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 Levine1,4

1Department of Molecular and Cell Biology
Division of Genetics and Development
401 Barker Hall
University of California, Berkeley
Berkeley, California 94720
2Auilix Biopharma, Inc.
950 Gilman Street
Berkeley, California 94710
3Committee 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 thresholds. 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 rhomboi), 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...
Figure 1. Summary of Microarray Experiments

(A) RNA was isolated from three different genetic backgrounds containing varying amounts of nuclear Dorsal protein. Toll10B is a dominant mutation in the Toll receptor that leads to high levels of nuclear Dorsal. Tollrm9/Tollrm10 is a recessive mutation in the Toll receptor that leads to low levels of nuclear Dorsal throughout the early embryo. In pipe (pipe/H11002) 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 Toll10B 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 Tollrm9/Tollrm10 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 in Toll10B relative to Tollrm9/Tollrm10 (369) and pipe (1116) but changed less than 3-fold in a comparison of Tollrm9/Tollrm10 versus pipe (12384). Genes responding to intermediate levels of nuclear Dorsal (141) were identified by determining which signals increased at least 3-fold in Tollrm9/Tollrm10 relative to Toll10B (578) and pipe (1280) but changed less than 3-fold in Toll10B versus pipe (12564). Genes repressed by Dorsal (123) were identified by determining which signals increased at least 3-fold in pipe relative to Toll10B (409) and Tollrm9/Tollrm10 (425) but changed less than 3-fold in Toll10B 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 (Toll10B) or low (Tollrm9/Tollrm10) 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 Toll9 mutations, 141 genes are upregulated in Toll9/Toll10 mutants, and 123 genes are upregulated in pipe /pipe mutants.

Validation of the Microarray Screens
Representative genes that exhibit upregulation in Toll9 mutations, Toll9/Toll10, 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 Toll9 mutants as compared with the other genetic backgrounds. Only 19 of the 89 genes that are significantly upregulated (greater than 3-fold) in Toll9 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 Toll9 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 Toll9 mutations, Toll9/Toll10 mutant embryos as compared with the other genetic backgrounds. Only 22 of the 114 genes that are upregulated in Toll9/Toll10 mutant embryos fulfilled this cutoff value (Table 2). One of these genes, Neu3, encodes a novel member of the ADAM family of endoproteases (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 Toll9/Toll10 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 Toll9 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 Toll9/Toll10 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 Toll9/Toll10 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 difference in Toll9/Toll10 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

