

Identification of a novel downstream *single-minded* midline regulatory element in *Drosophila melanogaster*

Geyenna Sterling-Lentsch¹, Marc S. Halfon^{1§}

¹Department of Biochemistry, University at Buffalo, State University of New York, Buffalo, New York, United States

[§]To whom correspondence should be addressed: mshalfon@buffalo.edu

Abstract

Development of the *Drosophila melanogaster* central nervous system midline depends on the gene *single-minded* (*sim*). Although *sim* regulation has been studied extensively, the fact that an enhancer mediating late embryonic *sim* transcription has not been identified suggests that additional regulatory sequences remain unknown. We tested several evolutionarily conserved sequences in the *sim* downstream region and isolated *sim_3pB*, whose midline activity in a reporter gene assay begins later than previously characterized *sim* enhancers. Its activity shares several key similarities with the *Aedes aegypti sim_5P3* enhancer, though is sufficiently different to warrant further investigation into how *sim_3pB* functions in its native context.

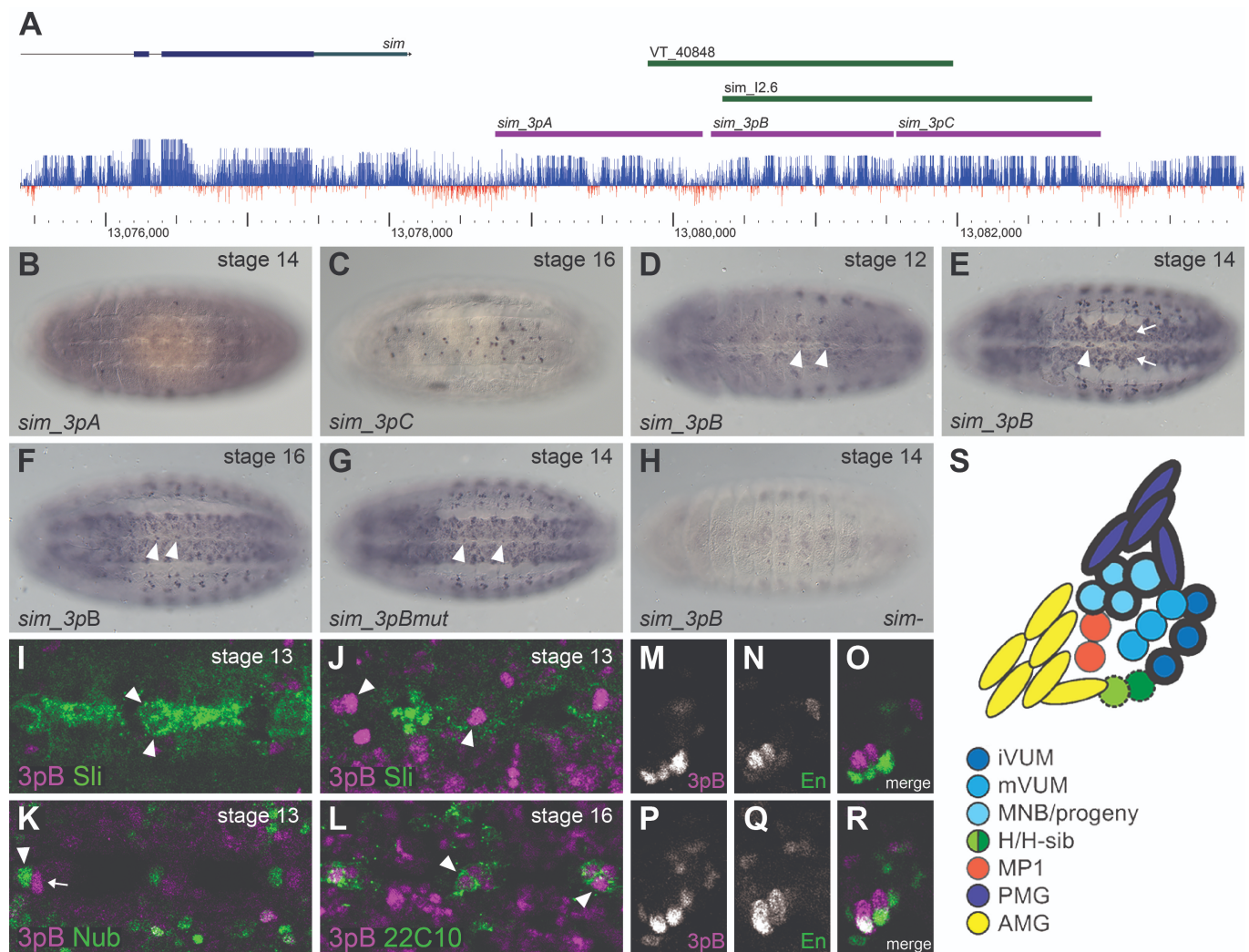


Figure 1. Characterization of putative midline enhancers in the *sim* 3' region:

(A) Schematic of the *D. melanogaster sim* downstream region. Genomic sequences chosen for testing are colored magenta; previously published regions overlapping *sim_3pB* are green. PhyloP sequence conservation across 27 insects lies at the bottom, with conserved sequence shown in blue and areas of low conservation shown in red. (B) The *sim_3pA* fragment fails to drive reporter expression in transgenic *D. melanogaster* embryos. Shown at stage 14. (C) Reporter expression driven by

