



A High Epigenetic Risk Score Shapes the Non-Inflamed Tumor Microenvironment in Breast Cancer

Dong Zhang^{1,2†}, Yingnan Wang^{1,2†} and Qifeng Yang^{1,3*}

¹Department of Breast Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, ²Department of Clinical Medicine, Cheeloo College of Medicine, Shandong University, Jinan, China, ³Department of Pathology Tissue Bank, Qilu Hospital, Shandong University, Jinan, China

Background: Epigenetic dysregulation via aberrant DNA methylation has gradually become recognized as an efficacious signature for predicting tumor prognosis and response to therapeutic targets. However, reliable DNA methylation biomarkers describing tumorigenesis remain to be comprehensively explored regarding their prognostic and therapeutic potential in breast cancer (BC).

Methods: Whole-genome methylation datasets integrated from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database were profiled ($n = 1,268$). A three-stage selection procedure (discovery, training, and external validation) was utilized to screen out the prominent biomarkers and establish a robust risk score from more than 300,000 CpG sites after quality control, rigorous filtering, and reducing dimension. Moreover, gene set enrichment analyses guided us to systematically correlate this epigenetic risk score with immunological characteristics, including immunomodulators, anti-cancer immunity cycle, immune checkpoints, tumor-infiltrating immune cells and a series of signatures upon modulating components within BC tumor microenvironment (TME). Multi-omics data analyses were performed to decipher specific genomic alterations in low- and high-risk patients. Additionally, we also analyzed the role of risk score in predicting response to several treatment options.

Results: A 10-CpG-based prognostic signature which could significantly and independently categorize BC patients into distinct prognoses was established and sufficiently validated. And we hypothesize that this signature designs a non-inflamed TME in BC based on the evidence that the derived risk score is negatively correlated with tumor-associated infiltrating immune cells, anti-cancer immunity cycle, immune checkpoints, immune cytolytic activity, T cell inflamed score, immunophenoscore, and the vast majority of immunomodulators. The identified high-risk patients were

Abbreviations: AUC, area under curve; BC, breast cancer; CI, confidence interval; c-index, concordance index; CpG, 5'-cytosine-phosphate-guanine-3'; DCA, decision curve analysis; ER, estrogen receptor; ESTIMATE, estimation of stromal and immune cells in malignant tumor tissues using expression data; FDR, false discovery rate; FISH, fluorescence *in situ* hybridization; GEO, gene expression omnibus; GSEA, gene set enrichment analysis; KNN, k-nearest neighbors; Her-2, human epidermal growth factor receptor-2; HR, hazard ratio; IHC, immunohistochemical; LASSO, least absolute shrinkage and selection operator; MCP-counter, microenvironment cell populations-counter; MSigDB, molecular signatures database; NES, normalized enrichment score; OS, overall survival; PR, progesterone receptor; ROC, receiver operator characteristic; SNP, single-nucleotide polymorphisms; TCGA, the cancer genome atlas; TPM, transcripts per kilobase million.

OPEN ACCESS

Edited by:

Guohui Wan,
Sun Yat-Sen University, China

Reviewed by:

Dong Ren,
The First Affiliated Hospital of Sun
Yat-Sen University, China
Jing Li,
Fujian Medical University, China

*Correspondence:

Qifeng Yang
qifengy_sdu@163.com

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Molecular Diagnostics and
Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 02 March 2021

Accepted: 14 July 2021

Published: 26 July 2021

Citation:

Zhang D, Wang Y and Yang Q (2021) A
High Epigenetic Risk Score Shapes the
Non-Inflamed Tumor
Microenvironment in Breast Cancer.
Front. Mol. Biosci. 8:675198.
doi: 10.3389/fmolb.2021.675198

characterized by upregulation of immune inhibited oncogenic pathways, higher TP53 mutation and copy number burden, but lower response to cancer immunotherapy and chemotherapy.

Conclusion: Our work highlights the complementary roles of 10-CpG-based signature in estimating overall survival in BC patients, shedding new light on investigating failed events concerning immunotherapy at present.

Keywords: breast cancer, prognosis, DNA methylation, tumor microenvironment, therapeutic target

INTRODUCTION

Breast cancer (BC) ranks third among the most common malignancies and is the leading cause of cancer-related death in females (Lin et al., 2019; Li et al., 2019b). Currently, the improvement of mammographic screening has made great progress in early-stage diagnosis of BC. Moreover, there are a series of systematic treatments for BC, including surgical resection, chemotherapy, radiotherapy, endocrine therapy, and alternative molecule-targeted therapy (e.g., trastuzumab and pertuzumab) (Akram et al., 2017). However, postoperative local or distant recurrence rate remains high, even for patients who have received conventional therapies in the early stage, causing a pessimistic mortality rate within BC patients at present (Spronk et al., 2018; Chen et al., 2019). This could be attributed to the restricted and incomprehensive understanding of BC heterogeneity concerning carcinogenesis, invasiveness, progression, and metastasis (Tazaki et al., 2013). Molecular characteristics that are reliably related to BC prognosis and patient survival would have tremendous value in guiding clinical management of BC. Hence, a deeper understanding of BC functional pathways, as well as the development of novel crucial biomarkers with biological background, for early diagnosis and prognostic prediction in BC patients, is urgently needed.

Although previous studies at different omics levels, such as somatic mutations, gene expression, non-coding RNA, and copy number variations, have revealed numerous promising biomarkers relevant to BC carcinogenesis (Zhang et al., 2018a; Fan et al., 2019; He et al., 2019; Li et al., 2019a), the contribution of epigenetic alterations including DNA methylation in human disease, particularly in cancer, has also been widely recognized (Khaled and Bidet, 2019; Luo et al., 2020; Yang et al., 2020). From the point of view of mechanism, DNA methylation, as inherently reversible changes to repress the transcriptional activities and interact with various negative and positive feedback processes, plays decisive roles in both physiological and pathological regulation of cellular fate (Drake and Søreide, 2019). Accumulated evidence highlights the multifaceted DNA methylation of 5'-cytosine-phosphate-guanine-3' (CpG) sites in various cancer hallmarks, including modulating energy metabolism and angiogenesis, sustaining proliferation signaling, epithelial-mesenchymal transition, invasiveness, and metastasis, mainly via promoting the activation of oncogenes and silencing of tumor suppressor genes (Pasculli et al., 2018; Drake and Søreide, 2019). The tumor microenvironment, intra-tumoral

cell typing, and subsequent response to immunotherapy could also be evaluated and characterized by DNA methylation profiling (Jeschke et al., 2015; Jeschke et al., 2017; Calle-Fabregat et al., 2020). More importantly, DNA methylation offers feasible clues to assist the early detection and prognosis of different cancers, using the early onset of global alterations within DNA methylation profiling in cancer initiation and significantly distinct methylation patterns between tumor and normal tissues (Szyf, 2012).

Recent advances of high-throughput technology, combined with the availability of public, large-scale sequencing datasets, have opened a new area for defining the genome-wide landscape of BC and provided opportunities but daunting challenges to identify potentially reliable BC biomarkers (Jiang et al., 2019; Xuchao et al., 2019). Preliminary investigation of profiling arrays has proposed several methylation-based signatures for survival stratification in patients with BC (Du et al., 2019; Tao et al., 2019). Unfortunately, none has yet been incorporated in clinical practice owing to issues such as model overfitting on a single cohort and insufficient external validation. Moreover, the aforementioned analyses have not elucidated the complicated regulatory network that governed BC biological processes, and clinical characteristics were not included in the predictive models for further assessment.

In the current study, we identified CpG sites that are specifically and significantly correlated with BC prognosis and then developed a multi-CpG classifier by mining DNA methylation profiles from the Cancer Genome Atlas (TCGA) project. With sufficient validation of two independent sample sets, we proved the stability and reliability of our established model. In addition, pivotal biological processes underlying the prognostic signature were revealed via a series of bioinformatics analyses, and the correlations of CpG-based signature with tumor microenvironment (TME) were also comprehensively evaluated. To further leverage the complementary value of molecular and clinical characteristics, we integrated the CpG-based signature and clinicopathological risk factors to build a composite nomogram, which improved the risk stratification of BC patients.

MATERIALS AND METHODS

Data Acquisition and Pre-Processing

Level 3 DNA methylation profiles using the Illumina Infinium HumanMethylation450 BeadChips Assay, including 785 BC patients, were downloaded and integrated from TCGA website

(<https://portal.gdc.cancer.gov>) via “TCGA-Assembler 2” R package (<https://github.com/compgenome365/TCGA-Assembler-2>) (Tomczak et al., 2015; Wei et al., 2018), and the genomic manifest of each CpG site was annotated using “IlluminaHumanMethylation450kanno.ilmn12.hg19” R package (version 0.6.0) (Aryee et al., 2010). For each CpG locus, the methylation level was represented by the β value ($\beta = M/(M+U)$), which ranges from 0 to 1, corresponding to unmethylated and fully methylated. Low quality probes were excluded if they met the following criteria: 1) failed detection in more than 25% samples; 2) not uniquely mapped to the human reference genome; 3) located on unstable genomic sites, including single-nucleotide polymorphisms (SNP) sites and Y chromosome; or 4) missed annotations to parent genes. Furthermore, the remaining CpG sites with unavailable methylation levels (NAs) were imputed based on k-nearest neighbors (KNN) algorithm via “impute” R package (version 1.56.0).

Level 4 RNA sequencing data (FPKM normalized) based on the Illumina HiSeq RNA-seq platform of BC samples were obtained from GDC data portal of TCGA website via “TCGAbiolinks” R package (version 2.10.5) (Colaprico et al., 2016). The ensemble IDs were annotated to gene symbols using “gencode.v22.annotation.gtf” file (www.gencodegenes.org/human/releases.html) (Paik and Hancock, 2012). And the FPKM values were further transformed into transcripts per kilobase million (TPM) values, which are more reliable and recommended for inter-group comparisons of expressive abundance (Wagner et al., 2012). Low-abundance profiles were eliminated. In addition, if multiple ensemble IDs correspond to the same gene symbol, the one with the highest mean value was extracted as the expression level of parent gene.

Subsequently, corresponding clinicopathological features of BC patients in TCGA cohort, including the gender, age, histological type, history of neoadjuvant treatment, regional lymph nodes involvement (H&E stain), TNM stage, pathological stage, estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor-2 (Her-2) status, and follow-up data, were also downloaded via “TCGAbiolinks” package and used for subsequent analysis. Therein, the Her-2 statuses of BC samples were firstly quantified by immunohistochemical (IHC) results, and we took an account of the results from fluorescence *in situ* hybridization (FISH) to determine the Her-2 status when the IHC diagnoses were missed or uncertain. The present study fully complies with the TCGA publication guidelines (Network, 2012).

The integrated methylation matrix files (GSE75067 and GSE72308) based on the platform GPL13534 were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo>). Among these two datasets, the missing values in methylation profiles were also imputed via KNN algorithm. And the corresponding clinical data were obtained either by directly downloading the attached files from the item page in the GEO website (GSE75067) or retrieving and manually organizing if available (GSE72308).

Study Population and Eligibility Criteria

The backbone of our study was comprised of three stages: discovery stage, training stage, and validation stage. In the discovery stage, we used paired samples (tumor/normal) from BC patients in the TCGA-BRCA project as discovery set ($n = 91$ pairs), which was used to identify the differential methylation CpG sites. In the training stage, a total of 654 BC patients from the entire TCGA cohort were enrolled as the training set with the following inclusion criteria: 1) female patients with a definitive diagnosis of BC, 2) without any neoadjuvant therapy, 3) with corresponding follow-up censoring data, 4) the follow-up time was no less than one month, and 5) with corresponding DNA methylation profiles. Furthermore, independent validations were conducted using two external GEO datasets in the validation stage. For validation set 1, the clinical and DNA methylation data were obtained from accession number GSE75067 (Holm et al., 2016). Only 180 BC patients were included in our analysis, after the removal of repetitions and samples with missing survival information. For validation set 2 (accession number: GSE72308), a subset of 237 BC patients (cohort 1 and cohort 2) treated with adjuvant therapies from 1995 to 2009 was included and the patients who had a history of neoadjuvant epirubicin monotherapy (TOP cohort) were excluded, as described by Jeschke et al. (2017).

Identification of Differentially Methylated CpG Sites

On the basis of the discovery set of 91 BC patients who had both tumor and adjacent-normal tissues, genome-wide differential methylation CpG sites were identified using paired Student's *t*-test. *p*-values were adjusted by Benjamini & Hochberg (BH) correction; the methylation differences were characterized by absolute differential methylation calculated as mean (β_{tumor}) – mean (β_{normal}) and adjusted *p*-values. An absolute differential methylation of >0.4 combined with adjusted $p < 0.05$ was set as the significance threshold. Additionally, the diagnostic values of specific CpG sites were evaluated by receiver operator characteristic (ROC) curve.

Construction and Validation of a CpG-Based Prognostic Model

In the training stage, the initial assessment of prognostic potential of each differential methylation CpG site was conducted using univariate Cox proportional hazard regression analysis, and those with *p*-values < 0.05 were selected for further analyses. Next, for solving the multicollinearity problem of highly correlated variables, we implemented the Cox regression model, with least absolute shrinkage and selection operator (LASSO) penalty, to reduce dimension (Friedman et al., 2010; Simon et al., 2011). Based on prognostic BC-specific CpG sites, which were significant in the univariate Cox regression analysis, a sub-selection of key biomarkers was determined by shrinkage of the regression coefficients via using a penalized weight to corresponding size. Finally, only the features with nonzero regression coefficient were extracted as the candidates for

constructing prognostic signature. In our LASSO analysis, the “lambda.min” criteria and 10-fold cross validation were utilized to conduct penalty parameter tuning. Subsequently, we computed an epigenetic risk score for each BC patient using the following regression equation: $\text{risk score} = \sum_{j=1}^n \text{Coe}f_j \beta_j$, with $\text{Coe}f_j$ indicating the coefficients derived from multivariate Cox regression model and β_j representing the methylation level (β value) of each CpG site. And the predictive performance of CpG-based risk score was assessed by time-dependent ROC curves with area under curve (AUC) values calculated (Blanche et al., 2013). Then we classified the BC patients in the training set into low- and high-risk subgroups based on the optimal cutoff value of risk scores computed using regression formula, which represented the point at which the Youden index (specificity + sensitivity - 1) reached a maximum value in 5-years ROC curve. Furthermore, the Kaplan-Meier survival analysis and log-rank test were applied to verify the classification performance of the model.

