






The UBP5 histone H2A deubiquitinase counteracts PRCs-mediated repression to regulate *Arabidopsis* development

Received: 15 November 2022

Accepted: 15 December 2023

Published online: 22 January 2024

 Check for updates

James Godwin ^{1,7}, Mohan Govindasamy ^{1,8}, Kiruba Nedounsejian^{1,8}, Eduardo March ¹, Ronan Halton¹, Clara Bourbousse², Léa Wolff², Antoine Fort³, Michal Krzyszton ⁴, Jesús López Corrales⁵, Szymon Swiezewski⁴, Fredy Barneche², Daniel Schubert⁶ & Sara Farrona ¹✉

Polycomb Repressive Complexes (PRCs) control gene expression through the incorporation of H2Aub and H3K27me3. In recent years, there is increasing evidence of the complexity of PRCs' interaction networks and the interplay of these interactors with PRCs in epigenome reshaping, which is fundamental to understand gene regulatory mechanisms. Here, we identified UBIQUITIN SPECIFIC PROTEASE 5 (UBP5) as a chromatin player able to counteract the deposition of the two PRCs' epigenetic hallmarks in *Arabidopsis thaliana*. We demonstrated that UBP5 is a plant developmental regulator based on functional analyses of *ubp5*-CRISPR Cas9 mutant plants. UBP5 promotes H2A monoubiquitination erasure, leading to transcriptional de-repression. Furthermore, preferential association of UBP5 at PRC2 recruiting motifs and local H3K27me3 gaining in *ubp5* mutant plants suggest the existence of functional interplays between UBP5 and PRC2 in regulating epigenome dynamics. In summary, acting as an antagonist of the pivotal epigenetic repressive marks H2Aub and H3K27me3, UBP5 provides novel insights to disentangle the complex regulation of PRCs' activities.

Histones that form the nucleosome, i.e. basic units of the chromatin, are marked by an array of covalent marks, especially on histone amino terminal tails but also on globular domains. Histone marks impact chromatin structure, modify its packaging and act as an anchor for chromatin-related proteins, transcription factors and other components of the transcriptional machinery¹. Therefore, different systems evolved in the eukaryotic nuclei to act as 'writers', able to deposit covalent chemical groups on specific histone residues, 'readers', which can directly bind and help to interpret histone marks, and 'erasers',

actively removing histone post-translational modifications. The orchestration of histone modifying enzymes allows for a highly dynamic chromatin regulation crucial to control nuclear structure and transcription². Two important histone modifications that are well conserved between plants and animals are the trimethylation on the lysine 27 of the histone H3 (H3K27me3)³ and the monoubiquitination of the histone H2A that in plants mostly occurs on the lysine 121 (H2Aub)⁴.

H3K27me3 and H2Aub are deposited, both in plants and animals, by two major types of Polycomb repressive complexes (PRCs),

¹School of Biological and Chemical Sciences, College of Science and Engineering, University of Galway, H91 TK33 Galway, Ireland. ²Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France. ³Dept. of Veterinary and Microbial Sciences, Technological University of The Shannon: Midlands, Athlone, Co., Roscommon, Ireland. ⁴Laboratory of Seeds Molecular Biology, Institute of Biochemistry and Biophysics, PAS, Warsaw 02-106, Poland. ⁵Molecular Parasitology Laboratory (MPL), Centre for One Health and Ryan Institute, School of Natural Sciences, University of Galway, Galway H91 DK59, Ireland. ⁶Institute of Biology, Freie Universität Berlin, 14195 Berlin, Germany. ⁷Present address: Donald Danforth Plant Science Center, St. Louis, MO 63132, USA. ⁸These authors contributed equally: Mohan Govindasamy, Kiruba Nedounsejian.

✉ e-mail: sara.farrona@universityofgalway.ie

respectively PRC2 and PRC1. PRC2 is a four-core subunit complex in which the catalytic component is a SET (Su(var), Enhancer of zeste, Trithorax) domain histone methyltransferase (HMT)^{5,6}. Analyses in different plant genomes showed that PRC2 decorates approximately 20–25% of euchromatic genes with H3K27me3, which switches them off in response to internal and external cues^{7,8}. In plants, PRC1 is formed by E3 ligases and other auxiliary proteins^{5,9}. Both PRCs maintain an intricate relationship in which members of the two complexes can directly interact, have common associated proteins, and share target genes. This is also reflected in their activities as H3K27me3 can precede H2Aub (i.e. hierarchical model) or oppositely follows this modification on the chromatin. Furthermore, both marks can independently regulate different set of genes^{7,9}.

In animals, H2AK119ub can be erased by the Polycomb Repressive-Deubiquitinase (PR-DUB) complex¹⁰. This complex contains a DUB protein of the ubiquitin carboxy-terminal (UCH) family, which does not have an obvious orthologous in plants¹¹. Indeed, the PR-DUB has not been described in plants so far, but two proteins of the UBIQUITIN PROTEASE (UBP) family, UBPI2 and UBPI3 redundantly mediate H2A deubiquitination^{12,13} and interact with LIKE HETEROCHROMATIN PROTEIN 1 (LHP1)¹², a H3K27me3 reader and interactor of both PRC2 and PRC1 components^{7,9}. UBPI2/13 regulate a similar set of genes with PRC2 and PRC1¹³.

To develop their activities, PRCs require a complex network of protein-protein interactions⁷. We and others recently demonstrated that PWWP-DOMAIN INTERACTOR OF POLYCOMBS1 (PWO1) is a key regulator of PRC2 activity, able to interact with the HMTs of the PRC2 complex¹⁴ and to form part of the PEAT complex (PWO/PWWP-EPCRs (ENHANCER OF POLYCOMB RELATED)-ARIDs (AT-RICH INTERACTION DOMAIN-CONTAINING)-TRBs (TELOMERIC REPEAT BINDING)) involved in heterochromatin dynamics¹⁵. Still, we are far from understanding the molecular impact of the PWO1-PRC2 interaction.

Here we show that UBPs regulates plant development and affects both H3K27me3 and H2Aub marks as well as the expression of a set of PRC2 target genes in *Arabidopsis thaliana* (Arabidopsis). Telobox and GAGA motifs, previously related to PRC2 recruitment^{16,17}, are among the most enriched signatures of UBPs binding to the chromatin. The vast majority of UBPs direct target genes showed either hyper-marking or de-novo marking by H2Aub in *ubp5* plants, altogether indicating that UBPs acts as an eraser of this epigenetic mark. Thus, our data uncovers UBPs as a new regulator of PRCs' activities, directly controlling H2Aub deubiquitination and affecting H3K27 trimethylation to regulate gene expression.

Results

UBP5 interacts with PRC2 and PWO1

We had identified the UBIQUITIN PROTEASE 5 (UBP5) protein as the most abundant interactor co-immunoprecipitated with Arabidopsis PWWP-DOMAIN INTERACTOR OF POLYCOMBS1 (PWO1)¹⁸. Furthermore, data mining of proteins in co-immunoprecipitation (co-IP) experiments with PEAT components also identified UBPs¹⁵. Therefore, we aimed to understand the link between UBPs, PWO1 and PRC2. Firstly, to elucidate the sub-cellular localisation of UBPs, transient inducible expression was performed using the β -estradiol-inducible 35S promoter (*i35S*) fused to an *UBP5-GFP* (*i35S::UBP5-GFP*) construct in *Nicotiana benthamiana* (*N. benthamiana*) and found that UBPs is exclusively nuclear, localises all over the nucleoplasm in a diffused way but not in the nucleolus (Fig. 1a). Further, we analysed the possibility of an interaction between UBPs and PWO1 *in planta*. Using a similar approach, we co-expressed *PWO1-GFP* and *UBP5-mCherry* fusion proteins in *N. benthamiana*. It is noteworthy that, as previously shown for CLF^{14,18}, co-expression of both proteins modified UBPs localisation recruiting it to PWO1-containing nuclear speckles. To a lower extent, co-localisation of both proteins was also observed all over the nucleoplasm (Fig. 1b)^{14,18}, but the formation of speckles was never

observed when *UBP5-GFP* was expressed alone (Fig. 1a). PWO1-UBPs association in the speckles was demonstrated by Foster resonance energy transfer with acceptor photobleaching (FRET-APB). FRET-APB efficiencies for co-expressed samples within the speckles were significantly higher than the negative controls (PWO1-GFP and UBPs-GFP expressed without donor mCherry construct) (Fig. 1c), which may indicate a preferential association within the speckles and/or higher contact probability between the two proteins within the speckles. Yeast two-hybrid (Y2H) assays not only confirmed the interaction of UBPs with PWO1 but also revealed its interaction with the PRC2 HMT subunit SWINGER (SWN) Δ SET (SWN clone lacking the SET domain¹⁹); (Fig. 1d). *In planta* interaction between SWN Δ SET and UBPs was further confirmed using co-IP assays in *N. benthamiana* (Fig. 1e). Therefore, our protein-protein interaction results suggest that UBPs is an interactor of PWO1-PRC2 and thus may play a role in PRC-mediated regulation of gene expression. Furthermore, Y2H assays showed interaction of UBPs with EMBRYONIC FLOWER 2 (EMF2), another PRC2 component²⁰, which further confirms the PRC2-UBPs connection (Fig. 1d).

UBP5 is an essential plant developmental regulator

To understand UBPs molecular functions in Arabidopsis, we generated an *ubp5* deletion mutant line via the CRISPR/Cas9 system with two guide RNAs, which partially deleted both DUSP and UBPs conserved domains (Supplementary Fig. 1a–c). The phenotypic analyses of *ubp5* mutant plants showed pleiotropic defects such as stunted growth due to the lack of apical dominance (Fig. 2a (i–iii)), shorter roots and hypocotyl length (Fig. 2a ii and 2b), floral architecture defects (Fig. 2a (v–vi)), fertilisation defects (Supplementary Fig. 1d) and poor pollen germination (Supplementary Fig. 1e). These results fit well with the fact that *UBP5* is expressed all over the plant (Supplementary Fig. 2a), suggesting that UBPs acts as a developmental regulator at different stages of the plant life cycle. Stable transformation of *UBP5pro::gUBP5-eGFP* was able to fully rescue the developmental pleiotropic phenotypes of *ubp5* (Fig. 2a (iv), 2b) and confirmed UBPs nuclear localisation (Supplementary Fig. 2b). qRT-PCR analyses further showed no significant difference in the relative expression of *UBP5* between the wild-type background Col-0 and the complementation line *UBP5pro::gUBP5-eGFP;ubp5* (Supplementary Fig. 3a, b). Transcriptional analyses of *ubp5* seedlings showed that 345 genes were up-regulated, and 478 genes were down-regulated (Fig. 2c; Supplementary data 1). Misregulation of developmental genes including *SAMBA*²¹, *URACIL PHOSPHORIBOSYLTRANSFERASE (UPP)*²², *GAMETOPHYTE DEFECTIVE 1 (GAF1)*²³, *GOLGI CANDIDATE 4 (GCA)*²⁴ and *ACTIN 1 (ACT1)*²⁵ may explain some of the observed *ubp5* mutant phenotypes (Supplementary Fig. 4a–e; Supplementary data 2). Gene Ontology (GO) analyses identified that genes associated with biotic and abiotic stress responses terms were significantly enriched among all *ubp5* mis-regulated genes (Fig. 2d). Consistently with previous studies showing that PRC2-associated components do not only regulate expression of genes related to plant development^{13,26–28}. Therefore, our phenotypic and transcriptomic data highlight UBPs key role in regulating Arabidopsis development and a possible dual role in regulating stress responses that will require further investigation.

