Developmental Cell, Volume *58*

Supplemental information

Chromatin accessibility in the *Drosophila* embryo is determined by transcription factor pioneering and enhancer activation

Kaelan J. Brennan, Melanie Weilert, Sabrina Krueger, Anusri Pampari, Hsiao-yun Liu, Ally W.H. Yang, Jason A. Morrison, Timothy R. Hughes, Christine A. Rushlow, Anshul Kundaje, and Julia Zeitlinger

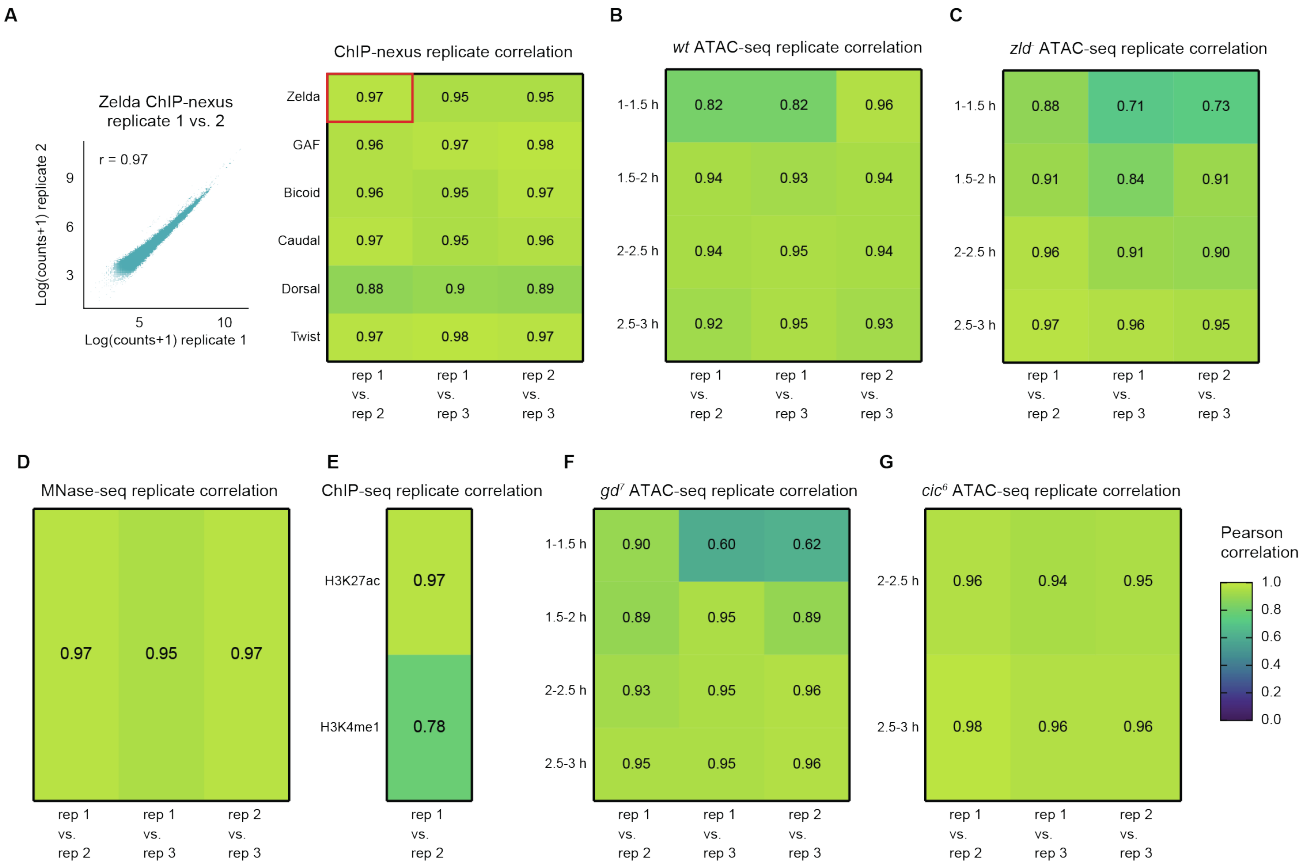
A
Zelda ChIP-nexus replicate 1 vs. 2
r = 0.97
Log(counts+1) replicate 2
Log(counts+1) replicate 1
ChIP-nexus replicate correlation
Zelda
GAF
Bicoid
Caudal
Dorsal
Twist
0.97 0.95 0.95
0.96 0.97 0.98
0.96 0.95 0.97
0.97 0.95 0.96
0.88 0.9 0.89
0.97 0.98 0.97
rep 1 vs. rep 2
rep 1 vs. rep 3
rep 2 vs. rep 3
B
wt ATAC-seq replicate correlation
1-1.5 h
1.5-2 h
2-2.5 h
2.5-3 h
0.82 0.82 0.96
0.94 0.93 0.94
0.94 0.95 0.94
0.92 0.95 0.93
C
zld ATAC-seq replicate correlation
0.88 0.71 0.73
0.91 0.84 0.91
0.96 0.91 0.90
0.97 0.96 0.95
D
MNase-seq replicate correlation
0.97 0.95 0.97
E
ChIP-seq replicate correlation
H3K27ac
H3K4me1
0.97
0.78
F
gd7 ATAC-seq replicate correlation
0.90 0.60 0.62
0.89 0.95 0.89
0.93 0.95 0.96
0.95 0.95 0.96
G
cic6 ATAC-seq replicate correlation
0.96 0.94 0.95
0.98 0.96 0.96
Pearson correlation
1.0
0.8
0.6
0.4
0.2
0.0

Figure S1 | ChIP-nexus, ATAC-seq, MNase-seq, and ChIP-seq replicates are highly correlated (related to Figures 1-5).
Pearson correlation values were determined for all pairwise comparisons between replicates of (**A**) TF ChIP-nexus experiments, (**B**) *wt* time course ATAC-seq experiments, (**C**) *zld*$^{-}$ time course ATAC-seq experiments, (**D**) MNase-seq experiments, (**E**) histone ChIP-seq experiments, (**F**) *gd*^{7} time course ATAC-seq experiments, and (**G**) *cic*^{6} time course ATAC-seq experiments. The Zelda ChIP-nexus replicates 1 and 2 correlation plot is shown as an example of how Pearson correlation values were derived. For ChIP-nexus, coverage for each replicate was calculated across a 400 bp window centered on the MACS2-called peaks for each TF. Because ChIP-nexus provides strand-specific information, the absolute value of the counts from the negative strand, which would otherwise be negative, was taken and added to the counts across the positive strand to determine the total region counts for a given replicate. For ATAC-seq, counts for each replicate were calculated across a 400 bp window centered on the MACS2-called peaks for each time point. For ChIP-seq, counts for each replicate were calculated across a 1000 bp window centered on the MACS2-called peaks for each histone mark. For MNase-seq, counts for each replicate were calculated across a 1000 bp window centered on *Drosophila* transcription start sites.

