

This Provisional PDF corresponds to the article as it appeared upon acceptance. Copyedited and fully formatted PDF and full text (HTML) versions will be made available soon.

DNA methylation epigenotypes in breast cancer molecular subtypes

Breast Cancer Research 2010, **12**:R77 doi:10.1186/bcr2721

Naiara G Bediaga (naiarabediaga@gmail.com)
Amelia Acha (ameachasag@gmail.com)
Isabel Guerra (ISABEL.GUERRAMERINO@osakidetza.net)
Amparo Viguri (MARIAAMPARO.VIGURIDIAZ@osakidetza.net)
Maria del Carmen Albaina (MARIADELCARMEN.ALBAINA@osakidetza.net)
Irene Ruiz (IRUNE.RUIZDIAZ@osakidetza.net)
Ricardo Rezola (patologia@oncologico.org)
Maria Jesus Alberdi (mjalberdi@onkologikoa.org)
Joaquin Dopazo (jdopazo@cipf.es)
David Montaner (dmontaner@cipf.es)
Mertxe Renobales (mertxe.derenobales@ehu.es)
Agustin F Fernandez (afernandezf@cniio.es)
Field K John (j.k.field@liv.ac.uk)
Mario F Fraga (mffraga@cnb.csic.es)
Triantafillos Liloglou (liloglout@roycastle.liv.ac.uk)
Marian M. de Pancorbo (marianpancorbo@gmail.com)

ISSN 1465-5411

Article type Research article

Submission date 22 April 2010

Acceptance date 29 September 2010

Publication date 29 September 2010

Article URL <http://breast-cancer-research.com/content/12/5/R77>

This peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Breast Cancer Research* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Breast Cancer Research* go to

© 2010 Bediaga *et al.*, licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DNA methylation epigenotypes in breast cancer molecular subtypes

Naiara G Bediaga¹, Amelia Acha², Isabel Guerra³, Amparo Viguri³, Maria del Carmen Albaina³, Iruñe Ruiz⁴, Ricardo Rezola⁵, Maria Jesus Alberdi⁵, Joaquín Dopazo⁶, David Montaner⁶, Mertxe Renobales⁷, Agustín F Fernandez⁸, Field K John⁹, Mario F Fraga¹⁰, Triantafillos Liloglou⁹ and Marian M de Pancorbo^{1*}

¹BIOMICs Research Group, Centro de Investigación y Estudios Avanzados 'Lucio Lascaray', University of the Basque Country UPV/EHU, Miguel de Unamuno 3,1006, Vitoria-Gasteiz, Spain

²Oral Medicine and Pathology Unit, Faculty of Medicine and Dentistry, University of the Basque Country UPV/EH, Barrio Sarriena s/n. E-48940 Leioa. Vizcaya, Spain

³Service of Anatomic Pathology, Hospital Txagorritxu, C/ José Achótegui s/n, E-01009 Vitoria-Gasteiz, Alava, Spain

⁴Service of Anatomic Pathology, Hospital Donostia, Paseo Beguiristain 107-115, E-20014 San Sebastián, Guipúzcoa, Spain

⁵Service of Anatomic Pathology, Instituto Oncológico, C/Aldakoenea 44, E-20012 San Sebastián, Spain

⁶Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe, Avda. Autopista del Saler, 164, E-6012 Valencia, Spain

⁷Molecular Biology and Biochemistry, Faculty of Pharmacy, University of the Basque Country UPV/EH, Paseo de la Universidad 7, E-01006 Vitoria-Gasteiz, Spain

⁸Cancer Epigenetics Laboratory, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), HUCA, Universidad de Oviedo, E- 33006 Oviedo, Spain

⁹Liverpool CR-UK Cancer Research Centre, University of Liverpool Cancer, School of Cancer Studies, 200 London Road, Liverpool L3 9TA, UK

¹⁰Department of Immunology and Oncology, National Center for Biotechnology, CNB-CSIC, Cantoblanco, E-28049 Madrid, Spain

*Corresponding author: Marian M de Pancorbo. Email: marianpancorbo@gmail.com

Abstract

Introduction: Identification of gene expression based breast cancer subtypes is considered as a critical means of prognostication. Genetic mutations along with epigenetic alterations contribute to gene expression changes occurring in breast cancer. So far, these epigenetic contributions to sporadic breast cancer subtypes have not been well characterized, and there is only a limited understanding of the epigenetic mechanisms affected in those particular breast cancer subtypes. The present study was undertaken to dissect the breast cancer methylome and deliver specific epigenotypes associated with particular breast cancer subtypes.

Methods: Using a microarray approach we analyzed DNA methylation in regulatory regions of 806 cancer related genes in 28 breast cancer paired samples. We subsequently performed substantial technical and biological validation by Pyrosequencing, investigating the top qualifying 19 CpG regions in independent cohorts encompassing 47 basal-like, 44 ERBB2+ overexpressing, 48 luminal A and 48 luminal B paired breast cancer/adjacent tissues. Using all-subset selection method, we identified the most subtype predictive methylation profiles in multivariable logistic regression analysis.

Results: The approach efficiently recognized 15 individual CpG loci differentially methylated in breast cancer tumor subtypes. We further identify novel subtype specific epigenotypes which clearly demonstrate the differences in the methylation profiles of basal-like and human epidermal growth factor 2 (HER2)-overexpressing tumors.

Conclusions: Our results provide evidence that well defined DNA methylation profiles enables breast cancer subtype prediction and support the utilization of this biomarker for prognostication and therapeutic stratification of patients with breast cancer.

INTRODUCTION

Breast cancer is a heterogeneous disease, both biologically and clinically. The molecular background behind breast cancer progression is not well understood, but is associated with the accumulation of genetic aberrations, leading to widespread gene expression changes in breast tumor cells. Consistent with this is the presence of at least four major breast cancer subtypes with distinct expression patterns and clinical outcomes. These subtypes are termed basal-like, ERBB2+, luminal B and luminal A [1,2]. Basal-like and ERBB2+ subtypes are hormone receptor negative and have poor prognosis. In contrast, luminal breast cancers are characterized by the expression of ER-associated genes, with luminal B tumors having poorer outcomes than luminal A tumors. Although this gene expression based approach has proven to add significant prognostic and predictive value to pathologic staging, histological grade and standard clinical molecular markers [3], the high cost of expression profiling and the molecular instability of the mRNA transcripts has limited its incorporation in clinical settings, and therefore, expression based breast cancer classification has not become a standardized method in the general practice [4]. Thus, although breast cancer stratification by gene expression is still considered the gold standard, there is an urgent need for well-defined biomarker panels allowing breast cancer subtyping in clinical diagnostics.

