

Supplemental material to:

**Uncovering *cis*-regulatory sequence requirements for
context specific transcription factor binding**

J. Omar Yáñez-Cuna¹, Huy Q. Dinh^{*2}, Evgeny Z. Kvon^{*1}, Daria Shlyueva^{*1} & Alexander Stark^{1#}

¹ The Research Institute of Molecular Pathology (IMP), 1030 Vienna, Austria

² Gregor Mendel Institute of Molecular Plant Biology (GMI), Austrian Academy of Sciences; Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories (MFPL), 1030 Vienna, Austria

* These authors contributed equally to this work.

To whom correspondence should be addressed (stark@starklab.org).

Table of Contents

Supplemental Discussion

Context-specific binding sites are associated with distinct biological functions. 2

Prediction performance on randomized datasets. 2

Identification of sequence requirements for context-specific Twist binding 3

Reporter and primer sequences for the *in vivo* validation experiments 4

Supplemental Figures

Fig. S1. Regulatory motifs ranked according to their predictive power 7

Fig. S2. The Twist motif is evolutionarily conserved in early and late Twist binding sites. 8

Fig. S3. Evolutionary conservation of putative Twist partner motifs 9

Fig. S4. Biological replicates for Twist ChIP-qPCR for the *cact* enhancer 10

Fig. S5. Motif requirements for context-specific binding sites of Twist 11

Fig. S6. Partner TF motif-counts per site for early Twist binding sites 12

Supplemental Tables

Table S1. TFs binding sites differ between time-points and cell types 13

Table S2. Functional analysis of genes flanking stage-specific Twist binding sites 14

Table S3. Motif-content predicts TF binding 20

Table S4. Motif-clusters for *Drosophila* motif dataset. 21

Table S5. Functional analysis of genes flanking stage-specific binding sites 26

Supplemental References 31

Supplemental Discussion

Context-specific binding sites are associated with distinct biological functions.

Similar to the analysis for stage-specific Twist binding (see main manuscript), context-specific binding sites of other TFs in different species are associated with genes of distinct functional classes as revealed by gene-ontology (GO) analyses (see also the corresponding analyses in the original ChIP publications): Genes close to peaks in time, cell-type and tissue-specific ChIP datasets from *C. elegans*, mouse, and human (Zinzen et al. 2009; Palii et al. 2011; Verzi et al. 2010; Lin et al. 2010; Zhong et al. 2010; Rada-Iglesias et al. 2011; Visel et al. 2009) show biological functions expected for each particular condition (Supplemental Table 1). For example, functional categories associated with genes close to the pro-B cell specific binding sites of TCF3 included lymphocyte and B-cell activation. Genes close to HNF4A, GATA6, or CDX2 binding sites in proliferating epithelial cells were associated with metabolism, whereas genes close to binding sites in differentiated epithelial cells are associated with apoptosis and morphogenesis (see also (Verzi et al. 2010)). Forebrain-specific EP300 binding sites were associated with neurogenesis. Neuronal differentiation and regulation of cellular development were observed for EP300 binding sites in neuroectodermal spheres (hNECs), while binding sites in embryonic stem cells (hESCs) were close to genes involved in cell migration, regulation of gene expression, and signaling (see also (Rada-Iglesias et al. 2011)). For the TF PHA-4, we found embryo development, biogenesis, and reproductive developmental processes for binding sites specific to embryonic stages (see also (Zhong et al. 2010)). Together these observations and the analyses presented in the original ChIP publications (see citations above) suggest that the context-specific binding for these TFs are not due to experimental artifacts but to dynamic regulation of context-specific TF binding and function.

Prediction performance on randomized datasets.

Modern machine learning methods for binary classification (e.g. support-vector-machines or SVMs used here) are very powerful in predicting class-membership based on shared features (Joachims 1999). Prediction performance typically differs between instances of the training set and instances not used for training (i.e. previously unseen instances). Therefore, training and test sets are typically separated, for example by a leave-one-out-crossvalidation (LOOCV) protocol to more accurately assess how a machine-learning approach generalizes (to unseen data).

However, some steps that are often performed to optimize prediction performance, including the choice of the machine-learning approach or its parameters (e.g. γ for SVMs with Gaussian kernels), feature or motif selection, and motif or PWM optimization (e.g. (Beer and Tavazoie 2004)) can lead to artificially increased prediction performances (e.g. (Yuan et al. 2007)). Even on randomized data, such optimizations can already lead to predictions better than random guessing ((Yuan et al. 2007) and our unpublished observation). As it is our goal to assess the information content of regulatory DNA sequences and extract biologically relevant features, we minimized such optimization steps (e.g. no motif/PWM-optimization performed, SVMs only used with default parameters, etc.) but also assessed the prediction performance on randomized datasets.

We firmly believe that any machine-learning approach should quote the prediction performance (accuracy and area under the receiver-operator-characteristic curve [AUC]) on randomized data for comparison. For this control, all steps – particularly all optimization steps (see above) – should be repeated for randomized datasets. This would allow

discriminating prediction performance achieved by the approach alone from prediction performance that stems from relevant generalizable information in the data (here: regulatory DNA sequences).

We for example randomly assigned the binding sites to the different classes (e.g. early and late) and repeated the entire protocol, including the feature selection and SVM predictions using LOOCV. We repeated this three times for three independent randomizations and in each case obtained prediction performances similar to random guessing, i.e. accuracies of ~50% and AUCs of 0.5. This suggests that the non-random predictions on the real data stem from differences in the regulatory DNA sequences' motif contents that might be biologically relevant. While non-random performance on randomized data could indicate that the randomization might not have been truly random, it more likely suggests that the prediction performance is overestimated (see e.g. (Yuan et al. 2007)). In any case, such an outcome means that the machine learning protocol should be critically reviewed.

Identification of sequence requirements for context-specific Twist binding

To identify the cis-regulatory requirements for Twist binding in different developmental contexts (time-point specific ChIP-seq from (Zinzen et al. 2009)), we repeated our approach (including the *in silico* mutations) for the following comparisons: early vs. continuous, late vs. continuous Twist binding and early, late, and continuously bound vs. not bound Twist motifs. The contributions of all motifs to all pairwise binary classifications are listed in Supplemental Table S3).

The main motifs that were required to successfully discriminate between bound and not bound Twist motif-instances were: The motif for Twist itself: 27% (i.e. the Twist motif affected 7 out of a total of 26 sites affected by any motif; hereafter abbreviated by (7/26)) in early vs. not bound; 40.5% (15/37) in continuous vs. not bound; 100% (18/18) in the late vs. not bound comparisons. This suggests that a difference between bound and not bound motifs is the presence of not one but several Twist motifs, i.e. clustering of Twist motifs in agreement with previous observations (Berman et al. 2002). Indeed we found between 1.6-fold and 2-fold more Twist motifs in the early, late and continuous sites respectively in comparison to the not-bound motifs-instances. The motif for Snail, an already well known TF-partner of Twist (He et al. 2011; Sandmann et al. 2007) was also important and affected 34.6% (9/26) of the predictions in early vs. not bound; 43.2% (16/37) in the continuous vs. not bound; and 5.6% (1/18) in the late vs. not bound comparisons (Supplemental Fig. S5).

In agreement with the results from the early vs. late comparison (see main text), we found the following motifs to be generally important for the classification of early-bound sites in other binary classifications: the Vielfaltig's TAGteam motif affected 62% (66/107) of the predictions in early vs. continuous and 11.5% (3/26) in early vs. not bound comparisons. The Tramtrack motif affected 19% (20/107) of the predictions in the early vs. continuous, and the Kruppel motif 19% (20/107) in the early vs. continuous comparison. In addition we also found the motif for the TF Dorsal, an essential TF for early embryonic development that, together with Twist and Snail, controls the dorso-ventral patterning. Dorsal is a known partner TF of Twist at early stages of development (Zeitlinger et al. 2007; Sandmann et al. 2007) and important for successful classifications: it affected 8% (9/107) of the predictions in the early vs. continuous and 31% (8/26) in the early vs. not bound comparisons (Supplemental Fig. S5).

The motifs required for the successful classification of regions continuously bound by Twist were: the Twist motif itself affected the 37% (51/138), 40.5% (15/37), and 89% (74/83) of the predictions in the early vs. continuous, continuous vs. not bound, and late vs. continuous comparisons, respectively. The motif for Trithorax-like (TRL) also known as GAGA or GAF affected 45% (62/138), 19% (7/37), and 5% (4/83) of the predictions in the early vs. continuous, continuous vs. not bound, and late vs. continuous comparisons, respectively. The motif for Zeste affected 32% (44/138) and 11% (4/37) of the predictions in the early vs. continuous and continuous vs. not bound comparisons (Supplemental Fig. S5).

Finally, the motifs ME17 and Brinker (BRK) affected 58% (19/33) and 27% (9/33) of the predictions in the comparison early vs. late (Fig. 4C) and 32.5% (10/38) and 16% (6/83) in the comparison late vs. continuous (Supplemental Fig. S5).

Reporter and primer sequences for the *in vivo* validation experiments

The *wild-type* and mutant sequences used for the *in vivo* validations of the enhancers close to the genes *breathless (btl)* (Ohshiro and Saigo 1997), *wnt inhibitor of Dorsal (wntD)* (Zeitlinger et al. 2007), *thisbe (ths)* (Markstein et al. 2004) and *cactus (cact)* are shown below. Each of them comprise the 401bp region used for the SVM predictions and the *in silico* mutations plus additional flanking regions for primer design. The TAGteam-like motifs are in lowercase and highlighted in a black box in the mutated sequences. The mutations performed are highlighted in yellow.

>chr3L_14069504_14069975 *btl*-enh wt

CGCATTTGTTCCCATAGCCCCTAACAAGTGCATCAAAAATGCATTTAACTATTTTTGGGTAGATT
ACGTGCCCGGCATCTGGATCTGGATTTTGGCCTATGCGGGAGCTAGCTGCTGCCATGCCACATGCAG
GCAACTGAGATTTTTGGATATCAGTCACGTAACGGATATTTCTTTTGCATAAATTAAGCCGCC
CATGCGCACGTACTGCACACCACTAAGACAATGGACACATTTTCTTTTGCATAAATTAAGCCGCC
AAGCAAACAAACAGGGAGGGAACCCGAAGGCCAGATAATAAAGAATTTCAACGAGTCGAAGCCA
CCGCAAACATGCTTAGAATAAAACCAACAAGAACCTAGCAGAAGCAGCAGCAGCAAACCAGCAAT
AAAAAGAAAGGAAACCGCTCAGGAAAACCTACCAGCGAAAACAATTGCAACATGCGGGAGGGAGA
CCACCTCGATCAGGTAGC

>chr3L_14069504_14069975 *btl*-enh mutant

CGCATTTGTTCCCATAGCCCCTAACAAGTGCATCAAAAATGCATTTAACTATTTTTGGGTAGATtgc
tgggcccggcatctggatctggatTTTGGCCTATGCGGGAGCTAGCTGCTGCCATGCCACATGCAG
CAACTGAGATTTTTGGATATCAGTatctgcaCGGATATTTCTTTATTTGATATGCGATTACCGCGCAT
GCGagcactTGCACACCACTAAGACAATGGACACATTTTCTTTTGCATAAATTAAGCCGCCAAGCA
AACAAACAGGGAGGGAACCCGAAGGCCAGATAATAAAGAATTTCAACGAGTCGAAGCCACCGCAA
ACATGCTTAGAATAAAACCAACAAGAACCTAGCAGAAGCAGCAGCAGCAAACCAGCAATAAAAA
GAAAGGAAACCGCTCAGGAAAACCTACCAGCGAAAACAATTGCAACATGCGGGAGGGAGACCACCT
CGATtagaacgC

>chr2L_16323473_16323946 *cact*-enh wt

GCCACCACGTCCACTGATCCCGAAATACAGGCCAGGTGCAAAGGTCTGAAATGAGAGAGAGAGAG
CAAAAAAACAACATGAATTTGTGGGTTTCTTTTACCTGGCGATCGAACACCCGCACTTTCTTTC
CTTCTGTACACTTGTGCGCGTTCATGTGGCGGCGTAAAAACCCAAGTAAGTTAAGTAGAGGTAA
AGTAATGTGAAGAAAGAAGAGGTGCCCCGAAAAACCTGTTCCCGTACCTGCCACATACCAGGAA
AACCCCAAAGCCTGACTAAAACACATGTGTGTCAGTCGCGTTTCTCGCCTCAGACAATCGTCTTTGA

GTGATTCACGAACTCTCCCCTTAGCTCGCTTGGTGGTCATCAGTGCAACGCGGCGTTGTCTTTGT
TATTTTGGGAAATTTTAGGATTCTCATCGGAAATACGGCACGTATAAGTTCATAGAAATTCGGAA
ATGATTCAAAGACTGTGC

>chr2L_16323473_16323946 *cact*-enh mutant

GCCACCACGTCCACTGATCCCGAAATACAGGCCAGGTGCAAAGGTCTGAAATGAGAGAGAGAGAG
CAAAAAAACAACATGAATTTGTGGGTTTCTT[tttccga]GCGATCGAACACCCGCACTTTCTTTCT
TCCTGTACACTTGTGCGCCGTTTCATGTGGCGGCGTAAAAACCT[taaccaa]GTTAAGTAGAGGTAAAGTA
ATGTGAAGAAAGAAGAGGTGCCCCGAAAAACCTGTTCCG[gttacca]CCACATACCAGGAAAAACCC
AAAGCCTGACTAAAACACATGTGTCAGTCGCGTTTCTCGCCTCAGACAATCGTCTTTGAGTGATT
CACGAAACTCTCCCCTTAGCTCGCTTGGTGGTCATCAGTGCAACGCGGCGTTGTCTTTGTTATTTT
GGGAAATTTTAGGATTCTCATCGGAAATACGG[agccgat]AAGTTCATAGAAATTCGGAAATGATTC
AAAGACTGTGC

>chr3R_9118937_9119597 *wntD*-enh wt

AGCACACTTATATAGCCTGCAAATCCCAAGCCAGGGCGCCCTCCTGGGGCCGGCCCGTGGGAATTT
CGGGCCTGCTCAAAAAACCGGAAATTTGCCGTTTTCCACTTGGAATTTTGCATGGGCAGGGGGT
AGGAACTCCCGGCAATGGACGGGTACAAAAACCCACTGGCAGCCCGAGACGCAATTGCGGAGCAG
CCCAGTTTCCTGGTTGACTACCTGCTCTCGTCTGCGCCGGCGGAGGTGAAGGATCCGCCTTGCTG
CGAGCAAGTTTCCCACGCTTAGGCAGGTAGAGCCGTAAACGGCACCCGACGTGCTCATGAATGAA
GCCCAGTCGAGTCCATTCAATACGGCCGGATTTTCCCGGACTCACACTGCACAACATCAATGCCCG
ATACGGGGACGGGTTTGTGTTGGGTTTGGACTGGTCAAGCCAATTATATAACAAAACATATGACC
AACAGTATATACAGTATAATCTGGGAAATTAATTGTCCTTTGTGGTGAGCCGGCGGGAAGTGC
AGTATTGTTGCTATATTATTTAAAAGTTAATACCCAGGATTATGGTGTTAAGCACCATTAGGAAC
AGATATCCAGATATTGGTAACGGATTTTTTTTGGGATTTTTTTTCAGATGAAAAATTTGTGCATGCCT
AAAAACAA