UBP5 deubiquitinates H2A

UBPs was shown to be involved in de-ubiquitination of hexa-ubiquitin substrates both *in vivo* and *in vitro*²⁹ and other UBPs family members have been linked to the histone monoubiquitination removal^{12,30,31}. In addition, the existence of the interaction between UBPs, PRC2 components and PWO1 made us speculate that UBPs may contribute to PRC-mediated histone monoubiquitination dynamics. Therefore, we analysed different histone marks abundance in *ubp5* and Col-0 seedlings by western blot (WB) assays and, in good agreement with UBPs acting in H2Aub removal, we found that H2Aub bulk levels were more

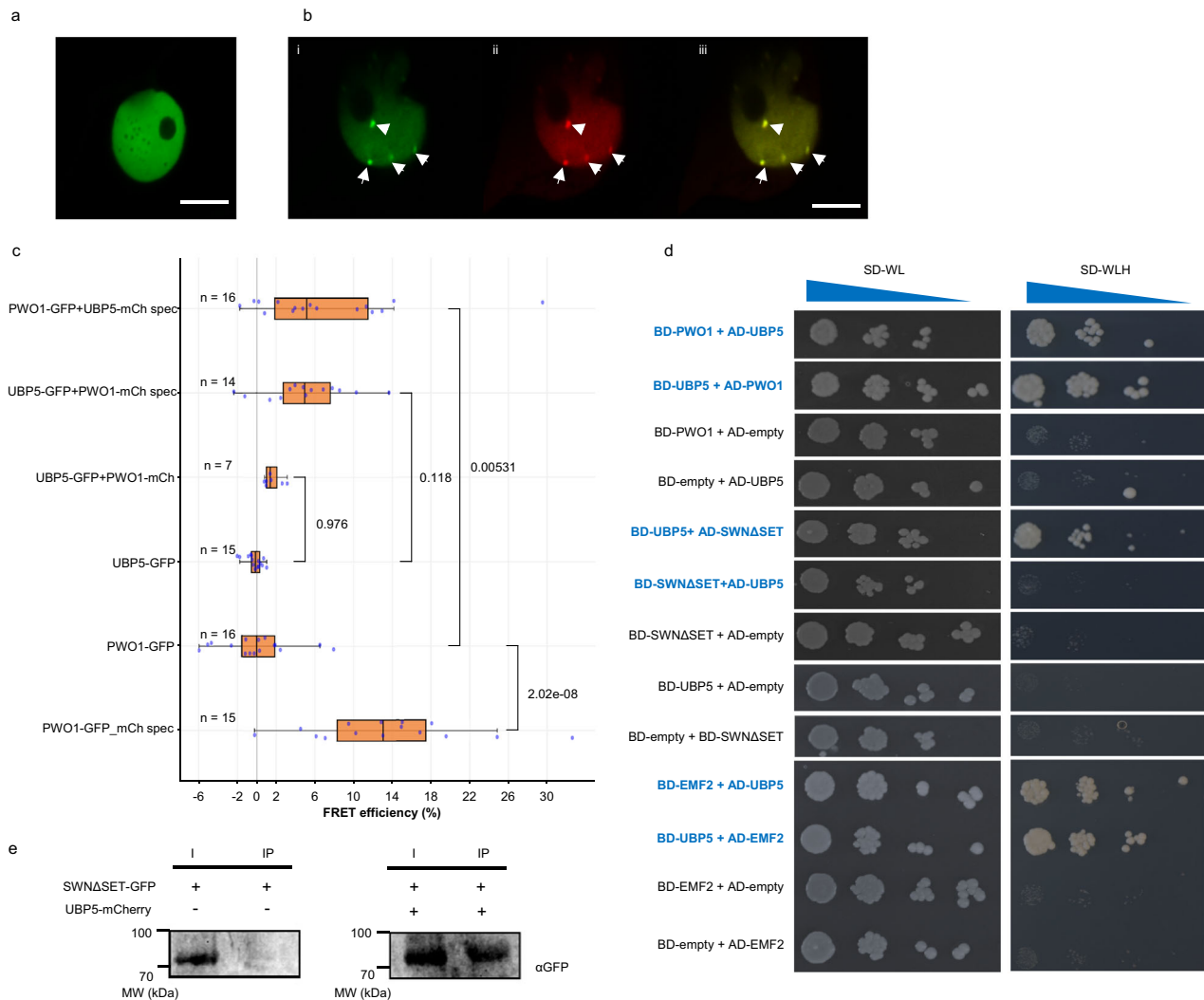


Fig. 1 | UBPS is a nuclear protein that interacts with PRC2 and colocalises with PW01. **a** Localisation of UBPS in the nucleus of *N. benthamiana* epidermal cells. Data represent results of three independent experiments, scale bar indicates 10 μ m. **b** Transient and inducible expression of *i35S::UBP5-GFP* and *i35S::PW01-mCherry* in *N. benthamiana* epidermal cells by co-transformation (i, *i35S::UBP5-GFP*; ii, *i35S::PW01-mCherry*; iii, overlay). Arrows indicate speckles. Data represent results of three independent experiments, scale bar indicates 10 μ m. **c** FRET-APB measurements for nuclei exemplified in **b**, with a distinction for speckle (spec) and non-speckle localisation. *i35S::PW01-GFP-mCherry* measurements in speckles were used as positive control. The respective average efficiency per nuclei (n) is given in the plot. For box plot, the middle line represents the median; the upper and lower lines are the first and third quartile (Q1 and Q3); the whiskers indicate the upper and

lower limits of data spread by subtracting 1.5* interquartile range (IQR) from Q1 and adding 1.5* IQR to Q3. Statistical significance was calculated using one-way ANOVA and the *p*-values are indicated in the figure. Error bars correspond to SE. **d** Y2H analyses confirm UBPS-PW01 interaction and show UBPS-SWN and UBPS-EMF2 interactions. Yeast cells containing the different construct combinations on selective medium for plasmids (-WL; -Tryptophan, Leucine) or for reporter gene activation (-WLH; -Tryptophan, Leucine, Histidine). Serial dilutions were used. BD, GAL4-DNA binding fusion; AD, GAL4-DNA activation domain fusion. SWN Δ SET, SWN construct lacking the SET domain. **e** Co-IP analyses confirming SWN-UBP5 interaction. IP was performed with anti-mCherry antibody and proteins were detected by western blot with anti-GFP. I, 5% input; IP, immunoprecipitation. Two biological replicates were performed.

than 3-fold higher in *ubp5* (Fig. 3a). To gain insight into the affected loci, we profiled the genome-wide distribution of H2Aub in *ubp5* and Col-0 seedlings using ChIP-seq. Our H2Aub data in Col-0 seedlings showed a good overlap with previous published data (Supplementary Fig. 5) and, when compared to Col-0 seedlings, we observed a large increase in the number of genes marked by H2Aub in *ubp5* (21,017 in *ubp5* instead of 15,615 genes in Col-0; Supplementary data 3-4), which includes genes that differentially gained H2Aub in *ubp5* ($n = 6201$, called from now on de-novo marked genes; Fig. 3b), hence UBPS is necessary to erase H2Aub in several thousands of genes. We then decided to conduct a comprehensive analysis on the H2Aub marked genes common to Col-0 and *ubp5* ($n = 14816$) and identified three different scenarios: genes that showed lower H2Aub levels in *ubp5* compared to Col-0 plants ($n = 7150$, hypo-marked genes), genes with higher H2Aub signal in *ubp5* compared to Col-0 ($n = 3055$, hyper-

marked genes) and genes with non-significant differences between *ubp5* and Col-0 ($n = 4611$, unchanged genes) (Fig. 3c; Supplementary data 5).

To test whether UBPS could act in H2Aub removal in *cis*, we further analysed the genome-wide association of UBPS-GFP in our *UBP5-pro::gUBP5-eGFP;ubp5* line. Notably, UBPS binding extends to a large part of the plant genome since the UBPS-GFP ChIP-seq profiling identified 8983 genes as direct UBPS targets (Supplementary Fig. 6a-c; Supplementary data 6), which corresponds to ~27% of the total number of Arabidopsis genes according to TAIR 10 annotation³². More precisely, UBPS directly targets 61% of hyper-marked genes, but only 9% of hypo-marked genes. In addition, among de-novo marked genes, 57% are also direct UBPS targets. Therefore, a statistically significant number of UBPS target loci gain the H2Aub mark in *ubp5* (1876 hyper-marked genes plus 3540 de-novo marked genes, 60% of UBPS targets)

