Ann Surg Oncol (2019) 26:3344–3353
https://doi.org/10.1245/s10434-019-07565-8

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – BREAST ONCOLOGY

Clinicopathological Features of Triple-Negative Breast Cancer Epigenetic Subtypes

Maggie L. DiNome, MD[1], Javier I. J. Orozco, MD[2], Chikako Matsuba, PhD[3], Ayla O. Manughian-Peter, MSc[2], Miquel Ensenyat-Mendez, MSc[4], Shu-Ching Chang, PhD[5], John R. Jalas, MD[6], Matthew P. Salomon, PhD[3], and Diego M. Marzese, PhD[2]

[1]Department of Surgery, David Geffen School of Medicine, University California Los Angeles (UCLA), Los Angeles, CA; [2]Cancer Epigenetics Laboratory, John Wayne Cancer Institute at Providence Saint John's Health Center, Santa Monica, CA; [3]Computational Biology Laboratory, John Wayne Cancer Institute at Providence St. John's Health Center, Santa Monica, CA; [4]Cancer Cell Biology Group, Balearic Islands Health Research Institute (IdISBa), Palma, Islas Baleares, Spain; [5]Medical Data Research Center, Providence Saint Joseph Health, Portland, OR; [6]Department of Pathology, Providence Saint John's Health Center, Santa Monica, CA

ABSTRACT

Background/Objective. Triple-negative breast cancer (TNBC) is a heterogeneous collection of breast tumors with numerous differences including morphological characteristics, genetic makeup, immune-cell infiltration, and response to systemic therapy. DNA methylation profiling is a robust tool to accurately identify disease-specific subtypes. We aimed to generate an epigenetic subclassification of TNBC tumors (epitypes) with utility for clinical decision-making.

Methods. Genome-wide DNA methylation profiles from TNBC patients generated in the Cancer Genome Atlas project were used to build machine learning-based epigenetic classifiers. Clinical and demographic variables, as well as gene expression and gene mutation data from the same cohort, were integrated to further refine the TNBC epitypes.

Results. This analysis indicated the existence of four TNBC epitypes, named as Epi-CL-A, Epi-CL-B, Epi-CL-C, and Epi-CL-D. Patients with Epi-CL-B tumors showed significantly shorter disease-free survival and overall survival [log rank; $P = 0.01$; hazard ratio (HR) 3.89, 95% confidence interval (CI) 1.3–11.63 and $P = 0.003$; HR 5.29, 95% CI 1.55–18.18, respectively]. Significant gene expression and mutation differences among the TNBC epitypes suggested alternative pathway activation that could be used as ancillary therapeutic targets. These epigenetic subtypes showed complementarity with the recently described TNBC transcriptomic subtypes.

Conclusions. TNBC epigenetic subtypes exhibit significant clinical and molecular differences. The links between genetic make-up, gene expression programs, and epigenetic subtypes open new avenues in the development of laboratory tests to more efficiently stratify TNBC patients, helping optimize tailored treatment approaches.

BACKGROUND

Triple-negative breast cancer (TNBC) is defined by the absence of estrogen receptor (ER) and progesterone receptor (PgR) expression and the lack of amplification and/or overexpression of human epidermal growth factor receptor 2 (HER2). This definition has resulted in a heterogeneous collection of tumors. Effective treatment of this subgroup, which constitutes 15–20% of breast cancers,[1] poses a clinical challenge. With no specific driver

Maggie L. DiNome and Javier I. J. Orozco have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1245/s10434-019-07565-8) contains supplementary material, which is available to authorized users.

First Received: 22 April 2019;
Published Online: 24 July 2019

D. M. Marzese, PhD
e-mail: MarzeseD@jwci.org

molecular alterations identified, cytotoxic chemotherapy remains the mainstay of therapy, but rates of metastasis for TNBC continue to be high. Identifying and characterizing TNBC subtypes would be an important step toward developing a more effective, targeted treatment strategy.

Recent gene expression analyses have categorized TNBC into distinct subtypes.[2–5] Lehmann et al. initially classified TNBC into six distinct subtypes,[1] but concern over tumor impurities caused by stromal cells and immune infiltration led to a refinement on this classification system.[6] In a similar and partially overlapping classification approach, Burstein et al. segregated TNBC into four transcriptomic subtypes.[2] Notwithstanding, these seminal classifiers have highlighted that distinct subtypes within TNBC exist and have served as the basis for additional preclinical work. In this context, further improvements of the classification system are necessary to enhance clinical usefulness. Assaying DNA methylation (DNAm) has been shown to be a robust tool for accurately identifying tumor-specific subtypes.[7–15] Diagnostic, prognostic, and predictive characteristics can be derived from DNAm signatures. DNAm profiling can also be performed using small amounts of routinely collected DNA samples, making it a readily applicable tool for the clinic. Our studies implementing supervised learning for DNAm data have resulted in a powerful set of epigenetic classifiers able to accurately stratify therapeutically relevant breast cancer subtypes in brain metastases.[13]

The objective of this study is to generate a machine learning-based epigenetic classification to better understand the molecular and clinical heterogeneity of TNBC, and to determine whether distinct epigenetic subtypes exist. To this end, we explored a large collection of clinically well-annotated DNA methylomes from the Cancer Genome Atlas (TCGA) breast cancer project.

METHODS

Patient Selection

We identified 161 samples from patients with primary stage I–III TNBC from TCGA. Of the 1108 breast cancer samples included in the TCGA-BRCA database from cBioPortal,[16] we excluded specimens obtained from metastatic sites ($n = 7$), breast sarcoma ($n = 2$), unavailable cancer types ($n = 5$), unavailable or indeterminate ER, PgR or HER2 statuses ($n = 163$), and samples from patients with de novo stage IV breast cancer ($n = 2$) (Fig. 1a).

