Convergent evolution between PALI1 and JARID2 for the allosteric activation of PRC2

Qi Zhang[1,3], Samuel C. Agius[1,3], Sarena F. Flanigan[1], Vitalina Levina[1], Brady M. Owen[1], Chen Davidovich[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia; [2]EMBL-Australia and the ARC Centre of Excellence in Advanced Molecular Imaging, Clayton, Victoria, Australia; [3]These authors contributed equally; *Correspondence to chen.davidovich@monash.edu

ABSTRACT

The polycomb repressive complex 2 (PRC2) is a histone methyltransferase that maintains cell identities. JARID2 is the only accessory subunit of PRC2 that known to trigger an allosteric activation of methyltransferase. Yet, this mechanism cannot be generalised to all PRC2 variants as, in vertebrates, JARID2 is mutually exclusive with most of the accessory subunits of PRC2. Here we provide functional and structural evidence that the vertebrate-specific PRC2 accessory subunit PALI1 emerged through a convergent evolution to mimic JARID2 at the molecular level. Mechanistically, PRC2 methylates PALI1 K1241, which then binds to the PRC2-regulatory subunit EED to allosterically activate PRC2. PALI1 K1241 is methylated in mouse and human cell lines and is essential for PALI1-induced allosteric activation of PRC2. High-resolution crystal structures revealed that PALI1 mimics the regulatory interactions formed between JARID2 and EED. Independently, PALI1 also facilitates DNA and nucleosome binding by PRC2. In acute myelogenous leukemia cells, overexpression of PALI1 leads to cell differentiation, with the phenotype abrogated by a separation-of-function PALI1 mutation, defective in allosteric activation and active in DNA binding. Collectively, we show that PALI1 facilitates catalysis and substrate binding by PRC2 and provide evidence that subunit-induced allosteric activation is a general property of holo-PRC2 complexes.

INTRODUCTION

Over the course of evolution, gene families tend to expand[1]. Accordingly, the number of genes linked to the same function is commonly increased in vertebrates with respect to invertebrates, especially in cases of genes coding for transcriptional regulators[2,3]. For instance, the histone H3K4 methyltransferase MLL/COMPASS complex expended from one in yeast to three sub-types in fly and six complexes in human, where additional subunits emerged through each expansion[4]. Similar expansion took place for the histone ubiquitin ligase polycomb repressive complex 1 (PRC1): from

two complexes in the fly to at least six variants in vertebrates[5,6]. Vertebrate-specific subunits of histone modifiers provide the opportunity to identify molecular mechanisms that are fundamental to chromatin biology and, therefore, re-emerged through the course of evolution.

The polycomb repressive complex 2 (PRC2) is a histone methyltransferase complex that is required for the maintenance of cell identity in all multicellular organisms. At the molecular level, PRC2 maintains the repressed state of developmentally expressed genes through the tri-methylation of lysine 27 in histone H3 (H3K27me3), a hallmark of facultative heterochromatin[6,7]. The core PRC2 complex includes four subunits[7-10], but it has a low histone methyltransferase activity and low affinity to DNA. Therefore, holo-PRC2 complexes include additional protein subunits—termed accessory subunits. Most of the accessory subunits of the vertebrate PRC2 emerged though gene duplication and some are vertebrate specific, with the latter poorly understood mechanistically.

The accessory subunits are collectively required for the recruitment of PRC2 to chromatin and for the regulation of its enzymatic activity[11,12]. Unbiased proteomic studies[13-17] revealed two distinct holo-PRC2 complexes—PRC2.1 and PRC2.2—defined by their mutually exclusive accessory subunits[16]. The PRC2.2 complex is nearly identical in fly and human, and includes the accessory subunits AEBP2 and JARID2. Contrarily to PRC2.2, the PRC2.1 complex went through a massive expansion over evolution: from one accessory subunit in fly, to at least five in vertebrates[13-17]. The fly PRC2.1 accompanies a single accessory subunit: Pcl. The vertebrate PRC2.1 is far more complex: it contains one of the three polycomb-like (PCL) proteins[16] (PHF1, MTF2 or PHF19) together with either EPOP[13,16] or PALI1[16,18] (also annotated as LCOR-CRA_b, LCOR isoform 3 or C10ORF12). Recent works indicate some non-redundant functions of the PRC2.1 and PRC2.2 complexes in mouse embryonic cells[11,12], but the molecular basis is unknown.

The PRC2.2-specific subunit JARID2 has two activities that were implicated in nucleating H3K27me3: chromatin binding[19,20] and allosteric stimulation of histone methyltransferase (HMTase)[21]. During JARID2-induced allosteric activation, PRC2 first di- or tri-methylates lysine 116 in JARID2 (JARID2-K116me2/3). Next, the di/tri-methyl-lysine binds to the regulatory subunit EED and triggers an allosteric activation of PRC2[21]. This mechanism is thought as a “jump-start” to activate PRC2[8]. After the nucleation of H3K27me3, histone tails carrying the H3K27me3 mark bind to the regulatory subunit EED[22] to trigger further allosteric activation of PRC2[23]. Yet, knockout of JARID2 in mouse ESC cells lacks major effect on H3K27me3 globally and locally[11,21]. This implies a parallel role taken by PRC2.1 during de novo introduction of H3K27me3, in agreement with the PRC2.1-specific subunit MTF2 being essential for this process[11 24]. Yet, a mechanism for subunit-induced allosteric activation of the PRC2.1 complex is yet to be discovered.

Multiple unbiased proteomic studies identified C10ORF12 as an accessory subunit of PRC2.1[15,16,25] and, thus, mutually exclusive with JARID2. More recently, C10ORF12 was annotated as PALI1 and identified as a vertebrate-specific protein, coded by a transcript of the *LCOR* locus[18]. Sequence homology pointed out a paralogue of PALI1, termed PALI2, encoded by the *LCORL* locus. PALI1 is required for mouse development[18] and promotes the histone methyltransferase (HMTase) activity of PRC2 *in vitro* and *in vivo*[16,18], but the molecular mechanism is unknown.

Here, we show that PALI1 allosterically activates PRC2 and facilitates substrate binding. Mechanistically, PALI1 lysine 1241 is a substrate for PRC2 *in vitro* and is methylated in multiple human and mouse cell lines. Once in a di- or tri-methyl-form, PALI1-K1241me2/3 binds to the regulatory subunit EED and allosterically activates PRC2. Structural and functional evidence indicates that PALI1 has emerged through a convergent evolution to mimic the function of JARID2 within the context of the PRC2.1 complex. We also show that the PRC2-binding domain of PALI1 increases the affinity of PRC2 to DNA and mono-nucleosome substrates by >10-fold. Allosteric activation and chromatin binding are two separate functions of PALI1, demonstrated by a separation-of-function PALI1 mutant: defective in allosteric activation but active in substrate binding. This separation of function PALI1 mutant abrogates PALI1-indudced cell differentiation, demonstrating a molecular function of subunit-induced allosteric activation in cells. Our results reveal how PRC2 is regulated by PALI1 at the molecular level and, more broadly, implies that subunit-induced allosteric activation is permitted in most variants of holo-PRC2 complexes in vertebrates.

RESULTS

PALI1 K1241 is methylated in mouse and human cell lines

In a search for a PRC2.1 accessory subunit that could trigger an allosteric activation of PRC2, we first set out to map the PRC2 methylome in mouse and human cells. We reasoned that affinity purification of PRC2 followed by tandem mass spectrometry (AP-MS), will allow for the identification of methyl-lysines residing within assembled PRC2 complexes *in vivo*. Hence, we analysed multiple publicly available LC-MS/MS data originating from AP-MS experiments, where PRC2 subunits were used as baits[15,26-29] (Fig. 1a and Supplementary Table 1). Although these studies[15,26-29] were not focused on the methylation of PRC2 subunits, the high quality of the raw data allowed us to detect methyl-lysines in tryptic peptides (Supplementary Table 1). As expected, the PRC2 methylome contains the previously reported methylations in JARID2 K116[21] and EZH2 K514 and K515[30-32].

The two most frequently detected di- and tri-methyl lysines in the accessory subunits of PRC2 were in JARID2 K116 and PALI1 K1241 (Fig 1a), with the former triggering an allosteric activation of PRC2[21]. Specifically, PALI1 K1241 is methylated in five of the seven cell lines that were tested, including both human and mouse cell lines: HEK293T (human embryonic kidney), STS26T (human malignant peripheral nerve sheath tumour), LnCAP (human prostate cancer), U2OS (human osteosarcoma) and mouse embryonic stem cells (mESC). In four of these cell lines, PALI1 K1241 identified either in its di- or tri-methyl form (i.e. PALI1 K1241me2/3).

In some of the cell lines, we also identified methylations in PALI1 K1214 and K1219, in agreement with a previous proteomic analysis in HCT116 cells[30]. The same study also identified methylations of EZH2 K510 and K515, that have more recently been shown to regulate PRC2[31,32]. Yet, K1241 was not identified in that study[30], which used antibodies against methyl-lysines for immunoaffinity purification ahead of the mass spectrometry[30].

Hence, PALI1 and JARID2 were the only accessory subunits that were identified with di- or tri-methyl lysin modifications that are evolutionary conserved in mouse and human (Fig. 1a). Di- and tri-methylated JARID2 K116 (JARID2 K116me2/3) allosterically activate the PRC2.2 complex through direct interactions with the regulatory subunit EED[21]. With that in mind, we set out to test the hypothesis that a methyl-lysines in PALI1 allosterically activates the PRC2.1 complex.

PALI1 K1241 is methylated by PRC2 *in vitro*

We next set out to determine if PALI1 can be methylated by PRC2. We expressed and purified a recombinant human PRC2 (EZH2, EED, SUZ12 and RBBP4) in a complex with the PRC2-interacting region from PALI1 ($PALI1_{PIR}$; Fig. 1b). *In vitro* HMTase assay using mononucleosome substrates (Fig. 1c) confirmed that the PRC2-$PALI1_{PIR}$ complex, comprising amino acids 1058-1250 from PALI1 (purple bar in Fig. 1a), is substantially more active than the core PRC2 complex. While this result is in agreement with previous reports[16,18,33], we also noted an additional band on the radiogram (Fig. 1c, marked with an asterisk). That band indicated a methylated protein that appeared only if PRC2 contained $PALI1_{PIR}$, and migrated with the apparent molecular weight of $PALI1_{PIR}$. In order to confirm that the methylated protein is indeed $PALI1_{PIR}$, we purified a PRC2-$PALI1_{PIR}$ complex with a 3C-cleavable MBP tag carried only by $PALI1_{PIR}$. We then performed the HMTase assay in the presence and absence of human rhinovirus 3C protease. The 3C-specific cleavage of the MBP-tag on $PALI1_{PIR}$ led to a large shift in the migration velocity of the methylated protein, confirming it is indeed $PALI1_{PIR}$ (Fig. 1c, lane 2 versus lane 3).

A similar result was obtained when we performed the same experiment using a longer truncation of PALI1 that was designed based on the previous mapping of the PRC2-interacting region from PALI1[18] (PALI1 1058-1329; termed PALI1$_{PIR\text{-}long}$ herein; long purple bar in Fig. 1a). PALI1$_{PIR\text{-}long}$ co-purified with PRC2 as a soluble complex (Supplementary Fig. 1a, b), albeit in multiple truncated forms. In-gel digestion with mass spectrometry subsequently identified PALI1$_{PIR}$ as a more stable truncation of PALI1$_{PIR\text{-}long}$, while both constructs co-purified with PRC2 and enhanced HMTase (Fig. 1c and Supplementary Fig. 1c). These experiments confirmed that PALI1$_{PIR}$ is sufficient to enhance the HMTase activity of PRC2 towered mononucleosome substrates and that PRC2 methylates PALI1$_{PIR}$ *in vitro*.

