Co-option of the piRNA pathway to regulate neural crest specification

Thesis by
Riley Galton

In Partial Fulfillment of the Requirements for
the Degree of
Doctor of Philosophy in Biology

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023

(Defended September 22, 2022)

© 2022

Riley Galton
ORCID: 0000-0001-6777-2177

ACKNOWLEDGEMENTS

The completion of my PhD would not have been possible without the support of many people to whom I owe a great deal of gratitude. First and foremost, I would like to thank my co-advisors, Kata Fejes-Toth and Marianne Bronner, whose faith in my abilities never wavered, even when mine did. They allowed me to join both of their labs and pursue a high-risk project of my choosing, without any preliminary data or idea where it would take us, and were always available to enthusiastically discuss new data, or cheer me up when it seemed like none of my experiments were working. When I had my son during my third year of grad school, they were more than supportive as I transitioned into motherhood, and were extremely understanding and patient during that long first year when I couldn't secure consistent childcare. I thank Marianne for the advocacy work she has since done on behalf of student parents, which has made a big difference for many families at Caltech, as well as for the ongoing support she has shown my family. I also thank Kata for her support, as well as for taking the time to personally teach me critical techniques in the lab, like piRNA cloning.

To those who have served on my candidacy and thesis committees over the years—Alexei Aravin, Ellen Rothenberg, Magdalena Zernicka-Goetz, and John Allman—I thank you all for your invaluable input on my work, as well as for the time you have put into serving on my committees. Thank you also to Justin Bois and Mitch Guttman, who decided to start the bioinformatics bootcamp after a conversation we had over beers during my biology recruitment. They then contacted me after recruitment to tell me it was really happening, which ultimately sealed my decision to come to Caltech, and the skills I learned during the bootcamp have carried me through my PhD.

I want to also thank Igor Antoshechkin from the Millard and Muriel Jacobs Genetics and Genomics Laboratory for advice on sequencing and analysis techniques over the years, as well as for making the special effort to send my small RNA libraries out to a third-party sequencing company during the Covid lockdown, which allowed me to analyze all of the small-RNA seq data for my paper before Caltech even opened again.

Thank you to my undergraduate advisor, Frederic Chedin, who enthusiastically nurtured my early interest in molecular biology, and personally trained me in many of the laboratory techniques that I have relied on throughout my PhD. My time in Fred's lab taught me that good science begins with

curiosity, and that getting results—expected or unexpected—is exciting, and strongly informed my decision to pursue a PhD.

I want to thank all the members of the Bronner, Fejes-Toth, and Aravin labs I have overlapped with over the years for their contributions to my training, advice on my work, and making the labs enjoyable places to work. Specifically, I want to thank Maria Ninova, Alicia Rogers, Mike Piacentino, and Meyer Barembaum for providing a great deal of training in bioinformatics or lab specific techniques, as well as always being available to chat and troubleshoot.

Outside of the lab, I want to thank the friends I have made during my PhD journey, who have been at my side through thick and thin. In particular, I want to give a shout-out to my ladies—Alicia Rogers, Maria Ninova, Katie Schretter, and Sofi Quinodoz—who were and always will be available to chat science or just decompress over wine and cheese (whether in person or on zoom from across the country).

Without the ongoing support of my family, completing my PhD would have been impossible. I thank my parents for always supporting me and nurturing my interest in science, as well as for the amazing grandparents they have become to my son, who is always ecstatic to get to stay with Grandma and Grandpa during Covid closures or a particularly busy week of experiments. Thank you to my grandmother, who is always ready to discuss science with me, and has cultivated my interest in biology from a young age. Thank you also to my in-laws, who have been so supportive of my PhD journey and my family over the years.

I perhaps owe the greatest debt of gratitude to my husband, Cory Reider, for always being my biggest support and an amazing father to our children. Cory, you followed me to grad school and now are following me across the country so that I may continue to pursue my passions, and you always pick up the slack in housework and childcare when I need to focus on work, and for this I am immensely grateful. Finally, I want to thank my sons. Indigo, you made me a mom and were such a good baby when you “worked” in the lab with me for almost a year after you were born. Picking you up from school after work is the highlight of my day, and being around you always washes away any stress I’m feeling. Jules, you are almost here, and we are so excited to meet you. Thank you for being an

easy baby to grow, and for allowing me to have the energy and focus to finish my thesis with you on board.

ABSTRACT

The piRNA pathway has persisted throughout evolution as an essential regulatory pathway to protect genomic integrity in the metazoan germline. It achieves this through the repression of transposable elements, or "selfish genes," which would otherwise jump throughout the genome unchecked, causing genomic instability and infertility. While transposable elements are generally deleterious in nature, their persistence in our genome remains an important driver of evolution, both as an agent of mutation and source of raw genetic material. Thus, a delicate balance must be struck to both maintain genomic integrity for the next generation but still enable enough mutation to allow for adaptation. The arms race between the ever-adapting piRNA pathway and its transposon targets provides this balance, and for a long time the piRNA pathway was considered to be germline specific, since that is where both transposon and piRNA pathway activity is highest. It has since become clear that the piRNA pathway is also active in somatic tissue of several invertebrate species, and may even target host genes in some. Whether the piRNA pathway plays a role outside of the germline in vertebrates, however, has remained elusive.

In this thesis, we demonstrate that the piRNA pathway is active in a vertebrate somatic cell type, the chick neural crest, where it has been co-opted into the gene regulatory network to repress the transposon-derived gene, ERNI. ERNI, in turn, supresses Sox2 when piRNA pathway protein Piwil1 is downregulated upon neural crest specification. Thus, the piRNA pathway functions to maintain Sox2 expression in the neural plate border stem cell niche, protecting its proliferative abilities and setting the timing of neural crest specification. We also provide preliminary evidence that the neural crest piRNA pathway might be conserved in other vertebrate species, and that two highly conserved transcription factors regulate its expression in the chick neural crest. Our work provides mechanistic insight into a novel function of the piRNA pathway as a regulator of somatic development in vertebrates, and raises the possibility that this ancient pathway may play a more significant role in evolution and transposon co-option by host genomes than previously thought.

PUBLISHED CONTENT AND CONTRIBUTIONS

Galton, R. et al. (2022). “Co-option of the piRNA pathway to regulate neural crest specification.” In: *Science Advances* 8, eabn1441. DOI: 10.1126/sciadv.abn1441.

R.G. conceptualized the project, designed and executed the experiments, performed the data analysis, and wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

Chapter 1

Introduction

The persistence of life relies on the integrity of genomic information being passed from one generation to the next. It also relies on the ability of species to adapt to new environments, for which a steady source of mutations is critical. This paradoxical necessity to both repress and maintain sources of genomic instability has been a hallmark of evolution since the beginning of time (*1*, *2*).

Transposable elements (TEs) are mobile genetic elements, or "selfish genes," that have the ability to self-replicate within host genomes. When left unchecked, TEs will devastate a genome, rendering the host sterile and thus disrupting the critical flow of genetic information to the next generation (*2*–*4*). To combat this, metazoa have evolved the piRNA pathway, a highly adaptive small RNA interference pathway that sequence specifically represses transposable elements via small RNA guides called piRNAs (*5*–*7*).

The genetic conflict between TEs and the piRNA pathway, both of which are constantly adapting to either evade or maintain control over the other, is often described as an arms race (*3*). It is this precise conflict that maintains life as we know it, striking the perfect balance of maintaining genomic integrity while still allowing for a baseline rate of mutation to drive evolution.

The canonical piRNA pathway

Though constantly adapting to combat new TE threats, the basic components of the piRNA pathway are highly conserved across metazoa (*8*). The core effector complex of this small RNA interference pathway consists of a Piwi clade argonaute protein and its bound piRNA guide, which recognizes target mRNAs through sequence complementarity. Due to the abundance of TE activity in the germline, the piRNA pathway is highly active in germ cells, and thus was long considered to be a germline specific pathway (*3*, *8*).

The most conserved component of this pathway is its cytoplasmic function, in which Piwi proteins, via their piRNA guides, target and cleave mRNAs with their catalytic Piwi domain, thus repressing

transposition post-transcriptionally. The piRNA pathway has also evolved a nuclear function, where specific Piwi proteins can enter the nucleus to target nascent mRNAs, recruiting machinery to instigate chromatin modifications for transcriptional silencing of TEs (*9*). While this nuclear arm of the piRNA pathway is found in many domains of life, some lineages, including birds and fish, appear to have independently lost the Piwi proteins capable of nuclear translocation and thus rely on the cytoplasmic facet of the piRNA pathway to maintain genomic stability (*10*).

Presumably due to its ancient origin and requirement for constant adaptability, the piRNA pathway is particularly complex compared to other small RNA pathways. It relies on a myriad of genes that are unique to the pathway, and new components continue to be uncovered regularly (*9*, *11*). As such, piRNA biogenesis is also complex and distinct from that of other small RNA pathways. It does not rely on the dicer endonuclease that mediates miRNA and siRNA biogenesis, but rather on piRNA pathway specific machinery.

piRNAs are typically classified as either “primary” or “secondary” based on their biogenesis. Primary piRNAs are processed from precursor transcripts transcribed from piRNA clusters, or discrete heterochromatic genomic regions that are enriched in TE sequences and either transcribed bidirectionally or in antisense. These long, piRNA precursor transcripts are then processed into mature primary piRNAs, with the antisense sequences capable of TE silencing when bound to a Piwi protein (*12*).

Secondary piRNAs are generated in the cytoplasm in a process called the ping-pong cycle, where sense and antisense piRNAs are cyclically generated between two Piwi proteins (*6*, *13*). The cycle begins when a Piwi protein bound with an antisense primary piRNA targets and cleaves a sense TE transcript. This cleaved transcript is then loaded onto another Piwi protein where it is trimmed and processed into a secondary piRNA, and can then target antisense transcripts to initiate another round of ping pong amplification, thus specifically amplifying piRNAs against the most active TEs.

Though piRNAs do not have a specific sequence motif or structural requirement, they do contain a bias for a U in position 1, and, in the case of sense oriented secondary piRNAs, an A in position 10. Secondary piRNAs also show an enrichment for a 10 base pair overlap of their 5’ ends, due Piwi’s

cleavage of the target transcript between nucleotides 10 and 11 of the complementary piRNA. piRNA length can vary based on the footprint of the Piwi proteins involved in biogenesis, but ranges from 23-35 nucleotides (*8*, *12*).

The lack of more concrete structural and sequence requirements for piRNAs lends to the pathway's adaptability. By design, any sequence can be processed into functional piRNAs, the only requirement for silencing being that the piRNA sequence is complementary to its target. This leaves room for the potential co-option of the piRNA pathway to regulate sequences far beyond TEs, and possibly play a much broader role in host gene regulation than originally thought.

Non-canonical roles for piRNA pathway

Despite the piRNA pathway's reputation as germline-specific, studies have shown that it is also active in somatic tissue of several invertebrate species. Regenerative models like hydra and planaria rely on the canonical piRNA pathway to repress transposons in their somatic stem cells, and inhibition of the pathway's activity impairs regeneration (*14*, *15*). Intriguingly, the piRNA pathway also targets host genes in planaria and may be involved in cell cycle regulation and immune response (*16*), while, in *Aplysia*, a piRNA mediated mechanism appears to regulate synaptic plasticity (*17*). In the more familiar *D. melanogaster*, piRNAs have also been associated with degradation of maternally deposited host gene mRNA during early embryogenesis, via sequences in their 3' UTRs complementary to transposon derived piRNAs (*18*). Additionally, expression of piRNA pathway components in somatic tissue of many arthropod species hints at a more ancestral role for the piRNA pathway in the soma (*19–21*).

In contrast to these invertebrate models, somatic and non-canonical functions for the piRNA pathway in vertebrate species are not as well understood, and have proven a source of intense debate (*22*, *23*). Piwi proteins and other piRNA pathway genes are upregulated in many somatic cancers and possibly correlate with poor prognostic outcomes (*24*). Piwi expression has also been reported in various vertebrate somatic tissues, including brain, heart and hematopoietic stem cells (*25–27*), and loss of Piwi proteins in the rodent brain have been correlated with behavioral deficits (*26*, *28*). In rat neurons, *Piwil1* has been implicated in regulation of migration-related genes (*29*), while in the mouse

germline it has also been implicated in host transcript stabilization (*30*). Despite this collective evidence that the piRNA pathway likely functions non-canonically in some capacity in vertebrates, mechanistic insights into what these functions are have remained elusive.

Transposable elements as drivers of evolution

As previously mentioned, TEs play a critical role in evolution by providing a steady source of mutation and raw genetic material for novel adaptations. At least 45% of the human genome is made up of TE sequences, and the genomes of most other organisms boast similarly significant proportions of repetitive elements (*2*, *31*, *32*). New TEs emerge often throughout evolution, and TE activity can vary significantly even between closely related species (*2*, *4*, *33*). Co-option of retroviral sequences into host genomes, as well as rewiring of host gene regulatory networks via domestication of long terminal repeats (LTRs), are surprisingly frequent occurrences (*2*, *34*).

TEs can contribute novel mutations to host genomes in several ways, the most intuitive perhaps being insertion into and thus disruption of a host gene or regulatory sequence (*32*, *35*). The imprecise nature of TE transposition mechanisms can also lead to duplication or deletion of nearby host genes, or portions of them, which can generate new genes by exon shuffling as well as lead to larger scale genome rearrangements over time (*35*, *36*). Even long-inactive TE sequences can cause mutation by providing regions of high homology throughout the genome that can facilitate recombination events (*2*, *32*, *35*, *37*).

In addition to the mutation causing effects of TE activity, components of TEs themselves are often co-opted or exapted by host genomes over the course of evolution (*2*, *35*, *38*). One of the most ancient examples of this might be the reverse transcriptase subunit of telomerase, which is thought to be derived from ancient retroelements (*39*, *40*). More recently, *Drosophila*, which lost the telomerase enzyme, have similarly co-opted three LINE-like TEs to maintain chromosome ends (*2*, *41*). Other notable examples of TE exaptation by host genomes include the *Rag1* and *Rag2* genes that catalyze V(D)J recombination in gnathostomes, which are derived from an ancient DNA transposon (*42*, *43*), and *syncytins*, key factors in placental development, which are derived from endogenous retrovirus (ERV) *env* genes (*44*). Additionally, ERV *gag* genes have been independently co-opted by tetrapods

and flies as neuronal *Arc* genes to mediate synaptic transfer of mRNA in the brain (*45*), and in the developing *Galliformes* embryo, *gag* derived gene *ERNI* functions to mediate neural induction (*46*, *47*).

Beyond contributions to gene products and chromosome structure, TEs are increasingly understood to be important modulators of host gene regulation. They contain their own promoters and regulatory elements, and can disperse these elements throughout a genome, potentially rewiring gene regulatory networks by dispersing the same bundle of transcription factor binding sites to new loci (*34*, *48*). LTR retrotransposons, which include ERVs, are often implicated in this process. Most LTR copies have lost their internal coding sequences to recombination, and thus exist solely as regulatory sequences scattered throughout the genome, which can lead to rewiring of host gene regulatory networks, even on a species-specific level (*34*, *48*, *49*).

The contribution of TEs to the evolution of host genomes is indisputable, but the relationship between the two must strike a delicate balance to remain functional. In the germline, this balance is maintained by piRNA pathway mediated repression, but how TEs are modulated in somatic tissues as well as whether there might be a host-mediated, systematic mechanism for TE co-option, remain unanswered.

The neural crest as a model to study evolution

The neural crest (NC) is a transient, embryonic population of pluripotent cells that gives rise to many different cell types in the developing vertebrate embryo (*50*). NC cells are induced at the neural plate border upon gastrulation, and during neurulation carry out a gene regulatory program called specification, which allows them to undergo an epithelial-to-mesenchymal transition and migrate extensively throughout the embryo (*51*, *52*). Upon reaching their final destinations, NC cells differentiate into a diverse array of cell types and structures, including the craniofacial bones and cartilage, much of the peripheral nervous system, and parts of the heart (*52*, *53*).

A vertebrate innovation, the neural crest is thought to have contributed greatly to the success of vertebrates, as outlined in the “new head” hypothesis put forth by Gans and Northcutt in 1983 (*54*,

55). The basis of this hypothesis is that the advent of the neural crest with the divergence of vertebrates led to several key adaptations, like jaws, that permitted the expansion and diversification of vertebrates (*54–56*).

Since then, much has been uncovered about the gene regulatory network underlying NC development (*52*, *57*). Comparative studies have shown that over the course of vertebrate evolution, the NC gene regulatory network has gradually acquired specific subcircuits to shape the "new head" of higher vertebrates (*56*, *58*). Similarly, while the core gene regulatory network governing NC development is highly conserved, lineage and species-specific alterations to it are thought to contribute to the immense morphological diversity among vertebrates (*51*, *52*, *59–61*).

The relatively recent and rapid evolution of this unique population of stem-like cells makes the neural crest an ideal model to dissect the processes that govern gene regulation and evolution at the molecular level. Furthermore, techniques to study neural crest development have been optimized in several systems, allowing for both experimental comparisons and access to a large pool of previously published data.

In this thesis, we investigate the role of the piRNA pathway in a somatic vertebrate tissue, the neural crest. In Chapter Two, we, for the first time, provide mechanistic evidence of an active piRNA pathway outside of the germline in vertebrates. Furthermore, we describe novel roles for the piRNA pathway as a regulator of somatic development and a key player in the co-option of a transposon derived gene. In Chapter Three, we provide preliminary evidence that the neural crest piRNA pathway might be a conserved subcircuit of the neural crest gene regulatory network, and in Chapter Four, we investigate the regulation of the neural crest piRNA pathway itself, identifying two highly conserved transcription factors as potential regulators of the pathway in the neural crest.

References

1. R. L. Cosby, N.-C. Chang, C. Feschotte, Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev*. **33**, 1098–1116 (2019).
2. G. Bourque, K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, D. L. Mager, C. Feschotte, Ten things you should know about transposable elements. *Genome Biol*. **19** (2018), doi:10.1186/s13059-018-1577-z.
3. A. A. Aravin, G. J. Hannon, J. Brennecke, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science (New York, N.Y.)*. **318**, 761–764 (2007).
4. T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A. H. Schulman, A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. **8**, 973–982 (2007).
5. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)*. **316**, 744–747 (2007).
6. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*. **128**, 1089–1103 (2007).
7. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*. **313**, 320–324 (2006).
8. D. M. Ozata, I. Gainetdinov, A. Zoch, D. O'Carroll, P. D. Zamore, PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*. **20**, 89–108 (2019).
9. B. Czech, M. Munafo, F. Ciabrelli, E. L. Eastwood, M. H. Fabry, E. Kneuss, G. J. Hannon, piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet.* **52**, 131–157 (2018).
10. J. Gutierrez, R. Platt, J. C. Opazo, D. A. Ray, F. Hoffmann, M. Vandewege, Evolutionary history of the vertebrate Piwi gene family. *PeerJ*. **9**, e12451 (2021).
11. A. Arif, S. Bailey, N. Izumi, T. A. Anzelon, D. M. Ozata, C. Andersson, I. Gainetdinov, I. J. MacRae, Y. Tomari, P. D. Zamore, GTSF1 accelerates target RNA cleavage by PIWI-clade Argonaute proteins. *Nature*. **608**, 618–625 (2022).

12. K. F. Tóth, D. Pezic, E. Stuwe, A. Webster, The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Adv Exp Med Biol*. **886**, 51–77 (2016).
13. L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M. C. Siomi, A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science*. **315**, 1587–1590 (2007).
14. C. E. Juliano, A. Reich, N. Liu, J. Gotzfried, M. Zhong, S. Uman, R. A. Reenan, G. M. Wessel, R. E. Steele, H. Lin, PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc Natl Acad Sci U.S.A*. **111**, 337–342 (2014).
15. D. Palakodeti, M. Smielewska, Y.-C. Lu, G. W. Yeo, B. R. Graveley, The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA (New York, N.Y.)*. **14**, 1174–1186 (2008).
16. I. V. Kim, E. M. Duncan, E. J. Ross, V. Gorbovytska, S. H. Nowotarski, S. A. Elliott, A. Sanchez Alvarado, C.-D. Kuhn, Planarians recruit piRNAs for mRNA turnover in adult stem cells. *Genes Dev*. **33**, 1575–1590 (2019).
17. P. Rajasethupathy, I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl, E. R. Kandel, A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell*. **149**, 693–707 (2012).
18. C. Rouget, C. Papin, A. Boureux, A.-C. Meunier, B. Franco, N. Robine, E. C. Lai, A. Pelisson, M. Simonelig, Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature*. **467**, 1128–1132 (2010).
19. S. Kawaoka, Y. Arai, K. Kadota, Y. Suzuki, K. Hara, S. Sugano, K. Shimizu, Y. Tomari, T. Shimada, S. Katsuma, Zygotic amplification of secondary piRNAs during silkworm embryogenesis. *RNA*. **17**, 1401–1407 (2011).
20. M. Ninova, S. Griffiths-Jones, M. Ronshaugen, Abundant expression of somatic transposon-derived piRNAs throughout Tribolium castaneum embryogenesis. *Genome Biol*. **18** (2017), doi:10.1186/s13059-017-1304-1.
21. S. H. Lewis, K. A. Quarles, Y. Yang, M. Tanguy, L. Frézal, S. A. Smith, P. P. Sharma, R. Cordaux, C. Gilbert, I. Giraud, D. H. Collins, P. D. Zamore, E. A. Miska, P. Sarkies, F. M. Jiggins, Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol*. **2**, 174–181 (2018).

22. E.-C. Cheng, D. Kang, Z. Wang, H. Lin, PIWI proteins are dispensable for mouse somatic development and reprogramming of fibroblasts into pluripotent stem cells. *PloS one*. **9**, e97821–e97821 (2014).
23. J. P. Tosar, C. Rovira, A. Cayota, Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol*. **1**, 1–8 (2018).
24. R. J. Ross, M. M. Weiner, H. Lin, PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*. **505**, 353–359 (2014).
25. A. K. Sharma, M. C. Nelson, J. E. Brandt, M. Wessman, N. Mahmud, K. P. Weller, R. Hoffman, Human CD34(+) stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi. *Blood*. **97**, 426–434 (2001).
26. S. Nandi, D. Chandramohan, L. Fioriti, A. M. Melnick, J. M. Hebert, C. E. Mason, P. Rajasethupathy, E. R. Kandel, Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. *Proc Natl Acad Sci U.S.A* (2016), doi:10.1073/pnas.1609287113.
27. B. P. U. Perera, Z. T.-Y. Tsai, M. L. Colwell, T. R. Jones, J. M. Goodrich, K. Wang, M. A. Sartor, C. Faulk, D. C. Dolinoy, Somatic expression of piRNA and associated machinery in the mouse identifies short, tissue-specific piRNA. *Epigenetics*. **14**, 504–521 (2019).
28. L. J. Leighton, W. Wei, P. R. Marshall, V. S. Ratnu, X. Li, E. L. Zajaczkowski, P. A. Spadaro, N. Khandelwal, A. Kumar, T. W. Bredy, Disrupting the hippocampal Piwi pathway enhances contextual fear memory in mice. *Neurobiol Learn Mem*. **161**, 202–209 (2019).
29. P.-P. Zhao, M.-J. Yao, S.-Y. Chang, L.-T. Gou, M.-F. Liu, Z.-L. Qiu, X.-B. Yuan, Novel function of PIWIL1 in neuronal polarization and migration via regulation of microtubule-associated proteins. *Mol. Brain*. **8**, 39–39 (2015).
30. A. Vourekas, Q. Zheng, P. Alexiou, M. Maragkakis, Y. Kirino, B.D. Gregory, Z. Mourelatos, Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat. Struct. Mol. Biol.* **19** (2012), doi:10.1038/nsmb.2347.
31. A. F. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev*. **9**, 657–663 (1999).
32. P. L. Deininger, J. V. Moran, M. A. Batzer, H. H. Kazazian, Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev*. **13**, 651–658 (2003).

