MOLECULAR ONCOLOGY 10 (2016) 292–302

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

An integrated genomic analysis of Tudor domain–containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer

Yuanyuan Jiang[a], *Lanxin Liu*[a], *Wenqi Shan*[a], *Zeng-Quan Yang*[a,b,*]

[a]*Department of Oncology, Wayne State University School of Medicine, Detroit, MI 48201, USA*
[b]*Molecular Therapeutics Program, Barbara Ann Karmanos Cancer Institute, Detroit, MI 48201, USA*

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form
15 October 2015
Accepted 16 October 2015
Available online 28 October 2015

Keywords:
Breast cancer
Copy number alteration
Tudor domain
Methylation
PHF20L1

ABSTRACT

Tudor domain–containing proteins (TDRDs), which recognize and bind to methyl-lysine/arginine residues on histones and non-histone proteins, play critical roles in regulating chromatin architecture, transcription, genomic stability, and RNA metabolism. Dysregulation of several TDRDs have been observed in various types of cancer. However, neither the genomic landscape nor clinical significance of TDRDs in breast cancer has been explored comprehensively. Here, we performed an integrated genomic and transcriptomic analysis of 41 TDRD genes in breast cancer (TCGA and METABRIC datasets) and identified associations among recurrent copy number alterations, gene expressions, clinicopathological features, and survival of patients. Among seven TDRDs that had the highest frequency (>10%) of gene amplification, the plant homeodomain finger protein 20-like 1 (*PHF20L1*) was the most commonly amplified (17.62%) TDRD gene in TCGA breast cancers. Different subtypes of breast cancer had different patterns of copy number and expression for each TDRD. Notably, amplification and overexpression of *PHF20L1* were more prevalent in aggressive basal-like and Luminal B subtypes and were significantly associated with shorter survival of breast cancer patients. Furthermore, knockdown of PHF20L1 inhibited cell proliferation in *PHF20L1*-amplified breast cancer cell lines. PHF20L1 protein contains N-terminal Tudor and C-terminal plant homeodomain domains. Detailed characterization of PHF20L1 in breast cancer revealed that the Tudor domain likely plays a critical role in promoting cancer. Mechanistically, PHF20L1 might participate in regulating DNA methylation by stabilizing DNA methyltransferase 1 (DNMT1) protein in breast cancer. Thus, our results demonstrated the oncogenic potential of PHF20L1 and its association with poor prognostic parameters in breast cancer.

Abbreviations: TDRD, Tudor domain–containing protein; PHD, plant homeodomain; PHF20L1, PHD finger protein 20-like 1; DNMT1, DNA (cytosine-5-)-methyltransferase 1; TCGA, The Cancer Genome Atlas; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; CNA, copy number alteration.

* *Corresponding author*. Barbara Ann Karmanos Cancer Institute, 4100 John R Street, HWCRC 815, Detroit, MI 48201, USA. Tel.: +1 (313)576 8339; fax: +1 (313)576 8029.
E-mail address: yangz@karmanos.org (Z.-Q. Yang).

http://dx.doi.org/10.1016/j.molonc.2015.10.013

1. Introduction

Methylation of lysine (K) and arginine (R) residues on histones and non-histone proteins plays critical roles in chromatin function, transcriptional regulation, genomic stability, and RNA metabolism (Barski et al., 2007; Chen et al., 2011; Greer and Shi, 2012; Hamamoto et al., 2015; Wei et al., 2014). These epigenetic methylations are mediated by antagonistic sets of enzymatic complexes—the methyltransferases, which catalyze methylation in a site-specific manner, and the demethylases, which remove the methylation marks (Greer and Shi, 2012). Such methylation marks are interpreted by "reader" proteins that specifically bind to the modified protein (Musselman et al., 2012). The largest and most diverse set of reader proteins includes the Tudor domain and plant homeodomain (PHD) (Gayatri and Bedford, 2014; Lu and Wang, 2013; Musselman et al., 2012; Sanchez and Zhou, 2011). In the human proteome, there are more than 40 Tudor domain–containing proteins (TDRDs) (Lu and Wang, 2013; Pek et al., 2012). The Tudor domain exists singly or in multiple copies, in the absence or in conjunction with other functional domains, such as the Jumonji C (JmjC) domain in histone demethylases KDM4A, KDM4B, and KDM4C; the SET domain in histone methyltransferase SETDB1, RING-finger type E3 ubiquitin ligases in UHRF1 and 2; the PHD domain in PHD finger protein 20 (PHF20) and PHF20L1; and the AT-rich interaction domain (ARID) in ARID4A and ARID4B (Lu and Wang, 2013; Pek et al., 2012). Broadly, TDRDs can be divided into two subfamilies based on their ability to bind methyl-lysine or methyl-arginine. Methyl-lysine-binding TDRDs are predominantly involved in histone modification and chromatin remodeling, and methyl-arginine-binding TDRDs are frequently associated with RNA metabolism, alternative splicing, small-RNA pathways, and germ cell development (Chen et al., 2011; Lu and Wang, 2013; Pek et al., 2012; Wei et al., 2014).

Dysregulation of several TDRDs has been observed in breast cancer. Recent studies in human cancer have identified frequent genetic alterations in genes encoding chromatin regulatory factors and histone proteins (Cancer Genome Atlas, 2012; Gonzalez-Perez et al., 2013; Tamborero et al., 2013). Those studies implicate such alterations as major players in the pathogenesis of both hematological malignancies and solid tumors, including breast cancer (Gonzalez-Perez et al., 2013; Pon and Marra, 2015; Roy et al., 2014; Schwartzentruber et al., 2012). All three Tudor domain–containing demethylases (KDM4A, B, and C) are frequently overexpressed in breast cancer (Berry and Janknecht, 2013; Labbe et al., 2013; Liu et al., 2009; Ye et al., 2015). We demonstrated that KDM4C is significantly amplified and overexpressed in aggressive basal-like breast cancers and functions as a transforming oncogene (Liu et al., 2009). Similarly, ARID4B is overexpressed in some estrogen receptor (ER)-positive breast cancers and can be used to predict the development of metastatic disease (Goldberger et al., 2013). Another TDRD, UHRF2, might contribute to tighter epigenetic control of key cell-cycle inhibitors and to regulating cell proliferation in breast cancer (Bronner et al., 2007; Wu et al., 2012).

Although dysregulation of TDRDs has been associated with the initiation and progression of cancer, the genomic landscape and clinical significance of TDRDs in breast cancer have not yet been comprehensively investigated. Thus, we interrogated cancer genomics data and functional small-interfering RNA (siRNA) screens to pinpoint potential oncogenes, focusing on TDRDs. We demonstrated the higher frequency of PHF20L1 amplification and overexpression in breast cancers and cell lines, association with patient survival and its role in viability, indicating that PHF20L1 is a novel oncogene in breast cancer.

2. Materials and methods

2.1. Cell culture

The cultures for the SUM series of breast cancer cell lines and the nontransformed human mammary epithelial cell line MCF10A have been described previously (Forozan et al., 1999; Yang et al., 2006). The Colo824 cell line was obtained from DSMZ (Braunschweig, Germany), SUM cell lines were obtained from Dr. Stephen P. Ethier, and all other cell lines in this study were obtained from ATCC (Manassas, VA, USA).