sim_3pC at stage 16 in the CNS, outside of the midline. (D-F) Reporter expression driven by *sim_3pB* at stages 12 (D), 14 (E), and 16 (F) in the CNS and other tissues. Midline expression begins at the end of stage 12 and continues to at least stage 16, with a peak in stages 13 and 14. Arrowheads show midline staining. Arrows in (E) point to ectopic CNS staining. (G) Some *sim_3pB* activity persists with mutation of CME sites; midline activity is reduced while background expression is mostly unchanged. Stage 14 is shown. Arrowheads show positive midline cells. (H) *sim_3pB* activity in the entire CNS is essentially absent in a *sim* null environment; stage 14 is shown. Embryos in B-H are all oriented ventral side up with anterior to the left. (I-J) Co-labelling of the *sim_3pB* reporter (3pB; magenta) and Slit (Sli; green) shows no overlap with PMG or AMG at stage 13 (ventral view; arrowheads point to midline cells). (K) Co-labelling of *sim_3pB* (3pB; magenta) and Nubbin (Nub; green) shows no overlap with MP1 neurons or H-cell/H-cell sib at stage 13 (ventral view; arrow and arrowhead point to a *sim_3pB* positive cell and nubbin positive cell, respectively). (L) Co-localization of *sim_3pB* (3pB; magenta) and 22C10 (green) suggests *sim_3pB* expression in the mVUMs at stage 16 (ventral view; arrowheads point to midline cells with colocalized marker). (M-R) Sagittal views of midline segments co-stained for *sim_3pB* and Engrailed in stage 13 embryos. (M-O) *sim_3pB* is active in the mVUMs, which are not Engrailed-positive. (P-R) *sim_3pB* sporadically overlaps Engrailed expression in the iVUMs. (S) Cartoon of a midline segment at stage 13, sagittal view (adapted from (Zhang et al., 2011)). Cells with heavy borders express Engrailed.

