfly eyespots by Distal-less is consistent with a reaction-
1417 diffusion process. *Dev.* **146**, 1–12 (2019).
- 1418 7. J. M. Musser, G. P. Wagner, Character trees from transcriptome data: Origin and
1419 individuation of morphological characters and the so-called “species signal.” *J. Exp.*
1420 *Zool. Part B Mol. Dev. Evol.* **324**, 588–604 (2015).
- 1421 8. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion
1422 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
- 1423 9. R. Suzuki, H. Shimodaira, Pvcust: An R package for assessing the uncertainty in
1424 hierarchical clustering. *Bioinformatics.* **22**, 1540–1542 (2006).
- 1425 10. N. Özsu, A. Monteiro, Wound healing, calcium signaling, and other novel pathways
1426 are associated with the formation of butterfly eyespots. *BMC Genomics.* **18** (2017),
1427 doi:10.1186/s12864-017-4175-7.
- 1428 11. J. D. Uhl, A. Zandvakili, B. Gebelein, A Hox Transcription Factor Collective Binds a
1429 Highly Conserved Distal-less cis-Regulatory Module to Generate Robust
1430 Transcriptional Outcomes. *PLoS Genet.* **12**, 1–26 (2016).
- 1431 12. J. T. Wagner-Bernholz, O. Wilson, G. Gibson, R. Schuh, W. J. Gehring, Identification
1432 of target genes of the homeotic gene Antennapedia by enhancer detection (Genes and
1433 Development 5 (2467-2480)). *Genes Dev.* **6**, 328 (1992).
- 1434 13. P. D. Si Dong, J. Chu, G. Panganiban, Coexpression of the homeobox genes Distal-
1435 less and homothorax determines Drosophila antennal identity. *Development.* **127**, 209–
1436 216 (2000).

- 1437 14. J. C. Oliver, X. L. Tong, L. F. Gall, W. H. Piel, A. Monteiro, A Single Origin for
1438 Nymphalid Butterfly Eyespots Followed by Widespread Loss of Associated Gene
1439 Expression. *PLoS Genet.* **8** (2012), doi:10.1371/journal.pgen.1002893.
- 1440 15. Y. Matsuoka, A. Monteiro, Hox genes are essential for the development of novel serial
1441 homologous eyespots on the wings of *Bicyclus anynana* butterflies. *Gentics*, iyaa005
1442 (2020).
- 1443 16. T. Das Banerjee, A. Monteiro, Molecular mechanisms underlying simplification of
1444 venation patterns in holometabolous insects. *Dev.* **dev.196394** (2020),
1445 doi:10.1242/dev.196394.
- 1446 17. L. Zhang, R. D. Reed, Genome editing in butterflies reveals that spalt promotes and
1447 Distal-less represses eyespot colour patterns. *Nat. Commun.* **7** (2016),
1448 doi:10.1038/ncomms11769.
- 1449 18. C. R. Brunetti, J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield, S. B. Carroll,
1450 The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**,
1451 1578–1585 (2001).
- 1452 19. S. V. Saenko, M. S. P. Marialva, P. Beldade, Involvement of the conserved Hox gene
1453 Antennapedia in the development and evolution of a novel trait. *Evodevo.* **2**, 9 (2011).
- 1454 20. G. Panganiban, Distal-less Function During *Drosophila* Appendage and Sense Organ
1455 Development. *Dev. Dyn.* **562**, 554–562 (2000).
- 1456 21. B. S. Emerald, S. M. Cohen, Spatial and temporal regulation of the homeotic selector
1457 gene Antennapedia is required for the establishment of leg identity in *Drosophila*. *Dev.*
1458 *Biol.* **267**, 462–472 (2004).
- 1459 22. Y. T. Lai, K. D. Deem, F. Borràs-Castells, N. Sambrani, H. Rudolf, K. Suryamohan, E.
1460 El-Sherif, M. S. Halfon, D. J. McKay, Y. Tomoyasu, Enhancer identification and
1461 activity evaluation in the red flour beetle, *Tribolium castaneum*. *Dev.* **145** (2018),
1462 doi:10.1242/dev.160663.
- 1463 23. G. Panganiban, J. L. R. Rubenstein, Developmental functions of the Distal-less/Dlx
1464 homeobox genes. *Development.* **129**, 4371–86 (2002).
- 1465 24. H. S. Bruce, N. H. Patel, *Knockout of crustacean leg patterning genes suggests that*
1466 *insect wings and body walls evolved from ancient leg segments* (2020), vol. 4.
- 1467 25. C. M. Clark-Hachtel, Y. Tomoyasu, Two sets of candidate crustacean wing
1468 homologues and their implication for the origin of insect wings. *Nat. Ecol. Evol.* **4**,
1469 1694–1702 (2020).
- 1470 26. G. Sabarís, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with Multiple Roles:
1471 Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**,
1472 423–433 (2019).
- 1473 27. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory
1474 evolution. *Light Evol.* **1**, 109–127 (2007).
- 1475 28. A. Monteiro, O. Podlaha, Wings, horns, and butterfly eyespots: how do complex traits
1476 evolve? *PLoS Biol.* **7**, e37 (2009).
- 1477 29. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler
1478 transform. *Bioinformatics.* **25**, 1754–1760 (2009).
- 1479 30. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G.
1480 Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools.

