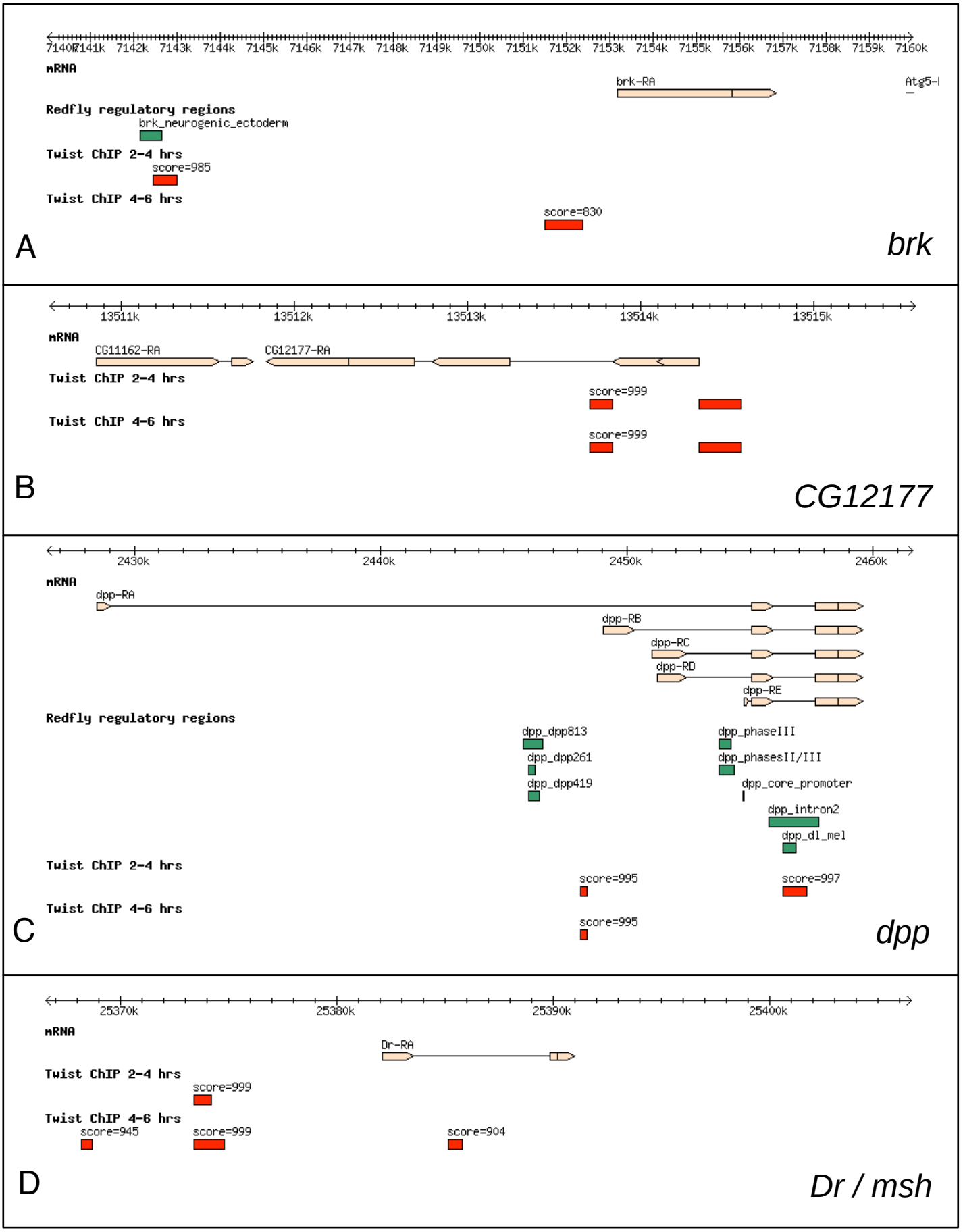
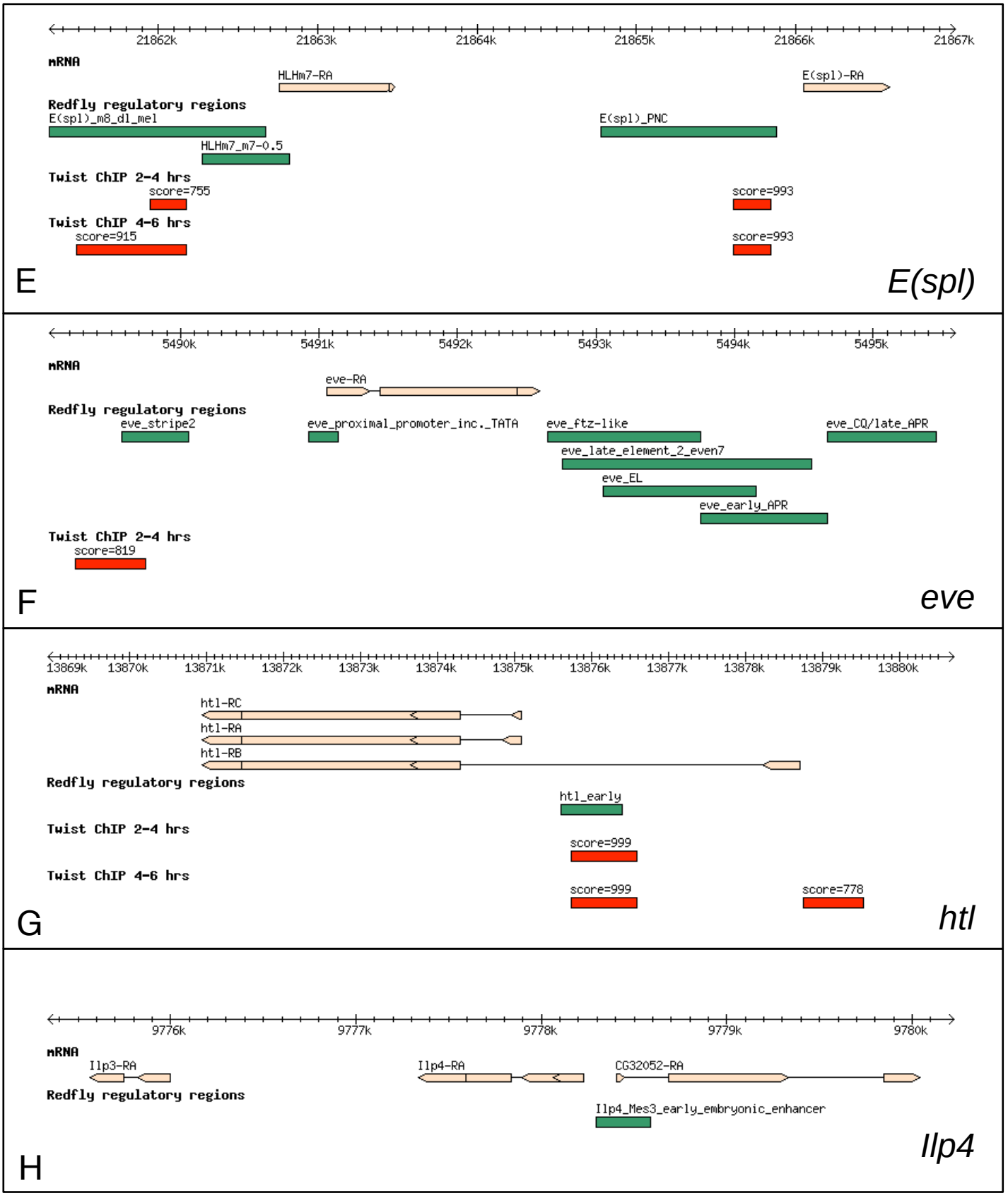
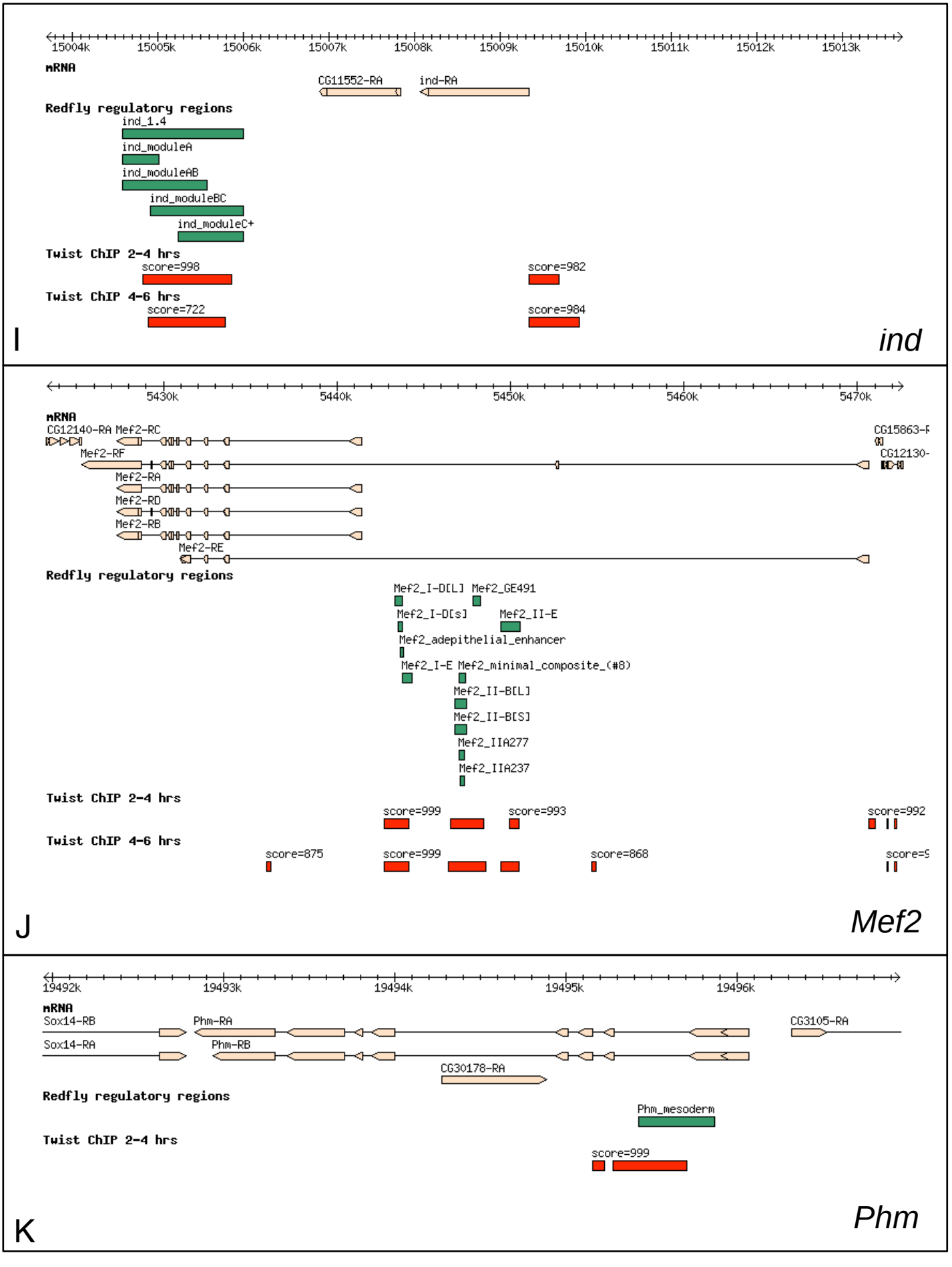
