



OPEN

# Whole organism integrated DNA methylation and transcriptomics analysis of butterfly metamorphosis

Andrew Hesketh<sup>1,3</sup>, Juned Kadiwala<sup>1,3</sup>, Ana Rita Garizo<sup>2</sup>, Patrícia Beldade<sup>2</sup> & Rameen Shakur<sup>1</sup>✉

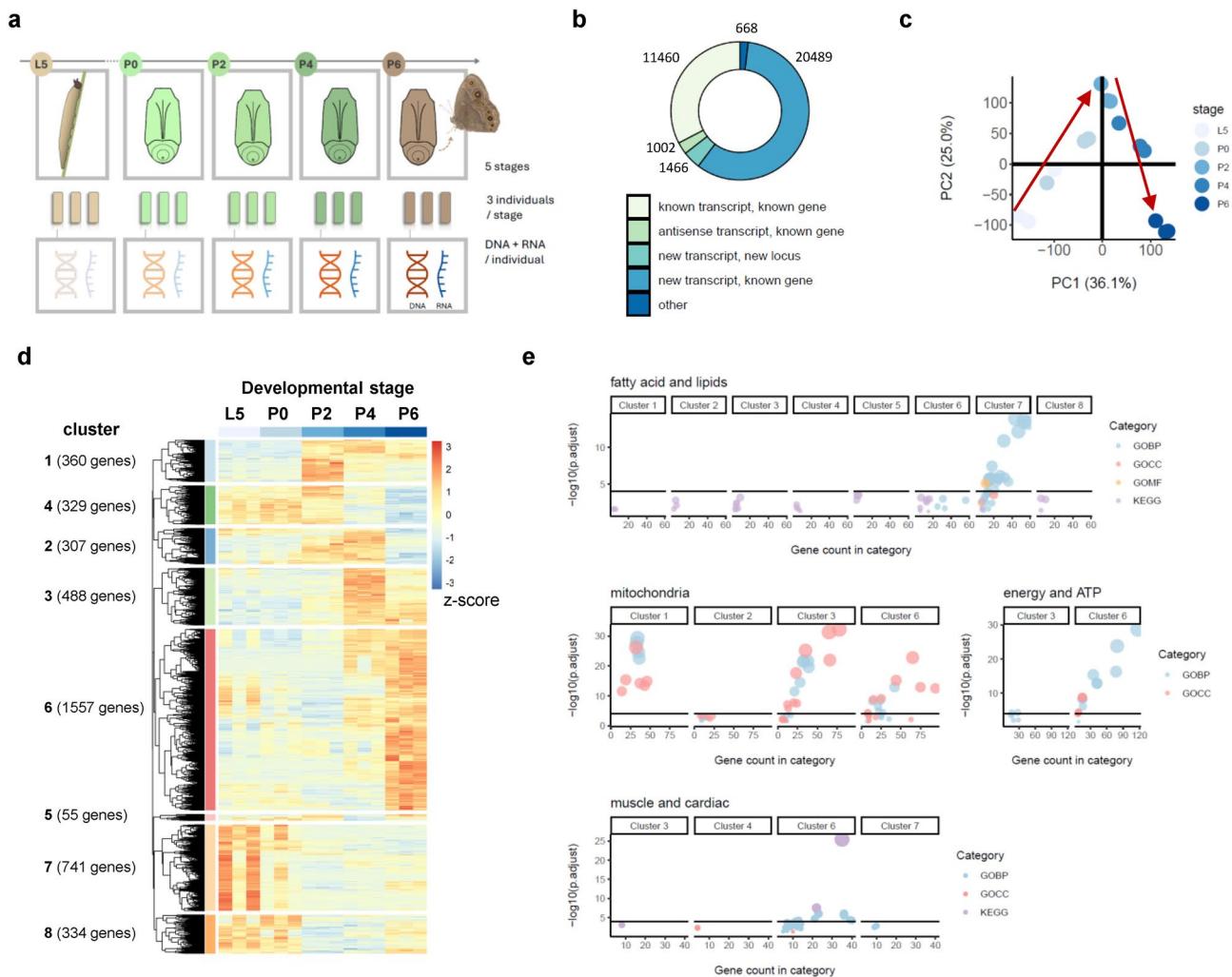
Metamorphosis is one of the most fascinating developmental processes in the natural world. The underlying molecular events resulting in the transformation of a caterpillar into a butterfly, including the epigenetic and transcriptomic programs, remain elusive. By integrating DNA and RNA long-read sequencing of the entire body of five consecutive stages from larval to late-pupal development in a laboratory butterfly, we characterise the fundamental metabolic and developmental transitions taking place. We identify a progression from lipid metabolism in larvae towards an up-regulation of muscle formation and mitochondrial energy generation in pupae. Intra-genic CpG methylation correlates with, but does not appear to dictate, gene expression. The level of 5-hydroxymethylcytosine modification detected was very low. The use of long-read mRNA sequencing provided access to complete transcript isoform sequences, and differential isoform usage was notably detected in genes for energy metabolism, and for muscle and neural development during the metamorphic process.

Insect metamorphosis is a complex developmental process of transition between life stages that differ greatly in morphology, behaviour, and life-style. Complete metamorphosis, which evolved over 400 million years ago and characterises the very successful group of holometabolous insects<sup>1</sup>, allowed the decoupling of growth and differentiation. During metamorphosis, insects undergo a dramatic remodelling of their body plan, with the breakdown and re-organization of larval tissues and the coordinated development of adult structures. The transformation of larvae that are specialised in feeding and growth into adults that can disperse and reproduce takes place through a non-feeding and immobile transitional stage called pupa. While often not featured in popular accounts of metamorphosis, the pupa is centre stage in metamorphosis. It provides a mould for the adult body and is typically formed from reprogrammed larval cells and cells from imaginal primordia, whose relative contribution varies between organs and between species<sup>2</sup>.

While the genetic programs associated with metamorphosis have been studied extensively in *Drosophila melanogaster* flies<sup>3–6</sup>, metamorphosis remains underexplored in other insects, including the Lepidoptera that include butterflies and moths. The transition from a (very hungry) terrestrial caterpillar to a (fluttering) airborne butterfly has long captivated people of all ages, but we know surprisingly little about how this is achieved. Much of what we do know comes from studies of specific body parts (e.g. metamorphosis of the wings, legs, head, eyes), and has focused primarily on comparison of a single transition event, from either larva to pupa or larva to butterfly<sup>7–10</sup>. Here, we use the butterfly model *Bicyclus anynana*<sup>11</sup> to characterise gene expression and DNA methylation of whole individuals as they transition from larva to pupa and then progress through pupal development until a fully developed adult is about to eclose out of the pupal case (Fig. 1, Supplementary Fig. S1). The use of triplicate samples obtained from the whole organism at five different developmental stages, combined with the sampling of RNA and DNA from the same individuals, provides a dynamic dataset describing butterfly metamorphosis.

We used Oxford Nanopore long-read sequencing to collect the data, allowing any DNA modifications to be directly observed without prior chemical modification, and providing access to complete mRNA transcripts. In this way, we sought to define the phase-dependent changes in gene expression taking place during metamorphosis

<sup>1</sup>Brighton Integrative Genomics (BIG) Unit and the Centre for Precision Health and Translational Medicine, School of Applied Sciences, University of Brighton, Brighton BN2 6DN, UK. <sup>2</sup>Center for Ecology, Evolution and Environmental Changes (cE3c) and Global Change and Sustainability Institute (CHANGE), Faculty of Sciences, University of Lisbon (FCUL), 1749-016 Lisbon, Portugal. <sup>3</sup>Andrew Hesketh and Juned Kadiwala have contributed equally to this work. ✉email: r.shakur@brighton.ac.uk



**Fig. 1.** Extensive re-organisation of gene expression during metamorphosis. **(a)** Male *Bicyclus anynana* individuals were sampled in triplicate for transcriptome and DNA methylome analysis during progression from 5<sup>th</sup> and final instar larvae (L5), to the day of pupation (P0), and 2 (P2), 4 (P4) and 6 (P6) days post-pupation when fully formed butterflies are about to eclose. **(b)** Overview of the known and novel transcripts identified in the RNA sequencing data compared to the reference annotation. The groups shown correspond to the following gffcompare class codes: “known transcript, known gene”, class code “=”; “antisense transcript, known gene” class code “x”; “new transcript, new locus” class code “u”; “new transcript, known gene”, class codes “c, j, k, n, m, o”; “other”, class codes “i, s, y” (Supplementary Fig. S1). **(c)** Unsupervised clustering using PCA separates the initial larval samples (L5) from the final late-stage pupal stage samples (P6) in PC1 (which describes 36% of the variance in the data) and shows a progression from stage L5 to P2 and from P2 to P6 in PC2 (indicated by the arrows). **(d)** Hierarchical clustering of the normalised expression data for the 4171 significantly differently expressed genes defines eight co-expression clusters and characterises the extensive reprogramming of transcription occurring during progression from L5 to P6. Gene cluster membership is listed in Supplementary Data S2. **(e)** Functional analysis of the genes in each co-expression cluster using annotation from Gene Ontology (Biological Process (GOBP), Molecular Function (GOMF), Cellular Compartment (GOCC)) and KEGG databases highlighting significant over-representation ( $\text{padj} < 0.05$ ) of categories in specific clusters. In each panel, the y-axis corresponds to  $-\log_{10}(\text{padj})$  such that more significant categories are displayed at higher values, and the x-axis indicates the number of genes in the input cluster list that generate the significant result. The horizontal lines mark a  $\text{padj} = 1e-04$  threshold. All significant functional enrichment results are provided in Supplementary Data S3, and additional similar plots for other functions referred to in the text are presented in Supplementary Fig. S2. The plots were generated by sub-setting all the significant functional categories using a search with the keywords indicated in the plot titles.

and to infer their functional significance; to detect differential usage of gene isoforms during metamorphosis; and to ascertain whether CpG methylation of DNA displays stage-specificity or any relationship with gene expression.

