PNAS
www.pnas.org

Supplementary Information for

"A Short ORF-encoded Transcriptional Regulator"

Minseob Koh, Insha Ahmad, Yeonjin Ko, Yuxiang Zhang, Thomas F. Martinez, Jolene K. Diedrich, Qian Chu, James J. Moresco, Michael A. Erb, Alan Saghatelian, Peter G. Schultz, and Michael J. Bollong

Email: mbollong@scripps.edu; schultz@scripps.edu

This PDF file includes:

Figures S1 to S4

Supplementary Table S1 to S2

Other supplementary materials for this manuscript include the following:

Supplementary Data Set 1

Supplementary Data Set 2

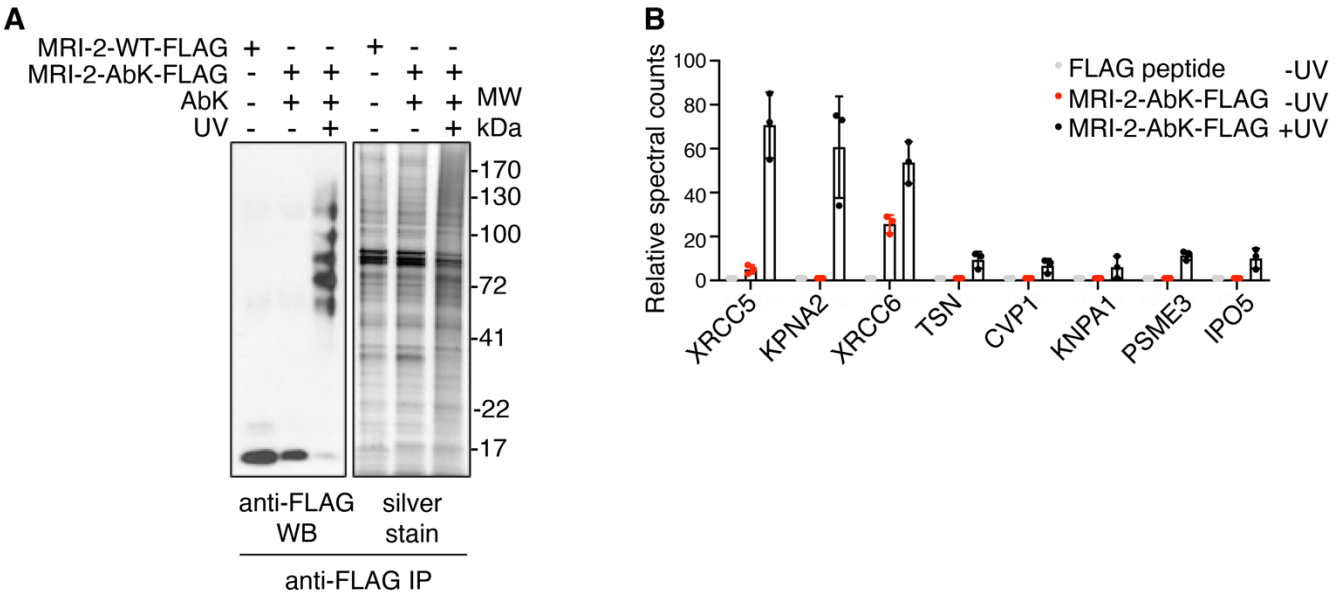

Figure S1. AbK incorporation allows for the identification of reported and novel interactors of MRI-2. A) Anti-FLAG Western blotting for anti-FLAG immunoprecipitated protein content from HEK293T cells after expression of the indicated MRI-2 transgenes in the presence or absence of UV. B) Relative spectral counts corresponding to the most enriched proteins identified by anti-FLAG immunoprecipitation after expression of MRI-2-AbK-FLAG in HEK293T cells for 48 hours with subsequent exposure to UV (*n*=3, mean and s.d.).