Data Access, Normalization, and Processing

TCGA-BRCA data deposited in the National Cancer Institute TCGA Data Portal were accessed using the *TCGAbiolinks* v1.2.5 R/Bioconductor package[17] in August 2018. A total of 119 DNA methylomes were retrieved from patients with TNBC subtype using the function *GDCdownload*. These analyses were compiled using the *GDCprepare* function into a single data file. Each of the successfully downloaded cases was assayed for overall Illumina Human Methylation 450 K beadchip (HM450K) data quality and effective tumor purity using the consensus estimate purity (CPE) algorithm.[18] Only cases with tumor content of at least 60% ($n = 79$) were selected (Fig. 1a).

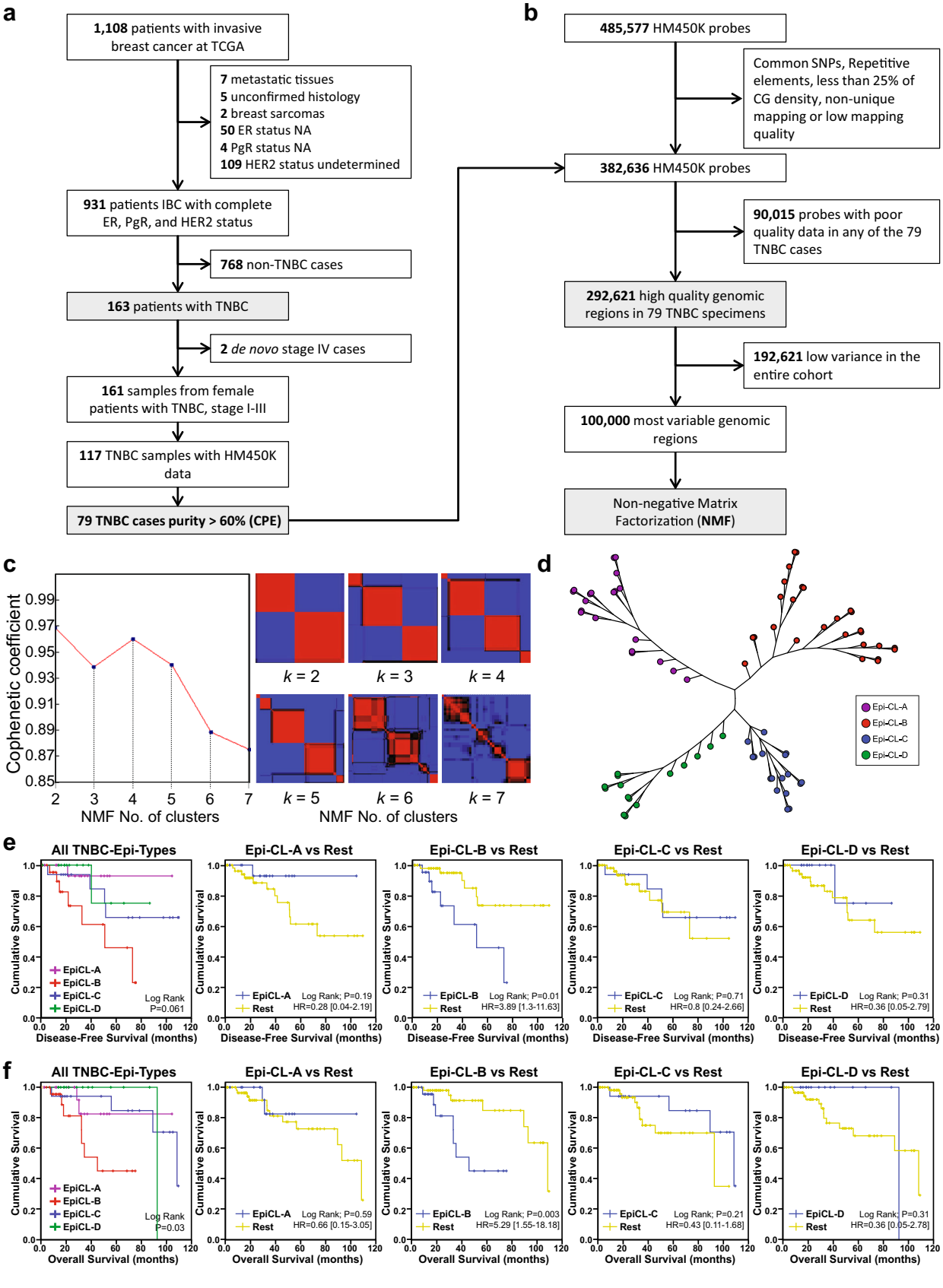
FIG. 1 Machine learning-based identification of TNBC DNA methylation subtypes. **a**, **b** Diagram describing the selection of TNBC patients from TCGA (**a**) and the normalization and filtering algorithms of Infinium HM450K probes (**b**) to generate the datasets to train the machine learning methods. **c** Nonnegative matrix factorization (NMF) analysis of the TNBC DNA methylomes. The plot indicates the variation of the cophenetic coefficient depending on the number of clusters; the heatmaps show the potential organization of the TNBC specimens according to the number of clusters (k). **d** Phylogenetic tree showing the distribution of the TNBC specimens ($n = 79$) classified into four TNBC epitypes using the uncorrelated shrunken centroid-based DNA methylation signature ($n = 103$ genomic regions, Supplementary Table S1). **e**, **f** Comparison of disease-free survival (**e**) and overall survival (**f**) intervals between the different TNBC epigenetic subtypes ▶

Identification of TNBC Epigenetic Subtypes

DNAm data from selected TNBC cases were extracted from Illumina.idat files using the Bioconductor package *minfi*.[19] The *preprocessNoob* function in *minfi* was used for normalization and dye-bias correction. DNAm levels were reported as β-values and calculated using the signal intensity value for each CpG site, as we previously showed.[13,20] HM450K probes with poor quality were excluded from the analysis. The nonnegative matrix factorization (NMF) machine learning approach was employed to identify the optimal number of subtypes based on the DNAm data of the top 100,000 most variable genomic regions. The uncorrelated shrunken centroids (USC) method [initial settings: delta value below 5 ($\Delta < 5$) and maximum correlation between the genomic regions = 0.7] was employed to select the most informative genomic regions to classify TNBC specimens into each epigenetic subtype (epitype).

