Structural basis for decoding active histone methylation marks by Polycomb Repressive Complex 2

One sentence summary: *"Nucleosome binding by PRC2 threads H3K27 into its active site via an interaction network set in register by unmodified H3K36."*

Ksenia Finogenova(1), Jacques Bonnet (1), Simon Poepsel(2), Ingmar B. Schäfer(3), Katja Finkl(1), Katharina Schmid(1), Claudia Litz(1), Mike Strauss(4,5), Christian Benda(3) and Jürg Müller(1)

1) Max Planck Institute of Biochemistry, Laboratory of Chromatin Biology, Martinsried, Germany
2) Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany Cologne Excellence Cluster for Cellular Stress Responses in Ageing-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
3) Max Planck Institute of Biochemistry, Department of Structural Cell Biology, Martinsried, Germany
4) Max Planck Institute of Biochemistry, cryoEM Facility, Martinsried, Germany
5) Current address: McGill University, Department of Anatomy and Cell Biology, Montreal, Canada

Correspondence should be addressed to J.M. (muellerj@biochem.mpg.de)

Short title: *"Trithorax-Polycomb antagonism"*

Keywords: Polycomb Repressive Complex 2 / cryo-EM/ nucleosome recognition/ H3K36me2/3

Repression of genes by Polycomb requires that PRC2 modifies their chromatin by trimethylating lysine 27 on histone H3 (H3K27me3). At transcriptionally active genes, di- and trimethylated H3K36 inhibit PRC2. Here, the cryo-EM structure of PRC2 on nucleosomes reveals how binding of its catalytic subunit EZH2 to nucleosomal DNA orients the H3 N-terminus via an extended network of interactions to place H3K27 into the active site. Unmodified H3K36 occupies a critical position in the EZH2-DNA interface. Mutation of H3K36 inhibits H3K27 methylation by PRC2. Accordingly, *Drosophila* H3K36 mutants show altered H3K27me3 profiles and deregulated expression of HOX genes. The relay of interactions between EZH2, nucleosomal DNA and the H3 N-terminus therefore creates the geometry for allosteric inhibition of PRC2 by methylated H3K36 in transcriptionally active chromatin.

Many post-translational modifications on histone proteins are essential for processes in the underlying chromatin. Typically, histone modifications themselves do not alter chromatin structure directly but function by binding effector proteins which alter chromatin or by interfering with such interactions. The histone methyltransferase Polycomb Repressive Complex 2 (PRC2) and its regulation by accessory proteins and histone modifications represent a prime example for understanding these interaction mechanisms. PRC2 trimethylates lysine 27 in histone H3 (H3K27me3), a modification that is essential for the transcriptional repression of developmental regulator genes that control cell fate decisions in metazoans (*1*, *2*). H3K27me3 marks chromatin for interaction with PRC1, an effector which compacts chromatin (*3*). H3K27me3 is also recognized by PRC2 itself, and this interaction allosterically activates the PRC2 enzyme complex to facilitate deposition of H3K27me3 across extended domains of chromatin (*4*-*6*). Genetic studies and subsequent biochemical work established that PRC2 is in addition subject to negative regulation. In particular, the H3K4me3, H3K36me2 and H3K36me3 marks present on nucleosomes in transcriptionally active chromatin directly inhibit H3K27 methylation by PRC2 (*7*-*11*). Importantly, while stimulation of PRC2 activity by H3K27me3 acts *in trans*, inhibition of PRC2 by H3K4me3, H3K36me2 and H3K36me3

requires that these modifications are present *in cis*, that is, on the same H3 molecule containing the K27 substrate lysine (*8*, *9*). While recent structural studies have uncovered the allosteric activation mechanism for PRC2 (*6*, *12*), the mechanism of PRC2 inhibition by active chromatin marks has remained enigmatic. Adding to the complexity, in nucleosome binding assays, histone methylation marks do not seem to have a major effect on the nucleosome-binding affinity of PRC2 (*8*, *13*, *14*). Instead, PRC2-DNA contacts make the largest contribution to nucleosome binding (*13*, *15*). Building on this earlier work, we investigated the molecular mechanism of how PRC2 binds to nucleosomes and this revealed how the complex gauges the presence of active chromatin marks.

We assembled recombinant full-length human PRC2 in complex with its accessory factor PHF1 (i.e. PHF1-PRC2) (*15*) (**Fig. 1A**) on a heterodimeric dinucleosome (di-Nuc), which consisted of a 'substrate' nucleosome with unmodified histone H3 and an 'allosteric' nucleosome containing H3 with a trimethyllysine analog (*16*) at K27, separated by a 35 base pair (bp) DNA linker (*17*). Single particle cryo-electron microscopy analysis yielded a reconstruction of the PHF1-PRC2:di-Nuc assembly with an overall resolution of 5.2 Å (**fig. S1-S3**). The map showed clear density for the catalytic lobe of PRC2 with similar chromatin interactions and binding geometry as previously described for the catalytic lobe of AEBP2-PRC2 (*17*) where PRC2 contacts the two nucleosomes via interactions with the DNA gyres (**Fig. 1B**). Specifically, the substrate nucleosome is bound by the $EZH2_{CXC}$ domain residues K563, Q565, K569 and Q570 (**Fig. 1C, fig. S3E**, cf. (*17*)), while the allosteric nucleosome is contacted by EED and by the SBD and SANT1 domain of EZH2 (**Fig 1D**, cf. (*17*)). We could not detect density for the 'bottom lobe' of PRC2 (*18*, *19*) or for the N-terminal winged-helix and tudor domains of PHF1 that bind DNA and H3K36me3, respectively (*15*, *20-23*).

Using focused refinement on the interface of EZH2 and the substrate nucleosome (**fig. S2-S3**), we then obtained an improved map at an apparent overall resolution of 4.4 Å which revealed well-defined density for the H3 N-terminus (**Fig. 1E, fig. S3F**). The visible sidechain density combined with the crystallographic models of the PRC2 catalytic lobe and of the

mononucleosome enabled us to build a pseudo-atomic model of the histone H3 N-terminus spanning residues R26 to K37 (**Fig. 1E**). This model revealed that EZH2 recognizes the H3 N-terminus via an extended network of contacts besides the previously described ionic interactions near the active site where H3 R26 interacts with EZH2 Q648/D652, and H3 K27 with the aromatic cage above the EZH2 catalytic center (*12*) (**Fig 1E**). Specifically, our structure suggests two hydrophobic hotspots, the first one involving H3 A29/P30 and EZH2 residues F667, A697, V699, I708 and F724 and the second one involving H3 V35 and F542, F557 and P558 of EZH2 (**Fig. 1E**). H3 G33/G34 is likely not recognized by PRC2 but might act as a flexible hinge between the two hydrophobic interaction sites (**Fig. 1E**). H3K36, finally, is directly juxtaposed to the $EZH2_{CXC}$-DNA interaction surface and appears to be involved in the EHZ2-DNA interface. The clear side chain density of H3K36 and the quality of the reconstruction in the DNA suggests that the epsilon-amino group of H3K36 contacts the phosphate backbone of the nucleosomal DNA (**Fig. 1E, fig. S3F**). In summary, an extensive network of interactions between EZH2, the nucleosomal DNA and the H3 N-terminus orients the latter into an extended conformation to thread H3K27 into the EZH2 active site.

We next analysed how the PRC2 surfaces contacting the substrate and the allosteric nucleosome contribute to the formation of productive PRC2-chromatin interactions. For these experiments, we used $PHF1_C$-PRC2 (*15*), which contains the minimal 5-kDa PRC2-interaction domain of PHF1 but lacks the tudor and the DNA-binding winged-helix domains (**Fig. 1A**, (*15*, *20*, *23*, *24*)) and therefore only retains the DNA-binding surfaces of the 4-subunit PRC2 core complex. For simplicity we shall refer to this $PHF1_C$-PRC2 complex as PRC2. We generated three mutant versions of PRC2. In $PRC2^{CXC>A}$ (K563A Q565A K569A Q570A), the $EZH2_{CXC}$ interface is mutated (**Fig. 1C**), in $PRC2^{EED>A}$ (K77A K83A K385A K398A K400A K408A), the EED interface contacting the allosteric nucleosome (**Fig. 1D**), is mutated, and $PRC2^{CXC>A/EED>A}$ carries the combination of these mutations. We first used electromobility shift assays (EMSA) to measure the binding affinity of wild-type and mutant PRC2 complexes on mononucleosomes. These mononucleosomes were assembled on a 215 bp long DNA fragment containing the 147-bp 601 nucleosome-positioning sequence (*25*) in the center and 34 bp of

linker DNA on both sides. Wild-type PRC2 bound this mononucleosome with an apparent K_d in the midnanomolar range (**Fig. 2A**, cf. (*15*)). The binding affinities of $PRC2^{CXC>A}$ or $PRC2^{EED>A}$ were two- to three-fold lower than that of wild-type PRC2 and that of $PRC2^{CXC>A/EED>A}$ was about five-fold lower compared to the wild-type (Fig. 2A, compare lanes 11-30 with 1-10, and **fig. S4A**). The residual nucleosome binding shown by $PRC2^{CXC>A/EED>A}$ (**Fig. 2A**, lanes 21-30) could in part be due to incomplete disruption of the mutated interfaces but in part is likely also due to the previously identified nucleosome-binding activity of the PRC2 bottom lobe (*18*, *26*). Therefore, the overall chromatin-binding affinity of the PRC2 core complex appears to be generated by the interactions of at least three distinct surfaces. Among those, binding of the $EZH2_{CXC}$ domain to the nucleosomal DNA only contributes modestly to the total binding affinity.

We then analysed the histone methyltransferase (HMTase) activity of $PRC2^{CXC>A}$. On the same mononucleosomes used above, $PRC2^{CXC>A}$ showed almost no detectable HMTase activity compared to wild-type PRC2 (**Fig. 2B**, compare lanes 5-7 with 2-4). On dinucleosomes, where EED binding to the juxtaposed nucleosome might be expected to facilitate the interaction of EZH2 with the H3 N-termini on the substrate nucleosome, $PRC2^{CXC>A}$ still shows significantly weaker HMTase activity than wild-type PRC2 (**Fig. 2B**, compare lanes 12-14 with 9-11). The contact of the $EZH2_{CXC}$ domain with the nucleosomal DNA is therefore critical for engaging the H3 N-terminus in a manner that allows H3K27 methylation.

The architecture of the $EZH2_{CXC}$-DNA interface around H3K36 (**Fig. 1E**) suggested that a bulkier side chain, such as that of a tri- or di-methylated lysine or an arginine may not be accommodated in this interface. In EMSAs, the affinity of PRC2 for binding to mononucleosomes containing a trimethyllysine analog at H3K36 (H3Kc36me3) was indistinguishable from that for binding to unmodified mononucleosomes (**Fig. 2C**). However, as previously reported (*8*, *9*), H3K27 methylation by PRC2 was strongly inhibited on H3Kc36me3-containing mononucleosomes (**Fig. 2D**, compare lanes 5-7 with 2-4). Methylation of H3K27 was also inhibited on mononucleosomes where H3K36 had been mutated to arginine or alanine ($H3^{K36R}$ and $H3^{K36A}$ mononucleosomes, respectively) (**Fig. 2D**,

compare lanes 8-13 with 2-4). PRC2 inhibition on $H3^{K36R}$ and $H3^{K36A}$ mononucleosomes was however less severe than on H3Kc36me3 mononucleosomes (**Fig. 2D**, compare lanes 8-13 with 5-7). Together with the structural data, this highlights that the unmodified side chain of H3K36 is critical for the productive positioning of H3K27 in the catalytic site of PRC2. Neither the bulkier side chains of trimethyllysine or arginine nor the short apolar side chain of alanine provide the correct fit at the position of H3K36. Importantly, on isolated histone $H3_{18-42}$ peptides, H3K27 methylation by PRC2 was not inhibited by the presence of H3K36me3 in the peptide (**Fig. 2E**). The allosteric inhibition of PRC2 by H3K36me3 therefore only occurs in the context of the geometric constraints of the nucleosome.

We next investigated how H3K27 trimethylation by PRC2 is affected in chromatin of *Drosophila* $H3^{K36A}$ or $H3^{K36R}$ mutants in which we had replaced wild-type histone H3 with $H3^{K36A}$ or $H3^{K36R}$. H3K27me3 is primarily found on canonical histone H3 (*1*) (*2*)and we used previously described strategies (*1*) to replace it and generate $H3^{K36A}$ and $H3^{K36R}$ mutant animals (see Methods). For some of the experiments, an independently generated $H3^{K36R}$ strain from Matera and colleagues (*2*) was used to corroborate results. Western blot analyses showed that bulk levels of H3K36me2 and H3K36me3 were reduced in $H3^{K36A}$ and $H3^{K36R}$ mutant embryos and larvae (**fig. S5A-B**). The residual H3K36me2 and H3K36me3 signals in the mutant embryos represented maternally-deposited wild-type H3 molecules that are modified and persist in their chromatin (**fig. S5A-B**). Accordingly, in $H3^{K36A}$ and $H3^{K36R}$ mutant embryos H3K36me2 levels were reduced across the genome but not eliminated (**Fig. 3A**, left panels, **Fig. 3B**, H3K36me2 tracks, **Table S2**). As shown above, on $H3^{K36A}$ and $H3^{K36R}$ nucleosomes, PRC2 methylates H3K27 less efficiently than on wild-type nucleosomes (**Fig. 2D**). Conversely, $H3^{K36A}$ and $H3^{K36R}$ nucleosomes are not as effective as H3Kc36me3 nucleosomes at inhibiting PRC2 (**Fig. 2D**). Accordingly, animals with $H3^{K36A}$ and $H3^{K36R}$ chromatin showed complex changes in their genome-wide H3K27me3 profiles, even though H3K27me3 bulk levels appeared largely unchanged compared to wildtype (**fig. S5A**). Firstly, H3K27me3 levels were reduced at the majority of genes that are located in chromatin domains with high-levels of H3K27me3 in wild-type animals (**Fig. 3A**, right panels). Among the genes with the most

pronounced H3K27me3 reduction were the HOX genes in the Bithorax- and Antennapedia-complexes (**Fig. 3A-B**, **fig. S5C**). In *Drosophila* with $H3^{K36A}$ or $H3^{K36R}$ chromatin, PRC2 therefore appears to be unable to generate high levels of H3K27me3 at Polycomb target genes. Secondly, in both mutants, H3K27me3 levels were increased at most of the transcriptionally active genes that are decorated with H3K36me2 in wild-type animals, even though these H3K27me3 levels remain at least one order of magnitude below those in high-level H3K27me3 domains (**Fig. 3A, B**). This suggests that, like in the *in vitro* assays, $H3^{K36A}$ and $H3^{K36R}$ nucleosomes are less effective than H3K36me2/3 at preventing the sporadic H3K27me3 deposition by PRC2 that occurs in these regions.

