

RESEARCH

Open Access



The genetic basis of wing spots in *Pieris canidia* butterflies

Jocelyn Liang Qi Wee^{1*†}, Suriya Narayanan Murugesan^{1*†}, Christopher W. Wheat² and Antónia Monteiro¹

Abstract

Spots in pierid butterflies and eyespots in nymphalid butterflies are likely non-homologous wing colour pattern elements, yet they share a few features in common. Both develop black scales that depend on the function of the gene *spalt*, and both might have central signalling cells. This suggests that both pattern elements may be sharing common genetic circuitry. Hundreds of genes have already been associated with the development of nymphalid butterfly eyespot patterns, but the genetic basis of the simpler spot patterns on the wings of pierid butterflies has not been investigated. To facilitate studies of pierid wing patterns, we report a high-quality draft genome assembly for *Pieris canidia*, the Indian cabbage white. We then conducted transcriptomic analyses of pupal wing tissues sampled from the spot and non-spot regions of *P. canidia* at 3–6 h post-pupation. A total of 1352 genes were differentially regulated between wing tissues with and without the black spot, including *spalt*, *Krüppel-like factor 10*, genes from the Toll, Notch, TGF- β , and FGFR signalling pathways, and several genes involved in the melanin biosynthetic pathway. We identified 14 genes that are up-regulated in both pierid spots and nymphalid eyespots and propose that spots and eyespots share regulatory modules despite their likely independent origins.

Introduction

Discovering how novel traits originate and diversify across different lineages of organisms is a long-standing pursuit of evolutionary biologists [1]. Most morphological traits are the result of complex gene-regulatory networks (GRNs), orchestrated by intercellular signals and their target transcription factors (TFs). The level of complexity underlying the development of any trait is usually so high that it has been proposed that novel traits originate via the rewiring of partial or whole pre-existent GRNs into new developmental contexts [2]. An example involves the gene *optix*, a gene essential for eye development which appears to have been redeployed, together with many other eye pigment enzyme genes, onto butterflies' wings to differentiate red wing patterns [3–6]. Another example involves the co-option of a general limb

[†]Jocelyn Liang Qi Wee and Suriya Narayanan Murugesan are the joint first authors.

*Correspondence:

Jocelyn Liang Qi Wee
jocelyn_wee@u.nus.edu
Suriya Narayanan Murugesan
suriya_nm@u.nus.edu

¹Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

²Department of Zoology, Stockholm University, Stockholm 10691, Sweden



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

GRN, regulating the development of antennae, legs, and wings, onto a small group of cells on the wing to underlie the evolution of nymphalid eyespot patterns [7].

Eyespot patterns, however, appear to have replaced pre-existing simpler spots that were already present on the margins of nymphalid wings. Eyespot and spot patterns could, thus, potentially share common positional information during development conferred by a common set of genes. This evolutionary transition, of spots to eyespots, inferred from the use of a large phylogeny of the butterfly family Nymphalidae and comparative methods [8], has not yet been fully explored at the genetic and developmental levels. A recent study, however, that examined the function of two essential genes for the differentiation of eyespot centers, *Distal-less*, and *spalt*, found that they played no such role in differentiating the centers of spot patterns in a separate family of butterflies, the Pieridae, which last shared a common ancestor with Nymphalidae nearly 100 million years ago [9]. This study suggested, instead, that spots in pierids are likely homologous to more marginal parafocal elements found on the wings of butterflies belonging to the family Nymphalidae and are not positional homologs of spot patterns in nymphalid lineages that eyespots are thought to have evolved from [10].

Yet, pierid spots and eyespots of *Bicyclus anynana* butterflies have a few features in common that might signal process homology at the level of the constituent gene-regulatory networks (GRNs) [11, 12]. They both have black scales that require the gene *spalt* for their differentiation [7, 10]; and both patterns become reduced in size when cells at their center are damaged in the early pupal stage [13, 14]. The full extent of shared genes, and larger GRNs, between spots in pierids and eyespots in nymphalids, however, is unknown.

To further probe the genetic basis of spots, and to compare it with that of eyespots, we first assembled and annotated a high-quality draft genome for *P. canidia*, generated from long reads obtained from the PacBio Sequel platform. RNA-seq was performed to identify the suite of transcription factors and signalling pathway

molecules that underlie the development of spot patterns in the wings of *Pieris canidia*, following the same protocol used to probe the genetic basis of nymphalid eyespots [15]. This involved micro-dissections of very small pieces of wing tissue containing the spot region, and of flanking tissue without spots during the early pupal stage, followed by sequencing and differential gene expression analysis. We then compared the gene expression profiles involved in the development of both spots and eyespots in butterflies.

Results

P. canidia genome assembly and annotation

The genome profile of *P. canidia* was determined using a 19-kmer distribution from Illumina short reads from single individual using GenomeScope (<http://qb.cshl.edu/genomescope/>) [16]. The k-mer counting was done using Jellyfish (v.2.2.3) [17] and the approximate genome size was determined to be around 224 MB (Figure S1). We assembled the PacBio long reads into contigs using wtdbg2 (version2.4) [18] and canu (version1.9) [19] assembler separately. Canu produced an assembly of 325 MB size with N50 length of 767 KB. Wtdbg2 produced a 257 MB assembly with a N50 length of 2.7 MB. The assemblies from both assemblers were merged using quickmerge [20] followed by purging of the haplotigs to remove heterozygosity using purge_haplotigs [21]. Subsequent output from purge_haplotigs was polished using long reads and short reads with three rounds of racon [22] and pilon [23] respectively, resulting in a final assembly of 264.9 MB size and a N50 length of 10.1 MB with 392 contigs (Table 1). The assembly was repeat masked using RepeatMasker [24] and annotated with four rounds of Maker [25] using the transcriptome constructed from RNA-seq from different tissues (detailed in Materials and Methods section) and transcriptome and proteins sequences from other species of butterflies as relative species information. This resulted in 14,518 genes with 26,376 transcripts. We checked the completeness of the assembly and gene set using a lepidoptera database with Benchmarking Universal Single-Copy Orthologs (BUSCO) (last accessed October 2022) [26]. The genome was very complete with 97.5% of single copy gene sets found in the *P. canidia* genome (Table 1).

Transcriptomic analysis performed on regions of spot patterns of *P. canidia*

To identify the genes and signalling pathways that are responsible for organising the spot patterns of *P. canidia*, we compared gene expression patterns between regions of pupal forewing tissues that will later develop spots (spot region in M3) versus two regions of the wings that do not develop spots (the proximal region next to the

Table 1 Genome assembly statistics of *P. canidia*

<i>P. canidia</i> Genome Assembly	
Number of Scaffolds	392
Total Base Pairs	264,993,535
N50	10,149,875
N90	4,915,149
GC Percentage	36.55
BUSCO Lepidoptera database	5286 genes
BUSCO Complete	5226 (98.9%)
BUSCO Complete and Single Copy	5154 (97.5%)
BUSCO Complete and Duplicated	72 (1.4%)
BUSCO Fragments	7 (0.1%)

spot in the same M3 sector and the Cu1 region, in a more posterior sector) (Fig. 1A and B). Wing tissues were dissected at 3–6 h after pupation, which was the time window previously sampled for eyespot tissue in *Bicyclus anynana* [7, 15]. This period is when the central focal cells are beginning the process of colour ring differentiation in nymphalid eyespots, because when these cells are transplanted elsewhere on the wing, they can set-up rings of colour around them [13, 27]. In total, 12 libraries were sequenced (4 biological replicates/sample type) with each replicate sequenced to a depth of between 40 and 90 million reads on average.

Principal component analysis (PCA) of the libraries revealed that the samples generally clustered together based on the location of the wing where the tissue was obtained, except for 1 library from each group (Fig. 1C). Samples from the same wing sector clustered together, i.e., the group ‘spot’ clustered closely together with the group ‘proximal’, whereas tissues belonging to a different wing sector, ‘Cu1’ clustered separately. Since no morphological marker distinguished the two wing locations

within the M3 sector, it is possible that the two dissected tissues might have overlapped the spot region in some cases. As the PCA plot revealed 3 outliers that deviated significantly from their respective groups, these were removed for downstream analyses. Subsequent hierarchical clustering of samples showed that samples belonging to the same group clustered together (Figure S2).

Genes differentially expressed in the wing spot pattern

Two comparisons were performed to identify differentially expressed genes (DEGs) in spots of *P. canidia*: we compared ‘spot’ vs. ‘proximal’ and ‘spot’ vs. ‘cu1’ to obtain the common set of genes that appeared in both comparisons. The comparison of ‘spot’ vs. ‘proximal’ yielded 15 upregulated genes and 15 downregulated genes in spots whereas the comparison of ‘spot’ vs ‘cu1’ tissues yielded 817 upregulated genes and 535 downregulated genes in spots (Fig. 1D). Overall, 16 DEGs were found in common across both comparisons with 4 genes being differentially regulated in a conflicting manner between the two comparisons (Table S1). Of these 16 genes, 4 genes