Independence of CpG-Based Signature From Other Clinicopathological Parameters in TCGA

To further validate whether the prognostic value of established CpG-based signature is independent of traditional clinicopathological parameters (including age, histological subtype, regional lymph nodes metastasis, T stage, N stage, pathological stage, ER, PR, and Her-2 status) for BC patients, stratification Cox analyses were conducted in various stratified cohorts. Additionally, among 654 BC patients with survival information in the training set, 537 BC patients with relatively complete clinical information, including age, histological type, regional lymph nodes metastasis, T stage, N stage, pathological stage, ER, PR, and Her-2 status, were subjected to subsequent analyses. Univariate followed by multivariate Cox regression analyses were performed.

Construction and Evaluation of a Predictive Nomogram

The independent clinicopathological covariates, identified by multivariate Cox regression analysis, were taken into consideration to assemble a nomogram, providing clinicians with a clinically relevant quantitative approach for predicting individualized survival probability in BC patients (Harrell et al., 1996; Zhang and Kattan, 2017). Additionally, the calibration plots were graphically depicted to evaluate the consistency between the predicted probability of derived nomogram and actual situation (Balachandran et al., 2015). And Harrell’s concordance index (c-index) was utilized to estimate the discrimination ability of the nomogram. Moreover, decision curve analysis (DCA) was performed to explore the clinical usefulness of the nomogram, further providing straightforward information concerning whether clinical decision-making based on a predictive model will do more good than harm, in contrast to abstract statistical concepts (Fitzgerald et al., 2015; Vickers and Elkin, 2016; Zhang et al., 2018d).

Gene Set Enrichment Analysis of CpG-Based Signature

In order to identify crucial biological mechanisms related to the final CpG-based signature in BC carcinogenesis, we implemented GSEA analysis (version 4.0.2, <http://software.broadinstitute.org/gsea/downloads.jsp>) using the adjusted RNA-sequencing profiles (TPM normalization) for all transcripts (Subramanian et al., 2005; Reimand et al., 2019). Annotated gene set files were downloaded from Molecular Signatures Databases (MSigDB, <http://software.broadinstitute.org/gsea/msigdb>), and the “c2.cp.v7.0.symbols.gmt” were selected to perform the quantification of pathway activity (Liberzon et al., 2011; Liberzon et al., 2015). Enrichment p -values were evaluated based on 1,000 permutations and subsequently adjusted by BH correction for multiple testing to control FDR values. Therein, nominal p -value < 0.05 and FDR < 0.25 were set as significance threshold.

Comprehensive Evaluation of TME in BC Patients

To decipher the role of established signature in modifying TME in BC patients, we comprehensively analyzed the association between the risk score and immunological characteristics within TME with respect to the below aspects. The immune score, stromal score, and tumor purity in BC samples were calculated by applying the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm designed by Yoshihara et al. (2013). Thereafter, we applied the Microenvironment Cell Populations-counter (MCP-counter) algorithm, an estimator for robust quantification of absolute abundance of two stromal and eight immune cell populations by gene expression profiles, to perform the immune infiltration estimation in BC tissues (Becht et al., 2016). The infiltration levels of immune cells were derived using “MCPcounter” R package (version 1.1.0). To avoid the calculation error due to distinct inner algorithm, additional independent algorithms, including the Cibersort (with LM22 signature) and single sample Gene Set Enrichment Analysis (ssGSEA) methods, were utilized to quantify the infiltration levels of immune cells within TME (Chen et al., 2018; Zhang et al., 2018b). A list of 24 immune inhibitory checkpoint molecules, 122 immunomodulators and effector genes of particular tumor-infiltrating immune cells were collected from hand-curated screenings of literature (Auslander et al., 2018; Hu et al., 2021). Moreover, based on the ssGSEA algorithm, we profiled the enrichment scores of seven steps that composed the “anti-cancer immunity cycle,” including release of cancer cell antigens (Step 1), cancer antigen presentation (Step 2), priming and activation (Step 3), trafficking of immune cells to tumors (Step 4), infiltration of immune cells into tumors (Step 5), recognition of cancer cells by T cells (Step 6), and killing of cancer cells (Step 7) (Xu et al., 2018).

Furthermore, immune cytolytic score (CLS), defined as the log-average (geometric mean) of PRF1 and GZMA in TPM expression values, was also estimated to represent in silico measurement of immune infiltration, as provided by Rooney et al. (2015). Pan-cancer T cell inflamed score (TIS), as a weighted

linear combination of scores from eighteen genes, which defines the pre-existing cancer immunogenicity and predicts response to anti-tumor immunotherapy, was also computed for each BC patient (Ayers et al., 2017). The stemness indices (mRNAsi) for BC patients were derived from the supplementary table by Tathiane et al., who measured the oncogenic dedifferentiation and were associated with Immune microenvironment contents and PD-L1 levels (Malta et al., 2018). Immunophenoscore (IPS), derived from a panel of immune-related genes belonging to four categories, namely effector genes, immunosuppressive genes, MHC-related molecules, and immunomodulators, represents the cancer antigenomes and immunogenicity (Charoentong et al., 2017). The IPSs of BC patients were obtained from The Cancer Immunome Atlas (TCIA) (<https://tcia.at/home>). Additionally, Vésteinn Thorsson and his colleagues identified and characterized six immune subtypes that encompass multiple tumor types, with distinct intratumoral immune states and immune response patterns impacting prognosis (Thorsson et al., 2018). In our present work, we also classified the BC samples into different immune subtypes using the ImmuneSubtypeClassifier R package (<https://github.com/Gibbsdavidl/ImmuneSubtypeClassifier>).

External Validation of CpG-Based Signature

To evaluate the robustness and practical application of the CpG-based signature in predicting the OS probability for BC patients, the performance of this prognostic signature was further validated based on two independent cohorts. With the consistent regression formula and cutoff value of CpG-based signature derived from the training set, we also stratified the BC patients in validation set 1 ($n = 180$) and set 2 ($n = 237$) into low-risk and high-risk subgroups, respectively. And we applied Kaplan-Meier survival analysis to compare the OS rates of these subgroups. Additionally, time-dependent ROC curves were conducted to investigate the prognostic efficacy of the model in these external cohorts.

Correlation of Epigenetic Risk Score With Mutations and Copy Number Aberrations

Somatic mutation profiles and copy number variation data of BC patients were downloaded from TCGA repositories and classified into two distinct groups according to the established risk score. Therein, the driver mutation genes were analyzed and visualized using “maftools” R package (Mayakonda et al., 2018), and significant amplifications or deletions among the whole genomic region were identified via GISTIC 2.0 software (Mermel et al., 2011), respectively. Additionally, tumor mutation burden (TMB) of each patient was calculated via counting the total number of non-synonymous mutation events per megabase within whole genome (38 Mb was utilized as the estimate of the exome size). The burden of copy number losses or gains was defined as the total number of genes with copy number aberrations at the arm and focal levels (Sia et al., 2017).

Therapeutic Response Prediction in BC Patients

To further investigate the potential therapeutic properties for BC patients with distinct prognoses, we collected a series of therapeutic signatures, including the oncogenic pathways shaping non-inflamed TME and gene signatures for targeted therapies and radiotherapy. Therein, GSEA algorithm was implemented to calculate the corresponding enrichment scores, quantifying the possibility for specific therapeutic response for each patient (Hänzelmann et al., 2013).

The programmed cell death 1 (PD-1) and cytotoxic T-lymphocyte associated protein 4 (CTLA-4) are involved in tumor immune evasion. We combined the unsupervised subclass mapping method (SubMap, <https://cloud.genepattern.org/gp/>) and pretreatment genomics to predict the clinical responses to immune checkpoint blockade for distinct risk BC patients (Hoshida et al., 2007; Lu et al., 2019).

Based on the Genomics of Drug Sensitivity in Cancer (GDSC) database, a large public pharmacogenomics repository, we predicted the potential drug resistance to six commonly used chemotherapeutic agents, namely, cisplatin, gemcitabine, docetaxel, paclitaxel, etoposide and vinorelbine. R package “pRRophetic,” implements ridge regression and 10-fold cross validation to help in estimating the half-maximum inhibitory concentration (IC50) of each sample, according to GDSC training cell lines (Geeleher et al., 2014).

The Broad Institute’s connectivity Map build 02 database (CMap, <https://portals.broadinstitute.org/cmap/>) was a comprehensive resource for investigating relationships among biomarkers, diseases, and therapeutics (Qu and Rajpal, 2012). To screen potential candidate compounds targeting crosstalk against CpG sites derived from established signature, we calculated the differential expression gene (DEG) list between low- and high-risk patients and selected the top 300 to query the CMap database. Furthermore, the shared mechanisms of actions (MoA) among perturbagens were revealed via specific analysis within CMap tools (<https://clue.io/>) (Subramanian et al., 2017).

Statistical Analysis

Shapiro-Wilk normality test was utilized to test the normality of variables. For pairwise comparisons, statistical significance for non-normally distributed variables was estimated by wilcoxon test (Mann-Whitney U test), whereas normally distributed variables were analyzed by Student’s *t*-test. The significance of correlations between variables was computed by Spearman or Pearson correlation analysis. Contingency tables were analyzed by two-sided Fisher’s exact tests. All the statistical analyses were conducted using R (version 3.5.3, <https://www.r-project.org/>) software and a two-tailed *p*-value < 0.05 was considered statistically significant. Hierarchical cluster analysis using the Euclidean distance method to calculate dissimilarity structure was graphically explored with the “pheatmap” package. The LASSO-penalized Cox regression model was performed with “glmnet” package. ROC curves were derived using “pROC” and “PRROC” package, whereas the estimation of time-dependent ROC curves and AUC values of censored survival

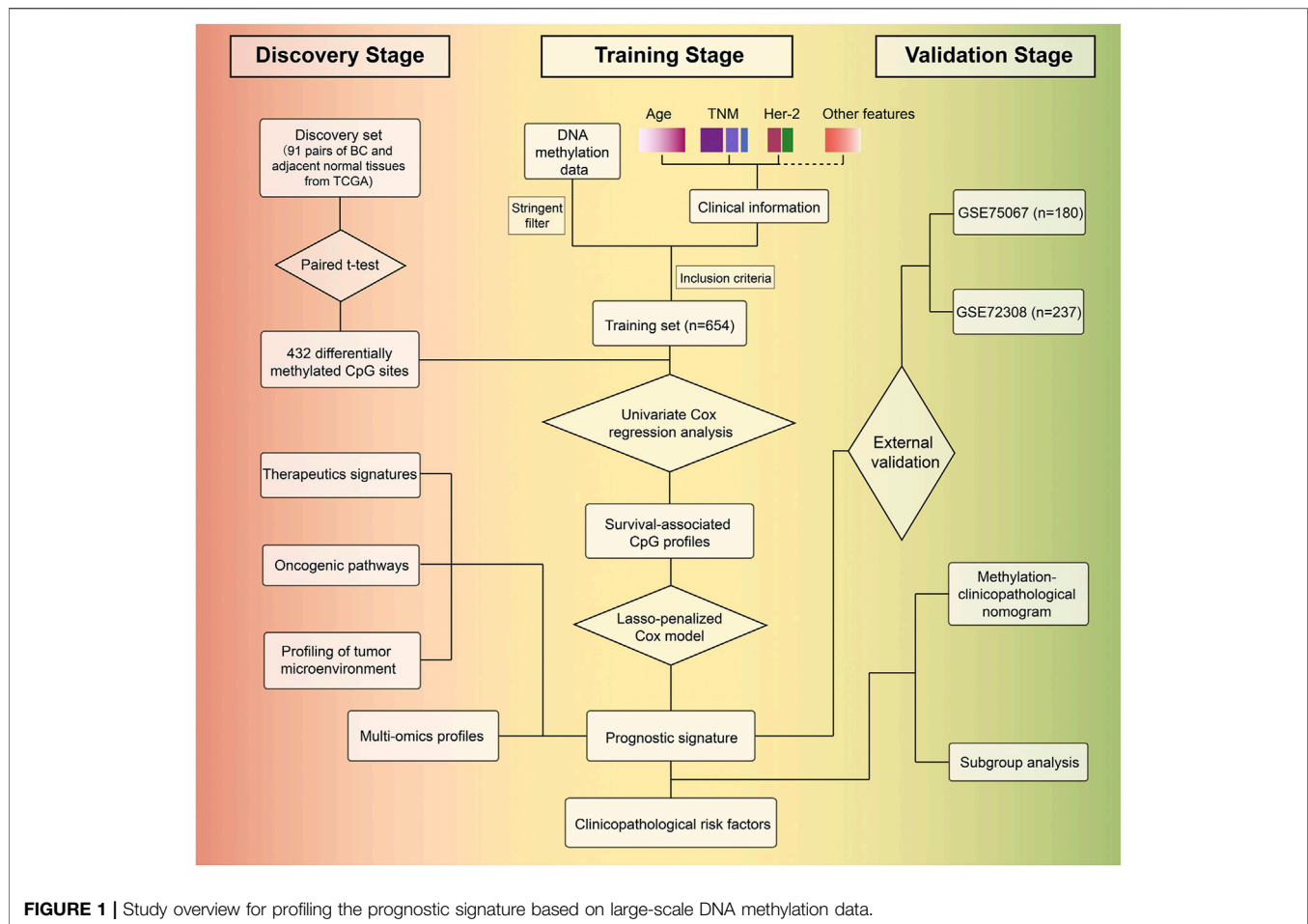


FIGURE 1 | Study overview for profiling the prognostic signature based on large-scale DNA methylation data.

data were computed using “timeROC” package. The nomogram and calibration plots were generated via “rms” package, and the c-index was calculated by “Hmisc” package. The DCA analyses were performed with the “dca.R” source code.