>chr3R_9118937_9119597 *wntD*-enh mutant

AGCACACTTATATAGCCTGCAAATCCCAAGCCAGGGCGCCCTCCTGGGGCCGGCCCGTGGGAATTT
CGGGCCTGCTCAAAAAACCGGAAATTTGCCGTTTTCCACTTGGAATTTTGCATGGGCAGGGGGT
AGGAACTCCCGGCAATGGACGGGTACAAAAACCCACTGGCAGCCCGAGACGCAATTGCGGAGCAG
CCCAGTTTCCTGGTTGAC[ctcgcca]CTCTCGTCTGCGCCGGCGGAGGTGAAGGATCCGCCTTGCTGCG
AGCAAGTTTCCCACGCTTAGG[ttgtaag]AGCCGTAAACGGCACCCGACGTGCTCATGAATGAAGCCC
AGTCGAGTCCATTCAATACGGCCGGATTTTCCCGGACTCACACTGCACAACATCAATGCCCGATAC
GGGGACGGGTTTGTGTTGGGTTTGGACTGGTCAAGCCAATTATATAACAAAACATATGACCAACA
GTATATA[tcctaai]AATCTGGGAAATTAATTGTCCTTTGTGGTGAGCCGGCGGGAAGTGCAGTATT
GTTGCTATATTATTTAAAAGTTAATACCCAGGATTATGGTGTTAAGCACCATTAGGAACAGATAT
CCAGATATTGGTAACGGATTTTTTTTGGGATTTTTTTTCAGATGAAAAATTTGTGCATGCCTAAAAAC
AA

>chr2R_7681627_7682251 *ths*-enh wt

CTCAACGCAACATCAGCGACCTTTTTTTGTTTTTGGATCTGACCATGGCCAATTAGAGCCGCTGCA
GCTGCTGCCTGAAAAGCCGAAAGAAGAGATACAAGGACCAGCAGAGCTACGCAGCCTCACACAG
CAGGATAATAGGGAAAGAAAGGACAGGACAATGGAGGTTTCAAGAAGCGAGCAAATGCTGGAA
AATGCAGTGACAACAGGTGCAAAATTATTTTTTGTGTTGCGAGTGCGCGTGAAAATTTCCAGC
TGGCCAGGGACAGGAATATGACCACTTAAGGCCTAATGTGCGAAAAGTTTCTTTGTCAATTTACA
CGCCCTCTCTCACCAGCGACCGTGAAAACTTCATTCAATTTGCATGGCTAAGCTCAGGTAGCCGG
GGATTATCCCTCGTTCTAACCAAAACCTCCTGCTACATTCGGGTTTATCCCACTTGTTTGGCTGGG
AATTTCCCCCGCAGATTTACGGTGGTCAGCCAAATCCCGGTTAGCTGGAAAAATATCGCAGAAAT

AAGGGAAGGTACGGCTGCTAATGAATCCTGACATCTCAATCAATTTTGGGGGAATCGAAACGCAA
GGGAGGTGGAACCTTCACCACCCCAAAACAAAAGTGAA

>chr2R_7681627_7682251 *ths*-enh mutant

CTCAACGCaaagcttaacGACCTTTTTTTGTTTTTGGATCTGACCATGGCCAATTAGAGCCGCtgacgtaa
ctGCCTGAAAAGCCGAAAGAAGAGATACAAGGACCAGCACGAGCTACGCAGCCTCACACAGCAGG
ATAATAGGGAAAAGAAAGGACAGGACAATGGAGGTTTCAAGAAGCGAGCAAATGCTGGAAAATG
CAGTGACaatagggccaaAAATTATTTTTTGTGTTGTTGCGAGTGCGCGTGAAAATTTcctccatgcCAGG
GACAGGAATATGACCACTTAAGGCCTAATGTGCGAAAAGTTCCTTTGTCAATTTACACGCCCTCT
CCTCACCAAGCGACCGTGAAAACCTTCATTCATTTGCATGGCTAAGCTCAGGTAGCCGGGGATTAT
CCCTCGTTCTAACCAAAAACCTCCTGCTACATTCGGGTTTATCCCACTTGTTTGGCTGGGAATTTCC
CCCGCAGATTTACGGTGGTCAGCCAAATCCCGTTAGCTGGAAAAATATCGCAGAAATAAGGGAA
GGtttggtcgcTAATGAATCCTGACATCTCAATCAATTTTGGGGGAATCGAAACGCAAGGGAGGTGG
AACTTTCACCACCCCAAAACAAAAGTGAA

List of primers

Sequences of the primers (all from 5' to 3') used for the PCR amplification of the wild type enhancers from genomic DNA are shown below.

cgcatttgttcccatagcccctaa	Forward	<i>btl</i>
gctacgtgatcgaggtggtctccc	Reverse	<i>btl</i>
gccaccacgtccactgatcccg	Forward	<i>cact</i>
gcacagtctttgaatcatttccgaatttct	Reverse	<i>cact</i>
agcacacttatatagcctgcaaatccca	Forward	<i>wntD</i>
tttttaggcgatgcacaaatttcatctg	Reverse	<i>wntD</i>
ctcaacgcaacatcagcgaccttt	Forward	<i>ths</i>
ttcactttttgtttggggtggtgaaa	Reverse	<i>ths</i>

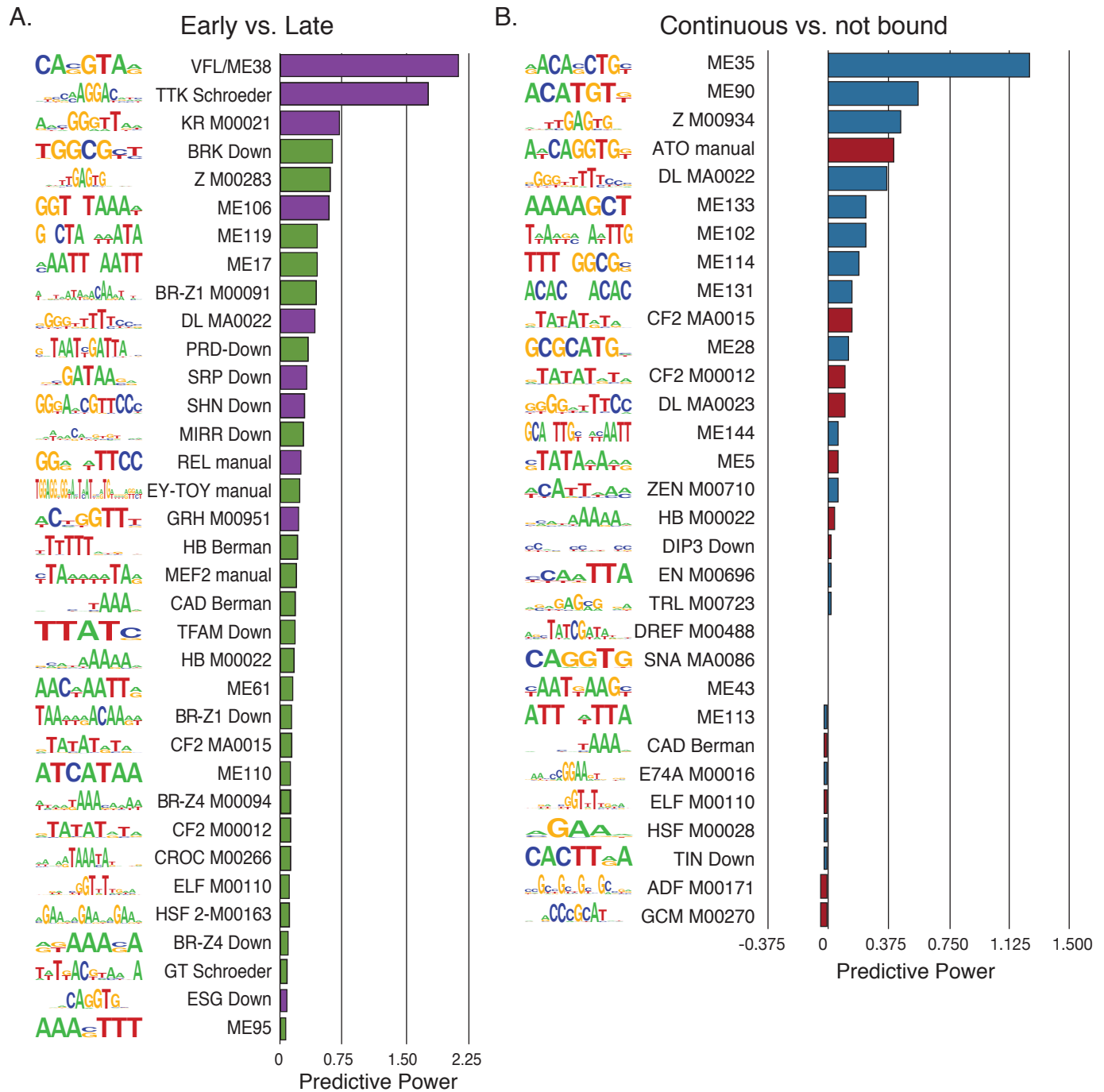
Sequences of the primers (all from 5' to 3') used for the ChIP qPCR are shown below. As a negative control we used the coding sequence of the gene *CG12362*

ttgcacctggcctgtatttcg	Forward	reporter
tcaccactttgtacaagaaagctgg	Reverse	reporter
atcgttctggatgttgagcagg	Forward	endogenous <i>cact</i> locus
ttgcacctggcctgtatttcg	Reverse	endogenous <i>cact</i> locus
atgacatgcaggaggagatgagc	Forward	Negative control
atcagagtgggttcgacattggc	Reverse	Negative control

Supplemental Fig. S1

Regulatory motifs ranked according to their predictive power.

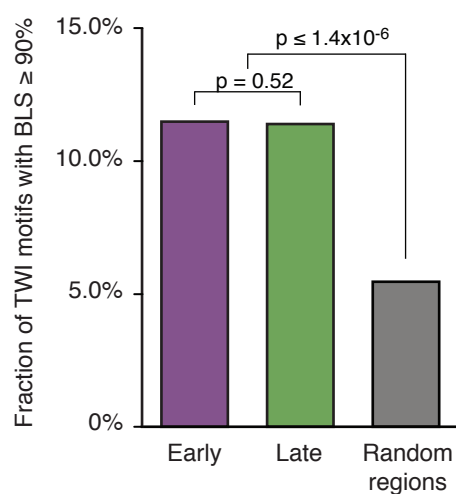
We defined the predictive power of each motif as the difference in prediction accuracy (percent correctly predicted sites) between cross-validated binary classification using all features versus all features except the motif of interest. Shown are the top motifs (used for all subsequent analysis), for the early vs. late comparison (A) and the continuous vs. not-bound comparison (B) ranked by their respective predictive power. The color of the bars correspond to the set of sites in which the respective motif is enriched (purple: early, green: late, blue: continuous, red: not-bound). All motifs were obtained from Stark *et al.* 2007. Schroeder, Down and Berman make reference to their original source: Schroeder *et al.* 2004, Down *et al.* 2007 and Berman *et al.* 2004. Motifs with an ID starting with MA were originally obtained from JASPAR (Bryne *et al.* 2009), with an ID starting with ME (Motif-in-Enhancers) are predicted motifs from Stark *et al.* 2007 and with an ID starting with M were originally obtained from TRANSFAC (Matys *et al.* 2003) (see Supplemental Table S4)



Supplemental Fig. S2**The Twist motif is evolutionarily conserved in early and late Twist binding sites.**

Shown is the fraction of Twist motif instances above a Branch Length Score (BLS) of 90% (i.e. a branch-length of 0.9×4.257582 substitutions per neutral site in a whole-genome alignment of 12 closely related *Drosophila* species; see Stark *et al.*, 2007; Kheradpour *et al.*, 2007) in the early and late binding sites of Twist and in random genomic regions. Statistical significance was calculated by a hypergeometric p-value.

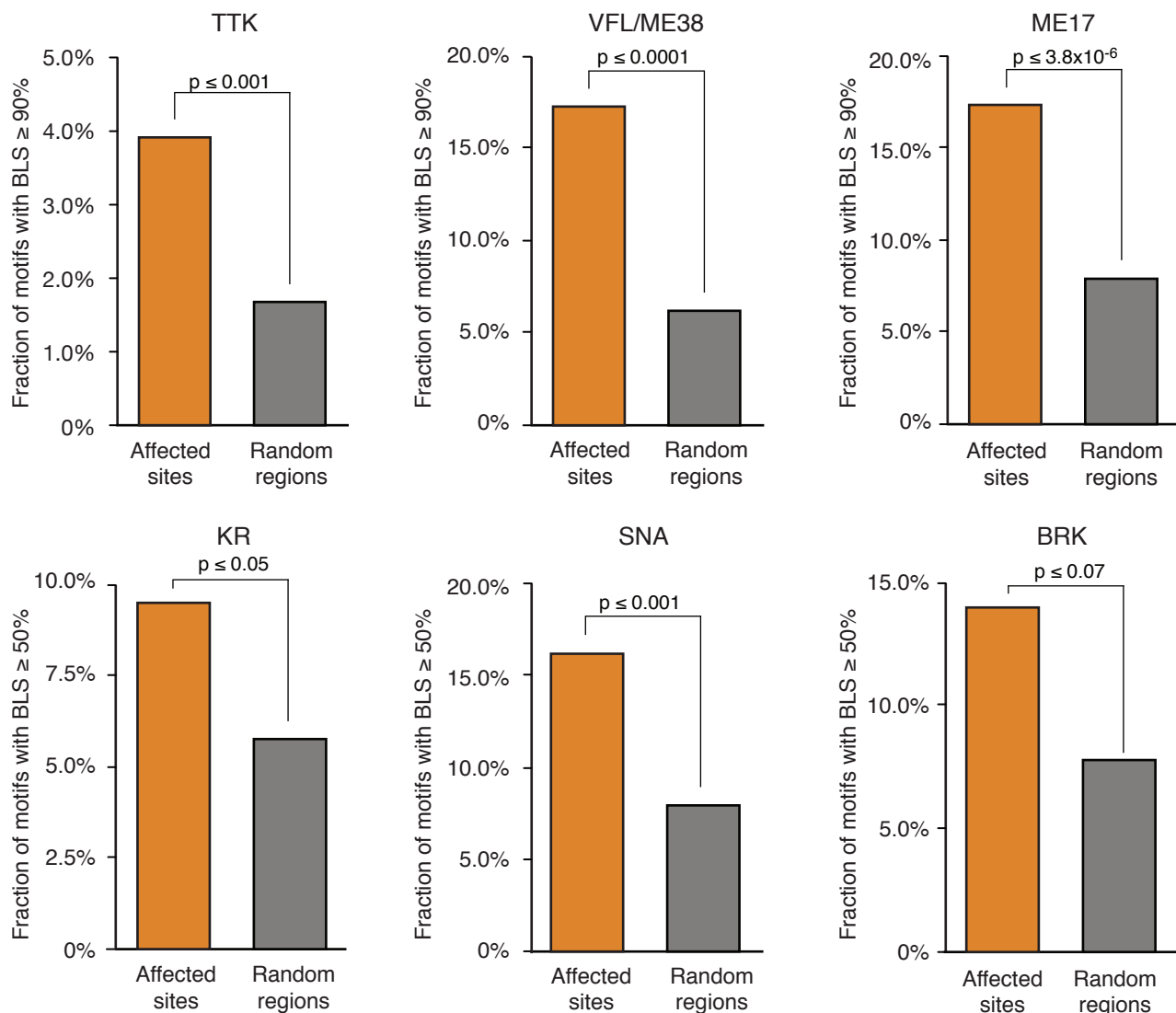
Twist motif instances in both early and late Twist binding sites (from ChIP-chip of Zinzen *et al.*, 2009) show a similar conservation levels that are significantly above the average conservation level of Twist motifs randomly selected from the the non-coding non-repeat *D. melanogaster* genome.



