METHODS

Molecular identification of the *calypso* gene. We previously mapped the *calypso* mutations to the right arm of chromosome 2 (2R)[8]. Complementation tests with a set of chromosomal deficiencies uncovering much of chromosome arm 2R allowed us to locate the *calypso* mutations to a short interval in the 52F-53A region. In particular, *calypso*^{1} failed to complement *Df(2R)Jp4*, *Df(2R)Jp5*, *Df(2R)Jp6*, *Df(2R)Jp7*, *Df(2R)Jp8* or *Df(2R)Exel6063* but complemented all other tested deficiencies on chromosome arm 2R. *calypso*^{2} also failed to complement *Df(2R)Jp4* and *Df(2R)Exel6063*. Because *Df(2R)Jp4* complemented mutations in *Khc8* (provided by P. Rørth), the *calypso* mutations were thus mapped to a chromosomal interval extending from the centromere-proximal end of *Df(2R)Exel6063* and the *Khc* locus. Moreover, we found that *calypso*^{1} and *calypso*^{2} both complemented *csul*^{RM} (provided by B. Mechler), a viable but female-sterile P-element insertion and *csul*^{RL}, a small chromosomal deletion that removes *csul*, *Khc* and *fidipidine*. Sequencing of the CG8445 open reading frame on the *FRT40 FRT42 y*$^{+}$ *calypso*^{1} and *FRT40 FRT42 y*$^{+}$ *calypso*^{2} chromosomes revealed a C-to-T transition that was present in both alleles and changed the CAA codon for Gln 41 into a TAA termination codon; in contrast, the wild-type CAA codon for Gln 41 was present on the isogenic *FRT40 FRT42 y*$^{+}$ chromosome on which these mutations had been induced by ethylmethanesulphonate (EMS)[8]. We note that even though *calypso*^{1} and *calypso*^{2} contain the same Gln 41>stop mutation, the two alleles had been isolated in two independently performed mutagenesis experiments[8]. No other mutations were detected in the CG8445 open reading frame, but we note that the *FRT40 FRT42 y*$^{+}$, *FRT40 FRT42 y*$^{+}$ *calypso*^{1} and *FRT40 FRT42 y*$^{+}$ *calypso*^{2} chromosomes all contain a TCA codon (serine) at residue 17 of the CG8445 open reading frame, whereas the database genomic sequence and the SM6b balancer chromosome contain a GCA codon for alanine at this position.

Molecular analysis of ASX alleles. *Asx*^{22P4} homozygous embryos fail to express detectable amounts of ASX protein (Supplementary Fig. 5). We have not been able to find a lesion in the ASX open reading frame on the *FRT40 FRT42 y*$^{+}$ *Asx*^{22P4} chromosome[8]. *Asx*^{27J6} homozygotes express a truncated ASX protein that migrates with an apparent molecular mass of 58 kDa (Supplementary Fig. 5). Sequencing of the ASX open reading frame on the *FRT40 FRT42 y*$^{+}$ *Asx*^{27J6} chromosome[8] showed a deletion of the cytosine in the CGT codon for Arg 433. This mutation results in a frameshift that is predicted to terminate the open reading frame after 45 nucleotides, that is, the predicted *Asx*^{27J6} open reading encodes ASX(1–432) fused to 15 further amino acids.

Immunostaining of *Drosophila* embryos and imaginal discs. Staining of *Drosophila* embryos and larval imaginal discs was performed following standard protocols. Induction of homozygous mutant cell clones in imaginal discs of *Drosophila* larvae and the rescue experiments with *hsp70-calypso* transgenes shown in Fig. 4 were performed as previously described[27].

Generation of H2Bub1-containing mononucleosomes. *Xenopus* octamers containing H2Bub1 were chemically generated as described[28]. Mononucleosome assembly was performed by stepwise salt dialysis as described earlier and the nucleosomes were directly used for deubiquitination assays. We note that the H2Bub1 isopeptide bond can be hydrolysed by UCH-L3 under appropriate experimental conditions[29].

Antibodies. The following previously described antibodies were used in this study: anti-UBX[30], anti-ABD-B[31], anti-H2A (Millipore, 07-146), anti-H2B (Millipore, 07-371), E6C5 (Millipore, 05-678, lot: DAM1514075), anti-Ubiquitin P4D1 (Santa Cruz Biotechnology, sc-8017), anti-Flag (Sigma, F7425), anti-HA (Sigma, H3663), anti-H3 (Abcam, AB1791), anti-H3K4me3 (provided by B. Turner), and anti-H3K27me3 (Upstate, 07-449).