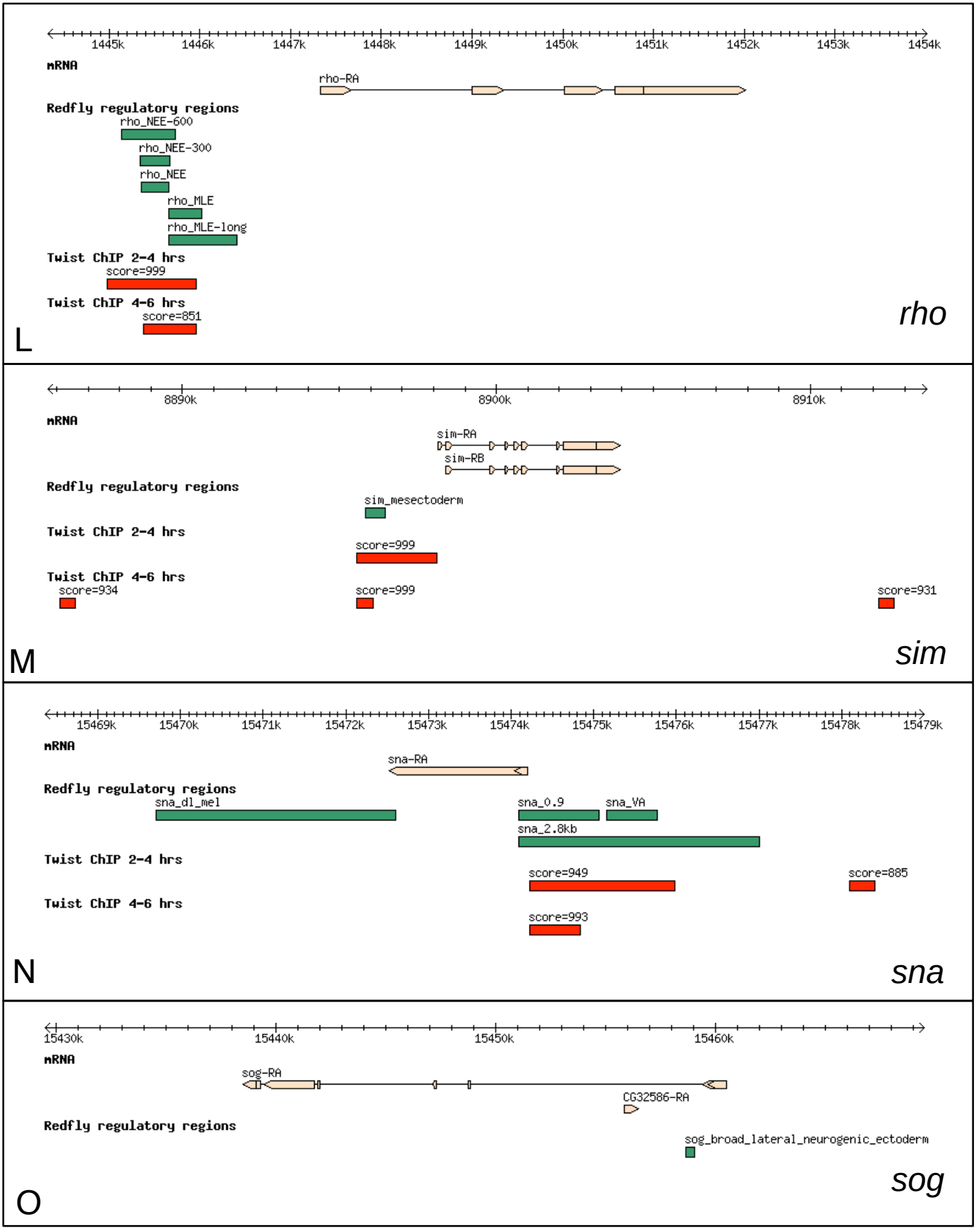
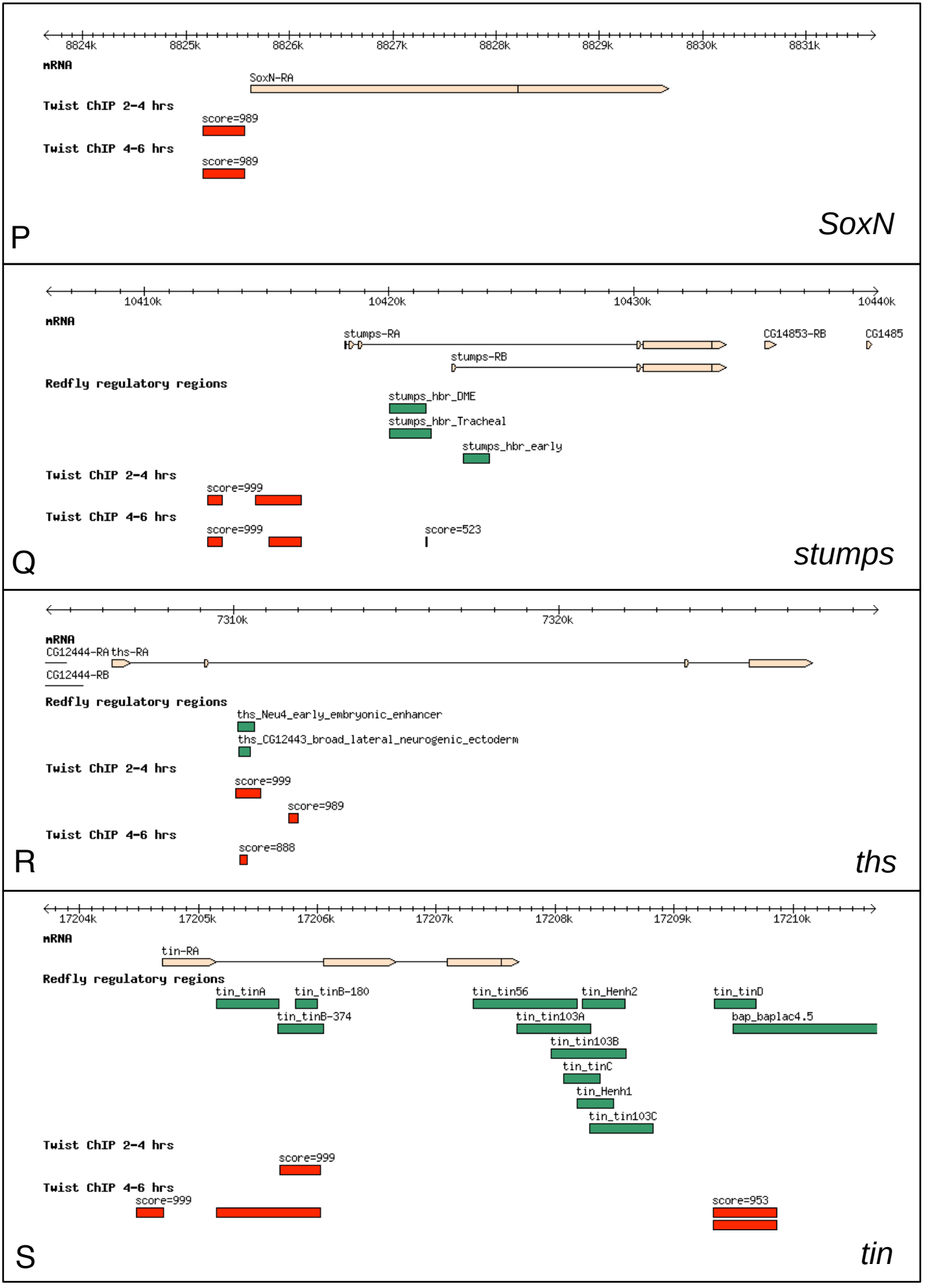
Suppl. figure 1