Epigenetic alterations such as aberrations in DNA methylation, microRNAs patterns and post-translational modifications of histones are common molecular abnormalities in cancer [5]. Furthermore, many studies suggest that epigenetic changes are involved in the earliest phases of tumorigenesis, and that they may predispose stem/progenitor cells to subsequent genetic and epigenetic changes involved in tumor promotion [6]. Cancer-related disruption of the DNA methylome involves global genomic hypomethylation and regional hypermethylation of cytosine-phosphate-guanine (CpG) islands. The first can lead chromosomal instability [7], while the second is frequently associated with promoters of tumor suppressor genes resulting in their transcriptional silencing. Both such alterations have been frequently observed in breast cancer [8-19]. Analogously to transcriptomic profiling, DNA methylation profiling is considered to enable the molecular classification of human malignancies and monitoring cancer progression based on a tumor-specific methylation signature [20]. At the same time it facilitates biomarker discovery for the clinical implementation of this process.

Previous epigenetic analyses have identified aberrant DNA methylation signatures associated with molecular subtypes of breast cancer via hormone receptor and human epidermal growth factor 2 (HER2) status [21-23]; however, to date there is very limited information on global methylation changes associated to each molecular subtype, as previous studies focused on individual candidate tumor suppressor genes using locus specific methods. Here we have applied an array-based method [24] for comprehensive DNA methylation profiling to identify differentially methylated genes in breast cancer molecular subtypes. This approach efficiently recognized 15 differentially DNA methylated loci, which were further validated through pyrosequencing in an independent cohort encompassing 47 basal-like, 44 ERBB2+ overexpressing, 48 luminal A and 48 luminal B paired breast cancer/adjacent tissues. Our results provide strong evidence for the existence of tumor subtype specific aberrant methylation profiles, which might be inductors of some transcriptional changes taking place in breast cancer molecular subtypes.

MATERIALS AND METHODS

Patients and tumor characteristics

Samples and associated clinicopathological data were obtained from the Anatomy Pathological Services of the Txagorritxu Hospital, Oncologic Institute of Donostia and Donostia Hospital (Basque Country). Samples of breast tumor and corresponding adjacent normal-appearing tissue (located at least 2 cm away from the site at which the tumor was sampled) were collected from 215 patients diagnosed with a ductal infiltrative breast carcinoma.

DNA-methylation measurements were performed on DNA isolated from paraffin embedded primary breast cancer. All breast specimens were reviewed by experienced pathologists. The inclusion criteria were the availability of the paraffin embedded tissue, tumor size between 1 and 3 cm, histological grade between 1 and 3, and estrogen receptor (ER) -, progesterone receptor (PR) -, HER2-, CK5/6+ or CK14+ or EFGR+ for the basal-like tumors, ER+, PR+/- and HER2+ for the Luminal B, ER-, PR- and HER2+ for the ERBB2+ tumor group and ER+, PR+ and HER2- for the Luminal A. Additional data such as Ki-67 status, p53 mutation and nodal involvement

were also registered. Ethical approval for the study was obtained from corresponding ethics committees of the institutions involved.

Macrodissection, DNA extraction and bisulphite modification.

To minimize contamination in the methylation analysis, we isolated breast cancer cells and paired normal breast epithelial cells from tissues by manual macrodissection. In brief, 10 µm sections were cut from each archival Formalin fixed Paraffin embedded (FFPE) tissue block. For each pair of tissues, the presence of tumor cells in malignant tissues and the absence of cancer cells in normal tissues were confirmed by histopathologic examination.

A total of 500 µl of buffer TE pH 9 was added to each sample and heated at 100°C for 20 min using a water bath. After heating, a cooling step of 5 min was allowed before adding 20 µl of proteinase K. Samples were incubated at 56°C overnight until all the tissue fragments were completely dissolved. Subsequent extraction and purification procedures were performed following the next steps: addition of 500 µl phenol:chloroform:isopropanol alcohol (25:24:1) to the digested tissue, followed by mixing during 10 minutes and centrifugation at 12,000 rpm for 10 minutes. The supernatant fluid was removed to an autoclaved microtube and one volume of chloroform: isopropanol alcohol (25:1) was added, mixed by vortexing, and centrifuged at 12,000 rpm for 10 minutes. The upper aqueous supernatant was purified using a DNA purification kit following the manufacturer's instructions (Qiagen, Valencia, Spain) and the final yield of DNA was dissolved in 50 µl of buffer TE. Sodium bisulphite modification of 1.5 µg DNA was done with the EZ DNA methylation kit (Zymo Research, Orange, CA, USA) following the manufacturer's protocol.

Marker discovery study

Illumina GoldenGate Methylation Cancer Panel 1

Illumina GoldenGate Methylation Cancer Panel was used to analyze 550 ng of starting bisulphite-modified genomic DNA. Methylation was represented as a continuous value from 0 (completely unmethylated) to 1 (completely methylated). This value is calculated by subtracting background hybridization levels obtained from

negative control probes on the array and calculating the ratio of the fluorescent signal from the methylated allele (M) to the sum of the fluorescent signals from both unmethylated (U) and methylated alleles ($|U|+|M|+100$).

Differential methylation analysis

Microarray data were analyzed to identify the most significant tumor subtype-specific changes relative to the adjacent tissue. Differential methylation was assessed by comparing the mean methylation level (b-value) of samples with the mean b-value of the corresponding adjacent tissue using BeadStudio (San Diego, CA, USA) and Qlucore Omics Explorer 2.0 (Qlucore AB, Lund, Sweden) softwares. Selection of the most significantly differentially methylated loci in each tumor subtype was based on: a) Δb value difference of at least 0.20 between the tumors and reference group, b) a FRD-corrected p value cut off of $p < 0.001$ as determined by two-tailed t-test [25] and c) a P-value below 0.05 when comparing mean methylation values among the studied tumor subtypes by using Bonferroni corrected Analysis of Variance (ANOVA). The methylation data have been deposited in NCBI's Gene Expression Omnibus (GEO) [26] and are accessible through GEO Series accession number [GEO:GSE22135].

Validation of the data by pyrosequencing

PCR and pyrosequencing reactions

Selected markers from microarray data analysis were further validated in a larger sample size by bisulphite/pyrosequencing. Additionally, four candidates from the literature were included; *LINE-1*, in order to measure the global hypomethylation of the tumors [27], and *Let-7a*, *Mir-10a* and *Mir-93* microRNAs, since they have been reported to be differently expressed in breast cancer molecular subtypes [28] and their expression might be regulated by methylation in some tumor subtypes, and thus provide new potential subtype specific biomarkers.

For the methylation analysis, 1.5 μg of genomic DNA was treated with sodium bisulphite, using the EZ DNA methylation KitTM (Zymo Research, USA) according to the manufacturer's protocol. All primers were designed using the Assay Design Software (Biotage, Uppsala, Sweden) and synthesized by MWG (Germany). PCR

amplifications were performed using Qiagen HotStarTaq Master Mix Kit (Qiagen, Valencia, Spain), 7.5 μ M biotinylated primer, 15 μ M nonbiotinylated primer and 2 μ l of bisulphite-treated DNA (60 ng). PCR primer sequences, PCR conditions and sequencing primer sequences are given in Table S1 in Additional file 1. The quality and quantity of the PCR product was confirmed by agarose gel (2%) electrophoresis before the cleanup and Pyrosequencing analysis. Pyrosequencing was carried out using the SQA kit (Biotage, Uppsala, Sweden) on a PSQ 96MA Pyrosequencer (Biotage, Uppsala, Sweden) and the methylation index was calculated using the Pyro Q CpG software (Biotage, Uppsala, Sweden).

