Marine Life Science & Technology (2020) 2:109–122
https://doi.org/10.1007/s42995-019-00025-y

RESEARCH PAPER

Functional analysis of the methyltransferase SMYD in the single-cell model organism *Tetrahymena thermophila*

Xiaolu Zhao[1,2,3] · Yuan Li[1,3] · Lili Duan[1,3] · Xiao Chen[1] · Fengbiao Mao[2] · Mina Juma[2] · Yifan Liu[2] · Weibo Song[1,3] · Shan Gao[1,3]

Received: 15 August 2019 / Accepted: 15 October 2019 / Published online: 5 March 2020

Abstract

Lysine methylation of histones and non-histones plays a pivotal role in diverse cellular processes. The SMYD (SET and MYND domain) family methyltransferases can methylate various histone and non-histone substrates in mammalian systems, implicated in HSP90 methylation, myofilament organization, cancer inhibition, and gene transcription regulation. To resolve controversies concerning SMYD's substrates and functions, we studied SMYD1 (TTHERM_00578660), the only homologue of SMYD in the unicellular eukaryote *Tetrahymena thermophila*. We epitope-tagged SMYD1, and analyzed its localization and interactome. We also characterized Δ*SMYD1* cells, focusing on the replication and transcription phenotype. Our results show that: (1) SMYD1 is present in both cytoplasm and transcriptionally active macronucleus and shuttles between cytoplasm and macronucleus, suggesting its potential association with both histone and non-histone substrates; (2) SMYD1 is involved in DNA replication and regulates transcription of metabolism-related genes; (3) HSP90 is a potential substrate for SMYD1 and it may regulate target selection of HSP90, leading to pleiotropic effects in both the cytoplasm and the nucleus.

Keywords *Tetrahymena thermophila* · Methyltransferase · SMYD · DNA replication · Transcription · HSP90

Introduction

Heritable changes can be attributed to changes not only in DNA sequences, but also to changes in epigenetic information, which mainly involve protein post-translational modifications (PTMs; especially histone PTMs), noncoding RNA, and DNA methylation (Allis et al. 2015; Blanc and Richard 2017; Edwards et al. 2017; Goldberg et al. 2007; Guillemette et al. 2011; Liu et al. 2016; Mao et al. 2018; Strzyz 2016; Walter and Hümpel 2017; Wang et al. 2017b, c, d; Zhao et al. 2019). In particular, protein lysine methylation plays pivotal roles in a range of cellular processes (Hamamoto et al. 2015; Jakobsson et al. 2017; Lanouette et al. 2014). Lysine methylation of histones contributes to the formation of various nuclear protein complexes, thus having significant impacts on transcription, DNA repair, and DNA replication (Black et al. 2012; Jakobsson et al. 2017; Moore and Gozani 2014; Ramadoss et al. 2017; Wozniak and Strahl 2014). Lysine methylation of non-histone substrates, while not as extensively studied as the histone counterpart, is increasingly being recognized for diverse roles ranging from transcription regulation to protein stabilization (Buuh et al. 2017; Cao et al. 2013; Chuikov et al. 2004; Huang et al. 2006; Ivanov et al. 2007; Wang et al. 2017a).

The lysine methyltransferase SMYD contains a conserved SET domain (*Drosophila* proteins Suppressor of variegation, Enhancer of Zeste, *Trithorax*) responsible for catalyzing lysine methylation (Herz et al. 2013; Qian and Zhou 2006). This SET domain is interrupted by a MYND domain (Myeloid, Nervy and DEAF-1) (Al-Shar'i and Alnabulsi 2016; Calpena et al. 2015; Li et al. 2017) which is mainly

Edited by Jiamei Li.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42995-019-00025-y) contains supplementary material, which is available to authorized users.

✉ Shan Gao
shangao@ouc.edu.cn

[1] Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao 266003, China

[2] Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA

[3] Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266003, China

responsible for protein–protein interactions with a proline-rich motif (PXLXP) (Arsenault et al. 2016; Calpena et al. 2015; Mazur et al. 2016; Spellmon et al. 2015). In mammals, there are five members of the SMYD family, SMYD1-5, with distinct biological functions (Al-Shar'i and Alnabulsi 2016; Bagislar et al. 2016; Giakountis et al. 2017; Li et al. 2017; Spellmon et al. 2015). SMYD1 is involved in cardiac development and in skeletal muscle growth and regeneration, by interacting with muscle-specific transcription factor skNAC or methylating histone H3K4 (Du et al. 2014; Rasmussen et al. 2015). SMYD2 is implicated in: (1) transcription activation, by methylating H3K4 (Boehm et al. 2017; Diehl et al. 2010); (2) transcription repression, by interacting with the Sin3A histone deacetylase complex (Brown et al. 2006); (3) tumorigenesis, by methylating p53 at K370 (Levine and Berger 2017; Ohtomo-Oda et al. 2016); and (4) HSP90 functions, by methylating different lysine residues of HSP90 to promote its binding to different substrates (Abu-Farha et al. 2011; Donlin et al. 2012; Hamamoto et al. 2014; Voelkel et al. 2013). SMYD3 regulates transcription as part of the RNA polymerase complex and has the ability to di/tri-methylate H3K4 in the presence of HSP90 (Hamamoto et al. 2004; Zhu et al. 2016). SMYD4, as a transcriptional modulator, recruits the HDAC co-repressor complex and represses transcription (Thompson and Travers 2008). Finally, SMYD5 regulates inflammation-related genes by tri-methylating H4K20 (Stender et al. 2012).

SMYD research has been largely limited to mammalian systems, often stereotypically associated with heart and muscle development and tumor suppression. However, SMYD family methyltransferases are widely distributed in eukaryotes, including protozoa and plants (Fig. 1). This suggests that the ancient and most likely conserved roles of SMYD methyltransferases may not be those animal-specific ones currently under extensive investigation. In this work, we characterized the sole homologue of SMYD methyltransferases in the unicellular eukaryote *Tetrahymena thermophila*. Our results provide insights into the evolution of the SMYD family and shed light on the conserved substrates and functions of SMYD methyltransferases, which may benefit studies in other systems.

Results

Phylogenetic analysis of SMYD family methyltransferases

There are five SMYD homologs (SMYD1-5) in vertebrates, such as *Danio rerio*, *Xenopus* (*Silurana*) *tropicalis*, *Gallus gallus*, *Mus musculus*, *Bos taurus*, *Bos mutu*, *Pan troglodytes*, *Pan paniscus*, and *Homo sapiens*. The SMYD family was divided into two main clades (Fig. 1). One branch consisted of SMYD1, 2 and 3, in which SMYD1 clustered with SMYD2 first (78% ML bootstrap support) forming a sister group to SMYD3 (89% ML). The other branch contained SMYD4 and SMYD5 (29% ML). The monophyly of SMYD1-5 was maximally supported (100% ML). This result was consistent with the domain similarity of the SMYD family (Spellmon et al. 2015): (1) SMYD1-3 shared similar domains, including the N terminal SET domain, MYND domain, and the C-terminal tetratricopeptide repeats (TPR) (Jiang et al. 2011; Sirinupong et al. 2010, 2011); (2) the domain structure of SMYD4 and SMYD5 was more divergent (Spellmon et al. 2015); SMYD4 had an extra TPR domain on its N terminus and an extended C-terminal domain, while SMYD5 lacked the C-terminal domain. The

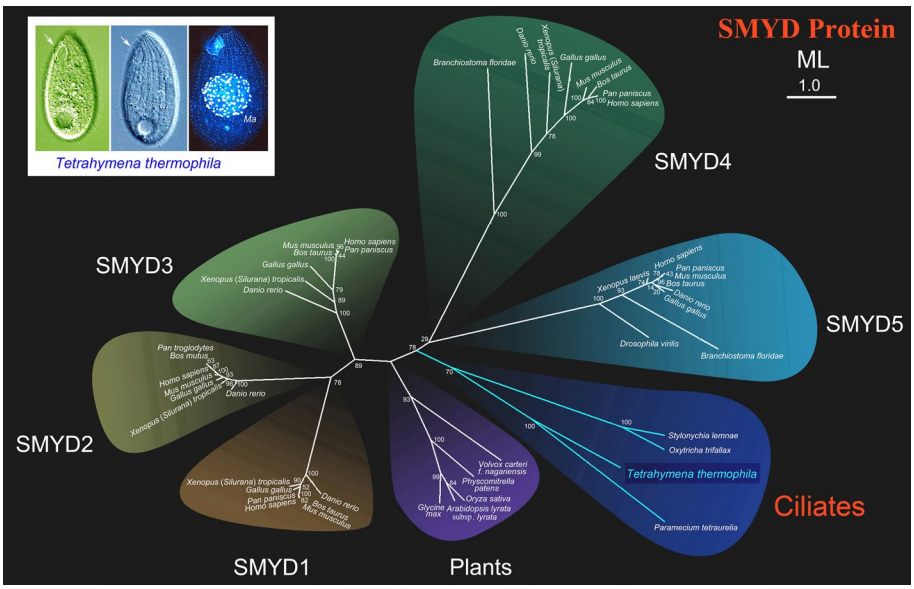

Fig. 1 Maximum likelihood (ML) tree of eukaryotes inferred from amino acid sequences of SMYD proteins. Blue branches denoted the phylogenetic positions of ciliates. Numbers given at nodes are ML bootstrap values from 1000 bootstrap replicates. The scale bar corresponds to 1 substitutions per 100 amino acid positions

sequence differences between SMYD4 and SMYD5 partially contributed to the low branch supporting value (29% ML).

In contrast to mammals, only one SMYD homologue is present in the genomes of many ciliates and plants (Spellmon et al. 2015). Ciliate SMYD and plant SMYD formed two well-separated clades (70% and 93% ML, respectively), indicating their diverged evolutionary paths from mammalian homologues. The clustering of ciliate SMYD also reflected the phylogenetic relationship of the constituent species which are divided between two fully supported sister branches: the oligohymenophoreans *Paramecium tetraurelia* and *Tetrahymena thermophila* and the hypotrichs *Stylonychia lemnae* and *Oxytricha trifallax* (Gao et al. 2016).