RESULTS

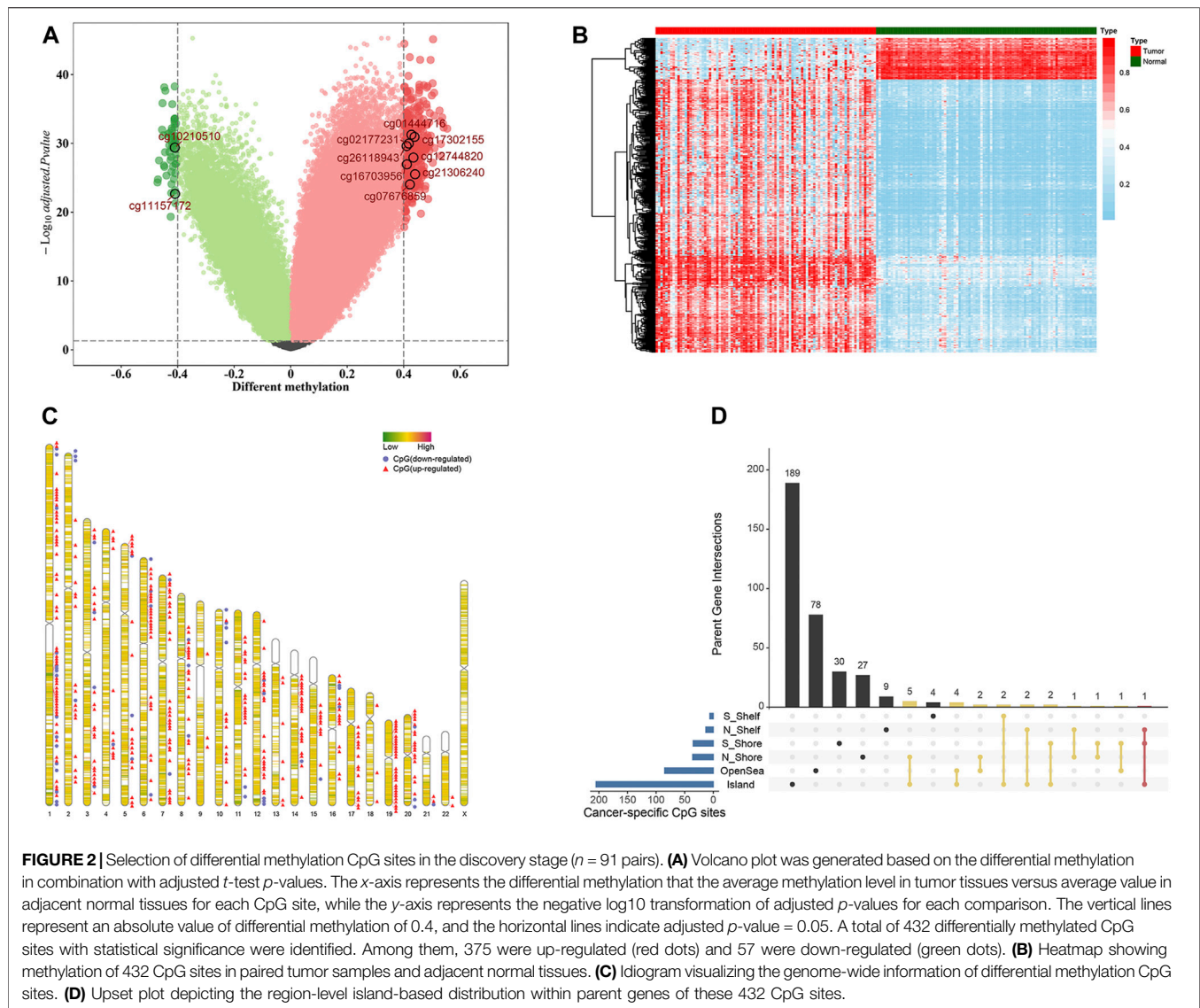
Selection of Cancer-Specific CpG Sites in BC Samples

We developed the model for the prognosis prediction of BC patients in three stages: discovery, training, and validation stage. The study flowchart is depicted in **Figure 1**. After a series of stringent filtering, a total of 308786 CpG sites were chosen as the background for the selection of candidate CpG sites. Firstly, in the discovery stage, volcano plot analysis (absolute differential methylation >0.4 and adjusted $p < 0.05$) identified 432 differentially methylated CpG sites (**Figure 2A**). Therein, 375 CpG sites were found to be hypermethylated and 57 CpG sites were hypomethylated, which corresponded to 308 and 54 genes, respectively. With regard to these cancer-specific CpG sites, hierarchical cluster analysis successfully segregated the 91 pairs of tumor/normal samples into two distinct clusters (**Figure 2B**). Then idiogram was used to map and visualize the genome-wide

information of these particular biomarkers into chromosomes (**Figure 2C**). Additionally, Upset plot was depicted to investigate the region-level island based distribution of these CpG sites across different genomic regions (**Figure 2D**). And it was found that the majority of these CpG sites were significantly enriched within the island and opensea, whereas only a few of them were located in the N shelf and S shelf.

Construction of a CpG-Based Signature and Evaluation of Its Predictive Ability in the TCGA BC Cohort

In the training stage, 654 BC patients with integrated methylation profiles and clinical information were retrospectively analyzed in depth, among which 66 (10.1%) died and the median survival time was 115.7 months. The median follow-up time of these BC patients was 17.3 months (range, 1.0–235.6 months). For the entire cohort, the 1-, 3-, 5-, and 8-years OS rates were 98.3, 91.5, 80.4, and 61.5%, respectively. To investigate the prognostic values of these cancer-specific biomarkers in BC patients, univariate Cox regression analyses were performed. As results, 18 CpG sites, the methylation levels of which were significantly correlated with OS ($p < 0.05$), were predominantly identified as the candidates (**Figure 3A**). To screen out the CpG sites with the

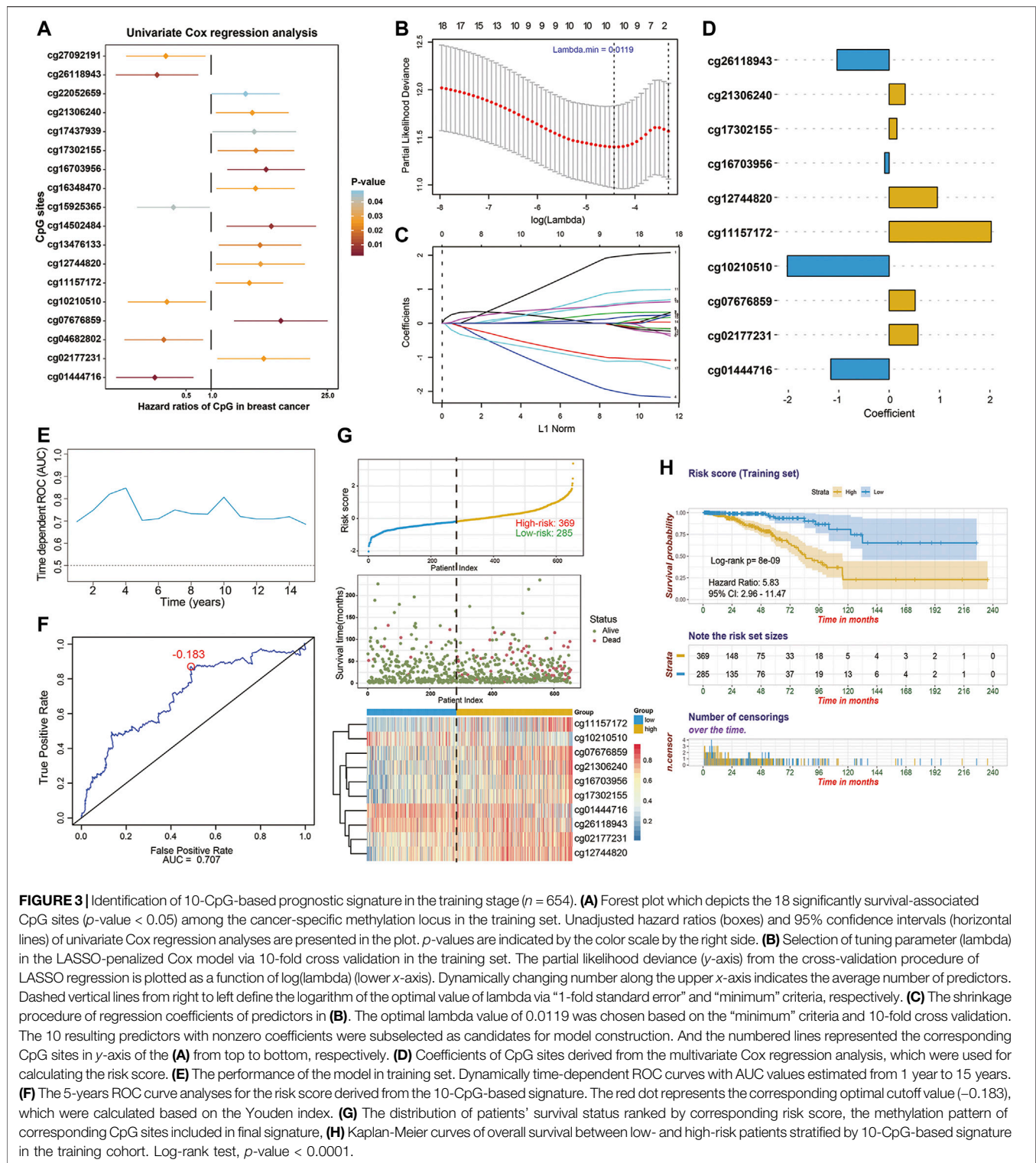


greatest potential prognostic values, we then applied a LASSO-penalized Cox regression model and narrowed down a methylation-based signature for BC patients in our training cohort (**Figures 3B,C**). Of these 18 candidate CpGs, 10 were finally identified, namely cg11157172, cg21306240, cg10210510, cg02177231, cg16703956, cg26118943, cg17302155, cg12744820, cg01444716 and cg07676859, mapped to PRDM16, EPHA10, COL9A2, TBX15, SLC6A3, SNX18, PRDM13, OLIG3, BTBD3, and SSTR4, respectively. Using coefficients weighted by multivariate Cox regression model (**Figure 3D**), an epigenetic score was calculated for each BC patient based on the individualized methylation levels of corresponding CpG sites as follows:

$$[2.0266 \times \text{Methylation level of cg11157172}] + [0.3172 \times \text{Methylation level of cg21306240}] + [(-2.0224) \times \text{Methylation level of cg10210510}] + [0.5711 \times \text{Methylation level of cg02177231}] + [(-0.0887) \times \text{Methylation level of cg16703956}] + [(-1.0346) \times \text{Methylation level of cg26118943}] + [0.1536 \times$$

$$\text{Methylation level of cg17302155}] + [0.9559 \times \text{Methylation level of cg12744820}] + [(-1.1586) \times \text{Methylation level of cg01444716}] + [0.5151 \times \text{Methylation level of cg07676859}]$$

Using the Youden index to generate the optimal cutoff value of the risk score, we assigned 369 BC patients (56.4%) with a derived risk score > -0.183 to a high-risk subgroup and others (43.6%) to a low-risk subgroup (**Figure 3G**). Kaplan-Meier survival analysis (**Figure 3H**) also revealed the risk of high-risk patients was 5.83-fold higher than that in the low-risk group (95%CI, 2.96–11.47; $p < 0.0001$). Next, time dependent ROC curves and C-index were performed to investigate the predictive efficiency of this established model (**Figures 3E,F**). The resulting AUC values of ROC curve remained above 0.7 even up to 15 years and C-index was 0.743 (95%CI, 0.675–0.811), indicating that our 10-CpG-based signature had superior predictive efficiency. The distribution of risk score, patients' survival status, and methylation patterns of corresponding CpG sites are shown in **Figure 3G**.



The CpG-Based Signature Is Independent of Conventional Clinicopathological Features

In order to investigate the prognostic value of 10-CpG-based signature in stratified cohorts, we classified BC patients into various subgroups and next performed stratification Cox

analyses for the entire cohort according to available clinicopathologic characteristics. As expected, our signature was able to classify patients with distinct prognoses, thus confirming its robustness for independently predicting BC prognosis (**Supplementary Figure S1**). Furthermore, 537 samples with