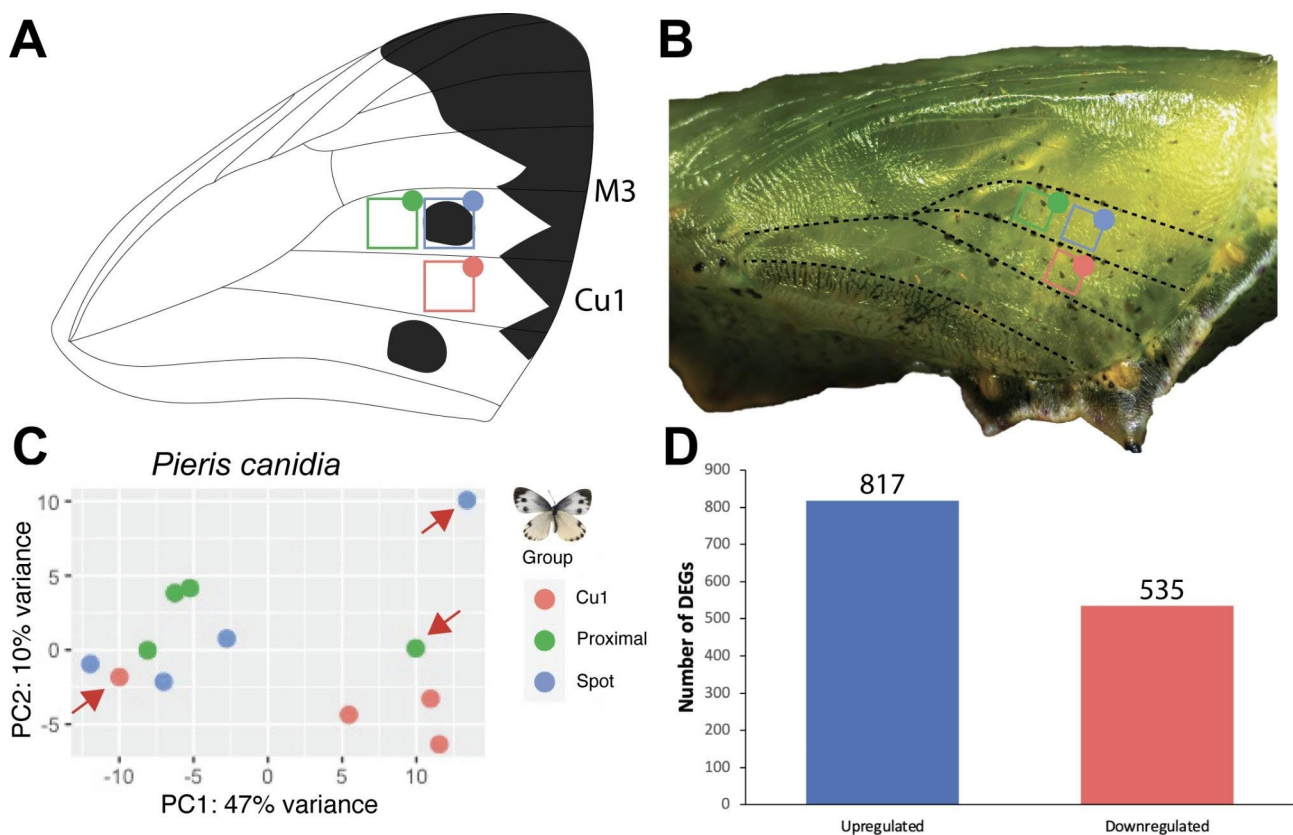


Fig. 1 Data analyses of RNA-seq samples. **(A)** Drawing of a *P. canidia* adult forewing showing the approximate location of the three dissected areas mapped to the wing. **(B)** *P. canidia* pupa (3 h old) with the three dissected tissues marked on the wing. **(C)** Principal component analyses of RNA samples clustered by the area of the wing of which samples were dissected. Apart from three outlier groups, the rest of the samples clustered first with others from the same wing location. Outlier groups were removed from downstream analyses. (Red: Wing tissues dissected from the Cu1 wing sector, Green: Wing tissues dissected from an area that is proximal to the spot pattern, and Blue: Wing tissues dissected from an area that contains the spot pattern). **(D)** Number of differentially expressed genes (DEGs) for the comparison of M3 ‘Spot’ versus Cu1 ‘Control’ tissue

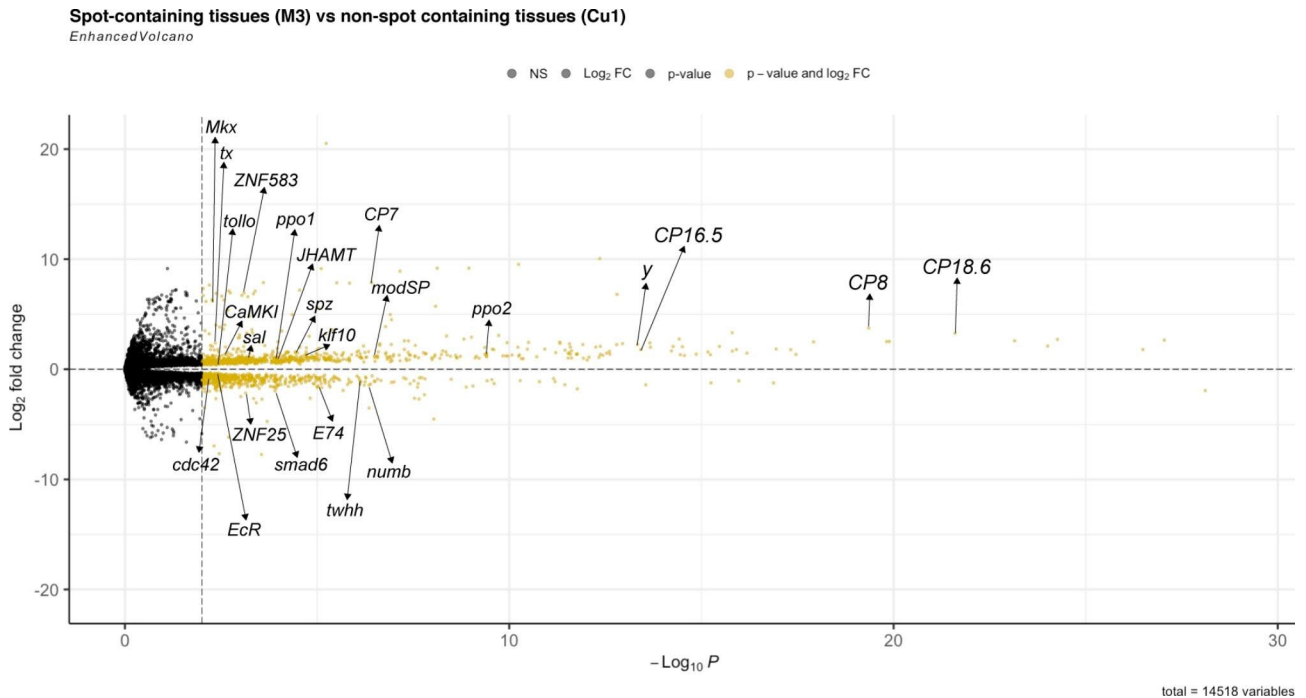


Fig. 2 Volcano plot of DEGs between “spot” and “cu1” tissues of 3–6h *P. canidia* pupal forewings. Log₂-fold change values were plotted against average values of $-\log_{10}(p\text{-value})$ with the threshold of DEGs set at $\text{padj} < 0.01$. Labeled genes represent a subset of DEGs present in spot-containing wing tissues. The full list of annotated DEGs is provided within the supplementary information. DEGs are marked as yellow dots. Genes upregulated in spot tissue are on the top half of the volcano plot. Downregulated genes are on the bottom half of the volcano plot. This plot is generated using the R package EnhancedVolcano [28]

Table 2 Genes annotated as part of pigment producing pathways expressed in spot tissues. Their log₂FC, padj values, and annotated biological processes. All genes described in this table are those that are upregulated in spot vs. cu1 samples

Gene ID	Gene Name	Log ₂ FC	padj	Biological Processes
Pcan_01059	<i>phenoloxidase subunit 1</i>	0.9761	0.0001	melanin biosynthetic process from tyrosine
Pcan_04990	<i>phenoloxidase subunit 2-like</i>	1.1925	3.91E-10	melanin biosynthetic process from tyrosine
Pcan_06980	<i>protein yellow-like</i>	1.5657	5.74E-05	melanin biosynthetic process
Pcan_06984	<i>major royal jelly protein 1-like isoform X2</i>	0.6703	0.0003	melanin biosynthetic process
Pcan_07144	<i>pyrimidodiazepine synthase-like</i>	0.7874	0.0034	eye pigment biosynthetic process; pteridine biosynthetic process
Pcan_09922	<i>L-dopachrome tautomerase yellow-f-like</i>	0.9512	0.0003	melanin biosynthetic process from tyrosine
Pcan_10393	<i>protein yellow-like</i>	2.2439	4.59E-14	melanin biosynthetic process from tyrosine
Pcan_11552	<i>protein yellow</i>	1.0028	0.0001	melanin biosynthetic process
Pcan_11721	<i>cysteine sulfinic acid decarboxylase</i>	1.0397	0.0071	developmental pigmentation
Pcan_12794	<i>protein yellow-like</i>	1.1784	3.40E-09	melanin biosynthetic process

were not annotated or were identified as uncharacterised proteins. The relatively low number of DEGs found in common across both sets of comparisons was primarily constrained by the low number of DEGs between ‘spot’ and ‘proximal’ tissues. Given that variations in spot size and location exist in natural populations, the possibility that wing tissue encompassing the spot pattern may have overlapped between the two tissue groups dissected side-by-side from the same wing sector, cannot be ruled out. Due to the constraints of the experimental design, we are not confident that the comparison of ‘spot’ versus ‘proximal’ will yield any informative insights regarding

spot pattern development and thus, did not pursue any further downstream analyses with this comparison.

To better elucidate the full suite of genes that are associated with spot development, we decided to focus on the 1352 DEGs, with a p-value adjusted (padj) of < 0.01 , that are up- or downregulated between ‘spot’ and ‘cu1’ tissues alone (Fig. 2). Upregulated DEGs were enriched for several GO terms such as “structural constituent of cuticle”, “structural constituent of chitin-based cuticle”, “chitin-based extracellular matrix”, and other GO terms. Downregulated DEGs were enriched for GO functions relating to “defense response to bacterium”, “regulation

Table 3 List of differentially expressed transcription factors. Their log₂FC and padj values. The four genes at the top of the table are upregulated in spot (vs. cu1 samples), the remaining genes are downregulated in spots

	Gene ID	Gene Name (Predicted)	Log ₂ FC	padj
Upregulated	Pcan_00226	Zinc finger protein 583	6.9395	0.0008
	Pcan_07555	homeotic protein spalt-major-like isoform X1	1.0576	0.0005
	Pcan_08360	helix-loop-helix protein delilah-like	2.3828	0.0045
	Pcan_12435	Krüppel-like factor 10	0.9471	2.58E-05
Downregulated	Pcan_01717	protein grainyhead isoform X1	-1.2947	1.26E-05
	Pcan_01823	Zinc finger protein 40	-0.5354	0.0024
	Pcan_03707	transcription factor E74 isoform B	-1.5034	9.40E-06
	Pcan_05425	POU domain, class 6, transcription factor 2 isoform X3	-1.6270	0.0011
	Pcan_05990	Mucin-2 isoform X2	-0.5291	0.0001
	Pcan_08419	REST corepressor 3 isoform X1	-1.0573	0.0010
	Pcan_08695	REPTOR-binding partner	-0.9769	0.0051
	Pcan_09441	T-box transcription factor TBX6	-4.518	9.22E-09
	Pcan_10341	transcriptional enhancer factor TEF-1 isoform X1	-1.0247	3.88E-05
	Pcan_12584	zinc finger protein MSN2	-1.5001	4.88E-06
	Pcan_12851	forkhead box protein binou	-1.1824	0.0011
	Pcan_13869	arginine-glutamic acid dipeptide repeats protein	-0.7630	0.0011

of transcription, DNA-templated”, and “protein phosphorylation” (Figure S3). Consistent with previous findings, *spalt* (*sal*) (Log₂FC: 1.05) is included in the set of DEGs and is shown to be over-expressed in wing tissues containing the spot pattern. This shows that our chosen method of analysis and thresholds applied to the dataset is sensitive enough to detect genes that are involved in wing spot development.