2.2. The Cancer Genome Atlas (TCGA) data for breast cancer

The DNA copy number, mutation, and overall survival datasets of 959 breast cancer samples used in this research were obtained from the cBio Cancer Genomics Portal (Cerami et al., 2012; Gao et al., 2013). The copy number for each TDRD was generated from the copy number analysis algorithm GISTIC (Genomic Identification of Significant Targets in Cancer) and categorized as copy number level per gene: "−2" is a deep loss (possibly a homozygous deletion), "−1" is a heterozygous deletion, "0" is diploid, "1" indicates a low-level gain, and "2" is a high-level amplification. For mRNA expression data, the relative expression of an individual gene and the gene's expression distribution in a reference population were analyzed. The reference population was either all tumors that are diploid for the gene in question, or, when available, normal adjacent tissue. The returned value indicates the number of standard deviations away from the mean of expression in the reference population (Z-score). Somatic mutation data were obtained from exome sequencing (Cerami et al., 2012; Gao et al., 2013). Breast cancer subtype and clinicopathologic information were extracted *via* the UCSC Cancer Genomics Browser (genome-cancer.ucsc.edu) and the cBio Cancer Genomics Portal (Cancer Genome Atlas, 2012; Cerami et al., 2012; Gao et al., 2013). Among the 959 breast cancer samples, 808 had subtype data available, including 22 normal-like, 405 Luminal A, 185 Luminal B, 66 HER2+, and 130 basal-like breast cancers (Supplementary Table S1) (Gao et al., 2013; Liu et al., 2015).

2.3. The METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset

The METABRIC dataset contains approximately 2000 primary breast cancers with long-term clinical follow-up. A detailed description of the dataset can be obtained from the original manuscript (Supplementary Table S1A) (Curtis et al., 2012). The copy number aberrations and normalized expression data of METABRIC were downloaded with access permissions from the European Genome-phenome Archive (https://www.ebi.ac.uk/ega) under accession number EGAC00000000005. In METABRIC dataset, copy number $\log_2$ ratios were segmented with two analytical methods, circular binary segmentation (CBS) and an adapted hidden Markov model (HMM). The median of the $\log_2$ ratio was computed and gene-centric alterations were categorized as amplification, gain, heterozygous loss and homozygous loss. The data for 41 TDRDs were based on the CBS-derived copy number profiles (Curtis et al., 2012). The normalized gene expression profiles were generated using the Illumina Human HT-12 platform (Curtis et al., 2012). For PHF20L1 expression analysis, we selected Illumina probes indicated as having "Perfect" evidence in the annotation.

2.4. Semiquantitative PCR reactions

mRNA was prepared from human breast cancer cell lines and the MCF10A cell line by using an RNeasy Plus Mini Kit (QIAGEN). mRNA was mixed with qScript cDNA SuperMix (Quanta Biosciences, Gaithersburg, MD, USA) then converted into cDNA through a reverse-transcription (RT) reaction for real-time PCR reactions. Primer sets were ordered from Life Technologies (Carlsbad, CA, USA). A PUM1 primer set was used as a control. Semiquantitative RT-PCR was performed using the FastStart Universal SYBR Green Master (Roche Diagnostics Indianapolis, IN, USA).

2.5. Immunoblotting and antibodies

Whole-cell lysates were prepared by scraping cells from dishes into cold RIPA lysis buffer. After centrifugation at high speed, protein content was estimated by the Bradford method. A total of 20–50 μg of total cell lysate was resolved by SDS–polyacrylamide gel electrophoresis and transferred onto a polyvinylidene difluoride membrane. Antibodies used in the study included anti-PHF20L1 (1:1000, HPA028417, Sigma–Aldrich, St. Louis, MO, USA), anti-DNMT1 (1:1000, #5032, Cell Signaling Technology, Beverly, MA, USA), and anti-β-actin (1:3000, T8328, Sigma–Aldrich).

2.6. PHF20L1 siRNA knockdown

Three PHF20L1 siRNAs were purchased from Sigma–Aldrich. A MISSION siRNA Universal Negative Control #1 (SIC001, Sigma–Aldrich) was used as the negative control in the knockdown experiment. For transfection, 10–30 nM of siRNA was transfected using the MISSION siRNA transfection reagent according to the manufacturer's protocol (Sigma–Aldrich). Then 5–7 days after siRNA transfection, mRNA expression and protein levels of the PHF20L1 were measured by qRT-PCR and western blot. Relative cell growth was measured using CellTiter-Blue® Cell Viability Assay (Promega, Sunnyvale, CA, USA).

2.7. Statistical analysis

Statistical analyses were performed using R software (http://www.r-project.org) and GraphPad Prism (version 6.03). Correlations between copy numbers and mRNA levels of each TDRD from 959 sequenced TCGA breast cancer specimens were analyzed using Spearman, Kendall, and Pearson correlation tests. The Spearman and Kendall tests are rank correlations—the Spearman coefficient relates the two variables while conserving the order of data points, and the Kendall coefficient measures the number of ranks that match in the dataset. We used the "cor" function in R statistical software for computation, specifying in the code which type of test we wanted (Spearman, Kendall, or Pearson). The significance of difference in mRNA expression level for each TDRD between the basal-like and the other cancer subtypes was calculated using Student's *t*-test. The association between the clinical outcome and individual TDRD copy number and expression level was evaluated using a log-rank test. To investigate how DNA copy number was associated with survival, samples were segregated into the following three groups for each TDRD: amp/gain (high-level amplification and low-level gain), diploid, or deletion (heterozygous and homozygous deletion). To analyze the relationships between TDRD mRNA expression and overall patient survival in breast cancer, samples were divided into lower and higher expression groups of each TDRD, based on mRNA expression Z-scores [RNA-Seq V2 RSEM (RNA-Seq by Expectation-Maximization)] in TCGA dataset or the $\log_2$ normalized expression level in METABRIC dataset (Supplementary Table S1). When we compared amp/gain vs. diploid or high vs. low mRNA expression for each TDRD in survival analysis, we segregated samples based on copy number or expression level in whole dataset independent of subtype. Multivariate survival analysis was conducted in TCGA breast cancer samples using the Cox regression function ("coxph") in the R statistical programming language. Factors included in the multivariate analysis model were age at diagnosis (continuous variable), ER status (positive vs. negative), progesterone receptor status (positive vs. negative), HER2 status (positive vs. negative), tumor size (>20 mm vs. ≤20 mm), lymph node status (positive vs. negative), metastasis status (positive vs. negative), MKI67 (marker of proliferation Ki-67) expression, and PAM50 subtype (basal vs. non-basal) (Supplementary Table S1B) (Bertucci et al., 2013).

3. Results and discussion

3.1. Genetic alteration of TDRDs in breast cancer

DNA copy number alterations (CNAs) can lead to the activation of oncogenes or inactivation of tumor suppressors in human cancers (Albertson, 2006; Albertson et al., 2003; Beroukhim et al., 2010). For example, GISTIC analysis revealed that *HER2*, *CCND1*, and *MYC*, three well-known and commonly amplified breast cancer oncogenes, exhibited high-level

amplification in 12.5%, 15.8%, and 22.1% of TCGA breast cancer samples, respectively (Cerami et al., 2012; Gao et al., 2013). We hypothesized that TDRDs with recurrent CNA might play important roles in breast cancer initiation and progression and could serve as therapeutic targets. Based on ChromoHub database, the human genome encodes 41 TDRDs (Supplementary Figure S1 and Table S2) (Liu et al., 2012). We first analyzed copy numbers and mutations of these 41 TDRDs compiled from 959 TCGA breast cancer specimens *via* cBioPortal (Cerami et al., 2012; Gao et al., 2013). As shown in Table 1, we discovered distinct patterns of CNAs and mutations of TDRDs in breast cancer. We found that seven TDRD genes (*PHF20L1*, *ARIB4B*, *SETDB1*, *LBR*, *TDRKH*, *TDRD10*, and *TDRD5*) exhibited high-level amplification in more than 10% of TCGA breast cancers. *PHF20L1* had the highest frequency (17.62%) of amplification. No TDRD genes showed homozygous deletion or somatic mutation in more than 2% of breast cancers. *SMN1* and *SMN2* had the highest frequency of homozygous deletion (1.46%), and *TP53BP1* had the highest frequency of mutation (1.46%) among these TCGA breast cancer samples.

