

## RESEARCH ARTICLE

Chromatin profiling of *Drosophila* CNS subpopulations identifies active transcriptional enhancersJoseph C. Pearson<sup>1,2,\*</sup>, Daniel J. McKay<sup>2,3,4,\*‡</sup>, Jason D. Lieb<sup>3</sup> and Stephen T. Crews<sup>1,2,‡</sup>

## ABSTRACT

One of the key issues in studying transcriptional regulation during development is how to employ genome-wide assays that reveals sites of open chromatin and transcription factor binding to efficiently identify biologically relevant genes and enhancers. Analysis of *Drosophila* CNS midline cell development provides a useful system for studying transcriptional regulation at the genomic level due to a large, well-characterized set of midline-expressed genes and *in vivo* validated enhancers. In this study, FAIRE-seq on FACS-purified midline cells was performed and the midline FAIRE data were compared with whole-embryo FAIRE data. We find that regions of the genome with a strong midline FAIRE peak and weak whole-embryo FAIRE peak overlap with known midline enhancers and provide a useful predictive tool for enhancer identification. In a complementary analysis, we compared a large dataset of fragments that drive midline expression *in vivo* with the FAIRE data. Midline enhancer fragments with a midline FAIRE peak tend to be near midline-expressed genes, whereas midline enhancers without a midline FAIRE peak were often distant from midline-expressed genes and unlikely to drive midline transcription *in vivo*.

**KEY WORDS:** Chromatin, CNS, *Drosophila*, Enhancer, FAIRE-seq, Midline

## INTRODUCTION

Tissue-specific regulation of gene expression is an important aspect of development, but the mechanisms that determine which segments of the regulatory genome are active in a given cell remain incompletely known. While genome-wide analyses identifying sites of open chromatin, transcription factor binding and chromatin states are commonly carried out, it is often unclear how well these assays are able to identify relevant regulatory elements and target genes. *Drosophila* CNS midline cells provide a useful experimental system for studying transcriptional regulation during CNS development. The development of the midline neurons and glia is well characterized with regards to both cellular and molecular mechanisms (Watson et al., 2011; Watson and Crews, 2012; Wheeler et al., 2008). Multiple large-scale screens have identified hundreds of genes that are expressed in midline cells (along with

associated enhancers), and several transcription factors are known that control midline cell transcription (Kearney et al., 2004; Kvon et al., 2014; Manning et al., 2012; Tomancak et al., 2002, 2007; Wheeler et al., 2006, 2009). In this study, we purify midline cells from a key developmental period in which cell fate specification occurs and perform formaldehyde-assisted isolation of regulatory elements (FAIRE) analysis (Giressi et al., 2007) to initiate a genome-wide characterization of midline cell chromatin states. These data allow us to use the unique breadth of existing *in vivo* midline cell data to examine the relationships between chromatin accessibility, transcriptional enhancer activity, and cell type-specific gene expression in developing animals.

The *Drosophila* embryonic CNS midline cells consist of ~22 neurons and glia per segment (Wheeler et al., 2006). The neurons are diverse and consist of GABAergic and glutamatergic interneurons and peptidergic and neuromodulatory motoneurons. The midline glia ensheath the commissural axons that cross the CNS and act as a key embryonic signaling center (Crews, 2009; Jacobs, 2000). There are three discrete phases of CNS midline cell development (Kearney et al., 2004). The initiation of midline cell development occurs during the mesectodermal stage (embryonic stages 5-8), mainly due to the activation of *single-minded* (*sim*) expression (Nambu et al., 1991). Formation of the midline primordium cells (stages 9-12) involves specification of neural precursors and midline glia by multiple signaling pathways (Watson et al., 2011; Watson and Crews, 2012; Wheeler et al., 2008), followed by division and development into neurons and glia. At the late embryonic midline stages (stages 13-17), the midline neurons complete terminal differentiation and the midline glia complete their apoptotic and migration steps to form the glial scaffold that supports the axon commissures.

One long-standing aspect of the study of *Drosophila* CNS midline cell development has been transcriptional control. This has been due, in part, to the multiple genome-level resources that exist to support the study of midline cell development and transcription. The Berkeley *Drosophila* Genome Project (BDGP) gene expression database has embryonic *in situ* hybridization (ISH) data on 7559 genes, including annotation for midline expression (Tomancak et al., 2002, 2007). The Midline Gene Expression Database (MidExDB) expanded on the BDGP effort and is focused on ISH data exclusively related to midline expression (Wheeler et al., 2009). Another resource is the transcriptomic profile of midline cells from RNA-seq analyses of FACS-purified midline cells at two developmental periods, namely 6-8 h after egg laying (AEL) and 14-16 h AEL (Fontana and Crews, 2012). In addition, two large-scale projects were carried out to identify the transcriptional enhancers that regulate gene expression. Through cloning of thousands of genomic fragments for transgenic reporter analysis, the Janelia Research Campus (JRC) FlyLight project identified 438 fragments with enhancer activity associated with 253 genes that drive strong embryonic midline expression (Manning et al., 2012).

<sup>1</sup>Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA. <sup>2</sup>Integrative Program for Biological and Genome Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA. <sup>3</sup>Department of Biology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA. <sup>4</sup>Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA.

\*These authors contributed equally to this work

‡Authors for correspondence (dmckay1@email.unc.edu; steve\_crews@unc.edu)

 S.T.C., 0000-0002-1432-401X

Similarly, the Vienna Fly Enhancer project identified 187 midline lines (Kvon et al., 2014). Although the FlyLight and Vienna collections have provided a deeper understanding of the transcriptional regulatory capacity of animal genomes, the extent to which the activity of DNA fragments in reporter assays reflects the activity of these DNAs in their natural genomic contexts is unclear.

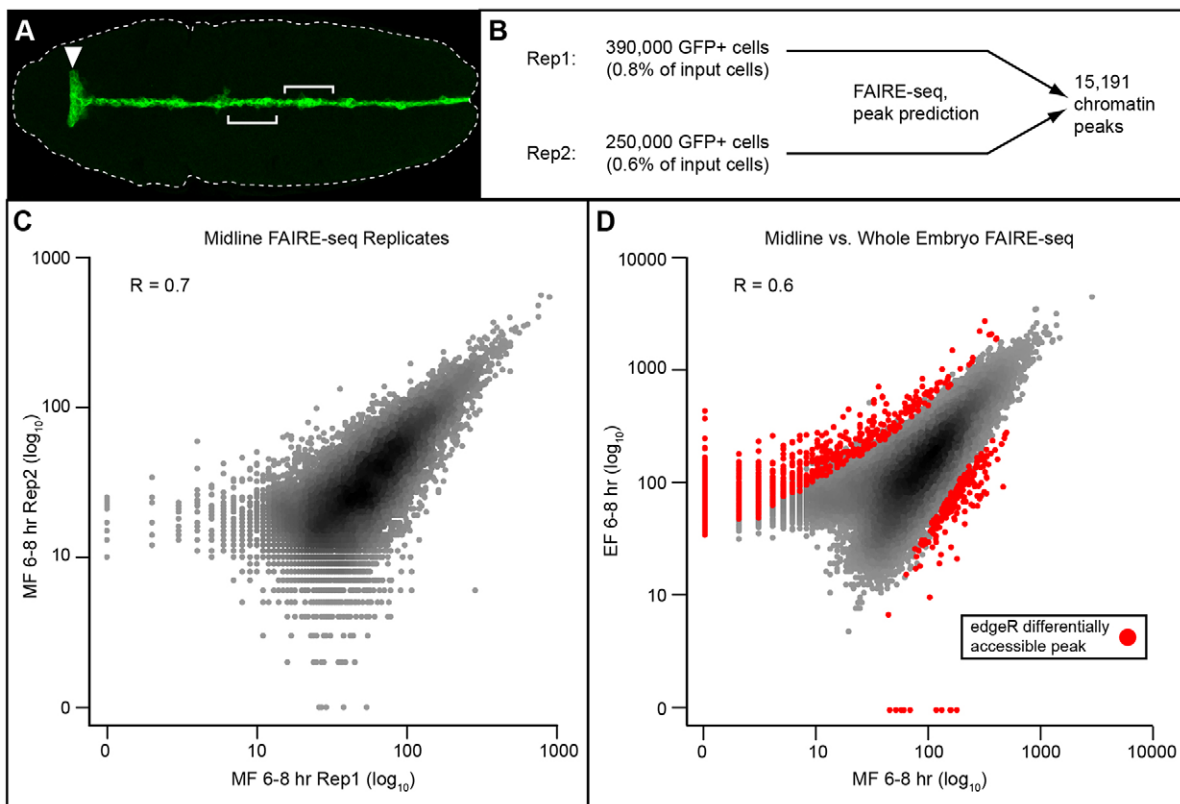
These genomic datasets, along with a well-characterized genetic understanding of midline cell development, provide a powerful system for mechanistically studying the transcriptional control of CNS development. To generate a genome-wide map of functional DNA regulatory elements in developing CNS midline cells, we performed FAIRE-seq (Giresi et al., 2007), which isolates nucleosome-depleted genomic regions, including transcriptional enhancers and promoters. The relatively low chromatin input required for FAIRE-seq allowed us to assay chromatin accessibility in FACS-isolated midline cells, which comprise less than 1% of *Drosophila* embryonic cells (Fontana and Crews, 2012). Chromatin accessibility profiles from FACS-isolated midline cells show a high degree of correspondence with previously identified midline enhancers. In this study, we show the utility and necessity of combining FAIRE-seq data from purified midline cells with data from whole-embryo cells to identify biologically relevant enhancers with a high degree of certainty. We further compare the FAIRE-seq data with the multiple midline expression and enhancer datasets to explore the predictive utility of the FAIRE data in identifying and understanding the biological relevance of midline enhancers.