Validation Data Analysis

Methylation status in tumor vs. adjacent tissue

Methylation status was assessed at the studied markers as previously described by Feng et al. [21]. Taking advantage of paired normal/tumor samples, normal tissues' value was considered as reference. If using the pooled normal samples' mean plus two times the standard deviation as a cut-off point (minimum 10%), we estimated that the probability of the methylation level for a normal appearing tissue to be lower than the cut-off point is less than 96%. Thus, it is reasonable to assume that samples with a methylation value larger than the cut-off point are likely to be abnormal (or positive). Paired t-test was used to determine whether there was a statistically significant change in the methylation of the markers examined between the tumors and adjacent tissues. Additionally, in order to enable the assessment of the observed methylation at multiple promoters as a continuous variable, Z score analysis was used [29,30]. A Z score for each gene was calculated using the given formula: (mean of CpG methylation density of the assessed promoter for each sample - mean of methylation density for the tumor panel) / SD of methylation density. A mean Z score was calculated integrating the promoter specific Z scores and used as a simple score characterizing mean methylation density. In this analysis, a Z score greater than 0 means methylation greater than the population mean.

Comparison between Methylation Status, Breast Cancer molecular Subtypes and Clinicopathological Characteristic

One-sample Kolmogorov-Smirnov test was used to evaluate fitness to normal distribution of continuous parameters. Differences in promoter methylation among tumor subtypes were analyzed by ANOVA or Kruskal Wallis tests as appropriate. The Wilcoxon signed-rank test was used to compare methylation in paired samples. If differences between two independent groups or clinicopathologic characteristics had to be considered, t test (Student t test) or nonparametric test (Mann–Whitney U test) was used. Comparisons of categorical variables were made using Fisher's exact and Pearson's chi-square tests. All reported p values are two-tailed and considered statistically significant if $p < 0.05$.

Subtype classification

Multivariable logistic Regression (MLR) analysis was performed on those biomarkers showing significance in univariate analysis to identify potential biomarker panels capable of discriminating breast cancer subtypes with the best sensitivity and specificity. Models including all possible combinations were constructed and tested by Mallows' Cp selection criterion. False discovery rate (FDR) of classifying breast cancer subtypes was determined in the best models, and selected those significant at $FDR < 0.2$ level.

Supervised hierarchical clustering based on genes selected in the models was performed by using ANOVA test (Benjamini-Hochberg FDR multiple testing corrected [25]) to confirm results obtained by MLR (Qlucore Omics Explorer 2.0 (Qlucore AB, Lund, Sweden)). DNA methylation profiles were standardized to have mean zero and standard deviation one, and clustering was performed using Euclidean method and average linkage.

Ethical considerations

The present study involved analysis of DNA from archival tissue with no subject intervention. No identities were linked to subject records. This study was approved by the Txagorritxu Hospital Review Board under the category of exempt status, and no consent form was required from the participants.

RESULTS

Breast cancer subtypes display distinct clinicopathologic characteristics

Patient characteristics are presented in Table 1. Basal-like and ERBB2+ subtypes were more aggressive tumors than luminal A and B ones, as indicated by their histological and Ki-67 grades ($p < 0.01$), but the basal-like subtype showed lower nodal involvement when compared with other subtypes ($p < 0.01$). Additionally, p53 mutation was significantly less frequent in the luminal A subgroup ($p < 0.05$).

Array-based DNA methylation profiling approach identifies some breast cancer subtype specific methylation changes

DNA methylation profiling analysis at 1537 CpG loci was performed in 30 paired breast cancer/adjacent tissues representing the four major breast cancer subtypes. Two of the tumor samples were eliminated from further analysis secondary to their low methylation levels. The final cohort consisted on 60 samples- 28 paired breast cancer/adjacent tissues and four samples from the peritumoral region. Extensive DNA methylation changes between tumor and adjacent tissues were revealed following supervised clustering of the samples (Figure 1, Table S2 in Additional file 2). In order to identify CpGs with highly significant tumor subtype-specific changes, stringent cut-offs ($p < 0.001$ and $\Delta b > 0.20$) were set for methylation level changes relative to controls for each tumor subtype. Distinct CpG loci groups were identified showing differential methylation profiles in tumor subtypes. We have focused in those genes that were significantly differently methylated in the basal-like tumor subtype (*HOXA9*, *SOX1*, *VAMP8* and *TNFRSF10D*) and those that were significantly hypermethylated in the luminal B tumors (*NPY*, *RASSF1*, *HS3ST2*, *DBC1*, *FGF2*, *CD40*, *SPARC*, *JAK3*, *SOX17*, *PRKCDBP* and *TAL1*). Additionally, we compared mean methylation values among the studied tumor subtypes by using Bonferroni corrected ANOVA and observed that *NPY*, *PRKCDBP* and *RASSF1* showed lower methylation levels in the basal-like tumor subgroup when compared with other subtypes, and that *DBC1*, *HS3ST2*, *FGF2*

and *CD40* displayed higher hypermethylation levels in the luminal B subtype than in the other subgroups.

The microarray was technically validated by means of pyrosequencing. Comparison of quantitative methylation values at 5 CpG loci by both GoldenGate array and pyrosequencing in 20 samples confirmed the accuracy of the array-based measurements (mean $r^2=0.75$, range 0.56-0.91, Table S3 in Additional file 3).

Validation of differential methylation events in a large panel of breast cancer tissue samples

For the validation study, pyrosequencing assays for 19 CpG loci were conducted in an independent cohort of 187 normal/breast cancer paired samples encompassing 47 basal-like, 44 ERBB2+, 48 luminal A and 48 luminal B. Differential methylation between tumor and adjacent tissue was confirmed in the validation panel (Figure 2). Among the loci displaying >95% specificity (i.e. very low DNA methylation frequency in normal tissue), the frequency of hypermethylation in tumor tissue ranged from 28% (*DBC1*, deleted in bladder cancer 1) to 75% (*JAK3*, Janus kinase 3), (see Table S4 in Additional file 4). Besides, *VAMP8* (Vesicle-associated membrane protein 8) was hypomethylated in 82 % of the tumors and *LINE-1*, the global methylation marker, was highly methylated in all normal tissues (mean=68% \pm 1.9%) while a statistically significant drop was observed in tumors (mean=63% \pm 5.4%, $p < 0.001$).

