Cancer Biomarkers 11 (2012) 75–88
DOI 10.3233/CBM-2012-0266
IOS Press

Quantitative DNA methylation and recurrence of breast cancer: A study of 30 candidate genes

Dae Cheol Kim[a,b,1], Mangesh A. Thorat[b,1], Mi Ri Lee[c], Se Heon Cho[c], Nataša Vasiljević[b], Dorota Scibior-Bentkowska[b], Keqiang Wu[b], Amar S. Ahmad[b], Stephen Duffy[b], Jack M. Cuzick[b] and Attila T. Lorincz[b,*]

[a]*Department of Pathology, College of Medicine, Dong-A University, Busan, Korea*

[b]*Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, London, UK*

[c]*Department of Surgery, College of Medicine, Dong-A University, Busan, Korea*

Abstract. *Background*: The need for new prognostic factors in breast cancer is ever increasing as breast cancer management evolves. Aberrant DNA methylation plays a pivotal role in cancer development and progression; DNA methylation-based biomarkers may provide independent prognostic information. We used pyrosequencing to investigate the prognostic potential of quantitative DNA methylation of a large set of candidate genes in a Korean single-institution series of operable breast cancer.
Methods: Absolute DNA methylation in 20 candidate genes from an initial set of 30 genes was measured by pyrosequencing of bisulfite converted DNA in 121 fresh frozen breast cancer cases. Survival analyses used continuous and categorized (quintile-based) gene methylation data with time to recurrence (TTR) as an endpoint. Prognostic abilities of gene-only and risk-score models were explored.
Results: Median follow-up was 5.1 years; 25 recurrences (21%) were observed. Nodal status, methylation of *TWIST1*, *SLIT2* (both as continuous and categorized variables) and *APC*, *HLA-A*, *NKX2-5*, *SERPINB5*, *SFN* (as categorized variables) were significantly prognostic; grade showed a prognostic trend. A multivariate model containing nodal status, grade and *TWIST1* was a best fit ($p < 0.001$) in stepwise regression; risk-score based on this model separated patients into 3 distinct risk-groups ($p < 0.001$). A gene-only model based on *TWIST1* and *SFN* also classified patients into distinct risk-groups ($p = 0.009$).
Conclusions: This study shows that accurate quantitative measurement of DNA methylation by pyrosequencing identifies a small set of genes with independent prognostic potential in breast cancer. These genes complement the current clinico-pathological prognostic factors and appear to be potential biomarkers that warrant further validation.

Keywords: Breast cancer, prognosis, DNA Methylation, pyrosequencing, biomarker

1. Background

Breast cancer (BC) is the most common cancer in women worldwide with an estimated 1.38 million new cancer cases diagnosed every year, accounting for 23% of all cancers in women and 10.9% of all cancers (second most common cancer) globally [1]. Overall incidence of BC is steadily increasing in most parts of the world, USA being one of the recent exceptions [2]. Trends in mortality from BC, however, show a mixed picture, with mortality declining in developed countries for the past two decades [3]. Evolved management of breast cancer and availability of various treatment options has played a major role in the welcome de-

[1]DCK and MAT contributed equally.

*Corresponding author: Attila T. Lorincz, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine, Queen Mary University of London, London, EC1M 6BQ, UK. E-mail: a.lorincz@qmul.ac.uk.

ISSN 1574-0153/12/$27.50

cline in mortality [4]. Adjuvant systemic treatment in BC includes various chemotherapy regimens, a choice of endocrine therapy drugs and targeted agents like Trastuzumab. Each therapeutic option has its own characteristic benefit and adverse effect profile and therefore selecting treatment options requires precise estimation of risk of recurrence or death due to BC, particularly in view of the heterogeneous nature of the disease. This underscores the continuing need for novel prognostic biomarkers.

A few new prognostic factors like uPA [5], Oncotype Dx and Mammaprint are being evaluated [6,7] and/or used [8] in clinical practice. The use of gene-signature based prognostic tools is still quite uncommon and some consider it premature [9] due in part to limitations like the need for centralized processing and high costs, which in the developing world often exceed the cost of the entire generic drug-based chemotherapy regimen [10]. Moreover, Oncotype Dx in particular is applicable only in the subset of breast cancer patients with ER positive node negative disease. uPA was recently recommended [5] as a prognostic factor in breast cancer, its use is also limited due to a requirement for frozen tumour tissue. Therefore, there is an urgent need for robust cost-effective prognostic biomarkers that can be used in different subtypes of breast cancers; employing assays performed locally on formalin-fixed paraffin embedded (FFPE) tissues.

Epigenetic alterations are one of the most common molecular changes in human cancer [11–13]. These changes differ from genetic mutations in that they occur at higher frequency and may be reversible upon treatment with pharmacological agents. DNA methylation is one well known epigenetic-mechanism and it has become clear in recent years that there is a synergy between genetic and epigenetic changes [13]. Recognizing that aberrant DNA methylation plays a pivotal role in carcinogenesis and disease progression we assumed that accurate measurement of differential gene methylation patterns among many genes is likely to reveal biomarkers for improved risk assessment. Since most new prognostic markers in breast cancer are based on RNA expression; DNA methylation based biomarkers may provide independent prognostic information and may even complement RNA expression based prognostic factors. DNA methylation status has the ability to reflect multiple aspects of disease; types of genes methylated reflecting the biology [14,15] and the number of genes methylated and quantitative level of methylation in these genes reflecting the chronology. Therefore, DNA methylation biomarkers can prove to be useful in both diagnosis and prognosis of breast cancer.

Various approaches including high-throughput approaches could be used for identification of candidate biomarkers. Alternatively numerous (hundreds) of candidate genes (*Web-appendix*) have been reported in literature as differentially methylated in cancer but have not yet been validated as a combined prognostic set. Thus, selecting from among the myriad of candidates initially provides an enriched population from which to discover a key set, which may have a higher chance of successfully completing validation [16]. Indeed such an approach has been productive in the past for RNA expression biomarkers, e.g. Oncotype Dx assay was developed using a similar approach [17]. For biomarkers to be useful and widely used, the method/assay should be accurate, reproducible, relatively simple to perform, not prone to biases, feasible on very small amounts of FFPE material and affordable. PCR followed by pyrosequencing (PSQ) is one such sensitive, highly reproducible and cost-effective method [18] for DNA methylation analyses. It provides absolute quantitative information on methylated bases at each interrogated CpG site (CG dinucleotides in DNA where methyl group can be enzymatically added to the 5' position of cytosine), which is not possible with most other methods.

We investigated the prognostic potential of DNA methylation of 20 preferred genes chosen from an initial set of 30 candidate genes using a PCR-PSQ method on frozen tissues in a single-institution series of Korean breast cancer patients with median follow-up of 5.1 years.

2. Materials and methods

2.1. Patients, pathology specimens and handling

The study included 121 breast cancer patients presenting at Dong-A University Medical Centre, Busan, Republic of Korea from January 2004 to December 2006. All consecutive operable breast cancer patients who consented to provide fresh breast tumour and adjoining normal tissue sample (measuring approximately 0.5 $\times$ 0.5 cm each) at the time of surgery and in whom such sample collection (and immediate freezing of samples) was possible without jeopardizing clinical diagnosis and management were included in the study. Patients receiving neo-adjuvant chemotherapy (NACT) were included. Patients with missing clinicopathological data for 2 or more variables were excluded ($n = 3$), resulting in 121 eligible participants.

