Supplemental Information

Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4

Qiang Zhang, Lei Zeng, Chen Shen, Ying Ju, Tsuyoshi Konuma, Chengcheng Zhao, Christopher R. Vakoc, & Ming-Ming Zhou

Supplemental Figures:

Figure S1. NMR structural analysis of the BRD4 ET domain-LANA complex. Related to Figure 1.

Figure S2. Mutational analysis of the BRD4 ET domain binding to the LANA peptide by NMR. Related to Figure 1.

Figure S3. Assessing effects of LANA mutations on binding to the Brd4 ET domain by NMR. Related to Figure 1.

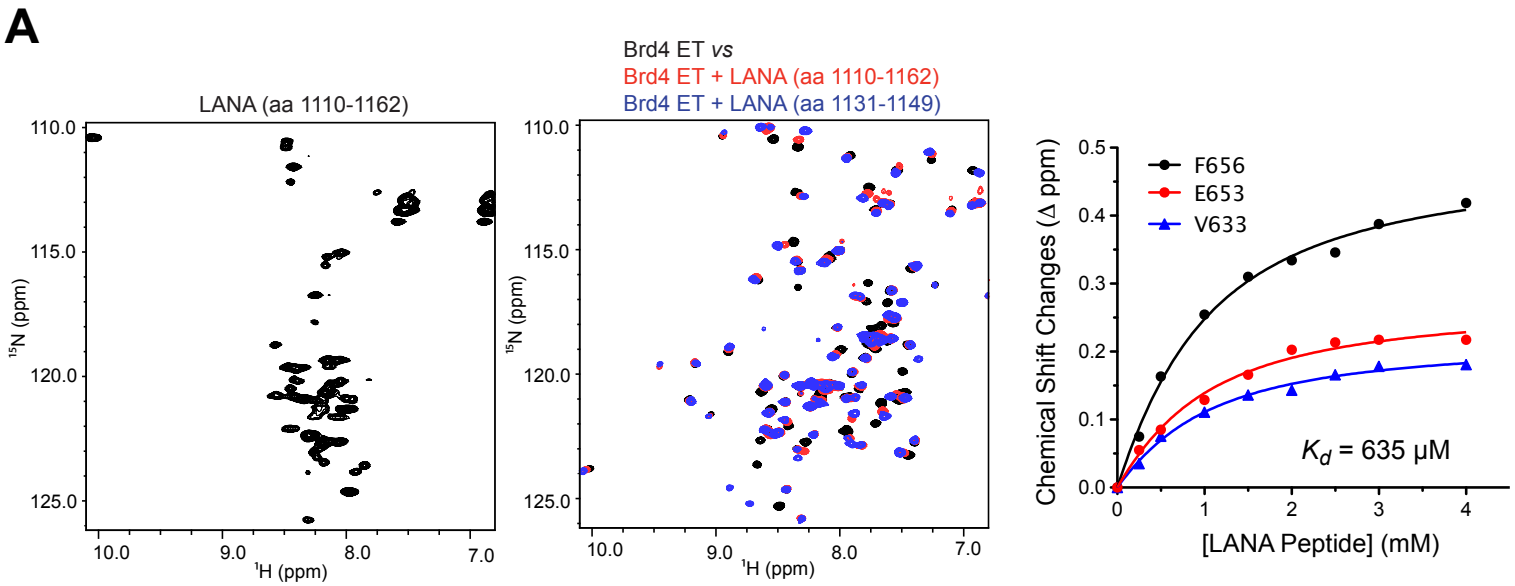
Figure S4. Characterization of the BRD4 ET domain binding motif in proteins in gene transcription. Related to Figure 2.

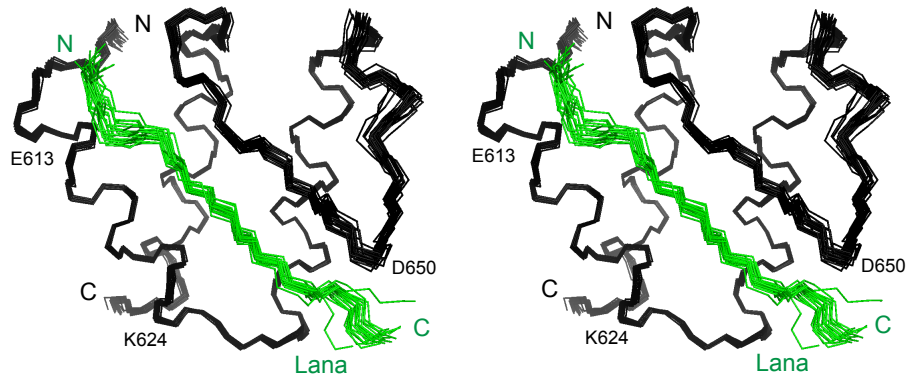
Figure S5. NMR structural analysis of the BRD4 ET domain binding to NSD3. Related to Figure 2.