33. R. Rahman, G. Chirn, A. Kanodia, Y. A. Sytnikova, B. Brembs, C. M. Bergman, N. C. Lau, Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. *Nucleic Acids Res*. **43**, 10655–10672 (2015).
34. G. Bourque, B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J.-L. Chew, Y. Ruan, C.-L. Wei, H. H. Ng, E. T. Liu, Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res*. **18**, 1752–1762 (2008).
35. R. Cordaux, M. A. Batzer, The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. **10**, 691–703 (2009).
36. N. Jiang, Z. Bao, X. Zhang, S. R. Eddy, S. R. Wessler, Pack-MULE transposable elements mediate gene evolution in plants. *Nature*. **431**, 569–573 (2004).
37. K. Han, J. Lee, T. J. Meyer, P. Remedios, L. Goodwin, M. A. Batzer, L1 recombination-associated deletions generate human genomic variation. *Proc Natl Acad Sci U S A*. **105**, 19366–19371 (2008).
38. Z. Joly-Lopez, T. E. Bureau, Exaptation of transposable element coding sequences. *Curr Opin Genet Dev*. **49**, 34–42 (2018).
39. M. Belfort, M. J. Curcio, N. F. Lue, Telomerase and retrotransposons: Reverse transcriptases that shaped genomes. *Proc Natl Acad Sci U S A*. **108**, 20304–20310 (2011).
40. N. Fulcher, E. Derboven, S. Valuchova, K. Riha, If the cap fits, wear it: an overview of telomeric structures over evolution. *Cell Mol Life Sci*. **71**, 847–865 (2014).
41. M.-L. Pardue, P. G. DeBaryshe, Retrotransposons that maintain chromosome ends. *Proc Natl Acad Sci U S A*. **108**, 20317–20324 (2011).
42. V. V. Kapitonov, E. V. Koonin, Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol Direct*. **10**, 20 (2015).
43. S. Huang, X. Tao, S. Yuan, Y. Zhang, P. Li, H. A. Beilinson, Y. Zhang, W. Yu, P. Pontarotti, H. Escriva, Y. Le Petillon, X. Liu, S. Chen, D. G. Schatz, A. Xu, Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell*. **166**, 102–114 (2016).
44. A. Dupressoir, C. Lavialle, T. Heidmann, From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta*. **33**, 663–671 (2012).
45. E. D. Pastuzyn, C. E. Day, R. B. Kearns, M. Kyrke-Smith, A. V. Taibi, J. McCormick, N. Yoder, D. Belnap, S. Erlendsson, D. R. Morado, J. A. G. Briggs, C. Feschotte, J. D. Shepherd, The

Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*. **172**, 275-288.e18 (2018).

46. A. Streit, A. J. Berliner, C. Papanayotou, A. Slrulnik, C. D. Stern, Initiation of neural induction by FGF signalling before gastrulation. *Nature*. **406**, 74–78 (2000).
47. C. Papanayotou, A. Mey, A. M. Birot, Y. Saka, S. Boast, J. C. Smith, J. Samarut, C. D. Stern, A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLoS Biology*. **6**, 0109–0123 (2008).
48. C. Feschotte, The contribution of transposable elements to the evolution of regulatory networks. *Nat Rev Genet*. **9**, 397–405 (2008).
49. D. R. Fuentes, T. Swigut, J. Wysocka, Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *Elife*. **7**, e35989 (2018).
50. N. M. Le Douarin, E. Dupin, The neural crest in vertebrate evolution. *Curr Opin Genet Dev*. **22**, 381–389 (2012).
51. S. A. Green, M. Simoes-Costa, M. E. Bronner, Evolution of vertebrates as viewed from the crest. *Nature*. **520**, 474–482 (2015).
52. M. Simoes-Costa, M. E. Bronner, Establishing neural crest identity: a gene regulatory recipe. *Development (Cambridge, England)*. **142**, 242–257 (2015).
53. H. C. Etchevers, E. Dupin, N. M. Le Douarin, The diverse neural crest: from embryology to human pathology. *Development*. **146** (2019), doi:10.1242/dev.169821.
54. C. Gans, R. G. Northcutt, Neural crest and the origin of vertebrates: a new head. *Science*. **220**, 268–273 (1983).
55. R. Glenn Northcutt, The new head hypothesis revisited. *J Exp Zool B Mol Dev Evol*. **304**, 274–297 (2005).
56. M. L. Martik, M. E. Bronner, Riding the crest to get a head: neural crest evolution in vertebrates. *Nat Rev Neurosci*. **22**, 616–626 (2021).
57. T. Sauka-Spengler, D. Meulemans, M. Jones, M. Bronner-Fraser, Ancient evolutionary origin of the neural crest gene regulatory network. *Dev Cell*. **13**, 405–420 (2007).
58. M. L. Martik, S. Gandhi, B. R. Uy, J. A. Gillis, S. A. Green, M. Simoes-Costa, M. E. Bronner, Evolution of the New Head by gradual acquisition of neural crest regulatory circuits. *Nature*. **574**, 675–678 (2019).

59. C. F. Kratochwil, L. Geissler, I. Irisarri, A. Meyer, Molecular Evolution of the Neural Crest Regulatory Network in Ray-Finned Fish. *Genome Biol Evol*. **7**, 3033–3046 (2015).

60. A. H. Newton, Marsupials and Multi-Omics: Establishing New Comparative Models of Neural Crest Patterning and Craniofacial Development. *Front Cell Dev Biol*. **10**, 941168 (2022).

61. Y. Wakamatsu, K. Suzuki, Sequence alteration in the enhancer contributes to the heterochronic Sox9 expression in marsupial cranial neural crest. *Dev Biol*. **456**, 31–39 (2019).

C h a p t e r 2

Co-option of the piRNA pathway to regulate neural crest specification

Riley Galton, Katalin Fejes-Toth*, Marianne E. Bronner*

This work was first published as:

Galton, R. et al. (2022). "Co-option of the piRNA pathway to regulate neural crest specification." In: *Science Advances* 8, eabn1441. DOI: 10.1126/sciadv.abn1441.

Abstract

Across Metazoa, Piwi proteins play a critical role in protecting the germline genome through piRNA-mediated repression of transposable elements. In vertebrates, activity of Piwi proteins and the piRNA pathway was thought to be gonad specific. Our results reveal the expression of Piwil1 in a vertebrate somatic cell type, the neural crest. Piwil1 is expressed at low levels throughout the chicken neural tube, peaking in neural crest cells just before the specification event that enables epithelial-to-mesenchymal transition (EMT) and migration into the periphery. Loss of Piwil1 impedes neural crest specification and emigration. Small RNA sequencing reveals somatic piRNAs with sequence signatures of an active ping-pong loop. RNA-seq and functional experiments identify the transposon-derived gene ERNI as Piwil1's target in the neural crest. ERNI, in turn, suppresses Sox2 to precisely control the timing of neural crest specification and EMT. Our data provide mechanistic insight into a novel function of the piRNA pathway as a regulator of somatic development in a vertebrate species.

Introduction

Small noncoding RNAs and their protein partners, Argonaute proteins, play central regulatory roles in transcriptional and posttranscriptional gene expression in all domains of life (*1*). The Piwi clade of Argonaute proteins is unique to Metazoa, where they are required for maintenance of stemness, self-renewal, and safeguarding of the genome by repressing transposable elements (TEs) in the germ cell lineage (*2–4*). TEs are mobile genetic elements that can replicate and reinsert themselves in the

genome, threatening genomic integrity. By keeping these "selfish genes" in check, the Piwi-interacting RNA (piRNA) pathway helps preserve genomic stability and thus plays a critical role in the arms race between TEs and their host genomes (*5*). piRNAs recognize TE transcripts via sequence complementarity, which, in the cytoplasm, leads to TE target cleavage by the Piwi protein. In several organisms, Piwi proteins have gained the ability to enter the nucleus and instigate chromatin modifications at target loci. piRNA biogenesis differs from that of their better-known relatives, microRNAs (miRNAs) and small interfering RNAs, and relies in part on cleavage by Piwi proteins themselves in an amplification cycle termed the ping-pong amplification loop (*3, 6*), as well as on other cytoplasmic factors that are unique to this pathway (*7, 8*). These differences lead to characteristics of piRNAs that are distinct from other small RNAs: piRNAs are slightly longer than miRNAs at around 23 to 30 nucleotides (nt), predominantly map to TE sequences, have a 1-U bias at their 5′ end, and, in the case of ping-pong–generated piRNAs, a 10-A bias and a characteristic 10-nt overlap between complementary sequences (*3, 6, 9*). Due to 3′ end processing, they also carry a 2′-O-methyl residue, which enables differential cloning (*10*).

TEs are extremely prevalent and found in all metazoan genomes. New lineage-specific TEs emerge often during the course of evolution, and the repertoire of active transposons can vary widely among species (*11, 12*). Despite their negative impact on genomic stability when left unchecked, TEs provide a steady source of germline mutations and are considered an important driver of evolution. There are many examples of host genome co-option of retroviral genes, as well as domestication of long terminal repeats (LTRs) that bind host transcription factors to rewire gene regulatory networks, many of which act in somatic tissues (*11, 13, 14*).

While Piwi proteins and the piRNA pathway perform critical functions in the germ line to repress TEs, their potential role in somatic cells is not as well understood. It has long been queried whether the piRNA pathway might be active outside the germ line or used to regulate genes other than TEs. In invertebrate models such as Hydra and Planaria with high regenerative capacity, somatic stem cells have been shown to use the canonical piRNA pathway to actively repress TEs (*15, 16*). In addition, Planaria Piwi protein SMEDWI-3 can regulate non-transposon mRNAs (*17*). In the sea slug Aplysia, a piRNA-mediated mechanism has been shown to mediate epigenetic regulation of

synaptic plasticity (*18*). piRNAs and piRNA pathway genes are also expressed in somatic tissues of numerous arthropods (*19–21*). These findings suggest that somatic piRNA-mediated regulation may be widespread. In contrast to these invertebrate models, a functional role for the piRNA pathway in somatic cells of vertebrates has been debated (*22, 23*). Piwi protein expression has been observed in cancer cells (*24*) and some adult tissues including hematopoietic stem cells (*25*), brain, and heart (*26, 27*). Loss of mouse Piwi protein Mili has been correlated with behavioral deficits (*27*), and Piwil1 has been implicated in neuronal migration in rats (*28*), although mechanistic insights remain elusive and whether there is piRNA involvement in these processes is unclear.

Here, we report somatic Piwi and piRNA expression in a vertebrate embryonic cell type, the neural crest. The neural crest is a rapidly evolving, migratory population of stem cells that is unique to vertebrates and essential for their development and evolution (*29*). While neural crest cells undergo specification within the forming central nervous system during neurulation, they subsequently leave the neural tube to migrate extensively throughout the embryo and contribute to a diverse array of tissues (*30, 31*). Our functional analysis shows that chick Piwil1 (Chiwi) is required for neural crest cell emigration from the neural tube. Chiwi regulates expression of TE-derived gene ERNI, a regulator of Sox2 expression during early nervous system formation, through an active somatic piRNA pathway. Our results indicate that the gene regulatory network controlling neural crest development has co-opted a transposon-derived sequence and its piRNA-mediated regulation to precisely time neural crest specification and initiation of the epithelial-to-mesenchymal transition (EMT) from the neural tube to begin neural crest migration.

Results

Chiwi exhibits a distinct expression pattern in the developing neural tube

As a first step in examining whether the piRNA pathway plays a role in regulating vertebrate developmental events, we assessed the expression of Piwi genes at early embryonic stages. Vertebrates have two conserved Piwi proteins, Piwil1 and Piwil2, which play central roles in the germline piRNA pathway (*32–34*). To test whether Piwi transcripts are expressed in somatic cells of the early vertebrate embryo, we examined Piwil1 (Chiwi) and Piwil2 (Chili) expression during chick

embryogenesis by quantitative reverse transcription polymerase chain reaction (qRT-PCR) using cDNA derived from whole embryos from stages ranging from Hamburger and Hamilton (HH) stages (*35*) HH4 to HH23 (fig. S1). We observed extremely low Chili transcript levels at all stages. In contrast, Chiwi transcript levels were markedly higher, especially at early stages, peaking at HH6 to HH8, corresponding to neurulation stages, and dropping after HH10 to HH12, corresponding to the time of active cranial neural crest migration. Primordial germ cells are extraembryonic until stage HH13 (~2 days of development), when they enter the embryonic bloodstream through which they migrate to the gonadal anlagen by around HH18 (3 days) (*36*). Thus, high Chiwi mRNA levels before this embryonic stage indicate that Chiwi is expressed in somatic cells.

To confirm the RT-qPCR results and assess Chiwi expression with spatiotemporal resolution, we performed hybridization chain reaction (HCR) on HH6 to HH12 embryos in whole mount (Fig. 1A and fig. S2A). Transverse sections revealed low but ubiquitous Chiwi expression throughout the cranial region at HH6, which then becomes primarily constrained to the neural tube at HH9 and onward (Fig. 1B and fig. S2B). This shift from ubiquitous to tissue-specific expression in the cranial region mirrors the drop in Chiwi expression that we observed between HH6 to HH8 and HH10 to HH12 cDNA via qPCR. Chiwi expression appears to be differentially regulated in the dorsal tube at HH9, where there are distinguishable subdomains of neural crest precursors that differ in their expression of neural crest specifier genes (Fig. 1, A and B) (*30, 37*). In the dorsal midline, there are cells that express both Snai2 and Pax7, corresponding to premigratory neural crest cells in the process of undergoing EMT (*30, 38*). In this domain, Chiwi expression is reduced compared to the rest of the neural tube. In contrast, in the immediately lateral domain that feeds into the specified neural crest pool (*30*), cells are marked by high Pax7 but no Snai2 and relatively high levels of Chiwi expression.

We next confirmed the presence of Chiwi in the neural crest region by RNA sequencing (RNA-seq). To this end, we used three RNA-seq datasets, one from dissected cranial neural folds (two replicates) and two previously published fluorescence-activated cell sorted specified neural crest datasets (*39*), which include early migrating cranial neural crest cells from HH9 stage embryos (three replicates) and trunk neural crest cells from HH18 embryos (three replicates). All three datasets reveal notable

expression levels of Chiwi mRNA, with respective transcripts per million (TPM) of 6.1 and 6.5 in the specified cranial and trunk datasets and a TPM of 13.9 in the cranial neural fold dataset. Chili was detected in neural folds with a TPM of 4.5 but was not detectable in sorted migrating neural crest cells (fig. S2C). Together, these results support the somatic expression of Chiwi in both neural crest precursors and early migrating neural crest cells of the chicken embryo, as well as low-level expression of Chili in dorsal neural folds.

Somatic piRNAs target transposons

To address whether Chiwi could exert a regulatory function in the neural crest that is directed by associated piRNAs, we tested for the presence of piRNAs in the cranial region. Due to a lack of appropriate antibodies, it was not feasible to immunoprecipitate Chiwi and directly analyze associated small RNAs. As an alternative, we cloned and sequenced small RNAs from the midbrain region of HH9 heads, where Chiwi expression is very pronounced and confined to the neural tube. In parallel, to specifically enrich for piRNAs, we performed small RNA cloning that included an oxidation step to select for RNA species that are 2′-O-methylated at their 3′ end, a characteristic of piRNAs (Fig. 1C).

Roughly 64 and 60% of the reads in the total and oxidized samples mapped to the genome with no mismatches. Among these, both the total and oxidized samples showed comparable numbers of reads mapping to ribosomal RNA (rRNA) at around 6% (Fig. 1D). Total small RNA reads mostly mapped to miRNAs, with some mapping to tRNAs and to genes in sense orientation (Fig. 1D). Sense mapping reads are unable to target the corresponding mRNAs and likely represent degradation products, while antisense mapping reads have the potential to silence transcripts via the piRNA pathway. Less than 1% of the mapped reads in the total small RNA samples corresponded to transposon sequences, mostly in antisense orientation, confirming that piRNA expression in this tissue is extremely low relative to other small RNA species. The representation of different TE families in total TE-mapping reads, however, was similar to the oxidized samples, indicating that the total small RNA library contains piRNAs, even if they represent only a small fraction of all reads (fig. S3A). Consistent with predominance of miRNAs and degradation products in the non-oxidized libraries, the size distribution of the total small RNA samples peaked at 22 nt and showed a broader

distribution across all sizes (Fig. 1E). In contrast, 44% of the reads in the oxidized samples mapped to TEs, with 96% of these in antisense orientation (Fig. 1D), and the size distribution of these libraries peaked at 28 nt (Fig. 1E), consistent with the size profile of previously published chick embryonic piRNA libraries (*40*). Interestingly, we noted an enrichment of tRNA mapping reads in addition to TE-mapping reads in the oxidized libraries, consistent with tRNAs being heavily 2′-O-methylated (Figs 1D) (*41*). Upon plotting the size distribution of tRNA mapping reads, we saw a strong enrichment of 23- and 25-nt sequences in the oxidized samples, possibly suggesting the presence of tRNA-derived small RNAs with 2′-O-methylated 3′ ends (fig. S3B).

To further characterize the presumptive piRNA population, we analyzed the size profiles of sense and antisense TE-mapping reads in the oxidized samples. Interestingly, we found that sense reads peaked between 25 and 27 nt, while the more abundant antisense reads had a strong peak at 28 nt (Fig. 1F and fig. S3C). While the RT-qPCR and RNA-seq data indicated very low Chili expression (Fig. 1A and fig. S2C), these size profiles suggest that more than one Piwi protein might be present, albeit at different expression levels, and contributing to piRNA biogenesis via the ping-pong amplification mechanism. Consistent with this idea, we observed a strong enrichment for 10–base pair (bp) overlap between complementary reads in the oxidized libraries when we analyzed 5′ to 5′ overlap of TE-mapping reads (Fig. 1G), as well as a notable 1-U bias of antisense and 10-A bias of sense sequences upon collapsing the libraries (Fig. 1H). These findings provide strong support for an active somatic piRNA pathway in the embryo and imply that somatic piRNAs are, at least in part, generated by a ping-pong mechanism, possibly through the interplay of two Piwi proteins.

To determine whether somatic piRNAs might be generated by a similar mechanism to germline piRNAs, we looked for expression of genes involved in germline piRNA biogenesis in RNA-seq data from dorsal neural folds (fig. S4). We found that several key factors are expressed at low levels, comparable to those of Chiwi and Chili. These include Hen1 (TPM 4.0), which imparts 2′-O-methylation to piRNAs (*7*), Tudor Domain Containing 9 (TDRD9) (TPM 1.4), which is involved in ping-pong piRNA biogenesis (*7*), PNLDC1 [poly(A)-specific ribonuclease-like domain containing 1] (TPM 7.6), which trims the 3′ ends of piRNAs (8), and, possibly, the Zucchini homolog phospholipase D6 (PLD6; TPM 9.0) (*7*), although RNA-seq tracks of PLD6 are noisier than those

of the other genes. Interestingly, the transcription factor A-MYB, which mediates expression of both piRNA clusters and piRNA pathway genes, is also present (TPM 3.0), as are the two Tudor domain–containing genes TDRD1 (TPM 7.7) and TDRD3 (TPM 53.5), which A-MYB has been shown to regulate along with Chiwi in rooster testes (*42*). Together, expression of piRNA pathway genes in the dorsal neural tube raises the possibility that a similar mechanism of piRNA biogenesis to that seen in the germ line might be at play.

Loss of Chiwi disrupts neural crest development

On the basis of Chiwi's distinct expression pattern, we hypothesized that it plays a role in neural crest development. To probe Chiwi's function in the neural crest, we first performed loss-of-function experiments using a translation-blocking Chiwi morpholino oligomer (MO). After electroporating Chiwi MO into the prospective neural crest region on one side of the embryo and control MO on the other side at HH4, we allowed the embryos to grow until HH9, when cranial neural crest begins to delaminate and migrate. Chiwi appears to be relatively strongly expressed in the dorsal neural tube at this stage (Fig. 2A and fig. S5). By immunostaining for the neural crest marker Pax7, we observed a reduction in neural crest migration distance from the midline on the Chiwi-depleted side. Upon sectioning, we saw significant reduction in the numbers of Pax7-positive cells on the Chiwi-depleted side (Fig. 2B).

To confirm the specificity of our results, we used a second approach to knock out Chiwi via a plasmid-based CRISPR-Cas9 system optimized for use in chicken (*43*). To this end, we designed two CRISPR guide RNAs against sequences corresponding to the first exon junction and the 3′ piRNA binding site of Chiwi. We electroporated a Cas9 construct along with control guide on one side and Chiwi guides on the other side of HH4 stage embryos. HCR for Chiwi after CRISPR-Cas9 knockout confirmed a reduction in Chiwi mRNA levels compared to the control side at stage HH9, just before the onset of neural crest migration (fig. S6). CRISPR-Cas9–mediated loss of Chiwi gave the same reduction in neural crest migration distance and cell count at HH9 as did the Chiwi morpholino (Fig. 2B).

Chiwi's piRNA binding activity is necessary for its function in the neural crest

Due to the presence of piRNAs in the cranial midbrain region, we hypothesized that Chiwi identifies its targets via the associated piRNAs. To test whether Chiwi's piRNA binding activity is necessary for its function in the neural crest, we created wild-type and mutant Chiwi overexpression constructs. The mCMV-YK-Chiwi construct contains two amino acid substitutions in positions 574 and 578, resulting in impaired piRNA binding (*44, 45*). To express these at levels similar to those of endogenous Chiwi, we cloned cDNA amplicons into a construct containing a minimal cytomegalovirus (mCMV) promoter, which has weak global expression in chick. The vector also contains an Histone2B-RFP (H2B-RFP) sequence following an internal ribosome entry site downstream of the Chiwi sequence. As a control, we used the empty vector, which expresses only the H2B-RFP under the mCMV promoter. While overexpression of wild-type Chiwi did not significantly alter neural crest migration, it did result in an increase in Pax7-positive cells. In contrast, overexpression of the piRNA binding mutant resulted in reduced neural crest migration distance and cell count (Fig. 2C) compared to the control, recapitulating the Chiwi loss of function phenotype and indicating that Chiwi's piRNA binding ability is required for its function in the neural crest.

Chiwi regulates a single, transposon-derived gene, ERNI, in the dorsal neural tube

To investigate Chiwi targets in the neural crest, we performed RNA-seq experiments after CRISPR-Cas9–mediated knockout. Differential mRNA expression analysis of control and knockout RNA-seq libraries generated from total RNA of neural folds at stage HH9 revealed that the most significantly up-regulated gene upon Chiwi knockout is the retrotransposon embryonic normal stem cell (ENS-1), also called Soprano (Fig. 3A). ENS-1 appears to have expanded recently within the Galliforme lineage and has numerous, highly homologous copies in the genome, most of which only contain the LTR (*46, 47*). Several copies, however, contain only the 5′ end of the internal domain, which codes for ERNI (*46, 48*), a chicken-specific transposon-derived gene and a known regulator of neural induction (*49, 50*).

To validate our RNA-seq results, we performed HCR against the ERNI mRNA sequence (Fig. 3B). In wild-type embryos, this revealed ERNI expression throughout the neural tube and ectoderm, albeit

at varying levels. In particular, we noted higher expression in the dorsalmost region of the neural tube, corresponding to the Snai2+ domain in which Chiwi transcripts are down-regulated. Consistent with this, ERNI was reduced compared to the control side after Chiwi overexpression, whereas Chiwi depletion and overexpression of the piRNA binding mutant resulted in an increase in ERNI (Fig. 3C).

Analysis of small RNA reads mapping to the ERNI mRNA sequence, which includes portions of both the ENS-1 internal domain and LTR, showed enrichment in the oxidized piRNA sample compared to total small RNAs, accounting for 0.01% of mapped total small RNA reads and 0.30% of mapped oxidized reads, and displayed the 5′ to 5′ complementarity signature of the ping-pong pathway (Fig. 3, D and E), indicating that Chiwi is regulating ERNI via a piRNA-mediated mechanism. Both the RNA-seq and the small RNA samples showed enrichment specifically over the LTR and ERNI sequence and did not map to other internal parts of the ENS-1 transposon (Fig. 3F). Together, these results indicate that ERNI expression in the dorsal neural tube is spatially regulated by Chiwi in a piRNA-dependent fashion.

Perturbation of ERNI recapitulates Chiwi phenotypes

To functionally test whether dysregulation of ERNI could account for the observed neural crest defects upon Chiwi perturbation, we directly disrupted ERNI expression in the dorsal neural tube and analyzed its effect on neural crest cell count and migration. To this end, we generated an overexpression vector, enh195-FLAG-ERNI, which encodes an N-terminally FLAG-tagged ERNI coding sequence under control of the Pax7 enhancer enh-195 (Fig. 4A) (*51*). We chose to perturb ERNI only within the PAX7+ domain to avoid disturbing the earlier process of neural induction, where ERNI is known to play a pivotal role. We electroporated this construct into one side of the embryo at HH4, with a control version driving citrine on the other side. Pax7 staining of HH9 embryos revealed a reduction in neural crest cell count and migration distance, similar to that seen after Chiwi knockout (Fig. 4B); this is consistent with Chiwi regulating ERNI expression. The ERNI protein harbors an N-terminal coiled-coil domain, which is responsible for recruitment of ERNI to the Sox2 N2 enhancer, and a C-terminal Heterochromatin Protein 1 (HP1) box, which recruits HP1-γ, thereby inducing transcriptional repression (*50*). To further test ERNI's function, we created a

dominant negative ERNI construct, enh195-N150-FLAG-ERNI-NLS (N150), which only contains the first 150 amino acids of ERNI. This includes the Sox2 localizing coiled-coil domain but lacks the HP1-γ binding domain. We also added a nuclear localization signal on the C terminus to ensure nuclear localization, which we confirmed by immunostaining (fig. S7). Overexpression of this dominant negative construct imparted a reciprocal phenotype to wild-type ERNI overexpression, with an increased number of Pax7+ neural crest cells at HH9 (Fig. 4B), recapitulating the Chiwi overexpression phenotype. Together, these data suggest that a primary role of Chiwi during neural crest development is to regulate ERNI expression in a spatiotemporal manner.