To further define those CpG loci differentially methylated in breast cancer subtypes, data were filtered on significance of differences using Bonferroni corrected ANOVA test ($p < 0.05$). We found that 15 out of the 19 genes involved displayed different methylation profiles in analyzed breast tumor subtypes (Table 2). *LINE-1*, *HOXA9*, *TAL1* and *SPARC* were excluded, as they showed no significant difference. During the validation study we confirmed that promoter methylation of *NPY*, *RASSF1*, *TNFRS10D*, *PRKCDBP*, *Let-7a* and *VAMP8* genes was lower in the basal-like subtype when compared with HER2 overexpressing tumors (luminal B and ERBB2+). *SOX1* and *SOX17* showed different methylation levels between luminal B and luminal A tumors. We also observed that *HS3ST2*, *DBC1*, *FGF2*, *CD40*, *JAK3*, *Mir-93* and *Mir-10a* markers displayed higher methylation levels in luminal B and ERBB2+ subtypes than in the basal-like and the luminal A tumors. Consequently, the validation study

confirmed that these markers were suitable for development of tumor subtype methylation profiles. One of the most interesting aspects of the validation study was to observe that while the microarray analysis showed no specific methylation profile for HER2 overexpressing tumors, the validation study (carried out in a larger sample size) demonstrated that HER2 overexpressing tumors (luminal B and ERBB2+ tumors) had a common methylation profile.

The analysis of DNA methylation profiles with regard to the tumors' clinicopathological features revealed that hypermethylation of *HS3ST2* was more prevalent in patients with nodal involvement ($\chi^2=10.61$, $p < 0.001$), while *RASSF1* hypermethylation was associated with the presence of p53 mutations ($\chi^2=7.51$, $p < 0.001$) and high ki-67 index ($\chi^2=15.63$, $p < 0.001$). In addition, promoter hypermethylation of *TALI* and *SPARC* genes was more frequent in tumors with high histological grade ($\chi^2=6.41$ and $\chi^2=6.21$ respectively, $p < 0.05$).

We next combined all the methylation data into a single variable, the mean Z score, to allow further analysis. As shown in the Figure 2c, mean methylation density in tumors was significantly higher than in adjacent tissue (-1.26 [CI, -1.26 to -1.17] and -0.01 [CI, -0.81 to 0.72], $p < 0.001$). Additionally, we found significant differences in the mean Z scores among the different breast cancer subtypes (basal-like -0.45 [-0.58 to -0.32], luminal B 0.24 [0.09 to 0.39], ERBB2+ 0.22 [0.05 to 0.41] and luminal A -0.02[-0.14 to 0.09], $p < 0.001$) (Figure 3c) and that significantly higher Z score epigenotypes were related to *LINE-1* hypomethylated tumors (0.27 [CI, 0.04 to 0.28] and -0.06 [CI, -0.13 to 0.02] and, $p < 0.05$), HER2-overexpressing breast cancers (0.25 [CI, 0.15 to 0.37] and -0.22 [CI, -0.33 to -0.22], $p < 0.001$) and node-positive tumors (0.11 [CI, -0.01 to 0.19] and -0.10 [CI, -0.25 to 0.4]), which are features associated with clinically more aggressive tumors. Paradoxically, highly proliferative and p53 mutated tumors, both associated to the basal-like phenotype, showed significantly lower Z scores of methylation (-0.17 [CI, -0.35 to 0.01] and 0.11 [CI, -0.03 to 0.26], $p < 0.001$) and (-0.23 [CI, -0.44 to -0.11] and 0.12 [CI, -0.02 to 0.26], $p < 0.001$) respectively.

Novel DNA methylation profiles with subtype predictive value

Selected loci were further processed to identify the best biomarker panel allowing classification in tumor subtypes. Multivariable Logistic Regression (MLR) analysis was performed to identify potential biomarker panels with the best sensitivity, specificity and false discovery rate (FDR). We first pruned all the variables showing significance in univariate analysis (Table 3) using all-subset selection with Mallows' Cp optimization criteria, and recognized biomarker panels capable of discriminating basal-like and HER2 overexpressing subtypes at FDR 0.2 level. Specifically, basal-like tumors were inversely associated to *NPY*, *FGF2*, *HS3ST2*, *RASSF1* and *Let-7a* markers hypermethylation (sensitivity 75%, specificity 95 % and FDR: 0.17) whereas HER2 overexpressing subtypes were strongly correlated to *NPY*, *FGF2*, *HS3ST2*, *RASSF1* and *Let-7a* hypermethylation (sensitivity 74%, specificity 80 %, and FDR: 0.18).

The supervised hierarchical clustering analysis based on these 5 genes, resulted in three different clusters of tumor groups (C1 to C3) (Figure 3). Samples fitting in cluster C1 (characterized by lack of methylation) were mostly basal-like tumors; cluster C2 was the most heterogeneous one with regard to both methylation status of the genes and samples fitting in it. This cluster contains the majority of the luminal A samples, but also includes some Luminal B and ERBB2+ tumors. On the other hand, samples in cluster C3 (hypermethylated in almost all the analyzed markers) was basically composed of luminal B and ERBB2+ samples (HER2 overexpressing subtypes).

DISCUSSION

Identification of gene expression based breast cancer subtypes is considered as a critical means of prognostication and further more as an important predictive marker for the response to treatment with endocrine therapy. However, analytical tests relying upon RNA measures are difficult to standardize and implement due to the instability of the mRNA transcripts. DNA methylation profiles reflect phenotypically important differences in gene transcription and in contrast to most mRNAs, it has a very stable structure, making DNA methylation profiles based diagnostic tests highly accurate and reproducible [31]. By use of two independent cohorts of invasive breast carcinomas, our study is, to our knowledge, one of the first to deliver specific methylation profiles

associated to basal-like, ERBB2+, luminal A and luminal B molecular subtypes of breast cancer.

Statistical analysis of the methylation microarray data revealed extensive DNA methylation changes between tumor and adjacent tissue (Figure 1, Table S2 in Additional file 2), and concurrence of some of these DNA methylation changes with particular breast cancer molecular subtypes and tumor morphologies. Applying stringent criteria, we identified significant differential methylation among the tumor subtypes in 15 of the 1537 screened CpG islands. Subsequent validation of the selected genes plus four more from literature by pyrosequencing-in an independent cohort of 187 breast tumor/normal pairs, confirmed that genes tested have significantly altered methylation profiles in comparison to normal appearing adjacent tissue, and that most of the candidate genes (15/19) displayed significantly different methylation profiles between different tumor subtypes. These results corroborate the potential usefulness of the studied markers for the development of a methylation marker panel defining tumor subtypes. Moreover, the current study represents the first evidence for DNA hypermethylation of *NPY*, *FGF2*, *CD40*, *TAL1*, *JAK3*, *SPARC*, *PRKCDBP*, *DBC1*, *SOX1*, *TNFRS10D*, *Let-7a*, *Mir-10a* and *Mir-93* and hypomethylation of *VAMP8* in breast cancer.

