



Published in final edited form as:

Dev Biol. 2014 August 15; 392(2): 466–482. doi:10.1016/j.ydbio.2014.05.011.

Enhancer Diversity and the Control of a Simple Pattern of *Drosophila* CNS Midline Cell Expression

Joseph C. Pearson and Stephen T. Crews*

Department of Biochemistry and Biophysics, Integrative Program in Biological and Genome Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280

Abstract

Transcriptional enhancers integrate information derived from transcription factor binding to control gene expression. One key question concerns the extent of trans- and cis-regulatory variation in how co-expressed genes are controlled. The *Drosophila* CNS midline cells constitute a group of neurons and glia in which expression changes can be readily characterized during specification and differentiation. Using a transgenic approach, we compare the cis-regulation of multiple genes expressed in the *Drosophila* CNS midline primordium cells, and show that while the expression patterns may appear alike, the target genes are not equivalent in how these common expression patterns are achieved. Some genes utilize a single enhancer that promotes expression in all midline cells, while others utilize multiple enhancers with distinct spatial, temporal, and quantitative contributions. Two regulators, Single-minded and Notch, play key roles in controlling early midline gene expression. While Single-minded is expected to control expression of most, if not all, midline primordium-expressed genes, the role of Notch in directly controlling midline transcription is unknown. Midline primordium expression of the *rhomboid* gene is dependent on cell signaling by the Notch signaling pathway. Mutational analysis of a *rhomboid* enhancer reveals at least 5 distinct types of functional cis-control elements, including a binding site for the Notch effector, Suppressor of Hairless. The results suggest a model in which Notch/Suppressor of Hairless levels are insufficient to activate *rhomboid* expression by itself, but does so in conjunction with additional factors, some of which, including Single-minded, provide midline specificity to Notch activation. Similarly, a midline glial enhancer from the *argos* gene, which is dependent on EGF/Spitz signaling, is directly regulated by contributions from both Pointed, the EGF transcriptional effector, and Single-minded. In contrast, midline primordium expression of other genes shows a strong dependence on Single-minded and varying combinations of additional transcription factors. Thus, Single-minded directly regulates midline primordium-expressed genes, but in some cases plays a primary role in directing target gene midline expression, and in others provides midline specificity to cell signaling inputs.

© 2014 Elsevier Inc. All rights reserved.

*Corresponding author: Stephen T. Crews, steve_crews@unc.edu, Tel: 919-423-8236, Fax: 919-962-4296.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Drosophila; CNS; enhancer; midline; *Notch*; *single-minded*

Introduction

Genomic enhancers control transcription by integrating the contributions of multiple DNA-binding transcription factors, leading to modification of chromatin and recruiting core transcriptional machinery to promoter elements (Ong and Corces, 2011; Spitz and Furlong, 2012). One of the most interesting aspects of enhancer function concerns how coordinately expressed eukaryotic genes are regulated, an issue that dates back to the groundbreaking contributions by Britten and Davidson (1971). Does the same group of transcription factors control transcription of each co-expressed gene in the same manner, utilizing a highly-conserved binding site code, or is the situation more variable (Yáñez-Cuna et al., 2013)? In the case of *Ciona*, genes co-expressed in muscle utilize a common set of transcription factors with a variable combination of binding sites to control transcription (Brown et al., 2007). Similarly, in the *Drosophila* embryo, a shared set of transcription factors control gene expression in the developing mesoderm, yet different co-expressed genes can utilize different transcription factors and binding site architectures, arguing against a strict mesoderm code (Zinzen et al., 2009). Still, the coexpression of genes may be idiosyncratic and variable depending on the regulatory proteins involved and how during evolution their respective target genes came to be expressed in the same cell type. Since evolution is dependent to a substantial degree on the control of gene expression (Carroll, et al. 2001; Davidson. 2006), it is important to study a broad sampling of the gene regulatory landscape.

Study of the *Drosophila* CNS midline cells has a number of advantages for understanding enhancer function and gene regulation during development. Midline cell expression of a gene is particularly easy to identify due to the characteristic midline stripe of embryonic expression. The development of the midline cells is well understood, with many of the key regulatory proteins identified (Thomas et al., 1988; Watson et al., 2011; Watson and Crews, 2012; Wheeler et al., 2008). In particular, the *Drosophila single-minded (sim)* gene, which encodes a bHLH-PAS transcription factor, controls early development of the CNS midline cells (Nambu et al., 1991) by acting as a transcriptional activator that promotes the midline transcriptional program (Nambu et al., 1990) and indirectly represses the lateral CNS program (Estes et al., 2001). The Sim protein forms a heterodimer with the Tango (Tgo) bHLH-PAS protein to bind the DNA sequence ACGTG, called the CNS Midline Element (CME) (Sonnenfeld et al., 1997; Wharton et al., 1994). The CME is present and functional in nearly all midline enhancers studied to date (Apitz et al., 2005; Estes et al., 2008; Hong et al., 2013; Long et al., 2014; Pearson et al., 2012; Wharton et al., 1994), consistent with the possibility that Sim:Tgo directly activates most midline-specific gene expression.

Recent studies have promoted a model in which *sim* initially commits ectodermal cells to a single midline neuronal precursor fate followed by a series of signaling events that further diversify midline cell fates (Watson and Crews, 2012). During midline cell differentiation, *sim* continues to control transcription in midline glia (MG) (Estes et al., 2008; Wharton et

al., 1994) and may interact with the Notch signaling pathway, another important regulator of MG transcription and development (Wheeler et al., 2008). At the midline primordium stage, Notch signaling through its transcriptional effector, Suppressor of Hairless (Su(H)), controls MG and midline neuronal cell fate and transcription. This raises a number of issues. Does Notch signaling mainly influence transcription in midline cells by maintaining *sim* expression? Alternatively, does Notch signaling, via Su(H), directly regulate many midline-expressed genes, and does it control transcription in conjunction with Sim and other midline-expressed transcription factors? How do these Notch-dependent midline genes restrict expression to the midline, given the number of tissues utilizing Notch signaling in the embryo?

In this paper, we use transgenic approaches to identify and analyze enhancers that drive midline expression from a variety of genes that all show similar expression in midline primordium cells. These experiments reveal considerable mechanistic diversity in controlling their midline primordium expression. In one mode, including the genes *Ectoderm-expressed 3 (Ect3)*, *midline fasciclin (mfas)*, *Sema-1b*, and *Toll (Tl)*, there is a single enhancer that drives midline primordium expression. In another mode (*rhomboid (rho)*, *sim* autoregulation), midline primordium expression is a sum of at least two distinct enhancers. In all tested cases, Sim binding contributes to midline expression. However, mutation of canonical Sim:Tgo binding sites (CMEs) results in different outcomes depending on the enhancer. Whereas midline primordium expression is absent when CMEs are mutated in the *Sema-1b*, *sim* autoregulatory, and *Tl* enhancers, the effect on two enhancers of genes strongly influenced by cell-cell signaling (*rho*, *argos*) is relatively minor. However, mutation of a *rho* Su(H) binding site results in a severe decrease of expression. Additional cis-control elements were identified in the *rho* enhancer, suggesting a model in which *rho* midline primordium expression is dependent on Notch/Su(H) signaling, but Su(H) activation is insufficient to activate expression by itself. The CMEs and additional functional cis-control elements, including Sox protein motifs, provide midline specificity in conjunction with Su(H) activation and drive *rho* expression in the midline cells. The *argos* gene, controlled by EGF/Spitz (Spi) signaling (Schweitzer et al., 1995), is regulated by Pointed, the transcriptional effector of Spi signaling, with a contribution by Sim, again suggesting that a “Signal plus Sim” code is common for midline-expressed genes controlled by cell signaling. This resembles the “Notch plus Proneural” code proposed for Notch signaling in lateral inhibition of neural fates (Castro et al., 2005). In summary, there is considerable diversity in the activity of enhancers controlling a common pattern of midline primordium gene expression, complicating attempts to derive a common “cis-regulatory code” underlying this simple expression pattern.

Materials and methods

Drosophila strains and genetics

Drosophila strains used include: *sim3.7-Gal4* (Xiao et al., 1996), *UAS-tau-GFP* (Wheeler et al., 2006), *sim-3.7-Gal4 UAS-tau-GFP* (Wheeler et al., 2006), *sim-1.0-Gal4* (Freer et al., 2011), *UAS-mCD8-GFP(LL6)* (Lee and Luo, 1999), *sim^{H9}* (Hilliker et al., 1980), *UAS-sim* (Xiao et al., 1996), *prd-Gal4* (BL 1947), *Dl³* (Uemura et al., 1989), *UAS-Su(H).VP16* (Kidd

et al., 1998), *12xSu(H)bs-lacZ* (Go et al., 1998), *rho-1.0-GFP* (*rho PE-A-GFP*; kindly provided by J. Posakony), *UAS-aop.Act* (BL 5789), *UAS-pnt.P1* (BL 869), *TM3-ftz-lacZ* (BL 3218), and *nanos-phiC31;;attP2* (68A1-B2) (Groth et al., 2004).

Sequence analysis and bioinformatics

Enhancer alignments and identification of motifs were performed as described previously (Pearson et al., 2012). Locus schematics were modified from GenePalette (Rebeiz and Posakony, 2004), and enhancer schematics were modified from Twine (Pearson and Crews, 2013). BDTNP DNase hypersensitivity (Thomas et al., 2011), blastoderm transcription factor ChIP-chip data (Li et al., 2008), and Zelda ChIP-Seq data (Harrison et al., 2011) were accessed on the UCSC Genome Browser (genome.ucsc.edu; Meyer et al., 2013).

DNA constructs

All enhancer fragments were PCR-amplified and cloned into pENTR or pCR8 (Life Technologies), then were cloned into pMintgate (Jiang et al., 2010) or pBPGw-UCP (Wheeler et al., 2012) using Gateway LR Clonase II (Life Technologies). Site mutations were introduced by PCR-based site-directed mutagenesis and PCR-sewing. PCR Primers were synthesized by IDT or Operon, and are listed in Supplementary Table 1.

In situ hybridization and immunostaining

Embryo collection, in situ hybridization, immunostaining, and confocal imaging were performed as previously described (Kearney et al., 2004; Wheeler et al., 2006; Wheeler et al., 2008). Digoxigenin- or biotin-labeled antisense RNA probes for *GFP*, *Gal4*, and *lacZ* were PCR-amplified from plasmids and cloned into pCR2.1. Probes for *Drosophila* genes were generated from the following *Drosophila* Gene Collection clones: *Tl* (GH03720), *mfas* (GH11519), *Ect3* (HL01076), *Sema-1b* (GH03186), and *rho* (LD06131). Primary antibodies used were rabbit anti-GFP (Abcam), mouse anti-tau (Sigma), mouse anti-En (4D9) (Patel et al., 1989), and guinea pig anti-Sim (Pearson et al., 2012). Segments T3-A2 were analyzed for all sagittal single segment images.

Imaging and image analysis

Imaging was performed on a Zeiss Pascal confocal microscope. Brightness and minimum signal threshold of each confocal channel were adjusted for clarity and to remove non-specific noise using ImageJ (Abramoff et al., 2004), except in cases where images were taken with identical confocal settings for comparison purposes. Image stacks were rotated along the X- or Y- axis using the TransformJ plugin for ImageJ (<http://www.imagescience.org/meijering/software/transformj/>) to generate sagittal views of midline segments. Original confocal stacks are available on request.

Results

Common patterns of midline primordium gene expression

During embryogenesis, the embryonic CNS midline cells give rise to ~22 neurons and glia per segment that play important roles in neurodevelopmental and neurobiological processes

(Wheeler et al., 2006). The midline glia (MG) consist of two subpopulations: anterior midline glia (AMG) and posterior midline glia (PMG) (Wheeler et al., 2006). The midline neurons arise from 6 precursors: Midline Precursors 1,3,4,5,6 (MP1,3,4,5,6) and the median neuroblast (MNB) stem cell (Fig. 1; Wheeler et al., 2006). At the mesectodermal (stages 5–8) and early midline primordium stages (stages 9–10), the midline cells appear relatively uniform. In the late midline primordium (stage 11), midline cells begin to divide, migrate, and develop into midline neurons and MG. In this paper, we analyze 5 genes that are expressed in a similar manner in the midline primordium (Fig. 2D–S) to address questions regarding their similarities and dissimilarities in midline gene regulation.

The *sim* gene is prominently expressed in all midline cells at the mesectodermal (stages 5–8) and midline primordium (stages 9–11) stages (Fig. 1,2A–C,S). At stage 10, all five genes analyzed (*Ect3*, *mfas*, *rho*, *Sema-1b*, *Tl*) are co-expressed in all midline primordium cells in thoracic and abdominal segments (Fig. 2D,G,J,M,P), although *Ect3* and *mfas* expression is reduced in the head segments compared to the other genes. After stage 10, their patterns become more complex. At early stage 11 (s11(e)), each gene is expressed in MPs and MG (Fig. 2E,H,K,N,Q,T). While delaminating MPs sometimes show a burst of enhanced expression (for example, *rho*; white arrowhead, Fig. 2H), expression is generally reduced upon MP delamination and division (Fig. 2E,K,N,Q,T; red bracket). By late stage 11 (s11(l)), midline expression is clearly discontinuous: present in AMG, often PMG, but generally absent in delaminating MPs and their neuronal progeny (Fig. 2F,I,L,O,R,T). This pattern is different than that of Sim protein, since at stage 11, Sim protein is abundant in all midline primordium cells, including divided MPs (Fig. 2B,C; red bracket). Consequently, the discontinuous nature of the 5-gene expression pattern is not due solely to *sim* function. However, the modulation of expression resembles the positive influence of *Notch* as indicated by expression of *Notch* signaling reporter transgenes (Wheeler et al., 2008). These reporters also show strong activation in AMG, PMG, and MNB, but are weakly activated in midline MPs and their neuronal progeny. These data indicate that the 5 genes analyzed are expressed in the midline primordium in a similar manner: first in all midline cells and then in a common subset. Based solely on common expression patterns, *sim* may directly control expression of these genes in all midline primordium cells at stages 9–11(e) (Fig. 2T), but this expression is modulated during stage 11 by other inputs, including *Notch* signaling. Although not well appreciated, *sim* expression is also modulated, changing from all midline cells after stage 11 to a *Notch*-dependent subset (AMG, PMG, H-cell sib, iVUMs, and MNB progeny) in stage 12 and later embryos (Wheeler et al., 2008).