ERNI regulates Sox2 during neural crest specification

Since ERNI has been shown to regulate Sox2 expression at early gastrula stages (*49, 50*), we next tested whether it also plays a role in Sox2 regulation in the dorsal neural tube during neural crest development. To address this, we performed HCR for Sox2 on embryos into which we electroporated the ERNI overexpression and ERNI dominant negative constructs. We noted a decrease in Sox2 expression in the Pax7+ domain of embryos, where ERNI expression is increased, and an increase in Sox2 expression in this domain of embryos that expressed the N150 dominant negative construct (Fig. 4C); this is consistent with ERNI inducing Sox2 repression. Together, these results suggest that ERNI functions to repress Sox2 in the dorsal neural tube, thereby enabling activation of the gene program to specify bona fide neural crest cells, which is inhibited by high levels of Sox2 (*30*). Chiwi, in turn, is responsible for repressing ERNI to maintain Sox2 expression in the Pax7+/Snai2− dorsal neural tube cells, prior to the onset of specification and emigration from the neural tube (Fig. 4D).

Discussion

Host co-option of TE components is a relatively common occurrence and an important driver of evolution. Until now, the piRNA pathway had not been implicated in this process. Here, we show that the piRNA pathway, canonically functioning in the germ line to repress deleterious expression of TEs, has been co-opted in the chicken embryo to spatiotemporally regulate the expression of TE-derived gene ERNI and precisely time neural crest specification and EMT. To our knowledge, this

is the first demonstration of co-option of the piRNA pathway to regulate a developmentally relevant gene, expanding our understanding of the piRNA pathway from solely being an antagonistic force against TEs to playing a critical role in the domestication of a TE-derived gene.

ERNI is known to regulate Sox2 at the onset of neural induction in the chicken embryo (*50*) and is derived from an endogenous retrovirus sequence that appears to be unique to the Galliforme lineage (*46*). ERNI expression has also been observed in chicken embryonic stem cells and in the embryonic gonads, where it appears to be a marker of pluripotency (*48, 52*). The fact that ERNI has embedded itself in the highly conserved processes of neural crest specification and neural induction raises the possibility that similar regulation of development and differentiation by piRNA-mediated regulation of TE sequences might be happening in other species. Intriguingly, another TE-derived gene, Crestin, is expressed in specified neural crest cells of zebrafish, although its function is unclear (*53*). We hypothesize that the piRNA pathway plays a conserved role in vertebrate somatic development, although the transposon sequences through which it exerts its function likely change more frequently, in concert with the ever-evolving TE landscape of vertebrates. By constantly updating the pool of piRNAs to match the evolving TE profiles, the piRNA pathway naturally provides the needed plasticity to repress whichever TE is co-opted by the gene regulatory network in a given species or tissue. Consistent with this idea, we observed piRNAs from several TE families in the chick midbrain, suggesting that despite the fact that only ERNI is being regulated by the piRNA pathway in this context, the machinery necessary to regulate other TEs in the same spatiotemporal manner is already present.

In the dorsal region of the chick neural tube, we posit that down-regulation of Sox2 by ERNI permits Pax7-positive cells to complete the neural crest specification program and undergo EMT, becoming migratory neural crest cells. However, when we overexpress ERNI, we see less Sox2 in the dorsal neural tube but fewer neural crest cells, which challenges our understanding of Sox2 as a promoter of neural fate and inhibitor of neural crest in this tissue (*54, 55*). One possible explanation for this observation is that ERNI might also regulate other genes in addition to Sox2 in the dorsal neural tube. We postulate, however, that very precise Sox2 levels are required to maintain the proliferative ability of Pax7+/Snai2− cells, such that reducing Sox2 in this region of the neural tube–which

expresses pluripotency markers and is thought to act as a stem cell niche, contributing to both neural and neural crest fates (*37, 55*)–leads to a loss of its ability to self-renew. This model of Sox2's role in neural crest development mirrors its function in other cell types. Sox2 is a well-known stem cell factor, required for maintenance of several developmental stem cell niches, and precise maintenance of its expression levels can have a notable effect on cell fate. For example, a recent study found that Sox2 levels in the chick tail bud modulate a stem cell population driving secondary neurulation, with very low levels of Sox2 required to maintain proliferation in the stem cell niche and overexpression of Sox2 instigating differentiation into neural epithelium and also reducing the self-proliferative properties of these cells (*56*). Thus, we hypothesize that the piRNA pathway may function as a guardian of the stem cell niche that feeds into the specified neural crest region, regulating proliferation by repressing ERNI to maintain Sox2 expression and only permitting ERNI to switch off Sox2 when it is time for a cell to undergo EMT and become bona fide migratory neural crest.

The piRNA pathway has long been thought to be confined to the germ line in most organisms because expression of its components is typically not observed in other tissues, particularly in vertebrates. Unlike the germline piRNA pathway, it is worth noting that the neural crest piRNA pathway requires relatively low levels of Piwil1 and its associated piRNAs to function. The fact that Piwil1 can have a substantial effect at these levels suggests that there may be other somatic piRNA pathways previously missed due to low expression profiles. This, in turn, raises the intriguing possibility that repurposing of the piRNA pathway for somatic gene regulation may be a widespread occurrence, perhaps shedding light on the myriad reports of Piwi protein expression in diverse cancers (*24*). Somatic regulation by Piwi proteins not only leads to interesting new functions of the piRNA pathway outside of the gonads, but also implies that activation of the piRNA pathway, and thus its adaptation as an epigenetic tool in research and therapy, may be an attainable goal.

Materials and Methods

Cloning of expression vectors

The mCMV-H2B-RFP construct was generated by replacing the rabbit β-actin promoter of PCI-H2B-RFP with an mCMV promoter sequence using the restriction sites Spe I and Xba I. This promoter also contains a 5xTetO site, which we left uninduced to achieve weak expression. The Chiwi coding sequence (Ensembl transcript ENSGALT00000004171.6) was PCR-amplified from HH10 to HH12 whole-embryo cDNA, prepared using oligo(dT) primers and the SuperScript III reverse transcriptase kit, and subsequently cloned into the mCMV-H2B-RFP vector using the restriction sites Asc I and Xho I to create mCMV-Chiwi. The mutant construct (mCMV-YK-Chiwi) was generated using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit to generate the Y574I and K578E amino acid substitutions, which have been previously described to inhibit piRNA binding (*44, 45*). As a marker for electroporation efficiency, pCI-H2B-RFP was coelectroporated alongside the mCMV constructs, as mCMV-H2B-RFP expression is too low to pick up after methanol dehydration of embryos.

The ERNI coding sequence [National Center for Biotechnology Information (NCBI) RefSeq: NM_001080874.1] was PCR-amplified from cDNA obtained from reverse transcribing HH9 head RNA using oligo(dT) and cloned with an N-terminal FLAG tag into the enh195-citrine vector (*51*) using fusion PCR, Eco RI, and Bsr GI (enh195-FLAG-ERNI). The coding sequence for the first 150 amino acids (N150) was similarly cloned but with an N-terminal FLAG tag and a C-terminal nuclear localization signal (enh195-N150-FLAG-ERNI-NLS), which was confirmed to localize to the nucleus by FLAG staining (fig. S7).

Electroporation

Fertilized chicken eggs were acquired from various providers, most recently from Sun State Ranch (Sylmar, CA), and grown at 37°C for 18 to 20 hours to reach HH4 and HH5. Ex ovo electroporations were performed on stage HH4 and HH5 embryos as previously described. Embryos were dissected onto rings of filter paper in Ringer's, and a solution of DNA expression construct or MO was injected

into the space between the vitelline membrane and ectoderm (Fig. 3A) and electroporated into the ectoderm with five pulses of 5.2 V for 50 ms, with 100 ms between each pulse. Embryos were then cultured at 37°C in thin albumen with penicillin-streptomycin until HH9. All embryos were bilaterally electroporated with the control on the left side and experimental on the right side, allowing for direct comparison, and electroporated with either a fluorescent reporter or tagged protein to allow for confirmation of electroporation coverage.

Expression constructs were injected at a concentration of 1 μg/μl, while MOs were used at a concentration of 0.25 mM with pCIG-GFP (green fluorescent protein) (1.0 μg/μl) as carrier DNA. Fluorescein isothiocyanate–labeled MOs used include standard control MO (5′-CCTCTTACCTCAGTTACAATTTATA-3′) and Chiwi translation blocking MO (5′-TCTGGCTCTAGCTCTTCCTGTCATG-3′) from Gene Tools.

CRISPR-mediated knockout was performed as described previously (43). A Cas9-expressing construct (pCAG-nls-hCas9-nls-eGFP) was coelectroporated alongside two guide RNA constructs targeting Chiwi, one in the first exon (5′-GGGAGGTCTCCCTCTCGCTC-3′) and the other in the piRNA binding region (GAATGTGACGGTAGGACCTG), or alongside a nonbinding control guide (5′-GCACTGCTACGATCTACACC-3′).

Reverse transcription quantitative polymerase chain reaction

RNA was extracted from batches of wild-type embryos dissected within the area pellucida using TRIzol and reverse transcribed using SuperScript III reverse transcriptase with random hexamers according to the manufacturer's suggestions. qPCR was performed on an Eppendorf RealPlex using MyTaq mix, SYBR Green, and the respective primers averaging the Ct values for technical triplicates. Chiwi and Chili Ct values at different developmental stages were normalized to 18S rRNA (delta Ct), and fold difference was calculated to Chiwi levels at HH4 to analyze relative expression.

Small RNA-seq

The midbrain region of wild-type HH9− chicken embryo heads was dissected, and RNA was purified using TRIzol. Two replicates of several pooled embryos each were collected. Two micrograms of total RNA was then run on a 15% denaturing polyacrylamide gel, and small RNAs within a 19- to ~38-nt range were isolated and gel extracted as previously described (*57*). Half of each sample was then oxidized in borate buffer [5× solution (pH 8.6); 150 mM borax and 150 mM boric acid] and 25 mM sodium periodate for 25 min at 25°C, while the other half of the sample was incubated in buffer only. Samples were then ethanol precipitated, and libraries were cloned using the NEBNext Small RNA Library Prep Set for Illumina (E7330S) according to the protocol with NEBNext Multiplex Oligos for Illumina (set 1 E7335S). To optimize for the extremely low concentration of RNA and avoid overamplification of adapter dimers, adapters were diluted 1:10 for cloning of the oxidized samples. Libraries were sequenced on an Illumina HiSeq X Ten (150-bp reads, single end) at a sequencing depth of ~20 million reads. Reads were trimmed and mapped to the chicken genome (galGal6) using Bowtie (*58*), with no mismatches end to end and multimapping reads included (-V 0 -k 1 --best). Reads were analyzed for sequence length, gene type that they mapped to, and whether they mapped sense or antisense to a feature. Reads mapping to TEs were extracted and analyzed for 5′ to 5′ complementarity using a previously published script that we altered to include reads ranging from 19 to 34 nt (*20*), and the ping-pong z score was calculated as described previously (*59*) by taking the difference of the value at position 10 and the mean of the background values (values of all positions but 10), divided by the SD of the background values. TE-mapping reads were then analyzed for mapping orientation by read length, and after deduplication of libraries, sequence logos were generated for the first 18 nt of each sequence using WebLogo (*60*). Due to the low number of sense mapping reads and overrepresentation of some sequences, collapsing of libraries was necessary to resolve the 10-A bias. Reads were normalized to reads per million mapped reads (to the genome) unless otherwise stated.

To generate a heatmap of small RNA reads mapping to TEs, reads mapping to different TE families were counted with FeatureCounts (*61*) using a RepeatMasker GTF. Reads were then normalized by reads per million mapped reads to the genome, and hierarchical gene clustering was performed using

Cluster 3.0 (*62*). Normalized read counts were then adjusted by log2, and a heatmap was generated using Morpheus software (*63*).

To analyze small RNAs mapping to the ENS-1 and ERNI loci, reads were mapped to chicken TE consensus sequences from Repbase (*64*) using Bowtie with three mismatches allowed and reporting all valid alignments (-V 3 -a --best --strata). Reads were separately mapped to the ERNI mRNA sequence (RefSeq: NM_001080874.1) using the same parameters. Reads were normalized to reads per million reads that map to the genome.

mRNA-seq

Cranial neural folds of two replicate batches of HH9− embryos electroporated at HH4 with CRISPR control constructs on the left side and Chiwi CRISPR constructs on the right side were dissected, and RNA was isolated using the RNAqueous-Micro Total RNA Isolation Kit (Ambion). Libraries were prepared and sequenced on an Illumina HiSeq 2500 by the Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech at a depth of ~60 million reads (50-bp reads, single end). Reads were trimmed for adapter sequences and mapped to the galGal6 genome using Bowtie2 (*65*), and reads mapping to genes and TEs were counted with FeatureCounts. Differential expression analysis was performed using DESeq2 (*66*). To analyze Chiwi and Chili levels in sorted cranial and trunk neural crest cells, previously published raw data (NCBI BioProject no. PRJNA497902) (*39*) were obtained and mapped to galGal6 using Bowtie2.

TPM counts were generated with TPMCalculator (*67*) using the Ensembl galGal6 GTF (GRCg6a, International Nucleotide Sequence Database Collaboration (INSDC) Assembly GCA_000002315.5) (*68*), with the coordinates for the predicted RefSeq Chili locus (XM_025142807.1) added, as Chili is not currently annotated in Ensembl. Read coverage plots were generated using the UCSC Genome Browser (*69*) and normalized to reads per million mapped reads. Ensembl tracks were used for gene models in the figures for all transcripts except Chili, which is not annotated in Ensembl, and PLD6, which is annotated slightly differently in RefSeq than Ensembl, with the RefSeq version more closely matching transcript sequences from other vertebrates and our RNA-seq data.

Immunofluorescence

All embryos were fixed for 20 min at room temperature in 4% paraformaldehyde, and subsequently blocked in 10% goat or donkey serum in PBST (phosphate-buffered saline and 0.5% Tween 20) for 2 hours at room temperature. Both primary and secondary antibody incubations occurred at 4°C for two nights in 10% goat or donkey serum, with four 1-hour washes in PBST at room temperature after primary and two 30-min washes in PBST after secondary antibody incubation. After imaging, whole-mount embryos were postfixed in 4% paraformaldehyde overnight at 4° before sectioning. The following primary antibodies were used: mouse immunoglobulin G1 anti-Pax7 (1:10; Developmental Studies Hybridoma Bank), rabbit anti-FLAG (1:1000; Sigma-Aldrich, F7425), and goat anti-GFP (1:500; Rockland, 600-101-215). The following secondary antibodies were used: Molecular Probes donkey or goat secondary antibody conjugated to Alexa Fluor 488, Alexa Fluor 568, or Alexa Fluor 647 (1:1000).

In situ HCR

All HCR was performed with probes designed by Molecular Technologies and following the published V3 protocol (*70*). Twenty probe sets were used for all genes except Sox2, for which a 12-probe set was used. Before sectioning, embryos were postfixed in 4% paraformaldehyde overnight at 4°C. Since Chiwi expression was very low and had not been previously reported in the neural tube, we confirmed that the signal was not due to autofluorescence or background amplification by performing a negative control with only the odd HCR probe set, which imparted no signal when imaged under the same conditions as the even+odd probe experiments (fig. S8).

Sectioning

Cryosectioning was performed at a thickness of 18 μm on a Microm HM550 cryostat. Embryo preparation included fixation in 4% paraformaldehyde overnight at 4°C (either from live embryos or postfix processed embryos), followed 15% sucrose overnight at 4°C and 7.5% gelatin overnight at 37°C before mounting in silicone molds and snap freezing in liquid nitrogen.

Imaging and statistical analysis

All images were taken using a Zeiss Axio Imager M2 with an ApoTome.2. Quantification of HCR intensity in wild-type neural tube images was performed using the Plot Profile feature in Fiji (*71*), which was then normalized to the length of the neural tube segment from ventral to dorsal midline and binned at 200. Fluorescent intensity was normalized to the highest value for each channel. Four to five cranial sections each from two HH9 embryos were analyzed. Whole-mount images of cranial neural crest stained for PAX7 were analyzed in Fiji by measuring the area of the migratory crest on the experimental (right) side of the embryo and dividing it by the area of the migratory neural crest on the left (control) side. Cell counts of PAX7-positive cells in sections of the cranial region were taken with the Analyze Particles feature in Fiji as previously described (*72*). Fluorescence intensity for experimental quantification was measured from maximum intensity projections of Z-stack images by manually drawing regions of interest to measure average intensity and subtracting average intensity of background regions. Experimental values were then divided by control values from the same image. For fluorescent intensity and cell count quantification, three nonadjacent cranial sections were measured and averaged to create a representative value for each embryo. For all statistical analysis on images, a paired two-tailed Student's t test was performed to compare two values (experimental and control) within single embryos.

Acknowledgments

We thank members of the M.E.B., K.F.-T., and Aravin laboratories for helpful discussions. We thank M. Ninova for advice on RNA-seq analysis and for making the ping-pong script available for our use. We also thank M. Piacentino for advice with imaging analysis and providing Fiji macros for our use. We thank Q. Tang for providing us with the mini CMV promoter. We acknowledge the Caltech Millard and Muriel Jacobs Genetics and Genomics Laboratory for library prep and sequencing of our CRISPR RNA-seq experiment and, in particular, thank I. Antoshechkin for advice on data analysis and ensuring that our small RNA libraries were sequenced during the COVID-19 lockdown.

Funding: This work is supported by NIH grants R01GM110217 to K.F.-T. and R35NS111564 to M.E.B. R.G. was supported by the NSF's GRFP fellowship.

Author contributions: Conceptualization: R.G., K.F.-T., and M.E.B. Investigation: R.G. Formal analysis: R.G. Visualization: R.G. Funding acquisition: R.G., K.F.-T., and M.E.B. Supervision: K.F.-T. and M.E.B. Writing (original draft): R.G. Writing (review and editing): R.G., K.F.-T., and M.E.B.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All raw sequencing data generated for this publication are available through Gene Expression Omnibus. Neural fold RNA-seq data (Fig. 3 and figs. S2 and S4) are available with accession number GSE171615 and small RNA-seq data (Figs. 1 and 3and fig. S3) with GSE171616. Previously published specified neural crest datasets (fig. S2) are available from NCBI BioProject no. PRJNA497902. All scripts used in this publication are included in table S2.

References

1. E. S. Cenik, P. D. Zamore, Argonaute proteins. *Curr. Biol.* **21**, R446–R449 (2011).

2. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).

3. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).

4. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).

5. A. A. Aravin, G. J. Hannon, J. Brennecke, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).

6. L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M. C. Siomi, A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).

7. T. Watanabe, H. Lin, Post-transcriptional regulation of gene expression by Piwi proteins and piRNAs. *Mol. Cell* **56**, 18–27 (2014).

8. D. Ding, J. Liu, K. Dong, U. Midic, R. A. Hess, H. Xie, E. Y. Demireva, C. Chen, PNLDC1 is essential for piRNA 3′ end trimming and transposon silencing during spermatogenesis in mice. *Nat. Commun.* **8**, 819 (2017).

9. W. Wang, M. Yoshikawa, B. W. Han, N. Izumi, Y. Tomari, Z. Weng, P. D. Zamore, The initial uridine of primary piRNAs does not create the tenth adenine that is the hallmark of secondary piRNAs. *Mol. Cell* **56**, 708–716 (2014).

10. T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. *Nat. Struct. Mol. Biol.* **14**, 349–350 (2007).

11. G. Bourque, K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, D. L. Mager, C. Feschotte, Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).

12. T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A. H. Schulman, A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).

13. G. Bourque, B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J.-L. Chew, Y. Ruan, C.-L. Wei, H. H. Ng, E. T. Liu, Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 (2008).

14. D. Jangam, C. Feschotte, E. Betrán, Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **33**, 817–831 (2017).

15. C. E. Juliano, A. Reich, N. Liu, J. Gotzfried, M. Zhong, S. Uman, R. A. Reenan, G. M. Wessel, R. E. Steele, H. Lin, PIWI proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. *Proc Natl Acad Sci U.S.A.* **111**, 337–342 (2014).

16. D. Palakodeti, M. Smielewska, Y.-C. Lu, G. W. Yeo, B. R. Graveley, The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* **14**, 1174–1186 (2008).

17. I. V. Kim, E. M. Duncan, E. J. Ross, V. Gorbovytska, S. H. Nowotarski, S. A. Elliott, A. Sanchez Alvarado, C.-D. Kuhn, Planarians recruit piRNAs for mRNA turnover in adult stem cells. *Genes Dev.* **33**, 1575–1590 (2019).

18. P. Rajasethupathy, I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl, E. R. Kandel, A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693–707 (2012).

19. S. Kawaoka, Y. Arai, K. Kadota, Y. Suzuki, K. Hara, S. Sugano, K. Shimizu, Y. Tomari, T. Shimada, S. Katsuma, Zygotic amplification of secondary piRNAs during silkworm embryogenesis. *RNA* **17**, 1401–1407 (2011).

20. M. Ninova, S. Griffiths-Jones, M. Ronshaugen, Abundant expression of somatic transposon-derived piRNAs throughout *Tribolium castaneum* embryogenesis. *Genome Biol.* **18**, 184 (2017).

21. S. H. Lewis, K. A. Quarles, Y. Yang, M. Tanguy, L. Frézal, S. A. Smith, P. P. Sharma, R. Cordaux, C. Gilbert, I. Giraud, D. H. Collins, P. D. Zamore, E. A. Miska, P. Sarkies, F. M. Jiggins, Pan-arthropod analysis reveals somatic piRNAs as an ancestral defense against transposable elements. *Nat. Ecol. Evol.* **2**, 174–181 (2018).

22. E.-C. Cheng, D. Kang, Z. Wang, H. Lin, PIWI proteins are dispensable for mouse somatic development and reprogramming of fibroblasts into pluripotent stem cells. *PLOS ONE* **9**, e97821 (2014).

23. J. P. Tosar, C. Rovira, A. Cayota, Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol.* **1**, 2 (2018).

24. R. J. Ross, M. M. Weiner, H. Lin, PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505**, 353–359 (2014).

25. A. K. Sharma, M. C. Nelson, J. E. Brandt, M. Wessman, N. Mahmud, K. P. Weller, R. Hoffman, Human CD34(+) stem cells express the hiwi gene, a human homologue of the *Drosophila* gene piwi. *Blood* **97**, 426–434 (2001).

26. B. P. U. Perera, Z. T.-Y. Tsai, M. L. Colwell, T. R. Jones, J. M. Goodrich, K. Wang, M. A. Sartor, C. Faulk, D. C. Dolinoy, Somatic expression of piRNA and associated machinery in the mouse identifies short, tissue-specific piRNA. *Epigenetics* **14**, 504–521 (2019).

27. S. Nandi, D. Chandramohan, L. Fioriti, A. M. Melnick, J. M. Hebert, C. E. Mason, P. Rajasethupathy, E. R. Kandel, Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. *Proc Natl Acad Sci U.S.A.* **113**, 12697–12702 (2016).

28. P.-P. Zhao, M.-J. Yao, S.-Y. Chang, L.-T. Gou, M.-F. Liu, Z.-L. Qiu, X.-B. Yuan, Novel function of PIWIL1 in neuronal polarization and migration via regulation of microtubule-associated proteins. *Mol. Brain* **8**, 39–39 (2015).

29. N. M. Le Douarin, E. Dupin, The neural crest in vertebrate evolution. *Curr Opin Genet Dev.* **22**, 381–389 (2012).

30. M. Simões-Costa, M. E. Bronner, Establishing neural crest identity: A gene regulatory recipe. *Development***142**, 242–257 (2015).

31. H. C. Etchevers, E. Dupin, N. M. Le Douarin, The diverse neural crest: From embryology to human pathology. *Development* **146**, dev169821 (2019).

32. S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Filippov, H. Blaser, E. Raz, C. B. Moens, R. H. A. Plasterk, G. J. Hannon, B. W. Draper, R. F. Ketting, A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebra fish. *Cell* **129**, 69–82 (2007).

33. A. Wilczynska, N. Minshall, J. Armisen, E. A. Miska, N. Standart, Two Piwi proteins, Xiwi and Xili, are expressed in the *Xenopus* female germline. *RNA* **15**, 337–345 (2009).

34. Y. H. Sun, L. H. Xie, X. Zhuo, Q. Chen, D. Ghoneim, B. Zhang, J. Jagne, C. Yang, X. Z. Li, Domestic chickens activate a piRNA defense against *Avian leukosis virus*. *eLife* **6**, e24695 (2017).

35. V. Hamburger, H. L. Hamilton, A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).

36. P. D. Nieuwkoop, L. A. Sutasurya, in *Primordial Germ Cells in the Chordates: Embryogenesis and Phylogenesis* (CUP Archive, 1979); https://books.google.com/books/about/Primordial_Germ_Cells_in_the_Chordates.html?id=Zu08AAAAIAAJ.

37. A. Lignell, L. Kerosuo, S. J. Streichan, L. Cai, M. E. Bronner, Identification of a neural crest stem cell niche by spatial genomic analysis. *Nat. Commun.* **8**, 1830 (2017).

38. D. Sakai, T. Suzuki, N. Osumi, Y. Wakamatsu, Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development. *Development* **133**, 1323–1333 (2006).

39. M. L. Martik, S. Gandhi, B. R. Uy, J. A. Gillis, S. A. Green, M. Simoes-Costa, M. E. Bronner, Evolution of the New Head by gradual acqu tion of neural crest regulatory circuits. *Nature* **574**, 675–678 (2019).

