

SUPPLEMENT

Supplement section overview:

Section I Primers used in this study

Section II Dorsal target enhancer sequences

Section III False-positive cluster/control sequences

Section IV Bioinformatics

Section IVa Results of MERmaid analysis

Section IVb Parameters used for *Drosophila melanogaster* genome query

Section I. Primers used in this studyPrimers for enhancers defined in this study:

CG12443left(SpeI):	aatactagtGGACAGGACAATGGAGGTTTCAGAAGAAGCGAGC
CG12443right(EcoRI):	aatgaattcGGCTGACCACCGTAAATCTGCGGGGGAAATCCCAGCC
BRK-left(182):	CTCTGGCACAAACCCTAAATGTGGATTACGCT
BRK-right(183):	GGGACAAGTGGCTCCCACAACGATCGGTTGGA
VND(348)-Lft-(SpeI):	aatactagtGACCCTGGCTTAGAAATCCCCTAGGTGAG
VND(348)-Rt-(EcoRI):	aatgaattcGCAAATCCGCTCGTAAGGATCGCAGCTACC
VND(743)-Lft-(SpeI):	ttcactagtGTCCCGCAAATTGTATATAATCGAGTGATTAATG
VND(743)-Rt-(EcoRI):	gaagaattcGTAGGTGAGAGCCAGGGAAACCCCAATCGGG
PHM-L(SpeI):	aatactagtGCTGCTCCTGCTTATCGGAGTGATCAGTGTGG
PHM-R(EcoRI):	aatgaattcACGGCGCAGGTTTCGAGAAGTGCAATCTACAGG
ADY-Left(186):	CCAAGGTGCAATTTTGTATGCAGTGC GGCT
ADY-Right(187):	CCGAAACTCCCTTGCCAAAAGCCCCACGAT
SIM-RZ F06(SpeI):	taaactagtCCCCGGCATATGTTACGCACATTTACAGCGTATG
SIM-RZ R07(EcoRI):	TAAGAATTCGGTTACAGGCAAACAGCAAAGTAAACAAATGG
VEIN-topAE:	TTATTGAAAGTGCCGAAGTTAGCGG
VEIN-botAE:	CTGAAACGTGGTTAAAGTGGCC
A.g.SIM-RZ F45(SpeI):	aagactagtACTGTTTACCGAGTGGAAAGGTCCTAACG
A.g.SIM-RZ R45(EcoRI):	tatgaattcATCGCGCCCGCACGATCGACTGCGCAGGAAGC

Primers for false-positive/control clusters C1-C10:

C1-TAK1-LEFT:	CGCGCTACCAAGCCCCATAGCCACCACCA
C1-TAK1-RIGHT:	CCACACACTCCCCTAGTCTGTGAATTGCA
C2-island1L(SpeI):	aatactagtACTGGAAATCATGTCACGTGCCTGGCGTGCC
C2-island2(RI):	aatgaattcGATTCCCATACCGATTCCAAATGCGAATGCGG
C3-RuntLEFT1(SpeI):	aatactagtGGTTTTCCCGATACTTTTAGGAAATCCCTTTCAGC
C3-RuntRIGHT1(RI):	aatgaattcCGTGATAAGCTATTCTAAATCGAATCGTTGGGG
C4-KrH2-L(SpeI):	attactagtCAGCTTAGAACAAGGGAATTTCCCTCTAGC
C4-KrH2-R(RI):	atagaattcGCCACTTGAGGGAAGATATATGTGGGTATGGG
C5-PpD5-LFT(SpeI):	aatactagtCTTTGAAGCCCACCGACCCATTTGTCCCGCCC
C5-PpD5-RT(RI):	aatgaattcAGGGGGAGCTGCTTTGCGGAGGGGTTTCGGA
C6-Fas3-L(SpeI):	aatactagtGACCAAGCTAATCACAAAAGGCTGAGAGAATGG
C6-Fas3-R(RI):	aatgaattcGGTCGAGGCTATGGAACGCTTCCGATTTCGGG
C7-CG5549-L(SpeI):	aatactagtCCTGACCGACGGGAAGGAAAAGTTTGCCGGCAC
C7-CG5549-R(RI):	aatgaattcGATCAAATGAGATGAGTTGAGATGAGATGGCGTTG
C8-Cli-L(Bgl2):	aatagatctCCACGGAGACAAAGACAAAGACAGAGAGTCC
C8-Cli-R(SpeI):	aatactagtGGCTAAGCCGCGGAGGTGTGGAGGGTCGGCCGG
C9-CG1924-L(SpeI):	aatactagtCCCCATGCATTTTACCATCCTTTACCATCGGC
C9-CG1924-R(RI):	aatgaattcTTGTGGCAATCTTGCCAACCATGTGTAGCAC
C10-Ben-L(SpeI):	aatactagtCCAATGGATAAGGCGTTCGGACTTCGGATCC
C10-Ben-R(RI):	aatgaattcAGGCAGCAAGAGCCAACGAACAAAGTTGGG

Primers for mutagenesis:

VND Δ E(CA)-RZ F04mut: GAAACCCCAATCGGGAATGgatccaaTACGACAGACATGGGACTCAG
VND Δ E(CA)-RZ F05mut: GCAAGCAAACCTTGCGCCAAtggaaaaGCACGACCTGTTTCGACCCG
BRK Δ E(CA)-(227): GCCAAAGTTAGAGGCACAGGCtatctaaCGTGTTTGGTTTGAACGGG
BRK Δ E(CA)-(229): GGCCAACCAACGGCtgctaaaGTTTCATGTTAGGACCG
VND Δ CTG-RZ F21mut: cctGTGctgacctGCAGtgtgtgGACCAGGTAGCTGCG
VND Δ CTG-RZ F22mut: CCCGTAAAGAGTCCCTGgggggggGTGccccccGCAGtgtgtgGACCAGG
BRK Δ CTG-RZ F46mut: CAATTAGCCGGACGGACACAAAaaaaaaaGAGCaaaaaaGCAAAAAGACACGGTTGTCC
BRK Δ CTG-RZ F47mut: GAAGGAAACTAAAGGAAACTGAGAGAGAGTTCTCGCTATGGATTGTGCAC
BRK Δ Su(H)-RZ F44mut: GCGactagtTGGTTTTTCCCAtttTTACAATTAGCCGGACGG
BRK Δ Su(H)-RZ F45mut: GATGCACCTCTGGGAGATTCCCtttACCGGCAGCAGGTCATG

Section II. Dorsal target enhancer sequences

>ZEN dorsal ectoderm

TCTAGAATGAACGAAAACAGTATCTGGTTTTCCCGAAAACTTATGAATTTAAAAATGCACCTTTATTCACATACTCACACATGCCTGCCATAAAAATG
 ATTCGCGATTTTTCGCGAACACCCCGCGGATCATAAAACATTTGCACCAGCTGCCTGTGTTTATTCACCTACCTGAAACCATACTCTTATCGCCTGATC
 CTCGCGCGGTCGCACTATTTAGGTAGACACTGTACAGGCAGCACTMCGCGGGCGCCGAGGCAACTTGTTCGCCCTATTTCTTTGATATAAGTTTGG
 GAAATCCAGAAGTCCAATAACGGGGCTATATGAACGAATATTGATTTGGGTTTCTCCAGTTATAGAGTTTATTTGATCTTGGGCGCGTTTATGATAGGT
 GGATTTATGGATGTCTGGGAAAACCAAGCTCTGAATCCATTTCTACTTATCACAAGACTTGTTCCTGTGCCCTTCATCACACCTAGTTTAACTACTAT
 AATTAATAAATCTTTTAAATCCGCCATTTATGAAGATAATTTATAGCATCTATGCATCTTCAAAATATCGCGAGATATCGCAAAGTATTCATAAATTA
 GTTTTTTATTGATATGTAAAGCT

>SOG broad lateral neurogenic ectoderm

GTTGCCAATGCCATTCGCGCATACGCCGTGTCTATATGGCTATATGGCTATATGGCTGTATGGTGGCGGGAAATCCCGTAATCGCAGGTAGAATCC
 AGCCGGTGGCAGGGCGGACTGCTCGCACCTCTAATCCCGCCAGGGTTTTCGGGACATGGGATATCCCGACGGCACAGCATAGCACTCCGTTTTCTTT
 TTTTTTTTTTATTATTATTGTGTCCAGTTTAAATCCGGAAAGCGGGAATTCCTTCCGCTCGCTGCCTGCAGCTGCGCTGCGCAGACGCATCGCCGTCCGT
 AAGCCGCTTACCAAAAAGTACGGGTATACCCAAATGGATGCCTGCCATGTATATAGACCATTGGGTGGTATGGACCATGGACCATAAAGC

>CG12443 broad lateral neurogenic ectoderm

GGACAGGACAATGGAGGTTTCCAGAAGAAGCGAGCAAAATGCTGGAAAAATGCAGTGACAACAGGTGCAAAATTAATTTTTTGTGTTGTTGCGAGTGGCGTGAAA
 ATTTCCAGCTGGCCAGGGACAGGAATATGACCACCTAAGGCCTAATGTGCGAAAAGTTCCTTTGTCAATTTACACGCCCTCTCCTACCAAGCGACCCGTG
 AAAACTTCATTCATTTGCACTGGCTAAGCTCAGGTAGCCGGGATTAATCCCTCGTCTAACCAAAAACCTCCTGCTACATTCGGGTTTATCCCACTGTTTTG
 GCTGGGAATTTCCCGCGAGATTTACGGTGGTCAGCCAAATTC

>BRK neurogenic ectoderm

CTCTGGCACAACCCCTAAATGTGGATTACGCTAATATTGCCCCCTAATAAAAAACGGTCTGTGTCAGGGCCGAATATTGCGTCTGATTGGTTTTTCCC
 ACGATTACAATTAGCCGGACGGACACAACTGACCTGAGCTGACCCGCAAAAAGACACGGTTGTCCGGCAGTCGGAACCTGAAGGAACTAAAGGAACTG
 AGGGCAGGTCAGCGTATGGATTGTGCACCTAAGTTGCTTAAATCCGACGGGAAATCCAAAACACAACCCGAGCCCGATCCTTCGCTCCTTCGATTTAAGCC
 AAAGTTAGAGGCACAGGCACACATGTGTGTTGGTTTGAACGGGAAAGCCCATTTTAAAGCTGGCCAACCAACGGCAACACATGTTCATGTTAGGACCC
 ATACAGGTTGACATTCCTTGAAGGATGCACCTCTGGGAGATTCCACAAACCGGCAGCAGGTCATGTCCAACCGATCGTTGCGGGAGCCACTTGTCCC

