Epigenetic and Genomic Biomarker Discovery in Breast Cancer

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

Vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Ramin Radpour

aus Tehran, Iran

Basel, 2011

creative commons

Attribution-Noncommercial-No Derivative Works 2.5 Switzerland

---

You are free:

to Share — to copy, distribute and transmit the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.

- Any of the above conditions can be waived if you get permission from the copyright holder.

- Nothing in this license impairs or restricts the author's moral rights.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en   Datum: 3.4.2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Christoph Dehio

Prof. Dr. Xiao Yan Zhong

Prof. Dr. Giulio C Spagnoli

Basel, den 22.02.2011

Prof. Dr. Martin Spiess, Dekan

Epigenetic and Genomic Biomarker Discovery in Breast Cancer

A thesis for the degree of

Doctor of Philosophy (Ph.D.) in Genetics

University of Basel

By

Ramin Radpour

From Tehran, Iran

This research was conducted at

Laboratory for Prenatal Medicine and Gynecological Oncology

Department of Biomedicine / Women's Hospital

University of Basel

Switzerland

To:
All of the patients who are suffering from cancer

To:
My great parents and my best friend Zeinab

Acknowledgements

Table of Contents

1. Introduction

Breast cancer is the most common malignancy among females and the fifth most common cause of cancer death worldwide which comprises about 10.4% of all cancers [1]. To be detected, a breast tumor should be at least a few millimeters in size. Only 60 percent of breast cancer is diagnosed at a local stage. At the time of primary diagnosis, usually tumor cells have already been shed and metastases had already occurred [2].

Localized breast cancer at an early stage has better prognosis and requires less severe treatment with a survival rate of 98% [3], however, diagnosis after tumor metastasis has lower survival rate to 27% [4]. This highlights the importance of early breast cancer detection which is dependent on sensitive and specific screening methods. The traditional triple test for breast cancer diagnosis includes physical examination, mammography and aspiration cytology. Unfortunately, all these methods are not sensitive enough in identifying breast cancer in early stages [3,5]. A minimally invasive screening test beside the triple test, or prior to biopsy, would lead to greater sensitivity. Many researchers have attempted to establish molecular and immunological methods for the detection of individual metastatic breast cancer cells in peripheral blood [6] but still finding an accurate and sensitive biomarker remains challenging [7].

The increasing knowledge about the various forms of cancer, especially for the breast cancer, including genetic, molecular and cellular mechanisms, is now providing clear objectives for their early detection, prevention, and therapy. The past 10 years have shown impressive progress in the field of large scale and high-throughput biology, resulting in a new area of technology development and improve of knowledge. During these years, protein- and nucleic acid -based high-throughput analyses have identified a substantial number of markers at the levels of proteins, genes, and gene expression, which can be linked to various forms of cancer.

Recently, the epigenetic of human cancer have become more visible due to increasing understanding of specific epigenetic mechanisms, including hypomethylation, hypermethylation, loss of imprinting, and chromatin modification [8]. It is becoming more evident that epigenetic, as well as genetic events might have important role in the initiation and progression of cancer. Hypermethylation of gene-promoter regions is being revealed as a loss of gene function, which is frequently found in cancers [9].

Hypermethylation of human tumor suppressor genes (TSGs) leads to the silencing of genes responsible for tumor suppression, thus causing cancers. Aberrant DNA methylation patterns have been suggested as biomarkers in cancer molecular diagnostics [10]. Methylation is the ideal parameter for comprehensive diagnostics in cancer management, with the potential to trigger a change in paradigm from single to multiple markers, and from individuals to a disease-management approach [11]. Recent technology development has provided the analysis of DNA methylation in a genome-wide scale [12,13] which may not be easily accessible for many institutions. Thereby, in most of the research centres methylation assays can be only determined on gene-by-gene analysis. Additionally clinical usefulness of methylation is still limited due to the fact that none of the available techniques are broadly accurate for quantification and sensitive detection of methylation changes in cancers [9]. Large-scale studies that enable evaluating quantitative alterations of methylation for multiple CpG sites in various gene regions and testing a large number of samples with automation are rare [9,14].

The use of methylation changes as a biomarker has a number of advantages compared with other approaches. Two of these advantages should be underlined: First, there is strong evidence that methylation is an early event in carcinogenesis, a characteristic highly desired in cancer biomarkers. Second, the DNA containing the methylation information is highly stable and can be easily isolated from most body fluids, as well as from archived fixed tissues. These include early detection, chemoprevention, and disease monitoring [15].

It is well recognized that solid malignant tumors release significant amounts of DNA into the systemic circulation through cellular necrosis or apoptosis [16]. The presence of cell-free DNA

(cfDNA) in plasma and serum has been known for over 60 years. Quantitative alteration of circulating cfDNA has been observed in several cancers including breast cancer [17]. The tumor released DNA in circulation might serve as biomarker for cancer [17]. Tumor-specific methylated DNA alterations have been found in the circulation of patients with different types of cancer [18,19]. The analysis of the methylation patterns of cfDNA by a blood-based test might enable to distinguish between benign and malignant tumors for diagnosis and surveillance of patients [19].

Abnormalities in non-coding genes can also contribute to cancer pathogenesis [20,21]. MicroRNAs (miRNAs) are endogenous non-coding RNAs with 19-25 nucleotides in size that deregulation of them contribute to cancer development and progression [22]. Recent evidence indicates that some miRNAs can function either as oncogenes or tumor suppressors [21,23], and expression profiling analyses have revealed characteristic miRNA signatures in certain human cancers [20,24]. However, the precise parts played by the expressed miRNAs in specific steps of malignant progression, including metastasis, are still unknown. Moreover, emerging studies reported that miRNAs are involved in promoter DNA methylation changes [25]. However, DNA sequences encoding miRNAs were found to be a target of aberrant DNA methylation as well as protein-coding genes [26].

Methylation causes the inactivation of numerous genes/miRNAs that are important in the development of most, or all, tumor types. Genetic changes such as mutation or deletion resulting permanent loss of gene expression whereas epigenetic changes such as DNA methylation can be pharmacologically reversible and could be a useful target to develop new therapeutic strategies for cancer therapy [27,28]. Reversal hypermethylation of silenced TSGs or miRNAs is increasingly being targeted for cancer therapy and prevention strategies [29,30]. Moreover, These approaches are particularly appealing because of the less toxicity of DNA methylation inhibitors in non-cancerous tissues than the other anti-cancer drugs [15].

References

1. WHO (2008) World Health Organization. Cancer. http://wwwwhoint/mediacentre/factsheets/fs297/en/: WHO [online].
2. Pantel K, Muller V, Auer M, Nusser N, Harbeck N, et al. (2003) Detection and clinical implications of early systemic tumor cell dissemination in breast cancer. Clin Cancer Res 9: 6326-6334.
3. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, et al. (2003) The case for early detection. Nat Rev Cancer 3: 243-252.
4. Ries L MD, Krapcho M, Mariotto A, Miller B, Feuer E (2006) SEER cancer statistics review, 1975–2004.: National Cancer Institute, Bethesda.
5. Radpour R, Barekati Z, Kohler C, Holzgreve W, Zhong XY (2009) New trends in molecular biomarker discovery for breast cancer. Genet Test Mol Biomarkers 13: 565-571.
6. Pantel K, Woelfle U (2005) Detection and molecular characterisation of disseminated tumour cells: implications for anti-cancer therapy. Biochim Biophys Acta 1756: 53-64.
7. Sidransky D (2002) Emerging molecular markers of cancer. Nat Rev Cancer 2: 210-219.
8. Feinberg AP, Tycko B (2004) The history of cancer epigenetics. Nat Rev Cancer 4: 143-153.
9. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.
10. Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253-266.
11. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, et al. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97: 5237-5242.
12. Shen L, Waterland RA (2007) Methods of DNA methylation analysis. Curr Opin Clin Nutr Metab Care 10: 576-581.
13. Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. Development 134: 3959-3965.
14. Issa JP (2004) CpG island methylator phenotype in cancer. Nat Rev Cancer 4: 988-993.
15. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.
16. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ (1977) Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 37: 646-650.
17. Zhong XY, Ladewig A, Schmid S, Wight E, Hahn S, et al. (2007) Elevated level of cell-free plasma DNA is associated with breast cancer. Arch Gynecol Obstet 276: 327-331.
18. Wong TS, Kwong DL, Sham JS, Wei WI, Kwong YL, et al. (2004) Quantitative plasma hypermethylated DNA markers of undifferentiated nasopharyngeal carcinoma. Clin Cancer Res 10: 2401-2406.
19. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683-692.
20. Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. Nat Rev Cancer 6: 857-866.
21. Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 6: 259-269.
22. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.
23. Slack FJ, Weidhaas JB (2006) MicroRNAs as a potential magic bullet in cancer. Future Oncol 2: 73-82.
24. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. (2005) MicroRNA expression profiles classify human cancers. Nature 435: 834-838.
25. Fabbri M, Ivan M, Cimmino A, Negrini M, Calin GA (2007) Regulatory mechanisms of microRNAs involvement in cancer. Expert Opin Biol Ther 7: 1009-1019.
26. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, et al. (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 67: 1424-1429.
27. Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, et al. (2008) Transient cyclical methylation of promoter DNA. Nature 452: 112-115.
28. Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, et al. (2008) Cyclical DNA methylation of a transcriptionally active promoter. Nature 452: 45-50.
29. Yoo CB, Jones PA (2006) Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov 5: 37-50.
30. Brueckner B, Kuck D, Lyko F (2007) DNA methyltransferase inhibitors for cancer therapy. Cancer J 13: 17-22.

2. Aim of the study

Method development approaches

I) The rapidly expanding interest in the involvement of DNA methylation in developmental mechanisms, human diseases and malignancies has highlighted the need for an accurate, quantitative and high-throughput assay. Existing methods are limited, and are often too laborious for high throughput or inadequate for quantitative analysis of methylation. Therefore, the primary aim of this thesis was to present an approach that allows for reduced costs, based on the SEQUENOM's EpiTYPER™ which is a high-throughput methylation quantification method and relies on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and our modified protocol.

II) Investigation of DNA, RNA and proteins for downstream genetic, epigenetic, transcriptomic and proteomic analysis holds an important place in the field of medical care and life science. This is often hampered by the limited availability of sample material. Therefore an increasing interest rises for simultaneous isolation of DNA, RNA and proteins from a single sample aliquot. As another aim we tried to establish applicable methodology of simultaneous isolation of DNA, RNA and/or proteins for quantitative and qualitative analysis of breast cancerous cell lines, whole blood, buffy coat, serum, plasma and formalin-fixed paraffin-embedded tissues for genetic, epigenetic, transcriptomic and proteomic profiling.

Biomarker discovery in cancerous and paired normal breast tissues

Alterations of DNA methylation patterns have been suggested as biomarkers for diagnosis and therapy of cancers. Every novel discovery in the epigenetic landscape and every development of an improved approach for accurate analysis of the events may offer new opportunity for the management of patients. Breast cancer is the most common type of cancer and most common

leading cause of cancer death in women. Therefore, the main aim of this study was to focus on the careful selection and analyses of candidate genes mainly involved in breast cancer pathways. In this study, we investigated quantitative methylation changes of 22 human genes (*APC, BIN1, BMP6, BRCA1, BRCA2, CADHERIN 1, CST6, DAPK1, EGFR, ESR2, GSTP1, NES1, Nm23-H1, P16, P21, Progesterone receptor, Prostasin, RAR-b, RASSF1, SRBC, TIMP3, TP53*) in the breast cancer tissue compared to the paired normal tissue using MALDI-TOF MS system purposing novel applicable biomarkers for diagnosis and prognosis of patients. The selected candidate cancer genes for this study have annotated functionality in the cell adhesion, cell interaction, invasion, metastasis, angiogenesis or gene expression during cancer development and progression. In order to understand the relationship between methylation and transcription events, individual assessment of the methylation status of each CpG dinucleotide was performed to check the position of hypermethylated sites along the gene's promoter.

Unregulated cell growth, a major hallmark of cancer, is associated with telomere shortening. Measurement of telomere length could provide important information on cell replication and proliferation status in cancer tissues. In another part of study we investigated a potential link between promoter hypermethylation of the *TP53*, *P21* and *P16* genes and telomere length shortening in the paired breast tumor and the adjacent normal breast tissue by quantitative PCR and MALDI-TOF MS to measure relative telomere length and promoter methylation level of aforementioned genes, respectively. Additionally, we compared telomere length in tumor tissues to traditional pathological parameters and clinical predictive markers.

Non-invasive biomarker discovery based on cancer specific methylation alteration

It is well recognized that solid malignant tumors release significant amounts of DNA into the systemic circulation through cellular necrosis or apoptosis. Quantitative alterations of circulating cell-free DNA have been observed in several cancers. This tumor-released DNA in blood might

serve as biomarker for cancer. In the present study, to achieve a reliable gene panel for developing a high-throughput blood-based test, we quantitatively assessed the DNA methylation profile of 10 breast cancer candidate genes (*APC, BIN1, BMP6, BRCA1, CST6, ESR2, GSTP1, P16, P21* and *TIMP3*) using MALDI-TOF MS in two different cohorts of patients with breast cancer on large-scale CpG sites.

Demethylation treatment

Genetic changes such as mutation or deletion resulting permanent loss of gene expression whereas epigenetic changes can be pharmacologically reversible and might be a useful target to develop new therapeutic strategies for cancer therapy. Reversal hypermethylation of silenced tumor suppressor genes or miRNAs is increasingly being targeted for cancer therapy and prevention strategies.

The 5-aza-2′-deoxycytidine (decitabine; DAC; Dacogen, Eisai, Inc.) as an effective demethylation drug is recently approved by the Food and Drug Administration (FDA) for the treatment of patients with Myelodysplastic Syndromes (MDS) and leukemia. The DAC is a nucleotide analog that is activated via phosphorylation by cellular deoxcytidine kinase and incorporated into the DNA at S phase, eventually lead to depletion of methyltransferase activity and demethylation of DNA.

Present study investigated the therapeutic value of DAC on the cancer specific hypermethylated genes. We could also provide a multidimensional model which showed complete early and late effects of DAC at the level of the genome, epigenome and proteome for subtypes of breast cancer. The present study initially assessed effective dosage of DAC for breast cancer therapy based on promoter demethylation considering cell viability, cytotoxicity and apoptosis for six breast cancer cell lines and a breast epithelial cell line as control. Then two breast cancer subtypes including highly aggressive and non-aggressive cell lines as well as the control cell line were investigated in more detail using 3-dimentional omics (gene expression, microRNA expression and proteomics analysis).

3. Published review article:

New trends in molecular biomarker discovery for breast cancer

Journal: Genet Test Mol Biomarkers. 2009 Oct;13(5):565-71. Review.

Summary:

Breast cancer is one of the most common leading causes of cancer death in women and early diagnosis, selection of appropriate therapeutic strategies and efficient follow up play an important role in reducing mortalities. Our review highlights the new trends and approaches in breast cancer biomarker discovery, which could be used for early diagnosis, development of new therapeutic approaches and follow up of patients.

First author's contribution:
Ramin Radpour was involved in writing the manuscript.

REVIEW ARTICLE

New Trends in Molecular Biomarker Discovery for Breast Cancer

Ramin Radpour,[1] Zeinab Barekati,[1] Corina Kohler,[1] Wolfgang Holzgreve,[2] and Xiao Yan Zhong[1]

Breast cancer is one of the most common and leading causes of cancer death in women. Early diagnosis, selection of appropriate therapeutic strategies, and efficient follow-up play an important role in reducing mortality. Recently, HER-2/neu in breast cancer has been routinely used to guide treatment of using Trastuzumab in less than 25–30% of patients. More new biomarkers will be still expected in the future to tailor treatments. However, there are still many obstacles in developing clinically useful biomarker tests for clinical practice. A lack of specificity of tumor markers and lack of sensitivity of testing systems have been noticed, which limit their clinical use. Finding biomarkers for breast cancer could allow physicians to identify individuals who are susceptible to certain types and stages of cancer to tailor preventive and therapeutic modalities based on the genotype and phenotype information. These biomarkers should be cancer specific, and sensitively detectable in a wide range of specimen(s) containing cancer-derived materials, including body fluids (plasma, serum, urine, saliva, etc.), tissues, and cell lines. This review highlights the new trends and approaches in breast cancer biomarker discovery, which could be potentially used for early diagnosis, development of new therapeutic approaches, and follow-up of patients.

Introduction

BREAST CANCER IS ONE of the most common malignancies accounting for 18% of all cancers in women, making it the leading cause of cancer-related deaths in women (World Health Organization, 2008). In the United States, estimates for 2008 indicate that over 240,000 new cases of *in situ* and invasive breast cancer will be diagnosed among women, and an additional 2030 new cases will be diagnosed among men (American Cancer Society, 2007). At present, routine mammography is the most widely used tool for the early detection of breast cancer. To be detected, however, a tumor should be at least a few millimeters in size. Only 60% of breast cancer is diagnosed at a local stage (American Cancer Society, 2007). At the time of primary diagnosis, usually tumor cells had already been shed and occult metastases had already occurred (Pantel *et al.*, 2003).

Earlier diagnosis and treatment of breast cancer could play an important role in reducing deaths. Many researchers have attempted to establish molecular and immunological methods for the detection of individual metastatic breast cancer cells in peripheral blood and bone marrow (Pantel and Woelfle, 2005), but still finding an accurate and sensitive biomarker remains challenging (Sidransky, 2002).

Our increasing knowledge about the various forms of cancer, especially for the breast cancer, including genetic, molecular, and cellular mechanisms, is now providing clear objectives for their early detection, prevention, and therapy. The past 10 years have shown impressive growth in the field of large-scale and high-throughput biology, resulting in a new era of technology development and the accumulation of new knowledge. During these years, protein-based and nucleic acid–based high-throughput analyses have identified a substantial number of markers at the level of proteins, genes, and gene expression, which can be linked to various forms of cancer. Further, the recently acquired knowledge of how chromatin organization moderates transcription has highlighted the importance of epigenetic mechanisms in the initiation and progression of cancer. These epigenetic changes, in particular, the aberrant promoter hypermethylation, which is associated with inappropriate gene silencing, virtually affect every step of tumor progression (Li *et al.*, 2001). This type of newly acquired knowledge has opened new windows for finding biomarkers for the management of cancer patients. Figure 1 shows general procedures in biomarker discovery.

Cancer biomarkers can be divided into three categories. (a) Diagnostic (classification) biomarkers are used to detect and identify a given type of cancer in an individual. This type of

---

[1]Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel, Basel, Switzerland.
[2]University Medical Center Freiburg, Freiburg, Germany.
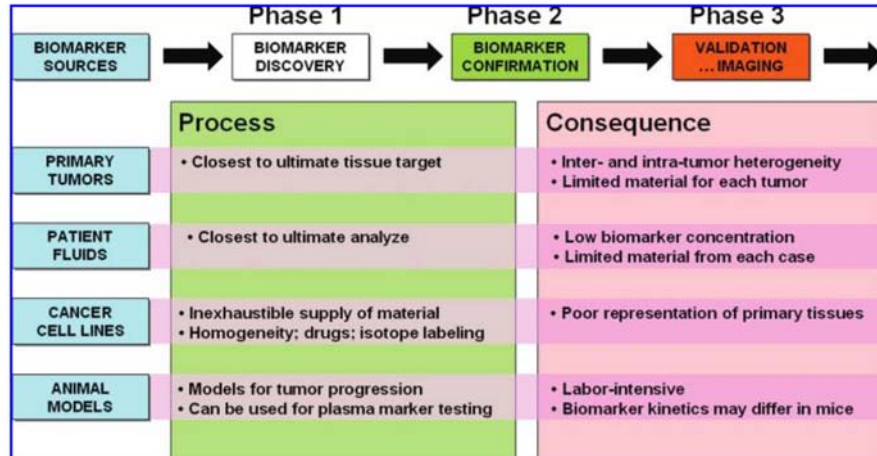
FIG. 1. General procedures in cancer biomarker discovery. Color images available online at www.liebertonline.com/gtmb.

biomarker is expected to have high levels of diagnostic sensitivity and specificity. (b) Prognostic biomarkers are commonly used once the disease status has been established. (c) Stratification (predictive) biomarkers can serve to predict the likely response to a drug before starting treatment, thus classifying individuals as responders or nonresponders (van de Vijver et al., 2002).

This review highlights the new trends in biomarker discovery for breast cancer, and how they could be used for the diagnosis of breast cancer, as well as for the management of breast cancer in terms of developing new diagnostic/therapeutic approaches.

DNA Copy-Number Variations

A human oncogenome has numerous alterations that include aberrations in the DNA sequence or structure and in the number of particular sequences, including genes or chromosomes (DNA copy number variants [CNVs]). These reversible and irreversible changes can affect hundreds to thousands of genes and/or regulatory transcripts. They result in the activation or inhibition of various biological events, thereby causing aspects of cancer pathophysiology, including angiogenesis, immune evasion, metastasis, altered cell growth, and death (Chin and Gray, 2008). Changes in the copy number of genetic regions or chromosomes across the entire genome of a cancer cell can be mapped onto a representation of the normal genome by using comparative genomic hybridization (Kallioniemi et al., 1992). Modern analysis platforms for comparative genomic hybridization map copy-number changes onto DNA sequences arranged in microarrays (Greshock et al., 2007), and allow these changes to be assessed quantitatively with subgene resolution.

The CNVs cover approximately 12% of the human genome, potentially altering gene dosage, disrupting genes or perturbing regulation of their expression. The overall dispersion of CNVs might also impact the interindividual differences in drug response (Ouahchi et al., 2006), as well as susceptibility to cancer (Cho et al., 2006), either directly, or by modulating penetrance or variability in the expression of the trait examined. In one recent study Han et al. (2008) could identify the DNA copy-number alterations and expression of relevant genes characteristic in triple-negative breast cancer, using two different gene expression array data sets. Triple-negative breast cancer is defined by a lack of expression of estrogen, progesterone, and HER2 receptors. Genetically, most of these fall into the basal subgroup of breast cancer. In that study they found 104 upregulated genes in the gained regions. In another study, Hwang et al. (2008) tried to establish models that predict systemic recurrence in breast cancer as predictive biomarkers. By selecting marker clones with CNAs, using an array comparative genomic hybridization containing 4044 human bacterial artificial chromosome clones in breast cancer, they could find significant CNAs in the chromosomal regions of 5p15.33, 11q13.3, 15q26.3, 17q25.3, including 18q23 and 21q22.3 with gain, and 9p12, 11q24.1, and 14q32.33 with loss.

CNVs were able to offer some of the missing clues to the genetic enigma of complex disorders like cancers. Currently, our knowledge of CNVs is still incomplete, and higher-resolution whole-genome tiling arrays are needed to capture CNVs of smaller size, that is, in the range of several kilobases.

DNA Mutations as Biomarkers

Genomic mutations screening

Human cancer is mostly caused by the accumulation of mutations in oncogenes and tumor suppressor genes. Discovery of the genes mutated in human cancer has provided key insights into the mechanisms underlying tumorigenesis, and has proven useful for the design of a new generation of targeted approaches for clinical intervention. Recent efforts in large-scale DNA sequence analysis have identified several hundred candidate genes that might have functional roles in various human cancers (Greenman et al., 2007; Wood et al., 2007). Some of these genetic alterations and mutations occur at a relatively high frequency, but most are present in only a few percent of tumors. Results from the extensive sequencing and mutation-validation efforts that are now underway will be necessary to establish the prevalence and clinicopathological associations for these genetic elements of interest. Several important discoveries regarding cancer mutation profiling have come from the systematic resequencing of genes, gene families, or genes in pathways that are relevant to cancer (Chin and Gray, 2008). In one recent study, Wood et al. (2007) revealed the complete mutation profile of breast and

colorectal cancers. As a result of this discovery, it is now easy to identify the genetic alterations in cancers on a genome-wide scale. It is still, however, much more difficult to elucidate the precise role of these alterations in tumorigenesis. The compendium of genetic changes in individual tumors provides new opportunities for individualized diagnosis and treatment of cancer.

Mitochondrial mutations screening

Mitochondrial DNA (mtDNA) has been proposed to be involved in carcinogenesis because of its high susceptibility to mutations and limited repair mechanisms in comparison to nuclear DNA (Copeland et al., 2002). The human eukaryotic cell contains hundreds or thousands of mitochondria, and each mitochondrium contains 1–10 copies of mtDNA. The high copy number in comparison with that of nuclear DNA enables the detection of rare target cells, even at low levels (Petros et al., 2005).

There have been a variety of mtDNA alterations found in breast tumor tissues (Brandon et al., 2006; Jakupciak et al., 2008; Losanoff et al., 2008). Therefore, instabilities and alterations of mtDNA in tumorigenesis may serve as earlier markers for cancer development, and may hold a potential role in tracking tumor progression and tumor metastasis (Radpour et al., 2008). By virtue of the clonal nature of mitochondria and high copy number, mitochondrial mutations may provide a powerful molecular biomarker for the detection of cancer. It has been suggested that the extent of mtDNA mutations might be useful in the prognosis of cancer outcome and/or the response to certain therapies (Copeland et al., 2002). The mutated mtDNA was readily detectable in cancer-paired body fluids (including urine, saliva, and sputum) from each type of cancer, and was 19–22 times more abundant than mutated nuclear DNA (Fliss et al., 2000).

DNA Methylation

Recently, the epigenetics of human cancer have become more visible, along with a growing understanding of specific epigenetic mechanisms, including hypomethylation, hypermethylation, loss of imprinting, and chromatin modification (Feinberg and Tycko, 2004). It is becoming more evident that epigenetic as well as genetic events might be central to the initiation and progression of cancer. Hypermethylation of gene-promoter regions is being revealed as a loss of gene function, which is frequently found in cancers. Methylation is the ideal parameter for comprehensive diagnostics in cancer management, with the potential to trigger a change in paradigm from single to multiple markers, and from individuals to a disease-management approach (Ramsahoye et al., 2000).

The use of methylation changes as a biomarker has a number of advantages compared with other approaches. Two of these advantages should be underlined: first, there is strong evidence that methylation is an early event in carcinogenesis, a characteristic highly desired in cancer biomarkers; second, the DNA containing the methylation information is highly stable and can be easily isolated from most body fluids, as well as from archived fixed tissues. These include early detection, chemoprevention, and disease monitoring (Suzuki and Bird, 2008).

Methylation probably causes the inactivation of numerous genes that are important in the development of most, or all,

tumor types. Thus, inhibition of DNA methylation and consequent re-activation of these genes is an attractive avenue for the development of novel therapeutics (Suzuki and Bird, 2008). Recently, two studies showed that the methylation process becomes transient, cyclical, and dynamic (Kangaspeska et al., 2008; Metivier et al., 2008). Unlike the stable and unchangeable pathological genetic alterations, this cyclical and dynamic epigenetic nature may be useful for developing new strategies in the treatment of cancer patients, and in risk assessment of cases before the onset, through modification of pathologic methylation patterns (Kangaspeska et al., 2008).

Recently, by analyzing methylation profiles of 42,528 CpG sites on the 22 genes that have a direct effect on breast carcinogenesis, we could show the methylation profile of breast cancer. Using thymidine-specific cleavage mass array (Radpour et al., 2008) on MALDI-TOF silico chips, 10 hypermethylated genes (*APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESRb*, *GSTP1*, *P16*, *P21*, and *TIMP3*) were identified as the top 10 mountains for breast cancer. These genes may be used as biomarkers for further understanding the nature of epigenetics in breast cancer, and for improved management of cancer patients in the future (Radpour et al., 2009).

Gene-Expression Profiling

Breast cancer develops as a result of multiple genetic defects, and individuals with the same type of cancer often have dissimilar genetic defects in their tumors. This finding explains why patients who seem to have similar cancers respond in a heterogeneous manner to anticancer agents. The use of DNA microarray technology has made it possible to assess the expression of tens of thousands of genes in a single experiment (Schena et al., 1995). Systematic analysis of the gene expression patterns of tumor samples enabled researchers to identify characteristic expression patterns of groups of genes that are associated with specific tumor traits. These patterns are known as gene expression signatures and can be used for the molecular classification of tumors and subtypes of cancer on the basis of gene expression (van't Veer and Bernards, 2008).

Two similar studies were designed to find a prognostic gene expression signature for breast cancer. In one, authors mentioned a 70-gene signature for breast cancer prognosis (Buyse et al., 2006). In the other study, authors showed a 76-gene signature (Desmedt et al., 2007), but this had only three genes in common with the 70-gene signature. This finding was interpreted by some to indicate that such gene expression signatures are highly unstable, and that these two signatures use different genes to monitor the same biological processes (Yu et al., 2007). From all the gene expression signatures for breast cancer that have been identified, only three are commercially available and can be used in the clinic: the 70-gene signature for breast cancer prognosis is available under the name MammaPrint (Agendia, Huntington Beach, CA); the 16-gene signature is available as Oncotype DX (Genomic Health, Inc., Redwood City, CA); and a 2-gene signature, which has recently been released, is available under the name the H/I test (AviaraDx, bioTheranostics, San Diego, CA).

Also, gene expression signatures can be used as predictive profiles for evaluating a patient's response to therapeutic strategies, as well as to find more specific therapies that are

targeted to each kind of tumor and might provide a new era of personalized medicine.

MicroRNA Expression Profiling

Abnormalities in noncoding genes can also contribute to cancer pathogenesis (Calin and Croce, 2006; Esquela-Kerscher and Slack, 2006). A class of small cellular RNAs, termed microRNAs (miRNAs), acts as agents of the RNA interference pathway, and can lead to silencing of their cognate target genes. miRNAs have been implicated in regulating diverse cellular pathways. Recent evidence indicates that some miRNAs can function either as oncogenes or tumor suppressors (Esquela-Kerscher and Slack, 2006; Slack and Weidhaas, 2006), and expression profiling analyses have revealed characteristic miRNA signatures in certain human cancers (Lu et al., 2005; Calin and Croce, 2006). However, the precise parts played by the expressed miRNAs in specific steps of malignant progression, including metastasis, are still unknown.

In one recent study, using a combination of mouse and human cells, Ma et al. (2007) could show that miRNA-10b is highly expressed in metastatic breast cancer cells and positively regulates cell migration and invasion. Overexpression of miRNA-10b in nonmetastatic breast tumors initiates invasion and metastasis.

Telomere Length Dynamics

Telomeres are special structures consisting of a variable number of repeated sequences (TTAGGG) at the ends of chromosomes (Blackburn, 1991). Telomeres play a key role in the maintenance of chromosomal stability (Blackburn, 1991). Short telomere length, as a measure of telomere dysfunction, was significantly related to baseline and mutagen-induced genetic instability (Wu et al., 2003; Blasco, 2005). Telomeric DNA is dynamic, and is progressively lost with each cell division due to incomplete replication of the termini of linear DNA molecules (the end replication problem).

Telomere length and telomerase activity play a dual role in tumorigenesis. In the early stage of malignant transformation, telomere loss limits cell proliferation, and telomerase activation protects the ends of the chromosome and suppresses tumorigenesis. In the late stages of tumorigenesis, telomere loss induces genomic instability, and telomerase activation promotes immortalization (Hahn, 2003; Harley, 2008). In cancer cells, the telomere length has a wide range of variability, and its equilibrium depends on the balance found between the telomere shortening from cell division, and the telomere elongating as a result of telomerase activity (Hahn, 2003). Therefore, the telomere length could serve as a useful indicator and biomarker in the risk assessment and prediction of different stages of breast cancer.

Several studies examining telomere length in humans found that breast carcinomas had shorter telomeres than normal breast tissue, and high-grade (grade III of III) invasive carcinomas had shorter telomeres than low-grade (grade I of III) invasive carcinomas (Rha et al., 1999). Dysfunctional telomeres are considered to be an early initiating event in breast cancer development, inducing chromosomal instability (Meeker et al., 2004). Since dynamic telomere length is determined by both genetic and environmental factors (Shen et al., 2007), its length in peripheral blood DNA is a potentially

useful biomarker, as a proxy of target tissue, to explore individual susceptibility to disease in epidemiologic studies. In one recent study, Svenson et al. (2008) showed that telomere length in peripheral blood cells differs between breast cancer patients and control subjects, and may serve as a significant prognostic biological marker.

Proteomic Profiling

Extensive activities dealing with protein profiling and analyses have generated a tremendous amount of data on the expression/modification of proteins under various forms of cancer, including breast cancer. Subsequent interpretation and assignment of such data allowed the identification of some biomarkers for early detection of breast cancer. Figure 2 shows the different steps in protein biomarker discovery and validation. Three of these protein families that have an important role in biomarker discovery of breast cancer are discussed below.

Human tissue kallikreins (KLK) are serine proteases encoded by 15 structurally similar, steroid hormone–regulated genes that colocalize to chromosome 19q13.4 in a 300-kb region (Diamandis et al., 2000; Yousef et al., 2001). Numerous reports have shown an association between dysregulated KLK expression and various types of cancer, as well as suggested their potential use as diagnostic/prognostic biomarkers for cancer (Esler and Wolfe, 2001; Krane, 2003). One example, which stands out in the role of KLK in breast cancer, is hk14. The authors reported that hK14 levels were elevated in a proportion of patients with ovarian (65%) and breast (40%) cancers (Borgono et al., 2003). As another example, *KLK10* was downregulated in breast and ovarian cancers due to the hypermethylation of coding exon 3 (Li et al., 2001; Roman-Gomez et al., 2004).

Protein family 14-3-3 is a large family of 25–30 kDa acidic proteins, and has seven homologous isoforms in mammalian cells, which were designated with the Greek letters β, γ, ε, η, σ, τ (sometimes referred to as θ, and ζ) (Fu et al., 2000). Processes that are relevant to cancer biology, and that are regulated by 14-3-3 protein interactions include cell-cycle progression, apoptosis, and mitogenic signaling, such as the ATM-*p53* pathway (Vogelstein et al., 2000). Of all the *14-3-3* genes, *14-3-3* σ has been most directly linked to cancer, and functions as a tumor suppressor by inhibiting cell-cycle progression and by causing cells to leave the stem cell compartment and undergo differentiation (Hermeking, 2003). In one study, authors reported that hypermethylation of this gene was a consistent alteration in invasive breast cancer, which is acquired in the late preinvasive phase of tumor progression (Umbricht et al., 2001).

Heat shock or stress proteins (HSPs) are one group of small proteins that dramatically increase in most tissues after exposure to environmental or physiological insults. HSPs encompass several groups of proteins, and may be divided into five major families on the basis of their size, structure, and function: HSP110, HSP90, HSP70, HSP60, and small HSP families (Lindquist and Craig, 1988). The increased expression of HSPs that is observed in many tumor types has been interpreted as an effort by the malignant cells to maintain homeostasis in a hostile environment (Whitesell and Lindquist, 2005). Elevated expression of Hsp90, Hsp70, and Hsp27, either individually or in combination, has been widely reported
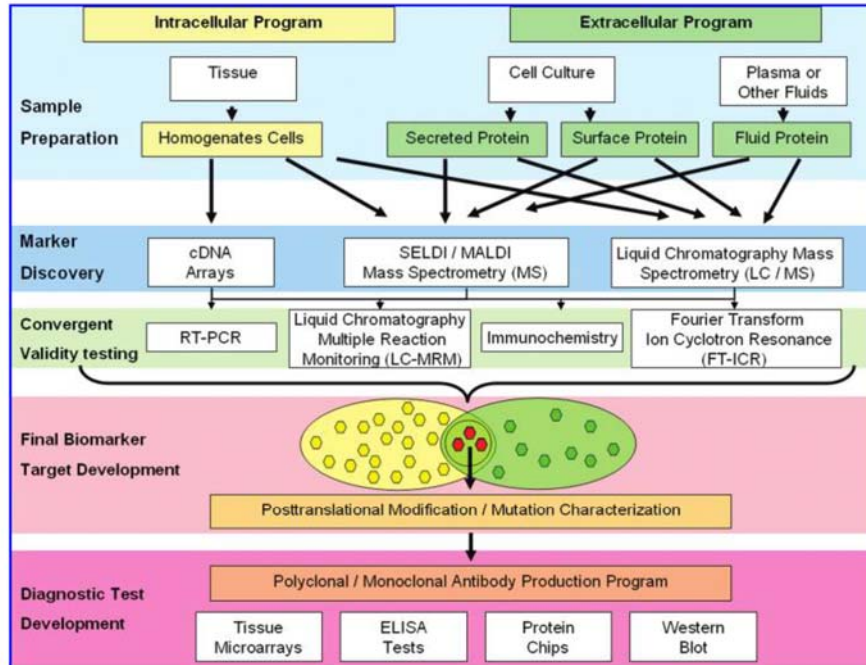
FIG. 2. Cascade of protein biomarker discovery and validation. Color images available online at www.liebertonline.com/gtmb.

in breast, uterine, renal, osteosarcoma, and endometrial cancers, and various leukemias (Helmbrecht *et al.*, 2000; Jolly and Morimoto, 2000).

Posttranslational Modification of Proteins as Biomarkers

Posttranslational modifications of proteins are important in many biological processes, and posttranslational changes have relevance to disease and cancer. The study of posttranslational modifications as a source of biomarkers in cancer is still at early stage. Cleavage products of tumor-derived proteins have been proposed as potential cancer biomarkers (Hanash *et al.*, 2008). Altered protein glycosylation in cancer is another source of potential cancer biomarkers (Kirmiz *et al.*, 2007). In the coming years, the study of posttranslational modifications will undoubtedly benefit from emerging mass-spectrometric technologies such as electron-transfer dissociation, enabling the study of modifications such as glycosylation and phosphorylation to provide a new generation of cancer biomarkers (Hanash *et al.*, 2008).

Discussion

The search for informative biomarkers for the early diagnosis of cancer has been going on for decades, yielding only a few major breakthroughs. Our increasing understanding of the biology of cancer, including genetic, molecular, and cellular mechanisms, and epigenetic background is now providing objectives for the early detection of some malignancies, including breast cancer. Such progress has had a direct impact on current activities dedicated to the search for sensitive and specific biomarkers for the early detection and diagnosis of cancers (Sawyers, 2008).

Although many papers are published each year describing genetic mutations or alterations in expression levels that are associated with various types of cancers, very few of these are developed into reliable molecular markers that can be used routinely in the clinical setting. Translating biomarker research into clinically useful tests has often been a discouraging activity. Many of the biomarkers identified in the initial tumor studies, which were retrospective, failed to be validated in subsequent studies. One area in which molecular biomarkers are expected to make a substantial contribution is in our capability to distinguish between subclasses of the same type of cancer. However, we can anticipate that current research efforts to discover sensitive and specific cancer biomarkers will provide a valuable contribution to future efforts to combat these malignancies.

The above-mentioned examples are an indication that molecular biomarkers hold great promise for enhancing our ability to establish early diagnosis, prognosis, and subclassification of some types of human cancers, including breast cancer. The same class of biomarkers will also contribute to a better prediction of response to therapy and metastasis.

Acknowledgments

Disclosure Statement

No competing financial interests exist.

References

American Cancer Society (2007) Cancer Facts and Figures, Atlanta, GA. www.cancer.org/downloads/STT/CAFF2007PWsecured.pdf
Blackburn EH (1991) Structure and function of telomeres. Nature 350:569–573.

Blasco MA (2005) Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet 6:611–622.

Borgono CA, Grass L, Soosaipillai A, *et al.* (2003) Human kallikrein 14: a new potential biomarker for ovarian and breast cancer. Cancer Res 63:9032–9041.

Brandon M, Baldi P, Wallace DC (2006) Mitochondrial mutations in cancer. Oncogene 25:4647–4662.

Buyse M, Loi S, van't Veer L, *et al.* (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 98:1183–1192.

Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. Nat Rev Cancer 6:857–866.

Chin L, Gray JW (2008) Translating insights from the cancer genome into clinical practice. Nature 452:553–563.

Cho EK, Tchinda J, Freeman JL, *et al.* (2006) Array-based comparative genomic hybridization and copy number variation in cancer research. Cytogenet Genome Res 115:262–272.

Copeland WC, Wachsman JT, Johnson FM, Penta JS (2002) Mitochondrial DNA alterations in cancer. Cancer Invest 20:557–569.

Desmedt C, Piette F, Loi S, *et al.* (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13:3207–3214.

Diamandis EP, Yousef GM, Luo LY, *et al.* (2000) The new human kallikrein gene family: implications in carcinogenesis. Trends Endocrinol Metab 11:54–60.

Esler WP, Wolfe MS (2001) A portrait of Alzheimer secretases—new features and familiar faces. Science 293:1449–1454.

Esquela-Kerscher A, Slack FJ (2006) Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer 6:259–269.

Feinberg AP, Tycko B (2004) The history of cancer epigenetics. Nat Rev Cancer 4:143–153.

Fliss MS, Usadel H, Caballero OL, *et al.* (2000) Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science 287:2017–2019.

Fu H, Subramanian RR, Masters SC (2000) 14-3-3 proteins: structure, function, and regulation. Annu Rev Pharmacol Toxicol 40:617–647.

Greenman C, Stephens P, Smith R, *et al.* (2007) Patterns of somatic mutation in human cancer genomes. Nature 446:153–158.

Greshock J, Feng B, Nogueira C, *et al.* (2007) A comparison of DNA copy number profiling platforms. Cancer Res 67:10173–10180. Epub 12007 Oct 10129.

Hahn WC (2003) Role of telomeres and telomerase in the pathogenesis of human cancer. J Clin Oncol 21:2034–2043.

Han W, Jung EM, Cho J, *et al.* (2008) DNA copy number alterations and expression of relevant genes in triple-negative breast cancer. Genes Chromosomes Cancer 47:490–499.

Hanash SM, Pitteri SJ, Faca VM (2008) Mining the plasma proteome for cancer biomarkers. Nature 452:571–579.

Harley CB (2008) Telomerase and cancer therapeutics. Nat Rev Cancer 8:167–179.

Helmbrecht K, Zeise E, Rensing L (2000) Chaperones in cell cycle regulation and mitogenic signal transduction: a review. Cell Prolif 33:341–365.

Hermeking H (2003) The 14-3-3 cancer connection. Nat Rev Cancer 3:931–943.

Hwang KT, Han W, Cho J, *et al.* (2008) Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. Int J Cancer 123:1807–1815.

Jakupciak JP, Maggrah A, Maragh S, *et al.* (2008) Facile whole mitochondrial genome resequencing from nipple aspirate fluid using MitoChip v2.0. BMC Cancer 8:95.

Jolly C, Morimoto RI (2000) Role of the heat shock response and molecular chaperones in oncogenesis and cell death. J Natl Cancer Inst 92:1564–1572.

Kallioniemi A, Kallioniemi OP, Sudar D, *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818–821.

Kangaspeska S, Stride B, Metivier R, *et al.* (2008) Transient cyclical methylation of promoter DNA. Nature 452:112–115.

Kirmiz C, Li B, An HJ, *et al.* (2007) A serum glycomics approach to breast cancer biomarkers. Mol Cell Proteomics 6:43–55. Epub 2006 Jul 2017.

Krane SM (2003) Elucidation of the potential roles of matrix metalloproteinases in skeletal biology. Arthritis Res Ther 5:2–4. Epub 2002 Oct 2008.

Li B, Goyal J, Dhar S, *et al.* (2001) CpG methylation as a basis for breast tumor-specific loss of NES1/kallikrein 10 expression. Cancer Res 61:8014–8021.

Lindquist S, Craig EA (1988) The heat-shock proteins. Annu Rev Genet 22:631–677.

Losanoff JE, Zhu W, Qin W, *et al.* (2008) Can mitochondrial DNA mutations in circulating white blood cells and serum be used to detect breast cancer? Mitochondrial mutations in cancer. Facile whole mitochondrial genome resequencing from nipple aspirate fluid using MitoChip v2.0. Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid. Breast 17:540–542. Epub 2008 Jun 2020.

Lu J, Getz G, Miska EA, *et al.* (2005) MicroRNA expression profiles classify human cancers. Nature 435:834–838.

Ma L, Teruya-Feldstein J, Weinberg RA (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 449:682–688. Epub 2007 Sep 2026.

Meeker AK, Hicks JL, Iacobuzio-Donahue CA, *et al.* (2004) Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. Clin Cancer Res 10:3317–3326.

Metivier R, Gallais R, Tiffoche C, *et al.* (2008) Cyclical DNA methylation of a transcriptionally active promoter. Nature 452:45–50.

Ouahchi K, Lindeman N, Lee C (2006) Copy number variants and pharmacogenomics. Pharmacogenomics 7:25–29.

Pantel K, Muller V, Auer M, *et al.* (2003) Detection and clinical implications of early systemic tumor cell dissemination in breast cancer. Clin Cancer Res 9:6326–6334.

Pantel K, Woelfle U (2005) Detection and molecular characterisation of disseminated tumour cells: implications for anticancer therapy. Biochim Biophys Acta 1756:53–64.

Petros JA, Baumann AK, Ruiz-Pesini E, *et al.* (2005) mtDNA mutations increase tumorigenicity in prostate cancer. Proc Natl Acad Sci USA 102:719–724. Epub 2005 Jan 2012.

Radpour R, Fan A, Kohler C, *et al.* (2008) Current understanding of mitochondrial DNA in breast cancer. Breast J 15:505–509.

Radpour R, Kohler C, Montazer Haghighi M, *et al.* (2009) Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28:2969–2978.

Radpour R, Montazer Haghighi M, Fan A, *et al.* (2008) High-throughput hacking of the methylation patterns in breast cancer by *in vitro* transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip. Mol Cancer Res 6:1702–1709.

Ramsahoye BH, Biniszkiewicz D, Lyko F, *et al.* (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci USA 97:5237–5242.

Rha SY, Park KH, Kim TS, *et al.* (1999) Changes of telomerase and telomere lengths in paired normal and cancer tissues of breast. Int J Oncol 15:839–845.

Roman-Gomez J, Jimenez-Velasco A, Agirre X, *et al.* (2004) The normal epithelial cell-specific 1 (NES1) gene, a candidate tumor suppressor gene on chromosome 19q13.3-4, is down-regulated by hypermethylation in acute lymphoblastic leukemia. Leukemia 18:362–365.

Sawyers CL (2008) The cancer biomarker problem. Nature 452:548–552.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470.

Shen J, Terry MB, Gurvich I, *et al.* (2007) Short telomere length and breast cancer risk: a study in sister sets. Cancer Res 67: 5538–5544.

Sidransky D (2002) Emerging molecular markers of cancer. Nat Rev Cancer 2:210–219.

Slack FJ, Weidhaas JB (2006) MicroRNAs as a potential magic bullet in cancer. Future Oncol 2:73–82.

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465–476.

Svenson U, Nordfjall K, Stegmayr B, *et al.* (2008) Breast cancer survival is associated with telomere length in peripheral blood cells. Cancer Res 68:3618–3623.

Umbricht CB, Evron E, Gabrielson E, *et al.* (2001) Hypermethylation of 14-3-3 sigma (stratifin) is an early event in breast cancer. Oncogene 20:3348–3353.

van de Vijver MJ, He YD, van't Veer LJ, *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009.

van't Veer LJ, Bernards R (2008) Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature 452:564–570.

Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. Nature 408:307–310.

Whitesell L, Lindquist SL (2005) HSP90 and the chaperoning of cancer. Nat Rev Cancer 5:761–772.

Wood LD, Parsons DW, Jones S, *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. Science 318:1108–1113. Epub 2007 Oct 1111.

World Health Organization (2008) Cancer. www.who.int/mediacentre/factsheets/fs297/en/. [Online.]

Wu X, Amos CI, Zhu Y, *et al.* (2003) Telomere dysfunction: a potential cancer predisposition factor. J Natl Cancer Inst 95: 1211–1218.

Yousef GM, Diamandis M, Jung K, Diamandis EP (2001) Molecular cloning of a novel human acid phosphatase gene (ACPT) that is highly expressed in the testis. Genomics 74:385–395.

Yu JX, Sieuwerts AM, Zhang Y, *et al.* (2007) Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. BMC Cancer 7:182.