<table>
<thead>
<tr>
<th>Known Targets</th>
<th>Mesoderm Function/Homology</th>
<th>10B/[rm9/10]</th>
<th>10B/pipe</th>
<th>pipe/[rm9/10]</th>
<th>Confirmed by In Situ</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>fog</td>
<td>CG9559 unknown</td>
<td>1.1</td>
<td>1.1</td>
<td>-1.0</td>
<td></td>
<td>Costa et al., 1994</td>
</tr>
<tr>
<td>heartless (htl)</td>
<td>CG7223 FGF receptor</td>
<td>14.9</td>
<td>7.7</td>
<td>1.4</td>
<td></td>
<td>Shisido et al., 1993</td>
</tr>
<tr>
<td>mef2</td>
<td>CG1429 MAD6-box TF</td>
<td>9.5</td>
<td>7.1</td>
<td>-1.5</td>
<td></td>
<td>Taylor et al., 1995</td>
</tr>
<tr>
<td>snail (sna)</td>
<td>CG3956 transcriptional repressor</td>
<td>3.1</td>
<td>14.2</td>
<td>-5.0</td>
<td></td>
<td>Ip et al., 1992a</td>
</tr>
<tr>
<td>tinman (tin)</td>
<td>CG7895 homeobox TF</td>
<td>10.2</td>
<td>7.0</td>
<td>-2.3</td>
<td></td>
<td>Bodmer et al., 1990</td>
</tr>
<tr>
<td>twist (twi)</td>
<td>CG2956 bHLH TF</td>
<td>63.8</td>
<td>31.3</td>
<td>-1.2</td>
<td></td>
<td>Jiang et al., 1991</td>
</tr>
<tr>
<td>zfh1</td>
<td>CG1322 Zn-finger/homeobox TF</td>
<td>3.6</td>
<td>1.8</td>
<td>2.2</td>
<td></td>
<td>Lai et al., 1991</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Microarray-Identified Targets</th>
<th>Mesoderm Function/Homology</th>
<th>10B/[rm9/10]</th>
<th>10B/pipe</th>
<th>pipe/[rm9/10]</th>
<th>Confirmed by In Situ</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RhoL</td>
<td>CG9366 Rho GTPase</td>
<td>11.9</td>
<td>7.1</td>
<td>-1.7</td>
<td>A, B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>ZPW domain</td>
<td>CG11100 Rho GTPase</td>
<td>11.0</td>
<td>6.5</td>
<td>-1.7</td>
<td>A</td>
<td>Brogiolo et al., 2001</td>
</tr>
<tr>
<td>insulin receptor ligand</td>
<td>CG6736 insulin receptor ligand</td>
<td>38.0</td>
<td>38.0</td>
<td>-2.1</td>
<td>A, B</td>
<td>Brogiolo et al., 2001</td>
</tr>
<tr>
<td>NF-YC homolog</td>
<td>CG11301 NF-YC homolog</td>
<td>10.7</td>
<td>5.0</td>
<td>1.4</td>
<td>A</td>
<td>Brogiolo et al., 2001</td>
</tr>
<tr>
<td>ABC transporter</td>
<td>CG3879 ABC transporter</td>
<td>11.6</td>
<td>24.1</td>
<td>-7.5</td>
<td>A</td>
<td>Brogiolo et al., 2001</td>
</tr>
<tr>
<td>cytoskeleton</td>
<td>CG10067 cytoskeleton</td>
<td>37.0</td>
<td>17.9</td>
<td>2.2</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>arginine kinase</td>
<td>CG4929 arginine kinase</td>
<td>9.8</td>
<td>6.9</td>
<td>-3.5</td>
<td>A</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>aspartyl β-hydroxylase</td>
<td>CG18658 aspartyl β-hydroxylase</td>
<td>5.1</td>
<td>11.6</td>
<td>-2.0</td>
<td>A</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>cytchrome P450</td>
<td>CG10391 cytchrome P450</td>
<td>28.9</td>
<td>7.5</td>
<td>2.1</td>
<td>A</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>gcm TF</td>
<td>CG12245 gcm TF</td>
<td>11.7</td>
<td>5.3</td>
<td>1.7</td>
<td>A</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>snRNA</td>
<td>CG3949 snRNA</td>
<td>9.7</td>
<td>3.7</td>
<td>2.8</td>
<td>B</td>
<td>Doyle et al., 1989</td>
</tr>
<tr>
<td>xanthine dehydrogenase</td>
<td>CG7642 xanthine dehydrogenase</td>
<td>13.6</td>
<td>6.8</td>
<td>1.3</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>cell signaling</td>
<td>CG18485 cell signaling</td>
<td>18.9</td>
<td>9.4</td>
<td>1.0</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>HLH/PAS TF</td>
<td>CG6883 HLH/PAS TF</td>
<td>10.5</td>
<td>4.7</td>
<td>-2.0</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>acyl-CoA synthetase</td>
<td>CG4500 acyl-CoA synthetase</td>
<td>15.0</td>
<td>5.9</td>
<td>2.6</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>Dnase</td>
<td>CG7780 Dnase</td>
<td>9.7</td>
<td>5.4</td>
<td>1.0</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>nucleoside hydrolase</td>
<td>CG12177 nucleoside hydrolase</td>
<td>8.3</td>
<td>13.5</td>
<td>1.3</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>unknown (Xenopus.hom)</td>
<td>CG14036 unknown (Xenopus.hom)</td>
<td>14.8</td>
<td>6.7</td>
<td>5.2</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
<tr>
<td>unknown (worm.hom)</td>
<td>CG14688 unknown (worm.hom)</td>
<td>25.2</td>
<td>5.7</td>
<td>3.2</td>
<td>B</td>
<td>Casal and Leptin, 1996</td>
</tr>
</tbody>
</table>

