Supporting Information

Busser et al. 10.1073/pnas.1210415109

SI Materials and Methods

Fly Stocks. *Drosophila* stocks containing the following transgenes were used: attP2, attP40, nos-phiC31intNLS (1) (gift of N. Perrimon, Harvard University, Boston, MA), and *twi*-Gal4 UAS-2EGFP (2).

Analysis of Transgenic Reporter Constructs and Embryo Staining. Enhancer regions were either synthesized in vitro (Integrated DNA Technologies) or PCR-amplified, sequence verified, and subcloned into pWattB-GFP (3) or pWattB-nLacZ (4). Constructs were targeted to attP40 or attP2 (5) with phiC31-mediated integration (6). Whole embryo immunohistochemistry followed standard protocols (7). The following antibodies were used: rabbit anti-Lmd (1:1,500; gift of H. Nguyen, University of Erlargen, Erlangen, Germany), chicken anti-GFP (1:2,000; Abcam), and mouse anti-β-gal (1:500; Promega).

ChIP-seq. A single cell suspension was prepared on ice from late stage 11 through early stage 12 *twi*-Gal4 UAS-2EGFP embryos according to previously published procedures (8) and fixed on ice in 1.8% formaldehyde (wt/vol). GFP-positive cells were isolated using flow cytometry. Chromatin was prepared, fragmented (to 200–500 bp), and independently immunoprecipitated with two different antibodies directed against Lmd (9). The sequencing library was constructed according to previously published procedures (10), and raw sequencing data were obtained with an Illumina HiSeq2000 sequencer. See FACS-ChIP Protocol for more details. Lmd-bound regions were identified with MACS using all of the uniquely mapped sequence tags for the two independent Lmd antibody immunoprecipitations as the treatment and the input sequence tags as control, with P values set to 10^{-5} and mfold set to 2 (11). Only peaks identified in both immunoprecipitations were scored as positive, with the minimum overlap between regions set at 50 bp.

Cloning, Expression, and Protein Binding Microarray Analysis of Lmd. The DNA binding domain (DBD) of Lame duck (Lmd) comprising residues 404–621 was cloned into Gateway-compatible vectors, and proteins were produced by in vitro transcription and translation (IVT PURExpress; New England BioLabs). Protein binding microarray (PBM) assays were performed as previously described with 250 nM of Lmd DBD protein applied to custom-designed microarrays from Agilent Technologies (AMADID 015681) (3). Contiguous and gapped 8-mer enrichment scores and Lmd^P and Lmd^S motifs were generated using the PBM analysis suite (12, 13). Logos were generated from the obtained PWMs using enoLOGOS (14).

Classifier Training. Control sequences were randomly sampled from the noncoding regions of the *Drosophila melanogaster* genomic DNA sequence (dm3) and exhibited similar GC contents, repeat densities, and identical lengths as Lmd-bound sequences. Using each Lmd-bound sequence as a reference, we generated two control samples. Thus, in this study, we have $5{,}119 \times 2 = 10{,}238$ control sequences. Each of the considered sequences was represented as a vector representing the number of occurrences of transcription factor (TF) binding motifs. Binding motifs were mapped by scanning the sequences using tfSearch (15) with appropriate position weight matrices (PWMs). In total, there were 1,358 PWMs compiled from TRANSFAC and JASPAR databases (16, 17). We used a linear support vector machine (SVM) to discriminate Lmd-bound sequences from controls. Given a training set of instances $\{x_1, x_2, \ldots, x_n\}$ with associated labels $\{y_1, y_2, \ldots, y_n\} \in \{-1, 1\}$, a linear SVM $y = w^T x + b$ is built through solving the optimization problem $min\ \left(\frac{1}{2}w^T w + C\sum_i \varepsilon_i\right)$ subject to $y_i(w^T x_i + b) \geq 1 - \varepsilon_i$ and $\varepsilon_i \geq 0$ (18). In a built linear SVM, a TF binding motif corresponds to a linear weight w_i. A large positive w_i indicates a binding motif that is heavily associated with the training set, whereas negative weights mark binding motifs associated with the control set.

Clustering of Lmd Peaks with Other Mesodermal Regulators. To analyze the interaction between Lmd and other mesodermal regulators, such as Twi, Mef2, and Tin, we investigated the genomic regions cooccupied by Lmd and at least one of Twi, Mef2, and Tin (19). A region was regarded as cooccupied by Lmd and other TFs when the centers of their individual peaks were separated by 1,000 bp or less. According to the presence/absence of the binding of other TFs, we then clustered these cooccupied Lmd-bound sequences into distinctive groups. The clustering method K-means (20) and Euclidean distance were used. After testing different settings of K (the number of clusters), we observed that the setting $K = 5$ produces the optimum clustering results (i.e., the lowest within-cluster distance and the highest between-cluster distance). Each of the indicated clusters represents a unique combination of Lmd with other mesodermal regulators (Fig. 3).

Association of Lmd Peaks to Genes. We associated Lmd peaks to the closest genes according to the genomic locations of the centers of the Lmd peaks. More specifically, given an intronic Lmd peak for which the center is located within a particular gene, we associated that peak with its host gene. In contrast, for an Lmd peak in which the center falls within an intergenic region, we associated it with the gene closest to the center of the Lmd peak under consideration.

Histone Signatures of Lmd Peaks. We investigated the association of published histone modification signals [available at http://furlonglab.embl.de/data/download (21)] with Lmd ChIP-seq peaks determined in the present study. Given the set of Lmd-bound sequences and a corresponding set of control sequences, we aligned all considered sequences at their centers and averaged histone modification signals across these sequences at each position relative to the center (from 500 bp upstream to 500 bp downstream of these middle positions). We evaluated the confidence of our average estimation by using Monte Carlo simulation. Specifically, for a sequence set, we randomly selected one third of the considered sequences and calculated the average signals along the selected sequences. After repeating this process 1,000 times, we ranked the 1,000 averaged results that were obtained and retained the results at the 25th and 75th percentiles, which indicate the variation of averages and thus suggest the confidence of the average estimation.

Data Access. ChIP-seq data are available from the Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) under study accession no. GSE38402.

FACS-ChIP Protocol. *Cell sorting. Cages.*

i) Set up 30 400-mL cages of ∼1,200 yw; twi-gal4, UAS-eGFP flies and one cage of yellow white (as a negative control).
ii) Pre-lay 1 h before collection.
iii) Two-hour collection at 25 °C and 13 h at 18 °C for stage 11/12 embryos. Each 2-h collection will yield $\sim 5 \times 10^6$ GFP-expressing cells. We perform two 2-h collections per day and pool them after sorting before chromatin preparation.