Address correspondence to:
Prof. Xiao Yan Zhong, M.D.
Laboratory for Prenatal Medicine and Gynecologic Oncology
Women's Hospital/Department of Biomedicine
University of Basel
Hebelstrasse 20, Room Nr. 416
Basel CH 4031
Switzerland

E-mail: zhongx@uhbs.ch

4. Published research article:

High-throughput hacking the methylation patterns in breast cancer by *in vitro* transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip

Summary:

DNA methylation is an important potential biomarker in cancer study but still an accurate, sensitive and reproducible high-throughput quantification of DNA methylation with compatibility of automation remains challenging. To find an efficient and more cost effective high-throughput method for analyzing the methylation profile in breast cancer, we assessed a methodological approach that allows for the simultaneous detection of multiple target CpGs residue and quantifies up to 5% of methylated sequence in unmethylated background by employing thymidine-specific cleavage mass array on MALDI-TOF slico chips. We evaluated the accuracy, variability and sensitivity of the approach, and implemented critical improvements in experimental design.

First author's contribution:
Ramin Radpour was involved in performing the experiment and data analysis.

High-Throughput Hacking of the Methylation Patterns in Breast Cancer by *In vitro* Transcription and Thymidine-Specific Cleavage Mass Array on MALDI-TOF Silico-Chip

Ramin Radpour,[1] Mahdi Montazer Haghighi,[2] Alex Xiu-Cheng Fan,[1] Peyman Mohammadi Torbati,[3] Sinuhe Hahn,[1] Wolfgang Holzgreve,[1] and Xiao Yan Zhong[1]

[1]Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel, Basel, Switzerland; [2]Department of Genetics, Azad University, East Tehran Branch; and [3]Department of Pathology, Shaheed Beheshti Medical University, Tehran, Iran

Abstract

Over the last decade, the rapidly expanding interest in the involvement of DNA methylation in developmental mechanisms, human diseases, and malignancies has highlighted the need for an accurate, quantitative, and high-throughput assay. Existing methods are limited and are often too laborious for high-throughput analysis or inadequate for quantitative analysis of methylation. Recently, a MassCLEAVE assay has been developed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry to analyze base-specific methylation patterns after bisulfite conversion. To find an efficient and more cost-effective high-throughput method for analyzing the methylation profile in breast cancer, we developed a method that allows for the simultaneous detection of multiple target CpG residues by using thymidine-specific cleavage mass array on matrix-assisted laser desorption/ionization time-of-flight silicon chips. We used this novel quantitative approach for the analysis of DNA methylation patterns of four tumor suppressor genes in 96 breast tissue samples from 48 patients with breast cancer. Each individual contributed a breast cancer specimen and corresponding adjacent normal tissue. We evaluated the accuracy of the approach and implemented critical improvements in experimental design. (Mol Cancer Res 2008;6(11):1702–9)

Introduction

DNA methylation is an important potential biomarker in cancer study and opens a new area in cancer therapy.

Although DNA methylation analysis is a rapidly developing field, studies about its clinical usefulness are limited due to the fact that no single technique is superior (1-4). An accurate, sensitive, and reproducible high-throughput quantification of DNA methylation with compatibility of automation remains challenging (4).

A novel EpiTYPER assay for high-throughput analysis of DNA methylation patterns using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has recently been introduced (5, 6). This assay is a tool for the detection and quantitative analysis of DNA methylation using MALDI-TOF MS and MassCLEAVE reagent, which enables base-specific (C/T) cleavage reactions (5, 7, 8). In a completely new and comprehensive study, by quantitative analysis of methylation patterns in a set of >400 candidate genes in 59 different cancer cell lines, the developer company showed robustness of (C/T) cleavage reactions in methylation studies (6).

The method uses T7-promoter–tagged PCR amplification of bisulfite-converted DNA, followed by generation of a single-stranded RNA molecules (9) and a subsequent procedure of base-specific cleavage (3′ to rUTP and rCTP) using RNase A (10-12). Bisulfite treatment of genomic DNA converts unmethylated cytosine into uracil while methylated cytosine remains unchanged. These C/T appear as G/A variations in the cleaved products generated from the reverse strand by C/T-specific cleavage. In the C-cleavage reaction, methylated regions are cleaved at every C to create fragments containing at least one CpG site each. For the T-cleavage reaction, both methylated and unmethylated regions are cleaved at every T to produce fragments (Fig. 1). These G/A variations result in a mass difference of 16 Da per CpG site, which is easily detected by the MassARRAY analyzer compact. In the mass spectrum, the relative amount of methylated sequence can be calculated by comparing the signal intensity between the mass signals of methylated and unmethylated templates to generate quantitative results for each cleavage product (5, 6, 13).

We present an approach that allows for reduced costs based on the above system for high-throughput DNA methylation analysis and our modified protocol. We performed a T-specific cleavage reaction on CpG islands after bisulfite conversion of target sequences to analyze the methylated sequences. The relative amounts of methylated DNA were automatically

calculated by the software. We applied this technology for high-throughput quantification of methylation patterns of four tumor suppressor genes (*Cadherin1, P16, RAR-b*, and *RASSF1*; Supplementary data 1) in the paired cancerous and adjacent normal breast tissues from 48 patients with breast cancer. The genes were observed to be involved in cell adhesion, cell interaction, or gene expression during cancer development and progression.
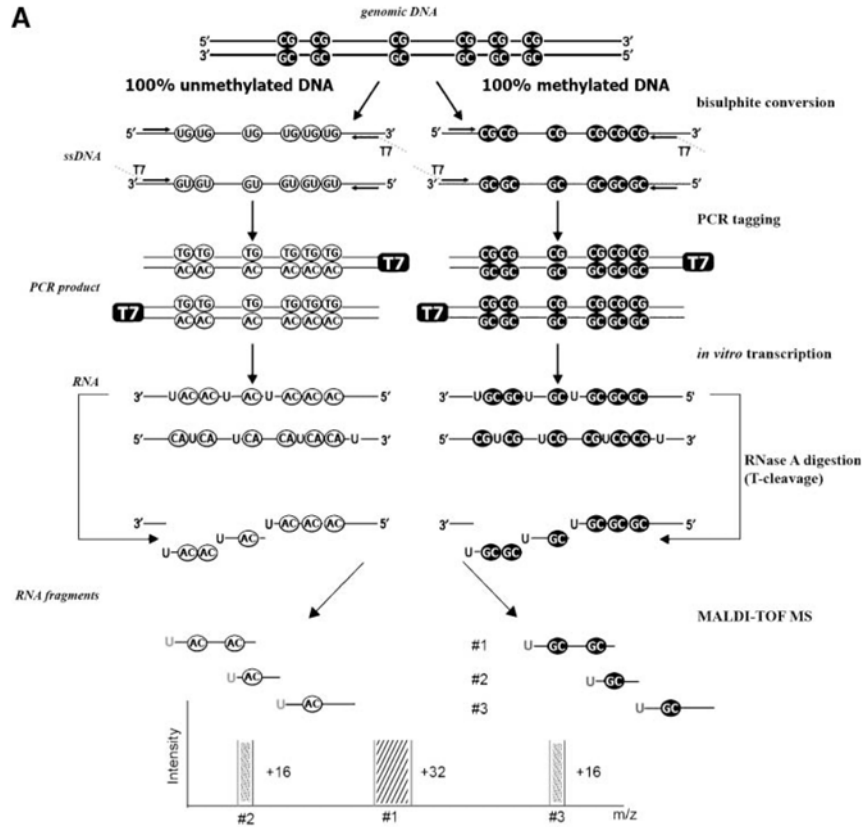
FIGURE 1. A. MALDI-TOF MS DNA methylation analysis. Overview of the MassCLEAVE assay. Genomic DNA is bisulfite treated and PCR tagged to include the T7-promoter sequence. As shown, either top or bottom strand can be used for amplification. Subsequent shrimp alkaline phosphatase treatment, *in vitro* transcription using T7R&DNA polymerase and a specific nucleotide mixture plus RNase A cleavage, results in specific fragmentation. The obtained mixture of fragments can be analyzed by MALDI-TOF-MS. **B.** In our newly developed method, we eliminated the C-cleavage, and by using T-cleavage we achieved the same result. In this method, spectra 1 and 3, which were related to C-cleavage, were deleted; spectrum 2 was related to the T-cleavage of unmethylated DNA; and spectrum 4 was related to the T-cleavage of methylated DNA.
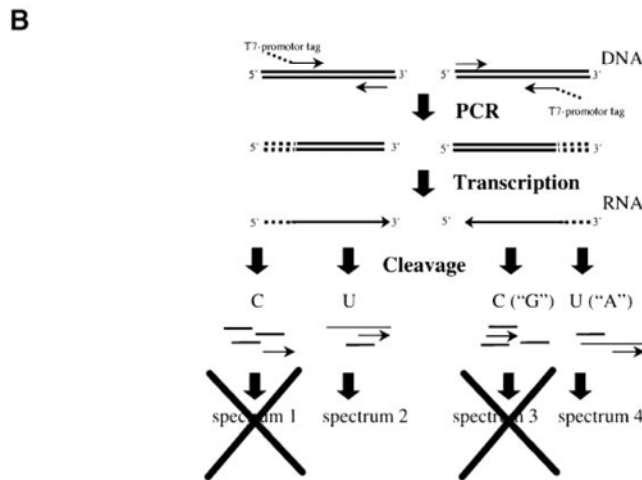
Table 1. Mechanism of T-Specific Cleavage for Methylated and Unmethylated DNA Sequences

T-Specific Cleavage Reaction	Methylated Sequence (TAACGATACGT)	Unmethylated Sequence (TAATGATATGT)	Type of Change
Forward cleavage	T-AACGAT-ACGT	T-AAT-GAT-AT-GT	Removal of cleaved nucleotide
Reverse cleavage	ACGT-AT-CGT-T-A	AT-GT-AT-T-GT-T-A	Mass shift

NOTE: Theoretical cleavage products of sequence TAACGATACGT, which will be converted through bisulfite treatment. Cleavage products are shown as DNA species. In actual practice, the DNA sequence is converted to a mixture of ribonucleotides and deoxynucleotides by transcription.

Results

Assay Development

The methylation quantification methodology presented here is schematically outlined in Fig. 1. In essence, the thymidine-specific reactions were realized by enzymatic cleavage of transcripts in which one specific base is present in the 2′-deoxy form (Table 1). In practice, not all spectral changes could be observed, and some fragments that were either too small or too large to be detected fall outside the utilizable mass window. Information may also be lost in the high-mass region. The length of the fragment may also influence the result of detection. We extended the utilizable mass range from 1,000 to <11,000 Da to increase the accuracy of detection. In the region <1,000 Da, we considered peaks detectable only when separated by ≥5 Da. Fragments that fall within this window and range in size from 4-mer to <30-mer can be calculated to cover ~76% of the target sequence. The use of a reference genomic sequence allows the resolution of remaining ambiguities spectra after reconstruction of sequence candidates from the cleavage pattern.

Accuracy of the Approach

We performed a number of experiments to confirm the accuracy of our modified T-cleavage assay for quantifying the methylation rate of the four tumor suppressor genes in breast cancer. In our design, a completely unmethylated template results in an RNA transcript that does not have any cleavable nucleotides and, hence, does not generate any cleavage products. A fully methylated template will result in an RNA transcript that contains several cleavable nucleotides and, thus, generate defined cleavage patterns. This distinct difference allows rapid and sensitive discovery of methylation sites.

We used the unique mass signals representing methylation events to test the specificity for detecting methylated DNA in mixed components. We mixed the fully methylated DNA of our designed positive control with the pure unmethylated DNA of our designed negative control using the sequence of tumor suppressor gene *P16* in ratios of 100:0; 50:50; 25:75; 5:95, and 0:100. Figure 2A shows the resulting spectra of the T-specific cleavage mixture on MALDI-TOF MS. The assay was able to discriminate the methylated and unmethylated components according to the designed ratios.

Sensitive Detection

To evaluate the robustness and sensitivity of the T-cleavage assay in this study, we prepared a dilution series of a DNA on the promoter region of *P16* gene, ranging from 80 to 2.5 ng DNA, to find the minimal input of target template necessary for a successful methylation analysis after *in vitro* transcription (Fig. 2B). The assay enabled the reliable detections of the

methylated DNA in as little as 5 ng per PCR reactions. At the lowest template input of 2.5 ng, some of the replicates failed to generate detectable fragments with distinctive signal-to-noise ratios in the spectrum.

High-Throughput Quantitative Discovery of Methylated Sites in Breast Cancer

We applied this T-cleavage assay in the analysis of breast cancer patients. We analyzed the methylation patterns of four
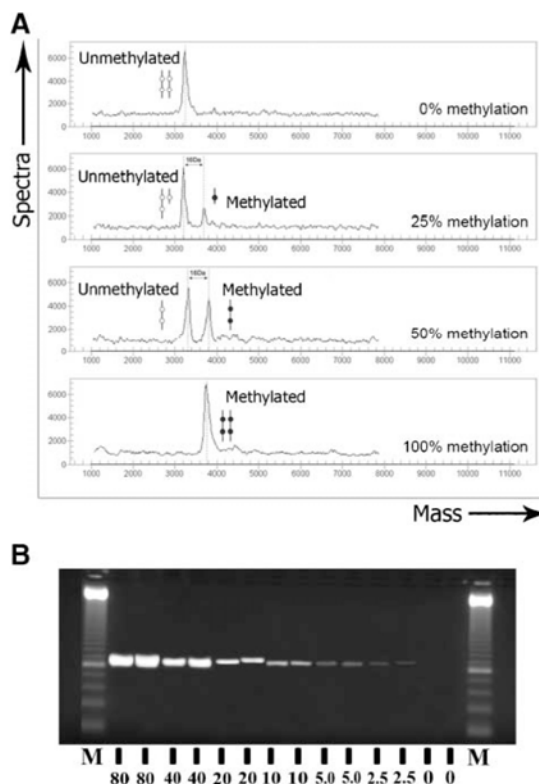
FIGURE 2. **A.** Unique mass signals representing methylation events to test the specificity of the method. For this purpose, we mixed the DNA from PCR amplification of positive and negative controls related to the tumor suppressor gene *P16* in different ratios (fully methylated template DNA, a mixture of 50:50 methylated/unmethylated DNA, 25% methylated DNA, 5% methylated DNA, and completely unmethylated DNA). **B.** Template dilution test for transcription and subsequent MALDI-TOF analysis. A 2-fold dilution series was prepared from T7-tagged PCR amplifications of *P16* CpG island ranging from 80 to 2.5 ng, and 8 μL of each dilution were visualized on 2.5% agarose gel.

Table 2.   High-Throughput Methylation Analysis of Informative CpG Sites in 48 Breast Cancerous Samples

Gene	Amplicon Size (bp)	Total No. of CpG Sites in Amplicon	No. of Analyzed CpG Sites in Amplicon	No. of Analyzed CpG Sites in 48 Samples			% of CpG sites Missed Due to	
				Total Sites	Single Sites	Combined Sites	Adjacent peak	Out of range
Cadherin1 (CDH1)	422	14	9	864	672	192	0	23
P16 (CDKN2A)	459	27	20	1,920	1,152	768	1	29
RAR-b	445	15	10	960	672	288	3	17
RASSF1	382	41	32	3,072	1,152	1,920	2	29

NOTE: The *in silico* digestion was done for the T-cleavage assay. The percentage of informative CpG sites (total sites) in the amplicon is divided into single CpG sites (single site) and combined sites when adjacent CpG sites fall within one fragment or when fragment masses are overlapping. The percentage of CpG sites that give no methylation information due to interfering neighboring peaks or partially overlapping peaks (adjacent peak) or that will be missed because they fall outside of the usable spectral range (out of range) is shown.

tumor suppressor genes (*Cadherin1, P16, RAR-b*, and *RASSF1*) in 96 cancerous and normal breast tissue samples from 48 patients with breast cancer. All of the CpG sites were analyzed by our T-cleavage assay using MALDI-TOF MS. The analyzed regions of the four tumor suppressor genes contained 71 CpG sites (total of 6,816 sites in 96 analyzed samples). The T-cleavage assay was able to detect >71% of CpG sits in amplicon (64% in *CDH1*, 74% in *P16*, 67% in *RAR-b*, and 78% in *RASSF1*; Table 2; Fig. 3). Nearly 44% of CpG sites were methylated at a very low degree, with average methylation <30%, and only 18% CpG units had mean methylation levels >90%. We performed two-way hierarchical clustering of the CpG unit methylation and the combined tumor and normal tissues in the training set to explore any natural groupings in this data set (Fig. 3). Using the cluster analysis, whereas high degrees of methylation in *P16* and *RASSF1* genes have been observed in cancerous tissues compared with paired normal tissues ($P < 0.001$), no differences of methylation patterns between cancerous and normal tissues could be found in the *CDH1* and *RAR-b* genes ($P > 0.05$).

Discussion

In this study, we present a quantitative high-throughput assay with reduced costs for methylation analysis. On the MALDI-TOF MS, mass spectrum information can be used to determine methylated and unmethylated DNAs to assess the degree of methylation for each CpG island independently, and to estimate the average methylation for the entire target region (5, 6, 13, 14). Using this method, both hypermethylation and hypomethylation can be detected in samples.

Recently, the company developed a product that showed robustness of (C/T) cleavage with the MassARRAY system for methylation analysis based on a study assessing a set of >400 candidate genes in 59 different cancer cell lines (6). In the MassCLEAVE kit (Sequenom, Inc.), the T-cleavage reagent contains dCTP, rUTP, rGTP, and rATP, and the C-cleavage reagent contains a mixture of dTTP, rCTP, rGTP, and rATP, resulting in a unique cleavage 3′ of rCTP or rUTP only, respectively. In EpiTYPER two cleavage reactions, the reverse strand is cleaved by RNase A at specific bases [uracil (U) or cytosine (C)]. Cleavage products are generated for the reverse transcription reactions for both U [thymidine (T)] and C in separate reactions. In the C-cleavage reaction, methylated

regions are cleaved at every C to create fragments containing at least one methylated CpG site. For the T-cleavage reaction, both methylated and unmethylated regions are cleaved at every T to produce fragments containing methylated and unmethylated CpG sites. For this reason, the C-cleavage will not be as informative for CpG-rich DNA, such as CpG islands. Therefore, only the T-cleavage assay was used in this study. This assay generates a defined experimental mass signal pattern wherein each mass signal represents at least one fragment of CpG site evolved from the target sequence. For analysis, this experimental pattern is subsequently compared with an *in silico* reference mass signal pattern derived from a reference genomic sequence (5). Differences between the expected and the observed mass signal pattern were interpreted, enabling quantitative analysis of methylated and unmethylated sequences. Finally, the data were compared with the original sequence. By using T-cleavage assay, the cost of this experiment can be reduced by half.

In our study, we assessed the accuracy of the T-cleavage assay by using positive (fully methylated DNA) and negative (unmethylated DNA) controls. The sensitivity and specificity of the assay were compared and confirmed by bisulfite sequencing analysis. The T-cleavage assay provides a rapid, automated discovery of multiple methylated CpG sites in regions of 200 to 600 bp and does not require cloning of PCR products for sequencing analysis.

Using the high-throughput robust T-cleavage assay on MALDI-TOF MS, we quantitatively analyzed methylation patterns of four tumor suppressor genes in 96 breast cancerous and adjacent normal tissues from patients with breast cancer. Hierarchical clustering identified substantial differences in the quantitative methylation profiling of tumor tissue compared with adjacent normal tissue (Fig. 1). We identified CpG units from the promoter regions of two genes (*P16* and *RASSF1*) that exhibited significantly different levels of methylation ($P < 0.001$) between normal and tumor samples. Our analysis shows that *RASSF1* is highly methylated in breast cancer tissue, and most significant methylation differences in this study were observed in genomic regions with relatively highly CpG density. Our data suggest that the high frequencies of methylation on the two tumor suppressor genes (*P16* and *RASSF1*) in breast cancer tissues may serve as biomarkers for the disease.

In this study, the large-scale analysis is the first high-throughput implementation of this method for quantification
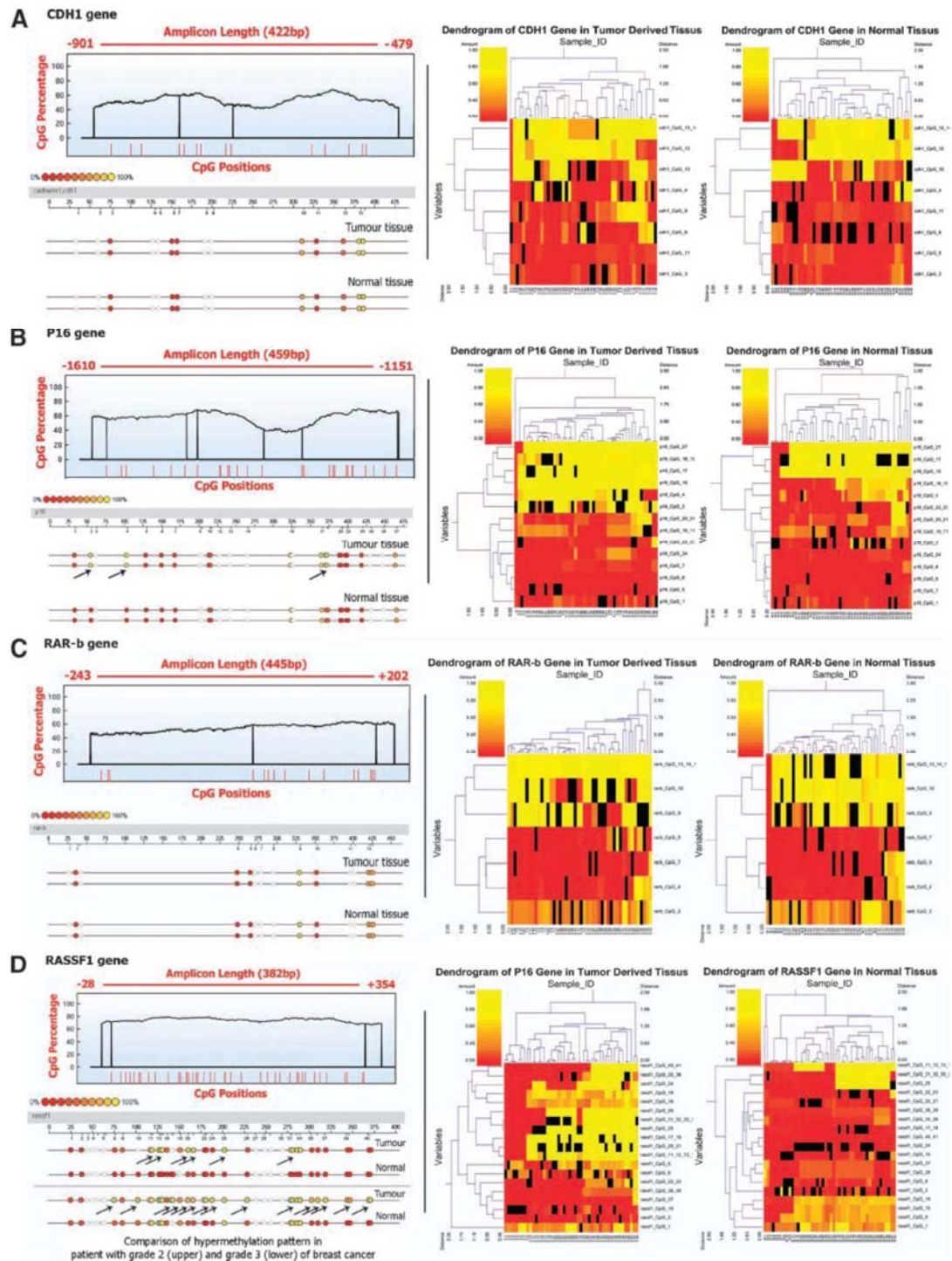
FIGURE 3. High-throughput analysis of informative CpG sites for the four studied tumor suppressor genes (**A-D**). The color of circles is related to percent of methylation in each CpG site. Arrows show the different methylation pattern between cancerous tissues and adjacent normal tissues (*left pictures*). Right pictures are two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 normal breast tissues.

A

1) 5' - cag**taatacgactcactatagggaga**aggct + gene specific sequence - 3'

                    T7-promoter tag

                    8 base pair insert (for prevention of
                    abortive cycling and constant 5'-
                    fragment for RNase A reaction)

2) 5' - aggaagagag + gene specific sequence - 3'

        10mer tag sequence

B

C

```
TTGTAAGGTTTGGGGGGTTTGACGTTTGATTTGTAAGGTTTTGTCGGTTTACGTT
GTGGTTTTATTTTTTTAATTAGGTTATTAGGTTTAAGCGTATGAATAGGAAGTTAT
TGAGAAGCGGGTTATATTTTTAGGTTTTCGGAGTTGGGTTTAGTTTCGAATTTTT
TGGTATATATTTATTTATTTAGGTCGCGGGTTTAGTTCGCGAGGTTTAGGACGGAT
TTAGGTAGACGGATAGAGTAATGAGATGATTTCGTTTTTTTTTTTTTTTTTTTTATT
TTTAAATAATTTAGTTTGAAGAATGGAAGATTTTCGACGAGGGGAGTTAGGAAT
AAAATAAGGGGAATAGGGGAGCGGGGACGCGAGTAGTATTA*GAATTC*GCGGGA
GCGCGGTTGTTTTTGGTAGGGTCGTGTTAGGTGACGGATGTAGTTAGGGGGCGA
GTTGTTTGGAGTTGCGTTT
```

D

```
TTGTAAGGTTTGGGGGGTTTGATGTTTGATTTGTAAGGTTTTGTTGGTTTATGTT
GTGGTTTTATTTTTTTAATTAGGTTATTAGGTTTAAGTGTATGAATAGGAAGTTATT
GAGAAGTGGGTTATATTTTTAGGTTTTGGAGTTGGGTTTAGTTTTGAATTTTTTG
GTATATATTTATTTATTTAGGTTGTGGGTTTAGTTTGTGAGGTTTAGGATGGATTTA
GGTAGATGGATAGAGTAATGAGATGATTTTGTTTTTTTTTTTTTTTTTTTTATTTTT
AAATAATTTAGTTTGAAGAATGGAAGATTTTGATGAGGGGAGTTAGGAATAAA
ATAAGGGGAATAGGGGAGTGGGGATGTGAGTAGTATTAGAATTTGTGGGAGTGT
GGTTGTTTTTGGTAGGGTTGTGTTAGGTGATGGATGTAGTTAGGGGGTGAGTTG
TTTGGAGTTGTGTTT
```

E

FIGURE 4.   **A.** Primers for *in vitro* transcription. **A**-1, reverse primer with T7-promoter tag. **A**-2, forward primer with 10-mer tag sequence as balance. **B.** PCR product cloning of *P16* tumor suppressor gene into T-vector as positive or negative control. **C.** Bisulfite-treated sequence after treatment with methyltransferase (M.SssI) enzyme as positive control. **D.** Bisulfite-treated sequence after first round of PCR as negative control. **E.** *Eco*RI restriction enzyme digested a 459-bp DNA amplified from *P16* tumor suppressor gene after bisulfite conversion. Lane 1, DNA treated with Bisulfite Kit and cytosine converted to thymine; therefore, it made a recognition site for *Eco*RI at position 379. Lane 2, in nontreated DNA, there was no recognition site for enzyme.

of methylation in breast cancer. We were able to conclude that T-specific assay in combination with MALDI-TOF MS is a sensitive, accurate, and reliable technique for cost-effective high-throughput methylation analysis. The T-cleavage assay is readily amenable to automation. Our results support the suitability and robustness of T-cleavage assay on MALDI-TOF MS for cancer studies as well as for understanding methylation events in life science.

Materials and Methods
Samples

The study was approved by the local institutional review board. DNA samples from human hepatocyte cells were used to develop a methylation-positive control and human genomic DNA samples were used to develop a methylation-negative control for the assessment of the sensitivity and specificity of the approach.

Samples were collected from 48 paraffin-embedded cancerous breast tissues and paired adjacent normal tissues. Paraffin-embedded sections from adjacent normal and cancerous breast tissue samples were examined by two experienced pathologists. DNA was extracted from three to five sections of each 10-μm-thick paraffin-embedded sample (around 0.01-0.02 g of tissue) using a High Pure PCR Template Preparation Kit (Roche Diagnostics) and eluted into 150 μL of elution buffer. The eluted DNA was stored at −20°C until further use.

Bisulfite Treatment

To perform bisulfite conversion of target sequence, the EZ-96 DNA Methylation Kit (Zymo Research) was used according to the instruction manual. During CT conversion reaction, the PCR program was used as follows: 95°C for 30 s and 50°C for 15 min, repeated for 46 cycles.

Primer Design and PCR Tagging for EpiTYPER Assay

CpG density and CpG sites of target sequences of the four tumor suppressor genes were analyzed with extreme precision for the PCR primer design. We designed primers using MethPrimer (15). For sequences that exceeded 450 bp in

length, we randomly assigned a 450-bp stretch within the annotated CpG islands. In PCR amplification, a T7-promoter tag was added to the reverse primer and a 10-mer tag sequence to the forward primer to balance the PCR primer length (Fig. 4A). The primer sequences, annealing temperatures (T_a), and PCR conditions are described in Supplementary data 2.

Synthesized Fully Methylated DNA as Positive Control and Unmethylated DNA as Negative Control

(*a*) To set up positive control, DNA sample from human hepatocyte cells was treated with the CpG methyltransferase (M.SssI) enzyme. M.SssI was isolated from a strain of *E. coli*, which contains the methyltransferase gene from *Spiroplasma* sp. strain MQ1 (New England Biolabs) and methylated all cytosine residues (C5) within the double-stranded dinucleotide recognition sequence 5′...CG...3′. S-Adenosylmethionine was used as source of methyl groups. After treatment with M.SssI, DNA was treated with bisulfite conversion.

(*b*) To set up negative controls, we amplified genomic DNA using selected primer pairs before bisufite conversion (because there was no methylase enzyme in the PCR reaction). Afterward, PCR products were converted by the bisulfite treatment, which changed whole cytosines into thymidines.

The sequence accuracy of the designed positive control was conformed by direct bisulfite sequencing, and the sequence accuracy of the negative controls was checked by bisulfite sequencing after cloning PCR products into T-vector. Clones were sequenced from the pooled PCR reactions using the Dye Terminator cycle sequencing kit with AmpliTaq DNA polymerase FS (Applied Biosystems) and the automated 373A NA Sequencer (Applied Biosystems; Fig. 4).

Dephosphorylation of Unincorporated Deoxynucleotide Triphosphates by Shrimp Alkaline Phosphatase

Following the PCR amplification, 5 μL of PCR products were mixed with 2 μL of shrimp alkaline phosphatase enzyme (Sequenom) to dephosphorylate unincorporated deoxynucleotide triphosphates from the PCR reactions. The mixtures were incubated at 37°C for 20 min and then at 85°C for 5 min to inactivate the shrimp alkaline phosphatase enzyme.

In vitro Transcription and RNase A Cleavage

In the first step, transcription/RNase A cocktail for T-cleavage reaction was prepared using MassCLIVE (hMC) kit (Sequenom) according to the instruction manual. After shrimp alkaline phosphatase treatment, 2.5 μL of the mixture were directly added to a 5 μL transcription cocktail, containing 20 units of T7 R&DNA polymerase (Epicentre) to incorporate dTTP in the transcripts. Ribonucleotides were used at 1 mmol/L and the deoxynucleotide triphosphate substrate at 2.5 mmol/L; other components in the reaction were as recommended by the supplier. Following the *in vitro* transcription, RNase was added to cleave the *in vitro* transcripts. In the cleavage reaction, the reverse strand was cleaved by RNase A at specific bases (U).

The mixture of cleaved fragments was further diluted with double-distilled water to a final volume of 27 μL. The phosphate backbone, before MALDI-TOF MS, was cleaned by adding 6 mg CLEAN Resin (Sequenom).

Mass Spectrometry

Twenty-two nanoliters of cleavage reaction were robotically dispensed (nanodispenser) onto silicon chips preloaded with matrix (SpectroCHIP, Sequenom). Mass spectra were collected using a MassARRAY mass spectrometer (Sequenom). Depending on the sequence of the target region and the distribution of CpGs, the mass spectrum may contain multiple signal pairs of cleavage products. These signal pairs can be used to estimate the ratio of methylated to unmethylated DNA. Finally, the results and spectra were analyzed with EpiTYPER v.1.0 software.

Statistical Methods

The two-way hierarchical cluster analysis clustered the 96 tissue samples (48 tumor derived and 48 paired normal tissues) and most variable CpG fragments for each gene based on pairwise Euclidean distances and the complete linkage clustering algorithm. The method involves first establishing the strength of the connection between two samples (called distance), followed by the reorganization of the samples according to their relationship to each other. The algorithm "clusters" samples divided into same clusters according to their similarity, and the double dendrogram was used to visualize the results. The method presented in this article clustered samples along the *x* axis and CpG units along the *y* axis. The procedure was carried out using the double dendrogram function of the Gene Expression Statistical System for Microarrays (GESS) version 7.1.5 (NCSS).

Additional Information

The list and complete sequences of CpG islands in selected tumor suppressor genes, the bisulfite converted sequences, the sequences of PCR-tagged primers for *in vitro* transcription, and the PCR conditions are available in supplementary data.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22:2990–7.

2. Fraga MF, Esteller M. DNA methylation: a profile of methods and applications. Biotechniques 2002;634:636–49.

3. Laird PW. The power and the promise of DNA methylation markers. Nat Rev 2003;3:253–66.

4. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. Nat Protoc 2006;1:2353–64.

5. Ehrich M, Nelson MR, Stanssens P, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 2005;102:15785–90.

6. Ehrich M, Turner J, Gibbs P, et al. Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci U S A 2008;105:4844–9.

7. Ehrich M, Böcker S, van den Boom D. Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. Nucleic Acids Res 2005;33:e38.

8. Stanssens P, Zabeau M, Meersseman G, et al. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome Res 2004;14:126–33.

9. Sousa R, Padilla R. A mutant T7 RNA polymerase as a DNA polymerase. EMBO J 1995;14:4609–21.

10. Rodi CP, Darnhofer-Patel B, Stanssens P, Zabeau M, van den Boom, D. A strategy for the rapid discovery of disease markers using the MassARRAY system. Biotechniques 2002;32:S62–9.

11. Hartmer R, Storm N, Boecker S, et al. RNase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis. Nucleic Acids Res 2003;31:e47.

12. Krebs S, Medugorac I, Seichter D, Förster M. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. Nucleic Acids Res 2003;31:e37.

13. Schatz P, Dietrich D, Schuster M. Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. Nucleic Acids Res 2004;32:e167.

14. Gut IG. DNA analysis by MALDI-TOF mass spectrometry. Hum Mutat 2004;23:437–41.

15. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics 2002;18:1427–31.

5. Published research article:

Simultaneous isolation of DNA, RNA and proteins for genetic, epigenetic, transcriptomic and proteomic analysis

Summary:

Analysis of DNA, RNA and proteins for downstream genetic, epigenetic, transcriptomic and proteomic analysis is often hampered by the limited availability of sample material. We present a methodological approach using the Qiagen AllPrep method that allows efficient isolation and purification of DNA, RNA and proteins from a single sample aliquot of cell lines, whole blood, buffy coat, serum, plasma and formalin-fixed paraffin-embedded tissues. The efficiency, quality, quantity and purity of the samples obtained by this strategy were confirmed by several genetic, epigenetic, transcriptomic and proteomic methods. This approach, when compared to the procedures based on separate extraction of samples, is convenient both in terms of its simplicity and cost-effectiveness.

First author s' contribution:

Ramin Radpour was involved in study design, making collaborations, data analysis and writing the manuscript.

Michal Sikora was involved in performing the experiment and data analysis.

Simultaneous Isolation of DNA, RNA, and Proteins for Genetic, Epigenetic, Transcriptomic, and Proteomic Analysis

Ramin Radpour,[†,#] Michal Sikora,[†,#] Thomas Grussenmeyer,[‡] Corina Kohler,[†] Zeinab Barekati,[†] Wolfgang Holzgreve,[§] Ivan Lefkovits,*,[‡] and Xiao Yan Zhong*,[†]

Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel, Switzerland, Department of Biomedicine and Department of Cardiac Surgery, University Hospital Basel, Switzerland, and University Medical Center Freiburg, Germany

Analysis of DNA, RNA, and proteins for downstream genetic, epigenetic, transcriptomic, and proteomic analysis holds an important place in the field of medical care and life science. This is often hampered by the limited availability of sample material. For this reason, there exists an increasing interest for simultaneous isolation of DNA, RNA and proteins from a single sample aliquot. Several kit-systems allowing such a procedure have been introduced to the market. We present an approach using the AllPrep method for simultaneous isolation of DNA, RNA and proteins from several human specimens, such as whole blood, buffy coat, serum, plasma and tissue samples. The quantification and qualification of the isolated molecular species were assessed by different downstream methods: NanoDrop for measuring concentration and purity of all molecular species; DNA and RNA LabChip for fractionation analysis of nucleic acids; quantitative PCR for quantification analysis of DNA and RNA; thymidine-specific cleavage mass array on MALDI-TOF silico-chip for epigenetic analysis; Protein LabChip and two-dimensional (2D) gel electrophoresis for proteomic analysis. With our modified method, we can simultaneously isolate DNA, RNA and/or proteins from one single sample aliquot. We could overcome to some method limitations like low quality or DNA fragmentation using reamplification strategy for performing high-throughput downstream assays. Fast and easy performance of the procedure makes this method interesting for all fields of downstream analysis, especially when using limited sample resources. The cost-effectiveness of the procedure when material is abundantly available has not been addressed. This methodological improvement enables to execute such experiments that were not performable with standard procedure, and ensures reproducible outcome.

Keywords: Column-based extraction • AllPrep method • genetic analysis • transcriptomic analysis • proteomic analysis • limited sample resources • MALDI-TOF mass spectrometry

Introduction

Genetic and proteomic analysis using limited sample resources is encountered in many fields of life science, while insufficient amounts of acquired molecular species of DNA, RNA, and proteins lead often to difficulties in obtaining conclusive results.[1,2] As an example in the field of cancer research, fine needle aspiration biopsies are minor surgical procedures which provide minute amounts of patient's specimens for diagnosis (genetic, epigenetic, expression, and proteomic analysis).[3] These biopsy samples yield in most instances amounts of material that do not allow to perform several

extractions, and therefore, DNA, RNA, and proteins are often isolated from different tissue sections or biopsies. Such a procedure makes it difficult to correctly evaluate genomic or epigenetic changes, which in turn makes their correlation with gene expression or proteome changes problematic and insecure. A method allowing simultaneous isolation of DNA, RNA, and proteins from the same population of cells is a prerequisite for overcoming such a limitation.

Methods for simultaneous isolation of DNA and RNA have been developed using TRIzol[4] or guanidinium-based reagent,[5,6] while the combined extraction of RNA and proteins using RNA spin column-based technology has been introduced by Morse and co-workers.[7] An improved method for isolation and solubilization of proteins after TRIzol extraction of RNA and DNA has been described by Hummon and colleagues.[8] Recently, column-based Qiagen AllPrep method has been developed for rapid simultaneous extraction of DNA, RNA, and proteins from the tissues and cultured cells.[9] To our knowledge, the Qiagen AllPrep method has not been tested for whole blood,

* To whom correspondence should be addressed. Prof. Dr. Xiao Yan Zhong, Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel. Tel, +41 61 265 9224; fax, +41 61 265 9399; e-mail, xzhong@uhbs.ch. Prof. Ivan Lefkovits, Department of Biomedicine, University Hospital Basel, Vesalianum, Vesalgasse 1, Switzerland. Tel, 004161 2673551; e-mail, ivan.lefkovits@unibas.ch.
† University of Basel.
# These authors contributed equally to this work.
‡ University Hospital Basel.
§ University Medical Center Freiburg.

Simultaneous Isolation of DNA, RNA, and Proteins   research articles

buffy coat, serum, plasma, and formalin-fixed/paraffin-embedded tissues samples.

Here, we describe our experience in applying the methodology of simultaneous isolation of DNA, RNA, and/or proteins for quantitative and qualitative analysis of breast cancerous cell lines, whole blood, buffy coat, serum, plasma, and formalin-fixed/paraffin-embedded tissues for genetic, epigenetic, transcriptomic, and proteomic profiling.

Materials and Methods

Samples. This study was performed in the laboratory of Prenatal Medicine and Gynecologic Oncology at the University Women's Hospital and Department of BioMedicine, University of Basel, Switzerland. A total of six breast cancer cell lines (MDA-MB-231, MCF-7, HS578T, BT549, T47D and SKBR3), six whole blood samples from healthy unrelated persons, six separated buffy coat, plasma, and serum samples from the same persons, and six formalin-fixed/paraffin-embedded cancerous breast tissues were subjected to this study. Use of the human samples has been approved by local ethical institutional review boards.

Cell Lines and Culture Conditions. MDA-MB-231, MCF-7, and HS578T were grown in Dulbecco's Modified Eagle Medium (DMEM; high glucose with L-glutamine). BT549 and T47D were cultured in RPMI 1640 medium, and SKBR3 was grown in McCoy's 5A. All media were supplemented with 10% fecal calf serum (FCS) and 1% penicillin—streptomycin. The cells were maintained in a humid incubator at 37 °C with 5% CO_2. Before using the cells for extraction, they were washed with phosphate buffered saline (PBS).

Simultaneous Isolation of DNA, RNA, and/or Proteins from a Single Sample Aliquot. The extraction was performed with the AllPrep DNA/RNA/Protein Mini Kit (QIAGEN AG, Basel, Switzerland) according to a modified manufacturer's protocols. For extraction from cell lines, 5×10^6 cells of each cell type were used, but harvesting was carried out using cell scraper instead of trypsinization. For extraction from whole blood, buffy coat, serum, and plasma, 500 μL of samples was disrupted by adding 600 μL of Buffer RLT (AllPrep kit). The RLT buffer was supplemented with β-mercaptoethanol as detailed in the manufacturer's instruction. For the cell lines, whole blood, and buffy coat, the homogenization was performed on ice with a blunt 20-gauge needle (0.9 mm diameter) fitted to an RNase-free syringe. Extraction from paraffin-embedded sections was performed from three to five sections of each 20 μm thick paraffin-embedded sample (around 0.02–0.04 g of tissue). For the deparaffinization of the samples, we treated them with 1000 μL of X-TRA-Solv (Medite Medizintechnik, Germany) and incubated for 1 h at 50 °C followed by washing with aqueous lower alcoholic solutions of decreasing concentration (100, 70, and 40% ethanol followed by washing with RNase-free water). Deparaffinized tissues were disrupted in 800 μL of RLT buffer and homogenized on ice for 1 min at maximum power using Polytron PT 1200E device (Kinematica AG, Switzerland). Further steps were carried out according to the AllPrep's instruction manual.

Following extractions, the absorbance at 260 and 280 nm was measured using a NanoDrop ND-1000 spectrophotometer (Biolab, Mulgrave, VIC, Australia), and the ratios were calculated. Size fractionation of DNA was evaluated by gel electrophoresis on a 1% agarose gel (Invitrogen) containing 0.5 mg/L

ethidium bromide (Sigma). Electrophoresis was carried out at 80 V for 1 h.

Quantity and Size Fractionation Analysis of Species Using LabChips. The aim of lab-on-a-chip or microfluidics technology is to shrink processes to very small dimensions, thus, allowing to handle very small sample volumes and to use very small amounts of reagents.

DNA: DNA 12000 LabChip Kit (Agilent Technologies, GmbH, Waldbronn, Germany) was used in this study to analyze the quantity and size fractionation. Briefly, microchannels were filled by pipetting 9.0 μL of the gel-dye mixture into the appropriate well and then forced the mixture into the microchannels by applying pressure to the well via a 1-mL syringe. The ladder well and each of the 12 sample wells were subsequently loaded with 5 μL of the DNA size marker mixture and 1 μL of either the molecular size ladder or sample. After vortexing, the chips were immediately inserted into the Agilent 2100 Bioanalyzer system and processed.

RNA: Total RNA quantity and quality were also determined by capillary chip electrophoresis using the RNA 6000 Pico LabChip (Agilent Technologies, GmbH, Waldbronn, Germany). The chip generated a quantitative measure of the amount of RNA loaded into the wells. According to the instructions provided in the kit, 9.0 μL of gel-dye mix was added into the marked well and dispensed by pressurizing (using the chip priming station) for 30 s. Finally, 5 μL of RNA 6000 Pico marker plus 1 μL of samples were added in each of the 11 sample wells. The chip was run in Agilent 2100 Bioanalyzer.

Protein: The molecular weight, as well as the concentration of proteins, was determined using the Agilent 2001 Bioanalyzer system (Agilent, Waldbronn, Germany) in combination with the Protein 230 LabChip kit (proteins between 14 and 230 kDa can be analyzed) and the Protein 80 LabChip (proteins between 5 and 80 kDa can be analyzed). The Protein pellets were solubilized in buffer containing 125 mM Tris, pH 7.2, and 2% SDS instead of using ALO buffer (supplied by Qiagen AllPrep kit). All the chips were prepared according to the protocol provided with the kits. Briefly, samples (4 μL) were diluted in sample buffer. Sample buffer included an upper marker that can be used for semiquantitative analysis. Separated proteins were detected by laser-induced fluorescence. The software automatically generated a standard curve of the ladder protein versus their known molecular weights, which was used to determine the size of the unknown proteins. To determine the relative concentration, the peak area of the unknown sample was compared to the peak area of the upper marker with known concentration. In order to check the efficiency of the extraction using AllPrep kit, the human serum albumin standard (code: K809AGL from Agilent) was used as control serum during the experiment.

Quantity assessment of extrcated DNA and RNA Using Quantitative PCR. The yield of DNA and RNA (after making cDNA using the QuantiTect Reverse Transcription kit, QIAGEN AG, Basel, Switzerland) was measured by quantitative PCR (Q-PCR). Real-time PCR was performed using ABI Prism 7000 Sequence Detector (Applied Biosystems). Each sample was amplified for glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*). The sequence of primers and probe for quantification of *GAPDH* in DNA and RNA samples and the Q-PCR conditions are summarized in Supplementary Data 1. Each sample was measured in duplicate. Standard and dissociation curves were generated with the ABI Prism 7000 SDS software. R^2 for each standard curve was >0.98.

Table 1. Quantity and Purity of Extracted DNA, RNA, and Proteins in Different Sample Types[a]

samples				
sample type (amount of input)	quantity and purity	DNA median (range)	RNA median (range)	proteins median (range)
Cell Line (5×10^6 cells)	Quantity	243.5 ng/μL (137.3 to 528.4)	1145.9 ng/μL (473.2 to 1751)	4.1 mg/mL (2.8 to 7.9)
	A 260/280	1.9 (1.78 to 2.1)	2.2 (2.1 to 2.3)	3.6 (−2.9 to 7.8)
	A 260/230	1.7 (0.9 to 2.1)	1.8 (0.9 to 2.1)	-
Whole Blood (500 μL)	Quantity	109.6 ng/μL (43.4 to 439.1)	29.1 ng/μL (24.3 to 42.1)	1.8 mg/mL (1.2 to 2.5)
	A 260/280	1.9 (1.7 to 2.0)	2.1 (1.9 to 2.5)	3.5 (1.2 to 3.8)
	A 260/230	0.5 (0.1 to 1.5)	0.2 (0.02 to 1.5)	-
Buffy Coat (500 μL)	Quantity	43.5 ng/μL (14.1 to 54.3)	20.2 ng/μL (7.8 to 32.1)	2.1 mg/mL (0.2 to 3.8)
	A 260/280	1.7 (1.7 to 2.1)	2.2 (1.7 to 3.2)	2.3 (−1.2 to 4.4)
	A 260/230	0.1 (0.02 to 0.58)	0.1 (0.02 to 0.2)	-
Serum (500 μL)	Quantity	6.1 ng/μL (4.1 to 6.8)	5.4 ng/μL (2.6 to 16.7)	5.7 mg/mL (4.2 to 7.5)
	A 260/280	1.4 (1.2 to 1.9)	1.9 (1.1 to 2.3)	4.5 (1.5 to 5.7)
	A 260/230	0.02 (0.01 to 0.04)	0.01 (0.01 to 0.2)	-
Plasma (500 μL)	Quantity	3.2 ng/μL (2.6 to 4.6)	9.8 ng/μL (5.6 to 13.5)	3.9 mg/mL (0.2 to 7.3)
	A 260/280	1.7 (1.3 to 1.8)	2.0 (1.8 to 2.8)	3.9 (0.2 to 7.3)
	A 260/230	0.01 (0.01 to 0.03)	0.3 (0.01 to 0.6)	-
Paraffin-Embedded Tissue (0.02−0.04 g)	Quantity	34.9 ng/μL (10.7 to 46.9)	8.7 ng/μL (18 to 60)	0.1 mg/mL (0.03 to 0.2)
	A 260/280	1.9 (1.6 to 2.0)	1.9 (1.7 to 2.1)	0.2 (−1.3 to 0.5)
	A 260/230	0.05 (0.03 to 0.3)	0.3 (0.1 to 0.7)	-

[a] A: Absorbance.