Supplemental Fig. S3**Evolutionary conservation of putative Twist partner motifs.**

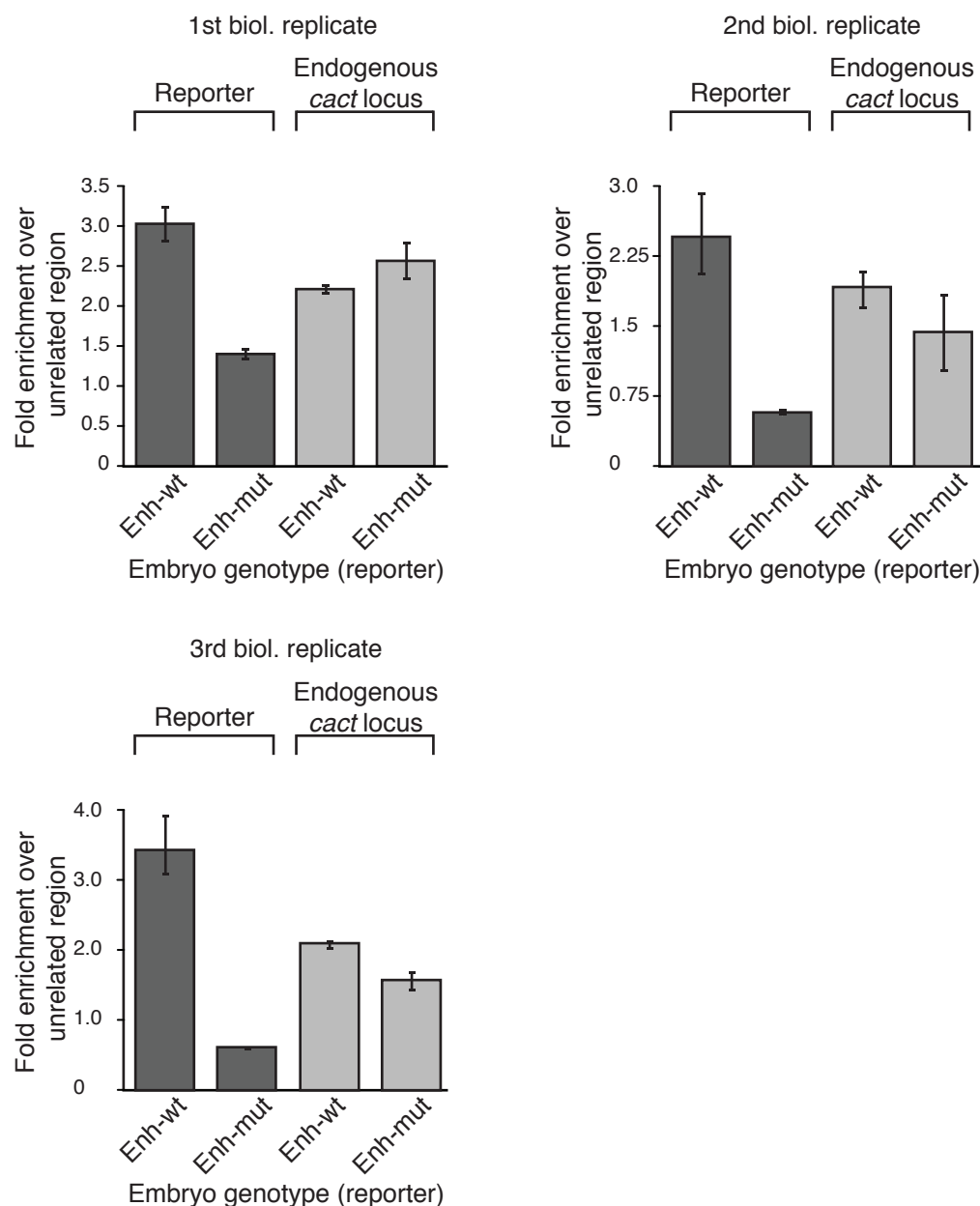
Shown is the fraction of regulatory motif instances above a Branch Length Score (BLS) of 90% (i.e. a branch-length of 0.9×4.257582 substitutions per neutral site in a whole-genome alignment of 12 closely related *Drosophila* species; see Stark *et al.*, 2007; Kheradpour *et al.*, 2007) in those Twist binding sites for which the correct classification depends on the presence of the respective TF motif and in random genomic regions. For the TF motifs KR, SNA and BRK we considered the motif instances above 70% of the BLS. Statistical significance was calculated by a hypergeometric p-value

Motif instances in Twist binding sites (from ChIP-chip of Zinzen *et al.*, 2009) for which the correct classification into early vs. late classes depended on the respective TF partner motif are significantly more strongly conserved than motif instances randomly selected from the the non-coding non-repeat *D. melanogaster* genome.



Supplemental Fig. S4**Biological replicates for Twist ChIP-qPCR for the *cact* enhancer.**

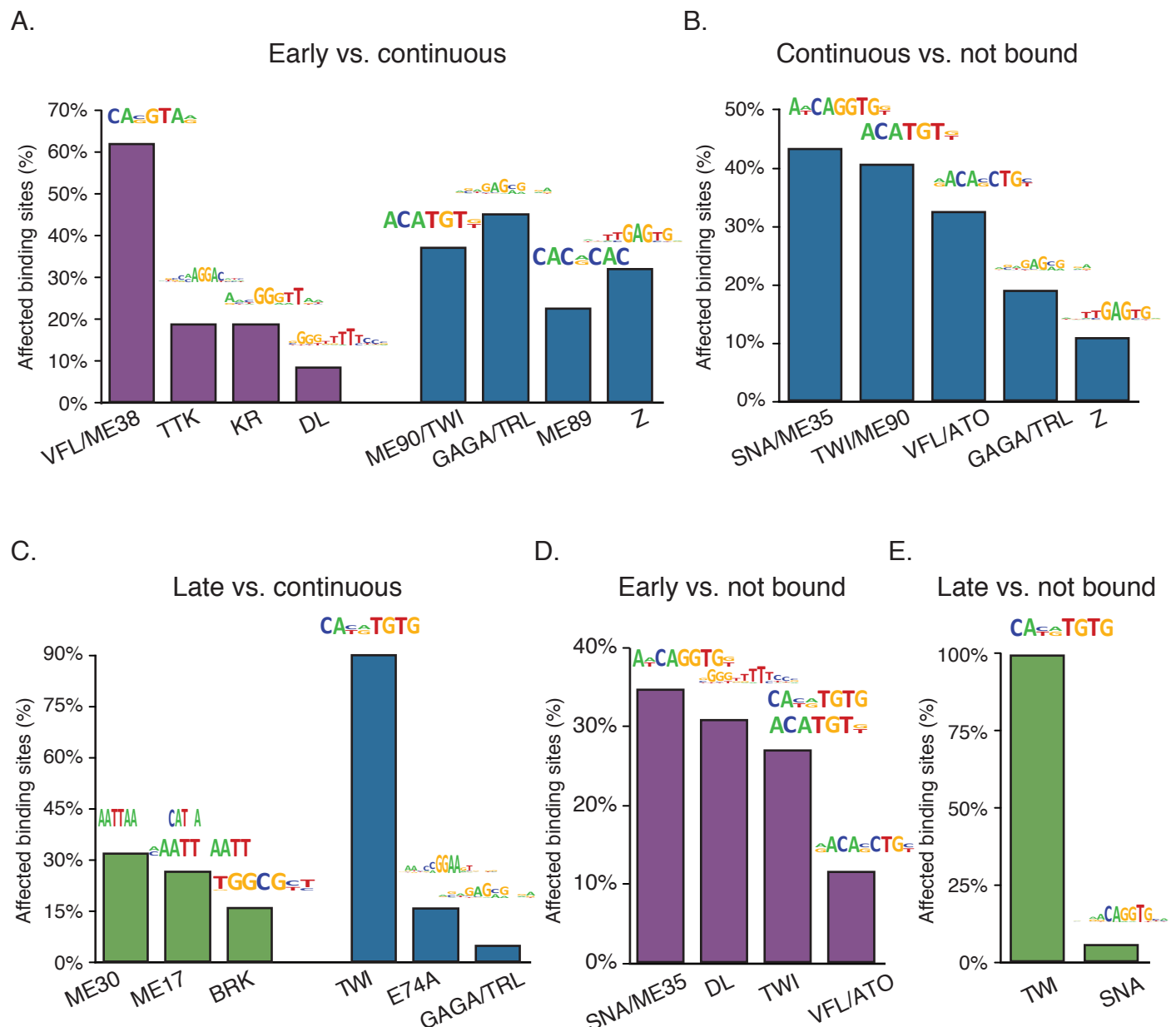
Twist ChIP enrichment at the reporter (dark bars) and the endogenous *cact* locus over an unrelated genomic region. Shown are three biological replicates from independent embryo collections carrying either the wildtype (Enh-wt) or the TAGteam mutant (Enh-mut) reporter enhancer (see methods). Error-bars show the minimum and maximum enrichment of two (1st biol. replicate) and three (2nd and 3rd biol. replicate) technical replicates. To discriminate between the endogenous *cact* locus and the transgenic reporter we designed specific qPCR primers pairs to match the *cact* enhancer (forward primer) and the respective flanking region (reverse primer; see schematic in Fig. 4E). Note that this places the reverse primer at a certain distance from the peak summit (301bp for the endogenous locus and 288bp for the reporter, which likely influences the absolute enrichment levels (note though that identical primers were used for the wt and mutant reporter).



Supplemental Fig. S5**Motif requirements for context-specific binding sites of Twist.**

Percent of the sites with a score ≥ 75 in the early (purple), late (green) or continuous (blue) sets that are affected (i.e. drop in score ≥ 20) by the *in silico* mutation of the different motifs for the following binary classifications (comparisons): early vs. continuous (A), continuous vs. not bound (B), late vs. continuous (C), early vs. not bound (D) and late vs. not bound (E).

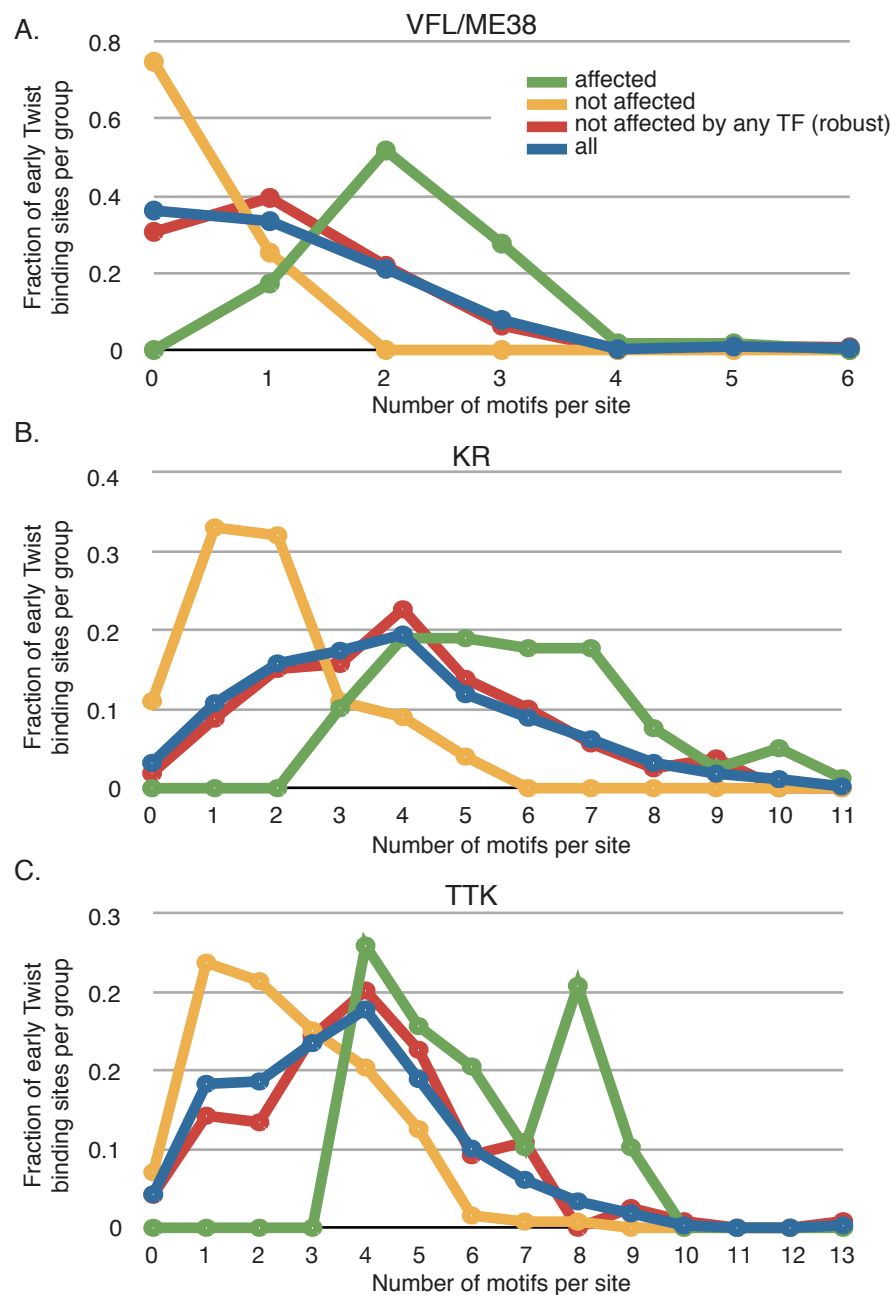
The type of regions that are affected by the respective motifs are color-coded (purple: early, blue: continuous, green: late). The logos for each motif are indicated above each bar. Note that comparisons with of bound with not bound Twist motifs only include bound regions that contain a Twist motif within 50 bp from the ChIP-chip peak summit (see methods).



Supplemental Fig. S6

Partner TF motif-counts per site for early Twist binding sites.

Distribution of partner TF motif-occurrences per site for early Twist binding sites dependent on whether the sites' predictability are affected by the in silico mutations for VFL/ME38 (A), KR (B), and TTK (C). Shown are normalized motif-counts (% of sites with that exact count) for sites affected by the in silico mutations (green), sites not affected (yellow), robust sites (red), and all sites (blue). As expected, affected Twist binding sites have at least one partner TF motif. However, even sites with a substantial number of partner motifs are in the not-affected and robust classes, suggesting redundancies between partner motifs and the corresponding TFs.