Localization and expression of SMYD protein in *Tetrahymena thermophila*

The expression levels of *SMYD1* during the life cycle of *Tetrahymena thermophila* were evaluated by RT-PCR (Fig. 2a). As previously reported (Miao et al. 2009), *SMYD1* is expressed during vegetative, starvation, and conjugation stages; in the present study, *SMYD1* expression was highest during conjugation (Fig. 2a); however, we mainly focus on SMYD1's function at the vegetative stage. To detect its cellular localization, we initially generated a somatic SMYD1-CHA strain by introducing a short sequence encoding the hemagglutinin (HA) tag to the C-terminus of the *SMYD1* gene, but we failed to detect any immunofluorescence signal, probably due to the low expression level of *SMYD1* during the vegetative stage (Fig. 2a). To facilitate tracking SMYD1's distribution, an *SMYD1* overexpression mutant was generated (SMYD1-CHA-overexpression) by placing *SMYD1* and the C-terminal HA tag coding sequence under the cadmium-inducible *MTT1* promoter (Fig. 2b). During the vegetative stage, SMYD1 was localized in both the cytoplasm and the macronucleus (Fig. 2c). As the confocal microscope scanned different layers of the *Tetrahymena* cell, SMYD1 showed different localization patterns: in layer 1, signals were mainly in the cytoplasm, and those in the macronucleus were weak; in layer 2, signals in the macronucleus became as strong as those in the cytoplasm; in layer 3, SMYD1 mainly appeared in the macronucleus and on the cell membrane. To remove the interference of SMYD1 signals from the cytoplasm and to reveal the presence of nuclear SMYD1, immunofluorescence staining was carried out on macronuclei purified from SMYD1-CHA-overexpression cells. Strong SMYD1 signals were detected throughout the transcriptionally active macronucleus and on the periphery of the transcriptionally inactive micronucleus (Fig. 2d), corroborating the localization of SMYD1 in the nucleus. The localization of SMYD1 in both the cytoplasm and the nucleus suggests that SMYD1 has the potential to methylate both histone and non-histone substrates (Al-Shar'i and Alnabulsi 2016; Calpena et al. 2015; Tracy et al. 2018).

To trace the dynamic change of SMYD1 protein, we performed a pulse-chase experiment by transiently inducing the expression of *SMYD1* with $CdCl_2$ (Fig. 2e). SMYD1 was initially detected in the cytoplasm immediately after the cadmium induction (0 h), suggesting that SMYD1 was present in the cytoplasm. Signals in the cytoplasm decayed, while the macronuclear signals increased 1 h later (1 h), demonstrating that SMYD1 can shuttle between the cytoplasm and the macronucleus.

Phenotypic analysis of Δ*SMYD1* cells

To reveal SMYD1's functions in *Tetrahymena*, we generated *SMYD1* knockout cells (Δ*SMYD1*) and characterized the phenotype. To investigate the roles of SMYD1 in regulating gene transcription, we carried out RNA-seq analysis in wild type (CU428) and the isogenic Δ*SMYD1* cells. In total, 21,461 well-annotated genes were included, among which 4841 (23%) were up-regulated (> 2-fold) and 1507 (7%) were down-regulated (< 0.5-fold). KEGG pathway analysis (Table 1) revealed that the most affected pathways in Δ*SMYD1* cells were a metabolic pathway (ko01100) and biosynthesis of secondary metabolites (ko01110), supporting the assertion that SMYD1 plays important roles in regulating metabolic genes in *Tetrahymena*. The result of Gene Ontology (GO) term enrichment analysis (Fig. 3) was consistent with that of the KEGG pathway analysis; this revealed the single organism metabolic process as the most enriched pathway for the up-regulated genes in Δ*SMYD1* cells. These results suggest that SMYD1 is involved in regulation of metabolism.

To examine if there was DNA replication deficiency in Δ*SMYD1* cells, immunofluorescence staining was performed for two indicative markers in the DNA damage response (DDR) system—γH2A.X (phosphorylation form of H2A.X, indicator of double strand DNA breakage) and RPA1 (single strand DNA binding protein, indicator of single strand DNA accumulation) (Gao et al. 2013). There was an increase of γH2A.X levels in Δ*SMYD1* cells, similar to the phenotype of the DNA replication deficient strain Δ*TXR1* (Gao et al. 2013), indicating accumulation of double strand DNA breakage produced by abnormal DNA replication (Fig. 4a). RPA1 was slightly induced at mRNA (Fig. 4b) and protein (more RPA1 foci) (Fig. 4c, white arrows) levels, weaker than Δ*TXR1* cells but stronger than WT cells. Moreover, genes significantly induced in Δ*TXR1* cells were mostly up-regulated (though to a more moderate degree) in Δ*SMYD1* cells, including many key players in ssDNA sensing/binding, DNA alkylation repair, nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), and non-homologous end joining (NHEJ) (Fig. 4b). Taken

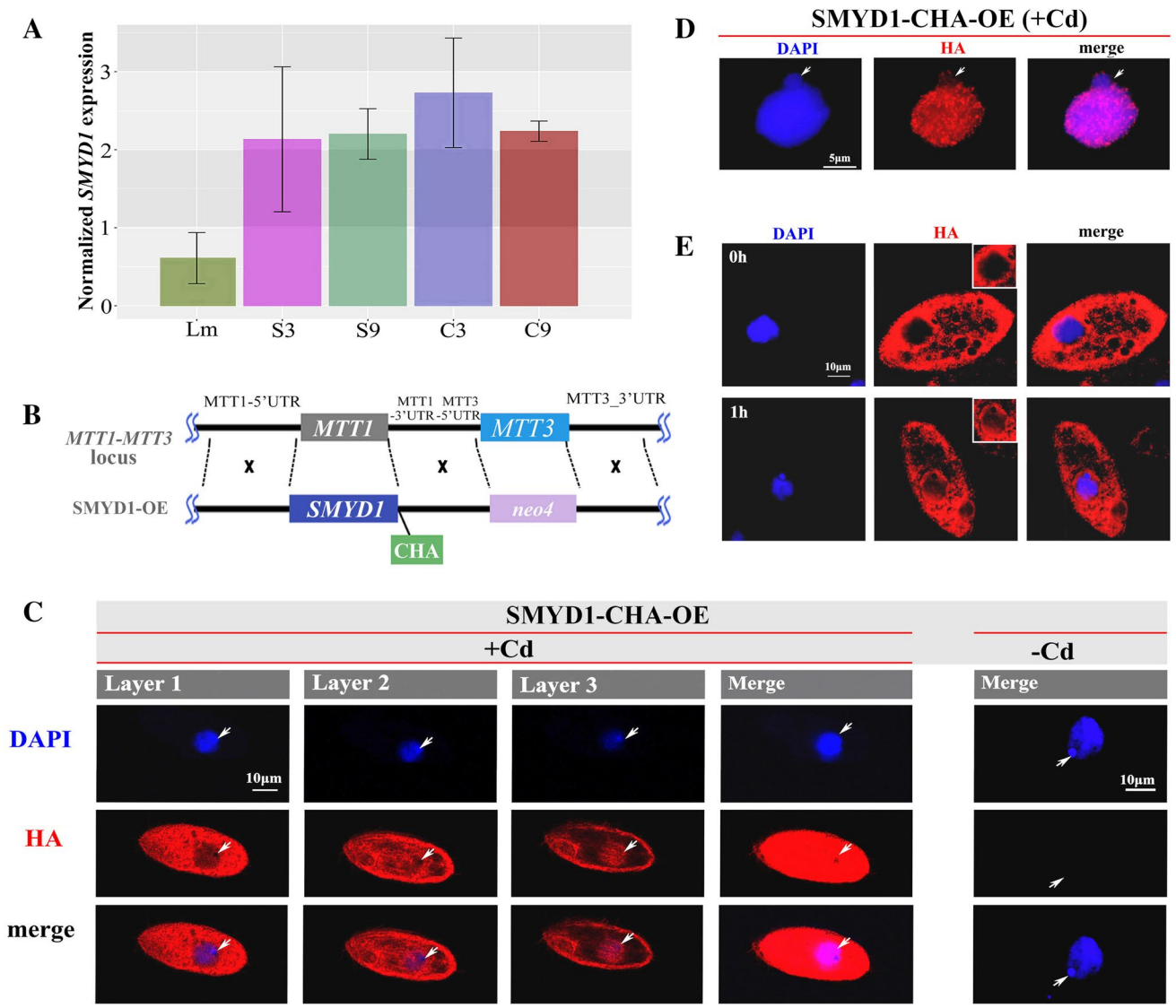

Fig. 2 Localization and expression of SMYD1 protein. **a** Gene expression profile for *SMYD1* in wide-type cell. The different time points are shown along the *X*-axis. Lm stands for growing cells at the density of ~ 3.5×10^5 cells/ml. S3, S9 stands for starved cells (~ 2×10^5 cells/ml) collected at 3 h and 9 h after starvation, respectively. C3, C9 stands for conjugative cells (equal volumes of CU427 and CU428) collected at 3 h and 9 h after mixing, respectively. The normalized gene expression levels, as previously described (Cheng et al. 2016), are shown along the *Y*-axis. **b** Schematic diagram of SMYD1-CHA-overexpression (SMYD1-CHA-OE) plasmid. **c** SMYD1 signals in different layers of the same SMYD1-CHA-OE cell (vegetative stage) scanned by confocal microscopy. Arrows represent the micronucleus. Layers 1–3 indicate three different layers of the same cell. **d** SMYD1 signals within the nucleus obtained by nucleus purification. Arrows show the micronucleus. **e** pulse-labeling of SMYD1-CHA-OE (vegetative stage) cells. 0 h and 1 h stand for SMYD1 signals of cells collected at 0 h and 1 h after the removal of $CdCl_2$, respectively. DAPI, nuclear signals; HA, signals of target protein; merge, combined signals of target protein and nucleus

together, these results argue that SMYD1 deletion resulted in mild replication stress.

