Comprehensive analysis based on DNA methylation and RNA-seq reveals hypermethylation of the up-regulated WT1 gene with potential mechanisms in PAM50 subtypes of breast cancer

Chongyang Ren[1,*], Xiaojiang Tang[2,*] and Haitao Lan[3]

[1] Department of Breast Cancer, Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, Guangzhou, Guangdong, China
[2] Department of Breast Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shanxi, China
[3] Academy of Medical Sciences, Sichuan Provincial People's Hospital, Chengdu, Sichuan, China
[*] These authors contributed equally to this work.

Corresponding author
Haitao Lan,
lanhaitao@med.uestc.edu.cn

ABSTRACT

Background. Breast cancer (BC), one of the most widespread cancers worldwide, caused the deaths of more than 600,000 women in 2018, accounting for about 15% of all cancer-associated deaths in women that year. In this study, we aimed to discover potential prognostic biomarkers and explore their molecular mechanisms in different BC subtypes using DNA methylation and RNA-seq.

Methods. We downloaded the DNA methylation datasets and the RNA expression profiles of primary tissues of the four BC molecular subtypes (luminal A, luminal B, basal-like, and HER2-enriched), as well as the survival information from The Cancer Genome Atlas (TCGA). The highly expressed and hypermethylated genes across all the four subtypes were screened. We examined the methylation sites and the downstream co-expressed genes of the selected genes and validated their prognostic value using a different dataset (GSE20685). For selected transcription factors, the downstream genes were predicted based on the Gene Transcription Regulation Database (GTRD). The tumor microenvironment was also evaluated based on the TCGA dataset.

Results. We found that Wilms tumor gene 1 (*WT1*), a transcription factor, was highly expressed and hypermethylated in all the four BC subtypes. All the *WT1* methylation sites exhibited hypermethylation. The methylation levels of the TSS200 and 1stExon regions were negatively correlated with *WT1* expression in two BC subtypes, while that of the gene body region was positively associated with *WT1* expression in three BC subtypes. Patients with low *WT1* expression had better overall survival (OS). Five genes including *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *IGFL2* were predicted as the downstream genes of *WT1*. Those five genes were dysregulated in the four BC subtypes. Patients with a favorable 6-gene signature (low expression of *WT1* and its five predicted downstream genes) exhibited better OS than that with an unfavorable 6-gene signature. We also found a correlation between *WT1* and tamoxifen using STITCH. Higher infiltration rates of CD8 T cells, plasma cells, and monocytes were found in the lower quartile *WT1* group and the favorable 6-gene signature group. In conclusion, we

demonstrated that *WT1* is hypermethylated and up-regulated in the four BC molecular subtypes and a 6-gene signature may predict BC prognosis.

INTRODUCTION

According to the World Health Organization's report, there were approximately 2.09 million breast cancer (BC) cases in 2018, making it the second most common cancer worldwide. Almost 627,000 women died from BC in 2018, accounting for almost 15% of all cancer-associated deaths in women (https://www.who.int/news-room/fact-sheets/detail/cancer). Therefore, BC is a severe medical burden that deserves extensive study. Our study aimed to discover potential prognostic biomarkers in different types of BC and explore their potential mechanisms in an epigenetic perspective.

The recent developments in sequencing technology has raised the interests in studying of the regulatory mechanism of BC progression. While normal gene expression is regulated by an intricate genetic and epigenetic regulatory system (*Arechederra et al., 2018*), dysregulation of oncogenes and tumor suppressor genes occurs in tumor cells (*Dawson & Kouzarides, 2012*). DNA methylation plays an important role in gene expression through various epigenetic mechanisms (*Karemaker & Vermeulen, 2018*). In tumor tissue, DNA hypomethylation shows a disperse distribution, whereas DNA hypermethylation is concentrated on CpG-rich regions, called CpG islands (*Feinberg & Tycko, 2004*). Gene promoter hypermethylation results in gene repression (*Xu et al., 2018*), but hypermethylation in the gene body (including exons and introns) (*Han et al., 2017*) elevates gene expression (*Jones, 2012*; *Renner et al., 2013*; *Wagner et al., 2014*; *Wang et al., 2016*). Wang et al. found that under the action of *DNMT3B*, the fully methylated body region turned refractory to *SP1* binding, releasing *SP1* for promoter binding and driving of gene expression (*Wang et al., 2016*). Gene methylation in the gene body might serve as a therapeutic target of cancers (*Yang et al., 2014*). For example, DNA methyltransferase inhibitors, approved for older acute myeloid leukemia (AML) patients, combined with 5-aza-2'-deoxycytidine predominant synergistic gene down-regulation is associated with gene body demethylation in AML cell line (*Blagitko-Dorfs et al., 2019*). Although the application of DNA methylation has also shown broad therapeutic potential in BC (*Downs et al., 2019*; *Li et al., 2019*; *Prajzendanc et al., 2020*; *Sasidharan Nair et al., 2018*), fewer than twenty papers about gene body methylation in BC were found in PubMed (*Croes et al., 2018*; *De Almeida et al., 2019*; *Flanagan et al., 2009*; *Jin et al., 2019*; *Kim et al., 2018*; *Leadem et al., 2018*; *Li et al., 2017b*; *Liu et al., 2018*; *Peiffer et al., 2019*; *Rodger et al., 2019*; *Shenker et al., 2015*; *Song et al., 2015*; *Stefansson et al., 2015*; *Taslim et al., 2012*; *Windhorst, Song & Gazdar, 2017*). As mentioned above, gene body hypermethylation elevates gene expression (*Jones, 2012*; *Renner et al., 2013*; *Wagner et al., 2014*; *Wang et al., 2016*), so we focused on the hypermethylated and up-regulated genes in our study and investigated whether gene body methylation is correlated to gene expression.

With the hope of achieving precision medicine, researchers have applied molecular biotechnology (defined by mRNA expression of 50 genes (PAM50) (*Pu et al., 2020*)) to classify BC into four molecular subtypes: luminal A, luminal B, basal-like (triple-negative), and HER2-enriched. However, most of the studies about gene body methylation were performed in only one or two selected subtypes of BC (*Leadem et al., 2018*; *Liu et al., 2018*; *Peiffer et al., 2019*; *Rodger et al., 2019*; *Shenker et al., 2015*). Hence, in this study, we specifically focused on the four BC molecular subtypes and combined gene methylation with expression profiles to find the common characteristics among different BC types. Arechederra et al. previously reported that hypermethylation could indicate elevated oncogene levels (*Arechederra et al., 2018*). Therefore, we expected that the screening of hypermethylated and up-regulated oncogenes could identify some potential prognostic biomarkers.

Our study mainly focused on an oncogene: Wilms' tumor 1 (*WT1*) and the potential mechanism that affecting its expression in the four types of BC. *WT1* was reported to be expressed in 87% of primary BC (*Loeb et al., 2001*) and was associated with a poor prognosis of BC (*Miyoshi et al., 2002*). In BC tissues, intragenic regions of *WT1* were hypermethylated (*Lian et al., 2012*). However, these previous studies about *WT1* were investigated in a single subtype of BC and had no subsequent analysis on its molecular mechanism. Hence, in our study, methylation level and its association with *WT1* expression in different types of BC were analyzed and the underlying oncogenic mechanism of *WT1* was further investigated.

In this study, we downloaded DNA methylation datasets and RNA expression profiles of primary BC tissues from The Cancer Genome Atlas (TCGA) to identify common hypermethylated and up-regulated genes among the four molecular subtypes (luminal A, luminal B, basal-like, and HER2-enriched). Based on the TCGA dataset, we also counted the number of methylated sites and examined the correlation between methylation sites and gene expression. The downstream genes of transcription factors were predicted using the Gene Transcription Regulation Database (GTRD). The prognostic value of signature genes was evaluated using the TCGA dataset and validated by the Gene Expression Omnibus (GEO) dataset. Finally, the protein-protein interaction network was constructed using Search Tool for Interacting Chemicals (STITCH).

MATERIALS & METHODS

Data collection

We obtained the DNA methylation data of 306 invasive breast carcinoma (BC) samples (beta values from methylation 450 K) from UCSC Xena (https://xenabrowser.net/datapages/?hub=https://tcga.xenahubs.net:443). The invasive BC datasets consisted of 98 normal tissue samples, 108 luminal A primary tumor tissue samples, 46 luminal B primary tumor tissue samples, 40 basal-like primary tumor tissue samples, and 14 HER2-enriched primary tumor tissue samples. The 450 K microarray contains not only CpG and CNG sites, CpG islands/shores/shelves/open sea, non-coding RNA (microRNAs and long non-coding RNAs), and sites surrounding the transcription start sites (-200 bp to $-1,500$ bp, 5′-UTRs and, exons 1) for coding genes, but also the corresponding gene bodies and 3′-UTR

(*Sandoval et al., 2011*). Based on the platform, the methylation probes can map to a wider variety of gene regions to obtain more data. The gene body region consisted of exons and introns (*Han et al., 2017*). The island (or CpG island) was defined as a 200-bp stretch of DNA with a C+G content of 50% and an observed CpG/expected CpG of over 0.6 according to that proposed by Gardiner-Garden and Frommer in 1987 (*Gardiner-Garden & Frommer, 1987*). The sequences up to 2 kb distant CpG islands were termed "CpG island shores" (*Irizarry et al., 2009*). The sequences from 2 to 4 kb distant CpG islands were denoted as shelves. The rest of the genome was defined as "open sea" (*Visone et al., 2019*).

