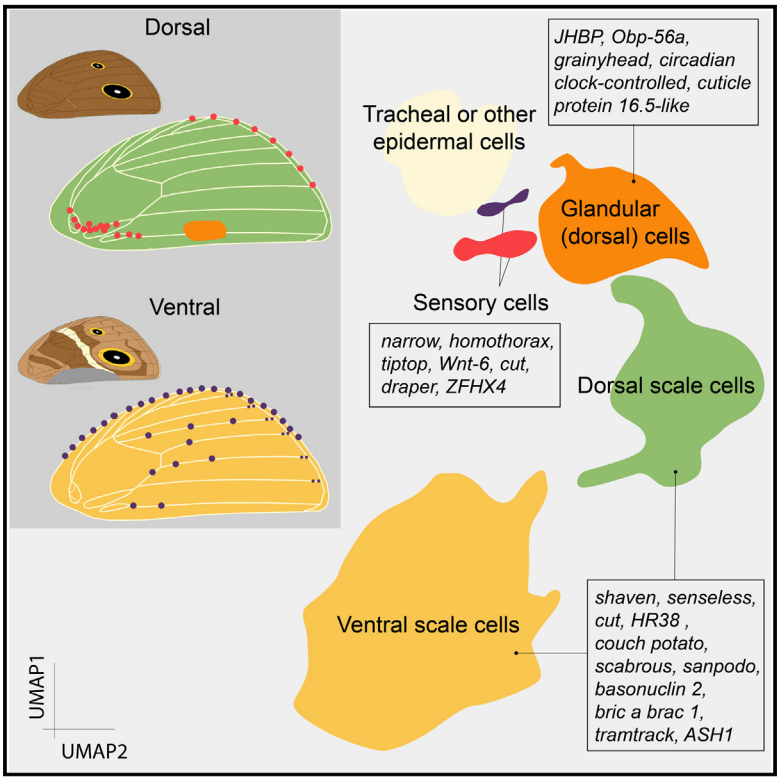
Cell Reports

Resource

The molecular basis of scale development highlighted by a single-cell atlas of *Bicyclus anynana* butterfly pupal forewings

Graphical abstract



Highlights

- *senseless* is important for both scale cell specification and differentiation
- *HR38* is necessary for the development of hair-like scales and regulates scale color and size
- *cut* mRNA expression is punctate and protein expression is non-homogenous in scale cells
- Identified marker genes for wing cell types lay a foundation for future comparative work

Authors

Anupama Prakash, Emilie Dion, Tirtha Das Banerjee, Antónia Monteiro

Correspondence

anupama@u.nus.edu (A.P.), antonia.monteiro@nus.edu.sg (A.M.)

In brief

Butterfly wing tissues are composed of various cell types including differently shaped and colored scales, sensory and glandular cells, and epidermal cells. Prakash et al. use single-cell sequencing to identify markers of different cell types and show that *HR38* and *senseless* regulate scale cell differentiation.

Prakash et al., 2024, Cell Reports *43*, 114147
May 28, 2024
https://doi.org/10.1016/j.celrep.2024.114147

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

The molecular basis of scale development highlighted by a single-cell atlas of *Bicyclus anynana* butterfly pupal forewings

Anupama Prakash,[1,2,*] Emilie Dion,[1] Tirtha Das Banerjee,[1] and Antónia Monteiro[1,3,*]
[1]Department of Biological Sciences, National University of Singapore, Singapore, Singapore
[2]Present address: School of Biosciences, University of Sheffield, Sheffield, United Kingdom
[3]Lead contact
*Correspondence: anupama@u.nus.edu (A.P.), antonia.monteiro@nus.edu.sg (A.M.)
https://doi.org/10.1016/j.celrep.2024.114147

SUMMARY

Butterfly wings display a diversity of cell types, including large polyploid scale cells, yet the molecular basis of such diversity is poorly understood. To explore scale cell diversity at a transcriptomic level, we employ single-cell RNA sequencing of ~5,200 large cells (>6 μm) from 22.5- to 25-h male pupal forewings of the butterfly *Bicyclus anynana*. Using unsupervised clustering, followed by *in situ* hybridization, immunofluorescence, and CRISPR-Cas9 editing of candidate genes, we annotate various cell types on the wing. We identify genes marking non-innervated scale cells, pheromone-producing glandular cells, and innervated sensory cell types. We show that *senseless*, a zinc-finger transcription factor, and *HR38*, a hormone receptor, determine the identity, size, and color of different scale cell types and are important regulators of scale cell differentiation. This dataset and the identification of various wing cell-type markers provide a foundation to compare and explore scale cell-type diversification across arthropod species.

INTRODUCTION

Butterfly wings contain exquisite natural mosaics of individual scale cells that exhibit incredible variation in size, shape, morphology, color, and function. Wing scales are homologous to sensory bristles found in other arthropods[1,2] but, in butterflies, these bristles differentiate into paddle-shaped, non-innervated projections, organized in neat rows of alternating larger cover scales overlapping smaller ground scales. Most scales that cover the wing membrane die upon adult emergence leaving behind their colored chitinous skeletons to serve a visual-signal function.[3–5] While scale cell color diversity is the most prominent feature of butterfly wings, wings also display a variety of other cell types. Wing epidermal cells (~3–4 μm in diameter) form most of the wing cells. Larger cells include the scale, muscle, glandular, and innervated sensory cells. Innervated sensory cells, often placed on top of veins,[6–8] are involved in mechanosensory or chemosensory functions[9–11] and glandular secretory cells are responsible for producing and secreting sex pheromones, mostly in males.[12] How such wing cell-type diversity is genetically and developmentally regulated remains largely unknown.

Bristle developmental progression has been well studied in the fruit fly, *Drosophila melanogaster*[13–18] (Figure 1A). In flies, each sensory bristle consists of four different cell types—the external socket and bristle cells and the internal sheath and neuron cells—that arise from a single sensory organ precursor (SOP) cell.[14] The SOP is selected from a proneural cluster of cells by the action of various proneural genes of the achaete-scute complex (AS-C) and it promotes an epidermal cell fate in neighboring cells by repressing AS-C genes through a Notch-mediated lateral inhibition process.[14–16,19] The SOP then undergoes two rounds of asymmetric cell divisions to produce four daughter cells (Figure 1A).[17] The first division leads to two daughter cells, pIIa and pIIb, where pIIa divides to form the future socket and bristle cells, and pIIb divides into the neuron and sheath cells. At each cell division, binary cell fate choices are regulated through Notch signaling, initiated by the asymmetric distribution of cell fate determinants such as *numb* and *neuralized* (*neur*).[20–22] In the first division, numb and neur are asymmetrically segregated to the pIIb cell, which acts as the signal sending cell, activating Notch in the pIIa cell. In the second set of divisions, the socket and sheath cells exhibit active Notch signaling (higher levels of Notch and its downstream targets), initiated by the bristle and neuron cell, respectively (Figure 1A).

The primary stages of lepidopteran scale development parallel the sensory bristle program, with certain modifications. Scales derive from a single SOP cell that is specified around 7% of pupal development (PD) (12–15 h after pupation).[1,23] As in flies, SOP cell spacing and organization is determined by a Notch-mediated lateral inhibition mechanism.[24] Unlike sensory bristle development, however, after the first cell division, one of the daughter cells (pIIb) dies (at ~17 h after pupation), eliminating the neuron and sheath progeny cells.[1] By ~14% PD (~24 h after pupation), neat rows of scale cell precursors (pIIa cells) are seen. In *Junonia coenia* butterflies, these pIIa cells express a butterfly homolog of

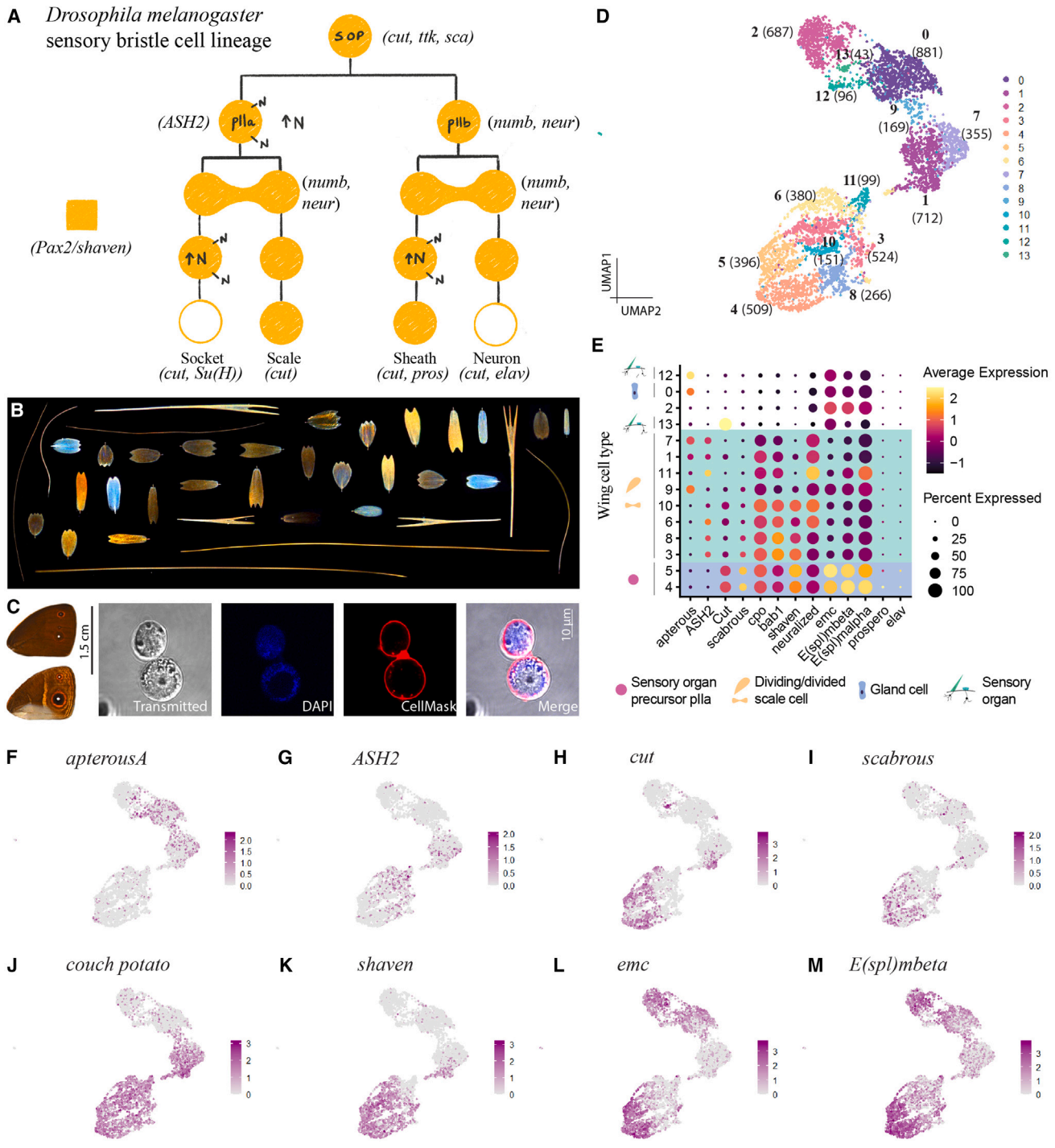


Figure 1. Scale cell diversity and a single-cell atlas of 24-h male pupal forewings of *B. anynana*

(A) A schematic of the sensory bristle cell lineage of *Drosophila melanogaster*. Genes expressed at different stages along the cell lineage are indicated in parentheses. *D. melanogaster Pax2/shaven* is expressed in the SOP, pIIa, and pIIb cells (filled circles) but, in the following cell divisions, its expression is restricted to the scale and sheath cells only (open circles). N, Notch receptor and increased Notch signaling.

(B) Diversity in scale color, morphology, and size seen on the forewings of male *B. anynana* butterflies.

(C) Male dorsal (top) and ventral (bottom) forewings of *B. anynana* and isolated single cells from a 24-h pupal forewing stained with DAPI and CellMask Plasma Membrane stain.

(legend continued on next page)

the AS-C genes, *ASH1* (*achaete-scute homolog 1*),[1] which is essential for scale development.[2] Division of the pIIa cell gives rise to a scale and socket cell (Figure S1A).[23] Alternating cover and ground scales can be distinguished based on size by ~21% PD.[23]

Given the established homology of sensory bristles and butterfly scales, and an understanding of the initial stages of scale development, we sought to explore how scales varying in morphology, color, and size (Figure 1B) vary in their differentiation program. In addition, we aimed to identify and characterize other large wing cell types beyond scale cells. We employed single-cell transcriptomics to collect gene expression information from individual large cells of developing pupal forewings of *Bicyclus anynana* butterflies. We obtained 14 biologically meaningful cell populations using unsupervised clustering of ~5,200 large cells and annotated the scale cell clusters using known marker genes from the extensive *Drosophila* sensory bristle literature. We further annotated epidermal and sensory cell types on the wings, like the marginal mechanosensory bristles and the glandular cells that produce male sex pheromones. We also uncovered important roles for the genes *senseless* (*sens*) and *hormone receptor 38* (*HR38*) in regulating characteristics and identities of scale cell types. With this dataset and functional characterization of marker genes, we provide a foundational wing cell atlas and candidate genes that can be used to further explore the molecular mechanisms of scale cell-type diversification in detail.