Several genes have been previously reported to be aberrantly methylated in breast cancer (reviewed in [32-34]). Furthermore, methylation in breast cancer has already been connected to breast cancer molecular subtypes, but these observations need further confirmation. As previously suggested, we have found that basal-like, ERBB2+, luminal A and luminal B molecular subtypes displayed specific methylation profiles. Specifically, HER2-enriched breast tumors (ERBB2+ and luminal B) were associated with the hypermethylation of several genes related to cancer development (Table 3). These results are in accordance with previous studies stating that HER2/neu breast cancers were associated with preferential hypermethylation of several genes [21-23]. In addition, Terada et al. [35] have recently found that frequent CpG islands methylation is highly associated with HER2 amplification. Conversely, we have observed that basal-like tumors were inversely related to promoter methylation of many of the studied genes (Table 3) and showed the lowest methylation levels among the studied subtypes as indicated by the mean Z score values (Figure 3c). These findings are consistent with the

recent observations made by Holm et al. [36] who have reported that basal-like tumors present low methylation levels of several CpG sites while luminal B tumors display high methylation levels.

On the other hand, our efforts to identify and validate breast cancer subtype-specific epigenotypes, resulted in a significant model based on five biomarkers, which is capable of discriminating basal-like and HER2 overexpressing subtypes. Basal-like tumors showed lack of methylation at *NPY*, *FGF2*, *HS3ST2*, *RASSF1* and *Let-7a* markers while HER2 overexpressing tumors (luminal B and ERBB2+ subtypes) were related to hypermethylation of these markers.

Several authors have speculated that these genomically defined tumor subtypes may represent transformation of stem cells. Some of these hypotheses suggest that mammary stem cell progresses to a luminal progenitor state (with a similar expression pattern to that identified in basal breast cancer), which then progresses to differentiated cells with more luminal characteristics [37]. As can be deduced, this hypothesis locate basal-like tumors in an intermediate differentiation step between the mammary stem cells and the more differentiated luminal subtypes, which would explain their poor prognosis despite responding to chemotherapy [38]. Many studies have been conducted to define methylation profiles in human stem cells [39-42]. Likewise, Calvanesse et al. [43] compared the methylation status in human embryonic stem cells (hESC), cancer cell lines and normal human primary tissues using the same platform as in our study. Their finding was consistent with the view that genes aberrantly hypermethylated in cancer (i.e., not hypermethylated in normal tissues) were not hypermethylated in hESC, termed by these authors as classical class A cancer hypermethylated genes. Interestingly, all the CpG sites related to basal-like tumors in univariate regression analysis with the exception of *RASSF1* (Table 3) belonged to this committed Class A category, and in a manner analogous to hESC; they were not hypermethylated in basal-like tumors, nor in normal breast, but they were hypermethylated in the rest of breast cancer subtypes, suggesting that basal-like tumors may share some similarities in their methylation pattern with that of the hESC cell lines. These shared methylation signatures may reinforce the hypothesis of basal-like tumors arising from a mammary stem cell and progressing to differentiated cells with more luminal characteristics (HER2 enriched, luminal A and B). In fact, a recent study by Holm et al. [36] has also

supported this hypothesis since they have observed that basal-like tumors seem to arise from luminal progenitors in which genes initiating a differentiated luminal cell fate are repressed by other mechanisms than promoter methylation such as the Polycomb repressive complex 2 (PRC2).

Finally, epigenetic therapy, including the use of demethylating agents and histone deacetylase inhibitors is now in clinical trials for myelodysplastic syndrome, leukemia and ovarian and lung cancers [21]. Recent studies have suggested that co-treatment of DNA methylation inhibitors and histone deacetylases might be an effective form of epigenetic therapy for breast cancer, as the interplay observed between DNA methylation and histone modifications can result in synergistic induction of tumor suppressor genes [21,44]. Thus, there is a possibility that epigenetic therapy could play an important role in the immediate future of breast cancer treatment. The information of subtype-specific methylation profiles described in the present study, might promote a better understanding of the epigenetic regulation mechanisms in breast cancer, thereby contributing to the improvement of epigenetic therapy.

CONCLUSIONS

In this study we utilized a microarray approach followed by substantial technical and biological validation to identify specific epigenotypes for particular subtypes in breast cancer. We clearly demonstrate the differences in the methylation profiles of basal-like and HER2-overexpressing tumors and provide evidence to support the utilization of this biomarker for prognostication and therapeutic stratification of patients with breast cancer.

Abbreviations

ANOVA: analysis of variance; CpG: cytosine-phosphate-guanine; ER: estrogen receptor; ERBB2+: ER-, PR- and HER2+ tumor subtype; FFPE: formalin fixed paraffin embedded; FDR: false discovery rate; GEO: Gene Expression Omnibus (GEO); hESC:

human embryonic stem cells; HER2: human epidermal growth factor 2; MLR: multivariable logistic regression; PR: progesterone receptor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NGB, TL, AFF and AA carried out the methylation assays and helped prepare the manuscript. IG, AV, RR and IR provided tissues and clinical data. MCA and MJA carried out HER2, CH5/6, CK14 and EGFR1 assays and provided pathologic supports for tissue microdissection. DM and JD performed the statistical analysis. MMP, TL, MFF, MR and JKF conceived the study and prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from project CGL2008-01131 (Departamento de Sanidad del Gobierno Vasco), S-PE08UN45 and PE09BF02 (Departamento de Ciencia y Tecnología del Gobierno Vasco), BIO2008-04212 and RD06/0020/1019 (Red Temática de Investigación Cooperativa en Cáncer, RTICC) from the MICINN. The CIBER de Enfermedades Raras is an initiative of the ISCIII. N.G.B. had a doctoral fellowship from the Basque Government (Departamento de Educación, Universidades e Investigación).

Reference List

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**: 747-752.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci U S A* 2001, **98**: 10869-10874.

3. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**: 1160-1167.
4. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO: **Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer.** *J Natl Cancer Inst* 2009, **101**: 736-750.
5. Mulero-Navarro S, Esteller M: **Epigenetic biomarkers for human cancer: the time is now.** *Crit Rev Oncol Hematol* 2008, **68**: 1-11.
6. Feinberg AP, Ohlsson R, Henikoff S: **The epigenetic progenitor origin of human cancer.** *Nat Rev Genet* 2006, **7**: 21-33.
7. Fraga MF, Esteller M: **Towards the human cancer epigenome: a first draft of histone modifications.** *Cell Cycle* 2005, **4**: 1377-1381.
8. Yuan J, Luo RZ, Fujii S, Wang L, Hu W, Andreeff M, Pan Y, Kadota M, Oshimura M, Sahin AA, Issa JP, Bast RC Jr, Yu Y: **Aberrant methylation and silencing of ARHI, an imprinted tumor suppressor gene in which the function is lost in breast cancers.** *Cancer Res* 2003, **63**: 4174-4180.
9. Fujii S, Luo RZ, Yuan J, Kadota M, Oshimura M, Dent SR, Kondo Y, Issa JP, Bast RC Jr, Yu Y: **Reactivation of the silenced and imprinted alleles of ARHI is associated with increased histone H3 acetylation and decreased histone H3 lysine 9 methylation.** *Hum Mol Genet* 2003, **12**: 1791-1800.
10. Dammann R, Yang G, Pfeifer GP: **Hypermethylation of the cpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers.** *Cancer Res* 2001, **61**: 3105-3109.
11. Krop IE, Sgroi D, Porter DA, Lunetta KL, LeVangie R, Seth P, Kaelin CM, Rhei E, Bosenberg M, Schnitt S, Marks JR, Pagon Z, Belina D, Razumovic J, Polyak K: **HIN-1, a putative cytokine highly expressed in normal but not cancerous mammary epithelial cells.** *Proc Natl Acad Sci U S A* 2001, **98**: 9796-9801.
12. Fackler MJ, McVeigh M, Evron E, Garrett E, Mehrotra J, Polyak K, Sukumar S, Argani P: **DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma.** *Int J Cancer* 2003, **107**: 970-975.
13. Murata H, Khattar NH, Kang Y, Gu L, Li GM: **Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability.** *Oncogene* 2002, **21**: 5696-5703.
14. Ferguson AT, Evron E, Umbricht CB, Pandita TK, Chan TA, Hermeking H, Marks JR, Lambers AR, Futreal PA, Stampfer MR, Sukumar S: **High frequency**