Pigmentation genes associated with spots

As the wing spot pattern of *P. canidia* consists of black scales, we looked through the list of DEGs to identify those known to function in pigment and melanin biosynthesis. In total, 10 out of 1352 DEGs are pigment-associated genes (Table 2). Of these 10 genes, 9 are involved in the melanogenic pathway in insects, and 1 in the pteridine biosynthesis pathway. All 10 genes involved in pigmentation processes are significantly upregulated in spot tissue compared to Cu1 control tissue, with *protein yellow-like* showing the largest fold change of 2.24 respectively. No genes involved in the melanin biosynthetic pathway or any other pigment pathways (i.e., ommochrome/pteridine) were downregulated at this stage of development.

Transcription factors and signalling pathways

The re-wiring of pre-existing gene regulatory networks to novel contexts has been proposed to drive the evolution of novel traits. Central to this regulation are ligands, receptors, and transcription factors involved in signalling pathways. To better understand the regulatory landscape that is controlling the expression of effector genes that are responsible for producing the dark wing spot, regulatory genes were identified by GO analysis. This was

achieved through searching through the set of annotated DE genes with corresponding GO terms using keywords such as i) “DNA-binding transcription factor activity” (GO:0003700, GO0000981, GO0001228, GO:0001217), (ii) “signalling pathway” and (iii) “signal transduction”. Sixteen transcription factors were differentially expressed between the two tissues (Spot vs. Cu1) (Table 3). Unlike the pigment pathway genes, most of the transcription factors were significantly downregulated in spots as summarised in Table 3. Other transcription factors, such as *Krüppel-like factor 10*, *homeotic protein spalt-major-like isoform X1*, and *helix-loop-helix protein delilah-like* were upregulated in spot tissues.

Recent studies have also pointed out the critical role that some signalling pathways have in organising butterfly wing patterns, providing positional information to surrounding cells to specify the development of colour patterns. Examples include but are not limited to Smad/TGF- β , Wnt, and the Hedgehog signalling pathway [29–31]. To infer the overall output of a signalling pathway, i.e., whether a pathway is active or repressed, expression profiles of all genes known to belong to signalling pathways in spot tissues were examined. In spots, the Toll, Notch, Transforming growth factor beta (TGF- β) pathways are likely to be associated with the development of spot patterns during early pupal development. It may also be possible that the Juvenile hormone and Fibroblast growth factor (FGFR) pathways play a role in the development of spot patterns during early pupal stages, but we stress that the evidence for this association is not strong given that only single genes from these pathways were upregulated in spot tissues (Table 4). Key positive regulators of these pathways were up-regulated in spots. In contrast, members of the Ecdysone and Hedgehog

Table 4 List of differentially expressed genes in spot vs. cu1 tissues belonging to different signalling pathways. Their log₂FC, padj values, and the signalling pathway they are associated with

	Gene ID	Gene Name (Predicted)	Log ₂ FC	padj	
Toll signalling	Pcan_04093	<i>protein spaetzle isoform X2</i>	1.4652	3.63E-05	
	Pcan_02090	<i>CLIP domain-containing serine protease 2-like</i>	1.2815	3.98E-08	
	Pcan_14273	<i>modular serine protease-like</i>	1.1835	3.22E-07	
	Pcan_12583	<i>serine protease inhibitor 88Ea-like</i>	1.0644	2.36E-09	
	Pcan_02088	<i>venom protease-like isoform X2</i>	1.0466	1.58E-05	
	Pcan_14259	<i>modular serine protease-like</i>	0.9910	1.15E-05	
	Pcan_05126	<i>plasma protease C1 inhibitor-like</i>	0.6930	6.59E-05	
	Pcan_10534	<i>protein toll-like</i>	0.6848	0.0004	
	Pcan_01305	<i>SUMO-activating enzyme subunit 1</i>	0.6812	0.0008	
	Pcan_08213	<i>phenoloxidase-activating enzyme-like</i>	0.6700	0.0008	
	Pcan_00871	<i>N-glycosylase/DNA lyase</i>	0.5928	0.0005	
	Pcan_01571	<i>toll-like receptor Tollo</i>	0.4321	0.0046	
	Pcan_08988	<i>beta-arrestin-1 isoform X3</i>	-0.6716	7.06E-06	
	Pcan_01487	<i>optineurin isoform X1</i>	-0.8099	0.0079	
	Wnt signalling	Pcan_05079	<i>transmembrane emp24 domain-containing protein 2</i>	0.8099	0.0029
		Pcan_10877	<i>transmembrane emp24 domain-containing protein 3</i>	0.7498	0.0027
Pcan_09890		<i>ruvB-like 2</i>	0.7422	0.0019	
Pcan_00025		<i>ruvB-like helicase 1</i>	0.7020	2.94E-06	
Pcan_10686		<i>tuberin isoform X2</i>	0.4477	0.0014	
Pcan_12849		<i>bifunctional heparan sulfate N-deacetylase/N-sulfotransferase</i>	-1.6603	7.55E-06	
Pcan_06691		<i>vacuolar protein sorting-associated protein 26B-like</i>	-1.2962	2.47E-10	
Pcan_13572		<i>F-box/WD repeat-containing protein 11 isoform X1</i>	-1.1326	0.0013	
Pcan_05017		<i>flotillin-2 isoform X1</i>	-0.8151	0.0091	
Pcan_04915		<i>serine/threonine-protein phosphatase alpha-2 isoform</i>	-0.7119	0.0068	
Pcan_13532		<i>transmembrane emp24 domain-containing protein 5-like</i>	-0.7038	0.0068	
Pcan_04914		<i>serine/threonine-protein phosphatase alpha-2 isoform</i>	-0.6798	0.0050	
Pcan_12157		<i>protein smoothed isoform X1</i>	-0.6092	0.0088	
Pcan_05769		<i>zinc finger protein 271 isoform X2</i>	-0.4434	0.0037	
Pcan_08442		<i>glycogen synthase kinase-3 beta isoform X1</i>	-0.4415	0.0063	
Notch signalling		Pcan_13084	<i>inorganic pyrophosphatase</i>	1.0857	1.18E-05
	Pcan_05048	<i>neurogenic locus notch homolog protein 1</i>	1.0814	0.0010	
	Pcan_08278	<i>eukaryotic translation initiation factor 3 subunit F</i>	0.9345	8.72E-06	
	Pcan_12581	<i>nicastrin</i>	0.7412	2.47E-05	
	Pcan_14339	<i>V-type proton ATPase catalytic subunit A</i>	0.7386	1.44E-06	
	Pcan_07331	<i>charged multivesicular body protein 2a</i>	0.6367	0.0030	
	Pcan_12310	<i>protein numb isoform X2</i>	-1.4288	4.45E-07	
	Pcan_08419	<i>REST corepressor 3 isoform X1</i>	-1.0573	0.0010	
	Pcan_00525	<i>importin subunit alpha-4</i>	-0.7257	3.92E-05	
	Pcan_10221	<i>neurogenic protein mastermind-like</i>	-0.6455	0.0002	
	Pcan_08957	<i>N6-adenosine-methyltransferase subunit METTL3</i>	-0.5183	0.0018	
	TGF-β signaling	Pcan_12435	<i>Krueppel-like factor 10</i>	0.9471	2.58E-05
		Pcan_01128	<i>peptide-N(4)-(N-acetyl-beta-glucosaminy)l asparagine amidase</i>	0.5512	0.0025
Pcan_06672		<i>calpain-A isoform X3</i>	0.5082	0.0001	
Pcan_06120		<i>mothers against decapentaplegic homolog 6</i>	-1.9599	0.0001	
Pcan_02675		<i>activin receptor type-1 isoform X2</i>	-1.1423	0.0026	
Pcan_05732		<i>E3 ubiquitin-protein ligase SMURF2</i>	-1.0781	0.0027	
Pcan_11478		<i>CTD nuclear envelope phosphatase 1 homolog</i>	-1.0736	0.0020	
Pcan_06119		<i>mothers against decapentaplegic homolog 6</i>	-1.0676	0.0001	
Pcan_04567		<i>protein expanded</i>	-0.8013	0.0028	
Pcan_03104		<i>cdc42 homolog</i>	-0.7415	0.0008	
Pcan_07516		<i>protein 60 A</i>	-0.6553	0.0089	
Pcan_01823		<i>Zinc finger protein 40</i>	-0.5354	0.0024	

Table 4 (continued)

	Gene ID	Gene Name (Predicted)	Log ₂ FC	padj
Hedgehog Signalling	Pcan_13241	<i>tiggy-winkle hedgehog protein</i>	-0.9714	7.82E-07
Juvenile Hormone Signalling	Pcan_04554	<i>juvenile hormone acid O-methyltransferase-like</i>	1.4772	0.0001
Ecdysone signalling	Pcan_00270	<i>cytochrome P450 6k1-like</i>	1.3776	2.39E-13
	Pcan_11395	<i>cytochrome P450 306a1</i>	0.7765	0.0079
	Pcan_03710	<i>ecdysone-induced protein 74EF isoform X1</i>	-1.3910	0.0010
	Pcan_05558	<i>transmembrane GTPase Marf</i>	-1.1996	0.0024
	Pcan_14165	<i>ecdysone receptor isoform X1</i>	-0.3772	0.0039
FGF Signalling	Pcan_04256	<i>acidic fibroblast growth factor intracellular-binding protein</i>	0.8700	1.50E-06

signalling pathway were down-regulated in spot tissues. The Wnt signalling pathway may also be repressed in spots (at 3–6 h), as key positive regulators of the Wnt pathway were down-regulated in spots (see discussion).

Genes known to be functioning or expressed in eyespot development

To explore the possibility that eyespots and spots share common developmental modules, we manually cross-listed the DE genes obtained in this study with the list of genes known to be associated with eyespot development from published eyespot transcriptomes and the literature (Fig. 3).

Discussion

Pioneers of pierid wing pattern research, such as Schwanwitsch [32] and Shapiro [33] proposed that spots of pierid butterflies are positional homologs of marginal nymphalid wing elements rather than of eyespots. This proposal was based on early wing pattern comparative work based on an extension of the nymphalid ground plan, a schematic of patterns homologies and symmetry systems deeply conserved amongst nymphalid species [34]. Subsequent studies on the molecular basis of these pierid wing spots showed that none of the known eyespot central (focal) genes were expressed in the centers of spots in both larval and pupal wings [10, 14, 35, 36]. The only exception was *spalt* which is expressed homogeneously in areas of the pupal wings that would eventually develop the complete black spot pattern. In addition to its role in differentiating eyespot centers, *spalt* is also an organiser of distal wing patterns in nymphalids. A robust expression of *Dpp*, an eyespot candidate morphogen, along the margins of *P. canidia* larval wings juxtaposed with the absence of *Dpp* signal in the center of spots [10], collectively supports the hypothesis that pierid spots may be variations of submarginal nymphalid pattern elements and not true homologs of nymphalid eyespot patterns. Here, however, we provide genetic evidence that the two

pattern elements may have genes and developmental modules in common.