>RHO NEE neurogenic ectoderm

AAATGGGAAAACATCGCGTGGGAAAACACACATCGCGAAACATTTGGCGCAACTTGGGAAAGACAAGTGGCGTGCACAAAAAGTTCGCGAAACGAAAC
 TCTGGGAAAGCGGAAAAGGACACCTTGTCTGTGCGCGGGGAAAGCGCAAGTGGCGGGCGGAATTTCTGATTGCGGATGCCATGAGGCACTCGCATATGTTG
 AGCACATGTTTGGGGGAAATTCGCGGGGACGGGCCAGGAATCAACGTCTGTCTGCGTGGGAAAAGCCACGTCCTACCCAGCCCACTCGGTTAC

>VND(348) neurogenic ectoderm

GACCTTGGCTTAGAAAATCCCGTAGGTGAGAGCCAGGGAAACCCCAATCGGGAATGACATGTGTACGACAGACATGGGACTCAGATGCCTTCGAGATACT
 GCGTACACTGTCTGGCAATGGGATTTCCGCTCAGGAGGACGGGGAATCCCGGTGTAGCCTGTCCATAGCGTGGGAAATTCGCGAGTCGGGGTCTTCGG
 GAAAACCGAAATGGGAAAACCGGAAGCAAGCAAACTTGGCCCAACATGTGGCAGCAGCTGTTTCGACCCGTAAGAGTCCCTGCTGACCTGTGCTGACC
 TGCACTGACCCGACAGGTAGCTGCGATCCTTACGAGGCGGATTTGCG

>VND(743) neurogenic ectoderm

GTAGGTGAGAGCCAGGAAAACCCCAATCGGGAATGACATGTGTACGACAGACATGGGACTCAGATGCCTTCGAGATACTGGCGTCACTGTCTGGCAAT
 GGGATTTCCGCTCAGGAGGACGGGGAATGCCCGTGTAGCCTGTCCATAGCGTGGGAAATTCGCGAGTCGGGGTCTTCGGGAAAACCTGAAATGGGAAAAC
 CGGAAGCAAGCAAACTTGCGCCAACATGTGGCAGCAGCTGTTTCGACCCGTAAGAGTCCCTGCTGACCTGTGCTGACCTGACCTGACCCGACAGGTAG
 CTGCGATCCTTACGAGGCGGATTTGCGTTAATTGTTGATGGTATTAGGCAAAATCAAACTCGGGGTCTGACCGGGACTAGGTGCAATAATCCAGCGAT
 TTGGGTGCACTTATTCAAAGTTAATTCGCGGGGAAATGTGCGCGTPTTCGGTTCCGAAGCATGCCTGCAGGATGCACACCCCCACCTCCTTATCTCTT
 AACAAACGGCAAGTGCAAAATCTGTGAAAGTCAGAGCGCTACAGGTAGTGCAGGTAGTTTCTTTGCATATCCCGACCAACAGGACCTCCTTTTGTAA
 ACCTTCCGGCATTTCACACGATTGACACAGGATGTCGCTGCAATAAGCATGAAACAGGGAAAATCGTTCCACGTCCTTAAGGAGCCATCTTTTACTC
 GGGGAGTCATTAATCACTCGATTATATGACAATTTGCCGGGAC

>VN neurogenic ectoderm

TTATTGAAAGTGGCGAAGTTAGCGGGCATTTCACTTACCTGCGTGGGAAAATCGACTAATCTGCGACCGCCCGAGGAGTCAGTTTTTGTTTTTAGAGCG
 GTAAAGGACAGGTAACGGGCCACATGTCTGGCCGGAAATTCGCCGTGACCCCTGACCCCGTGTCTTATGACGAATTCGTCACCTGGCGTGAGCACACC
 TGGATTTCCACCGCTTAGCCAGCGGAAATTCAAAACACTCCGGCCACTGGCCCTCAAAATGTTATATGCTCTGCTACGATGAAGCAGAAGCAGAA
 GCAGCAGTGTTTTATTGGCGGAAGCATCCGCCAAATTCACCCAACTCTGCAGTTTGAAGTGTCTCAAAACCCCAACCGCTCCCTGTGAATTTCCGCGGG
 CGCAAGGTGACCGTGTGCTAAAACAAAATTTTTATATCGAAATTCGCCCGGTCACCGCGCGCCTGCCAATGGCCACTTTAACACGTTTTAG

>SIM mesectoderm

CCCCGGCATATGTTACGCACATTTACAGCGTATGGCGATTTTCCGCTTTCCACGGCCACGGCCACAGCTTCCCACCTGATAGGACAGCTCGGCAATGTGT
 GGGAAATCGCAGTGAGGTGCCGGTAGGAGTGGCAGGTAAGCCTGGCCGCCCTCGCAAGTTTCTCACACTTCCAGGACATGTGCTGCTTTTTTGGCCGTTTTT
 CCCCAGCTGGTTATCAATTGGCCGATTGGAAATCCCCGATGGCGATGCGCTAGCGTGAGAACATGAGCTGCGAGCATCGGGTTTTAGCATATCCATAC
 CTGTGGCTCGTCTGATGGGAAGCGAGAAGCAGCAGGATCGGATGTAGGATGCAGGATATAGGGTATAGGCGCTGTTGCGCCTCACCCGCAACCCCACA
 TTAGCATCGGACCAGCGTCCAGTGTCTGTTAATTTGCTTTATGGACTCTCCACTTTCCGCTGCGTGGGAATCTTTGCTCATCTACCTGTTTCCATGCCA
 CACCAACCCATTCCCACAGCATTTGCTCCTTATGTGAAACTCTCTAGTTCAAGTTCAGTGTGAATATTTGTGTTGACTTTATTTTTAAACTTTTTGGCCA
 TTTGTTTTTTCAGTTTGCTGTTTGCCGTGAACC

>A.GAM. SIM mesectoderm

ACTGTTTACCAGAGTGAAAGGTCCTAACGAGTGTTTTAGTGCGGTGCAATAAAATCAGGTACGGCACCGATACCGAGCACGTTTGAAGGGTGGGAAATTT
 GGAACCGGCCCGGTTTCATGGCGGTTGATGATAGGCATAAGCGTGGGAATGAATGAACTTCGGTAAGGATTTTCCCGAAACGGCTCGATGCTGTGGAGCAG
 CAAATTTCTCAGTTTCTCAGCCTCGGTACACACATTTCCCGATGGGTTGCAGTGTAGAGACAAACATTTTTATGCAAAGTAGCATCTGGCGGAAGCA
 ACCTGCGAGGCATAAGATGGTGCCACGCCCCACTACCTGTCCCGCGGTTGTTTGGTATGGTTTCGAGCGTTAATACCTCTTCAAATATGGTGAACACAT
 GGTGACAAGCTGCTTACGCTTTTGACACAACACTAGCCAAGTACCCTTAGTTTTGTGTTTTATTGTAAGTACTGCCATGAGATTAGTCGTTGATTTTTATCAAAC
 AGCCTATCCTAAGGAAGGCTTATCTGGGAAGTTGTGTACGTATTTATAATCCGATCACTGCCCTTTCTCAACACCGATCAATTTGATGCTGATGCAT
 TTGTACACTCAACTATGTCATGTTTGCATAAAATACGCCATAAAATGCCTGCAAAAATGTGCACAGTGTGCATGATGGGAAGGCAACAGTTTCCATTTT
 CTTTTCGTCCATCGTATTTTTCTGCTTTCTTTCTCCTGTACCCTGACGACCCGCCATATTTTACAACCGACACCGATCGTGTGTCATTTTACCTGC
 AATACCCATCCCAGGCCACGCCGAGACCTGTGACCTGTTGTGCGCGTCAAACAATCGCAAAATGCAATTACAGTTGAGCTATTGGACGAGCAAGGTGCAA
 CGATACGGTTGGGGCAGTGATGTCAGTATTTCCCAACACACACGCTTCTGCGCAGTCGATCGTGGGGCGCGAT

>PHM mesoderm

GCTGCTCCTGCTTATCGGAGTGATCAGTGTGGATGGCCTTGTGAAAGAGGGGGATTACCAAAACTCCCTTTATCAACAGAATCTCGAGTCGAACTCCGCA
 ACAGGCGCAACGGCTTCGTTTCCATTCCTGATGCCAACGTTTCGCCCCAGACCGTAAGTCCAGGTTCCGTGAAAAACGCCATACATATACATCATGGAGT
 TCGCTCCATATGGATGAACAGATTGCCCCAGCTGCAGCTGTGCGCATACAATTCGAATCAGATTAGAAAAGTCCCTCGATATAGGCTCTGGTTTTTCCCGT
 ATCTTCCGCGGGGAAAACCCATAAGAGTACTGTGCAAAATTTTCCATATGTTTCAGTTGTACCCGAATACACACAATTCGCCACCCACATGTAAGGCC
 TGAAACCGAAACCTGTAGATTGCACCTCTCGAACCTGCGCCGT

>ADY mesoderm

CCAAGGTGCAATTTTGTATGCAGTGGCGCTATGCTGGGGGATTTCCCATTTGGCTGCGATTGGCTGCAGGGAAAACCCAGTTTCCAAAAGAGTTCTCTGC
 GAAAAACCCCAATGCAGTGCAGTTTATTTTTTCGATGCGTGGTTTTTCCCACTCCAACCCAGATACCATAATGTATTCTTTATTTATCGTGGGGCTTT
 TGGCAAGGGAGTTTCGG

Section III. False positive cluster/control sequences

False-positive Dorsal clusters of 3 or more occurrences of GGGW(D)WWWCC in 400 bp, which failed to display enhancer activity in the early *Drosophila* embryo

>FP-C1

CGCGCTACCAAGCCCCATAGCCACCACCACCCCTCTGGAATCACCCACTTTCGGGCCAGATTGCATCCACTTTAAAATCCATTTATGACTGAAA
AGCAACAAAGGGTGGCGCAAAAAGTTCGGATTGGACTTTCAGCTTTGGCTTCTGGGGTTTTCCAGCTCTTTGGGATCGAGGGTGGTTTCACCCCC
GGGGGATGGGGGTGGGTGGTTCGGGATTTCCCGTCTGCTCATTTACGCATGCAAAGGAATGGGAGTCCCGGTGCTGAAACCCCTTCGTAATG
AGTATTCAAAATGCAATTCACAGACTAGGGGAGTGTGTGG