We examined the amplification efficiency for *GAPDH* DNA and cDNA on experimental serial dilutions using HPLC-purified oligonucleotides specifying a 97-bp *GAPDH* amplicon for DNA and 226-bp *GAPDH* amplicon for RNA with concentration ranging from 3.125×10^4 to 10 pg/μL (including 31250, 6250, 1250, 250, 50 and 10 pg/μL). The amplifications of *GAPDH* on serial dilutions showed a good correlation with comparable efficiencies. In our study, the average threshold cycle number (Ct) values of the *GAPDH* were obtained from each case.

Methylation Analysis of Extracted DNA Using Thymidine-Specific Cleavage Mass Array on MALDI-TOF Silico-Chip. The EpiTYPER assay using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and MassCLEAVE reagent was assessed for high-throughput analysis of DNA methylation patterns of four tumor suppressor genes (*BIN1*, *P16*, *RAR-b* and *RASSF1*) according to the previously published methods.[10−12]

Bisulfite Treatment: To perform bisulfite conversion of the target sequence, the Epitect Bisulfite Kit (QIAGEN AG, Basel, Switzerland) was used.

Primer Designing and PCR-Tagging for EpiTYPER Assay: CpG density and CpG sites of target sequences in four tumor suppressor genes were analyzed for the PCR primer design. We designed primers to cover the promoter regions with the most CpG sites. The primers were designed using MethPrimer.[13] In PCR amplification, a T7-promoter tag was added to the reverse primer, and a 10mer-tag sequence was added to the forward primer to balance the PCR primer length. The primer sequences, annealing temperatures (T_a) and PCR conditions are described in Supplementary Data 1.

In Vitro **Transcription, T-Cleavage Assay and Mass Spectrometry:** Unincorporated dNTPs were dephosphorylated by adding 1.7 μL of H_2O and 0.3 units of shrimp alkaline phosphatase (SAP; SEQUENOM, Inc., San Diego, CA). The reaction mixture was incubated at 37 °C for 20 min and SAP was then heat inactivated for 10 min at 85 °C. Typically, 2 μL of the PCR was directly used as a template in a 5 μL transcription reaction. Twenty units of T7 R&DNA polymerase (Epicenter, Madison, WI) were used to incorporate dTTP in the

transcripts. Ribonucleotides were used at 1 mmol/L and the dNTP substrate at 2.5 mmol/L. In the same step, the *in vitro* transcription RNase A (SEQUENOM) was added to cleave the *in vitro* transcript (T-cleavage assay). The mixture was further diluted with H_2O to a final volume of 27 μL. Twenty-two nanoliters of cleavage reaction was robotically dispensed (nanodispenser) onto silicon chips preloaded with matrix (SpectroCHIP; SEQUENOM, San Diego, CA). Mass spectra were collected using a MassARRAY Compact MALDI-TOF (SEQUENOM) and spectra's methylation ratios were generated by the Epityper software v1.0 (SEQUENOM, San Diego, CA).

Efficiency Assessment of the Protein Extraction Using the Two-Dimensional (2D) Gel Electrophoresis. In order to check the efficiency of protein extraction using the AllPrep method in comparison with a normal conventional extraction method, proteins from two different breast cancerous cell lines (BT549 and SKBR3) were subjected to 2D gel electrophoresis. As conventional method, proteins from cell lines were extracted directly using 7 M urea, 2 M thiourea, 4% CHAPS, 20 mM DTT and 2% ampholines, pH 3−10 (Invitrogen) buffer. Per 5×10^6 cells, a volume of 0.1 mL of solubilization buffer was employed. Solubilization was done for 20 min at room temperature, supported by 3 short periods of vigorous vortexing. Extracts were cleared by centrifugation at 14 000g for 10 min. The protein pellets from AllPrep method also were solubilized in the same buffer containing (7 M urea, 2 M thiourea, 4% CHAPS, 20 mM DTT and 2% ampholines, pH 3−10).

2D Gel Electrophoresis and Spot Visualization: For 2D gel electrophoresis, the ISODALT system[14,15] was used (ampholines pH 3−10 (Invitrogen)) in the first dimension; and 11−19% linear acrylamide gradient in the second dimension). In this system, up to 20 gels can be processed simultaneously. Protein spots were visualized by silver staining. This was performed by some modifications of the Vorum protocol.[16] Protein fixation was done overnight in 40% ethanol and 10% acetic acid. The development step was performed for 10 min in 6.75% sodium carbonate decahydrate and 0.005% formaldehyde. Wet silver-stained gels were scanned by Pharmacia Image Scanner with 300 dpi, 16 bit.
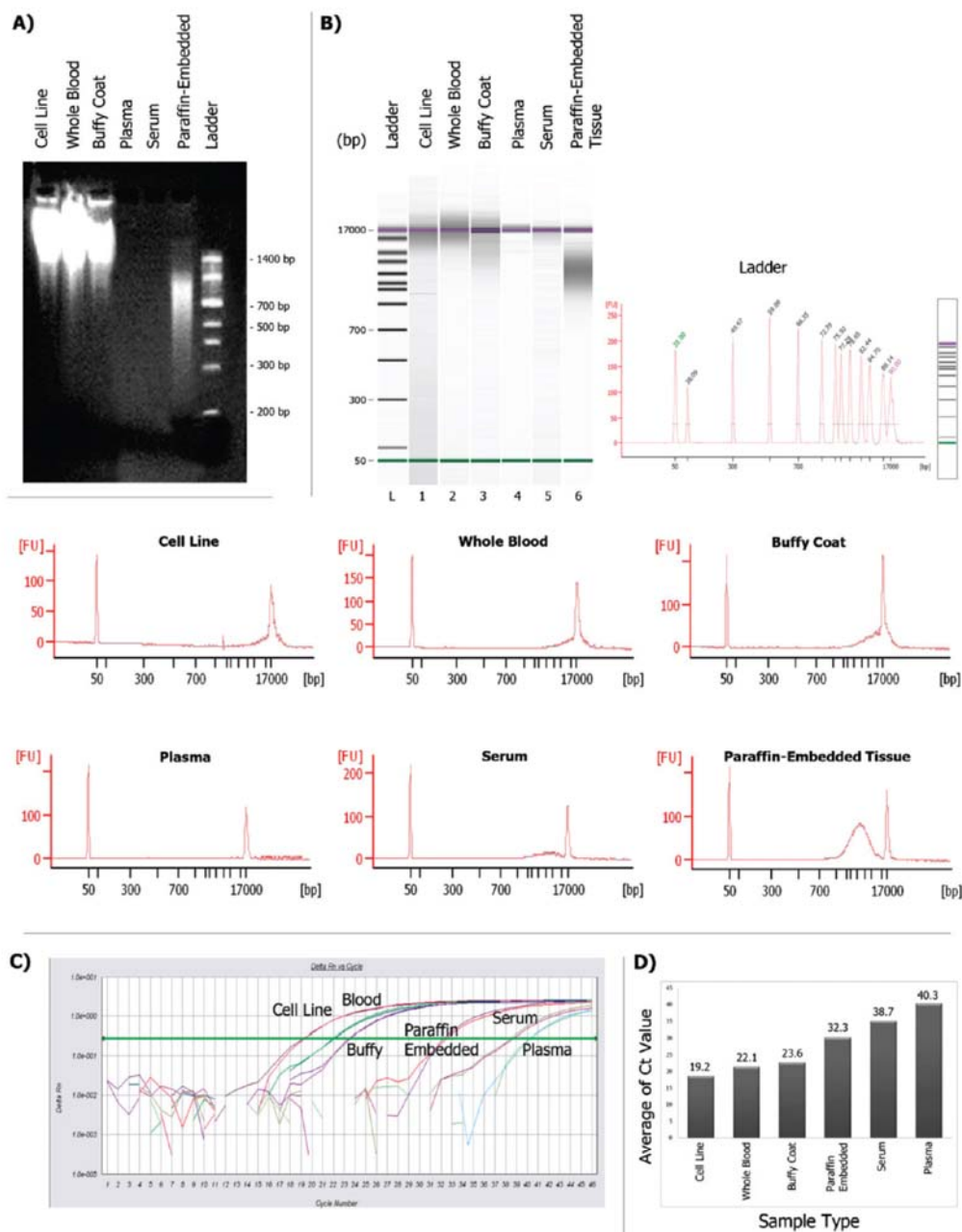
Figure 1. Quantitative and qualitative analysis of the extracted DNA. (A) Size fractionation analysis of samples using agarose gel electrophoresis. (B) Quantity and size fractionation determination of samples using DNA LabChip. (C) Quantity assessment of samples using Q-PCR. (D) Average of Ct values for different analyzed sample types.

Image Analysis and Spot Quantification: PDQuest image analysis software, version 7.2, was used for spot detection, matching and quantification. About 30−50 landmarks were usually introduced to achieve proper matching of spots between gels. Quantification results from PDQuest calculations based on Gaussian modeling of spots. No calibration for gray density values was used.

Correlation Analysis/Scatter Plot/Coefficient of Variation: The scatter plot (on log/log scale) is a convenient way to

identify quantitative relationship of compared gene products. The plot takes into consideration data from two patterns (*x*-axis for one pattern, *y*-axis for another one). Since each dot on the scatter plot is "clickable" (in the PDQuest program) and allows to inspect the actual spot (on the displayed pattern), it is rigorously used as an indispensable interactive method of proofreading to check for possible "spot mismatches".

Statistical Methods. The data analysis was performed using SPSS software (Statistical Software Package for Windows, V.17).

Table 2. High-Throughput Methylation Analysis of Informative CpG Sites Per Amplicons for Four Studied Genes in Different Sample Types[a]

sample	gene	amplicon size (bp)	total no. of CpG sites in amplicon	no. of analyzed CpG sites in amplicon		no. of analyzed CpG sites in amplicon after reamplification	
				single sites	composite sites	single sites	composite sites
Cell Line	*BIN1*	330	32	3	8	3	18
	P16 (CDKN2A)	580	62	5	5	9	20
	RAR-b	475	5	2	0	3	0
	RASSF1	346	22	2	2	4	3
Whole Blood	*BIN1*	330	32	2	10	4	18
	P16 (CDKN2A)	580	62	6	5	9	18
	RAR-b	475	5	1	0	2	0
	RASSF1	346	22	3	0	7	2
Buffy Coat	*BIN1*	330	32	1	12	3	18
	P16 (CDKN2A)	580	62	3	12	8	17
	RAR-b	475	5	2	0	3	0
	RASSF1	346	22	3	0	7	5
Serum	*BIN1*	330	32	2	4	2	13
	P16 (CDKN2A)	580	62	5	8	9	17
	RAR-b	475	5	1	0	3	0
	RASSF1	346	22	3	0	5	0
Plasma	*BIN1*	330	32	2	2	2	10
	P16 (CDKN2A)	580	62	3	5	6	13
	RAR-b	475	5	1	0	2	0
	RASSF1	346	22	2	0	5	0
Paraffin-Embedded Tissue	*BIN1*	330	32	2	8	3	15
	P16 (CDKN2A)	580	62	4	13	8	18
	RAR-b	475	5	1	0	3	0
	RASSF1	346	22	4	0	5	3

[a] The *in silico* digestion was performed for the T-cleavage assay. The percentage of informative CpG sites (total sites) in the amplicon is divided into single sites (single CpG sites) and composite sites (two or three adjacent CpG sites fall within one fragment, or when fragment masses are overlapping).

The Shapiro−Wilk test and the Kolmogorov−Smirnov test were used to analyze the normality. The two tests showed that our data set was not normally distributed ($p < 0.0001$ by Shapiro−Wilk test and $p > 0.0001$ by Kolmogorov−Smirnov). The Mann−Whitney U Test was used to compare methylation detection rate of amplification and reamplification steps. All P-values were two-sided; $p < 0.05$ were considered statistically significant.

Results

Quantitative and Qualitative Analysis of the Extracted DNA. Extracted DNA from different sample types was quantified using a NanoDrop spectrophotometer, and the 260/280 and 260/230 ratios were calculated. The median of 260/280 ratios for extracted DNA samples was shown to be between 1.4 and 1.9 and the median of 260/230 ratios was between 0.01 and 1.7 (Table 1). Size fractionation analysis of extracted DNA samples using 1% agarose gel electrophoresis showed the fractionation of DNA from paraffin-embedded tissues but >85% of DNA fragments had a size larger than 500-bp. The quantity of extracted DNA from serum and plasma using the Allprep method was less than the detection rate of the agarose gel and we could not see any band on the gel (Figure 1A) as it was shown in our many studies on serum and plasma DNA using conventional commercial kits.

Using the DNA LabChip, we assessed the quantity and size fractionation determination of extracted DNA samples. DNA LabChip assay also showed the fractionation of DNA from paraffin-embedded tissues and good quality of DNA for other types of analyzed samples (Figure 1B).

The Ct values of DNA-*GAPDH* amplification for different sample types were analyzed using Q-PCR. The average of Ct values for cell line, whole blood, buffy coat, paraffin-embedded tissue, serum, and plasma was 19.2, 22.1, 23.6, 32.3, 38.7, 40.3 (45−46 cycles for the reaction), respectively (Figure 1D). Using the same amount of input DNA, the lowest Ct values were observed for the whole blood, cell line, and buffy coat. The highest Ct values were obtained for the serum and plasma, which had the lowest quantity of extracted DNA (Figure 1D).

Methylation Status of Four Tumor Suppressor Genes Using MALDI-TOF MS. As a high-throughput quantitative methylation analysis of extracted DNA samples using AllPrep kit, we analyzed the methylation patterns of four tumor suppressor genes in six different types of analyzed samples. The CpG sites were analyzed by T-cleavage assay using MALDI-TOF MS. For all the studied genes, we analyzed one amplicon per gene. In total, we analyzed four amplicons in four genes, containing 121 CpG sites per sample (Table 2; Figure 2A; Supplementary Data 2). The sensitive T-cleavage assay was able to detect mean of 27.2% of CpG sites in the amplicon for the extracted DNA from cell lines in four studied genes (16.1% in *P16*, 18.2% in *RASSF1*, 40% in *RAR-b*, and 34.4% in *BIN1*), 22.2% of CpG sites in the amplicon for DNA from whole blood, 29.6% of CpG sites in the amplicon for DNA from buffy coat, 18.3% of CpG sites in the amplicon for DNA from serum, 13.6% of CpG sites in the amplicon for DNA from plasma and 24.2% of CpG sites in the amplicon for DNA paraffin-embedded tissues (Figure 2A; Supplementary Data 2). Using reamplification strategy (Supplementary Data 1) before *in vitro* transcription and T-cleavage assay, the detection rate of CpG sites were increased dramatically ($>50\%$; $p < 0.001$) and the assay was able to detect 51.1%, 49.1%, 55.1%, 42.9%, 32.7% and 49.7% of CpG sites in the amplicon for the DNA from cell lines, whole blood,

Simultaneous Isolation of DNA, RNA, and Proteins   research articles
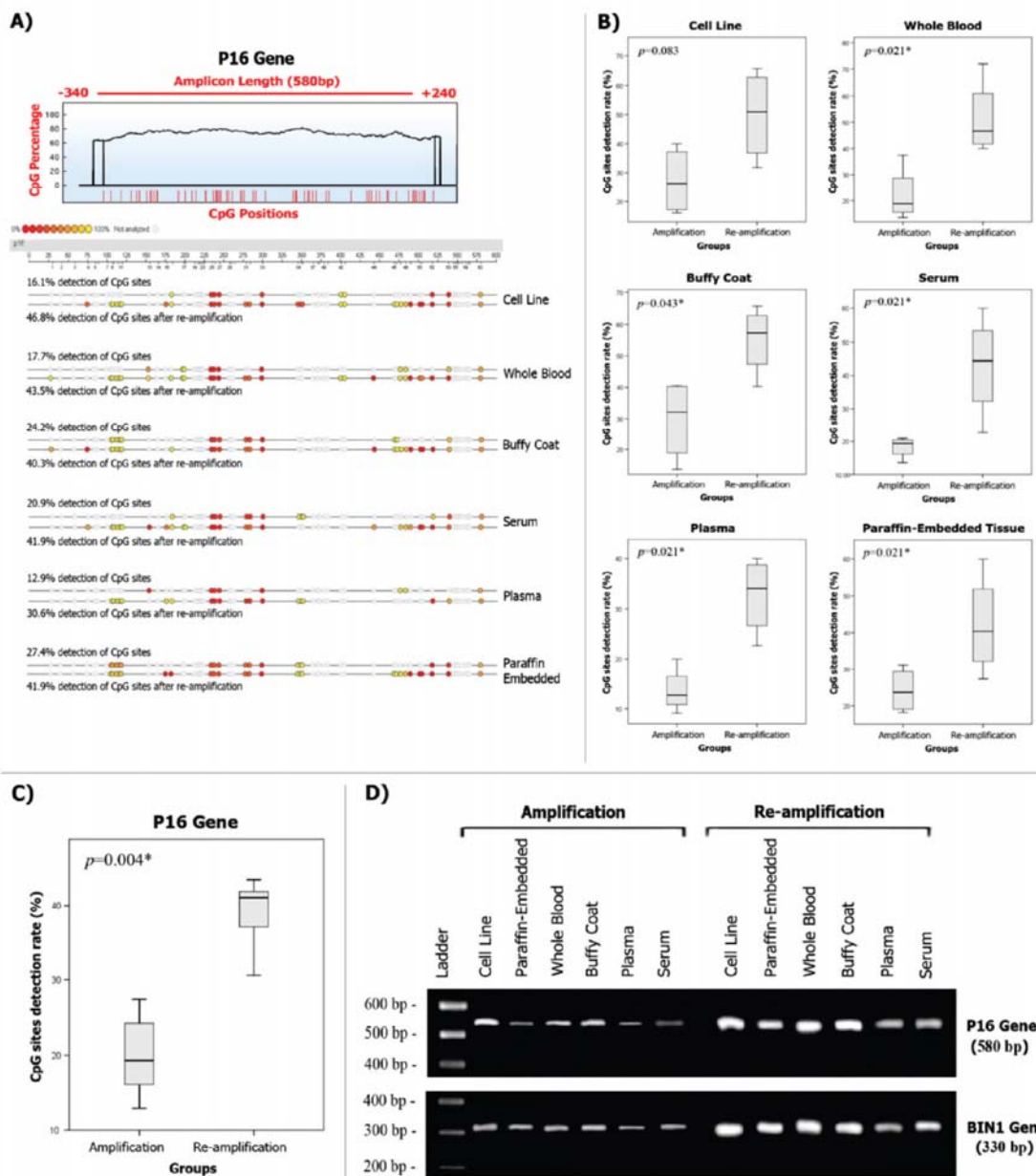
Figure 2. High-throughput methylation analysis of informative CpG sites for four tumor suppressor genes in the six sub groups of studied samples. (A) An example of high-throughput analysis of informative CpG sites for *P16* tumor suppressor genes before and after reamplification (colored circles representing analyzed and open circles representing not analyzed CpG sites). (B) The effect of reamplification on methylation analysis of four tumor suppressor genes in the six sub groups of studied samples. (C) The effect of reamplification on methylation analysis of *P16* gene. (D) PCR amplification/reamplification for the *BIN1* (330 bp) as example of the shortest amplicon and *P16* (580 bp) as the longest amplicon on 1.5% agarose gel. (* significant correlation; Mann−Whitney U Test). The complete data for the other four studied genes is summarized in Supplementary Data 2.

buffy coat, serum, plasma and paraffin-embedded tissues, respectively (Table 2; Figure 2B, C; Supplementary Data 2). The reamplification did not change the ratio of methylated to unmethylated component.
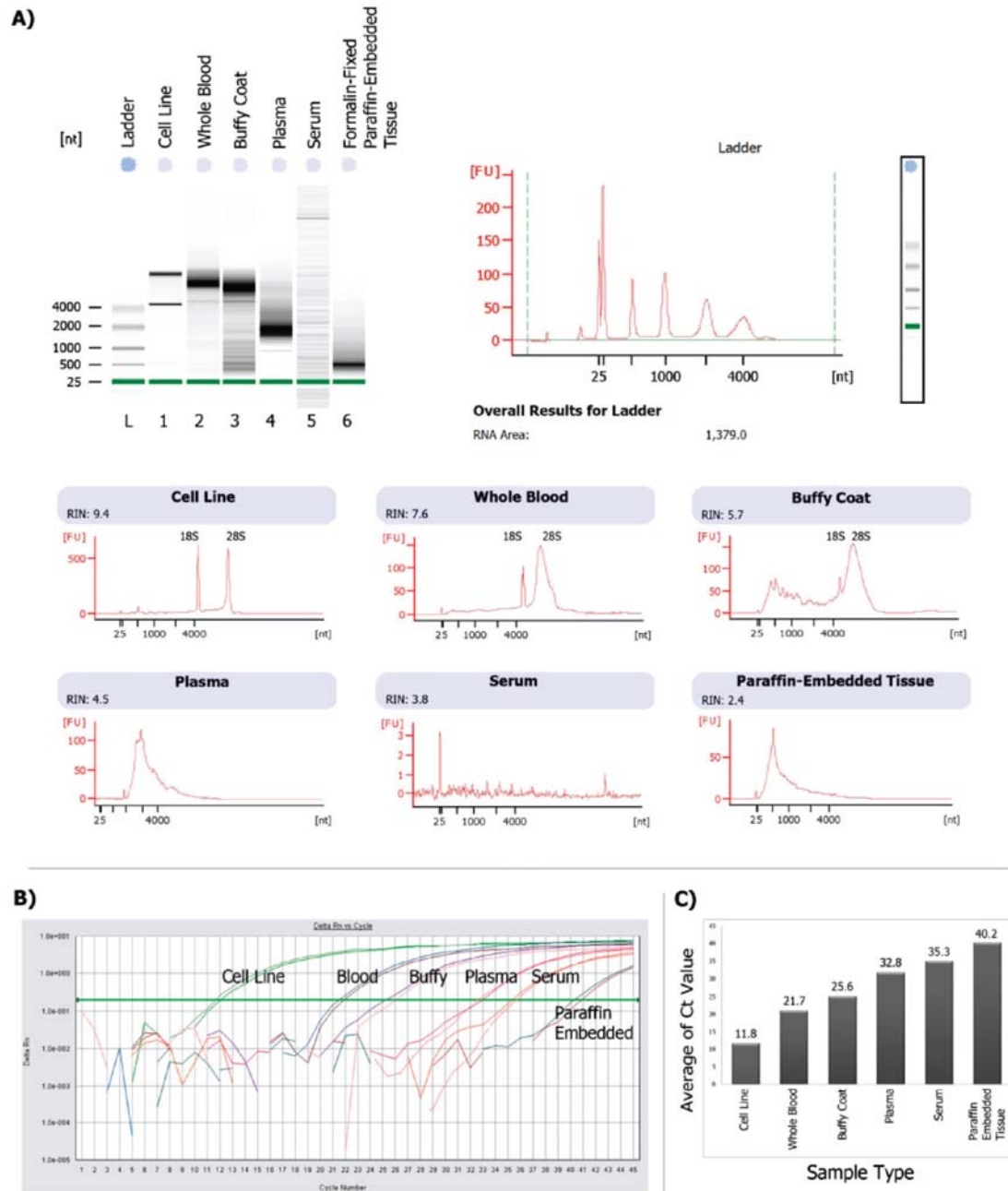
Quantitative and Qualitative Analysis of the Extracted RNA. The total amount of extracted RNA was quantified and analyzed by two different methods. The concentration and purity of the RNA samples were determined by measuring the

absorbance at 260 nm (A260) and 280 nm (A280) using a NanoDrop spectrophotometer. The calculated A260/A280 ratio of pure total RNA was between 1.9 and 2.2, and calculated A260/A230 ratio was between 0.01 and 1.8 (Table 1). The integrity of the RNA samples was further confirmed by RNA Pico LabChip.

Quantity and size fractionation analysis of total RNA was assessed by the Agilent RNA Pico LabChip. The total RNA

Figure 3. Quantitative and qualitative analysis of the extracted RNA. (A) Quantity and size fractionation determination of samples using DNA LabChip. (B) Quantity assessment of samples using Q-PCR. (C) Average of Ct values for different analyzed sample types.

extracted from cell lines by the AllPrep kit yielded fairly sharp peaks of the expected 18S and 28S sizes corresponding to human rRNA. The total RNA extracted from whole blood and buffy coat also displayed 18S and 28S rRNA peaks, but the peaks were always broad-based, indicative of substantial degradation of some rRNAs. The total RNA extracted from plasma, serum, and paraffin-embedded tissues could not display 18S and 28S

rRNA peaks and the RNA was mostly degraded and the RNA integrity number (RIN) was <5 for these samples (Figure 3A).

The Ct values of RNA-*GAPDH* amplification for different sample types were analyzed using Q-PCR after cDNA synthesis. The average of Ct values for cell line, whole blood, buffy coat, plasma, serum, and paraffin-embedded tissue was 11.8, 21.7, 25.6, 32.8, 35.3, and 40.2, respectively (Figure 3C). Using the
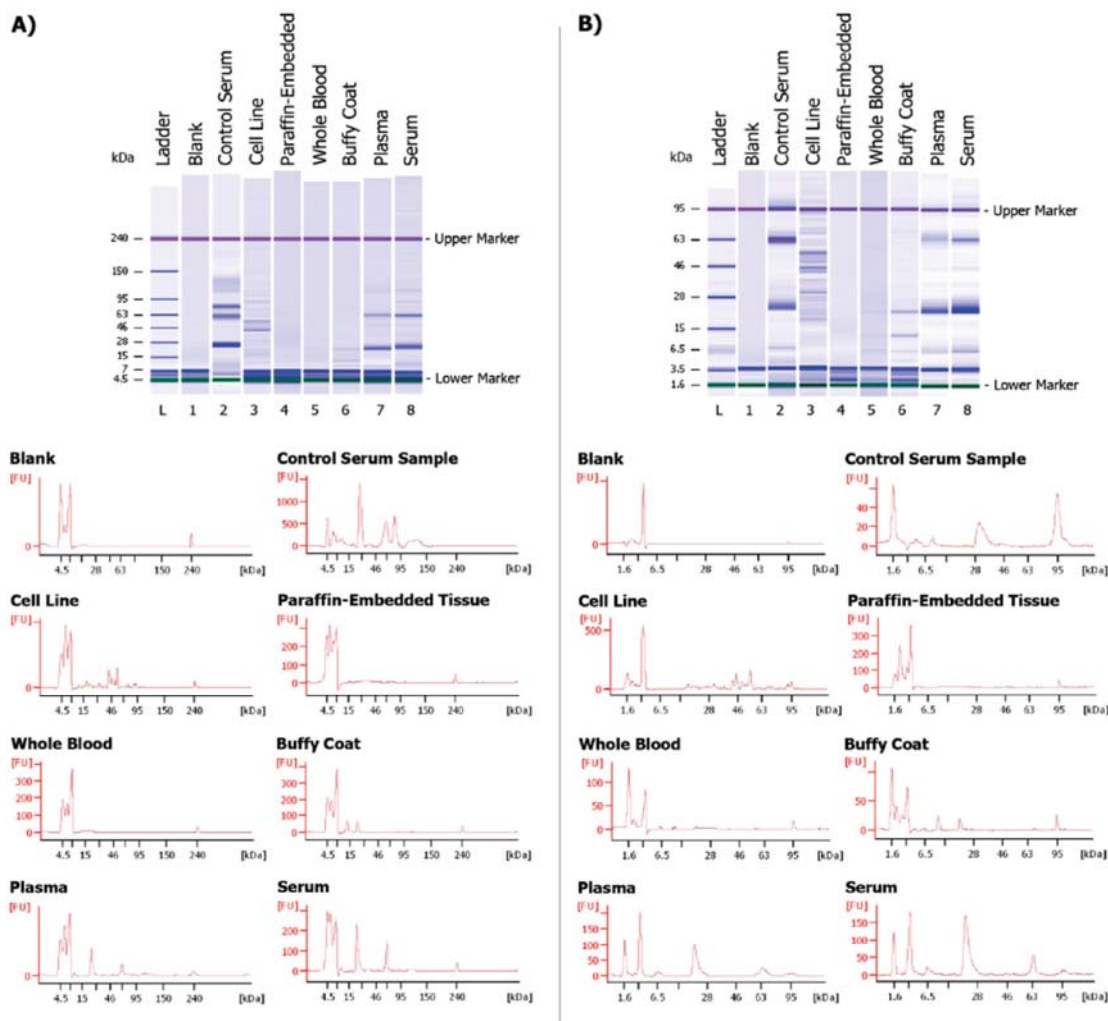
Figure 4. Quantitative analysis and determination of protein extraction efficiency using Protein LABChip. (A) Results for Agilent Protein 230 Kit. (B) Results for Agilent Protein 80 Kit.

same amount of input cDNA revealed the lowest Ct values for the whole blood, cell line, and buffy coat. The highest Ct values were observed for the paraffin-embedded tissue which had the lowest quantity of extracted RNA (Figure 3C).

Quantitative and Qualitative Analysis of the Extracted Proteins. The extracted proteins from different sample types were separated and molecular weight as well as concentration was determined by the Agilent 2100 Bioanalyzer system. Evaluation of the extracted proteins was performed by a computer-simulated gel-like image (Figure 4). The standard protein ladder was analyzed as the first sample and allowed an estimation of the appropriate molecular weight of the separated proteins (Figure 4, lane 1). The sample buffer for Protein 230 LabChip contained a lower (4.5 kDa) and an upper (240 kDa) marker, which allowed correct alignment of each lane (Figure 4A) and for Protein 80 LabChip contained a lower (1.6 kDa) and an upper (95 kDa) marker (Figure 4B). The first peak is the lower marker; the last peak is the upper marker. Molecular weight was calculated according to the protein ladder, migration time in seconds, and relative concentration

of each protein was calculated according to the upper marker. Proteins extracted from whole blood and buffy coat were not completely soluble in the buffer containing 125 mM Tris pH 7.2 and 2% SDS and we were unable to detect more proteins using the Protein LABChips. Using human serum albumin standard as control serum, we could get similar band patterns for the extracted serum proteins in the Agilent Protein 230 and Protein 80 Kits which confirmed the efficiency of extraction using AllPrep kit (Figure 4-a and b).

The proteins from BT549 and SKBR3 breast cancer cell lines were extracted using conventional method and AllPrep method in parallel, and the results obtained by both methods were compared by 2D gel electrophoresis. By the PDQuest-defined spot matching, we have obtained spot sets that are of the same size, since each spot in one population has a partner in the other one. To analyze gel similarities or experimental variations such as disparities in stain intensity or sample loading, we performed scatter plots for matched spots (Figure 5 and Supplementary Data 3). Scatter plots show the relationship between the spot values from two gels by searching for the
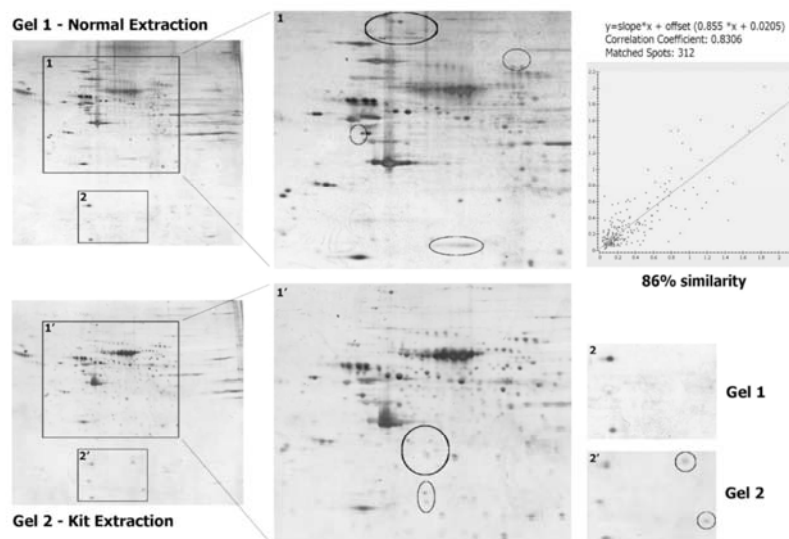
Figure 5. Assessment of protein extraction efficiency and quality using the two-dimensional (2D) gel electrophoresis for BT549 breast cancer cell line. Circles were used to define regions where differences could be observed between proteins extracted using traditional normal method in comparison with AllPrep method. Data regarding to SKBR3 breast cancer cell line is summarized in the Supplementary Data 3.

linear dependence between the spot values of one gel (variable X) and the corresponding values in the sheet reference (variable Y). The scatter plot for the cell line BT549 showed 86% similarity (312 matched spots) between conventional extraction and the AllPrep method (Figure 5) and 82% similarity (280 matched spots) was shown for the cell line SKBR3 (SupplementaryData 3). The linear dependence is defined as the best-fit line through the data points. The best-fit line is described by a slope and its offset from the equation $y = \text{slope} \times x + \text{offset}$. The correlation coefficient as goodness-of-fit was 0.8306 and 0.7572 for the BT549 and SKBR3 cell lines, respectively, showing that our result represented good matching between conventional extraction and the AllPrep method. The overall analysis of these gels showed a highly comparable pattern of proteins with a few exceptional differences. The most notable differences between the extraction methods were the relative abundance of some proteins preferentially extracted with one method in comparison to another. The column-based extraction methods in some cases enriched the protein profile compared to the conventional extraction methods particularly at low molecular weights. However, in some cases, the conventional extraction method appeared to enrich the proteins present in certain areas of the gel, particularly at high molecular weights (Figure 5 and Supplementary Data 3).

Discussion

We present a methodological approach that allows efficient isolation and purification of DNA, RNA, and proteins from a single sample aliquot. The isolated molecular species were found to be usable for a wide range of downstream analysis strategies which are especially valuable in instances of using limited sample resources. This is the first report providing information to meet the technical challenge of simultaneous isolating DNA, RNA, and/or proteins from the blood materials or formalin-fixed/paraffin-embedded tissue. The quality and

quantity of extracted nucleic acids and/or proteins allowed for downstream genetic, epigenetic, transcriptomic, and proteomic studies.

Isolation of DNA, RNA, and proteins separately is a time-consuming and laborious process. As in the past few years also the consideration of economical aspects has gained additional importance, the expenses for material as well as for human resources should be kept well under control and the simultaneous extraction method is indeed an efficient and cost-effective procedure. In addition, cancer tissues are highly heterogeneous and there may be some inherent differences between adjacent sections that could introduce errors when comparing genome and epigenome changes with gene expression and proteomic profiles.[17,18] Using the simultaneous isolation of species from a single sample aliquot, the relationship between genome, transcriptome and proteome profiles could be better approximated than if separate samples were interrogated. This strategy will be especially helpful for the researchers who intend to analyze profiles of heterogeneous cancerous tissues at the genomic, transcriptomic and proteomic levels. One of the important advantages of simultaneous isolation techniques at the proteomic level is removing the contamination of DNA and RNA from the proteins. This is particularly relevant in 2D gel analysis in which contaminating nucleic acids appear as streaks on the gels. The removal of these nucleic acids increases the clarity of the 2D gels and improves protein yield and spot resolution.[19] Whether some rare regulatory proteins remain bound to the DNA and could not be eluted during the column-based extraction (and are therefore lost for proteomic analysis) remains to be shown. We have used lysates from cell lines rather than from tissue samples, since proteomic patterns of cell lines are useful as benchmark comparison of the "completeness" of the patterns.

The quantity of circulatory cell-free DNA in human plasma and serum samples using conventional extraction methods was published by our group.[20] Using the AllPrep method, results

Simultaneous Isolation of DNA, RNA, and Proteins

confirmed that serum samples have a higher concentration of cell-free DNA than plasma (Figure 1, Table 1) which is in accordance with previous reports.[21] One limitation of the adapted AllPrep method for extraction from serum and plasma is that, although it recovers most of the DNA compared to the control method, it only recovers approximately 30% of the total RNA compared to the separate conventional RNA extraction methods (data not shown). The yield of the extracted RNA from serum and plasma (5.4 and 9.8 ng/μL, respectively) seems to be sufficient for Q-PCR or gene expression profiling for which only 5−30 ng of total RNA was needed. In general, using the AllPrep method, the quantity of extracted DNA and RNA from serum and plasma was considerably lower than when using conventional approaches in our group (Figures 1 and 3). It is very possible that this result might be affected by using lesser amount of starting material in the AllPrep method (500 μL) in comparison with other established methods (1000−1600 μL).[22,23]

Simple methods,[24] as well as commercial kits[25] are available for isolation of DNA from paraffin. DNA and RNA isolated from formalin-fixed, paraffin-embedded tissues are highly fragmented due to the formalin-induced cross-linking. Our results for simultaneous isolation from paraffin-embedded tissues confirmed these findings (Figures 1 and 3). This may require the development of novel PCR primers that produce smaller products. At present, formalin-fixed, paraffin-embedded tissues seem not to be suitable as starting materials for genome/epigenome-wide molecular profiling.[26,27] Using reamplification strategy before methylation analysis on MALDI-TOF MS, we could overcome this problem with consequences of dramatically increasing the detection rate (also for long amplicons such as 580-bp in length). Because this reamplification was done after bisulfite conversion of DNA samples, it did not affect the quantity of methylation status (Figure 2 and Supplementary Data 2).[10,11]

For protein extraction from blood and buffy coat, we could not use the ALO buffer (Laemmli-related buffer which contains 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromphenol blue, 0.125 M Tris HCl and 8 mg DTT per 1 μL; supplied by AllPrep kit). In the Protein LabChip sample buffer, there are also high levels of detergent which does interfere with the analysis. Bromphenol blue should be also omitted because it would create baseline humps and it interferes with the dye in the chip gel. Glycerol is useable because it is not charged. For this reason, we used buffer containing 125 mM Tris pH 7.2 and 2% SDS. Cellular proteins undergo a high degree of cross-linking during the formalin-fixation and paraffin-embedding steps which in turn makes proteins in the sample insoluble[28] with the consequence that we could not get any proteins from paraffin-embedded tissues (Figure 4). There have been several attempts to set them free from their covalent cage to enable a better protein extraction for proteomic analysis. According to our knowledge, there does not exist any well-documented method for this strategy and formalin-fixed, paraffin-embedded tissues are normally not considered to be suitable for the study of proteins, with the exception of immunohistochemistry (IHC) which is a semiquantitative protein analysis method.[28] Ikeda and co-workers have used heat-induced antigen retrieval (AR) technology for extracting proteins from paraffin-embedded tissues that provided material suitable for 2D gel electrophoresis.[29]

In summary, we have demonstrated that AllPrep methodology can be used for simultaneous isolation of DNA, RNA, and/or proteins from single sample aliquots of whole blood, buffy

coat, plasma, and serum and with some reservation also to paraffin-embedded samples. The quality and quantity of the samples obtained by this strategy were tested by several genetic, epigenetic, transcriptomic, and proteomic methods. The main benefit for our project is obtaining high quality results using limited sample resources.

Supporting Information Available: The sequence of PCR-tagged primers for *in vitro* transcription, PCR conditions for quantitative methylation analysis using MALDI-TOF MS, primers and probe sequences for quantitative PCR, are available in Supplementary Data 1. The complete data regarding the methylation quantification of four tumor suppressor genes is summarized in Supplementary Data 2. The results of 2D gel electrophoresis of SKBR3 breast cancer cell line are shown in Supplementary Data 3. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Donoso, P.; Devroey, P. PGD for aneuploidy screening: an expensive hoax. *Best Pract. Res., Clin. Obstet. Gynaecol.* **2007**, *21* (1), 157−68.

(2) Bianchi, D. W.; Wataganara, T.; Lapaire, O.; Tjoa, M. L.; Maron, J. L.; Larrabee, P. B.; Johnson, K. L. Fetal nucleic acids in maternal body fluids: an update. *Ann. N.Y. Acad. Sci.* **2006**, *1075*, 63−73.

(3) Noutsias, M.; Rohde, M.; Block, A.; Klippert, K.; Lettau, O.; Blunert, K.; Hummel, M.; Kuhl, U.; Lehmkuhl, H.; Hetzer, R.; Rauch, U.; Poller, W.; Pauschinger, M.; Schultheiss, H. P.; Volk, H. D.; Kotsch, K. Preamplification techniques for real-time RT-PCR analyses of endomyocardial biopsies. *BMC Mol. Biol.* **2008**, *9*, 3.

(4) Chomczynski, P. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *BioTechniques* **1993**, *15* (3), 532−4, 536−7.

(5) Jenner, M. W.; Leone, P. E.; Walker, B. A.; Ross, F. M.; Johnson, D. C.; Gonzalez, D.; Chiecchio, L.; Dachs Cabanas, E.; Dagrada, G. P.; Nightingale, M.; Protheroe, R. K.; Stockley, D.; Else, M.; Dickens, N. J.; Cross, N. C.; Davies, F. E.; Morgan, G. J. Gene mapping and expression analysis of 16q loss of heterozygosity identifies WWOX and CYLD as being important in determining clinical outcome in multiple myeloma. *Blood* **2007**, *110* (9), 3291–300.

(6) Walker, B. A.; Leone, P. E.; Jenner, M. W.; Li, C.; Gonzalez, D.; Johnson, D. C.; Ross, F. M.; Davies, F. E.; Morgan, G. J. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. *Blood* **2006**, *108* (5), 1733–43.

(7) Morse, S. M.; Shaw, G.; Larner, S. F. Concurrent mRNA and protein extraction from the same experimental sample using a commercially available column-based RNA preparation kit. *BioTechniques* **2006**, *40* (1), 54, 56, 58.

(8) Hummon, A. B.; Lim, S. R.; Difilippantonio, M. J.; Ried, T. Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. *BioTechniques* **2007**, *42* (4), 467−70, 472.

(9) Tolosa, J. M.; Schjenken, J. E.; Civiti, T. D.; Clifton, V. L.; Smith, R. Column-based method to simultaneously extract DNA, RNA, and proteins from the same sample. *BioTechniques* **2007**, *43* (6), 799–804.

(10) Radpour, R.; Kohler, C.; Haghighi, M. M.; Fan, A. X.; Holzgreve, W.; Zhong, X. Y. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. *Oncogene* **2009**, *28* (33), 2969–78.

(11) Radpour, R.; Haghighi, M. M.; Fan, A. X.; Torbati, P. M.; Hahn, S.; Holzgreve, W.; Zhong, X. Y. High-Throughput Hacking of the

Methylation Patterns in Breast Cancer by In vitro Transcription and Thymidine-Specific Cleavage Mass Array on MALDI-TOF Silico-Chip. *Mol. Cancer Res.* **2008**, *6* (11), 1702–9.

(12) Ehrich, M.; Nelson, M. R.; Stanssens, P.; Zabeau, M.; Liloglou, T.; Xinarianos, G.; Cantor, C. R.; Field, J. K.; van den Boom, D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (44), 15785–90.

(13) Li, L. C.; Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **2002**, *18* (11), 1427–31.

(14) Grussenmeyer, T.; Meili-Butz, S.; Dieterle, T.; Traunecker, E.; Carrel, T. P.; Lefkovits, I. Quantitative proteome analysis in cardiovascular physiology and pathology. I. Data processing. *J. Proteome Res.* **2008**, *7* (12), 5211–20.

(15) Anderson, N. L.; Anderson, N. G. Analytical techniques for cell fractionsXXII. Two-dimensional analysis of serum and tissue proteins: multiple gradient-slab gel electrophoresis. *Anal. Biochem.* **1978**, *85* (2), 341–54.

(16) Mortz, E.; Krogh, T. N.; Vorum, H.; Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **2001**, *1* (11), 1359–63.

(17) Torres, L.; Ribeiro, F. R.; Pandis, N.; Andersen, J. A.; Heim, S.; Teixeira, M. R. Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast Cancer Res. Treat.* **2007**, *102* (2), 143–55.

(18) Andersen, C. L.; Wiuf, C.; Kruhoffer, M.; Korsgaard, M.; Laurberg, S.; Orntoft, T. F. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* **2007**, *28* (1), 38–48.

(19) Kirkland, P. A.; Busby, J.; Stevens, S., Jr.; Maupin-Furlow, J. A. Trizol-based method for sample preparation and isoelectric focusing of halophilic proteins. *Anal. Biochem.* **2006**, *351* (2), 254–9.

(20) Zhong, X. Y.; Hahn, S.; Kiefer, V.; Holzgreve, W. Is the quantity of circulatory cell-free DNA in human plasma and serum samples associated with gender, age and frequency of blood donations. *Ann. Hematol.* **2007**, *86* (2), 139–43.

(21) Lee, T. H.; Montalvo, L.; Chrebtow, V.; Busch, M. P. Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma. *Transfusion* **2001**, *41* (2), 276–82.

(22) Zhong, X. Y.; Holzgreve, W.; Huang, D. J. Isolation of cell-free RNA from maternal plasma. *Methods Mol. Biol.* **2008**, *444*, 269–73.

(23) Wong, S. C.; Lo, E. S.; Cheung, M. T. An optimised protocol for the extraction of non-viral mRNA from human plasma frozen for three years. *J. Clin. Pathol.* **2004**, *57* (7), 766–8.

(24) Sjoholm, M. I.; Hoffmann, G.; Lindgren, S.; Dillner, J.; Carlson, J. Comparison of archival plasma and formalin-fixed paraffin-embedded tissue for genotyping in hepatocellular carcinoma. *Cancer Epidemiol., Biomarkers Prev.* **2005**, *14* (1), 251–5.

(25) Everson, R. B.; Mass, M. J.; Gallagher, J. E.; Musser, C.; Dalzell, J. Extraction of DNA from cryopreserved clotted human blood. *BioTechniques* **1993**, *15* (1), 18–20.

(26) Coudry, R. A.; Meireles, S. I.; Stoyanova, R.; Cooper, H. S.; Carpino, A.; Wang, X.; Engstrom, P. F.; Clapper, M. L. Successful application of microarray technology to microdissected formalin-fixed, paraffin-embedded tissue. *J. Mol. Diagn.* **2007**, *9* (1), 70–9.

(27) Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics* **2005**, *4* (11), 1741–53.

(28) Tapio, S.; Atkinson, M. J. Molecular information obtained from radiobiological tissue archives: achievements of the past and visions of the future. *Radiat. Environ. Biophys.* **2008**, *47* (2), 183–7.

(29) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Shiozaki, H.; Monden, M. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J. Histochem. Cytochem.* **1998**, *46* (3), 397–403.

PR900591W

Supplementary Data 1

**Table 1.** The sequence of PCR primers for Real-time PCR

Gene	Primer	Sequence (5'→3')	Length	T_a	Prodoct Size (bp)
GAPDH (for DNA)	Forward	CCCCACACACATGCACTTACC	21	60	97
	Reverse	CCTAGTCCCAGGGCTTTGATT	20		
	Probe	MGB–TAGGAAGGACAGGCAAC–FAM	17		
GAPDH (for RNA)	Forward	GAAGGTGAAGGTCGGAGT	18	60	226
	Reverse	GAAGATGGTGATGGGATTTC	20		

For the quantitative PCR (Q-PCR) of DNA, we used a reaction volume of 25μL, containing 5μL of template, primers at a concentration of 300nM, a probe at 200nM, and 12,5μL of 2xTaqMan® Master Mix (Applied Biosystems, Branchburg, New Jersey, USA). The reaction was performed at the following conditions: a first denaturation step at 95°C for 10 minutes, and 40 cycles of 1 minute at 60°C, followed by 15 seconds at 95°C. For the Q-PCR of RNA samples, an aliquot of 25ng (5μL) template cDNA was added to each reaction containing 12.5μL SYBR, Green PCR Master Mix (Applied Biosystems, Foster City, California, USA) and 7.5μL primers mixture. The thermal cycling profile for the _GAPDH_ was 95°C for 10 min, followed by 40 cycles of 95°C for 15s and 60°C for 1min. Following amplification of _GAPDH_ in cDNA samples, a dissociation curve was drawn in order to confirm the specificity of the reaction.

