A Comparison of Midline and Tracheal Gene Regulation during Drosophila Development

Sarah K. R. Long1, Eric Fulkerson1, Rebecca Breese1, Giovanna Hernandez1, Cara Davis1, Mark A. Melton2, Rachana R. Chandran3, Napoleon Butler1, Lan Jiang3, Patricia Estes1*

1 Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, United States of America, 2 Department of Biological and Physical Sciences, Saint Augustine’s University, Raleigh, North Carolina, United States of America, 3 Department of Biological Sciences, Oakland University, Rochester, Michigan, United States of America

Abstract

Within the Drosophila embryo, two related bHLH-PAS proteins, Single-minded and Trachealess, control development of the central nervous system midline and the trachea, respectively. These two proteins are bHLH-PAS transcription factors and independently form heterodimers with another bHLH-PAS protein, Tango. During early embryogenesis, expression of Single-minded is restricted to the midline and Trachealess to the trachea and salivary glands, whereas Tango is ubiquitously expressed. Both Single-minded/Tango and Trachealess/Tango heterodimers bind to the same DNA sequence, called the CNS midline element (CME) within cis-regulatory sequences of downstream target genes. While Single-minded/Tango and Trachealess/Tango activate some of the same genes in their respective tissues during embryogenesis, they also activate a number of different genes restricted to only certain tissues. The goal of this research is to understand how these two related heterodimers bind different enhancers to activate different genes, thereby regulating the development of functionally diverse tissues. Existing data indicates that Single-minded and Trachealess may bind to different co-factors restricted to various tissues, causing them to interact with the CME only within certain sequence contexts. This would lead to the activation of different target genes in different cell types. To understand how the context surrounding the CME is recognized by different bHLH-PAS heterodimers and their co-factors, we identified and analyzed novel enhancers that drive midline and/or tracheal expression and compared them to previously characterized enhancers. In addition, we tested expression of synthetic reporter genes containing the CME flanked by different sequences. Taken together, these experiments identify elements overrepresented within midline and tracheal enhancers and suggest that sequences immediately surrounding a CME help dictate whether a gene is expressed in the midline or trachea.

Citation: Long SKR, Fulkerson E, Breese R, Hernandez G, Davis C, et al. (2014) A Comparison of Midline and Tracheal Gene Regulation during Drosophila Development. PLoS ONE 9(1): e85518. doi:10.1371/journal.pone.0085518

Editor: Bassem A. Hassan, Vlaams Instituut voor Biotechnologie and Katholieke Universiteit Leuven, Belgium

Received September 15, 2013; Accepted November 28, 2013; Published January 20, 2014

Copyright: © 2014 Long et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants 0416812 and 0750620 from the National Science Foundation Grants website: http://www.nsf.gov/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pat_estes@ncsu.edu

Introduction

The genes expressed within a particular cell type control its developmental fate and physiological potential. Early in development, master control genes play pivotal roles in controlling cell fate and most master control genes are transcription factors that promote their own expression as well as a variety of downstream target genes. Each target gene, in turn, contributes to tissue development by regulating cellular processes, such as 1) morphology; 2) interactions with surrounding cells through signaling; 3) cell divisions and/or 4) the expression of additional genes. To understand how genes are differentially regulated within tissues, we compare the development and gene expression of two tissues in the Drosophila embryo: the central nervous system (CNS) midline and the trachea. In Drosophila, Single-minded (Sim) is the master control gene of CNS midline cells [1–3], while Trachealess (Trh) plays a large role in the development of the fly’s respiratory system, the trachea [4–6]. Both Sim and Trh are bHLH-PAS proteins and independently heterodimerize with a common partner, Tango (Tgo), before binding to DNA and activating transcription [7,8]. Tgo is also a bHLH-PAS protein and paradoxically, Sim/Tgo and Trh/Tgo both bind to a shared five base pair recognition sequence, ACGTG, called the CNS midline enhancer element (CME). Tgo is ubiquitously expressed, whereas Sim is restricted to the midline [9] and Trh to the trachea and a few other tissues, including the salivary gland, filzkorper and CNS [4,5]. In most cells, Tgo is located in the cytoplasm, but within cells that express one of its partner proteins, such as Sim or Trh, Tgo is transported to the nucleus and upregulated [10]. Once in the nucleus, Sim/Tgo and Trh/Tgo activate overlapping and distinct gene sets [7,8,11–13].

Single-minded and the Midline

The embryonic midline and trachea differ in many ways and the following is a brief summary and comparison of the development of these tissues during Drosophila embryogenesis. CNS midline cells are specified early in embryogenesis when sim is activated prior to gastrulation, in a single row of cells sandwiched between the mesoderm and ectoderm on each side of the embryo; cells called the mesectoderm [9]. Sim protein is first expressed during gastrulation as the two rows of mesectodermal
cells come together at the ventral midline. After meeting ventrally, midline cells invaginate to form a signaling center that organizes the CNS as it matures symmetrically on either side of the midline. As CNS axons differentiate, midline glia secrete Netrin (Net) A and B to attract axons to cross the midline [14–16] and then slit to prevent recrossing [17–19]. Some axons continually express roundabout (robo), the receptor for slit [18], at the growth cone surface and never cross the midline, whereas axons that cross the midline require commissureless (comm) to temporarily prevent robo localization at the growth cone, allowing them to cross [20–23]. During mid to late embryogenesis, midline cells differentiate into glia and six neural subtypes that can be distinguished based on their gene expression patterns (Fig. 1A–B) [11,24]. By the time the embryo hatches into a larva, most midline neurons have differentiated and begun to secrete subtype specific neurotransmitters and make connections with target tissues [24,25]. In addition, the midline glia have enwrapped and secured the CNS axons that cross the midline [1,26].

Trachealess and the Trachea

In Drosophila, the trachea are a network of air-filled tubes constructed during embryogenesis that function in gas exchange (reviewed in [27–30]). Tracheal cells can first be recognized during Drosophila gastrulation when ventral veinless (Vvl) and trh are activated by JAK/STAT signaling [31–34] within segmentally repeated tracheal pits or placodes [5,35]. Decapentaplegic (Dpp) and Epidermal Growth Factor (EGF) signaling limit the embryonic dorsal and ventral boundaries of the trachea, while wingless (wg) restricts the location of trachea within each segment [4,5,36]. As development progresses, terminal cells at the end of the growing tracheal tubes lead migration into tissues and specialized cells fuse to connect the separate, developing metameric trachea, creating a continuous tubular network. Fusion of lateral and dorsal trunks is facilitated by the Dysfusion (Dys) bHLH-PAS protein, another partner of Tgo [37–40] and after fusion, the two major tracheal tubes, called dorsal trunks, span the length of the embryo (Fig. 1C and D). Interestingly, insect trachea share functional and developmental similarities with the vertebrate vasculature. Both are interconnecting and branched tubular networks, function in gas exchange, and are patterned by related developmental genes and mechanisms [41]. For instance, signaling by fibroblast growth factor (FGF), called breathless (btl) in flies [42,43], plays a key role in the formation of both of these tissues. Btl is expressed in all tracheal cells and leading cells of nascent branches interact with neighboring tissues through their production of the FGF signal, branchless, which stimulates and guides branch formation [44]. FGF signaling,
together with the *Drosophila hypoxia inducible factor*, also guides later growth and branching of the trachea, driven, in part, by oxygen demands of tissues [41]. At the end of embryogenesis, the tracheal network fills with air and for the remainder of the fly’s life, the trachea delivers oxygen to its tissues.

Common and Distinct Genes and their Regulation within the Midline and Trachea

The functions and morphology of midline and tracheal cells differ, yet certain aspects of their embryonic development are similar. Both cell types are derived from the ectoderm (the midline is derived from the more specialized mesectoderm) and project long cellular extensions to form specialized contacts with many different cell types [45-48]. Moreover, midline glia and tracheal cells provide vital nutrients, growth factors and oxygen for active neurons within the mature embryo and larvae [6,26]. While Sim is restricted to the midline and Trh to tracheal cells within the embryo, many genes are expressed in both the midline and trachea, including the Vgl POU domain transcription factor, which is needed to activate genes in both tissues [49,50]. In addition, many signaling pathways, including Notch, FGF, EGF, engrailed, wg and hedgehog (hh) [11,12,26,47,51,52] provide positional cues to regulate development of various cell types within both tissues. Downstream components of these signaling pathways combine with Sim and Trh in unique ways to regulate different gene sets in the midline and tracheal cells. Differences between the two tissues are likely due to the presence of additional, unknown tissue specific proteins that combine with Sim and Trh in unique ways to control gene expression and alter cell activity. In support of this idea, exchanging the PAS domains between Sim and Trh indicates these domains determine target gene specificity, presumably by binding to co-factors restricted to either the midline or trachea [13]. This is consistent with the known properties of PAS domains, which bind many different molecules and co-factors to respond to the environment [53-55]. Such co-factors may cause the Sim/Tgo and Trh/Tgo heterodimers to recognize slightly different DNA binding sites within enhancer regions of target genes. The goal of these experiments is to understand how Sim and Trh bind the same protein partner and DNA sequence, yet activate different gene sets in midline and tracheal cells.