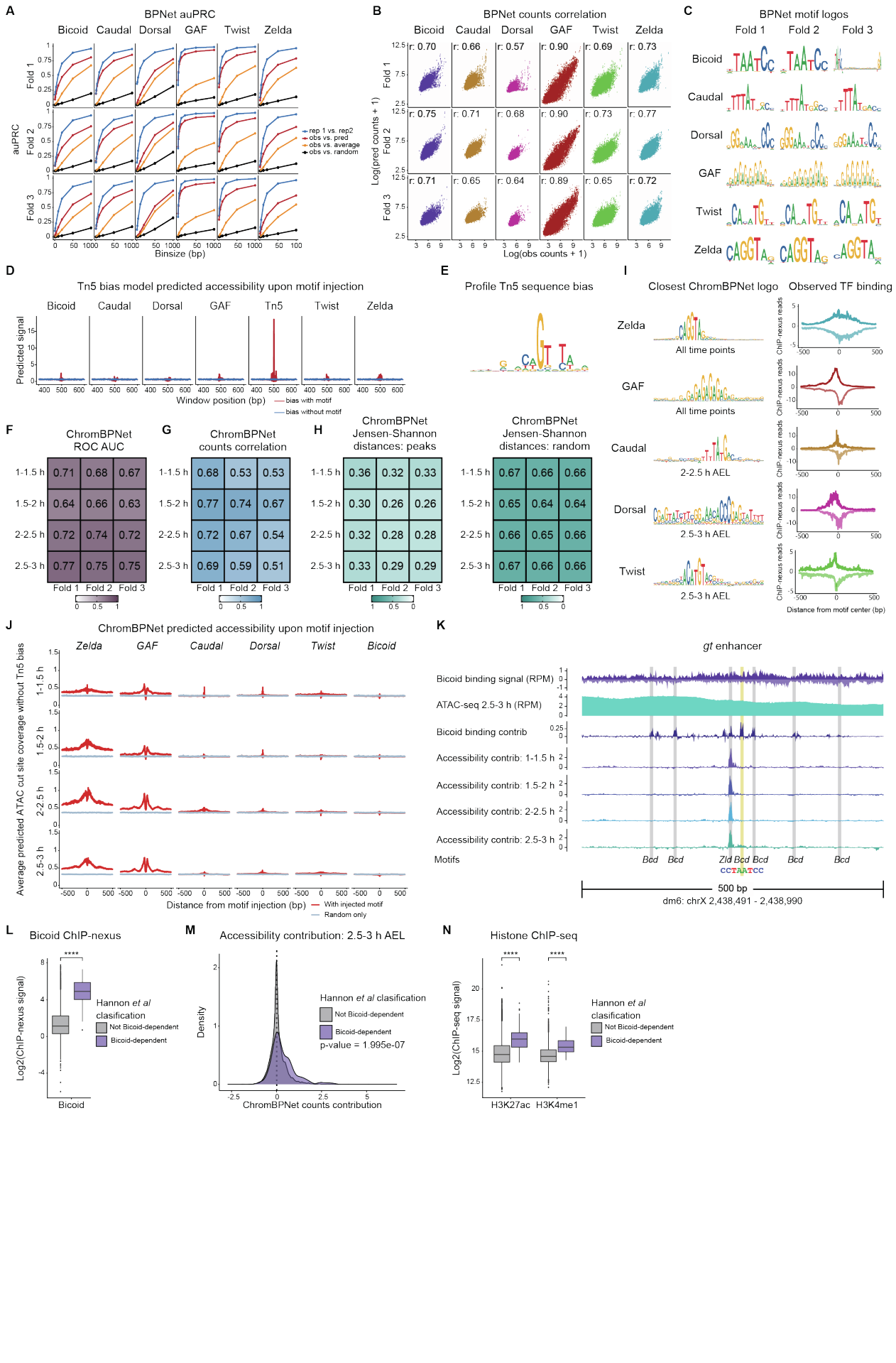
A
BPNet auPRC
Bicoid Caudal Dorsal GAF Twist Zelda
Fold 1
Fold 2
Fold 3
auPRC
Binsize (bp)
rep 1 vs. rep2
obs vs. pred
obs vs. average
obs vs. random
B
BPNet counts correlation
Bicoid Caudal Dorsal GAF Twist Zelda
r: 0.70 r: 0.66 r: 0.57 r: 0.90 r: 0.69 r: 0.73
r: 0.75 r: 0.71 r: 0.68 r: 0.90 r: 0.73 r: 0.77
r: 0.71 r: 0.65 r: 0.64 r: 0.89 r: 0.65 r: 0.72
Log(pred counts + 1)
Log(obs counts + 1)
C
BPNet motif logos
Fold 1 Fold 2 Fold 3
Bicoid
Caudal
Dorsal
GAF
Twist
Zelda
D
Tn5 bias model predicted accessibility upon motif injection
Bicoid Caudal Dorsal GAF Tn5 Twist Zelda
Predicted signal
Window position (bp)
bias with motif
bias without motif
E
Profile Tn5 sequence bias
I
Closest ChromBPNet logo
Observed TF binding
Zelda
All time points
GAF
All time points
Caudal
2-2.5 h AEL
Dorsal
2.5-3 h AEL
Twist
2.5-3 h AEL
ChIP-nexus reads
Distance from motif center (bp)
F
ChromBPNet ROC AUC
1-1.5 h 0.71 0.68 0.67
1.5-2 h 0.64 0.66 0.63
2-2.5 h 0.72 0.74 0.72
2.5-3 h 0.77 0.75 0.75
Fold 1 Fold 2 Fold 3
G
ChromBPNet counts correlation
1-1.5 h 0.68 0.53 0.53
1.5-2 h 0.77 0.74 0.67
2-2.5 h 0.72 0.67 0.54
2.5-3 h 0.69 0.59 0.51
Fold 1 Fold 2 Fold 3
H
ChromBPNet Jensen-Shannon distances: peaks
1-1.5 h 0.36 0.32 0.33
1.5-2 h 0.30 0.26 0.26
2-2.5 h 0.32 0.28 0.28
2.5-3 h 0.33 0.29 0.29
Fold 1 Fold 2 Fold 3
ChromBPNet Jensen-Shannon distances: random
1-1.5 h 0.67 0.66 0.66
1.5-2 h 0.65 0.64 0.64
2-2.5 h 0.66 0.65 0.66
2.5-3 h 0.67 0.66 0.66
Fold 1 Fold 2 Fold 3
J
ChromBPNet predicted accessibility upon motif injection
Zelda GAF Caudal Dorsal Twist Bicoid
Average predicted ATAC cut site coverage without Tn5 bias
1-1.5 h
1.5-2 h
2-2.5 h
2.5-3 h
Distance from motif injection (bp)
With injected motif
Random only
K
gt enhancer
Bicoid binding signal (RPM)
ATAC-seq 2.5-3 h (RPM)
Bicoid binding contrib
Accessibility contrib: 1-1.5 h
Accessibility contrib: 1.5-2 h
Accessibility contrib: 2-2.5 h
Accessibility contrib: 2.5-3 h
Motifs
Bcd Bcd Zld Bcd Bcd Bcd Bcd
CCTAATCC
500 bp
dm6: chrX 2,438,491 - 2,438,990
L
Bicoid ChIP-nexus
Log2(ChIP-nexus signal)
Hannon et al clasification
Not Bicoid-dependent
Bicoid-dependent
Bicoid
M
Accessibility contribution: 2.5-3 h AEL
Density
Hannon et al clasification
Not Bicoid-dependent
Bicoid-dependent
p-value = 1.995e-07
ChromBPNet counts contribution
N
Histone ChIP-seq
Log2(ChIP-seq signal)
Hannon et al clasification
Not Bicoid-dependent
Bicoid-dependent
H3K27ac H3K4me1