Table 2. The sequence of PCR tagged primers for _in vitro_ transcription and quantitative methylation study

| Gene	Primer	Sequence (5'→3')	Length	T_a	Product Size (bp)
BIN1	tag-EN1-FW	AGGAAGAGAGGGAGGTGAGTTTTTGGAA	18+10	58	330
	T7-EN1-RV	CAG**TAATACGACTCACTATAGGG**AGAAGGCTCTACCTTTTAAAAAACCACTCC	22+31		
_P16_ (CDKN2A)	tag-EN1-FW	AGGAAGAGAGGGGTTGTTTTTGGTAGGG	17+10	58	580
	T7-EN1-RV	CAG**TAATACGACTCACTATAGGG**AGAAGGCTATATAAACCACRAAAACCC	19+31		
RAR-b	tag-EN1-FW	AGGAAGAGAGGAGTGTATGTTAATGGGGGAG	21+10	56	475
	T7-EN1-RV	CAG**TAATACGACTCACTATAGGG**AGAAGGCTTTCCCAACATAATTTCTCTAC	21+31		
RASSF1	tag-EN1-FW	AGGAAGAGAGGGGYGGTAAAGTTGTTGA	18+10	56	346
	T7-EN1-RV	CAG**TAATACGACTCACTATAGGG**AGAAGGCTCRCAATAAAAACCTAAATACA	21+31		

For the PCR on bisulphite-treated genomic DNA (gDNA), the following PCR conditions were used: 1x: 95°C for 10 min; 48x: 95°C for 20s, Ta for 30s, 72°C for 1 min; 1x 72°C for 5 min. For the PCR on gDNA, the following cycling conditions were used: 1x: 95°C for 10 min; 40x: 95°C for 20 s, Ta for 30s, 72°C for 45s; 1x 72°C for 5 min. After first round of PCR amplification we used 1μl of the PCR product for re-amplification of selected amplicons with the same primer pairs.
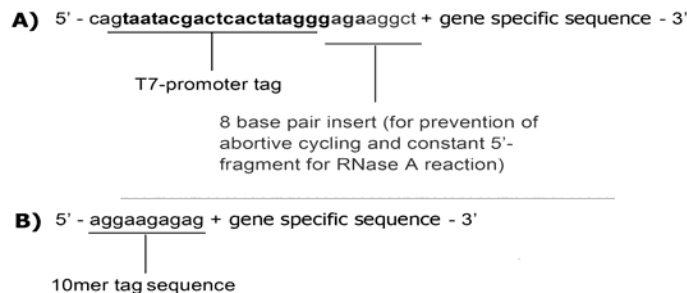

Primer taging for _in vitro_ transcription. (A) Reverse primer with T7-promoter tag. (B) Forward primer with 10mer tag sequence as balance.

Supplementary Data 2

Qualitative determination of extracted DNA using high-throughput methylation analysis of four tumor suppressor genes by thymidine-specific cleavage mass array on MALDI-TOF silico-chip, before and after re-amplification
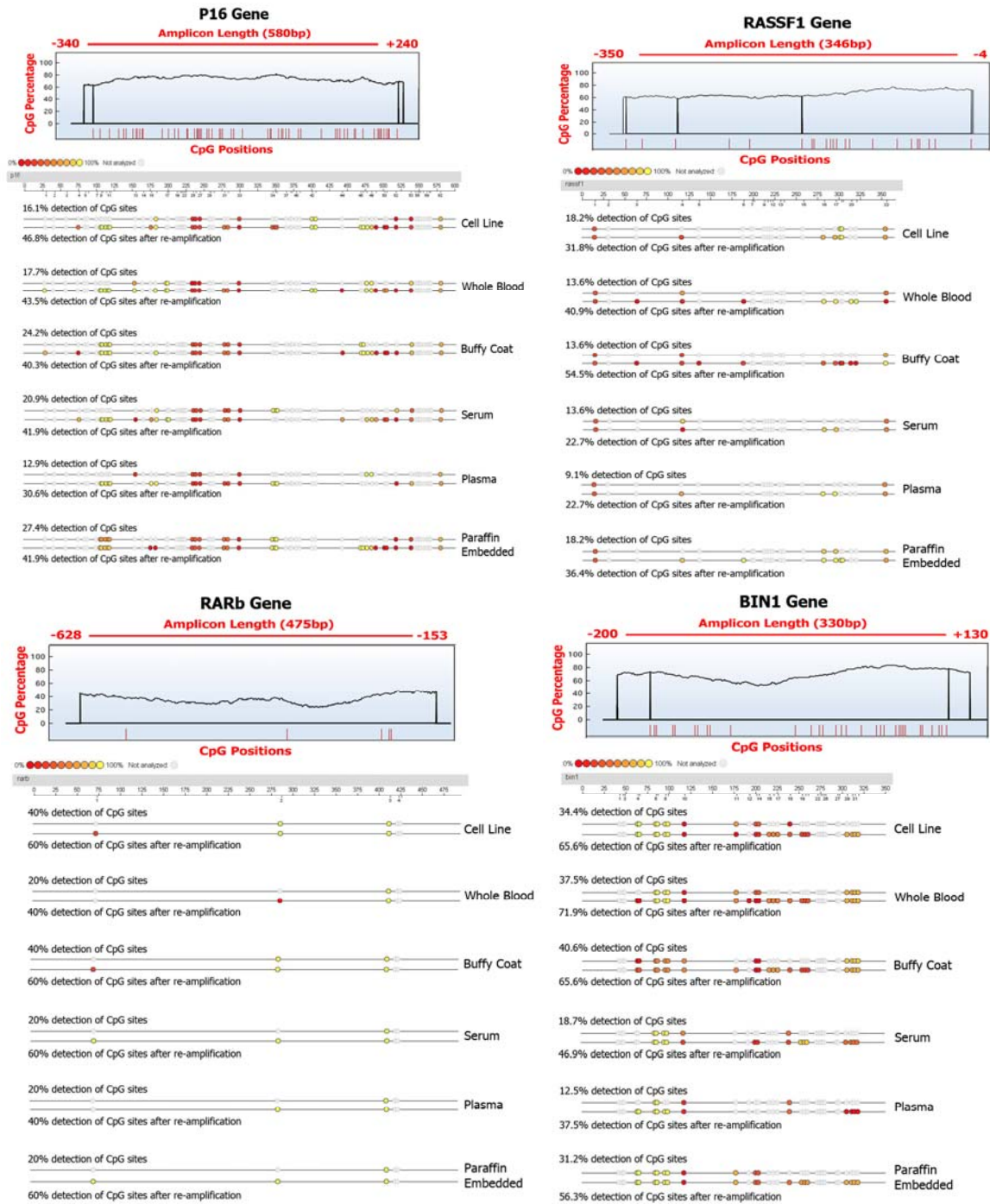
Colored circles representing analyzed CpG sites (Red color indicates 0% methylated, yellow color indicates 100% methylated and color gradient between red and yellow indicates methylation ranging from 0-100) and open circles representing not analyzed CpG sites.
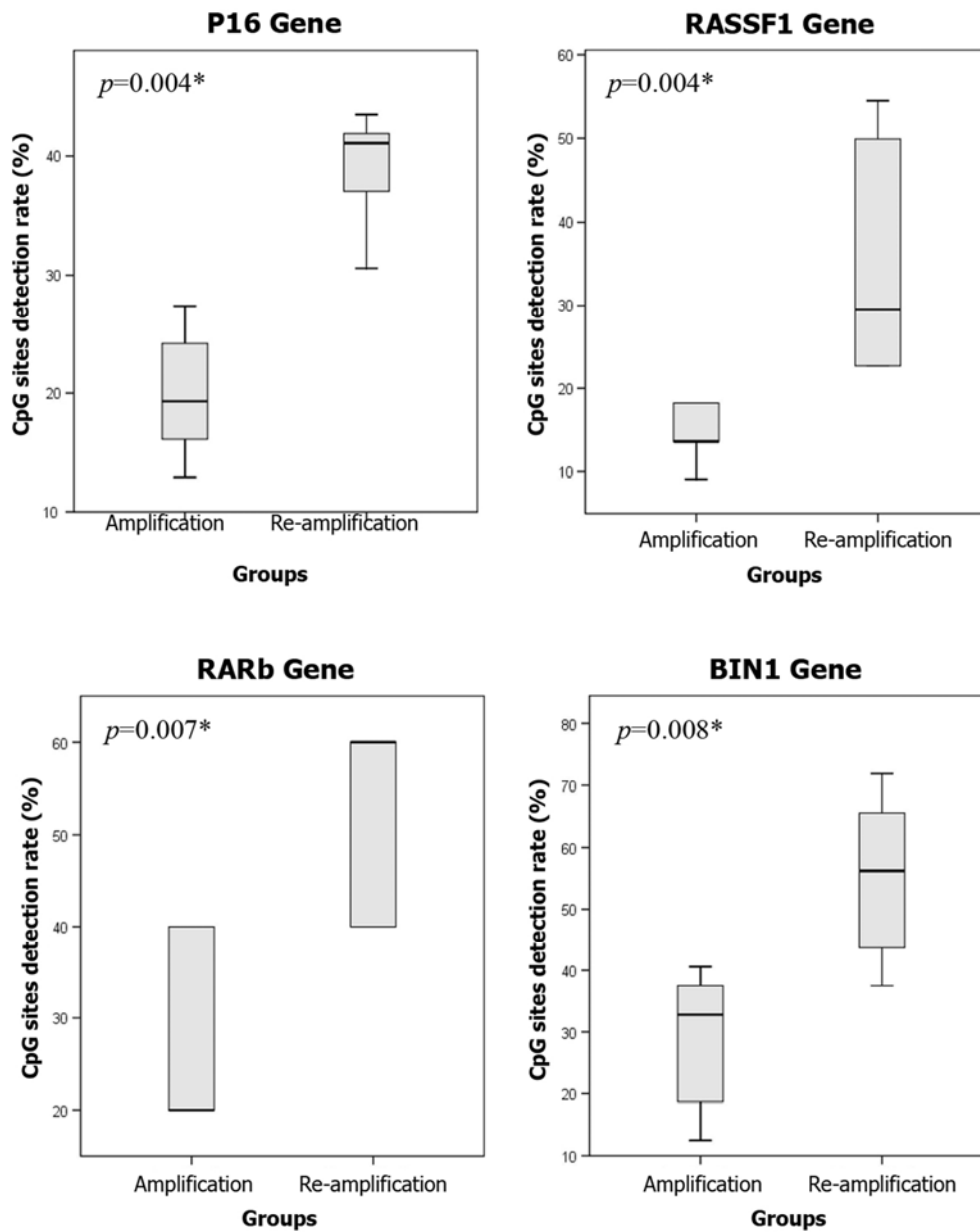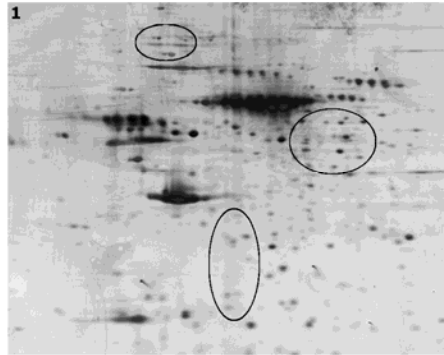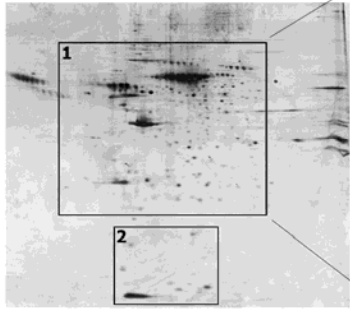
The effect of re-amplification on high-throughput methylation analysis of four studied tumor suppressor genes using MALDI-TOF MS (* significant correlation; Mann-Whitney U Test).

Supplementary Data 3

Two-dimensional (2D) gel electrophoresis for SKBR3 breast cancer cell line
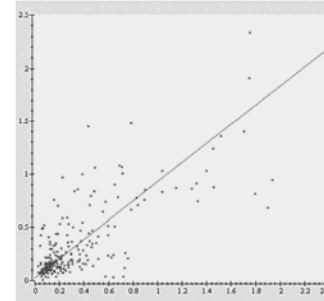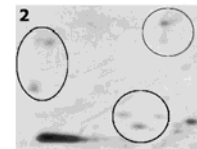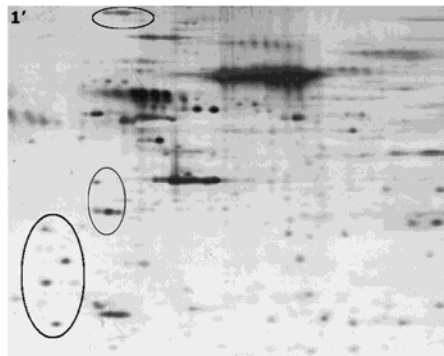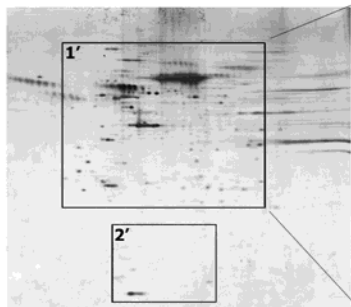
Assessment of protein extraction efficiency and quality using the two-dimensional (2D) gel electrophoresis for SKBR3 breast cancer cell line. Circles were used to define regions where differences could be observed between proteins extracted using traditional normal method in comparison with AllPrep method.

6. Published research article:

Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array

Journal: Oncogene. 2009 Aug 20;28(33):2969-78.

Summary:

For the first time, we used a high-throughput mass array to quantify the methylation status of 22 genes in breast cancer. We quantitatively analyzed 42,528 CpG sites on the genes in paired cancerous and adjacent normal breast tissues (96 samples) from 48 patients with breast cancer. Using the hierarchical clustering and distance analysis, 10 hypermethylated genes (*APC*, *BIN1*, *BMP6, BRCA1, CST6, ESRb, GSTP1, P16, P21* and *TIMP3*) were identified for breast cancer. The breast cancer tissue specific hypermethylated genes identified in this study might provide valuable information in their application for the classification, sensitive detection, risk and therapy assessment of breast cancer, as well as for developing novel targeted therapeutic strategies via modification of pathological epigenetic changes in cancer.

First author's contribution:

Ramin Radpour was involved in performing the experiment, data analysis and writing the manuscript.

ORIGINAL ARTICLE

Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array

R Radpour[1], C Kohler[1], MM Haghighi[2], AXC Fan[1], W Holzgreve[3] and XY Zhong[1]

[1]Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel, Basel, Switzerland; [2]Department of Biology, Faculty of Science, IAU, East of Tehran Branch, Tehran, Iran and [3]University Medical Center Freiburg, Freiburg, Germany

Alterations of DNA methylation patterns have been suggested as biomarkers for diagnostics and therapy of cancers. Every novel discovery in the epigenetic landscape and every development of an improved approach for accurate analysis of the events may offer new opportunity for the management of patients. Using a novel high-throughput mass spectrometry on matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) silico-chips, we determined semiquantitative methylation changes of 22 candidate genes in breast cancer tissues. For the first time we analysed the methylation status of a total of 42 528 CpG dinucleotides on 22 genes in 96 different paraffin-embedded tissues (48 breast cancerous tissues and 48 paired normal tissues). A two-way hierarchical cluster analysis was used to classify methylation profiles. In this study, 10 hypermethylated genes (*APC, BIN1, BMP6, BRCA1, CST6, ESRb, GSTP1, P16, P21* and *TIMP3*) were identified to distinguish between cancerous and normal tissues according to the extent of methylation. Individual assessment of the methylation status for each CpG dinucleotide indicated that cytosine hypermethylation in the cancerous tissue samples was mostly located near the consensus sequences of the transcription factor binding sites. These hypermethylated genes may serve as biomarkers for clinical molecular diagnosis and targeted treatments of patients with breast cancer.
Oncogene (2009) **28**, 2969–2978; doi:10.1038/onc.2009.149; published online 8 June 2009

Introduction

DNA methylation has been linked to carcinogenesis. Investigation of the events in human cancer has provided key insights into understanding the mechan-

isms of cancer development (Suzuki and Bird, 2008). Hypermethylation of human tumor suppressor genes (TSGs) leads to the silencing of genes responsible for tumor suppression, thus causing cancers. Aberrant DNA methylation patterns have been suggested as biomarkers in cancer molecular diagnostics (Laird, 2003) and for the design of new generations of targeted approaches for clinical intervention (Laird, 2003; Suzuki and Bird, 2008). For example, demethylating agents can be used for the treatments of hypermethylation-associated tumors (Laird, 2003; Issa, 2004; Suzuki and Bird, 2008). However, the clinical usefulness of methylation is still limited due to the fact that no one technique is broadly accurate for quantification and sensitive detection of methylation changes in cancers (Suzuki and Bird, 2008). Large-scale studies that enable evaluating quantitative alterations of methylation for multiple CpG sites in various gene regions and testing a large number of samples with automation are rare (Issa, 2004; Suzuki and Bird, 2008).

SEQUENOM's EpiTYPER assay created the opportunity for the high-throughput quantitative and semiquantitative analysis of DNA methylation status using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and MassCLEAVE reagent, which is based on base-specific (C/T) cleavage reactions (Stanssens *et al.*, 2004; Ehrich *et al.*, 2005). The robustness of the approach for quantifying methylated and unmethylated DNA has been confirmed earlier by SEQUENOM (Ehrich *et al.*, 2008) and also by our group (Radpour *et al.*, 2008).

Breast cancer is the most common type of cancer and most common leading cause of cancer death in women. In this study, for the first time, we investigated semiquantitative methylation changes of 22 human genes (*APC, BIN1, BMP6, BRCA1, BRCA2, CADHERIN 1, CST6, DAPK1, EGFR, ESR2, GSTP1, NES1, Nm23-H1, P16, P21, Progesterone receptor, Prostasin, RAR-b, RASSF1, SRBC, TIMP3* and *TP53*) in the malignancy using the high-throughput robust system with automation to search for novel biomarkers applicable to the management of patients. The selected candidate cancer genes for this study have annotated functionality in the cell adhesion, cell interaction, invasion, metastasis, angiogenesis or gene expression during cancer development and progression according to the OMIM (Online Mendelian Inheritance in

Man database; http://www.ncbi.nlm.nih.gov/omim) (Supplementary Data). A two-way hierarchical cluster analysis was used to identify breast cancer-specific hypermethylated genes according to differences in methylation between cancerous and normal breast tissues on large-scale CpG dinucleotide analysis in the 22 genes. In order to understand the relationship between methylation and transcription events, individual semiquantitative assessment of the methylation status of each CpG dinucleotide was carried out to check the position of hypermethylated CpG sites on the genes.

Results

Sensitive detection of methylation status using high-throughput MALDI-TOF MS with automation

The theory, practice and assay concept of the system have been published by the company and our group (Ehrich et al., 2008; Radpour et al., 2008). The system enables combining quantitative, sensitive, automated and high-throughput detection of methylation status in a single technique.

Quantitative methylation analysis. For every cleaved CpG site, the intensity of a pair of mass signals, one representing methylated and/or another one representing unmethylated DNA, was recorded and interpreted by MassARRAY EpiTYPER software. The relative methylation status was estimated by dividing the peak intensity, or area of the methylated DNA, by the sum of the intensities or areas of the methylated and unmethylated components. The ratios between methylated and unmethylated DNA were obtained as quantity of methylation for further analysis. In our group, we mixed full methylated DNA into the pure unmethylated DNA in different ratios of 100:0, 50:50, 25:75, 5:95 and 0:100. The assay was able to discriminate the methylated and unmethylated components according to the ratios (Radpour et al., 2008).

Sensitive detection. The assay enabled reliable detection of the methylated DNA in as little as 5 ng of DNA per PCR. At the lowest template input of 2.5 ng, some of the replicates failed to generate detectable fragments with distinctive signal-to-noise ratios in the spectrum (Radpour et al., 2008).

High-throughput automation. The approach is completely automated for high-throughput analysis of as much as 384 samples running on one silico-chip.

Using paraffin-embedded tissues for methylation analysis

A major limitation of using formalin-fixed and paraffin-embedded tissue sections is the significant degree of degradation, fragmentation and chemical modification of the nucleic acids recovered from fixed tissues (Srinivasan et al., 2002; Radpour et al., 2008). In this study, we explored the feasibility of using paraffin-embedded tissues for the methylation analysis on MALDI-TOF silico-chips. Before carrying out the methylation analysis, we quantified the yield of extracted DNA from formalin-fixed paraffin-embedded tissues by real-time PCR for *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) gene. The quality of the DNA allowed permitting successful amplification and quantification of the *GAPDH* gene in all samples. In our study, the calls and peaks for determining methylated and unmethylated fragments were successfully obtained from the mass spectrometry after the treatments of *in vitro* transcription and T-cleavage.

Semiquantitative methylation profiles of 22 genes in cancerous and normal breast tissues

In this study, we analysed the methylation patterns of 22 breast cancer candidate genes in 96 cancerous and normal breast tissue samples from 48 patients with breast cancer. For the 21 genes, one amplicon per gene and for the *NES1* gene, two amplicons were analysed. In total, we assessed 23 amplicons in 22 genes, containing 443 CpG sites per sample (total of 42 528 sites in 96 analysed samples) (Table 1, Figure 1 and Supplementary Data 2). From 23 analysed amplicons in 22 selected genes, 4 amplicons (*DAPK1, NES1, Prostasin* and *RAR-b*) contained CpG-poor islands (with number of CpG sites < 20) and others were CpG-rich islands (with number of CpG sites > 20) (Table 1). About 43% of CpG sites were defined as having a low degree of methylation with mean methylation levels below 30%, and 27% of CpG sites were defined as having a high degree of methylation with mean methylation levels above 90% (Supplementary Data 2).

Using the two-way hierarchical cluster analysis, we could find different levels of methylation in 17 genes between cancerous and adjacent normal tissues (Supplementary Data 2). No significant differences of methylation patterns between cancerous and normal tissues could be found in the other five studied genes (*BRCA2, EGFR, NM23-H1, Prostasin* and *TP53*) ($P > 0.05$). For these five genes, we could analyse > 60% of CpG sites in amplicons (55% in *BRCA2*, 47% in *EGFR*, 53% in *NM23-H1*, 70% in *Prostasin* and 77% in *TP53*) (Table 1).

Identifying the breast cancer-specific hypermethylated genes

From these 17 genes, the top 10 genes with a high degree of methylation across the 96 studied samples and significant different levels of methylation between cancerous and adjacent normal tissues ($P < 0.05$) were selected by distance statistic analysis using pair-wise clustering (Figure 2A(A1) and (A2)). The top 10 hypermethylated genes identified by the distance annotation from the 22 studied genes were *APC, BIN1, BMP6, BRCA1, CST6, ESRb, GSTP1, P16, P21* and *TIMP3* for breast cancer (Figures 2A(A3) and 3).

For the 10 hypermethylated genes, we could detect > 61% of CpG sites in amplicons (58% in *APC*, 60% in *BIN1*, 70% in *BMP6*, 57% in *BRCA1*, 50% in *CST6*,

Table 1 High-throughput methylation analysis of informative CpG sites per amplicons for 22 studied genes in breast cancer

Gene	Amplicon size (bp)	Total no. of CpG sites in amplicon	No. of analysed CpG sites in amplicon	No. of analysed CpG sites per amplicons	
				Single sites	Composite sites
APC	420	26	15	11	4
BIN1	330	32	19	4	15
BMP6	397	37	26	9	17
BRCA1	413	30	17	12	5
BRCA2	433	44	24	10	14
CADHERIN 1	500	34	18	7	11
CST6	445	49	24	12	12
DAPK1	408	14	12	8	4
EGFR	428	49	23	5	18
ESR2 (ER beta)	374	30	12	4	8
GSTP1	381	23	18	9	9
NES1 (CpG Island 1)	546	13	11	7	4
NES1 (CpG Island 2)	410	30	22	11	11
Nm23-H1 (NME1)	421	30	16	9	7
P16 (CDKN2A)	580	62	37	11	26
P21 (CDKN1A)	419	30	15	8	7
Progesterone receptor (PGR)	448	59	27	9	18
Prostasin (PRSS8)	450	10	7	7	0
RAR-b	475	5	3	3	0
RASSF1	346	22	12	8	4
SRBC (PRKCDBP)	435	30	17	12	5
TIMP3	441	51	38	14	24
TP53	449	39	30	15	15

The *in silico* digestion was performed for the T-cleavage assay. The percentage of informative CpG sites (total sites) in the amplicon is divided into single sites (single CpG sites) and composite sites (two or three adjacent CpG sites fall within one fragment, or when fragment masses are overlapping).

47% in *ESRb*, 78% in *GSTP1*, 60% in *P16*, 50% in *P21* and 75% in *TIMP3*) (Table 1). In the 10 hypermethylated genes, 39% of CpG sites were methylated with a low degree (average methylation below 30%) and 32% of CpG sites had mean methylation levels above 90% (Supplementary Data 2). The complete sequence and position of CpG sites that showed differences between cancerous and normal tissues are summarized in Supplementary Data 2.

Semiquantitative methylation profiles of 22 genes in three subgroups of patients
Gene clustering of all 22 studied genes was performed independently for each subgroup (invasive ductal carcinoma, infiltrating ductal carcinoma and infiltrating lobular carcinoma) using a two-dimensional hierarchical clustering algorithm (Figure 2B). No significant differences of methylation status between these three subgroups were found ($P > 0.05$). Independent distance analysis for the three subgroups of patients confirmed

that *APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESRb*, *GSTP1*, *P16*, *P21* and *TIMP3* can be identified as the top 10 hypermethylated genes in all subgroups (Figure 2B). In a study cohort, we had eight patients who were ER-β-positive (Table 2), and all of these patients had no hypermethylated profile for the estrogen receptor-beta (*ESR2*) gene.

*Comparison of methylation rate with consensus sequences and recognition sites of well-known transcription factors*
The methylation rate and place of each CpG site were compared with the consensus sequences and recognition sites of well-known transcription factors (upstream sites for regulatory enhancers, CAAT box, GC box, transcription factor IIB recognition elements, TATA box, initiation site of transcription and downstream promoter elements). In comparing the mean methylation value of CpG site positions of all 22 studied genes in the cancerous and normal tissue samples with the consensus sequences, we found that most of the hypermethylated CpG sites for top 10 genes in the cancerous samples were located in a range of −75 to +1, including the TATA box, BRE sequence and CAAT box, but in normal samples, the CpG sites were differentially methylated and located randomly in the 5′-UTRs (untranslated regions) of the studied genes (Figure 4). In the other 12 studied genes, we could not find significant association between place of the CpG sites and conserved sequences.

Discussion

The knowledge of methylation alterations in cancer has been collected for decades. Laird (2003) concluded that 'no one technique or general approach is superior, as the competing goals of quantitative accuracy, sensitive detection, and high local or global informational content, compatibility with formalin fixed tissues and compatibility with automation are not found in a single technique'. In this study, we explored the feasibility of combining accuracy, sensitive detection, high local or global informational content, compatibility with paraffin-embedded tissues and compatibility with automation in a single technique to identify breast cancer-specific DNA methylation changes.

For the first time, we assessed the methylation status of 22 genes in 96 paired normal and cancerous breast tissues from 48 patients (48 breast cancerous tissues and 48 paired normal tissues) using the high-throughput assay on MALDI-TOF MS. Using hierarchical clustering we found that CpG sites from the promoter regions of 17 genes exhibited significantly different methylation degrees ($P < 0.05$) between cancerous and adjacent normal tissues (Figure 1 and Supplementary Data 2). Distance analysis of the methylation differences showed the top 10 genes, such as *APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESRb*, *GSTP1*, *P16*, *P21* and *TIMP3*, selected from 17 significant genes with hypermethylation in breast cancer tissues (Figure 3).
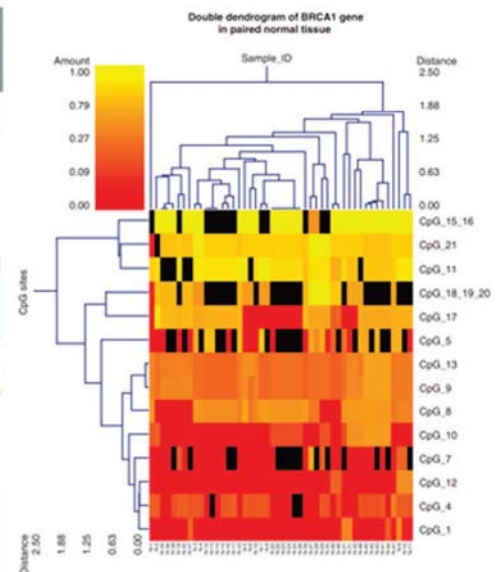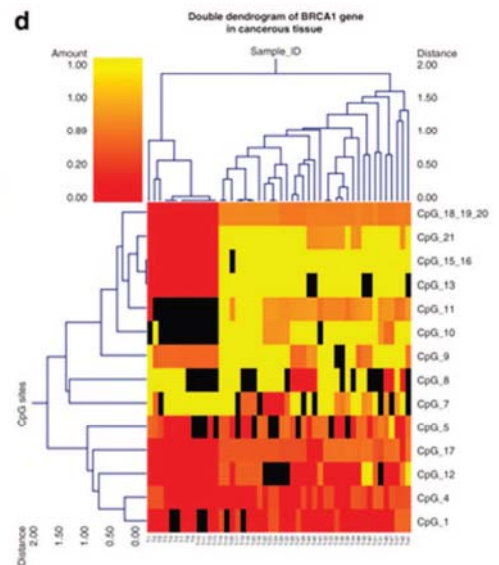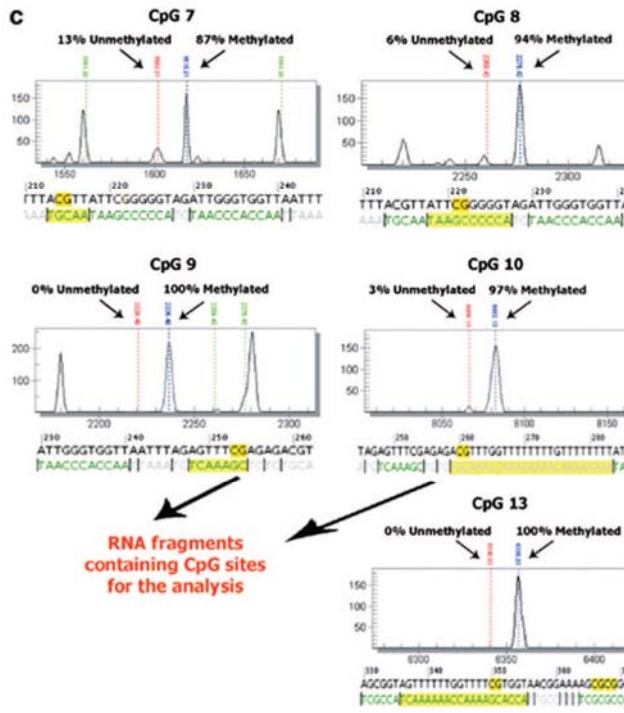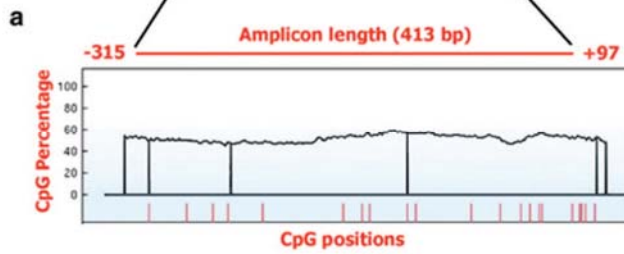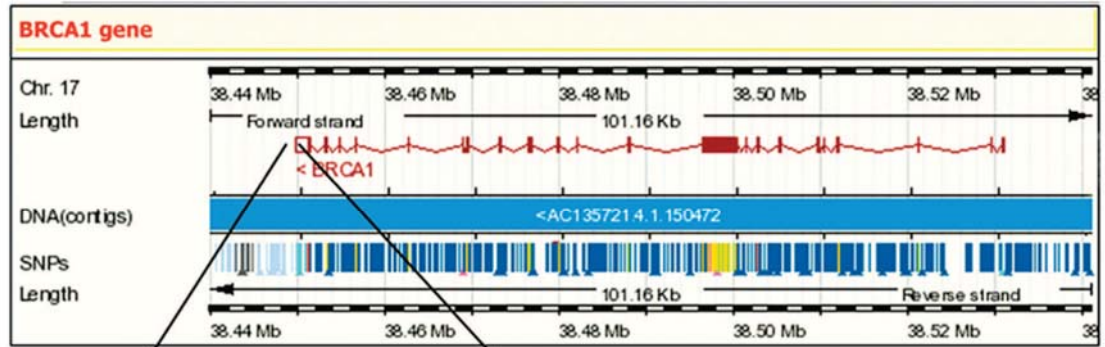
The top 10 hypermethylated genes are involved in cell cycle and DNA repair (*BRCA1*, *P16* and *P21*), invasion and metastasis (*CST6* and *TIMP3*), cell proliferation (*ESRb*), signal transduction (*APC*, *BIN1* and *BMP6*) and cell detoxification (*GSTP1*). Nine out of the 10 hypermethylated genes act as TSGs, whereas *ESRb* regulates cell proliferation and functions similar to the TSGs (OMIM database). Hypermethylation of TSGs causes the inactivation of the genes that are important in suppressing the development of most or all tumor types. Eckhardt *et al.* reported that about one-third of the differentially methylated 5′-UTRs are inversely correlated with transcription in normal tissues. In this study, our data showed that hypermethylated CpG sites on the top 10 genes in the cancerous samples were mostly located near the TATA box and BRE sequence (Figure 4), but in normal samples the CpG sites were differentially methylated and located randomly in the 5′-UTRs, which is similar to the observation in the previous study (Eckhardt *et al.*, 2006). The methylated CpG sites of the other 12 studied genes were distributed randomly in the 5′-UTRs and we could not find significant association between place of the CpG sites and conserved sequences. These data suggest that the hypermethylation may result in transcription alterations of those TSGs. Thus, inhibition of DNA hypermethylation and consequent reactivation of these genes is an attractive avenue for the development of novel therapeutics (Jones and Baylin, 2002; Suzuki and Bird, 2008). This strategy is particularly appealing because in normal cells, these genes are not normally regulated by DNA methylation. Therefore, the toxicity of DNA methylation inhibitors or DNA demethylation agents to noncancerous tissue could potentially be well below that of cytotoxic anticancer drugs. The targeted therapy should likely be more efficacious and safer for patients. Recently, two studies showed that the methylation process becomes transient, cyclical and dynamic (Kangaspeska *et al.*, 2008; Metivier *et al.*, 2008). Estrogen and anticancer drugs can cause changes in methylation status in breast cancer cell lines within 10 min (Kangaspeska *et al.*, 2008). Unlike the stable and unchangeable pathological somatic mutations, the cyclical and dynamic epigenetic nature may be useful for developing new strategies in the treatment of cancer patients, and in risk assessment of cases before the onset, through modification of pathological methylation patterns. To this end, our data identifying the top 10 hypermethylated genes would be important indicators in developing such strategies.

Aberrant DNA methylation patterns can also be used as biomarkers in cancer molecular diagnostics. In our study, the 10 hypermethylated genes enabled distinguishing cancerous breast tissues from normal breast tissues, suggesting that the genes may serve as classification markers for facilitating biopsy/diagnosis and staging and surgical margins in breast cancer management. Lofton-Day *et al.* (2008) successfully used colorectal cancer tissue-specific methylated DNA as a marker to identify tumor-derived methylated DNA in plasma for developing sensitive and specific blood-based tests applicable in screening and early detection of the disease. Tumor-specific cell-free hypermethylated p16 DNA in the serum of patients with colorectal cancer has shown prognostic value before and during therapy (Zou *et al.*, 2002; Nakayama *et al.*, 2003). Zhang *et al.* (2007) observed that hepatocellular carcinoma-specific methylation changes could be detectable in the patients' serum samples up to 9 years before the diagnosis of the condition. These studies suggest that cancer tissue-specific methylation changes can be used in developing blood-based tests for risk assessment, earlier diagnosis and monitoring of cancers. The breast cancer tissue-specific methylation changes identified by our study may provide valuable information for developing blood-based tests in management of the malignancy.

Technically, we successfully obtained the cancerous and normal breast tissue samples for this study by relative immunohistochemistry staining instead of using microdissection, which requires sophisticated and expensive equipment not applicable to many of diagnostic or research laboratories. Our method for the semiquantitative methylation analysis used in paraffin-embedded samples, which are mostly available to be obtained for subsequent analysis. In this study, we used only a single T-cleavage reaction instead of 2–4 cleavage reactions for assessment of the methylation status. Although the cost for a single cleavage reaction has been obviously reduced, the sensitivity of the modified approach has not been significantly decreased (Radpour *et al.*, 2008).

In this study, we did not find significant differences in methylation degrees between cancerous and normal tissues ($P > 0.05$) in the promoter regions of five analysed genes (*BRCA2*, *EGFR*, *NM23-H1*, *Prostasin* and *P53*) (Supplementary Data 2). According to previous reports, these genes were involved in breast carcinogenesis, mostly through mutations rather than hypermethylation (Sjoblom *et al.*, 2006; Trent and Touchman, 2007; Wood *et al.*, 2007).

As DNA methylation changes in human cancer are complicated and vary between the different types of cancer, every novel discovery in the epigenetic landscape and every development of an improved approach for accurate analysis of the events may provide a new

Figure 1 An example of high-throughput analysis of informative CpG sites for *BRCA1* tumor suppressor genes. (**a**) Gene location, amplicon size and place of CpG sites in the amplicon. (**b**) Methylation profile of CpG sites for *BRCA1* gene. The color of the circles is related to the percent of methylation in each CpG site. Arrows indicate the different methylation pattern between cancerous tissues and adjacent paired normal tissues. (**c**) Peaks show the methylation rates of the five CpG sites in the *BRCA1* gene for which we could find differences between cancerous and paired normal tissues. (**d**) The two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 normal breast tissues (red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0 to 100 and black clusters indicate not analysed CpG sites). The complete data for the other 21 studied genes is summarized in Supplementary Data 2. A full colour version of this figure is available at the *Oncogene* journal online.

A

Methylation Profiles of the 22 Genes

A1
Double dendrogram

A2
Dendrogram


A3

B

B1

B2

B3
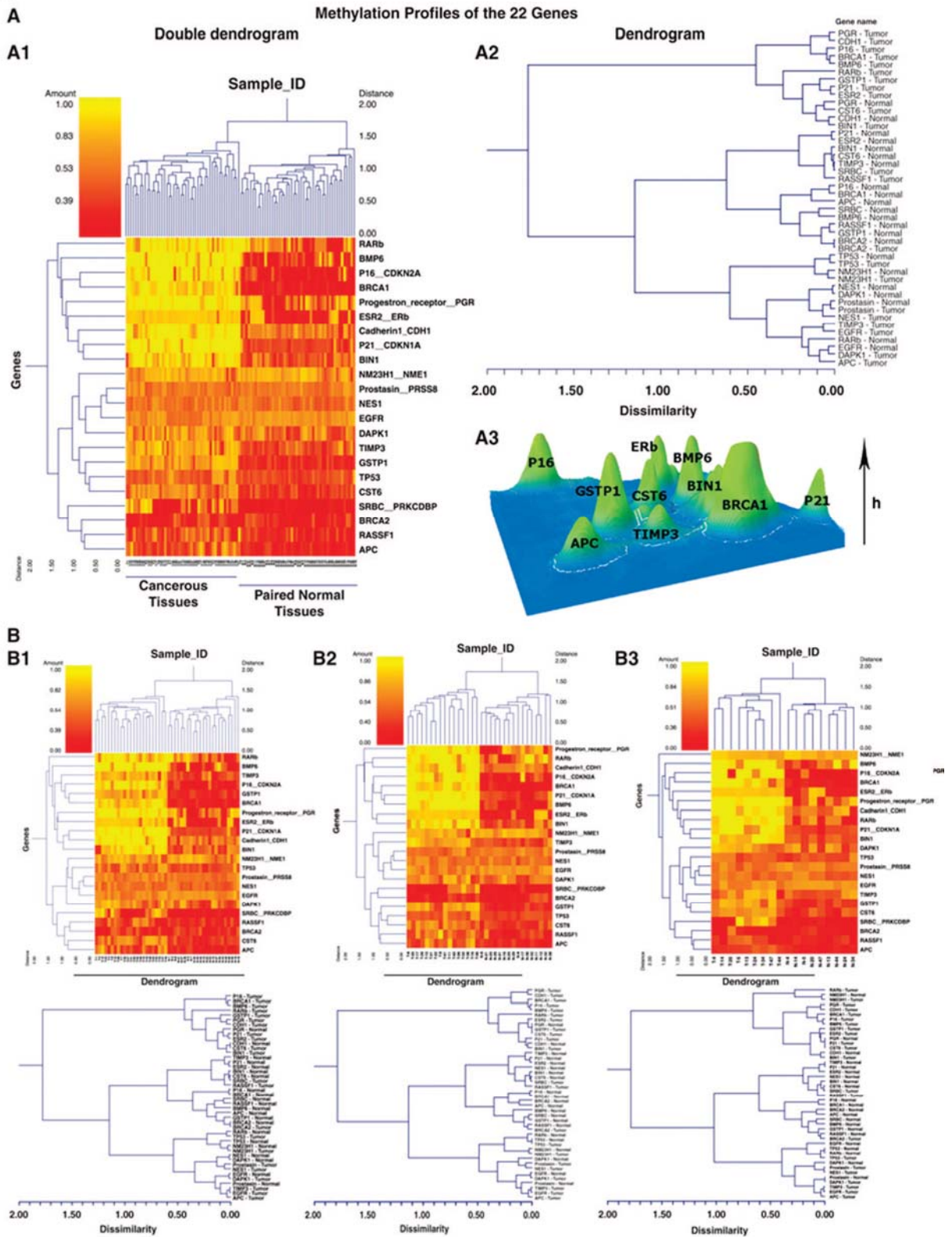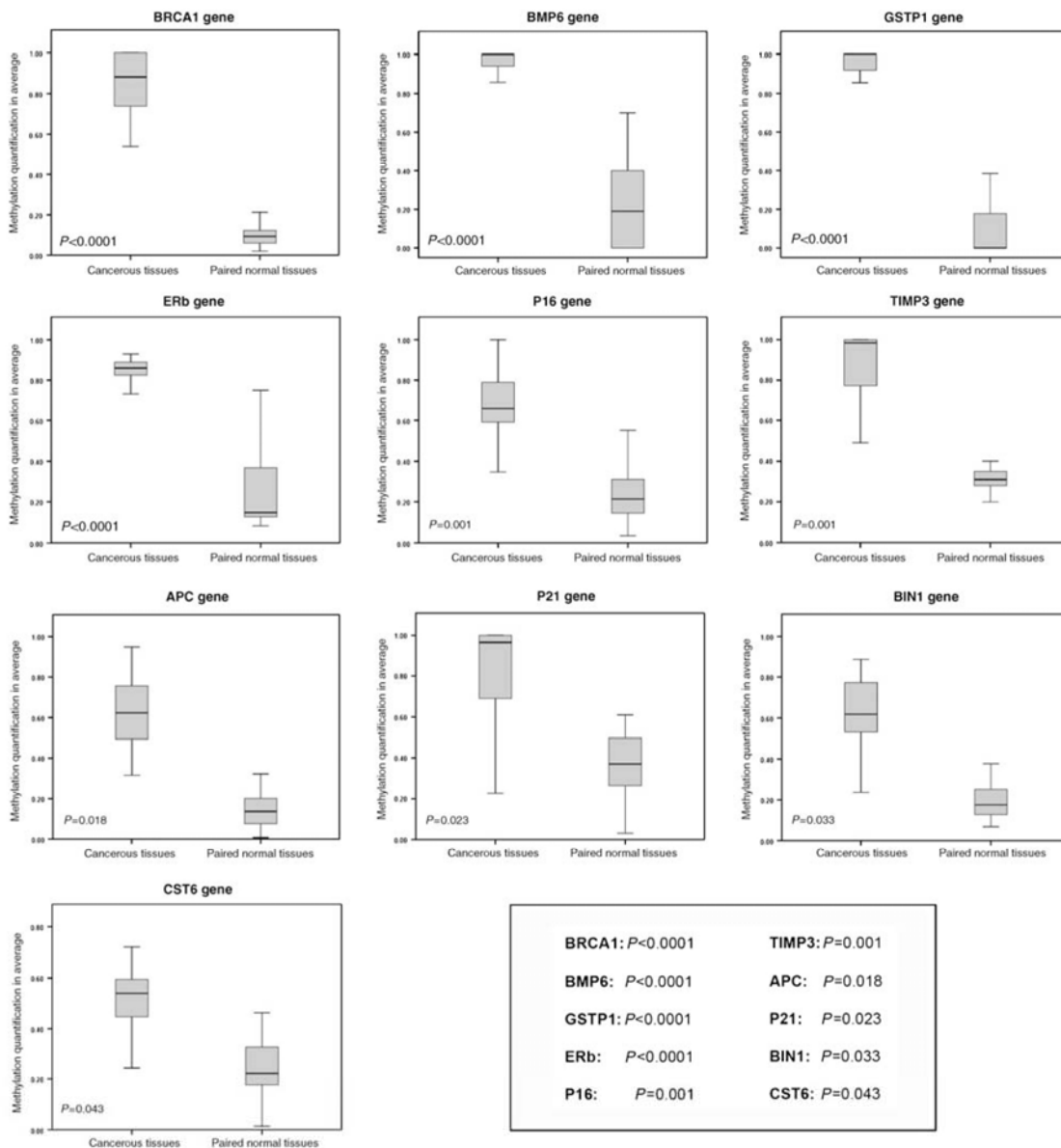
Figure 3 Comparison between semiquantitative analysis of methylation for the top 10 hypermethylated genes in the breast cancerous tissues and paired normal tissues. Our data showed the significant differences of methylation patterns between for cancerous and normal tissues ($P<0.05$).

Figure 2 (**A**) High-throughput methylation analysis of informative CpG sites for 22 breast cancer-related genes. (A1) Double dendrogram presents the methylation profiles of 22 studied genes in 46 cancerous tissues and 46 paired normal tissues. (A2) Dendrogram to calculate distances between methylation profile of 22 genes in cancerous and paired normal tissues. (A3) Schematic graph of top 10 hypermethylated genes in breast cancer. The height (h) and width of the mountains indicates the methylation grade of each gene. (**B**) High-throughput methylation analysis of informative CpG sites for 22 breast cancer-related genes in the three subgroups of patients. Double dendrograms present the methylation profiles of 22 studied genes. Dendrograms calculate distances between methylation profile of 22 genes. (B1) Data for 23 invasive ductal carcinoma tissues and paired normal tissues. (B2) Data for 16 infiltrating ductal carcinoma tissues and paired normal tissues. (B3) Data for nine infiltrating lobular carcinoma tissues and paired normal tissues.

Table 2 Patient's response to screening receptors/markers in the three breast cancer tumor subgroups

Breast cancer tumor type	Receptors and follow-up parameters for breast cancer											
	ER-α		ER-β		HER2		P53		Ki62		PS2	
	+	−	+	−	+	−	+	−	+	−	+	−
Invasive ductal carcinoma	8	15	6	17	13	10	2	21	0	23	7	16
Infiltrating ductal carcinoma	4	12	2	14	5	11	1	15	0	16	3	13
Infiltrating lobular carcinoma	2	7	0	9	2	7	0	9	0	9	1	8
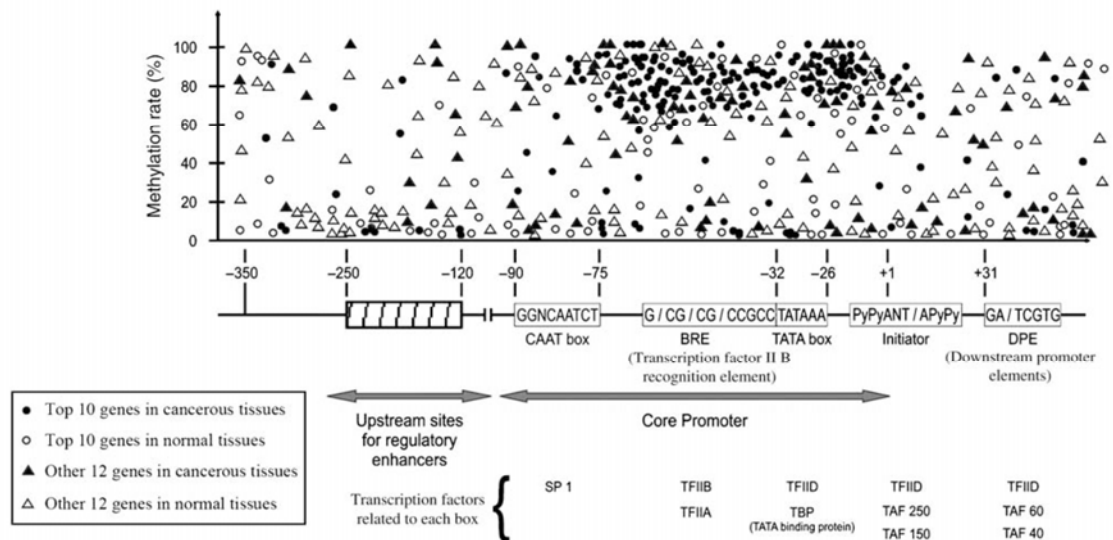
Figure 4 Comparison of the methylation rate and place of CpG sites according to the sequences and recognition sites of the transcription factors in the 22 studied genes for 48 cancerous and 48 normal tissue samples.

opportunity for targeted agents and accurate diagnosis in human cancer. We hope that the breast cancer tissue-specific hypermethylated genes identified in this study by the MALDI-TOF MS will provide valuable information in their application for the classification, sensitive detection, risk and therapy assessment of breast cancer, as well as for developing novel targeted therapeutic strategies through modification of pathological epigenetic changes in cancer.