All specimens were centrally reviewed to confirm diagnosis by the breast pathologist (DCK). Histopathological evaluation and immunohistochemistry for ER (DAKO, Clone 1D5, 1:50), PgR (DAKO, PgR 636, 1:100) and HER2 (Neomarkers, Clone e2-4001/3B5, Fremont, CA, USA,1:200) were done as per manufacturer recommended protocols. Histological grade was determined by the Nottingham Modification of Richardson Bloom Score (RBS). HER2 expression was assessed by immunohistochemistry (IHC), HER2 FISH was not performed in all cases of 2+ by IHC, therefore, only those 3+ on IHC were considered to be HER2 positive. Clinico-pathological and treatment variables including age, tumour size, type, histological grade, lymph node status, ER, PgR, HER2, type of surgery, neo-adjuvant/adjuvant treatment details (e.g. type of chemotherapy, no. of cycles, no. of fractions of radiotherapy, tumour bed boost etc.) were recorded in the study database for all eligible patients. Data on tumour size were a mixture of clinical and pathological tumour size, and some patients also received NACT. In view of the heterogeneous data on the tumour size variable, all analyses were done excluding this variable. Patients were regularly followed up every 6 months.

This study was approved by the Institutional Review Board of Dong-A University Medical Centre (IRB-2519 /SangYun- 07-04 Ho) and all patients gave written informed consent.

Laboratory methods were as previously described [19]; detailed description is included in the *Web-appendix*.[1]

2.2. DNA extraction and bisulfite conversion

A simple macrodissection of sections from frozen tissue was performed before DNA extraction to enrich for areas of cancer. Five consecutive sections per specimen (10 μm thickness) were obtained by cryosectioning the cancer tissues and staining the first and fifth sections by H&E for histopathology review to confirm the areas of cancer and to guide the dissections of the three central sections. Genomic DNA was extracted and bisulfite converted as previously described [19].

2.3. Primer design

Primer sets with one biotin-labelled primer were used to amplify bisulfite converted DNA. Thirty genes were identified from the literature as candidate genes for this study. These genes were differentially methylated in either breast or prostate cancer, a combined candidate gene pool was based on a rationale that both of these are endocrine sensitive cancers, therefore likely to have major overlap of important biological pathways [19]. New primers for each of the 30 genes (*Web-Table W2*[1]) were designed as previously described [19]. Most amplicons were kept short with lengths between 90 to 140 base pairs (bp) to facilitate later studies on FFPE specimens. Maximum permissible size of the amplicons was 210 bp.

2.4. PCR and pyrosequencing

PCRs followed by pyrosequencing were performed using 1.2 to 2.4 ng bisulfite converted DNA as previously described [19]. Optimized primer-specific annealing temperatures are given in *Web-Table W2*. All runs contained standard curves, which comprised a range of control methylated DNA (0%, 25%, 50%, 75%, and 100%) to allow standardized direct comparisons between different experiments.

A further selection of preferred genes from the initial 30 candidate genes was performed after the first 30 samples were processed. Genes ($n = 20$) correlating ($p < 0.1$) to any of Age, Nodal status, Histological grade, ER, PgR, HER2 were selected as preferred genes for further investigation. The remaining 10 genes (*Web-Table W2*) also had very low methylation frequency and levels and therefore were even more unlikely to succeed as biomarkers. These genes were therefore not investigated further in this study; we report findings on 20 selected preferred genes.

We have previously established reproducibility of the PCR-PSQ method [19]; therefore, all samples were assayed once except for the samples which did not yield pass results on first assay.

2.5. Statistical methods and data analyses

2.5.1. Data normalization

The main analyses used average of methylation values across all CG analyzed in a particular gene segment. The number of CGs analysed varied between two to six in each gene (*Web-Table W2*). Methylation data were adjusted for primer bias through re-scaling each gene's methylation measurements by the median standard curve obtained using control mixtures for each primer set as previously described [19].

[1] Web-appendix: http://bit.ly/CCP-MEL.

2.5.2. Correlations

Correlation analyses were performed using Spearman's test to investigate associations between the categorical clinico-histopathological variables: age, nodal status, grade, ER, PgR, HER2 and continuous percentage methylation (%MeC) of genes.

2.5.3. Univariate analyses and determination of gene cut-off values

Univariate analyses using log-rank test were performed with time to recurrence (TTR) as an endpoint to probe relationship between nodal status, grade, ER, PgR, and HER2.

2.5.3.1. Genes as continuous variables

Univariate analyses with time to recurrence (TTR) as an endpoint and gene methylation levels as continuous variables were performed using the likelihood-ratio test. Only genes with p-value at or below 0.05 were considered for further analyses.

2.5.3.2. Genes as categorized variables

Relationship between DNA methylation and gene function may be non-linear in some cases and therefore analyses as continuous variables may not identify significant relationships between DNA methylation and survival. Secondly, if a biomarker emerges as successful candidate, categorized values often simplify its clinical application. Therefore, we also analysed genes as categorized variables using cut-offs determined as described below.

Various approaches are used for biomarker assessment like fixed percentage or median-based cut-off. These may not yield an optimum cut-off. Outcome-based cut-point optimization, X-Tile [20] is another approach. Cut-off determination, however, is subjective in the X-Tile approach and may also result in overfitting.

We employed an outcome-based approach (log-rank test) with 4 cut-offs based on quintiles (at 20, 40, 60 and 80 percentile methylation values) to determine optimal cut-off. The cut-offs yielding a lowest p-value were chosen (optimal cut-off by p-value method). Only genes with p-value at or below 0.05 were considered for further analyses using gene methylation categorized variable. 10000-fold permutation analyses were performed to provide exact p values.

Multiple comparisons in a small sample can result in over-fitting and yield false positive results. However, correction for multiple testing in the early phase of biomarker discovery-credentialing can also be an excessively conservative approach. We, therefore, adopted a middle path approach of accepting those genes which showed a p value < 0.05 in a relevant significance test and a p value < 0.2 in 10000-fold permutation analyses.

2.5.4. Multivariate analyses

2.5.4.1. Genes as continuous variables

Clinico-pathological variables with p-value < 0.1 in univariate analysis and genes with p-value at or below 0.05 in univariate analyses using gene methylation as continuous variable were considered for multivariate analyses. Stepwise regression using the likelihood ratio test was performed to identify the best fitting multivariate model. Multivariate analyses were performed by Cox's Proportional Hazards method.

2.5.4.2. Genes as categorized variables

Genes with p-values at or below 0.05 in univariate analyses using gene methylation as categorized variable were considered for gene-only multivariate analyses. Stepwise regression using likelihood ratio test was performed to identify the best fitting gene-only model. Multivariate analyses were performed by the Cox's Proportional Hazards method.

2.5.5. Risk score stratification

A risk-score was calculated for each case based on a multivariate model (using β coefficients) and the cases were stratified into 3 groups, low risk-score (cases within 0–25 percentiles), intermediate risk-score (cases within 26–75 percentiles) and high risk-score (cases within 76–100 percentiles).

All statistical analyses were conducted using software R version 2.11.2 [21] and STATA version 11.2 (StataCorp LP, College Station, TX, USA). All tests were two-sided and p-values of $\leqslant 0.05$ were accepted as statistically significant.