Figure S2 | Deep learning models accurately learn TF ChIP-nexus and time course chromatin accessibility information in the early *Drosophila* embryo (related to Figures 1 and 2).
(**A**) Area under the Precision-Recall Curves (auPRC) show that BPNet predicts the profile positions with high accuracy. The ability of BPNet to identify positions of high ChIP-nexus signal is assessed at various resolutions up to 100 bp. Replicate experiments, average ChIP-nexus profiles, and randomized profiles are shown as controls. Three-fold validation was performed by applying the same model architecture from the original, optimized model (fold 1) to two additional models (fold 2 and fold 3) with the training, validation, and test chromosomes shuffled. These results show that the training regions are representative of the entire dataset and that the trained BPNet model is highly stable. (**B**) BPNet predicts ChIP-nexus counts with high accuracy. Pearson counts correlation values were determined by comparing the observed ChIP-nexus counts with BPNet's predicted counts at ChIP-nexus peaks for each of the TFs of interest. The stability of BPNet's counts predictions were assessed with three-fold validation. (**C**) BPNet re-discovered the known motifs for all TFs of interest irrespective of the distribution of the training, validation, and test chromosomes. BPNet CWMs are shown for each TF for the original optimized model (fold 1) and the additional models trained with the same architecture as part of three-fold validation (fold 2 and fold 3). (**D**) The Tn5 bias ChromBPNet model does not learn TF motif sequence grammar. The canonical sequences for each TF of interest were injected into 256 genomic sequences from ChromBPNet's test chromosome (chrX) and the Tn5 bias model was used to predict chromatin accessibility cut site signal. The effects were averaged across trials and show no predicted accessibility upon injection of any motif except the Tn5 preferred sequence. This confirms that the bias model's learning was limited to Tn5 bias and did not learn cis-regulatory grammar. (**E**) TF-MoDISco interpretations returned Tn5 sequence bias as the top Tn5 logo for profile contribution but not for counts contribution. These results show that the bias model only learned Tn5 positional information and was successfully trained to only represent Tn5 bias at closed genomic regions. (**F**) The time course ChromBPNet models accurately discriminates between ATAC-seq peak and non-peak regions. The models' predictions were assessed using area under the receiver operating characteristic curves (ROC AUC). Three-fold validation was performed by applying the original ChromBPNet architecture (fold 1) to two additional models with reshuffled training, test, and validation chromosome sets (fold 2 and fold 3). (**G**) The time course ChromBPNet models accurately predict chromatin accessibility counts. Pearson correlation values were calculated by comparing the observed ATAC-seq cut sites with the ChromBPNet predicted cut sites at ATAC-seq peak regions for all time points. (**H**) The time course ChromBPNet models have high profile prediction accuracy. Time course profile predictions were assessed by comparing to the observed ATAC-seq cut sites using Jensen-Shannon distances at peak regions, where lower values are better. Randomly shuffled region profiles were included as a control. (**I**) ChromBPNet identifies TF motifs in ATAC-seq data that are bound by their respective TFs. TF-MoDISco was run on all ChromBPNet models and sequence features with high counts contribution were consolidated into motifs. The closest motif logo for each TF of interest was manually identified with the exception of Bicoid, which ChromBPNet did not identify in the ATAC-seq data. Motifs for the pioneering TFs were unambiguous and identified at all time points, while the patterning TF motifs deviated from the BPNet-identified binding motifs for Caudal, Dorsal, and Twist and were identified only at later time points. Average observed ChIP-nexus binding profiles showed clear TF footprints on all motifs. Average footprints were anchored on and calculated across the accessibility-identified motifs for each TF. (**J**) ChromBPNet predicts time course chromatin accessibility in response to TF motif injection. The TF binding motifs identified by BPNet were injected into 512 randomized sequences. The ChromBPNet models were used to make chromatin accessibility cut site predictions before (blue) and after (red) TF motif injection, with the predicted effect centered on the injected motif. The predictions were averaged for all trials and show that pioneering motifs

have the largest predicted effect on chromatin accessibility. (**K**) The Bicoid motif contribution to chromatin accessibility increases over developmental time. Bicoid ChIP-nexus coverage (top track), 2.5-3 h ATAC-seq fragment coverage (second track), BPNet predicted counts contribution for Bicoid binding (third track), and the normalized ChromBPNet counts contribution for each time point (last four tracks) were plotted across the *gt* enhancer[1]. The central Bicoid motif (yellow) contributes to chromatin accessibility and does so most strongly at the latest time point. (**L**) Bicoid motifs that depend on Bicoid for chromatin accessibility are more bound by Bicoid. BPNet-mapped Bicoid motifs were overlapped with genomic regions previously described as dependent on Bicoid for chromatin accessibility[2] and categorized as Bicoid-dependent (n = 162) or Bicoid-independent (n = 5643). Bicoid ChIP-nexus signal was calculated in a 100 bp window surrounding each Bicoid motif instance for both Bicoid-dependent and Bicoid-independent motifs. The Wilcoxon rank-sum test was used to determine statistical significance between Bicoid-dependent and Bicoid-independent motifs (**** = $p < 0.0001$). (**M**) Bicoid motifs that depend on Bicoid for chromatin accessibility have higher accessibility contributions. The ChromBPNet counts contribution from the 2.5-3 h ATAC-seq time point was plotted for the two populations of Bicoid motifs, and the Kolmogorov-Smirnov test was performed and demonstrated a statistically significant ($p = 1.99e-07$) difference in accessibility contribution between Bicoid-dependent and Bicoid-independent motifs. (**N**) Bicoid motifs that depend on Bicoid for chromatin accessibility are more associated with features of enhancer activation. H3K27ac and H3K4me1 ChIP-seq signal was calculated in a 500 bp window surrounding each Bicoid motif instance for both Bicoid-dependent and Bicoid-independent motifs. The Wilcoxon rank-sum test demonstrated statistical significance between Bicoid-dependent and Bicoid-independent motifs for both histone modifications (**** = $p < 0.0001$), further suggesting that Bicoid's contribution to chromatin accessibility comes from enhancer activation.

