Cell, Vol. 111, 687–701, November 27, 2002, Copyright ©2002 by Cell Press

Whole-Genome Analysis of Dorsal-Ventral Patterning in the *Drosophila* Embryo

Angelike Stathopoulos,[1] Madeleine Van Drenth,[1] Albert Erives,[2] Michele Markstein,[3] and Michael Levine[1,4]
[1]Department of Molecular and Cell Biology
Division of Genetics and Development
401 Barker Hall
University of California, Berkeley
Berkeley, California 94720
[2]Auilix Biopharma, Inc.
950 Gilman Street
Berkeley, California 94710
[3]Committee on Developmental Biology
University of Chicago
Cummings Life Sciences Center, 755
920 East 58th Street
Chicago, Illinois 60637

Summary

The maternal Dorsal regulatory gradient initiates the differentiation of several tissues in the early *Drosophila* embryo. Whole-genome microarray assays identified as many as 40 new Dorsal target genes, which encode a broad spectrum of cell signaling proteins and transcription factors. Evidence is presented that a tissue-specific form of the NF-Y transcription complex is essential for the activation of gene expression in the mesoderm. Tissue-specific enhancers were identified for new Dorsal target genes, and bioinformatics methods identified conserved *cis*-regulatory elements for coordinately regulated genes that respond to similar thresholds of the Dorsal gradient. The new Dorsal target genes and enhancers represent one of the most extensive gene networks known for any developmental process.

Introduction

Gradient morphogens control a variety of metazoan patterning processes, including the primary axes of the *Drosophila* embryo (St. Johnston and Nusslein-Volhard, 1992; Courey and Huang, 1995; Rusch and Levine, 1996), the animal cap of the *Xenopus* embryo (Gurdon and Bourillot, 2001), the vertebrate neural tube (Jessell, 2000), the *Drosophila* wing imaginal disk (Strigini and Cohen, 1999), and the limb buds of chicks and mice (Martin, 2001). In most of these examples, extracellular gradients of cell signaling molecules such as Hedgehog, BMP, and FGF trigger the formation of transcription factor gradients by the differential activation of cell surface receptors, including Patched, Thickveins/Activin, and receptor tyrosine kinases (e.g., Podos and Ferguson, 1999; Van Buskirk and Schupbach, 1999; Ingham and McMahon, 2001). The resulting regulatory gradients initiate the formation of distinct cell types through the differential regulation of target genes that implement morphogenesis. This process is probably best understood for the dorsal-ventral patterning of the *Drosophila* embryo (reviewed by Stathopoulos and Levine, 2002a).

Dorsal is a member of the Rel family of sequence-specific transcription factors (Steward, 1987). It is initially distributed throughout the cytoplasm of developing oocytes but is transported into nuclei shortly after fertilization (reviewed by Belvin and Anderson, 1996). This regulated nuclear transport process leads to the formation of a broad Dorsal activity gradient, with peak activity in ventral regions and progressively lower levels in lateral and dorsal regions. The Dorsal nuclear gradient is formed by the differential activation of the Toll receptor, which probably depends on an extracellular gradient of the Spätzle ligand (reviewed by Roth, 1994; see Morisato, 2001). The resulting gradient initiates the differentiation of the mesoderm, neurogenic ectoderm, and dorsal ectoderm across the dorsal-ventral axis of the embryo through the differential regulation of ~10–15 previously characterized target genes (reviewed by Rusch and Levine, 1996; see Huang et al., 1997).

Most of the target genes encode sequence-specific transcription factors, including *twi*, *sna*, *vnd*, *brk*, and *zen* (reviewed by Stathopoulos and Levine, 2002a). In order to obtain a more complete understanding of how the Dorsal gradient controls development, Affymetrix chips containing the entire protein coding capacity of the *Drosophila* genome (greater than 13,500 genes) were screened with RNAs extracted from early mutant embryos that contain either no Dorsal protein, uniformly low levels of Dorsal, or uniformly high levels of Dorsal throughout the embryo. Mutant embryos that lack Dorsal overexpress target genes that are normally localized within the dorsal ectoderm, while embryos that contain uniformly low or high levels of Dorsal overexpress neurogenic genes or mesoderm genes, respectively. A total of 353 genes exhibit augmented expression in one or more of these mutant backgrounds; 57 of the genes display significant changes in expression and include as many as 40 new target genes that exhibit localized patterns of expression across the dorsal-ventral axis of wild-type embryos.

Previous studies identified four Dorsal target enhancers; two are activated by peak levels of the Dorsal gradient (*twist* and *snail*), one by intermediate levels (*rhomboid*), and one by low levels (*sog*; see Stathopoulos and Levine, 2002a). New Dorsal target enhancers were characterized in order to investigate the basis for gradient thresholds of gene expression. Three new enhancers were identified, thereby providing at least two enhancers for each of the three major Dorsal gradient thesholds. Bioinformatics methods identified conserved sequence motifs among coordinately regulated enhancers. For example, CACATGT is shared by target enhancers activated by intermediate levels of the Dorsal gradient (*vnd* and *rhomboid*), while GCTGGAA is present in enhancers activated by low levels of the gradient (*Neu4* and *sog*). The new target genes and associated *cis*-regulatory DNAs identified in this study constitute

[4]Correspondence: mlevine@uclink4.berkeley.edu

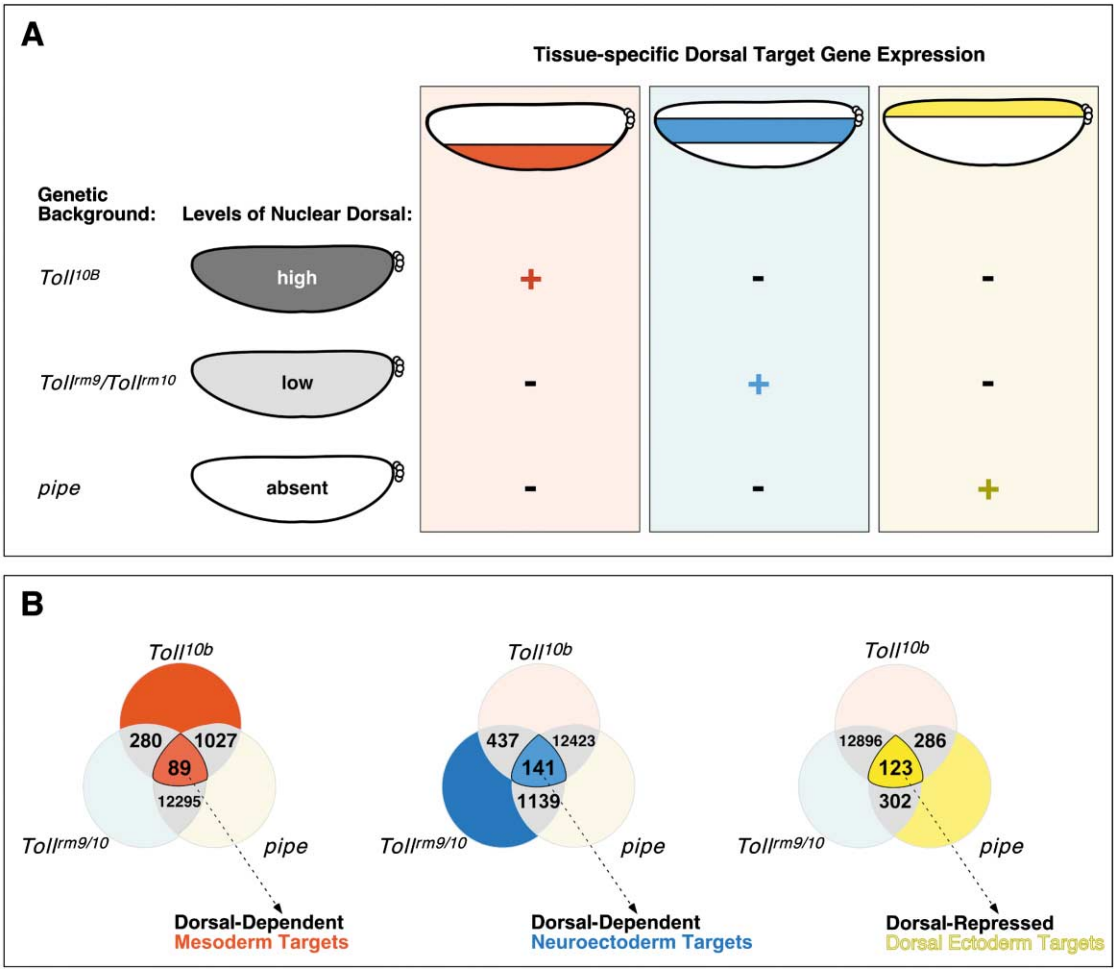

Figure 1. Summary of Microarray Experiments

(A) RNA was isolated from three different genetic backgrounds containing varying amounts of nuclear Dorsal protein. *Toll*^{10B} is a dominant mutation in the Toll receptor that leads to high levels of nuclear Dorsal. *Toll*^{rm9}/*Toll*^{rm10} is a recessive mutation in the Toll receptor that leads to low levels of nuclear Dorsal throughout the early embryo. In *pipe*$^{-}$ (*pipe*^{386}/*pipe*^{664}) mutant embryos, Dorsal fails to translocate to the nucleus and remains cytoplasmic. Dorsal targets requiring the highest levels of Dorsal for expression are normally only expressed in ventral parts of the embryo, but ubiquitously in the *Toll*^{10B} background. Dorsal targets responding to lower levels of nuclear Dorsal are expressed in a broader domain that extends to the lateral regions of the embryo but are absent from the ventral surface due to repression by Snail. These targets are expressed ubiquitously in the *Toll*^{rm9}/*Toll*^{rm10} background, which contains sufficient levels of Dorsal to promote activation but insufficient levels to activate the repressor Snail. Genes repressed by Dorsal in the dorsal ectoderm are expressed at the dorsal domain of the embryo. These genes are expressed ubiquitously in the *pipe*$^{-}$ mutant background due to an absence of nuclear Dorsal. A plus sign indicates the presence of specific targets gene expression in a certain mutant background, whereas a minus sign indicates its absence.