- of hypermethylation at the 14-3-3 sigma locus leads to gene silencing in breast cancer.** *Proc Natl Acad Sci U S A* 2000, **97**: 6049-6054.
15. Du Y, Carling T, Fang W, Piao Z, Sheu JC, Huang S: **Hypermethylation in human cancers of the RIZ1 tumor suppressor gene, a member of a histone/protein methyltransferase superfamily.** *Cancer Res* 2001, **61**: 8094-8099.
 16. Di Vinci A, Perdelli L, Banelli B, Salvi S, Casciano I, Gelvi I, Allemanni G, Margallo E, Gatteschi B, Romani M: **p16(INK4a) promoter methylation and protein expression in breast fibroadenoma and carcinoma.** *Int J Cancer* 2005, **114**: 414-421.
 17. Hu XC, Loo WT, Chow LW: **E-cadherin promoter methylation can regulate its expression in invasive ductal breast cancer tissue in Chinese woman.** *Life Sci* 2002, **71**: 1397-1404.
 18. Khan S, Kumagai T, Vora J, Bose N, Sehgal I, Koeffler PH, Bose S: **PTEN promoter is methylated in a proportion of invasive breast cancers.** *Int J Cancer* 2004, **112**: 407-410.
 19. Wei M, Grushko TA, Dignam J, Hagos F, Nanda R, Sveen L, Xu J, Fackenthal J, Tretiakova M, Das S, Olopade OI: **BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy.** *Cancer Res* 2005, **65**: 10692-10699.
 20. Orlando FA, Brown KD: **Unraveling breast cancer heterogeneity through transcriptomic and epigenomic analysis.** *Ann Surg Oncol* 2009, **16**: 2270-2279.
 21. Feng W, Shen L, Wen S, Rosen DG, Jelinek J, Hu X, Huan S, Huang M, Liu J, Sahin AA, Hunt KK, Bast RC Jr, Shen Y, Issa JP, Yu Y: **Correlation between CpG methylation profiles and hormone receptor status in breast cancers.** *Breast Cancer Res* 2007, **9**: R57.
 22. Fiegl H, Millinger S, Goebel G, Muller-Holzner E, Marth C, Laird PW, Widschwendter M: **Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer.** *Cancer Res* 2006, **66**: 29-33.
 23. Bertolo C, Guerrero D, Vicente F, Cordoba A, Esteller M, Ropero S, Guillen-Grima F, Martinez-Peñuela JM, Lera JM: **Differences and molecular immunohistochemical parameters in the subtypes of infiltrating ductal breast cancer.** *Am J Clin Pathol* 2008, **130**: 414-424.
 24. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, Goldmann T, Seifart C, Jiang W, Barker DL, Chee MS, Floros J, Fan JB: **High-throughput DNA methylation profiling using universal bead arrays.** *Genome Res* 2006, **16**: 383-393.

25. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: **Controlling the false discovery rate in behavior genetics research.** *Behav Brain Res* 2001, **125**: 279-284.
26. **Gene Expression Omnibus, GEO.** [<http://www.ncbi.nlm.nih.gov/geo/>]
27. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T: **Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer.** *Int J Cancer* 2009, **124**: 81-87.
28. Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavaré S, Caldas C, Miska EA: **MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype.** *Genome Biol* 2007, **8**: R214.
29. Geli J, Kiss N, Karimi M, Lee JJ, Backdahl M, Ekstrom TJ, Larsson C: **Global and regional CpG methylation in pheochromocytomas and abdominal paragangliomas: association to malignant behavior.** *Clin Cancer Res* 2008, **14**: 2551-2559.
30. Konishi K, Shen L, Wang S, Meltzer SJ, Harpaz N, Issa JP: **Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias.** *Gastroenterology* 2007, **132**: 1254-1260.
31. Christensen BC, Marsit CJ, Houseman EA, Godleski JJ, Longacker JL, Zheng S, Yeh RF, Wrensch MR, Wiemels JL, Karagas MR, Bueno R, Sugarbaker DJ, Nelson HH, Wiencke JK, Kelsey KT: **Differentiation of lung adenocarcinoma, pleural mesothelioma, and nonmalignant pulmonary tissues using DNA methylation profiles.** *Cancer Res* 2009, **69**: 6315-6321.
32. Hinshelwood RA, Clark SJ: **Breast cancer epigenetics: normal human mammary epithelial cells as a model system.** *J Mol Med* 2008, **86**: 1315-1328.
33. Lo PK, Sukumar S: **Epigenomics and breast cancer.** *Pharmacogenomics* 2008, **9**: 1879-1902.
34. Pogribny IP, Beland FA: **DNA hypomethylation in the origin and pathogenesis of human diseases.** *Cell Mol Life Sci* 2009, **66**: 2249-2261.
35. Terada K, Okochi-Takada E, Akashi-Tanaka S, Miyamoto K, Taniyama K, Tsuda H, Asada K, Kaminishi M, Ushijima T: **Association between frequent CpG island methylation and HER2 amplification in human breast cancers.** *Carcinogenesis* 2009, **30**: 466-471.
36. Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, Borg A, Ringnér M: **Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns.** *Breast Cancer Res* 2010, **12**: R36.
37. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ; kConFab, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE,

- Lindeman GJ: **Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15**: 907-913.
38. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes.** *Clin Cancer Res* 2007, **13**: 2329-2334.
 39. Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE, Gonzalez R, Crook J, Davidson B, Schulz TC, Robins A, Khanna A, Sartipy P, Hyllner J, Vanguri P, Savant-Bhonsale S, Smith AK, Chakravarti A, Maitra A, Rao M, Barker DL, Loring JF, Fan JB: **Human embryonic stem cells have a unique epigenetic signature.** *Genome Res* 2006, **16**: 1075-1083.
 40. Bibikova M, Laurent LC, Ren B, Loring JF, Fan JB: **Unraveling epigenetic regulation in embryonic stem cells.** *Cell Stem Cell* 2008, **2**: 123-134.
 41. Bloushtain-Qimron N, Yao J, Shipitsin M, Maruyama R, Polyak K: **Epigenetic patterns of embryonic and adult stem cells.** *Cell Cycle* 2009, **8**: 809-817.
 42. Altun G, Loring JF, Laurent LC: **DNA Methylation in Embryonic Stem Cells.** *Journal of Cellular Biochemistry* 2010, **109**: 1-6.
 43. Calvanese V, Horrillo A, Hmadcha A, Suarez-Alvarez B, Fernandez AF, Lara E, Casado S, Menendez P, Bueno C, Garcia-Castro J, Rubio R, Lapunzina P, Alaminos M, Borghese L, Terstegge S, Harrison NJ, Moore HD, Brüstle O, Lopez-Larrea C, Andrews PW, Soria B, Esteller M, Fraga MF: **Cancer genes hypermethylated in human embryonic stem cells.** *PLoS One* 2008, **3**: e3294.
 44. Pathiraja TN, Stearns V, Oesterreich S: **Epigenetic Regulation in Estrogen Receptor Positive Breast Cancer-Role in Treatment Response.** *J Mammary Gland Biol Neoplasia* 2010.