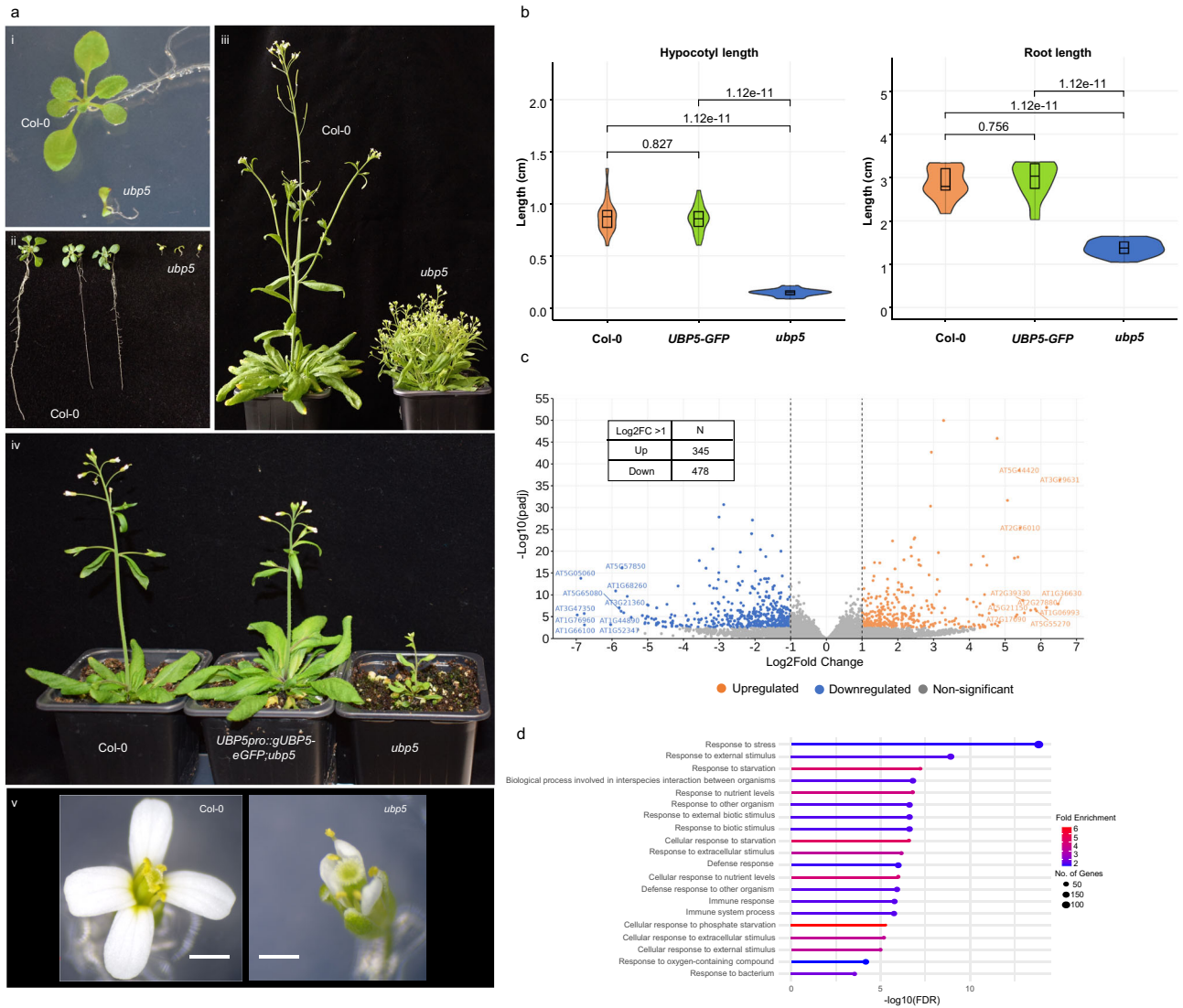


Fig. 2 | UBPS is an essential plant developmental regulator. a Phenotypic characterisation of *ubp5* mutant line: **i**, smaller seedlings; **ii**, shorter primary roots; and **iii**) stunted and bushy growth (Note: in **iii** *ubp5* plant was 2 weeks older than the *Col-0* plant) compared to *Col-0* plants; **iv**, complementation of *ubp5* mutant phenotypes in 4-week-old *Arabidopsis* plants (an *UBP5pro::gUBP5-eGFP* construct was used for the complementation of *ubp5*); **v**, floral phenotype of *Col-0* and **vi**, *ubp5*. Scale bar = 2.5 mm. **b** Violin cum box plot showing hypocotyl and root length of *Col-0*, *UBP5-GFP* (complementation line) and *ubp5* measured 10 days after germination. The values were obtained from twenty independent seedlings ($n = 20$). The violin plots show the distribution pattern of data and are overlaid with boxplots. For box plots, the middle line represents the median; the upper and lower lines are the first and third quartile (Q1 and Q3); the whiskers indicate the upper and lower limits of

data spread by subtracting 1.5* interquartile range (IQR) from Q1 and adding 1.5* IQR to Q3. Statistical significance was calculated with one-way ANOVA and Tukey HSD and *p*-values are indicated. **c** Volcano plot of the misregulated genes between *Col-0* vs *ubp5*. Genes with a *p* adjusted value (*padj*) lower than 0.05 are coloured. Up and down-regulated DEGs are shown in orange and blue colour dots respectively. The non-significant DEGs are represented as grey dots. The genes with $\text{Log}_2\text{FC} < 1$ or $\text{Log}_2\text{FC} > -1$ (False Discovery Rate (FDR) < 0.05) were considered for further analysis and 'n' denotes the number of genes. FDR values are derived from DESeq2 by adjusting *p*-values using Benjamini-Hochberg method. **d** Functional categorisation of *ubp5* misregulated (both upregulated and downregulated) genes based in ShinyGO v0.75 analysis. GO analysis of *ubp5* misregulated genes based on biological process (FDR < 0.05).

(Fig. 3d), pointing the key role of UBPS in regulating H2Aub removal. Importantly, there is a sharp co-localisation between UBPS chromatin association and domains where the H2Aub mark was gained in *ubp5* (Fig. 3e (i–ii); Supplementary Fig. 6a and 7a, b; Supplementary data 5). This frequent co-occurrence strongly argues in favour of a direct role for UBPS in H2Aub deubiquitination at its binding sites (Fig. 3f). Further supporting this observation, there is an anticorrelation between the intensity of UBPS binding and the presence of the H2Aub mark in *Col-0* plants (Fig. 3f). In addition, an increase in H2Aub levels in *ubp5* is more evident at UBPS target genes than for other, non-targets, H2Aub marked genes, in which even a slight significant decrease in the mark is observed (Fig. 3g and Supplementary Fig. 7c). Furthermore, overall increase of H2Aub levels was restored to *Col-0* levels in the *UBP5pro::*

gUBP5-eGFP;ubp5 line (Supplementary Fig. 7d) and inducing the expression of the *i35S::UBP5-GFP* construct in the nuclei of *N. benthamiana* leaves strongly decreased H2Aub levels (Fig. 3h). To confirm these observations, selected UBPS targets that are H2Aub hypermarked in *ubp5* were further validated by ChIP-qPCR (Supplementary Fig. 7e). Overall, these results indicate that UBPS acts in *cis* on the H2Aub mark by both maintaining H2Aub levels in a set of genes marked with this modification and erasing the H2Aub mark from a larger set of genes.

UBPS plays a role in transcriptional de-repression

Functional categorisation of UBPS direct targets revealed that genes related to chromosome organisation, histone binding and chromatin

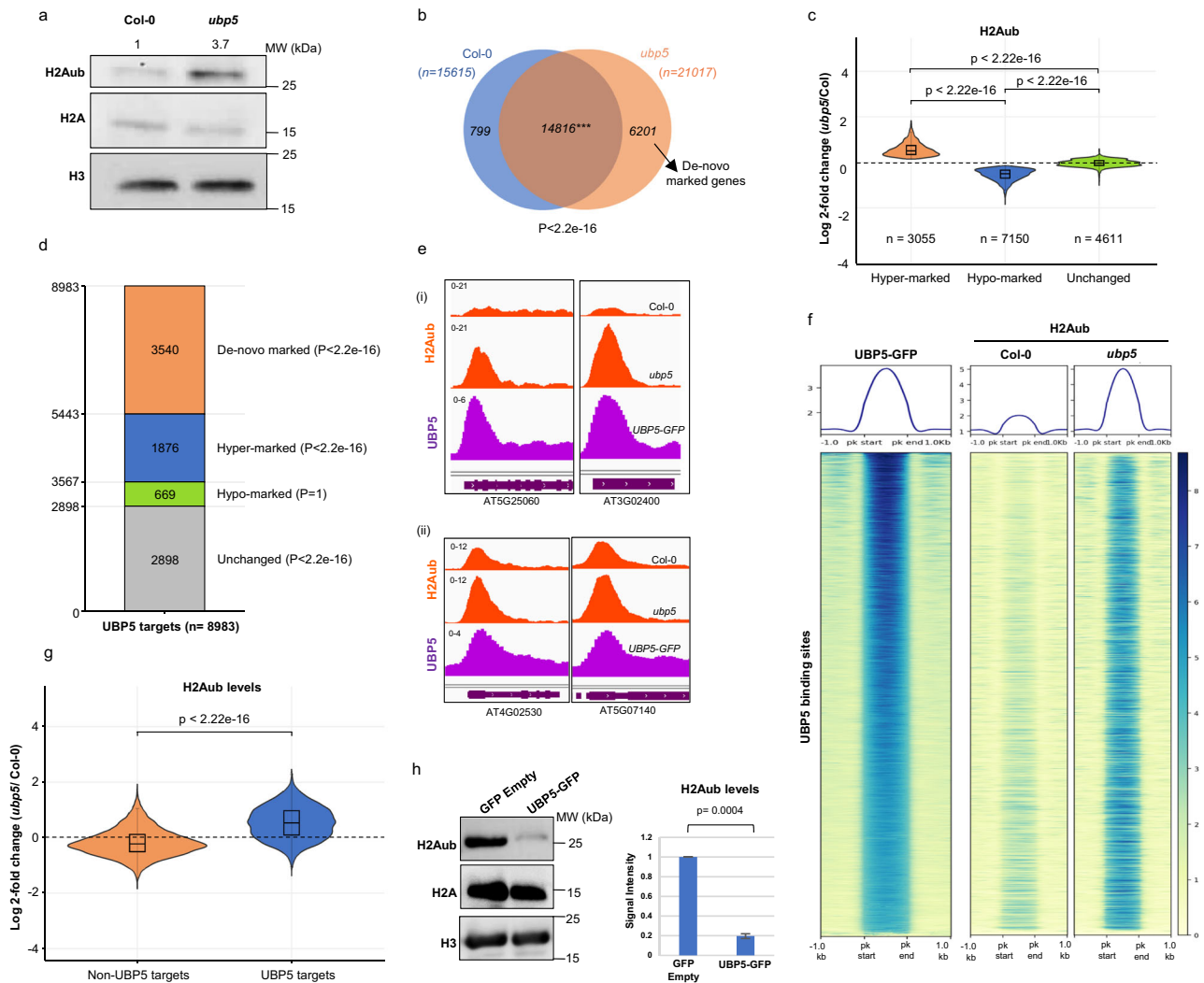


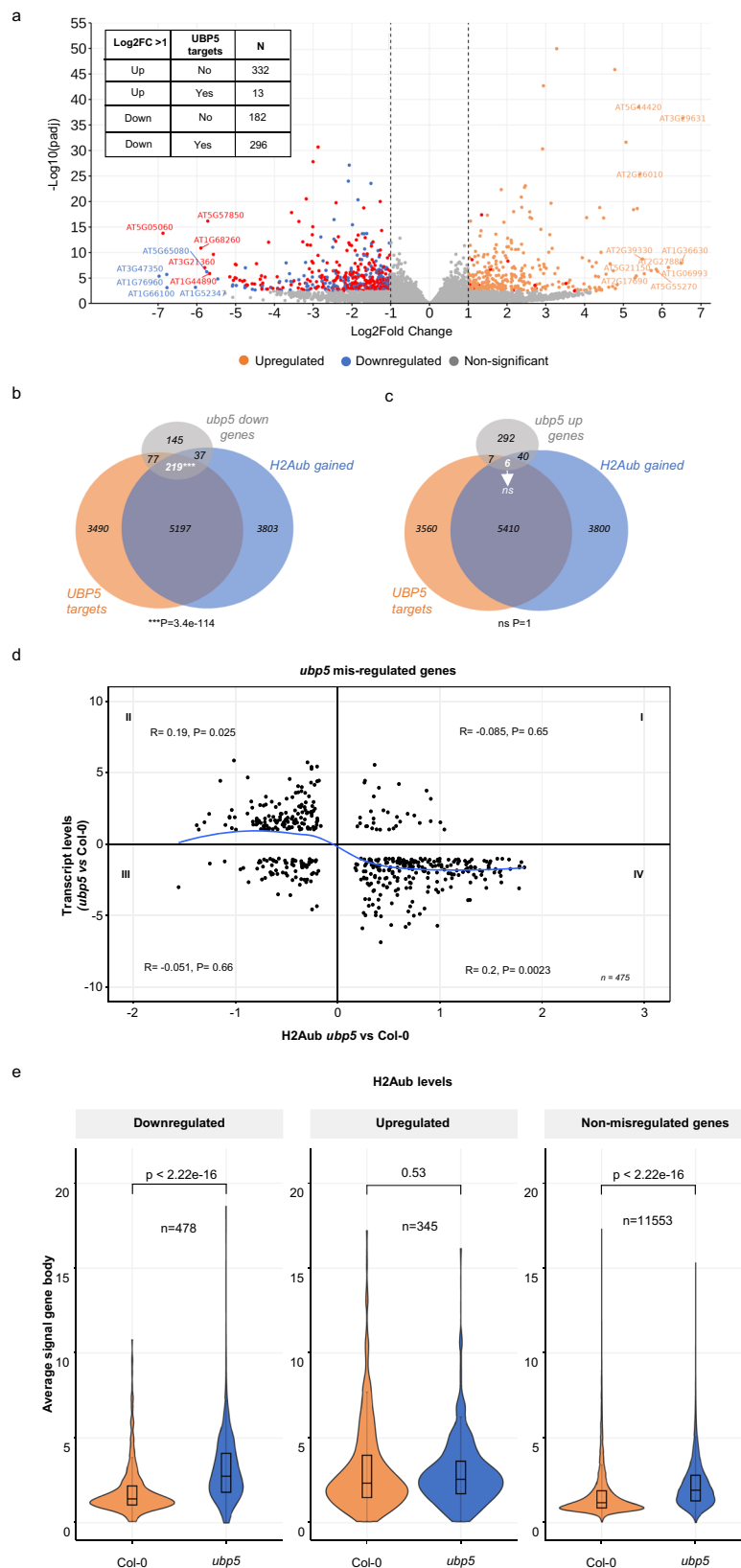
Fig. 3 | UBPS acts as a H2A deubiquitinase. a Western blot analysis of H2Aub and H2A in Col-0 and *ubp5* seedlings. Histone H3 is used as a loading control. Numbers above blots denote relative H2Aub levels determined by ImageJ. Three independent experiments yielded consistent results. **b** Venn diagram illustrates overlap between H2Aub marked genes in Col-0 and *ubp5*, n = no. of genes. The genes are considered as marked when an overlapping H2Aub peak is present in at least two biological replicates based on MACS3 peak calling ($q < 0.05$ and score > 30). Super exact test was conducted. **c** H2Aub levels in genes overlapping between Col-0 and *ubp5* ($n = 14816$) from Fig. 2B were categorised into hyper/hypo-marked and unchanged. Violin plots overlaid with boxplots show the distribution pattern of data. Displaying median (middle line), first and third quartiles (Q1 and Q3), with whiskers indicating data spread ($1.5 \times$ IQR from Q1 and Q3). Statistical significance was tested using the Kruskal-Wallis test followed by Dunn's post hoc analysis. **d** Bar chart shows the number of H2Aub hyper-, hypo-, de-novo- and unchanged-H2Aub