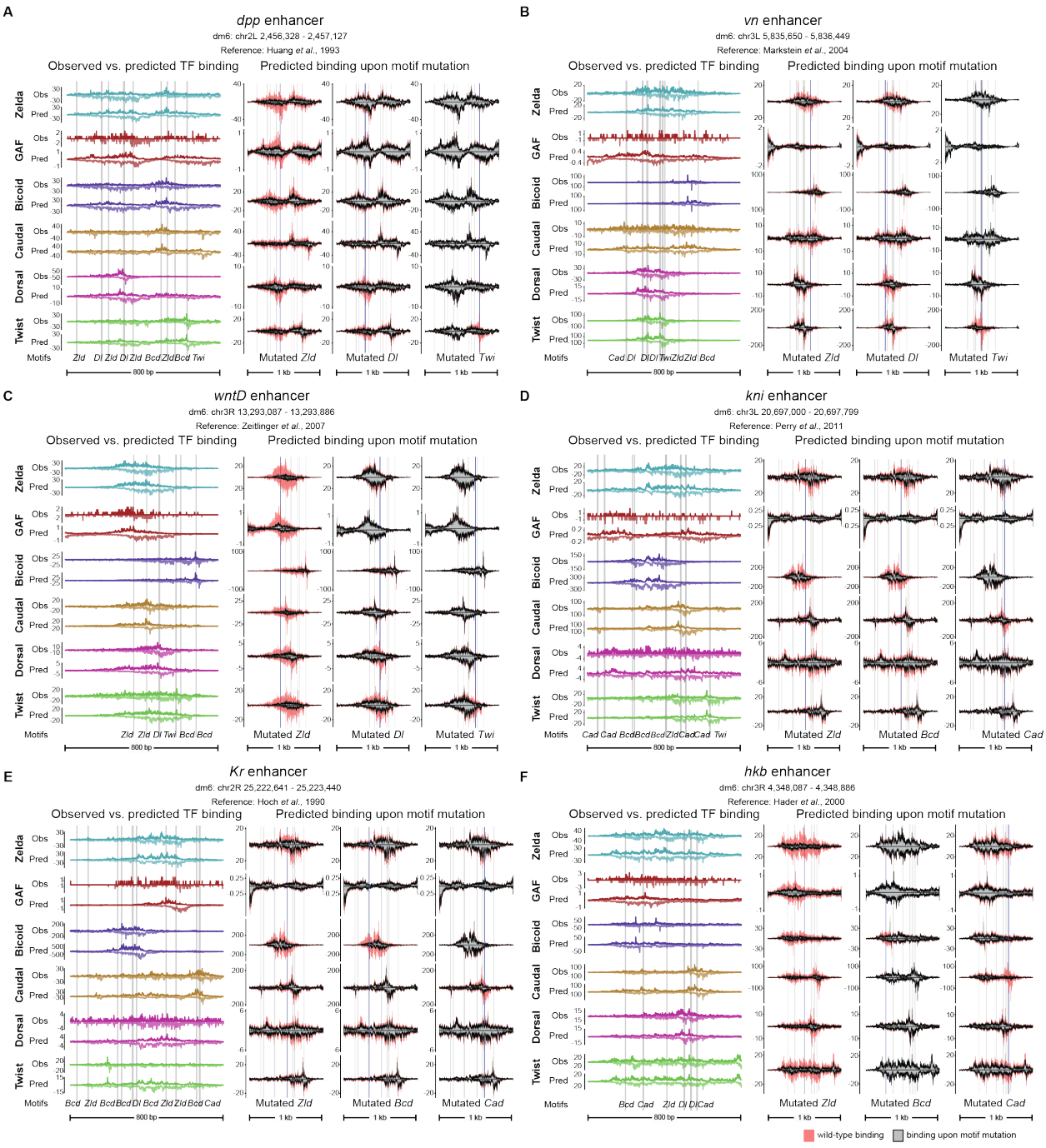
A
dpp enhancer
dm6: chr2L 2,456,328 - 2,457,127
Reference: Huang et al., 1993
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Zelda
GAF
Bicoid
Caudal
Dorsal
Twist
Obs
Pred
Motifs
Zld Dl Zld Dl Zld Bcd Zld Bcd Twi
800 bp
Mutated Zld
Mutated Dl
Mutated Twi
1 kb
B
vn enhancer
dm6: chr3L 5,835,650 - 5,836,449
Reference: Markstein et al., 2004
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Motifs
Cad Dl Dl Dl Twi Zld Zld Bcd
800 bp
Mutated Zld
Mutated Dl
Mutated Twi
1 kb
C
wntD enhancer
dm6: chr3R 13,293,087 - 13,293,886
Reference: Zeitlinger et al., 2007
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Motifs
Zld Zld Dl Twi Bcd Bcd
800 bp
Mutated Zld
Mutated Dl
Mutated Twi
1 kb
D
kni enhancer
dm6: chr3L 20,697,000 - 20,697,799
Reference: Perry et al., 2011
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Motifs
Cad Cad Bcd Bcd Bcd Zld Cad Cad Twi
800 bp
Mutated Zld
Mutated Bcd
Mutated Cad
1 kb
E
Kr enhancer
dm6: chr2R 25,222,641 - 25,223,440
Reference: Hoch et al., 1990
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Motifs
Bcd Zld Bcd Bcd Dl Bcd Zld Zld Bcd Cad
800 bp
Mutated Zld
Mutated Bcd
Mutated Cad
1 kb
F
hkb enhancer
dm6: chr3R 4,348,087 - 4,348,886
Reference: Hader et al., 2000
Observed vs. predicted TF binding
Predicted binding upon motif mutation
Motifs
Bcd Cad Zld Dl Dl Cad
800 bp
Mutated Zld
Mutated Bcd
Mutated Cad
1 kb
wild-type binding
binding upon motif mutation

Figure S3 | BPNet accurately maps TF motifs and predicts TF binding upon motif mutation at known *Drosophila* enhancers (related to Figure 1).
As in Figure 1D, the experimentally generated ChIP-nexus (top track) and BPNet predicted ChIP-nexus data (bottom track) for each TF (different colors) are plotted at known enhancers for the following genes: (**A**) *dpp*[3], (**B**) *vn*[4], (**C**) *wntD*[5], (**D**) *kni*[6], (**E**) *Kr*[7], and (**F**) *hkb*[8]. Motifs were discovered and mapped by BPNet. Enhancers across different developmental patterns and axes were deliberately selected for showcasing BPNet's predictive accuracy. As in Figure 1G, BPNet predicted the binding of all TFs at *wt* enhancer sequences and again at enhancers upon individual motif mutations for the following enhancers: (**A**) *dpp*, (**B**) *vn*, (**C**) *wntD*, (**D**) *kni*, (**E**) *Kr*, and (**F**) *hkb*. Shaded colors show TF binding across the *wt* enhancer, while the gray-filled profiles represent TF binding in response to the motif mutation. Blue bars indicate the mutated motifs that are highlighted under the predictions, and the gray bars are all other BPNet-mapped motifs across the enhancers.

