Aus der Klinik für Allgemein-, Viszeral-, Gefäß- und Transplantationschirurgie

der Ludwig-Maximilians-Universität München

(Direktor: Prof. Dr. med. Jens Werner)

The Epigenetic Role of PHF20/MOF Histone Acetyltransferase Complex in Lung Cancer

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Xiaoyan Wang

Aus Shandong, China

2017

Mit der Genehmigung der Medizinischen Fakultät

der Universität München

Berichterstatterin:	Prof. Dr. med. Christiane J. Bruns
Mitberichterstatter:	PD Thomas Düll
	Prof. Wolfgang Zimmermann
Mitbetreuung durch den promovierten Mitarbeiter:	Dr. rer. nat Yue Zhao
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	07. 12. 2017

Something attempted, something done.

Declaration

I hereby declare that the thesis is my original work.

The work and data presented in the thesis were performed independently.

X-ray crystallography, NMR titration experiments, NMR structure determination, and fluorescence spectroscopy were performed by our collaborator Dr. Brianna J. Klein (Department of Pharmacology, University of Colorado School of Medicine, USA) and Dr. Gaofeng Cui (Department of Biochemistry and Molecular Biology, Mayo Clinic, USA).

Parts of the results have been published in the following manuscripts:

Klein BJ*, **Wang X***, Cui G*, Yuan C, Botuyan MV, Lin K, Lu Y, Wang X, Zhao Y, Bruns CJ, Mer G, Shi X, Kutateladze TG. PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation. ***Cell Rep.*** 2016 Oct 18;17(4):1158-1170.

No unauthorized data were included.

Information from the literature was cited and listed in the reference.

All the data presented in the thesis will not be used in any other thesis for scientific degree application.

The work for the thesis began from Nov. 2014 with the supervision from Prof. Dr. Xiaobing Shi in UT M.D. Anderson Cancer Center, Houston, United States and Prof. Dr. med. Christiane J. Bruns and Dr Yue Zhao in Germany.

03.2017

Xiaoyan Wang

Table of Contents

1 INTRODUCTION

1.1 Overview of lung cancer

1.1.1 Incidence and mortality of lung cancer

Lung cancer was the most common cancer worldwide with nearly 1.8 million new cancer cases, and contribute 13% of the total number of new cancer cases diagnosed in 2012. In Germany, there are around 50,813 new cases of lung cancer, that's 140 cases diagnosed every day. The age-standardized incidence for lung cancer in Germany is about 10.3% (Figure 1.1) [1, 2].

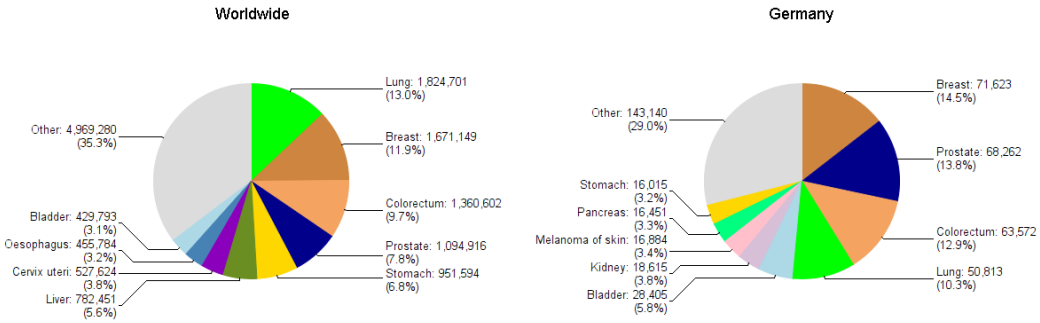


Figure 1.1. Incidence of lung cancer worldwide and in Germany based on the data from GLOBOCAN project.

Lung cancer is the leading cause of cancer-related death in men and the second leading cause in women worldwide [1, 3]. Lung cancer accounts for about 19.4% of all cancer deaths, with an estimated 1.6 million deaths in 2012 (1.1 million in men and 491,200 deaths in women). In Germany, there were around 43,420 lung cancer deaths in 2012, that's 119 deaths every day. Lung cancer accounts for 20% of all cancer deaths in Germany (Figure 1.2) [1, 2]. According to the latest WHO data published in May 2014, lung cancers deaths in Germany reached 45,785, and the age-adjusted death rate is 27.78 per 100,000 of population.

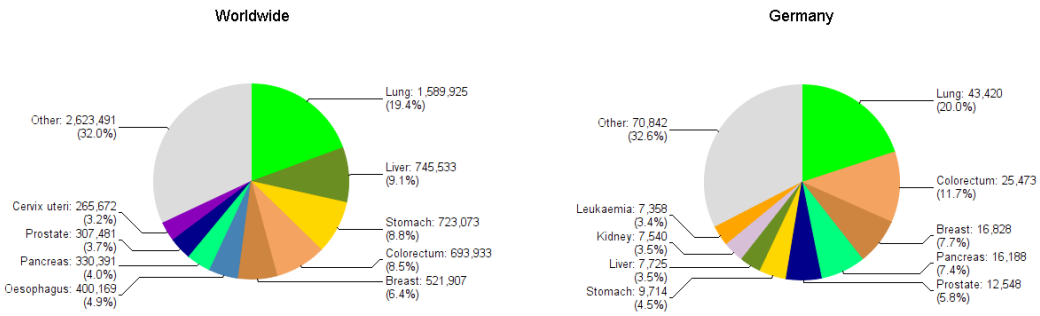


Figure 1.2. Mortality of lung cancer worldwide and in Germany.

1.1.2 Types of lung cancer

Based on the way the cells look under a microscope, the vast majority of lung cancers are classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [4]. Although these two types of cancer share a common organ of origin and some molecular attributes, they have unique growth and spread patterns and represent distinct diseases. Therefore it is important to make a distinction between these two types of cancer for the purpose of cancer treatment [4, 5].

SCLC is the most aggressive and rapidly growing tumor of all the types, and comprises around 15% of lung cancers. SCLCs can rapidly metastasize to other parts of the body and are most often diagnosed when they have spread extensively [6]. Approximately 85% of lung cancers are comprised of NSCLCs, which broadly includes squamous cell carcinoma (SqCC), adenocarcinoma (AdC), and large cell carcinoma subtypes. The AdC and SqCC subtypes represent about 85% of NSCLC cases. NSCLC is often insidious, producing no symptoms until the disease is well advanced [4, 7].

1.1.3 Lung cancer risk factors

The evolution of lung cancer is a multistep process involving genetic, epigenetic, and environmental factor interactions. Exposure to risk factors may increase the risk of lung cancer. Exposure to tobacco smoke has long been recognized as the principal risk factor for development of lung cancer. More than 80% of lung cancers are attributable to active smoking. The lifetime risk of developing lung cancer in smokers is much higher than the one

in nonsmokers [8, 9]. Passive smoking also increases the risk of lung cancer and is a major contributor to indoor pollution. Although smoking is the major risk factor for lung cancer, not all patients with lung cancer have a smoking history. Global statistics estimate that 15% of men and 53% of women with lung cancer are never-smokers [10]. So there are other risk factors for lung cancer, which include radon, asbestos, polycyclic aromatic hydrocarbons, quartz dust, nickel dust and other pollutants [11, 12].

1.1.4 Lung cancer survival rate

Most lung cancers are diagnosed at an advanced stage, where the prognosis is poor and therapeutic options are limited [7]. The survival rates for lung cancer vary depending on the cancer type and the stage of the disease. The overall 5-year survival rate for small cell lung cancer is only about 6%, while the one for non-small cell lung cancer is roughly 18%. The 5-year survival for localized lung cancer is around 55%, while the 5-year survival rate for distant tumors (spread to other organs) is only 4 percent [13, 14]. Detection of lung cancer at an early stage leads to a better prognosis. However, only 16% of lung cancer cases are diagnosed at an early stage [15]. In order to develop more effective strategies in lung cancer diagnosing and treatment, the underlying molecular and cellular mechanisms of lung cancer remains to be explored.

1.2 Cancer epigenetics

1.2.1 Epigenetics

As a relatively new scientific field, epigenetics has been intensely studied over the past two decades. Epigenetics represents a new frontier owing to its important role in regulation of normal biological processes and in development of human diseases, including cancer [16].

Historically, the term ‘‘epigenetics’’ was originally coined by Conrad Waddington to describe heritable changes in a cellular phenotype without altering the DNA sequence [17]. Nowadays, epigenetics can be defined as the study of any potentially stable and, ideally, heritable change in gene expression or cellular phenotype that occurs without alterations in

the DNA coding sequence [18, 19]. Epigenetics consists of heritable modifications including DNA methylation, histone modifications, nucleosome remodeling, and RNA-mediated targeting, all of which can influence overall chromatin structure [16]. Chromatin is the macromolecular complex of DNA and its intimately associated proteins, which provides the scaffold for the packaging of our entire genome. The fundamental repeating unit of chromatin is the nucleosome. The nucleosome contains 145–147 base pairs of DNA wrapped around a histone octamer, constructed from two copies of each of the core histones H2A, H2B, H3, and H4 [20]. The histone tails of all four core histones are subject to a variety of post-translational modifications including methylation, acetylation, phosphorylation, ubiquitylation, all of which occur at the site of a specific amino acid, such as K4 and K9 on the histone H3 tail [21, 22]. For now, at least 4 different DNA modifications and at least 16 distinct classes of histone modification have been described [18, 23]. These chemical modifications are dynamically laid down by "chromatin writers" and removed by "chromatin erasers" in a highly regulated manner. All of the "chromatin writers" and the "chromatin erasers" are chromatin-modifying enzymes [16]. These modifications can alter chromatin structure by altering non-covalent interactions within and between nucleosomes [18]. They also serve as docking sites for specialized proteins with unique domains that specifically recognize these modifications. These "chromatin readers" recruit additional chromatin modifiers and remodeling enzymes, which serve as the effectors of the modification [24]. All epigenetic processes work together to establish and maintain the global and local condensed or decondensed chromatin states that eventually determine gene expression.

Epigenetic modifications play an important role in the regulation of critical cellular processes, such as transcription, DNA repair, and replication. Aberrant epigenetic events are intimately linked to human diseases, most notably cancer [25, 26].

1.2.2 Cancer epigenetics

Cancer evolution at all stages is driven by both epigenetic abnormalities as well as genetic alterations [27]. In the last several years, the focus on the origin and progression of human cancers has shifted from genetic to epigenetic regulation, with particular attention to methylation and acetylation events that have profound effects on the eventual expression of oncogenes and the suppression of tumor suppressors [26, 28].

The cancer epigenome is characterized by global changes in DNA methylation and altered histone modification patterns [29]. For example, cancer cells show the global DNA hypomethylation and promoter-specific hypermethylation [30]. Disruption of the epigenetic machineries, either by mutation, deletion or the altered expression of any of their components, is known to provoke aberrant gene expression patterns that give rise to all typical cancer characteristics [31].

Aberrant epigenetic patterns associated with the development and progression of cancer have a potential clinical application. Since the aberrant DNA methylation events occurring in cancer cells have been extensively mapped, DNA methylation can be used as the biomarkers for cancer detection and prediction of therapy responses [32]. In contrast with the static alterations in DNA, epigenetic changes can potentially be reversed, making the epigenetic machineries involved in such processes promising therapeutic targets [29]. A few drugs targeting epigenetic changes have already been approved for cancer treatment. These epigenetic drugs include small molecules that target chromatin writers (DNA methyltransferase inhibitors), chromatin erasers (histone deacetylase inhibitors) and chromatin readers (histone acetylation reader inhibitors) [29, 33, 34].

Epigenetic alterations, including DNA methylation, histone modifications, and non-coding RNA expression, have widely been reported in the literature to play a critical role in the genesis of lung cancer [35]. With the recent advances in the push towards personalized therapy, questions have been asked about the possible targeting of epigenetic events for personalized lung cancer therapy. Some progress has been made but a lot of work still need to be done.

1.3 PHF20

1.3.1 PHF20 in human cancers

Plant homeodomainfinger protein 20 (PHF20, also termed glioma-expressed antigen 2, GLEA2) was originally identified as an antibody-reactive protein in patients suffering from glioblastoma [36]. In glioma patients, PHF20 elicit a frequent immune response, and autoantibodies against PHF20 led to a significant prolonged survival [37]. Interestingly, high

expression level of *PHF20* correlated with lower survival in NSCLC patients [38]. In addition, PHF20 was also identified as a homolog of hepatocellular carcinoma associated antigen 58 (HCA58) [36]. Moreover, *PHF20* was abundantly expressed in both advanced small-cell lung cancer and advanced adenocarcinoma [39]. Subsequently, abnormal expression of *PHF20* in various cancers indicates that PHF20 might be tumor-associated factor and could play a role in cancer progression. However, little is known about the function of PHF20 in various cancers.

1.3.2 Epigenetic role of PHF20

PHF20 is a multidomain protein, which contains several domains including two Tudor domains, a plant homeodomain (PHD) finger and putative DNA-binding domains AT hook and C2H2-type zinc finger (Figure 1.3) [40]. Since Tudor domain and PHD finger domain have been identified as methylated lysine binding domains [41-44], PHF20 was considered as a histone code reader. Indeed, the first Tudor (Tudor1) domain doesn't bind dimethyl-lysine residues, due to an atypical and occluded aromatic cage [45]. Whereas the second Tudor (Tudor2) domain can bind to several histone methylation, including H4K20me2 and H3K79me2 [46]. Besides histone modification, the Tudor domain 2 also recognizes the methylated p65 and methylated p53 [40, 47]. PHF20 positively regulates NF-κB signaling through interacting with the dimethylated p65 at Lys221 (p65K221me2) through the Tudor2 domain. The interaction of PHF20 with methylated p65 contributes to persistent p65 phosphorylation by limiting the recruitment of phosphatase PP2A [47]. PHF20 directly binds to p53 dimethylated at K370 or K382 (p53K370me2 and p53K382me2) through its second Tudor (Tudor2) domain. Tudor2 forms a dimer capable of associating with both p53 dimethyllysine marks simultaneously, thus greatly enhancing binding of PHF20 to p53. Interaction with PHF20 leads to the stabilization and activation of p53, because it blocks p53 ubiquitination and upregulates p53 protein level in response to DNA damage [40]. PHF20 is a potent transcriptional activator by an epigenetic-based mechanism.

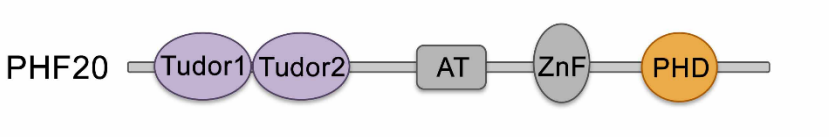


Figure 1.3 Schematic representation of PHF20. PHF20 contains two Tudor domains, a plant homeodomain (PHD) finger and putative DNA-binding domains AT hook and C2H2-type zinc finger.

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 [48-51]. 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 [49, 52, 53]. 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 [52, 53]. 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 [53].

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 [51, 54]. 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 [55]. Recent in vivo and in vitro studies demonstrate that PHF20 transcriptionally regulates p53 in an Akt-dependent manner and promotes nuclear factor κB (NF-κB) transcriptional activity [47, 56]. 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 [57].

1.4 Aims of this study

Abundant expression of PHF20 is reported in various cancers [38, 39, 58]. *PHF20* knockout mice die shortly after birth and display a variety of phenotypes within the skeletal and hematopoietic systems [57]. *PHF20* deficiency halts conversion of somatic cells into induced pluripotent stem cells (iPSCs), revealing a requirement of this factor for cell reprogramming [59]. Despite the vital role of PHF20 in survival and carcinogenesis, the molecular mechanism by which PHF20 contributes to transcription and exerts carcinogenic activity remains unclear.

In this study, we used lung cancer as a model to investigate the function of PHF20 in human cancer. We employ diverse approaches to elucidate how PHF20 affects cancer development at the cellular and molecular levels, including biochemistry, cell biology, next generation sequencing, and structure biology.

2 MATERIAL AND METHODS

2.1 Materials

2.1.1 Eukaryotic cell lines

Human lung fibroblast cell lines WI38 and IMR90, and human lung cancer cell lines A549, H1299, H1355, H1792 and H1993 were used in this study. These cell lines were kindly given from Dr. John Heymach, Dr. Seyed Moghaddam and Dr. Min Gyu Lee in UT M.D. Anderson Cancer Center in United States. The human lung cells information is listed in Table 2.1.

Human embryonic kidney cells HEK293T were used for virus packaging. The cell line was obtained from ATCC company (CRL-11268).

Table 2.1 Lung cell lines information

Cell line	Tissue	Histology	Tumor Source	Cell Type
WI38	Lung	Normal tissue	NA	Fibroblast
IMR90	Lung	Normal tissue	NA	Fibroblast
A549	Lung	NSCLC	Primary	Epithelial
H1299	lung	NSCLC	Metastasis (lymph node)	Epithelial
H1355	lung	NSCLC	Metastasis (pleural effusion)	Epithelial
H1792	lung	NSCLC	Metastasis (pleural effusion)	Epithelial
H1993	lung	NSCLC	Metastasis (lymph node)	Epithelial

2.1.2 Bacterial strains

Rosetta-2 (DE3) competent cells (Catalog No: 71400) were purchased from Novagen, and were used for the expression of *PHF20* WT and mutants. DH5α competent cells (Catalog No: 18265017) were purchased from Invitrogen, and used for general cloning and subcloning.

2.1.3 Plasmids constructs

cDNA/Construct	**Vector**	**Epitope tag**	**Source**
pMD2.G		-	Addgene
pPAX2		-	Addgene
pENTR3C		-	Invitrogen
pGEX-6P-1		GST	GE Healthcare
pCDH-Blasticidin-GW		Flag-HA	Dr. H Wen
PHF20	pENTR3C	-	Dr. X Shi
PHF20-Res	pENTR3C	-	X Wang
PHF20-Res	pCDH-BT-GW	Flag-HA	X Wang
PHF20 (E662K)-Res	pENTR3C	-	X Wang
PHF20 (E662K)-Res	pCDH-BT-GW	Flag-HA	X Wang
PHF20-PHD-L (aa 651-699)	pGEX-6P-1	GST	Dr. X Shi
PHF20-PHD-L (E662K)	pGEX-6P-1	GST	X Wang

PHF20-PHD (aa 646-699)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (E662K)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (E662D)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (E662Y)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (F665K)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (C694S)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (W675A)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (R654H)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (F665V)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (M666I)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (C680F)	pTEV	His	Dr. TG Kutateladze
PHF20-PHD (P689L)	pTEV	His	Dr. TG Kutateladze

2.1.4 shRNA

pLKO-scrambled shRNA negative control vector and *pLKO-PHF20 shRNA* constructs were purchased from Sigma. The *PHF20*-targeting shRNA sequences were 5′-CCCGAGAAATACACCTGTTAT-3′ (*shPHF20#1*) and 5′-ATTGTGCCACTGATGATAAAC-3′ (*shPHF20#2*).

2.1.5 Antibodies

Primary antibody	Species	Dilution for WB	Company	Catalog No.

anti-total H3	rabbit	1:5,000	Abcam	ab1791
anti-total H4	rabbit	1:5,000	Abcam	ab7311
anti-H4K16ac	rabbit	1:1,000	Millipore	07-329
anti-Flag	mouse	1:5,000	Sigma	F-1804
anti-GST	rabbit	1:5,000	Santa Cruz	sc-459
anti-tubulin	mouse	1:5,000	Sigma	T8328
anti-β-actin	mouse	1:5,000	Sigma	A1978
anti-PHF20	rabbit	1:1,000	Cell Signaling	3934S

Secondary antibody	**Company**	**Catalog No.**
donkey-anti-rabbit IgG (H+L)	Jackson ImmunoResearch	711-035-152
donkey-anti-mouse IgG (H+L)	Jackson ImmunoResearch	711-035-150
donkey-anti-goat IgG (H+L)	Jackson ImmunoResearch	711-035-147

*All the secondary antibodies were diluted at 1:10,000 for Western blotting.

2.1.6 Oligos for plasmids cloning and point mutagenesis

All synthetic oligonucleotides were purchased from Sigma-Aldrich. Lyophilized oligonucleotides were dissolved in water. The sequences of both forward (For) and reverse (Rev) primers are given in the 5' to 3' direction.