(B) To isolate additional Dorsal target genes, RNA was isolated from each of these three mutants and hybridized to Affymetrix *Drosophila* GeneChips. Genes exhibiting increased expression in response to only the highest levels of nuclear Dorsal (89) were identified by determining which genes are contained within the intersection of three comparisons: those signals that increased at least 3-fold (measured in $\log_2$) in *Toll*^{10b} relative to *Toll*^{rm9}/*Toll*^{rm10} (369) and *pipe* (1116) but changed less than 3-fold in a comparison of *Toll*^{rm9}/*Toll*^{rm10} versus *pipe* (12384). Genes responding to intermediate levels of nuclear Dorsal (141) were identified by determining which signals increased at least 3-fold in *Toll*^{rm9}/*Toll*^{rm10} relative to *Toll*^{10b} (578) and *pipe* (1280) but changed less than 3-fold in Toll^{10b} versus *pipe* (12564). Genes repressed by Dorsal (123) were identified by determining which signals increased at least 3-fold in *pipe* relative to *Toll*^{10b} (409) and *Toll*^{rm9}/*Toll*^{rm10} (425) but changed less than 3-fold in *Toll*^{10b} versus *pipe* (13019). These putative Dorsal targets represent genes expressed in the presumptive mesoderm, neuroectoderm, and dorsal ectoderm, respectively.

one of the most extensive gene regulation networks known for any developmental process.

Results

Because Dorsal exhibits strict maternal inheritance, it is possible to isolate homogenous populations of mutant embryos that either lack the gradient (*pipe*$^{-}$) or contain uniformly high (*Toll*^{10b}) or low (*Toll*^{rm9}/*Toll*^{rm10}) levels of Dorsal in all nuclei (summarized in Figure 1A). Embryos were collected from each of these three classes of mutant females and allowed to develop until cellularization, about 3 hr after fertilization. Dorsal enters nuclei between 90 min to 2 hr after fertilization and activates

target genes such as *twist*, *snail*, and *rhomboid* within the next 30 min to 1 hr (reviewed by Rusch and Levine, 1996). Cellularized embryos should express most of the direct Dorsal target genes, as well as genes that are indirectly regulated by Dorsal (see Discussion). Affymetrix chips containing the complete protein coding capacity of the *Drosophila* genome (greater than 13,500 genes) were separately hybridized with RNAs extracted from each mutant. A total of 353 genes exhibit at least a 3-fold increase in one of the mutant backgrounds (Figure 1B): 89 genes are upregulated in $Toll^{10B}$ mutants, 141 genes are upregulated in $Toll^{rm9}/Toll^{rm10}$ mutants, and 123 genes are upregulated in $pipe^{-}/pipe^{-}$ mutants.

Validation of the Microarray Screens

Representative genes that exhibit upregulation in $Toll^{10B}$, $Toll^{rm9}/Toll^{rm10}$, and $pipe^{-}/pipe^{-}$ mutant embryos were selected for further study. These genes fulfill one of the following criteria: availability of an EST, insight into function based on a known protein motif, or conservation in other organisms. Furthermore, the "cutoff" for new putative Dorsal target genes was assessed by examining known genes. For example, most of the known mesoderm-specific genes such as *twist*, *snail*, and *tinman* display at least a ~10-fold increase in $Toll^{10B}$ mutants as compared with the other genetic backgrounds. Only 19 of the 89 genes that are significantly upregulated (greater than 3-fold) in $Toll^{10B}$ mutant embryos fulfill this cutoff value (Table 1).

One such gene, *Mes3*, encodes an insulin-like growth factor (*dilp4*) (Brogiolo et al., 2001). In wild-type embryos, the gene is selectively expressed in ventral regions that will form the mesoderm (Figure 2C). The augmented expression of *Mes3* that is seen in $Toll^{10B}$ mutants (Table 1) can be explained by the marked expansion of the staining pattern in mutant embryos (Figure 2D). A similar expansion is seen for a known gene, *snail*, which is normally expressed in the ventral mesoderm (Figures 2A and 2B).

Dorsal target genes that are expressed in the lateral neurogenic ectoderm, such as *brinker* and *sog* (Table 2), exhibit just 6-fold increases in expression in $Toll^{rm9}/Toll^{rm10}$ mutant embryos as compared with the other genetic backgrounds. Only 22 of the 114 genes that are upregulated in $Toll^{rm9}/Toll^{rm10}$ mutant embryos fulfilled this cutoff value (Table 2). One of these genes, *Neu3*, encodes a novel member of the ADAM family of endopeptidases (see Qi et al., 1999; Lieber et al., 2002). It is expressed in broad lateral stripes in wild-type embryos (Figure 2G) but is expressed throughout the dorsal-ventral axis of mutant embryos derived from $Toll^{rm9}/Toll^{rm10}$ females (Figure 2H). This expanded expression is consistent with the results of the microarray assays, which indicate a 15-fold increase of *Neu3* expression in mutant embryos as compared with wild-type embryos (see Table 2). A similar expansion is observed for a known gene, *sog* (Figures 2E and 2F).

Genes that are specifically expressed within the dorsal ectoderm were identified on the basis of exhibiting augmented expression in $pipe^{-}/pipe^{-}$ embryos (Table 3). *zen* is directly repressed by the Dorsal gradient and is expressed in a progressively more refined pattern during cellularization (Ip et al., 1991). It exhibits a 22-fold increase in expression in $pipe^{-}/pipe^{-}$ mutant embryos (Table 3). However, other known genes such as *pannier* and *u-shaped* display just 6-fold increases in expression. Only 16 of the 123 genes that are upregulated in $pipe^{-}/pipe^{-}$ mutant embryos fulfill this cutoff value (Table 3). One of the genes, *Ect1*, encodes a TNF (tumor necrosis factor) signaling molecule (reviewed by Baud and Karin, 2001). This gene is normally expressed in the dorsal ectoderm of wild-type embryos (Figure 2K) but is expressed in both dorsal and ventral regions of mutant embryos derived from $pipe^{-}/pipe^{-}$ females (Figure 2L). This expanded staining pattern accounts for the 12-fold increase in the levels of *Ect1* expression observed in the microarray assays (Table 3). A similar expansion is observed for a known gene, *dpp* (Figures 2I and 2J).

Localized Expression of New Target Genes

In situ hybridization assays were done for five genes that exhibit at least a 10-fold increase in expression in $Toll^{10B}$ mutants (Table 1). All five genes display localized expression in the ventral mesoderm in wild-type embryos (Figures 3A–3E). Most of the genes encode proteins that might influence changes in cell size and shape during gastrulation (see Discussion).

Twelve of the genes that exhibit at least a 6-fold increase in expression in $Toll^{rm9}/Toll^{rm10}$ mutants were analyzed (Table 2). Seven of these genes exhibit localized expression within the neurogenic ectoderm, whereas five of the genes do not (data not shown). The expression patterns of four of the localized genes are shown (Figures 3F–3I). Two of the genes, *Neu3* and *Neu4*, exhibit more than a 10-fold increase in expression and display broad lateral stripes of expression in the presumptive neurogenic ectoderm (Figures 3H and 3I). In contrast, *Neu1* and *Neu2* exhibit only a 6-fold increase in expression in $Toll^{rm9}/Toll^{rm10}$ mutants and display narrower lateral stripes of expression (Figures 3F and 3G). To determine whether genes with smaller increases might respond to different thresholds of the Dorsal gradient, the *Neu5*/*Sulfated* gene was analyzed since it exhibits just a ~4-fold increase in $Toll^{rm9}/Toll^{rm10}$ embryos. *Neu5* displays weak, nonuniform staining in the lateral neurogenic ectoderm (Figure 3J).

In situ hybridization assays were done with six different genes that exhibit at least a 6-fold increase in expression in mutant embryos derived from $pipe^{-}/pipe^{-}$ females (see Table 3). Five of the genes that were tested exhibit localized expression within the dorsal ectoderm (Figures 3K–3O), whereas one of the genes does not (data not shown). Among the set of 123 upregulated genes, a homolog of *Ect2*, *CG5093*, displays 5-fold upregulation. *CG5093* exhibits a localized pattern of expression in the dorsal ectoderm that is virtually identical to the *Ect2* pattern (data not shown; Figure 3L).

A Tissue-Specific NF-Y Subunit Activates Gene Expression in the Mesoderm

The preceding microarray assays identified a number of genes that are likely to be important for the dorsal-ventral patterning of the early embryo. We selected *Mes4* for further analysis since it encodes a putative mesoderm-specific transcription factor and previous studies suggested that Dorsal is not sufficient for robust

Table 1. Dorsal Targets in the Mesoderm

Mesoderm		Function/Homology	10B/[rm9/10]	10B/pipe	pipe/[rm9/10]	Confirmed by In Situ	Reference
Known Targets							
fog	CG9559	unknown	1.1	1.1	−1.0		Costa et al., 1994
heartless (htl)	CG7223	FGF receptor	14.9	7.7	1.4		Shisido et al., 1993
mef2	CG1429	MADS-box TF	9.5	7.1	−1.5		Taylor et al., 1995
snail (sna)	CG3956	transcriptional repressor	3.1	14.2	−5.0		Ip et al., 1992a
tinman (tin)	CG7895	homeobox TF	10.2	7.0	−2.3		Bodmer et al., 1990
twist (twi)	CG2956	bHLH TF	63.8	31.3	−1.2		Jiang et al., 1991
zfh1	CG1322	Zn-finger/homeobox TF	3.6	1.8	2.2		Lai et al., 1991
Microarray-Identified Targets							
Mes1 = RhoL	CG9366	Rho GTPase	11.9	7.1	−1.7	A, B	Casal and Leptin, 1996
Mes2	CG11100	ZPW domain	11.0	6.5	−1.7	A	
Mes3 = dilp4	CG6736	insulin receptor ligand	38.0	38.0	−2.1	A, B	Brogiolo et al., 2001
Mes4	CG11301	NF-YC homolog	10.7	5.0	1.4	A	
Mes5 = Mdr49	CG3879	ABC transporter	11.6	24.1	−7.5	A	
Actin57B	CG10067	cytoskeleton	37.0	17.9	2.2	B	Casal and Leptin, 1996
Argk	CG4929	arginine kinase	9.8	6.9	−3.5		
Asph	CG18658	aspartyl β-hydroxylase	5.1	11.6	−2.0		
Cyp310a1	CG10391	cytochrome P450	28.9	7.5	2.1		
glial cells missing	CG12245	gcm TF	11.7	5.3	1.7		
hoi-polloi	CG3949	snRNP	9.7	3.7	2.6		
rosy	CG7642	xanthine dehydrogenase	13.6	6.8	1.3	B	Doyle et al., 1989
stumps	CG18485	cell signaling	18.9	9.4	1.0	B	Casal and Leptin, 1996
trachealess	CG6883	HLH/PAS TF	10.5	4.7	−2.0		
	CG4500	acyl-CoA synthetase	15.0	5.9	2.6		
	CG7780	Dnase	9.7	5.4	1.0		
	CG12177	nucleoside hydrolase	8.3	13.5	1.3		
	CG14036	unknown (Xenopus.hom)	14.8	6.7	5.2		
	CG14688	unknown (worm.hom)	25.2	5.7	3.2		

TF = transcription factor
A, predicted expression pattern confirmed (this study).
B, predicted expression pattern confirmed (referenced study).