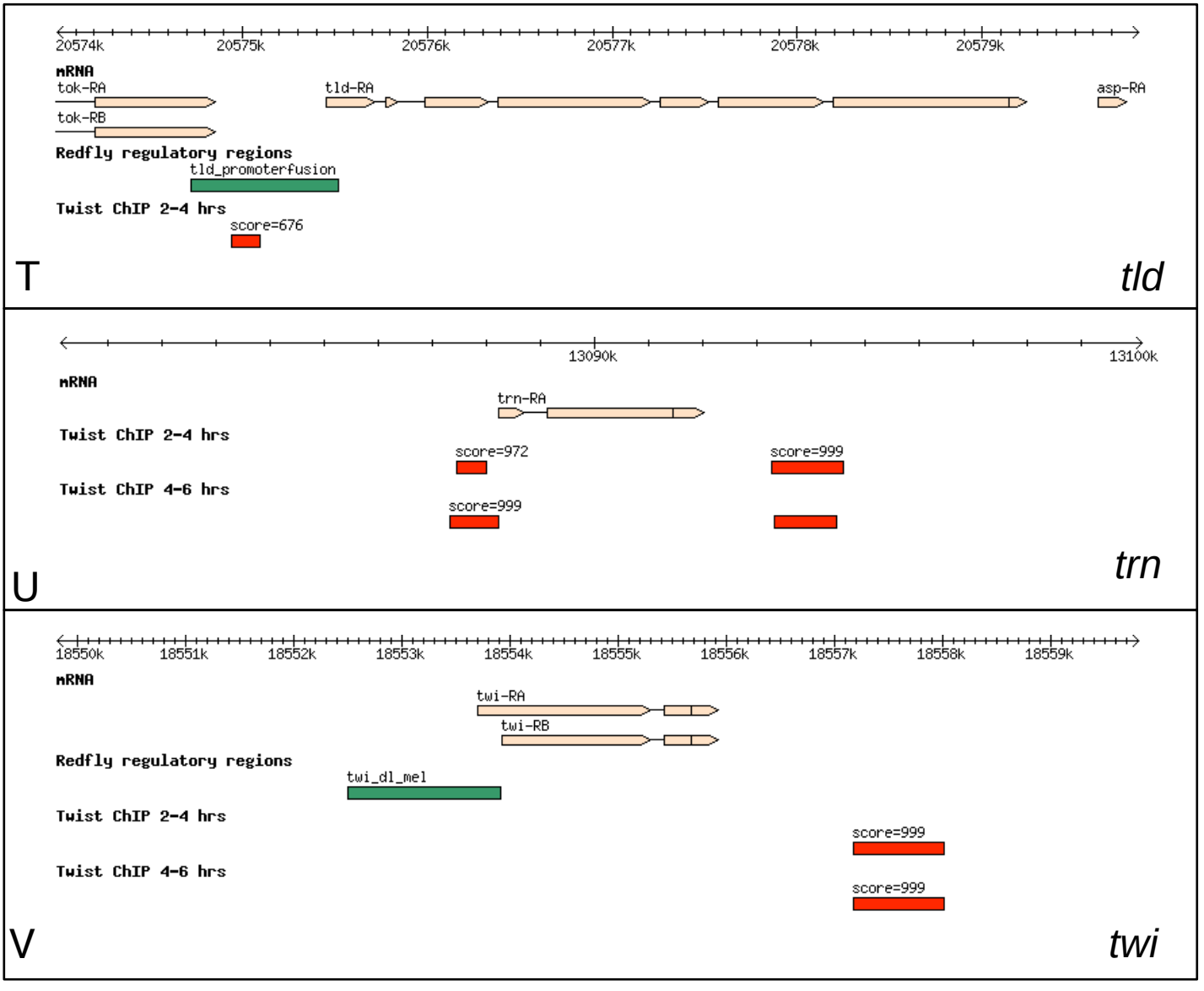


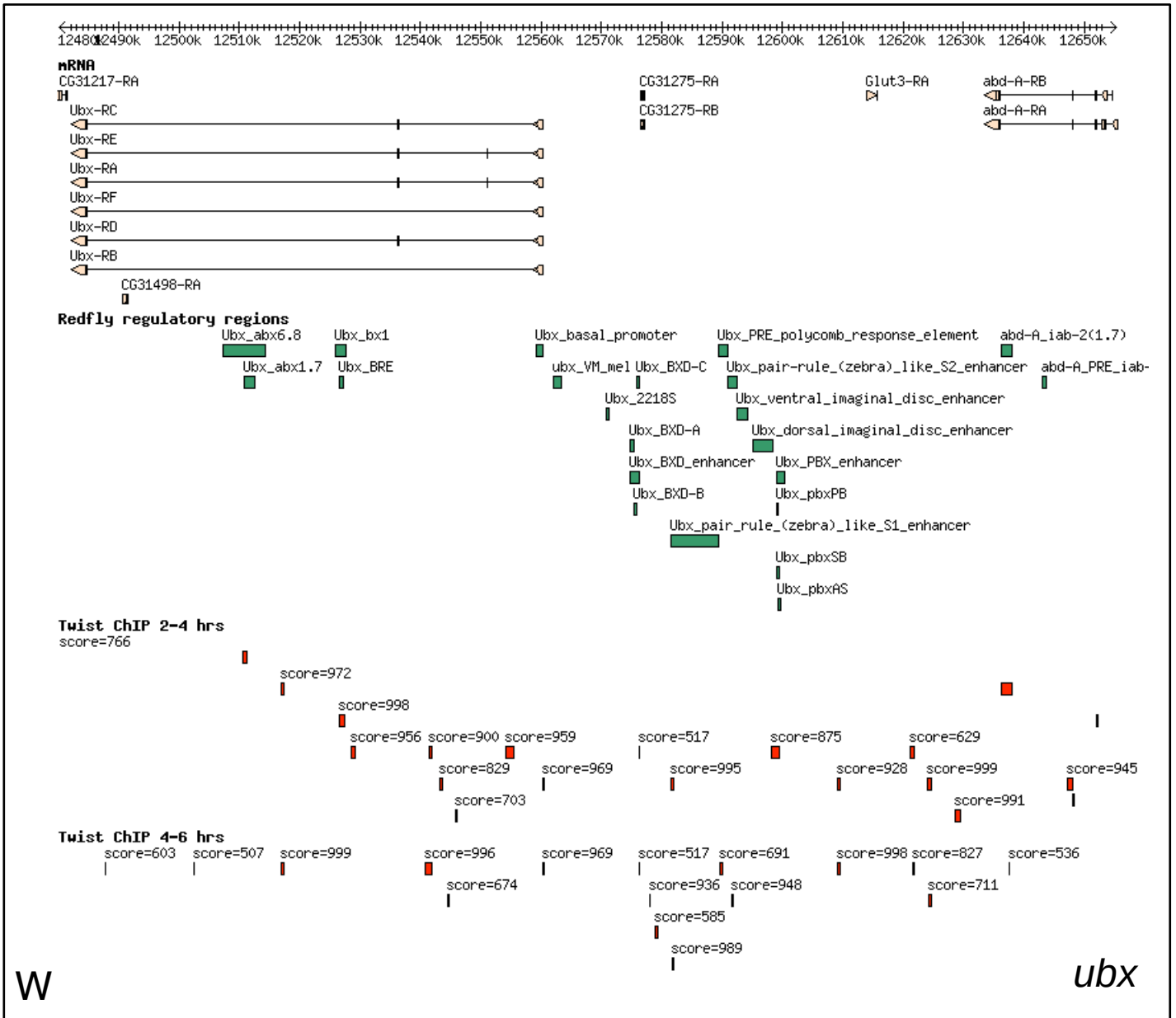

21862k
21863k
21864k
21865k
21866k
21867k
mRNA
HLHm7-RA
E(spl)-RA
Redfly regulatory regions
E(spl)_m8_dl_mel
E(spl)_PNC
HLHm7_m7-0.5
Twist ChIP 2-4 hrs
score=755
score=993
Twist ChIP 4-6 hrs
score=915
score=993
E
E(spl)
5490k
5491k
5492k
5493k
5494k
5495k
mRNA
eve-RA
Redfly regulatory regions
eve_stripe2
eve_proximal_promoter_inc._TATA
eve_ftz-like
eve_CQ/late_APR
eve_late_element_2_even7
eve_EL
eve_early_APR
Twist ChIP 2-4 hrs
score=819
F
eve
13869k
13870k
13871k
13872k
13873k
13874k
13875k
13876k
13877k
13878k
13879k
13880k
mRNA
htl-RC
htl-RA
htl-RB
Redfly regulatory regions
htl_early
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=999
score=778
G
htl
9776k
9777k
9778k
9779k
9780k
mRNA
Ilp3-RA
Ilp4-RA
CG32052-RA
Redfly regulatory regions
Ilp4_Mes3_early_embryonic_enhancer
H
Ilp4


15004k
15005k
15006k
15007k
15008k
15009k
15010k
15011k
15012k
15013k
mRNA
CG11552-RA
ind-RA
Redfly regulatory regions
ind_1.4
ind_moduleA
ind_moduleAB
ind_moduleBC
ind_moduleC+
Twist ChIP 2-4 hrs
score=998
score=982
Twist ChIP 4-6 hrs
score=722
score=984
I
ind
5430k
5440k
5450k
5460k
5470k
mRNA
CG12140-RA
Mef2-RC
Mef2-RF
Mef2-RA
Mef2-RD
Mef2-RB
Mef2-RE
CG15863-R
CG12130-
Redfly regulatory regions
Mef2_I-D[L]
Mef2_GE491
Mef2_I-D[s]
Mef2_II-E
Mef2_adepithelial_enhancer
Mef2_I-E
Mef2_minimal_composite_(#8)
Mef2_II-B[L]
Mef2_II-B[S]
Mef2_IIA277
Mef2_IIA237
Twist ChIP 2-4 hrs
score=999
score=993
score=992
Twist ChIP 4-6 hrs
score=875
score=999
score=868
J
Mef2
19492k
19493k
19494k
19495k
19496k
mRNA
Sox14-RB
Sox14-RA
Phm-RA
Phm-RB
CG30178-RA
CG3105-RA
Redfly regulatory regions
Phm_mesoderm
Twist ChIP 2-4 hrs
score=999
K
Phm