In order to identify the methylated amino acids within PALI1, we performed liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis of the recombinant PRC2-PALI1$_{PIR\text{-}long}$. As expected, we detected the previously reported methyl-lysines in EZH2 K514 and K515[31,32] (Supplementary Table 1). In PALI1$_{PIR\text{-}long}$, we identified mono- and di-methylations in K1214, K1219 and K1241 (Fig. 1a, Supplementary Table 1 and Supplementary notes for the MS/MS spectra), in agreement with our proteomic analysis of *in vivo* AP-MS data (Fig 1a). Additionally, K1214 and K1219 were detected also in tri-methyl forms. These methylations were identified either if the complex was pre-incubated with SAM or if not, indicating that a significant fraction of the complex was purified with these modifications (Supplementary Table 1). A similar observation was previously made for EZH2 automethylation, that occurs in the recombinant protein while co-expressed with other PRC2 subunits[31]. These results confirm that three lysines within the PRC2-interacting region of PALI1 serve as a substrate for PRC2, including K1214, K1219 and K1241.

PALI1 K1241 is required in order to enhance the HMTase activity of PRC2

If methyl-lysines in PALI1 allosterically activate PRC2, the corresponding lysine residues are expected to be required for PALI1-mediated enhancement of HMTase. We, therefore, aimed to determine if the candidate lysine residues that we identified using MS/MS (Fig. 1a) are required or dispensable for PALI1-mediated enhancement of HMTase. We expressed and purified PRC2-PALI1$_{PIR\text{-}long}$ mutant complexes, included all possible perturbations of the PALI1 mutations K1214A, K1219A or K1241A. Mutant complexes migrated on a gel filtration column similar to the wild type, excluding adverse effects on complex solubility (Supplementary Fig. 2a). To assess the ability of these lysine-to-alanine mutants to enhance HMTase, we carried out an *in vitro* HMTase assay using mononucleosomes as a substrate. The mutation K1241A in PRC2-PALI$_{PIR\text{-}long}$ lead to approximately 50% reduction in HMTase activity, compared to the wild type PRC2-PALI$_{PIR\text{-}long}$ complex (Fig. 2a). The two other lysine-to-alanine mutants, K1214A and K1219A, did not affect the HMTase activity of the PRC2-PALI$_{PIR\text{-}long}$

complex (Fig. 2a). The K1241A mutant significantly reduced PALI1-mediated enhancement of HMTase in all possible perturbations that we tested, with the other two lysines, K1214 and K1219, were dispensable for methyltransferase enhancement.

Of note, while the mutation K1241A in PALI1 significantly reduced the HMTase activity of PRC2, the PRC2-PALI$_{PIR\text{-}long}$ K1241A mutant complex was still about 10-fold more active than the core PRC2 complex (Fig. 2a). Collectively, these data indicate that K1241 is required for complete PALI1-mediated HMTase enhancement, and implies the presence of an additional mechanism, independent of K1241 methylation (more below).

PALI1-K1241me2/3 is sufficient in order to stimulate the HMTase activity of PRC2

If the methylation of PALI1 K1241 is sufficient to trigger an allosteric activation of PRC2, we expected to mimic these regulatory interactions by using a short peptide, including a tri-methyl-lysine K1241 flanked by amino acids of the corresponding sequence from PALI1 (termed PALI1-K1241me3 peptide herein). Indeed, the PALI1-K1241me3 peptide significantly stimulated the HMTase activity of PRC2 towered mononucleosome substrates (Fig. 2b). Similar observations were made in the past for H3K27me3 and JARID2-K116me2/3 peptides, which allosterically activate PRC2[21,23]. We also assayed a PALI1-K1219me3 peptide, which has a smaller positive effect on the HMTase activity of PRC2. The PALI1-K1214me3 peptide was ineffective in stimulating PRC2 (Fig. 2b). Another peptide, including tri-methylations on both K1219 and K1214, exhibited only a moderate stimulation of HMTase (Fig. 2b). As expected, no stimulation of HMTase observed by unmethylated wild type or lysine-to-arginine mutant K1241 peptides that were used as negative controls (Fig. 2c).

The level of PALI1 K1241 methylation ranges from mono- to tri-methyl-lysine in different human and mouse cell lines (Fig. 1a and Supplementary Table 1). To determine how the methylation level of PALI1 K1241 affects its ability to stimulate PRC2, we assayed the HMTase activity of the core PRC2 complex in the presence of unmethylated, mono-, di- and tri-methyl-lysine PALI1 K1241 peptides (PALI1 K1241me0-3, respectively; Fig. 2d). While the unmethylated peptide (PALI1-K1241me0) did not stimulate PRC2, any additional methyl up to the di-methyl form (PALI1-K1241me2) increased the HMTase activity of PRC2 (Fig. 2d). While the tri-methyl peptide (PALI1-K1241me3) was still efficient in HMTase stimulation, it did not increase the HMTase activity of PRC2 further beyond the di-methyl peptide (Fig. 2d). Taken together, these results indicate that PALI1-K1241me2/3 is sufficient to stimulate the HMTase activity of PRC2.

PALI1-K1241me2/3 binds to the aromatic cage of the regulatory subunit EED to stimulate PRC2

H3K27me3 and JARID2-K116me2/3 bind to the aromatic cage of the regulatory subunit EED to allosterically stimulate PRC2[21,23]. We, therefore, wished to determine if there is a direct link between PALI1-K1241me2/3 and EED. We first measured the affinity of EED (amino acids 40-441) for 5-carboxyfluorescein- (5-FAM, single isomer) labelled JARID2-K116me3 peptide using fluorescence anisotropy direct titrations. The JARID2-K116me3 peptide binds to EED with a dissociation constant (K_d) of 8.07 ± 0.49 μM (Fig. 3a and Supplementary Fig. 3a), consistent with previously published results[21]. Then, we quantified the K_d of unlabelled peptides, using fluorescence anisotropy displacement titrations (Fig. 3a). As a positive control, we first quantified the affinity of an H3K27me3 peptide for EED, resulting with K_d = 41.3 ± 2.8 μM, in agreement with a previous study[21]. The affinity of EED for PALI1-K1241me3 (K_d = 8.47 μM ± 0.71 μM) is similar to the affinity of EED for the JARID2-K116me3 peptide (K_d = 8.07 μM ± 0.49 μM; Fig. 3a and Supplementary Fig. 3a). We observed weak interaction between mono-methylated K1241 peptide to EED (K_d = 240 ± 38 μM), while the di-methyl form increased the affinity for EED by approximately 10-fold (K_d =18.8 ± 2.3 μM), almost to the level of the K1241 tri-methyl-lysine peptide (K_d = 8.47 μM ± 0.71 μM) (Fig. 3a). Qualitatively, these results are in agreement with the HMTase assays done in the presence of these peptides (Fig. 2d) and support a model were PALI1 K1241me2/3 binds to EED to stimulate the HMTase activity of PRC2.

The PALI1-K1219me3 peptide binds to EED with high affinity (*Kd* = 7.53 μM ± 0.71 μM; Fig. 3a), in agreement with its ability to stimulate PRC2 (Fig. 2b). Contrarily, the PALI1-K1214me3 peptides, which did not stimulate methyltransferase (Fig. 2b), binds to EED with a 4-fold lower affinity comparing PALI1-K1219me3 and PALI1-K1241me3 (Fig. 3a).

In order to directly link between PALI1 K1241me3 and the aromatic cage of EED within the context of PRC2, we reconstituted mutant PRC2 complex harbouring the defective cage mutation EED F97A[23]. The PALI1-K1241me3 peptide did not lead to the activation of the cage mutant PRC2 (Fig. 3b), in agreement with an EED-dependent allosteric activation of PRC2. On the same line of evidence, PALI1-K1241me3-induced activation of PRC2 was inhibited by an allosteric inhibitor of PRC2, A-395, but not the negative control A-395N[34] (Supplementary Fig. 3b). Collectively, our data support a mechanism where PALI1 K1241me2/3 binds to the aromatic cage in EED to trigger an allosteric activation of PRC2.

PALI1 and JARID2, but not H3, utilise the same interactions with the regulatory subunit EED

Given the functional identity between PALI1-K1241me3 and -K1219me3 to JARID2-K116me3, we wished to assess for structural resemblance. We, therefore, solved the crystal structures of EED_{76-441} co-crystallized with a PALI1-K1241me3 or PALI1-K1219me3 peptide (Fig. 3c and Table 1). We

compared the resulted structures with the crystal structures of the EED-H3K27me3 and EED-JARID2-K116me3 complexes. The structures indicate that the two tri-methyl-lysine PALI1 peptides (Fig. 3c, left two panels, and Supplementary Fig. 3c) bind to EED in a conformation resembling that seen for the JARID2-K116me3 peptide (Fig. 3c, the third panel). Specifically, in all three cases the tri-methyl-lysine and its adjacent aromatic residue, in the +1 position, adopting the same conformation when binding to EED (Fig. 3c, marked in dashed shapes). Contrarily, H3K27me3, does not have an aromatic residue at position +1, with respect to the tri-methyl-lysine, thus adopts a different binding mode to EED (Fig. 3c, right). Altogether, these structures indicate that PALI1 and JARID2 interact with EED using their tri-methyl-lysine and its adjacent aromatic residue, despite no other sequence similarity and no common ancestor (Fig. 3d and Supplementary Fig. 3d). Of note PALI1-K1241 and the adjacent F1242, but not PALI1-K1214 or PALI1-K1219, are fully conserved across different vertebrate species (Fig. 3d). This implies for a biological significance of PALI1-K1241me2/3 in regulating PRC2 across vertebrates.

PALI1 facilitates DNA binding by PRC2, with allosteric activation being dispensable for this function

The triple mutant complex, PRC2-PALI$_{PIR\text{-}long}$ K1241A, K1219A and K1241A, is defective in allosterically stimulating HMTase but was still more active than the core PRC2 complex (Fig. 2a). This mutant complex cannot harbour any of the methylations that we identified in PALI1 (Fig. 1a). Hence, we suspected that the PRC2-interacting region within PALI1 regulates HMTase in an additional mechanism. In vitro HMTase assays previously demonstrated that the HMTase activity of PRC2 is enhanced by several of its DNA-binding accessory subunits, including MTF2, PHF19 and AEBP2[35]. We therefore wished to determine if the PRC2-interacting region of PALI1 increases the affinity of PRC2 to DNA.

To directly test if the PRC2-interacting region of PALI1 can facilitate DNA binding, we first set out to quantify the affinity of PRC2-PALI1$_{PIR}$ to DNA using fluorescence anisotropy. We used a DNA probe designed to mimic 46 bases long dsDNA from a CpG island of the CDKN2B gene (termed CpG46 DNA, see methods section for the DNA sequence). The affinity of the PRC2-PALI1$_{PIR}$ to CpG46 DNA (K_d = 155 ± 26 nM) was >20-fold higher than the affinity to the PRC2 core complex to the same DNA probe (Kd > 4 µM) (Fig. 4a), indicating that PALI1 facilitates DNA binding. The DNA-binding activity of PALI1 was not specific to the CpG46 DNA probe: PRC2-PALI1$_{PIR}$ binds to DNA tightly even after the DNA probe was mutated to disrupt all the CpG sequences and to reduce the GC content from 79% to 21% (CpG46 mt DNA, see methods section for DNA sequence), with K_d = 73.7 ± 10.0 nM for CpG46 mt DNA (Fig. 4a). Conversely, PALI1$_{PIR}$ did not significantly increase the affinity of PRC2 to a G-tract RNA (Supplementary Fig. 4a), which interacts with core PRC2 subunits[35,36]. Accordingly, this

data indicates that the PRC2-interacting region of PALI1 facilitates high-affinity interactions between PRC2 and DNA, not RNA, without an apparent DNA-sequence selectivity.

Given that PALI1 facilitates DNA binding (Fig. 4a), we next wished to determine if PALI1 could facilitate substrate binding. We performed electrophoretic mobility-shift assays (EMSA) using a fluorescently labelled mononucleosome probe that was reconstituted using a 182 bp long DNA. The probe was reconstituted to allow for the simultaneous detection of the interactions between PRC2 and nucleosomes, at either the centred or the off-centred positions[37], or between PRC2 and the free DNA (Fig. 4b). In agreement with a previous work[38], the PRC2 core complex exhibited moderate affinity for mononucleosomes (K_d = 330 ± 30 nM). Remarkably, the $PRC2_{PIR\text{-}long}$ construct increased the affinity of PRC2 to nucleosomes by >15-fold (K_d = 19.0 ± 0.6 nM) compared to the PRC2 core complex (Supplementary Fig. 4b, c).