Transcriptomic Classification

TNBC cases were classified using the algorithm proposed by Burstein et al.[3] into basal-like immune-activated (BLIA), basal-like immunosuppressed (BLIS), luminal androgen receptor (LAR), and mesenchymal-like (MES). Whole transcriptome profiles of the same cases were submitted to the TNBCtype tool (http://cbc.mc.vanderbilt.edu/tnbc/prediction.php)[21] to


a
1,108 patients with invasive breast cancer at TCGA
7 metastatic tissues
5 unconfirmed histology
2 breast sarcomas
50 ER status NA
4 PgR status NA
109 HER2 status undetermined
931 patients IBC with complete ER, PgR, and HER2 status
768 non-TNBC cases
163 patients with TNBC
2 de novo stage IV cases
161 samples from female patients with TNBC, stage I-III
117 TNBC samples with HM450K data
79 TNBC cases purity > 60% (CPE)
b
485,577 HM450K probes
Common SNPs, Repetitive elements, less than 25% of CG density, non-unique mapping or low mapping quality
382,636 HM450K probes
90,015 probes with poor quality data in any of the 79 TNBC cases
292,621 high quality genomic regions in 79 TNBC specimens
192,621 low variance in the entire cohort
100,000 most variable genomic regions
Non-negative Matrix Factorization (NMF)
c
Cophenetic coefficient
NMF No. of clusters
k = 2
k = 3
k = 4
k = 5
k = 6
k = 7
d
Epi-CL-A
Epi-CL-B
Epi-CL-C
Epi-CL-D
e
All TNBC-Epi-Types
Epi-CL-A vs Rest
Epi-CL-B vs Rest
Epi-CL-C vs Rest
Epi-CL-D vs Rest
Cumulative Survival
Disease-Free Survival (months)
EpiCL-A
EpiCL-B
EpiCL-C
EpiCL-D
Rest
Log Rank P=0.061
Log Rank; P=0.19
HR=0.28 [0.04-2.19]
Log Rank; P=0.01
HR=3.89 [1.3-11.63]
Log Rank; P=0.71
HR=0.8 [0.24-2.66]
Log Rank; P=0.31
HR=0.36 [0.05-2.79]
f
All TNBC-Epi-Types
Epi-CL-A vs Rest
Epi-CL-B vs Rest
Epi-CL-C vs Rest
Epi-CL-D vs Rest
Cumulative Survival
Overall Survival (months)
EpiCL-A
EpiCL-B
EpiCL-C
EpiCL-D
Rest
Log Rank P=0.03
Log Rank; P=0.59
HR=0.66 [0.15-3.05]
Log Rank; P=0.003
HR=5.29 [1.55-18.18]
Log Rank; P=0.21
HR=0.43 [0.11-1.68]
Log Rank; P=0.31
HR=0.36 [0.05-2.78]

identify correlations with the centroids defining each of the six TNBC transcriptomic subtypes proposed by Lehmann:[2] basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR). PAM50 genomic signature classification of TNBC specimens was retrieved from TCGA. Cases with missing PAM50 annotation were classified using the *genefu* v1.12.0 R/Bioconductor package.

Bioinformatics Analyses

The *oncoprint* function of the R/Bioconductor package *TCGAbiolinks* was used to explore the mutational background and gene expression profile of each TNBC epigenetic subtype. Gene interaction network and Gene Ontology (GO) enrichments were analyzed using GeneMania.[22] The molecular-function-based weighting method was chosen for analysis with network categories of coexpression and pathways. The most significantly enriched pathways obtained from the GeneMania analysis were evaluated using Gene2drug,[23] and the top 10 most significant compounds with negative score (downregulation of the pathways) for each cluster were selected.

Biostatistics Analyses

Baseline characteristics were compared using Fisher's exact test and analysis of variance (ANOVA) or Kruskal–Wallis rank-sum test for categorical and continuous variables, respectively. The endpoint for overall survival (OS) represents the time to mortality or last follow-up date, while disease-free survival (DFS) represents the time to event related to the disease itself or death from disease-specific event (i.e., non-disease deaths are censored). Kaplan–Meier curves with log-rank test were generated to identify differences in survival rates, and Cox hazard models were used to assess the hazard ratio (HR). Cumulative incidence function with competing risks regression analysis were used to evaluate the differences in DFS, taking the competing risk of mortality into account, using the function *crr* in the statistical R package *cmprsk*.[24,25] Subdistribution hazard ratio (SHR) from the competing risk regression analysis was calculated for each epitype, taking the risk of noncancer deaths into account. TNBC epigenetic subtypes were compared with the transcriptomic subtypes using contingency tables, and the statistical significance of the differences observed in the proportions was estimated using the Chi square (χ^2) tests. Cramer's V coefficient was used to measure the strength of the associations.

RESULTS

Identification of Triple-Negative Breast Cancer Epigenetic Endotypes

From the total number of assayed genomic regions on the HM450K (n = 485,577), we initially excluded probes with potential design issues (n = 102,941) or poor quality (n = 90,015). A total of 292,621 genomic regions in 79 cases were employed as the initial dataset (Fig. 1b), selected based on the overall variance. The top 100,000 most variable genomic regions were selected to perform machine learning analysis using the NMF technique.[26] This analysis revealed four distinct groupings (k = 4; cophenetic coefficient c = 0.961; Fig. 1c). Using the USC supervised learning method to exclude highly correlated genomic regions, we produced a DNAm signature including 103 genomic regions (Supplementary Table S1) to classify the remaining cases (Δ = 0; ρ = 0.5). This DNAm signature classified the 79 tumors into four TNBC epitypes: Epi-CL-A, Epi-CL-B, Epi-CL-C, and Epi-CL-D (Fig. 1d).

