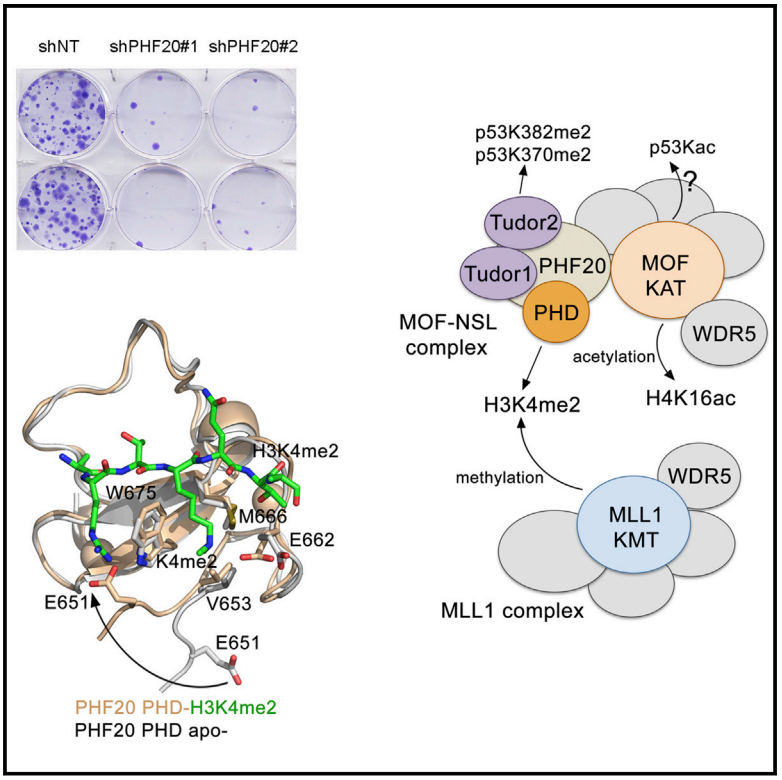
Article

Cell Reports

# PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation

## Graphical Abstract




## Authors

Brianna J. Klein, Xiaoyan Wang, Gaofeng Cui, ..., Georges Mer, Xiaobing Shi, Tatiana G. Kutateladze

## Correspondence

mer.georges@mayo.edu (G.M.), xbshi@mdanderson.org (X.S.), tatiana.kutateladze@ucdenver.edu (T.G.K.)

## In Brief

Klein et al. find that the PHD finger of PHF20, a core subunit of the MOF-NSL HAT complex responsible for H4K16 acetylation, is a reader of H3K4me2. The authors show that binding to H3K4me2 is required for PHF20-dependent histone acetylation, target gene activation, and cancer cell growth and survival.

## Highlights

- PHF20 is essential in cancer cell growth and survival
- The PHD finger of PHF20 exhibits high selectivity for H3K4me2
- H3K4me2 binding is necessary for MOF-dependent H4K16 acetylation and gene regulation
- PHF20 functions link MOF HAT, p53, and H3K4me2

## Accession Numbers

5TAB
5TBN
GSE82115

Klein et al., 2016, Cell Reports *17*, 1158–1170
October 18, 2016 
http://dx.doi.org/10.1016/j.celrep.2016.09.056

OPEN ACCESS
CellPress

Cell Reports
Article

# PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation

Brianna J. Klein,[1,6] Xiaoyan Wang,[2,4,6] Gaofeng Cui,[3,6] Chao Yuan,[2] Maria Victoria Botuyan,[3] Kevin Lin,[2] Yue Lu,[2] Xiaolu Wang,[2] Yue Zhao,[5] Christiane J. Bruns,[4,5] Georges Mer,[3,*] Xiaobing Shi,[2,*] and Tatiana G. Kutateladze[1,7,*]

[1]Department of Pharmacology, University of Colorado School of Medicine, Aurora, CO 80045, USA
[2]Department of Epigenetics and Molecular Carcinogenesis, Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[3]Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905, USA
[4]Faculty of Medicine, Ludwig Maximilian University of Munich, Munich 80336, Germany
[5]Department of General, Visceral and Tumor Surgery, University Clinic Cologne, Cologne 50931, Germany
[6]Co-first author
[7]Lead Contact
*Correspondence: mer.georges@mayo.edu (G.M.), xbshi@mdanderson.org (X.S.), tatiana.kutateladze@ucdenver.edu (T.G.K.)
http://dx.doi.org/10.1016/j.celrep.2016.09.056

## SUMMARY

**PHF20 is a core component of the lysine acetyltransferase complex MOF (male absent on the first)-NSL (non-specific lethal) that generates the major epigenetic mark H4K16ac and is necessary for transcriptional regulation and DNA repair. The role of PHF20 in the complex remains elusive. Here, we report on functional coupling between methylation readers in PHF20. We show that the plant homeodomain (PHD) finger of PHF20 recognizes dimethylated lysine 4 of histone H3 (H3K4me2) and represents an example of a native reader that selects for this modification. Biochemical and structural analyses help to explain this selectivity and the preference of Tudor2, another reader in PHF20, for dimethylated p53. Binding of the PHD finger to H3K4me2 is required for histone acetylation, accumulation of PHF20 at target genes, and transcriptional activation. Together, our findings establish a unique PHF20-mediated link between MOF histone acetyltransferase (HAT), p53, and H3K4me2, and suggest a model for rapid spreading of H4K16ac-enriched open chromatin.**

## INTRODUCTION

PHF20 is a component of the MOF (male absent on the first)-NSL (non-specific lethal) lysine acetyltransferase complex responsible for acetylation of histone H4 and non-histone proteins and is implicated in transcriptional regulation and Ataxia telangiectasia mutated (ATM)-dependent DNA damage response (Avvakumov and Côté, 2007; Cai et al., 2010; Li et al., 2009; Rea et al., 2007). The MOF-NSL complex shares the WD repeat domain 5 (WDR5) subunit with the H3K4-specific methyltransferase complex, MLL1, and stably associates with the MLL1 complex (Cai et al., 2010; Dou et al., 2005; Zhao et al., 2013b). The joint recruitment and coordinated activities of both MOF and MLL1 complexes are required for optimal transcriptional activation of a set of genes, and a synergistic distribution of H3K4me and H4K16ac marks at promoters of these genes is evident (Dou et al., 2005; Zhao et al., 2013b). Genomic and biochemical studies reveal that MOF-NSL stimulates MLL1 activity, enhancing dimethylation of H3K4 in an acetylation-dependent manner, and depletion of MOF or the NSL complex results in a reduction of H4K16 acetylation and H3K4 methylation (Zhao et al., 2013b).

The MOF-NSL complex is also capable of acetylating non-histone proteins, such as the transcription factor p53. Triggered by DNA damage, K120 acetylation in the DNA-binding domain of p53 stimulates expression of pro-apoptotic genes, promoting cell death (Li et al., 2009; Sykes et al., 2006). Additionally, p53 activity can be regulated by non-catalytic subunits of the MOF-NSL complex, OGT1 and PHF20. OGT1-catalyzed O-GlcNAcylation at S149 stabilizes p53 through impeding T155 phosphorylation and ubiquitin-dependent proteolysis of p53 (Yang et al., 2006). PHF20 directly binds to p53 dimethylated at K370 or K382 (p53K370me2 and p53K382me2) through its second Tudor (Tudor2) domain (Cui et al., 2012). Tudor2 forms a dimer capable of associating with both p53 dimethyllysine marks simultaneously, thus greatly enhancing binding of PHF20 to p53 (Cui et al., 2012). Interaction with PHF20 leads to the stabilization and activation of p53, because it blocks p53 ubiquitination and upregulates p53 in response to DNA damage (Cui et al., 2012). Recent in vivo and in vitro studies demonstrate that PHF20 transcriptionally regulates p53 in an Akt-dependent manner (Park et al., 2012) and promotes nuclear factor κB (NF-κB) transcriptional activity (Zhang et al., 2013). Loss of PHF20 results in a decreased expression of genes with elevated H4K16ac levels at their promoters, further supporting the notion that PHF20 acts as a transcriptional regulator (Badeaux et al., 2012).

Originally identified as an antigen in glioblastoma patients, PHF20 is highly expressed in a number of cancers and is implicated in the development and progression of glioma, adenocarcinomas, and lung cancer (Bankovic et al., 2010; Fischer et al.,

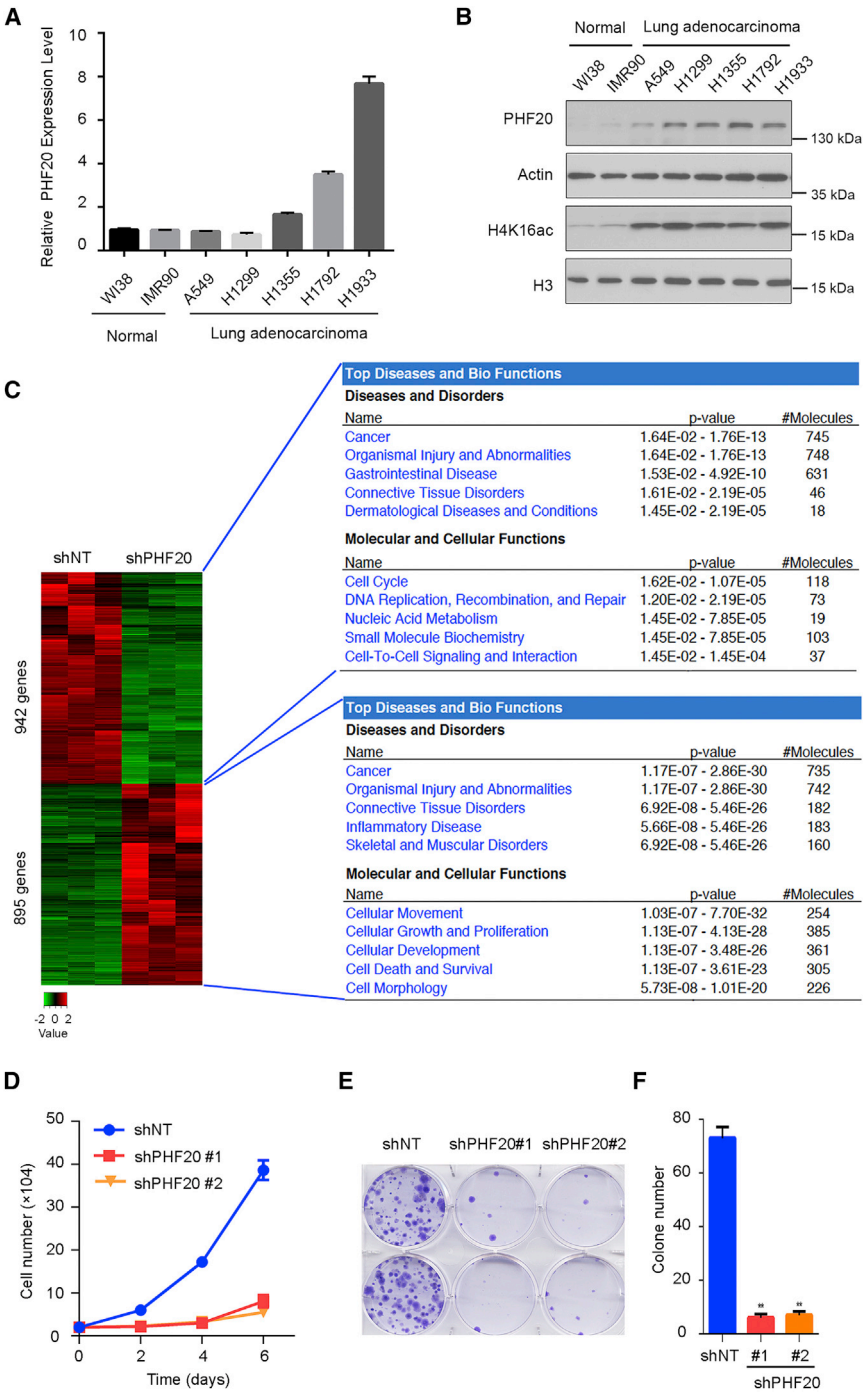



**Figure 1. PHF20 Is Amplified in Lung Cancer Cells and Is Required for Cancer Cell Growth and Survival**

(A and B) PHF20 is amplified in lung adenocarcinoma cell lines. qRT-PCR (A) and western blot analysis (B) of PHF20 gene expression and protein levels, respectively, in the indicated lung adenocarcinoma cell lines and control immortalized normal cells. GAPDH was used as internal control in (A), and actin was used as a loading control in (B). Error bars represent SEM of three replicates.