Materials and methods

Samples
The study was approved by the local institutional review board. A total of 96 paraffin-embedded different tissue samples (48 breast cancerous tissues and 48 paired normal tissues) were collected from 48 patients with breast cancer. Paraffin-embedded sections were examined by two experienced pathologists. None of the studied cases have a hereditary form of breast cancer. The criteria for checking hereditary patients were germline mutations in *BRCA1* or *BRCA2*, and in primary invasive breast carcinoma. According to the pathological tumor status and immunohistochemistry staining, we classified our patients in three subgroups (invasive ductal carcinoma,

infiltrating ductal carcinoma and infiltrating lobular carcinoma). Table 2 and 3 summarize the clinical data of patients and their response to the screening receptors/markers.

DNA extraction was performed from 3–5 sections of each 10-μm-thick paraffin-embedded sample (around 0.01–0.02 g of tissue), using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany).

Bisulfite treatment
The EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) was used for bisulfite conversion of the target sequences. The C/T conversion reaction was performed using the PCR program as follows: 95 °C for 30 s and 50 °C for 15 min, which was repeated for 46 cycles.

Primer design and PCR tagging for EpiTYPER assay
We designed primers for the 22 genes to cover the regions with the most CpG sites. Our selected amplicons were mostly located in the promoter region of genes or started from the promoter and ended in the first exon. For three specific genes (*ESR-b*, *PGR* and *TIMP3*), the amplicons that contained CpG sites were located in the first exon and there were no CpG sites in the promoter region. The primers were designed using MethPrimer (San Francisco, CA, USA) (Li and Dahiya, 2002). In PCR amplification, a T7-promoter tag was added to the reverse primer and a 10mer-tag sequence was added to the

Table 3 Clinical data of patients divided into three subgroups according to the type of breast cancer tumor

Breast cancer tumor type	Total no. of patients	Age (years) (mean ± s.d. [range])	Side of breast tumor		Tumor size (cm) (mean ± s.d. [range])	No. of patients with lymph node involvement	No. of patients with metastasis	Histological grade		
			R	L				Grade 1	Grade 2	Grade 3
Invasive ductal carcinoma	23	48 ± 11.2 (35–78)	10	13	3.6 ± 3.1 (1–12)	20	7	3	9	11
Infiltrating ductal carcinoma	16	49 ± 12.1 (32–74)	8	8	2.3 ± 1.9 (0.8–9)	12	2	4	5	7
Infiltrating lobular carcinoma	9	50 ± 10.9 (32–65)	5	4	3.25 ± 1.6 (1.5–6)	5	1	2	4	3

Abbreviations: L, left; R, right.

forward primer to balance the PCR primer length. The primer sequences, annealing temperatures (T_a) and PCR conditions are described in Supplementary Data 1.

In vitro transcription and T-cleavage assay

Unincorporated dNTPs were dephosphorylated by adding 1.7 μl H_2O and 0.3 U shrimp alkaline phosphatase (SEQUE-NOM, San Diego, CA, USA). The reaction was incubated at 37 °C for 20 min and shrimp alkaline phosphatase was then heat inactivated at 85 °C for 10 min. Typically, 2 μl of the PCR were directly used as a template in a 5-μl transcription reaction. T7 R&DNA Polymerase (20 U) (Epicentre, Madison, WI, USA) was used to incorporate dTTP in the transcripts. Ribonucleotides were used at 1 mmol/l and the dNTP substrate at 2.5 mmol/l. In the same step, the RNase-A enzyme (SEQUENOM) was added to cleave the in vitro transcripts (T-cleavage assay). The mixture was further diluted with H_2O to a final volume of 27 μl. Conditioning of the phosphate backbone was achieved by adding 6 mg of Clean Resin (SEQUENOM) before performing MALDI-TOF MS.

Mass spectrometry

After robotically dispensing 22 nl of cleavage reaction onto silicon matrix preloaded chips (SpectroCHIP; SEQUENOM), the mass spectra were collected using a MassARRAY Compact MALDI-TOF (SEQUENOM) and spectra's methylation ratios were generated by the EpiTYPER software v1.0 (SEQUENOM).

Statistical methods

Relative methylation was compared between cancerous and paired normal tissues using the Wilcoxon signed-rank test, a nonparametric counterpart of the paired t-test. The one-way analysis of variance test was used to compare relative methylation between three subgroups. Using the two-way hierarchical cluster analysis, the most variable CpG fragments for each gene were clustered based on pair-wise Euclidean distances and linkage algorithm for all of the 96 tissue samples

(48 cancerous tissues and 48 paired normal tissues), according to the previously developed method (Radpour et al., 2008). The gene clustering was performed independently using a hierarchical clustering algorithm. For gene clustering, pairwise similarity metrics were calculated for each gene separately on the basis of methylation ratio of cancerous tissues across the adjacent paired normal tissues. The procedure was performed using the double dendrogram function of the Gene Expression Statistical System for Microarrays (GESS, version 7.1.13; NCSS, Kaysville, UT, USA).

Additional information

The sequence of PCR-tagged primers for in vitro transcription and PCR conditions are available in Supplementary Data 1. The complete data for high-throughput methylation analysis of informative CpG sites in 22 breast cancer-related genes, including gene location, amplicon size and place of CpG sites in the amplicon and two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 normal breast tissues are available in Supplementary Data 2.

Conflict of interest

The authors declare no conflict of interest.

References

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M et al. (2006). DNA methylation profiling of human chromosomes 6 20 and 22. Nat Genet **38**: 1378–1385.

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G et al. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA **102**: 15785–15790.

Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M, Cantor C et al. (2008). Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci USA **105**: 4844–4849.

Issa JP. (2004). CpG island methylator phenotype in cancer. Nat Rev Cancer **4**: 988–993.

Jones PA, Baylin SB. (2002). The fundamental role of epigenetic events in cancer. Nat Rev Genet **3**: 415–428.

2978

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP *et al.* (2008). Transient cyclical methylation of promoter DNA. *Nature* **452**: 112–115.

Laird PW. (2003). The power and the promise of DNA methylation markers. *Nat Rev Cancer* **3**: 253–266.

Li LC, Dahiya R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**: 1427–1431.

Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M *et al.* (2008). DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem* **54**: 414–423.

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**: 45–50.

Nakayama H, Hibi K, Takase T, Yamazaki T, Kasai Y, Ito K *et al.* (2003). Molecular detection of p16 promoter methylation in the serum of recurrent colorectal cancer patients. *Int J Cancer* **105**: 491–493.

Radpour R, Haghighi MM, Fan AX, Torbati PM, Hahn S, Holzgreve W *et al.* (2008). High-throughput hacking of the methylation patterns in breast cancer by *in vitro* transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip. *Mol Cancer Res* **6**: 1702–1709.

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD *et al.* (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* **314**: 268–274.

Srinivasan M, Sedmak D, Jewell S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* **161**: 1961–1971.

Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N *et al.* (2004). High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res* **14**: 126–133.

Suzuki MM, Bird A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**: 465–476.

Trent JM, Touchman JW. (2007). Cancer. The gene topography of cancer. *Science* **318**: 1079–1080.

Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ *et al.* (2007). The genomic landscapes of human breast and colorectal cancers. *Science* **318**: 1108–1113.

Zhang YJ, Wu HC, Shen J, Ahsan H, Tsai WY, Yang HI *et al.* (2007). Predicting hepatocellular carcinoma by detection of aberrant promoter methylation in serum DNA. *Clin Cancer Res* **13**: 2378–2384.

Zou HZ, Yu BM, Wang ZW, Sun JY, Cang H, Gao F *et al.* (2002). Detection of aberrant p16 methylation in the serum of colorectal cancer patients. *Clin Cancer Res* **8**: 188–191.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)

Supplementary Data 1

The sequence of PCR tagged primers for *in vitro* transcription and PCR conditions

For the PCR on bisulphite-treated genomic DNA (gDNA), the following PCR conditions were used: 1x: 95°C for 10 min; 48x: 95°C for 20s, Ta for 30s, 72°C for 1 min; 1x 72°C for 5 min. For the PCR on gDNA, the following cycling conditions were used: 1x: 95°C for 10 min; 40x: 95°C for 20 s, Ta for 30s, 72°C for 45s; 1x 72°C for 5 min. After first round of PCR amplification we used 1.5µl of the PCR product for re-amplification of selected amplicons with the same primer pairs.

| Gene	Primer	Sequence (5'→3')	Length	T_a	Prodoct Size (bp)
APC	tag-EN1-FW	AGGAAGAGAGATTGTTTTTTTGTGTTGTAAAAATTAT	27+10	58	420
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCACCTCCATTCTATCTCCAATAAC	24+31		
BIN1	tag-EN1-FW	AGGAAGAGAGGGAGGTGAGTTTTTGGAA	18+10	58	330
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCTACCTTTTAAAAAACCACTCC	22+31		
BMP6	tag-EN1-FW	AGGAAGAGAGGGGGTAAATTTTATGGTGGTTT	22+10	57	397
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCCTTCCTAACCCTCAATCCTTA	22+31		
BRCA1	tag-EN1-FW	AGGAAGAGAGAATTGGAGATTTTTATTAGG	20+10	56	413
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAAATCTCAACRAACTCAC	18+31		
BRCA2	tag-EN1-FW	AGGAAGAGAGTTGGGATGTTTGATAAGGAATTT	23+10	61	433
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAAACAAAAAAACAAAAAACRC	22+31		
CADHERIN 1	tag-EN1-FW	AGGAAGAGAGGGTGAAAGAGTGAGATTTTATTTTTA	26+10	57	500
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCTCCAAAAACCCATAACTAACC	22+31		
CST6	tag-EN1-FW	AGGAAGAGAGGTTGGTAGTTTATTTTGGATAGTTT	25+10	59	445
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCAAATCCCRAAATTCTCC	18+31		
DAPK1	tag-EN1-FW	AGGAAGAGAGTTTGTAATTTTAGTATTTTGGGAGG	25+10	56	408
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAAATACACATTAAAATCCAAAACAA	25+31		
EGFR	tag-EN1-FW	AGGAAGAGAGTYGATTTGGATATAGGTTGGG	21+10	62	428
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCCRAAACTAACTCRAAACTCC	21+31		
ESR2 (ER beta)	tag-EN1-FW	AGGAAGAGAGTTTAGTTGTTGGTTTTTTGGAT	22+10	58	374
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAAAATTTCAAACAAAATAAAACAATT	26+31		
GSTP1	tag-EN1-FW	AGGAAGAGAGTTYGGGAGGTTGAAGTAGA	19+10	60	381
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAAACAAACAACAAAAAAAAACC	23+31		
NES1 (Island 1)	tag-EN1-FW	AGGAAGAGAGATATAGATGGTAGGGAGGGTG	21+10	56	546
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTATACAAACTTCTCCCAATACCC	22+31		
NES1 (Island 2)	tag-EN1-FW	AGGAAGAGAGGTTAGGGGTTTTTGGGTAGAG	21+10	61	410
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAACACAATTACCCTAATAACRCC	23+31		
Nm23-H1 (NME1)	tag-EN1-FW	AGGAAGAGAGTAGGTATTTAAAATTTTTTGTTTGA	25+10	57	421
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTATTTAAACTCCRACTACAACC	21+31		
P16 (CDKN2A)	tag-EN1-FW	AGGAAGAGAGGGTTGTTTTTGGTAGGG	17+10	58	580
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTATATAAACCACRAAAACCC	19+31		
P21 (CDKN1A)	tag-EN1-FW	AGGAAGAGAGGGTAAATTTTTGTTTGTTAGAGTGG	25+10	60	419
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTTAACTTCRACAACTACTCACACCT	24+31		
Progesterone receptor (PGR)	tag-EN1-FW	AGGAAGAGAGTTYGTTTTAAAGATAAAGGAGGAG	24+10	61	448
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCCTTAAACACRACRACCTAATAACC	25+31		
Prostasin (PRSS8)	tag-EN1-FW	AGGAAGAGAGGGGGTATAATTGGTTTGAGATA	22+10	57	450
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTTCCTCCCAAAAATAACTACACCT	23+31		
RAR-b	tag-EN1-FW	AGGAAGAGAGGAGTGTATGTTAATGGGGGAG	21+10	56	475
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTTTCCCAACATAATTTCTCTAC	21+31		
RASSF1	tag-EN1-FW	AGGAAGAGAGGGGYGGTAAAGTTGTTGA	18+10	56	346
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCRCAATAAAAACCTAAATACA	21+31		
SRBC (PRKCDBP)	tag-EN1-FW	AGGAAGAGAGTTTTTGTAGTGGAGAATTGAAATAGG	26+10	57	435
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCAACATAAAAACCAACTTCTCCAAC	25+31		
TIMP3	tag-EN1-FW	AGGAAGAGAGTTTTGTTATTGGTTTGAGGG	20+10	59	441
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTCCAAACTCCAACTACCCA	18+31		
TP53	tag-EN1-FW	AGGAAGAGAGATGGTTTYGAAGTTTTTAGGGAT	23+10	59	449
	T7-EN1-RV	CAGTAATACGACTCACTATAGGGAGAAGGCTAATACAAAACCTACTACRCCCTCT	24+31		

Supplementary Data 2

High-throughput methylation analysis of informative CpG sites in 22 genes related to breast cancer.

APC gene

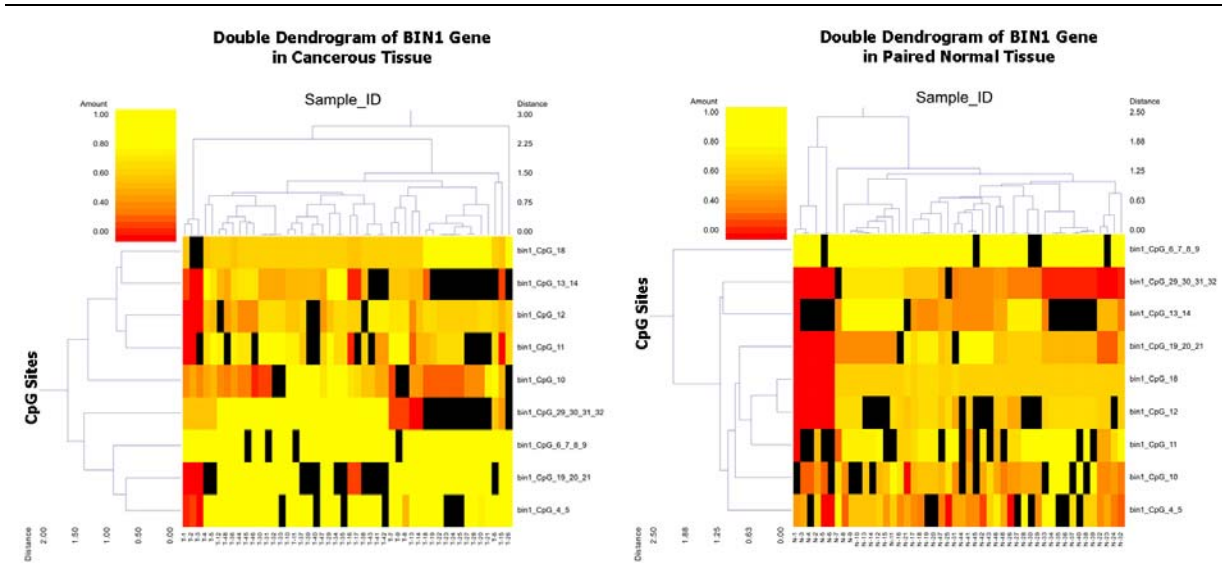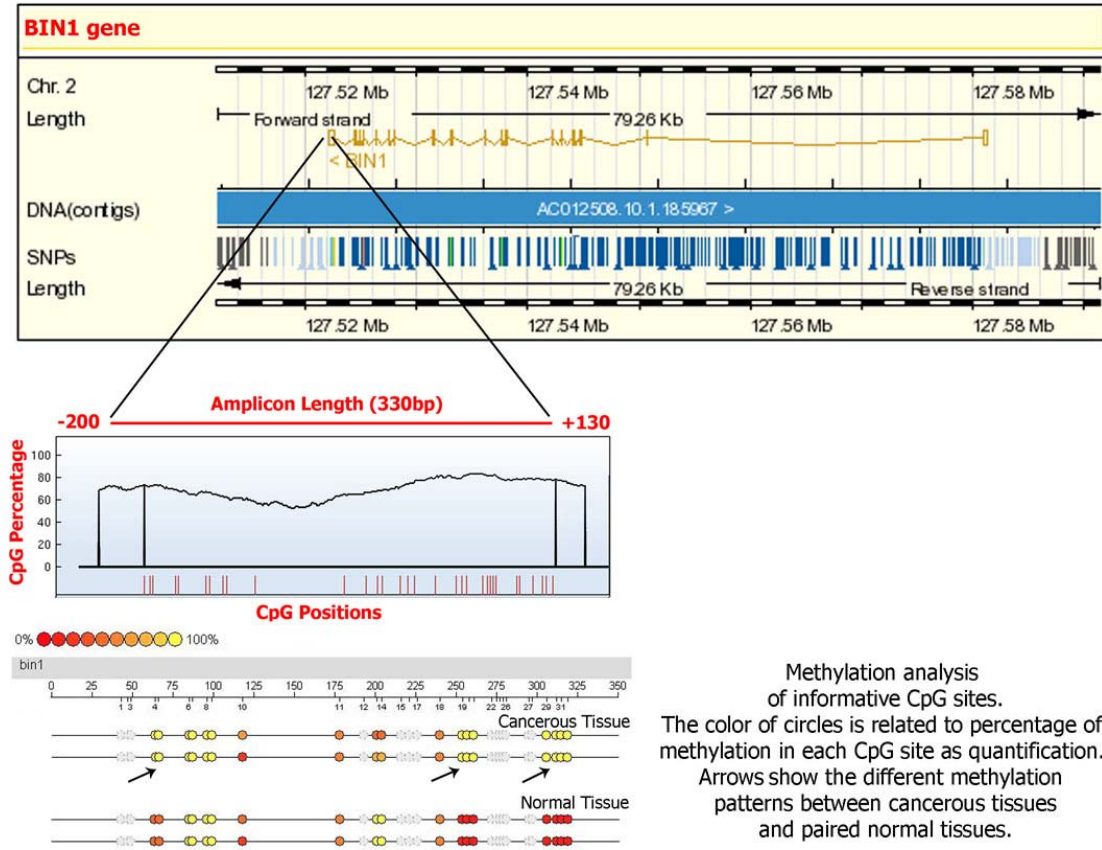| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
324	adenomatous polyposis coli	5q21-q22	Cell adhesion, signal transduction, stabilization of the cytoskeleton, regulation of cell cycle and apoptosis	direct
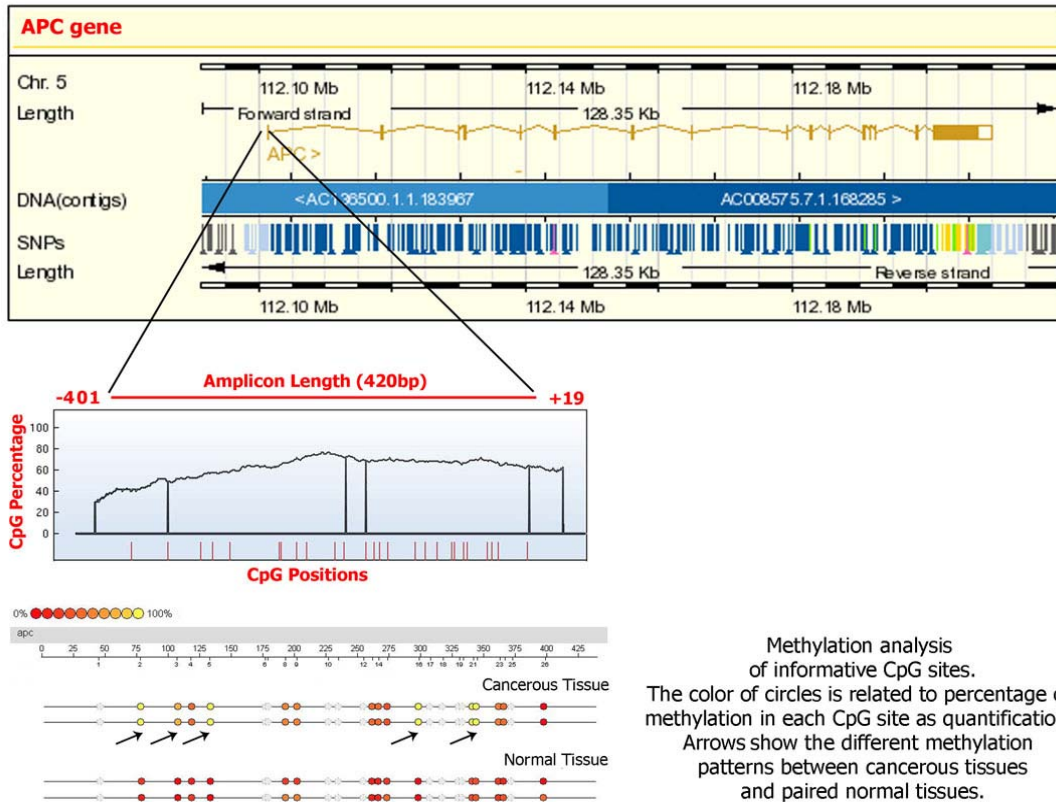
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
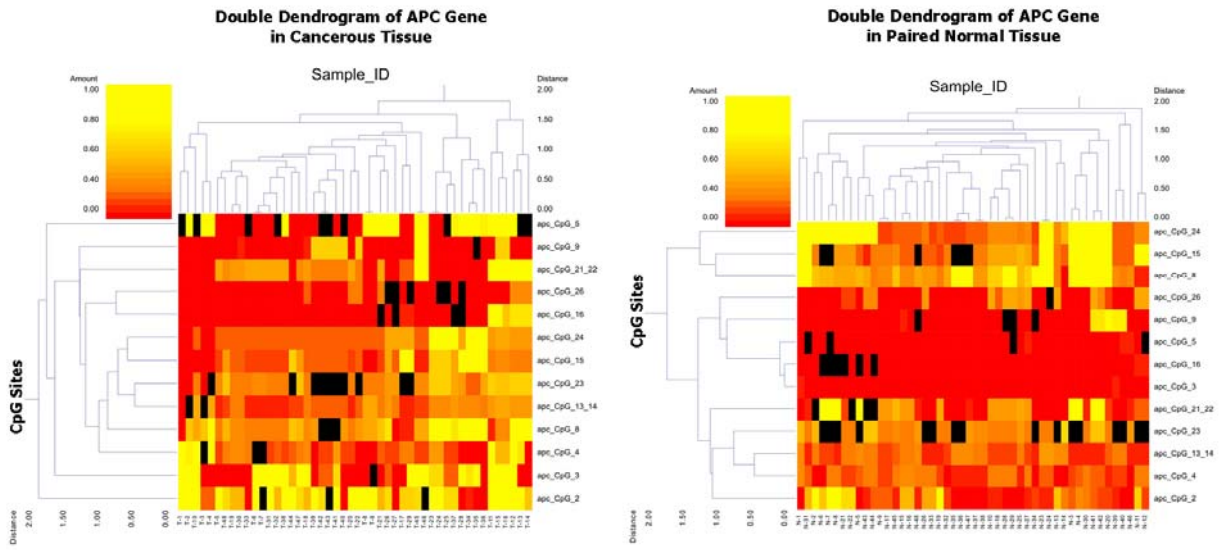
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

BIN1 gene

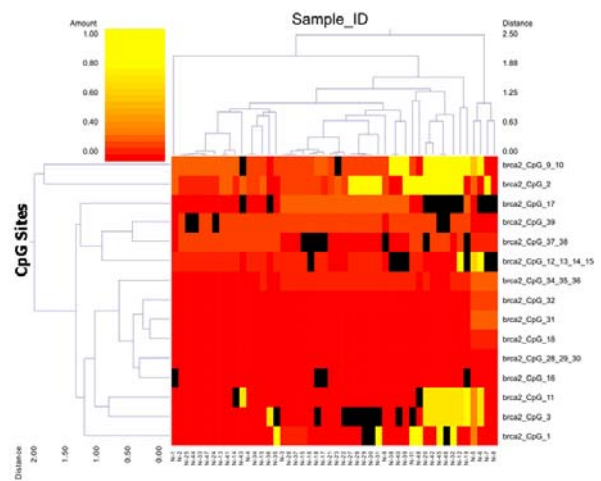| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
274	Bridging integrator 1	2q14	Encodes several isoforms of a nucleocytoplasmic adaptor protein, one of which was initially identified as a MYC-interacting protein with features of a tumor suppressor.	direct

Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

BMP6 gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
654	bone morphogenetic protein 6	6p24-p23	The bone morphogenetic proteins (BMPs) are a family of secreted signaling molecules.	direct
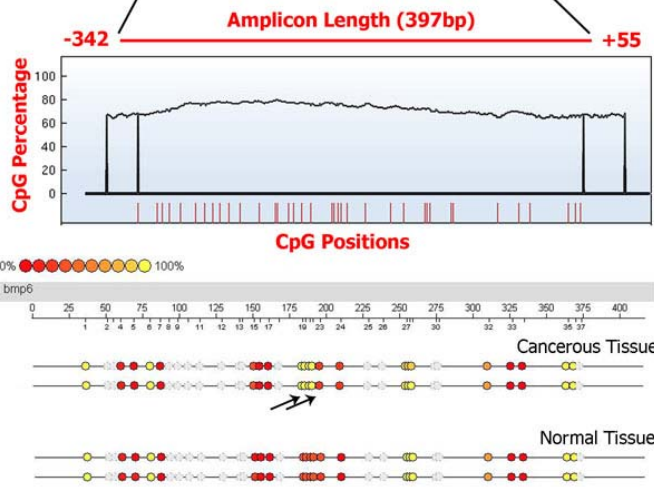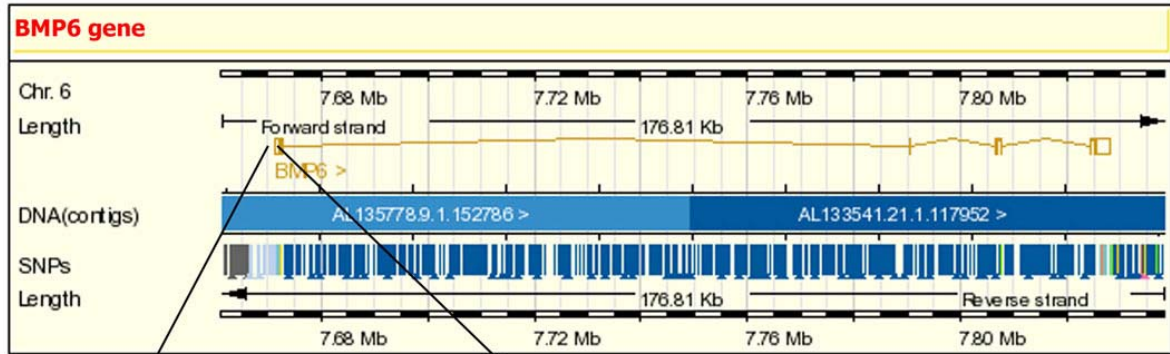
Methylation analysis
of informative CpG sites.
The color of circles is related to percentage of
methylation in each CpG site as quantification.
Arrows show the different methylation
patterns between cancerous tissues
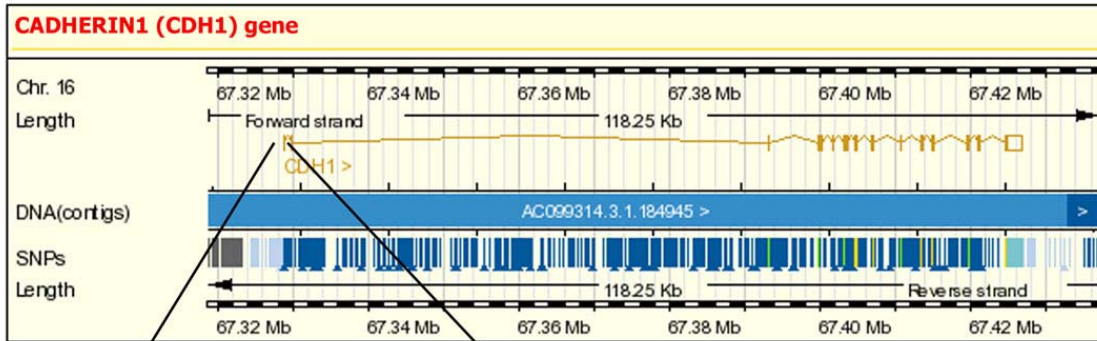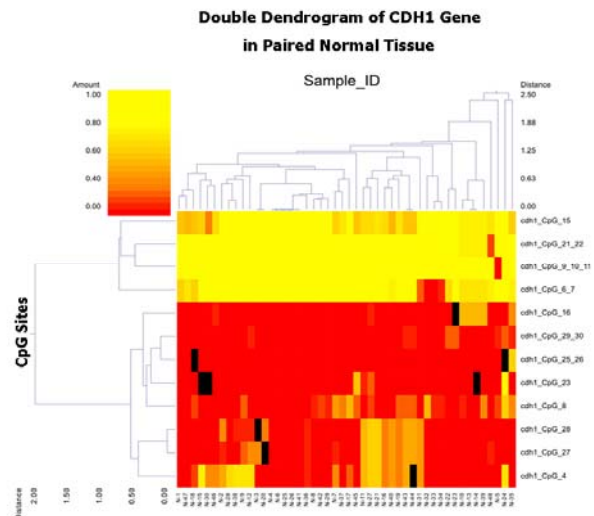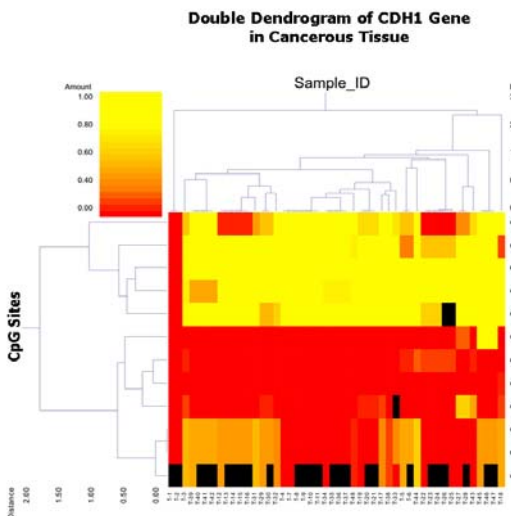and paired normal tissues.
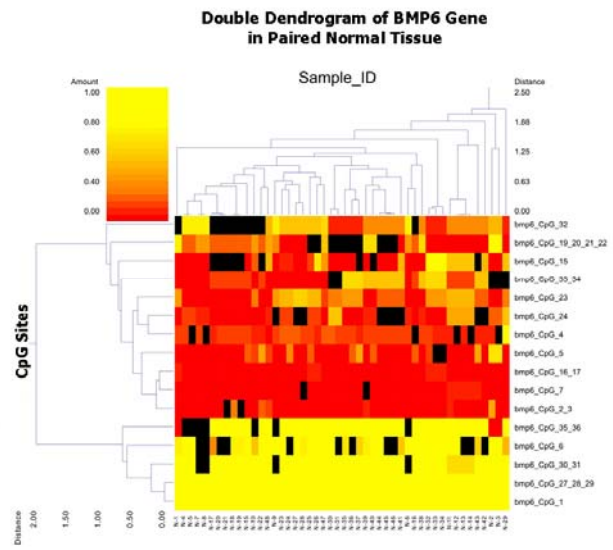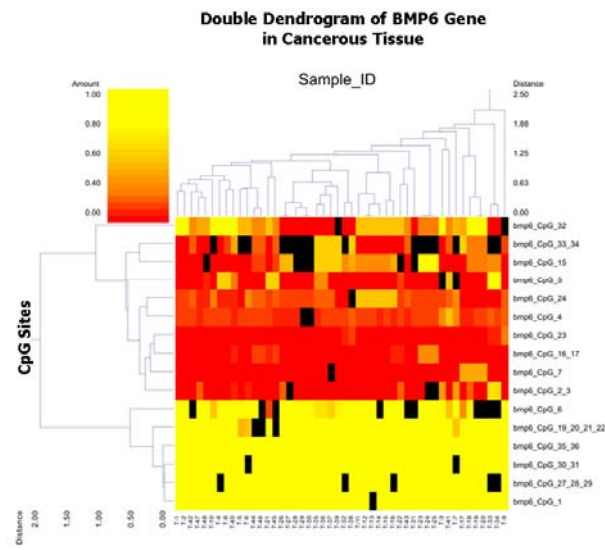
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

$BRCA1$ gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
672	Breast Cancer type 1	17q21	Involved in DNA repair, recombination, checkpoint control of the cell cycle and transcription. Interacts with p53, STAT-factors, SRBC, etc.	direct
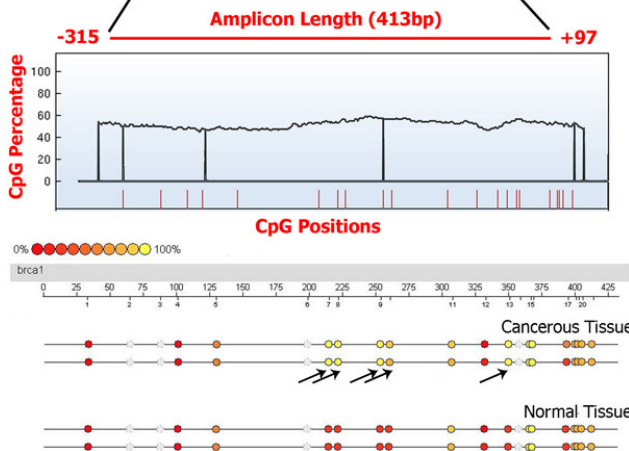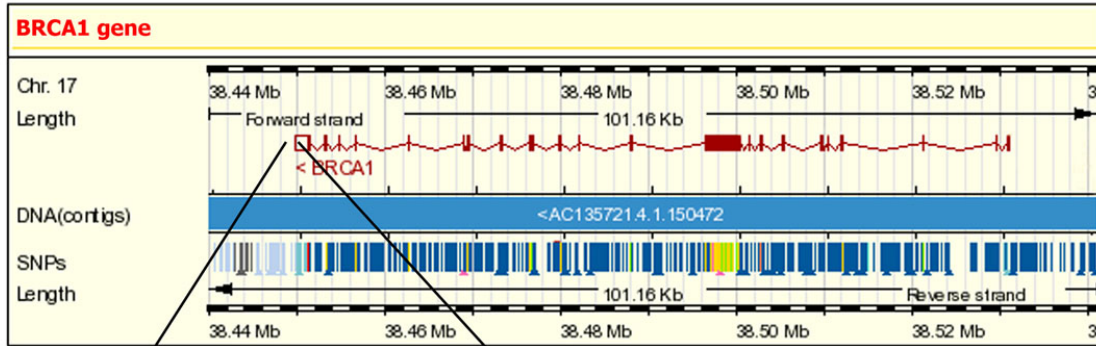
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
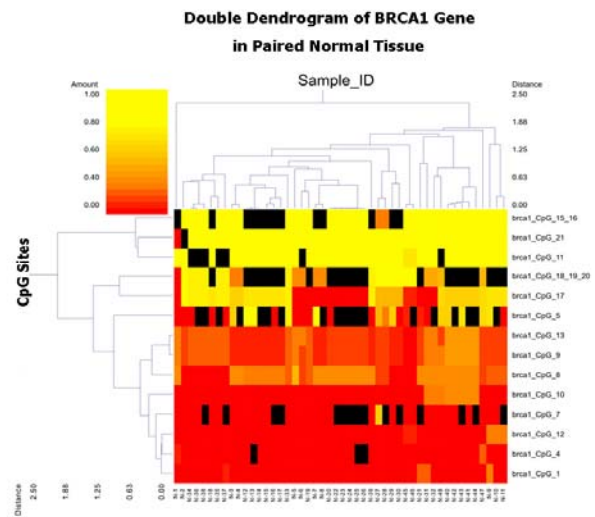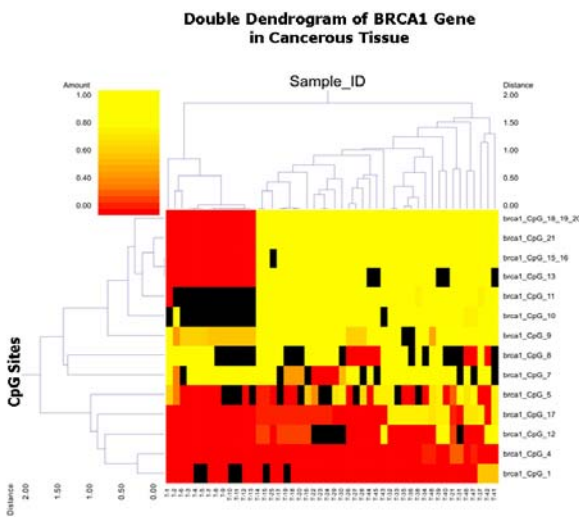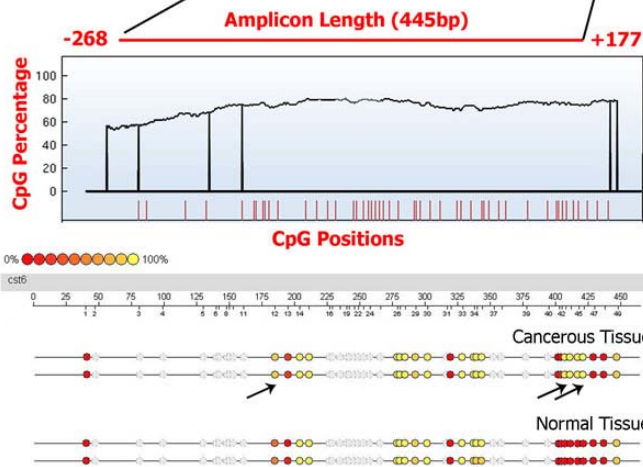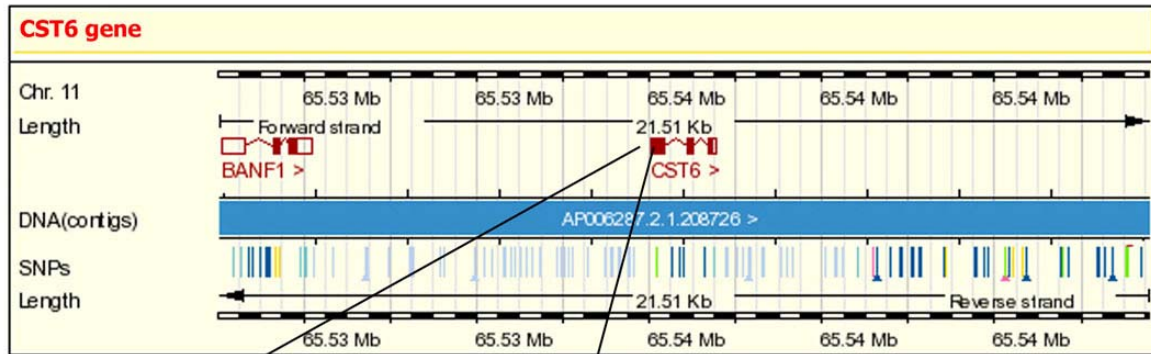

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

BRCA2 gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
675	breast cancer type 2	13q12.3	Involved in DNA repair, recombination, checkpoint control of the cell cycle and transcription.	direct
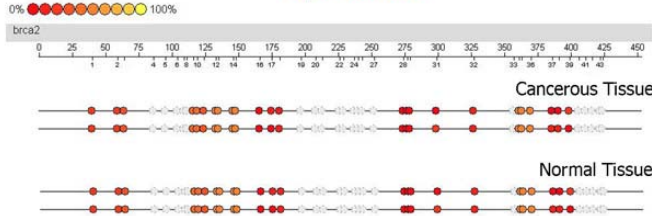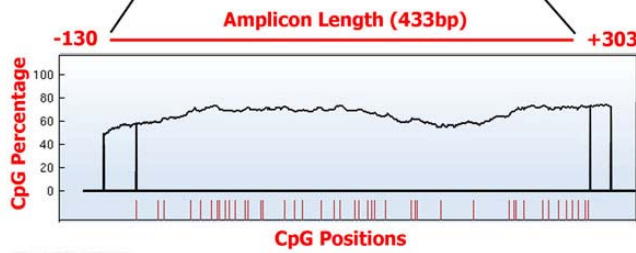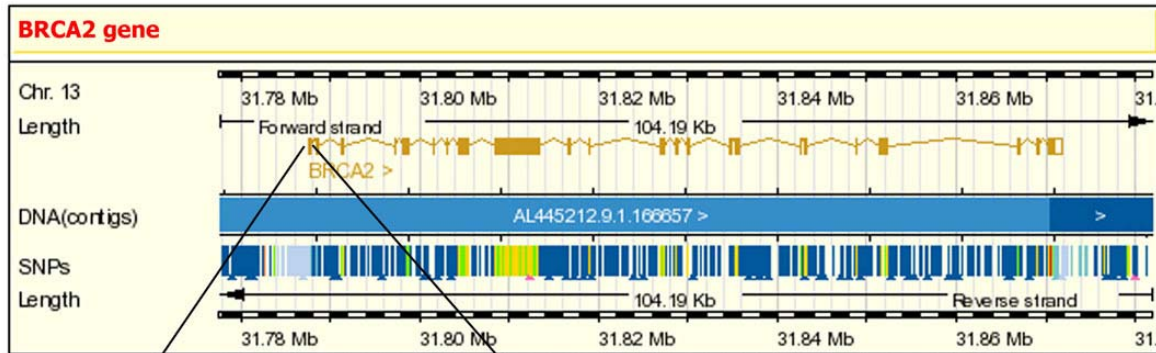
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. There is no difference in methylation patterns between cancerous tissues and paired normal tissues.
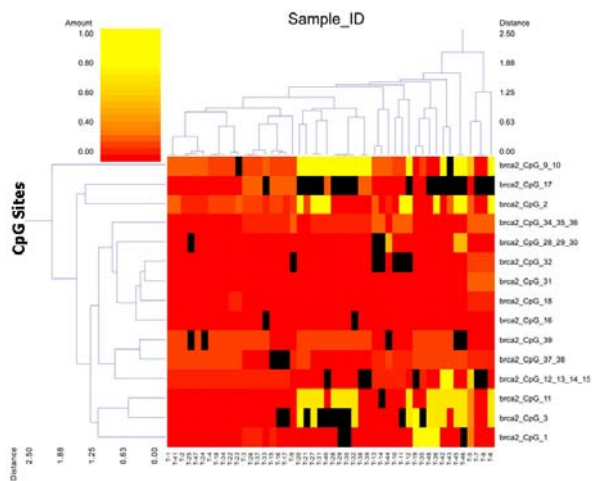
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

CADHERIN1 (CDH1) gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
999	cadherin 1, type 1, E-cadherin	16q22.1	The encoded protein is a calcium dependent cell-cell adhesion glycoprotein. Loss of function is thought to contribute to progression in cancer by increasing proliferation, invasion, and/or metastasis.	direct
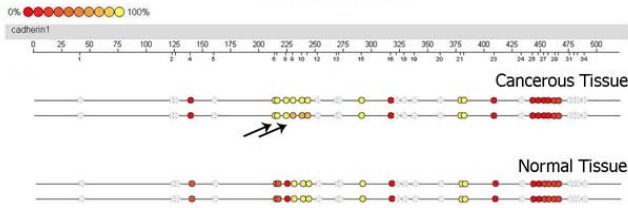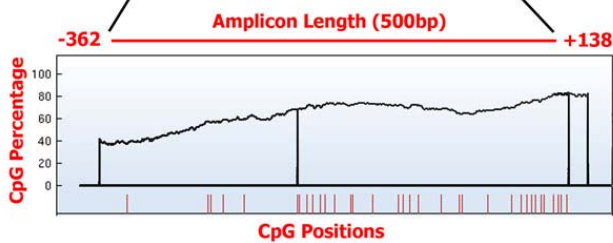
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

CST6 gene

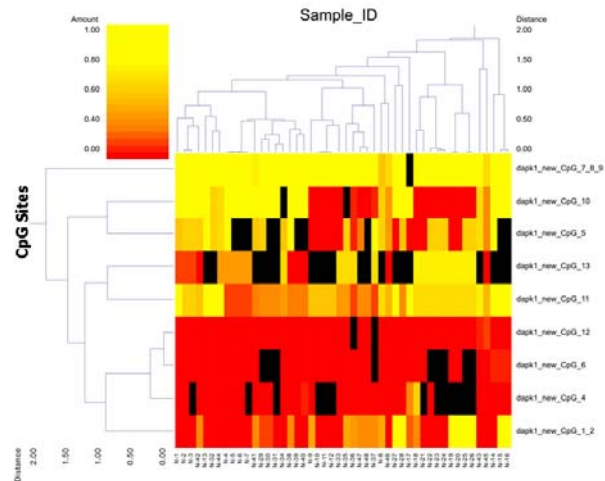| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
1474	Cystatin E/M	11q13	This gene encodes a cystatin from type 2 family, which is down-regulated in metastatic breast tumor cells as compared to primary tumor cells. Loss of expression is likely associated with the progression of a primary tumor to a metastatic phenotype.	direct

Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
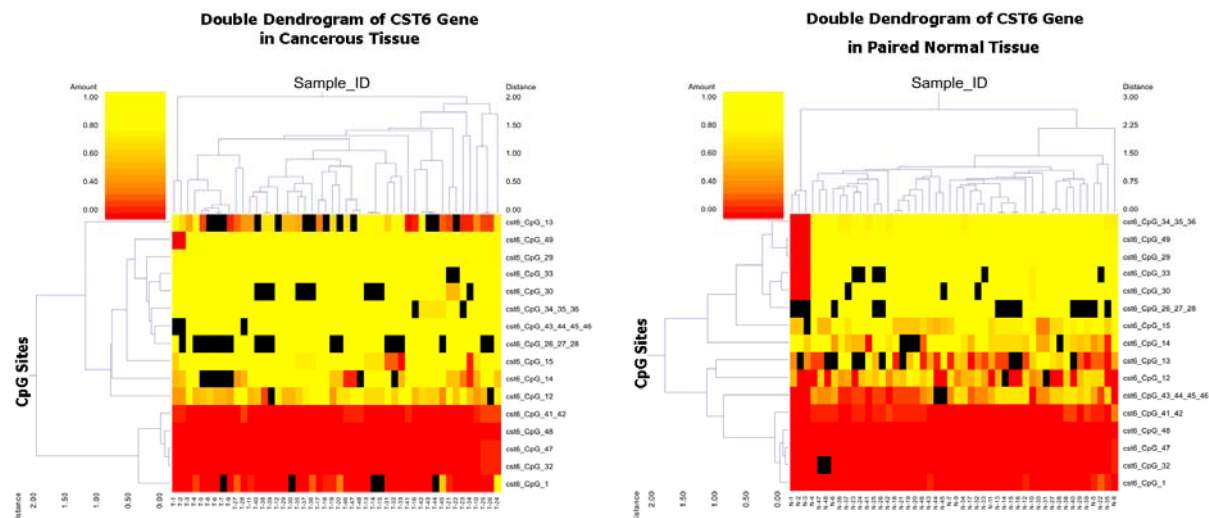
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