Clinical and Demographic Distribution of TNBC Epigenetic Subtypes

Baseline characteristics of the patients with the four TNBC epigenetic subtypes are summarized in Table 1. There were no significant differences in patient demographics or tumor features among the subtypes. Patients with TNBC Epi-CL-B presented significantly shorter DFS (log rank; P = 0.01; HR 3.89, 95% CI 1.3–11.63; Fig. 1e) and shorter OS (log rank; P = 0.003; HR 5.29, 95% CI 1.55–18.18; Fig. 1f) than the rest of the patients. After controlling for age, race, and pathological T and N stages, multivariable competing risk regression and Cox proportional-hazards regression analyses showed that Epi-CL-B TNBC subtype is an independent predictor of DFS (SHR 4.0, 95% CI 1.07–14.92; P = 0.04, Table 2) and OS (HR 10.49, 95% CI 2.28–48.14; P = 0.003; Table 2).

Integration of TNBC Transcriptomic and Epigenetic Subtypes

The majority of TNBC specimens showed a PAM50 basal-like signature; the Epi-CL-B subtype showed a significant association with HER2-enriched TNBC specimens (χ^2, P = 0.004, Cramer's V = 0.350; Fig. 2a). We found a significant association between the epigenetic subtypes and the Lehmann et al. TNBC classification (χ^2, P = 0.0003, Cramer's V = 0.433). The Epi-CL-A subtype was almost entirely composed of TNBC specimens classified as mesenchymal (M, 83.3%, Fig. 2b). All specimens classified as LAR were included in the Epi-CL-B subtype (Fig. 2b).

TABLE 1 Baseline characteristics of the patients in each TNBC epitype

Baseline characteristic	Epi-CL-A	Epi-CL-B	Epi-CL-C	Epi-CL-D	Total	*P* Value*
No. (%)	20 (25.3)	26 (32.9)	17 (21.5)	16 (20.3)	79 (100)	
Age, mean (SD), years	54.4 (9.8)	56.5 (13.2)	57.5 (16.9)	52.8 (10.7)	55.4 (12.8)	0.70
Race						0.48
White non-Hispanic	8 (40)	14 (53.8)	10 (58.8)	5 (31.2)	37 (46.8)	
Black	9 (45)	10 (38.5)	5 (29.4)	7 (43.8)	31 (39.2)	
Asian	2 (10)	2 (7.7)	0 (0)	1 (6.2)	5 (6.3)	
White Hispanic	0 (0)	0 (0)	2 (11.8)	1 (6.2)	3 (3.8)	
NA	1 (5)	0 (0)	0 (0)	2 (12.5)	3 (3.8)	
T stage, *n* (%)						0.17
T1	5 (25)	11 (42.3)	4 (23.5)	1 (6.2)	21 (26.6)	
T2	13 (65)	12 (46.2)	12 (70.6)	11 (68.8)	48 (60.8)	
T3	2 (10)	2 (7.7)	0 (0)	3 (18.8)	7 (8.9)	
T4	0 (0)	1 (3.8)	1 (5.9)	1 (6.2)	3 (3.8)	
N stage, *n* (%)						0.11
N0	16 (80)	17 (65.4)	9 (52.9)	10 (62.5)	52 (65.8)	
N1	3 (15)	1 (3.8)	6 (35.3)	3 (18.8)	13 (16.5)	
N2	1 (5)	6 (23.1)	1 (5.9)	3 (18.8)	11 (13.9)	
N3	0 (0)	2 (7.7)	1 (5.9)	0 (0)	3 (3.8)	
AJCC stage, *n* (%)						0.06
I	4 (20)	7 (26.9)	2 (11.8)	1 (6.2)	14 (17.7)	
II	15 (75)	10 (38.5)	13 (76.5)	12 (75)	50 (63.3)	
III	1 (5)	9 (34.6)	2 (11.8)	3 (18.8)	15 (19)	
Histology, *n* (%)						0.67
Ductal	15 (75)	22 (84.6)	14 (82.4)	14 (87.5)	65 (82.3)	
Lobular	2 (10)	1 (3.8)	1 (5.9)	0 (0)	4 (5.1)	
Metaplastic	2 (10)	3 (11.5)	0 (0)	1 (6.2)	6 (7.6)	
Medullary	0 (0)	0 (0)	1 (5.9)	1 (6.2)	2 (2.5)	
Other	1 (5)	0 (0)	1 (5.9)	0 (0)	2 (2.5)	
PAM50 subtype, *n* (%)						< 0.001
Basal-like	20 (100)	18 (69.2)	15 (88.2)	15 (93.8)	68 (86.1)	
HER2-enriched	0 (0)	7 (26.9)	0 (0)	0 (0)	7 (8.9)	
Luminal A	0 (0)	0 (0)	0 (0)	1 (6.2)	1 (1.3)	
Luminal B	0 (0)	1 (3.8)	0 (0)	0 (0)	1 (1.3)	
Normal-like	0 (0)	0 (0)	2 (11.8)	0 (0)	2 (2.5)	

*ANOVA analysis or Kruskal–Wallis rank-sum test for continuous variable; Fisher exact test for categorical variables

Furthermore, we found a significant association between the epigenetic subtypes and the Burstein et al. TNBC classification (χ^2, $P = 0.03$, Cramer's $V = 0.363$). The Epi-CL-B subtype was associated with the BLIS subtype, which correlates with worse DFS and OS as shown by Burstein et al.[3] In concordance with the previous observation, we found an association between the Epi-CL-A and the MES subtype (Fig. 2c). Neither the Epi-CL-C nor D subtypes showed a significant association with Burstein et al. subtypes.

