Supplemental Research Data

The ChIP-chip binding data of Dorsal, Twist and Snail can be downloaded from our web site at http://web.wi.mit.edu/young/dorsal/
The raw data files can be downloaded from ArrayExpress.

Table of contents:

Supplemental Information

How were putative enhancers of the Dorsal network identified?

While the binding profiles of Twist and Snail are overall extremely similar, Dorsal was found to bind to many regions that were not bound by Twist and Snail. These regions were preferentially found at promoters and were not enriched for Dorsal binding motifs (see below). Since the significance of these Dorsal-bound regions was not clear, we focused our efforts on identifying putative enhancers of the Dorsal network based on the binding patterns at known enhancers.

Known enhancers are typically bound by Dorsal, Twist and Snail, with the levels of Dorsal being comparatively low. We therefore first identified the set of regions bound by Twist and Snail, and subsequently determined whether Dorsal was also bound (based on a 2-fold cutoff).

How was the cut-off for binding determined?

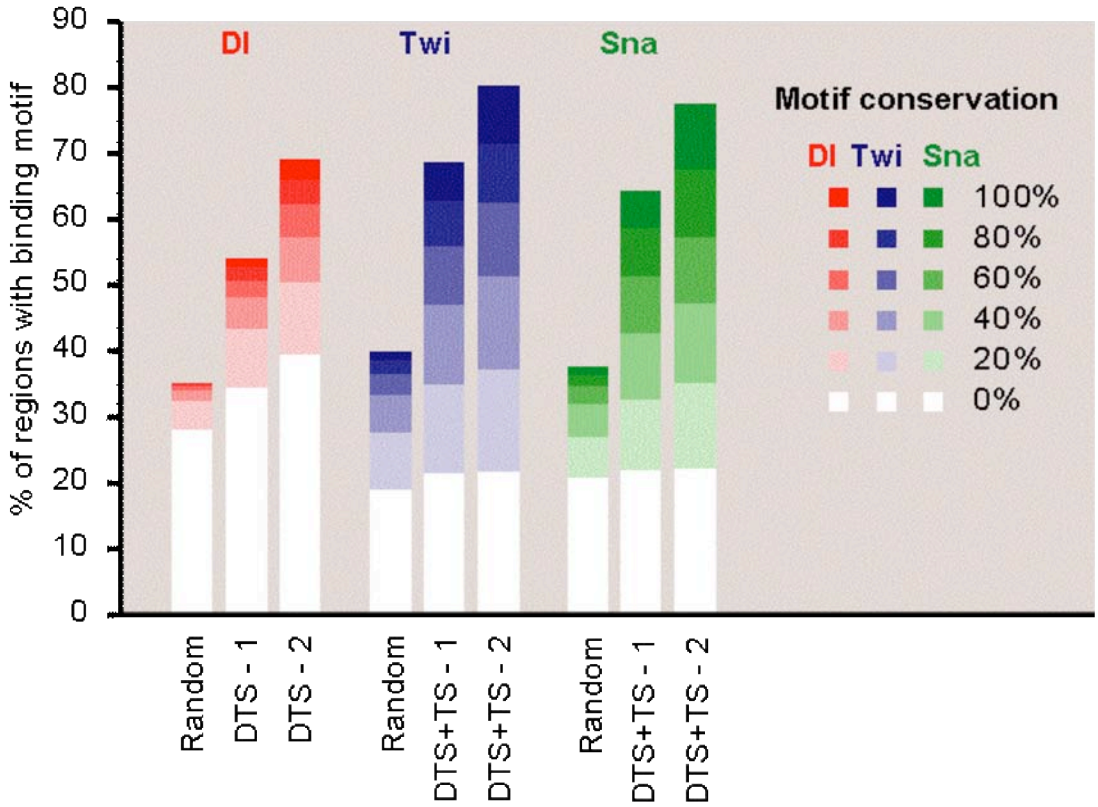
We first used a standard algorithm for identifying bound regions as described in Pokholok et al. 2005. This algorithm identifies bound regions with an estimated 1% of false positives based on known target genes in yeast (Pokholok et al. 2005) or 4% based on the verification of bound regions in human cells by gene-specific PCR (Lee et al. 2006a). Using this algorithm, we obtained 3000 regions that are bound by both Twist and Snail, 995 of which are also enriched in Dorsal IPs by at least 2-fold. In this data set, 20 of the known 23 DV enhancers were present. These data can be downloaded from our web site. Subsequent analysis, however, suggested that a more stringent cut-off would provide a better focus on the functionally relevant target genes.

We noticed that most of the known enhancers that were identified (17/20) showed strong enrichment of >5-fold in Twist and/or Snail IPs. This cut-off reduces the dataset from 3000 to 861. Thus, although the 5-fold cut-off increases false negative binding events, it also appears to reduce the number of false positives, i.e. there is a larger fraction of genes that are functionally relevant in DV axis specification.

When assaying the bound regions for Dorsal, Twist and Snail sequence motifs, we found that a larger fraction of these motifs are enriched in the more stringently defined dataset, as compared with lower cut-off values. It is feasible that many of the sequences discarded by the higher cut-off contain low-affinity binding sites or are bound by Dorsal, Twist and Snail through other transcription factors. Thus, Dorsal, Twist and Snail might not be the primary regulators of these regions.

The putative target genes with known DV-modulated expression patterns and those that are implicated to be DV-modulated by genome-wide expression data are disproportionally enriched in the more stringent dataset as compared to the lower cut-off (data not shown).

Fig. S1 Comparison of motif enrichments observed in the data set derived from a lower cutoff (1) and that of a more stringent cutoff (2)



Are all binding events functional? What about those regions that do not contain evolutionarily conserved binding motifs?