We also collected the gene expression profiles from 648 human tissues (including 114 normal tissues, 236 luminal A primary tumor tissues, 132 luminal B primary tumor tissues, 103 basal-like primary tumor tissues, and 63 HER2-enriched tumor tissues), which were sequenced using the Illumina HiSeq 2000 RNA Sequencing platform, from UCSC Xena. The validation cohort (GSE20685), which included 327 primary breast cancer samples, was downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/).

Selecting highly expressed and hypermethylated genes

We identified hypermethylated BC genes using the R package ChAMP. First, we loaded the beta value matrix and patient clinical information (including patient ID and tumor subtype). The champ.filter function was carried out to remove low-quality samples and probes. Second, we determined the NA value in the matrix using the Combine method of the champ.impute function, with a k value of 5, a probe cutoff of 0.1, and a sample cutoff of 0.5. Third, we performed quality control using the champ.QC function to ensure that the loaded data would be available for the subsequent analysis. Type II probe normalization was performed using the BMIQ method of the champ.norm function. Finally, the differentially methylated probes across the four BC subtypes were recognized using the champ.DMP function. They were selected with the following criteria: an absolute value of deltaBeta value > 0.2 and an adjusted P value < 0.05 (*Chang et al., 2019*).

The RNA expression profile of the dataset described above showed a log2 (x+1) transformed RSEM (RNA-Seq expression estimation by Expectation-Maximization) normalized count, which was directly downloaded from Xena (https://xenabrowser. net/datapages/?dataset=TCGA.BRCA.sampleMap%2FHiSeqV2&host=https%3A%2F% 2Ftcga.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu% 3A443). Before analysis, we performed gene filtering. Genes with extremely low expression (0 in more than 50% of samples) were removed from the subsequent analysis. We calculated the mean value for each gene in normal and tumor tissue, as well as the fold change of log2 (mean value of tumor tissue/mean value of normal tissue). The P value was computed using the t. test function and adjusted by FDR (false discovery rate) with p.adjust function in R. The gene with |log2 (fold change) |>1 and adjusted P value < 0.05 were defined as dysregulated genes.

Correlation and survival analysis

We calculated the correlation coefficient for methylation sites and gene expression and analyzed the gene co-expression using the Spearman's rank correlation test.

Survival analysis was performed in R Studio with the survminer package via the Cox proportional hazards regression model and visualized by survival package. After inputting patients' survival time and the endpoint information (dead or alive), Kaplan–Meier analysis was used to obtain the survival curve. *P* values were calculated using the log-rank *t*-test.

The download of breast cancer-related genes and human oncogenes

The gene list, containing 228 genes that have been reported to affect the development of BC, was downloaded from the Disease gene search engine (DigSee, http://210.107.182.61/geneSearch/). DigSee is a web tool developed to search MEDLINE abstracts for evidence sentences depicting that genes take part in the development of cancers via biological events (*Kim et al., 2013*). After selecting a type of cancer on DigSee, it will return all genes related to the cancer type and the corresponding references of each gene. On DigSee, 7,449 genes are documented for BC. We chose 228 genes that were supported by at least 10 references in BC to screen for the target genes of this study. The BC-related genes queried on DigSee were listed in Table S1.

The human oncogenes were downloaded from the Oncogene database (http://ongene.bioinfo-minzhao.org/download.html). Oncogene database is a literature-based genetic resource ground on a comprehensive review of research literature about oncogenes (*Liu, Sun & Zhao, 2017*). A total of 802 human oncogenes are recorded on the Oncogene database (Table S2). They were used to further screen for target genes of this study.

WT1 potential target genes prediction

We predicted the *WT1* target genes using GTRD (http://gtrd.biouml.org/). GTRD is a database that began in 2011, which contains transcription factor binding sites identified by ChIP-seq experiments for *Homo sapiens* (*Yevshin et al., 2019*). Using GTRD, we set transcription factor binding site location at promoter $[-1000, +100]$ to predict its target genes. The gene list was shown in Table S3.

Protein-protein interaction network construction

The protein-protein interaction network was generated using STITCH (http://stitch.embl.de/). STITCH is a web to explore the intersections between proteins and small molecules. It integrates these disparate data sources for 430,000 chemicals into a single, easy-to-use resource (*Szklarczyk et al., 2016*). After entering a gene, it will return its related genes and drugs.

Immune infiltration analysis

We analyzed the immune infiltration fraction according to the method as mentioned by *Thorsson et al. (2018)*, which estimated the relative fraction of 22 immune cells of each patient via CIBERSORT (*Newman et al., 2015*). CIBERSORT is a calculation method that can accurately calculate the relative level of multifarious immune cell types in a mixture of compound gene expression. CIBERSORT uses the gene expression signatures of 547 genes (LM22 files) as the input matrix to characterize and quantify immune cell subtypes. Using the TCGA's primary BC tissues RNA expression data and LM22 files as input data, CIBERSORT was implemented in "relative mode" to estimate the relative abundance of

tumor-infiltrating immune cells. Next, we calculated the difference in immune infiltration between various groups using an unpaired t-test.

Statistical analysis

All statistical analyses in our study were carried out using R. During the selection of highly expressed genes, the P value was computed using the t. test function and adjusted by FDR (false discovery rate) with p.adjust function in R. In survival analysis, P values were calculated using the log-rank t-test. The difference in immune infiltration between various groups was calculated using an unpaired t-test.

RESULTS

Selecting highly expressed, hypermethylated genes across the four BC subtypes

In order to find highly expressed and hypermethylated genes across the four BC subtypes, the differential analysis of gene expression and methylation levels were performed in R studio based on the TCGA's BC primary tissue RNA-seq and methylation datasets. We found 268, 321, 440, and 351 highly expressed genes and 5,462, 6,721, 3,357, and 5,805 hypermethylated genes in luminal A, luminal B, basal-like, and HER2-enriched BC tumor tissues, respectively (Fig. 1). Next, we selected genes that were both highly expressed and hypermethylated in BC tumor tissues. We found 84, 127, 80, and 103 genes that were highly expressed and hypermethylated in luminal A BC, luminal B BC, basal-like BC, and HER2-enriched BC, respectively (Fig. 1, Table S4). We downloaded 228 BC-related genes from DigSee and 802 oncogenes from the Oncogenes database. In order to find oncogenes that were highly expressed and hypermethylation across the four types of BC, we selected the genes commonly present in all the six gene lists (228 BC-related genes, 802 oncogenes, and highly expressed and hypermethylated genes in the four subtypes of BC). Ultimately, there was only one remaining gene which was the transcription factor Wilms tumor gene 1 (*WT1*). Therefore, *WT1* was chosen for the subsequent analysis.

WT1 methylation statuses at various methylation sites

To explore the cause of high *WT1* expression, we counted the methylated sites located on *WT1* based on the TCGA's BC primary tissue methylation datasets. For the four BC subtypes (luminal A, luminal B, basal-like, and HER2-enriched), the number of *WT1*-methylated sites were 38, 44, 38, and 36, respectively. Figure 2 showed the methylation levels of statistically discrepant methylated sites in luminal A (Fig. 2A), luminal B (Fig. 2B), basal-like (Fig. 2C), and HER2-enriched BCs (Fig. 2D). In the heatmap, each column represents an individual patient and each row is a cg probe. The left bar shows the cg probe's genomic region. The deep red color stands for high methylation level. We found that all methylation sites showed high methylation levels, particularly cg09695430, cg05940984, cg06516124, cg12006284, cg13638420, cg09234616, cg04456238, and cg10244666. Most of the methylated sites in luminal A, luminal B, basal-like, and HER2-enriched BCs were located in the body regions, and the percentages were 92.1% (35/38), 77.3% (34/44), 100% (38/38), and 100% (36/36), respectively. These might explain the strong *WT1* expression levels in BC tissues (*Yang et al., 2014*).
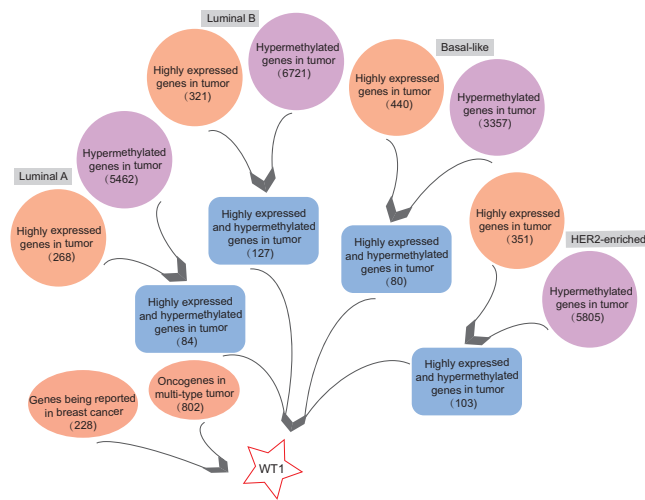
Figure 1 The workflow. Orange bubbles presented the number of highly expressed genes in different subtypes of BC based on TCGA RNA-seq analysis. Purple bubbles presented the number of hypermethylated genes in different subtypes of BC based on TCGA methylation data. Blue boxes showed the number of both up-regulated and hypermethylated genes in four subtypes of BC based on TCGA data. The left red bubble presented the number of gene reported in BC, while the right red bubble showed oncogene number in multi-type tumors. Tumor sample size of luminal A, luminal B, Basal-like, and HER2-enriched BC subtype is 108, 46, 40, and 14, respectively. Normal sample size is 98.