Based on exploration of differentially expressed genes (Supplementary Table S2) and differential pathway activation (Fig. 2), Epi-CL-A exhibited a significant upregulation of Wnt pathway-related genes (e.g., *LRP6*, *RNF8*), activation of Wnt signaling, and activation of mesenchymal cell differentiation and proliferation; these data are in agreement with the observed increased proportion of mesenchymal subtypes (Fig. 2b, c). In Epi-CL-B, some of the most differentially expressed genes were involved in maintaining chromatin structure (e.g., *KDM2B*) and DNA damage response (e.g., *ANKLE1*). Epi-CL-B showed significant enrichment of mitosis and cell division-related processes (Fig. 2d). *CDK2AP1*, a negative regulator of the *CDK2* oncogene,[27] was upregulated in Epi-CL-A but

TABLE 2 Independent predictors of disease-free survival and overall survival

Baseline characteristic	SHR	95% lower CI	Univariate analysis		Multivariable analysis			
			95% upper CI	*P* value	SHR	95% lower CI	95% upper CI	*P* value
Disease-free survival								
TNBC epitypes								
A/C/D	Ref				Ref			
B	3.16	1.09	9.18	0.04	2.11	0.66	6.73	0.21
Age (years)								
< 45	Ref				Ref			
≥ 45	0.55	0.18	1.72	0.31	0.7	0.14	3.57	0.67
Race								
White non-Hispanic	Ref				Ref			
Others	1.1	0.38	3.17	0.87	1.28	0.35	4.64	0.7
T stage, *n* (%)								
T1	Ref							
T2–T4	0.44	0.15	1.27	0.13				
N stage, *n* (%)								
N0	Ref							
N1–N3	3.05	0.99	9.36	0.05				
AJCC stage, *n* (%)								
I–II					Ref			
III	11.82	3.62	38.59	< 0.001	9.89	3.04	32.21	< 0.001
Baseline characteristics	HR	95% lower CI	95% upper CI	*P* value	HR	95% lower CI	95% upper CI	*P* value
Overall survival								
TNBC epitypes								
A/C/D	Ref				Ref			
B	5.31	1.55	18.18	0.01	4.05	1.11	14.72	0.03
Age (years)								
< 45	Ref				Ref			
≥ 45	1.16	0.32	4.26	0.82	0.99	0.23	4.27	0.99
Race								
White non-Hispanic	Ref				Ref			
Others	1.89	0.63	5.68	0.25	2.31	0.65	8.16	0.2
T stage, *n* (%)								
T1	Ref							
T2–T4	0.73	0.24	2.24	0.59				
N stage, *n* (%)								
N0	Ref							
N1–N3	2.04	0.7	5.95	0.19				
AJCC stage, *n* (%)								
I–II	Ref				Ref			
III	5.86	1.96	17.5	0.002	5.01	1.58	15.84	0.01

downregulated in Epi-CL-B. In line with the opposing survival intervals between these TNBC epitypes, *EPHA7*, an activator of the ERK pathway with proapoptotic activity,[28] and *CCNB1IP1*, a modulator of cyclin-B that delays the cell cycle,[29] were also upregulated in Epi-CL-A but downregulated in Epi-CL-B. Epi-CL-C exhibited activation of the cellular response to hypoxia pathway (and upregulation of related genes, e.g., *ARNT*), cellular stress,

