IJC
International Journal of Cancer

Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes

Balázs Győrffy[1,2,3], Giulia Bottai[4], Thomas Fleischer[5,6], Gyöngyi Munkácsy[1], Jan Budczies[7], Laura Paladini[4], Anne-Lise Børresen-Dale[5,6], Vessela N. Kristensen[5,6,8] and Libero Santarpia[4]

[1] MTA TTK Lendület Cancer Biomarker Research Group, Budapest, Hungary
[2] 2nd Dept. of Pediatrics, Semmelweis University, Budapest, Hungary
[3] MTA-SE Pediatrics and Nephrology Research Group, Budapest, Hungary
[4] Oncology Experimental Therapeutics Unit, IRCCS Clinical and Research Institute Humanitas, Rozzano - Milan, Italy
[5] Department of Genetics, Institute for Cancer Research, OUS Radiumhospitalet, Oslo, Norway
[6] The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Norway
[7] Institute of Pathology, Campus Charité Mitte, Charité Universitätsmedizin Berlin, Berlin, Germany
[8] Department of Clinical Molecular Biology and Laboratory Science (EpiGen), Akershus University Hospital, Division of Medicine, Lørenskog, Norway

DNA methylation has a substantial impact on gene expression, affecting the prognosis of breast cancer (BC) patients dependent on molecular subtypes. In this study, we investigated the prognostic relevance of the expression of genes reported as aberrantly methylated, and the link between gene expression and DNA methylation in BC subtypes. The prognostic value of the expression of 144 aberrantly methylated genes was evaluated in ER+/HER2−, HER2+, and ER−/HER2− molecular BC subtypes, in a meta-analysis of two large transcriptomic cohorts of BC patients ($n = 1{,}938$ and $n = 1{,}640$). The correlation between gene expression and DNA methylation in distinct gene regions was also investigated in an independent dataset of 104 BCs. Survival and Pearson correlation analyses were computed for each gene separately. The expression of 48 genes was significantly associated with BC prognosis ($p < 0.05$), and 32 of these prognostic genes exhibited a direct expression–methylation correlation. The expression of several immune-related genes, including *CD3D* and *HLA-A*, was associated with both relapse-free survival (HR = 0.42, $p = 3.5E\text{-}06$; HR = 0.35, $p = 1.7E\text{-}08$) and overall survival (HR = 0.50, $p = 5.5E\text{-}04$; HR = 0.68, $p = 4.5E\text{-}02$) in ER-/HER2- BCs. On the overall, the distribution of both positive and negative expression–methylation correlation in distinct gene regions have different effects on gene expression and prognosis in BC subtypes. This large-scale meta-analysis allowed the identification of several genes consistently associated with prognosis, whose DNA methylation could represent a promising biomarker for prognostication and clinical stratification of patients with distinct BC subtypes.

Breast cancer (BC) represents a heterogeneous disease, which includes several subtypes with different molecular and clinical features.[1] Distinct gene pathways, genomic aberrations, and gene expression profiles have been associated with pathological processes and prognosis in different BC subtypes.[1–4] Epigenetic alterations have recently emerged as a common hallmark of human cancer, including BC.[5,6] In particular, DNA methylation, which most frequently occurs at CpG dinucleotides, has been associated with clinicopathological features of BC patients, such as tumor stage, histological grade, and *TP53* status.[7–10] Furthermore, DNA hypomethylation and hypermethylation can influence BC progression and prognosis, contributing to the overexpression of oncogenes and downregulation of tumor suppressor genes, respectively.[5]

Key words: breast cancer subtypes, DNA methylation biomarkers, gene expression, immune genes, prognosis

Abbreviations: BC: breast cancer; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; HR: hazard ratio; OS: overall survival; RFS: relapse-free survival; TSS: transcription start site

Additional Supporting Information may be found in the online version of this article.

Conflict of Interest: The authors declare no conflict of interest.

Grant sponsor: Associazione Italiana Ricerca sul Cancro; **Grant number:** 6251; **Grant sponsor:** OTKA K; **Grant number:** 108655; **Grant sponsor:** Norwegian Research Council; **Grant number:** 193387/V50; **Grant sponsor:** Norwegian Cancer Society; **Grant number:** 4196163832

DOI: 10.1002/ijc.29684

History: Received 21 May 2015; Accepted 2 July 2015; Online 11 July 2015

Correspondence to: L. Santarpia, MD, PhD, Oncology Experimental Therapeutics Unit, IRCCS Clinical and Research Institute Humanitas, Via Manzoni 113 - 20089 Rozzano - Milan, Italy, Tel.: +39-02-8224-5173, Fax: +39-02-82245191, E-mail: libero.santarpia@humanitasresearch.it and V. N. Kristensen, PhD, Faculty of Medicine, Institute for Clinical Medicine, University of Oslo, 0372, Oslo, Norway, Tel.: +47-22-78-13-75/+47-920-68-432, Fax: +47-22-78-13-95, E-mail: v.n.kristensen@medisin.uio.no

What's new?

DNA methylation profiles may play an important role in the development and progression of breast cancer (BC) subtypes, but the prognostic value of aberrantly methylated biomarkers in distinct subtypes and the role of DNA methylation in distinct gene regions remain controversial. This study assesses the prognostic impact of the expression of aberrantly methylated genes and expression–methylation correlations in BC subtypes. Key methylated prognostic genes were identified, including immune-related genes, particularly in ER−/HER2− tumors. DNA methylation in specific gene regions differentially affects gene expression, supporting the importance of epigenetic biomarkers for prognostication and clinical stratification of patients with distinct BC subtypes.

Even though DNA hypermethylation is conventionally negatively associated with gene expression, a more complex scenario is emerging. Indeed, methylation has been demonstrated to positively correlate with gene expression, and the epigenetic modification of functionally different gene regions may consequently lead to distinct biological and clinical implications.[9–11] Noteworthy, specific methylation patterns have been associated with different BC subtypes.[12–15] Particularly, ER+/luminal BCs are characterized by a remarkably higher frequency of DNA methylation compared to ER−/basal-like tumors, and a substantial number of genes are differentially methylated in molecular BC subtypes.[3,12–15] Taken together, these data suggest that DNA methylation profiles may play an important role in the development and progression of distinct BC subtypes.