Oligo	**Sequence**
PHF20 PHD-BamHI-For	GCCGGGATCCGATGGGGATGACCGC

PHF20 PHD-EcoRI-Rev	GCCGGAATTCTTACCACTCCTTGTC
PHF20-E662K-For	TGTGAGGTCCAGGAGAAAAATGACTTCATG
PHF20-E662K-Rev	CATGAAGTCATTTTTCTCCTGGACCTCACA

2.1.7 Common chemical reagents

Chemical Name	**Catalog No.**	**Manufacturer**
Agarose	15510-027	Sigma-Aldrich
Bromophenol Blue	18030	Fluka
Coomassie Brilliant Blue R-250	161-0400	Bio-Rad
Ethylenediaminetetraacetic Acid Disodium Salt Dihydrate (EDTA)	E5134	Sigma-Aldrich
D-(+)-Glucose	G5400	Sigma-Aldrich
HEPES	BP310	Fisher Scientific
Magnesium Chloride Hexahydrate	M2670	Sigma-Aldrich
Magnesium Sulfate Heptahydrate	M5921	Sigma-Aldrich
Phenylmethanesulfonyl fluoride (PMSF)	P7626	Sigma-Aldrich
Potassium Chloride	P3911	Sigma-Aldrich
Sodium Bicarbonate	S6297	Sigma-Aldrich
Sodium Citrate Dihydrate	S279	Fisher Scientific
Sodium Dodecyl Sulfate (SDS)	L4390	Sigma-Aldrich

Sodium Hydroxide	S318	Fisher Scientific
Sodium Phosphate Monobasic	S8282	Sigma-Aldrich
Sodium Sulfate	S6547	Sigma-Aldrich
Sucrose	S0389	Sigma-Aldrich
Zinc Chloride	Z0152	Sigma-Aldrich
Dimethyl Sulphoxide (DMSO)	D8418	Sigma-Aldrich
GRAM Staining Solution I (crystal violet)	6100	NEG Scientific Inc.
10×Phosphate Buffered Saline (PBS)	BP399	Fisher Scientific
Triton X-100	T8787	Sigma-Aldrich
Tween-20	BP337	Fisher Scientific
Glycine	BP381	Fisher Scientific
Sodium Chloride	BP35810	Fisher Scientific
Tris base	BP152	Fisher Scientific
Ethyl Alcohol, 200 proof	61509	ACROS
2-Mercaptoethanol	M3148	Sigma-Aldrich
RNAse Away	10328-011	Invitrogen
Water (UltraPure, DNase- RNase- free)	10977-015	Invitrogen
Hydrochloric Acid	S93258	Fisher Scientific
Formaldehyde	4018	PolySciences Corp.
Glycerol	G5516	Sigma-Aldrich

Methanol	BP11054	Fisher Scientific
Isopropanol	BP2618	Fisher Scientific
Glutathione, reduced form	G4251	Sigma-Aldrich
Streptavidin Sepharose 4B	17-5113-01	GE Healthcare
UltraLink Immobulized Protein A/G	53133	Pierce
Complete mini, EDTA-free, protease inhibitor cocktail	12186500	Roche
Albumin Standard	23210	Thermo Scientific
Paraformaldehyde	19200	EMS
DTT (1,4-dithio-DL-threitol)	D9779	Sigma-Aldrich
Glutaraldehyde	G7526	Sigma-Aldrich
Lysozyme	L6876	Sigma-Aldrich
Proteinase K (Fungal)	25530-015	Invitrogen

2.1.8 Materials for PCR and quantitative real-time PCR

Kit Name	**Catalog No.**	**Manufacturer**
Taq DNA Polymerase	10342020	Invitrogen
Power SYBR Green PCR Master Mix	4367659	ABI
RNeasy Mini Kit	74104	Qiagen
iScript cDNA Synthesis Kit	1708890	Bio-Rad

Table 2.2 Primers for qRT-PCR

Gene	Forward sequence	Reverse sequence
CCNA1	TCCAGAACTTCACCTCCATATCAG	GATCCAACGTGCAGAAGCCTAT
HMGB2	CCCGGACTCTTCCGTCAA	TCCATCTCTCCGAACACTTCTTG
G6PD	TCCTGCGGGAAGAGCTTTT	GATGAATATGTGTGTATCCGACTGATG
MYB	GGGCGTTTTTCTGACTTGGA	GCAGTTCCGGCTCTGTACACT
LIG1	CGAGGTCCTGAAACGCTTTG	CTGCCCGTCATATTTGTATTCG
E2F8	GGCACGGGCGACACAA	CTTATTCTCCTCCCCGATGCT
MNS1	GTCAACAGTCAAATCAGGAATCAAA	GAAATTGCTTGCGCTGAACA
GINS2	GCGGGACTTTCCTCACACA	GGAGGTTCGTGCGGAGTTT
E2F2	GCCTATGTGACTTACCAGGATATCC	CCTTGACGGCAATCACTGTCT
MYBL2	TGCCCCTTCAAACTCTTCCA	CCTGGTTGAGCAAGCTGTTG
KRT7	CCATGCCTGGTCCCAAGA	TGCTTTATTGGAAGCTATTCTGACA
TEX15	CAGCTGACTCCCTTGCATTG	AAAGGCTGGGAATAAATGTCAAAT
MKI67	CAGGAAAACAAGAGTCAGGTTCAGA	GGACCTACGGCGTTGATCA
CHAF1B	GTCCCTCCAGCATTTTGCA	GTTTCTCGGGCACCGTTCTA
TK1	TTCCTACCTCTGGTGATGGTTTC	TGCCACCCATCTTGGTGAA
CCNE2	CCCAGCCAGACGGAATCC	TCCTGGGTAGTTTTCCTCTTCTTG

2.1.9 Reagents for Western blotting

Chemical Name	Catalog No.	Manufacturer
40% Acrylamide/Bis solution.29:1 (3.3%)	161-0146	Bio-rad
TEMED	1610801	Bio-rad
Ammonium Persulfate (APS)	BP179	Fisher Scientific
Pierce supersignal western pico chemilunimescent substrate	34080	Thermo Scientific
Amersham ECL western blotting detection reagents	RPN2106	GE Healthcare
Amersham ECL Prime western blotting detection reagents	RPN2236	GE Healthcare
Millipore Immobilon western chemiluminescent substrate	WBKLS0100	Millipore
Gelcode Blue Stain Reagent	24592	Pierce
PageRuler Plus Prestained 10-250kDa Protein Ladder	PI26620	Thermo Scientific

2.1.10 Materials for ChIP

Chemical Name	Catalog No.	Manufacturer
Glass beads, acid-washed, 212-300 um	G1277	Sigma-Aldrich
Lithium Chloride	L4408	Sigma-Aldrich

PIPES	P1851	Sigma-Aldrich
37% Formaldehyde	BP531	Fisher Scientific

Table 2.3 Primers for ChIP-PCR

Primer	For/Rev	Location (bp)	Sequence
CCNA1-p	Forward	-350	TTAGCGGCTGTTGGGAGAAC
	Reverse	-292	GATGGCCCCGGCTTTG
CCNA1-c	Forward	+4620	TGCCCACCTGCCCTGAT
	Reverse	+4675	TTAAGGTGAGCCATGCCAGAT
HMGB2-p	Forward	-947	TGGCTCGGGAAGGAAAGC
	Reverse	-1004	CCCCCAGTCTCCCGACTAAG
HMGB2-c	Forward	+1348	CAAGGGTTGCGATACAGCAGTA
	Reverse	+1275	GGACTATGGCACTGTGTGTGATG
G6PD-p	Forward	-1037	TCTTCGGAAATGCCTCACATATAG
	Reverse	-1101	CGAGCAAACAGGCATATGAAATC
G6PD-c	Forward	+8837	GACAGGGTTTCCCCATGTTG
	Reverse	+8774	GGCGGGCAGATCACTTGA
MYB-p	Forward	-583	TGTAAACCTTGACGAAAATCCAATC
	Reverse	-496	TCCTCCAGCTCCCACTCACT
MYB-c	Forward	+25047	GCACTCTAGCCTGGGAGACAA

	Reverse	+25113	CCACGGCTGGCCTTAAAAA
LIG1-p	Forward	-746	CACAGGTGAATACTGGTGACACTTC
	Reverse	-808	TTATCCACGCTTCGCCTACA
LIG1-c	Forward	+9994	CCCGGCGAAGGACAGAT
	Reverse	+10092	CACTCATGTCTTCCTGGTCTGTAAA
E2F8-p	Forward	+214	ACCTGTCAGTCCACCTCACG
	Reverse	+152	AGCGGTGCTGACGAGTCC
E2F8-c	Forward	+9532	AAACACTGCATGTTCTCACTCGTAA
	Reverse	+9469	CCTCCCTGTGTCCATGTGTTC
MNS1-p	Forward	+353	GCCAAAAAAGCTGGAAACCA
	Reverse	+297	CGGAGGCTTCAGGTGCTATT
MNS1-c	Forward	+27529	TCTGGCCAGGGCAGTCA
	Reverse	+27432	ATACAATCATGTCATCTGCAAACAAG
GINS2-p	Forward	+176	TCACCCCGATGAGGTAGATCTT
	Reverse	+176	CGAGAAGGAGCTGGTTACCATT
GINS2-c	Forward	+6897	TGGCACCATCACAGCTCACT
	Reverse	+6838	GAGGCGGGAGGATTGCTT
E2F2-p	Forward	-442	TGCATTGTTGCTTAGCTAGCACAT
	Reverse	-503	TGGCCTTCATTTTAGCAGGAA

E2F2-c	Forward	+16079	TGCTAAGAAGACAAAAAGTGGAACA
	Reverse	+16015	CGGCTAGACTGGCAGGATTC
MYBL2-p	Forward	+1185	TGCAGTGGCGCGATCTT
	Reverse	+1247	GGCAGGAGAATCGCTTGAAC
MYBL2-c	Forward	+32788	GAGCCCATCGGTACAGATCTG
	Reverse	+32851	GCTTCGGGAATTCCTCCAA
KRT7-p	Forward	-1482	GGCACAATTCCAGGTCGAA
	Reverse	-1424	CGGCTGTGGAGCATGTATGT
KRT7-c	Forward	+8053	GTCCCTTTCCTTATCCCCTTTAAC
	Reverse	+8113	GGCCCATTAGCAGTGTCTCAA
TEX15-p	Forward	+529	CTTCTCCTCACTGTGGTTTTGGT
	Reverse	+464	TCCTTCCCAATTTCTGTGTCAA
TEX15-c	Forward	+10060	CAGGCACCCGCCACTATG
	Reverse	+9992	GACCAACATGGTGAAACACTATCTCT
MKI67-p	Forward	+719	CAATCCTAGAGCGCGTTTCTG
	Reverse	+647	GCGGACGTTTTAGCTGAGAAAG
MKI67-c	Forward	+15091	CGGCCATCCCCACTTTAAA
	Reverse	+15028	CCATTCAATTCCCACTCTCAGTAA
CHAF1B-p	Forward	+763	GTGAAATAGCCTGGCACAACAA

	Reverse	+822	CCGTCCCATGCTGGAAGT
CHAF1B-c	Forward	+18384	TCTCAGAGGAGAATGTCTGCTCAT
	Reverse	+18455	GGGCTATGTTGTCATCTCTATGCA
TK1-p	Forward	+407	GAGAACATCGGCCCGAGAA
	Reverse	+351	GCGGGCCCTCAAATGAC
TK1-c	Forward	+9773	GCTCATTCCCAAGGTCTCATTATATT
	Reverse	+9704	AGAAGTGTTTCGGATTTCAGGTTT
CCNE2-p	Forward	-1647	GCCCAAGGTGACAGCATCA
	Reverse	-1717	GCGTTAGAAATGGCAGAAAGTTTC
CCNE2-c	Forward	+7887	GTGGTGGCAAGCACCTGTAA
	Reverse	+7821	CCCAGGTTCAAGCGATTCTTC

p: promoter, c: gene body

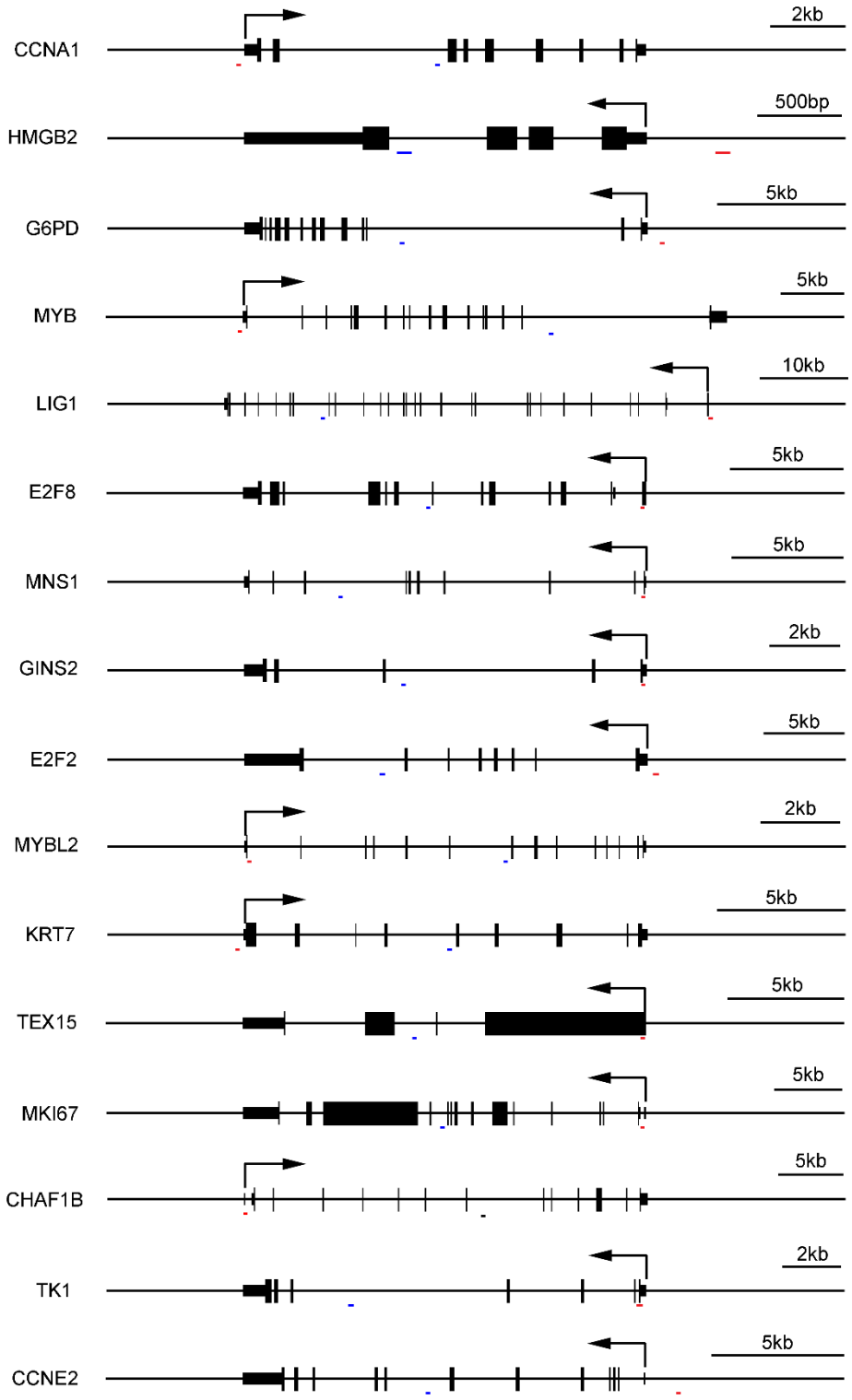


Figure 2.1 Schematic representation of ChIP-PCR primers location. The red line (-) represent the primer targeting gene promoter, while the blue line (-) represent the primers targeting gene body.

2.1.11 Reagents for bacterial culture

Chemical Name	Catalog No.	Manufacturer
Isopropyl b-D-1-thiogalactopyranoside (IPTG)	I6758	Sigma-Aldrich
Ampicillin sodium salt	BP1760	Fisher Scientific
Carbencillin sodium salt	C1389	Sigma-Aldrich
Chloramphenical	C1919	Sigma-Aldrich
Kanamycin sulfate	K1377	Sigma-Aldrich
Tryptone	BP1421	Fisher Scientific
Yeast Extract Powder	BP1422	Fisher Scientific

2.1.12 Reagents for mammalian cell culture

Chemical Name	Catalog No.	Manufacturer
RPMI 1640	10104CV	Corning
DMEM With 4.5g/L Glucose, Without L-Glutamine and Sodium Pyruvate	15017CM	Corning
Phosphate Buffered Saline (PBS), 1×	21040CV	Corning
Fetal Bovine Serum	35010CV	Corning
Penicillin-Streptomycin (10,000 U/mL)	15140122	Gibco
Trypsin-EDTA (0.25%)	25200056	Gibco
Trypan blue	T8154	Sigma-Aldrich

DMSO (Dimethylsulphoxide)	D2650	Sigma-Aldrich
X-tremeGENE™ HP DNA Transfection Reagent	06 366 546 001	Roche
Polybrene	107689	Sigma-Aldrich
Opti-MEM®I Reduced Serum medium	31985070	Gibco
Puromycin Dihydrochloride	A1113803	Gibco
Blasticidin S HCl	A1113903	Gibco
DMEM powder	12800-017	Gibco

2.2 Methods in molecular biology and biochemistry

2.2.1 Solution stock preparation

1. 1 M Tris-HCl Buffers

pH	**Volume (L)**	**Tris Base (g)**	**HCl (mL)**
pH 7.0	2	242.2	150-155
pH 7.5	2	242.2	120-125
pH 8.0	2	242.2	80-85

*The concentration of HCl is 36.5-38.0%.

2. EDTA 0.5 M (pH 8.0): Add 148 g of EDTA to 600 mL of H_2O, use about 35 g of NaOH to adjust the pH value to 8.0. Then add H_2O to 1000 mL.
3. 10% SDS: Add 200 g of SDS into 2 L of H_2O, heat to 68 ℃ for solubility.

4. NaCl solution (4 M): Add 467.5 g of NaCl into 2 L of H_2O. Autoclave and store at room temperature (RT).
5. NaOH solution (10 M): Add 200 g of NaOH into 0.5 L of H_2O.
6. NaAc (3 M): Add 204 g of $NaAc \cdot 3H_2O$ into 200 mL of H_2O, adjust pH by glacial acetic acid (~60 mL) to pH 5.2. Then adjust to 100 mL.
7. $MgCl_2$ (1 M): Add 101.65 g of $MgCl_2 \cdot 6H_2O$ into 500 mL of H_2O. Autoclave and store at room temperature.
8. $CaCl_2$ (1 M): Add 58.8 g of $CaCl_2 \cdot 2H_2O$ into 400 mL of H_2O, then filter for sterilization and store at room temperature.
9. $MgSO_4$ (1 M): Add 123.3 g of $MgSO_4 \cdot 7H_2O$ into 500 mL of H_2O. Autoclave and store at room temperature.
10. $ZnCl_2$ (0.1 M): Add 3.4 g of $ZnCl_2$ into 250 mL of H_2O. Store in -20°C.
11. IPTG (1 M): Disolve 1 g of IPTG in 4.2 mL of H_2O, then filter through 0.22 μm filters and aliquot 1 mL into Eppendorf tube. Store at -20°C.
12. DTT (1 M): Disolve 5 g of DTT in 32.5 mL of 10 mM NaAc (pH 5.2) solution, then filter through 0.22 μm filters and aliquot 1 mL in Eppendorf tube. Store at -20°C.
13. PMSF (100 mM, =17.4 mg/mL): Disolve 1.74 g of PMSF in 100mL of isoproponal. Aliquot and store at -20°C or RT.
14. Carbencillin or Ampcillin (50 mg/mL) in water: Disolve 2.5 g of carbencillin or ampicillin in 50 mL of H_2O. Aliquot and store at -20°C.
15. Kanamycin (10 mg/mL) in water: Disolve 0.5 g of kanamycin in 50 mL of H_2O. Aliquot and store at -20°C.
16. Chloramphenicol (34 mg/mL) in ethanol: Disolve 1.7 g of chloramphenicol in 50 mL of ethanol.
17. Lysozyme 50 mg/mL: Disolve 2.5 g of lysozyme in 50 mL of H_2O.

2.2.2 Buffer preparation

1. TAE DNA electrophoresis buffer (50×): 2 M Tris-HCl, 50 mM EDTA*;*
   Add 968 g of Tris base, 228.4 mL of glacial acetic acid and 400 mL of 0.5 M EDTA 8.0 to 4 L of H_2O. To make 1× TAE 20 L, add 400 mL of 50 × buffer into 19.6 L of ddH_2O.
2. SDS-PAGE gel solutions

	Vol (L)	Tris Base (g)	HCl (mL)	10% SDS (mL)
4× Lower gel buffer				
1.5 M Tris-HCl, pH 8.8, 0.4% SDS	2	363.3	50-60	80 mL
4× Upper gel buffer				
0.5 M Tris-HCl, pH 6.8, 0.4% SDS	2	121.1	70-80	80 mL

*The concentration of HCl is 36.5-38.0%.

3. 5 × SDS loading sample buffer

	Stock solution	Add volume
250 mM Tris-HCl pH6.8	1 M	25 mL
10% SDS		10 g
30% Glycerol		30 mL
5% β-mercapitalethanol		5 mL
0.02% Bromophenol blue	1%	2 mL

Then add H_2O to 100 mL.

4. 6× DNA loading sample buffer: 40% sucrose, 0.01-0.02% bromophenol blue
   Add 40 g of sucrose to 50 mL of H_2O, add 2 mL 1% bromophenol blue solution, adjust to 100 mL.
5. SDS-PAGE electrophoresis running buffer (10 ×)
   Add 303 g of Tris Base, 1440 g of glycine and 100 g of SDS to 10 L of H_2O.
   5.1 SDS-PAGE electrophoresis running buffer (1×): 25 mM Tris-HCl, 192 mM glycine, 0.1% SDS

Add 1L of 10× SDS-PAGE running buffer into 9 L of H_2O.

5.2 Transfer buffer (1×): 25 mM Tris-HCl, 192 mM glycine, 0.1% SDS, 20% methanol

Add 100 mL of 10× SDS-PAGE running buffer and 200 mL of methanol into 700 mL of H_2O.

6. TBS (10×):

Add 876.6 g of NaCl and 121.1 g of Tris Base to 9 L of H_2O, use about 55 mL of HCl to adjust the pH to 8.0. Then adjust the total volume to 10 L.

6.1 TBS-T (1×): 150 mM NaCl, 10 mM Tris-HCl pH8.0

Add 2 L of 10× TBS buffer, 200 mL of 10% Tween20 to 20 L of H_2O.

6.2 Blocking buffer: 5% nonfat milk in TBS-T

Add 5 g of milk into 100 mL of 1× TBS-T buffer.

7. Bacteria lysis buffer: 50 mM Tris-HCl 7.5, 150 mM NaCl, 0.05% NP-40

Add 50 mL of 1 M Tris-HCl pH 7.5, 37.5 mL of 4 M NaCl and 5 mL of 10% NP-40 into 1 L of H_2O.

8. RIPA buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.1% SDS

Add 50 mL of 1 M Tris-HCl pH 7.5, 37.5 mL of 4 M NaCl, 4 mL of 0.5 M EDTA, 10 mL of NP-40 and 10 mL of 10% SDS in 1 L of H_2O.

9. Cell lysis buffer: 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.5% Triton X100, 10% glycerol, 1 mM DTT

Add 50 mL of 1 M Tris-HCl pH 7.5, 62.5 mL of 4 M NaCl, 5 mL of Triton X-100, 1 mL of 1 M DTT and 100 mL glycerol in 1 L of H_2O.