## RESULTS

### Midline cell FAIRE-seq reveals cell type-specific open chromatin

To generate open chromatin profiles from midline cells, we used FACS to isolate GFP<sup>+</sup> cells from dissociated *sim3.7-Gal4 UAS-mCD8::GFP* embryos (Fontana and Crews, 2012). The *sim3.7-Gal4* transgene drives expression in only CNS midline cells and a small group of gut cells (Fig. 1A). We performed two replicates. Populations of midline cells (390,000 and 250,000 cells) were isolated from two independent collections of embryos at 6-8 h AEL, corresponding to mid-stage 11 to early stage 12 (Campos-Ortega and Hartenstein, 1997) (Fig. 1B). These cells are referred to as midline primordium cells and are just beginning to differentiate into midline neurons and glia. After sorting, cells were briefly fixed with formaldehyde, frozen, subjected to FAIRE, and the resulting DNA sequenced on an Illumina HiSeq. Despite the technical challenges presented by isolating GFP<sup>+</sup> cells from whole embryos, FAIRE replicates were well correlated (Fig. 1C), with a Pearson's coefficient of 0.7. MACS2 analysis (Zhang et al., 2008) of the pooled data identified 15,191 peaks of open chromatin (Fig. 1B).

Midline cells represent a small fraction of the embryo; for example, only 0.7% of the sorted cells were GFP<sup>+</sup>. Thus, it is possible that chromatin accessible only in midline cells would not be detected in assays performed on whole embryos. To identify regions of chromatin accessibility that are specific to midline cells, we compared midline FAIRE (MF) data with whole-embryo FAIRE (EF) data from the same 6-8 h developmental stage (McKay and



**Fig. 1. FAIRE-seq analysis of FACS-isolated *Drosophila* CNS midline cells.** (A) Stage 11 *sim3.7-Gal4 UAS-mCD8::GFP* embryo (ventral view, anterior left) stained with anti-GFP (green) shows midline cells (each bracket denotes a single embryonic segment; ~20 midline cells) and a small group of stomodeal cells (arrowhead). (B) Summary of FACS sorting data from two independent midline FACS samples (Rep1 and Rep2). (C) Scatterplot of midline FAIRE (MF) signals of the top 10,000 FAIRE peaks from replicates of FACS-sorted midline cells. Pearson's coefficient (R) is shown. (D) Scatterplot of FAIRE signals of the top 10,000 FAIRE peaks from FACS-sorted midline cells (MF) and stage-matched whole-embryo cells (EF). Red dots indicate differentially accessible peaks, as determined by edgeR (FDR  $\leq 0.05$ ).

Lieb, 2013). Genome-wide, the chromatin accessibility profiles in midline cells are similar to those from whole embryos at the same developmental stage (Fig. 1D), suggesting that many DNA regulatory elements are equivalently accessible in both midline and non-midline cells. However, we also noticed regions of high chromatin accessibility that were specific to midline cells or to whole embryos. Using edgeR to define FAIRE peaks with differential accessibility (Robinson et al., 2010), we identified 167 peaks with greater accessibility in midline cells (FDR  $\leq 0.05$ ; hereafter referred to as ‘midline-enriched’ FAIRE peaks; Table S1) and 942 peaks with greater accessibility in whole embryos (red dots in Fig. 1D). Thus, while chromatin accessibility is similar between whole embryos and midline cells, there are regions enriched only in midline cells and other regions that are relatively inaccessible in midline cells relative to whole embryos.

### Midline FAIRE peaks correspond to known midline enhancers

Chromatin accessibility correlates with DNA regulatory sequence activity of promoters, enhancers and other elements (e.g. McKay and Lieb, 2013). In the sections below, we employ five datasets of midline-expressing enhancers and genes (see Introduction) to address the significance of the MF data. Use of the midline enhancer, midline ISH gene expression, and RNA-seq datasets allows us to interpret the FAIRE data within the context of *in vivo* relevant midline-expressed genes and regulatory elements.

The first dataset that we examined included 19 midline enhancers associated with midline primordium-expressed genes (referred to as ‘canonical’ midline primordium enhancers) that were described in previous publications (Table S2) and are active during our 6-8 h FAIRE assay window (‘midline primordium’; 5-9 h AEL). All 19 midline primordium enhancers have an MF peak, whereas only ten have a corresponding EF peak (Fig. 2A). Consistent with this observation, ten of the midline primordium enhancers contain a midline-enriched peak (edgeR FDR  $\leq 0.05$ ), indicating that midline-specific enhancers often possess midline-enriched chromatin accessibility (Fig. 2A, ML-EN; Table S2). Five enhancers have been localized to a resolution of  $\sim 500$  bp, about the size of the FAIRE peaks (Table S2), indicating a strong correspondence between the MF peak and enhancer. Analysis of the 19 midline primordium enhancers indicates that these enhancers all have an MF signal that is larger than the EF signal (Fig. 2B), even though nine have an FDR  $> 0.05$ . We also analyzed a published set of 78 enhancers that do not drive midline expression when tested *in vivo* but are active in other cell types at the same stage 11-12 period. These fragments have significantly lower MF signals and MF/EF ratios (Fig. 2B) than the midline primordium fragments. Thus, enhancers active in midline primordium cells overlap regions of open chromatin in midline cells, whereas enhancers active in other tissues at the same stage do not.

### Midline FAIRE peaks overlap with enhancers in the *sim* master regulatory gene locus

The correspondence between MF peaks and enhancers that drive midline expression can be readily visualized by examination of well-characterized midline-expressed genes, such as *Drosophila sim*, a master regulator of CNS midline cell transcription and development (Nambu et al., 1990, 1991). *sim* is prominently expressed in midline cells throughout embryonic development, including the midline primordium stage (Thomas et al., 1988). It is also expressed in subsets of cells along the midline of the foregut and midgut (Nambu et al., 1990) and transiently in a subset of developing muscle cells (Lewis and Crews, 1994).