Several single genes whose global methylation status correlated with gene expression level and BC outcome have already been identified.[5] However, the prognostic value of these aberrantly methylated biomarkers in BC subtypes and the complex role of DNA methylation in distinct gene regions are still controversial topics, requiring further validation in larger independent cohorts of BC patients.

In this study, we performed a comprehensive literature review of genes whose aberrant DNA methylation status has been previously associated with prognosis in BC. Then we searched for the prognostic potential of the expression of these genes in a large-scale meta-analysis employing two independent transcriptomic cohorts of patients ($n = 1{,}938$ and $n = 1{,}640$) in distinct molecular BC subtypes. Finally, we assessed the correlation between the DNA methylation status in different gene regions and expression level of the identified genes in an independent cohort of 104 BC patients.

Materials and Methods

Study selection

A systematic search was conducted based on the Preferred Reporting Items for Systematics Reviews and Meta-Analyses (PRISMA) criteria.[16] The workflow of the study is summarized in Figure 1. PubMed, Web of Science, and Embase databases were screened for studies relating BC, prognosis, and aberrant DNA methylation until June 2014. Additional studies were identified through the references listed in review publications. Two independent reviewers (G.B. and G.M.) screened potentially relevant papers. To be eligible for inclusion, studies had to meet the following criteria: (*i*) gene methylation status determined by PCR-, sequencing-, or array-based techniques; (*ii*) DNA methylation detected from tissue or circulating tumor DNA methylation from the whole blood, plasma, or serum of BC patients; (*iii*) studies published as full papers. Exclusion criteria included (*i*) data from animal or cell line studies and (*ii*) studies published in any language other than English. Each gene was included only once in the final list and named according to the HUGO Gene Nomenclature Committee (Supporting Information, Table 1).

Construction of transcriptomic databases for meta-analysis

Gene expression data generated with Affymetrix U133A and U133 Plus2 gene chips were identified from transcriptomic studies comprising at least 30 BC patients. Quality control for all gene chips and removal of duplicate samples were performed as previously described.[17] Data were MAS5 normalized in the R statistical environment (http://www.r-project.org) using the Affy Bioconductor library. For genes targeted by multiple microarray probes, only the probe set with the highest JetSet score was selected.[18] ER and HER2 status were determined for each patient using the probe sets 205225_at and 216836_s_at, respectively.[19] Molecular subgroups were defined according to ER and HER2 status, and were evaluated separately. An additional independent analysis was performed in the Metabric cohort.[20] Due to batch effects between the Metabric training and validation sets, raw data were summarized using the beadarray package in the R environment.[21] For annotation, the Illumina Humanv3 database of Bioconductor was used (http://www.bioconductor.org). We then removed 319 unmapped probes and performed quantile normalization using the preprocess Core package. Finally, for genes with multiple probes, only the probe set with the highest dynamic range was retained. The clinical endpoints for the Affymetrix and Metabric transcriptomic cohorts were relapse-free survival (RFS) and overall survival (OS), respectively. Systemically untreated patients were not included in the analysis. Overall, we used expression data from 28 independent datasets of primary BC patients, for 3,578 patients analyzed. In the Affymetrix cohort ($n = 1{,}938$), 64.8% of the patients were ER+/HER2−, 19.8% were HER2+, and 15.4% were ER−/HER2− (Supporting Information, Table 2). The Metabric dataset ($n = 1{,}640$) was composed by 71.1% ER+/HER2−, 12.7% HER2+, and 16.2% ER−/HER2− tumors. The clinicopathological characteristics of

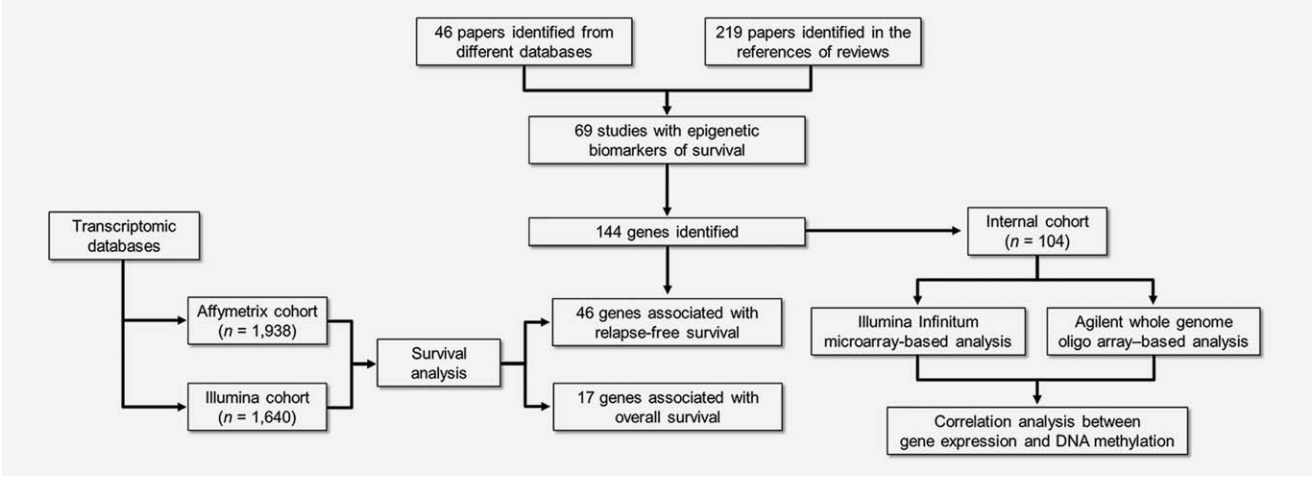

Figure 1. Workflow of the study and data processing.

patients included in the datasets used in this study are presented in the Supporting Information, Table 2.

Clinical sample collection

Fresh frozen tumor samples were obtained from 104 BC patients at Ullevål University Hospital, The Norwegian Radium Hospital, Bærum Hospital, Aker University Hospital, and Buskerud Hospital between 1995 and 1998. Patients were uniformly treated according to the National Guidelines, as previously described.[22] All samples were collected after obtaining informed consent and approval of the local ethical committee. Patients' characteristics are presented in the Supporting Information, Table 2.