Description

In Arthropods, development of the central nervous system relies heavily on the gene *single-minded* (*sim*; FlyBaseID: FBgn0004666) (Linne et al., 2012). *sim* encodes a bHLH-PAS transcription factor that acts as a key regulator in the development of the ventral midline (Crews et al., 1988; Nambu et al., 1991). In the fruit fly *Drosophila melanogaster*, *sim* is active throughout the whole of embryogenesis, with transcript first being detected at the cellular blastoderm stage and persisting until the end of embryo development (Thomas et al., 1988; Wheeler et al., 2006). Most of the noncoding DNA of the *sim* locus has been assayed for regulatory activity, and enhancers have been defined for early embryonic, larval, and adult stages of development (Freer et al., 2011; Hong et al., 2013; Jenett et al., 2012; Kvon et al., 2014; Markstein et al., 2004; Pearson et al., 2016; Sandmann et al., 2007). However, *sim* regulation is still not completely defined. For example, enhancers governing midline expression later in embryonic development have yet to be identified. Although much of the locus has already been characterized, we reasoned that a finer-scale mapping of specific regions might better resolve discrete regulatory activities and possibly discover unknown enhancers, improving our understanding of *sim* regulation.

In this study, we focus on the 3' end of *D. melanogaster sim*, where sequence comparison across 27 insects shows distinct blocks of high conservation (Fig. 1A). Using this conservation as a guide, we isolated three fragments—*sim_3pA*, *sim_3pB*, and *sim_3pC*—from the genome and characterized them by reporter gene assay in transgenic embryos. *sim_3pA* is negative in all embryonic tissues at all stages (Fig. 1B). *sim_3pC* shows limited activity in the central nervous system (CNS) of stage 16 embryos, but this expression is outside of the midline in regions ectopic to *sim* (Fig. 1C). Therefore, this element is likely not a true *sim* enhancer. Neither *sim_3pC* nor *sim_3pA* were explored further. *sim_3pB*, on the other hand, is active in the midline from the end of germband retraction (late stage 12) through at least stage 16, with the strongest signal at stages 13 and 14 (Fig. 1D-F, arrowheads). *sim_3pB* also displays extensive ectopic activity throughout the medial portion of the CNS, starting earlier at approximately stage 10 and continuing into stage 16 (Fig. 1E, arrows).

All known *sim* midline enhancers in *D. melanogaster* possess autoregulatory capability mediated by binding of Sim:Tango heterodimers to CNS Midline Elements (CMEs) (Pearson and Crews, 2014; Sonnenfeld et al., 1997; Wharton et al., 1994). Three matches to the CME consensus sequence, RCGTG (Hong et al., 2013; Sonnenfeld et al., 1997), are present in *sim_3pB*. However, mutation of these sites leads to only a modest reduction in reporter expression. Midline activity is slightly reduced, whereas background expression is mostly unchanged (Fig. 1G, arrowheads). Careful quantitative analysis will be required to confirm whether this difference in activity is significant. Surprisingly, when introduced into a *sim* null environment, *sim_3pB* activity is strongly reduced in the entire CNS (Fig. 1H). These contradictory results suggest that *sim_3pB* may be both directly and indirectly dependent on Sim for activity, although we cannot rule out additional direct Sim autoregulatory binding through non-canonical sites.

During early stage 11 in *D. melanogaster*, the embryonic midline consists of the anterior and posterior midline glia (AMG and PMG, respectively), the median neuroblast (MNB), and midline precursors (MP1, MP3, MP4, MP5, and MP6) (Wheeler et al., 2006). Toward the end of stage 11, the MP3 precursors divide into H-cell and H-cell sib while MP4-6 each give rise to one Ventral Unpaired Median motoneuron and one Ventral Unpaired Median interneuron (mVUM and iVUM, respectively). Each cell type expresses a specific set of genes that allows it to be uniquely identified (Wheeler et al., 2006). By co-staining *sim_3pB* embryos for the reporter gene and one of four known midline markers—Slit, Mab 22C10, Engrailed, and Nubbin—we were able to identify the midline cells in which *sim_3pB* is active. In stage 13/14 embryos, *sim_3pB* activity did not coincide with either Slit or Nubbin, which rules out midline glia (AMG/PMG), MP1 neurons, and H-cell/H-cell sib (Fig. 1I-

