

METHOD

Open Access



# Quantitative-enhancer-FACS-seq (QeFS) reveals epistatic interactions among motifs within transcriptional enhancers in developing *Drosophila* tissue

Colin T. Waters<sup>1,2†</sup>, Stephen S. Gisselbrecht<sup>1†</sup> , Yuliya A. Sytnikova<sup>1</sup>, Tiziana M. Cafarelli<sup>3,4,5</sup>, David E. Hill<sup>3,4,5</sup> and Martha L. Bulyk<sup>1,2,3,6\*</sup>

\* Correspondence: [mlbulyk@genetics.med.harvard.edu](mailto:mlbulyk@genetics.med.harvard.edu)

<sup>†</sup>Colin T. Waters and Stephen S. Gisselbrecht contributed equally to this work.

<sup>1</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

<sup>2</sup>Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA 02138, USA

Full list of author information is available at the end of the article

## Abstract

Understanding the contributions of transcription factor DNA binding sites to transcriptional enhancers is a significant challenge. We developed Quantitative enhancer-FACS-Seq for highly parallel quantification of enhancer activities from a genomically integrated reporter in *Drosophila melanogaster* embryos. We investigate the contributions of the DNA binding motifs of four poorly characterized TFs to the activities of twelve embryonic mesodermal enhancers. We measure quantitative changes in enhancer activity and discover a range of epistatic interactions among the motifs, both synergistic and alleviating. We find that understanding the regulatory consequences of TF binding motifs requires that they be investigated in combination across enhancer contexts.

**Keywords:** Enhancers, Transcription factor binding sites, Epistasis, *Drosophila* embryonic mesoderm, Reporter assay

## Background

Identifying sequence features of *cis*-regulatory elements that are important for their activities is a key challenge in understanding how regulatory information is encoded in genomes. Computational analysis of transcriptional enhancers with overlapping tissue and temporal expression patterns for enrichment of sequence matches to transcription factor (TF) DNA binding site motifs can aid in the identification of TFs that contribute to the enhancers' activities [1, 2]. Numerous studies have investigated the contributions of individual motifs to enhancer activity [3–6]. In many cases, combinatorial regulation by multiple TFs is required to specify an appropriate expression pattern, and the contribution of a given motif instance to transcriptional regulation is sensitive to context. This complicates the interpretation of motif enrichment. Experimental analysis of enhancers in which motif occurrences have been mutated in various combinations is



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

needed to determine the contributions of the motifs to the activities of the enhancers and whether a motif exerts different effects depending on the enhancer context [7, 8].

Various genomic sequence features surrounding a TF binding site can present a (dis-)favorable context for binding by the corresponding TF. Local sequence context can affect the probability of TF binding by various mechanisms, such as by providing a flanking sequence that adopts a DNA shape preferred by a particular TF [9], contains a motif that is preferential for co-binding with a protein partner(s) [10–12], contains a motif bound by a TF that affects local chromatin accessibility [13–15], contains a motif for competitive binding by a different TF [16, 17], or corresponds to sequence unfavorable for nucleosomes [18]. The potential for motifs present within enhancers to influence the activity of nearby motifs is referred to as epistasis or genetic interaction. Such epistatic interactions, where the regulatory output of a particular combination of motifs is different from that expected if the motifs are treated as acting independently, are important to identify: rules learned for individual motifs in one enhancer context might not be informative of those motifs' regulatory consequences in a different enhancer context. In a study of human TF binding sites, linear regression models that incorporated pairwise interactions between motifs in 145-bp sequences explained a higher fraction of the variance in activation of reporter expression, with five motif pairs showing significant interactions with regard to activity of both genomic and synthetic enhancer sequences [19]. A challenge for investigations to identify context-specific effects of motifs on enhancer activity is the need for quantitative data on wildtype enhancers and enhancers mutated for particular motifs across a panel of different enhancer sequence contexts.

Enabled by advances in high-throughput DNA sequencing and oligonucleotide synthesis, various technologies have been developed for highly parallel analysis of genomic or synthetic sequences for promoter or enhancer activity in a reporter assay [18, 20]. Massively parallel reporter assays (MPRAs) [21] and STARR-Seq [22] are powerful in their ability to assay tens to hundreds of thousands of short elements (typically less than 200 bp in length) for their ability to drive expression of a reporter in cultured cells. While these assays have allowed the identification of nucleotides within enhancers that are important for their *cis*-regulatory activity, they have a number of limitations. First, they employ episomal reporter constructs, but the activities of chromosomally integrated sequences can differ substantially from their activities when tested in episomes [23]. Furthermore, not all cell types are well modeled by cell lines, and the activities of some tissue-specific enhancers may not be cell autonomous. While MPRAs have been performed in retinal explants [24–26] and in mouse liver [27], there is a need to assay enhancers in additional tissues in the context of whole organisms, such as for tissue-specific developmental enhancers active in embryos [28, 29].

The enhancer-FACS-seq (eFS) technology addresses these needs by using (1) a genomically integrated enhancer-reporter system in live, developing animals; (2) cell-type-specific-expression of an ectopic cell-surface marker to isolate a cell population of interest; and (3) analysis of reporter gene expression driven from a site-specific integration site with low background activity. This technology is limited in throughput by the efficiency of *Drosophila* transgenesis, which permits the use of lower-throughput library construction methods that test the activity of longer sequences, making it a complement to higher-throughput episomal methods. eFS analysis of a pool of candidate

enhancers identified by various criteria, including enrichment for occupancy by mesodermal TFs or clustering of conserved TF binding motifs, validated 12 known and identified 45 novel enhancers active in *Drosophila* embryonic mesoderm [1].

Motif enrichment and pairwise motif analysis of these mesodermal enhancers revealed enrichment of the DNA binding specificity motifs of TFs with a known role in mesoderm development and also of factors not previously known to have a role in transcriptional regulation in mesoderm [1]. These putative novel mesodermal regulatory motifs included those corresponding to the factors Deaf1, Schlank, ZIPIC, and CG12236PB. Best characterized of these four factors is Deaf1, which was previously identified as a cofactor for the homeotic TF Deformed (Dfd) [30] and mutants of which led to embryonic segmentation defects [31]. Both repressive and activating activities of Deaf1 have been reported in different cell types [32], suggesting Deaf1 may exhibit cell type-specific activity, but its overall regulatory activity remains poorly understood. ZIPIC has been characterized as a putative insulator protein that physically interacts with CP190 [33] and was subsequently found to bind topologically associated domain (TAD) boundaries [34], supporting its putative role as an insulator. Schlank was initially characterized as a transmembrane protein with a role in fatty acid synthesis [35]; intriguingly, its homeodomain has been shown to play a role in fatty acid homeostasis independent of its enzymatic activity [36]. CG12236-PB has not been investigated beyond the identification of its DNA binding motif through a bacterial one-hybrid screen [37]. All of these factors have maternally deposited mRNA and ubiquitous expression early in embryogenesis [38], which may have contributed to their failure to be characterized by classical *Drosophila* genetics.

We sought to develop a method that would allow us to uncover the context-dependent, combinatorial impact of the DNA binding motifs of these four TFs on tissue-specific activities of the mesodermal enhancers that we had identified by eFS. To address this need, here we have developed Quantitative enhancer-FACS-Seq (QeFS) technology, which provides an RNA-Seq-based readout of enhancer activity. QeFS thus provides a quantitative readout of enhancer activity, allowing sensitive measurement of the effects of loss of individual or combinations of motifs within an enhancer on its activity. Comparisons of QeFS data among a series of enhancer variants enabled us not only to quantify the contributions of individual motifs but also to identify epistatic interactions between motifs through the analysis of combinatorially mutated enhancers. Such sensitive, quantitative readout is crucial for understanding the activity of factors that have subtle effects on relative gene expression levels. We attempted QeFS in whole mesoderm and separately in fusion competent myoblasts (FCMs), a small subset of the mesoderm, in developing *Drosophila* embryos to study the roles of the DNA binding motifs recognized by the TFs Deaf1, Schlank, ZIPIC, and CG12236PB on mesodermal enhancer activity and to identify epistatic interactions among these motifs.

## Results

### Quantitative measurement of the contributions of TF DNA binding motifs to enhancer activity in embryos

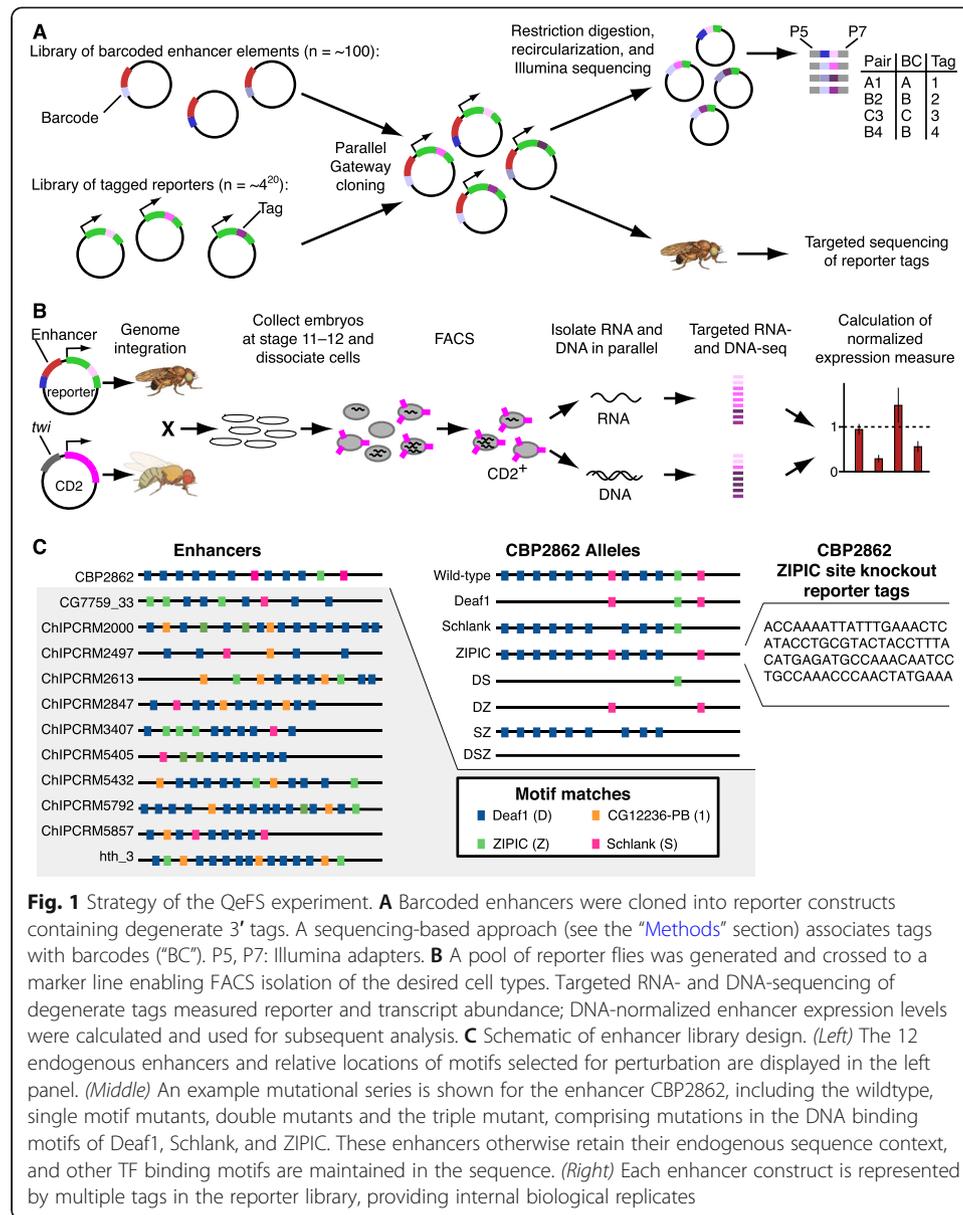
To assay *cis*-regulatory elements for enhancer activity in a quantitative, highly parallel fashion in a chromosomal context in the context of particular tissues in whole