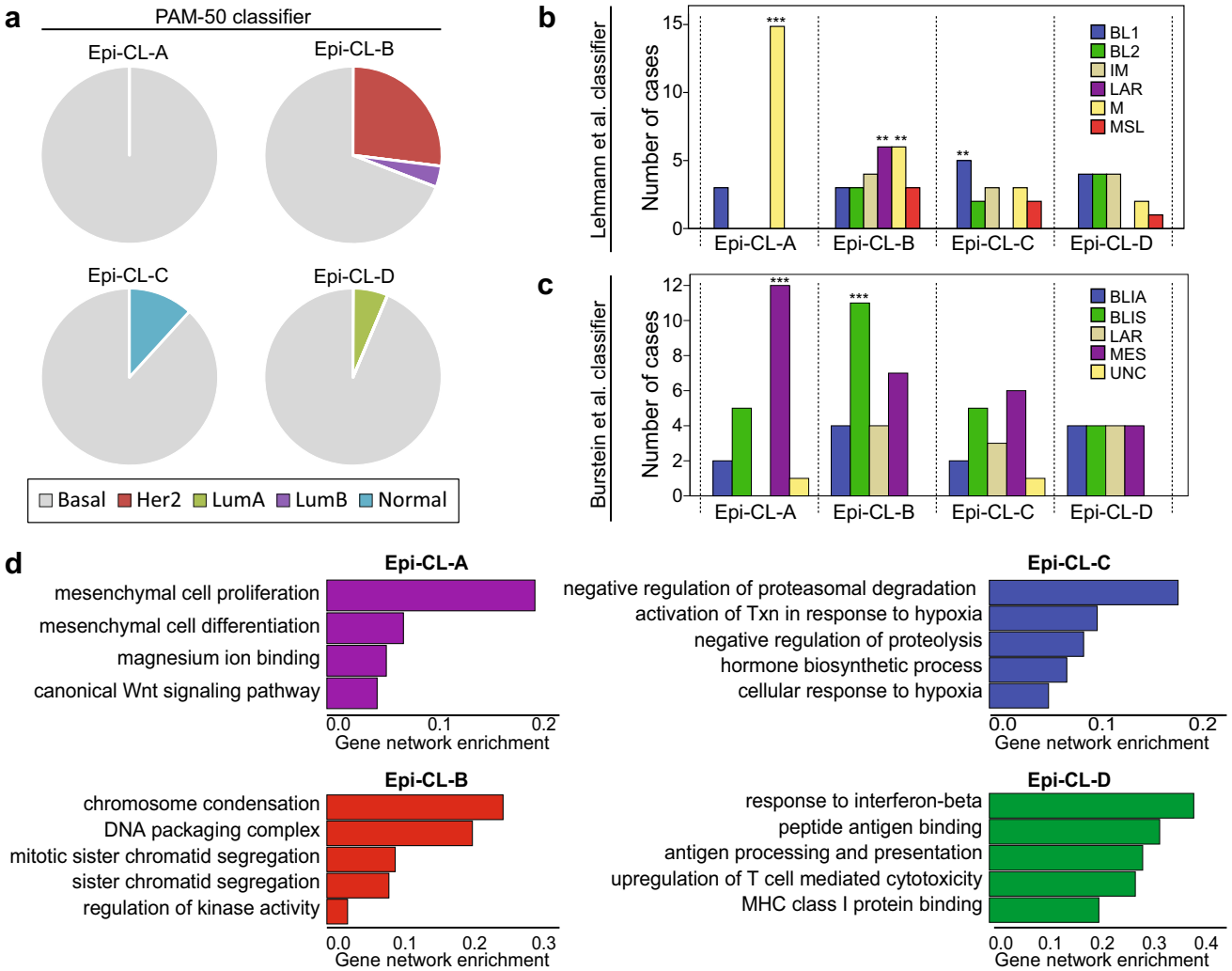


FIG. 2 Integration of the TNBC epitypes with the previously described TNBC transcriptomic subtypes. **a** Pie charts describing the overlap between the PAM50 genomic signatures and the TNBC epitypes. **b**, **c** Bar plots indicating the proportion of overlap between the TNBC epitypes and the Lehmann et al. TNBC classification (**b**) and the Burstein et al. TNBC classification (**c**). *BL1* Basal-like 1, *BL2* basal-like 2, *IM* immunomodulatory, *M* mesenchymal, *MSL* mesenchymal stem-like, *LAR* luminal androgen receptor, *BLIA* basal-like immuno-activator, *BLIS* basal-like immuno-suppressor, *LAR* luminal androgen receptor, and *MES* mesenchymal-like. $***P < 0.001$ and $**P < 0.01$. **d** Bar plots summarizing the gene networks differentially active in each TNBC epigenetic subtype

and negative regulation of proteasomal degradation and proteolysis pathways (WAC). Upregulation of *USP6NL* gene was observed in Epi-CL-C, which has been shown to contribute to breast tumor proliferation.[30] Finally, Epi-CL-D showed enriched response to interferon-beta (and upregulation of related genes, e.g., *CAGS*), antigen processing and presentation (with related genes such as *LILRA5* being upregulated), and positive regulation of T cell-mediated cytotoxicity and related genes such as *CCL18* and *IL15RA* (Fig. 2d). These features are compatible with a favorable immune response and partially explain the improved outcome. We also observed increased levels of *TP53* expression in Epi-CL-D. These differences prompted us to investigate epigenetic subtype-specific ancillary therapeutic approaches (Supplementary Table S3).

Genomic Landscapes of TNBC Epigenetic Subtypes

The mutational landscape was assessed using whole-exome sequencing (WES) data from the TCGA-BRCA cohort to identify the most commonly mutated genes in TNBC (Fig. 3a; $n = 143$ profiled cases). As widely reported, *TP53* was the most frequently mutated gene in TNBC cases (73%), followed by *PIK3CA* (10%). The mutational frequency differed among the four TNBC epi-types. Epi-CL-A exhibited a lack of mutations of the *PIK3CA* gene, which was mutated in almost 20% of TNBC Epi-CL-B

