



Title	Structural evidence for protein-protein interaction between the non-canonical methyl-CpG-binding domain of SETDB proteins and C11orf46
Author(s)	Mahana, Yutaka; Ariyoshi, Mariko; Nozawa, Ryu-Suke et al.
Citation	Structure. 2024, 32(3), p. 304-315.e5
Version Type	AM
URL	https://hdl.handle.net/11094/95866
rights	This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

1
2
3
4 **Structural evidence for protein–protein interaction between the non-canonical**
5 **methyl-CpG-binding domain of SETDB proteins and C11orf46**
6
7
8

9 **Authors**

10 Yutaka Mahana¹, Mariko Ariyoshi^{2,*†}, Ryu-Suke Nozawa^{3,‡}, Sachiko Shibata³, Koji Nagao³,
11 Chikashi Obuse³, and Masahiro Shirakawa^{1,*}
12
13
14

15 ¹Department of Molecular Engineering, Graduate School of Engineering, Kyoto University,
16 Nishikyo-Ku, Kyoto 615-8510, Japan
17

18 ²Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan
19

20 ³Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka,
21 Osaka 560-0043, Japan
22

23 *Correspondence to: ariyoshi.mariko.fbs@osaka-u.ac.jp (MA) or shirakawa@moleng.kyoto-
24 u.ac.jp (MS)
25

26 †Lead contact
27

28 ‡Present address: Division of Experimental Pathology, The Cancer Institute of JFCR, Koto-Ku,
29 Tokyo 135-8550, Japan
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5 **Summary**

6 SETDB1 and SETDB2 mediate trimethylation of histone H3 lysine 9 (H3K9), an epigenetic
7 hallmark of repressive chromatin. They contain a non-canonical methyl-CpG-binding domain
8 (MBD) and bifurcated SET domain, implying interplay between H3K9 trimethylation and DNA
9 methylation in SETDB functions. Here we report the crystal structure of human SETDB2 MBD
10 bound to the cysteine-rich domain of a zinc-binding protein, C11orf46. SETDB2 MBD
11 comprises the conserved MBD core and a unique N-terminal extension. Although the MBD core
12 has the conserved basic concave surface for DNA binding, it utilizes it for recognition of the
13 cysteine-rich domain of C11orf46. This interaction involves the conserved arginine finger motif
14 and the unique N-terminal extension of SETDB2 MBD, with a contribution from intermolecular
15 β -sheet formation. Thus, the non-canonical MBD of SETDB1/2 seems to have lost methylated
16 DNA-binding ability but gained a protein–protein interaction surface. Our findings provide
17 insight into the molecular assembly of SETDB-associated repression complexes.
18
19
20
21
22
23
24
25

26 **Keywords**

27 heterochromatin, histone methyltransferase, methyl-CpG-binding domain, SETDB1, SETDB2
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Introduction

Epigenetic modifications of the eukaryotic genome are fundamental regulators of various events in the nucleus. In particular, DNA methylation and post-translational modifications of histone proteins are major epigenetic marks that are widely involved in mammalian developmental processes and pathogenesis. Di- and trimethylation of lysine 9 of histone H3 (H3K9) are recognized as markers of repressed gene transcription and are essential for the formation of silenced heterochromatin.¹⁻³ SETDB2 and SETDB1 (also known as CLLD8 and ESET, respectively), which belong to the SUV39 histone methyltransferase subfamily, catalyze trimethylation of H3K9 (H3K9me3)⁴⁻⁷ and are required for germline development and early embryogenesis.^{5,8-11} The significance of SETDB1-mediated H3K9me3 has been intensively studied in murine embryonic stem cells, where depletion of SETDB1 leads to de-repression of transposable elements and other repetitive elements such as endogenous retroviruses^{7,12} and pericentric satellite repeats.¹³ SETDB2 is conserved only in higher species and thought to play an important role in the immune response by promoting transition from an inflammatory to a reparative state during viral infection¹⁴⁻¹⁶ and wound repair processes.¹⁷ As compared with SETDB1 and other SUV39 family members, SETDB2 is less studied and the mechanisms underlying its cellular localization, interactions with other proteins, and catalytic activity are poorly understood.³

The C-terminal catalytic module of SETDB proteins comprises three domains termed pre-SET, SET, and post-SET (Figure 1A). The SET domain is characterized by a large insertion in the vicinity of the active site. Previous studies have demonstrated that K867, which is located at the insertion of human SETDB1, as well as the equivalent lysine residue in mouse and fly orthologs, is monoubiquitylated.¹⁸⁻²¹ This modification has been shown to be essential for robust catalytic activity of SETDB1 and subsequent transcriptional silencing due to SETDB1-mediated H3K9me3 accumulation.^{18,19,22} The two SETDB proteins additionally possess a methyl-CpG-binding domain (MBD) that precedes the pre-SET domain (Figure 1A). MBD is an epigenetic reader motif that interprets the epigenetic status of DNA. In general, MBD of canonical MBD proteins such as MeCP2 and MBD1 recognizes the symmetrical methyl-CpG site in double-stranded DNA,²³⁻²⁹ although MBD of MBD4 shows preference for the methyl-CpG/TpG mismatch site.³⁰⁻³² The binding of canonical MBD proteins to either a gene promoter or repetitive region of transposable elements enriched with methyl-CpG sites leads to transcriptional repression of the genes encoded in these loci.³³⁻³⁶

Therefore, the presence of MBD in SETDB proteins implies the interplay of two repressive epigenomic markers: DNA methylation and H3K9 trimethylation. However, MBD of SETDB proteins shows less sequence similarity to canonical MBDs and has been classified as a non-canonical MBD.^{37,38} Recent studies suggest that the functional roles of non-canonical MBDs

1
2
3
4 may not be limited to methylated DNA binding. For example, the TIP5/ARBP/MBD (TAM)
5 domains of BAZ2A and BAZ2B proteins, which are components of the repressive chromatin
6 remodeling complex, recognize the stem-loop structure of promoter-associated RNA (pRNA)^{39–}
7 ⁴² and can bind non-specifically to double-stranded DNA.^{43,44} In addition, localization of SETDB
8 proteins to their target site has been proposed to partially rely on the DNA-binding ability of
9 MBD,^{4,45} while MBD of SETDB1 is required for the methyltransferase activity regulated by
10 monoubiquitylation.^{18,19,46} However, it is unclear how the non-canonical MBD of SETDB
11 proteins contributes to the SETDB-dependent accumulation of H3K9me3 on the genome.
12
13
14
15
16

17 In the present study, we have identified a human SETDB-associated protein, C11orf46, in
18 repression complexes, and determined the crystal structure of SETDB2 MBD in complex with
19 C11orf46. Our structure shows that SETDB2 MBD interacts directly with the cysteine-rich
20 domain of C11orf46 via the basic concave surface conserved in canonical MBDs. We identified
21 conserved and divergent structural characteristics of SETDB2 MBD mediating ~~To our~~
22 knowledge, this is the first structure showing a protein–protein interaction mediated with
23 C11orf46 rather than methylated DNA binding by MBD. Our findings imply that the non-
24 canonical MBD of SETDB proteins has a fundamental role in molecular assembly of the
25 repression complex and demonstrate the evolutionary and functional divergence of MBD proteins.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Results and Discussion

The methyl-CpG-binding domain of SETDB proteins directly interacts with the C-terminal cysteine-rich domain of C11orf46

To understand the functional roles of SETDB proteins in the context of enzyme–chromatin complexes, we first performed a proteomic analysis to identify binding partners. Consistent with previously published proteomic analyses,^{47,48} C11orf46 (also known as ARL14EP) was identified as a factor associated with both SETDB2 and SETDB1 in human cells (Figure S1A). We further verified the interaction between SETDB2 and C11orf46 by immunoprecipitation (IP) assay (Figure S1B). Because the two SETDB proteins share domain organization in the C-terminal region containing the non-canonical MBD and the catalytic module, but possess different functional domains in the extended N-terminal region (Figure 1A), we considered that C11orf46 must bind to the conserved region of SETDB proteins. Indeed, IP assays using several truncation mutants of SETDB2 demonstrated that C11orf46 interacts with the MBD-containing region of SETDB2 (residues 136–250; SETDB2^{136–250}) (Figures S1C–E). We also purified SETDB2^{136–250} and full-length C11orf46 (residues 1–260) as a complex using a bacterial co-expression system, suggesting their direct interaction (Figure S2A).

The exon structure of the human *C11orf46* gene, coupled with sequence-based prediction for disordered regions, suggested that C11orf46 consists of two globular regions separated by a flexible linker: an N-terminal domain (residues 1–142; NTD) and the C-terminal cysteine-rich domain (residues 186–260; CRD) (Figures 1A and 1B). To identify the minimal SETDB2-binding region of C11orf46, we designed four deletion variants of C11orf46 and examined their ability to form complexes with SETDB2^{136–250} by Ni affinity pull-down assay (Figure 1B). Deletion of CRD (Δ CRD) completely abolished complex formation, whereas deletion of NTD and the linker region did not affect it (Figure 1B), suggesting that C11orf46 CRD mediates the interaction with SETDB2^{136–250}. In addition, we found that a shorter fragment of SETDB2 containing residues 133–239 (hereafter “SETDB2 MBD”) was sufficient for stable complex formation with C11orf46 CRD. The equivalent SETDB1 MBD fragment (residues 565–674) was also found to form a complex with C11orf46 CRD (Figure S2B). In agreement with these findings, a previous study reported that a C11orf46 mutant lacking the cysteine-rich region (residues 209–237) does not associate with SETDB1 in the chromatin complex.⁴⁹ Using gel filtration chromatography, the molecular masses of the SETDB2 MBD–C11orf46 CRD and SETDB1 MBD–C11orf46 CRD complexes were estimated to be 26.9 kDa and 21.1 kDa, respectively, indicating formation of a heterodimer with 1:1 stoichiometry (Figure 1C).

Structure of the SETDB2 MBD–C11orf46 CRD complex

In the context of heterochromatin formation and transcriptional repression, MBDs have been

1
2
3
4 often described as methylated DNA-binding motifs with specificity for symmetrically
5 methylated CpG sites. However, our biochemical data demonstrated that the non-canonical MBD
6 of human SETDB1/2 directly interacts with C11orf46. In previous studies, human MBD3 and
7
8 *Arabidopsis thaliana* MBD2 and MBD6 were suggested to interact with other proteins *in*
9 *vitro*,^{50,51} while deletion of MBD of human MBD5 or MBD6 abolished their association with
10 components of a human Polycomb complex, PR-DUB.⁵² Nevertheless, no structural data
11 demonstrating a direct MBD–protein interaction have been reported. To understand the structural
12 basis of the protein–protein interaction mediated by the non-canonical MBD of SETDB proteins,
13 we therefore determined the crystal structure of SETDB2 MBD in complex with C11orf46 CRD
14 by using single-wavelength anomalous dispersion with intrinsic zinc atoms at 1.82 Å resolution
15 (Figure 2A and Table 1). In the crystal structure, SETDB2 MBD formed a heterodimer with
16 C11orf46 CRD. The complex contained three zinc ions (Figures 2A and S3A): one coordinated
17 by SETDB2 residues (Zn1), and two coordinated by C11orf46 residues (Zn2 and Zn3).
18 Stoichiometric binding of three zinc ions to the SETDB2 MBD–C11orf46 CRD complex was
19 validated by quantitative elemental analysis using inductively coupled plasma mass spectrometry
20 (ICP-MS) (Table S1). Two heterodimers were included in the asymmetric unit and their overall
21 structures were almost identical, with a root mean square deviation (RMSD) of 0.44 Å for the α -
22 carbon atoms of 151 residues (Figure S3B).
23
24

25
26
27
28
29
30
31
32
33 The structure shows that SETDB2 MBD comprises an atypical MBD fold (residues 161–
34 234; MBD core) and an additional extended region at the N-terminus (residues 140–152; N-
35 extension) (Figure 2A). Electron density of the region between the N-extension and the MBD
36 core (residues 153–160) is not observed, indicating the structural flexibility of this region. In the
37 crystal structure, the SETDB2 MBD core contains a short helix at the N-terminus, in addition to
38 the canonical MBD fold known as the α/β sandwich structure (Figures 2A and 2B).^{26,53} This core
39 structure can be well superimposed on previously reported MBD structures with an RMSD of 2–
40 3 Å for α -carbon atoms (Figure 2B). The N-extension of SETDB2 MBD, together with a residue
41 in the MBD core, forms a unique zinc-coordinating motif (Figure 2A). C11orf46 CRD is divided
42 into two structural moieties: an N-terminal zinc-binding core (residues 192–240; CRD core), and
43 a C-terminal region (residues 241–259; CRD tail) (Figure 2A). The CRD core adopts an intricate
44 loop-rich structure maintained via the coordination of two zinc ions (Zn2 and Zn3), while the
45 CRD tail forms a two-stranded antiparallel β -sheet. In a structural similarity search using Dali⁵⁴
46 and Foldseek⁵⁵ servers, C11orf46 CRD showed no significant similarity to any known protein
47 structure.
48
49
50
51
52
53
54
55
56

57 In canonical MBD proteins, MBD uses its basic concave surface for binding to double-
58 stranded DNA (Figure 2C, middle).^{28,29,32} The DNA-binding surface is positively charged and
59 thus highly attracted to negatively charged DNA. Intriguingly, the positively charged character of
60
61
62
63
64
65

1
2
3
4 the concave surface is also conserved in SETDB2 MBD (Figure 2C, left). Structural comparison
5 between the SETDB2 MBD–C11orf46 CRD and MeCP2 MBD–methylated DNA complexes
6 clearly shows that SETDB2 MBD binds to C11orf46 CRD in a manner similar to DNA
7 recognition by canonical MBDs (Figure 2C, left and middle). Regarding other non-canonical
8 MBD proteins, the TAM domain of BAZ2A and BAZ2B proteins contains an extra structural
9 motif composed of two β -strands and an α -helix at the C-terminal region (Figure 2B, right),^{42,44}
10 which is used to recognize the structure of pRNA (Figure 2C, right). This difference between
11 SETDB2 MBD and BAZ2A TAM in binding partner and molecular interface underlines the
12 diverse functional roles of MBDs.
13
14
15
16
17
18
19

20 **The loop-rich structure of C11orf46 CRD is maintained via two zinc-coordinating motifs**

21 An amino acid sequence alignment showed that C11orf46 CRD, which interacts with SETDB2
22 MBD, is widely conserved among animals (Figure 3A), whereas the N-terminal region of
23 C11orf46 varies among species. Notably, the CRD core of human C11orf46 contains no obvious
24 secondary structure elements but maintains a defined fold in the complex with SETDB2 MBD
25 (Figure S3B). *B*-factor profile analysis based on the crystal structure suggested that there is no
26 significant difference in folding stability between the two chains in the SETDB2 MBD–C11orf46
27 CRD complex (Figure S4): averaged *B*-factors of SETDB2 MBD and C11orf46 CRD were 38.6
28 Å² and 34.1 Å², respectively. In addition, no region with high flexibility was identified in the *B*-
29 factor profile of C11orf46 CRD in the complex (Figure S4).
30
31
32
33
34
35