RESULTS

A single-cell atlas of 24-h male pupal forewings of *B. anynana* and the identification of scale cell clusters

We applied single-cell transcriptomics to 22.5- to 25-h male pupal forewings of *B. anynana*, which corresponds roughly to 14%–16% of PD (Figure 1C). We dissociated pupal wing tissues to obtain single cells and sorted these cells based on size and viability (Figure 1C). We collected cells larger than ~6 μm in diameter to enrich for cell types we were most interested in.[25] At ~24 h, cells of this size or larger correspond to the pIIa cells, their progeny,[1] and other large cells on the wing, based on confocal images of wings from that time point stained with DAPI and CellMask Plasma Membrane stain (Figures 1C and S1B). Cells were processed via the 10X Genomics microfluidics system (10X Genomics Single Cell 3′ mRNAseq v.3) to generate the single-cell libraries. After quality control, 5,268 high-quality single-cell transcriptomes were processed downstream (see STAR Methods) to obtain a single-cell atlas consisting of 14 cell clusters that we deemed to provide a biologically meaningful resolution of the large wing cell types (Figure 1D; Table 1).

Within the wing cell atlas, we could first distinguish between dorsal and ventral cells based on the expression of the dorsal marker gene *apterous A* (*apA*) (Figures 1E and 1F).[26] Around 42% of the cells analyzed, i.e., 2,213 cells belonging to clusters 0, 1, 7, 9, and 12 had high levels of expression of *apA* (Figures 1E and 1F).

We next investigated the molecular identity of the scale cells and their diversity, and characterized the other cell types on the wing, including glandular and innervated cells. We deduced that cells in clusters 1 and 3–11 are future scale cells at different time points in their differentiation from a SOP (pIIa) to a differentiated scale and socket cell (Figures 1A, 1D, and 1E). These clusters expressed a neural cell-type marker, *couch potato* (*cpo*)[27] (Figure 1J), and *bric a brac 1* (*bab1*) (Figure S2), a gene known to mark both cover and ground scales in butterflies.[28] Clusters 4 and 5 are likely pIIa cells at earlier time points along the scale cell lineage, while cells in clusters 1, 3, and 6–11 are dividing/divided scale cells. The early pIIa cells expressed neural markers and high levels of *extramacrochaetae* (*emc*) and *enhancer of split mbeta* (*E(spl)mbeta*) which are downstream targets of Notch signaling (Figures 1E, horizontal blue box, 1L, and 1M).[15] In contrast, cell clusters 1, 3, and 6–11 showed reduced levels of Notch signaling alongside expression of some neural markers, suggesting that they are pIIa cells at later stages of the scale cell lineage or the progeny of the pIIa cell (Figure 1E, horizontal green box). This is supported by the higher levels of *neur* expression in cells of clusters 1, 6, 7, 10, and 11 (Figure 1E). *neur* increases over time in the pIIa cell and is asymmetrically distributed into a single progeny, the scale cell, in the following cell division. None of the cell types expressed pIIb markers *elav* and *prospero*, corresponding with the death of pIIb cells during lepidopteran scale cell development (Figure 1E).[29,30]

Cells in clusters 1 and 3–11 also expressed genes such as *cut*, *scabrous* (*sca*), *tramtrack* (*ttk*), and *sanpodo* (*spdo*) (Figures 1E, 1H, 1I, and S2), which play important roles in *Drosophila* sensory organ development.[31–35] *spdo* had a broad expression across clusters 1 and 3–11, similar to *cpo* and *bab1* (Figure S2), while a large number of cells, predominantly in clusters 4 and 5, expressed high levels of *cut*, *sca*, *ttk*, and a zinc finger transcription factor *basonuclin 2*, related to *Drosophila disco*[36] (Figures 1E, 1H, 1I, and S2). *cut*, *sca*, and *ttk* expression overlapped extensively with a putatively annotated *Ba-Pax5* gene, which is a homolog of *Drosophila Pax2/shaven* (Figure 1K), a gene essential for bristle development in *Drosophila*.[37] Cells in a subset of the above clusters (1, 3, 6, 7, 8, and 11) also expressed *BaASH2* (Figure 1G), whose homolog has a functional role in scale development of *Bombyx mori* (Figure S3).[2]

Cells in different clusters exhibit differences in cellular identity and activities

We next sought to identify groups of significantly co-expressed genes (gene expression programs or GEPs), akin to different "cellular processes," and infer their relative contributions to each of the previously identified cell clusters.[38] A GEP exclusively used in one cell type would indicate an identity GEP, while GEPs expressed in multiple cell types would be indicative of

(D) A Uniform Manifold Approximation and Projection (UMAP) plot of 5,268 large pupal forewing cells separated into 14 distinct clusters. The number of cells in each cluster is shown in brackets.

(E) Dotplot of different marker gene expressions within each cluster. Putative higher-order cell classes are indicated by the schematics. The horizontal blue and green boxes highlight the putative scale precursor cells and the dividing/divided scale cells, respectively.

(F–M) Expression of various genes used to annotate dorsal vs. ventral cells and identify scale and scale cell precursors. See also Figures S1–S3.

Table 1. Cell clusters from the *B. anynana* 24-h pupal wing single-cell atlas annotated based on expression of various marker genes and functional validation using *in situ* hybridization, immunofluorescence, and CRISPR-Cas9 gene editing

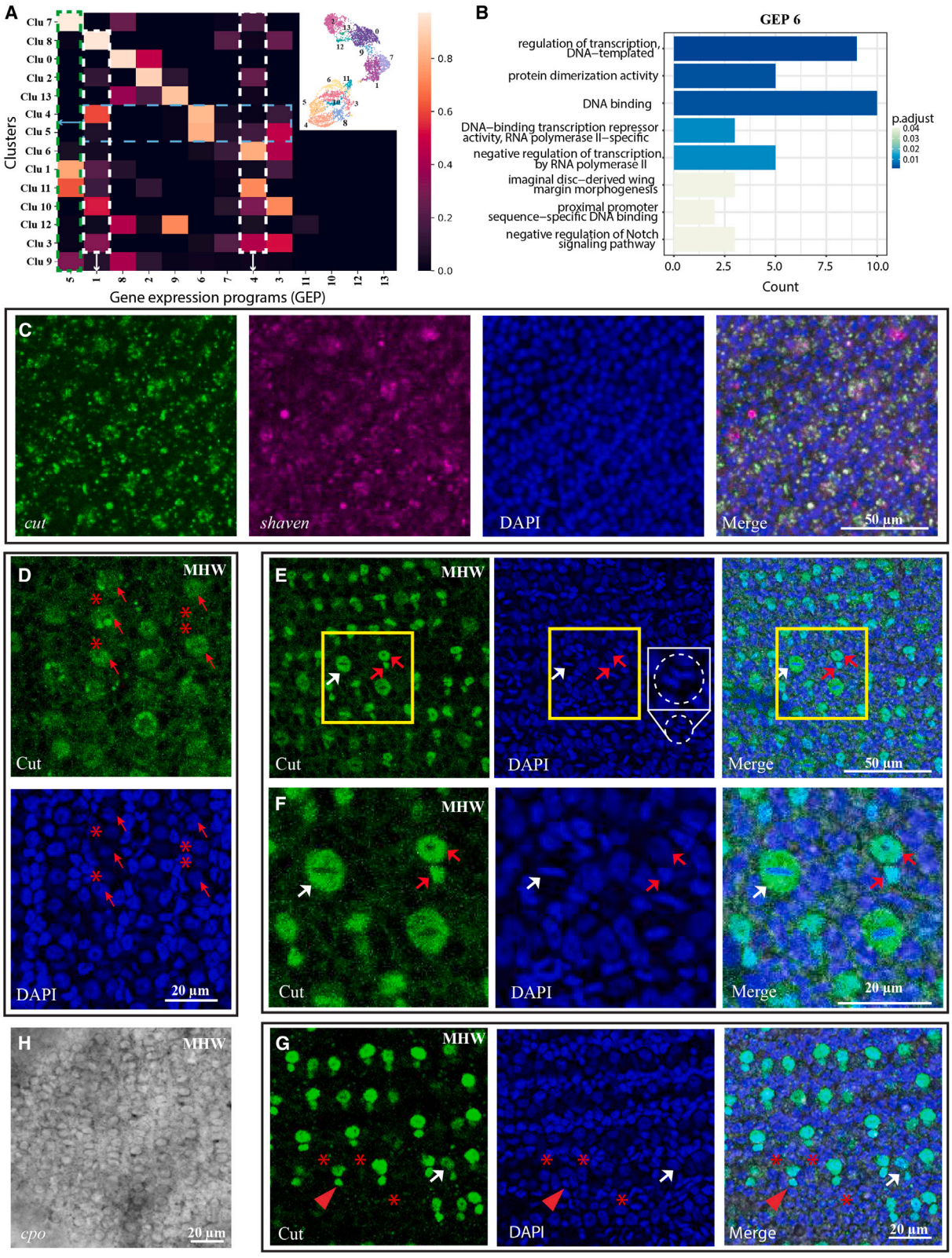
Cluster number	Cluster annotation	Dorsal/ventral	Key marker genes
0	pheromone-producing glandular cell	dorsal (*apterous A*+)	*JHBP*, *Obp-56a*, *circadian clock-controlled*, *cuticle protein 16.5-like*, *grainyhead*
1	dorsal scale cells	dorsal (*apterous A*+)	*neuralized*, *sanpodo*, *couch potato*, *bric a brac 1*, *ASH2*, *spire*, *HR38*, *senseless*
2	trachea, muscle, or other epidermal cells	ventral (*apterous A*–)	*grainyhead*, *ADAMTS7*
3	ventral scale precursor pIIa cells—growing/dividing	ventral (*apterous A*–)	*couch potato*, *bric a brac 1*, *shaven*, *ASH2*, *sanpodo*, *HR38*, *senseless*
4	ventral scale precursor pIIa cells—early	ventral (*apterous A*–)	*extramacrochaetae*, *enhancer of split mbeta*, *cut*, *scabrous*, *couch potato*, *shaven*, *tramtrack*, *bric a brac 1*, *basonuclin 2*, *sanpodo*
5	ventral scale precursor pIIa cells—early	ventral (*apterous A*–)	*extramacrochaetae*, *enhancer of split mbeta*, *cut*, *scabrous*, *couch potato*, *shaven*, *tramtrack*, *bric a brac 1*, *basonuclin 2*, *sanpodo*
6	ventral scale precursor pIIa cells—growing/dividing	ventral (*apterous A*–)	*neuralized*, *couch potato*, *bric a brac 1*, *shaven*, *ASH2*, *sanpodo*, *HR38*, *senseless*
7	dorsal scale cells	dorsal (*apterous A*+)	*neuralized*, *sanpodo*, *couch potato*, *bric a brac 1*, *ASH2*, *spire*, *HR38*, *senseless*
8	ventral scale precursor pIIa cells—growing/dividing	ventral (*apterous A*–)	*couch potato*, *bric a brac 1*, *shaven*, *ASH2*, *sanpodo*, *HR38*, *senseless*
9	dorsal scale cells—pheromone gland related	dorsal (*apterous A*+)	*couch potato*, *HR38*, *senseless*, *JHBP*, *Obp-56a*, *circadian clock-controlled*, *extramacrochaetae*, *enhancer of split mbeta*
10	ventral scale precursor pIIa cells—growing/dividing	ventral (*apterous A*–)	*neuralized*, *couch potato*, *bric a brac 1*, *shaven*, *sanpodo*, *HR38*, *senseless*
11	ventral scale cells	ventral (*apterous A*–)	*neuralized*, *sanpodo*, *couch potato*, *bric a brac 1*, *ASH2*, *spire*, *HR38*, *senseless*
12	innervated sensory organs—dorsal campaniform sensilla	dorsal (*apterous A*+)	*narrow*, *homothorax*, *tiptop*, *ZFHX4*, *draper*
13	innervated sensory organs—ventral marginal mechanosensory bristles	ventral (*apterous A*–)	*narrow*, *Wnt-6*, *cut*

common cellular programs like the cell cycle shared across the cell types.[38] We employed consensus non-negative matrix factorization and identified 13 GEPs that were informative for our analysis, along with their relative contributions to each cell cluster (Figures 2A and S4A–S4E).[39]

Different cell clusters exhibited varying combinations and levels of usage of GEPs that reflected their different states and activities, i.e., their identity or stage of cell-cycle progression. For example, clusters 4 and 5 uniquely expressed GEP 6 (Figures 2A and 2B), a module enriched in genes involved in regulation of transcription, DNA binding, imaginal disc-derived wing margin morphogenesis, and negative regulation of the Notch signaling pathway. GEP 6 consisted of many of the cell fate specification markers discussed above such as *cut*, *ttk*, and *shaven* along with *E(spl)*, *emc* and *basonuclin 2* (Table S1). The exclusive use of GEP 6 in only cells of clusters 4 and 5 suggests that this is an identity GEP, reflecting the early nature of these cells along the scale cell lineage. Clusters 4 and 5 also expressed GEPs 1 and 3, respectively (Figure 2A, blue dotted box), along with GEP 6. GEP 1 (and GEP 4), widely expressed across scale cell clusters (Figure 2A, white dotted boxes), consisted of numerous ribosomal proteins and ATP synthase subunits, suggesting that cells expressing this program were actively growing in either the G1 or G2 phase of the cell cycle (Figure S4F, G). GEP 3, which was also expressed in many cell clusters, was enriched in genes contributing to myosin binding (Figure S4H) as well as cell division (protein MIS12 homolog, protein Spindly) (Table S1), potentially correlating with the mitotic M phase of the cell cycle. Thus, the combinatorial expression of GEPs 6, 1, and 3 between clusters 4 and 5 illustrate their common identity but differing cellular activity profiles.