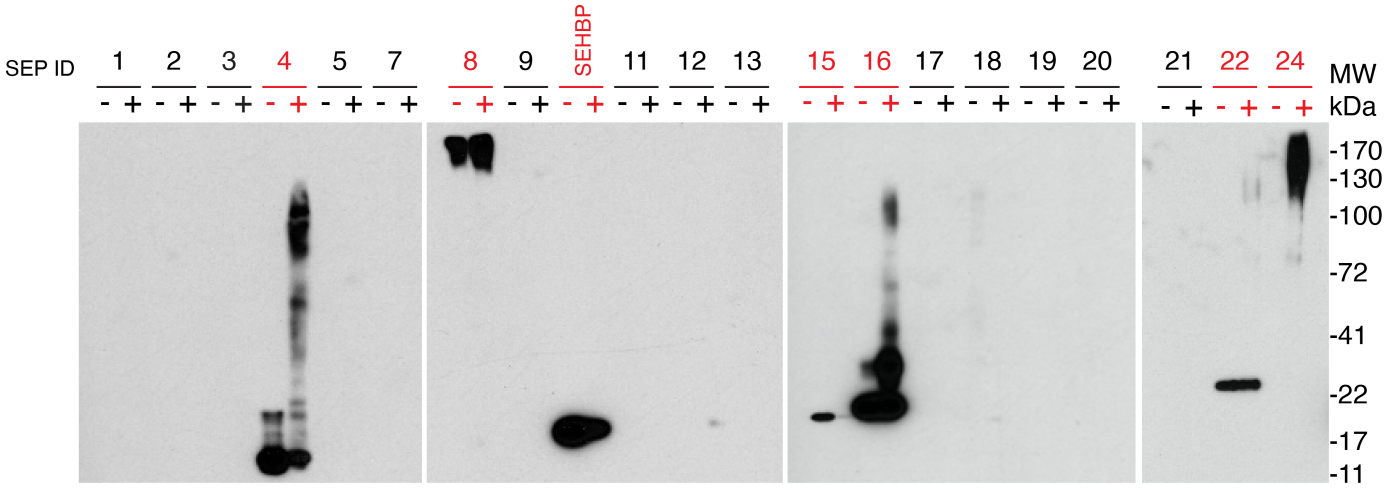

Figure S2. A Western blotting-based screen identifies high-expressing SEPs with crosslinking potential. Representative anti-FLAG Western blots from HEK293T cells expressing the indicated FLAG-tagged SEP transgenes in the presence or absence of UV. Identities of SEPs are listed in Table S1.