3. Results

3.1. Patient treatment details

One hundred and twenty-one patients met the inclusion criteria for this study. All patients underwent either a breast conservation surgery or a mastectomy. Axillary management was either a complete axillary clearance or a sentinel node biopsy (SNB); with complete axillary clearance in whom SNB showed metastatic node/s. Sixteen patients underwent breast conservation surgery,

Table 1
Patient characteristics

Variable	Groups	Number	Events*
Age	$\leqslant 50$	59	13
	> 50	62	12
Grade	I	18	2
	II	38	6
	III	61	17
	Not Avbl	4	0
Nodal status	Negative	60	6
	1–3 Positive	27	6
	> 4 Positive	34	13
ER	Positive	63	13
	Negative	58	12
PgR	Positive	60	12
	Negative	61	13
HER2	Positive	46	12
	Negative	72	11
	Not Avbl	3	2

*The recurrence of breast cancer.

104 underwent total mastectomy and radical mastectomy was performed in 1 patient. Forty nine patients underwent SNB as the initial axillary staging procedure. Thirteen patients received some NACT, while 99 patients received chemotherapy only post-operatively. Nine patients did not receive any chemotherapy, 46 received methotrexate-based chemotherapy (CMF), 19 received Anthracycline-based chemotherapy (FAC or FEC) and 32 patients received a sequential combination of anthracyclines and taxanes (AC/EC followed by Paclitaxel or Docetaxel). Twenty-three patients additionally received Doxifluridine after completion of standard adjuvant chemotherapy. Sixty patients received radiotherapy (all conservative surgery patients included), 16 of them received tumour bed boost. All premenopausal hormone receptor positive patients were prescribed Tamoxifen for 5 years, while postmenopausal patients were prescribed aromatase inhibitors. After a median follow-up of 5.1 years (4.87–5.4) 25 patients experienced recurrences, and there were 3 deaths; one patient died due to chemotherapy related complications after first cycle of CMF chemotherapy, cause of death was not known in the other two patients.

3.2. Patient characteristics and gene methylation

Patient characteristics are displayed in Table 1. Most ($n = 112$) patients had Invasive Ductal Carcinoma (IDC), 10 of these showed medullary features; 4 patients had Invasive Lobular Carcinoma (ILC), and 5 tumours had other morphologies like mucinous carcinoma or tubular carcinoma. Number of samples analysed and methylation values for 20 genes are displayed in *Supplementary Table 1*. Methylation distribution of 20 genes is also displayed in *Supplementary Fig. 1*; patients with and without recurrence are separated in two groups.

3.3. Correlations among histopathological and gene variables

ER showed a strong positive correlation with PR (Spearman's Rho $= 0.79$, $p < 0.001$), and negative correlation with HER2 (Rho $= -0.172, p < 0.042$), and grade (less ER positivity with increasing grade; Rho $= -0.361, p < 0.001$). PgR expectedly showed similar correlations; HER2 (Rho $= -0.252, p = 0.004$), grade (Rho $= -0.366$, $p < 0.001$). HER2 and grade did not correlate with each other and nodal status did not show correlation with any other variable.

Correlations between %MeC of genes and clinicopathological factors are displayed in *Supplementary Table 2*. None of the genes correlated with age. Only *SERPINB5* and *PDLIM4* correlated positively with grade and nodal status respectively. HER2 correlated positively with *PDLIM4*, *RARB* and *RASSF1A*. A majority of correlations were between genes and ER or PgR. ER and PgR correlated positively with *CDH13*, *EDNRB*, *EGFR5*, *HIN1*, *RASSF1A* and negatively with *RARB* and *SERPINB5*. ER alone correlated positively with *SLIT2* (*Supplementary Table 2*).

Correlations among the 20 genes are displayed as a heatmap in *Supplementary Fig. 2*. While most genes showed weak to moderate positive correlations, certain negative correlations were also observed, particularly of *SERPINB5*, *SFN* and *TWIST1*. *SERPINB5* correlated negatively with *EGFR5*, *MAL*, *SLIT2*, *HIN1*, *CDH13*, *EDNRB* and *RASSF1A*.

3.4. Univariate survival analyses

Univariate survival analysis explored relationships between age, histological grade, nodal status, ER, PgR, HER2 and TTR. Of the clinical variables, only nodal status ($p = 0.004$) was significantly associated with TTR (Fig. 1A). Histological grade showed a non-significant trend of association with TTR ($p = 0.086$). Both variables were considered for multivariate analysis.

Univariate survival analysis of genes as continuous variables and as categorized variables is displayed in Table 2. Resulting quintile-based cut-offs, number of cases and events in two groups created by dichotomization at these cut-offs are also displayed. In analyses