Breast cancers are highly heterogeneous, consisting of five subtypes—Luminal A, Luminal B, human epidermal growth factor receptor 2 (HER2/ERBB2)-enriched, basal-like, and normal-like breast cancers (Carey et al., 2006). Both Luminal A and Luminal B breast cancers are ER positive, but Luminal B cancers have poorer outcomes (Ades et al., 2014; Creighton, 2012). Basal-like breast cancer is a particularly aggressive subtype because it is more prevalent in young

Table 1 – Frequency (%) of TDRD genetic alterations and expression levels in TCGA breast cancers.

Gene	Location	DNA alterations						mRNA expression levels		
		Amp	Gain	Diploid	Hetloss	Homdel	Mutation	Z score ≥ 1	1 > Z score > −1	Z score ≤ −1
PHF20L1	8q24.22	17.62	42.02	35.56	4.48	0.31	0.73	41.40	52.03	6.57
ARID4B	1q42.1–q43	14.60	59.44	23.25	2.61	0.10	1.15	43.90	48.80	7.30
SETDB1	1q21	14.08	58.08	25.96	1.88	0.00	1.15	56.00	37.75	6.26
LBR	1q42.1	13.66	61.11	22.84	2.40	0.00	0.52	18.87	80.50	0.63
TDRKH	1q21	13.03	59.12	25.96	1.88	0.00	0.31	45.57	49.43	5.01
TDRD10	1q21.3	12.20	60.79	25.55	1.46	0.00	0.21	8.55	91.45	0.00
TDRD5	1q25.2	10.64	63.19	23.98	2.19	0.00	0.63	24.61	75.39	0.00
AKAP1	17q22	8.13	27.63	46.40	17.83	0.00	0.42	23.15	65.48	11.37
ZGPAT	20q13.3	6.36	42.23	47.13	4.28	0.00	0.21	17.00	71.22	11.78
KAT8	16p11.2	4.59	46.92	40.04	8.45	0.00	0.52	27.01	62.67	10.32
CCDC101	16p11.2	4.07	47.97	40.56	7.30	0.10	0.21	21.79	71.64	6.57
FXR1	3q28	3.86	26.80	63.82	5.53	0.00	0.63	30.03	57.14	12.83
PHF20	20q11.22–q11.23	3.13	40.98	50.89	5.01	0.00	0.31	27.53	59.12	13.35
TDRD12	19q13.11	2.71	22.31	60.90	14.08	0.00	0.52	6.36	93.64	0.00
KDM4C	9p24.1	2.40	14.91	52.76	29.20	0.73	0.83	18.35	56.20	25.44
UHRF2	9p24.1	2.40	14.81	52.87	29.20	0.73	0.10	16.89	63.61	19.50
KDM4A	1p34.1	1.98	13.24	57.04	27.74	0.00	0.83	16.58	59.75	23.67
KDM4B	19p13.3	1.88	14.18	58.19	25.65	0.10	0.42	13.03	60.06	26.90
STK31	7p15.3	1.77	30.14	57.14	10.84	0.10	1.36	10.74	89.26	0.00
KAT5	11q13	1.67	17.62	59.85	20.65	0.21	0.00	15.75	60.90	23.36
MORF4L1	15q24	1.67	12.41	58.29	27.53	0.10	0.31	18.87	58.71	22.42
SND1	7q31.3	1.46	25.23	58.81	14.49	0.00	0.31	18.14	65.28	16.58
TDRD6	6p12.3	1.46	21.27	61.52	15.54	0.21	1.15	14.18	84.57	1.25
UHRF1	19p13.3	1.25	13.87	58.71	26.07	0.10	0.31	16.27	78.83	4.90
FMR1	Xq27.3	1.15	15.33	66.32	17.00	0.21	0.63	17.73	73.62	8.65
MTF2	1p22.1	1.04	11.99	53.28	33.68	0.00	0.31	15.64	61.73	22.63
RNF17	13q12.12	1.04	11.37	48.18	38.58	0.83	0.52	0.73	99.27	0.00
MSL3	Xp22.3	0.83	15.85	64.55	18.04	0.73	0.52	16.68	75.50	7.82
PHF1	6p21.3	0.83	23.46	61.31	14.39	0.00	0.31	11.57	71.01	17.41
TDRD9	14q32.33	0.83	14.60	56.00	28.26	0.31	0.42	6.15	93.85	0.00
PHF19	9q33.2	0.73	14.49	58.39	26.38	0.00	0.31	16.79	69.55	13.66
ARID4A	14q23.1	0.73	14.29	58.08	26.90	0.00	0.31	14.29	63.30	22.42
TDRD3	13q21.2	0.73	8.65	47.86	41.61	1.15	0.31	6.26	83.94	9.80
TP53BP1	15q15–q21	0.42	6.67	56.83	35.56	0.52	1.46	13.14	59.75	27.11
TDRD15	2p24.1	0.42	14.29	67.26	17.94	0.10	0.00	N/A	N/A	N/A
TDRD1	10q25.3	0.31	9.91	60.17	29.20	0.42	0.42	9.59	90.41	0.00
TDRD7	9q22.33	0.31	12.10	57.87	29.41	0.31	0.31	14.81	66.01	19.19
SMNDC1	10q23	0.10	8.76	60.17	30.45	0.52	0.10	16.37	69.24	14.39
SMN1	5q13.2	0.10	17.94	54.33	26.17	1.46	0.00	10.64	80.08	9.28
SMN2	5q13.2	0.10	17.94	54.33	26.17	1.46	0.00	15.54	70.18	14.29
FXR2	17p13.1	0.10	4.90	33.37	60.48	1.15	0.52	7.40	46.30	46.30

Amp = high-level amplification; Gain = low-level gain; Hetloss = heterozygous deletion; Homdel = homozygous deletion.
Genes were ranked based on the frequency of high-level amplification.

women who often relapse rapidly, resulting in poor prognoses (Bertucci et al., 2012; Cancer Genome Atlas, 2012; Carey et al., 2006). To determine whether the genetic alteration of each TDRD is specific to a breast cancer subtype, we performed an independent analysis of CNA across the five subtypes in TCGA dataset (Supplementary Table S3). Notably, the most frequently amplified TDRD, *PHF20L1*, showed high-level amplification in 35.38% of basal, 25.76% of HER2+, and 27.57% of Luminal B, but only 8.15% of Luminal A and 13.64% of normal-like samples (Figure 1A).