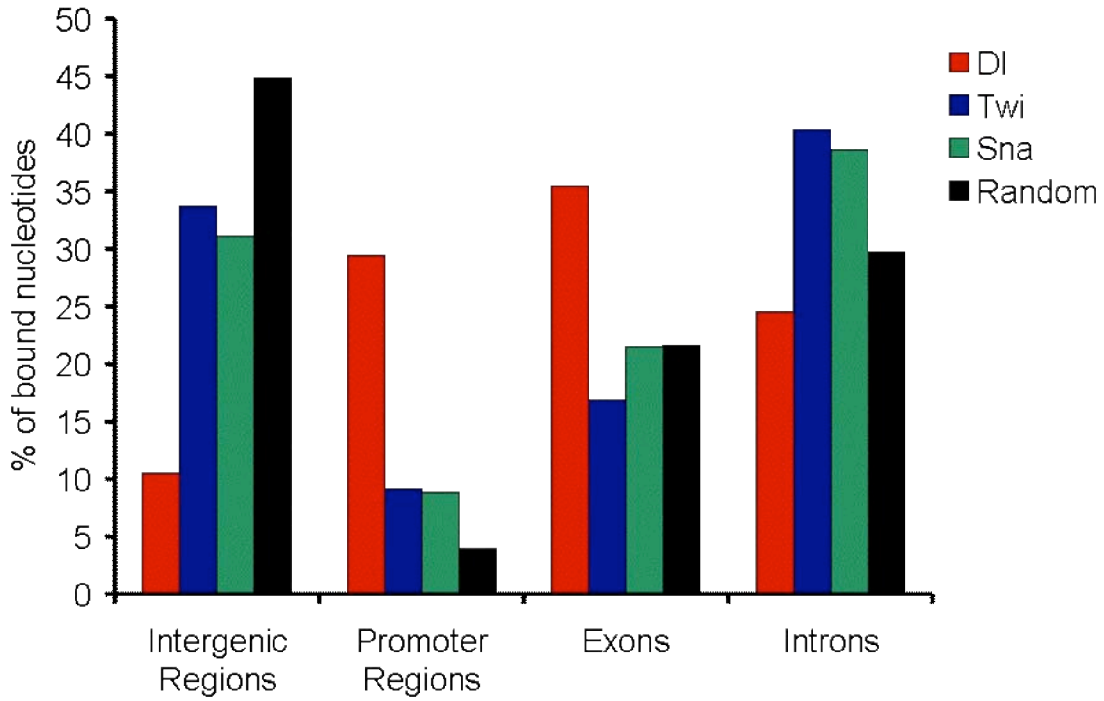
Based on our validation experiments, as well as our past experience in assessing microarray data quality, we believe that our data are of high quality and contain a minimal number of false positives due to errors in gene amplification and microarray hybridization. However, it is likely that not all binding events are functional, i.e. lead to changes in gene expression. In fact, it has been estimated that roughly half of the binding events detected by ChIP-chip in yeast may be non-functional (Gao et al. 2004).

Since non-functional binding sequences are less likely to be conserved across evolution, evolutionary conservation is a good filter for the identification of functional binding events. On the other hand, it is likely that some functional binding events are not conserved, either because the event is species-specific or because there is functional redundancy, i.e. other binding events replace its function. For example, we noted that the DTS-bound regions at segmentation genes were less well conserved than those from primary DV genes. Since segmentation and DV axis formation are controlled by genetically separate programs, it is possible that DV modulation of segmentation genes undergoes evolutionary changes more frequently than the regulation of primary DV genes.

Distribution of bound probes relative to gene models

While Twist and Snail are enriched at promoter regions (2 kb upstream of start sites) and in introns as expected, Dorsal is more highly enriched at promoters and exons (Fig. S2). Since many promoter and exon regions occupied by Dorsal do not contain Dorsal binding sites, the significance of these results is not clear.

Fig. S2 Distribution of bound regions

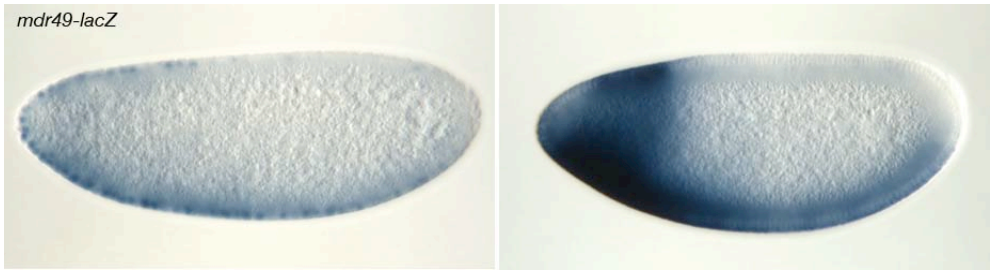


Further experimental validation

Mdr49 enhancer

A genomic sequence predicted by ChIP-chip to be associated with Dorsal, Twist, and Snail at the *Mdr49* locus was tested for enhancer activity. The reporter gene is expressed in the early mesoderm (Fig. S4), resembling the endogenous expression pattern (Stathopoulos et al. 2002).

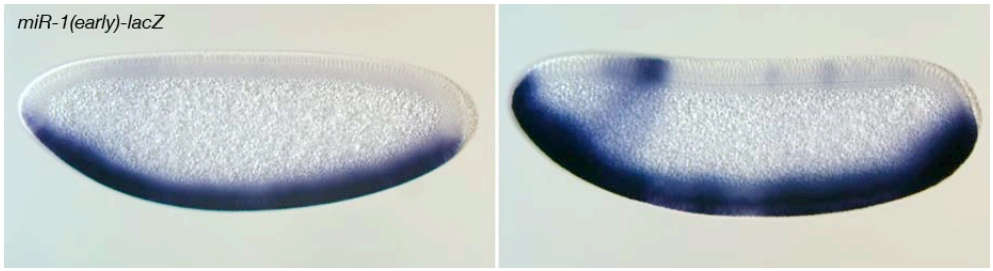
Fig. S3: *Mdr49* enhancer validation



miR-1 enhancer

Dorsal, Twist, and Snail were found to occupy a genomic sequence near *miR-1* that is significantly further upstream than either of the previously published *miR-1* enhancers (Fig.1B, Biemar et al. 2005; Sokol and Ambros 2005). When analyzed for enhancer activity, the genomic sequence directs *lacZ* reporter gene expression in the early mesoderm (Fig. S3), resembling the endogenous expression pattern.

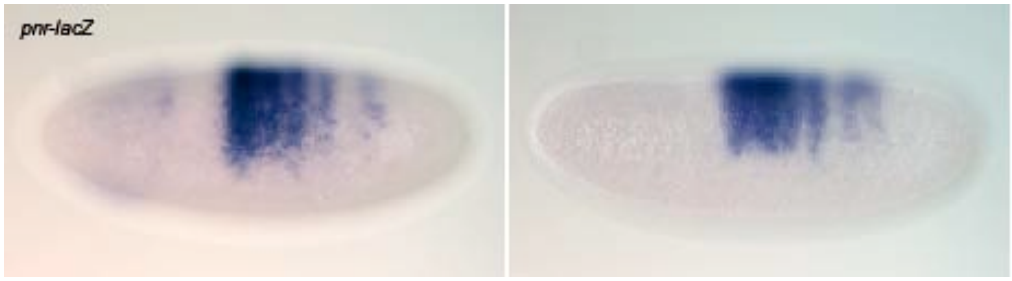
Fig. S4: *miR-1*(early) enhancer validation



pnr enhancer

An intronic sequence of the *pnr* gene identified by ChIP-chip was tested for enhancer activity. The reporter gene is expressed in the dorsal ectoderm in a pattern similar to that of the endogenous gene (Fig. S5).