- 1481 *Bioinformatics*. **25**, 2078–2079 (2009).
- 1482 31. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing
1483 genomic features. *Bioinformatics*. **26**, 841–842 (2010).
- 1484 32. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C.
1485 Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Shirley, Model-based analysis of
1486 ChIP-Seq (MACS). *Genome Biol.* **9** (2008), doi:10.1186/gb-2008-9-9-r137.
- 1487 33. Y. Naito, K. Hino, H. Bono, K. Ui-Tei, CRISPRdirect: Software for designing
1488 CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*. **31**, 1120–1123
1489 (2015).
- 1490 34. K. N. Eckermann, H. M. M. Ahmed, M. KaramiNejadRanjbar, S. Dippel, C. E.
1491 Ogaugwu, P. Kitzmann, M. D. Isah, E. A. Wimmer, Hyperactive piggyBac transposase
1492 improves transformation efficiency in diverse insect species. *Insect Biochem. Mol.*
1493 *Biol.* **98**, 16–24 (2018).
- 1494 35. C. J. Evans, J. M. Olson, K. T. Ngo, E. Kim, N. E. Lee, E. Kuoy, A. N. Patananan, D.
1495 Sitz, P. T. Tran, M. T. Do, K. Yackle, A. Cespedes, V. Hartenstein, G. B. Call, U.
1496 Banerjee, G-TRACE: Rapid Gal4-based cell lineage analysis in *Drosophila*. *Nat.*
1497 *Methods*. **6**, 603–605 (2009).
- 1498 36. B. Bushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner. *United States N. p*
1499 (2014), (available at [https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-](https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner)
1500 [aware-aligner](https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner)).
- 1501 37. E. Kopylova, L. Noé, H. Touzet, SortMeRNA: Fast and accurate filtering of ribosomal
1502 RNAs in metatranscriptomic data. *Bioinformatics*. **28**, 3211–3217 (2012).
- 1503 38. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level
1504 expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.
1505 *Nat. Protoc.* **11**, 1650–1667 (2016).
- 1506 39. M. S. Campbel, C. Holt, B. Moore, M. Yandell, *Genome Annotation and Curation*
1507 *Using MAKER and MAKER-P* (2008), vol. 48.
- 1508 40. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, D. Philip, J. Bowden, M. B.
1509 Couger, D. Eccles, B. Li, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N.
1510 Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N.
1511 Friedman, A. Regev, *De novo transcript sequence reconstruction from RNA-Seq:*
1512 *reference generation and analysis with Trinity* (2013), vol. 8.
- 1513 41. C. R. Fisher, J. L. Wegrzyn, E. L. Jockusch, Co-option of wing-patterning genes
1514 underlies the evolution of the treehopper helmet. *Nat. Ecol. Evol.* **4**, 250–260 (2020).
- 1515 42. R Core Team, R: A language and environment for statistical computing (2020).
- 1516 43. M. R. Corces, A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-
1517 Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria,
1518 S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A.
1519 Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf, H. Y. Chang, An improved
1520 ATAC-seq protocol reduces background and enables interrogation of frozen tissues.
1521 *Nat. Methods*. **14**, 959–962 (2017).
- 1522 44. A. P. Boyle, J. Guinney, G. E. Crawford, T. S. Furey, F-Seq: A feature density
1523 estimator for high-throughput sequence tags. *Bioinformatics*. **24**, 2537–2538 (2008).
- 1524 45. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program

- 1525 for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).
- 1526 46. F. Ramírez, F. Dünder, S. Diehl, B. A. Grüning, T. Manke, DeepTools: A flexible
1527 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).
- 1528 47. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E.
1529 L. Aiden, A. Mathematics, Juicer provides a one-click system for analyzing loop-
1530 resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2018).
- 1531 48. J. Ray, P. R. Munn, A. Vihervaara, J. J. Lewis, A. Ozer, C. G. Danko, J. T. Lis,
1532 Chromatin conformation remains stable upon extensive transcriptional changes driven
1533 by heat shock. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19431–19439 (2019).
- 1534 49. R. W. Nowell, B. Elsworth, V. Oostra, B. J. Zwaan, C. W. Wheat, M. Saastamoinen, I.
1535 J. Saccheri, A. E. van't Hof, B. R. Wasik, H. Connahs, M. L. Aslam, S. Kumar, R. J.
1536 Challis, A. Monteiro, P. M. Brakefield, M. Blaxter, A high-coverage draft genome of
1537 the mycalesine butterfly *Bicyclus anynana*. *Gigascience*. **6**, 1–7 (2017).
- 1538 50. P. Beldade, S. V. Saenko, N. Pul, A. D. Long, A gene-based linkage map for *Bicyclus*
1539 *anynana* butterflies allows for a comprehensive analysis of synteny with the
1540 lepidopteran reference genome. *PLoS Genet.* **5** (2009),
1541 doi:10.1371/journal.pgen.1000366.
- 1542 51. J. Catchen, A. Amores, S. Bassham, *G3: Genes|Genomes|Genetics*, in
1543 press, doi:10.1534/g3.120.401485.
- 1544 52. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov,
1545 BUSCO: Assessing genome assembly and annotation completeness with single-copy
1546 orthologs. *Bioinformatics*. **31**, 3210–3212 (2015).
- 1547 53. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J.
1548 Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M.
1549 Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale
1550 protein function classification. *Bioinformatics*. **30**, 1236–1240 (2014).
- 1551 54. J. Dainat, AGAT: Another Gff Analysis Toolkit to handle annotations in any
1552 GTF/GFF format.(Version v0.4.0) (2020), doi:10.5281/ZENODO.4205393.
- 1553 55. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using
1554 DIAMOND. *Nat. Methods*. **12**, 59–60 (2014).
- 1555 56. S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M.
1556 Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and
1557 data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
- 1558
- 1559