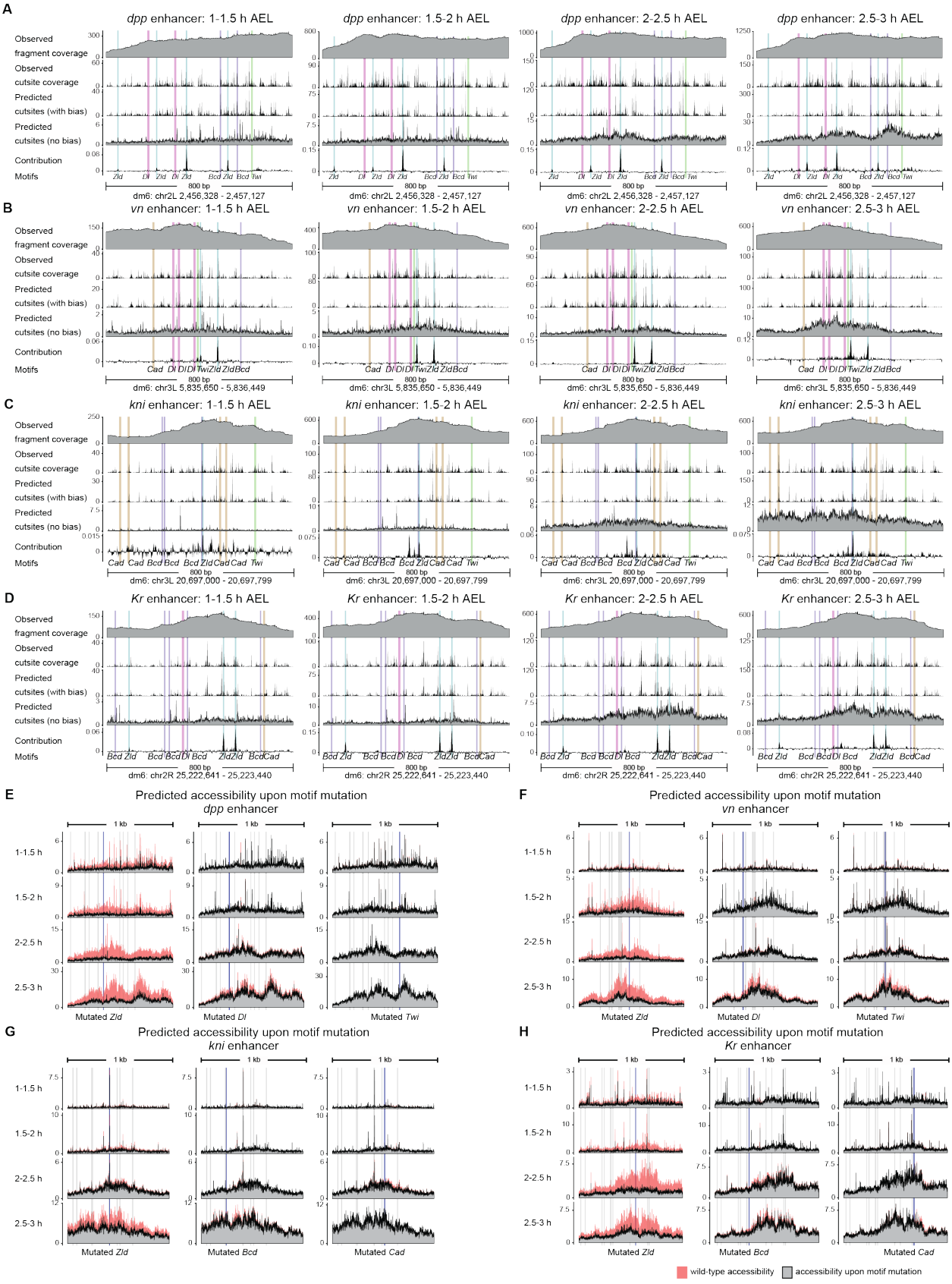
A
dpp enhancer: 1-1.5 h AEL
dpp enhancer: 1.5-2 h AEL
dpp enhancer: 2-2.5 h AEL
dpp enhancer: 2.5-3 h AEL
Observed fragment coverage
Observed cutsite coverage
Predicted cutsites (with bias)
Predicted cutsites (no bias)
Contribution
Motifs
Zld Dl Zld Dl Zld Bcd Zld Bcd Twi
800 bp
dm6: chr2L 2,456,328 - 2,457,127
B
vn enhancer: 1-1.5 h AEL
vn enhancer: 1.5-2 h AEL
vn enhancer: 2-2.5 h AEL
vn enhancer: 2.5-3 h AEL
Cad Dl Dl Dl Twi Zld Zld Bcd
dm6: chr3L 5,835,650 - 5,836,449
C
kni enhancer: 1-1.5 h AEL
kni enhancer: 1.5-2 h AEL
kni enhancer: 2-2.5 h AEL
kni enhancer: 2.5-3 h AEL
Cad Cad Bcd Bcd Bcd Zld Cad Cad Twi
dm6: chr3L 20,697,000 - 20,697,799
D
Kr enhancer: 1-1.5 h AEL
Kr enhancer: 1.5-2 h AEL
Kr enhancer: 2-2.5 h AEL
Kr enhancer: 2.5-3 h AEL
Bcd Zld Bcd Bcd Dl Bcd Zld Zld Bcd Cad
dm6: chr2R 25,222,641 - 25,223,440
E
Predicted accessibility upon motif mutation
dpp enhancer
1 kb
1-1.5 h
1.5-2 h
2-2.5 h
2.5-3 h
Mutated Zld
Mutated Dl
Mutated Twi
F
Predicted accessibility upon motif mutation
vn enhancer
G
Predicted accessibility upon motif mutation
kni enhancer
Mutated Bcd
Mutated Cad
H
Predicted accessibility upon motif mutation
Kr enhancer
wild-type accessibility
accessibility upon motif mutation

Figure S4 | ChromBPNet predicts time course chromatin accessibility at known enhancers and identifies the TF motif contribution to accessibility (related to Figure 2). As in Figure 2C, the experimentally generated ATAC-seq data (tracks one and two) are shown with ChromBPNet accessibility predictions with Tn5 bias (track three) and without Tn5 bias (track four) for known enhancers for the following genes: (**A**) *dpp*, (**B**) *vn*, (**C**) *kni*, and (**D**) *Kr*. Columns represent model predictions at each of the four ATAC-seq time points. The counts contribution for chromatin accessibility across each enhancer is shown as the fifth track, with spikes at BPNet-mapped TF motifs. TF motifs are highlighted with the colored bars. At the same enhancers (**E**) *dpp*, (**F**) *vn*, (**G**) *kni*, and (**H**) *Kr*, ChromBPNet predicted chromatin accessibility cut sites without Tn5 bias for all time points at *wt* enhancer sequences and when individual motifs are mutated, as in Figure 2D. Shaded colors show chromatin accessibility predictions across the *wt* enhancer sequence, and the gray-filled profiles represent chromatin accessibility upon mutation of the highlighted motif (blue bar). Gray bars are non-mutated motifs across each enhancer.