K). Co-localization with 22C10 in stage 16 embryos suggests activity in the mVUMs (Fig. 1L), which is confirmed by *sim_3pB* staining relative to Engrailed-positive cells at stage 13 (Fig. 1M-O, S). Additionally, sporadic activity is observed in the iVUMs (Fig. 1P-R, S). *sim* itself is not expressed in the mVUMs at stage 13 (Wheeler et al., 2006), but is it possible that activity begins weakly prior to stage 13 (in the MP4-6 precursor cells) and is already fading by the time they have divided. This could also explain the inconsistent activity seen in the iVUMs.

sim_3pB shares several characteristics with a recently discovered *sim* midline enhancer in the mosquito *Aedes aegypti*, *sim_5P3* (Schember et al., 2024). It is also dissimilar in key ways. Although both elements begin activity later than other known *sim* midline primordium enhancers, *sim_5P3* becomes active at the end of stage 9 and *sim_3pB* not until the end of stage 12 (with respect to its midline activity). Neither are active in the full set of midline cells and are instead restricted to the MP4 lineage, but whereas *sim_5P3* expression includes all the descendants of the initial MP4 equivalence group (PMG, MNB, the MP4-6 precursors and their VUM progeny), *sim_3pB* appears limited to the VUMs. Additionally, while neither element is directly autoregulated, they differ in their dependence on Sim protein. The *sim_5P3* sequence contains no CME sites and is fully functional in a *sim* null background. On the other hand, *sim_3pB* remains weakly active with the loss of three CME sites yet is not functional in a *sim* null background. Taken together, despite interesting similarities between these two regulatory elements, it cannot necessarily be concluded that *sim_3pB* is a functional counterpart to *A. aegypti sim_5P3*.

sim_3pB activity, though it includes *sim* positive cells, also extends to regions where *sim* is not expressed in the embryo. The sequence we isolated could be lacking repressors necessary for midline-only activity. Alternatively, *sim_3pB* may not act as an enhancer in its native context. Kalay et al. (2019) introduced the concept of “cryptic enhancers”—sequences that have the ability to drive gene expression in isolation but are repressed in their endogenous chromosomal environment. A cryptic enhancer might thus be located within a larger region that is negative for regulatory activity (Kalay et al., 2019). Such may be the case with *sim_3pB*. In two separate experiments, regions containing and extending beyond the *sim_3pB* sequence failed to drive reporter expression: *sim_I2.6*, which overlaps all but the first ~70bp of *sim_3pB* (Freer et al., 2011), and *VT40848*, which contains the entirety of the *sim_3pB* sequence (Kvon et al., 2014), were both reported to be inactive (Fig. 1A). Taking this into consideration, it is possible that *sim_3pB* is a cryptic enhancer for *sim*. An intriguing alternative possibility is that *sim_3pB* acts in its native location as a silencer element to prevent *sim* transcription in the mVUMs and cells outside of the midline, but manifests as an enhancer in an out-of-context reporter assay. This could explain the extensive activity observed in *sim*-negative cells. Further investigation will be necessary to test these various possibilities. Nonetheless, currently *sim_3pB* stands as the only identified embryonic *D. melanogaster sim* regulatory element with midline activity beginning at or after the late midline primordium stage. Continuing to explore the *sim* locus may reveal more enhancers governing midline activity at these later stages, providing a clearer picture of how *sim* is regulated throughout embryonic development in *D. melanogaster* and in Arthropods as a whole.