>FP-C2

ACTGGAATCATGTCACGTGCCTGGCGTGCCGAAAAACAAAAAATAAATAAAAAGGGGAAAAAACAATAAATGCGTCAACAGTGGCTTAATT
ATTTTTTATTCAGTTTATGTTGCCAGTGCAGCTCCTTTTGGCAGTGTCCGCTGGCATTCCCTGGCTTCTTTGTTGGTTTTTCCCGTTGTGCTTCCCTG
TTTATGGGATGCTCGTGTCTACCAGGATCTCTTCTAAGTATATGTTTTGGGATTACCAGGGGGATTACCATCCCTTTTATGAGTGGAGGGACGAGCGT
AGTGCATACCAATGTCCTGCTCCTTTGCTGGACATGTCACCTATTCCGGGCTGAAGGACTGAAGTTGCAACTGCCACACTGTGCTACAACCTGCTGAATTG
TTCCGAAACCAGGAGACCCGAGATGAAGTGCACCTACCATCCGCATTCGCATTTGGAATCGGTATGGGAATC

>FP-C3

GGTTTTCCCGATACCTTTTAGGAAATCCCTTTCAGCTTAACTACGAAACATTTGAAAACATTCCAATCCCTGAATCAAAATGAAAAGCCATTTTTCTTTT
AATTAAGTTTCCATAATATTTCTTAGATTCCTAACCTTTTAGGATTCACCTTAATGAACATAAACGCCCTATTTAATTTGGATTACCCCTTCAAAAA
TGAACCTCCATACAGTTGGCCAGCAGGCCAGCAGAAATCGAAAATTAATGATTTATAAAATGCGACCGTGAATTGCCTCACATACTGAGAATCTAAAT
AAAATATCAGACCAAAACAATGGAACCAACCCGCAATTAATATTTTCATTTTAACTTATCTCCGAGCATTAGCGATAACAACCTATTTGTGGTTCT
TAGCTGGGGGGAAATCCAAGTGTCTGGCACTTTCATCAATAATCTCTTTCTAATCAATCAAAACAATCGTTGAGTAACCCCAACGATTTCGATTTAGA
ATAGCTTATCACGA

>FP-C4

CAGCTTAGAACAAGGAATTTCCCTCTAGCAAAATATGGTTTCCACTCAAATGCCTTTAAATCGGTTGAAACGCGCTCAAAGGAATTTCCCTCCAATTGG
ATTCAATGGGCATGACATAACCGGCTGTGAGTCCGCATGTGCGTGTGTGCGTGTGTTGGAGTACAAGTGCTTCCATTTATTTATTTATTTATTTTTT
TTTTGGCTGTGGCCAGTGACCCCTTGGTTAACGGTATCCATAGTATATGTAAGTTGCCGCGGATTTCCCTTACTTCACTGATACTTTCTTATACTC
CATGATGATCTGGTTTTCAATCTACCCGAGTCTTTGGCTATTTCCATTTATTTGATGTTAGAATCATTCAACTTATGTTATATATGATGATCTC
TGATGGAGCTTTATTTAAATAGTTTTTTTTGTTGTATCGTCAAGTTTGGGGAGGAACACTTAACTATGTGGCGCCATACCCACATATATCTTCCCTCA
AGTGCC

>FP-C5

CTTTGAAGCCACCGACCCATTTGTCCCGCCAGAATGTATTTAATGCAAGCGCGCTCAAATAGGATTTGTGCATTGTTTTCTGTGGTGGATTTAATG
CATGGCCCGGGAAAAACCATTTCTCGCCACTCCCGGGTTTGTGATTTTTCTCAGCCACCATTCGGAATCTATTTTAAATGTCAGTTTAATTC
AATGCCCGAGCGGTTCTTAATGCGCTGCGGGAGCTTTGAGGGTTCCTTGTGCGGCAAAATGTAATTAATTTGGCCAAAGTGGGATAATCCCATCG
TTCTTTTCGATATGGGTACTTAGGCATATATGGGACTATGTGATATGGTTTTAGAAAGCAGCATTCGCGTTTAAAATCACACAATGAAATGCGAAGAATG
CAATCAACGCAATTTACAGTGTACTTGATTTGATGCGACCTAGTTTATAGAAAGTAAGCTATTTAATTAAGTGGAAAAACCCCTCTTTTCGTTCCATCTAA
ATACCAATTTAAGGAATTTCAACTCAATCGCTCACTGGGACTCAAAGTCAACAGCTTTCAACCGTCTCTGAAATCTGAAGCAGTCAAAGGCGCTT
GGCTTATTTCCGTTTCCGCTGCTGCCATATTTCCGAACCCCTCCGAAAGCAGCTCCCCCT