We next examined the relative mRNA expression levels of each TDRD in TCGA breast cancer samples (Cerami et al., 2012; Gao et al., 2013). As shown in Table 1, we found that mRNA was overexpressed in five TDRDs (*PHF20L1, ARID4B, SETDB1, TDRKH*, and *FXR1*) in more than 30% of breast cancers, and *SETDB1* exhibited the highest frequency (56%) of mRNA overexpression. In contrast, only one TDRD gene, *FXR2*, showed low expression in more than 30% of breast cancers. We also analyzed the correlation between copy number and mRNA level of 40 TDRDs (excluding that of *TDRD15*, as its RNA-sequencing data were not available in the cBioPortal database) from 959 sequenced TCGA breast cancer specimens. As shown in Supplementary Table S4, except for *TDRD10*, the correlations of DNA copy number and mRNA expression were positive for 39 TDRD genes, 11 of which (*SETDB1, FXR2, PHF20L1, KDM4C, TDRD3, FXR1, UHRF2, AKAP1, TDRKH, MORF4L1*, and *MTF2*) had a Spearman correlation coefficient (*r*) greater than 0.5.

Next, to determine whether mRNA expression is associated with a specific subtype of breast cancer, we compared expression levels of each TDRD in the 808 TCGA breast cancer samples that had subtype information available (Figure 1B and Supplementary Table S1 and S5). We found that the mRNA expression level of *PHF20L1* was significantly higher in basal and Luminal B subtypes (Figure 1C; $p < 0.001$). *ARID4B* had a higher expression level in both Luminal A and B subtypes, but *ARID4B* amplification exhibited a higher frequency in the basal subtype (Figure 1C). We found that three luminal breast cancer cell lines also exhibited higher mRNA expression of *ARID4B* (Supplementary Figure S2). The correlation

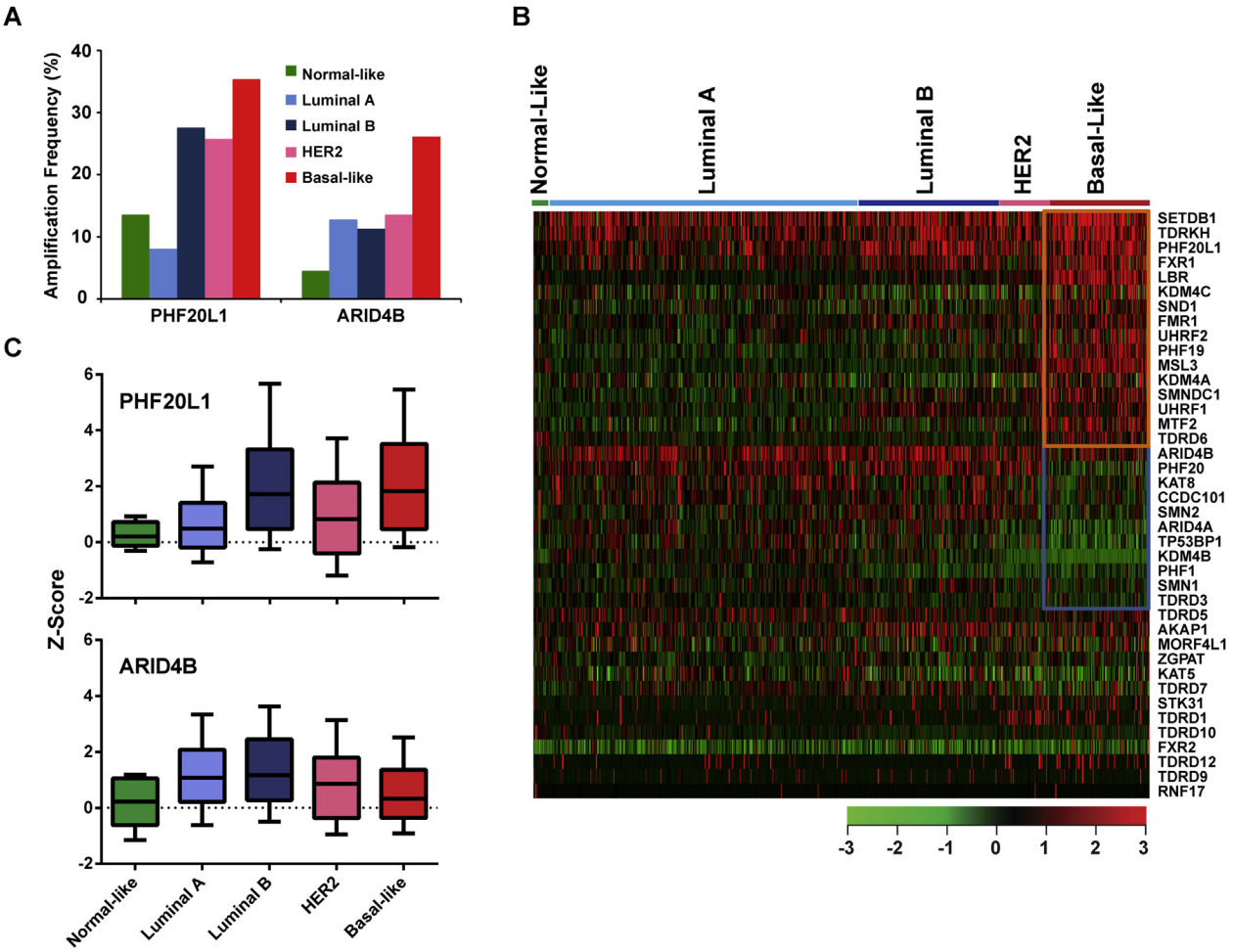


Figure 1 – (A) High-level amplification of *PHF20L1* and *ARID4B* across five subtypes of breast cancer. (B) Heatmap of TDRD expression profiles in different subtypes of breast cancer. The significance of difference in mRNA expression level for each TDRD between the basal-like and other subtypes was calculated using Student's *t*-test. Genes with significantly higher expression ($p < 0.001$) in basal-like tumors are highlighted in red, and genes with lower expression in basal-like tumors are in green. (C) Expression levels of *PHF20L1* and *ARID4B* across five subtypes of breast cancer, based on TCGA data. The differences in *PHF20L1* and *ARID4B* mRNA levels among breast cancer subtypes are statistically significant ($p < 0.001$).

between copy number and mRNA level of *ARID4B* was less than 0.5 in Spearman, Kendall, and Pearson analyses, suggesting that other regulators might contribute to the upregulation of ARID4B in luminal breast cancer (Supplementary Table S4). Indeed, a recent study demonstrated that microRNA miR-290 regulates ARID4B expression while simultaneously enhancing the ER signaling pathway in breast cancer (Goldberger et al., 2013).

To validate our findings of TDRD genomic alterations from the TCGA breast cancer dataset, we conducted an independent analysis using the METABRIC dataset. We found that, among 41 TDRD genes, *PHF20L1* also had the highest frequency of amplification in METABRIC breast cancer samples (Supplementary Table S6), although the frequency of high-level amplification identified in METABRIC dataset is lower than that of TCGA dataset, possibly due to the different CNA analysis platform and calling algorithms. We also found that basal and Luminal B breast cancer had the highest frequency of *PHF20L1* gain/amplification in METABRIC dataset (Supplementary Figure S3A). Again, mRNA expression level of *PHF20L1* was significantly higher in basal and Luminal B breast cancer (Supplementary Figure S3B; $p < 0.001$).

3.2. Association of PHF20L1 amplification/expression with breast cancer patient survival

To investigate the clinical relevance of genetic alterations of TDRDs in breast cancer, we first examined the relationship between TDRD copy number, mRNA expression, and overall patient survival in TCGA breast cancer samples, despite the relatively short follow-up time (median 2.1 years). We found that for only two TDRD genes (*PHF20L1* at 8q24.22 and *AKAP1* at 17q22), copy number gain/amplification was significantly associated ($p < 0.05$) with shorter survival in breast cancer patients (Figure 2A and Supplementary Table S7). For *PHF20L1*, the group with genetic gain/amplification had a hazard ratio (HR), a ratio of the probability of death, of 1.59 [95% confidence interval (CI), 1.035–2.305] compared to the breast cancer patients with diploid *PHF20L1*.