40. K.-W. Chang, Y.-T. Tseng, Y.-C. Chen, C.-Y. Yu, H.-F. Liao, Y.-C. Chen, Y.-F. E. Tu, S.-C. Wu, I.-H. Liu, M. Pinskaya, A. Morillion, B. Pain, S.-P. Lin, Stage-dependent piRNAs in

chicken implicated roles in modulating male germ cell development. *BMC Genomics* **19**, 425 (2018).

41. L. Ayadi, A. Galvanin, F. Pichot, V. Marchand, Y. Motorin, RNA ribose methylation (2′-*O*-methylation): Occurrence, biosynthesis and biological functions. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**, 253–269 (2019).

42. X. Z. Li, C. K. Roy, X. Dong, E. Bolcun-Filas, J. Wang, B. W. Han, J. Xu, M. J. Moore, J. C. Schimenti, Z. Weng, P. D. Zamore, An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell* **50**, 67–81 (2013).

43. S. Gandhi, M. L. Piacentino, F. M. Vieceli, M. E. Bronner, Optimization of CRISPR/Cas9 genome editing for loss-of-function in the early chick embryo. *Dev. Biol.* **432**, 86–97 (2017).

44. A. Webster, S. Li, J. K. Hur, M. Wachsmuth, J. S. Bois, E. M. Perkins, D. J. Patel, A. A. Aravin, Aub and Ago3 are recruited to nuage through two mechanisms to form a ping-pong complex assembled by Krimper. *Mol. Cell* **59**, 564–575 (2015).

45. S. Djuranovic, M. K. Zinchenko, J. K. Hur, A. Nahvi, J. L. Brunelle, E. J. Rogers, R. Green, Allosteric regulation of Argonaute proteins by miRNAs. *Nat. Struct. Mol. Biol.* **17**, 144–150 (2010).

46. E. Lerat, A.-M. Birot, J. Samarut, A. Mey, Maintenance in the chicken genome of the retroviral-like cENS gene family specifically expressed in early embryos. *J. Mol. Evol.* **65**, 215–227 (2007).

47. T. Wicker, J. S. Robertson, S. R. Schulze, F. A. Feltus, V. Magrini, J. A. Morrison, E. R. Mardis, R. K. Wilson, D. G. Peterson, A. H. Paterson, R. Ivarie, The repetitive landscape of the chicken genome. *Genome Res.* **15**, 126–136 (2005).

48. H. Acloque, V. Risson, A. M. Birot, R. Kunita, B. Pain, J. Samarut, Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo. *Mech. Dev.* **103**, 79–91 (2001).

49. A. Streit, A. J. Berliner, C. Papanayotou, A. Slrulnik, C. D. Stern, Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**, 74–78 (2000).

50. C. Papanayotou, A. Mey, A. M. Birot, Y. Saka, S. Boast, J. C. Smith, J. Samarut, C. D. Stern, A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLOS Biol.* **6**, 0109–0123 (2008).

51. R. M. Williams, I. Candido-Ferreira, E. Repapi, D. Gavriouchkina, U. Senanayake, I. T. C. Ling, J. Telenius, S. Taylor, J. Hughes, T. Sauka-Spengler, Reconstruction of the global neural crest gene regulatory network in vivo. *Dev Cell* **51**, 255–276.e7 (2019).

52. S. Intarapat, C. D. Stern, Sexually dimorphic and sex-independent left-right asymmetries in chicken embryonic gonads. *PLOS ONE* **8**, e69893 (2013).

53. A. L. Rubinstein, D. Lee, R. Luo, P. D. Henion, M. E. Halpern, Genes dependent on zebrafish cyclops function identified by AFLP differential gene expression screen. *Genesis* **26**, 86–97 (2000).

54. Y. Wakamatsu, Y. Endo, N. Osumi, J. A. Weston, Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Dev. Dyn.* **229**, 74–86 (2004).

55. D. Roellig, J. Tan-Cabugao, S. Esaian, M. E. Bronner, Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cells. *eLife* **6**, e21620 (2017).

56. T. Kawachi, E. Shimokita, R. Kudo, R. Tadokoro, Y. Takahashi, Neural-fated self-renewing cells regulated by Sox2 during secondary neurulation in chicken tail bud. *Dev. Biol.* **461**, 160–171 (2020).

57. A. K. Rogers, K. Situ, E. M. Perkins, K. F. Toth, Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes Dev.* **31**, 1858–1869 (2017).

58. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

59. Z. Zhang, J. Xu, B. S. Koppetsch, J. Wang, C. Tipping, S. Ma, Z. Weng, W. E. Theurkauf, P. D. Zamore, Heterotypic piRNA ping-pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol. Cell* **44**, 572–584 (2011).

60. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.***14**, 1188–1190 (2004).

61. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

62. M. J. L. de Hoon, S. Imoto, J. Nolan, S. Miyano, Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).

63. Morpheus, https://software.broadinstitute.org/morpheus/.

64. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

65. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

67. R. Vera Alvarez, L. S. Pongor, L. Mariño-Ramírez, D. Landsman, TPMCalculator: One-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960–1962 (2019).

68. A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. IIsley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek, Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).

69. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

70. H. M. T. Choi, M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis, J. Stegmaier, A. Cunha, N. A. Pierce, Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

71. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

72. M. L. Piacentino, M. E. Bronner, Intracellular attenuation of BMP signaling via CKIP-1/Smurf1 is essential during neural crest induction. *PLOS Biol.* **16**, e2004425 (2018).

Figures

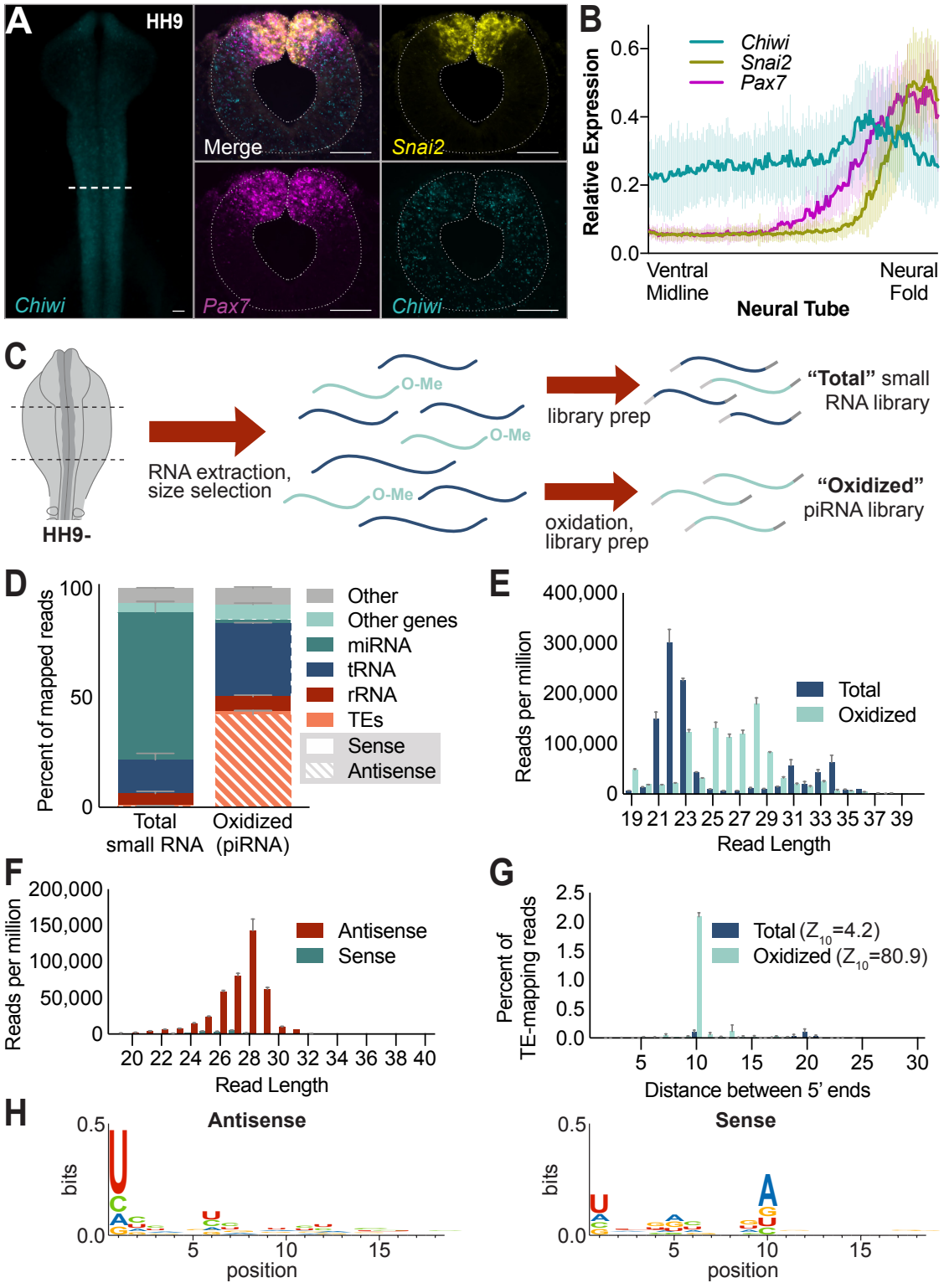
A
HH9
Merge
Snai2
Chiwi
Pax7
Chiwi
B
Chiwi
Snai2
Pax7
Relative Expression
0.6
0.4
0.2
0.0
Ventral Midline
Neural Fold
Neural Tube
C
HH9-
RNA extraction, size selection
O-Me
O-Me
O-Me
library prep
"Total" small RNA library
oxidation, library prep
"Oxidized" piRNA library
D
Percent of mapped reads
100
50
0
Total small RNA
Oxidized (piRNA)
Other
Other genes
miRNA
tRNA
rRNA
TEs
Sense
Antisense
E
Reads per million
400,000
300,000
200,000
100,000
0
Total
Oxidized
19 21 23 25 27 29 31 33 35 37 39
Read Length
F
Reads per million
200,000
150,000
100,000
50,000
0
Antisense
Sense
20 22 24 26 28 30 32 34 36 38 40
Read Length
G
Percent of TE-mapping reads
2.5
2.0
1.5
1.0
0.5
0.0
Total (Z10=4.2)
Oxidized (Z10=80.9)
5 10 15 20 25 30
Distance between 5' ends
H
Antisense
bits
0.5
0.0
5 10 15
position
Sense
bits
0.5
0.0
5 10 15
position

Figure 1. Piwi protein and piRNA expression in the cranial region.

(A) HCR reveals the expression of Chiwi and the neural crest markers, Pax7 and Snai2; scale bars, 50 μm. (B) Quantification of the intensity of HCR signal of Chiwi, Pax7, and Snai2 from the ventral midline to the dorsal neural folds of the neural tube (n = 4 to 5 sections each from two HH9 embryos). (C) Schematic diagram of the small RNA cloning strategy from the midbrain region of HH9− embryos (n = 2 biological replicates). (D) Annotation of small RNAs mapping to the genome. Orientation is relative to the annotated feature. "Other" category includes reads that could not be assigned to a feature, as well as reads mapping to simple repeats, satellite repeats, small cytoplasmic RNA, and small nuclear RNA, which together account for <1% of mapped reads in all samples. (E) Length distribution of all reads mapping to the genome in total and oxidized libraries. (F) Length distribution of reads from oxidized libraries mapping to TEs in sense and antisense orientation. (G) Analysis of 5′ to 5′ distance of complementary small RNA reads mapping to TEs in total and oxidized libraries. (H) Sequence logos of oxidized, collapsed sequences mapping to TEs in antisense (left) and sense (right) orientation.

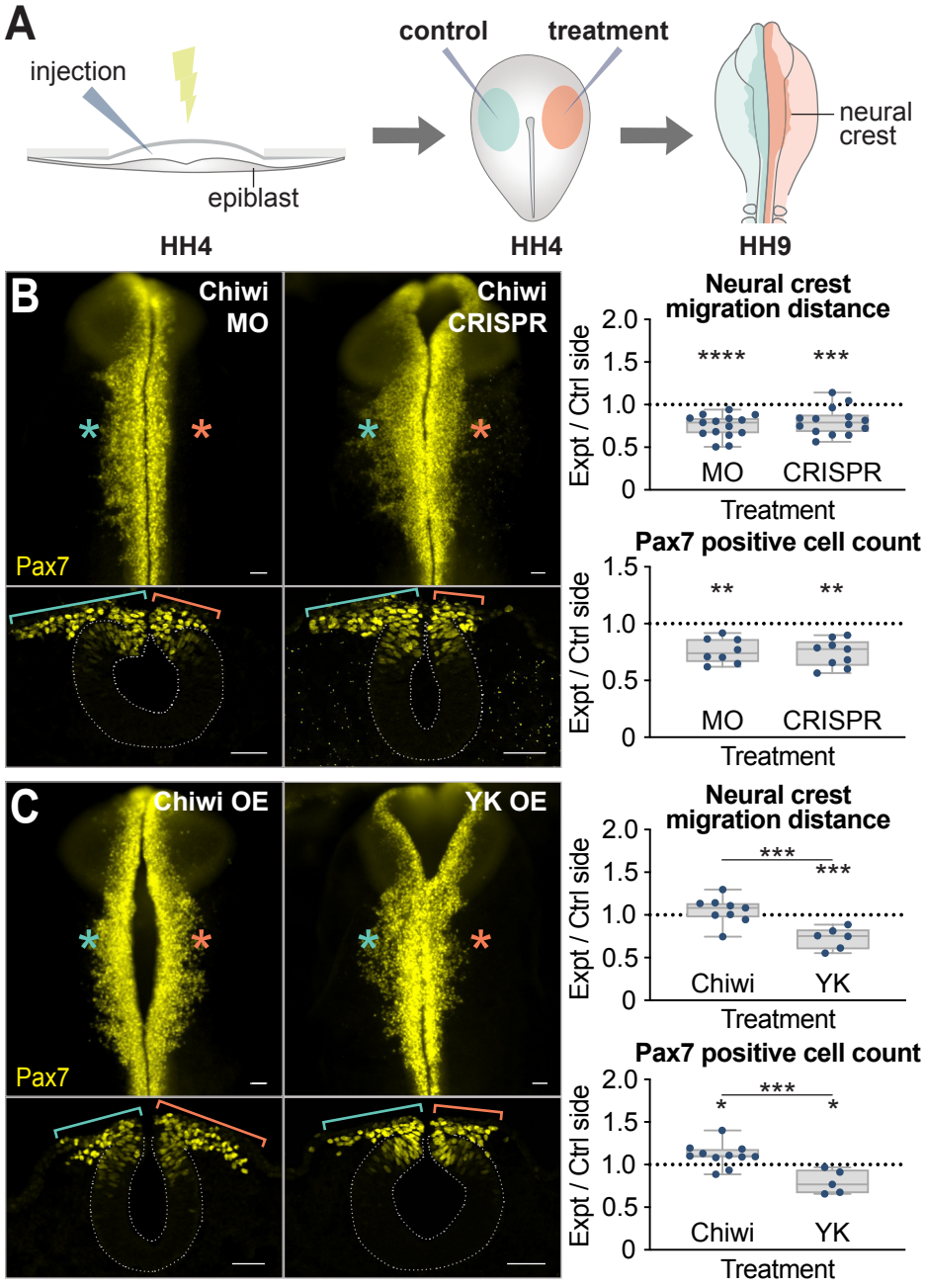

Figure 2. Perturbation of Chiwi disrupts neural crest development.

(A) Schematic diagram of the chick embryo electroporation strategy. (B) Loss of Chiwi impedes neural crest migration and reduces neural crest cell count, as measured by Pax7-expressing cells. Left: Examples of whole mount and cross sections upon MO and CRISPR knockout of Chiwi. Right:

Quantification of image analysis. Each data point represents measurements from a single embryo, with the right (experimental) divided by the left (control) side. (C) Overexpression (OE) of Chiwi increases the number of Pax7-positive cells, though migration distance is not significantly altered, while overexpression of the YK mutant, which is unable to bind piRNAs, impedes neural crest migration and reduces cell number. The blue stars and brackets denote the control side, while the red stars and brackets denote the experimental side. Scale bars, 50 μm. Box plots indicate the interquartile range, while whiskers extend to minimum and maximum values. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ****$P \leq 0.0001$. Asterisks (*) represent the difference between control and experimental measurements for each treatment.

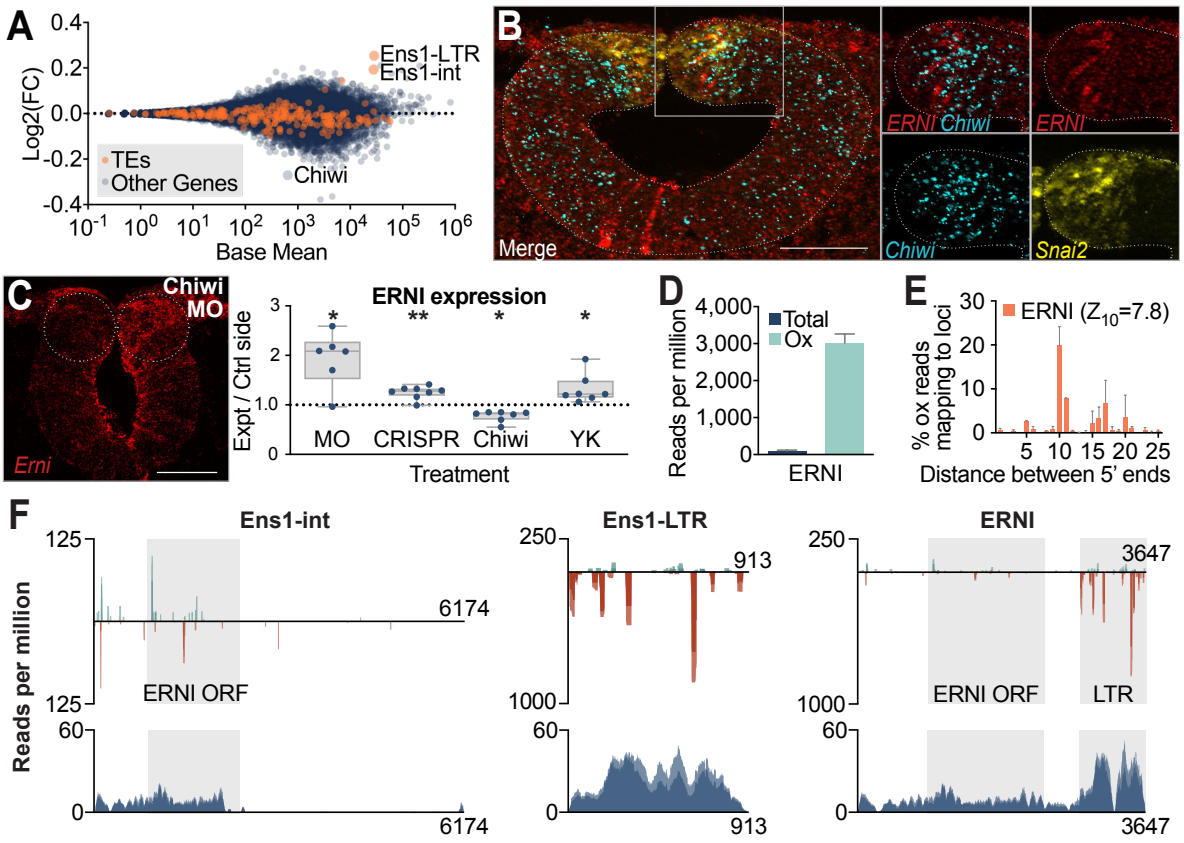

Figure 3. Chiwi regulates a single, transposon-derived gene, ERNI, in the neural crest.

(A) Differential expression analysis of RNA-seq from HH9− Chiwi CRISPR knockout versus control cranial neural folds. Dots represent genes (blue) and transposon families (orange). FC, fold change. (B) HCR depicting ERNI, Chiwi, and Snai2 expression in wild-type HH9 cranial midbrain region; scale bar, 50 μm. (C) HCR reveals changes to ERNI expression upon Chiwi loss (MO and CRISPR), as well as Chiwi and YK mutant overexpression. Each data point represents the average fluorescent intensity of the right dorsal fold region (experimental) divided by the left (control) side from three nonadjacent sections from the cranial region of a single embryo. Scale bar, 50 μm. Box plots indicate the interquartile range, while whiskers extend to minimum and maximum values. *P ≤ 0.05 and **P ≤ 0.01. Asterisks (*) represent the difference between control and experimental measurements for each treatment. (D) Normalized small RNA read counts mapping to the ERNI mRNA sequence in total versus oxidized (Ox) small RNA libraries. Error bars indicate SDs from

two biological replicates. (E) Analysis of 5′ to 5′ distance of complementary small RNA sequences mapping to the ERNI mRNA sequence in the oxidized small RNA libraries. (F) Small RNA-seq (top) and RNA-seq (bottom) tracks depicting sequences mapping to ENS-1 loci and the ERNI mRNA sequence (left). Oxidized small RNAs mapping in sense (teal) and antisense (red) orientation are depicted separately. Cranial neural fold total RNA-seq [control libraries from (A)] is depicted in blue. Replicate tracks are overlaid.

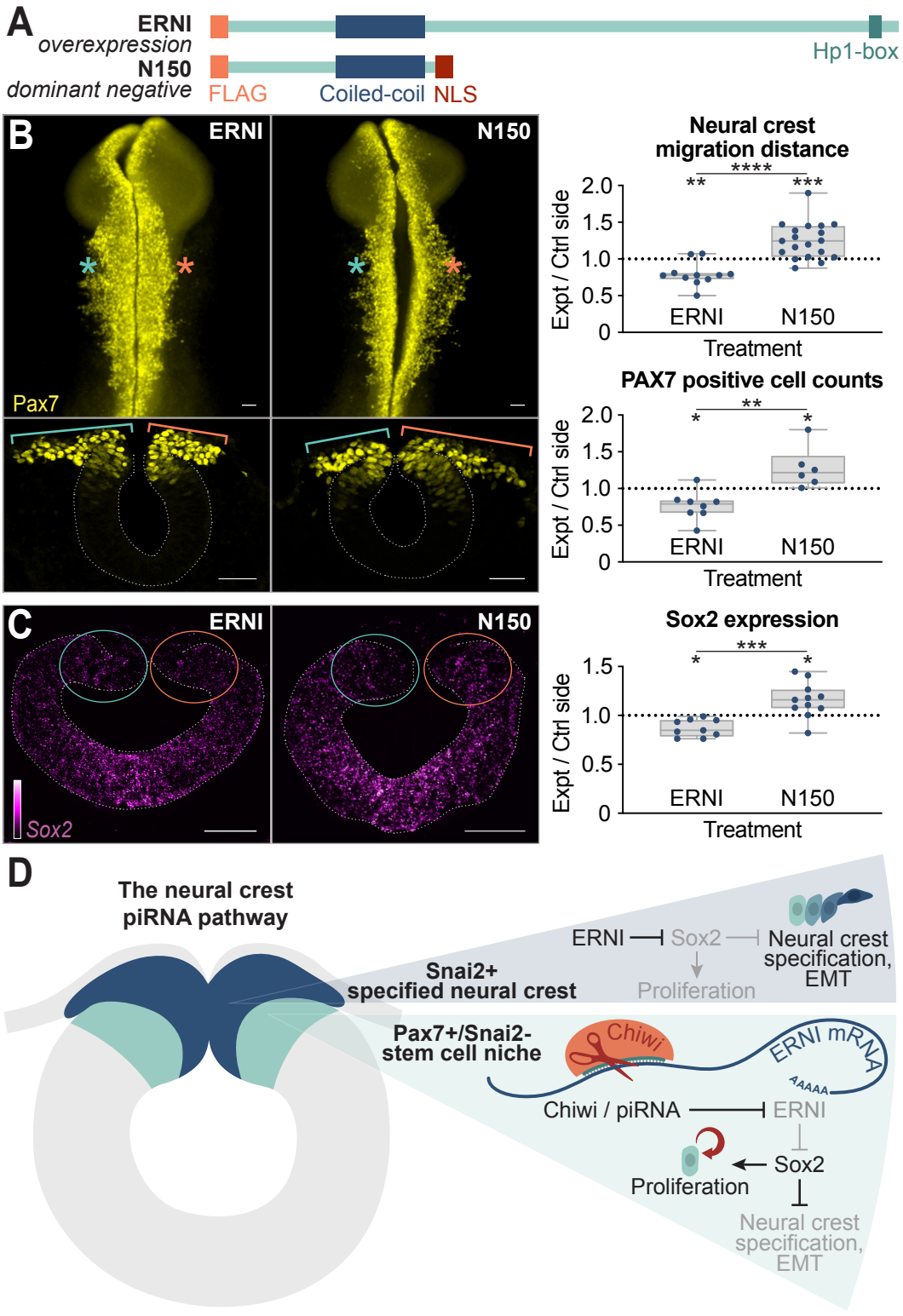

Figure 4. Perturbation of ERNI recapitulates Chiwi phenotypes.