1445k
1446k
1447k
1448k
1449k
1450k
1451k
1452k
1453k
1454k
mRNA
rho-RA
Redfly regulatory regions
rho_NEE-600
rho_NEE-300
rho_NEE
rho_MLE
rho_MLE-long
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=851
L
rho
8890k
8900k
8910k
mRNA
sim-RA
sim-RB
Redfly regulatory regions
sim_mesectoderm
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=934
score=999
score=931
M
sim
15469k
15470k
15471k
15472k
15473k
15474k
15475k
15476k
15477k
15478k
15479k
mRNA
sna-RA
Redfly regulatory regions
sna_dl_mel
sna_0.9
sna_VA
sna_2.8kb
Twist ChIP 2-4 hrs
score=949
score=885
Twist ChIP 4-6 hrs
score=993
N
sna
15430k
15440k
15450k
15460k
mRNA
sog-RA
CG32586-RA
Redfly regulatory regions
sog_broad_lateral_neurogenic_ectoderm
O
sog

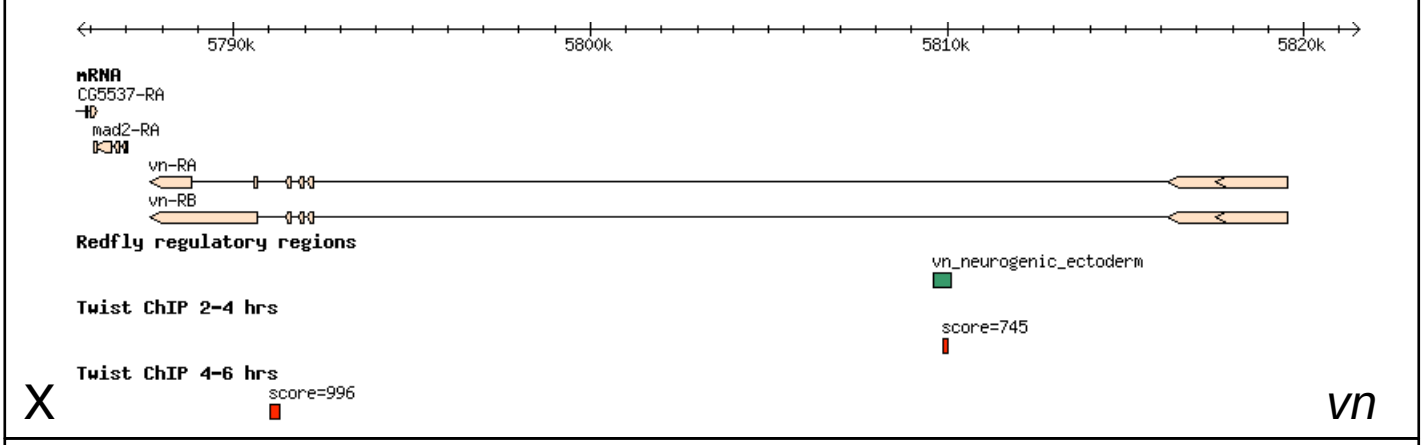

8824k
8825k
8826k
8827k
8828k
8829k
8830k
8831k
mRNA
SoxN-RA
Twist ChIP 2-4 hrs
score=989
Twist ChIP 4-6 hrs
score=989
P
SoxN
10410k
10420k
10430k
10440k
mRNA
stumps-RA
CG14853-RB
CG1485
stumps-RB
Redfly regulatory regions
stumps_hbr_DME
stumps_hbr_Tracheal
stumps_hbr_early
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=999
score=523
Q
stumps
7310k
7320k
mRNA
CG12444-RA
ths-RA
CG12444-RB
Redfly regulatory regions
ths_Neu4_early_embryonic_enhancer
ths_CG12443_broad_lateral_neurogenic_ectoderm
Twist ChIP 2-4 hrs
score=999
score=989
Twist ChIP 4-6 hrs
score=888
R
ths
17204k
17205k
17206k
17207k
17208k
17209k
17210k
mRNA
tin-RA
Redfly regulatory regions
tin_tinA
tin_tinB-180
tin_tinB-374
tin_tin56
tin_Henh2
tin_tin103A
tin_tinD
bap_baplac4.5
tin_tin103B
tin_tinC
tin_Henh1
tin_tin103C
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=999
score=953
S
tin


20574k
20575k
20576k
20577k
20578k
20579k
mRNA
tok-RA
tld-RA
asp-RA
tok-RB
Redfly regulatory regions
tld_promoterfusion
Twist ChIP 2-4 hrs
score=676
T
tld
13090k
13100k
mRNA
trn-RA
Twist ChIP 2-4 hrs
score=972
score=999
Twist ChIP 4-6 hrs
score=999
U
trn
18550k
18551k
18552k
18553k
18554k
18555k
18556k
18557k
18558k
18559k
mRNA
twi-RA
twi-RB
Redfly regulatory regions
twi_dl_mel
Twist ChIP 2-4 hrs
score=999
Twist ChIP 4-6 hrs
score=999
V
twi



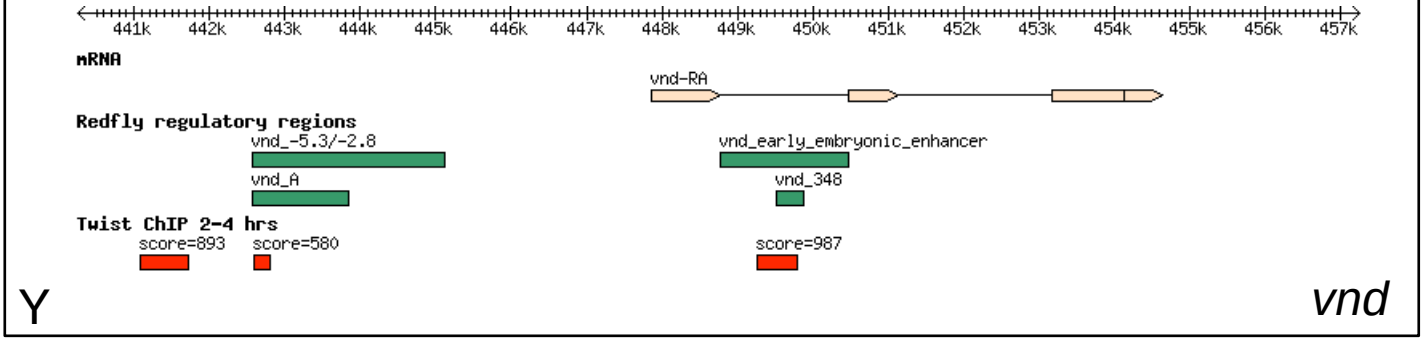
W

ubx



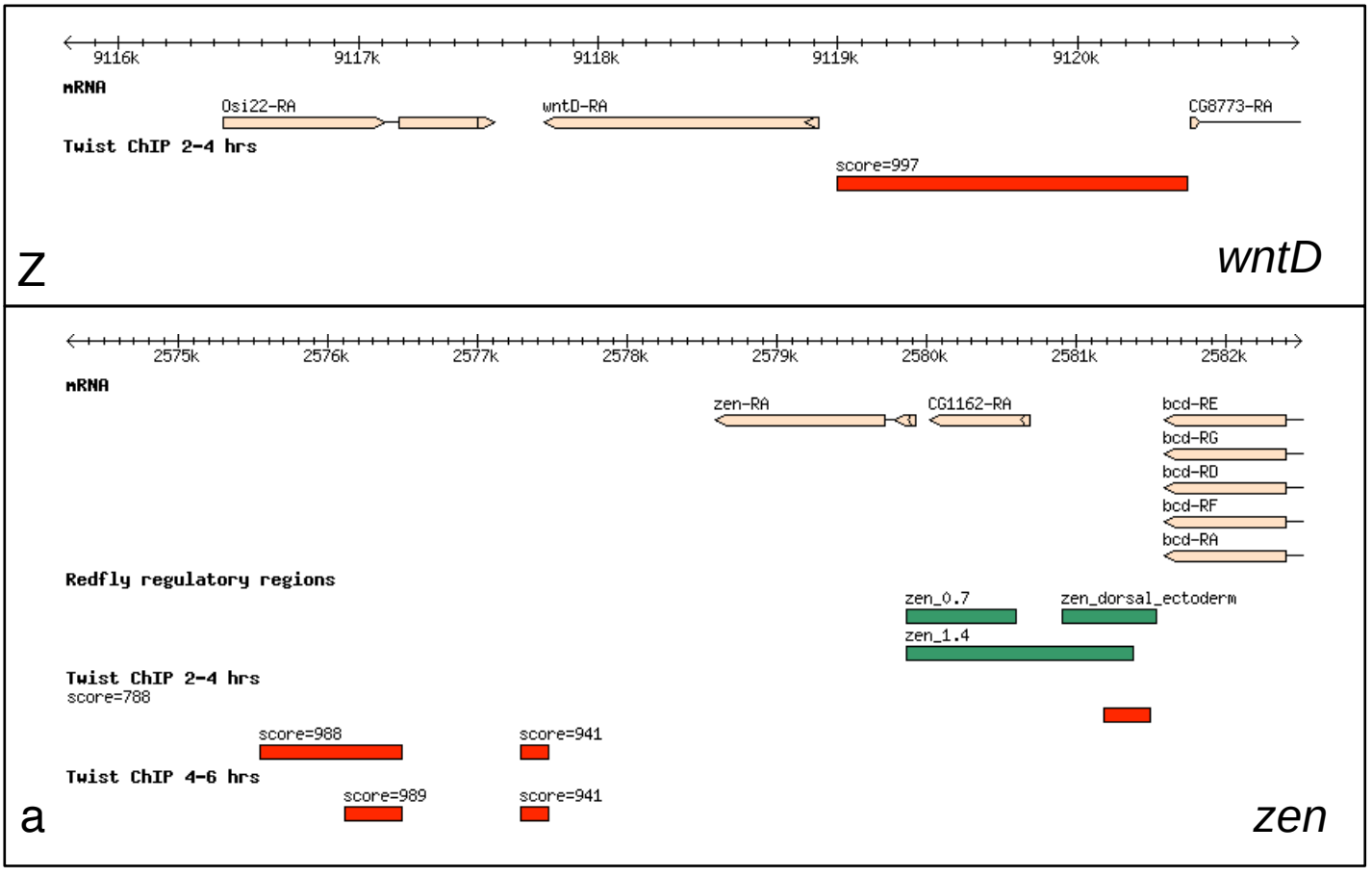
X

vn



Y

vnd



Suppl. Figure 1: Schematic overview of Twist-bound sequences overlapping known Twist binding sites (F,G,J,L,N,S,W) or in the vicinity of genes differentially expressed along the dorso-ventral axis of blastodermal embryos.

