Figure S1

A

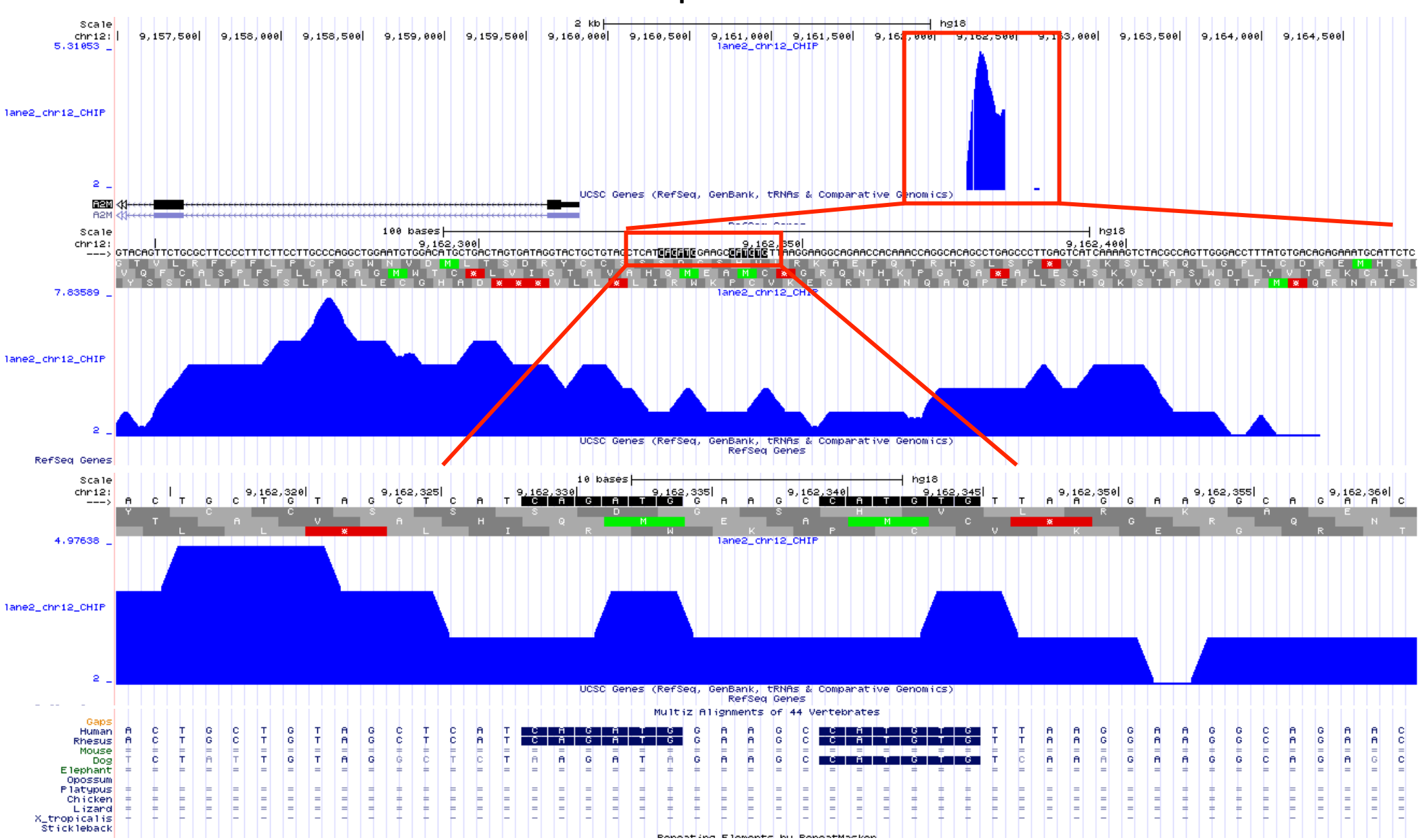

A2M promoter
Scale
chr12:
2 kb
hg18
lane2_chr12_CHIP
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
A2M
100 bases
RefSeq Genes
10 bases
Multiz Alignments of 44 Vertebrates
Gaps
Human
Rhesus
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X_tropicalis
Stickleback