To compare regulatory functions of Sim/Tgo and Trh/Tgo during fly development, we selected genes expressed in the midline, trachea or both tissues, identified enhancers that control the expression of each gene and compared them to previously identified midline and tracheal enhancers. To test the importance of previously identified sequence motifs, we generated synthetic reporters that contain the CME combined with binding sites for other factors expressed in the midline or trachea. To further analyze these enhancer sets, we searched for novel motifs common to both, as well as motifs unique to either midline or tracheal genes. The results identify sequence contexts, both proximal and distal to the CME, which promote midline and/or tracheal expression.

Materials and Methods

Production of Midline and Tracheal Reporter Genes and Transgenic Strains

*Drosophila melanogaster* genomic sequences encompassing select genes expressed in the midline and trachea were compared across the 12 sequenced *Drosophila* genomes [36] using the USCS genome browser (genome.ucsc.edu). The sequences examined included all introns within a gene and the intergenic regions located between the midline gene and its neighboring upstream and downstream gene. Identified regions conserved in at least 11 of the 12 genomes were first amplified within fragments ranging from ~200–3500 bp using the primers listed in Table S1 and genomic DNA isolated from the yw*2* *Drosophila melanogaster* strain. These fragments were either cloned into the pSTBlue intermediate vector and then into the pHiStinger vector [57] with Xhol/KpnI digestion, or cloned into pCR/GW/TOPO (Invitrogen) and transferred into pMinigate using the Gateway system [58]. Minor changes to this cloning scheme are noted below. Transgenic fly lines were generated with the pHiStinger constructs using standard procedures and three independent lines analyzed for each GFP reporter gene. pMinigate constructs were injected into the F031 genomic destination site attP2 (68A1-B2) as previously described [59].

**CG33275.** The CG33275 ML577 fragment was generated by first digesting the CG33275 ML2544-GFP construct in pSTBlue with BglII, re-ligating it and then subcloning the remaining 577 bp fragment into pHiStinger. The CG33275 ML1312 fragment was generated from the CG33275 ML2544-GFP construct using KpnI/SalI digestion and blunt end ligation, which removed 1232 bp from the original 2544 bp construct (Fig. 2B). The remaining 1312 bp fragment was then subcloned into pHiStinger.

**liprin γ.** The liprin γ 1761 fragment was generated from the liprin γ 3414-GFP construct using SacII/BamHI digestion, as previously reported [59].

Production of Synthetic Reporter Genes

To generate synthetic reporters, the forward and reverse primer pairs listed in Table S2 were phosphorylated, annealed, ligated and multimers consisting of four copies were separated on 12% polyacrylamide gels, excised and purified with isobutanol extraction. The multimers were first cloned into EcorI-digested Bluescript KS and subsequently into pHiStinger using KpnI/BamHI digestion. Each reporter gene was introduced into the *Drosophila* genome using P element mediated transformation and the GFP expression pattern of at least three transgenic lines examined.

Immunohistochemistry and Confocal Microscopy

Embryos were collected and labeled with antibodies as previously described [60]. The following antibodies were used to localize proteins in *Drosophila* embryos: rat anti-sim (1:100) [10], rabbit anti-GFP (1:500; Molecular Probes, Life Technologies), mouse anti-GFP (1:200; Promega), rabbit anti-lys (1:400) [39]; rabbit anti-actin-skipped (odd; 1:400), Jim Skeath, Washington University School of Medicine, St. Louis, MO, USA; and anti-reversed polarity (repo; 1:30), anti-engrailed (1:1), and 2A12 (1:5) MABs (Developmental Studies Hybridoma Bank). Secondary antibodies were used at 1:200 and included anti-rat-Alexa568, anti-mouse-Alexa638, anti-rabbit-Alexa405 and anti-rabbit-Alexa488 (Molecular Probes). Images were obtained on a Zeiss 710 in the Cellular and Molecular Imaging Facility at NCSU.

Motif Identification

MEME [61] was used to identify common motifs in midline and tracheal enhancers. We included previously identified midline enhancers for sim, Toll, slit [62,63], rhomboid (rho) [50], btl [8], Vgl [64], roughed (rst) [65,66], wrapper [67], glioectin, organic anion transporter protein26f, liprin γ [59] and link [68], as well as the new midline enhancers reported here: CG33275, NetB, comm, escargot (esg) and Ectoderm3 (Ect3). The following previously identified tracheal enhancers were included: trh, Vgl early [34], Vgl autoregulatory [64], CG13196, CG13252, dys [58], btl [8], rho [50],...
Results

To understand how diverse genes are transcriptionally regulated in the midline, trachea or both tissues, we identified and compared enhancers of seven genes that are expressed during *Drosophila* development. The seven genes studied include three genes that encode axon guidance and synaptic proteins: *lprin c*, *comm* and *NetB;* a gene in the EGFR signaling pathway, *CG33275;* a G protein coupled receptor, *moody;* a cell death gene, *Ectoderm 3 (Ect3);* and finally, the *esg* transcription factors. Several of these contain large introns and are separated from other genes by large intergenic regions and, therefore, to facilitate the identification of midline

Figure 2. *CG33275* contains a midline enhancer that is separable and distinct from a nearby tracheal enhancer. (A) Locations of regions within the fifth intron of *CG33275* used to generate the reporter constructs are shown. A scale is indicated on top and the thick lines represent the regions analyzed in (B–T). White boxes represent exons and thin lines represent introns. The indented boxes indicate flanking genes, *CG7716* and *snoRNA,* and arrows indicate the start and direction of transcription. (B) Fragments used to generate reporter constructs are shown. The size of each fragment is indicated below it, vertical lines indicate locations of CMEs and red lines represent sequence blocks conserved in at least 11 *Drosophila* species. (C–T) Whole mount embryos were double-stained with anti-GFP (green; D, G, J, L and S), anti-*sim* (red; E, H, K), anti-*dys* (red; T) antibodies and monoclonal antibody 2A12 (red; M–Q) and analyzed by confocal microscopy. The overlap in expression is shown in yellow in the merge images (C, F, I, M–Q and R). Reporters (C–E) *CG33275 ML2544:GFP,* (F–H) *CG33275 ML1312:GFP* and (I–K) *CG33275 ML753:GFP* drove expression in midline glia. Midline glia can be identified by the overlap in expression of GFP and *sim* (arrows C–K) and are located on the dorsal side of the nerve chord. Midline neurons are located on the ventral side of the nerve chord and labeled by *sim,* but do not express *CG33275* or any of the *CG33275* reporter genes. (L–T) Monoclonal antibody 2A12 labels the tracheal lumen and the anti-*dys* antibody labels tracheal fusion cells. *CG33275 TR1297:GFP* is expressed in all trachea, beginning in the tracheal pits at stage 12 (L) and extending to late embryogenesis (M–N), while *CG33275 TR265:GFP* is expressed only in fusion cells (O–T), indicated by co-localization of GFP and Dys (arrows R–T). Lateral views of stage 16 transgenic embryos are shown; anterior is in the top, left hand corner and ventral is on the left, except (L), which is a dorsal view of a stage 12 embryo.

doi:10.1371/journal.pone.0085518.g002

link [60], and the new tracheal enhancers described here: *CG33275, NetB, lprin c, esg* and *moody.*
and tracheal enhancers, we searched for sequences conserved in a relatively large number of the sequenced *Drosophila* species [36]. We tested the ability of the conserved regions to drive expression in midline and tracheal cells by fusing them to *GFP* within the pHstinger or Mintgate enhancer tester vectors and generating transgenic fly lines. In certain cases, we also identified a minimal region capable of driving tissue specific expression. The composition and expression patterns of the identified enhancers are briefly summarized below.