Methods

Reporter Constructs

sim_3pA and *sim_3pC* were isolated via PCR on genomic DNA from *D. melanogaster* strain OregonR (OrgR) and cloned into the *pattBnucGFP* vector (Kazemian et al., 2011). *sim_3pB* was synthesized as a gBlock (IDT, Coralville IA) and cloned into the *placZattB* vector (Bischof et al., 2013) using the NEBuilder HiFi DNA Assembly kit (New England Biolabs, Beverly, MA). To generate the *sim_3pBmut* line, MacVector (Apex, NC) was used to scan for Sim:Tgo binding sequences ACGTG (Sonnenfeld et al., 1997) and DDRCGTG (Hong et al., 2013). Three matches were found, and each site was mutated to GGATCC, as described by Wharton et al. (1994). The *sim_3pBmut* fragment was synthesized as a gBlock (IDT, Coralville IA) and inserted into *placZattB* using the NEBuilder HiFi DNA Assembly kit. Transgenic flies were generated by Rainbow Transgenic Flies (Camarillo, CA) by phiC31 recombination using line attP2 8622.

Sequence conservation

Sequence conservation was obtained from the UCSC Genome Browser (Raney et al., 2024) using the “phyloP 27 species” conservation track. Data in this track are obtained using phyloP (Pollard et al., 2010) and the following species: *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. biarmipes*, *D. suzukii*, *D. ananassae*, *D. bipunctata*, *D. eugracilis*, *D. elegans*, *D. kikkawai*, *D. takahashii*, *D. rhopaloa*, *D. ficusphila*, *D. pseudoobscura*, *D. persimilis*, *D. miranda*, *D. willistoni*, *D. virilis*, *D. mojavensis*, *D. albomicans*, *D. grimshawi*, *Musca domestica*, *Anopheles gambiae*, *Apis mellifera*, *Tribolium castaneum*.

Immunohistochemistry

Immunohistochemistry on *D. melanogaster* embryos was conducted using standard methods. Staining was done with primary antibodies mouse anti- β -galactosidase (1:500, Promega Z378) and rabbit anti-GFP (1:10000, Ab-cam ab290) then visualized

10/17/2024 - Open Access

using the Vectastain® Standard ABC kit (Vector Laboratories). Imaging was performed using a Zeiss Axioskop equipped with a Jenoptix Arctur camera. For fluorescent staining, primary antibodies were mouse anti-Slit C555.6D (1:10; Developmental Studies Hybridoma Bank), mouse anti-Nubbin 2D4 (1:10, DSHB), mouse anti-Engrailed 4D9 (DSHB), and Mab 22C10 (1:20, DSHB). Secondary antibodies were anti-mouse Alexa Fluor 488 (1:250, Molecular Probes, A11029) and anti-rabbit Alexa Fluor 647 (1:500, Molecular Probes, A21245). Fluorescent staining was visualized using a Leica SP8 confocal microscope.

Test for autoregulation

The *sim_3pB* construct was recombined onto a *sim* null chromosome (*sim*², Bloomington Stock Center stock #2055) and balanced with a *TM3* balancer containing *ftz-lacZ*. Loss of *sim* function was confirmed by complementation testing with a *sim*-null deficiency (*Df(3R)Exel6167/TM6B*, Bloomington Stock Center stock #7646).

Reagents

Plasmid	Features	Description	Available from
<i>pattBnucGFP</i>	white[+] SV40 3'UTR nucEGFP gene hsp70 minimal promoter MCS attB ampR	ϕC31-enabled vector containing <i>eGFP</i> under control of a minimal hsp70 promoter. Polylinker and <i>eGFP</i> cassette from pH-Stinger.	Authors, on request
<i>pLacZattB</i>	white[+] SV40 3'UTR lacZ gene hsp70 minimal promoter MCS attB loxP ampR	ϕC31-enabled vector containing <i>LacZ</i> under control of a minimal hsp70 promoter.	Drosophila Genomics Resource Center; stock 1421