Another identity GEP, GEP 5, was uniquely and highly expressed in clusters 1, 7, and 11 (Figure 2A, green dotted box). GEP 5 comprised many cuticle proteins, *neur*, a cytoskeleton binding protein, Klarsicht, and a gene related to *Drosophila insensible* that is involved in negatively regulating Notch signaling (Table S1). This GEP also comprised two proteins, Zasp-like and spire, which can bind actin and nucleate new actin filaments, respectively.[40,41] This suggested that cells in clusters 1, 7, and 11 are the progeny of pIIa cells, potentially

CellPress
OPEN ACCESS



(legend on next page)

undifferentiated scale cells, rather than socket cells, based on the size criterion used for initial cell sorting.

We further verified cells of the scale cell lineage (clusters 1 and 3–11) by examining the expression of three of the identified marker genes, *cpo*, *shaven*, and Cut, in 22- to 28-h developing pupal wings of *B. anynana*. Chromogenic *in situ* hybridization of *cpo* mRNA showed restricted expression in the cells of the scale cell lineage (Figures 2H and S5A), and not in the surrounding epidermal cells. Like *cpo*, we observed *cut* and *shaven* mRNA expression in scale building cells. Levels of *shaven* expression were low across the wing, while *cut* mRNA occurred in many small dots, some of which were tightly clustered. Using an antibody against Cut, we saw diffused Cut protein expression in the cytoplasm of pIIa cells (Figure 2D, red arrows). However, Cut was only expressed in some, but not all, pIIa cells (Figure 2D, red stars). As the pIIa cells divided into scale and socket cells, Cut protein expression changed from cytoplasmic to nuclear (Figure 2E). Cut cytoplasmic expression, clearly anti-localized with nuclear DAPI (Figures 2E and 2F, white arrows, and S6A), changed to strong nuclear localization in the scale and socket nuclei (Figures 2E and 2F, red arrows). Dramatic differences in Cut expression patterns correlated with cell-cycle progression, which was non-uniform across the wing. Different phases of the cell cycle were demonstrated by the shape of the nucleus (circular [red arrow] vs. disc shaped [white arrow] vs. separating disc shaped [dashed circle]) (Figures 2E and 2F). In regions of the wing where most of the pIIa cells had divided, Cut expression was still non-uniform, being strongly expressed in some scale and socket pairs (Figure 2G, red arrowheads), expressed at lower levels in a few pairs (Figure 2G, white arrow) and not expressed at all in others (Figure 2G, red stars).

HR38 and *senseless* are expressed in the scale cells

We next identified other highly expressed genes in the scale cell clusters, which comprised proteins Zasp, Spire, Paramyosin, and many non-coding RNAs (Figure S7). We attempted to visualize the expression patterns and/or functionally test over 25 candidates with varying levels of success (Table S5) and only present the following results that helped us successfully validate our cluster annotations. *Irregular chiasm C-roughest* (*rst*) was expressed in the scale cells (Figure S8). In addition, a *HR38* (Figure 3A) and a zinc-finger transcription factor, *zn271*, orthologous to *Drosophila senseless* (*sens*) (Figure 3C) were expressed in neat rows across the pupal wings corresponding to either the scale or scale precursor cells (Figures 3, S5B, and S5C). *HR38* and *sens* have known functions in cuticle formation and sensory organ development in *Drosophila*.[42,43] *sens* mRNA expression was uniform across the wing but lower in the eyespot centers (Figure 3), while *HR38* expression varied between cells, and levels were reduced in the eyespot center and rings (Figure 3).

HR38 and *senseless* regulate scale identity, spacing, color, and size

We then used CRISPR-Cas9 to functionally verify the role of *shaven* and identify the functions of the highly expressed *paramyosin*, *HR38*, and *sens* genes in scale development. Knockout of *shaven* in *B. anynana* led to large patches of the wings and body devoid of scales and sockets (Figures 4A, 4B, and S9), as also observed in *Drosophila* bristle development.[37] Similarly, knockout of *paramyosin* led to loss of scales across the wing (Figure S10).

Knockout of *sens* produced crispants that exhibited changes in scale identity and loss of regular spacing (Figures 4A and 4C). Within the targeted patches, scales no longer appeared in neat, overlapping rows and were instead non-overlapping, less dense, and randomly arranged (Figure 4C, SEM). Brown, black, beige, or orange ventral cover scales were transformed into large, yellow scales with highly serrated distal edges (Figures 4C, 4E, ventral, and S11) or, in some patches, into rounded silver scales (Figure S11, Ind 1, red box). Ground scales increased in size (Figure 4E, ventral). On the dorsal surface, there was a similar lower density of scale cells with cover and ground scales increasing in size and becoming yellower (Figure 4E, dorsal), while some became tiny and silver (Figure S11, Ind 1, and 9). The tiny scales were disproportionately present on the dorsal wing surfaces compared with the ventral surface. No transformations of the white eyespot center scales were visible even when surrounding areas were targeted by the knockout (Figure S11, Ind 3, 8, and 9), confirming the low levels of *sens* mRNA expression in these scales (Figure 3).

Unlike *sens* crispants, knockout of *HR38* led to the loss of all hair-like scales (Figure 4D), indicating that this gene is required to promote hair-like scale identities. Eyespot fields, where *HR38* mRNA expression levels were lower (Figure 3), usually

Figure 2. Gene expression programs in cell clusters and expression of *cpo*, *shaven*, and Cut

(A) Gene expression program (GEP) usage map indicating the contributions of 13 GEPs to each cell cluster. White and green dashed boxes highlight the usage of GEPs 1, 4, and 5 across different cell clusters. Blue dashed box indicates the usage of multiple GEPs within cell clusters 4 and 5. Inset shows a cropped UMAP plot of the *B. anynana* single-cell atlas with the corresponding cluster numbers marked.

(B) Gene enrichment plot of the top 60 genes within GEP 6.

(C) *cut* and *shaven* mRNA are seen in the scale building cells on 22-h pupal wings. n = 3–8 wings. (D–G) Co-immunostaining of ~24-h pupal wings of *B. anynana* with DAPI and anti-Cut antibody. n = 6 wings.

(D) Cut protein is seen in the cytoplasm of some pIIa cells (red arrows). Not all pIIa cells express Cut (red stars).

(E and F) Cut protein moves from the cytoplasm in pIIa cells to the nucleus in the divided scale and socket cells. Yellow boxed regions are magnified in (F). White arrows indicate Cut protein expression anti-colocalized with nuclear DAPI expression in the pIIa scale precursor cell. Red arrows show nuclear localization of Cut protein expression in the progeny of the pIIa cell, i.e., the scale and socket cells. There is spatial heterogeneity in the scale precursor pIIa cell division across the wing. The white dashed circle in (E), expanded in the inset, highlights a dividing cell.

(G) A wing region showing divided scale and socket cell pairs expressing Cut (red arrowheads). Cut protein is, however, not seen in all scale and socket cells (red stars). White arrow indicates a scale and socket pair expressing lower amounts of Cut protein. MHW, male hindwing.

(H) *cpo* mRNA expression in ~25- to 28-h pupal wings of *B. anynana*. Expression is seen in cells arranged in neat rows corresponding to the pIIa cells or their progeny. n = 6 wings. See also Figures S4–S6 and Table S1.

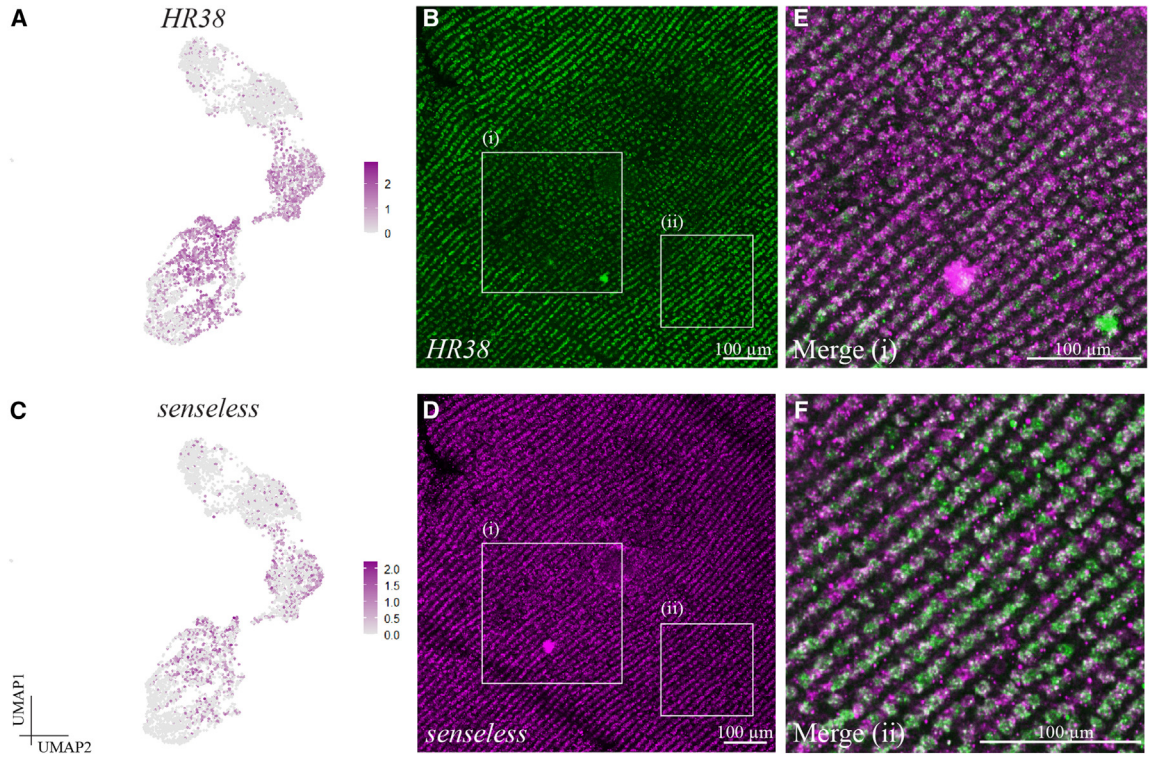


Figure 3. *HR38* and *senseless* are expressed in the scale cells

(A–D) UMAP expression plots of (A) *HR38* and (C) *senseless*. *HR38* (B) and *senseless* (D) mRNA are expressed in neat rows across 22-h pupal wings in the scale building cells.

(E and F) Enlarged merged images from regions (i) within and (ii) outside the eyespot, showing variations in levels of *HR38* mRNA between the two regions and also between different scale cells. *n* = 3–8 wings. See also Figures S5, S7 and S8.

have very low densities of hair-like scales (Figure S12). In addition, all cover and ground scales on both the dorsal and ventral surfaces decreased in size compared with wild-type scales (Figure 4E). Cover scales became similar in size to ground scales. Color was also affected, especially in the beige and brown areas of the ventral forewing, which became lighter (Figures 4D and S12). In one crispant, scales were transformed into iridescent silver scales, mostly restricted to the proximal and posterior regions of the wing (Figure S12 – Ind 3). Genotyping results for *shaven*, *senseless*, and *HR38* crispants are shown in Figure S13.

As a result of detailed comparisons of wild-type and crispant wings, we identified an intermediate scale cell type in wild-type wings among the alternating cover and ground scales (Figure S14). This scale cell type was most frequently seen on the ventral surfaces and was additionally found in other species we investigated, such as *Junonia orithya*, *Junonia almanac*, and *Catopsilia pomona* (Figure S14). This suggests that scale rows contain an extra cell type beyond ground and cover scales.

Identification of pheromone-producing glandular cells and innervated sensory cell types

Progressing beyond the scale cell clusters, we annotated cells in cluster 0 as the pheromone-producing glandular cells of *B. anynana* forewings. These are epidermal cell types based on the high level of Notch signaling in these clusters (Figures 1E, 1L, and 1M) and the lack of neural markers. GEP 8 was highly expressed in cluster 0 with some expression in clusters 9, 12, and 13 and consisted of genes involved in metabolic processes, positive regulation of Toll signaling, and structural constituent of cuticle (Figures 5A, dotted white box, and 5B; Table S1). Some of the highly expressed genes in cluster 0 (and 9, but not 2) also consisted of an odorant binding protein Obp-56a, various juvenile hormone binding proteins, cuticle proteins, and a trehalose transporter Tret-1 (Figure 5D; Table S1). Furthermore, a fatty acyl-CoA reductase (FAR), an important enzyme in the sex pheromone biosynthesis pathway,[44] was expressed at low levels in a few cells of cluster 0 (Figure S15A). Thus, cells in cluster 0 were active cuticle synthesizing, metabolic cells, which were dorsally located based on *apA* expression.