36 The CRD core contains nine cysteine residues, eight of which are completely conserved
37 among several organisms (Figure 3A). Of the conserved cysteine residues, the N-terminal four
38 (C209, C211, C216, and C219) and C-terminal four (C223, C226, C231, and C235) coordinate
39 Zn²⁺ and Zn³⁺ ions, respectively, and form major interfaces with SETDB2 MBD (Figure 3B);
40 simultaneously, the zinc-coordinating networks are thought to contribute to the structural
41 framework of the loop-rich region. Importantly, two aspartate–arginine pairs in C11orf46 CRD,
42 namely D207–R236 and D210–R239, form a pair of intradomain salt bridges at the center of the
43 CRD core, with the result that the two zinc-coordinating motifs are tightly packed against each
44 other (Figure 3B). These salt bridges bring the N- and C-terminal ends of the CRD core closer
45 and contribute to maintaining the compact fold of the core, which is followed by the CRD tail. In
46 accordance with our structural data and IP analysis (Figure S1B), an amino acid substitution in
47 C11orf46, R236H, has been reported to abolish the interaction of C11orf46 with SETDB1⁴⁹ and
48 is associated with an autosomal-recessive intellectual disability.⁵⁶
49
50
51
52
53
54
55
56
57

58 **Conserved and divergent structural properties of the non-canonical MBD of SETDB** 59 **proteins** 60

1
2
3
4 Although SETDB2 MBD shares the common MBD fold comprising β 1, β 2, β 3, α 2, and η 2 with
5 canonical MBD family proteins (Figures 4A and 4B), it seems to have gained additional unique
6 structural features. The core fold of SETDB2 MBD is stabilized by a hairpin loop formed on the
7 β -sheet (Figures 4B and 4C), as observed in the structures of human MeCP2 and MBD1: a
8 conserved arginine residue (R175) in the β 1 strand, equivalent to R106 in MeCP2 and R17 in
9 MBD1, maintains the hairpin structure via polar contacts with S223, N225, and V228 in the loop
10 (Figure 4C).^{26,27,29} However, the SETDB2 MBD core exhibits more structural similarity to the
11 MBD moiety in the TAM domains of BAZ2A and BAZ2B proteins, which also contain an N-
12 terminal α 1 helix (Figures 4B and 2B).⁴²⁻⁴⁴ Two conserved proline residues in the α 1 helix of
13 SETDB2 MBD, P164 and P168, contribute to the bent structure of the short helix, as observed in
14 the solution structure of BAZ2A TAM (Figures 4A and 2B).⁴²

15
16 Previous structural studies of MBD family proteins have demonstrated that the methyl-
17 CpG-binding mode is well conserved in canonical MBDs (Figures 4A and 4D). The methyl-CpG
18 dinucleotide is mainly recognized by two ‘arginine fingers’ (RFs) in MeCP2, namely R111 (RF1)
19 and R133 (RF2). The side-chain orientation of each arginine finger is fixed by an acidic residue,
20 for example, D121 for RF1 and E137 for RF2 in MeCP2. A tyrosine residue located in the
21 middle of the β 2 strand, Y123 in MeCP2, also recognizes the 5-methyl group of the methylated
22 cytosine via water-mediated hydrogen bonds. In SETDB2 MBD, the arginine residue
23 corresponding to RF1 is replaced with an uncharged threonine residue. In addition, the critical
24 aspartate that fixes RF1 and the tyrosine residue responsible for base recognition are missing in
25 both SETDB2 and SETDB1. We note that the conserved pair of arginine fingers are replaced by
26 lysine in BAZ2A TAM, and the side chains of these lysine residues contact the DNA phosphate
27 backbone in the DNA-bound form.⁴³ In canonical MBDs, two positively charged residues in the
28 L1 loop between the β 1 and β 2 strands contribute to the interaction with the DNA phosphate
29 backbone (Figure 4A). SETDB2 does not possess such positively charged residues in its L1 loop.
30 Thus, the non-canonical MBD of SETDB proteins seems to have lost the key structural elements
31 for binding to methylated DNA. In this regard, SETDB1 orthologs are found even in species
32 such as *D. melanogaster* and *C. elegans*, in which there is almost no cytosine methylation.^{57,58}

33
34 A multiple sequence alignment of SETDB proteins from various species highlighted an
35 additional conserved region, termed the N-extension, in SETDB proteins (Figure 4A). Indeed,
36 our structure shows that the N-extension of SETDB2 MBD is involved in zinc ion coordination.
37 Three of the four zinc-coordinating residues, H145, C147, and C151, are located in the N-
38 extension and completely conserved in SETDB proteins from various species (Figures 4A and
39 4E). The fourth ligand, C195, is provided by the loop region between the β 2 and β 3 strands in the
40 MBD core (Figure 4E). A homology model of the human SETDB1 MBD–C11orf46 CRD
41 complex suggests that SETDB1 MBD also contains the N-extension providing the zinc-binding
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 site (Figure S5). To date, no MBD other than those of SETDB proteins has been identified or
5 predicted to contain a coordinated zinc ion. Therefore, this conserved zinc-coordinating motif
6 may have further structural or functional roles unique to SETDB proteins.
7
8
9

10 **The binding interface between SETDB2 MBD and C11orf46 CRD**

11 The SETDB2 MBD–C11orf46 CRD complex is maintained via two major interfaces. First, the
12 unique spherical structure of the CRD core of C11orf46 fits into the cavity formed by SETDB2
13 MBD. Several conserved side chains of SETDB2 contribute to recognition of the CRD core. An
14 arginine finger of SETDB2 MBD, namely R200, makes hydrogen bonds with the backbone
15 carbonyl oxygen atoms of Y221, A222, C231, and G232 of C11orf46 (Figures 4A and 5A, left).
16 This structural feature resembles the mode of methylated DNA recognition conserved in
17 canonical MBDs. For instance, the guanidino group of R133 (RF2) of MeCP2 MBD, the
18 equivalent of SETDB2 R200, provides hydrogen bonds for one of the two guanine bases in the
19 symmetrical methyl-CpG dinucleotide (Figure 5A, right). Indeed, the spatial arrangement of
20 SETDB2 RF bound to C11orf46 CRD can be well superimposed on that of MeCP2 RF2 in
21 complex with methylated DNA. We speculate that SETDB2 RF possibly contributes to the
22 rigidity of the loop-rich structure of the C11orf46 CRD core. On this SETDB2–C11orf46
23 interface, the tyrosyl group of C11orf46 Y221 is inserted into a quadrupole-like cavity formed by
24 the side chains of R197, E204, R207, and E211 of SETDB2 (Figure 5B). The aromatic ring of
25 C11orf46 Y221 is recognized by cation– π interactions with the guanidino groups of the arginine
26 residues and further stabilized by hydrophobic interactions involving the aliphatic moieties of
27 R197, E204, and R207 side chains. SETDB2 Y142, which is in the N-terminal region of the
28 zinc-coordinating motif, acts as a hydrophobic wall against the aliphatic portion of C11orf46
29 C219 (Figure 5B). The backbone carbonyl oxygen of Y142 also interacts with the backbone
30 amide group of C11orf46 L217 via a water-mediated hydrogen bond (Figure 5B). In our IP
31 assay, a SETDB2 fragment lacking the N-extension region (residues 160–250) lost the ability to
32 bind C11orf46 (Figure S1E). These observations suggest that the N-extension of SETDB2 MBD
33 partly provides a scaffold to recognize the spherical shape of the C11orf46 CRD core. Therefore,
34 the coordination of Zn1 in SETDB2 MBD presumably contributes to stabilization of the complex
35 interface.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 The second SETDB2–C11orf46 major interface is an antiparallel intermolecular β -sheet
53 formed between the β 1 strand in the MBD core of SETDB2 (residues 175–181) and the β 1'
54 strand in the CRD tail of C11orf46 (residues 242–249) (Figure 5C). In addition to intermolecular
55 backbone contacts, several polar interactions involving side chains are found at both ends of the
56 β -sheet (Figures 5D and 5E). At one end, the acidic side chain of C11orf46 E249 is inserted into
57 a basic pocket of the SETDB2 MBD core and makes a salt bridge and hydrogen bonds with
58
59
60
61
62
63
64
65

1
2
3
4 SETDB2 residues including R175, which maintains the hairpin loop (Figure 5E). Collectively,
5 our structure suggests that the CRD tail of C11orf46 has a crucial role in recognition and
6 stabilization of the MBD structure of SETDB proteins.
7
8

9 To evaluate the significance of the associations observed in the crystal structure, we
10 performed interaction analyses using a bacterial co-expression system for SETDB2 MBD and
11 hexahistidine-tagged C11orf46 CRD mutants. We introduced an amino acid substitution of each
12 binding key residue and then examined formation of the SETDB2 MBD–C11orf46 CRD
13 complex by Ni affinity pull-down assay. As expected, the conserved arginine finger mutant,
14 SETDB2 R200E, showed loss of efficient binding to C11orf46 CRD (Figure 5F, left). Similarly,
15 the R636E mutation in SETDB1 MBD, which is equivalent to R200E in SETDB2 MBD, also
16 disturbed formation of the complex with C11orf46 CRD (Figure 5F, right). Consistent with our
17 structural observations, deletion of the C-terminal tail from C11orf46 CRD resulted in
18 significantly less efficient complex formation with both SETDB2 MBD and SETDB1 MBD
19 (Figure 5F). By contrast, C11orf46 Y221A, SETDB2 Y142A, and SETDB1 Y573A mutations
20 caused no significant change in complex formation (Figure 5F), suggesting that the interactions
21 mediated by these tyrosine residues are auxiliary in terms of binding affinity and complex
22 stability. Taken together, our results suggest that the conserved RF motif of SETDB proteins and
23 stable β -sheet formation between the MBD core and the CRD tail are required for the interaction
24 between SETDB proteins and C11orf46.
25
26
27
28
29
30
31
32
33
34
35

36 **Concluding remarks**

37 Given that SETDB proteins have been identified in multiple repressive chromatin complexes,^{59–}
38 ⁶¹ and their common domain organization harbors the non-canonical MBD alongside the histone
39 methyltransferase domain, we initially hypothesized that SETDB MBD may be involved in
40 methylated DNA binding. However, the unique structural features of SETDB2 MBD reported
41 herein suggest that it lacks DNA binding ability even in the absence of C11orf46 CRD: several
42 key residues observed in canonical MBDs, including RF1 and basic residues in the L1 loop, are
43 not conserved in the non-canonical MBD of SETDB proteins (Figure 4A). Previous studies have
44 suggested that the robust catalytic activity of human SETDB1 requires monoubiquitylation at
45 K867 in the bifurcated SET domain (Figure 1A).^{18,19} Intriguingly, the MBD-containing region of
46 human SETDB1 (residues 570–680) has been shown to be required for K867 ubiquitylation.¹⁹
47 Because MBD is located in the vicinity of the SET catalytic module in SETDB proteins, it is
48 possible that binding of C11orf46 to MBD might modulate the histone methyltransferase activity
49 of SETDB proteins. Deletion of the N-terminal region including the N-extension (residues 1–
50 584) in human SETDB1 has been shown to cause loss of monoubiquitylation at K867.¹⁸ Given
51 that the N-extension region that coordinates Zn1 is required for the interaction with C11orf46
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 CRD (Figure S1E), it seems probable that the interaction between SETDB1 MBD and C11orf46
5 CRD participates in the regulation of the ubiquitylation-dependent catalytic activity of SETDB1.
6

7 A recent study based on an IP assay showed that human C11orf46 interacts with histone
8 H3 in a SETDB1-independent manner.⁴⁹ In addition, an *in vitro* binding analysis using a histone
9 peptide array suggested that C11orf46 preferentially recognizes H3 peptides with methylation or
10 citrullination at R2.⁴⁹ Notably, a C11orf46 mutant lacking the cysteine-rich region (residues 209–
11 237) has been shown to maintain its interaction with H3.⁴⁹ Furthermore, our NMR titration
12 experiment showed no binding between the SETDB2 MBD–C11orf46 CRD complex and an H3
13 N-terminal peptide (residues 1–36) (Figure S6). These findings suggest that C11orf46, through
14 its interaction with the histone H3 tail via its NTD or flexible linker region (Figure 1A), may
15 play a role in the recruitment of SETDB proteins to their target site.
16

17 We note that the orthologs of SETDB1 and C11orf46 in *C. elegans*, MET-2 and ARLE-
18 14, have been shown to colocalize in perinuclear heterochromatic foci of early embryos.^{62,63} At
19 present, however, the significance of the interaction between C11orf46 CRD and the non-
20 canonical MBD of SETDB proteins in the repression complex remains an open question. Further
21 biochemical and structural characterization of SETDB proteins and C11orf46 in their full-length
22 context will be required to understand the functional role of the SETDB–C11orf46 association.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgments

We thank Profs. Yao Min and Yoshikazu Tanaka for protein expression in the first stage of the experiment, Emiko Noro and Teruaki Okada for technical assistance, and Dr. Naohito Nozaki for production of the C11orf46 and SENP7 antibodies. This work was supported by JSPS KAKENHI under Grant Numbers 18K06084 to M.A., 17H06426 and 23H02411 to K.N., 22H02546, 19H03156, and 18H05532 to C.O., and 20H00453 to M.S..

Author contributions

Y.M. prepared materials and performed all biochemical experiments and structural analysis under the supervision of M.A. and M.S.. R.N., S.S., K.N., and C.O. performed the proteomic analysis and IP experiments. Y.M. and M.A. wrote the manuscript, discussing with all authors.

Declaration of interests

The authors declare no competing interests.

Figure legends

Figure 1. Direct interaction of the non-canonical MBD of SETDB proteins with C11orf46 CRD.

(A) Domain organization of human SETDB2, SETDB1, and C11orf46 proteins. MBD, methyl-CpG-binding domain; SET, Su(var)3-9, Enhancer-of-zeste and Trithorax domain; HD, helical dimerization domain; TTD, tandem Tudor domain; NTD, N-terminal domain; CRD, cysteine-rich domain.

(B) Identification of the SETDB2 MBD-binding region of C11orf46. Upper, prediction of disordered regions in C11orf46 protein. The exon structure of human C11orf46 is shown on top, along with the C11orf46 deletion mutants used in biochemical assays. Lower, Ni affinity pull-down assay of the N-terminal hexahistidine-tagged C11orf46 fragments indicated in the upper panel after co-expression with SETDB2^{136–250}. Left lanes, supernatant of the cell lysate (s); right lanes, proteins bound to Ni-NTA agarose resin (b).

(C) Analytical gel filtration chromatography of SETDB MBD–C11orf46 CRD complexes using a Superdex 200 10/300 GL column. Upper left, overlaid chromatograms of protein samples. Upper right, estimation of the molecular weight of each complex based on the partition coefficient calculated from the elution volume. Lower, mass spectra of the peak fraction of SETDB2 MBD–C11orf46 CRD (left) and SETDB1 MBD–C11orf46 CRD (right) complexes. Asterisks indicate the divalent cation peaks of SETDB fragments.

Figure 2. Overall structure of the SETDB2 MBD–C11orf46 CRD complex.

(A) Cartoon representation of the crystal structure of the SETDB2 MBD–C11orf46 CRD complex. The MBD core, N-terminal extension, CRD core, and CRD tail are colored in pink, magenta, cyan, and blue, respectively. Zinc ions are represented as grey spheres. The side chains of zinc-coordinating residues are shown as stick models. The zinc-coordinating interactions are indicated by orange dashed lines. The schematic diagram at the bottom indicates the structured regions of the complex.