To determine if PALI1 enhances substrate binding in a mechanism linked to allosteric activation, we quantified the affinity of the PRC2 core complex to nucleosomes after pre-incubation with a PALI1-K1241me3 peptide (Fig. 4b and Supplementary Fig. 4b, c). The PALI1-K1241me3 peptide did not increase the affinity of the PRC2 core to nucleosomes (Supplementary Fig. 4b, c), thus suggesting that PALI1 facilitates an allosteric activation and substrate binding by independent mechanisms. Accordingly, the PRC2-$PALI1_{PIR\text{-}long}$ K1241A mutant, which is defective in the allosteric activation (Fig. 2a), binds nucleosomes with a similar affinity to the wild type PRC2-$PALI1_{PIR\text{-}long}$ (< 2-fold K_d) (Fig. 4b and Supplementary Fig. 4b, c). Collectively, these results indicate that the PRC2-binding domain of PALI1 facilitates DNA and substrate binding in addition to—and independent of—allosteric activation.

Overexpression of wild type PALI1 triggers cell differentiation in chronic myeloid leukaemia cells, with the phenotype abrogated by an allosteric-defective mutant.

One paper reported a large increment of global H3K27me3 in HeLa cells after C10ORF12 overexpression[33]. Yet, we did not detect a significant change of global H3K27me3 while overexpressing PALI1 in K562 (Supplementary Fig. 5a), HEK293T (Supplementary Fig. 5b, left) and HeLa (Supplementary Fig. 5b, right) cells. Little is known about the cellular function of PALI1, but our data thus far indicates a resemblance to JARID2 at the molecular level. JARID2 is frequently deleted in the leukemic transformation of chronic myeloid malignancies[39]. Accordingly, the overexpression of JARID2 leads to reduced proliferation in leukaemia cell lines and it has been proposed to serve as a tumour suppressor in leukaemia[40]. Given the functional resemblance with JARID2, we wished to determine if PALI1 has a negative effect on the proliferation of leukaemia cells. Indeed, competitive proliferation assay indicated that the overexpression of PALI1, but not the negative control LacZ, in a

human chronic myeloid leukaemia cell line (K562) leads to a strong reduction in cell proliferation (Fig. 5a, b).

In addition to reduced proliferation, we noticed that the overexpression of PALI1 in K562 cells, but not the negative control LacZ, led to the pelleted cells becoming red in colour (Fig. 5c). This observation suggested differentiation along the erythroid lineage[41], in accord with the reduced cell proliferation (Fig. 5a, b). We therefore set out to detect the erythroid differentiation marker CD235a, which increases during erythropoiesis[42], and the erythroid precursor marker CD71 that is lost in mature erythrocytes[43] (see illustration at Fig 5e). Indeed, the overexpression of PALI1, but not LacZ, led to increased expression of CD235a and reduction of CD71, in accord with differentiation along the erythroid lineage (Fig. 5d, e and Supplementary Fig. 5c, d).

We next assayed the PALI1 K1241A separation-of-function mutant, which is defective in allosteric activation (Fig. 2) and active in DNA binding (Fig. 4b and Supplementary Fig. 4b, c). The level of CD235a expression and red colour of the cells was comparable either if the PALI1 wild type or K1241A mutant was expressed (Fig 5c, d), and both the mutant and the wild type PALI1 significantly reduced K562 cell proliferation (Fig 5b). Yet, cells expressing the PALI1 K1241A mutant proliferated faster than cells expressing the wild type protein (Fig. 5b) and did not exhibit reduction of the CD71 erythroid precursor marker, similar to cells expressing the LacZ negative control (Fig. 5c, d and Supplementary Fig. 5c, d). These observations indicate that overexpression of PALI1 triggers K562 cell differentiation along the erythroid lineage, with the phenotype impaired in the K1241A mutant (Fig. 5e). These data, together with the data above (Fig. 1-4), imply a mechanistic resemblance between JARID2 to PALI1 at the molecular level, with a key role for PALI1 K1241 in allosterically activating PRC2.

DISCUSSION

Our data indicate that the PRC2-interacting domain of PALI1 is sufficient to enhance the HMTase activity of PRC2 by two independent mechanisms: (i) allosteric activation of catalysis (Fig. 2) and (ii) DNA binding (Fig. 4). Hence, as little as 193 amino acids in PALI1 (1058-1250) are sufficient to bind PRC2, bind DNA and stimulate methyltransferase, with the other >1500 amino acids in PALI1 likely available to engage in other tasks.

A convergent evolution between PALI1 and JARID2

The nucleation of H3K27me3 *de novo* takes place when transcription programs are changed during cell differentiation and newly repressed genes acquire the H3K27me3 mark. In the context of the PRC2.2, JARID2 facilitates the nucleation of H3K27me3[24], aided by its chromatin binding activity[19,20,44] and its ability to allosterically stimulate PRC2[21]. In the context of PRC2.1, MTF2 functions in the nucleation of H3K27me3[24], with a proposed contribution to its DNA-binding activity[45,46]. Although MTF2 does not allosterically stimulate PRC2, it coexists in the PRC2.1 complex with PALI1[16,18,25]. Through the discovery that PALI1 allosterically activates PRC2 (Fig. 2-3) and facilitates substrate binding (Fig. 4), we reveal a striking functional resemblance between PALI1 to JARID2: (i) Both PALI1 and JARID2 bind to nucleosomes: the former likely with the aid of its DNA binding activity (Fig. 4) and the latter through interactions with H2AK119-ubiquitinated chromatin[19,20]. (ii) JARID2 comprises the PRC2.2 complex together with the chromatin-binding subunit AEBP2 while PALI1 binds to the PRC2.1 complex together with a polycomb-like DNA binding subunit (PHF1, MTF2 or PHF19)[16,18,25]. (iii) PRC2 methylates PALI1 and JARID2, with the di- or tri-lysine then binds to EED for allosteric activation of PRC2 (Fig. 2-3 and [21]). (iv) For their interactions with EED, both PALI1 and JARID2 using an aromatic residue, located at the +1 position with respect to the methylated lysine (Fig. 3c). The importance of the +1 adjacent aromatic residue supported by the JARID2 F117A mutation that prevented both EED binding and the stimulation of PRC2[21].

Strictly, despite these mechanistic and structural similarities, PALI1 and JARID2 have no common ancestor: PALI1 is a vertebrate-specific protein[18] and JARID2 is conserved in fly and human (Fig. 3d). Therefore, we propose that PALI1 has emerged in vertebrates as the result of convergent evolution, under a selection pressure to mimic some of the molecular functions of JARID2 within the context of the PRC2.1 complex.

PALI1 provides PRC2 with means to gauge its own enzymatic activity before adding a stimulus

EZH2 automethylation is proposed to modulate the HMTase activity of PRC2 in response to molecular cues, including the presence of histone H3 tails and SAM concentration[31]. A similar principle might apply for PALI1. Future studies are still needed in order to identify whether PALI1 (and JARID2) are methylated in cis or by another PRC2 complex and if the degree of methylation changes during normal development or in cancer and other pathologies. Yet, our analysis indicates that PRC2 has the capacity to methylate PALI1 (Fig 1b, c) and implies that the degree of PALI1 K1241 methylation varies between cell lines (Fig. 1a). Given these observations, it is plausible that PALI1 provides PRC2 with means to gauge its own enzymatic activity before applying an additional stimulus.

356 **PALI1 as a potential bridge between the H3K27me3 and H3K9me2 repressive marks**

357 Amino acids in regions of PALI1 dispensable for the regulation of PRC2 bind to the H3K9-
358 methyltransferase G9a[18]. PRC2 and G9a share a portion of their genomic targets[47] and are physically
359 associated[47,48]. In ES cells, G9a contributes to H3K27 methylation *in vivo*, with the global
360 H3K27me1—but not H3K27me2/3—reduced upon knockout of G9a[49]. Affinity purification mass
361 spectrometry (AP-MS) with either C10ORF12[25] or PALI1[18] used as baits detected the subunits of the
362 G9a–GLP H3K9me1/2 methyltransferase complex, including G9a, GLP and WIZ. Importantly, the
363 PRC2-interacting domain of PALI1 is distinct to its G9a-interacting region[18]. Our data indicate that
364 the PRC2-binding region of PALI1 is sufficient (i) to bind to PRC2 (Fig. 1), (ii) to promote PRC2
365 nucleosome substrate binding (Fig. 4) and (iii) to trigger an allosteric stimulation of catalysis (Fig. 2):
366 three prerequisites for the nucleation of H3K27me3 *de novo*[21]. While experiments in cells are
367 required in order to determine if PALI1 could nucleate both H3K27me3 and H3K9me2 *de novo*
368 during cell differentiation, it does have the molecular characteristics for that.

369 Recent studies showed that PRC2.1 and PRC2.2 synergise and share most of their target genes[11],
370 with their accessory subunits collectively required[12]. While a previous attempt to identify the binding
371 sites of PALI1 on chromatin using ChIP was reported as unsuccessful[18], the PRC2.1 complex is
372 localised at a minority of unique genomic sites, independently of PRC2.2[11]. We measured small
373 variation between the affinity of the PRC2-PALI$_{PIR}$ complex for the CpG-reach DNA (CpG46 K_d = 155 ±
374 26 nM) comparing a size-matched DNA lacking CpG sequences (CpG46 mt; K_d = 73.7 ± 10 nM) (Fig.
375 4a). It is, therefore, possible that PALI1 could allow for some degree of target-specificity, utilising
376 some variations in affinities for DNA combined with context-specific chromatin binding. Such
377 context-specific binding could be attributed to interactions with the G9a complex and the CTBP
378 proteins binding to the N-terminal domain of PALI1[18]. This model is in agreement with the view that
379 a combination of factors and interactions are responsible for the recruitment of PRC2 to its target
380 genes[7].

381 Collectively, our data indicate that the PRC2-binding domain of PALI1 enhances H3K27-
382 methyltransferase by two independent mechanisms (Fig. 6): (i) DNA- and substrate-binding and (ii)
383 allosteric stimulation. The remarkable mechanistic resemblance between PALI1 and JARID2 indicates
384 convergent evolution for the emergence of mechanisms for the regulation of the PRC2.1 and PRC2.2
385 complexes, respectively. More broadly, it implies that subunit-induced allosteric activation is an
386 indispensable property of a holo-PRC2 complex in vertebrates.

387

ACKNOWLEDGMENTS

We would like to thank the Monash Biomedical Proteomics Facility for providing instrumentation and technical support, Monash FlowCore for sorting of cells and for assistance with flow cytometry analysis and Monash Molecular Crystallization Facility for initial crystallization screening. This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of Australia's Nuclear Science and Technology Organisation (ANSTO), and made use of the Australian Cancer Research Foundation (ACRF) detector. We also thank the support from the MASSIVE HPC facility (www.massive.org.au). We thank Ruby Law and her lab members for advice, discussions, training and assistance with the X-ray crystallography data collection and structure determination. Q.Z. holds an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE180100219). S.C.A. is supported through an Australian Government Research Training Program (RTP) Scholarship. B.M.O. is supported through an Australian Government RTP Scholarship and also by the Monash Graduate Excellence Scholarship. C.D. is an EMBL-Australia Group Leader and acknowledges support from the ARC (DP190103407) and the NHMRC (APP1162921 & APP1184637).