## Results

### Metamorphic shifts in transcription required for cellular energy metabolism and myogenesis

We sequenced complete mRNA transcripts using Oxford Nanopore long-read sequencing, obtaining 9–24 million reads per sample (average 17 million across the 15 samples). These describe a whole-organism transcriptome comprising 35,085 transcript isoforms derived from 11,728 gene loci, including 1466 transcripts derived from 734 potentially new loci (Fig. 1b, Supplementary Fig. S1). Principal components analysis of the gene-level transcript abundance data separated the five stages from each other in the first two principal components, clustering the P2 pupal time point samples together at a location on PC1 intermediate between the L5 and P6 samples (Fig. 1c). Grouping the data for all genes with significant expression differences between any two stages ( $\text{padj} < 0.05$ ,  $|\log\text{FC}| > 1$ , Supplementary Data S1) into co-expression profiles using hierarchical clustering identified eight clusters which define coordinated changes in expression taking place during metamorphosis (Fig. 1d, Supplementary Data S2).

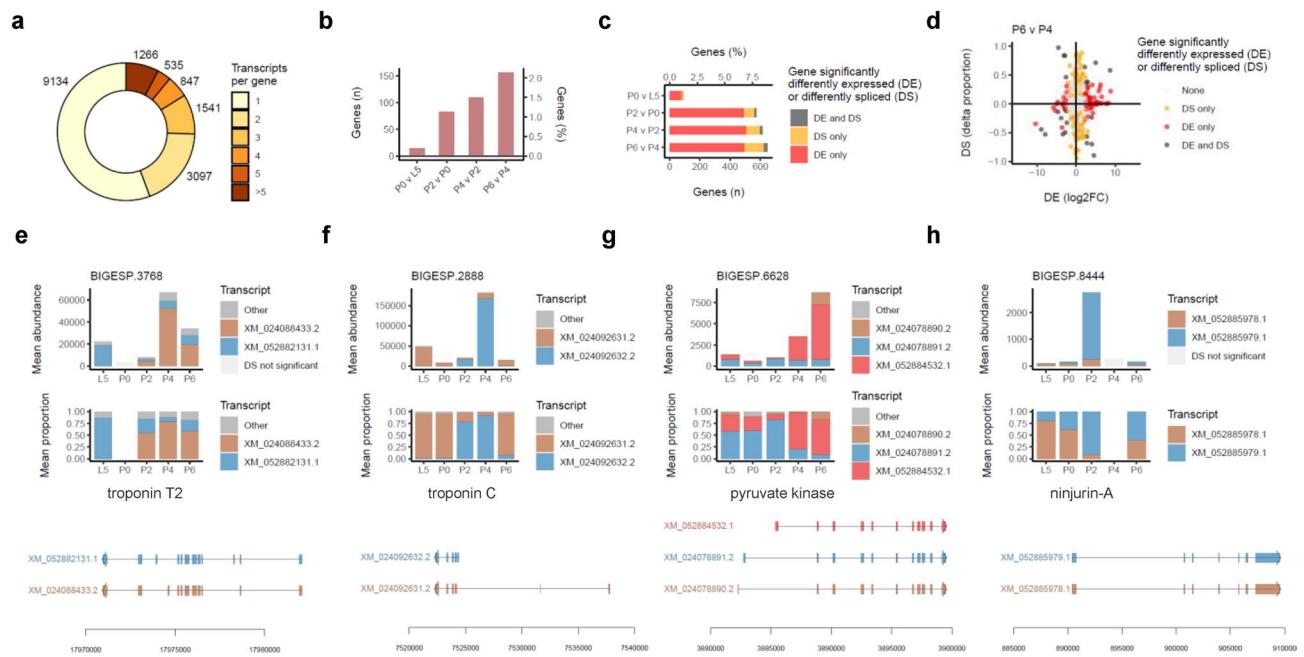
Gene co-expression typically reflects involvement in a shared biological process; for example, the production of components from the same pathway, protein complex or cell/tissue structure, regulation by the same transcriptional program, or a common functional role. Functional over-representation analysis of the gene clusters in Fig. 1d using a gene ontology and KEGG pathway<sup>12</sup> annotation produced by *eggnog-mapper v2*<sup>13</sup> indicates a progression from the use of fatty acid biosynthesis and lipid metabolism in the caterpillar stage L5 (cluster 7) towards an up-regulation in gene products required for muscle growth and energy generation in the pupal stages, P2–P6 (clusters 1, 3 and 6) (Fig. 1e, Supplementary Fig. S2, Supplementary Data S3). In particular, genes with functions required for the construction and functioning of mitochondria, cardiac processes, calcium ion signaling and metabolism, and formation of myelin are over-represented in the clusters showing up-regulation in P2–P6 (Fig. 1e, Supplementary Figs. S2 and S3, Supplementary Data S3). Hormone-related processes, including ecdysteroid and prostanoid biosynthesis, were enriched in the groups of genes most highly expressed in the L5 larval stage (cluster 7) or the mature pupal stage P6 (cluster 6), suggesting their greater importance at major developmental transitions (Supplementary Fig. S2). The expected up-regulation in genes encoding enzymes required for wing and body pigmentation was reassuringly observed in the late pupal stage samples (Supplementary Fig. S3), when coloration becomes visible.

### Differential isoform use during metamorphosis is prominent in genes for energy metabolism, and for muscle and neural development

We found that about 45% of the *B. anynana* gene loci expressed during metamorphosis encoded more than one mRNA transcript isoform (Fig. 2a). Over 20,000 previously unannotated, novel transcript isoforms were identified (Fig. 1b), originating from both previously known gene loci (3999 genes) and loci newly annotated in this study (734 genes). Antisense transcription to 652 previously known loci was also detected. The previously known genes that express novel transcripts are significantly enriched for functions associated with growth and primary metabolism, with gene ontology categories related to translation, respiration and protein-trafficking being the most significantly over-represented (Supplementary Fig. S4). Neither the completely novel loci nor the gene loci with antisense transcripts showed any significant functional enrichment.

Analysis of differential isoform expression identified significant changes in isoform usage in a total of 545 genes across all possible pairwise comparisons between the five target developmental stages, with 55% occurring in more than one comparison and the largest number involving the late pupal time point P6 (Fig. 2b, Supplementary Figs. S5 and S6, Supplementary Data S4). The genes exhibiting differential isoform usage between stages showed only a low level of overlap with the genes showing significant changes in expression (Fig. 2c, d, Supplementary Fig. S6). The differential splicing and expression of the genes in Fig. 2e–h was confirmed by qRT-PCR (Supplementary Fig. S7).

Genes exhibiting significant differential splicing in comparisons between the later stages of pupal development and stage P0 (P6 v P0, P4 v P0, P2 v P0) showed significant enrichment for functions related to ATP biosynthesis, glycolysis, and carbohydrate metabolism (Fig. 3, Supplementary Data S5). These include subunits of ATP synthase (BIGESP.3953, BIGESP.6855, BIGESP.7039), components of mitochondrial respiratory chain complexes (BIGESP.2612, BIGESP.7652, BIGESP.2959, BIGESP.9606), and key glycolytic enzymes including phosphofructokinase (BIGESP.302), pyruvate kinase (BIGESP.6628, Fig. 2g), fructose bisphosphate aldolase (BIGESP.4847), and also L-lactate dehydrogenase (BIGESP.13238) (Figs. 2 and 3). Significant differential isoform usage identified between other pairs of developmental stages included a number of genes encoding muscle filament fibres belonging to the troponin, tropomyosin and myosin families (Fig. 2 and Supplementary Fig. S8, Supplementary Data S4). Troponin T2 (BIGESP.3768) and troponin C (BIGESP.2888) exhibit both differential expression and differential splicing during metamorphosis, whereby total expression peaks during the later P4 and P6 pupal stages and there is a switch away from the isoforms that are predominantly used in the larval stage L5 (Fig. 2e, f). Genes encoding tropomyosin-1 (BIGESP.9197), myosin light chain 1 (BIGESP.5747), and myosin heavy chain (BIGESP.7889) are also both differentially spliced and differentially expressed during development, while myosin-2 essential light chain (BIGESP.5482) and myosin regulatory light chain sqh (BIGESP.4717) are differentially spliced but not differentially expressed (Supplementary Fig. S8). Other notable changes in isoform use include BIGESP.6531, encoding a putative angiotensin-converting enzyme required for the regulation of blood pressure in vertebrates<sup>14</sup>, and ninjurin-A (Fig. 2, Supplementary Fig. S8, Supplementary Data S4). Expression of the latter is strongly induced at the mid-point of the time course in the study, stage P2, and is from a family of cell–cell adhesion proteins whose vertebrate homologs have important functions in nerve