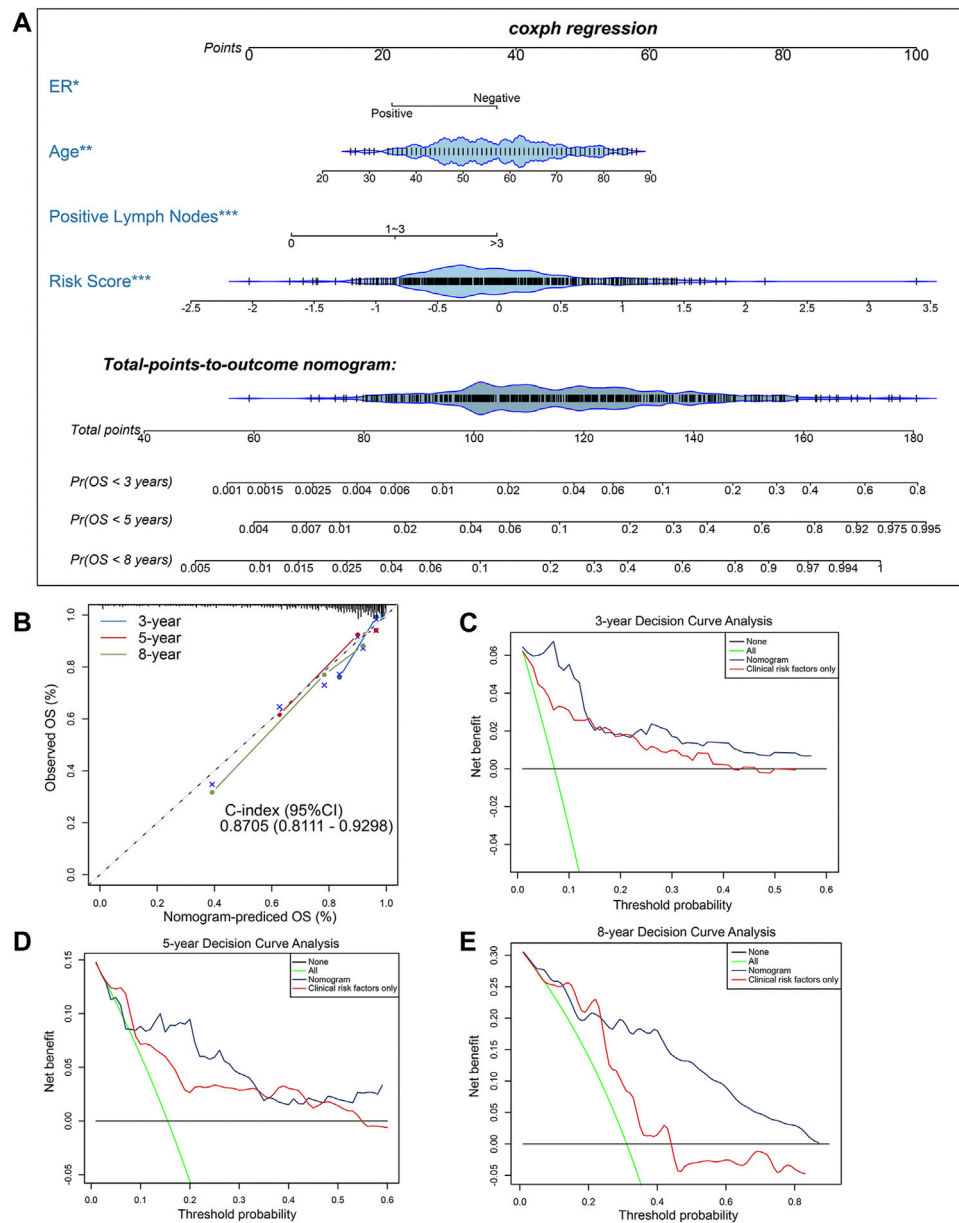


FIGURE 4 | Construction and evaluation of a clinical predictive model. **(A)** The methylation-clinicopathologic nomogram for predicting the 3-, 5- and 8-years overall survival probability for BC patients, which was developed in the TCGA cohort, with ER status, age, positive lymph nodes status and 10-CpG-based signature incorporated. **(B)** Calibration curves and C-index (95% CI) of the nomogram in TCGA cohort. The calibration plot depicts the calibration of the nomogram in term of agreement between observed and predicted 3-, 5- and 8-years clinical outcomes. The dashed 45-degree line indicates the ideal situation, and the dotted lines indicate the predictive performance of the model. **(C–E)** Decision curve analysis of the nomogram in TCGA cohort for evaluating the clinical usefulness in **(C)** 3 years, **(D)** 5 years and **(E)** 8 years. The blue line represents the CpG-clinicopathologic nomogram. The red line represents the model which integrated clinical risk factors only. The black line represents the treat-none scheme. And the green line represents the treat-all-patients scheme.

relatively complete clinical information within TCGA-BC cohort were subjected to verify whether our signature was an independent predictive factor for the prognosis of BC patients. The results of univariate analyses suggested that age, ER status, PR status, N stage, positive lymph nodes status, pathological stage, and risk score derived from 10-CpG-based signature were all remarkably correlated with OS of BC patients (**Supplementary Figure S2A**).

Hence, these risk factors were included to perform a multivariate Cox analysis and our results indicated the CpG-based signature was an independent prognostic factor when adjusted by those factors (**Supplementary Figure S2B**). Based on the PAM50 subtype classification of BC, the signature also shows its robust prognostic ability in basal-like and luminal subtype (**Supplementary Figures S2C,D**).

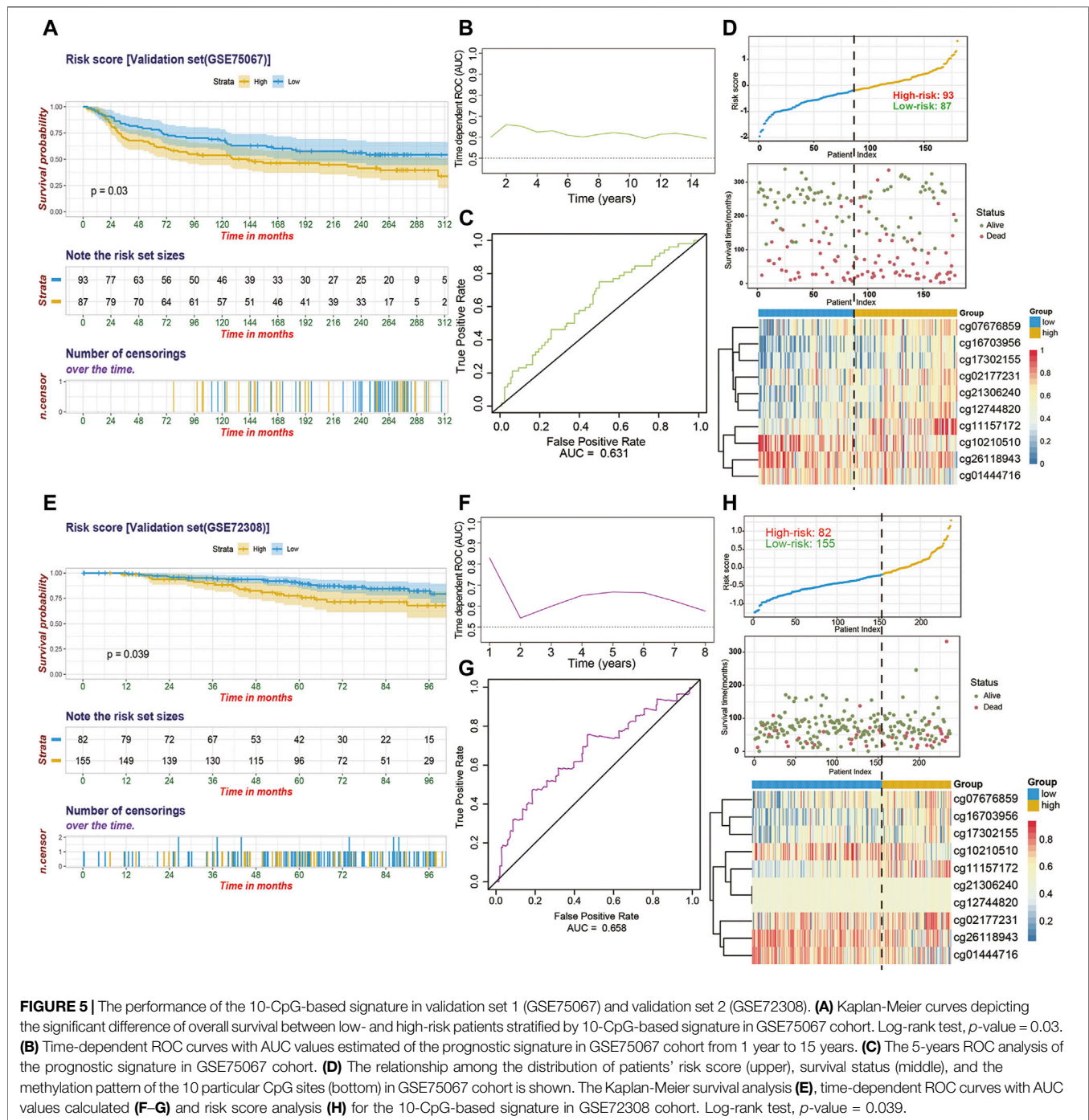


FIGURE 5 | The performance of the 10-CpG-based signature in validation set 1 (GSE75067) and validation set 2 (GSE72308). **(A)** Kaplan-Meier curves depicting the significant difference of overall survival between low- and high-risk patients stratified by 10-CpG-based signature in GSE75067 cohort. Log-rank test, p -value = 0.03. **(B)** Time-dependent ROC curves with AUC values estimated of the prognostic signature in GSE75067 cohort from 1 year to 15 years. **(C)** The 5-years ROC analysis of the prognostic signature in GSE75067 cohort. **(D)** The relationship among the distribution of patients' risk score (upper), survival status (middle), and the methylation pattern of the 10 particular CpG sites (bottom) in GSE75067 cohort is shown. The Kaplan-Meier survival analysis **(E)**, time-dependent ROC curves with AUC values calculated **(F-G)** and risk score analysis **(H)** for the 10-CpG-based signature in GSE72308 cohort. Log-rank test, p -value = 0.039.

Construction and Assessment of a Novel Methylation-Clinicopathological Nomogram

To further develop a clinically applicable approach for predicting OS probability in BC patients, a secondary multivariate Cox regression analysis that integrated the identified independent risk factors, consisting of age, ER status, positive lymph nodes status and risk score, was performed and depicted as an inclusive

nomogram (Figure 4A). It substantiated that the risk score contributes the greatest weight to the total points, whereas other clinicopathological features contribute much less, which were consistent with results from previous multivariate regression analyses. The C-index of our nomogram for OS prediction reached 0.8705 with 1,000 bootstrap replicates (95% CI: 0.8111–0.9298), indicating a favorable discriminatory performance of derived nomogram. And the results of calibration curves showed that the bias-corrected lines were

close to the ideal line (45-degree line), indicating good agreement between the predicted and actually observed outcomes (**Figure 4B**). Moreover, we performed DCA analysis to evaluate the real-world clinical usefulness of the inclusive nomogram by quantifying the net benefits against a range of threshold probabilities. The DCA results of the nomogram are presented (**Figures 4C–E**), revealing that the prognosis-related treatment decision-making based on our methylation-clinicopathological nomogram could add more net benefit than treat either none or all patients if the probability threshold for doctors or patients does not exceed 60%. Besides, the clinical usefulness of our nomogram significantly overwhelmed the conventional clinicopathological factors. Consequently, these findings suggested that our nomogram was an optimal prognostic model for predicting both short-term and long-term survival probability in BC patients.

External Validation and Evaluation of 10-CpG-Based Prognostic Signature in the GEO BC Cohorts

To substantiate the stability of final 10-CpG-based signature, external analyses were performed in two external validation cohorts: GSE75067 and a subset of GSE72308. Utilizing the same risk score formula and cutoff value obtained from the training cohort, the BC patients in the GSE75067 dataset were categorized into a high-risk (51.7%) subgroup and a low-risk (48.3%) subgroup. Consistent with the outcomes of TCGA-BC cohort, the patients who were assigned to the high-risk subgroup had significantly worse OS than low-risk BC patients (**Figure 5A**; HR, 1.576; 95%CI, 1.041–2.385; $p = 0.03$), demonstrating the robustness and clinical applicability of the established signature across different cohorts. And time-dependent ROC analysis revealed the CpG signature had favorable efficacy for predicting both short-term and long-term OS, with AUC value maintaining above 0.6 from 1 year to 15 years (**Figures 5B,C**). The distribution of risk score and methylation patterns of 10 particular CpG sites are shown (**Figure 5D**). Furthermore, similar analyses showed that the CpG signature could successfully divide 155 BC patients (65.4%) into low-risk subgroup and 82 patients (34.6%) into high-risk subgroup in term of OS (HR, 1.876; 95%CI, 1.023–3.44; $p = 0.039$) in the GSE72308 cohort (**Figure 5E**). Similarly, the CpG signature maintained its discriminative power for prognosis prediction spanning from 1 year to 8 years (**Figures 5F,G**). Additionally, the risk score distribution, survival status, and methylation patterns of 10 CpG biomarkers are displayed (**Figure 5H**). Taken together, these results suggested that our 10-CpG-based signature is robust for differentiating BC patients with a favorable or poor prognosis, which may possess the pivotal mechanism underlying the BC carcinogenesis, progression, and metastasis.

Decipher the Diagnostic Values of These CpG Sites in BC

These identified ten CpG sites were significantly differentially methylated in paired tumor/adjacent-normal tissues

(**Supplementary Figure S3**). ROC curves revealed that these CpG could individually or jointly show diagnostic value for BC in high efficacy in discovery set (**Supplementary Figure S4**) and entire TCGA-BC dataset (**Supplementary Figure S5**). Additionally, Spearman correlation analyses were performed to estimate the prospective methylation and expression quantitative trait loci (meQTL) relationships (**Supplementary Figure S6**). As results, hypermethylation of CpG sites (cg01444716, cg02177231, cg26118943, and cg07676859) inhibited expression of corresponding gene (BTBD3, TBX15, SNX18, and SSTR4), while hypermethylation (cg11157172, cg21306240, and cg10210510) was positively correlated with parent gene (PRDM16, EPHA10, and COL9A2) expression. However, methylation levels of three other CpG sites (cg12744820, cg16703956, and cg17302155), which were located in 1st exon of OLIG3, TSS1500 of SLC6A3, and gene body of PRDM13, respectively, did not display any correlation with the gene expression.

Altered Biological Processes and Pathways in High- and Low-Risk BC Patients

The strong risk stratification ability of our 10-CpG-based signature for BC patients could be attributed to their potential regulation in tumor development or metastasis. Therefore, GSEA analysis was performed to elucidate the association between potential biological phenotypes and our prognostic signature. In the GSEA enrichment results, we noticed that genes highly expressed in the high-risk patients were significantly enriched in multiple cancer-related pathways, such as Notch4 signaling pathway (normalized enrichment score, NES = 1.937, size = 53), cell cycle regulation (NES = 1.866, size = 85), stabilization of P53 (NES = 1.806, size = 55), regulation of DNA damage checkpoints (NES = 1.917, size = 66), and apoptosis (NES = 1.951, size = 52) ($p < 0.05$ and FDR < 0.25) (**Supplementary Figure S7**). In contrast, the CpG-based model was negatively correlated with immune relevant behaviors, including alteration of immune checkpoint molecules, activation of immune cell signaling pathways, and Interleukin-related signatures (**Figure 6**), further demonstrating low-risk BCs were characterized as an enhanced immune phenotype. Therefore, it could be speculated that our CpG-based signature might serve as an interface between epigenetic modification and immune activities. In summary, our GSEA analysis implied that these CpG sites play crucial roles in mammary carcinogenesis via the above-referenced biological processes and their functional dysregulation subtly affected the OS of BC patients.