The reduction of H3K27me3 at HOX genes prompted us to analyse whether their expression may be altered in $H3^{K36A}$ and $H3^{K36R}$ mutants. In embryos, both mutants showed stochastic misexpression of the HOX gene Abd-B, similar to but less severe than $H3^{K27R}$ mutant embryos or embryos lacking the PRC2 subunit Esc that were used as controls (**Fig. 3C, fig. S5D**, arrowheads). Moreover, in eye-antennal imaginal discs, $H3^{K36A}$ and $H3^{K36R}$ mutant cells showed misexpression of the HOX gene Antp in large clusters of cells, similar to those seen in $H3^{K27R}$ mutants (**Fig. 3D, fig. S5E**). The reduced H3K27me3 levels at *Abd-B* and *Antp* in $H3^{K36A}$ and $H3^{K36R}$ mutants (**Fig. 3B, fig. S5C**) therefore appears to be below the critical threshold needed to fully support Polycomb repression at these genes.

H3K36 di-methylation by the HMTase Ash1 antagonizes Polycomb repression by preventing PRC2 from depositing H3K27me3 at HOX gene chromatin in cells where they are expressed (*7*, *27*, *28*). Can $H3^{K36A}$ or $H3^{K36R}$ nucleosomes replace H3K36me2/me3 for this function? In embryos, $H3^{K36A}$ and $H3^{K36R}$ mutants showed stochastic loss of Abd-B expression within its normal expression domain, similar to but less severe than *ash1* mutants (**Fig. 3C**, **fig. S5D**, empty arrowheads). However, $H3^{K36A}$ and $H3^{K36R}$ mutant larvae did not show the loss of HOX gene expression characteristic of *ash1* mutants. From these findings it appears that $H3^{K36A}$ or $H3^{K36R}$ chromatin antagonize faulty H3K27me3 deposition and Polycomb repression at actively transcribed HOX genes, albeit somewhat less efficiently than H3K36me2/3

chromatin. In summary, in *Drosophila* with $H3^{K36A}$ or $H3^{K36R}$ chromatin HOX gene expression is deregulated and the predominant phenotype is stochastic loss of Polycomb repression.

Understanding how PRC2 binds chromatin and how it is regulated is essential for understanding how the complex marks genes for Polycomb repression to maintain cell fate decisions. Unlike many other histone-modifying enzymes (e.g. (*29*, *30*)), PRC2 does not recognize the nucleosome by docking on its acidic patch (*31*) to engage with the histone substrate. Instead, it is the interaction of the $EZH2_{CXC}$ domain with the DNA gyres on the nucleosome that sets the register for binding the H3 N-terminus and orienting it in a strechted-out conformation such that the H3K27 side chain reaches into the active site. Unmodified H3K36 is accommodated in a key position in this $EZH2_{CXC}$-DNA interface. Not only methylation of H3K36 but also its substitution with arginine or alanine inhibits H3K27 methylation by PRC2, both *in vitro* and *in vivo*. In conclusion, it is the exquisite geometry formed by the relay of interactions between EZH2, nucleosomal DNA and the H3 N-terminus that enables H3K36me2/3 to allosterically prevent PRC2 from trimethylating H3K27 at transcriptionally active genes.

REFERENCES AND NOTES

1. A. R. Pengelly, Ö. Copur, H. Jäckle, A. Herzig, J. Müller, A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*. **339**, 698–699 (2013).

2. D. J. McKay *et al.*, Interrogating the function of metazoan histones using engineered gene clusters. *Dev. Cell*. **32**, 373–386 (2015).

3. N. J. Francis, R. E. Kingston, C. L. Woodcock, Chromatin compaction by a polycomb group protein complex. *Science*. **306**, 1574–1577 (2004).

4. K. H. Hansen *et al.*, A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).

5. R. Margueron *et al.*, Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. **461**, 762–767 (2009).

6. L. Jiao, X. Liu, Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*. **350**, aac4383–aac4383 (2015).

7. T. Klymenko, J. Müller, The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).

8. F. W. Schmitges *et al.*, Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell.* **42**, 330–341 (2011).

9. W. Yuan *et al.*, H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989 (2011).

10. L. J. Gaydos, A. Rechtsteiner, T. A. Egelhofer, C. R. Carroll, S. Strome, Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell Rep*. **2**, 1169–1177 (2012).

11. G. Streubel *et al.*, The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell*. **70**, 371–379.e5 (2018).

12. N. Justin *et al.*, Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun*. **7**, 11316 (2016).

13. X. Wang *et al.*, Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA JA -, doi:10.1038/nsmb.3487.

14. N. Guidotti, C. C. Lechner, A. L. Bachmann, B. Fierz, A Modular Ligation Strategy for Asymmetric Bivalent Nucleosomes Trimethylated at K36 and K27. *Chembiochem*. **20**, 1124–1128 (2019).

15. J. Choi *et al.*, DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat. Struct. Mol. Biol.* **298**, 1039 (2017).

16. M. D. Simon *et al.*, The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell.* **128**, 1003–1012 (2007).

17. S. Poepsel, V. Kasinath, E. Nogales, Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* **25**, 154–162 (2018).

18. S. Chen, L. Jiao, M. Shubbar, X. Yang, X. Liu, Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol. Cell.* **69**, 840–852.e5 (2018).

19. V. Kasinath *et al.*, Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science*. **359**, 940–944 (2018).

20. H. Li *et al.*, Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* (2017), doi:10.1038/nature23881.

21. C. Ballaré *et al.*, Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* **19**, 1257–1265 (2012).

22. L. Cai *et al.*, An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell*. **49**, 571–582 (2013).

23. C. A. Musselman *et al.*, Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun.* **4**, 2969–9 (2013).

24. S. Chen, L. Jiao, X. Liu, X. Yang, X. Liu, A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Mol. Cell*. **77**, 1265–1278.e7 (2020).

25. P. T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

26. M. Nekrasov, B. Wild, J. Müller, Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Rep.* **6**, 348–353 (2005).

27. B. Papp, J. Müller, Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20**, 2041–2054 (2006).

28. S. Schmähling *et al.*, Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC. *Development*. **145**, dev163808 (2018).

29. R. K. McGinty, R. C. Henrici, S. Tan, Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*. **514**, 591–596 (2014).

30. E. J. Worden, N. A. Hoffmann, C. W. Hicks, C. Wolberger, Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell*. **176**, 1490–1501.e12 (2019).

31. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. **389**, 251–260 (1997).

32. J. Bonnet *et al.*, Quantification of Proteins and Histone Marks in Drosophila Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation. *Dev. Cell*. **51**, 632–644.e6 (2019).

We thank Eva Nogales for generous advice, sharing of expertise and for hosting K.F. for grid preparation. We thank Tom Cech for stimulating discussions. We thank J.R. Prabu for excellent computing support, S.Uebel, E.Weyher, R.Kim and A.Yeroslaviz of the MPIB core facilities for excellent technical support and S.Schkoelziger and S.Schmähling for help with some of the experiments. This work was supported by the Deutsche Forschungsgemeinschaft (SFB1064) and the MPG. ChIP-seq data have been deposited in GEO (accession number: GSE148254).

SUPPLEMENTARY MATERIALS

Materials and Methods

Fig S1 – S5

Tables S1 – S2

Movie S1

References

FIGURE LEGENDS

FIGURE 1

Interaction of the PRC2 catalytic lobe with nucleosomal DNA orients the H3 N-terminus for H3K27 binding to the active site.

(**A**) Domain organization in the five subunits of PHF1-PRC2. Dashed boxes indicate protein portions visible in the PHF1-PRC2:di-Nuc cryo-EM reconstruction and fitted in the structural model.

(**B**) Cryo-EM reconstruction of PHF1-PRC2:di-Nuc in two orientations with fitted crystal structures of human PRC2 catalytic lobe (PDB: 5HYN, (*12*)) and nucleosomes (1AOI, (*31*)) in a di-Nuc model with 35 bp linker DNA (see also **figs. S1-S3, Table S1, Move S1**). Density is colored as in (**A**) to show PRC2 subunits, DNA (blue) and octamers of substrate (pink) and allosteric (yellow) nucleosomes. Boxes indicate regions shown in (**C**), (**D**) and (**E**), respectively.

(**C**) Interaction of $EZH2_{CXC}$ residues with the DNA gyres of the substrate nucleosome; residues mutated in $PRC2^{CXC>A}$ are indicated. For the H3 N-terminus (pink), only the peptide backbone is shown in this view (see **E**).

(**D**) Interface formed by EED and the EZH2 SBD domain with DNA gyres on the allosteric nucleosome; residues mutated in $PRC2^{EED>A}$ are indicated. Asterisk indicates the approximate location of a residue, which is not built in the model.

(**E**) The H3 N-terminus (pink), shown as a pseudoatomic model fitted into the 4.4 Å density map, is recognized by EZH2 through an extensive interaction network (see text). Note the well-defined side-chain density of H3K36 (see also **fig S3F**).

FIGURE 2

Critical role of unmodified H3K36 in the EZH2$_{CXC}$-DNA interaction interface for H3K27 methylation on nucleosomes.

(**A**) Left, binding reactions with indicated concentrations of PRC2 (lanes 1-10), PRC2$^{CXC>A}$ (lanes 11-20) or PRC2$^{CXC>A/EED>A}$ (lanes 21-30) and 45 nM 6-carboxyfluorescein-labeled mononucleosomes, analyzed by EMSA on 1.2% agarose gels; for analysis of PRC2$^{EED>A}$ binding, see **fig S4A**. Right: quantitative analysis of EMSA data by densitometry of 6-carboxyfluorescein signals from independent experiments (n=3); error bars, SEM.

(**B**) Western Blot (WB) analysis of H3K27me1 and H3K27me3 formation in HMTase reactions with indicated concentrations of PRC2 and PRC2$^{CXC>A}$ on 446 nM mononucleosomes (lanes 1-7) or 223 nM dinucleosomes (lanes 8-14). Coomassie stained gel of reactions is shown in **fig. S4B**. H4 WB signal served as control for western blot processing.

(**C**) EMSA analysis and quantification as in (**A**), using PRC2 and mononucleosomes that were unmodified (lanes 1-10) or contained a trimethyllysine analog at H3K36 (H3Kc36me3, lanes 11-20). Coomassie stained gel of reactions is shown in **fig. S4C**.

(**D**) Western Blot (WB) analysis of HMTase reactions with PRC2 as in (**B**) on unmodified (lanes 1-4), H3Kc36me3 (lanes 5-7), H3^{K36R} (lanes 8-10) or H3^{K36A} (lanes 11-13) mononucleosomes. Bottom: quantification of H3K27me3 and H3K27me1 chemiluminescence signals, respectively, by densiometry analysis from three independent experiments. In each experiment, the methylation signal in lane 4 was defined as 100% and used to quantify the corresponding H3K27 methylation signals in the other lanes on the same membrane, circles show individual data points and error bars SEM.

(**E**) HMTase reactions monitoring H3K27me1 formation by PRC2 on H3$_{18-42}$ peptides that were unmodified (top) or contained K36me3 (bottom). Left: Deconvoluted ESI-MS spectra from data shown in **fig. S4D**. On both substrates, areas of the four colored peaks of H3K27me1-modified and unmodified substrate peptides were used for quantification of H3K27me1 formation. Right: Symbols represent percentages of peptides carrying H3K27me1 in three independent experiments, error bars show SEM; Welch's t-test showed no significant (ns)

difference between H3K27 monomethylation on the two peptide substrates.

FIGURE 3

***H3^{K36A}* and *H3^{K36R}* mutants show altered H3K27me3 profiles and deregulation of HOX genes.**

(**A**) Left, scatter plots showing fold change of H3K36me2 levels in *H3^{K36A}* (top) or *H3^{K36R}* mutant (bottom) embryos compared to *wildtype* (*wt*) embryos in relation to H3K36me2 read coverage across gene bodies in *wt* embryos. Embryos were 21-24 hr old. Right, similar representation for H3K27me3 level changes in relation to H3K27me3 read coverage. Green dots represent 10800 gene bodies overlapping with genomic intervals showing H3K36me2 enrichment, blue dots represent 1035 gene bodies overlapping with genomic intervals defined as high-level H3K27me3 domains (*32*), and grey dots represent 5400 gene bodies showing no enrichment for either methylation mark. HOX genes *Abd-B*, *abd-A*, *Ubx* and *Antp* in high-level H3K27me3 domains and six genes (*Ns1* (1), *mRpS11* (2), *Keap1* (3), *kuk* (4), *GckIII* (5), *cal1* (6)) in H3K36me2-enriched regions are indicated. See Supplementary Materials for information on genotypes.