Figure S6. NMR structural analysis of BRD4 ET domain binding to different sites in NSD3. Related to Figure 2.

Table S1. Putative ET domain binding sites in proteins involved in gene transcription. Related to Figure 2.

Table S2. Primers used in this study. Related to Figures 2 and 3.





C

```
                  601       610       620       630       640       650       660       670       680
                  | α1′       |  α1   |         |  α2     |         |  β1   |  α3   |         |
BRD4_Hs 601-  SEEEDK-CKPMSYEEKRQLSLDINKLPGEKLGRVVHIIQSREPSLKNSNPDEIEIDFETLKPSTLRELERYVTSCLRKK-RKPQA  -683
Brd4_Mm 602-  SEEEDK-CKPMSYEEKRQLSLDINKLPGEKLGRVVHIIQSREPSLKNSNPDEIEIDFETLKPSTLRELERYVTSCLRKK-RKPQA  -684
Brd4_Dr 654-  ---GEK-CKPMSYEEKRQLSLDINKLPGDKLGRVVHIIQSREPSLKNSNPDEIEIDFETLKPSTLRELERYVSSCLRKKKKPAVP  -731
Brd4_Xl 626-  SEEETQRCRPMSYEEKRQLSLDINKLPGEKLGRVVHIIQSREPSLKNSNPDEIEIDFETLKPSTLRELERYVTSCLRKKRKSQDK  -707
BRD2_Hs 633-  SEEEEES-RPMSYDEKRQLSLDINKLPGEKLGRVVHIIQAREPSLRDSNPEEIEIDFETLKPSTLRELERYVLSCLRKKPRKPYT  -716
Brd2_Mm 631-  SEEEEES-RPMSYDEKRQLSLDINKLPGEKLGRVVHIIQAREPSLRDSNPEEIEIDFETLKPSTLRELERYVLSCLRKKPRKPYT  -714
BRD3_Hs 563-  SEEEEEG-LPMSYDEKRQLSLDINRLPGEKLGRVVHIIQAREPSLRDSNPDEIEIDFETLKPTTLRELERYVKSCLQKKQRKPFS  -646
Brd3_Mm 564-  SEEEEEG-LPMSYDEKRQLSLDINRLPGEKLGRVVHIIQAREPSLRDSNPDEIEIDFETLKPTTLRELERYVKSCLQKKQRKPLS  -647
BRDT_Hs 501-  SEDEDNA-KPMNYDEKRQLSLNINKLPGDKLGRVVHIIQAREPSLSNSNPEEIEIDFETLKASTLRELEKYVSACLRKRPLKPPA  -584
BrdT_Mm 497-  SEEEDNA-KPMNYDEKRQLSLNINKLPGDKLGRVVHIIQAREPSLRNSNPEEIEIDFETLKASTLRELEKYVLACLRKRSLKPQA  -580
Prot_Dr 942-  SEEEDTA-KPMSYDEKRQLSLDINKLPGDKLGRVVHIIQNREPSLRDSNPDEIEIDFETLKPSTLRELESYVASCLRKKTHKKPS  -1026
GTE4_At 597-  PKANEPNKRDMTYEEKQKLSGHLQNLPPDKLDAIVQIVNKRNTAVKLRD-EEIEVDIDSVDPETLWELDRFVTNYKKGLSKKKRK  -670
GTE3_At 298-  LVEEASANRDLTFDEKRQLSEDLQDLPYDKLEAVVQIIKKRTPELSQQD-DEIELDIDSLDLETLWELFRFVTEYKESLSKKKEE  -380
Bdf1_Ys 526-  --------TVVTYDMKRIITERINDLPTSKLERAIDIIKKSMPNISED--DEVELDLDTLDNHTILTLYNTFFRQYESSSGASSN  -600
Bdf2_Ys 509-  RRKAMKH-SVDDLKKSITDKINELSDLEMNGMIRIIKNSLPADEILTSNEDEIEIDLDILDEATIARIYERYFEKKNNNNSKRKL  -587
```

Figure S1. NMR structural analysis of the BRD4 ET domain-LANA complex. Related to Figure 1.
(A) Identification of the ET domain binding site in the C-terminal region of LANA. Left, 2D ^{1}H-^{15}N HSQC spectrum of LANA C-terminal segment (residues 1110-1162). Middle, 2D ^{1}H-^{15}N HSQC spectra of the BRD4 ET domain in the free state (black) or in the presence of peptides derived from the C-terminal region of LANA (blue and red), as indicated. Right, determination of binding affinity of the LANA peptide (residues 1133-1144) to the BRD4 ET domain by NMR titriation using 2D ^{1}H-^{15}N HSQC spectroscopy.
(B) The structure of the BRD4 ET domain in complex with a LANA peptide (QSSIVKFKKPLP, residues 1133-1144) is shown as stereoview of the backbone atoms (N, Cα and C') of 25 superimposed NMR structures of the complexes (left).
(C) Structure-based sequence alignment of the ET domains from BRD4 proteins of human, mouse, fly, zebrafish and yeast. The conserved residues at the peptide binding site are indicated in bold, and highlighted in red for the absolutely conserved and blue for highly conserved. The key residues that are located in the peptide binding site are shown in bold and underlined.