DAPK1 gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
1612	Death-associated protein kinase 1	9q34.1	Mediator of interferon-γ induced apoptosis and it is a tumor suppressor candidate.	direct
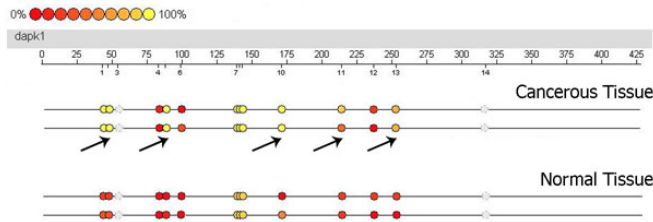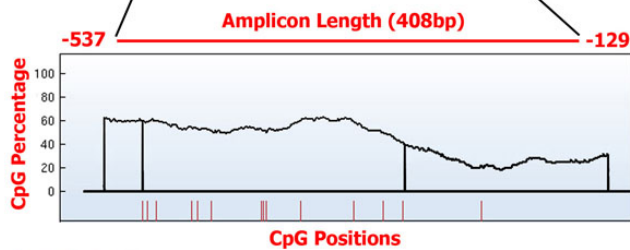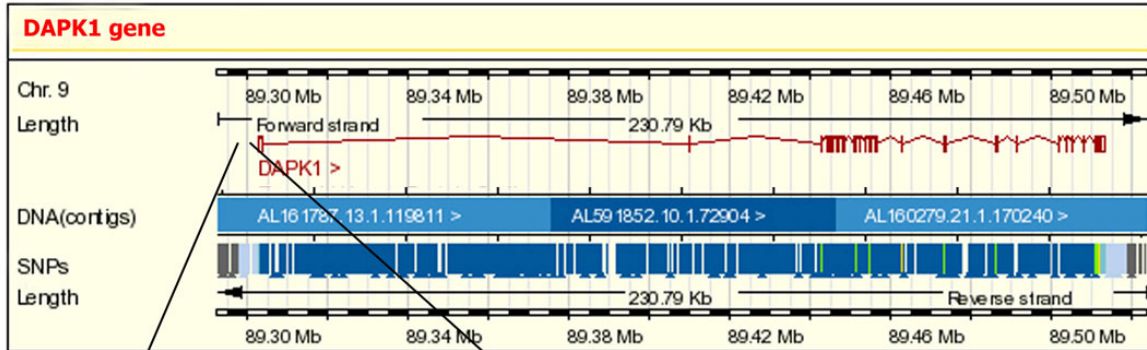
Methylation analysis
of informative CpG sites.
The color of circles is related to percentage of
methylation in each CpG site as quantification.
Arrows show the different methylation
patterns between cancerous tissues
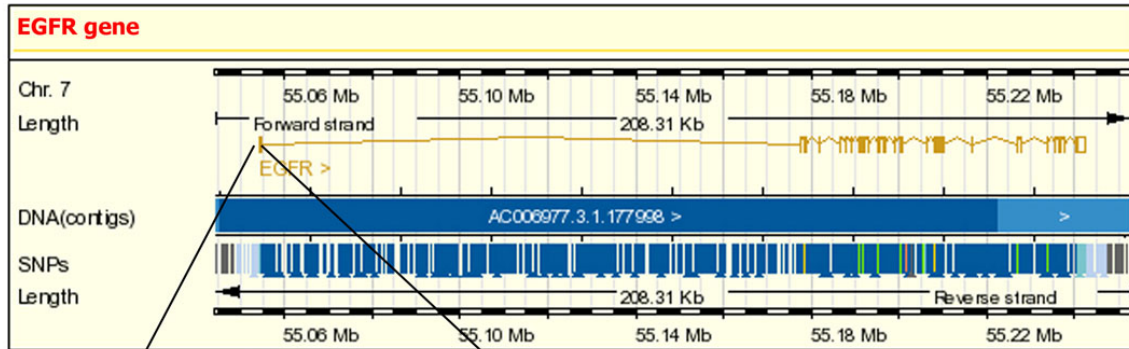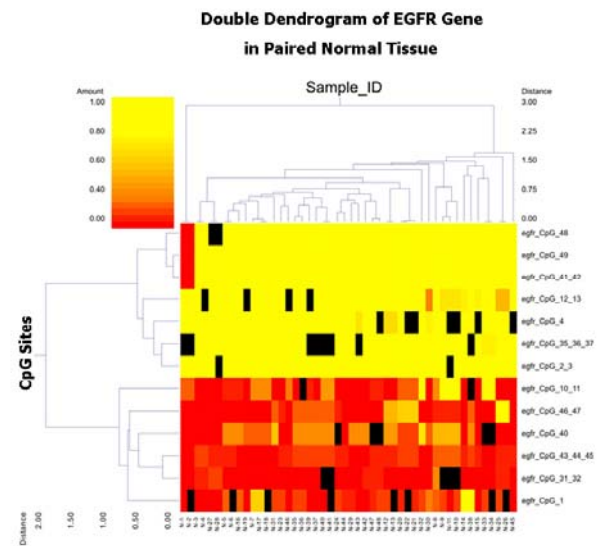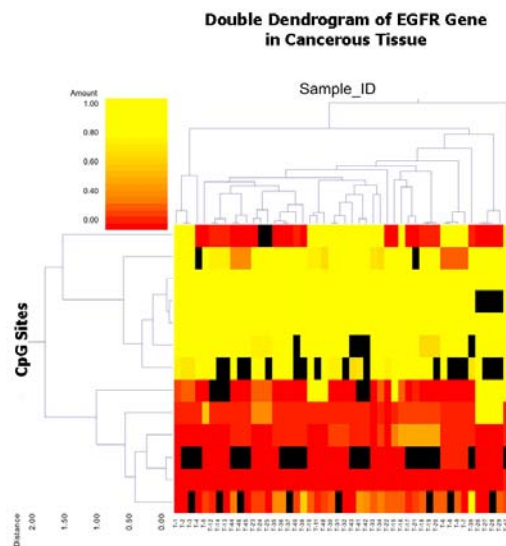and paired normal tissues.
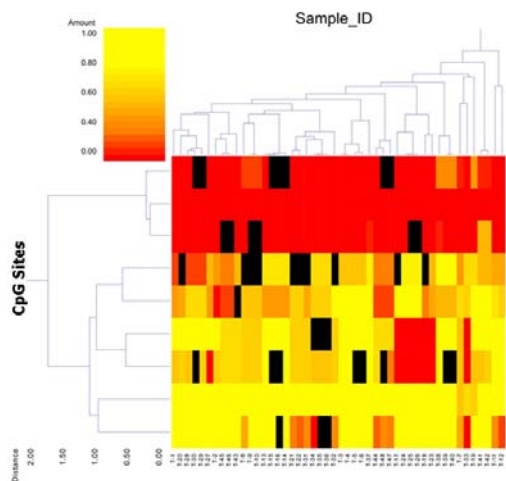
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

EGFR gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
1956	epidermal growth factor receptor	7p12	Receptor for epidermal growth factor and erythroblastic leukemia viral (v-erb-b) oncogene homolog	direct
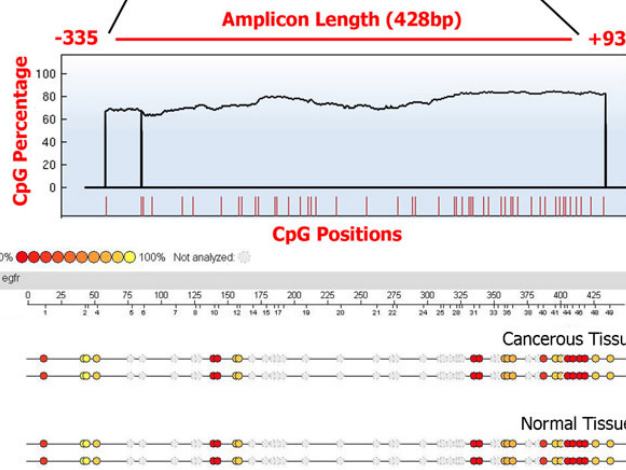
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. There is no difference in methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

ESR-b gene

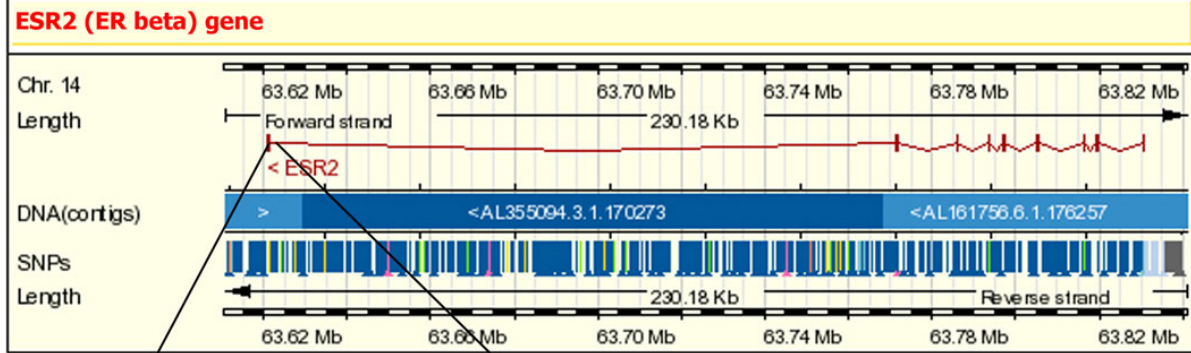| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
2100	estrogen receptor 2 (ER beta)	14q23.2	Regulation of cell proliferation, predictor of endocrine therapy	direct
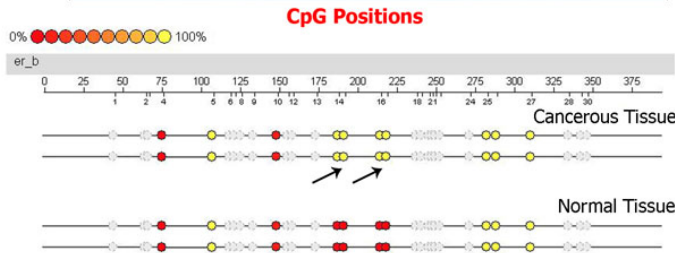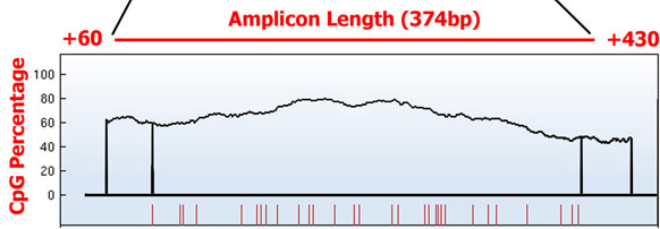
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

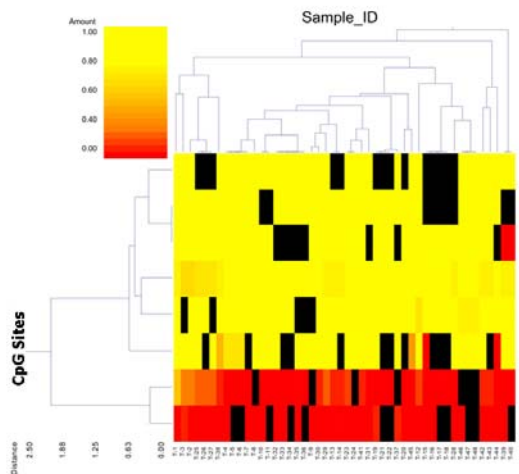**Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.**

GSTP1 gene

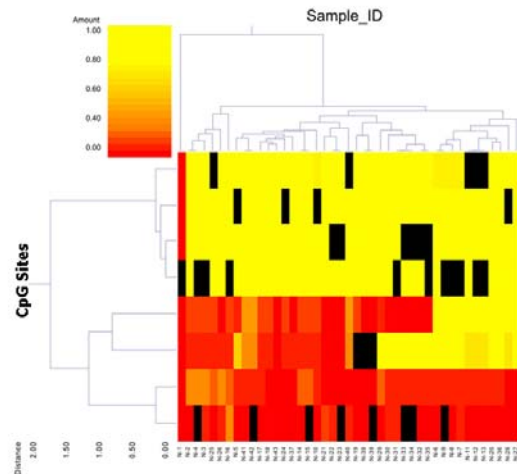| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
2950	Glutathione S-transferase P1	11q13	Carcinogen detoxification. GSTP1 is a polymorphic gene encoding active, functionally different GSTP1 variant proteins that are thought to function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases.	direct
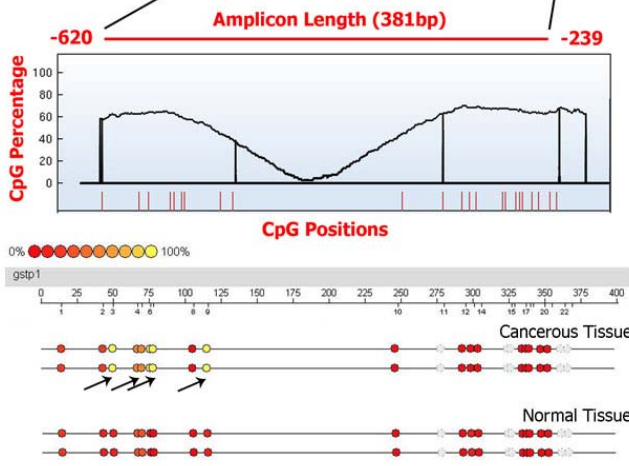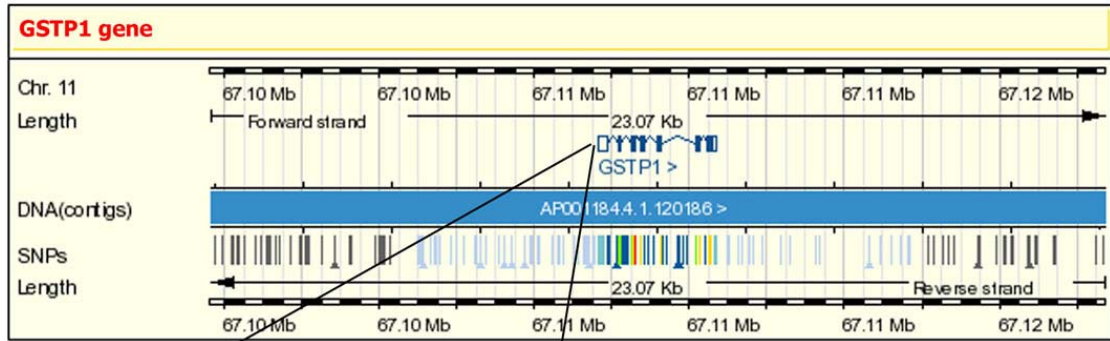
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

NES1 (KLK10) gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
5655	Kallikrein-related peptidase 10	19q13.3-q13.4	Inhibition of anchorage-independent growth and tumor formation.	direct

Methylation analysis of informative CpG sites.
The color of circles is related to percentage of methylation in each CpG site as quantification.
Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

NM23-H1 (NME1) gene
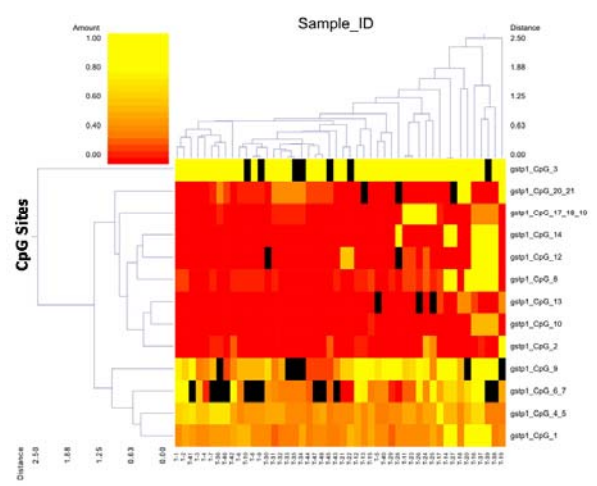
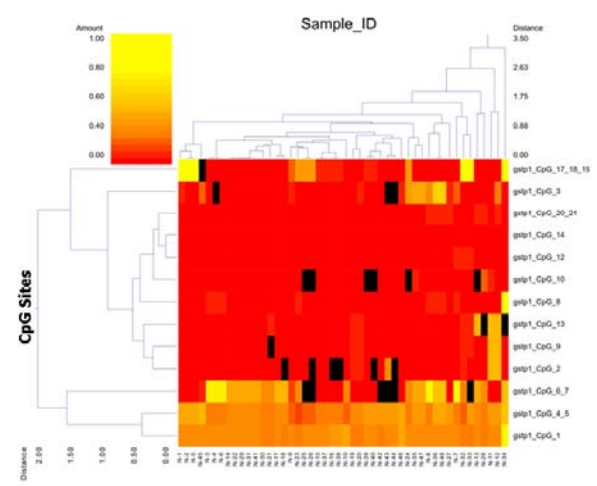| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
4830	non-metastatic cells 1	17q21.3	Metastasis suppressor activity.	direct
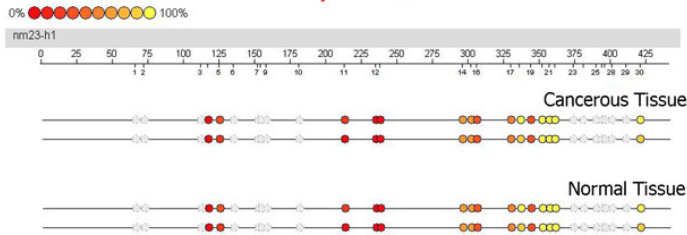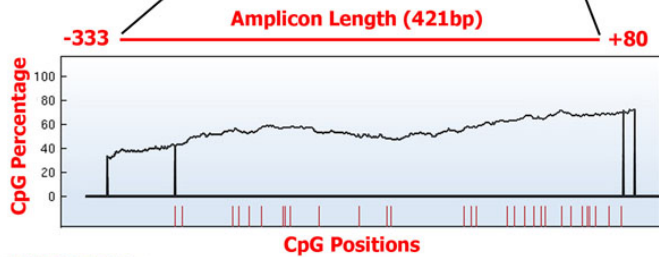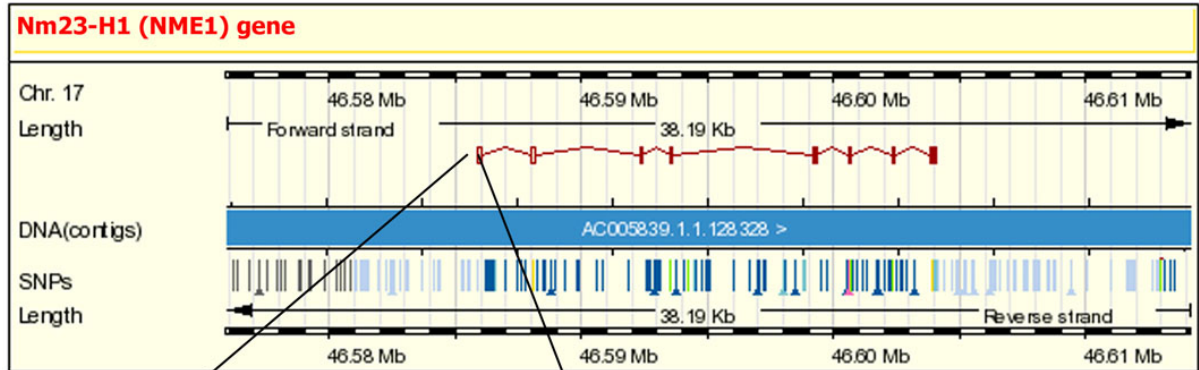
Methylation analysis
of informative CpG sites.
The color of circles is related to percentage of
methylation in each CpG site as quantification.
There is no difference in methylation
patterns between cancerous tissues
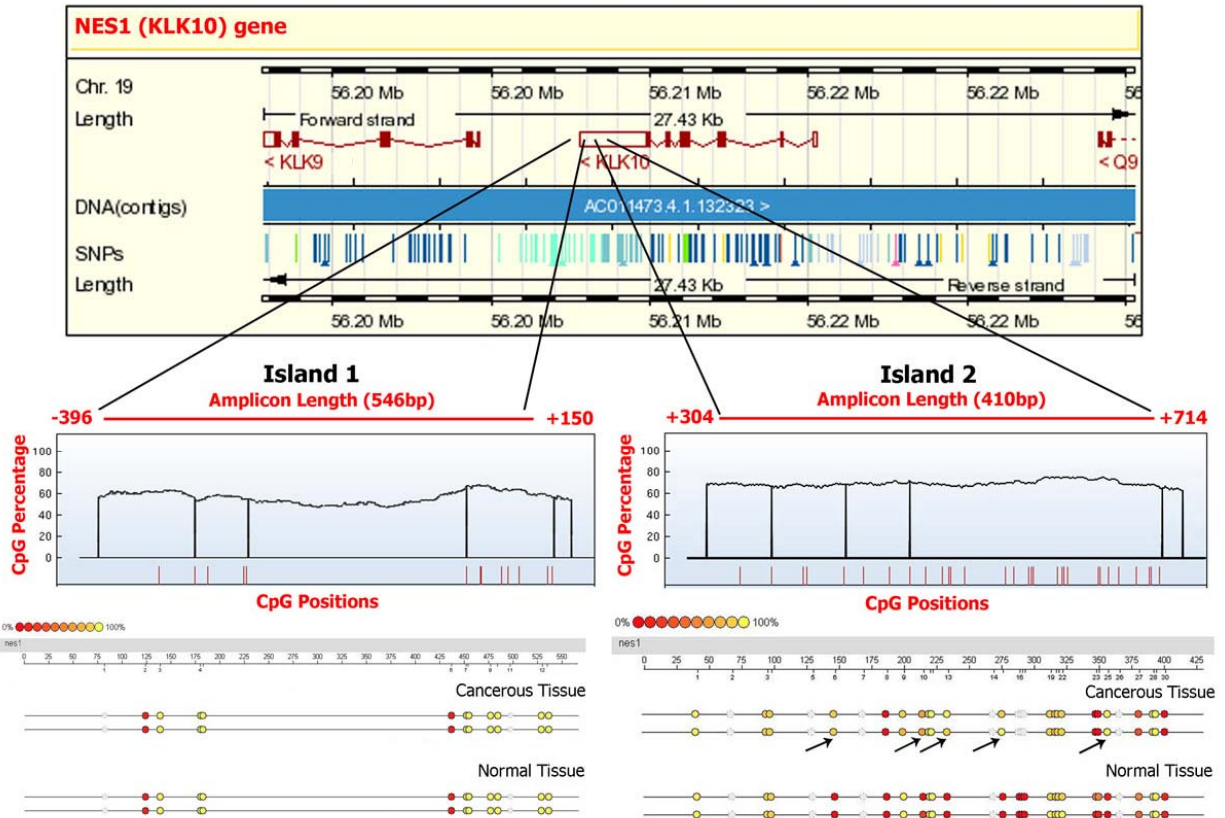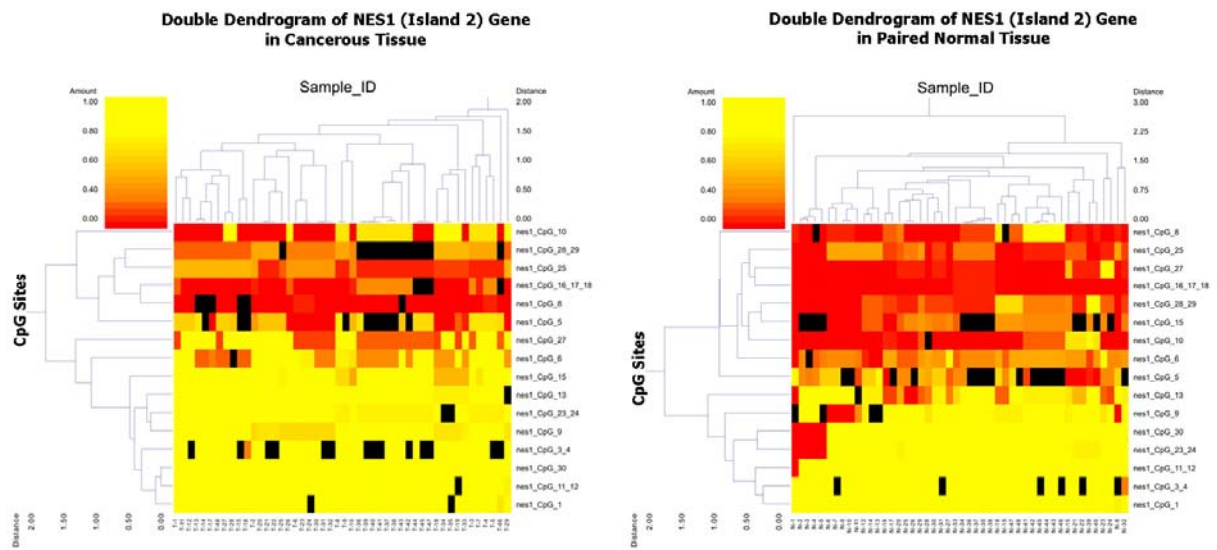and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

P16 (CDKN2A) gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
1029	Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	9p21	Cell cycle regulation, involved in senescence.	direct

Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
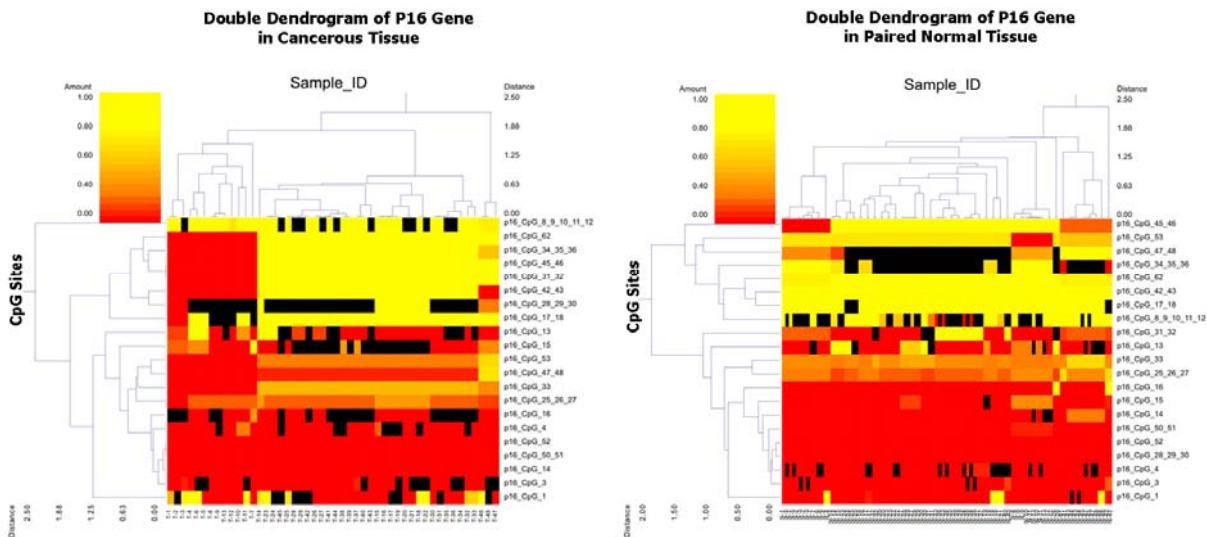
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

P21 (CDKN1A) gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
1026	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	6p21.2	Encodes protein that binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, and thus functions as a regulator of cell cycle progression at G1. This protein plays a regulatory role in S phase DNA replication and DNA damage repair.	direct
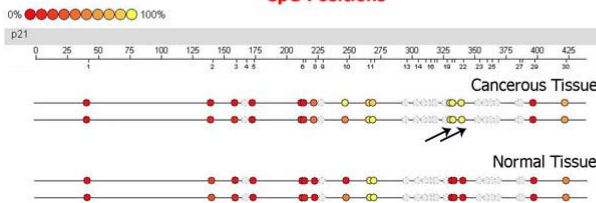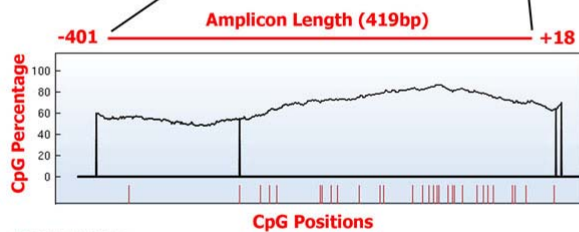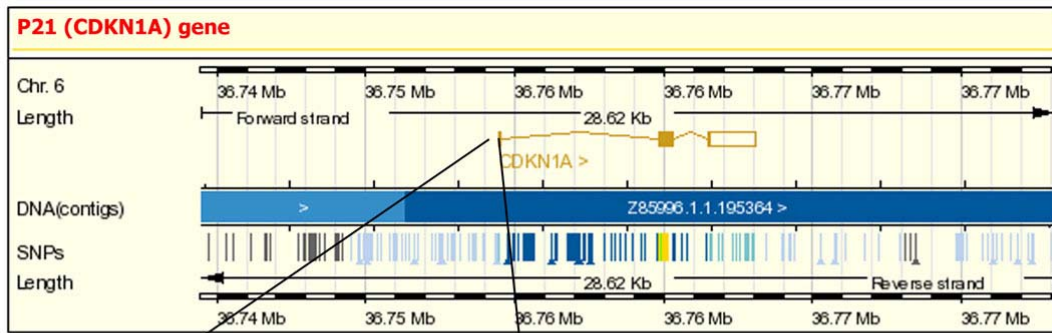
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
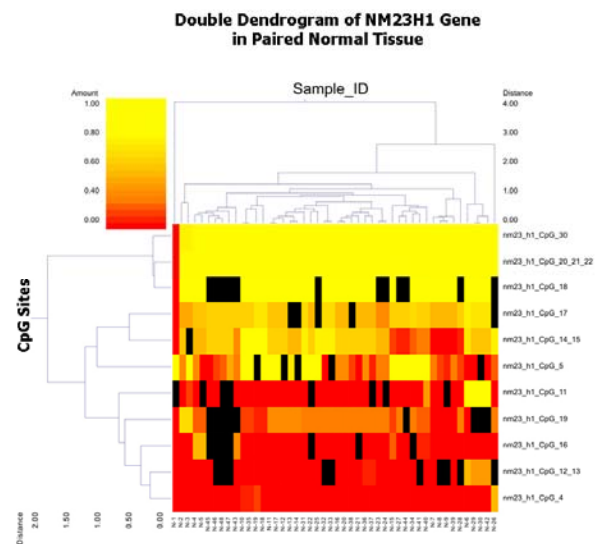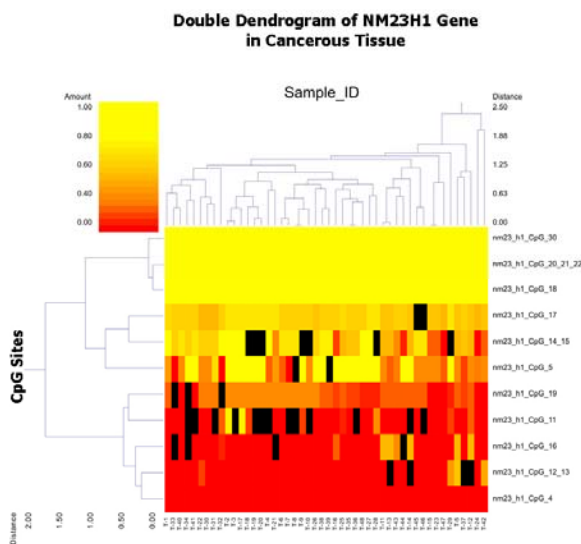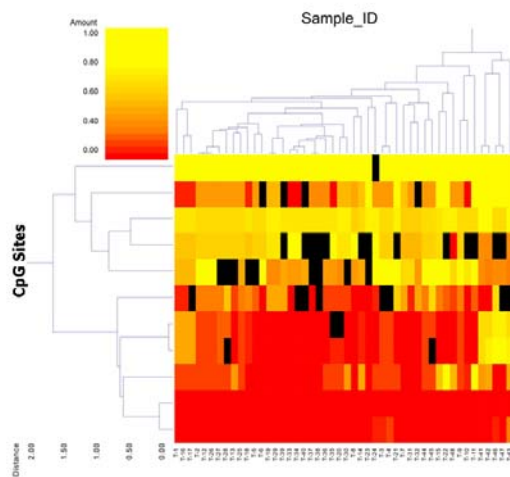

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

Progesterone Receptor (PGR) gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
5241	Progesterone receptor	11q22-q23	This gene encodes a member of the steroid receptor superfamily. Altered expression of this gene has been associated with the pathogenesis of breast cancers	direct

Methylation analysis
of informative CpG sites.
The color of circles is related to percentage of
methylation in each CpG site as quantification.
Arrows show the different methylation
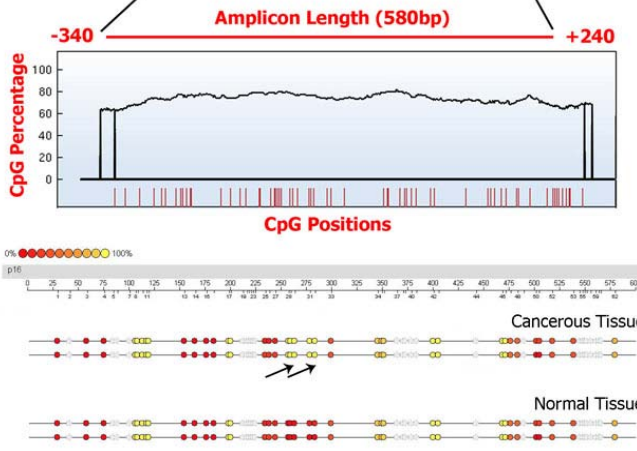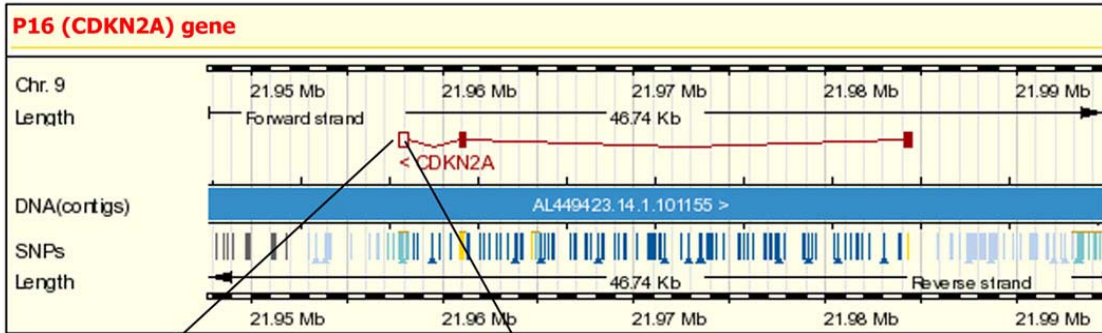patterns between cancerous tissues
and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

Prostasin (PRSS8) gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
5652	Protease serine 8	16p11.2	Suppression of invasion.	direct

Methylation analysis
of informative CpG sites.
The color of circles is related to percentage of
methylation in each CpG site as quantification.
There is no difference in methylation
patterns between cancerous tissues
and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

RAR-b gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
5915	Retinoic acid receptor β	3p24	Apoptosis, involved in senescence, inhibition of proliferation	direct
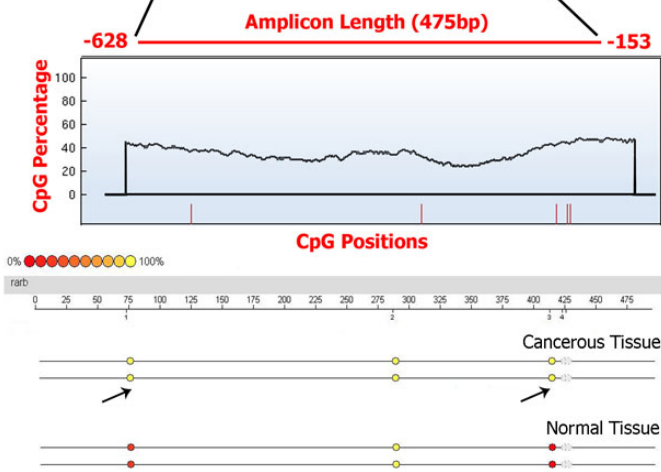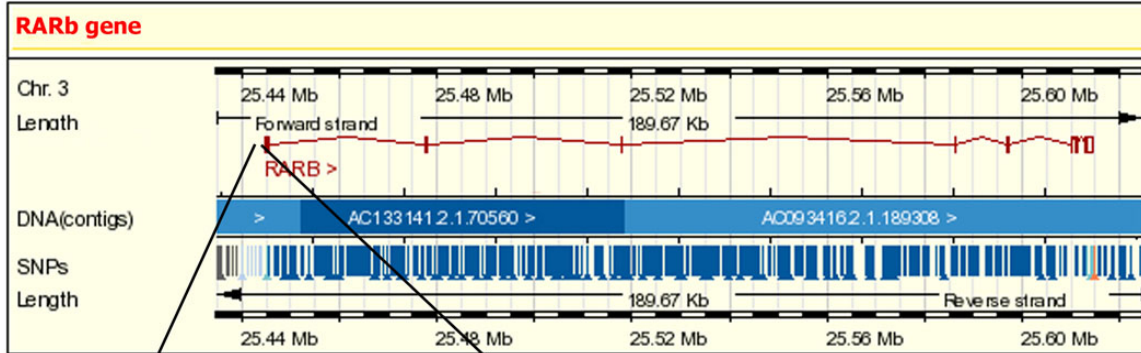
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.

Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

RASSF1 gene

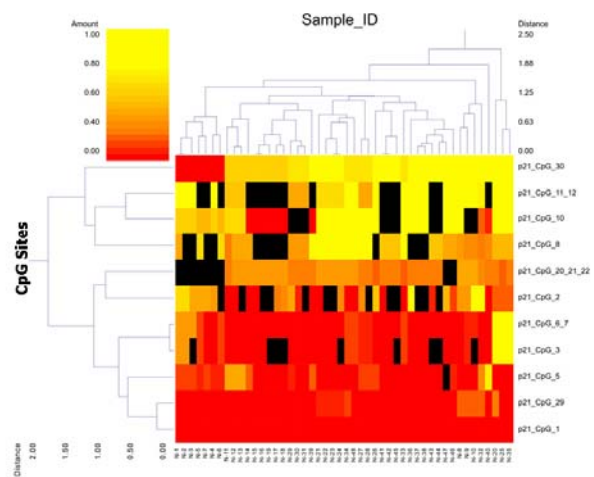| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
11186	Ras association (RalGDS/AF-6) domain family 1	3p21.3	This gene encodes a protein similar to the RAS effector proteins. Loss or altered expression of this gene has been associated with the pathogenesis of a variety of cancers, which suggests the tumor suppressor function of this gene.	direct

Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
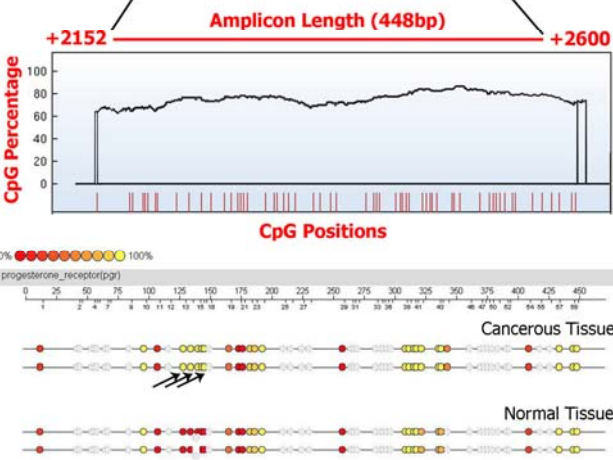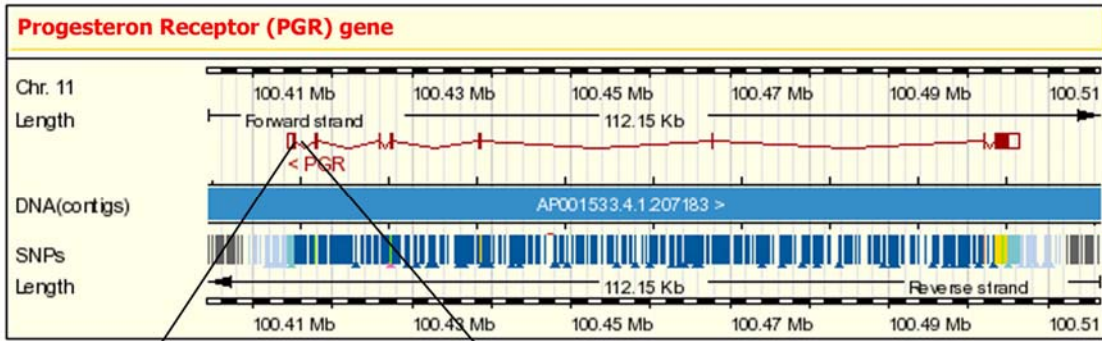

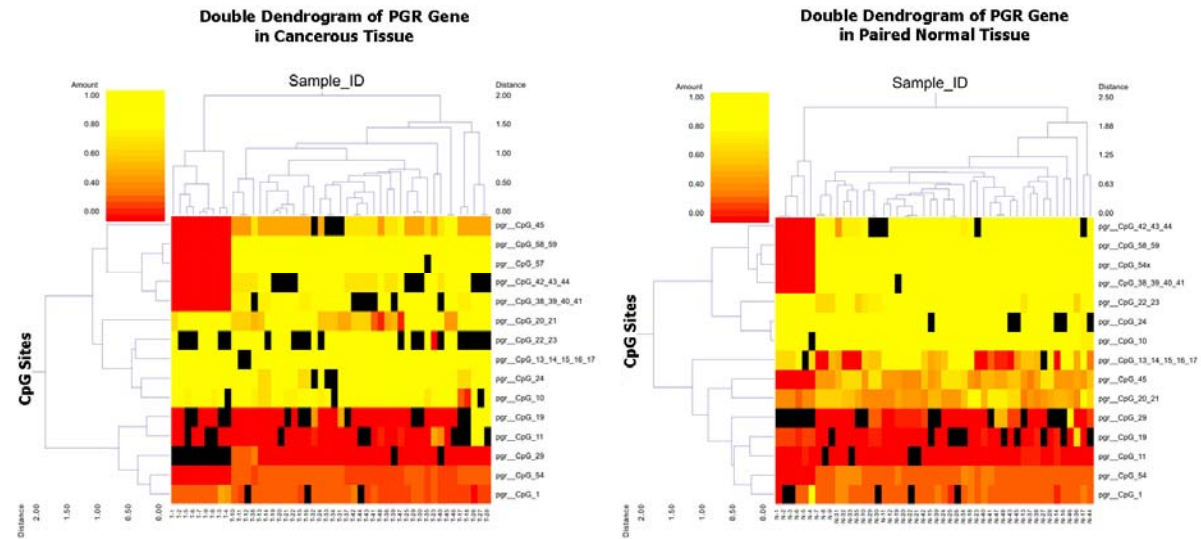**Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.**

SRBC (PRKCDBP) gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
112464	protein kinase C, delta binding protein	11p15.4	Interaction with BRCA1. The expression of this protein was found to be down-regulated in various cancer cell lines, suggesting the possible tumor suppressor function of this protein.	direct
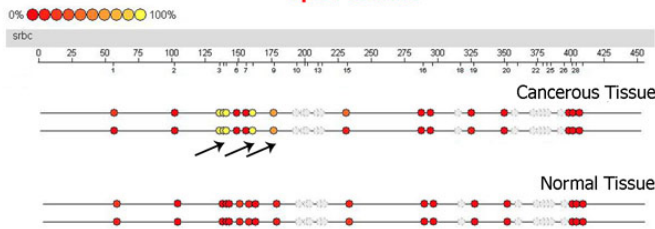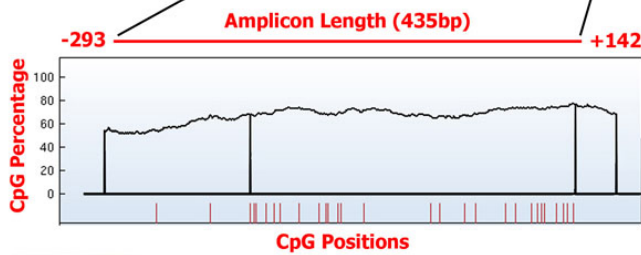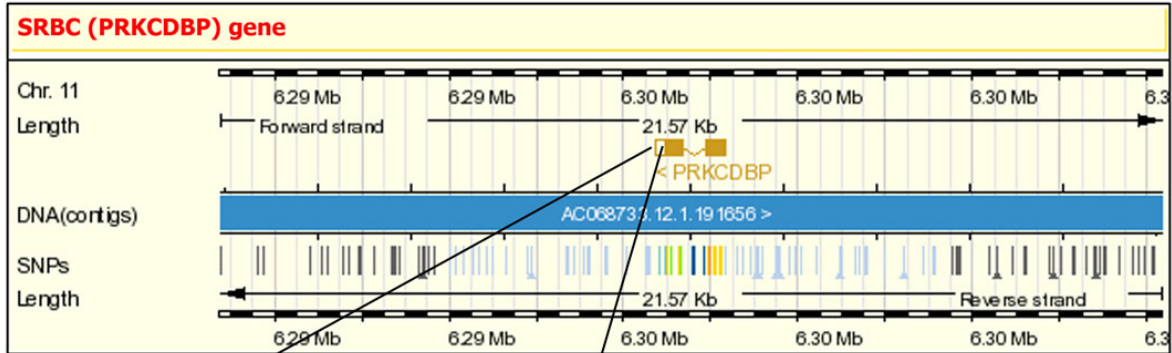
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
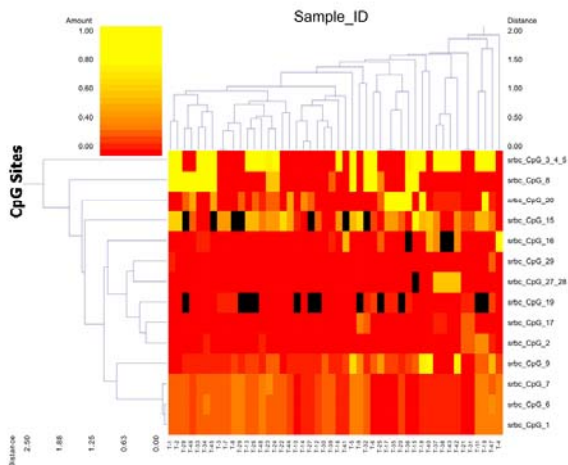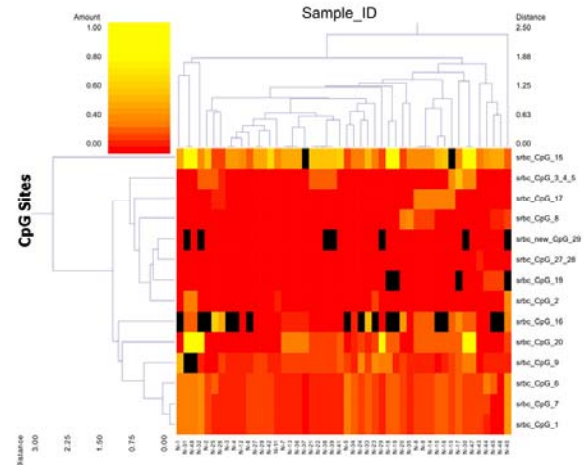
Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.