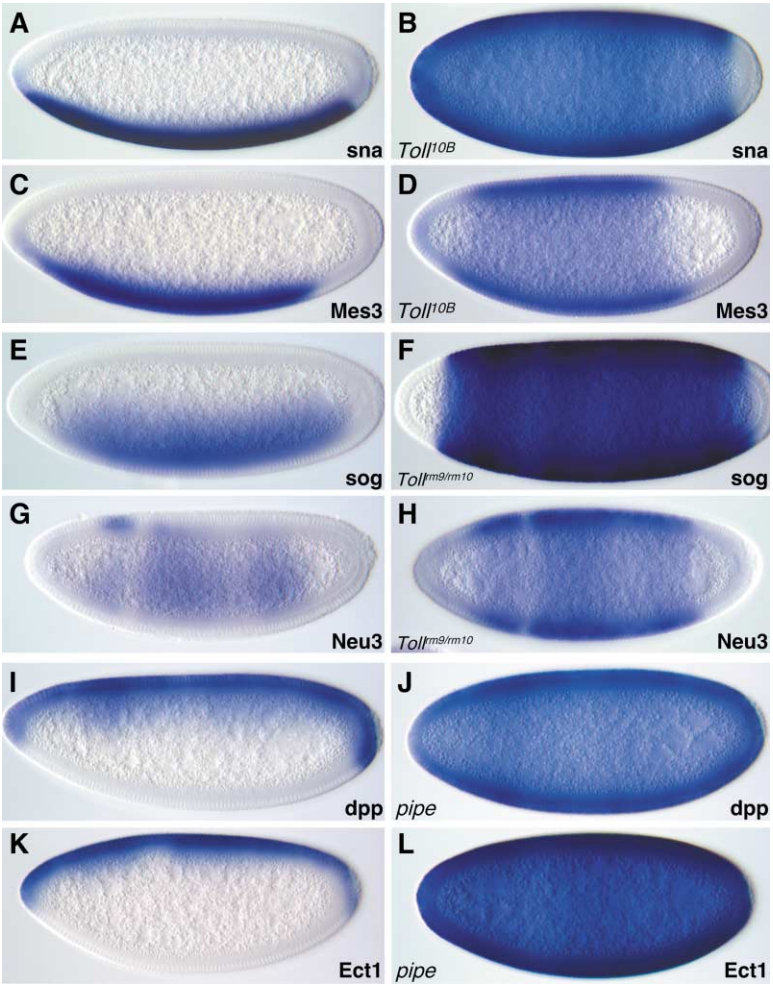

Figure 2. Expression of Dorsal Target Genes in Wild-Type and Mutant Embryos

Cellularizing embryos were hybridized with each of the indicated digoxigenin-labeled antisense RNA probes and stained to visualize the gene expression patterns. Embryos are oriented with anterior to the left and dorsal up.

(A and B) *snail (sna)* expression in wild-type (A) and *Toll*^{10B} embryos (B). *snail* is normally expressed in the ventral mesoderm of wild-type embryos (A) but is ubiquitously expressed in *Toll*^{10B} mutants (B).

(C and D) *Mes3* expression in wild-type (C) and *Toll*^{10B} mutant background (D). As seen for *snail, Mes3* is normally expressed in the mesoderm, but is greatly expanded in the mutant.

(E and F) *sog* expression in wild-type (E) and *Toll*^{rm9}/*Toll*^{rm10} (F) embryos. *sog* is normally expressed in lateral stripes in response to even low levels of nuclear Dorsal (E). The gene exhibits ubiquitous expression in mutant embryos (F).

(G and H) *Neu3* expression in wild-type (G) and *Toll*^{rm9}/*Toll*^{rm10} mutants (H). As seen for *sog*, there is a marked expansion of the expression pattern in mutant embryos.

(I and J) *dpp* expression in wild-type (I) and *pipe*$^{-}$ mutant (J) embryos. *Dpp* is normally expressed in the dorsal ectoderm (I) but is derepressed in mutant embryos (J).

(K and L) *Ect1* expression in wild-type (K) and *pipe*$^{-}$ (L) mutants. As seen for *dpp*, the expression pattern is expanded in mutant embryos.

activation of gene expression in the mesoderm (Jiang et al., 1991; Ip et al., 1992a; Szymanski and Levine, 1995). *Mes4* encodes a protein that contains a histone-fold multimerization domain and is related to the C subunit of the mammalian NF-Y transcription complex. NF-Y is a trimeric complex that is composed of three subunits, A, B, and C. It binds to conserved CCAAT motifs that are located between −100 bp and −60 bp 5′ of the transcription start site in ~30% of all mammalian promoters (Figure 4A). It has been proposed that NF-Y facilitates the binding of TFIID to adjacent core promoter elements such as TATA (reviewed by Mantovani, 1999). There is a curious absence of CCAAT sequences in *Drosophila* promoters, and there have been no reports of an NF-Y complex. However, it is likely that flies contain NF-Y since this complex is conserved in yeast, plants, and vertebrates.

A survey of the *Drosophila* genome reveals clear orthologs of all three NF-Y encoding genes. *CG3891* encodes a protein that contains 56/72 AA similarities with the histone-fold domain of the mammalian NF-YA subunit. *CG10447* contains 87/118 AA similarities with the histone-fold domain and adjacent regions of the mammalian NF-YB subunit. Finally, *CG3075* contains 91/113 similarities with the histone-fold region of the NF-YC subunit. The *Mes4* gene (*CG11301*, see Table 1) encodes a divergent copy of the NF-YC subunit that shares 49/78 similarities with the histone-fold region of *CG3075*. It is considerably more closely related to NF-YC than the other histone-fold proteins in *Drosophila*, such as Dr1, Drap1, $TAF_{II}80$, and Chrac-16 (Aoyagi and Wassarman, 2000).

The three primary NF-Y genes are maternally expressed, suggesting their encoded RNAs are ubiquitously distributed throughout early embryos (BDGP; http://www.fruitfly.org/cgi-bin/ex/insitu.pl). However, expression of the C subunit gene is rapidly lost in early embryos, at the time when the *Mes4* gene is first activated in the presumptive mesoderm (Figure 4B). This raises the possibility that a functional NF-Y complex is made only in the mesoderm during gastrulation and germband elongation. The *Mes4* staining pattern is initially uniform but exhibits segmental modulations along the anterior-posterior axis by the completion of cellularization. Expression is restricted to the ventral-most 18–20 cells, which coincides with the presumptive mesoderm (Figure 4H). Staining persists in the mesoderm during invagination and germband elongation (Figures 4C and 4D). At the time of invagination, *Mes4* transcripts exhibit clear segmental repeats and the staining

Table 2. Dorsal Targets in the Neuroectoderm

Neuroectoderm		Function/Homology	[rm9/10]/10B	pipe/10B	[rm9/10]/pipe	Confirmed by In Situ	Reference
Known Targets							
brinker (brk)	CG9653	transcriptional repressor	4.9	−1.2	5.7		Jazwinska et al., 1999
ind	CG11551	homeobox TF	17.8	−3.9	17.6		Weiss et al., 1998
lethal of scute (T3)	CG3839	HLH TF	4.2	1.4	3.0		Romani et al., 1987
rhomboid (rho)	CG1004	serine-type peptidase	2.9	−1.2	1.9		Ip et al., 1992b
single-minded (sim)	CG7771	HLH/PAS TF	−4.2	−2.2	1.5		Kasai et al., 1998
sog	CG9224	growth factor	6.4	−1.5	5.1		Francois et al., 1994
vnd	CG6172	homeobox TF	4.0	−2.2	3.1		McDonald et al., 1998
Microarray-Identified Targets							
Neu1 = Dscam	CG17800	axon guidance receptor	5.5	1.6	2.0	A	
Neu2	CG7204	zinc-finger protein	5.1	−1.1	5.7	A	
Neu3	CG7649	ADAM metalloprotease	15.4	−1.1	8.9	A	
Neu4	CG12443	unknown function	11.1	−1.1	6.3	A	
Neu5 = sulfated	CG6725	sulfotransferase	2.2	−2.3	3.7	A	below cutoff
echinoid	CG12676	cell adhesion	7.0	−2.7	5.3	NL	
leak	CG5574	axon guidance receptor	6.9	−2.8	7.7		
patched	CG2411	receptor	11.9	−2.9	8.1	NL	
PGRP-SC2	CG14745	defense/immunity	6.2	−2.0	5.8		
Scabrous	CG17579	receptor ligand	19.7	3.1	6.9	A, B	Mlodzik et al., 1990
Socs36E	CG15154	cell signaling	13.5	−2.1	6.7		
SoxNeuro	CG18024	HMG-box TF	18.0	−3.1	11.2	A, B	Cremazy et al., 2000
warts	CG12072	kinase	4.4	−2.7	6.0	NL	
	CG4589	calcium binding	5.6	−1.8	6.5		
	CG4749	methyltransferase	12.0	−2.1	7.4		
	CG2264	serine protease inhibitor	8.2	−1.8	5.3	A	
	CG5891	myosin binding	6.8	−2.4	5.2		
	CG6400	WD40 repeats	7.8	−1.9	5.0	NL	
	CG12467	myosin/kinesin motor	7.8	−1.8	5.9	NL	
	CG13797	unknown (embryoEST)	8.2	−2.3	6.7		
	CG14411	unknown (homologs)	7.1	−2.4	5.2		
	CG17229	G-protein signaling	7.6	−1.9	5.1		
	CG18353	zinc-finger protein	6.6	−2.6	5.3		

TF = transcription factor
A, predicted expression pattern confirmed (this study).
B, predicted expression pattern confirmed (referenced study).
NL = not localized