(A) Schematic diagram of ERNI overexpression and dominant negative (N150) expression construct products. (B) Overexpression of ERNI recapitulates loss of Chiwi phenotype, with a reduction in neural crest migration distance and cell number, while overexpression of the N150 truncated ERNI sequence increases cell count and migration distance, as measured by Pax7-expressing cells. Left: Examples of whole mount and cross sections upon ERNI and N150 overexpression. Right: Quantification of image analysis. Each data point represents measurements from a single embryo, with the right (experimental) divided by the left (control) side. (C) HCR reveals that ERNI overexpression leads to a reduction in Sox2 expression in the dorsal neural tube, while N150 dominant negative overexpression increases Sox2 expression. Each data point represents the average fluorescent intensity of Sox2 signal in the right dorsal fold region (experimental) divided by the left (control) side from three nonadjacent sections from the cranial region of a single embryo. The teal stars and circle denote the control side, while the orange stars and circle denote the experimental side. Scale bars, 50 μm. Box plots indicate the interquartile range, while whiskers extend to minimum and maximum values. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ****$P \leq 0.0001$. Asterisks (*) represent the difference between control and experimental measurements for each treatment. (D) Schematic diagram of the neural crest piRNA pathway, depicting a neural tube with the PAX7+/Snai2− stem cell niche in teal, which feeds into the Snai2+-specified neural crest region in dark blue. Chiwi represses ERNI in the PAX7+/Snai2− stem cell niche via a piRNA-mediated mechanism to maintain Sox2 expression and proliferation. In specified neural crest, Chiwi expression is reduced, permitting ERNI expression, which, in turn, represses Sox2 to allow for neural crest specification and EMT.

Supplementary Materials

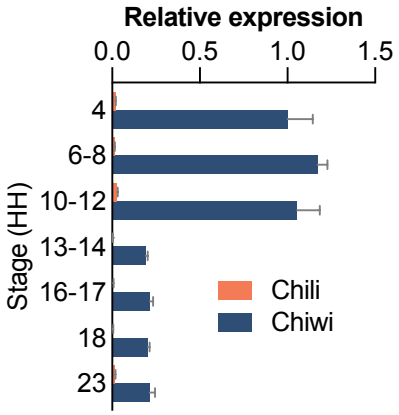

Figure S1.

RT-qPCR from whole embryo RNA across stages HH4-HH23 of chick development depicts relative expression of chick piwi transcripts Chiwi (Piwil1) and Chili (Piwil2) normalized to 18S rRNA. Error bars indicate st. dev., n=1 biological replicate (RNA extracted from 3 or more pooled embryos), 3 technical replicates.

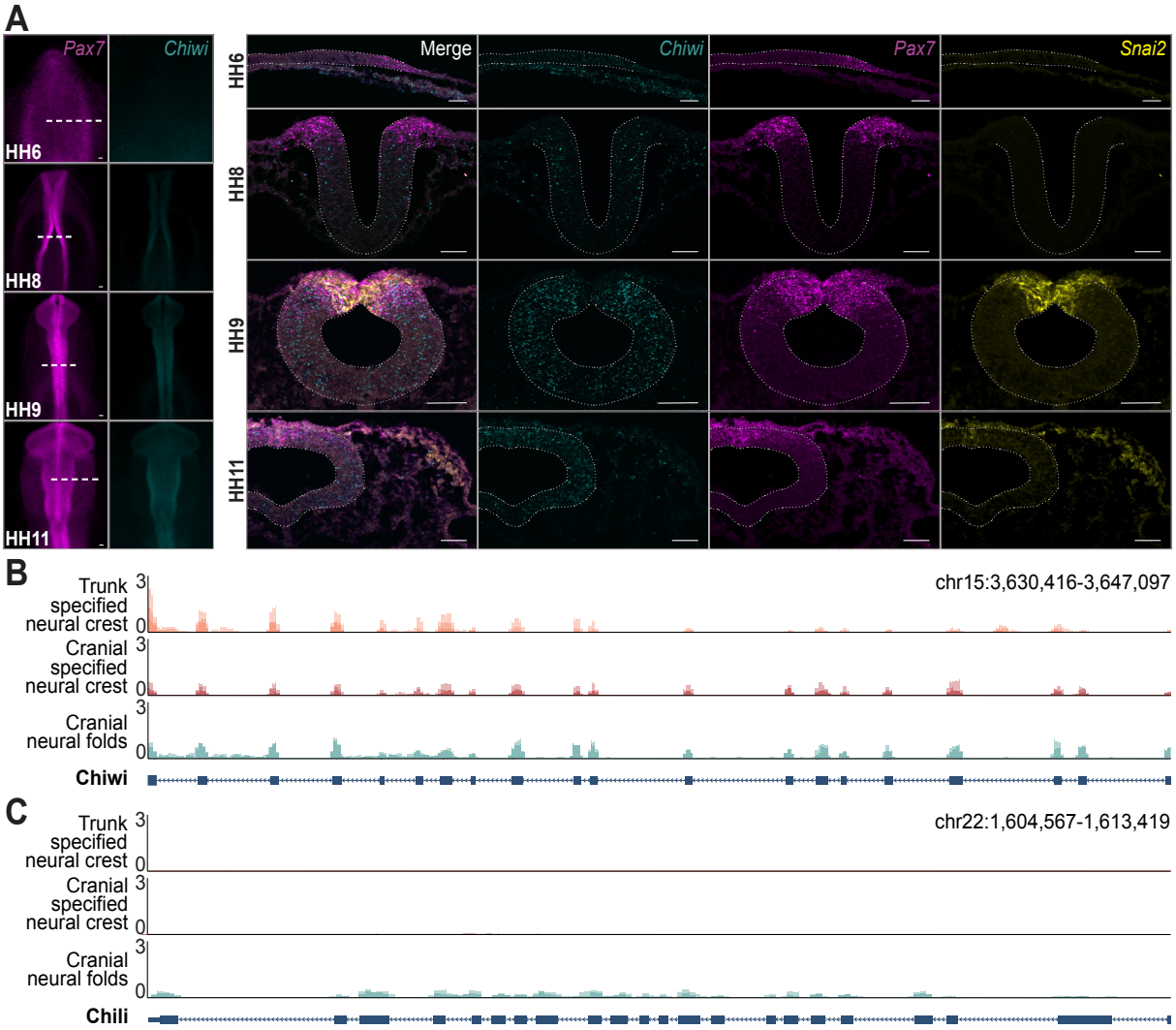

Figure S2.

A) HCR depicts expression of Chiwi and neural crest markers Pax7 and Snai2 across a range of stages from HH6-HH12. Scale bars = 50µm. B-C) RNA-seq tracks of Chiwi (ENSGALT00000004171.6) and Chili (XM_025142807.1) expression in dissected cranial neural folds (n=2 biological replicates, control libraries from Fig 3A) as well as specified, FACS-sorted trunk and cranial neural crest cells (n=3 biological replicates; previously published data). Y-axis is reads per million reads mapped to genome. Replicate tracks are overlaid.

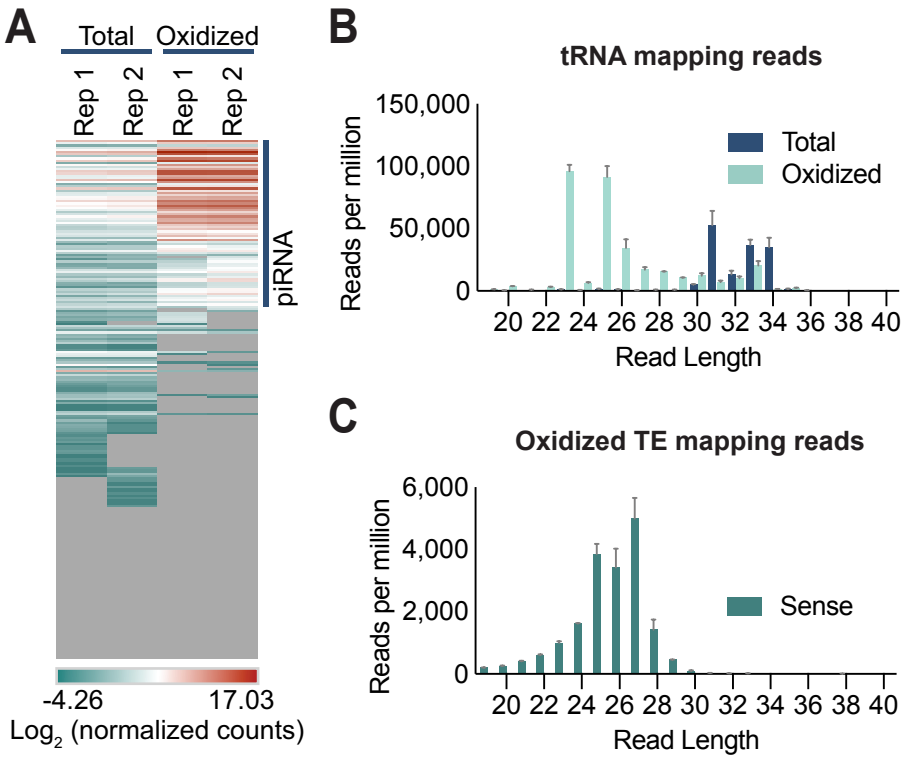

Figure S3.

A) Heatmap depicting normalized counts of small RNA sequences mapping to TEs. Each line represents a TE family. Counts are scaled by log2, after which the values of all cells range from -4.26 - 17.03. Cells with zero counts mapping are colored in gray (n=2 biological replicates). B) Length distribution of small RNA reads mapping to tRNAs (n=2 biological replicates). Error bars indicate st. dev. C) Length distribution of small RNA reads from oxidized libraries mapping to TEs in sense orientation (n=2 biological replicates). Error bars indicate st. dev.

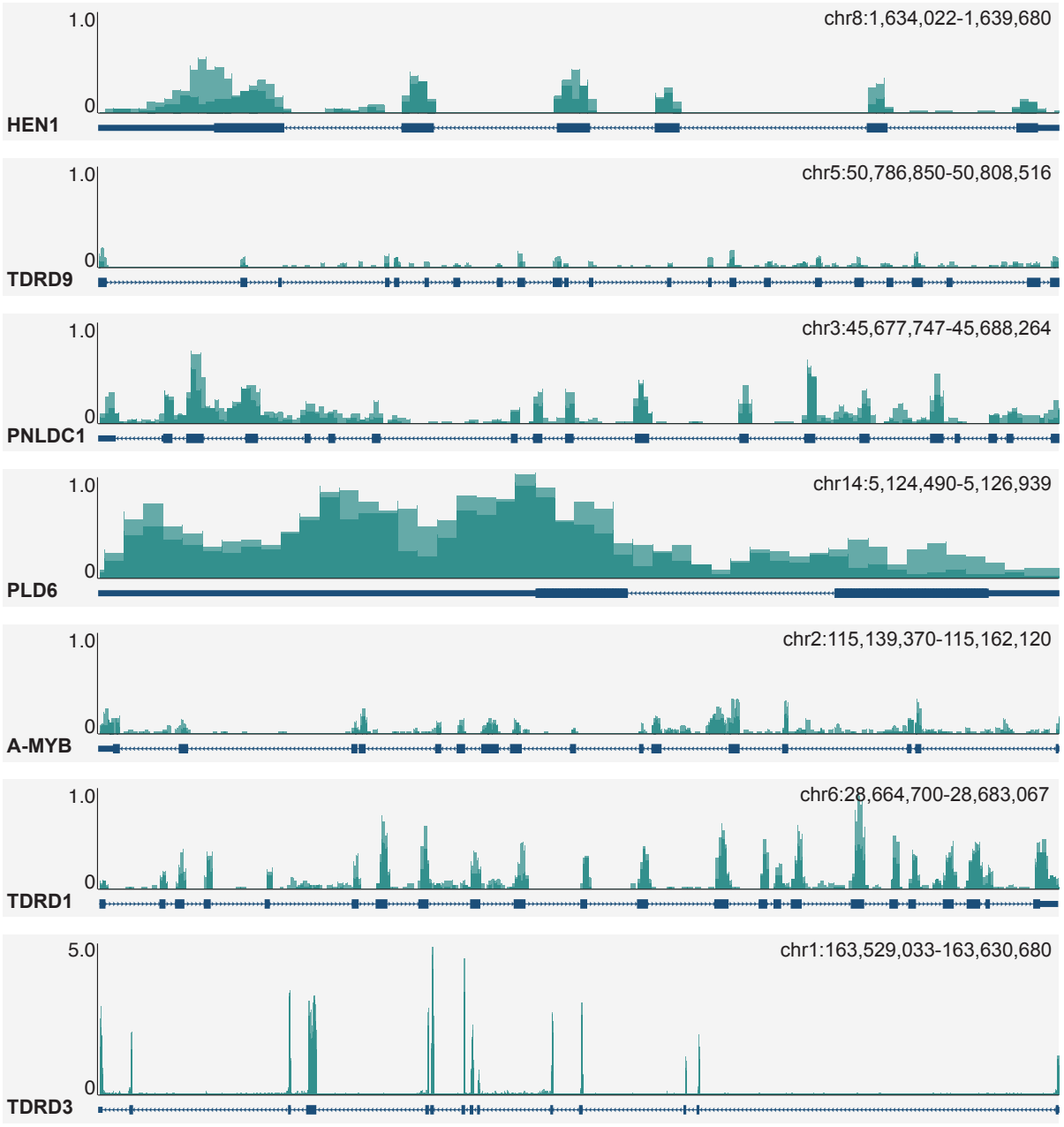
1.0
0
chr8:1,634,022-1,639,680
HEN1
1.0
0
chr5:50,786,850-50,808,516
TDRD9
1.0
0
chr3:45,677,747-45,688,264
PNLDC1
1.0
0
chr14:5,124,490-5,126,939
PLD6
1.0
0
chr2:115,139,370-115,162,120
A-MYB
1.0
0
chr6:28,664,700-28,683,067
TDRD1
5.0
0
chr1:163,529,033-163,630,680
TDRD3

Figure S4.

RNA-seq tracks of Hen1 (ENSGALT00000003030.5), TDRD9 (ENSGALT00000018863.7), PNLDC1 (ENSGALT00000018991.6), PLD6 (XM_015294407.2), A-MYB (ENSGALT00000066396.2), TDRD1 (ENSGALT00000092521.1), and TDRD3 (ENSGALT00000027376.6) expression in dissected cranial neural folds (n=2 biological replicates, control libraries from Fig 3A). Note that TDRD3 is shown on a larger scale than the other tracks. Y-axis is reads per million reads mapped to genome. Replicate tracks are overlaid.

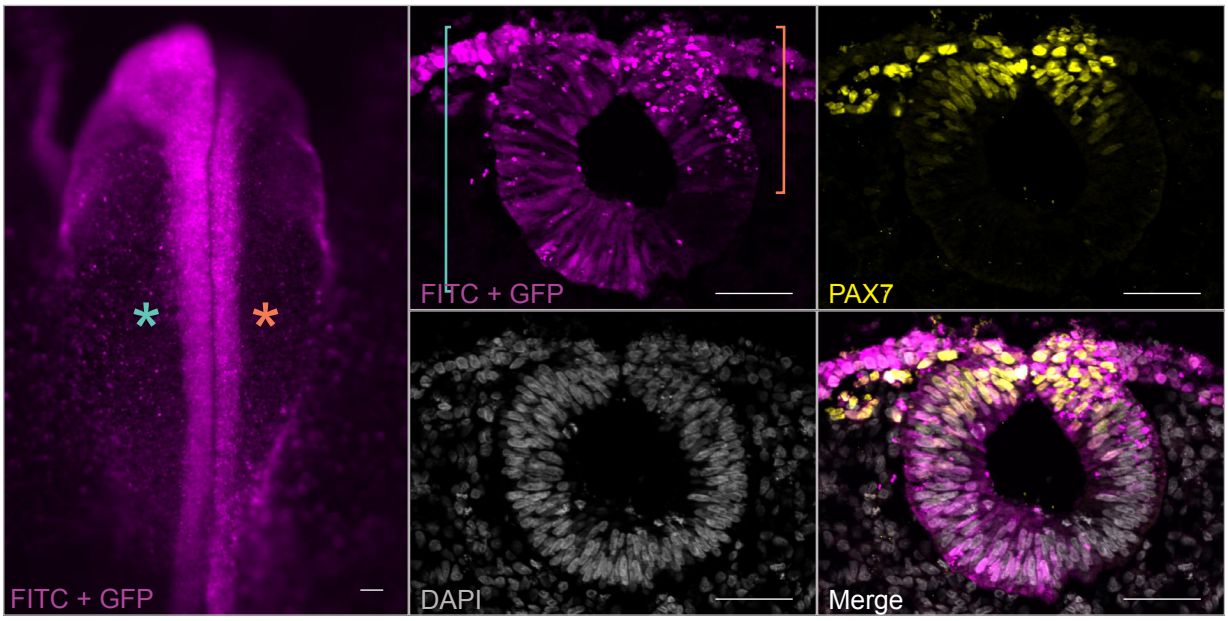

Figure S5.

A representative embryo electroporated with FITC labeled control MO + pCIG-GFP carrier plasmid on the left side, and FITC labeled Chiwi MO + pCIG-GFP on the right side. The left most panel depicts electroporation coverage in whole mount, while the right panels show it in cross-section in relation to the PAX7+ neural crest domain. The blue star and bracket denote the control side while the red star and bracket denote the Chiwi KD side. Scale bars = 50µm.

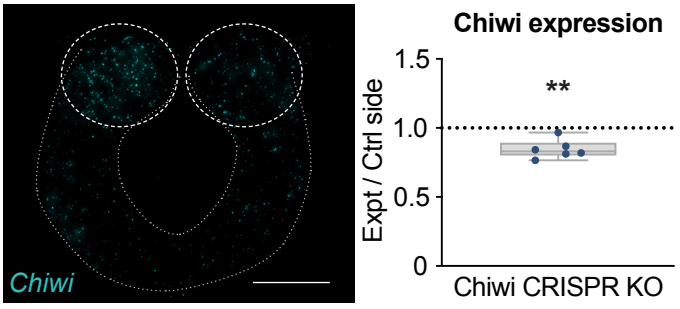

Figure S6.

HCR confirms a reduction in Chiwi expression upon electroporation of CRISPR/Cas9 plasmids targeting Chiwi. Left: example of cranial cross sections. Scale bar = 50µm. Right: quantification of image analysis. Each data point represents the average fluorescent intensity of the right dorsal fold region (knockout, KO) divided by the left (control) side from three non-adjacent sections from the cranial region of a single embryo. Box plots indicate the interquartile range, while whiskers extend to min and max values. *, **, *** and **** indicate p values of ≤ 0.05, 0.01, 0.001, and 0.0001, respectively, and represent the difference between control and experimental meausrements. The image is shown in conventional fluorescence for ease of visualization, while the quantifications were taken from maximum intensity projections of apotome slices.

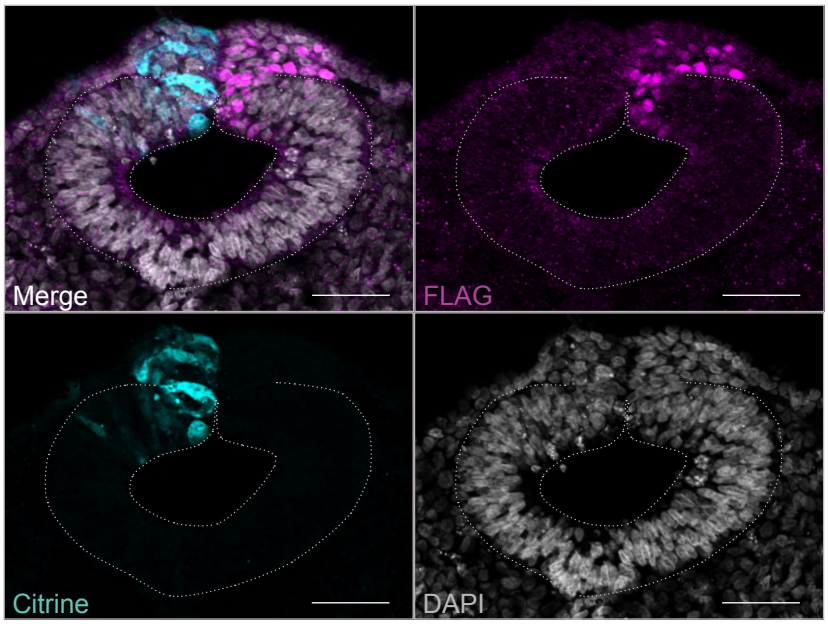

Figure S7.

Cross section of the cranial region from a representative embryo electroporated with enh195-citrine on the left and enh195-N150-FLAG-ERNI-NLS on the right. Immunostaining for FLAG confirms nuclear localization of the truncated N150 mutant. Scale bars = 50μm.

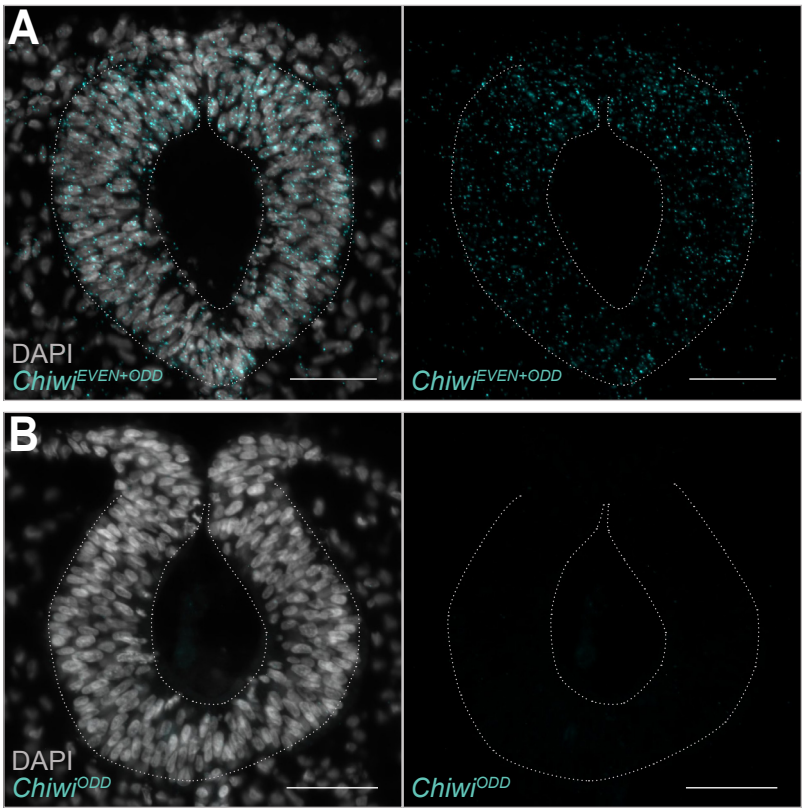

Figure S8.

HCR on HH9 embryos with both even and odd Chiwi probes depicts a distinct expression pattern in cross sections of the cranial neural tube (A), while HCR with only the odd probe imparts no signal (B), confirming that the low-level Chiwi expression picked up by HCR is not due to background or autofluorescence. Scale bars = 50µm.

Name	Sequence
Chiwi qPCR F	GCTTGCAAACTGCTCTAAGGG
Chiwi qPCR R	AGCAATCCAGAAACTGAGGCA
Chili qPCR F	TGCAGGACGTGACCAAGAAT
Chili qPCR R	GGATGGTTCGCACGTAGGTC3’
18s qPCR F	CCATGATTAAGAGGGACGGC
18s qPCR R	TGGCAAATGCTTTCGCTTT
mCMV Spe1 F	ATGCACTAGTGATCCTAGGAAGACGTCGATATACGCG
mCMV Xba1 R	TCAGTCTAGAGAATTCGAGAGCTCGGAAGCTTGG
Chiwi endogenous F	ATGACAGGAAGAGCTAGAGCCAGAGCGAGAGGGAGACCTCCAG
Chiwi alt nt sequence F	ATGACCGGTCGTGCACGTGCTCGTGCGAGAGGGAGACCTCCAG
Chiwi Xho1 F	ATGCCTCGAGATGACCGGTCGTGCACGT
Chiwi Asc1 R	TGAGGGCGCGCCTTAGAGATAGTAAAGTCTGTCTGAGA
YK mutation YI-KE-1	AATCCGTACATAGGTATTTCTCGATAGCATCAATCTTATCCTTTCGGGTACTAGAC
YK mutation YI-KE-2	GTCTAGTACCCGAAAGGATAAGATTGATGCTATCGAGAAATACCTATGTACGGATT
ERNI_FLAG_CDS_Nhe1_F	ATCGGCTAGCATGGACTACAAGGACGACGATGACA
P7ERNI_EcoR1_Fusion2_F	CGTCTTGTCATTGGCGAATTCGAACAC
V2_PAX7_ERNI_Fusion_R	TGTAGTCCATGCTAGCCGATGCTAGCTTGCGGCCGCTTA
ERNI_CDS_Bsrg1_R	TCAGTGTACAAAGGCATCAGCTCCCCTTGAG
erniN150NLS-Bsrg1_R	TCAGTGTACACTCATTATACCTTGCGCTTTTTCTTTGCACT
pCI sequencing primer F	TCAGGGTTATTGTCTCATGAGCGG
pCI sequencing primer R	GGGCTGCAGGAATTCGATATCAAGC
Chiwi mid sequencing primer 2	CCTGTGCAAAGGAGCTGATAGGTG
Chiwi mid sequencing primer 1	AGCACCTGAAGAGCCAGTTTT
Chiwi mid sequencing primer 3	ACCAAGATATCAGTGACTTGAACCAGCC

Chiwi mid sequencing primer 4	GGCTGGTTCAAGTCACTGATATCTTGGT
ERNI mid sequencing primer	CGTAGAGTTCTTTTCCTCTCGGAACTCC

Table S1.