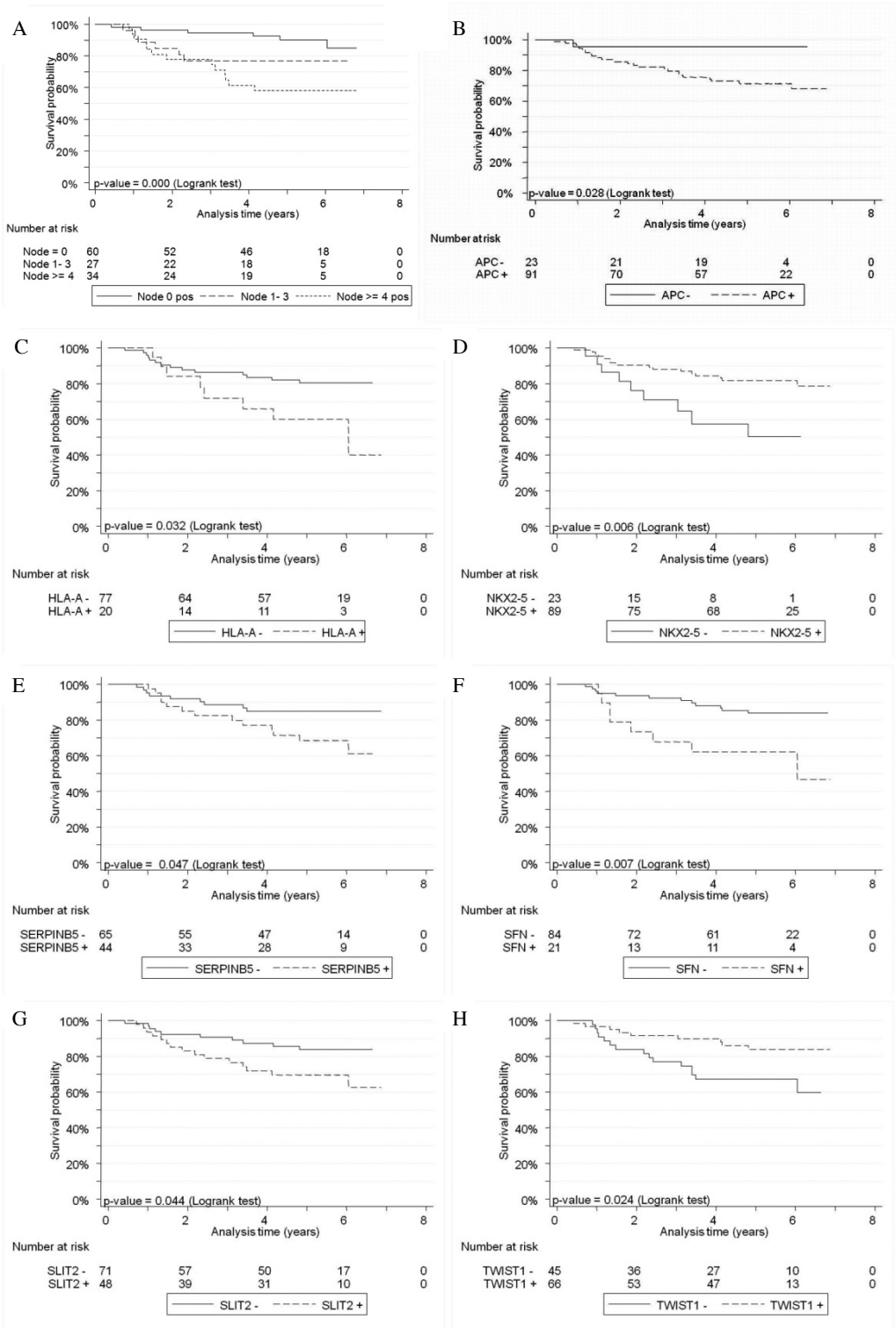


Fig. 1. Kaplan-Meier survival estimates of nodal status (A), *APC* (B), *HLA-A* (C), *NKX2-5* (D), *SERPINB5* (E), *SFN* (F), *SLIT2* (G) and *TWIST1* (H) in univariate analysis. $^{+}$ sign refers to cases with methylation values above cut-off and – sign refers to those below cut-off.

Table 2
Univariate survival analysis of genes as categorized variables, their derived cut-offs (percentile and corresponding level of methylation), resulting groups, p-values and as continuous variables (p-values in the last column)

Gene	Categorized					Continuous
	Cut-off percentile	Cut-off %MeC	Low Methylation N (Events)	High Methylation N (Events)	p-value	p-value
APC*	20	1.1	23 (1)	91 (24)	0.0094*	0.13
CDH13	20	3.8	23 (7)	90 (16)	0.1005	0.493
DAPK1	40	2.0	47 (12)	69 (10)	0.0831	0.681
DPYS	20	12.6	25 (7)	92 (18)	0.3847	0.735
EDNRB	80	42.5	94 (21)	24 (3)	0.2003	0.642
EGFR5	20	33.8	23 (4)	89 (20)	0.6318	0.967
GSTP1	80	19.8	94 (17)	24 (7)	0.1826	0.123
HIN1	40	9.1	45 (11)	67 (12)	0.2579	0.662
HLA-A*	80	26.6	77 (14)	20 (8)	0.0495*	0.069
MAL	20	10.5	24 (7)	92 (16)	0.2331	0.903
MDR1	40	4.8	45 (6)	68 (16)	0.2689	0.553
NKX2-5*	20	5.1	23 (9)	89 (16)	0.0141*	0.806
PDLIM4	60	23.3	69 (12)	46 (12)	0.1545	0.146
PTGS2	20	2.1	23 (7)	92 (16)	0.0983	0.405
RARB	80	15.5	94 (19)	24 (6)	0.5511	0.866
RASSF1A	20	2.4	24 (6)	95 (19)	0.3496	0.566
SERPINB5*	60	67.3	65 (9)	44 (13)	0.0506*	0.131
SFN*	80	86.2	84 (12)	21 (8)	0.0167*	0.182
SLIT2*	60	31.3	71 (10)	48 (15)	0.0473*	0.015*
TWIST1*	40	7.0	45 (15)	66 (9)	0.0259*	0.034*

Low methylation: $\leqslant$ Cut-off; High methylation: $>$ Cut-off; *p </= 0.05.

of genes as continuous variables, *TWIST1* ($p = 0.034$) and *SLIT2* ($p = 0.015$) were significantly associated with TTR, both genes were considered for subsequent multivariate analysis.

Seven genes (*APC*, *HLA-A*, *NKX2-5*, *SERPINB5*, *SFN*, *SLIT2* and *TWIST1*) were significantly associated with TTR (as categorised variables). Cut-offs derived by the quintile-based method were tested through permutations (B = 10000) and 5 of 7 genes showed p-values less than 0.1 while *SERPINB5* and *SLIT2* had p between 0.1 and 0.16 (Data not shown). All these genes except *NKX2-5*, showed p-values less than 0.2 in survival analyses as continuous variables as well (Table 2). Kaplan-Meier survival plots are displayed for *APC*, *HLA-A*, *NKX2-5*, *SFN*, *SLIT2* and *TWIST1* in Figs 1B, C, D, E, F, G, and H respectively.

3.5. Multivariate analyses

3.5.1. Genes as continuous variables

Stepwise regression using the likelihood ratio test and nodal status, grade, *SLIT2* and *TWIST1* as variables was performed on 109 cases (24 events, data available on all 4 variables); the best fitting multivariate model comprised of 3 variables, nodal status [HR (95%CI) = 1.81(1.12–2.95); $p = 0.016$], grade [HR (95%CI) = 1.77(0.88–3.55); $p = 0.108$], and *TWIST1* [HR of 1% change in methylation (95%CI) = 0.94(0.88–1.01); $p = 0.092$].

3.5.2. Genes as categorized variables: Gene-only model

Stepwise regression using the likelihood ratio test and 7 genes as categorised variables was performed on 72 cases (17 events, data available on all 7 variables); the best fitting gene-only model comprised *SFN*, and *TWIST1* (Fig. 2). Cases with both genes showing methylation either below or above their respective cut-offs belonged to an intermediate risk group, whereas those with *SFN* methylated below cut-off and *TWIST1* methylated above cut-off belonged to a low-risk group. A small group of cases with *SFN* methylated above cut-off and *TWIST1* methylated below cut-off showed very poor survival.

3.6. Risk score stratification

A risk-score was calculated for each case based on a multivariate model (using β coefficients). The information was entered in the model as follows: node negative = 0, 1-3 nodes positive = 1, 4 or more nodes positive = 2; grade 1 = 0, grade 2 = 1, grade 3 = 2, and *TWIST1* methylation values as follows:

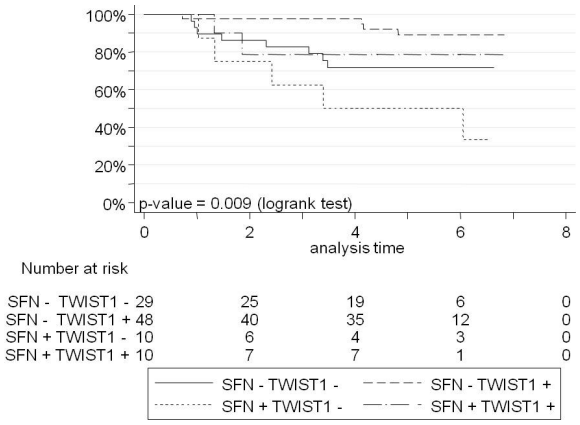


Fig. 2. Kaplan-Meier survival plots of gene-only model, genes as categorised variables.

