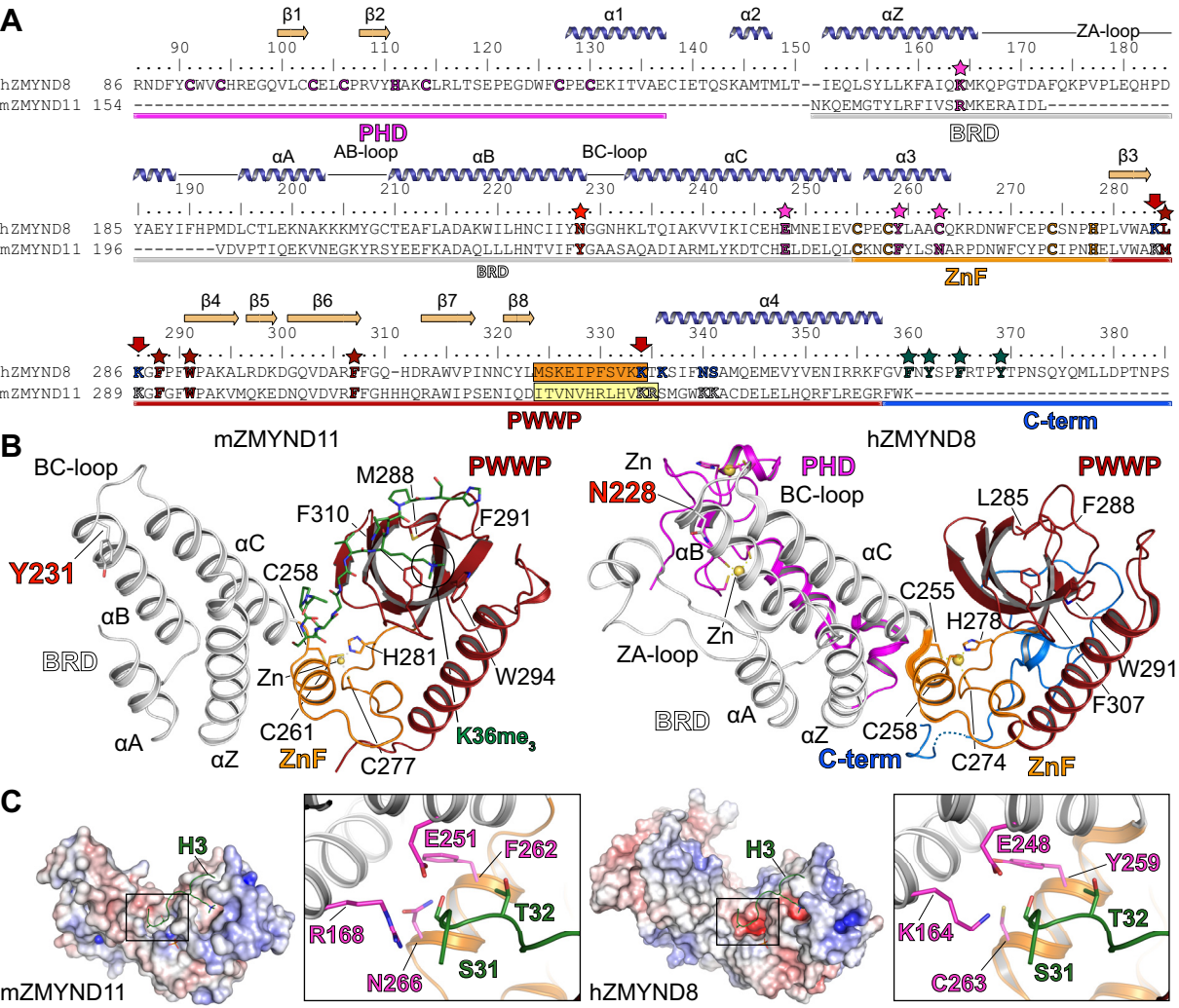
Cell Reports, Volume *17*

Supplemental Information

Multivalent Histone and DNA Engagement by a PHD/BRD/PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin

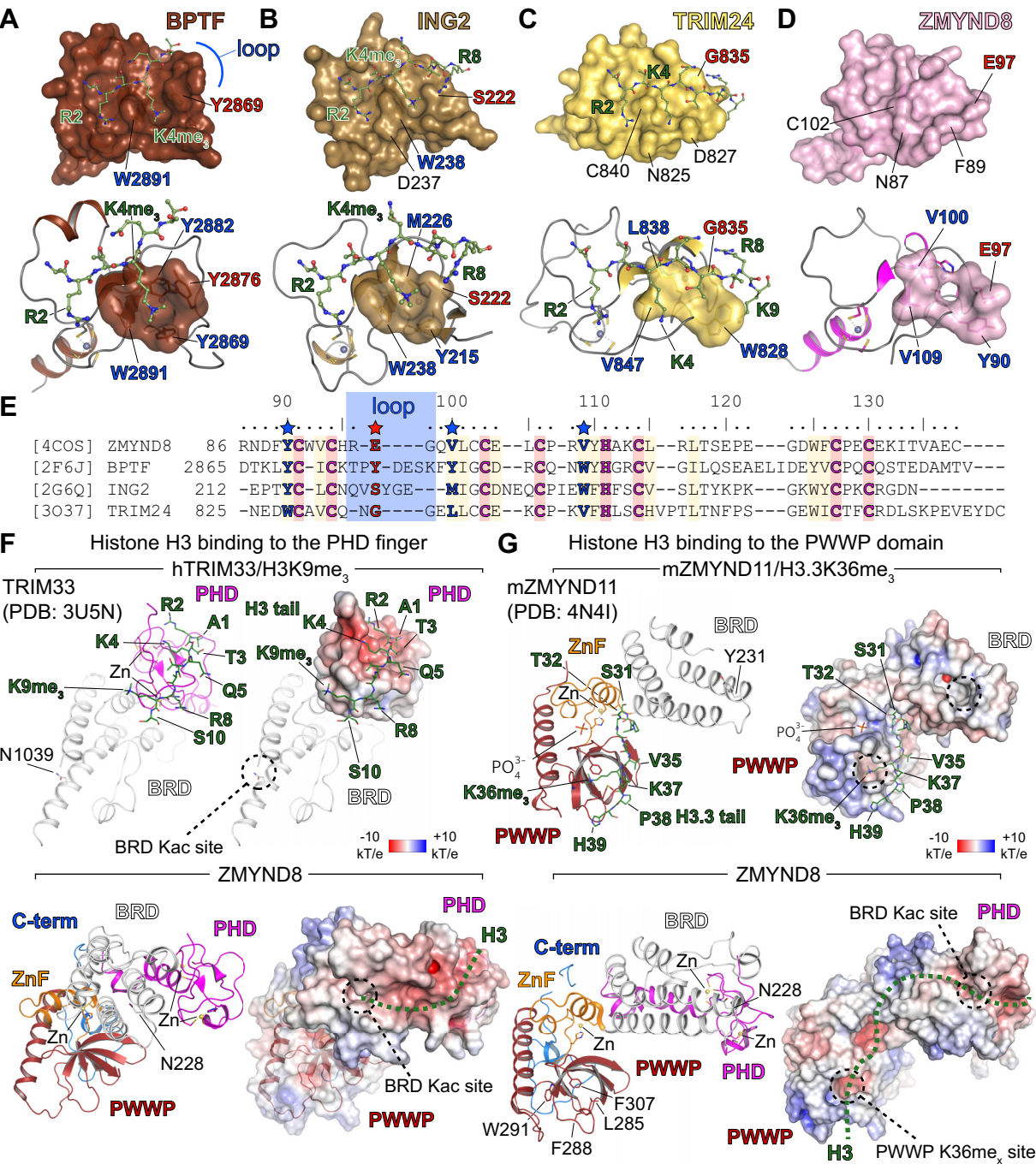
Pavel Savitsky, Tobias Krojer, Takao Fujisawa, Jean-Philippe Lambert, Sarah Picaud, Chen-Yi Wang, Erin K. Shanle, Krzysztof Krajewski, Hans Friedrichsen, Alexander Kanapin, Colin Goding, Matthieu Schapira, Anastasia Samsonova, Brian D. Strahl, Anne-Claude Gingras, and Panagis Filippakopoulos

Supplemental Figure S1



Supplemental Figure S1 – *Comparison of ZMYND8 and ZMYND11, related to Figure 1*. (**A**) Sequence alignment of human ZMYND8 and mouse ZMYND11. Both proteins share a common domain architecture, including three reader modules on their N-terminus, however ZMYND11 lacks the family-conserved asparagine (Y231 in mZMYND11, N228 in ZMYND8) which is responsible for binding to acetylated peptides (highlighted with a red * in the alignment). Secondary structural elements are highlighted and key residues are coloured in red or blue and highlighted with an arrow. The loop between the last β-sheet (β8) of the PWWP domain and the long helix (α4) before the C-terminal portion of the crystallized construct is highlighted (orange box for ZMYND11 and pale yellow for ZMYND8). C-terminal aromatic residues are highlighted in green and are annotated with a star. The PWWP aromatic/hydrophobic cage residues are also highlighted in dark red with a star. (**B**) Crystal structures of the BRD/PWWP domains of ZMYND11 in complex with an H3K36me$_3$ peptide (PDB: 4N4I) and the PHD/BRD/PWWP domains of ZMYND8 with structural elements annotated. Both proteins contain a zinc-finger insertion between the BRD and PWWP domains (shown in orange) which coordinates a zinc ion. The PHD domain of ZMYND8 packs behind the BRD module while the BRD/ZnF/PWWP arrangement is identical to that seen in ZMYND11. The C-terminal portion of the ZMYND8 crystallized construct is highlighted in blue and extends under the PWWP module towards the front of the protein, pointing towards the ZnF. Residues that form up the aromatic cage of the PWWP domains are highlighted (F291, W294, F310, capped by M288 in mZMYND11 and F288, W291, F307 capped by L285 in ZMYND8). (**C**) Residues responsible for S31 binding in mouse ZMYND11 (left) are mostly conserved in ZMYND8 with the exception of N266 which is replaced by C263. The H3.3 peptide found in ZMYND11 is overlaid on the ZMYND8 structure to highlight the position of S31.

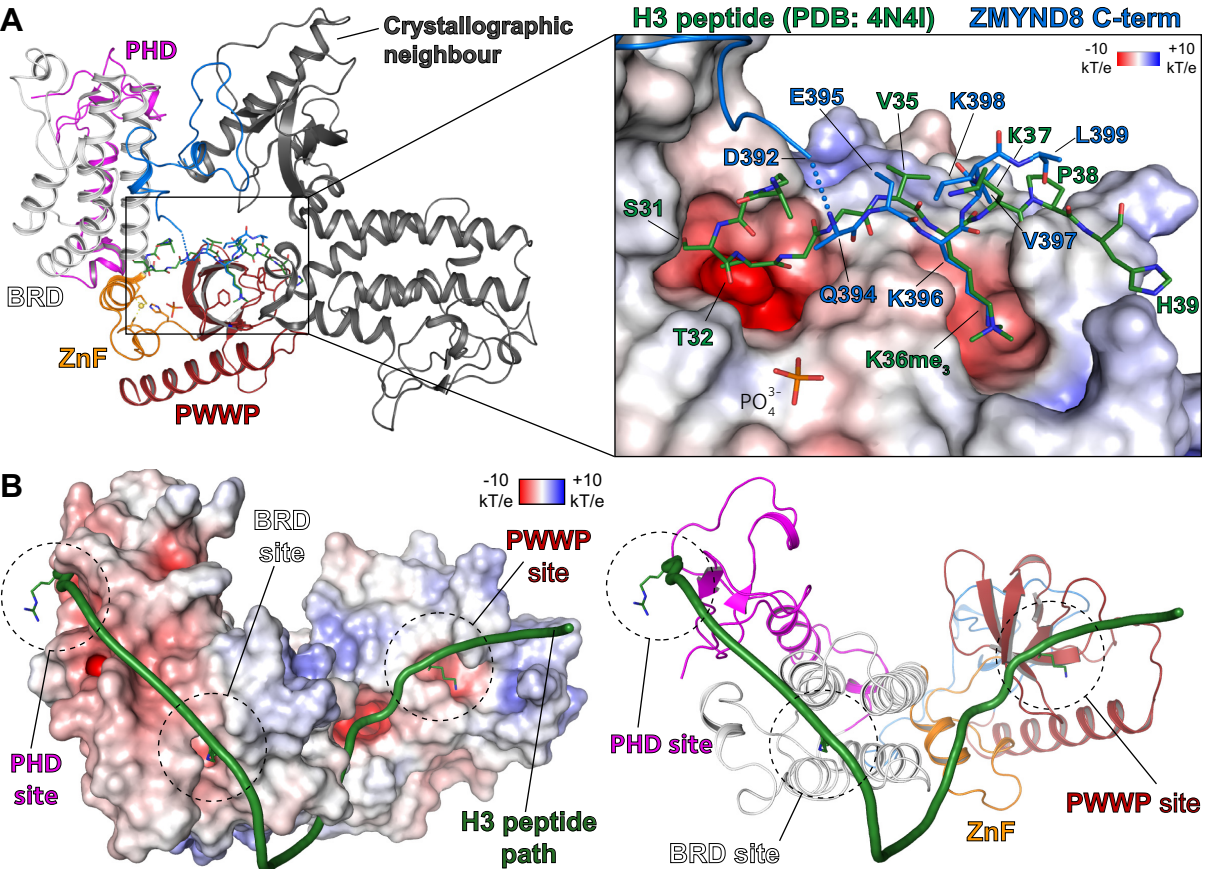
Supplemental Figure S2



Supplemental Figure S2 – *Structural rational for ZMYND8 histone tail binding, related to Figures 1 & 2*. (**A**) $K4me_3$ binding to the PHD finger of BPTF (PDB: 2F6J). Two grooves formed around W2891 separate the histone R2 and the modified $K4me_3$ which are accommodated on the protein surface (top panel). A typical full aromatic cage, formed by Y2882, Y2876, Y2869 and W2891 surrounds the $K4me_3$ (bottom panel). (**B**) $K4me_3$ binding to the PHD finger of ING2 (PDB: 2G6Q). R2 and $K4me_3$ are separated by the bulky W238 and are accommodated in two grooves on the PHD surface. A partial hydrophobic groove formed by M296, Y215 and W238 accommodates $K4me_3$ which packs next to S222. (**C**) Binding of K4 to the PHD finger of TRIM24 (PDB: 3O37). K4 inserts in the groove between C840, N825 and D827 (top panel). In this case the aromatic cage is replaced by L838, W828 and V847 which present an open surface to the unmodified lysine, capped by G835 on the flexible loop to the right of the pocket. (**D**) The PHD finger of ZMYND8 resembles that of TRIM24, lacking an aromatic cage. A shallow groove formed by C102, N87 and F89 is present, with hydrophobic residues completing the surface (including V100 and V109) which is capped by E97 on the loop side. (**E**) Structural alignment of the four PHD finger shown in (a-d) highlighting residues forming the cage (shown in blue and annotated with a '*' as well as those found on a variable length loop region and cap the right side of the cage (shown in red and highlighted by a '*'). Conserved residues are also highlighted and annotated. (**F**) Comparison of the PHD finger (in magenta) from the PHD/BRD readers of human TRIM33 (top) and the PHD finger (in magenta) from the PHD/BRD/PWWP triple readers of ZMYND8 (bottom). Both structures have been superimposed based on the coordinates of the PHD domain. The TRIM33 structure (PDB: 3U5N) in complex with a modified histone H3 peptide ($H3K9me_3$, shown in green) demonstrates how the N-terminal portion of the peptide engages the entire PHD domain, eventually positioning the rest of the tail towards the BRD domain (the conserved asparagine, N1039 is highlighted on the structure). In the case of the ZMYND8 structure, the topology of the PHD and BRD modules is different, with the BRD cavity now much closer to the end of the PHD finger, suggesting that a histone H3 peptide can follow a similar path, positioning a histone tail towards the conserved asparagine of the BRD module (N228 in ZMYND8) which can accept an acetylated lysine (Kac site), as shown on the electrostatic surface of the protein (histone peptide path shown as a dotted thick green line). (**G**) Comparison of the PWWP domains (shown in red) of the mouse ZMYND11 (PDB: 4N4I shown on top) and human ZMYND8 (bottom panel). The mZMYND11 structure carries a histone H3.3 peptide modified at lysine K36 ($H3.3K36me_3$) which binds into the aromatic cage of the PWWP domain. In this particular structure S31 and T32 pack in a groove formed between the BRD and PWWP modules with the rest of the peptide extending on the PWWP surface. The BRD and PWWP cavities are highlighted with a dotted circle. Using a similar path, an H3 peptide that binds to the PHD