(C) Heatmap and IPA of gene expression profiles in the control (shNT) and PHF20 KD (shPHF20) H1792 cells. The green and red colors indicate down- and upregulation, respectively. The right panel shows the IPA of the down- and upregulated genes in PHF20 KD cells. The top five hits in each category are listed. The full lists of up- and downregulated genes are provided in Table S1.

(D) Cell proliferation assays of the control and PHF20 KD H1792 cells. Live cells were counted over a 6-day time course. Error bars represent SEM of six replicates.

(E and F) Colony formation assays of the control and PHF20 KD H1792 cells. Representative crystal-violet-stained cells are shown in (E), and quantification is shown in (F).

Error bars represent SEM of six replicates. $^{**}p < 0.01$. See also Figure S1 and Table S1.

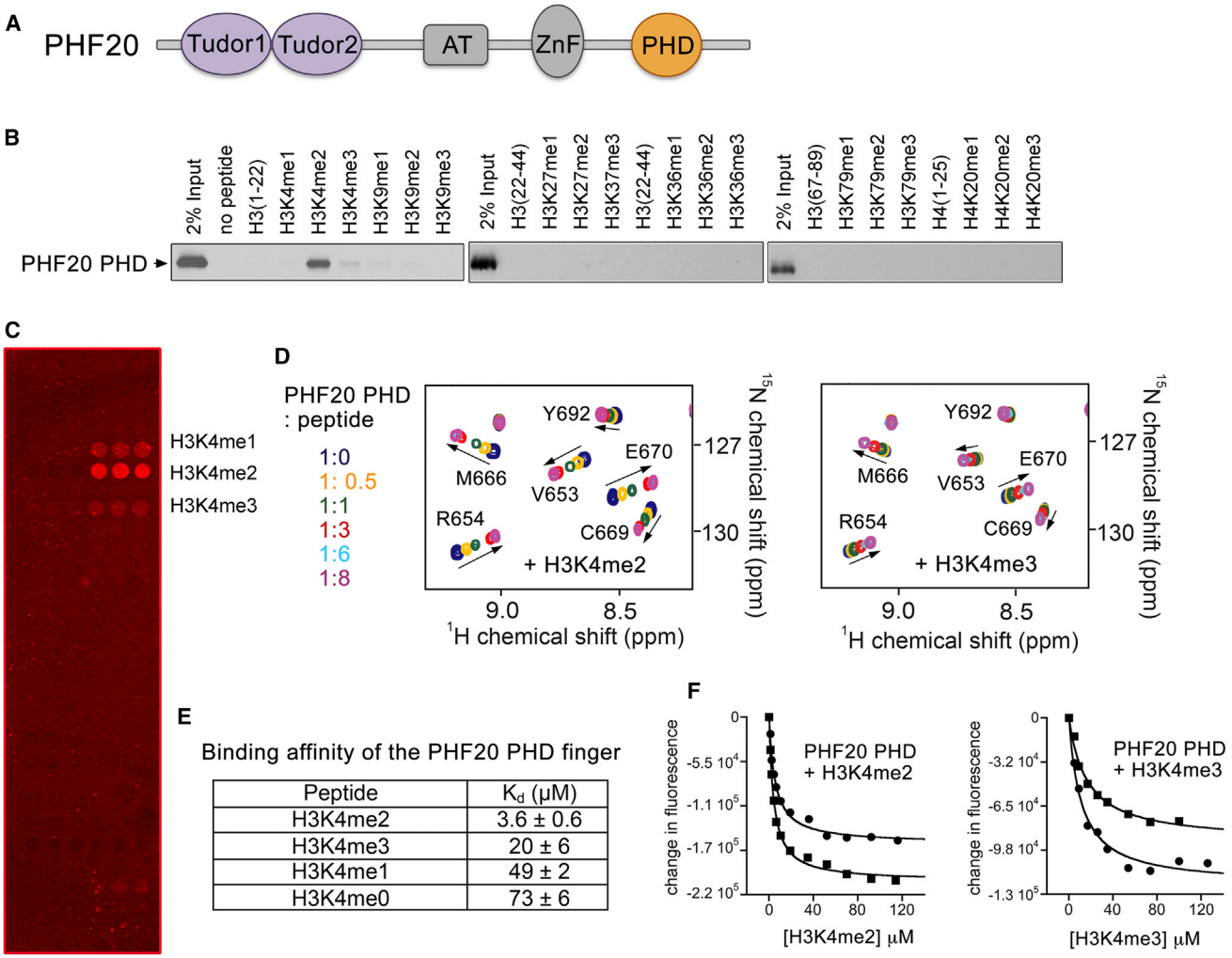
tal role of PHF20 in survival and carcinogenesis, the molecular mechanism by which PHF20 contributes to transcription and p53 regulation and exerts carcinogenic activity remains unclear. In this work, we report on a unique function of PHF20 that couples MOF histone acetyltransferase (HAT) activity with dimethylated lysine 4 of histone H3 (H3K4me2)–enriched chromatin and p53.

## RESULTS AND DISCUSSION

### PHF20 Is Required for Cancer Cell Growth and Survival

To establish the biological function of PHF20, we first analyzed the TCGA database and assessed *PHF20* gene alterations in human cancers. *PHF20* is amplified in a variety of human malignancies, including colorectal, uterine, cervical, bladder, and lung cancers (Figure S1A). We then carried out qRT-PCR and immunoblotting analyses to determine the PHF20 gene expression and protein levels across a number of lung adenocarcinoma cell lines. Compared with the normal lung fibroblast cells (WI-38 and IMR90), PHF20 is overexpressed in all lung cancer cell lines that we examined, especially at protein levels (Figures 1A and 1B). To identify the genes regulated by PHF20 genome-wide, we knocked down PHF20 gene expression in H1792 lung adenocarcinoma cells by short

2001; Heisel et al., 2008; Taniwaki et al., 2006). Overexpression of PHF20 is proposed to drive constitutive NF-κB activation in some tumors (Zhang et al., 2013). PHF20 knockout mice die shortly after birth and display a variety of phenotypes within the skeletal and hematopoietic systems (Badeaux et al., 2012). PHF20 deficiency halts conversion of somatic cells into induced pluripotent stem cells (iPSCs), revealing a requirement of this factor for cell reprogramming (Zhao et al., 2013a). Despite the vi-




| Peptide | $K_d$ (μM) |
|---|---|
| H3K4me2 | 3.6 ± 0.6 |
| H3K4me3 | 20 ± 6 |
| H3K4me1 | 49 ± 2 |
| H3K4me0 | 73 ± 6 |

**Figure 2. The PHF20 PHD Finger Binds to H3K4me2**

(A) Schematic representation of PHF20.

(B) Western blot analysis of histone peptide pull-downs with GST-PHF20 PHD and the indicated biotinylated peptides.

(C) Histone peptide array probed with GST-PHF20 PHD.

(D) Superimposed $^{1}H$,$^{15}N$ HSQC spectra of the PHF20 PHD finger collected upon titration with the indicated peptides. Spectra are color coded according to the protein:peptide molar ratio shown on the left.

(E) Binding affinities of WT PHF20 PHD for the indicated histone peptides measured by tryptophan fluorescence.

(F) Representative binding curves used to determine the $K_D$ values.

hairpins (shRNAs) (Figures S1B and S1C) and performed RNA sequencing (RNA-seq) analysis. PHF20 was reported as a transcriptional coactivator; however, we found similar numbers of genes were down- (942 genes) and upregulated (895 genes) in PHF20 KD cells compared with the control cells (Figure 1C). Ingenuity Pathway Analysis (IPA) revealed that the most affected genes are implicated in vital biological processes, including cell cycle and DNA replication, recombination and repair, cellular movement, and cell proliferation and death. Notably, both down- and upregulated genes are enriched for cancer, organismal injury, and connective tissue diseases (Figure 1C). Furthermore, cell proliferation assays showed that depletion of PHF20 markedly reduced the cell growth and colony formation of H1792 cells (Figures 1D–1F). Together, these data suggest that PHF20 has an oncogenic role in lung adenocarcinoma, promoting cancer cell growth and survival.

## PHF20 Recognizes Histone H3K4me2 via Its PHD Finger

PHF20 has a modular architecture consisting of tandem Tudor domains (Tudor1 and Tudor2), an AT-hook, a C2H2-type zinc finger, and a plant homeodomain (PHD) finger (Figure 2A). Biochemical and structural studies have shown that Tudor2 associates with dimethyllysine substrates (Adams-Cioaba et al., 2012; Cui et al., 2012; Kim et al., 2006), particularly with p53K370me2 and p53K382me2 (Cui et al., 2012); however, biological activities of other domains, including the PHD finger,

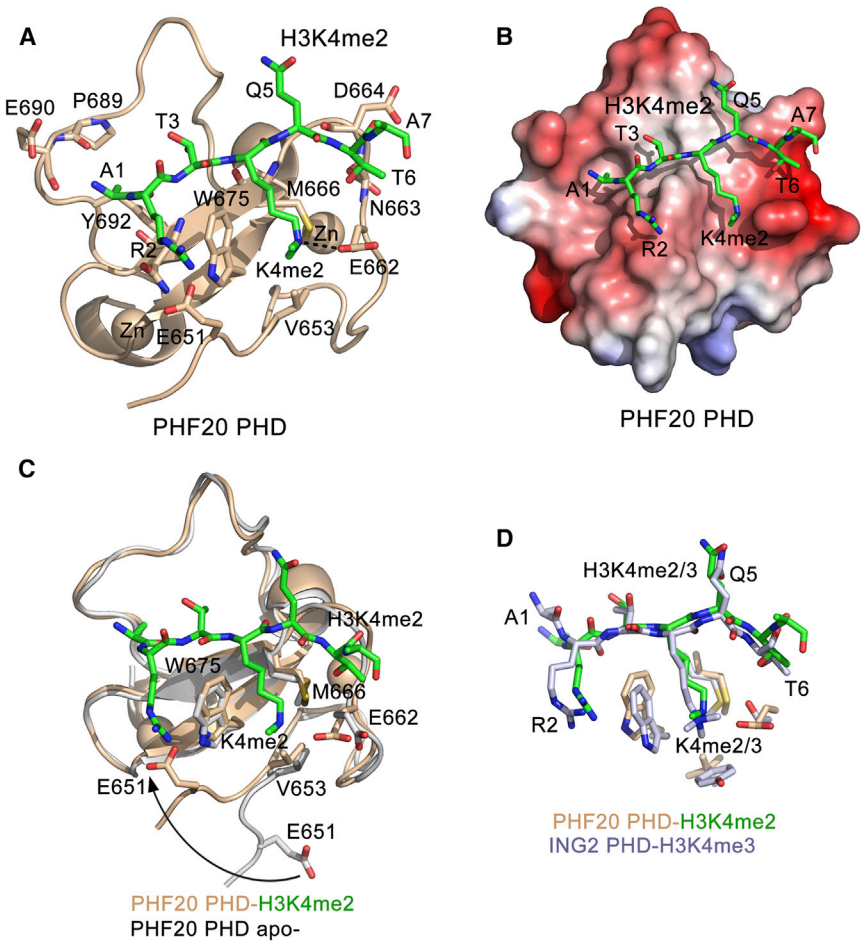



**Figure 3. Structural Mechanism for the Recognition of H3K4me2 by PHF20 PHD**

(A) The PHD finger is depicted in the ribbon diagram, and the H3K4me2 peptide is shown in a stick model. The histone peptide residues and the residues of the PHF20 PHD finger involved in the interaction are labeled.

(B) The electrostatic surface potential of the PHF20 PHD finger is shown using blue and red colors for positive and negative charges, respectively.