In this study we identified the first unbiased set of spot candidate genes in a pierid species at 3–6 h post pupation. We discovered a large suite of pigmentation genes, transcription factors, and signalling pathways, differentially expressed in spots and flanking wing areas, that are likely working together to pre-pattern the development of a melanin spot. Interestingly, some of these genes are also expressed in eyespot centers, suggesting process homology at the level of GRNs between these two traits.

Genes belonging to the melanin pathway are upregulated in spots

Most of the pigment-associated genes belong to the melanin biosynthesis pathway and were significantly upregulated in the presumptive spot area (Fig. 4). These genes are described below.

Several members of the *yellow* gene family, *yellow*, *L-dopachrome tautomerase yellow-f-like* and three copies of *yellow-like genes* were significantly up-regulated in wing spot tissues. Although the molecular functions of *yellow* genes are not yet fully determined, the expression of *yellow* prefigures melanic patterns and *yellow* genes are necessary for the formation of melanin pigments in *Drosophila*, *Bombyx mori*, numerous butterflies and many other insects [38]. *Yellow* was previously thought to have enzymatic properties that catalyses the conversion of melanin precursors into black melanin [39], but dopachrome conversion enzyme (DCE) activity was not detected in-vitro. Instead, a signal peptide that promotes the export of proteins is nestled within the N-terminus of the *yellow* protein coding sequence and *Yellow* protein has been visualised localising to extracellular regions. This suggests that *yellow* may regulate the formation of melanic patterns by acting as an anchoring protein that binds to melanin pigments in the cuticle [39–41]. However, in *Drosophila*, the two other related genes, *yellow-f* and *yellow f-2*, have DCE enzymatic activity so these

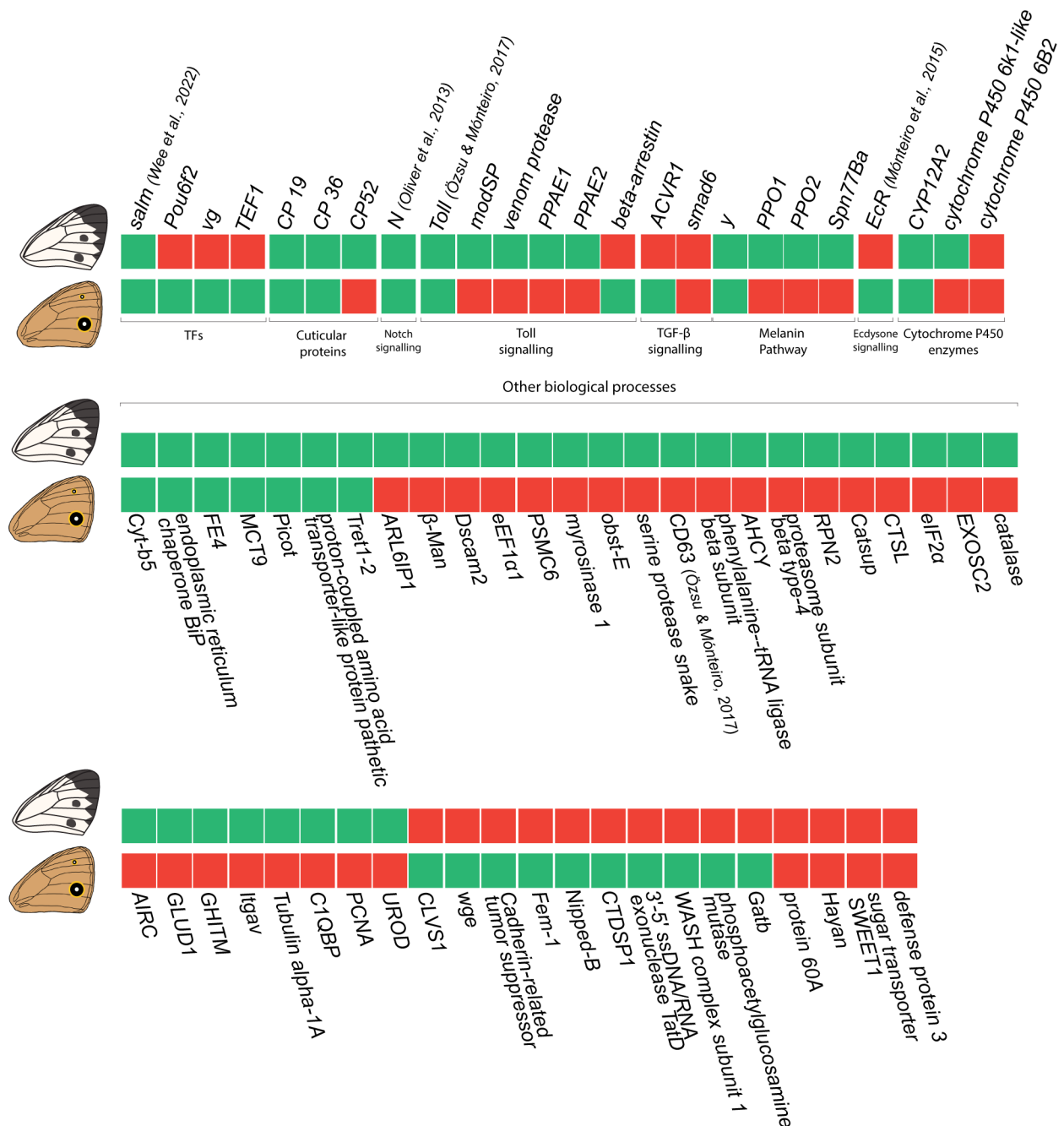


Fig. 3 Core set of genes shared or differentially expressed between spot and eyespot patterns. Differentially expressed genes in *P. canidia* spots versus control tissue that have corresponding homologs in *B. anynana* that are also differentially expressed in eyespots versus control tissue. The differentially expressed genes are arranged according to their annotated functions. Green blocks represent upregulated genes, while red blocks represent downregulated genes. The sum of expression patterns is: 14 genes are green-green, 16 genes are red-green, 35 genes are green-red and 6 genes are red-red. To improve readability, all genes within the figure that do not have an attached citation are referenced from [7]

genes would likely be directly involved in the synthesis of melanin pigments [42].

Like *yellow*, the two phenoloxidase (PO) enzymes, *phenoloxidase subunit 1*, and *phenoloxidase subunit 2*, that were upregulated in spot tissues, are also involved in the melanin pathway. POs are tyrosinases whose main role is to oxidise phenols to quinones, which polymerise to

form melanin. In the melanin pathway, POs hydroxylate monophenols, such as tyrosine, to produce the melanin precursor DOPA.

Although part of the pteridine pathway, the catalytic product of *pyrimidodiazepine (PDA) synthase-like*, may contribute to the production of melanin pigments for spot development. *PDA synthase*, catalyses a reversible

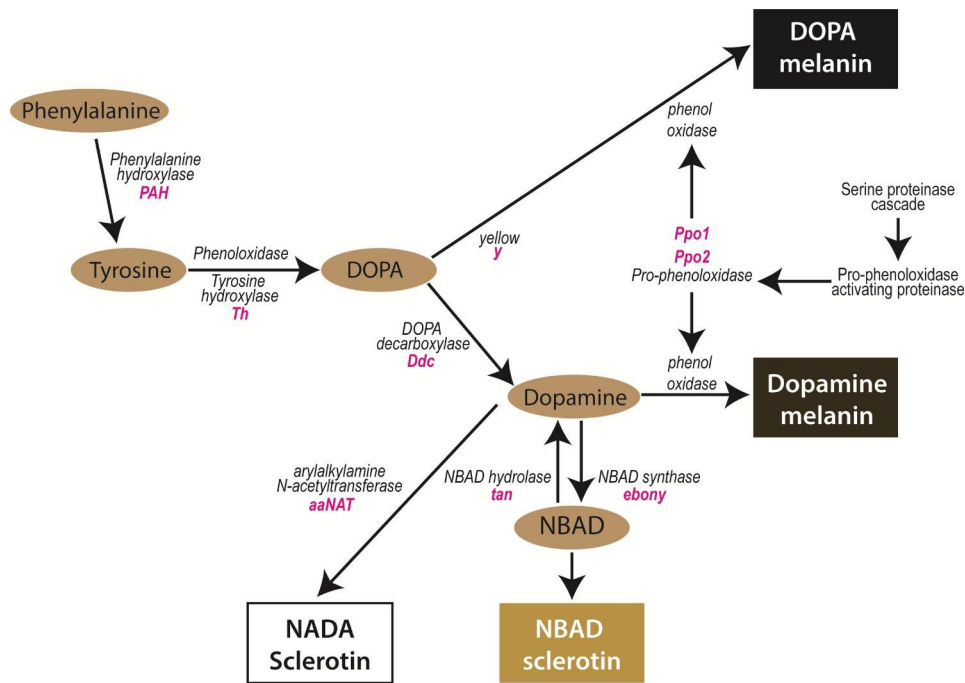


Fig. 4 Part of the melanin biosynthesis pathway in insects. The enzymatic reactions that produce two types of melanins, DOPA melanin and dopamine melanin are shown here. Enzymes involved in the catalytic reactions of intermediate substrates are in black, and the genes that encode these enzymes are italicised in pink. Adapted from [37]