Figure S1

B

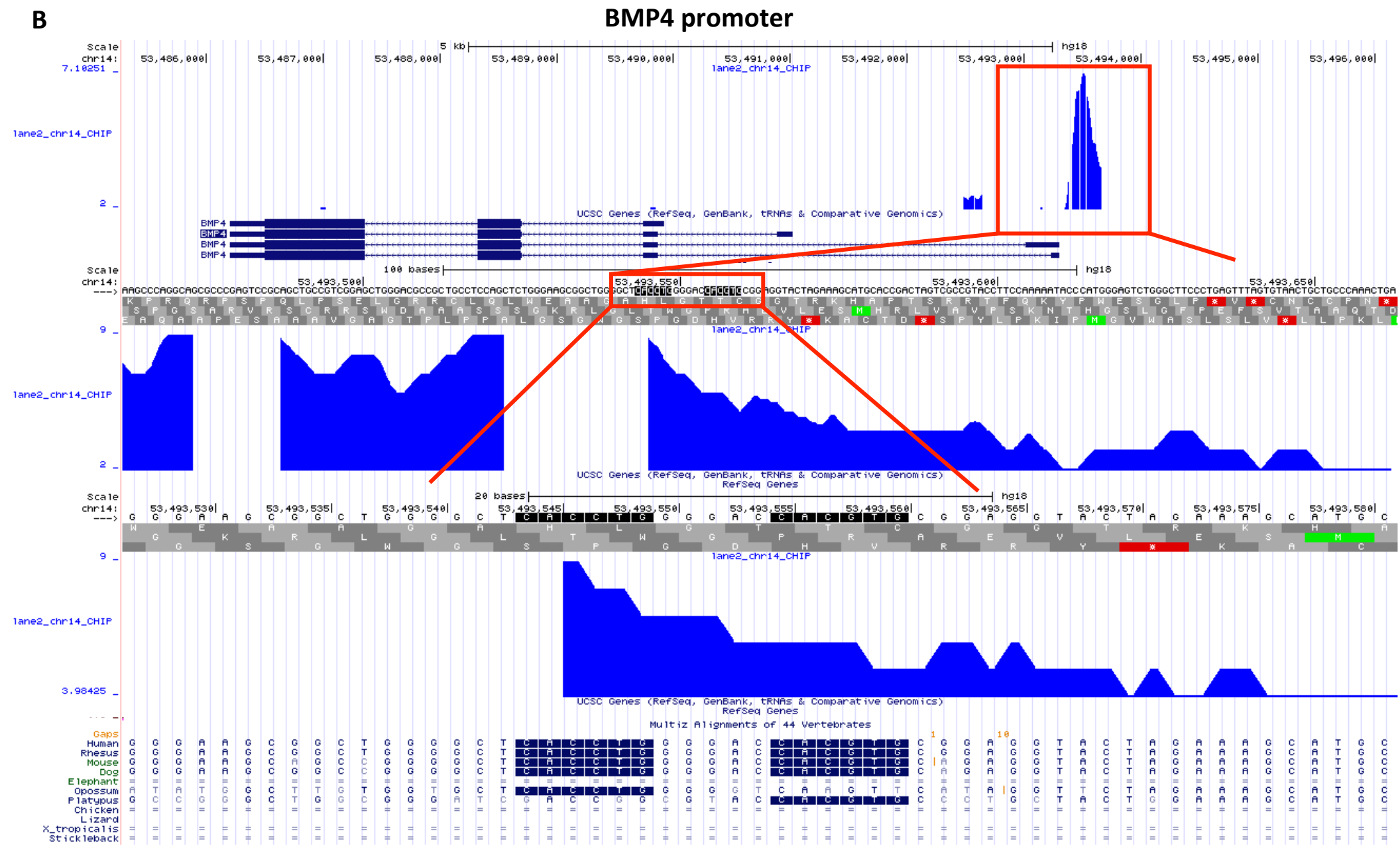

BMP4 promoter
Scale
chr14:
5 kb
hg18
lane2_chr14_CHIP
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
BMP4
100 bases
20 bases
RefSeq Genes
Multiz Alignments of 44 Vertebrates
Gaps
Human
Rhesus
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X_tropicalis
Stickleback