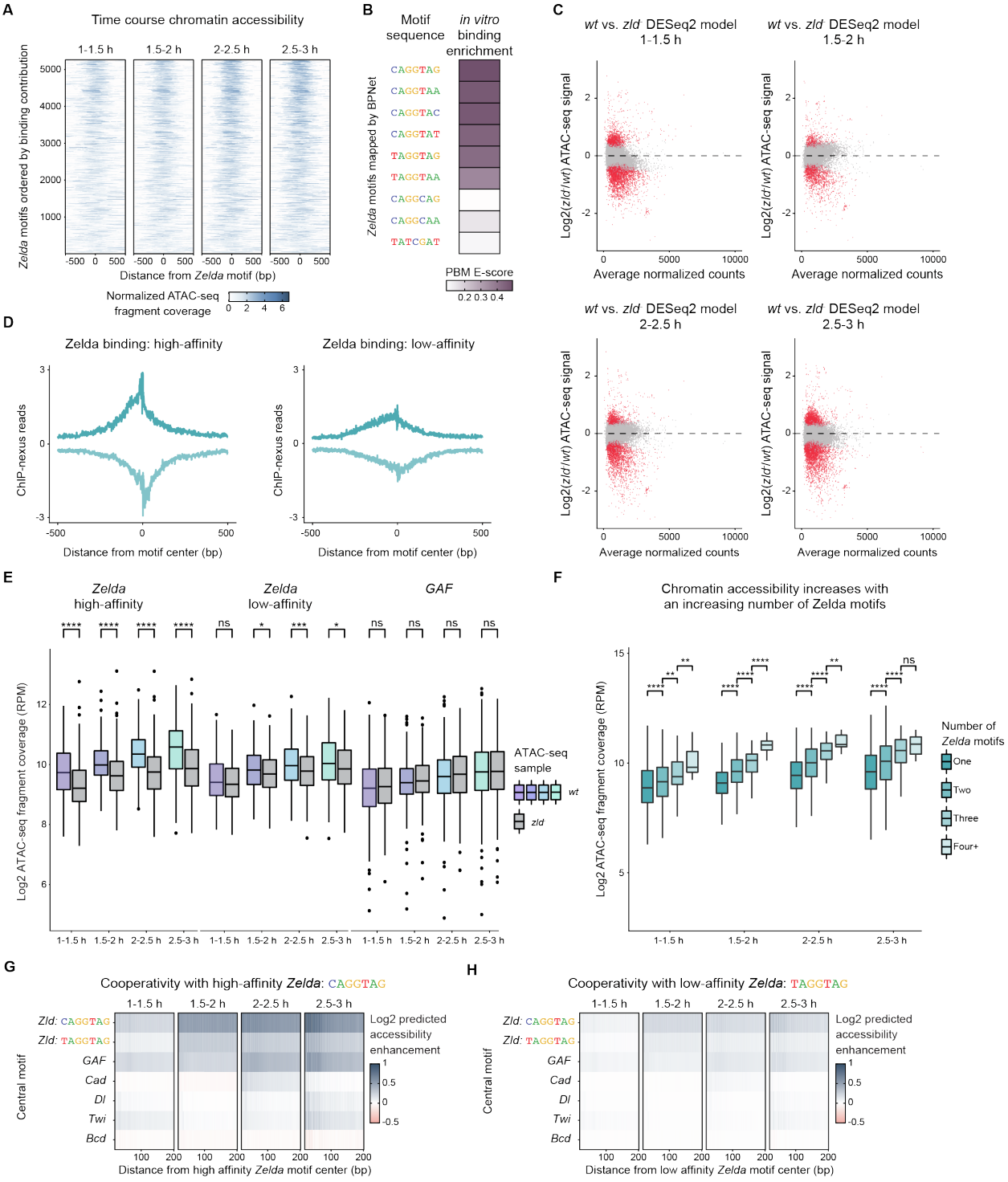
A
Time course chromatin accessibility
1-1.5 h
1.5-2 h
2-2.5 h
2.5-3 h
Zelda motifs ordered by binding contribution
Distance from Zelda motif (bp)
Normalized ATAC-seq fragment coverage
B
Motif sequence
in vitro binding enrichment
Zelda motifs mapped by BPNet
CAGGTAG
CAGGTAA
CAGGTAC
CAGGTAT
TAGGTAG
TAGGTAA
CAGGCAG
CAGGCAA
TATCGAT
PBM E-score
C
wt vs. zld DESeq2 model 1-1.5 h
wt vs. zld DESeq2 model 1.5-2 h
wt vs. zld DESeq2 model 2-2.5 h
wt vs. zld DESeq2 model 2.5-3 h
Log2(zld/wt) ATAC-seq signal
Average normalized counts
D
Zelda binding: high-affinity
Zelda binding: low-affinity
ChIP-nexus reads
Distance from motif center (bp)
E
Zelda high-affinity
Zelda low-affinity
GAF
Log2 ATAC-seq fragment coverage (RPM)
ATAC-seq sample
wt
zld
F
Chromatin accessibility increases with an increasing number of Zelda motifs
Number of Zelda motifs
One
Two
Three
Four+
G
Cooperativity with high-affinity Zelda: CAGGTAG
Zld: CAGGTAG
Zld: TAGGTAG
GAF
Cad
Dl
Twi
Bcd
Central motif
Log2 predicted accessibility enhancement
Distance from high affinity Zelda motif center (bp)
H
Cooperativity with low-affinity Zelda: TAGGTAG
Distance from low affinity Zelda motif center (bp)

Figure S5 | High-affinity Zelda motifs have greater effects on chromatin accessibility than low affinity Zelda motifs (related to Figure 3).
(**A**) Time course chromatin accessibility correlates with Zelda motif binding contribution. BPNet-mapped Zelda motifs were ordered by their counts contribution scores for Zelda binding as in Figure 3A. The experimentally generated ATAC-seq signal was calculated across a 1000 bp window, anchored on the Zelda motif, for each time point. The Zelda motifs that contribute most strongly to Zelda binding exhibit the highest chromatin accessibility. (**B**) Protein binding microarray (PBM) E-scores show differences between high- and low-affinity Zelda motifs. The E-score is a rank-based PBM statistic that is a variation on the area under the receiver operating characteristic curve (AUC) that ranges from -0.5 (lowest) to 0.5 (highest)[9]. E-scores for each Zelda heptad were calculated as done in Figure 3B. (**C**) Time course MA plots show differential chromatin accessibility between *wt* and *zld*$^{-}$ embryos. The differential chromatin accessibility was calculated between *wt* and *zld*$^{-}$ embryos using DESeq2 for all time points. Red highlighted dots are ATAC-seq peaks that are differentially accessible with statistical significance (FDR = 0.05). (**D**) Zelda is more strongly bound to high-affinity motifs than to low-affinity motifs. Average Zelda binding footprints were calculated and plotted across the same high- and low-affinity Zelda motifs as in Figure 3F. Average profiles were calculated across a 1000 bp window and were anchored on Zelda motifs. (**E**) Low-affinity Zelda motifs have a weaker effect on chromatin accessibility than high-affinity Zelda motifs. The average profiles in Figure 3F were quantified using boxplots and were tested for statistical significance using the using the Wilcoxon rank-sum test (* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$). Observed ATAC-seq fragment coverage was calculated across a 500 bp window centered on each Zelda motif using the same motif instances as in Figure 3F. There is an average of a five-fold weaker effect from low-affinity motifs than from high-affinity motifs, which was calculated using median values for accessibility for *wt* and *zld*$^{-}$ embryos at all time points. (**F**) Greater chromatin accessibility is associated with regions with more mapped Zelda motifs. All Zelda-containing islands were collected and separated based on how many Zelda motifs they contained. The observed normalized ATAC-seq fragment coverage for each time point was calculated across a 250 bp window anchored on the island center. Statistical significance was determined using the Wilcoxon rank-sum test (* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$). These results show that more Zelda motifs across a genomic region correlates with increased chromatin accessibility. This is consistent with previous results showing higher levels of nucleosome depletion for more Zelda motifs[10]. (**G**, **H**) ChromBPNet predicts that high-affinity Zelda motifs induce greater chromatin accessibility than low-affinity Zelda motifs. *In silico* motif injections into randomized sequences were performed as in Figure 3H, except the ChromBPNet models were used to predict chromatin accessibility at TF motifs upon injection of a (**G**) high-affinity or (**H**) low-affinity Zelda motif for each time point. Motif injections were repeated 512 times and predictions were averaged.