TF = transcription factor
A, predicted expression pattern confirmed (this study).
B, predicted expression pattern confirmed (referenced study).
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 TFIIID 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, TAFII80, 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 germ band 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 germ band elongation (Figures 4C and 4D). At the time of invagination, Mes4 transcripts exhibit clear segmental repeats and the staining

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 Toll10B embryos (B). snail is normally expressed in the ventral mesoderm of wild-type embryos (A) but is ubiquitously expressed in Toll10B mutants (B).

(C and D) Mes3 expression in wild-type (C) and Toll10B 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 Tollrm9/Tollrm10 (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 Tollrm9/Tollrm10 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 /H11002 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.
Table 2. Dorsal Targets in the Neuroectoderm

<table>
<thead>
<tr>
<th>Neuroectoderm Function/Homology</th>
<th>[m9/10]/10B</th>
<th>pipe/10B</th>
<th>[m9/10]/pipe</th>
<th>Confirmed by In Situ</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Targets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brinkerm (brk)</td>
<td>C9653</td>
<td>transcriptional repressor</td>
<td>4.9</td>
<td>-1.2</td>
<td>5.7</td>
</tr>
<tr>
<td>ind</td>
<td>C911551</td>
<td>homeobox TF</td>
<td>17.8</td>
<td>-3.9</td>
<td>17.6</td>
</tr>
<tr>
<td>lethal of scute (T3)</td>
<td>C93839</td>
<td>HLH TF</td>
<td>4.2</td>
<td>1.4</td>
<td>3.0</td>
</tr>
<tr>
<td>rhomboid (rho)</td>
<td>C1004</td>
<td>serine-type peptidase</td>
<td>2.9</td>
<td>-1.2</td>
<td>1.9</td>
</tr>
<tr>
<td>single-minded (sim)</td>
<td>C7771</td>
<td>HLH/PAS TF</td>
<td>-4.2</td>
<td>-2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>sog</td>
<td>C9224</td>
<td>growth factor</td>
<td>6.4</td>
<td>-1.5</td>
<td>5.1</td>
</tr>
<tr>
<td>vnd</td>
<td>C6172</td>
<td>homeobox TF</td>
<td>4.0</td>
<td>-2.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Microarray-Identified Targets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neu1 = Dscam</td>
<td>C17800</td>
<td>axon guidance receptor</td>
<td>5.5</td>
<td>1.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Neu2</td>
<td>C7204</td>
<td>zinc-finger protein</td>
<td>5.1</td>
<td>-1.1</td>
<td>5.7</td>
</tr>
<tr>
<td>Neu3</td>
<td>C7649</td>
<td>ADAM metalloprotease</td>
<td>15.4</td>
<td>-1.1</td>
<td>8.9</td>
</tr>
<tr>
<td>Neu4</td>
<td>C12443</td>
<td>unknown function</td>
<td>11.1</td>
<td>-1.1</td>
<td>6.3</td>
</tr>
<tr>
<td>Neu5 = sulfated</td>
<td>C66725</td>