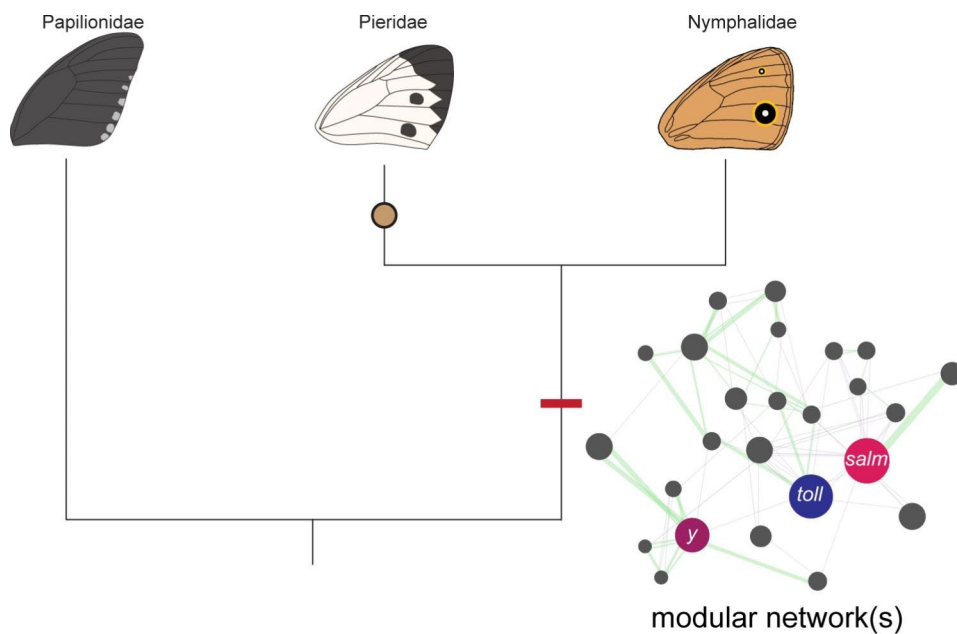


Fig. 5 Eyespots and spots are non-homologous traits that might share homology at the level of GRNs. Both pierid spots and nymphalid eyespots may share common developmental modules. The core set of genes, as represented by the red bar, identified as common to both wing patterns may be a representation of an ancient ‘melanism’ module that was used in both patterns to produce melanin pigments or an ancient ‘wing patterning’ module that has been used repeatedly amongst butterfly species to organise the formation of patterns on the wing. However, more work needs to be done to validate the role(s) of this core set of genes in wing pattern development. Also, the function of gene enhancers common to these two non-homologous traits would also have to be studied to confirm that homology exists at the level of GRNs. The modular network shown here is a hypothetical example of how some of the genes identified as upregulated in both spot and eyespot wing tissues may be arranged within a network

reaction of pyrimidodiazepine to 6-pyruvoyl-tetrahydropterin [43, 44]. 6-pyruvoyl-tetrahydropterin is a key intermediate required for the synthesis of tetrahydrobiopterin, also known as BH₄ [45]. In insects, BH₄ acts as an essential cofactor for other enzymes in several key reactions that involve the conversion of phenylalanine and tyrosine to melanin [46–48] (Fig. 4). Thus, the upregulation of *PDA synthase* in *Pieris* spots suggests that part of the pteridine biosynthesis pathway is also used to produce melanin pigments.

Another pigment associated gene, *cysteine sulfinic acid decarboxylase*, was also upregulated in black spots. *Csad* was suggested to be the locus responsible for a melanic mutation in *B. anynana* larvae, known as *chocolate* [49]. This gene belongs to the pyridoxal phosphate (PLP) aspartate aminotransferase superfamily. PLPs serve as co-enzymes in many biochemical reactions, and one of its most well-studied family members is *dopa decarboxylase (ddc)*, that catalyses the decarboxylation of L-dopa to produce dopamine [50]. Thus, it is highly likely that *csad* might also have direct enzymatic role on the melanin pathway.

Collectively, these genes are likely involved in the molecular processes required for the production of melanin pigments for the wing spot pattern in *P. canidia*.

Transcription factors regulating spot genes

Four transcription factors were significantly upregulated in spot tissues. They may be involved in the regulation of the pigmentation effector genes described above.

Helix-loop-helix protein delilah-like (tx), has been reported to be a marker gene that differentiates epidermal cells to facilitate muscle attachment in *Drosophila melanogaster* [51]. This gene is also important for specifying intervein regions, and adhesion of wing layers in *Drosophila*. *tx* is expressed in a homogenous fashion across the wing blade but is absent in the veins and is thought to be a mediator of several signalling pathways to specify the fate of intervein cells [52]. However, how this gene might be functioning in the context of the lepidopteran wing spot pattern is unclear.

spalt was also up-regulated in spot tissue in early pupal wings, which provided support for previous immunolocalization studies showing that *sal* expression correlates with melanic wing patterns in pierids [14, 36]. *sal*'s expression also supported functional studies, showing the essential role of this gene in the development of spots in *P. canidia* [10, 14]. However, it is still unclear how *spalt* is functionally connected to genes of the melanin pathway in this species.

Outside of pierids, however, *spalt* has also been associated with the development of both melanic and non-melanic patterns outside of pierids. In the pupal wings of the nymphalid *B. anynana*, *sal* expression maps to

the black scales in eyespots and brown scales along the wing margin. This expression of *sal* demarcating the boundaries of brown scales along the wing margin is also observed in another nymphalid species, *J. coenia* [53]. Additionally, in *J. coenia*, expression of *sal* is also associated with white and blue scales that are present within the eyespot pattern and in white spot patterns located on the forewing [53, 54] while in the lycaenid butterfly *L. melissa*, *sal* expression is associated with metallic-coloured scales located in chevrons at the margin [54]. These data suggest that *spalt* may play a primary pattern organisation role, rather than be a regulator of melanin pigmentation per se.

Krüppel-like factor 10 (Klf10) belongs to the Specificity protein 1 (Sp1)/Krüppel-like zinc finger family of transcription factors and it is up-regulated in spots. Members of this family are characterised by three deeply conserved C2H2 zinc finger motifs that preferentially bind to GC-rich promoter regions to modulate transcription of target genes [55, 56]. Sp1/Krüppel-like transcription factors, also known as transforming growth factor-beta (TGF- β) inducible genes (TIEG), modulate numerous cellular processes such as epithelial cell proliferation, differentiation, apoptosis, circadian rhythm, and regeneration, among other processes [57–59]. In *Drosophila*, the fly homolog of *Klf10* is *cabut (cbt)*. *cbt* is a positive regulator of TGF- β signalling and is known to be involved in wing disc morphogenesis, cell proliferation, and dorsal closure [60]. During *Drosophila* embryogenesis, *cbt* mutant embryos suffered from dorsal closure defects and expression of *dpp* was found to be downregulated in epidermal leading-edge cells during closure [61]. In *Drosophila* wing imaginal discs, *cbt* was found to be a positive regulator of *dpp* target genes such as *spalt* and *optomotor-blind* [60]. Although *dpp* was not detected in late larval wings, and 18 h pupal wings in *P. canidia* [10], *dpp* might be specifying the spot pattern in early pupal wings (3–6 h), either from the center of this pattern or from the wing margin [10]. *spalt* expression in spots may be responding to upstream signal input from both *dpp* and *cbt*.

Pathways associated with spot development

The most highly represented signalling pathway in spot tissue was the Toll signalling pathway. Fourteen genes belonging or interacting with this pathway were upregulated in spots. The Toll pathway is involved in wound repair and facilitating the production of melanin at wound sites [62–64]. To activate the Toll pathway, the extracellular ligand, *spaetzle* (up-regulated in spots) must first be proteolytically cleaved by proteases such as *serine protease easter-like* (also up-regulated in spots) [65]. The cleaved ligand subsequently binds to the Toll receptor (up-regulated in spots) inducing a receptor dimerization process. The dimerised Toll receptor will then interact

with a complex of several other downstream genes in a signalling cascade [66].

The Toll pathway has been previously implicated in the formation of wing patterns and melanic larval patterns in other lepidopteran species. Toll genes are overexpressed in eyespots of *B. anynana* [15], and in red marginal spots of *Papilio polytes* [67]. *Toll receptor 8* and *spätzle3* are required for the development of epidermal black stripes in the larvae of a silkworm mutant strain [68]. As mentioned previously, the melanin synthesis pathway is, in part, catalysed by the enzyme phenoloxidase (PO) which normally exists in its inactive form proPO. When required, proPO will be converted to active PO by a series of serine proteases, to eventually produce melanin pigments. In *Drosophila*, there is evidence for molecular cross-talks between the Toll pathway and melanin synthesis pathway as they are activated by a shared set of serine proteases working upstream of both pathways [69, 70] and it is likely that the melanin produced in *Pieris* black wing spots may have been synthesised through the PO-mediated melanin synthesis pathway.

Interestingly, the *Notch* receptor and other members of the Notch signalling pathway are overexpressed in spots. The Notch signalling pathway is known for its role in differentiating and patterning cells along the dorsal/ventral boundary that make up the wing margin in *Drosophila* [71–73]. One of the members of this pathway, *nicastarin*, shown here to be upregulated in spots, is an essential component for the activation of the pathway [74]. Other positive regulators expressed in spots include *eukaryotic translation initiation factor 3 subunit F (EIF3F)* and *V-type proton ATPase catalytic subunit A (Vha68-1)*, which are required for the de-ubiquitination of the activated Notch receptor [75], and intracellular trafficking of Notch signals [76], respectively. *Notch* is expressed in midline patterns in *Pieris rapae* in late larval wing discs [35], but its function in butterfly wings is still unknown. Larval integument markings, however, have been previously disrupted via the knockdown of both *Notch* and *delta*, using siRNA, in three different species of lepidopteran larvae, *Papilio xuthus*, *Papilio machaon*, and the multi lunar (L) mutant strain of *Bombyx mori* [77]. The Notch pathway may be patterning the wing spot in early pupal stages, but additional functional work is needed to functionally validate this hypothesis.

The Dpp/BMP (Bone morphogenetic proteins) branch of TGF- β signalling may also be upregulated in early pupal stages of pierid spot development. Within the TGF- β pathway, there are two subfamilies of ligands, namely the BMPs and the activins [78]. These ligands, upon binding their receptors, use different sets of Smad genes, transcription factors that translocate to the nucleus, to transcriptionally modulate target genes [79]. Identification of the Smads being up-regulated in tissues

allows the identification of which branch of the TGF- β signalling pathway is being used.

Within the list of DE genes, 2 copies of *mothers against decapentaplegic homolog 6 (Smad6)*, are downregulated in spot tissues. *Smad6*, whose fly homolog is *Daughters against dpp (Dad)*, belongs to a class of inhibitory Smads whose function is to downregulate *Dpp* signalling [80, 81]. The downregulation of *Smad6* in spots, suggests that *dpp* signalling is active in spot tissues at 3–6 h post pupation.

The up-regulation of *peptide-N(4)-(N-acetyl-beta-glucosaminyl) asparagine amidase*, known as *PngI*, also suggests that the *dpp* signalling is active in spot tissues at 3–6 h post pupation, as this protein is known to upregulate *dpp* expression in the *Drosophila* gut [82].

Wnt genes have long been associated with the organisation and development of butterfly wing patterns [30, 83–85] but it is inconclusive as to whether the Wnt pathway is involved in spot formation in pierids. In *P. canidia*, two modulators of the Wnt pathway, *casein kinase II subunit beta (CK2)* and *bifunctional heparan sulfate N-deacetylase/N-sulfotransferase (sulfateless)* are down-regulated in spots, but these modulators are not specific to the Wnt signalling pathway [86, 87]. Two other genes, *RuvB-like helicase 1* and *ruvB-like helicase 2*, are known antagonistic regulators of β -catenin activity, a signal transducer of Wnt signalling. *rvb1* is a positive regulator of Wnt signalling while the expression of *rvb2* represses Wnt/ β -catenin signalling [88]. Since both genes are upregulated in spot tissues, they provide limited evidence as to whether Wnt signalling is active in spots. Interestingly, *frizzled-2*, the receptor for *wingless* [89, 90], was down-regulated in spot-containing tissue, albeit at a higher *padj* value of 0.013 and with a Log_2FC of -0.64. Since Wnt signalling is known to repress the expression of *frizzled-2* on the *Drosophila* wing to shape and sustain the Wg morphogen gradient [91], future work should study the role of Wg signalling in organising and positioning pierid spot patterns along the margin.