Next, to analyze the relationship between TDRD mRNA expression and overall breast cancer patient survival, samples were divided into lower and higher expression groups based on the mRNA expression Z-scores of each TDRD in TCGA dataset. Supplementary Table S8 summarizes the results of a log-rank statistical analysis for 40 TDRDs in breast cancer. High mRNA levels of *PHF20L1* ($p = 0.0027$, HR = 1.78), *PHF20* ($p = 0.0114$, HR = 1.61), and *STK31* ($p = 0.0329$, HR = 1.50) were significantly associated ($p < 0.05$) with shorter survival in TCGA breast cancer patients (Figure 2B, Supplementary Table S8). Extending this analysis to the METABRIC dataset, which has a longer median survival time (7.3 years), confirmed that higher expression of *PHF20L1* was correlated with a poor prognosis ($p = 0.0003$, HR = 1.27) (Supplementary Figure S4). We then examined the relationship between *PHF20L1* expression and overall patient survival in Luminal B and basal subtypes, which exhibited higher frequencies of *PHF20L1* amplification and overexpression. We found that high mRNA levels of *PHF20L1* were significantly associated ($p = 0.041$, HR = 1.30) with shorter survival in Luminal B, but not in basal, METABRIC breast cancer patients.

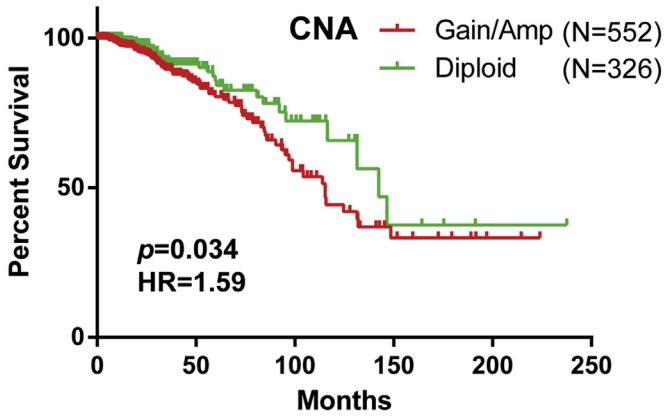
Next, to examine the prognostic value of PHF20L1, we performed a multivariate analysis to investigate whether the expression level of PHF20L1 was predictive of poor prognosis compared with standard prognostic markers. Again, a high mRNA level of *PHF20L1*, but not other TDRDs, was independently associated with shorter survival of TCGA breast cancer patients ($p = 0.029$, HR = 1.84) (Figure 2C and Supplementary Table S9).

3.3. Genetic and functional characterization of PHF20L1 in breast cancer

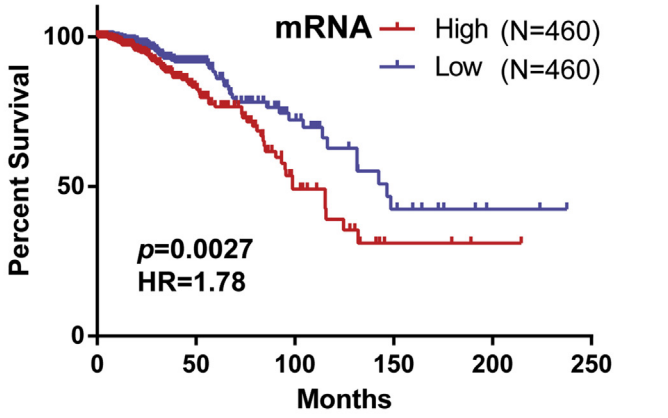
In our analysis of the genetic alterations of TDRDs in breast cancer, we were surprised by the importance of PHF20L1, because it had a higher frequency of amplification and mRNA overexpression in breast cancer, and its amplification/overexpression was associated with poor prognosis. Thus, we further examined the genetic alterations and transforming roles of PHF20L1 in breast cancer by using a panel of cell lines. Based on our data and published array-based comparative genomic hybridization (CGH) data, we found that, among 20 lines, 5 breast cancer cell lines had *PHF20L1* amplification and 10 had gain (Barretina et al., 2012; Wu et al., 2012; Yang et al., 2006). Consistent with the data from primary breast cancer specimens, we found that breast cancer cell lines that have genetic gain/amplification also exhibited higher mRNA expression of *PHF20L1* (Figure 3A). Previous studies found that PHF20L1 has three isoforms, PHF20L1a, -b, and -c, and they all contain the Tudor domain (Supplementary Figure S5) (Esteve et al., 2014). So next we performed western blotting to detect PHF20L1 protein levels in a panel of breast cancer cell lines by using an antibody (HPA028417, Sigma–Aldrich) raised from a region common to all isoforms. As shown in Figure 3B, we found that all three isoforms of PHF20L1 were overexpressed at the protein level, but only PHF20L1c was dramatically higher in several breast cancer cell lines, including the *PHF20L1*-amplified MDA-MB-468, HCC1937, and Colo824 cell lines (Figure 3B and Supplementary Figure S6).

To further assess the contribution of endogenous amplification/overexpression of TDRDs, particularly PHF20L1, to the proliferation of human breast cancer, we analyzed a large-scale shRNA-based growth and viability screen database of approximately 16,000 human genes in 78 cancer cell lines (29 breast, 28 pancreatic, 15 ovarian, and 6 colon cancer lines), in which shRNA data were available for 33 TDRDs (Note: the *SETDB1* gene was not in this shRNA screening) (Koh et al., 2012; Marcotte et al., 2012). In the pooled shRNA screen, cells carrying shRNAs that targeted 6 TDRD genes were significantly depleted from the cell population over time in more than 10 cancer cell lines; five of those genes (*PHF20L1*, *TDRD3*, *SMN2*, *SMNDC1* and *SMN1*) were in more than five breast cancer lines, based on the Gene Activity Ranking Profile (GARP) score (GARP < -0.1) (Supplementary Table S10). From the pooled shRNAs, PHF20L1 shRNA-depleted breast cancer lines included five basal lines (HCC1954, MDA-MB-468, HCC1395, HCC1143, and MDA-MB-231) and one HER2+ line (SK-BR-3). Next, we examined the effects of knocking down PHF20L1 by using siRNA in HCC1954 as well as HCC1937 and Colo824 cells, all of which exhibit high-level amplification/

A



B



C

	p value	HR	95% CI
PHF20L1 expression	0.029	1.976	1.073-3.639
Age , continuous	0.002	1.031	1.011-1.051
ER, positive versus negative	0.315	0.601	0.222-1.459
PR, positive versus negative	0.008	0.358	0.167-0.776
HER2, positive versus negative	0.083	0.422	0.159-1.120
MKI67 expression	0.407	0.763	0.402-1.447
pT, PT2-3 versus pT1	0.193	1.571	0.796-3.098
pN, positive versus negative	0.140	1.527	0.870-2.679
Metastasis, positive versus negative	0.018	3.045	1.208-7.677
PAM50, basal versus nonbasal	0.074	0.403	0.149-1.093

Figure 2 – Kaplan–Meier plots of overall survival associated with (A) copy number gain/amplification and (B) mRNA expression levels of *PHF20L1* in TCGA breast cancers. (C) Multivariate analysis revealed that *PHF20L1* expression was independent of prognostic variables influencing overall survival of TCGA breast cancer patients.

expression of PHF20L1. Three siRNAs that target the N-terminal Tudor-containing region were selected, and the knockdown of PHF20L1a, -b, and -c was confirmed by qRT-PCR and western blot assays (Figure 3C and D and Supplementary Figure S7). As shown in Figure 3C, PHF20L1 knockdown slowed HCC1937, Colo824, and HCC1954 cell growth to around 50%–90% of the growth of the non-silenced control. In contrast, knocking down PHF20L1 had only a minor effect on cell growth in MCF10A and SUM102 cells without PHF20L1 amplification (Supplementary Figure S8).