The Risk Score Shapes a Non-inflamed TME in BC

Considering that a vast number of immunological pathways were downregulated in high-risk BC patients, we sought to explore the intrinsic correlation between the TME and our established risk scores to elucidate the immunogenicity of each risk subgroup. To execute this, we first evaluated the correlation between the risk score and seven steps of cancer-immunity cycle, which

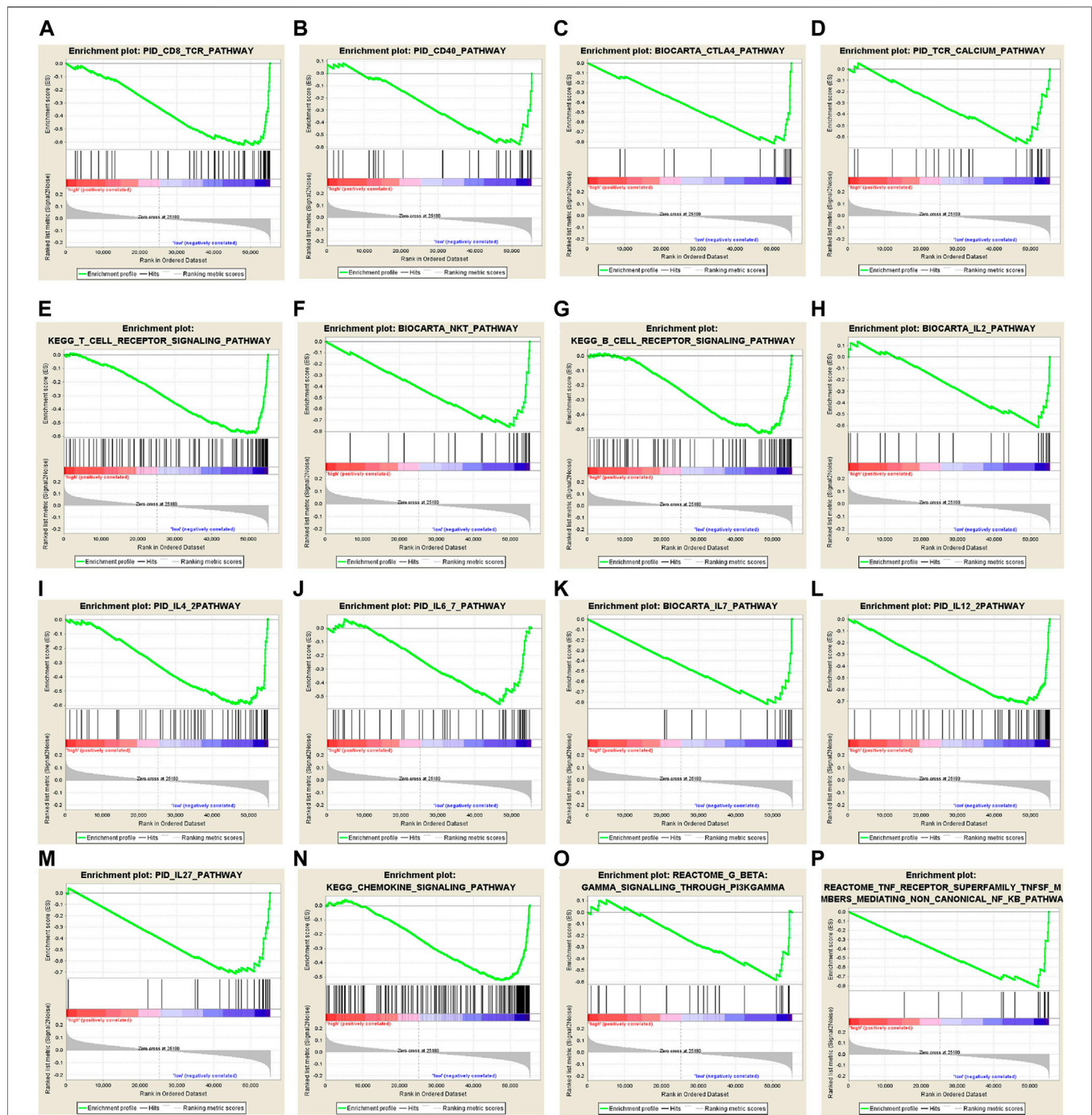


FIGURE 6 | Gene set enrichment analysis (GSEA) delineates the significantly enriched pathways which are negatively correlated with 10-CpG-based signature using the “c2.cp.v7.0. symbols” gene set, downloaded from the MSigDB database, as the reference. And the significance threshold was set at nominal p -value < 0.05 and FDR < 0.25.

conceptualized the anti-cancer immune response. Overall, activities of the majority of the steps in the cycle were negatively correlated with our prognostic signature (Figure 7A), including the release of cancer cell antigens (Step 1), cancer antigen presentation (Step 2), priming and activation (Step 3), trafficking of immune cells to tumors (The recruiting of

Th1 cell, dendritic cell, macrophage, NK cell, Th17 and Th2 cell, Step 4), and infiltration of immune cells into tumors (Step 5). Subsequently, for high-risk patients, the reduced activities of these steps may also impede the infiltration of tumor-associated immune cells and promote the formation of non-inflamed TME.

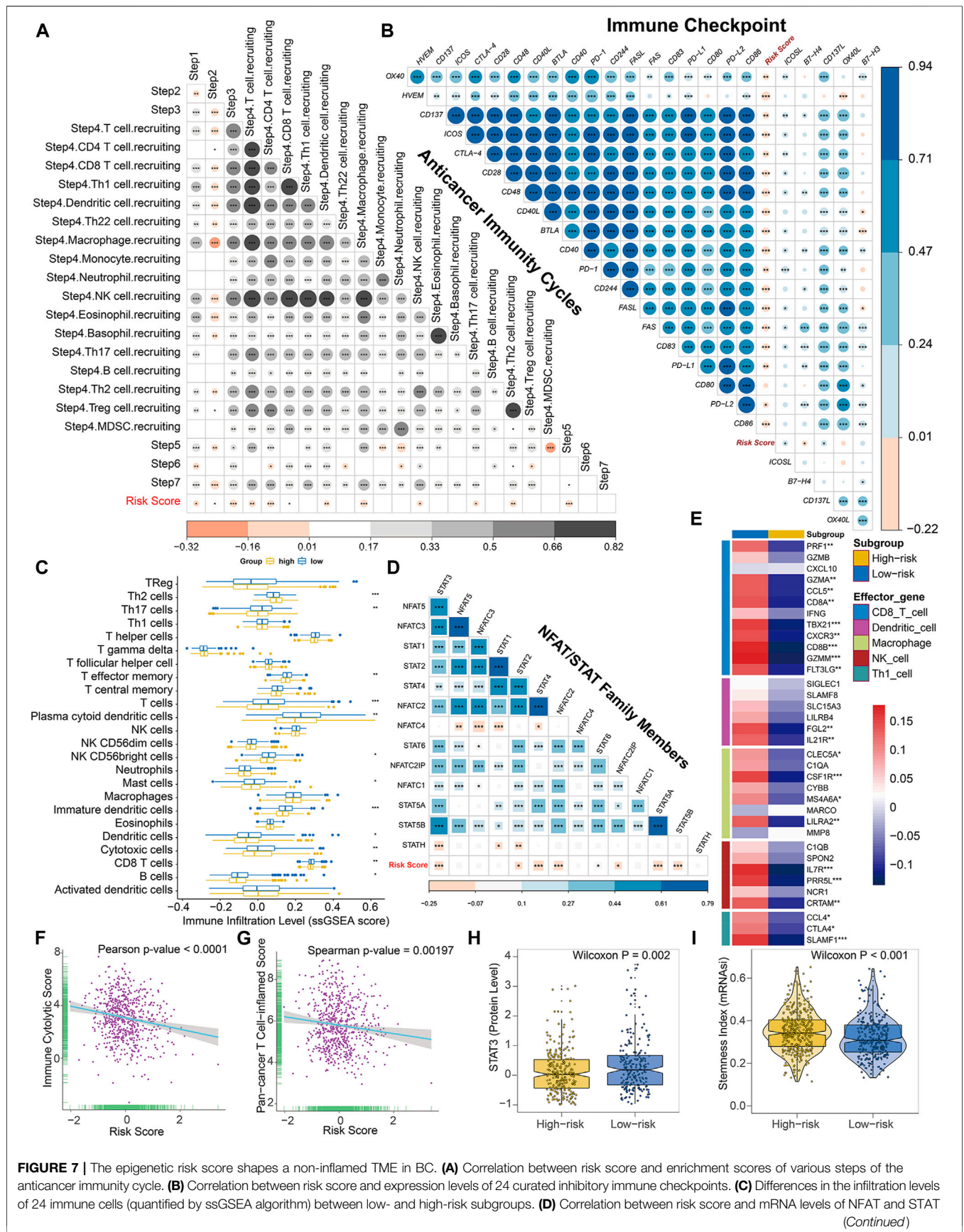


FIGURE 7 | The epigenetic risk score shapes a non-inflamed TME in BC. **(A)** Correlation between risk score and enrichment scores of various steps of the anticancer immunity cycle. **(B)** Correlation between risk score and expression levels of 24 curated inhibitory immune checkpoints. **(C)** Differences in the infiltration levels of 24 immune cells (quantified by ssGSEA algorithm) between low- and high-risk subgroups. **(D)** Correlation between risk score and mRNA levels of NFAT and STAT (Continued)

FIGURE 7 | family members. **(E)** Differences in expression levels (average z-scores) of the effector genes of tumor-associated immune cells between low- and high-risk patients. **(F)** Pearson correlation between immune cytolytic activities and risk score. **(G)** Spearman correlation between pan-cancer T cell inflamed signature and risk score. Both T cell inflamed score and immune cytolytic activities are positively correlated with clinical response to anti-cancer immunotherapy. **(H)** Differences in protein expression of STAT3 between low- and high-risk patients based on the reverse-phase protein array analysis. **(I)** Differences in stemness index (mRNAsi) between low- and high-risk patients. The statistical significance of pairwise correlations or comparisons is annotated with asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Thereafter, using the ESTIMATE algorithm in combination with the corresponding transcriptome profiles, we evaluated the differences in the TME between high- and low-risk BC patients, comprised of immune cell infiltration level (immune score), stromal cell content (stromal score), and tumor purity. Based on this algorithm, immune scores ranged from -1337.749 to 3747.411 , and stromal scores were distributed between -2142.388 and 2125.969 , respectively. And we found that both the immune and stromal scores were significantly elevated in low-risk subtype while an opposite trend was revealed when comparing the tumor purify of different risky subgroups (**Supplementary Figures S8A–C**), suggesting the risk score may reflect cancer immunogenicity in BC.

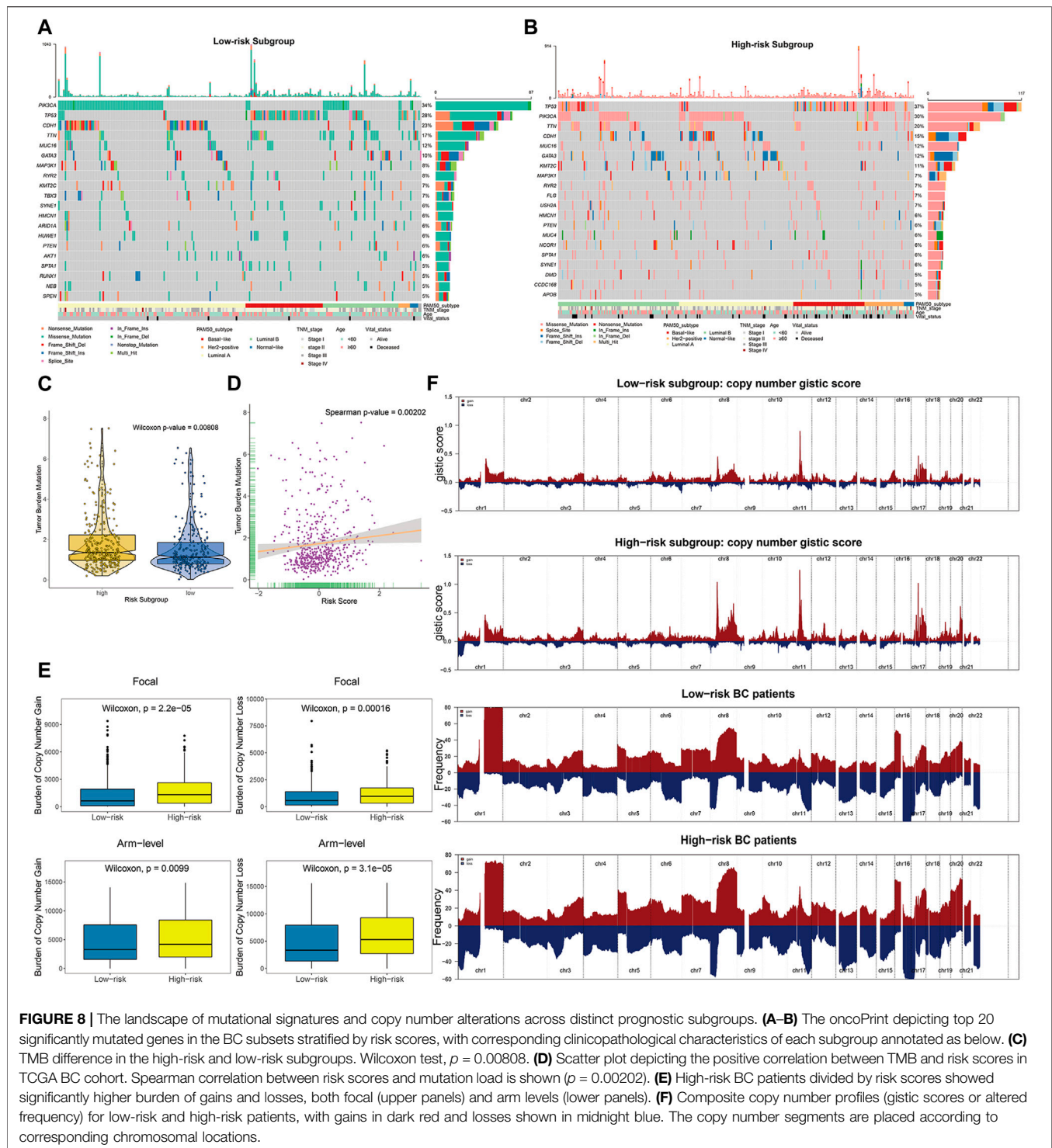
Consistently, the risk score was found to be mutually exclusive with a majority of immune checkpoint inhibitors, such as PD-1, PD-L1, CTLA-4, and so on (**Figure 7B**). Theoretically, the low abundance of checkpoint molecules was always indicative of resistance to immune checkpoint blockade. Moreover, the risk score was negatively correlated with a vast majority of immunomodulators (**Supplementary Figure S9**). Notably, a majority of MHC molecules were significantly downregulated in high-risk BC subgroup, suggesting potential regression on the capacity of antigen presentation and processing within these patients. Several critical chemokines, such as CCL19, CCL21, CCL22, XCR1, and paired receptors CCR7, CCR4, XCL1, and XCL2, were also negatively correlated with risk score. These chemokines and receptors could induce the recruitment of tumor-infiltrating immune cells such as CD8⁺ T cells and regulatory T cells (Treg). These outcomes indicated our risk score may exert a wide range of effects upon anticancer immune response.