AUTHOR CONTRIBUTION

Q.Z., S.C.A. and S.F.F. prepared reagents. Q.Z., S.C.A., and B.M.O. developed assays. Q.Z., S.C.A. and S.F.F. carried out experiments. Q.Z., S.C.A., S.F.F. and V.L. analysed data. Q.Z., S.C.A. and C.D. wrote the manuscript. Q.Z., S.C.A and C.D. designed the project. C.D. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

METHODS

Protein expression and purification

The cloning of constructs for the expression of the full-length sequences encoding for human EZH2, SUZ12, RBBP4, EED and AEBP2 (UniProtKB: Q15910-2, Q15022-1, Q09028-1, O75530-1 and Q6ZN18-2, respectively) into a pFastBac1 expression vectors, modified to include either a PreScission-cleavable N-terminal hexahistidine-MBP tag or TEV-cleavable N-terminal hexahistidine tag, were previously described[35,50,51]. Full-length PALI1 was assembled and subcloned into the pFB1.HMBP.A3.PrS.ybbR vector using Gibson Assembly (primers as indicated in Table S2) with gene synthesised N-terminal fragment (amino acids 1-310) of PALI1 (GeneScript) and commercially available C10orf12 cDNA clone encoding PALI1 amino acids 311-1557 (Millennium Science #MHS6278-202756878) as templates (see Supplementary Table 2 for the full-length PALI1 ORF sequence). The PIR (amino acids 1058-1250) and PIR-long (amino acids 1058-1329) fragments of PALI1 were subcloned into the pFB1.HMBP.A3.PrS.ybbR vector digested with XmaI and XbaI (NEB), under a PreScission-cleavable N-terminal hexahistidine-MBP tag, using Gibson Assembly (NEB #E2611L) with primers as indicated in Supplementary Table 2.

Mutations were introduced to plasmids coding for PALI1 and its truncations using Takara PrimeSTAR HS DNA Polymerase (Clontech #R045A), with primers indicated in Supplementary Table 2. Baculovirus production, titration, infection, and cell harvesting and the purification of PRC2, PRC2-PALI1$_{PIR}$, PRC2-PALI1$_{PIR\text{-}long}$ and their mutants performed as previously described[35]. The expression and purification of PRC2 in complexes with MBP-fused PALI1 truncations, PRC2-[MBP-PALI1$_{PIR}$] and PRC2-[MBP-PALI1$_{PIR\text{-}long}$] performed as above, with the exception that PRC2 core subunits were expressed under TEV-cleavable hexahistidine-tag and PALI1 truncations under PreScission-cleavable N-terminal hexahistidine-MBP tag. During the purification of these constructs, only TEV used to clave tags selectively from the PRC2 core subunits, with the MBP tag on the PALI1 construct remained intact. All the complexes were snap-frozen in liquid nitrogen and stored at -80 °C as single-use aliquots.

For the structure-function study of EED, two fragments of human EED (amino acids 40-441 and 76-441) were subcloned into a pGEX-MHL expression vector with a TEV-cleavable N-terminal GST-tag (a gift from the lab of Asst. Prof. Yufeng Tong, University of Windsor) using primers as in Supplementary Table 2. The recombinant proteins were overexpressed in *E.coli* BL21 (DE3) at 17 °C overnight and then purified by Glutathione-agarose (Sigma #G4510). Briefly, harvested cells were resuspended in an ice-cold lysis buffer (20 mM Tris-HCl pH 7.5 at 25 °C, 250 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM DTT) and lysed using sonication. The cleared lysate

was batch-bound to Glutathione-agarose and washed using ice-cold 10 column volumes (c.v.) of lysis buffer before proteins were eluted using ice-cold elution buffer (20 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 10 mM reduced glutathione). Tag cleaved using TEV, overnight at 4 °C. The protein was subsequently purified by heparin HP column (GE #17040701), using a buffer containing 20 mM Tris-HCl pH 7.5 at 4 °C and a 150 mM to 2000 mM NaCl gradient. Gel filtration purification carried out using HiLoad 16/600 Superdex 200 size exclusion column (GE #28-9893-35), using a buffer containing 20 mM HEPES pH 7.5 and 150 mM NaCl. The peak fractions were pooled, concentrated to a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), and snapped-frozen as single-use aliquots.

Nucleosome reconstitution

Recombinant human histones were purified from inclusion bodies and reconstituted into histone octamers as previously described[52]. For the generation of mononucleosomes, unlabelled or Cy5 labelled 182-base-pair DNA including one copy of the 601 Widom sequence was PCR amplified as previously described[35]. DNA was purified via HiTrap Q HP column (GE #17-1154-01) with 10 c.v. gradient starting with buffer A (20 mM Tris-HCl, pH 7.5 at 25 °C, 150 mM NaCl) into 50 % buffer B (20 mM Tris-HCl, pH 7.5 at 25 °C, 2 M NaCl). The peak fractions were pooled, DNA was concentrated by ethanol precipitation and was then dissolved in TE buffer. Mononucleosomes were reconstituted using the salt dialysis method[52]. Reconstituted mononucleosomes were dialyzed against a buffer consisting of 20 mM Tris-HCl pH 7.5 (at 25 °C), 25 mM KCl, 1 mM EDTA and 1 mM DTT and concentrated with Amicon Ultra-0.5 mL centrifugal filter (Merck #UFC503096). Nucleosomes were stored in 4 °C and the quality of nucleosomes was assessed by 5-6 % TBE gel.

***In vitro* HMTase activity assays using radiolabelled S-adenosyl-l-methionine**

HMTase activity assays were performed as previously described[35], with some modifications. In brief, each 10 µL HTMase reaction contained 500 nM PRC2, 2 µM mononucleosomes and 5.0 µM S-[methyl-14C]-adenosyl-l-methionine (PerkinElmer, #NEC363050UC) in the presence or absence of stimulatory or control peptides in concentrations as indicated in the text. For HMTase assays in the presence or absence of the PRC2 allosteric inhibitor A395 or the negative control A395N, PRC2 concentration was adjusted to 200 nM and the concentration of other reagents remained the same. For HMTase assays carried out for the identification of PALI1 methylations with the aid of 3C protease, 800 ng protease per reaction was added. All the reactions were incubated for 1 h at 30 °C in buffer containing 50 mM Tris-HCl pH 8.0 at 30 °C, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM 2-mercaptoethanol, 0.1 mg/ml BSA (NEB #B9000) and 5% v/v glycerol. The reactions were then stopped by adding 4× LDS sample buffer (Thermo Fisher Scientific, #NP0007) to a final concentration

of 1×. Samples were then heated at 95 °C for 5 min prior to being subjected to 16.5 % SDS-PAGE. Gels were stained with InstantBlue Coomassie protein stain (Expedeon, #ISB1L) before vacuum-drying for 1 h at 80 °C. Dried gels were then exposed to a storage phosphor screen (GE Healthcare) for 1-7 days before acquiring radiograms using Typhoon 5 Imager (GE Healthcare). Densitometry was carried out using ImageJ[53]. All experiments were performed in triplicate.

Liquid chromatography with tandem mass spectrometry (LC-MS/MS) and MS/MS data analysis

For the identification of methylations in the recombinant PRC2-PALI1$_{PIR\text{-}long}$ complex, 0.5 µM protein was incubated in the presence or absence of 20 µM SAM (NEB #B9003) in 25 mM HEPES pH 8.0, 50 mM NaCl, 2 mM $MgCl_2$, 2 mM 2-mercaptoethanol and 10 % glycerol for 1 hour at 30 °C, in a total final volume of 85 µL. The proteins were then alkylated, subjected to tryptic digestion and the tryptic peptides where treated and analysed by MS/MS as previously described[35]. Methylated residues were identified using MaxQuant[54] by searching against a database containing the protein subunits of the PRC2-PALI1$_{PIR\text{-}long}$ complex.

Publicly available A/IP-MS data (identifiers: PXD004462[27], PXD012547[28], PXD013390[29], PXD012354[26], and PXD003758[15]) were downloaded from ProteomeXchange and used to identify methylations of PRC2 subunits. Methylated residues were identified using MaxQuant[54] by searching against a database containing the human or mouse proteome from Uniprot (Proteome ID UP000005640 or UP000000589, respectively), with the sequences for human or mouse PALI1 and PALI2 appended to them (Uniprot identifiers: Q96JN0-3, A0A1B0GVP4, A0A571BF12 and A0A571BEC4). The amino acid numbering of methylated amino acids throughout the text and figures are based on the human numbering, with the mouse amino acids numbers converted to human numbers for compatibility, based on pairwise sequence alignment[55].

Crystallization and structure determination

Purified EED (amino acids 76-441) protein at 1.5-2.0 mg/mL was mixed with PALI1 peptides at 1:5 EED:PALI1-peptide molar ratio and the mixture were incubated at 4 °C for 1 h before subjected to crystallization trials. Crystals were grown using the vapour diffusion method in a buffer containing 3.5-3.9 M sodium formate, 10 mM TCEP and 5% glycerol. Crystals were soaked in a reservoir solution with 10-20% glycerol before flash-freezing in liquid nitrogen.

X-ray diffraction data were collected at MX2 beamline of the Australian Synchrotron[56]. All structures were determined by molecular replacement using PHASER[57,58], using EED-H3K27me3 structures[23,59] as the search model (PDB: 3JZG[59] and 3IIW[23]). REFMAC5[60] and PHENIX[61] were used for refinement and Coot[62] was used for manual structure building and visualization. Figures were generated with

PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Crystal diffraction data, refinement statistics and PDB accessions for these structures are in Table 1.

DNA binding assays using fluorescence anisotropy

3'-fluorescein-labelled CpG46 and CpG46 mt DNAs were synthesized by Integrated DNA Technologies, Inc. The CpG46 DNA probe was destined to form a hairpin including a 46 bases long double-stranded DNA with a nucleotide sequence originating from a CpG island of the human CDKN2B PRC2-target gene, with this sequence mutated in CpG46 mt DNA to remove the CpG sequences and reduce the GC content.

CpG46 sequence:
5'GGCGCCCTGCCCCGCCTCGCTCTGGCAGAGTGGGGAGCCAGCCGGCGCTAGCCGGCTGGCTCCCCACTCTG
CCAGAGCGAGGCGGGGCAGGGCGCC

CpG46 mt sequence:
5'AATATTTCATTTTATTCTATCTCAATGAGACAAAAGATTGATTAATGCTAATTAATCAATCTTTTGTCTCATTG
AGATAGAATAAAATGAAATATT

DNA was incubated for 2 min at 95 °C in 10 mM Tris-HCl pH 7.5 (at 25 °C) and was immediately snap-cooled on ice for 2 min. Next, DNA was allowed to fold at 37 °C for 30 min in binding buffer (50 mM Tris-HCl pH7.5 at 25 °C, 100 mM KCl, 2 mM 2-mercaptoethanol, 0.1 mg/mL BSA, 0.05 % NP-40). Two-fold serial dilutions of protein were made into a binding buffer and were added into the fluorescently-labelled DNA probe. The DNA probe concentration was 5 nM for each 40 µL reaction volume and the mixtures were equilibrated at 30 °C for 30 min before measurement. Fluorescence anisotropy data were collected using a PHERAstar plate reader (BMG Labtech) at 30 °C. The background was subtracted from protein-free samples. Data were fitted with GraphPad Prism 8 software using non-linear regression for specific binding with a Hill slope function. All experiments were performed in triplicates that were carried out on different days.

RNA binding assays using fluorescence anisotropy

The RNA-binding affinities of proteins were quantified using fluorescence anisotropy. Experiments were carried out as we previously described[35]. Briefly, a 3' fluorescein-labelled G4 24 RNA with the sequence $(UUAGGG)_4$ was used and the reaction took place as above, with the exception that the initial incubation at 95 °C was limited to 1 min and the binding buffer was 50 mM Tris-HCl pH7.5 at

25 °C, 200 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM 2-mercaptoethanol, 0.1 mg/mL BSA, 0.05% NP-40 and 0.1 mg/mL fragmented yeast tRNA (Sigma, #R5636).