CG33275

This gene is a guanyl-nucleotide exchange factor expressed in both the midline and trachea during embryogenesis [69,70]. The entire gene spans approximately 47 kb and consists mostly of large introns (Fig. 2A). We identified an enhancer within the fifth intron of CG33275 capable of driving high levels of *GFP* in midline glia and a separate and distinct tracheal enhancer downstream of the midline enhancer (Fig. 2A and B). The midline enhancer was identified by testing reporter genes CG33275 ML2544:GFP, ML753:GFP and ML577:GFP, and all but the CG33275 ML577:GFP reporter drove expression in midline glia (Fig. 2C–K), in a pattern similar to that of the endogenous gene [11]. The CG33275 ML753:GFP midline glial enhancer contains two regions conserved in 12 *Drosophila* species and one of these contains a CME (Fig. 2B). Sequences located just downstream of the midline enhancer drove high levels of *GFP* expression in a pattern similar to the endogenous gene [70]; in all tracheal cells beginning at stage 11 (Fig. 2L–N) and throughout larval stages (not shown). Both tracheal reporter genes CG33275 TRH1297:GFP (not shown) and the smaller CG33275 TRH1297:GFP reporter gene drove the same tracheal expression pattern (Fig. 2L–N). The CG33275 TRH1297:GFP reporter was not expressed in trachea or midline cells (not shown), whereas the CG33275 TRH265:GFP reporter was restricted to tracheal fusion cells (Fig. 2O–T), as demonstrated by the overlap in expression with *dys* (Fig. 2R–T). The CG33275 TRH1297:GFP reporter contains four CMEs, CG33275 TRH1297:GFP contains three of these and CG33275 TRH265:GFP contains one. All three of these reporters contain a region with a CME that is conserved across 12 *Drosophila* species (Fig. 2B). Dys, related to Trh, also heterodimerizes with Tgo and binds a site related to the CME, TCCTGT, and can weakly interact with the sequences, TCGTG as well as the CME (Table 1) [58]. Consistent with this, the CG33275 TRH265:GFP enhancer expressed in fusion cells contains two TCGTG Dys/Tgo sites conserved in 12 *Drosophila* species. In summary, CG33275 contains separable, but adjacent midline and tracheal enhancers, and the tracheal enhancer contains a subregion that drove expression restricted to fusion cells.

esg

esg is a zinc finger transcriptional repressor that regulates cell fate and development within the trachea and a subset of CNS cells, including the midline [35,71]. *esg* is expressed at high levels in the embryo and moderate levels in the larval central nervous system, larval/adult midgut and adult testis [70,72]. *esg* is rather isolated from other genes within the *Drosophila* genome and its next nearest upstream and downstream neighbors are ~15–25 kb away (Fig. 3A). We examined this entire region to search for midline and tracheal enhancers and identified two, separable tracheal enhancers, *esg* TR C1:GFP (Fig. 3A–F) and *esg* TR C7:GFP (Fig. 3A, B and G–J) downstream of the coding sequence and another separable and distinct midline enhancer, *esg* ML C2:GFP (Fig. 3A, B and K–N), adjacent to and downstream of the *esg* TR C1 tracheal enhancer. The endogenous *esg* gene is expressed in tracheal fusion cells during embryogenesis and first detected during branch migration [73]. The tracheal enhancers identified here drive expression only late in embryogenesis and during larval stages. The *esg* TR C1:GFP reporter is expressed sporadically in fusion cells (Fig. 3C–F), while the *esg* TR C7:GFP reporter is expressed in all tracheal branches and fusion cells and sporadically in the dorsal trunks (Fig. 3G–J).

In addition, the *esg* ML C2:GFP reporter drove a unique expression pattern in the midline, where it is expressed in both anterior and posterior midline glia and the median neuroblast and its progeny (Fig. 3K–N). This pattern is consistent with that of the endogenous *esg* gene, known to be expressed in a subset of mesectodermal and midline primordial cells [11]. In addition to these three enhancers, we found additional *esg* enhancers that drove expression in other embryonic tissues (Table S3).
182 bp core region contains a CME and conserved subregion that activates high levels of expression in both midline and trachea cells. Analysis of the expression pattern of the endogenous liprin \textit{c} gene indicates that it is either not expressed, or expressed at low levels, within the trachea during embryogenesis [70,74]. This, taken together with 1) the high level of \textit{GFP} expression in tracheal cells observed with the \textit{liprin c} \textit{182:GFP} reporter gene and 2) the diverse tracheal expression pattern of the larger \textit{liprin c} reporter genes, suggest that this region may only drive tracheal expression when isolated from surrounding sequences. Because multiple copies of the CME within a reporter gene, can drive expression in both the midline and trachea (see below), one function of sequences flanking the CME within enhancers is to limit expression of the gene to certain cell types. In summary, these

![Figure 3](image_url)

**Figure 3.** An \textit{esg} genomic region contains a midline enhancer that is separable and distinct from two \textit{esg} tracheal enhancers. (A) Genomic regions surrounding \textit{esg} used to generate the reporter constructs and (B) the \textit{esg} enhancers are shown as above. Gray boxes represent 5’ and 3’ untranslated regions. (C–J) Live larvae were analyzed by confocal and differential contrast microscopy and ventral views of the (C–F) \textit{esg TR C1:GFP} and (G–J) \textit{esg TR C7:GFP} reporters are shown. In larvae, the \textit{esg TR C1:GFP} reporter is expressed sporadically in fusion cells (arrow in E) and the \textit{esg TR C7:GFP} reporter is expressed in all tracheal branches and sporadically in the dorsal trunks (G–J), but consistently in fusion cells (arrows I). (K–N) Whole mount \textit{esg ML C2:GFP} reporter embryos were stained with an anti-\textit{GFP} antibody (green; L), engrailed monoclonal antibody (blue; M) and anti-\textit{sim} antibody (red; N) and analyzed by confocal microscopy. The overlap in expression is shown in the merge image (K). Anterior midline glia express \textit{GFP} and \textit{sim} and are located dorsally within the nerve chord. Posterior midline glia that normally undergo cell death during this time can still be visualized with \textit{GFP} (three cells surrounding star in N), but not \textit{sim} or \textit{engrailed}. The MNB and its progeny express \textit{sim}, \textit{engrailed} and \textit{GFP} and are located ventrally within the nerve chord (arrowheads in K–N). Lateral view of a stage 16 transgenic embryo is shown; anterior is in the top, left hand corner and ventral is on the left. doi:10.1371/journal.pone.0085518.g003
experiments further define the minimal sequences needed for expression in the CNS midline within the previously identified liprin\(^c\) enhancer [59]. Moreover, these minimal sequences, when isolated from the genome and placed within reporter genes, can activate expression in tracheal cells as well.

NetA and B

NetA and NetB are signaling molecules secreted by midline glia that attract axons to cross the midline and also function in glial migration [14,15,75,76]. Both genes are expressed in many tissues, including midline glia [15], the larval trachea and adult nervous system [77]. The Net797:GFP reporter identifies a midline and tracheal enhancer located between NetA and NetB (Fig. 5A) that

Figure 4. Liprin\(^c\) contains a conserved enhancer sufficient to drive expression in both midline glia and trachea. (A) The genomic regions within the first intron of liprin\(^c\) used to generate the reporter constructs and (B) the liprin\(^c\) enhancers are shown as in Fig. 1 and previously reported [59]. (C–T) Whole mount embryos were double-stained with an anti-GFP antibody (green: D, G, J, M, P and S) and monoclonal antibody 2A12 (red; E, H, K and N), anti-dys (red; Q) or anti-sim (red; T) and analyzed by confocal microscopy. The overlap in expression is shown in yellow in the merge column (C, F, I, L, O and R). Even though the liprin\(^c\) 3141:GFP and liprin\(^c\) 889:GFP reporters are expressed in the dorsal trunk (arrows C–E and I–K) and dorsal and ventral branches, liprin\(^c\) 1781:GFP is restricted to mostly fusion cells of the dorsal trunk (arrows F–H and O–Q). The liprin\(^c\) 182:GFP reporter is expressed in all tracheal cells (L–N), midline glia (arrows R–T) and a few midline neurons in certain segments (arrowheads R–T). Lateral views of stage 16 transgenic embryos are shown; anterior is in the top, left hand corner and ventral is on the left.

doi:10.1371/journal.pone.0085518.g004
drove expression in midline glia (Fig. 5C–E) and trachea cells outside the dorsal trunk (Fig. 5F–K). This enhancer contains three CMEs and three highly conserved regions (Fig. 5B). Therefore, in contrast to the CG33275 and eg enhancers and similar to the liprin γ enhancer described above, the single Net enhancer drove expression in both the midline and trachea. Moreover, the tracheal expression pattern provided by this enhancer is unique and highest in the visceral and dorsal branches and low or absent in the dorsal trunks (Fig. 5F–K). We also identified several Net enhancers that drive expression in tissues outside the midline and trachea (Table S3).

**comm**

comm functions in synapse assembly and axon guidance by controlling the subcellular localization of membrane receptors. In particular, comm controls the slit receptor, roundabout, as CNS axons navigate the midline to ensure they cross the midline only once [20,22,73–80]. comm is expressed at high levels in midline glia and transiently in lateral CNS axons [80]. A midline enhancer is located in the 3' untranslated region of comm (Fig. 6A and B), identified by testing comm2575:GFP (Fig. 6C–E), comm737:GFP (not shown), comm693:GFP (Fig. 6F–H), comm443:GFP (Fig. 6L–N) and comm267:GFP (Fig. 6I–K). All but the comm737:GFP and comm267:GFP reporters drove expression in midline cells. Each of the reporters that are active in the midline drive GFP expression in slightly different subsets of midline cells: comm2575:GFP is expressed in midline glia, with variable expression in midline neurons (Fig. 6C–E), comm693:GFP is expressed predominantly in a subset of midline neurons (Fig. 6F–H) and comm443:GFP is expressed in all midline cells (Fig. 6L–N). Only the comm443:GFP reporter drove expression in the trachea and tracheal expression initiated during early larval development and persisted throughout all larval stages (Fig. 6O–T). Therefore, comm contains a single enhancer that drives expression in both the embryonic midline and larval tracheal cells.

moody

moody is a rhodopsin and melatonin-like G-protein coupled receptor, found at the blood-brain barrier in adult flies [81] and that functions in germ cell migration in the embryo [82]. Moody is expressed in larval trachea, the larval/adult CNS, as well as many other tissues [77]. We tested three reporter genes: moody1970:GFP, moody1221:GFP and moody608:GFP (Fig. 7A and B) and found that moody1970:GFP is expressed in the dorsal vessel (Fig. 7C–E), but only moody1221:GFP drove expression in the dorsal trunks of the trachea, with expression highest in the posterior region of the embryo (Fig. 7F–H), similar to the liprin γ 3141 enhancer (Fig. 4C–E). Also similar to the liprin 3141:GFP enhancer [59], the moody1221:GFP (Fig. 7L–N) and moody1970:GFP (not shown) enhancers are expressed in lateral CNS glia. moody608:GFP drove expression in the fat body (not shown), but is not expressed in the trachea (Fig. 7I–K). The identified moody1221:GFP tracheal enhancer contains two CMEs, although they are not highly conserved. This enhancer does not drive midline expression, rather sequences within the moody enhancer restrict expression to the trachea.