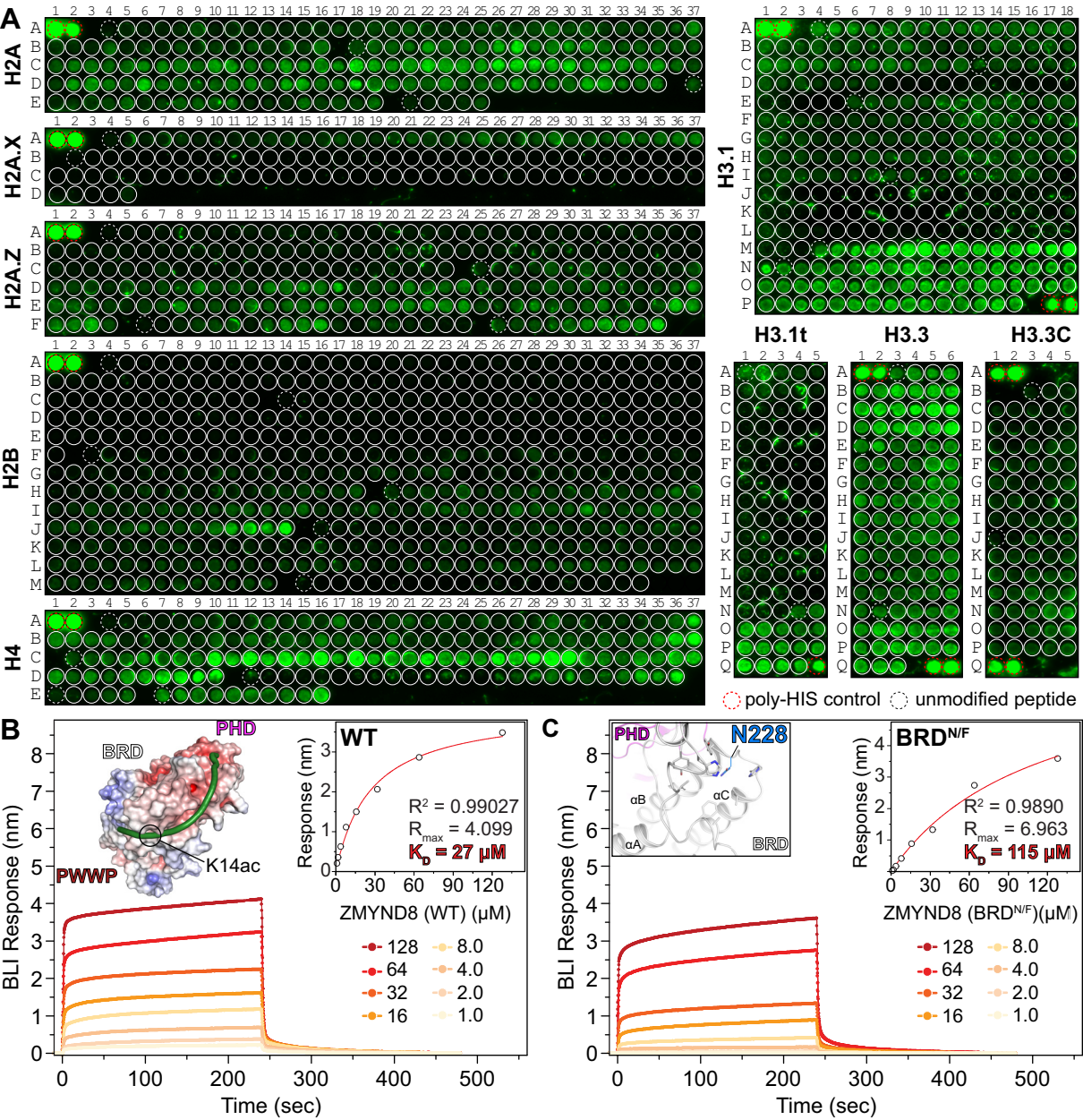
finger of ZMYND8 can extend to span the BRD cavity then the groove between the BRD and PWWP domains, potentially inserting K36 within the PWWP K36me_x site formed by F288, W291 and F307.

Supplemental Figure S3



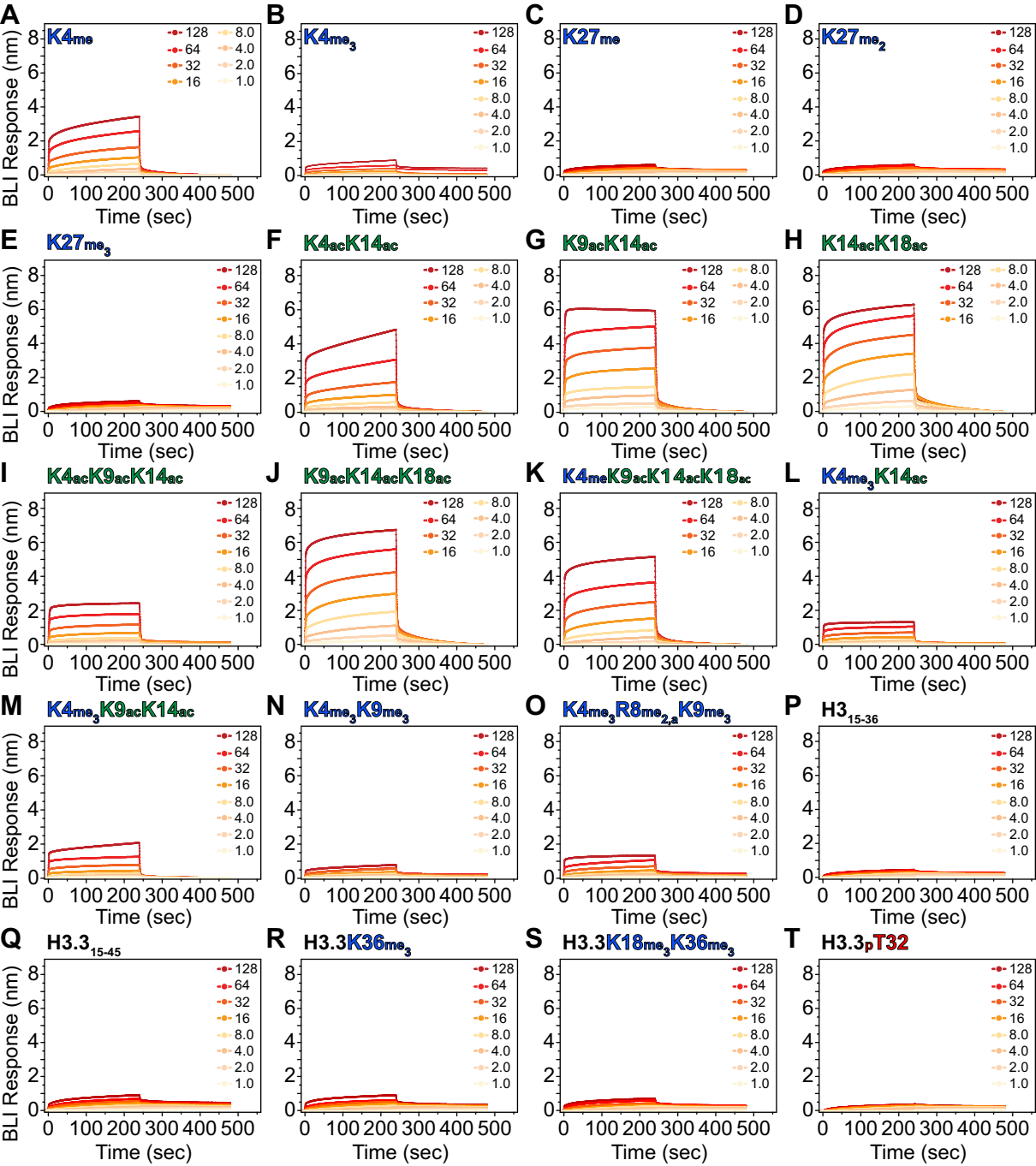
Supplemental Figure S3 – *Structural rational for ZMYND8 histone tail binding, related to Figures 1 & 2*. (**A**) Detail of the ZMYND8 structure highlighting packing of a symmetry equivalent (shown in gray) of the triple reader modules, inserting its C-terminal tail within the PWWP aromatic cage. Superimposition with the mouse ZMYND11 structure in complex with H3.3K36me$_3$ (PDB: 4N4I) shows good overlay of the ZMYND8 C-terminal K396 with the histone peptide's K36me$_3$ suggesting that the protein can bind to unmodified lysines. Key residues are highlighted coloured in green (histone H3.3 peptide) or blue (ZMYND8 crystallographic neighbour's C-terminal tail) on the electrostatic surface of the protein. A phosphate ion found on the mZMYND11 structure is positioned in a slightly positively charged pocket of the ZnF and PWWP domain surface, suggesting that small changes in peptide binding topology can be accommodated if T32 is phosphorylated. (**B**) Cartoon of hypothetical histone H3 peptide engagement by the triple reader modules of ZMYND8. The triple reader topology presents a charged surface that could potentially interact with a large portion of the histone N-terminal tail. Interfaces are presented in the PHD, BRD and PWWP sites, driving binding from all three reader domain sites.

Supplemental Figure S4



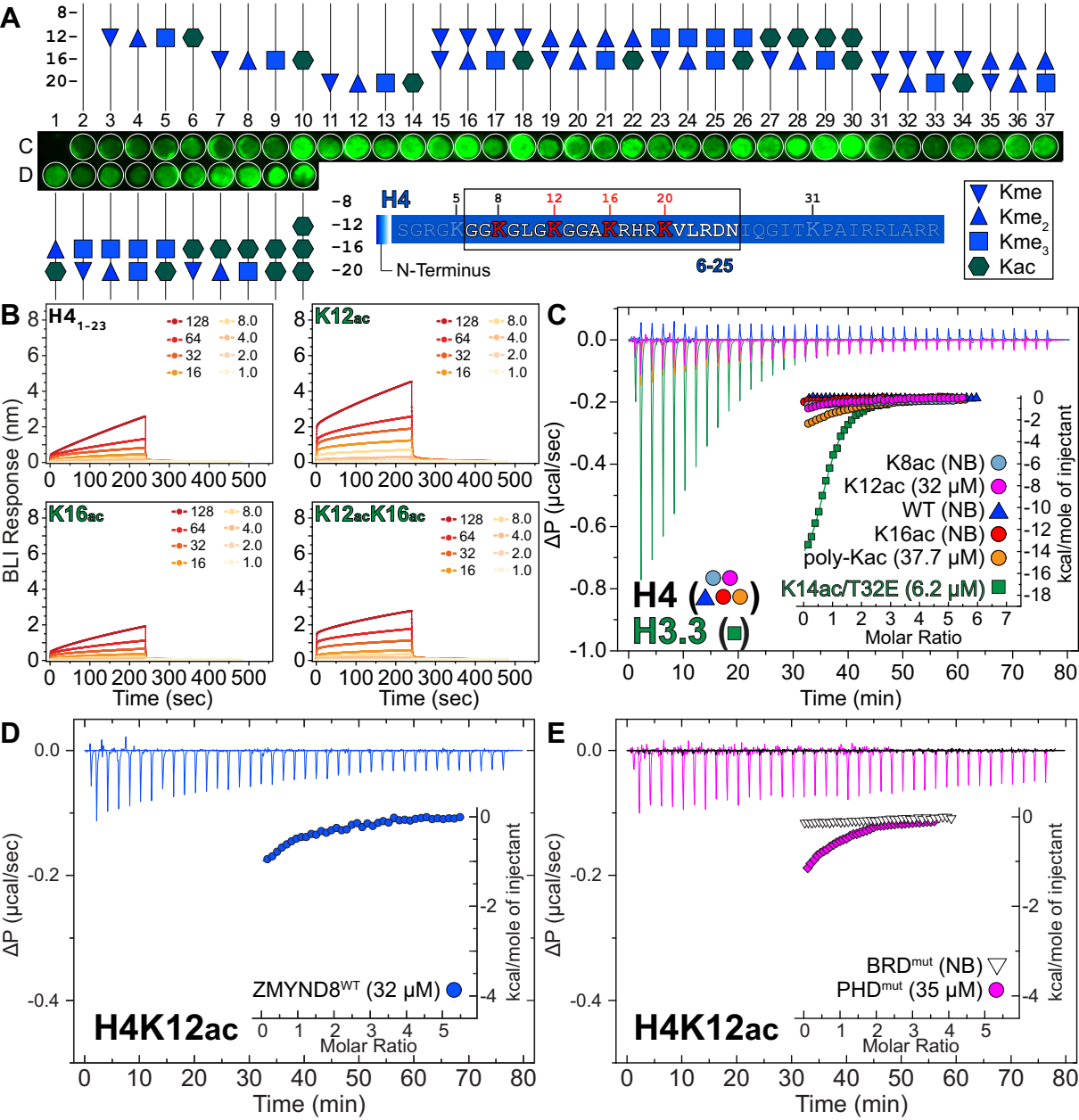
Supplemental Figure S4 – *ZMYND8 histone binding* in vitro*, related to Figure 2.* (**A**) Peptide SPOT arrays of human histones (H2, H3 and H4) carrying 20-aa long peptides with combinations of PTMs (including Kac, Kme_x (x =1,2,3), pS and pT). Histone isoforms (H2A, H2A.X, H2A.Z, H2B, H3.1t, H3.3 and H3.3C) were also tested to account for differences in sequence. Hexa-his-tagged recombinant ZMYND8 triple reader modules (PHD/PRD/PWWP) were incubated with the membranes and after washing were visualized with an anti-his antibody. The presence of multiple acetyl marks dominated the resulting binding that was observed. Peptide sequences are given in **Supplemental Table S2**. (**B**) Bio-layer interferometry (BLI) profiling of recombinant ZMYND8 (PHD/BRD/PWWP) binding to a synthetic H3K14ac biotinylated peptide (20-mer) immobilized on streptavidin sensors. The sensograms demonstrate binding and dissociation of the protein to the immobilized peptide at different concentrations. The inset shows the steady state response as a function of protein concentration calculated from the association and dissociation curves and fitted with a non-linear regression model. Fit parameters are given in the inset. A structural model highlighting the potential binding path of the H3 peptide crossing the BRD cavity where K14ac is predicted to bind is shown in the inset. Experiments were performed in 20 mM HEPES pH7.5, 200 mM NaCl, 20 mM TCEP. (**C**) Bio-layer inteferometry (BLI) evaluation of BRD mutant ($BRD^{N/F}$) ZMYND8 to an H3K14ac 40-mer peptide. Data and fitting is represented as in (B). The mutant protein exhibits weaker binding with slower and incomplete dissociation from the bound peptide suggesting non-specific binding.

Supplemental Figure S5



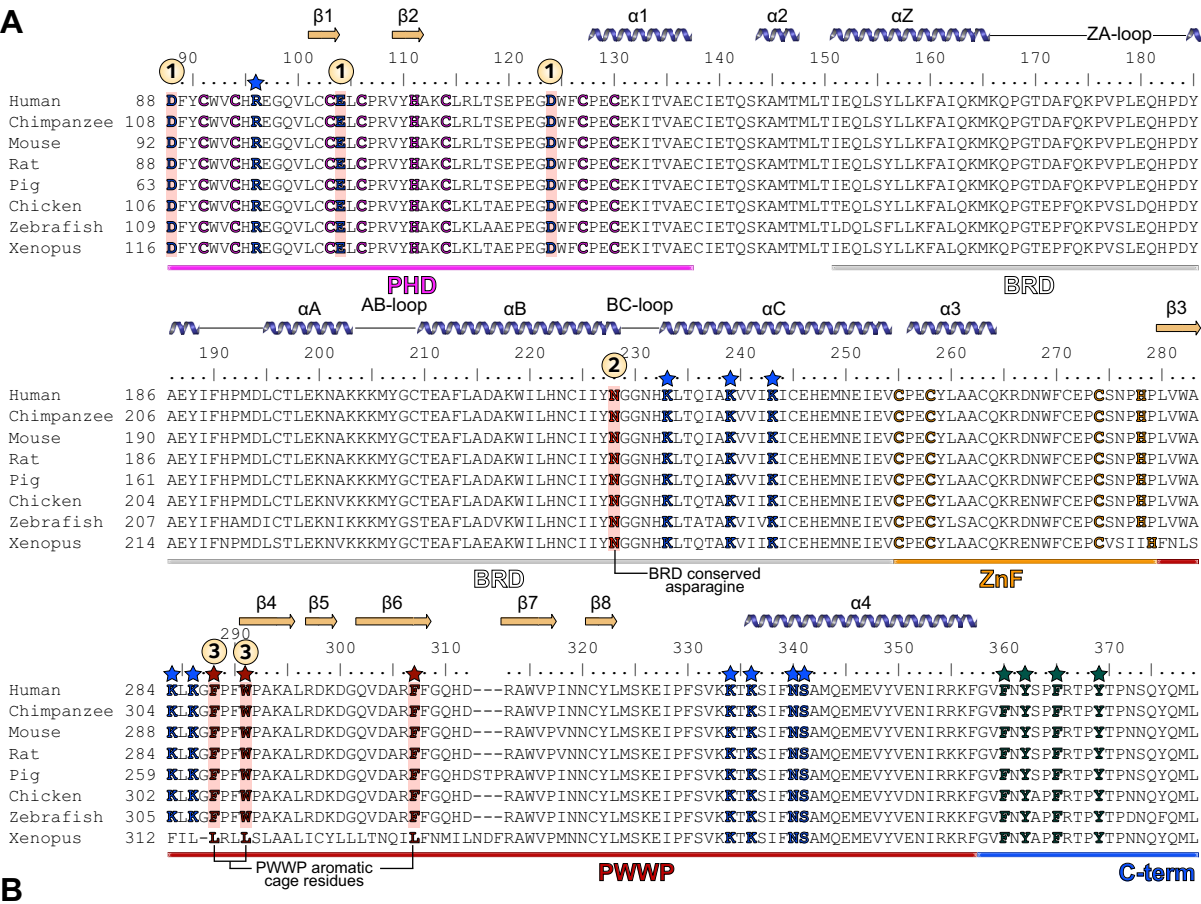
Supplemental Figure S5 – *In-solution validation of ZMYND8/histone H3 interactions employing biolayer interferometry (BLI), related to Figure 2*. Biotinylated histone H3 21-amino acid long peptides carrying different modification were immobilized on Super Streptavidin Biosensors and were profiled against a range of recombinant triple readers of ZMYND8. Association and dissociation measurements were performed in 20 mM HEPES pH7.5, 200 mM NaCl, 20 mM TCEP. Experiments were carried out at 25 °C with association and dissociation times of 240 sec using protein concentrations as indicated in the insets. Peptides tested (residues 1-21 unless otherwise stated) included: (**A**) K4me (**B**) K4me_3 (**C**) K27me (**D**) K27me_2 (**E**) K27me_3 (**F**) K4ac/K14ac (**G**) K9ac/K14ac (**H**) K14ac/K18ac (**I**) K4ac/K9ac/K14ac (**J**) K9ac/K14ac/K18ac (**K**) K4me/K9ac/K14ac/K18ac (**L**) K4me_3/K14ac (**M**) K4me_3/K9ac/K14ac (**N**) K4me_3/K9me_3 (**O**) K4me_3/R8$me_{2,a}$/K9me_3 (**P**) WT H3 (residues 15-36) (**Q**) WT H3.3 (residues 15-45) (**R**) H3.3 (residues 15-45) K36e_3 (**S**) H3.3 (residues 15-45) K18me_3/K36me_3 (**T**) H3.3 (residues 15-45) pT32.