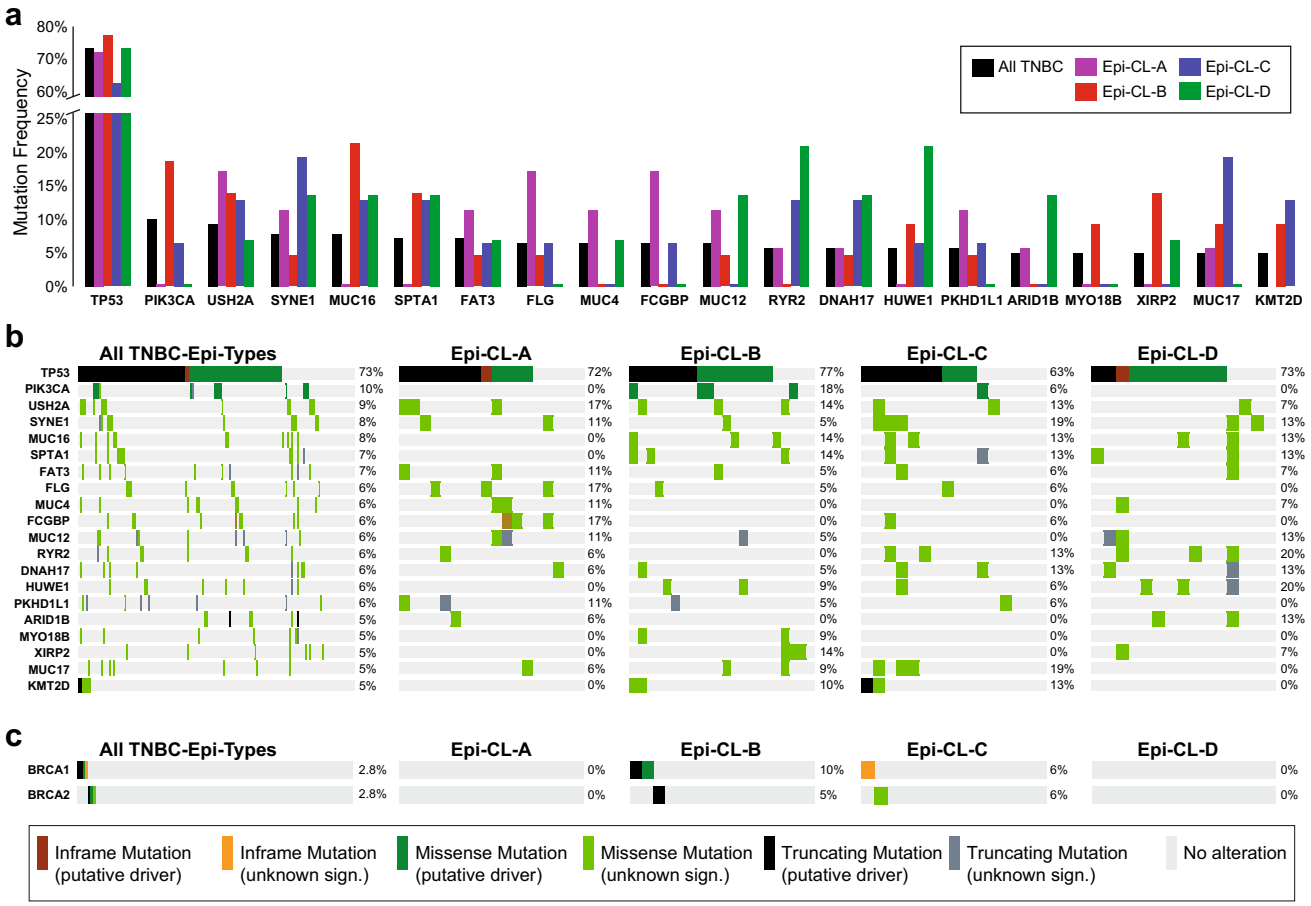


FIG. 3 Mutational landscapes of the TNBC epigenetic subtypes. **a** Bar plot indicating the frequency of gene mutation for each TNBC epitype. The subset of genes shown in this plot (n = 20) were selected based on the frequency of mutations in all the TNBC specimens profiled in TCGA by whole-exome sequencing (WES, n = 143). **b** Oncoprint analysis including the 20 most frequently mutated genes in TNBC for each epigenetic subtype, i.e., inframe mutations with putative drivers, inframe mutations with unknown significance, missense mutations with putative drivers, missense mutations with unknown impact, truncating mutations with putative drivers, and truncating mutations with unknown impact; represented in red, dark green, light green, black, and grey, respectively. **c** Oncoprint analysis for mutations impacting the *BRCA1* and *BRCA2* genes in each TNBC epitype

tumors (Fig. 3a, b). The *MUC16* gene, commonly known as cancer antigen 125 (CA-125), was frequently mutated in Epi-CL-B. We additionally evaluated the mutational spectrum of the *BRCA1* and *BRCA2* genes (representing 5.6% of all mutations). While tumors in Epi-CL-A and D showed no mutations of either *BRCA1* or *BRCA2* genes, 13% of tumors included in Epi-CL-B and C presented mutually exclusive mutations of these genes (Fig. 3c).

DISCUSSION

The TNBC definition based on ER, PgR, and HER2 has created a repository of highly variable tumors for which effective therapy is still lacking. To improve clinical outcomes, several studies have attempted to classify TNBC into distinct subtypes based on gene expression, immunogenic profiling, and transcriptome analysis.[2–6,24,25] Here, leveraging our previous DNAm-based classification systems, we demonstrated that machine learning-based identification of DNAm signatures can subclassify high-purity TNBC tumors. This study provides a step-by-step epigenetic data processing and reduction protocol to generate an informative DNAm signature (n = 103) to subclassify TNBC tumors (Supplementary Table S1). This resource can be utilized by other researchers to stratify TNBC patients according to the proposed method and explore the clinicopathological implications of this new stratification system.

The TNBC epitypes are in partial agreement with previously described transcriptomic classifications,[2,3,6] suggesting that combining these subtyping systems can provide further prognostic and therapeutic information. In fact, the characterization of differentially enriched gene networks in each epigenetic subtype could allow for

selection of subtype-specific treatments; For example, we demonstrated that the Epi-CL-A subtype is characterized by upregulation in the Wnt pathway and mesenchymal cell proliferation. Consideration then of CMF (cyclophosphamide, methotrexate, 5-fluorouracil) chemotherapy may be indicated, as methotrexate downregulates the Wnt pathway and mesenchymal cell proliferation and differentiation. New unexplored compounds such us esculetin and CMF could be investigated as combination therapy, as preclinical data suggest that esculetin decreases Wnt targeted genes[31] and also enhances the inhibitory effects of 5-fluorouracil.[32] For patients with the Epi-CL-B subtype, consideration of cell cycle inhibitor therapy could be explored given that CDK2AP1, a negative CDK2 regulator, is decreased in this subtype of TNBC. Treatment with a CDK2 inhibitor, e.g., SNS032, could be a potential option for targeted therapy. For patients with Epi-CL-C, the antibacterial agent enoxacin could be explored, since this agent acts specifically on pathways upregulated in this subtype involving cellular response to hypoxia, negative regulation of proteolysis, and negative regulation of proteasomal degradation. Finally, securinine could benefit patients of Epi-CL-D, by modulating antigen processing and presentation and major histocompatibility complex (MHC) class I protein binding, as has been shown for other cancer types (Supplementary Table S3).