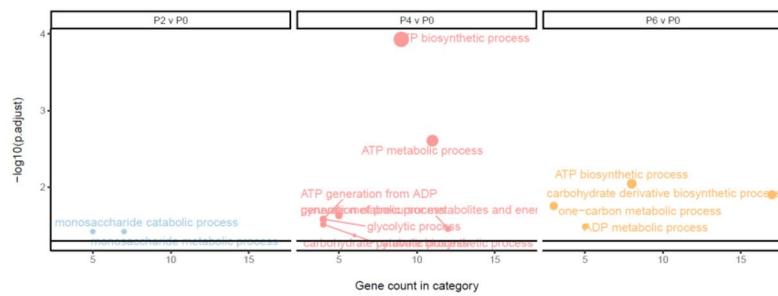
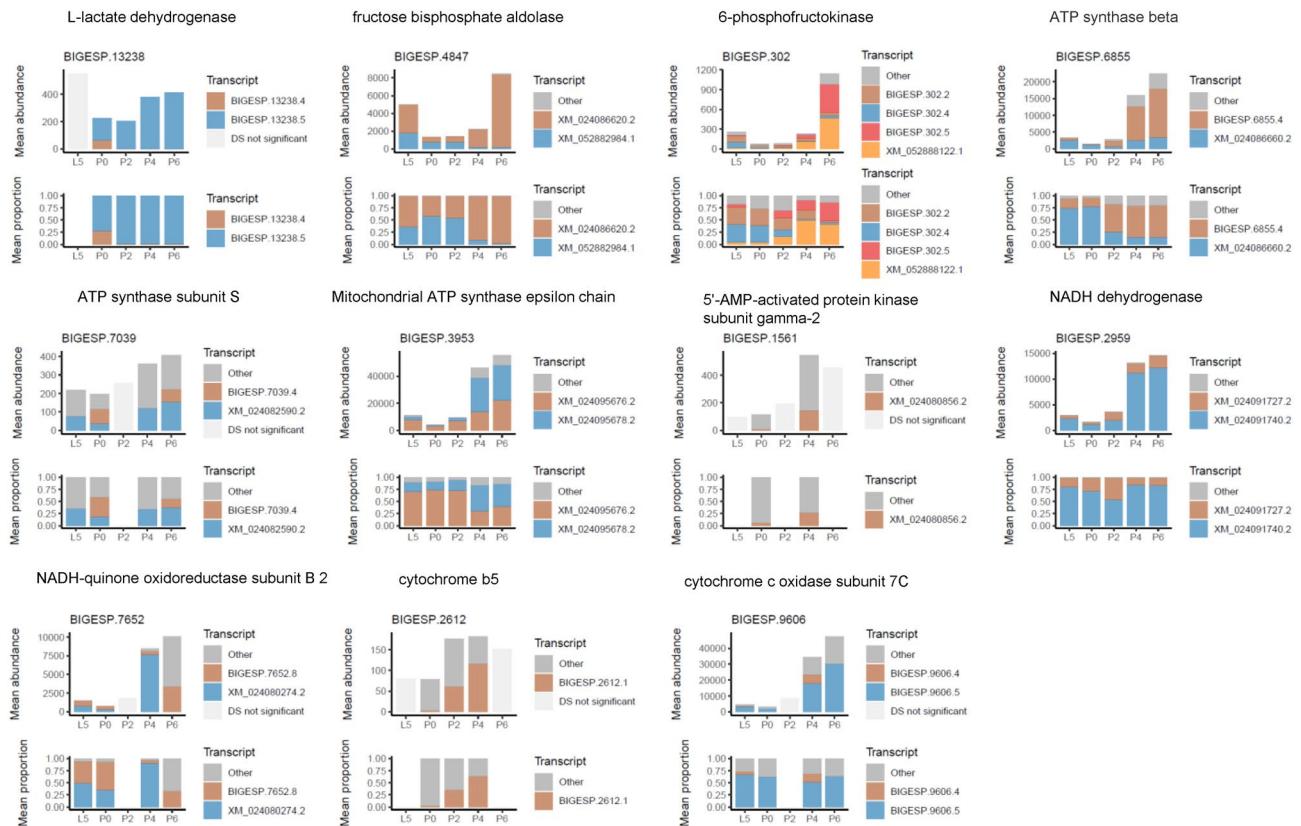
**Fig. 2.** Differential isoform use during metamorphosis. (a) Approximately 45% of genes in the *B. anynana* genome can generate more than one transcript, and potentially express specific isoforms at different developmental stages. (b) Genes exhibiting significantly different transcript splicing between consecutive developmental stages ( $\text{padj} < 0.05$ ,  $\text{delta isoform proportion} > 0.1$ ). The first y-axis provides the numbers of significant genes, the secondary y-axis shows the numbers as a percentage of the number of multi-exon genes in the genome, and the x-axis indicates the comparisons between developmental stages. The numbers for all pairwise comparisons are shown in Supplementary Fig. S5, and all significant genes are listed in Supplementary Data S4. (c) The genes exhibiting differential isoform usage between stages (DS) are largely distinct from the genes showing significant changes in expression (DE). Only multi-exon genes capable of being differentially spliced are included in this analysis, and only the comparisons between consecutive developmental stages in (b) are shown. Results for all comparisons are shown in Supplementary Fig. S6. (d) Scatterplot contrasting the changes in isoform usage (y-axis) with the changes in gene expression (x-axis) for the stage P6 v P4 comparison illustrating that differential splicing and differential expression largely affect different groups of genes (Supplementary Fig. S6 shows all comparisons between stages). (e–h) Changes in transcript abundance (top panel) and transcript isoform proportion (middle panel) across all the developmental stages for selected genes discussed in the text. For each gene shown, the lower panel illustrates the structure of the different isoforms. Only significant results are shown in the middle panel, where a zero value indicates no significant differential splicing and “Other” indicates the proportion of expression due to all other non-significant isoforms. In the upper, gene expression panel “DS not significant” corresponds to total gene expression for stages where differential splicing was not identified. The results were confirmed by qRT-PCR (Supplementary Fig. S7).

cell regeneration and repair, muscle and bone development and angiogenesis<sup>15–19</sup>. Intriguingly, the ninjurin-A homolog ninjurin-1 also exhibits a similar stage-specific up-regulation in expression, but at the later pupal stage P4 (Supplementary Fig. S8).

### Gene body methylation is associated with levels of gene expression

Analysis of CpG modifications in the DNA sequencing data revealed methylation in a mean of 2.95% of the detectable sites at each developmental stage, with 0.4% ( $+/- 0.016\text{sd}$ ) more than 50% methylated, and 97.05% ( $+/- 1.71\text{sd}$ ) completely unmethylated. CpG sites exhibiting more than 10% methylation were absent from intergenic regions, and most abundant over exons (Fig. 4a, Supplementary Fig. S9). Analysis of the average CpG site methylation over meta-exonic and meta-intronic regions indicates preferential methylation of the exons located towards the 5' end of transcripts (exons 1–4), with exons 2 and 3 typically showing the highest level (Fig. 4b, Supplementary Fig. S9). Methylation of introns was notably lower than exons and typically showed an uneven distribution, being highest towards upstream exon–intron junctions and decreasing towards the downstream intron–exon border. There were no notable differences in these trends between the developmental stages.

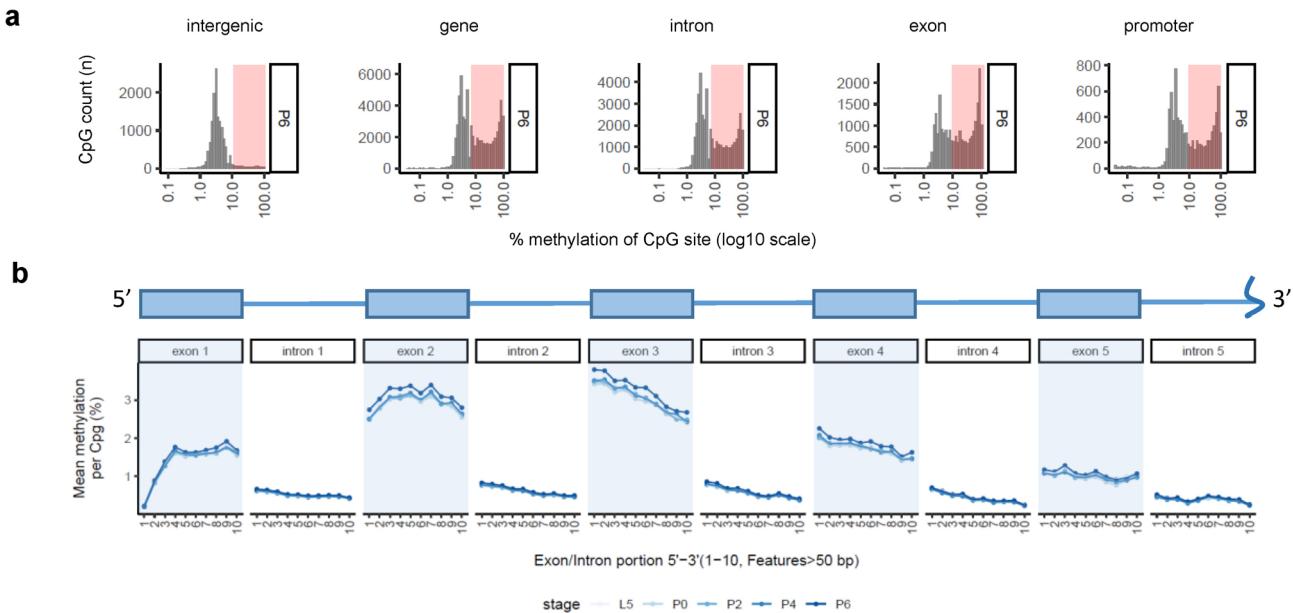
Low-level CpG site methylation (< 10%) was most prominent in the quintile group of genes with the lowest expression, with higher level methylation (> 10%) progressively increasing in frequency with increases in gene expression (Fig. 5a, Supplementary Fig. S9). For each developmental stage, the CpG sites that were 100% methylated were most frequent in the most highly expressed genes. This results in a low but significant positive correlation between the average CpG site methylation of a gene and its expression (Fig. 5b, Supplementary Fig.