Using the *WT1* RNA expression profiles of the four types of BC primary tumor tissue downloaded from TCGA, we analyzed the correlation between *WT1* expression and the methylated sites. In Fig. 3, the number displayed in each cell is the correlation coefficient of *WT1* expression and the methylated sites. The deep red color stands for a strong correlation. The bar on the left shows the genomic region of each cg probe. Cg13540960, cg05222924, cg20204986, and cg13638420 had fair positive associations (*Akoglu, 2018*) (correlation coefficient > 0.3, $P < 0.05$) with *WT1* expression in all the BC subtypes (Fig. 3). HER2-enriched BC had the highest correlation coefficient (median = 0.69) of all the common methylated sites. In contrast, luminal A, luminal B, and basal-like subtypes had median correlation coefficients of 0.36, 0.34, and 0.36, respectively.

To confirm whether the methylation levels of the different genomic regions were related to *WT1* expression, we calculated the mean beta value (an indicator of methylation level) of each patient in the 1stExon, Body, and TSS200 regions and performed correlation analysis between beta value and *WT1* expression based on the TCGA's BC primary tissue RNA-seq and methylation datasets. In luminal A and luminal B BCs, *WT1* expression was negatively associated with the methylation level of the 1stExon (Figs. 4C and 4D) and TSS200 regions (Figs. 4C and 4F). In luminal A, basal-like, and HER2-enriched BCs, *WT1* expression was positively associated with the methylation level of the gene body (Figs. 4B, 4G and 4H). However, in luminal B BC, *WT1* expression showed no statistically significant association with the methylation level of the gene body (Fig. 4E).
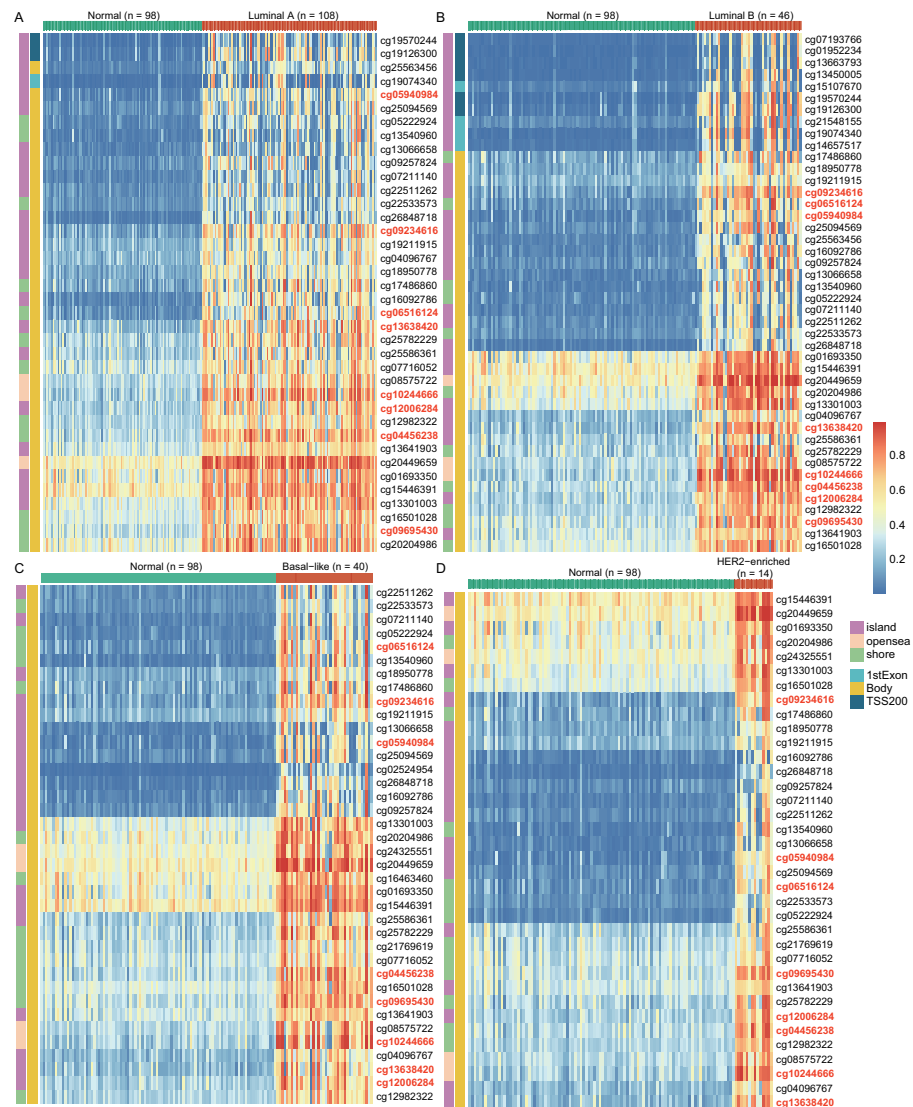
Figure 2 **The differentially *WT1* methylated sites in the four BC subtypes from TCGA methylation data (A-D).** In the heatmap, each column presents one patient and each row is a cg probe. The left bars show cg probe's genomic region. The deeper red means higher methylation levels.

Correlation analyses of the expression of *WT1* and its downstream genes

Since *WT1* is a transcription factor, we further examined the genes regulated by *WT1*. A total of 17,214 genes were predicted as *WT1*- target genes (set transcription factor binding site location on promoter [−1000, +100]) using GTRD. We then calculated the coefficient of association between the expression of the differentially expressed genes and *WT1* in the four types of BC tumors to identify *WT1*'s co-expressed genes based on the TCGA's BC primary tissue RNA-seq dataset. After screening with the selection criteria (*P* value < 0.05 and |correlation coefficient|>0.35), we identified 77, 223, 239, and 279 genes from luminal
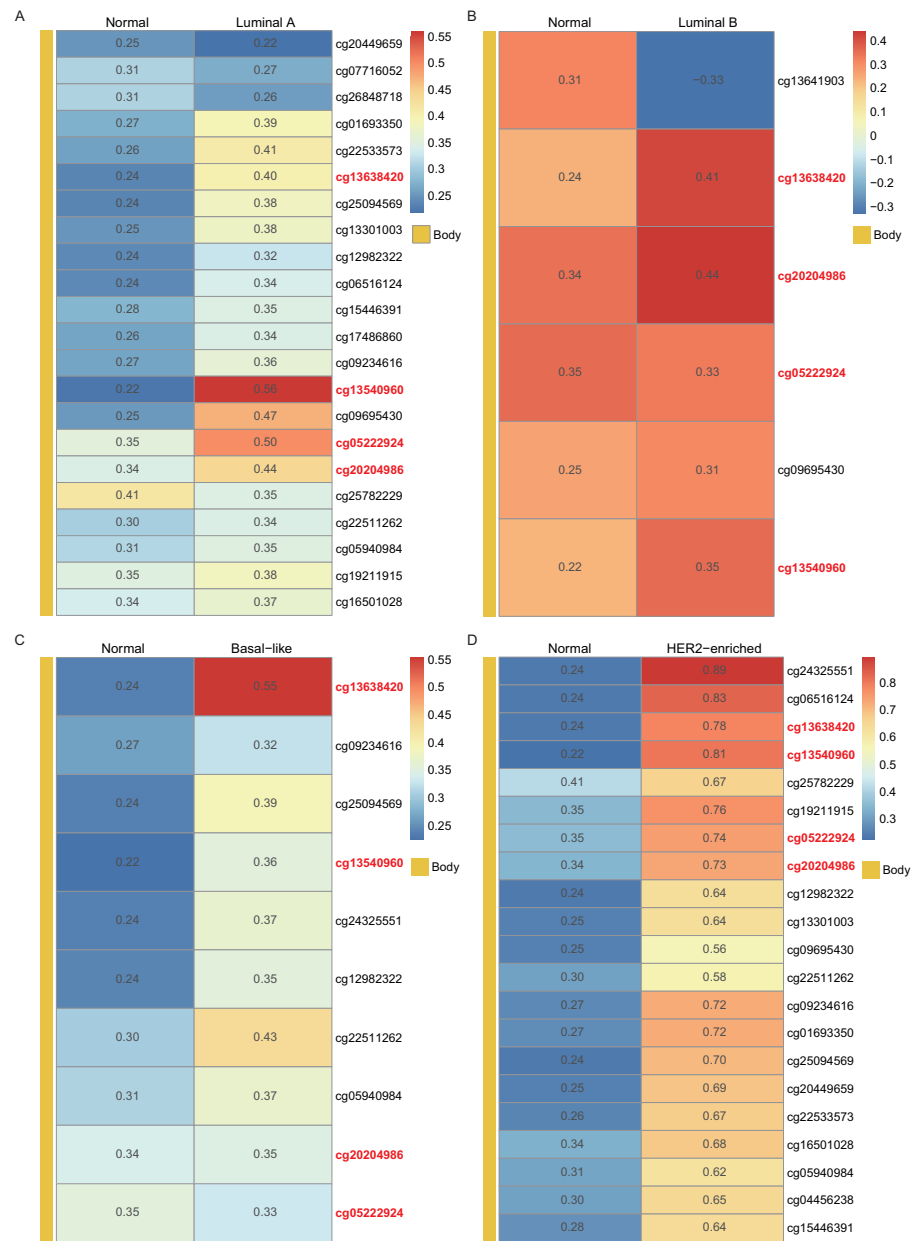
Figure 3 **The correlation of *WT1* methylated sites and *WT1* expression in four subtypes of BC was observed based on TCGA data (A-D).** The number in each small box was the correlation coefficient of *WT1* expression and the methylated sites. The deeper red means higher correlation. The bar in the left showed the genomic region of each cg probe.