>FP-C6

GACCAAGCTAATCACAAAAGGCTGAGAGAATGGATTTCCCATCCTTTTTGTTTGTTCGTTTTCCGCTTTTCCCTGATGACATAAATTCGGAGCGCTCGA
ATTTCTGCGCCTGACTTTACACACTCTGACATAATTAGGGTGAACACCCCTGCCACTGGCCATTTAAGTCCCTATCCGTGAACCCCTTTAAAATGG
CATCGACACACCCCTGAACACTTGTGTGCGCAGCTGGCTGTCTTCAATTTAAAAATCAAATGCCATGCGCTGCAATTGTTAATGAAAACCTCAGTTTTTT
GTGTAGAGTGAATATAAAATATGCAAAATTCGCCACGCTTTTCAATGAAGTAAACAAGGTGGGAAAGCTGAGTTTGTATTAAATAAGAAAGGATGTC
ATACTTTAATAAATGAATGTATTCCAATGCCATTTGGAATAGAAAAAGAAACGAAAGTGCACAACGACTTTTACACAAAAACTATTTTATTAGCATTGC
AAGTGTGCGCCCTAACATTTAAAATTCAAAATCTGCATACACTTTAACTGTTCAAAAATATATAACATTTTCTCCTTGGTTATTGCGGAGTCTTTGTCAT
TCGCTCTTGAATAAGGTATCGCATTATCTGACAAGACTTACCCCTCACCTCAACCGTAGTAAAGTTGGGGTATTCCCAAGAAGAAGTACATTTATTACG
GCTCCATTTCCCTTTTGGAGAAGAGTCCCTGTCTGCTGATAGAAATGCAAGTCAAGTAAAGTCCCTGAAAAATCCACAAGGATTTGCTCGGTGATAGCTGAAAGC
TTTCTCCCATTTCCCAACATCGAGGTTTGGATTTCCCTTTCCGGGATTTTCCACTCGTAGTTGATTTGTTGTCAAAGGTCAAGAAATGCAAAACAATATA
TAGCACATTTGTTTCCGATCCGATACGATTAGCTAATCGAAGCGATAAACTCAGTAGAAAAATCACGGCTGACCTGCTTTCCATTTTAAATGGCATT
GTAGGTCCTCGCTATTAGTAGATTAAAGTTTCGCAGGCTTTTGTGCAATGCAAGTGTGACTTTTCAATTTGCTGATTTTTGGCAATGGAAGTAG
CTACGGTTAAATGCTAATCACAAAATTTGCCATTTGATTGATAAACTGGCAGACACCCGCTCTCAATCGAAAGGACAAGCCACAAAAACCCGAATCGGAAG
CGTTTCCATAGCTCGACC

>FP-C7

CCTGACCGACGGGAAGGAAAAGTTTGCCGGCACTTTTCACGCTTTCTATGGAACTTTGGCGCTTCCGCTGCCACCTCTCGATTTTCATTTGTCAGGCG
ATTGTGTTTTCGACCATTTTCGAGCATTTCCACAAAAGTTCGCACAAAAGCTAAATGTTTATTTGTACTGCAGTTGGACTGCCCTCCGCTTTGTGTTGTT
TGATCGTTTTTTCCGCTTAGTTTTTTGTAGTTTTTTTTTTGCGTTTTGTTCAACTTTTCCGGTTTTTTCCCTGTGCGTCCGCGTGAATCGCATTTCTC
TTGGCCAGGAATCATTAGGAACTATTTGCTAAGCAAATGCCACTGGCATTTGCGGAGTGGGCTGAAATTTGTGAGCCGGAATGCTGGACCGCAAAAAA
AACTGAATGGCAATGGGAAAACCAGGCCAGGAGCAAAGCAACCTTGTGGAAAATGGCGCTGCGAATTTGGTATTTGTGGTTTTCCCTCATTTTCCCATA
TTCCCCAGCGGCGTGCCTCCGTTGCTGATTTTTCCCTGTCCCTGCGAAGTGTGATCATCATTTGCCGATTAACAGCCCCATTGGCACACTCACT
TGGCAAATGCACATTTGACTGGCTAATTACGAGGCTGTCAACAAATTTGGCATAAAAAACAACGCCATCTCATCTCAACTCATCTCATTTGATC

>FP-C8

CCACGGAGACAAAAGACAAAAGACAGAGAGTCCAGCGGACGAAGAAGCGTGCCATGACAATCTTGGGTCAACGGCTGCTTTAACCCCTTTCATGGGCAAGAGG
GTTCCCATCACACCCTCACTCACTCAAGAGGCTATGCAGACAAAAATAAATCCTAAGATCATGACAATCATTAAAGCTTTAATGATGACATGACATAC
AAATTTATGATAAGTTATTTACTTAAACTGCATAATCCAGTACCTATAGTCTGTTAAAATTTACGACCTGAAATGAAGCCAAAAAGTGATGCTCCCCA
AGGGTTAACTGGAATCTCCGATCTGAACCTATCTGGCCTGTTAATTTGATGTTACATTTCCCTTTGTGCGAGCGAATGGCTGTTTATGTTGGTGGCTGTGT
CCGGACAGCTTGCCTCACACTCCGTTGGGATTAACCCCAACACGGATTAACCTCAGAAGCTCCACTTCGCGCAGCTGCGATGCAGACGTTGTTTTTTATG
ATTGCCATTTTGTACAATGAATAAAAAATATATATAGTGTGCGACGTTTTGGTGGCTGTAATTTAATTAACACGATGTCGCCACGACCCGCATCACGATC
GCGCGTGTCCATGAGGGTTTTCCCGGAATTCAGCACTCCCCCCCCACCCCTGGCGAGCTTCATACAGGTGAGAAAATGGTGACAGTTTCGGAAAAAGT
GACTAATATATAAAGTACGATTTGTGCGTGCAGGAAAAATCCAGCTGGGAAAAAGGATGCGAATGCGGATGGGAAAAACAACGCAGCACTTGAGAAGT
CAGGACTTTTTCGAATATTTGCGCTTGTCTTTGGGAGAACCGAAGTGGCGCAAAATACAGCTGGATTCGTTAGGCAACAAAATCCGAATAAAATCATTTT
TGCCCTTCGTTAATGCAACTGCCCTCGGTGGCCAGATGGCTGAGGAAAAAGTGGAAAGTGGACGAGGACAAGGATGAGGATGAGCAGCCGGGAAAAACTT
GTGCGGCTTCTGTGCGTGTGAGCTGTGAACCTGTTTACACCAAACCTCGGAAAGAGAAAAGAAAAACGAACCGACTGAAACCATGAATGAACTGCAACTTG
GACTTGGGCGGGCATTTGCTACTGTGATACAGATATATAGAAAATGCACTCAAGTAACCTTCTACTTTGGCCGATCCGATTTTCATCCCAACCTCTTTTC
CCCTCACAGATTTATGCACCGCCGGCCGACCCTCCACACCTCCGCGCTTAGCC