For this study, antibodies against Calypso(1–471) and ASX(210–336) were raised in rabbits. In both cases, the epitopes for antibody production were expressed as 6×His-tagged fusion proteins in *E. coli* and purified under denaturing conditions.

As shown in Supplementary Fig. 6a, the E6C5 antibody (Millipore/Upstate) recognizes a multitude of protein bands in nuclear extracts from *Drosophila* embryos. For the western blot analyses shown in Supplementary Fig. 6b and c, H2Aub1-containing *Drosophila* mononucleosomes were generated as described earlier. K48-linked ubiquitin pentamers (Boston Biochem, UC-216) and K63-linked ubiquitin pentamers (Boston Biochem, UC-316) were used as a control.

ChIP assays. X-ChIP from *Drosophila* larval imaginal discs and qPCR analysis to determine protein binding at specific chromosomal locations were performed as described[19] with the following modifications. Washes of immunoprecipitated material on protein A-sepharose beads were performed for 10 min at 4 °C with 1 ml of the following buffers: five washes with RIPA buffer (350 mM NaCl (ASX ChIP)/140 mM NaCl (Calypso ChIP), 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 1 mM phenylmethylsulphonyl fluoride (PMSF)), followed by one wash with LiCl buffer (250 mM LiCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) and two washes with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA); each time, wash solution was removed after centrifugation for 2 min at 3,000*g* at 4 °C.

ChIP-chip in *Drosophila* using Affymetrix whole genome microarrays. Chromatin was prepared from imaginal disc (wing, haltere and third leg) and central nervous system tissues of third instar *Drosophila* larvae and ChIP was performed as described[17]. To determine regions bound by PR-DUB, three ChIP assays were performed with the anti- Calypso(1–471) antibody and three with the anti-ASX(210–336) antibody, using independently prepared batches of chromatin. The immunoprecipitated material was amplified by ligation-mediated PCR, fragmented, labelled and hybridized to Affymetrix whole genome microarrays as described[17]. Data analysis was done as described in ref. 17 and later.

Genome assembly and annotation. The third *Drosophila melanogaster* genome assembly (UCSC dm3 or Flybase 5.x) and the gff export of Flybase 5.23 for genome annotation was used.

Affymetrix probes re-mapping. The Dm35b_MR_v02-3_BDGPv4h.new.bpmap file from the CisGenome website (http://www.biostat.jhsph.edu/~hji/cisgenome/) was used. This file contains the re-mapped location to genome version 5 (that is, UCSC dm3) of all the original Affymetrix 25mer sequences but lacks probes that cannot uniquely map to the genome.

Detection of regions that are significantly bound by PR-DUB. A quantile normalization[32] was applied to normalize all six ChIP hybridizations (three anti-Calypso and three anti-ASX ChIP assays, see earlier) to three genomic DNA hybridizations. Regions that were significantly enriched for PR-DUB binding were identified using TileMap[16] with the Hidden Markov Model (HMM). A TileMap score where 50% of the previously identified 237 regions bound by PhoRC and Ph in larva[17,18] were recovered was used as threshold for considering significantly bound regions. In addition, the top 5% of regions below cutoff exhibiting overlap with Pho[17], dSfmbt[17], Ph[18] or Sce-bound regions (K.O., L. Gutierrez and J.M., unpublished observations) were rescued and included in the final set of high-confidence PR-DUB binding sites. We thus identified 879 regions (560 scoring above cutoff, plus 319 rescued) bound by PR-DUB in larval CNS and imaginal disc cells.

Venn diagram counts were obtained as follows. Two or more regions that overlap with at least one base were merged and defined as a 'common' region.

PR-DUB-bound regions, target gene assignment and Gene Ontology (GO) slim analysis. For the data shown in Supplementary Table 2, the relative distance of the midpoint of each PR-DUB-bound region with respect to the closest transcription start site (TSS) was computed. Target genes were assigned based on TSS-proximal location. Assigned genes to each data set were tested for enriched GO slim term annotations.