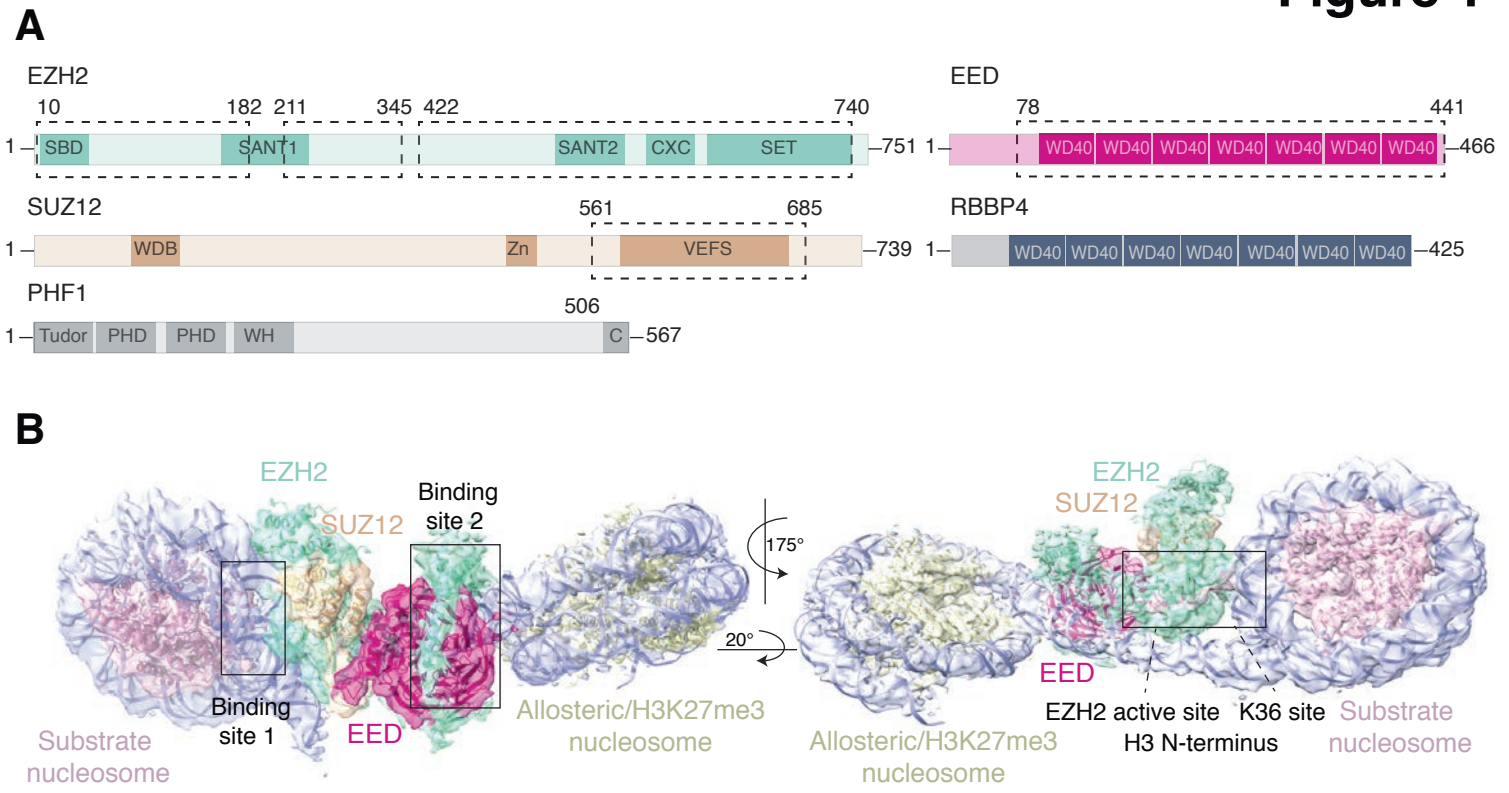
(**B**) Left: H3K27me3 and H3K36me2 ChIP-seq profiles at *Ubx*, *abd-A* and *Abd-B* in *wt* embryos (blue), and in *H3^{K36A}* (purple) and *H3^{K36R}* (yellow) mutant embryos with the *wt* profile superimposed (see **Table S2** and Supplementary Materials for information about normalization). Right: Zoomed-in version of the region marked by the dashed box in the left panel, harbouring genes 1 to 6 marked in (**A**). For the H3K27me3 profile, the y-scale was adjusted to visualize the increase of H3K27me3 levels in *H3^{K36A}* and *H3^{K36R}* mutants in regions showing H3K36me2 enrichment in *wt* embryos, i.e. genes 1-6 marked in (**A**).

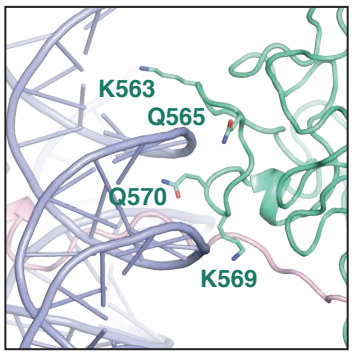
(**C**) Ventral views of stage 16 *wt* and *H3^{K36A}*, *H3^{K36R}* or *H3^{K27R}* homozygous mutant embryos, stained with Abd-B antibody. The vertical bar marks the anterior boundary of Abd-B expression in parasegment (ps) 10 in *wt* embryos. Note the stochastic misexpression of Abd-B in *H3^{K36A}* and *H3^{K36R}* (arrowheads) and widespread misexpression in *H3^{K27R}* mutant embryos.

In $H3^{K36A}$ and $H3^{K36R}$ mutants, stochastic loss of Abd-B expression is visible in ps10 cells (empty arrowheads); see **Fig. S5D** for further information.

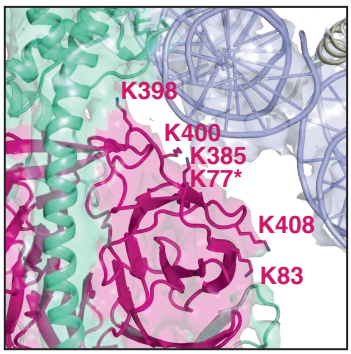
(**D**) Eye-antennal imaginal discs from third instar larvae with clones of $H3^{K36A}$, $H3^{K36R}$ or $H3^{K27R}$ homozygous mutant cells, or, as control, clones of *wt* cells, stained with antibody against Antp (red) and Hoechst (DNA). In all cases, clone cells are marked by the absence of GFP (green) and were induced 96 hrs before analysis. Antp is not expressed in the eye-antennal disc of *wt* animals but is misexpressed in $H3^{K36A}$, $H3^{K36R}$ or $H3^{K27R}$ mutant cell clones in the antenna primordium (arrowheads); misexpression only occurs in a subset of cells and not in all clones (empty arrowheads). See also **fig. S5E**.

Figure 1

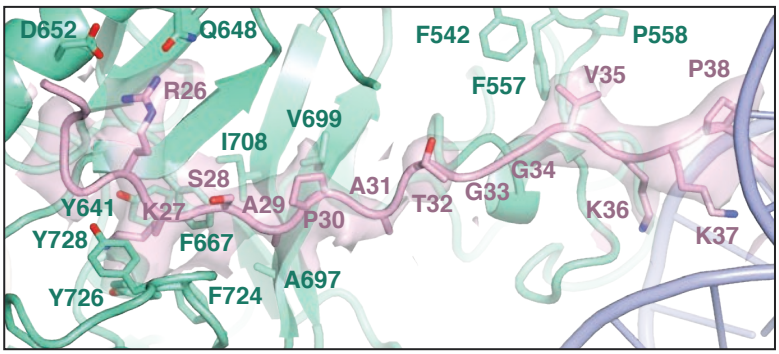




Binding site 1/ EZH2$_{CXC}$



Binding site 2/EED



EZH2 active site H3 N-terminus K36 site

Figure 2

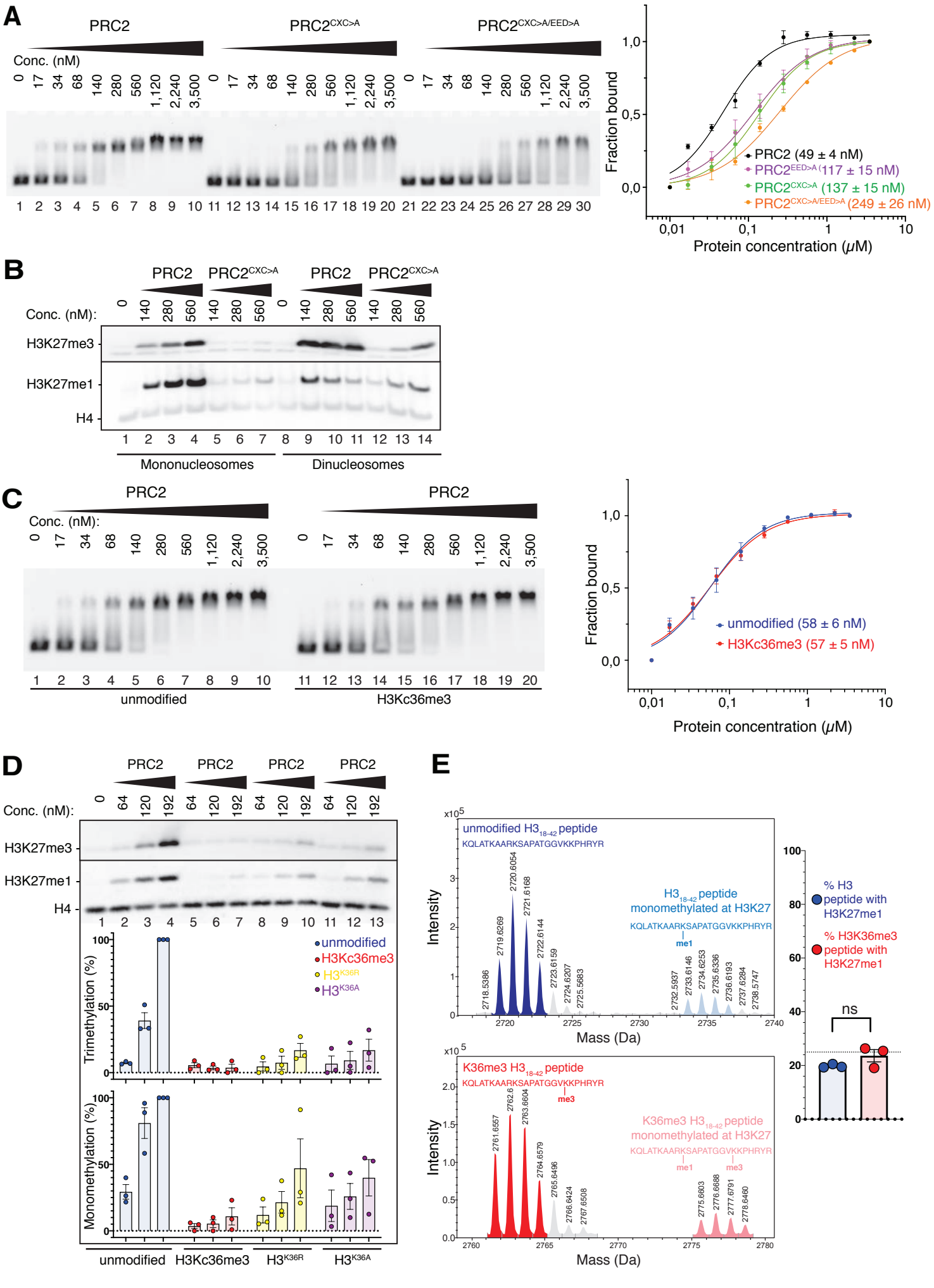
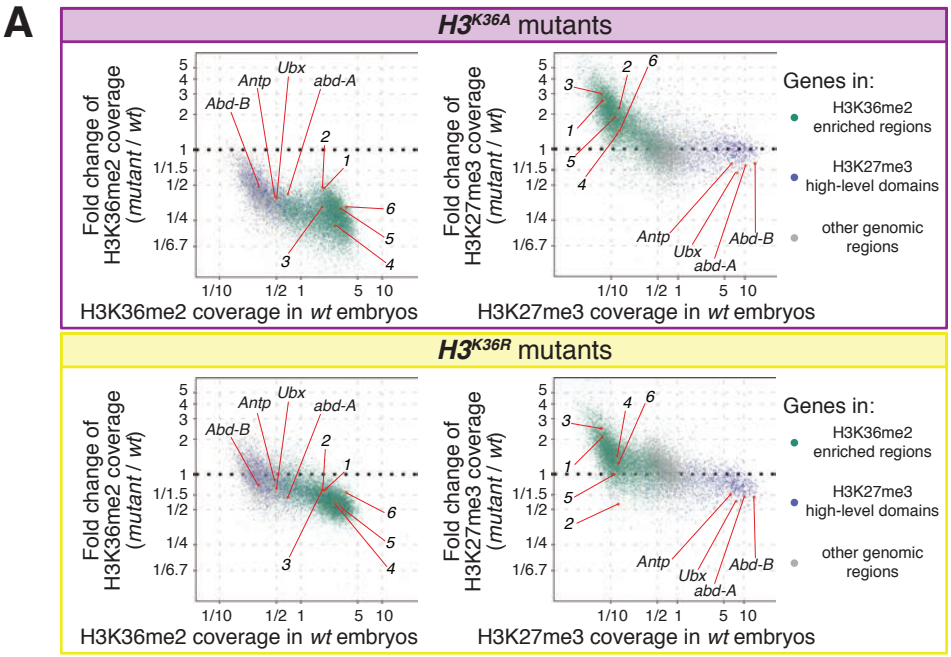
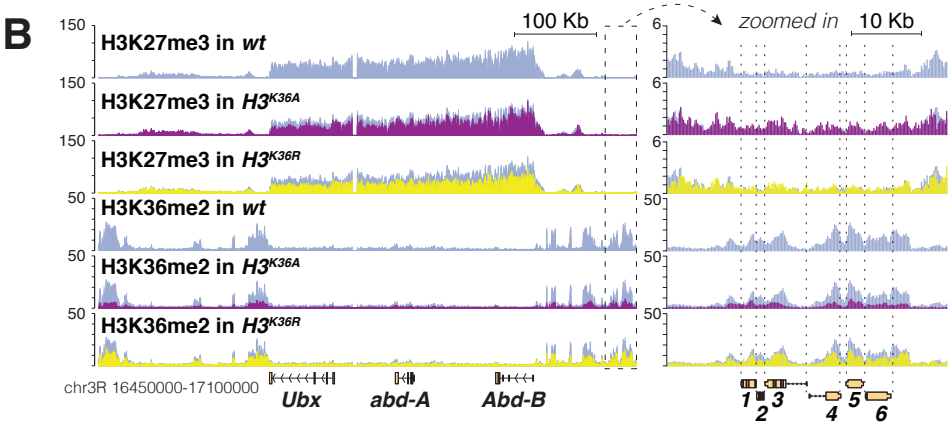
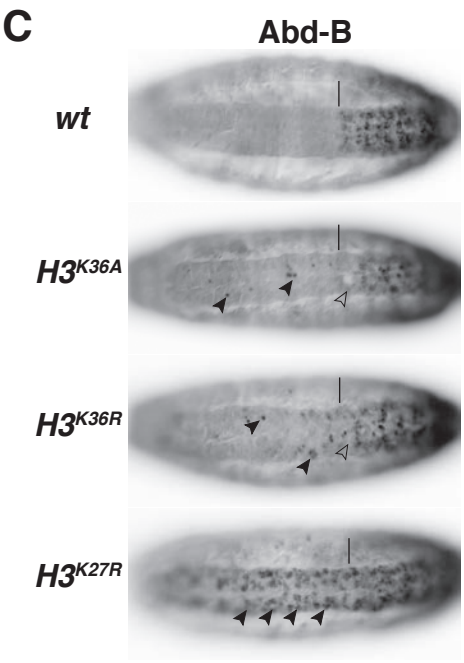
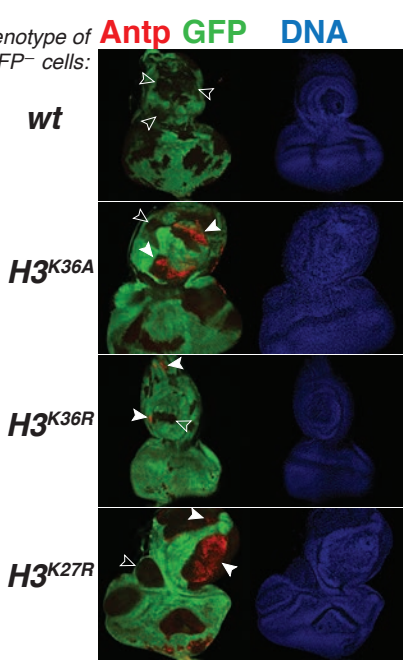