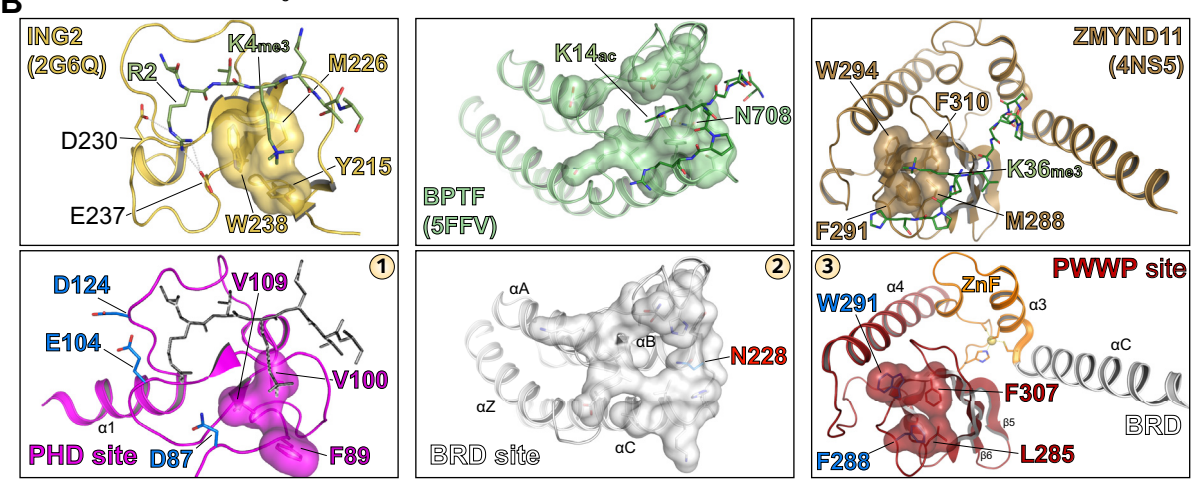
Supplemental Figure S6



Supplemental Figure S6 – *ZMYND8 binding to histone H4, related to Figure 2.* (**A**) SPOT peptide array detail exploring binding of modified histone H4 peptides to the recombinant triple reader module of ZMYND8. Peptides span residues 6-24 of the human H4 tail and modified as indicated in the inset. (**B**) In-solution biolayer interferometry (BLI) validation of ZMYND8/histone H4 interactions. Biotinylated histone H4 21-amino acid long peptides carrying different modifications were immobilized on Super Streptavidin Biosensors and were profiled against a range of recombinant triple readers of ZMYND8. Association and dissociation measurements were performed in 20 mM HEPES pH7.5, 200 mM NaCl, 20 mM TCEP. Experiments were carried out at 25 °C with association and dissociation times of 240 sec using protein concentrations as indicated in the insets. Peptides tested (residues 1-21 unless otherwise stated) included: WT H4 N-terminal tail (1-23); K12ac; K16ac; K12ac/K16ac. (**C**) In solution evaluation of histone H4 binding by isothermal titration calorimetry (ITC). Raw injection heats for titrations of modified peptides (recombinantly produced 40-mers carrying specific modifications as indicated in the inset, deposited by an Amber codon system, or *in vitro* poly-acetylated using acetic anhydride) into a solution of ZMYND8 are shown in the main panel. The inset shows the normalized binding enthalpies corrected for the heat of peptide dilution as a function of binding site saturation (symbols as indicated in the figure). Solid lines represent a nonlinear least squares fit using a single-site binding model. Histone H4 peptide binding is driven by K12ac (K_D = 32 μM), an interaction which appears to be much weaker than the H3K14ac/T32E (K_D = 6 μM). (**D/E**) ITC validation of histone H4K12ac binding to ZMYND8 WT (D) or mutant (PHD or BRD, (E)) in solution. Data is presented as in (D). All ITC titrations were carried out in 20 mM HEPES pH 7.5 (at 25 °C), 200 mM NaCl and 15 °C while stirring at 1000 rpm.

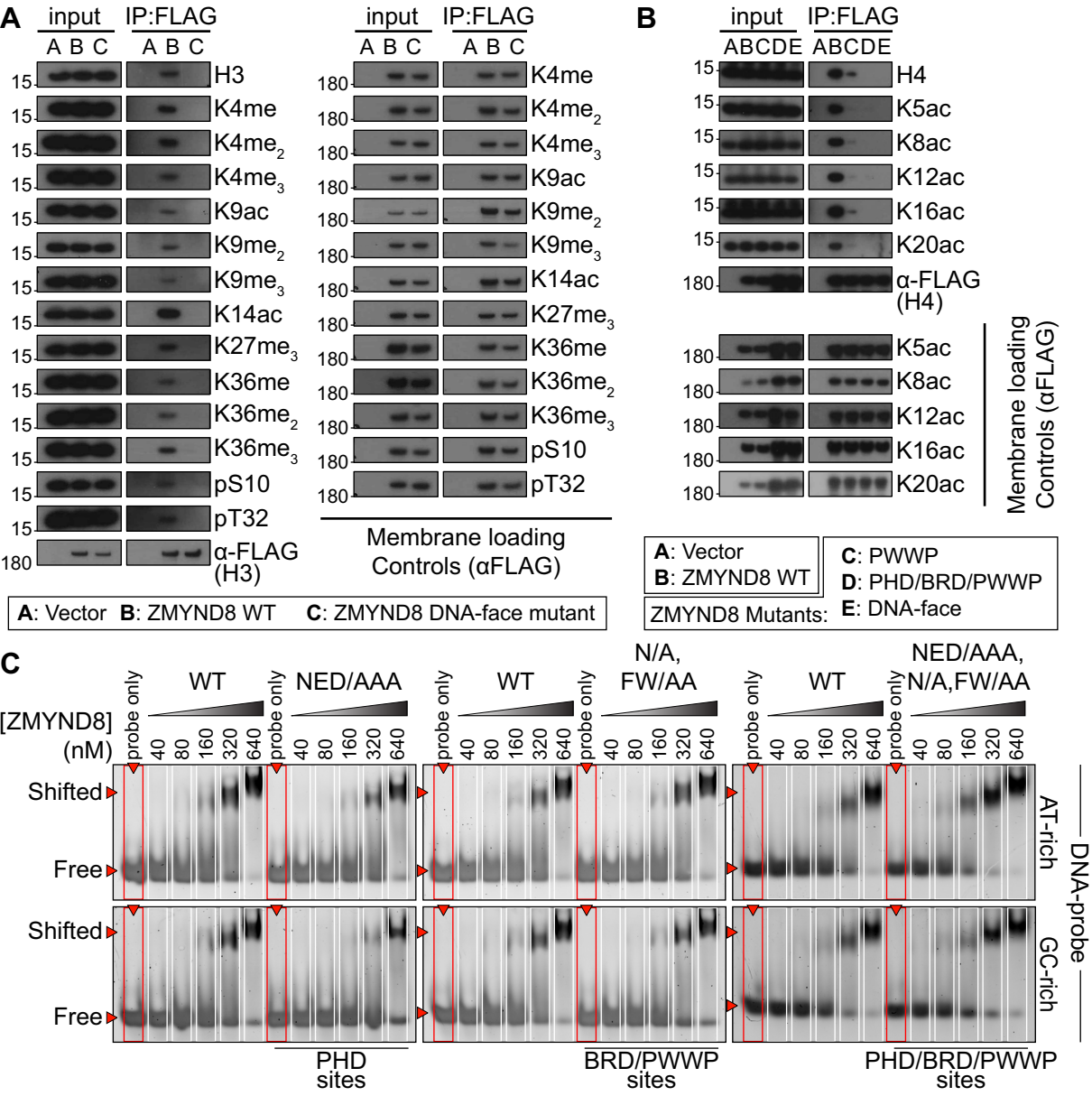
Supplemental Figure S7





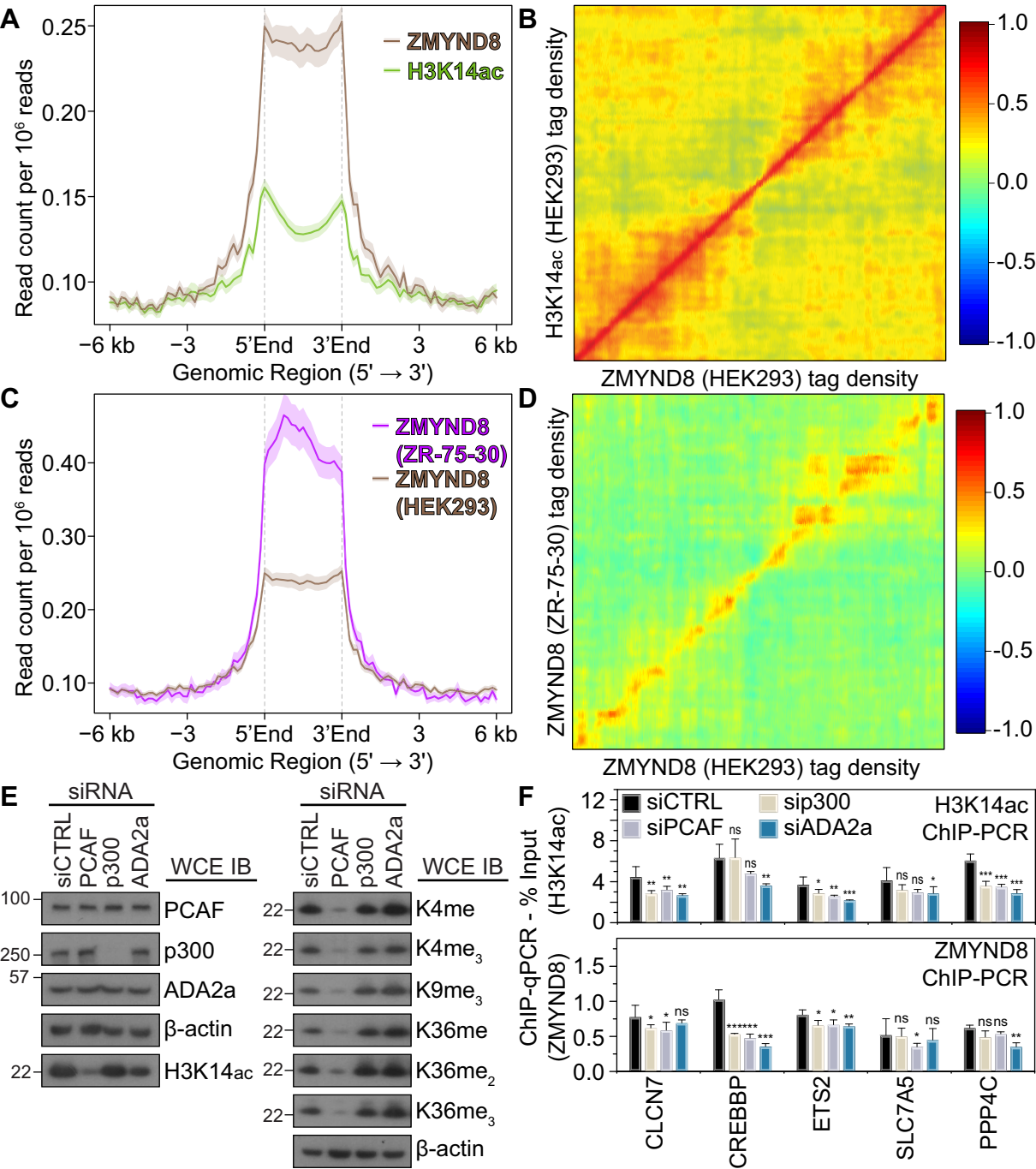
Supplemental Figure S7 – *Evolutionary conserved ZMYND8 residues and histone binding interfaces, related to Figure 2.* (**A**) Sequence alignment of human (ENSP00000418210), chimpanzee (ENSPTRP00000023370), mouse (ENSMUSP00000104892), rat (ENSRNOP00000025932), pig (ENSSSCP00000007942), chicken (ENSGALP00000007220), zebrafish (ENSDARP00000040234) and xenopus (ENSXETP00000063260) ZMYND8 triple reader modules, demonstrating high conservation of interaction interfaces. The secondary structure from the human ZMYND8 structure is displayed on top of the alignment and residues important for histone peptide binding are annotated. The three main sites (1: PHD, 2: BRD, 3: PWWP) are also highlighted in the inset. Residues implicated in DNA binding are highlighted in blue and annotated with a blue *. Residues implicated in C-terminal hydrophobic interactions are highlighted in green and annotated with a green *. The aromatic PWWP-cage residues are highlighted in dark red and indicated with a dark red *. (**B**) Structural overlays of histone peptide complexes with individual domains. Structural overlay of the ING2 PHD finger (top box) with the PHD domain of ZMYND8 (site '1'), highlighting residues important for K4me3 binding. Arg2 initiates interactions with the PHD loop region. The equivalent residues in ZMYND8 are annotated in blue (D87, E104 and D124), following the conservation highlighted in (a). A structural overlay of the BPTF bromodomain in complex with a K14ac peptide (top middle box) is shown compared to the ZMYND8 BRD module (middle lower box). The conserved asparagine (N228) is highlighted in blue. The PWWP domain of ZMYND11 (right top box) in complex with a K36me$_3$ peptide is shown in comparison to the ZMYND8 PWWP domain (right bottom box), highlighting the aromatic cage responsible for Kme$_x$ binding. Conserved residues are highlighted in blue (F288 and W291).

