Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to ESC genes during differentiation

Gerard L. Brien[1], Guillermo Gambero[3], David J. O'Connell[3], Emilia Jerman[1], Siobhán A. Turner[1], Chris M. Egan[1], Eiseart J. Dunne[1], Maike C. Jurgens[3], Kieran Wynne[3], Lianhua Piao[5], Amanda J. Lohan[3], Neil Ferguson[3], Xiaobing Shi[5], Krishna M. Sinha[4], Brendan J. Loftus[3], Gerard Cagney[3], Adrian P. Bracken[1, 2]

[1]The Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. [2]The Adelaide & Meath Hospital, including the National Children's Hospital, Dublin 16, Ireland. [3]Conway Institute, University College Dublin, Ireland. [4]Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [5]Department of Biochemistry and Molecular Biology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.

Contact:

Corresponding author: Dr. Adrian P. Bracken

e-mail: adrian.bracken@tcd.ie

Tel.: +353 1 8964121

Fax: +353 1 6798558

ABSTRACT (140 words)

Polycomb group proteins are repressive chromatin modifiers with essential roles in metazoan development, cellular differentiation and cell fate maintenance. How Polycomb proteins access active chromatin in order to confer transcriptional silencing during lineage transitions remains unclear. Here we show that the Polycomb Repressive Complex 2 (PRC2) component PHF19 binds the active chromatin mark H3K36me3 via its tudor domain. PHF19 associates with the H3K36me3 demethylase NO66, and is required to recruit the PRC2 complex and NO66 to stem cells genes during differentiation, leading to PRC2 mediated H3K27 tri–methylation, loss of H3K36me3 and transcriptional silencing. We propose a model whereby PHF19 functions during mouse embryonic stem cell differentiation to transiently bind the H3K36me3 mark via its tudor domain, forming essential contact points that allow recruitment of PRC2 and H3K36me3 demethylase activity to active gene loci during their transition to a Polycomb–repressed state.

INTRODUCTION

The Polycomb group proteins are transcriptional repressors that modulate chromatin structure to silence gene expression and are important regulators of cell fate transitions[1,2]. Biochemically, Polycomb group proteins have been characterized into two main complexes, known as Polycomb repressive complex 1 (PRC1) and PRC2 (ref. 3). The PRC2 complex contains three core subunits, EZH2, EED and SUZ12, as well as several sub-stoichiometric components, such as RBBP4, RBBP7, JARID2, AEBP2, PHF1, MTF2 and PHF19 (ref. 4). The exact combinations and stoichiometry of these components within PRC2 are not yet well characterized, although they may be combinatorially assembled in cell type–specific complexes[5]. EZH2 is the catalytically active component of PRC2 and, together with EED and SUZ12, it tri–methylates the N–terminal tail of histone H3 at lysine 27 (H3K27me3)[6,7]. The PRC1 complex is comprised of various combinations of at least five core Polycomb components, including the proteins: PC (CBX2, 4, 6–8); PSC (PCGF1–6); RING (RING1A and RINGB); PH (HPH1–3) and SCML (SCML1–2). The H3K27me3 mark acts as a docking site for the chromodomain of the CBX proteins within canonical PRC1 complexes. However, alternative recruitment pathways must exist for PRC1, since non-canonical complexes, which lack CBX components, have recently been reported[8,9]. PRC1 complexes mono–ubiquitinylate histone H2A at lysine 119 (H2AK119ub) through the enzymatic activity of the RING1A and RING1B subunits, and this is believed to be important for target gene repression, potentially by mediating chromatin compaction[13-16].

The mechanism(s) by which the PRC2 complex is recruited to active genes during lineage specification remain unclear[1,5]. The Drosophila PCL (Polycomb–like) protein is a sub-stoichiometric member of the PRC2 complex and has been reported to be required for the recruitment of the complex to target genes[4,17-20]. The three mammalian orthologues of Drosophila PCL (PHF19, MTF2 and PHF1) also associate with the PRC2 complex and have been implicated in the recruitment and maintenance of the complex on target

3

genes[21-26]. Like their Drosophila homologue, all three mammalian PCL proteins lack any known DNA binding domains, suggesting that their association with chromatin is not mediated by a direct association with DNA. All of these proteins contain a conserved domain structure consisting of a single N–terminal tudor domain and tandem PHD finger domains. These domain structures have previously been shown to bind to, or 'read', post–translationally modified histone tails and have been demonstrated to be functionally important both for protein targeting to chromatin and for transcription regulation[27,28]. This highlights the possibility that the PCL proteins may contribute to PRC2 recruitment through similar histone reading mechanisms.

The chromatin landscape at actively transcribed genes is marked by the histone modifications H3K4me3 and H3K36me3 (ref. 29). The H3K4me3 mark is tightly localized around the transcription start site (TSS), while the H3K36me3 modification is localized downstream, throughout the transcribed gene. Both the H3K4me3 and H3K36me3 histone modifications, when present on the N–terminal tails of both H3 components within a nucleosome, are known to inhibit the enzymatic activity and chromatin binding of the PRC2 complex[30,31]. Therefore, it is unclear how the PRC2 complex initiates repression of hitherto active genes during differentiation. Previous work suggests a potential link between Polycomb repressive mechanisms and histone demethylating activities directed towards the H3K4 and H3K36 residues. For example, Pasini and colleagues reported that the PRC2 complex is associated with the H3K4me3 demethylase RBP2 (also known as KDM5A) in mouse embryonic stem cells, and that this association is important for mediating the removal of H3K4me3 and for silencing of Polycomb target genes[32]. Additionally, the related H3K4me3 demethylase JARID1B (also known as KDM5B) shares a large proportion of PRC2 target genes in mouse embryonic stem cells[33]. Importantly, depletion of JARID1B resulted in higher levels of H3K4me3 at a subset of Polycomb target genes, concomitant with a reduction in PRC2 binding at these sites. Furthermore, in Drosophila, the H3K36me2 demethylase dKDM2 physically associates with a variant PRC1

complex, enhances Polycomb function and is required for target gene silencing[34]. Taken together, these data indicate that Polycomb repressive mechanisms, in part necessitate the specific concomitant modulation of H3K4 and H3K36 methylation status at target genes.

Here we set out to elucidate the role of the PHF19 molecule within the PRC2 complex. We show that a primary function of PHF19 is to recruit the PRC2 complex and the H3K36me3 demethylase NO66 (ref. 35, 36) to target genes. PHF19 achieves this, at least in part, by binding the H3K36me3 mark via its N–terminal tudor domain. While the function of PHF19 is dispensable for embryonic stem cell self–renewal, it potentiates normal embryonic stem cell differentiation, being required for the silencing of embryonic stem cell genes. Taken together, our results predict a model whereby PHF19 binds to an active histone modification to recruit the PRC2 complex and the H3K36me3 demethylase NO66 to embryonic stem cell genes during differentiation.

RESULTS:

PHF19 recruits the lysine demethylase NO66 to chromatin

In order to elucidate the molecular role of PHF19, we designed a screen to identify associated proteins in mammalian cells. We performed affinity purification mass spectrometry (APMS) on FLAG–immunoprecipitations of nuclear extracts prepared from both HEK293 and HMEC cell lines stably expressing a human FLAG–haemagglutinin (HA)–PHF19 fusion protein (Supplementary Fig. 1). As expected, this analysis identified the core components of the PRC2 complex, SUZ12, EED, EZH1 and EZH2 in addition to many additional proteins (Supplementary Fig. 1c and Supplementary Table 1). Interestingly, we also identified a novel interactor, NO66, in both cell types. This protein is a member of the Jumonji C (JmjC) family of demethylases and has been reported to demethylate both di– and tri–methylated histone H3 at lysine 4 (H3K4) and lysine 36 (H3K36)[35,36]. Therefore, we speculated that NO66 is a PHF19 co–repressor protein, which may play a role in Polycomb target gene repression. We subsequently validated that NO66 associates with PHF19 by Western blotting of endogenous NO66 on immunoprecipitations of FLAG–HA–PHF19 in both HEK293 and HMEC cells (Fig. 1a). Next, we also validated the interaction by Western blotting for endogenous PHF19 in a reciprocal purification of human FLAG–NO66, stably expressed in HEK293 cells (Fig. 1b). Finally, we performed immunoprecipitations of NO66, PHF19 and EZH2 in mouse embryonic stem cell nuclear protein lysates and showed that these proteins all interact with each other at the endogenous level, but not with the PRC1 component RING1B, included as a negative control (Fig. 1c).

We next asked if PHF19 is sufficient to recruit the PRC2 complex and NO66 to chromatin. We generated a tetracycline–inducible human GAL4-PHF19 fusion protein in a HEK293 cell line that contained an integrated heterologous GAL4–luciferase reporter construct (Fig. 1d)[38]. Addition of tetracycline to these cells resulted in GAL4–PHF19 protein induction and in efficient repression of luciferase activity. Notably, chromatin immunoprecipitation

(ChIP) analysis of the same cells showed that induction of GAL4–PHF19 binding to the luciferase promoter was sufficient to recruit the PRC2 component EZH2, the PRC1 component CBX8, and to induce a strong enrichment of H3K27me3 (Fig. 1e). Moreover, GAL4–PHF19 induction also led to NO66 recruitment, with a concomitant loss of H3K36me3 throughout the luciferase gene body. We included the *CCNA2* promoter as a negative control. Taken together, these data strongly suggest that PHF19 is sufficient to recruit both the PRC2 complex and NO66 to a specific target gene locus, concomitant with the transition from active expression, associated with high levels of H3K36me3, to gene repression and an associated increase in the levels of H3K27me3.

The PHF19 tudor domain binds H3K27me3 and H3K36me3 *in vitro*

The PHF19 protein contains an N-terminal tudor domain and tandem PHD finger domains (Fig. 2a). Previously, Hunkapiller and colleagues reported that a mouse PHF19 molecule, mutated in the tudor domain, was unable to rescue the loss of H3K27me3 observed upon depletion of endogenous PHF19 in embryonic stem cells, when compared to wild–type PHF19 (ref. 22). The tudor domain is an important 'reading' domain found in several epigenetic proteins[39]. To examine whether the tudor domain of human PHF19 reads modified histone tails, we purified it as a GST–fusion protein and performed *in vitro* peptide pull–down assays on biotinylated histone H3 peptides, modified by tri–methylation of lysine residues at positions 4, 9, 27 and 36 (Fig. 2b). This analysis demonstrated that the PHF19 tudor domain binds directly to the active H3K36me3 mark, in addition to binding the Polycomb–associated H3K27me3 mark, albeit to a lesser extent (Fig. 2b).

PHF19 localises with PRC2 and H3K27me3, not H3K36me3 *in vivo*

In order to investigate the potential *in vivo* relevance of these PHF19 histone binding capabilities, we next determined the genome–wide occupancy of

PHF19 in mouse embryonic stem cells by ChIP followed by DNA sequencing (ChIP–seq). This analysis showed that PHF19 has a highly comparable binding profile to EZH2 and SUZ12 on Polycomb target gene loci (Fig. 2c, Supplementary Fig. 2). Importantly, at a global level, PHF19 shares most of its target genes with SUZ12 and EZH2 (Fig. 2d,e). Several genes enriched for SUZ12 and EZH2 appeared not to have associated PHF19 in our ChIP–seq analysis, suggesting that PHF19 may execute a more restricted function in embryonic stem cells. However, when we subsequently tested several of these genes in quantitative ChIP analysis, we could show that they were in fact bound by PHF19 (Fig. 3a (see PHF19+ve and PHF19–ve) and Supplementary Fig. 3). Therefore, the smaller number of genes bound by PHF19 is likely a consequence of the lower affinity of the PHF19 antibody for its epitope, rather than a reflection of a restricted PRC2 function for PHF19. Consistent with its overlap with Polycomb target genes and our *in vitro* peptide binding experiments, we found that PHF19 target genes were enriched for the H3K27me3 modification (Fig. 2d,e). Importantly, despite the ability of the tudor domain of PHF19 to bind to the H3K36me3 mark, no PHF19 target genes were observed to be marked with H3K36me3 in mouse embryonic stem cells (Fig. 2d,e). This led us to speculate that *in vivo* PHF19 might only transiently interact with the H3 tail when tri–methylated at K36. Therefore, we next wished to explore the possibility that the lack of H3K36me3 on PHF19 target genes could be reconciled with the fact that PHF19 physically associates with an H3K36me3 demethylase.

PHF19 controls PRC2 and NO66 target gene occupancy

To explore the potential functional interplay between PHF19 and NO66, we first decided to determine if both proteins co-occupy the same target genes in mouse embryonic stem cells. We performed ChIP analysis followed by real–time PCR on a cohort of Polycomb target genes and found that PHF19 and NO66 co-occupied 8 of the 10 tested Polycomb target gene promoters (Fig. 3). We also analyzed a cohort of 10 "non–Polycomb" repressed genes in

these ChIP experiments. Consistent with these genes being repressed independently of Polycombs, they lacked both H3K27me3 and H3K4me3, but still retained Histone H3 on their promoters. Importantly, we did not observe PHF19 or NO66 on the promoters of these genes. Taken together, these results suggest that NO66 specifically associates with Polycomb and PHF19 target genes. Interestingly, our observation that NO66 co–occupied bivalent Polycomb target genes, which contained both H3K27me3 and H3K4me3, suggests that NO66 may primarily function as an H3K36me3 demethylase *in vivo*. Furthermore, these data suggest that the co-occupancy of NO66 and PHF19 on target genes is causally linked and not an indirect consequence of a more general role for NO66 as a transcriptional repressor.

We next wished to investigate if PHF19 is required for the association of the PRC2 complex and NO66 to target genes. To do this, we established mouse embryonic stem cell lines that stably expressed either scrambled (SCR) or one of two independent *Phf19*–specific short hairpin RNAs (shRNAs). Western blot and qPCR analysis confirmed efficient and specific inhibition of PHF19 expression (Fig. 4a and Supplementary Fig. 3e). These cells were only marginally affected in terms of their expression of embryonic stem cell genes, such as *Pou5f1* and *Nanog*, growth rate and alkaline phosphatase activity (Supplementary Fig. 3a,b,c). However, we did observe a global depletion of H3K27me3 and a global increase of H3K36me3 in PHF19 depleted cells (Fig. 4a and Supplementary Fig. 4a). We next performed ChIP analysis followed by quantitative real-time PCR of these cells and found that inhibition of PHF19 led to its depletion from the promoters of several Polycomb target genes, including markers of ectodermal (*Fgf5* and *Olig2*), mesodermal (*Gata4* and *Gata6*), and endodermal differentiation (*Pax3* and *Brachyury*) (Fig. 4b). This depletion of PHF19 correlated with a dramatic reduction in the binding of EZH2 and NO66 on the promoters of these genes. However, despite these changes on the chromatin level, we only detected moderate increases in the expression of a subset of these germ cells genes (Supplementary Fig. 3d). This is consistent with previous reports of the

activation of only a sub–cohort of Polycomb target genes upon inhibition of components of the PRC2 complex[40,41]. It has been speculated that the DNA binding, lineage specific transcription factors that normally activate this cohort of genes are absent in the undifferentiated cells[40]. Taken together, these results strongly suggest that PHF19, while not required for embryonic stem cell proliferation, is required for the maintenance of PRC2 and NO66 on the promoters of differentiation genes in embryonic stem cells.

In order to rule out the possibility that NO66 is required for targeting PHF19 and the PRC2 complex to chromatin, we established mouse embryonic stem cell lines that stably expressed either scrambled (SCR) or one of two independent *No66*–specific short hairpin RNAs (shRNAs). Western blot and quantitative RT-PCR analyses confirmed efficient inhibition of NO66 expression (Supplementary Fig. 4a,b). As expected, the depletion of NO66 led to loss of the protein on the promoters of Polycomb target genes, (Supplementary Fig. 4c). However, this loss did not lead to the displacement of PHF19 or the PRC2 complex. These NO66 depleted cells had no detectable proliferation or pluripotency defects and did not significantly activate the expression of Polycomb target genes (Supplementary Fig. 4d,e and data not shown). Taken together, our results show that NO66 localization is dependent on PHF19, but not the other way round and support a model in which PHF19 is targeted to chromatin independently of NO66.

PHF19 is necessary for *de novo* recruitment of NO66 and PRC2

To assess whether PHF19 is required for the *de novo* recruitment of the PRC2 complex and NO66 to target genes during embryonic stem cell differentiation, we generated mouse embryonic stem cells, expressing either scrambled (SCR) or *Phf19*–specific shRNAs, and differentiated them to embryoid bodies. Inhibition of PHF19 dramatically affected the ability of the embryonic stem cells to form large embryoid bodies with differentiation features (Fig. 5a, Supplementary Fig. 5a). Consistent with this, PHF19 inhibition led to impaired silencing of embryonic stem cell genes (Fig. 5b,

Supplementary Fig. 5b) and to reduced activation of differentiation genes (Supplementary Fig. 5d). We also observed that PHF19 inhibition led to a defect in silencing of embryonic stem cell genes when the cells were differentiated with retinoic acid addition for 2 days under adherent conditions (Supplementary Fig. 6). ChIP experiments of embryoid bodies, 8 days after they were induced to differentiate demonstrated that inhibition of PHF19 resulted in a decreased recruitment of the PHF19 protein to both the promoters and intragenic regions of the *Fgf4*, *Pou2f3* and *Fgf17* genes (Fig. 5c and Supplementary Fig. 5c). Importantly, PHF19 inhibition also led to impaired recruitment of the PRC2 complex and NO66 to these regions, concomitant with reduced accumulation of H3K27me3 and failure to remove the H3K36me3 mark. Conversely, mouse embryonic stem cells, expressing *No66*–specific shRNAs, when differentiated to embryoid bodies, were fully capable of repressing embryonic stem cell expressed genes (Supplementary Fig. 4f,g). Our observation that depletion of PHF19 led to a greater increase in H3K36me3 levels than depletion of NO66, suggests that PHF19 is required for the function of additional H3K36me3 demethylases (Supplementary Fig. 4a). Taken together, these results show that PHF19 is necessary for the *de novo* targeting of the PRC2 complex, as well as NO66, during embryonic stem cell differentiation.

A PHF19 W50C Y56A mutant cannot bind H3K27me3 or H3K36me3

We next wished to determine if the PHF19 tudor domain is required for targeting the PRC2 complex and NO66 to chromatin. Previously, Hunkapiller and colleagues reported that the mouse PHF19 molecule when mutated at positions W48 and Y54, within the tudor domain, was unable to rescue the loss of H3K27me3 observed upon depletion of endogenous PHF19 (ref. 22). We therefore generated the corresponding human PHF19 mutant tudor domain [W50C Y56A (corresponding to mouse W48 and Y54, respectively)], purified it as a GST–fusion protein and performed *in vitro* peptide pull–down assays on biotinylated histone H3 peptides, modified by mono–, di– and tri–

methylation of lysine residues at positions 27 and 36 (Fig. 6a). This revealed that the mutant tudor domain was incapable of binding the H3K27me3 and H3K36me3 modifications (Fig. 6a and Supplementary Fig. 7a). We then performed independent surface plasmon resonance (SPR) to accurately measure the *in vitro* binding properties of the wild–type and mutant GST–PHF19 tudor domain to the H3K36me3 peptide (Fig. 6b). We found the binding of GST–PHF19 to H3K36me3 was highly reproducible with an apparent K_d of ~188 nM, as determined from independent measurements at 5 concentrations of protein. Consistent with our peptide pull–down experiment, we observed negligible binding of GST–PHF19 mutant tudor domain to the H3K36me3 peptide.

PHF19 tudor domain is required for PRC2 and NO66 recruitment

To address whether recognition of H3K36me3 is important for PHF19 function *in vivo*, we performed rescue experiments in parallel, using lentiviral expression vectors, either expressing wild–type human PHF19 or human PHF19 containing the two point mutations (W50C Y56A), which result in loss of H3K36me3 recognition. The reintroduction of the wild–type PHF19 restored genome–wide levels of H3K27me3, as previously shown, as well as the association of NO66 and the PRC2 complex to target genes in embryonic stem cells (Fig. 6c and Supplementary Fig. 7b). It also rescued the defects in embryoid body differentiation (Fig. 6d) and restored the association of PRC2 and NO66 to hitherto active genes during embryonic stem cell differentiation (Fig. 6e and Supplementary Fig. 7c). In contrast, the PHF19 mutant (W50C Y56A) was incapable of associating with Polycomb target genes in embryonic stem cells (Fig. 6c), did not rescue the embryoid body differentiation defects (Fig. 6d) and was not recruited to embryonic stem cell genes during differentiation, leading to a failure to recruit both the PRC2 complex and NO66, and correlated with impaired gene silencing (Fig. 6e and Supplementary Fig. 7c,d).

Taken together, these results demonstrate that the ability of the PHF19 tudor domain to recognise the H3K27me3 and H3K36me3 modifications is essential for its function *in vivo*. We propose a model in which PHF19 'reads' the H3K36me3 mark at embryonic stem cells genes and recruits the core PRC2 complex, leading to H3K27me3 accumulation and gene repression during differentiation. The presence of H3K27me3 on these genes may contribute to stabilizing the association of PHF19 via its tudor domain. Furthermore, the demethylating activity of NO66 may contribute to stabilizing PRC2 activity on chromatin by preventing accumulation of the antagonistic H3K36me3 modification (Fig. 7).

DISCUSSION

A key unanswered question has been how the PRC2 complex is targeted to regions of active chromatin, in order to facilitate subsequent gene repression during lineage transitions. Our demonstration that PHF19 confers an H3K36me3 reading capability to PRC2 reveals a new aspect to the transition from an active gene to a Polycomb–repressed gene. Furthermore, we provide the first evidence revealing that an H3K36me3 demethylase activity is associated with the PRC2 complex.

The H3K36me3 and H3K27me3 histone modifications have never been observed to co–localize on chromatin in ChIP–seq experiments[29,42,43]. Furthermore, biochemical analyses of histone H3 modifications suggest that both marks rarely, if ever, co-exist on the same histone H3 tail[44]. Our demonstration that the H3K36me3 demethylase NO66 associates with PHF19 may explain this, at least in part. Tellingly, we found that depletion of PHF19 led to a much greater increase in global H3K36me3 levels, when compared to depletion of NO66 (Supplementary Fig. 4a). This could suggest that additional H3K36me3 demethylases are dependent on PHF19 and the PRC2 complex for their activity and (or) association on target sites. In addition to NO66, there are at least four other H3K36me3 demethylases known[35], yet we only found NO66 to be associated with the PHF19–PRC2 complex. Therefore, an alternative possibility is that the presence of the PHF19-PRC2 complex on chromatin is refractory to H3K36me3 methyltransferase binding and (or) enzymatic activity. However, previous observations have implicated other H3K36me3 demethylases in Polycomb gene repression. For example, the JMJD2A (also known as KDM2A) H3K36me3 demethylase is recruited to unmethylated CpG islands, which are enriched in Polycomb target genes, via its zinc finger CxxC (ZF–CxxC) domain[45,43]. This suggests that this particular H3K36me3 demethylase may be recruited to Polycomb target genes to collaborate in gene silencing, and it may be indicative of a more general link between Polycomb repressive mechanisms and H3K36me3 demethylase

activities. It will be important to determine if the other two mammalian PCL proteins, MTF2 and PHF1, associate with NO66 or indeed any other H3K36me3 demethylases.

Our results also have implications for the roles of the PCL proteins in early mammalian development. The phenotypes observed upon knockdown of PHF19 are similar to those previously observed upon loss of other PRC2 components and associated proteins. For example, the observation that PHF19 inhibition does not affect embryonic stem cell proliferation, but does reduce the association of Polycomb proteins to the promoters of repressed differentiation genes, is in keeping with previous observations on SUZ12 (ref. 41), EZH2 (ref. 46), EED[47] and JARID2 (ref. 48–50). Similarly, like PHF19, these proteins have been shown (by these and other studies) to be required during embryonic stem cell differentiation for the silencing of genes previously expressed in undifferentiated embryonic stem cells[4]. However, for the first time, our work now offers molecular insights into how the PRC2 complex is targeted to these genes and thereby initiates subsequent repression.

Our results also raise some intriguing new questions. For instance, PHF19 is expressed both in embryonic stem cells and during their differentiation towards embryoid bodies. Thus, while we have established that PHF19 functions in the early stages of the transition from active genes to PRC2 repressed genes, the actual trigger that initiates the *de novo* recruitment of PHF19 remains unknown. For instance, the H3K36me3 mark is present on the loci of these genes in embryonic stem cells, yet PHF19 is not recruited until these cells are induced to differentiate. Thus, additional changes to the local chromatin environment, such as deacetylation, may also be required. Alternatively, nucleosome compaction, lineage specific DNA binding transcription factors and ncRNAs may also play a role[1]. Furthermore, it is unclear how the demethylase activity of NO66 is sterically compatible with the simultaneous binding of the PHF19 tudor domain. It is possible that NO66 is not required for the removal of the H3K36me3 mark during initial gene silencing. Instead, other demethylase complexes might act in parallel with

PRC2 leading to demethylation. Therefore, the function of NO66 might be to prevent aberrant H3K36me3 accumulation on Polycomb target genes. Answering these questions will undoubtedly shed further light on the molecular mechanisms at the heart of not only embryonic stem cell differentiation, but also lineage changes in general.

ACCESSION CODES

PHF19 and control (whole genomic DNA) ChIP-seq data are available through the Gene Expression Omnibus (GEO) under accession code GSE41609.

K.M.S. contributed the NO66 antibody and advised on its use. G.L.B. and A.P.B. wrote the manuscript.

FIGURE and TABLE LEGENDS

Figure 1. **PHF19 recruits the H3K36me3 demethylase NO66 to chromatin to repress gene transcription.**

(a) Western blots of immunoprecipitates (IPs), showing that endogenous PRC2 components and NO66, but not the PRC1 component CBX8, associate with FLAG–HA–PHF19 in HEK293 and HMEC cells. **(b)** Western blots of immunoprecipitates showing that endogenous PHF19 and EZH2, but not CBX8, associate with FLAG–NO66 in HEK293 cells. **(c)** Immunoprecipitations of the indicated endogenous proteins show NO66, PHF19 and EZH2 co–associate in mouse embryonic stem cell nuclear protein lysates. **(d)** Left panel: Schematic representation of the GAL4-TKLuc reporter system, Middle panel: Western blot analysis of GAL4-PHF19 TKLuc cells, grown in the absence (–) or presence (+) of tetracycline for 48 h showing induction of the fusion protein. Right panel: GAL4–PHF19 TKLuc cells, but not control parental cells, show repression of luciferase activity upon addition of tetracycline for 48 h. **(e)** ChIP analyses using the indicated antibodies on GAL4–PHF19 TKLuc cells grown in the absence or presence (+) of tetracycline for 48 h. Precipitated DNA was analysed by quantitative RT–PCR with primers corresponding to the locations indicated in panel (**d**) and the *CCNA2* gene promoter, included as a negative control. ChIP enrichments are presented as percentage bound, normalized to input.

Figure 2. The PHF19 tudor domain binds to H3K36me3, but PHF19 co–localizes with PRC2 and H3K27me3, not H3K36me3, in mouse embryonic stem cells.

(a) Schematic representation of the domain structure of the human PHF19 protein **(b)** Western blot of an *in vitro* peptide pull-down assay performed using GST–PHF19$_{25-95}$ containing the tudor domain shows enrichment on the H3K27me3 and H3K36me3 residues. **(c)** Representative examples of PHF19 ChIP–seq results for the three Polycomb target genes and germ layer markers, *Gata6* (endoderm), *Brachyury* (mesoderm) and *Fgf5* (primitive ectoderm). The *Ccna2* gene is shown as a negative control. Binding profiles of ChIP–seq results for EZH2, SUZ12, H3K27me3, H3K36me3 and H3K4me3 obtained from previous studies are indicated for comparison[42,48,51]. **(d)** Venn diagram analysis showing overlaps of genes associated with PHF19, PRC2 component (SUZ12 and EZH2) and H3K27me3, but not H3K36me3, in mouse embryonic stem cells. **(e)** Heat maps depicting the binding patterns of PHF19, PRC2 components (EZH2 and SUZ12) and histone modifications (H3K27me3 and H3K36me3) within −5 and +5 Kb of the transcription start sites (TSS) of both 'active' and 'bivalent' genes in mouse embryonic stem cells. 'Bivalent' genes were ranked on their H3K27me3 read density, and 'active' genes were ranked on H3K4me3 read density around the TSS.

Figure 3: NO66 binds to Polycomb repressed genes in mouse embryonic stem cells

ChIP analyses of PHF19, NO66, H3K27me3, H3K4me3, Histone H3 and HA (negative control) in mouse embryonic stem cells. 'PHF19+ve' indicates genes identified as PHF19 targets in our genome–wide ChIP–seq analysis. 'PHF19–ve' indicates previously reported Polycomb target genes that were not identified as PHF19 targets in our ChIP–seq experiment. The precipitated DNA was analyzed by quantitative RT–PCR with primers corresponding to the promoter regions of the indicated actively expressed, Polycomb repressed, and non–Polycomb repressed genes and presented as percentage bound, normalized to input.

Figure 4. PHF19 controls PRC2 and NO66 target occupancy in mouse embryonic stem cells. (a) Western blot analysis of mouse embryonic stem cells expressing scrambled (shSCR), or one of two *PHF19*-specific (shPHF19 1. and 2.) shRNAs shows efficient depletion of PHF19, which correlates with a global decrease of H3K27me3 and a global increase of H3K36me3. **(b)** ChIP analysis of PHF19, NO66, EZH2 and H3K27me3 on promoters of lineage genes in cells from (a). ChIPs are presented as percentage bound, normalized to input. The *Ccna2* gene promoter is represented as a negative control.

Figure 5. PHF19 is required for PRC2 and NO66 recruitment and embryonic stem cell differentiation (a) Phase contrast microscope image of embryoid bodies formed after 8 days non–directed differentiation of embryonic stem cells, expressing either non-specific scrambled (shSCR) or PHF19 specific (shPHF19.1 and shPHF19.2) shRNAs reveals defects in differentiation in the absence of PHF19. **(b)** Defects in embryonic stem cell gene repression during differentiation are observed in PHF19 depleted cells. Presented are qPCR expression analysis of the indicated embryonic stem cell expressed genes in undifferentiated shSCR, shPHF19.1 or shPHF19.2 mouse embryonic stem cells (indicated with a "0"), or the same cells induced to differentiate towards embryoid bodies for 4 and 8 days (top panel). The relative fold differences in expression of each gene in shSCR, shPHF19.1 and shPHF19.2 cells at day 8 of embryoid body formation is indicated (bottom panel). **(c)** Dramatic chromatin defects are revealed in the absence of PHF19. Presented are ChIP analyses of PHF19, NO66, EZH2, H3K27me3 and H3K36me3 in undifferentiated shSCR or shPHF19 mouse embryonic stem cells (indicated with a "0"), or the same cells induced to differentiate towards embryoid bodies for 8 days. Precipitated DNA was analysed by quantitative RT–PCR with primers corresponding to the promoter (labelled 1 in the top panel) or intragenic (2 in the top panel) regions of the embryonic stem cell expressed genes *Fgf4* and *Pou2f3*, and presented as percentage bound, normalized to input.

Figure 6. A functional PHF19 tudor domain is required for the *de novo* recruitment of PRC2 and NO66 during embryonic stem cell differentiation. (a) Western blots of *in vitro* peptide pull–down assays using wild–type (WT) GST–PHF19$_{25-95}$, a mutant (W50C Y56A) GST–PHF19$_{25-95}$ or GST–only. **(b)** Representative SPR sensorgrams for WT–PHF19$_{25-95}$ (Top Panel), and WT and mutant–GST–PHF19$_{25-95}$ (Bottom Panel) and H3K36me3. Affinity rate constants were determined from a concentration series of WT–PHF19 binding to H3K36me3. **(c)** Rescue of PRC2 and NO66 recruitment to Polycomb target genes with WT, but not mutant PHF19. Presented are ChIP analyses of PHF19, NO66, EZH2 and H3K27me3 in mouse embryonic stem cells expressing shSCR or shPHF19 alone or in combination with WT (shPHF19+WT–hPHF19) or mutant (shPHF19+W50C Y56A–hPHF19) human PHF19. Precipitated DNA was analysed by quantitative RT–PCR with primers used in Fig. 4b and presented as percentage bound, normalized to input. The *Ccna2* gene promoter is presented as a negative control. **(d)** Rescue of embryoid body differentiation defects with WT but not mutant PHF19. Presented are phase contrast images of embryoid bodies formed after 8 days differentiation of cells in (c). **(e)** Rescue of PRC2 and NO66 recruitment to embryonic stem cell genes during differentiation with WT, but not mutant PHF19. Presented are ChIP analyses of PHF19, NO66, EZH2, H3K27me3 and H3K36me3 in cells from (c) (indicated with a "0"), or the same cells induced to differentiate towards embryoid bodies for 8 days. Precipitated DNA was analysed by quantitative

RT–PCR with the primers used in Figure 5c and presented as percentage bound, normalized to input.

Figure 7. Model for Polycomb PHF19 function in ES cells and during differentiation. (a) In undifferentiated embryonic stem cells, PHF19 and NO66 are co–bound on Polycomb repressed genes such as the germ cell markers, *Olig2* and *Gata4*. During differentiation PHF19 and other PRC2 components are displaced from differentiation genes as they become activated. **(b)** During differentiation of embryonic stem cells, PHF19 binds to H3K36me3–modified histones at active embryonic stem cell genes, such as *Fgf4* and *Pou2f3*. This binding results in the recruitment of the PRC2 complex and the H3K36me3 demethylase NO66 to hitherto active genes, resulting in a switch from K36me3 to K27me3 and consequent transcriptional repression.

REFERENCES

1.   Bracken, A.P. & Helin, K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat Rev Cancer* **9**, 773-84 (2009).
2.   Pietersen, A.M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr Opin Cell Biol* **20**, 201-7 (2008).
3.   Simon, J.A. & Kingston, R.E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).
4.   Sauvageau, M. & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **7**, 299-313 (2010).
5.   Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-9 (2011).
6.   Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-43 (2002).
7.   Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-905 (2002).
8.   Gao, Z. et al. PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol Cell* **45**, 344-56 (2012).
9.   Tavares, L. et al. RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* **148**, 664-78 (2012).
10.   O'Loghlen, A. et al. MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. *Cell Stem Cell* **10**, 33-46 (2012).
11.   Morey, L. et al. Nonoverlapping functions of the polycomb group cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62 (2012).
12.   Maertens, G.N. et al. Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. *PLoS One* **4**, e6380 (2009).
13.   Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-8 (2004).
14.   Francis, N.J., Kingston, R.E. & Woodcock, C.L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-7 (2004).
15.   de Napoles, M. et al. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**, 663-76 (2004).
16.   Shao, Z. et al. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46 (1999).
17.   Nekrasov, M. et al. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J* **26**, 4078-88 (2007).

18. O'Connell, S. et al. Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. *J Biol Chem* **276**, 43065-73 (2001).

19. Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A. & Harte, P.J. A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3. *Mol Cell Biol* **23**, 3352-62 (2003).

20. Savla, U., Benes, J., Zhang, J. & Jones, R.S. Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development* **135**, 813-7 (2008).

21. Boulay, G., Rosnoblet, C., Guerardel, C., Angrand, P.O. & Leprince, D. Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *Biochem J* **434**, 333-42 (2011).

22. Hunkapiller, J. et al. Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet* **8**, e1002576 (2012).

23. Walker, E., Manias, J.L., Chang, W.Y. & Stanford, W.L. PCL2 modulates gene regulatory networks controlling self-renewal and commitment in embryonic stem cells. *Cell Cycle* **10**, 45-51 (2011).

24. Li, X. et al. Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes. *Mol Cell Biol* **31**, 351-64 (2011).

25. Casanova, M. et al. Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development* **138**, 1471-82 (2011).

26. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28**, 2718-31 (2008).

27. Chan, D.W. et al. Unbiased proteomic screen for binding proteins to modified lysines on histone H3. *Proteomics* **9**, 2343-54 (2009).

28. Wysocka, J. et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).

29. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).

30. Schmitges, F.W. et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-41 (2011).

31. Voigt, P. et al. Asymmetrically modified nucleosomes. *Cell* **151**, 181-93 (2012).

32. Pasini, D. et al. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* **22**, 1345-55 (2008).

33. Schmitz, S.U. et al. Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J* **30**, 4586-600 (2011).

34. Lagarou, A. et al. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* **22**, 2799-810 (2008).

35. Kooistra, S.M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* **13**, 297-311 (2012).

36. Sinha, K.M., Yasuda, H., Coombes, M.M., Dent, S.Y. & de Crombrugghe, B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J* **29**, 68-79 (2010).

37. Mousavi, K., Zare, H., Wang, A.H. & Sartorelli, V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol Cell* **45**, 255-62 (2012).

38. Hansen, K.H. et al. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* **10**, 1291-300 (2008).

39. Maurer-Stroh, S. et al. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**, 69-74 (2003).

40. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-36 (2006).

41. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. & Helin, K. The Polycomb Group protein Suz12 is required for Embryonic Stem Cell differentiation. *Mol Cell Biol* (2007).

42. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).

43. Zhou, V.W., Goren, A. & Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18 (2011).

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

45. Blackledge, N.P. et al. CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* **38**, 179-90 (2010).

46. Shen, X. et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502 (2008).

47. Montgomery, N.D. et al. The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* **15**, 942-7 (2005).

48. Peng, J.C. et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290-302 (2009).

49. Pasini, D. et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-10 (2010).

50. Landeira, D. et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* **12**, 618-24 (2010).

51. Ku, M. et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).

MATERIALS AND METHODS:

Cell culture

HEK293T and Flp-In T-REx 293 cells (Invitrogen) were grown in DMEM media supplemented with 10% FBS (Hyclone), 100 U ml^{-1} penicillin and 100 U ml^{-1} streptomycin (Gibco). The stable GAL4-PHF19 Flp-In T-Rex cell line was generated from a previously described Flp-In T-Rex cell line[38]. GAL4-PHF19 protein expression was induced by addition of tetracycline (1 μg ml^{-1}) to the culture media for 48 h. HMEC cells were grown as previously described[52]. Embryonic stem cells were grown on gelatinized culture dishes in GMEM media (Sigma) supplemented with 10% FBS (Gibco) (v/v), 1000 U ml^{-1} leukemia inhibitory factor (Millipore), 100 U ml^{-1} penicillin, 100 U ml^{-1} streptomycin (Gibco), 1:100 GlutaMAX (Gibco), 1:100 non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco) and 50 μM β-mercaptoethanol (Sigma).

Cloning, plasmid generation and RNA interference

Human PHF19 was PCR amplified from low passage HMEC cDNA, inserted into the pCR8⁄GW⁄TOPO Gateway cloning entry vector (Invitrogen) following the protocol of the manufacturer and sequence verified. Sequence verified PHF19 was sub-cloned into Gateway cloning compatible expression vectors by recombination using the LR clonase enzyme (Invitrogen). Double point mutant forms of PHF19 were generated using the GeneArt® Site–Directed Mutagenesis System (Invitrogen) in accordance with the manufacturers instruction. Stable shRNA pLKO.1 vectors expressing control (Scrambled), PHF19–specific (Catalog number – TRCN0000096072 and TRCN0000096073) and NO66 specific shRNA sequences (Catalog number TRCN0000376816 and TRCN0000375742) were purchased from Sigma.

Luciferase assays

Luciferase assays were performed by lysing cells using 1x lysis buffer (Luciferase assay system – Promega). Cell lysates were pre–cleared by centrifugation and 10 µl of the lysate processed for luciferase measurements

in accordance with the manufacturers instructions. Overall protein measurements were performed using Bradford reagent (BioRad) and luminometer readings were normalized to protein content and expressed as arbitrary units.

Embryonic stem cell differentiation and viral transduction

Embryoid body differentiation of ES cells was induced by plating 1×10^5 embryonic stem cells ml^{-1} in non-adherent bacterial dishes (Greiner) in embryonic stem cell media without LIF. The media was changed every 2 to 3 days throughout the differentiation procedure. For embryonic stem cell differentiation with *all-trans* retinoic acid (ATRA), cells were cultured in media containing 1 μM of RA without LIF for 48 h, the media was replaced after 24 h. Retroviral and lentiviral particles were produced and used to transduce target cells as previously described[32,53].

Generation of PHF19 polyclonal antibodies

Polyclonal antibodies were generated by immunizing rabbits with 2 peptides corresponding to amino acids 362–376 (NSASSELRKRGKSKP +C) and 503–518 (SAEGASVPERPDEGID) of full–length human PHF19. These peptides were conjugated C terminally to KLH and subcutaneously injected into rabbits in accordance with standard procedures (Eurogentec). Whole rabbit serum was subsequently affinity purified on both peptide antigens.

Purification of recombinant GST-fusion protein

The PHF19–tudor (a.a. 25–95) fragment was cloned into pGEX–6P1 and transformed into the *E. coli* strain BL21–DE3. Protein expression was induced with 0.5 mM IPTG, and GST–PHF19 fusion proteins were purified over GSH–agarose beads (Pierce).

***In vitro* peptide pulldown**

Biotinylated histone H3 peptides (1μg) either unmodified or modified at the indicated lysine residues were incubated with 5 μg of GST–PHF19 (a.a. 25–95) in binding buffer (50 mM Tris·HCl pH 7.5, 300 mM NaCl, 20% glycerol, 0.1% NP40, 1 mM PMSF) overnight at 4°C. Strepavidin agarose beads

(Invitrogen) were added, and samples were incubated for a further 60 mins at 4°C. Beads were washed extensively in binding buffer, and bound protein was eluted using 2× Laemilli dye. Eluted protein was analysed by Western blot.

Surface plasmon resonance (SPR)

SPR experiments were performed at 25°C using a series S sensor chip SA with a BiaCore T200 surface plasmon resonance instrument (GE Healthcare). All experiments were performed with in HBS–P running buffer [10mM Hepes, 150mM NaCl, 0.05% surfactant P20]. Biotinylated H3, H3K4me3, H3K27me3 and H3K36me3 peptides were diluted in running buffer and then immobilised to a density of 600–770 RU. GST–PHF19 fusion-protein (at concentrations from 50 to 800nM) was injected onto the chip surface for 180 s at a flow rate of 20 μl min^{-1}. The dissociation phase was monitored for up to 600 s. Individual sensorgrams were double–referenced against injection onto an empty flow cell and GST alone injections at equivalent concentrations. Data were fitted to a 1:1 Langmuir model using Biaevaluate analysis software. The observed results and apparent K_d values were highly reproducible in replicate experiments.

ChIP and ChIP–Sequencing

ChIP analyses were performed as described previously[40]. ChIP against Haemaglutinin (HA tag) served as a negative control in all ChIP–qPCR experiments. For ChIP–seq, DNA from 10 independent PHF19 ChIP experiments was pooled, and 100 ng was prepared for sequencing using the ChIP–seq sample prep kit (Illumina). ChIP–seq library DNA was sequenced using the Genome Analyser II (Illumina). Base calling and mapping of the 42–bp sequence reads to the mouse genome (mm8, Feb 2006 release) was done using the Bowtie alignment tool, allowing for up to 2 mismatches in each read. Peak detection was performed using MACs[54], and input DNA was used as a control for normalization.

Quantitative real-time PCR

Total RNA was extracted from cells using the RNeasy kit (Qiagen) according to manufacturer's protocol. RNA was used to generate cDNA by reverse transcriptase PCR using the TaqMan Reverse Transcription kit (Applied Biosytems). Relative mRNA expression levels were determined using the SYBR Green I detection chemistry (Applied Biosystems) on the ABI Prism 7500 Fast Real–Time PCR System. The ribosomal constituent RPLPO was used as a control gene for normalization. All RT-PCR and ChIP experiments are representative results from a single experiment, which were performed multiple times. The error bars indicate standard deviation of triplicate qPCR data. Primer sequences used are available upon request.

Western blotting

Whole cell or nuclear protein samples were prepared in RIPA buffer containing protease inhibitors (25 mM Tris·HCl pH 7.6, 150 mM NaCl, 1% NP40, 1% sodium deoxycholate, 0.1% SDS, 1 μg ml^{-1} aprotinin, 10 μg ml^{-1} leupeptin, 1 mM PMSF). Protein lysates were separated on SDS–PAGE gels and transferred to nitrocellulose membranes. Membranes were subsequently probed using the relevant primary and secondary antibodies, and relative protein levels were determined by chemiluminescence.

Immunoprecipitations

Immunoprecipitations (IPs) were performed on nuclear protein lysates prepared in low salt buffer containing protease inhibitors (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% NP40, 1 μg ml^{-1} aprotinin, 10 μg ml^{-1} leupeptin, 1 mM PMSF). IPs of FLAG–tagged proteins were performed using M2 anti–FLAG agarose (Sigma) overnight at 4°C. Elution of FLAG–tagged proteins was performed at 4°C using 250 μg ml^{-1} of 3× FLAG peptide (Sigma) in 0.05% NP40 with horizontal shaking. Eluted protein fractions were separated by SDS–PAGE and analyzed by Western blot or liquid chromatography mass spectrometry.

Antibodies

The following antibodies were used for immunoblotting: rabbit αPHF19 (this work) (1:250), rabbit αNO66 (ref. 36) (1:500), mouse αEZH2 (BD43) (1:8)[55], mouse αEED (AA19) (1:10)[56], rabbit αSUZ12 (Abcam ab12073) (1:1000), rabbit αCBX8 (1:1000)[40], rabbit αGAL4 (Santa Cruz, sc-577) (1:500), rabbit αH3 (Abcam, ab1791) (1:5000), rabbit αH3K27me3 (Millipore, 07-449) (1:5000), rabbit αH3K36me3 (Abcam, ab9050) (1:5000), and mouse αFLAG (Sigma, F3165) (1:1000). The following antibodies were used for immunoprecipitation and ChIP experiments: rabbit αPHF19 (this work) (5μg per ChIP), rabbit αNO66 (10μg per ChIP)[36], mouse αEZH2 (AC22) (2.5μg per ChIP)[40], rabbit αCBX8 (2μg per ChIP)[40], rabbit αGAL4 (Santa Cruz, sc-577) (5μg per ChIP), rabbit αH3K27me3 (Millipore, 07-449) (1μg per ChIP), αH3K36me3 (Abcam, ab9050) (1μg per ChIP), and mouse αFLAG (Sigma, F3165) (10μg per Mass spec scale IP). 5μg of each antibody was used for all small scale IPs.

Mass Spectroscopy

IP samples were separated by SDS–PAGE and subsequently stained with GelCode Blue (Thermo Scientific). Protein bands were excised, processed for digestion with trypsin (Promega) and analysed by LCMS as previously described[57,58].

METHODS ONLY REFERENCES

52. Garbe, J.C. et al. Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res* **69**, 7557-68 (2009).
53. Bracken, A.P. et al. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**, 525-30 (2007).
54. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
55. Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* **23**, 4061-71 (2004).

56. Bracken, A.P. et al. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* **22**, 5323-35 (2003).

57. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-8 (1996).

58. 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).
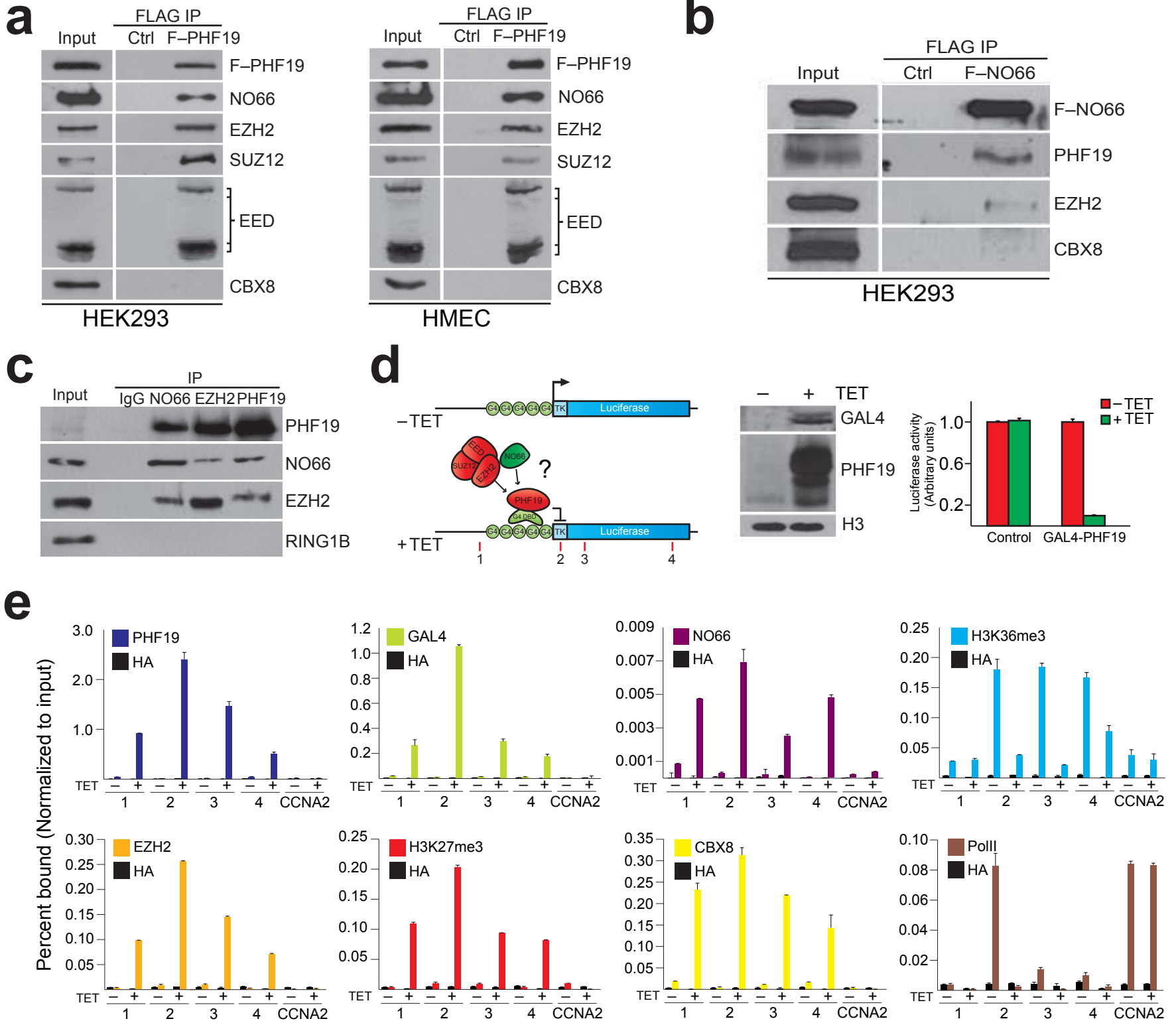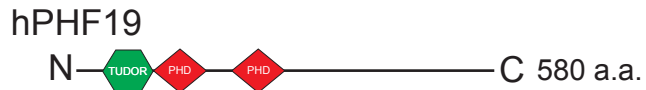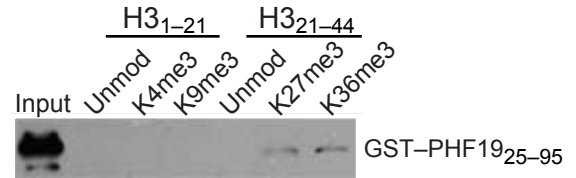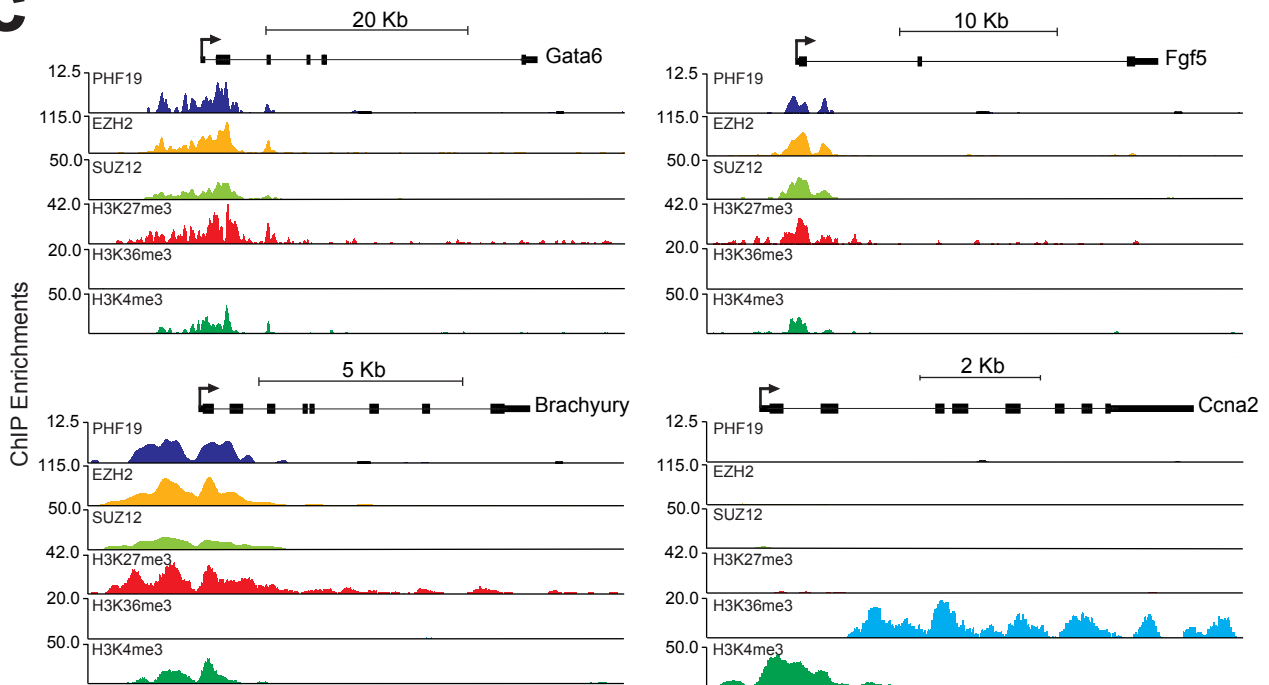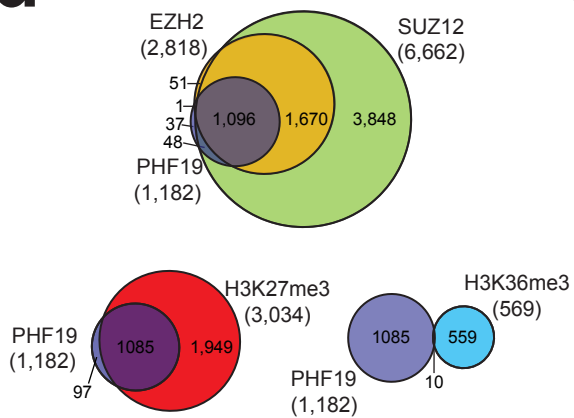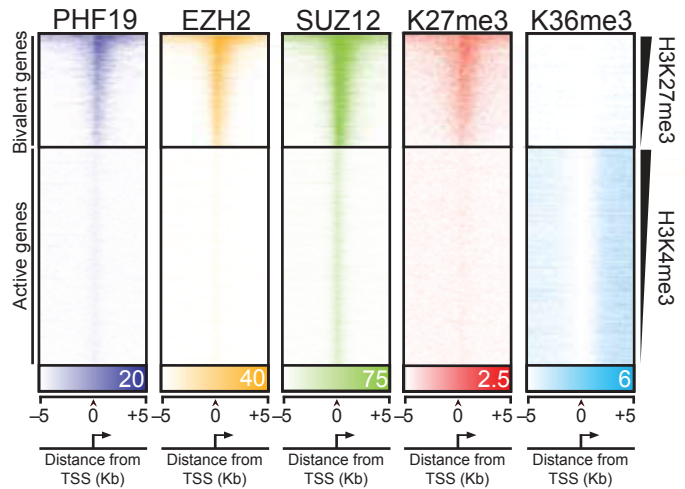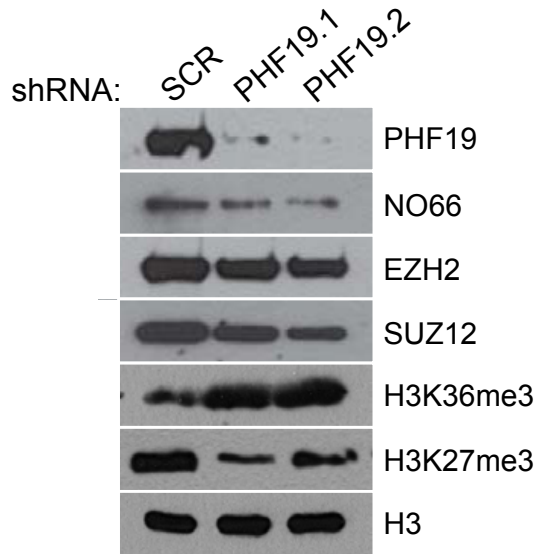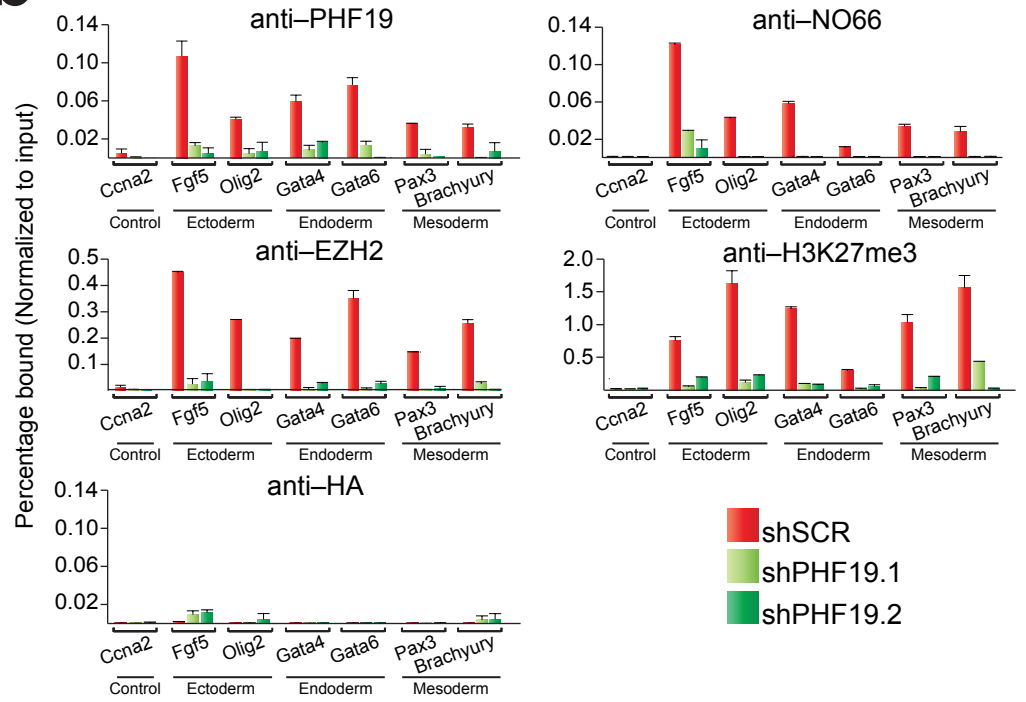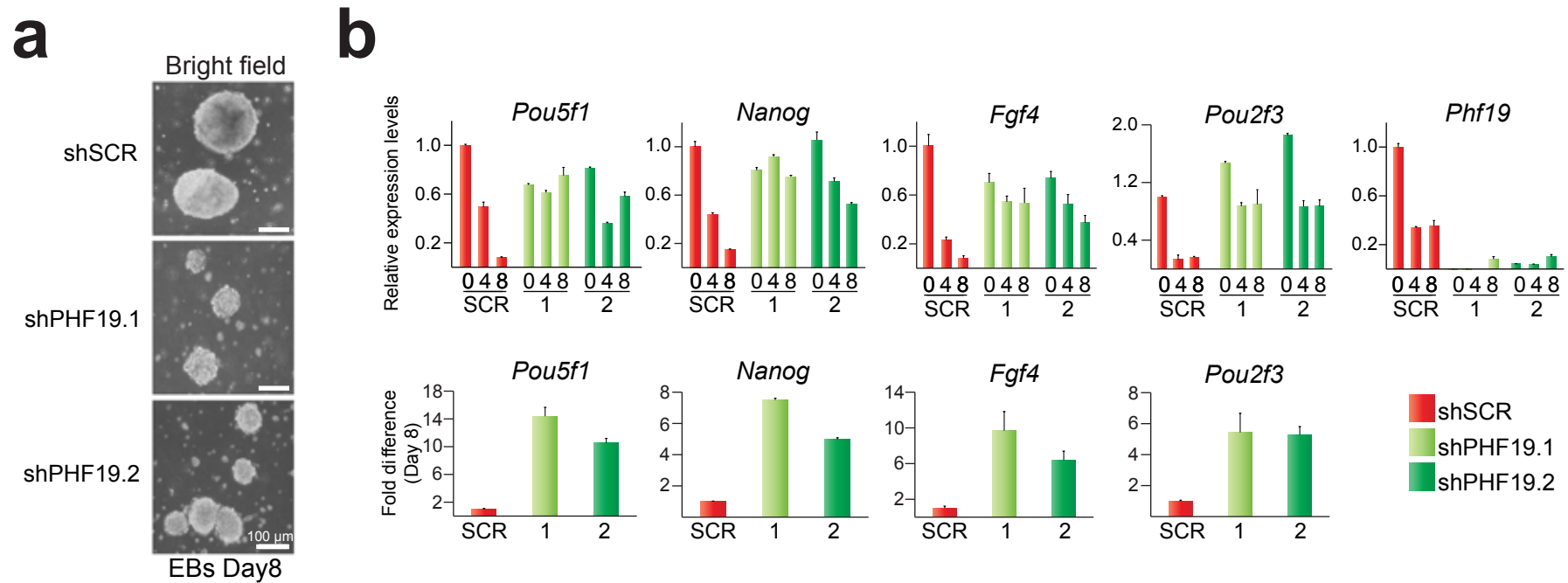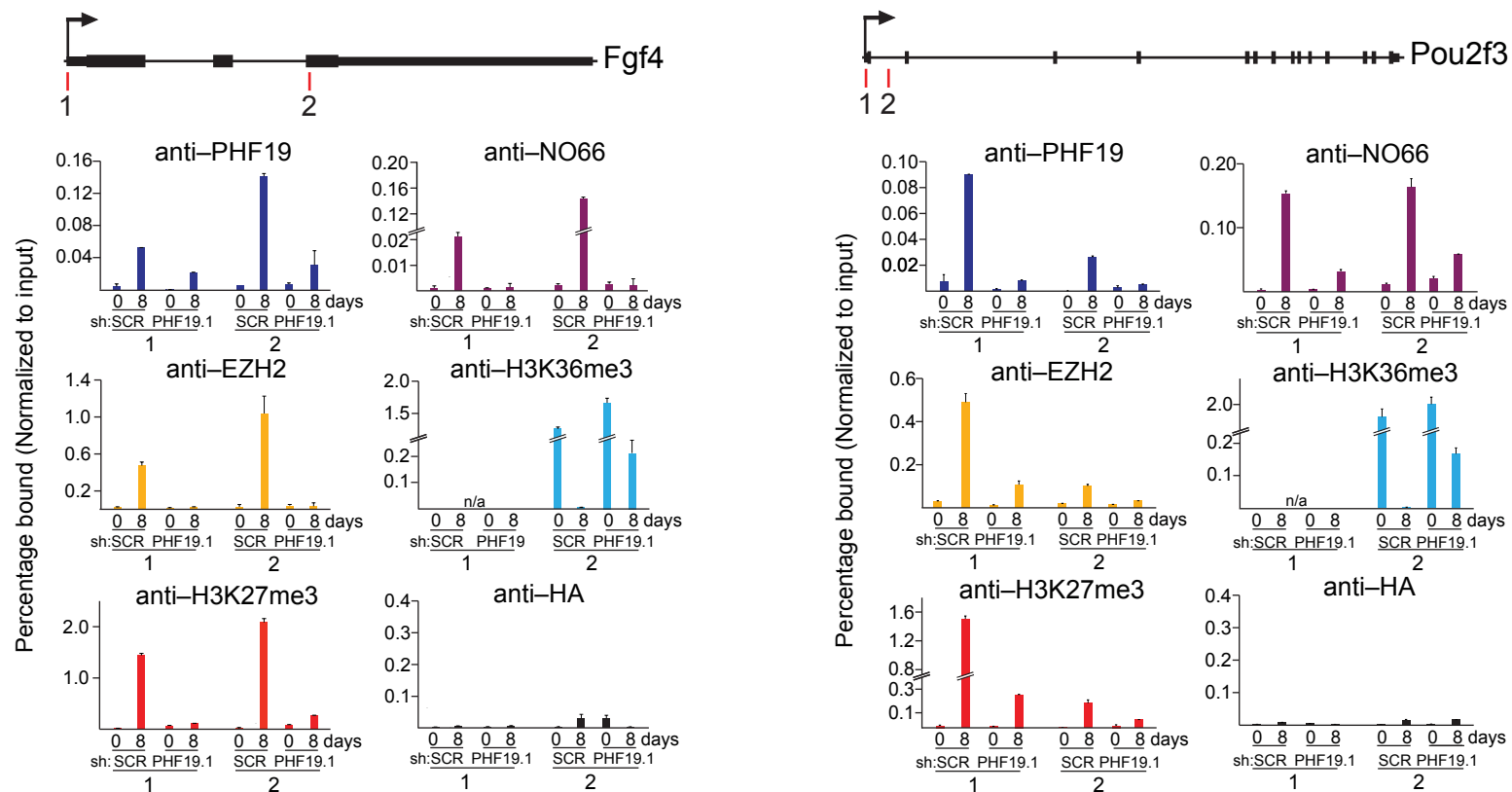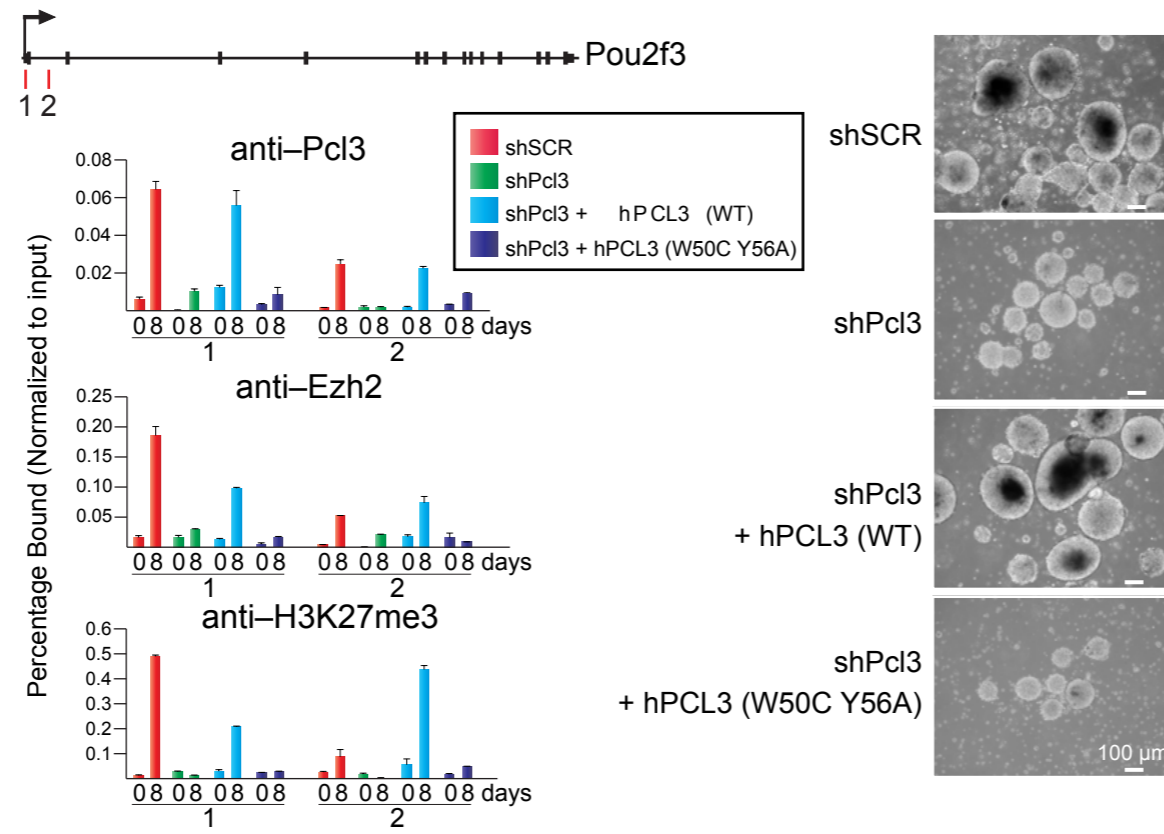
a

hPHF19

N — [TUDOR] [PHD] [PHD] —————— C 580 a.a.

b

c

ChIP Enrichments

d

e

a

b
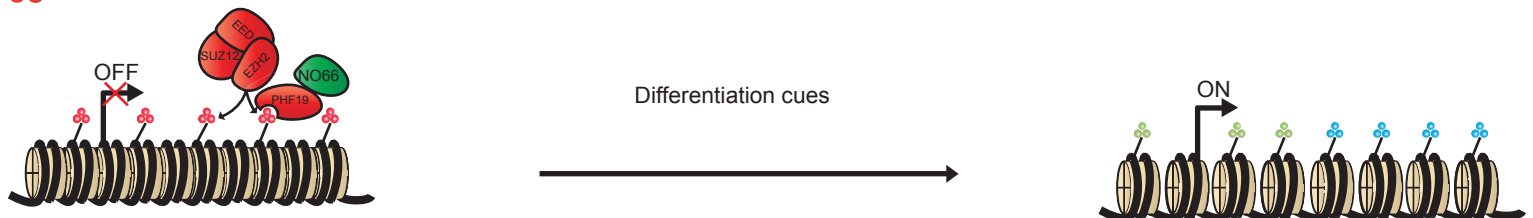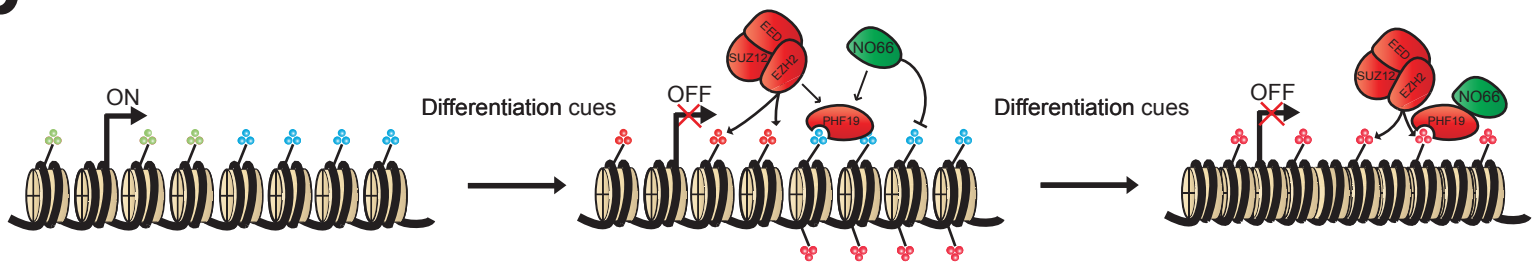
a
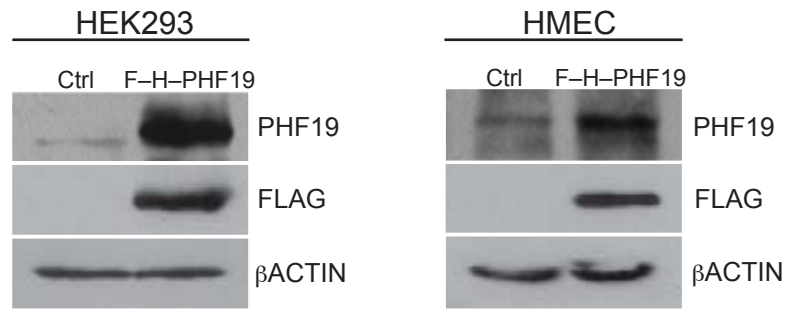
Western blots

b
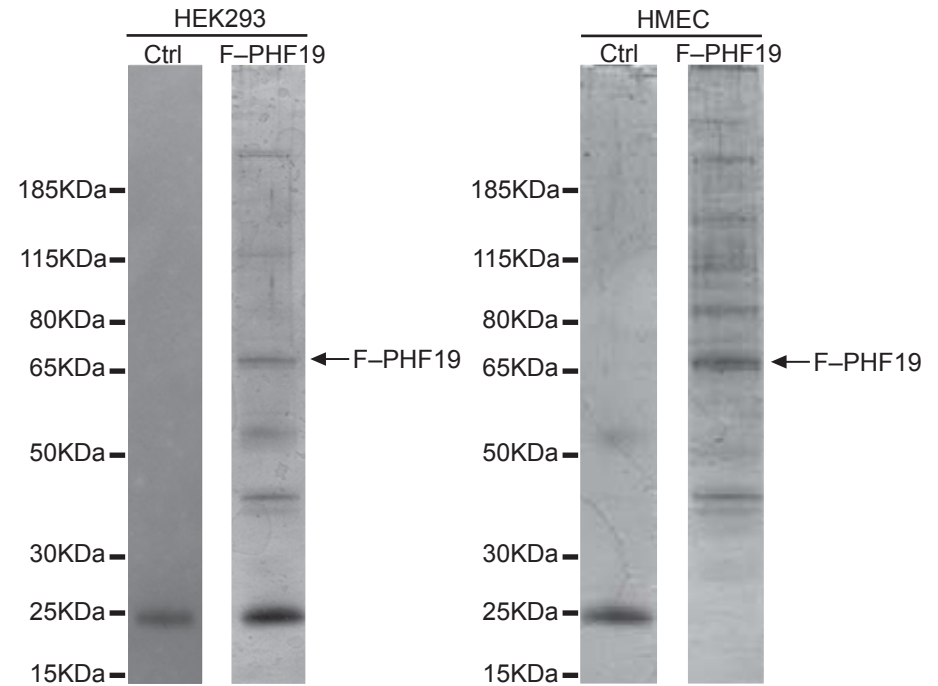
Silver stains

c

Mass spectrometric analysis of FLAG–PHF19 in HEK293 and HMEC cells						
			HEK293		HMEC	
Protein	Mass (Da)	Protein Accession	Average Peptides	Average Mascot	Average Peptides	Average Mascot
PHF19	**66803**	**gi|58331161**	**89**	**733**	**135**	**1506**
SUZ12	83744	gi|197333809	133	1390	34	330
EZH2	87247	gi|3334180	86	1060	26	193
EED	50907	gi|205790383	59	398	20	128
EZH1	87325	gi|19923202	35	166	6	25
RBBP4	47911	gi|1172846	33	312	15	111
RBBP7	48132	gi|49456363	34	241	12	61
MTF2	68301	gi|56203559	16	39	12	33
HDAC2	66294	gi|71051977	11	52	9	53
NO66	71086	gi|106879206	10	17	31	36

PRC2 (bracket spanning SUZ12 through HDAC2)