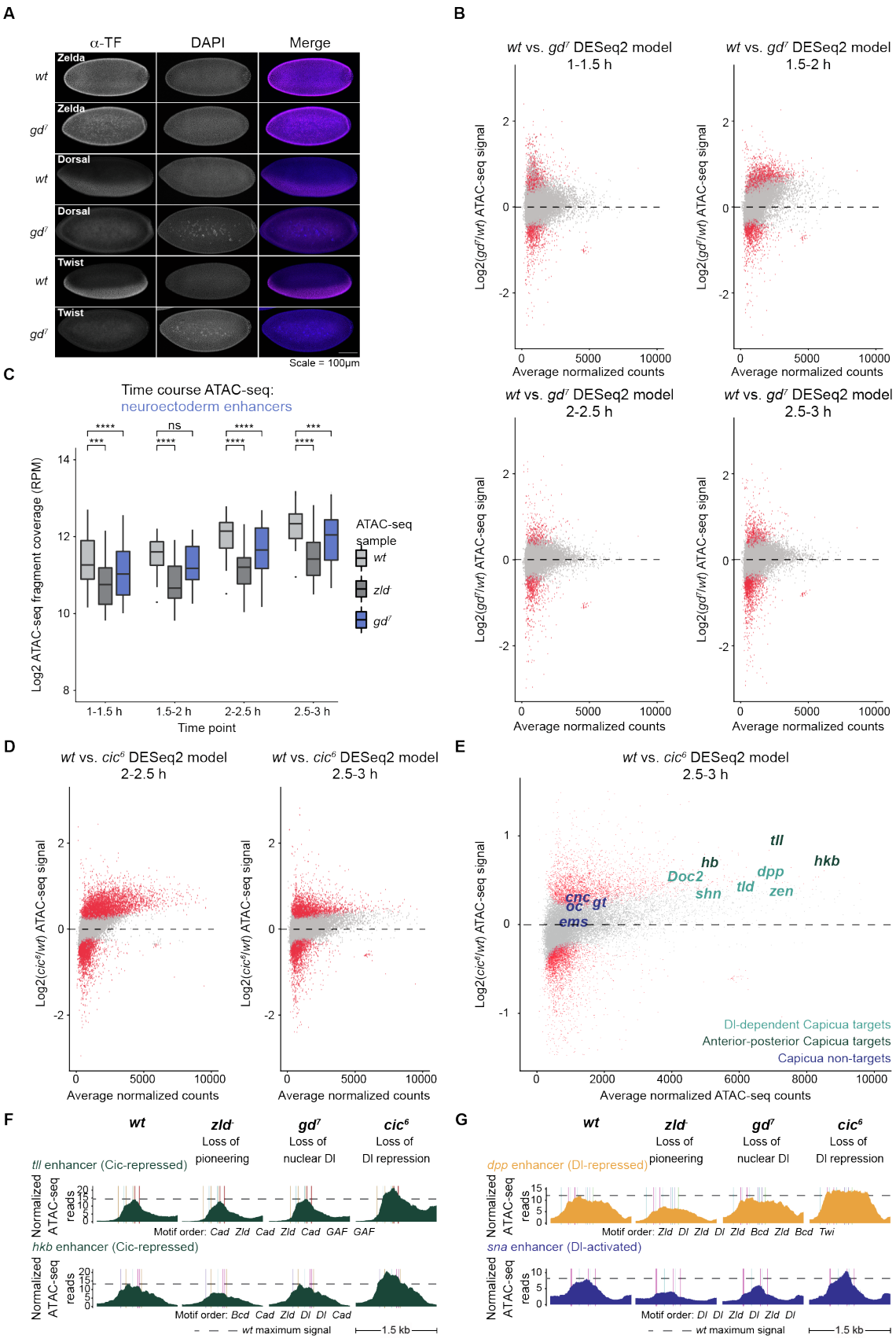
A
α-TF
DAPI
Merge
Zelda
wt
Zelda
gd7
Dorsal
wt
Dorsal
gd7
Twist
wt
Twist
gd7
Scale = 100μm
B
wt vs. gd7 DESeq2 model
1-1.5 h
wt vs. gd7 DESeq2 model
1.5-2 h
wt vs. gd7 DESeq2 model
2-2.5 h
wt vs. gd7 DESeq2 model
2.5-3 h
Log2(gd7/wt) ATAC-seq signal
Average normalized counts
C
Time course ATAC-seq:
neuroectoderm enhancers
Log2 ATAC-seq fragment coverage (RPM)
1-1.5 h
1.5-2 h
2-2.5 h
2.5-3 h
Time point
ATAC-seq sample
wt
zld-
gd7
D
wt vs. cic6 DESeq2 model
2-2.5 h
wt vs. cic6 DESeq2 model
2.5-3 h
Log2(cic6/wt) ATAC-seq signal
Average normalized counts
E
wt vs. cic6 DESeq2 model
2.5-3 h
Log2(cic6/wt) ATAC-seq signal
Average normalized ATAC-seq counts
tll
hb
Doc2
shn
tld
dpp
zen
hkb
cnc
oc
gt
ems
Dl-dependent Capicua targets
Anterior-posterior Capicua targets
Capicua non-targets
F
wt
zld-
Loss of pioneering
gd7
Loss of nuclear Dl
cic6
Loss of Dl repression
tll enhancer (Cic-repressed)
Normalized ATAC-seq reads
Motif order: Cad Zld Cad Zld Cad GAF GAF
hkb enhancer (Cic-repressed)
Motif order: Bcd Cad Zld Dl Dl Cad
wt maximum signal
1.5 kb
G
wt
zld-
Loss of pioneering
gd7
Loss of nuclear Dl
cic6
Loss of Dl repression
dpp enhancer (Dl-repressed)
Motif order: Zld Dl Zld Dl Zld Bcd Zld Bcd Twi
sna enhancer (Dl-activated)
Motif order: Dl Dl Zld Dl Zld Dl
wt maximum signal
1.5 kb

Figure S6 | Chromatin accessibility changes context-specifically in *gd*^{7} and *cic*^{6} embryos (related to Figure 5).
(**A**) *gd*^{7} embryos show a clear loss of dorsoventral patterning. Nuclear cycle 14 embryos were stained using the same Zelda, Dorsal, and Twist antibodies used in ChIP-nexus experiments. Confocal images of *wt* and *gd*^{7} embryos were collected using the same settings, maximum intensity projected, and processed in FIJI using the identical settings. (**B**) Time course MA plots show differential chromatin accessibility between *wt* and *gd*^{7} embryos. DESeq2 was used to determine differential chromatin accessibility for all time points, and the red points represent the ATAC-seq peaks that are significantly differentially expressed (FDR = 0.05). (**C**) Neuroectoderm enhancers lose chromatin accessibility in *gd*^{7} embryos. The normalized ATAC-seq fragment coverage was calculated in wt, *zld*$^{-}$, and *gd*^{7} embryos across known neuroectoderm enhancers (n = 23)[5,11] as in Figures 5C and 5D. Wilcoxon rank-sum tests were used to test for statistical significance (*** = $p < 0.001$; **** = $p < 0.0001$). In *gd*^{7} embryos, neuroectoderm enhancers are inactive. (**D**) MA plots show differential chromatin accessibility between *wt* and *cic*^{6} embryos at 2-2.5 h AEL and 2.5-3 h AEL. DESeq2 was run on *wt* and *cic*^{6} embryos for both time points to determine the differential chromatin accessibility. ATAC-seq peaks that are significantly differentially accessible are highlighted in red (FDR = 0.05). (**E**) In *cic*^{6} embryos, chromatin accessibility is increased at Capicua-repressed anterior-posterior patterning enhancers. In *wt* embryos, Capicua represses known anterior-posterior enhancers (e.g., *hb*, *tll*, *hkb*) in a Dorsal-independent fashion by binding high-affinity Capicua motifs[12]. Differential chromatin accessibility analysis was performed between *wt* and *cic*^{6} embryos as in Figure 5E. Both Dorsal-dependent (teal) and Dorsal-independent (green) enhancers gain accessibility in *cic*^{6} embryos, while enhancers not bound by Capicua (blue) do not. (**F**) Summary of chromatin accessibility at two Dorsal-independent Capicua-repressed enhancers (*tll* and *hkb*) upon loss of Zelda, nuclear Dorsal, and Dorsal-mediated repression as in Figure 5F. The dm6 enhancer coordinates are chr3R 30,851,400 - 30,852,900 for *tll*[13] and chr3R 4,347,821 - 4,349,321 for *hkb*[8]. Both enhancers do not significantly change accessibility in *gd*^{7} embryos but do show increased accessibility in *cic*^{6} as they are de-repressed. (**G**) Summary of chromatin accessibility at a Dorsal-repressed enhancer (*dpp*)[3] and Dorsal-activated enhancer (*sna*)[14] upon loss of Zelda, nuclear Dorsal, and Dorsal-mediated repression as in figure 5F. The dm6 enhancer coordinates are chr2L 2,456,160 - 2,457,660 (*dpp*) and chr2L 15,479,300 - 15,480,800 (*sna*).