A, luminal B, basal-like, and HER2-enriched BC groups, respectively. Among these, we found 5 shared genes, including *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *IGFL2*. Figs. 5A–5E showed the correlation of them with *WT1* expression, of which three genes exhibited fair positive correlations and two genes showed negative correlations (including *GFAP*). A comparison of the gene expression levels in BC tissues and normal tissues was shown in
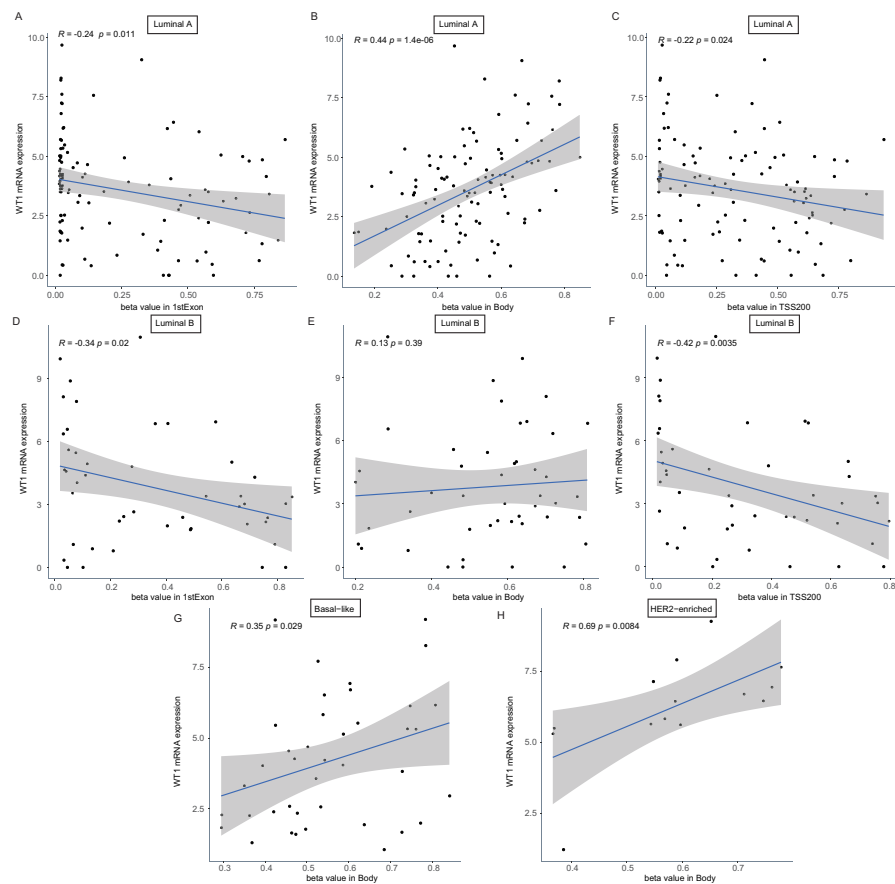
Figure 4 The association analysis between beta value of 1stExon, Body or TSS200 region and *WT1* expression in luminal A (A–C), luminal B (D–F), basal-like (G) or HER2-riched (H) BC based on TCGA data.

Figs. 5F–5K. *WT1*, *COL11A1*, *FGF5*, and *IGFL2* were up-regulated in tumor tissues, while *GFAP* and *CD300LG* were down-regulated. We also calculated the correlation coefficient of *COL11A1*, *GFAP*, and *FGF5* with *WT1* expression in GSE20685. *COL11A1* expression was fairly positively correlated with *WT1* expression and the expression of *FGF5* was poorly positively correlated with *WT1* expression. *GFAP* and *WT1* expressions had no statistically significant correlation (Fig. S1). The *WT1* binding sites of its potential downstream genes were shown in Fig. 6. The binding sites were predicted by GTRD based on ChIP-seq experiment. All the five genes possessed binding sites upstream of their protein-coding regions.

Survival and protein-protein interaction network analysis

Next, we examined *WT1*'s prognostic value based on the TCGA's BC primary tissue RNA-seq and survival datasets. When the median *WT1* expression in tumor tissue was set as the cutoff value to divide patients into high- and low-*WT1* expression groups, we observed no statistically significant differences in overall survival (OS) between the two
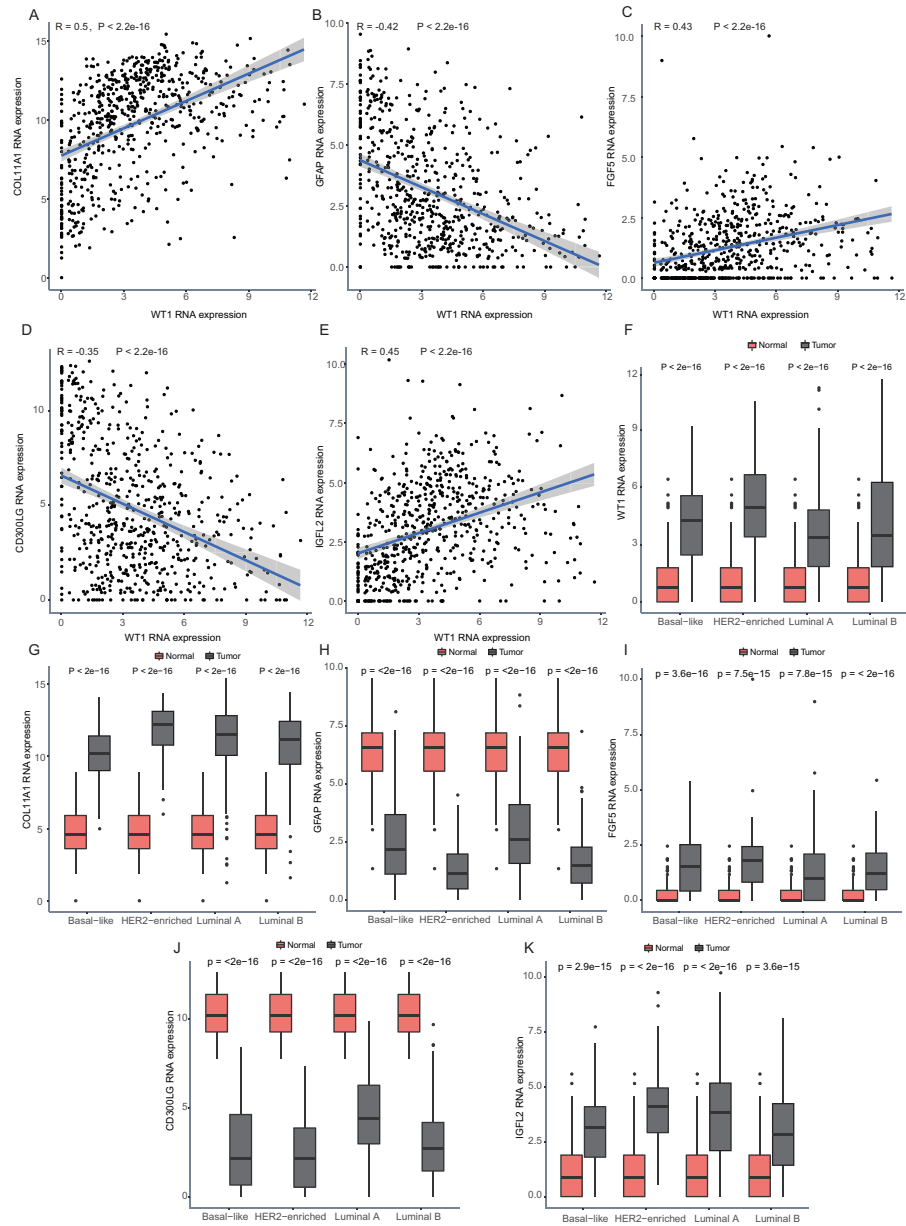
Figure 5 **Correlation analyses and gene expression in the four BC subtypes based on TCGA RNA-seq profiles.** (A–D) Correlation analysis showed *WT1* expression had a correlation with the downstream genes *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *IGFL2*. (E–I) *WT1*, *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *TDO2* expression was shown across the four BC subtypes of BC and normal tissues.