Figure S1: Biochemical analysis of PHF19.
(a) Western blot analysis of protein lysates prepared from control uninfected, or stably infected HEK293 and HMEC cell cultures expressing FLAG–HA–PHF19. **(b)** Silver-stain analysis of anti–FLAG IPs performed on cells in panel (a). **(c)** Summary of peptides identified by mass spectrometry in duplicate (HEK293) or triplicate (HMEC) anti–FLAG purifications of FLAG–HA–PHF19 stably expressed in cells from (a). Mass, accession number, average observed peptides numbers and average Mascot scores are indicated for each protein. All proteins were undetected in matched negative control experiments.
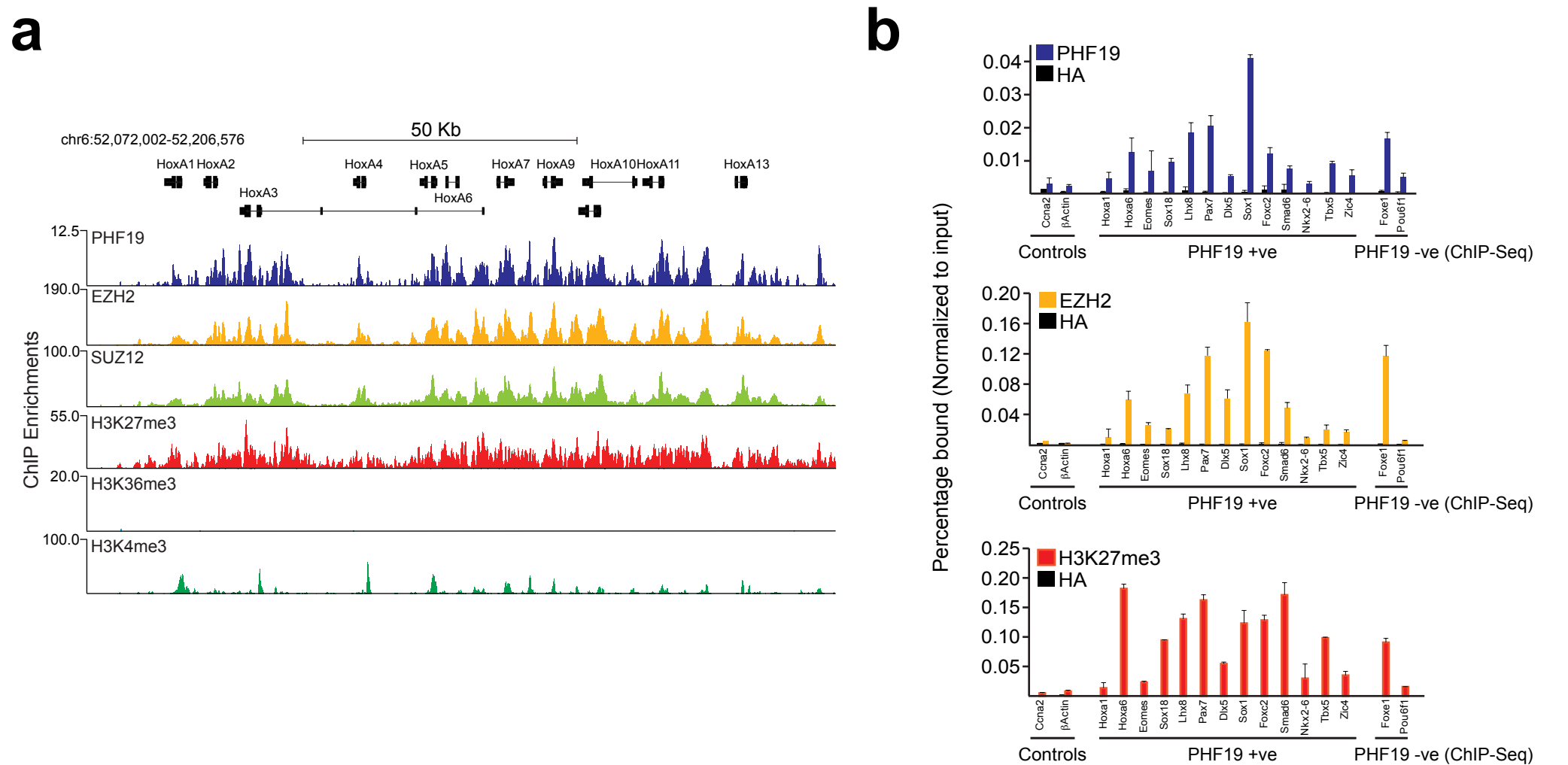
Figure S2: PHF19 co-localizes with PRC2 and H3K27me3 at target genes in mouse embryonic stem cells.
(a) Representation of PHF19 ChIP-seq results on the *HoxA* gene loci in mouse embryonic stem cells. Binding profiles of EZH2, SUZ12, H3K27me3, H3K36me3 and H3K4me3 obtained from previous studies are indicated for comparison. **(b)** Quantitative-ChIP analysis of PHF19, EZH2 and H3K27me3 in mouse embryonic stem cells. The *FoxE1* and *Pou6f1* genes were not identified as PHF19 targets in the ChIP-Seq analysis. The precipitated DNA was analyzed by quantitative RT-PCR with primers corresponding to the promoter regions of the indicated genes. The *Ccna2* and β*Actin* are shown as negative controls. ChIP enrichments are presented as percentage bound, normalized to input.
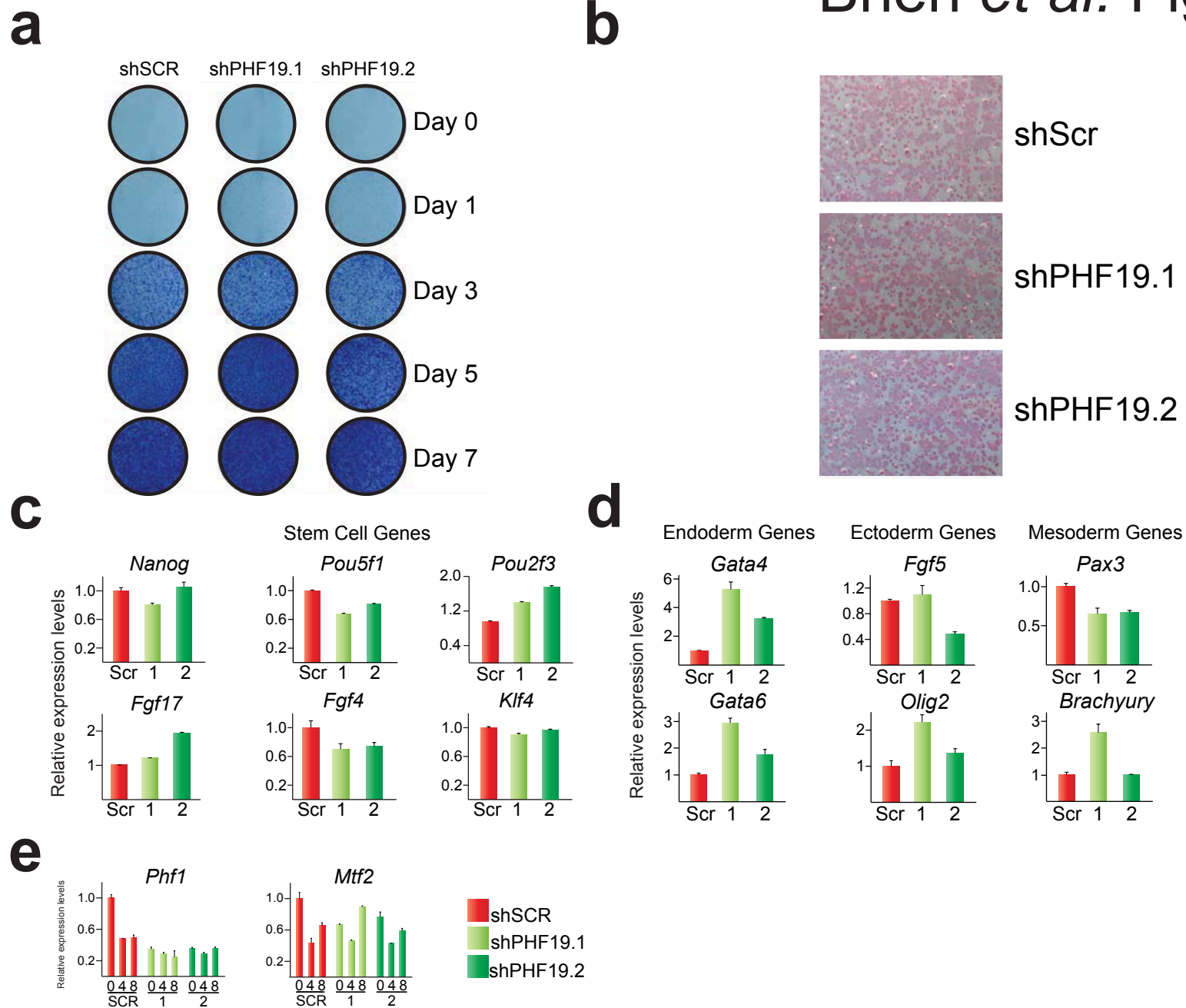
Figure S3: Depletion of PHF19 does not significantly affect the ES cell phenotype.
(a) Proliferation of embryonic stem cells is not affected by PHF19 inhibition. Mouse embryonic stem cells expressing either non–specific scrambled (shSCR) or PHF19 specific shRNAs (1 and 2) were seeded for growth assays and stained with crystal violet at the indicated time–points. (b) Alkaline phosphatase staining is not affected upon inhibition of PHF19. (c) qPCR expression analysis of embryonic stem cell expressed genes in cells from (a). (d) qPCR expression analysis of germ cell marker genes in cells from (a). (e) qPCR expression analysis of the *Phf19* homologues, *Phf1* and *Mtf2*, in cells from (a) marked with a "0", or the same cells induced to differentiate into embryoid bodies for 4 and 8 days.
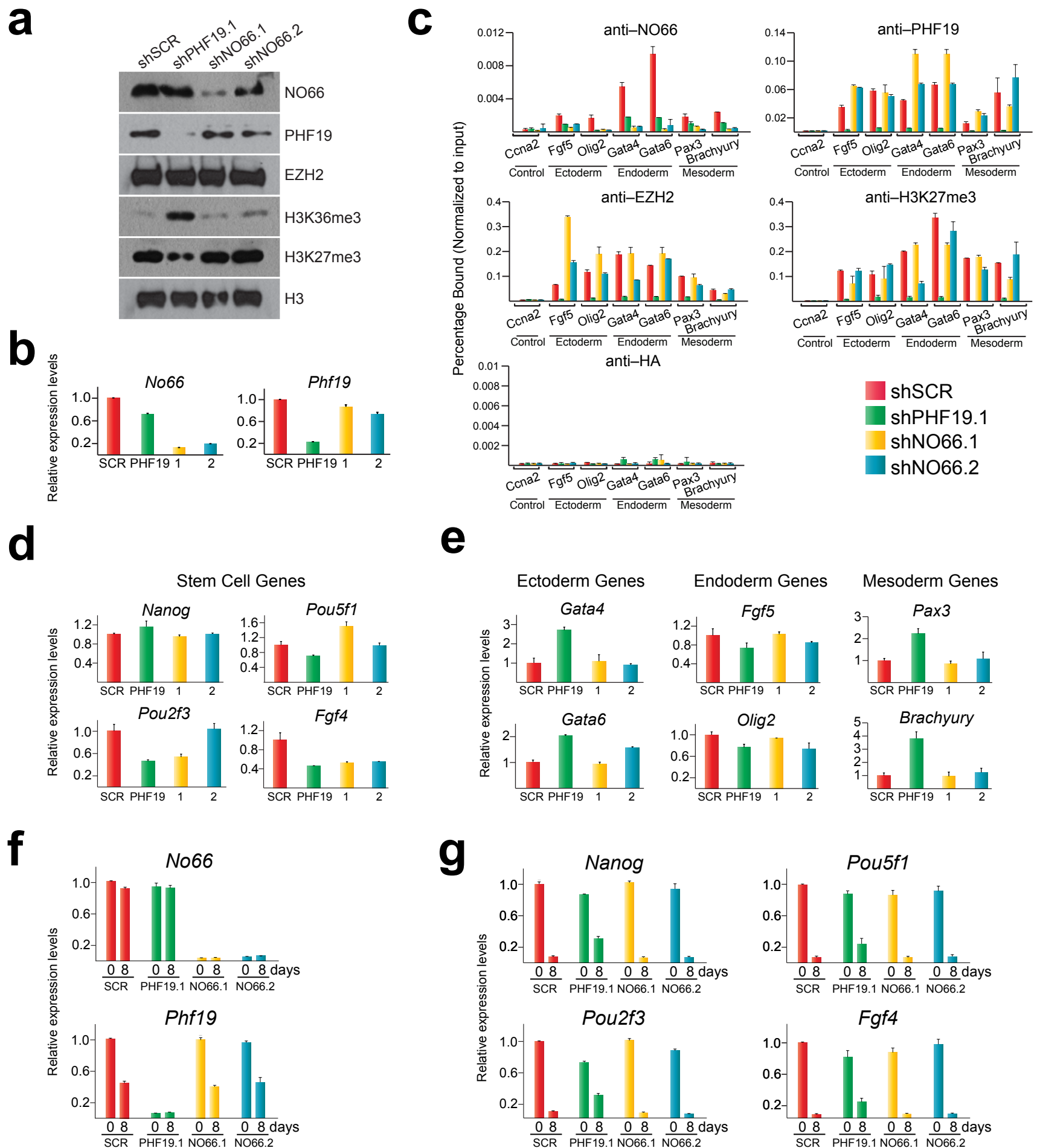
Figure S4: NO66 is not required for ES cell differentiation or PRC2 occupancy on target genes
(a) Western blot analysis of mouse embryonic stem cells expressing either non–specific scrambled (shSCR), *Phf19* specific (shPHF19.1) or one of two *No66* specific (shNO66.1 and shNO66.2) shRNAs. (b) qPCR expression analysis of the indicated genes in cells from (a). (c) Quantitative ChIP analysis of PHF19, NO66, EZH2 and H3K27me3 in cells from (a). The precipitated DNA was analysed by quantitative RT-PCR with primers corresponding to the promoter regions of the indicated germ cell marker genes and presented as percentage bound, normalised to input. The *Ccna2* gene promoter is presented as a negative control. (d) qPCR expression analysis of embryoinc stem cell expressed genes in cells from (a). (e) qPCR expression analysIs of germ cell marker genes in cells from (a). (f) qPCR expression analysis of the indicated genes in either undifferentiated mouse embryonic stem cells from (a) (indicated with a "0"), or the same cells induced to differentiate into embryoid bodies for 8 days. (g) qPCR expression analysIs of the indicated embryonic stem cell expressed genes in cells from (f).