Cell/chromatin preps.

i) Label douncers and 15-mL conical tubes and put on ice. Fill douncer with 7 mL Schneider's medium.
ii) Wash plates with embryo wash and collect in 70-μM cell strainers. Rinse with more embryo wash.
iii) Immerse strainers in 50% (vol/vol) bleach for 3–5 min while constantly streaming bleach over embryos using a pasteur pipet.
iv) Wash embryos with embryo wash and then H_2O.
v) Brush embryos into douncer. Try and keep it to five plates maximum per douncer.
vi) Smash embryos in douncer with seven strokes of the "loose" (type B) pestle. Put cells back on ice.
vii) Transfer dounced solution to a 15-mL conical tube (add two douncers worth of same cells per tube or fill with 7 mL more Schneider's).
viii) Centrifuge at 40 × *g* (4 °C, 5 min) to pellet the tissue debris, clumps, and vitelline membranes.
ix) Transfer supernatant to new 15-mL tube. Centrifuge at 380*g* (4 °C, 10 min) to spin down the cells.
x) Dump supernatant and resuspend pelleted cells in 1 mL Schneider's + 1.8% formaldehyde (wt/vol) to fix (15 min, on ice).
xi) Add 14 mL Schneider's medium to wash. Centifuge at 380*g* (4°C, 10 min). Repeat.
xii) Dump supernatant and resuspend pelleted cells in 1 mL Schneider's + 8% FBS (vol/vol), strain cell prep through a 40-μM cell strainer into a FACS tube, and bring to FACS facility.
xiii) Spin down sorted cells at 500 × *g* (20 min, 4 °C). Dump supernatant.
xiv) Resuspend cell pellets in 5 mL buffer A1 and put in 15-mL conical tube. Rock at 4 °C for 10 min. Fill tube with buffer A1 and spin (3,000 × *g*, 3 min, 4 °C).
xv) Wash with 10 mL buffer A1 and spin (3,000 × *g*, 3 min, 4 °C).
xvi) Resuspend sample in 0.3 mL buffer A2.
xvii) Sonicate using the Covaris sonicator (15 min).
xviii) Transfer sonicated sample to microfuge and spin for 10 min at high speed. Transfer supernatant to a fresh tube. Freeze chromatin at −80 °C.

ChIP. Day 1.

i) Add 0.1 g of Protein A beads (17-0780-01 1.5g; GE Healthcare) to every 1 mL of buffer A2 [plus 0.5% BSA (wt/vol)] to achieve a 50% by volume mixture of beads and buffer (wt/vol). Calculate amount of beads to be prepared (need ∼75 μL total per immunoprecipitation). Rotate (4 °C, 1 h).
ii) Pellet beads (3,380 × *g*, 4 °C, 2 min). Remove buffer A2 + BSA solution. Wash beads in ∼1 mL buffer A2, and then pellet beads (3,380 × *g*, 4 °C, 2 min), remove supernatant, and repeat two more times. When finished washing, resuspend in same volume of buffer A2 used in step 1. The beads are now "prepared."
iii) Add ∼25 μL of prepared beads to thawed chromatin (∼10 million cells worth) and rotate (4 °C, 1 h). Scale up amounts according to number of IPs to be done. This step preclears the chromatin of any nonspecific binding to beads. Pellet beads (3,380 × *g*, 4 °C, 2 min). Transfer supernatant to new tubes, adding ∼10 million cells worth per tube.
iv) Calculate amount of antibody to add (4 μg for affinity purified Abs, 1–3 μL for rabbit sera). Rotate (4 °C, overnight).

Day 2.

i) Add 50 μL prepared beads to antibody-chromatin sample and rotate (4 °C, 4 h).
ii) Wash IP with ∼1 mL RIPA wash buffer and rotate (4 °C, 5 min). Pellet beads (3,380 × *g*, 4 °C, 2 min) and remove supernatant. Repeat for a total of five RIPA washes. Wash twice with Tris-EDTA (TE) + 50 mM NaCl.
iii) Pellet beads (3,380 × *g*, 4 °C, 2 min) and remove supernatant.
iv) Add 100 μL elution buffer and incubate (10 min, 65 °C). Vortex periodically to keep beads in suspension. Pellet beads (18,400 × *g*, 1 min).
v) Transfer 100 μL supernatant into new tube with 100 μL TE.
vi) Add 100 μL elution buffer and 50 μL TE to 50 μL whole cell extract.
vii) Reverse cross-linking by incubation at 65 °C (overnight).

Day 3.

i) Add RNase A (to 0.05 μg/μL). Mix and incubate at 37 °C for 2 h.
ii) Add proteinase K (to 0.2 μg/μL). Mix and incubate at 55 °C for 2 h.
iii) Purify ChIP DNA and WCE DNA using ZYMO DNA clean and concentrator-5 kit. Elute in 30 μL H_2O.
iv) Ready for PCR or labeling.

Buffer A1	Buffer A2	Wash buffer (RIPA)	Elution buffer
15 mM Hepes, pH 7.5	15 mM Hepes, pH 7.5	50 mM Hepes, pH 7.5	50 mM Tris, pH 8
15 mM NaCl	140 mM NaCl	1 mM EDTA	10 mM 0.5 M EDTA
60 mM KCl	1 mM EDTA	0.7% Na deoxycholate (wt/vol)	1% SDS (vol/vol)
4 mM $MgCl_2$	0.5 mM EGTA	1% Nonidet P-40 (vol/vol)	
0.5% TritonX-100 (vol/vol)	1% Triton X-100 (vol/vol)	0.5 M LiCl	
0.5 mM DTT	0.1% sodium deoxycholate (wt/vol)		
Protease inhibitors	0.1% SDS (vol/vol)		
	0.5% *N*-lauroylsarcosine (wt/vol)		
	Protease inhibitors		

Reagents.

Reagents used were as follows: complete EDTA-free protease inhibitor mixture (Cat. no. 11873580001; Roche); RNase, DNase free (11119915001; Roche); and Proteinase K (03115879001; Roche). Add protease inhibitors (final concentration 1×) to all lysis buffers (A1 and A2) before use. (Dissolve one complete protease inhibitor mixture tablet in 1 mL H_2O to make a 50× solution. Store in aliquots at −20 °C.)