groups using the log-rank test ($P = 0.26$, Fig. 7A) and the Cox regression test (hazard ratio (HR) = 1.2, $P = 0.26$, Table 1). Therefore, we adjusted the cutoff value to the quartiles of *WT1* expression. We found that patients with *WT1* expression below the lower quartile showed better OS rates than patients with *WT1* expression above the upper quartile, and the log-rank test and Cox regression test (HR = 1.68) P values were 0.023 and 0.024, respectively (Fig. 7B, Table 1). When we combined the 5 co-expressed genes and

Figure 6 (A-E) *WT1* binding sites with its potential downstream genes predicted on GTRD.

WT1 as a gene signature to predict OS, the patients with low expression levels (favorable 6-gene signature) exhibited better OS compared to patients with high expression levels (unfavorable 6-gene signature). The log-rank test and Cox regression test (HR = 1.38) *P* values were 0.049 and 0.049, respectively (Fig. 7C, Table 1). The validation cohort showed a similar result, with $P = 0.049$ and HR = 1.55 (Fig. 7D, Table 1). Additionally, we used STITCH to construct the protein-protein interaction network and showed that *WT1* was linked with tamoxifen, a drug for BC treatment (Fig. 7E).

Immune infiltration analysis

To investigate the reason behind the better outcomes of the lower quartile *WT1* group and the favorable 6-gene signature group, we evaluated the differences in tumor microenvironments between groups based on the TCGA's BC primary tissue RNA-seq dataset. The infiltration rates of the CD8 T cells, plasma cells, and monocytes in the lower quartile *WT1* group were significantly higher than that of the higher quartile *WT1* group (Fig. 8A). The CD8 T cells, plasma cells, and monocytes in the favorable 6-gene signature group also had significantly higher infiltration rate than that of the unfavorable 6-gene signature group (Fig. 8B). Moreover, the lymphocytes, follicular helper T cells, and activated NK cells had higher infiltration rates in the favorable 6-gene signature group (Fig. 8B).
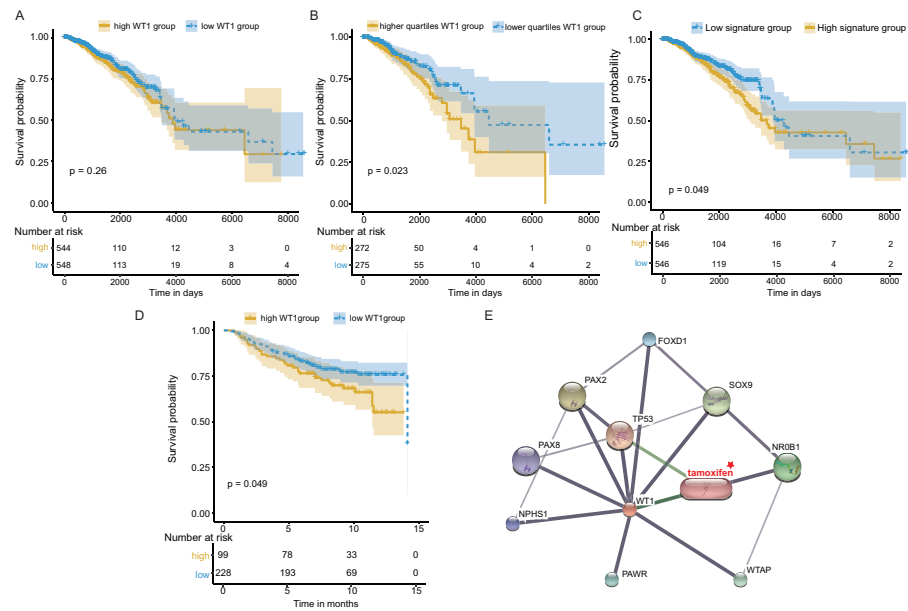
Figure 7 **Survival and protein-protein interaction network analysis.** (A) Survival analysis between the high and low *WT1* expression group was performed using TCGA RNA-seq and survival data. (B) Survival analysis between the upper and lower quartile *WT1* expression group was carried out using TCGA RNA-seq and survival data. (C) Survival analysis between the high and low signature expression group was performed with TCGA RNA-seq and survival data. (D) Survival analysis between the high and low *WT1* expression group in the validation cohort (GSE20685). (E) *WT1* protein-protein interaction network visualized on STITCH.

Table 1 **Hazard ratio and *P* value of each group.**

Data source	Group	Hazard ratio (HR)	*P* value	Lower 95% CI range	upper 95% CI range
TCGA	High WT1 group	1.2	0.26	0.87	1.65
TCGA	High quartiles WT1 group	1.68	0.024	1.07	2.65
TCGA	High signature group	1.38	0.049	1.0	1.90
GSE20685	High WT1 group	1.55	0.049	0.99	2.42

DISCUSSION

In this study, we analyzed the DNA methylation datasets and RNA expression profiles generated by RNA-seq across four molecular subtypes of BC. We found 84, 127, 80, and 103 genes that were highly expressed and hypermethylated in luminal A, luminal B, basal-like, and HER2-enriched BCs, respectively. *WT1,* a gene that was both highly expressed and hypermethylated across all the four subtypes, was selected as a potential prognostic biomarker of BC. McGregor et al. has previously reported that *WT1* expression was up-regulated in BC cells (*McGregor et al., 2018*), which was consistent with our findings. Furthermore, we found that *WT1* exhibited hypermethylation in BC tissues, which was consistent with the results reported by *Kim et al. (2012)*. Compared to Kim's research which
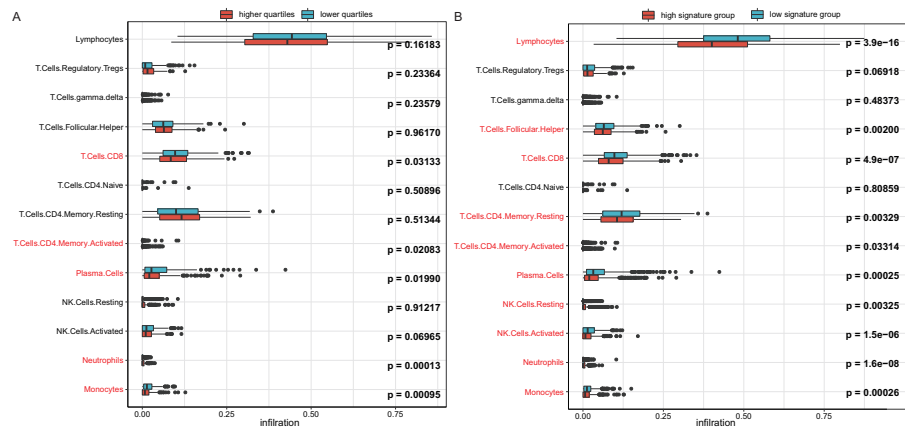
Figure 8 **Immune infiltration.** (A) Comparison between the immune infiltration of the lower quartile *WT1* expression group and the higher quartile group using TCGA RNA-seq profiles. (B) The immune infiltration of the low 6-gene signature group and the high 6-gene signature group using TCGA RNA-seq profiles.

was performed on BC cells and a single subtype of BC, our study additionally showed that *WT1* was highly expressed and hypermethylation in all the four subtypes of BC. Besides, this is the first research reporting the association of *WT1* expression and methylation in the four subtypes of BC. Since DNA methylation in the promoter region can suppress gene expression but up-regulate gene expression in the gene body region (*Yang et al., 2014*), we further examined the detailed distribution of the methylated sites. As expected, the *WT1* methylation sites were mostly or all located in the gene body region. In luminal A, basal-like, and HER2-enriched BCs, *WT1* expression was positively associated with the methylation level of the gene body, supporting that gene body methylation is related to high expression of *WT1*. This phenomenon might be owing to the modulation effects of gene body methylation on the binding of transcription factor to the promoter region (*Wang et al., 2016*). However, in luminal B BC, *WT1* expression showed no statistically significant association with the methylation level of the gene body, probably due to the fair negative correlation of cg13641903 methylation with *WT1* expression. A possible reason to explain the up-regulation of *WT1* in luminal B BC is that L1 transposon subfamilies that are up-regulated in ER+/HER- BC (*Yandim & Karakulah, 2019*) could regulate *WT1* gene expression (*Ramos et al., 2011*). Given that luminal B BC was characterized by ER+ and/or PR + as well as HER2-, hence, we supposed that *WT1* expression was influenced more by repetitive DNA in luminal B BC. This inspires scientists to put attention not only on the dysregulation at the gene level but also on the repetitive DNA.