Both separable and composite enhancers control midline primordium expression

Transgenic analysis of putative cis-regulatory regions was utilized to determine how midline primordium expression is controlled in vivo. This approach can address whether a single enhancer integrates transcription factor inputs to drive midline primordium expression or whether expression is dependent on multiple, separable enhancer elements each capable of driving different aspects of midline expression.

sim—Previous data for *sim* indicated that multiple enhancers control midline primordium expression (Supplementary Fig. S1A). An enhancer responsive to dorsal-ventral patterning

transcription factors Dorsal, Snail (Sna), Twist (Twi), and Notch signaling initiate midline expression in the mesectoderm from the *sim* early promoter (P_e) (Fig. 3A; (Cowden and Levine, 2002; Kasai et al., 1998; Morel and Schweisguth, 2000; Zinzen et al., 2006). Expression is maintained in the midline primordium from stages 9–11 by two distinct autoregulatory enhancers, one adjacent to P_e (*sim-3.7*, which also contains the above-mentioned initiation element) (Nambu et al., 1991; Wharton et al., 1994) and the other adjacent to the *sim* late promoter (P_l; *sim-1.0*) (Supplementary Fig. S1A; (Freer et al., 2011; Muralidhar et al., 1993). These enhancers together provide uniform expression in all midline cells from stages 5–11, as indicated by GFP protein accumulation throughout the midline primordium (Fig. 3B,C; green). The P_e mesectodermal/midline primordium enhancer is located in intron 1 within the *sim-3.7* fragment and the P_l midline primordium enhancer is located in the 5'-flanking *sim-1.0* fragment. While a *sim-3.7-Gal4* transgene can drive expression in the mesectoderm (Fig. 3A), it ceases activity in most midline cells by stage mid 11 (Fig. 3B; magenta). In contrast, *sim-1.0* activates expression in all midline cells starting at stage 9, and continues activity in all midline cells throughout stage 11, as detected by *Gal4* transcripts (Fig. 3C; magenta). Thus, the *sim* gene is an example of multiple enhancers functioning together to control levels of midline primordium expression, each active throughout the mesectoderm/midline primordium at successive stages. Through much of the stages 9–11 interval, their expression completely overlaps in all midline primordium cells, but by mid stage 11 (s11(m)), *sim-1.0* predominates as the *sim-3.7* enhancer fades. However, it is interesting that both *sim* enhancers are off after stage 11, indicating active processes that discontinue autoregulatory-induced transcription. How this is achieved at the cis-control level is unknown.

Sema-1b—Of the genes analyzed, *Sema-1b* most closely resembles *sim* due to its uniform expression in the mesectoderm from stages 5–8 while continuing to be expressed in midline primordium cells from stages 9–11 (Fig. 2D–F,S). Testing genomic fragments from the *Sema-1b* locus for midline expression (Supplementary Fig. S1B), we identified a single 580 bp fragment from the first intron that recapitulated *Sema-1b* mesectodermal (Fig. 3D) and midline primordium (Fig. 3E,F,H; green) expression, including a reduction in newly formed neurons (Fig. 3F; magenta). Like *sim*, the mesectodermal pattern of *Sema-1b* suggested that Dorsal, Sna, Twi, and Notch potentially initiate its expression at stage 5. Indeed, BDTNP ChIP-chip data for Dorsal, Sna, and Twi detected peaks centered on *Sema-1b-580* (Supplementary Fig. S1B; MacArthur et al., 2009). Two Twist consensus binding sites, T1 and T2 (CAYRTG) (Zeitlinger et al., 2007), were identified and separately mutated (the T2 site overlaps a CME; Sonnenfeld et al., 1997; Wharton et al., 1994) (Fig. 3G). In addition, a perfectly conserved E-box sequence (CANNTG) that could potentially bind Twi or another bHLH protein was present. Mutation of T1 along with the E-box site (mut-E-T1) had no effect on expression (Fig. 3G,I), whereas mutation of the T2/CME site in combination with two other CMEs (mut-CME-T2) resulted in an absence of *Sema-1b-580* expression (Fig. 3G,J). In contrast, mutation of two conserved Su(H) YRTGDGAA consensus sites (mut-Su(H)) (Barolo et al., 2000) did not alter *Sema-1b-580* expression (Fig. 3G,K). As mentioned, mutation of the three CMEs (*Sema-1b-580-mut-CME-T2*), which also included mutation of an overlapping Twi site, resulted in loss of both mesectodermal and midline primordium expression. However, analysis of homozygous *sim* null mutant embryos

indicated that loss of *sim* affects only midline primordium expression (stages 9–11) (Fig. 3N,O) but not mesectodermal expression (stages 5–8) (Fig. 3L,M). This result conflicts with a report that *Sema-1b* expression is *sim*-independent (Khare et al., 2000). This previous observation may have examined only early, *sim*-independent expression of *Sema-1b*. Our interpretation of these results is that Sim:Tgo directly regulates *Sema-1b* midline primordium expression, and Twi controls mesectodermal expression in conjunction with Dorsal and Sna. The absence of *Sema-1b-580* mesectodermal expression when all 3 CMEs were mutated is due to the loss of Twi binding (and not Sim:Tgo).

In summary, a single enhancer that is directly regulated by Sim:Tgo controls both *Sema-1b* mesectoderm and midline primordium expression, while *sim* utilizes at least three enhancers to achieve a similar pattern. *Sema-1b* expression is reduced in a subset of midline cells at stage 11, even though Sim protein is present. Thus, the influence of *sim* on *Sema-1b* expression is lost, just as with *sim* autoregulation. While the stage 11 pattern resembles *Notch* midline activity, the potential regulation of *Sema-1b* by *Notch* is not direct, since mutation of the two consensus Su(H) binding sites had no effect on transcription.

mfas|Ect-3—The *mfas* (Hu et al., 1998) and *Ect3* genes are closely linked, separated by 6.0 kb, and they lie in a head-to-head orientation (Supplementary Fig. S1C). There are two fragments on either side of the *mfas* promoter that drive expression in subsets of midline cells: *mfas-590* drives sporadic expression in PMG and MNB and its progeny starting at stage 13 (data not shown), while *mfas-780* drives early expression in MP4–6, as well as sporadic expression in AMG at stage 10 (Fig. 3P). The expression controlled by these enhancers was not sufficient to recapitulate the midline primordium expression of *mfas* during stages 10–11 (Fig. 2M–O,S). One fragment 5' to the *Ect3* promoter and another fragment from intron 1 (Supplementary Fig. S1C) are also active in midline cells: the intronic *Ect3-1.8* fragment promotes strong expression in all AMG starting at stage 12 (data not shown), while *Ect3-3.2* promotes expression in all midline primordium cells during stages 10–11 (Fig. 3Q; green), although there was a noticeable reduction in GFP protein accumulation in mVUM4 and iVUM4 (arrowhead), similar to endogenous *Ect3* expression (Fig. 2R). When *Ect3-3.2 GFP* expression is examined by in situ hybridization, we observed a reduction in expression in delaminated MPs (Fig. 3R; magenta), similar to endogenous *mfas* and *Ect3* expression (Fig. 2N,O,Q,R). Since only a single pan-midline primordium enhancer was detected in the regions of *mfas* and *Ect3* that can recapitulate endogenous expression of both genes, we suggest that the *Ect3-3.2* enhancer is shared between *mfas* and *Ect3*.

Tl—The *Tl-950* fragment drives midline reporter gene expression from stages 9–13 (Wharton and Crews, 1993). It contains 4 CMEs that are required for midline expression (Wharton et al., 1994). We examined *Tl-950* in greater detail, observing that it is expressed in all midline primordium cells beginning at stage 9 until stage 11 (Fig. 3S; green shows GFP accumulation from earlier stages). However, during stage 11 *Gal4* transcription is restricted from delaminating MPs and the posterior cells of each segment (Fig. 3S; magenta). Thus, the *Tl-950* fragment has a single midline enhancer capable of exactly recapitulating endogenous expression of *Tl* (Fig. 2J – L).

Summary—Gene expression in the midline primordium is characterized by expression in all midline cells followed by alterations in the expression pattern as midline cells begin to divide, migrate, and differentiate. Transgenic enhancer analysis of these 5 genes (and *rho*; see below) has successfully identified enhancer elements that recapitulate midline primordium expression, revealing that either a single enhancer or two enhancers that function together can generate the simple midline primordium expression pattern. In addition, two genes may share the same enhancer. The genes analyzed maintain expression in MG but not midline neurons, and this expression is controlled by the identified enhancers. Thus, although *Sim* is present in all midline cells through stage 11, its presence is unable to drive expression in midline neurons, indicating both a limitation of its activity and the existence of mechanisms that differentially control midline glial and neuronal expression.

Multiple enhancers generate the rhomboid midline primordium expression pattern

The *rho* gene encodes a protease involved in *Drosophila* Spi signaling and plays an important role in generating a diffusible midline Spi signal (Golembo et al., 1996; Kim and Crews, 1993). The *rho* gene is expressed throughout the midline from stages 6 to early stage 11 (Fig. 2G,H,S; Fig. 4A–D). Following a transient burst of expression in delaminated MPs before division (Fig. 4B,B'; yellow and white arrowheads), post-mitotic midline neurons show reduced expression at mid and late stage 11 (Fig. 4C,C arrows; D,D' bracket), while AMG and PMG maintain expression. There exist multiple *rho* midline enhancers (Fig. 4I, Supplementary Fig. S1D), and we initially focused on a previously identified enhancer that controls midline primordium expression (*rho-365*; Ip et al., 1992; Zelzer and Shilo, 2000). The *rho-365-GFP* transgene promoted only weak and sporadic *GFP* in the midline at stages 6–8 (not shown), resolving to expression in cells underneath the recently delaminated MP4 (Fig. 4E, arrowhead) during stages 9–10. These cells could be either MG or midline neural precursors (MPs or MNB) yet to delaminate. Expression progressed posteriorly in each midline segment (Fig. 4F, blue bracket), then into the anterior-most cells of each segment (Fig. 4G, white bracket). However, after each MP delaminated during stage 11, *GFP* expression ceased in that cell and its progeny (Fig. 4H, red bracket), similar to endogenous *rho*. Thus, *rho-365* progressively expands in midline cells during stages 10–11. This differs from other midline primordium enhancers that show uniform midline expression prior to stage 11, then lose expression in neurons after delamination.

While *rho-365* expression mirrored much of the spatial dynamics of *rho* expression during stage 11, it did not recapitulate the burst of expression seen occasionally in the MPs (compare Fig. 4B,B' to F,J), nor earlier mesectoderm or midline primordium expression. This lack of expression was not due to a partial enhancer, as similar results were observed with a larger fragment, *rho-2.2* (Fig. 4I,K, Supplementary Fig. S1D,S2), which includes neuroectodermal and tracheal enhancers and promoter sequences (Ip et al., 1992).

To find additional *rho* regulatory sequences, we used annotations of *Drosophila* embryonic DNaseI hypersensitive regions (Thomas et al., 2011) and noticed open chromatin regions at the: (1) *rho* promoter, (2) *rho-365*, and (3) an upstream region with enhancer activity in larval wing discs (Reeves and Posakony, 2005) (Fig. 4I). We tested this upstream enhancer (*rho-1.0*; a gift from J. Posakony) for embryonic activity, and observed sporadic expression

in all midline cells during stages 10–11, with the most consistent expression in the neurons derived from MP3 and MP4 (Fig. 4L), the cells with the most consistent “burst” of endogenous *rho* expression. Thus, together, the *rho-365* and *rho-1.0* fragments recapitulate the endogenous pattern of *rho* midline primordium expression during stage 11. The *rho-1.0* fragment has a single CME, and a subfragment lacking this site (*rho-785*, Fig. 4I) failed to drive midline expression, indicating essential sequences in the deleted region (data not shown). The larger *rho-2.3* fragment has a total of 4 CMEs and drives broader, more variable, expression in all midline cells compared to *rho-1.0* (Fig. 4M), but did not drive strong mesectoderm expression. In summary, multiple enhancers, including the *rho-1.0* enhancer, the *rho-365* enhancer, and an unidentified mesectodermal enhancer collaborate during stages 6–11 to drive dynamic midline expression of *rho*.

Multiple conserved cis-regulatory motifs reside within the *rho-365* midline enhancer

To identify potential sequences in *rho-365* required for midline expression, we compared the *rho-365* sequence to orthologous regions from all sequenced *Drosophila* species, analyzing these sequences for conserved motifs matching known midline regulatory sites (Estes et al., 2008; Fulkerson and Estes, 2010; Ma et al., 2000; Pearson et al., 2012; Wharton et al., 1994) and novel over-represented motifs (Fig. 5A). Two CMEs, both highly conserved in *Drosophila* species, are within the *rho-365* interval (also noted in Hong et al., 2013; Zelzer and Shilo, 2000; pink C1, C2). The second CME (C2) is not present in *D. pseudoobscura* or *D. persimilis*, but a potential compensatory site arose promoter-distal to the first CME (C1). Overlapping the C1 CME is a consensus high-affinity site for Su(H) (blue). In the conserved block promoter-distal to the first CME is a sequence ATGCGTG (Fig. 5A, Motif-A; magenta) that is ATACGTG in the distantly related *virilis* group, including *D. virilis* (note that ACGTG matches the CME). In all species where Motif-A contains GCGTG, a Sox-family binding site (ACAATG) is adjacent (Fig. 5A; green); the *Drosophila* Sox factor Dichaete regulates midline expression of several genes (Estes et al., 2008; Ma et al., 2000; Sanchez-Soriano and Russell, 1998) and Dichaete ChIP-chip experiments show a peak encompassing *rho-365* (Negre et al., 2011; Thomas et al., 2011). There is also a conserved TAGteam consensus binding site (YAGGYAR; purple), recognized by Zelda, a transcription factor shown to regulate midline expression of *link* (Pearson et al., 2012). Zelda ChIP-Seq detected binding to the *rho-365* region at stages 10–11 (Harrison et al., 2011). Finally, there is a conserved repeat sequence (GATTTAYGAWG; tan). To determine the contributions of these sites to *rho-365* midline expression, each site was mutated and the altered sequences tested in vivo (results summarized in Fig. 5B).

CME and Motif-A sequences synergistically control *rho* enhancer expression

It was previously reported that a *rho-lacZ* enhancer trap drove low levels of expression in homozygous *sim* null mutant embryos (Nambu et al., 1990). We observed similar results for *rho-365*; GFP levels were strongly reduced, but present, in the midline cells of *sim* mutant embryos (Fig. 6A,B). However, endogenous *rho* transcripts were not detected in midline cells in *sim* mutants at stages 10–11 (data not shown), suggesting that trace *rho-lacZ* and *rho-365* GFP were due to early, *sim*-independent expression in the mesectoderm. Thus, *sim* is required for midline primordium expression of *rho*. To confirm that *sim* directly regulates *rho* midline expression through *rho-365*, we mutated the two CMEs, both individually

(CME1, CME2) and together (CME12) (Fig. 5B) and tested the mutated transgenes in vivo. When both CMEs were mutated together, *rho-365* expression was only slightly reduced in 2–3 MG in the middle of each segment (compare Fig. 6C,C to Fig. 4J). We note that Hong et al. (2013) mutated the same sequences and observed a much greater reduction in expression; this could be due to a number of factors including differences in the mutated sequences, transformation systems, or staining conditions. We observed a similar result when CME1 alone was mutated, whereas mutation of CME2 alone had no effect. The lack of the strict requirement of the CMEs for *rho-365* expression, contrasted with *sim* mutants that have a much stronger effect, suggested that *sim* controls *rho-365* expression in part indirectly or via non-canonical Sim binding sites.

Motif-A consists of the sequence, ATGCGTG, that itself contains a sequence (GCGTG) that differs from a Sim:Tgo binding site (ACGTG) by a single base pair change. DNA binding experiments examining the in vitro DNA specificity of *Drosophila* Sim partnered with mammalian Arnt (ortholog to *Drosophila* Tgo) (Swanson et al., 1995) showed binding to GCGTG, although most other experiments have described ACGTG as being the unique binding site for Sim:Tgo or Sim:Arnt (see Discussion). In *D. virilis* and other distant *Drosophila* relatives of *D. melanogaster*, the orthologous Motif-A sequence is ACGTG (Fig. 5A) instead of GCGTG. While this sequence is recognized by both Spineless (Ss):Tgo (Emmons et al., 1999) and Dysfusion (Dys):Tgo bHLH-PAS proteins (Jiang and Crews, 2007), neither *ss* nor *dys* is expressed in embryonic midline cells as detected by in situ hybridization, immunostaining, or RNA-seq on purified midline cells (Fontana and Crews, 2012).