To verify the dorsal location of the glandular, pheromone-producing cells, we performed paraffin cross-sections of the wild-type gland in males. We indeed found large cells restricted to the dorsal side of the gland, overlaid by cuticle, indicating that cluster 0 comprised androconia-related, cuticle-synthesizing glandular cells (Figure 5C). To verify this, we knocked out three genes, *grainyhead*, *circadian clock-controlled* gene, and *FAR*, all three of which were expressed in cluster 0 (Figure S15A). Although the crispants showed mutations in the targeted genomic regions (Figure S15B), the frequency of indels was low and analysis of pheromone composition from

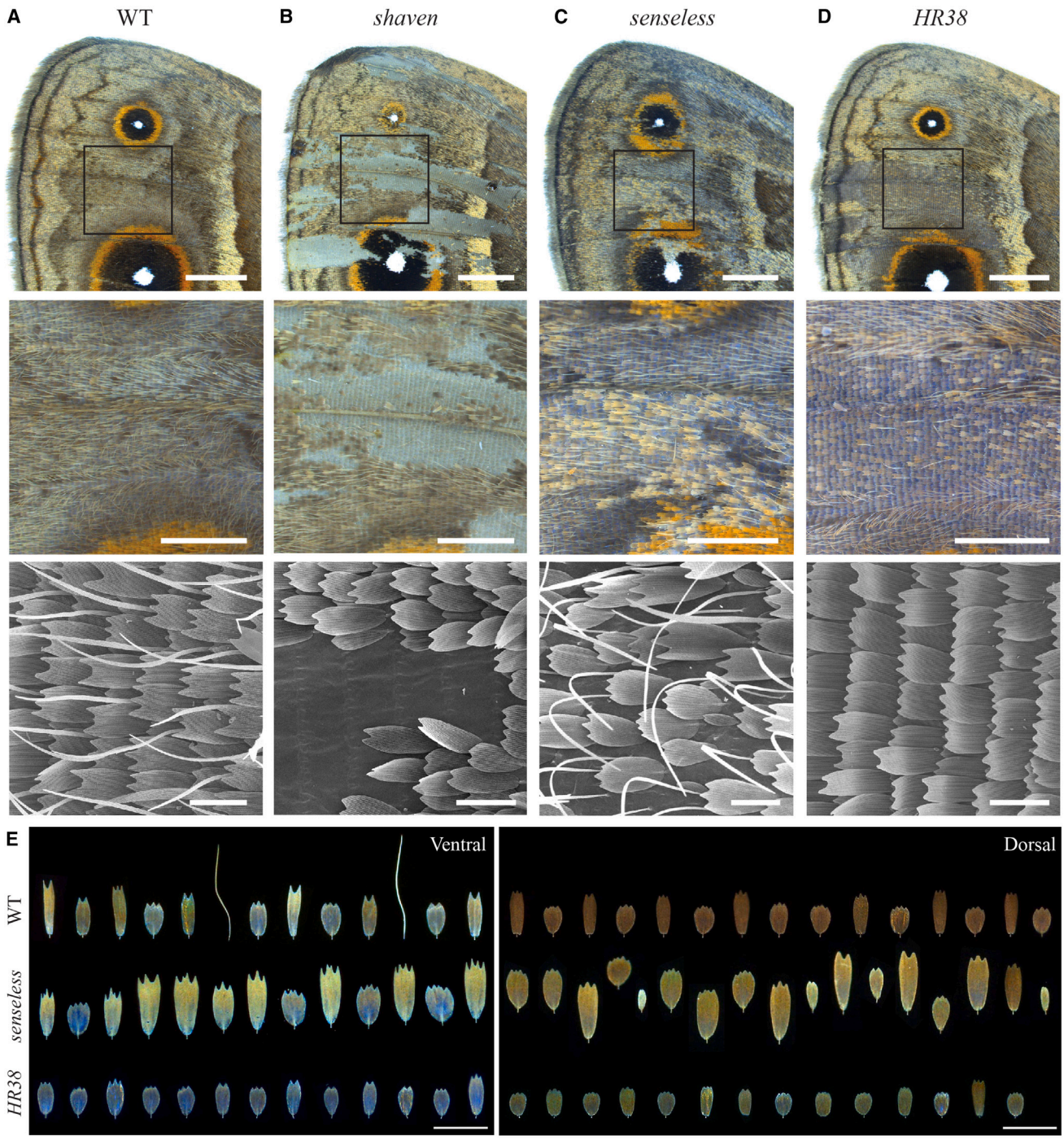


Figure 4. *shaven*, *senseless*, and *HR38* are important regulators of scale development

(A–D) (A) Wild-type (WT) ventral forewing of *B. anynana* at different magnifications. Homologous regions of the wing in (B) *shaven*, (C) *senseless*, and (D) *HR38* crispants. Scale bars, 2 mm. Black boxes in row 1 are magnified in row 2. Scale bars, 1 mm. Images in row 3 are SEM images. Scale bars, 100 μm.

(E) Arrangements of scales along a row in the WT wing compared with scale arrangements in a *senseless* and *HR38* crispant on both the dorsal and ventral surfaces. Lack of a neat, horizontal order in the *senseless* crispants indicates the loss of row arrangement and the disordered nature of the scales. Scale bars, 150 μm. See also Figures S9–S14 and Table S5.

CellPress
OPEN ACCESS

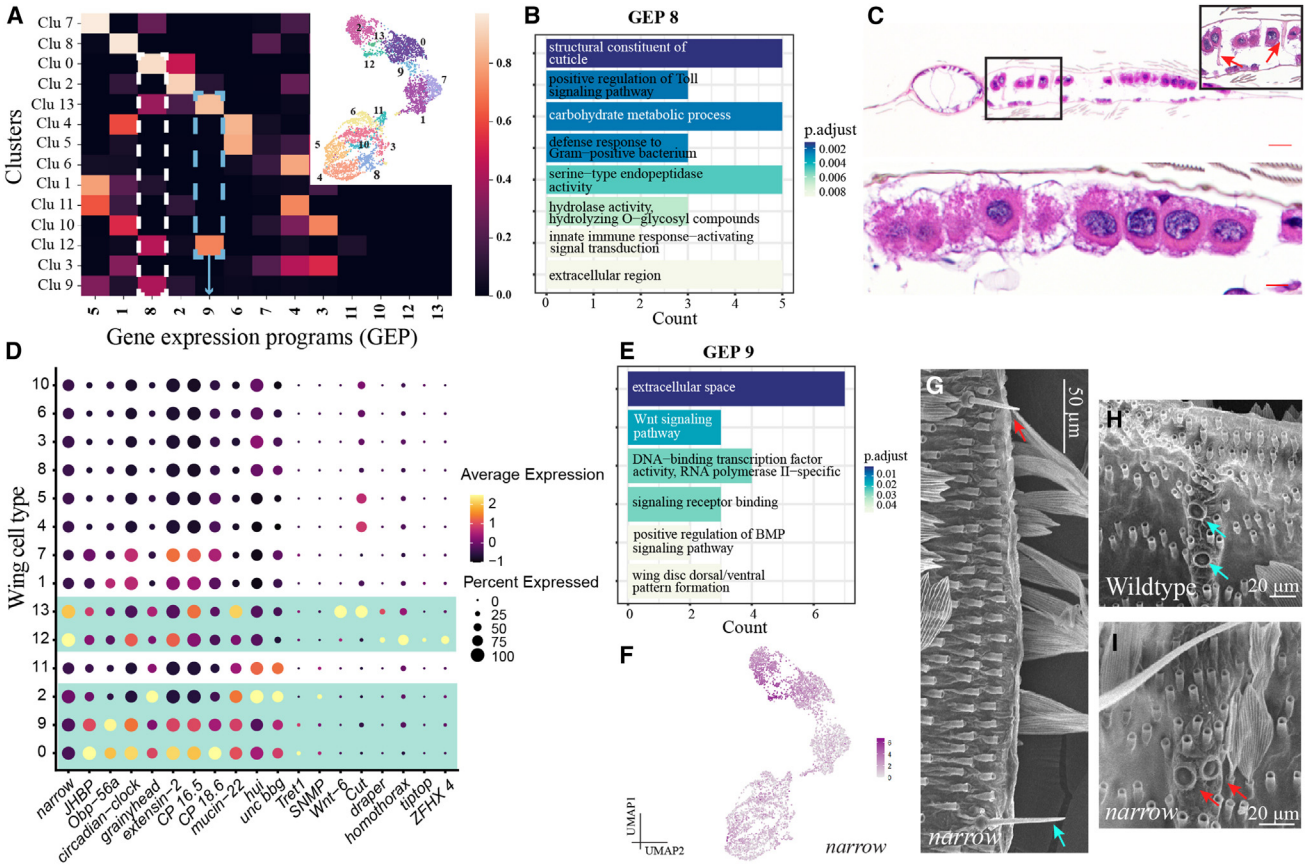


Figure 5. Epidermal and innervated sensory cells on *B. anynana* forewings

(A) GEP usage map highlighting the usage of two different GEPs, 8 and 9, across different clusters. Inset shows a cropped UMAP plot of the *B. anynana* single-cell atlas with the corresponding cluster numbers marked.

(B) Gene enrichment plot of the top 60 genes within GEP 8.

(C) Paraffin-embedded cross-sections of the WT forewing androconial gland stained with H&E showing large glandular cells restricted to the dorsal surface of the wing, overlaid by cuticle. Pillars of cuticle (red arrows in inset) appear to occur between the two surfaces of the gland that potentially make a reservoir for the pheromones. $n = 3$ individuals. Scale bars, 10 μm.

(D) Dotplot of different marker genes highly expressed in clusters 0, 2, 9, 12, and 13. The horizontal green boxes highlight the putative epidermal and innervated sensory structures on the wing.

(E) Gene enrichment plot of the top 60 genes within GEP 9.

(F) UMAP expression plot of an uncharacterized gene in *B. anynana* that is homologous to *Drosophila narrow*.

(G–I) *narrow* crispants display affected innervated sensory structure development in *B. anynana* (red arrows). Blue arrows mark WT-like marginal mechanosensory bristle in a *narrow* crispant as well as WT campaniform sensilla. To note, scales were manually brushed off the wings to be able to image the innervated sensory structures, which are often hidden by the scales. See also Figures S15 and S16.

these individuals did not show dramatic variations from the wild-type, potentially because of the low CRISPR efficiency or mistargeting of cells in the small androconial patches (Figure S15C).

We annotated cluster 2 as containing tracheal cells or muscle cells. Cells in this cluster strongly expressed GEP 2, which was enriched in genes involved in cell surface, heterophilic cell-cell adhesion, adherans junctions, and larval visceral muscle development (Figure S4I). The transcription factor *grainyhead* was also expressed at high levels in clusters 0 and 2 (Figures 5D and S15A). *grainyhead* is known to be important for cuticle formation and epidermal cell development in both vertebrates and invertebrates and in regulating apical cell membrane growth during tracheogenesis.[45–49]

Clusters 12 and 13 were annotated as different types of innervated sensory cells on the wing. These were the smallest clusters, with cluster 12 being *apA*-positive dorsal cells in contrast to the *apA*-negative ventral cluster 13 (Figure 1D). Both clusters shared the unique expression of GEP 9 (Figure 5A, dotted blue box), which was enriched for genes involved in Wnt signaling and dorsal-ventral wing patterning, suggesting that cells in clusters 12 and 13 lie close to the wing margin since the wing margin and the dorsal-ventral boundary coincide in butterflies (Figure 5E). An uncharacterized gene with a C-type lectin domain, homologous to *Drosophila narrow*, was strongly expressed in these two clusters (Figures 5D and 5F). In addition, cluster 13 expressed *cut* and *wnt-6* at high levels (Figure 5D). Supporting our annotation, strong *cut* expression

was seen in regularly spaced cells along the wing margin corresponding to the mechanosensory bristles (Figures S6C and S6D). High *cut* expression was also seen in the peripheral tissue cells close to the wing margin (Figure S6B), which may have also been captured in cluster 13. Cluster 12 differed from cluster 13 by the expression of a different group of genes such as *homothorax* and *tiptop. homothorax* is expressed along the margin and proximal wing and hinge regions in *Drosophila* and *Heliconius* butterflies (Figure 5D).[50–52] Thus, cells in cluster 12 are probably different sensory cell types close to the dorsal wing hinge.

To functionally validate cells in cluster 12 and 13, we first localized the sensory bristles on *B. anynana* pupal wings, using immunostainings with an anti-synapsin antibody, which stains neural synapses, and then knocked out a differentially expressed gene in these clusters. Various kinds of sensory cells were present across the wing disc (Figures S16A–S16D). Stout mechanosensory bristles occurred on the ventral side of the wing margin (Figures S16A and S16B, green arrow, and S16E), while dorsally, thin, hair-like bristles were present (Figures S16A and S16B, blue arrow). These thin sensory bristles were also present along the veins and near the base of the wings (Figure S16C, blue arrows) and were different from hair-like, non-innervated scales by their bases, which formed a rotatable structure within the socket (Figure S16D, blue arrow). Campaniform sensilla were also distributed across the wing, along the trachea (Figures S16C and S16D, red arrows). They occurred in groups at the base of the dorsal wing surface (Figures S16C and S16G) and in a characteristic pair, one below the other, on the ventral veins near the margin (Figures 5H and S16A, red arrows, and S16F). To verify if clusters 12 and 13 indeed correlated with the sensory cell types, we knocked out *narrow* using CRISPR-Cas9 (Figures S16I and S16L). Like *Drosophila* RNAi mutants, *B. anynana narrow* crispants had distorted wing shapes (Figure S16I).[53] In addition, they exhibited shorter ventral marginal mechanosensory bristles (Figures 5G, S16J, and S16K). In one individual, positioning of the marginal campaniform sensilla was also affected, with the pair now present side by side (Figures 5H and 5I). These data support clusters 12 and 13 mapping to various sensory cell types on the different surfaces of the wing.