Ect3

The Ect3 protein is a galactosidase expressed in midline glia, [11] as well as other tissues, that regulates autophagic cell death

---

**Figure 5. Net contains an enhancer that drives expression in both midline glia and trachea.** (A) A genomic region located between NetA and NetB was used to generate a reporter construct and (B) the Net enhancer is shown. (C–K) Whole mount embryos were double-stained with an anti-GFP antibody (green: D, G and J) and anti-sim (red: E) or monoclonal antibody 2A12 (red: F and K) and analyzed by confocal microscopy. The overlap in expression is shown in yellow in the merge column (C, F and I). The Net797:GFP reporter drove expression in midline glia (arrows C–E; ganglionic branches of trachea are also visible in the image in green) and occasionally midline neurons within some segments (arrowheads C–E). The tracheal expression pattern of this reporter is unique in that GFP is high in tracheal cells, except cells within the dorsal trunk (F–K). Lateral views of stage 16 transgenic embryos are shown; anterior is in the top, left hand corner and ventral is on the left.

doi:10.1371/journal.pone.0085518.g005
Because Ect3 is located within the first intron of Tachykinin (Tk), the identified midline enhancer is found just upstream of Ect3 as well as within the first intron of Tk (Fig. 8A and B). Tachykinin (Tk) is a neuropeptide hormone expressed at high levels during 18–24 hours of embryogenesis, early larval stages and in the adult male [77]. The midline enhancer identified here likely regulates expression of the endogenous Ect3 gene, because only Ect3, and not Tk, is expressed in the embryonic midline [11,70]. This midline enhancer is sensitive to small changes in sequence, such that various subregions drive different midline expression patterns.

The Ect3 3194:GFP reporter contains the region bordered by the first exon of Tk on the 5’ end and the Ect3 transcription start site on the 3’ end (Fig. 8A and B). This reporter (not shown), as well as the Ect3 1955:GFP (Fig. 8C–H) and Ect3 1456:GFP (Fig. 8I–N) reporters drive high levels of GFP expression in all midline cells, with the exception of one iVUM neuron. Four other reporters are expressed in a more limited set of midline cells: 1) the Ect3 2311:GFP (Fig. 8O–T) and 2) Ect3 517:GFP (Fig. 8U–Z) reporters drove expression in midline glia, MP1 neurons, the H-cell and spotty and variable expression in the H-cell sib, while the 3) Ect3

Figure 6. The comm cis-regulatory region contains an enhancer that drives expression in the embryonic midline and larval trachea. (A) Genomic regions surrounding and within comm were used to generate the reporter constructs and (B) the comm enhancer is shown. The closest gene upstream of comm is CG6224, which is 79,892 bp away. (C–N) Whole mount embryos were double-stained with anti-GFP (green: D, G, J and M) and anti-sim antibodies (red; E, H, K and N) and analyzed by confocal microscopy and the overlap in expression is shown in yellow in the merge column (C, F, I and L). (C–E) comm2575:GFP and (L–N) comm443:GFP are expressed in both midline glia (arrows) and midline neurons (arrowheads), while (F–H) comm693:GFP is restricted to some midline neurons (arrowheads). The comm267:GFP (i–k) and comm737:GFP (data not shown) reporters are not expressed in the midline. Lateral views of stage 16 transgenic embryos are shown; anterior is in the top, left hand corner and ventral is on the left. (O–T) comm443:GFP is also expressed in the tracheal dorsal trunks (arrows in O and S) as well as other tracheal branches (arrow in R). Live larvae containing the comm443:GFP reporter were analyzed by confocal and differential contrast microscopy and dorsal views are shown.

doi:10.1371/journal.pone.0085518.g006
572:GFP (Fig. 8A'-F') reporter drove expression in some midline glia, MP1 neurons and some of the progeny of the MNB and 4) the Ect3 1071:GFP (Fig. 8G'-L') reporter drove spotty and variable expression in only a few midline cells within some segments. Taken together, the data indicate the Ect3 enhancer contains sequences that promote expression in all midline cells. The endogenous Ect3 gene is expressed in midline glia [11] and all of the Ect3 reporters drive expression in these cells. The 572 bp Ect3 enhancer contains three CMEs and a highly conserved subregion that can combine with another, downstream subregion located within both the Ect3 1955:GFP and Ect3 1456:GFP reporters, to enhance expression in certain midline cells. In addition, the 517 bp region can drive expression in midline cells, despite the absence of any CMEs. Therefore, this region of the genome contains multiple subsections that combine to drive expression in midline cells.

In summary, ten enhancers were identified: six of the enhancers drove expression in the midline, seven in the trachea and three in both the midline and trachea (Table 2). CG33275 and esg each contain adjacent, separable midline and tracheal enhancers; whereas liprin, Net and comm each contain one enhancer that drove expression in both the midline and trachea. The moody
Figure 8. The Ect3 cis-regulatory region contains a midline enhancer sensitive to context. (A) The genomic region upstream of Ect3 (and within the first intron of Tk) used to generate the reporter constructs and (B) the Ect3 enhancers are shown. (C–L) Whole mount embryos were stained with anti-sim (red; F, I, R, X, D' and J'), anti-odd (blue; G, M, S, Y, E' and K') and anti-GFP (green; H, N, T, Z, F' and L') antibodies and analyzed by confocal microscopy. The overlap in expression is shown in the merge columns: all three antibodies (C, I, O, U, A' and G'), sim and GFP (D, J, P, V, B' and H') and GFP and odd (E, K, Q, W, C' and I'). Odd is expressed only in MP1 midline neurons. Both Ect3 1955:GFP (C–H) and Ect3 1456:GFP (I–N) drive expression in the all midline cells, with the exception of a single IVUM. Ect3 2311:GFP (O–T) and Ect3 517:GFP (U–Z) are restricted to some midline glia, MP1s neurons, the H cell and H cell sib. Ect3 572:GFP (A’–F’) is expressed in midline glia, some MP1 neurons and some of the MNB and its progeny. Finally, Ect3 1071:GFP (G’–L’) drives only spotty midline expression. Lateral views of stage 16 transgenic embryos are shown; anterior is in the top, left hand corner and ventral is on the left. The midline expression pattern of each reporter is shown schematically on the right.

doi:10.1371/journal.pone.0085518.g008
enhancer drove expression only in the trachea and the $Ect3$ enhancer drove expression only in the midline. Despite providing expression in overlapping cell types, each enhancer drove a unique expression pattern within the midline and trachea. Next, these enhancers, together with previously reported enhancers discovered by several groups, were combined to search for overrepresented motifs that may correspond to binding sites for transcription factors that activate or repress genes in the midline and/or trachea.

Proximal CME Sequences

A long-term goal is to use the midline and trachea as models to study how transcription factors combine with cell type-specific cofactors to regulate unique gene sets, and in this way, dictate development of unique tissues. Including the ten enhancers identified here, nineteen different midline enhancers and nineteen tracheal enhancers have been identified. To identify sequences that promote or inhibit CME utilization in either the midline or trachea, we analyzed the enhancers in two different ways. First, we searched sequences directly flanking the CME within defined enhancers to determine if these sequences could predict whether a particular CME is utilized by Sim/Tgo or Trh/Tgo and secondly, we searched the smallest region sufficient to drive expression in a tissue for reiterated motifs that may help restrict or promote gene expression in the midline and trachea.

Results from this analysis indicate that the nucleotide located directly upstream and downstream of the CME are strong, but not absolute, determinants of whether the CME is utilized in the midline or trachea (Table 3). We found sixty-six CMEs within all the enhancers examined here and 34/66 consisted of the sequences ACGTGTC, TACGTGA or TACGTGC (CME underlined), while the sequences ACGTGG, GACGTG, TACGTGG were not found in any of the enhancers, suggesting that Sim/Tgo and Trh/Tgo may not bind these sequences (Table S4). Enhancers that drive only midline expression, most often contain the sequence (A/G/T)ACGTGC, while enhancers that solely drive tracheal expression contain the sequence (A/T)ACGTGC/ACGTGC (Table 3). Therefore, the nucleotides immediately flanking the core CME may be one determinant that controls if a PAS heterodimer will bind this sequence within different cell types. We further investigated this by constructing and testing the expression pattern of synthetic reporter genes.