**a****b**

**Fig. 3.** Genes exhibiting significant differential isoform use are enriched for functions associated with ATP biosynthesis, glycolysis and carbohydrate metabolism. **(a)** Significant functional categories identified in the differential expression results for P2 v P0, P4 v P0, and P6 v P0. In each panel, the y-axis corresponds to  $-\log_{10}(\text{padj})$  such that more significant categories are displayed at higher values, and the x-axis indicates the number of genes in the input list that generate the significant result. The horizontal lines mark a  $\text{padj} = 0.05$  threshold. All the significant categories are from the GO BP ontology. All significant functional enrichment results for all the clusters in panel are provided in data S5. The plots were generated from these results after excluding redundant categories related to purine nucleoside biosynthesis. **(b)** Genes exhibiting significant differential splicing between stage P0 and the later developmental stages from the enriched functional categories ATP biosynthesis, glycolysis, and carbohydrate metabolism. Panels are as described in the legend to Fig. 1.

**S9**). Two distinct sets of highly expressed genes are, however, observed in the data, a group of 1077 genes with a mean methylation in the medium to high ranges (i.e. 10–100%) and a contrasting group of 1857 genes with no detectable methylation (Fig. 5c, marked with red boxes), revealing that the relationship between methylation and expression is clearly complex. These two groups encode distinct sets of cellular functions. The methylated set is enriched for functions associated with macromolecules and energy metabolism (e.g. the ribosome, translation, RNA splicing, mitochondrial gene expression, protein transport and processing), while the non-methylated group shows enrichment for functions mostly related to small molecule metabolism (including prostanoid and fatty acid biosynthesis, vitamin metabolism, carboxylic acid catabolism) (Fig. 5d, e, Supplementary Data S6).

Regions of the genome that are significantly differentially methylated (dmrs) between any of the five time points are not frequent, with typically less than 50 dmrs between each pairwise comparison (Fig. 5f, Supplementary



**Fig. 4.** Methylation of CpG sites in the *B. anynana* genome is prominent over exonic gene regions and tends to be associated with more highly expressed genes. (a) The density (y-axis) of percentage methylation of individual CpG sites (x-axis, log10 scale) over different regions of the genome (intergenic regions, genes, introns, exons, promoters, as defined in the Methods section) for the sample P6 data (all five developmental stages were similar, as shown in Supplementary Fig. S10). CpG methylation in the 10–100% range is highlighted by the red rectangle. Unmethylated CpG sites (0% values) are not included in the plot. (b) Mean CpG site methylation over meta-exons (shaded blue) and meta-introns (unshaded) ordered by transcript position, 5' to 3'. Only results for exons 1 to 5 and introns 1 to 5 are shown (all are shown in Supplementary Fig. S10).

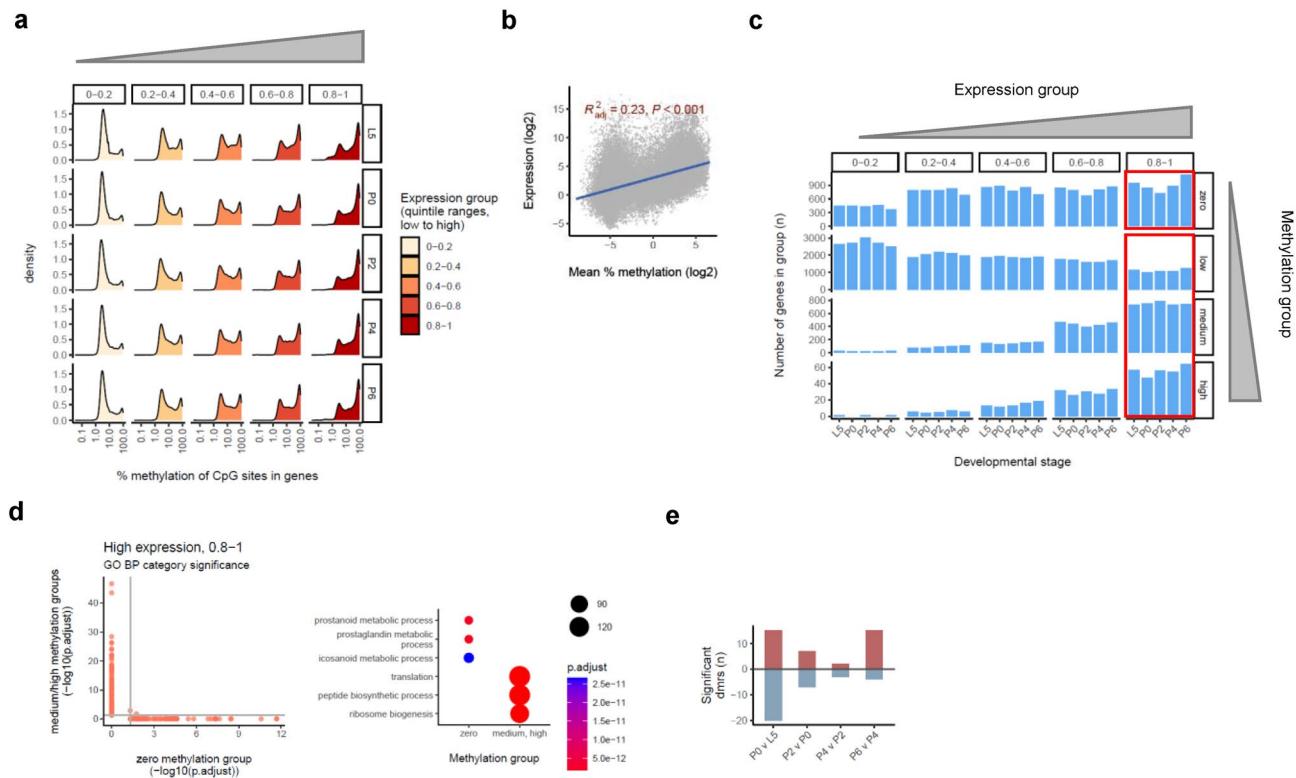
Fig. S9, Supplementary Data S7). The overlap between the identity of the genes in regions with differential methylation and those with either differential expression or differential splicing was very low, with only eight examples of the former and four of the latter (Supplementary Figs. S10 and S12, Supplementary Data S8). Thus, while DNA CpG methylation correlates with the strength of gene expression, it does not generally coincide with the up- or down-regulation of gene transcription during metamorphosis. No genes were found to be significant for all of differential expression, splicing, and methylation. Eight genes were both differentially methylated and differentially expressed. Of these, two identified between stages P4 v L5 both encode mitochondrial ribosomal proteins, MRPL11 and MRPS30 (BIGESP.10799 and BIGESP.9057, respectively (Supplementary Fig. S11).

In mammals, methylated CpG sites can be enzymatically oxidised to generate 5-hydroxymethylated (5Hm) CpG epigenetic marks with their own distinct influence on the regulation of gene expression<sup>20,21</sup>. During the metamorphosis of *B. anynana*, 5Hm of CpG sites was infrequent (only 0.036% of CpG sites exhibited higher than 10% 5Hm, and no sites were more than 50% modified; Supplementary Fig. S12), and no significant differentially 5-hydroxymethylated regions were detected in the comparisons between any of the developmental stages.

## Discussion

Butterfly metamorphosis describes a progression from a caterpillar feeding larval stage through to an adult capable of flight and reproduction, via a protected, encased and immobile pupal phase. This transition requires extensive changes to the organism's body plan to take place entirely dependent upon the nutrients acquired during larval growth. Many larval structures are therefore broken down and repurposed, while others, notably the imaginal discs that specify external structures of adult insects, undergo rapid growth and development<sup>1,22–24</sup>. These are both energy intensive anabolic processes, and our analysis reveals the step changes in gene expression and splicing associated with energy generation via glycolysis and respiration that accomplish this.

Gene products specifying the structure and function of mitochondria are notably up-regulated from the P2 pupal stage onwards, together with those which in vertebrates are associated with skeletal muscle, myelin, and cardiac muscle formation. The concomitant increases associated with muscle and mitochondria are consistent with the requirement for mitochondria to power muscle activity. Large and numerous mitochondria fill some 30–40% of insect flight muscle fibres, and typically occupy a third of the volume of cardiac myocytes<sup>25,26</sup>. The increases in transcription of genes encoding components of muscle fibres during pupation is accompanied by a gradual decrease in expression of an ortholog of myostatin (Supplementary Fig. S8), known to be progressively and positively associated with sarcopenia due to ageing or microgravity, and to decreased muscle mass and strength, and skeletal insertion sites<sup>27–31</sup>. As with other insects, the musculature of the butterfly larval stage is suited for slow crawling, while the adult requires powerful thoracic and abdominal muscles that enable flight. In *Drosophila*, extensive remodeling of the skeletal muscles takes place during pupal development and is accompanied