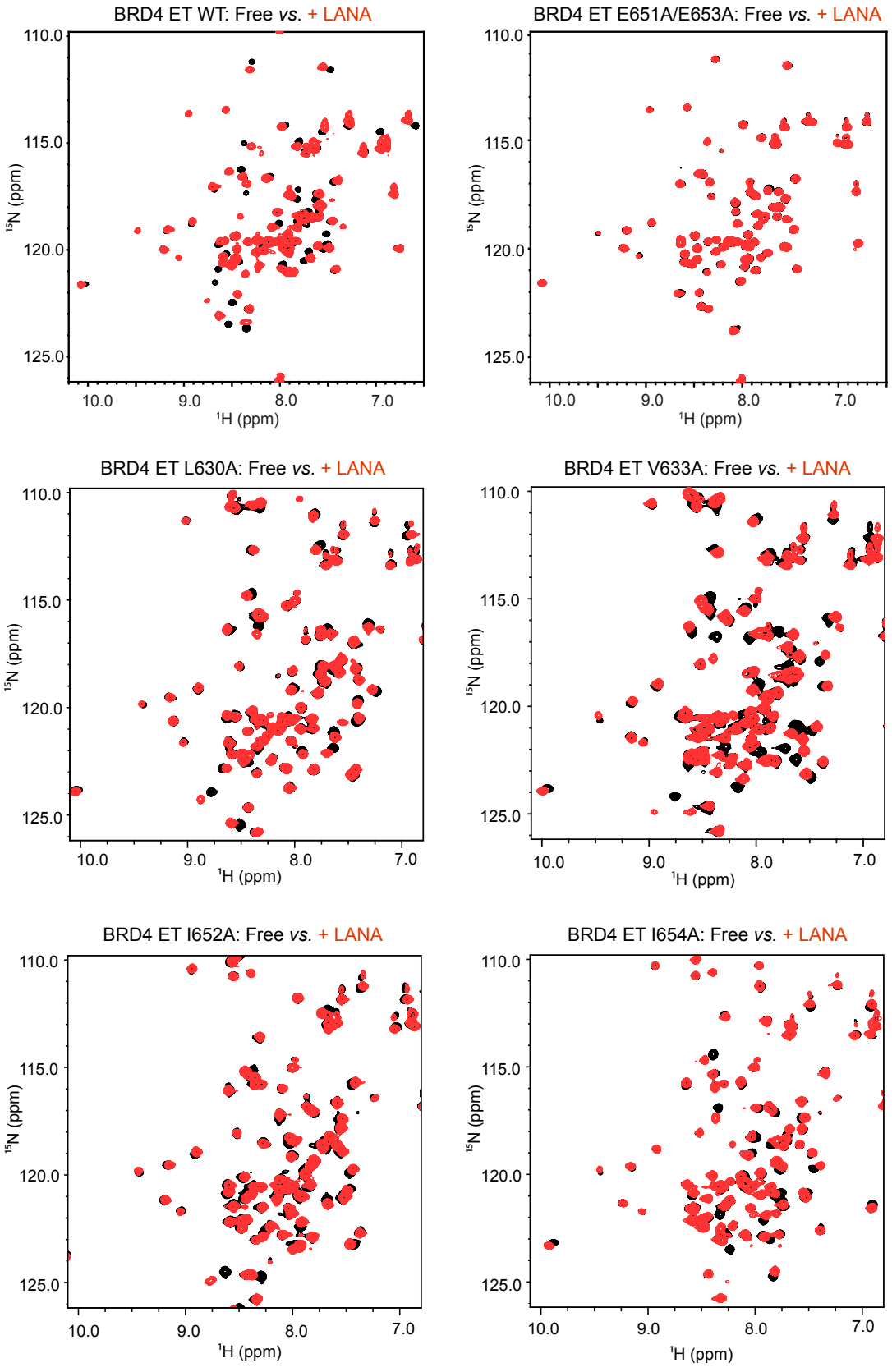


Figure S2. Mutational analysis of the BRD4 ET domain binding to the LANA peptide by NMR. Related to Figure 1. Superposition of 2D ^{1}H-^{15}N HSQC spectra of the Brd4 ET wild-type, or mutants (L630A, V633A, I652A, I654A, and E651A/E653A) in the free form (black) versus in the presence of the LANA peptide (red, residues 1033-1044), respectively. The protein concentration was 0.1 mM and the molar ratio of the protein to peptide was kept at 1:5.