Fig. S5: *pnr* enhancer validation



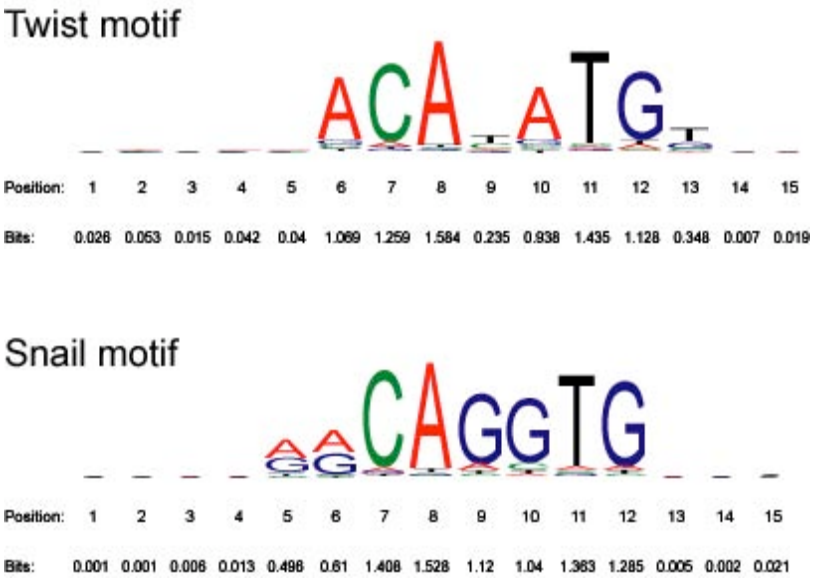
Motif analysis

Is there a difference between the Twist and Snail motif?

It has previously been shown *in vitro* and *in vivo* that the Twist and Snail binding sequences are very similar but not identical (Ip et al. 1992a). Since the ChIP-chip binding pattern of Twist and Snail is very similar, we tested whether the small differences between Twist and Snail binding might be explained by their different binding motifs. To do this, we performed *de novo* motif discovery in regions that were differentially bound by these factors (see the further description of Materials and Methods for details). We found similar but clearly distinct motifs (Fig. S6), which are highly similar to those found in SELEX studies (Zinzen et al. 2006).

These results suggest that while Twist and Snail may recognize the same sequence motifs in many instances, there are motif instances at which one binds with higher affinity than the other. In cases where Twist and Snail are able to bind to the same motif, we assume that this does not occur at the same time. We suggest that there is a binding equilibrium and that detection of Twist and Snail at the same site occurs because Twist is bound in some cells and Snail in others. It is also possible that Twist and Snail bind at the same time to the same region by binding to distinct motifs in close proximity to each other.

Fig. S6: The Twist and Snail motifs derived from ChIP-chip data *de novo*

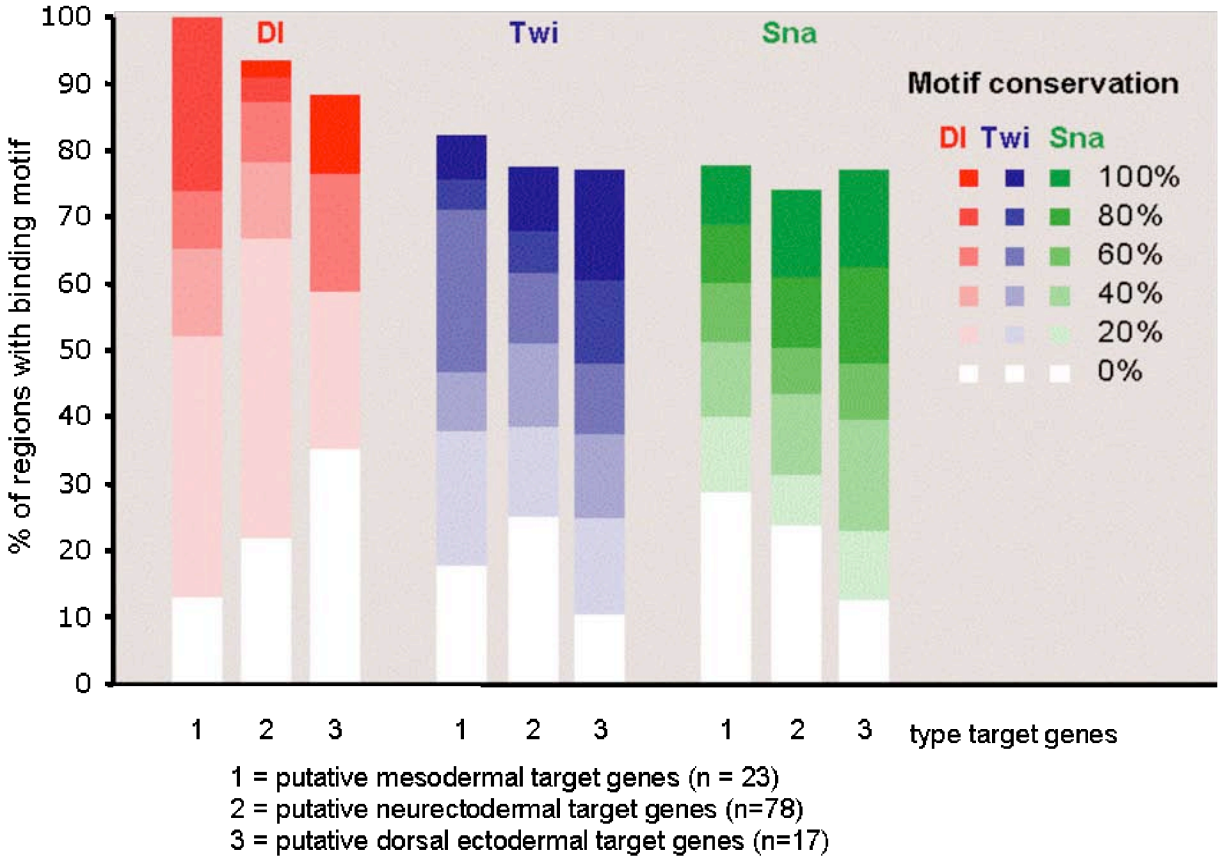


Analysis of type 1, 2 and 3 enhancers

Since Dorsal and Twist are predominantly activators and Snail is generally thought to be a dedicated repressor, we were surprised to find that all three proteins are found at both activated and repressed genes. Although activated genes often have higher enrichment levels of Twist as compared to Snail, this trend is not consistent and by itself insufficient to distinguish activated and repressed genes (data not shown).

We also tested whether activated and repressed genes could be distinguished based on the presence of Dorsal, Twist and Snail motifs. Although there are small differences among the three types of putative target genes based on expression analysis by Stathopoulos et al. 2002 (Fig.S7), the difference is not sufficient to distinguish activated and repressed genes.

Fig.S7 Dorsal, Twist and Snail motifs in type 1, 2 and 3 target genes



1 = putative mesodermal target genes (n = 23)
2 = putative neurectodermal target genes (n=78)
3 = putative dorsal ectodermal target genes (n=17)