DNA methylation and mRNA expression analysis

DNA methylation and gene expression data were available for 104 breast tumor samples. The DNA methylation status of [mt]450,000 CpG sites was interrogated using Illumina Infinium HumanMethylation450 microarray, as previously described.[8] Briefly, the returned value of each CpG probe (β) was calculated as the methylated signal divided by the sum of the methylated and the unmethylated signals, representing the fraction of methylated DNA molecules at a specific locus. Preprocessing and normalization steps of probe filtering, color bias correction, background subtraction, and subset quantile normalization were performed as previously described.[23] DNA methylation values for genes were summarized according to "gene regions" to focus the analysis on functionally relevant regions. A CpG located in a gene is mapped to one of the following six regions: (*i*) CpGs that are between 1500 and 200 bp upstream of the transcription start site (TSS1500); (*ii*) CpGs that are between 200 bp upstream of the TSS and the TSS itself (TSS200); (*iii*) CpGs in the 5′UTR (5′UTR); (*iv*) CpGs in the first exon (First exon); (*v*) CpGs in other exons or in introns (Body); and (*vi*) CpGs in the 3′UTR (3′UTR). Methylation levels for each region were summarized using the median. The DNA methylation data is available in the Gene Expression Omnibus (GEO) with the accession number GSE60185. Gene expression for the same patients was assessed using Agilent whole genome 4x44K oligo array as previously described.[24] The mRNA expression data is available in the GEO with accession number GSE19783.

Statistical analyses

Kaplan–Meier survival analysis was performed in R using Bioconductor packages. We used the median of gene expression across all samples as a cut-off for survival analysis. Multivariate analysis was adjusted for grade and lymph node status. Age was excluded because was not significant in univariate analysis. Tumor size and treatments were not included due to the low numbers of data available. Multivariate Cox regression analysis was performed using WinStat for Excel 2014 (R. Fitch Software, Staufen, Germany). DNA methylation of gene regions and gene expression level of each gene was tested for non-zero correlation using Pearson correlation (R function *corr.test*). Bonferroni or Benjamini–Hochberg (BH) corrections were applied for multiple testing, and statistical significance was set at $p < 0.05$.

Results

Identification of candidate epigenetic biomarkers

After retrieving search results from the PubMed, Web of Science, and Embase databases, 46 publications were identified. Additionally, 219 publications were found from the reference list of review articles. Overall, we identified 69 relevant studies correlating aberrantly methylated genes and survival in BC, and from these we selected 144 individual genes (Fig. 1; Supporting Information Table 1).

Survival analysis in breast cancer subtypes

To clarify the prognostic role of the 144 candidate epigenetic biomarkers in BC subtypes, we performed Kaplan–Meier survival analysis in both transcriptomic datasets, separately. The expression of 48 genes was significantly associated with RFS or OS in at least one BC subtype (Bonferroni-adjusted $p < 0.05$, Table 1). Twenty-seven genes correlated with the outcome of ER+/HER2− BC patients only, and *FLRT2*,

Table 1. Kaplan–Meier analysis of relapse-free survival and overall survival for the 48 significant genes in breast cancer subtypes

	ER+/HER2−				HER2+				ER−/HER2−			
	Relapse-free survival ($n = 1,257$)		Overall survival ($n = 1,167$)		Relapse-free survival ($n = 384$)		Overall survival ($n = 209$)		Relapse-free survival ($n = 297$)		Overall survival ($n = 264$)	
Gene symbol	*p*	HR	*P*	HR	*p*	HR	*p*	HR	*p*	HR	*p*	HR
ABCB1	<1.0E-05	0.56	7.0E-03[1]	0.74	–	–	–	–	<1.0E-05	0.42	3.7E-02[1]	0.67
ADAT1	–	–	–	–	–	–	–	–	<1.0E-05[1]	0.43	–	–
ALX1	8.3E-03	0.64	–	–	–	–	–	–	–	–	–	–
BRCA1	<1.0E-05[1]	1.67	–	–	–	–	–	–	–	–	–	–
CD3D	–	–	–	–	4.4E-04	0.57	–	–	<1.0E-05	0.42	5.5E-04[1]	0.50
CD3G	–	–	–	–	–	–	–	–	–	–	3.8E-03[1]	0.55
CD6	9.5E-03	0.76	–	–	1.1E-03	0.59	–	–	1.9E-04	0.50	8.9E-04[1]	0.51
CD79B	–	–	–	–	–	–	–	–	6.3E-03	0.60	7.1E-04[1]	0.52
CDO1	<1.0E-05	0.56	–	–	–	–	–	–	–	–	–	–
CHD9	8.3E-03	1.58	–	–	–	–	–	–	–	–	–	–
CKM	<1.0E-05[1]	0.59	–	–	–	–	–	–	–	–	–	–
DAAM1	<1.0E-05	1.63	–	–	–	–	–	–	–	–	–	–
DIRAS3	<1.0E-05[1]	0.43	–	–	–	–	–	–	–	–	–	–
ETS1	<1.0E-05	0.61	–	–	–	–	–	–	–	–	–	–
EYA4	<1.0E-05	0.59	–	–	–	–	–	–	–	–	–	–
FAM110A	<1.0E-05	0.56	–	–	–	–	–	–	–	–	–	–
FANCC	<1.0E-05	0.56	–	–	–	–	–	–	–	–	–	–
FGF4	<1.0E-05[1]	0.57	–	–	–	–	–	–	<1.0E-05	0.39	–	–
FLRT2	4.2E-03	0.62	1.5E-02[1]	0.76	–	–	–	–	–	–	–	–
FOXC1	8.3E-03	0.63	–	–	–	–	–	–	–	–	–	–
GADD45GIP1	<1.0E-05	1.64	–	–	–	–	–	–	4.2E-03	2.31	–	–
GBE1	4.2E-03[1]	1.60	–	–	–	–	–	–	–	–	–	–
HCLS1	–	–	8.0E-03	0.4	–	–	–	–	2.5E-05	0.45	4.1E-03[1]	0.57
HLA-A	–	–	–	–	<1.0E-05	0.42	–	–	<1.0E-05[1]	0.35	4.5E-02	0.68
HSP90AA1	<1.0E-05	1.98	1.2E-02[1]	1.3	–	–	–	–	–	–	–	–
ICOS	–	–	–	–	–	–	–	–	–	–	9.1E-04[1]	0.46
KLK10	4.2E-03	0.63	–	–	–	–	–	–	4.2E-03	0.44	–	–
LAMA1	<1.0E-05	0.59	–	–	–	–	–	–	–	–	–	–
LAX1	2.0E-05	0.63	–	–	5.0E-05	0.52	–	–	<1.0E-05	0.35	3.2E-03[1]	0.56
LCK	–	–	–	–	–	–	–	–	2.0E-05	0.45	1.8E-03	0.53