It is of note that Motif-A (ATGCGTG) also resembles a part of a POU homeodomain POU-specific box DNA binding site (ATGC), and the Ventral veins lacking (Vvl) POU homeodomain protein (binding site WWATKMR; Fly Factor Survey; Zhu et al., 2011) has been implicated in controlling midline transcription (Estes et al., 2008; Ma et al., 2000; Sanchez-Soriano and Russell, 1998). Mutation of the POU-homeobox motifs in a midline enhancer from the *wrapper* gene did not reduce expression (Estes et al., 2008), but was required for MG expression of a *vvl* autoregulatory midline-specific enhancer (Certel et al., 1996). Expression of *rho-365-Unmut* or *rho-365-CME12* was not affected in a *vvl* mutant (data not shown), suggesting that Vvl is not the sole regulator binding Motif-A; this is consistent with previous observations that examined expression of a larger fragment containing *rho-365* in *vvl* mutants (Zelzer and Shilo, 2000). One caveat to this interpretation is that other works showed only a small reduction of *slit* MG gene expression in *vvl* mutants except in the context of a *Dichaete* mutant (Ma et al., 2000; Sanchez-Soriano and Russell, 1998).

We tested the requirement of Motif-A for midline expression, alone and in conjunction with CMEs. Mutating Motif-A alone caused a reduction in GFP within central MG, similar to mutating both CMEs (Fig. 6D,D'). However, mutation of Motif-A and both CMEs together (Fig. 6E,E') caused a dramatic reduction in GFP protein in all cells, with trace expression remaining only in posterior cells with the strongest *Notch* signaling: MP6, MNB, and PMG (Wheeler et al., 2008). Thus, the Motif-A binding site synergizes with the CMEs to activate expression in many midline cells. Changing the Motif-A GCGTG to ACGTG CME along

with mutation of the two CMEs (Motif-A>CME, CME12) (Fig. 6F,F') did not abolish expression like the Motif-A CME12 mutation. Instead, expression was similar to CME12. Since it did not affect expression, it is consistent with Sim:Tgo being able to bind both ACGTG and GCGTG (Motif-A), although since the single G>A change may also not have affected binding of some other transcription factor.

To provide additional insights into whether Motif-A is recognized by Sim:Tgo in *rho-365-GFP*, we ectopically expressed *sim* in the lateral ectoderm using *prd-Gal4* (Estes et al., 2001), causing stripes of Sim protein to be produced in alternating segments (Fig. 6G; bracket) and assayed *rho* and various *rho-365* reporters. Ectopic *sim* expression in the ventral ectoderm is sufficient to cause ectopic expression of midline-specific genes, including *rho* (Fig. 6H) (Estes et al., 2001; Nambu et al., 1991; Zinzen et al., 2006). *rho-365* GFP expanded in a similar manner to *rho* (Fig. 6I), confirming that *sim* is genetically upstream of *rho-365*. *rho-365-CME* GFP also expanded, but with lower expression levels (Fig. 6J). Similar results were obtained with *Motif-A>CME CME12* GFP (Fig. 6K). In contrast, *rho-365-Motif-A* GFP did not expand laterally (Fig. 6L), although midline expression was maintained. This result adds additional complexities to *rho* midline expression, indicating the presence of a factor, in addition to Sim, that functions in the midline but not lateral CNS. The function of this factor is compensated by the activity of the Motif-A site, but in the absence of Motif-A, its function is revealed. This experiment also indicates that even if Sim:Tgo binds both Motif-A and the canonical Sim sites, the Motif-A (ATGCGTG) and Sim 12 (ACGTG) binding sites are not functionally equivalent.

Notch signaling directly regulates rhomboid midline expression

Notch signaling regulates multiple aspects of midline development and transcription during stage 11 (Wheeler et al., 2008), although no direct transcriptional targets have been identified. The presence of a conserved, high-affinity Su(H) binding site in *rho-365* (Fig. 5A) suggested that *Notch* signaling directly regulates *rho* midline expression. Furthermore, both endogenous *rho* gene expression and *rho-365-GFP* expression resembled expression of a *Notch* reporter, *12xSu(H)bs-lacZ* (Wheeler et al., 2008), as did *rho-365-CME12-GFP* (Fig. 7A'-D'), which we compared directly to *12xSu(H)bs-lacZ* transcription (Fig. 7A-D). The results indicated that: (1) both reporters are strongly expressed in the same midline cells underneath the delaminated MP4 at stage 10 (Fig. 4E, Fig. 7A,A'; white arrowheads), (2) both expand posteriorly in undifferentiated MPs, MNB, and PMG in early and mid stage 11 embryos (Fig. 4F, Fig. 7B,B',C,C; blue brackets); (3) both are present in AMG, but *12xSu(H)bs-lacZ* expression and *rho-365-CME12-GFP* are lower than *rho-365-GFP* (Fig. 7C,C; gray brackets; compare to Fig. 4G), and (4) expression of both reporters remains high in PMG and MNB but low elsewhere, including midline neurons and AMG (Fig. 7D,D'; beneath red brackets; compare to Fig. 4H). This pattern of activation explains the reduced accumulation of GFP protein in a subset of AMG in *rho-365-CME12-GFP* embryos (Fig. 6C, compare to Fig. 4J); the CMEs are required for maintaining transcription in midline cells with transient *Notch* signaling.

To further explore the role of *Notch* signaling in *rho-365* regulation, a transgene encoding a constitutively-active form of Su(H), *Su(H).VP16* (Kidd et al., 1998), was expressed

throughout the midline from stages 5–11 using *Sema-1b-580-Gal4*. This resulted in strong accumulation of *rho-365* GFP (Fig. 7E), including the presence of GFP in cells that normally do not express *rho-365-GFP*, such as MP1 (green arrowhead), MP3 (yellow arrowhead), and MP4 (white arrowhead). In contrast, levels of *rho-365* GFP were not significantly affected when *UAS-Su(H).VP16* was driven by *sim-1.0-Gal4* (Fig. 7F); *sim-3.7-Gal4* expanded GFP into MP1, MP3, and MP4 at a lower frequency (data not shown). The differences in the results with different midline drivers are likely due to the strength and timing of *Gal4* transcription; in particular, the later and weaker expression of *sim-1.0* compared to *Sema-1b-580*. This indicates that the timing and/or magnitude of *Notch* signaling are critical in driving midline expression.

Consistent with a regulatory role for *Notch* signaling, *rho-365* GFP accumulation was greatly reduced in embryos mutant for *Delta*, which encodes a *Notch* ligand (Fig. 7G). Finally, mutating a single nucleotide in the Su(H) binding site, previously shown to abolish binding (Bailey and Posakony, 1995), also strongly reduced midline GFP (Fig. 7H). These results indicate that *Notch* signaling and Su(H) directly activate *rho* expression in the midline primordium. However, other factors must restrict *rho* expression to midline cells, because *rho-365* is midline-specific, whereas *Notch* signaling is active throughout the embryo.

The relative positions of the Su(H) and Sim:Tgo binding sites influence rhomboid enhancer function

Sim:Tgo binding site CME1 overlaps the Su(H) binding site, which raises interesting possibilities regarding interactions between these transcription factors. One hypothesis is that competition between Sim and Su(H) may explain the *rho-365* expression dynamics at stage 11. For example, Su(H) acts as a repressor (Su(H)⁻) in the absence of *Notch* signaling (Barolo et al., 2000; Morel and Schweisguth, 2000); if Sim:Tgo binds to the shared sequence, formerly bound by Su(H) activator (Su(H)⁺) that was present during transient *Notch* signaling, it could compete with Su(H)^R to maintain expression. This model is consistent with the loss of *rho-365* MG expression when CMEs are mutated (*rho-365-CME12-GFP*) after the period of *Notch* signaling (Fig. 7C',D'). To test this model, we separated CME1 from the Su(H) site in the context of *rho-365*. This model predicts that if the CME1 and Su(H) sites are separated, then Su(H) will be more difficult to displace and *rho-365* expression levels will be reduced or delayed. First, we duplicated a 10 bp sequence containing both the CME1 and Su(H) sites, which slightly increased GFP levels but had no effect on the midline expression pattern (Fig. 7I,J). This control provided a platform for testing the significance of site position, as described below. Mutating either the first (Fig. 7K) or second pair (Fig. 7L) of CME/Su(H) sites had only small effects on midline expression, confirming that the absolute position of the shared CME/Su(H) site was flexible. Mutating the first Su(H) site along with the second CME site maintained wild-type expression (Fig. 7M), demonstrating that competition for binding did not control the *rho-365* expression pattern. Even flipping the orientation of the CME did not affect expression (Fig. 7N). However, mutating the first CME site and the second Su(H) site, which places the Su(H) site 5' to the CME, dramatically reduced expression levels (Fig. 7O,P). Thus, the relative positioning of the CME and Su(H) sites is important, and altering

the native positioning may interrupt interactions with nearby transcription factors. However, the overlap between the CME and Su(H) sites is unlikely to be important and more likely coincidental (Lusk and Eisen, 2010).

Sox factor binding sites and a novel repeated motif contribute to rhomboid enhancer function

Several additional sequences are present in *rho-365* (Fig. 5A) that are conserved and correspond to binding sites for known midline regulators. Both the TAGteam (YAGGYAR) and Sox (ACAATG) binding site sequences are similar to known binding sites for Zelda and Dichaete, respectively. Both of these proteins control midline transcription, and ChIP experiments detected binding by both factors in multiple locations within the *rho* locus, including *rho-365*. In addition, a repeated sequence (consensus GATTTAYGAWG), discovered by WinDotter analysis (Sonnhammer and Durbin, 1995), was found in two conserved positions, which is unlikely to occur by chance in such a short sequence.

To test the possible function of these sites on *rho-365* midline expression, each site was mutated and analyzed in vivo (Fig. 5B). The single site matching the TAGteam (Zelda consensus) sequence was mutated, but no alteration in GFP levels was observed in *rho-365-TAGteam* (Fig. 8A,A'). Similarly, no change in expression was observed with *rho-365-CME12-GFP* in a *zelda* null mutant strain (data not shown).

Several Sox family genes are expressed in the midline at stage 11, including *Dichaete*, *Sox-Neuro* (*SoxN*), and *Sox14* (Fontana and Crews, 2012; Ma et al., 2000; Sanchez-Soriano and Russell, 1998). Mutants for *Dichaete* are defective in MG transcription (Ma et al., 2000; Sanchez-Soriano and Russell, 1998), and the *wrapper* enhancer requires Sox consensus sites for robust midline expression (Estes et al., 2008). The single match to the Sox consensus in *rho-365* is between the Motif-A sequence and CME1; however, this site is present only in species where Motif-A contains GCGTG. With species such as *D. virilis*, that have a CME instead of Motif-A, there is no Sox site. This Sox site was tested for a role in *rho-365* function alone and in combination with both CMEs mutated. When the Sox site was mutated alone, expression was abolished in nearly all midline cells, but approximately one cell per segment maintained wild-type levels (Fig. 8B,B'). Mutating both CMEs along with the Sox site abolished detectable GFP (Fig. 8C,C). However, no changes in *rho* expression were observed in a *Dichaete* mutant. Thus, another transcription factor, not Dichaete, may bind to this sequence, or Dichaete's genetic function is compensated by other transcription factors, as previously shown (Buescher et al., 2002; Ma et al., 2000; Overton et al., 2002; Sanchez-Soriano and Russell, 1998).

We tested the requirement of the Repeat sequences, by mutating each site individually, as well as together. Mutation of either site by itself had little effect on expression (data not shown), but mutating both together dramatically reduced expression (Fig. 8D,D'). There was a low level of expression in all midline cells except the progeny of MP1, MP3, and MP4, suggesting that the unknown factor recognizing this sequence acts as a general activator without providing patterning information. There was relatively little sequence conservation in the 5' region of *rho-365*, suggesting an absence of additional important transcription factor binding sites in this region. Consistent with this observation, deletion of an *AvaII*

fragment that removes the 5' 72 bp (Fig. 5B) showed no alteration of midline expression, indicating no essential elements in that region.

Single-minded and Pointed directly control *argos* expression

Another midline-expressed gene regulated by intercellular cell signaling is *argos* (*aos*), which is a downstream target of Spi signaling in MG (Stermerdink and Jacobs, 1997). *Aos* inhibits Spi signaling in a negative feedback loop. The Pointed (Pnt) ETS transcription factor is the transcriptional effector of Spi signaling, and *aos* MG expression is absent in *pnt* mutants (Stermerdink and Jacobs, 1997). We identified a fragment, *aos-0.5*, in intron 1 (Supplementary Fig. S1E) that drives reporter gene expression in a subset of MG (Supplementary Fig. S3A,B), similar to endogenous *aos*; however, *aos-0.5* is not expressed in all cells that are responsive to Spi signaling (Gabay et al., 1997). Expression of this fragment responds to *spi* signaling, since expression is enhanced when *pnt* is overexpressed in midline cells (Supplementary Fig. S3C) and is greatly reduced when *pnt* function is inhibited by midline expression of *anterior open* (*aop*; *yan*), which encodes a competitive inhibitor of Pnt (Supplementary Fig. S3D). The *aos* gene has 1 CME, which is well conserved in *Drosophila* species (Supplementary Fig. S3E,F). When the CME was mutated, expression was delayed and slightly reduced in a small number of segments compared to wild-type, but was still present in MG (Supplementary Fig. S3F–H,J). This indicates, that similar to *rho*, *sim* directly controls *aos* expression, but the role of the CME is relatively weak. There are 9 potential ETS transcription factor binding sites (ETS consensus GGAW; (Wei et al., 2010) in *aos-0.5*. We mutated 4 ETS sites together: one high-affinity site (SMGGAWRY) (Wildonger et al., 2005; Zhu et al., 2011) and 3 well-conserved GGAW sequences. This resulted in a reduction or loss of expression (Supplementary Fig. S3I,J), and suggests that Pnt directly regulates *aos* expression. Since *aos* expression is absent in *pnt* mutant embryos, it is possible that the additional ETS sites that were not mutated also contribute to expression. These data provide additional evidence that midline-expressed genes responding to cell signaling inputs, such as Notch and Spi, utilize lineage-specific regulators, such as Sim, to restrict their response to the midline cells.

Discussion

Midline primordium expression is generated by diverse enhancer arrangements

We analyzed fragments from 5 genes, identified in this and previous studies, that recapitulate the midline primordium expression of each gene, including early expression in all midline cells followed by down-regulation in neurons. Detailed characterization of development in the CNS midline cells during differentiation from midline primordium into neurons and glia (Wheeler et al., 2006; Wheeler et al., 2008) allowed cell-level comparisons between enhancers with broadly similar midline expression. This analysis revealed considerable diversity in the architecture of enhancers driving midline primordium expression.