Supplemental Table S1**TFs binding sites differ between time-points and cell types**

Analysis of ChIP peaks for several TFs in *D. melanogaster* (BIN, MEF2, TIN and TWI), in mouse (TCF3 also known as E2A), human (TAL1, HNF4A, GATA7, CDX2) and *C. elegans* (PHA-4), for the cofactor EP300 (mouse) and the histone mark H3K4me2 (human). The ChIPs were performed at different time-points (BIN, MEF2, TIN, TWI, TCF3, PHA-4, HNF4A, H3K4me2, GATA and CDX2) and between distinct cell types (TAL1 and EP300) here referred as conditions (Cond) (For more information see methods). For example: Up to 94.2% and 93% of the binding sites for TAL1 are specific for erythroid cells and Jurkat cells as reported by (Palii et al. 2011). Data was obtained from: (Zinzen et al. 2009; Lin et al. 2010; Palii et al. 2011; Zhong et al. 2010; Verzi et al. 2010; Visel et al. 2009) (for more information see methods). The Gene Ontology analysis for each data set can be found at: Wilczynski et al. 2010 (for BIN, MEF2, TIN and TWI), Verzi et al. 2010 (for HNF4A, H3K4me2, GATA and CDX2), Palii et al. 2011 (for TAL1), Zhong et al. 2010 (for PHA-4), Supplemental Table S5 and Supplemental Discussion. "diffCell" and "prolCell" refers to differentiated and proliferating intestinal epithelial cells from Verzi et al. 2010 and hpf refers to "hours post fertilization".

Condition (Cond) 1	Condition (Cond) 2	Condition (Cond) 3	Total Peaks Cond 1	Total Peaks Cond 2	Total Peaks Cond 3	Overlap Cond1-Cond2	Overlap Cond1-Cond3	Overlap Cond2-Cond3	Overlap Cond1-Cond2-Cond3	Specific to Cond1 (%)	Specific to Cond2 (%)	Specific to Cond3 (%)
BIN 6-8 hpf	BIN 8-10 hpf	BIN 10-12 hpf	1084	613	1189	388	539	389	315	43.54%	24.63%	48.44%
MEF2 2-4 hpf	MEF2 4-6 hpf	MEF2 6-8 hpf	444	1834	2185	300	250	967	236	29.28%	43.78%	55.10%
MEF2 4-6 hpf	MEF2 6-8 hpf	MEF2 8-10 hpf	1834	2185	1737	967	739	1010	621	40.84%	37.94%	35.06%
MEF2 6-8 hpf	MEF2 8-10 hpf	MEF2 10-12 hpf	2185	1737	1807	1010	622	836	539	49.98%	24.76%	49.14%
MEF2 2-4 hpf	MEF2 6-8 hpf	MEF2 10-12 hpf	444	2185	1807	250	140	622	125	40.32%	65.81%	64.75%
MEF2 4-6 hpf	MEF2 8-10 hpf	MEF2 10-12 hpf	1834	1737	1807	739	481	836	406	55.62%	32.70%	49.58%
MEF2 2-4 hpf	MEF2 8-10 hpf	MEF2 10-12 hpf	444	1737	1807	201	140	836	129	52.25%	47.73%	53.13%
TIN 2-4 hpf	TIN 4-6 hpf	TIN 6-8 hpf	814	1154	1015	600	411	532	367	20.88%	33.71%	43.25%
TWI 2-4 hpf	TWI 4-6 hpf	TWI 6-8 hpf	1620	1750	1576	769	554	806	473	47.53%	37.03%	43.72%
TCF3 pre-pBcells	TCF3 pBcells	-	2902	16202	-	1506	-	-	-	48.10%	90.70%	-
TAL1 erythroid	TAL1 jurkat	-	6551	5390	-	381	-	-	-	94.18%	92.93%	-
PHA-4 embryo	PHA-4 L1	-	3793	6767	-	2857	-	-	-	24.68%	57.78%	-
HNF4A diffCell	HNF4A prolCell	-	46361	24941	-	17114	-	-	-	63.09%	31.38%	-
H3K4me2 diffCell	H3K4me2 prolCell	-	108725	101757	-	76888	-	-	-	29.28%	24.44%	-
GATA6 diffCell	GATA6 prolCell	-	2072	18446	-	824	-	-	-	60.23%	95.53%	-
CDX2 diffCell_2	CDX2 prolCell	-	28852	12072	-	7832	-	-	-	72.85%	35.12%	-
EP300 Forebrain	EP300 Limb	-	2453	561	-	274	-	-	-	88.83%	51.16%	-

Supplemental Table S2**Functional analysis of genes flanking stage-specific Twist binding sites**

Hypergeometric p-values (FDR-adjusted) for the enrichment of functional categories in genes close to early or late specific Twist binding sites. Only the Gene Ontologies that appeared significant in at least one condition are shown.

Functional Category	TWI 2-4 hpf	TWI 6-8 hpf
embryonic_pattern_specification	0.0000	0.0511
segmentation	0.0000	0.7836
cell_fate_specification	0.0000	0.0847
gland_development	0.0000	0.0831
central_nervous_system_development	0.0000	0.8412
nervous_system_development	0.0000	0.0563
negative_regulation_of_developmental_process	0.0000	0.7671
segment_specification	0.0000	1.0000
periodic_partitioning	0.0000	1.0000
germ_cell_migration	0.0000	0.6638
regulation_of_cell_differentiation	0.0000	0.0696
epithelial_cell_migration	0.0000	0.9409
epithelial_cell_migration,_open_tracheal_system	0.0000	0.9409
axis_specification	0.0000	0.3518
trunk_segmentation	0.0000	1.0000
genital_disc_development	0.0000	1.0000
ectoderm_development	0.0000	0.3197
dorsal/ventral_pattern_formation	0.0000	0.2104
midgut_development	0.0000	0.3697
neuroblast_fate_determination	0.0000	0.9564
regulation_of_cell_fate_specification	0.0000	0.2293
hindgut_morphogenesis	0.0000	1.0000
genital_disc_anterior/posterior_pattern_formation	0.0000	1.0000
genital_disc_pattern_formation	0.0000	1.0000
anterior/posterior_pattern_formation	0.0000	0.1769
imaginal_disc_development	0.0000	1.0000
head_segmentation	0.0000	1.0000
enzyme_linked_receptor_protein_signaling_pathway	0.0000	0.0538
anterior/posterior_axis_specification	0.0000	1.0000
periodic_partitioning_by_pair_rule_gene	0.0000	1.0000
salivary_gland_development	0.0000	0.8871
head_involution	0.0000	0.6362
specification_of_segmental_identity,_head	0.0000	1.0000
blastoderm_segmentation	0.0000	0.3503
gut_development	0.0000	0.4763
ventral_cord_development	0.0000	1.0000
cell_differentiation	0.0000	0.3611
posterior_head_segmentation	0.0000	1.0000
reproductive_developmental_process	0.0000	0.3697
regulation_of_cell_proliferation	0.0000	0.0831
spiracle_morphogenesis,_open_tracheal_system	0.0000	1.0000
regulation_of_tube_architecture,_open_tracheal_system	0.0000	1.0000
regulation_of_anatomical_structure_morphogenesis	0.0000	0.2098
transition_metal_ion_binding	0.0000	0.1603
negative_regulation_of_cell_differentiation	0.0000	1.0000
amnioserosa_maintenance	0.0000	1.0000
gonadal_mesoderm_development	0.0000	0.3049
germ-band_shortening	0.0000	0.1813
regulation_of_cell_cycle	0.0000	0.2754
imaginal_disc-derived_wing_vein_specification	0.0000	0.9564
asymmetric_cell_division	0.0000	0.5803
regulation_of_growth	0.0000	0.2180
wing_disc_anterior/posterior_pattern_formation	0.0000	0.0949
dendrite_morphogenesis	0.0000	0.1284
metal_ion_binding	0.0000	0.2266
regulation_of_biological_quality	0.0000	0.8429
pole_cell_migration	0.0000	0.2293
wing_disc_dorsal/ventral_pattern_formation	0.0000	0.5803
establishment_and/or_maintenance_of_cell_polarity	0.0000	1.0000
proximal/distal_pattern_formation,_imaginal_disc	0.0000	0.8316
tissue_homeostasis	0.0000	1.0000
embryonic_axis_specification	0.0000	1.0000
lymph_gland_development	0.0000	0.1351
negative_regulation_of_cell_fate_commitment	0.0000	0.6011
negative_regulation_of_cell_fate_specification	0.0000	0.6011
regulation_of_cell_fate_commitment	0.0000	0.6011
leg_disc_pattern_formation	0.0000	1.0000

leg_disc_proximal/distal_pattern_formation	0.0000	1.0000
anatomical_structure_homeostasis	0.0000	1.0000
reproductive_process	0.0000	0.5589
proximal/distal_pattern_formation	0.0001	0.8871
ion_binding	0.0001	0.3591
maintenance_of_epithelial_integrity_open_tracheal_system	0.0001	1.0000
R8_cell_fate_commitment	0.0001	1.0000
regulation_of_cellular_component_organization_and_biogenesis	0.0001	0.1732
transmembrane_receptor_protein_serine/threonine_kinase_signaling_pathway	0.0001	1.0000
optic_lobe_placode_development	0.0001	1.0000
segment_polarity_determination	0.0001	0.9564
cation_binding	0.0001	0.2111
foregut_morphogenesis	0.0001	1.0000
determination_of_anterior/posterior_axis_embryo	0.0001	1.0000
eye-antennal_disc_morphogenesis	0.0002	0.3697
eye_photoreceptor_cell_differentiation	0.0002	1.0000
regulation_of_apoptosis	0.0002	0.8731
germ-band_extension	0.0002	1.0000
gastrulation	0.0002	0.0949
wing_disc_morphogenesis	0.0002	0.0949
specific_transcriptional_repressor_activity	0.0002	0.2104
hemopoietic_or_lymphoid_organ_development	0.0003	0.2290
photoreceptor_cell_differentiation	0.0003	1.0000
regulation_of_programmed_cell_death	0.0003	0.1351
salivary_gland_boundary_specification	0.0003	0.1140
primary_branching_open_tracheal_system	0.0003	1.0000
regulation_of_cell_morphogenesis	0.0003	0.2389
cell_cycle	0.0004	0.7846
branched_duct_epithelial_cell_fate_determination_open_tracheal_system	0.0004	1.0000
terminal_region_determination	0.0004	1.0000
specification_of_segmental_identity_abdomen	0.0004	1.0000
specification_of_segmental_identity_trunk	0.0004	1.0000
transmembrane_receptor_protein_tyrosine_kinase_signaling_pathway	0.0005	0.0991
regulation_of_imaginal_disc_growth	0.0005	0.1351
compound_eye_photoreceptor_cell_differentiation	0.0006	1.0000
intercellular_junction	0.0007	0.9781
determination_of_genital_disc_primordium	0.0007	1.0000
determination_of_imaginal_disc_primordium	0.0007	1.0000
regulation_of_organ_growth	0.0007	0.1603
positive_regulation_of_signal_transduction	0.0007	1.0000
regulation_of_protein_kinase_cascade	0.0007	1.0000
border_follicle_cell_migration	0.0008	1.0000
morphogenesis_of_a_branching_structure	0.0010	0.3697
formation_of_primary_germ_layer	0.0011	1.0000
wing_disc_proximal/distal_pattern_formation	0.0011	1.0000
protein_kinase_cascade	0.0011	0.3627
compound_eye_photoreceptor_fate_commitment	0.0011	1.0000
eye_photoreceptor_cell_fate_commitment	0.0011	1.0000
photoreceptor_cell_fate_commitment	0.0011	1.0000
imaginal_disc_lineage_restriction	0.0012	0.4171
tracheal_outgrowth_open_tracheal_system	0.0012	1.0000
germ-line_stem_cell_division	0.0013	1.0000
stem_cell_division	0.0016	1.0000
regulation_of_mitotic_cell_cycle	0.0016	0.9973
ovarian_follicle_cell_migration	0.0017	1.0000
positive_regulation_of_cell-cell_adhesion	0.0017	0.7189
positive_regulation_of_cell_adhesion	0.0017	0.7189
regulation_of_cell-cell_adhesion	0.0017	0.7189
specification_of_segmental_identity_antennal_segment	0.0017	0.7189
compartment_specification	0.0017	1.0000
dorsal_closure_leading_edge_cell_fate_determination	0.0017	1.0000
leading_edge_cell_fate_determination	0.0017	1.0000
regulation_of_cell_shape	0.0018	0.3697
positive_regulation_of_apoptosis	0.0020	0.4171
cell-cell_adherens_junction	0.0020	1.0000
anterior_head_segmentation	0.0021	1.0000
morphogen_activity	0.0021	1.0000
nurse_cell_apoptosis	0.0021	1.0000
progression_of_morphogenetic_furrow_during_compound_eye_morphogenesis	0.0021	1.0000
somatic_stem_cell_division	0.0021	1.0000
regulation_of_neurogenesis	0.0024	1.0000
gland_morphogenesis	0.0024	1.0000
salivary_gland_morphogenesis	0.0024	1.0000
anti-apoptosis	0.0025	1.0000
cardiac_cell_differentiation	0.0026	0.0949
transforming_growth_factor_beta_receptor_signaling_pathway	0.0030	1.0000
branching_morphogenesis_of_a_tube	0.0033	0.2864
neuron_development	0.0034	1.0000