Furthermore, for reducing calculation errors caused by different inner algorithms and marker gene sets, several independent methods, including ssGSEA, MCP-counter, and Cibersort, were applied to infer the infiltration levels of immune cells among BC patients using bulk RNA-seq data. Consistently, the majority of immune subpopulations within TME exhibited reduced infiltration in high-risk BC patients, indicating an immune-deserted TME in high-risk subgroup (**Figure 7C**, **Supplementary Figures S8D–K**, **S10B**). Of note, the infiltration levels of dendritic cells and CD8⁺ T cells, which induce antigen presentation and immune activation in tumors, were significantly lower in the high-risk subgroup based on these three methods. Both MCP-counter and Cibersort revealed that the infiltration level of macrophage was significantly reduced in high-risk patients. In addition, the high-risk subgroup exhibited downregulation of the effector genes of these tumor-infiltrating immune cells (**Figure 7E**). Additionally, the ssGSEA algorithm also detected that a wide range of immune cells was significantly

decreased in high-risk BC subgroup, including T cells, Th17 cells, T effector memory cells, NK cells, mast cells, and B cells. We also found that cancer-associated fibroblasts were significantly downregulated in low-risk subtype. In addition, the risk score was negatively correlated with the gene signature of Th1 cells (**Figure 7E**). Collectively, for high-risk BC patients, the coordination role of these methylation biomarkers may restrain recruitment of tumor-associated infiltrating immune cells, transforming the inflamed surroundings into a non-inflamed TME.

Additionally, we observed that mRNA levels of STAT and NFAT family members, such as STAT3, STAT4, STAT5, and NFATC2, were also negatively correlated with our risk score (**Figure 7D**). The reverse-phase protein array data downloaded from cbiportal (<http://www.cbiportal.org/>) also validated that the protein level of STAT3 was significantly lower in high-risk patients compared with low-risk ones (**Figure 7H**). Given the evidence that NFAT and STAT family are well-known transcriptional regulators of immune checkpoints in T cells, these data indicated that DNA methylation biomarkers promote the downregulation of immune checkpoints synergistically in BC, potentially through the inhibition of NFAT/STAT signaling pathways.

To further authenticate our conjecture, we used CLS computed by Rooney et al., revealing the high-risk subgroup tends to exhibit downregulation of immune cytolytic activity (**Figure 7F**, **Supplementary Figure S8L**). The pan-cancer TIS is positively correlated with clinical response to anticancer immunotherapy and defines the pre-existing inflammatory status in TME. As expected, the risk score is negatively correlated with the T cell inflamed signature and the majority of genes within this signature (**Figure 7G**, **Supplementary Figure S10A**), further confirming its roles in shaping non-inflamed TME. And high-risk BC patients showed higher cancer stemlike indices (**Figure 7I**), indicating a potential negative regulation between stemness and anti-tumor immunity. Moreover, IPS was known as a superior immune response molecular marker to quantify the intratumoral immune landscape and predict response to immune checkpoint blockade therapy spanning multi-types of tumors. Of note, the epigenetic risk score was significantly negatively correlated with IPS (**Supplementary Figure S11A**) and the IPS, IPS-PD1-PD-L1-PD-L2, IPS-CTLA4, and IPS-PD1-PD-L1-PD-L2-CTLA4 scores in the high-risk subgroup were significantly lower compared with low-risk subgroup (**Supplementary Figure S11B**), indicating a less immunogenic phenotype and less sensitivity to immune checkpoint blockade treatment in high-risk patients. In addition, we found that in both the pre-defined low- and high-risk BC patients, C1 (Wound Healing) and C2 (IFN- γ



Dominant) immune subtypes were mostly enriched with absence of C5 (Immunologically Quiet), which were in accordance with results from Thorsson et al. It was noteworthy that the high-risk subgroup had considerably more C1 and C2 subtypes characterized by high proliferation rate and poor prognosis, whereas low-risk ones occupied a higher percentage of C3 (Inflammatory) and C6 (TGF- β Dominant) subtypes with a

high inflammatory signature and lymphocytic infiltrate (**Supplementary Figure S11C**). Thus, the distribution of immune subtypes between low- and high-risk patients was consistent with the prognosis and immunogenicity of each subgroup.

To further explore the contribution of each CpG site in shaping non-inflamed TME, we performed the correlation

analyses between methylation levels of biomarkers and adverse indicators of modifying non-inflamed TME, including TIS, CLS, and IPS. As shown in **Supplementary Figure S11D**, the methylation level of cg102110510 was significantly positively correlated with TIS, CLS, and IPS, indicating its negative role in modifying a non-inflamed TME, whereas the epigenetic modification of cg21306240, cg17302155, cg16703956, cg12744820, and cg022177231 was significantly positive correlated with shaping non-inflamed TME in BC.

In summary, these results indicated that compared with its counterparts, BC patients who were divided into high-risk subtype via 10-CpG-based signature have a distinct immune phenotype, characterized by lower immune activation and being less efficacious to immunotherapies. Hence, a prior treatment option for BC with higher risk score was to transform a non-inflamed TME into an inflamed microenvironment, consequently triggering anti-cancer immune response.

The Epigenetic Risk Score Is Associated With Distinct Genomic Imprints

To explore the intrinsic correlation between established risk score and specific genomic alterations in BC, we performed multi-omics data analyses, including somatic mutation and copy number variation analysis using the corresponding TCGA profiles. In terms of somatic mutation analysis, patients in the high-risk subgroup had significantly higher somatic mutation burdens in TP53, KMT2C, and GATA3 (**Figures 8A,B**), which have been shown to play crucial roles in the carcinogenesis of multiple tumors. Additionally, patients in the high-risk subgroup showed a significantly higher overall TMB than low-risk ones (Wilcoxon-test $p = 0.00808$; **Figure 8C**). Further correlation analyses confirmed that the risk score was significantly and positively correlated with TMB (Spearman $p = 0.00202$; **Figure 8D**).

We further explored the copy number burdens of distinct subgroups. Similar to the findings as previous described, the high-risk BC patients, occupied with non-inflamed TME, showed a higher burden of gain and loss both at the focal and arm-levels (**Figure 8E**). As illustrated within focal amplification and deletion peaks, we observed the deletions of chr1 (1p36) and amplifications of chr8 (8p11), chr17 (17q12), and chr20 (20q13) were significantly enriched in high-risk patients (**Figure 8F**, **Supplementary Figure S12**).

The Risk Score Distinguished Different Therapeutics Responses in BC

After comparing the differences in enrichment scores of therapeutic signatures, we could determine the roles of risk score in predicting clinical response to these therapies. As exhibited in **Figure 9A**, the majority of immune inhibited oncogenic pathways, such as WNT- β -catenin_network, cell cycle, mismatch repair, and p53 signaling pathway, were invigorated in the high-risk subgroup, suggesting an inhibited immune status in them. And it may be reasonable to take these

signatures as potential therapeutic targets for high-risk BC patients. However, for low-risk ones, immunotherapy should be regarded as candidates due to the inflamed TME within them. The submap analysis also revealed the low-risk patients were more sensitive to anti-PD-1 treatment (**Figure 9B**).

Since chemotherapy is a commonly used strategy for clinical treatment on BC patients, we selected three commonly used and three other useful chemicals agents (cisplatin, gemcitabine, docetaxel, paclitaxel, etoposide, and vinorelbine), further evaluating the clinical responses of each risk subgroup. As is shown in **Figure 9C**, the estimated IC50 for gemcitabine, etoposide, and vinorelbine was significantly higher in high-risk BC subgroup, indicating potential chemotherapeutic drug resistance for these patients.

Additionally, we intended to screen for the candidate compounds, especially targeting the high-risk BC patients identified by our risk score. Top 300 DEGs were obtained by comparing the corresponding mRNA expression profiles between low- and high-risk subgroups. The top 66 perturbagens that potentially target these DEGs were exhibited in **Figure 9D**, along with 48 MoAs derived from CMap analysis. The top hits revealed that six compounds (chlorpromazine, haloperidol, levomepromazine, prochlorperazine, thioridazine, and trifluoperazine) shared the MoA of dopamine receptor antagonist. Meanwhile, other specific MoAs, such as estrogen receptor agonist (diethylstilbestrol and estriol), mTOR inhibitor (LY-294002 and sirolimus), PKA inhibitor (H-7), and tyrosine kinase inhibitor (genistein), were also identified as potential therapeutic targets for high-risk BC patients. These findings may provide opportunities for optimizing treatment in BC.

DISCUSSION

As a highly complex and heterogeneous disease, BC involves intricate interwoven relationships (Reis-Filho and Pusztai, 2011). Currently, the conventional clinicopathologic indicators, including tumor size, lymph node metastasis, TNM stage, and cellular biomarkers (ER, PR, HER-2 status and Ki-67 index) of tumor biopsy, are still the gold standard for the risk stratification and subsequent formulating of therapeutic scheme. However, BC patients might possess different treatment responses and clinical outcomes spanning <6 months to beyond 10 years, even with the same histological grade and pathological stage (Network, 2012; Olsson et al., 2013). In this aspect, integrating reliable prognostic biomarkers into real-life work are a critical requisite to identifying the subset of patients who harbored a worse prognosis and might benefit from the systematic adjuvant therapy. Significant research on the high-throughput gene expression profiles has led to recent commercialization of two multi-gene-based signatures (Mamounas et al., 2010; Cardoso et al., 2016; Geyer et al., 2018), but their additional utility in aiding treatment decision in BC patients still needs validating based on long-term prospective studies (Sparano et al., 2015).

DNA methylation, as a changeable and possibly heritable epigenetic modifying mechanism, offers promising clues for early detection, monitoring treatment response, prognosis

prediction, and molecular subtyping in cancer (Franco et al., 2008; Wu and Ni, 2015; Zhang et al., 2018c). The aberrant methylation of particular subsets of CpG sites is prone to take place at the beginning of carcinogenesis, resulting in specific processes of tumorigenesis (Wang et al., 2019). Preliminary investigation indicated that using DNA methylation as a biomarker holds serious implications for cancer diagnosis and treatment, including the relative steady expression both *in vivo* and *ex vivo* (Keeley et al., 2013), ensuring precise judgement in a minimally invasive way instead of requiring bulk tissue samples (Dai et al., 2011), and higher accuracy in cancer management, especially distinguishing relatively indolent or aggressive tumors (Hao et al., 2017). Taking the potential epigenetic modification of DNA methylation in breast carcinogenesis into account, it is reasonable to consider the CpG sites to increase the effectiveness of adjuvant therapy. In the current study, we proposed a robust, individualized prognostic signature that can estimate OS probability in BC patients on the basis of 10 aberrant CpG sites, through integrating the well-established public, large-scale multi-omics sequencing cohorts. And our prognostic signature could further stratify clinically defined groups of patients (e.g., age, histological type, TNM stage) into subgroups with distinct survival outcomes. Moreover, attributed to the complementary value of molecular and conventional clinical characteristics, our inclusive nomogram could indicate a more accurate estimation of OS in both short-term and long-term prognosis prediction for BC patients.

We applied a three-stage selection procedure to screen out the prominent biomarkers from more than 300,000 CpG sites after quality control and rigorous filtering. Initially, in the discovery stage, differential methylation analyses using paired tissue profiles winnowed out nearly 99.86% of probes, and those CpG sites with cancer-specific properties in BC patients were preserved. Next, in the training stage, univariate Cox regression analyses were performed to assess their prognostic values, excluding probes unrelated to OS. Meanwhile, considering that the Cox model is insufficient for variable selection due to the high-dimensional data, the successive application of machine-learning methods, LASSO-penalized Cox regression with 10-fold cross validation, to screen the optimal combination of CpG sites for further modeling, markedly raised the accuracy of methylation-based signature. Finally, in the validation stage, the performance of the established model was further validated based on two external independent cohorts, confirming its robust and reliable prognosis prediction.