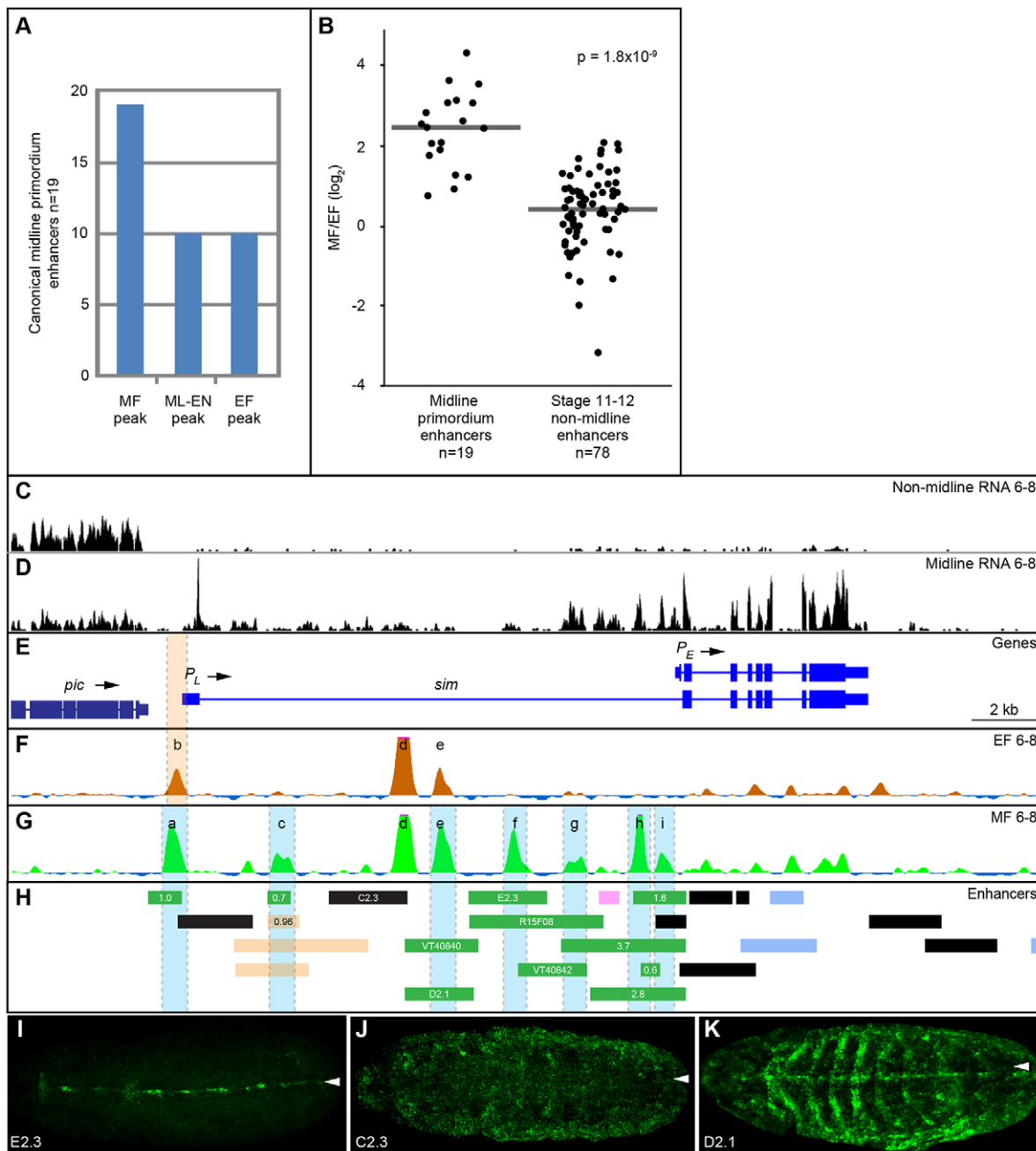
Embryonic transcription of *sim* proceeds from two promoters: the early promoter ( $P_E$ ) and the late promoter ( $P_L$ ) (Fig. 2C-E) (Kasai et al., 1998; Muralidhar et al., 1993; Nambu et al., 1990). Previous work has shown that, at 6-8 h, transcription is predominantly derived from  $P_L$ . This is confirmed by the 6-8 h MF and EF data (Fig. 2F,G), which show a peak of open chromatin at  $P_L$  (peak b) but not  $P_E$ . There are at least four distinct midline primordium enhancers at the *sim* locus, as indicated by *in vivo*-tested fragments (1.0, 0.7, R15F08, 1.6; Fig. 2H) (Freer et al., 2011; Manning et al., 2012; Muralidhar et al., 1993; Sandmann et al., 2007; Wharton et al., 1994). The MF data reveal a total of eight strong MF peaks (a, c-i). Five of the MF peaks (c, f, g, h, i) do not overlap a significant corresponding EF peak, and one peak (a) incompletely overlaps an EF peak (b) that is likely to reflect open chromatin at the  $P_L$  promoter. Peaks c, f, h and i are midline enriched (FDR  $\leq 0.05$ ). All four distinct midline primordium enhancers overlap an MF peak (1.0 with a, 0.7 with c, R15F08 with f/g, 1.6 with h/i). Interestingly, the 0.7 fragment drives midline expression, but larger fragments that encompass 0.7 do not; this includes 0.96, which is not much larger than 0.7. The presence of peak c overlapping the 0.7 enhancer provides support that this region is used *in vivo* to drive midline expression. Similarly, we propose that peak f is a midline enhancer and the previously uncharacterized E2.3 fragment that includes peak f drives midline expression (Fig. 2I). In addition, peaks h and i might represent distinct enhancers. Interestingly, the two MF peaks (d, e) with corresponding strong EF peaks do not drive strong midline primordium expression. Peak d does not drive embryonic 6-8 h expression (C2.3; Fig. 2J), and peak e overlaps two fragments (VT40840 and D2.1) that drive relatively weak midline expression, as well as non-midline expression (Fig. 2K) that does not correspond to endogenous *sim* expression.

In summary, all *sim* midline primordium enhancers identified by *in vivo* enhancer testing have corresponding 6-8 h MF peaks, including four midline-enriched MF peaks. Similarly, all 6-8 h MF peaks without significant corresponding EF 6-8 h peaks overlap a fragment with midline enhancer activity. Remarkably, the *sim* gene has potentially seven distinct enhancers that drive midline primordium expression in addition to another enhancer that drives mesectodermal expression.

### FAIRE identifies multiple new midline enhancers in the *engrailed/invected* locus

The preceding analysis of the *sim* locus demonstrated the utility of open chromatin profiling from purified cells to identify tissue-specific enhancers that might not be detected in whole embryo experiments. We next sought to test whether these data can be used to identify previously unknown midline enhancers. The *engrailed* (*en*) and *invected* (*inv*) genes are both expressed in midline and non-midline cells, as shown by ISH and RNA-seq from isolated cell populations (Fig. 3A-C,H). However, identification of midline enhancers has been hampered by the size of the locus (110 kb) and its complex organization (Gustavson et al., 1996). We used previously uncharacterized fragments from the FlyLight collection (Manning et al., 2012) to test whether midline FAIRE can predict midline enhancer activity.

There are eight substantial MF peaks (a-h) in the *en/inv* locus that we considered (Fig. 3D,E). Two of these regions (a, d) correspond to the *inv* and *en* promoters. Three of these regions (c, e, h) are accessible in both midline cells and whole embryos. Consistent with this observation, regions c and h overlap fragments from the FlyLight and Fly Enhancer (Kvon et al., 2014) collections that drive broad expression in both midline and non-midline cells (region e was