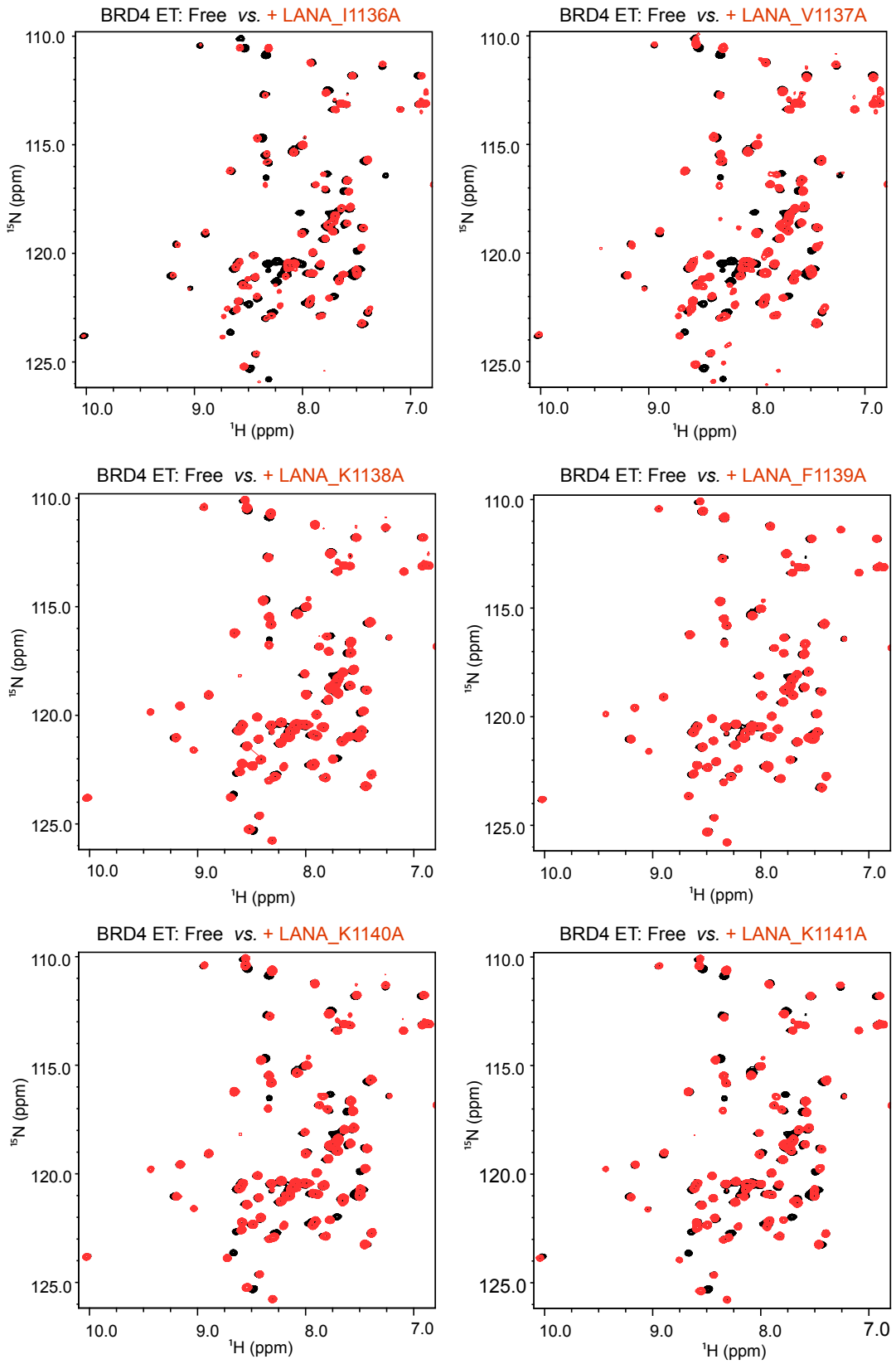


Figure S3. Assessing effects of LANA mutations on binding to the Brd4 ET domain by NMR. Related to Figure 1. Superposition of 2D ^{1}H-^{15}N HSQC spectra of the BRD4 ET domain in the free form (black) versus in the presence of mutant LANA peptide (red): LANA-I1136A; LANA-1137A; LANA-K1138A; LANA-F1139A; LANA-K1140A; LANA-K1141A. The protein concentration was 0.1 mM and the molar ratio of the protein to peptide was kept at 1:5.