Supplemental Tables

Table S1: DTS and TS regions at known DV genes

Region (Build April 2004)	Assigned gene	Gene symbol	Distance from TSS (bp)	Dorsal fold ChIP enrichment	Twist fold ChIP enrichment	Snail fold ChIP enrichment	known enhan-cer	confirm ed enhanc er	DV expre-ssion type*	References
2L_2447933_2448476	CG9885	dpp	transcript overlap	1.24	3.93	5.10			3	St Johnston and Gelbart 1987; Huang et al. 1993
2L_2456248_2457530	CG9885	dpp	transcript overlap	3.12	5.59	8.88	1		3	St Johnston and Gelbart 1987; Huang et al. 1993
2L_8809697_8810218	CG18024	SoxN	15407	1.71	3.57	5.77			2	Cremazy et al. 2000; Stathopoulos et al. 2002
2L_8832084_8832897	CG18024	SoxN	6459	2.75	4.89	6.54			2	Cremazy et al. 2000; Stathopoulos et al. 2002
2L_8841307_8842609	CG18024	SoxN	15682	4.20	9.72	13.74			2	Cremazy et al. 2000; Stathopoulos et al. 2002
2L_10548614_10549385	CG31721	Trim9	transcript overlap	2.14	4.23	7.28			2	Biemar et al. 2006
2L_12457269_12458378	CG5461	bun	transcript overlap	1.51	4.03	6.54			3	Biemar et al. 2006
2L_15471864_15476418	CG3956	sna	TSS overlap	8.21	12.11	19.24	1		1	Ip et al. 1992b
2L_17744588_17745109	CG7100	CadN	9138	2.55	6.74	4.75			1	Biemar et al. 2006
2L_17746765_17747317	CG7100	CadN	11315	2.11	3.50	5.03			1	Biemar et al. 2006
2L_18870798_18872110	CG10619	tup	transcript overlap	1.68	5.96	8.41		1	3	Frank and Rushlow 1996
2L_18877093_18877657	CG10619	tup	TSS overlap	3.39	7.96	14.52			3	Frank and Rushlow 1996
2L_18891098_18891898	CG10619	tup	13913	1.36	4.56	8.21			3	Frank and Rushlow 1996
2L_20468594_20469458	CR32958	mir-1	5631	2.02	12.59	11.67	1		1	Biemar et al. 2005
2L_20473425_20476067	CR32958	mir-1	TSS overlap	7.96	18.85	14.16			1	Biemar et al. 2005
2R_2891911_2894049	CG17800	Dscam	TSS overlap	7.26	17.78	27.80			2	Stathopoulos et al. 2002
2R_5443138_5443816	CG1429	Mef2	transcript overlap	1.92	6.15	3.76			1	Taylor et al. 1995
2R_7309632_7312483	CG12443	ths	transcript overlap	9.60	16.57	24.46	1		2	Stathopoulos et al. 2002
2R_8291749_8293745	CG17579	sca	2706	1.61	4.19	6.29			2	Mlodzik et al. 1990; Stathopoulos et al. 2002
2R_8295676_8296548	CG17579	sca	TSS overlap	2.50	6.74	7.47			2	Mlodzik et al. 1990; Stathopoulos et al. 2002

2R_16871284_16872649	CG15671	cv-2	transcript overlap	1.62	3.15	6.53			3	Biemar et al. 2006
2R_18552484_18552954	CG2956	twi	754	2.70	2.94	5.01			1	Jiang et al. 1991
2R_18553474_18554632	CG2956	twi	TSS overlap	2.94	3.55	5.57	1		1	Jiang et al. 1991
2R_19494333_19495879	CG3832	Phm	transcript overlap	2.95	2.59	5.24	1		1	Markstein et al. 2004
3L_1439750_1440303	CG1004	rho	7021	2.06	4.73	8.64			2	Ip et al. 1992a
3L_1445100_1446447	CG1004	rho	877	9.86	16.31	28.56	1		2	Ip et al. 1992a
3L_2863961_2866611	CG9973	CG9973	8719	3.31	6.61	9.20			2	Biemar et al. 2006
3L_5809480_5810317	CG10491	vn	transcript overlap	4.99	10.32	13.15	1		2	Markstein et al. 2004
3L_8977410_8977946	CG5093	Doc3	1199	1.77	9.30	15.66			3	Stathopoulos et al. 2002
3L_8978566_8979328	CG5093	Doc3	TSS overlap	1.91	7.08	8.74			3	Stathopoulos et al. 2002
3L_9016068_9016617	CG5133	Doc1	transcript overlap	1.39	2.44	5.83			3	Stathopoulos et al. 2002
3L_9017443_9018000	CG5133	Doc1	transcript overlap	1.19	4.37	7.45			3	Stathopoulos et al. 2002
3L_15004570_15006144	CG11551	ind	1685	4.06	6.79	9.25	1		2	Weiss et al. 1998, unpublished
3L_20334025_20334531	CG5408	trbl	transcript overlap	3.09	3.38	6.26			1	Casal and Leptin 1996
3L_20336151_20337244	CG5408	trbl	702	1.68	10.73	17.09			1	Casal and Leptin 1996
3L_20338381_20339810	CG5408	trbl	2932	2.20	4.23	6.31			1	Casal and Leptin 1996
3R_2577181_2578190	CG1046	zen	1731	1.74	5.39	5.65			3	Doyle et al. 1989; Jiang et al. 1991
3R_2579787_2581714	CG1046	zen	TSS overlap	2.66	5.49	10.05	1		3	Doyle et al. 1989; Jiang et al. 1991
3R_4847408_4849288	CG11988	neur	transcript overlap	3.89	3.39	6.27			1	Price et al. 1993
3R_4849848_4851522	CG11988	neur	transcript overlap	3.60	4.99	8.07			1	Price et al. 1993
3R_4855544_4856298	CG11988	neur	transcript overlap	2.99	2.55	5.25			1	Price et al. 1993
3R_4863101_4864713	CG11988	neur	transcript overlap	3.96	10.93	15.50			1	Price et al. 1993
3R_5328885_5329680	CG9366	RhoL	586	1.11	5.11	5.93			1	Casal and Leptin 1996; Stathopoulos et al. 2002
3R_8895104_8898505	CG7771	sim	transcript overlap	5.26	17.14	18.36	1		2	Kasai et al. 1992; Kasai et al. 1998; Markstein et al. 2004
3R_9118697_9120519	CG8458	wntD	TSS overlap	5.84	24.20	35.45		1	1	Gordon et al. 2005
3R_10412302_10413370	CG31317	stumps	4815	1.77	6.01	7.11			1	Casal and Leptin 1996, unpublished
3R_10504881_10506240	CG7649	Neu3	transcript overlap	4.52	13.91	16.14			2	Stathopoulos et al. 2002
3R_11854020_11855407	CG3978	pnr	transcript overlap	1.27	7.06	10.32		1	3	Winick et al. 1993