Table 1. Kaplan–Meier analysis of relapse-free survival and overall survival for the 48 significant genes in breast cancer subtypes (Continued)

	ER+/HER2−				HER2+				ER−/HER2−			
	Relapse-free survival ($n = 1,257$)		Overall survival ($n = 1,167$)		Relapse-free survival ($n = 384$)		Overall survival ($n = 209$)		Relapse-free survival ($n = 297$)		Overall survival ($n = 264$)	
Gene symbol	*p*	HR	*P*	HR	*p*	HR	*p*	HR	*p*	HR	*p*	HR
MYOD1	–	–	–	–	4.0E-05	0.47	–	–	<1.0E-05	0.37	–	–
NEUROG1	<1.0E-05	0.62	–	–	–	–	–	–	–	–	–	–
PIK3CA	<1.0E-05[1]	2.12	–	–	–	–	–	–	–	–	–	–
POU4F2	<1.0E-05	0.59	–	–	–	–	–	–	<1.0E-05	0.41	–	–
PPP2R2B	-	-	–	–	–	–	–	–	<1.0E-05	0.35	–	–
PTCH1	4.2E-03	0.62	–	–	–	–	–	–	8.3E-03	0.45	–	–
REM1	<1.0E-05	0.58	–	–	–	–	–	–	<1.0E-05	0.43	2.7E-03	0.56
SFRP1	<1.0E-05	0.61	3.3E-04[1]	0.67	–	–	–	–	–	–	–	–
SIT1	–	–	8.3E-03	0.75	–	–	–	–	9.5E-04	0.54	–	–
SLITRK2	–	–	–	–	–	–	–	–	<1.0E-05	0.37	–	–
SLITRK3	<1.0E-05	0.60	–	–	–	–	–	–	–	–	–	–
STMN1	<1.0E-05	1.85	–	–	–	–	–	–	–	–	–	–
SYDE1	<1.0E-05	0.60	–	–	–	–	–	–	–	–	–	–
TAC1	<1.0E-05	0.61	–	–	–	–	–	–	–	–	–	–
TOPBP1	<1.0E-05[1]	1.8	3.7E-03[1]	1.4	–	–	–	–	–	–	–	–
TUBB3	<1.0E-05[1]	1.79	–	–	–	–	–	–	–	–	–	–
UBASH3A	7.6E-03	1.32	–	–	–	–	–	–	–	–	–	–
UBE2C	<1.0E-05[1]	1.8	<1.0E-05[1]	2.0	–	–	–	–	–	–	–	–

[1]Significance retained in multivariate analysis including the gene, lymph node status, and grade.
HR, hazard ratio.

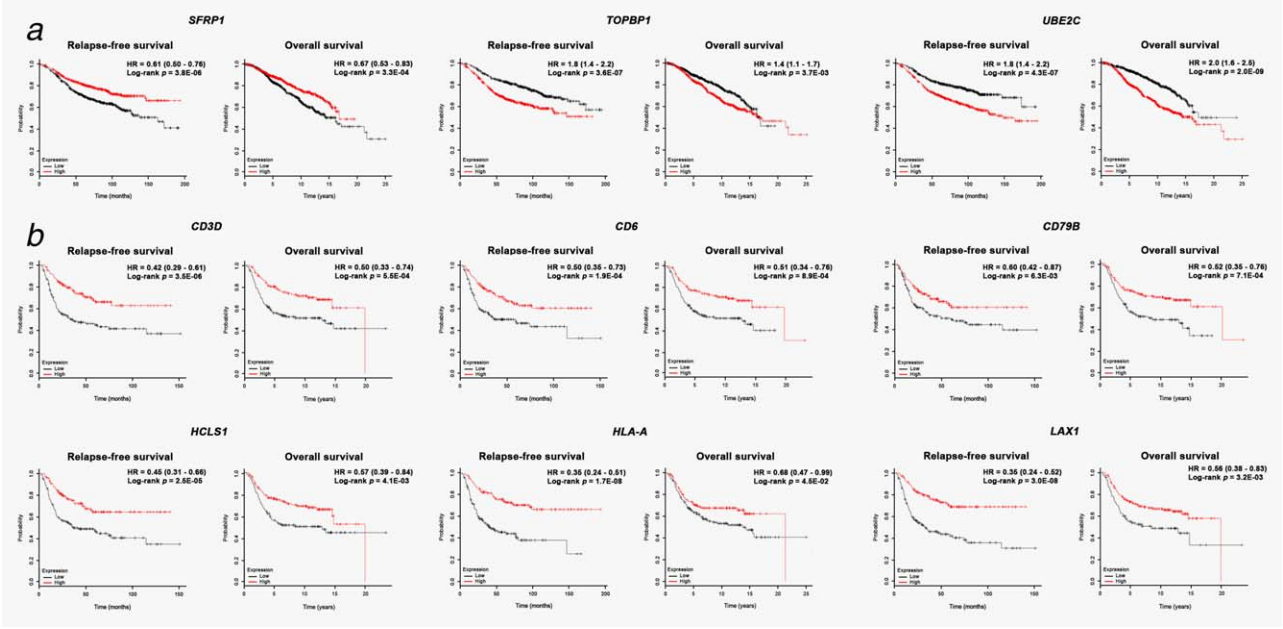

Figure 2. Survival analysis for the most significant genes in breast cancer subtypes. (a) Kaplan–Meier plots for *SFRP1*, *TOPBP1*, and *UBE2C* in ER+/HER2− BCs. (b) Kaplan–Meier plots for *CD3D*, *CD6*, *CD79B*, *HCLS1*, *HLA-A*, and *LAX1* in ER−/HER2− subtype. Log-rank *p* values and hazard ratios (HRs; 95% confidence interval in parentheses) are shown.