(B) Comparison of MBD structure between SETDB2 and other MBD family proteins. Left, crystal structure of SETDB2 MBD in complex with C11orf46 CRD; middle, crystal structure of MeCP2 MBD in complex with methylated DNA (Protein Data Bank entry 6C1Y); right, solution structure of BAZ2A TAM (Protein Data Bank entry 5AGQ). The conserved MBD core structure and the extended structural motif are shown in pale and dark colors, respectively. The indicated RMSD values were calculated for the positions of α -carbon atoms relative to SETDB2 MBD.

(C) Electrostatic potential of SETDB2 MBD, MeCP2 MBD, and BAZ2A TAM, ranging from $-5 k_B T e^{-1}$ (red) to $5 k_B T e^{-1}$ (blue), mapped on the surface representation. The binding partners, C11orf46 CRD for SETDB2 MBD and methylated DNA for MeCP2 MBD, are shown as grey

1
2
3
4 cartoon models. The pRNA-binding interface of BAZ2A TAM identified by NMR titration
5 experiments is indicated by a grey circle.
6
7
8

9 **Figure 3. Structure of C11orf46 CRD.**

10 (A) Amino acid sequence alignment of the CRD domain of human C11orf46 and its orthologs
11 from several species. Identical and similar residues are highlighted in black and grey,
12 respectively. Zinc-coordinating cysteine residues are colored in cyan and marked with filled
13 (Zn2) or open (Zn3) inverted triangles. The conserved arginine and aspartic acid residues that
14 form intradomain salt bridges are connected by black lines. Residues with side chains involved
15 in the interface with SETDB2 MBD are indicated by open black inverted triangles. Domain
16 segments and secondary structure elements identified in the crystal structure are displayed on
17 top; dashed lines indicate the unstructured regions. The ruler indicates the residue numbers of
18 human C11orf46 protein.
19

20 (B) Cartoon representation of C11orf46 CRD. The side chains of residues participating in the
21 two zinc-coordinating motifs and the two salt bridges of C11orf46 CRD are shown as stick
22 models. Zinc-coordinating interactions and salt bridges are indicated by orange dashed lines.
23 SETDB2 MBD is displayed as a cartoon and transparent surface colored in pale pink.
24
25
26
27
28
29
30
31
32

33 **Figure 4. Structural characteristics of SETDB2 MBD.**

34 (A) Sequence alignment of MBD of SETDB proteins from several species (upper), canonical
35 MBD family proteins (middle), and BAZ2A/B proteins (lower). The residue numbers correspond
36 to human SETDB2 protein. Zinc-coordinating residues are highlighted in magenta or pink and
37 marked with inverted triangles. Conserved structural motifs in the MBD moiety are annotated on
38 top of the sequence. Black open inverted triangles denote residues with side chains involved in
39 the interface with C11orf46 CRD. Residues in the canonical MBD that recognize the methyl-
40 CpG pair and residues in the L1 loop that make electrostatic contact with the DNA phosphate
41 backbone are indicated at the bottom by black and white circles, respectively. Hs, *Homo sapiens*;
42 Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Xl, *Xenopus laevis*; Dr, *Danio rerio*; Dm,
43 *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*.
44
45
46
47
48
49

50 (B) Cartoon representation of the structure of SETDB2 MBD in the complex with C11orf46
51 CRD. Side chains of SETDB2 residues forming the zinc-binding site and the hairpin loop are
52 displayed as stick models.
53
54

55 (C) The hairpin loop structure of SETDB2 MBD. For clarity, the C11orf46 CRD moiety is
56 shown as a surface model colored in pale cyan.
57

58 (D) Structural comparison between SETDB2 MBD (left) and MeCP2 MBD (right). Residues
59 critical for methyl-CpG-specific recognition by MeCP2 (indicated by black circles in (A)) and
60
61
62
63
64
65

1
2
3
4 the corresponding residues in SETDB2 are shown as stick models. The methyl-CpG pair is
5 shown as a stick model colored in grey, and the methyl groups are displayed as yellow spheres.
6 The N-extension of SETDB2 MBD is omitted for clarity.

7
8
9 (E) The zinc-coordinating motif of SETDB2 MBD. The C11orf46 CRD moiety is shown as a
10 surface model colored in pale cyan.
11
12
13

14 **Figure 5. Molecular interfaces between SETDB2 MBD and C11orf46 CRD.**

15 (A) Comparison of the mode of molecular recognition by the arginine finger between SETDB2
16 and MeCP2.
17

18 (B) Hydrophobic and indirect interactions of SETDB2 MBD with the C11orf46 CRD core. The
19 SETDB2 MBD core is shown as a transparent surface colored in pale pink. The aliphatic
20 portions of SETDB2 Y142 and C11orf46 C219 are shown as transparent spheres. A water
21 molecule that bridges the hydrogen bond between SETDB2 Y142 and C11orf46 L217 is shown
22 as a red sphere (labeled as 'Wat').
23

24 (C) Intermolecular β -sheet formed by the SETDB2 MBD core and the C11orf46 CRD tail.
25

26 (D–E) Magnified views of the intermolecular β -sheet interface shown in (C).
27

28 (F) Analysis of complex formation by structure-based mutants of C11orf46 CRD and MBD of
29 SETDB2 or SETDB1 using Ni affinity pull-down assay. Left lanes, supernatant of the cell lysate
30 (s); right lanes, proteins bound to Ni-NTA agarose resin (b).
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Tables

Table 1. Data collection and refinement statistics.

	Crystal 1	Crystal 2
Data collection		
Beamline	SPring-8 BL41XU	SPring-8 BL32XU
Wavelength (Å)	1.2800	1.0000
Resolution (Å) ^a	38.81–1.90 (1.94–1.90)	42.06–1.82 (1.86–1.82)
Space group	<i>P</i> 3 ₁ 21	<i>P</i> 3 ₁ 21
Lattice parameters	<i>a</i> , <i>b</i> , <i>c</i> (Å)	69.764, 69.764, 116.418
	<i>α</i> , <i>β</i> , <i>γ</i>	90°, 90°, 120°
Number of total observations	261996 (17225)	395803 (14670)
Number of unique reflections	26540 (1699)	30318 (1709)
Multiplicity	9.9 (10.1)	13.1 (8.6)
<i>R</i> _{meas}	0.086 (0.635)	0.098 (0.898)
<i>R</i> _{pim}	0.027 (0.199)	0.026 (0.294)
Completeness	1.00 (1.00)	0.998 (0.968)
Mean <i>I</i> / <i>σ</i> (<i>I</i>)	17.7 (3.7)	15.8 (2.3)
<i>CC</i> _{1/2}	0.999 (0.911)	0.999 (0.812)
Refinement		
Resolution (Å)		42.06–1.82
Number of reflections	work, free	28818, 1468
<i>R</i> _{work} , <i>R</i> _{free}		0.186, 0.227
Number of atoms	protein	2422
		(4 chains/2 complexes)
	Zn ²⁺	6
	water	110
	SO ₄ ²⁻	30
Mean <i>B</i> -factor (Å ²)	protein	36.9
	Zn ²⁺	28.7
	water	37.8
	SO ₄ ²⁻	60.1
Bond RMSD statistics	length (Å)	0.015
	angle	1.91°
MolProbity score		1.02
Ramachandran plot statistics	favored	96.26%
	allowed	3.74%

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

outliers

0%

^a Values in parentheses are for the highest resolution shell.

1
2
3
4 **STAR Methods**
5
6

7 **Resource availability**

8 **Lead contact**

9
10 Further information and requests for resources and reagents should be directed to and will be
11 fulfilled by the lead contact, Mariko Ariyoshi (ariyoshi.mariko.fbs@osaka-u.ac.jp).
12
13
14

15 **Materials availability**

16
17 All reagents generated in this study are available from the lead contact upon request.
18
19

20 **Data and code availability**

21
22 The atomic coordinates of the SETDB2 MBD–C11orf46 CRD complex and the structure factors
23 have been deposited in the Protein Data Bank with the accession code 8HFP and are publicly
24 available as of the date of publication.
25

26 This paper does not report original code.

27
28 Any additional information required to reanalyze the data reported in this paper is available from
29 the lead contact upon request.
30
31
32

33 **Experimental model and ~~study participant~~ subject details**

34 **Cell lines and bacterial strains**

35
36 Flp-In T-REx 293 cells were purchased from Invitrogen and cultured at 37°C in Dulbecco's
37 modified Eagle medium (DMEM) containing 5% fetal bovine serum (FBS).
38

39 *Escherichia coli* strains DH5 α and BL21(DE3) were purchased from Takara and Novagen,
40 respectively, and grown in lysogeny broth (LB) medium or M9 minimal medium. All
41 recombinant plasmids were amplified in DH5 α cells at 37°C. All recombinant proteins were
42 expressed in BL21(DE3) at 25°C or 17°C.
43
44
45
46

47 **Method details**

48 **Expression plasmids for transfection in mammalian cells**

49
50 The cDNA sequence of full-length human C11orf46 (NotI/XhoI), SETDB1 (NotI/KpnI), or
51 SETDB2 (NotI/BamHI) was cloned into a modified pcDNA5/FRT/TO vector (Invitrogen), which
52 attaches 3 \times FLAG-tags to the N-terminus of the protein, using the restriction enzyme cleavage
53 sites indicated in parentheses. All truncated or point mutants of C11orf46 and SETDB2 were
54 created by polymerase chain reaction (PCR), cloned into the same vector, and verified by
55 nucleotide sequencing. These plasmids were also used as templates in PCR to establish all
56 expression plasmids for the recombinant proteins used in this study (see below).
57
58
59
60
61
62
63
64
65

Antibodies

A human C11orf46-specific mouse monoclonal antibody (36A9) was raised against a mixture of two C11orf46 peptides, $_{18}\text{HKTYYYTRHTGFKTLQELSS}_{36}$ -Cys and Cys- $_{242}\text{LYEQIEIEGGEEIIHNKHAG}_{260}$. A human SENP7-specific mouse monoclonal antibody (27RE) was raised against a SENP7 fragment (residues F715–S1050), which was expressed in and purified from *E. coli* as a His₆-fusion protein. The other antibodies used in this study were anti-FLAG (mouse; M2; Sigma), anti-SETDB1 (rabbit monoclonal; C1C12; Cell Signaling), and anti-SETDB2 (mouse; 2F4; Abnova).

Cell culture

To establish stable cell lines for the inducible expression of FLAG-tagged C11orf46, SETDB1, and SETDB2, the Flp-In T-REx 293 system (Invitrogen) was used as described.⁶⁴ Cells were grown in DMEM (Wako) containing 5% FBS (Invitrogen) and penicillin-streptomycin (Invitrogen; 1% of stock solution) in 5% CO₂ at 37°C. For protein induction, 1 μg/mL of doxycycline (Sigma) was added and the cell culture was incubated for 24 hours.

Immunoprecipitation and mass spectrometry

Protein extraction, immunoprecipitation, and mass spectrometry were performed as described.⁶⁴ In brief, a cell extract was prepared using CSK buffer containing 0.5 M sodium chloride and incubated with anti-FLAG agarose beads (Sigma). The bound proteins were eluted by incubation with 0.2 mg/mL of 3×FLAG peptide (Sigma). Data analysis for LC/MS/MS was performed as described previously⁶⁴ with minor modifications. The International Protein Index database (Human, version 3.87; <ftp://ftp.ebi.ac.uk/pub/databases/IPI>) was used as a reference. The number of unique spectra for each protein was counted from data summarizing two measurements from all pieces of a gel and converted to the exponentially modified protein abundance index (emPAI) value.⁶⁵

Plasmid construction for bacterial co-expression

DNA fragments encoding full-length C11orf46 and a SETDB2 fragment containing MBD (residues S136–V250) were inserted into the pETDuet-1 vector (Novagen) using the BamHI/EcoRI and NdeI/XhoI restriction sites, respectively (pETDuet His₆-C11orf46/SETDB2^{136–250}). Next, cDNA fragments encoding His₆-tagged full-length C11orf46 and SETDB1 MBD (residues R565–K674) were cloned into the NdeI/XhoI and NcoI/NotI sites, respectively, of pETDuet-1 (pETDuet SETDB1 MBD/His₆-C11orf46). Co-expression plasmids for each deletion variant of C11orf46 were generated by inverse PCR based on pETDuet His₆-

1
2
3
4 C11orf46/SETDB2^{136–250} or pETDuet SETDB1 MBD/His₆-C11orf46. In the plasmid used for co-
5 expression of C11orf46 CRD (residues Q186–G260) with SETDB2 MBD (residues E133–E239)
6 or SETDB1 MBD (residues K565–K674), a human rhinovirus 3C protease (HRV3C) cleavage
7 site was inserted after the N-terminal His₆-tag. Mutations were introduced to these plasmids by
8 site-directed mutagenesis using inverse PCR. All constructed plasmids for bacterial
9 overexpression were amplified in *E. coli* DH5α cells and purified by using a NucleoSpin Plasmid
10 QuickPure kit (Macherey Nagel).
11
12
13
14
15

16 17 **Protein expression and purification**

18 Full-length C11orf46 in complex with SETDB2 was expressed in *E. coli* strain BL21(DE3).
19 Cells transformed with pETDuet His₆-C11orf46/SETDB2^{136–250} were grown at 37°C in LB
20 medium containing 200 mg/L of ampicillin. The temperature was lowered to 25°C when the
21 optical density at 600 nm (OD₆₀₀) reached 0.2–0.4. At OD₆₀₀ 0.6–0.8, isopropyl β-D-1-
22 thiogalactopyranoside and zinc acetate were added at final concentrations of 0.5 mM and 20 μM,
23 respectively. After protein induction for 20 hours, the cells were harvested by centrifugation for
24 20 min at 5180×g and 4°C, and stored at –80°C until use. Expression of the complex of SETDB2
25 MBD or SETDB1 MBD with each C11orf46 mutant was performed in the same way, except that
26 protein expression was induced at 17°C and zinc chloride was used instead of zinc acetate. For
27 preparation of isotope-labeled proteins, the cells were grown in modified M9 medium containing
28 1 g/L of U-¹⁵N ammonium chloride (Cambridge Isotope Laboratories) as the sole source of
29 nitrogen.
30
31
32
33
34
35
36
37

38 All protein purification steps were performed at 4°C unless otherwise specified. The pH
39 of buffer stock solutions was adjusted at room temperature. The SETDB2 MBD–C11orf46 CRD
40 complex was purified as follows. The cells were resuspended in lysis buffer containing 50 mM
41 Tris-HCl (pH 8.0), 300 mM sodium chloride, 5% glycerol, 0.5 mM Tris(2-
42 carboxyethyl)phosphine hydrochloride (TCEP), and 20 mM imidazole, supplemented with
43 cOmplete protease inhibitor cocktail (Roche) and 10 μM zinc chloride. The cells were sonicated
44 on ice and the lysate was clarified by centrifugation at 48384×g for 30 min. After filtration
45 through a 0.45-μm membrane, the supernatant was loaded onto a HisTrap HP column (Cytiva)
46 equilibrated with lysis buffer. The column was washed, and bound proteins were eluted with 20
47 mM Tris-HCl buffer (pH 8.0) containing 50 mM sodium chloride, 5% glycerol, 0.5 mM TCEP,
48 and 500 mM imidazole. After overnight cleavage of the His₆-tag with HRV3C protease, the
49 protein solution was applied to a HiTrap Q HP column (Cytiva) equilibrated with 20 mM Tris-
50 HCl buffer (pH 8.8) containing 20 mM sodium chloride, 0.5 mM TCEP, and 2 mM EDTA, and
51 then eluted by a sodium chloride gradient from 20 mM to 170 mM. Fractions containing the
52 SETDB2 MBD–C11orf46 CRD complex were combined and subjected to further purification by
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 gel filtration chromatography using a HiLoad 16/600 Superdex 75 pg column (Cytiva)
5 equilibrated with 20 mM HEPES-NaOH buffer (pH 7.4), containing 100 mM sodium chloride
6 and 1 mM TCEP. The final fractions containing the purified SETDB2 MBD–C11orf46 CRD
7 complex were concentrated to 21–23 mg/mL using an Amicon Ultra 10,000 MWCO filter
8 (Merck Millipore). The purity of the final fractions was verified by SDS-PAGE analysis.
9