Figure S1

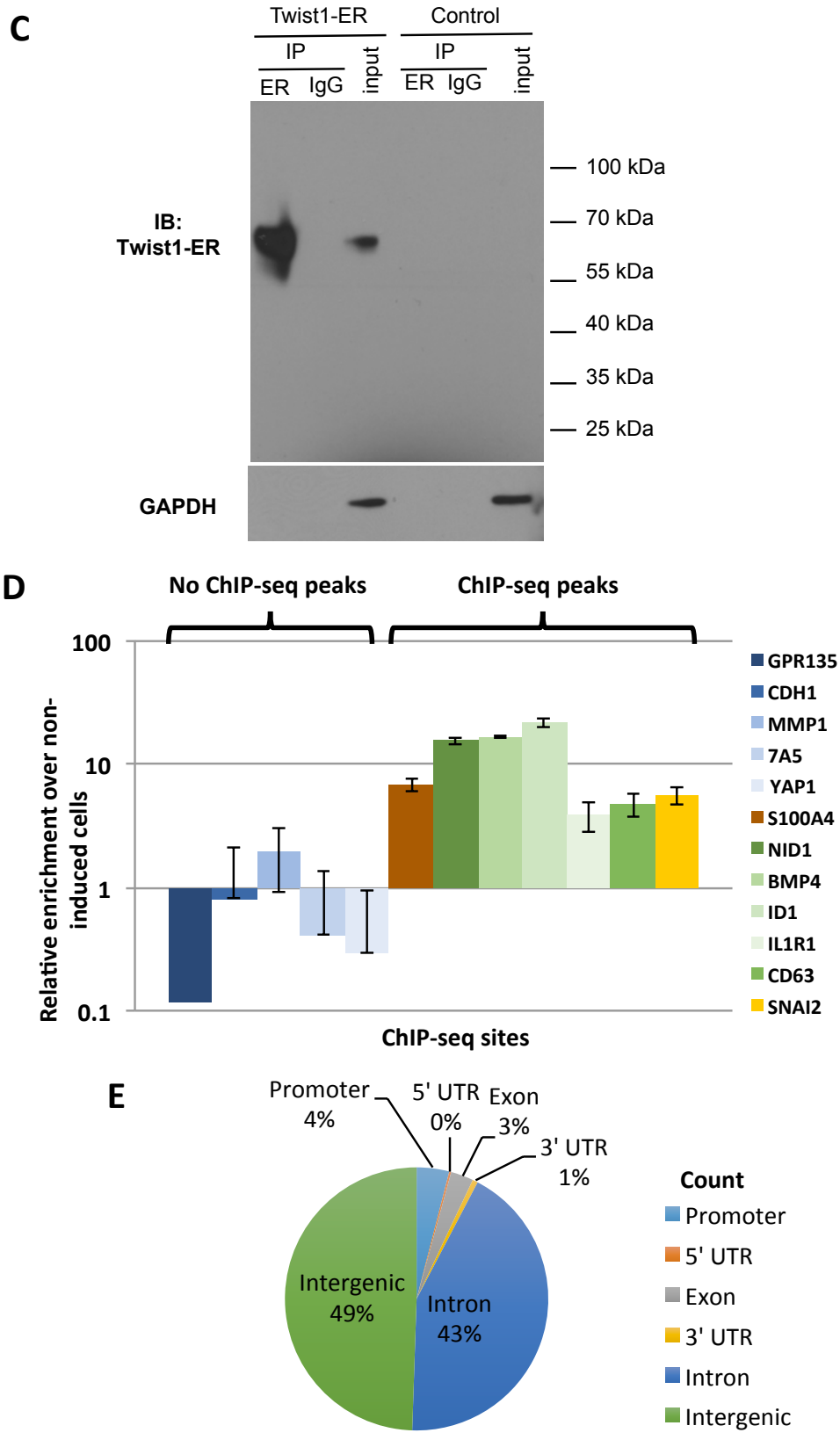

C
Twist1-ER
Control
IP
input
ER
IgG
IB:
Twist1-ER
100 kDa
70 kDa
55 kDa
40 kDa
35 kDa
25 kDa
GAPDH
D
No ChIP-seq peaks
ChIP-seq peaks
Relative enrichment over non-induced cells
100
10
1
0.1
ChIP-seq sites
GPR135
CDH1
MMP1
7A5
YAP1
S100A4
NID1
BMP4
ID1
IL1R1
CD63
SNAI2
E
Promoter 4%
5' UTR 0%
Exon 3%
3' UTR 1%
Intergenic 49%
Intron 43%
Count
Promoter
5' UTR
Exon
3' UTR
Intron
Intergenic

Figure S1: Related to Figure 1.