Day 4.

i) End-repair with END-IT kit from Epicentre.

dsDNA from previous step	30 μL
10× end-repair buffer (Epicentre)	5 μL
dNTP mix (Epicentre)	5 μL
ATP (Epicentre)	5 μL
End-IT enzyme mix (Epicentre)	1 μL
Nuclease-free H_2O	3 μL

ii) Incubate at room temperature for 45 min.
iii) Purify immediately with ZYMO DNA clean and concentrator-5 kit. Elute sample with 40 μL H_2O.
iv) A-tailing of DNA

DNA from previous step	40 μL
10× Klenow buffer (Epicentre)	5 μL
2.5 mM dATP (Bioline)	4 μL
Klenow (exo-) DNA polymerase (Epicentre)	1 μL

v) Incubate at 37 °C for 30 min.
vi) Purify with ZYMO DNA clean and concentrator-5 kit and elute with 6 μL H_2O.
vii) Ligation with PE Y-linker

DNA from previous step	7 μL
1 μM paired-end Y-linker (Illumina)	1 μL
10× T4 DNA ligase buffer (NEB)	1 μL
T4 DNA ligase, 2,000 U/μL (NEB)	1 μL

viii) Incubate at RT (25 °C) for 30 min.
xi) Purify the ligation reaction with the ZYMO kit and elute with 20 μL water. Gel purify a 250- to 450-bp piece by running on 2% agarose E-gel (Invitrogen), purifying with a ZYMO gel purification kit (elute with 20 μL H_2O).
x) PCR

DNA from previous step	20 μL
5× HF buffer (Finnzymes)	10 μL
10 μM PE_1.0 (IDT, HPLC)	2.5 μL
10 μM PE_2.0 (IDT, HPLC)	2.5 μL
10 mM dNTP	1 μL
Phusion DNA polymerase (Finnzymes)	0.5 μL
H_2O	13.5 μL

xi) Incubate at 98 °C for 30 s.
xii) Incubate for 16 cycles of 98 °C for 10 s, 67 °C for 10 s, and 72 °C for 30 s.
xiii) Incubate at 72 °C for 10 min. Hold at 4 °C.
xiv) Take 3 μL to run on a 2% agarose E-gel (Invitrogen). If a 300- to 500-bp smear can be seen, load 20 μL PCR product on 2% agarose E-gel (Invitrogen) and select the 300- to 500-bp band, purify with the ZYMO gel recovery kit (elute with 20 μL H_2O), and add 2 μL 1% Tween-20.
xv) If no band can be seen after 16 cycles, add 4 more cycles and check again.

1. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci USA* 104(9):3312–3317.
2. Halfon MS, et al. (2002) New fluorescent protein reporters for use with the *Drosophila* Gal4 expression system and for vital detection of balancer chromosomes. *Genesis* 34 (1-2):135–138.
3. Busser BW, et al. (2012) Molecular mechanism underlying the regulatory specificity of a *Drosophila* homeodomain protein that specifies myoblast identity. *Development* 139(6):1164–1174.
4. Busser BW, et al. (2012) A machine learning approach for identifying novel cell type-specific transcriptional regulators of myogenesis. *PLoS Genet* 8(3):e1002531.
5. Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40(4):476–483.
6. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* 166(4):1775–1782.
7. Halfon MS, et al. (2000) Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* 103(1):63–74.
8. Estrada B, et al. (2006) An integrated strategy for analyzing the unique developmental programs of different myoblast subtypes. *PLoS Genet* 2(2):e16.
9. Cunha PM, et al. (2010) Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. *PLoS Genet* 6(7):e1001014.
10. Nègre N, et al. (2011) A cis-regulatory map of the *Drosophila* genome. *Nature* 471 (7339):527–531.
11. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9): R137.
12. Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* 4(3): 393–411.
13. Berger MF, et al. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* 24(11):1429–1435.
14. Workman CT, et al. (2005) enoLOGOS: A versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* 33(Web Server issue):W389-92.
15. Ovcharenko I, et al. (2005) Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res* 15(1):184–194.
16. Wingender E, et al. (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res* 29(1):281–283.
17. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004) JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32(Database issue):suppl 1):D91–D94.
18. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297.
19. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM (2009) Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* 462(7269):65–70.
20. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Univ of California Press, Berkeley, CA), pp 281–297.
21. Bonn S, et al. (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44(2):148–156.