3.4. *Potential function of PHF20L1 Tudor domains in breast cancer*

In the TDRD family, there are 13 members (KDM4A, KDM4B, KDM4C, SETDB1, PHF20, PHF20L1, TP53BP1, FMR1, FXR1,

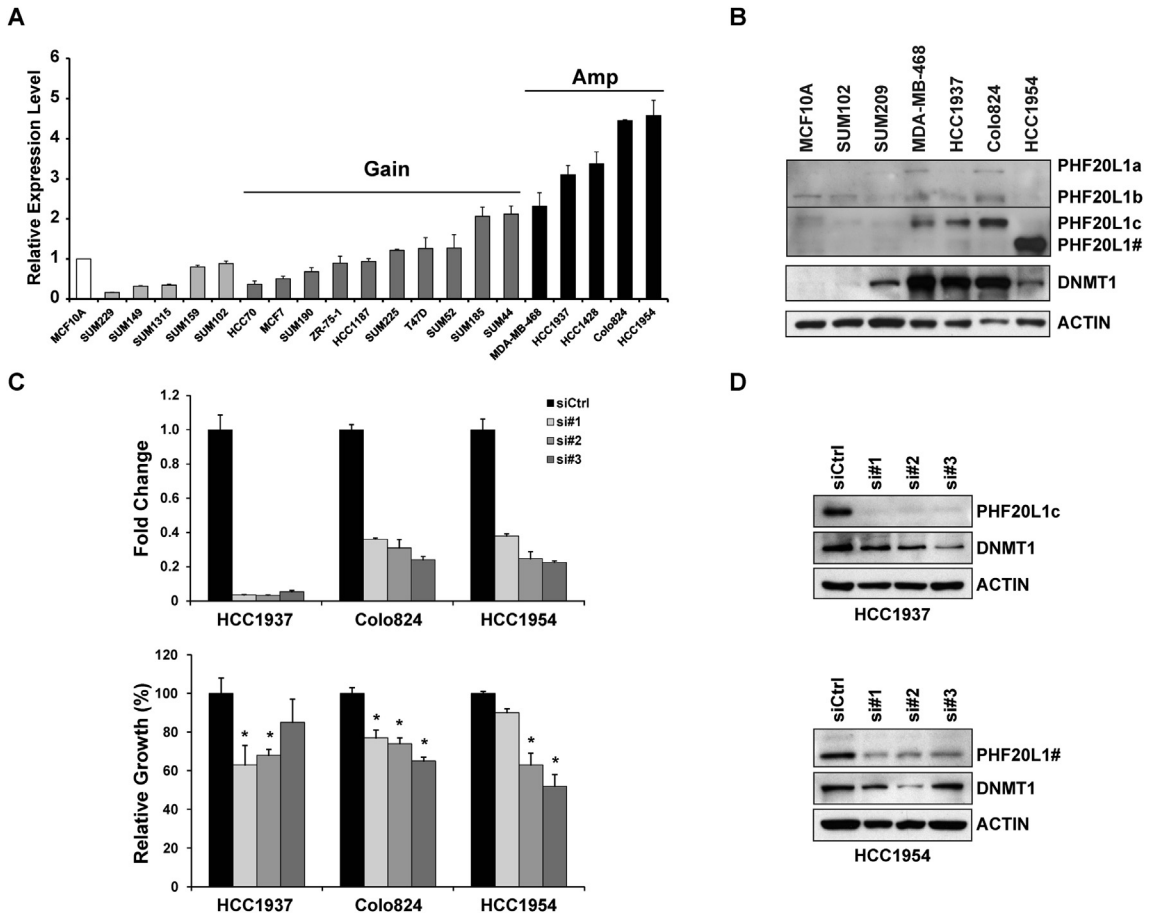


Figure 3 – (A) mRNA expression levels of the *PHF20L1*, measured by qRT-PCR, in a panel of 20 breast cancer cell lines. mRNA expression levels in MCF10A cells were arbitrarily set as 1. Relative expression levels are shown as fold changes compared with that in MCF10A cells (Amp: Amplification). (B) Protein levels of PHF20L1 and DNMT1 were analyzed by western blot in six breast cancer cell lines and the MCF10A cell line. HCC1954 line shows an additional protein band (PHF20L1#) with a lower molecular weight than PHF20L1c (Supplementary Figure S6). Previous studies using high-throughput DNA sequencing have shown that HCC1954 cells are amplified and genomically rearranged at the *PHF20L1* locus, which likely produces a truncated form of PHF20L1 that contains the Tudor domain (Asmann et al., 2011; Zhao et al., 2009). By using qRT-PCR, we confirmed that at the mRNA level, HCC1954 cells expressed a higher level of the N-terminal region but had low or moderate expression of the C-terminal region of PHF20L1 compared with MCF10A cells (Supplementary Figure S10). (C) Knockdown of PHF20L1 inhibits cell proliferation in breast cancer. Top: Knockdowns of PHF20L1 in HCC1937, Colo824, and HCC1954 cells with three different siRNAs were confirmed by qRT-PCR. Bottom: Bar graph shows relative cell growth after knocking down PHF20L1 in HCC1937, Colo824, and HCC1954 breast cancer cells (*$p < 0.05$). Data are shown as mean ± SE. (D) Protein levels of PHF20L1 and DNMT1 were analyzed by western blot in HCC1937 and HCC1954 cells after PHF20L1 knockdown.

FXR2, CCDC101, UHRF1, and UHRF2) that contain two Tudor domains in tandem (TTD: tandem Tudor domains), which provides the possibility of multivalent readout of unique methylation marks on histone and non-histone proteins (Lu and Wang, 2013; Musselman et al., 2012; Pek et al., 2012). Six of them (KDM4A, KDM4B, KDM4C, SETDB1, UHRF1, and UHRF2) possess catalytic domains and thereby also function as effectors; the TTD likely plays a regulatory role in recruiting them to specific chromatin regions (Musselman et al., 2012). PHF20L1 contains an N-terminal TTD and a C-terminal PHD finger domain (Supplementary Figure S5). The N-terminal TTD, but not others, are present in all three PHF20L1 isoforms. Notably, the first Tudor domain was reported to be a Tudor or a malignant brain tumor (MBT) domain, depending on which prediction algorithm was used. When a histone peptide array approach was used, the first Tudor domain of PHF20L1 was found to bind to H3K4me1 and H4K20me1 (Kim et al., 2006). More recently, Esteve et al. demonstrated that the first Tudor domain (also called MBT) of PHF20L1 binds to K142me1 on DNA (cytosine-5) methyltransferase 1 (DNMT1) and antagonizes DNMT1 proteasomal degradation (Esteve et al., 2014). DNMT1 is the major enzyme responsible for maintenance of DNA methylation patterns, and it methylates newly synthesized DNA (Jeltsch and Jurkowska, 2014). DNMT1 is overexpressed in various tumor types, including breast cancer, and it is essential for mammary and cancer stem cell maintenance and tumorigenesis (Montenegro et al., 2015; Pathania et al., 2015).