<td>sulfotransferase</td>
<td>2.2</td>
<td>-2.3</td>
<td>3.7</td>
</tr>
<tr>
<td>echinoid</td>
<td>C12676</td>
<td>cell adhesion</td>
<td>7.0</td>
<td>-2.7</td>
<td>5.3</td>
</tr>
<tr>
<td>leak</td>
<td>C6574</td>
<td>axon guidance receptor</td>
<td>6.9</td>
<td>-2.8</td>
<td>7.7</td>
</tr>
<tr>
<td>patched</td>
<td>C2411</td>
<td>receptor</td>
<td>11.9</td>
<td>-2.9</td>
<td>8.1</td>
</tr>
<tr>
<td>PGRP-SC2</td>
<td>C147579</td>
<td>defense/immunity</td>
<td>6.2</td>
<td>-2.0</td>
<td>5.8</td>
</tr>
<tr>
<td>Scabrous</td>
<td>C17579</td>
<td>receptor ligand</td>
<td>19.7</td>
<td>3.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Socs36E</td>
<td>C15154</td>
<td>cell signaling</td>
<td>13.5</td>
<td>-2.1</td>
<td>6.7</td>
</tr>
<tr>
<td>SoxNeuro</td>
<td>C18024</td>
<td>HMG-box TF</td>
<td>18.0</td>
<td>-3.1</td>
<td>11.2</td>
</tr>
<tr>
<td>warts</td>
<td>C12072</td>
<td>kinase</td>
<td>4.4</td>
<td>-2.7</td>
<td>6.0</td>
</tr>
<tr>
<td>C4589</td>
<td>calcitonin</td>
<td>binding</td>
<td>5.6</td>
<td>-1.8</td>
<td>6.5</td>
</tr>
<tr>
<td>C4749</td>
<td>methyltransferase</td>
<td>12.0</td>
<td>-2.1</td>
<td>7.4</td>
<td>NL</td>
</tr>
<tr>
<td>C2264</td>
<td>serine protease inhibitor</td>
<td>8.2</td>
<td>-1.8</td>
<td>5.3</td>
<td>NL</td>
</tr>
<tr>
<td>C5891</td>
<td>myosin binding</td>
<td>6.8</td>
<td>-2.4</td>
<td>5.2</td>
<td>NL</td>
</tr>
<tr>
<td>C6400</td>
<td>WD40 repeats</td>
<td>7.8</td>
<td>-1.9</td>
<td>5.0</td>
<td>NL</td>
</tr>
<tr>
<td>C12467</td>
<td>myosin/kinesin motor</td>
<td>7.8</td>
<td>-1.8</td>
<td>5.9</td>
<td>NL</td>
</tr>
<tr>
<td>C13797</td>
<td>unknown (embryocEST)</td>
<td>8.2</td>
<td>-2.3</td>
<td>6.7</td>
<td>NL</td>
</tr>
<tr>
<td>C14411</td>
<td>unknown (homologs)</td>
<td>7.1</td>
<td>-2.4</td>
<td>5.2</td>
<td>NL</td>
</tr>
<tr>
<td>C17229</td>
<td>G-protein signaling</td>
<td>7.6</td>
<td>-1.9</td>
<td>5.1</td>
<td>NL</td>
</tr>
<tr>
<td>C18553</td>
<td>zinc-finger protein</td>
<td>6.6</td>
<td>-2.6</td>
<td>5.3</td>
<td>NL</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dorsal Ectoderm</th>
<th>Function/Homology</th>
<th>[rm9/10]/10B</th>
<th>pipe/10B</th>
<th>pipe/[rm9/10]</th>
<th>Confirmed by In Situ</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Targets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>decapentaplegic (dpp)</td>
<td>CG9885</td>
<td>TGF-β receptor ligand</td>
<td>2.4</td>
<td>5.1</td>
<td>3.6</td>
<td>St. Johnston and Gelbart, 1987</td>
</tr>
<tr>
<td>pannier (pnr)</td>
<td>CG3978</td>
<td>GATA TF</td>
<td>1.4</td>
<td>6.4</td>
<td>6.3</td>
<td>Winick et al., 1993</td>
</tr>
<tr>
<td>race</td>
<td>CG8827</td>
<td>peptidyl dipeptidase A</td>
<td>−1.7</td>
<td>3.3</td>
<td>5.4</td>
<td>Tatei et al., 1995</td>
</tr>
<tr>
<td>tolloid (tld)</td>