The existence of different mutational profiles between the TNBC epitypes indicates that integrative studies of these molecular features could also generate more robust and clinically meaningful TNBC classification methods. We identified that the Epi-CL-B subtype resembles the HER2-enriched phenotype and has a high rate of *PIK3CA* mutations. The significance of HER2 enrichment for patients with HER2-negative disease is not clinically known, but exploring a role for anti-HER2 targeted therapy for this subset of TNBC patients may be warranted. Recently, alpelisib, a *PIK3CA* inhibitor, has demonstrated, in combination with fulvestrant significantly improved progression-free survival, and has been approved by the Food and Drug Administration (FDA) for patients with advanced ER-positive breast cancers *PIK3CA* mutated.[33] HER2-directed and PIK3CA inhibitor therapy, singly or in combination, would exploit the distinct mutational aberrations associated with this aggressive epitype. These links between genetic make-up and epigenetic subtypes open new avenues in the development of laboratory tests to more efficiently classify TNBC, helping to optimize treatment approaches.

The inability to effectively characterize TNBC into a targetable disease underscores the clinical challenge to successful treatment of this aggressive cancer subtype. Exploiting both genomic and epigenomic differences opens new doors for treatment currently not considered. Our data provide new information in a relatively unexplored area of TNBC molecular biology, which provides the foundation for additional investigation of DNAm epitypes as a reliable method of classifying TNBC into prognostic and therapeutically distinct subtypes.

ACKNOWLEDGMENT This work was supported by the Associates for Breast and Prostate Cancer Studies (ABCs) Foundation, the Fashion Footwear Association of New York (FFANY) Foundation, and the John Wayne Cancer Institute Translational Research Fund.

DISCLOSURE None of the authors have any financial disclosures. The authors declare no competing interests.

REFERENCES

1. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med.* 2010;363(20):1938–48.
2. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750–67.
3. Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res.* 2015;21(7):1688–98.
4. Jezequel P, Loussouarn D, Guerin-Charbonnel C, et al. Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res.* 2015;17:43.
5. Liu YR, Jiang YZ, Xu XE, et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res.* 2016;18(1):33.
6. Lehmann BD, Jovanovic B, Chen X, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE.* 2016;11(6):e0157368.
7. Jeschke J, Bizet M, Desmedt C, et al. DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest.* 2017;127(8):3090–102.
8. Mundbjerg K, Chopra S, Alemozaffar M, et al. Identifying aggressive prostate cancer foci using a DNA methylation classifier. *Genome Biol.* 2017;18(1):3.
9. Wu SP, Cooper BT, Bu F, et al. DNA methylation-based classifier for accurate molecular diagnosis of bone sarcomas. *JCO Precis Oncol.* 2017;1:1–11.
10. Klughammer J, Kiesel B, Roetzer T, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat Med.* 2018;24(10):1611–24.
11. Sahm F, Schrimpf D, Stichel D, et al. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol.* 2017;18(5):682–94.
12. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature.* 2018;555(7697):469–74.
13. Orozco JIJ, Knijnenburg TA, Manughian-Peter AO, et al. Epigenetic profiling for the molecular classification of metastatic brain tumors. *Nat Commun.* 2018;9(1):4627.
14. Marzese DM, Scolyer RA, Huynh JL, et al. Epigenome-wide DNA methylation landscape of melanoma progression to brain metastasis reveals aberrations on homeobox D cluster associated with prognosis. *Hum Mol Genet.* 2014;23(1):226–38.

15. Moran S, Martinez-Cardus A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* 2016;17(10):1386–95.
16. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–04.
17. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.
18. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun.* 2015;6:8971.
19. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10): 1363–69.
20. Salomon MP, Orozco JIJ, Wilmott JS, et al. Brain metastasis DNA methylomes, a novel resource for the identification of biological and clinical features. *Sci Data.* 2018;5:180245.
21. Chen X, Li J, Gray WH, et al. TNBCtype: a subtyping tool for triple-negative breast cancer. *Cancer Inform.* 2012;11:147–56.
22. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. *Nucleic Acids Res.* 2018;46(W1):W60–W64.
23. Napolitano F, Carrella D, Mandriani B, et al. gene2drug: a computational tool for pathway-based rational drug repositioning. *Bioinformatics.* 2018;34(9):1498–505.
24. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94(446):496–509.
25. de Glas NA, Kiderlen M, Vandenbroucke JP, et al. Performing survival analyses in the presence of competing risks: a clinical example in older breast cancer patients. *J Natl Cancer Inst.* 2016;108(5):djv366.
26. Devarajan K. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS Comput Biol.* 2008;4(7):e1000029.
27. He X, Xiang H, Zong X, et al. CDK2-AP1 inhibits growth of breast cancer cells by regulating cell cycle and increasing docetaxel sensitivity in vivo and in vitro. *Cancer Cell Int.* 2014;14(1):130.
28. Li S, Wu Z, Ma P, et al. Ligand-dependent EphA7 signaling inhibits prostate tumor growth and progression. *Cell Death Dis.* 2017;8(10):e3122.
29. Singh MK, Nicolas E, Gherraby W, Dadke D, Lessin S, Golemis EA. HEI10 negatively regulates cell invasion by inhibiting cyclin B/Cdk1 and other promotility proteins. *Oncogene.* 2007;26(33):4825–32.
30. Avanzato D, Pupo E, Ducano N, et al. High USP6NL levels in breast cancer sustain chronic AKT phosphorylation and GLUT1 stability fueling aerobic glycolysis. *Cancer Res.* 2018;78(13):3432–44.
31. Zhu X, Gu J, Qian H. Esculetin attenuates the growth of lung cancer by downregulating wnt targeted genes and suppressing NF-kappaB. *Arch Bronconeumol.* 2018;54(3):128–33.
32. Yan L, Yu HH, Liu YS, Wang YS, Zhao WH. Esculetin enhances the inhibitory effect of 5-fluorouracil on the proliferation, migration and epithelial-mesenchymal transition of colorectal cancer. *Cancer Biomark.* 2019;24(2):231–40.
33. Andre F, Ciruelos E, Rubovszky G, et al. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. *N Engl J Med.* 2019;380(20):1929–40.