marked genes represented within UBPS target genes. Significance tested using Super Exact test. **e** IGV browser views of representative UBPS target loci of (i) de-novo marked genes in *ubp5* and (ii) H2Aub hyper-marked genes in *ubp5*. **f** Heatmaps showing H2Aub distribution on genomic sequences targeted by UBPS, clustered based on higher to lower enrichment (top to bottom). **g** Average signal of H2Aub levels at non-UBPS targets in Col-0 and *ubp5*. The violin cum box plots display median (middle line), first and third quartiles (Q1 and Q3), with whiskers indicating data spread ($1.5 \times$ IQR from Q1 and Q3), two-sided Wilcoxon rank-sum test was performed. **h** H2A deubiquitination activity of UBPS. Transient expression of *i35S::UBPS-GFP* and *i35S::GFP* empty in *N. benthamiana* leaves. H2A deubiquitination was assessed by immunoblotting with α -H2Aub antibody; controls: α -H2A and α -H3 antibodies. Bar graph representing the relative H2Aub levels derived from band intensity. Error bars indicate SD between two biological replicates with two-tailed unpaired *t*-test, p values indicated above plot.

binding were significantly over-represented (Supplementary Fig. 8a–c). In addition to UBPS direct protein-protein interaction with PWO1 and SWN chromatin factors, we found that UBPS is also able to bind the chromatin of several loci encoding PRC2 subunits such as *CLF*, *EMF2*, *VERNALIZATION 2 (VRN2)*, *FERTILIZATION-INDEPENDENT ENDOSPERM (FIE)* and *MULTICOPY SUPPRESSOR OF IRA 1 (MSII)* as well as PRC1 subunit encoding gene *B LYMPHOMA Mo-MLV INSERTION REGION ONE HOMOLOG (BMIIB)* (Supplementary data 5). H2Aub mark was also gained in these genes (Supplementary data 5), although for most of the genes we did not observe transcriptional changes in *ubp5* under our analysed conditions. On the other hand, GO analyses of UBPS target genes that gained the H2Aub mark in *ubp5* revealed a

significant over-representation of genes involved in response to DNA damage and repair (Supplementary Fig. 8d).

At the genome-wide level, UBPS binding to chromatin typically occurs at the proximity of the transcription start site (TSS) and the start of the coding region (Supplementary Fig. 6a). Analyses of UBPS binding peaks showed that majority of these sites correspond to protein coding genes, particularly exons and 5'UTRs that respectively correspond to -51% and -23% of the binding sites (Supplementary Fig. 9). Hence, we evaluated the impact of UBPS in the transcriptional output of its target genes by integrating our ChIP-seq and RNA-seq data. We found a clear link between UBPS gene binding and repression, as a significant -62% (296/478) of the genes downregulated in *ubp5*



correspond to UBP5 targets, whereas only a few UBP5 target genes were upregulated (13/345 genes) (Fig. 4a). In addition, 52% of down-regulated genes in *ubp5* (219/478) were both UBP5 targets and gained H2Aub in the mutant, but these two conditions were almost never found associated to upregulated genes (Fig. 4b, c). More generally,

ubp5 related defects in transcription are associated to an increase in H2Aub (Fig. 4d). Figure 4e shows that down-regulated genes in *ubp5* have a significant probability to have increased H2Aub levels, while non-significant changes are observed in *ubp5* upregulated genes. Nevertheless, an increase in H2Aub does not always result in

Fig. 4 | UBP5 mediates transcriptional de-repression. **a** Volcano plot representing the UBP5 target genes which were downregulated and upregulated in *ubp5*. Up- and down-regulated DEGs are shown in orange and blue color dots respectively. Non-significant genes are shown as grey dots and UBP5 target genes as red dots. FDR < 0.05 and FDR values are derived from DESeq2 package by adjusting *p*-values using Benjamini-Hochberg method. **b** & **c** Venn diagrams showing overlap between UBP5 targets, H2Aub gained genes (hypermarked + de-novo) in *ubp5* and (**b**) downregulated or (**c**) upregulated genes in *ubp5* mutant. Super exact test was performed to check the statistical significance (***) represents significance with the *p* value = 3.4e-114, ns signifies non-significance). **d** Scatter plot showing the correlation between H2Aub and gene expression changes between Col-0 and *ubp5* plants. The x-axis shows Log₂FC levels of H2Aub marked genes (FDR < 0.05) and y-axis shows expression in Log₂FC of misregulated genes in *ubp5* as determined by

DESeq2 (>1 fold variation, FDR < 0.05). For each quadrant, the correlation coefficient (R) along with the significance (*p*-values) are shown. The blue curve shows trend-line from LOWESS smoother function. Quadrant IV shows higher correlation between low expressed genes and hyper-marking of H2Aub. **e** Violin cum box plots represents the average signal of H2Aub at gene body for downregulated, upregulated genes and non-misregulated genes in Col-0 and *ubp5*. The violin plots show the distribution pattern of data and are overlaid with boxplots. For box and whiskers plots the middle line represents the median; the upper and lower lines are the first and third quartile (Q1 and Q3); the whiskers indicate the upper and lower limits of data spread by subtracting 1.5* interquartile range (IQR) from Q1 and adding 1.5* IQR to Q3. Statistical significance was calculated with two-sided Wilcoxon rank sum test, *p*-values are indicated above the plot.

transcriptional changes as a slight but still significant increase in H2Aub was generally observed in *ubp5*.

Taken together, our transcriptomic and epigenomic data suggest a role of UBP5 in relieving H2Aub-mediated repression, thereby promoting gene expression. Therefore, UBP5 seems to be predominantly involved in H2Aub erasure, which, at least for a significant set of its targets genes, results in transcriptional de-repression.

UBP5-mediated H2A deubiquitination prevents deposition of H3K27me3

To explore whether UBP5 is targeted to chromatin in a sequence-specific manner, we analysed sequence motifs at UBP5 binding sites using MEME-ChIP³³ and identified a significant over-representation of GAGA and Telobox motifs (Fig. 5a). Notably, GAGA elements recognised by transcription activators/repressors and Telobox motifs typically recognised by TRBs are involved in recruiting PRC2^{17,34,35}. These results thus suggest the existence of sequence-specific mechanisms commonly recruiting UBP5 and PRC activity.

Therefore, to further unravel the relationship between UBP5 function and PRC2 activity, we analysed H3K27me3 bulk level by WB analysis and identified a 70% increase in its abundance in *ubp5* (Fig. 5b). We conducted ChIP-seq to further determine the genome-wide effects of UBP5 on H3K27me3. Our data showed a high overlap of H3K27me3 marked genes in Col-0 seedlings with previously published data³⁶ (Supplementary Fig. 10a; Supplementary data 3). Differential analysis of our H3K27me3 genome-wide data revealed 2779 gaining the mark (i.e. hyper-marked) and 2349 H3K27me3 hypo-marked genes in *ubp5* (Supplementary Fig. 11a; Supplementary data 7). Notably, at UBP5 target genes average H3K27me3 level was higher in *ubp5* (Fig. 5c–e), while no increase was observed in non-UBP5 targets; instead, only a slight but statistically significant decrease was noted (Supplementary Fig. 11b, c). Further analyses of ChIP-seq data based on differential analysis showed that in 640 genes the following conditions concurred: i) H3K27me3 and ii) H2Aub gained in *ubp5*, and iii) directly bound by UBP5 (Fig. 5f, g; Supplementary Fig. 10b), indicating that UBP5 not only erases H2Aub but also impairs the deposition of H3K27me3 at multiple sites. On the other hand, less than 4% of H3K27me3 hypo-marked genes were UBP5 targets (Supplementary Fig. 10c), suggesting that UBP5 may not play a direct role in H3K27me3 maintenance at these genes and therefore these changes might likely result from indirect effects in the regulation of H3K27me3 writers' or erasers' activity.

In agreement with a repressive role of H3K27me3 marking, average H3K27me3 levels in the gene body of *ubp5* downregulated genes was significantly higher than Col-0 levels, and there were no significant changes in the upregulated genes (Fig. 5h) and, similarly, there is a negative relationship between H3K27me3 and transcript levels in *ubp5* with almost all the down-regulated genes enriched with H3K27me3 (Fig. 5i). Nevertheless, as we have observed for H2Aub, gaining H3K27me3 does not always result in transcriptional changes as a subtle but significant H3K27me3 increase in non-misregulated *ubp5* genes

was observed (Fig. 5h). In summary, the combined analysis of our transcriptomic and epigenomic data points towards the activity of UBP5 in de-repressing a set of its target genes by preventing H3K27me3 enrichment.