<td>CG6868</td>
<td>endopeptidase</td>
<td>7.1</td>
<td>11.4</td>
<td>4.4</td>
<td>Kirov et al., 1994</td>
</tr>
<tr>
<td>tailup (tup)</td>
<td>CG10619</td>
<td>homeobox TF</td>
<td>1.3</td>
<td>3.2</td>
<td>3.2</td>
<td>Frank and Rushlow, 1996</td>
</tr>
<tr>
<td>u-shaped (ush)</td>
<td>CG2762</td>
<td>Zn-finger TF</td>
<td>1.2</td>
<td>6.0</td>
<td>5.2</td>
<td>Frank and Rushlow, 1996</td>
</tr>
<tr>
<td>zerknüllt (zen)</td>
<td>CG1046</td>
<td>homeobox TF</td>
<td>−1.0</td>
<td>21.9</td>
<td>22.7</td>
<td>Doyle et al., 1989</td>
</tr>
<tr>
<td>Microarray-Identified Targets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ect1</td>
<td>CG12919</td>
<td>TNF receptor ligand</td>
<td>2.1</td>
<td>11.9</td>
<td>11.4</td>
<td>A</td>
</tr>
<tr>
<td>Ect2 = Dorsocross</td>
<td>CG5133</td>
<td>T domain TF</td>
<td>1.1</td>
<td>28.6</td>
<td>32.5</td>
<td>A</td>
</tr>
<tr>
<td>Ect3</td>
<td>CG3132</td>
<td>galactosidase</td>
<td>−1.4</td>
<td>15.5</td>
<td>20.9</td>
<td>A</td>
</tr>
<tr>
<td>Ect4</td>
<td>CG7915</td>
<td>SAM,TIR, Pro-rich domains</td>
<td>−1.0</td>
<td>6.4</td>
<td>6.6</td>
<td>A</td>
</tr>
<tr>
<td>Ect5 = C15</td>
<td>CG7937</td>
<td>homeobox TF</td>
<td>1.6</td>
<td>6.8</td>
<td>5.2</td>
<td>A</td>
</tr>
<tr>
<td>Adult cuticle protein1</td>
<td>CG7216</td>
<td>structural cuticle protein</td>
<td>2.2</td>
<td>7.3</td>
<td>6.0</td>
<td>A</td>
</tr>
<tr>
<td>Lcp65Ag2</td>
<td>CG10534</td>
<td>structural cuticle protein</td>
<td>−3.1</td>
<td>13.3</td>
<td>19.9</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>CG5093</td>
<td>T domain TF</td>
<td>−1.0</td>
<td>4.9</td>
<td>5.1</td>
<td>below cutoff</td>
</tr>
<tr>
<td>CGS468(=CGS476)</td>
<td>unknown (embryo EST)</td>
<td>1.6</td>
<td>17.2</td>
<td>15.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGS476(=CGS468)</td>
<td>unknown (embryo EST)</td>
<td>1.6</td>
<td>11.4</td>
<td>11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG6180</td>
<td>phospholipid binding</td>
<td>−1.4</td>
<td>6.1</td>
<td>8.6</td>
<td>NL</td>
<td></td>
</tr>
<tr>
<td>CG9358</td>
<td>ligand binding/carrier</td>
<td>1.9</td>
<td>6.0</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG9466</td>
<td>alpha-mannosidase</td>
<td>2.6</td>
<td>6.4</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG11776</td>
<td>bipartite nls</td>
<td>−1.6</td>
<td>6.1</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG12011</td>
<td>unknown (embryo EST)</td>
<td>1.2</td>
<td>10.2</td>
<td>6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG14365</td>
<td>bipartite nls; Pro-rich</td>
<td>1.6</td>
<td>6.5</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG15282</td>
<td>unknown (embryo EST)</td>
<td>1.8</td>
<td>7.6</td>
<td>7.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.