*Drosophila* embryos, we developed quantitative enhancer-FACS-Seq (QeFS). In QeFS, each *cis*-regulatory element was cloned upstream of a reporter gene, in whose 3' UTR a degenerate sequence was inserted and used as an identifying tag in RNA-Seq readout of enhancer activity (Additional file 1: Fig. S1, S2, S3). To simplify the process by which highly similar *cis*-regulatory elements were distinguished, we barcoded the elements. Briefly, we synthesized a library of wildtype and mutant enhancers with flanking 8-mer barcodes and cloned them into a reporter backbone, with each reporter containing a unique, degenerate 20-mer tag in the 3' UTR of the reporter transcript (Fig. 1A). Sequence-validated reporter constructs were pooled to create a library of 303 unique enhancer-reporter constructs. We used a custom high-throughput sequencing library preparation protocol involving digestion, re-circularization, and sequencing [27] to generate a lookup table that associates reporter tags to the corresponding enhancer barcodes (Additional file 1: Fig. S4, S5, S6, S7). Typically at least 3 different tags were associated with each enhancer (Additional file 2: Table S1; Additional file 1: Fig. S8).

Next, as in eFS [1], we injected the enhancer-reporter library into *Drosophila* embryos, resulting in genomic integration of individual enhancer-reporter elements at a fixed landing site in each haploid genome of an embryo's germ cell progenitors. Assaying all the wildtype and mutant enhancer constructs from the same chromosomal integration site allowed us to determine the effects of a particular mutation or combination of mutations on the activity of an enhancer. Following a series of fly crosses, a population of flies was obtained, each with a single enhancer-reporter construct present in the genome of every cell in the adult fly. We marked our cell types of interest by crossing this population to a fly line expressing the rat cell surface marker CD2 under the control of a cell type-specific enhancer (either *twi* for whole mesoderm or *Mef2-I-E<sub>D5</sub>* for FCMs) (Fig. 1B). Embryos from this cross were aged to the desired stage (i.e., stages 11–12), harvested, and used to prepare single-cell suspensions. These suspensions were then subjected to FACS for the CD2 marker, to purify our cell types of interest (Additional file 1: Fig. S9; Additional file 2: Table S2). The small numbers (generally less than  $10^5$ ; see Additional file 2: Table S2) of cells recovered from the more specific *mef2*:CD2+ population gave information on very few tested elements after filtering to remove unreliable measurements (see the “Methods” section), establishing the limitations of the method for rare cell types at this scale. The complementary *mef2*:CD2- population was more reliably measured and provided a proxy for whole embryo measurements, as very few cells are excluded.

We processed the collected cell populations (CD2+ and CD2-) to prepare RNA-seq and DNA-seq libraries in parallel from the same collection of informative cells, representing the reporter tag in the 3' UTR of the reporter transcript, or the reporter tag in the 3' UTR of the reporter DNA template, respectively (Additional file 1: Fig. S10, S11). We used a unique molecular identifier (UMI) strategy to count individual molecules containing each tag, assigned tag counts to enhancers using the lookup table, and normalized the tag counts in the RNA-seq library to those in the DNA-seq library, thus scaling reporter transcript levels to the underlying abundance of individual reporters in the fly population (Additional file 1: Fig. S12, S13).

In order to test this experimental strategy, we designed a library of wildtype and mutant enhancers intended to take advantage of the strengths of this approach, including the resulting quantitative reporter data. We previously found that the DNA binding site



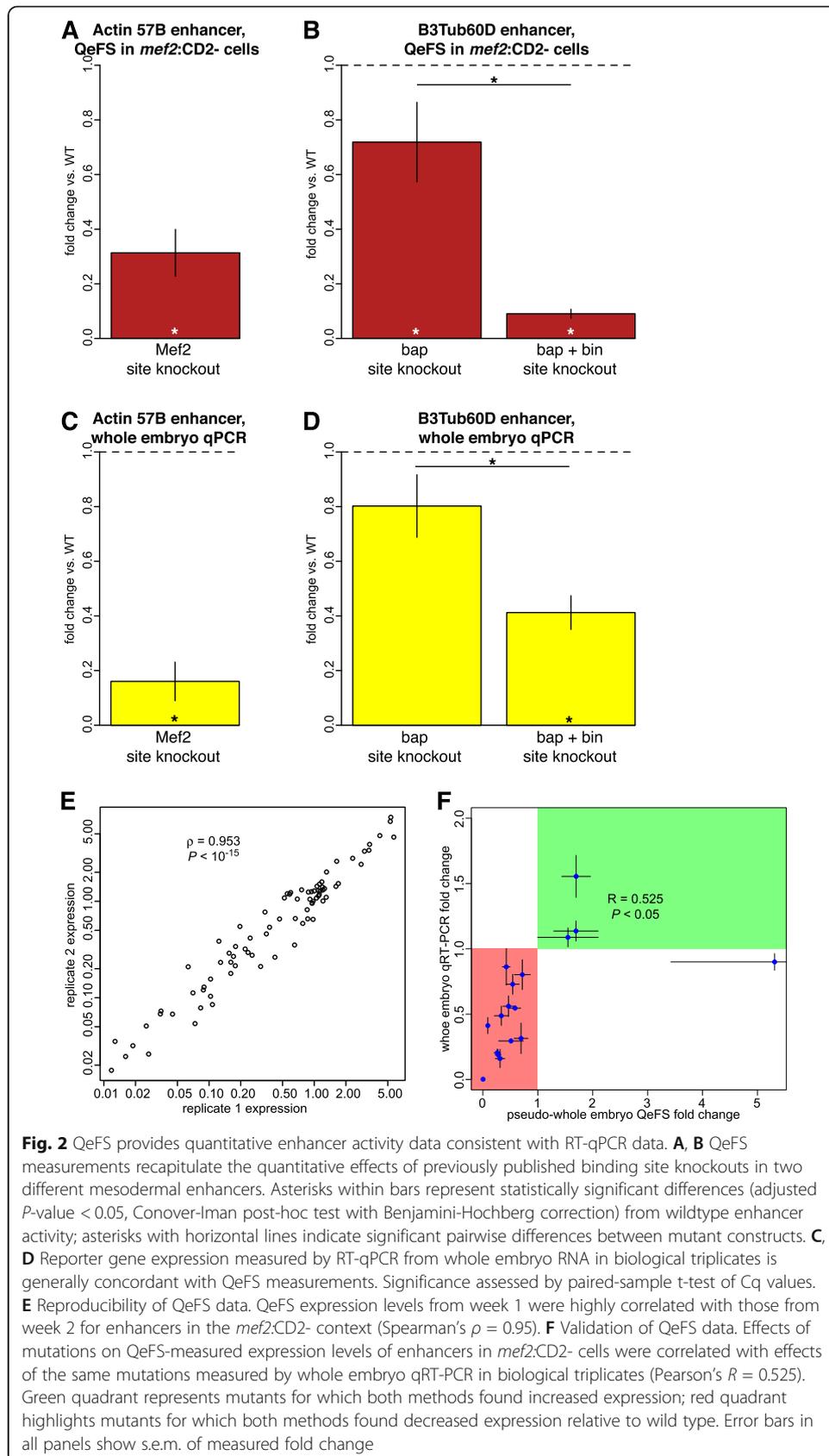
motifs of the ubiquitously expressed [39] TFs Deaf1, CG7928 (ZIPIC), CG12236-PB, and Schlank were enriched among mesodermally active enhancers [1]. Although these TFs were not known to regulate embryonic mesoderm development, the enrichment of their recognition motifs both individually and together with each other and with master regulators in either whole mesoderm or fusion-competent myoblasts (FCMs), an abundant mesodermal subpopulation, suggested that they may contribute to mesodermal enhancer activity. To investigate the roles of these motifs in mesodermal enhancer activity, we selected a panel of 12 endogenous enhancers of approximately 1 kb in length that are active in embryonic mesoderm during developmental stages 11–12 [1] and that each contain matches to three of these four motifs. To allow us to determine the contributions of each of the Deaf1, CG7928 (ZIPIC), CG12236-PB, and Schlank motifs to enhancer activity, we designed a series of systematic single-motif mutations, whereby in

each enhancer we mutated all the sites matching each of these motifs, considering each motif one at a time (“single-motif mutants”). Additionally, to identify potential regulatory interactions between these motifs, we designed combinatorial enhancer mutants comprising mutations in all possible pairs and three-way combinations of these motif mutants; thus, for each of the twelve enhancers, we synthesized eight different versions: wildtype, three single-motif mutants, three pairwise mutants, and one triple mutant (Fig. 1C, Additional file 3: Table S3). Mutant enhancers differed from wildtype by as little as 1 nucleotide out of ~ 1000, with a mean of 7.2 nucleotides mutated per enhancer. In total, we collected QeFS data for 88 wildtype or mutant enhancers and 5 spiked-in control elements (2 wildtype and 3 mutants) described below. Normalized reporter expression data generally correlated well across pairs of tags measuring the activity of the same enhancer construct in a given collection (Spearman’s  $\rho = 0.786$ ); enhancer construct activities were calculated as the mean normalized value across tags across replicate QeFS experiments.

