PNAS COMMENTARY

Hidden in plain sight: (Re)definition of a key lepidopteran color patterning gene

Nicholas W. VanKuren[a,1] and Marcus R. Kronforst[a,1]

Although the goal of scientific research may be to uncover ultimate truths, the routes to those truths are often circuitous, spiraling forward and occasionally backward as accumulating theory and experiment guide us to consistent answers. Some steps (both forward and backward) are driven by advances in technology or the availability of a new study system or a new theory. Other steps may be caused by a single serendipitous discovery that compels us to reevaluate established theories and previous experiments. But together, these cycles of self-evaluation conspire to propel science closer and closer to those ultimate truths. At each step, checks and balances between these different aspects of scientific discovery act to reinforce and, importantly, correct the interpretations and theories that came before (1). This iterative process, and therefore scientific progress itself, requires robust, reproducible results, and perhaps some serendipity. Two recent papers in PNAS, one by Livraghi et al. (2) and one by Fandino et al. (3), wonderfully highlight the mechanics and success of this self-correcting nature of science.

Together, these two studies convincingly show that melanic pattern variation in Lepidoptera is controlled by a novel, long noncoding RNA (lncRNA), *ivory*. These results are, frankly, shocking. The last two decades of evolutionary development research on butterfly and moth wing color patterns have shown that a surprisingly large portion of lepidopteran wing color pattern variation is controlled by just a small handful of critical transcription factors, signaling ligands, and receptors. In the hyper-diverse *Heliconius* butterflies, for example, yellow and white patterns are controlled by the transcription factor *aristaless-1* (4), red patterns are controlled by *Optix* (5), the boundaries of certain melanic patterns are controlled by *WntA* (6), and other melanic patterns were repeatedly associated with the protein-coding gene *cortex*.

In fact, over a dozen studies have implicated the *cortex* locus in the control of melanic patterns across Lepidoptera (Fig. 1). This locus controls the switch between mimetic color patterns in swallowtail butterflies (7), sex-limited polymorphism in *Hypolimnas misippus* (8), Müllerian mimicry in *Heliconius* (9, 10), leaf-mimicking camouflage in oakleaf butterflies (11), and melanic silkworm mutants, among others. Perhaps most famously, *cortex* has been thought to control melanic polymorphism in the peppered moth *Biston betularia*, which evolved in parallel with the industrial revolution and is a textbook example of evolution by natural selection (9, 12). Thus, we, as a community of lepidopteran color pattern researchers, have believed that melanic variation was controlled by *cortex* itself for nearly a decade. In fact, around 2020, our lab would joke that "of course it's *cortex*" because so many recent studies had implicated that gene in color pattern development.

Yet numerous, albeit small, gaps or inconsistencies plagued each of these studies. For one thing, each of these studies