branch_fusion,_open_tracheal_system	0.0037	1.0000
tube_fusion	0.0037	1.0000
compound_eye_cone_cell_fate_commitment	0.0037	1.0000
genitalia_development	0.0037	1.0000
negative_regulation_of_salivary_gland_boundary_specification	0.0037	1.0000
regulation_of_salivary_gland_boundary_specification	0.0037	1.0000
zygotoc_determination_of_anterior/posterior_axis,_embryo	0.0037	1.0000
cell_junction	0.0042	0.0696
dorsal/ventral_axis_specification	0.0042	1.0000
bristle_morphogenesis	0.0046	0.7836
torso_signaling_pathway	0.0054	1.0000
molecular_transducer_activity	0.0055	0.3631
signal_transducer_activity	0.0055	0.3631
anterior/posterior_lineage_restriction,_imaginal_disc	0.0055	0.8429
neuroblast_development	0.0055	0.8429
notum_cell_fate_specification	0.0055	0.8429
endoderm_formation	0.0055	1.0000
labial_disc_development	0.0055	1.0000
ventral_midline_development	0.0055	1.0000
protein_localization	0.0057	0.5653
morphogenesis_of_a_polarized_epithelium	0.0061	1.0000
negative_regulation_of_developmental_growth	0.0061	1.0000
adherens_junction	0.0094	0.0804
decapentaplegic_receptor_signaling_pathway	0.0097	1.0000
apoptosis	0.0114	1.0000
negative_regulation_of_neurogenesis	0.0123	1.0000
regulation_of_cell_adhesion	0.0125	0.1399
positive_regulation_of_cell_proliferation	0.0125	0.9564
amnioserosa_formation	0.0125	1.0000
establishment_and/or_maintenance_of_apical/basal_cell_polarity	0.0125	1.0000
genital_disc_sexually_dimorphic_development	0.0125	1.0000
negative_regulation_of_imaginal_disc_growth	0.0125	1.0000
regulation_of_R8_cell_spacing_in_compound_eye	0.0125	1.0000
sex_determination,_establishment_of_X:A_ratio	0.0125	1.0000
spermatid_development	0.0125	1.0000
fibroblast_growth_factor_receptor_signaling_pathway	0.0139	0.4171
mesodermal_cell_fate_specification	0.0139	0.4171
JAK-STAT_cascade	0.0139	1.0000
subapical_complex	0.0139	1.0000
regulation_of_mitosis	0.0147	1.0000
cell-cell_signaling	0.0147	0.1009
localization	0.0149	0.8871
regulation_of_phosphate_metabolic_process	0.0153	0.2293
regulation_of_phosphorus_metabolic_process	0.0153	0.2293
apical_plasma_membrane	0.0196	1.0000
epidermis_development	0.0196	0.4763
cell_projection_biogenesis	0.0196	1.0000
zonula_adherens_assembly	0.0196	1.0000
cellularization	0.0204	0.7314
positive_regulation_of_RNA_metabolic_process	0.0209	0.5397
positive_regulation_of_transcription,_DNA-dependent	0.0209	0.5397
cell_cortex_part	0.0225	1.0000
R8_cell_differentiation	0.0225	1.0000
analia_development	0.0225	1.0000
endoderm_development	0.0225	1.0000
oenocyte_development	0.0225	1.0000
regulation_of_hemocyte_differentiation	0.0239	0.8429
smoothened_signaling_pathway	0.0252	0.5491
plasma_membrane_part	0.0257	0.0767
imaginal_disc_fusion,_thorax_closure	0.0261	0.5397
antennal_morphogenesis	0.0261	1.0000
death	0.0270	0.8256
imaginal_disc-derived_leg_morphogenesis	0.0282	0.0582
anterior_midgut_development	0.0282	0.5803
dorsal_vessel_heart_proper_cell_fate_commitment	0.0282	0.5803
neuroblast_fate_specification	0.0282	0.5803
neuron_migration	0.0282	0.5803
notum_development	0.0282	0.5803
Malpighian_tubule_morphogenesis	0.0282	0.8871
L-lactate_dehydrogenase_activity	0.0282	1.0000
anterior_commissure_morphogenesis	0.0282	1.0000
apical_junction_complex	0.0282	1.0000
central_nervous_system_neuron_axonogenesis	0.0282	1.0000
central_nervous_system_projection_neuron_axonogenesis	0.0282	1.0000
cytoneme_biogenesis	0.0282	1.0000
lactate_dehydrogenase_activity	0.0282	1.0000
positive_regulation_of_nurse_cell_apoptosis	0.0282	1.0000
regulation_of_MAPKKK_cascade	0.0282	1.0000

specification_of_segmental_identity_mandibular_segment	0.0282	1.0000
specification_of_segmental_identity_thorax	0.0282	1.0000
regulation_of_protein_metabolic_process	0.0322	0.1009
negative_regulation_of_translation	0.0328	0.5803
JNK_cascade	0.0340	0.3503
ligand-dependent_nuclear_receptor_activity	0.0340	0.3503
spindle_assembly	0.0340	1.0000
BMP_signaling_pathway	0.0342	1.0000
chromatin-mediated_maintenance_of_transcription	0.0342	1.0000
cytokinesis_after_meiosis	0.0342	1.0000
early_endosome	0.0342	1.0000
imaginal_disc-derived_genitalia_development	0.0342	1.0000
negative_regulation_of_cellular_protein_metabolic_process	0.0342	1.0000
positive_regulation_of_gene_expression_epigenetic	0.0342	1.0000
regulation_of_JAK-STAT_cascade	0.0342	1.0000
retinal_ganglion_cell_axon_guidance	0.0342	1.0000
biopolymer_biosynthetic_process	0.0363	1.0000
receptor_binding	0.0371	0.0643
negative_regulation_of_cellular_biosynthetic_process	0.0371	0.6011
negative_regulation_of_protein_metabolic_process	0.0375	1.0000
regulation_of_cell_cycle_process	0.0393	0.9106
imaginal_disc-derived_limb_morphogenesis	0.0398	0.0775
germ_cell_development	0.0398	0.3697
negative_regulation_of_cell_proliferation	0.0398	0.3697
intracellular	0.0422	1.0000
programmed_cell_death	0.0434	0.7189
leg_morphogenesis	0.0476	0.0866
stress-activated_protein_kinase_signaling_pathway	0.0476	0.3697
photoreceptor_cell_morphogenesis	0.0488	0.6638
developmental_process	0.0000	0.0000
anatomical_structure_development	0.0000	0.0000
anatomical_structure_morphogenesis	0.0000	0.0000
organ_development	0.0000	0.0000
system_development	0.0000	0.0000
regulation_of_cellular_process	0.0000	0.0000
regulation_of_biological_process	0.0000	0.0000
biological_regulation	0.0000	0.0000
pattern_specification_process	0.0000	0.0000
cell_migration	0.0000	0.0000
cell_motility	0.0000	0.0000
regionalization	0.0000	0.0001
transcription_regulator_activity	0.0000	0.0000
transcription_factor_activity	0.0000	0.0000
regulation_of_macromolecule_metabolic_process	0.0000	0.0000
regulation_of_biosynthetic_process	0.0000	0.0000
regulation_of_macromolecule_biosynthetic_process	0.0000	0.0000
regulation_of_transcription	0.0000	0.0000
regulation_of_metabolic_process	0.0000	0.0000
regulation_of_cellular_metabolic_process	0.0000	0.0000
regulation_of_gene_expression	0.0000	0.0000
nucleus	0.0000	0.0000
regulation_of_nucleobase_nucleoside_nucleotide_and_nucleic_acid_metabolic_process	0.0000	0.0000
regulation_of_transcription_DNA-dependent	0.0000	0.0000
regulation_of_RNA_metabolic_process	0.0000	0.0000
regulation_of_developmental_process	0.0000	0.0000
intracellular_membrane-bounded_organelle	0.0000	0.0000
membrane-bounded_organelle	0.0000	0.0000
open_tracheal_system_development	0.0000	0.0263
sequence-specific_DNA_binding	0.0000	0.0000
intracellular_organelle	0.0000	0.0000
organelle	0.0000	0.0000
cellular_developmental_process	0.0000	0.0000
cell_fate_determination	0.0000	0.0056
axon_guidance	0.0000	0.0000
DNA_binding	0.0000	0.0000
imaginal_disc_pattern_formation	0.0000	0.0008
sensory_organ_development	0.0000	0.0000
cell_development	0.0000	0.0009
nucleic_acid_binding	0.0000	0.0004
positive_regulation_of_cellular_process	0.0000	0.0000
organ_morphogenesis	0.0000	0.0002
RNA_polymerase_II_transcription_factor_activity	0.0000	0.0000
specific_RNA_polymerase_II_transcription_factor_activity	0.0000	0.0057
morphogenesis_of_an_epithelium	0.0000	0.0230
positive_regulation_of_biological_process	0.0000	0.0000
tissue_development	0.0000	0.0007
heart_development	0.0000	0.0000
negative_regulation_of_biological_process	0.0000	0.0000

cellular_structure_morphogenesis	0.0000	0.0000
peripheral_nervous_system_development	0.0000	0.0000
tube_morphogenesis	0.0000	0.0499
negative_regulation_of_cellular_process	0.0000	0.0000
wing_disc_pattern_formation	0.0000	0.0057
regulation_of_transcription_from_RNA_polymerase_II_promoter	0.0000	0.0057
cell_fate_commitment	0.0000	0.0000
cell_communication	0.0000	0.0000
cell_part_morphogenesis	0.0000	0.0000
cell_projection_morphogenesis	0.0000	0.0000
neurite_morphogenesis	0.0000	0.0000
imaginal_disc-derived_wing_vein_morphogenesis	0.0000	0.0023
signal_transduction	0.0000	0.0001
embryonic_morphogenesis	0.0000	0.0088
imaginal_disc_morphogenesis	0.0000	0.0060
protein_binding	0.0000	0.0000
morphogenesis_of_embryonic_epithelium	0.0000	0.0055
binding	0.0000	0.0002
dorsal/ventral_pattern_formation,_imaginal_disc	0.0000	0.0425
eye_development	0.0000	0.0005
brain_development	0.0000	0.0102
anatomical_structure_formation	0.0000	0.0183
anterior/posterior_pattern_formation,_imaginal_disc	0.0000	0.0360
intracellular_part	0.0000	0.0014
appendage_development	0.0000	0.0000
imaginal_disc-derived_appendage_development	0.0000	0.0000
appendage_morphogenesis	0.0000	0.0000
dorsal_closure	0.0000	0.0293
muscle_development	0.0000	0.0001
mesoderm_development	0.0000	0.0180
cell_part	0.0000	0.0110
imaginal_disc-derived_appendage_morphogenesis	0.0000	0.0000
regulation_of_signal_transduction	0.0000	0.0036
zinc_ion_binding	0.0000	0.0149
cell_surface_receptor_linked_signal_transduction	0.0000	0.0007
cellular_process	0.0000	0.0057
plasma_membrane	0.0000	0.0004
compound_eye_development	0.0000	0.0223
formation_of_anatomical_boundary	0.0000	0.0003
axonogenesis	0.0000	0.0000
formation_of_organ_boundary	0.0000	0.0007
imaginal_disc-derived_wing_morphogenesis	0.0000	0.0005
cell_division	0.0000	0.0059
transcription_repressor_activity	0.0000	0.0271
cell_adhesion	0.0000	0.0005
motor_axon_guidance	0.0000	0.0425
biological_adhesion	0.0000	0.0010
cellular_component_organization_and_biogenesis	0.0001	0.0000
ommatidial_rotation	0.0001	0.0114
extracellular_matrix	0.0002	0.0064
regulation_of_developmental_growth	0.0002	0.0150
eye_morphogenesis	0.0003	0.0003
extracellular_region_part	0.0004	0.0007
haltere_development	0.0004	0.0372
ovarian_follicle_cell_development	0.0004	0.0060
embryonic_development_via_the_syncytial_blastoderm	0.0005	0.0066
positive_regulation_of_programmed_cell_death	0.0005	0.0128
female_gamete_generation	0.0005	0.0004
oogenesis	0.0005	0.0004
proteinaceous_extracellular_matrix	0.0006	0.0042
embryonic_development	0.0007	0.0017
positive_regulation_of_developmental_process	0.0010	0.0499
negative_regulation_of_macromolecule_metabolic_process	0.0011	0.0002
negative_regulation_of_cellular_metabolic_process	0.0011	0.0005
negative_regulation_of_metabolic_process	0.0012	0.0005
positive_regulation_of_biosynthetic_process	0.0013	0.0004
embryonic_development_ending_in_birth_or_egg_hatching	0.0013	0.0128
negative_regulation_of_macromolecule_biosynthetic_process	0.0014	0.0000
negative_regulation_of_biosynthetic_process	0.0015	0.0000
positive_regulation_of_macromolecule_biosynthetic_process	0.0018	0.0013
chemotaxis	0.0018	0.0071
cardioblast_differentiation	0.0021	0.0183
positive_regulation_of_cellular_metabolic_process	0.0022	0.0007
positive_regulation_of_metabolic_process	0.0022	0.0007
gamete_generation	0.0023	0.0006
cell_morphogenesis	0.0024	0.0261
reproduction	0.0024	0.0006
sexual_reproduction	0.0024	0.0006

positive_regulation_of_macromolecule_metabolic_process	0.0030	0.0021
negative_regulation_of_signal_transduction	0.0030	0.0092
positive_regulation_of_gene_expression	0.0030	0.0141
positive_regulation_of_nucleobase_nucleoside_nucleotide_and_nucleic_acid_metabolic_process	0.0030	0.0141
positive_regulation_of_transcription	0.0030	0.0141
glial_cell_migration	0.0033	0.0062
taxis	0.0037	0.0042
embryonic_heart_tube_development	0.0037	0.0127
tube_development	0.0037	0.0127
compound_eye_morphogenesis	0.0042	0.0107
neuron_recognition	0.0046	0.0001
cell_recognition	0.0054	0.0001
Bolwig's_organ_morphogenesis	0.0061	0.0372
locomotory_behavior	0.0063	0.0002
cell-cell_adhesion	0.0072	0.0151
cellular_component	0.0076	0.0169
gonad_development	0.0104	0.0150
reproductive_structure_development	0.0104	0.0150
behavior	0.0186	0.0000
negative_regulation_of_transcription_from_RNA_polymerase_II_promoter	0.0209	0.0119
negative_regulation_of_gene_expression	0.0254	0.0000
negative_regulation_of_transcription	0.0254	0.0000
cellular_morphogenesis_during_differentiation	0.0261	0.0141
post-embryonic_morphogenesis	0.0505	0.0263
sensory_organ_boundary_specification	0.0505	0.0263
negative_regulation_of_nucleobase_nucleoside_nucleotide_and_nucleic_acid_metabolic_process	0.0539	0.0001
biological_process	0.0545	0.0223
Wnt_receptor_signaling_pathway_calcium_modulating_pathway	0.0712	0.0372
Wnt_receptor_signaling_pathway	0.0877	0.0141
striated_muscle_development	0.1202	0.0055
negative_regulation_of_transcription_DNA-dependent	0.1460	0.0002
response_to_stimulus	0.1649	0.0003
axonal_fasciculation	0.1791	0.0007
extracellular_structure_organization_and_biogenesis	0.1791	0.0014
negative_regulation_of_RNA_metabolic_process	0.1802	0.0003
regulation_of_muscle_development	0.2499	0.0120
regulation_of_striated_muscle_development	0.2499	0.0120
homophilic_cell_adhesion	0.3120	0.0035
larval_central_nervous_system_remodeling	0.3120	0.0183
regulation_of_epidermal_growth_factor_receptor_signaling_pathway	0.3471	0.0004
synapse_organization_and_biogenesis	0.3471	0.0050
cell_migration_involved_in_gastrulation	0.3471	0.0263
mesoderm_migration	0.3471	0.0263
muscle_fiber_development	0.3471	0.0263
cell_surface	0.4087	0.0023
cardioblast_cell_fate_commitment	0.5727	0.0223
axon_extension	0.8459	0.0198
response_to_chemical_stimulus	0.8851	0.0032
response_to_external_stimulus	0.8851	0.0042
somatic_muscle_development	1.0000	0.0002
olfactory_learning	1.0000	0.0004
myoblast_fusion	1.0000	0.0005
plasma_membrane_fusion	1.0000	0.0005
syncytium_formation	1.0000	0.0005
syncytium_formation_by_plasma_membrane_fusion	1.0000	0.0005
learning	1.0000	0.0007
olfactory_behavior	1.0000	0.0010
learning_and/or_memory	1.0000	0.0013
chemosensory_behavior	1.0000	0.0015
membrane_fusion	1.0000	0.0057
asymmetric_neuroblast_division	1.0000	0.0062
neuroblast_division	1.0000	0.0062
multicellular_organismal_process	1.0000	0.0068
ecdysis_chitin-based_cuticle	1.0000	0.0120
ecdysis_protein-based_cuticle	1.0000	0.0120
circadian_rhythm	1.0000	0.0202
netrin_receptor_activity	1.0000	0.0223
rhythmic_process	1.0000	0.0223
negative_regulation_of_smoothed_signaling_pathway	1.0000	0.0263
skeletal_muscle_development	1.0000	0.0372
visceral_muscle_development	1.0000	0.0372

Supplemental Table S3**Motif-content predicts TF binding**

Prediction accuracies (ACC) and receiver-operating-characteristic (ROC) curves (area under the curve AUC) for the Early vs. Late binding, Stage-specific vs. Continuous binding and Bound vs. Not bound motifs for the mesodermal TFs: Biniou (BIN), Twist (TWI), Tinman (TIN) and Myocyte enhancer factor 2 (MEF2) obtained from Zinzen et al. 2009 and their respective controls (see methods). hpf refers to "hours post fertilization", cont. refers to continuous and NB refers to Not Bound motifs for the given TF (for details see methods). The total number of final selected motifs are shown in the 4th column.