DISCUSSION

Scale cell differentiation is temporally and spatially heterogeneous across the wing

In this scRNA-seq analysis of large cells (>6 μm) in ~24-h *B. anynana* pupal forewings, we annotated putative scale cell clusters and identified scale cells at different stages of their differentiation based on various marker genes. At this stage, Cut protein was not expressed in all pIIa cells and in all scale and socket cell pairs. Varying levels of Cut expression were seen in some precursor and divided scales, while others did not express Cut at all. It is possible that Cut could be marking a particular scale cell type. This hypothesis and the dynamics of Cut expression during butterfly scale development should be tested in future.

HR38 and *senseless* are important regulators of scale cell determination and differentiation

We found that *sens* is necessary for both the specification of scale cell precursors and also to specify size, color, and morphology of different scale cell types. Loss of *sens* caused a loss of row-like scale organization, lower scale density, larger scales, the appearance of smaller and scattered silver scales (dorsally) and to most scales becoming yellower. How *sens* determines surface-specific, and scale-type-specific variations (cover vs. ground, different colors and morphologies) remains unknown. Most likely, it acts in a combinatorial fashion with other transcription factors to determine cell-type specificity. In the *Drosophila* peripheral nervous system, *sens* acts as a binary switch during SOP selection, acting downstream of the proneural genes,[42,54] and later regulating terminal differentiation of the sensory cells.[55] Loss of *sens* also causes loss of SPOs via apoptosis,[42] potentially explaining the lower scale density phenotype observed in *B. anynana*.

The white eyespot center scales were, however, never affected in *sens* crispants. A possible hypothesis to explain this is the spatial repression of *sens* in the eyespot centers post-transcriptionally via the expression of an eyespot center-specific microRNA.[56] A microRNA, *miRNA-9a*, modulates levels of *sens* expression in a dynamic and complex manner in neural precursor cells during *Drosophila* sensory organ development leading to correct SOP specification.[57,58] Various such non-coding RNAs were expressed in different cell clusters in our dataset (Figure S7) and future work should investigate their function.

We identified functions for *HR38* in scale development. This hormone receptor was necessary for the development of hair-like scales, for the elongation of paddle-like scales, and for regulating their color. *Drosophila* HR38 (DHR38) protein, can form a heterodimeric complex with transactivated Ultraspiracles (USP) protein to mediate an atypical ecdysteroid signaling pathway, independent of that transduced by the ecdysone receptor/USP heterodimer.[59] Mutations in *DHR38* lead to flies with abnormal cuticle formation[60] and reduced levels of expression of various cuticle genes.[43] *HR38* has also been implicated in regulating the expression of genes in the melanin pathway.[61,62] Together, these results suggest that ecdysteroid hormonal regulation via *HR38* plays a role in lepidopteran scale cell-type development.

Pheromone-producing glandular cells are modified dorsal epithelial cells

We annotated dorsal-specific cells in our dataset with high metabolism as pheromone-producing glandular cells. These cells expressed high levels of cuticle producing genes as well as an odorant binding protein, Obp-56a. Odorant binding proteins are highly expressed in pheromone glands,[63] and Obp-56a binds to fatty acids such as palmitic acid and has been proposed as a solubilizer of fatty acids during digestion in blowflies.[64] Palmitic acid is a precursor of *B. anynana* sex pheromones,[44] and thus Obp-56a may be involved in solubilization and transport of sex pheromones in the gland cells.

Furthermore, although our analysis of pheromone composition in *grainyhead*, *circadian clock-controlled*, and *FAR* crispants did not identify any changes in levels of both MSPs 1 and 3, the

mosaic nature of CRISPR experiments might have failed to target the small number of glandular cells. Alternatively, genes such as *grainyhead* and *circadian clock-controlled* (which has a juvenile hormone binding protein domain) may function in cuticle production or transport and temporal determination of pheromone production, which was not captured in our analysis. In future, creating mutant lines would be a more precise way of investigating the roles of these genes in gland and pheromone development.

Scale cell types and a model for scale type differentiation

Butterfly wings are covered by a diverse array of scale cells that have largely been classified into a two-layer system consisting of large cover scales overlying smaller ground scales. In our study, we came across an intermediate scale cell type, distinct from the cover and ground scales. Although mentions of such intermediate scale types have been made in previous literature from moths,[65] its occurrence in butterflies had not been noted before. The presence of this intermediate scale type, which varies across and within wing surfaces, hints at the yet uncharacterized complexity of scale cell types on butterfly wings, and their developmental programs. Therefore, the first major challenge to be able to understand the evolution and development of scale cell diversity in the future, would be to systematically classify different scale types. Borrowing from concepts in neuronal cell-type classification and taxonomic principles, this would entail using multiple criteria (e.g., morphology, color, and molecular basis in the case of scales), adopting a hierarchical classification system to account for relationships between the different types and initially classifying types within specific wing regions such as dorsal vs. ventral.[66]

The generation of diverse scale cell types on butterfly wings will likely involve processes known to generate other diverse cell types such as neurons or sensory organs.[67,68] In most cases a core differentiation program, consisting of actin elongation and chitin synthesis, is modified by spatial or temporal control of regulatory factors, outside the core, that specify different scale cell fates. Genes such as *shaven* (and perhaps also *ASH2*) might represent top-level regulators of the core scale cell program. Once specified, genes such as *sens* and *HR38* appear to direct cell-type-specific differentiation. For example, low levels of *HR38* might be required to differentiate hair-like scales, whereas *sens* might be more highly expressed in cover scales, to make these larger than ground scales. These hypotheses require future testing.

Our analysis, done on large cells on the wings at ~24 h after pupation, identified many clusters of dorsal and ventral scale cells. However, although we annotated these clusters broadly as scale cells and functionally verified this annotation with a few highly expressed marker genes, we were unable to further annotate individual clusters, with markers for scale color, for instance. We were also unable to find genes or clusters that clearly differentiated cover and ground scales, such as the gene *defective proventriculus* (*dve*) that marks ground scales in *Colias* butterflies.[28] At this early pupal stage, scale cells diverged more strongly in their transcriptomic signals for genes involved in developmental progression coupled with broad morphological and functional diversity of scales, rather than genes involved in color production or cell-type identity per se. Furthermore, previously investigated transcription factors that are already marking the different colored scales at early stages of development (e.g., *spalt*, *Distal-less*, *engrailed*, *optix*, *doublesex*, etc.),[69–71] were not readily picked up by the single-cell sequencing, perhaps due to low levels of expression. To further annotate and distinguish various scale cell types, later stages of development could be sampled alongside increased depth of sequencing. Overall, our results provide a foundational atlas for future explorations of the genetic, molecular, evolutionary, and developmental basis of the various cell types in butterfly wings.

Limitations of the study

This study aimed to understand scale cell differentiation in butterfly wings by using single-cell transcriptomics of large cells at a single time point. However, scale cell determination and differentiation are temporal processes that occur over many days of PD, so we were unable to capture this dynamic process. In addition, our study investigated the population of cells >6 μm in diameter in ~24-h pupal wings. This was a limitation toward studying cells of a particular cell type and, while we captured scale cells and other large cells, we provide no information on the small epidermal cells that constitute most of the wing. In future, enrichment of scale cell populations can be done using antibodies against genes identified in this work, which will help capture specific cells of interest.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Butterfly husbandry
- METHOD DETAILS
  - Preparation of a single cell suspension, library preparation and sequencing
  - Data pre-processing and quality control
  - Analysis of scRNAseq data, unsupervised clustering and annotation of clusters
  - Consensus non-negative matrix Factoriztion (cNMF)
  - Immunostainings
  - Optical imaging and scanning electron microscopy
  - Phylogenetic analysis
  - Probe synthesis and *in situ* hybridization (digoxigenin based)
  - Hybridization chain reaction (HCR3.0 – Fluorescent based *in situ* hybridization)
  - CRISPR-Cas9 gene editing
  - Genotyping crispants
  - Histology
  - Sex pheromone composition and quantification

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114147.

ACKNOWLEDGMENTS

We thank Dr. Rajaganesan Ramaswamy for helping with the code to run cNMF (consensus non-negative matrix factorization). We acknowledge Tong Yan and Chong Ping Lee from the NUS Center for Bioimaging Sciences for their technical support in using the confocal and scanning electron microscopes. We thank the Advanced Molecular Pathology Laboratory at A*Star's Institute of Molecular and Cell Biology, Singapore, for their service in embedding and sectioning wing paraffin samples. We are grateful to Yi Yun and the Eunyoung Chae Lab at the Department of Biological Sciences, NUS, for the amplicon library preparation protocols and the iSeq runs. We thank the NUS Environmental Research Institute for the use of the GC-QQQ and associated equipment. We are grateful to Greenology for providing corn plants. This research was funded by the National Research Foundation, Singapore, Competitive Research Grant (awards NRF-CRP20-2017-0001 and NRF-CRP25-2020-0001).

AUTHOR CONTRIBUTIONS

Conceptualization, A.P. and A.M.; methodology, A.P., E.D., and T.D.B.; formal analysis, A.P.; investigation, A.P. and E.D.; writing – original draft, A.P.; writing – review & editing, A.P., E.D., and A.M.; visualization, A.P.; funding acquisition, A.P. and A.M.; supervision, A.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 18, 2023
Revised: February 26, 2024
Accepted: April 9, 2024
Published: April 23, 2024

REFERENCES

1. Galant, R., Skeath, J.B., Paddock, S., Lewis, D.L., and Carroll, S.B. (1998). Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. Curr. Biol. *8*, 807–813. https://doi.org/10.1016/S0960-9822(98)70322-7.
2. Zhou, Q., Yu, L., Shen, X., Li, Y., Xu, W., Yi, Y., and Zhang, Z. (2009). Homology of dipteran bristles and lepidopteran scales: Requirement for the Bombyx mori achaete-scute homologue ASH2. Genetics *183*, 619–627. https://doi.org/10.1534/genetics.109.102848.
3. Ghiradella, H. (1991). Light and color on the wing : butterflies and moths structural colors in. Appl. Opt. *30*, 3492–3500. https://doi.org/10.1364/AO.30.003492.
4. Greenstein, M.E. (1972). The ultrastructure of developing wings in the giant silkmoth, Hyalophora cecropia. II. Scale-forming and socket-forming cells. J. Morphol. *136*, 23–51. https://doi.org/10.1002/jmor.1051360103.
5. Overton, J. (1966). Microtubules and microfibrils in morphogenesis of the scale cells of Ephestia kühniella. J. Cell Biol. *29*, 293–305. https://doi.org/10.1083/jcb.29.2.293.
6. Yoshida, A., Noda, A., and Emoto, J. (2001). Bristle distribution along the wing margin of the small white cabbage butterfly (Lepidoptera: Pieridae). Ann. Entomol. Soc. Am. *94*, 467–470. https://doi.org/10.1603/0013-8746(2001)094[0467:BDATWM]2.0.CO;2.
7. Yoshida, A., and Emoto, J. (2011). Variations in the arrangement of sensory bristles along butterfly wing margins. Zool. Sci. (Tokyo) *28*, 430–437. https://doi.org/10.2108/zsj.28.430.
8. Dickerson, B.H., Aldworth, Z.N., and Daniel, T.L. (2014). Control of moth flight posture is mediated by wing mechanosensory feedback. J. Exp. Biol. *217*, 2301–2308. https://doi.org/10.1242/jeb.103770.
9. Fabian, J., Siwanowicz, I., Uhrhan, M., Maeda, M., Bomphrey, R.J., and Lin, H.T. (2022). Systematic characterization of wing mechanosensors that monitor airflow and wing deformations. iScience *25*, 104150. https://doi.org/10.1016/j.isci.2022.104150.
10. Aiello, B.R., Stanchak, K.E., Weber, A.I., Deora, T., Sponberg, S., and Brunton, B.W. (2021). Spatial distribution of campaniform sensilla mechanosensors on wings: form, function, and phylogeny. Curr. Opin. Insect Sci. *48*, 8–17. https://doi.org/10.1016/j.cois.2021.06.002.
11. Hartenstein, V., and Posakony, J.W. (1989). Development of adult sensilla on the wing and notum of Drosophila melanogaster. Development *107*, 389–405. https://doi.org/10.1242/dev.107.2.389.
12. Dion, E., Monteiro, A., and Yew, J.Y. (2016). Phenotypic plasticity in sex pheromone production in Bicyclus anynana butterflies. Sci. Rep. *6*, 39002–39013. https://doi.org/10.1038/srep39002.
13. Lees, A.D., Waddington, C.H., and Gray, J. (1942). The development of the bristles in normal and some mutant types of *Drosophila melanogaster*. Proc. R. Soc. Lond. Ser. B Biol. Sci. *131*, 87–110. https://doi.org/10.1098/rspb.1942.0019.
14. Hartenstein, V., and Posakony, J.W. (1990). A dual function of the Notch gene in Drosophila sensillum development. Dev. Biol. *142*, 13–30. https://doi.org/10.1016/0012-1606(90)90147-b.
15. Artavanis-Tsakonas, S., and Simpson, P. (1991). Choosing a cell fate: a view from the Notch locus. Trends Genet. *7*, 403–408. https://doi.org/10.1016/0168-9525(91)90264-q.
16. Furman, D.P., and Bukharina, T.A. (2008). How Drosophila melanogaster Forms its Mechanoreceptors. Curr. Genom. *9*, 312–323. https://doi.org/10.2174/138920208785133271.
17. Schweisguth, F. (2015). Asymmetric cell division in the Drosophila bristle lineage: From the polarization of sensory organ precursor cells to Notch-mediated binary fate decision. Wiley Interdiscip. Rev. Dev. Biol. *4*, 299–309. https://doi.org/10.1002/wdev.175.
18. Miller, S.W., Avidor-Reiss, T., Polyanovsky, A., and Posakony, J.W. (2009). Complex interplay of three transcription factors in controlling the tormogen differentiation program of Drosophila mechanoreceptors. Dev. Biol. *329*, 386–399. https://doi.org/10.1016/j.ydbio.2009.02.009.
19. Simpson, P. (1990). Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of Drosophila. Development *109*, 509–519. https://doi.org/10.1242/dev.109.3.509.
20. Le Borgne, R., and Schweisguth, F. (2003). Unequal segregation of neuralized biases Notch activation during asymmetric cell division. Dev. Cell *5*, 139–148. https://doi.org/10.1016/S1534-5807(03)00187-4.
21. Rhyu, M.S., Jan, L.Y., and Jan, Y.N. (1994). Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. Cell *76*, 477–491. https://doi.org/10.1016/0092-8674(94)90112-0.
22. Uemura, T., Shepherd, S., Ackerman, L., Jan, L.Y., and Jan, Y.N. (1989). numb, a gene required in determination of cell fate during sensory organ formation in Drosophila embryos. Cell *58*, 349–360. https://doi.org/10.1016/0092-8674(89)90849-0.
23. Dinwiddie, A., Null, R., Pizzano, M., Chuong, L., Leigh Krup, A., Ee Tan, H., and Patel, N.H. (2014). Dynamics of F-actin prefigure the structure of butterfly wing scales. Dev. Biol. *392*, 404–418. https://doi.org/10.1016/j.ydbio.2014.06.005.
24. Reed, R.D. (2004). Evidence for Notch-mediated lateral inhibition in organizing butterfly wing scales. Dev. Gene. Evol. *214*, 43–46. https://doi.org/10.1007/s00427-003-0366-0.
25. Prakash, A., and Monteiro, A. (2020). Cell dissociation from butterfly pupal wing tissues for single-cell RNA sequencing. Methods Protoc. *3*, 72. https://doi.org/10.3390/mps3040072.
26. Prakash, A., and Monteiro, A. (2018). *apterous A* specifies dorsal wing patterns and sexual traits in butterflies. Proc. Biol. Sci. *285*, 20172685. https://doi.org/10.1098/rspb.2017.2685.
27. Bellen, H.J., Kooyer, S., D'Evelyn, D., and Pearlman, J. (1992). The Drosophila couch potato protein is expressed in nuclei of peripheral