10
11
12 The SETDB1 MBD–C11orf46 CRD complex was purified as described above except that
13 in the ion exchange step, in which the protein solution was loaded onto a HiTrap SP HP column
14 (Cytiva) equilibrated with 20 mM HEPES-NaOH buffer (pH 7.4) containing 50 mM sodium
15 chloride, 5% glycerol, 0.5 mM TCEP, and 1 mM EDTA, and eluted by a sodium chloride
16 gradient from 50 mM to 340 mM.
17
18

19
20 For co-expression and affinity purification on a small scale, BL21(DE3) cells expressing
21 each complex were obtained from 10 mL of culture and sonicated in lysis buffer (2 mL of buffer
22 per 50 mg of wet cell pellet) on ice. After centrifugation, 500 μ L of supernatant was mixed with
23 50 μ L of Ni-NTA agarose (Qiagen) equilibrated with lysis buffer, incubated for 10 min with
24 rotation, and then centrifuged at 1000 \times g for 2 min at 4°C. The supernatant was removed, and the
25 resin was washed three times with 500 μ L of lysis buffer. Aliquots of cell lysate supernatant and
26 Ni-NTA agarose resin were boiled at 95°C for 5 min in sample loading buffer and analyzed by
27 SDS-PAGE using a 15% polyacrylamide gel. Gel images were acquired using ChemiDoc XRS+
28 system (Bio-Rad) and processed using Image Lab software (Bio-Rad).
29
30
31
32
33
34
35

36 Analytical gel filtration chromatography

37 Analytical gel filtration chromatography of SETDB2 MBD– or SETDB1 MBD–C11orf46 CRD
38 complex was performed at 4°C using a Superdex 200 10/300 GL column (Cytiva) equilibrated
39 with 20 mM HEPES-NaOH buffer (pH 7.4) containing 100 mM sodium chloride and 1 mM
40 TCEP. 0.8-1 mg of each protein sample was applied. Peak fractions were analyzed by MALDI-
41 TOF mass spectrometry. The partition coefficient K_{av} of the protein was calculated as follows:
42
43
44
45

$$46 K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

47
48 where V_c , V_o , and V_e denote the geometric column volume (24 mL), the void volume determined
49 by the elution volume of Blue Dextran 2000 (8.51 mL), and the elution volume of the protein,
50 respectively. For estimation of molecular weight of each complex, a calibration curve was
51 generated by linear regression using thyroglobulin (MW: 670000), γ -globulin (MW: 158000),
52 ovalbumin (MW: 44000), carbonic anhydrase (MW: 29000), myoglobin (MW: 17000),
53 ribonuclease A (MW: 13700), and aprotinin (MW: 6500) as standard proteins.
54
55
56
57
58
59

60 Elemental analysis

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Quantitative ICP-MS analysis of the SETDB2 MBD–C11orf46 CRD complex was performed at IDEA Consultants, Inc., Food and Life-Science Laboratory (Japan). The sample solution, containing 103 nmol of SETDB2 MBD–C11orf46 CRD complex purified from M9 culture, was injected into a 7700x ICP-MS system (Agilent) with the following settings: radio frequency power, 1550 W; reaction mode, helium mode with a flow rate of 4.3 mL/min; carrier gas flow, 1.07 L/min. The amount of zinc, cobalt, and manganese in the sample was quantified by monitoring ^{66}Zn , ^{59}Co , and ^{55}Mn , respectively, using ^{45}Sc as an internal standard. We note that zinc, cobalt, and manganese were present in the M9 medium at a final concentration of 25 μM , 2 μM , and 10 μM , respectively.

Protein crystallization and data collection

Crystallization of the SETDB2 MBD–C11orf46 CRD complex was performed at 20°C. Protein samples at a concentration of 21 mg/mL were centrifuged at 10000×g for 5 minutes at 4°C prior to preparing the crystallization droplets. Crystallization screening was performed by the sitting drop vapor diffusion method using three screening kits: Index HT, Crystal Screen HT, and PEG/Ion HT (Hampton Research). Single crystals of the complex were obtained from drops containing 0.1 M buffer component (pH 5.5–8.5) and 2 M ammonium sulfate (Index HT reagents A3–A6) within 5 days. Crystals were soaked in crystallization buffer containing 25% glycerol as a cryoprotectant and flash-frozen in liquid nitrogen. X-ray diffraction data were collected with a wavelength of 1.2800 Å or 1.0000 Å at the beamlines BL41XU or BL32XU in SPring-8 (Japan), respectively. Data processing was automatically performed by the program KAMO⁶⁶ installed in the beamlines. Crystallographic data are summarized in Table 1.

Structure determination and refinement

Structure determination and refinement were performed by the CCP4 software package using the CCP4i2 graphical interface.⁶⁷ The unmerged diffraction data were analyzed and merged by Pointless and Aimless.^{68,69} The initial phases were obtained by the single-wavelength anomalous diffraction technique with intrinsic zinc atoms using the programs SHELX C/D/E.⁷⁰ Diffraction data collected with a wavelength of 1.2800 Å to observe anomalous signals of zinc atoms was used for phase determination. The initial model was automatically built using Buccaneer⁷¹ and further refined by several cycles of manual correction on Coot⁷² and computational refinement using Refmac5.⁷³ The final model was refined against the diffraction data collected with a wavelength of 1.0000 Å. The stereochemistry of the final model was assessed by MolProbity.⁷⁴ The data collection and refinement statistics are shown in Table 1. Unless otherwise denoted, structural characteristics of the protein complex are found in both complexes in the asymmetric unit. The electrostatic potential maps of the SETDB2 MBD, MeCP2 MBD, BAZ2A TAM

1
2
3
4 domains were calculated using APBS software.⁷⁵ All graphics of molecular structure were
5 prepared using PyMOL (Schrödinger, LLC).
6
7
8

9 **Homology modeling**

10 The structural model of the SETDB1 MBD–C11orf46 CRD complex was built using the SWISS-
11 MODEL server⁷⁶ with the atomic coordinates of the SETDB2 MBD–C11orf46 CRD complex as
12 a template. The model calculation was well converged with a Q-mean value of –1.85.
13
14
15
16

17 **Solution NMR spectroscopy**

18 NMR experiments were performed at 30°C on an AVANCE II 700 MHz NMR spectrometer
19 (Bruker) equipped with a TCI cryogenic probe. ¹H-¹⁵N HSQC spectra of ¹⁵N-labeled SETDB2
20 MBD–C11orf46 CRD complex (initial concentration: 0.8 mM) dissolved in 20 mM HEPES-
21 NaOH buffer (pH 7.4) containing 100 mM sodium chloride, 1 mM TCEP, and 5% deuterium
22 oxide (Cambridge Isotope Laboratories) were measured in the absence and presence of an
23 equimolar amount of unlabeled H3 peptide (residues 1–36 with a tryptophan residue at the C-
24 terminus). The expression plasmid for the GST-SUMO-tagged H3 peptide was kindly provided
25 by Prof. Kyohei Arita in Yokohama City University, and the H3 peptide was purified as
26 described previously.⁷⁷ NMR data were processed using NMRPipe,⁷⁸ and the resulting spectra
27 were visualized using CcpNmr Analysis.⁷⁹
28
29
30
31
32
33
34
35

36 **Quantification and statistical analysis**

37 During structure refinement, 5% of total reflections were randomly selected and used for
38 calculation of the R_{free} value. All structural statistics are summarized in Table 1.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

1. Mozzetta, C., Boyarchuk, E., Pontis, J., and Ait-Si-Ali, S. (2015). Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat Rev Mol Cell Biol* *16*, 499–513. 10.1038/nrm4029.
2. Montavon, T., Shukeir, N., Erikson, G., Engist, B., Onishi-Seebacher, M., Ryan, D., Musa, Y., Mittler, G., Meyer, A.G., Genoud, C., et al. (2021). Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun* *12*. 10.1038/s41467-021-24532-8.
3. Padeken, J., Methot, S.P., and Gasser, S.M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol Cell Biol* *23*, 623–640. 10.1038/s41580-022-00483-w.
4. Falandry, C., Fourel, G., Galy, V., Ristriani, T., Horard, B., Bensimon, E., Salles, G., Gilson, E., and Magdinier, F. (2010). CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. *Journal of Biological Chemistry* *285*, 20234–20241. 10.1074/jbc.M109.052399.
5. Xu, P.-F., Zhu, K.-Y., Jin, Y., Chen, Y., Sun, X.-J., Deng, M., Chen, S.-J., Chen, Z., and Liu, T.X. (2010). Setdb2 restricts dorsal organizer territory and regulates left-right asymmetry through suppressing fgf8 activity. *Proceedings of the National Academy of Sciences* *107*, 2521–2526. 10.1073/pnas.0914396107.
6. Ryu, H., Lee, J., Hagerty, S.W., Soh, B.Y., McAlpin, S.E., Cormier, K.A., Smith, K.M., and Ferrante, R.J. (2006). ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington’s disease. *Proceedings of the National Academy of Sciences* *103*, 19176–19181. 10.1073/pnas.0606373103.
7. Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* *464*, 927–931. 10.1038/nature08858.
8. Hogarth, C.A., Mitchell, D., Evanoff, R., Small, C., and Griswold, M. (2011). Identification and expression of potential regulators of the mammalian mitotic-to-meiotic transition. *Biol Reprod* *84*, 34–42. 10.1095/biolreprod.110.086215.
9. Liu, S., Brind’Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev* *28*, 2041–2055. 10.1101/gad.244848.114.
10. Ninova, M., Chen, Y.C.A., Godneeva, B., Rogers, A.K., Luo, Y., Fejes Tóth, K., and Aravin, A.A. (2020). Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target

- 1
2
3
4
5 Recognition to Chromatin Silencing. *Mol Cell* 77, 556-570.e6.
6 10.1016/j.molcel.2019.11.012.
7
8 11. Dodge, J.E., Kang, Y.-K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9
9 Methyltransferase ESET Is Essential for Early Development. *Mol Cell Biol* 24, 2478–
10 2486. 10.1128/MCB.24.6.2478-2486.2004.
11
12 12. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai,
13 Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and
14 SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and
15 Chimeric Transcripts in mESCs. *Cell Stem Cell* 8, 676–687. 10.1016/j.stem.2011.04.004.
16
17 13. Cruz-Tapias, Robin, Pontis, Maestro, and Ait-Si-Ali (2019). The H3K9 Methylation
18 Writer SETDB1 and its Reader MPP8 Cooperate to Silence Satellite DNA Repeats in
19 Mouse Embryonic Stem Cells. *Genes (Basel)* 10, 750. 10.3390/genes10100750.
20
21 14. Schliehe, C., Flynn, E.K., Vilagos, B., Richson, U., Swaminathan, S., Bosnjak, B., Bauer,
22 L., Kandasamy, R.K., Griesshammer, I.M., Kosack, L., et al. (2015). The
23 methyltransferase Setdb2 mediates virus-induced susceptibility to bacterial superinfection.
24 *Nat Immunol* 16, 67–74. 10.1038/ni.3046.
25
26 15. Kroetz, D.N., Allen, R.M., Schaller, M.A., Cavallaro, C., Ito, T., and Kunkel, S.L. (2015).
27 Type I Interferon Induced Epigenetic Regulation of Macrophages Suppresses Innate and
28 Adaptive Immunity in Acute Respiratory Viral Infection. *PLoS Pathog* 11.
29 10.1371/journal.ppat.1005338.
30
31 16. Melvin, W.J., Audu, C.O., Davis, F.M., Sharma, S.B., Joshi, A., DenDekker, A., Wolf, S.,
32 Barrett, E., Mangum, K., Zhou, X., et al. (2021). Coronavirus induces diabetic
33 macrophage-mediated inflammation via SETDB2. *Proc Natl Acad Sci U S A* 118.
34 10.1073/pnas.2101071118.
35
36 17. Kimball, A.S., Davis, F.M., DenDekker, A., Joshi, A.D., Schaller, M.A., Bermick, J., Xing,
37 X., Burant, C.F., Obi, A.T., Nysz, D., et al. (2019). The Histone Methyltransferase Setdb2
38 Modulates Macrophage Phenotype and Uric Acid Production in Diabetic Wound Repair.
39 *Immunity* 51, 258–271. 10.1016/j.immuni.2019.06.015.
40
41 18. Sun, L., and Fang, J. (2016). E3-Independent Constitutive Monoubiquitination
42 Complements Histone Methyltransferase Activity of SETDB1. *Mol Cell* 62, 958–966.
43 10.1016/j.molcel.2016.04.022.
44
45 19. Ishimoto, K., Kawamata, N., Uchihara, Y., Okubo, M., Fujimoto, R., Gotoh, E.,
46 Kakinouchi, K., Mizohata, E., Hino, N., Okada, Y., et al. (2016). Ubiquitination of Lysine
47 867 of the Human SETDB1 Protein Upregulates Its Histone H3 Lysine 9 (H3K9)
48 Methyltransferase Activity. *PLoS One* 11, e0165766. 10.1371/journal.pone.0165766.
49
50 20. Tsusaka, T., Shimura, C., and Shinkai, Y. (2019). ATF7IP regulates SETDB1 nuclear
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 localization and increases its ubiquitination. *EMBO Rep* 20, e48297.
5 10.15252/embr.201948297.
6
7
8 21. Osumi, K., Sato, K., Murano, K., Siomi, H., and Siomi, M.C. (2019). Essential roles of
9 Windei and nuclear monoubiquitination of Eggless/ SETDB1 in transposon silencing.
10 *EMBO Rep* 20. 10.15252/embr.201948296.
11
12 22. Funyu, T., Kanemaru, Y., Onoda, H., and Arita, K. (2021). Preparation of the
13 ubiquitination-triggered active form of SETDB1 in *Escherichia coli* for biochemical and
14 structural analyses. *The Journal of Biochemistry* 170, 655–662. 10.1093/jb/mvab087.
15
16 23. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and
17 Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal
18 protein that binds to Methylated DNA. *Cell* 69, 905–914. 10.1016/0092-8674(92)90610-
19 O.
20
21 24. Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MeCP2, a vertebrate
22 DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* 20, 5085–
23 5092. 10.1093/nar/20.19.5085.
24
25 25. Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of
26 mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18, 6538–6547.
27 10.1128/mcb.18.11.6538.
28
29 26. Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution
30 structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional
31 repressor MBD1. *EMBO J* 18, 6653–6661. 10.1093/emboj/18.23.6653.
32
33 27. Wakefield, R.I.D., Smith, B.O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A.P., and
34 Barlow, P.N. (1999). The solution structure of the domain from MeCP2 that binds to
35 methylated DNA. *J Mol Biol* 291, 1055–1065. 10.1006/jmbi.1999.3023.
36
37 28. Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M.
38 (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in
39 complex with methylated DNA. *Cell* 105, 487–497. 10.1016/s0092-8674(01)00324-5.
40
41 29. Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D.
42 (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell* 29,
43 525–531. 10.1016/j.molcel.2007.12.028.
44
45 30. Hendrich, B., Hardeland, U., Ng, H.-H., Jiricny, J., and Bird, A. (1999). The thymine
46 glycosylase MBD4 can bind to the product of deamination at methylated CpG sites.
47 *Nature* 401, 301–304. 10.1038/45843.
48
49 31. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang,
50 X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and
51 erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 40, 4841–4849.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- 10.1093/nar/gks155.
32. Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013). Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J Biol Chem* *288*, 6351–6362. 10.1074/jbc.M112.431098.
33. Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* *20*, 116–117. 10.1038/2413.
34. Ng, H.H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. *Mol Cell Biol* *20*, 1394–1406. 10.1128/MCB.20.4.1394-1406.2000.
35. Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA Methylation-Related Chromatin Remodeling in Activity-Dependent *Bdnf* Gene Regulation. *Science* (1979) *302*, 890–893. 10.1126/science.1090842.
36. Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H., and Nakao, M. (2003). MCAF Mediates MBD1-Dependent Transcriptional Repression. *Mol Cell Biol* *23*, 2834–2843. 10.1128/MCB.23.8.2834-2843.2003.
37. Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* *19*, 269–277. 10.1016/S0168-9525(03)00080-5.
38. Roloff, T.C., Ropers, H.H., and Nuber, U.A. (2003). Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics* *4*. 10.1186/1471-2164-4-1.
39. Mayer, C., Schmitz, K.M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. *Mol Cell* *22*, 351–361. 10.1016/j.molcel.2006.03.028.
40. Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* *24*, 2264–2269. 10.1101/gad.590910.
41. Santoro, R., Schmitz, K.M., Sandoval, J., and Grummt, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep* *11*, 52–58. 10.1038/embor.2009.254.
42. Anosova, I., Melnik, S., Tripsianes, K., Kateb, F., Grummt, I., and Sattler, M. (2015). A novel RNA binding surface of the TAM domain of TIP5/BAZ2A mediates epigenetic regulation of rRNA genes. *Nucleic Acids Res* *43*, 5208–5220. 10.1093/nar/gkv365.
43. Chen, S., Zhou, M., Dong, A., Loppnau, P., Wang, M., Min, J., and Liu, K. (2021). Structural basis of the TAM domain of BAZ2A in binding to DNA or RNA independent of