To determine whether PHF20L1 regulates the stabilization of DNMT1 in breast cancer, we first determined the protein levels of DNMT1 in breast cancer cell lines with or without PHF20L1 amplification/overexpression. We found that DNMT1 was overexpressed at the protein level in *PHF20L1*-amplified HCC1937, Colo824, and MDA-MB-468 cell lines and was moderately expressed in HCC1954 cells (Figure 3B). Furthermore, we found that there are no significantly positive correlation between PHF20L1 and DNMT1 mRNA expression in both TCGA (Pearson's r = 0.12) and METABRIC (Pearson's r = 0.09) breast cancer samples. When PHF20L1 was knocked down in HCC1937 and HCC1954 cells, DNMT1 protein levels decreased, indicating that PHF20L1, most likely *via* a Tudor domain, regulates DNMT1 expression at the protein level (Figure 3D). Next, we analyzed the PHF20L1 amplification and expression in five distinct DNA methylation clusters that are grouped based on the Illumina Infinium DNA methylation array analysis in TCGA breast cancer samples (Cancer Genome Atlas, 2012). We found higher frequency of PHF20L1 amplification/gain and higher PHF20L1 expression in Clusters 3 and 5 breast cancers, in which Cluster 3 had a hypermethylated phenotype, while Cluster 5 was overlapped with the PAM50 basal-like subtype (Supplementary Figure S9) (Cancer Genome Atlas, 2012). These data indicate that PHF20L1 might participate in regulating DNA methylation by stabilizing DNMT1 protein in breast cancer.

Previous biochemical studies also revealed that PHF20L1 is a component of a histone acetyltransferase complex (Badeaux et al., 2012; Mendjan et al., 2006). In addition, the second Tudor (Protein Data bank code: 2EQU) of the PHF20L1 TTD has the canonical Tudor architecture that most likely recognizes dimethylated lysine. Thus, we speculate that PHF20L1 functions as the critical tethering factor, *via* its two Tudor domains, in regulating DNA and histone methylation signals in breast cancer.

4. Conclusion

In summary, we used an integrated genomics and functional screening of TDRDs and identified several novel putative oncogenic TDRDs in breast cancer that could be explored as therapeutic targets. We found that PHF20L1 is frequently amplified and overexpressed in breast cancer, particularly basal-like and Luminal B subtypes. Amplification/overexpression of PHF20L1 was significantly associated with shorter survival of breast cancer patients. Knockdown of PHF20L1 inhibits cell proliferation in breast cancer cell lines. Mechanistically, the PHF20L1 Tudor domain recognizes and binds DNMT1 and prevents its degradation; DNMT1 is a critical factor for maintenance of DNA methylation and breast tumorigenesis. Our results demonstrate the oncogenic potential of PHF20L1 and its association with poor prognostic parameters in breast cancer. Furthermore, our findings provide a strong foundation for further mechanistic research and for developing therapies that target PHF20L1 or other TDRDs in breast cancer.

Conflicts of interest

None of the authors declare conflicts of interest with this work.

Acknowledgments

This work was partially supported by grants from the NIH/NCI and the Office of Research on Women's Health (OWRH) R21CA175244-01A1, the Mary Kay Foundation Cancer Research Grant 094-11, and the Karmanos Cancer Institute-SRIG to Dr. Z-Q.Y.; and by funding from the China Scholarship Council Program fellowship to Wenqi Shan. We thank Hui Liu, Qin Ye, and Lihong Zhang for technical contributions. We thank Dr. Stephen P. Ethier for providing the SUM breast cancer cell lines and for his continuous encouragement.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2015.10.013.

REFERENCES

Ades, F., Zardavas, D., Bozovic-Spasojevic, I., Pugliano, L., Fumagalli, D., de Azambuja, E., Viale, G., Sotiriou, C., Piccart, M., 2014. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. J. Clin. Oncol. 32, 2794–2803.

Albertson, D.G., 2006. Gene amplification in cancer. Trends Genet. TIG 22, 447–455.

Albertson, D.G., Collins, C., McCormick, F., Gray, J.W., 2003. Chromosome aberrations in solid tumors. Nat. Genet. 34, 369–376.

Asmann, Y.W., Hossain, A., Necela, B.M., Middha, S., Kalari, K.R., Sun, Z., Chai, H.S., Williamson, D.W., Radisky, D., Schroth, G.P., Kocher, J.P., Perez, E.A., Thompson, E.A., 2011. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. Nucleic Acids Res. 39, e100.

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

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M.F., Monahan, J.E., Morais, P., Meltzer, J., Korejwa, A., Jane-Valbuena, J., Mapa, F.A., Thibault, J., Bric-Furlong, E., Raman, P., Shipway, A., Engels, I.H., Cheng, J., Yu, G.K., Yu, J., Aspesi Jr., P., de Silva, M., Jagtap, K., Jones, M.D., Wang, L., Hatton, C., Palescandolo, E., Gupta, S., Mahan, S., Sougnez, C., Onofrio, R.C., Liefeld, T., MacConaill, L., Winckler, W., Reich, M., Li, N., Mesirov, J.P., Gabriel, S.B., Getz, G., Ardlie, K., Chan, V., Myer, V.E., Weber, B.L., Porter, J., Warmuth, M., Finan, P., Harris, J.L., Meyerson, M., Golub, T.R.,