As positive controls to assess the sensitivity and cell-type specificity of QeFS, we analyzed two enhancers known to exhibit mesoderm-specific activity during embryonic stages 11–12 and previously found to be sensitive to either a single or a set of motif mutations. One element, for *Actin 57B*, showed early visceral mesoderm activity, spreading to whole mesoderm by stage 12 of development [40]. This element was reported to be very sensitive to loss of a single binding site for the TF Mef2, with 87% reduction in *lacZ* reporter activity [40]. The second enhancer, for  *$\beta$ 3Tub60D*, had been reported to be active in developing visceral mesoderm [41]. Imaging experiments using *lacZ* reporters revealed that this element was moderately sensitive to a single site mutation for the TF bagpipe (bap), while combinatorial mutations in two biniou (bin) sites and this single bap site led to complete loss of enhancer activity [41]. As previously described, both enhancers’ activities are restricted to the mesoderm, and none of the selected site mutations caused any detectable non-mesodermal expression.

We cloned each of these wildtype and mutant enhancers—wildtype and Mef2 site mutant *Actin 57B* enhancer, and wildtype, bap site mutant, and triple mutant  *$\beta$ 3Tub60D* enhancers—barcoded them, and spiked them into our enhancer library (see the “Methods” section; Additional file 1: Fig. S14; Additional file 3: Table S3). In parallel, we also created homozygous fly strains carrying reporter constructs for each of these five elements; embryos from each of these individual homozygous strains were harvested at developmental stages 11–12 for analysis of enhancer activity by imaging (Additional file 1: Fig. S15) to confirm the published effects of each mutation or by RT-qPCR.

Analysis of these control enhancers revealed high concordance of enhancer activity measured by QeFS from the spike-in constructs in *mef2:CD2-* cells (Fig. 2A,B) as a proxy for measurements from whole embryos, with that determined by RT-qPCR (Fig. 2C,D) from whole embryos from the homozygous strains. These reporter assay data of enhancer activity were consistent with imaging results of the activities of these enhancers [40, 41]. For example, the single Mef2 site mutation in the *Actin 57B* enhancer led to a significant decrease in enhancer activity by QeFS and of similar magnitude as that measured by RT-qPCR ( $P = 0.0007$ , Welch’s two-sample, one-tailed *t*-test). Similarly, the single Bap site mutation in the  *$\beta$ 3Tub60D* enhancer led to a detectable decrease in enhancer activity by both QeFS ( $P = 0.0258$ ) and RT-qPCR ( $P > 0.05$ ), while



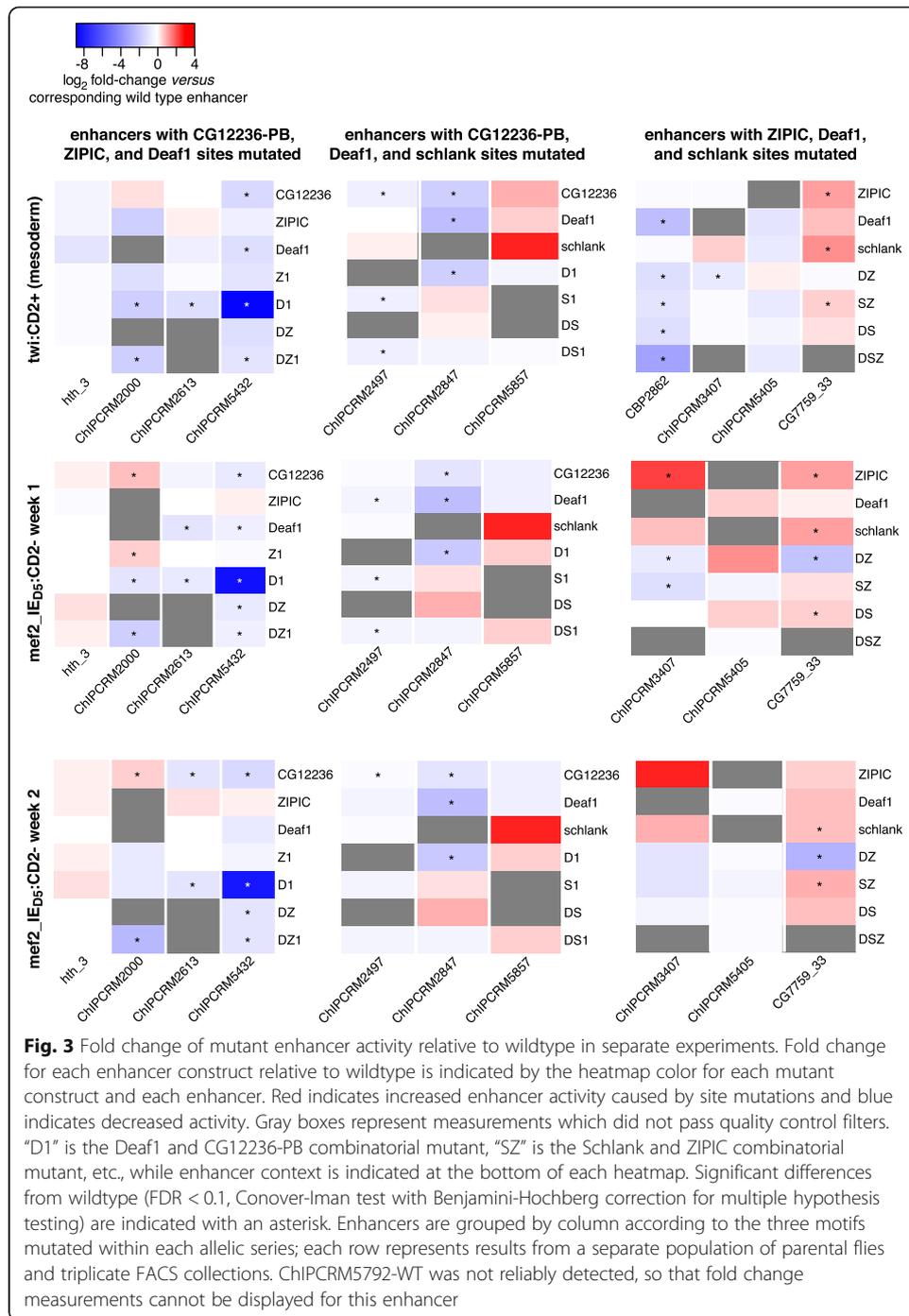
the Bin+bap triple site mutant led to a significant decrease in enhancer activity by both QeFS and RT-qPCR versus both the wildtype enhancer ( $P < 0.0001$ , Conover-Iman post-hoc test with Benjamini-Hochberg correction) and the single mutant ( $P = 0.0001$ ).

To assess the reproducibility of QeFS, we compared the mean measured expression level (normalized by input abundance) across all tested elements between two independent collections of *mef2:CD2-* material. The reporter transcript levels between the two experiments were highly reproducible (Spearman's  $\rho = 0.953$ ,  $P < 10^{-15}$ ) (Fig. 2E). We also compared the results of QeFS from *mef2:CD2-* cells to RT-qPCR measurement on whole embryos for 14 additional mutant constructs recovered at random or selected for replication (Fig. 2F). Fold change measurements vs. wild type were significantly correlated by both methods (Pearson's  $R = 0.525$ ,  $P < 0.05$ ) and concordant in 16/17 cases. Imaging of antibody-stained embryos for the one discordant case (Additional file 1: Fig. S16) confirmed that it is a false positive of QeFS; all other validations by imaging broadly supported the QeFS measurements (Additional file 1: Fig. S17, S18, S19, S20).

Altogether, these results demonstrate that QeFS enables highly accurate determination of the activities of *cis*-regulatory elements from a fixed, chromosomal context in *Drosophila*, allowing comparisons of the effects of TF binding site mutations on enhancer activity.

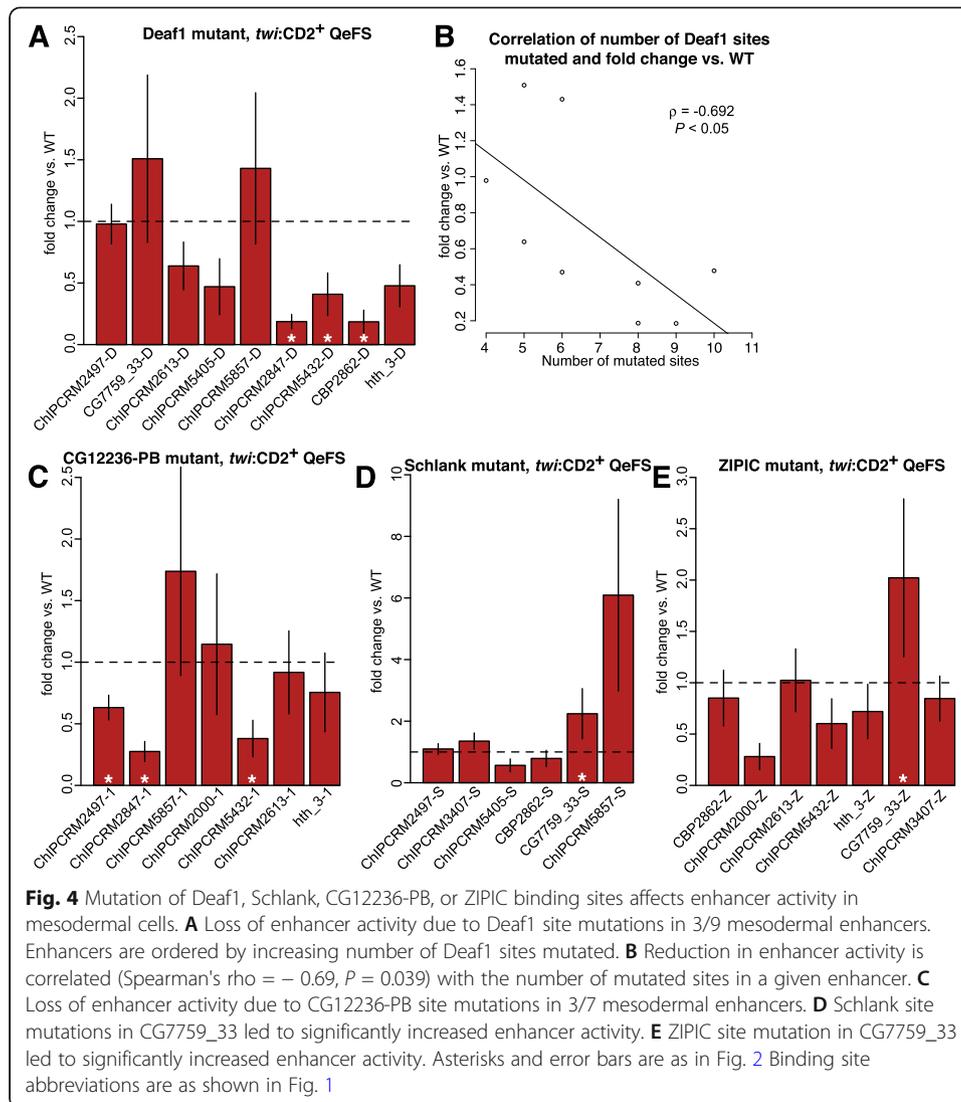
#### **Loss of Deaf1 or CG12236-PB sites causes context-dependent reduction in mesodermal enhancer activity**