HSP90AA1, *SFRP1*, *TOPBP1*, and *UBE2C* were significantly associated with prognosis in both the Affymetrix and the Illumina datasets. Particularly, *SFRP1*, *TOPBP1*, and *UBE2C* showed the most significant association both with RFS (HR = 0.61, $p = 3.8E\text{-}06$; HR = 1.8, $p = 3.6E\text{-}07$; HR = 1.8, $p = 4.3E\text{-}07$) and OS (HR = 0.67, $p = 3.3E\text{-}04$; HR = 1.4, $p = 3.7E\text{-}03$; HR = 2.0, $p = 2.0E\text{-}09$) in this BC subtype (Fig. 2a and Table 1). Multivariate analysis for RFS and OS showed that 13 genes, including *SFRP1*, *TOPBP1*, and *UBE2C*, retained their prognostic significance in ER+/HER2− BC (Table 1; Supporting Information, Table 3). In the ER−/HER2− subtype, we identified 21 prognostic genes (Table 1). Noteworthy, several genes related to immune functions, including *CD3D, CD6, CD79B, HCLS1, HLA-A,* and *LAX1,* correlated with both RFS and OS in the ER−/HER2− subtype, and remained independent prognostic factors in multivariate analysis (Fig. 2b and Table 1; Supporting Information, Table 3). Conversely, only few genes ($n = 5$), mainly involved in immune responses, were prognostic for RFS in univariate analysis in the HER2+ subtype, although no association was significant in the multivariate model (Fig. 2c and Table 1). A higher number of aberrantly methylated genes showed a prognostic value in the ER+/HER2− subtype ($n = 38$) compared to ER−/HER2− and HER2+ BCs ($n = 21$ and $n = 5$, respectively). Interestingly, the majority of genes ($n = 34$ out of 48) correlated with the outcome in only one specific BC subtype (Table 1). Collectively, these results suggest that the prognostic potential of genes affected by DNA methylation may be BC-subtype specific when measured through the effect on mRNA expression.

Correlation between DNA methylation and gene expression in breast cancer

To clarify the functional relevance of DNA methylation, we evaluated the direct correlation between the expression of the 144 selected genes and the methylation level of gene regions in an independent set of 104 BC patients, where both DNA methylation and mRNA expression data was available (Supporting Information, Table 2). Overall, we demonstrated that the expression of 88 out of 144 single genes was significantly associated with DNA methylation in at least one gene region (BH $p < 0.05$; Supporting Information, Table 4). Furthermore, correlative analysis revealed that the expression of a single gene could be concomitantly associated with methylation level in multiple gene regions. Specifically, the methylation status of 127 gene regions was found to negatively correlate with the expression of 67 single genes, and the methylation level of 50 regions was positively associated with the expression of 38 single genes (Supporting Information, Table 4). We showed that about 40% of all negative correlations were detected in the TSSs (Fig. 3). Conversely, positive correlations were enriched in the 3′UTR and body regions, and only 20% were located in TSSs (Fig. 3). Interestingly, positive correlations were never found in the first exon, suggesting that the methylation status of this subregion may have a specific silencing role in gene expression (Fig. 3).

Further, we investigated the potential relevance of DNA methylation on the expression of prognostic genes that emerged from our transcriptome meta-analysis. Among the 48 prognostic genes identified, we found a significant correlation between methylation status of at least one region and

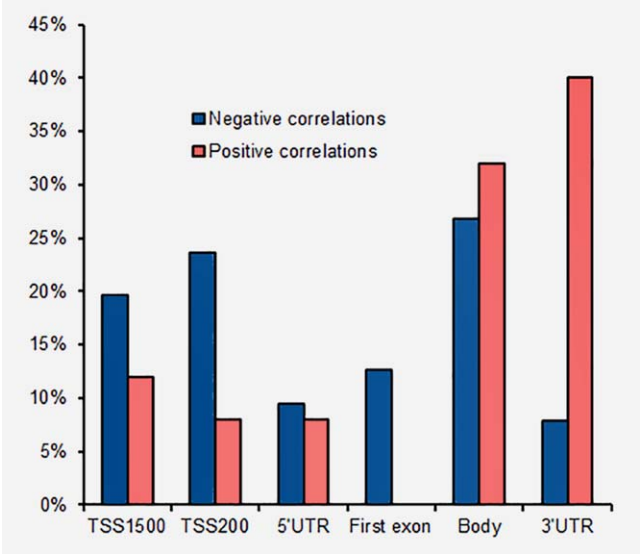

Figure 3. Distribution of significant negative and positive correlations between gene expression and DNA methylation status in different gene regions. A Benjamini–Hochberg corrected $p < 0.05$ was considered statistically significant. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

expression level for 32 genes (Table 2). Twenty-three of these genes, including *SFRP1*, were prognostic in ER+/HER2− subtype (Table 2). Moreover, a significant correlation was found for several genes (*CD3G*, *CD79B*, *ICOS*, and *LCK*) specifically associated with survival in ER−/HER2− BC patients, and for other immune-related genes, including *CD3D*, *HCLS1*, *HLA-A*, *CD6,* and *LAX1,* which were also associated with prognosis in other BC subtypes (Table 2). The percentage of all negative and positive correlations between prognostic (74% and 26%, respectively) and non-prognostic genes (70% and 30%, respectively) was comparable, resembling the overall methylation pattern (72% and 28%, respectively). However, the distribution of both positive and negative correlations was different between prognostic and non-prognostic genes (Fig. 4). Overall, correlations in the body and in the 3′UTR were lower in prognostic genes, whereas a reduction of the correlations distributed in the TSSs, 5′UTR, and the first exon was found in non-prognostic genes (Fig. 4a). Interestingly, prognostic and non-prognostic genes showed an inverse distribution of negative and positive correlations in the 3′UTR and in the TSS1500 regions (Fig. 4b). These results suggest that positive and negative correlations in distinct gene regions may have different effects on gene expression, thus differentially affecting the prognosis of patients with distinct BC molecular subtypes.

Discussion

Alterations in the DNA methylation profile have been associated with changes in gene expression and prognosis in BC.[3,5,7–10,12,25] This work represents the first study using a systematic literature-based review to identify genes whose methylation status was associated with the outcome of BC patients. A large-scale meta-analysis of gene expression was performed within different BC subtypes. Moreover, the potential role of DNA methylation pattern on gene expression and BC prognosis was investigated.