It should be noted that no obvious growth defect was observed for Δ*SMYD1* cells, since doubling time of Δ*SMYD1* strain (Fig. 4d) showed little difference from that of WT cells (4.4 vs. 4.7 h). This suggested that SMYD1 was not essential for vegetative growth, or SMYD1 deficiency could be successfully coped with by activating the DNA damage responses (Fig. 4b).

Interactome of SMYD1 protein

To further study the mechanism underlying SMYD1 functions, immunoprecipitation (IP) of SMYD1-CHA-overexpression cells was carried out to assess the interactome of SMYD1 protein (Fig. 5a). Mass spectrometry analysis (Supplementary Table 2) revealed the heat shock protein 90 (HSP90: TTHERM_00080030) as a potential interacting partner with SMYD1, which is consistent with previous

Table 1 KEGG pathway analysis for differentially expressed genes in Δ*SMYD1* strain

Pathways associated with up-regulated genes	
ko01100	Metabolic pathways (98)
ko01110	Biosynthesis of secondary metabolites (52)
ko01130	Biosynthesis of antibiotics (30)
ko01120	Microbial metabolism in diverse environments (25)
ko01200	Carbon metabolism (20)
ko01230	Biosynthesis of amino acids (17)
ko00970	Aminoacyl-tRNA biosynthesis (16)
ko00230	Purine metabolism (13)
ko00010	Glycolysis/gluconeogenesis (12)
ko04141	Protein processing in endoplasmic reticulum (12)
Pathways associated with down-regulated genes	
ko03008	Ribosome biogenesis in eukaryotes (17)
ko01100	Metabolic pathways (16)
ko04142	Lysosome (7)
ko01110	Biosynthesis of secondary metabolites (5)
ko01130	Biosynthesis of antibiotics (4)
ko00270	Cysteine and methionine metabolism (4)
ko00230	Purine metabolism (3)
ko04141	Protein processing in endoplasmic reticulum (3)
ko01120	Microbial metabolism in diverse environments (3)
ko04113	Meiosis–yeast (3)

Numbers within () correspond to genes mapped onto this pathway

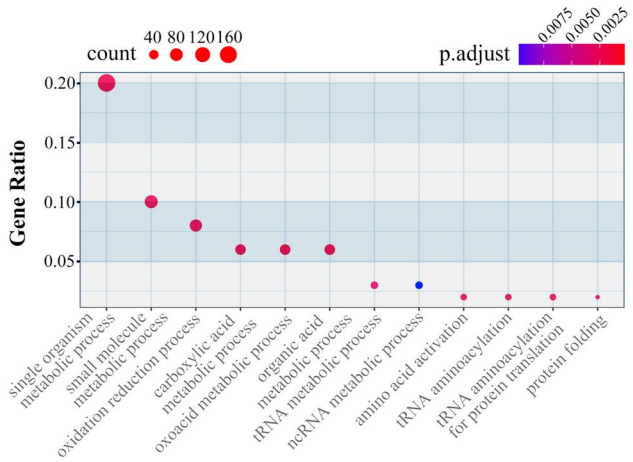

Fig. 3 GO (GO term: biological process) analysis of up-regulated genes in Δ*SMYD1*. Different biological processes are shown along the *X*-axis. *Y*-axis is the gene ratio. The higher the values, the closer the relationship between the SMYD1 protein and the target process

studies on mammalian systems (Donlin et al. 2012; Sima and Richter 2018; Spellmon et al. 2017). This mass spectrometry result was corroborated by the silver staining of the IP sample (Fig. 5a), which revealed a ~98 kDa band corresponding to the predicted size of HSP90 (blue arrow in Fig. 5a), as well as the bait protein (SMYD1) band (red arrow in Fig. 5a, ~53 kDa).

To further confirm the interaction between HSP90 and SMYD1, gel filtration chromatography was used to separate proteins in the IP sample of SMYD1-CHA-overexpression cells (Fig. 5b). As a control, the bait protein band (SMYD1) was enriched in fractions 29–32 (calculated size is ~53 kDa). It was also enriched in fractions 24 and 25 (calculated size ~150 kDa), corresponding to the predicted size of the SMYD1-HSP90 complex. This result raised the possibility of direct interaction between SMYD1 and HSP90. To support this, we introduced a flag tag to the N-terminus of *HSP90* in SMYD1-CHA-overexpression cells (HSP90_Nflag/SMYD1-CHA-OE) and carried out the co-precipitation of SMYD1 and HSP90 (Fig. 5c). By performing immunoprecipitation with anti-Flag M2 beads, we detected SMYD1 in the IP sample by the anti-HA antibody at the expected size (~53 kDa) (Fig. 5c lane 2). As negative controls, no SMYD1 band was detected without cadmium induction (Fig. 5c, lanes 3 and 4) or in IP samples from cells without flag-tagged HSP90 (Fig. 5c, lanes 5 and 6). These results support the interaction between SMYD1 and HSP90.

To investigate whether SMYD1 can methylate HSP90, SMYD1 in vitro methyltransferase assay was performed. We only detected a ~53 kDa radioactive band (Supplementary Fig. 1, red arrow), representing either the self-methylated SMYD1 or SMYD1 binding of H^3-labeled S-adenosylmethionine (SAM). Though we failed to detect any radioactive band of HSP90, the ~53 kDa radioactive band provides

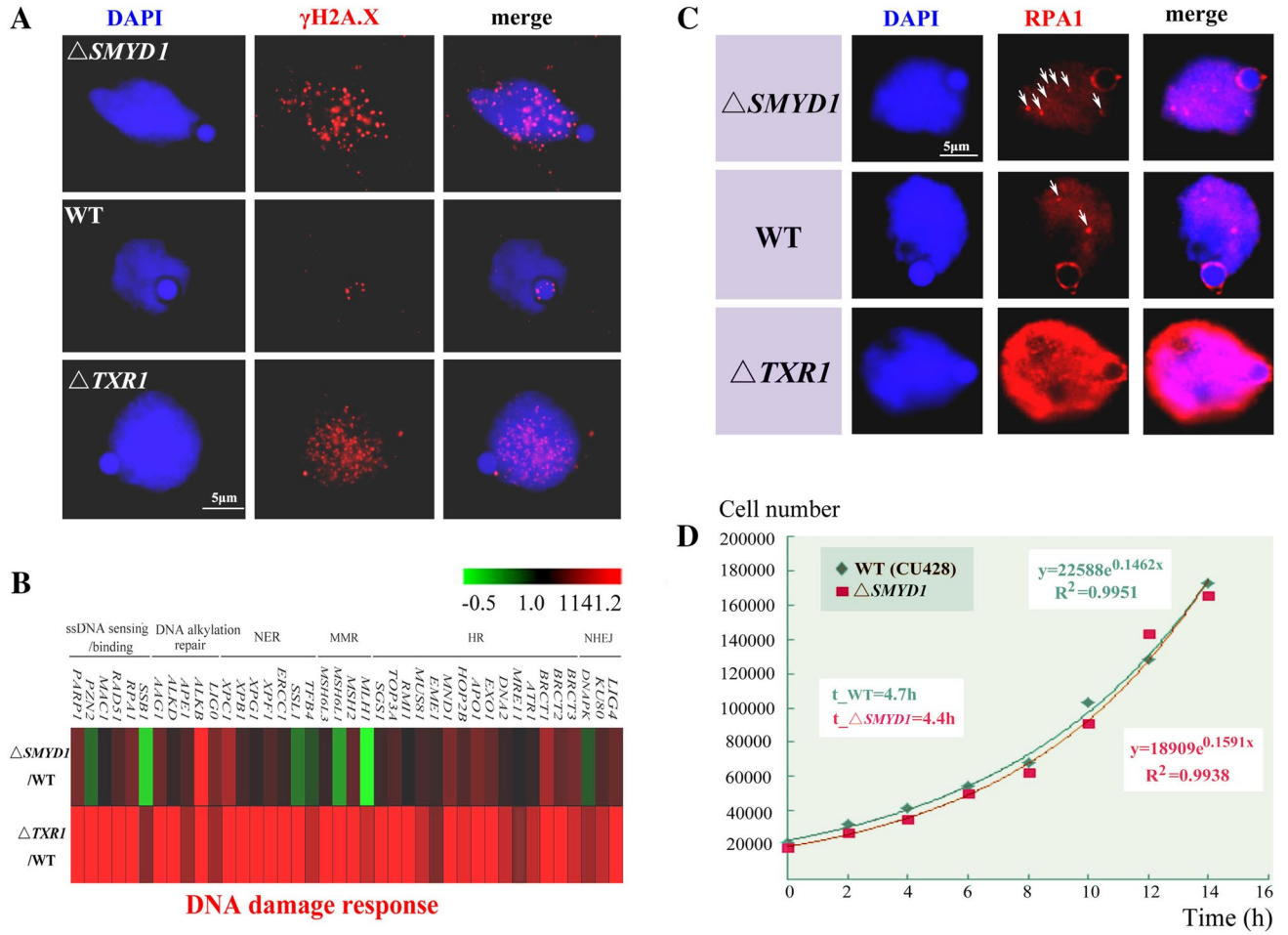

Fig. 4 Phenotype of Δ*SMYD1* cells. **a** Accumulation of double strand DNA breakage (DSBs) in Δ*SMYD1* cells. Δ*TXR1*, positive control; WT, negative control, wide type; γH2A.X., indicator of double strand DNA breakage, the phosphorylation form of the histone variant H2A.X. **b** Heat map of relative gene expression for DNA damage response-related genes in Δ*SMYD1* cells, normalized against WT cells. Δ*TXR1*/WT, positive control, data from (Gao et al. 2013); NER, Nucleotide excision repair; MMR, mismatch repair; HR, homologous recombination; NHEJ, non-homologous end joining. Up-regulated genes are the reddest (Δ*SMYD1*/WT > 2) down-regulated genes are the greenest (Δ*SMYD1*/WT < 0.5), constant genes are represented by color between them. **c** Accumulation of single strand DNA in Δ*SMYD1* strain. White arrows show RPA1 foci, indicating the accumulation of single strand DNA in the Δ*SMYD1* strain. Δ*TXR1*, positive control; WT, negative control, wide type; RPA1, indicator of single strand DNA, single strand DNA binding protein; DAPI, nuclear signals; merge, combined signals of the target protein and nucleus. **d** Growth curve of Δ*SMYD1* and WT strains. *X*-axis, different time points; *Y*-axis, cell number

evidence that SMYD1 protein possesses the methyltransferase activity.