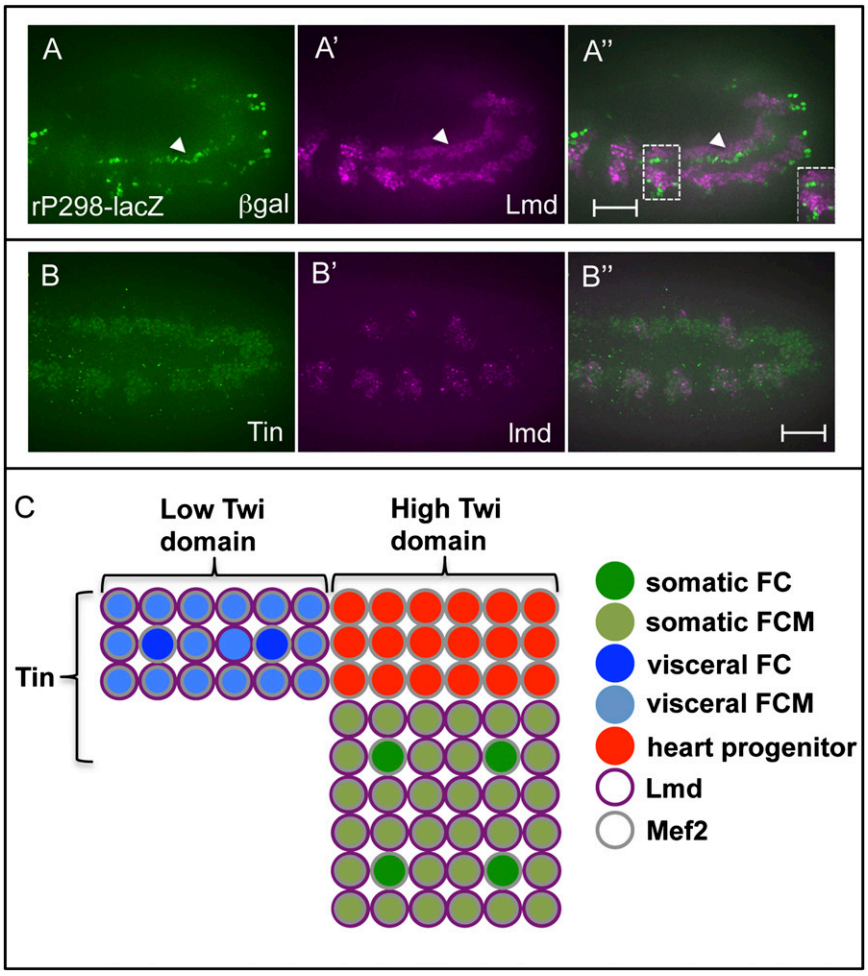

Fig. S1. Distribution of the founder cells (FCs) and fusion competent myoblasts (FCMs) of the somatic and visceral mesoderm. FCMs of the somatic and visceral mesoderm express Lmd. Costaining of stage 11 rP298-*lacZ* embryos for β-Gal (*A*) and Lmd (*A′*) proteins. *A″* is the merge of the color channels for the two different fluorescently labeled secondary antibodies that were used to detect the primary antibodies. rP298 marks FCs in the somatic and visceral mesoderm (1). Arrowhead denotes the visceral mesoderm, and the arrow marks a portion of the somatic mesoderm of the same hemisegment. Note that FC-expressed β-Gal does not coexpress with Lmd in either the somatic or visceral mesoderm, consistent with the unique localization of Lmd in FCMs. (*B*) Tin expression in the dorsal mesoderm coexpresses with *lmd*. Costaining of early stage 11 *yw* embryos for Tin (*B*) protein and *lmd* (*B′*) mRNA. *B″* is the merge of the two channels. (*C*) The somatic and visceral FCs and FCMs derive from unique progenitor pools, with Lmd being expressed in both populations of FCMs. All mesodermal cells derive from Twi-expressing cells, with cells of the visceral mesoderm deriving from a low Twi-expressing domain and the cells of the cardiac and somatic mesoderm deriving from a high Twi-expressing domain (2). At this developmental stage, Twi is expressed at high levels in the somatic and cardiac mesoderm and lower levels in the visceral mesoderm (3). Tin is expressed in the cardiac, visceral, and dorsal somatic mesoderm (4, 5). Mef2 is expressed in all cells of the depicted mesoderm. (Scale bar, 20 μm.)

1. Klapper R, et al. (2002) The formation of syncytia within the visceral musculature of the *Drosophila* midgut is dependent on *duf, sns* and *mbc*. *Mech Dev* 110(1-2):85–96.
2. Baylies MK, Bate M (1996) twist: A myogenic switch in *Drosophila*. *Science* 272(5267):1481–1484.
3. Borkowski OM, Brown NH, Bate M (1995) Anterior-posterior subdivision and the diversification of the mesoderm in Drosophila. *Development* 121(12):4183–4193.
4. Azpiazu N, Frasch M (1993) *tinman* and *bagpipe*: Two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev* 7(7B):1325–1340.
5. Yin Z, Frasch M (1998) Regulation and function of tinman during dorsal mesoderm induction and heart specification in Drosophila. *Dev Genet* 22(3):187–200.

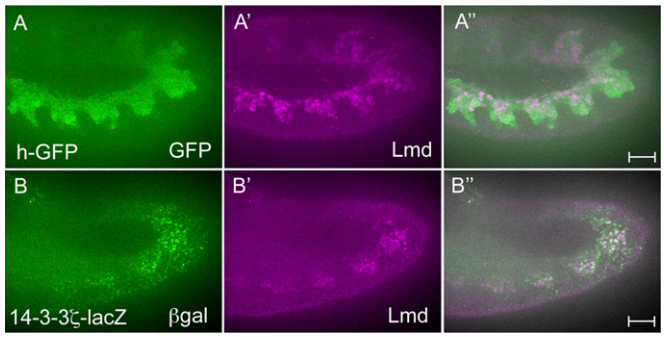

Fig. S2. Candidate enhancers bound by Twi and/or Mef2 are active in FCMs. Fluorescent antibody staining of stage 11 embryos containing FCM enhancer transgenes for *hairy* (h-GFP) (*A*) and 14–3-3ζ-lacZ (*B*). Reporters were detected using antibodies against GFP (*A*) or β-Gal(*B*) plus Lmd (*A′* and *B′*). *A″* and *B″* show the corresponding merged channels. (Scale bar, 20 μm.)

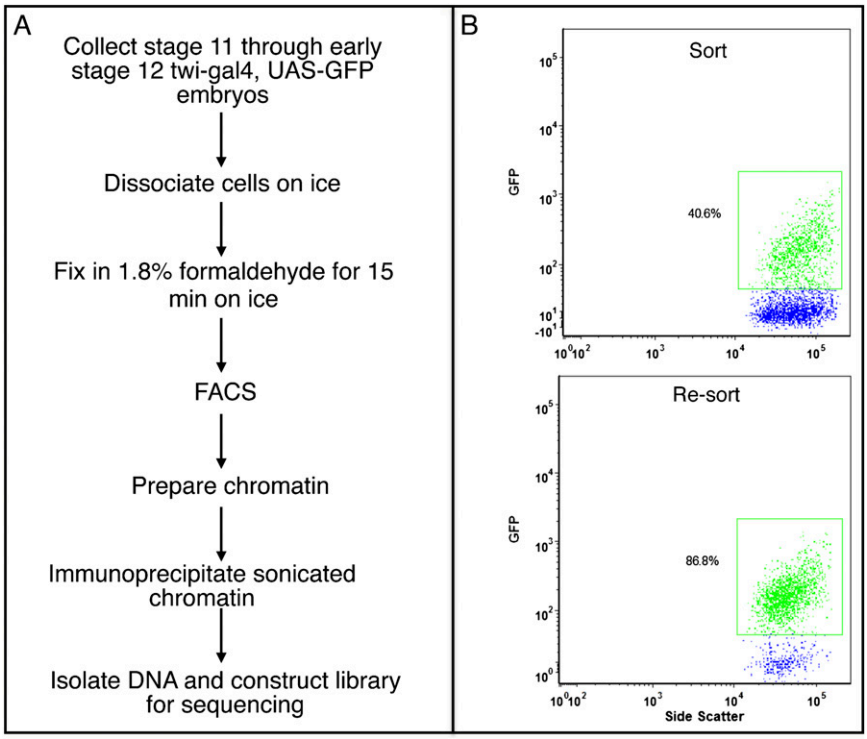

Fig. S3. Mesoderm-specific ChIP-seq. (*A*) Cells were dissociated from appropriately staged embryos (spanning a 2-h window when FCMs are specified but before extensive cell fusion with FCs) expressing a GFP transgene under *twi* promoter control to target the mesoderm, fixed in formaldehyde, and purified by FACS. Chromatin was prepared from the flow-sorted cells, immunoprecipitated with two distinct anti-Lmd antibodies, and subjected to massively parallel sequencing (1). Enriched sequences were identified using model-based analysis of ChIP-seq (MACS), and a region was only scored as bound if it was enriched using both antibodies (2). (*B*) Representative FACS plots showing side scatter on the *x*-axis and GFP fluorescence intensity on the *y*-axis for pre- and postsorted cells. Events shown in plots were first sorted for mononucleated cells based on forward and side scatter.

1. Cunha PM, et al. (2010) Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. *PLoS Genet* 6(7):e1001014.
2. Nègre N, et al. (2011) A cis-regulatory map of the *Drosophila* genome. *Nature* 471(7339):527–531.