Data derived from the meta-analysis performed on a large transcriptome dataset, and stratified on the three main molecular BC subtypes confirmed the prognostic role of only 48 genes. Twenty-seven genes were specifically associated with the outcome of ER+/HER2− BC patients, and the methylation status of 19 of these genes correlated with gene expression level in the internal cohort of patients. Interestingly, the expression of several of these genes, such as *FLRT2* and *SFRP1*, was previously found to negatively correlate with DNA methylation in a group of tumors that was enriched for ER+/luminal B BCs, supporting the correctness of the approach adopted in our study.[3] Even though *TOPBP1* and *UBE2C* remained significantly associated with both RFS and OS in multivariate analysis in the ER+/HER2− subtype, only the expression of *SFRP1* directly correlated with methylation status in the internal cohort. *SFRP1* has been demonstrated to antagonize the Wnt pathway, and to regulate the transcriptional activity of T-cell factor/lymphocyte enhancer factor, ultimately contributing to tumor initiation and progression.[26,27] Aberrant methylation of *SFRP1* has frequently been found in BC, and has been directly associated with the loss of *SFRP1* expression and poor prognosis in BC.[28] In agreement with our data, the methylation level of *SFRP1* was reported to be higher in ER+ BCs,[13,29] whereas its overexpression and hypomethylation was associated with the ER−/HER2− basal-like subtype.[30,31] Altogether, these previous findings support our results, confirming that DNA methylation negatively influences the expression of the putative tumor suppressor *SFRP1*, and that the methylation-driven reduction of *SFRP1* expression can be predictive of poor prognosis in ER+/HER2− BCs.

Interestingly, the 21 genes found to be prognostic in ER−/HER2− subtype were enriched for biological processes related to immune function. The expression of specific genes, such as *CD3D, CD6, CD79B, HCLS1, HLA-A,* and *LAX1* was consistently associated with both RFS and OS, and remained a significant independent predictor of survival after adjustment for clinicopathological variables in ER−/HER2− subtype. Furthermore, the methylation status of all these genes correlated with gene expression in the internal cohort of BC patients. The expression of *CD3D*, *CD6*, *CD79B*, *HCLS1*, and *LAX1* was previously found to negatively correlate with methylation status, and to be associated with prognosis in BC.[32] Interestingly, *CD3D* was included in a gene set predominantly expressed in a group of BC patients with good prognosis. The expression of this gene was also found to be associated with response to neoadjuvant chemotherapy particularly in ER- BCs, thus potentially identifying a subset of chemotherapy responder patients.[33–35] Our data demonstrated that the T-cell receptor component *CD3D* was the

Table 2. Significant correlations between methylation of gene regions and expression level among prognostic genes in the internal cohort of BC patients

		Pearson correlation coefficients					
Gene Symbol	Subtype	TSS1500	TS200	5'UTR	First exon	Body	3'UTR
BRCA1	ER+/HER2−		−0.365	−0.399			
CDO1	ER+/HER2−	−0.465	−0.409		−0.300		
CHD9	ER+/HER2−					−0.280	
CKM	ER+/HER2−			−0.250			
DAAM1	ER+/HER2−			−0.312		−0.311	
DIRAS3	ER+/HER2−						0.305
ETS1	ER+/HER2−	−0.253	0.397	0.452			0.578
FAM110A	ER+/HER2−	−0.339		−0.261	−0.420		
FANCC	ER+/HER2−			−0.254			
FLRT2	ER+/HER2−		−0.260			0.330	0.247
FOXC1	ER+/HER2−				−0.300		
GBE1	ER+/HER2−	−0.275			−0.278		
LAMA1	ER+/HER2−					0.258	
SFRP1	ER+/HER2−		−0.328		−0.259		0.392
STMN1	ER+/HER2−					−0.297	−0.383
SYDE1	ER+/HER2−	−0.317					−0.248
TAC1	ER+/HER2−	0.255		0.3			
TUBB3	ER+/HER2−					−0.328	0.461
UBASH3A	ER+/HER2−	−0.685	−0.590		−0.679	−0.660	
KLK10	ER+/HER2− and ER−/HER2−						0.310
PTCH1	ER+/HER2− and ER−/HER2−	−0.286					0.259
CD3G	ER−/HER2−	−0.430	−0.462			0.547	
CD79B	ER−/HER2−		−0.566		−0.531	−0.273	
ICOS	ER−/HER2−	−0.568					0.287
LCK	ER−/HER2−	−0.668	−0.535	−0.454			
PPP2R2B	ER−/HER2−			0.300		0.247	
CD3D	ER−/HER2− and HER2+	−0.491			−0.678	−0.541	
HCLS1	ER−/HER2− and HER2+		−0.586			0.265	
HLA−A	ER−/HER2− and HER2+		−0.249			−0.548	
SIT1	ER−/HER2− and HER2+		−0.636			−0.637	
CD6	ER+/HER2−, ER−/HER2− and HER2+		−0.742			−0.353	
LAX1	ER+/HER2−, ER−/HER2− and HER2+	−0.306	−0.586				

Statistical significance was set at $p < 0.05$ after Benjamini–Hochberg correction.

most significant gene associated with prognosis in ER−/HER2− BC. Noteworthy, we showed that the high expression of another immune-related gene, *HLA-A*, was associated with prolonged survival in patients with ER−/HER2− tumors. HLA-A belongs to the classical HLA class I (HLA-I) family, which modulates the function of the tumor-immune microenvironment, having a key role in cancer development orchestrated by tumor-associated macrophages and immune recognition of cancer cells by cytotoxic T lymphocytes (CTLs).[36,37] The downregulation of *HLA-I* expression by epigenetic silencing has been involved in tumor escape from immune surveillance, and has been reported in several types of human cancer, including BC.[38,39] Interestingly, the finding that low expression of *HLA-A* correlated with poor prognosis in ER−/HER2− BC in our meta-analysis is consistent with recent literature data, showing that immune evasion and tumor-infiltrating lymphocytes (TILs) have important functional and clinical implications in this BC subtype.[39–43] Indeed, TILs assessed at baseline were found to potentially stratify patients with ER−/HER2− BCs into