Supplemental Figure S8



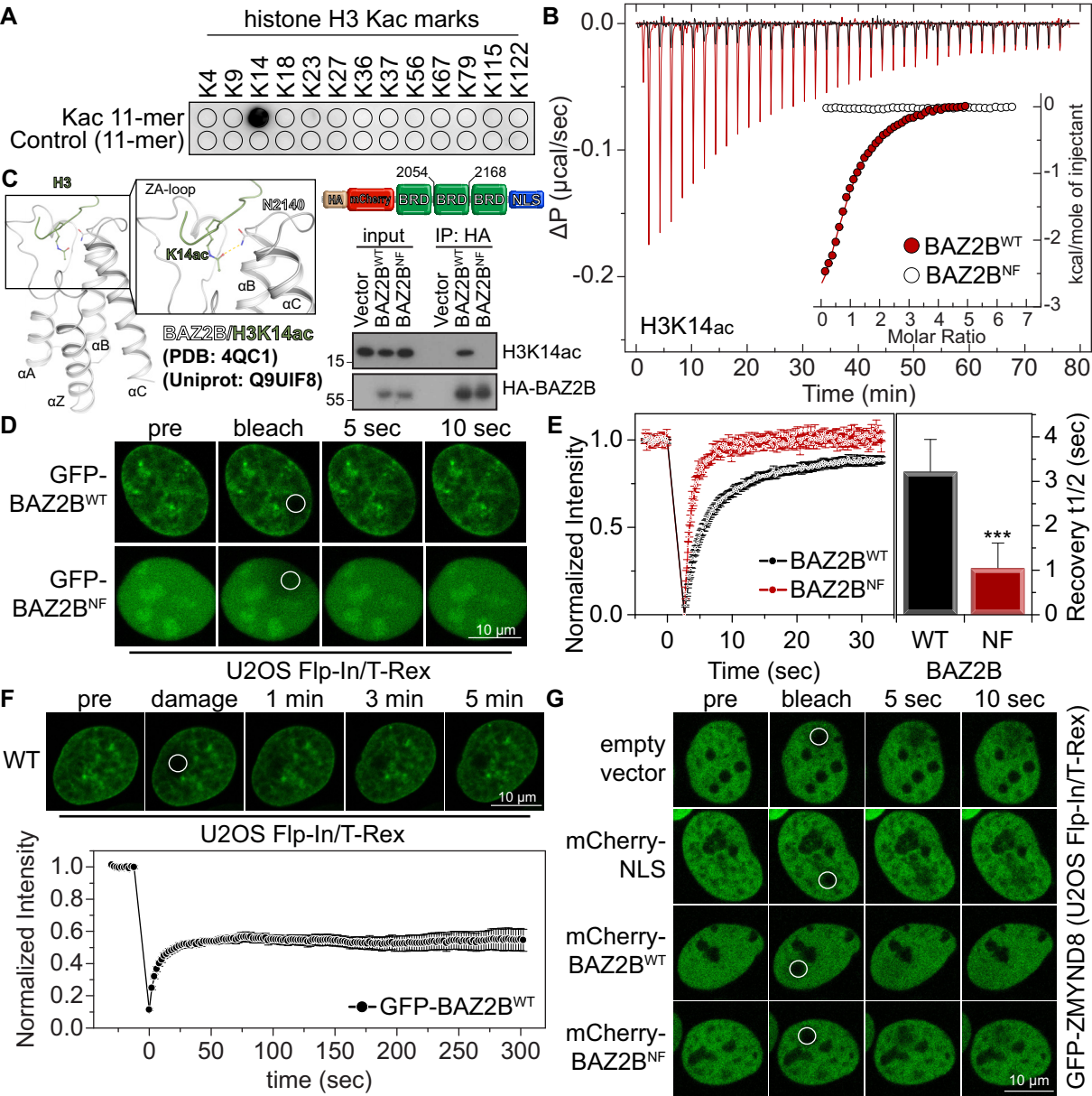
Supplemental Figure S8 – *ZMYND8 DNA binding is necessary for recruitment to DNA damaged sites and mutations of the triple reader ensemble do not affect recognition of DNA* in vitro*, related to Figures 3 & 4*. (**A**) Full-length ZMYND8 is unable to bind to histone H3 modifications in HEK293 cells when the DNA-binding sites identified in the structure are mutated to alanine (DNA-face mutant: R96A, K233A, K239A, K243A, K284A, K286A, K334A). HEK293 cells transiently transfected with full-length 3xFLAG ZMYND8 WT or DNA-face mutant and stained for different H3 modifications. Note that the DNA-face mutant of ZMYND8 shows much lower expression than the WT protein. (**B**) Effect of ZMYND8 mutations on histone H4 binding. HEK293 cells transiently transfected with full-length 3xFLAG ZMYND8 WT or PWWP (F288A/W291A), PHD/BRD/PWWP (N87A/E104A/D124A + N228F + F288A/W291A) or DNA-face (R96A, K233A, K239A, K243A, K284A, K286A, K334A) mutants, were stained for different H4 modifications. While mutations on the PWWP domain retained residual binding of all marks tested, mutation on all three domains completely abolished binding, as did mutations on the DNA-face of the protein. (**C**) Systematic mutation of domains within ZMYND8 did not have an effect in shifting a AT-rich (top) or GC-rich (bottom) DNA probe compared to the WT protein in non-radioactive EMSA assays. Mutations where introduced in the PHD domain (N87A/E104A/D124A), the BRD/PWWP domains (N228F; F288A/W291A), or all three reader domains (N87A/E104A/D124A/N228F/ F288A/W291A).

Supplemental Figure S9



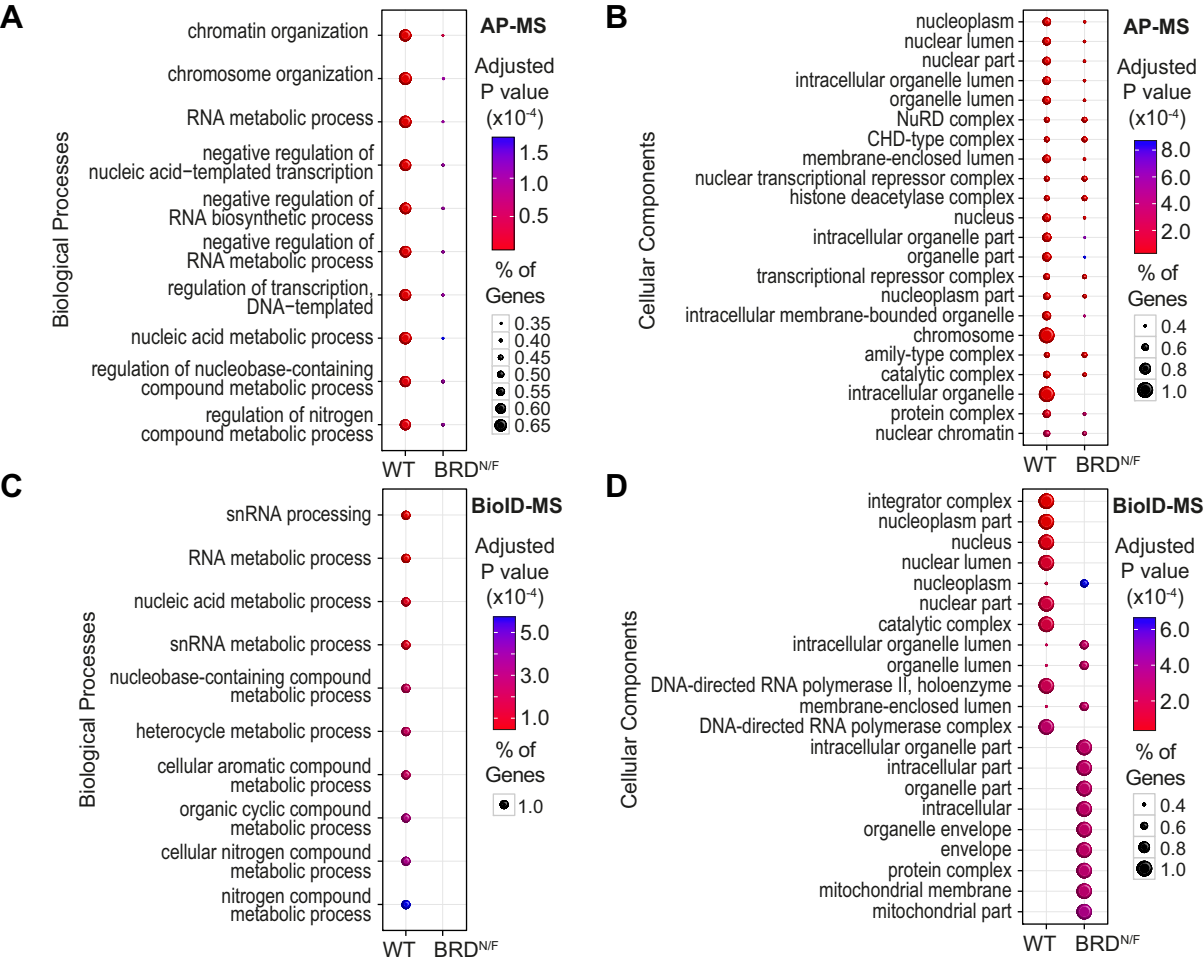
Supplemental Figure S9 – *ZMYND8 co-localizes with H3K14ac at Enhancers, related to Figure 5.* (**A**) Average profile of ZMYND8 and H3K14ac ChIP-seq signals on ±6 kb around enhancers (as defined by the FANTOM5 project) in HEK293 cells. (**B**) Canonical correlation matrix for ChIP-seq tag enrichment profiles observed in ±6 kb regions in ZMYND8-bound enhancers (defined by FANTOM5 annotations). High correlation (highlighted in red as indicated in the inset) was found between ZMYND8 and H3K14ac ChIP-seq signals in HEK293 cells. (**C**) Average profile of ZMYND8 ChIP-seq signals on ±6 kb around enhancers (as defined by the FANTOM5 project) in ZR-75-30 (data from GSE71323) and HEK293 cells. (**D**) Canonical correlation matrix for ZMYND8 ChIP-seq tag enrichment profiles observed in ± 6 kb regions in ZMYND8-bound enhancers (defined by FANTOM5 annotations), between HEK293 and ZR-75-30 (GSE71323) cells. Correlation is annotated as in (B). (**E**) siRNA depletion of HATs (PCAF, p300 and GCN5 via siRNA depletion of the ADA2a portion of the SAGA complex) resulted after 72h in reduction of H3K14ac (left panel). siHAT also (indirectly) resulted in variable attenuation of methylation marks found on histone H3 (right panel). (**F**) ChIP-qPCR enrichment of H3K14ac (top) and FLAG-ZMYND8 (bottom) signals following 72h siRNA depletion of HATs (siCTRL vs siPCAF, sip300, siADA2a). Signals are measured relative to IgG signal and presented as percentage of loaded input DNA. Data represent mean ± SEM from biological replicates (n=3). *P* values were calculated using Student's t-test and are represented so that $P < 0.005$, "***"; $P < 0.01$, "**"; $P < 0.05$, "*"; Not-significant, "ns".

Supplemental Figure S10



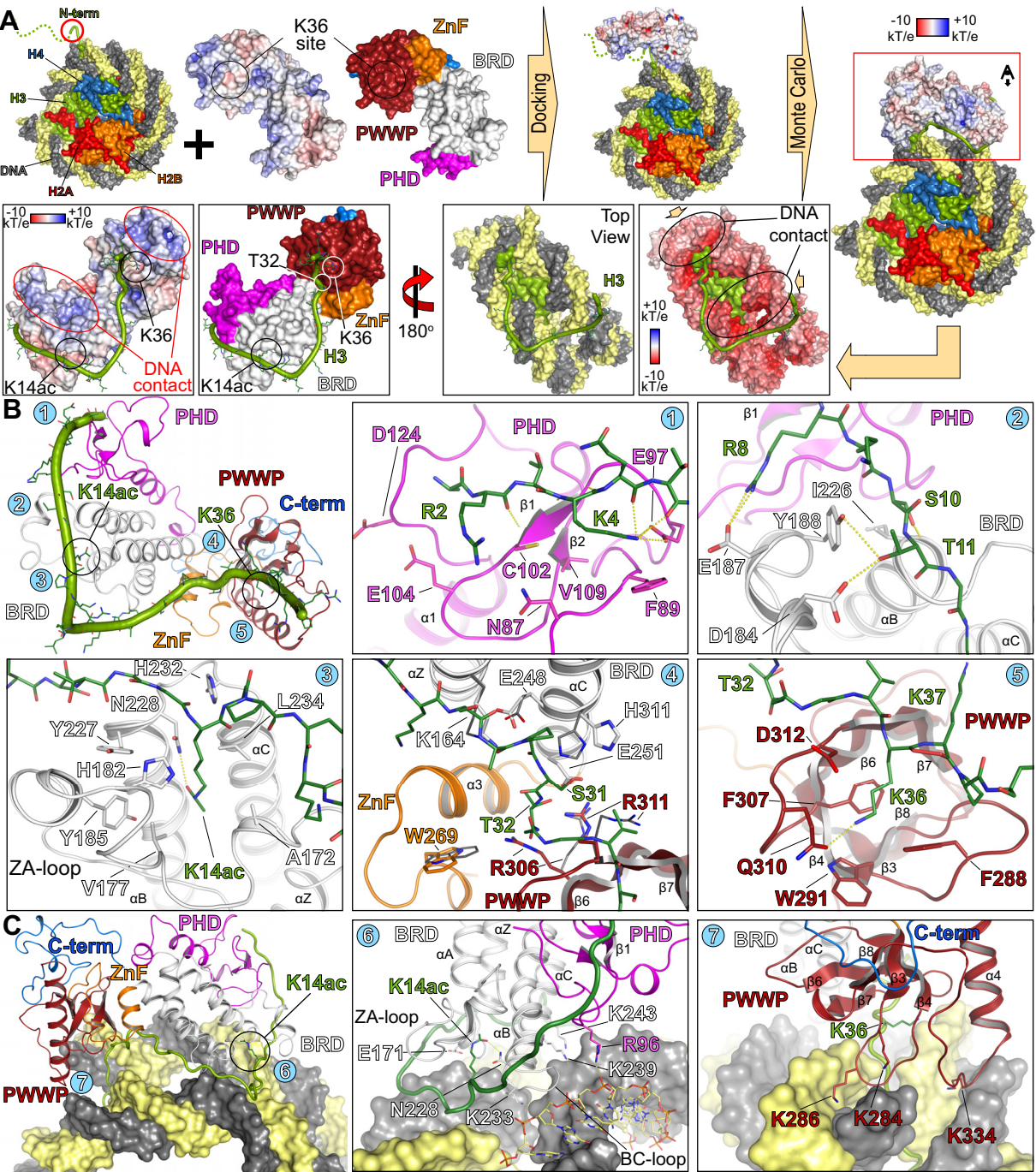
Supplemental Figure S10 – *Masking of H3K14ac using the bromodomain of BAZ2B prevents ZMYND8 binding, related to Figure 5.* (**A**) Cellulose-based SPOT validation of histone H3-peptide binding to recombinant BAZ2B bromodomain, showing high specificity for K14ac. (**B**) In solution evaluation of H3K14ac binding by ITC to recombinant BAZ2B WT or N2140F mutant. Raw injection heats for titrations of a 15-mer H3K14ac peptide into a solution of BAZ2B (WT or N2140F mutant) are shown in the main panel. The inset shows the normalized binding enthalpies corrected for the heat of peptide dilution as a function of binding site saturation (symbols as indicated in the figure). Solid lines represent a nonlinear least squares fit using a single-site binding model. While the WT proteins engages K14ac with low μM affinity the mutant shows no interaction at all. (**C**) Structural model of BAZ2B in complex with an H3K14ac peptide (PDB: 4QC1) showing the interaction between the conserved asparagine (N2140) and the acetylated lysine of histone H3. Mammalian constructs carrying 3x BRD modules of BAZ2B with N-terminal tags (HA/mCherry) and a C-terminal nuclear localization signal, bind by western blot to H3K14ac in HEK293 cells transiently transfected. The N2140F mutation abolishes the recognition of K14ac in cells. (**D**) Fluorescence recovery after photo-bleaching (FRAP) evaluation of GFP-3xBAZ2B (WT, top panels or N2140F mutant, bottom panels) dissociation from chromatin in U2OS cells transiently transfected with GFP-3xBAZ2B constructs. Target regions of photo-bleaching are indicated with a white circle. The scale bars correspond to a 10 μm region. (**E**) Quantitative comparison of time to half-maximal fluorescence recovery for GFP-3xBAZ2B. While the WT protein remains attached to chromatin, the mutant exhibits significant mobility as indicated by the faster recovery after bleaching. Data represent the mean ± SEM from multiple experiments in several cells (n = 15). (**F**) Data acquisition for longer times (5 min) did not show any accumulation of the GFP-BAZ2B protein onto the bleached site suggesting that there is no active recruitment to the damaged site. Data represent the mean ± SEM from multiple experiments in several cells (n = 15). (**G**) FRAP evaluation of full length GFP-tagged ZMYND8 dissociation from chromatin in U2OS Flp-In/TRex cells stably expressing GFP-ZMYND8. Nuclei of cells treated with vector, mCherry-NSL, mCherry-3xBAZ2B^{WT} or mCherry-3xBAZ2B^{NF} mutant.