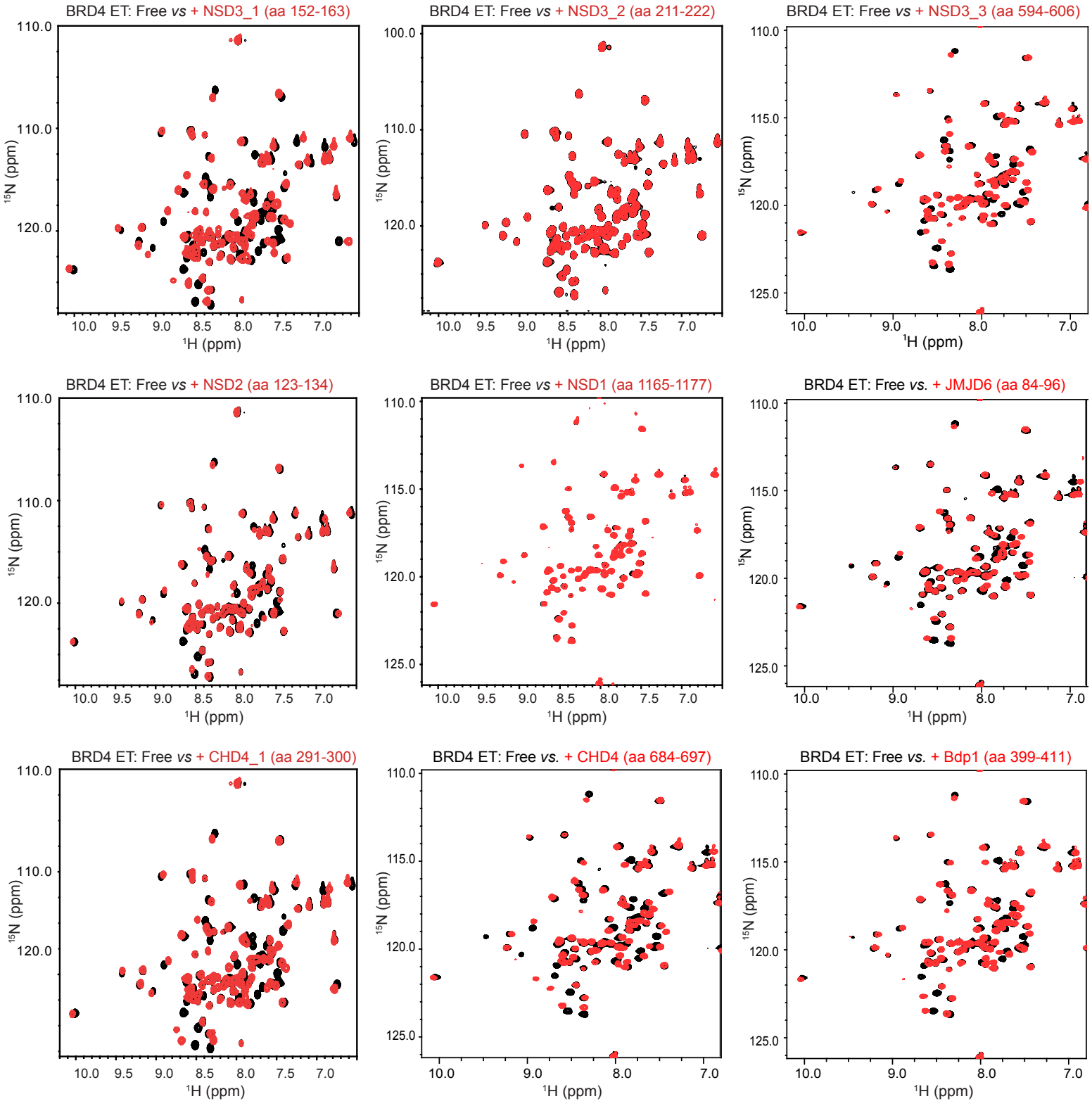


Figure S4. Characterization of the BRD4 ET domain binding motif in proteins in gene transcription. Related to Figure 2. Superposition of 2D ^{1}H-^{15}N HSQC spectra of the BRD4 ET domain in the free form (black) versus in the presence of a peptide derived various proteins as indicated (red). The protein concentration was 0.1 mM and the molar ratio of the protein to peptide was kept at 1:5.