Table 3. Dorsal Targets in the Dorsal Ectoderm

Dorsal Ectoderm		Function/Homology	[rm9/10]/10B	pipe/10B	pipe/[rm9/10]	Confirmed by In Situ	Reference
Known Targets							
decapentaplegic (dpp)	CG9885	TGF-β receptor ligand	2.4	5.1	3.6		St. Johnston and Gelbart, 1987
pannier (pnr)	CG3978	GATA TF	1.4	6.4	6.3		Winick et al., 1993
race	CG8827	peptidyl dipeptidase A	−1.7	3.3	5.4		Tatei et al., 1995
tolloid (tld)	CG6868	endopeptidase	7.1	11.4	4.4		Kirov et al., 1994
tailup (tup)	CG10619	homeobox TF	1.3	3.2	3.2		Frank and Rushlow, 1996
u-shaped (ush)	CG2762	Zn-finger TF	1.2	6.0	5.2		Frank and Rushlow, 1996
zerknüllt (zen)	CG1046	homeobox TF	−1.0	21.9	22.7		Doyle et al., 1989
Microarray-Identified Targets							
Ect1	CG12919	TNF receptor ligand	2.1	11.9	11.4	A	
Ect2 = Dorsocross	CG5133	T domain TF	1.1	28.6	32.5	A	
Ect3	CG3132	galactosidase	−1.4	15.5	20.9	A	
Ect4	CG7915	SAM,TIR, Pro-rich domains	−1.0	6.4	6.6	A	
Ect5 = C15	CG7937	homeobox TF	1.6	6.8	5.2	A	
Adult cuticle protein1	CG7216	structural cuticle protein	2.2	7.3	6.0		
Lcp65Ag2	CG10534	structural cuticle protein	−3.1	13.3	19.9		
	CG5093	T domain TF	−1.0	4.9	5.1	A	below cutoff
	CG5468(=CG5476)	unknown (embryo EST)	1.6	17.2	15.5		
	CG5476(=CG5468)	unknown (embryo EST)	1.6	11.4	11.2		
	CG6180	phospholipid binding	−1.4	6.1	8.6	NL	
	CG9358	ligand binding/carrier	1.9	6.0	6.4		
	CG9466	alpha-mannosidase	2.6	6.4	5.4		
	CG11776	bipartite nls	−1.6	6.1	6.4		
	CG12011	unknown (embryo EST)	1.2	10.2	6.8		
	CG14365	bipartite nls; Pro-rich	1.6	6.5	4.0		
	CG15282	unknown (embryo EST)	1.8	7.6	7.2		

TF = transcription factor
A, predicted expression pattern confirmed (this study).
NL = not localized

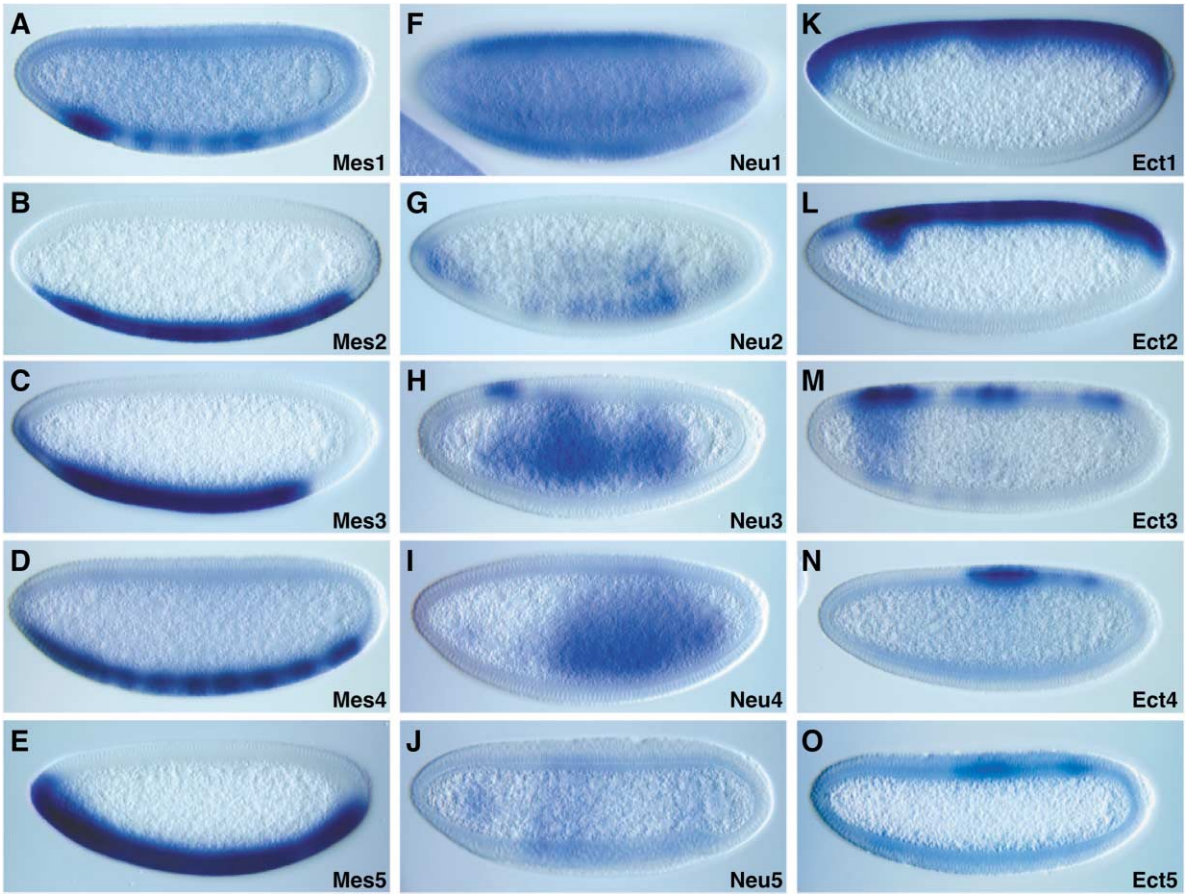

Figure 3. Expression of New Dorsal Target Genes

Cellularizing embryos were hybridized with each of the indicated digoxigenin-labeled antisense RNA probes and are oriented with anterior to the left and dorsal up.

(A–E) Genes expressed in the presumptive mesoderm responding to only the highest levels of nuclear Dorsal. (A) *Mes1/RhoL*, (B) *Mes2/CG11100*, (C) *Mes3/dilp4*, (D) *Mes4/CG11301*, and (E) *Mes5/Mdr49*.

(F–J) Genes expressed in the presumptive neuroectoderm responding to low levels of nuclear Dorsal. (F) *Neu1/Dscam*, (G) *Neu2/CG7204*, (H) *Neu3/CG7649*, (I) *Neu4/CG12443*, and (J) *Neu5/sulfated*.

(K–O) Genes expressed in the presumptive dorsal ectoderm presumably repressed by nuclear Dorsal. (K) *Ect1/CG12919*, (L) *Ect2/Dorsocross*, (M) *Ect3/CG3132*, (N) *Ect4/CG7915*, and (O) *Ect5/C15*.

pattern consists of six "stripes" between the cephalic furrow and presumptive abdomen; a seventh stripe is seen just anterior to the furrow (Figure 4H).

The distinctive *Mes4* staining pattern provides evidence that it functions as an activator of *snail* (see below). This was directly tested by placing the *Mes4* protein coding region under the control of the *Krüppel* 5′ regulatory region. Transgenic embryos that contain the *Krüppel-Mes4* fusion gene exhibit an ectopic band of staining in central regions, in addition to the normal pattern in ventral regions (Figure 4E; compare with Figure 4B). The ectopic *Mes4* expression pattern leads to the misexpression of *snail*. Normally, *snail* expression is restricted to ventral regions that form the mesoderm (Figure 4F), but it is weakly misexpressed in central regions of transgenic embryos carrying the *Krüppel-Mes4* fusion gene (Figure 4G).

snail is normally expressed uniformly, but exhibits nonuniform stripes in *twist*$^{-}$/*twist*$^{-}$ mutant embryos (Figure 4I; Ip et al., 1992a). These stripes are very similar to those seen for the normal *Mes4* expression pattern (compare Figures 4H and 4I). Thus, in the absence of Twist, Dorsal and Mes4 might function as the primary activators of the *snail* expression pattern.

Identification of New Dorsal Target Enhancers

Previous studies identified only four Dorsal target enhancers (reviewed in Stathopoulos and Levine, 2002a; see Markstein et al., 2002). This is not a sufficient collection to determine whether enhancers that respond to similar thresholds of the Dorsal gradient contain shared *cis*-regulatory elements. New enhancers were identified for the *Mes3* and *Neu4* genes by scanning ~25 kb of associated genomic DNA for clusters of Dorsal recognition sequences. Unlike previous whole-genome screens for clusters of optimal Dorsal binding sites (Markstein et al., 2002), the search for putative enhancers associated with known genes is much less stringent and permits

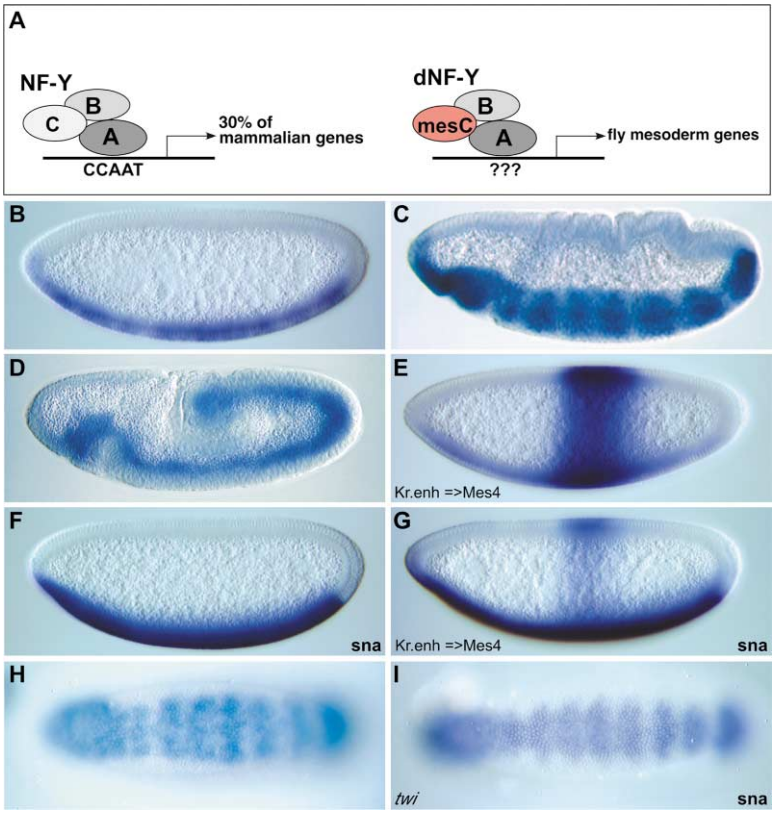

Figure 4. *Mes4*, a Novel Dorsal Target Gene Identified by Microarray, Contains Homology to NF-YC and Is Involved in Regulating *snail* Expression

(A) NF-Y is a heterotrimeric protein composed of subunits A, B, and C that binds to CCAAT motifs present from −100 to −60 bp upstream of the major transcriptional start site in 30% of mammalian promoters. *Drosophila* contains orthologs of all three subunits, and *Mes4* is the first described tissue-specific metazoan NF-YC homolog. We suggest that a tissue-specific NF-Y complex containing the Mes4 subunit helps activate the expression of mesoderm-specific genes, such as *snail*.