2.2.3 Plasmid cloning by PCR

Complementary DNA encoding full-length human *PHF20* was cloned into pENTR3C vector. The *PHF20* PHD finger region (aa651–699) was amplified by PCR and cloned from pENTR-PHF20 into pENTR3C.

2.2.3.1 Amplification of DNA fragments for cloning

To amplify DNA sequences, the *Taq* DNA polymerase purchased from Invitrogen was used for PCR reaction. Add the following components to a sterile PCR tube sitting on ice:

Reagents	Volume (µL)	Final Con.
10× PCR buffer (-Mg)	5	1×
10 mM dNTP mixture	1	0.2 mM each
50 mM MgCl2	1.5	1.5 nM
PHF20 PHD-BamHI-For (10 µM)	2.5	0.5 µM
PHF20 PHD-EcoRI-Rev (10 µM)	2.5	0.5 µM
pENTR-PHF20 vector (50 ng/µL)	1	1ng/µL
Taq DNA Polymerase (5 U/µL)	0.2	1 U/reaction
Autoclaved distilled water	to 50 µL	n/a

Mix and briefly centrifuge the components.

The PCR reaction was performed in a thermal cycler using the following program:

Step	Temperature (°C)	Time	
Initial Denaturation	94	3 min	
Denature	94	45 sec	32 PCR Cycles
Anneal	55	30 sec	
Extend	72	30 sec	
Final Extension	72	10 min	
Hold	4	indefinitely	

The amplification product was analyzed by agarose gel electrophoresis (2.2.3.2), purified (2.2.3.3) by QIAquick Gel Extraction Kit and and digested with restriction enzymes (2.2.3.4).

2.2.3.2 Agarose gel electrophoresis

Agarose gel electrophoresis of DNA samples was performed for separating DNA by size for visualization and purification after PCR reactions or restriction enzyme digestions. According to their size, 0.8–2.0 % (w/v) agarose gels were prepared using TAE buffer containing 0.2 µg/mL ethidium bromide (EtBr), which binds to DNA and makes the DNA visible under ultraviolet (UV) light. DNA samples were mixed with 6× DNA loading buffer and loaded on the gel. The gels were run in TAE buffer at a constant voltage of 100 V until the dye line is approximately 75-80% of the way down the gel. The DNA fragments can be visualized under UV light.

2.2.3.3 DNA extraction from agarose gel

After gel electrophoresis the DNA fragment of interest was excised from an agarose gel using a clean scalpel. The DNA was extracted using the QIAquick® Gel Extraction Kit according to the manufacturer's instructions. After binding to the QIAquick membrane, the DNA was washed and finally eluted by adding 30 µL EB buffer.

2.2.3.4 Digestion of DNA with restriction enzymes

The purified DNA fragment and the recipient plasmid pENTR-3C were digested by restriction enzymes BamHI and EcoRI purchased from Fermentas. According to the manufacturer's instruction, we used EcoRI buffer and 2-fold excess of BamHI to set up the digestion reaction system.

For the DNA fragment digestion, the digestion reaction system is

Reagents	**Volume (µL)**
nuclease-free water	5
10× EcoRI Buffer	4
BamHI (10 U/µL)	2
EcoRI (10 U/µL)	1

Purified DNA fragment	28

For the recipient plasmid digestion, the digestion reaction system is

Reagents	**Volume (μL)**
nuclease-free water	11.5
10× EcoRI Buffer	2
BamHI (10 U/μL)	1
EcoRI (10 U/μL)	0.5
pRNTR-3C (200ng/μL)	5

The mixture was incubated for 1 hour at 37 °C.

2.2.3.5 Isolation of the DNA fragment and recipient vector by gel purification

The digestion products were analyzed by agarose gel electrophoresis, then the DNA fragments and the linearized recipient vector were excised from an agarose gel using a clean scalpel. The DNA was extracted using the QIAquick® Gel Extraction Kit according to the manufacturer's instructions.

2.2.3.6 Ligation of DNA fragments

The DNA Ligation Kit purchased from Takara was used for the ligation reaction. The reaction was performed according the manufacturer's instructions. In brief, mix 5 μL of DNA fragment with 5 μL of linearized vector DNA in a PCR tube, then add 10 μL of Solution I and mix thoroughly. The mixture was incubated for 30 min at 16 °C. In parallel, a ligation of the recipient plasmid DNA without any insert was set up as a negative control.

2.2.3.7 Transformation of plasmid DNA

The ligation reaction mixture was used directly for transformation of DH5α competent cells. Add 10 μL of the ligation reaction into a vial of DH5α competent cells and mix gently. The vials were incubated on ice for 30 min, and heat-shocked for 90 seconds at 42°C without

shaking, followed by incubation on ice for 2 min. The bacteria were then cultured in 250 μL of LB medium at 37°C for 1 hour at 225 rpm in a shaking incubator. The cells were spread on the LB plate with kanamycin.

2.2.3.8 Isolation of plasmid DNA

Pick up 2 colonies and culture them overnight in 3 mL of LB medium containing 50 μg/mL kanamycin. The plasmid DNA was isolated using QIAprep Spin Miniprep Kit according the manufacturer's instructions.

The plasmids were analysed by restriction endonuclease digestion (2.2.3.4) and agarose gel electrophoresis (2.2.3.2) to confirm the presence of inserted cDNA. The BamHI and EcroRI restriction enzymes can be used. The plasmids containing inserted DNA were sequenced by Sanger Sequencing Method.

2.2.4 Plasmid cloning using Gateway technique

The full-length human *PHF20* gene and *PHF20* PHD finger (aa651–699) were subsequently cloned into destination vectors using the Gateway technique (Invitrogen). The Gateway Technology is a universal cloning method based on the site-specific recombination properties of bacteriophage lambda. The Gateway Technology provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression.

The sub-cloning reaction was performed using the Gateway LR Clonase II enzyme mix, which is a proprietary enzyme and buffer formulation containing the bacteriophage lambda recombination proteins Integrase and Excisionase, the E. coli-encoded protein Integration Host Factor (IHF). Gateway LR Clonase II enzyme mix catalyzes in vitro recombination between an entry clone and a destination vector to generate an expression clone.

For the LR reaction, add the following components to a 1.5 mL Eppendorf tube at room temperature and mix:

Reagents	Volume (µL)
Entry clone (100 ng/µL)	1 µL
Destination vector (150 ng/µL)	1 µL
TE buffer, pH 8.0	to 8 µL

Then add 2 µL of LR Clonase II enzyme mix to the reaction and mix well by vortexing briefly twice, followed by a brief spin. The mixture was incubated for 1 hour at 25°C. To terminate the reaction, 1 µL of the Proteinase K solution was added to each sample, followed by incubation at 37°C for 10 min.

The LR reaction mixture was used directly for transformation of DH5α competent cells (2.2.3.7).

2.2.5 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) according to the following protocol. Add the following components to a sterile PCR tube sitting on ice:

Reagents	Volume (µL)
10× Pfu Turbo Reaction mix	2.5
10 mM dNTP mixture	1
DMSO	1.25
PHF20-E662K-For (10 µM)	1
PHF20-E662K-Rev (10 µM)	1
pENTR-PHF20 vector (50 ng/µL)	1

Pfu turbo DNA Polymerase	1
Autoclaved distilled water	16.25

Mix and briefly centrifuge the components.

The PCR reaction was performed in a thermal cycler using the following program:

Step	**Temperature (°C)**	**Time**	
Initial Denaturation	95	2 min	
Denature	95	20 sec	19 PCR Cycles
Anneal	55	30 sec	
Extend	68	14 min	
Final Extension	68	10 min	
Hold	4	indefinitely	

After PCR, transfer 20 µL of amplification product to a new tube and incubate each sample with 1µL of Dpn I for 1 hour at 37˚C. At the meantime, 5 µL of amplification product was used to analyze the quality by agarose gel electrophoresis (2.2.3.2). 5 µL of DpnI digested mixture was used directly for transformation of DH5α competent cells (2.2.3.7). All constructs were confirmed by DNA sequencing.

2.2.6 Protein purification

The *PHF20* PHD finger construct (aa 651–699) was cloned into a pGEX-6P-1 expression vector using the Gateway technique. The WT and mutant proteins were expressed in E. coli Rosetta-2 (DE3) pLysS cells grown in Luria broth media supplemented with 150 µM $ZnCl_2$, and then purified using Glutathione Sepharose 4B purchased from GE Healthcare Life Sciences.

2.2.6.1 Protein expression in Rosetta-2 cells

1. 50 ng of pGEX-6P-PHF20-PHD(WT) or pGEX-6P-PHF20-PHD(E662K) plasmid was transformed into competent Rosetta 2 cells, which were plated on LB with 34 μg/mL chloramphenicol and 50 μg/mL ampicillin. The cells grown at 37 ˚C overnight.
2. Pick up single colony to inoculate a starter culture of 12 mL LB/Cam/Amp in a 50 mL conical tube. The bacteria were then cultured at 37°C overnight in a shaking incubator.
3. The next morning, transfer 10 mL overnight culture into a 250 mL flask with 100 mL LB medium containing with 50 μg/mL ampicillin and 150 μM $ZnCl_2$, followed by incubation at 37˚C for ~1 hour in a shaking incubator. OD600 should be around 0.6-0.8, if lower than 0.6, check OD every 20 min.
4. Once the OD600 >0.6, add 20 μL of 1M IPTG stock to a final concentration of 0.2 mM. The bacteria were then cultured at 20°C overnight in a shaking incubator.
5. Harvest the bacteria by centrifugation for 15 min at 4000 rpm. Pour off supernatant and either begin lysis or store pellet at –80˚C.

2.2.6.2 GST-tagged protein purification from bacteria

Bug lysis buffer: *50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.05% (v/v) NP-40; 1 mM PMSF; protease inhibitors (added prior to use)*

Elution Buffer: *100 mM Tris-HCl, pH 8.0; 30 mM reduced glutathione*

1. Resuspend the bacterial pellet from a 100 mL culture in 10 mL chilled bug lysis buffer containing protease inhibitor and 1 mM PMSF.
2. Add 100 μL of 50 mg/mL lysozyme solution to a final concentration of 0.5 mg/mL, leave on ice for 30-45 min, invert occasionally.
3. Sonicate with Branson Digital Sonifier at an output of 18% for 20 s. (1 s on, 1 s off), put back tube on ice. Repeat once.
4. Centrifuge at 12000 rpm at 4˚C for 15 min.
5. After spinning the lysed bacteria, transfer supernatant to 15 mL conical tube, and add 200 μL of glutathione beads equilibrated with PBS.
6. Rotate at 4˚C for 2-4 hours (can leave for overnight).
7. Wash the beads twice with bug lysis buffer, and once with 100 mM Tris-HCl (pH 8.0).

8. Re-suspend the beads in 1 mL of 100 mM Tris-HCl (pH 8.0) and transfer to a 1.5 mL Eppendorf tube.
9. Spin at 500 g at 4˚C for 5 min. Remove supernatant as much as possible.
10. Add 350 µL of elution buffer, rotate at 4˚C for 4 hours or overnight.
11. Spin at at 500 g at 4 ˚C for 5 min. Transfer 330 µL supernatant into a new 1.5 mL Eppendorf tube.
12. Spin the supernatant again at 13000 rpm at 4 ˚C for 10 min, transfer 300 µL supernatant into a new tube.
13. Determine protein concentration by Bio-Rad Bradford assay.
14. The purified protein can be used freshly for up to a week. For long term storage, add 200 µL of 50% glycerol to final concentration of 15-20%, make 30-50 µL aliquots depending on protein concentration and snap freeze with liquid nitrogen, store at -80 ˚C.

2.2.7 Peptide microarray assay

Binding buffer: *50 mM Tris-HCl, pH 7.5; 250 mM NaCl; 0.1% (v/v) NP-40; 1 mM PMSF; 20% (v/v) bovine serum albumin*

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, the slides were incubated with the GST-PHF20-PHD in binding buffer overnight at 4°C with gentle agitation. After being washed with the binding 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).

2.2.8 Pull-down assay of biotin-labeled histone peptides

1. Thaw the proteins GST-PHF20-PHD on ice. Spin at 13000 rpm for 10 min, and carefully take proteins from the top of the solution.
2. Dilute proteins for binding: each binding requires 1 µg of GST-tagged protein in 300 µL of binding buffer. Prepare an extra tube as input.
3. Add 1 µg different biotinylated histone peptides (1 mg/mL) into each tube.
4. Rotate at 4^{o}C for 4 hours or overnight.

5. Prepare the Streptavidin Sepharose beads. Use 12-15 μL Streptavidin beads for each binding assay. Invert the bottle gently to mix the beads well before taking beads. Use a pipette to take sufficient slurry and transfer to a 15 mL Falcon tube, bring the volume to 10 mL with cold binding buffer, spin the beads at 500 g for 3 min. Remove the supernatant and wash the beads at least two more times with binding buffer.
6. Add 30 μL of 50% slurry into each tube. Rotate at 4˚C for 1 hour.
7. Spin at at 500 g for 3 min, and carefully remove the supernatant.
8. Wash the beads with 1 mL of binding buffer for 4 times. Rotate at 4 ˚C for 5 min for each wash.
9. Resuspend beads in 60 μL of 1.5× SDS sample buffer. Boil it and ready for SDS-PAGE. For 10% input, take 30 μL of samples from the input tube, add 30 μL 2× SDS sample buffer and boil. Load 10 μL of boiled sample for each Western.

2.2.9 Western blotting

Proteins separated by SDS-PAGE (2.4.1) were immobilized on polyvinylidene fluoride (PVDF) membranes by the Western blot method, which allows the detection of proteins by specific antibodies.

2.2.9.1 SDS-PAGE gel preparation

Recipe of making SDS-PAGE

	SDS-PAGE	**7%**	**10%**	**12%**	**15%**
Running Gel 20 mL (For 4×1mm plates)	4 × buffer pH 8.8 (mL)	5	5	5	5
	40% Acr-Bis (mL)	3.5	5	6	7.5
	ddH_2O (mL)	11.5	10	9	7.5
	10% APS (μL)	100	100	100	100
	TEMED (μL)	20	20	20	20

	SDS-PAGE	4%
Stacking Gel 10 mL (For 6-8 plates)	4 × buffer pH 6.8 (mL)	2.5
	40% Acr-Bis (mL)	1
	ddH_2O (mL)	6.5
	10% APS (µL)	100
	TEMED (µL)	20

2.2.9.2 *Loading and running the gel*

Load equal amounts of protein into the wells of the SDS-PAGE gel, along with molecular weight marker. Load 20–30 µg of total protein from cell lysate or 10–100 ng of purified protein. Run the gel for 1–2 hour at 100 V.

2.2.9.3 Transfer

1. During electrophoresis, cut the membrane to the dimensions of the gel. Pre-wet PVDF membrane in 100% methanol for few seconds, and equilibrate in transfer buffer for 10 min before blotting. Completely saturate the filter paper by soaking in transfer buffer.
2. After electrophoresis, equilibrate gel in 1× transfer buffer for 10 -15 min.
3. Setup the transfer sandwich: 3 pre-soaked filter paper, gel, PVDF membrane, and 3 pre-soaked filter papers. Roll a pipet or test tube over the surface of the filter paper to exclude all air bubbles.
4. Set up the transfer condition: 15 V for 1 hour with SDS in transfer buffer.

2.2.9.3 Blotting

1. Following transfer, rinse the membrane with TBST buffer. Block the membrane for 1 hour at room temperature using blocking buffer.
2. Incubate the membrane with appropriate dilutions of primary antibody overnight incubation at 4°C in blocking buffer.
3. Wash the membrane with TBST for three times, 10 min each.

4. Incubate the membrane with the recommended dilution of horseradish peroxidase (HRP) labeled secondary antibody in blocking buffer for 30 min at room temperature.
5. Wash the membrane with TBST for three times, 10 min each.

2.2.9.3 Detection

For signal development, ECL western blotting detection reagents were used follow the kit manufacturer's recommendations.

1. Mix an equal volume of detection solution 1 with detection solution 2 allowing sufficient total volume to cover the membranes. The final volume required is 0.125 mL/cm^2 membrane.)
2. Drain the excess wash buffer from the washed membranes and place them (Notes: protein side up) on a sheet of SaranWrap.
3. Pipette the mixed detection reagent on to the membrane. Incubate for 1 min at room temperature.
4. Drain off excess detection reagent by holding the membrane gently with forceps and touching the edge against a tissue. Place the blots protein side down on to a fresh piece of SaranWrap, wrap up the blots and gently smooth out any air bubbles.
5. Place the wrapped blots, protein side up, in an X-ray film cassette.
6. Exposure for 5-10 s, 30 s, 1 min, 5-10 min each.

2.2.10 Quantitative real-time PCR and RNA-Seq analysis

2.2.10.1 RNA extraction

Total RNA was isolated from cells using the RNeasy Plus kit (QIAGEN) according to the manufacturer's instruction. Adherent cells grown in a 10 cm cell-culture dish were directly lysed by adding 600 µL RLT buffer. Cell lysates were directly transferred into a QIAshredder spin column placed in a 2 mL collection tube, and centrifuged for 2 min at maximum speed. The homogenized lysate was transfered to a gDNA Eliminator spin column placed in a 2 mL collection tube, and centrifuged for 30 sec at maximum speed. After adding 600 µL 70 % (v/v) ethanol to the flow-through, the samples were loaded on RNeasy spin columns, the bound RNA was washed with RW1 buffer and RPE buffer. The RNA was eluted with 30 µL

RNase-free water. RNA concentration and purity was determined in a spectrometer, and RNA was stored at –80 °C.

2.2.10.2 cDNA synthesis

The complementary DNA (cDNA) was synthesized by reverse transcription PCR using the iScript cDNA Synthesis Kit (Bio-Rad). The reaction was setup by adding flowing reagents into a PCR tube.

Reagent	**Volume (µL)**
5 × iScript Reaction mix	4
iScript Reverse Transcriptase	1
Nuclease-free water	15-x
RNA template (1 µg)	x

Incubate the complete reaction mix in a thermal cycler using the following protocol.

Step	**Temperature (°C)**	**Time**
Priming	25	5 min
Reverse Transcription	46	20 min
RT inactivation	95	1 min
Hold	4	indefinitely

After the reaction, the cDNA was diluted with sterile water to a volume of 100 µL and used as a template for quantitative real-time PCR.

2.2.10.3 Quantitative real-time PCR

Quantitative real-time RT-PCR was performed using the Power SYBR Green PCR Master Mix (Applied Biosystems). The reaction was setup by add the following components to a 96 well real time PCR plate.

Reagent	Volume (µL)
Power SYBR® Green PCR Master Mix (2×)	5
Forward Primer (0.25 µM)	1
Reverse Primer (0.25 µM)	1
Template cDNA (10 ng/µL)	2
Water	1

Quantitative real-time RT-PCR was performed on an ABI 7500-FAST Sequence Detection System using the following program:

Step	Temperature (°C)	Time	
Enzyme activation	95	10 min	
Denature	95	15 s	40 PCR Cycles
Anneal/ Extend	60	1 min	
Melt curve stage			

All experiments were performed in triplicates and gene expression was calculated after normalization to GAPDH cDNA levels using the comparative Ct (cycle threshold) method. The primer sequences used for RT-PCR are listed in Table 2.2. 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.

2.2.10.4 RNA-Seq Analysis

RNA-seq samples were sequenced using the Illumina Hiseq 2000. Three biological replicates were prepared for each condition. The sequencing reads were mapped to the human genome (hg19) by TopHat (version 2.0.10) [60]. The overall mapping rate was 95%–99%. The number of fragments in each known gene from RefSeq database [61] (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 [62] (version 1.18.0). Genes with false discovery rate ≤ 0.05 and fold change ≥ 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.

2.2.11 Cross-linking chromatin immunoprecipitation (X-ChIP)

2.2.11.1 ChIP buffer preparation

Cell Lysis Buffer: *5 mM PIPES pH 8.0; 85 mM KCl; 1% NP40; Add Protease Inhibitor and PMSF freshly.*

Nuclei Lysis buffer: *50 mM Tris-HCl pH 8.0; 10 mM EDTA; 1% SDS; Add Protease Inhibitor and PMSF freshly.*

Dilution buffer: *20 mM Tris-HCl pH 8.0; 150 mM NaCl; 1 mM EDTA; 1% Trition X 100; 0.01% SDS; Add Protease Inhibitor and PMSF freshly*

Low salt wash buffer: *20 mM Tris-HCl pH 8.0; 150 mM NaCl; 2 mM EDTA; 1% Trition X 100; 0.1% SDS*

High salt wash buffer: *20 mM Tris-HCl, pH 8.0; 500 mM NaCl; 2 mM EDTA; 1% Triton X-100; 0.1% SDS*

LiCl wash buffer: *20 mM Tris-HCl, pH 8.0; 250 mM LiCl; 1 mM EDTA; 1% NP40; 1% Na-deoxycholate*

TE buffer: *10 mM Tris-HCl 8.0; 1 mM EDTA*

ChIP Elution buffer: *50 mM $NaHCO_3$; 1% SDS. (Make freshly!)*

2.2.11.2 Procedure

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 cell lysis buffer for 20 min at 4 °C. The isolated nuclei were resuspended in nuclei lysis buffer 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 in low salt wash buffer, twice in high salt wash buffer, once in LiCl wash buffer. Reverse crosslinking was performed by incubation at 67 ℃ for 12 hours in the presence of 0.3 M NaCl. The DNA was eluted and purified using a PCR purification kit (QIAGENe).

2.3 Methods in cell biology

2.3.1 Cell culture

All cell lines were incubated at 37 °C in an atmosphere of 5 % CO_2. 293T cells, normal lung cell line WI-38 and IMR90 were cultured in DMEM supplemented with 10% fetal bovine serum. H1792 and all lung cancer cell lines were cultured in RPMI 1640 supplemented with 10% fetal bovine serum. Cells were grown in 10 cm cell culture dishes and routinely split every 2–3 days. For that purpose, cells were washed with PBS and incubated with 1 mL 0.25% trypsin-EDTA for 1–3 min at 37 °C. Detached cells were washed off by adding 5 mL fresh complete medium, which stops the trypsin reaction, and seeded into a new dish at an appropriate dilution.