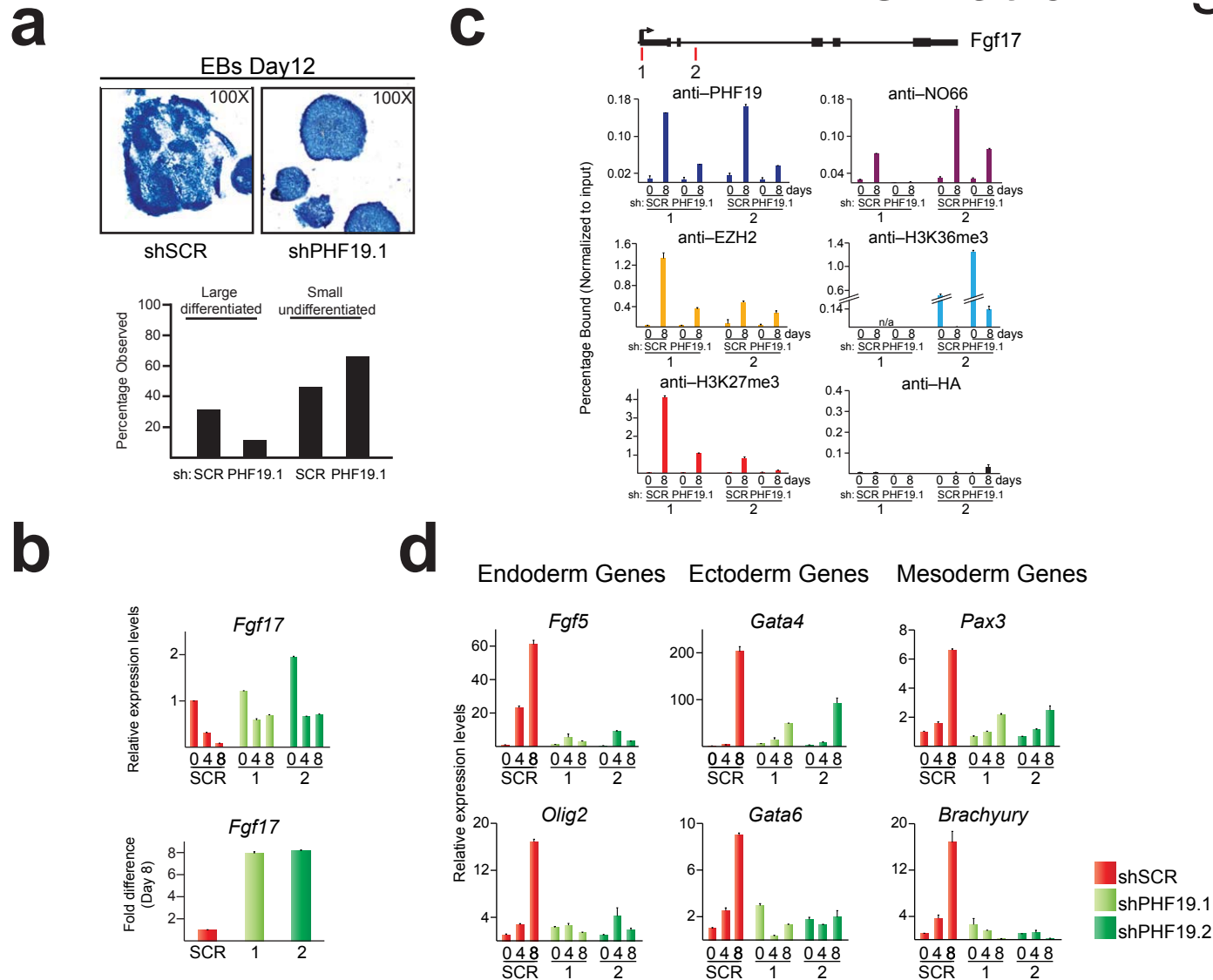
Figure S5: Depletion of PHF19 impairs the ability of ES cells to differentiate into embryoid bodies in suspension culture.
(a) Hematoxylin and Eosin (H&E) stains of embryoid bodies formed after 12 days from embryonic stem cell expressing either shSCR or shPHF19 (Top Panel). A total of 100 embryoid bodies were examined and the overall percentage of large embryoid bodies exhibiting structural features of differentiation 'Large differentiated' and small embryoid bodies which lacked these features 'Small undifferentiated' in shSCR and shPHF19 cells was examined (Bottom Panel). Presented are representative data from a single experiment which was performed three times **(b)** qPCR expression analysis of the indicated embryonic stem cell gene, *Fgf17*, in undifferentiated shSCR, shPHF19.1 or shPHF19.2 mouse embryonic stem cells (indicated with a "0") or the same cells induced to differentiate towards embryoid bodies for 4 and 8 days. **(c)** ChIP analysis of PHF19, NO66, EZH2, H3K27me3 and H3K36me3 in cells from (b). The precipitated DNA was analysed by quantitative RT-PCR with primers corresponding to the promoter (1 in the Top Panel) or intragenic (2 in the Top Panel) regions of the embryonic stem cell expressed gene *Fgf17*. Results are presented as percentage bound, normalised to input. **(d)** qPCR expression analysis of the indicated germ cell differentiation genes in cells from panel (b).
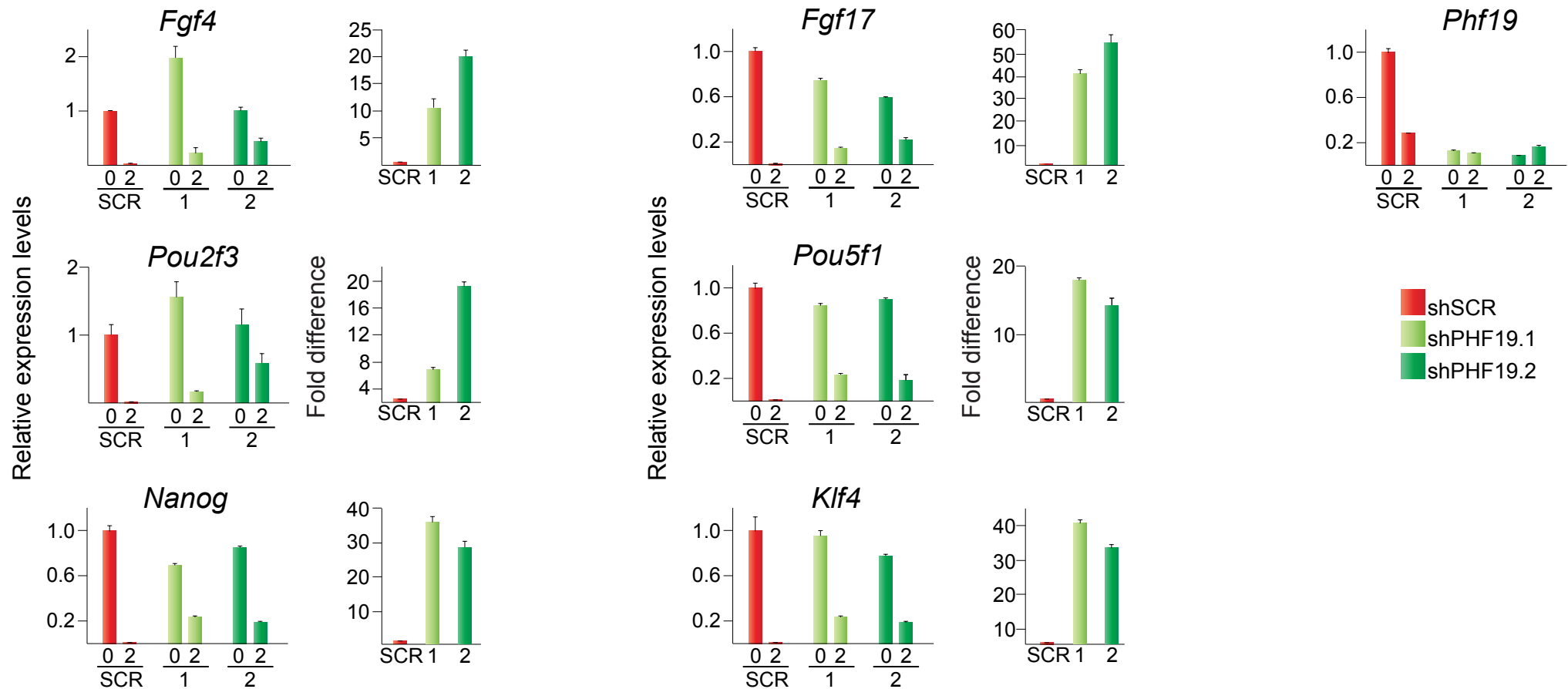
Figure S6: Depletion of PHF19 leads to impaired silencing of ES cell expressed genes following retinoic acid induced differentiation.
qPCR analysis of the indicated genes in mouse embryonic stem cells, expressing either non-specific scrambled (shSCR) or PHF19 specific (1 and 2) shRNAs (indicated with a "0"), or the same cells induced to differentiate with 1µM *all-trans* retinoic acid (ATRA) for 2 days. Also shown are the relative fold differences in expression of each embryonic stem cell expressed genes in shSCR, shPHF19.1 and shPHF19.2 cells after 2 days of RA treatment.
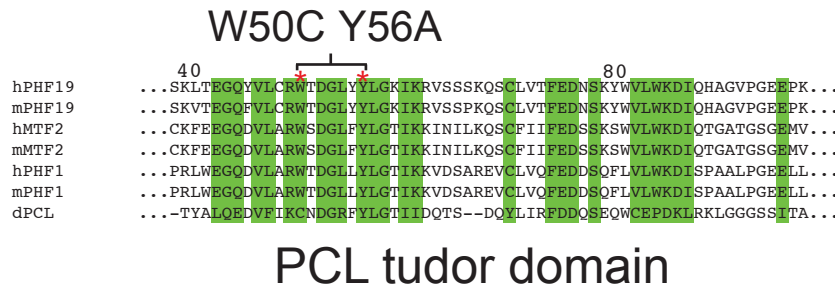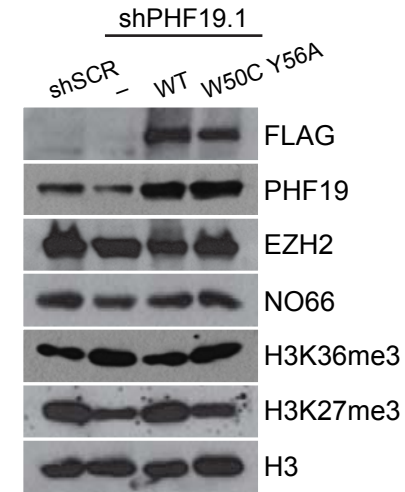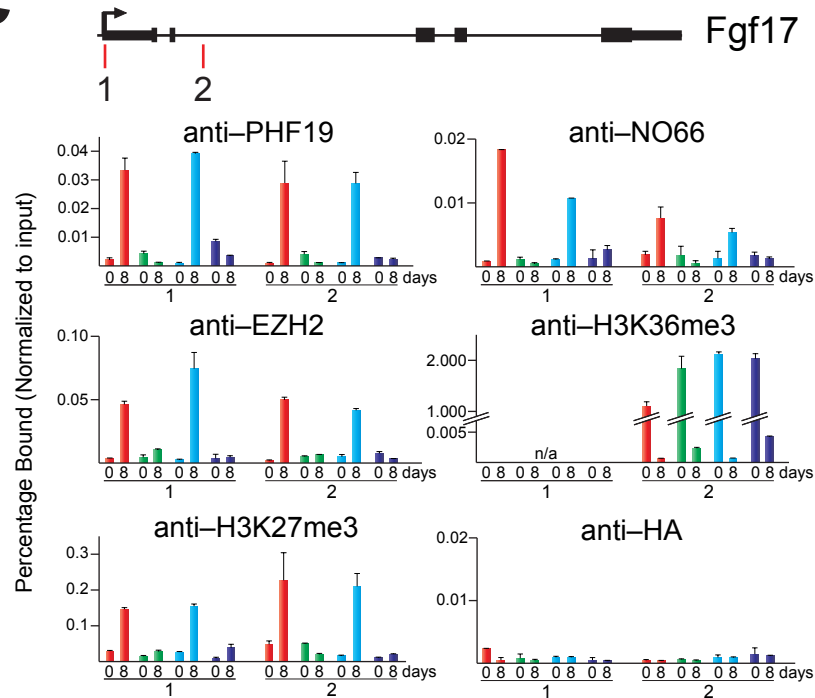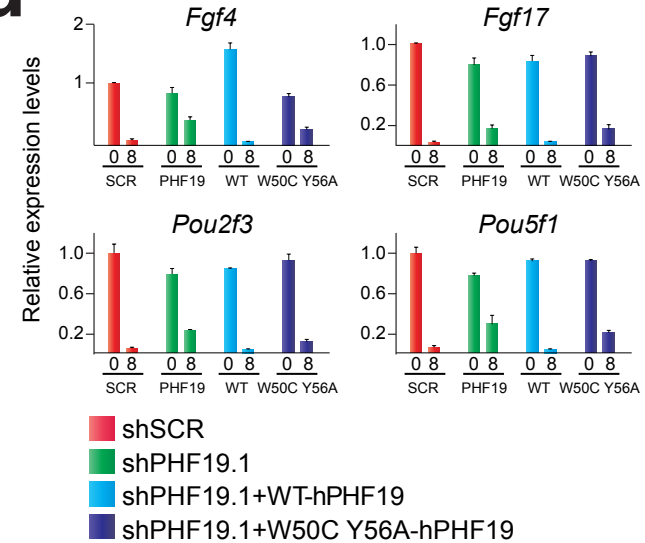
Figure S7: A PHF19 double point mutant is unable to rescue the differentiation defect seen in mouse embryonic stem cells depleted of endogenous PHF19. (a) Protein alignments of the tudor domains of human, mouse and Drosophila PCL proteins. Indicated are the residues targeted in the double point mutated form of PHF19. **(b)** Western blot analysIs of mouse embryonic stem cells expressing either non-specific scrambled (shSCR) or PHF19 specific (shPHF19) shRNAs, where PHF19 knockdown was rescued with wildtype (WT) or mutated (W50C Y56A) PHF19. **(c)** Quantitative ChIP analysis of PHF19, NO66, EZH2, H3K36me3 and H3K27me3 in undifferentiated cells mouse embryonic stem cells from panel (b) (indicated with a "0"), or the same cells induced to differentiate as embryoid bodies for 8 days. The precipitated DNA was analysed by quantitative RT-PCR with primers corresponding to the promoter (1 in the Top Panel) or intragenic (2 in the Top Panel) regions of the embryonic stem cell expressed gene *Fgf17*. Results are presented as percentage bound, normalised to input. **(d)** qPCR expression analysis of the indicated embryonic stem cell expressed genes in cells from panel (c).