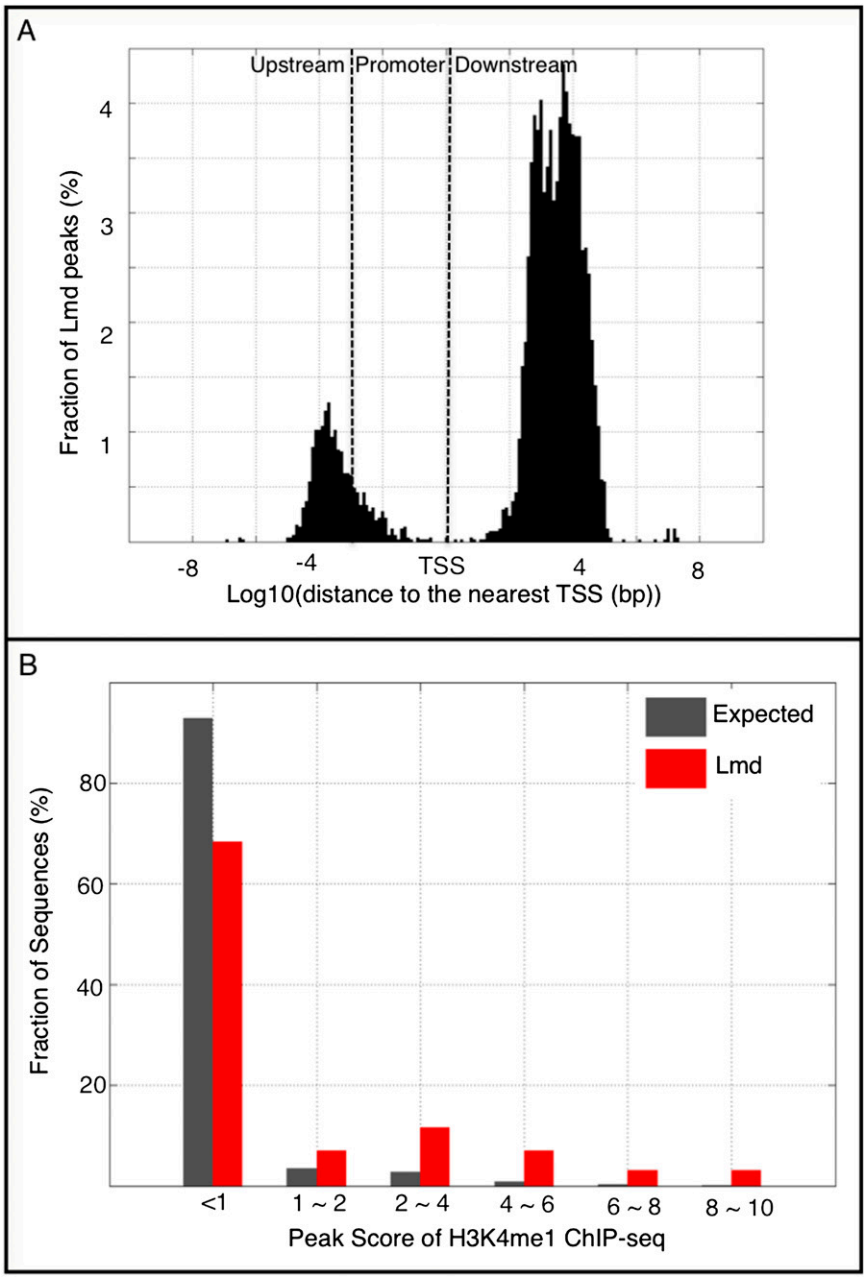

Fig. S4. Distribution of Lmd peaks. (*A*) Distribution of the distance between Lmd peaks and their most proximal RefSeq gene transcriptional start site (TSS). The RefSeq genes used here were annotated in the *D. melanogaster* Apr. 2006 (BDGP R5/*dm3*) assembly. (*B*) Distribution of ChIP-seq peak scores of H3K4me1. Given a Lmd-bound sequence, we calculated the average peak scores in a ChIP-seq experiment targeting H3K4me1 (1). Also, we calculated the expected distribution (i.e., the distribution of H3K4me1 peak scores along all noncoding DNA sequences). After sorting out all noncoding DNA sequences, we scanned these sequences with a window of 1,000 bp (length-matched to our Lmd peaks) in a step of 500 bp and calculated the average of peak scores within each window.