Quantitative PCR to determine binding at specific chromosomal locations. qPCR analysis was done as described[19], using the following primers. *Ubx*-1: forward, 5′-TGGGATTGCGATAGTGTTGGTC-3′, reverse, 5′-CGCAGCCATTATGAAACCTCCT-3′ (−37.7 kb); *Ubx*-2: forward, 5′-ATTTGGTCCGCAGTCGAGTGCATTT-3′, reverse, 5′-CCACATGCACATCCTGGATGCCAAT-3′ (−32.5 kb); *Ubx*-3: forward, 5′-GCAGCATAAAACCGAAAGGA-3′, reverse, 5′-CGCCAAACATTCAGAGGATAG-3′ (−30.9 kb); *Ubx*-4: forward, 5′-TAGTCTTATCTGTATCTCGCTCTTA-3′, reverse, 5′-CAGAACCAAAGTGCCGATAACTC-3′ (−29.8 kb); *Ubx*-5: forward, 5′-AAGGCGAAAGAGAGCACCAA-3′, reverse, 5′-CGTTTTAAGTGCGACTGAG-3′ (−29.6 kb); *Ubx*-6: forward, 5′-GCACGCACTAAACCCCATAA-3′, reverse, 5′-TCCACCTCCTCTTCCTCTCTC-3′ (−29.0 kb); *Ubx*-7: forward, 5′-GGTCAAAGGCCATACAATTCCA-3′, reverse, 5′-ATCTGTGAGAATGCGGCATCTAA-3′ (−16.0 kb); *Ubx*-8: forward, 5′-ATCGGTAGCTTGTTGCAGCA-3′, reverse, 5′-GGCTACTTGGACAGGTGTGAGC-3′ (−2.4 kb); *Ubx*-9: forward, 5′-TCCAATCCGTTGCCATCGAACGAAT-3′, reverse, 5′-TTAGGCCGAGTCGAGTGAGTTGAGT-3′ (0 kb); *Ubx*-10: forward, 5′-AATTGGTTTCCAGGGATCTGC-3′, reverse, 5′-ATCCAAAGGAGGCAAAGGAAC-3′ (+0.8 kb); *Ubx*-11: forward, 5′-ATGATATCTCGTCTGGCACTAC-3′, reverse, 5′- AGACATCCAGCAAACTGCGATA-3′ (+8.0 kb); *Ubx*-12: forward, 5′-GCCGTGGAGCAGTTCAAAGTA-3′, reverse, 5′-TCGTTGGTCGTGTCCTCTTAATT-3′ (+26.8 kb); *Ubx*-13: forward, 5′-CCATAAGAAATGCCACTTTGC-3′, reverse, 5′-CTCTCACTCTCTCACTGTGAT-3′ (+31.5 kb); *Ubx-14*: forward, 5′-GTCCTGGCCAAGGCAAATATT-3′, reverse, 5′-CGAAAGGAGAACGGAGAATGG-3′ (+34.4 kb); *Ubx*-15: forward, 5′-GGCATCTTCCAGGTTTTGAGT-3′, reverse, 5′-ACCAGCATTCGTCACTATTCG-3′ (+70.6 kb); *Ubx*-16: forward, 5′-GCCGAGGGTCAGAGAGTTTA-3′, reverse, 5′-CTGCATCCGACCACTTACCT-3′ (+72.3 kb); *Abd-B*: forward, 5′-GGAATACCGCACTGTCGTAGG-3′, reverse, 5′-GCAGCCATCATGGATGTGAA-3′ (+0.2 kb); *Scr*: forward, 5′-GAAGTGCGCCACGTTCAAT-3′, reverse, 5′-TCCTCTCTCTCGCACTCGTT-3′ (+0.2 kb); control-1: forward, 5′-TCAAGCCGAACCCTCTAAAAT-3′, reverse, 5′-AACGCCAACAAACAGAAAATG-3′ (−12.5 kb); control-2: forward, 5′-CCGAACATGAGAGATGGAAAA-3′, reverse, 5′-AAAGTGCCGAC

AATGCAGTTA-3′ (−3.1 kb); control-3: forward, 5′-CAGTTGATGGGATGAAT TTGG-3′, reverse, 5′-TGCCTGTGGTTCTATCCAAAC-3′ (+12.4 kb).

Tandem affinity purification of Calypso complexes. The *α-tubulin1-TAP-calypso* transgene in the *Drosophila* transformation vector pCaSpeR had the following structure: a 2.6-kb fragment of the *α-tubulin1* gene containing promoter and 5′-untranslated region sequences, followed by the N-terminal TAP-tag[33] was linked to a *calypso* cDNA fragment that contained the whole *calypso* open reading frame (detailed plasmid map available on request). Rescue function of the *α-tubulin1-TAP-calypso* transgene was tested by introducing the three independent transgene insertions into a *calypso*^{2}*/Df(2R)Exel6063* mutant background; in all cases we obtained *w; calypso*^{2}*/Df(2R)Exel6063; α-tubulin1-TAP-calypso/+* animals that were wild-type in appearance and fully viable and fertile.

TAP was performed on embryonic nuclear extracts prepared from wild-type or *a-tubulin1-TAP-calypso* transgenic embryos, following previously described protocols[26].