**Fig. 5.** Methylation of CpG sites in the *B. anynana* tends to be associated with more highly expressed genes. (a) The density (y-axis) of percentage methylation of individual CpG sites (x-axis, log10 scale) in genes grouped according to the strength of their expression (y-axis columns) in each of the developmental stages (x-axis rows). Results obtained using the mean % CpG site methylation per gene are similar (Supplementary Fig. S10). (b) Significant positive correlation between expression of a gene (y-axis, log2 scale) and its mean % CpG methylation (x-axis, log2 scale). Data points for all genes in all five developmental stages is shown. Correlation plots for the data from individual stages is shown in Supplementary Fig. S10. (c) Genes subdivided into expression groups (x-axis columns), gene-level mean CpG methylation groups (y-axis rows) and developmental stages (x-axis), showing both increasing numbers of highly methylated genes in the highly expressed group and a population of highly expressed genes that are unmethylated. The upper red box defines 1857 genes that are highly expressed but unmethylated, and the lower red box defines 1077 highly expressed genes with a mean methylation in the medium to high ranges (i.e. 10–100%). (d) Comparison of the significantly enriched GO BP categories represented by the zero methylated highly expressed genes (x-axis) with the medium-highly methylated highly expressed genes (y-axis) from the red boxed genes in (c) above. The left panel presents all significant categories identified using axes indicating the  $-\log_{10}(p_{adj})$  such that more significant categories are displayed at higher values. The horizontal and vertical lines mark a  $p_{adj} = 0.05$  threshold for each axis. The top three most highly significant GO BP categories for each group of input genes are shown in the right panel and all significant categories are listed in Supplementary Data S6. (e) Numbers of significantly differentially methylated regions identified between consecutive developmental stages using DSS ( $p_{adj} \leq 0.05$ ). Directionality is shown according to the first stage named in the pair minus the second stage, such that regions e.g. hypermethylated in stage P6 compared to L5 are coloured red in the P6 v L5 pair and those hypomethylated are coloured blue. The results for all inter-stage comparisons are shown in Supplementary Fig. S10.

by modification of the cardiac tube<sup>32–34</sup>. Leg muscle development is accompanied by the formation of myelin in the associated nerves which may enable high conductance velocity for fast and well-tuned locomotion in adult flies<sup>35</sup>. Remodelling of the cardiac tube may help meet the changes in demand on the circulatory system in the more active airborne adult. For all functional interpretations, we rely on a computationally generated annotation based on homology detection and inference of preservation of function across taxa which consequently may contain some functional predictions that turn out to be imperfect. The general trends observed using an annotation derived in this way are however robust to misassignments in individual genes.

Lipid metabolism in insects is predominantly associated with the fat body, which is regarded as a hub for the regulation of insect physiology via the sensing of both nutritional and hormonal signals and by responding with the mobilization of stored nutrient molecules<sup>23,36,37</sup>. The energy accumulated by lipid storage in the fat body during larval growth supports the growth and differentiation of adult structures during pupal development, and is released during the regulated remodeling of the fat body that takes place during metamorphosis<sup>23</sup>. Hormonal regulation is intimately linked with insect metamorphosis and nutrition, and although our experimental design

lacks the refinement to observe the production of, and responses to, the pulses of ecdysteroid and juvenile hormones that initiate many of the major developmental transitions, ecdysteroid metabolism was identified as a significant functional category in the clusters of genes showing higher expression in the larval stage L5 (Supplementary Fig. S2).

A striking feature of the changes in gene expression identified during *B. anynana* metamorphosis is the sequential up-regulation of two ninjurin homologs, 2 and 4 days after the onset of pupation, ninjurin-A (BIGESP.8444) and ninjurin-1 (BIGESP.10341), respectively (Fig. 2, Supplementary Fig. S8). Transcription of ninjurin family genes is up-regulated in response to injury, infection or stress in humans, mice, and *Drosophila*, and they have been shown to have roles in cell–cell adhesion, nonapoptotic cell death, peripheral nerve cell regeneration, and muscle growth and differentiation<sup>15–19,38</sup>. Ninjurin expression is also activated during metamorphosis of *Drosophila*<sup>5</sup>, and the observed orchestrated up-regulation of the ninjurin homologs at defined stages in *B. anynana* pupal development is intriguing, particularly since it is largely coincident with the observed changes in muscle development discussed above.

In addition to the regulation of the strength of gene expression, regulation of the use of different transcript isoforms also seems to play a role in metamorphosis, notably in relation to the processes of muscle growth and energy generation discussed above. Troponin T2, troponin C, myosin light chain I and myosin heavy chain all displayed a significant changeover in the use of one isoform in the larval stage to a differently spliced transcript isoform in one or more of the later pupal stages (Fig. 2, Supplementary Fig. S8). Alternative splicing of troponins and myosins in other organisms has been reported to occur in a cell type or tissue type dependent manner to generate variants that can exhibit different mechanical properties, intracellular localization, and functions<sup>39–41</sup>. Similar stage-dependent differential isoform usage was also evident in key glycolytic enzymes and respiratory chain complex subunits, including pyruvate kinase and phosphofructokinase (Figs. 2 and 3). In humans, modulation of the ratio between two isoforms of pyruvate kinase (M1 and M2) controls a preference for either aerobic glycolysis or complete glucose oxidation in the Krebs cycle, demonstrating functional and metabolic consequences for differential isoform use<sup>42</sup>. Three different isoforms of the phosphofructokinase enzyme, described as a gatekeeper for the control of glycolytic flux, are also expressed in humans, and show both different tissue specificities and biochemical activities<sup>43</sup>. The significant alternative splicing events observed here suggest the operation of alternative splicing programs that contribute to the coordination of the metabolic and developmental changes taking place during progression from larva to emergent adult butterfly. In a time-course studying the complexity of gene expression during *Drosophila* metamorphosis, the adult gonads, head and thorax tissues, particularly the brain, showed the greatest diversity of transcript isoforms, while embryos and larvae had fewer isoforms per gene on average, suggesting that changes in tissue complexity during metamorphosis may require the increased functional capacity afforded by increased differential transcript usage<sup>6</sup>. In this context, the identification of > 20,000 transcript isoforms not present in the published genome annotation, several hundred from potentially new gene loci, is notable. While we cannot exclude the possibility that some of these may be technical artefacts arising from preparation of the sequencing libraries, it appears that the deep sequencing of transcriptomes from different points in the life cycle are revealing more of the full repertoire of isoforms encoded by the genome of *B. anynana*. This is also facilitated by the technology used—long read sequencing of intact mRNA transcripts—which enables greater coverage of splice junction sequences than short read sequencing approaches. Overall, we found a limited overlap between genes differentially expressed between the developmental stages and genes exhibiting differential transcript utilization during development (Fig. 2c, d and Supplementary Fig. 5). In a study of gene expression during *Drosophila* metamorphosis, Ozerova et al. (2025)<sup>44</sup> similarly describe a modest overlap between these two types of genes, suggesting that it may be a common feature of insect metamorphosis. Decoupling gene expression from isoform usage could provide two independent layers of regulatory control that may enable cells/tissues to respond very precisely to developmental cues or environmental changes during metamorphosis. Methylation of CpG dinucleotides in the promoter regions of chromosomal DNA contributes to the regulation of gene expression in humans and other mammals, where it tends to have a repressive effect<sup>45–47</sup>. In *B. anynana*, we detected some degree of cytosine methylation of approximately 3% of the CpG sites observed during the developmental time course, but methylation was highest in intra-genic DNA (often referred to as gene body methylation), particularly at exons near the 5' end of transcripts, and shows a general positive correlation with the strength of gene expression (Fig. 4, Supplementary Fig. S9)<sup>7,48,49</sup>. The machinery for the modification and sensing of methylated CpG sites on DNA is not well conserved between insects and mammals<sup>50–52</sup>, and our observation that methylation of intra-genic CpG sites is prevalent in *B. anynana* is consistent with previous observations in other insects<sup>52–55</sup>. Gene body CpG methylation also occurs in humans and other mammals where it has been proposed to help specify the inclusion/exclusion of exons in gene transcripts<sup>45–47</sup>. A similar regulatory role has also been proposed for the intra-genic methylation observed in insect species<sup>52</sup>, but a consistent, definitive picture is yet to emerge. A genome-wide comparative analysis of gene expression and DNA methylation in caterpillars and adults of the tobacco hornworm *Manduca sexta* supports a role for epigenetic modification by CpG methylation in regulating the transcriptional differences between the two stages<sup>8</sup>. Moreover, potentially functional methylome shifts identified in the butterfly *Vanessa cardui* have suggested that DNA methylation forms part of this species' response to different environmental conditions during larval development<sup>7</sup>. However, the latter study also noted the absence of a relationship between changes in the CpG methylation status of a gene and changes in its expression<sup>7</sup>, and a comparative analysis between eight developmental stages of the honey bee also concluded that fluctuations in CpG methylation did not result in changes in gene expression<sup>48</sup>. In *B. anynana*, we observed only a relatively low number of significant changes in the methylation of genomic regions between the developmental stages, and little overlap between changes in methylation and changes in expression or transcript splicing (Supplementary Fig. S10).

While high levels of gene expression in *B. anynana* clearly do not exclusively associate with high levels of gene body methylation (Fig. 5a), the observation that the group of highly expressed genes that are also highly methylated are significantly enriched for functions related to ribosome biogenesis and translation (Fig. 5) agrees with results from other studies that linked high methylation to the expression of housekeeping genes<sup>7,56</sup>. Whether this indicates that methylation could play a role in maintaining open chromatin accessibility at highly expressed genes, or that the open structure is permissive for DNA methylation remains to be clarified. It seems unlikely that methylation status is only a passive reporter of the location of open chromatin or the passage of RNA polymerase, and it may play a role in maintaining stable expression of certain genes. In humans, the methylation of gene bodies has been reported to play a number of positive roles in transcription, including regulating alternative splicing, ensuring correct splicing and translation, and ensuring stable and orderly transcription by suppressing abnormal ectopic gene transcription<sup>57</sup>. Moreover, gradual oxidation of methylcytosine residues located within the bodies of transcriptionally active tissue-specific genes in mouse and human to 5-hydroxymethylcytosine has recently been shown to occur with aging, where it restricts the magnitude of gene expression changes with age, potentially preserving homeostatic expression of tissue-specific genes, and correlates with age-related alternative splicing events<sup>20</sup>. These reported functions in mammals, though still debated, suggest that there may be more to gene body methylation in insects than is currently understood. Stronger associations between DNA methylation and changes in transcript abundance or use may exist in specific tissues or cell types of *B. anynana*.