To understand how both H2Aub and H3K27me3 dynamics affect the transcriptional levels of genes we focussed on the set of genes which gained H2Aub in *ubp5*. In this set of genes, we analysed the transcriptional levels of H3K27me3/H2Aub marked genes in both Col-0 and *ubp5* and found that in both backgrounds, genes that are exclusively marked by H2Aub are more highly expressed than genes with the two marks or only H3K27me3, as previously shown (Zhou et al., 2017). On the other hand, while in Col-0 plants there is a significant difference in transcriptional levels of H2Aub/H3K27me3 versus H3K27me3 marked genes, this difference is lost in *ubp5* with both categories showing similar repressive levels (Supplementary Fig. 10d). Hence, UBP5 may contribute to pose H2Aub/H3K27me3 marked genes in a more responsive chromatin structure. Overall, we thus conclude that in the subset of 640 genes, UBP5-mediated H2Aub deubiquitination prevents the deposition of H3K27me3 mark leading to a de-repressed chromatin environment (Fig. 6).

Discussion

PRC2 interactors play a key role in regulating its molecular activities and recruitment to chromatin⁷. For instance, we previously showed that PWO1 may mediate in providing PRC2 with the right chromatin environment to methylate H3¹⁴. In addition, PWO1 was proposed to form part of the PEAT complex¹⁵. Therefore, unravelling the protein interactors associated with epigenetic pathways can provide important clues to understand their possible crosstalk and activities. UBP5 was identified co-immunoprecipitating with all main components of PEAT¹⁵ and, more recently, in a paper by Zheng et al. that was published during the revision of our manuscript, its role as a PEAT component has been confirmed. Notably, UBP5 molecular activity depends on its interaction to PWO1 and nucleosomal DNA³⁷. Here, we have shown that co-expression of *PWO1* and *UBP5* in *N. benthamiana* tethers UBP5 to nuclear speckles. Hence, it is tempting to speculate that PWO1-UBP5 interaction is required for its recruitment to specific chromatin regions. Furthermore, our results indicate that UBP5 and PWO1, two components of the PEAT complex, may also interact with PRC2¹⁴. In addition, UBP5 binding sites are enriched in Telobox motifs previously involved in PRC2 recruitment to the chromatin by TRBs¹⁷. Strikingly, TRBs also form part of PEAT¹⁵. Considering the evidence, a very intriguing hypothesis is that PRC2 and PEAT coordinate their activities to dynamically regulate chromatin (Fig. 6). Nevertheless, while the PWO1-UBP5 interaction was observed in plant extracts¹⁸ and has now been confirmed by our protein-protein interaction analyses, the UBP5-PRC2 binding has so far only been tested through heterologous systems (i.e. Y2H and co-IP in *N. benthamiana*). Hence, future *in planta* analyses will aid to confirm this interaction and to disentangle the complexity of UBP5's interacting network.

UBP5 belongs to the UBP family, which is part of the conserved DUB superfamily. Several DUBs are involved in the regulation of chromatin and some of them especially in H2A deubiquitination¹¹. For instance, *Drosophila* protein Calypso as well as its corresponding ortholog in humans, the tumour suppressor BRCA-1-associated protein 1 (BAP1), form part of a PR-DUB complex able to remove the H2AK119ub1 mark. Intriguingly, PR-DUB has been described as a type of PRC despite its opposite activity to PRC1. Therefore, it seems that a dynamic ubiquitination/deubiquitination counterbalance is key for maintaining PRCs' activities and proper H2A ubiquitination levels over the genome^{38–40}. The only proteins that have been related to H2A deubiquitination in Arabidopsis are the closely related UBPI2 and I3 proteins, which were identified interacting with LHPI¹², a protein that may act as an accessory protein in both PRC2 and PRC1⁷. UBPI2 was shown to be involved in the repression of a subset of PRC2 targets mediating H3K27me3 deposition and to be actively involved in H2A deubiquitination¹². However, UBPI5 has a much broader impact on H2Aub than the one so far described for UBPI2/I3 (Supplementary Fig. 12). Furthermore, UBPI2/I3-mediated H2Aub removal prevents loss of H3K27me3¹³. In contrast, our data indicate a role of UBPI5 in preventing H3K27me3 gain at specific loci (Fig. 6). Finally, the genes that are regulated by UBPI2/I3 (i.e. H2Aub gained genes in *ubp12/13*) and UBPI5 direct targets show little overlap (Supplementary Fig. 12), suggesting that they act through independent mechanisms or at different genome domains. However, as UBPI2/I3 direct target genes have not been described so far, this conclusion needs to be cautiously considered as indirect results in *ubp12/13* epigenomic data cannot be discarded¹³.

UBPI5 closest human orthologs are USP4, USP11 and USP15⁴¹. USP11 acts in both H2AK119 and H2BK120 deubiquitination and specifically deubiquitinates the histone variant γ H2AX, which is key in homologous recombination⁴¹, although no epigenomic data is available for the activity of these human proteins. Our ChIP-seq profiling in seedlings identified that UBPI5 is required for H2Aub deubiquitination at a majority of PRC1-regulated Arabidopsis genes, and, considering *ubp5* phenotypes, UBPI5 may have additional effects on H2Aub epigenome at other developmental stages. H2Aub ChIP-seq profile also points to a dual role of UBPI5 deubiquitination activity. In ~20% of genes showing a H2Aub gain in *ubp5*, UBPI5 acts to maintain a certain level of H2Aub in the plant; while, in ~40% of this set of genes, UBPI5 fully erases this histone mark (Fig. 3d). Overall, these results indicate that UBPI5 acts in *cis* to maintain the right H2Aub level at target genes with two possible scenarios for each locus: this modification is either 1) erased by UBPI5 in most cells and therefore not detected in Col-0 plants but only in *ubp5* (i.e. de novo marked genes) or 2) stably present in Col-0 seedlings but removed by UBPI5 only in certain genome copies or in certain cells (i.e. H2Aub hypermarked genes). Further studies will be required to fully understand how UBPI5 discerns between these different scenarios. Most UBPI5 targets gained H2Aub in *ubp5* plants (i.e., 60% of UBPI5 targets). However, a striking number of non-UBPI5 targets were hypomethylated suggesting that UBPI5 also indirectly reshapes the nuclear space. Understanding direct and indirect UBPI5's impact on chromatin may help to discover new molecular mechanisms controlling epigenetic regulation.

Mirroring the meta-gene pattern of H2Aub in Arabidopsis³⁶ (Fig. 3d), UBPI5 predominantly binds to chromatin in the vicinity of TSSs and at the start of protein coding regions. Furthermore, our transcriptional analyses show that UBPI5 target genes tend to be downregulated in the *ubp5* mutant. These results point to UBPI5 acting as a transcriptional activator, as shown for H2A deubiquitination in animals⁴². As UBPI5 acts in histone deubiquitination, we favour the possibility of its active role in promoting transcriptional de-repression through the erasure of H2Aub as it has been proposed for other erasers (e.g. histone demethylases⁴³). However, gain of H2Aub in *ubp5* is not always synonymous of changes in transcription in a comparable way as accessible chromatin is not always leading to activation⁴⁴.

The relationship between H2A and H3K27me3 deposition has been addressed in plants and metazoans⁴⁵, but much less is known about any possible crosstalk between the removal of these two marks. In line with the UBPI5-PRC2 protein interaction identified here, UBPI5 influences H3K27me3 levels at a majority of H3K27me3-marked genes (i.e., 5128 genes out of 7600 genes), ~20% of them corresponding to direct UBPI5 target sites at the seedling stage (i.e., 1028 genes out of 5128 genes). For these genes, deposition of H2Aub plausibly precedes H3K27 trimethylation on the same nucleosome, as suggested for several PRC1/PRC2 target genes⁴⁶, and hence UBPI5-mediated H2A deubiquitination will prevent H3K27me3 deposition by PRC2 (Fig. 6), probably making chromatin more accessible in these loci and leading to de-repression. Although further investigation will be required to fully address the UBPI5-PRC2 relationship, a sequential mechanism in which the putative interaction between UBPI5 and SWN is needed for 1) removal of H2Aub and 2) prevention of H3K27me3 deposition could explain how in a set of PRC2 target genes the removal of both marks is coordinated (Fig. 6). If prevention of H3K27me3 deposition occurs through UBPI5-direct impairment of SWN activity/recruitment/stability and/or requires the activity of demethylases are attractive hypotheses to test in the future. Our proposed functional model also fits well with evolutionary results linking the deposition of H3K27me3 to the ubiquitination of H2A in *Marchantia polymorpha*⁴⁷. Despite all our results suggest an UBPI5-PRC2 interaction, we should not forget that many UBPI5 target genes that are enriched in H2Aub do not gain H3K27me3, indicating that UBPI5 plays PRC2-independent functions uncoupling the removal of the two marks. This supports the possibility that PEAT can carry diverse activities or that different versions of PEAT can co-exist in the plant for a versatile chromatin regulation. In fact, Zheng et al. recently showed that PEAT is required for both H2A deubiquitination by UBPI5, confirming our results, and H4K5 acetylation by HISTONE ACETYLTRANSFERASE OF THE MYST FAMILY (HAM) proteins³⁷. Here, in addition to its H2A deubiquitination activity, we have demonstrated that UBPI5 is able to antagonistically regulate H3K27me3 deposition and may interact with PRC2 components. Therefore, these different results may indicate the multifunctionality of PEAT to activate transcription. It also opens further fascinating questions about UBPI5 alternative activities in the regulation of chromatin accessibility that we look forward to answering in future studies.

Methods

Plant materials and cultivation conditions

All *Arabidopsis thaliana* (Arabidopsis) lines used in this study were in the Columbia-0 (Col-0) ecotype background. For the generation of *ubp5* CRISPR-Cas9 mutant, double guide system of Cas9-directed mutagenesis was performed⁴⁸ to delete a fragment size of 3361 bp from UBPI5 gDNA sequence (Supplementary Fig. 2A). sgRNAs were designed using CRISPR-P tool⁴⁹. The P3-Cas9-mCherry vector for generating the *ubp5* line was kindly provided by Charles Spillane⁴⁸. Deletion of the genomic fragment from UBPI5 was confirmed using Sanger sequencing (LGC genomics, Germany). Transgenic plants were developed by *Agrobacterium*-mediated gene transformation with floral dip method⁵⁰. For genotyping, DNA extraction was done based on this described protocol⁵¹. Oligonucleotide primers used for CRISPR-Cas9 mutagenesis and genotyping are indicated in Supplementary Table 1. For the *UBP5pro::gUBP5-GFP;ubp5* line, a 1708-kb-upstream fragment and gene-body regions of UBPI5 without stop codon were amplified from genomic DNA of Col-0 with GW-compatible primers (Supplementary Table 1). *UBP5pro::gUBP5* was fused with a C-terminal GFP sequence in the (pGKGWG) vector⁵².