3R_11859695_11860268	CG3978	pnr	transcript overlap	1.62	3.31	7.37			3	Winick et al. 1993
3R_20574726_20575587	CG6868	tld	TSS overlap	2.41	4.36	9.10	1		3	Kirov et al. 1994
3R_21861181_21862991	CG8365	E(spl)	TSS overlap	5.52	15.62	16.42	1		2	unpublished
3R_21865619_21866301	CG8365	E(spl)	TSS overlap	1.93	3.49	5.61			2	unpublished
3R_25373440_25374228	CG1897	Dr	7881	1.55	6.06	8.59			2	D'Alessio and Frasch 1996; von Ohlen and Doe 2000
3R_26589558_26590308	CG1322	zfh1	1340	0.92	4.41	5.59			1	Lai et al. 1991
3R_26592529_26593072	CG1322	zfh1	transcript overlap	2.01	3.54	6.64			1	Lai et al. 1991
3R_26607274_26608079	CG1322	zfh1	transcript overlap	2.39	4.07	6.95			1	Lai et al. 1991
X_266022_266824	CG3839	l(1)sc	TSS overlap	5.08	9.43	15.00			2	Romani et al. 1987
X_440581_441651	CG6172	vnd	6212	5.15	10.40	10.22		1	2	McDonald et al. 1998; Stathopoulos et al. 2002; Markstein et al. 2004
X_442754_443577	CG6172	vnd	4286	6.41	11.11	14.16		1	2	McDonald et al. 1998; Stathopoulos et al. 2002; Markstein et al. 2004
X_448941_450341	CG6172	vnd	transcript overlap	24.06	34.00	42.16	1		2	McDonald et al. 1998; Stathopoulos et al. 2002; Markstein et al. 2004
X_7141851_7142904	CG9653	brk	5311	8.10	17.16	22.46	1		2	Jazwinska et al. 1999; Markstein et al. 2004
X_13513751_13514689	CG12177	CG12177	TSS overlap	4.11	10.20	11.01	1		1	unpublished
X_15458274_15459269	CG9224	sog	transcript overlap	4.89	9.32	14.03	1		2	Francois et al. 1994; Markstein et al. 2004
X_22039681_22040474	CG9559	fog	transcript overlap	1.46	3.86	6.25			1	Costa et al. 1994

* DV expression type: 1 = mesdermal target genes, 2 = neurectodermal target genes, 3 = dorsal ectodermal target genes

Table S2: TS and DTS regions at genes encoding anteroposterior determinants

Region (Build April 2004)	Assigned gene	Gene symbol	Distance from TSS (bp)	Dorsal fold ChIP enrichment	Twist fold ChIP enrichment	Snail fold ChIP enrichment	known enhan-cer	confir med enhan -cer	AP type*	References
2L_3602601_3603124	CG3851	odd	3624	1.25	4.12	6.11			P	
2L_3608131_3609727	CG3851	odd	1383	2.26	5.91	7.36			P	
2L_3821283_3823191	CG16738	slp1	2489	1.17	4.08	5.55			P	
2L_3833582_3834671	CG2939	slp2	2171	1.40	3.10	5.66			P	
2L_7299287_7300660	CG4889	wg	6501	2.46	4.89	7.25			S	
2L_11447614_11448150	CG6464	salm	2011	2.28	4.90	8.22			H	
2L_20765933_20766465	CG1759	cad	transcript overlap	1.57	4.23	6.75			M	
2L_21834858_21835698	CG1374	tsh	18670	1.60	4.32	5.27			H	
2L_21848032_21848773	CG1374	tsh	31844	1.43	4.10	7.18			H	
2R_4154841_4155926	CG2411	ptc	5522	1.72	4.68	7.17			S	
2R_4169474_4170261	CG2411	ptc	transcript overlap	1.78	4.34	7.86			S	
2R_7048168_7048725	CG9015	en	4051	1.27	4.94	7.69			S	
2R_7053070_7053844	CG9015	en	8953	1.58	4.26	6.79			S	
2R_7062435_7062989	CG9015	en	18318	1.92	3.93	5.31			S	
2R_7064543_7065075	CG9015	en	20426	1.30	3.50	6.75			S	
2R_20321033_20322290	CG3629	Dll	140	1.33	7.22	21.03			H	
2R_20564456_20565235	CG3388	gsb	4345	1.34	4.33	7.63			S	
2R_20572795_20573628	CG3388	gsb	3215	1.52	4.65	8.07			S	
3L_8644120_8645432	CG6494	h	4274	1.66	5.07	9.62			P	
3L_8649141_8650163	CG6494	h	TSS overlap	1.86	4.02	9.84			P	
3L_18998192_18999254	CG11614	nkd	transcript overlap	3.23	5.70	6.44			S	
3L_20627042_20628873	CG4717	kni	transcript overlap	2.36	4.31	7.02			G	
3L_20633387_20634232	CG4717	kni	4183	3.93	8.38	15.53		1	G	
3R_173772_174862	CG9768	hkb	1400	1.80	5.62	9.43	1		G	Hader et al.,2000
3R_671775_672589	CG1133	opa	5946	1.19	6.35	11.10			P	
3R_673147_673632	CG1133	opa	4903	1.57	3.55	5.09			P	

3R_674741_675542	CG1133	opa	2993	1.35	3.69	5.79			P	
3R_677966_679248	CG1133	opa	TSS overlap	2.40	4.00	6.55			P	
3R_679745_680348	CG1133	opa	transcript overlap	2.72	4.69	6.98			P	
3R_683739_684294	CG1133	opa	transcript overlap	2.34	3.49	7.14			P	
3R_2565178_2566035	CG31481	pb	transcript overlap	1.50	3.83	7.85			H	
3R_2689422_2690469	CG2047	ftz	TSS overlap	2.48	3.57	5.65			P	
3R_2691263_2692029	CG2047	ftz	1217	2.40	5.91	9.15			P	
3R_2714300_2714825	CG2047	ftz	24254	1.71	3.35	6.50			P	
3R_2718739_2719479	CG2047	ftz	28693	1.78	4.66	6.07			P	
3R_2742247_2743075	CG1028	Antp	transcript overlap	2.22	5.22	9.89			H	
3R_2774633_2775201	CG1028	Antp	transcript overlap	1.17	3.41	5.38			H	
3R_2819923_2821134	CG1028	Antp	transcript overlap	3.12	7.05	14.02			H	
3R_2832380_2833209	CG1028	Antp	7430	1.99	3.43	6.70			H	
3R_4524571_4525394	CG9786	hb	1031	1.12	2.11	5.22			G	
3R_9725409_9725851	CG2988	ems	1730	1.25	4.38	6.36			G	
3R_9735145_9735985	CG2988	ems	7564	1.50	3.40	5.34			G	
3R_12545317_12546139	CG10388	Ubx	transcript overlap	1.36	4.56	7.48			H	
3R_18957147_18957694	CG4637	hh	transcript overlap	2.10	8.00	9.10			S	
3R_19018827_19019617	CG17894	cnc	transcript overlap	2.46	12.47	14.15			H	
3R_19021189_19022009	CG17894	cnc	transcript overlap	1.39	5.51	7.51			H	
3R_26675168_26675975	CG1378	tll	2062	2.86	3.47	5.11	1		G	Liawa and Lengyel, 1993
3R_26681533_26682349	CG1378	tll	3496	1.92	4.05	6.28			G	
X_1755368_1756347	CG11579	arm	transcript overlap	2.37	3.97	5.20			S	
X_8498803_8500756	CG12154	oc	2917	1.96	5.80	7.17	1		G	Gao and Finkelstein, 1998
X_20502036_20503071	CG1849	run	2046	1.42	2.83	6.35			P	
X_20504075_20505996	CG1849	run	TSS overlap	2.67	4.09	6.10			P	

* AP type: M = maternal gene, G =gap gene, P = pair-rule gene, S = segment polarity, H = homeotic gene