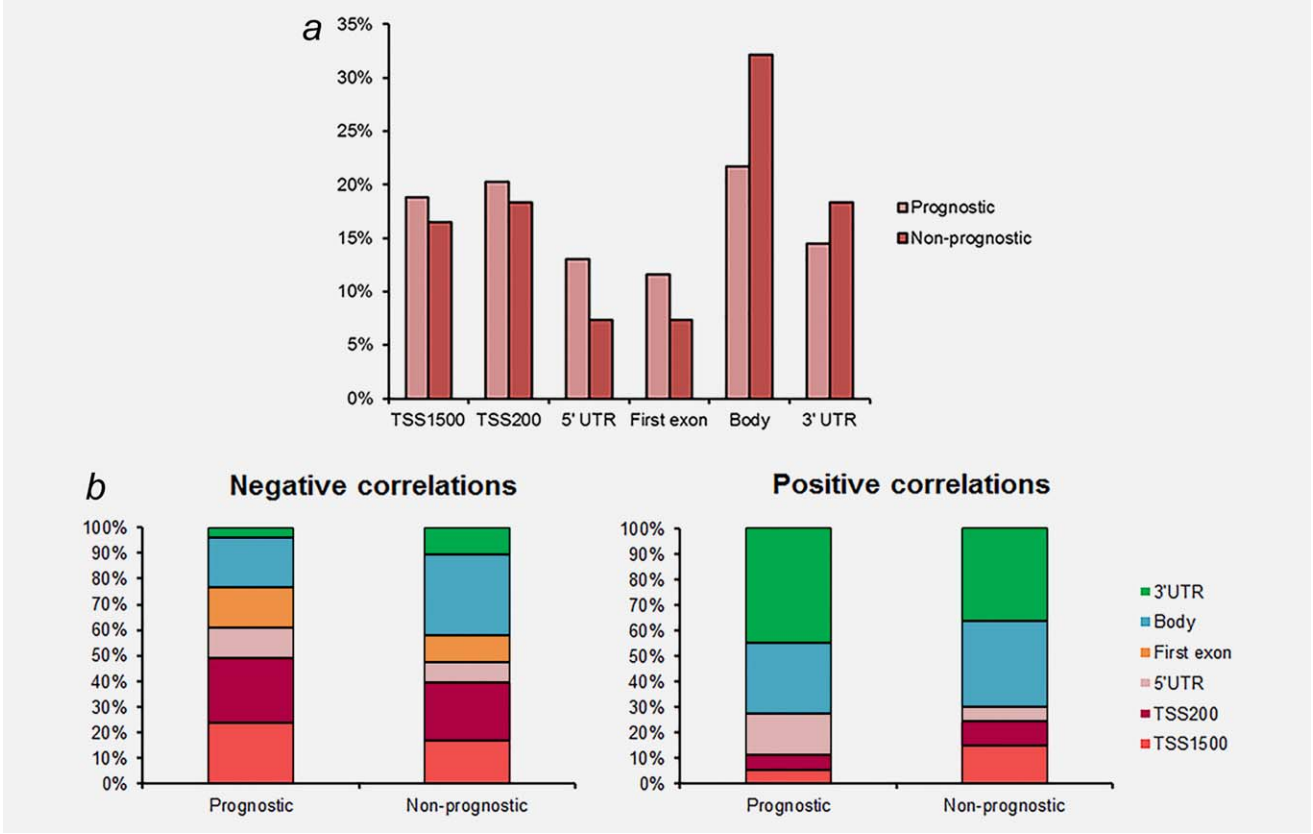

Figure 4. Significant correlations between methylation level of gene regions and expression of prognostic and non-prognostic genes in breast cancer. (a) Distribution of overall correlations relative to gene regions between prognostic and non-prognostic genes. (b) Distribution of negative and positive correlations in gene regions between prognostic and non-prognostic genes. A Benjamini–Hochberg corrected $p < 0.05$ was considered statistically significant. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

high- or low-risk populations.[39] Furthermore, high levels of TILs, especially of CD8+ CTLs, have been associated with favorable outcome in ER−/HER2− BC patients in large-scale studies.[39–42] Consistently, high expression, as well as low levels of DNA methylation, of T lymphocyte-related genes, including *CD3D*, *CD6*, and *LAX1*, correlated with high lymphocyte infiltration, and showed a positive prognostic value in basal-like BCs.[32] Overall, these data suggest that the methylation-dependent downregulation of the expression of immune-related genes, including *CD3D* and *HLA-A*, may be involved in the regulation of anti-tumor immune responses in a subset of ER−/HER2− BC patients. Furthermore, methylation of immune genes may reflect the presence of TILs in a highly sensitive manner suggesting that DNA methylation profiles might be used to predict immune cell proportions in the tumor microenvironment.[44] This approach may help to assess the balance between immune cell sub-populations within the tumor microenvironment, thus refining the prognostic and predictive effect of TILs in BC.

In the analysis of the internal cohort of BC patients, we demonstrated the expression of 88 genes significantly associated with the level of DNA methylation. Furthermore, the integrated approach adopted in this study revealed the expression of a high number of aberrantly methylated genes associated with prognosis in ER+/HER2−. These results are in line with previous findings and may be explained by a different pattern of methylation in distinct BC molecular subtypes.[12,13] Moreover, beyond the classical view that promoter methylation represses gene transcription, DNA methylation in other gene regions may have distinct functions. Accordingly, we found that the majority of identified genes showed a significant negative correlation between methylation of at least one region and gene expression. However, about 43% of genes also showed a significant positive correlation between the expression level and the DNA methylation status of at least one gene region. Given the traditional interpretation, we expected to find positive expression–methylation correlations in regions other than TSSs.[45] Conversely, positive correlations were distributed in all gene regions, with the exception of the first exon, suggesting that the biological role and the location of concordant associations warrant additional evaluation. Finally, we found several differences in the pattern of both positive and negative correlations between the genes that were associated with BC prognosis and the non-prognostic genes. Particularly, prognostic and non-prognostic genes showed a differential methylation profile of both negative and positive correlations in the 3′UTR and in the TSS1500 regions. Even though epigenetic regulation could impact BC

prognosis by mechanisms other than the direct effect on the expression of single genes (*e.g.* the alteration of oncogenic pathways and genomic stability), further efforts are required to understand the complex relation between the methylation status in a specific genomic and cellular context, gene expression, and prognosis in BC molecular subtypes.[45,46]

In conclusion, our large-scale meta-analysis and data integration allowed the identification of several genes consistently associated with prognosis, whose DNA methylation could represent a promising biomarker for prognostication and clinical stratification of patients with distinct subtypes of BC. In particular, the methylation of several immune-related genes may regulate anti-tumor immune responses and influence the outcome of ER−/HER2− BC patients. Since epigenetic mechanisms affect multiple aspects of cancer biology, the direct correlation with gene expression and the role of methylation in distinct gene regions in BC subtypes deserve further investigation.