To further understand the *WT1* regulatory mechanism, we predicted its downstream genes using GTRD. We confirmed that 5 genes, including *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *IGFL2*, were co-expressed with *WT1*. In agreement with our results, the high expressions of *WT1*, *COL11A1*, and *FGF5* in BC tumors have also been reported in previous literatures (*Huang, Wang & Yang, 2018*; *Li, Kong & Zou, 2017a*; *Malvia et al., 2019*). *WT1* overexpression could promote ERK1/2 phosphorylation and thus decreasing

E-cadherin expression and enhance EMT (epithelial-to-mesenchymal transition) (*Han et al., 2020*). Its potential downstream gene *COL11A1* was also EMT-related: *COL11A1* and *CK7* co-expression indicated that the cell is undergoing EMT process (*Garcia-Pravia et al., 2013*); *COL11A1* deficiency significantly induced the expression of E-cadherin (one of the epithelial markers) (*Zhang et al., 2018*). The induction of *GFAP* might be caused by the reduction of pERK (*Lind et al., 2006*). Therefore, we supposed that *WT1* might affect EMT by regulating *COL11A1* and *GFAP* expressions. *CD300LG* has diverse immunological functions and is capable of recognizing and interacting with extracellular lipids (*Borrego, 2013*; *Stoy et al., 2015*). *WT1* induces the generation of WT1-specific CD8+ T cells (*Lu et al., 2018*), suggesting that *WT1* and *CD300LG* have a functional dependency. *IGFL2* is a member of the insulin-like growth factor family, which plays a key role in cell energy metabolism, growth, and development, especially in prenatal growth (*Emtage et al., 2006*). *IGFL2* was up-regulated in the four types of BC in our result, which might contribute to tumor cell growth. Besides, *COL11A1* expression has been suggested as a promising marker for invasive breast lesions (*Freire et al., 2014*) and ovarian cancer and has a positive correlation with cisplatin treatment (*Rada et al., 2018*).

For *WT1*'s prognostic value, our results showed that the lower quartile *WT1* group exhibited better OS compared to the upper quartile *WT1* group based on the TCGA dataset, which was in agreement with a previous study (*Artibani et al., 2017*). Artibani et al. concluded that the poor prognosis for patients with high *WT1* expression may be caused by EMT. The expression of *COL11A1*, a downstream gene of *WT1*, has shown a correlation with the invasive capacity of BC cells. This may be one of the reasons for the poor prognosis of the upper quartile *WT1* group and the unfavorable 6-gene signature group. Additionally, *WT1* as a tumor-associated antigen stimulates the growth of WT1-specific CD8+ cytotoxic T cells (*Lehe et al., 2008*) which is associated with better BC outcomes (*Ziai et al., 2018*). However, the presence of regulatory T cells could inhibit the induction of anti-WT1-126 CD8+ CTL responses (*Lehe et al., 2008*). The depletion of CD4+CD25+ regulatory T cells is necessary for the generation of an effective WT1-specific cytotoxic response (*Asemissen et al., 2006*). In our immune infiltration analysis (Fig. 8), regulatory T cells were observed both in the upper/lower quartile *WT1* groups and the favorable/unfavorable signature groups, which might result in the inhibition of WT1-specific CD8+ T cell generation. CD8+ tumor-infiltrating lymphocytes demonstrated prognostic benefit only when present in combination with plasma cells (*Kroeger, Milne & Nelson, 2016*). The relatively higher plasma cell infiltrations in the lower quartile *WT1* group and the favorable signature group may, therefore, be a factor leading to the better outcomes of these groups. Finally, the protein-protein interaction network constructed using STITCH demonstrated an association between *WT1* and tamoxifen. *Han et al. (2008)* also found that *WT1* could regulate the resistance to antiestrogen, including tamoxifen.

CONCLUSION

In conclusion, we showed that *WT1* is highly expressed and hypermethylated in the four BC subtypes. All the *WT1* methylation sites exhibited hypermethylation and most of them were

located in the gene body region. Methylation levels of the TSS200 and 1stExon regions were negatively correlated with *WT1* expression in luminal A and luminal B BCs, while that of the gene body was positively associated with *WT1* expression in luminal A, basal-like, and HER2-enriched BCs. Furthermore, we found five dysregulated genes, including *COL11A1*, *GFAP*, *FGF5*, *CD300LG*, and *IGFL2*, in the four BC types that was predicted as *WT1*'s downstream genes. When we used the expression of the five genes and *WT1* as a signature, the group with low expression exhibited better OS than the high expression group.

Funding
The authors received no funding for this work.

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Chongyang Ren performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Xiaojiang Tang performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Haitao Lan conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability
Data is available at NCBI GEO: GSE20685 and at Xena.

Link to Xena: https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap%2FHiSeqV2&host=https%3A%2F%2Ftcga.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11377#supplemental-information.

REFERENCES

Akoglu H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* **18**:91–93 DOI 10.1016/j.tjem.2018.08.001.

Arechederra M, Daian F, Yim A, Bazai SK, Richelme S, Dono R, Saurin AJ, Habermann BH, Maina F. 2018. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nature Communications* **9**:3164 DOI 10.1038/s41467-018-05550-5.

Artibani M, Sims AH, Slight J, Aitken S, Thornburn A, Muir M, Brunton VG, Del-Pozo J, Morrison LR, Katz E, Hastie ND, Hohenstein P. 2017. WT1 expression in breast cancer disrupts the epithelial/mesenchymal balance of tumour cells and correlates with the metabolic response to docetaxel. *Scientific Reports* **7**:45255 DOI 10.1038/srep45255.

Asemissen AM, Keilholz U, Tenzer S, Muller M, Walter S, Stevanovic S, Schild H, Letsch A, Thiel E, Rammensee HG, Scheibenbogen C. 2006. Identification of a highly immunogenic HLA-A*01-binding T cell epitope of WT1. *Clinical Cancer Research* **12**:7476–7482 DOI 10.1158/1078-0432.CCR-06-1337.

Blagitko-Dorfs N, Schlosser P, Greve G, Pfeifer D, Meier R, Baude A, Brocks D, Plass C, Lubbert M. 2019. Combination treatment of acute myeloid leukemia cells with DNMT and HDAC inhibitors: predominant synergistic gene down-regulation associated with gene body demethylation. *Leukemia* **33**:945–956 DOI 10.1038/s41375-018-0293-8.

Borrego F. 2013. The CD300 molecules: an emerging family of regulators of the immune system. *Blood* **121**:1951–1960 DOI 10.1182/blood-2012-09-435057.

Chang D, Qian C, Li H, Feng H. 2019. Comprehensive analyses of DNA methylation and gene expression profiles of Kawasaki disease. *Journal of Cellular Biochemistry* **120**:13001–13011 DOI 10.1002/jcb.28571.

Croes L, Beyens M, Fransen E, Ibrahim J, Van den Berghe W, Suls A, Peeters M, Pauwels P, Van Camp G, Op de Beeck K. 2018. Large-scale analysis of DFNA5 methylation reveals its potential as biomarker for breast cancer. *Clinical Epigenetics* **10**:51 DOI 10.1186/s13148-018-0479-y.

De Almeida BP, Apolonio JD, Binnie A, Castelo-Branco P. 2019. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. *BMC Cancer* **19**:219 DOI 10.1186/s12885-019-5403-0.

Dawson MA, Kouzarides T. 2012. Cancer epigenetics: from mechanism to therapy. *Cell* **150**:12–27 DOI 10.1016/j.cell.2012.06.013.

Downs BM, Mercado-Rodriguez C, Cimino-Mathews A, Chen C, Yuan JP, Van Den Berg E, Cope LM, Schmitt F, Tse GM, Ali SZ, Meir-Levi D, Sood R, Li J, Richardson AL, Mosunjac MB, Rizzo M, Tulac S, Kocmond KJ, De Guzman T, Lai EW, Rhees B, Bates M, Wolff AC, Gabrielson E, Harvey SC, Umbricht CB, Visvanathan K, Fackler MJ, Sukumar S. 2019. DNA methylation markers for breast cancer detection in the developing world. *Clinical Cancer Research* **25**:6357–6367 DOI 10.1158/1078-0432.CCR-18-3277.