Supplemental Figure S11



Supplemental Figure S11 – *Gene Ontology analysis of ZMYND8 (WT & BRD$^{N/F}$ mutant) proteomic networks, related to Figure 7.* (**A-D**) Enrichment of gene ontology terms following analysis of AP-MS or BioID data from full length 3x-FLAG ZMYND8 WT or BRD (BRD$^{N/F}$) mutant proteins. Spheres indicate percentage of genes participating in each enriched cluster, coloured by *P* value adjusted using the Benjamini-Hochberg false discovery rate (FDR) with a threshold of 0.05. Functional enrichment is shown for Biological Processes (AP-MS data in (A) and BioID data in (C)) as well as Cellular Components (AP-MS data in (B) and BioID data in (D)) Gene Ontology terms. The BRD mutant systematically shows reduced enrichment in several functions

Supplemental Figure S12



Supplemental Figure S12 – *Molecular modelling of the ZMYND8/nucleosome interaction*, *related to Figure 7*. (**A**) ZMYND8 was docked onto the core nucleosome particle (coloured as indicated in the inset – model taken from PDB: 3AFA) in a stepwise fashion, first by docking its reader ensemble onto the nucleosome while restraining K36 within the aromatic PWWP cage, then by building an H3 tail covering residues 1-36 on top of the reader ensemble by tethering and restraining its N-terminal part (ARTK) within the PHD finger cavity, K14ac within the BRD cavity, T32 within the cavity at the interface of the PWWP and BRD modules and K36 within the aromatic cage of the PWWP domain (top panel). Subsequent energy minimization by Monte Carlo simulation converged towards a low-energy docked structure whereby the reader ensemble was found positioned on the nucleosome core with the predicted DNA-face of the molecule directly initiating interactions with the nucleosomal DNA, while the histone H3 tail was bound on the PHD, BRD and PWWP face of the structure. A side view of this complex mode of interaction is shown on the left panel. A top-down view is split into two view-points rotated by 180˚ compared to each other (right panel) highlighting domain location (top) or electrostatic properties (bottom). The H3 tail is wrapped around the reader modules which in turn are initiating direct contacts with the DNA backbone, suggesting that ZMYND8 acts as both nucleosome and histone binder. (**B**) The N-terminal portion of the H3 peptide retained mainly backbone interactions with the PHD domain, while R2 was found in a charged environment (N87, E104 and D124) and K4 packed between C102, V109 and F89 while forming a salt bridge to E97 (Panel 1). R8 engaged directly E187 on the BRD site via a salt bridge, and S10/T11 were accommodated on top of the BRD surface, engaging the protein either via backbone contacts (S10 with I226) or directly via hydrogen bonds to Y188 and D184 (Panel 2). Acetylated K14 inserted into the BRD cavity and directly engaged the conserved asparagine (N228) while packing between H232, L234, H182, Y185 and V177 (Panel 3). S31 was found in proximity to H311, while T32 inserted in a cavity formed between the PWWP, BRD and ZnF domains directly engaging E251 (BRD), R306 and R313 (PWWP) (Panel 4). This pose was highlighted by substantial side chain rearrangement in the T32 immediate environment (R306, R313 on the PWWP domain; W269 on the ZnF; K164, H247, E248 on the BRD domain), and we wondered if this would provide additional space for accommodating a bulkier phosphorylation on T32. Unmodified K36 inserted into the hydrophobic pocket of the PWWP domain, following side chain rotations on F288 as well as D312, with the latter forming a salt bridge to K36 (Panel 5). (**c**) Binding to the entire histone tail was accompanied by close interactions of the reader ensemble with the DNA wrapped around the core histones both on the PHD/BRD as well as the PWWP sites R96 on the PHD site contacted one DNA strand while a combination of histone interactions and direct DNA contact with the second stand via D171 on the BRD site resulted in a deep insertion of the BRD BC-loop between the two DNA turns and K233 initiating backbone phosphate contacts, positioning at the same time a charged face of

the BRD C-helix (N236, K239 and K243) in direct contact with the DNA backbone (Panel 6). On the PWWP site, the loop between sheets β3 and β4 inserted K286 into the minor groove of one DNA turn with K286 close to the phosphate backbone, while the loop between sheet β8 and helix α4 inserts between two DNA turns, allowing K334 to direct contact with the phosphate backbone (Panel 7). ZMYND8 reader domains and their residues are coloured as follows: PHD domain, magenta; BRD domain, white; ZnF domain, orange; PWWP domain, dark red; C-terminal tail, blue. Nucleosome elements are also shown with the histone H3 N-terminal tail coloured in green and the DNA strands shown as surfaces (yellow and grey respectively).

Supplemental Tables

Supplemental Table S1 - Histone Arrays - Peptide Sequences (K14 and K36 SPOT arrays) used in the membranes of **Figure 2**

Histone H3 Peptide Length (1 - 21)	Histone	Array Position	Peptide Sequence
CTRL	H3	**A1**	HHHHHHHHHHH
CTRL	H4	**A2**	HHHHHHHHHHH
	H5	**A3**	
WT	H6	**A4**	ARTKQTARKSTGGKAPRKQLA
K14ac	H7	**A5**	ARTKQTARKSTGG**Kac**APRKQLA
K4ac/K14ac	H8	**A6**	ART**Kac**QTARKSTGG**Kac**APRKQLA
K4me/K14ac	H9	**A7**	ART**Kme**QTARKSTGG**Kac**APRKQLA
K4me2/K14ac	H10	**A8**	ART**Kme$_2$**QTARKSTGG**Kac**APRKQLA
K4me3/K14ac	H11	**A9**	ART**Kme$_3$**QTARKSTGG**Kac**APRKQLA
K9ac/K14ac	H12	**B1**	ARTKQTAR**Kac**STGG**Kac**APRKQLA
K9me/K14ac	H13	**B2**	ARTKQTAR**Kme**STGG**Kac**APRKQLA
K9me2/K14ac	H14	**B3**	ARTKQTAR**Kme$_2$**STGG**Kac**APRKQLA
K9me3/K14ac	H15	**B4**	ARTKQTAR**Kme$_3$**STGG**Kac**APRKQLA
	H16	**B5**	
K4ac/K9ac/K14ac	H17	**B6**	ART**Kac**QTAR**Kac**STGG**Kac**APRKQLA
K4ac/K9me/K14ac	H18	**B7**	ART**Kac**QTAR**Kme**STGG**Kac**APRKQLA
K4ac/K9me2/K14ac	H19	**B8**	ART**Kac**QTAR**Kme$_2$**STGG**Kac**APRKQLA
K4ac/K9me3/K14ac	H20	**B9**	ART**Kac**QTAR**Kme$_3$**STGG**Kac**APRKQLA
K4me/K9ac/K14ac	H21	**C1**	ART**Kme**QTAR**Kac**STGG**Kac**APRKQLA
K4me/K9me/K14ac	H22	**C2**	ART**Kme**QTAR**Kme**STGG**Kac**APRKQLA
K4me/K9me2/K14ac	H23	**C3**	ART**Kme**QTAR**Kme$_2$**STGG**Kac**APRKQLA
K4me/K9me3/K14ac	H24	**C4**	ART**Kme**QTAR**Kme$_3$**STGG**Kac**APRKQLA
	H25	**C5**	
K4me2/K9ac/K14ac	H26	**C6**	ART**Kme$_2$**QTAR**Kac**STGG**Kac**APRKQLA
K4me2/K9me/K14ac	H27	**C7**	ART**Kme$_2$**QTAR**Kme**STGG**Kac**APRKQLA
K4me2/K9me2/K14ac	H28	**C8**	ART**Kme$_2$**QTAR**Kme$_2$**STGG**Kac**APRKQLA
K4me2/K9me3/K14ac	H29	**C9**	ART**Kme$_2$**QTAR**Kme$_3$**STGG**Kac**APRKQLA
K4me3/K9ac/K14ac	H30	**D1**	ART**Kme$_3$**QTAR**Kac**STGG**Kac**APRKQLA
K4me3/K9me/K14ac	H31	**D2**	ART**Kme$_3$**QTAR**Kme**STGG**Kac**APRKQLA
K4me3/K9me2/K14ac	H32	**D3**	ART**Kme$_3$**QTAR**Kme$_2$**STGG**Kac**APRKQLA
K4me3/K9me3/K14ac	H33	**D4**	ART**Kme$_3$**QTAR**Kme$_3$**STGG**Kac**APRKQLA

Histone H3.X Peptide Length (22 - 42)	Histone	Array Position	Peptide Sequence
K36me3	H3.1	**A1**	TKAARKSAPATGGV**Kme$_3$**KPHRYR
K36me3	H3.1t	**A2**	TKVARKSAPATGGV**Kme$_3$**KPHRYR
K36me3	H3.3	**A3**	TKAARKSAP**S**TGGV**Kme$_3$**KPHRYR
K36me3	H3.3C	**A4**	TKAARKSTP**S**T**C**GV**Kme$_3$**PHRYR

pT32/K36me3	H3.1	**A5**	TKAARKSAPA**pT**GGV**Kme$_3$**KPHRYR
pT32/K36me3	H3.1t	**A6**	TKVARKSAPA**pT**GGV**Kme$_3$**KPHRYR
pT32/K36me3	H3.3	**A7**	TKAARKSAP**SpT**GGV**Kme$_3$**KPHRYR
pT32/K36me3	H3.3C	**A8**	TKAARKSTP**SpTC**GV**Kme$_3$**PHRYR
K36me2	H3.1	**B1**	TKAARKSAPATGGV**Kme$_2$**KPHRYR
K36me2	H3.1t	**B2**	TKVARKSAPATGGV**Kme$_2$**KPHRYR
K36me2	H3.3	**B3**	TKAARKSAP**S**TGGV**Kme$_2$**KPHRYR
K36me2	H3.3C	**B4**	TKAARKSTP**S**T**C**GV**Kme$_2$**PHRYR
pT32/K36me2	H3.1	**B5**	TKAARKSAPA**pT**GGV**Kme$_2$**KPHRYR
pT32/K36me2	H3.1t	**B6**	TKVARKSAPA**pT**GGV**Kme$_2$**KPHRYR
pT32/K36me2	H3.3	**B7**	TKAARKSAP**SpT**GGV**Kme$_2$**KPHRYR
pT32/K36me2	H3.3C	**B8**	TKAARKSTP**SpTC**GV**Kme$_2$**PHRYR
K36me	H3.1	**C1**	TKAARKSAPATGGV**Kme**KPHRYR
K36me	H3.1t	**C2**	TKVARKSAPATGGV**Kme**KPHRYR
K36me	H3.3	**C3**	TKAARKSAP**S**TGGV**Kme**KPHRYR
K36me	H3.3C	**C4**	TKAARKSTP**S**T**C**GV**Kme**PHRYR
pT32/K36me	H3.1	**C5**	TKAARKSAPA**pT**GGV**Kme**KPHRYR
pT32/K36me	H3.1t	**C6**	TKVARKSAPA**pT**GGV**Kme**KPHRYR
pT32/K36me	H3.3	**C7**	TKAARKSAP**SpT**GGV**Kme**KPHRYR
pT32/K36me	H3.3C	**C8**	TKAARKSTP**SpTC**GV**Kme**PHRYR
		C9	
CTRL		**C10**	HHHHHHHHHHH
K36	H3.1	**D1**	TKAARKSAPATGGV**K**KPHRYR
K36	H3.1t	**D2**	TKVARKSAPATGGV**K**KPHRYR
K36	H3.3	**D3**	TKAARKSAP**S**TGGV**K**KPHRYR
K36	H3.3C	**D4**	TKAARKSTP**S**T**C**GV**K**PHRYR
pT32/K36	H3.1	**D5**	TKAARKSAPA**pT**GGV**K**KPHRYR
pT32/K36	H3.1t	**D6**	TKVARKSAPA**pT**GGV**K**KPHRYR
pT32/K36	H3.3	**D7**	TKAARKSAP**SpT**GGV**K**KPHRYR
pT32/K36	H3.3C	**D8**	TKAARKSTP**SpTC**GV**K**PHRYR
		D9	(blank)
CTRL		**D10**	HHHHHHHHHHH

Supplemental Table S2: Histone Arrays - Peptide Sequences used in the membranes of **Figure 4** (separate file).

Supplemental Table S3: Synthetic biotinylated histone peptides used to probe ZMYND8 histone binding preferences, related to **Figure 2**

#	Histone	Length	modifications	Sequence
1	H3	1-20	-	Ac-ARTKQTARKSTGGKAPRKQL-K(Biot)-NH2
2	H3	15-36	-	Ac-APRKQLATKAARKSAPSTGG-Peg-Biot
3	H3	1-20	K4me	ARTKmeQTARKSTGGKAPRKQL-K(Biot)-NH2
4	H3	1-20	$K4me_3$	$ARTKme_2$QTARKSTGGKAPRKQL-K(Biot)-NH2
5	H3	15-34	K27me	Ac-APRKQLATKAARKmeSAPSTGG-Peg-Biot
6	H3	15-34	$K27me_2$	Ac-APRKQLATKAARKme_2SAPSTGG-Peg-Biot
7	H3	15-34	$K27me_3$	Ac-APRKQLATKAARKme_3SAPSTGG-Peg-Biot
8	H3	1-20	K14ac	Ac-ARTKQTARKSTGGKacAPRKQL-K(Biot)-NH2
9	H3	1-20	K4acK14ac	Ac-ARTKacQTARKSTGGKacAPRKQL-K(Biot)-NH2
10	H3	1-20	K9acK14ac	Ac-ARTKQTARKacSTGGKacAPRKQL-K(Biot)-NH2
11	H3	1-20	K14acK18ac	Ac-ARTKQTARKSTGGKacAPRKacQL-K(Biot)-NH2
12	H3	1-20	K4acK9acK14ac	Ac-ARTKacQTARKacSTGGKacAPRKQL-K(Biot)-NH2
13	H3	1-20	K9acK14acK18ac	Ac-ARTKQTARKacSTGGKacAPRKacQL-K(Biot)-NH2
14	H3	1-20	K4meK9acK14acK18ac	Ac-ARTKmeQTARKacSTGGKacAPRKacQL-K(Biot)-NH2
15	H3	1-20	$K4me_3$K14ac	Ac-ARTKme_3QTARKSTGGKacAPRKQL-K(Biot)-NH2
16	H3	1-20	$K4me_3$K9acK14ac	Ac-ARTKme_3QTARKacSTGGKacAPRKQL-K(Biot)-NH2
17	H3	1-20	$K4me_3$K9me_3	Ac-ARTKme_3QTARKme_3STGGKAPRKQL-K(Biot)-NH2
18	H3	1-20	$K4me_3$R8$me_{2,a}$K9me_3	Ac-ARTKme_3QTAR$me_{2,a}$Kme_3STGGKAPRKQL-K(Biot)-NH2
19	H3.3	15-43	-	Ac-APRKQLATKAARKSAPSTGGVKKPHRY-GG-K(Biot)-NH2
20	H3.3	15-43	$K36me_3$	Ac-APRKQLATKAARKSAPSTGGVKme_3KPHRY-GG-K(Biot)-NH2
21	H3.3	15-43	$K18me_3$K36me_3	Ac-APRKme_3QLATKAARKSAPSTGGVKme_3KPHRY-GG-K(Biot)-NH2
22	H3.3	27-45	pT32	Ac-KSAPSpTGGVKKPHRYRPGT-G-K(Biot)-NH2
23	H3.3	27-45	pT32/K36me	Ac-KSAPSpTGGVKmeKPHRYRPGT-G-K(Biot)-NH2
24	H4	1-23	-	Ac-SGRGKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2
25	H4	1-23	K12ac	Ac-SGRGKGGKGLGKacGGAKRHRKVLR-K(Biot)-NH2
26	H4	1-23	K16ac	Ac-SGRGKGGKGLGKGGAKacRHRKVLR-K(Biot)-NH2
27	H4	1-23	K12acK16ac	Ac-SGRGKGGKGLGKacGGAKacRHRKVLR-K(Biot)-NH2

Supplemental Table S4 – Isothermal Titration Calorimetry of human ZMYND8 reader ensemble with recombinant 40-amino-acid long histone H3 peptides bearing AMBER codon deposited acetyl-lysine residues. Titrations were carried out in 20 mM HEPES pH 7.5 (at 25 °C), 200 mM NaCl and 15 °C while stirring at 1000 rpm. Peptides were titrated into protein solutions. Data related to **Figure 2**.

	Peptide	[P] (μM)	[Pep] (μM)	K_D (μM)	ΔH^{obs} (kcal/mol)	N	TΔS (kcal/mol)	ΔG (kcal/mol)
H3.1	WT	30	777	19.0 ± 0.8	-7.42 ± 0.26	0.99 ± 0.028	-1.20	-6.22
	T32E	31	789	20.7 ± 0.9	-8.02 ± 0.26	1.05 ± 0.027	-1.84	-6.18
	K14ac	31	790	15.1 ± 0.6	-12.02 ± 0.37	0.79 ± 0.020	-5.64	-6.38
	K14ac/T32E	31	795	17.5 ± 0.8	-13.15 ± 0.47	0.81 ± 0.024	-6.88	-6.27
H3.3	WT	30	720	19.6 ± 0.8	-8.00 ± 0.26	1.03 ± 0.026	-1.79	-6.21
	T32E	29	761	17.4 ± 1.1	-8.62 ± 0.43	1.00 ± 0.040	-2.34	-6.28
	K14ac	27	800	6.8 ± 0.2	-16.49 ± 0.19	0.87 ± 0.079	-9.68	-6.81
	K14ac/T32E	32	779	6.2 ± 0.2	-17.26 ± 0.19	0.83 ± 0.069	-10.40	-6.86
	polyKac	29	471			NB		
	K14ac (20-mer)	29	742	6.6 ± 0.2	-22.88 ± 0.25	0.83 ± 0.070	-16.04	-6.84
	K14ac (mBRD)[a]	30	785	12.2 ± 0.3	-2.70 ± 0.05	0.95 ± 0.013	3.77	-6.47
	K14ac (mPHD)[b]	34.1	772			NB		
	K14ac (mPHD/BRD)[c]	35.4	772			NB		
	K14ac (mPWWP)[d]	27	761	8.6	-15.84 ± 0.72	0.49 ± 0.019	-9.16	-6.68
	K14ac/T32E (mPWWP)[d]	27	761	23.0	-22.61 ± 0.86	0.88 ± 0.029	-16.47	-6.14

[a] mBRD: BRD mutant (N228F); [c] mPHD/BRD: PHD+BRD mutant (N87A, E104A, D124A, N228F)
[b] mPHD: PHD mutant (N87A, E104A, D124A); [d] mPWWP: PWWP mutant (F288A, W291A)

Supplemental Table S5 – Isothermal Titration Calorimetry of human ZMYND8 reader ensemble with recombinant 40-amino-acid long histone H4 peptides bearing AMBER codon deposited acetyl-lysine residues. Titrations were carried out in 20 mM HEPES pH 7.5 (at 25 °C), 200 mM NaCl and 15 °C while stirring at 1000 rpm. Peptides were titrated into protein solutions. Data related to **Figure 2**.