Sterilised seeds were sown on Murashige & Skoog medium (MS Base) supplemented with 1% Sucrose, 0.1% MES, 0.8% agar with pH

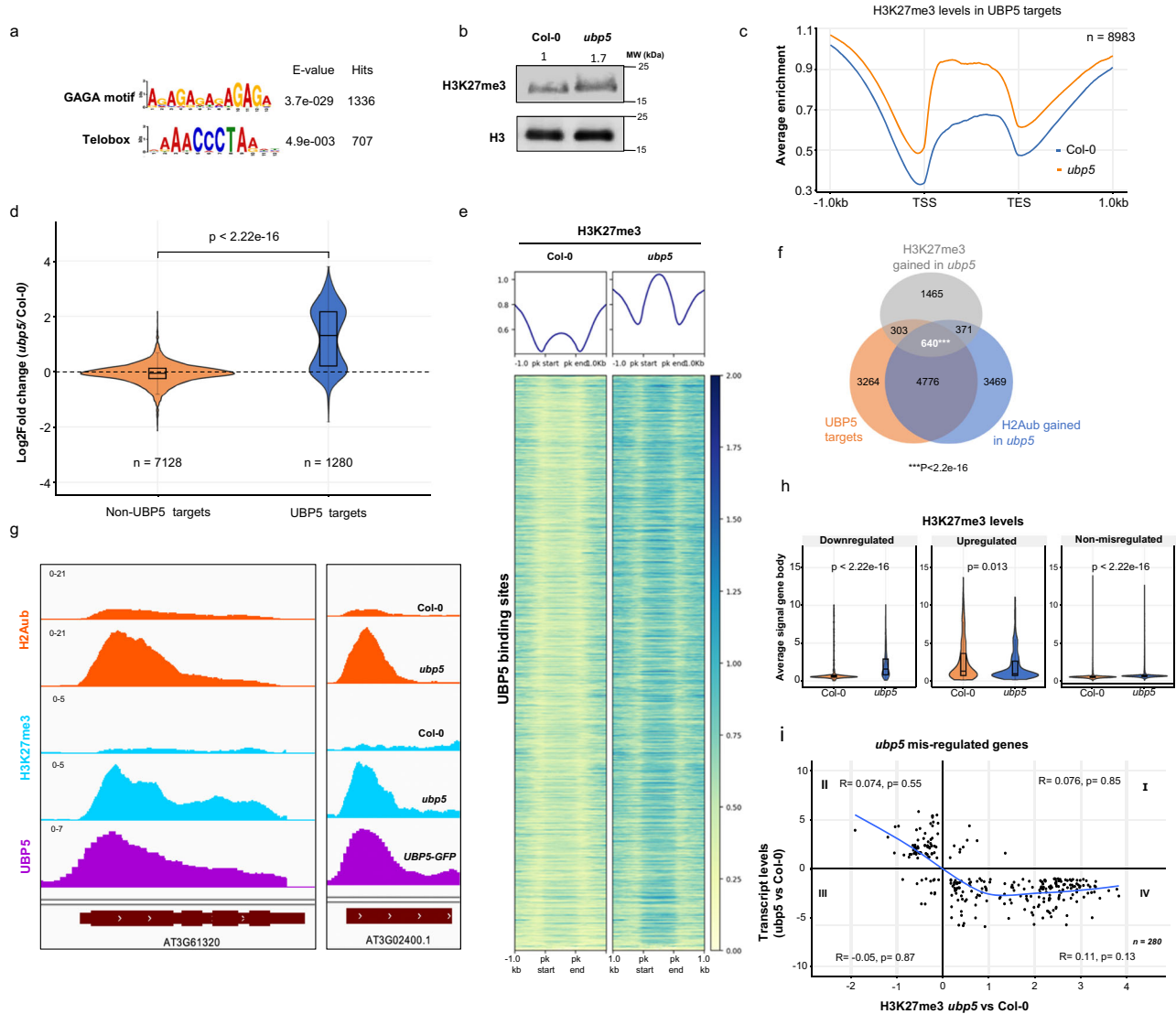


Fig. 5 | UBPs-mediated H2Aub deubiquitination prevents deposition of H3K27me3. **a** Motif enrichment analysis of UBPs target genes. The sequence logos, accuracies and hits of the best motifs found by MEME-ChIP. **b** H3K27me3 levels in *ubp5* seedlings, Histone H3 used as a loading control. Numbers above blots indicate relative H3K27me3 levels. Three independent experiments were performed. **c** Metagen plot of average H3K27me3 enrichment over the UBPs target genes. **d** H3K27me3 ChIP seq signal in *ubp5*/Col-0 between UBPs targets and non-UBPs targets. Violin overlaid boxplots display median (middle line), first and third quartiles (Q1 and Q3), with whiskers indicating data spread (1.5*IQR from Q1 and Q3). Statistical significance was determined by two-sided Wilcoxon rank sum test. **e** Heatmap showing the distribution of H3K27me3 on UBPs binding sites for Col-0 and *ubp5*. UBPs binding peaks are clustered based on higher to lower enrichment (top to bottom). **f** Venn diagram representing the overlap between UBPs targets, H2Aub and H3K27me3 gained genes in *ubp5* (FDR < 0.05). Super exact test is

performed to test the overlap (***) represents significant overlap with $p < 2.2e-16$). **g** IGV browser snapshots of representative UBPs target genes in which H2Aub and H3K27me3 are gained in the *ubp5* mutant. **h** Average signal of H3K27me3 at gene body for downregulated, upregulated and non-misregulated genes. Median (middle line), first and third quartiles (Q1 and Q3), with whiskers indicating data spread (1.5*IQR from Q1 and Q3). Statistical significance was determined by two-sided Wilcoxon rank sum test. **i** Scatter plot shows correspondence between H3K27me3 levels and gene expression changes between Col-0 and *ubp5* plants. The x-axis shows Log2FC levels of H3K27me3 marked genes by DESeq2 analysis (FDR < 0.05), y-axis shows expression of misregulated genes in *ubp5* (Log2FC > 1, FDR < 0.05). Blue curve shows trend-line from LOWESS smoother function, with Pearson correlation coefficient (R) and significance (p values) displayed. Quadrant IV shows significant enrichment of low expressed genes that have gained H3K27me3.

adjusted to 5.6, stratified at 4 °C for three days and placed in a Percival tissue culture cabinet under a 16:8 h light: dark (21 °C/18 °C) regime until they were transferred to soil. Arabidopsis plants were grown on pots containing compost, vermiculite and perlite (5:1:1 proportion) with the same photoperiod under fluorescent lamps at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$. For hypocotyl and root length measurements, Col-0 and *ubp5* seeds were sown on MS medium, and the plates were placed vertically in the growth chamber in LD conditions. Photographs were taken at the end of 10 days, hypocotyl and root length were measured using the Fiji image processing software.

Yeast two hybrid assay

For yeast two hybrid assays, untransformed *Saccharomyces cerevisiae* AH109 cultures were grown at 28 °C, on solid or liquid Yeast Peptone Dextrose (YPD) media supplemented with adenine (80 mg/L). The *S. cerevisiae* AH109 competent cells were obtained as previously described⁵³. Yeast was co-transformed using a heat shock method at 42 °C for 30 min⁵⁴. For plating, 3 μL of culture were plated at the same concentration on drop-out media (minimal medium) in the absence of leucine and tryptophan (SD-L-W) or more restrictive media without histidine as well (SD-L-W-H) in

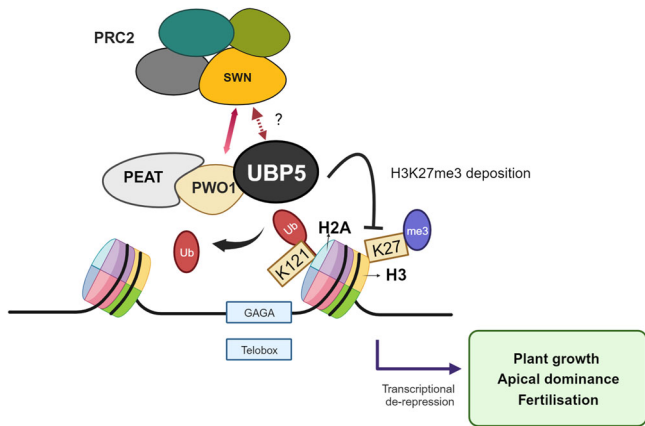


Fig. 6 | Working model for UBP5 function. UBP5 interacts with PWO1 (a PEAT complex subunit) and possibly with PRC2 through a direct interaction with SWN. UBP5 recruitment to chromatin associates with PRC2-related *cis*-elements (light blue boxes). UBP5 acts as a H2A deubiquitinase and prevents the deposition of H3K27me3, leading to transcriptional de-repression and changes in plant development. Created with BioRender.com.

serial dilutions. Yeast colony growth was analysed after 3 to 4 days growing at 28 °C. Both bait and prey empty vectors were used as negative controls.

Co-immunoprecipitation assay

Modified versions of pMDC7 carrying the GFP or mCherry tags⁵⁵ were used to insert the coding sequence of *UBP5* and *SWN/ASE1* via Gateway cloning (Invitrogen). Vectors were transformed in *Agrobacterium tumefaciens* (*Agrobacterium*) GV3101 pMP90. For transient expression assays, the abaxial sides of leaves of 4/5-week-old *N. benthamiana* plants were infiltrated with transformed *Agrobacterium* cell culture suspension in log phase growth. Expression was induced by spraying 20 µM β-estradiol in 0.1% Tween onto infiltrated leaves 48 to 72 h after *Agrobacterium* infiltration. Fluorescence was monitored in leaf epidermis cells after a short induction period (4–6 h when fluorescence was visible) using an Olympus BX51 epifluorescence microscope. After 6 h from the second induction of β-estradiol, the samples were frozen in liquid N₂. The samples were ground in a liquid N₂ pre-cooled mortar followed by 20 min at 4 °C in a shaker in 10 mL of protein extraction buffer (10% glycerol, 150 mM NaCl, 2.5 mM EDTA, 20 mM Tris-HCl pH 8, 1% Triton and Complete® EDTA-free protease inhibitor cocktail (1 tablet/50 mL; Roche)). After resuspension, samples were filtered through two Miracloth (Calbiochem®) layers and centrifuge at 4 °C for 15 min at 4000 rpm. After centrifugation, the supernatants were transferred to a new 15 mL tube, and the extracts were taken, mixed with 3X Laemmli buffer (0.3 M Tris-HCl (pH 6.8); 10 % (w/v) SDS; 30% (v/v) glycerol; 0.6 M DTT; 0.01% (w/v) bromophenol blue) and heated at 95 °C for 5 min. Co-IPs were carried out by incubating the samples with 30 µL of protein A agarose bead slurry for 4 h at 4 °C in a rotating wheel and with anti-mCherry (Takara 632496, Dilution-1:1,000). After 4 h incubation, a centrifugation at 4 °C at 500 g for 2 min was carried out to precipitate the beads. The beads were washed 3 times with protein extraction buffer, resuspended in 3X Laemmli buffer and denatured at 95 °C for 10 min. Proteins were loaded in 10% SDS-PAGE gels and transferred to a PVDF membrane. Membranes were developed with anti-GFP (Roche 11814460001, Dilution-1:1000).

Subnuclear localisation and FRET assay

For subnuclear localization in *N. benthamiana*, estradiol-inducible pMDC7-derivatives plasmid vectors containing our coding sequences were transformed into *Agrobacterium* (GV3101 PMP90 strain with

p19 silencing suppressor plasmid). FRET-APB assays were performed in *N. benthamiana* *Agrobacterium*-infiltrated with the corresponding estradiol-inducible pMDC7 vectors carrying the appropriate cDNAs. After infiltration (following the same protocol as described for co-IPs) and after 16–20 h after β-estradiol induction, leaf pieces were observed under a LSM780 (Zeiss) by excitation at 488 nm (argon laser) for eGFP and 551 nm (helium laser) for mCherry tagged proteins. FRET efficiency was calculated as the intensity increase in GFP fluorescence signal after photobleaching of the acceptor mCherry only using measurements with less than 10% GFP intensity fluctuations before acceptor bleaching⁵⁵.