This study defines the growth phase-dependent changes in gene expression, gene splicing and DNA methylation taking place during metamorphosis of *B. anynana*, and provides a set of data that raises a number of avenues for future study. In a broader context, it also provides the basis for functional genomics studies investigating the impact of temperature on the butterfly life-cycle, thereby providing a useful route towards a better understanding of some of the ecological consequences of global climate change. We deliberately designed our study to measure the changes in gene expression taking place during the *B. anynana* life cycle at the level of the whole organism to be able to characterize the major trends. Not all tissues are present to the same readily accessible extent throughout every stage during metamorphosis, limiting the application of dissection and analysis of specific tissues to study the larva to adult transition in its entirety. Tissue dissection would almost certainly result in loss of information. Sampling the whole organism thus provided the most straightforward initial method to capture data reflecting the developmental program as a whole, generating an unbiased view of the changes in the gene expression landscape that occur during this complex transitional process. This approach however, comes with the built-in limitation of generating data that is an average across all tissues present in the sample, with limited opportunity for insight into spatial or tissue-specific transcriptional or DNA methylation programs. In addition, for practical reasons, only five time points were sampled during the life cycle, generating relatively coarse-grained datasets that contained only one sample from the larval stage. The sampling of more time points, particularly from the pre-pupal phase, and the development of methods to dissect out organs and tissues of interest from all of the growth stages under study would be useful future goals for further defining butterfly metamorphosis. In addition, quantification of changes in the abundance of microRNAs, a molecular species that is of increasing interest in insect developmental biology but absent from the transcriptome data reported here, would also be desirable<sup>58–63</sup>. The DNA methylation analysis we report provides only a partial view of the CpG sites in the genome, with DNA sequencing to a coverage depth of approximately 10× per sample generating replicate data of sufficient depth from all growth stages of interest for 40% of these dinucleotides. Nevertheless, at the whole-organism level, CpG methylation of DNA was found to display only limited stage-specificity, and was associated with levels, not dynamic changes, in gene expression. It remains to be seen whether this also holds true in specific tissues or cell types.

## Materials and methods

### Insect rearing

We used *Bicyclus anynana* butterflies from a captive population maintained at 27 °C with 65% humidity, and a 12 h light/12 h dark photoperiod<sup>11</sup>. Larvae were fed ad libitum with young maize plants. In the 3<sup>rd</sup>/4<sup>th</sup> instar, 200 male larvae were selected and maintained in the same cage. Three specimens were collected from each of the following stages: 5<sup>th</sup> and final instar larvae (L5), and pupae on the day of pupation (P0), as well as 2 (P2), 4 (P4) and 6 (P6) days post-pupation (Fig. 1a). The P6 stage corresponds to what are essentially a fully formed butterfly that typically ecloses out of the pupal case on that same day. Each individual was placed inside a microtube, killed by flash freezing (30 s in liquid nitrogen), and subsequently stored at –80 °C until further processing.

### Sample preparation

Disruption and homogenization of tissue samples were performed using TissueLyser (Qiagen) as per the manufacturer's instructions. The tissue homogenates were divided to achieve extraction of DNA and RNA from the same individual sample. To prevent degradation, RNA was extracted immediately after homogenization. RNA was extracted and purified using the QIAzol Lysis Reagent and Qiagen RNeasy Kits following the manufacturer's protocols. DNA was extracted as per the manufacturer's instructions using the Qiagen DNeasy® Blood & Tissue Kit.

### Sample sequencing

RNA sequencing libraries representing full-length transcripts were prepared from total RNA using the Oxford Nanopore Technology (ONT) PCR-cDNA Barcoding Kit (SQK-PCB111-24) according to the manufacturer's protocol. A 15–25 fmol aliquot of each RNAseq library was sequenced to a read depth of 9–24 million reads per sample (average 17 million) on the ONT GridION platform using minion FLO-MIN106D flow cells. The DNA libraries were prepared using ONT's Native Barcoding Kit 24 V14 (SQK-NBD114.24) following the

manufacturer's protocol, and sequenced on the ONT GridION platform using 10–20 fmol of each library loaded on to minion FLO-MIN114 flow cells, collecting 7.5–14  $\times$  coverage of the genome per sample (average 10.7x).

### RNAseq data analysis

Basecalling and demultiplexing of raw signal data was performed using the GridION minKNOW software (version 22.07.9) using guppy 6.3.9 with the high accuracy basecalling model. Reads were quality filtered with the default q-score setting of 9. The basecalled fastq files passing the filters were processed using pycruncher (version 2.7.2) with option '-k PCS111' to identify, orient and trim full-length nanopore cDNA reads, and minimap2 (version 2.24-r1122) to align to the reference genome GCF\_947172395.1\_ilBicAnyn1.1<sup>64</sup> using part of the epi2me labs wf-transcriptomes nextflow workflow (version 0.1.13). The reference genome and annotation for GCF\_947172395.1\_ilBicAnyn1.1 was downloaded from NCBI RefSeq in August 2023. To identify and quantify novel transcript splicing compared to the reference genome, the sample-level alignment bam files were analysed using espresso (version 1.4), increasing the -read\_num\_cutoff setting to 4 to require a minimum of 4 perfect reads to call de novo detected candidate splice junctions. This procedure in espresso produced an updated unified annotation based on the transcripts detected in all samples. This was subsequently merged with the existing annotation using stringtie (version 2.1.7)<sup>65</sup>, generating new unique identifiers beginning with a prefix "BIGESP". The updated annotation is available from <https://doi.org/10.5281/zenodo.15837799>, and was compared to the reference annotation using gffcompare (version 0.11.2)<sup>66</sup>. The script agat\_sp\_extract\_sequences.pl from agat (version 1.2.0) was used to generate a transcriptome.fasta file of all transcript sequences present in the new annotation. Transcript abundances in each sample were counted by first aligning the pycruncher RNA-seq reads to the transcriptome mapping index using minimap2 (version 2.24-r1122) with input options '-ax splice -uf -p 1.0', then quantifying with salmon (version 1.10.2) using the quant\_-ont alignment model for ONT long reads<sup>67</sup>. Differential gene expression was quantified using DESeq2<sup>68</sup> with the salmon transcript abundances as input, and differential isoform usage was analysed using rMATS-long (version 1.0.0) with the espresso values for expression of detected isoforms as input. In contrast to the analysis of changes in genes expression which combines the abundances from all transcripts, for the analysis of transcript isoform usage, the abundance of a particular transcript isoform for a specific gene is expressed as a proportion of the total of the abundances of all transcripts for that gene (ranging from 1 to many), generating values limited to the range between 0 (not expressed at all) and 1 (only this isoform expressed). This captures the relative use of each isoform transcribed from each gene, independent from the magnitude of its expression. Significant genes in the DESeq2 pipeline were defined as those with a Benjamini and Hochberg corrected *p*-value  $<= 0.05$  when tested against a null hypothesis requiring a minimum absolute log<sub>2</sub> foldchange in expression of at least 1 (denoted in the text as padj  $<= 0.05$ ,  $|\log FC| > 1$ ). Significant transcript isoforms from the rMATS-long analysis were defined as those with a Benjamini and Hochberg corrected *p*-value  $<= 0.05$  and exhibiting a change in isoform proportion  $> 0.1$ . For data integration to assess the relationship between differential expression, differential methylation and differential splicing, only the data for genes theoretically capable of being differentially spliced (encoding more than one transcript isoform) were used.

### DNAseq data analysis

Raw signal data was collected and samples demultiplexed using the GridION minKNOW software (version 22.07.9). Modified basecalling and genome alignment was repeated off-line using the epi2me labs wf-basecalling nextflow workflow (version 0.7.1, running dorado version 0.3.0 and minimap2 version 2.26-r1175) with the models -basecaller\_cfg dna\_r10.4.1\_e8.2\_400bps\_hac@v4.1.0 and -remora\_cfg dna\_r10.4.1\_e8.2\_400bps\_hac@v4.1.0\_5mCG\_5hmCG@v2. The basecalled cram alignment files were analysed for CpG site methylation and 5-hydroxymethylation modifications using modkit (version 0.1.6), aggregating data across both strands using the -combine-strands option. Differential modification of genome regions was analysed in R using the DSS package (version 2.42.0)<sup>69</sup>, requiring a CpG site to be covered by at least 4 reads to be considered in the analysis. Genome-wide DNA modification profiles were analysed using the R package methylKit<sup>70</sup> (version 1.34), and genome coverage was assessed using deitools (version 3.5.1)<sup>71</sup>.

### Analysing DNA methylation across regions of the genome and in relation to gene expression

To assess CpG site methylation in different functional regions of the genome gene, exon and intron regions were obtained from the start and end coordinates for these features listed in the updated genome annotation. Promoter regions were defined as 2Kbp regions from 1500 bp upstream to 500 bp downstream of the start location for each gene in the annotation. Intergenic regions were defined as those at least 2Kbp from any annotated gene. To analyse CpG site methylation across exons in transcripts, meta exons were constructed by first combining each exon according to its position in the annotated transcript (e.g. an exon number 1 group, an exon number 2 group, an exon number 3 group, etc.) and then dividing these 5' to 3' into 10 portions of equal length numbered 1 to 10 (e.g. exon number 1, parts 1–10; exon number 2, parts 1–10). Only exons with an original length greater than 50 bp were included. The mean of the methylation proportion values for all CpG sites in each portion for all exon numbers were then calculated with the data for each developmental stage. A similar process was applied for analysis of methylation across introns.