Our data also hints at a possible role of the insect hormone JH in the regulation of spot size. JH is one of the two major insect hormones that regulate the development of seasonal polyphenisms [92, 93], and *P. canidia* spot patterns are seasonally plastic. In colder seasons, these wing spots are bigger and darker [94]. Our data shows that the gene *juvenile hormone acid O-methyltransferase-like*, is upregulated in spot tissues. This gene encodes an enzyme that mediates the conversion of inactive precursors to JH [95]. Future work attempting to dissect the seasonal plasticity of *P. canidia* wings should look to studying the possible functions of JH in regulating the size and intensity of the melanised spots between the two seasonal forms.

Spots and eyespots are non-homologous wing pattern elements that may share developmental modules in common (Fig. 5). We compared the spot transcriptome of *P. canidia* to two different eyespot transcriptomes from *B. anynana*, all sampled at 3–6 h after pupation [7, 15]. This current spot transcriptome was also cross checked with previous studies of genes associated with eyespot development. Collectively, these comparisons revealed a set of shared genes that are expressed in both traits (Fig. 4).

Genes and signalling pathways shared across the two non-homologous traits include *spalt*, *yellow*, *notch*, Toll, and BMP signalling. Some of these genes (*yellow*, *spalt*, and Toll genes) may simply reflect the independent deployment of an ancient GRN, that was deeply conserved across numerous taxa (i.e., melanin gene regulatory network), onto the wing to produce black pigments needed to form the black circles in both spots and eyespots. Other genes and pathways such as (*Notch* and TGF- β) may likely also have conserved roles in organising wing patterns along the tip of the midline of each wing sector in insects. It remains to be seen, however, whether genes shared between eyespots and spots, and their regulatory interactions, were recruited once, in the ancestors of both lineages, or evolved convergently to function in wing patterning, independently in each lineage. To help solve this question, a more comprehensive analysis of spot evolution is necessary, using a denser sampling of species across all butterflies. Alternatively, discovering that genes common to spots and eyespots sharing the same enhancers might also confirm process homology across both wing patterns.

Conclusion

Here, we report a complete genome assembly of *Pieris canidia* and identify transcripts associated with early wing spot formation in this species. The list of DEGs identified here, represents the first unbiased set of spot candidate genes in a pierid species and provides a comparative framework to study the origin and diversification of butterfly wing patterns. These analyses, however, represents a single snapshot of spot development during a specific developmental stage. Future studies should functionally validate the list of identified spot genes using CRISPR/Cas9 system for further insights on the developmental basis of spot patterns.

Materials and methods

Animal husbandry

Pieris canidia used in this study were filial descendants from wild-caught individuals found along West Seletar Road, Singapore. Caterpillars were reared on potted *Brassica sp.* plants, at 27 °C, 80% humidity following a 12 h light:12 h dark cycle. *P. canidia* pre-pupae are readily identified by several changes to their morphology

and behaviour. During the last stage of larval development, *P. canidia* larvae will shrink slightly in length and move away from the potted plants. Once a suitable substrate for pupation is found, the larvae will produce a silk girdle around their thoracic segments and attach to the substrate. These pre-pupae were carefully removed from their pupation site and transferred to petri dishes where their time of pupation was monitored using a camera (Olympus Stylus Tough TG-5 Digital Camera) with the in-built time-lapse function. Pupae that were 3–6 h old were sexed, with only female pupae used for subsequent dissections to exclude any gender bias in gene expression profiles as *P. canidia* wing spots are sexually dimorphic. Female spots are larger than male spots.

Genome Assembly

High Molecular weight (HMW) DNA was extracted from two female pupae using a phenol-chloroform method with little modification [96]. Extracted DNA was stored in –80 °C before being pooled and sent for sequencing. Long reads library preparation and sequencing was performed at AIT Novogene, Singapore. 20GB of data, roughly of about (70x) coverage was sequenced using the PacBio sequel platform.

Initial genome assembly was carried out using canu [19] and wtdbg2 [18] assemblers separately. The canu assembler was run with the default option with minimum read length set to 3000 for the assembly, and wtdbg2 was run with default settings. For both assemblies, the relative genome size of 320 MB was set based on the sister species genome sizes (*P. napi* and *P. rapae*) [97]. The two assemblies were merged using quickmerge [20], with the canu assembly as the reference, and the wtdbg2 assembly as the query. Heterozygosity in the contigs of the resulting assembly was purged using purge_haplotigs [21], and only the haploid contigs were kept. The obtained contigs were polished with three rounds of racon [22] and Pilon [23] using corrected PacBio long reads and Illumina short reads, respectively, to get the final genome assembly, Pcan.v1. Illumina short reads were generated using standard protocol. DNA was extracted from a female adult individual. Library preparation was done using the Illumina TruSeq Nano DNA kit, with cluster amplification performed using the Illumina cBot. Sequencing was done through the Illumina HiSeqX platform (HiSeq Control Software 3.3.76/RTA 2.7.6) using sequencing-by-synthesis (SBS) kits to generate 151 bp paired-end reads. A BUSCO score [26] was used to check the completeness of the gene sets in the assembly resulting in 14,518 genes with 26,376 transcripts.

Repeat masking of genome

The genome was repeat masked for transposable elements, small and tandem repeats. Repeat Masking was

performed using a similar approach as that used for *B. anynana* genome masking [7]. De novo transposable elements and repeats were modelled and predicted using RepeatModeler v1.0.5 [98]. The repeat library created using the *de novo* approach sometimes had non-repeats and transposable elements in them, and to filter the non-repeats and coding regions we used the protein sequences from the genome annotation performed earlier. The resulting library was merged with the Lepidoptera repeat library obtained from RepBase v20.05 [99]. The genome was masked using the final merged library using Repeatmasker v4.0.7 [24] resulting in 30.05% of the genome being masked.

Genome annotation

The soft-repeatmasked genome was annotated, using four rounds of Maker v3 [25]. The transcriptome assembled from the RNA-Seq data was used as transcripts for the species, with transcriptome and protein sequences from *Pieris napi*, *Pieris rapae*, *Junonia coenia*, *Bombyx mori*, and *Bicyclus anynana* as relative species transcripts and protein homology evidence for the first round of gene predictions. Output gene predictions from each round were used as input for the next round. Snap and Augustus [100] were used for the second round of gene predictions, followed by Genemark [101] for the third round of gene modelling, and one final round of Snap and Augustus predictions. The minimum length of 35 amino acids was set for gene predictions. The predicted gene models were kept for genes that had an Annotation Edit Distance (AED) score of <1 and/or had a gene ontology obtained from Interproscan [102]. This resulted in 14,594 genes with 26,463 transcripts. In order to correct the annotations and produce a standardized gff3 file, the gff file obtained from Maker was run through `agat_convert_sp_gxf2gxf.pl` script, which is a part of AGAT tools [103]. This step resulted in the removal of 77 identical isoforms and added the missing gene features, leading to a total of 14,518 genes with 26,376 transcripts. Functional annotation was performed by locally blasting the transcripts against a non-redundant (nr) protein database, using diamond blast [104], and a gene ontology analysis was performed using Interproscan in Blast2Go [105]. Finally, the blast results were merged with the Interproscan results in Blast2Go to produce a final functional annotation for the genome.

Tissue collection and RNA extraction

Small squares of wing tissues of approximately 0.5×0.5 cm containing the spot centers from the M3 and Cu1 sector of the pupa were manually dissected from the dorsal forewing using a homemade dissecting tool which consists of thin strips of cut razor blades glued to a wooden handle (Figure S4). All dissections were made

in cold diethyl pyrocarbonate (DEPC) treated 1xPBS and dissecting tools were wiped with RNaseZap® (Invitrogen™) prior to dissection. Dissections of the wing tissues of different wing sectors were done in a random fashion to avoid any bias in gene expression levels. Harvested wing tissues from the pupa were immediately placed into individual 1.5ml Eppendorf tubes containing RNAlater® (Ambion) to preserve the integrity of the RNA. Each pupa yielded 6 pieces of wing tissues, 2 tissues each from the 3 biological groups ('spot', 'proximal' and 'cu1') from the left and right forewing respectively. Tubes were labelled individually according to the tissue group they belonged to and assigned numbers in a sequential manner. These were then stored at 4 °C for at least 24 h to allow for complete RNA stabilisation and transferred to -80 °C storage until all samples required for four biological replicates for each of the three wing tissue groups were obtained. To obtain biological replicates for each tissue type, a random group generator was used to divide 120 tubes into 4 groups in a random fashion. Each tube contained one piece of wing tissue, stored in RNAlater® solution. 30 pieces of tissue belonging to the same wing sector were retrieved from their individual tubes, pooled together, snap-frozen in liquid nitrogen, and homogenised using disposable polypropylene pellet pestles. RNA was subsequently extracted using the RNeasy Mini Kit (Qiagen), according to manufacturer's instructions. Preliminary concentration of the extracted RNA was measured using a NanoDrop 100 Spectrophotometer (ND-1000) and RNA integrity was assessed through agarose gel electrophoresis and Agilent 2100 Bioanalyser by an external vendor (AITbiotech).

Next generation sequencing

Library preparation, and sequencing was performed by Novogene. A total of 12 libraries was sequenced across all three different groups of wing tissues at 150-bp paired end using the Illumina NovaSeq6000 platform (Table S3).

Pre-processing of raw sequencing data

Quality control, removal of adaptors, and read filtering was performed using *fastp* [106]. As the Novaseq sequencing platform was used to sequence the libraries, we chose *fastp* to remove overrepresented polyG tails that might appear from a lack of signal in a two-colour system used in Novaseq sequencing technologies.

Genome alignment

A genome-guided transcriptome assembly was constructed using the HISAT2/StringTie pipeline [107] against the assembled *Pieris* genome (Table S3). Filtered reads were mapped to the *P. canidia* genome using HISAT2. The resulting alignments were later processed and quantified using StringTie which produces estimates

of transcript abundances and generated read count tables. A python script prepDe.py was used to convert the output files from Stringtie to produce genes and transcripts count matrices in the form of CSV files.

Differential gene expression analysis

Differential expression of genes was calculated using the Bioconductor R package, DESeq2 [108]. Pre-filtering of the gene count matrix was conducted prior to downstream analyses. Genes that had read counts present in less than three samples, and with normalised counts lesser or equal to 10 were removed from the dataset. A pairwise DE analysis was then made between the M3 spot tissue and the Cu1 non-spot control tissue. Genes with an adjusted P value of <0.01 were considered as significantly differentially expressed in this dataset. In all, a total of three biological replicates for each of the two different tissue groups (Spot & Cu1) were used in the final analysis.