To determine the contribution of the Deaf1, ZIPIC, CG12236-PB, and Schlank DNA binding site motifs to each enhancer's activity in each of two different tissue types (whole mesoderm, FCMs), we compared the normalized RNA-Seq level of each mutant enhancer to that of the corresponding wildtype enhancer in each tissue. We then inspected the resulting QeFS fold change values for trends in motif activity across the analyzed panel of enhancers (Fig. 3; Additional file 4: Table S4). We observed that loss of Deaf1 sites led to decreased enhancer activity in whole mesoderm when it had any reliably detectable effect, with significant reduction in 3/9 enhancers (Fig. 4A; Additional file 1: Fig. S21, S22, S23, S24, S25, S26, S27, S28, S29), suggesting that Deaf1 generally acts as an activator in mesodermal cells at stages 11–12. Notably, different enhancers were differentially sensitive to Deaf1 binding site mutations. Five of the 12 enhancers were sensitive to mutation of Deaf1 sites in at least one cell type. For example, Deaf1 site mutation within the ChIPCRM2847 enhancer led to a significant decrease in activity in whole mesoderm ( $P < 0.0001$ ) (Fig. 4A) and in non-FCM cells ( $P < 0.0001$ ) (Fig. 3), which approximate whole embryo as described above. In other cases, the effect was tissue-specific; for example, Deaf1 site mutation in ChIPCRM2497 resulted in loss of enhancer activity only in non-FCM cells ( $P = 0.0079$ ). In two enhancers (ChIPCRM5857 and CG7759\_33), Deaf1 site mutations led to a measurable, but not statistically significant, increase in enhancer activity in whole mesoderm (Fig. 4A). This variability in response level and direction is consistent with a prior analysis of Deaf1 activity at a single promoter region, in which single site knockouts led to variable effects on reporter expression, with the strongest effect being loss of activity following site mutation [42], as we observed here. Interestingly, in whole mesoderm, there was a correlation (Spearman's  $\rho = -0.692$ ,  $P < 0.05$ ) between the number of sites mutated in a



given enhancer and the overall reduction in enhancer activity, suggesting that homotypic clustering of Deaf1 binding sites within an enhancer contributes to larger increases in mesodermal enhancer activity (Fig. 4B).

Mutation of CG12236-PB sites also tended to reduce enhancer activity, with 3/7 tested enhancers showing significant reduction in whole mesoderm (Fig. 4C; Additional file 1: Fig. S21, S23, S25, S26, S27, S29, S30). For example, ChIPCRM2847 activity in mesodermal cells decreased upon mutation of CG12236-PB sites ( $P < 0.0001$ ). To our knowledge, a functional activity for CG12236-PB had not been demonstrated



previously. Our data suggest that CG12236-PB is a novel, sequence-specific transcriptional activator in embryonic mesoderm development.

#### Effects of loss of ZIPIC or Schlank binding sites are highly context-dependent

In contrast to reductions in enhancer activity due to Deaf1 or CG12236-PB binding site mutation, the QeFS data revealed few consistent changes in enhancer activity resulting from mutation of ZIPIC or Schlank binding sites alone. We found that Schlank binding site mutations in CG7759\_33 led to a significant increase in enhancer activity in whole mesoderm ( $P = 0.0048$ ) (Fig. 4D; Additional file 1: Fig. S21, S22, S24, S25, S28, S31), consistent with a prior report that Schlank contributes to transcriptional repression of the *Drosophila* gene *lip3* [36]. A significant increase in enhancer activity was also observed for ZIPIC mutations in CG7759\_33 in mesodermal cells ( $P = 0.0099$ ) (Fig. 4E; Additional file 1: Fig. S22, S23, S27, S28, S29, S30, S31) and in ChIPCRM3407 in non-FCM cells ( $P = 0.0087$ ) (Fig. 3). Where changes in expression are detected upon

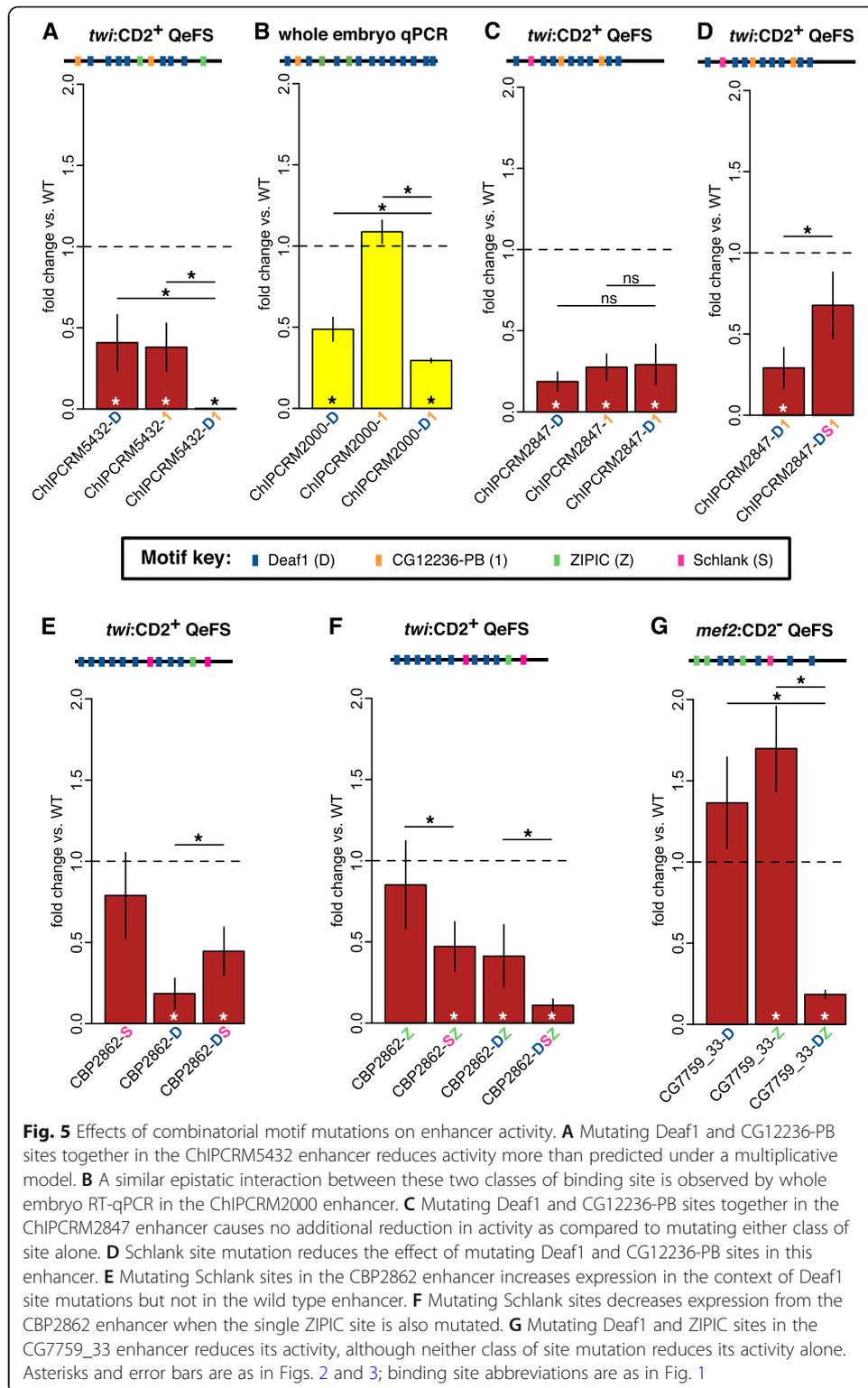
mutation of ZIPIC sites alone, they do not appear to reflect changes in the timing or pattern of expression (Additional file 1: Fig. S16, S19), suggesting an effect on the per-cell expression level.

### QeFS reveals epistatic interactions among sites within enhancers

Bolstered by the ability of QeFS to determine the effects of mutations of an individual TF's binding sites within enhancers, we next analyzed the QeFS data for evidence of potential regulatory dependencies (also known as genetic interactions, or epistatic interactions) among motifs within the enhancers. Briefly, we examined our data for “synergistic” genetic interactions, where mutation of two motifs results in a greater change in reporter gene expression than is predicted from the product of the changes observed upon mutation of sites of the individual motifs [43], and for “alleviating” (also termed “antagonistic”) genetic interactions, where mutation of two motifs results in a less severe change in reporter gene expression than is predicted from the product of the changes observed upon mutation of sites of the individual motifs [44]. Among our panel of combinations of motif mutations across these 12 mesodermal enhancers, we observed cases of both synergistic and alleviating genetic interactions. Notably, all of the epistatic interactions that we observed between motifs were dependent on the enhancer context.

In *twi*:CD2+ cells, where mutation of the 8 Deaf1 sites or of a single CG12236-PB site in the ChIPCRM5432 enhancer both led to a significant decrease in expression relative to the wildtype enhancer ( $P = 0.0147$  or  $0.0148$ , respectively) (Fig. 5A, Additional file 1: Fig. S27), mutation of both the Deaf1 and CG12236-PB sites resulted in a synergistic reduction in enhancer activity ( $P < 10^{-15}$ ). Like the effects of ZIPIC mutations described above, this epistasis appears to occur uniformly in all the cells in which this enhancer drives expression (Additional file 1: Fig. S18). We examined the other enhancers in which Deaf1 and CG12236-PB sites were mutated (Fig. 3) to see if this synergy was observed in other contexts. As QeFS data were not available for ChIPCRM2000-D, we generated transformant lines and observed synergistic effects between these two motifs in ChIPCRM2000 by whole-embryo qRT-PCR (Fig. 5B) and by imaging of stained embryos (Additional file 1: Fig. S17); mutating a single CG12236-PB site in this enhancer caused a significant reduction of activity ( $P = 0.0382$ ) in the context of Deaf1 site knockouts but not on its own. In contrast, mutation of either Deaf1 or CG12236-PB sites in ChIPCRM2847 reduced its enhancer activity in mesodermal cells by over 70% ( $P < 0.0001$ ), but mutation of sites for both motifs in that enhancer had no greater effect than mutation of sites for either factor alone ( $P > 0.2$ , Kruskal-Wallis rank sum test with Benjamini-Hochberg correction for multiple hypothesis testing) (Fig. 5C, Additional file 1: Fig. S26), suggesting that the regulatory outputs of CG12236-PB and Deaf1 for this enhancer are dependent on each other. These results demonstrate a range of context-dependent interactions between Deaf1 and CG12236-PB sites in mesodermal enhancers.