To analyse the relationship between gene body methylation and gene expression, length-corrected normalised gene expression values were first generated for each developmental stage by dividing the DESeq2 rlog normalised gene-level abundance data for each gene by the length of each gene according to the formula  $\log_2(\text{abundance}) - \log_2(\text{length})$ . For each stage, genes were then divided into 5 expression groups of approximately equal size based on the quintile values. The groups generated were designated 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0 and represent lowest to highest expression quintile groups for each stage. The frequency of percentage methylation

of the individual CpG sites in all the genes in each expression group for each of the developmental stages was then visualised using density plots.

### Other data analysis

To generate a gene ontology and KEGG pathway annotation for the genome, an automated Diamond database search and functional annotation of the gene and transcript models generated by espresso was performed using eggNOG-mapper-v2 (version 2.1.9, Diamond version 2.0.15)<sup>13</sup>, constraining the orthologous groups from which to retrieve annotations to the *Eukaryota* taxon and specifying insects as the target taxa from which to transfer annotation. This annotation, and the eggNOG-mapper output, is available from <https://doi.org/10.5281/zenodo.15837799>, and, like all computational homology-based annotations, comes with the caveat that it may contain some functional predictions that turn out to be incomplete or inaccurate. The identity of the majority of individual genes (15,114/15,858, Supplementary Fig. 1F) was obtained from the reference genome GCF\_947172395.1\_ilBicAny1.1<sup>64</sup>. Unsupervised clustering of gene expression, transcript expression and DNA methylation data to assess the inter-relationship and the reproducibility of biological replicates was performed using principal components analysis (PCA). The R package clusterProfiler<sup>72</sup> was used for functional over-representation analysis of groups of significant genes of interest, applying a significance threshold of 5% to the Benjamini and Hochberg corrected p-values to select significantly over-represented functional categories. Hierarchical clustering of data was performed in R using the pheatmap package, scaling the data to produce row-wise z-scores before clustering<sup>73</sup>.

### qRT-PCR

qRT-PCR was performed using standard protocols based on cDNA synthesis using random primers followed by quantification using the Sensi Mix SYBR® Low-ROX kit (QT625-20, Bioline) according to the manufacturer's guidelines. Gene transcript levels were normalized to housekeeping genes selected from the RNAseq data as being stably expressed across the developmental stages under study (calpain-B (BIGESP.9282) and casein kinase I (CK1, BIGESP.12112)), and data analysis was carried out using the comparative delta-delta Ct method. Primer sequences are listed in Supplementary Table S1.

### Data availability

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE276476 and GSE276477, and are available from the corresponding author upon reasonable request.

### Code availability

Code used in this study is available from zenodo and can be accessed via the link <https://doi.org/10.5281/zenodo.15837799>.

Received: 22 April 2025; Accepted: 3 September 2025

Published online: 13 October 2025

### References

1. Rolff, J., Johnston, P. R. & Reynolds, S. Complete metamorphosis of insects. *Philos. Trans. R Soc. Lond B Biol. Sci.* **374**, 20190063. <https://doi.org/10.1098/rstb.2019.0063> (2019).
2. Truman, J. W. The evolution of insect metamorphosis. *Curr. Biol.* **29**, R1252–R1268. <https://doi.org/10.1016/j.cub.2019.10.009> (2019).
3. Arbeitman, M. N. et al. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275. <https://doi.org/10.1126/science.1072152> (2002).
4. Mod, E. C. et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797. <https://doi.org/10.1126/science.1198374> (2010).
5. Graveley, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479. <https://doi.org/10.1038/nature09715> (2011).
6. Brown, J. B. et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399. <https://doi.org/10.1038/nature12962> (2014).
7. Boman, J. et al. Environmental stress during larval development induces DNA methylation shifts in the migratory painted lady butterfly (*Vanessa cardui*). *Mol. Ecol.* **32**, 3513–3523. <https://doi.org/10.1111/mec.16957> (2023).
8. Gegner, J., Vogel, H., Billon, A., Förster, F. & Vilcinskas, A. Complete metamorphosis in *manduca sexta* involves specific changes in DNA methylation patterns. *Front. Ecol. Evol.* **9**, 646281. <https://doi.org/10.3389/fevo.2021.646281> (2021).
9. Tian, S. & Monteiro, A. A transcriptomic atlas underlying developmental plasticity of seasonal forms of *Bicyclus anynana* butterflies. *Mol. Biol. Evol.* **39**, 126. <https://doi.org/10.1093/molbev/msac126> (2022).
10. Pruijscher, P., Lehmann, P., Nylin, S., Gotthard, K. & Wheat, C. W. Extensive transcriptomic profiling of pupal diapause in a butterfly reveals a dynamic phenotype. *Mol. Ecol.* **31**, 1269–1280. <https://doi.org/10.1111/mec.16304> (2022).
11. Brakefield, P. M., Beldade, P. & Zwaan, B. J. The African butterfly *Bicyclus anynana*: a model for evolutionary genetics and evolutionary developmental biology. *Cold Spring Harb. Protoc.* **2009**, pdb emo122. <https://doi.org/10.1101/pdb.emol22> (2009).
12. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677. <https://doi.org/10.1093/nar/gkae099> (2025).
13. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829. <https://doi.org/10.1093/molbev/msab293> (2021).
14. Bader, M., Steckelings, U. M., Alenina, N., Santos, R. A. S. & Ferrario, C. M. Alternative renin-angiotensin system. *Hypertension* **81**, 964–976. <https://doi.org/10.1161/HYPERTENSIONAHA.123.21364> (2024).
15. Araki, T. & Milbradt, J. Ninjurin, a novel adhesion molecule, is induced by nerve injury and promotes axonal growth. *Neuron* **17**, 353–361. [https://doi.org/10.1016/s0896-6273\(00\)80166-x](https://doi.org/10.1016/s0896-6273(00)80166-x) (1996).