Peptide binding assays using fluorescence anisotropy

For assaying the affinity of the JARID2-K116me3 peptide for EED, various concentrations of EED (amino acids 40-441) were incubated with 40 nM of 5-FAM labelled JARID2-K116me3 peptide in binding buffer (50 mM Tris-HCl pH 7.5 at 25 °C, 100 mM KCl, 2 mM 2-mercaptoethanol, 0.1 mg/mL BSA, 0.05 % NP-40, 2.5% glycerol) at 30 °C for 30 min before a fluorescence anisotropy measurement took place using a PHERAstar plate reader. Data processing was carried out as previously described[63], with some modifications. Specifically, with changing the concentration of EED protein (P), we recorded $\Delta Robs$: the observable anisotropy of the mixtures after the subtraction of the observable anisotropy of the 5-FAM labelled JARID2-K116me3 peptide ligand. Data were fitted to equation (1), below, by using nonlinear least-squares fit (Matlab, MathWorks) to estimate the anisotropy difference Δr and the dissociation constant between the protein EED to the ligand JARID2-K116me3 peptide, K_L:

$$(1)\quad \Delta Robs = \frac{\Delta r[\mathrm{PL}]}{[\mathrm{PL}]+[\mathrm{L}]}$$

where the concentrations of the protein-ligand complex [PL] and the free ligand [L] are calculated from equations (2) and (3) below, respectively, with P_0 and L_0 indicates the total concentration of the protein and the ligand, respectively.

$$(2)\quad [\mathrm{PL}] = \frac{(K_L+[\mathrm{P_0}]+[\mathrm{L_0}])-\sqrt{(K_L+[\mathrm{P_0}]+[\mathrm{L_0}])^2-4[\mathrm{P_0}][\mathrm{L_0}]}}{2}$$

$$(3)\quad [\mathrm{L}] = [\mathrm{L_0}] - [\mathrm{PL}]$$

Fluorescence anisotropy displacement titrations were used to assay the dissociation constants of the unlabelled peptides (N) and the protein (P). Assays were carried out as described above, with the exception that two-fold serial dilutions of unlabelled peptides were combined with EED at a final concentration of 10 µM and 5-FAM labelled JARID2-116me3 peptide at a final concentration of 40 nM before anisotropy data were collected as described above. Data processing was carried out as previously described[63]. Specifically, $\Delta Robs$ were recorded for each peptide [N] concentration point and equation (1) used to estimate K_N and Δr as the fitting parameters. K_N is the equilibrium

dissociation constant for binding of the unlabelled peptide N to the protein P. N_0 is the total concentration of N. [PL] and [L] are calculated in a different way:

$$[\mathrm{PL}] = \frac{[\mathrm{L_0}][\mathrm{P}]}{K_L + [\mathrm{P}]}$$

and $[\mathrm{L}] = [\mathrm{L_0}] - [\mathrm{PL}]$

where K_L are obtained from the measurement above, and $[\mathrm{P}]$ is one of the roots of the cubic equation:

$$[\mathrm{P}]^3 + c_2[\mathrm{P}]^2 + c_1[\mathrm{P}] + c_0 = 0$$

where

$$c_2 = K_L + K_N + [\mathrm{N_0}] + [\mathrm{L_0}] - [\mathrm{P_0}]$$
$$c_1 = K_L K_N - K_N[\mathrm{P_0}] + K_L[\mathrm{N_0}] - K_L[\mathrm{P_0}]$$
$$c_0 = -K_L K_N[\mathrm{P_0}]$$

Electrophoretic mobility shift assay (EMSA)

Cy5 labelled nucleosomes were diluted using binding buffer (50 mM Tris-HCl, pH7.5 at 25 °C, 100 mM KCl, 2 mM 2-mercaptoethanol, 0.05 % v/v NP-40, 0.1 mg/mL BSA, 5 % glycerol). Two-fold serial dilutions of protein in binding buffer were combined with nucleosome probes, to a final probe concentration of 5 nM. The reaction mixtures were incubated at 4 °C for 30 min and then subjected to non-denaturing gel electrophoresis at 6.6 V/cm over a 0.7 % agarose gel buffered with 1 × TBE at 4 °C for 1 h. Gels were imaged using Typhoon 5 Imager (GE Healthcare) to record signals from the Cy5 dye. The fractions of bound nucleosomes were calculated based on the unbound nucleosomes band, with the densitometry analysis carried out using ImageJ[53]. Data were fitted with GraphPad Prism 8 software using non-linear regression for specific binding with Hill slope function. All experiments were performed in triplicate.

Cell culture

K562 cells were cultured in RPMI-1640 (Merck #R8758) growth medium and HEK293T and HeLa were cultured in DMEM growth medium. In all cases, growth media were supplemented with 10 % FBS (Cellsera AU-FBS/SF) and 1 % (v/v) penicillin-streptomycin (Thermo Scientific #15140122) and incubated at 37°C with 5% CO2. K562 and HeLa cells were acquired from ATCC and cells were tested periodically for mycoplasma contamination.

Plasmid transfection, generation of lentiviruses and lentiviral transduction

Flag-PALI1 WT and K1241A mutant, and Flag-LacZ (ORF originated from Addgene #25893), were subcloned into SmaI (NEB #R0141) linearized pHIV-EGFP (Addgene #21373) or pHIV-dTomato (Addgene #21374) vectors using Gibson Assembly (see Table S2 for primers) and NEB stable Competent *E.coli* (NEB #C3040). For plasmid transfection, approximately 10^6 HEK293T or HeLa cells were seeded in a 6 well plate. The following day, the medium was replaced with 2 ml of antibiotic-free DMEM. The transfection mixture for HEK293T cells was created by adding 9 µL Lipofectamine™ LTX Reagent with 3 µL PLUS™ Reagent (Thermo Scientific #15338100) and 3 µg of a plasmid to 500 µL Opti-MEM™ (Thermo Scientific #31985062). The transfection mixture for HeLa cells instead consisted of 6 µL Lipofectamine™ LTX Reagent with 3 µL PLUS™ Reagent and 1.25 µg of transfer plasmid in 500 µL Opti-MEM™ (Thermo Scientific #31985062). The transfection mixture was incubated at room temperature for 25 minutes and then added to the cells before returning them to the incubator. The growth medium was replaced after 24 hours. For cell harvesting ahead of immunoblotting, 48 hours after transfection the media was removed and replaced with 350 µL of Laemmli buffer (1 % (v/v) SDS, 12.5 % (v/v) glycerol, 35 mM Tris pH 7.5 at 25 °C, 0.01 % (w/v) bromophenol blue, 5 mM $MgCl_2$, 1 % (v/v) 2-mercaptoethanol) and 25 U/mL Benzonase (Merck #70746).

For the generation of lentiviruses, HEK293T cells were transfected as above, with 0.5 µg pMD2.G plasmid, 1 µg psPAX2 plasmid and 1.5 µg of transfer plasmid. After 48 and 72 hours, the culture supernatant containing the lentivirus was collected and stored at -80 °C. For lentiviral transduction, 200 µL of lentiviral supernatant was added to $3x10^4$ K562 cells to a final volume of 500 µL, with polybrene at a final concentration of 8 µg/mL.

Immunoblotting

For nuclear fractionation of K562 cells, cells were washed twice with phosphate-buffered saline (PBS) by centrifugation at 500 g for 5 minutes then resuspended in cytoplasmic extraction buffer (20 mM Tris pH 7.5 at 25 °C, 0.1 mM EDTA, 2 mM $MgCl_2$, 20 mM BME and protease inhibitor cocktail (Sigma #4693132001)) to a density of $2x10^7$ cells/mL. The cells were incubated for 2 minutes at room temperature then 10 minutes on ice before adding NP-40 to a concentration of 1% (v/v) and mixing. Samples were centrifuged at 4 °C and 500 g for 3 minutes and the supernatant was kept as the cytoplasmic fraction. The nuclear fraction was washed in cytoplasmic extraction buffer with 1% (v/v) NP-40 twice, by centrifugation at 4 °C at 500 g for 3 minutes, and was then resuspended in Laemmli buffer.

Samples contain 50 µg total protein were loaded on a 10 % or 16.5 % acrylamide gel for SDS-PAGE and then transferred to a nitrocellulose membrane (GE Life Sciences #10600002). Membranes were incubated in blocking buffer (Thermo Scientific #37539) for 1 hour at room temperature before applying antibodies. Signal was developed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific #34580) and images were taken on a ChemiDoc™ imager. All experiments were performed in triplicates.

The antibodies used for immunoblotting include: anti-Flag HRP-conjugated (Sigma #A8592, 1:1000), anti-LCOR (Merck #ABE1367, 1:250), anti-actin (Sigma #A2066, 1:800), anti-EZH2 (Active Motif #39875, 1:10000), anti-H3 (Abcam #Ab1791, 1:100000), anti-H3K27me3 (Merck #07-449, 1:25000), anti-mouse HRP-conjugated (Jackson Immuno-Reasearch #715-035-150, 1:5000), anti-rabbit HRP-conjugated (Santa Cruz Biotechnology #sc-2357, 1:5000).

Competitive cell proliferation assay

Cells were transduced with lentiviruses carrying the specified gene constructs and were then cultured, each sample separately, for 7 days. At this point, an equal number of eGFP or dTomato positive cells were sorted using flow cytometry, as described below, and combined into the same collection tube and placed in the same well for the competition experiment. Cells of the two competing treatments were cultured together in the same well for 7 days. Next, the number of eGFP- and dTomato-expressing cells counted using flow cytometry with the B530-A and YG586-A detectors, respectively. Three independent biological replicates, starting from lentivirus transduction, were initiated on three different days and were carried out as described above.

Detection of eGFP and dTomato using Flow cytometry

Before sorting or analysis by flow cytometry, cells were centrifuged at 500 g for 5 minutes, and the supernatant removed. The cells were then resuspended in flow cytometry buffer (PBS supplemented with 10 % FBS and 615 µM EDTA) to a density of approximately 10^7 cells/mL, ran through a cell strainer (Falcon 352235) and kept on ice. For the detection of GFP or dTomato, the cells were sorted by flow cytometry on a BD Influx™ cell sorter using the 488 nm or the 561 nm lasers, respectively. For the selection of transduced cells, gates for sorting were set to include the top 10 or 20 % of the GFP or dTomato positive cells in the samples transduced with PALI1 wild type, based on the first replicate. For the analysis of transduced cells, approximately 0.5-1.0x10^5 intact single cells were analysed on a BD LSRFortressa™ X-20 analyser, with the threshold for GFP or dTomato positive cells was defined based on the intensity observed at the top 0.1 % of untransduced K562 cells. Data were analysed using BD FACSDiva™ and GraphPad Prism.

675 **Detection of CD235a or CD71 using Flow Cytometry**

676 Cells were transduced with lentiviruses carrying the specified gene constructs and were then
677 cultured for 7 days. Cells were sorted for high GFP expression, as described above, and then cultured
678 for 7 additional days before the growth medium was removed by spinning the cells at 500 g for 5
679 minutes. Cells at a density of $2x10^7$ cells/mL were incubated for 1 hour in flow cytometry buffer with
680 5 µL of Pacific Blue™ conjugated anti-CD235a antibody (BioLegend #349108) and 4 µg of
681 unconjugated mouse IgG2a (Merck #M5409) per 100 µL for detection of CD235a, or 2.5 µL of
682 Brilliant Violet 421™ conjugated anti-CD71 antibody (Biolegend #113813) per 100 µL for the
683 detection of CD71. The cells were centrifuged again at 500 g for 5 minutes and the supernatant
684 removed, then washed with antibody-free flow cytometry buffer. The cells were then analysed by
685 flow cytometry for the quantification of CD235a in the GFP positive cells using the V450-A and B530-
686 A detectors, respectively. Three independent biological replicates, starting from lentivirus
687 transduction, were initiated on three different days and were carried out as described above. The
688 data was analysed using FlowJo and GraphPad Prism.