1. Bonn S, et al. (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44(2):148–156.

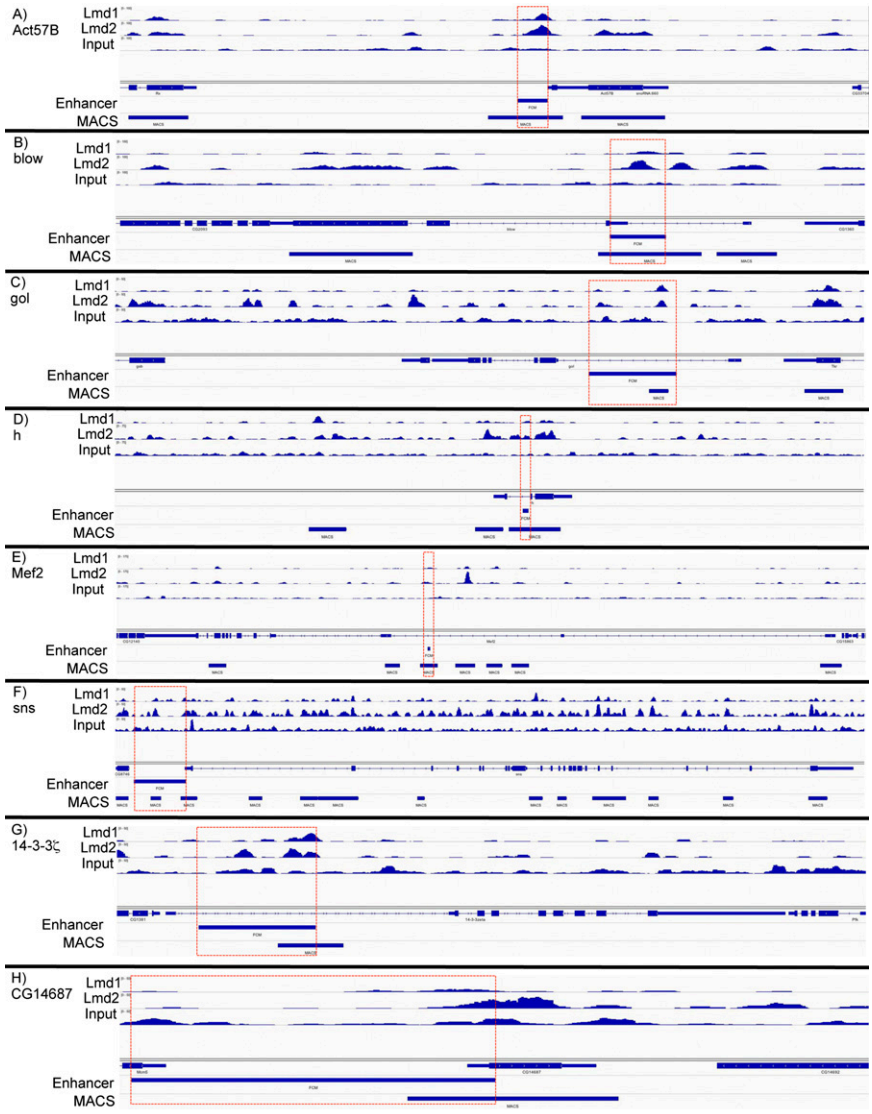

Fig. S5. Lmd binding to eight previously characterized FCM enhancers. Lmd ChIP-seq data were analyzed by MACS, and the generated wiggle files were imported to the IGV 2.0 browser for input chromatin and anti-Lmd1 and anti-Lmd2 immunoprecipitated chromatin for genomic regions surrounding *Act57B* (*A*), *blow* (*B*), *gol* (*C*), *h* (*D*), *Mef2* (*E*), *sns* (*F*), *14-3-3ζ* (*G*), and *CG14687* (*H*). For each of these genes, the previously characterized enhancers and the peaks identified by MACS for both immunoprecipitations are shown. For some genes (*gol*, *sns*, *14-3-3ζ*, and *CG14687*), the Lmd peaks can be used to predict minimal candidate enhancers. For other genes (*mef2* and *sns*), the additional Lmd peaks that are apparent may correspond to other FCM enhancers for the gene required to drive additional aspects of the gene's FCM expression pattern (1) or to provide redundancy and/or robustness of enhancer activity (2).

1. Nguyen HT, Xu X (1998) *Drosophila mef2* expression during mesoderm development is controlled by a complex array of *cis*-acting regulatory modules. *Dev Biol* 204(2):550–566.
2. Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321(5894):1314.

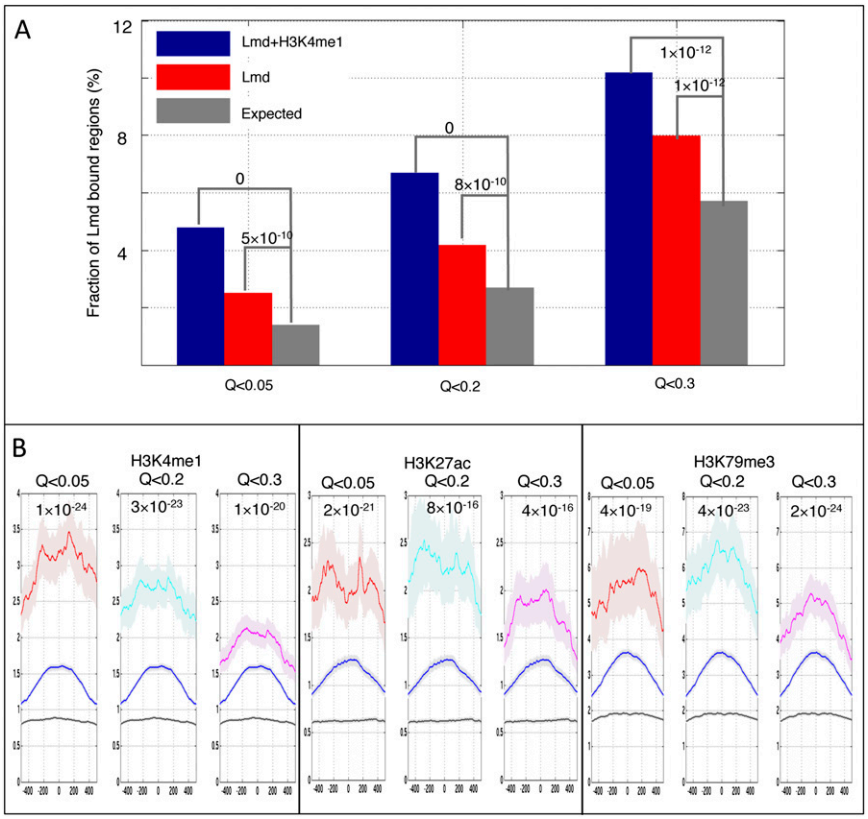

Fig. S6. Lmd-bound CREs associated with all *lmd*[1]-responsive genes determined in prior genome-wide gene expression profiling experiments. (*A*) Fraction of Lmd-bound CREs associated with genes down-regulated in response to *lmd*[1] compared with WT at the indicated Q-value cutoffs. Lmd-bound elements were related to their target genes such that intergenic sequences are associated with the nearest genes and intronic sequences are linked to their host gene. The binomial distribution was used to estimate the significance of the observations. (*B*) Enrichment of H3K4me1, H3K27ac, and H3K79me3 signals across Lmd-bound sequences [the vertical axis shows the peak score (1)]. The average signal along control (black line), all Lmd-bound (blue line), Q < 0.05 Lmd-bound (red line), Q < 0.2 Lmd-bound (cyan line), and Q < 0.3 Lmd-bound (magenta line) sequences. Shading highlights the 25th–75th percentile intervals. The *P* values are estimated by comparing the average signals of Lmd-bound sequences for the *lmd*[1]-responsive genes at different Q values with controls using the Wilcoxon rank-sum test.