16. Kny, M. et al. Ninjurin1 regulates striated muscle growth and differentiation. *PLoS ONE* **14**, e0216987. <https://doi.org/10.1371/journal.pone.0216987> (2019).
17. Liu, K., Wang, Y. & Li, H. The role of Ninjurin1 and its impact beyond the nervous system. *Dev. Neurosci.* **42**, 159–169. <https://doi.org/10.1159/000512222> (2020).
18. Tomita, Y. et al. Ninjurin 1 mediates peripheral nerve regeneration through Schwann cell maturation of NG2-positive cells. *Biochem. Biophys. Res. Commun.* **519**, 462–468. <https://doi.org/10.1016/j.bbrc.2019.09.007> (2019).
19. Zhang, S. et al. An MMP liberates the Ninjurin A ectodomain to signal a loss of cell adhesion. *Genes Dev.* **20**, 1899–1910. <https://doi.org/10.1101/gad.1426906> (2006).
20. Occean, J. R. et al. Gene body DNA hydroxymethylation restricts the magnitude of transcriptional changes during aging. *Nat. Commun.* **15**, 6357. <https://doi.org/10.1038/s41467-024-50725-y> (2024).
21. He, B. et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat. Commun.* **12**, 4249. <https://doi.org/10.1038/s41467-021-24425-w> (2021).
22. Hall, M. J. R. & Martin-Vega, D. Visualization of insect metamorphosis. *Philos. Trans. R Soc. Lond B Biol. Sci.* **374**, 20190071. <https://doi.org/10.1098/rstb.2019.0071> (2019).
23. Li, S., Yu, X. & Feng, Q. Fat body biology in the last decade. *Annu. Rev. Entomol.* **64**, 315–333. <https://doi.org/10.1146/annurev-ento-011118-112007> (2019).
24. Truman, J. W. & Riddiford, L. M. The evolution of insect metamorphosis: a developmental and endocrine view. *Philos. Trans. R Soc. Lond B Biol. Sci.* **374**, 20190070. <https://doi.org/10.1098/rstb.2019.0070> (2019).
25. Nguyen, B. Y. et al. Mitochondrial function in the heart: the insight into mechanisms and therapeutic potentials. *Br. J. Pharmacol.* **176**, 4302–4318. <https://doi.org/10.1111/bph.14431> (2019).
26. Ellington, C. P. Power and efficiency of insect flight muscle. *J. Exp. Biol.* **115**, 293–304. <https://doi.org/10.1242/jeb.115.1.293> (1985).
27. Amthor, H. et al. The regulation and action of myostatin as a negative regulator of muscle development during avian embryogenesis. *Dev. Biol.* **251**, 241–257. <https://doi.org/10.1006/dbio.2002.0812> (2002).
28. Elkasrawy, M. N. & Hamrick, M. W. Myostatin (GDF-8) as a key factor linking muscle mass and bone structure. *J. Musculoskelet. Neuronal Interact.* **10**, 56–63 (2010).
29. Tarantino, U. et al. Effects of Simulated Microgravity on Muscle Stem Cells Activity. *Cell Physiol Biochem* **54**, 736–747 (2020). <https://doi.org/10.33594/000000252>
30. Tarantino, U. et al. Saropenia: a histological and immunohistochemical study on age-related muscle impairment. *Aging Clin. Exp. Res.* **27**(Suppl 1), S51–60. <https://doi.org/10.1007/s40520-015-0427-z> (2015).
31. Zhao, Z., Yan, K., Guan, Q., Guo, Q. & Zhao, C. Mechanism and physical activities in bone-skeletal muscle crosstalk. *Front Endocrinol (Lausanne)* **14**, 1287972. <https://doi.org/10.3389/fendo.2023.1287972> (2023).
32. Schaub, C., Marz, J., Reim, I. & Frasch, M. Org-1-dependent lineage reprogramming generates the ventral longitudinal musculature of the Drosophila heart. *Curr. Biol.* **25**, 488–494. <https://doi.org/10.1016/j.cub.2014.12.029> (2015).
33. Molina, M. R. & Cripps, R. M. Ostia, the inflow tracts of the Drosophila heart, develop from a genetically distinct subset of cardiac cells. *Mech. Dev.* **109**, 51–59. [https://doi.org/10.1016/s0925-4773\(01\)00509-3](https://doi.org/10.1016/s0925-4773(01)00509-3) (2001).
34. Bilecky, C., Blotz, B. & Cripps, R. M. Drosophila as a model to understand second heart field development. *J. Cardiovasc. Dev. Dis.* **10**, 494. <https://doi.org/10.3390/cbdd10120494> (2023).
35. Rey, S. et al. Glial-dependent clustering of voltage-gated ion channels in Drosophila precedes myelin formation. *Elife* **12**, e85752. <https://doi.org/10.7554/elife.85752> (2023).
36. Arrese, E. L. & Soulages, J. L. Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol.* **55**, 207–225. <https://doi.org/10.1146/annurev-ento-112408-085356> (2010).
37. Song, Y. et al. The key role of fatty acid synthase in lipid metabolism and metamorphic development in a destructive insect pest, spodoptera litura (Lepidoptera: Noctuidae). *Int. J. Mol. Sci.* **23**, 9064. <https://doi.org/10.3390/ijms23169064> (2022).
38. Broderick, S., Wang, X., Simms, N. & Page-McCaw, A. Drosophila Ninjurin A induces nonapoptotic cell death. *PLoS ONE* **7**, e44567. <https://doi.org/10.1371/journal.pone.0044567> (2012).
39. Cooper, T. A. & Ordahl, C. P. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J. Biol. Chem.* **260**, 11140–11148 (1985).
40. Sheng, J. J. & Jin, J. P. Gene regulation, alternative splicing, and posttranslational modification of troponin subunits in cardiac development and adaptation: a focused review. *Front. Physiol.* **5**, 165. <https://doi.org/10.3389/fphys.2014.00165> (2014).
41. Fili, N. & Toseland, C. P. Unconventional myosins: how regulation meets function. *Int. J. Mol. Sci.* **21**, 67. <https://doi.org/10.3390/ijms21010067> (2019).
42. Biamonti, G., Maita, L. & Montecucco, A. The krebs cycle connection: reciprocal influence between alternative splicing programs and cell metabolism. *Front. Oncol.* **8**, 408. <https://doi.org/10.3389/fonc.2018.00408> (2018).
43. Campos, M. & Albrecht, L. V. Hitting the sweet spot: how glucose metabolism is orchestrated in space and time by phosphofructokinase-1. *Cancers (Basel)* **16**, 16. <https://doi.org/10.3390/cancers16010016> (2023).
44. Ozerova, A. M., Kulikova, D. A., Evgen'ev, M. B. & Gelfand, M. S. Temporal dynamics of gene expression during metamorphosis in two distant drosophila species. *Genome Biol. Evol.* **17**, evaf100. <https://doi.org/10.1093/gbe/evaf100> (2025).
45. Lev Maor, G., Yearim, A. & Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet.* **31**, 274–280. <https://doi.org/10.1016/j.tig.2015.03.002> (2015).
46. Li, S., Zhang, J., Huang, S. & He, X. Genome-wide analysis reveals that exon methylation facilitates its selective usage in the human transcriptome. *Brief. Bioinform.* **19**, 754–764. <https://doi.org/10.1093/bib/bbx019> (2018).
47. Schubeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326. <https://doi.org/10.1038/nature14192> (2015).
48. Harris, K. D., Lloyd, J. P. B., Domb, K., Zilberman, D. & Zemach, A. DNA methylation is maintained with high fidelity in the honey bee germline and exhibits global non-functional fluctuations during somatic development. *Epigenetics Chromatin* **12**, 62. <https://doi.org/10.1186/s13072-019-0307-4> (2019).
49. Xiang, H. et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat. Biotechnol.* **28**, 516–520. <https://doi.org/10.1038/nbt.1626> (2010).
50. Albalat, R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus Branchiostoma floridae. *Dev. Genes Evol.* **218**, 691–701. <https://doi.org/10.1007/s00427-008-0247-7> (2008).
51. Yoon, K., Williams, S. & Duncan, E. J. DNA methylation machinery is involved in development and reproduction in the viviparous pea aphid (Acyrtosiphon pisum). *Insect Mol. Biol.* **33**, 534–549. <https://doi.org/10.1111/imb.12936> (2024).
52. Provataris, P., Meusemann, K., Niehuis, O., Grath, S. & Misof, B. Signatures of DNA Methylation across Insects suggest reduced DNA methylation levels in holometabola. *Genome Biol. Evol.* **10**, 1185–1197. <https://doi.org/10.1093/gbe/evy066> (2018).
53. Hunt, B. G., Glastad, K. M., Yi, S. V. & Goodisman, M. A. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. *Genome Biol. Evol.* **5**, 591–598. <https://doi.org/10.1093/gbe/evt030> (2013).
54. Hunt, B. G., Glastad, K. M., Yi, S. V. & Goodisman, M. A. The function of intragenic DNA methylation: insights from insect epigenomes. *Integr. Comp. Biol.* **53**, 319–328. <https://doi.org/10.1093/icb/ict003> (2013).
55. Xu, G. et al. Intragenic DNA methylation regulates insect gene expression and reproduction through the MBD/Tip60 complex. *iScience* **24**, 102040. <https://doi.org/10.1016/j.isci.2021.102040> (2021).
56. Jones, C. M., Lim, K. S., Chapman, J. W. & Bass, C. Genome-wide characterization of DNA methylation in an invasive lepidopteran pest, the cotton bollworm helicoverpa armigera. *G3 (Bethesda)* **8**, 779–787. <https://doi.org/10.1134/g3.117.1112> (2018).

57. Wang, Q. et al. Gene body methylation in cancer: molecular mechanisms and clinical applications. *Clin. Epigenetics* **14**, 154. <https://doi.org/10.1186/s13148-022-01382-9> (2022).
58. Belles, X. MicroRNAs and the evolution of insect metamorphosis. *Annu. Rev. Entomol.* **62**, 111–125. [https://doi.org/10.1146/annu\\_rev-ento-031616-034925](https://doi.org/10.1146/annu_rev-ento-031616-034925) (2017).
59. Bendena, W. G., Hui, J. H. L., Chin-Sang, I. & Tobe, S. S. Neuropeptide and microRNA regulators of juvenile hormone production. *Gen. Comp. Endocrinol.* **295**, 113507. <https://doi.org/10.1016/j.ygcen.2020.113507> (2020).
60. Mukherjee, K., Baudach, A., Vogel, H. & Vilcinskas, A. Seasonal phenotype-specific expression of microRNAs during metamorphosis in the European map butterfly *Araschnia Levana*. *Arch. Insect Biochem. Physiol.* **104**, e21657. <https://doi.org/10.1002/arch.21657> (2020).
61. Qu, Z., Bendena, W. G., Tobe, S. S. & Hui, J. H. L. Juvenile hormone and sesquiterpenoids in arthropods: biosynthesis, signaling, and role of MicroRNA. *J. Steroid. Biochem. Mol. Biol.* **184**, 69–76. <https://doi.org/10.1016/j.jsbmb.2018.01.013> (2018).
62. Roberts, K. T., Steward, R. A., Suess, P., Lehmann, P. & Wheat, C. W. A time course analysis through diapause reveals dynamic temporal patterns of microRNAs associated with endocrine regulation in the butterfly *Pieris napi*. *Mol. Ecol.* **34**, e17348. <https://doi.org/10.1111/mec.17348> (2024).
63. Tian, S. et al. A micro-RNA is the effector gene of a classic evolutionary hotspot locus. *Science* **386**(6726), 1135–1141. <https://doi.org/10.1101/2024.02.09.579741> (2024).
64. Saccheri, I. J., Wellcome Sanger Institute Tree of Life, p., Wellcome Sanger Institute Scientific Operations, D. N. A. P. c., Tree of Life Core Informatics, c. & Darwin Tree of Life, C. The genome sequence of the Squinting Bush Brown, *Bicyclus anynana* (Butler, 1879). *Wellcome Open Res* **8**, 280 (2023). <https://doi.org/10.12688/wellcomeopenres.19432.1>
65. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie Ballgown. *Nat. Protoc.* **11**, 1650–1667. <https://doi.org/10.1038/nprot.2016.095> (2016).
66. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Res* **9** (2020). <https://doi.org/10.12688/f1000research.23297.2>
67. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
69. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic. Acids Res.* **42**, e69. <https://doi.org/10.1093/nar/gku154> (2014).
70. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87. <https://doi.org/10.1186/gb-2012-13-10-r87> (2012).
71. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic. Acids Res.* **42**, W187–191. <https://doi.org/10.1093/nar/gku365> (2014).
72. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141. <https://doi.org/10.1016/j.xinn.2021.100141> (2021).
73. Kolde, R. *Pheatmap: Pretty Heatmaps R package version 1.0.12*. <https://CRAN.R-project.org/package=pheatmap> (2019).

## Acknowledgements

This work was supported by Brighton Integrative Genomics, University of Brighton, and the Portuguese Science Funding Agency, Fundação para a Ciência e Tecnologia (research grant PTDC/BIA-EVL/0321/2021 to P.B.).

## Author contributions

R.S conceived the study, and co-supervised the research with P.B. A.R.G and J.K. performed the experiments and collected data, and A.H processed and analysed data. A.H, J.K, and R.S drafted the manuscript. All authors revised the manuscript and contributed to data interpretation.

## Funding

Portuguese Science Funding Agency, grant PTDC/BIA-EVL/0321/2021, Brighton Integrative Genomics.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-18794-1>.

**Correspondence** and requests for materials should be addressed to R.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025