	Conditions		Final # Motifs	SVM Prediction		Control	
	Condition 1	Condition 2		ACC	AUC	ACC	AUC
Stage-specific Binding (Early vs. Late)	BIN 6-8 hpf	BIN 10-12 hpf	50	70.00	0.73	53.37	0.53
	TIN 2-4 hpf	TIN 6-8 hpf	25	72.67	0.77	54.33	0.54
	TWI 2-4 hpf	TWI 6-8 hpf	35	76.23	0.76	53.38	0.54
	MEF2 2-4 hpf	MEF2 8-10 hpf	44	74.33	0.78	52.13	0.50
	MEF2 2-4 hpf	MEF2 10-12 hpf	19	69.59	0.73	51.79	0.52
	MEF2 4-6 hpf	MEF2 8-10 hpf	37	72.26	0.78	51.36	0.52
	MEF2 4-6 hpf	MEF2 10-12 hpf	42	76.83	0.82	52.17	0.52
Stage-specific Binding (Stage-specific vs. Continuous)	BIN 6-8 hpf	BIN cont.	42	69.55	0.74	51.88	0.51
	BIN 8-10 hpf	BIN cont.	39	80.75	0.86	55.49	0.55
	BIN 10-12 hpf	BIN cont.	26	68.23	0.72	53.34	0.54
	TIN 2-4 hpf	TIN cont.	23	72.50	0.77	50.47	0.49
	TIN 4-6 hpf	TIN cont.	39	67.25	0.70	52.51	0.52
	TIN 6-8 hpf	TIN cont.	26	71.51	0.75	53.61	0.54
	TWI 2-4 hpf	TWI cont.	41	76.21	0.81	51.81	0.51
	TWI 4-6 hpf	TWI cont.	57	73.67	0.78	52.97	0.52
	TWI 6-8 hpf	TWI cont.	39	72.29	0.76	51.70	0.53
	MEF2 2-4 hpf	MEF2 cont.	18	74.54	0.81	58.88	0.60
	MEF2 4-6 hpf	MEF2 cont.	47	69.60	0.74	54.90	0.55
	MEF2 6-8 hpf	MEF2 cont.	41	71.45	0.76	54.15	0.55
	MEF2 8-10 hpf	MEF2 cont.	24	66.98	0.71	55.96	0.57
	MEF2 10-12 hpf	MEF2 cont.	53	69.88	0.74	54.99	0.56
Bound vs. Not-Bound (NB) Motifs	BIN 6-8 hpf	BIN NB	38	84.02	0.91	52.23	0.52
	BIN 8-10 hpf	BIN NB	14	76.89	0.82	53.51	0.56
	BIN 10-12 hpf	BIN NB	50	82.53	0.89	51.97	0.53
	BIN cont.	BIN NB	52	87.97	0.93	52.90	0.54
	TIN 2-4 hpf	TIN NB	22	72.50	0.76	51.97	0.50
	TIN 4-6 hpf	TIN NB	37	72.34	0.76	51.62	0.52
	TIN 6-8 hpf	TIN NB	53	75.90	0.81	54.32	0.55
	TIN cont.	TIN NB	40	81.78	0.87	53.70	0.53
	TWI 2-4 hpf	TWI NB	35	84.93	0.90	51.25	0.51
	TWI 4-6 hpf	TWI NB	47	73.44	0.79	51.74	0.51
	TWI 6-8 hpf	TWI NB	40	77.76	0.84	53.06	0.54
	TWI cont.	TWI NB	31	85.96	0.93	50.88	0.49
	MEF2 2-4 hpf	MEF2 NB	22	77.59	0.82	49.94	0.44
	MEF2 4-6 hpf	MEF2 NB	25	77.65	0.83	53.31	0.53
	MEF2 6-8 hpf	MEF2 NB	36	78.83	0.84	53.28	0.54
	MEF2 8-10 hpf	MEF2 NB	36	79.62	0.84	52.85	0.53
	MEF2 10-12 hpf	MEF2 NB	43	76.1	0.81	53.33	0.53
	MEF2 cont.	MEF2 NB	59	82.33	0.88	53.08	0.52

Supplemental Table S4**Motif-clusters for *Drosophila* motif dataset.**

The cluster was formed by grouping motifs according to their co-occurrences in random genomic regions by a k-mean clustering (see methods). Motifs were obtained from Stark et al. 2007. The original source is shown in the last column.

Motif Name	IUPAC	Cluster ID	Source
br-Z1_M00091	WDRTMAWWRACAARWNHV	1	TRANSFAC
br-Z4_MA0013	TWRTAACWAW	2	JASPAR
br-Z1_MA0010	RTMAWWRACAARW	2	JASPAR
h_Schroeder	GRCACGCKVS	3	Schroeder et al. 2004
run_Schroeder	YGYGGTY	3	Schroeder et al. 2004
ttk_M00009	GSTCCTGS	3	TRANSFAC
Cf1_M00111	GRGGTCAYS	3	TRANSFAC
Cf1_M00112	GGGGTCACS	3	TRANSFAC
Sry-beta_M00666	CKCAKCWCT	3	TRANSFAC
usp_MA0016	GGGGTCACSS	3	JASPAR
E-spl_manual	TGGCACGTGYA	3	Stark et al. 2007
Ets_manual	CHGGAW	3	Stark et al. 2007
ac_manual	RRNNMCACCTGC	3	Stark et al. 2007
grh_Down	YNAACYGGTYTGC	3	Down et al. 2007
prd_PD-Down	DBYCGTCACGSTYSRBV	3	Down et al. 2007
shn-ZFP1_Down	GGKAWCGTTCCC	3	Down et al. 2007
usp_Down	GGGGTCACS	3	Down et al. 2007
ME35	RACASCTGY	3	Stark et al. 2007
h_M00067	VNGRCACGCKVSHK	3	TRANSFAC
h_Down	VNGRCACGCKVSHK	3	Down et al. 2007
Adf-1_M00171	BMGYBGYYGYNGMYBV	4	TRANSFAC
Adf-1_M00923	VSGYYGCMGYCGYYGMMKKYG	4	TRANSFAC
sc_manual	GCAGSTGK	4	Stark et al. 2007
Adf-1_Down	GCTSCYKCYGYH	4	Down et al. 2007
z_M00283	BNWDTYGAGTGRNHDD	5	TRANSFAC
Zeste_M00934	WDTYGAGTGR	5	TRANSFAC
eyg_Down	KVRKRNTCACTSRNTVHDB	5	Down et al. 2007
ey-toy_manual	TGGAGDGGWAHTMATBVRTGWDDDRKKMW	6	Stark et al. 2007
vnd_manual	TYAAGTGS	7	Stark et al. 2007
tin_Down	CACTTRA	7	Down et al. 2007
vnd_Down	WDWBAAGTRSBNB	7	Down et al. 2007
acj6_manual	TGCATAATTAATTAC	8	Stark et al. 2007
ems_Down	AANTNTAATGACA	8	Down et al. 2007
ME6	VATTWGCAT	8	Stark et al. 2007
ME12	TTAATGATG	8	Stark et al. 2007
ME25	TTAYGTAA	8	Stark et al. 2007
ME36	TGTCAATTG	8	Stark et al. 2007
ME55	TAATATGCRA	8	Stark et al. 2007
ME57	RCGTGYAAAT	8	Stark et al. 2007
ME77	TGAYWWWTGCA	8	Stark et al. 2007
ME84	TRTCAWNWRTCA	8	Stark et al. 2007
ME110	ATCATAA	8	Stark et al. 2007
ME112	MAACAA	8	Stark et al. 2007
ME117	MAAMNNCAA	8	Stark et al. 2007
ME125	TAAATAG	8	Stark et al. 2007
ME133	AAAAGCT	8	Stark et al. 2007
ME143	TNRCGCNYNATTAATY	8	Stark et al. 2007
tin_manual	CTYAAGTGG	9	Stark et al. 2007
ase_manual	CAGSTG	9	Stark et al. 2007
gl_manual	CAATGCACCTTCTGGGGCTTCCAC	9	Stark et al. 2007
Su-H_Down	RTGRGAR	9	Down et al. 2007
ME2	AWNTGGGTCA	9	Stark et al. 2007
ME8	HRTCAATCA	9	Stark et al. 2007
ME10	RCGTGNNNNGCAT	9	Stark et al. 2007
ME31	GCGTSAAA	9	Stark et al. 2007
ME32	YCGGYRTCAWT	9	Stark et al. 2007
ME33	GCGTTGAYA	9	Stark et al. 2007
ME46	RAGTGAAAGT	9	Stark et al. 2007
ME50	TRACRYGCA	9	Stark et al. 2007

ME52	RKGTCAAGK	9	Stark et al. 2007
ME56	TNATGNNTGACA	9	Stark et al. 2007
ME59	GCGYNWNAWTGAY	9	Stark et al. 2007
ME63	AAATCAAT	9	Stark et al. 2007
ME64	CNNNGCGYRTGANYNAT	9	Stark et al. 2007
ME65	ACGYNGCGTATGM	9	Stark et al. 2007
ME66	AAAATGCA	9	Stark et al. 2007
ME67	CATTANYGTCA	9	Stark et al. 2007
ME71	TGTCAMNTGCA	9	Stark et al. 2007
ME74	YGATAAGC	9	Stark et al. 2007
ME78	AATTNNNTCACGY	9	Stark et al. 2007
ME79	GCTNMTTAA	9	Stark et al. 2007
ME97	GCGNNTNNNTTA	9	Stark et al. 2007
ME99	GNMCTTGAA	9	Stark et al. 2007
ME101	ATGANNNNTCA	9	Stark et al. 2007
ME103	YRAAMGTGM	9	Stark et al. 2007
ME104	CTTNNATAC	9	Stark et al. 2007
ME106	GGTNTAAAW	9	Stark et al. 2007
ME111	ATCTNATC	9	Stark et al. 2007
ME114	TTTNGGCGS	9	Stark et al. 2007
ME115	GATNTCAT	9	Stark et al. 2007
ME116	TGGATTA	9	Stark et al. 2007
ME126	CAANNNA	9	Stark et al. 2007
ME129	ATGTGAT	9	Stark et al. 2007
ME135	AAANNNNNNCAA	9	Stark et al. 2007
ME136	AAANNNNNNNNNAAT	9	Stark et al. 2007
ME137	GAGAGAG	9	Stark et al. 2007
ME139	ATANANNCGC	9	Stark et al. 2007
ME140	TCANNNTGGA	9	Stark et al. 2007
ME142	CTATNNNAAG	9	Stark et al. 2007
dTCF_M00362	CCTTTGATCTT	9	TRANSFAC
pan_Down	CCTTTGATCTT	9	Down et al. 2007
kni_Berman	AAHKMTHBCA	10	Berman et al. 2004
croc_Down	RTAAMA	10	Down et al. 2007
mirr_Down	RWWWASWBDYSTDMW	10	Down et al. 2007
ME81	CTCRTAAAW	10	Stark et al. 2007
TorRE_Schroeder	TGCTCAATGAA	11	Schroeder et al. 2004
sna_M00044	MSMASBTGHTDVSW	11	TRANSFAC
sd_manual	RRAYATTYBKSGVATKVCA	11	Stark et al. 2007
ME94	RAGTKCAANG	11	Stark et al. 2007
mtTFA_M00487	KNMTTATSVNH	12	TRANSFAC
pnr_Down	ANHDDBGATAASSDNNB	12	Down et al. 2007
srp_Down	KHGATAASR	12	Down et al. 2007
ap_M00199	VTKASTCAK	13	TRANSFAC
byn_Down	RWWNTNRCACYT	13	Down et al. 2007
ME9	TGACANNNNNTGACA	13	Stark et al. 2007
ME13	WTGACANBT	13	Stark et al. 2007
ME19	MATTRRCACNY	13	Stark et al. 2007
ME27	AATTRYGWCA	13	Stark et al. 2007
ME41	TYRACACTTK	13	Stark et al. 2007
ME44	TGACANNNTTGAC	13	Stark et al. 2007
ME70	GACAATK	13	Stark et al. 2007
ME76	TGACAWWTWTGC	13	Stark et al. 2007
ME144	GCANTTGYNWYAATT	13	Stark et al. 2007
Dstat_Schroeder	TTCCSGGAA	14	Schroeder et al. 2004
prd_HD-Schroeder	BKKGDTAATTKMBKD	15	Schroeder et al. 2004
slbo_Down	TTNNRCAATM	15	Down et al. 2007
ME51	RCAAWTTR	15	Stark et al. 2007
Antp_M00692	TTWKYMW	16	TRANSFAC
zen_M00710	WCATTWMM	17	TRANSFAC
cad_Schroeder	BRDHTTTTWWRRHH	18	Schroeder et al. 2004
br-Z1_Down	TAAWWRACAAARW	18	Down et al. 2007
br-Z2_Down	TTWWCTATTT	18	Down et al. 2007
br-Z3_Down	WAAACWWRW	18	Down et al. 2007
br-Z3_M00093	DRVWAAACWARWHS	19	TRANSFAC
br-Z3_MA0012	WAAACWARWHS	19	JASPAR
ME95	AAASTTT	20	Stark et al. 2007
sna_M00060	VHRRCAGGTGYMV	21	TRANSFAC
sna_MA0086	CAGGTG	21	JASPAR
sna_Down	VHRRCAGGTGYMV	21	Down et al. 2007