Twist binds to genomic sequences in the vicinity of the majority of loci expressed differentially along the dorso-ventral axis in the early blastoderm. Regions enriched at 2-4 hrs or 4-6 hrs, as reported by Tilemap are shown in red, with their associated significance score indicated above. Known regulatory sequences are indicated in green, with their corresponding RedFly identifier shown above.
These maps have been generated using the gBrowse tool available at Flybase.org.

Suppl. Table 1: Nine genes were previously identified as direct Twist target genes

Characterised direct Twist target gene	Number of Twist binding sites	Reference	Twist binding to region shown in
snail	Two	Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. 1992. *Genes Dev* **6**(8): 1518-1530.	Supple. Fig.1N
rhomboid	Two	Ip, Y.T., Park, R.E., Kosman, D., Bier, E., and Levine, M. 1992. *Genes Dev* **6**(9): 1728-1739.	Supple. Fig.1L
single-minded	Two	Kasai, Y., Stahl, S., and Crews, S. 1998. *Gene Expr* **7**(3): 171-189.	Fig. 3G Supple. Fig.1M
Ultrabithorax	Six	Qian, S., Capovilla, M., and Pirrotta, V. 1993. *Embo J* **12**(10): 3865-3877. Pirrotta, V., Chan, C.S., McCabe, D., and Qian, S. 1995. *Genetics* **141**(4): 1439-1450.	Supple. Fig.1W
tinman	Three	Lee, Y.M., Park, T., Schulz, R.A., and Kim, Y. 1997. *J Biol Chem* **272**(28): 17531-17541. Yin, Z., Xu, X.L., and Frasch, M. 1997. *Development* **124**(24): 4971-4982.	Supple. Fig.1S
even-skipped	Two	Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jimenez, F., Baylies, M.K., and Michelson, A.M. 2000. *Cell* **103**(1): 63-74.	Supple. Fig.1F
Mef2	One	Cripps, R.M., Black, B.L., Zhao, B., Lien, C.L., Schulz, R.A., and Olson, E.N. 1998. *Genes Dev* **12**(3): 422-434.	Supple. Fig.1J
mir-1	Several	Sokol, N.S. and Ambros, V. 2005. *Genes Dev* **19**(19): 2343-2354. Biemar, F., Zinzen, R., Ronshaugen, M., Sementchenko, V., Manak, J.R., and Levine, M.S. 2005. *Proc Natl Acad Sci U S A* **102**(44): 15907-15911.	Fig.3E
heartless	Two	Stathopoulos A, Tam B, Ronshaugen M, Frasch M, Levine M. 2004. *Genes Dev*. **18**(6): 687-99.	Supple. Fig.1G

Suppl. Table 1 (continued): Eleven genes were previously identified as direct Twist target genes

Characterised direct Twist target gene	Number of Twist binding sites	Reference	Twist binding to region shown in
vnd	Two	A regulatory code for neurogenic gene expression in the Drosophila embryo Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A and Michael Levine 2004 Development **131**(10):2387-94.	Supple. Fig.1Y
brk	Two	A regulatory code for neurogenic gene expression in the Drosophila embryo, Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A and Michael Levine 2004 Development **131**(10):2387-94.	Supple. Fig.1A

Suppl. Table 3: A literature survey of genes implicated in mesoderm gastrulation

Gene	***in vivo* binding of Twist to enhancer region(s)**
twist	yes
snail	yes
fog	yes
tribbles	yes
hkb	yes
cad	yes
shg	yes
Rho1	yes
Z600	yes
pbl	yes
dia	yes
pnut	yes
nullo	yes
T48	yes
dnt	yes
RhoGEF2	No *
cta	N/A **

* RhoGEF2 expression is strictly maternal and therefore not expected to be regulated by Twist

** This region is in heterochromatin and therefore could not be assayed

NimbleGen chip design

This ChIP-on-chip study has been conducted with 60mer oligonucleotide microarrays, custom produced using maskless array synthesizer (MAS) technology (Nuwaysir et al. 2002). We designed a custom array to ensure that each putative E-BOX site found in the *Drosophila melanogaster* genome could be assayed. The following two conditions were used as a starting point for the design:

1. EBOX sites should not be found within the last 10 bases of a 60mer oligonucleotide.
2. Each precipitated DNA sequence should be detectable by at least two neighboring oligonucleotide probes. In other words, one or more subsequent probes should be found within the maximum gap distance allowed. This condition allows differentiating between specific and unspecific hybridization events.

To design the 388706 oligonucleotidic probes, we followed a four-step strategy. First, we identified all E-BOX sites found within intergenic or intronic regions of the *Drosophila melanogaster* genome. Second, 60mer sequences were designed for each of them. Third, redundant probes were discarded. Finally, additional probes were designed to fulfill the second condition mentioned above.

To identify all potential E-BOX sites (matching the CANNTG motif) sites in the *Drosophila melanogaster* genome (assembly version 2, release 4.0), we masked the genome for low-complexity regions and repeats using RepeatMasker, version 2004/03/06 (Smit 1996-2004). To satisfy the probe design conditions, only sites with at least 100 bases of unmasked sequence on either side were kept, resulting in 550800 E-BOX sites. We then excluded sites found in exon boundaries (extracted from release 4.0 Flybase GFF files). The remaining 398292 E-BOX sites represented the complete set of motifs to be assayed with this array platform.

In the second step, a 60mer was designed for each E-BOX sites using a modified version of ArrayOligoSelector v3.8 (Bozdech et al. 2003). For each E-BOX site, a 94 bases long genomic region, centered on the test site, was used as input for ArrayOligoSelector (evaluating a set of 34 different 60-mers for each putative E-BOX site). For each of the 391084 obtained probes an occurrence score, reflecting the cross-hybridization potential of a probe, was computed (defined as the average occurrence (perfect match) in the genome of each 17-mer composing the 60-mer). Probe with and occurrence score > 2 or those that could not be synthesized for technical reasons (160 sequences) were discarded.

In the third step we recursively reduced the set of probes required to assay all remaining E-BOX sites by increasing the maximum gap parameter until the probe set size matched the chip capacity. Probes were declared redundant when multiple E-BOX sites were close enough to each other be detected by a single oligonucleotide. The algorithm converged to a gap of 290 bases; in addition to E-BOX containing sequences, an additional set of 53570 “neighboring” probes was required to satisfy condition 2.

These additional probes were designed following the same procedure described above (ArrayOligoSelector followed by occurrence score and synthesizability filtering). Note that this additional probe set (53570 probes) has been obtained from a larger set of 60719 candidate probes; i.e. reflecting situations where the additional probe could be designed at either side of the existing probe. Whenever a choice was possible, we selected the best probe based on its occurrence score.

Hybridization of 60mer oligonucleotide arrays

(Stolc et al. 2004)

Prehybridization

1. *For each microarray used prepare the following hyb and prehyb solutions:*

Component	***Prehyb. Solution***	***Hyb. Solution***
*labeled cDNA***	*none*	*2ug*
(6 Oligo (Cy5 and Cy3), 0.1mM	*none*	*0.4ul each*
Herring Sperm DNA, 10mg/ml	*1.0 ml, add later !*	*0.4ul*
BSA, 50mg/ml	*1.0 ml*	*0.4ul*
2X MES Hybridization Buffer	*100 mL*	*20ul*
Water	*to 300 ml*	*to 40ul*
Total	*300ml*	*40 ul*

2. Warm the prehybridization solution to 45°C, heat the Herring sperm DNA for 5 min at 100 degr and add to the prewarmed solution.
3. Prehybridize the array(s) in the prehybridization solution for 20 minutes at 45°C.
4. Transfer the slide rack into a container with destilled water and move up and down 20x.
5. Quickly transfer the slide rack into the centrifuge and dry the slides at 800 g for 4 min.

Hybridization

6. Resuspend the speed-vac dried probes in the hybridization solution, place into a 95 °C heat block and incubate them for 3 minutes.
7. Centrifuge hybridization solution at > 12.000g for 0.5 min at RT in a tabletop centrifuge. Keep at 65 °C afterwards until use.
8. Apply the hybridization solution onto a 30x25 mm coverslip, place an array on top and seal in the hybridization chamber. (Add 2-3 15 ul of 0.6x MES Hybridization Buffer into the chamber to keep the chamber humid.)
9. Incubate in a waterbath at 65 °C for 16 hrs.

Washing steps

10. Prewarm ca. 800 ml of stringent wash buffer (SWB) buffer to 45 °C and add DTT
11. Prepare 200 ml of the remaining wash buffers and add DTT where required:

Buffer 1	**Buffer 2**	**Buffer 3**	**Buffer 4**	**Buffer 5**
NSWB (to remove cover slips)	NSWB 2 min	45 degr SWB 20 min 3 changes	NSWB 2 min	0.2x SSC 2 min in glass chamber

12. Transfer the slides from the hyb chamber into NSWB and protect the container from light.
13. After the cover slips have fallen off, transfer the slides into fresh NSWB buffer and wash under agitation for 3 min.
14. Transfer the slides into the prewarmed container with SWB and wash the slides under agitation for 20 min. Keep the temperature at 45 °C by exchanging the buffer every 5 minutes !
15. Transfer the slides into a container with NSWB and wash for 3 minutes.
16. Finally wash the slides in 0.2x SSC for 2 minutes.
17. Quickly transfer the slide rack into the centrifuge and dry the slides at 800 g for 4 min.
18. Scan the slides a.s.a.p. with 5 μm resolution.

Recipes for Buffers (Stolc et al. 2004)

2X MES Hybridization Buffer *(100mM MES, 1M [Na+], 20mM EDTA, 0.01% tween20)*

41.5mL	*12X MES Stock Buffer (below)*
88.5 mL	*5M NaCl*
20.0mL	*0.5M EDTA*
0.5mL	*10% Tween20*
99.5mL	*Water (not DEPC WATER!)*
250mL	

Combine listed components and bring to volume in a graduated cylinder
Sterile Filter, 0.2um. Store protected from light and at 4°C.