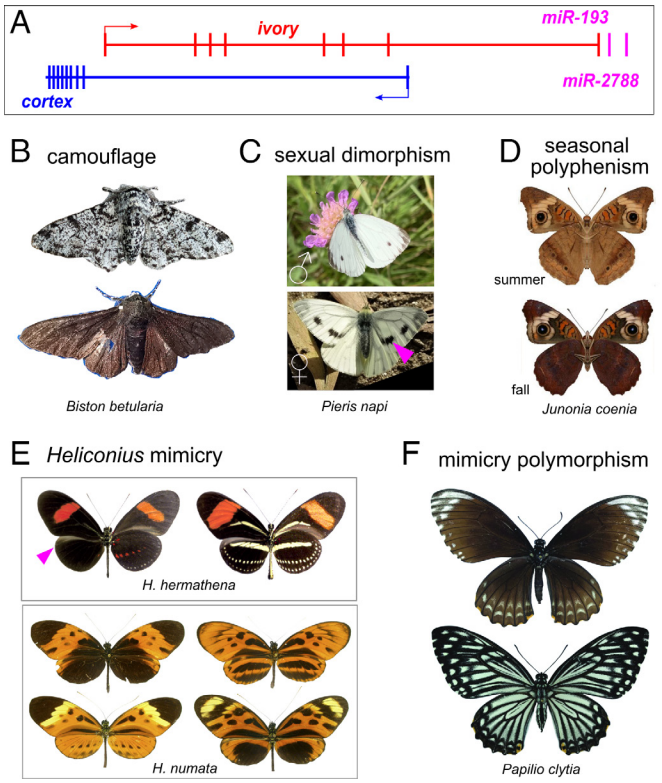

Fig. 1. Melanic variation associated with the *cortex* locus. (*A*) The newly discovered lncRNA *ivory* overlaps with the protein-coding gene *cortex*. (*B–F*) Key examples of melanic variation across Lepidoptera associated with genetic variation in the *cortex* locus. (*B*) The peppered moth provides a classic example of adaptation by natural selection. A transposable element in the *cortex* locus is associated with the evolution of a melanic morph (*Bottom*). (*C*) Sexual dimorphism, including black forewing spots in *Pieris* species such as *Pieris napi,* has been associated with variation in the *cortex* locus. (*D*) Seasonal polyphenism in the common buckeye (covered in detail in ref. 3). (*E*) Many mimetic color patterns in *Heliconius* butterflies are mapped to the *cortex* locus, including introgressed *cortex* alleles in *Heliconius hermathena;* band versus stripe patterns; and complex polymorphism in *Heliconius numata*. (*F*) Mimicry polymorphism in some *Papilio* swallowtails such as *Papilio clytia* is associated with a putative inversion allele in the *cortex* locus. Image credits: (*B*) iNaturalist/Matthew Wilkinson. (*C, Top*) iNaturalist/abumadsen. (*C, Bottom*) iNaturalist/José Antonio León Mangado. (*D*) Wikipedia/Megan McCarty. (*E, Top*) Darli Massardo (University of Chicago, Chicago, IL). (*E, Bottom*) Modified from ref. 13, which is licensed under CC BY 4.0. (*F*) Darli Massardo (University of Chicago, Chicago, IL).

Author affiliations: [a]Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637

Author contributions: N.W.V. and M.R.K. wrote the paper.

The authors declare no competing interest.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See companion articles, "A long noncoding RNA at the *cortex* locus controls adaptive coloration in butterflies," 10.1073/pnas.2403326121 and "The *ivory* lncRNA regulates seasonal color patterns in buckeye butterflies," 10.1073/pnas.2403426121.

[1]To whom correspondence may be addressed. Email: nvankuren@uchicago.edu or mkronforst@uchicago.edu.

Published November 25, 2024.

simply found statistical associations between the *cortex* locus and the melanic pattern of interest. Second, many color patterning genes are expressed in larval or pupal development in wing regions that prefigure the adult color patterns. Although several studies found that *cortex* was differentially expressed as expected, *cortex* expression poorly prefigured adult color patterns (11, 12). Most importantly, recent attempts by Livraghi et al. to disrupt *cortex* function using CRISPR/Cas9 largely failed to produce consistent phenotypes (14). Experiments aimed at disrupting the *cortex* coding sequence or deleting putative *cortex cis*-regulatory elements (CREs) both yielded very rare color pattern phenotypes, with only ~2% efficiency. The most efficient knockouts were recovered from CRISPR/Cas9 experiments targeting the extreme 3′ end of the *cortex* coding sequence (11, 14, 15). This is in stark contrast to knockout experiments for almost every other color patterning gene, where efficiencies are typically 30 to 60%. Surprisingly, the true push to resolve these discrepancies came from a serendipitous mutation in a breeder's population of *Heliconius melpomene* (15). Individual butterflies hetero- or homozygous for the so-called *ivory* deletion were mostly or completely devoid of melanic scales, respectively. Yet this deletion removed a region 50 kb upstream of the *cortex* coding sequence, encompassing a potential alternative *cortex* transcription start site and two microRNAs (*miR-193* and *miR-2788*).

Together, these two studies convincingly show that melanic pattern variation in Lepidoptera is controlled by a novel, long non-coding RNA, ivory.

This luck, combined with persistence and rigorous experiments by Livraghi et al. (2) and Fandino et al. (3) has now largely resolved each of these issues. These two studies provide a comprehensive picture of how a previously undetected gene, the lncRNA *ivory*, not *cortex*, is the lepidopteran melanism gene. These groups performed exceptional CRISPR/Cas9 experiments and established homozygous mutant lines for both *ivory* and *cortex*, firmly establishing *ivory* as the melanism gene and ruling out *cortex*'s role in wing pattern development. Both groups show that *ivory* expression perfectly prefigures adult melanic patterns, while Fandino et al. performed additional experiments testing the function of conserved CREs that control *ivory* expression and its function in color pattern development. In addition to the points above, these authors go the extra step of identifying genes that regulate and are regulated by this novel lncRNA, really getting at the gene regulatory networks underlying these adaptive phenotypes rather than just identifying a causal gene.

Identifying a major color patterning gene is an exciting and impressive achievement on its own. However, the additional fact that *ivory* is a noncoding RNA (ncRNA)—the first ncRNA known to be involved in lepidopteran color patterning—has widespread implications for a variety of questions in evolutionary developmental biology. lncRNAs are known to perform many important roles in animals and plants, some famous examples being the *Xist* RNA essential for mammalian dosage compensation and *COOLAIR* in the regulation of flowering time in *Arabidopsis* (16, 17). There is growing recognition that ncRNAs play important roles in critical developmental processes—evidenced by the most recent Nobel Prize in Medicine or Physiology to V. Ambros and G. Ruvkun. Yet the fact that most ncRNAs are not poly-adenylated means that they are not detected using traditional RNA-sequencing approaches, and this is the main reason that they are frequently missed or ignored in gene annotation sets and downstream exploration. Few lepidopteran genomes have been comprehensively annotated for ncRNAs, with silkworm being a notable exception. Fandino et al. (3) and Livraghi et al. (2) make it clear that it is worth investing the time and effort to ensure that the full complement of functional genes is annotated.