Figure 3









Supplementary Materials for

Structural basis for decoding active histone methylation marks by Polycomb Repressive Complex 2

Ksenia Finogenova, Jacques Bonnet, Simon Poepsel, Ingmar Schäfer, Katja Finkl, Katharina Schmid, Claudia Litz, Mike Strauss, Christian Benda and Jürg Müller*

Correspondence to: muellerj@biochem.mpg.de

This PDF file includes:

Materials and Methods

Protein expression and purification

Human PHF1-PRC2 wild-type (wt) complex was expressed and purified as previously described (*15*). In brief, an optimized ratio of the baculoviruses for the different PHF1-PRC2 subunits was used to infect HiFive cells (Invitrogen). Cell were lysed using a glass Dounce homogenizer and the complex was purified using affinity chromatography (Ni-NTA and Strep-tag), followed by simultaneous TEV mediated protease tag cleavage and Lambda Phosphatase treatment (obtained from the MPI of Biochemistry Protein Core facility) and a final size exclusion chromatography (SEC) step in a buffer containing 25 mM Hepes, pH 7.8, 150 mM NaCl, 10% glycerol, 2 mM DTT.

$PRC2^{CXC>A}$, $PRC2^{EED>A}$ and $PRC2^{CXC>A/EED>A}$ mutants were generated by PCR with primers containing the desired mutations, subsequent ligation and transformation. Expression and purification were performed as above.

Xenopus laevis (*X.l.*) and *Drosophila melanogaster* (*D.m.*) histones were expressed and purified from inclusion bodies as described in (*33*). To mimic the inhibitory mark H3K36me3 or the allosteric activating mark H3K27me3, the cysteine side chain of a mutated *D.m.* histone $H3^{C110A\ K36C}$ or *X.l.* histone $H3^{C110A\ K27C}$ was alkylated with (2-bromoethyl) trimethylammonium bromide (Sigma Aldrich) as described previously (*16*). Nucleosomes containing these modifications are abbreviated with e.g. H3Kc36me3.

For histone octamers, equimolar amounts of histones H2A, H2B, H4 and H3 (wt, $H3^{K36A}$, $H3^{K36R}$, H3Kc27me3 or H3Kc36me3) were mixed and assembled into octamers in high salt buffer containing 10 mM Tris-HCL pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol. Subsequent SEC was performed to separate octamers from H3/H4 tetramers or H2A/H2B dimers (*33*).

Reconstitution of nucleosomes:

For *X.l.* and *D.m* mononucleosomes used in biochemical assays, 6-carboxyfluorescein (6-FAM)-labeled 215 bp 601 DNA (*25*) was PCR amplified from the p601 plasmid, purified on a MonoQ column (GE Healthcare), precipitated with ethanol and dissolved in the same high salt buffer used for octamers. Optimized ratios of octamer to DNA (usually ranging between 0.8-1.3 : 1) were mixed and nucleosomes were reconstituted by gradient and stepwise dialysis against low salt buffers to a final buffer containing 25 mM Hepes, pH 7.8, 60 mM NaCl, 2 mM DTT.

X.l. asymmetrical dinucleosomes for cryo-EM studies containing one unmodified substrate nucleosome and one H3K27me3-modified (allosteric) nucleosome connected with a 35 bp linker DNA were reconstituted using the protocol described in (*17*). In brief, substrate nucleosomes and allosteric nucleosomes were separately assembled on the respective *DraIII* digested nucleosomal DNA. The latter was generated by PCR with primers introducing the desired linker and *DraIII* recognition sites and purified as described above. The assembled nucleosomes were purified on a preparative native gel system (Biorad 491 prep cell). After ligation using T4 ligase (Thermo Fisher Scientific) the resulting dinucleosomes were purified from aberrant or non-ligated mononucleosomes by a second preparative native gel system (Biorad 491 prep cell). In contrast to (*17*), the dinucleosome DNA used in this study contained an additional 30 bp overhang on the substrate nucleosome, thus resulting in the following DNA sequence:

5′–601 binding (allosteric nucleosome) – agcgatctCACCCCGTGatgctcgatactgtcata – 601 binding (substrate nucleosome) – atgcatgcatatcattcgatctgagctcca –3′ (after DraIII digestion, assembly of substrate/allosteric nucleosome and ligation to dinucleosomes).

X.l. symmetrical unmodified dinucleosomes used for the HMTase assays with the $PRC2^{CXC}$ mutants were obtained by reconstituting octamers with a 377 bp DNA containing two 601

sequences connected by a 35 bp linker DNA. A vector containing the 377 bp sequence was ordered from Invitrogen GeneArt and was used for PCR resulting in:

5′–atatctcgggcttatgtgatggac – 601 binding (substrate nucleosome 1) – agcgatctcaacgagtgatgctcgatactgtcata – 601 binding (substrate nucleosome 2) – gtattgaacagcgactcgggatat–3′.

The PCR products were purified as described above. Optimized ratios of octamer : DNA (usually ranging between 1.8-2.3 : 1) were mixed and nucleosomes were reconstituted by gradient and stepwise dialysis against low salt buffers to a final buffer containing 25 mM Hepes, pH 7.8, 60 mM NaCl, 2 mM DTT.

Cryo-EM Data acquisition

Complexes of PHF1-PRC2 and asymmetrically modified 35 bp dinucleosomes were assembled and grids were prepared essentially as described previously (*17*). Cryo-EM data were collected on an FEI Titan Krios microscope operated at 300 kV and equipped with a post-column GIF and a K2 Summit direct detector (Gatan) operated in counting mode. A total of 3467 movies were collected at a nominal magnification of 81,000x (1.746 Å/pixel) at the specimen level using a total exposure of 52 $e^-/Å^2$ distributed over 60 frames and a target defocus range from 1.5–3 µm. Data acquisition was carried out with SerialEM.

Cryo-EM Data processing

Movies were aligned and corrected for beam-induced motion as well as dose compensation using MotionCor2 (*34*). CTF estimation of the summed micrographs was performed with Gctf (*35*) and particles were picked in Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/ K. Zhang, MRC LMB, Cambridge, UK) using templates created from the AEBP2-PRC2-dinucleosome cryo-EM structure (EMD-7306, (*17*). All subsequent image processing steps were performed in Relion 3.0 (*36*) as shown in Fig. S2. A total of 1,028,229 candidate particles were subjected to two rounds of initial 3D classification against a reference map (AEBP2-PRC2-dinucleosome low-pass filtered to 60 Å) and the Bayesian fudge factor (T value) set to 8. 330,482 remaining particles were subjected to two more rounds of 3D classification, this time using the best 3D model from the previous run as reference. Finally, the two best 3D models were 3D refined and further classified into 10 classes without translational and spatial sampling, using a T value of 4. From this run, the best 3D classes with the highest nominal overall resolution and rotational and translational accuracy were subjected to iterative rounds of 3D refinement, this time applying a soft mask for solvent flattening, per particle CTF refinement and Bayesian polishing. The final map after postprocessing had an overall nominal resolution of 5.2 Å, as determined from the gold-standard FSC criterion of 0.143 (*37*) (Fig. S1D). The density (Overall PHF1-PRC2:di-Nuc) with fitted models is shown in Fig.1A and in Fig. S1E using UCSF ChimeraX (*38*). Local resolution estimation was performed in Relion 3.0 and is shown in Fig. S1B. The spherical angular distribution of all particles in the final model is shown in Fig. S1C.

To further improve the resolution and map details of the region around the H3 N-terminus, particle subtraction and focused 3D refinement was applied (*39-41*). Using a mask generated with UCSF Chimera (*42*) and Relion 3.0 the signal of the allosteric nucleosome as well as parts of PRC2 (EED and $EZH2_{allo}$) was subtracted from all particle images. These signal subtracted particles were then subjected to focused 3D refinement using a soft mask around the substrate nucleosome and $EZH2_{sub}$. This yielded a 4.4 Å map ($EZH2_{sub}$-Nuc_{sub}) (Fig. S3B). Local resolution estimation is shown in Fig.S3A. For model building and depiction, the final density was further sharpened using the Multisharpen function in Coot (*43*) (e.g. in Figs. 1E, S4D, E and F).

Cryo-EM data fitting, modeling and refinement

Available crystal structures were fitted into the final maps using rigid-body fitting in UCSF Chimera and all manual remodeling and building was performed in Coot. For PRC2, the crystal

structure of the catalytic lobe of human PRC2 (PDB: 5HYN (*12*)) was used. Since the SBD helix and the SANT1 helix bundle of the crystal structure was not accommodated well by the corresponding EM density, this region was fitted separately. A model of a dinucleosome with linker DNA (Supplementary dataset 1 in (*17*), including crystal structures of nucleosomes, PDB 3LZ1, also PDB 1AOI) was fitted.

The above described overall model was then used as a starting model for fitting and building $EZH2_{sub}$-Nuc_{sub} into the focused map. Where possible, missing parts in the model were built de-novo, i.e. the H3 N-terminal tail (residues 30-37) between the catalytic site of PRC2 and the substrate histone. Available information from crystal structures was used as a guide (PRC2 with H3 peptide bound: PDB: 5HYN (*12*), and high resolution crystal structures of nucleosomes (PDB 1AOI) (*31*). Parts of $EZH2_{sub}$-Nuc_{sub} model were then fitted using the morph fit routine in Coot or manually (*44*). Secondary structure restraints for real-space refinement were generated automatically with phenix.secondary_structure_restraints (*45*) and manually curated. Hydrogens were added and the model was real-space refined with Phenix (*46*), using reference structures (PDB 6T9L and PDB 1AOI for nucleosome and one copy of the human PRC2 crystal structure generated from PDB 5HYN), applying strict secondary structure and Ramachandran restraints. Structures were visualized with UCSF ChimeraX (*38*) and PyMOL2 (https://pymol.org/2/).

Electrophoretic mobility shift assay (EMSA)

EMSAs on a 1.2% agarose gel in 0.4x TBE Buffer with 45 nM 6-FAM - labeled mononucleosomes (unmodified wt *X.l.* for bandshifts with the $PRC2^{CXC}$ mutants, unmodified wt *D.m*. and *D.m* H3Kc36me3 trimethyllysine analog containing nucleosomes) and increasing PRC2 concentrations (concentrations indicated in the figures above the gels) were performed in triplicates as described in (*15*). A Typhoon FLA 9500 scanner and the Fiji software was used for densitometric analysis of the 6-FAM signal (*47*). Background correction and calculation of the fractions of bound nucleosomes was performed with R using tidyverse (https://www.r-project.org/). In detail: two parts were boxed out in each lane: 1. unbound nucleosomes ('unbound' box) and 2. shifted nucleosomes ('bound', everything above 'unbound'). The boxed-out signals were integrated and background corrected by subtracting the respective control ('bound' background of lane 1 for 'bound' boxes and 'unbound' background of lane 10 for 'unbound' boxes). To calculate the fraction of bound vs. unbound nucleosomes, the value for 'bound' nucleosome in each lane was divided by the total signal (sum of bound and unbound) of the same lane. Hill function fitting and illustration of the plot were subsequently performed with Prism 8 (GraphPad).

Histonemethyltransferase (HMTase) assay

For all HMTase assays, 446 nM of mononucleosomes or 223 nM of dinucleosomes were incubated with indicated amounts of the different PRC2 complexes, in a reaction buffer containing 20 mM HEPES pH 7.8, 50 mM NaCl, 2.5 mM $MgCl_2$, 5% glycerol, 0.25 mM EDTA, 0.5 mM DTT and 80 µM S-adenosylmethionine (SAM). Reactions were allowed to proceed for 90 min at RT before quenching by the addition of 1x (final concentration) SDS loading buffer and heat inactivation at 95 °C for 5 min. Proteins were separated by electrophoresis on a 16% (w/v) SDS gel, transferred to a nitrocellulose membrane and probed with antibodies against H3K27me3 (Millipore, 07-449), H3K27me1 (Millipore, 07-448) and H4 (Abcam, ab10158). For quantification, HMTase reactions and the corresponding western blots on *D.m.* unmodified, H3Kc36me3, $H3^{K36A/R}$ mononucleosomes were performed in triplicates and subjected to densitometric analysis (Chemiluminescence signal, ImageQuant LAS 4000). The integrated densitometric signal (band) in each lane was background corrected against the control lane (lane 1, no PRC2 in the reaction) and normalized with respect to the lane containing the highest amount (i.e. 100%) of PRC2 on unmodified nucleosomes (lane 4). The relative amounts of trimethylation/monomethylation for all other lanes were calculated with respect to lane 4. Graphical representations were made with Prism 8 (GraphPad).

