

Basis of the H2AK119 specificity of the Polycomb repressive deubiquitinase

<https://doi.org/10.1038/s41586-023-05841-y>

Received: 26 February 2022

Accepted: 14 February 2023

Published online: 29 March 2023

 Check for updates

Weiran Ge^{1,2,5}, Cong Yu^{1,2,5}, Jingjing Li^{1,2,5}, Zhenyu Yu¹, Xiaorong Li¹, Yan Zhang¹, Chao-Pei Liu¹, Yingfeng Li¹, Changlin Tian³, Xinzheng Zhang^{1,2}, Guohong Li^{1,2}, Bing Zhu^{1,2,4} & Rui-Ming Xu^{1,2,4}✉

Repression of gene expression by protein complexes of the Polycomb group is a fundamental mechanism that governs embryonic development and cell-type specification^{1–3}. The Polycomb repressive deubiquitinase (PR-DUB) complex removes the ubiquitin moiety from monoubiquitinated histone H2A K119 (H2AK119ub1) on the nucleosome⁴, counteracting the ubiquitin E3 ligase activity of Polycomb repressive complex 1 (PRC1)⁵ to facilitate the correct silencing of genes by Polycomb proteins and safeguard active genes from inadvertent silencing by PRC1 (refs. 6–9). The intricate biological function of PR-DUB requires accurate targeting of H2AK119ub1, but PR-DUB can deubiquitinate monoubiquitinated free histones and peptide substrates indiscriminately; the basis for its exquisite nucleosome-dependent substrate specificity therefore remains unclear. Here we report the cryo-electron microscopy structure of human PR-DUB, composed of BAP1 and ASXL1, in complex with the chromatosome. We find that ASXL1 directs the binding of the positively charged C-terminal extension of BAP1 to nucleosomal DNA and histones H3–H4 near the dyad, an addition to its role in forming the ubiquitin-binding cleft. Furthermore, a conserved loop segment of the catalytic domain of BAP1 is situated near the H2A–H2B acidic patch. This distinct nucleosome-binding mode displaces the C-terminal tail of H2A from the nucleosome surface, and endows PR-DUB with the specificity for H2AK119ub1.

Polycomb-group (PcG) proteins repress gene expression through post-translationally modifying histones and promoting the formation of facultative heterochromatin^{1,2,10}. The archetypal PcG complexes PRC1 and PRC2 affect developmental programs and human diseases through H2AK119 ubiquitination and H3K27 methylation, respectively^{11–13}. The distinct PR-DUB complex fulfils its PcG function through a mechanism antagonistic to that of PRC1; namely, by deubiquitination of H2AK119ub1 on the nucleosome^{4,14}. This property of PR-DUB ensures the correct levels of H2AK119ub1 for compaction of the Polycomb chromatin domain, and protects active genes from silencing by PRC1 (refs. 6–9). Human PR-DUB is composed of BAP1—a ubiquitin C-terminal hydrolase (UCH)—and one of the ASXL (additional sex combs-like) proteins (ASXL1–ASXL3). Both BAP1 and the ASXL proteins are evolutionarily conserved, multidomain proteins that are frequently mutated in cancer^{15–19} (Fig. 1a). Notably, some cancer-associated truncations of ASXL1–ASXL3 behave as gain-of-function mutations of PR-DUB, making BAP1 a potential therapeutic target in leukaemia^{20–22}.

BAP1 is inactive on its own, but binding to the deubiquitinase adaptor (DEUBAD) domain of ASXL1–ASXL3 activates BAP1's deubiquitinase activity^{4,23}. The activated BAP1 complex specifically deubiquitinates H2AK119ub1 in the context of the nucleosome, despite being capable of promiscuously deubiquitinating free histones and peptide substrates^{4,23} (Fig. 1b). The crystal structure of *Drosophila* PR-DUB,

consisting of Calypso (Caly) and the DEUBAD domain of Asx (Asx^{DEU}), shows that Asx^{DEU} interacts with the UCH37-like (ULD) domain of Caly to form a complete ubiquitin-binding pocket^{24,25}. The Caly–Asx^{DEU} structure lacks a bound substrate, although we can learn a great deal about ubiquitin binding from the highly homologous structure of the UCH37–RPN13^{DEU} complex (UCH37 is also known as UCH-L5) bound to ubiquitin^{26,27}. However, what determines the nucleosome-dependent H2AK119ub1 specificity remains unclear. To solve this problem, we determined the structure of PR-DUB in complex with the H2AK119ub1 chromatosome, which is formed by the binding of linker histone H1 (H1.4) to the 187-bp DNA nucleosome core particle (NCP) with H2AK119ub1, termed NCP(H2AK119ub1).

Overall structure

The complexes of a human BAP1–ASXL1 heterodimer, with a catalytically inactive C91S mutant version of full-length BAP1 and a 378-residue N-terminal fragment of ASXL1, bound to NCP(H2AK119ub1) with and without linker histone H1 (H1.4), were prepared for structural determination by single-particle cryo-electron microscopy (cryo-EM) (Fig. 1a and Extended Data Fig. 1). The structure of the H1-bound complex was determined at an overall resolution of 3.0 Å, and mask-based local refinement improved the resolution of the relatively flexible region

¹National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. ²School of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. ³Division of Life Sciences and Anhui Provisional Engineering Laboratory of Peptide Drugs, University of Science and Technology of China, Hefei, China. ⁴Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China. ⁵These authors contributed equally: Weiran Ge, Cong Yu, Jingjing Li. ✉e-mail: zhubing@ibp.ac.cn; rmxu@ibp.ac.cn

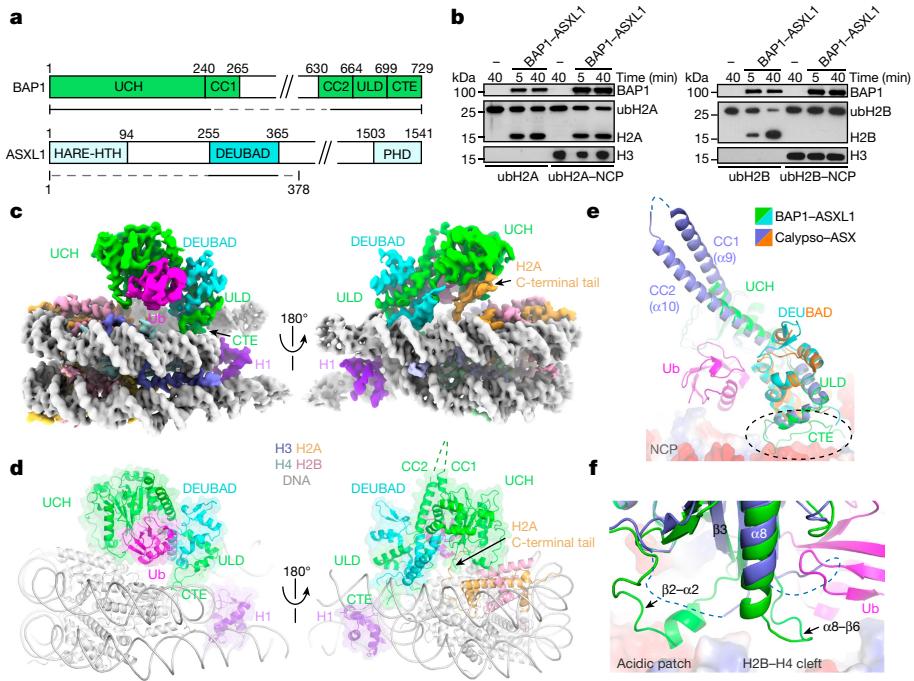


Fig. 1 | Structure of PR-DUB bound to the H2AK119ub1 chromatosome. **a**, Domain organization of human BAP1 and ASXL1. Domain names and residue numbers marking their boundaries are labelled. The underlines denote the fragments used in this study, and the dashed segments indicate disordered regions. CC, coiled-coil domain. **b**, Western blot detection of deubiquitination by the BAP1-ASXL1(1–378) complex using free histone and nucleosome substrates, monoubiquitinated at H2AK119 (left) or H2BK120 (right). Lanes with a ‘-’ sign are mock reactions without enzyme added. The top, middle and bottom blots used anti-BAP1, anti-H2A (or anti-H2B) and anti-H3 antibodies, respectively. H3 lanes were run on a separate gel as sample-processing controls.

c, Two views of the composed density map of the BAP1-ASXL1-H1-NCP complex. The colour code is used throughout. Ub, ubiquitin. **d**, Atomic model of the complex structure, with the cartoon diagrams of BAP1, ASXL1, H1 and ubiquitin superimposed with a transparent surface rendition, shown in the same views as above. The dashed lines denote disordered regions. **e,f**, Superposition with the *Drosophila* Caly-Asx^{DEU} structure (PDB_id, 6HGC) reveals structural differences in the coiled-coil domains (CC1 and CC2) and the CTE (e), and in the β2-α2 and α8-β6 loops (f). NCP is shown in an electrostatic-potential surface representation.

surrounding BAP1, ASXL1 and the ubiquitin moiety to 3.9 Å (Fig. 1c, Extended Data Fig. 2 and Extended Data Table 1). The high-quality density maps allowed reliable building of the atomic model (Extended Data Fig. 3). In the absence of H1, a relatively lower resolution (3.8 Å overall) complex is visualized (Extended Data Fig. 4). The two structures are highly similar (Extended Data Fig. 4e), and we use the higher-resolution one for most of the analysis here, unless explicitly specified otherwise.

A heterodimer of BAP1 and the DEUBAD domain of ASXL1 (ASXL1^{DEU}) is bound on one side of the nucleosome (Fig. 1d). The catalytic UCH domain, the ASXL1-binding ULD domain, an approximately 15-residue positively charged segment of BAP1’s C-terminal extension (CTE) and a nearly intact ASXL1^{DEU} (amino acids (aa) 249–341) are well resolved, whereas a large non-conserved internal region (aa 245–645) and several exposed loops of BAP1, and around 250 N-terminal residues of ASXL1, are disordered (Fig. 1a). The overall structure of the BAP1-ASXL1^{DEU} heterodimer is similar to that of Caly-Asx^{DEU}. Major differences include the longer antiparallel helices, α9 and α10, in the coiled-coil domain of Caly, and the ordered CTE and β2-α2 and α8-β6 loops in BAP1 (Fig. 1e,f and Extended Data Fig. 5a–c). Despite the shorter ordered α9–α10 helices in BAP1, their role in bringing the UCH and ULD domains spatially close—thus allowing the formation of a complete ubiquitin cleft by the BAP1 UCH and the ASXL1^{DEU} domains—is maintained (Fig. 1d,e). It has been suggested that α9–α10 of Caly might facilitate the homodimerization and deubiquitination of nucleosomal H2AK119ub1 (refs. 24,25), but our multi-angle light-scattering (MALS) analysis does not provide evidence of dimerization of the human BAP1-ASXL1(1–378) complex (Extended Data Fig. 1d). In the middle of the ordered β2-α2 loop of BAP1, a conserved positively charged sequence motif spanning

residues 56–60 adopts a helical form and is located in the acidic patch of histones H2A and H2B (Fig. 1f and Extended Data Fig. 5b). In addition, the tip of the α8-β6 loop (aa 220–223) is located above the cleft formed between histones H2B and H4, and the C-terminal portion of this loop interacts with ubiquitin (Fig. 1f and Extended Data Fig. 5c).

Nucleosome anchoring through the BAP1 CTE

A section of the positively charged BAP1 CTE has previously been shown to be important for nucleosome binding and deubiquitination of H2AK119ub1 (ref. 23). In the structure, residues 699–715 of the BAP1 CTE interact with both histones and DNA near the nucleosomal dyad (superhelical location 0; SHL 0) (Figs. 1e and 2a). Residues 699–706 form a finger-like structure that pokes into a nucleoprotein pocket composed of the N-terminal ends of α3 of H3 and α2 of H4 of the same heterodimer, and two DNA strands spanning the minor groove at SHL 0 (Fig. 2b). Although Arg699 and Arg700 only contact the DNA backbone, Arg701 is inserted into the DNA minor groove and makes contacts with the bases and the sugar group (Fig. 2b,c). The side chain of Gln702 interacts with the DNA backbone and Ser47 of H4, and its main chain, together with that of Gly703, contacts Met120 and Lys122 of H3 (Fig. 2b). Outside the finger region, the remaining ordered CTE portion stretches along the DNA backbone, with Arg708 and Arg713 making direct contacts. The CTE sequence is strictly conserved among vertebrate BAP1 proteins, but a 35-residue insertion is found in *Drosophila* Caly (Extended Data Fig. 5a).

In agreement with previous findings, a partial truncation of CTE from residue 711 (ref. 23), or removing the entire CTE from residue 697,

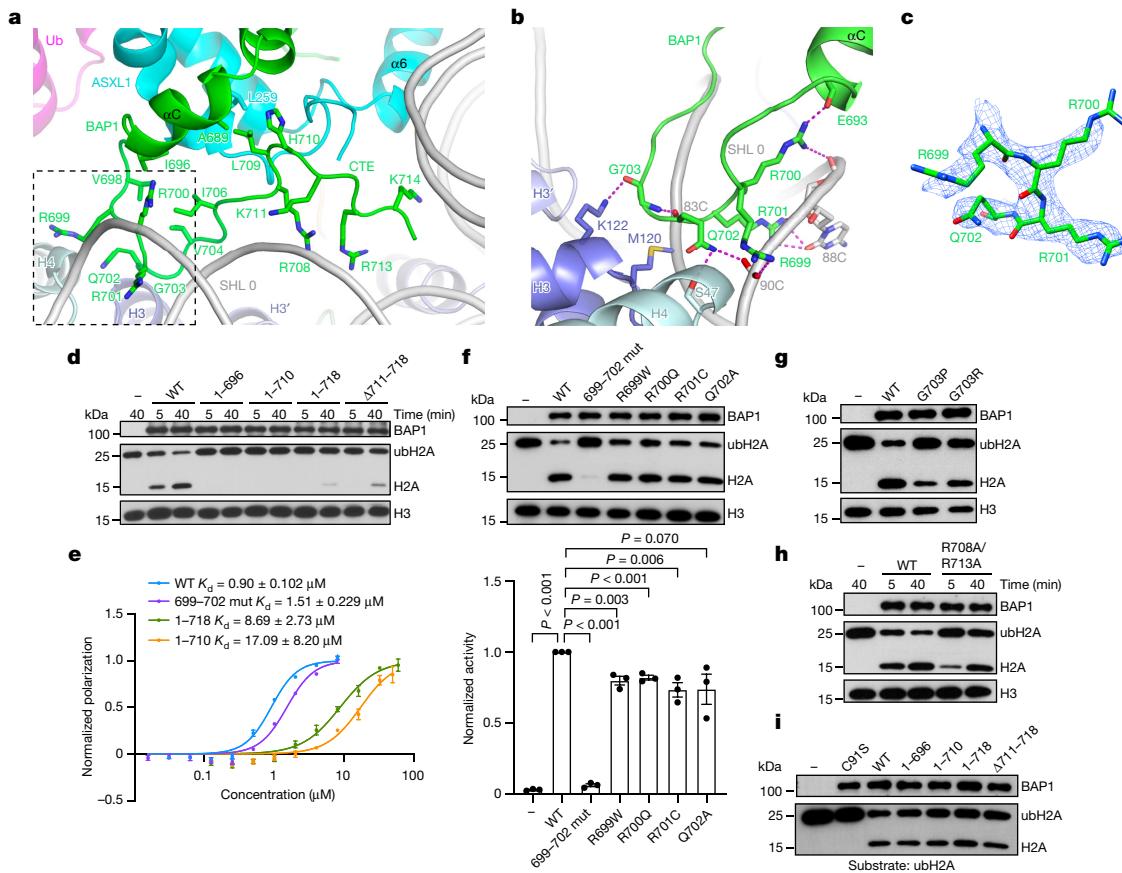


Fig. 2 | Structure and function of the BAP1 CTE. **a**, The BAP1 CTE is ordered from I696 to P715, and side chains of key residues are depicted in a stick model. **b**, Enlarged view of the interface between BAP1's CTE finger motif and NCP. Residues and nucleotides that are involved are indicated. Magenta dashed lines represent hydrogen bonds. **c**, EM density map surrounding residues 699–702 of the BAP1 CTE finger motif, contoured at 7σ . **d**, Western blot detections of nucleosomal H2AK119 deubiquitination of PR-DUB assembled with wild-type (WT) BAP1 or the indicated BAP1 mutants and ASXL1(1–378). Deubiquitination reactions were monitored at 5 min and 40 min; ‘-’ signs indicates lanes without enzyme added. The top, middle and bottom blots used anti-His tag, anti-H2A and anti-H3 antibodies. Quantification shown in Extended Data Fig. 5d. **e**, Fluorescence polarization measurements of the binding affinities of the

indicated BAP1-ASXL1(1–378) complexes to NCP(H2AK119ub1). K_d values were obtained by fitting the data to a model of specific binding with Hill slope in GraphPad ($n = 3$ biologically independent experiments, mean \pm s.e.m.).

f–h, Western blot analyses of the indicated BAP1 mutants, assayed as described in **d**. Blotting was performed at 10 min (**f,g**), or at 5 min and 40 min (**h**). H3 lanes are sample-processing controls. The bottom panel in **f** is the quantification of the top blot ($n = 3$ biological independent experiments, mean \pm s.e.m.), normalized against the value of the wild-type complex. P values were derived from an unpaired two-tailed t -test, with no adjustment made for multiple comparisons. Quantifications of **g,h** are shown in Extended Data Fig. 5e,f.

abolished the nucleosomal H2AK119 deubiquitination activity of PR-DUB in vitro (Fig. 2d and Extended Data Fig. 5d). A shorter deletion from residue 719, or removal of an internal segment spanning residues 711–718, yielded substantially reduced levels of enzymatic activity. BAP1 CTE truncation mutants also exhibited reduced nucleosome-binding abilities, as judged by fluorescent polarization experiments (Fig. 2e). BAP1-ASXL1(1–378) complexes assembled with BAP1(1–710) and BAP1(1–718) bound to NCP approximately 18- and 10-fold more weakly, respectively, than the wild-type complex, which has a dissociation constant (K_d) of $0.9 \mu\text{M}$ (Fig. 2e).