NSWB Non-Stringent Wash Buffer *(6X SSPE, 0.01% Tween-20)*

300mL	*20X SSPE*
1.0mL	*10% Tween-20*
698mL	*water (not DEPC Water!)*
1000ml + 1 ml 1M DTT before use	

Combine listed components and bring to volume in a graduated cylinder
Sterile filter, 0.2μm. Store at Room Temperature.

SWB Stringent Wash Buffer *(100mM MES salt and free acid solution (see 12X MES below), 0.1M [Na+], 0.01% Tween-20):*

83.3mL	*12 X MES Stock Buffer (above)*
5.2mL	*5M NaCl*
1.0mL	*10% Tween-20*
910.5mL	*water (not DEPC Water!)*
1000ml + 1 ml 1M DTT before use	

Combine listed components and bring to volume in a graduated cylinder
Sterile filter, 0.2μm. Store protected from light and at 4°C.

12X MES Stock Buffer, 1L (1.22M MES, 0.89M [Na+])

35.2g	*MES, free acid monohydrate*
96.65g	*MES, sodium salt*
molecular biology grade water to volume (not DEPC Water!)	
500ml	

Combine MES salt and free acid, bring to volume with water
Stir until well blended, final pH should be 6.5 – 6.7
Sterile filter, 0.2μm. Store protected from light and at 4°C.

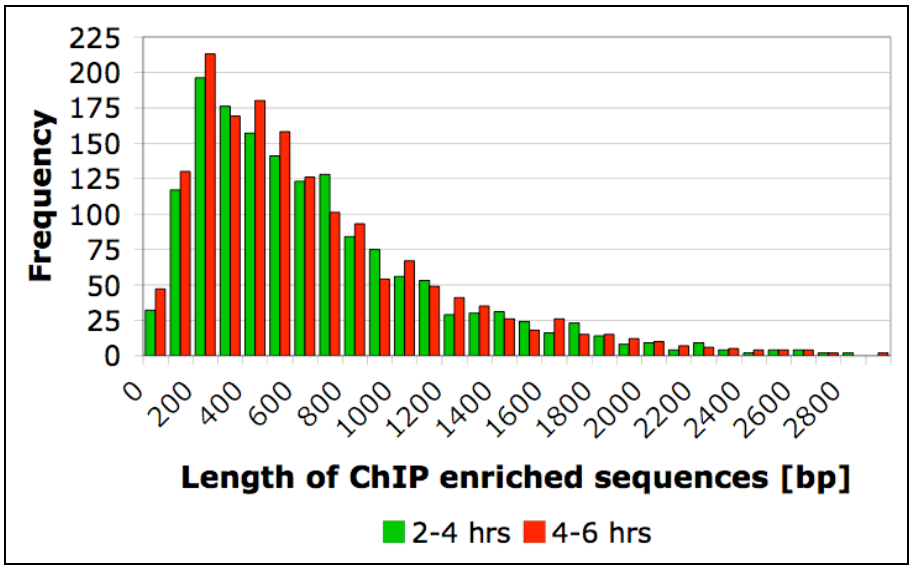
Identification of significantly enriched regions using the Tilemap algorithm (Ji and Wong 2005)

The Tilemap algorithm was used with the following adjusted parameters, while otherwise accepting the default settings:

- Range of test-statistics = 1
- Zero cut = 0.001
- Method to combine neighboring probes = 0
- Expected hybridization length = 2
- Half-window size = 1
- Selection offset = 1

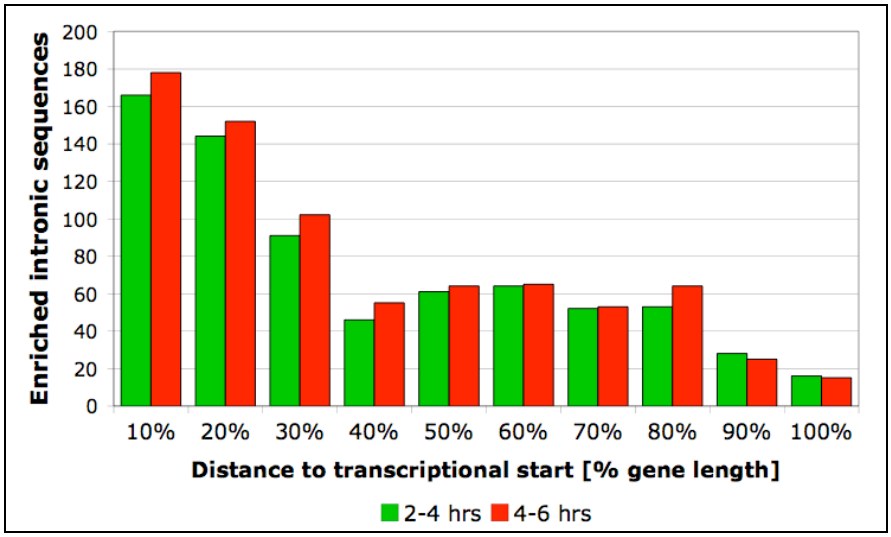
The Ebox-array does not cover coding or repetitive sequences of the genome. Therefore tilemap regions spanning introns were split into two distinct regions (indicated by the _sub# suffix of the region identifier). Split regions were excluded from the calculation of fragment size distributions or the analysis of relative positioning of enriched sequences relative to gene loci or transcriptional start sites.

As all known Twist binding sites except within the *eve*-cardiac enhancer are detected within regions receiving a score of at least 875 at one or both developmental timepoints, this threshold was chosen as a stringent significance cut-off for all further analysis.



Suppl. figure 2: Size distribution of significant tilemap regions

Tilemap regions (score ≥ 875) show a very similar sequence length distribution at both 2-4 (green) and 4-6 hrs (red) of development, with a median size of ca. 650 bps.



Suppl. figure 3: Positioning of ChIP-enriched tilemap regions within introns with respect to the transcriptional start site

The positioning of ChIP-enriched sequences within gene loci is skewed towards the transcriptional start site of the genes. The distance between the start of each tilemap region and the start coordinate of the respective locus was expressed as a percentile to allow comparisons of loci with very different (intron) lengths.

Assignment of target genes

Significantly enriched Twist-bound sequences were assigned to the most likely target genes using a scoring scheme taking into account both the distance between an enriched region and gene loci in its vicinity as well as loss-of-function expression profiling data and information about the expression domains of the candidate target genes. While this is similar to the approach used in a previous study identifying targets of Mef2 (Sandmann et al. 2006), changes were made to account for the severe consequences of the *twist* mutant phenotypes (see below).

Distance score

The association between a positive fragment and a gene was scored based on the distance between the two. Because fragments may fully or partially overlap with the gene in question and can have different lengths, each basepair of a fragment was first scored separately. Basepairs located within the gene were assigned a perfect score of 1 whereas all other basepairs receive the score 1/(1+exp(0.25*(d-15))), where d is the distance between the basepair and the nearest end of the gene. To calculate the score for an entire fragment, the average score of all basepairs was calculated.

Expression score

Expression profiling data comparing $twist^{ey53}$ loss of function mutant embryos with stage matched wild-type embryos was used to assay whether a gene is genetically downstream of *twist*. Expression data was collected assayed at four consecutive one-hour time-periods using both cDNA microarrays as well as INDAC oligonucleotide microarrays, together covering the vast majority of all annotated *Drosophila* genes. At each time point, q-values were calculated for all probes using SAM (Tusher et al. 2001; Saeed et al. 2003). To down-weigh genes that change sporadically at a single time point, q-values were averaged over pairs of neighboring time points. For each probe, the minimum average q-value was identified and the average expression ratio was calculated for the corresponding time points. A probe was considered to change significantly only if the minimum average q-value

was below 0.1 and the corresponding absolute expression ratio was over 0.5; for these probes an expression score was calculated as 1-4*qvalue. An expression score of 0.1 was assigned for the probes that did not change significantly. For each gene, the expression score was defined from the best scoring probe.

Supporting score

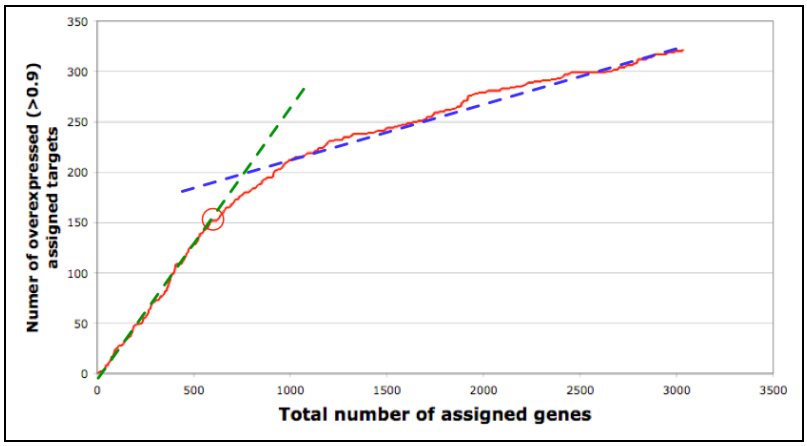
Supporting evidence was gathered from different sources and scored as follows: The highest score (1.0) was given to genes with known early mesodermal or embryonic muscle expression. Genes with respective annotations in flybase or in the BDGP in situ database (Tomancak et al. 2002)were extracted and supplemented with information manually collected from the literature. Additionally, genes without expression in early mesoderm or any muscle derivative were extracted from the BDGP database and down weighed by assigning a score of 0.2. Finally, for genes without any information about expression in Drosophila but with known muscle expression of mouse orthologs (Delgado et al. 2003; Kuninger et al. 2004; Masino et al. 2004; Tomczak et al. 2004), an intermediate score of 0.8 was chosen.