(C) An overlay of the NMR structure of the PHF20 PHD-H3K4me2 complex (wheat-green) with the crystal structure of the apo-state of PHF20 PHD (gray).

(D) Superimposed H3K4me2 (green)-binding site of PHF20 PHD (wheat) and H3K4me3 (light blue)-binding site of ING2 PHD (light blue) (PDB: 2G6Q) are shown.

See also Figure S2 and Tables S2 and S3.

remain unknown. Because a set of PHD finger-containing proteins was found to bind histone sequences (Li et al., 2006; Musselman and Kutateladze, 2011; Peña et al., 2006; Shi et al., 2006; Wysocka et al., 2006), we tested the PHF20 PHD finger in pull-down assays, histone peptide microarrays, and nuclear magnetic resonance (NMR). In pull-down assays, glutathione S-transferase (GST) fusion PHD finger was incubated with biotinylated unmodified histone H3 and H4 peptides or histone peptides containing single methylation marks. The histone-bound PHD finger was captured using streptavidin Sepharose beads and detected by western blot analysis (Figure 2B). Unexpectedly, we found that the PHF20 PHD finger binds to H3K4me2 peptide but does not recognize either H3K4me3 or unmodified H3 that were previously reported as ligands for PHD fingers. This specificity of the PHF20 PHD finger toward H3K4me2 was confirmed using an extended peptide microarray (Figure 2C).

To compare binding of the PHF20 PHD finger to the methylated H3K4 species in more detail, we carried out $^{1}H$,$^{15}N$ heteronuclear single quantum coherence (HSQC) titration experiments. Gradual addition of the H3K4me2 peptide to the $^{15}N$-labeled PHD finger caused large chemical shift perturbations (CSPs) in the protein, indicating direct binding (Figure 2D). However, interaction with H3K4me3 was considerably weaker judging by small CSPs observed upon titration with this peptide. Quantitative measurements of binding affinities using fluorescence spectroscopy further corroborated these results. The dissociation constant ($K_D$) for the interaction of the PHF20 PHD finger with H3K4me2 was found to be 3.6 μM; however, interaction with H3K4me3, H3K4me1, and H3K4me0 was ~6-, ~14-, and ~20-fold, respectively, weaker (Figures 2E and 2F). Collectively, our data derived from four orthogonal biochemical approaches demonstrate that the PHF20 PHD finger exhibits high selectivity for H3K4me2, representing an example of the native PHD finger capable of reading the dimethylated species.

## Molecular Basis for the Recognition of H3K4me2

To determine the mechanism underlying recognition of H3K4me2, we obtained a 1.25-Å resolution crystal structure of the PHF20 PHD finger in the apo-state and the solution structure of the PHD finger in complex with the H3K4me2 peptide (Figure 3; Tables S2 and S3). The NMR ensemble shows a canonical fold of the PHD finger, consisting of a short double-stranded antiparallel β sheet, two α-helical turns, and several loops connecting two zinc-binding clusters (Figure 3A). The histone H3K4me2 peptide is bound in an extended conformation in a branched negatively charged channel (Figure 3B). The peptide is aligned with β1 of the PHD finger, allowing for the formation of a three-stranded antiparallel β sheet (Figure 3A).

The first six residues of the H3K4me2 peptide, from A1 to T6, are in direct contact with the PHF20 PHD finger. The methyl group of A1 lies in a deep cavity lined with Y692 of the protein, whereas the amino terminus of A1 is located within a short distance to the backbone carbonyl oxygen atoms of P689 and E690. The guanidino group of Arg2 is in close proximity to the carboxyl group of E651. An overlay of the structures of the

PHF20 PHD finger in complex with the H3K4me2 peptide and in the apo-state reveals that the binding is accompanied by a large conformational change in the protein N-terminal loop where E651 resides (Figure 3C). Upon complex formation, E651 swings toward the core of the PHD finger, most likely forming electrostatic and/or hydrogen bonding contacts with the positively charged R2 of the H3K4me2 peptide and thereby stabilizing the complex.

The fully extended side chain of K4me2 occupies an elongated groove, framed by V653, E662, M666, and W675. The side chains of E662 and W675, which are oriented perpendicularly to the protein surface, create the walls of the groove, whereas V653 and M666 line the bottom. To test the idea that the specificity of the PHF20 PHD finger for H3K4me2 over H3K4me3 is due to a smaller size of the methyllysine binding site, we superimposed the K4me2-binding groove of PHF20 with the K4me3-binding pocket of the inhibitor of growth, member 2 (ING2) PHD finger (Peña et al., 2006) and computed solvent-accessible surface areas (Figure 3D). We found that the K4me2-binding groove in PHF20 is slightly larger (~273 Å) compared with the K4me3-binding pocket of ING2 (~200 Å). These data imply that the size of the binding pocket is not a factor responsible for the preference of PHF20 for the dimethyllysine mark.

## The Presence of a Glutamate, but Not an Aspartate, Defines the Specificity for H3K4me2

The current predominant view that helps explain specificity of some methyllysine-recognizing readers, such as Tudor and MBT domains, for mono- or dimethylated lysine sequences is the presence of a negatively charged aspartate or glutamate residue in the methyllysine binding pocket (Li et al., 2007; Musselman et al., 2012; Taverna et al., 2007). However, this view is not pertinent to PHD fingers because the PHD fingers of MLL5, TAF3, and Pygo have an aspartate residue (D128, D877, and D352, respectively) in the methyllysine-binding cage (Figures 4A and 4B), yet MLL5 and TAF3 prefer H3K4me3 over H3K4me2 by ~5- to 10-fold (Ali et al., 2013; van Ingen et al., 2008; Vermeulen et al., 2007), and Pygo binds both marks with the same affinity of 2.4–2.5 μM (Fiedler et al., 2008). Because the PHF20 PHD finger contains the glutamate residue (E662) instead, we tested whether the longer glutamic side chain accounts for the specificity toward H3K4me2. We generated the E662D mutant of the PHD finger and examined its binding to H3K4me2 and H3K4me3 peptides by NMR and tryptophan fluorescence (Figures 4C and 4D). We found that the E662D mutant of PHF20-PHD interacts with both peptides equally well, exhibiting 3.6 μM affinity for H3K4me2 and 3.7 μM affinity for H3K4me3. These results suggest that the shorter aspartic side chain in PHF20-PHD E662D does not discriminate between the tri- and dimethylation states of H3K4, and a longer glutamic side chain is indeed necessary to select for H3K4me2.

Although we cannot validate hydrogen bond formation in NMR-derived structures, a short distance between the dimethylammonium group of H3K4 and the carboxylic group of E662 suggests that these groups interact through hydrogen bonding and ionic contacts. Unlike the trimethylammonium group of H3K4 in the MLL5 and TAF3 complexes that is located roughly at equal distance between the aromatic tryptophan residue and the aspartic acid (~4.7 and ~4.0 Å, respectively, in the MLL5 complex) (Figure 4B), the position of the dimethylammonium group of H3K4 in the PHF20 complex is shifted toward E662 (Ali et al., 2013; Lemak et al., 2013; van Ingen et al., 2008). A similar shift toward the glutamate residue is observed in the crystal structure of the H3K4me2-bound Y17E mutant of the PHD finger of bromodomain PHD finger transcription factor (BPTF) (Li et al., 2007). This mutant prefers H3K4me2 as opposed to wild-type (WT) BPTF that selects for H3K4me3 (Li et al., 2007) and, in support of the binding mode of PHF20, the distances between K4me2 and W32 and E17 in the Y17E BPTF complex are 4.5 and 2.7 Å, respectively (Figure 4B). These structural and biochemical analyses demonstrate that the glutamate residue present in the K4me2-binding pocket is the major determinant of the low-methylation state specificity of the PHF20 PHD finger.

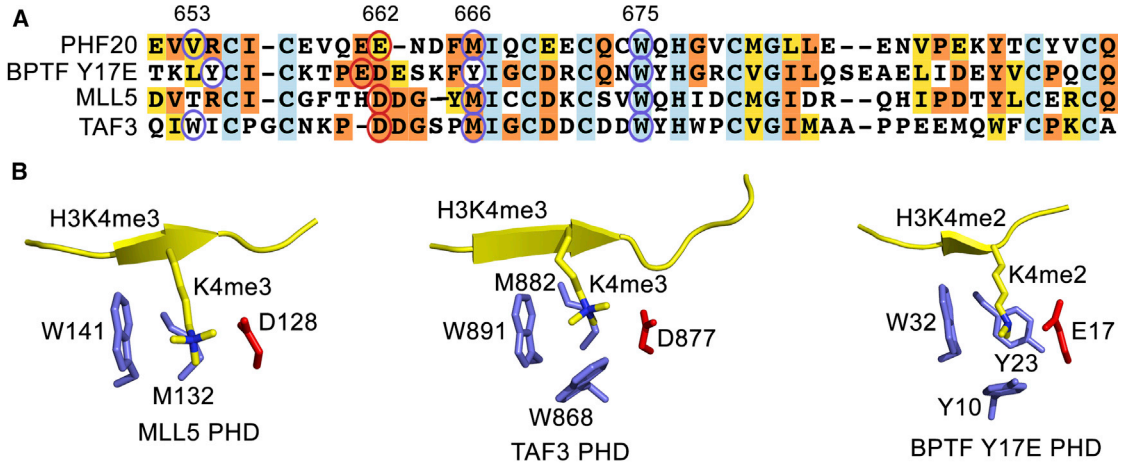
## Mutations in the PHD Finger Decrease Binding to H3K4me2

A number of cancer-associated mutations and deletions have been identified in *PHF20*, several of which occur in the PHD finger (Catalogue of Somatic Mutations in Cancer [COSMIC]; Figures 4F and 4G). To determine whether the cancer-relevant mutations affect binding of the PHD finger to H3K4me2, we produced R654H, F665V, M666I, and P689L mutants of PHF20 found in stomach, breast, kidney, and lung cancers, and we tested the corresponding $^{15}$N-labeled proteins by NMR (Figure 4H). We found that binding of R654H, F665V, and P689L to H3K4me2, compared with WT, was decreased ~8-, ~9-, and ~17-fold, respectively, whereas M666I was unstructured. Mapping the cancer-relevant mutations on the structure of the PHD-H3K4me2 complex revealed that some of the mutated residues can be essential for proper folding of the domain (such as the hydrophobic core residue M666 or a zinc-coordinating residue C680), or they make contacts with the histone peptide (such as F665 and P689). Overall, these data demonstrate that some cancer-relevant mutants have compromised histone-binding activity and suggest a relationship between this function of PHF20 and certain cancers.

Substitution of residues in the aromatic dimethyllysine-binding pocket of the PHF20 PHD finger, including W675 to an alanine and F665 to a lysine, led to protein misfolding, pointing to the importance of these aromatic residues for structural stability of the PHD finger (Figure 4C). Mutation of E662 to a lysine resulted in a folded, soluble, and stable protein, but completely abrogated interaction with both H3K4me2 and H3K4me3, allowing us to utilize this point mutation in characterizing the functional significance of the PHD-H3K4me2 interaction in full-length PHF20 in vivo (Figures 4C and 4E).