1. Bonn S, et al. (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44(2):148–156.

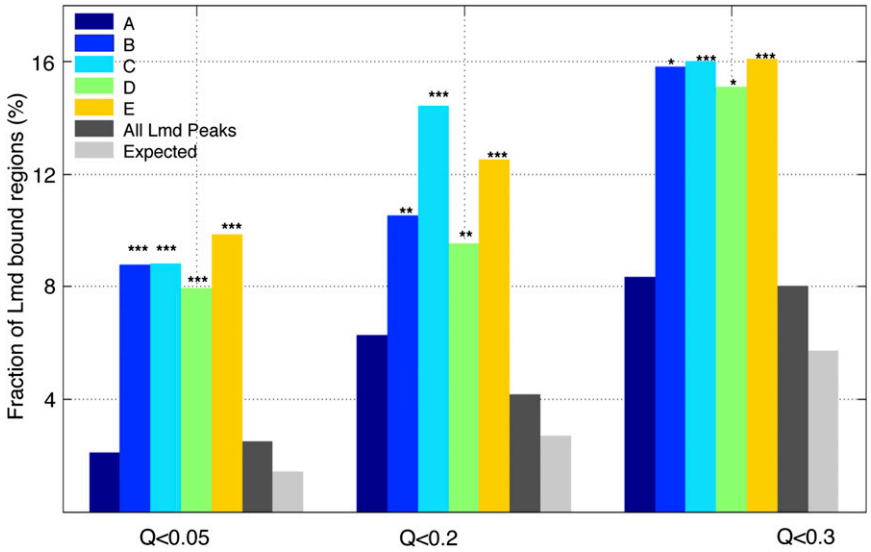

Fig. S7. Interaction between Lmd and other mesodermal TFs with all *lmd*[1]-responsive genes determined in prior genome-wide gene expression profiling experiments. (*A*) Clustering Lmd-bound sequences cooccupied by Tin, Twi, and Mef2. K-means clustering based on Euclidean distance was performed according to the presence/absence of the binding of other mesodermal TFs. (*B*) Distribution of the indicated Lmd peaks (from Fig. 2*A*) associated with genes down-regulated in response to *lmd*[1] compared with WT at the indicated Q-value cutoffs (see Dataset S1 for gene lists). The specificity *P* value is estimated by comparing the indicated cluster with "all Lmd peaks" using the hypergeometric distribution. (*$P < 10^{-2}$; **$P < 10^{-3}$; ***$P < 10^{-4}$.)

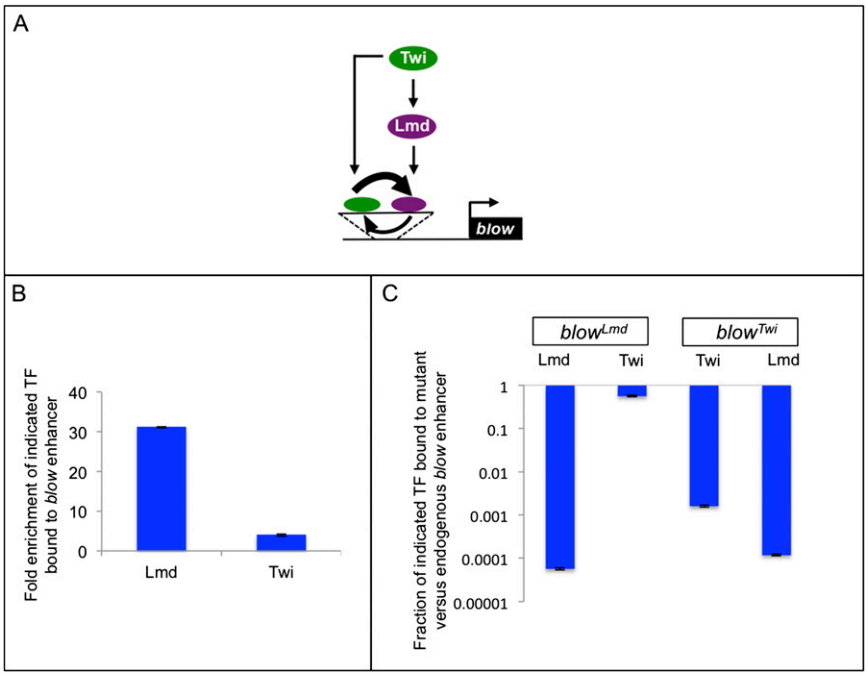

Fig. S8. Combinatorial regulation of FCM genes by Twi and Lmd. (*A*) Schematic depicting the feed-forward regulation of blow by Lmd and Twi and the cooperative influence of these two transcription factors in binding to the *blow* enhancer (the thicker arrow connecting Lmd and Twi on the enhancer is intended to represent the larger effect of Twi on Lmd binding than vice versa). A similar mechanism is likely responsible for the coregulation of other FCM genes by these two TFs (Fig. 2). (*B*) Quantitative real-time PCR analysis showing the fold enrichment of Lmd or Twi binding to a genomic sequence that encompasses the endogenous *blow* enhancer compared with a genomic region associated with *rp49* as a negative control. The fold enrichment was calculated in comparison with total input DNA using previously described procedures (1). The Twi antibody used in these studies was a gift of E. Furlong (European Molecular Biology Laboratory, Heidelberg, Germany) (2). Representative results of two experiments are shown. (*C*) Quantitative real-time PCR analysis showing the fraction of Lmd or Twi binding to a transgenic *blow* enhancer with mutated Lmd ($blow^{Lmd}$) or mutated Twi ($blow^{Twi}$) binding sites compared with the endogenous *blow* enhancer. As expected, $blow^{Twi}$ affects the ability to be bound by Twi, and $blow^{Lmd}$ affects the ability to be bound by Lmd. Interestingly, $blow^{Twi}$ leads to a greater than 1,000-fold reduction in Lmd binding, whereas $blow^{Lmd}$ leads to a modest 2-fold reduction in Twi binding. These results suggest that protein-protein interactions between these two TFs may aid their combinatorial occupancy of the *blow* enhancer. However, the significantly greater reduction in Lmd binding to the $blow^{Twi}$ mutant transgene strongly suggests that Twi binding precedes and aids the binding of Lmd to the *blow* enhancer, a hypothesis that is supported by the developmental timing of gene expression and loss-of-function analyses of these two TFs (3, 4). Chromatin was prepared from stage 11–12 embryos, and ChIP conditions were as detailed in *SI Materials and Methods*. The quantitative real-time PCR protocol was guided by a previous study (5) and require the use of allele-specific PCR primers and a blocking primer to suppress amplification of the inappropriate allele. PCR primers were as follows: $blow^{WT}$ (GGCATATCCACCAACTGCACACC and CCCGTCATCGCAATCGAA to detect the WT endogenous *blow* locus versus the $blow^{Lmd}$ locus, and CTGGCATATCCACCAACTG and CCCGTCATCGCAATCGAA to detect the WT endogenous *blow* locus versus the $blow^{Twi}$ locus), $blow^{Lmd}$ (GGCATATCCACCAACTGCACATT and CCCGTCATCGCAATCGAA to detect $blow^{Lmd}$ versus the WT endogenous *blow* locus), $blow^{Twi}$ (CTGGCATATCCACCCACTT and CCCGTCATCGCAATCGAA to detect $blow^{Twi}$ versus the WT endogenous *blow* locus), and *rp49* (CGGATCGATATGCTAAGCTG, TCTGTTGTCGATACCCTTGG). Blocking primers were phosphorylated at the 3′end and were as follows to discriminate the $blow^{Lmd}$ enhancer versus the endogenous *blow* enhancer ($blow^{endo}$, CAACTGCACACCACACGGCGTTTTGC and $blow^{Lmd}$, CAACTGCACATTACACGGCGTTTTGC) and for discriminating the $blow^{Twi}$ enhancer versus the endogenous *blow* locus ($blow^{endo}$, TATCCACCAACTGCACACCACA and $blow^{Twi}$, TATCCACCcACTtCACACCACA). Representative results of two experiments are shown.