neuronal precursors and shows homology to RNA-binding proteins. Genes Dev. *6*, 2125–2136. https://doi.org/10.1101/gad.6.11.2125.

28. Ficarrotta, V., Hanly, J.J., Loh, L.S., Francescutti, C.M., Ren, A., Tunström, K., Wheat, C.W., Porter, A.H., Counterman, B.A., and Martin, A. (2022). A genetic switch for male UV iridescence in an incipient species pair of sulphur butterflies. Proc. Natl. Acad. Sci. USA *119*, e2109255118. https://doi.org/10.1073/pnas.2109255118.
29. Bier, E., Ackerman, L., Barbel, S., Jan, L., and Jan, Y.N. (1988). Identification and characterization of a neuron-specific nuclear antigen in Drosophila. Science *240*, 913–916. https://doi.org/10.1126/science.3129785.
30. Manning, L., and Doe, C.Q. (1999). Prospero distinguishes sibling cell fate without asymmetric localization in the Drosophila adult external sense organ lineage. Development *126*, 2063–2071. https://doi.org/10.1242/dev.126.10.2063.
31. Blochlinger, K., Bodmer, R., Jack, J., Jan, L.Y., and Jan, Y.N. (1988). Primary structure and expression of a product from cut, a locus involved in specifying sensory organ identity in Drosophila. Nature *333*, 629–635. https://doi.org/10.1038/333629a0.
32. Blochlinger, K., Bodmer, R., Jan, L.Y., and Jan, Y.N. (1990). Patterns of expression of cut, a protein required for external sensory organ development in wild-type and cut mutant Drosophila embryos. Genes Dev. *4*, 1322–1331. https://doi.org/10.1101/gad.4.8.1322.
33. Guo, M., Bier, E., Jan, L.Y., and Jan, Y.N. (1995). tramtrack acts downstream of numb to specify distinct daughter cell fates during asymmetric cell divisions in the drosophila PNS. Neuron *14*, 913–925. https://doi.org/10.1016/0896-6273(95)90330-5.
34. Mlodzik, M., Baker, N.E., and Rubin, G.M. (1990). Isolation and expression of scabrous, a gene regulating neurogenesis in Drosophila. Genes Dev. *4*, 1848–1861. https://doi.org/10.1101/gad.4.11.1848.
35. Renaud, O., and Simpson, P. (2001). scabrous Modifies Epithelial Cell Adhesion and Extends the Range of Lateral Signalling during Development of the Spaced Bristle Pattern in Drosophila. Dev. Biol. *240*, 361–376. https://doi.org/10.1006/dbio.2001.0482.
36. Vanhoutteghem, A., Maciejewski-Duval, A., Bouche, C., Delhomme, B., Hervé, F., Daubigney, F., Soubigou, G., Araki, M., Araki, K., Yamamura, K.i., and Djian, P. (2009). Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. Proc. Natl. Acad. Sci. USA *106*, 14432–14437. https://doi.org/10.1073/pnas.0905840106.
37. Kavaler, J., Fu, W., Duan, H., Noll, M., and Posakony, J.W. (1999). An essential role for the Drosophila Pax2 homolog in the differentiation of adult sensory organs. Development *126*, 2261–2272. https://doi.org/10.1242/dev.126.10.2261.
38. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. Elife *8*, e43803–e43826. https://doi.org/10.7554/eLife.43803.
39. Brückner, A., Badroos, J.M., Learsch, R.W., Yousefelahiyeh, M., Kitchen, S.A., and Parker, J. (2021). Evolutionary assembly of cooperating cell types in an animal chemical defense system. Cell *184*, 6138–6156.e28. https://doi.org/10.1016/j.cell.2021.11.014.
40. Ashour, D.J., Durney, C.H., Planelles-Herrero, V.J., Stevens, T.J., Feng, J.J., and Röper, K. (2022). Zasp52 strengthens whole embryo tissue integrity through supracellular actomyosin networks. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.11.511783.
41. Quinlan, M.E., Heuser, J.E., Kerkhoff, E., and Mullins, R.D. (2005). Drosophila Spire is an actin nucleation factor. Nature *433*, 382–388. https://doi.org/10.1038/nature03241.
42. Nolo, R., Abbott, L.A., and Bellen, H.J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. Cell *102*, 349–362. https://doi.org/10.1016/S0092-8674(00)00040-4.
43. Kozlova, T., Lam, G., and Thummel, C.S. (2009). Drosophila DHR38 nuclear receptor is required for adult cuticle integrity at eclosion. Dev. Dynam. *238*, 701–707. https://doi.org/10.1002/dvdy.21860.
44. Liénard, M.A., Wang, H.L., Lassance, J.M., and Löfstedt, C. (2014). Sex pheromone biosynthetic pathways are conserved between moths and the butterfly Bicyclus anynana. Nat. Commun. *5*, 3957. https://doi.org/10.1038/ncomms4957.
45. Mace, K.A., Pearson, J.C., and McGinnis, W. (2005). An Epidermal Barrier Wound Repair Pathway in Drosophila Is Mediated by grainy head. Science *308*, 381–385. https://doi.org/10.1126/science.1107573.
46. Moussian, B., and Uv, A.E. (2005). An ancient control of epithelial barrier formation and wound healing. Bioessays *27*, 987–990. https://doi.org/10.1002/bies.20308.
47. Gangishetti, U., Veerkamp, J., Bezdan, D., Schwarz, H., Lohmann, I., and Moussian, B. (2012). The transcription factor Grainy head and the steroid hormone ecdysone cooperate during differentiation of the skin of Drosophila melanogaster. Insect Mol. Biol. *21*, 283–295. https://doi.org/10.1111/j.1365-2583.2012.01134.x.
48. Yao, L., Wang, S., Westholm, J.O., Dai, Q., Matsuda, R., Hosono, C., Bray, S., Lai, E.C., and Samakovlis, C. (2017). Genome-wide identification of Grainy head targets. In Drosophila Reveals Regulatory Interactions with the POU Domain Transcription Factor Vvl, pp. 3145–3155. https://doi.org/10.1242/dev.143297.
49. Hemphälä, J., Uv, A., Cantera, R., Bray, S., and Samakovlis, C. (2003). Grainy head controls apical membrane growth and tube elongation in response to Branchless/FGF signalling. Development *130*, 249–258. https://doi.org/10.1242/dev.00218.
50. Hanly, J.J., Wallbank, R.W.R., McMillan, W.O., and Jiggins, C.D. (2019). Conservation and flexibility in the gene regulatory landscape of heliconiine butterfly wings. EvoDevo *10*, 15. https://doi.org/10.1186/s13227-019-0127-4.
51. Azpiazu, N., and Morata, G. (2000). Function and regulation of homothorax in the wing imaginal disc of Drosophila. Development *127*, 2685–2693. https://doi.org/10.1242/dev.127.12.2685.
52. Bessa, J., Carmona, L., and Casares, F. (2009). Zinc-finger paralogues tsh and tio are functionally equivalent during imaginal development in Drosophila and maintain their expression levels through auto- and cross-negative feedback loops. Dev. Dynam. *238*, 19–28. https://doi.org/10.1002/dvdy.21808.
53. Ray, R.P., Matamoro-Vidal, A., Ribeiro, P.S., Tapon, N., Houle, D., Salazar-Ciudad, I., and Thompson, B.J. (2015). Patterned Anchorage to the Apical Extracellular Matrix Defines Tissue Shape in the Developing Appendages of Drosophila. Dev. Cell *34*, 310–322. https://doi.org/10.1016/j.devcel.2015.06.019.
54. Jafar-Nejad, H., Acar, M., Nolo, R., Lacin, H., Pan, H., Parkhurst, S.M., and Bellen, H.J. (2003). Senseless acts as a binary switch during sensory organ precursor selection. Genes Dev. *17*, 2966–2978. https://doi.org/10.1101/gad.1122403.
55. Xie, B., Charlton-Perkins, M., McDonald, E., Gebelein, B., and Cook, T. (2007). Senseless functions as a molecular switch for color photoreceptor differentiation in Drosophila. Development *134*, 4243–4253. https://doi.org/10.1242/dev.012781.
56. Tian, S., and Monteiro, A. (2022). A Transcriptomic Atlas Underlying Developmental Plasticity of Seasonal Forms of Bicyclus anynana Butterflies. Mol. Biol. Evol. *39*, msac126. https://doi.org/10.1093/molbev/msac126.
57. Li, Y., Wang, F., Lee, J.A., and Gao, F.B. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. Genes Dev. *20*, 2793–2805. https://doi.org/10.1101/gad.1466306.
58. Gallicchio, L., Griffiths-Jones, S., and Ronshaugen, M. (2021). Single-cell visualization of miR-9a and Senseless co-expression during Drosophila melanogaster embryonic and larval peripheral nervous system development. G3 (Bethesda). *11*, 1–11. https://doi.org/10.1093/G3JOURNAL/JKAA010.

59. Baker, K.D., Shewchuk, L.M., Kozlova, T., Makishima, M., Hassell, A., Wisely, B., Caravella, J.A., Lambert, M.H., Reinking, J.L., Krause, H., et al. (2003). The Drosophila orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. Cell *113*, 731–742. https://doi.org/10.1016/S0092-8674(03)00420-3.

60. Kozlova, T., Pokholkova, G.V., Tzertzinis, G., Sutherland, J.D., Zhimulev, I.F., and Kafatos, F.C. (1998). Drosophila hormone receptor 38 functions in metamorphosis: A role in adult cuticle formation. Genetics *149*, 1465–1475. https://doi.org/10.1093/genetics/149.3.1465.

61. Kalay, G., Lusk, R., Dome, M., Hens, K., Deplancke, B., and Wittkopp, P.J. (2016). Potential Direct Regulators of the Drosophila yellow Gene Identified by Yeast One-Hybrid and RNAi Screens. G3 (Bethesda). *6*, 3419–3430. https://doi.org/10.1534/g3.116.032607.