- methylation status. *Journal of Biological Chemistry* 297, 101351.
10.1016/j.jbc.2021.101351.
44. Feng, Y., Chen, S., Zhou, M., Zhang, J., Min, J., and Liu, K. (2022). Crystal structure of the BAZ2B TAM domain. *Heliyon* 8. 10.1016/j.heliyon.2022.e09873.
45. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16, 919–932. 10.1101/gad.973302.
46. Maeda, R., and Tachibana, M. (2022). HP1 maintains protein stability of H3K9 methyltransferases and demethylases. *EMBO Rep* 23. 10.15252/embr.202153581.
47. Rolland, T., Taşan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226. 10.1016/j.cell.2014.10.050.
48. Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* 163, 712–723. 10.1016/j.cell.2015.09.053.
49. Peter, C.J., Saito, A., Hasegawa, Y., Tanaka, Y., Nagpal, M., Perez, G., Alway, E., Espeso-Gil, S., Fayyad, T., Ratner, C., et al. (2019). In vivo epigenetic editing of *Sema6a* promoter reverses transcallosal dysconnectivity caused by *C11orf46/Ar114ep* risk gene. *Nat Commun* 10, 4112. 10.1038/s41467-019-12013-y.
50. Saito, M., and Ishikawa, F. (2002). The mCpG-binding Domain of Human MBD3 Does Not Bind to mCpG but Interacts with NuRD/Mi2 Components HDAC1 and MTA2. *Journal of Biological Chemistry* 277, 35434–35439. 10.1074/jbc.M203455200.
51. Zemach, A., Li, Y., Wayburn, B., Ben-Meir, H., Kiss, V., Avivi, Y., Kalchenko, V., Jacobsen, S.E., and Grafi, G. (2005). DDM1 binds Arabidopsis methyl-CpG binding domain proteins and affects their subnuclear localization. *Plant Cell* 17, 1549–1558. 10.1105/tpc.105.031567.
52. Baymaz, H.I., Fournier, A., Laget, S., Ji, Z., Jansen, P.W.T.C., Smits, A.H., Ferry, L., Mensinga, A., Poser, I., Sharrocks, A., et al. (2014). MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* 14, 2179–2189. 10.1002/pmic.201400013.
53. Buck-Koehntop, B.A., and Defossez, P.-A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics* 8, 131–137. 10.4161/epi.23632.
54. Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res* 50, W210–W215. 10.1093/nar/gkac387.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
55. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. *Nat Biotechnol.* 10.1038/s41587-023-01773-0.
 56. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 478, 57–63. 10.1038/nature10423.
 57. Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-methyladenine DNA modification in *Drosophila*. *Cell* 161, 893–906. 10.1016/j.cell.2015.04.018.
 58. Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.H., Aravind, L., He, C., and Shi, Y. (2015). DNA methylation on N6-adenine in *C. elegans*. *Cell* 161, 868–878. 10.1016/j.cell.2015.04.005.
 59. Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y.R., and Almouzni, G. (2009). The HP1 α -CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep* 10, 769–775. 10.1038/embor.2009.90.
 60. Minkovsky, A., Sahakyan, A., Rankin-Gee, E., Bonora, G., Patel, S., and Plath, K. (2014). The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. *Epigenetics Chromatin* 7. 10.1186/1756-8935-7-12.
 61. Tchakovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* (1979) 348, 1481–1485. 10.1126/science.aaa7227.
 62. Mutlu, B., Chen, H.-M., Moresco, J.J., Orelo, B.D., Yang, B., Gaspar, J.M., Keppler-Ross, S., Yates, J.R., Hall, D.H., Maine, E.M., et al. (2018). Regulated nuclear accumulation of a histone methyltransferase times the onset of heterochromatin formation in *C. elegans* embryos. *Sci Adv* 4, eaat6224. 10.1126/sciadv.aat6224.
 63. Delaney, C.E., Methot, S.P., Guidi, M., Katic, I., Gasser, S.M., and Padeken, J. (2019). Heterochromatic foci and transcriptional repression by an unstructured MET-2/SETDB1 co-factor LIN-65. *Journal of Cell Biology* 218, 820–838. 10.1083/jcb.201811038.
 64. Nozawa, R.-S., Nagao, K., Masuda, H.-T., Iwasaki, O., Hirota, T., Nozaki, N., Kimura, H., and Obuse, C. (2010). Human POGZ modulates dissociation of HP1 α from mitotic chromosome arms through Aurora B activation. *Nat Cell Biol* 12, 719–727. 10.1038/ncb2075.
 65. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially Modified Protein Abundance Index (emPAI) for Estimation of

- 1
2
3
4 Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per
5 Protein. *Molecular & Cellular Proteomics* 4, 1265–1272. 10.1074/mcp.M500061-
6 MCP200.
7
8
9 66. Yamashita, K., Hirata, K., and Yamamoto, M. (2018). KAMO: towards automated data
10 processing for microcrystals. *Acta Crystallogr D Struct Biol* 74, 441–449.
11 10.1107/S2059798318004576.
12
13 67. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan,
14 R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4
15 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235–242.
16 10.1107/S0907444910045749.
17
18 68. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol*
19 *Crystallogr* 62, 72–82. 10.1107/S0907444905036693.
20
21 69. Evans, P.R. (2011). An introduction to data reduction: Space-group determination, scaling
22 and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282–292.
23 10.1107/S090744491003982X.
24
25 70. Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: Combining chain
26 tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66, 479–485.
27 10.1107/S0907444909038360.
28
29 71. Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing
30 protein chains. *Acta Crystallogr D Biol Crystallogr* 62, 1002–1011.
31 10.1107/S0907444906022116.
32
33 72. Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics.
34 *Acta Crystallogr D Biol Crystallogr* 60, 2126–2132. 10.1107/S0907444904019158.
35
36 73. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A.,
37 Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of
38 macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67, 355–367.
39 10.1107/S0907444911001314.
40
41 74. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray,
42 L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: All-atom
43 contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35.
44 10.1093/nar/gkm216.
45
46 75. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H.,
47 Wilson, L., Chen, J., Liles, K., et al. (2018). Improvements to the APBS biomolecular
48 solvation software suite. *Protein Science* 27, 112–128. 10.1002/pro.3280.
49
50 76. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer,
51 F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: Homology
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 modelling of protein structures and complexes. *Nucleic Acids Res* 46, W296–W303.
5 10.1093/nar/gky427.

- 6
7
8 77. Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai,
9 N., Matsumura, R., Kawakami, T., Mishima, Y., et al. (2017). Structure of the Dnmt1
10 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3
11 Reveals the Basis for DNA Methylation Maintenance. *Mol Cell* 68, 350-360.e7.
12 10.1016/j.molcel.2017.09.037.
13
14
15 78. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995).
16 NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol*
17 *NMR* 6, 277–293. 10.1007/BF00197809.
18
19
20 79. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L.,
21 Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR
22 spectroscopy: development of a software pipeline. *Proteins* 59, 687–696.
23 10.1002/prot.20449.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