1. Busser BW, et al. (2012) Molecular mechanism underlying the regulatory specificity of a *Drosophila* homeodomain protein that specifies myoblast identity. *Development* 139(6): 1164–1174.
2. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM (2009) Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* 462(7269):65–70.
3. Furlong EE, Andersen EC, Null B, White KP, Scott MP (2001) Patterns of gene expression during Drosophila mesoderm development. *Science* 293(5535):1629–1633.
4. Ruiz-Gómez M, Coutts N, Suster ML, Landgraf M, Bate M (2002) *myoblasts incompetent* encodes a zinc finger transcription factor required to specify fusion-competent myoblasts in *Drosophila*. *Development* 129(1):133–141.
5. Morlan J, Baker J, Sinicropi D (2009) Mutation detection by real-time PCR: a simple, robust and highly selective method. *PLoS ONE* 4(2):e4584.

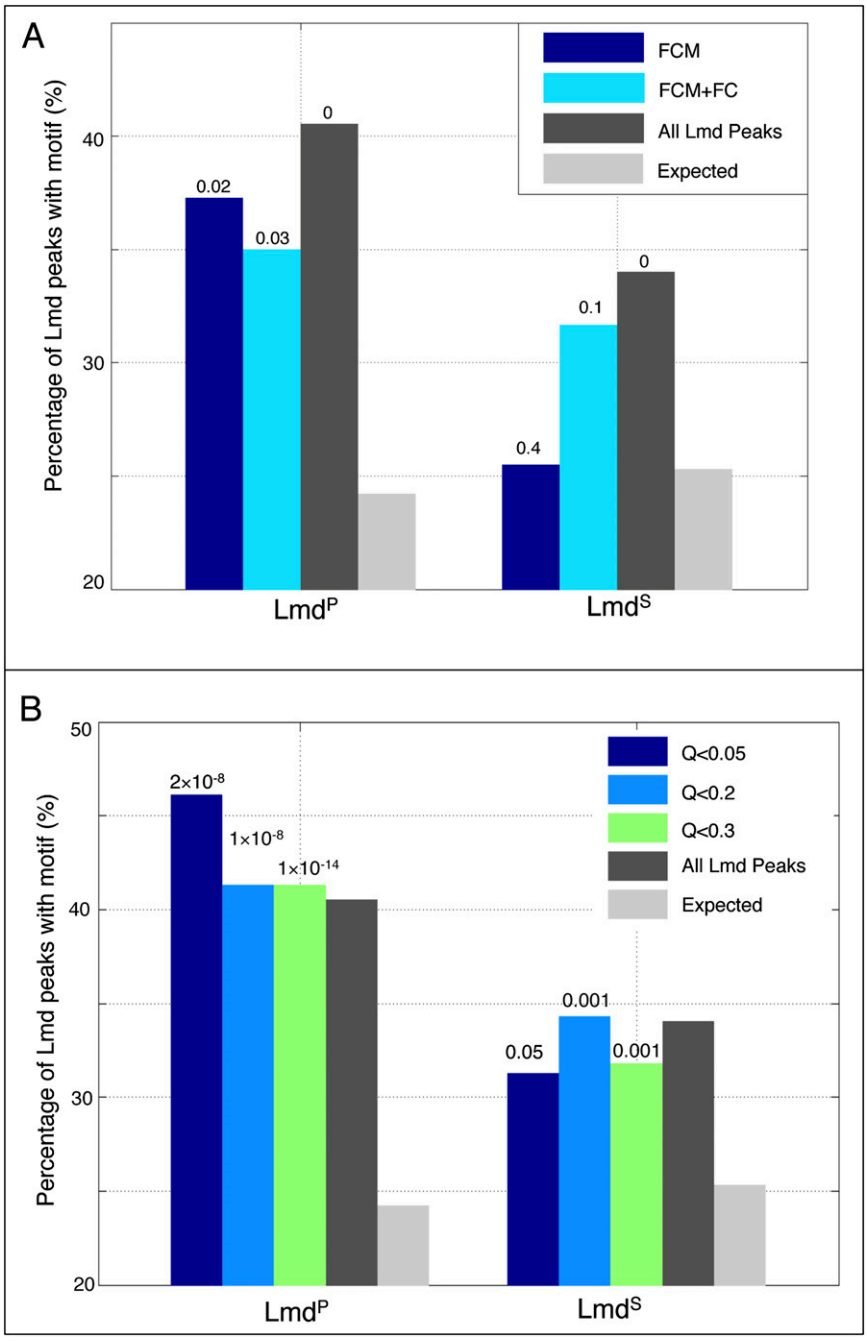

Fig. S9. Enrichments of Lmd^P and Lmd^S motifs among Lmd peaks. Lmd motifs were mapped to FCM and FCM+FC sequences (*A*), sequences associated with lmd^1 down-regulated genes at the indicated Q-value cutoffs (*B*), and "all Lmd peaks" and control (Expected) sequences using MAST (1). The percentages of sequences having Lmd^P or Lmd^S are shown separately. The expected here means the percentages of control sequences containing the indicated motifs. The control sequences, which were randomly sampled from noncoding DNA sequences, exhibited similar GC content, repeat density, and length with respect to Lmd peaks. *P* values are estimated under a binomial distribution by comparing with the expected. Note that, although Lmd^P motifs were significantly associated with both FCM and FCM+FC gene sets, there was a 1.5-fold enrichment of the Lmd^S motif among the FCM+FC gene set that was not statistically significant ($P = 0.1$), possibly because of the small sample size. The latter finding suggests the possibility that Lmd^S motifs are preferentially used to regulate FCM+FC gene activity, a hypothesis that will need to be further tested by identifying the regulatory elements that are targeted by Lmd and that contain functional Lmd^S motifs.

1. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue):W202-8.

Dataset S1. Lists of FCM and FCM+FC genes validated by in situ hybridization, all genes down-regulated in *lmd*^{1} loss-of-function mutant embryos as determined by prior genome-wide expression profiling, and expression, genomic coordinates, and binding of Mef2, Twi, and Lmd to tested and previously characterized enhancers.

Dataset S1

Dataset S2. DNA sequence motifs and weighting factors identified by the machine learning classifier

Dataset S2

Dataset S3. PBM E-scores and PWMs for Lmd

Dataset S3

Dataset S4. Mapping of TF binding sites onto the *blow* and IED5 *Mef2* FCM enhancers

Dataset S4