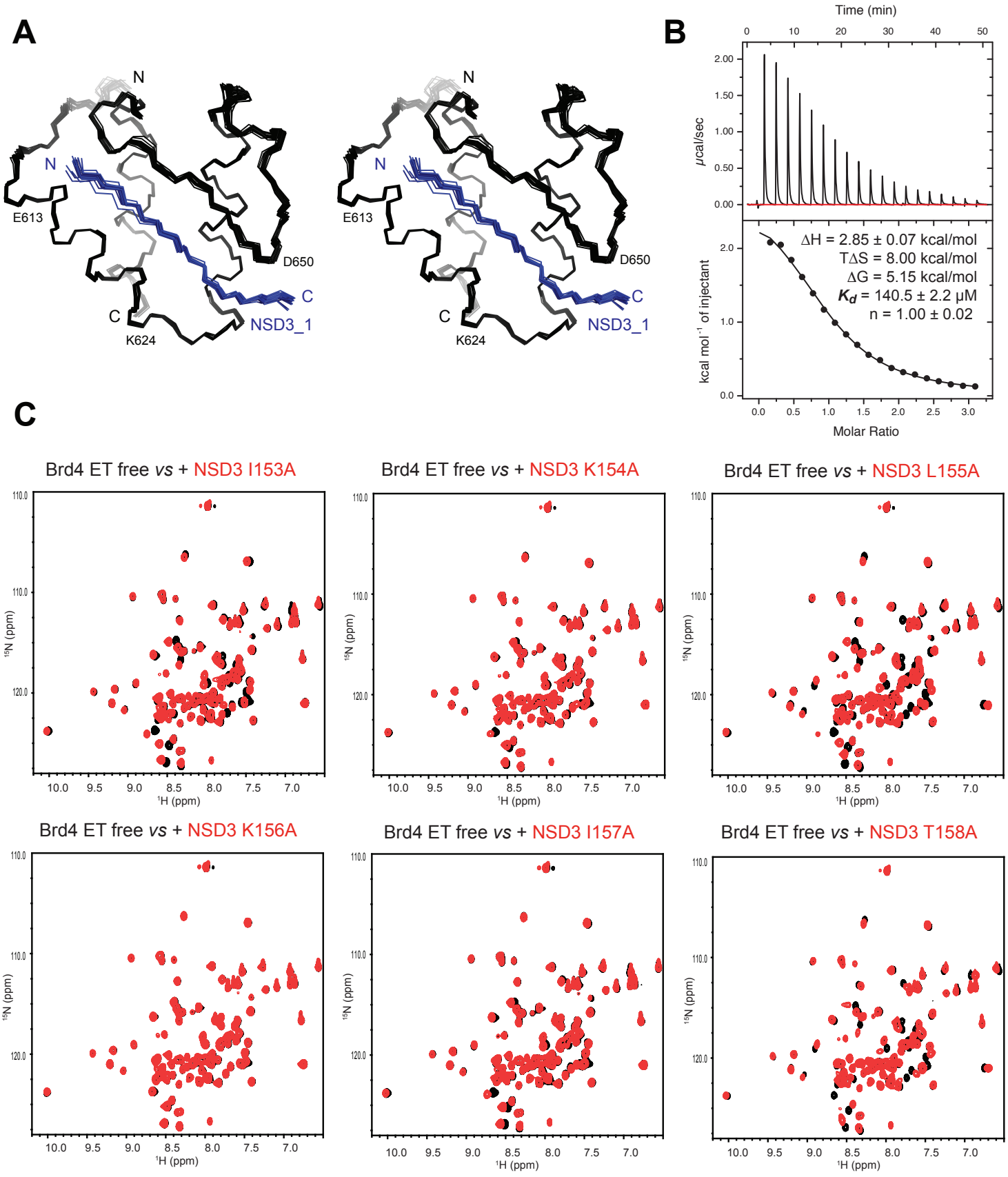
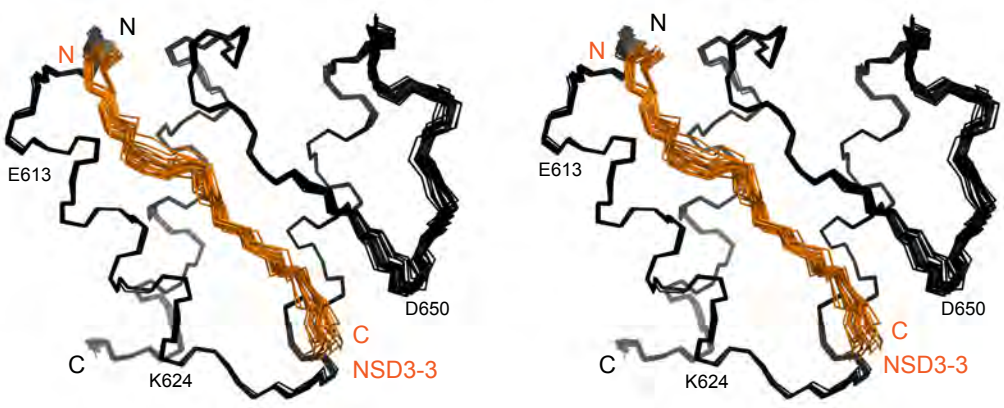