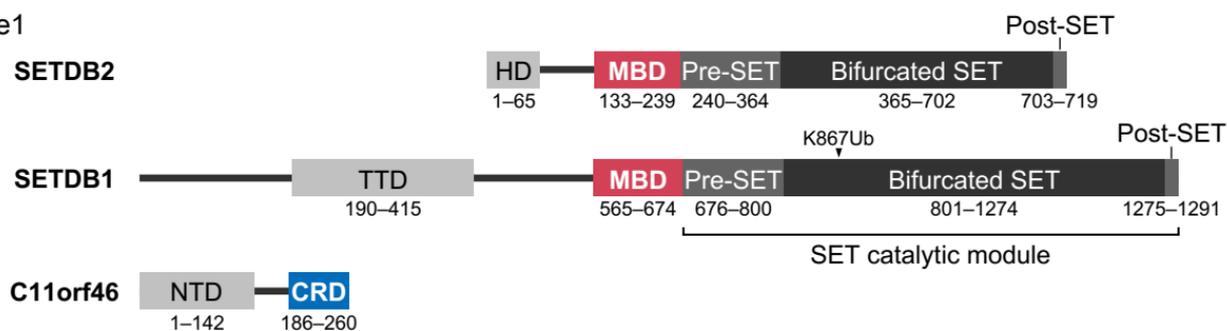
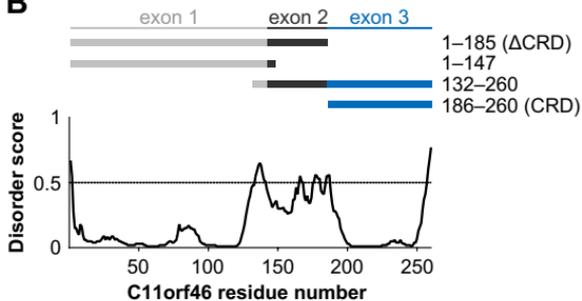
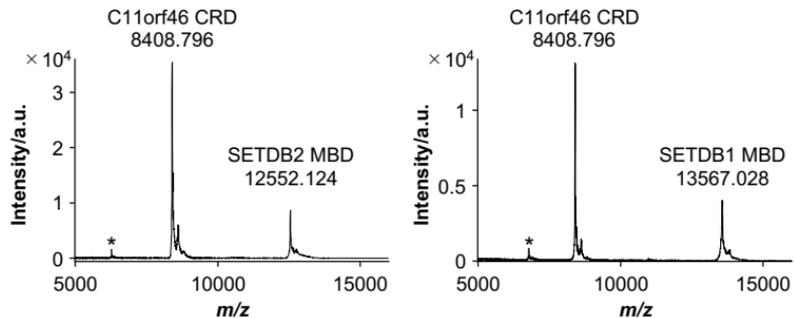
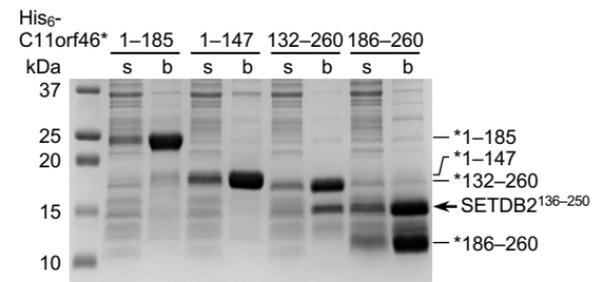
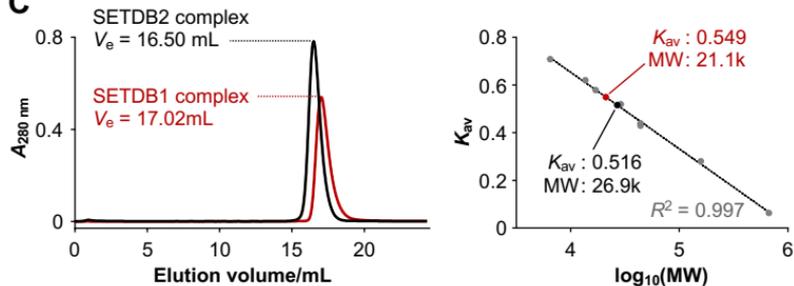
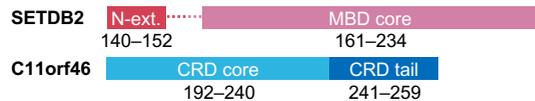
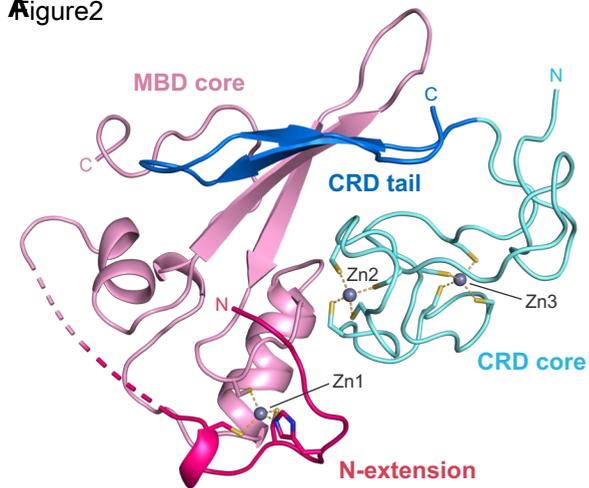
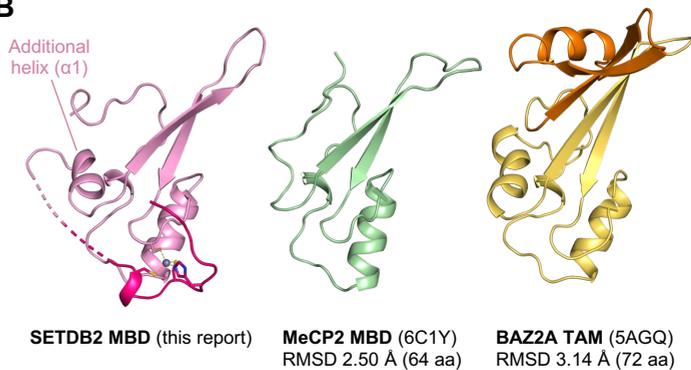
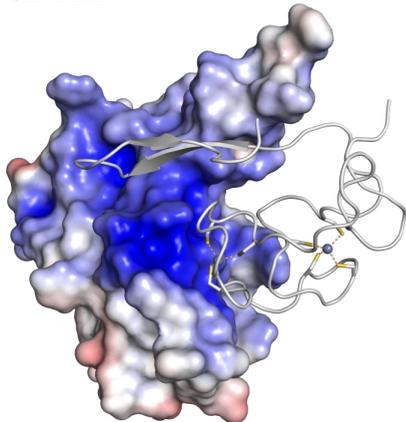
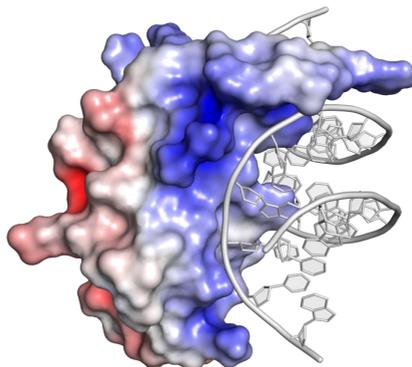
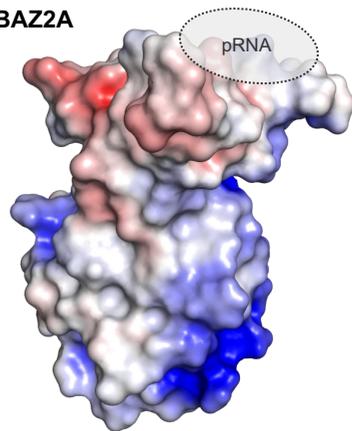
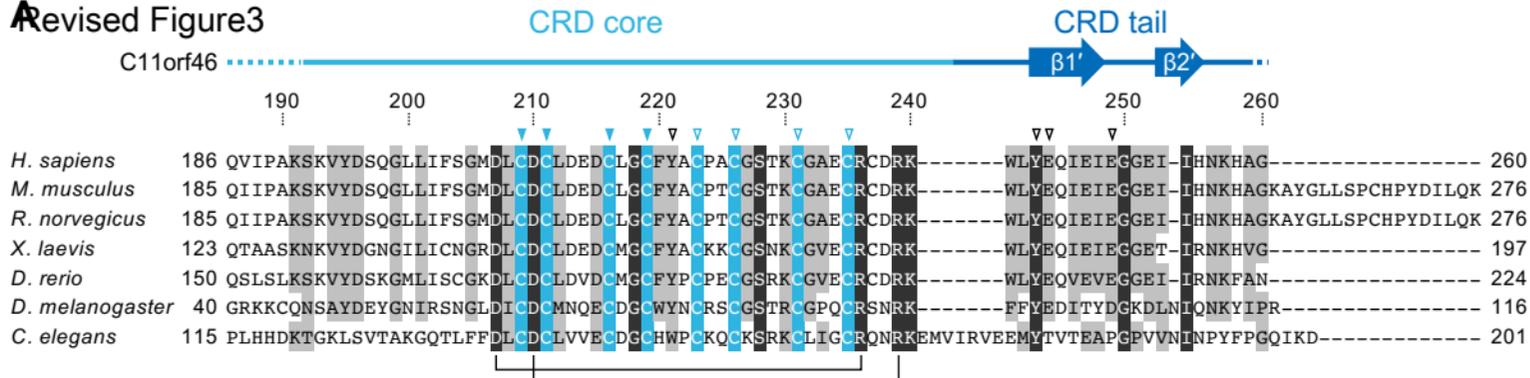
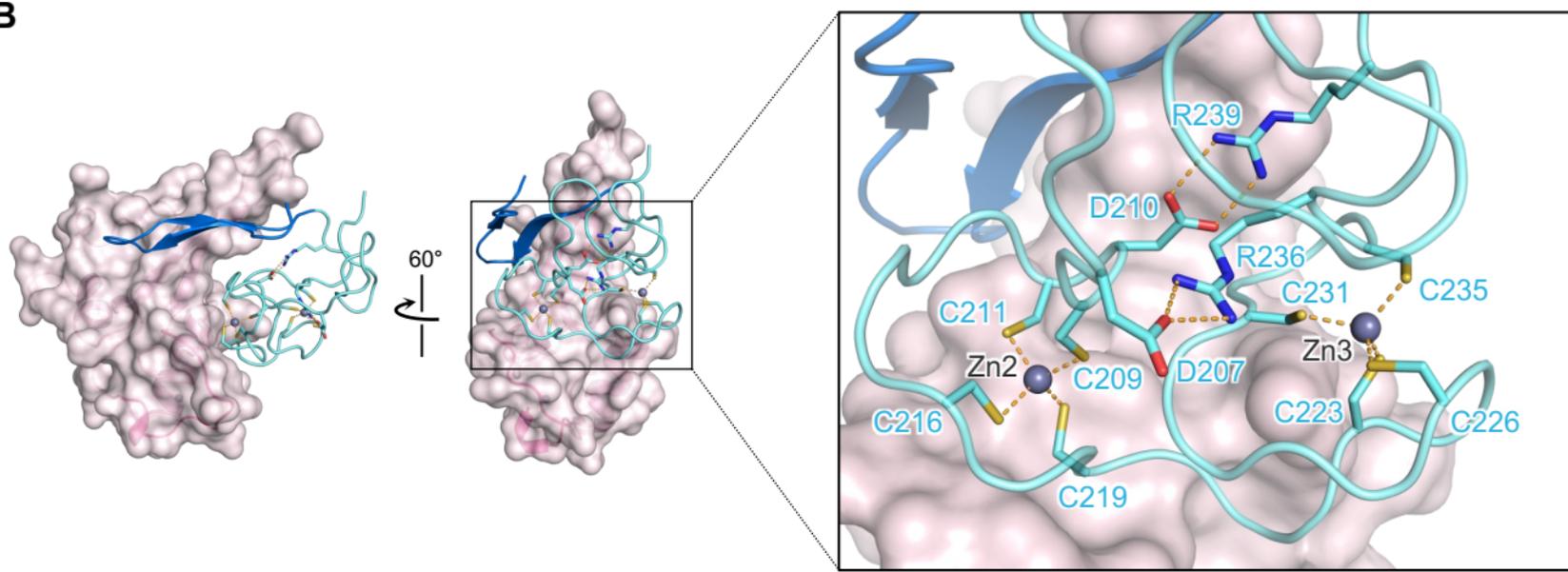
Figure 1**B****C**

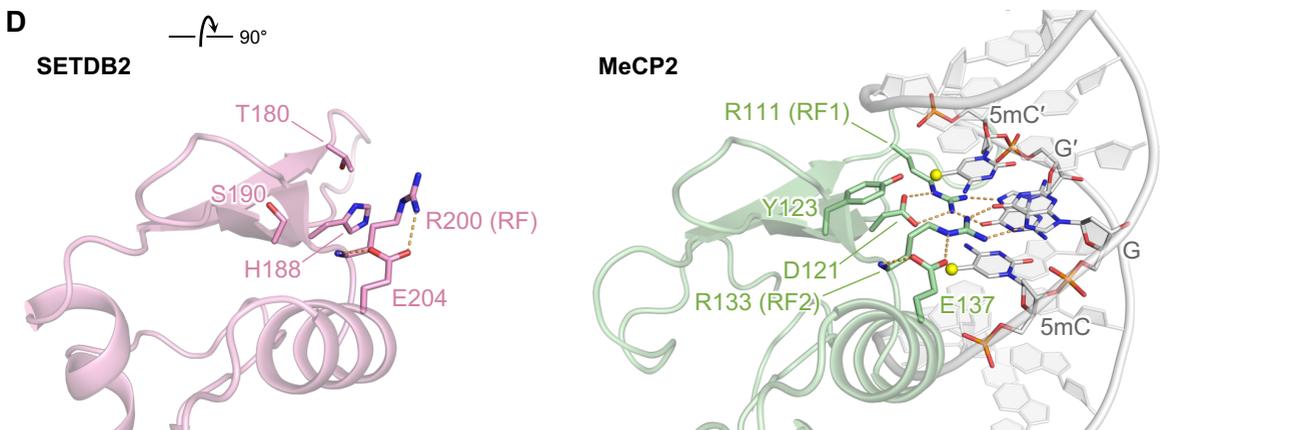
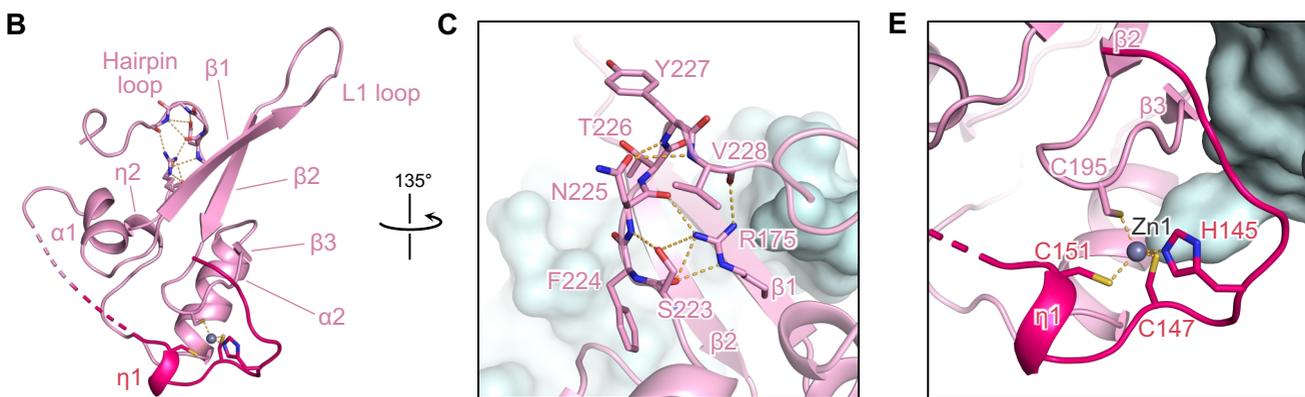
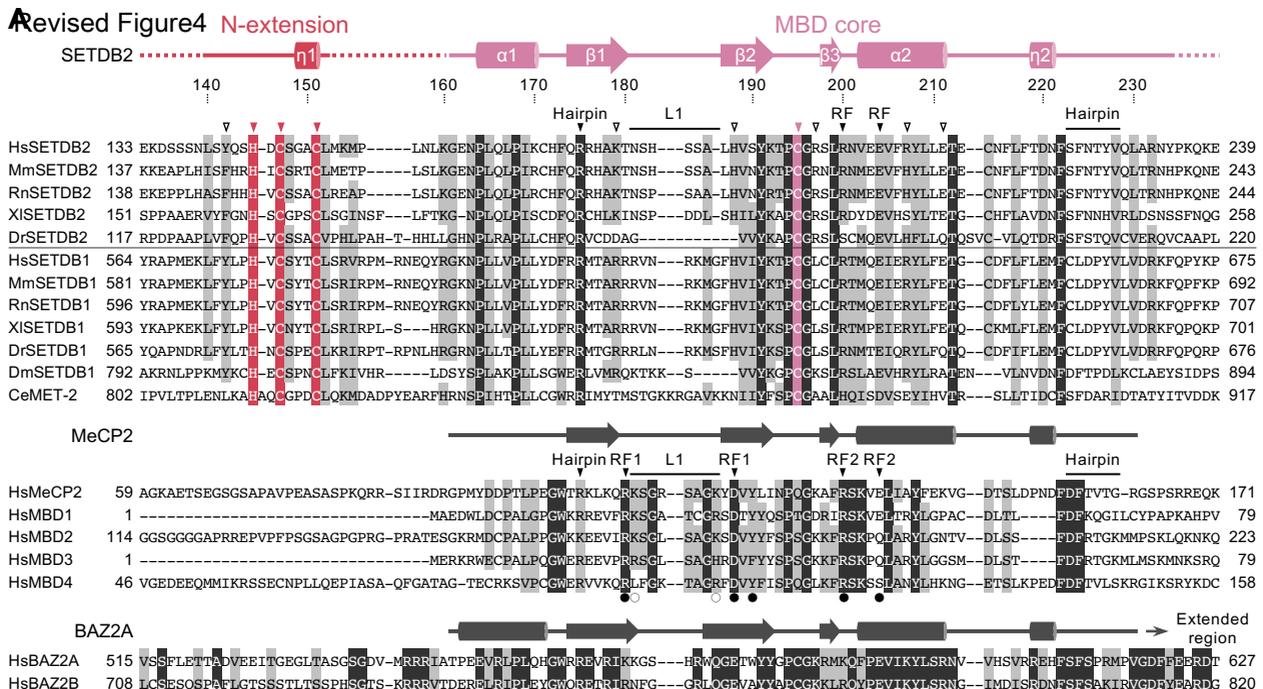
Figure 2**B****C****SETDB2****MeCP2****BAZ2A**