Table S1 | Motif island summary statistics (related to Figure 4)

Number of each island type	
Island composition	**Number**
Bcd	7036
Bcd_Cad	404
Bcd_Cad_Dl_Zld	31
Bcd_Cad_Twi	135
Bcd_Cad_Twi_Zld	43
Bcd_Cad_Zld	107
Bcd_Dl	75
Bcd_Dl_Twi	69
Bcd_Dl_Twi_Zld	32
Bcd_Dl_Zld	43
Bcd_GAF	214
Bcd_GAF_Twi	78
Bcd_Twi	981
Bcd_Twi_Zld	161
Bcd_Zld	393
Cad	6405
Cad_Dl	64
Cad_Dl_Twi	59
Cad_GAF	244
Cad_GAF_Twi	107
Cad_Twi	867
Cad_Twi_Zld	72
Cad_Zld	268
Dl	449
Dl_GAF	33
Dl_Twi	235
Dl_Twi_Zld	75
Dl_Zld	70
GAF	4136
GAF_Twi	681
GAF_Twi_Zld	46
GAF_Zld	141
Twi	11763
Twi_Zld	366
Zld	1673
Island summary statistics	
Minimum	200
1st quartile	200
Median	200
Mean	245
3rd quartile	266
Maximum	945

Summary of motif islands generated according to the scheme in Figure 4A. Only motifs identified and mapped by BPNet and that are bound by their associated TF are used for island generation. Islands with fewer than 30 genomic instances are excluded. Motif islands are not separated according to how many motifs they contain but are instead classified based on which TF motifs compose them. Summary statistics for motif islands show that islands are approximately the sizes as *Drosophila* enhancers, with no island greater than 945 bp. Islands that are 200 bp wide are single-motif islands.

Table S2 | Oligonucleotides used for ChIP-nexus (related to STAR Methods).

Name	Identity	Modification	Barcode	Sequence
Nex_adapter_U	Adaptor: universal	5' phosphate	/	/5Phos/GATCGGAAGAGCACACGTCTG GATCCACGACGCTCTTCC
Nex_adapter_1	Adaptor: barcoded	5' phosphate	AGTCNNNNN	/5Phos/AGTCNNNNNAGATCGGAAGAG CGTCGTGGATCCAGACGTGTGCTCTT CCGATCT
Nex_adapter_2	Adaptor: barcoded	5' phosphate	CAGTNNNNN	/5Phos/CAGTNNNNNAGATCGGAAGAG CGTCGTGGATCCAGACGTGTGCTCTT CCGATCT
Nex_adapter_3	Adaptor: barcoded	5' phosphate	GTCANNNNN	/5Phos/GTCANNNNNAGATCGGAAGAG CGTCGTGGATCCAGACGTGTGCTCTT CCGATCT
Nex_adapter_4	Adaptor: barcoded	5' phosphate	TCAGNNNNN	/5Phos/TCAGNNNNNAGATCGGAAGAG CGTCGTGGATCCAGACGTGTGCTCTT CCGATCT
Nex_primer_U	Primer: universal	3' phosphoro-thioate bond	/	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATC*T
Nex_primer_01	Primer: indexed	3' phosphoro-thioate bond	ATCACG	CAAGCAGAAGACGGCATACGAGATCG TGATGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_02	Primer: indexed	3' phosphoro-thioate bond	CGATGT	CAAGCAGAAGACGGCATACGAGATAC ATCGGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_03	Primer: indexed	3' phosphoro-thioate bond	TTAGGC	CAAGCAGAAGACGGCATACGAGATGC CTAAGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_04	Primer: indexed	3' phosphoro-thioate bond	TGACCA	CAAGCAGAAGACGGCATACGAGATTG GTCAGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_05	Primer: indexed	3' phosphoro-thioate bond	ACAGTG	CAAGCAGAAGACGGCATACGAGATCA CTGTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_06	Primer: indexed	3' phosphoro-thioate bond	GCCAAT	CAAGCAGAAGACGGCATACGAGATAT TGGCGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_07	Primer: indexed	3' phosphoro-thioate bond	CAGATC	CAAGCAGAAGACGGCATACGAGATGA TCTGGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T
Nex_primer_08	Primer: indexed	3' phosphoro-thioate bond	ACTTGA	CAAGCAGAAGACGGCATACGAGATTC AAGTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATC*T

ChIP-nexus adapters are generated using the universal adapter (Nex_adapter_U) and the barcoded adapters (Nex_adapter_1, Nex_adapter_2, Nex_adapter_3, Nex_adapter_4). The universal (Nex_primer_U) and indexed (Nex_primer_01 - Nex_primer_08) primers are used for amplification of the library. Oligonucleotides can be ordered from Integrated DNA Technologies (IDT).

Supplemental reference list

1. Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D., and Small, S. (2005). The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proc Natl Acad Sci USA 102, 4960–4965. 10.1073/pnas.0500373102.
2. Hannon, C.E., Blythe, S.A., and Wieschaus, E.F. (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. eLife 6. 10.7554/eLife.28275.
3. Huang, J.D., Schwyter, D.H., Shirokawa, J.M., and Courey, A.J. (1993). The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. Genes Dev. 7, 694–704. 10.1101/gad.7.4.694.
4. 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. 10.1242/dev.01124.
5. Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., 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 Dev. 21, 385–390. 10.1101/gad.1509607.
6. Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci USA 108, 13570–13575. 10.1073/pnas.1109873108.
7. Hoch, M., Schröder, C., Seifert, E., and Jäckle, H. (1990). cis-acting control elements for Krüppel expression in the Drosophila embryo. EMBO J. 9, 2587–2595. 10.1002/j.1460-2075.1990.tb07440.x.
8. Häder, T., Wainwright, D., Shandala, T., Saint, R., Taubert, H., Brönner, G., and Jäckle, H. (2000). Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal. Curr. Biol. 10, 51–54. 10.1016/s0960-9822(99)00265-1.
9. Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. 24, 1429–1435. 10.1038/nbt1246.
10. Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., and Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res. 25, 1703–1714. 10.1101/gr.192542.115.
11. Koenecke, N., Johnston, J., He, Q., Meier, S., and Zeitlinger, J. (2017). Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res. 27, 64–74. 10.1101/gr.209486.116.
12. Ajuria, L., Nieva, C., Winkler, C., Kuo, D., Samper, N., Andreu, M.J., Helman, A., González-Crespo, S., Paroush, Z., Courey, A.J., et al. (2011). Capicua DNA-binding sites are general response elements for RTK signaling in Drosophila. Development 138, 915–924. 10.1242/dev.057729.
13. Liaw, G.J., Steingrimsson, E., Pignoni, F., Courey, A.J., and Lengyel, J.A. (1993). Characterization of downstream elements in a Raf-1 pathway. Proc Natl Acad Sci USA 90, 858–862. 10.1073/pnas.90.3.858.
14. Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. Genes Dev. 6, 1518–1530. 10.1101/gad.6.8.1518.