Mass Spectrometry (MS)

500 nM of PRC2 were incubated with 2 µM of either unmodified or $H3_{18-42}$ peptide containing the K36me3 modification in HMTase reaction buffer (described above) and methyltransferase activity was allowed to proceed over night at RT. Reactions were then quenched with 1% trifluoroacetic acid (TFA). Home-made stage tips with poly(styrenedivinylbenzene) copolymer (SDB-XC) were used to remove PRC2 from the reactions (*48*). First, stage tips were washed with methanol, followed by a second wash with buffer B (0.1% (v/v) formic acid, 80% (v/v) acetonitrile). The SDB-XC material was then equilibrated with buffer A (0.1% (v/v) formic acid) and 40 µl of sample was applied and washed several times. Finally, samples were eluted using buffer B and introduced into the Bruker maXis II ETD mass spectrometer by flow injection of 20 µl sample using an Agilent HPLC at a flow rate of 250 µl/min and 0.05% TFA in 70% acetonitril:H2O as solvent for ESI-MS time-of-flight analysis.

Peptides were ionized at a capillary voltage of 4500 V and an end plate offset of 500 V.

Full scan MS spectra (200-1600 m/z) were acquired at a spectra rate of 1 Hz and a collision cell energy of 15 eV.

Raw data files were processed using Bruker Compass DataAnalysis. The m/z spectra were deconvoluted (maximum entropy method) with an instrument resolving power of 10,000 and the resulting neutral spectra peaks were integrated. For quantification, the experiment was performed in triplicates. The sum of the monomethylation peak areas was divided by the sum of the first 4 peaks of the input peptide together with the sum of the monomethylation peak areas. Illustration of the quantification was subsequently performed with Prism 8 (GraphPad). A Welch's t-test was calculated to show the nonsignificant difference between the activity of PRC2 on unmodified or H3K36me3 peptide.

Construction of histone transgenes to generate $H3^{K36A}$ and $H3^{K36R}$ strains

Site directed mutagenesis on *pENTR221-HisGU.WT, pENTRL4R1-HisGU.WT* and *pENTRR2L3-HisGU.WT* (*49*) was used to mutate histone H3K36 to alanine or arginine. The final constructs *pfC31-attB-3xHisGU.H3K36A* and *pfC31-attB-3xHisGU.H3K36R* were generated by Gateway LR recombination of above vectors and integrated at attP sites VK33 (BDSC 9750) and 86Fb (BDSC 130437). The full genotypes of animals used in the study is described below.

Drosophila strains and genotypes

The following strains were used in this study:

Oregon-R

w; Df(2L)His^{C} FRT40A/ CyO ubi:GFP; 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb)/ 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb) (*1*)

w; Df(2L)His^{C} FRT40A/ CyO twi:Gal4 UAS:GFP; 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)/ 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb) (generated in this study)

w; Df(2L)His^{C} FRT40A/ CyO twi:Gal4 UAS:GFP; 3xHisGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)/ 3xH isGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb) (generated in this study)

w; Df(2L)His^{C} FRT40A/ CyO ubi:GFP; 3xHisGU^{H3K27R}(68E) 3xHisGU^{H3K27R} (86Fb)/ 3xHisGU^{H3K27R} (68E) 3xHisGU^{H3K27R} (86Fb) (*1*)

w hs-flp; M(2)25A ubi-GFP FRT40A/CyO

w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 12xHisGU^{wt}/ 12xHisGU^{wt} (*2*)

w; Df(2L)His^{C} FRT40A/ CyO ubi-GFP; 12xHisGU^{H3K36R}/TM6B (*2*)

yw; esc^{6} b pr / CyO, P[esc^{+}]

In(2LR) Gla / CyO, esc^{2}

w; ash1^{22} FRT2A/ TM3 twi::EGFP

w hsp70-flp; ovo^{D} FRT2A/ TM2

The following genotypes were used for the experiments shown in:

Fig. 3A, B

wt: *Oregon-R*
$H3^{K36A}$: *w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)/ 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)*
$H3^{K36R}$: *w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)/ 3xH isGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)*

Fig. 3C

wt: *w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb)/ 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb)*
$H3^{K36A}$: *w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)/ 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)*
$H3^{K36R}$: *w;Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)/ 3xH isGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)*
$H3^{K27R}$: *w; Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K27R}(68E) 3xHisGU^{H3K27R} (86Fb)/ 3xHisGU^{H3K27R} (68E) 3xHisGU^{H3K27R} (86Fb)*

Fig. 3D

wt: *w hs-flp; M(2)25A ubi-GFP FRT40/ Df(2L)His^{C} FRT40A; 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb)/ 3xHisGU^{wt}(68E) 3xHisGU^{wt}(86Fb)*
$H3^{K36A}$: *w hs-flp; M(2)25A ubi-GFP FRT40/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)/ 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)*
$H3^{K36R}$: *w hs-flp; M(2)25A ubi-GFP FRT40/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)/ 3xH isGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)*
$H3^{K27R}$: *w hs-flp; M(2)25A ubi-GFP FRT40/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K27R}(68E) 3xHisGU^{H3K27R} (86Fb)/ 3xHisGU^{H3K27R} (68E) 3xHisGU^{H3K27R} (86Fb)*

fig. S5A

wt: *Oregon-R*
$H3^{K36A}$: *Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)/ 3xHisGU^{H3K36A}(VK33) 3xHisGU^{H3K36A}(86Fb)*
$H3^{K36R}$: *Df(2L)His^{C} FRT40A/ Df(2L)His^{C} FRT40A; 3xHisGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)/ 3xH isGU^{H3K36R}(VK33) 3xHisGU^{H3K36R}(86Fb)*

fig. S5B

wt: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{wt}(VK33)/ 12xHisGU^{wt}(VK33)*
$H3^{K36R}$: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B*

fig. S5C

as in Fig. 3A, B

fig. S5D

wt: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{wt}(VK33)/ 12xHisGU^{wt}(VK33)*
$H3^{K36R}$: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B*
esc^{-}: *esc^{6} b pr / CyO, esc^{2}* (*esc$^{mat-zyg-}$* obtained as progeny from *esc^{6} b pr/CyO, esc^{2}* parents)
ash1^{-}: *ash1^{22} FRT2A/ ash1^{22} FRT2A (ash1$^{mat-zyg-}$* obtained as progeny from *w hsp70-flp; ash1^{22} FRT2A/ ovo^{D} FRT2A* mothers with germ line clones and
w; ash1^{22} FRT2A/ TM3 twi::EGFP fathers)

fig. S5E

wt: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{wt}(VK33)/ 12xHisGU^{wt}(VK33)*
$H3^{K36R}$: *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B*

Analysis of lethality phase of *$H3^{K36A}$* and *$H3^{K36R}$* mutants

Animals that are homozygous for a deletion of the histone gene cluster *HisC (i.e. Df(2L)HisC* homozygotes) arrest development at the blastoderm stage after exhaustion of the pool of maternally-deposited histones but transgene cassettes providing 12 copies of the wild-type *HisGU* (*12xHisGU^{WT}*) rescue *Df(2L)HisC* homozygotes develop into viable adults (*2*, *49*). Matera and colleagues previously reported that *Df(2L)HisC* homozygotes that carry a single transgene array with 12 *HisGU^{H3K36R}* copies (i.e. *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B*) arrest development as early pupae (*2*). Using these strains developed by Matera and colleagues (*2*), Schwartz and colleagues claimed that a fraction of *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B* animals would develop into adults with normal morphology (*50*). Using the strains from Matera and colleagues (*2*), we found that of 600 first instar larvae with the genotype *Df(2L) HisC FRT40/ Df(2L) HisC FRT40; 12xHisGU^{H3K36R}(VK33)/TM6B*, 86% arrested development during the larval stages, while 14% developed to the pupal stage and then died, with no adults eclosing.

Our observations are therefore consistent with those from Matera and colleagues (*2*) and disagree with the claims made by Schwartz and colleagues (*50*).
With the histone replacement strategy that was developed by Herzig and colleagues (*49*) and used here to generate $H3^{K36A}$ and $H3^{K36R}$ mutant animals, we found that 12 copies of $HisGU^{H3K36R}$ or of $HisGU^{H3K36A}$ rescue *Df(2L)HisC* homozygotes only up to the first larval instar, with most $Df(2L)His^{C}$ *FRT40A/* $Df(2L)His^{C}$ *FRT40A; 3xHisGU*^{H3K36R}*(VK33) 3xHisGU*^{H3K36R}*(86Fb)/ 3xH isGU*^{H3K36R}*(VK33) 3xHisGU*^{H3K36R}*(86Fb)* animals, or $Df(2L)His^{C}$ *FRT40A/* $Df(2L)His^{C}$ *FRT40A; 3xHisGU*^{H3K36A}*(VK33) 3xHisGU*^{H3K36A}*(86Fb)/ 3xHisGU*^{H3K36A}*(VK33) 3xHisGU*^{H3K36A}*(86Fb)* animals arresting development shortly after completion of embryogenesis. The reason for this earlier lethality is not known.

Immunohistochemistry and immunofluorescence stainings
Embryos of the appropriate genotypes listed above were identified by the lack of GFP marked balancer chromosomes, fixed and stained with Abd-B antibody, following standard protocols. Imaginal discs from third instar larvae were stained with Antp and Cy3-labeled secondary antibodies following standard protocols. For clonal analysis (**Fig. 3D**), clones were induced 96 hrs before analyses by heat-shocked induced expression of Flp recombinase in the genotypes listed above.

ChIP-seq analysis in *Drosophila* embryos
Embryo collection, chromatin preparation and ChIP: 21-24 hr old *wt*, $H3^{K36A}$ and $H3^{K36R}$ (see above the detailed genotypes) embryos were dechorionated, quick-frozen in liquid N2 and stored at -80˚C. 5 µL of thawed embryos were homogenized in 5 mL of fixing solution (60 mM KCl, 15 mM NaCl, 4 mM $MgCl_2$, 15 mM Hepes pH 7.6, 0.5% Triton X-100, 0.5 mM DTT, protease inhibitors, 0.9% Formaldehyde) at r.t.. The homogenate was filtered through a strainer (Greiner Bio-One, EASYstrainer^{TM} 100 µm, #542 000) and incubated for 10 min with frequent gentle shaking. Cross-linking was stopped by the addition of 450 µL of 2.5 M Glycine. Fixed nuclei were washed with 1 mL of buffer A1 (60 mM KCl, 15 mM NaCl, 4 mM $MgCl_2$, 15 mM Hepes pH 7.6, 0.5% Triton X-100, 0.5 mM DTT, protease inhibitors), washed with 1 mL of pre-lysis buffer (140 mM NaCl, 15 mM Hepes pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% Na Deoxycholate, protease inhibitors), resuspended in 1 mL of lysis buffer (140 mM NaCl, 15 mM Hepes pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% Na Deoxycholate, protease inhibitors, 0.1% SDS, 0.5% N-laurylsarcosine), incubated at least 10 min at 4˚C with shaking, and transferred into milliTUBES 1 mL AFA Fiber (100) (Covaris, #520130) for sonication. Sonication was performed in a Covaris S220 AFA instrument using the following setup: 140W (peak incident power) / 5% (duty cycle) / 200 (cycle per burst) / 15 min. Insoluble material was removed by centrifugation in an Eppendorf centrifuge at 14000 rpm (10 min at 4°C). Input chromatin was quantified by measuring DNA concentration after decrosslinking using Qubit (Thermo Scientific) and 250 ng of chromatin were used for each ChIP experiment. 250 ng of an independently prepared batch of *D. pseudoobscura* chromatin were spiked-in in each ChIP experiment for subsequent normalization of the ChIP-seq datasets. The rest of the ChIP protocol was performed as described **(*32*)**. For each condition, the ChIP experiment was performed in duplicates from two biologically independent chromatins.
Library preparation and sequencing: Library preparation for sequencing was performed with TruSeq kits from Illumina. Illumina systems (NextSeq 500) were used for paired-end DNA sequencing. All reads were aligned using STAR **(*51*)** to the *D. melanogaster* dm6 genome assembly **(*52*)** and to the *D. pseudoobscura* dp3 genome assembly (Nov. 2004, FlyBase Release 1.03). Only sequences that mapped uniquely to the genome with a maximum of two mismatches were considered for further analyses.
Identification of H3K36me2 and H3K27me3 enriched regions: The Bioconductor STAN-package **(*53*)** was used to define the location of H3K36me2-enriched regions. The seven chromosome arms (X, 2L, 2R, 3L, 3R, 4 and Y) defined in the dm6 genome assembly were segmented in 200 bp bins. STAN annotated each of these bins into 1 of 3 'genomic states' based on the number of H3K36me2 ChIP-seq reads and the number of input reads overlapping with

each bin, in 21-24 hr wild-type embryos. These 3 'genomic states' corresponded to: 'H3K36me2 enriched' regions; 'low or no H3K36me2' regions and 'no input' regions. The Poisson Lognormal distribution was selected and fitting of hidden Markov models was performed with a maximum number of 100 iterations. Stretches of consecutive bins annotated as 'H3K36me2 enriched' regions were sometimes separated by a few bins showing another type of annotation (i.e. 'no input'). To define a relevant set of H3K36me2 enriched regions, we considered that if stretches of consecutive bins annotated as 'H3K36me2 enriched' regions are not separated by more than 7 Kb, they can be fused. High-level H3K27me3 domains previously defined using the same Bioconductor STAN-package in Bonnet et al **(*32*)** were used in this study.

Normalization and visualisation of H3K27me3 and H3K36me2 ChIP-Seq datasets: The proportion of *D. pseudoobscura* reads as compared to *D. melanogaster* reads in input and in samples was used to normalize the H3K36me2 and H3K27me3 ChIP-seq datasets from $H3^{K36A}$ and $H3^{K36R}$ mutants to the wild-type H3K36me2 and H3K27me3 ChIP-seq datasets respectively (see Table S2). Chip-seq tracks shown in Fig. 3B and fig. S5B show the average of the two biological replicates that were performed for each condition. Y-axes of ChIP-seq tracks correspond to normalized numbers of mapped reads per million reads per 200 bp bin.