Emtage P, Vatta P, Arterburn M, Muller MW, Park E, Boyle B, Hazell S, Poli-zotto R, Funk WD, Tang YT. 2006. IGFL: a secreted family with conserved cysteine residues and similarities to the IGF superfamily. *Genomics* **88**:513–520 DOI 10.1016/j.ygeno.2006.05.012.

Feinberg AP, Tycko B. 2004. The history of cancer epigenetics. *Nature Reviews Cancer* **4**:143–153 DOI 10.1038/nrc1279.

Flanagan JM, Munoz-Alegre M, Henderson S, Tang T, Sun P, Johnson N, Fletcher O, Dos Santos Silva I, Peto J, Boshoff C, Narod S, Petronis A. 2009. Gene-body

hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Human Molecular Genetics* **18**:1332–1342 DOI 10.1093/hmg/ddp033.

Freire J, Dominguez-Hormaetxe S, Pereda S, De Juan A, Vega A, Simon L, Gomez-Roman J. 2014. Collagen, type XI, alpha 1: an accurate marker for differential diagnosis of breast carcinoma invasiveness in core needle biopsies. *Indian Journal of Pathology: Research and Practice* **210**:879–884 DOI 10.1016/j.prp.2014.07.012.

Garcia-Pravia C, Galvan JA, Gutierrez-Corral N, Solar-Garcia L, Garcia-Perez E, Garcia-Ocana M, Del Amo-Iribarren J, Menendez-Rodriguez P, Garcia-Garcia J, De Los Toyos JR, Simon-Buela L, Barneo L. 2013. Overexpression of COL11A1 by cancer-associated fibroblasts: clinical relevance of a stromal marker in pancreatic cancer. *PLOS ONE* **8**:e78327 DOI 10.1371/journal.pone.0078327.

Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196**:261–282 DOI 10.1016/0022-2836(87)90689-9.

Han F, Zhang X, Liu X, Su H, Kong C, Fang Z, Yang L, Zhuang M, Zhang Y, Liu Y, Li Z, Lv H. 2017. Comparative analysis of genome wide DNA methylation profiles for the genic male sterile cabbage line 01-20S and its maintainer line. *Genes* **8(6)**:159 DOI 10.3390/genes8060159.

Han Y, Song C, Zhang T, Zhou Q, Zhang X, Wang J, Xu B, Zhang X, Liu X, Ying X. 2020. Wilms' tumor 1 (WT1) promotes ovarian cancer progression by regulating E-cadherin and ERK1/2 signaling. *Cell Cycle* **19**:2662–2675 DOI 10.1080/15384101.2020.1817666.

Han Y, Yang L, Suarez-Saiz F, San-Marina S, Cui J, Minden MD. 2008. Wilms' tumor 1 suppressor gene mediates antiestrogen resistance via down-regulation of estrogen receptor-alpha expression in breast cancer cells. *Molecular Cancer Research* **6**:1347–1355 DOI 10.1158/1541-7786.MCR-07-2179.

Huang Y, Wang H, Yang Y. 2018. Expression of Fibroblast Growth Factor 5 (FGF5) and its influence on survival of breast cancer patients. *Medical Science Monitor* **24**:3524–3530 DOI 10.12659/MSM.907798.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyan S, Feinberg AP. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics* **41**:178–186 DOI 10.1038/ng.298.

Jin W, Li QZ, Zuo YC, Cao YN, Zhang LQ, Hou R, Su WX. 2019. Relationship between DNA methylation in key region and the differential expressions of genes in human breast tumor tissue. *DNA and Cell Biology* **38**:49–62 DOI 10.1089/dna.2018.4276.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**:484–492 DOI 10.1038/nrg3230.

Karemaker ID, Vermeulen M. 2018. Single-Cell DNA methylation profiling: technologies and biological applications. *Trends in Biotechnology* **36**:952–965 DOI 10.1016/j.tibtech.2018.04.002.

Kim D, Shivakumar M, Han S, Sinclair MS, Lee YJ, Zheng Y, Olopade OI, Kim D, Lee Y. 2018. Population-dependent intron retention and DNA methylation in breast cancer. *Molecular Cancer Research* **16**:461–469 DOI 10.1158/1541-7786.MCR-17-0227.

Kim J, So S, Lee HJ, Park JC, Kim JJ, Lee H. 2013. DigSee: disease gene search engine with evidence sentences (version cancer). *Nucleic Acids Research* **41**:W510–W517 DOI 10.1093/nar/gkt531.

Kim MS, Lee J, Oh T, Moon Y, Chang E, Seo KS, Hoehn BD, An S, Lee JH. 2012. Genome-wide identification of OTP gene as a novel methylation marker of breast cancer. *Oncology Reports* **27**:1681–1688 DOI 10.3892/or.2012.1691.

Kroeger DR, Milne K, Nelson BH. 2016. Tumor-infiltrating plasma cells are associated with tertiary lymphoid structures, cytolytic T-Cell responses, and superior prognosis in ovarian cancer. *Clinical Cancer Research* **22**:3005–3015 DOI 10.1158/1078-0432.CCR-15-2762.

Leadem BR, Kagiampakis I, Wilson C, Cheung TK, Arnott D, Trojer P, Classon M, Easwaran H, Baylin SB. 2018. A KDM5 inhibitor increases global H3K4 trimethylation occupancy and enhances the biological efficacy of 5-Aza-2'-Deoxycytidine. *Cancer Research* **78**:1127–1139 DOI 10.1158/0008-5472.CAN-17-1453.

Lehe C, Ghebeh H, Al-Sulaiman A, Qudaihi GAl, Al-Hussein K, Almohareb F, Chaudhri N, Alsharif F, Al-Zahrani H, Tbakhi A, Aljurf M, Dermime S. 2008. The Wilms' tumor antigen is a novel target for human CD4+ regulatory T cells: implications for immunotherapy. *Cancer Research* **68**:6350–6359 DOI 10.1158/0008-5472.CAN-08-0050.

Li GS, Kong GY, Zou Y. 2017a. Protective role of LRRC3B in preventing breast cancer metastasis and recurrence post-bupivacaine. *Oncology Letters* **14**:5013–5017 DOI 10.3892/ol.2017.6773.

Li S, Kim E, Wong EM, Joo JE, Nguyen TL, Stone J, Song YM, Flander LB, Saffery R, Giles GG, Southey MC, Sung J, Hopper JL. 2017b. Twin birth changes DNA methylation of subsequent siblings. *Scientific Reports* **7**:8463 DOI 10.1038/s41598-017-08595-6.

Li SY, Wu HC, Mai HF, Zhen JX, Li GS, Chen SJ. 2019. Microarray-based analysis of whole-genome DNA methylation profiling in early detection of breast cancer. *Journal of Cellular Biochemistry* **120**:658–670 DOI 10.1002/jcb.27423.

Lian ZQ, Wang Q, Li WP, Zhang AQ, Wu L. 2012. Screening of significantly hypermethylated genes in breast cancer using microarray-based methylated-CpG island recovery assay and identification of their expression levels. *International Journal of Oncology* **41**:629–638 DOI 10.3892/ijo.2012.1464.

Lind CR, Gray CW, Pearson AG, Cameron RE, O'Carroll SJ, Narayan PJ, Lim J, Dragunow M. 2006. The mitogen-activated/extracellular signal-regulated kinase kinase 1/2 inhibitor U0126 induces glial fibrillary acidic protein expression and reduces the proliferation and migration of C6 glioma cells. *Neuroscience* **141**:1925–1933 DOI 10.1016/j.neuroscience.2006.05.038.

Liu Q, Kulak MV, Borcherding N, Maina PK, Zhang W, Weigel RJ, Qi HH. 2018.
A novel HER2 gene body enhancer contributes to HER2 expression. *Oncogene*
37:687–694 DOI 10.1038/onc.2017.382.

Liu Y, Sun J, Zhao M. 2017. ONGene: a literature-based database for human oncogenes.
Journal of Genetics and Genomics **44**:119–121 DOI 10.1016/j.jgg.2016.12.004.

**Loeb DM, Evron E, Patel CB, Sharma PM, Niranjan B, Buluwela L, Weitzman SA,
Korz D, Sukumar S. 2001.** Wilms' tumor suppressor gene (WT1) is expressed
in primary breast tumors despite tumor-specific promoter methylation. *Cancer
Research* **61**:921–925.

Lu X, Liu J, Cui P, Liu T, Piao C, Xu X, Zhang Q, Xiao M, Liu X, Wang Y, Yang L. 2018.
Co-inhibition of TIGIT, PD1, and Tim3 reverses dysfunction of Wilms tumor
protein-1 (WT1)-specific CD8+ T lymphocytes after dendritic cell vaccination in
gastric cancer. *American Journal of Cancer Research* **8**:1564–1575.

**Malvia S, Bagadi SAR, Pradhan D, Chintamani C, Bhatnagar A, Arora D, Sarin R,
Saxena S. 2019.** Study of gene expression profiles of breast cancers in indian women.
Scientific Reports **9**:10018 DOI 10.1038/s41598-019-46261-1.