Mass spectrometry, capillary LC–MS/MS analysis and protein identification. Silver-stained protein bands were excised and trypsin digested as previously described[34]. The samples were loaded and separated on a nano-flow 1D-plus Eksigent (Eksigent) HPLC system coupled to a qStar Pulsar *i* quadrupole time-of-flight MS (Applied Biosystems). The peptides were separated by a linear gradient that started at 100% mobile phase A and increased the mobile-phase composition to 50% B (0.5% acetic acid in 98% acetonitrile) over a span of 45 (single protein band) or 120 (complex samples) minutes at a constant flow rate of 200 nl min^{-1}. The mass spectrometer was operated in data-dependent positive-ion mode. MS spectra were acquired over *m*/*z* range from 350 to 1,300 for 1 s and one subsequent MS/MS spectrum from 60 to 1,800 *m*/*z* for 1.5 s.

MS/MS data was extracted using the AnalystQS software (Applied Biosystems). Peptides were identified by searching the peak-list against the NCBI database using the MASCOT (Matrix Science) algorithm. Peptide tolerance was limited to 50 p.p.m., and peptides with a score below 15 were excluded. Protein identifications were accepted if a single peptide with an individual MASCOT score above 45 or at least two peptides with a summed MASCOT score of more than 40 were identified.

A detailed list of peptide sequences obtained from mass spectrometry analysis of the protein bands shown in Fig. 1 is shown in Supplementary Fig. 1 and in Supplementary Table 1. Peptides identified by LC–MS/MS analysis of total purified material are also shown in Supplementary Table 1.

Constructs for baculovirus production and purification of recombinant proteins. Baculoviruses expressing Flag–ESC, Flag–Sce, BMI1–Flag and Flag–RING1A have been described previously[35–37]. For this study we generated baculovirus vectors (pFastBac) that encode the following proteins: Flag–Calypso and Flag–Calypso(C131S), both containing the entire Calypso(1–471) open reading frame, HA–ASX, containing the entire ASX(1–1668) open reading frame, HA–ASX(2–337), Flag–BAP1(2–729) and HA–ASXL1(2–365). Detailed plasmid maps for all constructs are available on request.

Flag-affinity purification of proteins and complexes was carried out as described[26] with the only modification that protein complexes were reconstituted by infecting cells with individual viruses and then mixing cell lysates and incubating them for 2–3 h at 4 °C under mild agitation before Flag-affinity purification.

Deubiquitination assay on Ub-AMC. Ub-AMC (25 pmol) (BostonBiochem) was coincubated with 10 pmol of protein or protein complex in 1× assay buffer (25 mM HEPES, pH 7.5, 0.25 mM EDTA, 0.05% Chaps, 10% dimethylsulphoxide (DMSO) and 10 mM DTT)[38] at 25 °C. Fluorescence spectroscopy measurements were done 1×/20 s at excitation wavelength = 380 nm, and emission wavelength = 436 nm.

Generation of H2Aub1-containing mononucleosomes. As a template for assembly of mononucleosomes we used PCR to generate a 288-bp-long 5′-biotinylated DNA fragment that included 87 bp of pUC19 DNA followed by a unique EcoRV restriction site and a 201-bp-long '601' nucleosome positioning sequence[39] (full fragment sequence available on request).

Recombinant *Xenopus* or *Drosophila* octamers were prepared as described[40] and mononucleosomes were assembled from the 5′-biotinylated 288-bp DNA fragment by stepwise salt dialysis, as follows. Octamers and DNA were mixed in a ratio of 1.0 μg octamers:1.0 μg DNA (288 bp) in TE buffer containing 2 M NaCl (2 M NaCl, 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, pH 8.0). The mixture was incubated at room temperature for 30 min, then stepwise dialysed against TE buffer containing 1.2 M NaCl, 1.0 M NaCl, 0.8 M NaCl and 0.6 M NaCl at 4 °C for 2 h each. The final dialysis step was carried out overnight against buffer C (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 1 mM DTT)[24]. Nucleosomal H2A was ubiquitinated using recombinant human E1, recombinant human E2 (UbcH5c) and recombinant partial human E3 complex (RING1B(1–159)–BMI1(1–109)) as described before[24]. (Recombinant E1, E2 and E3 enzymes were a gift from G. Buchwald and T. Sixma.) For purification of ubiquitinated nucleosomes, the reaction mixture was incubated with streptavidin-coupled Dynabeads (Dynabeads M-280 Streptavidin, Invitrogen) for 1 h at room temperature, followed by removal of the supernatant and washes of the bead-bound nucleosomes with buffer C (4 × 2-min washes). Nucleosomes were released from beads by cleavage with the restriction enzyme EcoRV for 1 h at 37 °C. The supernatant containing H2Aub1-mononucleosomes was then used for deubiquitination assays.