References

1. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med* 2009; 360:790–800.
2. Santarpia L, Qi Y, Stemke-Hale K, et al. Mutation profiling identifies numerous rare drug targets and distinct mutation patterns in different clinical subtypes of breast cancers. *Breast Cancer Res Treat* 2012; 134:333–43.
3. Cancer Genome Atlas Network Comprehensive molecular portraits of human breast tumours. *Nature* 2012; 490:61–70.
4. Santarpia L, Iwamoto T, Di Leo A, et al. DNA repair gene patterns as prognostic and predictive factors in molecular breast cancer subtypes. *Oncologist* 2013; 18:1063–73.
5. Jovanovic J, Rønneberg JA, Tost J, et al. The epigenetics of breast cancer. *Mol Oncol* 2010; 4: 242–54.
6. Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med* 2011; 17:330–9.
7. Rønneberg JA, Fleischer T, Solvang HK, et al. Methylation profiling with a panel of cancer-related genes: association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. *Mol Oncol* 2011; 5:61–76.
8. Fleischer T, Frigessi A, Johnson KC, et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol* 2014; 15:435
9. Fleischer T, Edvardsen H, Solvang HK, et al. Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. *Int J Cancer* 2014; 134:2615–25.
10. Fackler MJ, Umbricht CB, Williams D, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res* 2011; 71:6195–207.
11. Kulis M, Heath S, Bibikova M, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* 2012; 44:1236–42.
12. Kamalakaran S, Varadan V, Giercksky Russnes HE, et al. DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol Oncol* 2011; 5:77–92.
13. Stefansson OA, Moran S, Gomez A, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol* 2014; pii:S1574–7891(14)00261-0.
14. Holm K, Hegardt C, Staaf J, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res* 2010; 12:R36
15. Bediaga NG, Acha-Sagredo A, Guerra I, et al. DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Res* 2010; 12:R77
16. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010; 8:336–41.
17. Gyorffy B, Schafer R. Meta-analysis of gene expression profiles related to relapse-free survival in 1,079 breast cancer patients. *Breast Cancer Res Treat* 2009; 118:433–41.
18. Li Q, Birkbak NJ, Gyorffy B, et al. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* 2011; 12:474
19. Gyorffy B, Benke Z, Lanczky A, et al. RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res Treat* 2012; 132:1025–34.
20. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; 486:346–52.
21. Dunning MJ, Smith ML, Ritchie ME, et al. Beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* 2007; 23:2183–4.
22. Naume B, Zhao X, Synnestvedt M, et al. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol* 2007; 1: 160–71.
23. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 2012; 4:325–41.
24. Enerly E, Steinfeld I, Kleivi K, et al. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* 2011; 6: e16915
25. Stirzaker C, Zotenko E, Song JZ, et al. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat Commun* 2015; 6:5899
26. Ugolini F, Charafe-Jauffret E, Bardou VJ, et al. WNT pathway and mammary carcinogenesis: loss of expression of candidate tumor suppressor gene SFRP1 in most invasive carcinomas except of the medullary type. *Oncogene* 2001; 20:5810–7.
27. Suzuki H, Toyota M, Carraway H, et al. Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer. *Br J Cancer* 2008; 98:1147–56.
28. Veeck J, Niederacher D, An H, et al. Aberrant methylation of the Wnt antagonist SFRP1 in breast cancer is associated with unfavourable prognosis. *Oncogene* 2006; 25:3479–88.
29. Park SY, Kwon HJ, Choi Y, et al. Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Mod Pathol* 2012; 25:185–96.
30. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003; 100:10393–8.
31. Jeong YJ, Jeong HY, Bong JG, et al. Low methylation levels of the SFRP1 gene are associated with the basal-like subtype of breast cancer. *Oncol Rep* 2013; 29:1946–54.
32. Dedeurwaerder S, Desmedt C, Calonne E, et al. DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* 2011; 3:726–41.
33. Finak G, Bertos N, Pepin F, et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* 2008; 14:518–27.
34. Denkert C, Loibl S, Noske A, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2010; 28:105–13.
35. West NR, Milne K, Truong PT, et al. Tumor-infiltrating lymphocytes predict response to anthracycline-based chemotherapy in estrogen receptor-negative breast cancer. *Breast Cancer Res* 2011; 13:R126
36. Marchesi M, Andersson E, Villabona L, et al. HLA-dependent tumour development: a role for tumour associate macrophages?. *J Transl Med* 2013; 11:247
37. Bukur J, Jasinski S, Seliger B. The role of classical and non-classical HLA class I antigens in human tumors. *Semin Cancer Biol* 2012; 22:350–8.
38. Kaneko K, Ishigami S, Kijima Y, et al. Clinical implication of HLA class I expression in breast cancer. *BMC Cancer* 2011; 11:454
39. Andre F, Dieci MV, Dubsky P, et al. Molecular pathways: involvement of immune pathways in the therapeutic response and outcome in breast cancer. *Clin Cancer Res* 2013; 19:28–33.
40. Loi S, Sirtaine N, Piette F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol* 2013; 31:860;:-
41. Adams S, Gray RJ, Demaria S, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol* 2014; 32: 2959–66.
42. Loi S, Michiels S, Salgado R, et al. Tumor infiltrating lymphocytes are prognostic in triple

negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol* 2014; 25:1544–50.
43. Mahmoud SM, Paish EC, Powe DG, et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol* 2011; 29:1949–55.
44. Jeschke J, Collignon E, Fuks F. DNA methylome profiling beyond promoters - taking an epigenetic snapshot of the breast tumor microenvironment. *FEBS J.* 2015; 282:1801–4.
45. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012; 13:484–92.
46. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?. *Nat Rev Cancer* 2006; 6:107–16.