(A-B) A Twist1-binding site by ChIP-seq is mapped on the UCSC Genome Browser at the promoter regions of BMP4 (A) and A2M (B). A magnified view shows the double E-box sequence in the binding site highlighted with a red square. (C) HMLE-Twist1ER or HMLE-control cells were lysed and immunoprecipitated with the anti-ER antibody or control IgG and processed for western analysis for Twist1 and GAPDH. The data show that the ER antibody specifically recognizes and immunoprecipitate the Twist1-ER fusion protein from HMLE-Twsit1ER cells. (D) HMLE-Twist1ER cells were treated with 4-OH Tamoxifen for four days to activate Twist1 and then processed for ChIP using an anti-ER antibody. Enriched Twist1-bound DNA fragments were used for qPCR analyses of randomly selected Twist1-binding peaks and non-bound regions identified by ChIP-seq. (E) Human Twist1-bound DNA fragments were categorized based on their locations in the human genome.

Figure S2

Motif Name	Consensus	Species	P-value	Log P-value	# of Target Sequences with Motif(of 13876)	% of Target Sequences with Motif	# of Background Sequences with Motif(of 31415)	% of Background Sequences with Motif	Fold enrichment above background
Ebox-x2-Nx5	CANNTGNNNNNNCANNTG	Human	1e-2097	-4.83E+03	1994	14.37%	166.5	0.53%	27.11
Ebox-x2-Nx7	CANNTGNNNNNNNCANNTG	Human	1.00E-175	-4.04E+02	389	2.80%	138.4	0.44%	6.36
Ebox-x2-Nx6	CANNTGNNNNNNNCANNTG	Human	1.00E-90	-2.09E+02	382	2.75%	252.8	0.80%	3.44
Ebox-x2-Nx3	CANNTGNNNCANNTG	Human	1.00E-90	-2.08E+02	348	2.51%	214.9	0.68%	3.69
Ebox-x2-Nx8	CANNTGNNNNNNNNCANNTG	Human	1.00E-82	-1.90E+02	301	2.17%	177.3	0.56%	3.88
Ebox-x2-Nx2	CANNTGNNCANNTG	Human	1.00E-72	-1.66E+02	308	2.22%	206.6	0.65%	3.42
Ebox-x2-Nx1	CANNTGNCANNTG	Human	1.00E-45	-1.05E+02	248	1.79%	196.1	0.62%	2.89
Ebox-x2-Nx9	CANNTGNNNNNNNNNCANNTG	Human	1.00E-38	-8.88E+01	214	1.54%	172.8	0.54%	2.85
Ebox-x2-Nx4	CANNTGNNNNCANNTG	Human	1.00E-36	-8.31E+01	251	1.81%	230.2	0.73%	2.48
Ebox-x2-Nx10	CANNTGNNNNNNNNNNCANNTG	Human	1.00E-16	-3.86E+01	223	1.61%	274.3	0.86%	1.87
Ebox-x2-Nx0	CANNTGCANNTG	Human	1.00E-12	-2.81E+01	188	1.35%	244.4	0.77%	1.75
Ebox-x2-Nx5	CANNTGNNNNNNCANNTG	fly	1.00E-35	-8.07E+01	52	5.65%	188.2	0.52%	10.87
Ebox-x2-Nx7	CANNTGNNNNNNNCANNTG	fly	1.00E-11	-2.67E+01	25	2.71%	162.8	0.45%	6.02
Ebox-x2-Nx6	CANNTGNNNNNNCANNTG	fly	1.00E-01	-3.11E+00	14	1.52%	326.5	0.91%	1.67
Ebox-x2-Nx3	CANNTGNNNCANNTG	fly	1.00E-02	-6.69E+00	19	2.06%	333.2	0.92%	2.24
Ebox-x2-Nx8	CANNTGNNNNNNNNCANNTG	fly	1.00E-07	-1.69E+01	22	2.39%	208.8	0.58%	4.12
Ebox-x2-Nx2	CANNTGNNCANNTG	fly	1.00E-04	-1.14E+01	19	2.06%	229.3	0.64%	3.22
Ebox-x2-Nx1	CANNTGNCANNTG	fly	1.00E-02	-5.34E+00	13	1.41%	218.4	0.61%	2.31
Ebox-x2-Nx9	CANNTGNNNNNNNNNCANNTG	fly	1.00E-02	-5.90E+00	14	1.52%	228.4	0.63%	2.41
Ebox-x2-Nx4	CANNTGNNNNCANNTG	fly	1.00E-04	-9.36E+00	14	1.52%	159.4	0.44%	3.45
Ebox-x2-Nx10	CANNTGNNNNNNNNNNCANNTG	fly	1.00E-05	-1.33E+01	18	1.95%	181.7	0.50%	3.90
Ebox-x2-Nx0	CANNTGCANNTG	fly	1.00E-02	-6.10E+00	23	2.50%	460.2	1.28%	1.95