	Peptide	[P] (μM)	[Pep] (μM)	K_D (μM)	ΔH^{obs} (kcal/mol)	N	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)
H4	WT	30	1025			NB		
	K8ac	33	814			NB		
	K12ac	29	760	31.6 ± 5.0	-2.33 ± 0.59	0.83 ± 0.179	3.41	-5.74
	K16ac	30	736			NB		
	poly-Kac	32	805	37.7 ± 1.2	-5.65 ± 0.24	1.00 ± 0.036	0.18	-5.83
	K12ac (mBRD)[a]	37	727			NB		
	K12ac (mPHD)[b]	41	706	35.0 ± 2.2	-2.51 ± 0.20	0.75 ± 0.048	3.37	-5.88

[a] mBRD: BRD mutant (N228F); [b] mPHD: PHD mutant (N87A, E104A, D124A)

Supplemental Table S6 – Isothermal Titration Calorimetry of human BAZ2B (UniProt: Q9UIF8 residues 2054-2168 WT or N2140F mutant) with recombinant histone H3 peptides bearing K14ac. Titrations were carried out in 20 mM HEPES pH 7.5 (at 25 °C), 200 mM NaCl and 15 °C while stirring at 1000 rpm. Peptides were titrated into protein solutions. Data related to **Figure 5**.

BAZ2B	Peptide	[P] (μM)	[Pep] (μM)	K_D (μM)	ΔH^{obs} (kcal/mol)	N	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)
WT[¥]	$H3_{(9-19)}$K14ac	50	1550	7.6 ± 0.3	-8.91 ± 0.08	0.99 ± 0.007	-2.27	-6.64
WT[‖]	$H3_{(5-25)}$K14ac	40	1400	6.3 ± 0.2	-6.86 ± 0.06	1.00 ± 0.007	-0.12	-6.74
WT	$H3_{(1-20)}$(no Kac)	60	750			NB		
WT	$H3_{(1-20)}$K14ac	60	750	17.3 ± 0.9	-4.22 ± 0.15	0.94 ± 0.026	2.06	-6.28
N2140F	$H3_{(1-20)}$K14ac	60	750			NB		

[¥] data from (Filippakopoulos et al., 2012)
[‖] data from (Philpott et al., 2011)

Supplemental Table S7: Data collection and refinement statistics, related to the crystal structure shown in **Figure 1**.

Data Collection		
PDB	4COS	
Protein/Ligand	ZMYND8	ZMYND8 (Zn SAD)
Beamline	DLS I04	DLS I03
Wavelength	0.9796	1.2829
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions: a, b, c (Å) α, β, γ (deg)	66.96 68.83 70.35 90.00 90.00 90.00	66.06 69.01 70.34 90.00 90.00 90.00
Resolution* (Å)	1.67 (1.77-1.67)	1.78 (1.83-1.78)
Unique observations*		
Completeness* (%)	99.9 (99.7)	99.5 (98.8)
Redundancy*	6.0 (4.5)	12.5 (11.7)
R_{sym}*	0.077 (0.767)	0.115 (0.990)
I/ σI*	12.2 (2.0)	18.5 (3.6)
Refinement		
Resolution (Å)	1.67	
R_{work} / R_{free} (%)	17.2/21.0	
Number of atoms (protein/other/water)	2553/15/333	
B-factors (Å^2) (protein/other/water)	31.5/59.7/40.2	
r.m.s.d bonds (Å) r.m.s.d angles (^{o})	0.015 1.661	
Ramachadran Favoured (%) Allowed (%) Disallowed (%)	97.09 2.59 0.32	

* Values in parentheses correspond to the highest resolution shell.

Rsym is the unweighted R-value on I between symmetry mates.

Rwork = Σhkl | |Fobs (hkl)| - k |Fcalc (hkl)| | / Σhkl |Fobs (hkl)| for the working set of reflections

Rfree is the R-value for 5% of the reflections excluded from refinement.

Supplemental Table S8: Primers used for EMSA assays related to **Figure 4**

DNA Probe	Primer	Sequence
AT-rich	fwd	CATGTTTTAATAGTATCTGCAATTTGTAATATGTCTA**c**atg
	rvs	gta**c**AAAATTATCATAGACGTTAAACATTATACAGATGTAC
GC-rich	fwd	CATGGCGGGGCCCCGCGACGCGGCTGCCTGCGGGGTG**c**atg
	rvs	gta**c**CGCCCCGGGGCGCTGCGCCGACGGACGCCCCACGTAC

fwd: forward primer; **rvs**: reverse primer;

Supplemental Table S9: Primers used ChIP-PCR experiments related to **Figure 5**.

Gene	NCBI		Sequence (5' →3')	Tm (˚C)	Location	Transcript Length
CLCN7	NM_001287	fwd	AAGAACAGAAGGCGAGT	50.3	chr16 1525587-1525658	72
		rvs	CAGTTTCCAGAGGCGAC	55.0		
CREBBP	NM_004380	fwd	GTTACAGACCGCCCTTG	54.9	chr16 3931258-3931342	85
		rvs	GACCACTTTGGAGCAGTTG	56.3		
ETS2	NM_001256295	fwd	CCTTATTACCCAAGCCCG	58.0	chr21 40177294-40177382	89
		rvs	CCCTAACGCCAAACCTG	57.6		
PPP4C	NM_002720	fwd	TGAAAGAGGGAGGCAGG	57.3	chr16 30087365-30087437	73
		rvs	TAGAGGAGTCTCGCACTATTAC	53.2		
SLC7A5	NR_002594	fwd	TATGTCGGGCATTCCTCT	56.5	chr16 87904472-87904572	101
		rvs	ACAAGTTAGTATCTGTGTGACC	51.0		

fwd: forward primer; **rvs**: reverse primer

Online Methods

Cloning and Mutagenesis

cDNA encoding the full length ZMYND8 (Uniprot: Q9ULU4, isoform 13) was reconstructed following the Gibson assembly protocol (Gibson et al., 2010; Gibson et al., 2009) from a PCR fragment spanning residues T95-D1188, amplified from the MGC cDNA clone (IMAGE 4844473) and oligonucleotides encoding the first 94 amino-acids that were missing in the MGC clone. The construct was subsequently cloned into a pDONR221 vector. Single point mutations or their combinations were introduced into the full length ZMYND8 Gateway entry clone using 15 cycles of the QuikChange II PCR protocol (Agilent Technologies). Full length ZMYND8 mammalian expression vectors encoding either an N-terminal 3x-FLAG, EGFP or BirA*-FLAG tags were constructed following the Gateway LR recombination reaction between relevant destination vectors previously described (Lambert et al., 2014) and wild type or mutated ZMYND8 Gateway entry clones. Wild type or mutated forms of the N-terminal reader modules of ZMYND8 (PHD, Bromo and PWWP domains, aa Q83-S406) for bacterial expression were amplified by PCR and sub-cloned into the pNIC-ZB vector (Genbank: GU452710) using the ligation-independent cloning (LIC) protocol (Aslanidis and de Jong, 1990).

Gateway entry clones containing three copies of either the wild type or N2140F substituted bromodomain of BAZ2B (S2054 – S2168 of Uniprot accession Q9UIF8) and C-terminal NLS sequence from the SV40 Large T-antigen (PKKKRKV) were constructed by multiple ligation independent cloning (LIC) protocol (Aslanidis and de Jong, 1990) in the pENTR-221 vector (Invitrogen). The GFP, mCherry or 3*HA tagged 3x-BAZ2B bromodomain constructs were constructed by Gateway LR recombination reaction between appropriate destination vector and an entry clone encoding for a wild type or mutated triple bromodomains.

Generation of Stable Cell Lines

Constructs for the genes of interest were generated via Gateway cloning into pDEST 5’ Triple-FLAG-pcDNA5-FRT-TO or pDEST 5’ BirA*-FLAG-pcDNA5-FRT-TO vectors. Proteins of interest were stably expressed in T-REx Flp-In HEK293 cells as previously described (Couzens et al., 2013).

Protein Expression and Purification

Recombinant wild type or mutated ZMYND8 (Q83-S406) proteins were expressed and purified as previously described (Savitsky et al., 2010). Briefly, E.coli BL21(DE3) cells containing a pRARE2 plasmid were transformed with a pNIC-ZB vector encoding the N-terminal triple reader domains of ZMYND8. Small (50 ml) cultures inoculated with several colonies were grown overnight at 37 ˚C in terrific broth (TB) media and were used to inoculate 2 l of TB with K/Na phosphates substituted with 5 g/l NaCl to prevent $ZnCl_2$ precipitation at 37 ˚C with aeration until the optical density (λ = 600 nm) reached a value of 3. The temperature was reduced to 18 ˚C and the cultures were allowed to cool down for 1 h before 100 μM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and 1 mM of $ZnCl_2$ were added (final concentrations) to induce protein expression overnight. Cultures were harvested and cell pellets were resuspended in lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) and 5 % w/v glycerol) and mechanically lysed following 3 passes through an Avestin C5 high pressure homogeniser at 4 ˚C. The lysates were span down and supernatants were collected and loaded onto 5 ml gravity flow Ni-IDA columns. Each column was washed with at least 10 column volumes of lysis buffer containing 50 mM imidazole and eluted with elution buffer (20 mM HEPES, pH7.5, 250 mM NaCl, 300 mM imidazole, 5 % w/v glycerol, 0.5 mM TCEP) in 10 ml fractions. Fractions containing recombinant ZMYND8 protein were directly loaded onto an HP SP column on an AKTA Purifier, were eluted with a 0.25 – 1 M NaCl gradient and were combined. TCEP was added to a final concentration of 20 mM and the Z-Basic (ZB) tag was removed by overnight incubation at 4 ˚C with TEV protease (at 1:100 w/w). Following tag cleavage, the protein was re-purified through a Ni-IDA column in order to remove the ZB tag. The flow-through fraction was collected, concentrated and loaded onto a Superdex S75 16/60 column equilibrated with GF buffer (50 mM HEPES, pH 7.5, 200 mM NaCl, 2 % w/v glycerol, 20 mM TCEP). Recombinant ZMYND8 eluted as a single symmetrical monomeric peak, fractions were pooled and were either flash frozen in liquid nitrogen and stored in -80C freezer or used directly in crystallisation or biophysical experiments. All purification steps were performed on ice or in a cold room at 4 ˚C.

Custom Peptide Synthesis

Custom histone peptides carrying specific modifications were synthesised as previously described (Rothbart et al., 2012b). Sequences are tabulate in **Supplemental Table S3**.

Recombinant Histone Peptides

Histone H3 and H4 peptides encoding the N-terminal portion of the proteins (20 or 40 amino-acid long) were amplified in a bacterial expression system using a pCDFDuet™-1 (Novagen) plasmid carrying the ORF for the histone-tail fragments of interest, in order to yield sufficient amounts for biophysical characterization of histone interactions with ZMYND8. In brief, all peptides were expressed as his-tagged proteins and were purified by nickel affinity and size exclusion chromatography. Peptide integrity was assessed by SDS-PAGE and electro-spray mass spectrometry on an Agilent 6530 QTOF (Agilent Technologies Inc. Palo Alto, CA) mass spectrometer. Poly-acetylation of the peptides was performed essentially as previously described (Smith, 2005). In addition, histone peptides carrying specific acetylations (histone H3 at K14; histone H4 at K8 or K12 or K16) were produced *in vivo* using the amber stop codon/suppressor tRNA technology developed by Neumann and colleagues (Neumann et al., 2009) with several modifications that increased overall yields from 10% to above 70 %. In brief, ORF for the histone-tail fragments of interest presenting an amber codon (TAG) engineered at the desired sites were co-expressed in Tuner™ BL21 DE3 cells (Novagen, #70622) with an orthogonal N(epsilon)-acetyl-lysyl-tRNA synthetase/tRNA(CUA) pair, leading to a site specific incorporation of N(epsilon)-acetyl-lysine. Cells were grown at 37 ˚C in Terrific Broth (TB) medium previously inoculated with 10 ml of an overnight culture.

At OD_{600} of ~0.7 the culture was supplemented with 20 mM nicotinamide (NAM, Sigma, #N3376-100G) and 25 mM acetyl-lysine (H-Lys-Ac-OH, GL Biochem Shangai Ltd, #GLS140325). Protein expression was induced 30 min later by addition of 50 µM IPTG and cells were harvested by centrifugation (8,700 x *g*, 15 min, 4 ˚C; Beckman Coulter Avanti J-20 XP centrifuge) after 6 h at 37 ˚C. Peptides were then purified as described above.

SPOT Peptide Assays

Cellulose-bound peptide arrays were prepared employing Fmoc solid phase peptide synthesis using a MultiPep-RSi-Spotter (INTAVIS, Köln, Germany) according to the SPOT synthesis method

provided by the manufacturer as previously described (Picaud and Filippakopoulos, 2015). Human histone peptides (using UniProt accession codes Q6FI13 (histone H2A), P16104 (histone H2A.X), P0C0S5 (histone H2A.Z), P62807 (histone H2B), P68431 (histone H3.1), Q16695 (histone H3.1t), P84243 (histone H3.3), Q6NXT2 (histone H3.3C) and P62805(histone H4)) were synthesized on amino-functionalized cellulose membranes (Whatman™ Chromatography paper Grade 1CHR, GE Healthcare Life Sciences #3001-878) and the presence of SPOTed peptides was confirmed by ultraviolet light (UV, λ = 280 nM). The assay was performed as previously described (Filippakopoulos, 2012) using the his-6-tagged recombinant triple readers of ZMYND8 as bait protein. ZMYND8 protein bound to peptides was detected using antibody HPR conjugated (Novagene, # 71841) and the Pierce® ECL Western blotting Substrate (Thermo Scientific, # 32106). Chemiluminescence was detected with an image reader (Fujifilm LAS-4000 ver.2.0) typically using an incremental exposure time of 5 min for a total of 80 min (or until saturation was reached, in the case of very strong signal). Peptide locations on the arrays and their sequences are given in **Supplemental Tables S1 and S2**.

Biolayer Interferometry (BLI)

Experiments were performed on an Octet RED384 system (FortéBio) at 25 ˚C in 20 mM HEPES, pH 7.5, 200 mM NaCl and 20 mM TCEP using the FortéBio data acquisition software V.7.1.0.100. Biotinylated peptides were first immobilized onto Super Streptavidin biosensors (SuperStreptavidin (SSA) Dip and Read Biosensors for kinetic #18-0011, FortéBio), pre-equilibrated in the BLI buffer then quenched in a solution of 5 µM Biotin. (baseline equilibration 60 sec, peptide loading for 240 sec, quenching for 60 sec, 1000 x rpm shake speed, at 25 ˚C). The immobilized peptides were subsequently used in association and dissociation measurements performed within a time window of 600 sec (base line equilibration 120 sec, association for 240 sec, dissociation for 240 sec, 1000 x rpm shake speed, at 25 ˚C). Interference patterns from peptide-coated biosensors without protein were used as controls. After referencing corrections, the subtracted binding interference data were analyzed using the FortéBio analysis software (FortéBio data analysis software V.7.1.0.38) provided with the instrument following the manufacturer's protocols. Histone peptide binding to ZMYND8 was first explored against a set of biotinylated histone peptides commercially available (AltaBioSciences Histone array, Set 3 & Set 4 Histone Acetyl-Lysine library) carrying 21-mers with

single or multiple modifications, including lysine acetylation and methylation. Custom made histone H3 and H4 20-amino-acid long peptides with specific modification were then tested by employing a range of ZMYND8 concentrations (1, 2, 4, 8, 16, 32, 64 and 128 μM) in order to determine binding constants. Peptide sequences are given in **Supplementary Table S3**.

Isothermal Titration Calorimetry (ITC)

Experiments were carried out on an ITC200 titration microcalorimeter from MicroCal™, LLC (GE Healthcare) equipped with a Washing module, with a cell volume of 0.2003 ml and a 40 μl microsyringe. Experiments were carried out at 15 ˚C while stirring at 1000 rpm, in ITC buffer (20 mM HEPES pH 7.5 (at 25 °C), 200 mM NaCl). The microsyringe was loaded with a solution of peptide sample (470 - 1100 μM, in ITC buffer) and was carefully inserted into the calorimetric cell which was filled with an amount of the protein (0.2 ml, 27-33 μM in ITC buffer). Following baseline equilibration an additional delay of 60 sec was applied. All titrations were conducted using an initial control injection of 0.3 μl followed by 38 identical injections of 1 μl with a duration of 2 sec (per injection) and a spacing of 120 sec between injections. The titration experiments were designed as to ensure complete saturation of the proteins before the final injection. The heat of dilution for the peptides were independent of their concentration and corresponded to the heat observed from the last injection, following saturation of ligand binding, thus facilitating the estimation of the baseline of each titration from the last injection. The collected data were corrected for peptide heats of dilution (measured on separate experiments by titrating the proteins into ITC buffer) and deconvoluted using the MicroCal™ Origin software supplied with the instrument to yield enthalpies of binding (ΔH) and binding constants (K_B) in the same fashion to that previously described in detail by Wiseman and co-workers (Wiseman et al., 1989). Thermodynamic parameters were calculated using the basic equation of thermodynamics ($\Delta G = \Delta H - T\Delta S = -RT\ln K_B$, where ΔG, ΔH and ΔS are the changes in free energy, enthalpy and entropy of binding respectively). In all cases a single binding site model was employed, supplied with the MicroCal™ Origin software package. Dissociation constants and thermodynamic parameters are listed on **Supplementary Tables S4** and **S5**. Thermodynamic parameters for BAZ2B interactions with H3K14ac are listed in **Supplementary Table S6**.