Table S3: TS and DTS regions at genes encoding developmental signaling molecules

Region (Build April 2004)	Assigned gene	Gene symbol	Distance from TSS (bp)	Dorsal fold ChIP enrichment	Twist fold ChIP enrichment	Snail fold ChIP enrichment	Signaling pathway
2L_1062806_1063327	CG4385	S	transcript overlap	1.7	3.8	3.8	EGF
2L_1065637_1067268	CG4385	S	transcript overlap	2.3	6.0	8.6	EGF
2L_1067819_1069122	CG4385	S	transcript overlap	2.3	5.9	7.8	EGF
2L_1075636_1077929	CG4385	S	TSS overlap	5.2	13.1	15.1	EGF
2L_2162783_2165538	CG3166	aop	transcript overlap	3.9	7.5	10.9	EGF
2L_2171847_2173170	CG3166	aop	transcript overlap	1.8	7.4	10.3	EGF
2L_2178240_2178982	CG3166	aop	TSS overlap	1.7	3.3	5.5	EGF
2L_2447933_2448476	CG9885	dpp	transcript overlap	1.2	3.9	5.1	Dpp
2L_2453415_2453904	CG9885	dpp	transcript overlap	2.1	2.2	3.1	Dpp
2L_2454392_2454768	CG9885	dpp	transcript overlap	2.1	2.4	3.2	Dpp
2L_2456248_2457530	CG9885	dpp	transcript overlap	3.1	5.6	8.9	Dpp
2L_277990_278805	CG11561	smo	transcript overlap	1.6	2.6	3.3	Hh
2L_3158653_3159389	CG12399	Mad	transcript overlap	4.2	2.3	3.8	Dpp
2L_5239532_5240570	CG14026	tkv	transcript overlap	1.9	5.7	4.6	Dpp
2L_7299287_7300660	CG4889	wg	6501	2.5	4.9	7.2	Wg
2L_9454180_9455027	CG3779	numb	transcript overlap	2.3	10.2	10.9	N
2R_17031153_17033313	CG10079	Egfr	transcript overlap	2.7	7.6	11.9	EGF
2R_17051799_17054492	CG10079	Egfr	transcript overlap	4.2	10.6	21.6	EGF
2R_4154841_4155926	CG2411	ptc	5522	1.7	4.7	7.2	Hh
2R_4169474_4170261	CG2411	ptc	transcript overlap	1.8	4.3	7.9	Hh
2R_6694844_6695084	CG7734	shn	transcript overlap	1.3	12.0	21.1	Dpp
2R_6707354_6707906	CG7734	shn	transcript overlap	3.2	3.2	5.8	Dpp
2R_6713007_6716096	CG7734	shn	transcript overlap	1.9	5.0	6.8	Dpp
2R_8046410_8046931	CG8581	fra	transcript overlap	1.4	3.4	6.8	EGF
3L_1439750_1440303	CG1004	rho	7021	2.1	4.7	8.6	EGF
3L_1445100_1446447	CG1004	rho	877	9.9	16.3	28.6	EGF
3L_14932886_14935033	CG5185	Tom	TSS overlap	7.9	10.7	13.0	N
3L_14937109_14939212	CG3096	Brd	TSS overlap	3.3	5.8	8.3	N
3L_16440643_16441511	CG4531	argos	transcript overlap	2.4	4.4	8.4	EGF
3L_3387415_3387967	CG1921	sty	transcript overlap	1.6	6.6	12.0	STAT
3L_3392148_3393969	CG1921	sty	transcript overlap	2.4	7.0	9.5	STAT
3L_732910_733380	CG1007	emc	TSS overlap	3.3	1.9	5.8	Dpp
3L_736881_737721	CG1007	emc	3943	5.9	4.1	7.2	Dpp
3L_8803491_8804296	CG4974	dally	transcript overlap	1.6	4.6	6.9	Wg
3L_8806639_8807425	CG4974	dally	transcript overlap	1.4	2.6	5.4	Wg
3R_10412302_10413370	CG31317	stumps	4815	1.8	6.0	7.1	FGF
3R_13874578_13875050	CG7223	htl	transcript overlap	1.6	1.9	2.3	FGF

3R_13875558_13876595	CG7223	htl	transcript overlap	1.8	4.5	2.9	FGF
3R_1510161_1511458	CG2108	Rab23	TSS overlap	1.9	5.6	9.8	Hh
3R_15131038_15131876	CG3619	Dl	transcript overlap	2.2	1.9	2.6	N
3R_15140852_15141125	CG3619	Dl	transcript overlap	2.1	2.7	3.3	N
3R_15146804_15147051	CG3619	Dl	transcript overlap	1.2	1.5	2.3	N
3R_15151378_15151986	CG3619	Dl	TSS overlap	2.1	1.6	2.0	N
3R_15155533_15156312	CG3619	Dl	3598	1.8	6.8	7.9	N
3R_15157497_15158780	CG3619	Dl	5562	2.0	3.0	3.1	N
3R_15160250_15160530	CG3619	Dl	8315	1.2	1.4	2.2	N
3R_15164424_15165483	CG3619	Dl	12489	2.0	4.7	7.7	N
3R_15166283_15168070	CG3619	Dl	14348	3.3	14.0	19.3	N
3R_15170820_15172829	CG3619	Dl	18885	1.5	5.4	9.1	N
3R_16368496_16369600	CG4257	Stat92E	transcript overlap	2.8	7.1	7.8	STAT
3R_16375977_16376973	CG4257	Stat92E	transcript overlap	2.7	7.6	10.8	STAT
3R_18957147_18957694	CG4637	hh	transcript overlap	2.1	8.0	9.1	Hh
3R_19124440_19127283	CG17077	pnt	transcript overlap	1.9	9.1	12.8	EGF
3R_19169083_19170192	CG17077	pnt	transcript overlap	2.8	7.9	9.2	EGF
3R_20574726_20575587	CG6868	tld	TSS overlap	2.4	4.4	9.1	Dpp
3R_3939141_3939926	CG7850	puc	transcript overlap	1.7	4.2	5.9	JNK
3R_3942101_3942380	CG7850	puc	transcript overlap	1.6	2.5	5.0	JNK
3R_4667547_4668085	CG31349	pyd	transcript overlap	3.3	7.2	11.8	FGF
3R_4682860_4684183	CG31349	pyd	transcript overlap	2.1	3.5	5.6	FGF
3R_4847408_4849288	CG11988	neur	transcript overlap	3.9	3.4	6.3	N
3R_4849848_4851522	CG11988	neur	transcript overlap	3.6	5.0	8.1	N
3R_4855544_4856298	CG11988	neur	transcript overlap	3.0	2.6	5.2	N
3R_4863101_4864713	CG11988	neur	transcript overlap	4.0	10.9	15.5	N
3R_5328885_5329680	CG9366	RhoL	586	1.1	5.1	5.9	ERK
3R_6425355_6426322	CG17117	hth	transcript overlap	3.2	4.1	6.9	FGF
3R_6439800_6440941	CG17117	hth	transcript overlap	2.4	5.5	12.3	FGF
3R_6442333_6443741	CG17117	hth	transcript overlap	1.9	4.7	5.6	FGF
3R_9118697_9120519	CG8458	wntD	TSS overlap	5.8	24.2	35.4	Wnt
4_76522_77315	CG2125	ci	transcript overlap	9.4	4.3	5.7	Hh
X_1167018_1167841	CG14622	DAAM	transcript overlap	1.4	4.6	6.9	Wg
X_12488043_12488909	CG2028	Cklalpha	transcript overlap	2.0	5.5	8.8	Hh
X_15458274_15459269	CG9224	sog	transcript overlap	4.9	9.3	14.0	Dpp
X_1755368_1756347	CG11579	arm	transcript overlap	2.4	4.0	5.2	Wg
X_2992423_2992958	CG3936	N	transcript overlap	2.1	4.1	5.9	N
X_2995280_2996295	CG3936	N	transcript overlap	2.0	3.6	5.2	N
X_3008126_3009129	CG3936	N	transcript overlap	1.6	5.7	6.5	N
X_3012003_3012747	CG3936	N	transcript overlap	2.3	8.2	10.9	N