In comparison, BAP1 CTE point mutations had less effect (Fig. 2f). Changing all four residues of BAP1699–702 to alanines (699–702 mut) seriously compromised DUB activity, although nucleosome binding was only mildly affected (Fig. 2e). Three cancer missense mutations, R699W, R700Q and R701C (ref. 28), as well as a Q702A mutation, led to modest reductions of activity compared to that of the wild-type complex (Fig. 2f). Gly703 at the tip of the CTE finger is situated in a shallow depression on the nucleosome surface near the interface between two H3–H4 heterodimers (Fig. 2a,b). Introducing a sharp bend with a G703P mutation or a G703R bulky side-chain substitution reduced

the nucleosomal H2AK119ub1 DUB activity by about half (Fig. 2g and Extended Data Fig. 5e). Furthermore, alanine substitutions of the DNA-backbone-interacting CTE residues Arg708 and Arg713 also resulted in a marked reduction of BAP1's nucleosomal DUB activity (Fig. 2h and Extended Data Fig. 5f).

Because the BAP1 CTE is located distant from the ubiquitin-binding pocket and the enzymatic active site, it is possible that it influences H2AK119ub1DUB activity by positioning the BAP1-ASXL1(1–378) complex correctly with respect to the nucleosome, rather than by directly affecting the enzymatic ability of PR-DUB. Indeed, various BAP1 CTE truncation complexes showed similar levels of enzymatic activity to that of the wild-type BAP1 complex toward free histone H2AK119ub1 (Fig. 2i), unlike their activities toward the nucleosomal substrate (Fig. 2d).

ASXL1 directs the nucleosome binding of the CTE

The structure shows that stabilization of BAP1's α C, to which the CTE is immediately joined, through interaction with ASXL1^{DEU}, is crucial for the spatial positioning of the CTE (Fig. 3a). In particular, Thr262 and Leu267 of ASXL1 contact α C residues through hydrophobic and van der

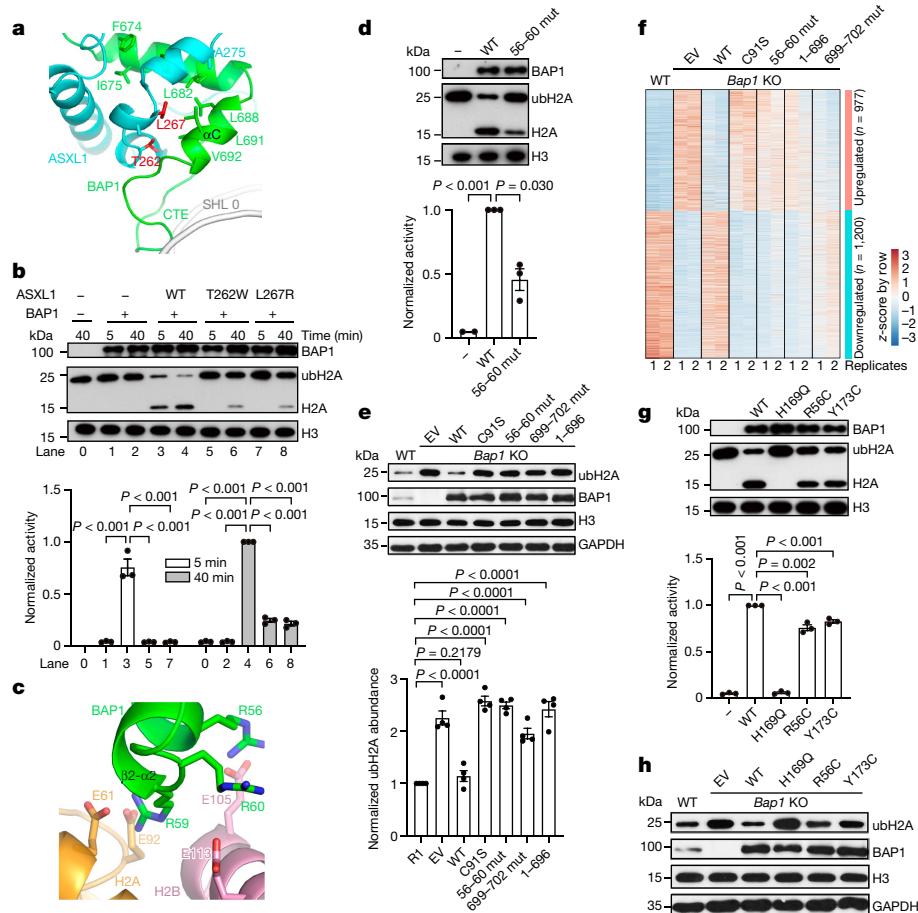


Fig. 3 | Nucleosome binding is key to H2AK119ub1 deubiquitination.

a, Interaction between α C of BAP1^{ULD} (green) and ASXL1^{DEU} (cyan), involving Thr262 and Leu267 (red), directs the proper nucleosome binding of the BAP1 CTE. **b**, Top, western blot detection of H2AK119 deubiquitination by BAP1 or its complexes with wild-type or mutant forms of ASXL1(1–378), detected with anti-His tag (top), anti-H2A (middle) and anti-H3 (bottom) antibodies at 5 min and 40 min. H3 lanes here and in **d,g** were run on separate gels as sample-processing controls. Bottom, quantification of the western blot, normalized against the wild-type value at 40 min. Unshaded and grey-shaded bars indicate values at 5 min and 40 min, respectively. $n = 3$ biologically independent replicates; data are mean \pm s.e.m. P values were derived from an unpaired two-tailed t -test, with no adjustments made for multiple comparisons. **c**, Magnified view of the interface between the RRSRR motif (aa 56–60) of BAP1 and the H2A–H2B acidic patch. **d**, Deubiquitination activity with the five RRSRR

residues mutated to alanine (56–60 mut), measured as in **b** at 10 min. **e**, Top, western blot detection of H2AK119ub1 and BAP1 in wild-type R1 mouse ES cells and in *Bap1* KO cells transfected with an empty vector (EV) or with plasmids re-expressing the indicated BAP1 constructs, using anti-H2AK119ub1, anti-BAP1, anti-H3 and anti-GAPDH antibodies. Bottom, quantification of the levels of ubH2A, normalized against wild-type R1. Data are mean \pm s.e.m. ($n = 4$ biologically independent experiments). **f**, Heat maps of transcriptional changes in BAP1-regulated genes in wild-type cells and the indicated rescued cell lines (two replicates each). **g**, In vitro deubiquitination activity of the indicated BAP1 cancer mutations, detected as in **b** at 10 min. **h**, Western blot detection of the levels of ubH2A in *Bap1* KO cells rescued with the indicated BAP1 constructs. See Extended Data Fig. 9 for quantification and transcriptional changes.

Waals interactions. Replacing them with bulky residues should interfere with the packing of BAP1 α C and alter CTE's mode of NCP binding. Not unexpectedly, T262W or L267R ASXL1 mutant complexes showed clear reductions in H2AK119ub1 DUB activity (Fig. 3b), but no appreciable difference in nucleosome-binding affinity was exhibited by the T262W mutant complex (Extended Data Fig. 5g). This result indicates that the stable positioning of BAP1 α C against ASXL1 directs the CTE to enable correct nucleosome binding. On another note, ASXL1^{DEU} does not seem to be involved in nucleosome binding, as a shorter ASXL1 fragment, ASXL1(1–330)—which lacks the entire C-terminal helix that is closest to NCP within ASXL1—did not show any appreciable difference in nucleosome binding (Extended Data Fig. 5g).

Targeting the H2A–H2B interface

Almost diametrically opposite to the CTE-finger-binding site on the nucleosomal surface, the β 2– α 2 loop in the BAP1 UCH domain is

situated next to the H2A–H2B acidic patch (Fig. 1f). The acidic patch is a common protein–nucleosome interaction site²⁹, but a previous report found that it had no obvious involvement in binding the BAP1–ASXL1 complex²³. The density for the loop segment encompassing residues 55–70 of BAP1 is relatively weak, but sufficient to locate a highly conserved positively charged sequence motif, RRSRR (spanning residues 56–60), in the vicinity of the H2A–H2B acidic patch (Fig. 3c and Extended Data Fig. 5b). Owing to the ambiguity of side-chain positions, we mutated all five residues to alanine (56–60 mut) and found that this change significantly compromised the H2AK119ub1 DUB activity (Fig. 3d), although no notable difference in nucleosome binding was detected (Extended Data Fig. 5g). As with the CTE mutations, the β 2– α 2 loop mutation did not affect the deubiquitination of the free histone H2AK119ub1 substrate (Extended Data Fig. 5h), providing evidence of the importance of the RRSRR motif for nucleosomal H2AK119ub1 deubiquitination that involves the H2A–H2B acidic patch.

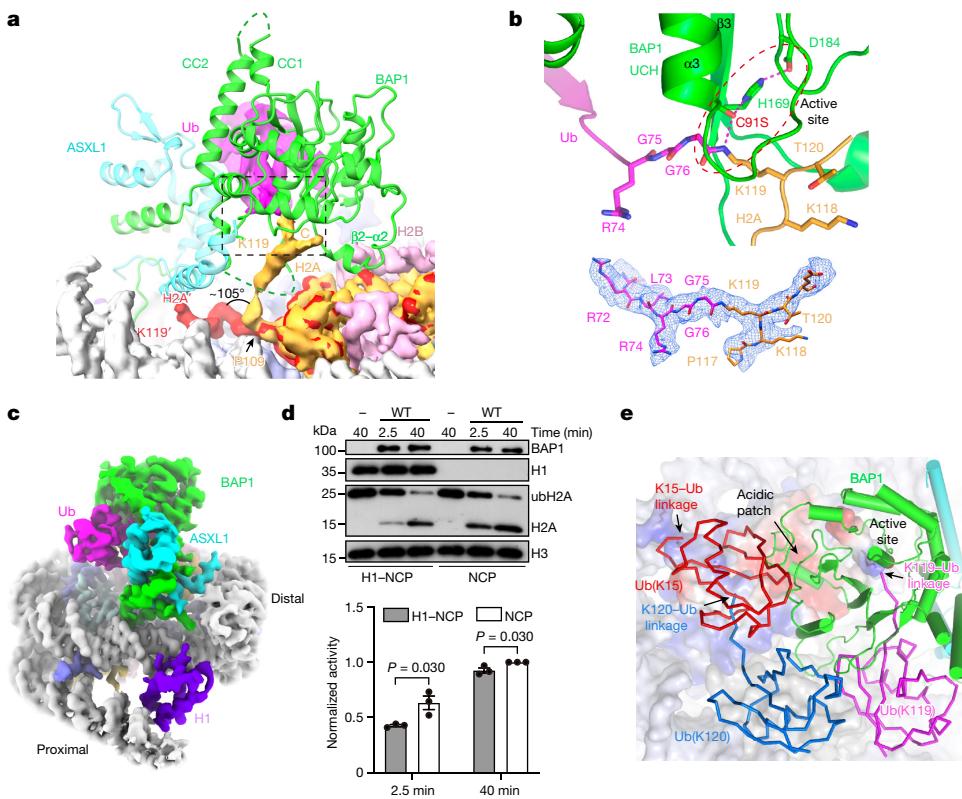


Fig. 4 | Conformational change of the H2A C-terminal tail and the role of H1 in PR-DUB activity. **a**, Comparison of the two H2A C-terminal tails from the same NCP reveals a large conformational difference. The density map of H2A in the BAP1-ASXL1-bound side is shown in yellow-orange and that in the opposite side (H2A') is shown in red. **b**, Top, magnified view of the H2AK119-Ub isopeptide bond next to the BAP1 catalytic triad (C91, H169 and D184). Bottom, the composite density map, contoured at 4σ , showing the covalent linkage of H2AK119 and ubiquitin. **c**, Head-on view of H1 binding near the nucleosome dyad. Density regions for H1, NCP, BAP1, ASXL1 and ubiquitin are coloured in purple, grey, green, cyan and magenta, respectively. Linker-DNA arms near

(proximal) and far from (distal) H1 are labelled. **d**, Western blot detection (top) and quantification (bottom) comparing the deubiquitinase activity with NCP and chromatosome substrates ($n = 3$ biologically independent replicates, mean \pm s.e.m.). The H3 lanes are sample-processing controls. Statistical testing was performed in the same way as in Fig. 2f. **e**, Distinct spatial locations of H2AK15ub (red; Protein Data Bank (PDB) ID: 7E8I), H2BK120ub (blue; PDB ID: 4ZUX) and H2AK119ub (magenta), and the histone–ubiquitin isopeptide bonds relative to the NCP-bound BAP1-ASXL1. The NCP is shown in a surface representation, with H2A and H2B coloured according to electrostatic potential.

BAP1 mutations alter gene regulation

Consistent with the in vitro results, knockout (KO) of *Bap1* in mouse embryonic stem (ES) cells yielded an increased level of H2AK119ub1. Western blot analyses showed that transfecting the *Bap1* KO cells with mammalian expression vectors carrying coding sequences for the BAP1(C91S), BAP1(56–60 mut), BAP1(699–702 mut) and BAP1(1–696) mutants all resulted in higher levels of H2AK119ub1 as compared to transfection of the wild-type BAP1 plasmid (Fig. 3e). RNA-seq analysis identified several hundred up- and downregulated genes in *Bap1* KO cells compared to wild-type ES cells, in agreement with previous data^{7,8} (Extended Data Fig. 6). Gene expression analyses of PR-DUB-regulated genes by RNA-seq revealed similar effects to those that were observed in the western blot analyses, with the CTE-finger alanine-substitution mutant showing partial rescues, consistent with the cumulative effect of the CTE constituents (Figs. 2d and 3e,f and Extended Data Fig. 7a,b).