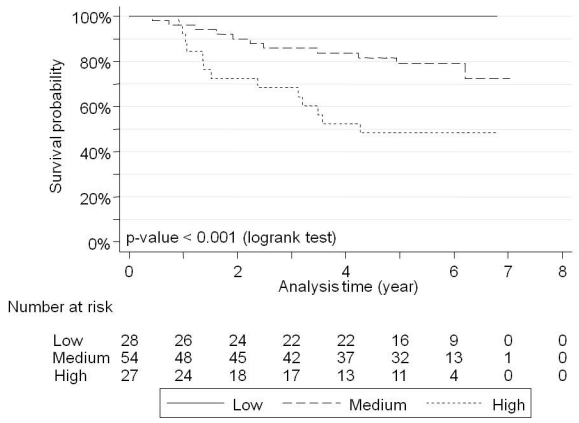


Fig. 3. Kaplan-Meier survival estimates of risk-score based stratification; stratified in 3 groups, low risk-score (cases within 0–25 percentiles), intermediate risk-score (cases within 26–75 percentiles) and high risk-score (cases within 76–100 percentiles). Nodal status, grade and *TWIST1* methylation level used to estimate risk-score.

$$\text{Risk-Score} = (0.60 \times \text{Node}) + (0.57\,\text{Grade}) - (0.06 \times \textit{TWIST1})$$

A risk-score based stratification of the cases in 3 groups, low risk-score (cases within 0–25 percentiles), intermediate risk-score (cases within 26–75 percentiles) and high risk-score (cases within 76–100 percentiles) resulted in a very clear separation of patients with different risk profiles, with low-risk group patients being virtually recurrence free (having received systemic treatment) and high-risk group patients having less than 50% 5-year survival despite receiving systemic treatment (Fig. 3).

4. Discussion

Alterations in DNA methylation contribute to development and progression of invasive breast cancer [22], and studying the methylation of genes is likely to reveal useful novel biomarkers. Several methods of determination of promoter methylation have been described including the use of restriction enzymes [23], genomic bisulfite sequencing [24] and microarray-based methylation analysis [25]. The overwhelming majority of published data use methylation-specific PCR following bisulfite treatment (MSP) [26,27]. MSP is highly sensitive and has been used widely in this context. However, as there is no in-built measure of adequacy of bisulfite treatment; the possibility of false positives due to inadequate conversion of non-methylated cytosine to uracil exists. Another potential source of false positives is mis-priming, and this may be a greater problem when high numbers of PCR cycles or if nested primers are used. Previously described methods of controlling for mis-priming include re-analysis by methylation-sensitive restriction enzymes [28] or subsequent bisulfite DNA sequencing [29]. Accurate quantitative analysis is very important for DNA methylation analysis and MSP has many weaknesses in this regard. One of the major issues for MSP is the use of relative quantification by means of an external control with a different PCR of a control gene. The accuracy and reproducibility of this approach is still not adequately answered for a large diversity of genes. Development of DNA methylation risk prediction panels based on diverse genes will however require an accurate and reproducible method. High reliability and flexibility have made pyrosequencing an analysis platform that has been widely used for various diagnostic applications such as routine (multiplex) genotyping [30], sensitive detection of mutations [31] and microbacterial identification [32]. Pyrosequencing combines the ability of direct quantitative sequencing, reproducibility, speed and ease-of-use and is becoming more widely used [33].

Our aim was to investigate and credential a set of candidate genes previously reported as differentially methylated in endocrine-responsive cancers [19] in order to identify potential prognostic biomarkers in breast cancer. This is one of the first studies to quantify methylation of a large number of candidate genes in a common set of breast cancer specimens by means of a single highly accurate and validated DNA methylation assay.

To ensure accuracy and reproducibility, we used PSQ, an approach particularly suited to situations where the available specimen set for molecular studies is highly limited in either quality or amount. As-

says were specifically designed for future use in FFPE material by limiting the amplicon size but covering a reasonable number of CG sites. This will, in future, allow us to validate our findings in larger sets of samples from clinical trials and develop the biomarkers for clinical use. In addition, our preliminary data show that methylation levels in frozen and matched FFPE specimens are equivalent [19]. To fine-tune the PSQ method, predefined quantitative controls were included within each experiment to allow the adjustment of all data and thus improve the accuracy of comparisons in levels of methylation between the investigated genes. Internal controls to check completeness of bisulfite conversion were also included in the assays. This is an important advantage of the PSQ method, increasing the measurement accuracy compared to other commonly used methods [19].

We used a set of prospectively collected well-annotated Korean breast cancer samples from patients treated at Dong-A University hospital. There are certain important differences between Korean and Western breast cancers, both in risk factors as well as presentation. Apart from dietary [34,35] and ethnic differences, Korean women with breast cancer present at younger mean age of 48 years, approximately 15 years younger than the median age of white women with breast cancer in the United States [36]. Expectedly, ER positivity rates are lower [37], whereas HER2 positivity is much higher in Korean breast cancers [38,39], as seen in our data (ER positive 52%, HER2 positive 39%). Fifty percent of our patients had node positive breast cancer and the same proportion had grade 3 disease. This is consistent with a higher proportion (49% were 50 years or younger) of young patients [37]. Correlations among various histopathological variables were consistent with general breast cancer populations. Although breast conservation rates were low, patients received standard adjuvant treatments. Use of multiple standard chemotherapy regimens and different numbers of cycles (6 or 8), however, did not permit us to use chemotherapy regimen as one of the variables in survival and other analyses as it would have resulted in multiple small subgroups.

We explored the relationship between genes and TTR in two ways, genes treated as continuous variables and as categorized variables (optimum cut-off determination) for the reasons described in methods. Seven genes, *TWIST1*, *SFN*, *SLIT2*, *APC*, *NKX2-5*, *HLA-A*, and, *SERPINB5* were univariately associated with TTR when analysed as categorical variables; *TWIST1* and *SLIT2* were also associated with TTR when analysed as continuous variables. Optimum cut-off determination can result in over-fitting and false-discovery, we therefore tested robustness of our discovery by testing cut-offs derived by the quintile-based method through 10000 permutations and 5 of 7 genes showed p-values less than 0.1 and two were between 0.10 and 0.16 (*SERPINB5* and *SLIT2*). We believe that these 7 genes and the cut-offs derived are sufficiently robust for the biomarker discovery-credentialing phase of our study. Of these 7 genes, only *SERPINB5* correlated with ER, PgR and grade, *SLIT2* correlated with ER (*Supplementary Table 2*). Correlation among these 7 genes was none to weak (*Supplementary Fig. 2*), suggesting that the genes can provide independent prognostic information.

Univariate analysis of clinic-pathological variables revealed nodal status as strongly prognostic and histological grade exhibiting a prognostic trend. ER, PgR, HER2, which are very strong predictive but weak prognostic factors, were not prognostic, perhaps due to moderate sample size and high HER2 positivity or due to biological differences between Korean and Western breast cancers.

A step-wise approach was used to identify the best-fit model for primary multivariate analysis, using *TWIST1* and *SLIT2* as continuous variables, in addition to nodal status and grade. A model with nodal status, grade and *TWIST1* was the best model ($p < 0.0001$ vs. null) and *TWIST1* may have better prognostic ability than grade although neither reached significance when combined with nodes alone.

Additional exploratory analyses were performed to see if a risk-score based on this 3 variable model or alternatively a gene-only model can classify patients into different risk categories. The gene-only model was developed by a step-wise approach from a larger number of genes (7 significant genes in analysis using genes as categorized variables). *TWIST1* and *SFN* produced the best model ($p = 0.009$ for model vs. null) that could identify a small group of cases with very high risk of recurrence. Such models may be useful when risk-prediction based on nodal status etc is not available, for example, in tailoring neo-adjuvant therapies based on individual patient risk. Stratification of patients into 3 groups based on risk-score showed excellent separation of patients with different risks of recurrence, with none of the low-risk group patients experiencing a recurrence within 5 years and the highest risk patients having less than 50% 5-year survival.

All genes except *NKX2-5* have a proven important role in carcinogenesis and tumour progression. For ex-

ample, *TWIST1* acts as an oncogene and is a key driver of epithelial-mesenchymal transition (EMT) [40,41] and is directly involved in breast cancer stem cell phenotype [42]. Higher methylation of *TWIST1* was associated with better prognosis in our study as expected. *SFN* acts as a tumour suppressor gene to produce protein 14-3-3 sigma, which acts as a regulator of mitotic translation; deficiency of this protein may cause cells to progress on the path to aneuploidy and tumorigenesis [43]. *SLIT2*, a tumour suppressor gene, suppresses breast cancer cell growth [44].

Ideally, the question of prognostic potential should be addressed in a systemic therapy naive population. Unfortunately, getting such contemporary sample sets is becoming increasingly difficult to impossible since most patients now receive systemic therapy and those who do not are not representative of a general pool of breast cancer patients; therefore most major biomarker studies now have to resort to sample sets from systemically treated patients. Our study has some other limitations as well. The study was small and not blinded and there were no comparisons to other well established DNA methylation methods such as bisulfite sequencing. However, inclusion of quantitative controls in every experimental run and inclusion of non-CpG C controls wherever possible ensured reliability of our assays. Since, this was an exploratory study of several markers, there were no formal *a-priori* sample size calculations and we used all samples available from the pathology archives. With high HER2 positivity and a younger population, our patient set is representative of a Korean breast cancer population, but probably not of a western breast cancer population. Although breast conservation rates were lower patient management was generally as per international standards, Dong-A University hospital being a leading cancer centre in Korea. Information on tumour size was not uniform and use of several chemotherapy regimens did not allow us to investigate these variables and their impact on survival, this may have resulted in sub-optimal multivariate analyses. The cut-offs for genes were derived by univariate analysis of the same specimens, thus there were no separate training and validation sets. We chose not to divide our patient set into training and validation sets in view of modest sample size, however, we confirmed robustness of the cut-offs through permutations. Also 2 of 7 genes were significantly associated with TTR when analysed as continuous variables an approach that is less prone to overfitting bias. It should, however, be emphasized that, additional extensive validation of our findings is important in order to establish the clinical utility of these biomarkers and any models based on them.

We identified a set of 7 genes univariately associated with TTR of which *TWIST1*, *SFN* and *SLIT2* either complemented node status and grade in providing additional prognostic information or stratified patients into different risk groups when evaluated as a gene-only model (Figs 2 and 3). Since these 7 genes did not correlate with conventional clinico-pathological variables, they are likely to provide independent prognostic information. If these data are confirmed in larger independent studies it may be reasonable to explore an approach where a gene-only model helps in tailoring neo-adjuvant therapy and risk-score helps in tailoring adjuvant therapy, with a possibility that low-risk group patients may forgo chemotherapy and avoid associated side-effects.

In summary, PSQ is a useful and accurate method for studying methylation of genes. We identified seven differentially methylated genes in breast cancer that were informative of breast cancer recurrence and deserve further validation as prognostic markers in larger sets of specimens.

Competing interests

ATL is an advisor to Qiagen Inc. and owns Qiagen stock. JMC has given lectures sponsored by Qiagen on an occasional basis. No other relevant competing interests exist.

Acknowledgments

This study was supported by grants from Cancer Research UK (Grant number C569/A10404), by a Korea Breast Cancer Foundation grant to DCK and by an unrestricted research grant from Qiagen Inc Hilden Germany.

Sponsors had no role in study design, execution, data analyses or the content of the paper.

References

[1] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin, Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*, **127**(**12**), (2010), 2893-917.

[2] M. A. Thorat, What caused the decline in US breast cancer incidence? *Nat Clin Pract Oncol*, **5**(**6**), (2008), 314-5.

[3] I. Jatoi and A. B. Miller, Why is breast-cancer mortality declining? *Lancet Oncol*, **4**(**4**), (2003), 251-4.

[4] D. A. Berry, K. A. Cronin, S. K. Plevritis, D. G. Fryback, L. Clarke, M. Zelen, et al., Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*, **353**(**17**), (2005), 1784-92.

[5] L. Harris, H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. Taube, et al., American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*, **25**(**33**), (2007), 5287-312.

[6] J. A. Sparano, TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer*, **7**(**4**), (2006), 347-50.

[7] J. Bogaerts, F. Cardoso, M. Buyse, S. Braga, S. Loi, J. A. Harrison, et al., Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol*, **3**(**10**), (2006), 540-51.

[8] F. O. Ademuyiwa, A. Miller, T. O'Connor, S. B. Edge, M. A. Thorat, G. W. Sledge, et al., The effects of oncotype DX recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort. *Breast Cancer Res Treat*, **126**(**3**), (2011), 797-802.

[9] M. A. Thorat, Gene-signature-based prognostic tools in breast cancer: not yet. *Lancet*, **369**(**9571**), (2007), 1428.

[10] M. A. Thorat, Herceptin in early breast cancer: a call for judicious use. *Natl Med J India*, **18**(**6**), (2005), 315-7.

[11] S. B. Baylin, S. A. Belinsky and J. G. Herman, Aberrant methylation of gene promoters in cancer – concepts, misconcepts, and promise. *J Natl Cancer Inst*, **92**(**18**), (2000), 1460-1.

[12] P. A. Jones, DNA methylation errors and cancer. *Cancer Res*, **56**(**11**), (1996), 2463-7.

[13] P. A. Jones and P. W. Laird, Cancer epigenetics comes of age. *Nat Genet*, **21**(**2**), (1999), 163-7.

[14] J. A. Ronneberg, T. Fleischer, H. K. Solvang, S. H. Nordgard, H. Edvardsen, I. Potapenko, 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*, **5**(**1**), 61-76.

[15] S. Kamalakaran, V. Varadan, H. E. Giercksky Russnes, D. Levy, J. Kendall, A. Janevski, 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*, **5**(**1**), 77-92.

[16] D. F. Ransohoff, Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer*, **4**(**4**), (2004), 309-14.

[17] M. A. Cobleigh, P. Bitterman, J. Baker, M. Cronin, M. L. Liu, R. Borchik, et al., Tumor Gene Expression Predicts Distant Disease-Free Survival (DDFS) in Breast Cancer Patients with 10 or More Positive Nodes: High Throughput RT-PCR Assay of Paraffin-Embedded Tumor Tissues. *Proc Am Soc Clin Oncol*, **22**, (2003), 850.

[18] S. Marsh, Pyrosequencing applications. *Methods Mol Biol*, **373**, (2007), 15-24.

[19] N. Vasiljevic, K. Wu, A. R. Brentnall, D. C. Kim, M. A. Thorat, S. C. Kudahetti, et al., Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. *Dis Markers*, **30**(**4**), 151-61.

[20] R. L. Camp, M. Dolled-Filhart and D. L. Rimm, X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*, **10**(**21**), (2004), 7252-9.

[21] R. D. C. Team. R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*: Vienna, Austria, 2009.

[22] M. Esteller, P. G. Corn, S. B. Baylin and J. G. Herman, A gene hypermethylation profile of human cancer. *Cancer Res*, **61**(**8**), (2001), 3225-9.

[23] J. Singer-Sam, M. Grant, J. M. LeBon, K. Okuyama, V. Chapman, M. Monk, et al., Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol Cell Biol*, **10**(**9**), (1990), 4987-9.