Revised Figure3

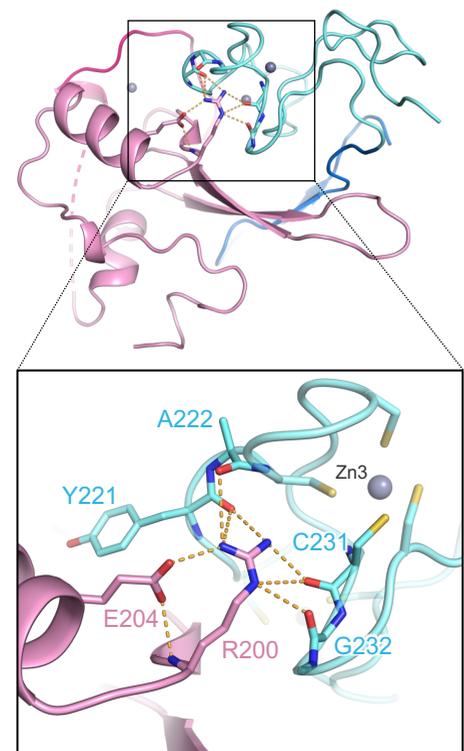


B

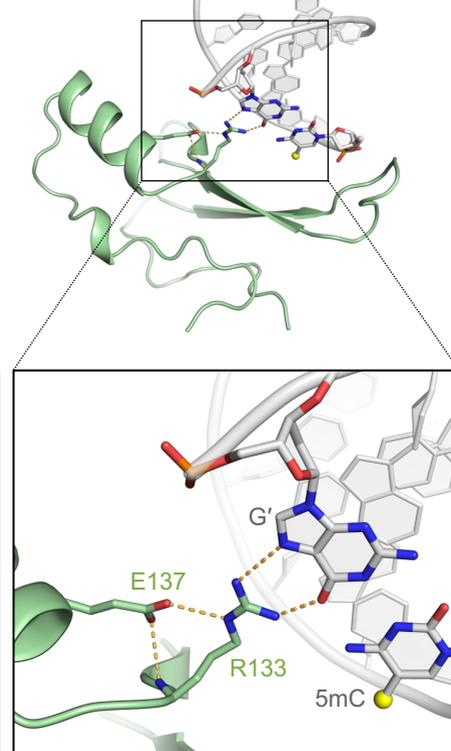




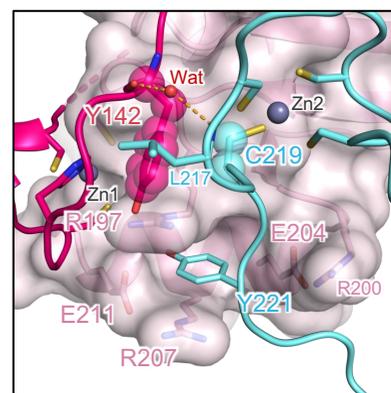
A Revised Figure5
SETDB2



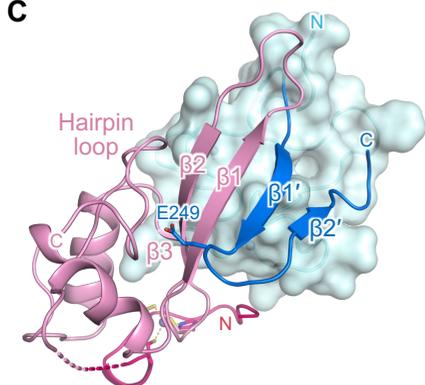
MeCP2



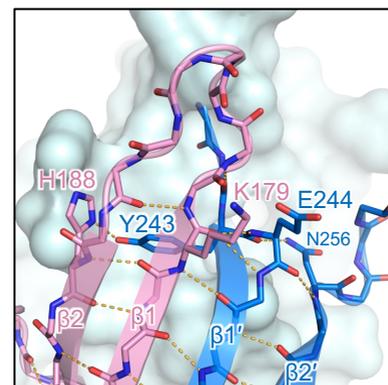
B



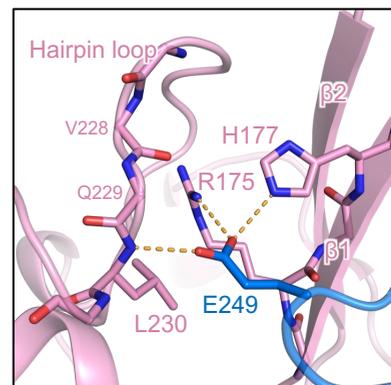
C



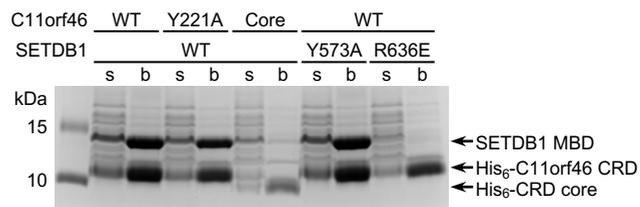
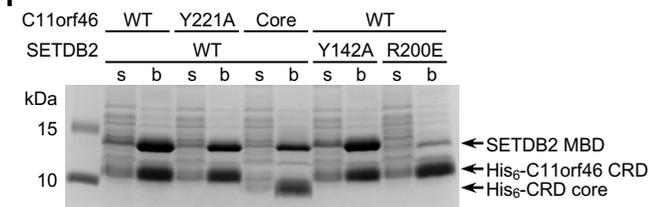
D



E



F

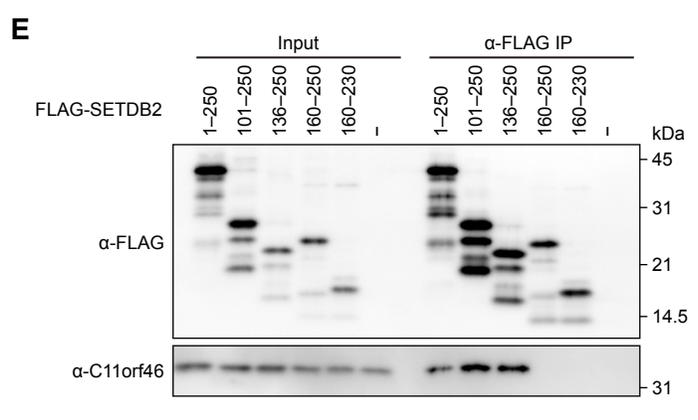
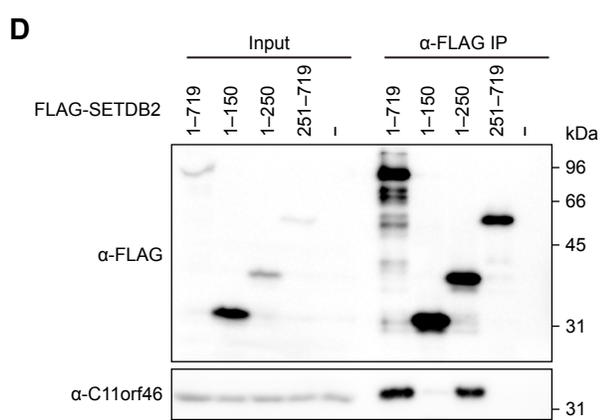
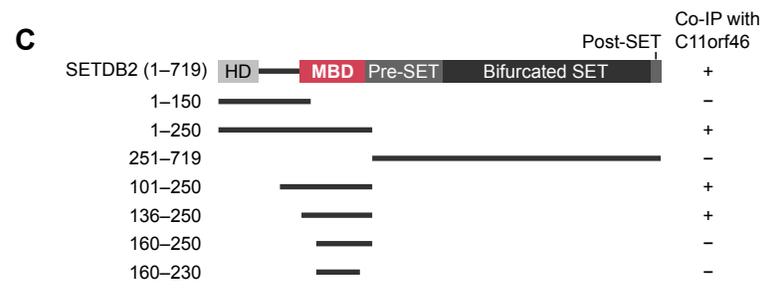
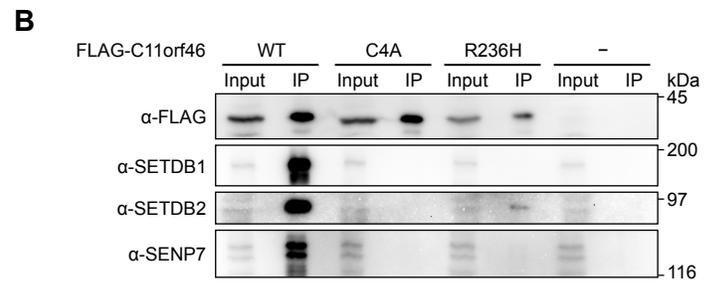
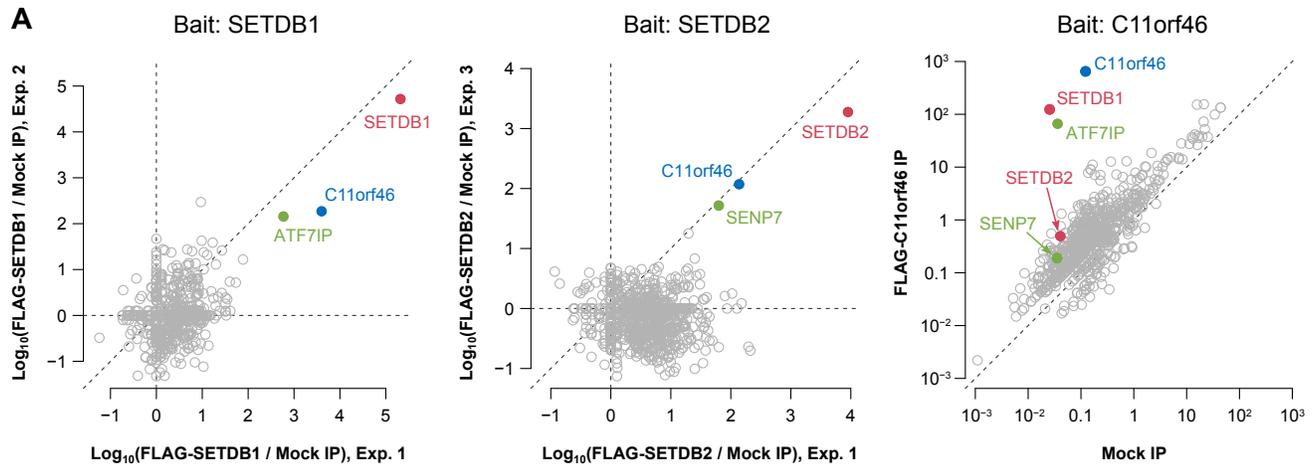


KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-C11orf46	This paper	N/A
Mouse monoclonal anti-SEN7	This paper	N/A
Mouse monoclonal anti-FLAG	Sigma-Aldrich	Cat#F1804; RRID: AB_262044
Rabbit monoclonal anti-SETDB1	Cell Signaling Technology	Cat#2196; RRID: AB_823637
Mouse monoclonal anti-SETDB2	Abnova	Cat#H00083852-M17; RRID: AB_1713857
Bacterial and virus strains		
<i>Escherichia coli</i> : DH5 α	Takara	Cat#9057
<i>Escherichia coli</i> : BL21(DE3)	Novagen	Cat#69450
Biological samples		
Chemicals, peptides, and recombinant proteins		
Doxycycline	Sigma-Aldrich	Cat#D9891
3xFLAG peptide	Sigma-Aldrich	Cat#F4799
Isopropyl β -D-1-thiogalactopyranoside	Nacalai Tesque	Cat#19742-94
Tris(2-carboxyethyl)phosphine hydrochloride	Nacalai Tesque	Cat#07277-16
cOmplete protease inhibitor cocktail	Roche	Cat#11836170001
HRV3C protease	This paper	N/A
Critical commercial assays		
Index HT	Hampton Research	Cat#HR2-134
Crystal Screen HT	Hampton Research	Cat#HR2-130
PEG/Ion HT	Hampton Research	Cat#HR2-139
Gel Filtration Calibration Kit LMW	Cytiva	Cat#28403841
Gel Filtration Standard	Bio-Rad	Cat#1511901
Deposited data		
Atomic coordinates of the SETDB2 MBD-C11orf46 CRD complex	This paper	Protein Data Bank: 8HFP
Experimental models: Cell lines		
Human: Flp-In T-REx 293 cells	Invitrogen	Cat#R78007
Experimental models: Organisms/strains		
Oligonucleotides		
Recombinant DNA		
pcDNA5/FRT/TO 3xFLAG-C11orf46 and mutants	This paper	N/A
pcDNA5/FRT/TO 3xFLAG-SETDB1	This paper	N/A
pcDNA5/FRT/TO 3xFLAG-SETDB2	This paper	N/A
pETDuet His ₆ -C11orf46/SETDB2 S136-V250	This paper	N/A
pETDuet His ₆ -C11orf46 CRD/SETDB2 MBD and mutants	This paper	N/A
pETDuet SETDB1 MBD/His ₆ -C11orf46	This paper	N/A
pETDuet SETDB1 MBD/His ₆ -C11orf46 CRD and mutants	This paper	N/A
pGEX-ST-1 H3 A1-K36+W	Ishiyama et al. ⁷⁷	N/A
Software and algorithms		

Dali	Holm ⁵⁴	http://ekhidna2.biocebioc.helsinki.fi/dali/ ; RRID: SCR_013433
Foldseek	van Kempen et al. ⁵⁵	https://search.foldsefo.com/search
Image Lab v4.1	Bio-Rad	RRID: SCR_014210
KAMO	Yamashita et al. ⁶⁶	N/A
CCP4	Winn et al. ⁶⁷	https://www.ccp4.ac.uk/ ; RRID: SCR_007255
Pointless	Evans ^{68,69}	https://www.ccp4.ac.uk/html/pointless.html ; RRID: SCR_014218
Aimless	Evans ^{68,69}	https://www.ccp4.ac.uk/html/aimless.html ; RRID: SCR_015747
SHELX C/D/E	Sheldrick ⁷⁰	http://shelx.uni-ac.gwdg.de/SHELX/ ; RRID: SCR_014220
Buccaneer	Cowtan ⁷¹	https://www.ccp4.ac.uk/html/cbuccaneer.html ; RRID: SCR_014221
Coot v0.9.3	Emsley and Cowtan ⁷²	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/cool/ ; RRID: SCR_014222
Refmac5	Murshudov et al. ⁷³	https://www.ccp4.ac.uk/html/refmac5.html ; RRID: SCR_014225
MolProbity	Davis et al. ⁷⁴	http://molprobity.biocchem.duke.edu/ ; RRID: SCR_014226
APBS	Jurrus et al. ⁷⁵	https://server.poissonboltzmann.org/apbs ; RRID: SCR_008387
PyMOL v2.0.4	Schrödinger, LLC	https://pymol.org/2/ ; RRID: SCR_000305
SWISS-MODEL	Waterhouse et al. ⁷⁶	https://swissmodel.expasy.org/ ; RRID: SCR_018123
NMRPipe	Delaglio et al. ⁷⁸	https://www.ibbr.umd.edu/nmrpipe/index.html
CcpNmr Analysis v2.5.2	Vranken et al. ⁷⁹	https://ccpn.ac.uk/ ; RRID: SCR_016984
Other		
Anti-FLAG agarose beads	Sigma-Aldrich	Cat#A2220
Ni-NTA agarose	Qiagen	Cat#30230
Superdex 200 10/300 GL	Cytiva	Cat#17517501

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



1
2
3
4 **Figure S1. Identification of C11orf46 as a SETDB-binding factor in human cells,**
5 **related to Figure 1.**
6

7 (A) Proteins enriched in FLAG-SETDB1, FLAG-SETDB2, and FLAG-C11orf46
8 immunoprecipitates assessed by mass spectrometry. Left and middle, scatter plots
9 showing the log₁₀-transformed ratios of emPAI values for proteins immunoprecipitated
10 from cells expressing FLAG-SETDB1 (left) or FLAG-SETDB2 (middle) to those from
11 parental cells (mock) in two independent experiments. Right, scatter plot showing
12 emPAI values for proteins immunoprecipitated from FLAG-C11orf46-expressing cells
13 and parental control cells.
14
15
16
17

18 (B) T-REx 293 cells stably expressing FLAG-tagged wild-type (WT) C11orf46 or its
19 mutants were subjected to immunoprecipitation using anti-FLAG antibody and analyzed
20 by western blotting using α-FLAG, α-SETDB1, α-SETDB2, and α-SEN7 antibodies.
21 C4A indicates C11orf46 harboring four alanine substitutions at C216, C219, C223, and
22 C226.
23
24
25
26

27 (C) Schematic diagram of the SETDB2 deletion mutants used in the
28 immunoprecipitation assay.
29

30 (D–E) T-REx 293 cells stably (D) or transiently (E) expressing FLAG-tagged SETDB2
31 deletion mutants were subjected to immunoprecipitation using anti-FLAG antibody and
32 analyzed by western blotting using α-FLAG and α-C11orf46 antibodies.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