Again consistent with the in vitro results (Fig. 2f), the R699W and R701C cancer-associated mutations in the BAP1 CTE led to mild reductions of DUB activity in cells (Extended Data Fig. 8a), and RNA-seq analysis revealed corresponding changes of BAP-regulated genes (Extended Data Fig. 8b–d). Three non-CTE cancer-related mutations—H169Q in the catalytic active site, R56C in the β_2 - α_2 helix (Fig. 3c) and Y173C on β_3 , which is involved in stabilizing α_8 (Fig. 1f)—also affected H2AK119 DUB activity to varying degrees both in vitro and in cells (Fig. 3g,h and

Extended Data Fig. 9a), with the catalytic H169Q mutation showing the most significant defect, and resulted in corresponding gene expression changes (Extended Data Fig. 9b–d).

In addition, frameshift and premature truncation mutations of ASXL1 are frequently associated with myeloid leukaemias and Bohring–Opitz syndrome. Ectopic expression of these mutants resulted in reduced levels of H2AK119ub1, leading to the hypothesis that ASXL1's C-terminal region, which is absent in these mutants, might have an autoinhibitory activity²². We directly tested this hypothesis by fusing two C-terminal segments, one spanning residues 1401–1541 and the other residues 1306–1541, to the N-terminal 378 residues of ASXL1. Both fusion proteins contain N- and C-terminal domains that are conserved in ASXL1–ASXL3, and their complexes with BAP1 showed in vitro DUB activities that were indistinguishable from that attained with the ASXL1 N-terminal fragment only (Extended Data Fig. 10a,b), suggesting that there are alternative mechanisms for the elevated DUB activity that is associated with these ASXL1 cancer mutations.

The H2A tail in the active site of BAP1

The covalent linkage between the lysine side chain and ubiquitin is flexible and has not (to our knowledge) previously been observed in protein–nucleosome complexes with ubiquitinated histones^{30,31}. Our 3.7-Å-resolution unsharpened global refinement map shows that the C-terminal tail of H2A, starting from Pro109, turned by around 105°

toward the active site of BAP1 (Fig. 4a and Extended Data Fig. 2c), with the C α of Lys119 rising by around 26 Å above the nucleosomal surface from the C α of Pro109. By contrast, the H2A tail on the opposite side of the nucleosome, shown as H2A' in the superimposed figure, remains in the same conformation as that in the free NCP (Fig. 4a). H2A C-terminal residues beyond Lys118, as observed in the high-resolution NCP structure³², will clash with the BAP1 CTE, suggesting that nucleosome binding of BAP1–ASXL1 displaces the C-terminal tail of H2A in the NCP structure, and that this enables the active deubiquitination of H2AK119ub1 (Fig. 4a). It is worth noting that H2AK118 may be monoubiquitinated inside the cell, although it appears that only one of the two consecutive lysines—predominantly Lys119 (K119)—is modified at a time^{33,34}. The structure shows that the C-terminal tail of H2A may be further extended to place an ubiquitinated H2AK118 at the active site for deubiquitination (Fig. 4b). Finally, the 3.9-Å-resolution locally refined map clearly shows the placement of the isopeptide bond between the C terminus of ubiquitin and the ε amino group of H2AK119, next to the catalytically inactivated C91S residue of BAP1 (Fig. 4b), thus providing a rare glimpse into the geometric arrangement of the scissile isopeptide bond and the active-site residues of hydrolases of the UCH family. No interaction is observed between BAP1 and the side chains of other H2A residues that neighbour Lys119, supporting the notion that the H2AK119 specificity of PR-DUB results from the particular spatial configuration of nucleosome binding, and not from its intrinsic preference for the amino acid sequence of H2A.

Functional roles of the linker histone

H1 is present in both euchromatic and heterochromatic regions, and is generally associated with gene repression³⁵. It can coexist with H2AK119ub1 in the same nucleosome³⁶, and we showed that PR-DUB is active toward a chromatosome substrate (Extended Data Fig. 1c). In our structure, the globular domain of H1 is visualized bound to nucleosomal DNA in an on-dyad mode³⁷ (Fig. 4c). It is situated closer to the proximal arm of the linker DNA than to the distal arm, and there is no direct contact with the PR-DUB complex. However, an H2AK119ub1 chromatosomal substrate exhibited a modestly reduced DUB activity (Fig. 4d). This difference might be attributable to the asymmetrical contact between the linker H1 histone and the DNA, which appears to limit the binding of PR-DUB to one side of the nucleosomal surface owing to steric constraint (Extended Data Fig. 10c). Without the linker histone, independent binding of PR-DUB to either side of the nucleosomal surface seems to be possible, although only at a low percentage (Extended Data Fig. 4c,f). Nevertheless, our demonstration that PR-DUB can deubiquitinate nucleosomal and chromatosomal substrates alike suggests that H2AK119ub1 and H1 can separately modulate gene expression. They might also work synergistically, as indicated by the observation that deubiquitination of H2AK119 promotes the dissociation of H1 from the nucleosome³⁸.

Conclusions

Our study reveals that the distinct nucleosome-binding mode of PR-DUB displaces the H2A C-terminal tail on the nucleosome surface and allows H2AK119ub1 access to the enzymatic active site of BAP1. In this mode of binding, the isopeptide bond between ubiquitin and other nucleosomal ubiquitination sites—such as H2BK120 and H2AK13 or H2AK15—cannot reach the active site of PR-DUB owing to distance and geometric constraints (Fig. 4e), thus specifying the physiologically relevant H2AK119ub1 specificity of PR-DUB. Consistent with the structural findings, mutations of BAP1 residues that are involved in catalysis or NCP binding, including cancer-associated mutations, all resulted in the reduction of DUB activity and the alteration of gene expression, to varying degrees. The DEUBAD domain of ASXL1 has at least two roles in H2AK119 deubiquitination: formation of the

ubiquitin-binding cleft together with the UCH domain of BAP1; and securing the CTE finger for interaction with the nucleosome, through contact with the αC of BAP1. Frequent cancer-associated truncations of ASXL1–ASXL3 retain an intact DEUBAD domain, and ectopic expression of ASXL1 truncation mutants resulted in substantially reduced levels of cellular H2AK119ub1, indicative of an autoinhibitory effect of ASXL1's C-terminal domain. Our biochemical analysis suggests that further effects of the C-terminal domains of ASXL proteins might be important for tumorigenesis.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-05841-y>.

1. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708 (2009).
2. Schuetzengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
3. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).
4. Scheuermann, J. C. et al. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243–247 (2010).
5. Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
6. Campagne, A. et al. BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat. Commun.* **10**, 348 (2019).
7. Conway, E. et al. BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. *Mol. Cell* **81**, 3526–3541 (2021).
8. Fursova, N. A. et al. BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. *Genes Dev.* **35**, 749–770 (2021).
9. Bonnet, J. et al. PR-DUB preserves Polycomb repression by preventing excessive accumulation of H2Aub1, an antagonist of chromatin compaction. *Genes Dev.* **36**, 1046–1061 (2022).
10. Beck, D. B. et al. Chromatin in the nuclear landscape. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 11–22 (2010).
11. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–2534 (2013).
12. Chen, Z., Djekidel, M. N. & Zhang, Y. Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat. Genet.* **53**, 551–563 (2021).
13. Parreño, V., Martínez, A. M. & Cavalli, G. Mechanisms of Polycomb group protein function in cancer. *Cell Res.* **32**, 231–253 (2022).
14. Gaytán de Ayala Alonso, A. et al. A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics* **176**, 2099–2108 (2007).
15. Harbour, J. W. et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410–1413 (2010).
16. Pena-Llopis, S. et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–579 (2012).
17. Dey, A. et al. Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541–1546 (2012).
18. Carbone, M. et al. BAP1 and cancer. *Nat. Rev. Cancer* **13**, 153–159 (2013).
19. Katoh, M. Functional and cancer genomics of ASXL family members. *Br. J. Cancer* **109**, 299–306 (2013).
20. Wang, L., Birch, N. W., Zhao, Z., Nestler, C. M. & Shilatifard, A. Epigenetic targeted therapy of stabilized BAP1 in ASXL1 gain-of-function mutated leukemia. *Nat. Cancer* **2**, 515–526 (2021).
21. Abdel-Wahab, O. et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180–193 (2012).
22. Balasubramani, A. et al. Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1–BAP1 complex. *Nat. Commun.* **6**, 7307 (2015).
23. Sahtoe, D. D., van Dijk, W. J., Elkебus, R., Ovaa, H. & Sixma, T. K. BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nat. Commun.* **7**, 10292 (2016).
24. De, I. et al. Structural basis for the activation of the deubiquitinase Calypso by the Polycomb protein ASX. *Structure* **27**, 528–536 (2019).
25. Foglizzo, M. et al. A bidentate Polycomb repressive-deubiquitinase complex is required for efficient activity on nucleosomes. *Nat. Commun.* **9**, 3932 (2018).
26. Sahtoe, D. D. et al. Mechanism of UCH-L5 activation and inhibition by DEUBAD domains in RPN13 and INO80G. *Mol. Cell* **57**, 887–900 (2015).
27. Vander Linden, R. T. et al. Structural basis for the activation and inhibition of the UCH37 deubiquitinase. *Mol. Cell* **57**, 901–911 (2015).
28. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
29. McGinty, R. K. & Tan, S. Principles of nucleosome recognition by chromatin factors and enzymes. *Curr. Opin. Struct. Biol.* **71**, 16–26 (2021).
30. Morgan, M. T. et al. Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**, 725–728 (2016).
31. Wang, H. et al. Structure of the transcription coactivator SAGA. *Nature* **577**, 717–720 (2020).

32. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
33. Lan, L. et al. Monoubiquitinated histone H2A destabilizes photolesion-containing nucleosomes with concomitant release of UV-damaged DNA-binding protein E3 ligase. *J. Biol. Chem.* **287**, 12036–12049 (2012).
34. Nakata, S. et al. The dynamics of histone H2A ubiquitination in HeLa cells exposed to rapamycin, ethanol, hydroxyurea, ER stress, heat shock and DNA damage. *Biochem. Biophys. Res. Commun.* **472**, 46–52 (2016).
35. Zlatanova, J. & Vanholde, K. Histone H1 and transcription—still an enigma. *J. Cell Sci.* **103**, 889–895 (1992).
36. Jason, L. J. M., Finn, R. M., Lindsey, G. & Ausio, J. Histone H2A ubiquitination does not preclude histone H1 binding, but it facilitates its association with the nucleosome. *J. Biol. Chem.* **280**, 4975–4982 (2005).
37. Zhou, B. R. et al. Distinct structures and dynamics of chromatosomes with different human linker histone isoforms. *Mol. Cell* **81**, 166–182 (2021).
38. Zhu, P. et al. A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. *Mol. Cell* **27**, 609–621 (2007).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Methods

Plasmid construction

Full-length cDNA of human BAP1 was amplified by PCR, and bacterial expression plasmids for full-length BAP1 and BAP1 fragments encompassing residues 1–696, 1–710, 1–718 and an internal deletion of residues 711–718 (Δ 711–718) were cloned into a modified pRSFDuet-1 vector (Novagen), termed pRSFDuet-Smt, in which a 6×His-SUMO cassette was inserted between the NcoI and BamHI sites. A plasmid with the coding sequence for ASXL1 residues 1–480 was purchased from Sangon Biotech, and bacterial expressing plasmids for ASXL1 fragments spanning residues 1–330 or 1–378 were cloned into the pGEX-6P-1 vector (GE Healthcare). Plasmids for the expression of C-terminal glutathione S-transferase (GST)-fused ASXL1(1–378)-(GGGGS)₂-(1401–1541) and ASXL1(1–378)-(GGGGS)₂-(1306–1541) fragments, termed ASXL1-C1 and ASXL1-C2, respectively, were constructed by linking the C-terminal pieces, which were amplified from human cDNA, with the N-terminal region by PCR. Expression plasmids for BAP1 and ASXL1 mutants were generated by site-directed mutagenesis (Supplementary Table 1).

A bacterial expression plasmid for human H2B was generated by cloning the cDNA into pET30a (Novagen), and a co-expression plasmid for human H3 and H4 was obtained by inserting their cDNAs into the same pCDFDuet-1 vector (Novagen). A plasmid expressing C-terminal 6×His-tagged full-length human H1.4 was constructed using a pET-28a vector (Novagen).

DNA fragments of 187 bp for nucleosome assembly were generated from a plasmid with four tandem copies of the 187-bp fragment cloned into a pBluescript vector (Addgene). Each DNA fragment contains a 147-bp Widom 601 positioning sequence flanked by a 17-bp and a 23-bp linker-DNA sequence³⁹ (Supplementary Table 1). An EcoRV restriction site is engineered at both ends of the DNA for isolation of the 187-bp fragment.

Expression and purification of BAP1–ASXL1 complexes

Wild-type or mutant BAP1 and ASXL1(1–378) proteins were co-expressed in *Escherichia coli* BL21(DE3) RIL cells. The plasmids were co-transformed into the *E. coli* cells, then a single colony was picked and cultured overnight in 150 ml LB medium containing 50 μ g ml⁻¹ kanamycin, 100 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol at 37 °C. The start-up culture was then used to inoculate 5-l or 2.4-l cultures, and protein production was induced with the addition of 0.25 mM isopropylthio-β-galactoside (IPTG) when the cell density reached optical density at 600 nm ($OD_{600\text{nm}}$) values between 0.6 and 0.8, followed by shifting the temperature to 16 °C and continuing shaking the culture for 17 h. Cells were then collected by centrifugation, resuspended in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole and 1 mM phenylmethylsulfonyl fluoride (PMSF), and lysed by pressure rupture followed by ultrasonication. Cell debris was then cleared by centrifugation, and the supernatant was incubated with Ni-excel affinity resins (GE Healthcare), followed by washing the resins first with the lysis buffer, and subsequently with 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 25 mM imidazole, and the bound proteins were eluted in this buffer by adjusting the imidazole concentration to 200 mM.

For purification of the BAP1(C91S)–ASXL1(1–378) complex, GST and His-SUMO tags were cleaved after the Ni-excel step by the addition of PreScission and SUMO proteases during overnight dialysis in buffer A-100 (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM dithiothreitol (DTT)). The dialysed sample was then loaded onto a HiTrap SP cation exchange column (GE Healthcare) pre-equilibrated with buffer A-100, and a linear gradient of 0–70% buffer A-1000, which differs from A-100 in NaCl concentration (100 mM to 1 M), was developed to elute the bound proteins. The eluted fractions were analysed by SDS-PAGE, and the ones enriched with the BAP1(C91S)–ASXL1(1–378) complex were pooled and concentrated, before loading onto a Superdex 200 HiLoad 16/60

sizing column in the A-500 buffer. The second peak from the sizing column contains the protein complex of interest, and the fractions were pooled and concentrated for complex formation with the nucleosome.

Wild-type and mutant BAP1–ASXL1(1–378) complexes used in fluorescence polarization experiments and H2A and H2B deubiquitinase assays were purified following similar procedures to that outlined for the BAP1(C91S)–ASXL1(1–378) complex above, except that the size-exclusion step was performed with a Superdex 200 10/300 GL in the A-100 buffer.

For the nucleosomal H2AK119ub1 deubiquitination assay, mutant BAP1–ASXL1 complexes were first purified through Ni-excel resins, followed by cleavage of the GST tag using PreScission protease, leaving the His-SUMO on BAP1. The GST tag was removed by applying the sample through an SP column. The target protein complex was eluted with a linear gradient of 0–70% buffer A-1000. The BAP1 complexes with ASXL1-C1 and ASXL1-C2 fragments were successively purified through Ni-excel and glutathione (GSH) sepharose (GE Healthcare) resins. The target protein was finally eluted by 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 30 mM GSH. These proteins were then flash-frozen in liquid nitrogen and stored at –80 °C before use. All of the purification steps were performed at 4 °C.

Preparation of histones, nucleosomal DNA and nucleosomes

Chemically synthesized full-length human H2A monoubiquitinated at K119 (H2AK119ub1) and H2B monoubiquitinated at K120 (H2BK120ub1) were purchased from Hefei KS-V Peptide Biotechnology.