2.3.2 Protein extraction and concentration measurement

2.3.2.1 Cell Lysate Collection (For 10 cm dish of adherent cells)

1. Place the cell culture dish on ice and wash the cells with cold PBS.

2. Suck the PBS, then add 1 mL of PBS-containing 1 mM PMSF. Scrap the cells off and transfer to 1.5 mL Eppendorf tubes. (Keep on ice from now on)
3. Spin at 1000 g at 4°C for 3 min.
4. Aspirate the supernatant. The cell pellets can be submitted to making cell lysate immediately or snap frozen in liquid nitrogen and stored at -80°C.
5. Add x μL of cell lysis buffer (x=5 times the volume of cell pellets), resuspend the cell pellets by pipette up and down for several times. Briefly vortex for 5 s, then put on ice for 20 min.
6. Break the cells by sonication under the following condition: 15% output, 20 burst, 0.5 sec on, 0.5 sec off.
7. Centrifuge at 12000 rpm for 10 min at 4°C, and supernatants were transferred into a fresh tube.

2.3.2.1 Protein concentration measurement (Dc Protein Assay kit).

According to the instruction, microplate assay protocol was used to measure the concentration.

1. Preparation of working reagent. Add 20 μL of reagent S to each mL of reagent A that will be needed for the run.
2. Prepare 3-5 dilutions of a protein standard containing from 0.2 mg/mL to about 1.5 mg/mL protein. For best results, the standard should be prepared in the same buffer as the sample.
3. Pipet 5 μL of standards and samples into a clean, dry microplate.
4. Add 25 μL of working reagent A into each well.
5. Add 200 μL reagent B into each well. Gently agitate the plate to mix the reagents.
6. After 15 min, absorbance can be read at 750 nm.
7. Based on the standard curve, calculate the protein concentration and adjust the concentration to 1mg/mL. Add 40 mL of 5× SDS loading buffer into 200 mL 1mg/mL cell lysate, boil for 10 min.

2.3.3 Lenti-virus packaging

The lentiviral systems were used to introduce stable knockdown or overexpression of *PHF20* gene in lung cancer cell line via short-hairpin RNA (shRNA) and complementary DNA (cDNA) expression.

2.3.3.1 Plate cells (Day 0)

1. 24 hours before transfection, plate 293T cells in 10 cm cell culture dish. Cells should be 40–70% confluent the next day prior to transfection.
2. Incubate the cell cultures overnight at 37 °C in an atmosphere of 5 % CO_2.

2.3.3.2 Transfection (Day 1)

1. Warm Opti-MEM reduced serum medium at 37℃ and FuGENE HD transfection reagent to room temperature. Vortex FuGENE HD transfection reagent briefly and gently before using.
2. Place 500 µL of Opti-MEM reduced serum medium in a sterile 1.5 mL Eppendorf tube. Add 10 µg plasmid DNA (4 µg shRNA or cDNA: 4 µg pPAX: 2 µg PMD2.G) into the Opti-MEM medium, pipet gently to mix completely.
3. Place 500 µL of Opti-MEM in another 1.5 mL Eppendorf tube, add 24 µL FuGENE HD transfection reagent. Vortex briefly or pipet gently to mix completely.
4. Combine the plasmid solution and FuGENE HD transfection reagent solution together. Incubate at room temperature for 20–30 min to allow sufficient time for complex formation. Inverse gently a few times during incubation.
5. Add the mix into a 10 cm dish cells in a drop-wide fashion, gently shake the dish to mix well.
6. Incubate the cells at 37℃ and 5% CO_2.

2.3.3.3 Virus collection (Day 3)

The virus was collected 2 days after transfection. The medium with virus was collected using 10 mL syringe and filtered through 0.45 µm filters. The virus can be stored at 4℃ for short time use or split into into 2 mL aliquot which can be frozen at -80 ℃ for long-term storage.

2.3.4 Virus transduction

2.3.4.1 Plate cells (Day 0)

H1792 cells were plated into 10 cm cell culture dishes 24 hours before virus infection. The cells should be around 40-70% confluent the next day.

2.3.4.2 Virus infection (Day 1)

The next day, remove media and add 6 mL of new medium containing 64 μg polybrene. Add 2 mL virus into each plate, and the final concentration of polybrene is 8 μg/mL. After 4-6 hours, add 4 mL of complete medium into each plate.

2.3.4.3 Medium replacement (Day 2)

24 hours after infection, change medium or split if necessary.

2.3.4.4 Selection with drugs (Day 3)

To select the infected cells, the proper drugs were added into the medium. For H1792 cells, the proper concentration of puromycin is 1.5 μg/mL, while the proper concentration of blasticidin is 10 μg/mL.

2.3.4.5 Stable cell line maintenance

Change medium every 1-2 days or split cells when necessary. In 2-3 days all non-transduced cells should have been killed and the stable cell lines can be maintained in culture for at least 2 months.

2.3.5 Colony formation assay

Colony formation assay is an in vitro cell survival assay based on the ability of a single cell to grow into a colony [63]. As described previously, 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 cut off of 75 μm. Results were quantitated from at least three independent replicates.

2.3.6 Soft agar assay

1.2% agarose: weigh 1.2 g agarose and add 100 mL of ddH_2O, autoclave and mix well to make a uniform gel solution after autoclaving. This can be stored at room temperature for a couple of months.

2× RPMI 1640 medium: make from RPMI 1640 powder (Gibco Cat. No. 51800019) according to the manufacturer's instruction. One bag of powder is for making 1L of 1× RPMI

1640, so only use it to make 500 mL of 2× RPMI 1640. Double all the supplements you need to make them 2×.

2.3.6.1 Preparation of base agarose layer

All steps must be done sterilely and use cell culture grade water

1. Melt 1.2% agarose in microwave. Let it cool down in a 42°C water bath, and allow at least 30 min to equilibrate the temperature.
2. Using conical tubes, warm 2× complete RPMI1640 containing 20% FBS and 2× all the other supplements to 42°C in water bath. Allow at least 30 min for temperature to equilibrate.
3. Quickly but thoroughly mix equal volumes of the two solutions to give 0.6% agarose + 1× RPMI 1640 + 10% FBS.
4. Quickly add 1.5 mL of mixture from step 3 to each well of 6-well plate, swirl a bit on the bench to allow the gel solution to completely cover the culture surface. Put the plates aside for 5 min to allow agarose to solidify.

2.3.6.2 Preparation of top agarose

1. Melt 1.2% agarose in a microwave. Mix well and aliquot 2.5 mL into each 15 mL conical tubes and put them in a 42°C water bath. Allow about 30 min for temperature to equilibrate.
2. Put an aliquot of complete 2× RPMI 1640 medium in the 42°C water bath.
3. Trypsinize H1792 cells, collect dis-attached cells and spin down, resuspend pellet in suitable volume of medium to make the cell density fall in the range of 5×10^5-1×10^6 cells/mL. Count cells to determine the exact cell density.
4. Calculate the volume of cell suspension that will yield the number of cells you want to seed in one well. For each cell density, seed three wells as replica. Calculate the volume of every reagent as in Table 2.4 to make the top agarose-cell mixture layer for three wells.
5. Mix the cell suspension prepared in step 4 with 1 mL of pre-warmed 2× complete RPMI 1640 medium and 1 mL of 1.2% agarose quickly and thoroughly. Then add 1 mL of this mixture to each well and let it spread evenly on top of the base agarose.
6. Allow the top agarose-cell layer to solidify for 30 min at RT.

7. Add 1-1.5 mL of regular medium to the well, and put them back in the 37°C incubator.
8. Change medium every 4 days. Culture them until colonies can be seen with naked eye.
9. Colony numbers can be scored under light microscope, adjust the depth of the field by changing the focus to count 8-10 three-dimensional fields.

Table 2.4 Recipe of top agarose

No. of cells be seeded in each well	2×10^4	5×10^4	1×10^5
Vol. of cell suspension			
Vol. of cells for triplicate			
Vol. of 1×DMEM for triplicate			
Vol. of 2×DMEM for triplicate	1 mL	1 mL	1 mL
Vol. of 1.2% agarose for triplicate	1 mL	1 mL	1 mL

Vol. of cells for triplicate=3.5× vol. of cell suspension

Vol. of 1×DMEM for triplicate= (0.4-vol of cells) ×3.5

2.4 Methods in structural biology

2.4.1 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 (aa1–12) 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 [64], and the model was built with Coot [65].

2.4.2 NMR titration experiments

The $^{1}H,^{15}N$ HSQC spectra of 0.1 mM uniformly ^{15}N-labeled WT or mutant PHD finger 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. 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_ds 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, Δδ 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 Δδ is the change in chemical shift in parts per million (ppm).

2.4.3 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 and NMRViewJ. 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 were recorded with a 120-ms mixing time to unambiguously identify intermolecular nuclear Overhauser enhancement (NOEs).

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 and CSI analysis. Two hundred structures of free PHF20 were initially calculated using the simulated annealing protocol of CYANA. Twenty of these structures with the lowest energies were then used to analyze for violations and obtain additional SANE-assisted NOE assignments for PHF20. 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. In AMBER refinement, the structures were subjected to 30 ps 30,000 steps of simulated annealing using a generalized Born solvent model. 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.

2.4.4 Fluorescence spectroscopy

Spectra were recorded at room temperature on a Fluoromax-3 spectrofluorometer (HORIBA). The samples containing the PHF20 PHD finger 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, ΔI is the observed change of signal intensity, and ΔI_{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.

2.5 Online databases and analysis

2.5.1 GLOBOCAN project and online analysis

GLOBOCAN is a project of the International Agency for Research on Cancer, and provides contemporary estimates of the incidence, mortality and prevalence from major types of cancer at national level. These estimates are based on the best available data in each country and on information publically available on the Internet. In order to obtain the incidence and mortality of lung cancer worldwide and in Germany, the online analysis tools (http://globocan.iarc.fr/Pages/online.aspx) of GLOBOCAN project

(http://globocan.iarc.fr/Default.aspx) were used. Pie charts were used to present the incidence and mortality of lung cancer by population.

2.5.2 TCGA database and cBioPortal for Cancer Genomics

The Cancer Genome Atlas (TCGA) project is supervised by the National Cancer Institute's Center for Cancer Genomics and the National Human Genome Research Institute funded by the US government. TCGA program applies high-throughput genome analysis techniques to generate, analyze, and made available genomic sequence, gene expression, gene methylation, and copy number variation (CNV) data. The aim is to improve our ability to diagnose, treat, and prevent cancer through a better understanding of the genetic basis of cancer. TCGA (https://cancergenome.nih.gov/) is one of the largest and most commonly used public resource to date.

The cBioPortal for Cancer Genomics was originally developed at Memorial Sloan Kettering Cancer Center (MSK). The public cBioPortal site (http://cbioportal.org) is hosted by the Center for Molecular Oncology at MSK. The cBioPortal software enables interactive, exploratory analysis of large-scale cancer genomics data sets including TCGA database.

To investigate *PHF20* gene alterations in human cancers, the cBioportal software was used to analyze *PHF20* gene mutation and CNV mainly within TCGA database. Briefly, on the cBioPortal site (http://cbioportal.org), TCGA studies were selected and "Mutation and CNV" was chosen for "Data Type Priority", then enter *PHF20* gene and submit query. The result was downloaded and used directly.

2.5.3 Kaplan-Meier plotter analysis

Kaplan-Meier (KM) Plotter (http://kmplot.com/analysis/) is a database that integrates gene expression data and clinical data. Till 3rd December 2017, the KM Plotter is capable to assess the effect of 54,675 genes on survival using 10,461 cancer samples, which include 5,143 breast, 1,816 ovarian, 2,437 lung and 1,065 gastric cancer.

To determine the correlation between expression level of *PHF20* and the overall survival (OS) of lung cancer patients, KM plotter analysis was carried out. Briefly, on the KM Plotter site (http://kmplot.com/analysis/), Click "Start KM Plotter for lung cancer" from mRNA project, and goes into lung cancer specific website. Set the parameters as follows,

Affy id/Gene symbol: 1569906_s_at (Input PHF20 and select 1569906_s_at)
Split patients by: median
Survival: OS
Set other parameters as defaults.

Then draw Kaplan-Meier Plot and download the result directly. To determine the correlation with different types of lung cancer, set the parameters as before, and set Histology as adenocarcinoma or squamous cell carcinoma.

3 RESULTS

3.1 PHF20 is required for cancer cell growth and survival

3.1.1 PHF20 gene is amplified in human cancers including lung cancer

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 3.1).

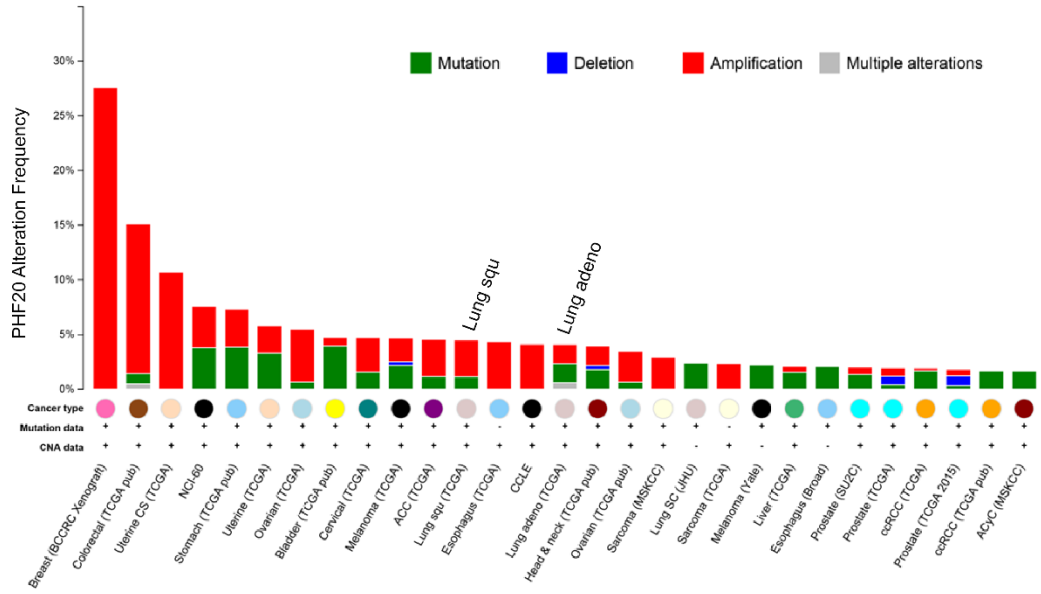


Figure 3.1 *PHF20* is amplified in human cancers. The gene alterations frequency data is obtained from the TCGA database.

In order to investigate whether *PHF20* expression level correlates with patient survival, we utilized Kaplan-Meier survival plot to analyze the correlation between *PHF20* expression and lung cancer. It was shown that high expression of *PHF20* was significantly correlated with lower survival rate (P=1.1e-05, hazard ratio (HR) =1.33 with 95% confidence interval 1.17–1.51) in lung cancer patients (Figure 3.2). Moreover, this correlation only existed in lung adenocarcinoma patients (P=1.2e-06, Figure 3.3 left), but not in lung squamous carcinoma (P=0.46, Figure 3.3 right).

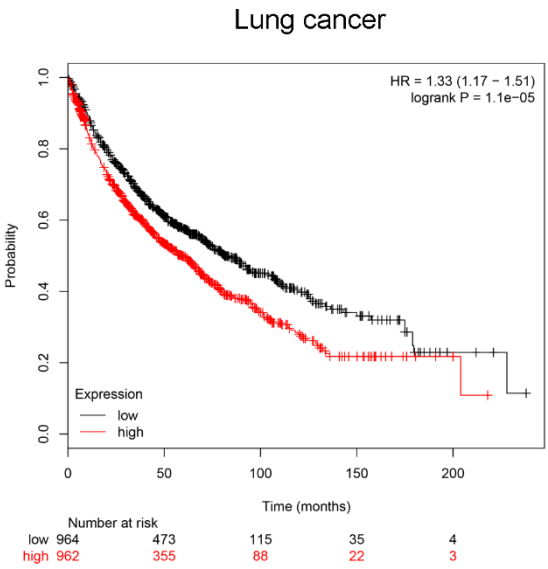


Figure 3.2 Kaplan–Meier analysis of survival of lung cancer patients with low (black) or high (red) *PHF20* expression. The number of surviving patients at 0-, 50-, 100-, 150- and 200- month time points is indicated below the graph.

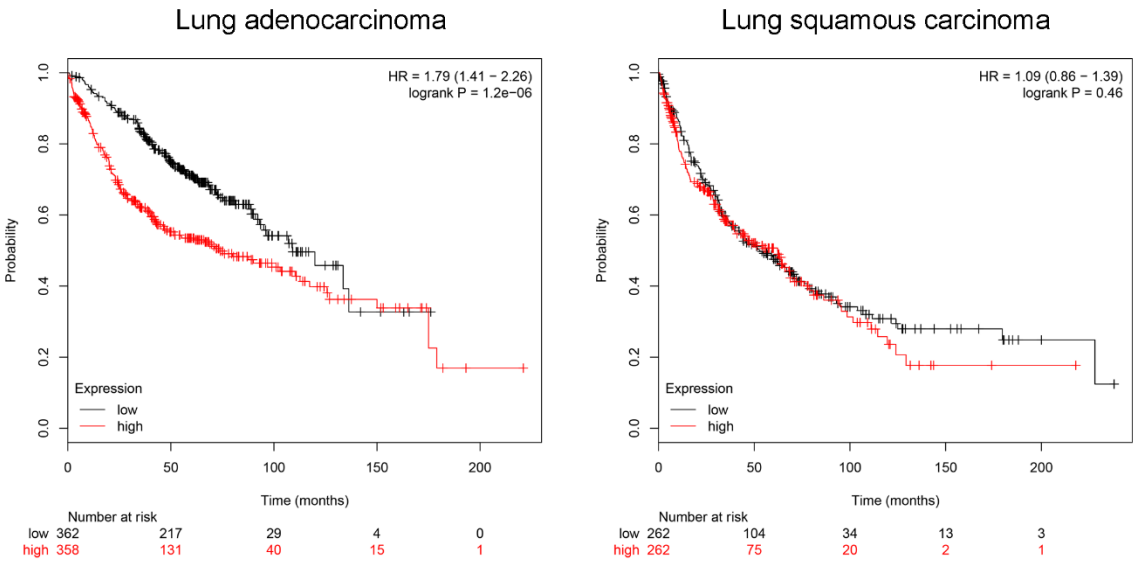


Figure 3.3 Kaplan–Meier analysis of survival of lung adenocarcinoma patients (left) and lung squamous carcinoma (right) with low (black) or high (red) *PHF20* expression. The number of surviving patients at 0-, 50-, 100-, 150- and 200- month time points is indicated below the graph.

3.1.2 PHF20 is overexpressed in lung adenocarcinoma cell lines

In order to study the biological function of PHF20 in lung adenocarcinoma, we firstly determine the *PHF20* gene expression and protein levels across a number of lung adenocarcinoma cell lines together with normal lung fibroblast cell lines. Our qRT-PCR and immunoblotting results showed that, compared with the normal lung fibroblast cells (WI-38 and IMR90), *PHF20* is overexpressed in all lung cancer cell lines that we examined at protein levels (Figure 3.4B).

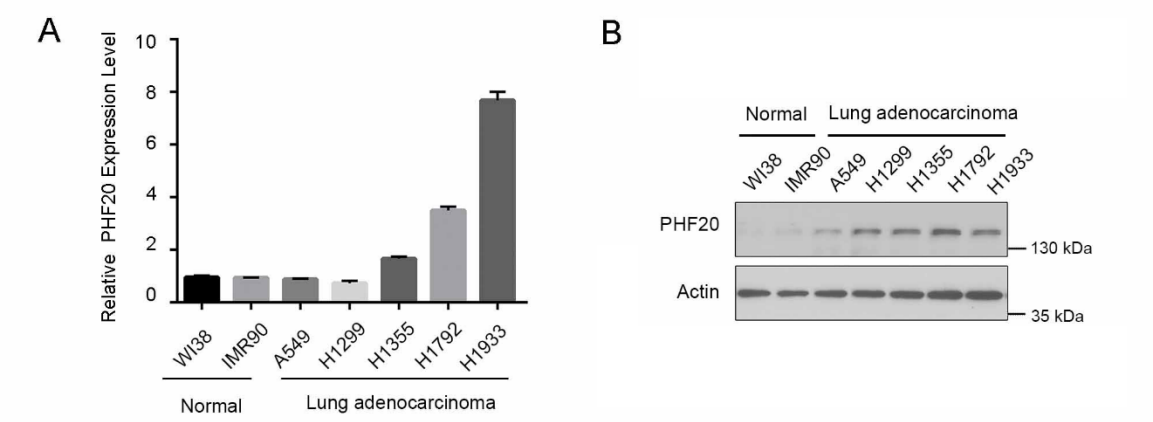


Figure 3.4 *PHF20* expression is amplified in lung adenocarcinoma cell lines. qRT-PCR (A) and western blot analysis (B) of *PHF20* gene expression and protein levels, respectively, in the indicated lung adenocarcinoma cell lines and control immortalized normal cells. GAPDH expression was used as internal control in (A), and actin was used as a loading control in (B). Error bars represent SEM of three replicates.

3.1.3 Depletion of PHF20 inhibits cell proliferation and division in H1792 cells

To further evaluate the contribution of the up-regulation of *PHF20* to the oncogenesis of lung adenocarcinoma, we knocked down *PHF20* gene expression in H1792 lung adenocarcinoma cells using lentiviral *PHF20*-specific short hairpins (shRNAs) constructs. qRT-PCR and immunoblotting analyses were performed to assess the efficiency of shRNAs. Compared with control non-targeting shRNA (*shNT*), both *shPHF20 #1* and *shPHF20 #2* effectively knocked down endogenous *PHF20* expression at mRNA level and protein level (Figure 3.5).

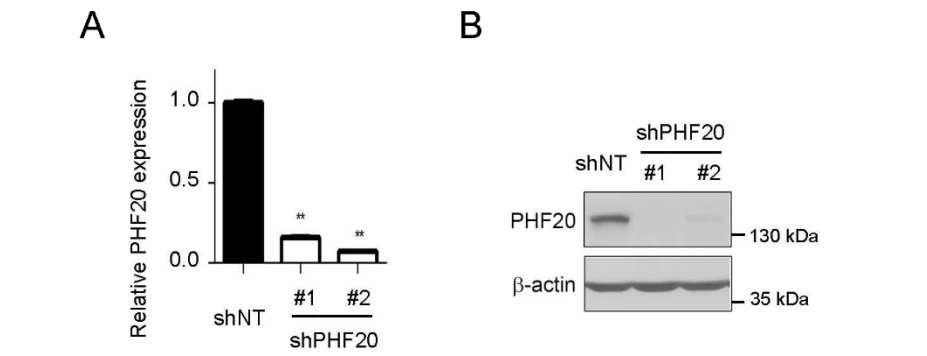


Figure 3.5 *shPHF20* effectively knocked down endogenous *PHF20* expression. (A) qRT-PCR analysis of *PHF20* knockdown in H1792 cells. **p<0.01 in comparison with control; (B) Western blot analysis of *PHF20* knockdown in H1792 cells.

Interestingly, we found that the cells notably grow slower after *PHF20* depletion. To confirm the effect of PHF20 on cell growth, we performed cell proliferation assay, which showed that depletion of *PHF20* markedly inhibited the cell growth of H1792 cells. (Figure 3.6 left). Similarly, knockdown of *PHF20* expression obviously reduce cell growth in H1299 (Figure 3.6 middle) and H1993 cells (Figure 3.6 right).

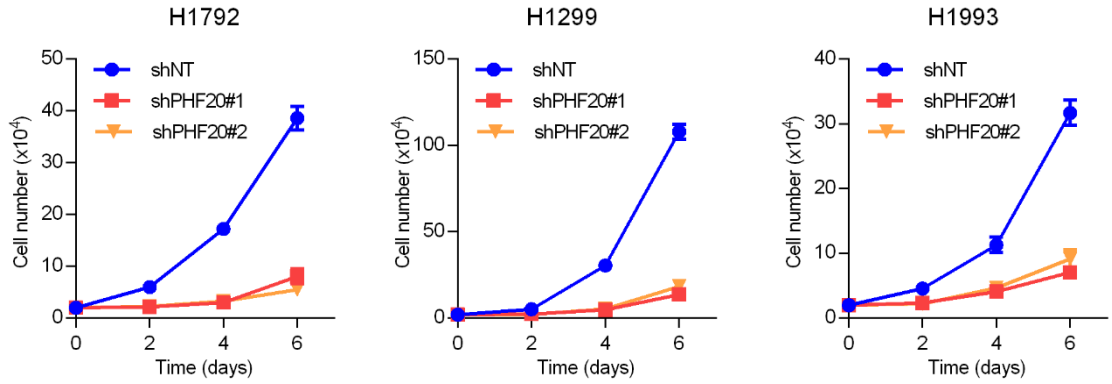


Figure 3.6 Cell proliferation assays of the control and *PHF20* KD H1792 cells (left), H1299 cells (middle) and H1993 cells (right). Live cells were counted over a 6-day time course. Error bars represent SEM of six replicates.

To determine the effect of PHF20 on the cancer cell division, we performed the clonogenic assay. The control and *PHF20* KD H1792 cells were plated at low density, and grow under normal culture condition. The cell colonies were fixed, stained, and counted two weeks later.

As shown in Figure 3.7A, the *PHF20* KD H1792 cells formed less colonies compared with control cells. The colonies were counted and summarized in Figure 3.7B.

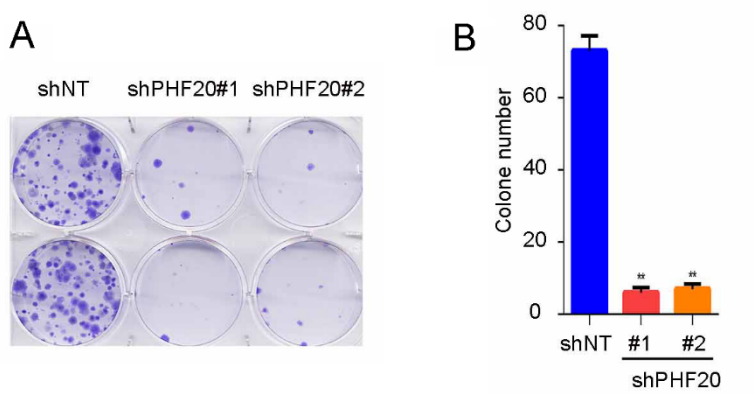


Figure 3.7 Colony formation assays of the control and *PHF20* KD H1792 cells. Representative crystal-violet-stained cells are shown in (A), and quantification is shown in (B). Error bars represent SEM of six replicates. **p < 0.01.

3.1.4 Depletion of PHF20 inhibits colony formation in H1792 cells

The ability to exhibit anchorage-independent growth of cancer cells in vitro has been considered as a key aspect of the tumor phenotype, particularly with respect to metastatic potential [66]. To further demonstrate the effect of PHF20 on tumor cell aggressiveness, we performed the soft agar colony formation assay in control and *PHF20* depleted H1792 cells. As expected, compared with well-developed colonies observed in control cells, *PHF20*-depleted cell colonies were fewer and relatively smaller in size (Figure 3.8).

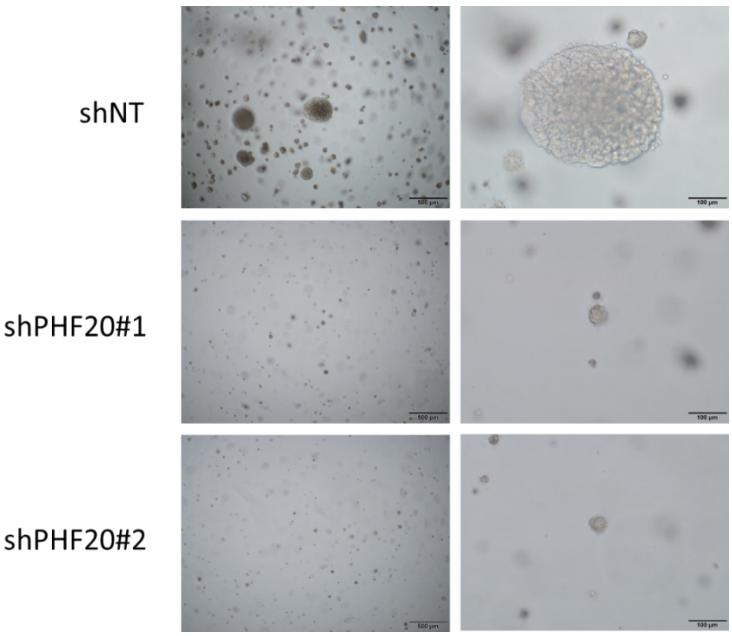


Figure 3.8 Soft agar assays in the control and *PHF20* KD H1792 cells.

PHF20 is upregulated in several human cancers including lung cancer, indicating *PHF20* is a candidate oncogene. Depletion of *PHF20* significantly inhibited cell growth and colony formation. **These data suggest that PHF20 has an oncogenic role in lung adenocarcinoma, promoting cancer cell growth and survival.**

3.2 PHF20 regulates the expression of genes involved in cancer and cell cycle

3.2.1 RNA-seq in PHF20-depleted cells

PHF20 is a component of the MOF-NSL lysine acetyltransferase complex and is implicated in transcriptional regulation [48, 49, 51]. In order to study the molecular mechanism under which PHF20 promotes cell growth, firstly we utilized RNA sequencing (RNA-seq) analysis to identify PHF20-regulated genes genome-wide. We knocked down *PHF20* gene expression and performed RNA-seq analysis in H1792 lung adenocarcinoma cells. PHF20 was reported as a transcriptional coactivator [67]; however, we found similar numbers of genes were down- (942 genes) and upregulated (895 genes) in *PHF20* KD cells compared with the control cells. 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 up-regulated genes are enriched for cancer, organismal injury, and connective tissue diseases (Figure 3.9).

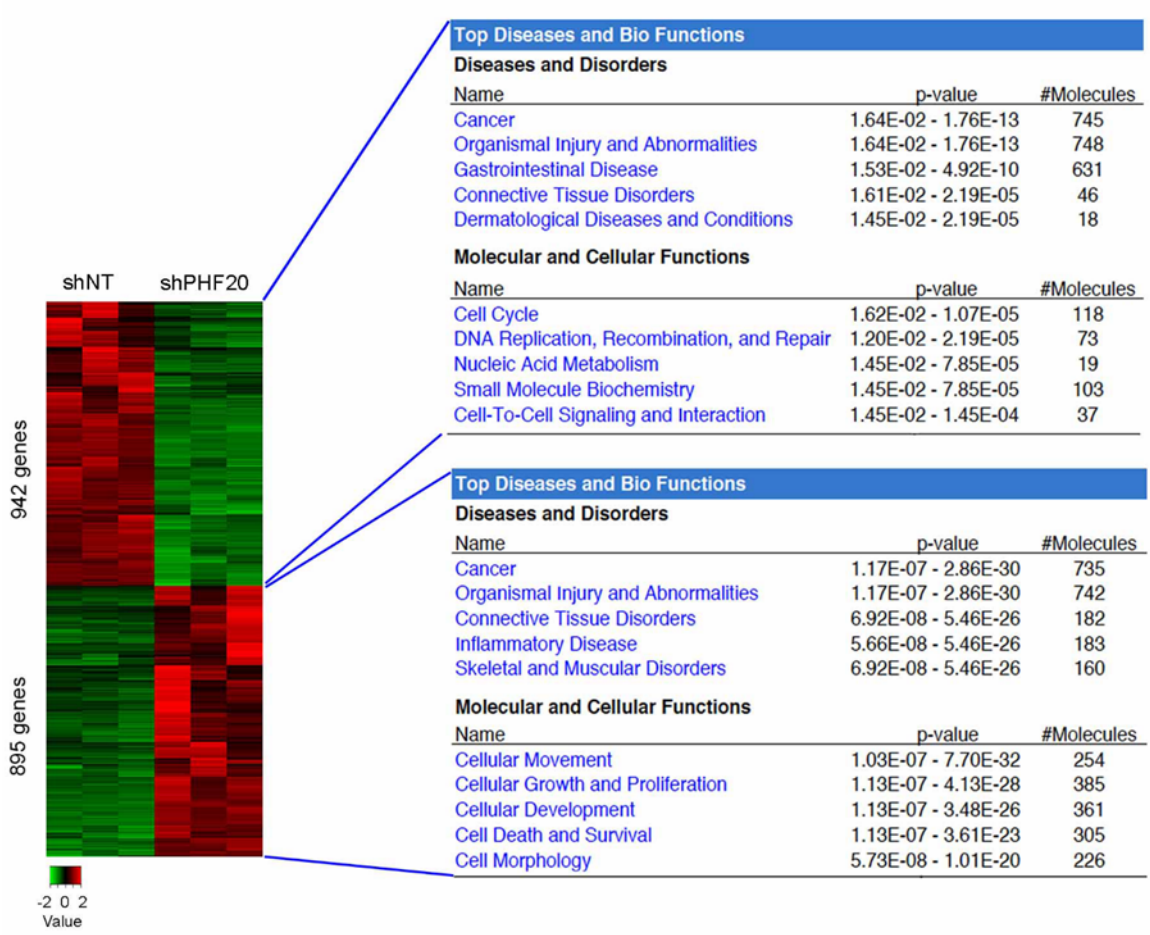


Figure 3.9 Heatmap and IPA of gene expression profiles in the control (*shNT*) and *PHF20* KD *(shPHF20)* H1792 cells. The green and red colors indicate down- and upregulation, respectively. The right panel shows the IPA of the down- and upregulated genes in *PHF20* KD cells. The top five hits in each category are listed.

3.2.2 Validation of RNA-seq results via qRT-PCR

In order to validate the RNA-seq results, we first randomly selected 16 downregulated genes in *PHF20* KD cells and performed qPCR analysis. All the genes tested showed considerable reductions of gene expression upon *PHF20* knockdown by two independent shRNAs (Figure

3.10). Our RNA-seq results can explain how PHF20 affect cancer cell growth and cell proliferation through regulating gene transcription.

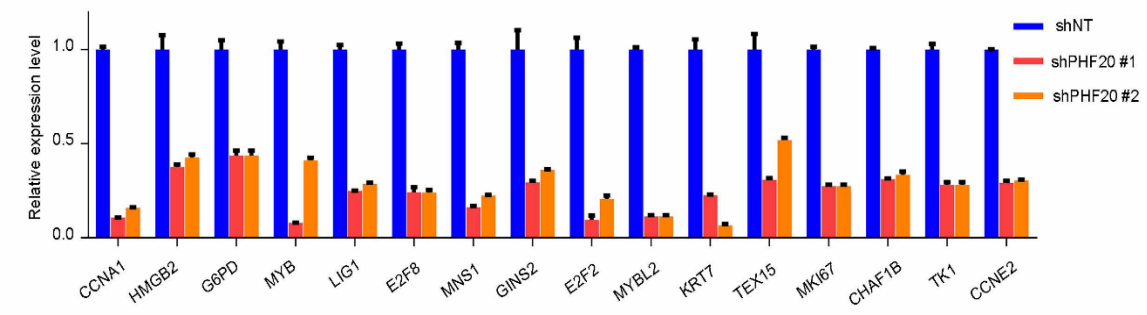


Figure 3.10 qRT-PCR analysis of gene expression in control (*shNT*) and *PHF20* KD (*shPHF20*) cells.

3.3 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 1.3) [40]. Biochemical and structural studies have shown that Tudor2 associates with dimethyllysine substrates [40, 45, 46], particularly with p53K370me2 and p53K382me2 [40]; however, biological activities of other domains, including the PHD finger, remain unknown. Because a set of PHD finger-containing proteins was found to bind histone sequences [41, 42, 68-70], we tested the PHF20 PHD finger in histone peptide microarrays, pull-down assays and nuclear magnetic resonance (NMR).

3.3.1 PHF20 PHD finger binds to H3K4me2 in histone peptide microarray assay

We used histone peptide microarray contains unmodified histone H3 and H4 peptides or histone peptides bearing single methylation marks to screen the PHF20 PHD finger. Unexpectedly, we found that the PHF20 PHD finger strongly binds to H3K4me2 peptide but does not recognize either H3K4me3 or unmodified H3 that were previously reported as ligands for PHD fingers. PHF20 PHD finger domain specifically recognized methylated H3 but not methylated H4 or methylated H2A (Figure 3.11).

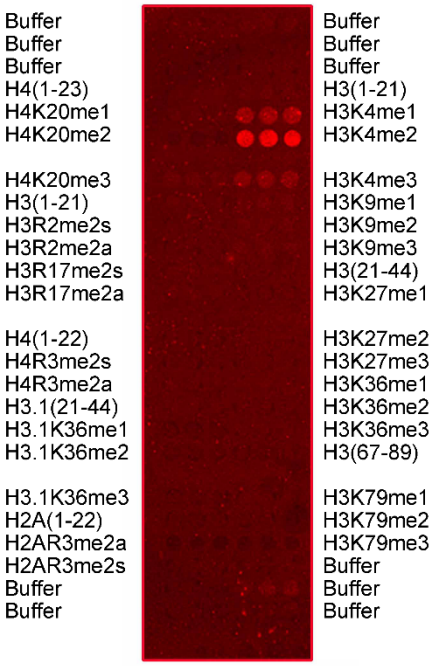


Figure 3.11 Histone peptide array probed with GST-PHF20 PHD.

3.3.2 PHF20 PHD finger specifically recognizes H3K4me2 in histone peptides pull-down assay

To confirm the specificity of the PHF20 PHD finger toward H3K4me2, we performed pull-down assays. The 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. The result was consistent with the peptides microarray assay, the PHF20 PHD finger specifically binds to H3K4me2 peptide (Figure 3.12).

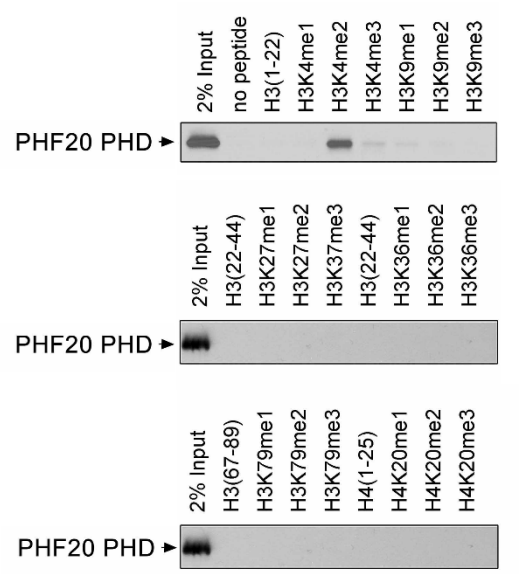


Figure 3.12 Western blot analysis of histone peptide pull-downs with GST-PHF20 PHD and the indicated biotinylated peptides.

3.3.3 HSQC titration experiments indicate direct binding between PHF20 PHD finger and H3K4me2

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 3.13, left). However, interaction with H3K4me3 was considerably weaker judging by small CSPs observed upon titration with this peptide (Figure 3.13, right).

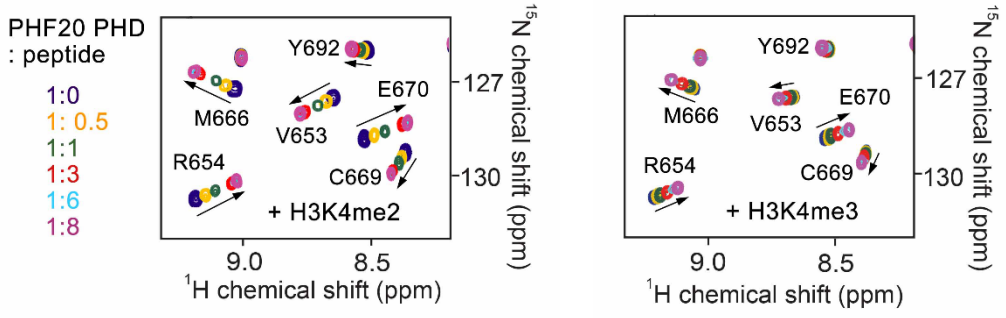


Figure 3.13 Superimposed $^{1}H,^{15}N$ HSQC spectra of the PHF20 PHD finger collected upon titration with the indicated peptides. Spectra are color coded according to the protein: peptide molar ratio shown on the left.

3.3.4 The PHF20 PHD finger binds to the H3K4me2 peptide with highest affinity in fluorescence spectroscopy assay

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 (Figure 3.14).

A

Binding affinity of the PHF20 PHD finger

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

B

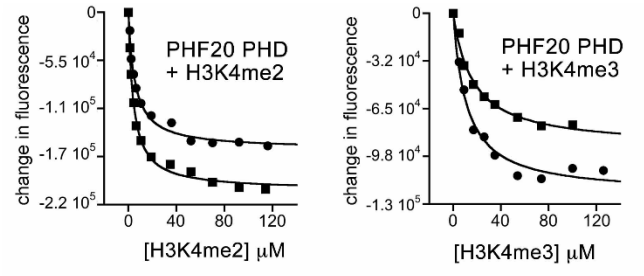


Figure 3.14 (A) Binding affinities of WT PHF20 PHD for the indicated histone peptides measured by tryptophan fluorescence. (B) Representative binding curves used to determine the K_d values.

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**.

3.4 Molecular basis for the recognition of H3K4me2

3.4.1 Solution structure of the PHD finger in complex with the H3K4me2 peptide

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 3.15; Tables 3.1 and 3.2). 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 3.15A). The histone H3K4me2 peptide is bound in an extended conformation in a branched negatively charged channel (Figure 3.15B). The peptide is aligned with β1 of the PHD finger, allowing for the formation of a three-stranded antiparallel β sheet (Figure 3.15A).

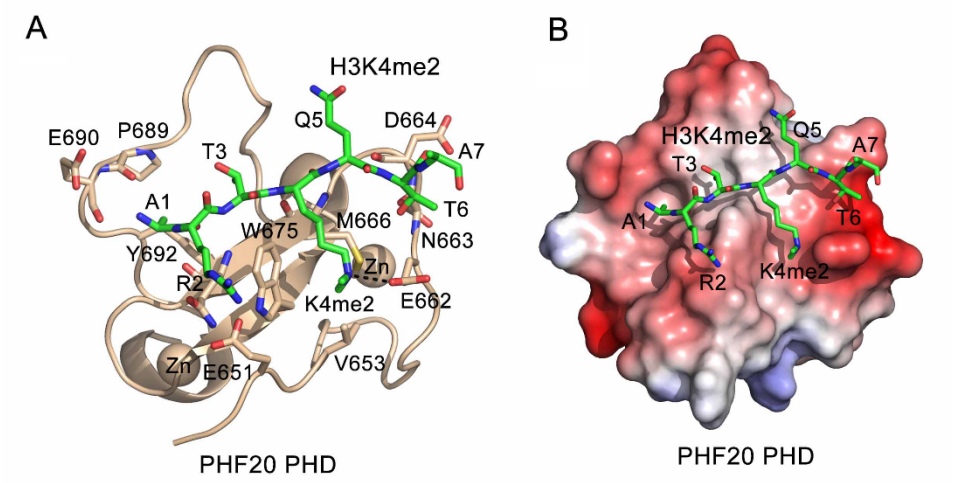


Figure 3.15 Solution structure of the PHD finger in complex with the H3K4me2 peptide. (A) The PHD finger is depicted in the ribbon diagram, and the H3K4me2 peptide is shown in a stick model. The histone peptide residues and the residues of the PHF20 PHD finger involved in the interaction are labeled. (B) The electrostatic surface potential of the PHF20 PHD finger is shown using blue and red colors for positive and negative charges, respectively.

Table 3.1 X-ray crystallographic data collection and refinement statistics.

	PHF20 PHD
Data collection	
Space group	*P61*
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.

Table 3.2 NMR restraints and refinement statistics.

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
ϕ	44
ψ	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.

3.4.2 The binding between PHF20 PHD finger and H3K4me2 is accompanied by a large conformational change

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 (Figure 3.15A). 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 3.16). 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.

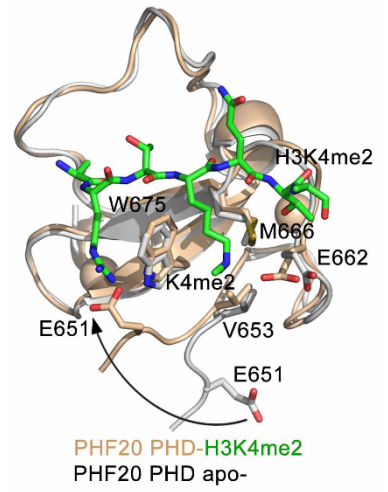


Figure 3.16 An overlay of the NMR structure of the PHF20 PHD-H3K4me2 complex (wheat-green) with the crystal structure of the apo-state of PHF20 PHD (gray).

3.4.3 The preference of PHF20 for the dimethyllysine mark is independent on the size of the methyllysine binding pocket.

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 (Figure 3.15). 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 [71] and computed solvent-accessible surface areas (Figure 3.17). 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.

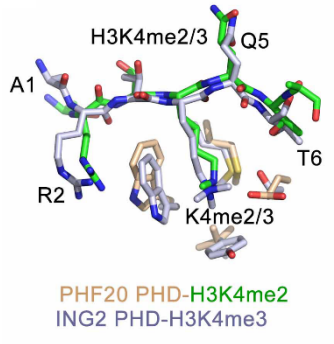
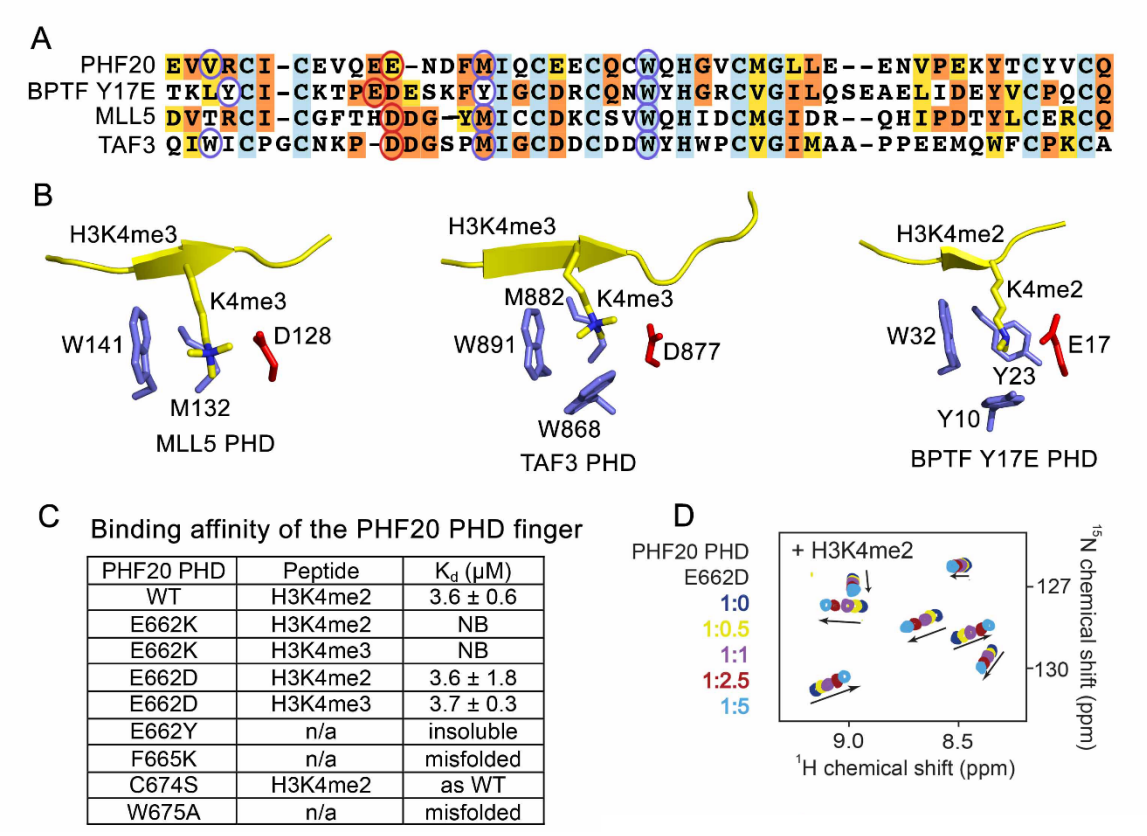


Figure 3.17 Superimposed H3K4me2 (green)-binding site of PHF20 PHD (wheat) and H3K4me3 (light blue)-binding site of ING2 PHD (light blue) (PDB: 2G6Q) are shown.

3.4.4 The glutamate residue of the K4me2-binding pocket determines the specificity of the PHF20 PHD finger

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 [24, 72, 73]. 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 3.18A and 3.18B), yet MLL5 and TAF3 prefer H3K4me3 over H3K4me2 by ~5- to 10-fold [74-76], and Pygo binds both marks with the same affinity of 2.4–2.5 μM [77]. 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 3.18C and 3.18D). 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.



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

Figure 3.18 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. NB, no binding. (D) Overlays of ^{1}H,^{15}N HSQC spectra of the mutated PHF20 PHD fingers collected upon titration with the indicated peptides.

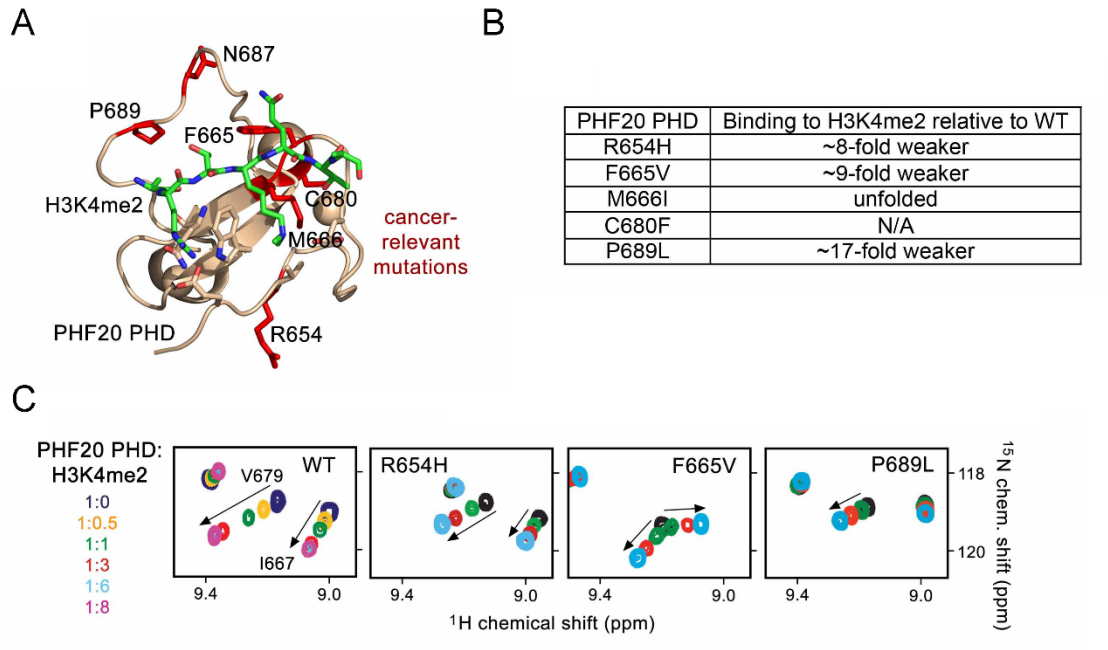
Although we cannot validate hydrogen bond formation in NMR-derived structures, a short distance between the dimethylammonium group of H3K4 and the carboxylic group of E662 suggests that these groups interact through hydrogen bonding and ionic contacts. Unlike the trimethylammonium group of H3K4 in the 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 3.18B), the position of the dimethylammonium group of H3K4 in the PHF20 complex is shifted toward E662 [74, 78, 79]. 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) [73]. This mutant prefers H3K4me2 as opposed to wild-type (WT) BPTF that selects for H3K4me3 [73] 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 3.18B). 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.

3.5 Mutations in the PHD finger decrease binding to H3K4me2

3.5.1 Cancer related PHD finger mutants 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 3.19A and 3.19B). 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 3.19C). 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.



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

Figure 3.19 Cancer related PHD finger mutants decrease binding to H3K4me2 (A) Residues of the PHF20 PHD finger, found mutated in cancer, are shown as sticks in red and labeled. (B) 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. (C) 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.

3.5.2 Mutations in the PHD finger decrease binding to H3K4me2

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 3.18C). 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 3.18C and 3.20).

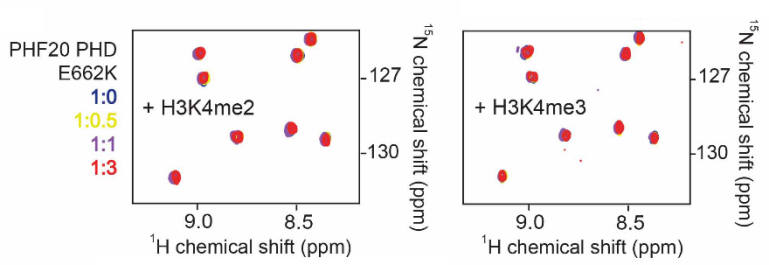


Figure 3.20 Overlays of ^{1}H,^{15}N HSQC spectra of the PHF20 PHD fingers E662K mutant collected upon titration with the indicated peptides.

Then we used PHD E662K point mutant to do histone peptides pull-down. The results showed that E662K mutation abolished the binding between PHF20 PHD finger and H3K4me2 (Figure 3.21).

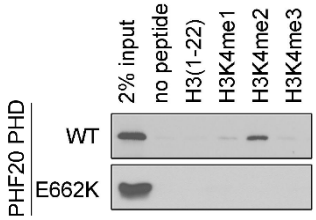


Figure 3.21 PHF20 PHD fingers E662K mutant abolished the binding to H3K4me2. Western blot analysis of the peptide pulldown analysis using the WT PHF20 PHD finger and E662K mutant.

3.6 Binding of the PHF20 PHD finger to H3K4me2 is required for MOF-dependent H4K16 acetylation and transcriptional regulation

3.6.1 H4K16ac abundance is relative to the PHF20 level in lung adenocarcinoma cell lines

PHF20 is a core subunit of the MOF-NSL lysine acetyltransferase complex, which is responsible for acetylating histone H4 at lysine 16 (H4K16) [49, 80]. Previous study reported that loss of acetylation at Lys16 of histone H4 is a common hallmark of human cancer [66]. In order to study the correlation between PHF20 and H4K16ac level, we compared the level

of H4K16Ac as well as PHF20 level in normal cells to that in lung adenocarcinoma cancer cell lines. Surprisingly, our results showed that the lung adenocarcima cell lines have high level of H4K16ac, which was positive correlated with the PHF20 level (Figure 3.22).

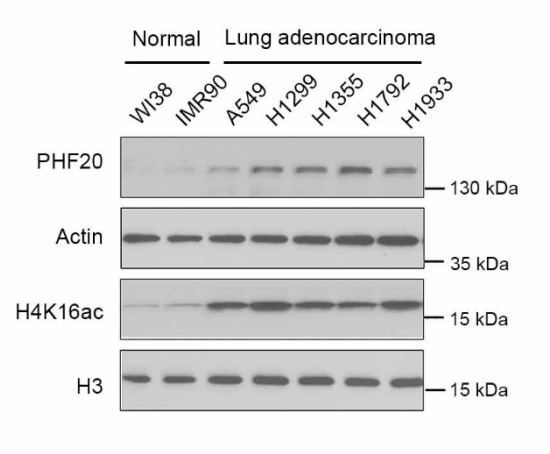


Figure 3.22 H4K16ac abundance is positive correlated with the PHF20 level in lung adenocarcinoma cell lines. Western blot analysis of H4K16ac abundance in the indicated lung adenocarcinoma cell lines and control immortalized normal cells. Histone H3 was used as a loading control.

3.6.2 PHF20 depletion reduces the H4K16ac levels in promoters of target genes

Because PHF20 has been implicated in transcription [47, 56, 57], we explored whether the histone-binding activity of the PHD finger is necessary for PHF20/MOF-dependent gene expression.

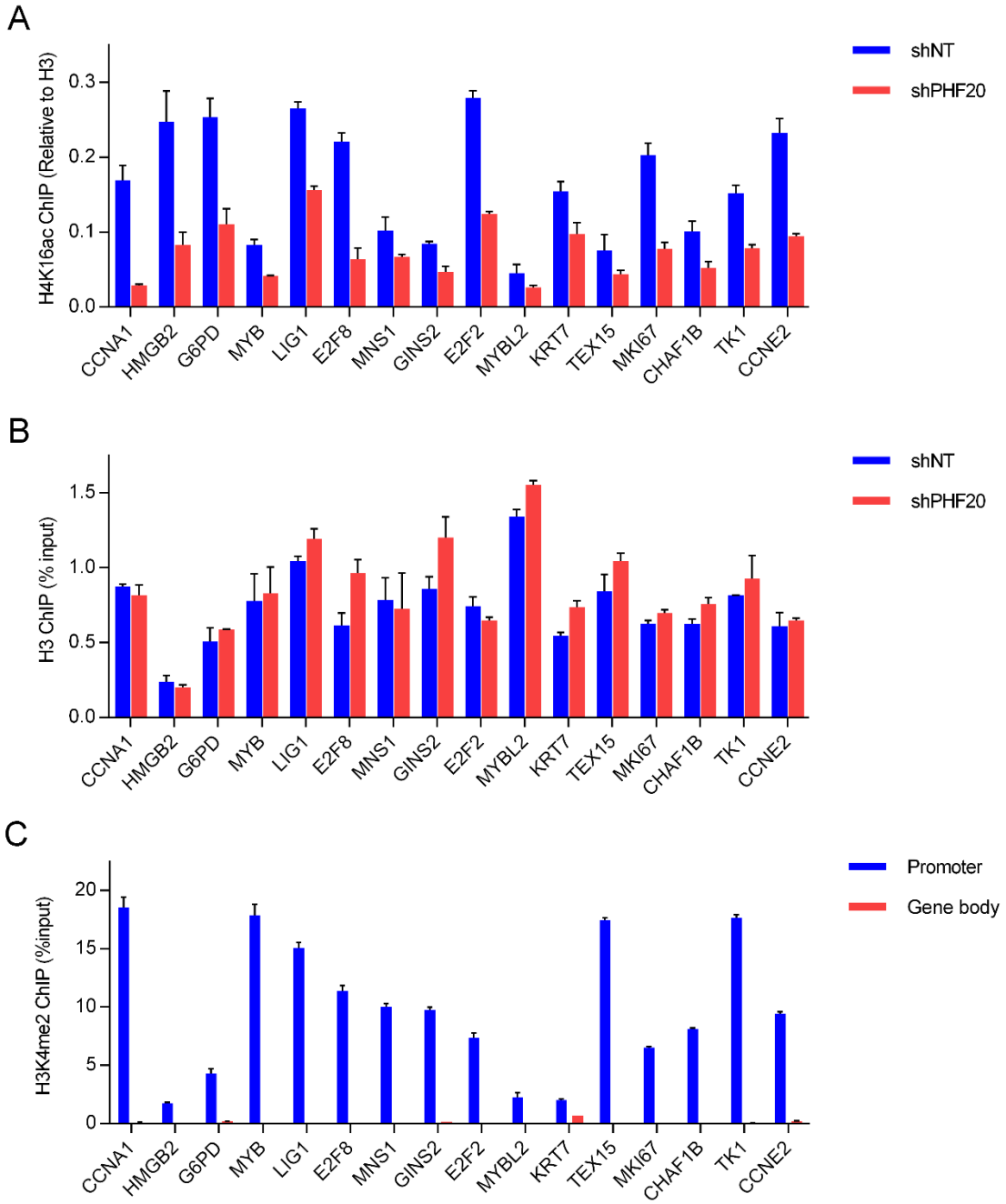


Figure 3.23 *PHF20* depletion reduced the H4K16ac levels in promoters of target genes. (A) qRT-PCR analysis of H4K16ac ChIP on gene promoters in control (*shNT*) and *PHF20* KD (*shPHF20*) cells. (B) qRT-PCR analysis of total H3 ChIP on promoters of the indicated genes in cells as in (A). (C) qRT-PCR analysis of H3K4me2 ChIP on promoters and gene bodies of the indicated genes in H1792 cells. Error bars represent SEM of three biological replicates.

Firstly, we performed the chromatin immunoprecipitation (ChIP) analysis to detect the H4K16ac occupancy on PHF20 up-regulated genes, and the results revealed a substantial

drop in H4K16ac levels in promoters of these genes, whereas the total H3 levels remained unchanged (Figures 3.23A and 3.23B). Notably, all the genes tested have strong H3K4me2 levels on their promoters (Figure 3.23C). These findings suggest that PHF20 regulates gene expression likely through MOF-mediated H4K16ac.

3.6.3 Binding of the PHF20 PHD finger to H3K4me2 is required for H4K16 acetylation.

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 re-introduced into the *shPHF20* cells. As shown in Figure 3.24, knockdown of *PHF20* led to reduced H4K16ac levels; importantly, WT *PHF20*, but not the E662K mutant, rescued the global reduction of H4K16ac levels.

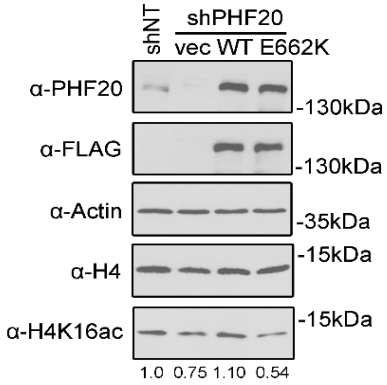


Figure 3.24 Western blot analysis of PHF20 and H4K16ac levels in control (*shNT*), *PHF20* KD (*shPHF20*), and KD cells rescued with the WT *PHF20* or E662K mutant. Total H4 and actin were used as loading controls. The density of H4K16ac bands was quantified using ImageJ.

This was further substantiated by measuring H4K16ac levels in the promoters of PHF20 target genes. *PHF20* knockdown resulted in decreased H4K16ac levels, which were fully restored by WT *PHF20*, but were not restored by the E662K mutant (Figure 3.25).

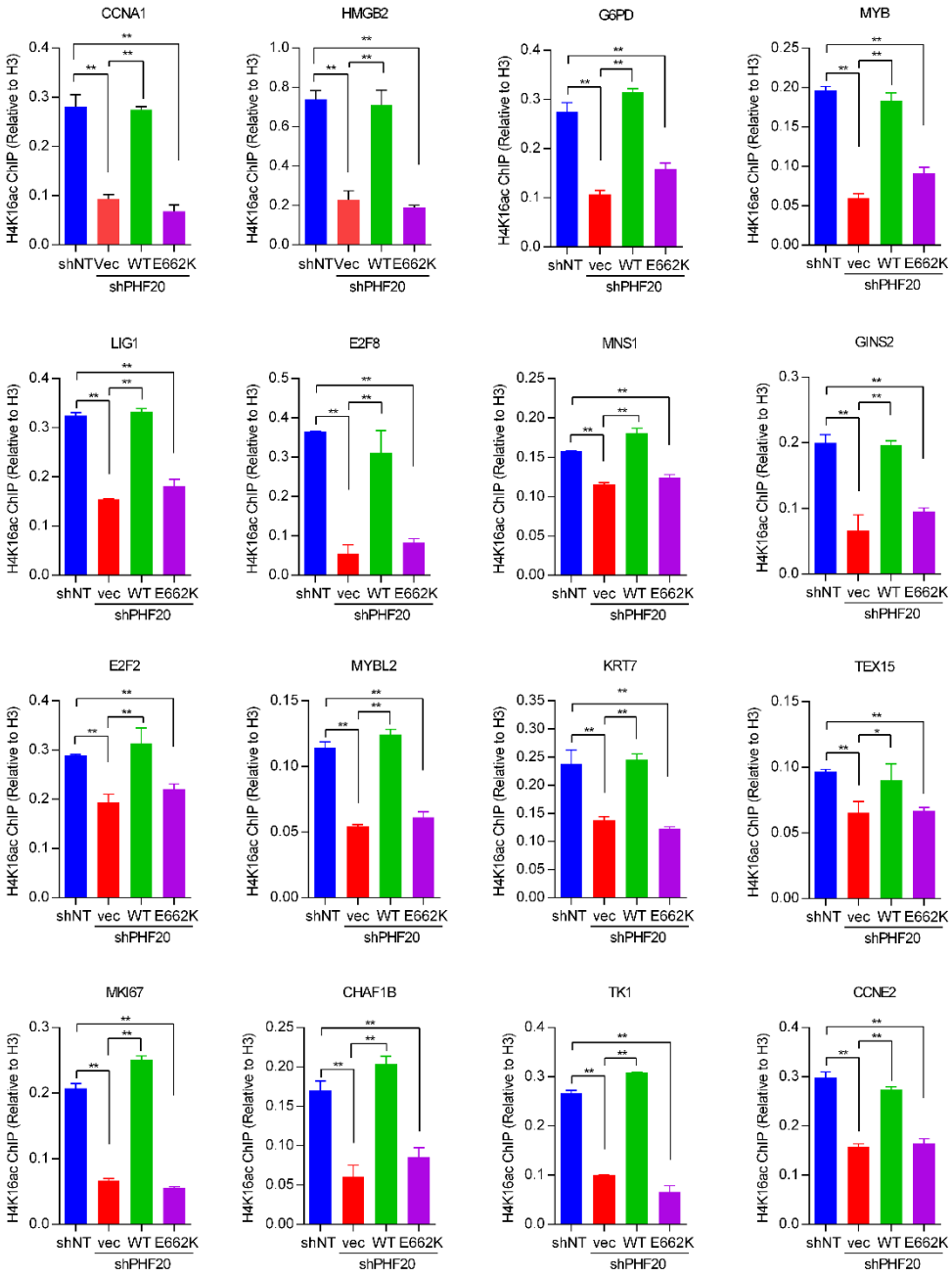


Figure 3.25 qRT-PCR analysis of H4K16ac ChIP on promoters of the indicated genes in control (*shNT*), *PHF20* KD (*shPHF20*) and KD cells rescued with the WT *PHF20* or E662K mutant. Error bars represent s.e.m. of 3 biological replicates. **p values< 0.01.