cad_Berman	DDNYNYAAAD	22	Berman et al. 2004
Su-H_M00234	RBYGTGRGAAMCB	23	TRANSFAC
ovo_M00461	HHHSWAACHGHV	23	TRANSFAC
Su-H_MA0085	BYGTGRGAAMCBNDVD	23	JASPAR
ovo_Down	HSWAACHGH	23	Down et al. 2007
gcm_M00270	VDRCCCGCATVHD	24	TRANSFAC
gcm_Down	ATGCGGGY	24	Down et al. 2007
grh_M00951	WCYGGTTT	25	TRANSFAC
ato_manual	AWCAGGTGK	25	Stark et al. 2007
ME83	MACMDGTTK	25	Stark et al. 2007
ME87	MACTTGTYR	25	Stark et al. 2007
hb_Schroeder	VTTTTTWBKHBH	26	Schroeder et al. 2004
hb_M00022	SMVHAAAAAW	26	TRANSFAC
hb_MA0049	SMVHAAAAAW	26	JASPAR
Antp_manual	TTAATT	27	Stark et al. 2007
ap_manual	TGANTCA	27	Stark et al. 2007
CrebA_manual	TGACGTCA	27	Stark et al. 2007
Unknown-2_manual	GGGGGATT	27	Stark et al. 2007
pnt_manual	GAGGAAGC	27	Stark et al. 2007
repo_manual	CAATTA	27	Stark et al. 2007
sim_manual	GTACGTG	27	Stark et al. 2007
so_manual	YGATAC	27	Stark et al. 2007
Aef1_Down	CAACAA	27	Down et al. 2007
Hr46_Down	GGGTCA	27	Down et al. 2007
cad_Down	TTTATG	27	Down et al. 2007
pho_Down	GCCATT	27	Down et al. 2007
ME1	GTCACGTD	27	Stark et al. 2007
ME16	TGACGTCAT	27	Stark et al. 2007
ME28	GCGCATGH	27	Stark et al. 2007
ME38	CASGTAR	27	Stark et al. 2007
ME39	WCACGTGC	27	Stark et al. 2007
ME45	VCACGCRH	27	Stark et al. 2007
ME49	CGTGNGAA	27	Stark et al. 2007
ME54	RCGYRCGY	27	Stark et al. 2007
ME62	TGGCGCC	27	Stark et al. 2007
ME73	AAAGTGANA	27	Stark et al. 2007
ME80	GTCANTNAAC	27	Stark et al. 2007
ME82	TNAGCATAA	27	Stark et al. 2007
ME88	ACGNNAATTG	27	Stark et al. 2007
ME91	CATNNNNCGCG	27	Stark et al. 2007
ME92	TAACCTC	27	Stark et al. 2007
ME93	GCAACA	27	Stark et al. 2007
ME107	ACTNACCT	27	Stark et al. 2007
ME118	GTCNNNNGACA	27	Stark et al. 2007
ME120	AACTGA	27	Stark et al. 2007
ME121	ACANACA	27	Stark et al. 2007
ME123	CACNNNNNNACA	27	Stark et al. 2007
ME124	GCANGTCC	27	Stark et al. 2007
ME128	CACGAGNC	27	Stark et al. 2007
ME134	CAGNNGCA	27	Stark et al. 2007
ME138	TCCTNNNNNNNGGA	27	Stark et al. 2007
Unknown-1_manual	TATCGATA	27	Stark et al. 2007
Dref_Down	TATCGATA	27	Down et al. 2007
Elf-1_M00110	DDNBKGTDTHDHV	28	TRANSFAC
dl_MA0022	SGGKKTTTYCV	29	JASPAR
Hsf_1-M00028	MGAAD	30	TRANSFAC
dl_M00120	HSRGAAAAHYV	30	TRANSFAC
Hsf_3-M00164	MGAADMGAADHTTCK	31	TRANSFAC
Hsf_4-M00165	MGAADHTTCKMGAAD	31	TRANSFAC
Hsf_5-M00166	HTTCKMGAADMGAAD	31	TRANSFAC
en_Down	TAATTR	32	Down et al. 2007
dri_M00455	WHRATTAAW	32	TRANSFAC
dri_Down	WHRATTAAW	32	Down et al. 2007
Hsf_2-M00163	MGAADMGAADMGAAD	33	TRANSFAC
Dref_M00488	WVYTATCGWWDDH	34	TRANSFAC
Unknown-3_manual	RTAAACAA	35	Stark et al. 2007
br-Z4_Down	RKAAASA	35	Down et al. 2007
ME96	RTAAACA	35	Stark et al. 2007
GAGA_M00723	RSWGAGMRHRR	36	TRANSFAC
STAT_M00259	HBDTTTCCSGGAAABNNHD	37	TRANSFAC

Stat92E_Down	HBDTTTCCSGGAAABNNHD	37	Down et al. 2007
ME5	STATAWAWR	38	Stark et al. 2007
ME100	GTATNWATA	38	Stark et al. 2007
ttk_Schroeder	MKSCMAGGACVHH	39	Schroeder et al. 2004
ftz-f1_Schroeder	YSAAGGWCRCHRM	39	Schroeder et al. 2004
EcR-usp_Down	VRGKTYAWTGAMMY	39	Down et al. 2007
ME14	YGACMTTGA	39	Stark et al. 2007
ME86	YGTCAWTGAC	39	Stark et al. 2007
kr_Schroeder	RAAMGRGTT	40	Schroeder et al. 2004
kr_Berman	RAAMGGRTTA	40	Berman et al. 2004
kr_M00021	AMYGGGTAW	40	TRANSFAC
kr_Down	AACGGGTAA	40	Down et al. 2007
ME22	WAATGCGCNT	41	Stark et al. 2007
ME26	YGCGTHAATTR	41	Stark et al. 2007
ME29	WAATCARCGC	41	Stark et al. 2007
ME42	RCGCMATTW	41	Stark et al. 2007
ME68	GCATAHWWNNNGCGY	41	Stark et al. 2007
ME109	ATCWATG	41	Stark et al. 2007
ME145	RTTRCGYATRCGCM	41	Stark et al. 2007
tll_Schroeder	MTTWGRCTT	42	Schroeder et al. 2004
eve_hd-Schroeder	TCAWTTAAMT	43	Schroeder et al. 2004
ME58	TAATTNWMATT	43	Stark et al. 2007
br-Z4_M00094	WWWRKAAASAWAW	44	TRANSFAC
ME141	AATNNNNNNNA	45	Stark et al. 2007
gt_Schroeder	TWTKACKTAAWA	46	Schroeder et al. 2004
bab1_Down	WHWWWWWWWWKK	47	Down et al. 2007
br-Z2_M00092	DYBTKWWCTATTTKKY	48	TRANSFAC
br-Z2_MA0011	WWCTATTT	48	JASPAR
d-mtTFA_Down	TTATS	48	Down et al. 2007
ME3	BCATAAATYA	48	Stark et al. 2007
croc_M00266	HWNARTAAAYAKHNKR	49	TRANSFAC
ftz_Schroeder	KYTAATKDN	50	Schroeder et al. 2004
ftz_M00020	VVDVYAATTAAG	50	TRANSFAC
en_M00696	YYWWTTA	50	TRANSFAC
ME7	BYAATTARH	50	Stark et al. 2007
Cf2_M00013	RTATATRTA	51	TRANSFAC
Cf2_M00012	RTATATRTR	51	TRANSFAC
Cf2_Down	RTATATRTR	51	Down et al. 2007
Cf2_MA0015	RTATATRTRB	52	JASPAR
Mef2_manual	YTAWWWWTAR	53	Stark et al. 2007
hb_Berman	TTTTTTRHBNB	54	Berman et al. 2004
hb_Down	SMATAAAAAA	54	Down et al. 2007
TATA_M00252	STATAWAWRSVVV	55	TRANSFAC
ME30	AATTAANNNNNCATNA	56	Stark et al. 2007
ME61	AACWAATTR	56	Stark et al. 2007
ME113	ATTNWTTA	56	Stark et al. 2007
bin_Down	RTAAATA	57	Down et al. 2007
sd_Down	WRNATKTNT	57	Down et al. 2007
ME119	GNCTANWWATA	57	Stark et al. 2007
ME4	HAATTAYGCRH	58	Stark et al. 2007
ME11	MATTAAWNATGCR	58	Stark et al. 2007
ME23	MATTWRTCA	58	Stark et al. 2007
ME24	YAATTWNRYGC	58	Stark et al. 2007
ME53	YGTCWAATTA	58	Stark et al. 2007
ME69	AATKACA	58	Stark et al. 2007
ME85	TAATTRNNNNYGACA	58	Stark et al. 2007
ME105	GYATGMGWAATKA	58	Stark et al. 2007
en_manual	TAATTAA	59	Stark et al. 2007
tll_Down	AAATTAA	59	Down et al. 2007
ME15	AATTRNNNNCAATT	59	Stark et al. 2007
ME21	TAATTRANNTTNATG	59	Stark et al. 2007
ME60	CATNANTYAAA	59	Stark et al. 2007
kni_Schroeder	DDRDMWMSKRMDVR	60	Schroeder et al. 2004
eve_Down	TCAWTTAAMT	61	Down et al. 2007
zen_Down	TCAWTTAAMT	61	Down et al. 2007
ME89	CACRCAC	62	Stark et al. 2007
ME122	ACACNNNNRCAC	62	Stark et al. 2007
ME127	CACNNRNNNNNNCAC	62	Stark et al. 2007
ME131	ACACNNACAC	62	Stark et al. 2007
ME132	RCACNNNNNNNCACA	62	Stark et al. 2007

prd_HD-Down	SVTAATYGATTANS	63	Down et al. 2007
tlI_M00679	RAKWYWAV	64	TRANSFAC
dsx_Down	BYRHBACAAWGTDDDB	64	Down et al. 2007
bcd_Schroeder	BYTAAKCYSBV	65	Schroeder et al. 2004
bcd_Berman	HWAACCYB	65	Berman et al. 2004
bcd_M00140	SGGATTAW	65	TRANSFAC
bcd_Down	VVVSTAATCC	65	Down et al. 2007
oc_Down	MVHTAAKCCS	65	Down et al. 2007
ftz_Down	VNRYTAATGRBM	66	Down et al. 2007
ME75	TAATKRNGTCATTA	66	Stark et al. 2007
Ubx_M00018	KDVNVBYTAATKRSBHNK	66	TRANSFAC
Ubx_Down	KDVNVBYTAATKRSBHNK	66	Down et al. 2007
ME20	YTAATGAVS	67	Stark et al. 2007
ME37	WAATKNNNNNCRGCGY	67	Stark et al. 2007
ME40	CATTANNNNWAATT	67	Stark et al. 2007
ME43	YAATKAAGY	67	Stark et al. 2007
ME47	AATTANWNRCGC	67	Stark et al. 2007
ME48	AATTWWNAYGCR	67	Stark et al. 2007
ME72	KCAATAAA	67	Stark et al. 2007
ME98	AATTRNNNNNNCAA	67	Stark et al. 2007
ME108	AATNNNNNCATNR	67	Stark et al. 2007
ME130	WTGGNNNNNTAAY	67	Stark et al. 2007
abd-B_M00090	VSNKTDATKRCNV	67	TRANSFAC
abd-B_Down	VSNKTDATKRCNV	67	Down et al. 2007
E74A_MA0026	CMGGAAR	68	JASPAR
E74A_M00016	HVAWYCMGGAARTDNKS	68	TRANSFAC
Eip74EF_Down	HVAWYCMGGAARTDNKS	68	Down et al. 2007
ME17	MAATTNAATT	69	Stark et al. 2007
ME34	AAATKKCATTA	69	Stark et al. 2007
ME102	TWAWKMNAWTTG	69	Stark et al. 2007
Dfd_M00019	RATTAMY	70	TRANSFAC
Ubx_MA0094	TAAT	70	JASPAR
Ap_Down	TAAT	70	Down et al. 2007
dl_MA0023	GKGGWWTTC	71	JASPAR
Rel_manual	GGRNWTTCC	71	Stark et al. 2007
dif-rel_Down	GGGGAWTCCCY	71	Down et al. 2007
dl_B-Down	KGGGWWWHCYV	71	Down et al. 2007
Rel_Down	GGGGAWYCMC	71	Down et al. 2007
shn-ZFP2_Down	GGGGAMWWCCM	71	Down et al. 2007
dl_M00043	GGGWTTTCCV	71	TRANSFAC
dl_A-Down	GGGWTTTCCV	71	Down et al. 2007
eve_t2-Schroeder	YKNWSMGCVSCRHD	72	Schroeder et al. 2004
dve_M00629	KBWSSYCHKC	72	TRANSFAC
Dip3_Down	YYWVNYWWDNYS	72	Down et al. 2007
pros_Down	CWYBDCY	72	Down et al. 2007
brk_Down	TGGCGYY	72	Down et al. 2007
ME18	MRYTTCCGYY	72	Stark et al. 2007
twi_manual	CAYRTGTG	73	Stark et al. 2007
esg_Down	RRCAGGTGB	73	Down et al. 2007
ME90	ACATGTK	73	Stark et al. 2007

Supplemental Table S5**Functional analysis of genes flanking stage-specific binding sites**

Hypergeometric p-values (FDR-adjusted) for the enrichment of functional categories in genes close to context-specific binding sites. A selection of Gene Ontologies significant in at least one of the condition is shown. Gene Ontology analysis can also be found at: Wilczynski et al. 2010 for BIN, MEF2, TIN and TWI, Verzi et al 2010 for HNF4A, H3K4me2, GATA and CDX2, Pali et al. 2011 for TAL1, Zhong et al. 2010 for PHA-4. diffCell and prolCell correspond to the conditions: differentiating intestinal epithelial cells and proliferating intestinal epithelial cells respectively. "class I" correspond to the binding sites of EP300 in embryonic stem cells (hESC) and "class_II-I" to the cell-type-dependent EP300 binding sites found in the neuroectodermal spheres (hNECs) as defined by Rada-Iglesias et al. 2011.

Lin et al. 2010**Functional Category**

	TCF3 pre-pBcells	TCF3 pBcells
lymphocyte_activation	0.1211	0.0002
cell_activation	0.1957	0.0021
B_cell_activation	0.6873	0.0025
positive_regulation_of_molecular_function	0.1237	0.0036
leukocyte_activation	0.0438	0.0001
cellular_response_to_stimulus	0.0149	0.0021
signal_transduction	0.0363	0.0021
regulation_of_cellular_process	0.0135	0.0054
regulation_of_molecular_function	0.0147	0.0054

Verzi et al. 2010**Functional Category**

	HNF4A diffCell	HNF4A prolCell
positive_regulation_of_macromolecule_metabolic_process	0.175	0.000
positive_regulation_of_metabolic_process	0.161	0.000
positive_regulation_of_macromolecule_biosynthetic_process	0.114	0.000
positive_regulation_of_cellular_metabolic_process	0.104	0.000
positive_regulation_of_biosynthetic_process	0.079	0.000
positive_regulation_of_nitrogen_compound_metabolic_process	0.070	0.000
positive_regulation_of_nucleobase-containing_compound_metabolic_process	0.056	0.000
positive_regulation_of_cellular_biosynthetic_process	0.066	0.001
regulation_of_cell_morphogenesis_involved_in_differentiation	0.873	0.001
positive_regulation_of_transcription-DNA-dependent	0.147	0.001
regulation_of_cell_proliferation	0.079	0.001
locomotion	0.964	0.001
regulation_of_cell_development	0.130	0.001
regulation_of_metabolic_process	0.365	0.001
proteoglycan_metabolic_process	0.061	0.001
regulation_of_macromolecule_metabolic_process	0.364	0.001
positive_regulation_of_transcription_from_RNA_polymerase_II_promoter	0.263	0.001
positive_regulation_of_gene_expression	0.301	0.002
regulation_of_phosphorus_metabolic_process	0.520	0.002
positive_regulation_of_biological_process	0.594	0.002
regulation_of_phosphate_metabolic_process	0.520	0.002
regulation_of_molecular_function	0.557	0.002
regulation_of_protein_metabolic_process	0.964	0.002
regulation_of_anatomical_structure_morphogenesis	0.029	0.000
regulation_of_multicellular_organismal_development	0.013	0.000
regulation_of_cell_differentiation	0.008	0.000
anatomical_structure_morphogenesis	0.001	0.000
negative_regulation_of_biological_process	0.001	0.000
regulation_of_developmental_process	0.005	0.000
regulation_of_multicellular_organismal_process	0.036	0.000
negative_regulation_of_cellular_process	0.001	0.001
biological_regulation	0.030	0.002
regulation_of_signaling	0.007	0.002
regulation_of_biological_process	0.006	0.002
regulation_of_signal_transduction	0.002	0.002
system_development	0.049	0.003
regulation_of_cellular_process	0.007	0.004
regulation_of_response_to_stimulus	0.044	0.005
anatomical_structure_development	0.018	0.006
developmental_process	0.021	0.008
organ_development	0.001	0.012
negative_regulation_of_signaling	0.001	0.634
epithelial_tube_formation	0.001	0.540
embryonic_epithelial_tube_formation	0.001	0.540

morphogenesis_of_an_epithelium	0.001	1.000
negative_regulation_of_cell_communication	0.001	0.634
sympathetic_nervous_system_development	0.001	0.471
tissue_morphogenesis	0.001	0.525
negative_regulation_of_cell_death	0.001	0.432
tube_morphogenesis	0.001	0.092
negative_regulation_of_signal_transduction	0.002	0.614
negative_regulation_of_response_to_stimulus	0.002	0.193
glial_cell_development	0.003	1.000
core_promoter_sequence-specific_DNA_binding	0.003	1.000
negative_regulation_of_apoptotic_process	0.005	0.505
negative_regulation_of_programmed_cell_death	0.006	0.471
regulation_of_epithelial_cell_differentiation	0.007	0.634
gland_development	0.007	0.084
RNA_polymerase_II_core_promoter_proximal_region_sequence-specific_DNA_binding_t	0.008	0.057