Human histones H2B and H2A were expressed in an insoluble form in *E. coli*, and purified following a method described before⁴⁰. Human H3 and H4 were co-expressed in soluble forms in the BL21 (DE3) RIL strain of *E. coli*⁴¹, and purified through successive HiTrap Heparin and Superdex 200 HiLoad 16/60 (GE Healthcare) column chromatography steps. Purified H3–H4 tetramers in 20 mM Tris-HCl pH 7.5, 2 M NaCl, 1 mM EDTA and 0.1% β-mercaptoethanol were concentrated and flash-frozen in liquid nitrogen and stored at –80 °C.

Linker histone H1 was expressed in the BL21 (DE3) Rosetta strain of *E. coli*. Collected cells were resuspended in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 6 M guanidine-HCl and 1 mM PMSF, followed by cell rupture by pressure and ultrasonication. Cellular debris was cleared by centrifugation, and the supernatant was loaded onto a Ni-excel column. Then, the loaded column was washed with buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% glycerol and 7 M urea, first without imidazole, then with the addition of 15 mM imidazole, and finally with 200 mM imidazole to elute the bound H1. The eluted sample was dialysed against 20 mM Tris-HCl pH 7.5, 500 mM NaCl and 1 mM EDTA three times to remove urea. Then, the sample was purified through a HiTrap Heparin HP cation exchange column (GE Healthcare). Highly purified H1 fractions were pooled and flash-frozen in liquid nitrogen and stored at –80 °C.

The plasmid carrying the tandem 187-bp nucleosomal DNA fragment was produced using the DH5α strain of *E. coli* in cell culture with TB medium. Large-scale plasmid purification was performed following a published procedure⁴². The 187-bp fragment was excised from the plasmid by EcoRV digestion, isolated by polyethylene glycol (PEG) precipitation and further purified through ethanol precipitation.

The 5'-FAM-labelled 167-bp DNA fragment used for the assembly of nucleosomes in fluorescent polarization experiments was generated by PCR using a 5'-FAM-labelled primer. The sequence of the 167-bp fragment, which contains a centrally located 147-bp Widom sequence, is shown in Supplementary Table 1. The PCR product was purified using the Omega E.Z.N.A. gel extraction kit in the dark.

The assembly of histone octamers and reconstitution of NCPs by salt dialysis were performed according to a published protocol⁴⁰, and the assembly of the H1-containing chromatosome was performed according to a previously described procedure⁴³. In brief, H1 was added to the NCP dialysis mixture at a molar ratio of 0.8:1 when the salt concentration

Article

reached 600 mM, followed by further dialysis at 600 mM NaCl for 3 h, and then against the HE buffer (10 mM HEPES pH 8.0 and 1 mM EDTA) for 4 h.

Assembly of the BAP1–ASXL1–NCP complexes

For assembly of the BAP1(C91S)–ASXL1(1–378)–H1–NCP complex, purified BAP1–ASXL1 and H1–NCP complexes were separately dialysed into 20 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM DTT, and precipitants were removed by centrifugation. Then, BAP1–ASXL1 and H1–NCP complexes were mixed at a 3:1 molar ratio and incubated on ice for 30 min.

For the BAP1(C91S)–ASXL1(1–378)–NCP complex, purified BAP1–ASXL1 and NCP were first dialysed into 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA and 1 mM DTT separately. The BAP1–ASXL1 complex was then added in a threefold molar excess to NCP, and the mixture was dialysed into 20 mM HEPES pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM DTT.

The above samples were then subjected to GraFix (ref. 44). One hundred microlitres of sample was loaded to the 5-ml tubes with a 10–40% linear glycerol gradient and a 0–0.15% linear glutaraldehyde gradient in 20 mM HEPES pH 8.0 and 50 mM NaCl, followed by centrifugation at 30,000 rpm for 17 h using a Hitachi SW50 rotor at 4 °C. After centrifugation, the sample was separated into 25 fractions, 200 µl each, from top to bottom, and the fractions were examined by 4% native PAGE and negative-staining EM. Fractions with homogeneous, properly sized particles were picked and subjected to glycerol removal by dialysis and concentration by centrifugation.

Cryo-EM sample preparation and data collection

Quantifoil R2/1300 mesh grids were glow-discharged for 1 min using a Gatan Plasma System. Three microlitres of BAP1–ASXL1–NCP sample, with or without H1, was placed onto the grid, which was plunge-frozen in liquid ethane by using a Vitrobot (Thermo Fisher Scientific) with settings at 10 °C, 100% humidity, 10 s incubation time and 3.5 s blot time.

Automated data acquisition was performed using a modified SerialEM (ref. 45) script to perform beam-image shift data collection⁴⁶. Images for the histone-H1-containing complex were collected on a Talos Arctica 200 kV machine (Thermo Fisher Scientific) equipped with a K2 Summit camera (Gatan) in super-resolution mode at a nominal magnification of 130,000 and a pixel size of 0.5 Å, with a defocus range from −1.5 µm to −2.0 µm. The exposure time was 5.2 s, with a total exposure dose of 50 electrons per Å² over 32 frames. Images for the nucleosome complex were collected using a Titan Krios 300 kV machine (FEI) equipped with the same type of camera, and a dosage of 60 electrons per Å² was used for data collection. A total of 21,466 micrographs for the BAP1(C91S)–ASXL1(1–378)–H1–NCP sample and 6,269 micrographs for the BAP1(C91S)–ASXL1(1–378)–NCP sample were collected (Extended Data Table 1).

Image processing and 3D reconstruction

For the BAP1(C91S)–ASXL1(1–378)–H1–NCP complex, all micrographs were motion-corrected and dose-weighted with patch-based schemes in cryoSPARC (ref. 47), and contrast transfer function (CTF) parameters were estimated with the patch-based method in cryoSPARC. Micrographs with poor Thon rings were discarded, and around 100 micrographs were randomly selected for reference-free particle picking using cryoSPARC blob picker. The picked particles were used for an initial round of two-dimensional (2D) classification, and 2D class averages with clear features were selected as templates for particle auto-picking of the whole dataset. A total of 12,435,535 particles were auto-picked and extracted with twofold binning using a box size of 256 pixels. After several rounds of reference-free 2D classification, a total of 4,818,641 particles were selected for cryoSPARC ab initio reconstruction. After converting the cryoSPARC particle location file to a star file using the csparc2star.py script in pyem (ref. 48), several

rounds of global three-dimensional (3D) classification and one round of no-alignment 3D classification in RELION 3.0 were performed⁴⁹. A total of 443,790 particles were selected and re-extracted without binning using a box size of 352 pixels and subjected to non-uniform refinement, yielding a cryo-EM density map at an overall resolution of 2.75 Å. Multibody refinement analysis revealed that the BAP1–ASXL1 and ubiquitin module is placed, with respect to NCP, relatively dynamically (data not shown). To improve the quality of the density for this module, the density belonging to the NCP and H1 was subtracted from the 2D images and these images were subjected to two rounds of no-alignment 3D classification with a soft-edged mask covering the BAP1–ASXL1–Ub region. A total of 95,225 particles in classes with obviously improved densities for the BAP1–ASXL1–Ub module were selected and re-extracted using a box size of 352 pixels for local refinement with the soft-edged mask, which resulted in a much-improved density at a resolution of 3.9 Å. Meanwhile, global refinement using these particles without density subtraction yielded a 3.0-Å-resolution density map for the entire complex, as evidenced in a region near the catalytic active site of BAP1. The map was post-processed in deepEM-hancer to improve interpretability⁵⁰. All resolutions were estimated according to the gold-standard Fourier shell correlation criterion value of 0.143. The 2.75-Å and 3.0-Å global refinement maps and the 3.9-Å local refinement map were used to generate a composite map in PHENIX (ref. 51). A flow chart schematically summarizing the procedures of image processing and 3D reconstruction is shown in Extended Data Fig. 2.

For the BAP1(C91S)–ASXL1(1–378)–NCP complex without H1, image pre-processing steps similar to that described above were used. A total of 5,181,645 particles were picked, and 1,006,755 of these were kept after 2D classification. Then, global 3D classification, particle subtraction, 3D classification without alignment and particle re-extraction were performed and 56,370 particles were kept for the final refinements. Global refinement using these particles yielded a 4.2-Å-resolution density map, and the density was further improved by density modification in PHENIX (ref. 52). Meanwhile, local refinement using these 56,370 particles yielded a 4.4-Å-resolution density map in the region of the BAP1–ASXL1–Ub module. A flow chart schematically summarizing the procedures of image processing and 3D reconstruction is shown in Extended Data Fig. 4.

Model building, refinement and validation

An initial model of the BAP1(C91S)–ASXL1(1–378)–H1–NCP complex was built with the help of homologous substructures, including the cryo-EM structure of the H1-containing chromatosome (PDB ID: 7K5Y; ref. 37), and crystal structures of Caly–ASX (PDB ID, 6HGC) and UCH-L5–RPN13–Ub (PDB ID, 4UEL) complexes^{24,26}. UCSF Chimera was used for docking the homologous structures into the density, and model adjustment and editing were performed with Coot (ref. 53). De novo building of various regions without previous structural knowledge, including the BAP1 CTE region and the H2A tail surrounding the K119–Ub isopeptide bond, and multiple rounds of manual model adjustment and rebuilding, were performed in Coot. Subsequently, the model was subjected to real-space refinement using PHENIX and density-constrained stereochemistry refinement using Rosetta (ref. 54), and, finally, the model was validated with MolProbity (ref. 55). Statistics for model refinement and validation are shown in Extended Data Table 1. Cryo-EM density map and structural model figures were prepared using Chimera (ref.⁵⁶), ChimeraX (ref. 57), Coot and PyMOL (Schrödinger).

In vitro deubiquitination assay

Deubiquitination reactions with free histone H2BK120ub1 and H2AK119ub1 substrates were performed with 10 pmol of histone mixed with 1 pmol of the BAP1–ASXL1(1–378) complex in 20 µl buffer at 25 °C for defined durations; reactions with nucleosomal substrates carrying corresponding ubiquitinated histones were performed using 3.3 pmol

nucleosome and 3.3 pmol of the BAP1–ASXL1(1–378) complex in 20 µl buffer. Reactions were stopped by the addition of 10 µl SDS loading buffer.

Separate deubiquitination assays of nucleosomal H2AK119ub1 were performed with 3.2–3.5 pmol of NCP mixed with wild-type or mutant His-SUMO-tagged BAP1–ASXL1(1–378) complexes at enzyme-to-substrate molar ratios of 0.3:1 or 0.5:1 in 20-µl reactions. These reactions were allowed to proceed at 25 °C for a defined duration before being stopped by the addition of 20 µl SDS loading buffer. When comparing the activities of BAP1(WT)–ASXL1(1–378) and BAP1(C91S)–ASXL1(1–378) for chromatosomal H2AK119ub1, 4.4 pmol NCP was deubiquitinated by the enzyme complex in a 40-µl system at 25 °C for 20 min. The enzyme was twofold serially diluted from the double molar amount of substrate, and the reaction was terminated by 20 µl SDS loading buffer. The results were analysed by western blot with anti-His tag (Invitrogen, MA1-21315, 1:1,000 dilution), anti-GST (Beyotime, AF0174, 1:4,000 dilution), anti-H2A (Millipore, 07-146; Cell Signaling Technology, 12349s, 1:5,000 dilution), anti-H2B (Abcam, ab1790, 1:5,000 dilution), anti-H3 (Abcam, ab1791, 1:10,000 dilution) and anti-BAP1 (Cell Signaling Technology, 13271s, 1:1,000 dilution) antibodies.

All of the reactions above were performed in a buffer with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgCl₂ and 1 mM DTT, except for the reactions in Extended Data Fig. 8b, in which 150 mM NaCl was used and 0.1% NP40 was added. The uncropped and unprocessed scans are supplied in Supplementary Fig. 1.

Quantification of deubiquitination activity was based on scanned intensities of western blots from three biologically independent experiments: catalytic activity = [H2A]/([H2Aub] + [H2A]), and the value was then normalized with that of the wild type in the same blot. Bar charts represent plots of mean ± s.e.m., and *P* values denote the results of an unpaired two-tailed *t*-test.

Fluorescence polarization assay

The of 5'-FAM-labelled 167-bp NCP (100 nM) was mixed with twofold serially diluted samples of BAP1–ASXL1 in buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM DTT. The mixtures were incubated at 4 °C for 30–60 min before being measured with an Envision multimode plate reader (Perkin Elmer). The mP values after subtracting background were fitted with GraphPad Prism v.9 using the specific binding model with Hill slope: normalized polarization value = (measured mP value – mP value of background)/mP_{max}. The calculated K_d shows symmetrical confidence intervals with a 95% confidence level. Measurements were taken from three distinct samples and each sample was measured three times.

MALS

One hundred microlitres of purified BAP1(C91S)–ASXL1(1–378) complex at a concentration of around 2 mg ml⁻¹ was injected into the Superdex 200 Increase 10/300 GL column (GE Healthcare) at a flow rate of 0.4 ml per min in a buffer of 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 1 mM EDTA. The chromatography system was coupled with the 18-angle light-scattering detector (DAWN HELEOS II, Wyatt Technology) and differential refractive index detector (Optilab rEX, Wyatt Technology), and the data were analysed by the ASTRA software.

Cell lines and cell culture

R1 mouse ES cells were grown on 0.1% gelatin-coated dishes in 2i/LIF-containing DMEM medium (Gibco) supplemented with 15% fetal bovine serum (Hyclone), 2 mM glutamine (Sangon Biotech), 100 U ml⁻¹ penicillin, 0.1 mg ml⁻¹ streptomycin (Sangon Biotech), 0.1 mM non-essential amino acids (Sigma), 1 mM sodium pyruvate (Gibco), 50 mM β-mercaptoethanol and 1,000 U ml⁻¹ leukaemia inhibitory factor (Novoprotein), and GSK3b and MEK1/2 inhibitors (Selleckchem) to final concentrations of 3 µM and 1 µM, respectively.

The stable *Bap1* KO cell line was generated by CRISPR–Cas9. sgRNA sequences were as follows: sgRNA1: ACCGAAGTCGTGGTAACG; sgRNA2: AGGGCGAGAGCGTTCCGC.

Bap1 rescue cells were generated by the PiggyBac system. *Bap1* KO cells were mock-transfected or transfected with an EF1α-driven plasmid encoding wild-type BFP-P2A-HA-BAP1 or BAP1 truncations or mutations, and another vector carrying the transposase using Lipofectamine 3000 (Thermo Fisher Scientific), according to the manufacturer's instructions. BFP-positive cells were selected by fluorescence-activated cell sorting (FACS) 24 h after transfection. After further culturing for about seven days, BFP-positive cells within a specific BFP signal range were collected as stable BAP1 rescue cells. A representative FACS gating strategy is shown in Supplementary Fig. 2.

Cellular western blotting

For the analysis of BAP1 with total protein lysates, mouse ES cells were lysed in SDS lysis buffer (50 mM Tris-HCl pH 6.8, 2% SDS and 5 mM DTT) and boiled at 99 °C for 20 min. Lysates were centrifuged at 16,000g for 20 min, and the supernatant was collected for western blot analysis. For analysis of histone modifications, cells were washed with 1×PBS and resuspended in 0.4 M HCl overnight at 4 °C. Cells were then pelleted by centrifugation at 16,000g for 30 min at 4 °C, and the supernatant containing histones was precipitated on ice with 33% trichloroacetic acid for at least 30 min, followed by centrifugation at 16,000g for 20 min at 4 °C. The pellets were washed twice with ice-cold acetone and then dissolved in double-distilled water. The Bradford protein assay was used to determine protein concentration, and primary antibodies included: anti-BAP1 (Proteintech, 10398-1-AP; dilution, 1:500), anti-H2AK119ub1 (Cell Signaling Technology, 8240; dilution: 1:2,000), anti-H3 (Abcam, ab1791; dilution 1:80,000) and anti-GAPDH (ABclonal, AC001; dilution: 1:100,000).

The quantification of the levels of ubH2A in wild-type R1 and BAP1 mutant rescue cell lines was derived from scanned intensities of western blots. The grey values of ubH2A and H3 were collected by ImageJ v.2.1.0. H3 is a loading control. [ubH2A]/[H3] is used to represent the ubH2A level, and the value is normalized relative to that of the wild-type R1. The values were analysed using Microsoft Excel and plotted using GraphPad Prism v.9.0.0. Bar charts represent plots of mean ± s.e.m. (*n* = 4 biologically independent experiments). *P* values denote the results of an unpaired two-tailed *t*-test.