Histone extraction and western blot

Nuclei were extracted from 1.5 g of 12 days after germination (DAG) seedlings using nuclei extraction buffer (0.4 M Sucrose, 10 mM Tris-HCl pH 8.0, 5 mM β-Mercaptoethanol, 10 mM MgCl₂, 0.1 mM PMSF). Extracted nuclei were treated overnight with 0.4 N H₂SO₄ to obtain a histone-enriched extract. The extracted proteins were precipitated with 33% trichloroacetic acid and then washed 3 times with acetone, air-dried, and re-suspended in 100 µL 3X Laemmli buffer. The samples were boiled for 10 min, separated on 15% sodium dodecyl sulfate-polyacrylamide electrophoresis gels and transferred to a polyvinylidene difluoride membrane (Immobilon-P Transfer membrane, Millipore) by wet blotting in transfer buffer (25 mM Tris-HCl, 192 mM glycine, and 10% methanol). Primary and secondary antibodies used were anti-H2Aub antibody (Cell Signalling Technology D27C4, dilution- 1:2000), anti-H2A antibody (Active Motif 91325, dilution- 1:1000), anti-H3K27me3 antibody (Millipore 07-449, dilution- 1:5000), anti-H3 (Abcam ab1791, dilution- 1:5000), anti-mouse IgG (H+L) HRP conjugated (Chemicon International AP308P, dilution- 1:3000) and Anti-Rabbit IgG (whole molecule)-Peroxidase (Sigma Aldrich A9169, dilution- 1:63000). Chemiluminescence detection was done with Super-Signal West Pico or Femto (Thermo Fischer Scientific) following the manufacturer's instructions.

ChIP-qPCR, ChIP-seq and data analyses

Chromatin immunoprecipitations (ChIP) were carried out using 12-DAG seedlings. Chromatin was extracted from formaldehyde fixed tissue and fragmented using a Bioruptor® Pico (Diagenode) in fragments of 200–500 bp. Antibodies used for ChIP-qPCR in this study were H3K27me3 (Millipore 07-449) and H2Aub (Cell Signalling Technology D27C4). 30 µL/sample of Protein A Dynabeads (10002D) were used for preclearing before IP. The IP was performed with 60 µL/sample of Protein A Dynabeads and 5 µL of antibodies in the ChIP dilution buffer at 4 °C overnight. Following IP, chromatin was washed with four different wash buffers- Low Salt, High salt, LiCl and TE wash buffer sequentially. Then, the chromatin was eluted, and crosslinking was reversed overnight at 65 °C. After IP, DNA was eluted and purified using ultrapure phenol:chloroform:isoamyl alcohol (25:24:1) pH 8.05 followed by ethanol precipitation. Input DNA was diluted to 1:10, and 1 µL of IP DNA was used for quantitative PCR (qPCR). ChIP-qPCRs were carried out in a CFX96TM Real-Time PCR Detection System (Bio Rad) using TakyonTM No Rox SYBR MasterMix dTTP Blue (Eurogentec).

For ChIP-seq experiments, chromatin extraction and immunoprecipitation of histones were done in three biological replicates for H2Aub and two biological replicates for H3K27me3 at 12-DAG-old Col-0 and *ubp5* seedlings. Two IPs were carried out for each biological replicate using 100 µg of chromatin, quantified using Pierce BiCinchoninic Acid (BCA) assay kit (Thermo Fisher Scientific). After IP, DNA was eluted and purified. Library preparation and paired end sequencing was performed using DNA Nanoballs (DNB™) sequencing technology from BGI (Sequencing method: DNBSEQ-G400_PE100). Reads were mapped using STAR v2.7.8a⁵⁶ onto TAIR10 Arabidopsis with parameters align intron max as 1 and align ends type as EndToEnd. The organelle genomes were excluded from the mapped reads. Duplicated

reads were removed using Picard tool MarkDuplicates option. Only uniquely mapped reads were retained for further analysis. Marked peaks for each IP were obtained using MACS3⁵⁷ with parameters broad peak and q value cut off as 0.05. Browser tracks were obtained using the bamCoverage function by scaling with the parameter --normalizeUsing RPGC. Tracks were visualised using IGV v2.12.3⁵⁸. Bedtools Utility Intersect⁵⁹ was used to intersect the MACS3 peaks obtained from the biological replicates. The resulting peaks from the biological replicates were merged and annotated with TAIR10 gene coordinates. To determine gain or depletion of H2Aub or H3K27me3 marks, the number of reads mapping into the peak coordinates was calculated using Bedtools Utility Multicov and the peaks from all samples were grouped by gene-ID to obtain unique peak coordinates per marked gene using Bedtools Utility Groupby v2.26.0⁵⁹. Differential enrichment of respective marks between samples were done using DESeq2 analysis⁶⁰. The comparison between biological replicates of H2Aub and H3K27me3 are shown in Supplementary Fig. 13a, b.

UBP5-GFP ChIP-seq and data analyses

UBP5-GFP ChIP was performed with the *UBP5pro::gUBP5-GFP;ubp5* line using a double crosslinking protocol⁶¹. Two biological replicates with 2 g each from 12-DAG seedlings were ground in liquid N₂ to fine powder and resuspended in nuclei isolation buffer (60 mM HEPES pH 8.0, 1 M Sucrose, 5 mM KCl, 5 mM MgCl₂, 5 mM EDTA, 0.6% Triton X-100, 0.4 mM PMSF, pepstatin and complete protease inhibitors (Roche)). Then, the samples were cross-linked with 25 mM ethylene glycol bis succinimidyl succinate (EGS) by rotating for 20 min and with 1% formaldehyde by rotating for 10 min. The crosslinking of samples was stopped by 2 M glycine for 10 min at room temperature. The chromatin was isolated and sheared into 200–500 bp fragments by sonication. For IP, the sonicated chromatin was incubated with 20 µl of anti-GFP antibody (Thermo Fisher #A11122) overnight at 4°C while gentle rotating. Followed by IP, eluted and purified DNA of two independent biological replicates along with input control without antibody was used for library preparation and paired end sequencing was performed using DNBTM sequencing technology from BGI.

For UBP5-GFP ChIP-seq data analysis, raw data with adapter sequences or low-quality sequences was filtered using SOAPnuke software (BGI). The reads were mapped to the Arabidopsis genome (TAIR10) using Bowtie2 2.4.5⁶² with default parameters. Only uniquely mapped reads were retained for further analysis. Peaks were called using MACS3⁵⁷. The peaks were converted to bigwig files using deepTools⁶³. bamCoverage was done using RPGC normalisation. The intersections of common peaks between two biological replicates with FDR < 0.01 was obtained using Bedtools Utility Intersect v2.30.0⁵⁹. Comparison between ChIP-seq replicates were shown in Supplementary Fig. 13a, b.

For DNA motifs analyses, we considered -500 bp to +250 bp from TSS for the UBP5 target genes using 'getfasta' function. We searched for enriched DNA motifs using the fasta file as a input for MEME-ChIP³³ with discriminative mode using the negative control sequences wherein UBP5 targeting regions were removed.

H2A deubiquitination assay

N. benthamiana plants were infiltrated with Agrobacterium strain GV3101 carrying appropriate binary vectors. Before infiltration, bacterial cells were pelleted and resuspended in MES buffer (10 mM MgCl₂, 10 mM MES, 200 µM acetosyringone, pH 5.7) in the dark for 2 h at room temperature (RT). The suspended Agrobacterium cells were mixed with P19 silencing suppressor at an appropriate ratio to a final OD₆₀₀ of 0.6, followed by infiltration in 6-week-old *N. benthamiana* leaves. After 24 h of infiltration, an induction was performed by spraying the leaves on their abaxial side with 50 µM of β-estradiol solution in 0.1% (v/v) Tween-20. Post induction, the transient expression of agroinfiltrated cells was examined under an Olympus BX51

epifluorescence microscope. Histone extraction was performed from the infiltrated leaves and immunoblotted using anti-H2Aub antibody (Cell Signalling Technology D27C4, dilution- 1:2000), and anti-H2A (Active Motif 91325, dilution- 1:1000) and anti-H3 (Abcam ab1791, dilution- 1:5000) as controls.

RNA isolation, quantitative RT PCR

Total RNA was isolated from 12-DAG seedlings (Col and *ubp5*) using E.Z.N.A. Plant RNA Kit (OMEGA biotek) following manufacturer instructions. The RNA concentration was determined using the Nanophotometer (IMPLEN). RNA was examined by electrophoresis on a 1.2% agarose gel. For cDNA synthesis, RNA samples were subjected to DNase treatment and cDNA synthesis was performed using (Thermo Scientific). Quantitative real time PCR (qRT-PCR) was performed in a CFX96TM Real-Time PCR Detection System (Bio Rad) using TakyonTM No Rox SYBR MasterMix dTTP Blue (Eurogentec). Expression levels were normalised to the reference genes At5G25760 and At4G34270⁶⁴. Relative enrichment was calculated using the 2^{-ΔΔCT} method⁶⁵.

RNA-seq library preparation, sequencing, and bioinformatics

For RNA-seq, RNA was extracted from 12-DAG seedlings with four biological replicates for each background (Col-0 and *ubp5*). Library preparation and RNA-seq was performed according to the protocol described recently⁶⁶. 500 ng DNase-treated RNA was used for reverse transcription with 50 mM different barcoded oligo(dT) primers and SuperScript III (Thermo Fisher Scientific). Each reaction was pooled, pools were Ampure purified (1.5x beads to sample volumes) and then eluted. Second-strand synthesis was carried out using nick translation protocol (Krzyszton et al. 2022). Tagmentation reaction⁶⁷ was performed using recovered dsDNA sample incubated with homemade Tn5 enzyme in a freshly prepared 2x buffer (20 mM Tris-HCl pH 7.5, 20 mM MgCl₂, 50% DMF). Illumina indexing PCR was performed using the tagged DNA. Libraries were sequenced on Illumina NextSeq 500 system using the paired-end mode to obtain 21 nt R1 (contain barcode and Unique Molecular Identifier (UMI)) and 55 nt R2 (contain mRNA sequences).

After quality control using fastqc, reads R1 and R2 were processed separately. In our oligo(dT) primers two parts of UMI are split by barcode sequence, therefore we transformed read R1 fastq file using awk command. Read R2 was trimmed to remove potential contamination with poly(A) tail using BRBseqTools v 1.6 Trim⁶⁸. Reads were mapped using STAR v 2.7.8a⁵⁶ to TAIR 10 genome with Araport11 genome annotation. Finally, the count matrix for each library and each gene was obtained using BRBseqTools (v 1.6) CreatedGEMatrix⁶⁸ with parameters *-p UB -UMI 14 -s yes*, using Araport11 genome annotation and a list of barcodes. The differential gene expression analysis was done using the DESeq2⁶⁹. Further, the genes were filtered based on log₂ fold-change of ±1 and an adjusted *p*-value of less than 0.05 and categorised as upregulated, downregulated, and non-misregulated genes. GO enrichment analyses were performed in different gene set using ShinyGO tool v0.75⁷⁰.

Statistics and reproducibility

Statistical tests performed on experimental data and sample sizes are noted in figure legends. No statistical method was used to pre-determine sample size. All data points are derived from biological replicates. Exact *p*-values for each pairwise comparisons are mentioned in the figures. Plants were placed randomly in the plant growth facility. No data were excluded from the analyses. No blinding was applied for sampling.