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 ventral 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
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 (dip4) 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 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 to generate transgenic lines that misexpress 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 (dip4) 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 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), and 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).

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

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).
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 3P).

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-W-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, Tollmut/Ompat mutants were used to isolate genes that are specifically 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 TAF80 derivative that is required for spermatogenesis (Hiller et al., 2001), and TAF105, a TAF130 derivative that is required for the differentiation of follicle cells in the mouse ovary (Freiman et al., 2001). TAF80 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 amnioblasts 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 amnioblasts 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 RGCNCAG 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 recessive *Toll* mutations, *Toll* 

, were obtained from S. Govind (Schneider et al., 1991). Files containing recessive *Toll* mutations, *Toll* 

, were obtained from K. Anderson (Schneider et al., 1991). Files containing recessive *pipe* mutations, *pipe* 

, and *pipe* 

, were obtained from D. Stein (Sen et al., 1998). A transgenic line homozygous for a *Pheat* ( *pipe-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* 

 flies.

Females of the genotype *Toll* 

 were obtained directly from the balanced stock (*Toll* 

/TM3 Sb Ser and *Toll* 

/Or60). To generate *Toll* 

/Toll* 

 females, females of the genotype *Toll* 

/TM3 Sb were mated with males of the genotype *Toll* 

/TM3 Ser. Non-Sb, non-Ser females of the genotype *Toll* 

 were selected. To generate *pipe* / *pipe* females, *pipe* 

/TM3 Sb females were mated with males of the genotype *pipe* 

/TM3 Sb. Non-Sb females of the genotype *pipe* 

 were selected. Embryos were collected from females of the selected genotypes, either *Toll* 

, *Toll* 

, or *pipe* 

.

**Transgenic Lines**

To ectopically express *Mes4* within the Kruppel domain, females containing the misexpression construct *Kr-FSF-Mes4* were mated with males carrying the *i2-tubulin-flp* gene to obtain males containing both transgenes. In these males, *i2-tubulin-flp* catalyzes the activation of the misexpression construct by the spermatocyte-specific removal of stop codons contained within a flip-out cassette (FSF). These males were mated to *yw* 

 females to establish a fliped 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 transfectants 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* 

, *Toll* 

, *Toll* 

, or *pipe* 

 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.chp 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* 

 and *pipe*, the *Toll* 

 and *Toll* 

, and the *pipe* / *pipe* and *Toll* 

/Toll* 

 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 log2, 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. Jacobs 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'-ggtagggcacaaggttagcagt-3' and 5'-gacctgctacgccaacacc-3'. A 260 bp genomic DNA fragment located 236 bp upstream of the Mes3 gene start codon was amplified using the primers 5'-cccatatatgatattggtggt-3' and 5'-gctgctacgccaacacctcagt-3'. A 510 bp genomic region approximately 15.2 kb upstream of the Neuf4 gene start codon was amplified using primers 5'-ggacgcagagctagcatagcagctc-3' and 5'-gtgtgaaagttccactctcctgttggg-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-Neuf4enh, respectively. pGEM-vndenh, pGEM-Mes3enh, and pGEM-Neuf4enh were digested with EcoRI, and EcoRI fragments containing the respective enhancers were isolated and cloned into the unique EcoRI site of pBSAscII (Small et al., 1990), which places the enhancer upstream of the even-skipped minimal promoter driving lacZ reporter expression, creating vndenh, Mes3enh, and Neuf4enh, respectively.

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 Not1 digestion and replaced with a ~1.5 kb Not1 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'-gctcagtagggccagggcgacgaattttg-3' and 5'-gttagctgaaagttccactctcctg-3' and cloned into the unique XbaI site of pBSAscII, a modified form of pBluescript in which the unique Sacl and HincII sites have been converted to formation of a 3' Dorsal half-site with extended 3' sequences.

**References**


**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 (*Tollrm9/rm10* versus *Toll10b*), GSM2453 (*pipe* versus *Toll10b*), and GSM2454 (*pipe* versus *Tollrm9/rm10*).

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