Discussion

SMYD homologues in ciliates

The phylogenetic position of ciliate SMYD1 was revealed in this study for the first time, on the basis of phylogenetic analysis in metazoans (Calpena et al. 2015; Jiang et al. 2017). Ciliate SMYD1 occupied a basal position in the phylogenetic tree, which is consistent with the early branching position of ciliates in the evolutionary history of eukaryotes (Adl et al. 2012; Sheng et al. 2018; Zhang et al. 2018; Zhao et al. 2018; Zheng et al. 2018), The presence of a single homologue is likely an ancestral state, given that only one homologue was identified in ciliates, while several occur in animals.

Interestingly, SMYD homologues of the four ciliates included in this study showed relatively low similarity (in the range of 10.4%–41%). This high diversity of ciliate SMYD is consistent with the fact that ciliates underwent radiation after their early branching and harbored a rich pool of morphological and genetic diversities (Chen et al. 2014, 2018, 2019; Guerin et al. 2017; Hamilton et al. 2016;

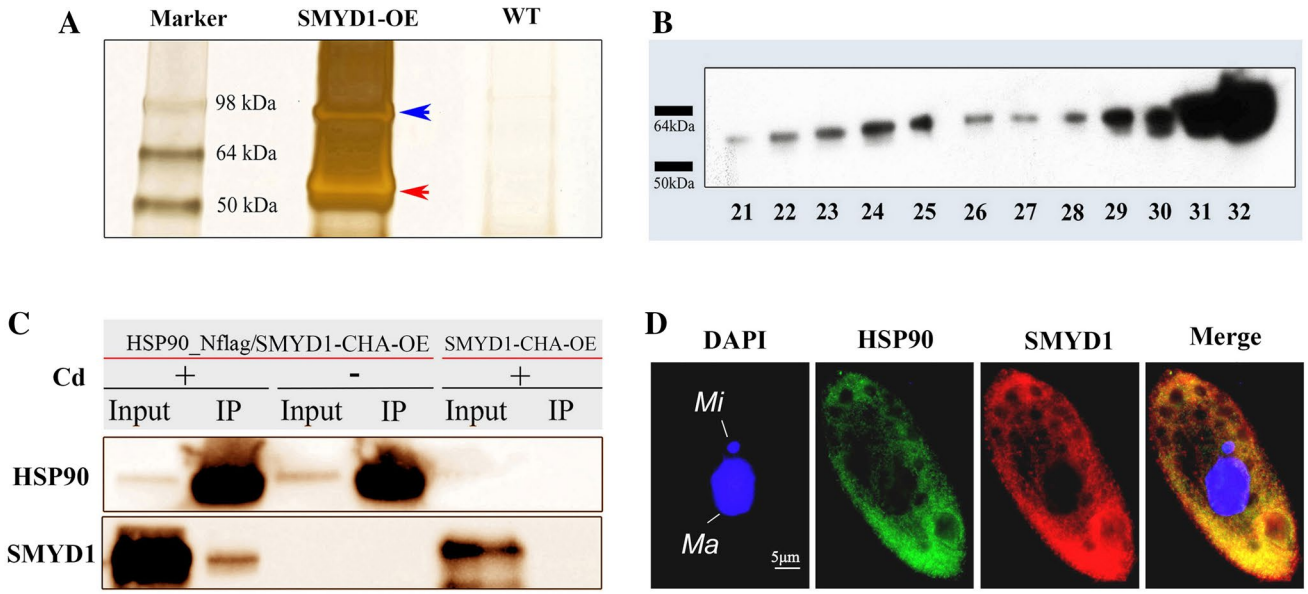

Fig. 5 Interactome analysis of SMYD1 protein. **a** Silver staining of the SMYD1-CHA-OE IP sample. Red arrow indicates SMYD1 protein (53 kDa); blue arrow indicates the potential HSP90 protein. WT, wild type (negative control). **b** Purification of proteins interacting with SMYD1 using gel filtration chromatography. IP sample (without TCA precipitation) of SMYD1-CHA-OE strain was used as input. 21–32, different sized protein complex in different collection tubes: 24–25, protein complex of ~150 kDa; 29–32, bait protein of ~53 kDa (SMYD1). **c** Western blot of IP samples for HSP90-Nflag/SMYD1-CHA-overexpression strains with or without Cd^{2+} induction. SMYD1-CHA-overexpression strain with Cd^{2+} induction was used as negative control. **d** Co-localization of HSP90 and SMYD1 in Cd^{2+}-induced HSP90-Nflag/SMYD1-CHA-overexpression strain. DAPI, nuclear signals; HSP90, HSP90 protein signals, detected by the α-Flag antibody; SMYD1, SMYD1 protein signals, detected by the α-HA antibody; Merge, combined signals of target proteins and nucleus

Huang et al. 2018; Luo et al. 2018; Noto and Mochizuki 2017; Wang et al. 2019; Xu et al. 2019a, b; Yan et al. 2018).

SMYD is involved in the regulation of metabolism-related genes and DNA replication

SMYD proteins in mammals are located in both the cytoplasm and the nucleus, and correspondingly have the ability to methylate both histone and non-histone targets (Brown et al. 2006; Gottlieb et al. 2002; Hamamoto et al. 2004; Huang et al. 2006; Mazur et al. 2014; Tracy et al. 2018; Yi et al. 2017). SMYD1 was also detected in both the cytoplasm and the macronucleus (and possibly micronucleus) in *Tetrahymena* cells, which is consistent with its localization in mammalian systems.

The SMYD1 localization in the transcriptionally active macronucleus is consistent with its potential role in regulating gene expression. Extensive studies have pointed out that mammalian SMYD proteins can affect gene accessibility by histone methylation and interaction with transcription factors (Abu-Farha et al. 2008; Chen et al. 2017; Gottlieb et al. 2002). In the current study, we revealed that ciliate SMYD1 is involved in the regulation of metabolism-related genes and consequently plays roles in *Tetrahymena* metabolism. Therefore, we proposed that functions of mammalian SMYD in cancer development (Giakountis et al. 2017; Hu et al. 2009; Leinhart and Brown 2011; Mazur et al. 2016) might have evolved from the function of SMYD1 in regulating metabolism-related genes in ciliates. However, the underlying mechanism in ciliates is yet to be explored.

The regulation of DNA replication by SMYD has not been reported before. In this study, we demonstrate that lack of SMYD1 in *Tetrahymena* causes mild DNA replication stress, manifested by the accumulation of DNA double strand breaks (DSBs) and single strand DNA (ssDNA), and activation of DNA damage response.

HSP90 is a conserved SMYD substrate

HSP90 is an essential chaperone protein involved in a variety of biological processes, including stabilizing proteins against heat stress, stabilizing a quantity of tumor proteins, and enhancing the loading process of small RNAs into Argonaute proteins (Bachman et al. 2018; Karras et al. 2017; Taipale et al. 2010; Woehrer et al. 2015). In mammalian cells, SMYD2 was shown to regulate biological functions of HSP90 by methylating its different domains (Abu-Farha et al. 2011). Our data show that *Tetrahymena* SMYD1 and HSP90 are both localized in the cytoplasm. More importantly, *Tetrahymena* SMYD1 can physically interact with HSP90 and

may have the capability to catalyze methylation. Thus, we propose that HSP90 is one of the conserved substrates for SMYD methyltransferases. More studies are needed to explore the functions of SMYD in regulating HSP90.

In conclusion, our study represents the first report of the functions of methyltransferase SMYD in the single-cell model organism *Tetrahymena thermophila*. We revealed the localization and dynamics of SMYD1 in *Tetrahymena* cytoplasm and nucleus, and demonstrated the roles of *Tetrahymena* SMYD1 in DNA replication and transcription regulation. Additionally, we show that accumulating evidence supports the possibility that HSP90 is a conserved SMYD substrate. These findings support a conserved function in ciliate SMYD and shed light on the mechanisms that underlie the roles that SMYD family proteins play in the development of cancer in higher eukaryotes.

Materials and methods

Phylogenetic analysis of SMYD

A total of 47 SMYD amino acid sequences (Table 2) of representative eukaryotic species were downloaded from the National Center for Biotechnology Information (NCBI) Database (https://www.ncbi.nlm.nih.gov/). Sequences were aligned by MUSCLE 3.7 (Edgar 2004), provided on the web server "Phylogeny.fr Robust Phylogenetic Analysis For The Non-Specialist" (http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=muscle). The alignment was used for the subsequent phylogenetic tree construction.

A Maximum-Likelihood (ML) tree was constructed with RAxML-HPC2 on XSEDE v 7.2.8 (Stamatakis 2006; Stamatakis et al. 2008), provided by the CIPRES Science Gateway (Miller et al. 2010), using plant species, *Volvox carteri* f. *nagariensis*, *Physcomitrella patens*, *Oryza sativa*, *Arabidopsis lyrata* subsp. *lyrata* and *Glycine max*, as the outgroup. The MTART model of Protein Substitution Matrix selected by ProtTest 3 (Darriba et al. 2011) and other defaulted parameters were used for the maximum-likelihood (ML) analysis. The robustness of internal branches was estimated by 1000 bootstrap replicates.

Strains and culture conditions

The wide-type *Tetrahymena thermophila* strain CU428 (provided by *Tetrahymena* Stock Center, Cornell University,

Table 2 Accession numbers of species used in the phylogenetic tree

Species name and protein name	GenBank accession no.	Species name and protein name	GenBank accession no.
Danio rerio SMYD1	NP_001034725	*Gallus gallus* SMYD4	NP_001025886
Xenopus (*Silurana*) *tropicalis* SMYD1	XP_012811600	*Mus musculus* SMYD4	AAH95952
Gallus gallus SMYD1	NP_989486	*Bos taurus* SMYD4	XP_005220153
Mus musculus SMYD1	NP_001153599	*Pan paniscus* SMYD4	XP_003816904
Bos taurus SMYD1	DAA24603	*Homo sapiens* SMYD4	NP_443160
Pan paniscus SMYD1	XP_003805903	*Drosophila virilis* SMYD5	XP_002053397
Homo sapiens SMYD1	NP_938015	*Branchiostoma floridae* SMYD5	XP_002609030
Danio rerio SMYD2	NP_001013568	*Xenopus laevis* SMYD5	NP_001085635
Xenopus (*Silurana*) *tropicalis* SMYD2	XP_00293475NP_9380151	*Danio rerio* SMYD5	NP_001004614
Gallus gallus SMYD2	NP_001264500	*Gallus gallus* SMYD5	NP_001012912
Mus musculus SMYD2	EDL13024	*Bos taurus* SMYD5	NP_001073717
Bos mutus SMYD2	NP_001069832	*Mus musculus* SMYD5	NP_659167
Pan troglodytes SMYD2	XP_003308794	*Pan paniscus* SMYD5	XP_003808242
Homo sapiens SMYD2	NP_064582	*Homo sapiens* SMYD5	NP_006053
Danio rerio SMYD3	NP_001032477	*Stylonychia lemnae*	CDW88943
Xenopus (*Silurana*) *tropicalis* SMYD3	XP_004914684	*Oxytricha trifallax*	EJY75465
Gallus gallus SMYD3	XP_015139481	*Paramecium tetraurelia*	XP_001437474
Bos mutus SMYD3	XP_005216902	*Tetrahymena thermophila*	XP_001022867
Mus musculus SMYD3	NP_081464	*Volvox carteri* f. *nagariensis*	XP_002956692
Pan paniscus SMYD3	XP_514316	*Physcomitrella patens*	XP_001778213
Homo sapiens SMYD3	NP_001161212	*Oryza sativa*	AAS07242
Branchiostoma floridae SMYD4	XP_002589088	*Arabidopsis lyrata* subsp. *lyrata*	XP_002886136
Danio rerio SMYD4	NP_001070062	*Glycine max*	XP_006578981
Xenopus (*Silurana*) *tropicalis* SMYD4	XP_012812692		

Ithaca, NY), from which all mutant strains were derived, was cultured in SPP medium (Orias et al. 1999) at 30 °C. Cells in mid-exponential growth phase (~ 2×10^5 cells/ml) were used for subsequent experiments.

Generation of the transgenic strains

In the current study, 6 constructs, including Δ*SMYD1*, SMYD1-CHA, Δ*SMYD1*-RPA1-CHA, SMYD1-CHA-overexpression, and HSP90-Nflag were generated, and all primers used here are listed in Table 3. Construct RPA1-CHA was generated as previously described (Gao et al. 2013). Constructs Δ*SMYD1* and SMYD1-CHA were generated according to previous studies (Feng et al. 2017; Liu et al. 2007; Noto et al. 2015). For the SMYD1-CHA-overexpression construct, the target fragments were cloned into the newly constructed CHA-overexpression vectors. CHA-overexpression vectors were generated based on the inducible MTT1 and MTT3 promotors (Cd^{2+} inducible) to investigate proteins of low expression level. Primers used are shown in Supplementary Table 1. The MTT1–MTT3 region, including the 3′ and 5′ untranslated regions (5.4 kb in total), were amplified with Platinum Tag DNA Polymerase (Invitrogen, 11304-011) and cloned into pBlueScript SK (−) vector. The MTT3 locus was replaced with *neo4* coding region, providing the paromomycin resistance for cells. The MTT1 locus was replaced by Flag-HA tag with Sbf I cutting site on its N terminus.

All of the above constructs except RPA1-CHA and HSP90-Nflag were introduced into CU428 by standard biolistic transformations (Cassidy-Hanley et al. 1997). RPA1-CHA was introduced into Δ*SMYD1* strain, and HSP90-Nflag into SMYD1-CHA-overexpression strain, respectively. Paromomycin, Cycloheximide or Blasticidin was used for subsequent transformant selection according to the drug cassette. Complete somatic replacement was validated by quantitative-PCR as previously reported (Zhao et al. 2017).

Macronucleus purification

SMYD1-CHA-overexpression stain was cultured overnight (~ 18 h) in 1L 1 × SPP containing 0.5 μg/ml $CdCl_2$. Mid-log-phase (~ 2×10^5 cells/ml) cells were collected and the macronucleus purification was carried out as described (Chen et al. 2016). The purified macronuclei were washed with nuclear wash buffer (50 mmol/L pH 7.4 Tris, 2 mmol/L $MgCl_2$) once and resuspended in 200 μl nuclear wash buffer for subsequent immunofluorescence staining.

Table 3 Primers used for plasmid construction and RT-PCR

Name	Sequence
SMYD1_5f715_NotI	*AGTTCTAGAGCGGCCGC*GATTATTCGCCTATAGTTGATGG
SMYD1_3r5096_Not I	*ACCGCGGTGGCGGCCGC*GTTTCCTACTCAGTCTCTTGC
SMYD1_Nf1849_GHA	*CCCTACGACGTCCCCGACTACGCC*TACTAGGTAGAAAATCTTAATAATTAGTATCG
SMYD1_Nr1848_GHA	*GTCGGGGACGTCGTAGGGGTATCC*CATATATATCTTTGATTTTCTAAATTAATTG
SMYD1_Cf3335_GHA	*CCCTACGACGTCCCCGACTACGCC*TGATATTCTTTAAAATAAAATAAAAAAAAG
SMYD1_Cr3334_GHA	*GTCGGGGACGTCGTAGGGGTATCC*ATTATATTTCATTTTTATCTCAC
SMYD1_3f3956_neo4	*CTGACGTCGCACCATCCC*GTTGTTGCATAGGATTGTTTTCG
SMYD1_3r3928_neo4	*GTCAGGTGCCTGGTACCC*GTTTATTTAAAAAGCAGTAGC
SMYD1_f2734	CTTTTTGTGTGAATGTAAAAGGTG
SMYD1_r2862	CCTTAAGTGTTACAGTCAGAGC
α-Tubulin_f	TCAGTAACCTTCTTCTTCACC
α-Tubulin_r	CACTGGTTTCAAGGTCGGTAT
MTT1_SMYD1_f1724_NHA	*CGTCCCCGACTACGCC*TACTAGGTAGAAAATCTTAATAATTAG
MTT1_SMYD1_r3241_NHA	*CATATTTATTTCACCT*ATTATATTTCATTTTTATCTCACTTTTTATATC
MTT1_SMYD1_f1698_CHA	*CTTAAAATAATGGATCCT*TACTAGGTAGAAAATCTTAATAATTAG
MTT1_SMYD1_r3217_CHA	*CGTCGTAGGGGTATCC*ATTATATTTCATTTTTATCTCACTTTTTATATC
HSP90_5f1223_Not I	*AGTTCTAGAGCGGCCGC*ATCAAAGTATGAAGAAGACAGG
HSP90_3r6007_Not I	*ACCGCGGTGGCGGCCGC*TAATCAAATAAATCTCTCTGTTCTG
HSP90_Nf2158_Flag	*GGAGACTACAAGGACGACGATGACAAG*TCTCAACAAGCTGAACACTTTGC
HSP90_Nr2157_Flag	*GTCATCGTCGTCCTTGTAGTCTCC*CATTTCTTATGATATATCTTTTTTTTTAAT
HSP90_3f5043_BSR	*CTGACGTCGCACCATCCC*GTGAAGTTTTTTGATATTATCACAC
HSP90_3r5004_BSR	*GTCAGGTGCCTGGTACCC*ACTTTTATATCAGTGAAAATGGAG

Italic characters represent the adaptors

Immunofluorescence staining

A volume of 15 ml of target cells in mid-exponential growth phase was collected and fixed in 2% paraformaldehyde (diluted with 1 × PBS). Permeabilization was then accomplished with 0.4% Triton X-100 (diluted with 1 × PBS), after which antibodies (details in Table 4) were incubated with cells (Gao et al. 2013; Liu et al. 2007). Digital images were captured using an Olympus BX43 microscope and an Olympus DP73 camera.

Immunoprecipitation and quantitative liquid chromatography–mass spectrometry (LC–MC) analysis

SMYD1-CHA-overexpression cells were cultured overnight in 800 ml 1 × SPP containing 0.5 μg/ml $CdCl_2$. Cells at mid-log phase were collected by centrifugation and immunoprecipitation (IP, details shown in Supplementary Methods) was carried out, after which the IP sample was sent to Proteomics Resource Facility (Department of Pathology, University of Michigan) for LC–MC analysis.

The HSP90-Nflag/SMYD1-CHA-overexpression strain was cultured overnight in 800 ml 1 × SPP with or without 0.5 μg/ml $CdCl_2$. At the same time, the SMYD1-CHA-overexpression strain was cultured overnight in 800 ml 1 × SPP as a negative control. IP was carried out on these strains as described above. IP samples were used for western blotting with primary antibodies α-Flag (Mouse monoclonal, Sigma, F1804, 1:5000) and α-HA (Rabbit monoclonal, Cell Signaling, 3724S, 1:2000).

In vitro methyltransferase assay

An IP sample (without TCA precipitation) of SMYD1-CHA-overexpression strain was used for the in vitro methyltransferase activity test. Immunoprecipitation was modified from the protocol outlined in the Supplementary Methods, in which T0 buffer (30 mmol/L Tris HCl, 30 mmol/L Tris Base, 20 mmol/L KCl and 2 mmol/L $MgCl_2$) was used instead of T150 buffer (30 mmol/L Tris HCl, 30 mmol/L Tris Base, 20 mmol/L KCl, 2 mmol/L $MgCl_2$ and 150 mmol/L NaCl) and HA elution (eluted with 250 μg/ml HA peptide at RT for 15 min) was used as a final product for the subsequent methyltransferase activity test.

The in vitro methyltransferase activity test was modified from previous studies (Wu et al. 2013; Zhou et al. 2016). For each test, H^3-labeled S-adenosylmethionine (SAM, final concentration 10 mmol/L, from Perkin Elmer, NET155H250UC) was added to 20 μl IP sample (without TCA precipitation, in T0 buffer) or 20 μl T0 buffer (negative control). The reaction was carried out at 25 °C overnight, after which the methylation was detected by autoradiography.

Total RNA extraction and RT-PCR

A volume of 10 ml of cells was collected at indicated time points. TRIzol™ Reagent (Invitrogen, 15596026) was used to extract the total RNA, after which DNA was removed with the Turbo DNA-free kit (Ambion, AM1907). Complementary DNA (cDNA) was reverse-transcribed using Superscript III Reverse Transcriptase kit (Invitrogen, 18080-051) with parameters as follows: 50 °C for 50 min, 85 °C for 5 min, and maintained at 4 °C.

RT-PCR, with cDNA as template, was carried out on a CFX96™ Real-Time System (BIO-RAD, USA) with the Radiant™ Green Lo-Rox qPCR Kit (Alkali Scientific, QS1020). The reaction was carried out as previously described (Gao et al. 2013). Each reaction was performed in duplicate using primers SMYD1_f2734 and SMYD1_r2862 (Table 3). α-Tubulin_f and α-Tubulin_r (Table 3) were used as internal controls (Cheng et al. 2016). The $2^{-\Delta\Delta Ct}$ method was used to analyze the real-time PCR data.

RNA extraction, library preparation, Illumina sequencing and data analysis

Total RNA of log-phase Δ*SMYD1* and wild-type CU428 cells (negative control) were extracted using Qiashredder (Qiagen, 79654) and RNeasy Kit (Qiagen, 74624) according to the protocol provided in TetraFGD (Xiong et al. 2013). The Qubit RNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, CA, USA, Q32852) and the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent

Table 4 Antibodies for immunofluorescence staining

	Antibody	dilution	Incubation condition
Primary antibody	α-HA (Rabbit monoclonal, Cell Signaling, 3724S)	1:2000	4 °C, overnight
	α-γH2A.X(Mouse monoclonal, Millipore, 05636)	1:5000	RT, 2 h
	α-Flag (Mouse monoclonal, Sigma, F1804)	1:5000	4 °C, overnight
Secondary antibody	Goat Anti Mouse 555 IgG (Invitrogen, A32727)	1:5000	RT, 1 h
	Goat Anti Rabbit 555 IgG (Invitrogen, A27017)	1:5000	RT, 1 h
	Goat Anti Rabbit 488 IgG (Invitrogen, A32731)	1:5000	RT, 1 h

Technologies, CA, USA, 5067-1511) were used to measure the RNA concentration and integrity, respectively.

In total, 3 μg RNA was used for sample preparation. The sequencing library was produced by NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, MA, USA, E7530L) according to manufacturer's recommendations, and details are as previously described (Zhang et al. 2015). The library was sequenced on the Illumina Hiseq 2500 platform provided by Novogene Bioinformatics Institute (Beijing, China), and 125 bp paired-end reads were generated.

FASTX-Toolkit (Gordon and Hannon 2010) was used to remove adapters and reads of low quality from the raw data, after which the remaining reads were mapped onto the *Tetrahymena thermophila* genome (http://ciliate.org/index.php/home/downloads, June 2014). Gene expression levels were calculated by RSEM v1.2.7 (Li and Dewey 2011). A heat map of genes differentially expressed between Δ*SMYD1* and CU428 was generated with MultiExperiment Viewer (v4.9) (Mar et al. 2011). The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis was carried out using the KAAS server (http://www.genome.jp/kegg/kaas/). Gene Ontology (GO) term enrichment analysis on the up-regulated genes was performed using BiNGO v3.0.3 (p.adjust < 0.05), which was integrated in Cytoscape v3.4.0, and the plot was generated by the R package, ggplot2 (Kohl et al. 2011; Maere et al. 2005; Wickham 2016).

Gel filtration chromatography

The IP sample (without TCA precipitation) of the SMYD1-CHA-overexpression strain was used as input of the purification. Purification was carried out by the AKTApurifier HPLC system (including Pump P-900 and Superose 6 column, GE Healthcare) according to previous studies (Hubert et al. 2014; Hughes et al. 2013). Firstly, filtered distilled water was pumped into the system to remove any contamination and bubbles, after which filtered T0 buffer was pumped into the system as loading buffer. Secondly, formula between the Retention Volume and Molecular Weight was calculated using two marker proteins, BSA (66 kDa) and WDR5 (36 kDa); that is Rv = (2.9661-logMw)/0.073, where Rv is Retention Volume (ml) and Mw is Molecular Weight (kDa). Thirdly, 300 μl IP samples of the SMYD1-CHA-overexpression strain were added into the system by 1 ml syringe and run at a flow rate of 0.5 ml/min to purify proteins of different sizes. Samples (10 μl) from each fraction (500 μl/tube) were retained for western blot analysis.

Pulse-chase experiment

The SMYD1-CHA-overexpression strain was cultured in 60 ml 1 × SPP at 30 °C overnight. The mid-log-phase cells were pulse-labeled by adding $CdCl_2$ (final concentration 1.5 μg/ml) into the medium. Two hours later, $CdCl_2$ was removed by washing with 1 × SPP twice. Cells were resuspended in 60 ml fresh 1 × SPP and collected at 1 h intervals during the chase.

Acknowledgements This work was supported by the Natural Science Foundation of Shandong Province (JQ201706 to SG), Fundamental Research Funds for the Central Universities (201841013 to SG), National Science Foundation [MCB 1411565 to YL], and National Institutes of Health Foundation [R01 GM087343 to YL]. XZ was supported by China Scholarship Council Scholarship for joint PhD students. Our thanks are given to Ms. Yuanyuan Wang (Laboratory of Protozoology, Ocean University of China), for helping with the preparation of Fig. 1.

Author contributions SG, YFL and XLZ participated in study design. XLZ carried out most of the experiments. YL conducted the establishment of HSP90-Nflag/SMYD1-CHA-overexpression strain and LLD prepared the RNA-seq samples. XC and FBM conducted the bioinformatics analysis. Manuscript writing was conducted by XLZ with assistance from SG, YFL, WBS and MJ. All authors have read and approved the final manuscript.

Data availability RNA-seq datasets have been deposited to NCBI with accession number GEO: GSE138246.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Animal and human rights statement This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abu-Farha M, Lambert JP, Al-Madhoun AS, Elisma F, Skerjanc IS, Figeys D (2008) The tale of two domains proteomics and genomics analysis of SMYD2, a new histone methyltransferase. Mol Cell Proteomics 7:560–572

Abu-Farha M, Lanouette S, Elisma F, Tremblay V, Butson J, Figeys D, Couture JF (2011) Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. J Mol Cell Bol 3:301–308

Adl SM, Simpson AG, Lane CE, Lukes J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J et al (2012) The revised classification of eukaryotes. J Eukaryot Microbiol 59:429–493

Allis CD, Caparros M, Jenuwein T, Reinberg D (2015) Epigenetics, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Al-Shar'i NA, Alnabulsi SM (2016) Explaining the autoinhibition of the SMYD enzyme family: a theoretical study. J Mol Graph Model 68:147–157

Arsenault PR, Song D, Chung YJ, Khurana TS, Lee FS (2016) The zinc finger of prolyl hydroxylase domain protein 2 is essential for efficient hydroxylation of hypoxia-inducible factor α. Mol Cell Biol 36:2328–2343

Bachman AB, Keramisanou D, Xu W, Beebe K, Moses MA, Kumar MV, Gray G, Noor RE, van der Vaart A, Neckers L (2018) Phosphorylation induced cochaperone unfolding promotes kinase recruitment and client class-specific Hsp90 phosphorylation. Nat Commun 9:265

Bagislar S, Sabò A, Kress TR, Doni M, Nicoli P, Campaner S, Amati B (2016) Smyd2 is a Myc-regulated gene critical for MLL-AF9 induced leukemogenesis. Oncotarget 7:66398–66415

Black JC, Van Rechem C, Whetstine JR (2012) Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol Cell 48:491–507

Blanc RS, Richard S (2017) Arginine methylation: the coming of age. Mol Cell 65:8–24

Boehm D, Jeng M, Camus G, Gramatica A, Schwarzer R, Johnson JR, Hull PA, Montano M, Sakane N, Pagans S (2017) SMYD2-mediated histone methylation contributes to HIV-1 latency. Cell Host Microbe 21:569–579

Brown MA, Sims RJ, Gottlieb PD, Tucker PW (2006) Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. Mol Cancer 5:1

Buuh ZY, Lyu Z, Wang RE (2017) Interrogating the roles of post-translational modifications of non-histone proteins. J Med Chem 61:3239–3252

Calpena E, Palau F, Espinós C, Galindo MI (2015) Evolutionary history of the smyd gene family in metazoans: a framework to identify the orthologs of human smyd genes in *Drosophila* and other animal species. PLoS ONE 10:e0134106

Cao XJ, Arnaudo AM, Garcia BA (2013) Large-scale global identification of protein lysine methylation in vivo. Epigenetics 8:477–485

Cassidy-Hanley D, Bowen J, Lee JH, Cole E, VerPlank LA, Gaertig J, Gorovsky MA, Bruns PJ (1997) Germline and somatic transformation of mating *Tetrahymena thermophila* by particle bombardment. Genetics 146:135–147

Chen X, Bracht JR, Goldman AD, Dolzhenko E, Clay DM, Swart EC, Perlman DH, Doak TG, Stuart A, Amemiya CT, Sebra RP, Landweber LF (2014) The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. Cell 158:1187–1198

Chen X, Gao S, Liu YF, Wang YY, Wang YR, Song WB (2016) Enzymatic and chemical mapping of nucleosome distribution in purified micro-and macronuclei of the ciliated model organism, *Tetrahymena thermophila*. Sci China Life Sci 59:909–919

Chen Y, Tsai CH, Wang PY, Teng SC (2017) SMYD3 promotes homologous recombination via regulation of H3K4-mediated gene expression. Sci Rep 7:3842

Chen X, Wang YY, Sheng YL, Warren A, Gao S (2018) GPSit: an automated method for evolutionary analysis of nonculturable ciliated microeukaryotes. Mol Ecol Resour 18:700–713

Chen X, Jiang YH, Gao F, Zheng WB, Krock TJ, Stover NA, Lu C, Katz LA, Song WB (2019) Genome analyses of the new model protist *Euplotes vannus* focusing on genome rearrangement and resistance to environmental stressors. Mol Ecol Resour 19:1292–1308

Cheng CY, Young JM, Lin CYG, Chao JL, Malik HS, Yao MC (2016) The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in *Tetrahymena thermophila*. Gene Dev 30:2724–2736

Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ (2004) Regulation of p53 activity through lysine methylation. Nature 432:353–360

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165

Diehl F, Brown MA, Van Amerongen MJ, Novoyatleva T, Wietelmann A, Harriss J, Ferrazzi F, Böttger T, Harvey RP, Tucker PW, Engel FB (2010) Cardiac deletion of Smyd2 is dispensable for mouse heart development. PLoS ONE 5:e9748

Donlin LT, Andresen C, Just S, Rudensky E, Pappas CT, Kruger M, Jacobs EY, Unger A, Zieseniss A, Dobenecker MW, Voelkel T, Chait BT, Gregorio CC, Rottbauer W, Tarakhovsky A, Linke WA (2012) Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. Genes Dev 26:114–119

Du SJ, Tan X, Zhang J (2014) SMYD proteins: key regulators in skeletal and cardiac muscle development and function. Anat Rec 297:1650–1662

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Edwards JR, Yarychkivska O, Boulard M, Bestor TH (2017) DNA methylation and DNA methyltransferases. Epigenet Chromatin 10:23

Feng LF, Wang GY, Hamilton EP, Xiong J, Yan GX, Chen K, Chen X, Dui W, Plemens A, Khadr L (2017) A germline-limited piggyBac transposase gene is required for precise excision in *Tetrahymena* genome rearrangement. Nucleic Acids Res 45:9481–9502

Gao S, Xiong J, Zhang CC, Berquist BR, Yang RD, Zhao M, Molascon AJ, Kwiatkowski SY, Yuan DX, Qin ZH, Wen JF, Kapler GM, Andrews PC, Miao W, Liu YF (2013) Impaired replication elongation in *Tetrahymena* mutants deficient in histone H3 Lys 27 monomethylation. Genes Dev 27:1662–1679

Gao F, Warren A, Zhang QQ, Gong J, Miao M, Sun P, Xu DP, Huang J, Yi ZZ, Song WB (2016) The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci Rep 6:24874

Giakountis A, Moulos P, Sarris ME, Hatzis P, Talianidis I (2017) Smyd3-associated regulatory pathways in cancer. Semin Cancer Biol 42:70–80

Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. Cell 128:635–638

Gordon A, Hannon G (2010) Fastx-toolkit. FASTQ/A short-reads preprocessing tools. Unpublished Available online at: http://hannonlab.cshl.edu/fastx_toolkit/

Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, Harriss JV, Maika SD, Kuziel WA, King HL, Olson EN (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. Nat Genet 31:25–32

Guerin F, Arnaiz O, Boggetto N, Denby Wilkes C, Meyer E, Sperling L, Duharcourt S (2017) Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. BMC Genom 18:327

Guillemette B, Drogaris P, Lin HHS, Armstrong H, Hiragami-Hamada K, Imhof A, Bonneil E, Thibault P, Verreault A, Festenstein RJ (2011) H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. PLoS Genet 7:e1001354

Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 6:731–740

Hamamoto R, Toyokawa G, Nakakido M, Ueda K, Nakamura Y (2014) SMYD2-dependent HSP90 methylation promotes cancer cell proliferation by regulating the chaperone complex formation. Cancer Lett 351:126–133

Hamamoto R, Saloura V, Nakamura Y (2015) Critical roles of non-histone protein lysine methylation in human tumorigenesis. Nat Rev Cancer 15:110–124

Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler EV, Russ C, Zeng Q, Fan L, Levin JZ, Shea T, Young SK, Hegarty R, Daza R, Gujja S, Wortman JR et al (2016) Structure of the germline

genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. eLife 5:e19090
Herz HM, Garruss A, Shilatifard A (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem Sci 38:621–639
Hu LP, Zhu YT, Qi C, Zhu YJ (2009) Identification of smyd4 as a potential tumor suppressor gene involved in breast cancer development. Cancer Res 69:4067–4072
Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, Kubicek S, Opravil S, Jenuwein T, Berger SL (2006) Repression of p53 activity by Smyd2-mediated methylation. Nature 444:629–632
Huang JB, Zhang TT, Zhang QQ, Li Y, Warren A, Pan HB, Yan Y (2018) Further insights into the highly derived haptorids (Ciliophora, Litostomatea): phylogeny based on multigene data. Zool Scr 47:231–242
Hubert Á, Mitani Y, Tamura T, Boicu M, Nagy I (2014) Protein complex purification from *Thermoplasma acidophilum* using a phage display library. J Microbiol Meth 98:15–22
Hughes MA, Langlais C, Cain K, MacFarlane M (2013) Isolation, characterisation and reconstitution of cell death signalling complexes. Methods 61:98–104
Ivanov GS, Ivanova T, Kurash J, Ivanov A, Chuikov S, Gizatullin F, Herrera-Medina EM, Rauscher F, Reinberg D, Barlev NA (2007) Methylation-acetylation interplay activates p53 in response to DNA damage. Mol Cell Biol 27:6756–6769
Jakobsson ME, Małecki J, Nilges BS, Moen A, Leidel SA, Falnes PØ (2017) Methylation of human eukaryotic elongation factor alpha (eEF1A) by a member of a novel protein lysine methyltransferase family modulates mRNA translation. Nucleic Acids Res 45:8239–8254
Jiang YY, Sirinupong N, Brunzelle J, Yang Z (2011) Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain. PLoS ONE 6:e21640
Jiang F, Liu Q, Wang YL, Zhang J, Wang HM, Song TQ, Yang ML, Wang XH, Kang L (2017) Comparative genomic analysis of SET domain family reveals the origin, expansion, and putative function of the arthropod-specific SmydA genes as histone modifiers in insects. GigaScience 6:1–16
Karras GI, Yi S, Sahni N, Fischer M, Xie J, Vidal M, D'Andrea AD, Whitesell L, Lindquist S (2017) HSP90 shapes the consequences of human genetic variation. Cell 168:856–866
Kohl M, Wiese S, Warscheid B (2011) Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol 696:291–303
Lanouette S, Mongeon V, Figeys D, Couture JF (2014) The functional diversity of protein lysine methylation. Mol Syst Biol 10:724
Leinhart K, Brown M (2011) SET/MYND lysine methyltransferases regulate gene transcription and protein activity. Genes 2:210–218
Levine AJ, Berger SL (2017) The interplay between epigenetic changes and the p53 protein in stem cells. Genes Dev 31:1195–1201
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323
Li LX, Fan LX, Zhou JX, Grantham JJ, Calvet JP, Sage J, Li XG (2017) Lysine methyltransferase SMYD2 promotes cyst growth in autosomal dominant polycystic kidney disease. J Clin Invest 127:2751–2764
Liu YF, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD (2007) RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. Genes Dev 21:1530–1545
Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA methylation in the mammalian genome. Cell 167:233–247
Luo XT, Yan Y, Shao C, Al-Farraj SA, Bourland WA, Song WB (2018) Morphological, ontogenetic and molecular data support strongylidiids as being closely related to Dorsomarginalia (Protozoa, Ciliophora) and reactivation of the family Strongylidiidae Fauré-Fremiet, 1961. Zool J Linn Soc-Lond 184:237–254
Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448–3449
Mao FB, Liu Q, Zhao XL, Yang HN, Guo S, Xiao LY, Li XF, Teng HJ, Sun ZS, Dou YL (2018) EpiDenovo: a platform for linking regulatory de novo mutations to developmental epigenetics and diseases. Nucleic Acids Res 46:D92–D99
Mar JC, Wells CA, Quackenbush J (2011) Defining an informativeness metric for clustering gene expression data. Bioinformatics 27:1094–1100
Mazur PK, Reynoird N, Khatri P, Jansen PW, Wilkinson AW, Liu S, Barbash O, Van Aller GS, Huddleston M, Dhanak D, Tummino PJ, Kruger RG, Garcia BA, Butte AJ, Vermeulen M, Sage J, Gozani O (2014) SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. Nature 510:283–287
Mazur PK, Gozani O, Sage J, Reynoird N (2016) Novel insights into the oncogenic function of the SMYD3 lysine methyltransferase. Transl Cancer Res 5:330–333
Miao W, Xiong J, Bowen J, Wang W, Liu Y, Braguinets O, Grigull J, Pearlman RE, Orias E, Gorovsky MA (2009) Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. PLoS ONE 4:e4429
Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway computing environments workshop, Institute of Electrical and Electronics Engineers, New Orleans, Louisiana
Moore KE, Gozani O (2014) An unexpected journey: lysine methylation across the proteome. BBA Gene Regul Mech 1839:1395–1403
Noto T, Mochizuki K (2017) Whats, hows and whys of programmed DNA elimination in *Tetrahymena*. Open Biol 7:170172
Noto T, Kataoka K, Suhren JH, Hayashi A, Woolcock KJ, Gorovsky MA, Mochizuki K (2015) Small-RNA-mediated genome-wide trans-recognition network in *Tetrahymena* DNA elimination. Mol Cell 59:229–242
Ohtomo-Oda R, Komatsu S, Mori T, Sekine S, Hirajima S, Yoshimoto S, Kanai Y, Otsuji E, Ikeda E, Tsuda H (2016) SMYD2 overexpression is associated with tumor cell proliferation and a worse outcome in human papillomavirus-unrelated nonmultiple head and neck carcinomas. Hum Pathol 49:145–155
Orias E, Hamilton EP, Orias JD (1999) *Tetrahymena* as a laboratory organism: useful strains, cell culture, and cell line maintenance. Methods Cell Biol 62:189–211
Qian C, Zhou MM (2006) SET domain protein lysine methyltransferases: structure, specificity and catalysis. Cell Mol Life Sci 63:2755–2763
Ramadoss S, Guo G, Wang CY (2017) Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. Oncogene 36:47–59
Rasmussen TL, Ma Y, Park CY, Harriss J, Pierce SA, Dekker JD, Valenzuela N, Srivastava D, Schwartz RJ, Stewart MD (2015) Smyd1 facilitates heart development by antagonizing oxidative and ER stress responses. PLoS ONE 10:e0121765
Sheng YL, He M, Zhao FQ, Shao C, Miao M (2018) Phylogenetic relationship analyses of complicated class Spirotrichea based on transcriptomes from three diverse microbial eukaryotes: *Uroleptopsis citrina*, *Euplotes vannus* and *Protocruzia tuzeti*. Mol Phylogenet Evol 129:338–345
Sima S, Richter K (2018) Regulation of the Hsp90 system. Biochem Biophys Acta 1865:889–897

Sirinupong N, Brunzelle J, Ye J, Pirzada A, Nico L, Yang Z (2010) Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. J Biol Chem 285:40635–40644

Sirinupong N, Brunzelle J, Doko E, Yang Z (2011) Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. J Mol Biol 406:149–159

Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z (2015) Structure and function of SET and MYND domain-containing proteins. Int J Mol Sci 16:1406–1428

Spellmon N, Sun X, Xue W, Holcomb J, Chakravarthy S, Shang W, Edwards B, Sirinupong N, Li C, Yang Z (2017) New open conformation of SMYD3 implicates conformational selection and allostery. AIMS Biophys 4:1–18

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758–771

Stender JD, Pascual G, Liu W, Kaikkonen MU, Do K, Spann NJ, Boutros M, Perrimon N, Rosenfeld MG, Glass CK (2012) Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. Mol Cell 48:28–38

Strzyz P (2016) Non-coding RNA: 7SK dampens transcription at super-enhancers. Nat Rev Mol Cell Biol 17:202

Taipale M, Jarosz DF, Lindquist S (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol 11:515–528

Thompson EC, Travers AA (2008) A *Drosophila* Smyd4 homologue is a muscle-specific transcriptional modulator involved in development. PLoS ONE 3:e3008

Tracy C, Warren JS, Szulik M, Wang L, Garcia J, Makaju A, Russell K, Miller M, Franklin S (2018) The smyd family of methyltransferases: role in cardiac and skeletal muscle physiology and pathology. Curr Opin Microbiol 1:140–152

Voelkel T, Andresen C, Unger A, Just S, Rottbauer W, Linke WA (2013) Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function. BBA Mol Cell Res 1833:812–822

Walter J, Hümpel A (2017) Introduction to epigenetics. In: Heil R, Seitz S, König H, Robienski J (eds) Epigenetics, futures of technology, science and society. Springer, Wiesbaden, pp 11–29

Wang Q, Wang KY, Ye ML (2017a) Strategies for large-scale analysis of non-histone protein methylation by LC-MS/MS. Analyst 142:3536–3548

Wang YR, Wang YY, Sheng YL, Huang JB, Chen X, Al-Rasheid KAS, Gao S (2017b) A comparative study of genome organization and epigenetic mechanisms in model ciliates, with an emphasis on *Tetrahymena*, *Paramecium* and *Oxytricha*. Eur J Protistol 61:376–387

Wang YY, Chen X, Sheng YL, Liu YF, Gao S (2017c) N^6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*. Nucleic Acids Res 45:11594–11606

Wang YY, Sheng YL, Liu YF, Pan B, Huang JB, Warren A, Gao S (2017d) N^6-methyladenine DNA modification in the unicellular eukaryotic organism *Tetrahymena thermophila*. Eur J Protistol 58:94–102

Wang YR, Wang CD, Jiang YH, Katz LA, Gao F, Yan Y (2019) Further analyses of variation of ribosome DNA copy number and polymorphism in ciliates provide insights relevant to studies of both molecular ecology and phylogeny. Sci China Life Sci 62:203–214

Wickham H (2016) ggplot2: elegant graphics for data analysis, 2nd edn. Springer, New York

Woehrer SL, Aronica L, Suhren JH, Busch CJL, Noto T, Mochizuki K (2015) A *Tetrahymena* Hsp90 co-chaperone promotes siRNA loading by ATP-dependent and ATP-independent mechanisms. EMBO J 34:559–577

Wozniak GG, Strahl BD (2014) Hitting the "mark": interpreting lysine methylation in the context of active transcription. BBA Gene Regul Mech 1839:1353–1361

Wu LP, Lee SY, Zhou B, Nguyen UT, Muir TW, Tan S, Dou YL (2013) ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. Mol Cell 49:1108–1120

Xiong J, Lu YM, Feng JM, Yuan DX, Tian M, Chang Y, Fu CJ, Wang GY, Zeng HH, Miao W (2013) *Tetrahymena* functional genomics database (TetraFGD): an integrated resource for *Tetrahymena* functional genomics. Database. https://doi.org/10.1093/database/bat008

Xu J, Bo T, Song WB, Wang W (2019a) Metabolomic fingerprint of the model ciliate, *Tetrahymena thermophila* determined by untargeted profiling using gas chromatography-mass spectrometry. J Ocean Univ China 18:654–662

Xu J, Li X, Song WB, Wang W, Gao S (2019b) Cyclin Cyc2 is required for elongation of meiotic micronucleus in the unicellular eukaryotic model organism *Tetrahymena thermophila*. Sci China Life Sci 62:668–680

Yan Y, Fan YB, Luo XT, El-Serehy HA, Bourland W, Chen XR (2018) New contribution to the species-rich genus *Euplotes*: morphology, ontogeny and systematic position of two species (Ciliophora; Euplotia). Eur J Protistol 64:20–39

Yi X, Jiang XJ, Li XY, Jiang DS (2017) Histone lysine methylation and congenital heart disease: from bench to bedside. Int J Mol Med 40:953–964

Zhang XD, Huang L, Wu T, Feng YF, Ding YY, Ye PF, Yin ZJ (2015) Transcriptomic analysis of ovaries from pigs with high and low litter size. PLoS ONE 10:e0139514

Zhang TT, Wang CD, Laura AK, Gao F (2018) A paradox: rapid evolution rates of germline-limited sequences are associated with conserved patterns of rearrangements in cryptic species of *Chilodonella uncinata* (Protist, Ciliophora). Sci China Life Sci 61:1071–1078

Zhao XL, Wang YY, Wang YR, Liu YF, Gao S (2017) Histone methyltransferase TXR1 is required for both H3 and H3. 3 lysine 27 methylation in the well-known ciliated protist *Tetrahymena thermophila*. Sci China Life Sci 60:264–270

Zhao Y, Yi ZZ, Warren A, Song W (2018) Species delimitation for the molecular taxonomy and ecology of the widely distributed microbial eukaryote genus *Euplotes* (Alveolata, Ciliophora). Proc Biol Sci 285:20172159

Zhao XL, Xiong J, Mao FB, Sheng YL, Chen X, Feng LF, Dui W, Yang WT, Kapusta A, Feschotte C (2019) RNAi-dependent Polycomb repression controls transposable elements in *Tetrahymena*. Genes Dev 33:348–364

Zheng WB, Wang CD, Yan Y, Gao F, Doak TG, Song WB (2018) Insights into an extensively fragmented eukaryotic genome: de novo genome sequencing of the multinuclear ciliate *Uroleptopsis citrina*. Genome Biol Evol 10:883–894

Zhou B, Wang JY, Lee SY, Xiong J, Bhanu N, Guo Q, Ma PL, Sun YQ, Rao RC, Garcia BA (2016) PRDM16 suppresses MLL1r leukemia via intrinsic histone methyltransferase activity. Mol Cell 62:222–236

Zhu Y, Zhu MX, Zhang XD, Xu XE, Wu ZY, Liao LD, Li LY, Xie YM, Wu JY, Zou HY (2016) SMYD3 stimulates EZR and LOXL2 transcription to enhance proliferation, migration, and invasion in esophageal squamous cell carcinoma. Hum Pathol 52:153–163