Synthetic Genes

Enhancers are modular and contain multiple binding sites for many activators and repressors that work together in large multigene complexes to regulate transcription in different cell types. Nevertheless, individual binding sites of a limited number of transcription factors are sufficient to drive expression in certain tissues, particularly when present in more than one copy. Relevant to this study, four copies of the CME fused to β-galactosidase or GFP is sufficient to drive reporter expression in both the midline and trachea [7,8,59,63]. Our previous results indicated that the context surrounding the CME within such multimerized constructs had a large impact on the reporter gene expression pattern [59]. To confirm and extend the results obtained with endogenous enhancers, we analyzed the expression pattern of additional

### Table 3. Proximal CME context in midline and tracheal enhancers.

| Most common | Number | Never found | Tissue | Predicted Consensus |
|-------------|--------|-------------|--------|---------------------|
| A ACGTG C  | 14     | A ACGTG G  | midline (A/G/T)ACGTGC |        |
| T ACGTG A  | 10     | G ACGTG T  | trachea (A/T)ACGTG (A/C/T) |        |
| T ACGTG C  | 10     | T ACGTG G  | both (A/T)ACGTG C |        |

| Tissue | Predicted Consensus |
|--------|---------------------|
| midline | (A/G/T)ACGTGC |
| trachea | (A/T)ACGTG (A/C/T) |
| both | (A/T)ACGTG C |

Sixty-six CMEs were found in all midline and tracheal enhancers examined. The nucleotides found directly 5’ and 3’ of the CME within the enhancers and the number of times that sequence was found in all the midline and tracheal enhancers are listed. The three sequence contexts found in the left column represent 52% of the CMEs found within all enhancers (34/66; Table S4), while other sequence contexts were not found in any enhancers. Sequences flanking the sixty-six CMEs were used to derive a consensus sequence for genes expressed in the midline, trachea or both tissues.

doi:10.1371/journal.pone.0085518.t003

Table 2. Ten identified midline and tracheal enhancers.

| Gene | Size | Position | Tissue | Midline cells | Tracheal cells |
|------|------|----------|--------|---------------|---------------|
| CG32275 | 753 | intron 5 | midline | glia (st. 13) | – |
|       | 1297/265 | 5 | trachea | – | all (st. 11)/fusion cells (st. 13) |
| e95920 | 3092 | ~5 kb downstream | midline | glia, MNB and progeny (st.12) | – |
| esg | 3586 | downstream | trachea | – | larval fusion cells (1st instar) |
|       | 3468 | ~20 kb downstream | trachea | – | larval fusion cells, secondary branches (1st instar) |
| lpr | 889/182 | intron 1 | both | glia (st. 12)/glia and sporadic in MNB and progeny (st. 12) | DT(st. 15)/all (st. 13) |
| Net | 797 | ~10 kb downstream | both | mostly glia (st. 12) | all but DT (st. 12) |
| comm | 443 | 3’ untranslated region | both | all midline cells (st. 10) | larval dorsal trunk and some secondary branches (st. 17) |
| moody | 1221 | intron 1 | both | – | posterior DT (st. 15) |
| Ect3 | 517 | upstream of Ect3/1 intron 1 of Tk | midline | glia, MP1s, H cell and H cell sib (st. 10) | – |

The size of the minimal fragment with enhancer activity.

The position of the enhancer relative to the gene and 5’midline and 4’tracheal cells that exhibit enhancer activity are indicated.

The stage of development when reporter expression is first observed is indicated in parentheses. The absence of expression in the midline or trachea is indicated with a dash.

doi:10.1371/journal.pone.0085518.t002
The synthetic sequences were modifications of CMEs derived from either the wrapper (synth 1–4 and synth 6) or Toll midline enhancers (synth 5, Toll and synth 7–12). We chose these particular sequences because both had been tested previously within synthetic reporter genes and drove different patterns of expression. The CME and flanking sequences found in the wrapper enhancer, when multimerized four times, drives expression only in the midline [59], while the CME and flanking sequences found in the Toll enhancer drives expression in both the midline and trachea [62].

To understand which sequences within the previously published synthetic constructs are responsible for the two different expression patterns, we tested additional synthetic reporter genes. The sequence context surrounding the CME in the wrapper enhancer was tested using two approaches. First, the 70 bp minimal wrapper enhancer was divided into two sections and tested independently: synth 1 contained sequences 7–53 and synth 2 contained sequences 28–70 of the wrapper minimal enhancer. Both of these constructs contain the single CME and flanking sequences found in the wrapper enhancer and both of these multimerized reporters were expressed in the midline, but not the trachea (Fig. 9A–H). However, the expression pattern within the midline differed and synth 1 drove expression in all midline cells (Fig. 9A–C), while synth 2 drove expression restricted to the midline glia (Fig. 9K’–M’). These nucleotide changes also did not eliminate the midline expression pattern, although synth 6, containing the Vvl binding site, drove expression in midline neurons, but not midline glia (Fig. 9K’–M’).

In summary, four of five synthetic constructs containing the sequence AAGCTGC, were expressed in midline cells only (Fig. 9A–P and Table 4), while the fifth was not expressed in either the midline or trachea (Fig. 9Y–B’). Four of six synthetic constructs containing the related sequence, TAGCTGC, drove expression in both midline and tracheal cells (Fig. 9C’–R’) and the other two drove expression only in trachea (Fig. 9W’–D’). Finally, three synthetic genes containing the sequence, GACGTGC, each exhibited a different expression pattern: synth 3 was expressed only in the midline (Fig. 9Q–T); Sox in neither tissue (Fig. 9U–X) and synth 10 in both the midline and trachea (Fig. 9S’–V’), suggesting changing the first three nucleotides within this 14 bp region from GTT found in synth 1 to GCC within synth 6 almost completely eliminated midline expression (Fig. 9Y–B’).

As mentioned above, the multimerized Toll CME and flanking sequences drives reporter expression in all midline and tracheal cells [62] (Fig. 9C’–F’). We tested whether adding binding sites of known midline transcription factors affected the expression pattern of this synthetic reporter gene. For this, an Engrailed binding site (TAATTA; [84]) was added to synth 7, a binding site for the POU domain transcription factor, Vvl (GGTGCAT; [64]) was added to synth 8 and binding sites for the Suppressor of Hairless transcription factor (CGTGGGAACCGAGCTGAAAAG-TAAAGTTTCTCACACA; [83]) within synth 9 (Table 4).

Table 4. Sequence of synthetic reporter constructs.

| Midline only          | Sox CACAATGACGTGCCCAGAGA | synth 6 | ATTACACTCTGCCGTCAGGAGAACGTGCTGTGGCATATTCCGA |
|-----------------------|--------------------------|---------|---------------------------------------------|
| synth 1               | ATTACACTCTGCCGTCAGGAGAACGTGCTGTGGCATATTCCGA |
| synth 2               | TTCAGAGAACEGTCGCTGTGGCATATTCCGAGATAATACTCATTTT |
| synth 3               | GCCACACTCTGCCGTCAGAGAACGTGCTGTGGCATATTCCGA |
| synth 4               | ATTACACTCTGCCGTCAGGAGAACGTGCTGTGGCATATTCCGA |
| synth 5               | TATGCCAAATGACATTTGAGAAATTACGCCAGTGGGCCAGACCA |
| Neither Midline nor Trachea | Sox CACAATGACGTGCCCAGAGA | synth 6 | ATTACACTCTGCCGTCAGGAGAACGTGCTGTGGCATATTCCGA |
| Both Midline and Trachea | Toll AAATTTGACGTGCCCAGAGA | synth 7 | AATAATTTGACGTGCCCAGAG TaoaA |
| synth 8               | AAATTTGACGTGCCCAGAGAATTGACGTGCCCAGAGA |
| synth 9               | AAATTTGACGTGCCCAGAGAATTGACGTGCCCAGAGA |
| synth 10              | AATAATTTGACGTGCCCAGAGAATTGACGTGCCCAGAGA |
| Trachea only          | synth 11 AAATTTGACGTGCCCAGAGAATTGACGTGCCCAGAGA |
| synth 12              | AAATTTGACGTGCCCAGAGAATTGACGTGCCCAGAGA |

Sequences of synthetic constructs multimerized four times and fused to GFP within reporter constructs are listed according to the tissue that expressed each synthetic reporter (Fig. 9). The CME is enlarged within each sequence. Sites important for midline expression within the wrapper enhancer [67] are underlined in synth 1–6 and include putative binding sites for Sox (ATTTG), pointed (CTCTCCG) and unknown (AAAA) transcription factors. Binding sites for engrailed (TAATTA), Vvl (GGTGCAT) and Suppressor of Hairless (GTGGGAACCGAGCTGAAAAG-TAAAGTTTCTCACACA) were added to the Toll CME sequence and shown in bold in synths 7–9.

doi:10.1371/journal.pone.0085518.t004
that this sequence is more sensitive to effects of additional sequences flanking the CME, compared to the other contexts. Taken together, the results suggest that the nucleotides immediately upstream and downstream of the CME had the largest impact on whether GFP was expressed in the midline or trachea. In most cases, the spacing and sequences between the CMEs did not affect whether or not the synthetic reporter was expressed in the midline or trachea, but instead, these sequences controlled which cell types within the midline or trachea, expressed GFP.