Crystallization

Aliquots of the purified protein were set up for crystallization using a mosquito® crystallization robot (TTP Labtech, Royston UK). Coarse screens were typically setup onto Greiner 3-well plates using three different drop ratios of precipitant to protein per condition (100+50 nl, 75+75 nl and 50+100 nl). Initial hits were optimized further using Greiner 1-well plates and scaling up the drop sizes in steps. All crystallizations were carried out using the sitting drop vapour diffusion method at 4 °C. Crystals of ZMYND8 were grown by mixing 100 nl of the protein (11 mg/ml in 10 mM HEPES pH 7.5, 500 mM NaCl, 5 % Glycerol) with 50 nl of reservoir solution containing 1.8 M $(NH_4)_2SO_4$, 0.1 M MES pH 6.3 and 15 % (v/v) dioxane. Plate-like crystals appeared within several days from sitting drop plates at 4 ˚C. Prior to data collection, all crystals were transferred to a solution consisting of the precipitation buffer supplemented with 2M Li_2SO_4 and subsequently flash frozen in liquid nitrogen.

Data Collection and Structure Determination

Zn-SAD data were collected at Diamond Lightsource from one crystal close to the Zn K-edge at 1.2829 Å on beamline I03 and a second crystal was used to obtain higher resolution on beamline I04 at a wavelength of 0.9796 Å. Both datasets were integrated with XDS (Kabsch, 2010) and scaled with SCALA or AIMLESS (Evans and Murshudov, 2013). SHELXD (Sheldrick, 2010) was used to identify 3 putative Zn sites in the initial dataset. The electron density map after phase refinement with autoSHARP (Vonrhein et al., 2007) and subsequent density modification with SOLOMON (Abrahams and Leslie, 1996) was of excellent quality. Automated model building with BUCCANEER (Cowtan, 2006) resulted in a more than 90 % complete model, which was from then refined against the higher resolution dataset. Refinement was carried out with REFMAC (Murshudov et al., 2011) and after several rounds of manual rebuilding with COOT (Emsley et al., 2010), the model converged to a final Rcryst/Rfree of 17.2 % and 21.0 %, respectively. The quality of the final model was validated with MOLPROBITY (Chen et al., 2010). Thermal motions were analyzed using TLSMD (Painter and Merritt, 2006) and hydrogen atoms were included in late refinement cycles. Data collection and refinement statistics can be found in **Supplemental Table S7**. The model and structure factors have been deposited with PDB accession code: **4COS**.

Molecular Modelling

The ZMYND8 structure was manually positioned on the core nucleosome structure (PDB code 3AFA) (Tachiwana et al., 2010) so that the aromatic cage of the PWWP domain overlapped with the expected location of H3K36. While electron density for H3K37 is present in the experimental nucleosome structure, no density is available for H3K36. The manually positioned model was further optimized in the internal coordinates space with ICM (version 3.8-2) with loose distance restraints (10 Å < d < 20 Å) imposed between the side-chains of W291, Q310, F307 of ZMYND8 and the phosphate groups of T6, C84 (of strand i) and T292 (of strand j) on the nucleosome respectively, during two independent rigid docking Monte Carlo energy minimization cycles of 1.5 million steps each that converged towards the same low-energy docked conformation. Next, a histone H3K14ac peptide tail (residues 1 to 36) was docked with ICM to the nucleosome-ZMYND8 model with the following constraints: the epsilon nitrogen of H3K36 was restrained in the PWWP aromatic cage, the side-chain of T32 was restrained in the cavity at the interface of the PWWP and BRD domains (methyl group of T32 less than 5 Å from the A262 side-chain), the side-chain of K14ac was restrained in the bromodomain binding pocket (acetyl oxygen of K14 less than 5 Å from the amide nitrogen of N228) and the side-chain of A1 and K4 were restrained at the PHD finger (A1 side-chain less than 5 Å from P121 and K4 side-chain less than 5 Å from F89). A 36 million step Monte Carlo energy minimization simulation was conducted at this step with fully flexible H3 and flexible ZMYND8 side-chains.

In order to merge the histone H3 tail (docked to the nucleosome-ZMYND8 complex) with the experimental structure of the nucleosomal H3 core, a full-length H3 structure was built, residues 1 to 36 were tethered to the docking model, residues 39 to 135 were tethered to the H3 core (using PDB code 3AFA) and the energy of the system was again relaxed with ICM. Finally, a phosphate group was added to T32 and the energy of the system was minimized with the ICM “loop” algorithm where the loop was defined as residues 29 to 34 of H3.

Chromatin Fractionation Assay

Chromatin fractionation assays were performed with HEK293T cells transfected with 1 µg of the indicated plasmids as previously described (Rothbart et al., 2012a). Cells were collected 36 h post-transfection, and western blots were performed with 8 µg of protein. The following antibodies were

used to assess ZMYND8 chromatin association: Flag (Sigma), H4 (Abcam), and b-tubulin (Cell Signaling Technologies).

Cell Culture

Flp-In T-REx HEK293 cells, stably expressing full length 3xFLAG- or BirA*-FLAG-tagged ZMYND8 and 3xFLAG-ZMYND8 N228F mutant, were maintained in DMEM, GlutaMAX^{TM} supplement (Cat.#10566-016; Gibco), supplemented with 10 % heat-inactivated foetal bovine serum (FBS - Sigma), 100 U/ml penicillin (Gibco), 100 µg/ml streptomycin (Gibco), and 200 µg/ml Hygromycin (Invitrogen). Cells were grown at 37 °C in a humidified cabinet at 5 % CO_2. Protein expression was induced by treatment of cells with 1 µg/ml Tetracycline for 24 h.

Histone Immuno-precipitation

Cells were lysed on ice in a buffer containing 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 % NP-40, 1 mM PMSF (P7626, Sigma) and protease inhibitor cocktail (#539134, Calbiochem). Lysates were treated with 100 U of TurboNuclease (T4330, Sigma) and 1 mM $MgCl_2$ for 1 h at 4 °C to further digest chromatin. Finally lysates were centrifuged at 21,130xg for 30 min at 4 °C, and the supernatant was collected. Co-immunoprecipitations were performed with anti-FLAG M2 antibody (F1804, Sigma) and Dynabeads Protein G (10003D, Life Technologies). Cell lysates were mixed with 1 µg of anti-FLAG M2 and 10 µl of Dynabeads Protein G, and incubated 2 h at 4 °C (constant rotation). The beads were washed with washing buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM EGTA and 1 % Triton X-100) and washing buffer 2 (20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM EGTA). Samples were then dissolved for electrophoresis in Laemmli sample buffer (300 mM Tris-HCl pH6.8, 10 % SDS, 50 % glycerol and 0.05 % Bromophenol blue) and incubated at 98 °C for 3 min. Samples were fractionated by electrophoresis in a 4-12 % Bis-Tris Midi Protein Gels (WG1403, Life technologies) and transferred to a PVDF membrane (10600023, GE healthcare). The membrane was blocked for 1 h with 5 % semi-skimmed milk in T-TBS and was then immunoblotted with indicated primary antibodies overnight at 4 °C in 5 % BSA (A7906, Sigma) in T-TBS at a 1:2000 dilution. After 3 washes with T-TBS the membrane was incubated with anti-rabbit (7074, Cell Signalling) or anti-mouse (A4416, Sigma) HRP-linked antibodies at 1:5000 dilution in 5 % BSA in T-TBS for 2 h at room temperature. After 5-10 washes for 2 h with T-TBS the membrane was incubated with ECL (RPN2209, GE healthcare), and exposed to film.

Antibodies used for immunoblotting: Histone H3 (ab1791, Abcam); H3K4me (pAb-194-050, Diagenode); H3K4me2 (ab7766, Abcam); H3K4me3 (pAb-003-050, Diagenode); H3K9me2 (ab1220, Abcam); H3K9me3 (ab8898, Abcam); H3K9ac (07-352, Millipore); H3K14ac (ab52946, abcam); H3K27me3 (07-449, Millipore); H3K36me (ab9048, Abcam); H3K36me2 (07-369-I, Millipore); H3K36me3 (Lys36) (ab9050, Abcam); H3pT32 (ab4076, Abcam); H3pS10 (ab5176, Abcam); Histone H4 (05-858, Millipore); H4K5ac (07-327, Millipore); H4K8ac (ab15823, Abcam); H4K12ac (07-595, Millipore); H4K16ac (07-329, Millipore); H4K20ac (61531, Active Motif).

Fluorescent Recovery After Photo-bleaching (FRAP)

FRAP experiments were performed using a protocol modified from previous studies (Filippakopoulos et al., 2010). Briefly, U2OS Flp-In/T-REx cells were transfected (FuGENE6, Qiagen) with mammalian over-expression constructs encoding GFP chimeras with WT or mutant ZMYND8 (BRD mutant: N228F; PWWP mutant: F288A/W291A; BRD/PWWP mutant: N228F + F288A/W291A; PHD/BRD/PWWP mutant: E104A/N87A/D124A, N228F, F288A/W291A). Transfected cells were plated onto an 8-well imaging chamber (Miltenyi Biotec) and pre-sensitized with 10 μM 5-bromo-2'-deoxyuridine (BrdU) for 20-40 h. The FRAP and imaging system consisted of a Zeiss LSM 710 scan-head (Zeiss GmbH, Jena, Germany) coupled to an inverted Zeiss Axio Observer Z1 microscope equipped with a high-numerical-aperture (N. A. 1.40) 63x oil immersion objective (Zeiss GmbH, Jena, Germany) equipped with an incubator XLmulti S1 set to 37 °C and 5% CO_2. GFP fluorescence imaging was carried out with an argon-ion laser (λ = 488 nm) and with a piezomultiplier tube (PMT) detector set to detect fluorescence between 493-555 nm. A region of the nucleus was selected (3.09 μm^2) and after 10 pre-scans, the region was bleached (60 % laser power of 30 mW 405 nm diode laser; 12.61 μsec pixel dwell; 30 iterations). A time-lapse series was taken to record GFP recovery using 0.07-0.40 % of the power with an interval time of 2.0 sec. The image datasets and fluorescence recovery data were exported from the control software of the microscope (ZEN v.2.1) into Origin v.7 and analysed. The average intensity of pre-scans was normalized to 1. Mean signal values and SEM were determined from 15 cells per condition tested, over multiple experiments and *P* values were calculated using two-tailed Student's *t*-test or one-way analysis of variance followed by Dunnett's multiple comparison test.

Electrophoretic mobility shift assay (EMSA)

Oligonucleotide probes encoding AT-rich or GC-rich sequences (20 µg each – **Supplemental Table S8**) were annealed in 1 X React buffer 3 (Invitrogen) using the following PCR program: 94 ˚C, 2 min; 65 ˚C, 10 min; 37 ˚C, 10 min; 25 ˚C, 10 min. The probe (200 ng) was then labelled for 1 h at 25 ˚C in 1 X React buffer 1 (Invitrogen) supplemented with 200 µg/ml bovine serum albumin (BSA), 2 µM dithiothreitol (DTT), 200 µM each of dATP, dTTP, and dGTP, 1 µl Klenow polymerase (Invitrogen), and 1.5 µl α32-P-dCTP (15 µCi). Glycogen (20 µg) was then added, and the reaction mixture was brought up to 50 µl with deionised water. The DNA probe was extracted twice through 1 volume of phenol-chloroform and precipitated overnight in 1 volume ammonium acetate (5 M) and 6 volumes of ethanol at -20 ˚C. The probe was then recovered by centrifugation (16,000 x *g*; 30 min) and suspended in water. Purified recombinant protein encoding the three N-terminal reader domains of ZMYND8 (1 µl) was then added to 14 µl of band shift buffer (25 mM HEPES pH 7.4, 150 mM KCl, 200 µg/ml BSA, 5 µM DTT) on ice. The mixture was then incubated at 25 ˚C for 15 min with 1 µl of labelled probe and loaded onto a native gel containing 6 % acrylamide (44 % stock solution; 29:1 acrylamide:bis-acrylamide; Severn Biotech) in 0.5 X Tris/Borate/EDTA buffer (TBE, Gibco). Gels were polymerized by adding 0.09 % ammonium persulfate (APS) and N,N,N',N'-tetramethyl-ethylene diamine (TEMED). After running for 90 min at 180 V, gels were dried under vacuum for 1 h at 80 ˚C and were then exposed to autoradiography film, which was developed after overnight exposure at -20 ˚C.

Non radioactive EMSAs were run by annealing the DNA probes (20 µg each) in in oligo annealing buffer (1 M NaCl, 100 mM $MgCl_2$ and 500 mM Tris-HCl pH 8.0) using the following PCR program: 94 ˚C, 2 min; 65 ˚C, 10 min; 37 ˚C, 10 min; 25 ˚C, 10 min. The probe (500 nM) and purified recombinant protein encoding the three N-terminal reader domains of ZMYND8 (WT, PHD mutant, BRD/PWWP mutant or PHD/BRD/PWWP mutant) was then added to band shift buffer (25 mM HEPES pH 7.4, 150 mM KCl, 200 µg/ml BSA, 5 µM DTT) on ice. The mixture was incubated at 25 ˚C for 15 min and loaded onto a native gel containing 6 % acrylamide (44 % stock solution; 29:1 acrylamide:bis-acrylamide; Severn Biotech) in 0.5 X Tris/Borate/EDTA (TBE) buffer. Gels were polymerized by adding 0.09 % ammonium persulfate (APS) and N,N,N',N'-tetramethyl-ethylene diamine (TEMED). After running for 20 min at 150V, gels were stained with SYBR® Gold Nucleic

Acid Gel Stain (Thermo Scientific, S11494) for 1 h at room temperature. The images were acquired with ChemiDoc Imaging system (Bio-Rad).

Affinity Purification and Mass Spectrometry (AP-MS)

Stable cell lines (full length ZMYND8 WT or BRD mutant, N228F) were selectively grown in the presence of 200 µg/ml hygromycin up to 80 % confluence before expression was induced via 1 µg/ml tetracycline for 24 h and the cells were harvested. Two 150-mm plates were induced with tetracycline and treated with 50 µM biotin for 24 h before harvesting. Cells were pelleted at low speed, washed with ice-cold PBS and frozen at -80 °C until purification. Protein interactions were investigated following the FLAG AP-MS protocol previously described (Lambert et al., 2014). Briefly, cells from two 150-mm plates were pelleted, frozen down and lysed in 1.5 ml ice cold 50 mM HEPES-NaOH pH 8.0, 100 mM KCl, 2 mM EDTA, 0.1 % NP40, and 10 % glycerol with 1 mM PMSF, 1 mM DTT and Sigma protease inhibitor cocktail (P8340, 1:500) added immediately prior to processing. To aid with lysis, the cells were frozen on dry ice and thawed in a 37 ˚C water bath, and then put back on ice. Samples were sonicated at 4 ˚C using three 10 sec bursts with 2 sec pauses at 35 % amplitude. 100 U of benzonase were then added and the lysates were incubated at 4 ˚C for 1 h with rotation, then centrifuged (20,817 x *g* for 20 min at 4 ˚C) and the supernatant was added to tubes containing 25 µl of 50 % magnetic anti-FLAG M2 bead (Sigma, M8823) slurry prewashed in lysis buffer. FLAG immunoprecipitation was allowed to proceed at 4 ˚C for 2 h with rotation. Beads were pelleted by centrifugation (1000 rpm for 1 min) and magnetized, then the unbound lysate was aspirated. The beads were demagnetized, washed with 1 ml lysis buffer and magnetized again to aspirate off the wash buffer. The beads were then washed with 1 ml of 20 mM Tris-HCl (pH 8.0) containing 2 mM $CaCl_2$, then any excess wash buffer was removed by centrifugation, magnetizing and pipetting off the liquid. The now-dry magnetic beads were removed from the magnet and re-suspended in 7.5 µl of 20 mM Tris-HCl (pH 8.0) containing 750 ng of trypsin (Sigma, T7575) and the mixture was incubated at 37 ˚C with agitation overnight. After the initial incubation, beads were magnetized and the supernatant was transferred to a fresh tube. Another 250 ng of trypsin was added to the mixture followed by incubation, without agitation, for 3-4 h. The sample was acidified with formic acid to a final concentration of 2 % and the tryptic digests were stored at -40 ˚C until ready for mass spectrometry analysis. Parental Flp-In T-REx HEK293

cells and cells expressing NLS-BirA* fused to FLAG tag (not treated with excess biotin) were used as negative controls for AP-MS experiments and processed in parallel to the bait-expressing cell lines

Proximity biotinylation coupled to mass spectrometry (BioID-MS)

BioID was performed following the previously published protocol (Lambert et al., 2015). Briefly, cell pellets were thawed in 1.5 ml ice cold RIPA buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0.1 % SDS and 0.5 % sodium deoxycholate. PMSF (1 mM), DTT (1 mM) and Sigma protease inhibitor cocktail (P8340, 1:500) were added immediately before use. The lysates were sonicated, treated with benzonase and centrifuged as described in the FLAG AP-MS section. For each sample, 60 µl of streptavidin-sepharose bead slurry (GE Healthcare, Cat 17-5113-01) were pre-washed three times with 1 ml of lysis buffer by pelleting the beads with gentle centrifugation and aspirating off the supernatant before adding the next wash. Biotinylated proteins were captured on pre-washed streptavidin beads for 3 h at 4 ˚C with rotation. The beads were then gently pelleted and the unbound supernatant was saved for further analysis. The beads were then washed 2 x 1 ml with RIPA buffer and 3 x 1 ml with 50 mM ammonium bicarbonate (pH 8.0). Following the final wash, the beads were pelleted and any excess liquid was aspirated off. Beads were then re-suspended in 100 µl of 50 mM ammonium bicarbonate (pH 8.0), and 1 µg of trypsin solution was added. The samples were incubated overnight at 37 ˚C with rotation and an additional 1 µg of trypsin was then added, followed by a further incubation for 2-4 h. The beads were pelleted and the supernatant was transferred to a fresh tube. The beads were rinsed with 2 x 100 µl HPLC grade water and the wash fraction was combined with the supernatant. The peptide solution was acidified with 50 % formic acid to a final concentration of 2 % and the samples were placed in a speedvac to dry. Tryptic peptides were resuspended in 25 µl of 5 % formic acid and stored at -80 ˚C until analyzed by mass spectrometry. Parental Flp-In T-REx HEK293 cells, and stable cells expressing BirA*-FLAG fused either to a green fluorescent protein (GFP) or to a nuclear localization sequence (NLS) were used as negative controls for the BioID experiments and processed in parallel to the bait proteins.