(B–I) In situ hybridization experiments were performed using antisense probes to either *Mes4* (B–E and H) or *snail* (*sna;* F, G, and I). *Mes4* is normally expressed in the presumptive mesoderm, in ventral regions of the embryo. (B) and (C) are lateral views of *Mes4* expression in mid-nuclear cleavage cycle 14 (B) and gastrulating (C) embryos. The expression pattern exhibits pair-rule modulations along the anterior posterior axis (H; ventral view) and persists in the mesoderm through germband elongation (D). To study its function, *Mes4* was cloned into a P element insertion vector that promotes ectopic expression in the region of *Krüppel* expression (Kr-FSF; see Experimental Procedures) and was used to generate transgenic lines that misexpress this gene (*Kr.enh=>Mes4*; compare [E] with [B]). *snail*, which is normally expressed only in ventral regions of the embryo in the presumptive mesoderm (F), is ectopically expressed in the *Krüppel* domain when *Mes4* is misexpressed (compare [G] with [F]). Interestingly, the fact that *snail* expression, which normally extends 18–20 cells in a broad ventral band in wild-type embryos, is reduced in width and exhibits stripes of expression in a *twist* (*twi*) mutant background (I) implies that Mes4 may regulate the expression of *snail* in the absence of Twist.

the use of degenerate, low-affinity Dorsal binding sites.

We investigated the feasibility of this approach by analyzing the previously identified *ventral nervous system defective (vnd)* gene. Recent studies identified multiple enhancers in the 5′ flanking region that mediate *vnd* expression in specific neuroblasts of advanced-stage embryos (Shao et al., 2002). However, these enhancers do not direct lateral stripes of *vnd* expression within the ventral neurogenic ectoderm. The best cluster of potential Dorsal binding sites in the *vnd* genomic interval is located within the first intron (Figure 5A). A ~1.7 kb DNA fragment that encompasses these binding sites was placed 5′ of a minimal *eve-lacZ* fusion gene and expressed in transgenic embryos. The fusion gene exhibits lateral stripes of *lacZ* expression that are virtually identical to the endogenous *vnd* expression pattern (Figures 5B and 5C).

A cluster of four low-affinity Dorsal binding sites were identified in a 260 bp region of the *Mes3* (*dilp4*) 5′ flanking sequence (Figure 5D). The *Mes3* gene is selectively expressed in the ventral mesoderm (Figures 2C and 3C), and the associated putative Dorsal binding sites are similar to those seen in the 5′ flanking regions of the *snail* and *twist* genes (Jiang et al., 1991; Ip et al., 1992a). The 260 bp fragment from *Mes3* was placed 5′ of a minimal *eve-lacZ* fusion gene and expressed in transgenic embryos. *LacZ* staining is detected in ventral regions at the onset of nuclear cleavage cycle 14 (Figure 5E). This staining pattern persists during cellularization, gastrulation, and germband elongation (Figure 5F). The expression profile generated by the 260 bp DNA fragment is similar to that observed for the endogenous *Mes3* gene (e.g., Figure 3C). For example, *lacZ* staining is excluded from the posterior pole as seen for the endogenous pattern (compare Figure 5E with Figure 3C). Thus, it was possible to identify an authentic mesoderm-specific enhancer from the *Mes3* 5′ flanking region by simply identifying the best potential Dorsal binding cluster in the vicinity of the gene.

A new neurogenic enhancer was identified in the 5′ flanking region of the *Neu4* gene, which is expressed in broad lateral stripes that encompass both ventral and dorsal regions of the neurogenic ectoderm in the presumptive thorax and abdomen (Figure 3I). Previous studies identified 16 regions in the entire genome that contain 3 or 4 optimal Dorsal binding sites within a stretch of 400 bp or less (Markstein et al., 2002). The highest density binding cluster is located 15.2 kb 5′ of the *Neu4* gene. There are three evenly spaced binding sites within a stretch of just 75 bp (Figure 5G). A 500 bp fragment that encompasses these sites was placed 5′ of an *eve-lacZ* fusion gene and expressed in transgenic embryos. *LacZ* staining is detected in the neurogenic ectoderm at the onset of nuclear cleavage cycle 14 (Figure 5H). As seen for the endogenous gene, staining is strongest in the presumptive thorax and abdomen (see Figure 3I). This staining pattern persists during gastrulation and germband elongation (Figure 5I). It is possible

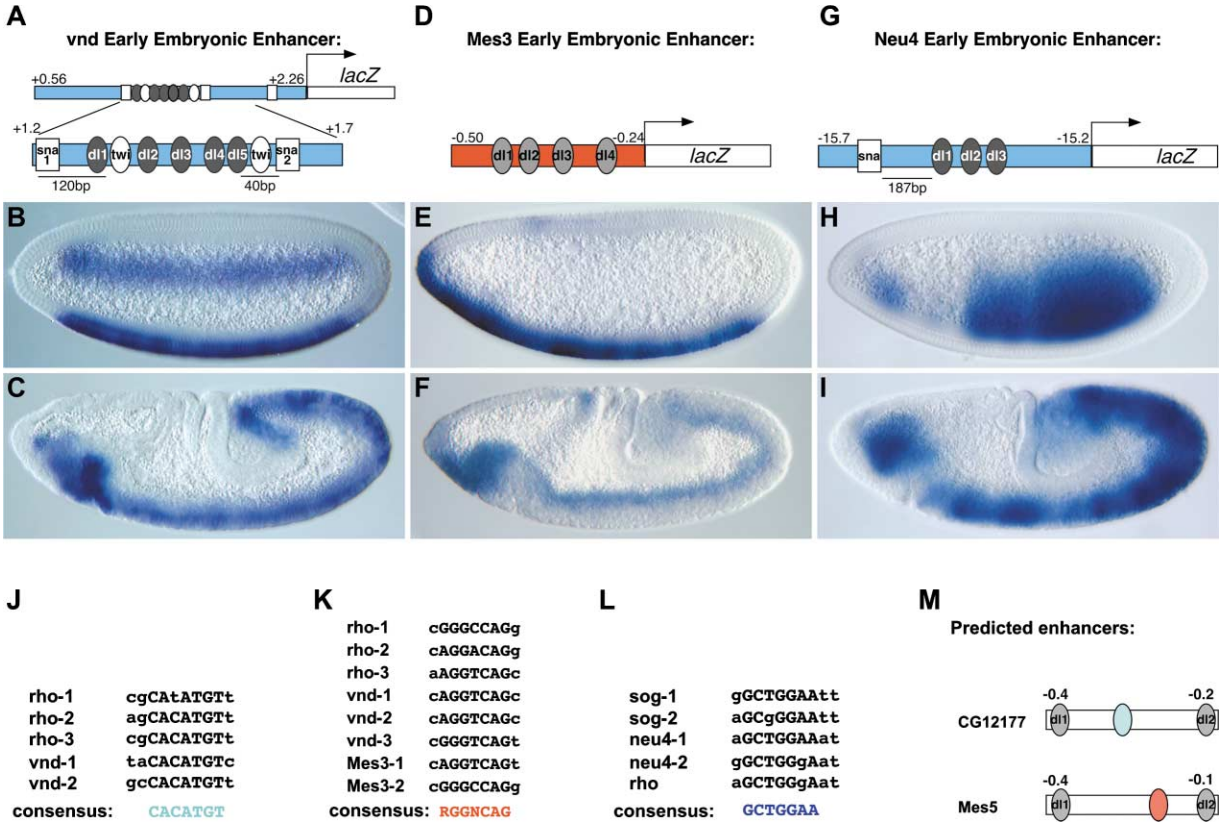

Figure 5. The Identification of New Dorsal Target Enhancers and Novel *cis*-Regulatory Elements

(A, D, and G) Putative enhancers were initially identified by the presence of potential clusters of Dorsal binding sites.

(A–C) The best cluster in the *vnd* genomic region is located in the first intron, between 560 bp and 2.26 kb downstream of the transcription start site (A). The diagram shows the location of putative high-affinity Dorsal binding sites (black ovals), two Twist sites (white ovals), as well as two potential Snail repressor sites (white squares) that might keep the gene off in the ventral mesoderm. This 1.7 kb DNA fragment was placed 5′ of an *eve-lacZ* fusion gene and expressed in transgenic embryos. Embryos were hybridized with a digoxigenin-labeled *lacZ* RNA probe. Staining is initially detected in lateral stripes that straddle the mesoderm (B). These stripes correspond to the ventral half of the neurogenic ectoderm and persist during gastrulation (data not shown) and germband elongation (C).

(D–F) The best Dorsal binding cluster in the *Mes3/dilp4* region is located between 500 bp and 240 bp 5′ of the transcription start site. The diagram shows the location of low-affinity Dorsal binding sites (gray ovals). This 260 bp fragment was placed 5′ of the *eve-lacZ* fusion gene and expressed in transgenic embryos (D). Staining is detected in the presumptive mesoderm during mid-nuclear cleavage cycle 14 (data not shown) and persists during cellularization (E) and gastrulation (F). Staining is restricted to the ventral mesoderm.

(G–I) The best Dorsal binding cluster associated with the *Neu4* gene was identified in a whole-genome search for optimal Dorsal clusters. The highest density cluster (three optimal Dorsal sites in 75 bp; black ovals) is located ~15 kb 5′ of the *Neu4* transcription start site (G). A 500 bp fragment that encompass these sites was placed 5′ of the *eve-lacZ* reporter gene and expressed in transgenic embryos. Staining is detected in broad lateral stripes in cellularizing embryos (H) and persists during gastrulation and germband elongation (I). Staining is excluded from the ventral mesoderm, possibly by binding of the Snail repressor to a site located in the 500 bp *Neu4* 5′ fragment (see diagram in [G]; white square).

(J–M) Comparative analyses of two enhancers from each of these representative thresholds led to the identification of conserved sequence elements. Exact matches to the computed consensus are depicted in uppercase; sequences not conforming to the consensus in lowercase.

(J) CACATGT is present in *rho* and *vnd* but not the other enhancers.

(K) RGGNCAG is present in multiple enhancers (*rho*, *vnd*, and *Mes3*).

(L) Perfect matches to GCTGGAA are present in *sog* and *Neu4*; a divergent copy of this sequence is also present in *rho*.

(M) Using these novel *cis*-regulatory motifs, enhancers were predicted in other putative Dorsal targets that contain only two low-affinity Dorsal sites. These include *CG12177*, which contains low-affinity Dorsal binding sites and one CACATGT site (blue oval), and *Mes5*, which contains low-affinity Dorsal binding sites and one RGGNCAG site (red oval).

that *Neu4* is repressed in the ventral mesoderm by the Snail repressor since the 500 bp enhancer contains at least one optimal Snail binding site (Figure 5G).