Figure S5. NMR structural analysis of the BRD4 ET domain binding to NSD3. Related to Figure 2.
(A) The structure of the BRD4 ET domain in complex with a NSD3-1 peptide (EIKLKITKTIQN, residues 152-163) is shown as stereoview of the backbone atoms (N, Cα and C') of 25 superimposed NMR structures of the complex (left).
(B) ITC determination of binding affinity of NSD3-1 peptide (SPEIKLKITKT, residues 150-160) to the ET domain of BRD4.
(C) Superposition of 2D ^{1}H-^{15}N HSQC spectra of the BRD4 ET domain in the free form (black) versus in the presence of various NSD3 peptides carrying point Ala mutation as indicated (red). The protein concentration was 0.2 mM and the molar ratio of the protein to peptide was kept at 1:5. The NMR sample sodium phosphate buffer used consists of 50 mM Na_2HPO_4-NaH_2PO_4 at pH6.5, containing 2.0 mM EDTA and 2.0 mM DTT.

A



B

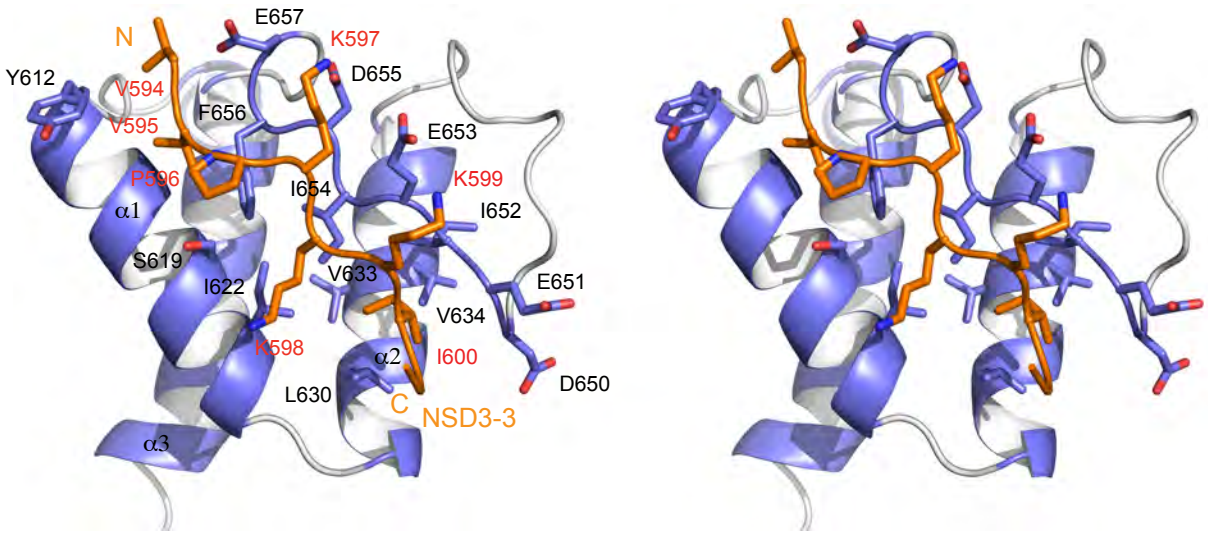


Figure S6. NMR structural analysis of BRD4 ET domain binding to different sites in NSD3. Related to Figure 2.
(A) The structure of the BRD4 ET domain in complex with a NSD3-3 peptide (VVPKKKIKKEQVE) is shown as stereoview of the backbone atoms (N, Cα and C') of 25 superimposed NMR structures of the complex.
(B) Ribbons representation of the average minimized NMR structure of the BRD4 ET domain/NSD3-3 peptide complex in a similar orientation, prepared using Pymol (middle).

Table S1. Putative ET domain binding sites in proteins involved in gene transcription. Related to Figure 2.

The identificatoin of putative ET domain binding motifs in cellular proteins was done using Scan Prosite program (http://prosite.expasy.org), as described in detail in Experimental Procedures. Note that RBBP4 is identified as one of Brd4 ET domain associated proteins reported by Rahman and colleagues (2011).