## Binding of the PHF20 PHD Finger to H3K4me2 Is Required for MOF-Dependent H4K16 Acetylation and Gene Regulation

Because PHF20 has been implicated in transcription (Badeaux et al., 2012; Park et al., 2012; Zhang et al., 2013), we explored whether the histone-binding activity of the PHD finger is necessary for PHF20/MOF-dependent gene expression. We first randomly selected 16 downregulated genes in PHF20 KD cells (see Figure 1C) and performed qPCR analysis to validate the



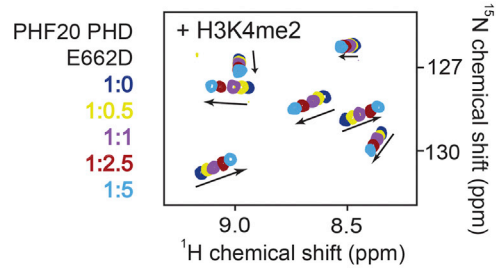


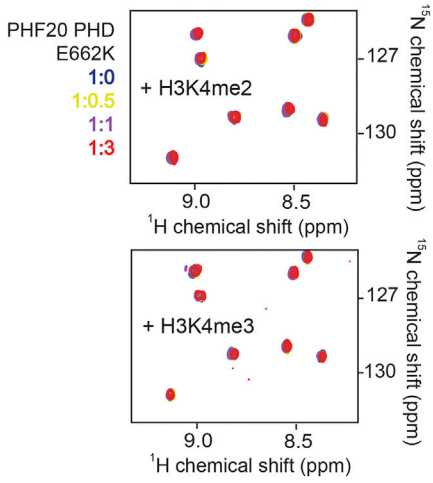


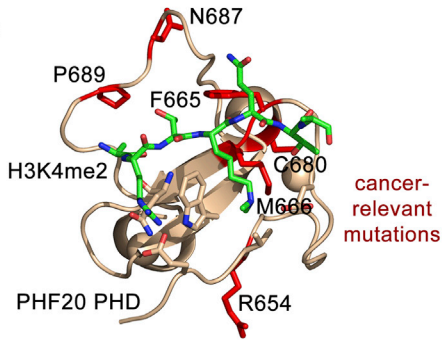
C

Binding affinity of the PHF20 PHD finger

| PHF20 PHD | Peptide | $K_d$ (μM) |
|---|---|---|
| WT | H3K4me2 | 3.6 ± 0.6 |
| E662K | H3K4me2 | NB |
| E662K | H3K4me3 | NB |
| E662D | H3K4me2 | 3.6 ± 1.8 |
| E662D | H3K4me3 | 3.7 ± 0.3 |
| E662Y | n/a | insoluble |
| F665K | n/a | misfolded |
| C674S | H3K4me2 | as WT |
| W675A | n/a | misfolded |

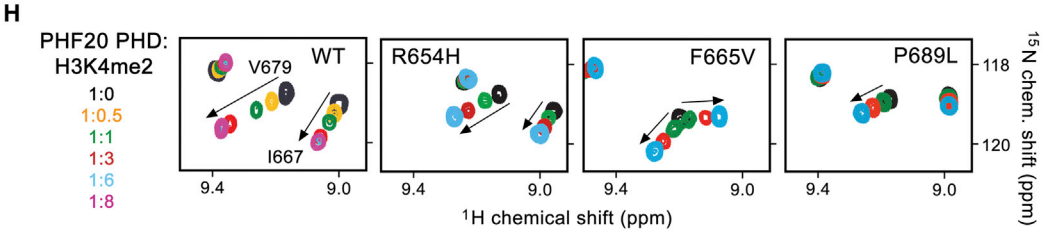



G

| PHF20 PHD | Binding to H3K4me2 relative to WT |
|---|---|
| R654H | ~8-fold weaker |
| F665V | ~9-fold weaker |
| M666I | unfolded |
| C680F | N/A |
| P689L | ~17-fold weaker |




(legend on next page)

RNA-seq results. All the genes tested showed considerable reductions of gene expression upon PHF20 KD by two independent shRNAs (Figure 5A). Concomitantly, chromatin immunoprecipitation (ChIP) analysis revealed a substantial drop in H4K16ac levels in promoters of these genes, whereas the total H3 levels remained unchanged (Figures 5B and S3A). Notably, all the genes tested have strong H3K4me2 levels on their promoters (Figure S3B). These findings suggest that PHF20 regulates gene expression likely through MOF-mediated H4K16ac. To characterize the role of the PHD finger in this regulation, we designed a reconstitution system in which shRNA-resistant WT PHF20 or E662K mutant, impaired in binding to H3K4me2, was reintroduced into the shPHF20 cells. As shown in Figure 5C, KD of PHF20 led to reduced H4K16ac levels; importantly, WT PHF20, but not the E662K mutant, rescued the global reduction of H4K16ac levels. This was further substantiated by measuring H4K16ac levels in the promoters of PHF20 target genes, such as *CCNA1* and *HMGB2*. PHF20 KD resulted in decreased H4K16ac levels, which were fully restored by WT PHF20, but were not restored by the E662K mutant (Figures 5D and S3C). Subsequently, WT PHF20 rescued the defects in target gene expression, cell proliferation, and colony formation of the PHF20 KD H1792 cells, whereas the PHD finger mutant E662K, which is incapable of binding to H3K4me2, failed to do so (Figures 5E–5H and S3D). Collectively, these results suggest that the H3K4me2-binding function of the PHD finger is essential for PHF20-dependent histone acetylation, target gene activation, and cancer cell growth and survival.

## Functional Link between Tudor2 and PHD of PHF20

Along with the PHD finger, PHF20 contains another methyllysine reader, the tandem Tudor domains, in which Tudor2 recognizes dimethyllysine substrates, such as p53K370me2 and p53K382me2 (Adams-Cioaba et al., 2012; Cui et al., 2012; Kim et al., 2006). To determine whether Tudor2 competes with the PHD finger for the same posttranslational modification (PTM), we generated $^{15}$N-labeled Tudor2 and examined its interactions with H3K4me2 and p53K382me2 by $^{1}$H,$^{15}$N HSQC (Figure 6). Substantial CSPs in PHF20 Tudor2 were observed upon addition of the p53K382me2 peptide, and in agreement with previous reports (Cui et al., 2012), $K_D$ for this interaction was found to be ~100 μM (Figures 6A and 6C). In contrast, H3K4me2 peptide induced small CSPs in Tudor2 and was bound much weaker, with affinity of ~1 mM. To test whether the PHD finger can compete with Tudor2 for binding to dimethylated p53, we next titrated the p52K382me2 peptide into the $^{15}$N-labeled PHD finger (Figure 6B). Lack of any CSPs in the spectrum demonstrated that the PHF20 PHD finger does not recognize methylated p53. Despite some similarities in the amino acid (aa) sequences of dimethylated H3K4 and p53 (Tudor2 associates with p53K370me2 and p53K382me2 almost equally well; Cui et al., 2012), superimposition of the H3K4me2 peptide with the p53K370me2 peptide bound to Tudor2 provides a possible explanation for the inability of Tudor2 to recognize H3K4me2 (Figure 6D). Most likely, the positively charged N terminus of H3A1 and the side chain of H3R2 would be repelled from the positively charged surface of Tudor2, where p53S367 and p53H368 are bound, and the neutral H3Q5-T6 sequence lacks the capacity to electrostatically interact with the negatively charged surface of Tudor2, where p53K372 and p53K373 are bound. In contrast, the absence of the AR sequence N-terminal to dimethyllysine likely precludes binding of the PHF20 PHD finger to dimethylated p53. Together, these results demonstrate that, although both reader domains in PHF20 exhibit preference for dimethyllysine species, their functions do not overlap.

## Conclusions

Both histone marks, H4K16ac and H3K4me, are vital epigenetic PTMs linked to transcriptional regulation. Accumulating evidence suggests that these modifications and the enzymatic complexes that produce them, such as MOF-NSL and MLL1, function in a highly cooperative manner (Dou et al., 2005; Zhao et al., 2013b). Although the mechanism underlying this close correlation is poorly understood, it has been suggested that WDR5, a component of both MOF-NSL and MLL1 complexes, plays a role (Dou et al., 2005; Li and Dou, 2010). Structural analysis of the WD40 domain of WDR5, however, reveals that this protein associates with the H3 N terminus regardless of the methylation state of H3K4 (Couture et al., 2006; Ruthenburg et al., 2006; Schuetz et al., 2006), leaving the question of its possible role in functional cooperation of the enzymatic activities of the two complexes open. In this study, we identified a relationship between the H4K16ac- and H3K4me-generating complexes. We show that the PHD finger of PHF20, a core subunit of the MOF-NSL HAT complex responsible for writing H4K16ac, recognizes H3K4me2, a mark generated by the MLL1 methyltransferase complex (Milne et al., 2002) (Figure 6E). Binding of the PHD finger to H3K4me2 is required for PHF20-dependent histone acetylation, occupancy of PHF20 at target genes, and transcriptional activation of these genes. Our results indicate that PHF20 links MOF to H3K4me2-mediated cellular processes and suggest a model for rapid propagation of H4K16ac-enriched

**Figure 4. Molecular Basis for the Selectivity toward Dimethyllysine**

(A) Alignment of the PHD finger sequences: absolutely, moderately, and weakly conserved residues are colored light blue, orange, and yellow, respectively. The H3K4me3-binding site residues of MLL5 and TAF3, as well as the H3K4me2-binding site residues of PHF20 and the Y17E mutant of BPTF, are circled.

(B) A close view of the Kme-binding sites of the PHD fingers of MLL5, TAF3, and the Y17E mutant of BPTF. The histone peptide is yellow. PDB codes are 4L58, 2K17, and 2RI7.

(C) Binding affinities of mutated PHF20 PHD for the indicated histone peptides as measured by tryptophan fluorescence.

(D and E) Overlays of $^{1}$H,$^{15}$N HSQC spectra of the mutated PHF20 PHD fingers collected upon titration with the indicated peptides.

(F) Residues of the PHF20 PHD finger, found mutated in cancer, are shown as sticks in red and labeled.

(G) Binding affinities of the cancer-relevant mutants of the PHF20 PHD finger for H3K4me2 relative to the binding affinity of the WT PHD finger as measured by NMR.

(H) Overlays of $^{1}$H,$^{15}$N HSQC spectra of the cancer-relevant mutants of the PHF20 PHD finger collected upon titration with the H3K4me2 peptide.

NB, no binding.

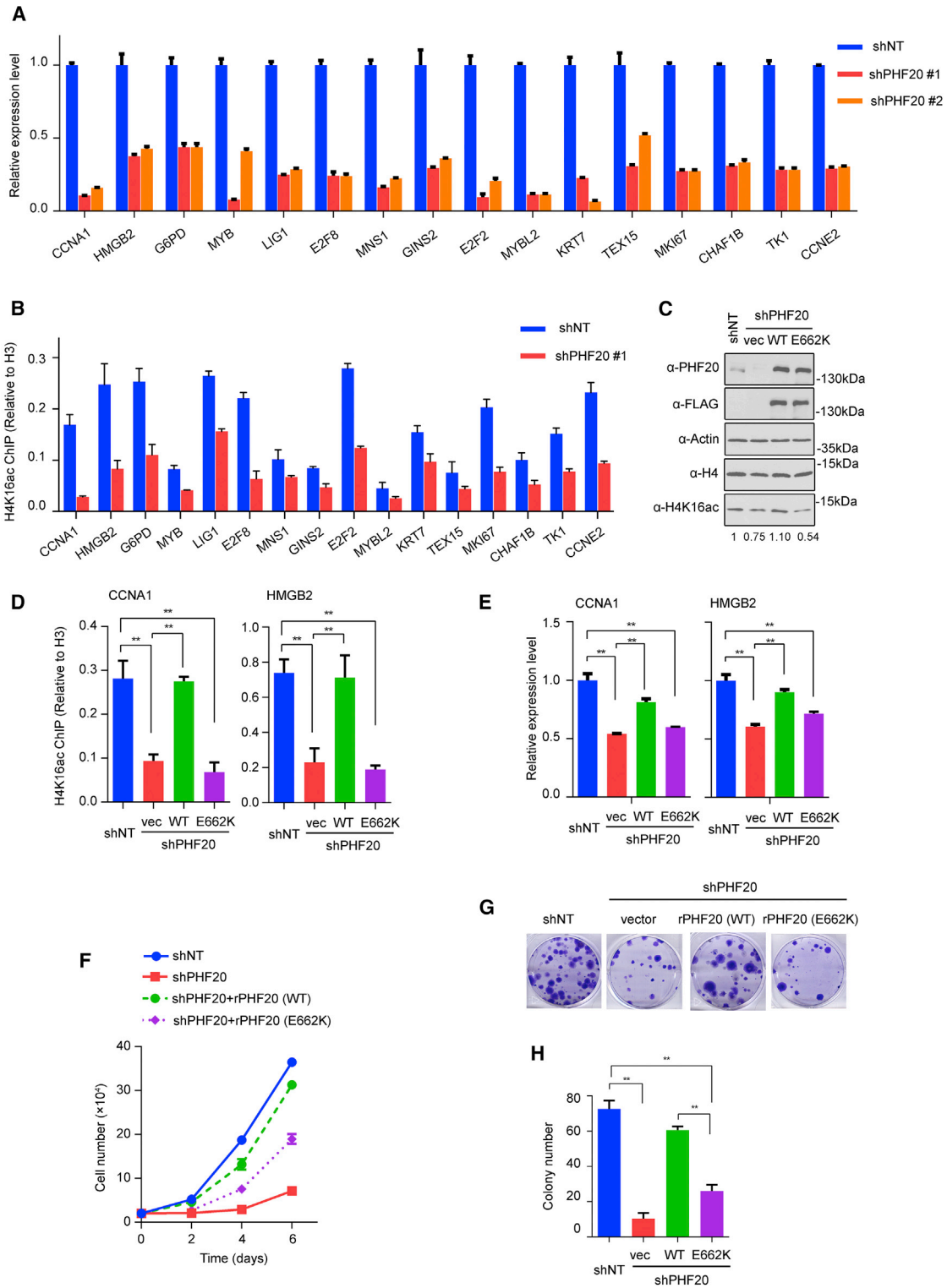
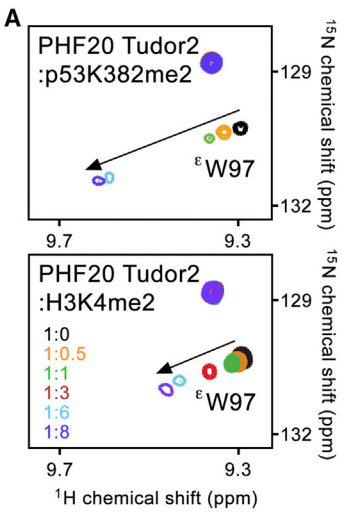