Finally, ncRNAs function in a variety of ways to regulate gene expression, from regulating local chromatin environments to directly binding messenger RNAs and blocking their translation (18). How does *ivory* function in the developing wing to control melanic pattern formation? Fandino et al. (3) provide a detailed first pass on answering this question, identifying potential direct targets of *ivory* and genes that function downstream, and it will be an important next step to test these interactions. It is also important to note that a preprint by Tian et al. provides strong evidence that a miRNA, *miR-193*, at the 3′ end of *ivory*, may be the functional element in this noncoding region (19). It is possible that the *ivory* lncRNA could simply serve as a scaffold (i.e., a primary miRNA) for the expression and processing of *miR-193*. Butterfly wing color patterns depend on tight coordination of temporal and spatial gene expression patterns throughout larval and pupal development. This additional layer of regulation from ncRNAs like *ivory* could provide even more avenues to fine-tune when and where target genes function, and it will be exciting to see how these different types of transcripts comprise the complex developmental gene networks controlling color pattern development.

Thus, even as Fandino et al. (3) and Livraghi et al. (2) solve many mysteries surrounding the control of melanic patterns in Lepidoptera, some important questions remain. But this is the hallmark of good science. Studies such as these reinforce the argument that reproducibility, consistency, and skepticism are all required for a healthy cycle of scientific inquiry. Reassuringly, these two papers highlight the self-correcting nature of scientific research. Inconsistent results cause an itch that must be scratched until it is relieved—when we get as close to ultimate truth as current systems and technologies allow. As is obvious with the *cortex* story, it can often take years of orthogonal experiments to finally home in on the causative variants underlying even the most extreme phenotypes. "Being wrong" has never been a bad thing in science because good scientists cannot let go of those little things that just do not line up. When science is really working, our mistakes serve as jumping-off points for more research that ultimately corrects the record and sets us on a straighter path to the ultimate truth. These two papers show just how robust the self-correcting nature of science remains.

1. T. S. Kuhn, *The Structure of Scientific Revolutions*, O. Neurath, Ed. (University of Chicago Press, 1962).
2. L. Livraghi *et al.*, A long noncoding RNA at the *cortex* locus controls adaptive coloration in butterflies. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2403326121 (2024).
3. R. A. Fandino *et al.*, The *ivory* lncRNA regulates seasonal color patterns in buckeye butterflies. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2403426121 (2024).
4. E. L. Westerman *et al.*, Aristaless controls butterfly wing color variation used in mimicry and mate choice. *Curr. Biol.* **28**, 3469–3474 (2018).
5. A. Martin *et al.*, Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo* **5**, 7 (2014).
6. A. Martin *et al.*, Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12632–12637 (2012).
7. N. W. VanKuren, D. Massardo, S. Nallu, M. R. Kronforst, Butterfly mimicry polymorphisms highlight phylogenetic limits of gene reuse in the evolution of diverse adaptations. *Mol. Biol. Evol.* **36**, 2842–2853 (2019).
8. A. Orteu *et al.*, Optix and cortex/ivory/mir-193 again: The repeated use of two mimicry hotspot loci. *Proc. Biol. Sci.* **291**, 20240627 (2024).
9. N. J. Nadeau *et al.*, The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* **534**, 106–110 (2016).
10. D. Massardo *et al.*, The roles of hybridization and habitat fragmentation in the evolution of Brazil's enigmatic longwing butterflies, Heliconius nattereri and H. hermathena. *BMC Biol.* **18**, 84 (2020).
11. S. Wang *et al.*, The evolution and diversification of oakleaf butterflies. *Cell* **185**, 3138–3152.e20 (2022).
12. A. E. Van't Hof *et al.*, The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105 (2016).
13. A. Meyer, Repeating patterns of mimicry. *PLoS Biol.* **4**, e341 (2006), 10.1371/journal.pbio.0040341.
14. L. Livraghi *et al.*, Cortex cis-regulatory switches establish scale colour identity and pattern diversity in Heliconius. *Elife* **10**, e68549 (2021).
15. J. J. Hanly *et al.*, A large deletion at the cortex locus eliminates butterfly wing patterning. *G3 (Bethesda)* **12**, jkac021 (2022).
16. F. Liu, S. Marquardt, C. Lister, S. Swiezewski, C. Dean, Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* **327**, 94–97 (2010).
17. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
18. L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
19. S. Tian *et al.*, A micro-RNA is the effector gene of a classic evolutionary hotspot locus. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.02.09.579741 (Accessed 1 October 2024).