62. Sekine, Y., Takagahara, S., Hatanaka, R., Watanabe, T., Oguchi, H., Noguchi, T., Naguro, I., Kobayashi, K., Tsunoda, M., Funatsu, T., et al. (2011). p38 MAPKs regulate the expression of genes in the dopamine synthesis pathway through phosphorylation of NR4A nuclear receptors. J. Cell Sci. *124*, 3006–3016. https://doi.org/10.1242/jcs.085902.

63. Zhang, Y.-N., Zhu, X.-Y., Fang, L.-P., He, P., Wang, Z.-Q., Chen, G., Sun, L., Ye, Z.-F., Deng, D.-G., and Li, J.-B. (2015). Identification and Expression Profiles of Sex Pheromone Biosynthesis and Transport Related Genes in Spodoptera litura. PLoS One *10*, e0140019–e0140022. https://doi.org/10.1371/journal.pone.0140019.

64. Ishida, Y., Ishibashi, J., and Leal, W.S. (2013). Fatty Acid Solubilizer from the Oral Disk of the Blowfly. PLoS One *8*, e51779–9. https://doi.org/10.1371/journal.pone.0051779.

65. Kristensen, N. (2012). Lepidoptera, Moths and Butterflies: Vol 2 Morphology, Physiology, and Development (De Gruyter).

66. Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: Challenges, opportunities and the path forward. Nat. Rev. Neurosci. *18*, 530–546. https://doi.org/10.1038/nrn.2017.85.

67. Arendt, D. (2008). The evolution of cell types in animals: Emerging principles from molecular studies. Nat. Rev. Genet. *9*, 868–882. https://doi.org/10.1038/nrg2416.

68. Klann, M., Schacht, M.I., Benton, M.A., and Stollewerk, A. (2021). Functional analysis of sense organ specification in the Tribolium castaneum larva reveals divergent mechanisms in insects. BMC Biol. *19*, 22. https://doi.org/10.1186/s12915-021-00948-y.

69. Brunetti, C.R., Selegue, J.E., Monteiro, A., French, V., Brakefield, P.M., and Carroll, S.B. (2001). The generation and diversification of butterfly eyespot color patterns. Curr. Biol. *11*, 1578–1585.

70. Prakash, A., and Monteiro, A. (2020). Doublesex Mediates the Development of Sex-Specific Pheromone Organs in Bicyclus Butterflies via Multiple Mechanisms. Mol. Biol. Evol. *37*, 1694–1707. https://doi.org/10.1093/molbev/msaa039.

71. Banerjee, T.D., Ramos, D., and Monteiro, A. (2020). Expression of Multiple engrailed Family Genes in Eyespots of Bicyclus anynana Butterflies Does Not Implicate the Duplication Events in the Evolution of This Morphological Novelty. Front. Ecol. Evol. *8*. https://doi.org/10.3389/fevo.2020.00227.

72. Tian, L., Su, S., Dong, X., Amann-Zalcenstein, D., Biben, C., Seidi, A., Hilton, D.J., Naik, S.H., and Ritchie, M.E. (2018). scPipe: A flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. PLoS Comput. Biol. *14*, 10063611–e1006415. https://doi.org/10.1371/journal.pcbi.1006361.

73. R Core Team (2021). R: A Language and Environment for Statistical Computing.

74. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

75. Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122. https://doi.org/10.12688/f1000research.9501.2.

76. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096.

77. Park, J., Lim, K., Kim, J.-S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. Bioinformatics *33*, 286–288. https://doi.org/10.1093/bioinformatics/btw561.

78. Challi, R.J., Kumar, S., Dasmahapatra, K.K., Jiggins, C.D., and Blaxter, M. (2016). Lepbase: the Lepidopteran genome database. Preprint at bioRxiv. https://doi.org/10.1101/056994.

79. Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L., and Marioni, J.C. (2018). Detection and removal of barcode swapping in single-cell RNA-seq data. Nat. Commun. *9*, 2667. https://doi.org/10.1038/s41467-018-05083-x.

80. Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and participants in the 1st Human Cell Atlas Jamboree; and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63. https://doi.org/10.1186/s13059-019-1662-y.

81. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS A J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

82. Macdonald, W.P., Martin, A., and Reed, R.D. (2010). Butterfly wings shaped by a molecular cookie cutter: Evolutionary radiation of lepidopteran wing shapes associated with a derived Cut/wingless wing margin boundary system. Evol. Dev. *12*, 296–304. https://doi.org/10.1111/j.1525-142X.2010.00415.x.

83. Zhou, Q., Zhang, T., Xu, W., Yu, L., Yi, Y., and Zhang, Z. (2008). Analysis of four achaete-scute homologs in Bombyx mori reveals new viewpoints of the evolution and functions of this gene family. BMC Genet. *9*, 24. https://doi.org/10.1186/1471-2156-9-24.

84. Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. *8*, 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

85. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.

86. Banerjee, T.D., and Monteiro, A. (2018). Crispr-cas9 mediated genome editing in bicyclus anynana butterflies. https://doi.org/10.3390/mps1020016.

87. Nieberding, C.M., de Vos, H., Schneider, M.V., Lassance, J.M., Estramil, N., Andersson, J., Bång, J., Hedenström, E., Löfstedt, C., and Brakefield, P.M. (2008). The male sex pheromone of the butterfly Bicyclus anynana: Towards an evolutionary analysis. PLoS One *3*, 27511–e2812. https://doi.org/10.1371/journal.pone.0002751.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
2B10 mouse anti-Cut	DSHB	Cat# 2b10; RRID: AB_528186
mouse 3C11 (anti-SYNORF1)	DSHB	Cat# 3C11; RRID:AB_528479
Alexa Fluor 488 AffiniPure Goat anti-mouse IgG	Jackson ImmunoResearch Labs	Cat# 115-545-003; RRID:AB_2338840
Chemicals, peptides, and recombinant proteins		
10X TrypLE Select	Thermo Fisher Scientific	Cat# A12177-01
Q5 High-Fidelity DNA Polymerase	NEB	Cat# M0491S
T7 RNA Polymerase	NEB	Cat# M0251S
HCR Hairpins	Molecular Instruments	https://www.molecularinstruments.com/hcr-rnafish-products
Critical commercial assays		
Kaneka Easy DNA Extraction Kit 2	Kaneka, Japan	Cat# KN-T110005
Deposited data		
Raw scRNA-seq reads and processed matrix files	This paper	GEO: GSE241536
HCR, *insitu* hybridization, IHC and crispant images	This paper	Mendeley Data: https://doi.org/10.17632/rhvrsdzc66.2
Experimental models: Organisms/strains		
Bicyclus anynana wild-type butterflies	Antonia Monteiro lab	N/A
Oligonucleotides		
Primers to amplify mRNA sequences for *in situ* hybridization probes	This paper	Table S3
Primers for Hybridization Chain Reaction	This paper	Table S4
Guide RNA sequences to generate the different *Bicyclus anynana* crispants	This paper	Table S6
Genotyping primers for crispants	This paper	Table S7
Recombinant DNA		
PGEM-T plasmids with partial mRNA sequences for *couch potato*, *HR38*, *senseless* and *rst*-scales	This paper	N/A
Software and algorithms		
scPipe R package	Tian et al.[72]	https://bioconductor.org/packages/release/bioc/html/scPipe.html
RStudio 2023.06.0 + 421 with R 4.3.1	R core team[73]	https://www.rstudio.com/; https://www.r-project.org/
Seurat v3.1	Satija lab; Stuart et al.[74]	https://satijalab.org/seurat/
Scran	Lun et al.[75]	https://bioconductor.org/packages/release/bioc/html/scran.html
cNMF code	Brückner et al.[39]	https://doi.org/10.1016/j.cell.2021.11.014; CaltechData https://doi.org/10.22002/D1.1918
MEGA X	Kumar et al.[76]	https://www.megasoftware.net/
CRISPR-RGEN Tools	Park et al.[77]	http://www.rgenome.net/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Antonia Mónteiro (antonia.monteiro@nus.edu.sg).

Materials availability

All resources and reagents generated in this study are listed in this paper or are available upon request to the lead contact.

Data and code availability

- Single-cell RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. HCR, *insitu* hybridization, IHC and crispant images have been deposited in Mendeley Data. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Butterfly husbandry

Bicyclus anynana butterflies were reared in a temperature-controlled room at 27°C, 65% humidity and a 12:12 h light:dark cycle. Caterpillars were fed on corn plants and adults were fed on mashed bananas. For this experiment, male pupal forewings of *B. anynana* butterflies at 22.5–25 h after pupation were used.

METHOD DETAILS

Preparation of a single cell suspension, library preparation and sequencing

The preparation of a single-cell suspension from 22.5–25-h male pupal forewings of *B. anynana* butterflies has been documented in-depth in a protocol paper along with the FACS results.[25] Briefly, 14 male pupal forewings were dissected into multiple tubes with sterile PBS, washed, then dissociated in warm 5X TrypLE with continuous trituration for a total of 15–20 min. The dissociated cell mixture was filtered with a 41 μm filter and collected by centrifuging at 300 g, 4°C for 5 min. Samples were pooled and gently resuspended into 1mL of cold 1X PBS +0.01% BSA. Cells were counted in a Countess Automated cell counter by mixing 5 μL of sample with 5 μL of 0.4% trypan blue and loading it into a Countess cell counting chamber. Before going for fluorescence-activated cell sorting (FACS), 5 μL of a 300 μM intermediate concentration of DAPI (Invitrogen by Thermo Fisher Scientific, Cat. no.: D1306) was added per mL of sample. Cells larger than 6 μm in diameter were sorted using FACS and collected in a 1.5 mL LoBind Eppendorf tube prefilled with 50 μL of cold 1X PBS +0.01% BSA. We excluded the large majority of ~3 μm wing epithelial cells using this strategy. Around 104,000 large cells were collected and precipitated by centrifuging at 900g, 4°C for 3 min. Cells were resuspended in 30 μL of cold 1X PBS +0.01% BSA, with a final cell count of 1650 cells/μL and 88% viability.

Cells were processed via the 10X Genomics microfluidics system at the Genome Institute of Singapore. Single-cell libraries were generated using the Chromium Single cell 3′ Library Kit V3 with a targeted cell recovery of 7000 cells. The quality of the generated cDNA library was quantified using an Agilent Bioanlyzer and sequenced on one lane of HiSeq4K 2x151bp (multiplex) at the Genome Institute of Singapore.

Data pre-processing and quality control

Barcode and UMI trimming, demultiplexing, and quality assessment of the reads was performed using the scPipe R package.[72] The workflow was as follows: The FASTQ files containing the reads were first reformatted to trim the barcodes and UMI sequences and move it to the header. Reads were then aligned to the *Bicyclus anynana* genome (v1.2) and mapped to exons using an annotation file - Bicyclus_anynana_v1.2.gff3 downloaded from Lepbase.[78] Demultiplexing was done based on barcodes. Of the 346,685,500 initial reads, 72.36% had a barcode match. The generated gene count matrix consisted of 13631 genes and 9986 cells, with a mean count per cell of ~10700 reads and mean number of genes per cell of ~1400.

We next discarded low-quality cells, empty droplets with no cells, and potential doublets from the gene count matrix. Cells with counts per cell>100000 were removed and genes with less than 10 counts summed across all cells were discarded. The DropletUtils R package[79,80] was used to filter our empty droplets using an FDR $\leq$0.001. The detect_outlier function in scPipe was then used to detect outliers (both low quality cells and potential doublets) using a Gaussian mixture model. Additionally, we removed 18 histone and 53 ribosomal protein genes because an initial analysis of the data indicated unsupervised clustering based on these genes, masking underlying patterns (Table S2). This led to a gene count matrix of 5268 cells and 10743 genes used in the initial exploratory analysis detailed below. A Seurat object[74] was created using the count matrix with min.cells = 5 i.e., keeping features that were present in at least 5 cells. The count matrix for the Seurat analysis comprised 5268 cells and 10696 genes.

Analysis of scRNAseq data, unsupervised clustering and annotation of clusters

Downstream analysis was carried out in two ways in R v4.3.1[73]: using Seurat v3.1[74] (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html) or scran[75] (https://bioconductor.org/packages/devel/bioc/vignettes/scran/inst/doc/scran.html). Initial data exploration was carried out using scran. First, scaling normalization was performed using a deconvolution strategy with the quickCluster, computeSumFactors and logNormCounts functions. Highly variable genes were identified by variance modeling using the modelGenVarByPoisson function and getTopHVGs function was used to select the top 50% (prop = 0.5; var.field = "bio") of genes. The reduced data matrix with the highly variable genes was used as an input for a Principal Component Analysis (PCA). The number of principal components to retain in the analysis was computed using getClusteredPCs function. A shared nearest neighbor graph was constructed using buildSNNGraph function, following which, clusters were identified using the igraph package. Varying values of k changed cluster resolution with smaller k values giving finer clusters. Different k values were used for the initial exploration of the data. runUMAP and plotUMAP functions helped visualize the clusters.

In the Seurat analysis, transcript counts for each cell were log normalized using the NormalizeData function. The top 1000 highly variable transcripts across all cells were selected using the FindVariableFeatures function and 'vst' as the selection method. Transcript expressions across cells were then centered and scaled using the ScaleData function. Dimensionality reduction was done using PCA on the scaled data matrix with the previously determined variable features. Thereafter, FindNeighbors function was used to construct a shared nearest neighbor graph using the first 25 principal components and prune.SNN = 1/15. Clustering was performed using the FindClusters function and a resolution of 0.8. The wing cell type atlas was generated using the RunUMAP function on the first 50 principal components and plotted using DimPlot.