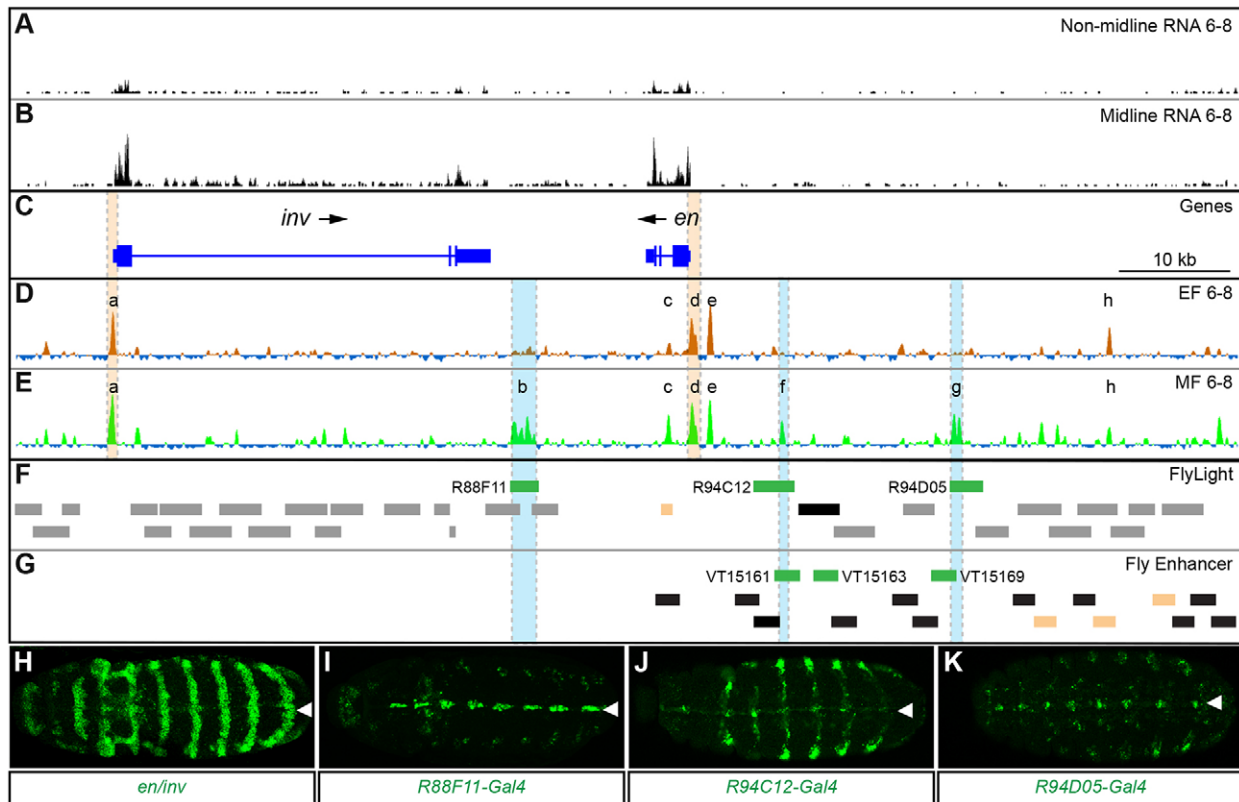


**Fig. 2. Midline enhancers contain midline-enriched FAIRE peaks.** (A) Canonical midline primordium enhancers from 19 midline-expressed genes were analyzed for the presence of: (1) an MF peak; (2) a midline-enriched MF peak ( $FDR \leq 0.05$ ) (ML-EN peak); and (3) an EF peak. (B) Plots of the FAIRE  $\log_2$  signal ratio (MF/EF) for canonical enhancers active during the 6–8 h FAIRE assay window. Each dot represents one enhancer. Midline primordium enhancers have significantly higher MF/EF values than non-midline enhancers (Mann–Whitney U-test,  $P=1.8 \times 10^{-9}$ ). The two horizontal lines in each data column represent the median value for each set of enhancers. (C–H) UCSC Genome Browser tracks at the *sim* locus. (C,D) Non-midline and midline 6–8 h embryonic RNA-seq data indicating strong midline expression of *sim*. (E) Schematic of *sim* gene structure with early ( $P_E$ ) and late ( $P_L$ ) promoters of *sim* indicated, with arrows showing transcriptional orientation. Boxes indicate exons and lines signify introns. Tall boxes indicate coding sequence and shorter boxes denote untranslated regions. (F,G) FAIRE-seq data from (F) 6–8 h whole embryo FAIRE (EF) and (G) 6–8 h midline FAIRE (MF). Letters (a–i) identify significant FAIRE peaks. (H) Boxes represent FlyLight, Fly Enhancer and published enhancers. Colors indicate activity: green, midline primordium expression; pink, mesectodermal expression; blue, midline expression after stage 12; orange, broad expression that may include midline expression; black, no midline expression. Vertical dashed columns indicate predicted midline enhancers from MF peaks (blue) and the  $P_L$  promoter corresponding to FAIRE peak b (orange). (I–K) Expression of selected *sim* gene enhancers at stage 11. Expression is revealed by *Gal4* fluorescent ISH of ventral views of strains (I) *simE2.3-Gal4*, (J) *simC2.3-Gal4* and (K) *simD2.1-Gal4*.

not cloned in either project, and its pattern of activity was not assayed). Most interesting are three regions that are more highly accessible in MF than in EF (b, f, g), although none of the three peaks is midline enriched ( $FDR \leq 0.05$ ). Each of these midline-specific

accessible regions overlaps a fragment that was cloned to generate transgenic *Gal4* lines by FlyLight (Fig. 3F). The expression patterns controlled by these fragments were previously untested, providing an opportunity to test the predictive value of midline-specific open





**Fig. 3. Midline FAIRE reveals multiple midline enhancers in the *inv/en* locus.** (A-G) UCSC Genome Browser tracks corresponding to the *inv/en* locus. (A,B) RNA-seq indicating strong expression of *inv* and *en* in midline cells at 6-8 h AEL. (C) Schematic of the *inv/en* locus. (D,E) Data from (D) 6-8 h EF and (E) 6-8 h MF. Letters (a-h) identify significant MF peaks. Blue vertical columns correspond to MF peaks that match *in vivo*-tested midline enhancers and orange columns indicate FAIRE peaks that correspond to *inv* and *en* promoters. (F,G) Boxes indicate (F) FlyLight and (G) Fly Enhancer fragments. Colors indicate activity: green, midline primordium expression; orange, broad expression that may include midline expression; black, no midline expression; gray, untested. (H-K) Ventral views of stage 11 embryos hybridized *in situ* to (H) an *en* probe and (I-K) *Gal4* probes; arrowheads indicate midline cells. (H) Fluorescent ISH with *en* probe showing ectodermal stripes and midline expression. (I) Strong midline expression in posterior midline cells of an *R88F11-Gal4* strain. (J) *R94C12-Gal4* drives expression in a subset of posterior midline cells and weakly in ectodermal stripes. (K) *R94D05-Gal4* drives expression in a small group of posterior midline cells.

chromatin for transcriptional regulatory activity. We assayed the expression of each line by ISH for *Gal4* and anti-GFP immunostaining of embryos also possessing *UAS-mCD8::GFP*. All three fragments drive strong midline expression during stages 11-12 (Fig. 3I-K). More detailed analysis of the FlyLight *Gal4* lines revealed that all three fragments drive expression in distinct posterior midline subsets of the endogenous *en/inv* pattern (Fig. S1), which is likely to reflect the complexities of *en/inv* midline function (Watson et al., 2011; Watson and Crews, 2012; Wheeler et al., 2006). Subsequently, the publication of the Fly Enhancer database (Kvon et al., 2014) (Fig. 3G) and a transgenic dissection of the *en/inv* locus (Cheng et al., 2014) also revealed that peaks f and g overlap fragments that drive midline expression; the overlaps between R94C12 and VT15161, and between R94D05 and VT15169, occur over the MF peaks (Fig. 3E-G).

In summary, FAIRE analysis from sorted midline cells reveals three prominent accessible regions at the *en/inv* locus that are specific to midline cells, and each of these regions overlaps a fragment that controls expression in a subset of the endogenous *en/inv* pattern. Thus, by leveraging high-throughput transcriptional reporter databases, such as FlyLight and the Fly Enhancer collection, sorted-cell FAIRE data can be used to successfully pinpoint the location of *in vivo* relevant enhancers that are active even in a small percentage of cells from a complex genetic locus.

### Midline-enriched FAIRE peaks efficiently identify midline-expressed genes

The results described above indicate that 19 canonical midline primordium enhancers each contain a high-ranking MF peak and a low-ranking or absent EF peak. Ten of the enhancers had a midline-enriched MF peak (FDR  $\leq 0.05$ ). We also showed in the case of *en/inv* how the appearance of high MF/EF ratio peaks could be used to predict midline enhancers. Consequently, we asked what fraction of the 167 midline-enriched peaks is associated with midline-expressed genes (and presumed midline enhancers). This analysis is greatly facilitated by the existence of the BDGP and MidExDB ISH data and our midline and whole-embryo RNA-seq data, since these resources reliably detect midline expression for most of the genome. Each peak was scanned in the UCSC browser for nearby genes and these genes assessed for embryonic expression using RNA-seq and ISH data; ISH expression data were present for 89% of the nearby genes analyzed. The peaks were sorted into four groups (see Fig. S2 for examples). The first group contained peaks near or within prominent midline-expressed genes (Table S1, 'Midline'; Fig. S2A,B). This group consists of 59 midline-enriched MF peaks and constitutes 35% of the total midline-enriched peaks (Table S3), demonstrating the strong utility of this identifier for recognizing midline-expressed genes. Of these 59 peaks, 31 overlap known enhancers from canonical, FlyLight or the Fly Enhancer collections; 24 of those enhancers drive strong midline expression, whereas the

other seven do not (Table S3). The 59 peaks correspond to 43 distinct gene loci, many of which are well-studied, prominent midline-expressed genes. However, we also were introduced to a number of genes whose midline expression was not known to us (*Ank2*, *dnr1*, *mirr*, *sick*), further indicating the utility of the MF data. We documented the midline expression of *mirr* by ISH (Fig. S3).