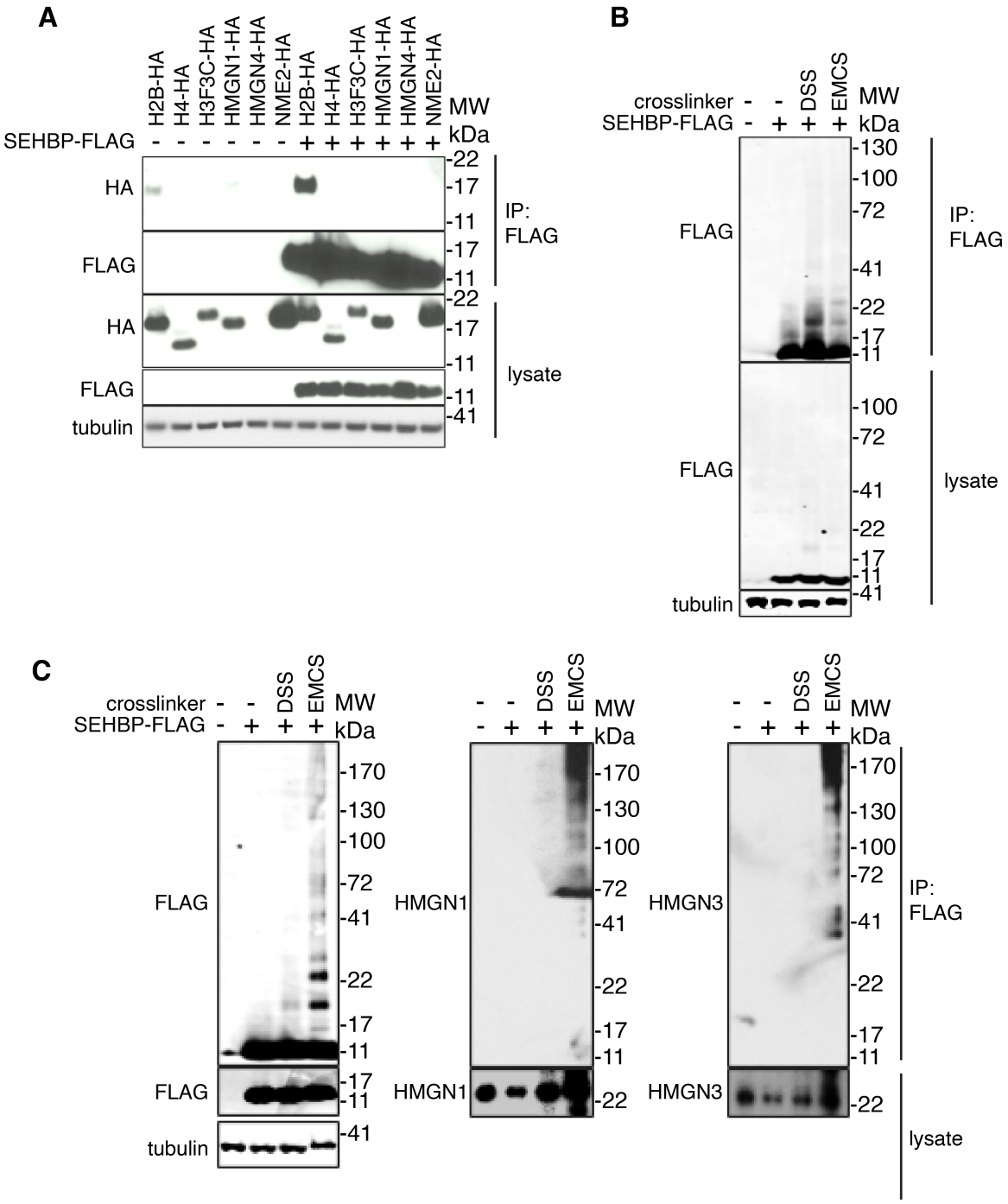

Fig. S3. SEHBP interacts with Histone H2B and HMGN proteins. A) Anti-HA Western blotting analysis after anti-FLAG immunoprecipitated protein content from HEK293T cells expressing SEHBP-FLAG and the indicated HA-tagged transgene preys. B) Anti-FLAG Western blotting analysis of anti-FLAG immunoprecipitated protein content after 48-hour exposure of SEHBP-FLAG and subsequent 1-hour treatment with the indicated chemical crosslinkers (1 mM). C) Western blotting analysis for endogenous HMGN1 and HMGN3 from HEK293T cells expressing SEHBP-FLAG for 48-hours and then exposed to the indicated chemical crosslinkers for 1 hour (1 mM).