689

690

691 **REFERENCES**

692 1. Holland, L.Z. et al. The amphioxus genome illuminates vertebrate origins and cephalochordate
693 biology. *Genome Res* **18**, 1100-11 (2008).
694 2. Prachumwat, A. & Li, W.H. Gene number expansion and contraction in vertebrate genomes
695 with respect to invertebrate genomes. *Genome Res* **18**, 221-32 (2008).
696 3. Larroux, C. et al. Genesis and expansion of metazoan transcription factor gene classes. *Mol*
697 *Biol Evol* **25**, 980-96 (2008).
698 4. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in
699 development and disease pathogenesis. *Annu Rev Biochem* **81**, 65-95 (2012).
700 5. Scheuermann, J.C., Gutierrez, L. & Muller, J. Histone H2A monoubiquitination and Polycomb
701 repression: the missing pieces of the puzzle. *Fly (Austin)* **6**, 162-8 (2012).
702 6. Schuettengruber, B., Bourbon, H.M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb
703 and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57 (2017).
704 7. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**,
705 343-9 (2011).
706 8. Yu, J.R., Lee, C.H., Oksuz, O., Stafford, J.M. & Reinberg, D. PRC2 is high maintenance. *Genes*
707 *Dev* **33**, 903-935 (2019).
708 9. Laugesen, A., Hojfeldt, J.W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment
709 and H3K27 Methylation. *Mol Cell* **74**, 8-18 (2019).
710 10. Deevy, O. & Bracken, A.P. PRC2 functions in development and congenital disorders.
711 *Development* **146**(2019).
712 11. Healy, E. et al. PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell* **76**,
713 437-452 e6 (2019).

12. Hojfeldt, J.W. et al. Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol Cell* **76**, 423-436 e3 (2019).
13. Beringer, M. et al. EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol Cell* **64**, 645-658 (2016).
14. Zhang, Z. et al. PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* **29**, 229-40 (2011).
15. Kloet, S.L. et al. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* **23**, 682-690 (2016).
16. Hauri, S. et al. A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* **17**, 583-595 (2016).
17. Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A. & Vermeulen, M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* **41**, e28 (2013).
18. Conway, E. et al. A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol Cell* **70**, 408-421 e8 (2018).
19. Cooper, S. et al. Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat Commun* **7**, 13661 (2016).
20. Kalb, R. et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-71 (2014).
21. Sanulli, S. et al. Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell* **57**, 769-783 (2015).
22. Hansen, K.H. et al. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* **10**, 1291-300 (2008).
23. Margueron, R. et al. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-7 (2009).
24. Oksuz, O. et al. Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* **70**, 1149-1162 e5 (2018).
25. Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V. & Kuroda, M.I. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* **111**, 2488-93 (2014).
26. Ragazzini, R. et al. EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat Commun* **10**, 3858 (2019).
27. Oliviero, G. et al. Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells. *Mol Cell Proteomics* **15**, 3450-3460 (2016).
28. Wassef, M. et al. EZH1/2 function mostly within canonical PRC2 and exhibit proliferation-dependent redundancy that shapes mutational signatures in cancer. *Proc Natl Acad Sci U S A* **116**, 6075-6080 (2019).
29. Jain, S.U. et al. PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nat Commun* **10**, 2146 (2019).
30. Guo, A. et al. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol Cell Proteomics* **13**, 372-87 (2014).
31. Wang, X. et al. Regulation of histone methylation by automethylation of PRC2. *Genes Dev* **33**, 1416-1427 (2019).
32. Lee, C.H. et al. Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. *Genes Dev* **33**, 1428-1440 (2019).
33. Shi, Y. et al. C10ORF12 modulates PRC2 histone methyltransferase activity and H3K27me3 levels. *Acta Pharmacol Sin* **40**, 1457-1465 (2019).
34. He, Y. et al. The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nat Chem Biol* **13**, 389-395 (2017).
35. Zhang, Q. et al. RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* **26**, 237-247 (2019).

36. Long, Y. et al. Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *Elife* **6**(2017).
37. Luger, K., Rechsteiner, T.J. & Richmond, T.J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16 (1999).
38. Wang, X. et al. Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Struct Mol Biol* **24**, 1028-1038 (2017).
39. Puda, A. et al. Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am J Hematol* **87**, 245-50 (2012).
40. Su, C.L., Deng, T.R., Shang, Z. & Xiao, Y. JARID2 inhibits leukemia cell proliferation by regulating CCND1 expression. *Int J Hematol* **102**, 76-85 (2015).
41. Andersson, L.C., Jokinen, M. & Gahmberg, C.G. Induction of erythroid differentiation in the human leukaemia cell line K562. *Nature* **278**, 364-5 (1979).
42. Andersson, L.C., von Willebrand, E., Jokinen, M., Karhi, K.K. & Gahmberg, C.G. Glycophorin A as an erythroid marker in normal and malignant hematopoiesis. *Haematol Blood Transfus* **26**, 338-44 (1981).
43. Marsee, D.K., Pinkus, G.S. & Yu, H. CD71 (transferrin receptor): an effective marker for erythroid precursors in bone marrow biopsy specimens. *Am J Clin Pathol* **134**, 429-35 (2010).
44. Son, J., Shen, S.S., Margueron, R. & Reinberg, D. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* **27**, 2663-77 (2013).
45. Perino, M. et al. MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet* **50**, 1002-1010 (2018).
46. Li, H. et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287-291 (2017).
47. Mozzetta, C. et al. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol Cell* **53**, 277-89 (2014).
48. Maier, V.K. et al. Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol Cell Proteomics* **14**, 1435-46 (2015).
49. Wu, H. et al. Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell Res* **21**, 365-7 (2011).
50. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250-7 (2013).
51. Davidovich, C., Goodrich, K.J., Gooding, A.R. & Cech, T.R. A dimeric state for PRC2. *Nucleic Acids Res* **42**, 9236-48 (2014).
52. Luger, K., Rechsteiner, T.J. & Richmond, T.J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* **304**, 3-19 (1999).
53. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-5 (2012).
54. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-72 (2008).
55. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641 (2019).
56. Aragao, D. et al. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J Synchrotron Radiat* **25**, 885-891 (2018).
57. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
58. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).
59. Xu, C. et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A* **107**, 19266-71 (2010).

60. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).
61. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
62. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
63. Justin, N. et al. Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* **7**, 11316 (2016).
64. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259 (2019).

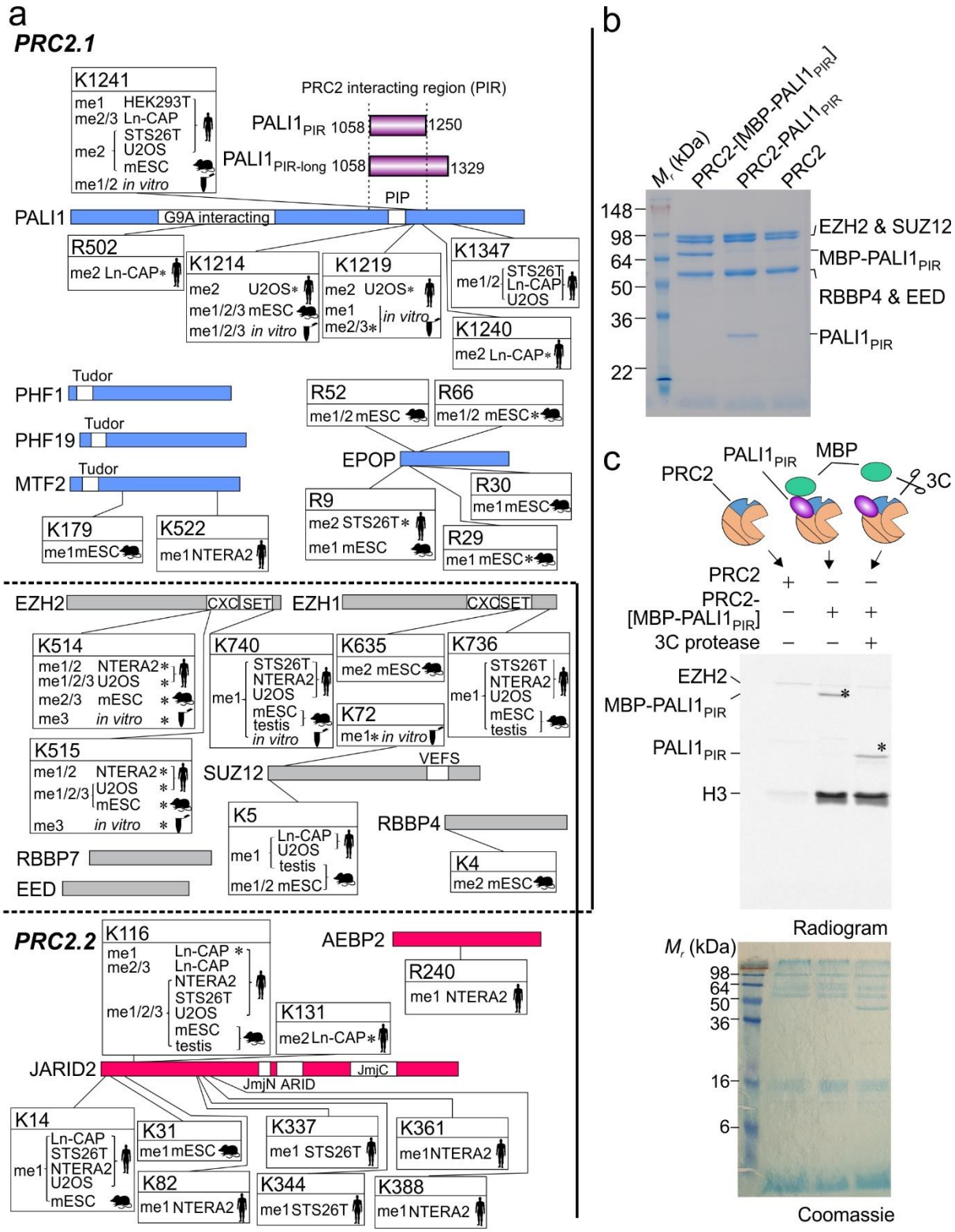


Fig. 1. PALI1 is methylated *in vitro* and *in vivo*.

a, Schematic representation of the PRC2 methylome *in vivo* and *in vitro,* as identified from MS/MS data. Mouse and human icon represent the organism of origin and cell lines are indicated (see methods section for references and accession numbers of the raw MS/MS data). Test tube icons representing methylations in the purified recombinant human PRC2-PALI1$_{PIR\text{-}long}$. PALI truncations used in this study are indicated in purple (upper right). Residues are indicated with asterisks where the position probability of the methylation is less than 0.95 (see Supplementary Table 1 for the

values). PRC2.1 and PRC2.2 accessory subunits are in blue and red, respectively, and core subunits are in grey. **b,** Coomassie blue-stained SDS-PAGE of recombinant human PRC2-PALI1$_{PIR}$ complexes, as indicated. **c,** HTMase assay of the PRC2-[MBP-PALI1$_{PIR}$] complex using mononucleosomes substrate were carried out in the presence or absence of C3 protease to confirm that PALI1 $_{PIR}$ is methylated. The MBP-cleaved and uncleaved PALI1 $_{PIR}$ bands are indicated on the radiogram with asterisks.