In the simplest case, both *Sema-1b* and *Tl* appear to have a single enhancer that controls all aspects of midline primordium expression (“single enhancer” is defined here as a region of DNA that, to date, has not been subdivided into multiple fragments with midline

primordium expression). An interesting variation of the single enhancer mode is the linked *mfas* and *Ect3* genes: we propose that they share a single midline primordium enhancer that resides between them. While additional midline enhancers may be present in sequences not yet analyzed, fragments identified for each gene generally recapitulate the endogenous embryonic midline primordium expression pattern. In contrast, midline primordium expression of *rho* and *sim* autoregulation is controlled by at least two distinct enhancers. However, the relative contributions of the two enhancers are different for the two genes. Midline primordium expression of *sim* involves two enhancers that are both expressed in all midline primordium cells at distinct but overlapping times. Together, they cover the entire period from stages 9–11. In contrast, the two *rho* enhancers are expressed in largely distinct sets of midline cells that together result in expression in the midline primordium at stages 9–11. Thus, the two *sim* enhancers are spatially identical, but differ in a temporal and quantitative manner, whereas the two *rho* enhancers differ spatially. In summary, analysis of midline primordium enhancers from multiple genes indicates the existence of diverse ways to control a simple expression pattern.

The direct role of Single-minded in the control of midline transcription

The *sim* gene is considered to be a master regulator of midline transcription based on its requirement for midline cell specification, development, and transcription, and its ability to convert neuroectodermal cells to midline cells upon ectopic expression (Crews, 2003). However, questions remain regarding how *sim* functions at the molecular level. Experiments from this paper and previous studies confirm that enhancers that drive midline primordium expression utilize CMEs, matching the known binding site for Sim:Tgo. In midline primordium enhancers tested in vivo to date (*sim-1.0*, *sim-3.7*, *rho-9/rjF6D*, *o c t 7 n c n o 1 con l j.1 onn*, *11 1 J J ± ' ' J 365*, *rst -2.5*, *Tl-950*, *Sema-lb-580*, *btl-390*), all showed reductions in midline primordium expression when CMEs were mutated (Aplitz et al., 2005; Freer et al., 2011; Ohshiro and Saigo, 1997; Wharton et al., 1994). Subsequently, *sim* is expressed in all MG, and several MG-expressed enhancers analyzed (*link-285*, *slit-380* and *wrapper -166*) also require CMEs for expression (Estes et al., 2008; Pearson et al., 2012; Wharton et al., 1994). While it is reasonable to propose that every gene expressed in Sim⁺ midline primordium cells and MG is directly regulated by Sim:Tgo, this needs to be more rigorously analyzed by comparing whole-genome Sim binding to enhancers of midline-expressed genes.

The *rho-365* enhancer has two Motif-A sequences that contain GCGTG sequences, and their mutation in the context of mutations of the two CMEs results in a strong reduction in expression. We and others (Hong et al., 2013) considered whether Sim:Tgo may also recognize GCGTG to activate midline expression. If true, this would be the first example of Sim:Tgo binding a non-CME sequence in vivo, and significantly expand the possibilities for Sim control of midline transcription. Numerous studies using in vivo and cell-based assays have consistently demonstrated ACGTG to be a recognition sequence of Sim:Tgo (Jiang and Crews, 2007; Probst et al., 1997; Zhu et al., 2011). Swanson et al. (1995) used in vitro DNA binding assays to demonstrate that a heterodimer of *Drosophila* Sim and mammalian Arnt also bound GCGTG, as well as ACGTG (Swanson et al., 1995). In contrast, transient transfection experiments in *Drosophila* S2 cells showed that changing ACGTG to GCGTG

abolished Sim:Tgo transcriptional activation of a reporter containing multimerized copies of those sequences (Jiang and Crews, 2007). However, further work has revealed that Sim:Tgo binding to CMEs is sensitive to surrounding DNA sequences (Long et al., 2014) ; J. Pearson, unpubl.), suggesting that sequence context can influence binding. Ultimately, the “Does Motif-A binds Sim” issue will be resolved by examining whether Sim:Tgo bind *rho* Motif-A in vivo using ChIP-seq.

rhomboid is directly regulated by Notch signaling

Notch plays an important role during the midline primordium stage, generating midline neural precursors by lateral inhibition and activating the MG program. A key result from mutational analyses of *rho-365* is that *Notch* signaling directly regulates *rho-365* enhancer expression via its transcriptional effector, Su(H). Mutating the Su(H) binding site in *rho-365* nearly eliminates reporter expression, similar to the reduction of expression in a *Dl* mutant embryo. However, other midline primordium-expressed genes are unlikely to be directly regulated by Su(H). The *Ect3-3.2*, *mfas-780*, and *sim-1.6* (Hong et al., 2013) fragments do not contain high-affinity Su(H) consensus binding sites. Mutation of two Su(H) sites had no effect on *Sema-1b-580* enhancer expression, nor did mutation of a Su(H) binding site in *sim-1.0* (J.C.P., unpubl.). Nonetheless, these genes are genetically downstream of Notch signaling, in that expression is down-regulated in Notch-independent MPs soon after the *Notch* signaling event that specifies MG (Wheeler et al., 2008). The lack of Notch signaling in MPs and nascent midline neurons leads to a change in expression or activity of the direct regulators of these midline primordium genes.

In addition to direct Notch/Su(H) activation of *rho-365*, we identified at least 4 other motifs: CME (Sim:Tgo), Motif-A, Repeats-1,2, and Sox. The requirements for each individual motif site was relatively weak (except for the Sox site), but stronger in combination. Still, together they are not sufficient for *rho-365* midline expression, since mutation of the Su(H) site nearly eliminates expression. One or more of these transcription factors are midline-specific (e.g. Sim and MSR; see below) and provide the midline specificity of *rho-365* midline expression. Notch/Su(H) controls the precise timing and pattern of expression within the midline, maintaining *rho* expression within unspecified midline primordium cells, as well as Notch-specified MG and MNB. This model resembles the “Notch plus Proneural” code established for Notch signaling in lateral inhibition, and found in other examples of cell signaling control of transcription (for example: Halfon et al., 2000; Hayashi et al., 2008).

The *rho-365* Su(H) site overlaps a CME, and we carried out a series of experiments to determine if this relationship was functional. We noted that while many alterations to the arrangement of the CME and Su(H) sites were tolerated, *rho-365* expression was strongly reduced in one particular arrangement of sites (Fig. 7O,P), demonstrating some constraints on the position and order of the CME and/or Su(H) sites, relative to adjacent sequences, to promote midline expression of *rho*. This resembles the conserved architectural arrangement of Su(H) and sites of Dorsal and Twist required for regulating neuroectodermal gene expression (Erives and Levine, 2004).

The *Drosophila aos* gene possesses a signaling-dependent midline enhancer whose MG expression is directly regulated by Spi signaling and its transcriptional effector, Pnt (Fig.

9E). Mutation of the *aos* CME site, similar to *rho-365*, did not have a strong effect on reporter gene expression. In contrast, mutating a subset of putative ETS binding sites (4/8) reduced expression. Since *aos-500* is quite specific for MG relative to EGF signaling, these results are consistent with the hypothesis that Sim helps provide midline specificity to Pnt, whose activity may be insufficient to activate *aos* by itself.

Progressive modes of midline gene control

The analysis of midline gene transcription described in this and previous works can begin to describe the progression of midline transcription during development and demonstrate the diverse modes for achieving equivalent patterns of expression (Fig. 9). The *sim* gene likely plays a role in controlling transcription of all midline primordium and MG-expressed genes. The Sim protein is present at high levels at stages 8–11, then is restricted to *Notch*-dependent midline cells (AMG, PMG, MNB/progeny, H-cell sib, and iVUMs 4–6) (Fig. 9A) (Wheeler et al., 2008). The *sim* gene is initially transcribed in midline cells by the actions of dorsal-ventral patterning transcription factors and *Notch* signaling (Crews, 2003), and expression is then maintained through stage 11 by two *sim*-dependent autoregulatory elements residing in *sim-1.0* and *sim-3.7*. However, these *sim* enhancer elements cease activity after stage 11 after which *sim* expression is present in only the *Notch*-dependent cell types. We propose that later *sim* expression is dependent on *Notch* signaling and may also be maintained by *sim* autoregulation.

In one common mode of midline target expression (Fig. 9B), genes such as *Ect3*, *mfas*, *Tl*, *rst*, and *btl* are activated by Sim in all midline primordium cells until stage 11. During stage 11, although Sim protein is present in midline neurons, it is unable to activate transcription of these target genes. These genes generally have 3 or more CMEs, and this multiple number may be essential to the early midline primordium expression as suggested by Hong et al. (2013). In these enhancers, when the CMEs are mutated, expression is highly reduced or absent, indicating the strong requirement of these sites for expression.

Other genes, including *link*, *slit* and *wrapper*, are strongly expressed beginning at stage 11 and continue expression in MG (Fig. 9C). Although not expressed at high levels at stages 9–10, their regulation mechanistically resembles that of genes, such as *Ect3* and *Tl* (Fig. 9B). Their expression in MG is dependent on Sim, with additional inputs by transcription factors such as Sox proteins, Vvl (Ma et al., 2000; Sanchez-Soriano and Russell, 1998), and Zld (Pearson et al., 2012). MG enhancers from *slit*, *wrapper*, and *link* have a single CME each that are required for MG expression, and the absence of additional CMEs may be a factor in why their expression begins later than genes such as *sim*, *Ect3*, and *Tl*, which have additional Sim:Tgo sites (Hong et al., 2013). However, *link* expression is not entirely dependent on its CME, as multiple binding sites for the MG-expressed transcription factor Zld compensate (Pearson et al., 2012).

Another mode of transcription described in this paper and Hong et al. (2013) concerns the *rho* gene, which encodes an important signaling protein whose midline expression is directly initiated by *Notch* signaling (Fig. 9D). There are additional transcriptional regulators of *rho*, including Sim:Tgo, Sox proteins, and unknown proteins that bind to Motif-A (Motif-A Binding Protein; MABP; possibly Sim:Tgo), the Repeat sequences (Repeat Binding Protein;

RBP), and another postulated transcription factor that confers midline specificity (Midline Specificity Regulator; MSR). Interestingly, the role of the ACGTG binding sites is modest by themselves, although expression is strongly reduced when other sequences, such as Motif-A, are also mutated.

In experiments testing the response of *rho-365* variants to *sim* misexpression, several results suggested the existence of additional midline-specific regulatory factors (Fig. 9F). When the Motif-A mutant variant of *rho-365* was examined in *sim* misexpression embryos, *GFP* expression was not induced laterally, although it was present in midline cells (Fig. 6L). One interpretation of this result is that there exists a midline regulatory factor (MSR) present in midline cells that is absent in the ventral ectoderm and cannot be induced ectopically by *sim*. However, MSR is not sufficient to activate *rho-365* expression in the absence of Sim and MABP, since combined mutation of Motif-A and both CMEs abolishes expression.

Another observation from the *rho-365* misexpression experiments suggests that *sim* induces an additional transcription factor(s) (Sim-Induced Regulator; SIR) required for *rho* expression. Expression of the *rho-365* variant with both CMEs mutated is still expanded in *prd-Gal4 UAS-sim* embryos. Since this *sim*-dependent expansion cannot be working through the ACGTG sequences, this suggests that ectopic *sim* induces expression of an additional factor(s) activating *rho-365*; alternatively, Sim:Tgo is able to bind Motif-A in this context to expand expression. The SIR protein may be acting in a feed-forward mechanism with Sim. Similar to MSR, SIR is not sufficient to active *rho-365* expression in the absence of Sim and MABP. In summary, multiple midline specificity factors exist that drive midline lineage-specific transcription even for individual enhancers like *rho-365*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This paper is dedicated to the memory of Larysa Pevny and her groundbreaking work on mammalian Sox proteins. The authors are grateful to Michael Perry, Michael Levine, Nick Reeves, and Jim Posakony for *Drosophila* strains and information regarding *rho* enhancers, Daniel Lau regarding *sim* enhancers, Lan Jiang, Derek Applewhite, Steve Rogers for reagents, and Scott Wheeler for helpful advice. We would particularly like to acknowledge James Oh for cloning *Sema-1b* constructs, and Rachel Tyson for assistance in cloning. We are grateful to the Bloomington *Drosophila* Stock Center for providing *Drosophila* stocks. The project was supported by NRSA postdoctoral awards to JCP (NICHHD), a fellowship from the UNC/NIH Developmental Biology Training Program to JCP, and NIH grants R01 NS64264 (NINDS) and R37 RD25251 (NICHHD) to STC.

References

- Abramoff MD, Magalhaes PJ, Ram SJ. Image processing with ImageJ. *Biophotonics International*. 2004; 11:36–42.
- Apitz H, Strunkelberg M, de Couet HG, Fischbach KF. Single-minded, Dmef2, Pointed, and Su(H) act on identified regulatory sequences of the *roughest* gene in *Drosophila melanogaster*. *Dev. Genes Evol.* 2005; 215:460–469. [PubMed: 16096801]
- Bailey AM, Posakony JW. Suppressor of Hairless directly activates transcription of *Enhancer of split* complex genes in response to Notch receptor activity. *Genes Dev.* 1995; 9:2609–2622. [PubMed: 7590239]

- Barolo S, Walker RG, Polyanovsky AD, Freschi G, Keil T, Posakony JW. A Notch-independent activity of Suppressor of Hairless is required for normal mechanoreceptor physiology. *Cell*. 2000; 103:957–969. [PubMed: 11136980]
- Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 1971; 46:111–138. [PubMed: 5160087]
- Brown CD, Johnson DS, Sidow A. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science*. 2007; 317:1557–1560. [PubMed: 17872446]
- Buescher M, Hing FS, Chia W. Formation of neuroblasts in the embryonic central nervous system of *Drosophila melanogaster* is controlled by *SoxNeuro*. *Development*. 2002; 129:4193–4203. [PubMed: 12183372]
- Carroll SB, Grenier JK, Weatherbee SD. *From DNA to Diversity*. Blackwell Science. 2001
- Castro B, Barolo S, Bailey AM, Posakony JW. Lateral inhibition in proneural clusters: cis-regulatory logic and default repression by Suppressor of Hairless. *Development*. 2005; 132:3333–3344. [PubMed: 15975935]
- Certel K, Anderson MG, Shrigley RJ, Johnson WA. Distinct variant DNA-binding sites determine cell-specific autoregulated expression of the *Drosophila* POU domain transcription factor drifter in midline glia or trachea. *Mol. Cell. Biol.* 1996; 16:1813–1823. [PubMed: 8657157]
- Cowden J, Levine M. The Snail repressor positions Notch signaling in the *Drosophila* embryo. *Development*. 2002; 129:1785–1793. [PubMed: 11923213]
- Crews ST. Crews ST. *Drosophila* bHLH-PAS developmental regulatory proteins. PAS Proteins: Regulators and Sensors of Development and Physiology, Kluwer. 2003:69–108.
- Davidson, EH. *The Regulatory Genome*. Academic Press; 2006.
- Emmons RB, Duncan D, Estes PA, Kiefel P, Mosher JT, Sonnenfeld M, Ward MP, Duncan I, Crews ST. The Spineless-Aristopedia and Tango bHLH-PAS proteins interact to control antennal and tarsal development in *Drosophila*. *Development*. 1999; 126:3937–3945. [PubMed: 10433921]
- Erives A, Levine M. Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:3851–3856. [PubMed: 15026577]
- Estes P, Mosher J, Crews ST. *Drosophila single-minded* represses gene transcription by activating the expression of repressive factors. *Dev. Biol.* 2001; 232:157–175. [PubMed: 11254355]
- Estes P, Fulkerson E, Zhang Y. Identification of motifs that are conserved in 12 *Drosophila* species and regulate midline glia vs. neuron expression. *Genetics*. 2008; 178:787–799. [PubMed: 18245363]
- Fontana JR, Crews ST. Transcriptome analysis of *Drosophila* CNS midline cells reveals diverse peptidergic properties and a role for *castor* in neuronal differentiation. *Dev. Biol.* 2012; 372:131–142. [PubMed: 23010511]
- Freer SM, Lau DC, Pearson JC, Talsky KB, Crews ST. Molecular and functional analysis of *Drosophila single-minded* larval central brain expression. *Gene Expr. Patterns*. 2011; 11:533–546. [PubMed: 21945234]
- Fulkerson E, Estes PA. Common motifs shared by conserved enhancers of *Drosophila* midline glial genes. *J. Exp. Zool.* 2011; 316B:61–75.
- Gabay L, Seger R, Shilo BZ. In situ activation pattern of *Drosophila* EGF receptor pathway during development. *Science*. 1997; 277:1103–1106. [PubMed: 9262480]
- Go MJ, Eastman DS, Artavanis-Tsakonas S. Cell proliferation control by Notch signaling in *Drosophila* development. *Development*. 1998; 125:2031–2040. [PubMed: 9570768]
- Golembo M, Raz E, Shilo BZ. The *Drosophila* embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development*. 1996; 122:3363–3370. [PubMed: 8951053]
- Groth AC, Fish M, Nusse R, Calos MP. Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics*. 2004; 166:1775–1782. [PubMed: 15126397]
- Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jimenez F, Baylies MK, Michelson AM. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*. 2000; 103:63–74. [PubMed: 11051548]

- Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* 2011; 7:e1002266. [PubMed: 22028662]
- Hayashi T, Xu C, Carthew RW. Cell-type-specific transcription of *prospero* is controlled by combinatorial signaling in the *Drosophila* eye. *Development.* 2008; 135:2787–2796. [PubMed: 18635611]
- Hilliker AJ, Clark SH, Chovnick A, Gelbart WM. Cytogenetic analysis of the chromosomal region immediately adjacent to the rosy locus in *Drosophila melanogaster*. *Genetics.* 1980; 95:95–110. [PubMed: 6776006]
- Hong JW, Won Park K, Levine MS. Temporal regulation of *single-minded* target genes in the ventral midline of the *Drosophila* central nervous system. *Dev. Biol.* 2013; 380:335–343. [PubMed: 23701883]
- Hu S, Sonnenfeld M, Stahl S, Crews ST. Midline Fasciclin: a *Drosophila* Fasciclin-I-related membrane protein localized to the CNS midline cells and trachea. 1998; 35:77–93.
- Ip YT, Park RE, Kosman D, Bier E, Levine M. The *dorsal* gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev.* 1992; 6:1728–1739. [PubMed: 1325394]
- Jiang L, Crews ST. Transcriptional specificity of *Drosophila* *dysfusion* and the control of tracheal fusion cell gene expression. *J. Biol. Chem.* 2007; 282:28659–28668. [PubMed: 17652079]
- Jiang L, Pearson JC, Crews ST. Diverse modes of *Drosophila* tracheal fusion cell transcriptional regulation. *Mech. Dev.* 2010; 127:265–280. [PubMed: 20347970]
- Kasai Y, Stahl S, Crews S. Specification of the *Drosophila* CNS midline cell lineage: direct control of *single-minded* transcription by dorsal/ventral patterning genes. *Gene Expr.* 1998; 7:171–189. [PubMed: 9840810]
- Kearney JB, Wheeler SR, Estes P, Parente B, Crews ST. Gene expression profiling of the developing *Drosophila* CNS midline cells. *Dev. Biol.* 2004; 275:473–492. [PubMed: 15501232]
- Khare N, Fascetti N, DaRocha S, Chiquet-Ehrismann R, Baumgartner S. Expression patterns of two new members of the Semaphorin family in *Drosophila* suggest early functions during embryogenesis. *Mech. Dev.* 2000; 91:393–397. [PubMed: 10704872]
- Kidd S, Lieber T, Young MW. Ligand-induced cleavage and regulation of nuclear entry of Notch in *Drosophila melanogaster* embryos. *Genes Dev.* 1998; 12:3728–3740. [PubMed: 9851979]
- Kim SH, Crews ST. Influence of *Drosophila* ventral epidermal development by the CNS midline cells and *spitz* class genes. *Development.* 1993; 118:893–901. [PubMed: 8076524]
- Kvon EZ, Stampfel tG, Yáñez-Cuna JO, Dickson BJ, Stark A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev.* 2012; 26:908–913. [PubMed: 22499593]
- Lee T, Luo L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron.* 1999; 22:451–461. [PubMed: 10197526]
- Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* 2008; 6:e27. [PubMed: 18271625]
- Long SK, Fulkerson E, Breese R, Hernandez G, Davis C, Melton MA, Chandran RR, Butler N, Jiang L, Estes P. A Comparison of Midline and Tracheal Gene Regulation during *Drosophila* Development. *PLoS One.* 2014; 9:e85518. [PubMed: 24465586]
- Lusk RW, Eisen MB. Evolutionary mirages: selection on binding site composition creates the illusion of conserved grammars in *Drosophila* enhancers. *PLoS Genet.* 2010; 6:e1000829. [PubMed: 20107516]
- Ma Y, Certel K, Gao Y, Niemitz E, Mosher J, Mukherjee A, Mutsuddi M, Huseinovic N, Crews ST, Johnson WA, Nambu JR. Functional interactions between *Drosophila* bHLH/PAS, Sox, and POU transcription factors regulate CNS midline expression of the *slit* gene. *J. Neurosci.* 2000; 20:4596–4605. [PubMed: 10844029]

- MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keranen SV, Knowles DW, Stapleton M, Bickel P, Biggin MD, Eisen MB. Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 2009; 10:R80. [PubMed: 19627575]
- Manning L, Heckscher ES, Purice MD, Roberts J, Bennett AL, Kroll JR, Pollard JL, Strader ME, Lupton JR, Dyukareva AV, Doan PN, Bauer DM, Wilbur AN, Tanner S, Kelly JJ, Lai SL, Tran KD, Kohwi M, Lavery TR, Pearson JC, Crews ST, Rubin GM, Doe CQ. A resource for manipulating gene expression and analyzing *cis*-regulatory modules in the *Drosophila* CNS. *Cell. Rep.* 2012; 2:1002–1013. [PubMed: 23063363]
- Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A, Levine M. A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development.* 2004; 131:2387–2394. [PubMed: 15128669]
- Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, Haussler M, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Dreszer TR, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, Kent WJ. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* 2013; 41:D64–D69. [PubMed: 23155063]
- Morel V, Schweisguth F. Repression by Suppressor of Hairless and activation by Notch are required to define a single row of *single-minded* expressing cells in the *Drosophila* embryo. *Genes Dev.* 2000; 14:377–388. [PubMed: 10673509]
- Muralidhar MG, Callahan CA, Thomas JB. Single-minded regulation of genes in the embryonic midline of the *Drosophila* central nervous system. *Mech. Dev.* 1993; 41:129–138. [PubMed: 8518191]
- Nambu JR, Franks RG, Hu S, Crews ST. The *single-minded* gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell.* 1990; 63:63–75. [PubMed: 2242162]
- Nambu JR, Lewis JL, Wharton KA, Crews ST. The *Drosophila single-minded* gene encodes a helix-loop-helix protein which acts as a master regulator of CNS midline development. *Cell.* 1991; 67:1157–1167. [PubMed: 1760843]
- Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, Li Z, Ishii H, Spokony RF, Chen J, Hwang L, Cheng C, Auburn RP, Davis MB, Domanus M, Shah PK, Morrison CA, Zieba J, Suchy S, Senderowicz L, Victorson A, Bild NA, Grundstad AJ, Hanley D, MacAlpine DM, Mannervik M, Venken K, Bellen H, White R, Gerstein M, Russell S, Grossman RL, Ren B, Posakony JW, Kellis M, White KP. A *cis*-regulatory map of the *Drosophila* genome. *Nature.* 2011; 471:527–531. [PubMed: 21430782]
- Ohshiro T, Saigo K. Transcriptional regulation of *breathless* FGF receptor gene by binding of TRACHEALESS/dARNT heterodimers to three central midline elements in *Drosophila* developing trachea. *Development.* 1997; 124:3975–3986. [PubMed: 9374395]
- Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 2011; 12:283–293. [PubMed: 21358745]
- Overton PM, Meadows LA, Urban J, Russell S. Evidence for differential and redundant function of the *Sox* genes *Dichaete* and *SoxN* during CNS development in *Drosophila*. *Development.* 2002; 129:4219–4228. [PubMed: 12183374]
- Patel NH, Martin-Blanco E, Coleman KG, Poole SJ, Ellis MC, Kornberg TB, Goodman CS. Expression of *engrailed* proteins in arthropods, annelids, and chordates. *Cell.* 1989; 58:955–968. [PubMed: 2570637]
- Pearson JC, Watson JD, Crews ST. *Drosophila melanogaster* Zelda and Single-minded collaborate to regulate an evolutionarily dynamic CNS midline cell enhancer. *Dev. Biol.* 2012; 366:420–432. [PubMed: 22537497]
- Pearson JC, Crews ST. Twine: Display and Analysis of Cis-Regulatory Modules. *Bioinformatics.* 2013; 29:1690–1692. [PubMed: 23658420]

- Probst MR, Fan CM, Tessier-Lavigne M, Hankinson O. Two murine homologs of the *Drosophila* single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* 1997; 272:4451–4457. [PubMed: 9020169]
- Rebeiz M, Posakony JW. GenePalette: a universal software tool for genome sequence visualization and analysis. *Dev. Biol.* 2004; 271:431–438. [PubMed: 15223345]
- Reeves N, Posakony JW. Genetic programs activated by proneural proteins in the developing *Drosophila* PNS. *Dev. Cell.* 2005; 8:413–425. [PubMed: 15737936]
- Sanchez-Soriano N, Russell S. The *Drosophila* SOX-domain protein Dichaete is required for the development of the central nervous system midline. *Development.* 1998; 125:3989–3996. [PubMed: 9735360]
- Schweitzer R, Howes R, Smith R, Shilo BZ, Freeman M. Inhibition of *Drosophila* EGF receptor activation by the secreted protein Argos. *Nature.* 1995; 376:699–702. [PubMed: 7651519]
- Sonnenfeld M, Ward M, Nystrom G, Mosher J, Stahl S, Crews S. The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development.* 1997; 124:4571–4582. [PubMed: 9409674]
- Sonnhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene.* 1995; 167:GC1–GC10. [PubMed: 8566757]
- Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 2012; 13:613–626. [PubMed: 22868264]
- Stemerdink C, Jacobs JR. Argos and Spitz group genes function to regulate midline glial cell number in *Drosophila* embryos. *Development.* 1997; 124:3787–3796. [PubMed: 9367434]
- Swanson HI, Chan WK, Bradfield CA. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. 1995; 280:26292–26302.
- Thomas JB, Crews ST, Goodman CS. Molecular genetics of the *single-minded locus*: a gene involved in the development of the *Drosophila* nervous system. *Cell.* 1988; 52:133–141. [PubMed: 3345559]
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, Biggin MD, Stamatoyannopoulos JA. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* 2011; 12:R43. [PubMed: 21569360]
- Uemura T, Shepherd S, Ackerman L, Jan LY, Jan YN. *numb*, a gene required in determination of cell fate during sensory organ formation in *Drosophila* embryos. *Cell.* 1989; 58:349–360. [PubMed: 2752427]
- Watson JD, Wheeler SR, Stagg SB, Crews ST. *Drosophila hedgehog* signaling and *engrailed-runt* mutual repression direct midline glia to alternative ensheathing and non-ensheathing fates. *Development.* 2011; 138:1285–1295. [PubMed: 21350018]
- Watson JD, Crews ST. Formation and specification of a *Drosophila* dopaminergic precursor cell. *Development.* 2012; 139:3316–3325.
- Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, Yan J, Talukder S, Turunen M, Taipale M, Stunnenberg HG, Ukkonen E, Hughes TR, Bulyk ML, Taipale J. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* 2010; 29:2147–2160. [PubMed: 20517297]
- Wharton JKA, Crews ST. CNS midline enhancers of the *Drosophila slit* and *Toll* genes. *Mech. Dev.* 1993; 40:141–154. [PubMed: 8494768]
- Wharton JKA, Franks RG, Kasai Y, Crews ST. Control of CNS midline transcription by asymmetric E-box elements: similarity to xenobiotic responsive regulation. *Development.* 1994; 120:3563–3569. [PubMed: 7821222]
- Wheeler SR, Kearney JB, Guardiola AR, Crews ST. Single-cell mapping of neural and glial gene expression in the developing *Drosophila* CNS midline cells. *Dev. Biol.* 2006; 294:509–524. [PubMed: 16631157]
- Wheeler SR, Stagg SB, Crews ST. Multiple *Notch* signaling events control *Drosophila* CNS midline neurogenesis, gliogenesis and neuronal identity. *Development.* 2008; 135:3071–3079. [PubMed: 18701546]

- Wheeler SR, Stagg SB, Crews ST. MidExDB: a database of *Drosophila* CNS midline cell gene expression. *BMC Dev. Biol.* 2009; 9:56–64. [PubMed: 19903351]
- Wheeler SR, Pearson JC, Crews ST. Time-lapse imaging reveals stereotypical patterns of *Drosophila* midline glial migration. *Dev. Biol.* 2012; 361:232–244. [PubMed: 22061481]
- Wildonger J, Sosinsky A, Honig B, Mann RS. Lozenge directly activates *argos* and *klumpfuss* to regulate programmed cell death. *Genes Dev.* 2005; 19:1034–1039. [PubMed: 15879554]
- Xiao H, Hrdlicka LA, Nambu JR. Alternate functions of the *single-minded* and *rhomboid* genes in development of the *Drosophila* ventral neuroectoderm. *Mech. Dev.* 1996; 58:65–74. [PubMed: 8887317]
- Yáñez-Cuna JO, Kvon EZ, Stark A. Deciphering the transcriptional cis-regulatory code. *Trends Genet.* 2013; 29:11–22. [PubMed: 23102583]
- Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev.* 2007; 21:385–390. [PubMed: 17322397]
- Zelzer E, Shilo B. Interaction between the bHLH-PAS protein Trachealess and the POU-domain protein Drifter, specifies tracheal cell fates. *Mech. Dev.* 2000; 19:163–173. [PubMed: 10704841]
- Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasfield JA, Zhu C, Asriyan Y, Lapointe DS, Sinha S, Wolfe SA, Brodsky MH. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 2011; 39:D111–7. [PubMed: 21097781]
- Zinzen RP, Cande J, Ronshaugen M, Papatsenko D, Levine M. Evolution of the ventral midline in insect embryos. *Dev. Cell.* 2006; 11:895–902. [PubMed: 17141163]
- Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature.* 2009; 462:65–70. [PubMed: 19890324]