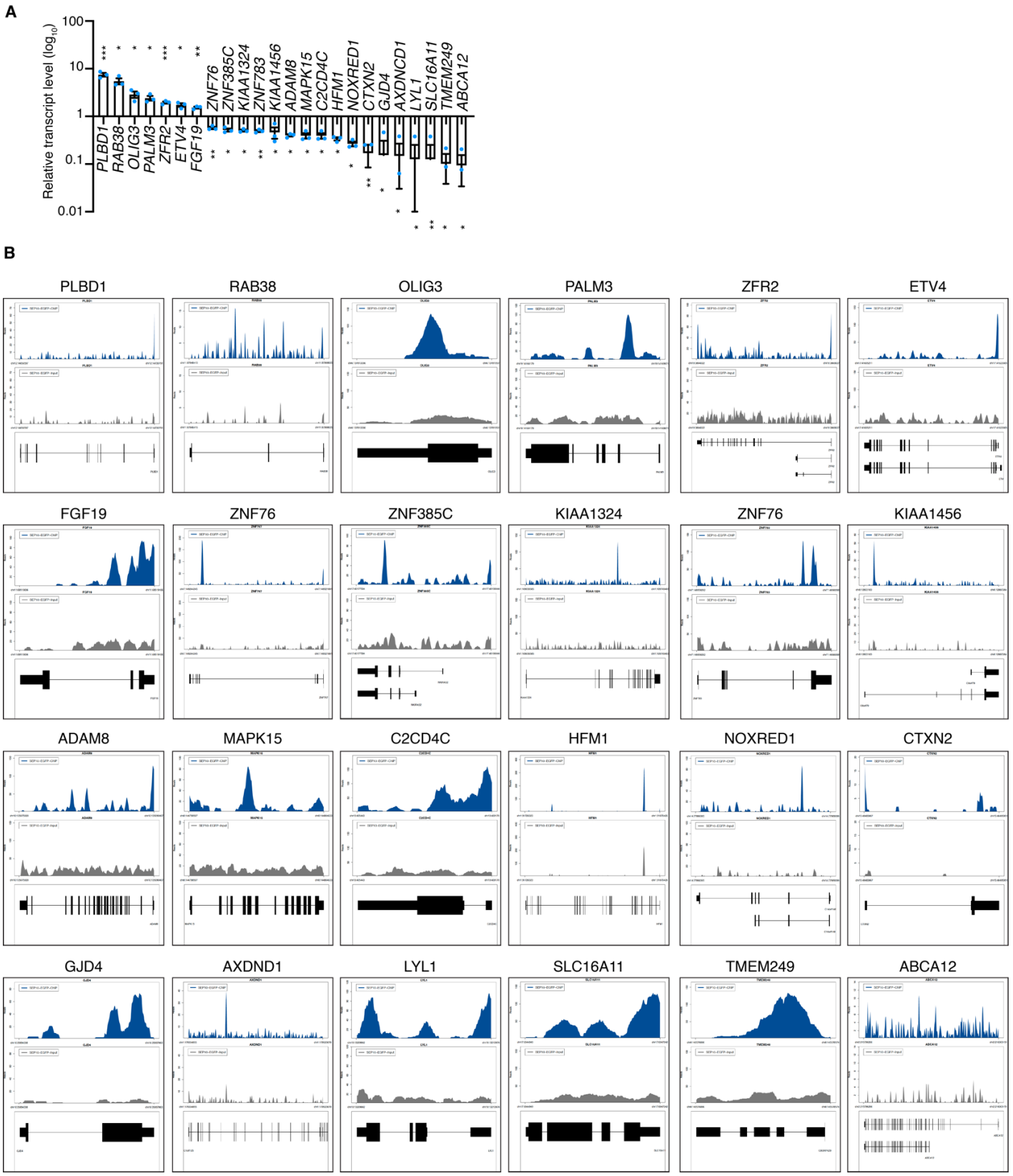

Figure S4. SEHBP modulates transcription at associated genomic loci. A) Relative abundance of the indicated transcripts derived from RNA-seq in response to SEHBP expression in HEK293T cells (*n*=3; mean and s.d.; **P* < 0.05, ***P* < 0.005, ***P* < 0.005, *t*-test). B) Plots depicting specific ChIP-seq-derived read abundance from SEHBP (blue) and background (grey) in relation to their genomic position with exons depicted as black boxes.

Table S1- Nomenclature and sequences of the SEPs used in this work. A lower case 'm' depicts an introduced ATG start site from a SEP using a noncanonical start codon.

ID	SEP amino acid sequence	SEP length (aa)	Parent RNA
1	mVRPDAISRSCAGPARLLVPGARVNARPRPRSPD	34	PMSD3
2	mTVRTGRSRGSEAVSGLGKKCLLDLALGGTLATRVGVIG	39	KLF9
3	MCLRRTTGYGHHDFRVPSFRSPAADPRRQTEALQ	34	MEIS2
4	MASEGKPRPGQAWRLPLEIGRAPFPAPRQQLCNSQPG	37	HOXA3
5	MTAPGLLKMSRICINIQQESIIVCPSSRAAKNGGIITSSKATL	43	FOXP4-AS1
6	MVHCCHSPDCIFETQALSNLQRTKRQPPRYVCWEGVIVTAL	41	CLASP1
7	MEDRWTAVWPVTSPLKPLESGPLSSAGHIIGILTHSLNSLFIKT	44	PIK3R1
8	mLPERANKKGAGVGSKINTSEFMMASTEALIGLKVSKQTITLWKKG	46	ZBTB37
9	MIRPQSSMSKHIPQFCGVLGHTFMEFLKGSGDYCQAQHDLYADK	44	PTP4A1
SEHBP (10)	MALRSIKSIAGSCLCSRQRRCGSSAAIFPEGIFRCLSPKFGQEFPE	46	ZNF689
11	MLAFWAPNWHWEGVTESTRLIPGPTFKRSSTYNPSISLKTKTSLN	45	MBNL1
12	MGACLIVNPSAWTPEPTQLSVCRGPRVRGDAQAVAETTLEMLMKTGLSC	49	AHDC1
13	MSPANGEQSRSRCLGSEGSCLRHRHGRWIQPPPAAAPGAMRRGAAATATARL	52	GSK3B
14	MASCPASVTSGSAFTKAFTRARGNLERRAPGTLVGAWGCEAAASPAGHYNYILRM	55	LOC105374809
15	MRQRLNSRRANQRHPESHVGLGAQRPSANPLKVSQRNLRDDLDPAERNCF	50	PCDHGC3
16	MYRFRSQLFTGISAAATAHSYPRRFSTLLLAEDSPLSRPPHRRTSKKCSSIG	52	IFRD1
17	mLWFGNVVDQLDMLAKCAPCNKMKRRDKMMSFSHIVKPKTNAFCETKLTNL	51	FZD3
18	MEGKWTICPGLQPDTIFQSQNLLHFLCKCQKNETPSLPSQRRGSNGSHVGTILSGS	56	HOXB6
19	mLSRCRRLPEERGGVPLALRPSRAESTRHSRTSEASRPRGTPPRRPPAGPRPLFPL	56	antisense chr17
20	MASDTQAVPDVSGRSRARRYAEWSSGLRRGAEPSGWAGLAGPAEGRRRLSGGEHALLTHWRLS	63	NDEL1
21	mVQGSSGRHFRELVDAVAGVDPELPKPPSCSPAWASLIPLFFLDSLGPSRPWNPIFGLGGSAAWA	65	RBM39
22	MRAGPRERQRRWRSGRRSPGREVWMRLTRGRAASPALVQRSPSASDQSPVIMQDSGPISSASS	63	DLX1
23	MGDQPCASGRSTLPPGNAREAKPPKKRCLLAPRWDYPEGTPNGGSTTLPSAPPPASAGLKSHPPPPEK	68	LINC01420
24	mLEELLPLPAPFLQKGGGNIHLFMPVCCMQAFWLPTLQQNNCTNSLVPIPPTESPGATVFFALHCRRGTKC	71	SOX4