RNA-seq and data analysis

Total RNA was extracted from cultured cells using TRIzol reagent (Thermo Fisher Scientific), according to the manufacturer's instructions. The generation and sequencing of RNA-seq libraries of wild-type and *Bap1* KO mouse ES cells, and *Bap1* KO rescue cell lines (*Bap1* KO + BAP1(R699W); *Bap1* KO + BAP1(R701C); and *Bap1* KO + BAP1(699–702 mut)) were performed by Gene+; and those of the *Bap1* KO rescue cell lines (*Bap1* KO + empty vector; *Bap1* KO + wild-type BAP1; *Bap1* KO + BAP1(C91S); *Bap1* KO + BAP1(56–60 mut); *Bap1* KO + BAP1(699–702 mut); *Bap1* KO + BAP1(I–696); *Bap1* KO + BAP1(H169Q); *Bap1* KO + BAP1(R56C); and *Bap1* KO + BAP1(Y173C)) were performed by ANOROAD. All samples were collected in two replicates.

RNA-seq analysis was performed as previously described⁵⁸. After removing adapters and low-quality bases with a Phred quality score of less than 20 using Trimmomatic (v.0.36)⁵⁹, paired-end reads were aligned to the annotated mouse transcripts (mm10 Gencode vM15 release) using Hisat2 (v.2.1.0)⁶⁰. The expression levels of genes were calculated using StringTie (v.1.3.3b)⁶¹. Read counts were calculated using featureCounts (ref. 62). The R package edgeR (ref. 63) was used for identifying differentially expressed genes, and only those with a read count higher than 50 in at least one sample of a comparison were kept. A given gene was considered differentially expressed if the false discovery rate (FDR) was <0.05, *P* < 0.01 and fold change ≥ 2.

Statistics and reproducibility

Deubiquitination assays in Figs. 1b and 2i and Extended Data Fig. 5h were repeated three times independently with similar results. Consistent results were obtained for that of Extended Data Fig. 1c from three independent experiments under different conditions. Similar chromatography and electrophoresis results to those shown in Extended Data Fig. 1a,b were obtained in six separate protein purifications. The reconstitution of nucleosome or chromatosome experiments shown in Extended Data Fig. 1e were consistently repeated seven times. The GraFix results shown in Extended Data Fig. 1f were repeated three times independently, all having similar electrophoresis patterns.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) with accession numbers EMD-34431 and EMD-34432 for the chromatosomal and nucleosomal complexes, respectively. Constituent and consensus maps for the chromatosomal complex have been deposited with accession numbers EMD-35179, EMD-35180, EMD-35181 and EMD-35182. The atomic coordinates for the structure of PR-DUB bound to the chromatosome have been deposited in the PDB with the accession code 8H1T. The coordinates for the structures of *Drosophila* Caly in complex with DEUBAD of ASX (PDB ID: 6HGC), H2AK15-ubiquitinated NCP in complex with the BRCA1-BARD complex (PDB ID: 7E8I), the H2BK120-ubiquitinated NCP in complex with the SAGA DUB module (PDB ID: 4ZUX), the UCH-L5 complex with DEUBAD of RPN13 and ubiquitin (PDB ID: 4UEL) and the chromatosome with linker histone H1.4 (PDB ID: 7K5Y) were downloaded from the RCSB Protein Data Bank (<https://www.rcsb.org>). The RNA-seq data have been deposited in the Genome Sequence Archive at the National Genomics Data Center, China National Center for Bioinformation—Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA006013) and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. Two public RNA-seq datasets with the accession numbers GSE162739 and GSE161995 were used in this study. The list of differentially expressed genes (DEGs) defined in a previous study⁷ was downloaded from ref. 7 (<https://www.sciencedirect.com/science/article/pii/S1097276521005001?via%3Dihub#app2>), with the original file name of 1-s2.0-S1097276521005001-mmc3.xlsx. The list of DEGs defined in a previous study⁸ was downloaded from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161995>), with the original file name of GSE161995_BAPIff_cRNA_seq_DESeq2_Table_GEO.txt. A threshold of $P_{adj} < 0.05$ and fold change > 1.5 was used to extract the DEGs. Source data are provided with this paper.

39. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
40. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
41. Fang, Q. et al. Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome. *eLife* **5**, e11911 (2016).
42. Dyer, P. N. et al. in *Chromatin and Chromatin Remodeling Enzymes Part A* (eds Allis, C. D. & Wu, C.) 23–44 (2004).
43. Song, F. et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetraneulosomal units. *Science* **344**, 376–380 (2014).
44. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
45. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
46. Wu, C., Huang, X., Cheng, J., Zhu, D. & Zhang, X. High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. *J. Struct. Biol.* **208**, 107396 (2019).
47. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
48. Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. Zenodo <https://doi.org/10.5281/zenodo.3576630> (2019).
49. Zivanović, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
50. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
51. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
52. Terwilliger, T. C., Lutcke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
54. Wang, R. Y. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
55. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
56. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
58. Zhang, J. et al. Highly enriched BEND3 prevents the premature activation of bivalent genes during differentiation. *Science* **375**, 1053–1058 (2022).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
61. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
62. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Acknowledgements We thank B. Zhu, X. Huang, L. Chen, X. Li and other staff members at the Center for Biological Imaging of the Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS), for support in cryo-EM data collection; X. Yu at the Protein Science Core Facility of IBP for help with MALDI analyses; Y. Gu for participation at an early stage of the study; and J. Chen for logistic support. The study was supported in part by grants from the Chinese Ministry of Science and Technology (2019FA0508900), the Natural Science Foundation of China (31991162, 92153302 and 91853204), the K. C. Wong Education Foundation (GJTD-2020-06), the Youth Innovation Promotion Association (2018125) and the Strategic Priority Research Program (XDB37010100) of CAS.

Author contributions R.-M.X. & B.Z. designed and supervised the project. G.L. contributed experimental designs and advised on enzymatic activity assays. X.Z. advised on cryo-EM analysis. C.T. advised on the synthesis of ubiquitinated histones. W.G. performed protein expression and purification and cryo-EM sample preparation, analysed the structure and performed biochemical experiments. C.Y. prepared cryo-EM samples, collected data and solved and analysed the structures. J.L. analysed the cellular effects of BAP1 mutations. Z.Y. supervised the cryo-EM sample preparation and fluorescent polarization experiments. X.L. advised on enzymatic activity assays. Y.Z. performed bioinformatic analysis. C.-P.L. helped with structural model building. Y.L. contributed to the analysis of cellular effects of BAP1 mutations. W.G., C.Y. and R.-M.X. wrote the manuscript, and all authors read and edited the manuscript.

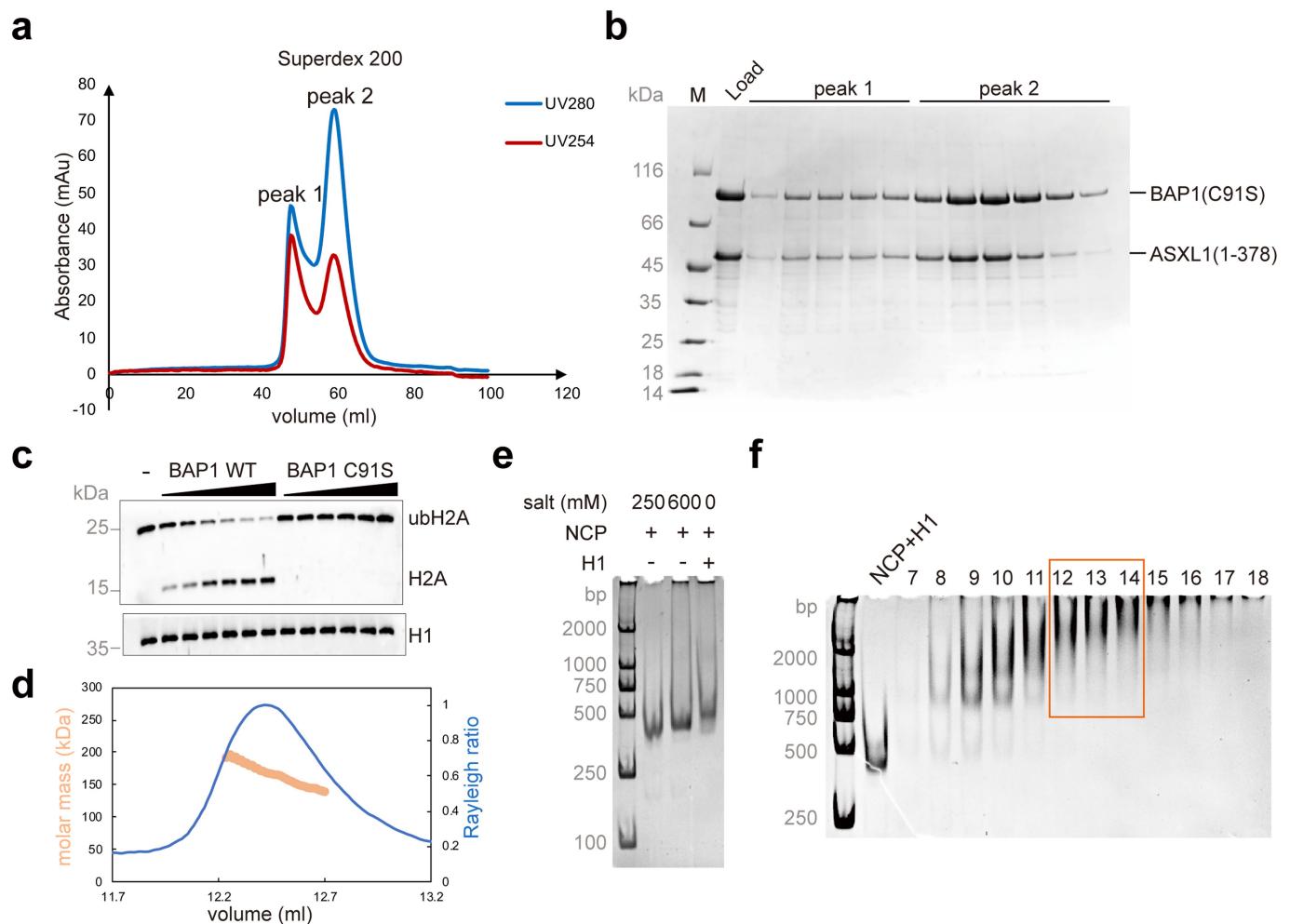
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-05841-y>.

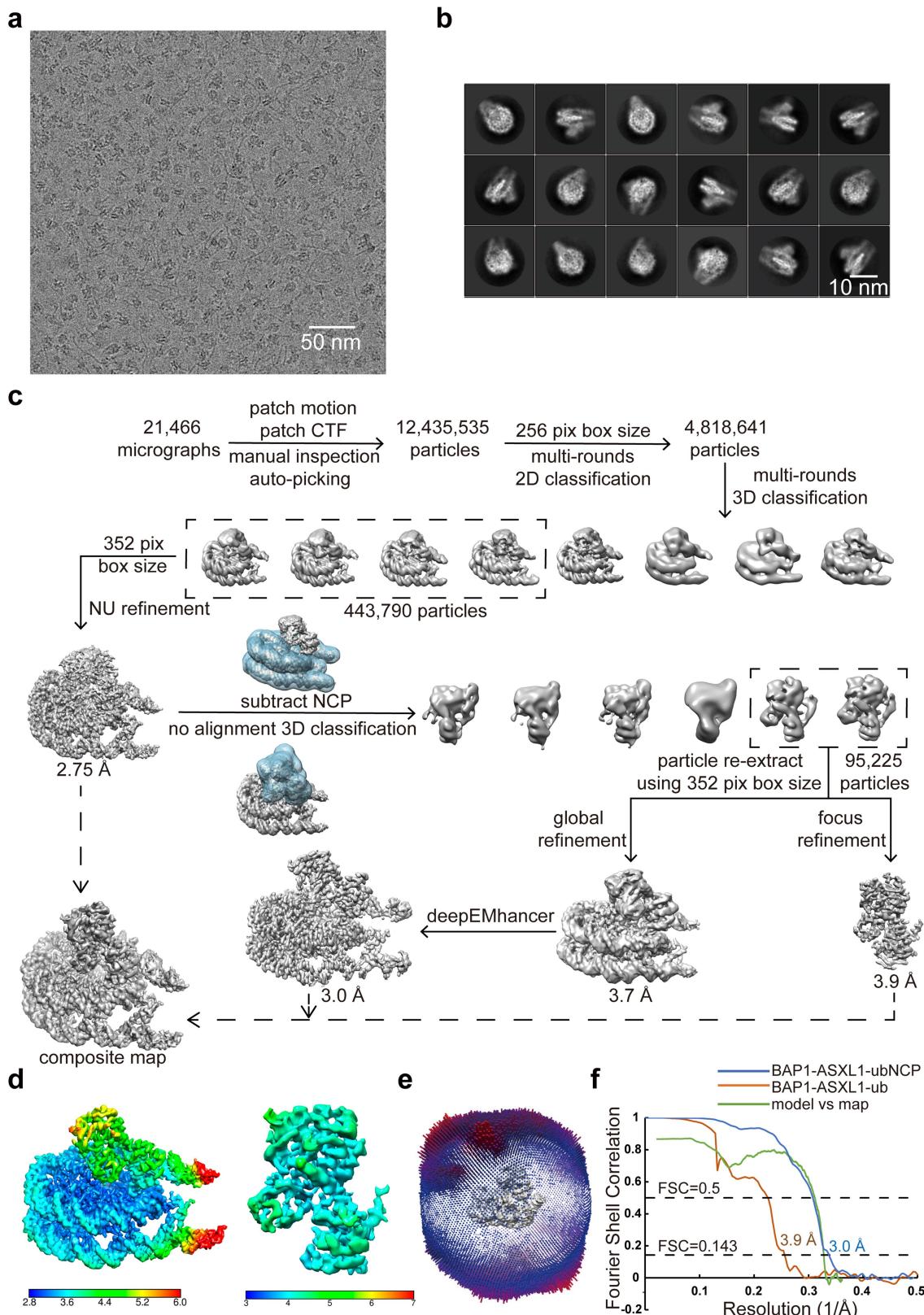
Correspondence and requests for materials should be addressed to Bing Zhu or Rui-Ming Xu. **Peer review information** *Nature* thanks Robert Klose and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



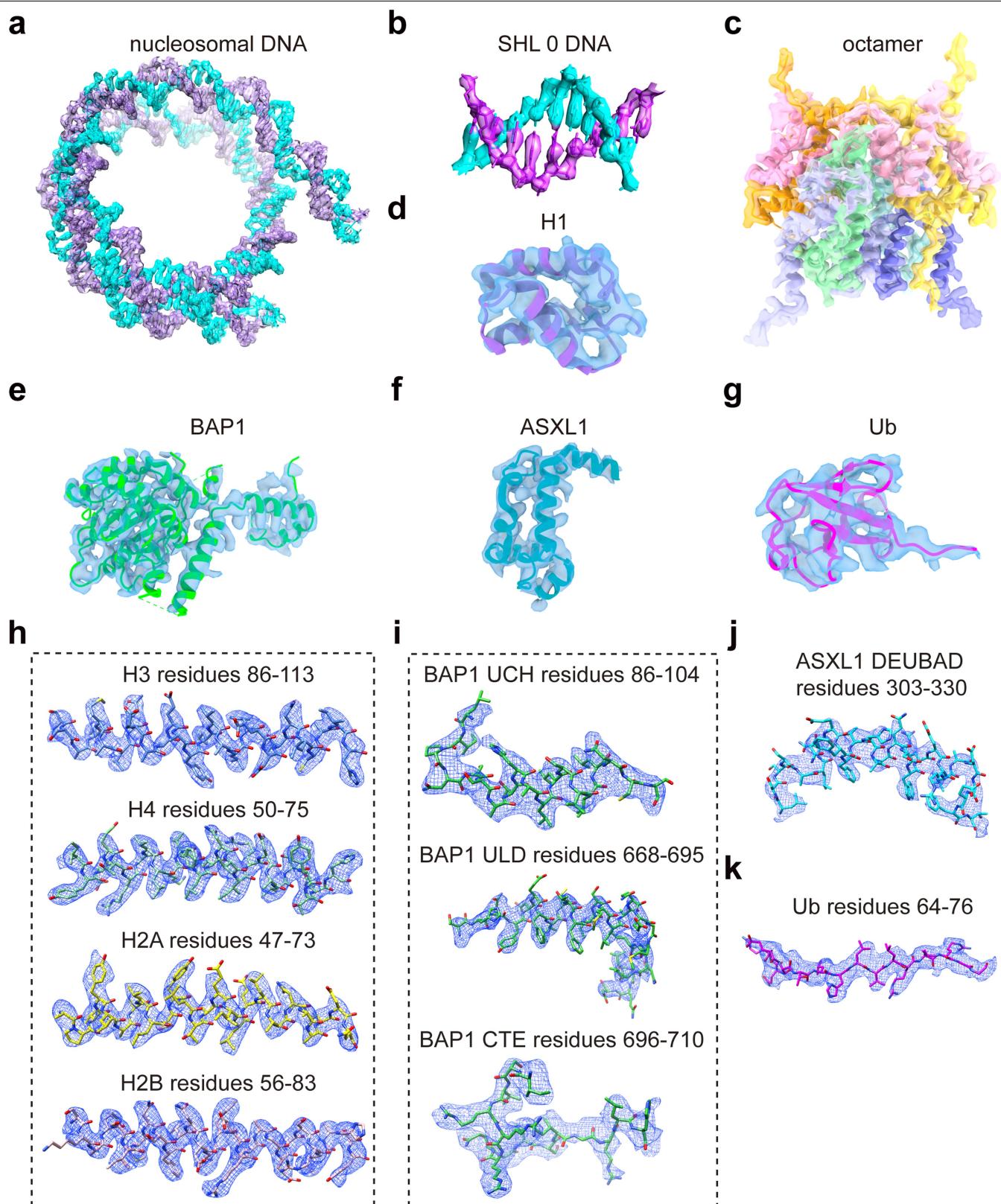
Extended Data Fig. 1 | Cryo-EM sample preparation. **a**, Elution profile of size-exclusion column chromatography (SEC) of the BAP1–ASXL1(1–378) complex. Peak 1 was eluted close to the void volume of the column. **b**, Coomassie-stained SDS–PAGE analysis of fractions from SEC shown in (a). M, molecular weight marker; values are labelled on the left. Peak 2 fractions were collected for assembling complex with NCP. **c**, Western blot analyses of DUB activities of the wild-type and the C91S mutant BAP1 complexes with ASXL1(1–378). Same amount of the 187-bp DNA H2AK119ub1 chromatosome substrate was used in each reaction. From left to right, lane 1, no enzyme added; lanes 2–7 and lanes 8–13, successive twofold dilutions of the wild-type and C91S BAP1 enzyme

complexes, respectively, detected with anti-H2A (top panel) and anti-His tag (bottom panel) antibodies. **d**, SEC-MALS analysis of BAP1C91S–ASXL1(1–378) complex at the concentration of 2.2 mg/ml. Calculated molar mass = 164±8.2 kDa, which is far below the 244.2 kDa calculated molecular mass of a dimer of the BAP1–ASXL1(1–378) complex. The data were analysed by the ASTRA software. **e**, GoldView-stained native PAGE shows the reconstituted NCP, NCP in a 600 mM NaCl buffer, at which point H1 was added and dialysed to zero salt to assemble the chromatosome. **f**, GoldView-stained native PAGE of fractions after GraFix. The chromatosome alone was used as a control. Fractions 12–14, highlighted in the red box, were used for cryo-EM sample preparation.

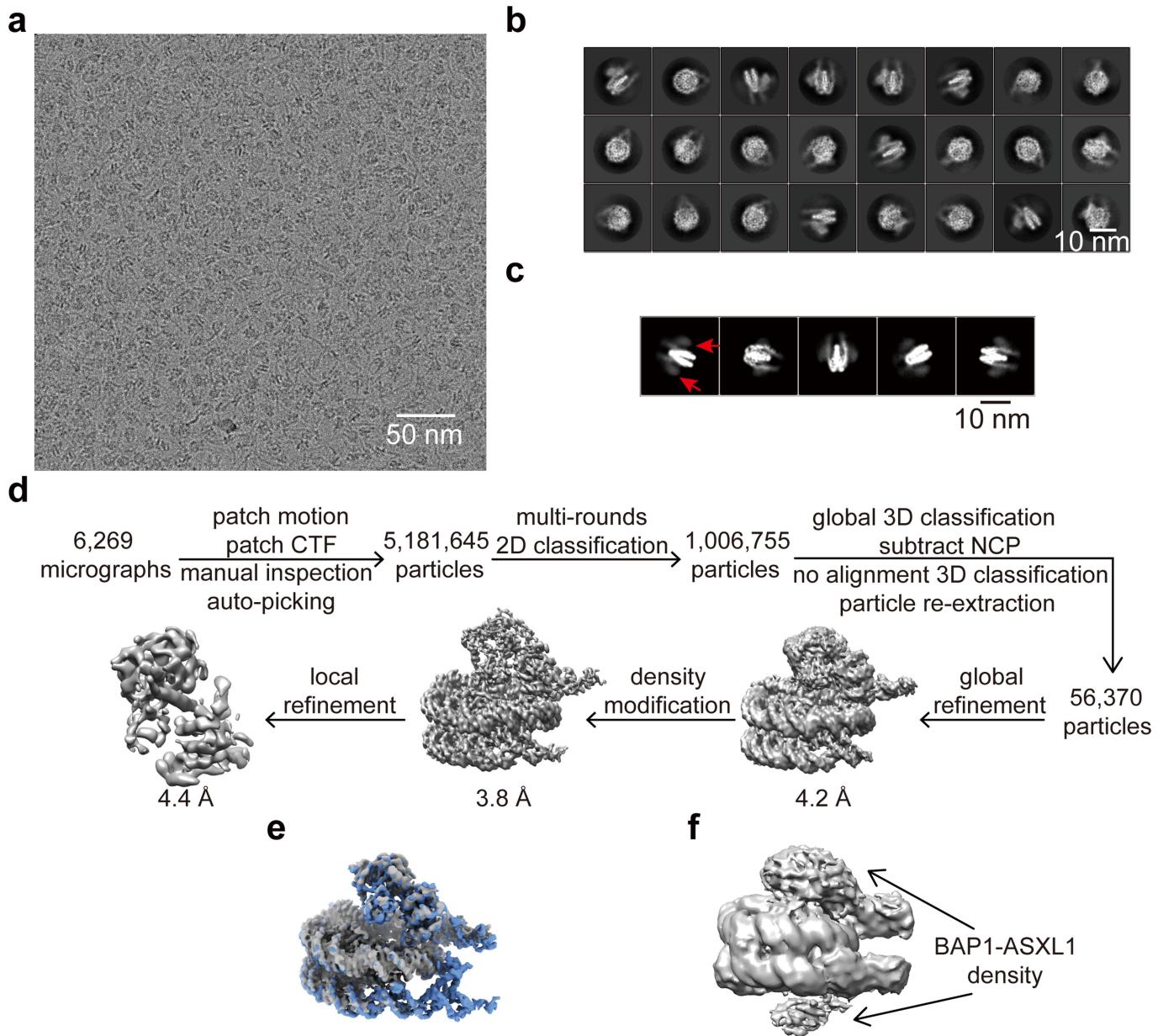


Extended Data Fig. 2 | Structure determination of the BAP1-ASXL1(1-378)-H1-NCP(H2AK119ub1) complex. **a**, A representative electron micrograph low-pass-filtered to 15 Å from 21,466 micrographs. **b**, Selected reference-free 2D class average. **c**, Workflow of cryo-EM data processing. **d**, The 3.0-Å global refinement map of the full complex (left), and the 3.9-Å local refinement

BAP1-ASXL1(1-378)-ubiquitin subcomplex (right). For convenience of presentation, a composite map was constructed by combining indicated global and focused refinement maps. **e**, Angular distribution of the full complex particles in the final reconstructions. **f**, Gold-standard Fourier correlation of the BAP1-ASXL1(1-378)-H1-NCP(H2AK119ub1) complex.

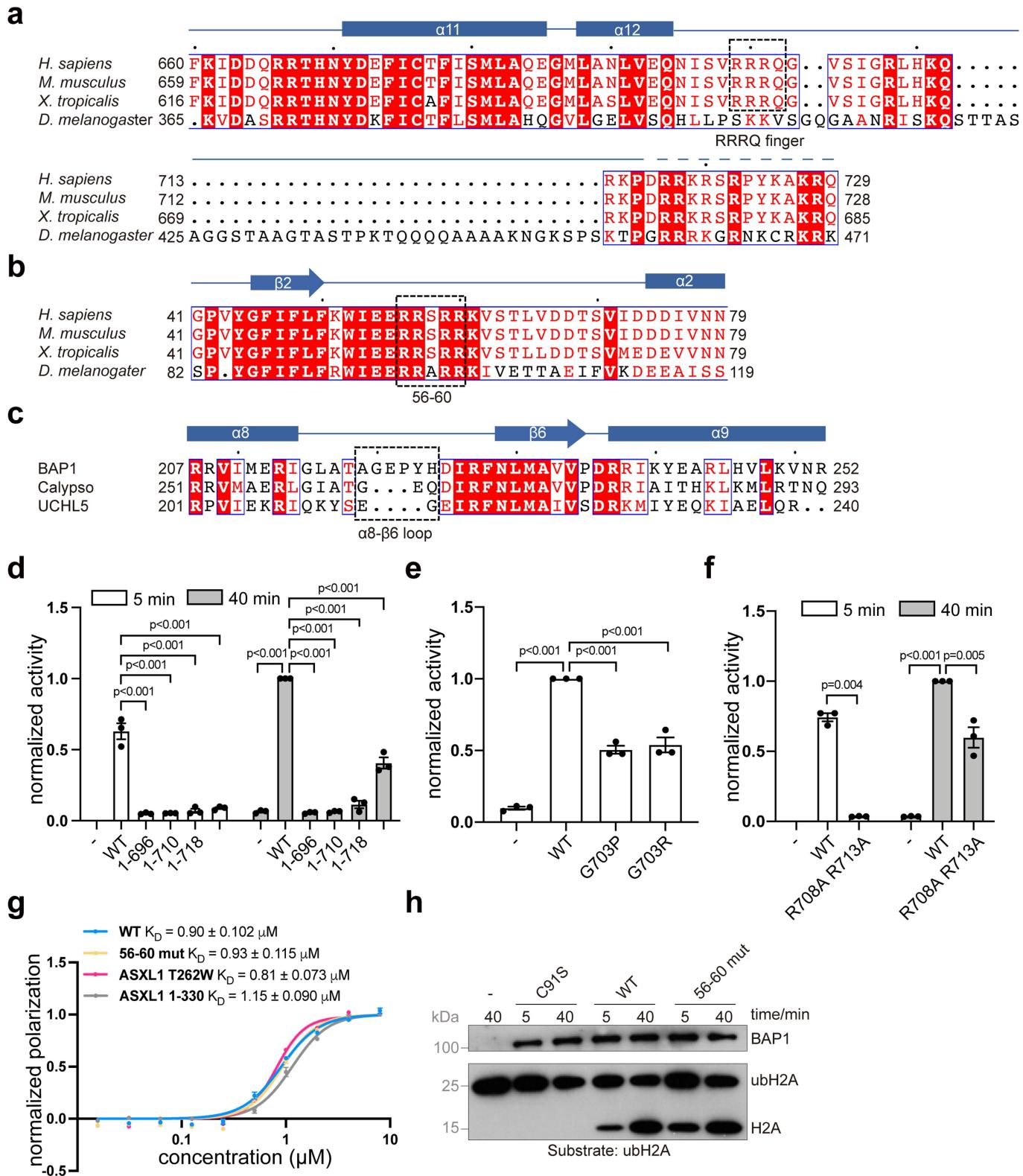


Extended Data Fig. 3 | EM density maps of various components or regions of the PR-DUB–substrate complex. **a**, 601 nucleosomal DNA. **b**, The SHL 0 region of nucleosomal DNA. **c**, Human histone octamer. Two H3s are coloured in blue and light blue, H4s in pale cyan and pale green, H2As in yellow-orange and bright orange, and H2Bs in pink and light pink. **d**, The globular domain of linker histone H1.4. **e**, Human BAP1. **f**, The DEUBAD domain of ASXL1. **g**, The ubiquitin moiety of H2AK119ub1. **h**, Selected regions of core histones. **i**, BAP1 regions. **j**, The DEUBAD domain of ASXL1. **k**, The C-terminal region of the ubiquitin moiety of H2AK119ub1. Atomic models are shown in a stick representation, and the maps are contoured at 5σ – 7σ .



Extended Data Fig. 4 | Cryo-EM analysis of the BAP1-ASXL1(1-378)-NCP(H2AK119ub1) complex without H1. **a**, A representative electron micrograph low-pass-filtered to 15 Å from 6,269 micrographs. **b**, Selected reference-free 2D class average. **c**, 2D class averages of two BAP1-ASXL1 complexes bound to NCP in higher contrast. The red arrows indicate the

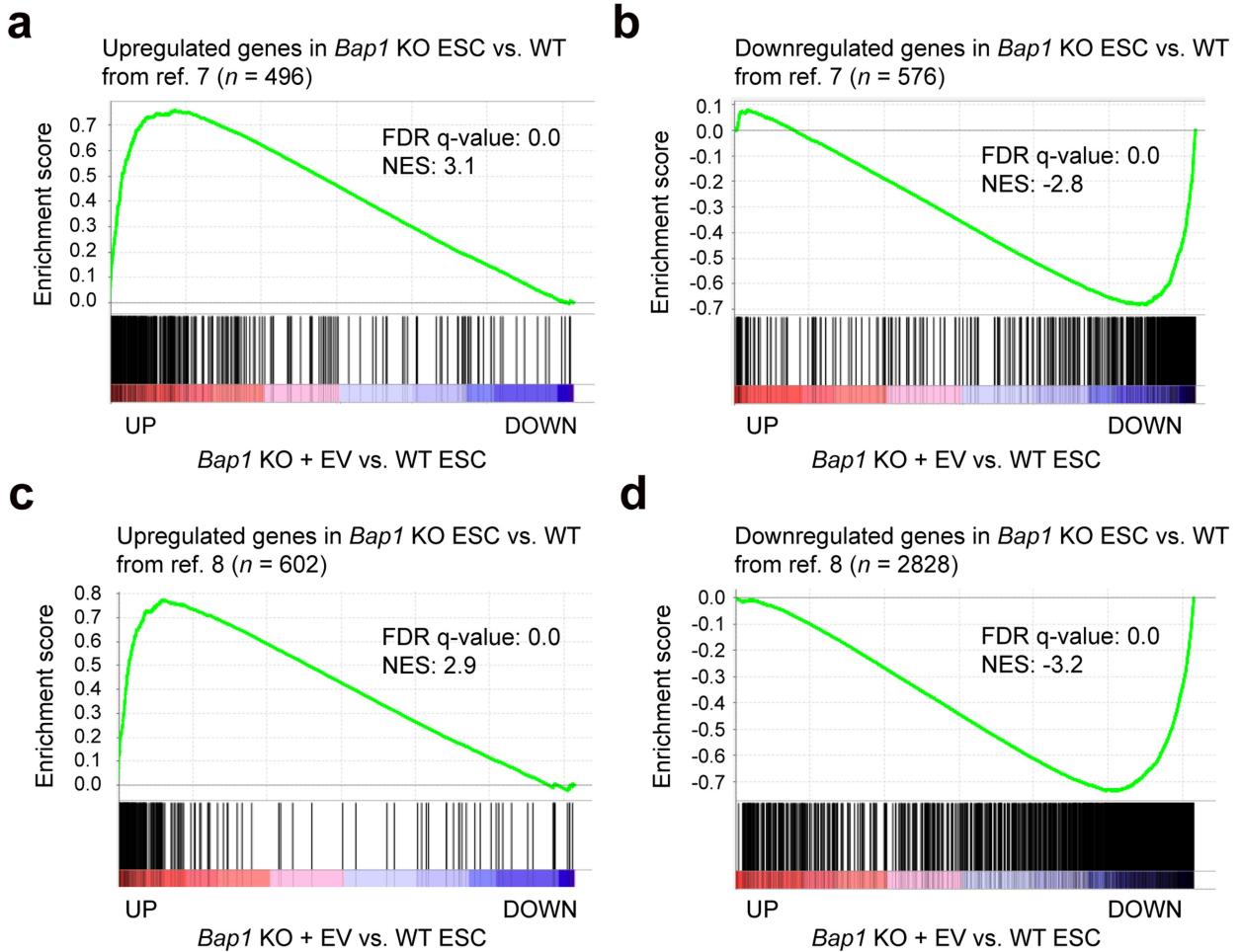
density for the bound BAP1-ASXL1 complexes. **d**, Workflow of cryo-EM data processing. **e**, Alignment of the EM densities for the BAP1-ASXL1(1-378)-NCP(H2AK119ub1) complexes with (blue) and without (grey) H1. **f**, 3D classification map of two BAP1-ASXL1(1-378) complexes bound to the nucleosome.



Extended Data Fig. 5 | See next page for caption.

Article

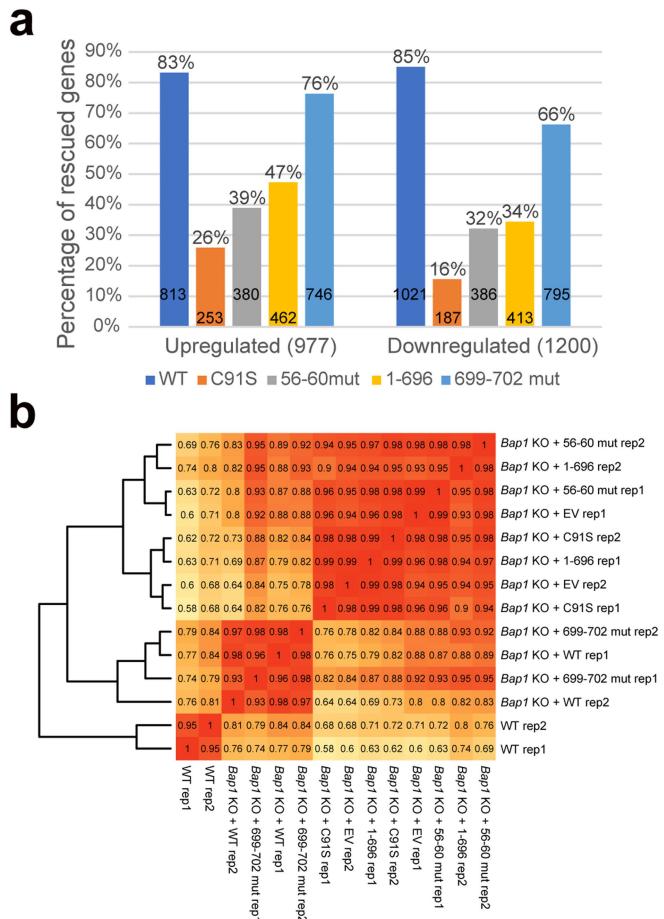
Extended Data Fig. 5 | BAP1 residues in distinct regions involved in H2AK119 deubiquitination and nucleosome binding. **a**, Sequence alignment of the CTE region of BAP1 orthologues. The RRRQ (aa 699–702) CTE finger motif is boxed in dashed lines. **b**, Alignment of the RRSRR motif (aa 56–60) in the β 2– α 2 loop of BAP1. **c**, Alignment of the sequences encompassing the α 8– β 6 loop of BAP1, Caly and UCH-L5. Above the sequence, BAP1 secondary structure elements are depicted, and period signs denote whole 10th residues of BAP1. **d–f**, Quantification of catalytic activities of western blot measurements in Figs. 2d, g and h, respectively. The values for indicated wild-type and BAP1 mutant complexes measured at 5 and 40 min in Fig. 2d,h are shown with clear and grey-shaded bars, respectively. The wild-type values (at 40 min) were set to unity, and all values were derived from three replicates, and represented as mean \pm s.e.m. *P* values, displayed at the top of the bar charts, denote the results of an unpaired two-tailed *t*-test. **g**, Fluorescence polarization measurements of binding affinities of the BAP1–ASXL1 complexes carrying indicated BAP1 (56–60 mut) or ASXL1 (T262W, 1–330) mutations to the NCP(H2AK119ub1) assembled with 167-bp 601 DNA containing a 5'-FAM fluorophore. The data were analysed in the same way as in Fig. 2e ($n = 3$ biologically independent experiments; mean \pm s.e.m.). **h**, Western blot analysis of free histone H2AK119 deubiquitination by wild-type and indicated BAP1 mutants in complex with ASXL1(1–378). A ‘–’ sign indicates a mock reaction without enzyme added. Reactions were monitored at the indicated time points, and detected with anti-His tag (top) or anti-H2A (bottom) antibodies.



Extended Data Fig. 6 | Gene set enrichment analysis. **a,b**, Upregulated (**a**) and downregulated (**b**) genes in *Bap1* KO ES cells versus wild-type ES cells from a previous study⁷ with respect to the global transcriptional changes observed

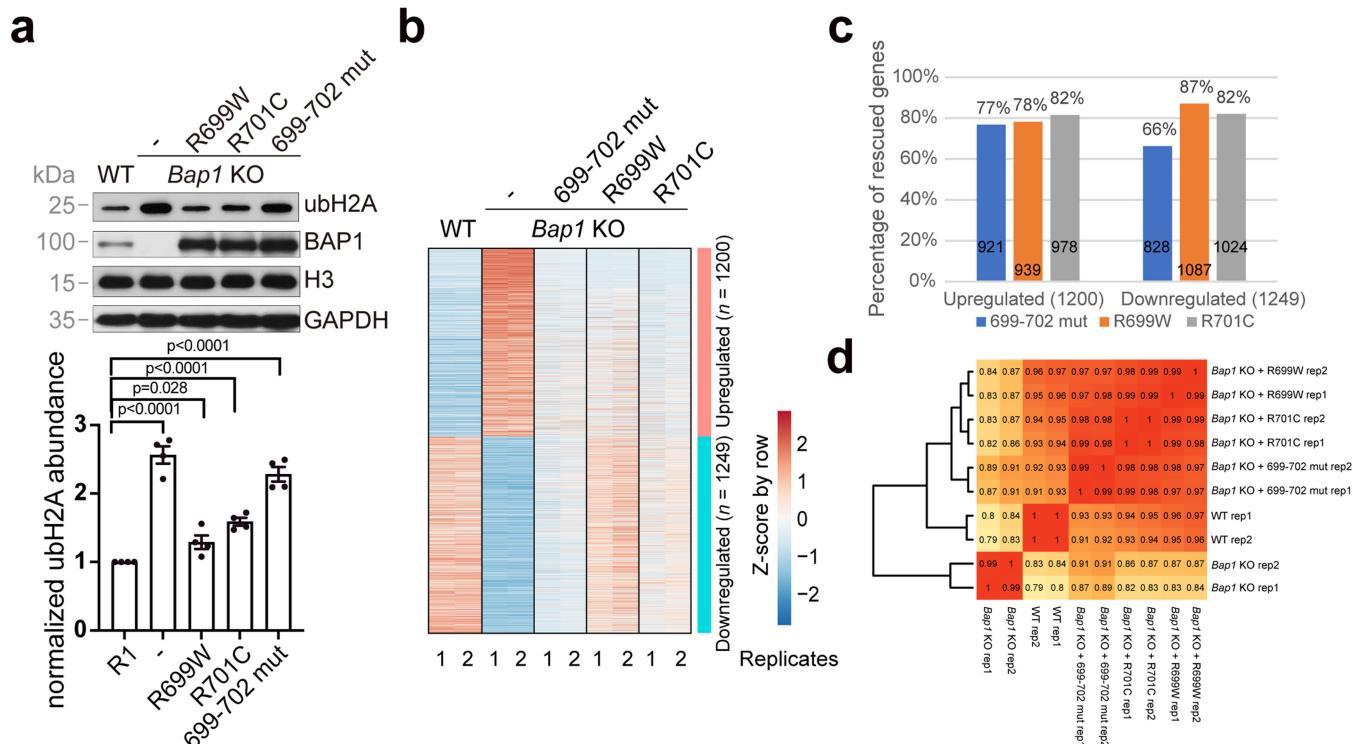
in our *Bap1* KO + EV ES cells versus wild-type ES cells. **c,d**, Upregulated (**c**) and downregulated (**d**) genes analysed in a previous study⁸ relative to the global transcriptional changes we observed.

Article



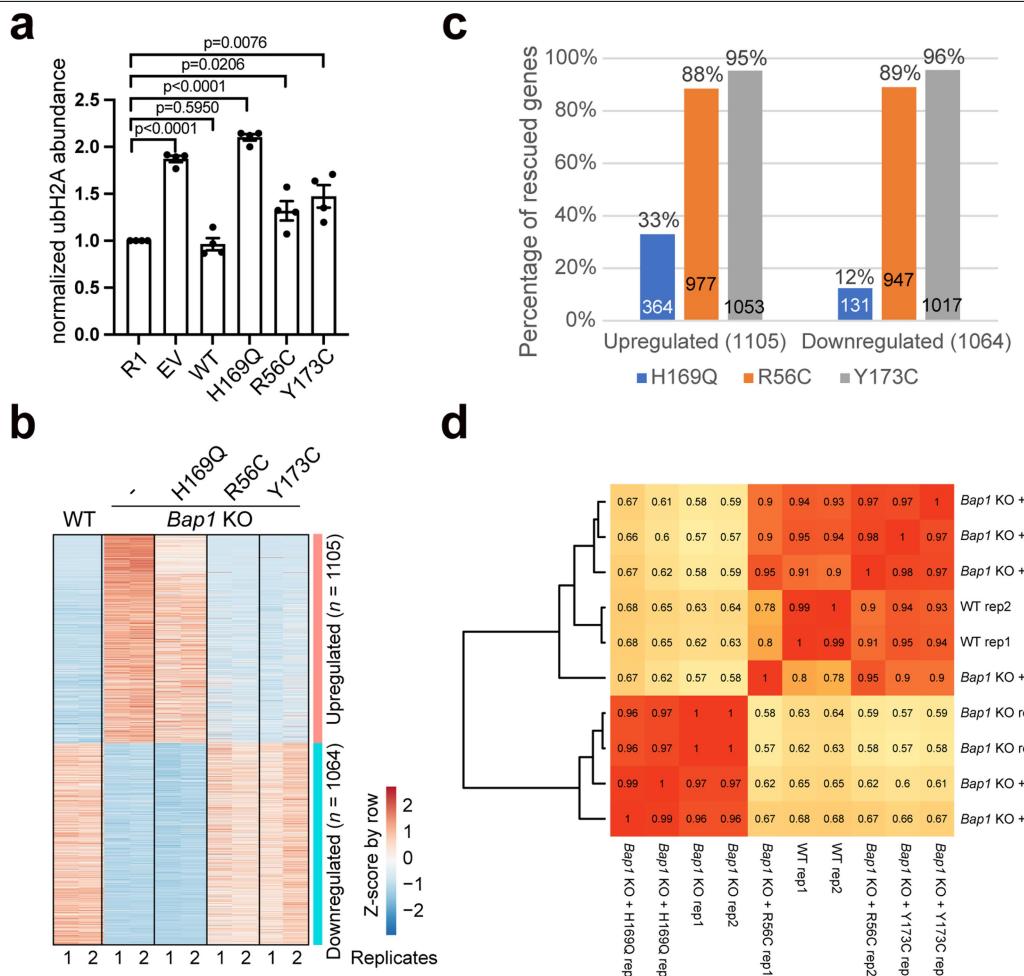
Extended Data Fig. 7 | Gene expression analysis of *Bap1* mutants.

a, Percentage of rescued BAP1-regulated genes (differentially expressed genes between *Bap1* KO + EVES cells and wild-type ES cells) in the indicated wild-type and mutant BAP1 rescue cells. **b**, Expression correlation among the 14 RNA-seq samples of mouse ES cells on BAPI-regulated genes (differentially expressed genes between *Bap1* KO + EV and wild-type R1). There are two replicates for each cell line. Pearson correlation coefficients are indicated.



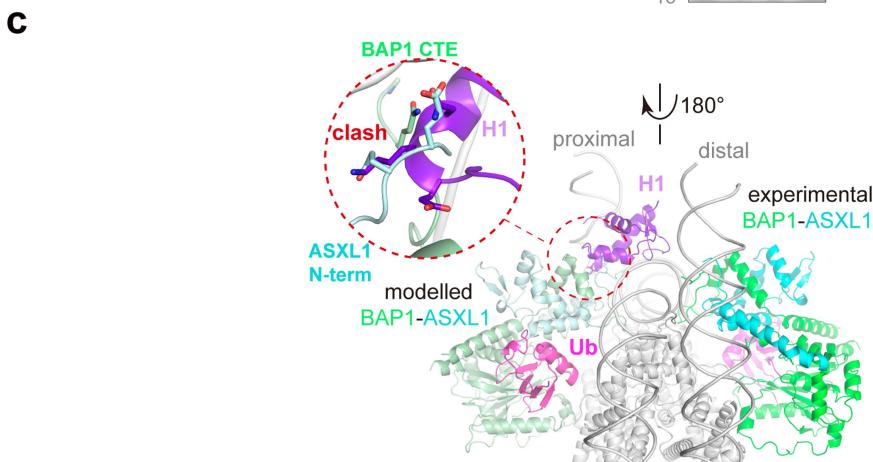
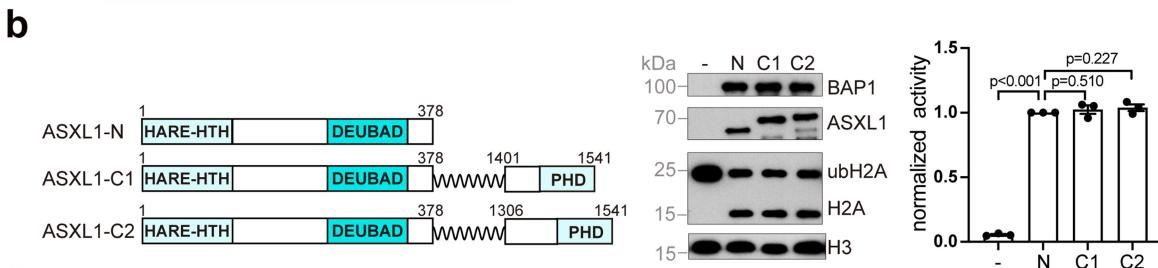
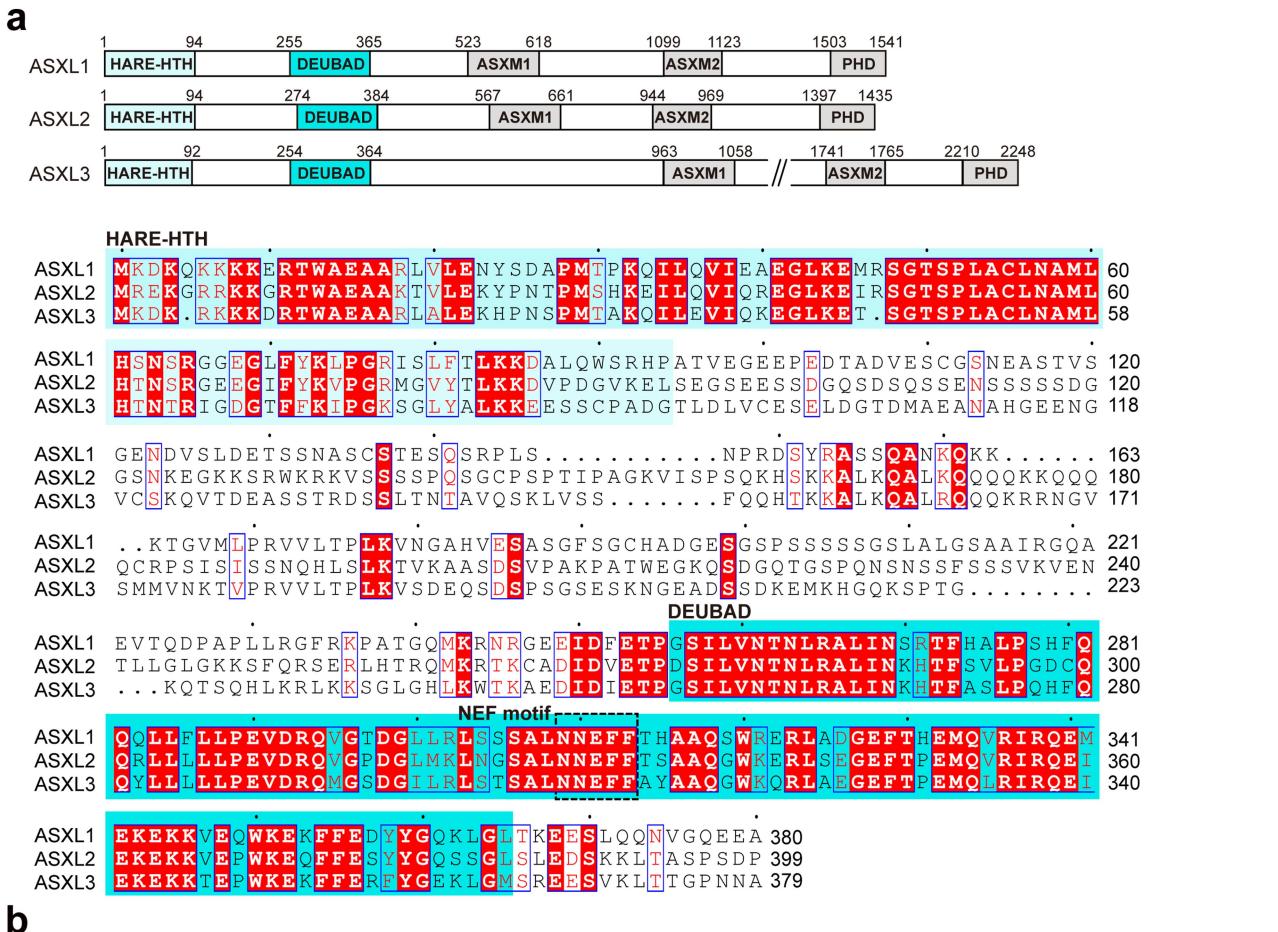
Extended Data Fig. 8 | Cellular DUB activities and changes in gene expression for BAP1 CTE cancer-associated mutations. **a**, Western blot detection of the levels of ubH2A and BAP1 in *Bap1* KO cells rescued with the indicated BAP1 mutant plasmids. Quantification of the western blot results is represented in a bar chart of mean \pm s.e.m. ($n = 4$ biologically independent experiments). P values, displayed at the top of the bar chart, denote the results of an unpaired