Figure S2: Related to Figure 1 &2.

Permutation analyses for the enrichment of the motifs containing two perfect E-boxes separated by 0-10 nucleotides among human and Drosophila Twist-binding peaks. The results show that two E-boxes separated by 5nt spacing are more common than any other spacing configurations for both human and Drosophila binding.

Figure S3

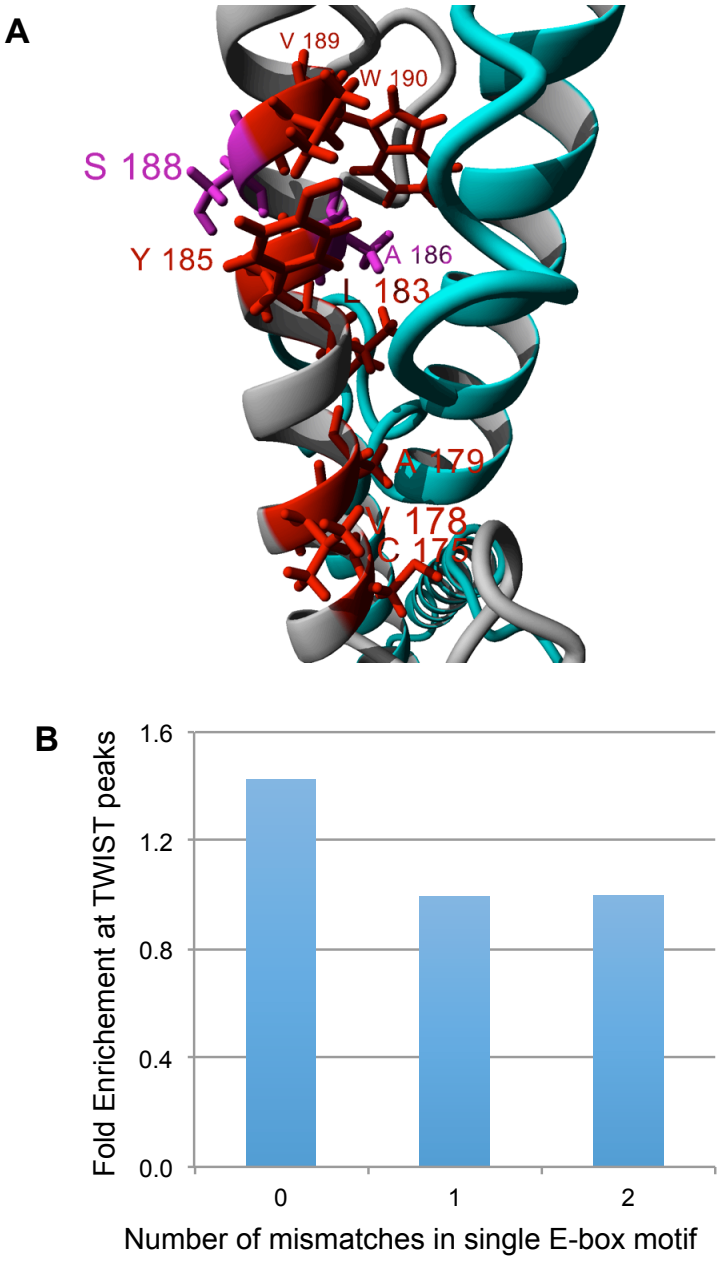


Figure S3: related to Figure 3.

A) A high-resolution view of the interacting face between the two WR domains of the human Twist1 protein. B) Computational analysis of the single E-box motif at human Twist1-binding peaks shows that the CANNTG sequence with 1 or 2 mismatched nucleotide are not enriched above the calculated random occurrence frequency of each motif.