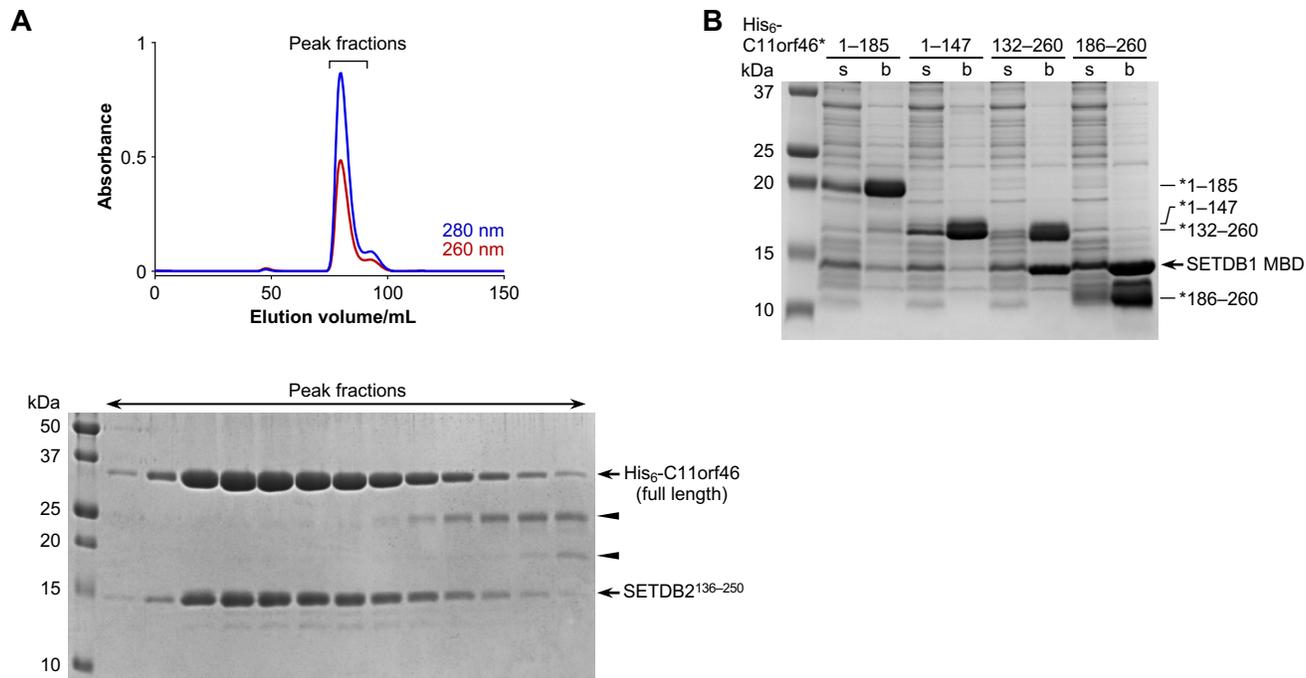


Figure S2. Confirmation of the direct interaction between MBD of SETDB proteins and C11orf46, related to Figure 1.

(A) Purification of full-length C11orf46 as a complex with SETDB2¹³⁶⁻²⁵⁰ by gel filtration chromatography. Black triangles indicate C11orf46 fragments probably cleaved by trace protease activity.

(B) Co-expression of N-terminal hexahistidine-tagged C11orf46 fragments with SETDB1 MBD (residues 565-674). Left lanes, supernatant of cell lysate (s); right lanes, proteins bound to Ni-NTA agarose resin (b).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

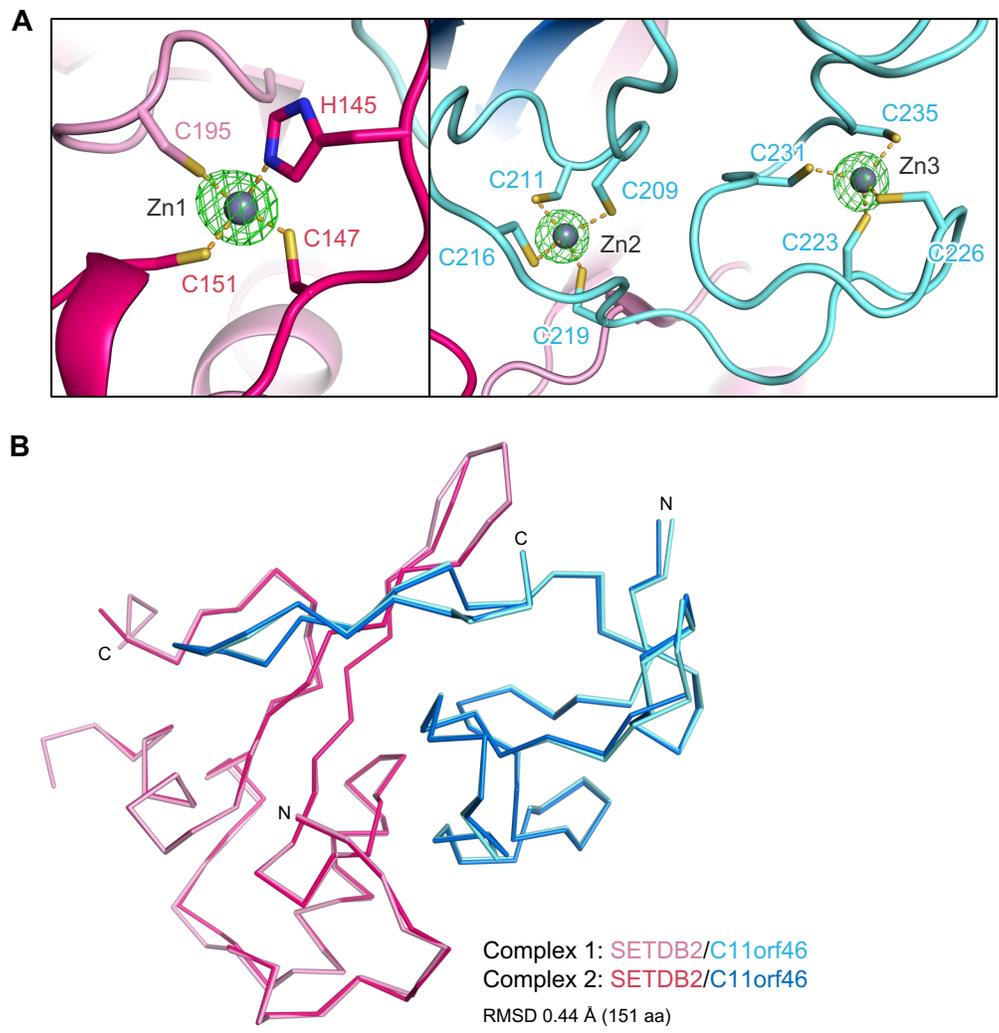


Figure S3. Structural properties of the SETDB2 MBD–C11orf46 CRD complex, related to Figure 2.

(A) Anomalous difference maps of zinc ions found in the crystal structure contoured at 8σ .

(B) Superimposition of the two heterodimeric complexes found in the asymmetric unit. Complexes 1 and 2 are shown as ribbon representations of α -carbon atoms in pale and deep colors, respectively.

Complex 1

Complex 2

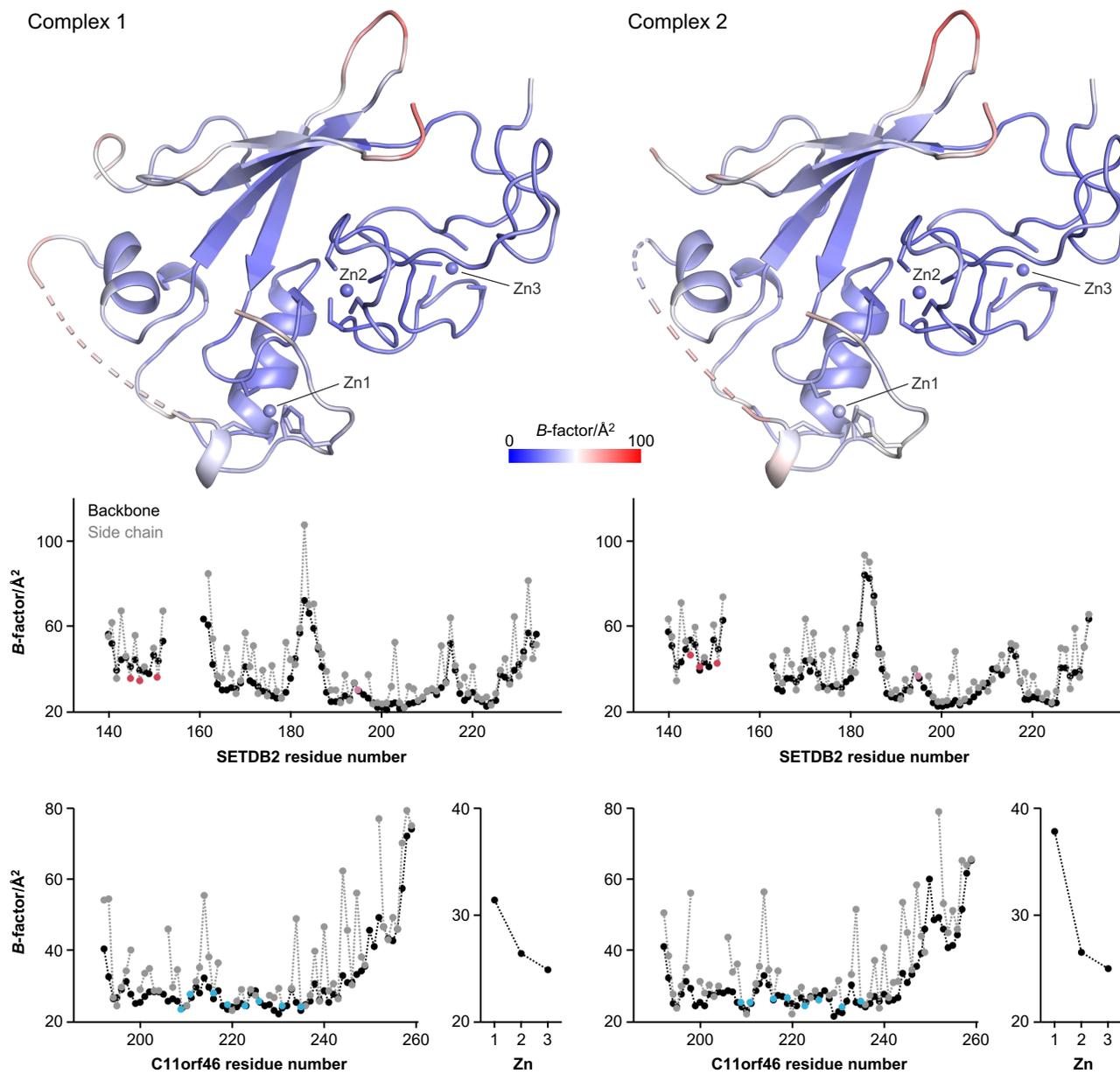
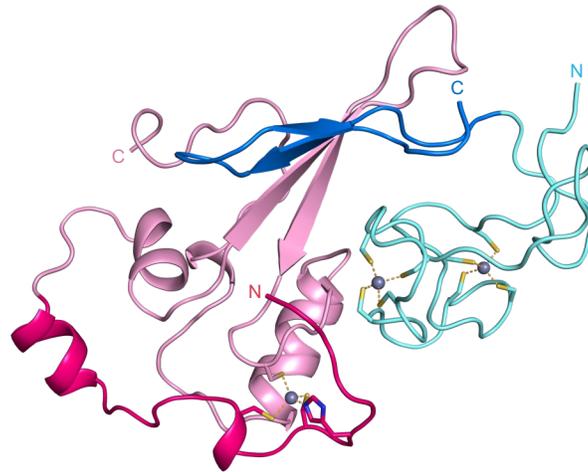


Figure S4. B-factor values of individual residues in the SETDB2 MBD-C11orf46 CRD complex, related to Figures 2-4.

Upper, B-factor-colored structure of the two SETDB2 MBD-C11orf46 CRD complexes in the asymmetric unit. Lower, averaged B-factor values of the backbone and the side chain plotted for each residue. The B-factors of the side chains of zinc-coordinating residues are highlighted in color.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



```
HsSETDB1 564 YRAPMEKLFVLPVCSYTCLSRVRPMPRNEQY 594  
HsSETDB2 133 EKDSSSNLSVQSQDCSGACLMKMP----LNL 159
```

Figure S5. Homology model of the SETDB1 MBD–C11orf46 CRD complex, related to Figures 2 and 4.

The structure is colored as in Figure 2A. The amino acid sequence of the N-extension region of human SETDB proteins is shown at the bottom.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

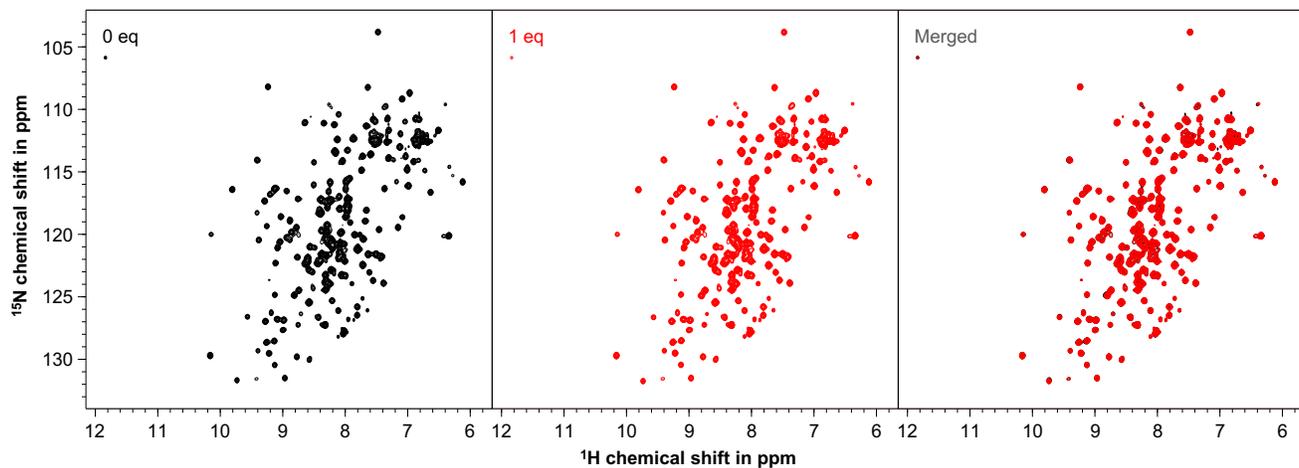


Figure S6. Examination of H3 binding by the SETDB2 MBD–C11orf46 CRD complex using solution NMR spectroscopy, related to Figures 1 and 2.

^1H - ^{15}N HSQC spectra of the SETDB2 MBD–C11orf46 CRD complex in the absence (left) or presence (middle) of an equimolar amount of H3 peptide (residues 1–36). The overlaid spectrum is shown on the right. No obvious changes in the chemical shift of each cross-peak were observed, indicating that the complex does not interact with H3 peptide.

Table S1. Identification of the metal ions incorporated in the SETDB2 MBD–C11orf46 CRD complex by ICP-MS elemental analysis, related to Figures 2–4.

	Zinc	Cobalt	Manganese
Amount	20.8 µg	53.5 ng	1.92 ng
Molar quantity	318 nmol	908 pmol	34.9 pmol
Metal/protein ratio	3.09	8.82×10^{-3}	3.39×10^{-4}