Combined scores

Twist loss-of-function mutants lack any mesoderm. The loss of the complete germ layer in *twist* mutant embryos differs markedly from the phenotype encountered in *Mef2* mutant embryos. All genes with known expression in this germ layer are expected to show reduced expression levels when compared to wildtype controls. Nevertheless, not all of these changes are reflected in the expression profiling data, e.g. due to additional expression in other germ layers/ tissues of the embryo or low overall expression levels. We therefore adjusted the scoring algorithm to reflect the *twist* phenotype: In case of known mesodermal genes, the expression profiling data represents a redundant source of information. The combined score for a target gene was therefore derived by multiplying the scores for distance and supporting evidence. For the majority of gene loci, information about the expression domains is lacking. For these candidate loci, the combined score was instead

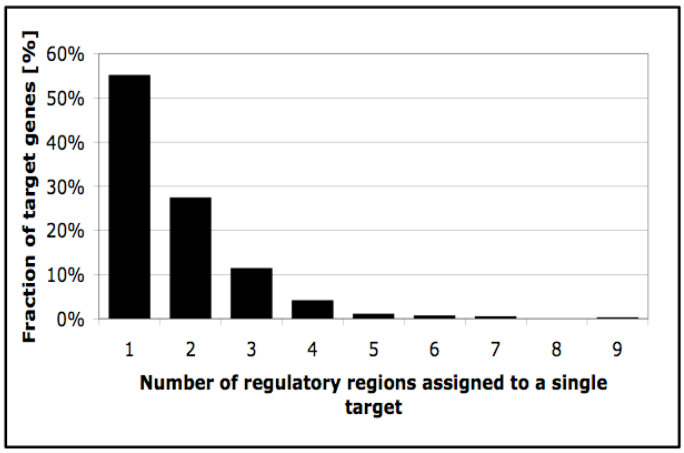
derived by multiplying the scores for distance and expression evidence.

In case of missing scores for expression or supporting evidence, the value 0.6 was used (corresponding to the midpoint the minimum score of 0.2 and the maximum of 1). To choose a score-cutoff for defining the high-confidence set of targets, we evaluated the frequency of assigning target genes significantly overexpressed in *Toll10B* mutant embryos. In this genetic background, a dominant active form of the Toll receptor leads to ventralization of the embryo and triggers the expression of mesoderm specific genes in all cells (Furlong et al. 2001).



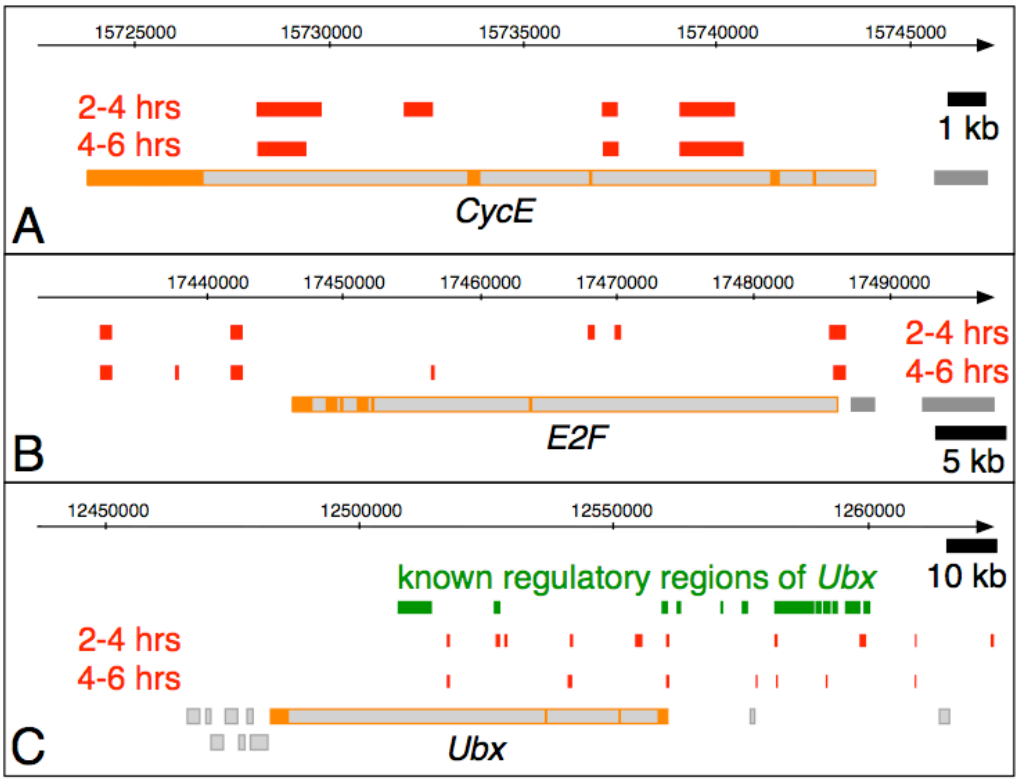
Suppl. figure 4: A stringent threshold for high-confidence target assignments was chosen by evaluating the frequency assigned loci significantly upregulated in *Toll10B* mutant embryos.

The number of assigned target genes significantly enriched in *Toll*^{10B} mutant embryos (expression score > 0.9) was plotted against the total number of genes assigned at the same total score cutoff. Among the 494 target genes with the highest scores (equivalent to a score cutoff of 0.94, red circle), genes misregulated in Toll10B mutants are assigned with high frequency (green dashed line). This threshold was therefore chosen to identify high-confidence target genes.



Suppl. figure 5: 50% of Twist target genes have more than one Twist-bound CRM

ChIP-enriched regions were assigned separately to likely candidate genes. Almost half of the assigned Twist target genes are associated with two or more significantlyl enriched ChIP regions.



Suppl. figure 6: Examples of target gene loci assigned to multiple Twist-enriched tilemap regions

Among the loci associated with more than one ChIP-enriched regions are the *CycE*, *E2F* and *Ubx* genes (A-C). All three possess multiple intronic or closely positioned intergenic enriched sequences, several of which are differentially bound over time. Some of the regions identified in the vicinity of the *Ubx* locus overlap with known regulatory regions of this gene (C, green).

Quantitative real-time PCR

The relative enrichment of known Twist or Mef2 binding after chromatin immunoprecipitation with either specific or mock antisera was evalulated by quantitative real-time PCR (qPCR). Primer pairs were designed flanking predicted or known binding sites (see below for oligonucleotide sequences). Two microlitres of ChIP or mock amplicons (PCR amplification reaction A, diluted 1:50) were assayed using the following reaction setup:

Quantitative real-time polymerase chain reaction, reaction setup

Volume	Reagent
2.5 µl	2.5 µM primer A
2.5 µl	2.5 µM primer B
12.5 µl	SYBR green reaction mix
2 µl	ChIP or mock amplicon
5.5 µl	Water

For each primer pair, a standard curve was determined in duplicate using serial dilutions from ca. 40 ng/ml sheared genomic DNA (1:10 to 1:10.000 dilutions). The amplification reactions were performed and recorded on an abi7500 real-time PCR system (Applied Biosystems, Foster City, USA) using standard settings for absolute quantitation. Dissociation curves were recorded after each run to evaluate the amplification of uniform products. The results were converted into enrichment ratios:

ChIP-enrichment = $[ChIP_{known\_binding\_site}/ChIP_{control\_region}]$ and mock-enrichment = $[mock_{known\_binding\_site}/mock_{control\_region}]$ by referring to the respective standard curve for each primer pair.

Oligonucleotide primers used for quantitative real-time assays:

a) Verification of predicted Tinman sites

Mef2_enhancer	**2R**	**5447301**	**5447310**
Forward	ACTCGACTGCGGATTCTCTG		
Reverse	AAACAACCGCACACGGATAC		
Tl6.5_enhancer	**3R**	**22621301**	**22621310**
Forward	ACTGCAACTGCGAACTGCTA		
Reverse	GCGTCGAATGGTTTTTGTTT		
18w	**2R**	**15618137**	**15618146**
Forward	CATTCCCTCGCATTTTGAAT		
Reverse	TGGGTTTTCCTCCTTTTTCC		
gene_desert	**2L**	**19318616**	**19318625**
Forward	GCGGCAATTAAGATTTCCTTT		
Reverse	GGCTGGGATCTACAGTGAGC		
Doc3	**3L**	**8977692**	**8977701**
Forward	CCTTTTCCATCCCGTCCTAC		
Reverse	AAGACACTGTCGCCTTCGAG		
E2f	**2L**	**15740133**	**15740142**
Forward	TCTAAAAGGATGCCCACAGC		
Reverse	GTCCGACTGGCGATTTGT		
epac	**2R**	**2301110**	**2301119**
Forward	TGATCTCGATGGGTCAGATG		
Reverse	GCTGTCGGATGTCTGAATCTC		
gsb-n_gsbnE_enh	**2R**	**20561434**	**20561443**
Forward	GAGTCCCTGCGATAATGAGC		
Reverse	TGCATGGCAAGTTCTATTGC		
hh	**3R**	**18972433**	**18972442**
Forward	CAGACGCAGACGAGTCACAT		
Reverse	AAGAAAATCCCCCTGTGGAC		
Hs6st	**3R**	**15823580**	**15823589**
Forward	TTCCCTTTGTTTTCGTACTGC		
Reverse	ACATTCACCGGACGACTTTC		
jeb	**2R**	**7629595**	**7629604**
Forward	TTTGACAGGAGCAAGGGACT		
Reverse	TTTCATACACTGGCGGTTGA		
nuf	**3L**	**14161711**	**14161720**
Forward	GCGATCTTTCGAGCAGGTAG		
Reverse	AAGCGAGTAAAGTGCGGAGA		
Sa-2	**3L**	**1414866**	**1414875**
Forward	GCTGACAGGGCACGACTAAT		
Reverse	GGAGTAGTGGAGGCCAAGG		
syx7	**3L**	**21387351**	**21387360**
Forward	GAGCCCGATTCAAATTCAAG		