Figure S4

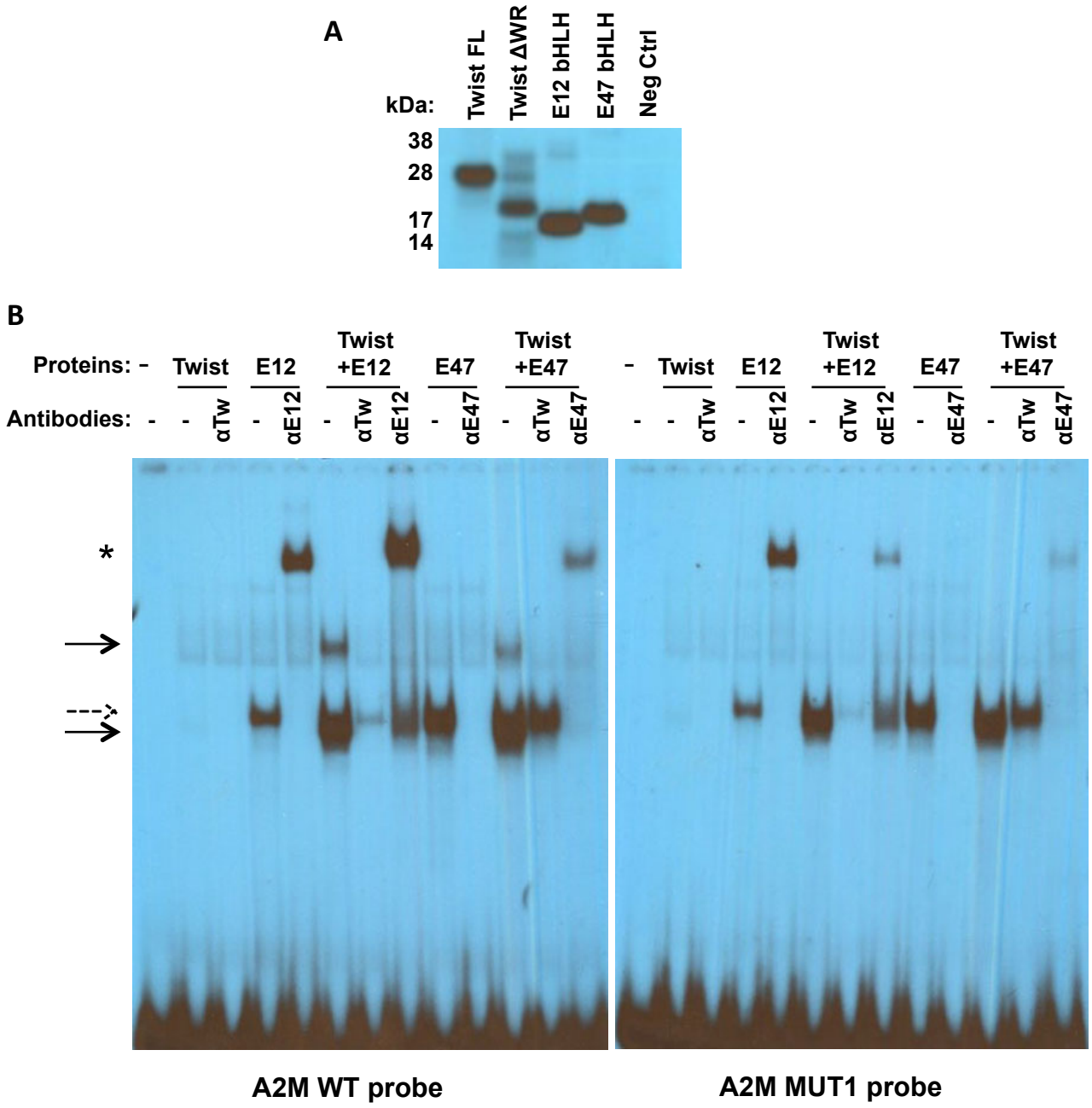