Morrissey, M.P., Sellers, W.R., Schlegel, R., Garraway, L.A., 2012. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607.
Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K., 2007. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.
Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., Meyerson, M., 2010. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899–905.
Berry, W.L., Janknecht, R., 2013. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. Cancer Res. 73, 2936–2942.
Bertucci, F., Finetti, P., Birnbaum, D., 2012. Basal breast cancer: a complex and deadly molecular subtype. Curr. Mol. Med. 12, 96–110.
Bertucci, F., Finetti, P., Roche, H., Le Doussal, J.M., Marisa, L., Martin, A.L., Lacroix-Triki, M., Blanc-Fournier, C., Jacquemier, J., Peyro-Saint-Paul, H., Viens, P., Sotiriou, C., Birnbaum, D., Penault-Llorca, F., 2013. Comparison of the prognostic value of genomic grade index, Ki67 expression and mitotic activity index in early node-positive breast cancer patients. Ann. Oncol. 24, 625–632.
Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., Schini-Kerth, V.B., 2007. The UHRF family: oncogenes that are drugable targets for cancer therapy in the near future? Pharmacol. Ther. 115, 419–434.
Cancer Genome Atlas, N, 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.
Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S., Deming, S.L., Geradts, J., Cheang, M.C., Nielsen, T.O., Moorman, P.G., Earp, H.S., Millikan, R.C., 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. J. Am. Med. Assoc. 295, 2492–2502.
Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401–404.
Chen, C., Nott, T.J., Jin, J., Pawson, T., 2011. Deciphering arginine methylation: Tudor tells the tale. Nature reviews. Mol. Cell Biol. 12, 629–642.
Creighton, C.J., 2012. The molecular profile of luminal B breast cancer. Biol. Targets Ther. 6, 289–297.
Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352.
Esteve, P.O., Terragni, J., Deepti, K., Chin, H.G., Dai, N., Espejo, A., Correa Jr., I.R., Bedford, M.T., Pradhan, S., 2014. Methyllysine reader plant homeodomain (PHD) finger protein 20-like 1 (PHF20L1) antagonizes DNA (cytosine-5) methyltransferase 1 (DNMT1) proteasomal degradation. J. Biol. Chem. 289, 8277–8287.
Forozan, F., Veldman, R., Ammerman, C.A., Parsa, N.Z., Kallioniemi, A., Kallioniemi, O.P., Ethier, S.P., 1999. Molecular cytogenetic analysis of 11 new breast cancer cell lines. Br. J. Cancer 81, 1328–1334.
Gao, J.J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y.C., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6.
Gayatri, S., Bedford, M.T., 2014. Readers of histone methylarginine marks. Biochim. Biophys. Acta 1839, 702–710.
Goldberger, N., Walker, R.C., Kim, C.H., Winter, S., Hunter, K.W., 2013. Inherited variation in miR-290 expression suppresses breast cancer progression by targeting the metastasis susceptibility gene Arid4b. Cancer Res. 73, 2671–2681.
Gonzalez-Perez, A., Jene-Sanz, A., Lopez-Bigas, N., 2013. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. Genome Biol. 14, r106.
Greer, E.L., Shi, Y., 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. 13, 343–357.
Hamamoto, R., Saloura, V., Nakamura, Y., 2015. Critical roles of non-histone protein lysine methylation in human tumorigenesis. Nat. Rev. Cancer 15, 110–124.
Jeltsch, A., Jurkowska, R.Z., 2014. New concepts in DNA methylation. Trends Biochem. Sci. 39, 310–318.
Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., Bedford, M.T., 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep. 7, 397–403.
Koh, J.L., Brown, K.R., Sayad, A., Kasimer, D., Ketela, T., Moffat, J., 2012. COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. Nucleic Acids Res. 40, D957–D963.
Labbe, R.M., Holowatyj, A., Yang, Z.Q., 2013. Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. Am. J. Transl. Res. 6, 1–15.
Liu, G., Bollig-Fischer, A., Kreike, B., van de Vijver, M.J., Abrams, J., Ethier, S.P., Yang, Z.Q., 2009. Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. Oncogene 28, 4491–4500.
Liu, H., Liu, L., Holowatyj, A., Jiang, Y., Yang, Z.Q., 2015. Integrated genomic and functional analyses of histone demethylases identify oncogenic KDM2A isoform in breast cancer. Mol. Carcinog. (in press).
Liu, L., Zhen, X.T., Denton, E., Marsden, B.D., Schapira, M., 2012. ChromoHub: a data hub for navigators of chromatin-mediated signalling. Bioinformatics 28, 2205–2206.
Lu, R., Wang, G.G., 2013. Tudor: a versatile family of histone methylation 'readers'. Trends Biochem. Sci. 38, 546–555.
Marcotte, R., Brown, K.R., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P.M., Sircoulomb, F., Medrano, M., Fedyshyn, Y., Koh, J.L., van Dyk, D., Fedyshyn, B., Luhova, M., Brito, G.C., Vizeacoumar, F.J., Vizeacoumar, F.S., Datti, A., Kasimer, D., Buzina, A., Mero, P., Misquitta, C., Normand, J., Haider, M., Ketela, T., Wrana, J.L., Rottapel, R., Neel, B.G., Moffat, J., 2012. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. Cancer Discov. 2, 172–189.
Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., Wilm, M., Stunnenberg, H.G., Saumweber, H., Akhtar, A., 2006. Nuclear pore components are involved in the

transcriptional regulation of dosage compensation in Drosophila. Mol. Cell 21, 811–823.
Montenegro, M.F., Sanchez-del-Campo, L., Fernandez-Perez, M.P., Saez-Ayala, M., Cabezas-Herrera, J., Rodriguez-Lopez, J.N., 2015. Targeting the epigenetic machinery of cancer cells. Oncogene 34, 135–143.
Musselman, C.A., Lalonde, M.E., Cote, J., Kutateladze, T.G., 2012. Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. 19, 1218–1227.
Pathania, R., Ramachandran, S., Elangovan, S., Padia, R., Yang, P.Y., Cinghu, S., Veeranan-Karmegam, R., Arjunan, P., Gnana-Prakasam, J.P., Sadanand, F., Pei, L.R., Chang, C.S., Choi, J.H., Shi, H.D., Manicassamy, S., Prasad, P.D., Sharma, S., Ganapathy, V., Jothi, R., Thangaraju, M., 2015. DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. Nat. Commun. 6.
Pek, J.W., Anand, A., Kai, T., 2012. Tudor domain proteins in development. Development 139, 2255–2266.
Pon, J.R., Marra, M.A., 2015. Driver and passenger mutations in cancer. Ann. Rev. Pathol. 10, 25–50.
Roy, D.M., Walsh, L.A., Chan, T.A., 2014. Driver mutations of cancer epigenomes. Protein Cell 5, 265–296.
Sanchez, R., Zhou, M.M., 2011. The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. 36, 364–372.
Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M., Hovestadt, V., Albrecht, S., Kool, M., Nantel, A., Konermann, C., Lindroth, A., Jager, N., Rausch, T., Ryzhova, M., Korbel, J.O., Hielscher, T., Hauser, P., Garami, M., Klekner, A., Bognar, L., Ebinger, M., Schuhmann, M.U., Scheurlen, W., Pekrun, A., Fruhwald, M.C., Roggendorf, W., Kramm, C., Durken, M., Atkinson, J., Lepage, P., Montpetit, A., Zakrzewska, M., Zakrzewski, K., Liberski, P.P., Dong, Z., Siegel, P., Kulozik, A.E., Zapatka, M., Guha, A., Malkin, D., Felsberg, J., Reifenberger, G., von Deimling, A., Ichimura, K., Collins, V.P., Witt, H., Milde, T., Witt, O., Zhang, C., Castelo-Branco, P., Lichter, P., Faury, D., Tabori, U., Plass, C., Majewski, J., Pfister, S.M., Jabado, N., 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 482, 226–231.
Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M.S., Getz, G., Bader, G.D., Ding, L., Lopez-Bigas, N., 2013. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci. Rep. 3, 2650.
Wei, H., Mundade, R., Lange, K.C., Lu, T., 2014. Protein arginine methylation of non-histone proteins and its role in diseases. Cell Cycle 13, 32–41.
Wu, J., Liu, S., Liu, G., Dombkowski, A., Abrams, J., Martin-Trevino, R., Wicha, M.S., Ethier, S.P., Yang, Z.Q., 2012. Identification and functional analysis of 9p24 amplified genes in human breast cancer. Oncogene 31, 333–341.
Yang, Z.Q., Streicher, K.L., Ray, M.E., Abrams, J., Ethier, S.P., 2006. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer Res. 66, 11632–11643.
Ye, Q., Holowatyj, A., Wu, J., Liu, H., Zhang, L., Suzuki, T., Yang, Z.Q., 2015. Genetic alterations of KDM4 subfamily and therapeutic effect of novel demethylase inhibitor in breast cancer. Am. J. Cancer Res. 5, 1519–1530.
Zhao, Q., Caballero, O.L., Levy, S., Stevenson, B.J., Iseli, C., de Souza, S.J., Galante, P.A., Busam, D., Leversha, M.A., Chadalavada, K., Rogers, Y.H., Venter, J.C., Simpson, A.J., Strausberg, R.L., 2009. Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. Proc. Natl. Acad. Sci. U. S. A. 106, 1886–1891.