Highlights

- *Drosophila* CNS midline enhancers show architectural and mechanistic diversity
- All tested midline-expressed enhancers are directly regulated by Single-minded
- Notch signaling directly regulates expression of the EGF protease *rhomboid* gene
- Multiple midline-specific regulators control expression along with Single-minded

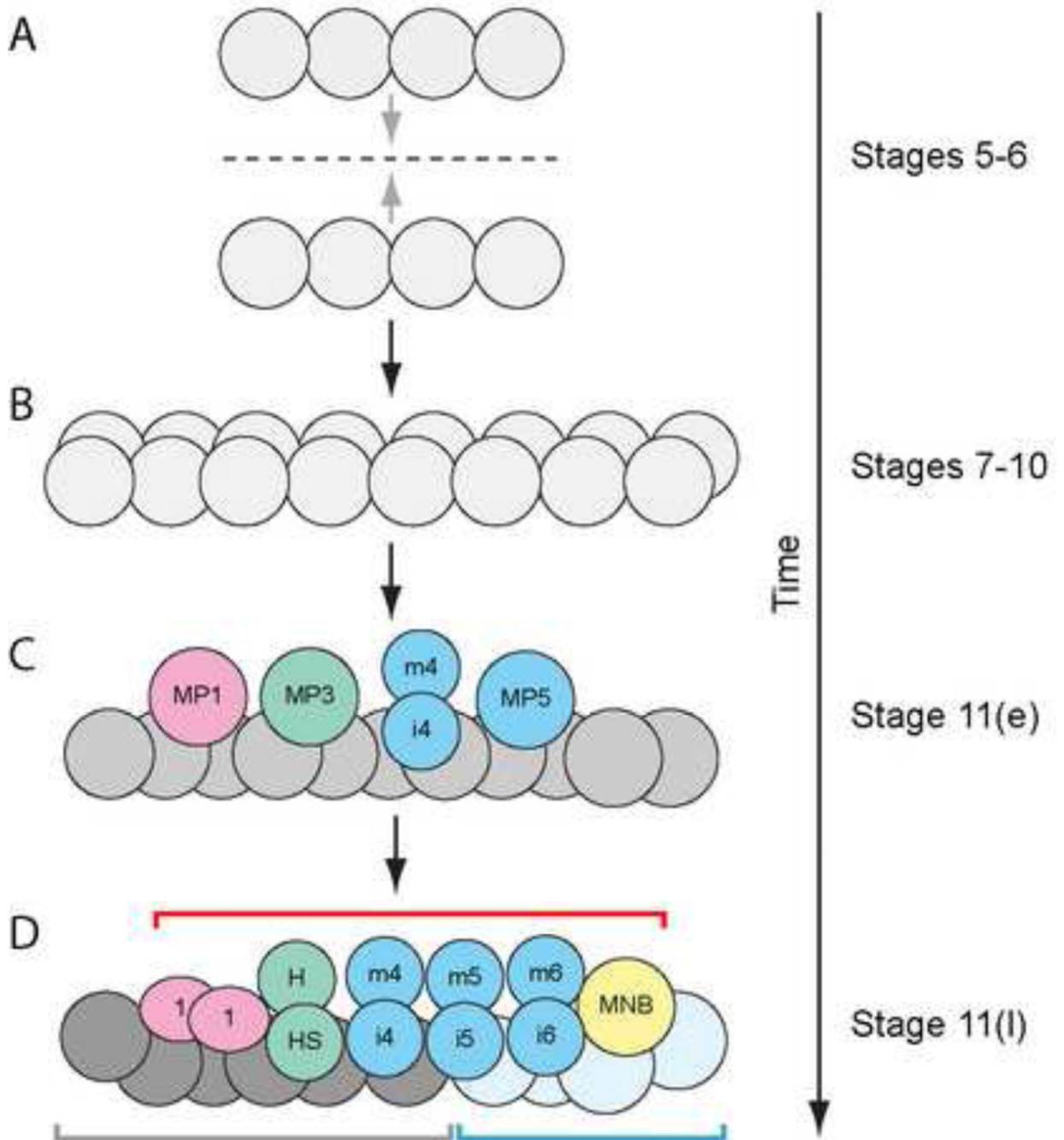


Figure 1. Overview of early CNS midline cell development. Single segments are depicted at successive stages of development

(A) Horizontal view of stages 5–6 midline cells (referred to as mesectoderm at these stages). Two single cell-wide stripes of 4 cells each converge at the embryonic midline (dotted line) during gastrulation (stages 6–7). (B) After gastrulation, the mesectodermal cells undergo a synchronous cell division to give rise to 16 cells/segment at stages 8–10 (referred to as the midline primordium) and intercalate. Sagittal view is depicted. (C) At early stage 11 (11(e)), MP4 has delaminated and divided into mVUM4 (m4) and iVUM4 (i4), while MP1, MP3, and MP5 have delaminated but not yet divided into their neuronal progeny. The remaining

midline cells (gray) are a mixture of the not-yet-delaminated MP6, newly formed MG, and the MNB. (D) By late stage 11 (11(l)), MP1,3–6 have delaminated and divided. MP1 has divided symmetrically into the 2 MP1 neurons (1), MP3 has divided into the H-cell (H) and H-cell sib (HS) neurons, and MP4–6 have divided into iVUM4–6 and mVUM4–6. The MNB has delaminated and will soon initiate division. AMG (dark gray) and PMG (light blue) are beginning their inward migration. In this figure and others, the red bracket shows the extent of the midline neurons and precursors, the gray bracket indicates AMG and blue bracket indicates PMG.

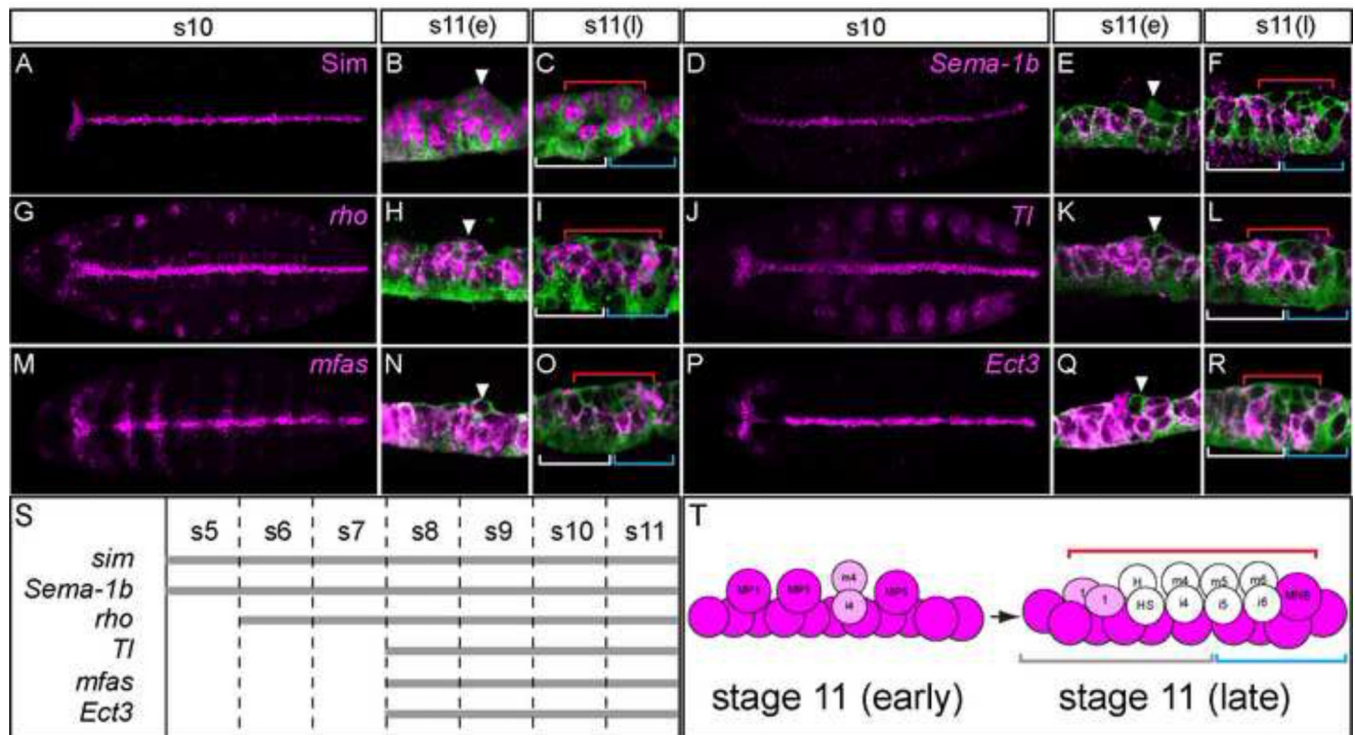


Figure 2. Common patterns of midline primordium gene expression

(A–R) Midline primordium expression (magenta) of 6 genes revealed by (A–C) anti-Sim staining, or (D–R) in situ hybridization of *sim-Gal4 UAS-tau-GFP* embryos. Stages shown are: stage 10 (s10), early stage 11 (s11(e)), and late stage 11 (s11(l)). Horizontal views of whole embryo maximum projections, with first and last confocal slices selected to emphasize CNS midline expression, are shown for stage 10 embryos, and sagittal views of single segments are shown for stage 11 embryos. Embryos are co-stained with anti-tau (green) to outline midline cells. The red, white, and blue lines correspond to the lines in Fig. 1 and mark midline neural precursors and neurons (red), AMG (white), and PMG (blue). Arrowheads point to delaminating MP4s or their iVUM4/mVUM4 progeny, which act as developmental timing and positional markers. Anterior is to the left. (S) Timeline of midline expression of 6 genes depicted in A–R. Stages are shown at the top. (T) Schematic of dynamic midline primordium gene expression during stage 11. Midline cells are depicted as in Fig. 1, and all cells initially express each gene (magenta). As MPs delaminate and divide, they reduce expression, resulting in regions of low (pink) or no (white) expression in the interior of each segment by late stage 11.

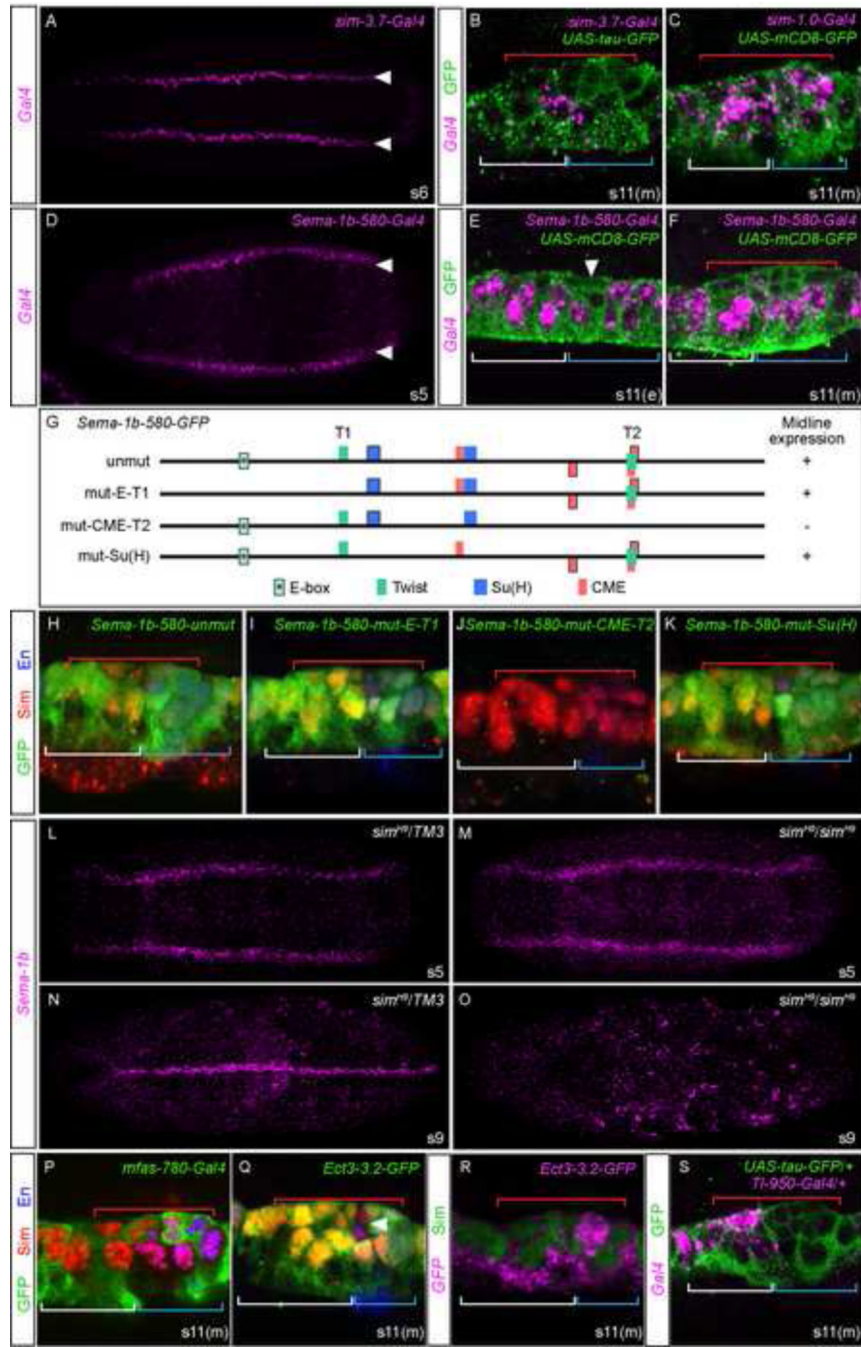


Figure 3. Midline primordium enhancers control broad and restricted midline expression (A–F) In situ hybridizations of *sim-3.7-Gal4*, *sim-1.0-Gal4*, and *Sema-1b-580-Gal4* embryos with a *Gal4* probe showing midline expression (magenta) at stages 5 (s5), 6 (s6), early 11 (s11(e)) and mid 11 (s11(m)). (A,D) Horizontal views are shown with mesectodermal stripes indicated by arrowheads. (B,C,E,F) Shown are maximum projections of two adjacent sagittal view confocal slices of abdominal midline segments from embryos containing either *UAS-tau-GFP* or *UAS-mCD8-GFP* and stained with anti-GFP to visualize GFP accumulation due to the activity of each enhancer throughout the midline primordium

(green). White arrowhead indicates delaminated MP4, red brackets cover midline neurons and MPs, white and blue brackets show the locations of AMG and PMG, respectively. (G) Schematics of *Sema-1b-580* and mutant variants showing potential CME, Twist (T1, T2), and Su(H) binding sites. Width of each block corresponds to the length of consensus binding sites and the presence of a black border indicates an aligned sequence with 100% identity in all species examined. A conserved E-box (CANNTG) that was tested in addition to canonical Twist sites is indicated with (*). Midline expression of the reporter transgene is indicated at right. (H–K) Sagittal views of constructs schematized in G are shown stained with anti-GFP (reporter expression; green), anti-Sim (midline cells; red), and anti-Engrailed (En) (posterior midline cells; blue). (L–O) Hybridization of a *Sema-1b* probe to (L,N) *sim* /*TM3* heterozygous embryos and (M,O) *sim* homozygous mutant embryos. (P–Q) Expression analysis of *mfas*, *Ect3*, and *Tl* reporter constructs stained for (P,Q) GFP protein, (R) *GFP* RNA, and (S) *Gal4* RNA. Degenerate nucleotide code: R=A/G, Y=C/T, S=G/C, W=A/T, K=G/T, M=A/C, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, N=A/C/G/T.