Identification of Conserved Sequence Motifs

Bioinformatics methods were used to identify shared sequence motifs among the expanded collection of Dorsal target enhancers (see Experimental Procedures). Particular efforts focused on coordinately regulated enhancers that respond to similar levels of the Dorsal gradient. The newly identified *vnd* enhancer directs an expression pattern that is virtually identical to the one produced by the 300 bp *rhomboid* NEE (Ip et al., 1992b). Both enhancers are activated by intermediate levels of

the Dorsal gradient and direct lateral stripes of expression within ventral regions of the neurogenic ectoderm. In addition to Dorsal binding sites, the two enhancers share an additional sequence motif: CACATGT. There are multiple copies of this motif in each enhancer (Figure 5J). The motif is probably recognized by the bHLH protein Twist, which is distributed in a steep gradient in ventral regions of the neurogenic ectoderm (Kosman et al., 1991). Clustered Dorsal binding sites and the CACATGT motif are observed within the intron of the *Dscam* (*Neu1*) gene, which also exhibits lateral stripes in the ventral neurogenic ectoderm (see Figure 3F).

The newly identified *Neu4* enhancer directs broad lateral stripes of gene expression in response to low levels of the Dorsal gradient. The staining pattern is very similar to the one produced by the 400 bp *sog* intronic enhancer (see Markstein et al., 2002). Both the *Neu4* and *sog* enhancers contain a series of evenly spaced, optimal Dorsal binding sites (GGG-$W_{4\text{-}5}$-CCC, where W = A or T). They also share a novel sequence motif, GCTGGAA (Figure 5L). There are two copies of this motif in each enhancer, but it is generally absent in Dorsal target enhancers that are activated by higher concentrations of the Dorsal gradient (although there is one copy of the motif in the *rhomboid* NEE). It is conceivable that the GCTGGAA motif interacts with an unknown transcription factor, which permits low levels of the Dorsal gradient to activate gene expression in dorsal regions of the neurogenic ectoderm.

A third sequence motif was identified in the *vnd* enhancer, the *rhomboid* NEE, and the *Mes3* enhancer: RGGNCAG (Figure 5K). These enhancers are activated by either high or intermediate levels of the Dorsal gradient. This new motif helps identify putative mesoderm enhancers, which are often activated by degenerate, low-affinity Dorsal binding sites that are difficult to discern. For example, a 300 bp DNA fragment located 110 bp 5′ of the *Mes5* transcription start site contains two putative Dorsal binding sites and one copy of the RGGNCAG motif (Figure 5M). Similarly, a putative enhancer was identified in the 5′ regulatory region of *CG12177* (see Table 1) based on the occurrence of low-affinity Dorsal binding sites and a copy of the CACATGT motif (Figure 5M). It is conceivable that *CG12177* is expressed in both the ventral mesoderm and ventral regions of the neurogenic ectoderm (data not shown).

Discussion

Microarray assays identified a large number of new Dorsal target genes. Approximately two-thirds of the genes that were tested (19/25) exhibit localized patterns of gene expression across the dorsal-ventral axis of wild-type embryos. Moreover, an additional 32 untested genes fulfill the cutoff criteria (Tables 1–3), and ~20 (two-thirds) would be expected to display localized expression. Thus, this study identified as many as 40 new Dorsal target genes. It is likely that at least half correspond to direct targets of the Dorsal gradient since they are activated during early embryogenesis, within an hour of Dorsal nuclear transport.

There have been earlier attempts to identify genes that are specifically expressed in the mesoderm (Casal and Leptin, 1996; Furlong et al., 2001). Several such genes were identified, including *Mes1*. However, the screens were not performed to saturation and many of the genes are activated in mesoderm derivatives after the Dorsal gradient initiates dorsal-ventral patterning. In contrast, the present study focused on early embryos in order to identify direct target genes. In addition, $Toll^{rm9}/Toll^{rm10}$ mutants were used to isolate genes that are selectively expressed in the neurogenic ectoderm, while the *pipe* mutant permitted the identification of genes that are selectively expressed in the dorsal ectoderm.

Mesoderm cells undergo extensive changes in cell shape during gastrulation (reviewed by Leptin and Roth, 1994). Previous studies identified a putative G-coupled signaling pathway that influences changes in cell shape, including *folded gastrulation*, a Rho GTPase (DRhoGEF2), and *concertina*, an α subunit of a heteromeric G protein complex (reviewed by Leptin and Roth, 1994; see Morize et al., 1998; Hacker and Perrimon, 1998). Many of the mesoderm genes that were identified in this screen (Table 1) encode proteins implicated in changes in cell growth and proliferation, including a Rho GTPase (Mes1), an insulin-like growth factor (Mes3), an ABC transporter (Mes5), acyl-CoA synthetase (CG4500), and a nucleoside hydrolase (CG12177). It seems likely that one or more of these genes influence changes in cell shape or size, possibly by interacting with the *fog* signaling pathway.

Evidence that at least one of the new mesoderm target genes has a function in the early embryo was obtained for *Mes4*. It is the first example of a tissue-specific NF-Y subunit in metazoans. It joins a growing list of general transcription factors that have duplicated to produce a tissue-specific variant, which controls the differentiation of specific cell types. Other examples include tissue-specific TAFs, such as *cannonball*, which is a $TAF_{II}80$ derivative that is required for spermatogenesis (Hiller et al., 2001), and $TAF_{II}105$, a $TAF_{II}130$ derivative that is required for the differentiation of follicle cells in the mouse ovary (Freiman et al., 2001). $TAF_{II}80$ and Mes4 contain histone-fold dimerization motifs, which are conserved in a small subset of general transcription factors that function at or near the core promoter (e.g., Aoyagi and Wassarman, 2000).

Ect1 represents the best-conserved TNF homolog in the fly genome. TNFs have been implicated as key mediators of JNK signaling and apoptosis in a variety of mammalian tissues (reviewed by Baud and Karin, 2001). *Ect1* might play a similar role in *Drosophila* since the gene is specifically expressed in the presumptive amnioserosa during gastrulation. The *Drosophila* homologs of *fos* and *jun* exhibit similar patterns of expression (Rusch and Levine, 1997; Riesgo-Escovar and Hafen, 1997), thereby raising the possibility that TNF triggers the histolysis of the amnioserosa by inducing JNK signaling (Baud and Karin, 2001). *Ect1* corresponds to a recently identified gene called *Eiger*, which was shown to trigger cell death upon overexpression in the adult eye (Igaki et al., 2002).

The relatively late onset of *Ect2* expression (see Figure 3L) suggests that it might be regulated by the Dpp activ-

ity gradient present in the dorsal ectoderm of cellularized embryos (reviewed by Podos and Ferguson, 1999). *Ect2* encodes a T domain transcription factor, thereby raising the possibility that there is an evolutionarily conserved link between TGF-β signaling and T box transcription factors. TGF-β gradients regulate T box genes in the *Xenopus* animal cap (reviewed by Gurdon and Bourillot, 2001) and the *Drosophila* wing imaginal disk (reviewed by Strigini and Cohen, 1999). A second T box gene, *CG5093*, exhibits an expression pattern that is virtually identical to the *Ect2* pattern (see Table 3). Potential redundancy in the activities of the *Ect2* and *CG5093* genes might explain why these genes were not identified in previous genetic screens. Similar arguments apply to some of the other genes identified in this screen, such as *Mes1* and *Mes4*. Perhaps mutations in the *Mes1* Rho GTPase are compensated by the maternal expression of the *DRhoGEF2* GTPase (Hacker and Perrimon, 1998), while mutations in *Mes4* are compensated by the general NF-YC subunit encoded by *CG3075*.

The combination of microarray assays and bioinformatics methods provided a highly effective means for identifying new target enhancers and potential *cis*-regulatory elements that respond to different thresholds of the Dorsal gradient. The analysis of coordinately regulated enhancers provided an opportunity to identify shared sequence elements that might help specify different threshold readouts of the Dorsal gradient. The identification of the CACATGT motif in target enhancers that are activated by intermediate levels of the gradient reinforce the view that Twist is not a dedicated mesoderm determinant, but is also essential for specifying the ventral neurogenic ectoderm (see Stathopoulos and Levine, 2002b). The RGGNCAG motif facilitates the identification of mesoderm enhancers, which are generally regulated by poorly conserved, low-affinity Dorsal binding sites. Future studies will determine whether the newly identified GCTGGAA motif is essential for the activation of target enhancers by the lowest levels of the Dorsal gradient.

In terms of sheer number of potential target genes and associated *cis*-regulatory DNAs, Dorsal represents the most thoroughly characterized morphogen in development. Target genes were identified for every dorsal-ventral patterning threshold by analyzing mutant embryos that express different concentrations of the Dorsal protein. In principle, a similar strategy could be used for other patterning processes, such as the specification of different neurons in the vertebrate neural tube by a gradient of Sonic Hedgehog (Jessell, 2000).

Experimental Procedures

Drosophila Stocks and Genetic Crosses

Flies containing a dominant gain-of-function mutation of the maternal gene *Toll*, $Toll^{10B}$, were obtained from S. Govind (Schneider et al., 1991). Flies containing recessive *Toll* mutations, $Toll^{rm9}$ and $Toll^{rm10}$, were obtained from K. Anderson (Schneider et al., 1991). Flies containing recessive *pipe* mutations, $pipe^{386}$ and $pipe^{664}$, were obtained from D. Stein (Sen et al., 1998). A transgenic line homozygous for a P[ry^{+} β2-tubulin-flp] insertion was provided by G. Struhl. $twist^{-}$ flies (*cn twi bw sp/CyO*; #2381) were obtained from the Bloomington Stock Center. Other transgenic lines described in this work were generated using yw^{67c23} flies.

Females of the genotype $Toll^{10B}/+$ were obtained directly from the balanced stock ($Toll^{10b}$*/TM3 Sb Ser* and $Toll^{10b}$*/OR60)*. To generate $Toll^{rm9}/Toll^{rm10}$ females, females of the genotype $Toll^{rm10}$*/TM3 Sb* were mated with males of the genotype $Toll^{rm9}$*/TM3 Ser*. Non-*Sb*, non-*Ser* females of the genotype $Toll^{rm9}/Toll^{rm10}$ were selected. To generate $pipe^{-}/pipe^{-}$ females, $pipe^{664}$*/TM3 Sb* females were mated with males of the genotype $pipe^{386}$*/TM3 Sb*. Non-*Sb* females of the genotype $pipe^{386}/pipe^{664}$ were selected. Embryos were collected from females of the selected genotypes, either $Toll^{10B}/+$, $Toll^{rm9}/Toll^{rm10}$, or $pipe^{386}/pipe^{664}$.