Mass Spectrometry and Data Analysis

AP-MS samples, BioID samples and controls were analyzed by mass spectrometry in at least two biological replicates. 5 µl of each sample were directly loaded at 400 nl/min onto a 75 µm x 12 cm emitter packed with 3 µm ReproSil-Pur C_{18}-AQ (Dr.Maisch HPLC GmbH, Germany). Peptides were eluted from the column over a 90 min gradient generated by a NanoLC-Ultra 1D plus (Eksigent, Dublin CA) nano-pump and analyzed on a TripleTOF^{TM} 5600 instrument (AB SCIEX, Concord, Ontario, Canada). The gradient was delivered at 200 nl/min, starting at 2 % acetonitrile with 0.1 % formic acid and ending at 35 % acetonitrile with 0.1 % formic acid over 90 min, followed by a 15 min clean-up at 80 % acetonitrile with 0.1 % formic acid, and a 15 min equilibration period back to 2 % acetonitrile with 0.1 % formic acid for a total of 120 min. To minimize carryover between each sample, the analytical column was washed for 3 h by running an alternating "saw-tooth" gradient from 35 % acetonitrile with 0.1 % formic acid to 80 % acetonitrile with 0.1 % formic acid, holding each gradient concentration for 5 min. Analytical column and instrument performance were verified after each sample by loading 30 fmol BSA tryptic peptide standard (Michrom Bioresources Inc. Fremont, CA) with 60 fmol α-Casein tryptic digest and running a short 30 min gradient. TOF MS calibration was performed on BSA reference ions before running the next sample in order to adjust for mass drift and verify peak intensity. The instrument method was set to a discovery or IDA mode which consisted of one 250 ms MS1 TOF survey scan from 400-1300 Da followed by twenty 100 ms MS2 candidate ion scans from 100-2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 4+ which exceeded a threshold of 200 cps were selected for MS2, and former precursors were excluded for 10 sec after 1 occurrence.

Mass spectrometry data generated were stored, searched and analyzed using the *ProHits* laboratory information management system (LIMS) platform (Liu et al., 2010). Within *ProHits*, the resulting WIFF files were first converted to an MGF format using WIFF2MGF converter and to an mzML format using *ProteoWizard* (v3.0.4468) and the *AB SCIEX MS Data Converter* (V1.3 beta) and then searched using *Mascot* (v2.3.02) and *Comet* (v2012.02 rev.0). The spectra were searched with the RefSeq database (version 53, May 28th, 2014) acquired from NCBI against a total of 34374 human and adenovirus sequences supplemented with "common contaminants" from the Max Planck Institute (http://141.61.102.106:8080/share.cgi?ssid=0f2gfuB) and the Global

Proteome Machine (GPM; http://www.thegpm.org/crap/index.html). The database parameters were set to search for tryptic cleavages, allowing up to 2 missed cleavage sites per peptide with a mass tolerance of 40 ppm for precursors with charges of 2+ to 4+ and a tolerance of +/- 0.15 amu for fragment ions. Variable modifications were selected for deamidated asparagine and glutamine and oxidized methionine. The results from each search engine were analyzed through the *Trans-Proteomic Pipeline* (TPP v4.6 OCCUPY rev 3) (Deutsch et al., 2010) via the *iProphet* pipeline (Shteynberg et al., 2011). *SAINTexpress* (v3.3) (Teo et al., 2014) was used as a statistical tool to calculate the probability value of each potential protein-protein interaction from background contaminants using default parameters. Unless otherwise specified, controls were compressed to 6 samples and a FDR of 1 % or lower were required for proteins to be classified as significant interaction partners of ZMYND8.

AP-MS and BioID-MS Data Functional Enrichment, Visualization and Deposition

Functional enrichment analysis was performed using *DAVID Bioinformatics Resources* (v6.7) (Huang da et al., 2009). Functional Annotation Clustering (July 2015) was performed and Biological Process (BP FAT) are reported, with enrichment probabilities adjusted for GO terms with Benjamini-Hochberg FDR correction of 0.01 or smaller.

Dot plots of FL-ZMYND8 WT and FL-ZMYND8 N228F mutant interactomes obtained by BioID and AP-MS were generated using custombuilt proteomics data visualization tools (Knight et al., 2015). All MS files used in this study were deposited at MassIVE (http://massive.ucsd.edu) and assigned ID **MSV000079336**.

RNA Interference (ZMYND8, HATs) and ZMYND8 knock-down

The knock-down experiments of ZMYND8 or histone acetyl-transferases (HATs, including PCAF, p300, TADA2A) were performed by RNA interference using Lipofectamine RNAiMAX reagent (Cat.#13778-150). In the case of ZMYND8, HEK293 cells expressing inducible 3xFLAG-ZMYND8 were reverse transfected with 100 pmol Stealth siRNAs targeting ZMYND8 (Cat.#HSS119060; Invitrogen) or Stealth Negative Control for 48 h. Cells were harvested for RNA extraction, western blot or ChIP-qPCR analysis. In the case of HATs, cells were reverse transfected with 20 pmol siRNAs (Dharmacon ON TARGET SMARTpool; L005055-00 for PCAF; L003486-00 for p300, and L017516-00 for TADA2A) or Non-targeting siRNA (D001810-01-20) 72 h, as well as 10 nM

Doxycycline to induce 3xFLAG-ZMYND8 expression. Cells were harvested for western blot and ChIP-qPCR analysis

RNA extraction

Total RNA was extracted and prepared using the RNeasy Mini Kit (Cat.#74106; Qiagen) including an on-column DNase I digestion (Cat.#79254; Qiagen), according to the manufacturer's instructions. The resulting RNA was quantified and quality controlled using a Nanodrop spectrophotometer (model ND1000; Thermo Scientific) and the Agilent RNA 6000 Nano Kit (Cat.#5067-1511, Agilent Technologies).

Chromatin Immunoprecipitation (ChIP) and ChIP-seq

The ChIP DNA was collected using the SimpleChIP™ Enzymatic Chromatin IP Kit (Cell Signaling Technology #9005), according to the manufacturer's protocol. Briefly, HEK293 cells were collected from 15 cm plates and treated with 1 % formaldehyde (cross-linking) for 10 min and 125 mM Glycine for another 5 min (quenching) at room temperature. After washing three times with 1X PBS, cells were lysed in buffer A for 10 min on ice, then washed once in buffer B. Cells were re-suspended in buffer B, then treated with MNase (2000 gel units) at 37 ˚C for 30 min, before the digestion was stopped by adding EDTA (10 µl 0.5 mM EDTA per 100 µl reaction). Chromatin extracts were harvested following complete lysis of nuclei by sonication. 10 µg of chromatin samples were incubated with each antibody (2 µl anti-ZMYND8 (Bethyl Laboratory, A302-089A); 4 µl anti-H3K14ac (Millipore, 07-353); 4 µl anti-H3K27ac (Millipore, 07-595); 2 µl anti-Flag (Sigma, F1804)) for 12 h at 4 °C with rotation. Then 30 µl of ChIP-Grade ProteinG Magnetic Beads (#9006, Cell Signalling Technology) were added to each sample and incubated for 2 h at 4 °C with rotation. The beads were washed with low salt buffer and high salt buffer. Chromatin samples were eluted from antibody/beads in 150 µl of ChIP elution buffer for 30 min at 65 °C. The samples were reverse cross-linked by adding 6 µl 5M NaCl and 2 µl Proteinase K (20 mg/ml) for 2 h at 65 °C and then purified in DNA purification columns.

Library construction and sequencing were performed at the Wellcome Trust Centre for Human Genetics; Samples were quantified using the Quant-iT™ PicoGreen® dsDNA Kits (Invitrogen) and sample integrity was assessed using Agilent's Tapestation system both in accordance to manufacturer specifications. Libraries were constructed using PrepX Complete ILMN 32i DNA

Library Kit (#400076) with an input of 5 ng where available. Ligation of adapters was performed using Illumina Adapters within the Multiplexing Sample Preparation Oliogonucleotide Kit. Each library was PCR enriched with 25 µM each of the following custom primers: Multiplex PCR primer 1.0 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and Index primer 5'-CAAGCAGAAGACGGCATACGAGAT[INDEX]CAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'. Indexes used were 8 bp long as previously described (Lamble et al., 2013). Enrichment and adapter extension of each preparation was obtained using 15 µl of size selected library in a 50 µl PCR reaction using NEBNext High-Fidelity 2X PCR Master Mix. After 18 cycles of amplification (one cycle at 98 ˚C for 30 sec followed by 18 cycles of the round (10 sec at 98 ˚C, 30 sec at 65 ˚C and then 30 sec at 72 ˚C) and then one cycle at 72 ˚C for 5 min. Hold at 4 ˚C), the PCR products were purified with Ampure beads (Agencourt/Beckman) on a Biomek NXp. The concentrations used to generate the multiplex pool were determined by Picogreen and pooling was performed using a Biomek 3000. The final size distribution of the pool was determined using a Tapestation 1DK system (Agilent), quantification for sequencing was determined using Library Quant Primers + KSF (Illumina #KK4923) and an MX3005P instrument (Agilent) before sequencing on an Illumina HiSeq 2500 as 51 single end. Base-calling was performed with bclToFastq (Illumina base-caller) version 1.8.4A.

Chromatin Immunoprecipitation & Quantitative PCR (ChIP-qPCR)

Eluted chromatin supernatants were purified using the QIAquick PCR Purification Kit (QIAGEN, #28106), and analysed by quantitative PCR. All reactions were run on a LightCycler 480 Instrument II (Roche) using LightCycler® 480 SYBR Green I Master (# 04887352001; Roche) according to the manufacturer's protocol. Samples were prepared on 384-well plates with a final reaction volume of 10 µl, containing 5 µl LightCycler® 480 SYBR Green I Master, 0.4 µM forward and reverse primers (**Supplemental Table S9**) and 2 µl of diluted samples. Data were normalized as % of input and are presented as mean ± SD from three biological replicates. Statistical significance was analysed using two-tailed Student's *t*-test or one-way analysis of variance followed by Dunnett's multiple comparison test. *P* values are presented such that: P<0.001 ****; P<0.005 ***; P<0.01 **; P<0.05, *; Not-significant, ns.

ChIP-seq Data Analysis

FASTQ files for ZMYND8, H3K14ac, H3K27ac were mapped to the human reference genome (hg19) by *BWA-MEM* (v.0.7.13) (Li and Durbin, 2009) using default parameters allowing up to two mismatches. SRA files were downloaded using the *SRA toolkit* (v.2.6.2) from NCBI's Gene Expression Omnibus for H3K14ac (GSM521881) and H3K27ac (GSM521887) in IMR90 cells; RACK7 and H3K27ac in ZR-75-30 cell lines (GSE71327); SRA files were processed with *bowtie2* (v.2.2.9) (Langmead and Salzberg, 2012) using default parameters. The genome ChIP-seq profiles for ZMYND8, H3K14ac, H3K27ac were generated using *HOMER* (v.4.8) (Heinz et al., 2010). The ChIP-seq profiles were normalized to GC content, 10^9 total tag numbers and peaks were called at FDR ≤ 10^{-3}. Peaks were annotated by REFSEQ genes. Bigwig coverage profiles were generated with the UCSD toolbox. Enrichment profiles of DNA-interacting proteins at functionally important regions were visualized using *ngs.plot* (v.2.6.1) (Shen et al., 2014). Enhancer regions in HEK293 and computation of enrichment profiles in these loci were defined employing annotations from the FANTOM5 consortium (Andersson et al., 2014). Heatmaps of read coverage in ZMYND8-bound TSS and FANTOM5 enhancers were constructed with *deepTools* (v.2.2.4) (Ramirez et al., 2014). Canonical correlation analysis in ZMYND8-bound enhancers was performed with CCA (v.1.2) (Gonzalez et al., 2008).

ChIP-seq tags enrichment in promoter and enhancer regions relative to local background was computed with the *localEnrichmentBed* toolkit from https://github.com/dariober/bioinformatics-cafe/tree/master/localEnrichmentBed. The coordinates of the respective regulatory genomic elements were downloaded from the FANTOM5 data repository (http://fantom.gsc.riken.jp/5/data/). The *P*-value computed by the algorithm was filtered using a threshold of 0.01 to determine significantly enriched genomic loci. ChIP-sequencing data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO accession number **GSE81696**.

Supplemental References

Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. Acta Crystallogr D *52*, 30-42.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res *18*, 6069-6074.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D *66*, 12-21.

Couzens, A.L., Knight, J.D., Kean, M.J., Teo, G., Weiss, A., Dunham, W.H., Lin, Z.Y., Bagshaw, R.D., Sicheri, F., Pawson, T.*, et al.* (2013). Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. Sci Signal *6*, rs15.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D *62*, 1002-1011.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B.*, et al.* (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics *10*, 1150-1159.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D *66*, 486-501.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D *69*, 1204-1214.

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Muller, S., Pawson, T.*, et al.* (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell *149*, 214-231.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I.*, et al.* (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067-1073.

Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C., and Merryman, C. (2010). Chemical synthesis of the mouse mitochondrial genome. Nat Methods *7*, 901-903.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Gonzalez, I., Dejean, S., Martin, P.G.P., and Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. Journal of Statistical Software *23*, 1-14.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kabsch, W. (2010). XDS. Acta Crystallogr D *66*, 125-132.

Knight, J.D., Liu, G., Zhang, J.P., Pasculescu, A., Choi, H., and Gingras, A.C. (2015). A web-tool for visualizing quantitative protein-protein interaction data. Proteomics *15*, 1432-1436.

Lambert, J.P., Tucholska, M., Go, C., Knight, J.D., and Gingras, A.C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. J Proteomics *118*, 81-94.

Lambert, J.P., Tucholska, M., Pawson, T., and Gingras, A.C. (2014). Incorporating DNA shearing in standard affinity purification allows simultaneous identification of both soluble and chromatin-bound interaction partners. J Proteomics *100*, 55-59.

Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B., and Piazza, P. (2013). Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol *13*, 104.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Liu, G., Zhang, J., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z.Y., Breitkreutz, B.J., Ding, Y., Colwill, K., Pasculescu, A.*, et al.* (2010). ProHits: integrated software for mass spectrometry-based interaction proteomics. Nat Biotechnol *28*, 1015-1017.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D *67*, 355-367.

Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J.W. (2009). A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell *36*, 153-163.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D *62*, 439-450.

Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A.*, et al.* (2011). Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. Mol Biosyst *7*, 2899-2908.

Picaud, S., and Filippakopoulos, P. (2015). SPOTing Acetyl-Lysine Dependent Interactions. Microarrays *4*, 370-388.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M.*, et al.* (2012a). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol *19*, 1155-1160.

Rothbart, S.B., Krajewski, K., Strahl, B.D., and Fuchs, S.M. (2012b). Peptide microarrays to interrogate the "histone code". Methods Enzymol *512*, 107-135.

Savitsky, P., Bray, J., Cooper, C.D., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., and Gileadi, O. (2010). High-throughput production of human proteins for crystallization: the SGC experience. J Struct Biol *172*, 3-13.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D *66*, 479-485.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics *10*, M111 007690.

Smith, C.M. (2005). Quantification of acetylation at proximal lysine residues using isotopic labeling and tandem mass spectrometry. Methods *36*, 395-403.

Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., Kimura, H., and Kurumizaka, H. (2010). Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. Proc Natl Acad Sci U S A *107*, 10454-10459.

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J Proteomics *100*, 37-43.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol Biol *364*, 215-230.

Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.N. (1989). Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. Anal Biochem *179*, 131-137.