FIGURE LEGENDS

Figure 1. Supervised hierarchical clustering of 28 paired breast cancer samples using methylation microarray data from differentially methylated genes. Heat map and clustering of 76 significant (FRD-corrected $p < 0.001$) differences between tumors (red) compared to adjacent tissues (green) shows that 30 of the differences (corresponding to 25 genes) are increases of methylation, while 47 (corresponding to 37 genes) are decreases of promoter methylation in tumors. Heat map colors symbolize DNA methylation as indicated in the color key. The full list of genes is presented in Table S2 in Additional file 2.

Figure 2. Supervised hierarchical clustering of 187 paired breast cancer samples using methylation data from selected genes in the validation study. A) Heat map and clustering of significant (FRD-corrected $p < 0.01$) differences between tumors (red) compared to adjacent tissues (green). Heat map colors symbolize DNA methylation as indicated in the color key. B) Principal Component Analysis (PCA) plots showing separation between tumor and adjacent tissue, C) Box plots showing lower methylation density in adjacent tissue compared to tumors.

Figure 3. Supervised Hierarchical clustering of the five candidate genes (Let-7a, NPY, RASFF1, FGF2 and HS3ST2). Differential methylation of patterns is shown by red (hypermethylated) versus green (none methylated) colors. As can be noticed, three differentiated clusters (C1 to C3) are generated by this statistical analysis, C) Box plots showing differential methylation density among tumors subtypes.

Table 1. Patient characteristics of both marker discovery and validation studies

	Type			
	Basal	ERBB2+	Luminal B	Luminal A
Number of patients	59	49	54	53
Age(years)				
<i>Mean</i>	56.1	57.1	56.9	62.9
<i>Median</i>	53	57.5	51	61
<i>Range(95%CI)</i>	51.5-60.7	52.5-61.6	51.8-61.7	56.9-68.7
Tumour size(cm)				
<i>Mean</i>	2.4	2.8	3.2	3.1
<i>Range(95%CI)</i>	2.1-2.7	2.1-3.5	2.5-3.5	2.4-3.7
Histological grade (%)				
<i>I</i>	5.0	0.0	0.0	6.0
<i>II</i>	47.1	24	40.0	55.5
<i>III</i>	48.9	76	60.0	38.5
Nodal involvement (%)				
<i>Positive</i>	35.5	63.4	72.2	57.3
<i>Negative</i>	64.4	36.6	27.8	42.7
Ki-67 overexpression (%)				
<i><10%</i>	0.0	8.9	21.9	27.8
<i>10%-25%</i>	0.0	36.1	31.2	66.7
<i>>25%</i>	100.0	56.0	46.9	5.6
p53 mutation (%)				
<i>Positive</i>	47.2	44.1	21.8	14.8
<i>Negative</i>	52.8	55.9	88.2	85.2

Table 2. Methylation levels in breast cancer subtypes

Gene	Methylation level (mean \pm SD) (Positive rate*)				pvalue**
	Basal like	Luminal B	ERBB2+	Luminal A	
CD40	24.3 \pm 12.3(21.7)	30.8 \pm 11.9(36.2)	30.8 \pm 11.9(33.4)	27.4 \pm 7.9(23.9)	< 0.001
DBC1	12.8 \pm 9.2(15.9)	19.9 \pm 10.3(34.1)	17.8 \pm 8.2(42.7)	14.0 \pm 7.3(19.5)	< 0.001
FGF2	7.4 \pm 7.2(11.1)	23.1 \pm 11.2(63.8)	32.2 \pm 13.9(67.5)	15.0 \pm 13.4(30.8)	< 0.001
HOXA9	21.1 \pm 12.0(61.8)	22.5 \pm 9.4(64.7)	28.4 \pm 10.3(83.8)	20.3 \pm 8.0(78.1)	NS
HS3ST2	10.3 \pm 11.9(31.0)	23.9 \pm 12.9(75.0)	26.8 \pm 12.3(79.8)	17.5 \pm 13.2(40.4)	< 0.001
JAK 3	28.0 \pm 15.3(53.5)	36.8 \pm 15.2(87.5)	33.3 \pm 13.7(74.1)	28.8 \pm 10.3(83.0)	< 0.001
Let-7a	33.9 \pm 11.3(19.1)	55.1 \pm 10.9(72.9)	48.8 \pm 15.7(65.7)	53.5 \pm 8.4(80.0)	< 0.001
LINE-1	63.8 \pm 3.2(30.2)	63.1 \pm 6.2(40.4)	63.8 \pm 7.8(40.2)	62.8 \pm 4.7(30.4)	NS
Mir-10a	15.5 \pm 4.2(36.4)	25.2 \pm 5.8(70.8)	23.1 \pm 10.1(77.4)	20.8 \pm 11.3(63.2)	< 0.001
Mir-93 a	55.6 \pm 7.3(23.9)	60.6 \pm 6.5(57.8)	61.2 \pm 8.9(62.1)	61.3 \pm 6.5(47.5)	< 0.001
NPY	14.9 \pm 4.9(35.6)	38.9 \pm 14.6(95.8)	34.5 \pm 14.5(80.9)	27.4 \pm 14.8(78.6)	< 0.001
PRKCDBP	7.6 \pm 6.1 (4.3)	13.2 \pm 10.0(12.5)	15.6 \pm 10.1(22.3)	10.6 \pm 8.4(15.0)	< 0.001
RASSF1	20.4 \pm 11.9(35.7)	33.9 \pm 12.9(79.2)	35.6 \pm 10.8(74.5)	23.5 \pm 5.9(73.9)	< 0.001
SOX 1	19.8 \pm 10.7(75.0)	24.5 \pm 11.2(81.3)	21.5 \pm 10.1(78.6)	20.7 \pm 10.1(53.3)	< 0.001
SOX17	19.1 \pm 15.6(59.1)	22.4 \pm 14.2(70.8)	23.2 \pm 14.7(62.4)	14.1 \pm 7.2(32.4)	< 0.001
SPARC	35.3 \pm 10.7(48.9)	33.7 \pm 10.2(60.4)	34.5 \pm 12.2(67.3)	31.1 \pm 11.0(68.1)	NS
TAL1	38.8 \pm 12.5(59.1)	33.9 \pm 11.3(56.5)	37.7 \pm 11.6(60.4)	27.4 \pm 7.9(52.3)	NS
TNFRS10D	12.5 \pm 12.6(27.8)	18.9 \pm 8.8(60.9)	23.5 \pm 11.3(65.7)	19.6 \pm 9.2(60.0)	< 0.001
VAMP8	8.8 \pm 3.8(93.1)	12.4 \pm 4.4(72.1)	11.4 \pm 3.9(77.2)	11.9 \pm 4.1(79.2)	< 0.001

*Positive rate using the sample mean plus two times the SD of the pooled normal samples (and minimum 10% methylation) as a cut-off point .