Calculation of read coverage: In wild-type and $H3^{K36A\ and\ R}$ mutant conditions, H3K36me2 and H3K27me3 ChIP-seq read coverages across gene bodies were computed on genomic intervals starting 750 bp upstream transcription start sites and ending 750 bp downstream transcription termination sites. Read coverage is defined as the normalized number of mapped reads per million reads from a ChIP-seq dataset divided by the number of mapped reads per million reads from the corresponding input dataset across a genomic region. Among the 17660 *D. melanogaster* Refseq genes, approximately 10800 are overlapping with H3K36me2 enriched regions and approximately 1030 genes are overlapping with high level H3K27me3 domains. Approximately 400 genes with null or extreme coverage values are not represented in Fig3A.

<u>*Drosophila* nuclear and cell extracts for western blot analysis</u>

For embryonic total nuclear extracts, nuclei from 21-24 hr old *wt*, $H3^{K36A}$ or $H3^{K36A}$ mutant embryos were purified and quantified as described (*32*). Pellets of nuclei were resuspended in appropriate volumes of SDS sample buffer proportional to the number of nuclei in each pellet. Extracts were then sonicated in a Bioruptor instrument (Diagenode) (8 cycles (30 sec ON / 30 sec OFF), high power mode), incubated at 75˚C for 5 min and insoluble material was removed by centrifugation at 14000 rpm for 1 mn at r.t..

Total cell extracts from imaginal disc tissues were prepared by resuspending hand-dissected disc tissues in SDS sample buffer. Extracts were then sonicated, incubated at 75˚C for 5 min and insoluble material was removed by centrifugation.

<u>Antibodies</u>

For ChIP analysis:

Rabbit monoclonal anti-H3K27me3	Cell Signaling Technology	#9733
Rabbit polyclonal anti-H3K36me2	Abcam	#9049

For western blot analysis on embryonic and larval extracts:

Rabbit monoclonal anti-H3K27me3	Cell Signaling Technology	#9733
Rabbit polyclonal anti- H3K27me3	Millipore	#07-449
Rabbit polyclonal anti-H3K27me1	Millipore	#07-448

Rabbit monoclonal anti-H3K36me3	Cell Signaling Technology	#4909
Rabbit monoclonal anti-H3K36me2	Cell Signaling Technology	#2901
Rabbit polyclonal anti-H2B	(against full-length recombinant D.m. H2B)	
Rabbit polyclonal anti-H4	Abcam	#10158
Rabbit polyclonal anti-Caf1	**(*54*)**	

For immunohistochemistry and immunofluorescence analysis:

Mouse monoclonal anti-Abd-B	DSHB (1A2E9)
Mouse monoclonal anti-Antp	DSHB (8C11)

Fig. S1.

Initial Cryo-EM analysis of the PHF1-PRC2:di-Nuc complex (related to Fig. 1).
(**A**) Representative micrograph of the cryo-EM dataset (left) and reference-free 2D classes from particles picked without templates (right). Circles indicate particles, which were picked with templates and directly subjected to 3D analysis (see **fig. S2**).
(**B**) Local resolution estimation of the 5.2 Å overall PHF1-PRC2:di-Nuc map. The substrate nucleosome and the adjacent part of EZH2 are well resolved (colors red to yellow).
(**C**) Spherical angular distribution of particles included in the final reconstruction of PHF1-PRC2:di-Nuc.
(**D**) Output from the 3DFSC Processing Server (https://3dfsc.salk.edu/ (*55*)) showing the Fourier Shell Correlation (FSC) as a function of spatial frequency, generated from masked independent half maps of PRC2:diNuc: global FSC (red), directional FSC (blue histogram) and deviation from mean (spread, green dotted line). The nominal overall resolution of 5.24 Å was estimated according to the gold standard FSC cutoff of 0.143 (grey dotted line) (*37*). Sphericity is an indication for anisotropy and amounts to 0.806 in this data. The minor directional anisotropy of the data can be explained by the slightly preferred orientation and missing views as seen in (**C**).
(**E**) Top: Refined and postprocessed cryo-EM density map of overall PHF1-PRC2:di-Nuc colored according to the subunit organization. Bottom: pseudoatomic model of fitted crystal structure of the human PRC2 catalytic lobe (PDB: 5HYN) and a di-Nuc model with 35 bp linker DNA (*17*), including PDB 1AOI.

Figure S1

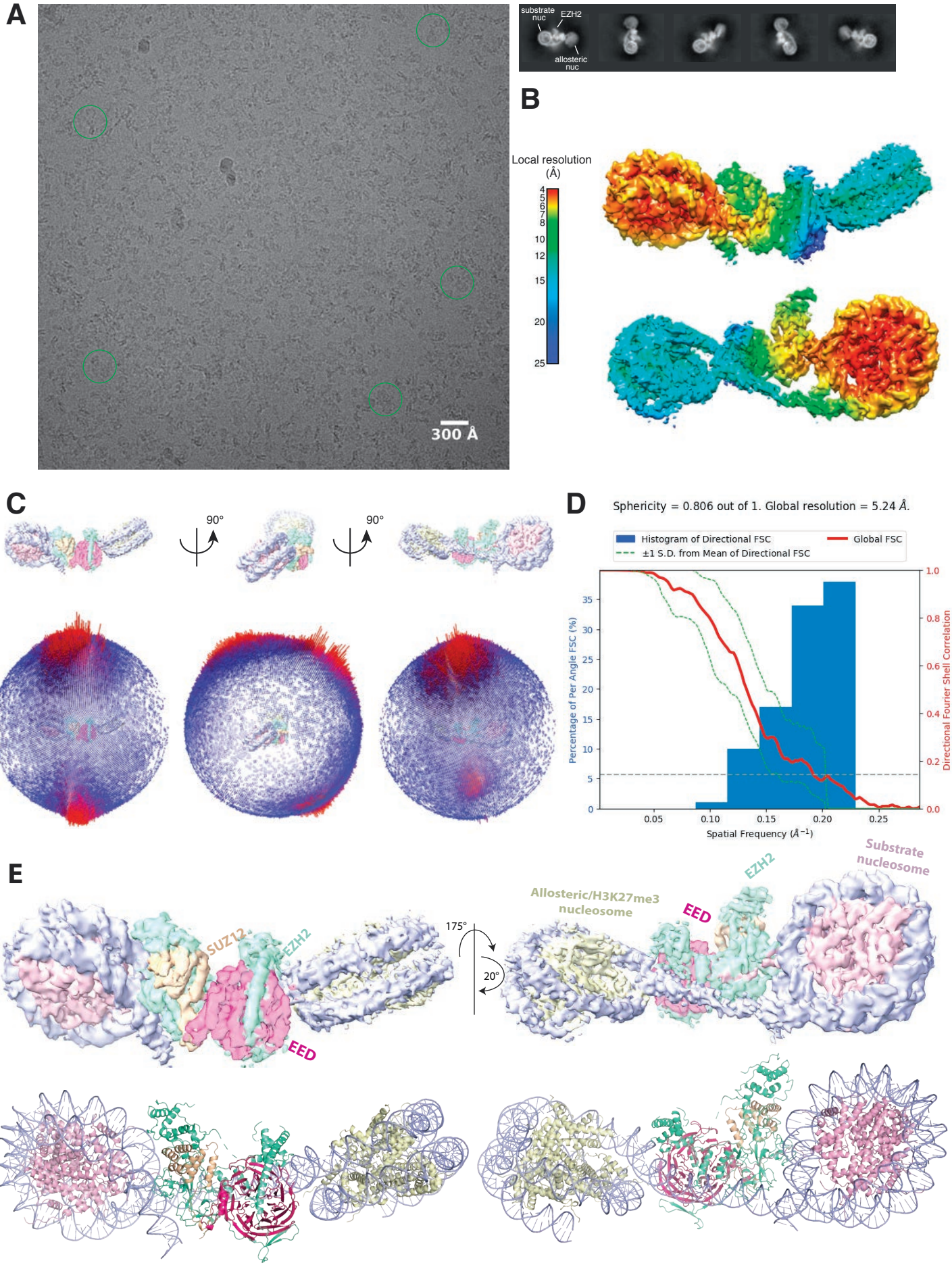


Fig. S2.

Overview of the cryo-EM Data-Processing and Particle Sorting Scheme (related to Fig. 1).

Processing and particle sorting scheme, also described in Methods. Squares indicate 3D classes (and corresponding particles) chosen for further processing steps based on their nominal global resolution values, translational and rotational accuracy and the presence of detailed structural information. Two final reconstructions were obtained in this study: Overall PHF1-PRC2:di-Nuc, and $EZH2_{sub}$-Nuc_{sub} after performing signal subtraction (mask indicated in pink) and focused refinement.

Figure S2

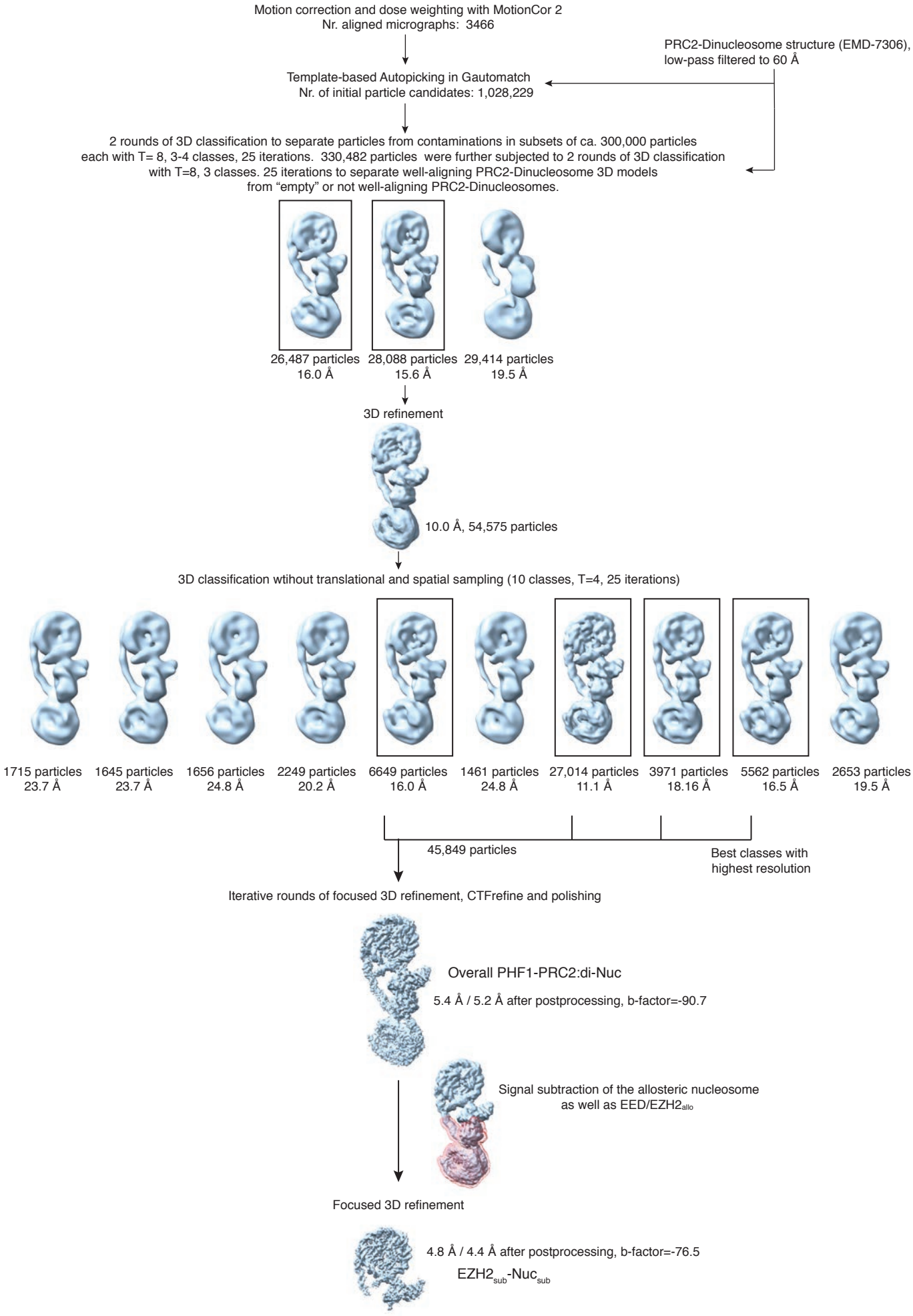


Fig. S3.

The improved map of the interaction between EZH2 and the substrate nucleosome after focused refinement reveals location of H3K36 (related to Fig. 1).

(**A**) Local resolution estimation of the focused 4.4 Å $EZH2_{sub}$:Nuc_{sub} reconstruction. Regions in the nucleosome core as well as the adjacent regions including parts of the H3 N-terminus close to the exit side of the nucleosome are well resolved (4.4 – 5.5 Å). Regions close to the mask, especially the nucleosomal DNA and parts of EZH2, are less well resolved (colors green to blue).

(**B**) Output from the 3DFSC Processing Server (https://3dfsc.salk.edu/ (*55*)) showing the global FSC (red) generated from masked independent half maps of $EZH2_{sub}$-Nuc_{sub} map, and the histogram of the directional FSC (blue) together with the spread (green dotted line) plotted against spatial frequency. The resolution of 4.36 Å was estimated according to the gold standard FSC cutoff of 0.143 (grey dotted line) (*37*). Sphericity is an indication for anisotropy and amounts to 0.748 in this data. The minor directional anisotropy of the data can be explained by the slightly preferred orientation and missing views as seen in **fig. S1C**.