Reporting summary

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

Data availability

All data supporting the findings of this study are available within the manuscript and its supplementary files. All high-throughput sequencing data generated in this study have been deposited in GEO with accessions codes GSE217614 and GSE209707 and data lists are available in the Source Data file. Seeds of the *ubp5* CRISPR/Cas9 line are available under request. Source data are provided with this paper.

Code availability

The custom code used for the analysis has been deposited at [https://github.com/mohang13/ubp5_nat_comms]. Any additional information required to reanalyse the data reported in this paper is available from the corresponding author upon request.

References

- Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* **32**, 42–56 (2016).
- Morgan, M. A. J. & Shilatifard, A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat. Genet.* **52**, 1271–1281 (2020).
- Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **43**, 31–37 (2017).
- Barbour, H., Daou, S., Hendzel, M. & Affar, E. B. Polycomb group-mediated histone H2A monoubiquitination in epigenome regulation and nuclear processes. *Nat. Commun.* **11**, 5947 (2020).
- Mozgova, I. & Hennig, L. The polycomb group protein regulatory network. *Annu Rev. Plant Biol.* **66**, 269–296 (2015).
- Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708 (2009).
- Godwin, J. & Farrona, S. The Importance of Networking: Plant Polycomb Repressive Complex 2 and Its Interactors. *Epigenomes* **6**, 8 (2022).
- Shen, Q., Lin, Y., Li, Y. & Wang, G. Dynamics of H3K27me3 Modification on Plant Adaptation to Environmental Cues. *Plants* **10**, 1165 (2021).
- Merini, W. & Calonje, M. PRC1 is taking the lead in PcG repression. *Plant J.* **83**, 110–120 (2015).
- Chittock, E. C., Latwiel, S., Miller, T. C. & Muller, C. W. Molecular architecture of polycomb repressive complexes. *Biochem Soc. Trans.* **45**, 193–205 (2017).
- March, E. & Farrona, S. Plant Deubiquitinases and Their Role in the Control of Gene Expression Through Modification of Histones. *Front. Plant Sci.* **8**, 2274 (2017).
- Derkacheva, M. et al. H2A deubiquitinases UBP12/13 are part of the Arabidopsis polycomb group protein system. *Nat. Plants* **2**, 16126 (2016).
- Kralemann, L. E. M. et al. Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis. *Genome Biol.* **21**, 144 (2020).
- Hohenstatt, M. L. et al. PWWP-DOMAIN INTERACTOR OF POLYCOMBS1 Interacts with Polycomb-Group Proteins and Histones and Regulates Arabidopsis Flowering and Development. *Plant Cell* **30**, 117–133 (2018).
- Tan, L. M. et al. The PEAT protein complexes are required for histone deacetylation and heterochromatin silencing. *EMBO J.* **37**, e98770 (2018).
- Deng, W. et al. Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC genom.* **14**, 593 (2013).
- Zhou, Y. et al. Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nat. Genet.* **50**, 638–644 (2018).
- Mikulski, P. et al. The Chromatin-Associated Protein PWO1 Interacts with Plant Nuclear Lamin-like Components to Regulate Nuclear Size. *Plant Cell* **31**, 1141–1154 (2019).
- Chanvittana, Y. et al. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development* **131**, 5263–5276 (2004).
- Yoshida, N. et al. EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. *Plant Cell* **13**, 2471–2481 (2001).
- Eloy, N. B. et al. SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization. *Proc. Natl Acad. Sci. USA* **109**, 13853–13858 (2012).
- Mainguet, S. E. et al. Uracil salvage is necessary for early Arabidopsis development. *Plant J.* **60**, 280–291 (2009).
- Wang, S. Q., Shi, D. Q., Long, Y. P., Liu, J. & Yang, W. C. GAMETOPHYTE DEFECTIVE 1, a putative subunit of RNases P/MRP, is essential for female gametogenesis and male competence in Arabidopsis. *PLoS One* **7**, e33595 (2012).
- Takahashi, H. et al. MAG4/Atp115 is a golgi-localized tethering factor that mediates efficient anterograde transport in Arabidopsis. *Plant Cell Physiol.* **51**, 1777–1787 (2010).
- An, Y. Q., Huang, S., McDowell, J. M., McKinney, E. C. & Meagher, R. B. Conserved expression of the Arabidopsis ACT1 and ACT 3 actin subclass in organ primordia and mature pollen. *Plant Cell* **8**, 15–30 (1996).
- Ramirez-Prado, J. S. et al. The Polycomb protein LHP1 regulates Arabidopsis thaliana stress responses through the repression of the MYC2-dependent branch of immunity. *Plant J.* **100**, 1118–1131 (2019).
- Kleinmanns, J. A., Schatlowski, N., Heckmann, D. & Schubert, D. BLISTER Regulates Polycomb-Target Genes, Represses Stress-Regulated Genes and Promotes Stress Responses in Arabidopsis thaliana. *Front. Plant Sci.* **8**, 1530 (2017).
- Shu, J. et al. Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings. *Plant Direct* **3**, e00100 (2019).
- Rao-Naik, C., Chandler, J. S., McArdle, B. & Callis, J. Ubiquitin-specific proteases from Arabidopsis thaliana: cloning of AtUBP5 and analysis of substrate specificity of AtUBP3, AtUBP4, and AtUBP5 using Escherichia coli in vivo and in vitro assays. *Arch. Biochem. Biophys.* **379**, 198–208 (2000).
- Sridhar, V. V. et al. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* **447**, 735–738 (2007).
- Nassrallah, A. et al. DET1-mediated degradation of a SAGA-like deubiquitination module controls H2Bub homeostasis. *Elife* **7**, e37892 (2018).
- Berardini, T. Z. et al. The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis* **53**, 474–485 (2015).
- Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
- Xiao, J. et al. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* **49**, 1546–1552 (2017).
- Zhou, Y., Hartwig, B., James, G. V., Schneeberger, K. & Turck, F. Complementary Activities of TELOMERE REPEAT BINDING Proteins and Polycomb Group Complexes in Transcriptional Regulation of Target Genes. *Plant Cell* **28**, 87–101 (2016).
- Zhou, Y. et al. H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. *Genome Biol.* **18**, 1–13 (2017).
- Zheng, S. Y. et al. Dual roles of the Arabidopsis PEAT complex in histone H2A deubiquitination and H4K5 acetylation. *Mol. Plant* **16**, 1847–1865 (2023).

38. Foglizzo, M. et al. A bidentate Polycomb Repressive-Deubiquitinase complex is required for efficient activity on nucleosomes. *Nat. Commun.* **9**, 3932 (2018).
39. Scheuermann, J. C. et al. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243–247 (2010).
40. Sahtoe, D. D., van Dijk, W. J., Ekkebus, R., Ovaas, H. & Sixma, T. K. BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nat. Commun.* **7**, 10292 (2016).
41. Yu, M. et al. USP11 Is a Negative Regulator to gammaH2AX Ubiquitylation by RNF8/RNF168. *J. Biol. Chem.* **291**, 959–967 (2016).
42. Vissers, J. H., Nicassio, F., van Lohuizen, M., Di Fiore, P. P. & Citterio, E. The many faces of ubiquitinated histone H2A: insights from the DUBs. *Cell Div.* **3**, 8 (2008).
43. Lu, F., Cui, X., Zhang, S., Jenuwein, T. & Cao, X. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat. Genet.* **43**, 715–719 (2011).
44. Yin, X. et al. H2AK121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots. *Nat. Commun.* **12**, 315 (2021).
45. Mozgova, I. & Hennig, L. The polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* **66**, 269–296 (2015).
46. Calonje, M. PRC1 Marks the Difference in Plant PcG Repression. *Mol. Plant* **7**, 459–471 (2014).
47. Liu, S. et al. H2A ubiquitination is essential for Polycomb Repressive Complex 1-mediated gene regulation in *Marchantia polymorpha*. *Genome Biol.* **22**, 253 (2021).
48. Dupouy, G. et al. Plastid ribosome protein L5 is essential for post-globular embryo development in *Arabidopsis thaliana*. *Plant Reprod.* **35**, 189–204 (2022).
49. Lei, Y. et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* **7**, 1494–1496 (2014).
50. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
51. Edwards, K., Johnstone, C. & Thompson, C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* **19**, 1349 (1991).
52. Zhong, S., Lin, Z., Fray, R. G. & Grierson, D. Improved plant transformation vectors for fluorescent protein tagging. *Transgenic Res* **17**, 985–989 (2008).
53. Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 35–37 (2007).
54. Kawai, S., Hashimoto, W. & Murata, K. Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioeng. Bugs* **1**, 395–403 (2010).
55. Bleckmann, A., Weidtkamp-Peters, S., Seidel, C. A. & Simon, R. Stem cell signaling in *Arabidopsis* requires CRN to localize CLV2 to the plasma membrane. *Plant Physiol.* **152**, 166–176 (2010).
56. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
58. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
59. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
60. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
61. Bourbousse, C., Vegesna, N. & Law, J. A. SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **115**, E12453–E12462 (2018).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
63. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–191 (2014).
64. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
65. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ Method. *Methods* **25**, 402–408 (2001).
66. Krzyszton, M. et al. Single seeds exhibit transcriptional heterogeneity during secondary dormancy induction. *Plant Physiol.* **190**, 211–225 (2022).
67. Hennig, B. P. et al. Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. *G3 (Bethesda)* **8**, 79–89 (2018).
68. Alpern, D. et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.* **20**, 1–15 (2019).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).

Acknowledgements

S.F. acknowledges support from the College of Science and Engineering and the Research Office (University of Galway). S.F. is grateful to Jennifer Siobal for technical support. S.F. thanks Miguel de Lucas Torres (Durham University) for his support with Y2H experiments. S.F., R.H., K.N. and M.G. were supported by 20/FFP-P/8693 grant from Science Foundation Ireland (SFI) and by a NUI Galway Research Grant for Returning Academic Careers QA151. M.G. was also supported by the SFI Centre for Research Training in Genomics Data Science (grant N. 18/CRT/6214). J.G. was supported through the University of Galway Hardiman Scholarship programme and Thomas Crawford Research Grant. J.G. internship at IBENS was supported by the COST Action CA16212 INDEPTH (EU). E.M. was funded by a College of Science and Engineering scholarship (University of Galway). Work in FB and CB laboratory was supported by ANR-18-CE13-0004-01 and ANR-20-CE13-0028 grants from the French National Research Agency. S.S. was supported by Foundation for Polish Science (TEAM POIR.04.04.00-00-3C97/16) and by Polish National Science Centre (SONATA BIS UMO-2018/30/E/NZ1/00354). M.K. was supported by Polish National Science Centre (OPUS UMO-2021/41/B/NZ3/02605).

Author contributions

J.G. and S.F. conceptualised the experiment approach and designed the methodology; J.G., K.N., E.M., L.W., R.H. and J.L. performed the experiments; J.G., M.G., A.F., M.K., F.B. and C.B. performed the genomic data curation, analysis, and visualisation; J.G. and S.F. wrote the original manuscript; J.G., F.B., C.B., S.S., D.S. and S.F. contributed to the interpretation of results; all the authors contributed to manuscript revision and approved the final 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/s41467-023-44546-8>.

Correspondence and requests for materials should be addressed to Sara Farrona.

Peer review information *Nature Communications* thanks Yijing Zhang and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024