The second, smaller group of midline-enriched enhancers are found in genes with broader CNS expression as determined by ISH; midline expression is present but much less prominent compared with the Midline group genes mentioned above (Table S1, 'Broad'; Fig. S2C). Within this group, the midline-enriched MF peak may potentially contribute to the midline expression component. There were 18 peaks corresponding to 18 gene loci in the Broad group, accounting for 11% of the total peaks (Table S3). RNA-seq data for these genes (Table S1) consistently show significant midline expression. Known enhancers overlap five peaks; one drives midline expression and four do not. Combined with the Midline group, these two groups constitute 46% of the midline-enriched MF peaks.

The third group of midline-enriched MF peaks resides near genes expressed in the gut (Table S1, 'Gut'; Fig. S2D). This is not surprising, since the *3.7sim-Gal4 UAS-mCD8::GFP* strain used to isolate the midline cells also expresses *GFP* in a subset of gut cells (Fig. 1A) and gut expression of *sim* was previously described (Nambu et al., 1990). There are 40 peaks (24% of total) that correspond to 35 genes with prominent gut expression (Table S3). When the Gut, Midline and Broad group peaks are combined, 70% of the midline-enriched peaks correlate with relevant enhancers (midline and gut) active in the sorted cells. This leaves only 50 peaks (30% of the total) that are not obviously associated with a potentially relevant enhancer or gene (Table S1, 'None').

Overall, these data indicate the utility of comparing FAIRE-seq of purified cells with whole embryo FAIRE-seq to identify cell type-specific enhancers. In addition, the data indicate that the purer the starting cells, the more useful the results. Why don't all midline-enriched peaks correspond to midline or gut-expressed genes? Some midline-enriched peaks and corresponding open chromatin might simply not act as a midline or gut enhancer; other reasons include: (1) a lack of ISH expression data that, if present, would indicate the existence of a nearby midline or gut-expressed gene and (2) the ability of enhancers to act over long distances.

#### DNA binding motifs are variably enriched among different classes of midline-enriched FAIRE peaks

The four groups of midline-enriched FAIRE peaks were each analyzed for over-represented DNA-binding site motifs of prominent midline primordium transcription factors using a collection of tools from the MEME suite and transcription factor position weight matrices (PWMs) from Fly Factor Survey (<http://mccb.umassmed.edu/ffs>) (Fig. 4) (Bailey et al., 2015; Meng et al., 2005). In the Midline group of peaks, the most prominent over-represented motif is ACGTG, which corresponds to the binding site for Sim-Tgo (Bailey et al., 2015; Meng et al., 2005). This reinforces the association of these peaks with midline enhancers and the prominent role of Sim-Tgo in controlling midline gene regulation. The next highest motif corresponds to the binding sites for homeobox transcription factors, including midline-expressed *Ocelliless* (*Oc*). Three additional over-represented sites correspond to *Pointed* (*Pnt*), *Ventral veins lacking* (*Vvl*) and *En*, which play prominent roles in midline transcription. Motif analysis of the Broad, Gut and None groups did not reveal any strongly over-represented motifs other than a modest over-representation of a homeobox-related motif in the Gut and None

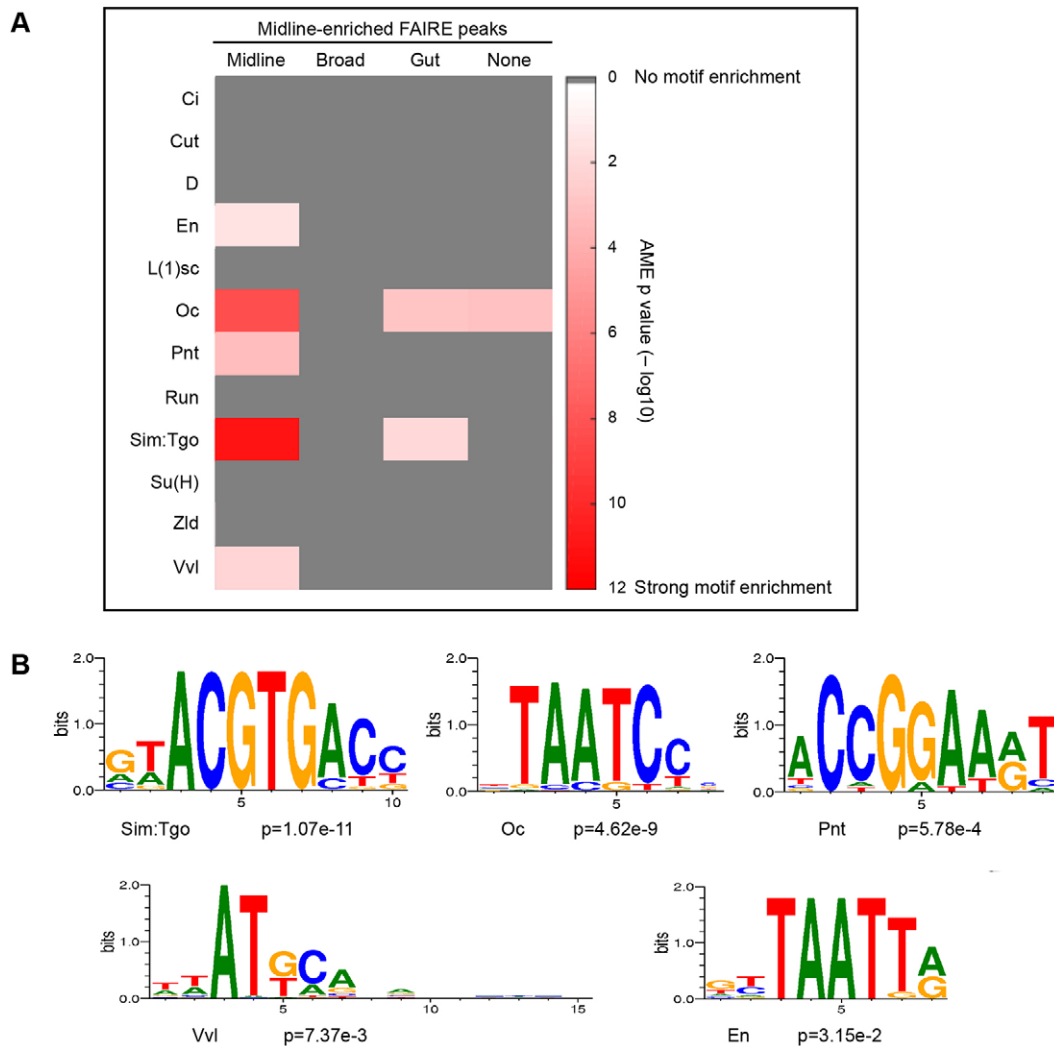
groups and for the Sim-Tgo motif in the Gut group. Directly counting the instances of ACGTG motifs revealed that 47/59 Midline group peaks contained an ACGTG motif (80%), with a mean of 1.56 ACGTGs/peak, significantly greater than the background mean of 0.66 ( $P < 0.0001$ ). The Midline group ACGTG frequency was also significantly greater ( $P = 0.01$ ) than the occurrence of ACGTG motifs in the Broad group, in which 10/18 peaks had an ACGTG (56%), with a mean of 0.83 ACGTGs/peak, not significantly different from the background mean of 0.63 ( $P = 0.33$ ). Similarly, the Gut group had 21/40 (52%) peaks with an ACGTG motif, with a mean of 0.85 ACGTGs/peak, similar to the background mean of 0.60 ( $P = 0.14$ ).