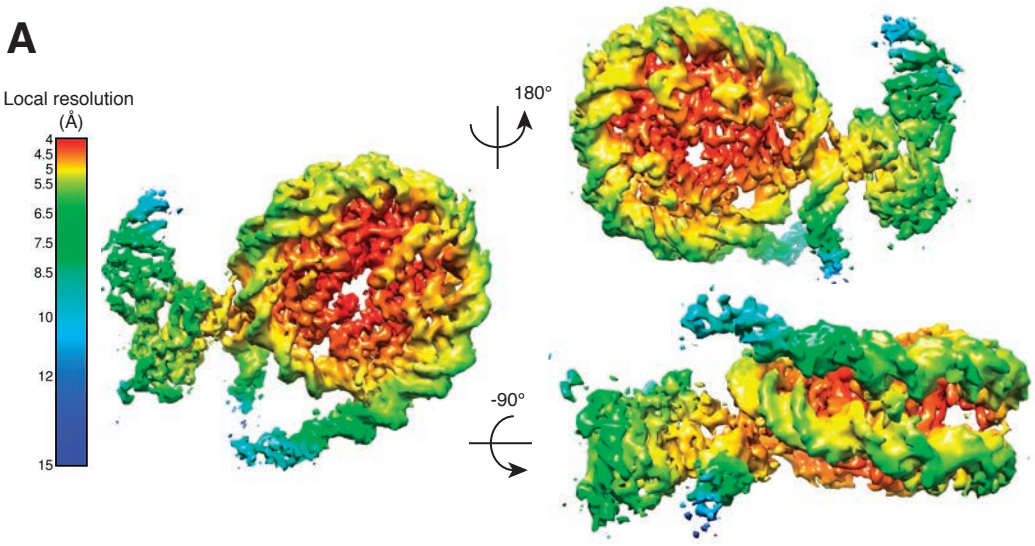
(**C**) FSC between the atomic model and the masked (applied in Phenix) map of $EZH2_{sub}$-Nuc_{sub} after real-space refinement (*46*). Orange line represents the cut-off at 0.5 (6.2 Å) and blue line represents the cut-off at 0.143 (4.3 Å) (see also Table 1) at which map and model Fourier coefficients are most similar (*37*, *56*, *57*).

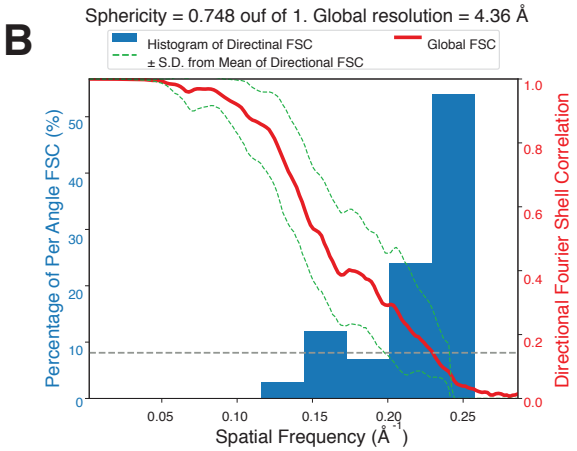
(**D**) Selected regions within $EZH2_{sub}$-Nuc_{sub} showing side chain density, e.g. K36 (red arrow). A red asterisk indicates the last residue of the H3 tail visible in known crystal structures (usually P38 or H39).

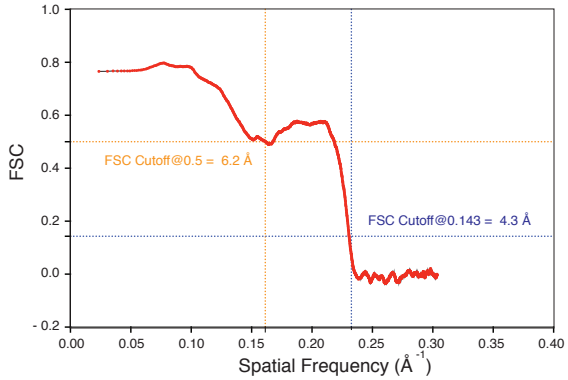
(**E**) Details of the $EZH2_{CXC}$ interaction with nucleosomal DNA with the corresponding cryo-EM density of the $EZH2_{sub}$-Nuc_{sub} map.

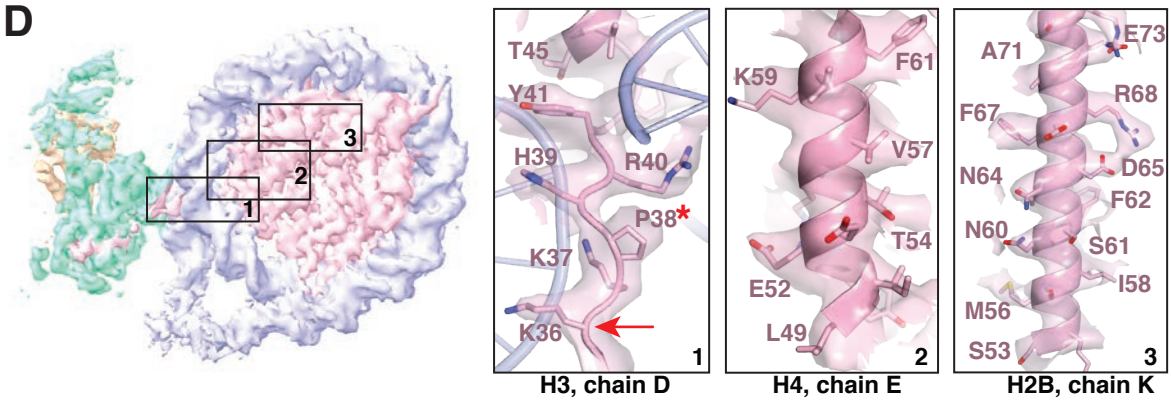
(**F**) Details of the vicinity of K36 with the corresponding density for the H3 tail, EZH2 and nucleosomal DNA.

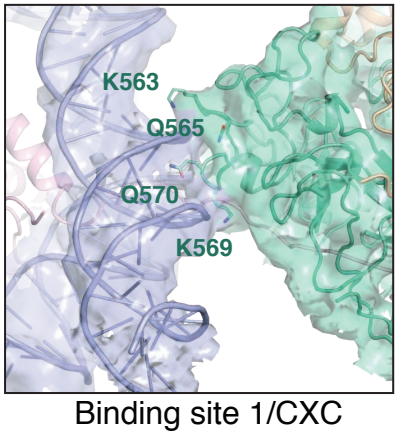
Figure S3











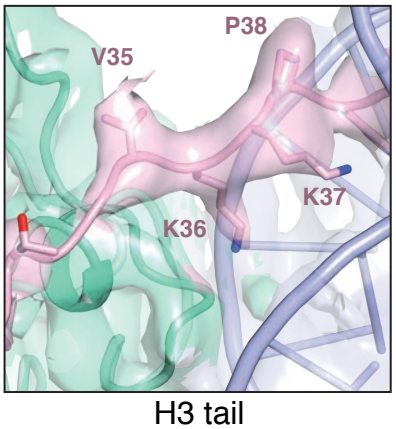


Fig. S4.

Accommodation of unmodified H3K36 in the EZH2$_{CXC}$-DNA interaction interface is essential for H3K27 methylation on nucleosomes (related to Fig. 2).
(A) Binding reactions with indicated concentrations of PRC2 (lanes 1-10) or PRC2$^{EED>A}$ (lanes 11-20) and 45 nM 6-carboxyfluorescein-labeled mononucleosomes, analyzed by EMSA on 1.2% agarose gels.
(**B**) Coomassie-stained 4-12% SDS-PAGE of the HMTase reactions shown in **Fig. 2B**. *Xenopus laevis* (*X.l.*) nucleosomes were used for these experiments. The short 5-kDa PHF$_C$ fragment is not visible on this gel.
(**C**) Coomassie-stained 4-12% SDS-PAGE of the HMTase reactions shown in **Fig. 2D**. *Drosophila melanogaster* (*D.m.*) nucleosomes were used for these experiments. The short 5-kDa PHF$_C$ fragment is not visible on this gel.
(**D**) Full ESI MS spectra (upper part) and full deconvoluted MS spectra (lower part) shown for input peptides without PRC2 as a control (left) and with PRC2 (right) to ensure no overlapping between possible adduct peaks and monomethylation peaks.

Figure S4

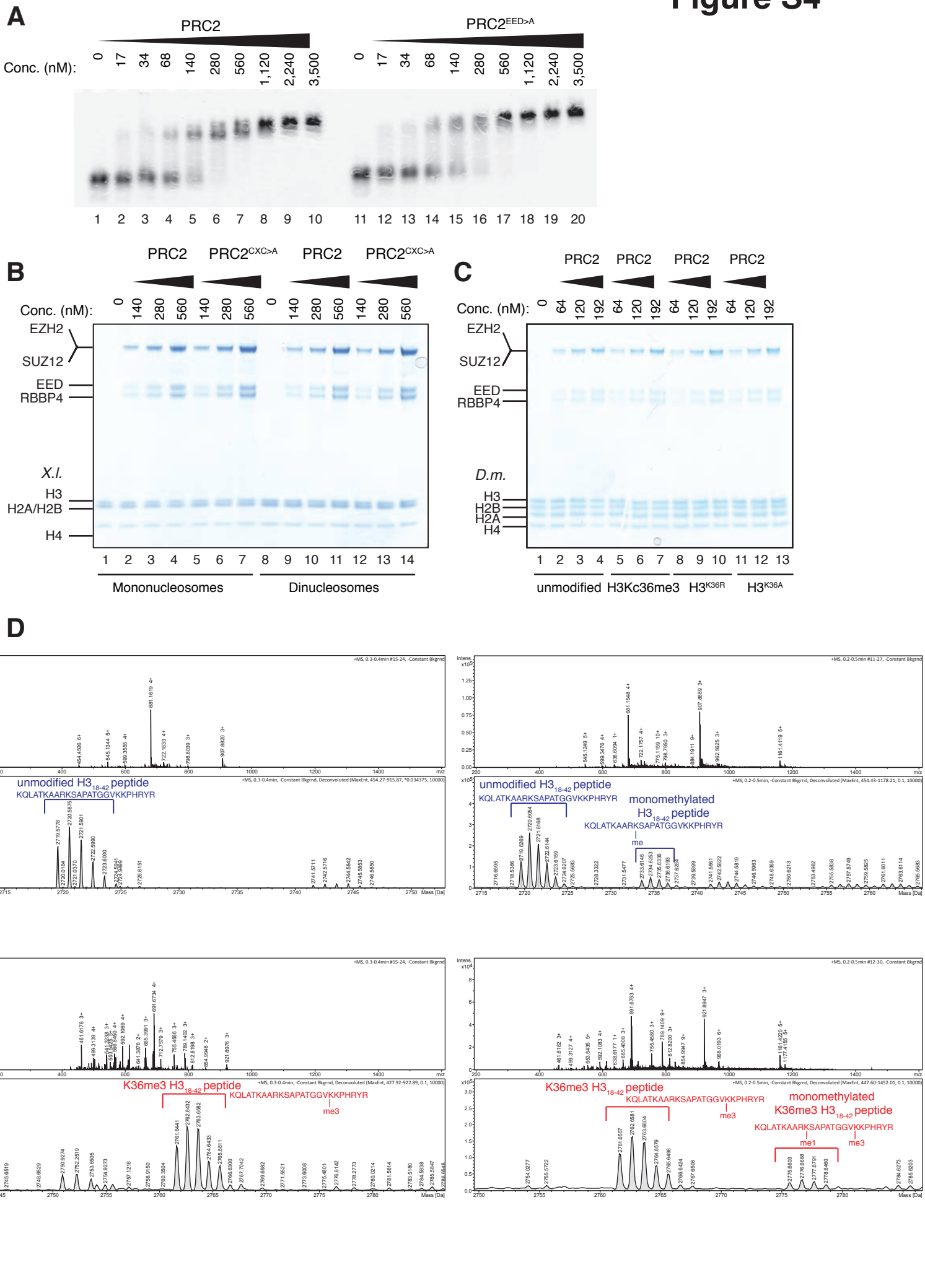


Fig. S5.

$H3^{K36A}$ and $H3^{K36R}$ mutants show altered H3K27me3 profiles and deregulation of HOX genes (related to Fig. 3)

(**A**) Western blot analysis on serial dilutions (8:4:2:1) of total nuclear extracts from 21-24 hr old embryos, probed with antibodies against H3K36me3, H3K36me2 or H3K27me3; in each case, probing of the same membranes with antibodies against Caf1-55 and H2B served as controls for loading and western blot processing. The residual H3K36me3 and H3K36me2 signals in the mutants (lanes 5-12) mostly represent the modified forms of maternally-deposited wild-type H3 (compare with (**B**)). It is not clear why H3K36me3 and H3K36me2 signals in $H3^{K36R}$ mutants (lanes 9-12) are less severely reduced than in $H3^{K36A}$ mutants (lanes 5-8) but the H3K36me2 ChIP-seq profile in $H3^{K36R}$ mutant embryos is also less severely reduced than in $H3^{K36A}$ mutants (**Fig. 3A**). H3K27me3 levels were not detectably reduced in either mutant.

(**B**) Western blot analysis on serial dilutions (4:2:1) of total cell extracts from wing, haltere and 3rd leg imaginal disc tissues from third instar larvae with the genotypes *Df HisC/Df HisC; 12xHisGU^{wt}/ 12xHisGU^{wt}* (*wt*, lanes 1-3) and *Df HisC/Df HisC; 12xHisGU^{H3K36R}/ +* ($H3^{K36R}$, lanes 4-6) from the strains provided by Matera and colleagues (*2*). Blots were probed with antibodies against H3K36me3 or H3K36me2 and, as control, against Caf1-55 and H4. The reduction of H3K36me3 and H3K36me2 signals in $H3^{K36R}$ mutants compared to *wt* is more pronounced than in (**A**) because maternally-deposited wild-type H3 has been strongly diluted by cell division and/or become depleted by protein turnover by this stage of development. The faint residual H3K36me3 and H3K36me2 signals in lane 4 could represent the methylated forms of *H3.3*, or they could be due to cross-reactivity of the antibodies with other epitopes. The more complete replacement of wild-type H3 by H3K36R in $H3^{K36R}$ mutant larvae might also account for the reduction of H3K27me3 bulk levels detected in larvae (*58*), as opposed to embryos (**A**).

(**C**) Left: H3K27me3 and H3K36me2 ChIP-seq profiles at the genomic region of the *Antennapedia* complex containing the HOX genes *lab*, *pb*, *Dfd*, *Scr* and *Antp* with tracks and representation as in **Fig. 3B**. Note the reduction of H3K27me3 levels across the entire *Antennapedia* complex in $H3^{K36A}$ and $H3^{K36R}$ mutants. Right: Zoomed-in version of the region marked by the dotted line in the left panel with the y-scale of the H3K27me3 profile adjusted as in **Fig. 3B** to visualize the increase of H3K27me3 levels in $H3^{K36A}$ and $H3^{K36R}$ mutants in regions showing H3K36me2 enrichment in *wt* embryos.