Table S2- Amino acid sequences of the SEP transgenes used in this work.

Name	Amino acid sequence
3XFLAG-GGSGX-SEP01	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*VRPDAISRSCAGPARLLVPGARVNAR PRPRSPD
3XFLAG-GGSGX-SEP02	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*TVRTGRSRGSEAVSGLGKKCLLDLAL GGTLATRVGVIG
3XFLAG-GGSGX-SEP03	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*CLRRTTGYGHHDFRVPSFRSPAADPR RQTEALQ
3XFLAG-GGSGX-SEP04	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*ASEGKPRPGQAWRLPLEIGRAPFPAP RQQLCNSQPG
3XFLAG-GGSGX-SEP05	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*TAPGLLKMSRICINIQQESIIVCPSSRA AKNGGIITSSKATL
3XFLAG-GGSGX-SEP06	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*VHCCHSPDCIFETQALSNLQRTKRQP PRYVCWEGVIVTAL
3XFLAG-GGSGX-SEP07	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*EDRWTAVWPVTSPLKPLESGPLSSAG HIIGILTHSLNSLFIKT
3XFLAG-GGSGX-SEP08	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*LPERANKKGAGVGSKINTSEFMMAST EALIGLKVSKQTITLWKKG
3XFLAG-GGSGX-SEP09	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*IRPQSSMSKHIPQFCGVLGHTFMEFL KGSGDYCQAQHDLYADK
3XFLAG-GGSGX-SEP10	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*ALRSIKSIAGSCLCSRQRRCGSSAAIF PEGIFRCLSPKFGQEFPE
3XFLAG-GGSGX-SEP11	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*LAFWAPNWHWEGVTESTRLIPGPTFK RSSTYNPSISLKTKTSLN
3XFLAG-GGSGX-SEP12	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*GACLIVNPSAWTPEPTQLSVCRGPRV RGDAQAVAETTLEMLMKTGLSC
3XFLAG-GGSGX-SEP13	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*SPANGEQSRSRCLGSEGSCLRHRHG RWIQPPPAAAPGAMRRGAAATATARL
3XFLAG-GGSGX-SEP14	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*ASCPASVTSGSAFTKAFTRARGNLER RAPGTLVGAWGCEAAASPAGHYNYILRM
3XFLAG-GGSGX-SEP15	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*RQRLNSRRANQRHPESHVGLGAQRP SANPLKVSQRNLRDDLDPAERNCF
3XFLAG-GGSGX-SEP16	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*YRFRSQLFTGISAAATAHSYPRRFSTL LLAEDSPLSRPPHRRTSKKCSSIG
3XFLAG-GGSGX-SEP17	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*LWFGNVVDQLDMLAKCAPCNKMKRR DKMMSFSHIVKPKTNAFCETKLTNL
3XFLAG-GGSGX-SEP18	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*EGKWTICPGLQPDTIFQSQNLLHFLCK CQKNETPSLPSQRRGSNGSHVGTILSGS
3XFLAG-GGSGX-SEP19	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*LSRCRRLPEERGGVPLALRPSRAEST RHSRTSEASRPRGTPPRRPPAGPRPLFPL
3XFLAG-GGSGX-SEP20	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*ASDTQAVPDVSGRSRARRYAEWSSG LRRGAEPSGWAGLAGPAEGRRRLSGGEHALLTHWRLS
3XFLAG-GGSGX-SEP21	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*VQGSSGRHFRELVDAVAGVDPELPKP PSCSPAWASLIPLFFLDSLGPSRPWNPIFGLGGSAAWA
3XFLAG-GGSGX-SEP22	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*RAGPRERQRRWRSGRRSPGREVWM RLTRGRAASPALVQRSPSASDQSPVIMQDSGPISSASS
3XFLAG-GGSGX-SEP23	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*GDQPCASGRSTLPPGNAREAKPPKK RCLLAPRWDYPEGTPNGGSTTLPSAPPPASAGLKSHPPPPEK
3XFLAG-GGSGX-SEP24	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*LEELLPLPAPFLQKGGGNIHLFMPVCC MQAFWLPTLQQNNCTNSLVPIPPTESPGATVFFALHCRRGTKC
3XFLAG-GGSGX-mScarlet	MDYKDHDGDYKDHDIDYKDDDDKGGSG*AbK*VSKGEAVIKEFMRFKVHMEGSMNGH EFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQFMYGSRAFIKHPADIPDYY

	KQSFPEGFKWERVMNFEDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPPDGPVMQKK TMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNV DRKLDITSHNEDYTVVEQYERSEGRHSTGGMDELYKSGLRSRAQASNSAVDGTAGP GSTGSR
FLAG-GGSGX-MRI2	MDYKDDDDKSGG*AbK*ETLQSETKTRVLPSWLTAQVATKNVAPMKAPKRMRMAAVPV AAARCDSSGQKTPANLTPCDKDCVLHE
HA-HIST1H2BB	MYPYDVPDYAGGGGSPEPSKSAPAPKKGSKKAITKAQKKDGKKRKRSRKESYSIYVY KVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRLAHYNKRSTITSREIQTAVRLLL PGELAKHAVSEGTKAVTKYTSSK
HA-HIST1H4A	MYPYDVPDYAGGGGSSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRG GVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGF GG
HA-H3F3C	MYPYDVPDYALMARTKQTARKSTGGKAPRKQLATKAARKSTPSTCGVKPHRYRPGT VALREIRRYQKSTELLIRKLPFQRLVREIAQDFNTDLRFQSAAVGALQEASEAYLVGLLE DTNLCAIHAKRVTIMPKDIQLARRIRGERA
HA-HMGN1	MYPYDVPDYAGGGGSPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAK DKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAENGETKTEESPASDEAGEKEAK SD
HA-HMGN4	MYPYDVPDYAGGGGSPKRKAKGDAKGDKAKVKDEPQRRSARLSAKPAPPKPEPRP KKASAKKGEKLPKGRKGKADAGKDGNNPAKNRDASTLQSQKAEGTGDAK
HA-NME2	MYPYDVPDYAGGGGSANLERTFIAIKPDGVQRGLVGEIIKRFEQKGFRLVAMKFLRAS EEHLKQHYIDLKDRPFFPGLVKYMNSGPVVAMVWEGLNVVKTGRVMLGETNPADSK PGTIRGDFCIQVGRNIIHGSDSVKSAEKEISLWFKPEELVDYKSCAHDWVYE
MRI2-FLAG	METLQSETKTRVLPSWLTAQVATKNVAPMKAPKRMRMAAVPVAAARCDSSGQKTPA NLTPCDKDCVLHEDYKDDDDK
3XFLAG-SEP10	MDYKDHDGDYKDHDIDYKDDDDKALRSIKSIAGSCLCSRQRRCGSSAAIFPEGIFRCL SPKFGQEFPE
SEP10-EGFP	MALRSIKSIAGSCLCSRQRRCGSSAAIFPEGIFRCLSPKFGQEFPEGSVSKGEELFTG VVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQ NTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK