

A functional single nucleotide polymorphism of SET8 is prognostic for breast cancer

Ben Liu^{1,*}, Xining Zhang^{1,*}, Fengju Song¹, Qun Liu^{1,2}, Hongji Dai¹, Hong Zheng¹, Ping Cui¹, Lina Zhang¹, Wei Zhang^{3,4}, Kexin Chen¹

¹Department of Epidemiology and Biostatistics, Key Laboratory of Breast Cancer Prevention and Therapy, Ministry of Education, Key Laboratory of Cancer Prevention and Therapy, Tianjin, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

²Departments of Neurosurgery, Key Laboratory of Breast Cancer Prevention and Therapy, Ministry of Education, Key Laboratory of Cancer Prevention and Therapy, Tianjin, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

³Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁴Department of Cancer Biology, Comprehensive Cancer Center of Wake Forest Baptist Medical Center, Winston-Salem, NC 27157, USA

*These authors contributed equally to this work

Correspondence to: Kexin Chen, **email:** chenkexin@tjmuch.com

Wei Zhang, **email:** wzhang@mdanderson.org, wezhang@wakehealth.edu

Keywords: SET8, single nucleotide polymorphisms (SNP), rs16917496, prognosis, breast cancer

Received: November 30, 2015

Accepted: April 10, 2016

Published: April 29, 2016

ABSTRACT

A single-nucleotide polymorphism (SNP) locus rs16917496 (T > C) within the 3'-untranslated region (3'-UTR) of SET8 was associated with susceptibility in several malignancies including breast cancer. To further elucidate the prognostic relevance of this SNP in breast cancer, we conducted a clinical study as well as SET8 expression analysis in a cohort of 1,190 breast cancer patients. We demonstrated the expression levels of SET8 in TT genotype were higher than in CC genotypes, and high levels of SET8 were associated with poor survival. SET8 expression was significantly higher in breast tumor tissue than in paired adjacent normal tissue. In addition, survival analysis in 315 patients showed SNP rs16917496 was an independent prognostic factor of breast cancer outcome with TT genotype associated with poor survival compared with CC/CT genotypes. Thus, this SNP may serve as a genetic prognostic factor and a treatment target for breast cancer. Future studies are warranted.

INTRODUCTION

Breast cancer is the most common malignancy in women in many countries, including China [1, 2]. In spite of the generally good prognosis of breast cancer patients, wide variation in survival indicates that genetic factors may affect the outcomes of breast cancer [3, 4]. Thus, exploring potential genetic biomarkers could benefit to a more appropriate prediction and effective therapeutic strategy of breast cancer prevention and treatment.

In 2008, the first epidemiological study demonstrated that single-nucleotide polymorphisms (SNPs) at microRNA (miRNA) -binding sites within the 3'-untranslated (3'-UTR) region were associated with cancer risk [5], prompting us to propose the polymorphisms in microRNA targets were a gold mine for molecular epidemiology [6].

Afterward numerous investigations have revealed the SNPs within miRNA-binding region may contribute to cancer risk and outcomes by altering miRNA-mRNA binding affinities and miRNA-targeted gene expression [7, 8]. For example, He *et al.* found that rs8752 located with *miR-485-5p*-binding site of *HPGD* gene was related to the risk of breast cancer [9]. Zhang *et al.* found that *miR-367*-binding site rs1044129 in *RYR3* gene was associated with poor survival of patients with breast cancer [10].

SET8 (also known as SETD8, PR-SET7, KMT5A), located on chromosome 12q24.31, is a specific histone H4 lysine 20 methyltransferase (H4K20me1) [11]. Prior studies indicated SET8 may exert various functions in a series of biological processes, including maintaining genome integrity [12], controlling cell-cycle progression and development [13], regulating gene transcription [14],

and mediating DNA repair and damage through its histone monomethylating activity [15]. Moreover, SET8 was also found to bind and methylate nonhistone proteins such as p53, TWIST, Wnt-activated genes, PCNA, ER α and AR [16–21]. All of the above discoveries suggested that SET8 may have a link with carcinogenesis and cancer progression.

Based on a large case control cohort, we first demonstrated that SNP rs16917496-T/C located in the 3'UTR of the SET8 mRNA was associated with the risk of early onset of breast cancer, and this SNP region was predicted as a potential binding site of *miR-502* [22]. This SNP was subsequently shown by others to be a susceptibility factor for a number of cancers, including non-small cell lung cancer [23], epithelial ovarian cancer [24], childhood acute lymphoblastic leukemia [25], and cervical cancer [26]. This broad spectrum of association suggests that this SNP is a robust genetic regulatory factor fundamental to cells. However, the role of the SET8 3'-UTR SNP in breast cancer prognosis has remained unclear and has not been reported, which is the main motivation of this investigation.

RESULTS

Clinical characteristics of breast cancer patients

A total of 1,190 pathologically confirmed breast cancer patients were enrolled in the study. The demographic and clinical characteristics of patients were summarized in Table 1. The median age at diagnosis was 54 years (range, 29–89 years). The median follow-up time of the 315 breast cancer patients who had complete follow-up information from this cohort was 82 months (range, 78–115 months). During the follow-up period, 26 patients died from breast cancer, and 14 patients were loss follow-up. Of these 315 patients, the expression of SET8 mRNA in 30 pairs of tumor and adjacent normal sample was analyzed.

The association of SET8 expression with T allele and poor outcome

In order to evaluate the biological relevance of the rs16917496 polymorphisms, we examined SET8 relative expression through semi-quantitative RT-PCR (qRT-PCR) in 77 breast cancer patients with different SET8 genotypes. The results showed that breast tumors had higher expression of SET8 mRNA in TT genotype than CC genotype ($P = 0.024$) (Figure 1A). In addition, we also measured the protein expression of SET8 in tumor tissues by Western blot in 44 patients, the results indicated that the SET8 protein in TT genotype was higher than CC genotype ($P = 0.015$) (Figure 1B and 1C).

To further elucidate the correlation of SET8 expression with overall survival (OS) and disease-free

survival (DFS), we performed Kaplan-Meier analysis by stratifying patients according to SET8 median expression. Kaplan-Meier survival curves suggested that patients with high expression of SET8 had poor OS and DFS compared with the SET8 low expression group ($P = 0.009$ and $P = 0.029$, respectively) (Figure 1D and 1E). Analysis by univariate Cox proportional hazard model also showed similar results both in OS (HR = 5.90; 95% CI: 1.29–26.96) and DFS analysis (HR = 5.90; 95% CI: 1.08–14.29) (Figure 1D and 1E).

The mRNA expression of SET8 in breast cancer tissues

We subsequently evaluated the relative expression of SET8 in 30 pairs of breast cancer tissues and paired normal tissues. The result suggested that the relative expression of SET8 was up-regulated in breast cancer tissues than in paired normal tissues ($P < 0.001$) (Figure 2A). This result was also validated in the TCGA cohort of 166 pairs of breast cancer tissues and normal tissues ($P = 0.03$) (Figure 2B).

The protein expression of SET8 in breast cancer tissues

In addition, expression of SET8 protein was also examined in 25 breast cancer tissues (Figure 3A–3F) and 10 adjacent non-cancerous tissues (Figure 3G and 3H) by immunohistochemical staining. SET8 protein was strongly expressed in the nucleus in breast cancer tissue. The results showed that the protein levels of SET8 were significantly higher in breast cancer tissues compared with the adjacent non-cancerous tissues ($P = 0.022$) (Figure 3I).

Association of T allele with poor outcome

The association between the rs16917496 genotypes and the clinical and pathological features of 1,190 breast cancer patients was shown in Table 1. We found that the rs16917496 genotypes were associated with ER status (TT vs. CC + CT, $P = 0.011$), and the expression of p53 was significantly lower in CC genotype than in CT and TT genotypes ($P = 0.047$) (Supplementary Figure 1). To determine whether the rs16917496 genotype was associated with the outcome of breast cancer patients, we analyzed the association of rs16917496 genotypes with overall survival (OS) and disease-free survival (DFS) in 315 breast cancer patients who had sufficient follow-up data. Kaplan-Meier survival curves suggested that compared with the TT genotype, the genotypes with C allele was significantly associated with longer OS and DFS of breast cancer patients (Figure 4A and 4C). After adjustment for potential confounding factors including age, TMN stage, lymph node metastasis, and molecular subtype, the multivariate Cox regression analysis was

Table 1: Association of SNP in SET8 3'-UTR and clinicopathological features of 1190 breast cancer patients

Variable	Number	CC + CT (%)	TT (%)	P value
Age at diagnosis(years)				
< 50	582	335(57.6)	247(42.4)	0.138
≥ 50	608	324(53.3)	284(46.7)	
Tumor size (cm)				
< 2	372	209(56.2)	163(43.8)	0.647
≥ 2	538	294(54.6)	244(45.4)	
TNM stage				
I + II	511	288(56.4)	223(43.6)	0.700
III	85	46(54.1)	39(45.9)	
Lymph node metastasis^a				
No	682	386(56.6)	296(43.4)	0.302
Yes	480	257(53.5)	223(46.5)	
Histological grade^a				
I + II	665	368(55.3)	297(44.7)	0.295
III	131	79(60.3)	52(39.7)	
Family history of breast cancer^a				
No	779	423(54.3)	356(45.7)	0.097
Yes	323	193(59.8)	130(40.2)	
Menopause^a				
No	525	292(57.0)	226(43.0)	0.426
Yes	570	311(54.6)	259(45.4)	
ER^a				
Negative	495	292(59.0)	203(41.0)	0.011*
Positive	606	311(51.3)	295(48.7)	
PR^a				
Negative	446	252(56.5)	194(43.5)	0.352
Positive	656	352(53.7)	304(46.3)	
HER2^a				
Negative	697	380(54.5)	317(45.5)	0.420
Positive	334	191(57.2)	143(42.8)	
Molecular subtype^a				
Luminal type	494	259(52.4)	235(47.6)	0.118
HER-2 overexpression	137	80(58.4)	57(41.6)	
Basal-like	171	104(60.8)	67(39.2)	
PCNA^a				
+	130	71(54.6)	59(45.4)	0.767
++	746	414(55.5)	332(44.5)	
+++	63	32(50.8)	31(49.2)	
P53^a				
Negative	626	335(53.5)	291(46.5)	0.293
Positive	392	223(56.9)	169(43.1)	

^aTotal case number was less than 1190 owing to missing data;

*Indicate statistically significant ($P < 0.05$).

performed to evaluate the factors for breast cancer prognosis. Among all prognosis factors, TT genotype of rs16917496 was found to be significantly associated with increased risk of cancer-related death. The relative risk (RR) was 3.61, with 95% confidence interval (CI) ranging from 1.05 to 12.35. As expected, age at diagnosis, TNM stage, Lymph node metastasis, p53 status and molecular subtype was also significant predictors of outcome (Figure 4B). Multivariate Cox regression analysis

showed that rs16917496 genotype (TT vs. CC + CT) was marginally associated with DFS of breast cancer (HR = 2.56, 95% CI: 0.96–7.65) (Figure 4D).

DISCUSSION

In this study, we uncovered that TT genotype of rs16917496 on SET8 3'-UTR region was significantly associated with poor outcome of breast cancer in a Chinese

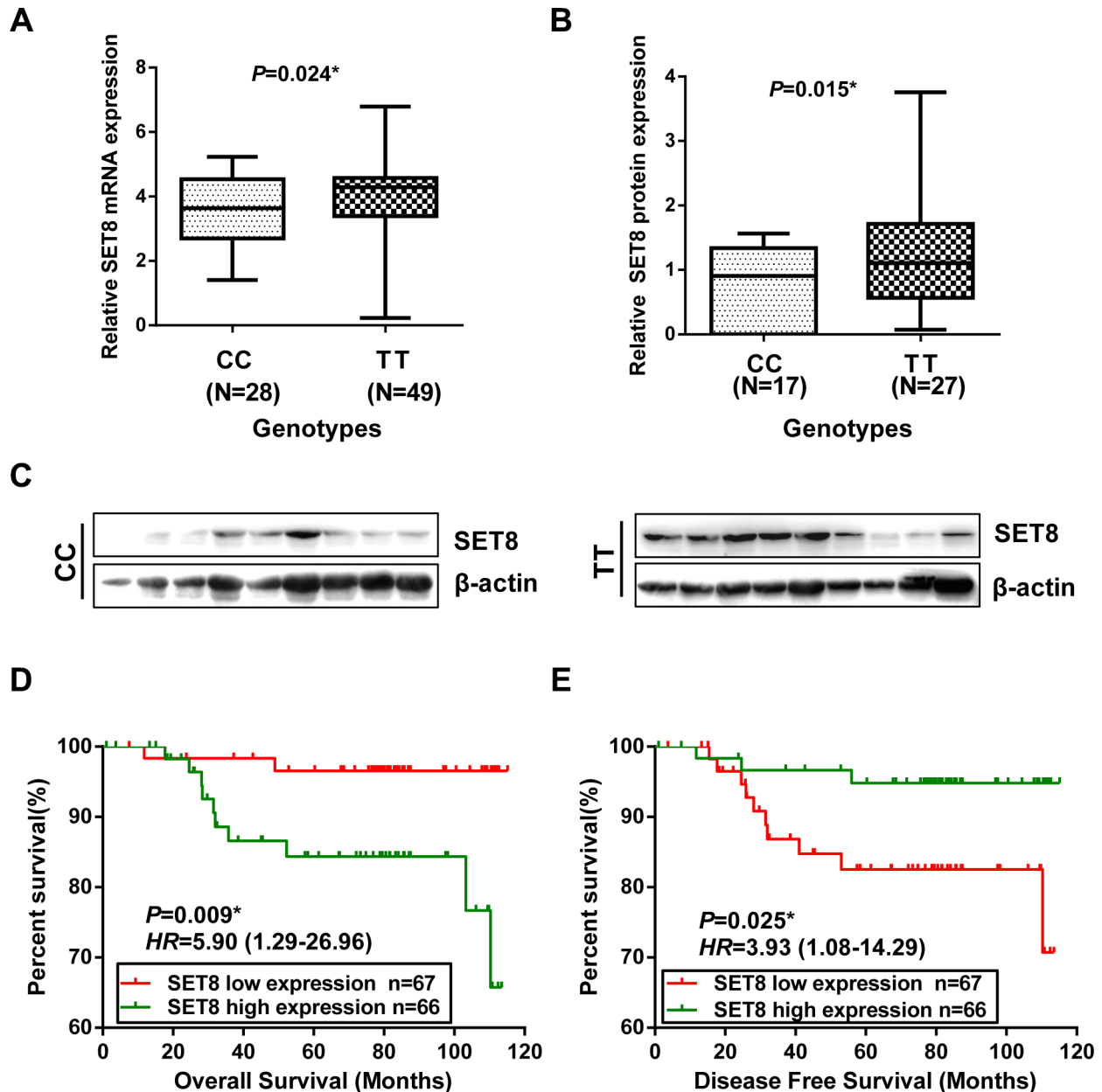


Figure 1: Functional relevance of SET8 3'-UTR SNP on SET8 expression and the association of SET8 expression with the prognosis of breast cancer patients. (A) The relative expression level of SET8 mRNA in breast cancer tumor tissues measured by taqman qRT-PCR in 77 breast cancer patients. Housekeeping gene GAPDH was used as the reference. (B) The graph represents the relative quantitation of protein expression of SET8 in tumor tissues by Western blot in 44 breast cancer cases. (C) The image shows the representative Western blot result of SET8 expression between two genotypes from panel B. (D) Kaplan–Meier curves of overall survival of breast cancer patients with high or low SET8 expression level. (E) Kaplan–Meier curves of disease free survival of breast cancer patients with high or low SET8 expression level. *P* values are from the log-rank test. HR with 95% CI was from Univariate analysis of OS and DFS.

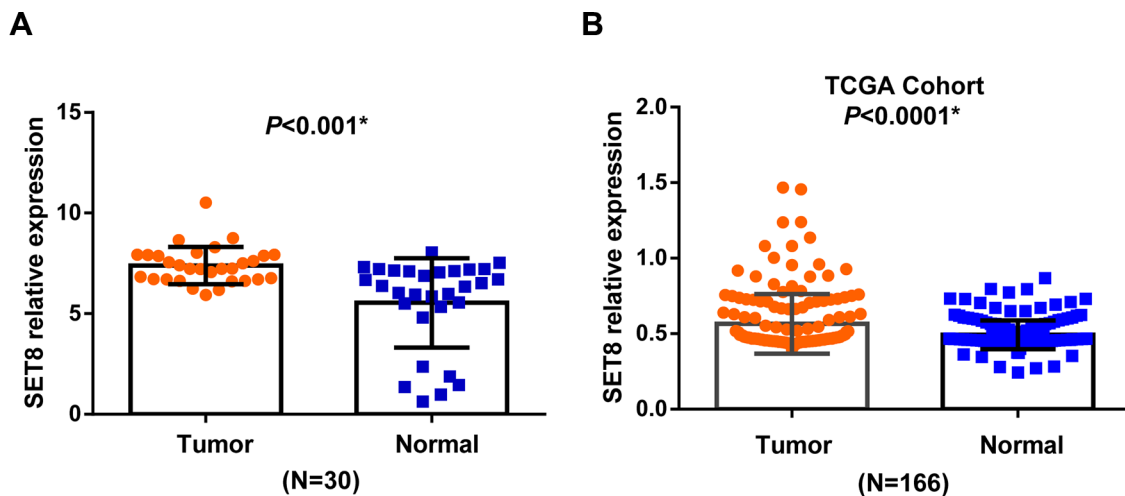


Figure 2: The mRNA expression of SET8 in breast cancer tissues and paired normal tissues. (A) The relative expression SET8 in 30 pairs of breast cancer tissues and paired normal tissues by qRT-PCR. (B) The validation results in 166 pairs of breast cancer tissues and paired normal tissues from TCGA database.

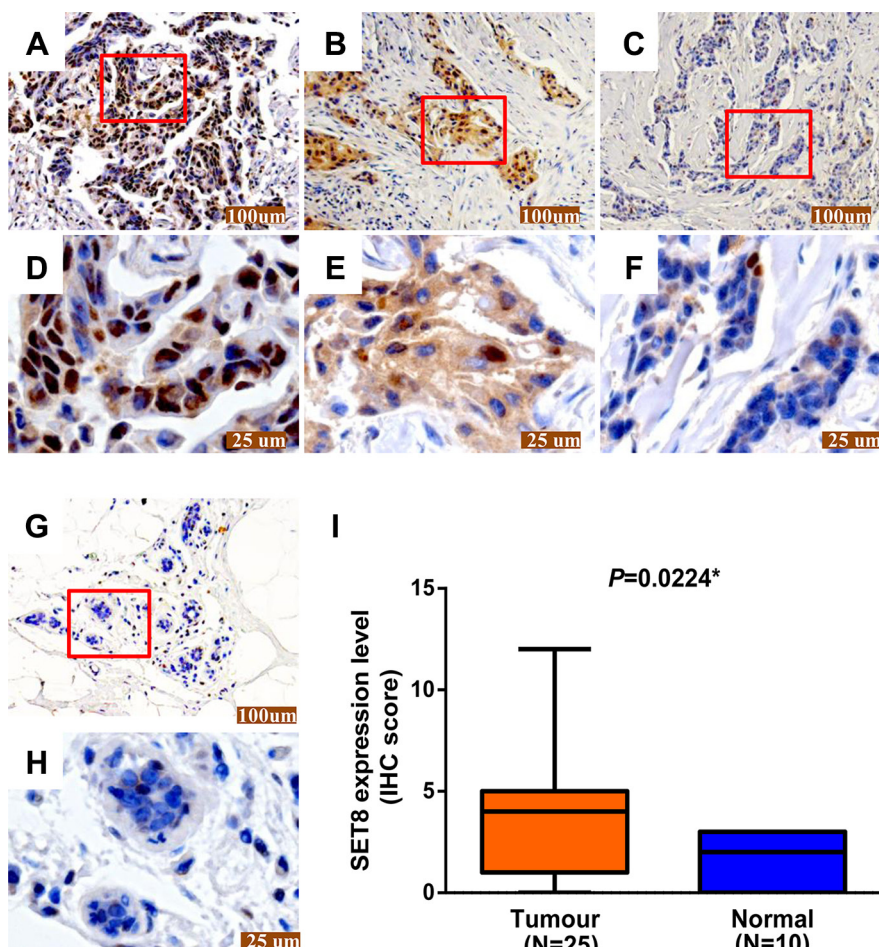


Figure 3: The SET8 protein expression in breast cancer tissues and adjacent normal tissues. (A–C) Representative images of different degrees of staining was detected by immunohistochemical staining in breast cancer tissue sections: (A) strong, (B) moderate, (C) weak. (D–F) Higher magnification images of A–C, respectively. (G) Representative images of SET8 protein expression in breast adjacent normal tissues were detected by immunohistochemical staining. (H): Higher magnification images of G. Scale bar represents 100 μm in 100 \times magnification (A–C and G) and 25 μm in 400 \times magnification (D–F and H), respectively. (I) Semi-quantitative levels of immunohistochemical staining (IHC score) in samples of breast cancer tissues and adjacent normal tissues. $*P < 0.05$

population, and had allele-specific increased on SET8 expression. More importantly, these results remained significant after adjustment for other potential predictors of patient outcome in this patient cohort study. To the best of our knowledge, this is the first epidemiological study investigating the association of this SNP with breast cancer survival.

A growing number of researchers suggest that polymorphisms within miRNA-binding sites in 3'-UTR of genes contribute to carcinogenesis and progress by affecting miRNA-regulated gene expression [27]. In addition to the risk of cancer, the SNP rs16917496 has been related to the clinical outcome of several cancer types. For example, Ding et al found the small cell lung cancer patients with SET8 CC + CT genotypes have good prognosis [28]. In another study, Guo et al. has demonstrated liver cancer patients with SET8 CC genotype postoperatively live longer than those without [29]. In non-Hodgkins lymphoma (NHL), Diao et al. found patients carrying SET8 CC genotype have significantly

longer survival time than the patients with SET8 CT or TT genotype [30]. In line with these studies, our results discovered that the functional SNP associated with breast cancer survival via influencing SET8 expression. One rational explanation was *miR-502* bound more tightly to the C allele of rs16917496 than to the T allele. To test this hypothesis, We used four miRNA target predicting databases (miRWalk, TargetScan, miRMap and miRanda) to identify the candidate targets of miR-502. We found that the four programs all predicted SET8 as candidate targets of miR-502 and SNP16917496 located within the miR-502-binding site (Supplementary Figure 2A and 2B). In addition, we computationally determined the different minimum free energy (MFE) for both the SET8 genotype using an online thermodynamic software RNAHYBRID. The MFE of genotype C [mfe = -26.8 kcal/mol] was smaller than that of the T genotype [mfe = -25.1 kcal/mol], which suggested that miR-502 has a higher binding affinity for the C genotype (Supplementary Figure 2C and 2D). We went further to examine SET8 protein

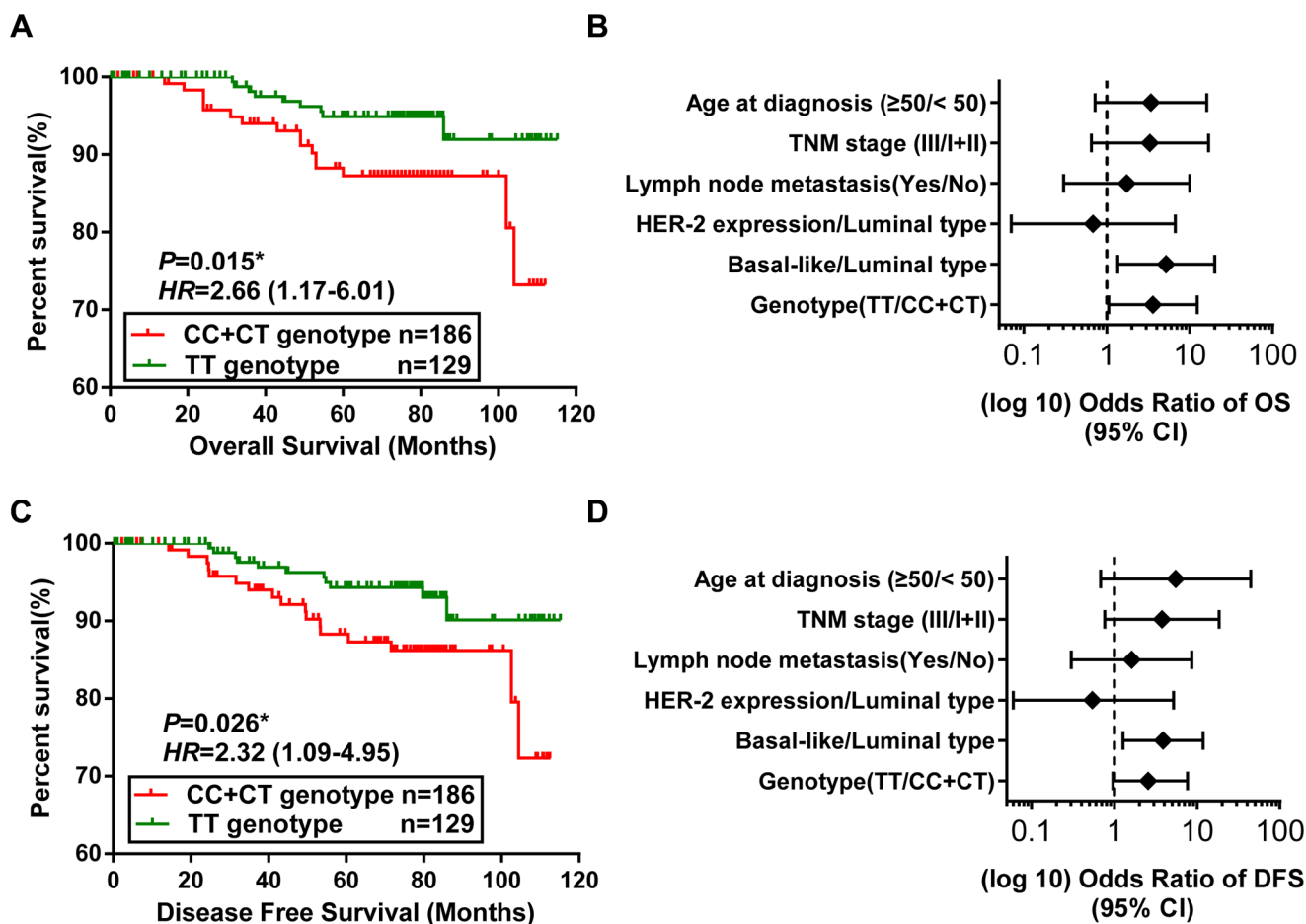


Figure 4: Association between the SNP in SET8 3'-UTR and breast cancer survival. (A, C) Kaplan-Meier analysis of overall survival (OS) (A) and disease free survival (DFS) (C) of patients with the CC + CT phenotypes vs. the TT genotype in 315 cases. Red line represents CC and CT genotypes, and the green line represents TT genotypes. *P* values are from the log-rank test. HR with 95% CI was from Univariate analysis of OS and DFS. (B, D) Forest plot of the association of breast cancer prognosis factors with OS (B) and DFS (D) of 315 breast cancer patients. The results were from Cox regression analysis. The confounding factors include age, TMN stage, lymph node metastasis, and molecular subtype.

level in breast cancer patients with different rs16917496 genotypes. A significant increase was observed in both SET8 mRNA and protein levels in TT genotype relative to the CC genotype. One immunohistochemical study on 192 NSCLC tissue aligned with our results in which SET8 protein expression level of CC genotype found to be low [31]. By searching the Genotype-Tissue Expression (GTEx) portal database (<http://www.gtexportal.org/home/>), we performed an expression quantitative trait loci (eQTL) analysis to verify the association between the genotype and SET8 gene expression level in a cohort of 183 normal breast tissues. The result showed the SET8 expression significantly reduce in CC genotype compare to TT genotype ($P = 0.021$), which is consistent with our finding. The above evidence supports the hypothesis that functional 3'-UTR SNP harbor in miRNA target sites may regulate the miRNA-target gene expression through affecting the binding affinity, which may in turn affect the inhibitory regulation ability of miRNA on mRNA's translation to protein. In addition, these empirical findings consistently corroborate our result that the T allele of rs16917496 alters SET8 functions and therefore is associated with breast cancer outcome.

In our results, the data also suggested the expression of SET8 has a significant association with prognosis of breast cancer. SET8 is known as the sole methyltransferase for H4K20me1, which usually associates with the higher-order chromatin structure and the variability of actively transcribed genes in cancerous conditions as the epigenetic mark [32, 33]. Accumulative clinical data showed that the abnormal expression of histone methyltransferase is correlated to cancer occurrence as well as prognosis of various cancers. For example, the expression level of histone lysine methyltransferase EZH2 which mainly catalyze H3K27me3 is higher in the breast cancer, prostate cancer and bladder cancer, and has been used as a prognostic marker in breast cancer and metastatic prostate cancer [34, 35]. Elevated expression of Suv39h1, another histone methyltransferase which can catalyze H3K9me2, has been found to associate with higher incidence of hepatocellular carcinoma recurrence [36]. As recently more and more research groups devote to explore molecule targeting inhibitors of histone methyltransferase, such as EZH2- H3K27 inhibitor EPZ – 6438 [37] and DOT1L- H3K79 methylation inhibitors EPZ– 5676 [38], it has become a promising target for anti-tumor therapy [39, 40]. In summary, the result from our current study as well as that in other researcher's work revealed that SET8 closely associated with the occurrence of a wide variety of tumor development and prognosis, and therefore, is believed to become potential new targets for the treatment of tumor.

While interpreting these results and presenting the contributions of this study, several limitations should be considered and discussed as well. First, based on the accumulated evidence, we infer that SET8 has a key

role in breast cancer progression by inducing cancer cell proliferation and migration. This implication needs further in-depth functional experimental research to investigate and help understand the complicated mechanism *in vitro* and *in vivo*. Second, because the sample size of survival and SET8 expression analysis was relatively modest or small, replication studies in a larger population are needed in order to validate our results. Third, our results suggested the SET8 rs16917496 genotype was associated with ER and p53 expression, which cannot be easily explained for regulation of protein methylation by SET8 due to the fact that it will not change the transcriptional level of these genes. Thus, the role of this SNP as a modifier needs further examination and clarification through laboratory-based functional studies.

In summary, we confirmed that the rs16917496 polymorphism can predict breast cancer patients' survival in Chinese population. The genetic variation rs16917496 with SET8 3'-UTR could modify the breast cancer outcome by regulating the expression of SET8. These findings present new evidence that the genetic variant might be a molecular switcher to fine tune the miRNA-mRNA binding and related gene function through a genetic regulatory circuit. Therefore, SET8 may be used as a new treatment target and prognosis biomarker for breast cancer.

MATERIALS AND METHODS

Ethics statement

The study had been approved by the Ethical Committee of Tianjin Medical University Cancer Institute and Hospital. The informed consent form was signed by each participant.

Study subjects

A total of 1,190 newly diagnosed primary breast cancer cases was recruited from Tianjin Medical University Cancer Institute and Hospital between 2001 and 2008. Demographic features and clinical information were separately collected from structured questionnaire and medical records. The clinical information included tumor features and severity such as tumor size, lymph node metastasis and estrogen receptor (ER). We followed up 315 patients among this cohort until July 2014 through regular telephone, clinical visits or email. The overall survival (OS) time was calculated from the diagnostic date to the date of death or last follow-up date.

Genomic DNA samples

Each participant donated 20 ml of blood into an ethylene diamine tetraacetic acid (EDTA) vacutainer tube used for DNA extraction and genotyping. Total genomic

DNA was isolated from whole blood using the QIAGEN DNA Blood Mini Kit (QIAGEN, Inc.). The isolated DNA was stored at -20°C until analysis.

Genotypes

RFLP-PCR was used to identify the genotypes of the SET8 (rs16917496 C/T) polymorphism within the 3'-UTR in patients and breast cancer cells. The amplification conditions were: 95°C for 5 min, 35 cycles of 95°C for 45 s, 66°C for 40 s, and 72°C for 30 s, and a final extension step of 72°C for 10 min. The primers on the *miR-502* binding site were 5'-GGCCTCACGACGGTGCTAC-3' and 5'-GTTCCCCAGGAGGATGCT-TAC-3'. A 308-bp DNA fragment was produced in this process, then digested the DNA produced by SWAI (New England BioLabs, Inc.) overnight at 25°C . Allel C lacks the SWAI restriction site and only produces a 308-bp band, allel T produces two bands including 149-bp and 159-bp, TC heterozygote produces three bands (149-bp, 159-bp, and 308-bp). The detailed method of RFLP PCR was described in our previous study [22].

RNA extraction and quantitative real-time PCR

Total RNA was extracted using Trizol reagents (Invitrogen, USA), according to the manufacturer's instructions. Reverse transcription was performed using M-MLV Reverse Transcriptase (Applied Biosystems, USA). QRT-PCR was performed using ABI 7900 Real-time PCR (Applied Biosystems, USA). Probe for SET8 was 5'-FAM-CCCTGTCCGAAGGAGCTCCAGGAAGA-TAMRA-3'. Forward and reverse primers for SET8 are 5'-CGAAACTTACGGATTCT-3' and 5'-CGATGAGGTCAATCTTCATT-3', respectively. Samples were done in triplicate. GAPDH were used as an endogenous control to normalize the level of SET8 expression. The relative expression of SET8 was calculated using the $2^{-\Delta\Delta\text{ct}}$ method. SDS 2.4 Software (Applied Biosystems, USA) was used to analyze data. The detailed protocol was described in our previous study [22].

Western blot

Protein was extracted from breast cancer tissues and cells at 48 h after transfection with siSET8 using RIPA buffer (Thermo, USA). Protein concentration was measured using the BCA protein quantitation kit by Protein Assay Kit (Bio-Rad). Total 40ug of protein mixed with $5\times$ loading buffer was separated by SDS-PAGE electrophoresis and transferred to PVDF membranes. Membranes were blocked with 5% non-fat milk for 1 h, and incubated with diluted 1:2000 SET8 monoclonal antibodies (Thermo, USA) at 4°C overnight. Secondary antibodies were added at concentrations of 1:4000. β -actin was used as an internal reference (Santa Cruz Biotechnology, USA). C-DiGit Chemiluminescent

Western Blot Scanner (LI-COR, USA) was used to visualize the protein bands.

Immunohistochemical (IHC) Staining

Formalin-fixed and paraffin-embedded blocks the breast cancer specimens and were analyzed using anti-SET8 antibody (Abgent, China). The deparaffinized sections were boiled in a sodium citrate buffer (pH 6.0) for 15min as an antigen retrieval method. After quenching endogenous peroxidases with hydrogen peroxide, the sections were rinsed with Tris-HCl buffer twice and incubated with anti-SET8 antibody diluted at 1:800 overnight at 4°C . The primary antibodies were detected using a secondary antibody with HRP polymer (EnVision™ Detection Systems, DAKO, Carpinteria, CA). Diaminobenzidine was used as chromogen according to the manufacture's instruction. Hematoxylin was used for nuclear counterstaining, and the sections were mounted and coverslipped. The tissue staining results were scored based on signal distribution (distribution score) and intensity (intensity score). The distribution score includes 0 (0–5%), 1 (6–25%), 2 (26–50%), 3 (51–75%) and 4 (76–100%), which indicates the percentage of positive cells in all the tumor cells present in a sample. The signal intensity consists of 0 (no signal), 1 (weak), 2 (moderate), or 3 (strong). The final staining score was the product of the distribution and intensity scores.

TCGA mRNA dataset analysis

SET8 relative expression data of breast cancer patients were obtained from the Cancer Genome Atlas (TCGA) data portal (tcga-data.nci.nih.gov/tcga/). The relative expression data of SET8 consists of normalized (level 3) from the Illumina HiSeq_RNASeqV2 platform and Agilent4502A_07 platform (paired tumor tissue and adjacent normal tissues, $n = 166$). Level 3 normalized mRNA expression was normalized with the quantile method using the R/Bioconductor package of preprocessCore [41].

Bioinformatic miRNA target prediction and analysis

miRWalk 2.0 database was used to predict potential target genes for miR-502, which combines the information with a comparison of binding sites resulting from other three existing miRNA-target prediction programs: TargetsCan6.2, miRNAMap, miRanda-rel2010 [42]. The Venn diagram was drawn by Venny2.1 software. The minimum free energy (MFE) hybridization of miR-502 with C or T genotype of SET8 3'-UTR was predicted using RNAHYBRID software (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html>) [43].

Statistical analysis

The expression of SET8 was calculated using the $2^{-\Delta\Delta ct}$ method, and the variance of SET8 relative expression and the expression of breast cancer marker in different genotypes was calculated using one-way ANOVA method. A chi-squared test was used to show the association between SNP rs16917496 genotypes and clinicopathological features of patients. The differences in survival were examined using the log-rank test, and Kaplan-Meier curve and Cox regression were used to analyze the relationship between genotype and survival time of breast cancer. Statistical analyses were performed with SPSS 19.0 (SPSS Inc., Chicago, IL, USA), and graphs were generated by Graphpad Prism 5.0 statistic software (Graphpad Software, Inc.) and STATA 12.0 (version 12.0; StataCorp, College Station, Texas, USA). All statistical tests were two-sided, and *P* value less than 0.05 was considered statistically significant.

ACKNOWLEDGMENTS

The tissue bank is jointly supported by Tianjin Cancer Institute and Hospital and the National Foundation for Cancer Research (USA).

CONFLICTS OF INTEREST

The authors indicate no potential conflicts of interest.

GRANT SUPPORT

This work was supported by grants from the National Natural Science Foundation of China (no. 81172762 to K.C., no. 81071627, 81572445, 81202275 to B.L. and no. 81473039 to F.S.) and the program for the Tianjin City High School Science & Technology Fund Planning Project (no. 20090137). This work was partially supported by the Program for Changjiang Scholars and Innovative Research Team in University in China (IRT_14R40 to K.C.), the Chinese 863 Program (no. 2012AA02A207 to K.C. and J.W.), and the National Key Scientific and Technological Project (2015BAI12B15 to K.C.)

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015; 65:87–108.
2. Chen W, Zheng R, Zeng H, Zhang S. The updated incidences and mortalities of major cancers in China, 2011. *Chin J Cancer*. 2015; 34:53.
3. Yu ZB, Li Z, Jolicoeur N, Zhang LH, Fortin Y, Wang E, Wu MQ, Shen SH. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic acids research*. 2007; 35:4535–4541.
4. Yu JC, Ding SL, Chang CH, Kuo SH, Chen ST, Hsu GC, Hsu HM, Hou MF, Jung LY, Cheng CW, Wu PE, Shen CY. Genetic susceptibility to the development and progression of breast cancer associated with polymorphism of cell cycle and ubiquitin ligase genes. *Carcinogenesis*. 2009; 30:1562–1570.
5. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, Novotny J, Forsti A, Hemminki K, Canzian F, Landi S. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis*. 2008; 29:579–584.
6. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W. Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. *Carcinogenesis*. 2008; 29:1306–1311.
7. Saetrom P, Biesinger J, Li SM, Smith D, Thomas LF, Majzoub K, Rivas GE, Alluin J, Rossi JJ, Krontiris TG, Weitzel J, Daly MB, Benson AB, et al. A risk variant in an miR-125b binding site in *BMPR1B* is associated with breast cancer pathogenesis. *Cancer Res*. 2009; 69:7459–7465.
8. Martinez E, Silvy F, Fina F, Bartoli M, Krahn M, Barlesi F, Figarella-Branger D, Iovanna J, Laugier R, Ouaiissi M, Lombardo D, Mas E. Rs488087 single nucleotide polymorphism as predictive risk factor for pancreatic cancers. *Oncotarget*. 2015. doi: 10.18632/oncotarget.5627.
9. He N, Zheng H, Li P, Zhao Y, Zhang W, Song F, Chen K. miR-485-5p binding site SNP rs8752 in *HPGD* gene is associated with breast cancer risk. *PloS one*. 2014; 9:e102093.
10. Zhang L, Liu Y, Song F, Zheng H, Hu L, Lu H, Liu P, Hao X, Zhang W, Chen K. Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (*RYR3*) affects breast cancer risk and calcification. *Proc Natl Acad Sci U S A*. 2011; 108:13653–13658.
11. Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, Frenkiel TA, Martin SR, Sarma K, Reinberg D, Gambin SJ, Wilson JR. Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev*. 2005; 19:1444–1454.
12. Houston SI, McManus KJ, Adams MM, Sims JK, Carpenter PB, Hendzel MJ, Rice JC. Catalytic function of the PR-Set7 histone H4 lysine 20 monomethyltransferase is essential for mitotic entry and genomic stability. *J Biol Chem*. 2008; 283:19478–19488.
13. Wu S, Wang W, Kong X, Congdon LM, Yokomori K, Kirschner MW, Rice JC. Dynamic regulation of the PR-Set7 histone methyltransferase is required for normal cell cycle progression. *Genes Dev*. 2010; 24:2531–2542.
14. Congdon LM, Houston SI, Veerappan CS, Spektor TM, Rice JC. PR-Set7-Mediated Monomethylation of Histone H4 Lysine 20 at Specific Genomic Regions Induces Transcriptional Repression. *J Cell Biochem*. 2010; 110:609–619.
15. Oda H, Hubner MR, Beck DB, Vermeulen M, Hurwitz J, Spector DL, Reinberg D. Regulation of the histone H4

- monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage. *Mol Cell*. 2010; 40:364–376.
16. Yao L, Li Y, Du F, Han X, Li X, Niu Y, Ren S, Sun Y. Histone H4 Lys 20 methyltransferase SET8 promotes androgen receptor-mediated transcription activation in prostate cancer. *Biochem Biophys Res Commun*. 2014.
 17. Yang F, Sun L, Li Q, Han X, Lei L, Zhang H, Shang Y. SET8 promotes epithelial-mesenchymal transition and confers TWIST dual transcriptional activities. *EMBO J*. 2012; 31:110–123.
 18. Takawa M, Cho HS, Hayami S, Toyokawa G, Kogure M, Yamane Y, Iwai Y, Maejima K, Ueda K, Masuda A, Dohmae N, Field HI, Tsunoda T, et al. Histone Lysine Methyltransferase SETD8 Promotes Carcinogenesis by Deregulating PCNA Expression. *Cancer Res*. 2012; 72:3217–3227.
 19. Shi X, Kachirskaja I, Yamaguchi H, West LE, Wen H, Wang EW, Dutta S, Appella E, Gozani O. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol Cell*. 2007; 27:636–646.
 20. Li Z, Nie F, Wang S, Li L. Histone H4 Lys 20 monomethylation by histone methylase SET8 mediates Wnt target gene activation. *Proc Natl Acad Sci U S A*. 2011; 108:3116–3123.
 21. Li Y, Sun L, Zhang Y, Wang D, Wang F, Liang J, Gui B, Shang Y. The histone modifications governing TFF1 transcription mediated by estrogen receptor. *J Biol Chem*. 2011; 286:13925–13936.
 22. Song F, Zheng H, Liu B, Wei S, Dai H, Zhang L, Calin GA, Hao X, Wei Q, Zhang W, Chen K. An miR-502-binding site single-nucleotide polymorphism in the 3'-untranslated region of the SET8 gene is associated with early age of breast cancer onset. *Clin Cancer Res*. 2009; 15:6292–6300.
 23. Yang S, Guo H, Wei B, Zhu S, Cai Y, Jiang P, Tang J. Association of miR-502-binding site single nucleotide polymorphism in the 3'-untranslated region of SET8 and TP53 codon 72 polymorphism with non-small cell lung cancer in Chinese population. *Acta Biochim Biophys Sin (Shanghai)*. 2014; 46:149–154.
 24. Wang C, Guo Z, Wu C, Li Y, Kang S. A polymorphism at the miR-502 binding site in the 3' untranslated region of the SET8 gene is associated with the risk of epithelial ovarian cancer. *Cancer Genet*. 2012; 205:373–376.
 25. Hashemi M, Sheybani-Nasab M, Naderi M, Roodbari F, Taheri M. Association of functional polymorphism at the miR-502-binding site in the 3' untranslated region of the SETD8 gene with risk of childhood acute lymphoblastic leukemia, a preliminary report. *Tumour Biol*. 2014.
 26. Yang SD, Cai YL, Jiang P, Li W, Tang JX. Association of a miR-502-Binding Site Single Nucleotide Polymorphism in the 3'-Untranslated Region of SET8 and the TP53 Codon 72 Polymorphism with Cervical Cancer in the Chinese Population. *Asian Pac J Cancer Prev*. 2014; 15:6505–6510.
 27. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, Manoukian S, Secreto G, Ravagnani F, et al. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res*. 2010; 70:2789–2798.
 28. Ding C, Li R, Peng J, Li S, Guo Z. A polymorphism at the miR-502 binding site in the 3' untranslated region of the SET8 gene is associated with the outcome of small-cell lung cancer. *Exp Ther Med*. 2012; 3:689–692.
 29. Guo Z, Wu C, Wang X, Wang C, Zhang R, Shan B. A polymorphism at the miR-502 binding site in the 3'-untranslated region of the histone methyltransferase SET8 is associated with hepatocellular carcinoma outcome. *Int J Cancer*. 2012; 131:1318–1322.
 30. Diao L, Su H, Wei G, Li T, Gao Y, Zhao G, Guo Z. Prognostic value of microRNA 502 binding site SNP in the 3'-untranslated region of the SET8 gene in patients with non-Hodgkin's lymphoma. *Tumori*. 2014; 100:553–558.
 31. Xu J, Yin Z, Gao W, Liu L, Yin Y, Liu P, Shu Y. Genetic variation in a microRNA-502 binding site in SET8 gene confers clinical outcome of non-small cell lung cancer in a Chinese population. *PLoS one*. 2013; 8:e77024.
 32. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 2008; 40:897–903.
 33. Knijnenburg TA, Bismeyjer T, Wessels LF, Shmulevich I. A multilevel pan-cancer map links gene mutations to cancer hallmarks. *Chin J Cancer*. 2015; 34:48.
 34. Kleer CG, Cao Q, Varambally S, Shen RL, Ota L, Tomlins SA, Ghosh D, Sewalt RGAB, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A*. 2003; 100:11606–11611.
 35. Beguelin W, Popovic R, Teater M, Jiang YW, Bunting KL, Rosen M, Shen H, Yang SN, Wang L, Ezponda T, Martinez-Garcia E, Zhang HK, Zheng YP, et al. EZH2 Is Required for Germinal Center Formation and Somatic EZH2 Mutations Promote Lymphoid Transformation. *Cancer cell*. 2013; 23:677–692.
 36. Chiba T, Saito T, Yuki K, Zen Y, Koide S, Kanogawa N, Motoyama T, Ogasawara S, Suzuki E, Ooka Y, Tawada A, Otsuka M, Miyazaki M, et al. Histone lysine methyltransferase SUV39H1 is a potent target for epigenetic therapy of hepatocellular carcinoma. *Int J Cancer*. 2015; 136:289–298.
 37. Baker T, Nerle S, Pritchard J, Zhao B, Rivera VM, Garner A, Gonzalez F. Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. *Oncotarget*. 2015; 6:32646–32655. doi: 10.18632/oncotarget.5066.
 38. Daigle SR, Olhava EJ, Therkelsen CA, Majer CR, Sneeringer CJ, Song J, Johnston LD, Scott MP, Smith JJ,

- Xiao Y, Jin L, Kuntz KW, Chesworth R, et al. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer cell*. 2011; 20:53–65.
39. Wee S, Dhanak D, Li H, Armstrong SA, Copeland RA, Sims R, Baylin SB, Liu XS, Schweizer L. Targeting epigenetic regulators for cancer therapy. *Ann NY Acad Sci*. 2014; 1309:30–36.
40. Tan J, Yang X, Zhuang L, Jiang X, Chen W, Lee PL, Karuturi RK, Tan PB, Liu ET, Yu Q. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev*. 2007; 21:1050–1063.
41. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*. 2010; 11:94.
42. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. 2015; 12:697.
43. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA*. 2004; 10:1507–1517.