These results, together with those of the endogenous enhancers suggest that sequences proximal to the CME are strong, but not absolute, predictors of midline or trachea expression. Additional sequences, more distal to the CME, also impact CME utilization, as well as control which cell types express the gene.

Identification of Overrepresented Midline and Tracheal Motifs

To identify motifs other than the CME overrepresented within midline and tracheal enhancers, we used MEME (http://meme. ebi.edu.au/meme/cgi-bin/meme.cgi  [61]). Examination of enhancers that drive expression in both tissues together with enhancers that drive expression only in the midline identified
three overrepresented midline motifs (Fig. 10 and Table 5). In addition, MEME analysis of enhancers expressed in both tissues together with enhancers that drive expression only in the trachea, led to the identification of a single overrepresented tracheal motif. All four motifs consist of simple sequence repeats: midline motif 1 is 22 bp long, consists of repeating TG residues and is present 18 times in the 19 midline enhancers; midline motif 2 is 31 bp, consists mostly of T residues and is found 30 times; and midline motif 3 is 12 bp, consists of four repeats of the trinucleotide TGC and is found 25 times. The identified tracheal motif is 22 bp long, consists mostly of G residues and is found 16 times in the 19 tracheal enhancers examined. To ensure these results were not biased by including enhancers of variable sizes (336–3586 bp), we compared the above results to those obtained after restricting the search to only the smallest midline and tracheal enhancers identified, and excluded enhancers that function in both tissues. The same motifs were identified using this approach.

Discussion

DNA sequences located within introns and intergenic regions are known to regulate transcription and package DNA; however, many aspects of these processes remain unknown. Enhancers that control gene expression patterns are modular and contain binding sites for transcription factors that function in a combinatorial manner [86–90]. The array of transcription factors expressed within a particular cell type, and available to bind enhancers, depends upon the cell's position in the embryo as well as its developmental history. Identifying shared properties of enhancers active within a given cell type is challenging because most genes display their own unique expression pattern. Moreover, transcription factor binding sites can be combined in multiple ways to generate a similar expression pattern [91]. As a result, the complexity of gene expression patterns is often reflected by a complex and unpredictable organization of cis-regulatory sequences. Untangling this complexity to reveal how enhancers integrate positional, environmental and physiological information to regulate gene expression is needed to understand how organisms adapt to their internal and external environments at the molecular level.

Each enhancer described here contained a unique constellation of transcription factor binding sites and, as a result, drove a unique expression pattern in midline and tracheal cells. By analyzing and comparing available midline and tracheal enhancers, we have identified sequences, both proximal and distal to the CME, which provide expression in one tissue or the other. These reporters can be exploited in the future to identify transcription factors that bind to the enhancers using techniques such as chromatin immunoprecipitation, the yeast one hybrid assay and mutant genetic backgrounds. In addition, over one thousand GAL4 lines have been identified that drive expression in embryonic midline cells [92], providing a rich resource for extending these studies.

We have identified enhancers that drive expression restricted to midline glia, midline neurons, all embryonic tracheal cells, tracheal fusion cells, the posterior dorsal trunk, lateral tracheal branches, terminal cells or larval trachea cells and are activated at different stages of development (Table 2). In addition to identifying new motifs that may bind characterized or novel transcription factors, these studies provide tools for expressing transgenes in specific midline and tracheal subtypes for experimental purposes. When combined with toxins, RNAi or fluorophores, these sequences can be used to ablate cells, knockdown expression of specific genes and/or specifically label midline or tracheal subtypes. Moreover, genes within orthologous vertebrate tissues, such as glia and blood vessels, are regulated by similar regulatory networks [6]. Comparing midline and tracheal regulatory networks with networks that impact related tissues in other organisms will reveal how functionally distinct tissues are generated.

### Table 5. Midline and tracheal motifs identified with MEME.

| Motif Name | Number of enhancers examined | Number of enhancers containing site | E value | Total number of sites | Length of site |
|------------|-----------------------------|------------------------------------|---------|-----------------------|---------------|
| Midline 1  | 19                          | 9                                  | 1.2 e-05| 18                    | 22            |
| Midline 2  | 19                          | 14                                 | 7.2 e-34| 50                    | 31            |
| Midline 3  | 19                          | 12                                 | 1.2 e-10| 25                    | 12            |
| Trachea    | 19                          | 8                                  | 1.4 e-10| 16                    | 22            |

MEME analysis was used to identify motifs overrepresented in midline and tracheal enhancers. Three motifs were found in midline enhancers and one in tracheal enhancers (Fig. 10). Results from twelve enhancers that drive expression in the midline together with seven enhancers that drive expression in the midline and trachea or twelve enhancers that drive expression in the trachea and seven enhancers that drive expression in the midline and trachea are shown, as well as. The number of enhancers examined, number of enhancers containing the motif, likelihood of finding the motif by chance, number of times the site was found in all the enhancers examined and length of the identified site. doi:10.1371/journal.pone.0085518.t005
heterodimer binds the site. The first crystal structure of a bHLH-PAS heterodimer bound to DNA reveals that the recognition sequence of the human Clock/Brain bHLH-PAS heterodimer actually consists of seven base pairs, rather than five [100]. This is consistent with results reported here that suggest Sim/Tgo and Trh/Tgo heterodimers preferentially bind highly related, but slightly different seven base pair sequences (Tables 3 and 4). In addition, experiments with fly Sim and human Tgo, called Aryl hydrocarbon receptor nuclear translocator protein (Arnt), using the Systematic Evolution of Ligands by Evolutionary Enrichment (SELEX) approach, identify the sequence DRCGTG (D = A, C or T and R = either purine) as the Sim/Tgo binding site [101]. Our results agree with this, although the consensus sequence we identify by examining known enhancers, is shifted by one nucleotide (DACGTCG; Table 3). In the midline and tracheal enhancers, we found sixty-six copies of the CME, ACGTG, and forty-eight copies of the related sequence, GCGTG, also identified in the SELEX experiments (Table S5). Half of these GCCGTG sites fit the seven bp consensus TGGCTGR and future experiments are needed to determine their importance within the various enhancers. Our results indicated that the CME context favored within midline and tracheal enhancers as well as enhancers active in both tissues, was very similar (Table 3), yet clearly distinct from binding sites of other bHLH and bHLH-PAS heterodimers (Table 1). Based on the expression pattern of certain reporter genes examined here, the same CME may be bound by Sim/Tgo in the midline and Trh/Tgo in the trachea within certain enhancers. Within other contexts, the CME appears to be discriminated by these different heterodimers, because some enhancers drive expression in only one tissue or the other.

Enhancer Complexity

Results from both endogenous enhancers and the synthetic reporter genes confirm the importance of the proximal sequences in limiting expression to either the midline or trachea. While the proximal context of the CME plays a role, additional sequences clearly combine with the CME to ultimately determine if an enhancer is functional in the midline or trachea. Taken together, these results indicate that proximal motifs combine with additional sequences not only to determine whether or not a gene is expressed in the midline or trachea, but also to determine which cellular subtypes express the gene and when it is activated within a tissue. Future experiments will reveal if 1) changing the sequence, AACGTGC, to TACGTGC within a midline enhancer will cause the enhancer to drive expression in trachea as well and 2) if changing the sequence, TACGTGC, to AACGTGC within an enhancer that drives expression in both the midline and trachea, will restrict expression to only the trachea. Sequences proximal to the CME likely affect the affinity of either Sim/Tgo and/or Trh/Tgo heterodimers to the DNA, but binding sites for additional factors that interact cooperatively to stabilize an entire transcription complex are needed for high levels of expression within a particular cell. Moreover, recent experiments indicate that enhancers containing multiple CMEs are activated earlier in the embryonic midline than enhancers containing only one CME [102]. The authors of this study suggest Sim/Tgo binding sites may be sufficient for activation in the early embryo, but that binding sites for additional transcription factors must combine with the CME to drive expression within the later, more complex embryo.