Line	Genotype	Available from
<i>attP2</i> 8622	P{CaryP}attP2	Bloomington Drosophila Stock center; stock #8622
<i>sim</i> ²	<i>sim</i> [2] <i>kar</i> [1]/ <i>TM3</i> , P{ry[+7.2]=ftz-lacC}SC1, Sb[1] ry[RK]	Bloomington Drosophila Stock center; stock #2055
(<i>Df(3R)Exel6167/TM6B</i>)	w[1118]; <i>Df(3R)Exel6167</i> , P{w[+mC]=XP-U}Exel6167/ <i>TM6B</i> , Tb[1]	Bloomington Drosophila Stock center; stock #7646
<i>Dr/TFLZ</i>	<i>TM3</i> , P{ry[+7.2]=ftz-lacZ.ry[+]} <i>TM3</i> , Sb[1] ry[*]/ <i>Dr</i> [Mio]	Bloomington Drosophila Stock Center; stock #3218

Antibody	Animal and clonality	Description	Available from
Mab 22C10	Mouse; monoclonal	Raised against full-length <i>Drosophila melanogaster</i> Futsch sequence	Developmental Studies Hybridoma Bank
anti-Engrailed 4D9	Mouse; monoclonal	Antibody recognizes both the Engrailed and Invected gene products of <i>Drosophila</i> .	Developmental Studies Hybridoma Bank
anti-Nubbin 2D4	Mouse; monoclonal	Raised against recombinant Nub/Pdm sequence from <i>Artemia</i>	Developmental Studies Hybridoma Bank
anti-Slit C555.6D	Mouse; monoclonal	Raised against recombinant fusion protein containing the C-terminal region of <i>Drosophila</i> Slit, aa 1311-1480 (isoform A; Uniprot ID P24014-2)	Developmental Studies Hybridoma Bank
anti-GFP ab290	Rabbit; polyclonal		Ab-cam
anti-β-galactosidase	Mouse; monoclonal		Promega #Z378
anti-mouse Alexa Fluor 488	Goat; polyclonal		Molecular Probes #A11029
anti-rabbit Alexa Fluor 647	Goat; polyclonal		Molecular Probes #A21245
anti-mouse IgG (biotinylated)	Goat; polyclonal		Vector Laboratories
anti-rabbit IgG (biotinylated)	Goat; polyclonal		Vector Laboratories

Acknowledgements:

We thank Jack Leatherbarrow for assistance with cloning and fly genetics, Bill Reid for helpful advice and discussion, and Cassandra Rohrdanz for fly maintenance. Fly stocks were obtained from the Bloomington *Drosophila* Stock Center. Monoclonal antibodies used in this study were obtained from the Developmental Studies Hybridoma Bank. Confocal images were obtained using the Leica SP8 confocal microscope at the Jacobs School of Medicine and Biomedical Sciences Confocal Microscopy Core Facility.

References

- Bischof J, Björklund M, Furger E, Schertel C, Taipale J, Basler K. 2013. A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*. *Development* 140(11): 2434-42. PubMed ID: [23637332](#)
- Crews ST, Thomas JB, Goodman CS. 1988. The *Drosophila* single-minded gene encodes a nuclear protein with sequence similarity to the per gene product. *Cell* 52(1): 143-51. PubMed ID: [3345560](#)
- Freer SM, Lau DC, Pearson JC, Talsky KB, Crews ST. 2011. Molecular and functional analysis of *Drosophila* single-minded larval central brain expression. *Gene Expr Patterns* 11(8): 533-46. PubMed ID: [21945234](#)
- Hong JW, Park KW, Levine MS. 2013. Temporal regulation of single-minded target genes in the ventral midline of the *Drosophila* central nervous system. *Dev Biol* 380(2): 335-43. PubMed ID: [23701883](#)

- Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, et al., Zugates CT. 2012. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep* 2(4): 991-1001. PubMed ID: [23063364](#)
- Kalay G, Lachowiec J, Rosas U, Dome MR, Wittkopp P. 2019. Redundant and Cryptic Enhancer Activities of the *Drosophila* yellow Gene. *Genetics* 212(1): 343-360. PubMed ID: [30842209](#)
- Kazemian M, Zhu Q, Halfon MS, Sinha S. 2011. Improved accuracy of supervised CRM discovery with interpolated Markov models and cross-species comparison. *Nucleic Acids Res* 39(22): 9463-72. PubMed ID: [21821659](#)
- Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* 512(7512): 91-5. PubMed ID: [24896182](#)
- Linne V, Eriksson BJ, Stollewerk A. 2012. Single-minded and the evolution of the ventral midline in arthropods. *Dev Biol* 364(1): 66-76. PubMed ID: [22306923](#)
- Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A, Levine M. 2004. A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development* 131(10): 2387-94. PubMed ID: [15128669](#)
- Nambu JR, Lewis JO, Wharton KA Jr, Crews ST. 1991. The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* 67(6): 1157-67. PubMed ID: [1760843](#)
- Pearson JC, Crews ST. 2014. Enhancer diversity and the control of a simple pattern of *Drosophila* CNS midline cell expression. *Dev Biol* 392(2): 466-82. PubMed ID: [24854999](#)
- Pearson JC, McKay DJ, Lieb JD, Crews ST. 2016. Chromatin profiling of *Drosophila* CNS subpopulations identifies active transcriptional enhancers. *Development* 143(20): 3723-3732. PubMed ID: [27802137](#)
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20(1): 110-21. PubMed ID: [19858363](#)
- Raney BJ, Barber GP, Benet-Pagès A, Casper J, Clawson H, Cline MS, et al., Haeussler M. 2024. The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res* 52(D1): D1082-D1088. PubMed ID: [37953330](#)
- Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, Furlong EE. 2007. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev* 21(4): 436-49. PubMed ID: [17322403](#)
- Schember I, Reid W, Sterling-Lentsch G, Halfon MS. 2024. Conserved and novel enhancers in the *Aedes aegypti* single-minded locus recapitulate embryonic ventral midline gene expression. *PLoS Genet* 20(4): e1010891. PubMed ID: [38683842](#)
- Sonnenfeld M, Ward M, Nystrom G, Mosher J, Stahl S, Crews S. 1997. The *Drosophila* tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development* 124(22): 4571-82. PubMed ID: [9409674](#)
- Thomas JB, Crews ST, Goodman CS. 1988. Molecular genetics of the single-minded locus: a gene involved in the development of the *Drosophila* nervous system. *Cell* 52(1): 133-41. PubMed ID: [3345559](#)
- Wharton KA Jr, Franks RG, Kasai Y, Crews ST. 1994. Control of CNS midline transcription by asymmetric E-box-like elements: similarity to xenobiotic responsive regulation. *Development* 120(12): 3563-9. PubMed ID: [7821222](#)
- Wheeler SR, Kearney JB, Guardioli AR, Crews ST. 2006. Single-cell mapping of neural and glial gene expression in the developing *Drosophila* CNS midline cells. *Dev Biol* 294(2): 509-24. PubMed ID: [16631157](#)
- Zhang Y, Wheatley R, Fulkerson E, Tapp A, Estes PA. 2011. Mastermind mutations generate a unique constellation of midline cells within the *Drosophila* CNS. *PLoS One* 6(10): e26197. PubMed ID: [22046261](#)

Funding:

This work was funded by National Science Foundation grant IOS-1911723 to MSH.

Supported by National Science Foundation (United States) IOS-1911723 to MSH.

Author Contributions: Geyenna Sterling-Lentsch: investigation, writing - original draft, conceptualization. Marc S. Halfon: conceptualization, funding acquisition, project administration, supervision, writing - review editing.

Reviewed By: Anonymous

History: Received August 8, 2024 **Revision Received** October 7, 2024 **Accepted** October 10, 2024 **Published Online** October 17, 2024 **Indexed** October 31, 2024

Copyright: © 2024 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Sterling-Lentsch, G; Halfon, MS (2024). Identification of a novel downstream *single-minded* midline regulatory element in *Drosophila melanogaster*. microPublication Biology. [10.17912/micropub.biology.001317](https://doi.org/10.17912/micropub.biology.001317)