3.6.4 Binding of the PHF20 PHD finger to H3K4me2 is required for gene transcription

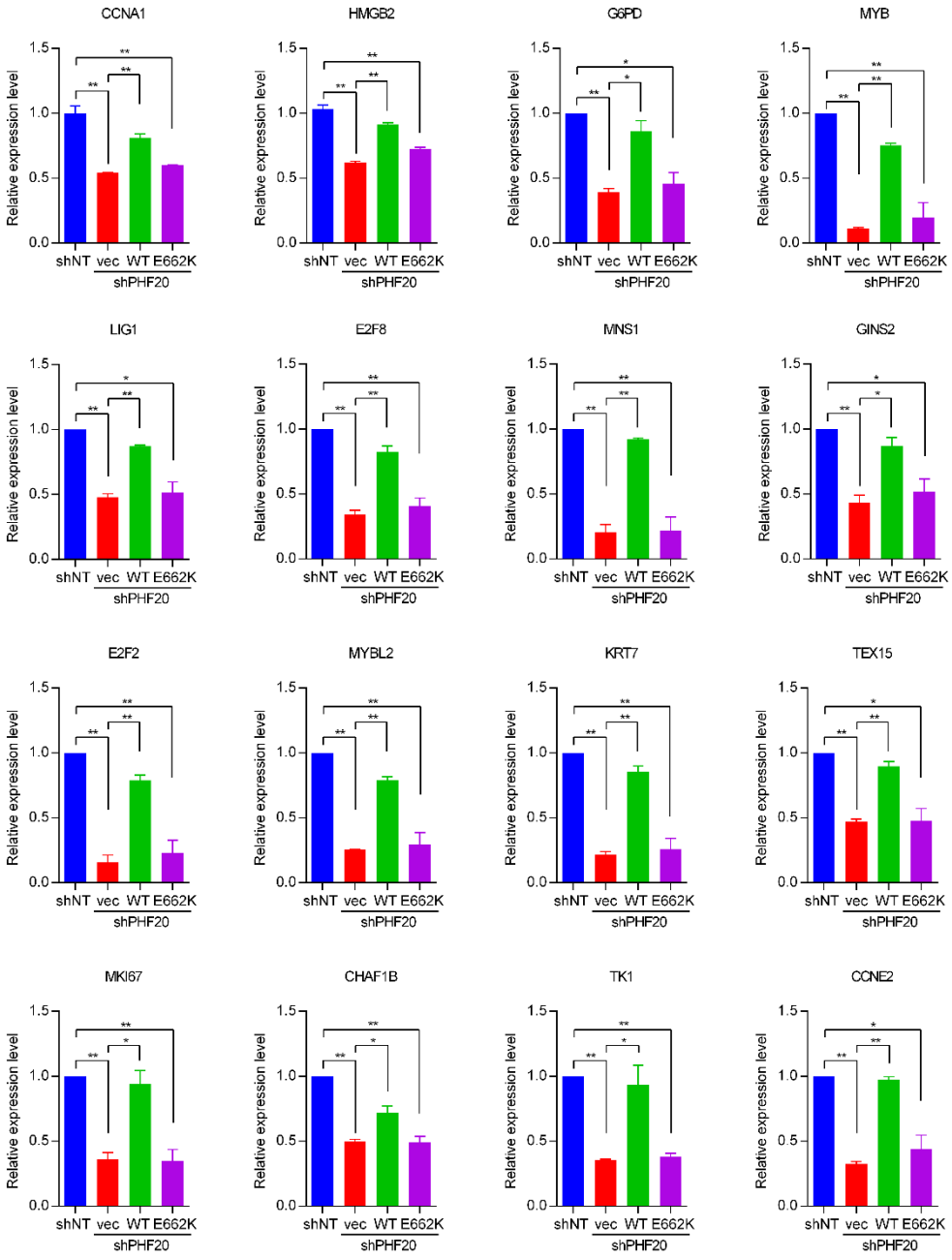


Figure 3.26 qRT-PCR analysis of gene expression in control (*shNT*), *PHF20* KD (*shPHF20*) and KD cells rescued with the WT *PHF20* or E662K mutant. Error bars represent s.e.m. of 3 biological replicates. **p values< 0.01.

Subsequently, WT *PHF20* rescued the defects in target gene expression of the *PHF20* KD H1792 cells, whereas the PHD finger mutant E662K, which is incapable of binding to H3K4me2, failed to do so (Figure 3.26)

Collectively, these results suggest that **the H3K4me2-binding function of the PHD finger is essential for PHF20-dependent histone acetylation and target gene activation.**

3.7 PHD finger is required for the biological function of PHF20

3.7.1 PHF20 PHD finger is required for cell proliferation

Since shRNA-resistant WT *PHF20* but not PHD finger mutant can restore cell cycle related genes expression in *PHF20*-depleted cells, we expected that WT *PHF20* can rescue the cell growth in *PHF20*-depleted genes. The cell proliferation assay showed that WT *PHF20* rescued the defects in cell proliferation of the *PHF20* KD H1792 cells, whereas the PHD finger mutant E662K partly rescued.

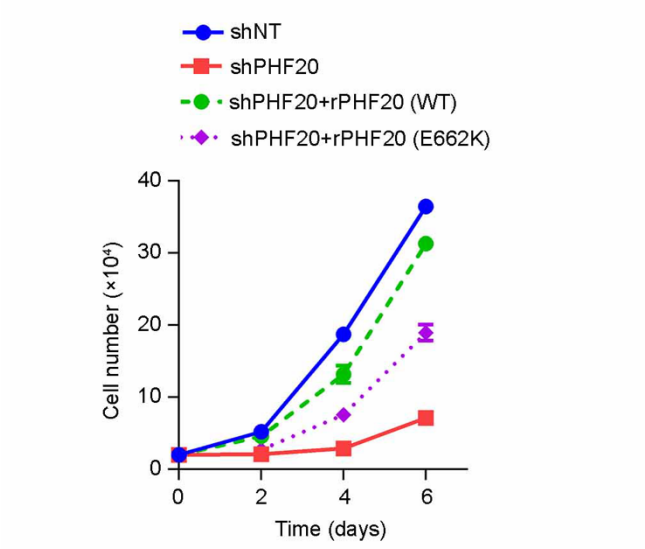


Figure 3.27 Cell proliferation assays of cells as in in control (*shNT*), *PHF20* KD (*shPHF20*), and KD cells rescued with the WT *PHF20* or E662K mutant.

3.7.2 PHF20 PHD finger is required for colony formation

In accord with the resuts of the cell proliferation assay, the colony formation assay showed that WT *PHF20* rescued the defects in colony formation of the *PHF20*-depleted H1792 cells, whereas the PHD finger mutant E662K failed to do so.

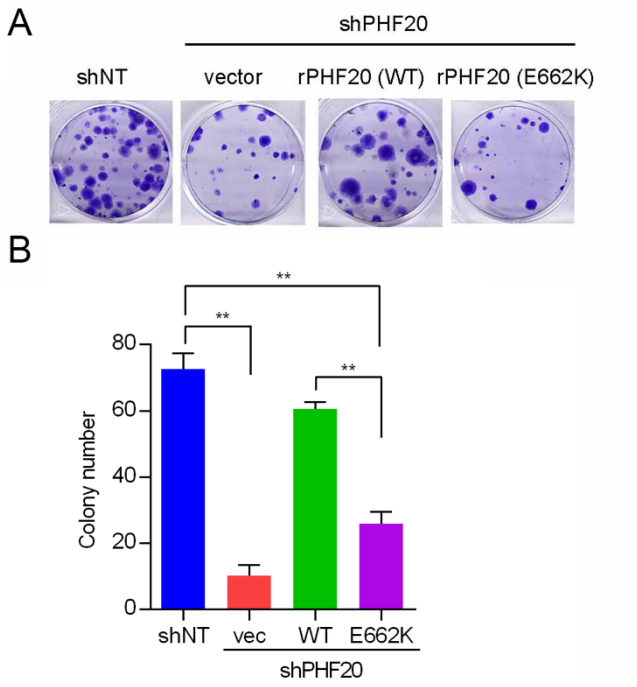


Figure 3.28 Colony formation assays of cells as in control (*shNT*), *PHF20* KD (*shPHF20*), and KD cells rescued with the WT *PHF20* or E662K mutant. Representative crystal-violet-stained cells are shown in (A), and quantification of six replicates is shown in (B).

Taken together, these results indicate that PHD finger is essential for *PHF20*-dependent cancer cell growth and survival.

4 DISCUSSION

4.1 PHF20 PHD finger represents an example of a native reader capable of reading H3K4me2

The plant homeodomain (PHD) finger is one of the largest families of epigenetic reader domains present in chromatin-related proteins, which are involved in the regulation of chromatin structure and dynamics [81]. The PHD finger is a small protein domain of ~65-residue cysteine-rich sequence, which binds two zinc ions in a cross-braced topology. The PHD finger is an evolutionarily conserved module and is present either as a single module or in multiple copies [68].

In the past few years, a wide array of PHD finger ligands have been identified. PHD fingers can mainly recognize different post-translational modifications (PTMs) on the H3 tail, including trimethylation of K4, trimethylation of K9, acetylation of K14 as well as unmodified histone H3 [41, 69, 70, 82-94]. The original discovery of their role in gene transcription is attributed to the recognition of trimethylated lysine 4 of histone H3 (H3K4me3) [41, 42, 69, 70]. The PHD fingers of numerous proteins have been shown to bind H3K4me3 with high specificity and affinity. These readers of H3K4me3 together comprise one of the well-established subsets of PHD fingers. In 2007, the PHD fingers of BHC80 and DNMT3L were characterized that recognize unmodified histone H3 tail [93, 94], which represented the second major subset of PHD fingers. A smaller number of PHD fingers displays specificity for trimethylated lysine 9 of histone H3 (H3K9me3) [87-92]. The double PHD finger (DPF) domain of DPF3b, MORF and MOZ select for acetylated histone H3 at lysine 14 (H3K14ac) [82-86].

In addition to trimethylated histone H3, the PHD finger can recognize dimethylated histone H3 at lysine 4 (H3K4me2), which is exemplified by the PHD finger of Pygo [77]. Pygo and BCL9 contribute to efficient β-catenin-mediated transcription in Wnt-stimulated mammalian cells [95, 96]. BCL9 proteins function as adaptors between Pygo and β-catenin, by binding to the PHD finger in the C terminus of Pygo through their homology domain 1 (HD1), and to the Armadillo repeat domain of β-catenin through their homology domain 2 (HD2) [97, 98]. Efficient methylated histone binding of Pygo PHD finger requires BCL9 HD1 association.

The PHD finger from human Pygo1 binds to H3K4me3 and H3K4me2 with similar high affinities (K_d, 2.5 and 2.4 μM, respectively). After binding to the HD1 domain of BCL9, the human PHD-HD1 complex showed higher affinity to H3K4me3 and H3K4me2 (K_d, 1.2 and 0.9 μM, respectively) [77]. The PHD finger of Pygo is the only one reported to bind to H3K4me2. In this study, we utilized biochemical and structural analyses to identify that the PHD finger of PHF20 specifically binds to H3K4me2, which represents an example of a native reader that selects for this modification.

4.2 Functional link between Tudor2 and PHD of PHF20

The PHD fingers are frequently found adjacent to other chromatin-binding modules, such as bromodomain [99, 100], tudor domains [40, 101], chromodomains [89] and other PHD fingers [82, 85], which aid combinatorial, multivalent readout of histone tails. The interplay between multiple readers and the DNA-binding domains creates an intricate network of contacts with nucleosomes and adds another layer of complexity to the chromatin-targeting mechanisms.

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 [40]. 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. Substantial CSPs in PHF20 Tudor2 were observed upon addition of the p53K382me2 peptide, and in agreement with previous reports, K_d for this interaction was found to be ~100 μM (Figures 4.1A and 4.1C). 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 4.1B). 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), 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 4.1D). 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. These results demonstrate that, although both reader domains in PHF20 exhibit preference for dimethyllysine species, their functions do not overlap.

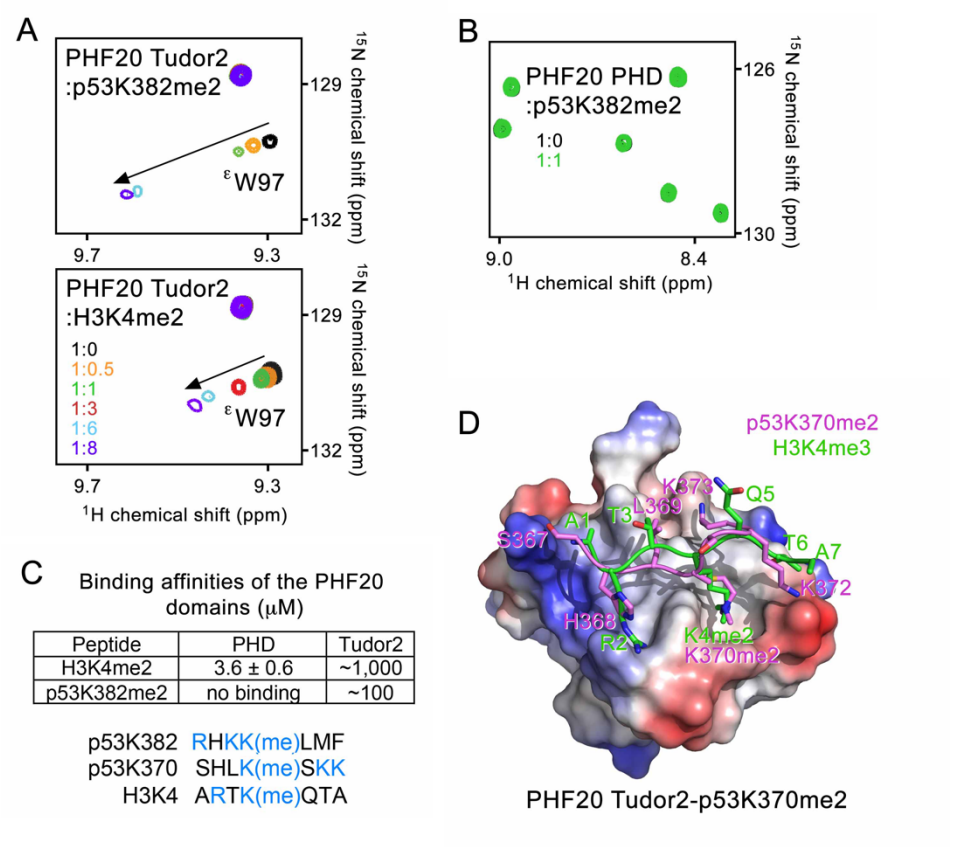


Figure 4.1. Functional Coupling between Tudor2 and PHD finger of PHF20. (A and B) Superimposed $^{1}H,^{15}N$ HSQC spectra of the Tudor2 and PHD finger domains of PHF20 collected upon titration with the indicated peptides. Spectra are color coded according to the protein:peptide molar ratio (inset). (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 (PDB2LDM). 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.

Besides being a subunit of NSL, MOF is present in another evolutionarily conserved HAT complex, MSL [51, 53, 102]. Although both NSL and MSL complexes efficiently acetylate H4K16, only NSL is able to acetylate non-histone proteins, including p53 [50, 51]. Because the NSL-MOF complex, but not the MSL-MOF complex, contains the PHF20 subunit whose Tudor2 binds to dimethylated p53 [40], it is tempting to suggest that PHF20 might have a role in promoting p53 acetylation by MOF (Figure 4.2). 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.

4.3 PHF20 links H3K4 methylation and H4K16 acetylation, interpreting the language of histone

Post-translational modifications on histones act singly and in combination to form a language or 'code' that is read by specialized proteins, which in turn recruit additional epigenetic regulators to modulate chromatin dynamics [103, 104]. The dynamic nature of chromatin and its myriad modifications play a crucial role in transcriptional regulation [105]. In the current study, we found that recognition of H3K4me2 by PHF20 PHD finger is required for histone acetylation and accumulation of PHF20 target genes. We propose that H3K4me2 recruits the PHF20/MOF complex, which subsequently catalyzes acetylation of histone 4 at lysine 16 (H4K16ac), leading to chromatin decondensation and transcriptional activation.

Both histone marks, H4K16ac and H3K4me2, 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 [52, 53]. 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 [52, 102]. 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 [106-108], 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 [109] (Figure 4.2). 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 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.

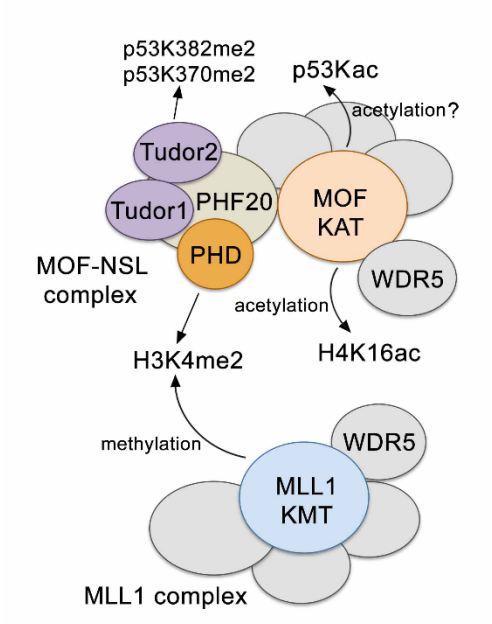


Figure 4.2 A model depicting a potential role of PHF20 readers in linking H3K4 methylation with H4K16 acetylation as well as p53 methylation and acetylation.

SUMMARY

Lung cancer is the most common cancer and the leading cause of cancer death worldwide. Most lung cancers are diagnosed at an advanced stage, where the prognosis is poor and the therapeutic options are limited. In order to develop more effective novel strategies in lung cancer diagnosis and treatment, the underlying molecular and cellular mechanisms of lung cancer remains to be explored.

PHF20 was originally identified as a target of autoantibodies in patients suffering from glioblastoma. Then abnormal expression of *PHF20* in various cancers including non-small cell lung cancer, small cell lung cancer was reported, indicating that PHF20 might be tumor-associated factor. However, little is known about the function of PHF20 in cancer development. PHF20 is a core component of the lysine acetyltransferase complex MOF-NSL that generates the major epigenetic mark H4K16ac and is essential for transcriptional regulation and DNA repair. As a multi-domain protein, PHF20 contains several domains including two Tudor domains and a PHD finger, which are putative methylation reader modules. However, the role of PHF20 in the MOF complex remains elusive.

In this study, we use non-small cell lung cancer as a model to investigate the epigenetic role of PHF20 in cancer development. The abundant expression of *PHF20* in lung adenocarcinoma cell lines was validated. Depletion of *PHF20* mRNA in lung adenocarcinoma inhibits cell growth and colony formation. RNA sequencing (RNA-seq) analysis showed that PHF20 affected genes are implicated in cell cycle and DNA replication. Biochemical and structural analyses showed that the 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. 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 and H3K4me2. We propose that PHF20 recognizes H3K4me2 through its PHD finger and recruits the MOF complex to the chromatin, which subsequently catalyzes acetylation of histone 4 at lysine 16 (H4K16ac), leading to chromatin decondensation and transcriptional activation.

ZUSAMMENFASSUNG

Lungenkrebs ist die häufigste Krebsart mit der höchsten Mortaliät weltweit. Die meisten Lungenkrebsarten werden erst in einem fortgeschrittenen Stadium der Erkrankung diagnostiziert, wo die Prognosen ungünstig und die therapeutischen Optionen begrenzt sind. Um neue effektivere Strategien bei der Lungenkrebsdiagnose und -behandlung zu entwickeln, müssen die zugrunde liegenden molekularen und zellulären Mechanismen des Lungenkrebses weiter aufgeklärt werden.

PHF20 wurde ursprünglich als ein Target für Autoantikörper von Patienten mit einem Glioblastom identifiziert. Die abnormale Expression von *PHF20* in verschiedenen Krebsarten, einschließlich im kleinzelligen und nicht-kleinzelligen Lungenkrebs, weist auf eine potentielle Rolle dieses Proteins als Tumor-assoziierter Faktor hin. Allerdings ist wenig über die Funktion von PHF20 in der Tumorentwicklung bekannt. PHF20 ist eine Kernkomponente des MOF-NSL Histon-Acetyltransferase-Komplexes, das für die wichtige epigenetische Markierung H4K16ac, und für die transkriptionelle Regulation und DNA-Reparatur essentiell ist. Als multidomänales Protein enthält PHF20 zwei Tudor-Domänen und eine PHD-Finger-Domäne, die vermutlich in der Methylierungserkennung involviert sind. Jedoch ist die Rolle von PHF20 im MOF-Komplex noch nicht vollständig geklärt.