#### Genes with strong midline expression have abundant midline FAIRE peaks

Because of the ease of identifying midline primordium gene expression, most genes with strong midline expression are known. The MidExDB database currently lists 164 genes with midline primordium expression in all or a subset of midline cells. These genes vary significantly in the levels of midline expression. Reanalyzing the ISH data, 37 genes are noteworthy for having strong midline primordium expression (Table S4). This is reinforced by a median midline cell RNA-seq FPKM value of 142.52 and median ratio of midline RNA-seq/whole-embryo RNA-seq FPKMs of 8.9 for the 37 genes. Examining each gene for the occurrence of MF peaks indicates that each has an MF peak varying in number from 1 to 39. The median number of MF peaks/gene is 10.1. There are 22 genes (59%; 22/37) with at least one midline-enriched (FDR  $\geq 0.05$ ) peak. There are a total of 50 midline enhancers associated with 28 of the genes and 90% (45/50) overlap with an MF peak. These data further indicate that genes with strong midline expression have enhancers that can be recognized by the occurrence of midline-enriched or other high-ranking MF peaks.

#### Not all midline enhancers have a midline FAIRE peak, but those with midline-enriched peaks reside near midline-expressed genes

Having established that most midline-expressed genes have a corresponding MF peak, we addressed a different, but related question: what fraction of midline enhancers identified by *in vivo* testing have MF peaks and are associated with (and likely control the expression of) midline-expressed genes? We focused our analysis on the Fly Enhancer *Gal4* lines, which were analyzed by ISH and thus closely reflect the dynamics of enhancer activity. We omitted the FlyLight data because analysis was by anti-GFP immunostaining, and in some cases, the 'midline primordium' expression might be due to GFP protein perdurance from enhancers active at earlier developmental stages. We examined each enhancer for the presence of an MF peak and its location with respect to midline-expressed genes (Table S5). In the Fly Enhancer collection there are 102 non-overlapping fragments that drive expression at the midline primordium stage. Of these fragments, 20 (20% of total) have a midline-enriched MF peak, of which 17/20 (85%) reside near a midline-expressed gene (Table S6). This indicates that regulatory elements in the genome with a midline-enriched MF peak that drive expression in a transgenic assay are likely to be used *in vivo*. More common are fragments that have MF peaks but with edgeR FDR values  $> 0.05$ . There were 49 members of this category (48% total) and 22 reside near a midline-expressed gene (45%). Thus, a midline enhancer with a lower ranking MF peak still has almost a 50% chance of being associated with a midline-expressed gene. The final class is midline enhancers that do not have an MF peak. There were 33 members of this class and only eight reside near a midline-expressed



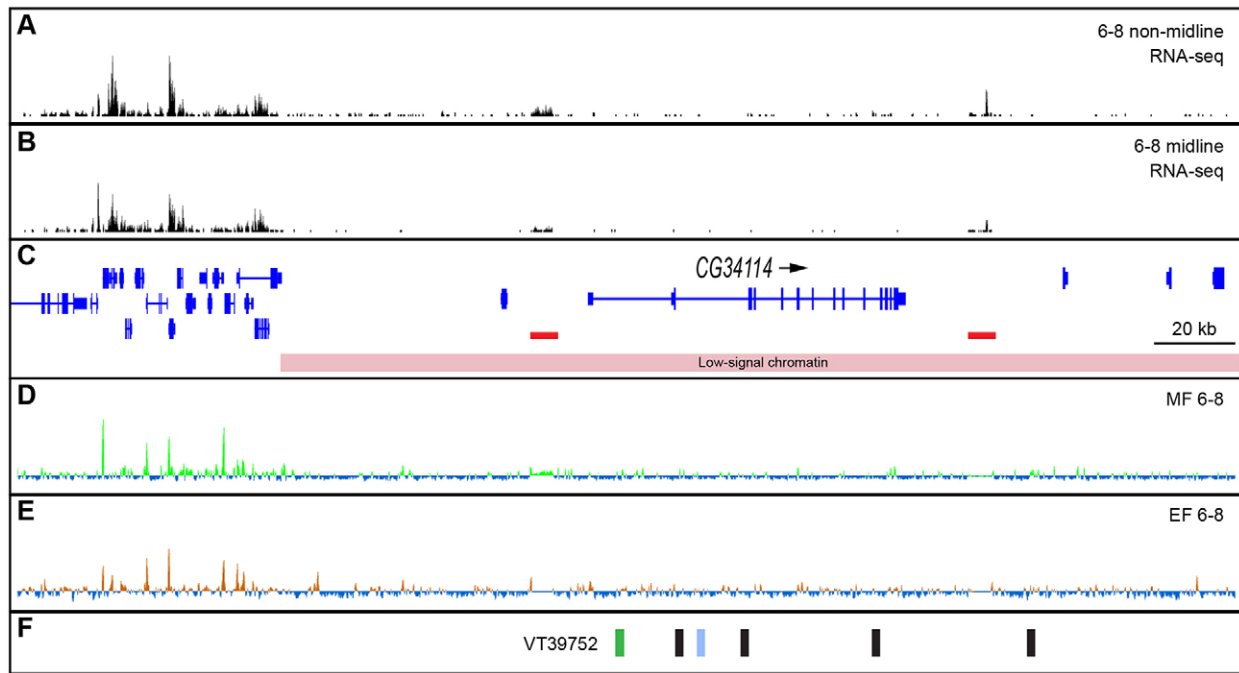
**Fig. 4. Motif analysis of midline-enriched peaks reveals binding sites for Sim-Tgo.** (A) MEME analysis of motif enrichment (AME) of the four groups of midline-enriched peaks with motifs from prominent midline primordium-expressed transcription factors revealed over-representation of Sim-Tgo binding sites in the 'Midline' group, along with sites for the midline-localized transcription factors Oc, Pnt, Vvl and En. Motifs corresponding to the midline transcription factors were not over-represented in the other groups, except for a modest over-representation of the Oc homeobox motif in the 'Gut' and 'None' groups and Sim-Tgo in the 'Gut' group. Heat map shows motif enrichment for the corresponding transcription factor binding sites. (B) Position weight matrix sequence logos are shown for the five transcription factors, along with corresponding *P*-values.

gene (24%). As an example, a fragment (VT39752) from intron 1 of *CG34114* (Fig. 5C,F) drives strong midline expression from late stage 11 through stage 16 (Kvon et al., 2014). Yet, the *CG34114* gene does not have an MF or EF peak (Fig. 5D,E), and *CG34114* expression is not detected in 6-8 h midline or non-midline cells (Fig. 5A,B); nor are the adjacent *CG4683* and *CG6629* genes expressed in midline cells as assessed by RNA-seq or ISH. Consistent with the weak expression, the *CG34114* locus lies within a 174 kb region lacking histone modifications in whole embryos, which typically corresponds to regions of low gene activity (Ho et al., 2014) (Fig. 5C; see below). Not surprisingly, the average strength of the midline enhancers as measured by the Fly Enhancer group is significantly higher for those with a midline-enriched MF peak (3.0) than for enhancers with a lower ranking MF peak (2.4;  $P=0.0037$ ) or without a peak (2.3;  $P=0.0117$ ) (on a scale of 1-4, with 4 having the highest levels; Table S5) (Kvon et al., 2014).

Overall, 55/102 (54%) of the midline enhancer fragments do not reside near a known midline-expressed gene and presumably are not active in midline cells; instead, the midline activity of these

fragments might be a consequence of being removed from the natural genomic context. In summary, fragments that drive midline primordium expression and contain midline-enriched FAIRE peaks generally reside near midline-expressed genes (85%), those fragments with lower ranking MF peaks are closely split between those associated with a midline-expressed gene (45%) and those without (55%), whereas fragments lacking an MF peak tend not to reside near midline-expressed genes (76%) and are unlikely to drive midline expression *in vivo*.

Examination of the location of the enhancers with respect to modified histone chromatin states (Ho et al., 2014) reinforces the distinction between enhancers with an MF peak and those without. Of the midline enhancers with an MF peak, whether midline enriched or not (Fig. S4; MF), 25% lie within chromatin with an 'enhancer' chromatin profile and 39% have a 'Polycomb-repressed' chromatin profile – both chromatin states are associated with developmentally regulated genes. Only 6% of the enhancers without an MF peak reside in 'enhancer' chromatin and 24% reside within 'Polycomb-repressed' chromatin (Fig. S4; No MF).