UniProKB_ID	Protein name	Putative ET domain binding sites	Residue numbers
P51532	Transcription activator BRG1	SRSVKVKIKLG	1590-1600
		SVKVKIKLGRK	1592-1602
Q9NR48	Histone-lysine N-methyltransferase ASH1L	EGNLKLKIGLQ	82-92
Q12873	Chromodomain-helicase-DNA-binding protein 3	MAPLKIKLGLL	289-299
Q03164	Histone-lysine N-methyltransferase 2A	FPGVKIKITHG	228-238
Q8NEZ4	Histone-lysine N-methyltransferase 2C	KPKLKLKIINQ	1200-1211
Q8WVC0	RNA polymerase-associated protein LEO1	RTRLKLKVENT	409-419
Q5JPI9	Protein-lysine N-methyltransferase METTL10	LSNIKLKVEDF	127-137
Q6ZTZ1	Myb/SANT-like DNA-binding domain-containing protein 1	GEEIKIKITNM	92-102
P35579	Myosin-9	NKELKVKLQEM	1787-1797
Q00872	Myosin-binding protein C	HASIKVKVVDF	609-619
Q9H3P2	Negative elongation factor A	GGAVKLKLLLG	49-59
Q96D46	60S ribosomal export protein NMD3	SKRLKVKLTIQ	107-117
Q9ULI1	NACHT and WD repeat domain-containing protein 2	DAALKIKIATS	1650-1660
Q86SE9	Polycomb group RING finger protein 5	FLSLKLKLPSS	179-189
Q9NUG6	p53 and DNA damage-regulated protein 1	RKQLKVKVNRL	92-102
Q92576	PHD finger protein 3	NKEIKVKVDNI	1520-1530
Q09028	Histone-binding protein RBBP4*	LRNLKLKLHSF	303-313
Q16576	Histone-binding protein RBBP7	LRNLKLKLHTF	302-312
Q9Y580	RNA-binding protein 7	GPVIKVKIPKD	34-44
Q2KHR2	DNA-binding protein RFX7	SPDIKVKLEGS	657-667
Q9H1D9	DNA-directed RNA polymerase III subunit RPC6	MAEVKVKVQPP	1-11
Q9H9Y6	DNA-directed RNA polymerase I subunit RPA2	AMNIKVKLDVV	1125-1135
Q86TU7	Histone-lysine N-methyltransferase setd3	HDRVKIKLGVS	334-344
Q9C086	INO80 complex subunit B.	KPQLKLKIKLG	66-76
		QLKLKIKLGGQ	68-78
O60583	Cyclin-T2	TSPIKMKIPIA	478-488
		KEELKMKIKVS	520-530
		ELKMKIKVSSS	522-532
Q9BZH6	WD repeat-containing protein 11.	ADVVKVKWARE	62-72
O75083	WD repeat-containing protein 1.	GPPFKFKFTIG	177-187
Q9P2L0	WD repeat-containing protein 35	NFMIKMKLSCL	188-198
Q6NYC1	Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6	GYSVKMKMKYY	107-117
Q96N64	PWWP domain-containing protein 2A	SNSLKMKVFSK	630-640
Q96EB1	Elongator complex protein 4	ESNIKMKIAWR	153-163
Q8TD26	Chromodomain-helicase-DNA-binding protein 6	MKMKIQKK	1-8
O14867	Transcription regulator protein BACH1	FQFLKFKFLDS	123-133
Q9P281	BAH and coiled-coil domain-containing protein 1	NMVVKVKWFYH	2516-2526
P56524	Histone deacetylase 4	STEVKMKLQEF	168-178
Q9UGU5	HMG domain-containing protein 4	PDGLKMKLILS	187-197
Q9UIG0	Tyrosine-protein kinase BAZ1B	EKMLKVKIVKI	137-147
Q9NYF8	Bcl-2-associated transcription factor 1	RPEVKLKMAPV	544-554
P49336	Cyclin-dependent kinase 8	DYDFKVKLSSE	2-12

Table S2. Primers used in this study. Related to Figures 2 and 3

The names of genes that the primers were designed to target are as shown. F and R stand for forward and reverse primers, respectively.

Gene	Primer sequence
m*Gapdh* RT_F	TTCACCACCATGGAGAAGGC
m*Gapdh* RT_R	CCCTTTTGGCTCCACCCT
m*Nsd3* RT_F	TCCTTACCAGCCTCCATCAC
m*Nsd3* RT_R	CCCATCTCCTGTTGCATTCT
m*Myc* RT_F	GCCGATCAGCTGGAGATGA
m*Myc* RT_R	GTCGTCAGGATCGCAGATGAAG
m*Myb* RT_F	GCTGAAGAAGCTGGTGGAAC
m*Myb* RT_R	CAACGCTTCGGACCATATTT
m*Check1* RT_F	ATTCTATGGCCACAGGAGGG
m*Check1* RT_R	ATAAACCACCCCTGCCATGA
m*Chst13* RT_F	CAGTGTTCGTTGAAGGGCTC
m*Chst13* RT_R	TTGTGTGCCCAAGAAGATGC
m*Elane* RT_F	TGGCCTCAGAGATTGTTGGT
m*Elane* RT_R	TACCTGCACTGACCGGAAAT
m*Hmgb2* RT_F	GAACACCCAGGCCTGTCTAT
m*Hmgb2* RT_R	TTCCTGCTTCACTTTTGCCC