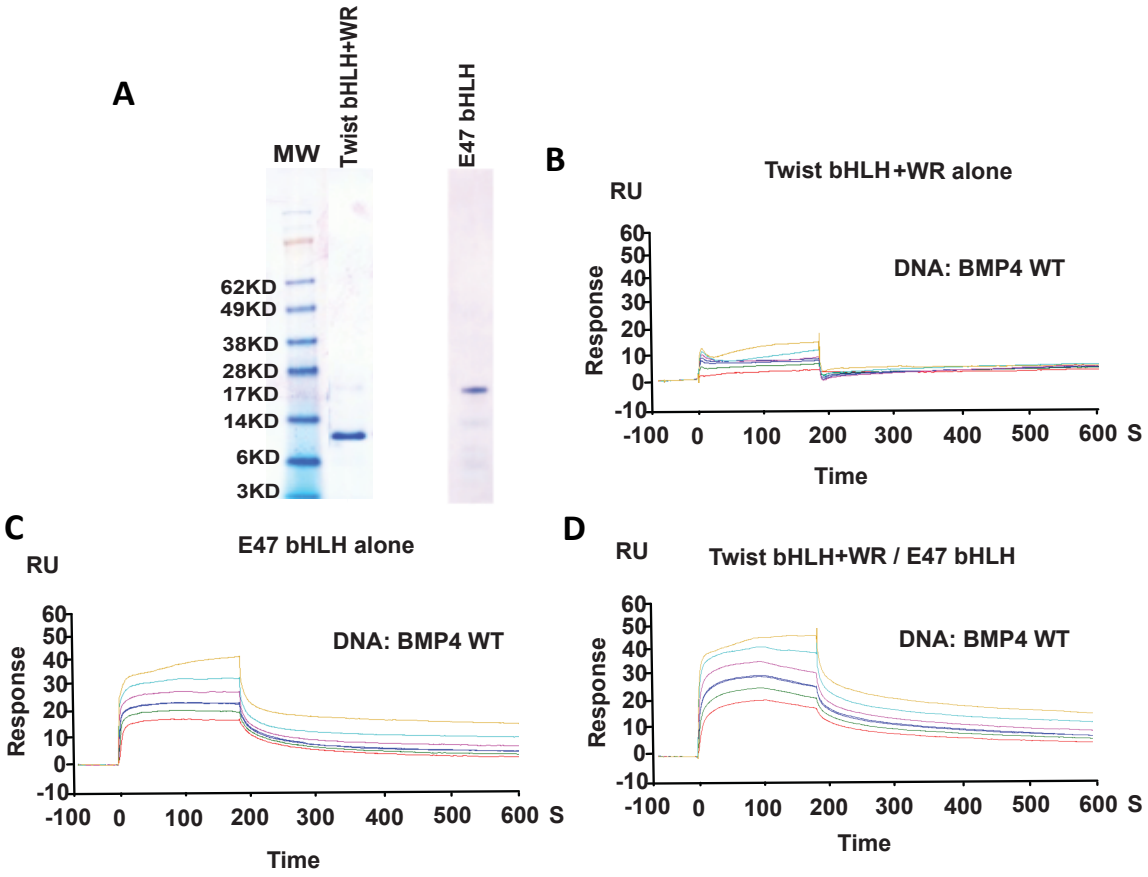