**p value computed using the ANOVA or Kruskal Wallis tests as appropriate.

Table 3. Subtype specific methylation profiles in univariate analysis

<i>Tumor subtype</i>	<i>Related genes*</i>
<i>Basal-like</i>	<i>JAK3 (-), NPY (-), RASSF1 (-), SPARC (-), DBC1 (-), FGF2 (-), HS3ST2 (-), PRKCDBP(-), TNFRS10D (-) and VAMP8 (+)</i>
<i>HER2 overexpressing</i>	<i>NPY (+), DBC1 (+), FGF2 (+), RASSF1 (+), HS3ST2 (+), SOX17(+), and TNFRS10D (+)</i>
<i>Luminal A</i>	<i>SOX1(-), SOX17(-) and VAMP8(+)</i>

*Related genes in univariate logistic regression analysis at 0.05 significance level.

**(+), (-) positively and (-) negatively related.

Additional files

Additional file 1

Table S1. PCR primer and sequencing primer sequences and PCR conditions.

Additional file 2

Table S2. Genes differently methylated that provided optimal distinction between tumor and normal samples ($p < 0.001$ and $\Delta b > 10.20$).

Additional file 3

Table S3. Comparison of quantitative methylation measurements at five CpG loci by both GoldenGate array and pyrosequencing.

Additional file 4

Table S4. Methylation levels in tumors and adjacent tissues.

TUMOR

ADJACENT TISSUE

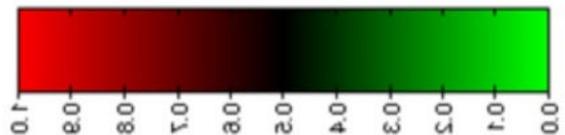
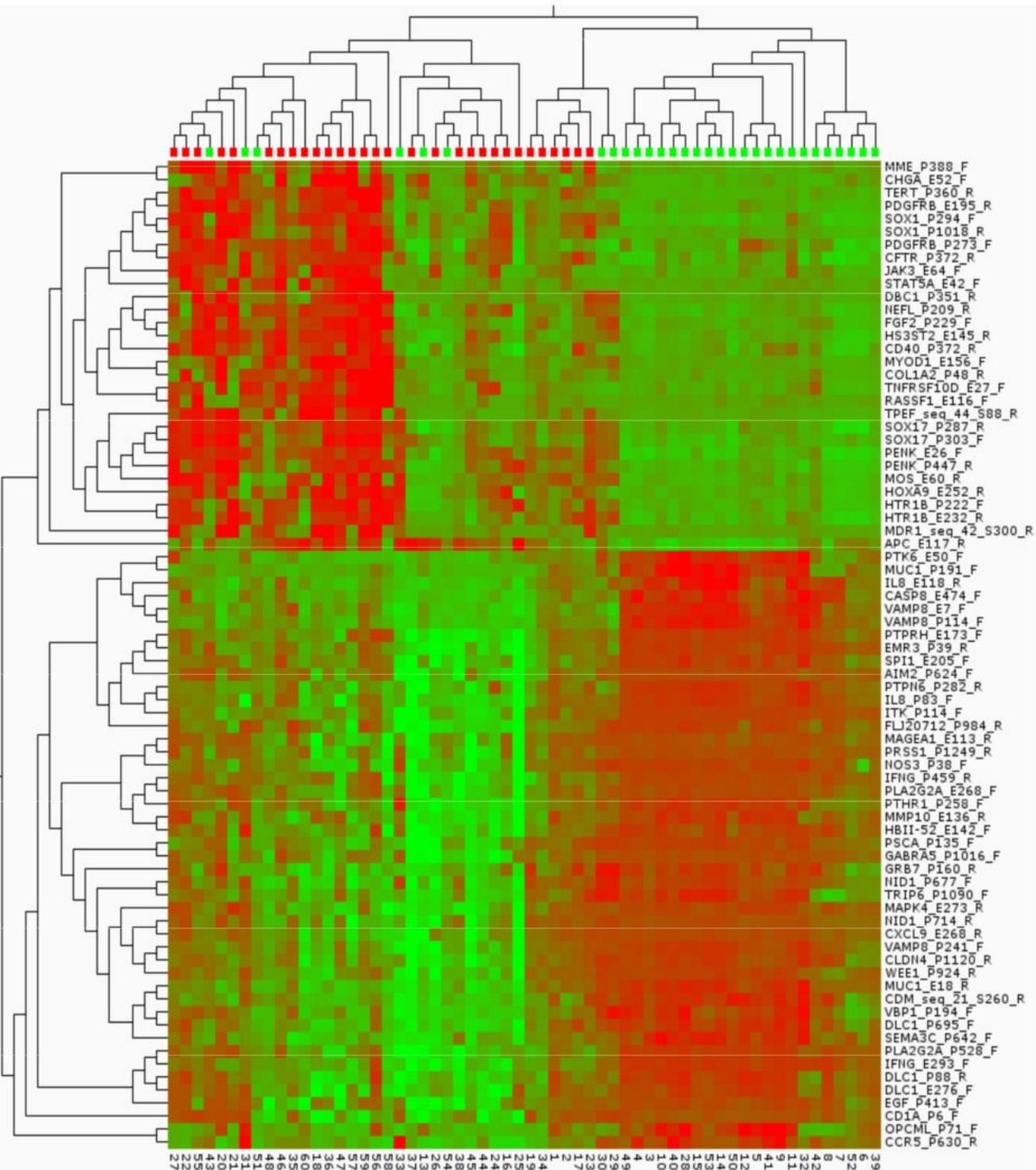


Figure 1

Additional files provided with this submission:

Additional file 1: Supplementary Table S1.xls, 57K

<http://breast-cancer-research.com/imedia/1146158605382968/supp1.xls>

Additional file 2: Supplementary Table S2.xlsx, 18K

<http://breast-cancer-research.com/imedia/1416152468387174/supp2.xlsx>

Additional file 3: Supplementary Table S3.xls, 31K

<http://breast-cancer-research.com/imedia/6211889038296879/supp3.xls>

Additional file 4: Table S4.docx, 14K

<http://breast-cancer-research.com/imedia/1956090025455290/supp4.docx>