McGregor RJ, Chau YY, Kendall TJ, Artibani M, Hastie N, Hadoke PWF. 2018. WT1
expression in vessels varies with histopathological grade in tumour-bearing and con-
trol tissue from patients with breast cancer. *British Journal of Cancer* **119**:1508–1517
DOI 10.1038/s41416-018-0317-1.

**Miyoshi Y, Ando A, Egawa C, Taguchi T, Tamaki Y, Tamaki H, Sugiyama H, Noguchi S.
2002.** High expression of Wilms' tumor suppressor gene predicts poor prognosis in
breast cancer patients. *Clinical Cancer Research* **8**:1167–1171.

**Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M,
Alizadeh AA. 2015.** Robust enumeration of cell subsets from tissue expression
profiles. *Nature Methods* **12**:453–457 DOI 10.1038/nmeth.3337.

**Peiffer DS, Wyatt D, Zlobin A, Piracha A, Ng J, Dingwall AK, Albain KS, Osipo
C. 2019.** DAXX suppresses tumor-initiating cells in estrogen receptor-positive
breast cancer following endocrine therapy. *Cancer Research* **79**:4965–4977
DOI 10.1158/0008-5472.CAN-19-1110.

**Prajzendanc K, Domagala P, Hybiak J, Rys J, Huzarski T, Szwiec M, Tomiczek-
Szwiec J, Redelbach W, Sejda A, Gronwald J, Kluz T, Wisniowski R, Cybulski C,
Lukomska A, Bialkowska K, Sukiennicki G, Kulczycka K, Narod SA, Wojdacz TK,
Lubinski J, Jakubowska A. 2020.** BRCA1 promoter methylation in peripheral blood
is associated with the risk of triple-negative breast cancer. *International Journal of
Cancer* **146**:1293–1298 DOI 10.1002/ijc.32655.

**Pu M, Messer K, Davies SR, Vickery TL, Pittman E, Parker BA, Ellis MJ, Flatt SW,
Marinac CR, Nelson SH, Mardis ER, Pierce JP, Natarajan L. 2020.** Research-based
PAM50 signature and long-term breast cancer survival. *Breast Cancer Research and
Treatment* **179**:197–206 DOI 10.1007/s10549-019-05446-y.

Rada M, Nallanthighal S, Cha J, Ryan K, Sage J, Eldred C, Ullo M, Orsulic S, Cheon DJ. 2018. Inhibitor of apoptosis proteins (IAPs) mediate collagen type XI alpha 1-driven cisplatin resistance in ovarian cancer. *Oncogene* **37**:4809–4820 DOI 10.1038/s41388-018-0297-x.

Ramos KS, Montoya-Durango DE, Teneng I, Nanez A, Stribinskis V. 2011. Epigenetic control of embryonic renal cell differentiation by L1 retrotransposon. *Birth Defects Research Part A: Clinical and Molecular Teratology* **91**:693–702 DOI 10.1002/bdra.20786.

Renner M, Wolf T, Meyer H, Hartmann W, Penzel R, Ulrich A, Lehner B, Hovestadt V, Czwan E, Egerer G, Schmitt T, Alldinger I, Renker EK, Ehemann V, Eils R, Wardelmann E, Buttner R, Lichter P, Brors B, Schirmacher P, Mechtersheimer G. 2013. Integrative DNA methylation and gene expression analysis in high-grade soft tissue sarcomas. *Genome Biology* **14**:r137 DOI 10.1186/gb-2013-14-12-r137.

Rodger EJ, Chatterjee A, Stockwell PA, Eccles MR. 2019. Characterisation of DNA methylation changes in EBF3 and TBC1D16 associated with tumour progression and metastasis in multiple cancer types. *Clinical Epigenetics* **11**:114 DOI 10.1186/s13148-019-0710-5.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. 2011. Validation of a DNA methylation microarray for 450, 000 CpG sites in the human genome. *Epigenetics* **6**:692–702 DOI 10.4161/epi.6.6.16196.

Sasidharan Nair V, Salhat HEl, Taha RZ, John A, Ali BR, Elkord E. 2018. DNA methylation and repressive H3K9 and H3K27 trimethylation in the promoter regions of PD-1, CTLA-4, TIM-3, LAG-3, TIGIT, and PD-L1 genes in human primary breast cancer. *Clinical Epigenetics* **10**:78 DOI 10.1186/s13148-018-0512-1.

Shenker NS, Flower KJ, Wilhelm-Benartzi CS, Dai W, Bell E, Gore E, Bahrawy MEl, Weaver G, Brown R, Flanagan JM. 2015. Transcriptional implications of intragenic DNA methylation in the oestrogen receptor alpha gene in breast cancer cells and tissues. *BMC Cancer* **15**:337 DOI 10.1186/s12885-015-1335-5.

Song MA, Brasky TM, Marian C, Weng DY, Taslim C, Dumitrescu RG, Llanos AA, Freudenheim JL, Shields PG. 2015. Racial differences in genome-wide methylation profiling and gene expression in breast tissues from healthy women. *Epigenetics* **10**:1177–1187 DOI 10.1080/15592294.2015.1121362.

Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, Hilmarsdottir H, Olafsdottir E, Tryggvadottir L, Jonasson JG, Eyfjord J, Esteller M. 2015. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Molecular Oncology* **9**:555–568 DOI 10.1016/j.molonc.2014.10.012.

Stoy J, Kampmann U, Mengel A, Magnusson NE, Jessen N, Grarup N, Rungby J, Stodkilde-Jorgensen H, Brandslund I, Christensen C, Hansen T, Pedersen O, Moller N. 2015. Reduced CD300LG mRNA tissue expression, increased intramyocellular lipid content and impaired glucose metabolism in healthy male carriers of Arg82Cys in CD300LG: a novel genometabolic cross-link between CD300LG and common metabolic phenotypes. *BMJ Open Diabetes Research & Care* **3**:e000095 DOI 10.1136/bmjdrc-2015-000095.

Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. 2016. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Research* **44**:D380–D384 DOI 10.1093/nar/gkv1277.

Taslim C, Lin S, Huang K, Huang TH. 2012. Integrative genome-wide chromatin signature analysis using finite mixture models. *BMC Genomics* **13**(**Suppl 6**):S3 DOI 10.1186/1471-2164-13-S6-S3.

Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Yang THOu, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS. 2018. The immune landscape of cancer. *Immunity* **48**:812–830 DOI 10.1016/j.immuni.2018.03.023.

Visone R, Bacalini MG, Di Franco S, Ferracin M, Colorito ML, Pagotto S, Laprovitera N, Licastro D, Marco MDi, Scavo E, Bassi C, Saccenti E, Nicotra A, Grzes M, Garagnani P, De Laurenzi V, Valeri N, Mariani-Costantini R, Negrini M, Stassi G, Veronese A. 2019. DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. *Epigenomics* **11**:587–604 DOI 10.2217/epi-2018-0153.

Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. 2014. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biology* **15**:R37 DOI 10.1186/gb-2014-15-2-r37.

Wang YW, Ma X, Zhang YA, Wang MJ, Yatabe Y, Lam S, Girard L, Chen JY, Gazdar AF. 2016. ITPKA gene body methylation regulates gene expression and serves as an early diagnostic marker in lung and other cancers. *Journal of Thoracic Oncology* **11**:1469–1481 DOI 10.1016/j.jtho.2016.05.010.

Windhorst S, Song K, Gazdar AF. 2017. Inositol-1, 4, 5-trisphosphate 3-kinase-A (ITPKA) is frequently over-expressed and functions as an oncogene in several tumor types. *Biochemical Pharmacology* **137**:1–9 DOI 10.1016/j.bcp.2017.03.023.

Xu D, Du M, Zhang J, Xiong P, Li W, Zhang H, Xiong W, Liu F, Liu J. 2018. DNMT1 mediated promoter methylation of GNAO1 in hepatoma carcinoma cells. *Gene* **665**:67–73 DOI 10.1016/j.gene.2018.04.080.

Yandim C, Karakulah G. 2019. Dysregulated expression of repetitive DNA in ER+/HER2- breast cancer. *Cancer Genetics* **239**:36–45 DOI 10.1016/j.cancergen.2019.09.002.

Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**:577–590 DOI 10.1016/j.ccr.2014.07.028.

Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. 2019. GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Research* **47**:D100–D105 DOI 10.1093/nar/gky1128.

Zhang B, Zhang C, Yang X, Chen Y, Zhang H, Liu J, Wu Q. 2018. Cytoplasmic collagen XIalphaI as a prognostic biomarker in esophageal squamous cell carcinoma. *Cancer Biology & Therapy* **19**:364–372 DOI 10.1080/15384047.2018.1423915.

Ziai J, Gilbert HN, Foreman O, Eastham-Anderson J, Chu F, Huseni M, Kim JM. 2018. CD8+ T cell infiltration in breast and colon cancer: a histologic and statistical analysis. *PLOS ONE* **13**:e0190158 DOI 10.1371/journal.pone.0190158.