Reverse	GAAAATGAGCGAAGGAATCG

b) Verification of predicted Dorsal sites

ind	**3L**	**15004888**	**15005049**
Forward	TTTAACAGGCCCAAAGAACC		
Reverse	TTTCTTTTTGAATTGGCCTCA		
rho	**3L**	**1445493**	**1445654**
Forward	GAATTTCCTGATTCGCGATG		
Reverse	CAGGACAGGACGTTGATTCC		
dpp	**2L**	**2456468**	**2456545**
Forward	AATGCGAATGAAGAGCCAGT		
Reverse	TCTAGGGATCGGCAGGTATG		
pbl	**3L**	**7883455**	**7883616**
Forward	AAGTGCCGAGACTCACAGGT		
Reverse	CAGCCAGCGAGGAAAAGTAG		
hkb	**3R**	**174262**	**174423**
Forward	TAGGTTTGGACTTGGGCTTG		
Reverse	GTTATGAGTGCCGCATTGTC		
CG8117	**X**	**15480740**	**15480901**
Forward	CATAAAAGGGGCGCAGATAA		
Reverse	ACTGCTTCGTCCTGGTCCT		
wntD	**3R**	**9119140**	**9119449**
Forward	GAATGAAGCCCAGTCGAGTC		
Reverse	ACCAGTCCAAAACCCAAACA		
phm	**X**	**18518755**	**18519064**
Forward	CTTCCTTTCCCACTCGCTAA		
Reverse	GCAGCCCTCTGTAGAAATGC		
CG13897	**3L**	**718794**	**719103**
Forward	GCGAAACTAGGCAGAAAAGG		
Reverse	TCATTCCCATTTTCCAGAGC		
CG5718	**3L**	**11899212**	**11899521**
Forward	TCGAGCAAACAGACGAGAAA		
Reverse	TGTCCTTTGAGCGCAATTAG		
Delta	**3L**	**14115258**	**14115567**
Forward	GTTGGCATTGTCTTGGCTTT		
Reverse	TGACTTTTGTTGAGCCTTGC		

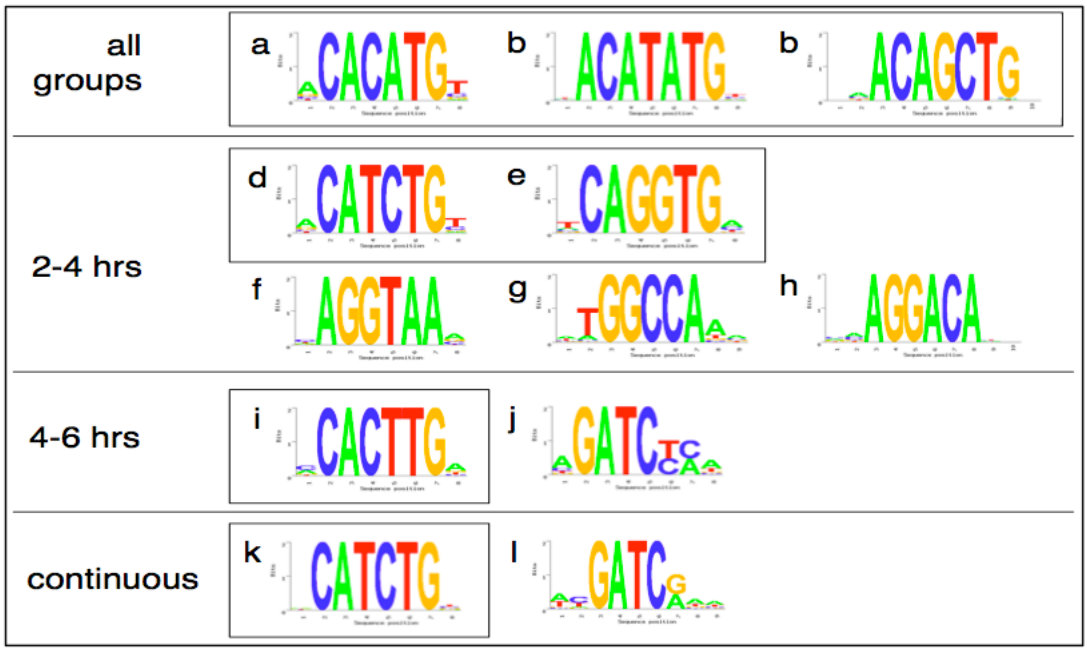
De novo pattern discovery

De novo pattern discovery in Twist bound regions was conducted using the RSAT package (van Helden 2003). In detail, enriched patterns were identified using the “oligo-analysis” RSAT tool with the following parameters: word length = 7, E-value cut-off < 0.01. As input sequences, we used either the 60mer oligonucleotide sequences present on the microarray chip features or the full Tilemap regions. In both cases, the input sequences were split into three temporal groups (2-4 hrs, 4-6 hrs or continuously bound).

First, the pattern discovery was performed for each group separately using either the complete chip feature set (in case of 60mer input sequences) or the *Drosophila melanogaster* genome (in case of Tilemap regions as input sequences), masked for repeats and exons.

Second, differential over-representation of sequence motifs was evaluated by comparing each temporal group (60mer or tilemap sequences as input, as above) to the other two temporal groups.

Evaluation of discovered motifs was done separately for each temporal group / comparison. To facilitate the anlysis of the large number of discovered patterns, we subdivided the motifs into two groups: a) patterns containing the CANNTG signature and b) others. Overlapping sequences were assembled into groups and the WebLogo tool was used to generate sequence logos (Crooks et al. 2004).



Suppl. figure 7: De novo motif discovery using RSAT

Unbiased analysis of overrepresented motifs identifies differential enrichment of E-box motifs along with novel oligonucleotide assemblies: Motifs were discovered either in each of the three temporal groups independently (top row, “all groups”) or show temporal specificity (2-4 hrs, 4-6 hrs, continuous). Motifs within the black boxes conform to the E-box consensus (CANNTG). Several motifs resemble known transcription factor consensus motifs: a,b) Twist, d,k) Twist/Daughterless, e) Snail or Daughterless, h) Tramtrack, i) Tinman.

Identification of over-represented known TF sites

Twist bound regions (as reported by Tilemap) were searched for over-represented known transcription factor binding motifs (TFBSs) using the Clover program (Frith et al., 2004). A set of 104 matrices for transcription factors from *Drosophila melanogaster* was obtained from FlyReg(Bergman et al. 2005), Transfac (Matys et al. 2006), Jaspar (Vlieghe et al. 2006) and the literature. Clover was used to search each temporal group (2-4 hrs, 4-6 hrs or continuously bound) using default parameters with all matrices and reported 44 different transcription factor signatures to be over-represented in at least one of the three temporal groups.

For each matrix reported as significantly enriched by Clover, we performed a second motif prediction round using the Patser tool (G.Z.Hertz).

The hits reported by Patser were binned according to their score (lower score cutoff =4, bins in steps of 0.25) and an enrichment ratio (ER) was calculated for every bin as: [number of TFBSs per base in the test set] / [number of TFBS per base found in full *Drosophila melanogaster* genome masked for repeats and exons].

A threshold score for Patser was chosen based on the following two criteria:

(1) at the chosen threshold hits in at least 10% of the input sequences must be reported and

(2) the cumulative enrichment fold above this score threshold must exceed 1.5.

We consider all Clover-reported matrices passing the Patser-based filter above as significantly enriched.

References

Bergman, C.M., Carlson, J.W., and Celniker, S.E. 2005. Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. *Bioinformatics* **21**(8): 1747-1749.

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**(2): R9.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**(6): 1188-1190.

Delgado, I., Huang, X., Jones, S., Zhang, L., Hatcher, R., Gao, B., and Zhang, P. 2003. Dynamic gene expression during the onset of myoblast differentiation in vitro. *Genomics* **82**(2): 109-121.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., and Scott, M.P. 2001. Patterns of gene expression during Drosophila mesoderm development. *Science* **293**(5535): 1629-1633.

Ji, H. and Wong, W.H. 2005. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**(18): 3629-3636.

Kuninger, D., Kuzmickas, R., Peng, B., Pintar, J.E., and Rotwein, P. 2004. Gene discovery by microarray: identification of novel genes induced during growth factor-mediated muscle cell survival and differentiation. *Genomics* **84**(5): 876-889.

Masino, A.M., Gallardo, T.D., Wilcox, C.A., Olson, E.N., Williams, R.S., and Garry, D.J. 2004. Transcriptional regulation of cardiac progenitor cell populations. *Circ Res* **95**(4): 389-397.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., and Wingender, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**(Database issue): D108-110.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F., and Green, R.D. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* **12**(11): 1749-1755.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**(2): 374-378.

Sandmann, T., Jensen, L.J., Jakobsen, J.S., Karzynski, M.M., Eichenlaub, M.P., Bork, P., and Furlong, E.E. 2006. A temporal map of transcription factor

activity: mef2 directly regulates target genes at all stages of muscle development. *Dev Cell* **10**(6): 797-807.

Smit, A., Hubley, R & Green, P. 1996-2004. *RepeatMasker Open-3.0*. In.

Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., Bussemaker, H.J., and White, K.P. 2004. A gene expression map for the euchromatic genome of Drosophila melanogaster. *Science* **306**(5696): 655-660.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., and Rubin, G.M. 2002. Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* **3**(12): RESEARCH0088.

Tomczak, K.K., Marinescu, V.D., Ramoni, M.F., Sanoudou, D., Montanaro, F., Han, M., Kunkel, L.M., Kohane, I.S., and Beggs, A.H. 2004. Expression profiling and identification of novel genes involved in myogenic differentiation. *Faseb J* **18**(2): 403-405.

Tusher, V.G., Tibshirani, R., and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**(9): 5116-5121.

van Helden, J. 2003. Regulatory sequence analysis tools. *Nucleic Acids Res* **31**(13): 3593-3596.

Vlieghe, D., Sandelin, A., De Bleser, P.J., Vleminckx, K., Wasserman, W.W., van Roy, F., and Lenhard, B. 2006. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res* **34**(Database issue): D95-97.