Further description of Materials and Methods

Design of the *Drosophila* whole-genome arrays

Genome sequence and handling of masked regions

Build April 2004 downloaded from UCSC browser, using the repeat-masked (-s option) version of the sequences and coordinates. Regions that were repeat masked were not considered for oligo design.

Oligo selection criteria

The array was designed for 60-mer oligonucleotide arrays as manufactured by Agilent. To define the quality of oligos, we used the four criteria found in the ArrayOligoSelector program: GC content, self-binding, complexity and uniqueness (Bozdech et al. 2003). Instead of BLAST we used BLAT for sequence alignment. Based on signal intensities obtained from test arrays, we derived thresholds for top quality and medium quality oligos:

Top quality oligos (stringent filter):

- GC content between 30 percent and 100 percent
- Self-binding score less than 90
- Complexity score less than or equal to 25
- Uniqueness equal to 0 (no other BLAT match)

Medium quality oligos (relaxed filter):

- GC content between 30 percent and 100 percent
- Self-binding score less than 110
- Complexity score less than or equal to 26
- Uniqueness unequal to F

Spacing

After sorting all oligos by chromosomal order, the oligo search starts with the lowest chromosomal coordinate on the Watson strand and then selected a

qualified probe that is furthest away within a window of 150 -280 bp. If there were no probes within this limit, the next acceptable probe was selected. The process was then repeated with the most recently selected probe.

The search was first performed to select top quality oligos. Gaps were then filled with medium quality probes, and later oligos of any quality (if there were no sequence uncertainties). Test arrays showed that even low quality oligos respond to signal and overall enhance the confidence of binding signals without increasing noise.

Handling of duplicated regions

To avoid selecting different oligos for identical regions, oligos for duplicated regions (100% sequence identity) were selected only once and replicated and assigned to the other identical regions

Control probes

866 negative control spots were added to each array:

Normalization controls	
8+5 oligos from non-genic (desert) regions and from the middle of long exons, both of which should not be enriched in most ChIP-chip experiments. Each oligo is present 10 times. They hybridize at different intensities and are used for normalization.	130
More non-genic probes	357
More long exon probes	273
Non-specific hybridization controls (Arabidopsis sequences)	40
Non-specific hybridization controls (Repeat sequences)	66
	866

There are also positive control spots specific to the experiment (Dorsal, Twist and Snail targets), blank spots and controls added by Agilent (standard).

ChIP-chip protocol

The protocol was developed based on protocols from the Young and Maschat labs (Chanas et al. 2004; Lee et al. 2006b). Briefly, embryos were dechorionated, rinsed in isopropanol and cross-linked for 5 min. in a 5% formaldehyde/hexane solution (Toth and Biggin 2000), washed twice in 1X PBS + 0.5% Triton X-100, flash frozen and weighed. Upon thawing, embryos were resuspended in buffer A1 (15 m*M* HEPES pH 7.5, 15 mM NaCl, 60 m*M* KCl, 4 m*M* $MgCl_2$, 0.5 % Triton X-100, 0.5 m*M* DDT, EDTA-free protease inhibitor), pooled and disrupted in a 7 ml Wheaton Dounce homogenizer. The homogenate was transferred to a 15 ml Falcon tube, centrifuged for 3 min. at 3000 g at 4°C, the pellet was resuspended in 5 ml fresh buffer A1 by stirring and washed in this manner thrice with buffer A1 and once with buffer A2 (15m*M* HEPES pH 7.5, 140 m*M* NaCl, 1 m*M* EDTA, 0.5 m*M* EGTA, 1 % Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 0.5% N-lauroylsarcosine, proteinase inhibitors). At the second wash with buffer A2, the sample volume was adjusted to 0.5 ml per 0.15g of cross-linked embryos. Samples were sonicated 4 times for 30 sec (power 3, Brandson sonifier 250), transferred to microfuge tubes, and centrifuged for 10 min at high speed at 4°C. The supernatant was transferred to a fresh tube on ice. For each IP, 200 µl of extract was incubated with magnetic beads that had been pre-incubated with antibodies overnight at 4°C. Washing of the beads, as well as purification, amplification and labeling of the DNA was performed as previously described (Lee et al. 2006b). The Cy5-labeled IP DNA and Cy3-labeled input DNA were hybridized to whole-genome *Drosophila* arrays using the CGH protocol and CGH reagents provided by Agilent Technologies.

Sequence motifs used for motif analysis

The 12 sequenced *Drosophila* species are found at http://hanuman.math.berkeley.edu/genomes/drosophila.html.

The binding motifs for Dorsal used for the analysis were 'GGGWNNNNCCM', 'GGGWWWCCC', 'GGGWWWWCCA', 'GGGWWWWCYS', 'GGGWDWWWCCM', 'WWWWWWWWWWCCC', 'GGGWWWWCCM', 'GGGWDWWWCCM' (Markstein et al. 2002; Markstein et al. 2004), 'GGAATTTCC', (Papatsenko and Levine 2005), 'SGGKKTTTYYCV' (Jaspar MA0022), 'GKGGWWTTCC' (Jasper MA0023), 'GGGWWTTTCCV' (Transfac M00043), 'HSRGAAAAHYV' (Transfac M00120).

The binding motifs for Twist were 'CACATGT', 'CAYRTGT', 'CAYNTGT' (Zinzen et al. 2006).

The binding motifs for Snail were: 'MMRCAWGT', 'RCARGT' (Zinzen et al. 2006), 'CAGGTG' (Jaspar MA0086), 'MSMASBTGHTDVSW' (Transfac M00044), 'VHRRCAGGTGYMN' (Transfac M00060).

De novo sequence analysis

We extracted the genomic sequences corresponding to snail and twist bound regions from the genomes of the *melanogaster* group of Drosophila (i.e. *dmel, dana, dere, dyak, dsec,* and *dsim*). In addition, we extracted regions exclusively bound by Snail or Twist and random intergenic regions as control. In each region, we counted all instances of all possible motifs consisting of 6 defined nucleotides and a central gap between 0 and 5 nucleotides. From all motifs, we selected those that were significantly conserved and/or enriched in the ChIP-chip regions and clustered them based on sequence similarity to remove redundant highly similar motif variants. In Snail and Twist bound regions, we discovered motifs with high similarity to the known Snail motif. The Twist motif was only discovered in regions that were exclusively bound by Twist. We re-scanned the bound regions for all the motif instances and their close sequence variants and constructed position-specific weight matrices from the recovered sequences. We also used Bioprospector to find motifs that were enriched in bound regions vs.

control regions but found exclusively highly repetitive sequences of the form (AC)n.