The reduction of ChIPCRM2847 enhancer activity resulting from mutation of Deaf1 and/or CG12236-PB sites was significantly abated ( $P < 0.002$ ) when they were mutated in combination with mutation of a Schlank site (Fig. 5D, Additional file 1: Fig. S26). This result is consistent with Schlank acting primarily as a transcriptional repressor in



the mesoderm, as we observed above for the CG7759\_33 enhancer. Unlike other mutations for which we generated imaging data, the mutation of a Schlank site in this context appears to have somewhat cell type-specific effects, rescuing expression

in only a small subset of the cells in which the wild type enhancer is active (Additional file 1: Fig. S20).

In addition to pairwise epistatic effects, we also observed higher-order interactions (Additional file 4: Table S4), such as in the case of the CBP2862 mutation series. For example, we observed an alleviating epistatic interaction ( $P = 0.017$ ) between Schlank and Deaf1 site mutations. Even though mutation of Schlank sites on their own did not increase activity of the enhancer CBP2862, Schlank site mutations partially rescued ( $P = 0.0062$ ) the diminished enhancer activity ( $P = 0.0001$ ) that arose from Deaf1 site mutations in the CBP2862 enhancer (Fig. 5E, Additional file 1: Fig. S28). In contrast, when the single ZIPIC site was also mutated in this enhancer, mutation of Schlank sites decreased enhancer activity (Fig. 5F, Additional file 1: Fig. S28). The Schlank+ZIPIC double site knockout element drove significantly lower expression than did the ZIPIC alone knockout ( $P = 0.029$ ), and the Deaf1+Schlank+ZIPIC triple site knockout also showed significantly reduced activity relative to the Deaf1+ZIPIC double knockout ( $P = 0.0169$ ). These results demonstrate an example of “sign epistasis,” where the effect of mutating a motif on its own was in the opposite direction of the effect when it was mutated together with another motif, here suggesting that in the context of ZIPIC site mutations, Schlank can act as an activator. Such context-dependent activity of Schlank may be due to altered protein-protein interactions induced by the presence of ZIPIC. Context-dependent activity of a TF as either an activator or a repressor has been observed previously in *Drosophila* [45].

Finally, we also observed sign epistasis between Deaf1 and ZIPIC motifs. Mutation of ZIPIC sites alone in the CG7759\_33 enhancer resulted in increased expression in non-FCM cells ( $P = 0.0005$ ), while mutation of Deaf1 sites in CG7759\_33 did not exert a statistically significant effect. Mutating Deaf1 and ZIPIC sites in combination, however, resulted in strongly reduced enhancer activity ( $P < 0.0001$ ) (Fig. 5G, Additional file 1: Fig. S32).

## Discussion

In this study, we developed and applied QeFS to assay the effects of multiple TF binding site mutations on enhancer activity in the developing mesoderm in *Drosophila* embryos. QeFS provides: (1) a direct, quantitative read-out of enhancer activity by sequencing reporter transcripts; (2) a barcoding and tagging strategy that enables discrimination of reporter transcripts driven by very similar (> 99% sequence similarity) enhancer elements; (3) multiple distinct reporter constructs per enhancer element, enabling analysis of variance-based comparisons of activity driven by enhancers within a given mutation series; and (4) the use of UMI tagging of reporter constructs, reducing biases introduced during library preparation steps. As with the prior eFS technology [1], QeFS allows for parallel assays of the activities of hundreds of genomically integrated tissue-specific enhancers in the context of a whole organism. It thus provides a compromise approach between the extreme throughput of MPRA carried out with episomal reporters in cultured cells and the exquisitely detailed information available from conventional reporter assays, for which informative transgenic lines must be made individually.

Our enhancer barcoding approach enabled the analysis of pools of enhancers that varied at only a few key nucleotide positions, and our degenerate reporter tagging

strategy allowed us to use a non-parametric analysis-of-variance statistical approach to distinguish the activity level driven by different enhancers in the same cell type and developmental stage. This approach could be applied to any cell type in *Drosophila* which can be collected in sufficient numbers and for which a well-defined, cell and developmental stage-specific enhancer is available to drive CD2 expression. Additionally, this approach could be extended to other model organisms for which tools for genomic integration are available [46, 47].

The quantitative nature of the QeFS data is the key advance over the prior eFS technology. QeFS allows the identification of not just varying degrees of decreased enhancer activity, but also increased enhancer activity. The library of enhancer elements that we analyzed in this study was specifically designed to assess the contributions of the DNA binding site motifs of four TFs—Deaf1, Schlank, ZIPIC, and CG12236-PB—to the activities of 12 mesodermally active enhancers. Notably, all 4 motifs that we investigated in this study are those of poorly characterized regulators. Most TFs that have strong effects have in all likelihood already been identified from earlier genetic studies and screens in the literature; many poorly characterized TFs remain, and they likely are regulators that exert effects that are subtle and/or highly context-dependent in a manner that depends on the particular enhancer sequence and/or the particular tissue or cell type in which activity is being assessed. Examining the effects of these motifs in combination and in the context of several different enhancer sequences (all assayed at a common genomic location) in the *Drosophila* mesoderm permitted the detection of such subtle effects.

Our results provide insights into the roles of Deaf1, Schlank, ZIPIC, and CG12236-PB DNA binding site motifs in mesodermal enhancer activity. The reductions in enhancer activity that we observed upon mutation of Deaf1 or CG12236-PB binding site motifs suggest that they act as activators in mesoderm. These results are consistent with a prior study that found decreased reporter activity in the *Drosophila* SL2 cell line from a Deaf1-dependent promoter that was mutated for a Deaf1 binding site [42]. In contrast, the increases in enhancer activity that we observed upon mutation of Schlank or ZIPIC binding sites suggest that they act as repressors in mesoderm. Our results support a previous study that found Schlank to act as a transcriptional repressor in *Drosophila* S2 cells and larvae, where it regulates lipid metabolism and growth [36]. Notably, among the enhancers that we found to be sensitive to mutation of Deaf1 binding sites, are as follows: CBP2862, located ~ 150 bp upstream of *four wheel drive* (*fwd*), which encodes a Golgi-localized lipid kinase that is expressed in mesoderm, and ChIPCRM2847, located in an intron of the insulin receptor, which is also expressed in mesoderm, and which controls the growth of *Drosophila* larval skeletal muscles [48]. Both of these Deaf1 motif effects are sensitive to the presence or absence of Schlank motifs (Fig. 5 and Additional file 4: Table S4). Altogether, these results suggest a role for Schlank as a transcriptional regulator of fat metabolism and growth in the developing embryo. One prior study reported ZIPIC to act as an insulator protein [33], while another found it to support long-range genomic interactions in *Drosophila* [34]. Our results suggest that the zinc finger protein ZIPIC may also act as a transcriptional repressor and/or modulate the activity of nearby TFs. Such multifunctionality in transcriptional regulation by a DNA-binding zinc finger protein has been observed before for CTCF [49]. Finally, although we previously found the CG12236-PB DNA binding

motif to be enriched among mesodermal enhancers [1], to our knowledge, our QeFS results are the first to demonstrate that the CG12236-PB motif regulates enhancer activity and suggest that CG12236-PB may act as a previously unknown transcriptional regulator in *Drosophila* embryonic mesoderm. In all cases, the effects of mutating binding sites of a TF were highly context-dependent, in that effects varied across different wild type enhancers and when different combinations of additional motifs were also mutated.

The large number of mutations that showed no detectable effect on enhancer activity is surprising in light of a recent report of densely encoded regulatory information in a *Drosophila* enhancer [50]. The *shavenbaby E3N* enhancer chosen for that study was selected for its small size and the large amount of patterning information that it integrates, and thus may not be similar to the mesodermal enhancers we mutated in larger fragments. Alternatively, the motifs that we chose to mutate, for relatively understudied factors with ubiquitous expression, might be atypical in their tolerance of mutation. Future experiments could investigate how the observed contributions of these *cis*-regulatory motifs might differ among more specific subpopulations of mesodermal cells.

Our results across these 12 developmental enhancers active in embryonic mesoderm development indicate that they integrate a combination of positive and negative regulatory inputs from Deaf1, Schlank, ZIPIC, and CG12236-PB. Furthermore, the use of QeFS in combination with synthesis of a specifically designed set of mutant enhancers allowed us to screen for epistatic interactions between *cis*-regulatory motifs in a highly parallel manner in a developing organism (Fig. 6). Although prior studies have systematically analyzed sequence features of enhancers that are important for their activities [21, 25], to our knowledge, our study is the first to assay the effects of combinatorial motif mutations using genomically integrated enhancers in a whole organism. Notably, a prior study found that the chromatin accessibility profiles, as determined by ATAC-Seq, of genomically integrated reporters are the same as those of the endogenous enhancers [52], suggesting that this strategy is likely to measure activity in the appropriate chromatin context. Here, we observed apparent synergy between motifs, finding in one case that Deaf1 and CG12236-PB pairwise mutations appear to potentiate each other's individual mutations, leading to an additional loss of activity at ChIPCRM5432. Interestingly, we also found that Schlank site mutations within ChIPCRM2847 suppress the effects of Deaf1 and/or CG12236-PB mutations.

Our results support prior studies that have demonstrated synergy between TF binding motifs [19], such as long-distance cooperative interactions between the Dorsal and Twist activators' motifs in the *rhomboid* enhancer in *Drosophila* embryos [53]. Such synergy between TF motifs was also observed in an imaging-based analysis of *eve* enhancer activity in a subset of muscle and cardiac cells in developing *Drosophila* embryos [54], where combinatorial mutations between dTCF and Twi1 sites or dTCF and ETS3 sites led to a decrease in the number of cells with reporter activity. This is a possible phenomenon underlying some of our observations of synergistic activity, although by QeFS alone we cannot distinguish between altered levels of reporter activity across all cells and altered numbers of cells exhibiting reporter activity. As the GFP reporter transcript we use here expresses a functional protein, complementary imaging studies using our backbone can be performed to test these two different hypotheses.