The experiments described here as well as previous experiments indicate that the CME is not always necessary for either midline or tracheal expression. A number of enhancers that drive expression in both tissues do not contain a CME, including: 1) a 517 bp autoregulatory Vel enhancer that drove expression in both the midline and trachea [64], 3) another, separate tracheal enhancer of Vel [34], 5) a trh autoregulatory enhancer 4) the link enhancer, after its sole CME has been destroyed [68], 5) a vvi tracheal enhancer [58], 6) a tracheal enhancer of CG15252, 7) a tracheal enhancer of CG15196, and 8) the 517 bp EBi midline enhancer described here (Fig. 8). These sequences may be capable of driving midline and tracheal expression due to the presence of unknown, low affinity binding sites for Sim/Tgo and Trh/Tgo, or binding sites for other midline and tracheal transcription factors that can help recruit PAS heterodimers to the enhancer. To understand how a combination of binding sites that does not include the CME can drive expression in the midline and trachea, as well as how CMEs are distinguished by Sim/Tgo and Trh/Tgo heterodimers, we searched and found other regulatory motifs, both proximal and distal to the CME in midline and tracheal enhancers. Future experiments are needed to understand how Sim and Trh interact with additional factors to modify chromatin structure, and ongoing mutagenesis experiments will help reveal roles for the identified T, TG and G rich regions within midline and tracheal enhancers (Fig. 10). These repetitive motifs are found scattered throughout the enhancers and do not appear to have a fixed location relative to the CMEs. AT rich regions bend and denature relatively easily, facilitating DNA looping and are often found in cis-regulatory regions. The short, repetitive regions identified here may interact with specific transcription factors, such as Sox, Forkhead-type or other remodeling proteins to open chromatin [103,104]. Alternatively, these regions may be involved in 1) recruiting transcription factors after replication, 2) nucleosome positioning and/or 3) binding of histone modification enzymes to enhance transcription; all of which may affect quantitative and qualitative genetic variation in expression [105]. In addition, results with multiple transgenic lines indicate the synthetic constructs show little variation in patterns and levels and consistently recruit Sim/Tgo and/or Trh/Tgo regardless of insertion site. This suggests that factors interacting with these relatively small multimerized sequences (20–57 bp) are sufficient to open chromatin to allow for efficient transcription. Taken together, results from a number of labs suggest the following enhancer characteristics combine to determine if a gene will be expressed in the midline or trachea: 1) the number of CMEs within the enhancer, 2) the proximal context surrounding each CME and 3) binding sites for additional activators, repressors and/or factors that affect chromatin structure.

Evolution of Sim and Trh Developmental Functions

While these experiments focus on the cis-regulatory sequences that control the expression of genes within the midline and trachea, they do not address why many genes are expressed in both of these tissues and regulated by related PAS heterodimers. It is predominantly genes expressed in the CNS midline glia, rather than the midline neurons, that are also expressed in tracheal cells. PAS proteins perform diverse functions across all biological kingdoms and most characterized members function as environmental sensors [53–55]. Historically, Sim and Trh have been considered exceptions and their developmental functions have been emphasized [106]. However, functions of Sim and Trh may have arisen in ancestral organisms that more closely resemble the adult form of Drosophila, a stage when Sim and Trh may function more similarly. For instance, in adult flies, both glia and trachea provide support and energy to neurons and trh is expressed in the CNS late in embryogenesis and throughout the remainder of the fly’s life. In the adult fly brain, tracheal development is guided by glial cells, and ablating glia causes the trachea to branch more
extensively within this tissue [107]. Related mechanisms that guide glia and trachea distribution in the brain may explain, in part, shared gene regulatory pathways, including those regulated by the related PAS proteins, Sim and Trh. Most of the Drosophila PAS proteins that interact with Tgo are expressed in the trachea, including Trh, Dys and Similar (the fly version of HIF-1α), and Sim likely descended from a common ancestral gene. Developmental functions of Sim and Trh may have arisen later than their adult functions and common ancestral functions of these two tissues in the adult may explain why many enhancers drive expression in both the midline and trachea and why other midline and tracheal enhancers are closely linked. Further dissecting the similarities and differences in gene regulation within the CNS midline and trachea will reveal novel molecular mechanisms used to construct these tissues during development. Additional experiments are also needed to understand how signaling pathways combine with Sim and Trh to regulate genes in midline glia and trachea, not only in embryos, but also in larvae and adults, under different environmental conditions.

Supporting Information

Table S1 List of PCR primers used to generate fragments of the CG33275, esg, liprin γ, Netrin, comm, moody and Ect3 genes that were tested for their ability to drive midline and tracheal transcription. Restriction sites introduced for cloning purposes are indicated in lower case. (DOC)

Table S2 List of PCR primers used to generate the synthetic reporter genes tested for their ability to drive midline and tracheal transcription. The Toll and Sox synthetic reporters have been previously reported [59]. Engineered restriction sites used to ligate and subclone the synthetics are shown in lower case. (DOC)

Table S3 For each enhancer, 1 the name of the enhancer, 2 the tissue that expressed GFP driven by the enhancer, and 3,4 PCR primers used to generate the enhancers derived from esg and Netrin genes are listed. Restriction sites introduced for cloning purposes are indicated in lower case. (DOC)

Table S4 Listed is the immediate context of CMEs found within enhancers that drive expression in 1 the midline, 2 the trachea or 3 both tissues. For each CME (ACGTG), 4 the gene where it is found, 5 the number of CMEs within each enhancer and the 6 seven bp sequence of the site are shown. 7 The total number of CMEs found within the enhancers examined is indicated at the bottom of the table. (DOC)

Table S5 Listed is the immediate context of GCGTG motifs found within identified enhancers that drive expression in 1 the midline, 2 the trachea or 3 both tissues. For each GCGTG motif, 4 the gene where it is found, 5 the number found within each enhancer and the 6 seven bp sequence of the site are shown. 7 The total number of CMEs found within the enhancers examined is indicated at the bottom of the table. (DOC)

Acknowledgments

We would like to thank Trudy MacKay and Faye Lawrence for support and help with fly maintenance, Jim Skeath for antibodies and Laura Mathies for critically reading the manuscript.

Author Contributions

Conceived and designed the experiments: SL EF RB CD MM LJ PE. Performed the experiments: SL EF RB CD MM CN RC NB LJ PE. Analyzed the data: SL EF RB CD MM CN NB LJ PE. Contributed reagents/materials/analysis tools: LJ PE. Wrote the paper: SL PE.

References

1. Nambu JR, Lewis JO, Crews ST (1993) The development and function of the Drosophila CNS midline. Comparative Biochemistry and Physiology a-Physiology 104: 399–409.

2. Nambu JR, Lewis JO, Wharton KA, Crews ST (1991) The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. Cell 67: 1157–1167.
23. Tear G, Harris R, Sutaria S, Kilomanski K, Goodman CS, et al. (1996) The Drosophila tango gene encodes a MLL-H-PAS protein that is orthogonal to mammalian Amt and controls CNS midline and tracheal development. Development 124: 4571–4582.

24. Ohashi T, Saijo K (1997) Transcriptional regulation of breakdown LGF receptor gene by binding of TRACHEALESS/GARNT heterodimers to three central midline elements in Drosophila developing trachea. Development 124: 8739–8746.

25. Crews S, Thomas J, Goodman C (1988) The Drosophila single-minded gene encodes a nuclear protein with sequence similarity to the per gene product. Cell 15: 145–151.

26. Jacobs JR (2000) The Midline Glia of Drosophila: a molecular genetic model for the developmental functions of Glia. Progress in Neurobiology 62: 475–508.

27. Kidd T, Biese K, Mitchell JK, Fetter RD, Tessier-Lavigne M, et al. (1998) Roundabout controls axon crossing of the CNS midline and defines a novel gene family of evolutionarily conserved guidance receptors. Cell 92: 205–215.

28. Jiang L, Crews ST (2007) Transcriptional specificity of drosophila dysfusion for the developmental functions of Glia. Trends in Genetics 13: 191–196.

29. Seeger M, Tear G, Ferresmarco D, Goodman CS (1993) Dosage-sensitive and complementary functions of roundabout and commissuresless control axon crossing at the midline. Neuron 20: 23–33.

30. Seeger M, Tear G, Ferresmarco D, Goodman CS (1993) Mutations affecting growth cone guidance in Drosophila - genes necessary for guidance toward or away from the midline. Neuron 10: 409–426.

31. Tear G, Harris R, Sotaria S, Kilomanski K, Goodman CS, et al. (1996) commissuresless controls growth cone guidance across the CNS midline in Drosophila and encodes a novel membrane protein. Neuron 16: 501–514.

32. Wheeler SR, Kearney JB, Guardiola AR, Crews ST (2005) Single-cell mapping of neural and glial gene expression in the developing Drosophila CNS midline cells. Developmental Biology 294: 509–524.

33. Affolter M, Causinu E (2008) Tracheal branching morphogenesis in Drosophila: new insights into cell behaviour and organ architecture. Developmental Dynamics 135: 2053–2064.

34. Casanova J (2007) The emergence of shape: notions from the study of the Drosophila tracheal system. Embryo Reports 8: 335–339.

35. Brown S, Hu N, Hombria JC (2001) Identification of the first intervertebral interleukin-17a receptor, the Drosophila gene domeless. Current Biology 11: 1700–1705.

36. Chen HW, Chen X, Oh SW, Marinissen MJ, Gutkind JS, et al. (2002) mom identifies a receptor for the Drosophila JAK/STAT signal transduction pathway and encodes a protein distantly related to the mammalian cytokine receptor family. Genes & Development 16: 388–398.

37. Li J, Li W, Callhoun HC, Xia F, Gao FB, et al. (2003) Patterns and functions of STAT activation during Drosophila embryogenesis. Mechanisms of Development 120: 1455–1466.