Primers used for amplification of newly identified enhancers

wntD	F137 5'-GCAAATCCCAAGCCAGGGCGCCCTCC-3'
	R137 5'-ACTGGCAGTTCCCGCCGGCTCACCAC-3'
vnd(vNE)	F141 5'-CGGACTTGAATGGTCGGTCAC-3'
	R141 5'-GCCTATGCTCGTCGTTGATGTTC-3'
vnd(mc)	F140 5'-CATGTCCTCGTTCAGGAAACTGTTAC-3'
	R140 5'-GTTATGTAAGGGGATGGGTCCTAAC-3'
kni	F144 5'-GGTTCGGTTTCGCCTGACAAATGTCTTGTG-3'
	R144 5'-GGTATCCGGTCAGAATTTTTATGGGCGATC-3'
tup	tupF 5'-GAATGCCTCTCTTTCCGTCTGGCCG-3'
	tupR 5'-AACCACTCCACTCAATGTCAAGTGGAG-3'
pnr	pnrF 5'-ATAAATTCATGCTCCTTGAACTG-3'
	pnrR 5'-TAAATTAAAAGGCGCTTAAACACGCAC-3'
Mdr49	F20 5'–CGTATAACTAATCGGGATTCCCGCCTGGCAACCCG–3'
	R20 5'–CCTCTCGGTTAGTCCCCCGATCTGGCGAATTCAGC–3'
miR-1	F139 5'–CTCGAAAATGAAACGGACAATAAGC–3'
	R139 5'–GAAGGACTCAGGACTCAAGCCTC–3'

References

Biemar, F., Nix, D.A., Piel, J., Peterson, B., Ronshaugen, M., Sementchenko, V., Bell, I., Manak, J.R., and Levine, M.S. 2006. Comprehensive identification of Drosophila dorsal-ventral patterning genes using a whole-genome tiling array. *Proc Natl Acad Sci U S A* **103**: 12763-8.

Biemar, F., Zinzen, R., Ronshaugen, M., Sementchenko, V., Manak, J.R., and Levine, M.S. 2005. Spatial regulation of microRNA gene expression in the Drosophila embryo. *Proc Natl Acad Sci U S A* **102**: 15907-11.

Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B.,and DeRisi, J.L. 2003. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome Biol* **4**: R9.

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 U S A* **93**: 10327-32.

Chanas, G., Lavrov, S., Iral, F., Cavalli, G., and Maschat, F. 2004. Engrailed and polyhomeotic maintain posterior cell identity through cubitus-interruptus regulation. *Dev Biol* **272**: 522-35.

Costa, M., Wilson, E.T., and Wieschaus, E. 1994. A putative cell signal encoded by the folded gastrulation gene coordinates cell shape changes during Drosophila gastrulation. *Cell* **76**: 1075-89.

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**: 215-9.

D'Alessio, M. and Frasch, M. 1996. msh may play a conserved role in dorsoventral patterning of the neuroectoderm and mesoderm. *Mech Dev* **58**: 217-31.

Doyle, H.J., Kraut, R., and Levine, M. 1989. Spatial regulation of zerknullt: a dorsal-ventral patterning gene in Drosophila. *Genes Dev* **3**: 1518-33.

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**: 2602-16.

Frank, L.H. and Rushlow, C. 1996. A group of genes required for maintenance of the amnioserosa tissue in Drosophila. *Development* **122**: 1343-52.

Gao, F., Foat, B.C., and Bussemaker, H.J. 2004. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics* **5**: 31.

Gordon, M.D., Dionne, M.S., Schneider, D.S., and Nusse, R. 2005. WntD is a feedback inhibitor of Dorsal/NF-kappaB in Drosophila development and immunity. *Nature* **437**: 746-9.

Huang, J.D., Schwyter, D.H., Shirokawa, J.M., and Courey, A.J. 1993. The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. *Genes Dev* **7**: 694-704.

Ip, Y.T., Park, R.E., Kosman, D., Bier, E., and Levine, M. 1992a. The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the Drosophila embryo. *Genes Dev* **6**: 1728-39.

Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. 1992b. dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev* **6**: 1518-30.

Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S., and Rushlow, C. 1999. The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**: 563-73.

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**: 1881-91.

Kasai, Y., Nambu, J.R., Lieberman, P.M., and Crews, S.T. 1992. Dorsal-ventral patterning in Drosophila: DNA binding of snail protein to the single-minded gene. *Proc Natl Acad Sci U S A* **89**: 3414-8.

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**: 171-89.

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**: 713-22.

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**: 123-34.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H.L., Zucker, J.P., Yuan, B., Bell, G.W., Herbolsheimer, E., Hannett, N.M., Sun, K., Odom, D.T., Otte, A.P., Volkert, T.L., Bartel, D.P., Melton, D.A., Gifford, D.K., Jaenisch, R., and Young, R.A. 2006a. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-13.

Lee, T.I., Johnstone, S.E., and Young, R.A. 2006b. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature Protocols* **1**: 729 - 748.

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 U S A* **99**: 763-8.

Markstein, M., Zinzen, R., Markstein, P., Yee, K.P., Erives, A., Stathopoulos, A., and Levine, M. 2004. A regulatory code for neurogenic gene expression in the Drosophila embryo. *Development* **131**: 2387-94.

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**: 3603-12.

Mlodzik, M., Baker, N.E., and Rubin, G.M. 1990. Isolation and expression of scabrous, a gene regulating neurogenesis in Drosophila. *Genes Dev* **4**: 1848-61.

Papatsenko, D. and Levine, M. 2005. Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* **102**: 4966-71.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., and Young, R.A. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517-27.

Price, B.D., Chang, Z., Smith, R., Bockheim, S., and Laughon, A. 1993. The Drosophila neuralized gene encodes a C3HC4 zinc finger. *Embo J* **12**: 2411-8.

Romani, S., Campuzano, S., and Modolell, J. 1987. The achaete-scute complex is expressed in neurogenic regions of Drosophila embryos. *Embo J* **6**: 2085-2092.

Sokol, N.S. and Ambros, V. 2005. Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev* **19**: 2343-54.

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**: 2785-91.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., and Levine, M. 2002. Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell* **111**: 687-701.

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.

Toth, J. and Biggin, M.D. 2000. The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in drosophila embryos. *Nucleic Acids Res* **28**: e4.

von Ohlen, T. and Doe, C.Q. 2000. Convergence of dorsal, dpp, and egfr signaling pathways subdivides the drosophila neuroectoderm into three dorsal-ventral columns. *Dev Biol* **224**: 362-72.

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**: 3591-602.

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**: 1055-65.

Zinzen, R.P., Senger, K., Levine, M., and Papatsenko, D. 2006. Computational models for neurogenic gene expression in the Drosophila embryo. *Curr Biol* **16**: 1358-65.