[24] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, et al., A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA*, **89**(**5**), (1992), 1827-31.

[25] S. Huang, Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *J Mol Med*, **77**(**6**), (1999), 469-80.

[26] J. G. Herman, J. R. Graff, S. Myohanen, B. D. Nelkin and S. B. Baylin, Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA*, **93**(**18**), (1996), 9821-6.

[27] S. Derks, M. H. Lentjes, D. M. Hellebrekers, A. P. de Bruine, J. G. Herman and M. van Engeland, Methylation-specific PCR unraveled. *Cell Oncol*, **26**(**5-6**), (2004), 291-9.

[28] S. A. Belinsky, D. M. Klinge, J. D. Dekker, M. W. Smith, T. J. Bocklage, F. D. Gilliland, et al., Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res*, **11**(**18**), (2005), 6505-11.

[29] K. Kawamoto, H. Enokida, T. Gotanda, H. Kubo, K. Nishiyama, M. Kawahara, et al., p16INK4a and p14ARF methylation as a potential biomarker for human bladder cancer. *Biochem Biophys Res Commun*, **339**(**3**), (2006), 790-6.

[30] T. Langaee and M. Ronaghi, Genetic variation analyses by Pyrosequencing. *Mutat Res*, **573**(**1-2**), (2005), 96-102.

[31] S. Ogino, T. Kawasaki, M. Brahmandam, L. Yan, M. Cantor, C. Namgyal, et al., Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *J Mol Diagn*, **7**(**3**), (2005), 413-21.

[32] S. C. Clarke, Pyrosequencing: nucleotide sequencing technology with bacterial genotyping applications. *Expert Rev Mol Diagn*, **5**(**6**), (2005), 947-53.

[33] J. Tost and I. G. Gut, DNA methylation analysis by pyrosequencing. *Nat Protoc*, **2**(**9**), (2007), 2265-75.

[34] Y. A. Cho, J. Kim, K. S. Park, S. Y. Lim, A. Shin, M. K. Sung, et al., Effect of dietary soy intake on breast cancer risk according to menopause and hormone receptor status. *Eur J Clin Nutr*, **64**(**9**), 924-32.

[35] A. Shin, J. Kim, S. Y. Lim, G. Kim, M. K. Sung, E. S. Lee, et al., Dietary mushroom intake and the risk of breast cancer based on hormone receptor status. *Nutr Cancer*, **62**(**4**), 476-83.

[36] B. A. Miller, L. N. Kolonel, L. Bernstein, J. J. L. Young, G. M. Swanson, D. West, et al. Racial/Ethnic Patterns of Cancer in the United States, 1988–1992. In *National Cancer Institute. NIH Pub. No. 96-4104*, Miller BA, ed.; Diane Publishing: Bethsda, MD, 1996.

[37] L. A. Carey, C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, et al., Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*, **295**(**21**), (2006), 2492-502.

[38] M. L. Telli, E. T. Chang, A. W. Kurian, T. H. Keegan, L. A. McClure, D. Lichtensztajn, et al., Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat*, **127**(**2**), (2011), 471-478.

[39] D. H. Choi, D. B. Shin, M. H. Lee, D. W. Lee, D. Dhandapani, D. Carter, et al., A comparison of five immunohistochemi-

cal biomarkers and HER-2/neu gene amplification by fluorescence in situ hybridization in white and Korean patients with early-onset breast carcinoma. *Cancer*, **98**(**8**), (2003), 1587-95.

[40] R. A. Whipple, M. A. Matrone, E. H. Cho, E. M. Balzer, M. I. Vitolo, J. R. Yoon, et al., Epithelial-to-mesenchymal transition promotes tubulin detyrosination and microtentacles that enhance endothelial engagement. *Cancer Res*, **70**(**20**), 8127-37.

[41] E. Foubert, B. De Craene and G. Berx, Key signalling nodes in mammary gland development and cancer. The Snail1-Twist1 conspiracy in malignant breast cancer progression. *Breast Cancer Res*, **12**(**3**), 206.

[42] F. Vesuna, A. Lisok, B. Kimble and V. Raman, Twist modulates breast cancer stem cells by transcriptional regulation of CD24 expression. *Neoplasia*, **11**(**12**), (2009), 1318-28.

[43] E. W. Wilker, M. A. van Vugt, S. A. Artim, P. H. Huang, C. P. Petersen, H. C. Reinhardt, et al., 14-3-3sigma controls mitotic translation to facilitate cytokinesis. *Nature*, **446**(**7133**), (2007), 329-32.

[44] A. Dallol, N. F. Da Silva, P. Viacava, J. D. Minna, I. Bieche, E. R. Maher, et al., SLIT2, a human homologue of the Drosophila Slit2 gene, has tumor suppressor activity and is frequently inactivated in lung and breast cancers. *Cancer Res*, **62**(**20**), (2002), 5874-80.

Supplementary material

Supplementary Table 1

Number of samples analysed and methylation values for 20 genes

Gene	N	Q1	Median	Q3	MAD	Min	Max
APC	114	1.62	5.64	14.06	6.998	0	98.5
CDH13	113	7.2	15.79	31.523	18.421	0	100
DAPK1	116	0	3.43	8.68	5.085	0	37.8
DPYS	117	14.044	21.812	39.312	16.735	0	95.2
EDNRB	118	7.64	21.08	37.56	22.15	0	94.1
EGFR5	111	35.78	48.725	60.19	18.725	0	100
GSTP1	118	2.2	6.35	15.65	8.273	0	100
HIN1	112	6.23	14.8	42.275	16.457	0	78
HLA-A	97	0	0	15.927	0	0	100
MAL	116	12.11	22.58	36.08	16.783	0	80
MDR1	113	2.875	7.1	16.68	9.133	0	75.88
NKX2-5	112	5.37	10.12	22.925	8.332	0	61.75
PDLIM4	115	10.407	19.44	33.778	15.953	0	85.58
PTGS2	115	4.075	11.35	22.333	12.639	0	80.83
RARB	118	0	2.48	12.8	3.677	0	71.4
RASSF1A	118	3.9	27.18	61.7	36.857	0	100
SFN	105	54.853	70.2	82.8	21.349	0	100
SERPINB5	108	50.575	63.1	73.1	17.05	0	100
SLIT2	119	14.998	25.705	40.195	18.347	0	95.88
TWIST1	111	4.253	8.85	13.84	7.369	0	52.78

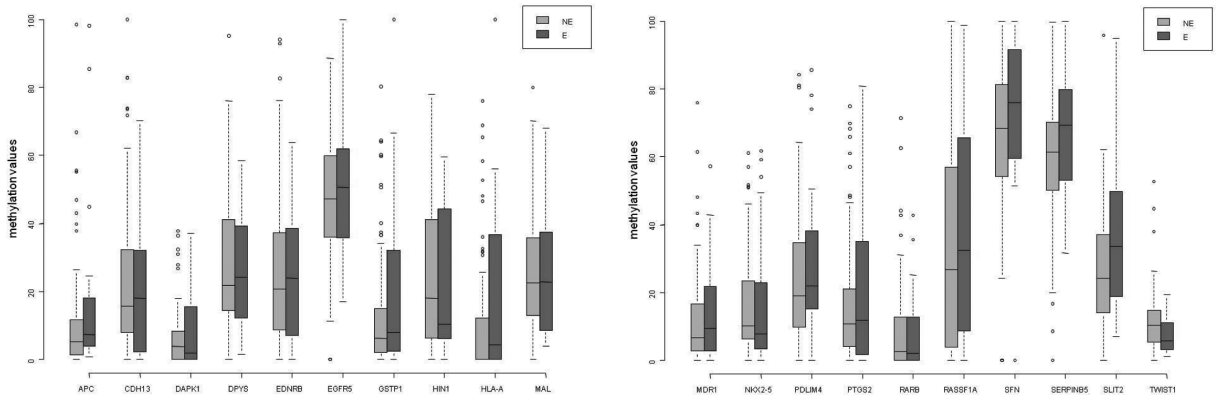
Q1: First quartile, Q3: Third quartile, MAD: Median Absolute Deviation.