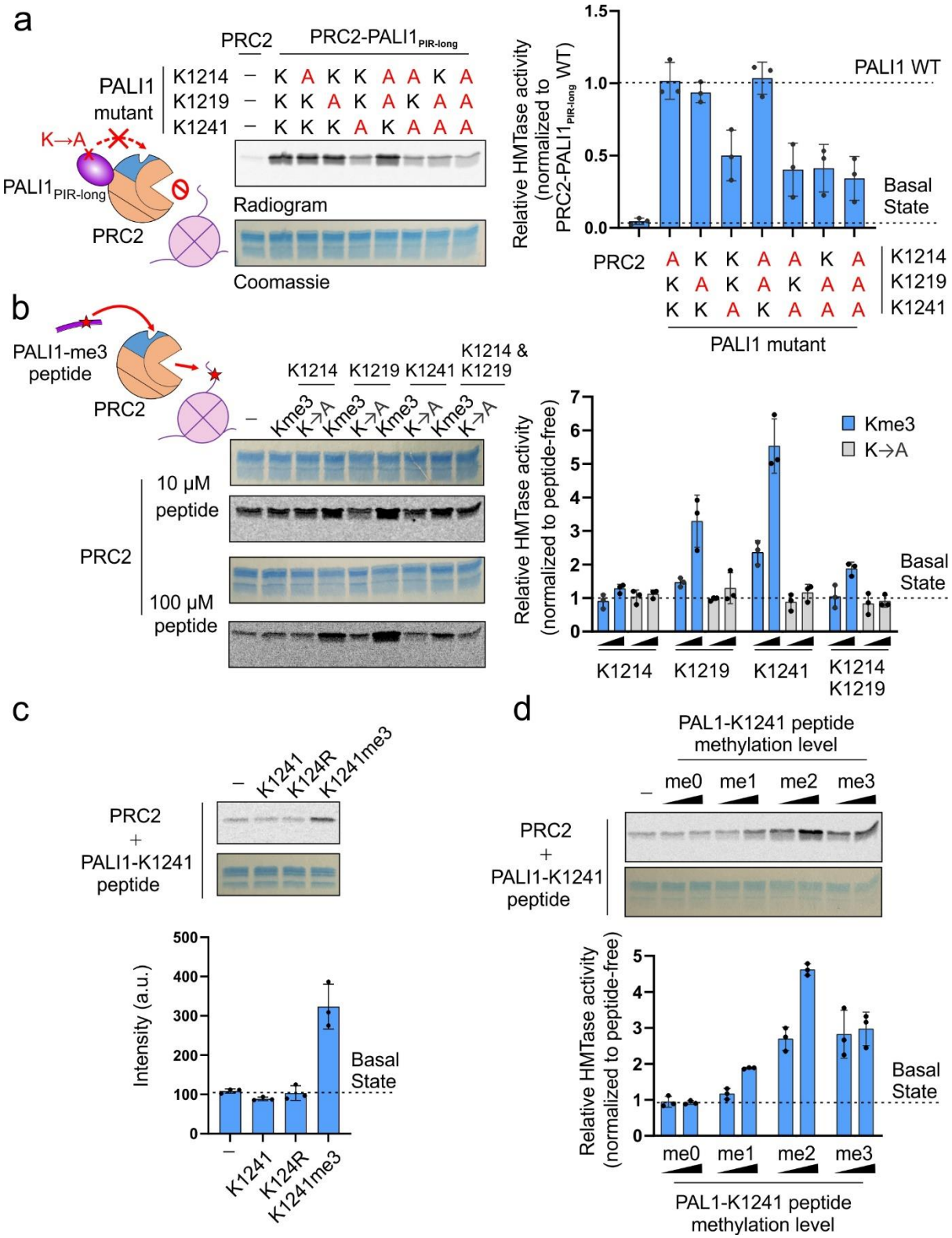
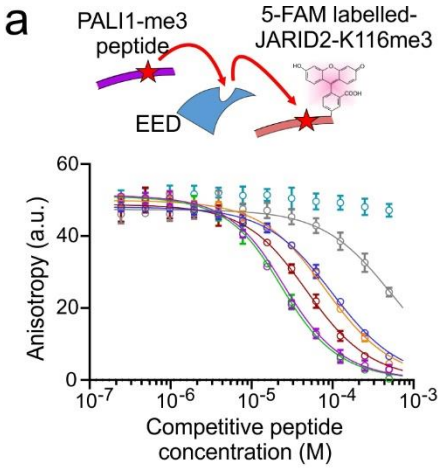


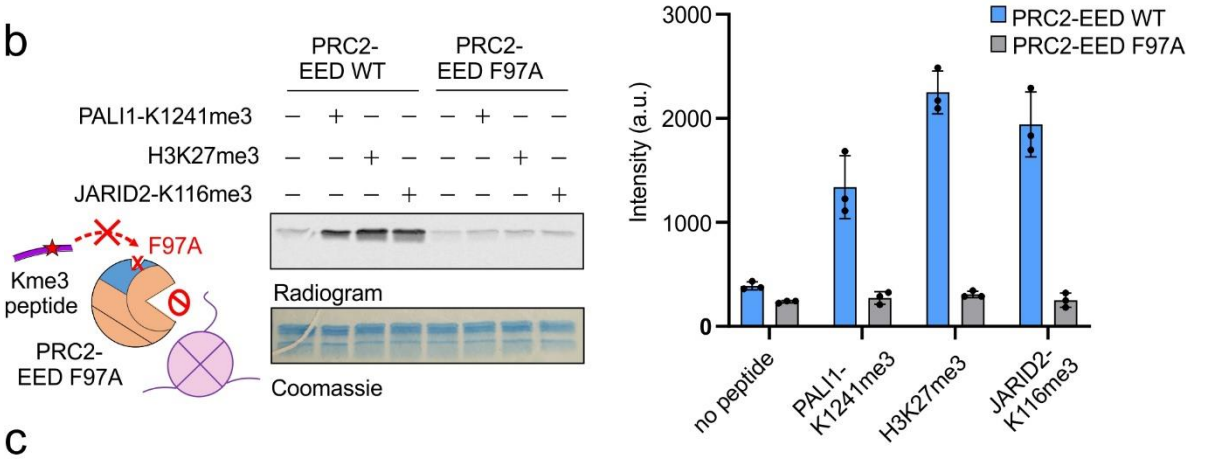
Fig. 2. PALI1 K1241 is required and PALI1-K1241me2/3 is sufficient to stimulate the HMTase activity of PRC2.

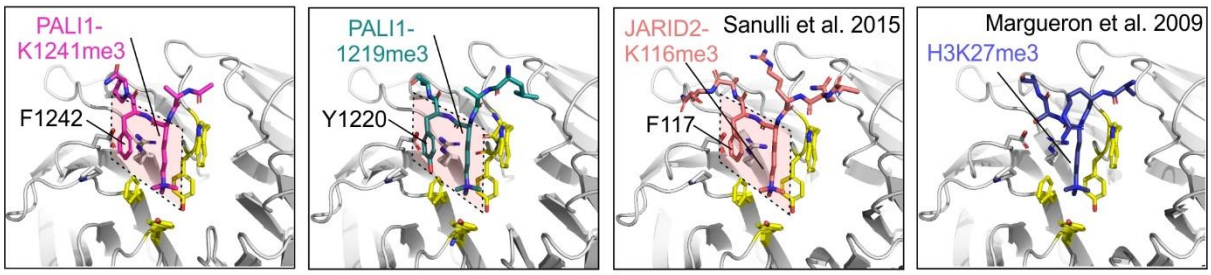
a, HMTase assays were carried out using 500 nM of wild type or mutant recombinant complexes, as indicated, using 2 μM mononucleosomes substrate. The bar plot (right) represents mean HTMase activities, quantified using densitometry and normalized to the activity of the wild-type PRC2-PALI1$_{PIR-long.}$ Dash lines indicate the activity of the wild-type PRC2-PALI1$_{PIR-long}$ (upper line) and the core PRC2 (bottom line) complexes. **b,** HMTase assay performed with 500 nM PRC2, 2 μM

mononucleosomes and in the presence or absence of either 10 µM or 100 µM PALI1 peptide, as indicated. The bar plot (right) represents the relative HMTase activities of PRC2 in the presence of tri-methylated (blue) or K-to-A mutated (grey) PALI1 peptides, as indicated. **c,** HMTase assays of PRC2 performed as above, in the presence or absence of PALI1-K1241 peptides, as indicated. **d,** HMTase assays of PRC2 performed as above, in the presence or absence of PALI1-K1241 peptides with different methylation states, as indicated. The bar plots in (**b**-**d**) represents the relative HMTase activities, normalized to the HMTase activity of PRC2 in its basal state (dashed line). The bar plots in all panels represent the mean of the quantification performed using densitometry over three independent replicates. Error bars shown in this Figure represent standard deviation with the observed values plotted as dots. Uncropped gel images used to generate this figure are in Supplementary Fig. 2.



Key	Peptide	Sequence	K_d (μM)
○	PALI1-K1241me0	KKHLKKFPGATY	n.d.
	PALI1-K1241me1	KKHLKK(me1)FPGATY	240 ± 38
	PALI1-K1241me2	KKHLKK(me2)FPGATY	18.8 ± 2.3
	PALI1-K1241me3	KKHLKK(me3)FPGATY	8.47 ± 0.71
	PALI1-K1214me3	DVPPVK(me3)HPLQKY	33.8 ± 2.6
	PALI1-K1219me3	KHPLQK(me3)YAPSSY	7.53 ± 0.71
	H3K27me3	TKAARK(me3)SAPATY	41.3 ± 2.8
	5-FAM labelled-JARID2-K116me3	RLQAQRK(me3)FAQSQ	8.07 ± 0.49





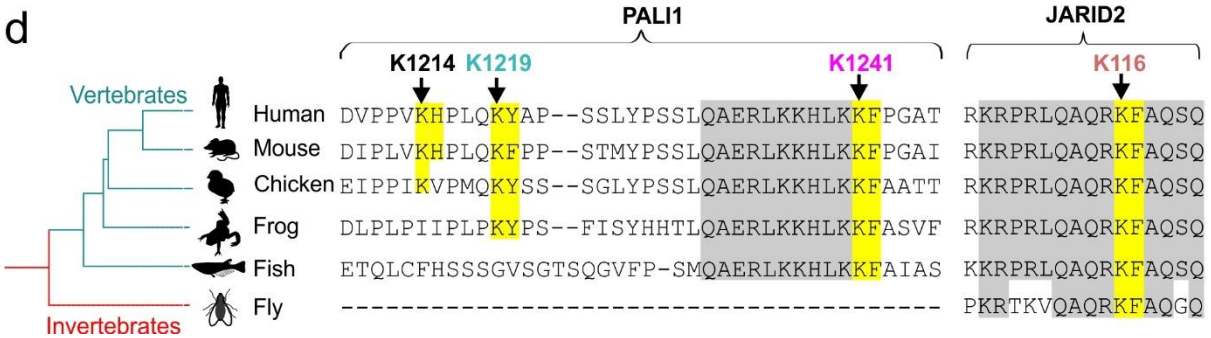
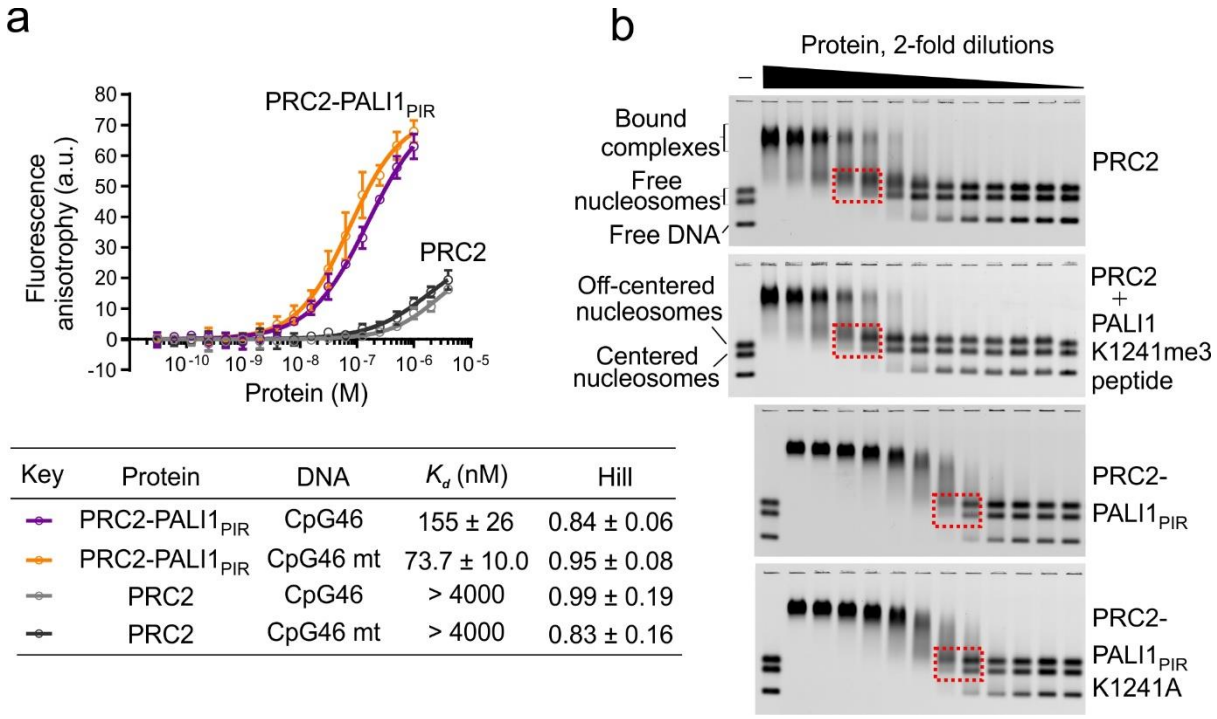