Functional Category

	GATA6 diffCell	GATA6 prolCell
positive_regulation_of_RNA_metabolic_process	1.000	0.001
regulation_of_cell_differentiation	1.000	0.001
regulation_of_anatomical_structure_morphogenesis	0.921	0.004
positive_regulation_of_gene_expression	1.000	0.004
regulation_of_multicellular_organismal_development	1.000	0.004
anatomical_structure_morphogenesis	0.910	0.004
regulation_of_locomotion	0.910	0.005
positive_regulation_of_transcription-DNA-dependent	0.921	0.006
negative_regulation_of_developmental_process	0.921	0.009
regulation_of_developmental_process	0.921	0.009

Functional Category

	CDX2 diffCell	CDX2 prolCell
positive_regulation_of_cellular_metabolic_process	0.050	0.000
positive_regulation_of_macromolecule_metabolic_process	0.058	0.001
positive_regulation_of_macromolecule_biosynthetic_process	0.123	0.001
positive_regulation_of_cellular_biosynthetic_process	0.240	0.001
regulation_of_transcription_from_RNA_polymerase_II_promoter	0.509	0.001
positive_regulation_of_transcription-DNA-dependent	0.359	0.001
positive_regulation_of_biosynthetic_process	0.224	0.001
positive_regulation_of_nitrogen_compound_metabolic_process	0.263	0.002
positive_regulation_of_RNA_metabolic_process	0.425	0.002
positive_regulation_of_gene_expression	0.273	0.003
regulation_of_macromolecule_metabolic_process	0.080	0.003
regulation_of_primary_metabolic_process	0.158	0.004
positive_regulation_of_nucleobase-containing_compound_metabolic_process	0.281	0.004
regulation_of_metabolic_process	0.068	0.005
cellular_component_organization_at_cellular_level	0.246	0.005
cell_projection	0.630	0.005
morphogenesis_of_an_epithelium	0.086	0.005
positive_regulation_of_protein_metabolic_process	0.318	0.006
regulation_of_cellular_metabolic_process	0.076	0.006
cellular_component_organization_or_biogenesis_at_cellular_level	0.238	0.007
cell_projection_organization	0.949	0.008
tissue_morphogenesis	0.107	0.008
regulation_of_locomotion	0.155	0.008
cytoplasm	0.805	0.009
positive_regulation_of_cell_proliferation	0.540	0.010
anatomical_structure_morphogenesis	0.000	0.000
multicellular_organismal_process	0.023	0.000
developmental_process	0.013	0.000
regulation_of_cell_proliferation	0.039	0.000
cellular_developmental_process	0.032	0.000
protein_binding	0.030	0.000
negative_regulation_of_biological_process	0.000	0.000
regulation_of_cellular_process	0.001	0.001
positive_regulation_of_metabolic_process	0.030	0.001
negative_regulation_of_cellular_process	0.000	0.001
regulation_of_biological_process	0.002	0.001
cellular_component_organization	0.016	0.001
positive_regulation_of_biological_process	0.013	0.001
positive_regulation_of_cellular_process	0.018	0.001
regulation_of_developmental_process	0.013	0.001
cellular_component_organization_or_biogenesis	0.016	0.001
regulation_of_protein_metabolic_process	0.000	0.001
embryonic_morphogenesis	0.001	0.001
biological_regulation	0.001	0.001
anatomical_structure_development	0.009	0.001
regulation_of_cellular_protein_metabolic_process	0.000	0.002
locomotion	0.012	0.005
regulation_of_cell_communication	0.016	0.007

regulation_of_multicellular_organismal_process	0.039	0.007
organ_morphogenesis	0.000	0.009
regulation_of_signaling	0.000	0.011
positive_regulation_of_molecular_function	0.004	0.013
regulation_of_phosphorylation	0.001	0.024
regulation_of_protein_phosphorylation	0.001	0.028
regulation_of_protein_modification_process	0.001	0.029
regulation_of_phosphorus_metabolic_process	0.000	0.030
regulation_of_phosphate_metabolic_process	0.000	0.030
regulation_of_molecular_function	0.004	0.053
cellular_process	0.000	0.059
positive_regulation_of_catalytic_activity	0.004	0.069
regulation_of_response_to_stimulus	0.001	0.077
cell_activation	0.000	0.101
regulation_of_signal_transduction	0.000	0.118
negative_regulation_of_signaling	0.003	0.199
regulation_of_transferase_activity	0.002	0.201
regulation_of_catalytic_activity	0.003	0.201
negative_regulation_of_cell_communication	0.003	0.206
negative_regulation_of_response_to_stimulus	0.004	0.227
positive_regulation_of_transferase_activity	0.002	0.230
response_to_organic_substance	0.006	0.241
response_to_chemical_stimulus	0.000	0.274
sequence-specific_DNA_binding_RNA_polymerase_II_transcription_factor_activity	0.005	0.325
regulation_of_kinase_activity	0.001	0.356
tube_formation	0.010	0.417
positive_regulation_of_kinase_activity	0.001	0.465
regulation_of_protein_kinase_activity	0.001	0.572
positive_regulation_of_protein_kinase_activity	0.001	0.620
phosphatidylinositol_3-kinase_cascade	0.000	0.629
response_to_endogenous_stimulus	0.004	0.750
response_to_hormone_stimulus	0.001	0.808
regulation_of_biological_quality	0.004	0.923
stress_fiber_assembly	0.004	1.000
organelle_lumen	0.005	1.000
epithelial_tube_formation	0.003	1.000
embryonic_epithelial_tube_formation	0.003	1.000

Visel et al. 2009**Functional Category**

	EP300 Forebrain	EP300 Limb
developmental_process	0.000	0.087
anatomical_structure_development	0.000	0.229
cellular_developmental_process	0.000	0.872
DNA_binding	0.000	0.093
regulation_of_transcription-_DNA-dependent	0.000	0.248
regulation_of_RNA_biosynthetic_process	0.000	0.248
positive_regulation_of_transcription-_DNA-dependent	0.000	0.238
regulation_of_RNA_metabolic_process	0.000	0.292
sequence-specific_DNA_binding	0.000	0.238
multicellular_organismal_process	0.000	1.000
cell_differentiation	0.000	0.571
regulation_of_gene_expression	0.000	0.238
positive_regulation_of_RNA_metabolic_process	0.000	0.238
system_development	0.000	0.338
regulation_of_multicellular_organismal_process	0.000	1.000
negative_regulation_of_biological_process	0.000	1.000
regulation_of_transcription_from_RNA_polymerase_II_promoter	0.000	0.262
positive_regulation_of_gene_expression	0.000	0.154
regulation_of_cellular_macromolecule_biosynthetic_process	0.000	0.238
regulation_of_macromolecule_biosynthetic_process	0.000	0.238
pattern_specification_process	0.000	0.093
olfactory_receptor_activity	0.000	1.000
nervous_system_development	0.000	1.000
regulation_of_neuron_differentiation	0.000	0.087
organ_development	0.000	0.460
regulation_of_neurogenesis	0.000	0.087
locomotion	0.000	1.000
positive_regulation_of_neuron_differentiation	0.000	0.616
axon_guidance	0.000	0.872
behavior	0.000	1.000
response_to_external_stimulus	0.000	1.000
brain_development	0.003	1.000
cell_migration	0.004	1.000
negative_regulation_of_cell_communication	0.004	1.000
negative_regulation_of_response_to_stimulus	0.004	1.000
neuron_differentiation	0.004	0.460

forebrain_development	0.005	0.468
neuron_fate_commitment	0.009	1.000

Rada-Iglesias et al. 2010

Functional Category	class I	class II-I
neuron_differentiation	0.123	0.002
beta-catenin_binding	0.102	0.004
positive_regulation_of_cell_development	0.358	0.006
regulation_of_nervous_system_development	0.000	0.002
anatomical_structure_morphogenesis	0.000	0.002
anatomical_structure_development	0.000	0.002
organ_development	0.000	0.002
regulation_of_multicellular_organismal_development	0.000	0.002
canonical_Wnt_receptor_signaling_pathway	0.002	0.002
cell_differentiation	0.000	0.002
regulation_of_cell_development	0.000	0.002
negative_regulation_of_developmental_process	0.000	0.002
embryonic_morphogenesis	0.000	0.002
regulation_of_Wnt_receptor_signaling_pathway	0.000	0.002
regulation_of_developmental_process	0.000	0.002
cellular_developmental_process	0.000	0.002
regulation_of_cell_differentiation	0.000	0.004
negative_regulation_of_cell_differentiation	0.000	0.008
developmental_process	0.000	0.009
regulation_of_multicellular_organismal_process	0.000	0.012
regulation_of_signal_transduction	0.000	0.016
sequence-specific_DNA_binding	0.000	0.023
pattern_specification_process	0.001	0.026
regulation_of_signaling	0.000	0.027
regulation_of_neurogenesis	0.000	0.027
positive_regulation_of_cell_differentiation	0.008	0.027
regulation_of_response_to_stimulus	0.000	0.032
system_development	0.000	0.039
positive_regulation_of_gene_expression	0.001	0.039
regulation_of_neuron_differentiation	0.000	0.039
positive_regulation_of_cellular_process	0.000	0.042
negative_regulation_of_biological_process	0.000	0.320
negative_regulation_of_cellular_process	0.000	0.266
regulation_of_biological_process	0.000	0.940
regulation_of_cellular_process	0.000	1.000
regulation_of_cell_migration	0.001	1.000
biological_regulation	0.000	1.000
response_to_external_stimulus	0.000	0.811
positive_regulation_of_gene_expression	0.001	0.824
response_to_chemical_stimulus	0.000	1.000
regulation_of_anatomical_structure_morphogenesis	0.000	0.320
positive_regulation_of_biological_process	0.000	0.070
negative_regulation_of_macromolecule_metabolic_process	0.000	0.334
signal_transduction	0.000	0.967
multicellular_organismal_process	0.000	0.159
regulation_of_cellular_component_organization	0.000	0.993
negative_regulation_of_metabolic_process	0.000	0.321
negative_regulation_of_biosynthetic_process	0.000	0.411
olfactory_receptor_activity	0.000	1.000
negative_regulation_of_macromolecule_biosynthetic_process	0.000	0.508
negative_regulation_of_cellular_metabolic_process	0.000	0.385
negative_regulation_of_cellular_biosynthetic_process	0.000	0.408
cell_surface_receptor_linked_signaling_pathway	0.000	0.196
cell_junction	0.000	0.582

Zhong et al. 2010

Functional Category	PHA-4 embryo	PHA-4 L1
binding	0.000	0.727
developmental_process	0.000	1.000
larval_development	0.000	1.000
anatomical_structure_morphogenesis	0.000	1.000
nematode_larval_development	0.000	1.000
post-embryonic_development	0.000	1.000
intracellular_part	0.000	0.460
developmental_process_involved_in_reproduction	0.000	1.000
embryo_development	0.000	1.000
cellular_component_organization_at_cellular_level	0.000	1.000
embryo_development_ending_in_birth_or_egg_hatching	0.000	1.000
cellular_component_organization_or_biogenesis_at_cellular_level	0.000	1.000

organelle_organization	0.000	1.000
reproduction	0.000	1.000
protein_binding	0.000	0.963
organ_development	0.000	1.000
tissue_morphogenesis	0.000	1.000
locomotion	0.000	1.000
cellular_component_organization	0.000	1.000
cellular_component_organization_or_biogenesis	0.000	1.000
regulation_of_biological_process	0.000	1.000
anatomical_structure_development	0.000	1.000
morphogenesis_of_an_epithelium	0.000	1.000
cell_migration	0.001	1.000
dauer_entry	0.001	1.000
post-embryonic_organ_development	0.001	1.000
Wnt_receptor_signaling_pathway	0.001	1.000
body_morphogenesis	0.003	1.000
axon	0.003	1.000
cell_cycle_process	0.005	1.000
dormancy_process	0.007	1.000

Supplemental References

- Beer MA, and Tavazoie S. 2004. Predicting gene expression from sequence. *Cell* **117**: 185–198.
- Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, and Eisen MB. 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc Natl Acad Sci USA* **99**: 757–762.
- He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, and Zeitlinger J. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* **43**: 414–420.
- Joachims T. 1999. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.). *MIT-Press*.
- Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, et al. 2010. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* **11**: 635–643.
- 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**: 2387–2394.
- Ohshiro T, and Saigo K. 1997. Transcriptional regulation of breathless FGF receptor gene by binding of TRACHEALESS/dARNT heterodimers to three central midline elements in *Drosophila* developing trachea. *Development* **124**: 3975–3986.
- Palii CG, Perez-Iratxeta C, Yao Z, Cao Y, Dai F, Davison J, Atkins H, Allan D, Dilworth FJ, Gentleman R, et al. 2011. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J* **30**: 494–509.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, and Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.
- Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, and Furlong EEM. 2007. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes & Development* **21**: 436–449.
- Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, Montgomery RK, Fleet JC, Brown M, Liu XS, and Shivdasani RA. 2010. Differentiation-Specific Histone Modifications Reveal Dynamic Chromatin Interactions and Partners for the Intestinal Transcription Factor CDX2. *Developmental Cell* **19**: 713–726.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–858.
- Yuan Y, Guo L, Shen L, and Liu JS. 2007. Predicting gene expression from sequence: a

reexamination. *PLoS Comp Biol* **3**: e243.

Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, and Levine M. 2007. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes & Development* **21**: 385–390.

Zhong M, Niu W, Lu ZJ, Sarov M, Murray JI, Janette J, Raha D, Sheaffer KL, Lam HYK, Preston E, et al. 2010. Genome-wide identification of binding sites defines distinct functions for *Caenorhabditis elegans* PHA-4/FOXA in development and environmental response. *PLoS Genet* **6**: e1000848.

Zinzen RP, Girardot C, Gagneur J, Braun M, and Furlong EEM. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65–70.