TIMP3 gene

Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
7078	Tissue inhibitor of metalloproteinase-3	22q12.3	Suppresses tumor growth, angiogenesis, invasion and metastasis	direct
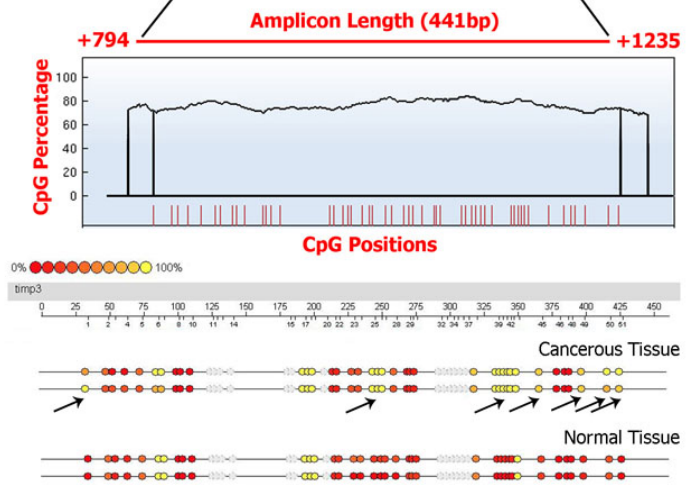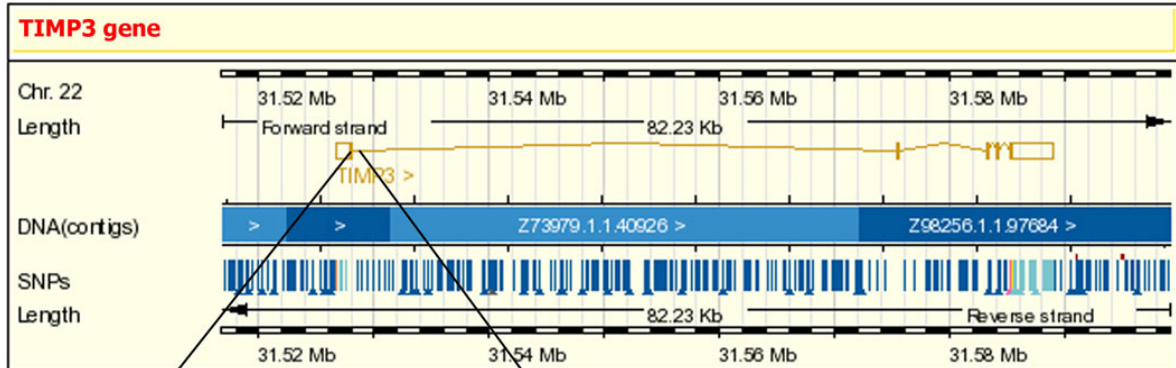
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. Arrows show the different methylation patterns between cancerous tissues and paired normal tissues.
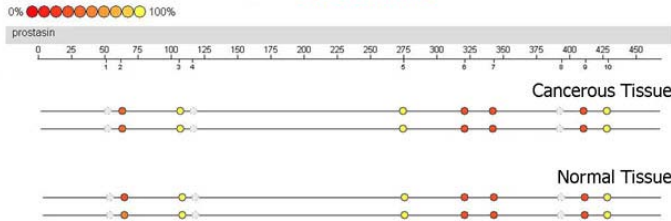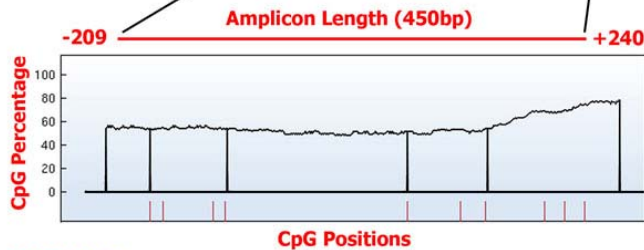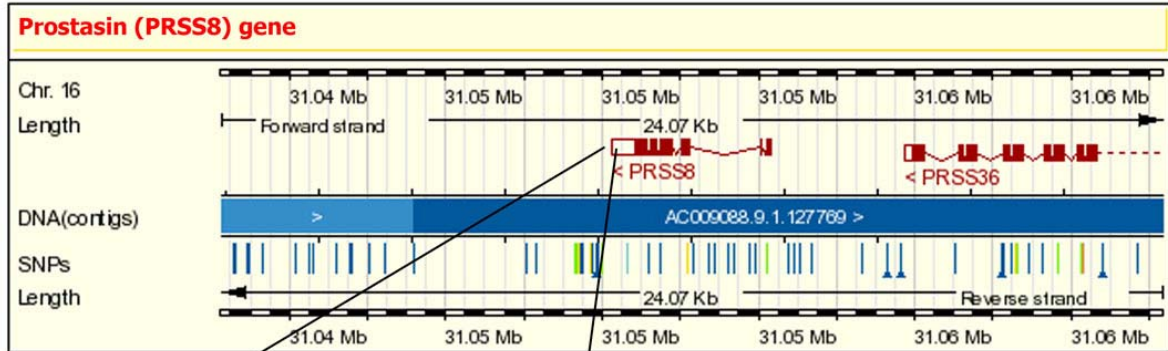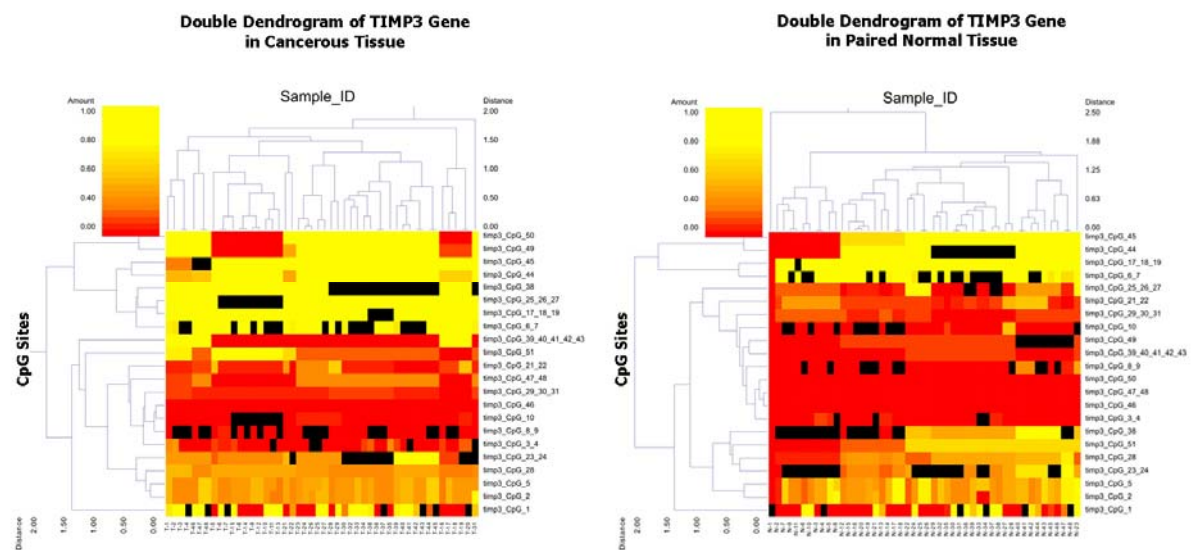

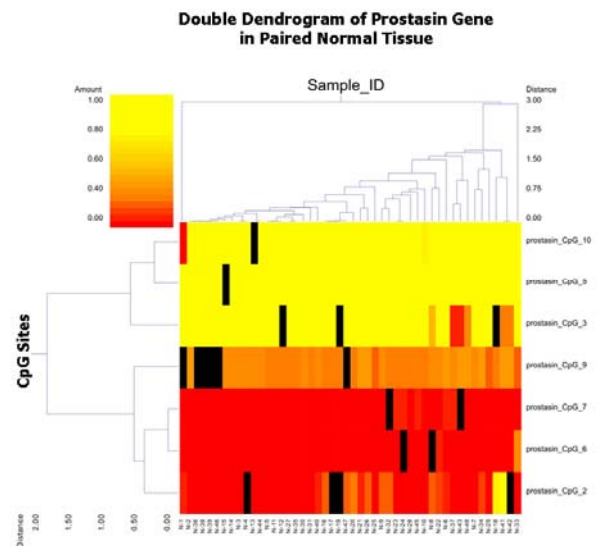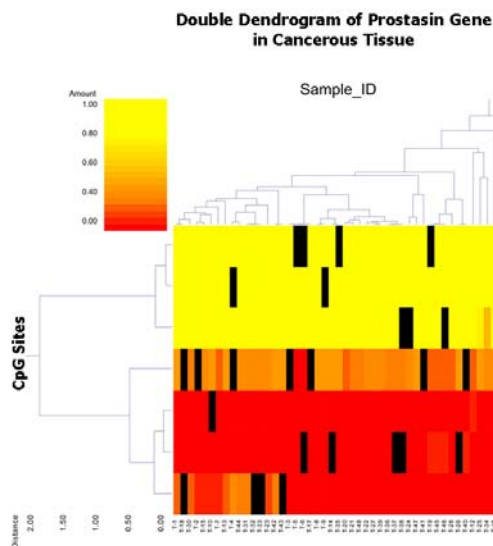**Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.**

TP53 gene

| Gene ID	Alternate gene name	locus	Function	Methylation effect on breast cancer
7157	Transformation-related protein 53	17p13.1	Apoptosis, cell cycle regulation, inhibition of growth and invasion.	direct
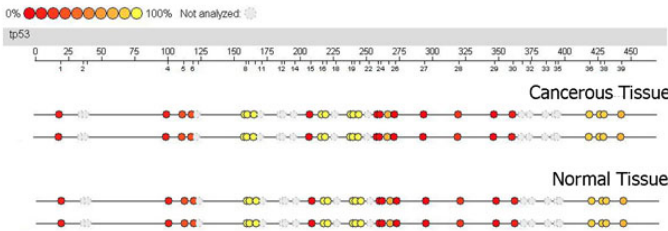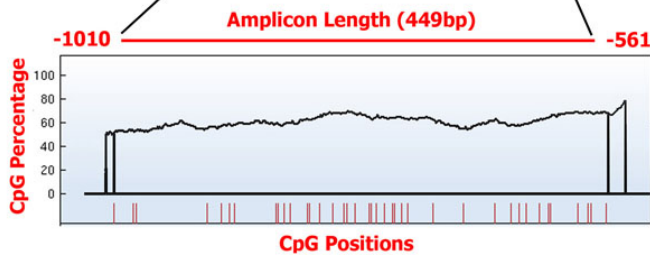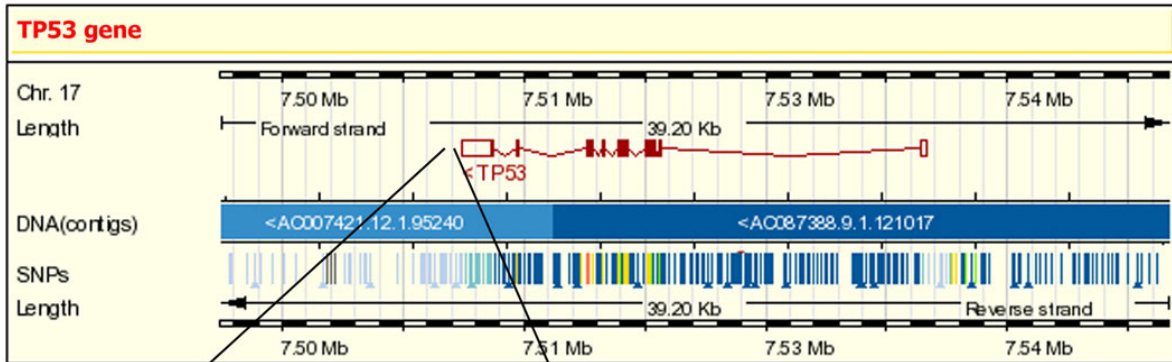
Methylation analysis of informative CpG sites. The color of circles is related to percentage of methylation in each CpG site as quantification. There is no difference in methylation patterns between cancerous tissues and paired normal tissues.
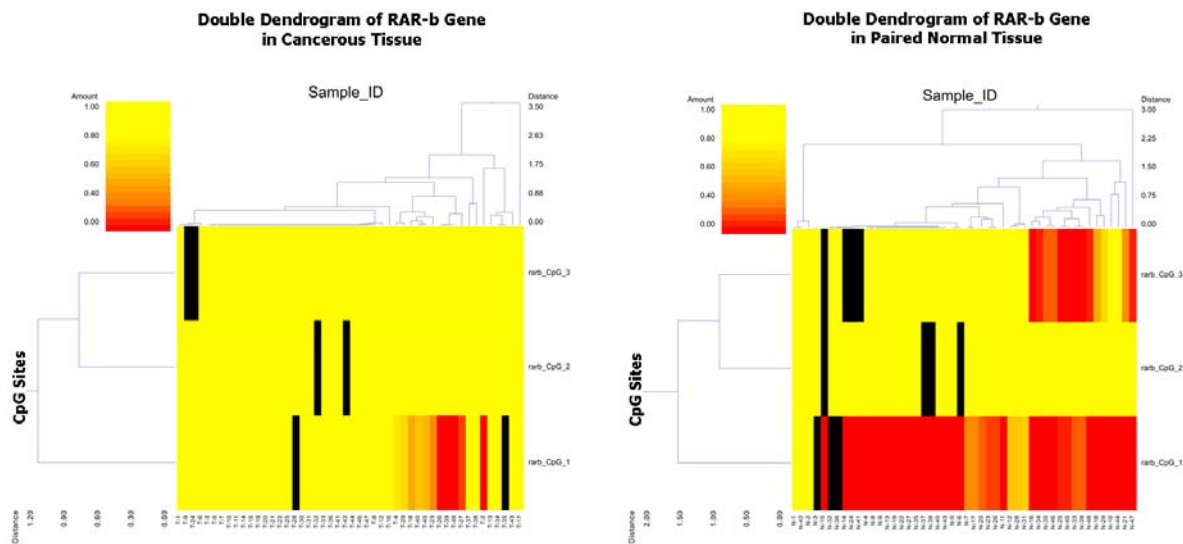

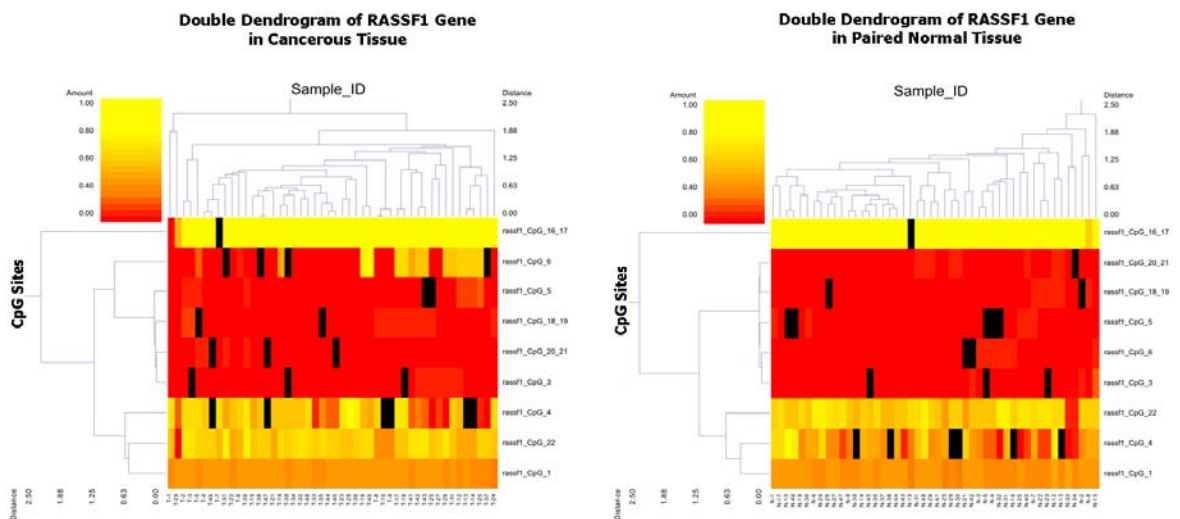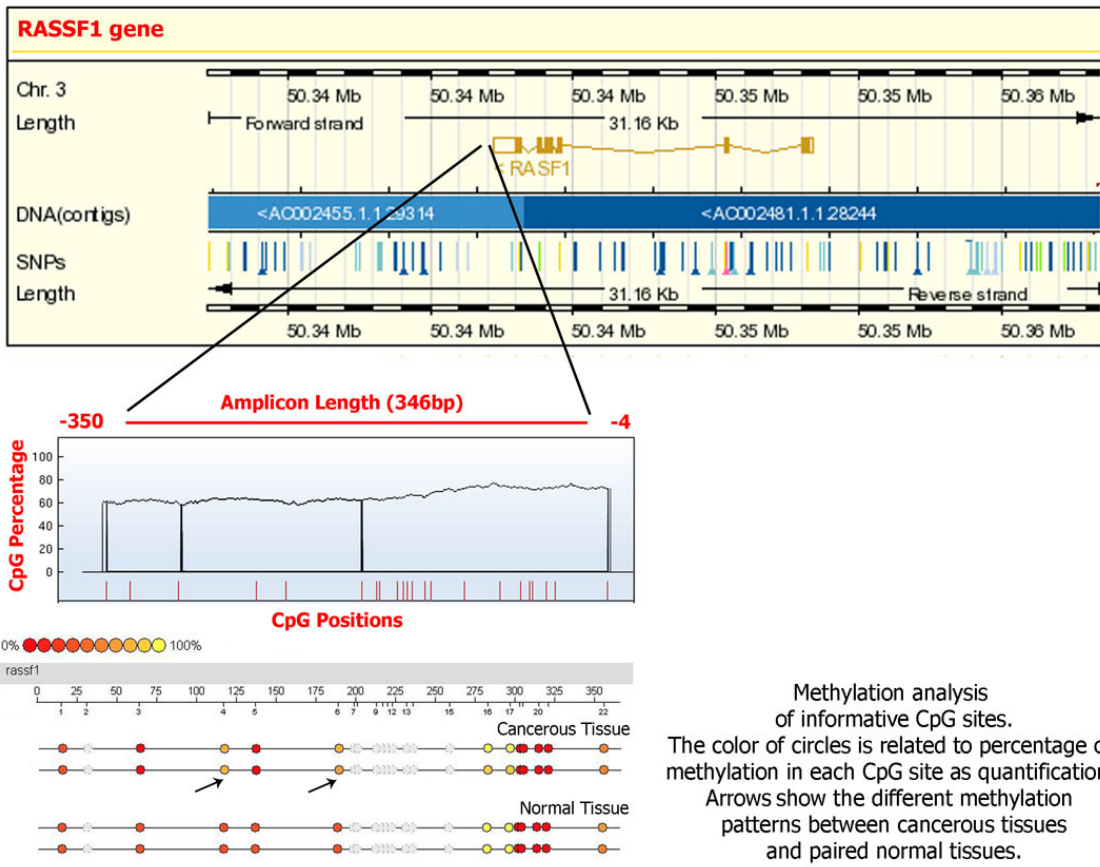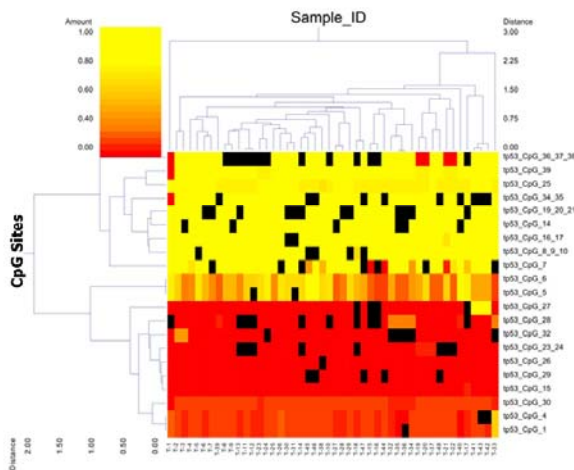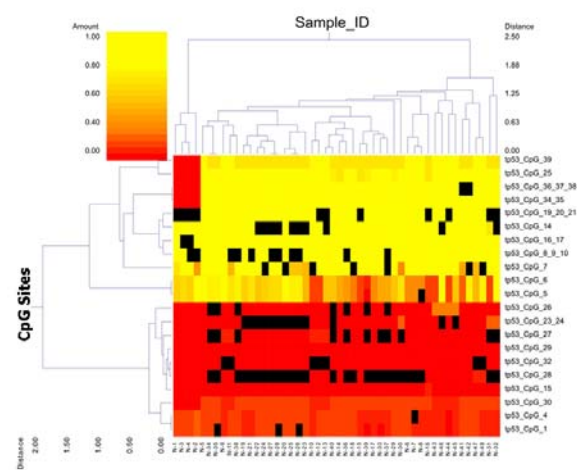**Two-way hierarchical cluster analysis of 48 cancerous breast tissues and 48 adjacent normal tissues.**

7. Published research article:

Hypermethylation of tumor suppressor genes involved in critical regulatory pathways for developing a high-throughput blood-based test in breast cancer

Summary:

In this study, the methylation profile of 10 tumor suppressor genes (*APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESR-b*, *GSTP1*, *P16*, *P21* and *TIMP3*) involved in critical cellular and molecular regulatory pathways were quantitatively assessed using a mass spectrometry on MALDI-TOF MS in circulating cell-free DNA, paired breast tumor and normal tissue of different 126 samples. The possibility of using the cancer specific alterations in plasma and serum samples as an epigenetic blood-based test for breast cancer was explored. Our study on clinical samples suggests that the selected TSG panel combined with the high-throughput technology is a useful tool to develop epigenetic based biomarker for breast cancer relies on pathologic methylation changes in tumor tissue, as well as in circulation.

First author's contribution:

Ramin Radpour was involved in performing the experiment, data analysis and writing the manuscript.

PLoS one

Hypermethylation of Tumor Suppressor Genes Involved in Critical Regulatory Pathways for Developing a Blood-Based Test in Breast Cancer

Ramin Radpour[1], Zeinab Barekati[1], Corina Kohler[1], Qing Lv[2], Nicole Bürki[3], Claude Diesch[3], Johannes Bitzer[4], Hong Zheng[5], Seraina Schmid[4]*, Xiao Yan Zhong[1]*

1 Laboratory for Gynecological Oncology, Department of Biomedicine, Women's Hospital, University of Basel, Basel, Switzerland, 2 Department of Breast Surgery, West China Hospital, West China School of Medicine, Sichuan University, Chengdu, China, 3 Department of Obstetrics and Gynecology, Kantonsspital, Liestal, Switzerland, 4 Department of Obstetrics and Gynecology, Women's Hospital, University of Basel, Basel, Switzerland, 5 Laboratory of Molecular Diagnosis of Cancer, Department of Oncology, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China School of Medicine, Sichuan University, Chengdu, China

Abstract

Background: Aberrant DNA methylation patterns might be used as a biomarker for diagnosis and management of cancer patients.

Methods and Findings: To achieve a gene panel for developing a breast cancer blood-based test we quantitatively assessed the DNA methylation proportion of 248 CpG sites per sample (total of 31,248 sites in all analyzed samples) on 10 candidate genes (*APC, BIN1, BMP6, BRCA1, CST6, ESR-b, GSTP1, P16, P21* and *TIMP3*). The number of 126 samples consisting of two different cohorts was used (first cohort: plasma samples from breast cancer patients and normal controls; second cohort: triple matched samples including cancerous tissue, matched normal tissue and serum samples). In the first cohort, circulating cell free methylated DNA of the 8 tumor suppressor genes (TSGs) was significantly higher in patients with breast cancer compared to normal controls ($P<0.01$). In the second cohort containing triple matched samples, seven genes showed concordant hypermethylated profile in tumor tissue and serum samples compared to normal tissue ($P<0.05$). Using eight genes as a panel to develop a blood-based test for breast cancer, a sensitivity and specificity of more than 90% could be achieved in distinguishing between tumor and normal samples.

Conclusions: Our study suggests that the selected TSG panel combined with the high-throughput technology might be a useful tool to develop epigenetic based predictive and prognostic biomarker for breast cancer relying on pathologic methylation changes in tumor tissue, as well as in circulation.

Introduction

Breast cancer is one of the most common types of cancer among women. Localized breast cancer at an early stage has better prognosis and requires less severe treatment with a survival rate of 98% [1]. However, diagnosis after tumor metastasis lowers the survival rate to 27% [2]. This highlights the importance of early breast cancer detection which is dependent on sensitive and specific screening methods. The traditional triple test for breast cancer diagnosis includes physical examination, mammography and aspiration cytology. Unfortunately, all these methods are not sensitive enough in identifying breast cancer in early stages [1,3]. A minimally invasive screening test beside the triple test, or prior to biopsy, would lead to greater sensitivity.

It is well recognized that solid malignant tumors release significant amounts of DNA into the systemic circulation through cellular necrosis or apoptosis [4]. The presence of cell-free DNA (cfDNA) in plasma and serum has been known for over 60 years.

Quantitative alteration of circulating cfDNA has been observed in several cancers, such as prostate cancer [5], lung cancer [6], pancreatic cancer [7], and breast cancer [8]. The tumor released DNA in circulation might serve as biomarker for cancer [8].

Aberrant promoter methylation pattern of tumor suppressor genes (TSGs) is known to be a frequent and early event in carcinogenesis [9,10,11,12]. Tumor-specific methylated DNA alterations have been found in the circulation of patients with different types of cancer [12,13]. The analysis of the methylation patterns of cfDNA by a blood-based test might enable to distinguish between benign and malignant tumors for diagnosis and surveillance of patients [12].

The SEQUENOM's EpiTYPER™ assay is a high-throughput methylation quantification method which relies on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [14]. The sensitivity, specificity and assay concept of the method have been previously described by ulterior studies [14,15,16,17,18,19,20]. Recently, we analyzed the methylation profiles of more than 42,528 CpG sites on 22 genes of which

Table 1. Clinical characteristics of patients in the two study cohorts.

Sample type	Total no. of patients	Age mean ± S.D. (range)	Pathologic stage		No. of patients with lymph node involvement	No. of patients with metastasis	Histological grade			ER		PR	
			Early*	Late**			1	2	3	Positive	Negative	Positive	Negative
Plasma samples	36	67±13.4 (38–89)	27	9	19	0	11	18	7	28	8	23	13
Triple samples	20	50±11.7 (33–77)	12	8	13	0	0	5	15	16	4	10	10

*The pathologic stage<III was considered as "Early stage".
**The pathologic stage III and IV was considered as "Late stage".
ER: Estrogen receptor; PR: Progesterone receptor.
doi:10.1371/journal.pone.0016080.t001

10 were shown to be hypermethylated genes (*APC, BIN1, BMP6, BRCA1, CST6, ESR-b, GSTP1, P16, P21* and *TIMP3*) in cancerous breast tissue in comparison with matched normal tissue [17]. These 10 hypermethylated genes were considered as methylation signature of breast cancer and were used in this study for further investigations to develop an epigenetic blood-based assay for breast cancer.

In the present study, to achieve a reliable gene panel for developing a blood-based test, we quantitatively assessed the DNA methylation profile of 10 breast cancer candidate genes using MALDI-TOF MS in two different cohorts of patients with breast cancer on large-scale CpG sites.

Materials and Methods

The study was performed at the Laboratory for Gynecological Oncology, Department of Biomedicine, Women's Hospital, Basel and approved by the local institutional review board (Ethic commission beider Basel). Written consent forms were collected from all patients who were involved in this study.

Sampling and pathological classification

In total 126 samples were used in this study. For analysis we divided these samples in two different cohorts. The first cohort consisted of 36 plasma samples of breast cancer patients and 30 plasma samples of healthy non-relative women. The second cohort consisted of 60 triple samples (cancerous tissue, matched normal tissue and serum samples) from 20 patients with non-familial breast cancer. Staging and grading was evaluated according to the WHO histological classification. Breast cancer characteristics, such as staging, histological grading, and hormone receptor expression from the two different cohorts are summarized in Table 1.

Isolation of circulating cfDNA from plasma and serum

A total of 20 mL blood samples were collected in both EDTA tubes (for plasma) and EDTA-free tubes (for serum) and processed immediately after collection. The plasma samples were centrifuged at 1,600×g (10 min), and supernatant was carefully transferred into 2 mL microtubes. Samples were centrifuged in a microcentrifuge at full speed (10 min), and supernatants were stored at −80°C until analysis was performed. The serum tubes were coagulated during approximately 1h, after which the serum was harvested and stored using the above mentioned procedure.

DNA extraction was performed from 25–50 mg of frozen tissue and 600 µL of plasma and serum using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) and eluted in a final volume of 100 µL. The median

Table 2. High-throughput methylation analysis of CpG sites per amplicon for the 10 studied genes.

Genes	Amplicon size (bp)	Total No. of CpG sites in amplicon	No. of analyzed CpG sites in amplicon	No. of informative CpG sites in amplicon	No. of analyzed CpG sites per amplicons	
					Single sites	Composite sites
APC	420	26	18	12	12	6
BIN1	330	32	18	12	3	15
BMP6	397	37	30	9	11	19
BRCA1	413	30	15	7	10	5
CST6	445	49	27	6	15	12
ESR-b (ER beta)	374	30	24	6	7	17
GSTP1	381	23	17	8	10	7
P16 (CDKN2A)	580	62	36	14	13	23
P21 (CDKN1A)	419	30	19	8	10	9
TIMP3	441	51	44	14	11	33

The *in silico* digestion was performed for the T-cleavage assay. The percentage of total CpG sites in the amplicon is divided into single sites (single CpG sites) and composite sites (two or more adjacent CpG sites fall within one fragment, or when fragment masses are overlapping).
doi:10.1371/journal.pone.0016080.t002
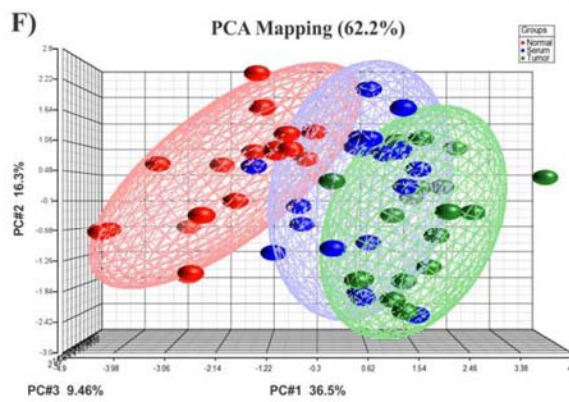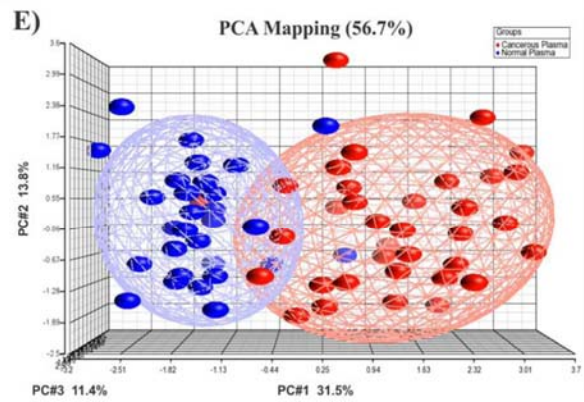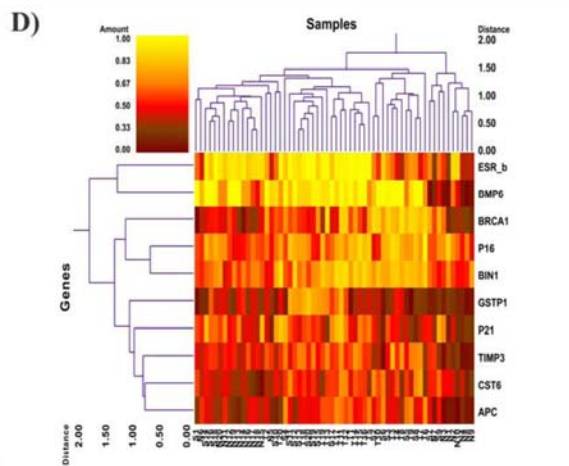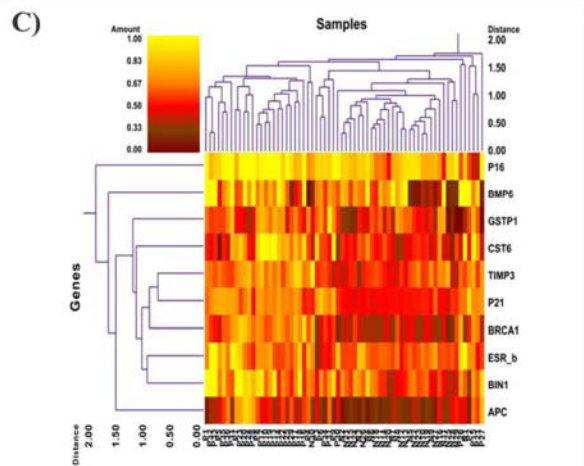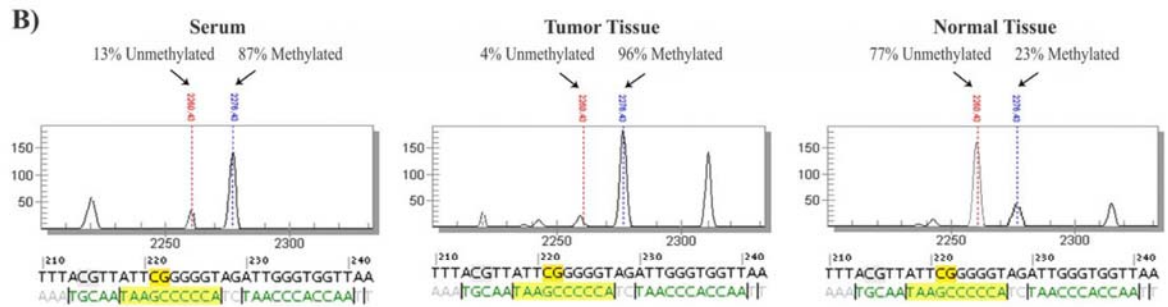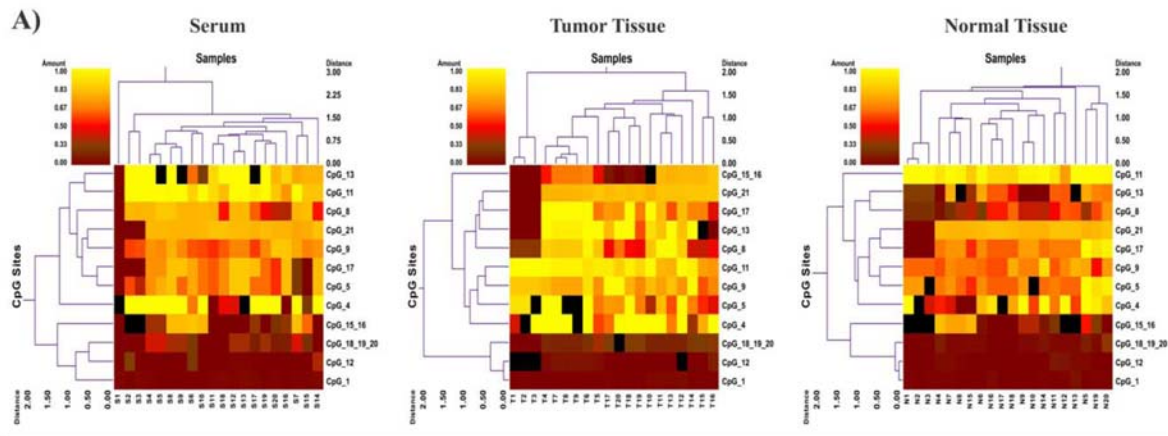
Figure 1. Methylation profiling of 10 candidate genes in two studied cohorts. A) An example of high-throughput methylation analysis of CpG sites for the *BRCA1* gene for the 60 triple samples (cancerous tissue, matched normal tissue and serum samples). The complete data for the other genes is summarized in Dataset S1. B) Peaks show percentage of methylation extent obtained from an informative CpG site of *BRCA1* gene with a significant difference between serum and tumor with normal tissue in a triple case. C) Double dendrogram profiles the mean methylation proportion of all 10 studied genes in plasma samples from breast cancer patients and normal subjects. D) Double dendrogram profiles the mean methylation proportion of all 10 studied genes in triple matched samples. E) PCA mapping of the mean methylation proportion of analyzed genes in plasma samples. F) PCA mapping of the mean methylation proportion of analyzed genes in triple matched samples.
doi:10.1371/journal.pone.0016080.g001

quantity of extracted cfDNA in plasma and serum were 5.7 ng/μL (range 2.6 to 12.1) and 7.1 ng/μL (range 5.4 to 14.8) respectively. The median quantity of extracted DNA from frozen tissue was 65.7 ng/μL (range 28.3 to 186.1).

Before performing the methylation analysis, we quantified the yield of extracted DNA by quantitative PCR for the *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) gene. The good quality of the extracted DNA allowed successful amplification and quantification of the *GAPDH* gene in all samples (data not shown).

Bisulfite Treatment

To perform bisulfite conversion of the target sequence, the Epitect® Bisulfite Kit (QIAGEN AG, Basel, Switzerland) was used according to the manufacturer's protocol.

Primer design and PCR-tagging for EpiTYPER™ assay

We used previously designed and tagged primers (reverse primer with T7-promoter tag and forward primer with 10mer tag sequence as balance) for the 10 candidate genes [17]. Selected amplicons were mostly located in the promoter regions, or started from the promoter and partially covered the first exon [17]. For the PCR on bisulfite-treated genomic DNA (gDNA), the following PCR conditions were used: 1×: 95°C for 10 min; 48×: 95°C for 30s, Ta for 40s, 72°C for 1 min; 1× 72°C for 5 min. The PCR cocktail was: 2μL DNA (2.00μL of at least 10 ng/μL DNA for a final concentration of 2ng/μL per reaction) in a 10μL total volume using 1pmol of each primer, 200μM dNTP, 0.2 unit Hot Start Taq DNA polymerase, 1.5mM MgCl2 and the buffer supplied with the enzyme.

In vitro transcription and T-cleavage assay

In vitro transcription and T-cleavage assay were assessed according to the previously published methods [14,16,17]. Briefly, unincorporated dNTPs were removed by shrimp alkaline phosphatase (SAP; SEQUENOM, Inc., San Diego, CA) treatment. Typically, 2 μL of the PCR product were used as template for the transcription reaction. Twenty units of T7 R&DNA polymerase (Epicentre, Madison, WI) were used to incorporate dTTP in the transcripts. Ribonucleotides and dNTPs were used at concentrations of 1 mmol/L and 2.5 mmol/L, respectively. In the same step, RNase-A (SEQUENOM Inc., San Diego, CA) was added to cleave the *in vitro* transcripts (T-cleavage assay). Samples were diluted with H₂O to a final volume of 27 μL. Conditioning of the phosphate backbone was achieved by adding 6 mg of Clean Resin (SEQUENOM) before performing MALDI-TOF MS analysis.

Mass spectrometry

Twenty-two nanoliters of the RNase-A treated product were robotically dispensed onto silicon matrix preloaded chips (Spectro-CHIP; SEQUENOM, San Diego), the mass spectra were collected using a MassARRAY® Compact MALDI-TOF (SEQUENOM) and spectra's methylation proportion were generated by the EpiTYPER™ software v1.0 (SEQUENOM, San Diego).

Statistical methods

Data analysis was performed using the PASW Statistics software v.18. The Shapiro-Wilk and Kolmogorov-Smirnov tests were used for data distribution analysis. Both tests similarly demonstrated that our data set was not normally distributed (Shapiro-Wilk test; $P<0.001$ and Kolmogorov-Smirnov test; $P<0.001$). Quantitative methylation proportion of 10 genes was analyzed in two different study cohorts. Using two-way hierarchical cluster analysis, the most variable CpG sites for each gene were clustered based on pair-wise Euclidean distances and linkage algorithm for all studied samples according to the previously developed method by Gene Expression Statistical System (GESS) version 7.1.19 (NCSS, Kaysville, Utah, USA) [16,17,18]. The Mann-Whitney U test was used to compare the promoter methylation between study groups and also with clinicopathological parameters. The non-parametric Spearman's rho test was used to find out the correlation of methylation proportion in serum versus tumor and normal samples. Three dimensional principal component analysis (PCA) was accessed for both different cohorts based on the methylation proportion of 10 studied genes to transform a number of possibly correlated variables into a smaller number of uncorrelated variables.

Results

Quantitative methylation profiling of the 10 studied genes

In this study, we analyzed the methylation proportion of 10 breast cancer candidate genes in 126 different samples consisting of two different cohorts (36 plasma samples from patients with breast cancer and 30 plasma samples from normal controls, as well as 60 triple matched samples containing cancerous tissue, normal tissue and serum from 20 breast cancer patients). For all of the studied genes one amplicon per gene was analyzed and all amplicons contained CpG rich islands (with the number of CpG sites higher than 20) (Table 2). In total, we assessed 10 amplicons, containing 248 CpG sites per sample (total of 31,248 sites in all analyzed samples) (Table 2; Fig. 1; Dataset S1). From several analyzed CpG sites per amplicon few of them could represent valuable differences in the studied cases which were considered as informative CpG sites (Table 2). The mean methylation quantity of the informative CpG sites per each gene was used to figure out the methylation proportion of the candidate genes (Dataset S1).

Methylation proportion of candidate genes in plasma samples. Methylation proportion of each CpG site is given on a scaling 0 to 100 percent. Using two-way hierarchical cluster analysis, we found different methylation pattern of the candidate genes in plasma samples between patients with breast cancer and normal controls (Fig. 1; Dataset S1). Cell free methylated DNA levels of 8 genes (*APC*, *BIN1*, *BRCA1*, *CST6*, *GSTP1*, *P16*, *P21* and *TIMP3*) were significantly higher in the plasma samples from patients with breast cancer in comparison with those from normal controls ($P<0.01$), while the other two genes *BMP6* and *ESR-b* showed the same tendency but was not significant ($P>0.05$) (Fig. 2a; Dataset S1). PCA mapping based on the mean

Circulating Methylated DNA in Breast Cancer

A)

P: Plasma of breast cancer patients   N: Plasma of normal subjects

B)

S: Serum   T: Tumor   N: Normal
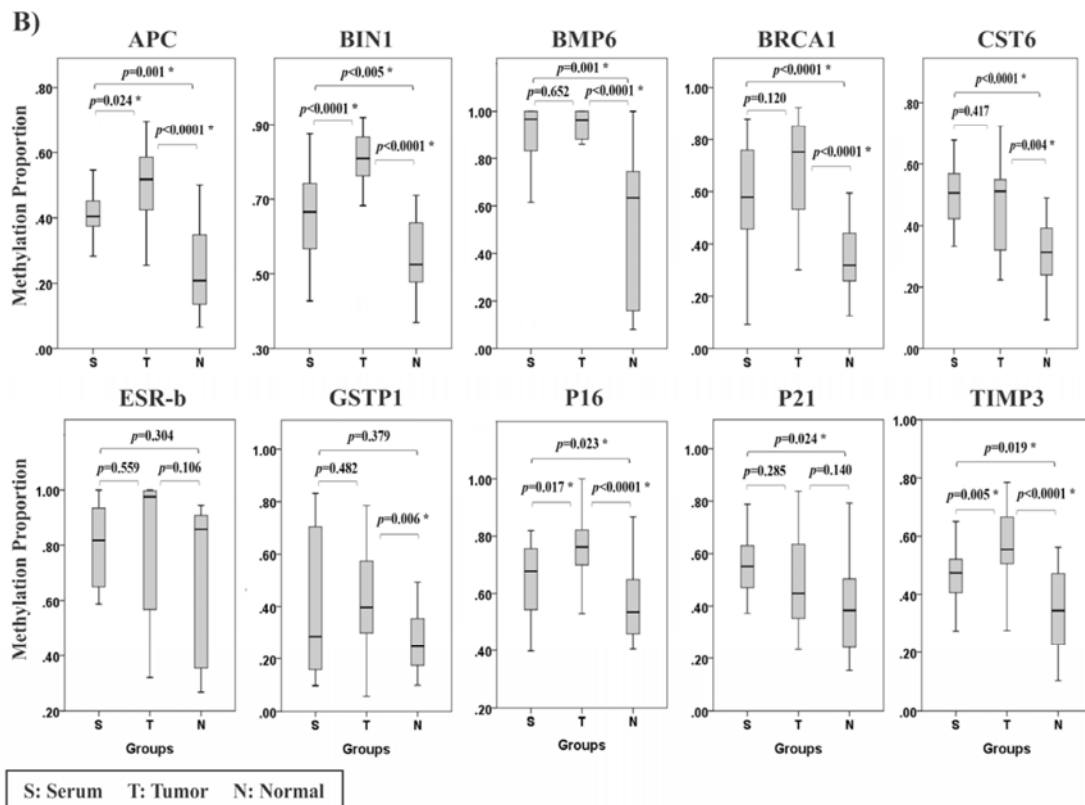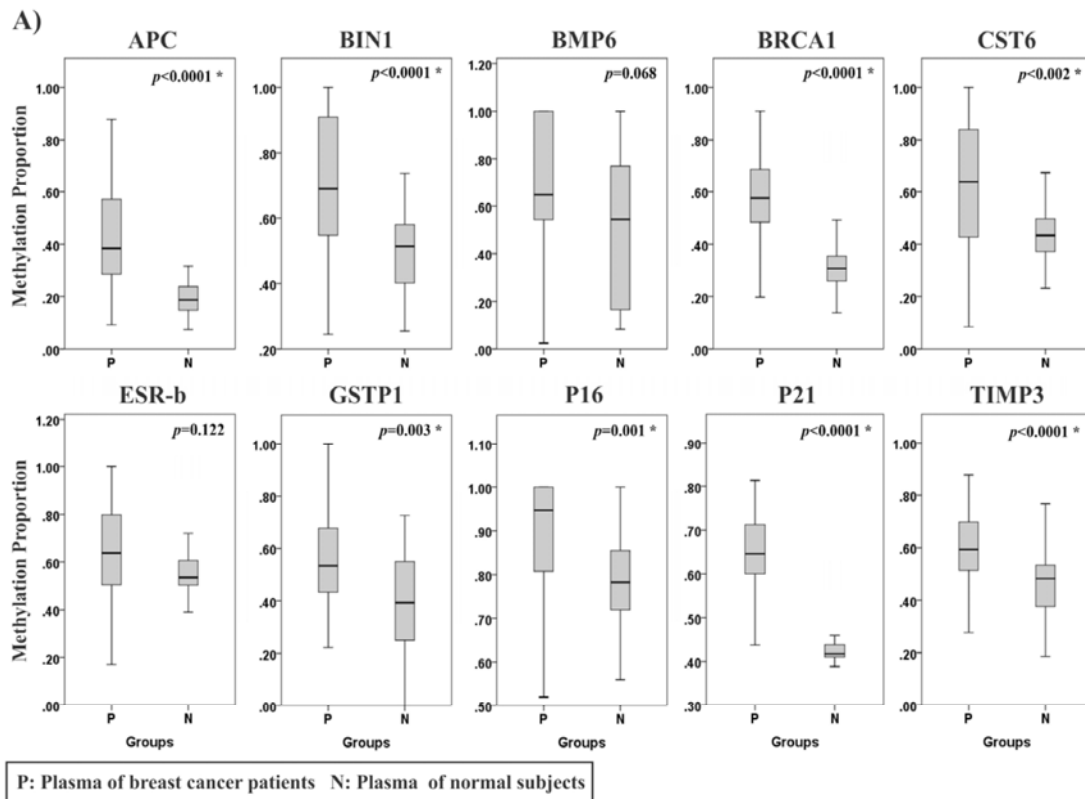
Figure 2. Comparison between quantitative methylation analyses of 10 candidate genes. A) Thirty six plasma samples of breast cancer patients and 30 plasma samples of normal subjects as control. B) Triple matched samples from 20 breast cancer patients. (* significant difference; Mann-Whitney U Test).
doi:10.1371/journal.pone.0016080.g002

methylation proportion of eight genes in the cohort of plasma samples showed a number of possibly correlated samples into a smaller number of uncorrelated samples (Fig. 1E).