Transgenic Lines

To ectopically express *Mes4* within the *Krüppel* domain, females containing the misexpression construct *Kr-FSF-Mes4* were mated with males carrying the β*2-tubulin-flp* gene to obtain males containing both transgenes. In these males, β*2-tubulin-flp* catalyzes the activation of the misexpression construct by the spermatocyte-specific removal of stop codons contained within a flp-out cassette (FSF). These males were mated to yw^{67c23} females to establish a flipped line, which we determined was viable. Embryos were collected from these homozygosed, flipped lines and analyzed by in situ hybridization (Tautz and Pfeifle, 1989; Jiang et al., 1991). To analyze the *vnd*, *Mes3*, and *Neu4* enhancer reporters, embryos were collected from transformants and analyzed by in situ hybridization using a *lacZ* antisense RNA probe. For the *vnd* enhancer transgenic lines, one transformant was analyzed. For *Kr-FSF-Mes4*, *Mes3*, *Neu4* transgenic lines, at least three transformants were analyzed for each. The staining patterns depicted in figures represents the staining pattern observed for the majority of embryos examined for each line.

Microarray Experiments

Sample Preparation

Two hour embryo collections were made from either $Toll^{10B}$, $Toll^{rm9}/Toll^{rm10}$, or $pipe^{386}/pipe^{664}$ mutant females on apple juice/yeast plates at 25°C. Plates were removed and the embryos were aged an additional 2 hr at 25°C. Carefully staged embryos which had been aged 2–4 hr in this manner were collected, dechorionated, and frozen in liquid nitrogen for storage at −80°C until RNA was to be isolated. Multiple collections from different days were pooled for each sample in order to better normalize the age of these embryo populations. Once sufficient amounts of embryos had been collected (~500 embryos), total RNA was extracted from them using Trizol Reagent (GIBCO-BRL) according to the manufacturer's protocol. For each sample, 100 μg of total RNA was further purified using the RNeasy Mini Kit (Qiagen) following the Rneasy Mini Protocol for RNA cleanup. Total RNA was prepared independently three times from embryos of each genetic background.

Probe Preparation

cDNA synthesis was carried out as described in the Expression Analysis Technical Manual (Affymetrix) using 7 μg of total RNA for each sample. The cRNA reactions were carried out using the BioArray High-Yield Transcript Labeling kit (Enzo). 20 μg of labeled cRNA was fragmented for 35 min at 94°C using fragmentation buffer (200 mM Tris-acetate [pH 8.1], 500 mM KOAc, 150 mM MgOAc). Affymetrix high-density oligonucleotide arrays for *Drosophila melanogaster* were probed, hybridized, stained, and washed according to the manufacturer's protocol. Greater than 13,500 gene sequences predicted from the annotation of the *Drosophila* genome (version 1) are represented on the array.

Data Analysis

Hybridized arrays were scanned using Affymetrix Microarray Suite software as described in the manufacturer's protocol. GeneChip Analysis Suite Software was used to normalize the data contained in each experimental GeneChip.dat file, creating GeneChip.chp files for each experiment. For comparison analysis, GeneChip.dat files were analyzed relative to the GeneChip.chp file defined as baseline. In this way, pair-wise comparisons were made using the GeneChip program between the $Toll^{10B}$ and $pipe^{-}/pipe^{-}$, the $Toll^{10B}$ and $Toll^{rm9}/Toll^{rm10}$, and the $pipe^{-}/pipe^{-}$ and $Toll^{rm9}/Toll^{rm10}$ microarray data. Data output of these comparisons was formatted in Microsoft Excel and then imported into FileMaker Pro for further analysis. Fold-differences are reported in $\log_2$ such that increasing and decreasing levels of RNA can be compared directly. As discussed in Casal and Leptin

(1996), we examined the levels of a gene uniformly expressed in the embryo βTub56D (Natzle and McCarthy, 1984) as a control for our RNA samples. We found that the levels of this transcript changed less than 1-fold for each microarray experimental comparison, as expected for a uniformly expressed gene.

Plasmid Construction, P Element-Mediated Germline Transformation, and Whole-Mount In Situ Hybridization

An ~1.7 kb genomic DNA fragment located 560 bp downstream of the *vnd* start codon was amplified from *Drosophila melanogaster* genomic DNA using polymerase chain reaction (PCR) and the primers 5′-gggtaagcacaaggattccaatg-3′ and 5′-cgaaaagctgcaaggagat caatatg-3′. A 260 bp genomic DNA fragment located 236 bp upstream of the *Mes3* gene start codon was amplified using the primers 5′-cccatagatatatgtaaaagttgttg-3′ and 5′-ggtcagtcaaccaaccaaa cagtc-3′. A 510 bp genomic region approximately 15.2 kb upstream of the *Neu4* gene start codon was amplified using primers 5′-ggac cagcacgagctacgcagcctcac-3′ and 5′-gtggtgaaagttccacctcccttgcg-3′. These 1.7 kb, 260 bp, and 510 bp PCR products were cloned directly into pGEM-T Easy Vector (Promega) using the manufacturer's directions, creating pGEM-vndenh, pGEM-Mes3enh, and pGEM-Neu4enh, respectively. pGEM-vndenh, pGEM-Mes3enh, and pGEM-Neu4enh were digested with EcoRI, and EcoRI fragments containing the respective enhancers were isolated and cloned into the unique EcoRI site of −42evelacZCasper (Small et al., 1992), which places the enhancer upstream of the even-skipped minimal promoter driving lacZ reporter expression, creating vndenh.lacZ-Casper, Mes3enh.lacZCasper, and Neu4enh.lacZCasper.

22FPE (Kosman and Small, 1997) provided by S. Small was modified to promote ectopic expression in the *Krüppel* domain. The *even-skipped (eve)* stripe 2 enhancer was removed from 22FPE by NotI digestion and replaced with a ~1.5 kb NotI fragment containing two copies of the *Krüppel* CD1 enhancer identified by Hoch et al. (1990), thereby creating a *Kr-FSF* ectopic expression construct. The *Mes4* open reading frame was PCR amplified from *Drosophila melanogaster* genomic DNA using primers 5′-gctctagatgcgcaatggccagcgag gaactatttg-3′ and 5′-gctctagatcaaaagttcatggctccatcaaggaat-3′ and cloned into the XbaI site of pBsAscII, a modified form of pBluescript in which the unique SacI and HincII sites have been converted to AscI sites, thus creating pBsAscII-*Mes4*. An AscI fragment containing *Mes4* was isolated by AscI digestion of pBsAscII-*Mes4* and inserted into the unique AscI site of *Kr-FSF*, creating *Kr-FSF-Mes4*.

P element plasmids were introduced into the *Drosophila* germline using standard methods (Spralding and Rubin, 1982). Plasmid DNA was injected into a strain carrying a mutation in *white* (yw^{67}), together with the δ2,3 transposase helper plasmid. Multiple independent transformed lines were examined. Appropriately staged embryos were fixed and hybridized with a digoxigenin-UTP-labeled antisense RNA probes as described previously (Tautz and Pfeifle, 1989; Jiang et al. 1991). Antisense probes to previously characterized Dorsal targets (*snail*, *sog*, and *dpp*) have been described (Huang et al., 1997). Antisense probes to other genes were made using ESTs available from Research Genetics (Huntsville, AL); if no EST was available, primers were designed to amplify by PCR 400–1000 bp fragments of coding sequence from either *Drosophila melanogaster* genomic DNA or cDNA (made by reverse transcription of genomic DNA). These fragments were subsequently cloned into pGEM-T Easy Vector, and the resulting plasmids used to generate antisense probes. More information on the generation of probes for specific genes will be made available upon request.

Computational Identification of Novel Sequence Motifs in Dorsal Target Enhancers

Genomic DNA fragments (1 kb apiece) that encompass the *twist*, *Mes3*, *rho*, *vnd*, *sog*, and *Neu4* enhancers were compared with a 20 kb control sequence located 5′ of the *Abd-B* locus. All possible 7–11 bp sequence motifs were identified using Auilix Biopharma's *cis*-regulatory bioinformatics package, GeneGrokker version 0.29. Motifs were allowed to include up to two "wild-card" positions denoted by the following symbols: R = A/G, Y = C/T, S = C/G, W = A/T, K = G/T, M =A/C, V = A/C/G, H = A/C/T, D = A/G/T, B = C/G/T, N = A/C/G/T. Motifs were identified that are overrepresented in the enhancer fragments as compared with the control *Abd-B* sequence. These motifs include Dorsal and Snail recognition sequences (not shown). Novel motifs were also identified, including CACATGT (Figure 5J), which is likely to correspond to a binding site for the bHLH protein, Twist. Two additional motifs were also identified, RGGNCAG (Figure 5K) and GCTGGAA (Figure 5L). The latter sequence motif resembles a 3′ Dorsal half-site with extended 3′ sequences.

Acknowledgments

We thank M. Blanchette, A. Hochheimer, E. Laborier, D. Rio, and R. Tjian for advice and technical assistance with the microarray experiments, and L. Mirels and R. Zinzen for critically reading the manuscript. We also thank R. Zinzen for help with Figure 5M and K. Senger for sharing SELEX data. P. Markstein independently identified the CACATGT motif using a software package, Mermaid, that is under development. We are grateful to H. Ashe for the Kr-FSF expression vector and probes and Y. Nibu for providing ESTs. This work was funded by a grant from the NIH (GM46638) to M.L. A.S. is supported by a postdoctoral fellowship from the NIH (GM20352).

Received: May 30, 2002
Revised: October 3, 2002

References

Aoyagi, N., and Wassarman, D.V. (2000). Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. J. Cell Biol. *150*, F45–F50.

Baud, V., and Karin, M. (2001). Signal transduction by tumor necrosis factor and its relatives. Trends Cell Biol. *11*, 372–377.

Belvin, M.P., and Anderson, K.V. (1996). A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. Annu. Rev. Cell Dev. Biol. *12*, 393–416.

Bodmer, R., Jan, L.Y., and Jan, Y.N. (1990). A new homeobox-containing gene, msh-2, is transiently expressed early during mesoderm formation of *Drosophila*. Development *110*, 661–669.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. Curr. Biol. *11*, 213–221.

Casal, J., and Leptin, M. (1996). Identification of novel genes in *Drosophila* reveals the complex regulation of early gene activity in the mesoderm. Proc. Natl. Acad. Sci. USA *93*, 10327–10332.

Costa, M., Wilson, E.T., and Wieschaus, E. (1994). A putative cell signal encoded by the folded gastulation gene coordinates cell shape changes during *Drosophila* gastrulation. Cell *76*, 1075–1089.

Courey, A.J., and Huang, J.D. (1995). The establishment and interpretation of transcription factor gradients in the *Drosophila* embryo. Biochim. Biophys. Acta *1261*, 1–18.