Primers used in this study.

Ping-pong analysis (Perl, available at https://github.com/ninova/pingpong/blob/master/pair_overlap_selected_regions.pl)	#!/usr/bin/perl/
	use warnings;
	use strict;
	
	#usage perl script.pl mapped_reads.bam region.bed (bed file with regions of interest to be pre-selected)
	
	my $file=shift; #bam file with mapped reads
	my $region=shift; #file with given regions
	
	system "bedtools intersect -abam $file -b $region |samtools view - > $file.pis.tmp.ubam";
	
	open (FILE, "$file.pis.tmp.ubam");
	
	open (NEG, ">$file.neg1.tmpbed");
	open (POS, ">$file.pos1.tmpbed");
	
	my %positive=();
	my %negative=();
	
	while (<FILE>) {
	
	my (@tmp) = split (/\t/,$_);
	my $strand=$tmp[1];
	my $chrom=$tmp[2];
	my $seq=$tmp[9];
	my $start=$tmp[3];
	my $end=$start+length($seq);
	if ($strand==0) {if ((length($seq)>18) and (length($seq)<35))
	{$positive{"$chrom\t$start\t$end\t$seq"}+=1}}
	
	elsif ($strand==16) {if ((length($seq)>18) and (length($seq)<35))
	{$negative{"$chrom\t$start\t$end\t$seq"}+=1}}
	
	}
	
	foreach (sort (keys %positive)) {print POS "$_\;$positive{$_}\n" }
	foreach (sort (keys %negative)) {print NEG "$_\;$negative{$_}\n" }
	close FILE;

```
close POS;
close NEG;

system "bedtools intersect -a $file.pos1.tmpbed -b
$file.neg1.tmpbed -wo > $file.inters1.tmptab";
system "bedtools intersect -a $file.neg1.tmpbed -b
$file.pos1.tmpbed -wo > $file.inters2.tmptab";

my %overlaps=();
my %counts=();
my %totals=();
my %counts10=();
open (TAB, "$file.inters1.tmptab");

while (<TAB>) {
   chomp($_);
   my (@tmp) = split (/\t/,$_);
   my $name1=join(":",$tmp[0],$tmp[1],$tmp[2],$tmp[3]);
   my $name2=join(":",$tmp[4],$tmp[5],$tmp[6],$tmp[7]);
   my $start1=$tmp[1];
   my $start2=$tmp[5];
   my $overlap=$tmp[8];
   if ($start1>$start2) {
  my ($else1,$count1)=split(/\;/,$name1,2);
  my ($else2,$count2)=split(/\;/,$name2,2);

  push(@{$overlaps{$name1}}, $overlap);
  push(@{$counts{$name1}}, $count2);
  $totals{$name1}+=$count2;
  if ($overlap==10) {$counts10{$name1}=1;
$counts10{$name2}=1};
   }
}
close TAB;
open (TAB, "$file.inters2.tmptab");

while (<TAB>) {
   chomp($_);
   my (@tmp) = split (/\t/,$_);
   my $name1=join(":",$tmp[0],$tmp[1],$tmp[2],$tmp[3]);
   my $name2=join(":",$tmp[4],$tmp[5],$tmp[6],$tmp[7]);
   my $start1=$tmp[1];
   my $start2=$tmp[5];
   my $overlap=$tmp[8];
```

	if ($start1<$start2){
	my ($else1,$count1)=split(/\;/,$name1,2);
	my ($else2,$count2)=split(/\;/,$name2,2);
	
	push(@{$overlaps{$name1}}, $overlap);
	push(@{$counts{$name1}}, $count2);
	$totals{$name1}+=$count2;
	}
	}
	
	my %freqs=();
	
	foreach (sort (keys %totals))
	{#print "$_\t$totals{$_}\t";
	my ($else1,$count1)=split(/\;/,$_,2);
	if (@{$counts{$_}} != @{$overlaps{$_}}) {print "something went wrong" and die}
	for (my $i=0;$i<@{$counts{$_}} ;$i++) {
	
	$freqs{$overlaps{$_}[$i]}+=$counts{$_}[$i]/$totals{$_}*$count1;
	}
	# foreach my $m(@{$counts{$_}}) {print "$m:";}
	# print "\t";
	# foreach my $n(@{$overlaps{$_}}) {print "$n:"; }
	# print "\t";
	# print "\n";
	}
	
	foreach (sort {$a<=>$b} (keys %freqs)) {print "$_\t$freqs{$_}\n"}
	
	my $tot10=0;
	foreach (keys %counts10) {my ($name,$n) = split (/\;/,$_) ; $tot10+=$n}
	print STDERR $tot10,"\n";
	
	unlink ("$file.pis.tmp.ubam");
	unlink ("$file.pos1.tmpbed");
	unlink ("$file.neg1.tmpbed");
	unlink ("$file.inters1.tmptab");
	unlink ("$file.inters2.tmptab");

Cell counting (IJ1 Macro, available at	name=File.nameWithoutExtension; rename("A");

https://github.com/mpiacentino/fiji_macros/blob/4d20899ece32056769422d9646c728ac944c5c59/CellCount/CellCount_2ROI_1Channel.ijm)	setTool("freehand");
	
	if (isOpen("Results")) {
	selectWindow("Results");
	run("Close");
	}
	if (isOpen("Summary")) {
	selectWindow("Summary");
	run("Close");
	}
	if (isOpen("ROI Manager")) {
	selectWindow("ROI Manager");
	run("Close");
	}
	
	run("Rotate... ", "angle=180 grid=1 interpolation=Bilinear stack");
	
	waitForUser("Draw ROI 0 (Control), then press ok");
	roiManager("Add");
	roiManager("Select",0);
	roiManager("Rename","Cntl");
	roiManager("Show All");
	waitForUser("Draw ROI 1 (Experimental), then press ok");
	roiManager("Add");
	roiManager("Select",1);
	roiManager("Rename","Expt");
	
	resetMinAndMax();
	run("Split Channels");
	
	selectWindow("C2-A");
	run("Median...", "radius=3 slice");
	resetMinAndMax();
	run("8-bit");
	rename("RAW");
	run("Auto Local Threshold", "method=Bernsen radius=20 parameter_1=20 parameter_2=0 white");
	roiManager("Show All");
	run("Analyze Particles...", "size=5.00-Infinity show=Masks");
	rename("CntlSide");
	run("Duplicate...", "title=ExptSide");
	
	selectWindow("CntlSide");
	roiManager("Show All");
	roiManager("Select", 0);

	run("Analyze Particles...", "size=2-Infinity show=Nothing summarize");
	selectWindow("ExptSide");
	roiManager("Show All");
	roiManager("Select", 1);
	run("Analyze Particles...", "size=2-Infinity show=Nothing summarize");
	
	dir = getDirectory("Choose a directory to save ROI sets.");
	roiManager("Save", dir+name+".zip");
	
	selectWindow("CntlSide");
	close();
	selectWindow("ExptSide");
	close();
	//selectWindow("C2-A");
	//close();
	selectWindow("C1-A");
	close();
	selectWindow("RAW");
	close();
	selectWindow("C3-A");
	close();
	selectWindow("C4-A");
	close();

Fluorescent intensity, area (IJ1 Macro)	name=File.nameWithoutExtension;
	rename("A");
	run("Rotate... ", "angle=180 grid=1 interpolation=Bilinear stack");
	setTool("freehand");
	if (isOpen("Results")) {
	selectWindow("Results");
	run("Close");
	}
	if (isOpen("Summary")) {
	selectWindow("Summary");
	run("Close");
	}
	if (isOpen("ROI Manager")) {
	selectWindow("ROI Manager");
	run("Close");
	}
	waitForUser("Draw ROI 0 (Background), then press ok");
	roiManager("Add");
	roiManager("Select",0);

```
roiManager("Rename","background");
roiManager("Show All");
waitForUser("Draw ROI 1 (Background), then press ok");
roiManager("Add");
roiManager("Select",1);
roiManager("Rename","background");
waitForUser("Draw ROI 2 (Background), then press ok");
roiManager("Add");
roiManager("Select",2);
roiManager("Rename","background");

waitForUser("Draw ROI 3 (Experimental ROI), then press
ok");
roiManager("Add");
roiManager("Select",3);
roiManager("Rename","Expt");
waitForUser("Draw ROI 4 (Control ROI), then press ok");
roiManager("Add");
roiManager("Select",4);
roiManager("Rename","Cntl");
run("Split Channels");
selectWindow("C1-A");
rename("1");
resetMinAndMax();
run("Median...", "radius=3");
roiManager("Show All");
roiManager("Select", 0);
run("Measure");
roiManager("Select", 1);
run("Measure");
roiManager("Select", 2);
run("Measure");
roiManager("Select", 4);
run("Measure");
roiManager("Select", 3);
run("Measure");
selectWindow("1");
close();
selectWindow("C3-A");
rename("2");
resetMinAndMax();
run("Median...", "radius=3");
roiManager("Show All");
roiManager("Select", 0);
```

	run("Measure");
	roiManager("Select", 1);
	run("Measure");
	roiManager("Select", 2);
	run("Measure");
	roiManager("Select", 4);
	run("Measure");
	roiManager("Select", 3);
	run("Measure");
	selectWindow("2");
	close();
	selectWindow("C4-A");
	rename("3");
	resetMinAndMax();
	run("Median...", "radius=3");
	roiManager("Show All");
	roiManager("Select", 0);
	run("Measure");
	roiManager("Select", 1);
	run("Measure");
	roiManager("Select", 2);
	run("Measure");
	roiManager("Select", 4);
	run("Measure");
	roiManager("Select", 3);
	run("Measure");
	selectWindow("3");
	close();
	selectWindow("C2-A");
	close();
	waitForUser("Choose a directory to save ROIs, then press ok");
	dir = getDirectory("Choose a directory to save ROI sets.");
	roiManager("Save", dir+name+".zip");
	saveAs("Results", dir+name+".csv");
	selectWindow("ROI Manager");
	run("Close");

Fluorescent intensity, plot profile (IJ1 Macro)	name=File.nameWithoutExtension;
	rename("A");
	run("Rotate... ", "angle=180 grid=1 interpolation=Bilinear stack");
	setTool("polyline");
	if (isOpen("Results")) {
	selectWindow("Results");

```
        run("Close");
    }
if (isOpen("Summary")) {
        selectWindow("Summary");
        run("Close");
    }
if (isOpen("ROI Manager")) {
        selectWindow("ROI Manager");
        run("Close");
    }
waitForUser("Draw ROI 0 (Right), then press ok");
roiManager("Add");
roiManager("Select",0);
roiManager("Rename","Right");
roiManager("Show All");
waitForUser("Draw ROI 1 (Left), then press ok");
roiManager("Add");
roiManager("Select",1);
roiManager("Rename","Left");
run("Subtract Background...", "rolling=50 stack");
run("Split Channels");
selectWindow("C1-A");
rename("Chiwi");
resetMinAndMax();
roiManager("Show All");
roiManager("Select", 0);
run("Plot Profile");
Plot.getValues(x, y);
str = "";
for ( i = 0; i < x.length; i++ ) { str +=  "" + x[i] + "\t" + y[i] +
"\n"; }
dir = getDirectory("Choose a directory to save plot data");
File.saveString( str, dir+name+"_right_chiwi.txt" );
selectWindow("Plot of Chiwi");
close();
selectWindow("Chiwi");
roiManager("Select", 1);
run("Plot Profile");
Plot.getValues(x, y);
str = "";
for ( i = 0; i < x.length; i++ ) { str +=  "" + x[i] + "\t" + y[i] +
"\n"; }
File.saveString( str, dir+name+"_left_chiwi.txt" );
selectWindow("Chiwi");
```

```
close();
selectWindow("Plot of Chiwi");
close();
selectWindow("C2-A");
rename("Pax7");
resetMinAndMax();
roiManager("Show All");
roiManager("Select", 0);
run("Plot Profile");
Plot.getValues(x, y);
str = "";
for ( i = 0; i < x.length; i++ ) { str +=  "" + x[i] + "\t" + y[i] +
"\n"; }
File.saveString( str, dir+name+"_right_PAX7.txt" );
selectWindow("Plot of Pax7");
close();

selectWindow("Pax7");
roiManager("Select", 1);
run("Plot Profile");
Plot.getValues(x, y);
str = "";
for ( i = 0; i < x.length; i++ ) { str +=  "" + x[i] + "\t" + y[i] +
"\n"; }
File.saveString( str, dir+name+"_left_PAX7.txt" );
selectWindow("Pax7");
close();
selectWindow("Plot of Pax7");
close();
selectWindow("C3-A");
rename("Snail");
resetMinAndMax();
roiManager("Show All");
roiManager("Select", 0);
run("Plot Profile");
Plot.getValues(x, y);
str = "";
for ( i = 0; i < x.length; i++ ) { str +=  "" + x[i] + "\t" + y[i] +
"\n"; }
File.saveString( str, dir+name+"_right_Snail.txt" );
selectWindow("Plot of Snail");
close();
selectWindow("Snail");
roiManager("Select", 1);
```

	run("Plot Profile");
	Plot.getValues(x, y);
	str = "";
	for (i = 0; i < x.length; i++) { str += "" + x[i] + "\t" + y[i] + "\n"; }
	File.saveString(str, dir+name+"_left_Snail.txt");
	selectWindow("Snail");
	close();
	selectWindow("Plot of Snail");
	close();
	selectWindow("C4-A");
	close();
	selectWindow("C2-A");
	close();
	selectWindow("ROI Manager");
	run("Close");

Plot profile binning script (Python)	import os import sys import numpy as np import pandas as pd from pathlib import Path import fileinput bin_num=0.5 i=0 avg=0 mysum=0 grey = 0 mylist = [] maxscale1=0 maxscale2=0 with open(sys.argv[1], 'r') as my_file: df = pd.read_csv(my_file, sep='\t', header=None) maxscale1 = df[0].max() / 100 maxscale2 = df[1].max() df[0] = df[0]/maxscale1 df[1] = df[1]/maxscale2 df.to_csv('temp.txt', sep='\t', header=None, index=False) with open('temp.txt', 'r') as f: for line in f:

	col = line.split('\t')
	if float(col[0])<=bin_num and float(col[0])!=100:
	mylist.append(float(col[1]))
	i+=1
	
	elif i!=0:
	avg = sum(mylist) / i
	
	print(str(bin_num) + '\t' + str(avg))
	mylist = []
	i=0
	bin_num+=0.5
	elif float(col[0])!=100:
	
	bin_num+=0.5
	else:
	print(str(bin_num) + '\t' + str(avg))
	f.close()

Table S2.

Scripts used in this study.

Chapter 3

Piwil1 expression in the neural crest is conserved across vertebrates

Riley Galton, Marianne Bronner*, and Kata Fejes-Toth*

Results

In Chapter 2, we demonstrated that the piRNA pathway is active in the chick neural crest, and serves to modulate neural crest specification by spatiotemporally regulating the transposon derived gene ERNI. ERNI is derived from a Galliformes specific transposon sequence, raising the question of whether piRNA pathway activity in the neural crest is unique to chickens and closely related birds, or conserved in other organisms. Intriguingly, another transposon derived gene, Crestin, is highly expressed in zebrafish neural crest cells, though its function is unclear (*1*). Crestin is of similar length and structure to ERNI, containing an N-terminal coiled-coil domain, though no HP1-box sequence (Fig. 1A) (*2*) (Cedric Feschotte, personal communication). To examine whether there is a reciprocal expression pattern of Piwi and Crestin, we performed HCR in zebrafish embryos with probes for *Piwil1* (Ziwi), *Crestin*, and chick *Snai2* homolog *Snai1b* at the onset of cranial neural crest migration at 13 hour post fertilization (hpf), which is comparable to HH9 in chick. Our results indeed reveal a similar pattern of expression as in chicken, with *Ziwi* peaking in the dorsal half of the neural keel, then dropping in the *Snai1b*+/*Crestin*+ specified neural crest region (Fig. 1B).

To further explore whether the neural crest piRNA pathway might be conserved across vertebrates, we similarly examined the expression pattern of *Snai1*, the Snail homolog expressed during neural crest EMT in mouse(*3*), and *Piwil1* (Miwi) in the cranial region of mouse embryos at E8.5, corresponding to HH9-10 in chick. As in chick, we observed *Miwi* expression in the neural tube with expression peaking just lateral to the *Snai1*+ specified neural crest domain, then dropping in the *Snai1*+ domain itself (SupFig. 5C). These results raise the intriguing possibility that the piRNA

pathway is a conserved facet of the vertebrate neural crest gene regulatory network while acting on different transposon-derived genes.

Acknowledgements

We thank David Mayorga for help acquiring fish embryos, and Magdalena Zernicka-Goetz and Estefania Sanchez Vasquez for providing mouse embryos.

References

1. A. L. Rubinstein, D. Lee, R. Luo, P. D. Henion, M. E. Halpern, Genes dependent on zebrafish cyclops function identified by AFLP differential gene expression screen. *Genesis*. **26**, 86–97 (2000).

2. A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences. *Science*. **252**, 1162–1164 (1991).

3. M. Sefton, S. Sánchez, M. A. Nieto, Conserved and divergent roles for members of the Snail family of transcription factors in the chick and mouse embryo. *Development*. **125**, 3111–3121 (1998).

Figures

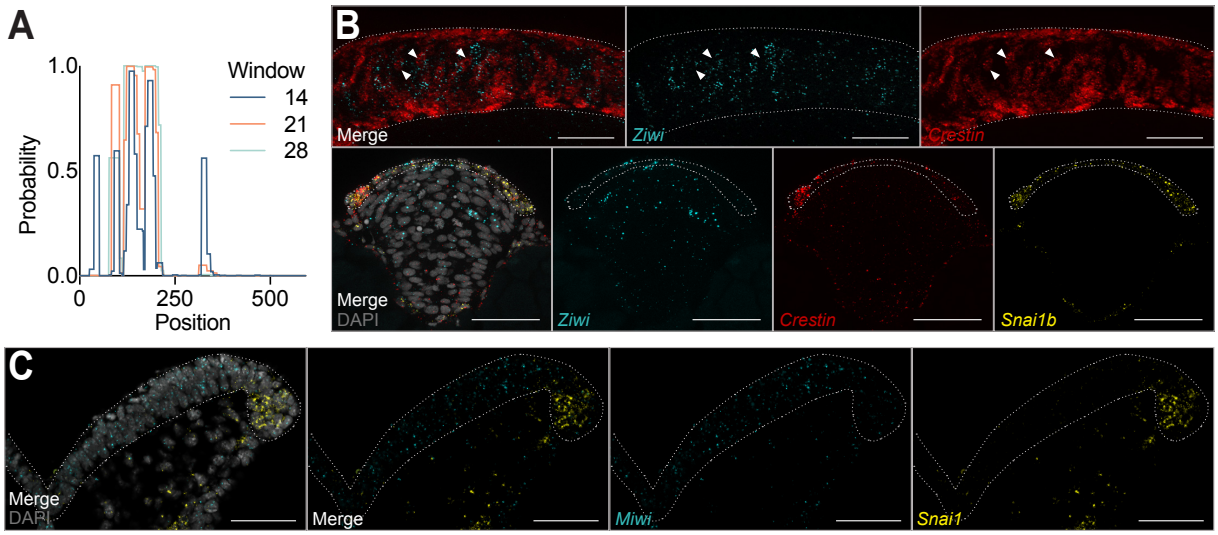

Figure 1: Piwil1 expression in the dorsal neural tube is conserved in zebrafish and mouse.

(A) Predicted coiled-coil structure of the Crestin protein sequence over three window sizes. (B) HCR for *Crestin*, *Snai1b* and *Ziwi* in zebrafish cranial neural keel at 13hpf. The top panel is a dorsal view with arrows pointing to *Crestin*+/*Ziwi*- cells and the bottom is a section with the dotted line outlining the *Snai1b*+ neural crest region. (C) HCR for Snai1 and Miwi in the mouse cranial neural tube at E8.5, with dotted line outlining the neural tube. All scale bars = 50µm.

Chapter 4

Myb genes act as a molecular switch to regulate piRNA pathway control of neural crest specification

Riley Galton, Marianne Bronner*, and Kata Fejes-Toth*

Abstract

The canonical function of the highly conserved piRNA pathway is to repress transposon activity in the metazoan germline. Recently, we reported a novel function of the piRNA pathway in the chick neural crest, where it has been co-opted into the gene regulatory network to mediate neural crest specification by spatiotemporally regulating the transposon-derived gene ERNI. In investigating the regulation of the piRNA pathway in the neural crest, here we show preliminary evidence that it is regulated by two conserved vertebrate Myb genes, A-Myb and C-Myb. We find that A-Myb is expressed at low levels in the neural tube, and depletion of A-Myb results in a reduction of Chiwi expression. C-Myb, on the other hand, is transiently upregulated at the onset of neural crest specification, and depletion of C-Myb results in a slight increase in Chiwi expression in the neural crest. We find that the Chiwi promoter sequence contains a Myb binding site, which, when mutated, results in a reduction of Chiwi expression. Analysis of the effects of Myb gene perturbation on a reporter construct under control of the Chiwi promoter serves to further validate the relationship between A-Myb and C-Myb in Chiwi regulation. Together, our data provide preliminary evidence supporting the hypothesis that Myb proteins antagonistically regulate piRNA pathway activity in the neural crest, which allows for the spatiotemporal regulation of ERNI and precise timing of neural crest specification.

Introduction

In developing vertebrate embryos, the formation and subsequent migration of neural crest cells is a highly regulated and complex process, the outcome of which is critical to many structures and organs in the vertebrate body (*1–3*). This unique population of stem-like cells is induced at the neural plate

border during gastrulation. Upon neurulation, neural crest cells carry out a gene regulatory program known as specification, which subsequently allows them to undergo an epithelial-to-mesenchymal transition (EMT) and migrate extensively throughout the embryo. Eventually, neural crest cells differentiate into a wide array of critical cell types and structures, including the craniofacial skeleton and much of the peripheral nervous system (*3*, *4*). As such, precise regulation of neural crest development is critical to the formation of the vertebrate embryo, and a complex gene regulatory network (GRN) has evolved for this purpose (*2–4*).

Recently we showed that, in the chicken embryo, the piRNA pathway has been co-opted by the neural crest GRN to spatiotemporally regulate a transposon-derived gene, ERNI (*5*). The piRNA pathway is a highly conserved small RNA pathway that canonically functions to silence transposable elements in the animal germline to maintain genomic integrity for the next generation (*6–8*). During neural crest development, however, the piRNA pathway, via chicken Piwil1 (Chiwi) and its associated piRNAs, silences ERNI transcripts at the neural plate border to maintain Sox2 levels in the neural plate border stem cell niche prior to neural crest specification. Upon specification, Chiwi expression rapidly drops, permitting ERNI expression which in turn represses Sox2, allowing for completion of the neural crest specification program, which is inhibited by high levels of Sox2 (*5*, *9*, *10*).

Many questions remain about how the piRNA pathway and one of its transposon targets were co-opted to perform such a critical function in the highly conserved developmental process of neural crest specification. The piRNA pathway, even in its most basic form, relies on the expression and coordination of many pathway-specific genes (*11–15*). Antisense piRNA precursors are transcribed from discrete genomic regions called piRNA clusters, processed by pathway-specific machinery into mature piRNAs, and then loaded onto pathway-specific Piwi clade argonaute proteins, which identify and cleave target transcripts based on complementarity to their bound piRNA (*11*, *13*).

While it is not fully understood how the piRNA pathway is regulated in the germline, the transcription factor A-Myb (MybL1) has been implicated as a master regulator of the pathway, driving expression of both piRNA cluster transcripts and key piRNA pathway proteins, including Piwil1, during the pachytene stage of meiosis in both mouse and rooster testes (*16*). We previously

reported that A-Myb is expressed in the dorsal neural folds of the developing chick embryo, raising the possibility that it might play a similar role in the neural crest to that of the germline (*5*).

A-Myb is a member of the Myb family of proto-oncogenes, the most well-studied of which are A-, B- and C-Myb (*17*). While A-Myb has not been previously implicated in neural crest development, C-Myb has been shown to be a regulator of neural crest specification (*18–20*). Members of the Myb gene family contain highly homologous N-terminal DNA-binding domains and are thought to target the same DNA binding sites (*17*, *21*). The rest of their structures are less conserved, suggesting that they may interact with different transcriptional partners. All Mybs, however, contain an internal transactivation domain, as well as a C-terminal regulatory domain which can negatively regulate their transcriptional transactivation function (*17*). Due to their varying transactivation abilities in different systems, it is thought that the function of Myb proteins relies heavily on cell-type specific cofactors, and in some contexts Myb proteins appear to have an antagonistic relationship, competing for binding sites and reciprocally regulating the same genes (*17*, *22*, *23*).

