MOLECULAR ONCOLOGY XXX (2015) 1–11

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 ■

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
1574-7891/

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).

Please cite this article in press as: Jiang, Y., et al., An integrated genomic analysis of Tudor domain–containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.10.013