Cremazy, F., Berta, P., and Girard, F. (2000). *Sox Neuro*, a new *Drosophila* Sox gene expressed in the developing central nervous system. Mech. Dev. *93*, 215–219.

Doyle, H.J., Kraut, R., and Levine, M. (1989). Spatial regulation of *zerknullt*: a dorsal-ventral patterning gene in *Drosophila*. Genes Dev. *3*, 1518–1533.

Francois, V., Solloway, M., O'Neill, J.W., Emery, J., and Bier, E. (1994). Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the short gastrulation gene. Genes Dev. *8*, 2602–2616.

Frank, L.H., and Rushlow, C. (1996). A group of genes required for maintenance of the amnioserosa tissue in *Drosophila*. Development *122*, 1343–1352.

Freiman, R.N., Albright, S.R., Zheng, S., Sha, W.C., Hammer, R.E., and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. Science *293*, 2084–2087.

Furlong, E.M., Anderson, E.C., Null, B., White, K.P., and Scott, M.P.

(2001). Patterns of gene expression during *Drosophila* mesoderm development. Science *293*, 1629–1633.

Gurdon, J.B., and Bourillot, P.Y. (2001). Morphogen gradient interpretation. Nature *413*, 797–803.

Hacker, U., and Perrimon, N. (1998). DRhoGEF2 encodes a member of the Dbl family of oncogenes and controls cell shape changes during gastrulation in *Drosophila*. Genes Dev. *12*, 274–284.

Hiller, M.A., Lin, T.Y., Wood, C., and Fuller, M.T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. Genes Dev. *15*, 1021–1030.

Hoch, M., Schröder, C., Seifert, E., and Jackle, H. (1990). Cis-acting control elements for Krüppel expression in the *Drosophila* embryo. EMBO J. *8*, 2587–2595.

Huang, A.D., Rusch, J., and Levine, M. (1997). An anterior-posterior Dorsal gradient in the *Drosophila* embryo. Genes Dev. *11*, 1963–1973.

Igaki, T., Kanda, H., Yamamoto-Goto, Y., Kanuka, H., Kuranaga, E., Aigaki, T., and Miura, M. (2002). Eiger, a TNF superfamily ligand that triggers the *Drosophila* JNK pathway. EMBO J. *21*, 3009–3018.

Ingham, P.W., and McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev. *15*, 3059–3087.

Ip, Y.T., Kraut, R., Levine, M., and Rushlow, C.A. (1991). The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. Cell *64*, 439–446.

Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. (1992a). *dorsal-twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. Genes Dev. *6*, 1518–1530.

Ip, Y.T., Park, R.E., Kosman, D., Bier, E., and Levine, M. (1992b). The dorsal gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. Genes Dev. *6*, 1728–1739.

Jazwinska, A., Rushlow, C., and Roth, S. (1999). The role of brinker in mediating the graded response to Dpp in early *Drosophila* embryos. Development *126*, 3323–3334.

Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat. Rev. Genet. *1*, 20–29.

Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant *twist* in early *Drosophila* embryos. Genes Dev. *5*, 1881–1891.

Kasai, Y., Stahl, S., and Crews, S. (1998). Specification of the *Drosophila* CNS midline cell lineage: direct control of *single-minded* transcription by dorsal/ventral patterning genes. Gene Expr. *7*, 171–189.

Kirov, N., Childs, S., O'Connor, M., and Rushlow, C. (1994). The *Drosophila* dorsal morphogen represses the tolloid gene by interacting with a silencer element. Mol. Cell. Biol. *14*, 713–722.

Kosman, D., and Small, S. (1997). Concentration-dependent patterning by an ectopic expression domain of the *Drosophila* gap gene knirps. Development *124*, 1343–1354.

Kosman, D., Ip, Y.T., Levine, M., and Arora, K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. Science *254*, 118–122.

Lai, Z.C., Fortini, M.E., and Rubin, G.M. (1991). The embryonic expression patterns of zfh-1 and zfh-2, two *Drosophila* genes encoding novel zinc-finger homeodomain proteins. Mech. Dev. *34*, 123–134.

Leptin, M., and Roth, S. (1994). Autonomy and non-autonomy in *Drosophila* mesoderm determination and morphogenesis. Development *120*, 853–859.

Lieber, T., Kidd, S., and Young, M.W. (2002). kuzbanian-mediated cleavage of *Drosophila* Notch. Genes Dev. *16*, 209–221.

Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. Gene *239*, 15–27.

Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA *99*, 763–768.

Martin, G. (2001). Making a vertebrate limb: new players enter from the wings. Bioessays *23*, 865–868.

McDonald, J.A., Holbrook, S., Isshiki, T., Weiss, J., Doe, C.Q., and Mellerick, D.M. (1998). Dorsoventral patterning in the *Drosophila* central nervous system: the vnd homeobox gene specifies ventral column identity. Genes Dev. *12*, 3603–3612.

Mlodzik, M., Baker, N.E., and Rubin, G.M. (1990). Isolation and expression of *scabrous*, a gene regulating neurogenesis in *Drosophila*. Genes Dev. *4*, 1848–1861.

Morisato, D. (2001). Spatzle regulates the shape of the Dorsal gradient in the *Drosophila* embryo. Development *128*, 2309–2319.

Morize, P., Christiansen, A.E., Costa, M., Parks, S., and Wieschaus, E. (1998). Hyperactivation of the folded gastrulation pathway induces specific cell shape changes. Development *125*, 589–597.

Natzle, J.E., and McCarthy, B.J. (1984). Regulation of *Drosophila* alpha- and beta-tubulin genes during development. Dev. Biol. *104*, 187–198.

Podos, S.D., and Ferguson, E.L. (1999). Morphogen gradients: new insights from DPP. Trends Genet. *15*, 396–402.

Qi, H., Rand, M.D., Wu, X., Sestan, N., Wang, W., Rakic, P., Xu, T., and Artavanis-Tsakonas, S. (1999). Processing of the Notch ligand Delta by the metalloprotease Kuzbanian. Science *283*, 91–94.

Riesgo-Escovar, J.R., and Hafen, E. (1997). Common and distinct roles of Dfos and Djun during *Drosophila* development. Science *278*, 669–672.

Romani, S., Campuzano, S., Macagno, E.R., and Modolell, J. (1987). Expression of achaete and scute genes in *Drosophila* imaginal discs and their function in sensory organ development. Genes Dev. *3*, 997–1007.

Roth, S. (1994). Axis determination. Proteolytic generation of a morphogen. Curr. Biol. *4*, 755–757.

Rusch, J., and Levine, M. (1996). Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. Curr. Opin. Genet. Dev. *6*, 416–423.

Rusch, J., and Levine, M. (1997). Regulation of a dpp target gene in the *Drosophila* embryo. Development *124*, 303–311.

Schneider, D.S., Hudson, K.L., Lin, T.Y., and Anderson, K.V. (1991). Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo. Genes Dev. *5*, 797–807.

Sen, J., Goltz, J.S., Stevens, L., and Stein, D. (1998). Spatially restricted expression of pipe in the *Drosophila* egg chamber defines embryonic dorsal-ventral polarity. Cell *95*, 471–481.

Shao, X., Koizumi, K., Nosworthy, N., Tan, D.P., Odenwald, W., and Nirenberg, M. (2002). Regulatory DNA required for vnd/NK-2 homeobox gene expression pattern in neuroblasts. Proc. Natl. Acad. Sci. USA *99*, 113–117.

Shishido, E., Higashijima, S., Emori, Y., and Saigo, K. (1993). Two FGF-receptor homologues of *Drosophila*: one is expressed in mesodermal primordium in early embryos. Development *117*, 751–761.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the *Drosophila* embryo. EMBO J. *11*, 4047–4057.

Spralding, A.C., and Rubin, G.M. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348–353.

St. Johnston, R.D., and Gelbart, W.M. (1987). Decapentaplegic transcripts are localized along the dorsal-ventral axis of the *Drosophila* embryo. EMBO J. *6*, 2785–2791.

St. Johnston, D., and Nusslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. Cell *68*, 201–219.

Stathopoulos, A., and Levine, M. (2002a). Dorsal gradient networks in the *Drosophila* embryo. Dev. Biol. *246*, 57–62.

Stathopoulos, A., and Levine, M. (2002b). Linear signaling in the Toll-Dorsal pathway of *Drosophila*: activated Pelle kinase specifies all threshold outputs of gene expression while the bHLH protein Twist specifies a subset. Development *129*, 3411–3419.

Steward, R. (1987). Dorsal, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, c-rel. Science *238*, 692–694.

Strigini, M., and Cohen, S.M. (1999). Formation of morphogen gradients in the *Drosophila* wing. Semin. Cell Dev. Biol. *10*, 335–344.

Szymanski, P., and Levine, M. (1995). Multiple modes of dorsal-bHLH synergy in the *Drosophila* embryo. EMBO J. *14*, 2229–2238.

Tatei, K., Cai, H., Ip, Y.T., and Levine, M. (1995). Race: a *Drosophila* homologue of angiotensin converting enzyme. Mech. Dev. *51*, 157–168.

Tautz, D., and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. Chromosoma *98*, 81–85.

Taylor, M.V., Beatty, K.E., Hunter, H.K., and Baylies, M.K. (1995). *Drosophila* MEF2 is regulated by twist and is expressed in both the primordia and differentiated cells of the embryonic somatic, visceral and heart musculature. Mech. Dev. *50*, 29–41.

Van Buskirk, C., and Schupbach, T. (1999). Versatility in signalling: multiple responses to EGF receptor activation during *Drosophila* oogenesis. Trends Cell Biol. *9*, 1–4.

Weiss, J.B., Von Ohlen, T., Mellerick, D.M., Dressler, G., Doe, C.Q., and Scott, M.P. (1998). Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. Genes Dev. *12*, 3591–3602.

Winick, J., Abel, T., Leonard, M.W., Michelson, A.M., Chardon-Loriaux, I., Holmgren, R.A., Maniatis, T., and Engel, J.D. (1993). A GATA family transcription factor is expressed along the embryonic dorsoventral axis in *Drosophila melanogaster*. Development *119*, 1055–1065.

Accession Numbers

The array data have been deposited in the Gene Expression Omnibus at NCBI (GEO: http://www.ncbi.nlm.nih.gov/geo) as series GSE86: GSM2452 (*Toll*$^{rm9/rm10}$ versus *Toll*^{10b}), GSM2453 (*pipe* versus *Toll*^{10b}), and GSM2454 (*pipe* versus *Toll*$^{rm9/rm10}$).

The *vnd*, *Mes3*, and *Neu4* embryonic enhancer DNA sequences have been deposited into GenBank with accession numbers BK000635, BK000634, and BK000636, respectively.