During neural crest development, C-Myb is expressed at low levels in the neural plate border and appears to be upregulated in specified neural crest cells (*18*). Prior to neurulation, C-Myb maintains Pax7 expression at the neural plate border through direct interaction with a Pax7 enhancer sequence, though overexpression of C-Myb represses Pax7, lending credence to the hypothesis that cofactors are critical to the transactivation potential of C-Myb (*19*). Upon neural crest specification, loss of C-Myb results in a reduction in expression of the neural crest specifier genes Ets1 and Snai2, as well as an increase in neural plate border gene Zic1 (*18*). Upon delamination of the neural crest, C-Myb also plays an important role in directly activating Sox10 via a binding site in the Sox10E2 enhancer (*20*).

Here, we provide preliminary data suggesting that A-Myb and C-Myb may reciprocally regulate piRNA pathway activity in the developing neural crest. We find that A-Myb is expressed at low levels throughout the neural tube and functions to promote Chiwi expression. C-Myb, on the other hand, is transiently upregulated at the onset of neural crest specification and appears to repress Chiwi, thus permitting ERNI expression and its subsequent repression of Sox2 to allow for the completion

of the neural crest specification program. Thus, we hypothesize that a Myb "switch" is at play, and performs a critical role in neural crest specification via regulation of the Piwi pathway.

Results

A reporter driven by the Chiwi promoter sequence recapitulates Chiwi's expression pattern in the neural crest

Chiwi is expressed at low levels in the developing neural tube, with its expression peaking in the Pax7+/Snai2- neural plate border region, then dropping again in the Snai2+ specified neural crest domain (*5*) (Fig. 1A). This distinct expression pattern is critical to the piRNA pathway's function in the neural crest, where it maintains the neural plate border stem-cell niche and spatiotemporally regulates the transposon-derived gene ERNI to time neural crest specification (*5*).

Upon analysis of our previously published RNA-seq data from cranial neural folds at HH9 (*5*), we noted that Chiwi transcription appears to initiate ~600 base pairs (bp) upstream of the translation start site, in the middle of a 463bp long CpG island which we hypothesized functioned as the Chiwi promoter (Fig. 1B). To see if this region functioned as a promoter, as well as how it is regulated in vivo, we cloned a segment of genomic DNA 1131 bp upstream of the Chiwi start codon, which includes the CpG island in the middle, into a promoter-less vector and upstream of a GFP coding sequence. We named this genomic region chP, for Chiwi promoter, and the plasmid chP-GFP (Fig. 1B-C).

To determine whether this sequence was capable of driving GFP expression in the chicken embryo, we electroporated chP-GFP into the neural plate border region of one side of the embryo at HH4, and allowed the electroporated embryos to grow until HH9, when cranial neural crest specification occurs and delamination begins. Our results indicated that the chP sequence is indeed capable of driving low level GFP expression throughout the embryo (Fig. 1D). Upon immunostaining for neural crest marker Pax7 and sectioning the embryos, we found that chP-GFP expression appears to emulate the distinct *Chiwi* expression pattern in the neural crest–peaking just within the Pax7+ domain and appearing to drop again in the presumptive specified neural crest region. We also noted sporadic

expression throughout the non-neural ectoderm, where we don't see endogenous *Chiwi* expression, suggesting that other regulatory factors or elements outside of this genomic region likely contribute to *Chiwi* regulation as well.

We then used Jaspar to identify putative neural crest transcription factor binding sites within the chP sequence. We scanned chP for Msx1, Zic1, Snai2, Pax7, C-Myb, Ets1, Sox9, FoxD3, TFAP2A and Sox10 sites, using available matrices from mouse, rat, and human. At a relative profile score threshold of 95%, putative binding sites were identified for Msx1, C-Myb, Ets1, TFAP2A and Sox10, but when we increased the relative profile score threshold to 100%, one Msx1 site, one C-Myb site, and three Sox10 sites remained. Given A-Myb's role in piRNA pathway regulation in the germline and the overlapping DNA binding preferences of Myb proteins, we were particularly interested in the relevance of the Myb binding site, which was predicted to be 7 bp long and spanned the start of the CpG island. Comparison of this region to the chicken A-Myb binding site identified by Li et al via chip-seq of rooster testes (*16*) also revealed a match, and expanded the putative binding site to a length of 10bp.

A-Myb regulates Chiwi in the developing neural crest

To determine if A-Myb could be regulating the piRNA pathway during neural crest development, we first performed HCR to resolve the expression pattern of *A-Myb* in the developing neural tube. Our results indicate that *A-Myb* is expressed at very low levels throughout the neural tube and neural folds (Fig. 2A), which is consistent with the low expression levels observed in our neural fold RNA-seq data (*5*).

We then directly tested the function of A-Myb in the neural crest by knocking it out using a translation blocking morpholino oligomer (MO). We electroporated A-Myb MO on the right side of HH4 embryos, with a standard control MO on the left, and grew the embryos to HH9 (Fig. 2B). HCR for *Chiwi* revealed a significant decrease in *Chiwi* expression of 19% on average (Fig. 2C).

To confirm the specificity of our results, we employed a second approach to deplete A-Myb in the neural crest using a single-plasmid based CRISPR/Cas9 system optimized for use in the chicken

embryo (*24*). To this end, we designed a CRISPR guide RNA targeting the second exon junction and start of the DNA-binding domain. We electroporated this construct into the right side of the embryo at HH4, along with a CRISPR/Cas9 control guide construct on the left side, and grew the embryos to HH9. HCR for *Chiwi* revealed a slight but significant reduction in *Chiwi* expression upon *A-Myb* knockout (KO), with a 10% reduction in *Chiwi* HCR signal on average (Fig. 2D). To confirm the KO, we also did HCR for *A-Myb*, which similarly indicated a slight reduction in *A-Myb* expression, with 3/4 embryos displaying an average reduction of 15%, suggesting that the CRISPR KO worked but was not very efficient (Fig. 2D).

To determine if A-Myb or other Myb proteins directly regulate Chiwi expression via the putative binding site we identified in the chP sequence, we designed a CRISPR guide RNA targeting that site (Fig. 2E). Upon electroporating HH4 embryos with this and a control construct, growing them to HH9 and performing HCR for *Chiwi*, we found that disruption of this binding site indeed reduces Chiwi expression significantly, at an average reduction of 11% (Fig. 2E). While this reduction is significant, it is small, which might be explained by the inefficiency of the CRISPR approach in this system, or the possibility that this particular site is one of many factors contributing to *Chiwi* regulation in the neural crest, or perhaps a combination of the two.

C-Myb is transiently upregulated at the onset of neural crest specification

Though the expression pattern of *C-Myb* in the neural crest has been previously reported (*18*), techniques for visualization of mRNA have significantly improved since then (*25*). To better resolve the precise expression pattern of *C-Myb* in neural crest cells, we performed HCR for *C-Myb* on wildtype embryos at HH8 and HH9 (Fig. 3A). With this increased resolution, we found that C-Myb is transiently upregulated at the onset of neural crest specification. Specification of the neural crest occurs in a rostral-caudal wave, and peak *C-Myb* expression is visible just caudal to peak Snai2 expression—a marker for specification—in the dorsal neural tube, indicating that C-Myb expression peaks just prior to peak expression of neural crest specification markers.

Co-hybridization of *Chiwi* and *C-Myb* revealed that *C-Myb* expression peaks in the same specified neural crest cells in which *Chiwi* appears to be downregulated (Fig. 3B). This raises the intriguing possibility that A-Myb and C-Myb play reciprocal roles in the developing neural crest, with A-Myb activating piRNA pathway activity, and C-Myb transiently downregulating it via competitive binding to allow for the completion of the neural crest specification circuit.

Depletion of C-Myb in the neural crest increases Chiwi expression

To determine whether C-Myb regulates Chiwi in the neural crest, we again turned to a single plasmid based CRISPR/Cas9 system. We designed two guide RNAs targeting C-Myb's DNA binding domain, and co-electroporated them into the right side HH4 embryos along with a control construct on the left, before allowing them to grow to HH9. HCR for *Chiwi* revealed a small but significant increase in Chiwi expression of about 4% (Fig. 3C). HCR for *C-Myb* indicated a significant but minor reduction in C-Myb signal of 11% on average, again suggesting that the KO worked, but was not very efficient (Fig. 3C). Despite such a minor reduction in *C-Myb*, the fact that we did see a significant change in *Chiwi* expression supports the hypothesis that C-Myb acts to repress *Chiwi* in the neural folds at the onset of specification.

To further validate this hypothesis, we attempted to deplete C-Myb using a second method. To this end, we employed a splice blocking MO targeting exon 6 of C-Myb, which should cause a frameshift in the downstream sequence and lead to nonsense mediated decay of the transcript. We electroporated this MO into HH4 embryos along with a control, and subsequently performed HCR for *C-Myb* at HH9 to determine if the knockdown worked. Unfortunately, we saw a strong increase in *C-Myb* signal on the side of the embryos containing C-Myb MO (N=4/4), rather than the decrease we would expect if nonsense mediated decay were induced (Fig. 3D). HCR for *Chiwi* showed an apparent decrease in *Chiwi* signal, which would be consistent with an upregulation of *C-Myb* repressing *Chiwi* expression, though at this point we only have an N of 1 embryo (Fig. 3D). We hypothesize that these results are due to a compensation effect triggered by an earlier decrease in C-Myb activity caused by the MO, and that by HH9 the embryo has upregulated *C-Myb* transcript in response, to the point that it overcomes the concentration of unbound MO. This could be investigated

further by performing RT-qPCR to determine the ratio of spliced to unspliced transcript. It might also be possible to overcome this issue by increasing the concentration of MO.

To examine whether *C-Myb* expression might change in response to A-Myb perturbation, we performed HCR for *C-Myb* on embryos electroporated with A-Myb MO, and found that this results in a decrease in C-Myb expression upon A-Myb knockdown (Fig. 3E). This supports the hypothesis that A-Myb promotes *Chiwi* expression while C-Myb represses it, and that to compensate for a reduction in A-Myb and thus Chiwi, the embryo might downregulate C-Myb in response.

A-Myb and C-Myb knockdown have opposite effects on chP-GFP reporter expression

To further explore the role of Myb proteins in Chiwi regulation, we tested whether A-Myb depletion affects expression of the Chiwi promoter reporter construct, chP-GFP. We hypothesized that since loss of A-Myb reduces endogenous Chiwi expression, it should similarly reduce expression of chP-GFP, which contains a Myb binding site. As our CRISPR constructs already express GFP as an electroporation control, we utilized the A-Myb MO for this experiment. We co-electroporated A-Myb MO and chP-GFP on the right side of HH4 embryos with control MO and chP-GFP on the left side, and grew them to HH9. Since the MO is FITC tagged, which fluoresces at the same wavelength as GFP, we immunostained for and quantified GFP expression on the far-red channel, in the left versus right neural folds of cranial sections. Indeed, our data confirmed that A-Myb knockdown significantly reduces chP-GFP expression, with an average decrease of 25% (Fig. 4A).

We then asked whether C-Myb knockdown affected chP-GFP expression. Analysis of HH9 embryos co-electroporated with C-Myb MO and chP-GFP indicates an increase in reporter expression compared to the control MO side (Fig. 4A).

Given our hypothesis that the C-Myb knockdown is inducing a compensatory response, we wondered if looking at its effects on the chP-GFP reporter at a slightly earlier timepoint would elucidate things further. To this end, we grew embryos co-electroporated as previously described to HH8 instead of HH9, and analyzed them for GFP expression. Though we currently only have an N of 2, we saw significantly stronger activation of the chP-GFP reporter at HH8 compared to HH9

(Fig. 4B), supporting the hypothesis that the embryo activates a compensatory response to C-Myb knockdown, prior to which loss of C-Myb does in fact de-repress the Chiwi promoter. In contrast, analysis of A-Myb MO and chP-GFP electroporated embryos at HH8 does not appear significantly different to those analyzed at HH9, though a higher N is needed to confirm this (Fig. 4C).

Discussion

Our previous work demonstrating that the piRNA pathway has been co-opted into the chick neural crest GRN has raised many intriguing questions regarding how the pathway is regulated and if there is conservation in regulatory mechanisms with those in play in the germline. Here, we provide preliminary evidence that the Myb family genes A-Myb and C-Myb may act antagonistically to regulate Chiwi in the neural crest. A-Myb, which is expressed at low levels, activates Chiwi throughout the neural tube, while C-Myb is transiently upregulated at the onset of neural crest specification, where it appears to function to downregulate Chiwi, possibly by outcompeting A-Myb for the Myb binding site within the Chiwi promoter. Thus, Myb proteins may contribute to the sharp downregulation of *Chiwi* in the neural crest, which allows for ERNI to be expressed and repress Sox2, and consequently the completion of the neural crest specification program that is otherwise inhibited by high levels of Sox2.

Though we use *Chiwi* here as a readout for piRNA pathway activity in the neural crest, in the germline A-Myb activates not just Piwil1, but several other piRNA pathway proteins as well as piRNA precursor transcripts (*16*). We think it is likely functioning similarly in the neural crest, as an important regulator of the neural crest piRNA pathway. Likewise, due to the overlapping binding preferences of A-Myb and C-Myb and their competitive regulation of Chiwi, it is possible that C-Myb functions as a critical repressor of the piRNA pathway in this context, downregulating not just Chiwi, but piRNA expression and other pathway genes as well.

Given the strong conservation of A-Myb and C-Myb throughout the vertebrate lineage, as well as A-Myb's conserved function in piRNA pathway regulation in the germline, the incorporation of these proteins into the neural crest GRN might lend support to the hypothesis that the piRNA pathway is a conserved feature of neural crest development (*16*, *26*). Furthermore, C-Myb is known

to play a role in the regulation of several other critical neural crest genes, suggesting that it is a well-established member of the neural crest GRN, rather than a recent addition to regulate Chiwi (*18–20*).

Prior to neurulation C-Myb is known to promote Pax7 expression in the neural plate border (NPB) via direct binding to a Pax7 enhancer, and loss of C-Myb leads to a reduction in Pax7 (*18*, *19*). Interestingly, Pax7 is inhibited by Sox2 expression, and as such, Sox2 must be repressed in the neural plate border to maintain its identity. In chick, this is partially achieved by ERNI, which serves to repress Sox2 in the neural plate prior to neural induction, but is also maintained in the neural plate border prior to neurulation (*27*). We now know that the piRNA pathway functions to regulate ERNI during neural crest specification later in development, but Chiwi is also expressed at low levels at the stages prior to neurulation (*5*). If C-Myb is functioning at these earlier stages to maintain neural plate border identity by promoting Pax7 expression, it is possible that it is also repressing Chiwi in these cells in order to preserve ERNI expression, thus mounting a two-edged defense of neural plate border identity.

Until recently, piRNA pathway activity was largely considered to be confined to the germline in vertebrates, where it serves to protect genomic integrity for the next generation by repressing transposons. Our recent work has demonstrated that the piRNA pathway also has the potential to interact with transposons in a much more nuanced manner and possibly drive evolution of new genes and regulatory circuits. Many fascinating questions remain, including how this may have occurred given the number of unique components that must be expressed to activate a functional piRNA pathway, as well as whether it is a conserved facet of the vertebrate neural crest gene regulatory network. Though ERNI, the target of the piRNA pathway in the chick neural crest, is derived from a chicken specific transposon, substitution of the transposon derived target throughout evolution could be easily enabled by the unique adaptations of the piRNA pathway, which is constantly updating to target the most active transposons in a given species (*5*). Understanding how the piRNA pathway is regulated in the neural crest not only sheds light on how it was incorporated into a somatic gene regulatory network to regulate a developmentally relevant gene, but might also help us detect similar co-option events to better understand how the piRNA pathway may drive evolution.

Materials and Methods

Cloning of expression vectors

The chP-GFP construct was generated by PCR amplifying a 1131bp long region directly upstream of the Chiwi start codon (chP), which contains a ~450bp CpG island, from chicken genomic DNA. The coding sequence for eGFP was subsequently fused downstream of the chP promoter via fusion PCR, and the resulting amplification product was cloned into the pTK vector using Acc651 and Xba1 restriction sites to entirely replace the basal promoter in the pTK vector (chP-GFP).

CRISPR/Cas9 guide RNA sequences were designed using CHOPCHOP(*28*) and cloned into gRNA/Cas9 expression vectors as previously described (*24*).

Electroporation

Fertilized chicken eggs were acquired from Sun State Ranch (Sylmar, CA) and Petaluma Egg Farm (Petaluma, CA), and grown at 37°C for 18-20 hours to reach HH4-5. Ex ovo electroporations were performed on stage HH4-5 embryos. Embryos were dissected onto rings of filter paper in Ringer's and a solution of DNA expression construct or morpholino (MO) was injected into the space between the vitelline membrane and ectoderm (Fig. 3A) and electroporated into the ectoderm with 5 pulses of 5.2V for 50ms, with 100ms between each pulse. Embryos were then cultured at 37°C in thin albumen with penicillin/streptomycin, or on a thin albumen/agar mixture with penicillin/streptomycin until HH9. Knock down and knockout embryos were bilaterally electroporated with the control on the left side and experimental on the right side, allowing for direct comparison.

The chP-GFP expression construct was injected at a concentration of 1μg/ul, while CRISPR constructs were electroporated at a concentration of 2.5ug/ul when using a single guide, and 1.5ug/ul each when using two guides. MOs were used at a concentration of 0.5-1.5mM with 1.0μg/ul pCIG-GFP or chP-GFP as carrier DNA. FITC labelled MOs used include standard control MO (5′–

CCTCTTACCTCAGTTACAATTTATA–3′), A-Myb translation blocking MO (5’-GCCATCCTCGGGTCGCAGCATA-3’), and C-Myb splice blocking MO (5’-GCGCTTTCAGGAGCTTACATTTTGT-3’) from Gene Tools.

CRISPR mediated knockout was performed as described previously, with a non-binding control guide construct electroporated contralateral to targeted experimental guide constructs (*24*). One guide RNA was designed to target the Myb binding site in the chP sequence (5’-CCTTCAACGGCCGCTCGTTCCGC-3’). One guide RNA was designed to target the second exon junction and DNA binding site of A-Myb (5’-GGTGAAATGGACACGTGATG-3’), and two guide RNAs were designed to target the C-Myb DNA binding site (5’-AGGTCCATGGACTAAAGAGG-3’, 5’-ATACGGTCCAAAGCGCTGGT-3’).

Immunofluorescence

All embryos were fixed for 30 minutes at room temperature in 4% paraformaldehyde, and subsequently blocked in 10% donkey serum in PBST (PBS, 0.5% Tween-20) for two hours at room temperature. Both primary and secondary antibody incubations occurred at 4°C for two nights in 10% donkey serum, with four one hour washes in PBST at room temperature after primary, and two 30 minute washes in PBST after secondary antibody incubation. After imaging, whole mount embryos were post-fixed in 4% paraformaldehyde overnight at 4°C prior to sectioning. Primary antibodies used: Goat anti-GFP (1:2500), Rockland 600-101-215, Rabbit anti-FITC (1:500), Thermo-Fisher 71-1900. Secondary antibodies used: Molecular Probes donkey secondary antibody conjugated to Alexa Fluor 488 or 647 (1:1000).

In situ Hybridization Chain Reaction (HCR)

All HCR was performed with probes designed by Molecular Technologies and following the published V3 protocol (*25*). 20-probe sets were used for all genes. Prior to sectioning, embryos were postfixed in 4% paraformaldehyde overnight at 4°C.

Sectioning

Cryosectioning was performed at a thickness of 18μm on a Microm HM550 cryostat. Embryo preparation included fixation in 4% paraformaldehyde overnight at 4°C (either from live embryos to post-fix processed embryos), followed 15% sucrose overnight at 4°C and 7.5% gelatin overnight at 37°C prior to mounting in silicone molds and snap freezing in liquid nitrogen.

Imaging and statistical analysis

All images were taken using a Zeiss AxioImager.M2 with an Apotome.2. Fluorescence intensity for experimental quantification was measured from maximum intensity projections of Z-stack images by manually drawing regions of interest to measure average intensity and subtracting average intensity of background regions. Experimental values were then divided by control values from the same image. For fluorescent intensity quantification, three non-adjacent cranial sections were measured and averaged to create a representative value for each embryo. For all statistical analysis on images, a paired two-tailed Student *t* test was performed to compare two values (experimental and control) within single embryos, or an unpaired two-tailed Student *t* test was used to compare values from different sets of embryos.

Acknowledgements

We thank members of the M.E.B., K.F.T., and Aravin laboratories for helpful discussions. We thank Michael Piacentino for providing Fiji macros for our use. We thank Fred He for help with cloning and maxi prepping constructs.

References

1. N. M. Le Douarin, E. Dupin, The neural crest in vertebrate evolution. *Curr Opin Genet Dev*. **22**, 381–389 (2012).

2. M. E. Bronner, N. M. LeDouarin, Development and evolution of the neural crest: an overview. *Dev Biol*. **366**, 2–9 (2012).

3. M. E. Bronner, M. Simoes-Costa, The Neural Crest Migrating into the Twenty-First Century. *Curr Top Dev Biol*. **116**, 115–134 (2016).

4. M. Simoes-Costa, M. E. Bronner, Establishing neural crest identity: a gene regulatory recipe. *Development (Cambridge, England)*. **142**, 242–257 (2015).

5. R. Galton, K. Fejes-Toth, M. E. Bronner, Co-option of the piRNA pathway to regulate neural crest specification. *Sci Adv*. **8**, eabn1441 (2022).

6. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)*. **316**, 744–747 (2007).

7. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*. **128**, 1089–1103 (2007).

8. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*. **313**, 320–324 (2006).

9. Y. Wakamatsu, Y. Endo, N. Osumi, J. A. Weston, Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Dev Dyn*. **229**, 74–86 (2004).

10. D. Roellig, J. Tan-Cabugao, S. Esaian, M. E. Bronner, Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cells. *eLife*. **6**, doi:10.7554/eLife.21620.

11. B. Czech, M. Munafo, F. Ciabrelli, E. L. Eastwood, M. H. Fabry, E. Kneuss, G. J. Hannon, piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet.*. **52**, 131–157 (2018).

12. A. K. Rogers, K. Situ, E. M. Perkins, K. F. Toth, Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes Dev*. **31**, 1858–1869 (2017).

13. K. F. Tóth, D. Pezic, E. Stuwe, A. Webster, The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Adv Exp Med Biol*. **886**, 51–77 (2016).

14. D. Ding, J. Liu, K. Dong, U. Midic, R. A. Hess, H. Xie, E. Y. Demireva, C. Chen, PNLDC1 is essential for piRNA 3′ end trimming and transposon silencing during spermatogenesis in mice. *Nat Commun*. **8**, 819 (2017).

15. T. Watanabe, H. Lin, Post-transcriptional regulation of gene expression by Piwi proteins and piRNAs. *Mol Cell*. **56**, 18–27 (2014).

16. X. Z. Li, C. K. Roy, X. Dong, E. Bolcun-Filas, J. Wang, B. W. Han, J. Xu, M. J. Moore, J. C. Schimenti, Z. Weng, P. D. Zamore, An Ancient Transcription Factor Initiates the Burst of piRNA Production During Early Meiosis in Mouse Testes. *Mol Cell*. **50**, 67–81 (2013).

17. I. H. Oh, E. P. Reddy, The myb gene family in cell growth, differentiation and apoptosis. *Oncogene*. **18**, 3017–3033 (1999).

18. P. Betancur, M. Simoes-Costa, T. Sauka-Spengler, M. E. Bronner, Expression and function of transcription factor cMyb during cranial neural crest development. *Mech Dev*. **132**, 38–43 (2014).

19. S. Vadasz, J. Marquez, M. Tulloch, N. A. Shylo, M. I. García-Castro, Pax7 is regulated by cMyb during early neural crest development through a novel enhancer. *Development*. **140**, 3691–3702 (2013).

20. P. Betancur, M. Bronner-Fraser, T. Sauka-Spengler, Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proc Natl Acad Sci U.S.A*. **107**, 3570–3575 (2010).

21. G. Foos, S. Grimm, K. H. Klempnauer, The chicken A-myb protein is a transcriptional activator. *Oncogene*. **9**, 2481–2488 (1994).

22. G. Foos, S. Grimm, K. H. Klempnauer, Functional antagonism between members of the myb family: B-myb inhibits v-myb-induced gene activation. *EMBO J*. **11**, 4619–4629 (1992).

23. R. J. Watson, C. Robinson, E. W. Lam, Transcription regulation by murine B-myb is distinct from that by c-myb. *Nucleic Acids Res*. **21**, 267–272 (1993).

24. S. Gandhi, Y. Li, W. Tang, J. B. Christensen, H. A. Urrutia, F. M. Vieceli, M. L. Piacentino, M. E. Bronner, A single-plasmid approach for genome editing coupled with long-term lineage analysis in chick embryos. *Development*. **148**, dev193565 (2021).

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

26. C. J. Davidson, R. Tirouvanziam, L. A. Herzenberg, J. S. Lipsick, Functional Evolution of the Vertebrate Myb Gene Family. *Genetics*. **169**, 215–229 (2005).

27. A. Streit, A. J. Berliner, C. Papanayotou, A. Slrulnik, C. D. Stern, Initiation of neural induction by FGF signalling before gastrulation. *Nature*. **406**, 74–78 (2000).

28. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing | Nucleic Acids Research | Oxford Academic, (available at https://academic.oup.com/nar/article/47/W1/W171/5491735).

Figures

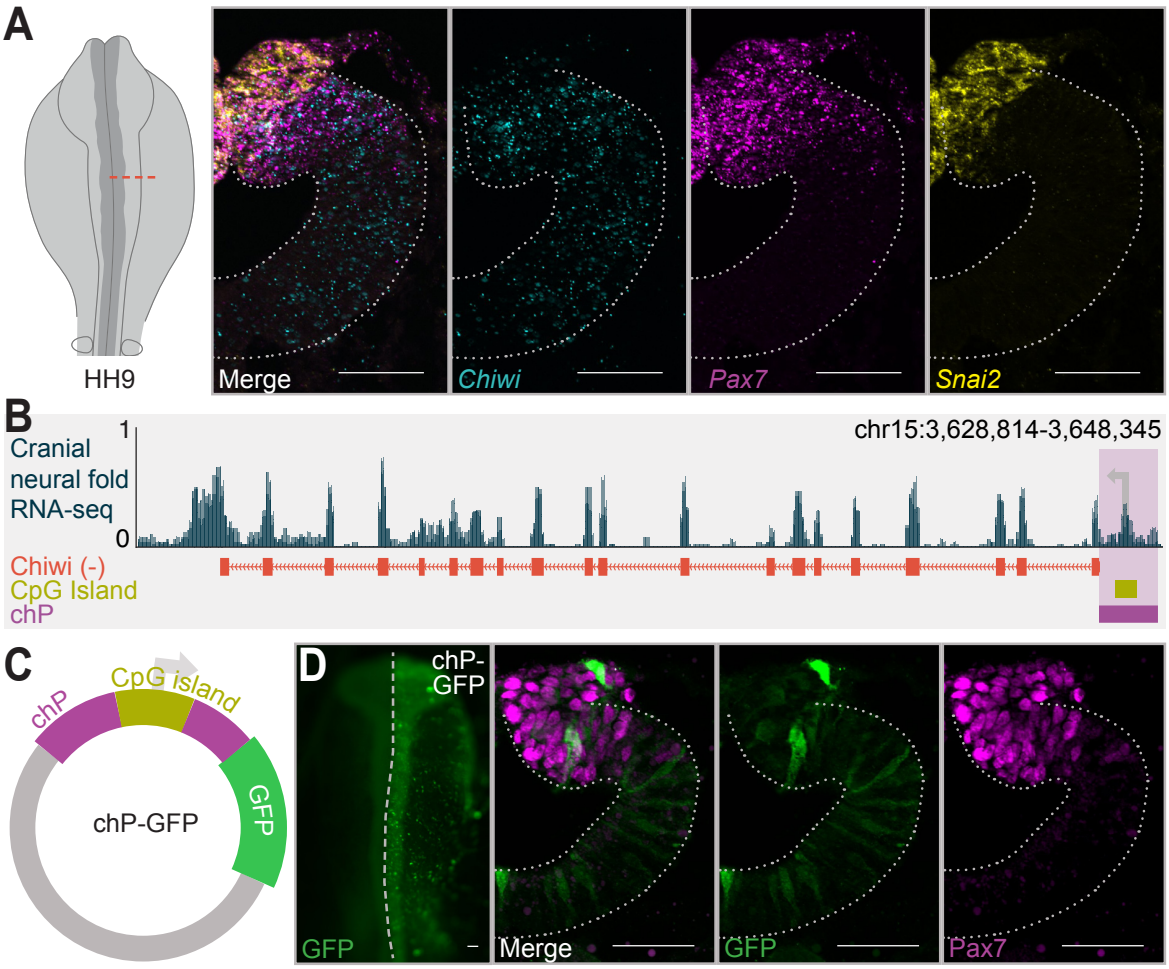

Figure 1. A reporter driven by the Chiwi promoter sequence recapitulates Chiwi's expression pattern in the neural crest. (A) HCR reveals the endogenous expression of *Chiwi* and neural crest markers *Pax7* and *Snai2* at HH9; scale bars = 50µm. (B) RNA-seq tracks from HH9 cranial neural folds depicting reads mapping to the Chiwi locus and its promoter region, chP. Reads are normalized to reads per million mapped reads, replicate tracks are overlaid. (C) Schematic of the chP-GFP reporter construct. (D) Expression of the chP-GFP reporter construct at HH9. Immunofluorescence for neural crest marker Pax7 in cranial sections (right panels) reveals that the expression pattern of chP-GFP mirrors that of endogenous *Chiwi*; scale bars = 50µm.

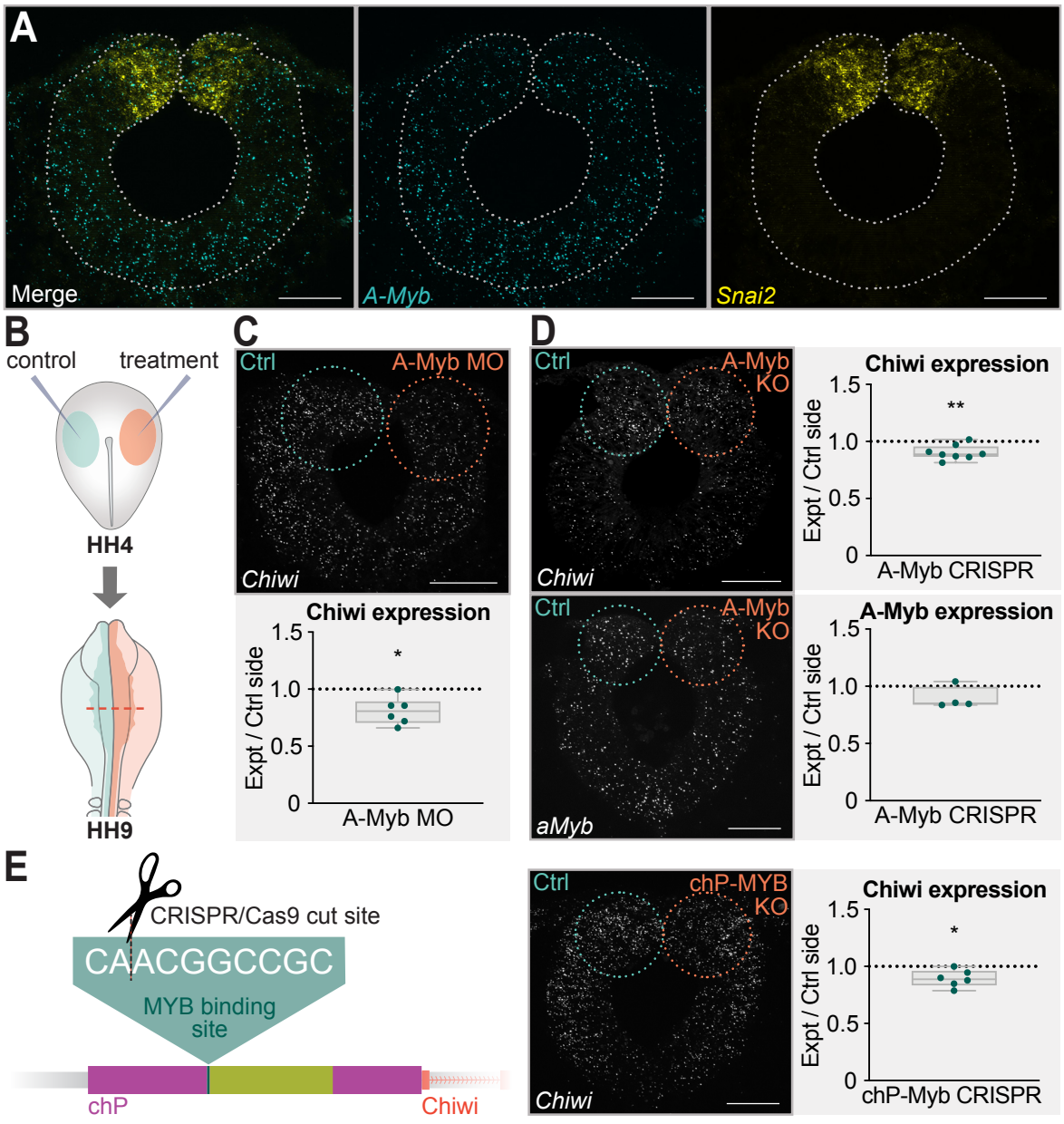

Figure 2. A-Myb is expressed in the neural tube and regulates Chiwi expression. (A) HCR reveals the endogenous expression of *A-Myb* and neural crest marker *Snai2* at HH9. (B) Schematic diagram of the chick embryo bilateral electroporation strategy. (C) Morpholino mediated knockdown of A-Myb reduces *Chiwi* expression at HH9, as indicated by HCR. (D) HCR reveals

that CRISPR/Cas9 mediated knockout of A-Myb reduces *Chiwi* and *A-Myb* expression at HH9. (E) HCR reveals that CRISPR/Cas9 mediated disruption of the Myb binding site in the chP region reduces Chiwi expression at HH9. All scale bars = 50μm. Each data point in the graphs represents the average fluorescent intensity of HCR signal in the right dorsal fold region (experimental) divided by the left (control) side from three non-adjacent sections from the cranial region of a single embryo. The teal circles denote the control side while the orange circles denote the experimental side. Box plots indicate the interquartile range, while whiskers extend to min and max values. * and ** indicate p values of $\leq$ 0.05 and 0.01, respectively, and represent the difference between control and experimental measurements for each treatment.

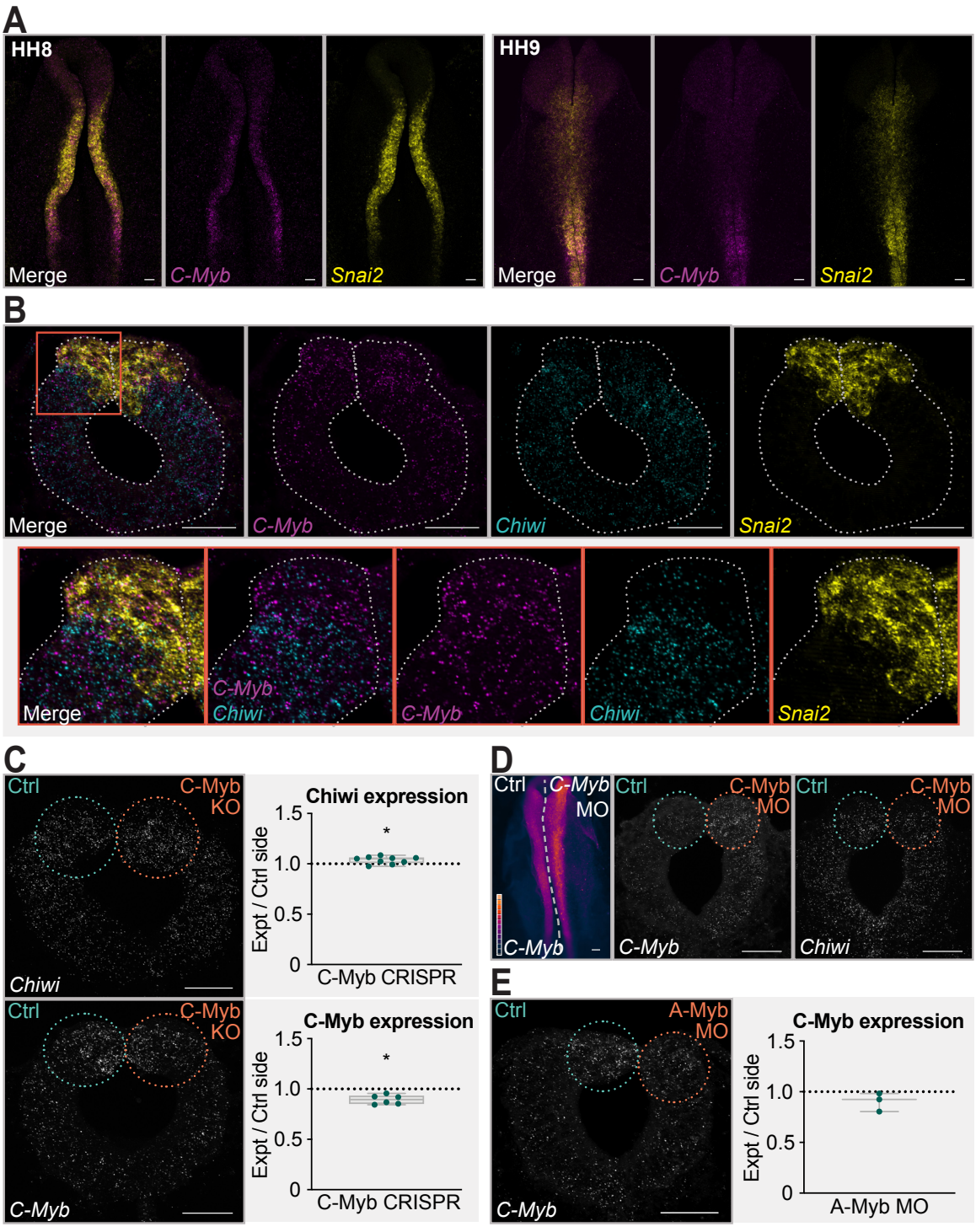
A
HH8
Merge
C-Myb
Snai2
HH9
Merge
C-Myb
Snai2
B
Merge
C-Myb
Chiwi
Snai2
Merge
C-Myb
Chiwi
C-Myb
Chiwi
Snai2
C
Ctrl
C-Myb KO
Chiwi
Chiwi expression
Expt / Ctrl side
1.5
1.0
0.5
0
*
C-Myb CRISPR
Ctrl
C-Myb KO
C-Myb
C-Myb expression
Expt / Ctrl side
1.5
1.0
0.5
0
*
C-Myb CRISPR
D
Ctrl
C-Myb MO
C-Myb
Ctrl
C-Myb MO
C-Myb
Ctrl
C-Myb MO
Chiwi
E
Ctrl
A-Myb MO
C-Myb
C-Myb expression
Expt / Ctrl side
1.5
1.0
0.5
0
A-Myb MO

Figure 3. C-Myb is transiently upregulated upon neural crest specification and downregulates Chiwi expression. (A) HCR reveals the endogenous expression pattern of *C-Myb* alongside specified neural crest marker *Snai2* in whole mount at HH8 (left) and HH9 (right). (B) HCR reveals the endogenous expression pattern of *C-Myb* alongside *Chiwi* and *Snai2* in a cranial neural tube section at HH9 (top). Enlargement of the neural fold region indicates that *C-Myb* expression appears to peak in the same cells where *Chiwi* expression drops (bottom). (C) HCR reveals that CRISPR/Cas9 mediated knockout of C-Myb reduces *Chiwi* and *C-Myb* expression at HH9. (D) HCR reveals that a splice blocking MO against C-Myb induces upregulation of *C-Myb* transcript at HH9 (N=4/4), which appears to reduce *Chiwi* expression (N=1/1). (E) Morpholino mediated knockdown of A-Myb reduces *C-Myb* expression at HH9, as indicated by HCR. All scale bars = 50μm. Each data point in the graphs represents the average fluorescent intensity of HCR signal in the right dorsal fold region (experimental) divided by the left (control) side from three non-adjacent sections from the cranial region of a single embryo. The teal circles denote the control side while the orange circles denote the experimental side. Box plots indicate the interquartile range, while whiskers extend to min and max values. * indicats a p value of ≤ 0.05 and represents the difference between control and experimental measurements for each treatment.

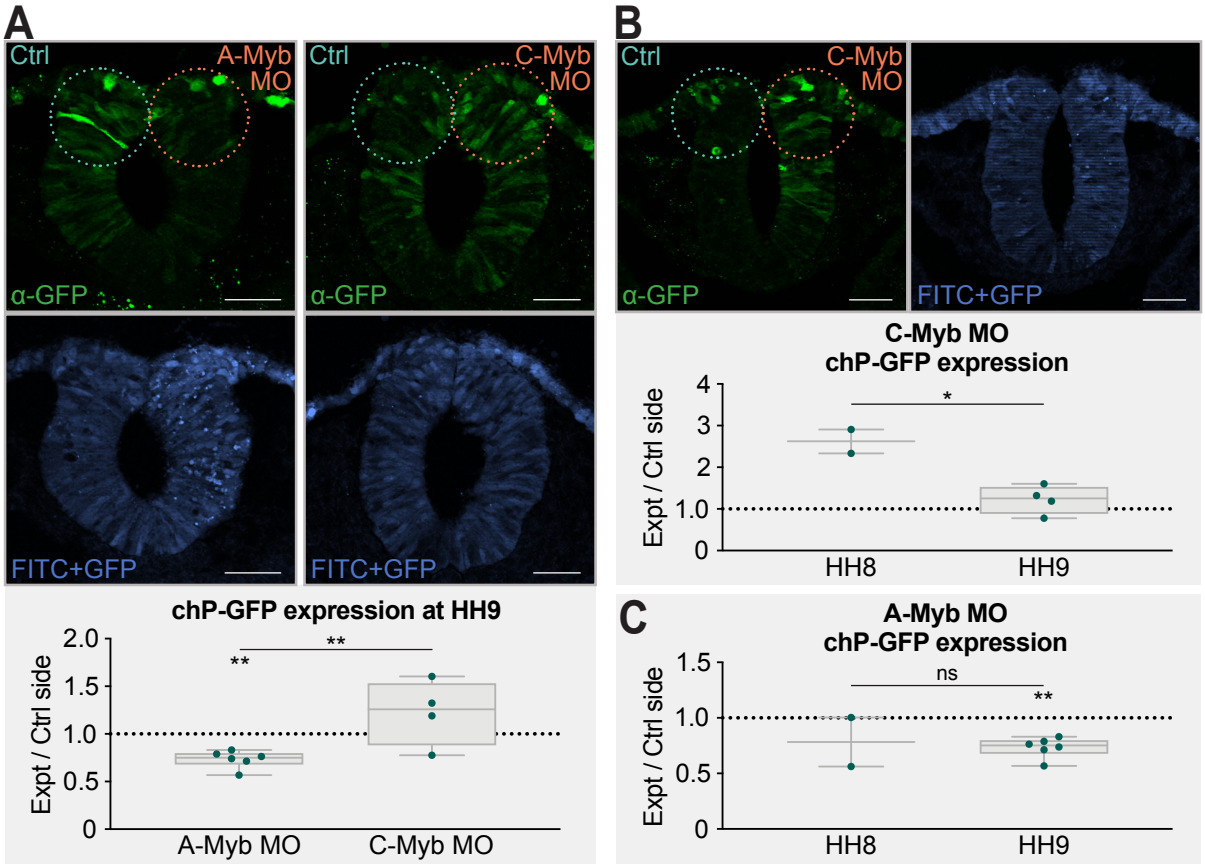

Figure 4. A-Myb and C-Myb knockdown have opposite effects on chP-GFP reporter expression. (A) Immunostaining for GFP in A-Myb and C-Myb knockdown embryos at HH9 reveals opposite regulation the chP-GFP reporter construct by A-Myb and C-Myb. (B) Immunostaining for GFP in C-Myb knockdown embryos at HH8 reveals an even stronger upregulated on the chP-GFP reporter construct than at HH9. (C) Comparison of immunostaining for GFP in A-Myb knockdown embryos at HH8 and HH9. All scale bars = 50μm. Each data point in the graphs represents the average fluorescent intensity of HCR signal in the right dorsal fold region (experimental) divided by the left (control) side from three non-adjacent sections from the cranial region of a single embryo. The teal circles denote the control side while the orange circles denote the experimental side. Box plots indicate the interquartile range, while whiskers extend to min and max values. * and ** indicate p values of ≤ 0.05 and 0.01, respectively, and represent the difference between control and experimental measurements for each treatment.

Chapter 5

Conclusion

Transposable elements have long been considered a major driving force of evolution, providing genomes with a steady source of de novo mutations and raw genetic material to mold into new adaptations. Despite this, transposons wreak havoc on our genomes when left unchecked, as their primary evolutionary purpose, like all of ours, is self-propagation (*1*). The piRNA pathway, conserved across metazoa and designed to rapidly adapt to new transposon invasions, is our genome's main line of defense against transposon-induced genomic instability in the germline, where preservation of genomic information is critical to the perpetuation of the species (*1–4*). The genetic conflict between transposons and the piRNA pathway has often been described an arms race, with each side constantly evolving new offensive and defensive strategies against the other (*1*).

The work presented here establishes an entirely novel role for the piRNA pathway as a mediator of transposon co-option into a host gene regulatory network. We show, for the first time, that the piRNA pathway is not only active in a vertebrate somatic tissue, but that it is capable of spatiotemporal gene regulation to modulate a key developmental process: neural crest specification (*5*). Furthermore, we present preliminary data suggesting that regulation of the pathway's expression in the neural crest is mediated by the highly conserved Myb gene family, which are critical components of many other developmental processes. Intriguingly, the neural crest piRNA pathway functions at expression levels significantly below those required to adequately repress transposon activity in the germline, potentially explaining why piRNA and Piwi protein expression has rarely been noted in somatic tissues.

We hypothesize that our results represent the tip of the iceberg, and that the piRNA pathway plays a much more significant role in evolution and development than previously thought. Our work expands the potential evolutionary relevance of the piRNA pathway beyond its previously understood purpose of globally repressing transposons in the germline. We propose, instead, that the piRNA pathway serves as an interface between our genomes and the world of mobile genetic

elements, providing an adaptive framework through which our genomes can safely test new transposon-derived sequences for co-option into host processes.

One avenue for further exploration that may lend credence to this hypothesis is whether the neural crest piRNA pathway is conserved in other species. While its transposon-derived target in the neural crest is unique to chickens and closely related birds (*5*), the piRNA pathway's unique biology allows it to rapidly evolve repressive function against newly introduced transposon species while concurrently maintaining control over older ones (*1*). When incorporated into a gene regulatory network, this might allow the piRNA pathway to employ a cut-and-paste type mechanism whereby new transposon-derived genes can be swapped out as transposon activity fluctuates between species. Intriguingly, zebrafish neural crest cells express Crestin, a transposon-derived sequence of unknown function, but similar structure to ERNI (*6*). Whether Crestin expression is similarly regulated by the piRNA pathway, or whether it functions to repress Sox2 in the neural crest, are intriguing questions that might lead to direct evidence of a cut-and-paste, piRNA-mediated mechanism for co-option of new genes.

Another fascinating question that requires further study is that of how prevalent piRNA pathway incorporation into host gene regulatory networks is. While our work represents the first report of it, many studies have reported Piwi expression in various somatic tissues. One example of this is in the rodent hippocampus, where both Piwil1 and Piwil2 are expressed and lead to behavioral changes when repressed, which the authors note could be explained by defective neurogenesis (*7*). Line1 expression also correlates with Sox2 downregulation at the onset of neuronal differentiation in the hippocampus (*8*), and Sox2 downregulation is considered a critical step in the initiation of hippocampal neurogenesis, possibly suggesting that Sox2 is maintaining stem-like cells that give rise to adult neuronal precursors, much like its function in the neural crest (*9*). Together with our data, these reports raise the question of whether the piRNA pathway might be repressing a transposon-derived gene to maintain Sox2 expression in neural stem cells in the hippocampus, and upon the onset of differentiation, piRNA pathway activity is downregulated which in turn leads to the reported upregulation of Line1 elements.

The above are just few examples of areas for further study where our results might help to answer questions raised by previous work in other systems. Investigating how the piRNA pathway might be involved in these systems, as well as the many others where Piwi expression has been reported or known transposon co-option events have occurred, could further establish its role as a mediator of transposon co-option into host gene regulatory networks, or perhaps elucidate entirely new functions for this highly adaptable pathway and its unique components, broadening our understanding of the piRNA pathway's unique place in evolutionary history even more.

References

1. A. A. Aravin, G. J. Hannon, J. Brennecke, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science (New York, N.Y.)*. **318**, 761–764 (2007).

2. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)*. **316**, 744–747 (2007).

3. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*. **128**, 1089–1103 (2007).

4. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*. **313**, 320–324 (2006).

5. E. Lerat, A.-M. Birot, J. Samarut, A. Mey, Maintenance in the Chicken Genome of the Retroviral-like cENS Gene Family Specifically Expressed in Early Embryos. *J Mol Evol*. **65**, 215–227 (2007).

6. A. L. Rubinstein, D. Lee, R. Luo, P. D. Henion, M. E. Halpern, Genes dependent on zebrafish cyclops function identified by AFLP differential gene expression screen. *Genesis*. **26**, 86–97 (2000).

7. L. J. Leighton, W. Wei, P. R. Marshall, V. S. Ratnu, X. Li, E. L. Zajaczkowski, P. A. Spadaro, N. Khandelwal, A. Kumar, T. W. Bredy, Disrupting the hippocampal Piwi pathway enhances contextual fear memory in mice. *Neurobiol Learn Mem*. **161**, 202–209 (2019).

8. A. R. Muotri, V. T. Chu, M. C. N. Marchetto, W. Deng, J. V. Moran, F. H. Gage, Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. **435**, 903–910 (2005).

9. T. Kuwabara, J. Hsieh, A. Muotri, G. Yeo, M. Warashina, D. C. Lie, L. Moore, K. Nakashima, M. Asashima, F. H. Gage, Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci*. **12**, 1097–1105 (2009).