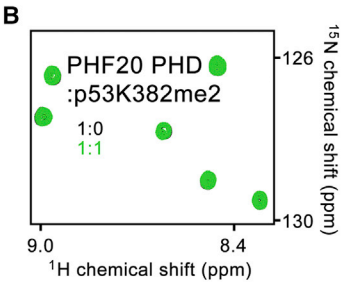

**Figure 5. H3K4me2-Binding Activity of PHD Is Essential for PHF20 to Regulate Gene Expression and Cancer Cell Growth and Survival**

(A) qRT-PCR analysis of gene expression in control (shNT) and PHF20 KD (shPHF20) cells.

(B) qRT-PCR analysis of H4K16ac ChIP on gene promoters in control (shNT) and PHF20 KD (shPHF20) cells. (A and B) Error bars represent SEM of three biological replicates.

*(legend continued on next page)*

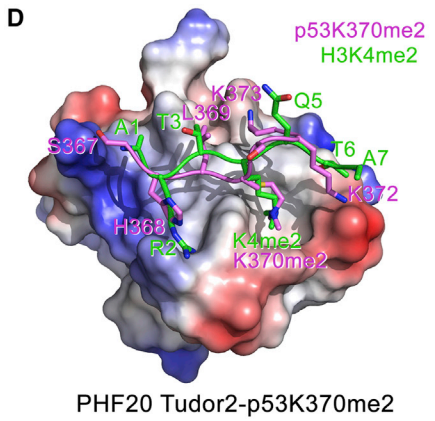



C Binding affinities of the PHF20 domains (μM)

| Peptide | PHD | Tudor2 |
|---|---|---|
| H3K4me2 | 3.6 ± 0.6 | ~1,000 |
| p53K382me2 | no binding | ~100 |

p53K382 RHKK(me)LMF
p53K370 SHLK(me)SKK
H3K4 ARTK(me)QTA



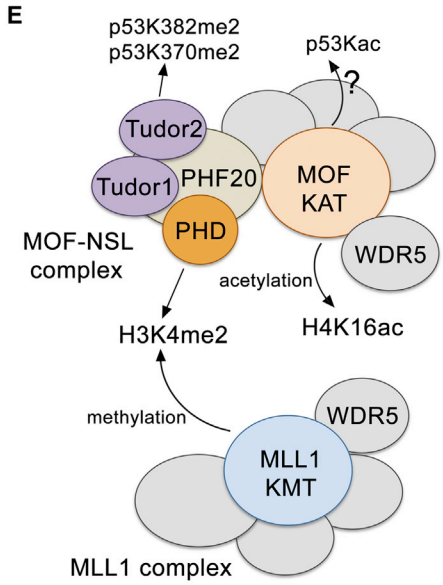




**Figure 6. Functional Coupling between Tudor2 and PHD of PHF20**

(A and B) Superimposed $^{1}H$,$^{15}N$ HSQC spectra of the Tudor2 and PHD domains of PHF20 collected upon titration with the indicated peptides. Spectra are color coded according to the protein:peptide molar ratio (inset).

(C) Binding affinities of PHD and Tudor2 and alignment of the p53 and H3 sequences. Basic residues are in blue.

(D) The electrostatic surface potential of the PHF20 Tudor2 domain bound to the p53K370me2 peptide is shown (PDB 2LDM). Blue and red colors represent positive and negative charges, respectively. The p53K370me2 peptide (magenta) is manually superimposed with the histone H3K4me2 peptide (green) derived from the PHF20 PHD-H3K4me2 complex.

(E) A model depicting a potential role of PHF20 readers in linking H3K4 methylation with H4K16 acetylation as well as p53 methylation and acetylation.

open chromatin. Because the MOF complex contains several subunits capable of binding to histones or DNA, including H3-recognizing WDR5 and putative DNA-binding domains in the NSL2 subunit, we expect that multiple contacts contribute and/or fine-tune the MOF recruitment to specific chromatin loci. Further studies are needed to investigate the interplay of the histone- and DNA-binding activities within the MOF complex.

Besides being a subunit of NSL, MOF is present in another evolutionarily conserved HAT complex, MSL (Li and Dou, 2010; Li et al., 2009; Zhao et al., 2013b). Although both NSL and MSL complexes efficiently acetylate H4K16, only NSL is able to acetylate non-histone proteins, including p53 (Li et al., 2009; Rea et al., 2007). Because the NSL-MOF complex, but not the MSL-MOF complex, contains the PHF20 subunit (Zhao et al., 2013b) whose Tudor2 binds to dimethylated p53 (Cui et al., 2012), it is tempting to suggest that PHF20 might have a role in promoting p53 acetylation by MOF (Figure 6E). Another possibility is that the distinctly different dimethyllysine-binding activities of Tudor2 and PHD in PHF20 tether or stabilize p53 at the genomic sites enriched in H3K4me2. It will be interesting in future studies to explore whether this functional coupling involving the two PHF20 readers exists. It is also essential to establish the role of PHF20 in regulating MOF activity on p53 as well as in p53-mediated DNA damage response, cell cycle regulation, and apoptosis.

## EXPERIMENTAL PROCEDURES

### Materials

shRNA constructs were purchased from Sigma. The PHF20-targeting shRNA sequences were 5′-CCCGAGAAATACACCTGTTAT-3′ and 5′-ATTGTGC CACTGATGATAAAC-3′. The following antibodies were purchased: anti-PHF20 (3934S, 1:1,000) antibody from Cell Signaling, anti-FLAG (F-1804,

(C) Western blot analysis of PHF20 and H4K16ac levels in control (shNT), PHF20 KD (shPHF20), and KD cells rescued with the WT PHF20 or E662K mutant. Total H4 and actin were used as loading controls. The density of H4K16ac bands was quantified using ImageJ.

(D) qRT-PCR analysis of H4K16ac ChIP on promoters of CCNA1 and HMGB2 genes in cells as in (C).

(E) qRT-PCR analysis of CCNA1 and HMGB2 gene expression in cells as in (C).

(F) Cell proliferation assays of cells as in (C).

(G and H) Colony formation assays of cells as in (C). Representative crystal-violet-stained cells are shown in (G), and quantification of six replicates is shown in (H).

**$p < 0.01$. See also Figure S3 and Tables S4 and S5.

1:5,000) and anti-tubulin (T8328, 1:5,000) antibodies from Sigma, anti-H4K16ac (07-329, 1:1,000) antibody from Millipore, and anti-total H3 (ab1791, 1:5,000) antibody from Abcam.

### Cloning and Protein Purification

The PHF20 PHD domain construct (aa 651–699) was cloned from cDNA of PHF20 (Open Biosystems) into a pGEX-6P-1 (GE Healthcare) expression vector with ampicillin resistance. The WT and mutant proteins were expressed in *E. coli* Rosetta-2(DE3)pLysS cells grown in either Luria broth or $^{15}NH_4Cl$ minimal media, supplemented with 150 μM $ZnCl_2$. After induction with isopropyl-ß-D-1-thiogalactopyranoside (IPTG; 0.5 mM) for 16–18 hr at 18°C, cells were harvested by centrifugation at 5,000 rpm and lysed by sonication. GST fusion proteins were purified on glutathione Sepharose 4B beads (Pierce, Fisher). The GST tag was cleaved with PreScission protease (Amersham). When necessary, the proteins were further purified by size-exclusion chromatography over a HiPrep 16/60 Sephacryl S-100 column (GE Healthcare) and concentrated in Millipore concentrators (Millipore). Tudor2 (aa 58–148) was purified as described previously (Zhang et al., 2016).

For NMR structure determination of the PHD finger in complex with an H3K4me2 peptide, a DNA fragment encoding PHF20 residues 646–699 was inserted in a pTEV vector encoding a tobacco etch virus (TEV) protease-cleavable N-terminal hexahistidine tag. The protein was overexpressed in *E. coli* Rosetta (DE3) pLysS cells cultured in $^{15}N$- and $^{15}N/^{13}C$-enriched M9 media supplemented with 100 μM $ZnSO_4$. Cells were grown at 37°C to an $OD_{600}$ (optical density measured at a wavelength of 600 nm) of about 0.6, transferred to 15°C, induced with IPTG (0.5 mM) after 45 min, and cultured for another 16–20 hr before harvesting by centrifugation. Cell pellets were resuspended in 50 mM sodium phosphate (pH 7.5), 300 mM NaCl with 5 mM imidazole, and 5 mM PMSF and then lysed using an Emulsiflex C-5 high-pressure homogenizer. Cell debris was removed by centrifugation. The protein was purified by metal chelation chromatography using $Ni^{2+}$-NTA resin and further purified by size-exclusion chromatography using a Superdex 75 column after cleavage of the hexahistidine tag with TEV protease overnight at 4°C. Protein concentration was determined by UV absorption at 280 nm with the absorption coefficient predicted by ProtParam (http://www.expasy.org/).

### PCR Mutagenesis

Point mutants of the PHF20 PHD finger were generated using the Stratagene QuikChange XL Site Directed Mutagenesis Kit and Pfu turbo DNA polymerase (Stratagene). All constructs were confirmed by DNA sequencing.

### Peptide Microarray and Peptide Pull-Down Assay

Peptide microarray and peptide pull-down assays were performed as described previously (Wen et al., 2014). In brief, biotinylated histone peptides were printed in triplicate onto a streptavidin-coated slide (PolyAn) using a VersArray Compact Microarrayer (Bio-Rad). After a short blocking with biotin (Sigma), the slides were incubated with the GST-PHF20 PHD in binding buffer (50 mM Tris-HCl [pH 7.5], 250 mM NaCl, 0.1% Nonidet P-40 [NP-40], 1 mM PMSF, 20% fetal bovine serum) overnight at 4°C with gentle agitation. After being washed with the same buffer, the slides were probed with an anti-GST primary antibody and then a fluorescein-conjugated secondary antibody and visualized using a GenePix 4000 scanner (Molecular Devices). For the peptide pull-down assays, 1 μg biotinylated histone peptides with different modifications was incubated with 1 μg GST-PHF20 PHD in binding buffer (50 mM Tris-HCl [pH 7.5], 250 mM NaCl, 0.1% NP-40, 1 mM PMSF) overnight. Streptavidin beads (Amersham) were added to the mixture, and the mixture was incubated for 1 hr with rotation. The beads were then washed three times and analyzed using SDS-PAGE and western blotting.

