Supplementary Table 2

Unique insert sequences from the prey isolated using the B1H system with each bait. Bold letters indicate an overrepresented motif identified by MEME analysis. Italic letters indicate the inclusion of constant bases neighboring the randomized region to complete a binding site consistent with an overrepresented motif identified in the other sequences by MEME analysis. The E-value calculated by MEME for each identified motif is indicated. An asterisk following the sequence indicates that it is the reverse complement of the insert sequence. The absence of bold letters indicates a sequence that does not contain a binding site motif based on computational analysis. Numbers in parenthesis indicate how many times a sequence occurred among the analyzed clones.

Zif268 sites (Two step selection; E-value=3.4e-020)
TTGAG**CGCGTGGGTGG**AT
CAACGT**TGTGTGGGCGT**G(2)
TTGCCC**CGCGTGGGCGT**C
CA**TACGGAAGCGG**CGTGC*
TTGTTG**TGCGTAGGCGG**C
AATCG**AGCGTGGGAGG**GG
ATCGAG**GGCGTGGGTGG**C
TGTAGT**TGCGTGGGAGG**G(2)
CTAACC**CGCGTGGGCGT**A
CGAGGA**TGCGGGGGTGT**G
TTGCCC**CACGTGGGCGT**G
CGATAA**TCCGGGGGCGT**A
GAATG**TGTGTGGGCGG**CT
ATTTGT**AGCGTGGGCGA**T
GAGCGA**AGCGTAGGCGT**C

PLAG1 sites (Motif score = 2.278 calculated by Bioprospector). The Sequence logo for PLAG1 in Figure 2 is the reverse complement of the overrepresented motifs in these sequences. 18 of 21 sequences contain the identified PLAG1 motif.
C**CCCCCGA**AAA**GGGCCCC**(6)
CCCCCGATTTG**GGGCCTC**(3)
TT**CCCCATT**T**GGGCCCC**T(4)
AACA**CCCCTGTGGGCCCC**
CCCCCTAAT**GCACCCC**AT
CCCCCGTGG**AGGCCCC**AC(2)
C**TACCTTT**TTC**TGGCCCC**(2)
AA**CCCCCTA**AG**GAGCCCC**(2)
CCCCCCTCCT**AGGCCCC**C(2)
CCCCCTACCA**GGGCCTC**A(2)
CCCCCTTGG**AGGCCAC**TA
CCC**CCCCCGTGGCCCCC**T
CCC**CACCCCTGGGCCCC**T
CCCCCTAAC**GCGCCTC**CG

A**CCCCCTT**AG**TGCCCCC**C
AC**CCCCATTTACCCCC**CC
GT**TCCCCTT**CC**CGGCCCC**
CCCCCTTCAC**GGGCCCC**C
GTATCCCGGTGTATGCGA
GCGGAACCGGCTAGGAAT
GGGCCCCCAGAATTCCGG

Zif268 sites (One-step selection; E-value=3.0e-06)
TGTAC**TGCGTGGGTGT**GG*(2)
ATGCTTGTCGC**TACGTGG*GCGG***(2)
CAGCC**TGCGTGGGAGG**TT*
TGGACAT**TGCGTAGGCGG***
CCGCG**TGTGTGGGCGG**AA*
TTGCCC**CACGTGGGCGT**G*
CTCGCA**CGTGTGGGCGT**C*
ACTATC**CACGTGGGCGT**G*

ZnFp53 sites (E-value=1.7e-045). The Sequence logo for Znfp53 in Figure 2 is the reverse complement of the overrepresented motifs in these sequences
ACGTGTCCGTCAGACGGG
A**GCGTGTCC**CTCAGTCCG
TG**ACGTGGCC**TGTCACTC
GACGTGGCC**GGATGTCC**G
AGC**TCGTGTCC**ATGATCA
GACGTGCCC**GGCTATCC**C
G**ACATGTCC**ACGCCCTAG
TAA**GCGTGTCC**GTAGCGG
CCCCCGAC**ACGTGTCC**TA*
G**GCGTGTCC**CGCCTGTTA
AC**CCGTGTCC**AGCTTTAG
CC**GCGTGTCC**GTAAAGGG
G**TCGTGTCC**CCATGGTCG
G**TCGTGTCC**TTGCACGGG
AGT**GCGTGTCC**TGGTGCT
A**GCGTGTCC**TGGTCTGCT
G**ACGTGTCC**GGATTTCCC
GGTG**ACGTGTCC**TGGGCA
GCGTGTCCCGCAGCCTGT
A**TCGTGTCC**GCTCACCTG

LAG-1 sites (E-value=1.4e-023). The Sequence logo for LAG-1 in Figure 3 is the reverse complement of the overrepresented motifs in these sequences. 12 of 17 sequences contain the identified LAG-1 motif.
TGTGAGA**ATTCCCACGGT**(2)
GA**GTTCCCACGGT**CGACA*(3)
GGTCAC**TTTCCCACGGT**G(3)
TCAACTT**TTTCCCACTGA***(3)

CC**TTTCCCACGGT**TACCC*(3)
AGTTGTT**TTTCTCACGGT**(2)
CA**GTTCCCACGAT**TCCGC*
GAT**TTTCCCACGTG**TGGC*
CA**TTTCCCACGCT**TCAGA*(2)
TC**ATTCCCACGGT**AGCAG*
CTC**ATTCCCACGGT**CCAG*
TCAACTT**TTTCCCACCTG***
CTTCACGGCTCTGACCGG
GGTATGCAATCGATAGAG
GTCTGACCATGTCCTTCC
ATCCGATGTGTTGATCGC
CCCGCGGACACGAACGGT

Dorsal sites (E-value = 5.5e-058). The Sequence logo for Dorsal in Figure 2 is the reverse complement of the overrepresented motifs in these sequences. 32 of 35 sequences contain the identified Dorsal motif.

CCACCACG**GGGAATTCCC**
AGGCACAGA**GGGAATTCC*C***
CCCACCACG**GGGAATTCC*C***(3)
CTGTGGACA**GGGATTTCC*C***
ATAGCCAGA**GGGATTTCC*C***
CCGGATTCT**GGGATTTCC*C***
TTTCGCGGG**GGGTTTTCC*C***
ATCCGGCG**GGGAAAACCC**
CGCGTACC**GGGAAAACCC**
GGTGCGTCT**GGGAAAACC*C***
TAATCACGG**GGGAAAACC*C***
GGTGCGTCT**GGGAAAACC*C***(2)
GGTGTGGCG**TGGAAAACC*C***
CCACCGGA**CGGAATTCCC**
GCCATGTAC**CGGAATTCC*C***
TCGTATACT**GGGAAAAAC*C***
CGCCTGCGC**TGGAAAAAC*C***
ACCTCGCCT**GGAAAAACC*C***
CGGTGCCT**TGGGATTCCC**
ACCTGCTCT**GGGAATTTC*C***(2)
CCTACCTCG**GAGAATTCC*C***
CAGTGAAC**TGGGAAACCC**
CCTGGTAGC**CGGAAAAAC*C***
CCCTAGC**GGGGTTTCCC**C
TGTTCACCC**GGGGTTTCC*C***
AACCTGGAG**CGGAAAAAC*C***(2)
ACTGCATCA**CGGGATTCC*C***(2)
TGG**TAGAATTCCC**ATGGT*
TCAATCGGA**GGGGAAAAC*C***
ACCCAGCTT**GGGGAAATC*C***
TCGTCACGG**TGGTTTTCC*C***

CATCGGTTT**GGGGTTTCC*C***
GCCATGCTATCGGCCGCG
CACCCCCATCAGAATCCTC
CGAACCGTTAGCAGCTTAC

Paired site (E-value = 3.0e-028). 22 of 32 sequences contain the identified Paired motif.

AATTCGTCACGGTTGATG
GATTCGTCACGCTTGCTT
TG**ATTTCGTTACGGT**CTG
C**GATTAGTCACAGT**GCCA
CT**AATTCGTTCCGCT**CCC(4)
CC**GATTCGTTCCACT**ACC
TT**AATTCTTTCCGCT**CCT
CC**GATTCGTTCCTCT**ACC
AC**AATTCGCTACGGT**GGT
CC**GATTCGACCCAGT**CAC
CC**GATTAGTCTCGGT**TCA
CC**GCTTCGTCACGCT**ATG
CT**GATTTGATACGCT**TCG
CA**TATCCGTCACGGT**GTT
A**AGAACGTCACGGT**TCAC
CT**CGCTCGTCACGGT**TGG
AC**AATTCTCTACTGT**GGT
CACTAGTTACTGTACCGC*
C**GTTTCTATAGGCT**GCGT*
A**AGAACGTCCCGTT**TCAC
AT**AAATCTGACCGGT**GGG*
T**CATACGGCCCCTT**TCTC
GCGGTACAGTATGTCCCA(2)
TGAGTTATTATGGCGGGG
AGTGATTGTATGGTATAC
AACCTGTGCTAGAGTACC
AGTAAAGACCGTGCTGCC
CAAGCTGAGCCCAAAATG
TCTTATTTGTCAGTAGGA
CGGGCGGCTCGGATTAAT(2)
GGAATACTCTGAAAACAC
GCCCCGCTGCAGAATTAA

Runt/Bgb sites (Two step selection; E-value = 1.1e-067) The Sequence logo for Runt in Figure 2 is the reverse complement of the overrepresented motifs in these sequences.

<u>Sequences prior to counterselection</u>
TAGCCGCAGATTGACTGC
GATCTACAGGCGCTATGA

TCTCGGACGCAGAGTACG
AAACTGTCCCACCCCTAC
CACGCTATACTGCAACAC
ACATGTATACTTTGAGAG
AGCTACTCCTGGGAGAAT
ACTGGTATATTAGTGCGG

Sequences after counterselection
C**AAACCGCAA**TCCTCGCC(2)
AAGTTCC**AAACCGCAA**GC
GGTAGCC**TAACCGCAA**GT*
ACT**TAACCGCAA**AGGATG(2)
AA**TAACCGCAA**ACTACCC(3)
A**AAACCACAA**GCGCTCCT
AT**AAACCGCAG**GCTCTAC
CTCGAA**AAACCGCAG**CCT*
CTG**AAACCGCAG**TTTCGG
T**AAACCGCAG**GTTCCAGC
AAGC**TAACCGCAG**AAGTG
C**TAACCGCAG**CGCTATTC
ACCCCAG**TAACCACAA**GT*
CT**AAACCACAG**CTCCCCG
CT**CAACCGCAA**GGACGAC
ACC**CAACCACAA**AAGGTG
ACCGCAA**AAGCCACAA**TC*(2)
AA**CGACCACAA**GTCCTTG

Odd sites

1.5 mM 3-AT (E-value=4.3e-031)
_G_CTACTGGTCTTAGGATCT
CT**GCTACTGTA**ATGGTAAC
_G_CTACCGGTTCCTTACAGA
_G_CTACCGGAAGTCATCGTA
GGATCTGTTTGTTCGGTG
GA**GCTACTGTC**TTGCGGG
_G_CTACTGCTTAGTCTTGGT
_G_CTACCGGACTAGTTACCC
A**GCTTCTGAA**ATTCCAGA
GCTACTGGACCCACCTAG
_G_CTACTGGCGGGGCAACTC
_G_CTACTGGACTGTTGTCGC
GCTACCGATGCTCACTCC
GA**GGATCCGGA**AGCAACG
GGTCCTTCCGT**GCTACTG_GC_**
GCTACTGGTGTCCTTAGT
_G_CTACTTGTCACGGTCCGA
T**GCTACCGAG**CATGAGCG
_G_CTACTGCAAGCGGTGCGTA

ACTACCGTTTAGTTGGCG
G**ACTACTGGA**AGCCTGGT
C**GCTTCCGGA**AACTACCG
G**GGATCCGTT**GGAGCGAT

2.5 mM 3-AT (E-value=9.5e-43)
GCTACCGTTTTACCCGGG
_G_CTACTGGTCTTAGGATCT(2)
T**GCTACTGTG**TAATGCGC
T**GCTACTGGA**AGTTGGTC
C**GCTACTGTT**GTGTGTGG
GCTTCTGTATTGTTGCGA
TG**GCTACTGTC**TGCCACC
TT**GCTACTGTA**CCCGGGG
GCTACCGTGTGACTCGGA
GCTACTGTAGGCGAACAA
GCTACCGTAGACGCTCGG
GCTACCGTGTGTACCGCA
A**GCTACCGGA**GCCCTAAG
GCTATCTGGTGGAAGTTG
GCTACCGGAAGTTCCTAG
GCTACCGTTATGTCGTCG
_G_CTACTGATTCCCCCTTTA
_G_CTACTGATTGCGGAAGCA
GCTTCCGGAAAAGGCGTG
GCTACTGGTGCTGCACCC
G**GCTACTGGT**AACAGAGT
GG**GCTACCGGG**GGAATTC
_G_CTACCGTGAAGTTGTTGT
GCTACCGGACGGTCCCTC
GGATCCGTATGAGTGCCTC

5 mM 3-AT (E-value=3.2e-049)
_G_CTACCGGTTCCTTACAGA
C**GCTACCGAA**TTTAAGGG
GCTACCGTATACTTTCCA
GC**GCTACTGTT**TGGGGTG
_G_CTACTGTTTTCTGTCACC
GCTACTGGTAAAACCAGT
_G_CTACTGTTGCGTGCAAGG
T**GCTACTGGA**CAAAATCG
T**GCTACTGTT**CACTCGGG(3)
T**GCTACTGGA**AGTTGGTC
T**GCTACTGGA**AGGGGTGA
A**GCTACTGTA**TCAGTGGT(3)
_G_CTACTGTTTACGCATGGG
T**GCTACCGTA**GCGACGAG
_G_CTACTGGTCTTAGGATCT

T**GCTACTGTG**CAATGCGG
T**GCTACTGTG**GATGAGCG
T**GCTACTGTG**TAATGCGC(3)
CG**GCTACTGTG**CAGCTAG
GCTACTGTGCAGTTGCTG
CG**GCTACTGTA**TACGTCC
GT**GCTACTGTA**AGCTCAG
GT**GCTACCGTT**TGTTCGG
GGACTAGGCTGTCTACCC
GACACATCCTCGGCGGCT
GGCCTAGGGAAAGTCTGT
GCCCCGAGGAGCATAAAG