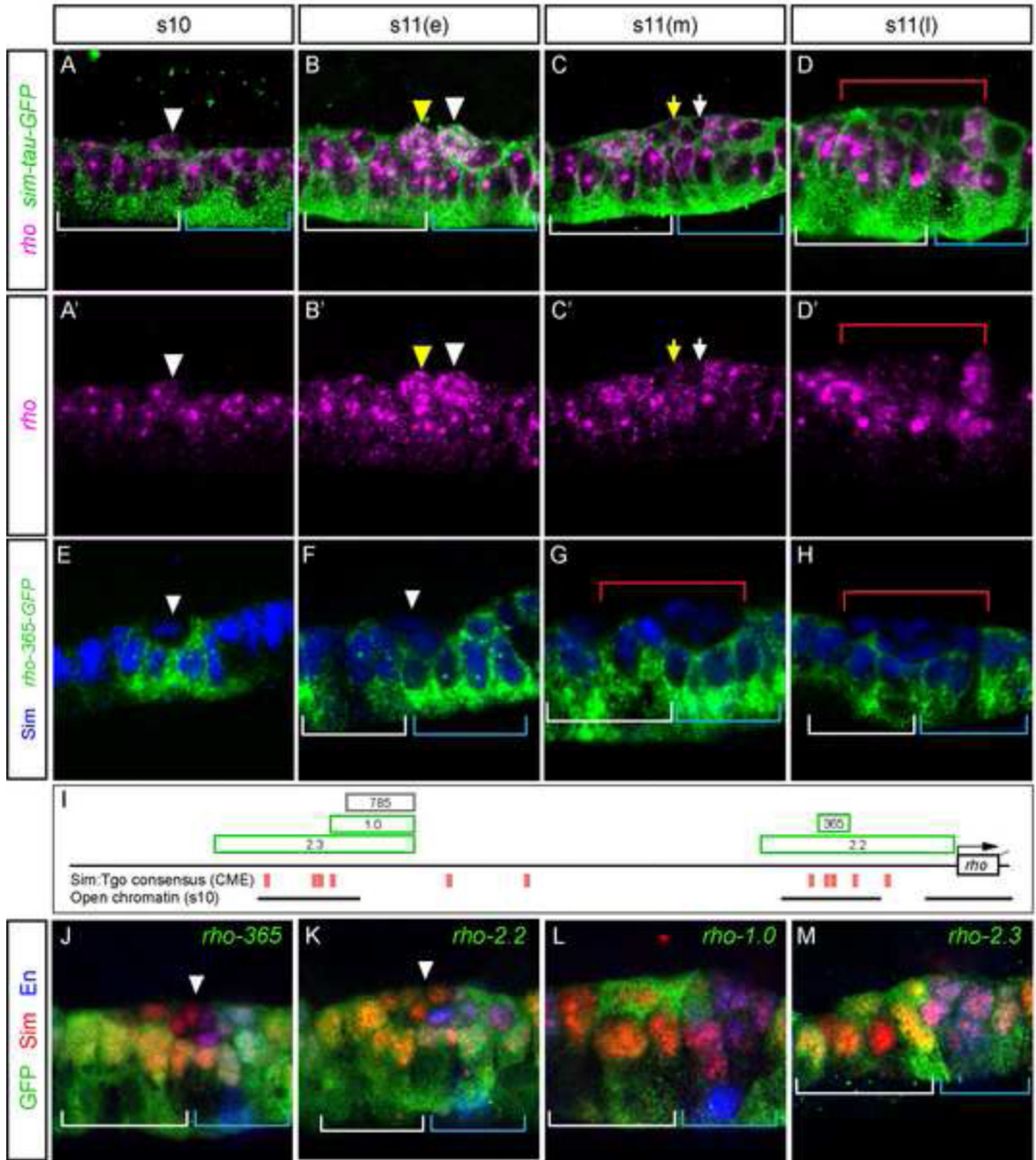


Figure 4. Multiple enhancers collaborate to control *rhomboid* midline primordium expression
 (A-D') Expression dynamics of *rho* (magenta) is shown in representative segments in *sim-Gal4 UAS-tau-GFP* (*sim-tau-GFP*) embryos also stained with anti-tau (green). Stages shown are: (A,A') stage 10, (B,B') early stage 11 showing a burst of *rho* expression in MP3 (yellow arrowhead) and MP4 (white arrowhead), (C,C) mid stage 11 in which neuronal progeny (arrows) show reductions in expression, and (D,D') late-stage 11 showing low levels of *rho* expression in neurons (below red bracket). (E-H) *rho-365-GFP* embryos stained for *GFP* RNA (green) and Sim protein (blue). (E) Arrowhead indicates the presence

of strong *GFP* in a small number of midline cells at stage 10. (F) *GFP* expression expands posteriorly in midline cells, but is restricted from iVUM4/mVUM4 (white arrowhead). (G,H) Reduction of *GFP* expression is apparent in midline neurons beneath the red bracket, but is maintained in AMG (white bracket) and PMG (blue bracket). (I) Schematic of *rho* locus showing features related to gene expression. Genomic fragments with midline activity are indicated as green boxes; *rho-785* does not have midline activity. Consensus Sim:Tgo binding site matches (ACGTG) are indicated as red boxes, and the presence and extent of open chromatin regions (DNase I hypersensitive sites, 5% FDR, adapted from genome.ucsc.edu) at stage 10 are indicated by lines. (J–M) Embryos containing *rho* midline fragments driving *GFP* expression at mid-stage 11 stained for GFP, Sim, and En. (J,K) Arrowheads point to neurons with reduced expression.

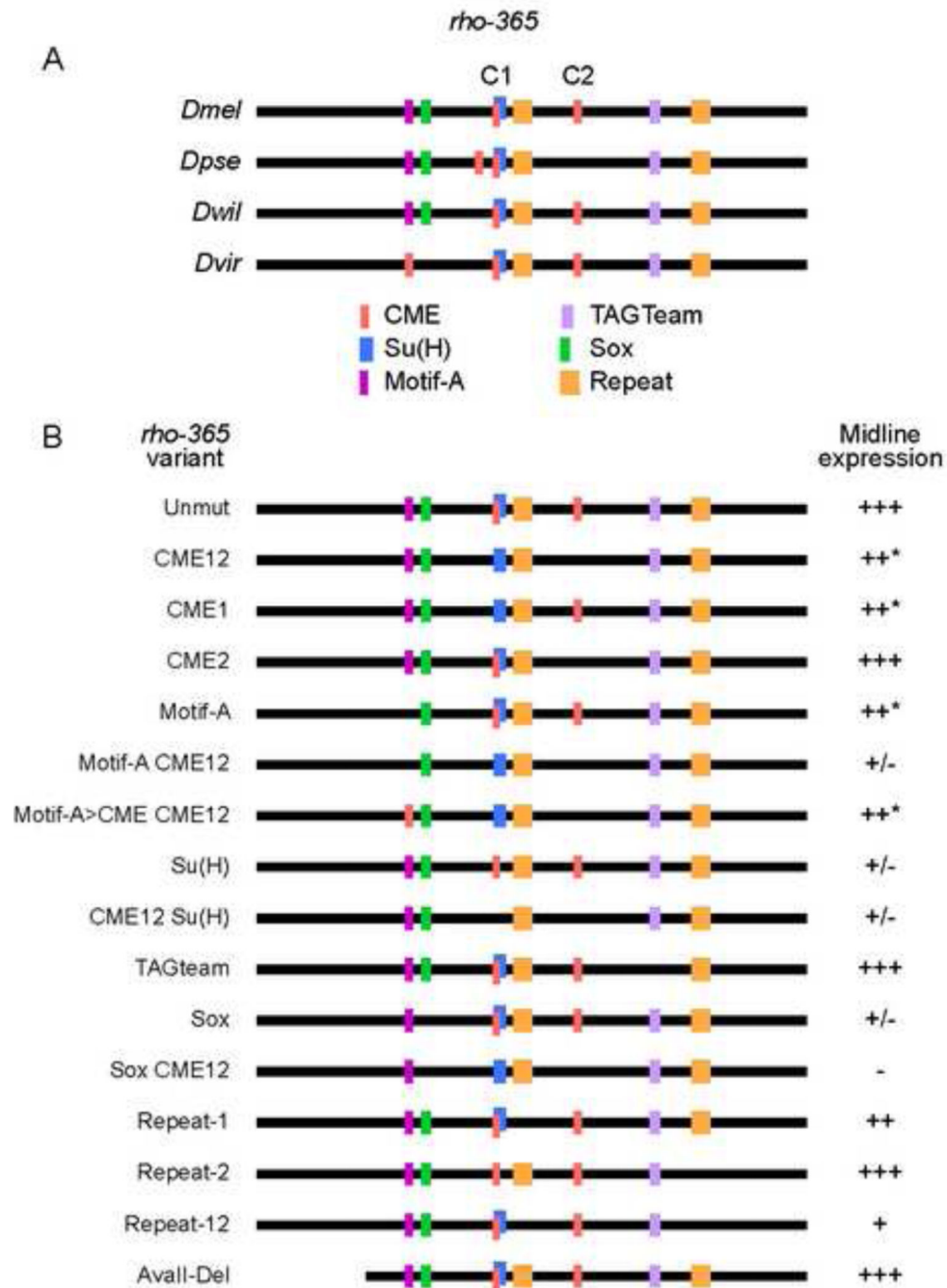


Figure 5. Functional analysis of *rho-365* DNA sequence motifs

(A) Schematic of the *rho-365* sequence and representative orthologous sequences among sequenced *Drosophila* species. Vertically aligned motifs indicate likely orthologous sequences; sequence lines are not drawn to scale. Species shown are: *D. melanogaster* (*Dmel*), *D. pseudoobscura* (*Dpse*); *D. willistoni* (*Dwil*), and *D. virilis* (*Dvir*). The two CMEs are indicated by C1 and C2. Note that the first Repeat motif site (orange) in *D. willistoni* contains a mismatch compared to the *D. melanogaster* sequence. (B) Mutations were generated in *Dmel rho-365* and tested in vivo for midline expression. Absence of a colored

block indicates mutation of that sequence; conversion to a different color indicates that the motif was changed to a different motif. Midline expression levels (GFP protein accumulation at *s11(m)*) were indicated as follows: (+++) normal levels and pattern, (++*) slight reduction in a subset of cells, (++) slightly reduced expression, (+) weak but consistent expression; (+/-) nearly undetectable expression, and (-) no expression.

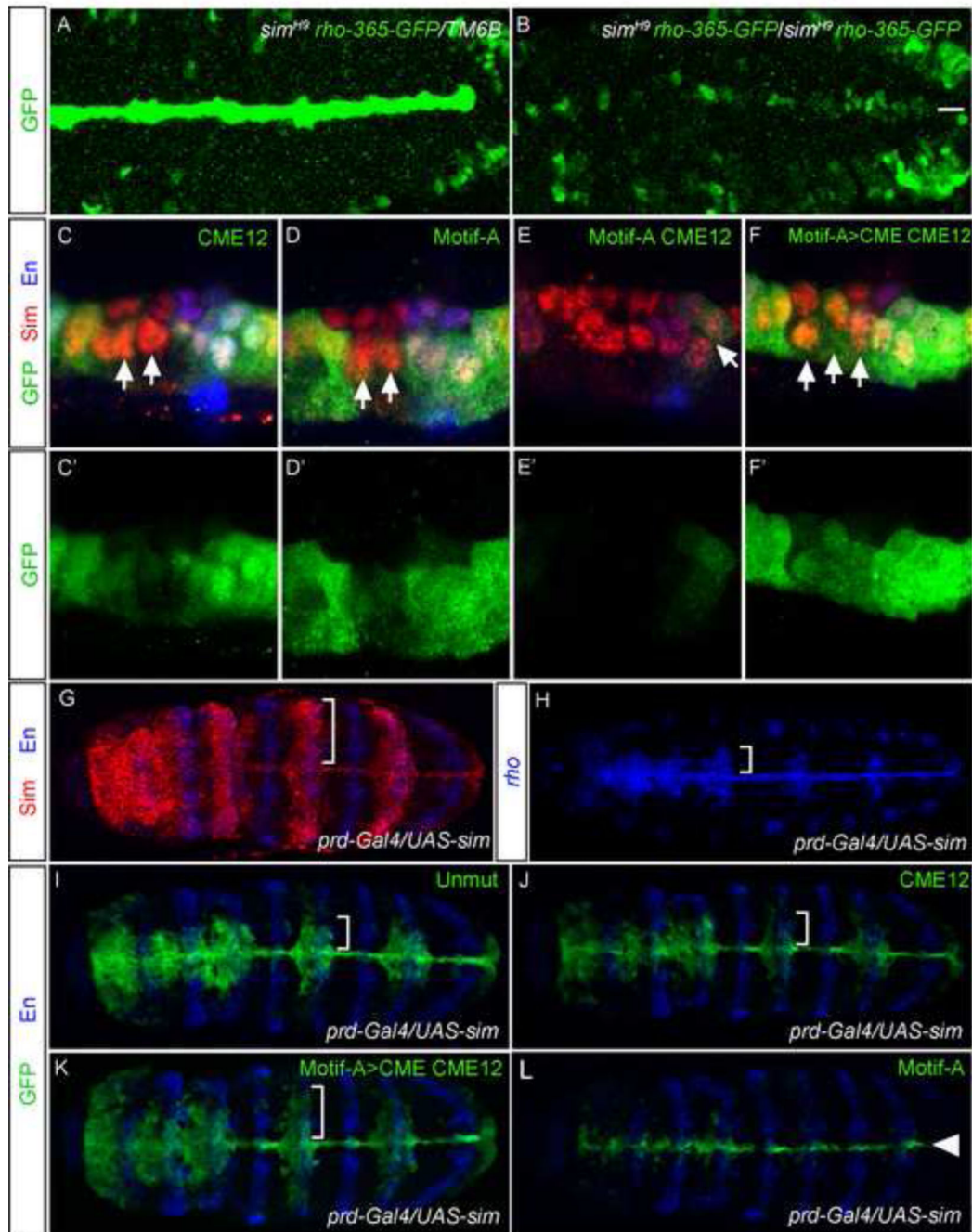


Figure 6. Direct and indirect regulation of *rho* by Sim

(A,B) Horizontal views of *rho-365-GFP* embryos in posterior segments at stage 11 in: (A) *sim* heterozygous and (B) *sim* homozygous mutant embryos, detecting GFP. In B, the white line indicates the location of the ventral midline. (C–F) Sagittal views of stage 11 single segments of *rho-365* site variants, detecting GFP (green), Sim (red), and En (blue) (compare to Fig. 4J): (C) *rho-365-Sim12*; (D) *rho-365-Motif-A*; (E) *rho-365-Motif-A Sim12*; (F) *rho-365-Motif-A>Sim Sim12*. (C'–F') Same images as (C–F), but showing only the GFP channel. (C,D,F) Arrows point to MG with reduced GFP. (E) Arrow points to several

posterior midline cells with weak GFP. (G,H) Horizontal views of stage 11 *prd-Gal4/UAS-sim* embryos stained for (G) Sim and En or (H) *rho*; brackets indicate expanded expression. (I–L) Stage 11 *prd-Gal4 UAS-sim* embryos containing *rho-365-GFP* either: (I) unmutated (Unmut) or with (J) Sim12, (K) Motif-A>Sim Sim12, or (L) Motif-A mutations. The bracket indicates a domain of expanded expression, except for (L) in which there is no significant expanded expression, while expression remains strong in the midline primordium (white arrowhead).