Figure S4: related to Figure 4 and 6.

(A) In vitro translated S^{35}-labeled Twist1 full-length, Twist1ΔWR, E12bHLH and E47bHLH proteins were visualized by SDS-PAGE and used to normalize the amount of proteins used in individual EMSA assays. (B) EMSA analysis of human Twist1/E12 or Twist1/E47 bound to a 30nt oligo containing the Double E-box motif from the human *A2M* promoter or the same oligo

with one of the E-boxes mutated. Antibodies against Twist1, E12, or E47 were added into the corresponding reactions to either supershift (for E12) or block (for Twist1 and E47) the protein/DNA complexes. These are the replicate experiments shown in Fig. 4E and 4F, and these gels were run for a shorter period of time to show equal amounts of excess free probes.

Figure S5



E **Binding Parameters:**

Analyte	Ligand (DNA)	$k_a(M^{-1}s^{-1})$	$k_d(s^{-1})$	$K_D(M)$
Twist bHLH+WR	BMP4 WT	/	/	/
E47 bHLH	BMP4 WT	6731.403	0.000981	1.46E-07
Twist bHLH+WR / E47 bHLH	BMP4 WT	16989.7	0.00133	7.83E-08

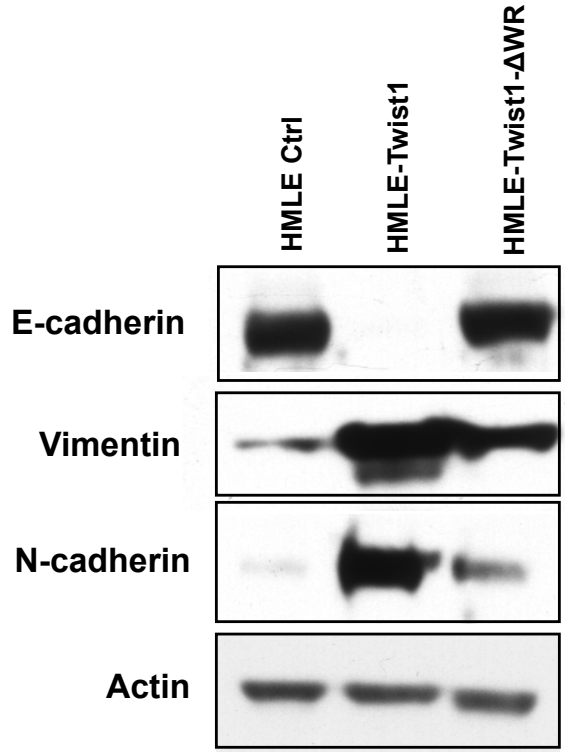
Figure S5: related to Figure 5.

(A) Purified His-tagged Twist1 bHLH+WR and E47 proteins were expressed in E.coli by using codon optimization strategies to the fully synthetic hTwist1 coding sequence and inserted into the pQE30 vector. (B-D) SPR-based kinetic analysis of the association and dissociation rates between the double E-box motif with 5nt spacing and Twist1 bHLH+WR (A), E47 bHLH (B),

and Twist1 bHLH+WR and E47 bHLH together (C). (D) Summary of the binding parameters obtained by the SRP-based kinetic analysis shown in (B-D).

Figure S6

A



B

Gene Name	Distance from TSS(Double E-box Sequence)
BMP4	238(CTCACCTGGGGACCACGTGC)
SYT12	403(CACACTTGCTCCCCATCTGG)
LAMB1	924(GCCACATGGCCCCCATCTGT)
FILIP1L	28(TGCCTGTGCCCAGCAGCTGC)
CSF2	4268(GTCATCTGGTGCCCTGCTGT)
NTSR1	5464(ACCATGTGGACAGCAGGCGG), 5296(GCCAGGTGGAAGGCAGATGG)
A2M	2502(ATCAGATGGAAGCCATGTGT)
SERPINF1	2413(GCCGGGTGGAGACCAGCTGT)

Figure S6: related to Figure 6.

(A) HMLE control cells and HMLE cells expressing Twist1 and Twist1ΔWR were processed for Western blot analyses for expression of epithelial marker; E-cadherin, and mesenchymal markers; vimentin and N-cadherin. (B) A list of positions of individual double E-box motif sites relative to Transcription Start Site (TSS) for individual genes tested in Figure 7C.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES: related to Experimental Procedures.

Primers used for qPCR analysis of Twist1 target genes used in Fig. 7C.

Gene Name	Primer Sequence
BMP4 Forward	CCGCAGCCTAGCAAGAGTG
BMP4 Reverse	GCTCAGGATACTCAAGACCAGTG
SYT12 Forward	CAGAATACCATCTGAGCGTCATC
SYT12 Reverse	TAGTCGTAATTGGGGAACGGA
LAMB1 Forward	CACAAGCCCGAACCCTACTG
LAMB1 Reverse	GACCACATTTTCAATGAGATGGC
FILIP1L Forward	ACAGCTCACCCTTCAAAGACA
FILIP1L Reverse	TCCTTCTCTAGTCTGGTTGCC
CSF2 Forward	TCCTGAACCTGAGTAGAGACAC
CSF2 Reverse	TGCTGCTTGTAGTGGCTGG
SERPINF1 Forward	GGAAATTCCCGATGAGATCAGC
SERPINF1 Reverse	AGTCAAACTTTGTTACCCACTGC
NTSR1 Forward	AGCAGTGGACTCCGTTCCT
NTSR1 Reverse	GTTGGCAGAGACGAGGTTGT
A2M Forward	AGGAAATCGCATCGCACAATG
A2M Reverse	ACGGTGAAAGGGTGCTCTG
GAPDH Forward	GACCCCTTCATTGACCTCAAC
GAPDH Reverse	CTTCTCCATGGTGGTGAAGA