Our finding of sign epistasis between Deaf1 site mutations and ZIPIC site mutations in the CG7759-33 enhancer, where we found that ZIPIC putatively acts as a repressor, is consistent with a prior study of point mutations in the canonical lambda bacteriophage promoter  $P_R$  that affect binding by RNA Polymerase and the TF cI, where thermodynamic modeling of that system indicated that sign epistasis cannot exist in the absence of a repressor [55]. Sign epistasis may be an outcome of combinatorial regulation by multiple TFs in a cell-type-specific or condition-dependent manner, in which the regulatory output of a TF is modulated by its co-regulatory factors. For example, in the scenario of a TF that accelerates a rate-limiting step in the transcription cycle and inhibits another step, a second TF that also activates the first step to the degree that the second step now becomes rate-limiting, may result in the overall output of the first TF being switched from activation to repression [56].

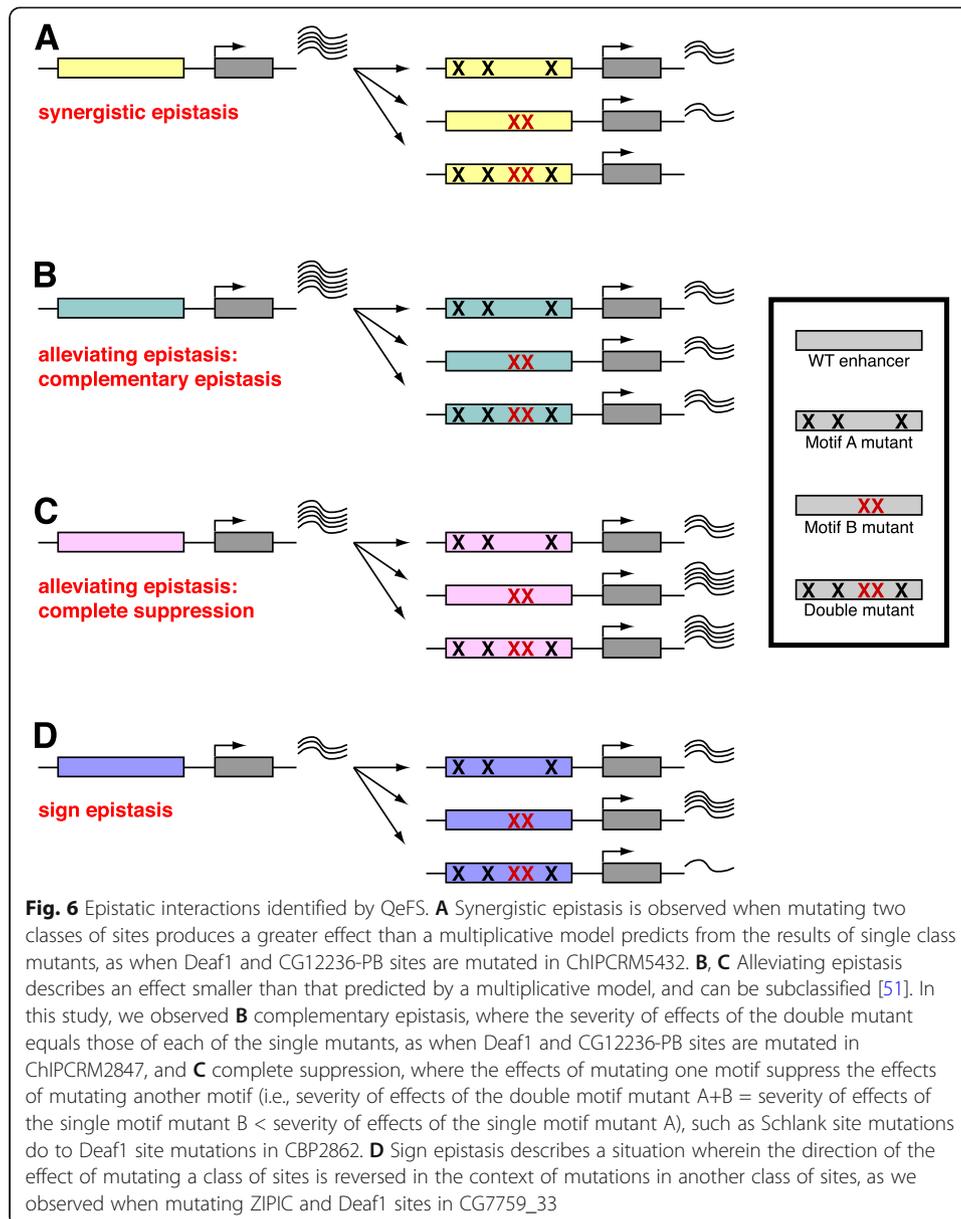
## Conclusions

Epistatic interactions between TF binding sites show extreme context dependence, so that even the direction of epistasis observed between the same motifs can be different in the context of different enhancers (Fig. 3). Uncovering the logic of combinatorial regulation will thus require sensitive, quantitative methods for parallel assay of large numbers of sequences. Analysis of a broader set of motifs and enhancers may reveal the sequence context features that influence when particular combinations of motifs act synergistically. Here, we focused on investigating four TF binding site motifs in a set of 88 wildtype and mutant mesodermal enhancers. The use of degenerate sequence tags and enhancer barcodes enables QeFS to be scaled up in future studies to assay mutational effects in hundreds of enhancers. QeFS could be applied to investigate the effects of SNPs within enhancers on enhancer activity in living organisms. Application of QeFS to other cell types and species should aid in elucidating the *cis*-regulatory codes, including context-dependent epistatic interactions among *cis*-regulatory motifs, underlying cell-type-specific enhancer activity.

## Methods

### Design and synthesis of barcoded enhancer constructs for QeFS

We designed each version (WT, 3 single-motif mutants, 3 motif-pair mutants, triple motif mutant) of each enhancer to be flanked by a unique 8-mer barcode. Briefly, barcodes were chosen to reduce the introduction of potential functional DNA binding sites adjacent to the enhancer sequence and to enable distinguishing of similar enhancer sequences despite potential mutations in the sequence during cloning or errors during lookup table library sequencing. All 8-mer sequences were first filtered against a set of fly TF protein binding microarray (PBM) data [1, 57], and any 8-mer with a PBM E-score  $> 0.3$  from any PBM experiment was eliminated as a candidate barcode. A similar filter was applied based on empirical thresholding against a set of 86 representative position weight matrices (PWMs) [1]. The remaining barcodes and their reverse complements were filtered to achieve a minimum Hamming distance of 3 between any two remaining barcodes. Filtering with these criteria yielded 101 high-stringency, 8-mer barcodes, 100 of which we used in the construction of our experimental and control libraries.



Selected enhancers were scanned with the *Deaf1*, *ZIPIC*, *Schlank*, and CG12236-PB PWMs [1]. Matches with a log-likelihood ratio, calculated relative to the background nucleotide distribution, greater than 6 were considered further. For each motif's sites, occupancy scores were calculated, with an occupancy of 1 corresponding to the optimal site. *Deaf1* and CG12236-PB have compact motifs with little degeneracy, making motif matches unambiguous. *Schlank* sites with predicted occupancy scores less than 0.5 and *ZIPIC* sites with predicted occupancy scores less than 0.05 were eliminated from further consideration. Mutations to selected sites were designed manually to maximally disrupt predicted binding affinity according to PWM score while avoiding creation or disruption of overlapping sites for TFs anticipated to be relevant to mesoderm development. In most cases (118 out of 135 motif instances), a single nucleotide change was sufficient to disrupt predicted binding, while the largest change involved mutation of

just 3 nucleotide positions. Mutant enhancers differed from wildtype by a minimum of 1 nucleotide and a maximum of 19.

Wildtype and mutant enhancers were synthesized with flanking barcodes and Gateway Clonase recognition sequences with the following structure for cloning into our reporter constructs: Gateway attB1 site, primer binding site (SQ3), enhancer barcode, Bmt I restriction enzyme site, enhancer element (wildtype or mutant), reverse complement of the same barcode, primer binding site (rc(SQ5)), and the reverse complement of the Gateway attB2 site. All but one (the *Deaf1* and *ZIPIC* double mutant allele for the enhancer ChIPCRM2000) of the resulting 96 sequences were successfully synthesized and cloned into the pUC19 vector (Gen9). The *Act57B* and  $\beta$ 3Tub60D control enhancers described in the text were prepared by PCR from OreR genomic DNA followed by site-directed mutagenesis, with cloning sites and barcodes added as above. Sequences of all tested constructs are provided in Additional file 3: Table S3.

### Construction of enhancer-reporter library

As the original enhancer-FACS-seq approach was based upon a DNA-seq readout of abundance of the enhancer element, but QeFS is based upon paired RNA- and DNA-seq readout of reporter transcripts and templates through targeted sequencing of degenerate tags in the 3' UTR of the reporter element, the original pEFS reporter backbone [1] needed to be modified in several ways. Ligation-free mutagenesis was used to introduce two sets of primer binding sites into the 3' UTR of the reporter gene template. These primers facilitate later construction of the lookup table sequencing library and the RNA- and DNA-seq libraries. Additionally, conventional restriction digest cloning was used to introduce a library of degenerate 20-mer sequences into the 3' UTR of the reporter template in a parallel fashion, thus creating a library of reporter backbones, each with a unique 20-mer tag.

Each enhancer construct was cloned independently into the pQeFS reporter backbone and individually validated by Sanger sequencing before being pooled at equimolar concentrations, creating a final enhancer-reporter library of ~320 unique enhancer-reporter constructs suitable for injections. The enhancer-reporter library was constructed such that there are multiple reporter tags per enhancer, but with each tag linked to a unique enhancer. We determined the uniformity of enhancer representation in this pool and generated a lookup table that associates enhancer barcodes with reporter tags, by a custom high-throughput sequencing approach.

### Generation and sequencing of lookup table

To generate the lookup table, a portion of the library pool was first digested with BmtI, which cleaves (1) in the 3' UTR of the reporter transcript, directly adjacent to the degenerate 20-mer tag, and (2) directly adjacent to the enhancer barcode that is most distal to the reporter promoter, and then re-circularized by intramolecular ligation. The re-circularized product juxtaposes the enhancer barcode in close proximity to the degenerate 20-mer. The ligation product was purified by a Zymo DNA Clean and Concentrator column. PCR was then used to amplify the product of interest and label the product with adapters (Illumina P5 and P7, a custom sequencing primer, and the

Illumina index read 1 primer) necessary for Illumina sequencing. Following amplification, the library was size-selected with SPRI beads at a ratio of 0.8x beads:library products.