38. Sorillon S, Espinoza-Vazquez JM, Foglia F, Hu N, Hombria JC (2010) An efficient approach to isolate STAT regulated enhancers uncovers STAT92E fundamental role in Drosophila tracheal development. Developmental Biology 340: 571–582.

39. Samakovlis C, Manning G, Steeneberg P, Hacohen N, Cantera R, et al. (1996) Genetic control of epithelial tube fusion during Drosophila tracheal development. Development 122: 3331–3336.

40. de Celis JF, Linsalata M, Casanova J (1995) Ventral veinless, the gene encoding the Cfa transcription factor, links positional information and cell differentiation during embryonic and imaginal development in Drosophila melanogaster. Development 121: 3456–3461.

41. Jiang L, Rogers SL, Crews ST (2007) The Drosophila Dead end Arf-like/3Pase controls vesicle-mediated tracheal fusion cell morphogenesis. Developmental Biology 311: 487–499.

42. Jiang L, Crews ST (2006) dysfusion transcriptional control of Drosophila tracheal migration, adhesion, and fusion. Molecular and Cellular Biology 26: 6547–6556.

43. Jiang L, Crews ST (2003) The Drosophila dysfusion basic helix-loop-helix (bHLH)-PAS gene controls tracheal fusion levels and of the tracheal cells bHLH-PAS protein. Molecular and Cellular Biology 23: 5625–5637.

44. Jiang L, Crews ST (2007) Transcriptional specificity of dysfusion dysfusion and the control of tracheal fusion cell gene expression. Journal of Biological Chemistry 282: 28659–28668.

45. Metzger KJ, Krausna MA (1999) Development - Genetic control of branching morphogenesis. Science 284: 1635–1639.

46. Klambt G, Glazer L, Shilo BZ (1992) Breathless, a Drosophila FGF receptor homolog, is essential for migration of tracheal and specific midline glial-cells. Genes & Development 6: 1668–1678.

47. Glazer L, Shilo BZ (1991) The Drosophila FGFR homolog is expressed in the embryonic tracheal system and appears to be required for directed tracheal cell extension. Genes & Development 5: 697–705.

48. Gherardi A, Krausna MA (2006) Social interactions among epithelial cells during tracheal branching morphogenesis. Nature 441: 746–749.

49. Hombria JC, Hacohen N, Cantera R, Manning G, Steeneberg P, Hacohen N (2002) Cave, a conserved Rae/Cdc62 GAP mediating Robo repulsion in tracheal cells and axons. Genes & Development 16: 2161–2171.

50. Englund C, Steeneberg P, Fahlke L, Nyborgnis G, Samakovlis C (2002) Attractive and repulsive functions of Slit are mediated by different receptors in the Drosophila trachea. Development 129: 4941–4951.

51. Watson JD, Wheeler SR, Stagg SB, Crews ST (2011) Drosophila hedgeshog signaling and en格uated-run mutual repression direct midline glia to alternative ensheathing and non-ensheathing fates. Development 138: 1283–1295.

52. Bossing T, Technau GM (1994) The fate of the CNS midline progenitors in Drosophila as revealed by a new method for single-cell labeling. Development 120: 1893–1906.

53. Ma Y, Cerdà K, Gao VP, Niemietz E, Musher J, et al. (2000) Functional interactions between Drosophila bHLH/PAS, Sex, and POU transcription factors regulate CNS midline expression of the slit gene. Journal of Neuroscience 20: 4596–4605.

54. Zeiler E, Shilo BZ (2009) Interaction between the bHLH-PAS protein Tracheless and the POU-domain protein Drifter, specifies tracheal cell fates. Mechanisms of Development 91: 163–173.

55. Bergmann A, Tugemann S, Shilo BZ, Steller H (2002) Regulation of cell number by MAPK-dependent control of apoptosis. A mechanism for trachic survival signaling. Developmental Cell 2: 159–170.

56. Schweitzer R, Shilo BZ (1997) A thousand and one roles for the Drosophila EGF receptor. Trends in Genetics 13: 191–196.

57. Taylor BL, Zhulin IB (1999) PAS domains: Internal sensors of oxygen, redox potential, and light. Microbiology and Molecular Biology Reviews 63: 479–506.

58. Henry JT, Crosston S (2011) Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. Annual Review of Microbiology, Vol 65 65: 255–286.

59. McIntosh BE, Hogemese NB, Bradfield CA (2010) Mammalian Per-Arnt-Sim Proteins in Environmental Adaptation. Annual Review of Physiology 72: 625–643.

60. Clark AG, Eiser MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–218.

61. Barolo S, Carver LA, Posakony JW (2000) GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. Biotechniques 29: 732–736.

62. Jiang L, Pearson JC, Crews ST (2010) Diverse modes of Drosophila tracheal cell transmigration regulation. Mechanisms of Development 127: 263–276.
60. Patel NH (1994) Imaging neuronal subsets and other cell types in whole-mount Drosophila embryos and larvae using antibody probes. Methods in Cell Biology 44: 445–487.

61. Bailey T, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Nucleic Acids Research 22: 413–418.

62. Wharton KA, Franko RG, Kasai Y, Crews ST (1994) Control of CNS midline transcription by asymmetric E-box-like elements - similarity to xenobiotic responsive regulation. Development 120: 3563–3569.

63. Wharton KA, Franks RG, de Couet HG, Fischbach KF (2005) Single-minded, DmeZ, Pointed, and Su(H) act on identified regulatory sequences of the roughist gene in Drosophila melanogaster. Development Genes and Evolution 215: 469–469.

64. Apitz H, Kambacheld M, Hohe M, Ramos RGP, Straube A, et al. (2004) Identification of regulatory modules mediating specific expression of the roughest gene in Drosophila melanogaster. Development Genes and Evolution 214: 434–459.

65. Estes P, Fullerton E, Zhang Y (2006) Identification of motifs that are conserved in 12 Drosophila species and regulate midline glia vs. neuron expression. Genetics 172: 787–799.

66. Pearon JC, Watson JD, Crews ST (2012) Drosophila melanogaster Zelda and Single-minded-correlate to regulate an evolutionarily dynamic CNS midline cell enhancer. Developmental Biology 366: 420–432.

67. McQuillan P, St Pierre SE, Thurmond J, Fly Base C (2012) FlyBase 101-the basis of navigating Flybase. Nucleic Acids Research 40: D706–D714.

68. Kumar S, Komikoff C, Van Enden B, Buicic C, Davis KT, et al. (2011) FlyExpress: visual mining of spatiotemporal patterns for genes and publications in Drosophila embryogenesis. Bioinformatics 27: 3319–3320.

69. Sambouris C, Hohlen N, Manning G, Sutherland DC, Guillem M, et al. (1996) Development of the Drosophila tracheal system occurs by a series of morphologically distinct but genetically coupled branching events. Development 122: 1395–1407.

70. Roy S, Erm L, Krabbecken PK, Vherdadpur P, Negre N, et al. (2010) Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science 330: 1767–1779.

71. Cenoffur T, Oda H, Kraut R, Hayashi S, Kotaoka Y, et al. (1996) Zygotic expression of the Drosophila kradavin expression is required for processes of dynamic epithelial cell rearrangement in the Drosophila embryo. Genes & Development 10: 659–671.

72. Sturgarraga S, Hofmeyer K, Farajian R, Treisman JE (2013) Three Drosophila Lipins Interact to Control Synapse Formation. Journal of Neuroscience 33: 15358–15368.

73. Moore SW, Tessier-Lavigne M, Kennedy TE (2007) Neuritins and their receptors. Axon Growth and Guidance 6: 17–31.

74. Dickson BJ (2002) Molecular mechanisms of axon guidance. Science 298: 1959–1964.

75. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster enhancers of the Drosophila slit and Toll genes. Mechanisms of Development 140: 151–154.

76. Cervenka R, Shugrue DJ, Ayer A, Nislow HC (2007) Breast cancer stem cells. Journal of Biological Chemistry 282: 1910–1917.

77. McLaurin JA, Ramsey MJ, Li H, Zhang X, et al. (2004) The rules of developmental gene regulation. Cell Reports 2 1002–1013.

78. Nestorov D, Shapovalov Y, Kuznetsova E, et al. (2010) The roughist gene in Drosophila melanogaster. Development Genes and Evolution 220: 499–510.

79. Wang Y, Wu Y, Liu L, Su X (2013) Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL basic helix-loop-helix domains with E-box DNA. Cell Research 23: 213–224.

80. Swanson HI, Chan WK, Bradfield CA (1995) DNA binding specificities and pairing rules of the Ah receptor, ARNT and SIM proteins. Journal of Biological Chemistry 270: 26292–26302.

81. Hong J, Park KW, Levine M (2013) Temporal regulation of single-minded target genes in the ventral midline of the Drosophila central nervous system. Developmental Biology 380: 355–365.

82. Lammens AS, Karmakar S, Jin Y, Nagaih AK (2012) Multiple modes of transcriptional repression in response to Notch receptor activity. PLoS One 6: e26197.