### X-Ray Crystallography

PHF20 PHD C674S (aa 651–699) was concentrated to ~16 mg/ml in 20 mM Tris-HCl (pH 7.5), supplemented with 130 mM NaCl and 5 mM dithiothreitol. The protein solution was incubated overnight with the H3K4me2 peptide (aa 1–12) in a 1:1.5 molar ratio before crystallization. Crystals were grown using the sitting-drop diffusion method at 4°C by mixing 600 nL protein-peptide solution with 600 nL well solution composed of 0.1 M bis-tris propanol (pH 6.5), 2.55 M ammonium sulfate titrated with 6 μL/1 mL 1 N HCl. Although the histone peptide was present, PHF20 PHD C674S was crystallized in an apo-state. Cryoprotected crystals (soaked in mother liquor with ~30% glycerol) were picked and plunged into liquid nitrogen. X-ray diffraction data were collected from a single crystal at National Synchrotron Light Source X25 (BNL) beamline. The protein structure was solved using Single-wavelength Anomalous Dispersion method with Zn anomalous signal. Datasets were processed with HKL2000 and CCP4. Refinement was carried out using Phenix (Adams et al., 2010), and the model was built with Coot (Emsley and Cowtan, 2004). Data and refinement statistics are summarized in Table S2.

### NMR Titration Experiments

The $^{1}H,^{15}N$ HSQC spectra of 0.1 mM uniformly $^{15}N$-labeled WT or mutant PHD domain of PHF20 in PBS (pH 6.7), 5 mM dithiothreitol buffer or Tudor2 in 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM dithiothreitol buffer were collected at 298K on Varian INOVA 600- and 500-MHz spectrometers as described previously (Klein et al., 2014). The binding was characterized by monitoring chemical shift changes as 12-mer histone tail peptides (synthesized by the University of Colorado Denver Peptide Core Facility) or p53 (aa 377–387) was added stepwise. The $K_D$s were determined using a nonlinear least-squares analysis in KaleidaGraph and the equation

$$\Delta\delta = \Delta\delta_{max}\left(([L]+[P]+K_d) - \sqrt{([L]+[P]+K_d)^2 - 4[P][L]}\right)\Big/2[P],$$

where [L] is concentration of the peptide, [P] is concentration of the protein, $\Delta\delta$ is the observed chemical shift change, and $\Delta\delta_{max}$ is the normalized chemical shift change at saturation. Normalized chemical shift changes were calculated using the equation $\Delta\delta = \sqrt{(\Delta\delta H)^2 + (\Delta\delta N/5)^2}$, where $\Delta\delta$ is the change in chemical shift in parts per million (ppm).

### NMR Structure Determination

NMR experiments were recorded at 25°C using a 700-MHz Bruker Avance III spectrometer equipped with a cryoprobe. The NMR data were processed and analyzed using NMRPipe (Delaglio et al., 1995) and NMRViewJ (Johnson and Blevins, 1994). Protein samples were in 25 mM sodium phosphate buffer (pH 7.0), 1.5 mM $NaN_3$, 0.3 mM sodium 4,4-dimethyl-4-silapentane-1-sulphonate, 10% $D_2O$, and 90% $H_2O$. The 11-residue C-terminally amidated H3K4me2 peptide [ART(Kme2)QTARKST] was purchased from GenScript and purified by reverse-phase high-performance liquid chromatography (HPLC) using a Jupiter 5u C18 300A preparative column (Phenomenex). For structure determination, PHF20 PHD and H3K4me2 peptide were at concentrations of 1.5 and 6 mM, respectively. A series of standard NMR experiments was recorded for backbone and side-chain resonance assignments including 2D correlation spectroscopy (COSY), $^{1}H,^{15}N$ and $^{1}H,^{13}C$ HSQC, and 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH, CCH-TOCSY, $^{15}N$ NOESY-HSQC, and $^{13}C$ NOESY-HSQC. $^{15}N$ NOESY-HSQC and $^{13}C$ NOESY-HSQC optimized for aliphatic and aromatic spectral regions were recorded with a mixing time of 120 ms to generate distance restraints. 3D $^{13}C,^{15}N$-filtered, $^{13}C/^{15}N$-edited NOESY experiments (Zwahlen et al., 1997) were recorded with a 120-ms mixing time to unambiguously identify intermolecular nuclear Overhauser enhancement (NOEs) (Figure S2).

For structure calculations, interproton distances were derived from $^{15}N$ NOESY-HSQC, $^{13}C$ NOESY-HSQC, and $^{13}C,^{15}N$-filtered, $^{13}C/^{15}N$-edited NOESY experiments and categorized into five distance ranges with upper limits of 2.8, 3.5, 4.5, 5.5, and 7.5 Å; the 7.5 Å is to account for possible spin diffusion in very weak NOE signals. Hydrogen bonds were determined from deuterium exchange experiments and given upper limits of 3.2 Å (for N-O) and 2.2 Å (for HN-O). For all distance restraints, a lower limit of 1.8 Å was used. Dihedral angles φ and ψ were derived from TALOS (Cornilescu et al., 1999) and CSI (Wishart and Sykes, 1994) analysis. Two hundred structures of free PHF20 were initially calculated using the simulated annealing protocol of CYANA (Güntert, 2004). Twenty of these structures with the lowest energies were then used to analyze for violations and obtain additional SANE-assisted NOE assignments for PHF20 (Duggan et al., 2001). Several iterations of CYANA and SANE were carried out until distance and angle violations fell

below 0.3 Å and 5°, respectively. Next, the above procedure was repeated but introducing the H3K4me2 peptide in the calculations. Two hundred structures of the PHF20-H3K4me2 complex were calculated, and 100 of those with the lowest energies were refined using AMBER (Case et al., 2005). In AMBER refinement, the structures were subjected to 30 ps 30,000 steps of simulated annealing using a generalized Born solvent model (Botuyan et al., 2001; Tsui and Case, 2000-2001). The system was first heated to 1,000 K in the first 8 ps and remained at 1,000 K for 2 ps. It was then annealed and cooled to 0 K for the next 20 ps. Force constants were set as follows: 20 kcal $mol^{-1}$ $Å^2$ for NOE-derived distance restraints, 40 kcal $mol^{-1}$ $Å^2$ for hydrogen bond restraints, 70 kcal $mol^{-1}$ $rad^{-2}$ for dihedral angle restraints, 100 kcal $mol^{-1}rad^{-2}$ for chirality restraints, and 150 kcal $mol^{-1}rad^{-2}$ for omega angle restraints. At the end of the refinement, 20 structures with the lowest AMBER energies and violations were selected for final structure statistics (Table S3).

### Fluorescence Spectroscopy

Spectra were recorded at room temperature on a Fluoromax-3 spectrofluorometer (HORIBA). The samples containing the PHF20 PHD domain in PBS (pH 6.8), 10 mM DTT buffer, and progressively increasing concentrations of histone peptide were excited at 280 nm. Emission spectra were recorded over a range of wavelengths between 310 and 405 nm with a 0.5-nm step size and a 1-s integration time and averaged over three scans. The $K_D$ values were determined using a nonlinear least-squares analysis and the equation

$$\Delta I = \Delta I_{max}\left(([L]+[P]+K_d) - \sqrt{([L]+[P]+K_d)^2 - 4[P][L]}\right)\Big/2[P],$$

where [L] is the concentration of the histone peptide, [P] is the concentration of the protein, $\Delta I$ is the observed change of signal intensity, and $\Delta I_{max}$ is the difference in signal intensity of the free and bound states of the protein. The $K_D$ value was averaged over three separate experiments, with error calculated as the SD between the runs.

### Cell Culture and RNA Interference

H1792 and all lung cancer cell lines were cultured in RPMI 1640 (Cellgro) supplemented with 10% fetal bovine serum (Sigma). For shRNA knockdown (KD), 293T cells were cultured in DMEM (Cellgro) and co-transfected with pMD2.G and pPAX2 (Addgene) together with pLKO-shRNA constructs or a non-targeting pLKO-shRNA (shNT). Viral supernatants were harvested after 48 hr. For infections, H1792 cells were incubated with viral supernatants in the presence of 8 μg/ml polybrene. After 48 hr, puromycin (2 μg/ml) was added to the medium. Cells were grown, with selection for 5–7 days to select for stable KD cells.

### Colony Formation Assays

H1792 cells treated with shRNAs were seeded in six-well plates (800 cells/well) and grown in RPMI 1640 plus 10% FBS at 37°C for 14 days. Cells were fixed, stained with 0.005% crystal violet blue, and photographed. Colony numbers were counted using ImageJ software with size cutoff of 75 μm. Results were quantitated from at least three independent replicates.

### ChIP

ChIP analysis was performed essentially as described previously (Wen et al., 2014). In brief, cells were crosslinked with 1% formaldehyde for 10 min at room temperature, and the reaction was stopped with 125 mM glycine. Nuclei were isolated by resuspending the cells in swelling buffer containing 5 mM piperazine-N,N'-bis(ethanesulfonic acid) (PIPES; pH 8.0), 85 mM KCl, 1% NP-40, and complete protease inhibitors for 20 min at 4°C. The isolated nuclei were resuspended in nuclei lysis buffer (50 mM Tris [pH 8.0], 10 mM EDTA, 1% SDS) and sonicated using a Bioruptor Sonicator (Diagenode). Samples were immunoprecipitated with 2–4 μg of the appropriate antibodies overnight at 4°C. Immunoprecipitates were washed twice with dialysis buffer (50 mM Tris [pH 8.0], 2 mM EDTA, 0.2% Sarkosyl) and four times with IP wash buffer (100 mM Tris [pH 8.0], 500 mM LiCl, 1% NP-40, and 1% deoxycholic acid sodium salt). After reverse crosslinking was performed, the DNA was eluted and purified using a PCR purification kit (QIAGENe). The primer sequences used for ChIP analyses are listed in Table S4.

### RT-PCR, Real-Time PCR, and RNA-Seq Analysis

RT-PCR and real-time PCR were performed as described previously (Wen et al., 2014). mRNA was prepared using the RNeasy Plus kit (QIAGEN) and reverse-transcribed using the First Strand Synthesis Kit (Invitrogen). Quantitative real-time RT-PCR was performed on an ABI 7500-FAST Sequence Detection System using the Power SYBR Green PCR Master Mix (Applied Biosystems). Gene expression was calculated after normalization to GAPDH levels using the comparative Ct (cycle threshold) method. The primer sequences used for RT-PCR are listed in Table S5. Experimental data are presented as means ± SD unless stated otherwise. Statistical significance was calculated by two-tailed unpaired t test on two experimental conditions with $p < 0.05$ considered statistically significant.

RNA-seq samples were sequenced using the Illumina Hiseq 2000. Three biological replicates were prepared for each condition. The sequencing reads were mapped to human genome (hg19) by TopHat (version 2.0.10) (Kim et al., 2013). The overall mapping rate is 95%–99%. The number of fragments in each known gene from RefSeq database (Pruitt et al., 2012) (downloaded from UCSC Genome Browser on June 2, 2014) was enumerated using htseq-count from HTSeq package (version 0.6.0) (http://www-huber.embl.de/users/anders/HTSeq/). The differential expression between conditions was statistically assessed by R/Bioconductor package DESeq (Anders and Huber, 2010) (version 1.18.0). Genes with false discovery rate $\leq$ 0.05 and fold change $\geq$ 2 were called significant. Hierarchical clustering was performed by hclust function in R using their expression values estimated by DESeq. The expression values of each gene across samples were centered by median and scaled by SD before clustering. Euclidean distance and ward.D2 clustering method were used. The heatmap was plotted by heatmap.2 function in R.

## ACCESSION NUMBERS

The accession numbers for the atomic coordinates for the structures reported in this paper are PDB: 5TAB and 5TBN. The accession number for the RNA-seq data reported in this paper is GEO: GSE82115.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.09.056.

## AUTHOR CONTRIBUTIONS

B.J.K., X.W., G.C., C.Y., M.V.B., K.L., Y.L., X.W., and Y.Z. performed experiments and together with C.J.B., G.M., X.S., and T.G.K. analyzed the data. B.J.K., G.M., X.S., and T.G.K. wrote the manuscript with input from all authors.

## ACKNOWLEDGMENTS

We thank the MD Anderson Science Park Next-Generation Sequencing Facility (grant RP120348) for sequencing. This work is supported by NIH grants GM100907, GM106416, and GM101664 (T.G.K.), CA204020-01 (X.S.), and CA132878 and GM116829 (G.M.); CPRIT grant RP140323 (X.S.); Welch grant G1719 (X.S.); American Cancer Society grant RSG-13-290-01-TBE (X.S.); and a Mayo Clinic Cancer Center Fraternal Order of Eagles Fellowship Award (G.C.). X.W. was supported by a Ludwig Maximilian University-China Scholarship Council scholarship. X.S. is a recipient of a Leukaemia & Lymphoma Society Career Development Award and is a R. Lee Clark Fellow and Faculty Scholar of MD Anderson Cancer Center.

Received: June 13, 2016
Revised: September 5, 2016
Accepted: September 16, 2016
Published: October 18, 2016

## REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010).

PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Adams-Cioaba, M.A., Li, Z., Tempel, W., Guo, Y., Bian, C., Li, Y., Lam, R., and Min, J. (2012). Crystal structures of the Tudor domains of human PHF20 reveal novel structural variations on the Royal Family of proteins. FEBS Lett. *586*, 859–865.

Ali, M., Rincón-Arano, H., Zhao, W., Rothbart, S.B., Tong, Q., Parkhurst, S.M., Strahl, B.D., Deng, L.W., Groudine, M., and Kutateladze, T.G. (2013). Molecular basis for chromatin binding and regulation of MLL5. Proc. Natl. Acad. Sci. USA *110*, 11296–11301.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Avvakumov, N., and Côté, J. (2007). The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene *26*, 5395–5407.

Badeaux, A.I., Yang, Y., Cardenas, K., Vemulapalli, V., Chen, K., Kusewitt, D., Richie, E., Li, W., and Bedford, M.T. (2012). Loss of the methyl lysine effector protein PHF20 impacts the expression of genes regulated by the lysine acetyltransferase MOF. J. Biol. Chem. *287*, 429–437.

Bankovic, J., Stojsic, J., Jovanovic, D., Andjelkovic, T., Milinkovic, V., Ruzdijic, S., and Tanic, N. (2010). Identification of genes associated with non-small-cell lung cancer promotion and progression. Lung Cancer *67*, 151–159.

Botuyan, M.V., Mer, G., Yi, G.S., Koth, C.M., Case, D.A., Edwards, A.M., Chazin, W.J., and Arrowsmith, C.H. (2001). Solution structure and dynamics of yeast elongin C in complex with a von Hippel-Lindau peptide. J. Mol. Biol. *312*, 177–186.

Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J. Biol. Chem. *285*, 4268–4272.

Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J. Comput. Chem. *26*, 1668–1688.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Couture, J.F., Collazo, E., and Trievel, R.C. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. Nat. Struct. Mol. Biol. *13*, 698–703.

Cui, G., Park, S., Badeaux, A.I., Kim, D., Lee, J., Thompson, J.R., Yan, F., Kaneko, S., Yuan, Z., Botuyan, M.V., et al. (2012). PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. Nat. Struct. Mol. Biol. *19*, 916–924.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873–885.

Duggan, B.M., Legge, G.B., Dyson, H.J., and Wright, P.E. (2001). SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. J. Biomol. NMR *19*, 321–329.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fiedler, M., Sánchez-Barrena, M.J., Nekrasov, M., Mieszczanek, J., Rybin, V., Müller, J., Evans, P., and Bienz, M. (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. Mol. Cell *30*, 507–518.

Fischer, U., Struss, A.K., Hemmer, D., Pallasch, C.P., Steudel, W.I., and Meese, E. (2001). Glioma-expressed antigen 2 (GLEA2): a novel protein that can elicit immune responses in glioblastoma patients and some controls. Clin. Exp. Immunol. *126*, 206–213.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. *278*, 353–378.

Heisel, S.M., Ketter, R., Keller, A., Klein, V., Pallasch, C.P., Lenhof, H.P., and Meese, E. (2008). Increased seroreactivity to glioma-expressed antigen 2 in brain tumor patients under radiation. PLoS ONE *3*, e2164.

Johnson, B.A., and Blevins, R.A. (1994). NMR View: a computer program for the visualization and analysis of NMR data. J. Biomol. NMR *4*, 603–614.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep. *7*, 397–403.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Klein, B.J., Piao, L., Xi, Y., Rincon-Arano, H., Rothbart, S.B., Peng, D., Wen, H., Larson, C., Zhang, X., Zheng, X., et al. (2014). The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. Cell Rep. *6*, 325–335.

Lemak, A., Yee, A., Wu, H., Yap, D., Zeng, H., Dombrovski, L., Houliston, S., Aparicio, S., and Arrowsmith, C.H. (2013). Solution NMR structure and histone binding of the PHD domain of human MLL5. PLoS ONE *8*, e77020.

Li, X., and Dou, Y. (2010). New perspectives for the regulation of acetyltransferase MOF. Epigenetics *5*, 185–188.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91–95.

Li, H., Fischle, W., Wang, W., Duncan, E.M., Liang, L., Murakami-Ishibe, S., Allis, C.D., and Patel, D.J. (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. Mol. Cell *28*, 677–691.

Li, X., Wu, L., Corsa, C.A., Kunkel, S., and Dou, Y. (2009). Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. Mol. Cell *36*, 290–301.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol. Cell *10*, 1107–1117.

Musselman, C.A., and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res. *39*, 9061–9071.

Musselman, C.A., Lalonde, M.E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227.

Park, S., Kim, D., Dan, H.C., Chen, H., Testa, J.R., and Cheng, J.Q. (2012). Identification of Akt interaction protein PHF20/TZP that transcriptionally regulates p53. J. Biol. Chem. *287*, 11151–11163.

Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100–103.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. *40*, D130–D135.

Rea, S., Xouri, G., and Akhtar, A. (2007). Males absent on the first (MOF): from flies to humans. Oncogene *26*, 5385–5394.

Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J., and Verdine, G.L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat. Struct. Mol. Biol. *13*, 704–712.

Schuetz, A., Allali-Hassani, A., Martín, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H., and Min, J. (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. EMBO J. *25*, 4245–4252.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Sykes, S.M., Mellert, H.S., Holbert, M.A., Li, K., Marmorstein, R., Lane, W.S., and McMahon, S.B. (2006). Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Mol. Cell *24*, 841–851.

Taniwaki, M., Daigo, Y., Ishikawa, N., Takano, A., Tsunoda, T., Yasui, W., Inai, K., Kohno, N., and Nakamura, Y. (2006). Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. Int. J. Oncol. *29*, 567–575.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tsui, V., and Case, D.A. (2000-2001). Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers *56*, 275–291.

van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure *16*, 1245–1256.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., et al. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature *508*, 263–268.

Wishart, D.S., and Sykes, B.D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR *4*, 171–180.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90.

Yang, W.H., Kim, J.E., Nam, H.W., Ju, J.W., Kim, H.S., Kim, Y.S., and Cho, J.W. (2006). Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. Nat. Cell Biol. *8*, 1074–1083.

Zhang, T., Park, K.A., Li, Y., Byun, H.S., Jeon, J., Lee, Y., Hong, J.H., Kim, J.M., Huang, S.M., Choi, S.W., et al. (2013). PHF20 regulates NF-κB signalling by disrupting recruitment of PP2A to p65. Nat. Commun. *4*, 2062.

Zhang, X., Peng, D., Xi, Y., Yuan, C., Sagum, C.A., Klein, B.J., Tanaka, K., Wen, H., Kutateladze, T.G., Li, W., et al. (2016). G9a-mediated methylation of ERα links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression. Nat. Commun. *7*, 10810.

Zhao, W., Li, Q., Ayers, S., Gu, Y., Shi, Z., Zhu, Q., Chen, Y., Wang, H.Y., and Wang, R.F. (2013a). Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. Cell *152*, 1037–1050.

Zhao, X., Su, J., Wang, F., Liu, D., Ding, J., Yang, Y., Conaway, J.W., Conaway, R.C., Cao, L., Wu, D., et al. (2013b). Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 di-methylation activity by MLL/SET complexes. PLoS Genet. *9*, e1003940.

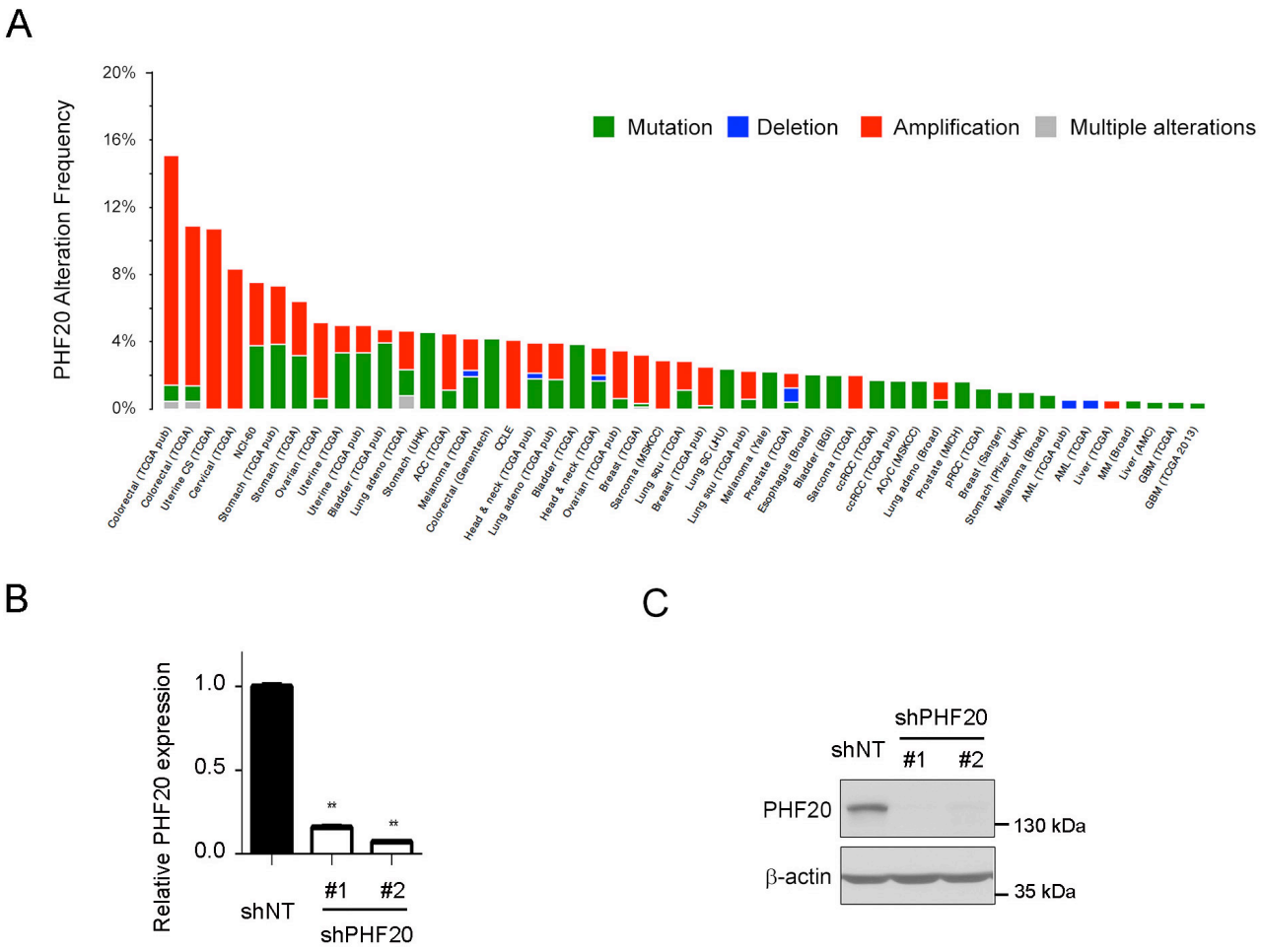
Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/*boxB* RNA complex. J. Am. Chem. Soc. *119*, 6711–6721.