two-tailed *t*-test. **b**, Heatmaps of transcriptional levels of BAP1-regulated genes in wild-type and indicated rescue cell lines. **c**, Percentage of rescued BAP1-regulated genes (differentially expressed genes between *Bap1* KO and wild-type ES cells) in the indicated mutant BAP1 rescue cells. **d**, Expression correlation among the 10 RNA-seq samples.



Extended Data Fig. 9 | Quantification of cellular DUB activities and RNA-seq analysis for non-CTE BAP1 cancer-associated mutations. **a**, Quantification of western blot detection of H2AK119ub and BAP1, as shown in Fig. 3h, in *Bap1* KO cells transfected with an empty vector (EV), or plasmids re-expressing wild-type or the indicated BAP1 mutants. Bar charts represent plots of mean \pm s.e.m. ($n = 4$ biologically independent experiments). P values, displayed at the top of

the bar chart, denote the results of an unpaired two-tailed t-test. **b**, Heatmaps of transcriptional levels of BAP1-regulated genes in wild-type and indicated rescue cell lines. There are two replicates for each cell line. **c**, Percentage of rescued BAP1-regulated genes (differentially expressed genes between *Bap1* KO and wild-type ES cells) in each indicated mutant BAP1 rescue cells. **d**, Expression correlation among the 10 RNA-seq samples.



Extended Data Fig. 10 | Structure and function of ASXL1 domains and the effect of H1 on H2AK119 deubiquitination. **a**, Domain structures of ASXL1-ASXL3 and alignment of their N-terminal sequences encompassing HARE-HTH and DEUBAD domains. **b**, Left, schematic representation of ASXL1 constructs used in the deubiquitinase assay. Wavy lines represent a 2×(GGGGS) linker. Middle, western blot analysis of nucleosomal H2AK119 deubiquitination of BAP1 complexes with the indicated ASXL1 fragments.

b, Right, quantification of catalytic activities based on western blot measurements ($n=3$ biologically independent experiments, mean \pm s.e.m.). P values, displayed at the top of the bar chart, denote the results of an unpaired two-tailed t -test. No adjustment was made for multiple comparisons. **c**, A BAP1-ASXL1(1–378) complex modelled to the proximal DNA side (left) of the chromatosome based on the observed binding mode in the distal DNA side (right) sterically clashes with H1 (see inset).

Article

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

Samples	BAP1-ASXL1-H1-NCP ^{K119ub}	BAP1-ASXL1-NCP ^{K119ub}
Data collection and processing		
Microscope	Talos Arctica	Titan Krios
Detector	Gatan Bio-Quantum K2	Gatan Bio-Quantum K2
Magnification	130,000	130,000
Voltage (kV)	200	300
Electron exposure (e ⁻ /Å ²)	50	60
Defocus range (μm)	-1.5 to -2.0	-1.5 to -2.0
Pixel size (Å)	0.5	0.52
Symmetry imposed	C1	C1
Initial particle images (no.)	12,435,535	5,181,645
Final particle images (no.)	95,225	56,370
Map resolution (Å)	3.0 (global) 3.9 (local)	3.8 (global) 4.4 (local)
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	7k5y/6hgc/4uel	
Model resolution range (Å)	2.8 - 6	
Map sharpening B factor (Å ²)	-65	
Model composition		
Nonhydrogen atoms	17400	
Protein residues	1323	
Nucleotides	334	
B factors (Å ²)		
Protein	110.95	
Nucleotides	149.03	
R.m.s. deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.512	
Validation		
MolProbity score	1.53	
Clashscore	11.0	
Poor rotamers (%)	0	
Ramachandran plot		
Favored (%)	97.76	
Allowed (%)	2.24	
Disallowed (%)	0	

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM 3.8

Grey values of bands in western blots were collected by ImageJ 2.1.0.
Data of FP experiments were collected by EnVision Manager 1.13.3009.1409.
Data of SEC-MALS was collected by ASTRA 5.3.4.20.
RNA-seq data was collected by Illumina NovaSeq 6000 and DNBSEQ-T7
Flow cytometry data was collected by BD FACS Aria III

Data analysis

CryoSPARC v3.2

RELiON-3.0.8

UCSF pyem v0.5

DeepEMhancer

Rosetta 3.13

Coot 0.8.9

PHENIX 1.18.2

Pymol 1.7

UCSF Chimera 1.13

ChimeraX 1.4

MolProbity 4.3

Quantifications of western blots were analyzed by Excel and plotted by GraphPad Prism 9.0.0.

Data of FP was firstly calculated by Microsoft Excel, then analyzed and fitted to specific binding model with Hill slope by GraphPad Prism 9.0.0.

Data of SEC-MALS was analyzed by ASTRA 5.3.4.20 and plotted by Microsoft Excel.

Trimmomatic v0.36

Hisat2 v2.1.0

StringTie v1.3.3.b

featureCounts v1.6.3
edgeR v3.38.4
GSEA v4.0.3
R v3.4.0
Flow cytometry data was analyzed using FlowJo v10.8.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) with accession numbers 34431, 34432, 35179, 35180, 35181 and 35182. The atomic coordinates have been deposited in the Protein Data Bank (PDB) under the accession code 8H1T.

The coordinates for the structures of Drosophila CALY in complex with DEUBAD of ASX (PDB id: 6HGC), H2AK15-ubiquitinated NCP in complex with BRCA1-BARD complex (PDB id: 7E8I), the H2BK120-ubiquitinated NCP in complex with the SAGA DUB module (PDB id: 4ZUX), the UCH-L5 complex with DEUBAD of RPN13 and ubiquitin (PDB id: 4UEL), and the chromatosome with linker histone H1.4 (PDB id: 7K5Y) were downloaded from the RCSB Protein Data Bank (<https://www.rcsb.org>).

The RNA-Seq data have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA006013) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.

Two public RNA-seq data sets under the accession numbers GSE162739 and GSE161995 were used in this study. The list of differentially expressed genes (DEGs) defined in ref. 7 was downloaded from <https://www.sciencedirect.com/science/article/pii/S1097276521005001?via%3Dihub#app2> and the original file name is 1-s2.0-S1097276521005001-mmcc3.xlsx. The list of DEGs defined in ref. 8 was downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161995> and the original file name is GSE161995_BAP1ff_cRNA_seq_DESeq2_Table_GEO.txt. A threshold of P-adj < 0.05 and fold change > 1.5 was used to extract the DEGs.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size Sample sizes are reported in the figure legends or the Methods section. For western blots and fluorescence polarization, each assay was conducted at least 3 times independently. No sample-size calculation was performed.

Data exclusions In Extended Data Fig. 5g, two points of ASXL1 T262W at the concentration of 8 uM were excluded because at this concentration only 2 samples were measured, which is not sufficient for statistical analysis. However, the exclusion does not affect the calculation of the Kd value appreciably, as these data points are located in the already plateaued region.

Replication All replications were successful, and replication numbers are either described in the figure legends or the Methods section. Protein purifications were conducted at least 6 times and showed similar chromatography and electrophoresis patterns. Reconstruction of nucleosome and chromatosome were repeated at least for 6 times and gave similar results. GraFix was repeated 3 times independently and showed similar electrophoresis patterns. For RNA-seq, two replications showed similar results.

Randomization For cryo-EM structure determination, all particles were randomly split into two half sets during 3D refinement. For all other experiments, all data were used for analysis, therefore no randomization was needed.

Blinding RNA-seq data were generated by commercial vendors who did not have the knowledge about the cell lines being sequenced. Blinding does not apply to other experiments, as no groups were assigned.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.</i>
Research sample	<i>Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.</i>
Sampling strategy	<i>Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.</i>
Data collection	<i>Describe the data collection procedure, including who recorded the data and how.</i>
Timing and spatial scale	<i>Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Reproducibility	<i>Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.</i>
Blinding	<i>Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access & import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used

- anti-BAP1, Proteintech, 10398-1-AP, western blot (1:500)
- anti-BAP1, Cell Signaling Technology, 13271s, western blot (1:1000)
- anti-H2AK119ub1, Cell Signaling Technology, 8240, western blot (1:2000)
- anti-H3, Abcam, ab1791, western blot (1:80k)
- anti-GAPDH, ABclonal, AC001, western blot (1:100k)
- anti-H2A, Millipore, 07-146, western blot (1:5000)
- anti-H2A, Cell Signaling Technology, 12349s, western blot (1:2000)
- anti-H2B, Abcam, ab1970, western blot (1:5000)
- anti-His tag, Invitrogen, MA1-21315, western blot (1:1000)
- anti-GST, Beyotime, AF0174, western blot (1:4000)

Validation

- Commercially available products, validations are available on manufacturers' websites.
- anti-BAP1, Proteintech, 10398-1-AP: <https://www.ptgcn.com/products/BAP1-Antibody-10398-1-AP.htm>
- anti-BAP1, Cell Signaling Technology, 13271s: <https://www.cellsignal.com/products/primary-antibodies/bap1-d7w7o-rabbit-mab/13271>
- anti-H2AK119ub1, Cell Signaling Technology, 8240: https://www.cellsignal.cn/products/primary-antibodies/ubiquityl-histone-h2a-lys119-d27c4-xp-rabbit-mab/8240?site-search-type=Products&N=4294956287&Ntt=8240&fromPage=plp&_requestid=4288435
- anti-H3, Abcam, ab1791: <https://www.abcam.cn/histone-h3-antibody-nuclear-marker-and-chip-grade-ab1791.html>
- anti-GAPDH, ABclonal, AC001: <https://abclonal.com/catalog-antibodies/GAPDH-RabbitpAb/AC001>
- anti-H2A, Millipore, 07-146: <https://www.sigmaldrich.cn/CN/zh/product/mm/07146>
- anti-H2A, Cell Signaling Technology, 12349s: https://www.cellsignal.cn/products/primary-antibodies/histone-h2a-d6o3a-rabbit-mab/12349?site-search-type=Products&N=4294956287&Ntt=12349s&fromPage=plp&_requestid=4289536
- anti-H2B, Abcam, ab1970: <https://www.abcam.cn/prolactinprl-antibody-me-122-ab1970.html>
- anti-His tag, Invitrogen, MA1-21315: <https://www.thermofisher.cn/cn/zh/antibody/product/6x-His-Tag-Antibody-clone-HIS-H8-Monoclonal/MA1-21315>
- anti-GST, Beyotime, AF0174: <https://beyotime.com/product/af0174.htm>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Mouse ES cell line R1 (ATCC) was a gift from Guohong Li, Institute of Biophysics. Stable Bap1 KO cell line was generated by CRISPR/Cas9. BAP1 WT or mutation rescue cell lines (Bap1 KO + empty vector, Bap1 KO + WT BAP1, Bap1 KO + C91S BAP1, Bap1 KO + 56-60mutant BAP1, Bap1 KO + 699-702mutant BAP1, Bap1 KO + 1-696 BAP1, Bap1 KO + R56C BAP1, Bap1 KO + H169Q BAP1, Bap1 KO + Y173C BAP1, Bap1 KO + R699W BAP1, Bap1 KO + R701C BAP1) were generated by PiggyBac system, Bap1 KO cells were mock transfected or transfected with EF1α-driven vectors encoding BFP-P2A-HA-BAP1 wild type or BAP1 truncation/mutations and another vector encoding transposase using Lipofectamine 3000 (Thermo Fisher Scientific),

according to manufacturer's instructions. BFP-positive cells were selected by FACS 24 h post-transfection. After further cultured for about seven days, BFP-positive cells within specific BFP signal range were collected as stable BAP1 rescue cells.

Authentication

Mouse ES cell line R1 was frequently checked by the morphological features, but not authenticated. Bap1 KO cell line and BAP1 WT or mutation rescue cell lines were verified by western blot and RNA-seq

Mycoplasma contamination

The cell lines were not tested for mycoplasma contamination but no indication of contamination was observed.

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

BAP1 rescue cells were generated by PiggyBac system. Bap1 KO cells were mock transfected or transfected with EF1 α -driven vectors encoding BFP-P2A-HA-BAP1 wild type or BAP1 truncation/mutations and another vector encoding transposase using Lipofectamine 3000 (Thermo Fisher Scientific). BFP-positive cells were sorted by FACS 24h post-transfection. After further cultured for about seven days, the cells were collected and sorted by FACS the second time and BFP-positive cells within specific BFP signal range were collected as stable BAP1 rescue cells

Instrument

Cell sorting was done using BD FACS Aria III or BD Influx

Software

BD FACS Diva v8.0.1 software was used for data collection and FlowJo v10.8.1 was used for data analysis

Cell population abundance

For the first FACS sorting, there were about 4-10% BFP-positive cells; For the second FACS sorting, there were about 50% BFP-positive cells. In order to make all BAP1 rescue cells has similar BAP1 expression level, the cells in the same BFP signal range were collected, which were about 25%

Gating strategy

Untransfected cells as negative control to set gating strategy. BAP1 WT or mutations rescue cells were gated using BFP signal

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.