Supplementary Table 2

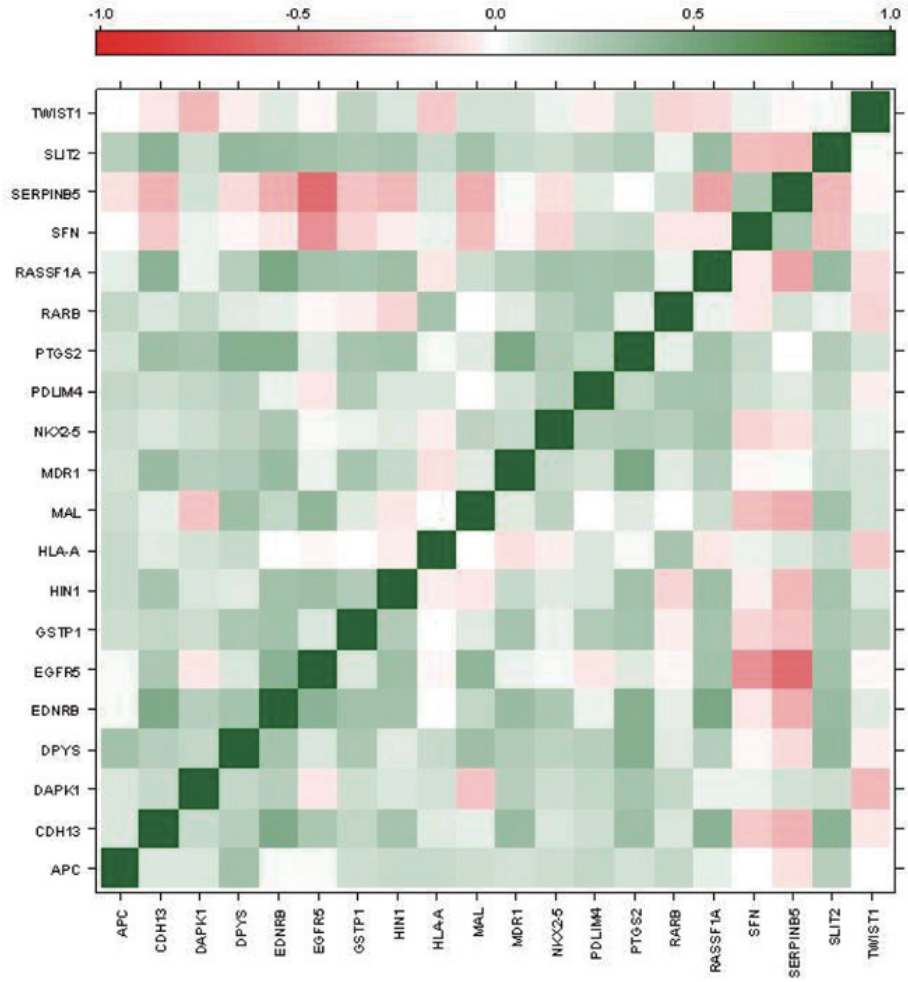
Correlation of various clinico-pathological variables with genes

Gene	Age		ER		PgR		Grade		HER2		Nodal Status	
	Rho	P-value	Rho	P-value	Rho	P-value	Rho	P-value	Rho	P-value	Rho	P-value
APC	0.146	0.122	0.005	0.957	−0.124	0.187	**0.222**	**0.019**	0.139	0.152	−0.035	0.712
CDH13	0.014	0.887	**0.275**	**0.003**	**0.2**	**0.034**	−0.185	0.051	0.088	0.359	−0.08	0.401
DAPK1	0.076	0.415	−0.061	0.514	−0.019	0.836	0.016	0.86	−0.019	0.856	−0.035	0.708
DPYS	0.055	0.556	−0.029	0.757	−0.02	0.831	0.14	0.14	0.061	0.524	0.034	0.712
EDNRB	−0.044	0.634	**0.397**	**0**	**0.373**	**0**	−0.105	0.263	0.102	0.276	0.076	0.411
EGFR5	−0.12	0.208	**0.343**	**0**	**0.301**	**0.001**	−0.016	0.892	−0.125	0.192	0.057	0.547
GSTP1	−0.134	0.148	0.155	0.093	0.07	0.448	−0.032	0.756	0.178	0.058	0.169	0.067
HIN1	−0.036	0.709	**0.449**	**0**	**0.375**	**0**	−0.062	0.505	−0.131	0.175	0.142	0.135
HLA-A	−0.037	0.72	−0.163	0.111	−0.157	0.124	0.139	0.177	0.066	0.529	0.033	0.749
MAL	0.023	0.809	0.002	0.987	−0.059	0.531	0.099	0.306	0.128	0.178	0.004	0.969
MDR1	0.049	0.604	0.056	0.557	0.058	0.54	0.06	0.545	0.181	0.059	0.072	0.451
NKX2-5	−0.031	0.746	−0.093	0.327	0	0.998	0.132	0.171	0.154	0.11	0.027	0.777
PDLIM4	−0.042	0.658	−0.165	0.078	−0.146	0.119	0.168	0.077	**0.212**	**0.024**	**0.215**	**0.021**
PTGS2	0.098	0.298	0.054	0.566	0.159	0.09	0.032	0.745	0.083	0.389	0.026	0.779
RARB	0.046	0.622	*−0.292*	*0.001*	*−0.311*	*0.001*	0.139	0.143	**0.23**	**0.013**	−0.129	0.165
RASSF1A	−0.15	0.102	**0.374**	**0**	**0.358**	**0**	−0.049	0.602	**0.203**	**0.027**	0.06	0.52
SFN	0.129	0.188	−0.19	0.052	−0.177	0.071	0.004	0.962	0.05	0.618	0.031	0.753
SERPINB5	0.099	0.306	*−0.376*	*0*	*−0.265*	*0.005*	**0.262**	**0.006**	0.042	0.661	−0.008	0.93
SLIT2	0.021	0.818	**0.254**	**0.005**	0.169	0.066	0.063	0.506	0.114	0.219	0.117	0.204
TWIST1	0.047	0.622	0.039	0.688	0.033	0.727	−0.025	0.785	−0.006	0.958	−0.094	0.325

All significant correlations have been shown in bold and significant negative correlations in italics.
P-values have not been corrected for multiple comparisons.



Supplementary Fig. 1. Methylation distribution of 20 genes, patients separated in two groups, one with recurrence events (E) and other without recurrence events (NE).



Supplementary Fig. 2. Heatmap of correlations among 20 genes. Negative correlations are indicated by "-" sign in this B&W figure. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/CBM-2012-0266)