**Fig. 5. The VT39752 midline primordium enhancer does not reside near a midline-expressed gene.** (A-F) Example of a gene locus with a strong midline enhancer that resides far from a known midline-expressed gene. The *CG34114* gene (C) contains a strong 11-12 h midline intronic enhancer (F; VT39752) (Kvon et al., 2014), yet the gene is not expressed in midline cells (A,B) and contains no significant (D) MF or (E) EF peaks. The *CG34114* gene is flanked by two expressed repeats (red boxes), as detected by RepeatMasker (<http://www.repeatmasker.org/>, current version: open-4.0.5) and resides in an embryonic region lacking histone modifications (Low-signal chromatin; bar at bottom of C). (F) Boxes indicate FlyLight enhancer fragments. Colors indicate activity: green, midline primordium expression; purple, mesectodermal expression; black, no midline expression.

By contrast, regions associated with low gene activity ('low' histone modified or 'heterochromatin' chromatin) are present in enhancers with an MF peak at a low frequency of 7%, whereas 45% of the enhancers without an MF peak were present in the low gene activity chromatin states. These data support the view that midline enhancers with an MF peak (enriched and not-enriched are similar; Fig. S4) tend to reside in chromatin regions associated with active transcription (64%) and infrequently in inactive regions (7%), whereas midline enhancers without an overlapping MF peak are commonly in inactive chromatin regions (45%) and less frequently in active regions (30%).

## DISCUSSION

We have examined the utility and limits of cell type-specific FAIRE-seq as an assay to identify biologically relevant enhancers throughout the genome. Our analysis utilized purified *Drosophila* CNS midline cells for FAIRE-seq, and compared the data with large-scale, developmentally matched ISH, RNA-seq, and enhancer expression datasets. These data allow us to draw a number of conclusions about how to assess the *in vivo* validity of enhancers.

The first point regards the use of FAIRE data to identify *in vivo* relevant, cell type-specific enhancers, which is one of its most important functions. In this study, we demonstrate the utility of matching midline FAIRE (MF) data to whole-embryo FAIRE (EF) data for enhancer recognition. We identified 15,191 6-8 h MF peaks. Most of the MF peaks are unlikely to correspond to relevant midline primordium enhancers and promoters. However, comparing the MF peaks with whole-embryo (EF) peaks using edgeR revealed a group of only 167 peaks that were 'midline-enriched' to an edgeR FDR value  $\leq 0.05$ . Comparing the presence of these peaks with known midline enhancers and midline-expressed genes indicates that a relatively large number are likely to correspond to *in vivo* relevant

enhancers. They are over-represented in Sim-Tgo binding sites and have predictive value. There are additional MF peaks that possess a high ME/EF ratio but have an edgeR value  $>0.05$  that correspond to *in vivo* relevant midline enhancers. Our data strongly indicate that genes with prominent midline primordium expression contain one or more MF peaks and these peaks generally correspond to midline enhancers. Since the GFP-sorted midline cells also contained some gut cells, some midline-enriched peaks were near genes expressed in gut cells. Although this extends the success of FAIRE-seq to identify relevant genes and enhancers, it also reinforces the need to use highly purified cell populations.

While a significant fraction (59%) of the 167 midline-enriched edgeR peaks reside near strong midline (and gut)-expressed genes and most peaks are likely to correspond to cell type-specific enhancers, the significance of the other 41% of midline-enriched peaks remains an open question. One class of midline-enriched peaks was associated with genes that are broadly expressed, including in midline cells, although the midline expression is generally weak (Broad group; 11% of the midline-enriched peaks). One hypothesis is that the midline-enriched MF peak is an enhancer element that contributes to the midline component of the expression pattern of the gene. However, these peaks are not over-represented in Sim-Tgo binding sites (although this might be expected for an enhancer with modest midline activity), and one of the five tested overlapping DNA fragments has significant midline enhancer activity. Thus, it is presently unclear whether the Broad group of midline-enriched peaks commonly contributes to midline gene expression. The members of the final group (30%) are not clearly associated with any midline or gut-expressed genes and their significance is unknown. Still, by combining the MF and EF data to identify a small set of midline-enriched peaks, we can identify genes with relevant expression and enhancers with a likelihood of success



approaching 50% or higher. It is worth emphasizing that all well-characterized midline enhancers of midline-expressed genes (Table S2) have a corresponding MF peak that is either midline enriched or has a high MF/EF ratio. Thus, combining cell type-specific and whole embryo data can identify candidate cell type-specific enhancers with a strong degree of certainty.

Another related issue concerns the relationship between genomic fragments that drive midline expression using *in vivo* enhancer assays to the presence or absence of MF peaks. The Fly Enhancer dataset contains 102 non-overlapping midline primordium enhancers, allowing a robust assessment. Midline-enriched peaks are present on 20% of the midline enhancer fragments and these generally correspond to midline enhancers residing near midline-expressed genes (85%). Another 48% of the fragments have an MF peak with an edgeR value  $>0.05$  and 45% of these enhancers reside near a midline-expressed gene. Thus, most midline enhancer fragments have MF peaks and the majority reside near midline-expressed genes. A third class of midline enhancer fragments (32%) does not have an MF peak, and 76% of these fragments do not reside near a midline-expressed gene. Although it is possible that some function *in vivo* as midline enhancers for genes that reside a large distance away (Kvon et al., 2014), it seems likely that many of these fragments do not function *in vivo* as midline enhancers. This is reinforced by their common occurrence in regions of the genome associated with low transcriptional activity. Considering all 102 fragments with midline primordium activity, 55% of the fragments do not reside near a known midline-expressed gene. Other studies have concluded that gene expression is frequently unaffected by transcription factor binding (Cusanovich et al., 2014) and fragments tested for enhancer activity occasionally show patterns that do not accurately reflect expression of the nearby gene (Kvon et al., 2014). However, in this study we provide evidence that enhancers with specific patterns of activity (midline) frequently reside in inactive chromatin regions and are unlikely to be employed *in vivo* despite their enhancer potential.

In summary, we find that the identification of biologically relevant cell type-specific enhancers requires the purification of the relevant cell type and the use of open chromatin profiling (FAIRE-seq, ATAC-seq, DNase-seq) on chromatin from both purified cells and whole embryos. Only in this manner can one identify cell type-specific enhancers with a fair degree of certainty.

## MATERIALS AND METHODS

### Fluorescence-activated cell sorting (FACS) and cell fixation

Embryo collection, dissociation, and sorting of GFP<sup>+</sup> cells from *sim3.7-Gal4 UAS-mCD8::GFP* embryos were performed as previously described (Fontana and Crews, 2012), except that a Dounce homogenizer was used to dissociate embryos, and cells were resuspended in Hemolymph-Like (HL) buffer (Salmand et al., 2011). GFP<sup>+</sup> cells were sorted as 100,000 cell batches into cold HL buffer, fixed for 5 min with 1% formaldehyde, and quenched with 125 mM glycine for 5 min. Fixed cells were pelleted for 5 min at 2000 *g* and washed twice in 10 mM HEPES in 1×PBS, and then frozen in liquid nitrogen.

### FAIRE-seq

FAIRE was performed as previously described (McKay and Lieb, 2013). FAIRE-seq libraries were sequenced on an Illumina HiSeq 2000 at the UNC High-Throughput Sequencing Facility. Reads were mapped to the dm3 *D. melanogaster* reference genome with bowtie2 (Langmead and Salzberg, 2012). Reads with a map quality score  $<10$  were removed using SAMtools (Li et al., 2009). FAIRE peaks were called on pooled MF data with MACS2 (Zhang et al., 2008) using genomic DNA reads as an input control sample and an extension size of 125 bp. The top 10,000 sorted MF peaks were used

in downstream analysis. For consistency, an equal number of EF peaks was also used. Differentially accessible peaks were determined with edgeR (Robinson et al., 2010) using the ‘classic analysis’. FAIRE peaks with an edgeR FDR  $\leq 0.05$  were called as ‘midline-enriched’.

### RNA-seq data

RNA-seq data from Fontana and Crews (2012) were mapped to *D. melanogaster* genome release dm3 with TopHat (Kim et al., 2013) and BedGraph files were generated using BEDTools (Quinlan and Hall, 2010).

### Motif prediction

DNA sequences corresponding to the 167 midline-enriched MF peaks were analyzed using the AME tool within the MEME suite (McLeay and Bailey, 2010) for over-representation of the binding site motifs of a set of midline-primordium-expressed transcription factors. DNA sequences corresponding to the top 10,000 EF peaks were used as the control. The position weight matrices (PWMs) for these midline transcription factors were obtained from the Fly Factor Survey database (Zhu et al., 2011).