In dieser Studie verwendeten wir nicht-kleinzelligen Lungenkrebs als Modell, um die epigenetische Rolle von PHF20 in der Tumorentwicklung zu untersuchen. Die erhöhte Expression von *PHF20* wurde in Lungen-Adenokarzinom-Zelllinien validiert. Eine Depletion von *PHF20*-mRNA in diesen Zelllinien führte zu einer Hemmung des Zellwachstums und der Koloniebildung. Expressionsanalysen durch quantitative RNA-Sequenzierung (RNA-seq) offenbarten, dass PHF20 in der Regulation von Genen des Zellzyklus und der DNA-Replikation involviert ist. Biochemische und strukturelle Analysen zeigten, dass der PHD-Finger von PHF20 spezifisch am dimethylierten Lysinrest in Position 4 am N-terminalen Ende im Histon H3 (H3K4me2) bindet, was zugleich ein Beispiel für die native Funktionsweise dieser Domäne ist. Die Bindung des PHD-Fingers an H3K4me2 ist für die Histon-Acetylierung sowie für die Akkumulation von PHF20 an Zielgenen und die Transkriptionsaktivierung erforderlich. Zusammenfassend lässt sich feststellen, dass unsere Erkenntnisse eine einzigartige PHF20-vermittelte Verbindung zwischen der MOF-Histon-Acetyltransferase (HAT) und H3K4me2 darstellen. Wir nehmen an, dass PHF20 H3K4me2

durch einen PHD-Finger erkennt und den MOF-Komplex zum Chromatin rekrutiert, welcher anschließend die Acetylierung von Histon 4 am Lysin 16 (H4K16ac) katalysiert. Dies führt zu einer Chromatindekondensation und Aktivierung der Transkription.

ABBREVIATION

PHF20, plant homeodomain finger protein 20

GLEA 2, glioma-expressed antigen 2

PHD, plant homeodomain

MOF, male absent on the first

NSL, non-specific lethal

H3K4me2, dimethylated lysine 4 of histone H3

HAT, histone acetyltransferase

SCLC, small cell lung cancer

NSCLC, non-small cell lung cancer

SqCC, squamous cell carcinoma

AdC, adenocarcinoma

HCA58, hepatocellular carcinoma associated antigen 58

ATM, ataxia telangiectasia mutated

WDR5, WD repeat domain 5

MLL1, mixed-lineage leukemia 1

IPSCs, induced pluripotent stem cells

PBS, phosphate buffered saline

FBS, fetal bovine serum

PMSF, phenylmethanesulfonyl fluoride

SDS, sodium dodecyl sulfate

DTT, 1,4-dithio-DL-threitol

APS, ammonium persulfate

IPTG, isopropyl b-D-1-thiogalactopyranoside

NMR, nuclear magnetic resonance

GST, glutathione S-transferase

HSQC, $^{1}H,^{15}N$ heteronuclear single quantum coherence

CSPs, chemical shift perturbations

ING2, inhibitor of growth, member 2

ChIP, chromatin immunoprecipitation

RIPA, radio immunoprecipitation assay

DNA, desoxyribonucleic acid

RNA, ribonucleic acid

shRNA, short hairpin RNA

RNA-seq, RNA sequencing

qRT-PCR, real time quantitative reverse transcriptase polymerase chain reaction

IPA, Ingenuity Pathway Analysis

KD, knockdown

TAF3, TATA-Box Binding Protein Associated Factor 3

BPTF, bromodomain PHD finger transcription factor

WT, wild-type

PTM, post translational modification

REFERENCE

1. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.
2. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
3. Addario, B.J., *Lung cancer is a global epidemic and requires a global effort.* Ann Transl Med, 2015. **3**(2): p. 26.
4. Herbst, R.S., J.V. Heymach, and S.M. Lippman, *Lung cancer.* N Engl J Med, 2008. **359**(13): p. 1367-80.
5. Walsh, K. and W.A. Wallace, *Molecular pathology in lung cancer: a guide to the techniques used in clinical practice.* Histopathology, 2014. **65**(6): p. 731-41.
6. Brody, H., *Lung cancer.* Nature, 2014. **513**(7517): p. S1.
7. Ansari, J., R.E. Shackelford, and H. El-Osta, *Epigenetics in non-small cell lung cancer: from basics to therapeutics.* Transl Lung Cancer Res, 2016. **5**(2): p. 155-71.
8. Brennan, P., et al., *High cumulative risk of lung cancer death among smokers and nonsmokers in Central and Eastern Europe.* Am J Epidemiol, 2006. **164**(12): p. 1233-41.
9. Villeneuve, P.J. and Y. Mao, *Lifetime probability of developing lung cancer, by smoking status, Canada.* Can J Public Health, 1994. **85**(6): p. 385-8.
10. Herceg, Z. and T. Vaissiere, *Epigenetic mechanisms and cancer: an interface between the environment and the genome.* Epigenetics, 2011. **6**(7): p. 804-19.
11. Tomioka, K., et al., *Risk of Lung Cancer in Workers Exposed to Benzidine and/or Beta-Naphthylamine: A Systematic Review and Meta-Analysis.* J Epidemiol, 2016. **26**(9): p. 447-58.
12. Sheen, S., et al., *An updated review of case-control studies of lung cancer and indoor radon-Is indoor radon the risk factor for lung cancer?* Ann Occup Environ Med, 2016. **28**: p. 9.
13. Berman, A.T., et al., *Use of Survivorship Care Plans and Analysis of Patient-Reported Outcomes in Multinational Patients With Lung Cancer.* J Oncol Pract, 2016. **12**(5): p. e527-35.
14. Vijayvergia, N., P.C. Shah, and C.S. Denlinger, *Survivorship in Non-Small Cell Lung Cancer: Challenges Faced and Steps Forward.* J Natl Compr Canc Netw, 2015. **13**(9): p. 1151-61.
15. Pozo, C.L., M.A. Morgan, and J.E. Gray, *Survivorship issues for patients with lung cancer.* Cancer Control, 2014. **21**(1): p. 40-50.
16. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.
17. Riddihough, G. and E. Pennisi, *The evolution of epigenetics.* Science, 2001. **293**(5532): p. 1063.

18. Pfeifer, G.P., *Epigenetics: An elusive DNA base in mammals.* Nature, 2016. **532**(7599): p. 319-20.

19. Stricker, S.H., A. Koferle, and S. Beck, *From profiles to function in epigenomics.* Nat Rev Genet, 2017. **18**(1): p. 51-66.

20. Margueron, R. and D. Reinberg, *Chromatin structure and the inheritance of epigenetic information.* Nat Rev Genet, 2010. **11**(4): p. 285-96.

21. Villanueva, M.T., *Epigenetics: Chromatin marks the spot.* Nat Rev Cancer, 2015. **15**(4): p. 196-7.

22. Venkatesh, S. and J.L. Workman, *Histone exchange, chromatin structure and the regulation of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 178-89.

23. Tan, M., et al., *Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification.* Cell, 2011. **146**(6): p. 1016-28.

24. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone readers.* Nat Struct Mol Biol, 2012. **19**(12): p. 1218-27.

25. Sapienza, C. and J.P. Issa, *Diet, Nutrition, and Cancer Epigenetics.* Annu Rev Nutr, 2016. **36**: p. 665-81.

26. Feinberg, A.P., M.A. Koldobskiy, and A. Gondor, *Epigenetic modulators, modifiers and mediators in cancer aetiology and progression.* Nat Rev Genet, 2016. **17**(5): p. 284-99.

27. Tsai, H.C. and S.B. Baylin, *Cancer epigenetics: linking basic biology to clinical medicine.* Cell Res, 2011. **21**(3): p. 502-17.

28. Shortt, J., et al., *A chemical probe toolbox for dissecting the cancer epigenome.* Nat Rev Cancer, 2017. **17**(4): p. 268.

29. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nat Rev Genet, 2016. **17**(10): p. 630-41.

30. Kanwal, R., K. Gupta, and S. Gupta, *Cancer epigenetics: an introduction.* Methods Mol Biol, 2015. **1238**: p. 3-25.

31. Avgustinova, A. and S.A. Benitah, *The epigenetics of tumour initiation: cancer stem cells and their chromatin.* Curr Opin Genet Dev, 2016. **36**: p. 8-15.

32. Campbell, R.M. and P.J. Tummino, *Cancer epigenetics drug discovery and development: the challenge of hitting the mark.* J Clin Invest, 2014. **124**(1): p. 64-9.

33. Shinjo, K. and Y. Kondo, *Targeting cancer epigenetics: Linking basic biology to clinical medicine.* Adv Drug Deliv Rev, 2015. **95**: p. 56-64.

34. Hatzimichael, E., et al., *Epigenetics in diagnosis, prognostic assessment and treatment of cancer: an update.* EXCLI J, 2014. **13**: p. 954-76.

35. Langevin, S.M., R.A. Kratzke, and K.T. Kelsey, *Epigenetics of lung cancer.* Transl Res, 2015. **165**(1): p. 74-90.

36. Fischer, U., et al., *Glioma-expressed antigen 2 (GLEA2): a novel protein that can elicit immune responses in glioblastoma patients and some controls.* Clin Exp Immunol, 2001. **126**(2): p. 206-13.

37. Pallasch, C.P., et al., *Autoantibodies against GLEA2 and PHF3 in glioblastoma: tumor-associated autoantibodies correlated with prolonged survival.* Int J Cancer, 2005. **117**(3): p. 456-9.

38. Bankovic, J., et al., *Identification of genes associated with non-small-cell lung cancer promotion and progression.* Lung Cancer, 2010. **67**(2): p. 151-9.

39. Taniwaki, M., et al., *Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer.* Int J Oncol, 2006. **29**(3): p. 567-75.

40. Cui, G., et al., *PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53.* Nat Struct Mol Biol, 2012. **19**(9): p. 916-24.

41. Shi, X., et al., *ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression.* Nature, 2006. **442**(7098): p. 96-9.

42. Li, H., et al., *Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF.* Nature, 2006. **442**(7098): p. 91-5.

43. Yun, M., et al., *Readers of histone modifications.* Cell Res, 2011. **21**(4): p. 564-78.

44. Huang, Y., et al., *Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A.* Science, 2006. **312**(5774): p. 748-51.

45. Adams-Cioaba, M.A., et al., *Crystal structures of the Tudor domains of human PHF20 reveal novel structural variations on the Royal Family of proteins.* FEBS Lett, 2012. **586**(6): p. 859-65.

46. Kim, J., et al., *Tudor, MBT and chromo domains gauge the degree of lysine methylation.* EMBO Rep, 2006. **7**(4): p. 397-403.

47. Zhang, T., et al., *PHF20 regulates NF-kappaB signalling by disrupting recruitment of PP2A to p65.* Nat Commun, 2013. **4**: p. 2062.

48. Avvakumov, N. and J. Cote, *The MYST family of histone acetyltransferases and their intimate links to cancer.* Oncogene, 2007. **26**(37): p. 5395-407.

49. Cai, Y., et al., *Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex.* J Biol Chem, 2010. **285**(7): p. 4268-72.

50. Rea, S., G. Xouri, and A. Akhtar, *Males absent on the first (MOF): from flies to humans.* Oncogene, 2007. **26**(37): p. 5385-94.

51. Li, X., et al., *Two mammalian MOF complexes regulate transcription activation by distinct mechanisms.* Mol Cell, 2009. **36**(2): p. 290-301.

52. Dou, Y., et al., *Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF.* Cell, 2005. **121**(6): p. 873-85.

53. Zhao, X., et al., *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, 2013. **9**(11): p. e1003940.

54. Sykes, S.M., et al., *Acetylation of the p53 DNA-binding domain regulates apoptosis induction.* Mol Cell, 2006. **24**(6): p. 841-51.

55. Yang, W.H., et al., *Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability.* Nat Cell Biol, 2006. **8**(10): p. 1074-83.

56. Park, S., et al., *Identification of Akt interaction protein PHF20/TZP that transcriptionally regulates p53.* J Biol Chem, 2012. **287**(14): p. 11151-63.

57. Badeaux, A.I., et al., *Loss of the methyl lysine effector protein PHF20 impacts the expression of genes regulated by the lysine acetyltransferase MOF.* J Biol Chem, 2012. **287**(1): p. 429-37.

58. Heisel, S.M., et al., *Increased Seroreactivity to Glioma-Expressed Antigen 2 in Brain Tumor Patients under Radiation.* Plos One, 2008. **3**(5).

59. Zhao, W., et al., *Jmjd3 Inhibits Reprogramming by Upregulating Expression of INK4a/Arf and Targeting PHF20 for Ubiquitination.* Cell, 2013. **152**(5): p. 1037-1050.

60. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.

61. Pruitt, K.D., et al., *NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy.* Nucleic Acids Res, 2012. **40**(Database issue): p. D130-5.

62. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.

63. Franken, N.A., et al., *Clonogenic assay of cells in vitro.* Nat Protoc, 2006. **1**(5): p. 2315-9.

64. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

65. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

66. Fraga, M.F., et al., *Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer.* Nat Genet, 2005. **37**(4): p. 391-400.

67. Zhang, X., et al., *G9a-mediated methylation of ERalpha links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression.* Nat Commun, 2016. **7**: p. 10810.

68. Musselman, C.A. and T.G. Kutateladze, *Handpicking epigenetic marks with PHD fingers.* Nucleic Acids Res, 2011. **39**(21): p. 9061-71.

69. Pena, P.V., et al., *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2.* Nature, 2006. **442**(7098): p. 100-3.

70. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

71. Pena, P.V., et al., *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2.* Nature, 2006. **442**(7098): p. 100-103.

72. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-40.

73. Li, H., et al., *Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger.* Mol Cell, 2007. **28**(4): p. 677-91.

74. Ali, M., et al., *Molecular basis for chromatin binding and regulation of MLL5.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11296-301.

75. van Ingen, H., et al., *Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3.* Structure, 2008. **16**(8): p. 1245-1256.

76. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.

77. Fiedler, M., et al., *Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex.* Mol Cell, 2008. **30**(4): p. 507-18.

78. Lemak, A., et al., *Solution NMR structure and histone binding of the PHD domain of human MLL5.* PLoS One, 2013. **8**(10): p. e77020.

79. van Ingen, H., et al., *Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3.* Structure, 2008. **16**(8): p. 1245-56.

80. Su, J., et al., *The Functional Analysis of Histone Acetyltransferase MOF in Tumorigenesis.* Int J Mol Sci, 2016. **17**(1).

81. Bortoluzzi, A., et al., *Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains.* Biochem J, 2017. **474**(10): p. 1633-1651.

82. Zeng, L., et al., *Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b.* Nature, 2010. **466**(7303): p. 258-62.

83. Qiu, Y., et al., *Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription.* Genes Dev, 2012. **26**(12): p. 1376-91.

84. Lange, M., et al., *Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex.* Genes Dev, 2008. **22**(17): p. 2370-84.

85. Dreveny, I., et al., *The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification.* Nucleic Acids Res, 2014. **42**(2): p. 822-35.

86. Ali, M., et al., *Tandem PHD fingers of MORF/MOZ acetyltransferases display selectivity for acetylated histone H3 and are required for the association with chromatin.* J Mol Biol, 2012. **424**(5): p. 328-38.

87. Xi, Q., et al., *A poised chromatin platform for TGF-beta access to master regulators.* Cell, 2011. **147**(7): p. 1511-24.

88. Musselman, C.A., et al., *Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications.* Biochem J, 2009. **423**(2): p. 179-87.

89. Mansfield, R.E., et al., *Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9.* J Biol Chem, 2011. **286**(13): p. 11779-91.

90. Iwase, S., et al., *ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome.* Nat Struct Mol Biol, 2011. **18**(7): p. 769-76.

91. Eustermann, S., et al., *Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin.* Nat Struct Mol Biol, 2011. **18**(7): p. 777-82.

92. Dhayalan, A., et al., *The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9.* Hum Mol Genet, 2011. **20**(11): p. 2195-203.

93. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.

94. Lan, F., et al., *Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression.* Nature, 2007. **448**(7154): p. 718-22.

95. Brembeck, F.H., et al., *Essential role of BCL9-2 in the switch between beta-catenin's adhesive and transcriptional functions.* Genes Dev, 2004. **18**(18): p. 2225-30.

96. Adachi, S., et al., *Role of a BCL9-related beta-catenin-binding protein, B9L, in tumorigenesis induced by aberrant activation of Wnt signaling.* Cancer Res, 2004. **64**(23): p. 8496-501.

97. Stadeli, R. and K. Basler, *Dissecting nuclear Wingless signalling: recruitment of the transcriptional co-activator Pygopus by a chain of adaptor proteins.* Mech Dev, 2005. **122**(11): p. 1171-82.

98. Kramps, T., et al., *Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex.* Cell, 2002. **109**(1): p. 47-60.

99. Zeng, L., et al., *Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing.* Nat Struct Mol Biol, 2008. **15**(6): p. 626-33.

100. Tsai, W.W., et al., *TRIM24 links a non-canonical histone signature to breast cancer.* Nature, 2010. **468**(7326): p. 927-32.

101. Xie, S., J. Jakoncic, and C. Qian, *UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail.* J Mol Biol, 2012. **415**(2): p. 318-28.

102. Li, X. and Y. Dou, *New perspectives for the regulation of acetyltransferase MOF.* Epigenetics, 2010. **5**(3): p. 185-8.

103. Rothbart, S.B. and B.D. Strahl, *Interpreting the language of histone and DNA modifications.* Biochim Biophys Acta, 2014. **1839**(8): p. 627-43.

104. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

105. Tessarz, P. and T. Kouzarides, *Histone core modifications regulating nucleosome structure and dynamics.* Nat Rev Mol Cell Biol, 2014. **15**(11): p. 703-8.

106. Couture, J.F., E. Collazo, and R.C. Trievel, *Molecular recognition of histone H3 by the WD40 protein WDR5.* Nat Struct Mol Biol, 2006. **13**(8): p. 698-703.

107. Ruthenburg, A.J., et al., *Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex.* Nat Struct Mol Biol, 2006. **13**(8): p. 704-12.

108. Schuetz, A., et al., *Structural basis for molecular recognition and presentation of histone H3 by WDR5.* EMBO J, 2006. **25**(18): p. 4245-52.

109. Milne, T.A., et al., *MLL targets SET domain methyltransferase activity to Hox gene promoters.* Mol Cell, 2002. **10**(5): p. 1107-17.

ACKNOWLEDGEMENT

I would appreciate this three-year doctoral study in LMU, which I will keep as one of the most unforgettable experience in my life.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr. Christiane J. Bruns for her continuous support and encouragement of my doctoral study and research, for her motivation, enthusiasm, inspiration and immense knowledge. She provided nice guidance and creative ideas. She taught me to be independent and have an open mind during my study. To me, she is not only a supervisor for my doctoral study, but also a perfect mentor in my career development and life plan.

Foremost, I would like to express my deeply gratitude to my advisor Prof. Dr. Xiaobing Shi in U.T. MD Anderson Cancer Center in the United States. He gave me the great opportunity to do my doctoral research as a visiting student in his Lab. He was always there for the discussion of all the data analysis, graph design, manuscript preparation and thesis writing with big encouragement and nice advice. Moreover, he made great efforts to explain things clearly and simply, which helped me develop an understanding of the project and basic scientific research. Without his help, this thesis could not be finished.

In particular, I am sincerely grateful to Dr. Yue Zhao for her patient guidance and great support. She helped me a lot in both my study and life. She always provided me encouragement, big support and lots of good ideas. To me, she is more than a colleague, a teacher and a friend.

I would like to express my sincere gratitude to the cooperators for this project. I thank Dr. Brianna J. Klein, Dr. Gaofeng Cui, Prof. Dr. Georges Mer, Prof. Dr. Tatiana G. Kutateladze, Maria Victoria Botuyan, Kevin Lin, Yue Lu and Xiaolu Wang. This thesis could not be finished without their work.

I sincerely thank Dr. Yan Wang, Dr. Lu Zhao and Jiwei Qin. They do not only give me help in my doctoral work, but also in my life. It is lucky for me to work with them. I learnt a lot from the communications and interactions with them for their intelligence and actions. Without their big support and sound advice, I could not get to this step.

I would like to thank all my lab members: Dr. Peter Camaj, Dr. Dominik Modest, Susanne Bonifatius, Dr. Christoph Härtl, Dr. Popp Felix, Dr. Thomas Wartmann, Ms. Franziska Braun, Prof. Dr. Walter Halangk, for their help on my work, including suggestions, discussion and technique assistance.

I am grateful to the lab members in Shi Lab: Dr. Hong Wen, Danni Peng, Dr. Shiming Jiang, Dr. Chih-Chao Hsu, Dr. Yun Seong Jeong, Dr. Leilei Shi, Dr. Dan Su, Dr. Wenyi Mi, Yongming Xue, XiaoLu Wang. Thanks for their patience and enthusiasm whenever I had questions or got into difficulties. It was really a wonderful experience to be with them for my visiting in MD Anderson Cancer Center.

Special thank Prof. Dr. Alexender Bathemann, for his full recommendation and big encouragement during my study and life in Munich.

I would also appreciate Chinese Scholarship Council to provide the economic support since 2013. Thank Mr. Ji Qiang Dai and Mr. Chongling Huang from Chinese Consulate Munich for their concern and help. Same thank to Dr. Stefan Lauterbach, Mrs. Monique Esnouf, Dr. Dong Mei Zhang from the International Office of LMU. They provided a well-organized orientation course and living assistance since I arrived in Munich.

I want to give especial thank to my friends in Munich: Hao Feng, Shuxia Fan, Wenwen Fu, Liang Jin, Fang Zhou, Yifei Zhang and Hongshuo Chen. They are always glad to help me, willing to share the happiness with me. They enriched my life.

Most important, I would thank my family, for their endless love, inspiration and support. And special gratitude to my husband Dr. Chao Yuan for he is always there, with faith, hope and love. To them, I dedicated this thesis.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office
Medical Faculty

Affidavit

Wang, Xiaoyan

Surname, first name

Street

Zip code, town

Country

I hereby declare, that the submitted thesis entitled

The Epigenetic Role of PHF20/MOF Histone Acetyltransferase Complex in Lung Cancer

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Gießen 14.12.2017

Place, date

Xiaoyan Wang

Signature doctoral candidate