Annotation and GO annotation of DE genes

Differentially expressed genes were blasted against the nr database in NCBI using blastx in Blast2GO with an E-value of <1e-3 with the application of a taxonomy filter Blast2GO and against both the InterPro and Egnog databases. All GO annotations obtained through the three different databases for each transcript were then merged in the Blast2GO pipeline [105].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09261-0>.

Additional file 1: Dataset containing annotated differentially regulated genes associated with *P. canidia* spot development across all comparisons.

Additional file 2: Table S1. List of differentially regulated genes common to both set of comparisons: i) spot vs cu1 and ii) spot vs proximal. **Table S2.** RNA concentrations obtained from dissected tissues for the 12 different libraries. **Table S3.** Summary of sequencing statistics and percentage of reads mapped to *P. canidia* genome using HISAT2. **Figure S1.** *P. canidia* genome size estimation using GenomeScope. **Figure S2.** Hierarchical clustering and heat map. **Figure S3.** Gene Ontology (GO) enrichment analyses for spot DEGs. **Figure S4.** Microdissections of early pupal wing tissues using dissecting tools fashioned from cut razor blades.

Acknowledgements

We would like to thank Heidi Connahs for providing initial comments for the manuscript.

Author Contribution

JW, SNM and AM conceptualised and designed the study. SNM and CWW performed whole-genome sequencing of *Pieris canidia*. JW performed RNA-seq of *P. canidia* wing tissues. Both JW and SNM analysed the transcriptomic and genomic data respectively and wrote the manuscript with input from CWW and AM. JW and SNM contributed equally to this work. All authors have read and approved the manuscript for submission.

Funding

This work was funded by the National Research Foundation (NRF) Singapore, under its Investigatorship Programme (NRF-NRFI05-2019-0006 Award). JW and

SNM were supported by graduate scholarships awarded by the Department of Biological Sciences, National University of Singapore, and by Yale-NUS, Singapore respectively. CWW would also like to acknowledge support from the Science for Life Laboratory, the National Genomics Infrastructure (NGI), Sweden, for genome sequencing and lastly, the Knut and Alice Wallenberg Foundation (grant number 2012.0058).

Data Availability

All data used in the analyses are included within the article and its accompanying supplementary files. Raw Illumina and PacBio sequence reads used for the assembly of the genome and for the wing spot transcriptomic analyses are available under NCBI Bioproject PRJNA895987. (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA895987>). The genome assembly for *P. canidia* can be downloaded from the Dryad Digital Repository (https://datadryad.org/stash/share/WY0iCxNeFNJ_TDYOKAPCWaKCPOXZJtgvyS7YzYyMCLI).

Declarations

Ethical approval

Not applicable.

Consent to publish

Not applicable.

Conflict of interest

The authors have no conflicts of interest to declare.

Received: 17 November 2022 / Accepted: 20 March 2023

Published online: 04 April 2023

References

- Wagner GP, Lynch VJ. Evolutionary novelties. *Curr Biol*. 2010;20(2):R48–R52.
- Halfon MS. Perspectives on gene regulatory network evolution. *Trends Genet*. 2017;33(7):436–47.
- Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Kronforst MR et al. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 2011;333.
- Monteiro A. Gene regulatory networks reused to build novel traits: co-option of an eye - related gene regulatory network in eye - like organs and red wing patches on insect wings is suggested by optix expression. *BioEssays*. 2012;34(3):181–6.
- Martin A, McCulloch KJ, Patel NH, Briscoe AD, Gilbert LE, Reed RD. Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *EvoDevo*. 2014;5(1):7.
- Zhang L, Mazo-Vargas A, Reed RD. Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proceedings of the National Academy of Sciences* 2017, 114(40):10707–10712.
- Murugesan SN, Connahs H, Matsuoka Y, Das Gupta M, Tiong GJL, Huq M, Gowri V, Monroe S, Deem KD, Werner T et al. Butterfly eyespots evolved via cooption of an ancestral gene-regulatory network that also patterns antennae, legs, and wings. *Proceedings of the National Academy of Sciences* 2022, 119(8):e2108661119.
- Oliver JC, Beaulieu JM, Gall LF, Piel WH, Monteiro A. Nymphalid eyespot serial homologues originate as a few individualized modules. *Proceedings of the Royal Society B: Biological Sciences* 2014, 281(1787):20133262.
- Chazot N, Wahlberg N, Freitas AVL, Mitter C, Labandeira C, Sohn J-C, Sahoo RK, Seraphim N, De Jong R, Heikkilä M. Priors and posteriors in bayesian timing of divergence analyses: the age of butterflies revisited. *Syst Biol*. 2019;68(5):797–813.
- Wee JIQ, Das Banerjee T, Prakash A, Seah KS, Monteiro A. Distal-less and spalt are distal organisers of pierid wing patterns. *EvoDevo*. 2022;13(1):12.
- Gilbert SF, Bolker JA. Homologies of process and modular elements of embryonic construction. *J Exp Zool*. 2001;291(1):1–12.
- DiFrisco J, Jaeger J. Homology of process: developmental dynamics in comparative biology. *Interface focus*. 2021;11(3):20210007.

13. French V, Brakefield PM. The development of eyespot patterns on butterfly wings: morphogen sources or sinks? *Development* 1992; 116.
14. Stoehr AM, Walker JF, Monteiro A. Spalt expression and the development of melanic color patterns in pierid butterflies. *EvoDevo*. 2013;4(1):1–11.
15. Özsü N, Monteiro A. Wound healing, calcium signaling, and other novel pathways are associated with the formation of butterfly eyespots. *BMC Genomics*. 2017;18(1):788.
16. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33(14):2202–4.
17. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764–70.
18. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. 2020;17(2):155–8.
19. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722–36.
20. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 2016;44(19):e147–7.
21. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19(1):1–10.
22. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;27(5):737–46.
23. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963.
24. Smith AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2015.
25. Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. *Current protocols in bioinformatics* 2014, 48(1):4.11. 11–14.11. 39.
26. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
27. Nijhout HF. Pattern formation on lepidopteran wings: determination of an eyespot. *Dev Biol*. 1980;80(2):267–74.
28. Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 2019, 1.12.0.
29. Tong X, Lindemann A, Monteiro A. Differential involvement of hedgehog signaling in Butterfly Wing and Eyespot Development. *PLoS ONE*. 2012;7(12):e51087.
30. Martin A, Reed RD. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Dev Biol*. 2014;395(2):367–78.
31. Connahs H, Tlili S, van Creijl J, Loo TY, Banerjee TD, Saunders TE, Monteiro A. Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development*. 2019;146(9).
32. Schwanwitsch B. Wing pattern of pierid butterflies (Lepidoptera, Pieridae). *Entomol Obozrenie*. 1956;35.
33. Shapiro AM. Seasonal polyphenism. *Evol Biol*. 1976;9.
34. Schwanwitsch B. On the Ground-plan of Wing - pattern in Nymphalids and certain other Families of the Rhopalocerous Lepidoptera. In: *Proceedings of the Zoological Society of London: 1924: Wiley Online Library*; 1924: 509–528.
35. Reed RD, Serfas MS. Butterfly wing pattern evolution is associated with changes in a Notch/Distal-less temporal pattern formation process. *Curr Biol*. 2004;14(13):1159–66.
36. Monteiro A, Glaser G, Stockslager S, Glansdorp N, Ramos D. Comparative insights into questions of lepidopteran wing pattern homology. *BMC Dev Biol*. 2006;6(1):1–13.
37. Zhang L, Martin A, Perry MW, van der Burg KRL, Matsuoka Y, Monteiro A, Reed RD. Genetic basis of melanin pigmentation in Butterfly Wings. *Genetics*. 2017;205(4):1537.
38. Futahashi R, Sato J, Meng Y, Okamoto S, Daimon T, Yamamoto K, Suetsugu Y, Narukawa J, Takahashi H, Banno Y. Yellow and ebony are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics*. 2008;180(4):1995–2005.
39. Wittkopp PJ, True JR, Carroll SB. Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development*. 2002;129(8):1849–58.
40. Geyer PK, Spana C, Corces VG. On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *EMBO J*. 1986;5(10):2657–62.
41. Hinaux H, Bachem K, Battistara M, Rossi M, Xin Y, Jaenichen R, Le Poul Y, Arnoult L, Kobler JM, Kadow ICG. Revisiting the developmental and cellular role of the pigmentation gene yellow in *Drosophila* using a tagged allele. *Dev Biol*. 2018;438(2):111–23.
42. Han Q, Fang J, Ding H, Johnson JK, Christensen BM, Li J. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem J*. 2002;368(1):333–40.
43. Pyrimidodiazepine synthase. In: *Springer Handbook of Enzymes: Class 1 • Oxidoreductases VIII EC 15*. Edited by Schomburg D, Schomburg I. Berlin, Heidelberg: Springer Berlin Heidelberg; ; 2005: 323–325.
44. Kim J, Suh H, Kim S, Kim K, Ahn C, Yim J. Identification and characteristics of the structural gene for the *Drosophila* eye colour mutant sepia, encoding PDA synthase, a member of the omega class glutathione S-transferases. *Biochem J*. 2006;398(3):451–60.
45. Thöny B, Auerbach G, Blau N. Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem J*. 2000;347(1):1–16.
46. O'Donnell JM, McLean JR, Reynolds ER. Molecular and developmental genetics of the Punch locus, a pterin biosynthesis gene in *Drosophila melanogaster*. *Dev Genet*. 1989;10(3):273–86.
47. Fitzpatrick PF. Tetrahydropterin-dependent amino acid hydroxylases. *Annu Rev Biochem*. 1999;68(1):355–81.
48. Chen P, Wang J, Li H, Li Y, Chen P, Li T, Chen X, Xiao J, Zhang L. Role of GTP-CHI links PAH and TH in melanin synthesis in silkworm, *Bombyx mori*. *Gene*. 2015;567(2):138–45.
49. Saenko S, Jerónimo M, Beldade P. Genetic basis of stage-specific melanism: a putative role for a cysteine sulfinic acid decarboxylase in insect pigmentation. *Heredity*. 2012;108(6):594–601.
50. Richardson G, Ding H, Rocheleau T, Mayhew G, Reddy E, Han Q, Christensen BM, Li J. An examination of aspartate decarboxylase and glutamate decarboxylase activity in mosquitoes. *Mol Biol Rep*. 2010;37(7):3199–205.
51. Armand P, Knapp AC, Hirsch AJ, Wieschaus EF, Cole MD. A novel basic helix-loop-helix protein is expressed in muscle attachment sites of the *Drosophila* epidermis. *Mol Cell Biol*. 1994;14(6):4145–54.
52. Egoz-Matia N, Nachman A, Halachmi N, Toder M, Klein Y, Salzberg A. Spatial regulation of cell adhesion in the *Drosophila* wing is mediated by Delilah, a potent activator of β PS integrin expression. *Dev Biol*. 2011;351(1):99–109.
53. Reed RD, Selegue JE, Zhang L, Brunetti CR. Transcription factors underlying wing margin color patterns and pupal cuticle markings in butterflies. *EvoDevo*. 2020;11:1–10.
54. Brunetti CR, Selegue JE, Monteiro A, French V, Brakefield PM, Carroll SB. The generation and diversification of butterfly eyespot color patterns. *Curr Biol*. 2001;11(20):1578–85.
55. Kaczynski J, Cook T, Urrutia R. Sp1- and Krüppel-like transcription factors. *Genome Biol*. 2003;4(2):1–8.
56. Spittau B, Krieglstein K. Klf10 and Klf11 as mediators of TGF- β superfamily signaling. *Cell Tissue Res*. 2012;347(1):65–72.
57. Cook T, Gebelein B, Mesa K, Mladek A, Urrutia R. Molecular cloning and characterization of TIEG2Reveals a new subfamily of transforming growth Factor- β -inducible Sp1-like Zinc Finger-encoding genes involved in the regulation of cell growth. *J Biol Chem*. 1998;273(40):25929–36.
58. Chalaux E, López-Rovira T, Rosa JL, Pons G, Boxer LM, Bartrons R, Ventura F. A zinc-finger transcription factor induced by TGF- β promotes apoptotic cell death in epithelial Mv1Lu cells. *FEBS Lett*. 1999;457(3):478–82.
59. Hirota T, Kon N, Itagaki T, Hoshina N, Okano T, Fukada Y. Transcriptional repressor TIEG1 regulates Bmal1 gene through GC box and controls circadian clockwork. *Genes Cells*. 2010;15(2):111–21.
60. Rodriguez I. *Drosophila* TIEG is a modulator of different signalling pathways involved in wing patterning and cell proliferation. *PLoS ONE*. 2011;6(4):e18418.
61. Muñoz-Descalzo S, Terol J, Paricio N. Cabut, a C2H2 zinc finger transcription factor, is required during *Drosophila* dorsal closure downstream of JNK signaling. *Dev Biol*. 2005;287(1):168–79.
62. Hashimoto C, Gerttula S, Anderson KV. Plasma membrane localization of the toll protein in the syncytial *Drosophila* embryo: importance of transmembrane signaling for dorsal-ventral pattern formation. *Development*. 1991;111(4):1021–8.
63. Jasper H, Bohmann D. *Drosophila* innate immunity: a genomic view of pathogen defense. *Mol Cell*. 2002;10(5):967–9.