### Chromatin state analysis

The chromatin state of Fly Enhancer fragments with reported midline activity was determined by intersecting with the hiHMM dataset from late embryos (Ho et al., 2014) using BEDTools, requiring a greater than 50% overlap.

### Generation and analysis of transgenic *Drosophila* strains

*sim0.96* was cloned into the *hsp70* promoter-*nucGFP* pMintgate vector (Jiang et al., 2010) to assay its enhancer ability. The *Gal4* lines *simC2.3*, *simD2.1* and *simE2.3* were described previously (Freer et al., 2011). FlyLight *Gal4* enhancer lines were obtained from the Bloomington *Drosophila* Stock Center, and embryos were subjected to fluorescent ISH with a *Gal4* probe or crossed to *UAS-mCD8::GFP* and immunostained with anti-GFP as previously described (Pearson and Crews, 2014; Wheeler et al., 2006).

### Acknowledgements

We are grateful to the Bloomington *Drosophila* Stock Center for contributing *Drosophila* stocks, Sue Celniker and the Berkeley *Drosophila* Genome Project for use of embryonic images, and Joseph Fontana and Barry Udis (UNC Flow Cytometry Core Facility) for providing helpful advice on the FACS isolation of CNS midline cells. The UNC Research Computing Center provided support on installing and running computer applications.

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

J.C.P. and D.J.M. designed and performed the experiments. J.C.P., D.J.M. and S.T.C. analyzed the data and wrote the manuscript. J.D.L. supported the initiation of this study.

### Funding

The project was supported by a National Institutes of Health/National Institute of Neurological Disorders and Stroke grant [R01 NS64264] to S.T.C., and D.L.M. was supported by University of North Carolina institutional start-up funds. Deposited in PMC for release after 12 months.

### Data availability

FAIRE-seq data have been deposited at Gene Expression Omnibus (GEO) under accession number GSE83463.

### Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.136895.supplemental>

### References

- Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* **43**, W39–W49.
- Campos-Ortega, J. A. and Hartenstein, V. (1997). *The Embryonic Development of Drosophila melanogaster*. Berlin; Heidelberg: Springer Verlag.

- Cheng, Y., Brunner, A. L., Kremer, S., DeVido, S. K., Stefaniuk, C. M. and Kassisi, J. A. (2014). Co-regulation of *invected* and *engrailed* by a complex array of regulatory sequences in *Drosophila*. *Dev. Biol.* **395**, 131-143.
- Crews, S. T. (2009). Axon-glia interactions at the *Drosophila* CNS midline. *Cell Adh. Migr.* **4**, 66-70.
- Cusanovich, D. A., Pavlovic, B., Pritchard, J. K. and Gilad, Y. (2014). The functional consequences of variation in transcription factor binding. *PLoS Genet.* **10**, e1004226.
- Fontana, J. R. and Crews, S. T. (2012). Transcriptome analysis of *Drosophila* CNS midline cells reveals diverse peptidergic properties and a role for *castor* in neuronal differentiation. *Dev. Biol.* **372**, 131-142.
- Freer, S. M., Lau, D. C., Pearson, J. C., Talsky, K. B. and Crews, S. T. (2011). Molecular and functional analysis of *Drosophila single-minded* larval central brain expression. *Gene Expr. Patterns* **11**, 533-546.
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877-885.
- Gustavson, E., Goldsborough, A. S., Ali, Z. and Kornberg, T. B. (1996). The *Drosophila engrailed* and *invected* genes: partners in regulation, expression and function. *Genetics* **142**, 893-906.
- Ho, J. W. K., Jung, Y. L., Liu, T., Alver, B. H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M. Y., Appert, A. et al. (2014). Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449-452.
- Jacobs, J. R. (2000). The midline glia of *Drosophila*: a molecular genetic model for the developmental functions of glia. *Prog. Neurobiol.* **62**, 475-508.
- Jiang, L., Pearson, J. C. and Crews, S. T. (2010). Diverse modes of *Drosophila* tracheal fusion cell transcriptional regulation. *Mech. Dev.* **127**, 265-280.
- Kasai, Y., Stahl, S. and Crews, S. (1998). Specification of the *Drosophila* CNS midline cell lineage: direct control of *single-minded* transcription by dorsal/ventral patterning genes. *Gene Expr.* **7**, 171-189.
- Kearney, J. B., Wheeler, S. R., Estes, P., Parente, B. and Crews, S. T. (2004). Gene expression profiling of the developing *Drosophila* CNS midline cells. *Dev. Biol.* **275**, 473-492.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.
- Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91-95.
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359.
- Lewis, J. O. and Crews, S. T. (1994). Genetic analysis of the *Drosophila single-minded* gene reveals a central nervous system influence on muscle development. *Mech. Dev.* **48**, 81-91.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Manning, L., Heckscher, E. S., Purice, M. D., Roberts, J., Bennett, A. L., Kroll, J. R., Pollard, J. L., Strader, M. E., Lupton, J. R., Dyukareva, A. V. et al. (2012). A resource for manipulating gene expression and analyzing *cis-regulatory* modules in the *Drosophila* CNS. *Cell Rep.* **2**, 1002-1013.
- McKay, D. J. and Lieb, J. D. (2013). A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev. Cell* **27**, 306-318.
- McLeay, R. C. and Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165.
- Meng, X., Brodsky, M. H. and Wolfe, S. A. (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.* **23**, 988-994.
- Muralidhar, M. G., Callahan, C. A. and Thomas, J. B. (1993). Single-minded regulation of genes in the embryonic midline of the *Drosophila* central nervous system. *Mech. Dev.* **41**, 129-138.
- Nambu, J. R., Franks, R. G., Hu, S. and Crews, S. T. (1990). The *single-minded* gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell* **63**, 63-75.
- Nambu, J. R., Lewis, J. L., Wharton, K. A. and Crews, S. T. (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.
- Pearson, J. C. and Crews, S. T. (2014). Enhancer diversity and the control of a simple pattern of *Drosophila* CNS midline cell expression. *Dev. Biol.* **392**, 466-482.
- Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Salmand, P.-A., Iché-Torres, M. and Perrin, L. (2011). Tissue-specific cell sorting from *Drosophila* embryos: application to gene expression analysis. *Fly* **5**, 261-265.
- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. M. (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436-449.
- Thomas, J. B., Crews, S. T. and Goodman, C. S. (1988). Molecular genetics of the *single-minded* locus: a gene involved in the development of the *Drosophila* nervous system. *Cell* **52**, 133-141.
- Tomancak, P., Beaton, A., Weiszmam, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, research0088.1.
- Tomancak, P., Berman, B. P., Beaton, A., Weiszmam, R., Kwan, E., Hartenstein, V., Celniker, S. E. and Rubin, G. M. (2007). Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145.
- Watson, J. D. and Crews, S. T. (2012). Formation and specification of a *Drosophila* dopaminergic precursor cell. *Development* **139**, 3316-3325.
- Watson, J. D., Wheeler, S. R., Stagg, S. B. and Crews, S. T. (2011). *Drosophila hedgehog* signaling and *engrailed-runt* mutual repression direct midline glia to alternative ensheathing and non-ensheathing fates. *Development* **138**, 1285-1295.
- Wharton, J. K. A., Franks, R. G., Kasai, Y. and Crews, S. T. (1994). Control of CNS midline transcription by asymmetric E-box elements: similarity to xenobiotic responsive regulation. *Development* **120**, 3563-3569.
- Wheeler, S. R., Kearney, J. B., Guardiola, A. R. and Crews, S. T. (2006). Single-cell mapping of neural and glial gene expression in the developing *Drosophila* CNS midline cells. *Dev. Biol.* **294**, 509-524.
- Wheeler, S. R., Stagg, S. B. and Crews, S. T. (2008). Multiple *Notch* signaling events control *Drosophila* CNS midline neurogenesis, gliogenesis and neuronal identity. *Development* **135**, 3071-3079.
- Wheeler, S. R., Stagg, S. B. and Crews, S. T. (2009). MidExDB: a database of *Drosophila* CNS midline cell gene expression. *BMC Dev. Biol.* **9**, 56.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.
- Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., Brasfield, J. A., Zhu, C., Asriyan, Y., Lapointe, D. S. et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39**, D111-D117.