**Cell Reports, Volume *17***

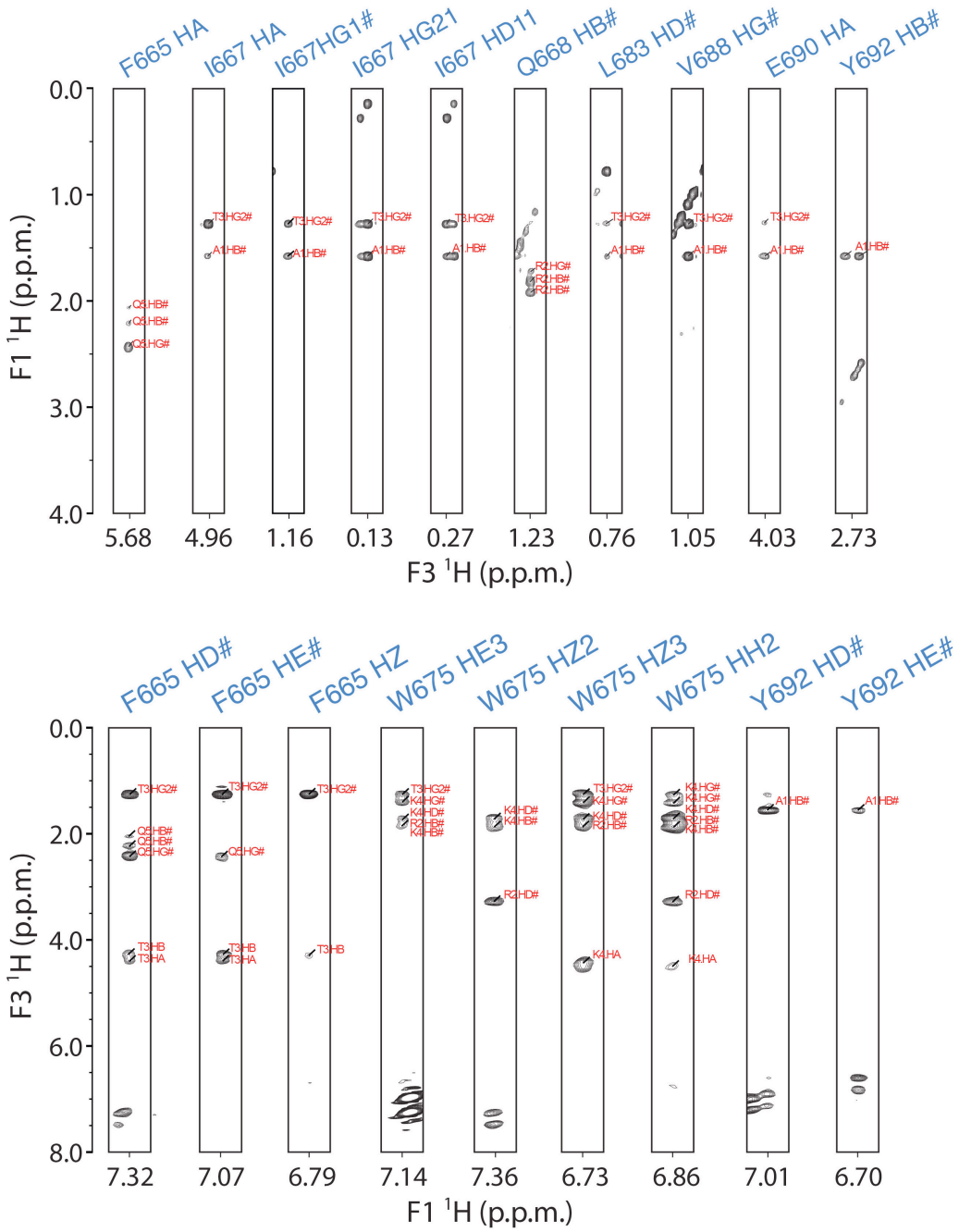
## Supplemental Information

# PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation

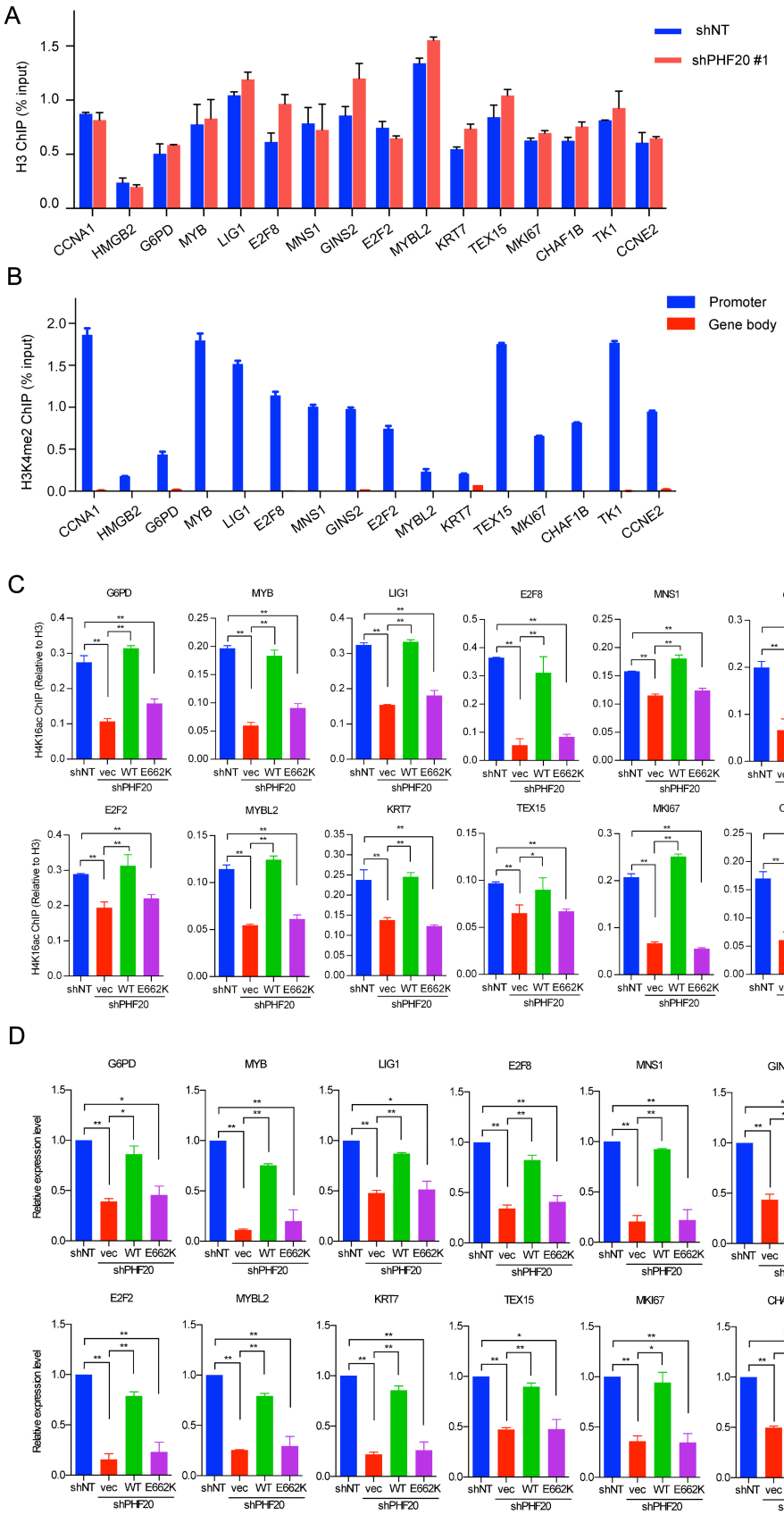
**Brianna J. Klein, Xiaoyan Wang, Gaofeng Cui, Chao Yuan, Maria Victoria Botuyan, Kevin Lin, Yue Lu, Xiaolu Wang, Yue Zhao, Christiane J. Bruns, Georges Mer, Xiaobing Shi, and Tatiana G. Kutateladze**




**Supplemental Figure S1. PHF20 is amplified in human cancers.** (A) PHF20 is amplified in human cancers. The gene alterations frequency data is obtained from the TCGA database. (B) qRT-PCR analysis of PHF20 knockdown in H1792 cells. (C) Western blot analysis of PHF20 knockdown in H1792 cells. Related to Figure 1.




**Supplemental Figure S2. Intermolecular NOE signals between PHF20 PHD and H3K4me2.** Related to Figure 3.


A
shNT
shPHF20 #1
H3 ChIP (% input)
CCNA1
HMGB2
G6PD
MYB
LIG1
E2F8
MNS1
GINS2
E2F2
MYBL2
KRT7
TEX15
MKI67
CHAF1B
TK1
CCNE2
B
Promoter
Gene body
H3K4me2 ChIP (% input)
C
H4K16ac ChIP (Relative to H3)
shNT
vec
WT
E662K
shPHF20
D
Relative expression level

**Supplemental Figure S3. The H3K4me2-binding activity of PHD is essential for PHF20 to regulate MOF-dependent H4K16ac and gene expression.** (A) qRT-PCR analysis of total H3 ChIP on promoters of the indicated genes in control (shNT) and PHF20 KD (shPHF20) cells. (B) qRT-PCR analysis of H3K4me2 ChIP on promoters and gene bodies of the indicated genes in cells as in (A). (C) qRT-PCR analysis of H4K16ac ChIP on promoters of the indicated genes in control (shNT), PHF20 KD (shPHF20) and KD cells rescued with the WT PHF20 or E662K mutant. (D) qRT-PCR analysis of gene expression in cells as in (C). Error bars represent s.e.m. of 3 biological replicates. ** *P-values*<0.01**.** Related to Figure 5.

**Supplemental Table S2. X-ray crystallographic data collection and refinement statistics.** Related to Figure 3.

| | PHF20 PHD |
|---|---|
| **Data collection** | |
| Space group | $P_{61}$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 46.74, 46.74, 50.93 |
| α, β, γ (°) | 90, 90, 120 |
| Resolution (Å) | 31.69-1.25 (1.29-1.25) * |
| $R_{sym}$ or $R_{merge}$ | 5.8 (24.3) |
| $I / \sigma I$ | 7.4 (3.1) |
| Completeness (%) | 99.7 (97.8) |
| Redundancy | 1.9 (1.8) |
| | |
| **Refinement** | |
| Resolution (Å) | 31.69-1.25 |
| No. reflections | 17326 |
| $R_{work}$ / $R_{free}$ | 11.8/14.4 |
| No. atoms | 510 |
| Protein | 425 |
| Ligand/ion | 14 |
| Water | 71 |
| *B*-factors | 13.6 |
| Protein | 11.5 |
| Ligand/ion | 35.2 |
| Water | 21.8 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.012 |
| Bond angles (°) | 1.47 |

*Values in parentheses are for highest-resolution shell.

**Supplemental Table S3. NMR restraints and refinement statistics.** Related to Figure 3.

| **NMR distance and dihedral restraints** | PHF20 PHD–H3K4me2 |
|---|---|
| Distance restraints | |
| Total NOE | 2294 |
| Intra-residue | 545 |
| Inter-residue | 1635 |
| Sequential (\|i-j\|=1) | 515 |
| Medium-range (\|i-j\|<5) | 380 |
| Long-range(\|i-j\|>4) | 740 |
| Intermolecular | 114 |
| Hydrogen bonds | 10 |
| Total dihedral angle restraints | 86 |
| $\phi$ | 44 |
| $\psi$ | 42 |
| | |
| **Structure statistics** | |
| Violations(mean ± s.d) | |
| Distance constraints (Å) | 0.08±0.06 |
| Dihedral angle constraints (°) | 0.60±0.71 |
| Max. dihedral angle violation (°) | 4.10 |
| Max. distance restraint violation (Å) | 0.27 |
| Deviations from idealized geometry | |
| Bond lengths (Å) | 0.0116±0.0001 |
| Bond angles (°) | 2.62±0.02 |
| Impropers (°) | 0.38±0.02 |
| Average pairwise r.m.s. deviation (Å) [a] | |
| Heavy | 0.73±0.11 |
| Backbone | 0.19±0.05 |

[a] Calculated from residues 651-699 of PHF20 and residues 1-5 of H3K4me2.