Fig. 3. Structural basis for a convergent evolution between PALI1 to JARID2.

a, Fluorescence anisotropy displacement titrations, where unlabelled peptides competed a 5-FAM labelled JARID2-K116me3 peptide (40 nM) for binding to EED (10 μM). Error bars represent standard deviation over three independent replicates that were carried out on different days. Dissociation constants (K_d) and 95% confidence bounds on the coefficient are indicated in the table. The sequence of the peptides indicated in the table, with the methyl-lysine in red. See Supplementary Fig. 3a for the binding curve of the 5-FAM labelled JARID2-K116me3 to EED. **b,** HMTase assay

performed using wild type (WT) or cage-mutant PRC2 (EED F97A) in the presence or absence of stimulatory peptides, as indicated. The bar plot represents the means of quantification using densitometry done on three independent replicates. Error bars represent standard deviation and the observed values indicated in dots. **c,** High-resolution crystal structures of EED in a complex with either PALI1-K1241me3 (PDB 6V3X; this study), PALI1-K1219me3 (PDB 6V3Y; this study), JARDI2-K116me3 (PDB 4X3E; Sanulli et al 2015) and EED-H3K27me3 (PDB 3IIW; Margueron et al. 2009), as indicated. The methylated lysines and their +1 adjacent conserved aromatic residues are labelled and marked in quadrilaterals. The side chains of the -1 adjacent residues to the tri-methyl-lysines—K1240 in PALI1-K1241me3 and Q1218 in PALI1-K1219me3—could not be traced and are likely disordered. The tri-methyl-lysine peptides are in sticks representation in assorted colours and EED is in grey cartoon representation, with the exception of yellow sticks that represent the aromatic cage amino acids of EED and grey sticks that represent amino acids of EED at the vicinity of the +1 conserved aromatic residue of the peptides. See Supplementary Fig. 3c for the omit electron densities of PALI1-K1241me3 and PALI1-K1219me3 peptides and Supplementary Fig. 3d, e for the sequence alignments and phylogenetic analysis. **d,** Multiple sequence alignments was carried out using T-coffee[55] on entire protein sequences with the relevant regions presented (the following species were used: Homo sapiens, Mus musculus, Gallus gallus, Xenopus laevis, Oryzias latipes and Drosophila melanogaster. Phylogenetic tree was constructed based on NCBI taxonomy and visualized using iTOL[64]. The methylated lysines of JARID2 and PALI1 are labelled with arrows coloured using the same colour code as in panel c, with both the lysines and their adjacent aromatic residues are highlighted in yellow. Other vertebrate-conserved amino acids that are highlighted in grey.



Key	Protein	DNA	K_d (nM)	Hill
	PRC2-PALI1$_{PIR}$	CpG46	155 ± 26	0.84 ± 0.06
	PRC2-PALI1$_{PIR}$	CpG46 mt	73.7 ± 10.0	0.95 ± 0.08
	PRC2	CpG46	> 4000	0.99 ± 0.19
	PRC2	CpG46 mt	> 4000	0.83 ± 0.16

Fig. 4. PALI1 facilitates DNA binding by PRC2.

a, Fluorescence anisotropy used to quantify the affinity of PRC2 complexes to fluorescein-labelled CpG46 or CpG46 mt DNA. Data represent the mean of three independent experiments that were carried out on different days and error bars represent standard deviation. Dissociation constants (K_d) and Hill coefficients are indicated in the table, including their standard error. **b,** EMSA used to quantify the affinity of the indicated PRC2 complexes for a mixture of Cy5-labelled mononucleosomes and free DNA of the same sequence. Dashed boxes indicate the mononucleosome bands near the K_d concentration of the protein, where half of the labelled mononucleosomes are shifted (for quantification, see Supplementary Fig. 4b and c).

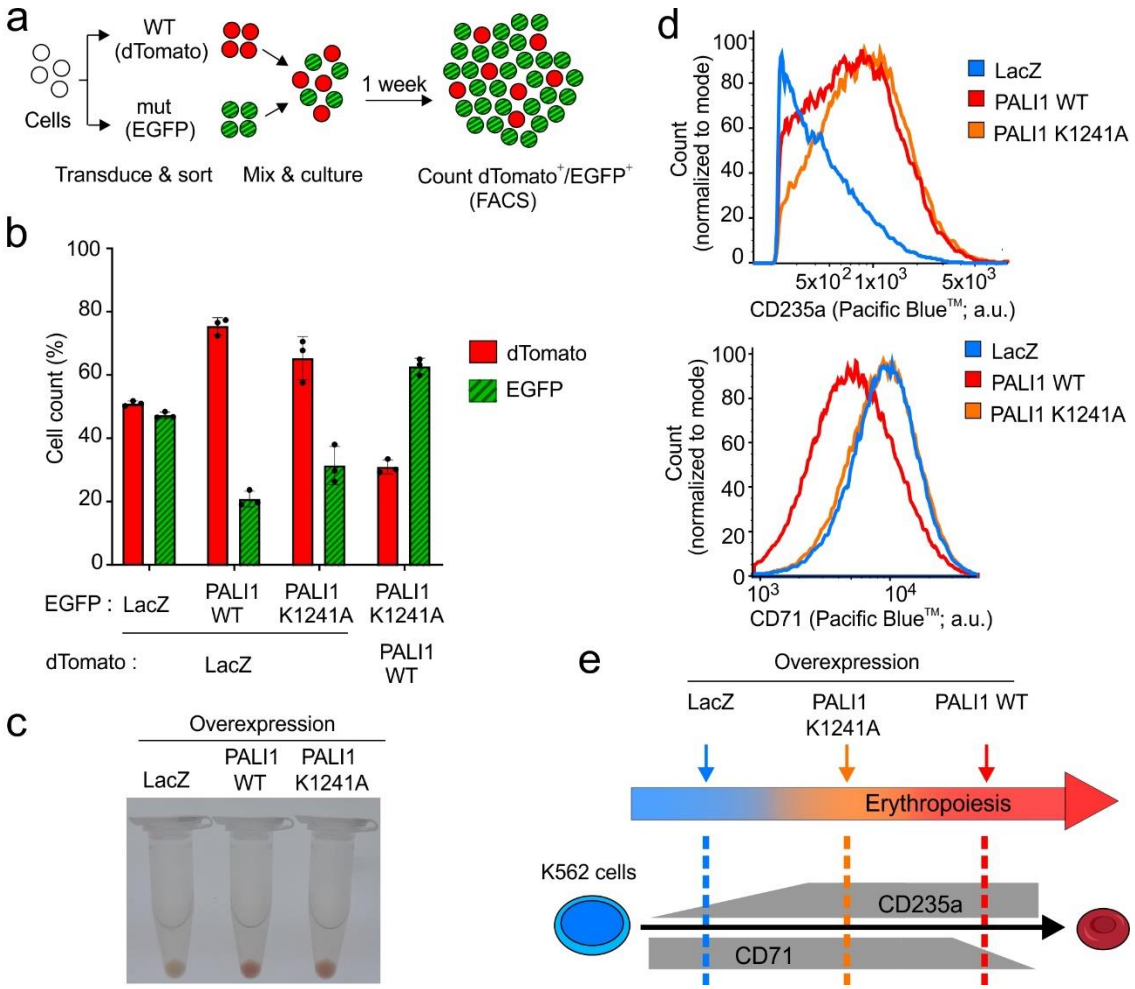


Fig. 5. Overexpression of PALI1 triggers K562 cell differentiation along the erythroid lineage, with the effect alleviated by a separation-of-function allosteric-defective PALI1 mutant (K1241A)

a, Schematic illustration of the competitive cell proliferation assay used to directly compare the proliferation of cells transduced with different constructs. Cells overexpressing different proteins and fluorophores are co-cultured, with the higher-proliferating cells outcompete the lower-proliferating cells. At the end of the experiment, the number of cells carries each vector is quantified using flow cytometry. **b,** Competitive cell proliferation assay of human chronic myeloid leukemia (K562) cells overexpressing either PALI1 wild type (WT), the separation-of-function PALI1 mutant (K1241A) or the LacZ negative control, as indicated. The bar plot represents the mean percentage of each cell population after 7 days of competition, as quantified using flow cytometry. Error bars represent standard deviations derived from three independent experiments carried out on different days with the observed values indicated by dots. Evidence for protein expression and nuclear localization are in Supplementary Fig. 5a. **c,** Colour photographs of pelleted K562 cells overexpressing different proteins: LacZ (left), PALI1 WT (middle) and PALI1-K1241A mutant (right). **d,** Representative histograms generated from flow cytometric analysis of the differentiation markers CD235a (top) and CD71 (bottom) exhibited by K562 cells overexpressing PALI1 WT (red), PALI1 K1241A mutant (orange) or LacZ (blue). The other two replicates shown in Supplementary Fig. 5c, d and evidence for protein expression are in Supplementary Fig 5a. **e,** Illustration of the expected expression level of CD235a and CD71 during erythropoiesis (grey) and the actual observed values represented in dashed lines coloured in red, orange and blue for PALI1 WT, PALI1 K1241A and the negative control LacZ, respectively.

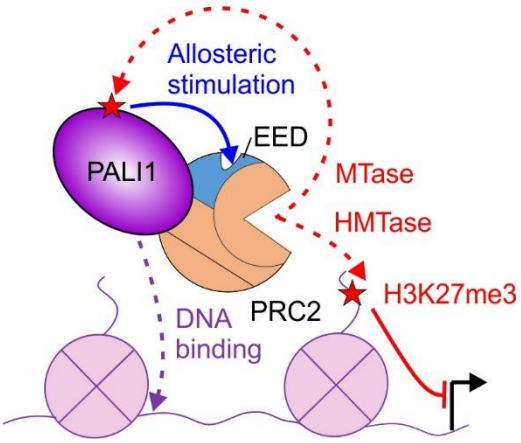


Fig. 6. A model for PALI1-mediated regulation of PRC2: PALI1 regulates PRC2 through two independent mechanisms: (i) PRC2 methylates PALI1 K1241, and possibly K1219, which then binds to the regulatory subunit EED to trigger an allosteric activation of PRC2. (ii) PALI1 facilitates DNA binding and nucleosome substrate binding. PALI1 is in deep purple, EED in light blue and other PRC2 core subunits are in orange. Dashed red arrows represent methylation, with the red stars represent methyl-lysines. The blue arrow represents a positive regulation by an allosteric activation, the dashed purple arrow represents DNA binding and the continues red arrow represents a transcriptional repression.

1028 **Table 1. X-ray crystallography data collection and refinement statistics.**

	PALI1-K1241me3 (PDB:6V3X)	PALI1-K1219me3 (PDB:6V3Y)
Data collection		
Space group	*P* $2_12_12_1$	*P* $2_12_12_1$
Cell dimensions		
a, *b*, *c*(Å)	56.2, 84.7, 90.4	57.8, 85.3, 91.1
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	47.76-1.70	48.79-1.63
R_{sym} or R_{merge}	0.066 (0.631)	0.074 (0.643)
$I/\sigma I$	16.5 (2.8)	16.2 (3.2)
Completeness (%)	97.8 (96.5)	99.7 (97.9)
Redundancy	7.6 (8.0)	8.1 (8.1)
Refinement		
Resolution (Å)	35.2-1.70	48.8-1.63
No. reflections	47039(4553)	56904(5630)
R_{work} / R_{free}	0.174/0.200	0.165/0.193
No. atoms		
Protein (chain A)	2850	2898
Ligand (chain B)	39	42
Water	178	267
B-factors		
Protein	22.9	18.0
Ligand/ion	29.1	24.6
Water	30.1	27.4
R.M.S. deviations		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	0.84	0.89
Ramachandran plot		
Favoured regions (%)	96.6	96.4
Allowed regions (%)	3.4	3.6
Disallowed regions (%)	0.0	0.0

1029 Values in parentheses are for the highest-resolution shell.

1030