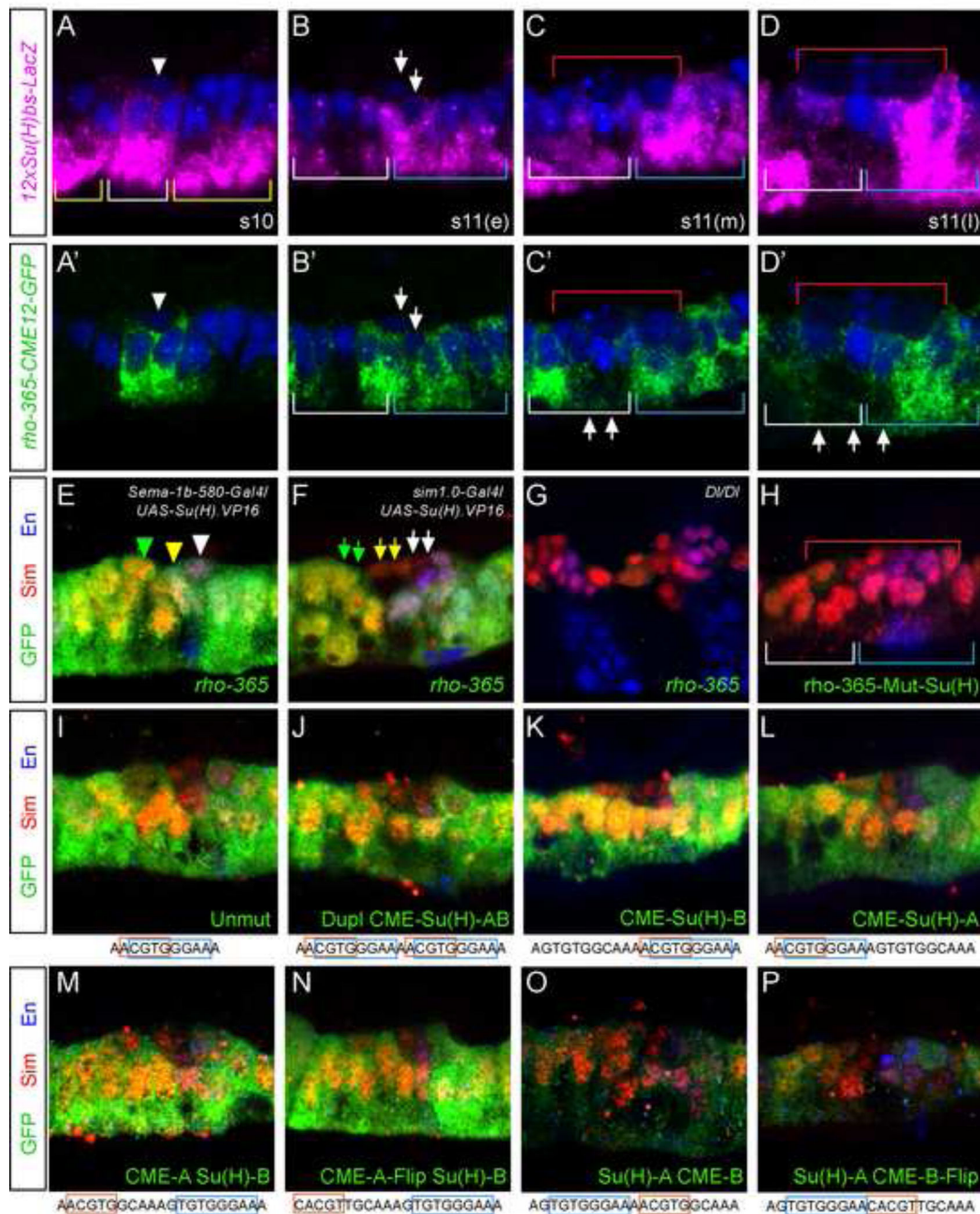


Figure 7. Notch/Su(H) directly regulates *rho* midline expression

(A–D') Sagittal views of single segments of *12xSu(H)bs-lacZ*; *rho-365-Sim12-GFP* embryos, stained for Sim (blue) and (A–D) *lacZ* (magenta) or (A'–D') *GFP* (green). Embryonic stages examined were (A,A') stage 10, (B,B') early stage 11 (s11(e)), (C,C) mid stage 11 (s11(m)), and (D,D') late stage 11 (s11(l)). (A) Yellow brackets indicate *lacZ* expression in non-midline ectodermal cells, and white brackets indicates *lacZ* expression in midline cells. (C',D') White arrows point to reduced apical *GFP* expression in posterior-most AMG compared to *rho-365-GFP* (Fig. 4G,H). (E–H) Stage 11 *rho-365-GFP* in genetic

backgrounds with altered *Notch* signaling, stained for GFP (green), Sim (red), and En (blue). (E) *Sema-1b-580-Gal4/UAS-Su(H).VP16; rho-365-GFP* midline segment with strong GFP staining in all midline cells, including MP1 (green arrowhead), MP3 (yellow arrowhead), and MP4 (white arrowhead). (F) *sim1.0-Gal4/UAS-Su(H).VP16; rho-365-GFP* embryo showing GFP levels similar to *rho-365*; note absence of staining in midline neurons from MP3 and MP4. (G) *rho-365-GFP* in a *Dl* mutant. (H–P) *rho-365-GFP* Su(H) and CME site variants, stained for GFP, Sim, and En. (H) *rho-365-GFP* with a single base change in the Su(H) binding site. (I–P) *rho-365-GFP* with the overlapping CME/Su(H) sequences altered as indicated below each image. CME is boxed in red, and the Su(H) binding site is boxed in blue. Positions “A” and “B” indicate the first or second positions in duplicated sequences.

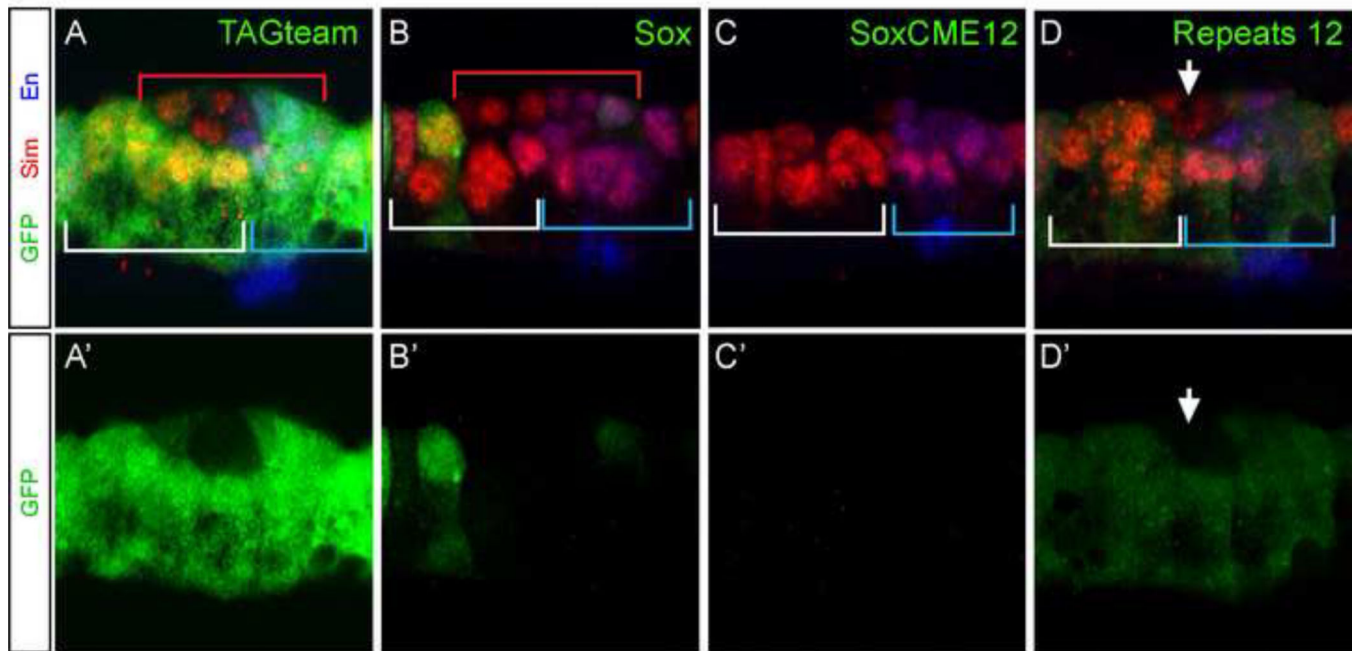


Figure 8. Sox binding sites and Repeat sequences are required for *rho* midline expression
 (A–D) Sagittal views of embryos containing the indicated *rho-365-GFP* variants, stained for GFP, Sim, and En. Variants include mutations in the (A) TAGteam sequence, (B) Sox site, (C) Sox CME12, and (D) Repeats 12. (A'–D') Same images as A–D, showing only GFP.

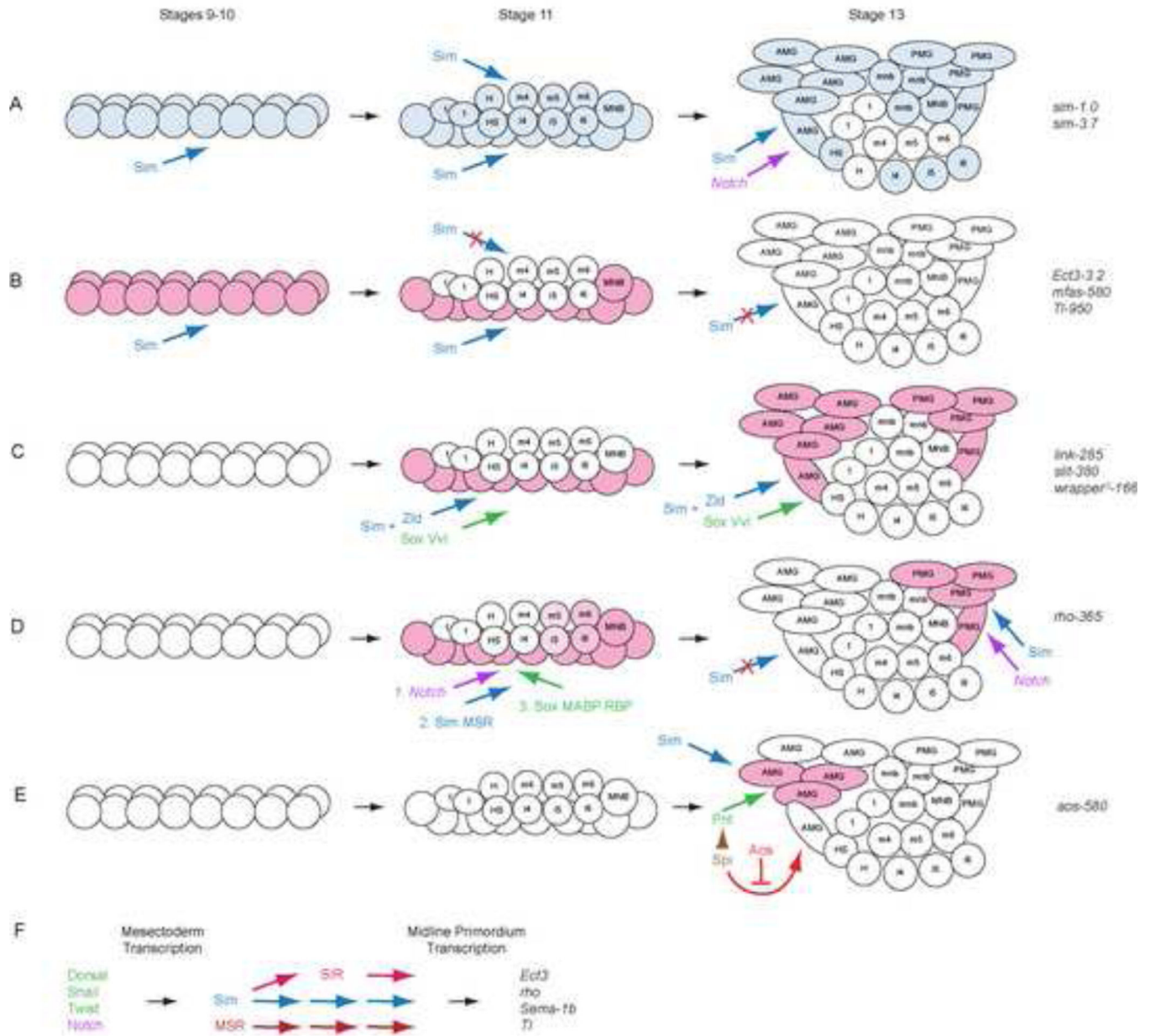


Figure 9. Different modes of midline gene control

(A) Sim protein (blue) is present in all midline cells at stages 8–11, and in *Notch*-dependent cell types at stage 13 (purple arrow); direct autoregulation is a major aspect of maintained *sim* expression. Sim arrows indicate direct regulation of the enhancers listed on the right. At stage 11, Sim regulates midline neuronal (top arrow) and MG (bottom arrow) expression. (B) Sim-Primary role: midline primordium mode of expression in which Sim protein directly activates high-level expression at stages 9–10 (red). At stage 11, Sim protein function is blocked in neurons, with continued activity in MG and MNB. Activity of these enhancers is absent during stages 12 and later, despite continued presence of Sim in most

midline cells. (C) MG-expressed genes, including *link*, *slit* and *wrapper*, are regulated by Sim in combination with other transcription factors, such as Zld (blue), Vvl, and Sox proteins (green). Sim provides both temporal (initiation at stage 11) and spatial (MG-restricted) control. (D) Sim-Secondary role: another model of midline primordium expression, depicted by *rho* and *aos* expression, requires initiation by *Notch* signaling (magenta), restriction to midline cells by Sim and other midline-specific factors, Motif-A Binding Protein (MABP) and Midline Specificity Regulator (MSR). Sox and Repeat Binding Protein (RBP) (green) also promote activity. (E) The *aos* gene is activated by a combination of Spi signaling (magenta) via direct binding of Pnt (green) and Sim (blue). Aos protein inhibits EGF signaling, restricting *spi*-dependent activity to those MG receiving the most Spi signal. (F) Sim and additional factors confer midline specificity. Mesectodermal specification is due to the combined action of dorsal-ventral patterning proteins (Dorsal, Twist, and Snail) and *Notch* signaling. This results in the midline expression of *sim* and two postulated regulatory genes whose identities are unknown. MSR is expressed only in midline cells and is not ectopically induced by *sim*. Another factor, SIR, is expressed in midline cells as a consequence of *sim* regulation, since it can be ectopically expressed by *sim* misexpression. Together these regulators are involved in directing midline-specific expression of the *rho-365* enhancer and combinations may be important in regulating additional midline-expressed genes.