(**D**) Ventral views of stage 16 embryos stained with an Abd-B antibody. Vertical bar marks the anterior boundary of Abd-B expression in parasegment 10 in *wt* embryos and arrowheads and empty arrowheads mark misexpression and loss of expression of Abd-B, respectively, as in **Fig. 3C**.

Explanatory notes:

In *wt* embryos, Abd-B expression levels are stepwise increased from ps10 to ps14 (i.e. expression levels in ps10 < ps11 <ps12 < ps13 = ps14). In $esc^{mat\text{-}zyg\text{-}}$ (esc^{-}) mutant embryos Abd-B is expressed at high levels in all ps, and the graded pattern is lost, i.e. compare Abd-B levels in ps10 of the esc^{-} mutant and *wt* embryos (small arrows). In *wt* animals, Polycomb repression requiring PRC2 function therefore dampens Abd-B expression in ps10 to 12. In $ash1^{mat\text{-}zyg\text{-}}$ ($ash1^{-}$) mutant embryos, the loss of Abd-B expression in ps10 to ps12 (empty arrowheads) is due to ectopic repression by Polycomb, because HOX gene expression in these segments is restored in *ash1* mutants lacking PRC2 (*7*). As illustrated in **Fig. 3C**, $H3^{K36A}$ and $H3^{K36R}$ mutant embryos show misexpression of Abd-B (arrowheads) as well as loss of expression in ps10 (empty arrowheads), indicative of a failure to maintain Polycomb repression anterior to ps10 but also compromised ability to counteract Polycomb repression in ps 10. The $H3^{K36R}$ embryo shown here (panel **D**) is from the strain generated by Matera and colleagues (*2*) (genotype: *Df HisC/ Df HisC; 12xHisGU^{H3K36R}/+*) and the *wt* embryo is from the corresponding *Df HisC/ Df HisC; 12xHisGU^{wt}/ 12xHisGU^{wt}* control strain from the same study (*2*). See Supplementary Materials for full description of all genotypes.

(**E**) Eye imaginal discs from third instar larvae stained with Antp (red) and Hoechst (DNA) to label all nuclei. The same strains from Matera and colleagues (*2*) used in (**D**) were used here (*wt* genotype: Df *HisC/ Df HisC; 12xHisGU^{wt}/ 12xHisGU^{wt}*, and $H3^{K36R}$ genotype: *Df HisC/ Df HisC; 12xHisGU^{H3K36R}/ +*). Note the misexpression of Antp in the antennal disc portion of the disc (arrowheads) of the $H3^{K36R}$ mutant (compare with **Fig. 3D**).

Figure S5

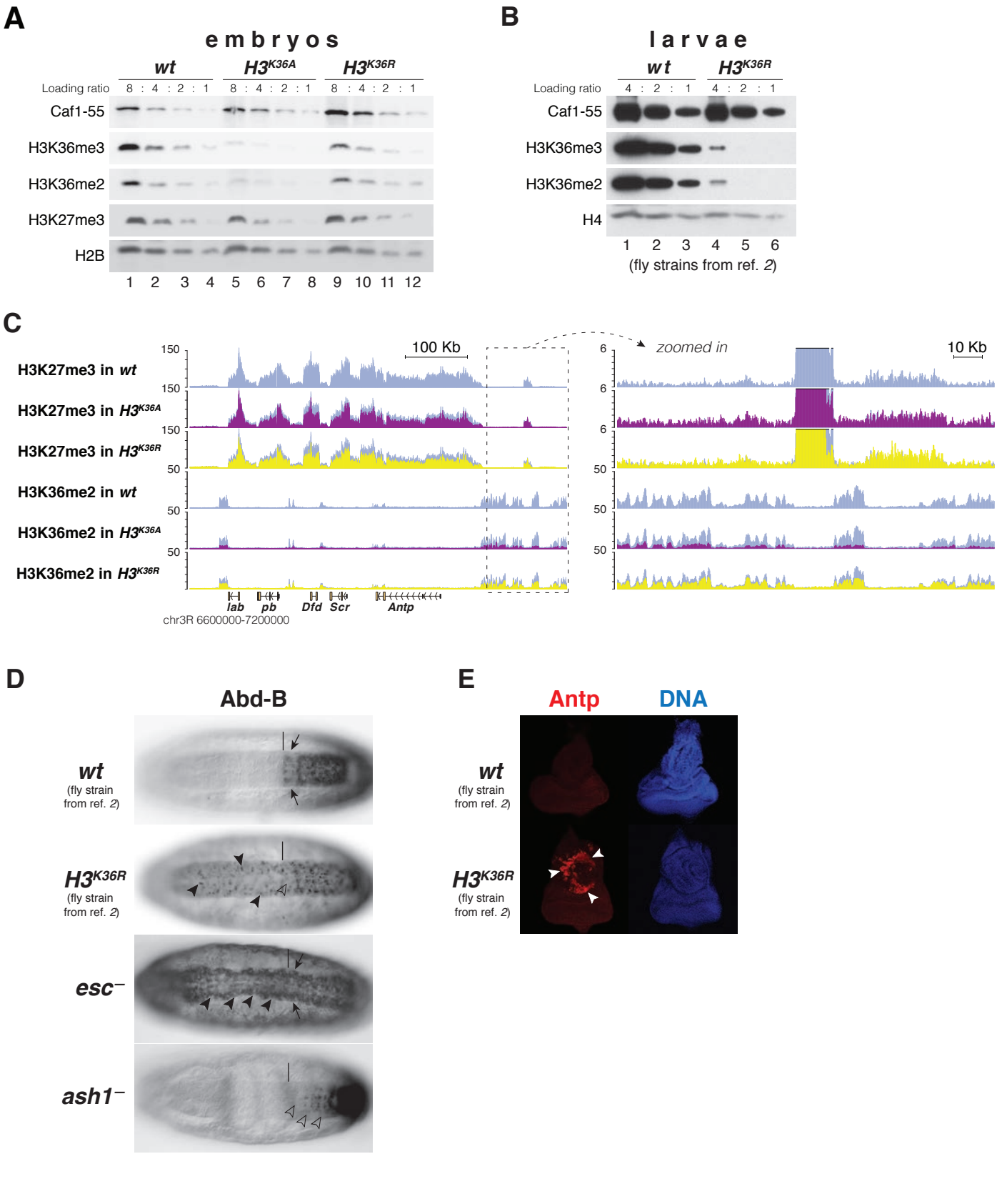


Table S1.

Cryo electron microscopy data collection summary, processing statistics and model

Cryo electron microscopy data collection		
Microscope	FEI Titan Krios GII	
Voltage (kV)	300	
Camera	Gatan K2-Summit	
Energy Filter	Gatan Quantum-LS (GIF)	
Pixel size (Å/pix)	1.75	
Preset target global defocus range (µm)	0.5 - 3.5	
3D reconstruction		
Number of movies	3466	
Initially selected particle candidates	1,028,229	
Final number of particles	45,849	
	Overall PHF1-PRC2:diNuc	**$EZH2_{sub}$-Nuc_{sub}**
Resolution $_{FSC\ independent\ halfmaps}$ (Å)[a]	5.24	4.36
Local resolution range (Å)	4.01 - 24.97	4.01 – 15.00
Sharpening B-factor ($Å^2$)	-90.7	-76.5
Refinement		**$EZH2_{sub}$-Nuc_{sub}**
No. atoms		30893 (Hydrogens: 13818)
Residues		Protein: 1276 Nucleotide: 336
Ligands		ZN: 8
CC_{mask}, CC_{box}, CC_{peaks}, CC_{volume}[b]		0.61, 0.73, 0.61, 0.64
Mean CC for ligands		0.57
Resolution$_{FSC\ map\ vs.\ model\ (0/0.143/0.5)}$ (Å)[b]		4.2/4.3/6.2
R.m.s. deviations		
Bond lengths (Å)		0.014
Bond angles (°)		1.054
Ramachandran favored (%)		93.40
Ramachandran gen. allowed (%)		6.36
Ramachandran disallowed (%)		0.24
MolProbity score		3.02
Clash score		24.14

[a]according to the Fourier Shell Correlation (FSC) cut-off criterion of 0.143 defined in (Rosenthal and Henderson, 2003)
[b]according to the map-vs.-model Correlation Coefficient definitions in (Afonine et al., 2018a)

Table S2.

Number of aligned reads to the *D. melanogaster* and *D. pseudoobscura* genomes from ChIP and input datasets and normalization process (related to Fig. 3 and fig. S5).

Movie S1

Cryo-EM structure of the PHF1-PRC2:di-Nuc complex (related to Fig. 1).

References

1. A. R. Pengelly, Ö. Copur, H. Jäckle, A. Herzig, J. Müller, A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*. **339**, 698–699 (2013).

2. D. J. McKay *et al.*, Interrogating the function of metazoan histones using engineered gene clusters. *Dev. Cell*. **32**, 373–386 (2015).

3. N. J. Francis, R. E. Kingston, C. L. Woodcock, Chromatin compaction by a polycomb group protein complex. *Science*. **306**, 1574–1577 (2004).

4. K. H. Hansen *et al.*, A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).

5. R. Margueron *et al.*, Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. **461**, 762–767 (2009).

6. L. Jiao, X. Liu, Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*. **350**, aac4383–aac4383 (2015).

7. T. Klymenko, J. Müller, The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).

8. F. W. Schmitges *et al.*, Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell*. **42**, 330–341 (2011).

9. W. Yuan *et al.*, H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989 (2011).

10. L. J. Gaydos, A. Rechtsteiner, T. A. Egelhofer, C. R. Carroll, S. Strome, Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell Rep*. **2**, 1169–1177 (2012).

11. G. Streubel *et al.*, The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell*. **70**, 371–379.e5 (2018).

12. N. Justin *et al.*, Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun*. **7**, 11316 (2016).

13. X. Wang *et al.*, Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA JA -, doi:10.1038/nsmb.3487.

14. N. Guidotti, C. C. Lechner, A. L. Bachmann, B. Fierz, A Modular Ligation Strategy for Asymmetric Bivalent Nucleosomes Trimethylated at K36 and K27. *Chembiochem*. **20**, 1124–1128 (2019).

15. J. Choi *et al.*, DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat. Struct. Mol. Biol.* **298**, 1039 (2017).

16. M. D. Simon *et al.*, The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell*. **128**, 1003–1012 (2007).

17. S. Poepsel, V. Kasinath, E. Nogales, Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* **25**, 154–162 (2018).

18. S. Chen, L. Jiao, M. Shubbar, X. Yang, X. Liu, Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol. Cell*. **69**, 840–852.e5 (2018).

19. V. Kasinath *et al.*, Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science*. **359**, 940–944 (2018).

20. H. Li *et al.*, Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* (2017), doi:10.1038/nature23881.

21. C. Ballaré *et al.*, Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* **19**, 1257–1265 (2012).

22. L. Cai *et al.*, An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell*. **49**, 571–582 (2013).

23. C. A. Musselman *et al.*, Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun*. **4**, 2969–9 (2013).

24. S. Chen, L. Jiao, X. Liu, X. Yang, X. Liu, A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Mol. Cell*. **77**, 1265–1278.e7 (2020).

25. P. T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

26. M. Nekrasov, B. Wild, J. Müller, Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Rep.* **6**, 348–353 (2005).

27. B. Papp, J. Müller, Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20**, 2041–2054 (2006).

28. S. Schmähling *et al.*, Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC. *Development*. **145**, dev163808 (2018).

29. R. K. McGinty, R. C. Henrici, S. Tan, Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*. **514**, 591–596 (2014).

30. E. J. Worden, N. A. Hoffmann, C. W. Hicks, C. Wolberger, Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell*. **176**, 1490–1501.e12 (2019).

31. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. **389**, 251–260 (1997).

32. J. Bonnet *et al.*, Quantification of Proteins and Histone Marks in Drosophila Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation. *Dev. Cell*. **51**, 632–644.e6 (2019).

33. K. Luger, T. J. Rechsteiner, T. J. Richmond, Preparation of nucleosome core particle from recombinant histones. *Meth. Enzymol.* **304**, 3–19 (1999).

34. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*. **14**, 331–332 (2017).

35. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

36. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. **7**, 163 (2018).

37. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

38. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

39. X.-C. Bai, E. Rajendra, G. Yang, Y. Shi, S. H. W. Scheres, Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife*. **4**, 1485 (2015).

40. Q. Zhou *et al.*, Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Res.* **25**, 551–560 (2015).

41. S. L. Ilca *et al.*, Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat Commun*. **6**, 8843–8 (2015).

42. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. **25**, 1605–1612 (2004).

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

44. A. Casañal, B. Lohkamp, P. Emsley, Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).

45. O. V. Sobolev, P. V. Afonine, P. D. Adams, A. Urzhumtsev, Programming new geometry restraints: parallelity of atomic groups. *J Appl Crystallogr*. **48**, 1130–1141 (2015).

46. P. V. Afonine *et al.*, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol*. **74**, 531–544 (2018).

47. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).

48. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc.* **2**, 1896–1906 (2007).

49. U. Günesdogan, H. Jäckle, A. Herzig, A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. *EMBO Rep.* **11**, 772–776 (2010).

50. E. Dorafshan *et al.*, Ash1 counteracts Polycomb repression independent of histone H3 lysine 36 methylation. *EMBO Rep.*, e46762 (2019).

51. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

52. G. dos Santos *et al.*, FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* **43**, D690–7 (2015).

53. B. Zacher *et al.*, Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN. *PLoS ONE*. **12**, e0169249 (2017).

54. M. C. Gambetta, K. Oktaba, J. Müller, Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science*. **325**, 93–96 (2009).

55. Y. Z. Tan *et al.*, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods*. **14**, 793–796 (2017).

56. R. Henderson *et al.*, (2012), vol. 20, pp. 205–214.

57. P. B. Rosenthal, J. L. Rubinstein, Validating maps from single particle electron cryomicroscopy. *Curr. Opin. Struct. Biol.* **34**, 135–144 (2015).

58. M. P. Meers *et al.*, Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity. *Elife*. **6**, 1191 (2017).