>FP-C9

CCCCATGCATTTTACCATCCTTTACCATCGGCTTAACTTTAGTTTTGCGGCTTTTGGTCAATTTTTCAGCTTCAACTTCTTCTTCTTCTTCTTTGGGGT
TTTCCCCCTTGTTTTTCCGGTTTTTCGGCATTTCTTCTTCTTCTTTTTTCCGTTTTTCGGTCCGTTTTTCGGCAAAGTTAATTAATAATGCATTAGTT
GATTTCCCTAAAGGGATAAACAGTTGGCCGGTGTCTGGGCAAACTCCTTGTGCTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGCTTTGTGGG
TGTGTTCTGTACATGTGTATGTGTGTTTTTTGATGTGTGTGATTTACATGTGGGACTCATAAATGTTTTGGGATACTTACATTACTATTTGCTTGAA
ATTCGTTTCGCACTTTTCGATTTTCAGTGGGAAAAATCCCGTACGCATTTTCACTTTTCACATAAGAAAATTTTCCAATATGGGGAATGAAATGCCATTTAT
TTGTGCTACACATGGTTGGCAAGATTGCCACAA

>FP-C10

CCCAATGGATAAGGCGTCGGACTTCGGATCCGAAGATTGCAGGTTTCAGTCTGTACCGGTCGAAGCTCAGGCTACATTTTTTTTTAAATATATATTTTGT
CGTCTAGAAATATATTAATATGGGAGATTCCCTAGCCCAACCCATTTGTGTAACCTGAGAAAATGGGAATTTGGGACGGCGGTTGCGTAACTGACCGTGT
GGCCCAATGGATAAGGCGTCGGACTTCGGATCCGAAGATTGCAGGTTTCAGTCTGTACCGGTCGAAGCTCAGGCTATATTTTTTTTTAAATATATATTTG
TTTCGCTCTAGAAATATATTTAATATGGGAGATTCCCTAGCCCAACCCATTTGTGTAACCTGAGAAAATGGGAATTTGGGACGGGGTTACGTAACCGACCGT
GTGGCCCAATGGATAAGGCGTCGGACTTCGGATCCGAAGATTGCAGGTTTCAGTCTGTACCGGTCGTACCTCAGTATTTAATTTTTTTTGAACCTATTTT
TTTCGTTTCGTTCTATAATATATTAATATGGGAGATTCCCTAGCCCAACCCATTTGTGTAACCTGAGTGTGCGGTAAGCAGCAATCGTAACCAATTGGCATAAC
GAATGAAAGATTTATTTGACTTTTACATGGGTCGTTCCATGGACGAATCAACATGTTGGTGTCCACCGCAAGAAGCCCAACTTTGTTCGTTGGCTCTTGCT
GCCT

Section IV. Bioinformatics

OpenGenomics software, Fly Enhancer (www.flyenhancer.org) and Mosquito Enhancer (www.mosquitoenhancer.org), were used to conduct queries for clustering of specified DNA sequence motifs of the *Drosophila melanogaster* and *Anopheles gambiae* genomes, respectively. Diagrams shown in Figure 1 are adapted from Fly Enhancer output. The OpenGenomics MERmaid program (freely available for download at www.opengenomics.org) identifies n-mers of any length that are present or absent in specified groups of sequences. In this study, we considered two classes of motifs: “exact match” motifs in which every position in the motif is filled by one specific nucleotide (A, C, T, or G), and “fuzzy” motifs in which up to two positions in the motif can be occupied by any of the four nucleotides (denoted by the IUPAC symbol “N”).

Section IVa. Results of MERmaid analysis

Table S1. Frequency of shared motifs found in comparisons of *brk*, *vnd*, and *rho* enhancers

nMER	Condition 1			Condition 2			Condition 3		
	A	B	C	A	B	C	A	B	C
6	157	731	0	16	536	0	17	578	0
7	75	2711	38	8	778	2	8	588	3
8	40	3725	1434	4	435	102	1	218	48
9	18	1979	1411	1	141	68	0	61	48
10	10	809	672	0	35	22	–	8	5
11	3	311	281	–	8	7	–	2	2
12	1	118	114	–	1	1	–	1	1
13	0	47	47	–	0	0	–	0	0
14	–	26	26						
15	–	15	15						
16	–	9	9						
17	–	7	7						
18	–	4	4						
19	–	2	2						
20	–	1	1						
21	–	0	0						

Beginning with 6-mers, this table shows the frequencies of all n-mer motifs that meet the following conditions: Condition 1 – motifs occur at least once in any pair of enhancers for *brk*, *vnd* and *rho*; Condition 2 – motifs occur at least once in each of the enhancers for *brk*, *vnd*, and *rho*; and Condition 3 – motifs occur at least twice in each enhancer or any pair of enhancers for *brk*, *vnd* and *rho*.