64. Carvalho L, Jacinto A, Matova N. The Toll/NF- κ B signaling pathway is required for epidermal wound repair in *Drosophila*. *Proceedings of the National Academy of Sciences* 2014, 111(50):E5373–E5382.
65. Jang I-H, Chosa N, Kim S-H, Nam H-J, Lemaître B, Ochiai M, Kambris Z, Brun S, Hashimoto C, Ashida M. A Spätzle-processing enzyme required for toll signaling activation in *Drosophila* innate immunity. *Dev Cell*. 2006;10(1):45–55.
66. Lemaître B, Meister M, Govind S, Georgel P, Steward R, Reichhart JM, Hoffmann JA. Functional analysis and regulation of nuclear import of dorsal during the immune response in *Drosophila*. *EMBO J*. 1995;14(3):536–45.
67. Nishikawa H, Iga M, Yamaguchi J, Saito K, Kataoka H, Suzuki Y, Sugano S, Fujiwara H. Molecular basis of wing coloration in a batesian mimic butterfly, *Papilio polytes*. *Sci Rep*. 2013;3(1):1–9.
68. KonDo Y, Yoda S, Mizoguchi T, Ando T, Yamaguchi J, Yamamoto K, Banno Y, Fujiwara H. Toll ligand Spätzle3 controls melanization in the stripe pattern formation in caterpillars. *Proceedings of the National Academy of Sciences* 2017, 114(31):8336–8341.
69. Kan H, Kim C-H, Kwon H-M, Park J-W, Roh K-B, Lee H, Park B-J, Zhang R, Zhang J, Söderhäll K. Molecular control of phenoloxidase-induced melanin synthesis in an insect. *J Biol Chem*. 2008;283(37):25316–23.
70. Dudzic JP, Hanson MA, Iatsenko I, Kondo S, Lemaître B. More than black or white: melanization and toll share regulatory serine proteases in *Drosophila*. *Cell Rep*. 2019;27(4):1050–61. e1053.
71. de Celis J, Garcia-Bellido A. Roles of the notch gene in *Drosophila* wing morphogenesis. *Mech Dev*. 1994;46(2):109–22.
72. Go MJ, Eastman DS, Artavanis-Tsakonas S. Cell proliferation control by notch signaling in *Drosophila* development. *Development*. 1998;125(11):2031–40.
73. Baonza A, Garcia-Bellido A. Notch signaling directly controls cell proliferation in the *Drosophila* wing disc. *Proceedings of the National Academy of Sciences* 2000, 97(6):2609–2614.
74. Hu Y, Ye Y, Fortini ME. Nicastrin is required for γ -secretase cleavage of the *Drosophila* notch receptor. *Dev Cell*. 2002;2(1):69–78.
75. Moretti J, Chastagner P, Gastaldello S, Heuss SF, Dirac AM, Bernards R, Masucci MG, Israel A, Brou C. The translation initiation factor 3f (eIF3f) exhibits a deubiquitinase activity regulating notch activation. *PLoS Biol*. 2010;8(11):e1000545.
76. Vaccari T, Duchi S, Cortese K, Tacchetti C, Bilder D. The vacuolar ATPase is required for physiological as well as pathological activation of the notch receptor. *Development*. 2010;137(11):1825–32.
77. Jin H, Yoda S, Liu L, Kojima T, Fujiwara H. Notch and Delta Control the switch and formation of camouflage patterns in caterpillars. *Iscience*. 2020;23(7):101315.
78. Hamarotoglu F, Affolter M, Pyrowolakis G. Dpp/BMP signaling in flies: from molecules to biology. In: *Seminars in cell & developmental biology: 2014*:Elsevier; 2014:128–136.
79. Heldin C-H, Miyazono K, Ten Dijke P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature*. 1997;390(6659):465–71.
80. Tsuneizumi K, Nakayama T, Kamashida Y, Kornberg TB, Christian JL, Tabata T. Daughters against dpp modulates dpp organizing activity in *Drosophila* wing development. *Nature*. 1997;389(6651):627–31.
81. Kamiya Y, Miyazono K, Miyazawa K. Specificity of the inhibitory effects of Dad on TGF- β family type I receptors, thickveins, saxophone, and Baboon in *Drosophila*. *FEBS Lett*. 2008;582(17):2496–500.
82. Galeone A, Han SY, Huang C, Hosomi A, Suzuki T, Jafar-Nejad H. Tissue-specific regulation of BMP signaling by *Drosophila* N-glycanase 1. *Elife*. 2017;6:e27612.
83. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, Jiggins CD, Kronforst MR, Long AD, McMillan WO. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences* 2012, 109(31):12632–12637.
84. Özsü N, Chan QY, Chen B, Gupta MD, Monteiro A. Wingless is a positive regulator of eyespot color patterns in *Bicyclus anynana* butterflies. *Dev Biol*. 2017;429(1):177–85.
85. Fenner J, Benson C, Rodriguez-Caro L, Ren A, Papa R, Martin A, Hoffmann F, Range R, Counterman BA. Wnt genes in wing pattern development of Coliadinae butterflies. *Front Ecol Evol*. 2020;8:00197.
86. Litchfield DW, Lüscher B. Casein kinase II in signal transduction and cell cycle regulation. Reversible Protein Phosphorylation in Cell Regulation. 1993:187–199.
87. Lin X, Buff EM, Perrimon N, Michelson AM. Heparan sulfate proteoglycans are essential for FGF receptor signaling during *Drosophila* embryonic development. *Development*. 1999;126(17):3715–23.
88. Bauer A, Chauvet S, Huber O, Usseglio F, Rothbächer U, Aragnol D, Kemler R, Pradel J. Pontin52 and Reptin52 function as antagonistic regulators of β -catenin signalling activity. *EMBO J*. 2000;19(22):6121–30.
89. Tomlinson A, Strapps WR, Heemskerk J. Linking frizzled and wnt signaling in *Drosophila* development. *Development*. 1997;124(22):4515–21.
90. Chen C, Struhl G. Wingless transduction by the Frizzled and Frizzled2 proteins of *Drosophila*. *Development*. 1999;126(23):5441–52.
91. Cadigan KM, Fish MP, Rulifson EJ, Nusse R. Wingless repression of *Drosophila* frizzled 2 expression shapes the wingless morphogen gradient in the wing. *Cell*. 1998;93(5):767–77.
92. Hartfelder K. Endocrine control of insect polyphenism. *Compr Mol Insect Sci*. 2005;3:651–703.
93. Gotoh H, Miyakawa H, Ishikawa A, Ishikawa Y, Sugime Y, Emlen DJ, Lavine LC, Miura T. Developmental link between sex and nutrition; doublesex regulates sex-specific mandible growth via juvenile hormone signaling in stag beetles. *PLoS Genet*. 2014;10(11):e1004098.
94. Gautam S, Kunte K. Adaptive plasticity in wing melanisation of a montane butterfly across a himalayan elevational gradient. *Ecol Entomol*. 2020;45(6):1272–83.
95. Shinoda T, Itoyama K. Juvenile hormone acid methyltransferase: a key regulatory enzyme for insect metamorphosis. *Proceedings of the National Academy of Sciences* 2003, 100(21):11986–11991.
96. Green MR, Sambrook J. Isolation of high-molecular-weight DNA using organic solvents. *Cold Spring Harbor Protocols* 2017, 2017(4):pdb. prot093450.
97. Hill J, Rastas P, Hornett EA, Neethiraj R, Clark N, Morehouse N, Paz Celorio-Mancera M, Cols JC, Dirksen H, Meslin C. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science advances* 2019, 5(6):eaau3648.
98. Robert Hubley AS. RepeatModeler Open-1.0. 2010.
99. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005;110(1–4):462–7.
100. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006;34(suppl2):W435–9.
101. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR genomics and bioinformatics*. 2020;2(2):lqaa026.
102. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.
103. J.D. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. (Version v0.7.0). 2020.
104. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
105. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35.
106. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
107. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650.
108. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1–21.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.