A custom sequencing primer (“LUT\_sequencing”) was designed and used to sequence the lookup table library on an Illumina MiSeq instrument with 70 cycles of read 1 sequencing and a 6 cycle index 1 read. A total of 35,247,345 raw reads were sequenced, and 33,876,153 Passed Filter reads were identified, a rate of 96%. A total of 33,322,759 reads had the expected Illumina index.

Enhancer barcode-tag pairs obtained by lookup table sequencing were compared to expected barcode-tag pairs obtained by Sanger sequencing during library cloning to define true positive pairs. In seven cases, additional tags were defined as positive following empiric thresholding of barcode-tag sequencing read counts for individual enhancer barcodes. This final list of barcode-tag pairs was used to build a pseudo-genome for mapping QeFS sequencing reads.

### FACS and preparation of QeFS sequencing libraries

The enhancer-reporter library was injected into flies (Rainbow Transgenic Flies) from the strain *w- nos-ΦC31; attP40*. Progeny of injected flies, each of which contained a single genomically integrated enhancer-reporter, were crossed to CD2 driver lines expressing rat CD2 under the control of a cell type-specific driver (the *twi* promoter for whole mesoderm, or the *Mef2-I-E<sub>D5</sub>* enhancer and the *Hsp70* TATA box for fusion-competent myoblasts (referred to as *mef2*)). Embryos from these crosses were harvested, aged to the desired stage, and processed to prepare single-cell suspensions suitable for FACS. We performed 2 weeks of collections of both CD2+ and CD2- populations, with three collections per week, for *twi*:CD2 crosses and for *mef2*:CD2 crosses (Additional file 2: Table S2).

We used a custom high-throughput sequencing library preparation method to generate two libraries (paired RNA- and DNA-seq) for both CD2+ and CD2- collections for each day. For the QeFS collections, each week’s libraries were indexed, pooled, and sequenced on an Illumina NextSeq instrument, yielding ~ 5 million to ~ 25 million reads per library.

### Processing of sequencing reads

Processing of sequencing data began with recovery of the UMI sequence from each Illumina sequencing read. We wrote a custom Python script to process individual fastq file entries, parsing the first 8 bases off the beginning of each read and appending this sequence to the read name. The utility cutadapt was then used to trim off the ~ 20-nt constant region from the new 5’ end of each read (example command: cutadapt -g CGCGGGATGCTAGCACGCGG -e 0.3 -o CTW30OutFive.fq.gz --format fastq CTW30.out.gz). Cutadapt was then used to trim the constant 3’ end of each read off as well, leaving the ~ 20-nt sequence that should be unique to each reporter element (example command: cutadapt -a ACATATAGGACCAG -e 0.3 -m 10 -o CTW30OutFive-Three.fq CTW30OutFive.fq.gz). In preparation for mapping the degenerate reporter tags back to their corresponding enhancer, the bowtie-build command was used to build a custom “genome” index of the barcode-tag pairs obtained from the lookup

table, where each pseudo-chromosome was a particular enhancer, and each pseudo-chromosome sequence was the degenerate tag sequence. Thus, the pre-processed FASTQ file containing the trimmed degenerate tags could be aligned to this custom index with bowtie allowing 2 mismatches along the ~ 20-nt sequence. The output was then parsed using a second custom Python script to generate a count of per-tag UMI reads, which are the final read-out of enhancer activity.

### Analysis of QeFS sequencing data

The sequencing library preparation incorporated a UMI label for each cDNA or DNA template that enabled quantification of starting RNA or DNA abundance, respectively, to minimize effects of potential amplification bias. Sequencing libraries were processed to obtain a final per-reporter UMI count for each tag in both the DNA- and RNA-seq libraries. Each UMI count was scaled by the total number of processed tags in a given sequencing library, thereby scaling for overall read depth of the library. To scale each RNA count (corresponding to enhancer activity) to the abundance of a given enhancer in the underlying fly population, per-tag UMI totals for RNA-seq libraries were scaled by corresponding per-tag UMI totals for paired DNA-seq libraries from the same collection, scaling the reporter signal to the underlying abundance of a given reporter in the fly population (Eq1).

$$Expression_{tag_A} = \frac{\frac{UMI_{RNA,tag_A}}{\sum UMI_{RNA,tag_n}}}{\frac{UMI_{DNA,tag_A}}{\sum UMI_{DNA,tag_n}}} \quad (1)$$

Tags that did not meet a minimum empirical threshold of 65 DNA read counts per tag per day were removed; this type of thresholding is consistent with other quality control thresholds used in a prior MPRA study [24]. In addition, some enhancers were represented by only a small number of tags, either per day or across all three days of collections for a given week. A minimum of at least 4 tags per enhancer per week of aggregated data was required for inclusion of that enhancer construct in the final dataset. Overall, UMI read counts and DNA- or RNA-seq read counts correlated well (e.g., Spearman's  $\rho = 0.997$  and Pearson's  $r = 0.997$  for DNA and Spearman's  $\rho = 0.999$  and Pearson's  $r = 0.951$  for RNA), indicating that there was minimal amplification bias during library preparation (Additional file 1: Figure S12).

### Statistical analysis

Data were aggregated by treating all observations for a given enhancer in a given week as independent biological replicates. Thus, if an enhancer was represented by three different tags on three consecutive days, there would be nine total observations for this enhancer. Differences in expression within libraries due to site mutations were determined through nonparametric one-way analysis of variance testing using the Kruskal-Wallis (KW) test. This test was applied across each series of enhancer mutant constructs (WT and individual and combinatorial mutants) following aggregation within each week to determine if there were significant differences in expression across the series. Post-hoc pairwise testing within an allelic series was performed using the Conover-Iman test [58], which is valid if and only if the null hypothesis of no difference

between groups is rejected following the KW test, and preserves the ranks used in the KW omnibus test. Benjamini-Hochberg multiple comparison adjustment was applied to calculated  $P$ -values, and an alpha value of 0.1 was used for calling significantly different expression levels. All results from this test are provided in Additional file 4: Table S4. All  $P$ -values reported for QeFS pairwise comparisons between constructs were obtained by this method.

For display purposes, mean enhancer activity was calculated for each enhancer over all replicate measurements, and normalized to “Fold Change” by dividing by the activity of the corresponding wildtype enhancer. Standard error of the mean was also calculated, and a propagated standard error of the mean was calculated for the Fold Change measurement as described previously [24].

To test for epistasis, here understood as statistically significant deviation from the prediction of a multiplicative model [44] for combining the effects of separately mutating individual classes of sites, all replicate expression measurements for WT, motif A mutant, motif B mutant, and motifs A+B mutant constructs were log-transformed and modeled with the `lm` function in R. All  $P$ -values reported for such effects represent the  $P$ -values associated with the interaction term.

#### Additional method details

For additional methods, please see Additional file 2: Supplemental Methods [59].

#### Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02574-x>.

**Additional file 1:** Supplementary Figures. **Figures S1–S32.**

**Additional file 2:** Supplemental Methods and **Tables S1** and **S2.**

**Additional file 3:** **Table S3.** Sequences of tested elements.

**Additional file 4:** **Table S4.** QeFS processed data.

**Additional file 5:** Peer review history.

#### Acknowledgements

We thank Yi-Dong Lin for technical assistance, Raehoon Jeong and Luis Barrera for computational assistance, and Kaia Mattioli and Kian Hong Kock for critical reading of the manuscript.

#### Peer review information

Anahita Bishop and Tim Sands were the editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

#### Review history

The review history is available as Additional file 5.

#### Authors' contributions

C.T.W., S.S.G., Y.A.S., D.E.H., and M.L.B. designed the research; C.T.W., S.S.G., and T.M.C. performed the experiments; C.T.W. and S.S.G. analyzed the data; C.T.W., S.S.G., and M.L.B. wrote the paper. The authors read and approved the final manuscript.

#### Funding

This project was supported in part by a U.S. National Science Foundation Graduate Research Fellowship to C.T.W. and by grants R01 HG005287 and R01 HG009723 from the U.S. National Institutes of Health to M.L.B.

#### Availability of data and materials

Sequences of the tested enhancers described in this study are available in Additional file 3: Table S3. The normalized expression measurements and results of all pairwise statistical tests appear in Additional file 4: Table S4. The raw sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus under accession number GSE149908 [60].

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

### Author details

<sup>1</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA 02138, USA. <sup>3</sup>Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. <sup>4</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. <sup>5</sup>Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA. <sup>6</sup>Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.