Furthermore, the GSEA analysis provided more opportunities for deciphering the largely untapped mechanisms which our identified CpG sites might participate in. For instance, Nothch4 signaling pathway is considered as playing pivotal roles in malignant potential, including proliferation, invasiveness, and metastasis, by sustaining epithelial-mesenchymal transition and controlling BC stem cell activity (Nagamatsu et al., 2014; Kim and Singh, 2015; Wang et al., 2018; Giuli et al., 2019), and activation of it could contribute to endocrine therapy assistance in BC cells (Lombardo et al., 2014; Simões et al., 2015; Bui et al., 2017). Upregulation of ORC1 (origin recognition complex 1), the largest unit of ORC

required in the initiation of DNA replication, was gradually confirmed resulting in DNA re-replication to trigger DNA damage response and control cancer cell-cycle (McNairn and Gilbert, 2005; Kara et al., 2015; Chen et al., 2019). Altered tumor metabolism, including increased glycolysis in cancer cells, determined the malignant biological behaviors and cancerous phenotype (Ma and Zong, 2020). Inhibitor of glycolysis pathway could sensitize the anti-cancer effect of chemotherapeutic agents and delay the occurrence of acquired drug resistant in hypoxia (Le Calvé et al., 2010; Jiang et al., 2018). In addition, dysregulation of other pathways, such as regulation of apoptosis (Jordan, 2015; Ahmed et al., 2019), DNA damage checkpoints (Lord and Ashworth, 2012; del Rincón et al., 2014), and P53 signaling pathway (Silwal-Pandit et al., 2017), were also involved in cancer cell growth, proliferation, invasion, and metastasis and played crucial roles in BC cells. Hence, the results mentioned above added more evidence for the interactions between our established signature and cancers, highlighting its clinically transitional potential.

Nowadays, exploiting the intrinsic interplay of host immune system and malignant tumors has achieved impressive success (Chung et al., 2017; Plava et al., 2019). Moreover, preliminary reports have provided elegant analyses on the cross-talk between tumor-intrinsic genes and BC microenvironment (Fox et al., 2019; Patel et al., 2019), hailing cancer immunotherapy as a vast breakthrough in the combat against malignancies (Jia et al., 2017; Szekely et al., 2018). However, currently, the immune checkpoint blockades such as PD-1, PD-L1, and CTLA-4 do not exhibit an overwhelming situation for conquering cancers. Only a small portion of BC patients with suitable immunogenicity could show clinical response to immunotherapy. And it is necessary for the success of immunotherapy that to hold an inflamed TME in conjunction with pre-existing anticancer immunity. Particular molecules or pathways, inducing the formation of non-inflamed TME and making the recruitment of tumor-cytotoxic T cells invalid, may cause resistance to immunotherapeutic agents. Intriguingly, an extensive immunogenomic analysis has revealed that our established risk score could distinguish subgroups with different prognoses and remarkably distinct phenotypes within TME. For high-risk BC patients with a non-inflamed TME, transforming it into an inflamed TME is one of the top priorities along with re-ignition of tumor-infiltrating immune cells to drive tumor regression. To some extent, these findings suggested the CpG sites included in our signature hold great promise for identifying novel molecular targets and improving cancer management in the era of immunotherapy.

Although our study proposed a robust 10-CpG-based prognostic signature and shed new light on the epigenetic microenvironment for possible therapeutic potential, several limitations should be addressed. Firstly, due to the retrospective nature of our work, it is imperative to design multicenter prospective clinical trials with large sample sizes for further providing high-level evidence for clinical application. Secondly, due to the lack of normal mammary samples, the differential methylation profiles merely derived from the comparison between tumor and adjacent normal

tissues need to be further validated. Thirdly, our attempts were based on single-omics data (DNA methylation), so that the understanding of BC-related properties could be inevitably incomprehensive in the presence of tumor heterogeneity. Finally, concrete biological mechanisms of the candidate CpG sites still need to be experimentally verified, especially underlying the immune microenvironment.

CONCLUSION

Overall, we constructed and externally validated a novel predictive model by sufficiently integrating and analyzing the DNA methylation profiles. And we expect that application of our model will not only greatly contribute to the personalized follow-up and decision-making process of clinicians, but also facilitate further understanding of the basic biology of BC and thereby inform the drafting of appropriate therapeutic strategies in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The data sets involved in our study are publicly available in the TCGA database (<https://cancergenome.nih.gov/>) and the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

AUTHOR CONTRIBUTIONS

QY and DZ conceived and designed the project. QY provided administrative support. DZ provided study materials or patients. DZ and YW collected and assembled the data. DZ analysed and

REFERENCES

- Ahmed, S., Khan, H., Fratantonio, D., Hasan, M. M., Sharifi, S., Fathi, N., et al. (2019). Apoptosis Induced by Luteolin in Breast Cancer: Mechanistic and Therapeutic Perspectives. *Phytomedicine* 59, 152883. doi:10.1016/j.phymed.2019.152883
- Akram, M., Iqbal, M., Daniyal, M., and Khan, A. U. (2017). Awareness and Current Knowledge of Breast Cancer. *Biol. Res.* 50 (1), 33. doi:10.1186/s40659-017-0140-9
- Aryee, M. J., Wu, Z., Ladd-Acosta, C., Herb, B., Feinberg, A. P., Yegnasubramanian, S., et al. (2010). Accurate Genome-Scale Percentage DNA Methylation Estimates from Microarray Data. *Biostatistics* 12 (2), 197–210. doi:10.1093/biostatistics/kxq055
- Auslander, N., Zhang, G., Lee, J. S., Frederick, D. T., Miao, B., Moll, T., et al. (2018). Robust Prediction of Response to Immune Checkpoint Blockade Therapy in Metastatic Melanoma. *Nat. Med.* 24 (10), 1545–1549. doi:10.1038/s41591-018-0157-9
- Ayres, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN- γ -related mRNA Profile Predicts Clinical Response to PD-1 Blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/JCI91190
- Balachandran, V. P., Gonen, M., Smith, J. J., and DeMatteo, R. P. (2015). Nomograms in Oncology: More Than Meets the Eye. *Lancet Oncol.* 16 (4), e173–e180. doi:10.1016/s1470-2045(14)71116-7
- Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune

and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1). doi:10.1186/s13059-016-1070-5

Blanche, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-dependent Areas under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Statist. Med.* 32 (30), 5381–5397. doi:10.1002/sim.5958

Bui, Q. T., Im, J. H., Jeong, S. B., Kim, Y.-M., Lim, S. C., Kim, B., et al. (2017). Essential Role of Notch4/STAT3 Signaling in Epithelial-Mesenchymal Transition of Tamoxifen-Resistant Human Breast Cancer. *Cancer Lett.* 390, 115–125. doi:10.1016/j.canlet.2017.01.014

Calle-Fabregat, C. d. I., Morante-Palacios, O., and Ballestar, E. (2020). Understanding the Relevance of DNA Methylation Changes in Immune Differentiation and Disease. *Genes* 11 (1), 110. doi:10.3390/genes11010110

Cardoso, F., van't Veer, L. J., Bogaerts, J., Slaets, L., Viale, G., Delaloge, S., et al. (2016). 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N. Engl. J. Med.* 375 (8), 717–729. doi:10.1056/NEJMoa1602253

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, X., Xiong, D., Ye, L., Wang, K., Huang, L., Mei, S., et al. (2019). Up-regulated lncRNA XIST Contributes to Progression of Cervical Cancer via Regulating miR-140-5p and ORC1. *Cancer Cel Int* 19, 45. doi:10.1186/s12935-019-0744-y

FUNDING

This work was supported by the National Key Research and Development Program (No. 2020YFA0712400), Special Foundation for Taishan Scholars (No. ts20190971), National Natural Science Foundation of China (No. 81874119; No. 82072912), Special Support Plan for National High Level Talents (Ten Thousand Talents Program W01020103), National Key Research and Development Program (No. 2018YFC0114705), Foundation from Clinical Research Center of Shandong University (No.2020SDUCRCA015), and the Qilu Hospital Clinical New Technology Developing Foundation (No. 2018-7; No. 2019-3).

ACKNOWLEDGMENTS

Special thanks to Yi Duan and Zhenqiu Liu. Additionally, we acknowledge TCGA and GEO databases for providing their platforms and contributors for uploading their meaningful datasets.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.675198/full#supplementary-material>

- Chung, W., Eum, H. H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., et al. (2017). Single-cell RNA-Seq Enables Comprehensive Tumour and Immune Cell Profiling in Primary Breast Cancer. *Nat. Commun.* 8 (1), 15081. doi:10.1038/ncomms15081
- Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507
- Dai, W., Teodoridis, J. M., Zeller, C., Graham, J., Hersey, J., Flanagan, J. M., et al. (2011). Systematic CpG Islands Methylation Profiling of Genes in the Wnt Pathway in Epithelial Ovarian Cancer Identifies Biomarkers of Progression-free Survival. *Clin. Cancer Res.* 17 (12), 4052–4062. doi:10.1158/1078-0432.Ccr-10-3021
- del Rincón, S. V., Widschwendter, M., Sun, D., Ekholm-Reed, S., Tat, J., Teixeira, L. K., et al. (2014). Cks Overexpression Enhances Chemotherapeutic Efficacy by Overriding DNA Damage Checkpoints. *Oncogene* 34 (15), 1961–1967. doi:10.1038/onc.2014.137
- Drake, T. M., and Søreide, K. (2019). Cancer Epigenetics in Solid Organ Tumours: A Primer for Surgical Oncologists. *Eur. J. Surg. Oncol.* 45 (5), 736–746. doi:10.1016/j.ejso.2019.02.005
- Du, T., Liu, B., Wang, Z., Wan, X., and Wu, Y. (2019). CpG Methylation Signature Predicts Prognosis in Breast Cancer. *Breast Cancer Res. Treat.* 178 (3), 565–572. doi:10.1007/s10549-019-05417-3
- Fan, C. N., Ma, L., and Liu, N. (2019). Comprehensive Analysis of Novel Three-long Noncoding RNA Signatures as a Diagnostic and Prognostic Biomarkers of Human Triple-negative Breast Cancer. *J. Cel Biochem* 120 (3), 3185–3196. doi:10.1002/jcb.27584
- Fitzgerald, M., Saville, B. R., and Lewis, R. J. (2015). Decision Curve Analysis. *JAMA* 313 (4), 409–410. doi:10.1001/jama.2015.37
- Fox, N. S., Haider, S., Harris, A. L., and Boutros, P. C. (2019). Landscape of Transcriptomic Interactions between Breast Cancer and its Microenvironment. *Nat. Commun.* 10 (1), 3116. doi:10.1038/s41467-019-10929-z
- Franco, R., Schoneveld, O., Georgakilas, A. G., and Panayiotidis, M. I. (2008). Oxidative Stress, DNA Methylation and Carcinogenesis. *Cancer Lett.* 266 (1), 6–11. doi:10.1016/j.canlet.2008.02.026
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01
- Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468
- Geyer, C. E., Tang, G., Mamounas, E. P., Rastogi, P., Paik, S., Shak, S., et al. (2018). 21-Gene Assay as Predictor of Chemotherapy Benefit in HER2-Negative Breast Cancer. *npj Breast Cancer* 4 (1), 37. doi:10.1038/s41523-018-0090-6
- Giuli, M. V., Giuliani, E., Screpanti, I., Bellavia, D., and Checquolo, S. (2019). Notch Signaling Activation as a Hallmark for Triple-Negative Breast Cancer Subtype. *J. Oncol.* 2019, 1–15. doi:10.1155/2019/8707053
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7
- Hao, X., Luo, H., Krawczyk, M., Wei, W., Wang, W., Wang, J., et al. (2017). DNA Methylation Markers for Diagnosis and Prognosis of Common Cancers. *Proc. Natl. Acad. Sci. USA* 114 (28), 7414–7419. doi:10.1073/pnas.1703577114
- Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist. Med.* 15 (4), 361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4
- He, Y., Li, X., Meng, Y., Fu, S., Cui, Y., Shi, Y., et al. (2019). A Prognostic 11 Long Noncoding RNA Expression Signature for Breast Invasive Carcinoma. *J. Cel Biochem* 120 (10), 16692–16702. doi:10.1002/jcb.28927
- Holm, K., Staaf, J., Lauss, M., Aine, M., Lindgren, D., Bendahl, P.-O., et al. (2016). An Integrated Genomics Analysis of Epigenetic Subtypes in Human Breast Tumors Links DNA Methylation Patterns to Chromatin States in normal Mammary Cells. *Breast Cancer Res.* 18 (1), 27. doi:10.1186/s13058-016-0685-5
- Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2007). Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets. *PLoS One* 2 (11), e1195. doi:10.1371/journal.pone.0001195
- Hu, J., Yu, A., Othmane, B., Qiu, D., Li, H., Li, C., et al. (2021). Siglec15 Shapes a Non-inflamed Tumor Microenvironment and Predicts the Molecular Subtype in Bladder Cancer. *Theranostics* 11 (7), 3089–3108. doi:10.7150/thno.53649
- Jeschke, J., Bizet, M., Desmedt, C., Calonne, E., Dedeurwaerder, S., Garaud, S., et al. (2017). DNA Methylation-Based Immune Response Signature Improves Patient Diagnosis in Multiple Cancers. *J. Clin. Invest.* 127 (8), 3090–3102. doi:10.1172/jci91095
- Jeschke, J., Collignon, E., and Fuks, F. (2015). DNA Methylome Profiling beyond Promoters - Taking an Epigenetic Snapshot of the Breast Tumor Microenvironment. *Febs J.* 282 (9), 1801–1814. doi:10.1111/febs.13125
- Jia, H., Truica, C. I., Wang, B., Wang, Y., Ren, X., Harvey, H. A., et al. (2017). Immunotherapy for Triple-Negative Breast Cancer: Existing Challenges and Exciting Prospects. *Drug Resist. Updates* 32, 1–15. doi:10.1016/j.drug.2017.07.002
- Jiang, T., Zhou, M.-L., and Fan, J. (2018). Inhibition of GLUT-1 Expression and the PI3K/Akt Pathway to Enhance the Chemosensitivity of Laryngeal Carcinoma Cells *In Vitro*. *Ott Vol.* 11, 7865–7872. doi:10.2147/ott.S176818
- Jiang, Y.-Z., Ma, D., Suo, C., Shi, J., Xue, M., Hu, X., et al. (2019). Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell* 35 (3), 428–440.e425. doi:10.1016/j.ccell.2019.02.001
- Jordan, V. C. (2015). The New Biology of Estrogen-Induced Apoptosis Applied to Treat and Prevent Breast Cancer. *Endocr. Relat. Cancer* 22 (1), R1–R31. doi:10.1530/erc-14-0448
- Kara, N., Hossain, M., Prasanth, S. G., and Stillman, B. (2015). Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. *J. Biol. Chem.* 290 (19), 12355–12369. doi:10.1074/jbc.M114.625012
- Keeley, B., Stark, A., Pisanic, T. R., 2nd, Kwak, R., Zhang, Y., Wrangle, J., et al. (2013). Extraction and Processing of Circulating DNA from Large Sample Volumes Using Methylation on Beads for the Detection of Rare Epigenetic Events. *Clinica Chim. Acta* 425, 169–175. doi:10.1016/j.cca.2013.07.023
- Khaled, N., and Bidet, Y. (2019). New Insights into the Implication of Epigenetic Alterations in the EMT of Triple Negative Breast Cancer. *Cancers* 11 (4), 559. doi:10.3390/cancers11040559
- Kim, S.-H., and Singh, S. V. (2015). The Role of Polycomb Group Protein Bmi-1 and Notch4 in Breast Cancer Stem Cell Inhibition by Benzyl Isothiocyanate. *Breast Cancer Res. Treat.* 149 (3), 681–692. doi:10.1007/s10549-015-3279-5
- Le Calvé, B., Rynkowski, M., Le Mercier, M., Bruyère, C., Lonz, C., Gras, T., et al. (2010). Long-term *In Vitro* Treatment of Human Glioblastoma Cells with Temozolomide Increases Resistance *In Vivo* through Up-Regulation of GLUT Transporter and Aldo-Keto Reductase Enzyme AKR1C Expression. *Neoplasia* 12 (9), 727–739. doi:10.1593/neo.10526
- Li, J., Liu, C., Chen, Y., Gao, C., Wang, M., Ma, X., et al. (2019a). Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated with Prognosis. *Front. Genet.* 10, 1119. doi:10.3389/fgene.2019.01119
- Li, N., Deng, Y., Zhou, L., Tian, T., Yang, S., Wu, Y., et al. (2019b). Global burden of Breast Cancer and Attributable Risk Factors in 195 Countries and Territories, from 1990 to 2017: Results from the Global Burden of Disease Study 2017. *J. Hematol. Oncol.* 12 (1), 140. doi:10.1186/s13045-019-0828-0
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* 27 (12), 1739–1740. doi:10.1093/bioinformatics/btr260
- Lin, L., Yan, L., Liu, Y., Yuan, F., Li, H., and Ni, J. (2019). Incidence and Death in 29 Cancer Groups in 2017 and Trend Analysis from 1990 to 2017 from the Global Burden of Disease Study. *J. Hematol. Oncol.* 12 (1), 96. doi:10.1186/s13045-019-0783-9
- Lombardo, Y., Faronato, M., Filipovic, A., Vircillo, V., Magnani, L., and Coombes, R. C. (2014). Nicastrin and Notch4 Drive Endocrine Therapy Resistance and Epithelial to Mesenchymal Transition in MCF7 Breast Cancer Cells. *Breast Cancer Res.* 16 (3), R62. doi:10.1186/bcr3675
- Lord, C. J., and Ashworth, A. (2012). The DNA Damage Response and Cancer Therapy. *Nature* 481 (7381), 287–294. doi:10.1038/nature10760