Clusters were annotated using a combination of different methods. Initially, expression of known marker genes from the *Drosophia melanogaster* bristle development literature was visualized using the FeaturePlot function. In parallel, genes highly and differentially expressed between various clusters were identified from both the Seurat and scran analyses using FindAllMarkers (Seurat), FindMarkers (Seurat) or findMarkers (scran) functions.

Consensus non-negative matrix Factoriztion (cNMF)

cNMF was run on the cells-by-transcripts count matrix (5268 cells by 10743 transcripts) to identify gene expression programs.[38] The steps followed were based on the code provided in.[39] We used the 2000 most over-dispersed transcripts for the factorizations and ran 200 replicates of NMFs for each value of K ranging from 5 to 20. Replicates for each K were then combined to obtain a consensus, and this was used to estimate the stability and error for each K. The stability-error plot as a function of K was used to select the best choice of K for our analysis. Though K = 10 had the maximum stability solution, we found that this K value provided poorer resolution of our clusters in the downstream analysis. We chose to go for K = 13 which had the second highest stability vs. error. We re-ran the consensus with K = 13 and a density threshold of 0.18, to obtain the consensus GEP usage scores for each cell. The normalized GEP usage scores were combined with the cluster annotations for each cell obtained from Seurat to calculate the proportional GEP usage in each cluster, visualized as a heatmap.

Gene scores were extracted from the generated gene_spectra_score text file and the top 60 enriched genes in each GEP were obtained. Gene enrichment analysis was carried out using the enricher function (pvalueCutoff and qvalueCutoff = 0.05, pAdjustMethod = "BH") from the clusterProfiler package.[81]

Immunostainings

We used the 2B10 mouse anti-Cut primary antibody raised against a *Drosophila* Cut antigen. 2B10 was deposited to the DSHB by Rubin, G.M. (DSHB Hybridoma Product 2B10). This primary antibody has been previously used to target Cut protein expression in many Lepidopteran species.[82] The mouse 3C11 (anti-SYNORF1) was used to target synapsin. 3C11 (anti SYNORF1) was deposited to the DSHB by Buchner, E. (DSHB Hybridoma Product 3C11 (anti SYNORF1)). An Alexa Fluor 488-conjugated goat anti-mouse antibody (Jackson ImmunoResearch Laboratories, Inc.) was used as a secondary antibody.

Pupal wing tissues at various time points were dissected and fixed in fix buffer (0.1M PIPES pH 6.9, 1mM EGTA pH 6.9, 1% Triton X-100, 2mM MgSO4) at room temperature. Fixation was done for 30 min following the addition of formaldehyde (final concentration of 4% directly to the wells). The wings were then moved onto ice, washed five times with PBS and transferred to block buffer (50mM 626 Tris pH 6.8, 150mM NaCl, 0.5% IGEPAL, 1 mg/ml BSA) at 4°C overnight. Incubation with primary antibody at a concentration of 2–2.5 μg/mL in wash buffer (50mM Tris pH 6.8, 150mM NaCl, 0.5% IGEPAL, 1 mg/ml BSA) was carried out at room temperature for 1 h, followed by four washes with wash buffer. Wings were then incubated with secondary antibody (1:500) for 30 min at room temperature followed by 5–10 washes to remove the secondary antibody. Wing tissues were counterstained with DAPI (1:100) for 5–10 min, followed by further washing. Control stains used only secondary antibodies. Mounted wings were imaged on an Olympus FLUOVIEW FV3000 confocal microscope.

For staining wings with CellMask Orange Plasma membrane stain (Invitrogen, Cat no. C10045) and DAPI, wings were dissected and immediately placed in PBS with the CellMask Orange stain (1:300) for 10 min at room temperature. This was because the CellMask Plasma membrane stain does not survive permeabilization. The stain was then removed, and wings were fixed in 4% formaldehyde in PBS at room temperature for 20 min, followed by 5 washes with cold PBS. Tissues were counterstained with DAPI (1:100) for 5–10 min, then washed with PBS 3–5 times. Mounted wings were imaged on an Olympus FLUOVIEW FV3000 confocal microscope.

Optical imaging and scanning electron microscopy

Optical images of single scales and wildtype or crispant wings were captured using a Leica DMS 1000 microscope.

For scanning electron microscopy, desired regions of the wildtype or crispant adult wings were cut using scissors and mounted onto carbon tape. For imaging different sensory cell types, the scales were carefully and gently brushed off the wings using a paintbrush. Wing pieces were sputter coated with gold using a JFC-1100 Fine Coat Ion Sputter (JEOL Ltd. Japan) and imaged on a JEOL JSM-6510LV scanning electron microscope (JEOL Ltd. Japan) located at the Center for Bioimaging Sciences (CBIS, NUS).

Phylogenetic analysis

To identify the evolutionary relatedness of the *Bicyclus anynana* AS-C homolog, identified in the scRNA-seq analysis, to other Lepidopteran, Dipteran and Coleopteran AS-C genes, we compiled a list of protein sequences of AS-C homologs from NCBI, based on.[83] The final dataset consisted of 12 amino acid sequences with a total of 503 positions. Evolutionary analyses were conducted in MEGA X.[76] The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model.[84] Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree with the highest log likelihood (−7192.28) was selected.

Probe synthesis and *in situ* hybridization (digoxigenin based)

mRNA sequences were amplified from cDNA using the primers specified in Table S3 and cloned into a PGEM-T Easy vector (Promega). Colonies containing plasmids with inserts of the right size were selected using blue-white screening. Purified plasmids were used to prepare digoxigenin-labelled sense and anti-sense riboprobes using T7 and SP6 RNA polymerases (Roche). Probes were purified using ethanol precipitation and resuspended in 1:1 volume of DEPC-treated water:formamide.

Pupal wings at various time points were dissected and placed in glass well plates containing PBST (PBS+0.1% Tween 20). Tissues were fixed at room temperature for 30 min in 4% formaldehyde in PBST followed by 3 washes in cold PBST. The wings were then incubated in 25 μg/mL proteinase K in cold PBST for 5 min, washed twice with 2 mg/mL glycine in cold PBST, then 5 times with cold PBST before being gradually transferred to a prehybridization buffer (5X saline sodium citrate (pH 4.5), 50% formamide, 0.1% Tween 20 and 100 μg/mL denatured salmon sperm DNA). Tissues in prehybridization buffer were incubated at 60°C for 1 h then transferred to a hybridization buffer (prehybridization buffer with 1 mg/mL glycine and 180 ng/mL DIG labeled riboprobe) overnight. This was followed by 5–6 rounds of washes with prehybridization buffer at 60°C, a gradual transfer back to PBST at room temperature, 3 washes in PBST and blocking overnight at 4°C (PBST +1% BSA). Probe detection was by tissue incubation in 1:3000 anti-DIG alkaline phosphatase (Roche) in block buffer for 1 h, 2 washes in alkaline phosphatase buffer (100 mM Tris (pH 9.5), 100 mM NaCl, 5 mM MgCl2, 0.1% Tween) and a final incubation at room temperature in NBT/BCIP solution (Promega) till color developed. The reaction was stopped by washing in PBST. Wings were mounted on slides and imaged on a Leica DMS 1000 microscope. High magnification images were captured using a 100X lens of a uSight-2000-Ni microspectrophotometer (Technospex Pte. Ltd., Singapore) and a Touptek U3CMOS-05 camera.

Hybridization chain reaction (HCR3.0 – Fluorescent based *in situ* hybridization)

Fluorescent *in situ* hybridization of the mRNA species was carried out either using an inbuilt fluidics robot or manually based on the protocol described in Choi et al., 2018[85] with a few modifications in buffers and incubation conditions. Timed pupal wings were fixed at room temperature in glass spot plates in 4% formaldehyde for 30 min. Wings were thereafter washed twice using 1x PBST and treated with a detergent solution. Afterward, the wings were washed twice using 1x PBST and twice using 5x SSCT. Wings were then either transferred to a 3D printed incubation chamber where the hybridization reaction was carried out at 42°C, automatically using the fluidics robot, or performed manually. For both the manual and the automatic rounds, the hybridization involved incubation in a solution containing 20 μL (100 μM) of probe set against each gene (IDT) in 1000 μL of 30% probe hybridization buffer followed by rigorous washing with 30% probe wash buffer. Afterward, wings were washed with 5X SSCT and incubated in an Amplification buffer for 30 min. For the chain reaction, a solution with HCR hairpins (Molecular instruments) in the amplification buffer was added to the tissues followed by washes in 5x SSCT. Wings were stained with DAPI and mounted in an in-house mounting media. Imaging was carried out manually under an Olympus FV3000 confocal microscope. Images were post-processed using the Levels function in Adobe Photoshop to improve contrast. The primers for HCR are specified in Table S4.

CRISPR-Cas9 gene editing

Expression and/or functional validation for 25 candidate genes listed in Table S5 were carried out to help in the annotation of various clusters. Briefly, the CRISPR-Cas9 gene editing workflow involved the following steps.[86] Guide DNA templates targeting genes of interest, preferably protein domains, were manually designed by searching for $GGN_{18}NGG$ sequences and used to design forward primers 5′-GAAATTAATACGACTCACTATAGG-xxxxxxxxxxxxxxxxxxxx-GTTTTAGAGCTAGAAATAGC-3' (Table S6). Double stranded DNA templates were generated by a Q5 High-Fidelity DNA polymerase PCR (Cat. No. M0491S, NEB) using a common reverse primer 5′-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG CTATTTCTAGCTCTAAAAC-3'. Guide RNAs were prepared from purified DNA templates using *in vitro* transcription reactions with a T7 RNA polymerase (Cat. No. M0251S, NEB) and purified using ethanol precipitation. Injection mixtures containing 500 ng/μL of Cas 9 protein

(NEB; Cat. no.: M0641) and 300 ng/μL of guide RNA were prepared and injected into eggs within 4 h of egg laying. Caterpillars were reared on corn plants and adults were scored for their phenotypes.

Genotyping crispants

Genomic DNA from thoracic tissue and legs of potential crispant individuals was isolated using the Kaneka Easy DNA Extraction Kit 2 (Cat. No. KN-T110005, Kaneka, Japan). Next generation sequencing was used to verify the CRISPR edits. Indexed libraries were prepared in a two-step PCR process. First, targeted regions were amplified using gene specific primers appended to reading primers: Forward: 5′ACACTCTTTCCCTACACGACGCTCTTCCGATCT3'; Reverse: 5′GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT3' (Table S7). Gene of interest amplicon size was kept to <300 bp. In the second PCR step, indices and Illumina adapter sequences were appended. The Illumina adapters D501-D508 and D701-D712 were used to multiplex samples of different genes and different individuals together. PCR products were then purified using ethanol purification, pooled together and sequenced on an iSeq 100 (2 x 150 bp) at the Chae lab, National University of Singapore. Demultiplexed FASTQ files were visualized online using the web-tool "CRISPR-RGEN Tools" (http://www.rgenome.net/).[77]

Histology

Androconial glands from the forewings of adult male butterflies were cut from the wings using a dissection scissors and descaled using a brush. Tissue pieces were fixed in 4% formaldehyde in PBS overnight. Samples were then passed to the Advanced Molecular Pathology Laboratory (AMPL) at the A*STAR Institute of Molecular and Cell Biology, Biopolis, Singapore for downstream processing. Tissues were embedded in paraffin blocks, sectioned and stained with H&E. Sections were imaged using the imaging system on a uSight-2000-Ni microspectrophotometer (Technospex Pte. Ltd., Singapore) and a Touptek U3CMOS-05 camera.

Sex pheromone composition and quantification

The composition of the sex pheromone blend and the amount of the two components (called MSP1 and MSP3 for Male Sex Pheromone) were compared between wildtype and crispant individuals using gas chromatography – mass spectrometry. Three days-old crispant butterflies were anesthetized at −80°C for 5 min. The forewing was cut at the base of the thorax with fine scissors and placed for 15 min in a glass vial containing 500uL of hexane. 400uL of the solution containing the dissolved sex pheromones was extracted to a new vial, to which 10 μg/mL of methyl stearate (Merck, Singapore) was added as an internal standard. All tools and vials were pre-rinsed with hexane, and all extractions were done at 2 p.m. to prevent the effects of daily fluctuations in pheromone titers.

The extracts were analyzed with a Shimadzu Gas-Chromatography-QQQ Mass Spectrometer equipped with a DB-5 column, using the following set up: electron ionization (EI) was done at 70 eV, 1 μL of each extract was injected splitless, with the injector temperature set up at 250°C. Helium was used as carrier gas, with the flow set at 1.9 mL/min. The column temperature gradient began at 50°C, increased to 210°C at a rate of 35°C/min, then increased to 280°C at a rate of 3°C/min. The detector was set to unit mass resolution and 3 scans/sec, from m/z 37 to 500. Chromatograms and mass spectra were analyzed using the Shimadzu GCMSsolution software v. 4.11. The components and precursors are well documented and were identified based on their known mass spectrum and retention time.[12,44,87] The amount of sex pheromone components was calculated by normalizing the area of the component peak to the peak of the internal standard.