Received: 30 October 2020 Accepted: 10 December 2021

Published online: 20 December 2021

## References

- Gisselbrecht SS, Barrera LA, Porsch M, Aboukhalil A, Estep Ilii PW, Vedenko A, et al. Highly parallel assays of tissue-specific enhancers in whole *Drosophila* embryos. *Nat Methods*. 2013;10(8):774–80. <https://doi.org/10.1038/nmeth.2558>.
- Stathopoulos A, Van Drenth M, Erives A, Markstein M, Levine M. Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell*. 2002;111(5):687–701. [https://doi.org/10.1016/S0092-8674\(02\)01087-5](https://doi.org/10.1016/S0092-8674(02)01087-5).
- Bodner M, Karin M. A pituitary-specific *trans*-acting factor can stimulate transcription from the growth hormone promoter in extracts of nonexpressing cells. *Cell*. 1987;50(2):267–75. [https://doi.org/10.1016/0092-8674\(87\)90222-4](https://doi.org/10.1016/0092-8674(87)90222-4).
- Lichtsteiner S, Schibler U. A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. *Cell*. 1989;57(7):1179–87. [https://doi.org/10.1016/0092-8674\(89\)90055-X](https://doi.org/10.1016/0092-8674(89)90055-X).
- Herrera R, Ro HS, Robinson GS, Xanthopoulos KG, Spiegelman BM. A direct role for C/EBP and the AP-1-binding site in gene expression linked to adipocyte differentiation. *Mol Cell Biol*. 1989;9(12):5331–9. <https://doi.org/10.1128/MCB.9.12.5331>.
- Rhodes SJ, Chen R, DiMattia GE, Scully KM, Kalla KA, Lin SC, et al. A tissue-specific enhancer confers Pit-1-dependent morphogen inducibility and autoregulation on the pit-1 gene. *Genes Dev*. 1993;7(6):913–32. <https://doi.org/10.1101/gad.7.6.913>.
- Lehming N, Thanos D, Brickman JM, Ma J, Maniatis T, Ptashne M. An HMG-like protein that can switch a transcriptional activator to a repressor. *Nature*. 1994;371(6493):175–9. <https://doi.org/10.1038/371175a0>.
- Pan D, Courey AJ. The same dorsal binding site mediates both activation and repression in a context-dependent manner. *EMBO J*. 1992;11(5):1837–42. <https://doi.org/10.1002/j.1460-2075.1992.tb05235.x>.
- Gordân R, Shen N, Dror I, Zhou T, Horton J, Rohs R, et al. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep*. 2013;3(4):1093–104. <https://doi.org/10.1016/j.celrep.2013.03.014>.
- Fromental C, Kanno M, Nomiya H, Chambon P. Cooperativity and hierarchical levels of functional organization in the SV40 enhancer. *Cell*. 1988;54(7):943–53. [https://doi.org/10.1016/0092-8674\(88\)90109-2](https://doi.org/10.1016/0092-8674(88)90109-2).
- Siggers T, Duyzend MH, Reddy J, Khan S, Bulyk ML. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol*. 2011;7(1):555. <https://doi.org/10.1038/msb.2011.89>.
- Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, et al. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*. 2011;147(6):1270–82. <https://doi.org/10.1016/j.cell.2011.10.053>.
- Polach KJ, Widom J. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J Mol Biol*. 1996;258(5):800–12. <https://doi.org/10.1006/jmbi.1996.0288>.
- Buck MJ, Lieb JD. A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nat Genet*. 2006;38(12):1446–51. <https://doi.org/10.1038/ng1917>.
- Mirny LA. Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci*. 2010;107(52):22534–9. <https://doi.org/10.1073/pnas.0913805107>.
- Han K, Levine MS, Manley JL. Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. *Cell*. 1989;56(4):573–83. [https://doi.org/10.1016/0092-8674\(89\)90580-1](https://doi.org/10.1016/0092-8674(89)90580-1).
- Wasson T, Hartemink AJ. An ensemble model of competitive multi-factor binding of the genome. *Genome Res*. 2009;19(11):2101–12. <https://doi.org/10.1101/gr.093450.109>.
- Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol*. 2012;30(6):521–30. <https://doi.org/10.1038/nbt.2205>.
- Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, et al. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci*. 2017;114(7):E1291–300. <https://doi.org/10.1073/pnas.1621150114>.
- Gertz J, Siggia ED, Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*. 2009;457(7226):215–8. <https://doi.org/10.1038/nature07521>.
- Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30(3):271–7. <https://doi.org/10.1038/nbt.2137>.

22. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339(6123):1074–7. <https://doi.org/10.1126/science.1232542>.
23. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27(1):38–52. <https://doi.org/10.1101/gr.212092.116>.
24. Kwasniewski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc Natl Acad Sci*. 2012;109(47):19498–503. <https://doi.org/10.1073/pnas.1210678109>.
25. White MA, Myers CA, Corbo JC, Cohen BA. Massively parallel *in vivo* enhancer assay reveals that highly local features determine the *cis*-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci*. 2013;110(29):11952–7. <https://doi.org/10.1073/pnas.1307449110>.
26. Hughes AEO, Myers CA, Corbo JC. A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites *in vivo*. *Genome Res*. 2018;28(10):1520–31. <https://doi.org/10.1101/gr.231886.117>.
27. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat Biotechnol*. 2012;30(3):265–70. <https://doi.org/10.1038/nbt.2136>.
28. Nam J, Dong P, Tarpine R, Istrail S, Davidson EH. Functional *cis*-regulatory genomics for systems biology. *Proc Natl Acad Sci*. 2010;107(8):3930–5. <https://doi.org/10.1073/pnas.1000147107>.
29. Kwon EZ, Zhu Y, Kelman G, Novak CS, Plajzer-Frick I, Kato M, et al. Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. *Cell*. 2020;180:1262–1271.e1215.
30. Gross CT, McGinnis W. DEAF-1, a novel protein that binds an essential region in a Deformed response element. *EMBO J*. 1996;15(8):1961–70. <https://doi.org/10.1002/j.1460-2075.1996.tb00547.x>.
31. Veraksa A, Kennison J, McGinnis W. DEAF-1 function is essential for the early embryonic development of *Drosophila*. *Genesis*. 2002;33:67–76.
32. Czesak M, Lemonde S, Peterson EA, Rogaeva A, Albert PR. Cell-specific repressor or enhancer activities of Deaf-1 at a serotonin 1A receptor gene polymorphism. *J Neurosci*. 2006;26(6):1864–71. <https://doi.org/10.1523/JNEUROSCI.2643-05.2006>.
33. Maksimenko O, Bartkuhn M, Stakhov V, Herold M, Zolotarev N, Jox T, et al. Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res*. 2015;25(1):89–99. <https://doi.org/10.1101/gr.174169.114>.
34. Zolotarev N, Fedotova A, Kyrchanova O, Bonchuk A, Penin AA, Lando AS, et al. Architectural proteins Pita, Zw5, and ZIPIC contain homodimerization domain and support specific long-range interactions in *Drosophila*. *Nucleic Acids Res*. 2016;44(15):7228–41. <https://doi.org/10.1093/nar/gkw371>.
35. Voelzmann A, Wulf A-L, Eckardt F, Thielisch M, Brondolin M, Pesch Y-Y, et al. Nuclear *Drosophila* CerS Schlank regulates lipid homeostasis via the homeodomain, independent of the lag1p motif. *FEBS Lett*. 2016;590(7):971–81. <https://doi.org/10.1002/1873-3468.12125>.
36. Sociale M, Wulf A-L, Breiden B, Klee K, Thielisch M, Eckardt F, et al. Ceramide synthase Schlank is a transcriptional regulator adapting gene expression to energy requirements. *Cell Rep*. 2018;22(4):967–78. <https://doi.org/10.1016/j.celrep.2017.12.090>.
37. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enameh MS, Basciotta MD, et al. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res*. 2011;39(suppl\_1):D111–7. <https://doi.org/10.1093/nar/gkq858>.
38. Hammonds AS, Bristow CA, Fisher WW, Weiszmann R, Wu S, Hartenstein V, et al. Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biol*. 2013;14(12):R140. <https://doi.org/10.1186/gb-2013-14-12-r140>.
39. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, Lewis SE, et al. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol*. 2002;3:research0088.0081.
40. Kelly KK, Meadows SM, Cripps RM. *Drosophila* MEF2 is a direct regulator of Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. *Mech Dev*. 2002;110(1-2):39–50. [https://doi.org/10.1016/S0925-4773\(01\)00586-X](https://doi.org/10.1016/S0925-4773(01)00586-X).
41. Zaffran S, Frasch M. The  $\beta 3$  tubulin gene is a direct target of bagpipe and biniou in the visceral mesoderm of *Drosophila*. *Mech Dev*. 2002;114(1-2):85–93. [https://doi.org/10.1016/S0925-4773\(02\)00063-1](https://doi.org/10.1016/S0925-4773(02)00063-1).
42. Kutenkeuler D, Pelte N, Ragab A, Gesellchen V, Schneider L, Blass C, et al. A large-scale RNAi screen identifies *Deaf1* as a regulator of innate immune responses in *Drosophila*. *J Innate Immun*. 2010;2(2):181–94. <https://doi.org/10.1159/000248649>.
43. Phillips PC. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008;9(11):855–67. <https://doi.org/10.1038/nrg2452>.
44. Mani R, St. Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. *Proc Natl Acad Sci*. 2008;105(9):3461–6. <https://doi.org/10.1073/pnas.0712255105>.
45. Jiang J, Cai H, Zhou Q, Levine M. Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. *EMBO J*. 1993;12(8):3201–9. <https://doi.org/10.1002/j.1460-2075.1993.tb05989.x>.
46. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP. Site-specific genomic integration in mammalian cells mediated by phage  $\phi$ C31 integrase. *Mol Cell Biol*. 2001;21(12):3926–34. <https://doi.org/10.1128/MCB.21.12.3926-3934.2001>.
47. Tasic B, Hippenmeyer S, Wang C, Gamboa M, Zong H, Chen-Tsai Y, et al. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc Natl Acad Sci*. 2011;108(19):7902–7. <https://doi.org/10.1073/pnas.1019507108>.
48. Dementis F, Perrimon N. Integration of Insulin receptor/Foxo signaling and dMyc activity during muscle growth regulates body size in *Drosophila*. *Development*. 2009;136(6):983–93. <https://doi.org/10.1242/dev.027466>.
49. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137(7):1194–211. <https://doi.org/10.1016/j.cell.2009.06.001>.
50. Fuqua T, Jordan J, van Breugel ME, Halavatyi A, Tischer C, Polidoro P, et al. Dense and pleiotropic regulatory information in a developmental enhancer. *Nature*. 2020;587(7833):235–9. <https://doi.org/10.1038/s41586-020-2816-5>.
51. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet*. 2007;39(2):199–206. <https://doi.org/10.1038/ng1948>.
52. Jacobs J, Atkins M, Davie K, Imrichova H, Romanelli L, Christiaens V, et al. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat Genet*. 2018;50(7):1011–20. <https://doi.org/10.1038/s41588-018-0140-x>.

53. Sayal R, Dresch JM, Pushel I, Taylor BR, Arnosti DN. Quantitative perturbation-based analysis of gene expression predicts enhancer activity in early *Drosophila* embryo. *eLife*. 2016;5:e08445. <https://doi.org/10.7554/eLife.08445>.
54. Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jimenez F, Baylies MK, et al. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*. 2000;103(1):63–74. [https://doi.org/10.1016/S0092-8674\(00\)00105-7](https://doi.org/10.1016/S0092-8674(00)00105-7).
55. Lagator M, Paixão T, Barton NH, Bollback JP, Guet CC. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *eLife*. 2017;6:e25192. <https://doi.org/10.7554/eLife.25192>.
56. Scholes C, DePace AH, Sánchez Á. Combinatorial gene regulation through kinetic control of the transcription cycle. *Cell Syst*. 2017;4:97–108.e109.
57. Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, et al. A comprehensive *Drosophila melanogaster* transcription factor interactome. *Cell Rep*. 2019;27:955–970.e957.
58. Conover WJ, Iman RL. On multiple-comparison procedures. Technical Report LA-7677-MS, Los Alamos Scientific Laboratory. 1979. <https://doi.org/10.2172/6057803>.
59. Li C, Wen A, Shen B, Lu J, Huang Y, Chang Y. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol*. 2011;11(1):92. <https://doi.org/10.1186/1472-6750-11-92>.
60. Waters CT, Gisselbrecht SS, Sytnikova YA, Cafarelli TM, Hill DE, Bulyk ML. Quantitative-enhancer-FACS-seq (QeFS) reveals epistatic interactions among motifs within transcriptional enhancers in developing *Drosophila* tissue. *Series GSE149908*. *Gene Expr Omnibus* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149908>. 2021.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