For each condition, three types of motifs are listed: (A) Exact motifs (non-degenerate); (B) Fuzzy motifs which can have up to two wildcard or “N” positions; and (C) Fuzzy motifs from column B that do not occur three or more times in the 10 false-positive sequences. The numbers in this Table represent all motifs that meet the specified conditions. Several of these motifs are related by shifts and sequence. This table represents raw numbers of motifs, without attempts at alignment.

Table S2. Motif-classes shared between *brk*, *vnd*, and *rho*

class	feature	aligned motifs	Tot	brk	vnd	rho	FP	genome
1	Extended E-box(CA)	CNCATGTNT	6	2	1	2	0	12583
		CACATGT	10	3	2	1	3	16384
2	Match to Su(H)	CGTGGGAAAWNC	3	1	1	1	0	149
		CGTGGGAAA	3	1	1	1	0	910
3	Match to Dorsal	TGGGAAAWWC	5	1	1	1	0	2622
		CGGGAAWTYC	4	1	1	1	0	1004
		CNGGNAAT	15	2	2	2	0	60582
		CGGGAA	18	2	3	2	6	40668
		GGGCWTTYCC	3	1	1	1	0	1118
		GGGAAAT	19	1	1	1	9	28614
		GGGAAT	28	2	3	1	1	258514
		GGAAWGNC	5	2	2	0	1	26120
		GAAATTC	11	0	2	2	5	29259
4	Dorsal Half-sites	CKNCGGGAA	7	1	2	1	1	4352
		AAATGGG	9	1	1	1	5	30229
		AAYGGGAAARC	3	1	1	1	0	708
		GACGGG	8	1	1	1	1	21962
		GACGGGVMA	5	1	1	1	1	3470
		CCAGGG	8	2	2	0	2	33813
5	Clustered Motifs	GCTGACCYGY	5	2	2	0	1	465
		CTGNCCY	16	4	3	2	9	82829
		CTGWCCY	9	3	3	1	2	33022
6	Palindromic	TTRCGCYAA	3	1	1	1	0	2987
		RANTCGCGA	4	0	2	2	0	4623
7	Other	CCTGWWTCG	3	1	1	1	0	1076
		GNCWTTCCC	4	2	2	0	0	6733
		AAANGAMAC	5	2	0	2	1	21169
		CATTTNGNG	4	1	1	1	1	24568
		ANTTSCGC	9	0	2	3	1	30178
		CGGNNGGA	7	2	0	3	1	35770
		CTGTGC	10	1	1	1	2	55970
		CNGNAGG	11	2	3	0	2	161358

The motifs shown in this table are representative of computer-assisted and manually adjusted alignments of *all* the motifs meeting Conditions 2 or 3 in Table S1. The motifs were manually grouped into seven classes. Motifs in classes 1–3 match consensus sequences for the known transcription factor binding sites Twist, (CACATGT E-box), Su(H), and Dorsal, respectively. Motifs in class 4 include portions of Dorsal half-sites; motifs in class 5 are related to the previously identified CTGNCCY (Stathopoulos et al., 2002) motif. This motif is tightly clustered in the *brk* and *vnd* enhancers. Motifs in class 6 contain partial symmetry; and the remaining motifs meet conditions 2 or 3 but do not have obvious distinguishing features. Within each class, representative motifs are shown. The frequency of each motif in the enhancers for *brk*, *vnd*, *rho*, and the false-positive clusters are shown as a sum total and for each enhancer separately. In addition the frequency of each motif in the *D. melanogaster* genome is shown.

Section IVb. Parameters used for *Drosophila melanogaster* genome query

Using Fly Enhancer, the *D. melanogaster* genome was scanned for composite clusters of Dorsal, Twist, Su(H) and CTGWCCY motifs using the parameters below

Input:

Site A:	GGGWWWCCC	Site F:	CACATGTH
Site B:	GGGWWWCCA	Site G:	CACATGTG
Site C:	GGGWWWWCYS	Site H:	
Site D:	GGGWDWWCCM	Site I:	YGTGDGAA
Site E:		Site J:	CTGWCCY

Chromosome Arm:

Cluster Size: at least binding sites

Window Size: within base pairs

Boolean Condition:

Output Summary (for graphics, enter input at www.flyenhancer.org):

Chromosome arm: 2L 2R 3L 3R 4 X all

```

-----
site A (GGGWWWCCC):  447  401  466  566  15  441  2336
site B (GGGWWWCCA):   693  645  763  872  29  663  3665
site C (GGGWWWWCYS): 1919 1664 1993 2217  47 1769  9609
site D (GGGWDWWCCM):  734  618  732  795  20  665  3564
site E (WWWWWWWWCCC): 4633 3696 4888 5514 337 4452 23520
site F (CACATGTH): 2097 1876 2174 2761 100 2020 11028
site G (CACATGTG):  464  444  537  664  19  550  2678
site H (YGTGDGAA): 4039 3618 4264 4989 173 3859 20942
site J (CTGWCCY):   6217 5988 6466 7901 253 6197 33022
-----

```

```

-----
Clusters of at least
5 sites in 250 bases:                58
-----

```

```

-----
Clusters satisfying
1(ABCDE) and 1(FG) and 1(H) and 1(J):                12
-----

```

```

-----
After merging overlapping clusters:                7
-----

```