- Lu, X., Jiang, L., Zhang, L., Zhu, Y., Hu, W., Wang, J., et al. (2019). Immune Signature-Based Subtypes of Cervical Squamous Cell Carcinoma Tightly Associated with Human Papillomavirus Type 16 Expression, Molecular Features, and Clinical Outcome. *Neoplasia* 21 (6), 591–601. doi:10.1016/j.neo.2019.04.003
- Luo, H., Zhao, Q., Wei, W., Zheng, L., Yi, S., Li, G., et al. (2020). Circulating Tumor DNA Methylation Profiles Enable Early Diagnosis, Prognosis Prediction, and Screening for Colorectal Cancer. *Sci. Transl. Med.* 12 (524), eaax7533. doi:10.1126/scitranslmed.aax7533
- Ma, L., and Zong, X. (2020). Metabolic Symbiosis in Chemoresistance: Refocusing the Role of Aerobic Glycolysis. *Front. Oncol.* 10, 5. doi:10.3389/fonc.2020.00005
- Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173 (2), 338–e15.e315. doi:10.1016/j.cell.2018.03.034
- Mamounas, E. P., Tang, G., Fisher, B., Paik, S., Shak, S., Costantino, J. P., et al. (2010). Association between the 21-Gene Recurrence Score Assay and Risk of Locoregional Recurrence in Node-Negative, Estrogen Receptor-Positive Breast Cancer: Results from NSABP B-14 and NSABP B-20. *Jco* 28 (10), 1677–1683. doi:10.1200/jco.2009.23.7610
- Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118
- McNairn, A. J., and Gilbert, D. M. (2005). Overexpression of ORC Subunits and Increased ORC-Chromatin Association in Transformed Mammalian Cells. *J. Cel. Biochem.* 96 (5), 879–887. doi:10.1002/jcb.20609
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 Facilitates Sensitive and Confident Localization of the Targets of Focal Somatic Copy-Number Alteration in Human Cancers. *Genome Biol.* 12 (4), R41. doi:10.1186/gb-2011-12-4-r41
- Nagamatsu, I., Onishi, H., Matsushita, S., Kubo, M., Kai, M., Imaizumi, A., et al. (2014). NOTCH4 Is a Potential Therapeutic Target for Triple-Negative Breast Cancer. *Anticancer Res.* 34 (1), 69–80.
- Network, C. G. A. (2012). Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* 490 (7418), 61–70. doi:10.1038/nature11412
- Olsson, N., Carlsson, P., James, P., Hansson, K., Waldemarson, S., Malmström, P., et al. (2013). Grading Breast Cancer Tissues Using Molecular Portraits. *Mol. Cell Proteomics* 12 (12), 3612–3623. doi:10.1074/mcp.M113.030379
- Paik, Y.-K., and Hancock, W. S. (2012). Uniting ENCODE with Genome-wide Proteomics. *Nat. Biotechnol.* 30 (11), 1065–1067. doi:10.1038/nbt.2416
- Pasculli, B., Barbano, R., and Parrella, P. (2018). Epigenetics of Breast Cancer: Biology and Clinical Implication in the Era of Precision Medicine. *Semin. Cancer Biol.* 51, 22–35. doi:10.1016/j.semcancer.2018.01.007
- Patel, S., Alam, A., Pant, R., and Chattopadhyay, S. (2019). Wnt Signaling and its Significance within the Tumor Microenvironment: Novel Therapeutic Insights. *Front. Immunol.* 10, doi:10.3389/fimmu.2019.02872
- Plava, J., Cihova, M., Burikova, M., Matuskova, M., Kucerova, L., and Miklikova, S. (2019). Recent Advances in Understanding Tumor Stroma-Mediated Chemoresistance in Breast Cancer. *Mol. Cancer* 18 (1), 67. doi:10.1186/s12943-019-0960-z
- Qu, X. A., and Rajpal, D. K. (2012). Applications of Connectivity Map in Drug Discovery and Development. *Drug Discov. Today* 17 (23–24), 1289–1298. doi:10.1016/j.drudis.2012.07.017
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14 (2), 482–517. doi:10.1038/s41596-018-0103-9
- Reis-Filho, J. S., and Pusztai, L. (2011). Gene Expression Profiling in Breast Cancer: Classification, Prognostication, and Prediction. *The Lancet* 378 (9805), 1812–1823. doi:10.1016/s0140-6736(11)61539-0
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160 (1–2), 48–61. doi:10.1016/j.cell.2014.12.033
- Sia, D., Jiao, Y., Martinez-Quetglas, I., Kuchuk, O., Villacorta-Martin, C., Castro de Moura, M., et al. (2017). Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* 153 (3), 812–826. doi:10.1053/j.gastro.2017.06.007
- Silwal-Pandit, L., Langerød, A., and Børresen-Dale, A.-L. (2017). TP53 Mutations in Breast and Ovarian Cancer. *Cold Spring Harb Perspect. Med.* 7 (1), a026252. doi:10.1101/cshperspect.a026252
- Simões, B. M., O'Brien, C. S., Eyre, R., Silva, A., Yu, L., Sarmiento-Castro, A., et al. (2015). Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-dependent Cancer Stem Cell Activity. *Cel Rep.* 12 (12), 1968–1977. doi:10.1016/j.celrep.2015.08.050
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Soft.* 39 (5), 1–13. doi:10.18637/jss.v039.i05
- Sparano, J. A., Gray, R. J., Makower, D. F., Pritchard, K. I., Albain, K. S., Hayes, D. F., et al. (2015). Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N. Engl. J. Med.* 373 (21), 2005–2014. doi:10.1056/NEJMoa1510764
- Spronk, I., Schellevis, F. G., Burgers, J. S., de Bock, G. H., and Korevaar, J. C. (2018). Incidence of Isolated Local Breast Cancer Recurrence and Contralateral Breast Cancer: A Systematic Review. *The Breast* 39, 70–79. doi:10.1016/j.breast.2018.03.011
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171 (6), 1437–1452.e1417. doi:10.1016/j.cell.2017.10.049
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Szekely, B., Bossuyt, V., Li, X., Wali, V. B., Patwardhan, G. A., Frederick, C., et al. (2018). Immunological Differences between Primary and Metastatic Breast Cancer. *Ann. Oncol.* 29 (11), 2232–2239. doi:10.1093/annonc/mdy399
- Szyf, M. (2012). DNA Methylation Signatures for Breast Cancer Classification and Prognosis. *Genome Med.* 4 (3), 26. doi:10.1186/gm325
- Tao, C., Luo, R., Song, J., Zhang, W., and Ran, L. (2019). A seven-DNA Methylation Signature as a Novel Prognostic Biomarker in Breast Cancer. *J. Cel Biochem* 121 (3), 2385–2393. doi:10.1002/jcb.29461
- Tazaki, E., Shishido-Hara, Y., Mizutani, N., Nomura, S., Isaka, H., Ito, H., et al. (2013). Histopathological and Clonal Study of Combined Lobular and Ductal Carcinoma of the Breast. *Pathol. Int.* 63 (6), 297–304. doi:10.1111/pin.12065
- Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48 (4), 812–e14.e814. doi:10.1016/j.immuni.2018.03.023
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *wo* 1A, 68–77. doi:10.5114/wo.2014.47136
- Vickers, A. J., and Elkin, E. B. (2016). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Making* 26 (6), 565–574. doi:10.1177/0272989x06295361
- Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent among Samples. *Theor. Biosci.* 131 (4), 281–285. doi:10.1007/s12064-012-0162-3
- Wang, J. W., Wei, X. L., Dou, X. W., Huang, W. H., Du, C. W., and Zhang, G. J. (2018). The Association between Notch4 Expression, and Clinicopathological Characteristics and Clinical Outcomes in Patients with Breast Cancer. *Oncol. Lett.* doi:10.3892/ol.2018.8442
- Wang, Y., Deng, H., Xin, S., Zhang, K., Shi, R., and Bao, X. (2019). Prognostic and Predictive Value of Three DNA Methylation Signatures in Lung Adenocarcinoma. *Front. Genet.* 10, doi:10.3389/fgene.2019.00349
- Wei, L., Jin, Z., Yang, S., Xu, Y., Zhu, Y., Ji, Y., et al. (2018). TCGA-assembler 2: Software Pipeline for Retrieval and Processing of TCGA/CPTAC Data. *Bioinformatics* 34 (9), 1615–1617. doi:10.1093/bioinformatics/btx812
- Wu, Q., and Ni, X. (2015). ROS-mediated DNA Methylation Pattern Alterations in Carcinogenesis. *Cdt* 16 (1), 13–19. doi:10.2174/1389450116666150113121054
- Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 78 (23), 6575–6580. doi:10.1158/0008-5472.CAN-18-0689
- Xuchao, Z., Zhiyong, L., Shengyue, W., Shun, L., Yong, S., Ying, C., et al. (2019). Application of Next-Generation Sequencing Technology to Precision Medicine in Cancer: Joint Consensus of the Tumor Biomarker Committee of the Chinese Society of Clinical Oncology. *Cancer Biol. Med.* 16 (1), 189. doi:10.20892/j.issn.2095-3941.2018.0142

- Yang, Y., Wu, L., Shu, X.-O., Cai, Q., Shu, X., Li, B., et al. (2020). Genetically Predicted Levels of DNA Methylation Biomarkers and Breast Cancer Risk: Data from 228 951 Women of European Descent. *J. Natl. Cancer Inst.* 112 (3), 295–304. doi:10.1093/jnci/djz109
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612
- Zhang, L., Feizi, N., Chi, C., and Hu, P. (2018a). Association Analysis of Somatic Copy Number Alteration Burden with Breast Cancer Survival. *Front. Genet.* 9, 421. doi:10.3389/fgene.2018.00421
- Zhang, L., Zhao, Y., Dai, Y., Cheng, J.-N., Gong, Z., Feng, Y., et al. (2018b). Immune Landscape of Colorectal Cancer Tumor Microenvironment from Different Primary Tumor Location. *Front. Immunol.* 9, 1578. doi:10.3389/fimmu.2018.01578
- Zhang, S., Wang, Y., Gu, Y., Zhu, J., Ci, C., Guo, Z., et al. (2018c). Specific Breast Cancer Prognosis-subtype Distinctions Based onDNAmethylation Patterns. *Mol. Oncol.* 12 (7), 1047–1060. doi:10.1002/1878-0261.12309
- Zhang, Z., and Kattan, M. W. (2017). Drawing Nomograms with R: Applications to Categorical Outcome and Survival Data. *Ann. Transl. Med.* 5 (10), 211. doi:10.21037/atm.2017.04.01
- Zhang, Z., Rousson, V., Rousson, V., Lee, W.-C., Ferdynus, C., Chen, M., et al. (2018d). Decision Curve Analysis: a Technical Note. *Ann. Transl. Med.* 6 (15), 308. doi:10.21037/atm.2018.07.02

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Wang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.