Deubiquitination assays on H2Aub1- or H2Bub1-containing mononucleosomes. Flag–Calypso, Flag–Calypso(C131S), Flag–Calypso–HA–ASX(2–337) or Flag–Calypso(C131S)–HA–ASX(2–337) were mixed on ice with H2Aub1- or H2Bub1-containing mononucleosomes in an enzyme-to-substrate ratio of 1:1 (40 pmol Calypso and 20 pmol H2Aub1 or H2Bub1-containing nucleosome) in 50 μl deubiquitination buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 1 mM DTT) and reactions were incubated at 25 °C, 12.5 μl aliquots (that is, containing 5 pmol nucleosomes) were removed at indicated time points and reactions were stopped by the addition of SDS sample loading buffer and incubation at 95 °C for 5 min.

Deubiquitination assays on K48- and K63-linked ubiquitin polymers. Flag–Calypso, Flag–Calypso(C131S), Flag–Calypso–HA–ASX(2–337) or Flag–Calypso(C131S)–HA–ASX(2–337) were mixed on ice with K48- or K63-linked ubiquitin hexamers (Boston Biochem) in a molar enzyme to substrate ratio of roughly 1:1 in which the molarity of substrate is calculated according to the number of cleavable bonds within the ubiquitin polymers. The reaction was carried out for 40 min at 25 °C in 50 μl deubiquitination buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 1 mM DTT) and was stopped by the addition of SDS sample loading buffer and incubation at 95 °C for 5 min.

Total embryo extract preparation. Embryos (16–18-h old) were dechorionated and immediately incubated in SDS-gel-loading buffer for 5 min at 95 °C. The suspension was sonicated, debris was pelleted and the supernatant was analysed by SDS–PAGE. Embryos that were homozygous for PcG mutations were in each case hand-picked by selecting GFP-negative embryos from strains that contained the mutant chromosome and an appropriate balancer chromosome carrying a *ubi-nGFP* transgene.

Acid-extraction of histones from *Drosophila* embryos. Embryos (16–18-h old) were dechorionated, homogenized and the homogenate was filtered through Miracloth tissue (Calbiochem). Nuclei were pelleted, resuspended in 10 volumes of 0.4 M HCl containing 10% glycerol, and incubated for 30 min on ice. Insoluble material was pelleted, and histones were precipitated from the supernatant with acetone and resuspended in SDS gel loading buffer. For isolation of histones from *Asx*^{22P4}*/Asx*^{22P4} embryos, embryos were collected from a *FRT40 FRT42D y*$^{+}$ *Asx*^{22P}*/CyO ubi-nGFP* strain by selecting for GFP-negative embryos using an embryo sorter (COPAS, Union Biometrica).

29. McGinty, R. K. *et al.* Structure-activity analysis of semisynthetic nucleosomes: mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem. Biol.* **4,** 958–968 (2009).
30. White, R. A. H. & Wilcox, M. Protein products of the Bithorax Complex in *Drosophila*. *Cell* **39,** 163–171 (1984).
31. Celniker, S. E., Sharma, S., Keelanm, D. J. & Lewis, E. B. The molecular genetics of the bithorax complex of *Drosophila*: *cis*-regulation in the *Abdominal-B* domain. *EMBO J.* **9,** 4277–4286 (1990).
32. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193 (2003).
33. Nekrasov, M. *et al.* Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* **26,** 4078–4088 (2007).
34. Fraterman, S., Zeiger, U., Khurana, T. S., Wilm, M. & Rubinstein, N. A. Quantitative proteomics profiling of sarcomere associated proteins in limb and extraocular muscle allotypes. *Mol. Cell. Proteomics* **6,** 728–737 (2007).
35. Müller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111,** 197–208 (2002).
36. Francis, N. J., Saurin, A. J., Shao, Z. & Kingston, R. E. Reconstitution of a functional core polycomb repressive complex. *Mol. Cell* **8,** 545–556 (2001).
37. Levine, S. S. *et al.* The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol. Cell. Biol.* **22,** 6070–6078 (2002).
38. Mason, D. E., Ek, J., Peters, E. C. & Harris, J. L. Substrate profiling of deubiquitin hydrolases with a positional scanning library and mass spectrometry. *Biochem.* **43,** 6535–6544 (2004).
39. Thåström, A. *et al.* Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* **288,** 213–229 (1999).
40. Luger, K., Rechsteiner, T. J. & Richmond, T. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304,** 3–19 (1999).