Methylation proportion of candidate genes in triple matched samples. To confirm that hypermethylation of cell free DNA in circulation of breast cancer patients are derived from

tumor tissue, triple matched samples including cancerous tissue, matched normal tissue and serum samples were analyzed. Hierarchical clustering showed significant hypermethylation patterns for serum and tumor tissue compared with normal tissue for seven genes (*APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *P16* and *TIMP3*). The *GSTP1* gene was significantly hypermethylated in
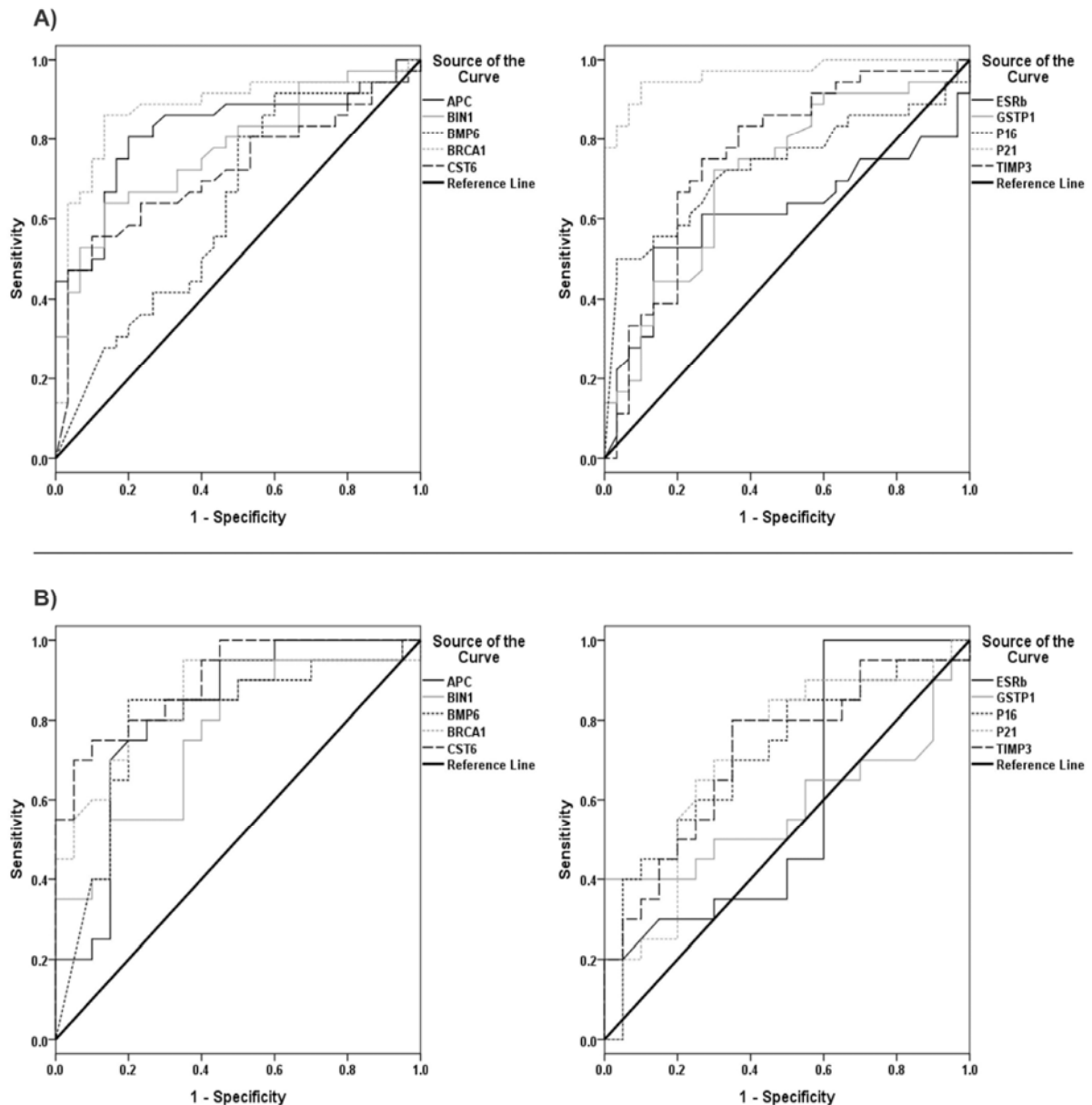
Figure 3. ROC curve analysis using cfDNA for discriminating between cancerous and normal samples based on methylation patterns of 10 candidate genes. A) ROC curves of cfDNA to discriminate between plasma sample of breast cancer patients and plasma samples of normal subjects. B) ROC curves of cfDNA to discriminate between serum with and matched normal tissue samples.
doi:10.1371/journal.pone.0016080.g003

tumor tissue and *P21* gene was mostly hypermethylated in serum samples compared to the normal tissue. For the *ESR-b* gene no significant differences in the methylation extent was observed between studied groups (Fig. 2b). PCA mapping based on the mean methylation proportion of seven genes in the cohort of triple matched samples showed a number of possibly correlated samples into a smaller number of uncorrelated samples (Fig. 1F).The complete methylation quantification data for 10 studied genes is summarized in dataset S1.

A correlation study of the methylation proportion of DNA in serum versus cancerous and normal breast tissue revealed a correlation between tumor tissue and serum for *BMP6*, *BRCA1*, *CST6*, *GSTP1*, *P16* and *TIMP3* genes but not with the matched normal tissue (Dataset S2).

Sensitivity and specificity of a blood based assay to distinguish tumor derived hypermethylated DNA with non-hypermethylated DNA

To find a reliable gene panel which could serve as sensitive and specific blood-based methylation test, gene coverage analysis was assessed for 8 genes with significant different methylation pattern between cancerous and normal plasma. To evaluate the applicability of circulating cfDNA as a biomarker for breast cancer, receiver operating characteristic (ROC) curve analysis was used. Cut-off points, sensitivity, area under the curve (AUC) and confidence interval were calculated for each gene respectively based on at least 90% specificity (Figure 3; Table 3). The methylation quantity over cut-off points were considered as hypermethylation per gene to calculate the methylation frequency in all studied cases (Table 4). As blood-based marker, our designed

panel could cover 91.7% of the plasma samples (92.6% for early stage and 88.9% for late stage of breast cancer) and also covered 95% of serum samples (91.7% for early stage and 100% for late stage of breast cancer) (Table 4).

Relationship between promoter methylation and clinicopathological parameters

In this study, associations between the promoter methylation of the 10 studied genes in cfDNA of breast cancer patients and clinicopathological parameters, such as age, histological grade, pathologic stage, lymph node involvement and receptor status were analyzed (Dataset S3).

In plasma samples, promoter hypermethylation of *GSTP1* gene was significantly correlated with higher age ($\geq$50), hypermethylation of *P16* gene was correlated with pathologic early stage of cancer and hypermethylation of *BMP6* gene was correlated with lymph node involvement ($P<0.05$) (Dataset S3). In normal cases, there was no significant correlation between methylation proportion of candidate genes and clinicopathological parameters.

In triple matched samples, promoter hypermethylation extent of four TSGs in serum samples showed correlation with clinical parameters (*APC* with histological grade G2; *BMP6* and *CST6* with lymph node involvement; *TIMP3* with pathological late stage and lymph node involvement) ($P<0.05$) (Dataset S3).

Comparison of methylation proportion with recognition sites of well-known transcription factor regions

The methylation proportion and localization of each CpG site in the range of -400 to $+200$ was schematically compared to the

Table 3. ROC curve analysis of plasma and serum samples based on methylation proportion of the 10 genes.

Sample type	Genes	Specificity (%)	Sensitivity (%)	AUC[*]	Asymptotic 95% confidence interval (Lower - Upper Bound)	Cut-off points[**] (methylation quantification)
Plasma	APC	90	50	0.824	0.720–0.928	0.42
	BIN1	90	53	0.776	0.665–0.888	0.67
	BMP6	90	30	0.631	0.493–0.768	0.71
	BRCA1	90	75	0.874	0.780–0.967	0.57
	CST6	90	56	0.718	0.592–0.843	0.58
	ESR-b (ER beta)	90	31	0.611	0.471–0.752	0.76
	GSTP1	90	35	0.712	0.585–0.838	0.66
	P16 (CDKN2A)	90	50	0.732	0.608–0.856	0.91
	P21 (CDKN1A)	90	88	0.965	0.915–1.000	0.65
	TIMP3	90	35	0.761	0.640–0.882	0.68
Serum	APC	90	25	0.814	0.676–0.951	0.46
	BIN1	90	40	0.757	0.608–0.907	0.69
	BMP6	90	45	0.795	0.647–0.943	0.93
	BRCA1	90	60	0.854	0.730–0.977	0.70
	CST6	90	75	0.870	0.809–0.991	0.53
	ESR-b (ER beta)	90	30	0.595	0.409–0.781	0.92
	GSTP1	90	40	0.581	0.393–0.769	0.43
	P16 (CDKN2A)	90	40	0.710	0.546–0.874	0.69
	P21 (CDKN1A)	90	25	0.709	0.540–0.877	0.66
	TIMP3	90	35	0.717	0.557–0.878	0.48

[*]AUC: area under the curve.
[**]Cut-off points were calculated according to 90% specificity.
doi:10.1371/journal.pone.0016080.t003

Table 4. Frequency and coverage of promoter methylation in plasma and serum cfDNA.

Type of tumor	Promoter methylation frequency in *plasma*			Promoter methylation frequency in *serum*		
	Genes	Methylation frequency	Coverage[a]	Genes	Methylation frequency	Coverage[a]
Samples with early stage* of cancer	APC	48.1%	92.6%	APC	25%	91.7%
	BIN1	55.6%		BIN1	41.7%	
	BRCA1	48.1%		BRCA1	58.3%	
	CST6	55.6%		CST6	41.7%	
	GSTP1	29.6%		GSTP1	8.4%	
	P16 (CDKN2A)	59.2%		P16 (CDKN2A)	41.7%	
	P21 (CDKN1A)	40.7%		P21 (CDKN1A)	16.7%	
	TIMP3	33.3%		TIMP3	25%	
Samples with late stage** of cancer	APC	44.4%	88.9%	APC	25%	100%
	BIN1	44.4%		BIN1	12.5%	
	BRCA1	55.5%		BRCA1	12.5%	
	CST6	44.4%		CST6	12.5%	
	GSTP1	22.2%		GSTP1	75%	
	P16 (CDKN2A)	22.2%		P16 (CDKN2A)	37.5%	
	P21 (CDKN1A)	33.3%		P21 (CDKN1A)	37.5%	
	TIMP3	22.2%		TIMP3	62.5%	
All analyzed samples	APC	47.2%	91.7%	APC	25%	95%
	BIN1	52.8%		BIN1	30%	
	BRCA1	50%		BRCA1	40%	
	CST6	52.8%		CST6	30%	
	GSTP1	27.8%		GSTP1	35%	
	P16 (CDKN2A)	50%		P16 (CDKN2A)	40%	
	P21 (CDKN1A)	38.9%		P21 (CDKN1A)	25%	
	TIMP3	30.6%		TIMP3	40%	

[a]Coverage: percentage of cases with methylation in at least one gene in the given panel (i.e., coverage of 100% means that all samples had methylation of at least one gene in the panel).
*The pathologic stage<III was considered as "Early stage".
**The pathologic stage III and IV was considered as "Late stage".
doi:10.1371/journal.pone.0016080.t004

consensus sequences of well-known transcription factors (upstream sites for regulatory enhancers, CAAT box, GC box, transcription factor II-B recognition elements, TATA box, initiation site of transcription and downstream promoter elements) for the both cohorts (Fig. 4).

Hypermethylated CpG sites in the plasma sample from breast cancer patients were almost located between TATA box and initiator, however, in plasma of normal subjects hypermethylated CpGs were randomly distributed and did not show significant association between location of the CpG sites and conserved sequences (Fig. 4a).

The analysis of triple matched samples revealed that the hypermethylated CpG sites in tumor tissue and serum samples were mostly located in a range of −40 to +1 (TATA box and initiator) and −26 to +1 (TATA box), respectively. While in normal samples the CpG sites were differentially methylated and located randomly in the 5′UTRs of the studied genes (Fig. 4b).

Discussion

In this study, we assessed the methylation proportion of more than 31,248 CpG sites on 10 breast cancer candidate genes in 126 different samples consisting of two different cohorts. Using hierarchical clustering in the plasma samples cohort, we found significant promoter hypermethylation of eight genes (APC, BIN1, BRCA1, CST6, GSTP1, P16, P21 and TIMP3) in patients' plasma of breast cancer patients compared with plasma of normal subjects (Fig. 2a). To proof that hypermethylation pattern in plasma cfDNA of breast cancer patients is derived from tumor tissue, triple matched samples including breast cancerous tissue, matched normal tissue and serum samples were analyzed. Two-way hierarchical clustering in serum and tumor tissue showed significant hypermethylation patterns of seven genes (APC, BIN1, BMP6, BRCA1, CST6, P16 and TIMP3) compared with normal tissue (Fig. 2b; Dataset S1). This data revealed the potential of the candidate genes for developing a blood-based test as a predictive and prognostic biomarker for breast cancer.

The pathway analysis showed the involvement of the 10 candidate genes in cell cycle and DNA repair (BRCA1, P16 and P21), invasion and metastasis (CST6 and TIMP3), cell proliferation (ESR-b), signal transduction (APC, BIN1 and BMP6) and cell detoxification (GSTP1), and highlighted their role in breast carcinogenesis. Approximately one-third of the differentially
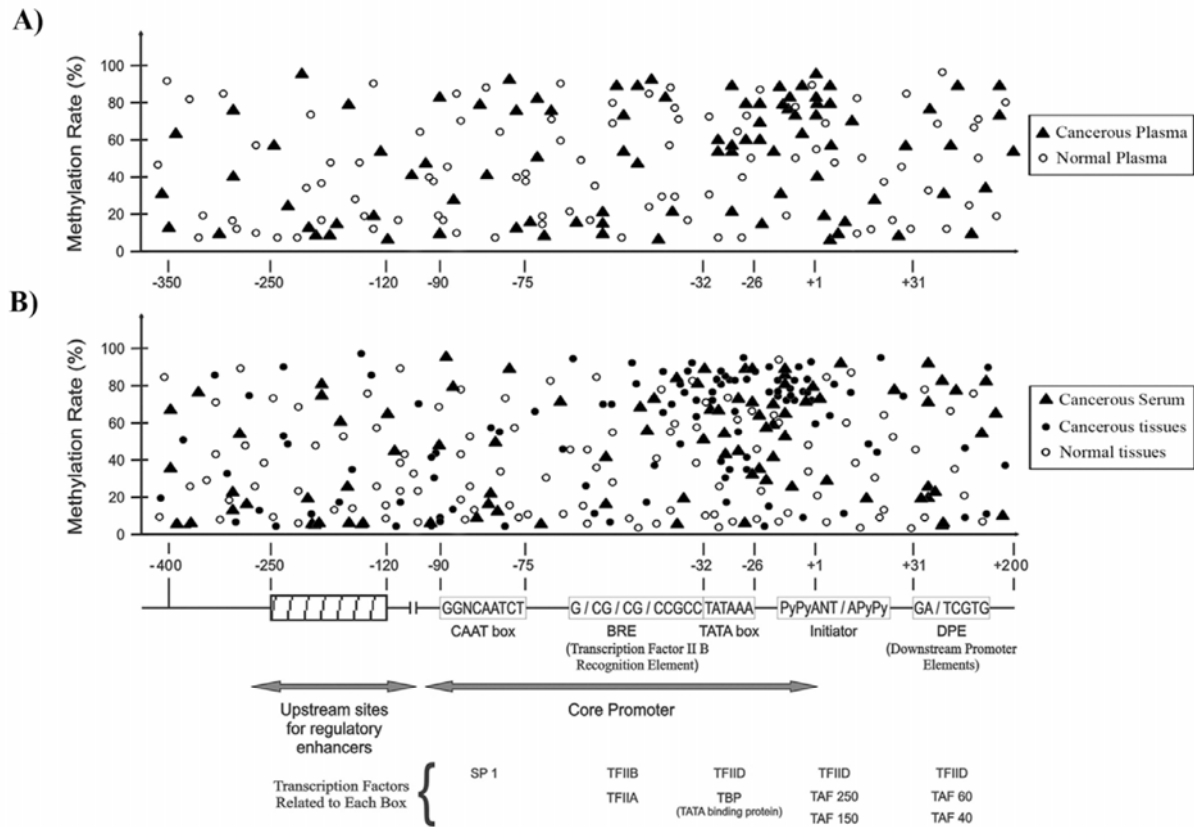
Figure 4. Comparison of the mean methylation proportion and approximate position of informative CpG sites in the range of −400 to +200 according to the recognition sites of the transcription factors in the 10 candidate genes. A) Comparison of methylation proportion in 36 plasma samples of breast cancer patients and 30 plasma samples of normal subjects. B) Comparison of methylation proportion in 60 triple samples (cancerous breast tissue, matched normal tissue and serum samples) from 20 breast cancer patients. (Dots in the map are corresponding to the mean methylation quantity of each CpG site in all analyzed cases).
doi:10.1371/journal.pone.0016080.g004

methylated 5′UTRs are inversely correlated with transcription in normal tissue [17,21]. Similarly, present study showed accumulation of hypermethylated CpG sites of 10 studied genes nearby TATA box and initiator (−40 to +1) in the cancerous and serum samples, however, in normal samples the CpG sites were differentially methylated and located randomly in the 5′-UTRs (Fig. 4). This data verified that hypermethylation of critical part of promoter regions might be the major mechanism for transcription alterations leading to silencing or down regulation of candidate genes. Also hypermethylation of several genes in the same pathway might contribute to tumor aggressiveness.

It has been estimated that more than 90% of the total circulating cfDNA is derived from tumor tissue [22,23]. Several studies reported tumor related genetic and epigenetic alterations in serum and plasma cfDNA of breast cancer patients [24,25,26,27], but studies comparing methylation patterns in tumor and serum DNA in early or late stages of breast tumorigenesis are limited. According to the origin of plasma or serum cfDNA which is released during cell necrosis or apoptosis, it appears that serum tends to contain more DNA than plasma. However, some of this DNA in serum could be due to DNA contamination derived from leukocytes [28]. In our study, there was significant concordance regarding the methylation patterns of seven analyzed genes in serum sample with tumor tissue (Dataset S1 & 2). This result

suggested that cancer specific methylation changes in plasma and serum could be used in developing blood-based tests, which could be applied for risk assessment, earlier diagnosis and monitoring of cancers.

Methylation changes in the process of tumorigenesis are often very heterogeneous and still no single gene has been found to be methylated in all breast cancer types. Therefore it is necessary to use a panel of genes as biomarkers to screen certain type of cancer. Different studies have shown a wide range of gene panels according to the frequently methylated genes in different type of cancer including breast cancer. The coverage and sensitivity of reported panels to detect different types of breast cancer, ranges from 40 to 90% depended on the selected genes [10,17,26,29,30,31,32,33]. In the recent study, applying the eight genes panel could achieve to 91.7% of coverage in sensitive methylation quantification for plasma and 95% for serum samples with more than 90% specificity in both studied cohorts (Table 4).

The variability of the reported gene panels in different studies makes it difficult to compare or combine them and to interpret how promoter methylation would serve as biomarker [34]. The inclusion of genes that may have a key role in breast cancer might help to improve the specificity of a gene panel. Our finding highlights the necessity of using different genes in one panel which increases the coverage of detected cases nearly to 100% (Table 4).

The correlation analysis between methylation proportion of 10 candidate genes in plasma of breast cancer patients and clinicopathological parameters revealed significant correlation of *GSTP1* hypermethylation with higher age (≥50), *P16* hypermethylation with early stage of breast cancer and *BMP6* with lymph node involvement. In the serum samples, promoter hypermethylation of some studied genes was correlated with clinical parameters (*APC* with histological grade G2; *TIMP3* with late stage of breast cancer; *BMP6*, *CST6*, and *TIMP3* with lymph node involvement) (Dataset S3). This data might require further validation by using bigger sample size cohorts.

Technically, we quantified methylation proportion of the candidate genes in cfDNA derived from plasma and serum using T-cleavage assay on MALDI-TOF MS. According to the origin of cfDNA in plasma or serum which is released during cell necrosis or apoptosis, the majority of isolated cfDNA should be poor with regard to quantity and quality and it is difficult to deal with long amplicons for further downstream experiments [5,18,35]. We could overcome this limitation of dealing with fractionated, low concentrated and poor quality DNA, using a specialized re-amplification strategy for performing high-throughput methylation analysis on MALDI-TOF MS based on our established method [18].

Presented data is promising to design a gene panel and develop a blood-based screening method for breast cancer which relies on pathologic methylation changes. Tissue specific and blood-based methylation markers might provide valuable information as prognostic and predictive markers for breast cancer, as well as for developing novel targeted therapeutic strategies.

Additional information

The complete data for high-throughput methylation analysis of informative CpG sites in 10 breast cancer-related genes, including: gene location, amplicon size and two-way hierarchical cluster analysis of two different studied cohorts are illustrated in dataset S1. Scatterplot Matrix (SPLOM) analysis and correlation of the methylation extent of DNA from different are summarized in dataset S2. Correlation study between promoter methylation extent and clinicopathological parameters is shown in dataset S3.

Supporting Information

Dataset S1

- Double dendrogram of analyzed genes: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects.

- Comparison of informative CpG sites in two groups of plasma samples.
- Double dendrogram of analyzed genes: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients.

(PDF)

Dataset S2

- Scatterplot Matrix (SPLOM) analysis for mean methylation proportion of 10 genes in triple samples from 20 breast cancer patients (breast cancerous tissue, matched normal tissue and serum samples).
- Correlation study of the mean methylation proportion of informative CpG sites for ccfDNA in serum versus tumor and normal samples (S: Serum, N: Normal, T: Tumor).
- Scatterplot Matrix (SPLOM) analysis for mean methylation proportion of 10 genes in 36 plasma samples of breast cancer patients and 30 plasma samples of normal subjects.

(PDF)

Dataset S3

- Correlation study between promoter methylation of 10 studied genes and clinicopathological parameters in 36 plasma and 20 serum samples.

(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RR XYZ. Performed the experiments: RR ZB. Analyzed the data: RR ZB CK. Contributed reagents/materials/analysis tools: QL NB CD HZ SS. Wrote the paper: RR ZB CK XYZ. Other: EW JB.

References

1. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, et al. (2003) The case for early detection. Nat Rev Cancer 3: 243–252.
2. Ries L MD, Krapcho M, Mariotto A, Miller B, Feuer E (2006) SEER cancer statistics review, 1975–2004. Bethesda: National Cancer Institute.
3. Radpour R, Barekati Z, Kohler C, Holzgreve W, Zhong XY (2009) New trends in molecular biomarker discovery for breast cancer. Genet Test Mol Biomarkers 13: 565–571.
4. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ (1977) Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 37: 646–650.
5. Chan KC, Zhang J, Hui AB, Wong N, Lau TK, et al. (2004) Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 50: 88–92.
6. Gautschi O, Bigosch C, Huegli B, Jermann M, Marx A, et al. (2004) Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. J Clin Oncol 22: 4157–4164.
7. Giacona MB, Ruben GC, Iczkowski KA, Roos TB, Porter DM, et al. (1998) Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. Pancreas 17: 89–97.
8. Zhong XY, Ladewig A, Schmid S, Wight E, Hahn S, et al. (2007) Elevated level of cell-free plasma DNA is associated with breast cancer. Arch Gynecol Obstet 276: 327–331.
9. Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3: 415–428.
10. Fackler MJ, McVeigh M, Mehrotra J, Blum MA, Lange J, et al. (2004) Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer Res 64: 4442–4452.
11. Widschwendter M, Jones PA (2002) DNA methylation and breast carcinogenesis. Oncogene 21: 5462–5482.
12. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683–692.
13. Wong TS, Kwong DL, Sham JS, Wei WI, Kwong YL, et al. (2004) Quantitative plasma hypermethylated DNA markers of undifferentiated nasopharyngeal carcinoma. Clin Cancer Res 10: 2401–2406.
14. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, et al. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 102: 15785–15790.
15. Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M, et al. (2008) Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci U S A 105: 4844–4849.
16. Radpour R, Haghighi MM, Fan AX, Torbati PM, Hahn S, et al. (2008) High-Throughput Hacking of the Methylation Patterns in Breast Cancer by In vitro

Transcription and Thymidine-Specific Cleavage Mass Array on MALDI-TOF Silico-Chip. Mol Cancer Res 6: 1702–1709.

17. Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, et al. (2009) Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28: 2969–2978.

18. Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, et al. (2009) Simultaneous Isolation of DNA, RNA, and Proteins for Genetic, Epigenetic, Transcriptomic, and Proteomic Analysis. J Proteome Res 8: 5264–5274.

19. Barekati Z, Radpour R, Kohler C, Zhang B, Toniolo P, et al. (2010) Methylation profile of TP53 regulatory pathway and mtDNA alterations in breast cancer patients lacking TP53 mutations. Hum Mol Genet 19: 2936–2946.

20. Radpour R, Barekati Z, Haghighi MM, Kohler C, Asadollahi R, et al. (2010) Correlation of telomere length shortening with promoter methylation profile of p16/Rb and p53/p21 pathways in breast cancer. Mod Pathol 23: 763–772.

21. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38: 1378–1385.

22. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, et al. (2001) DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 61: 1659–1665.

23. Diehl F, Li M, Dressman D, He Y, Shen D, et al. (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A 102: 16368–16373.

24. Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253–266.

25. Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, et al. (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res 63: 7641–7645.

26. Dulaimi E, Hillinck J, Ibanez de Caceres I, Al-Saleem T, Cairns P (2004) Tumor suppressor gene promoter hypermethylation in serum of breast cancer patients. Clin Cancer Res 10: 6189–6193.

27. Shukla S, Mirza S, Sharma G, Parshad R, Gupta SD, et al. (2006) Detection of RASSF1A and RARbeta hypermethylation in serum DNA from breast cancer patients. Epigenetics 1: 88–93.

28. Sunami E VA-T, Nguyen SL, Hoon DSB (2009) Analysis of methylated circulating DNA in cancer patients' blood DNA Methylation. pp 349–356.

29. Parrella P, Poeta ML, Gallo AP, Prencipe M, Scintu M, et al. (2004) Nonrandom distribution of aberrant promoter methylation of cancer-related genes in sporadic breast tumors. Clin Cancer Res 10: 5349–5354.

30. Fackler MJ, McVeigh M, Evron E, Garrett E, Mehrotra J, et al. (2003) DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. Int J Cancer 107: 970–975.

31. Shinozaki M, Hoon DS, Giuliano AE, Hansen NM, Wang HJ, et al. (2005) Distinct hypermethylation profile of primary breast cancer is associated with sentinel lymph node metastasis. Clin Cancer Res 11: 2156–2162.

32. Tao MH, Shields PG, Nie J, Millen A, Ambrosone CB, et al. (2009) DNA hypermethylation and clinicopathological features in breast cancer: the Western New York Exposures and Breast Cancer (WEB) Study. Breast Cancer Res Treat 114: 559–568.

33. Li S, Rong M, Iacopetta B (2006) DNA hypermethylation in breast cancer and its association with clinicopathological features. Cancer Lett 237: 272–280.

34. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465–476.

35. Li Y, Zimmermann B, Rusterholz C, Kang A, Holzgreve W, et al. (2004) Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. Clin Chem 50: 1002–1011.

Dataset S1

High-throughput methylation analysis of CpG sites in 10 candidate genes related to breast cancer.

APC **Gene**

Double dendrogram of *APC* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *APC* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).

Double dendrogram of *BIN1* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *BIN1* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).

BMP6 Gene

Double dendrogram of *BMP6* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *BMP6* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).

Double dendrogram of *BRCA1* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *BRCA1* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).

Double dendrogram of *CST6* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *CST6* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).

Double dendrogram of *ESR-b* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *ESR-b* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).
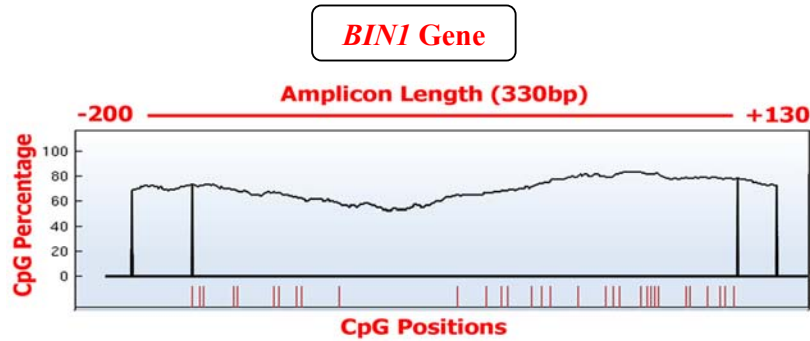
GSTP1 Gene

Double dendrogram of *GSTP1* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).
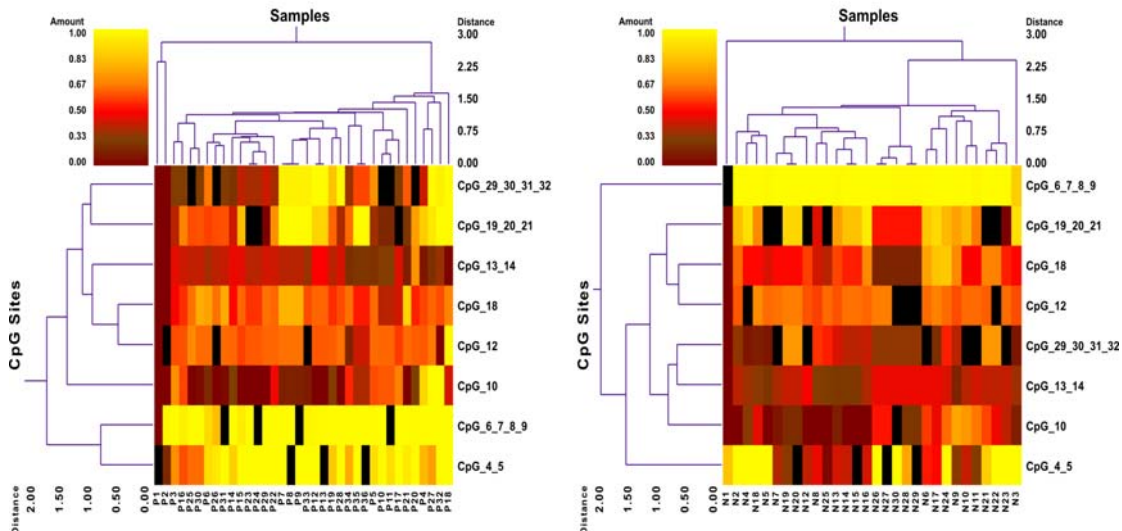

Double dendrogram of *GSTP1* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).
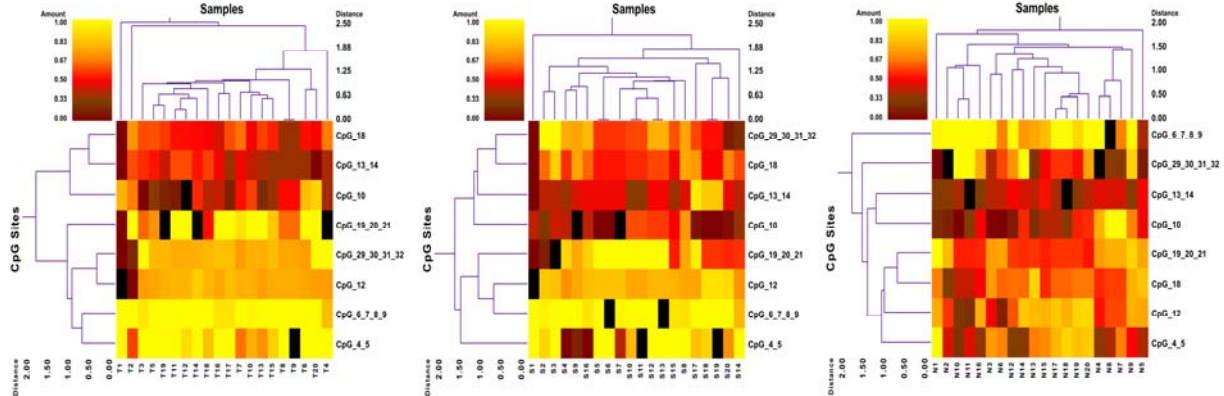
Double dendrogram of *P16* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *P16* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).
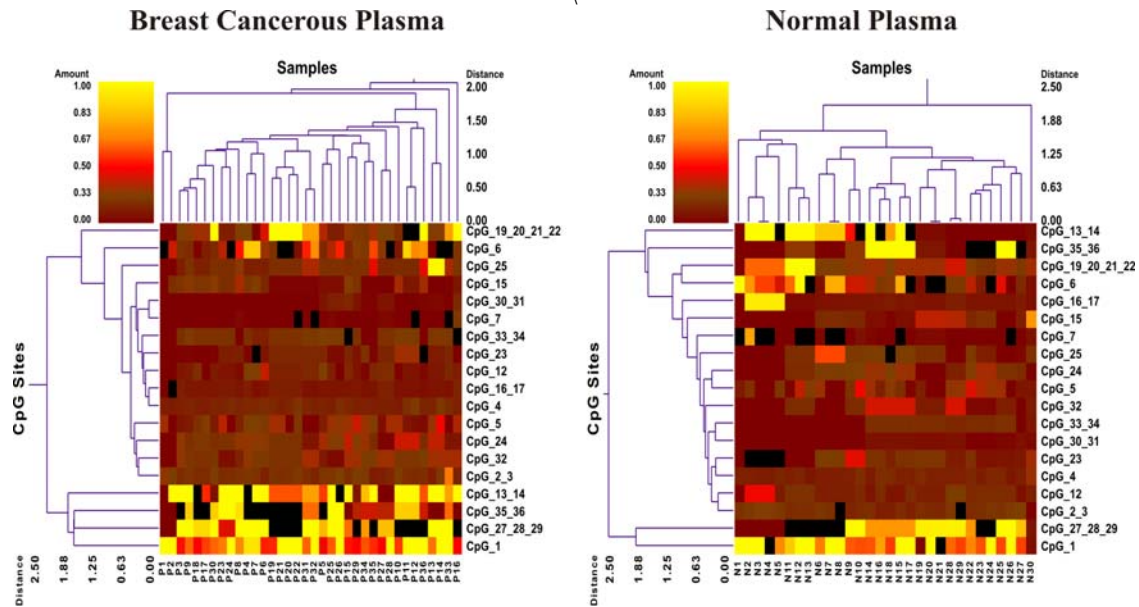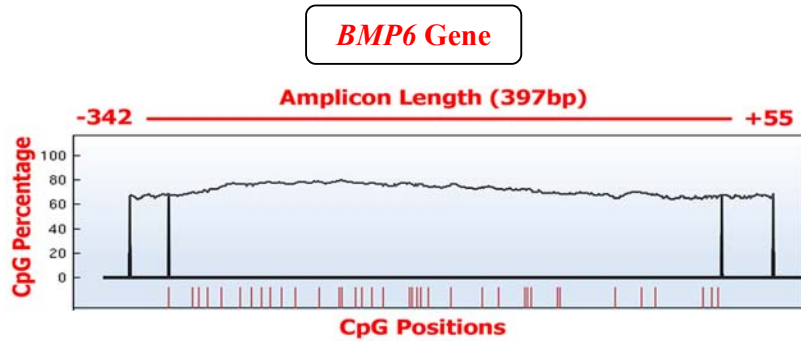
Double dendrogram of *P21* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).
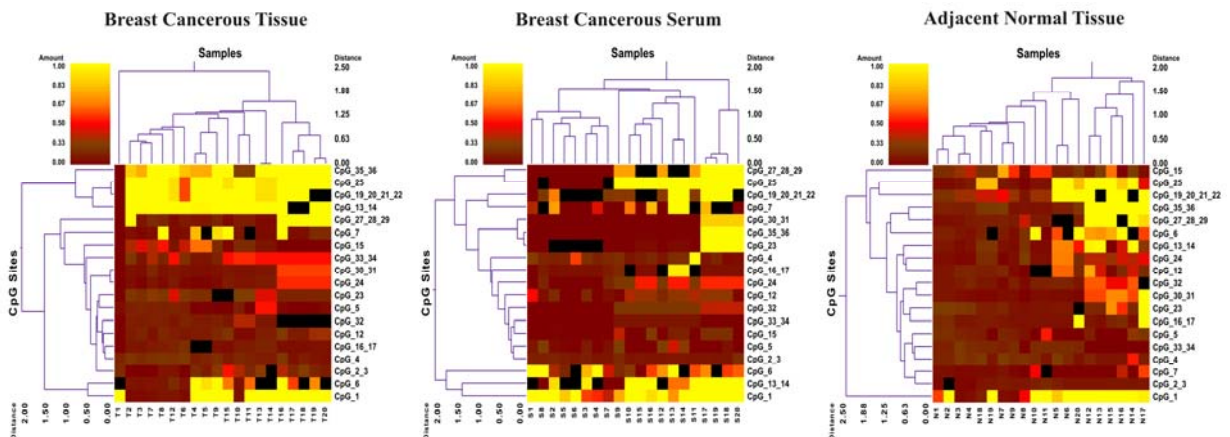

Double dendrogram of *P21* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).
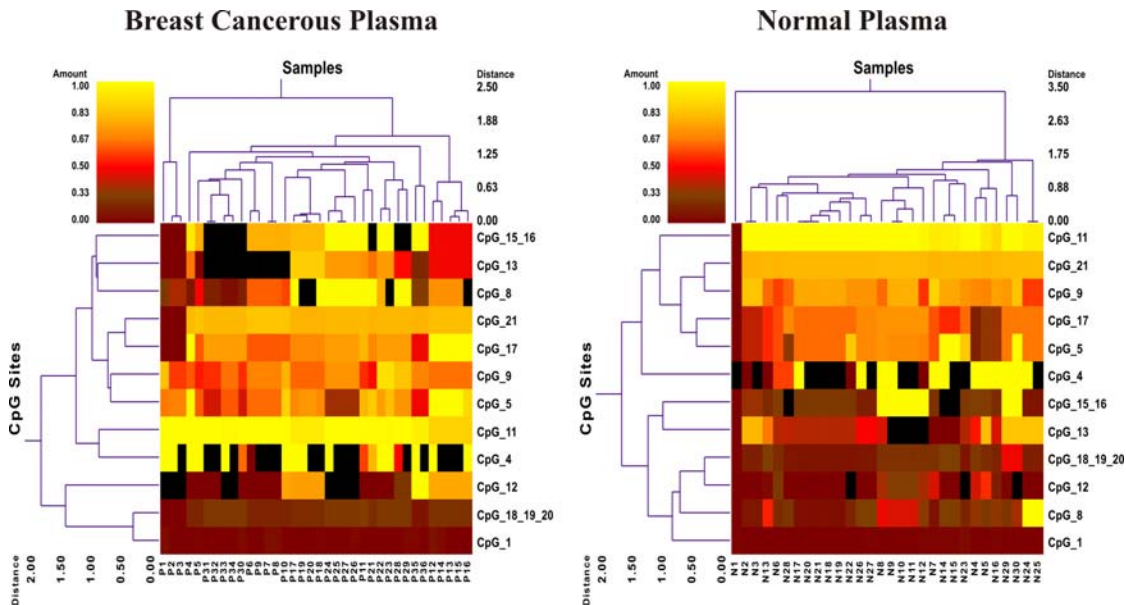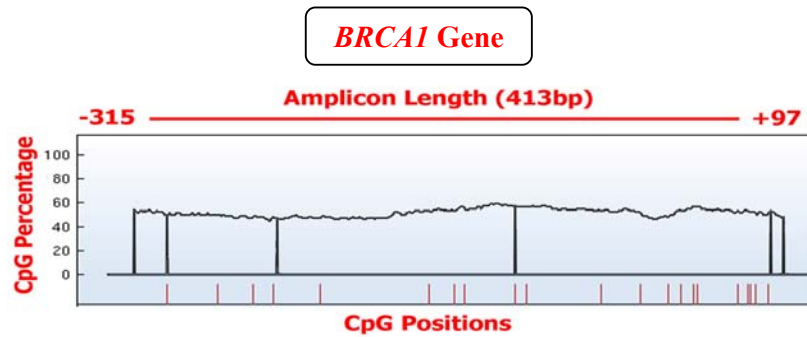
Double dendrogram of *TIMP3* gene: Two-way hierarchical cluster analysis of 36 plasma samples from breast cancer patients and 30 plasma samples of normal subjects. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).

Double dendrogram of *TIMP3* gene: Two-way hierarchical cluster analysis of 60 triple samples (breast cancerous tissue, matched normal tissue serum and samples) from 20 breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpGs).
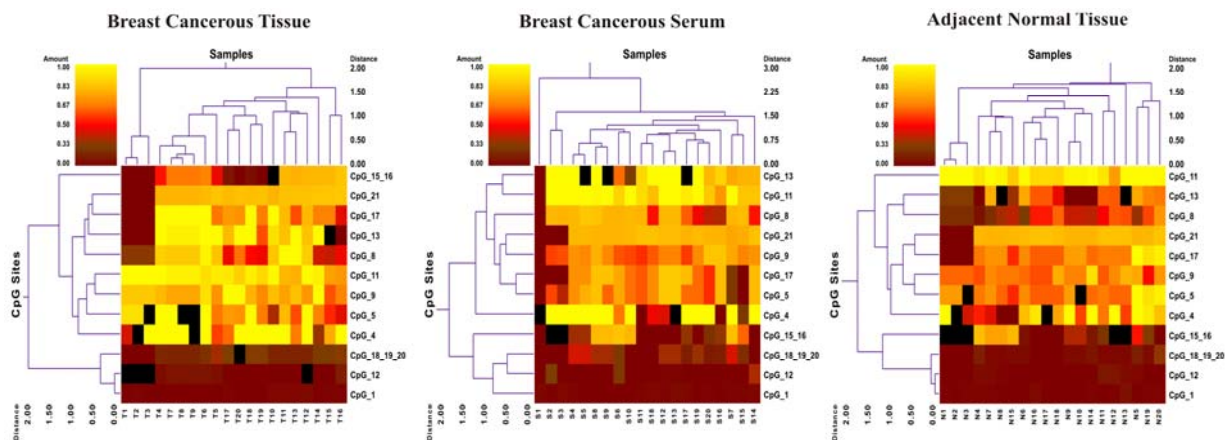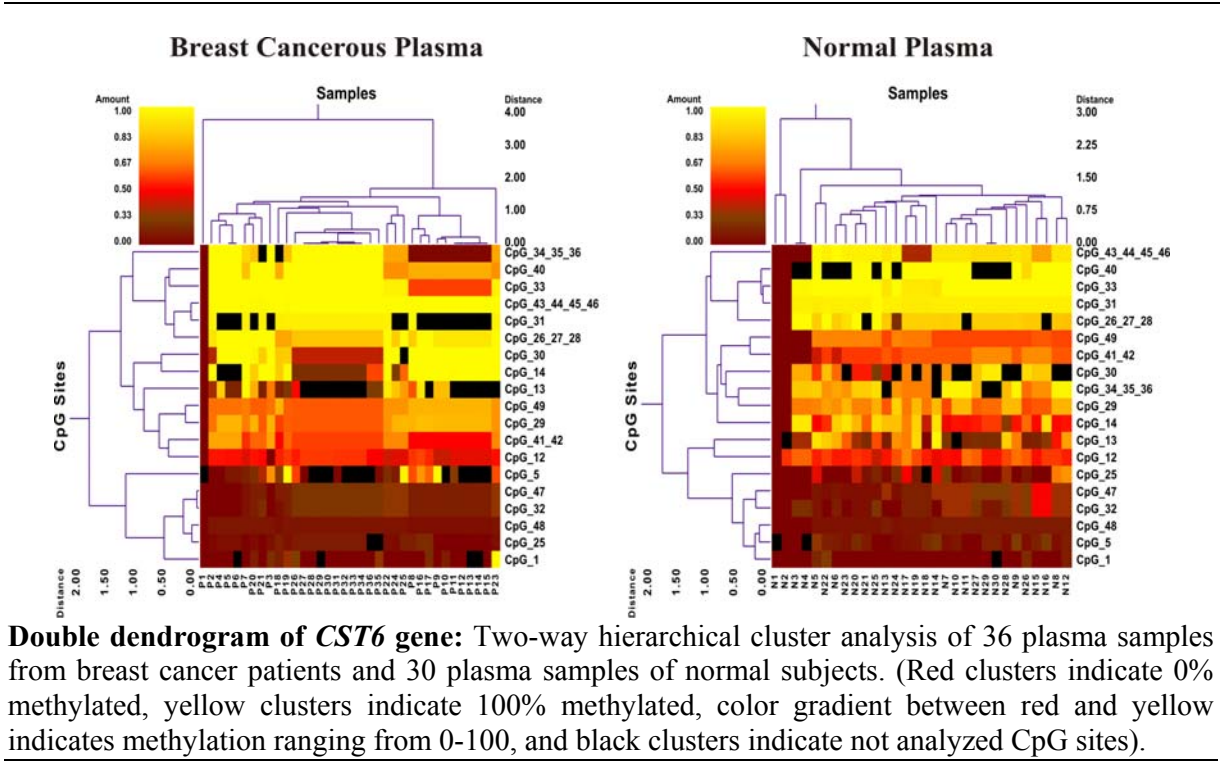
Dataset S2
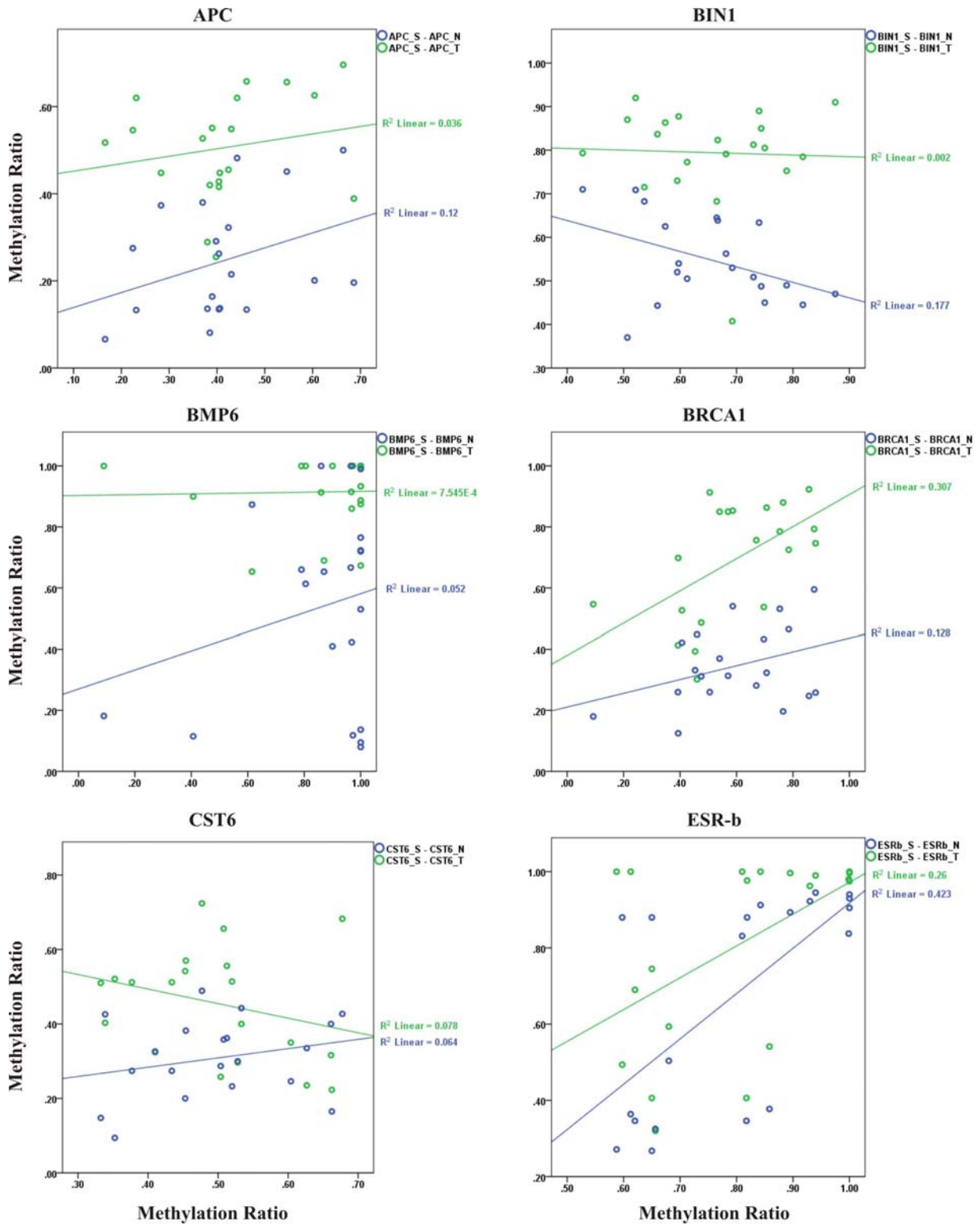
Correlation study of the mean methylation proportion of informative CpG sites for ccfDNA in serum versus tumor and normal samples (S: Serum, N: Normal, T: Tumor).
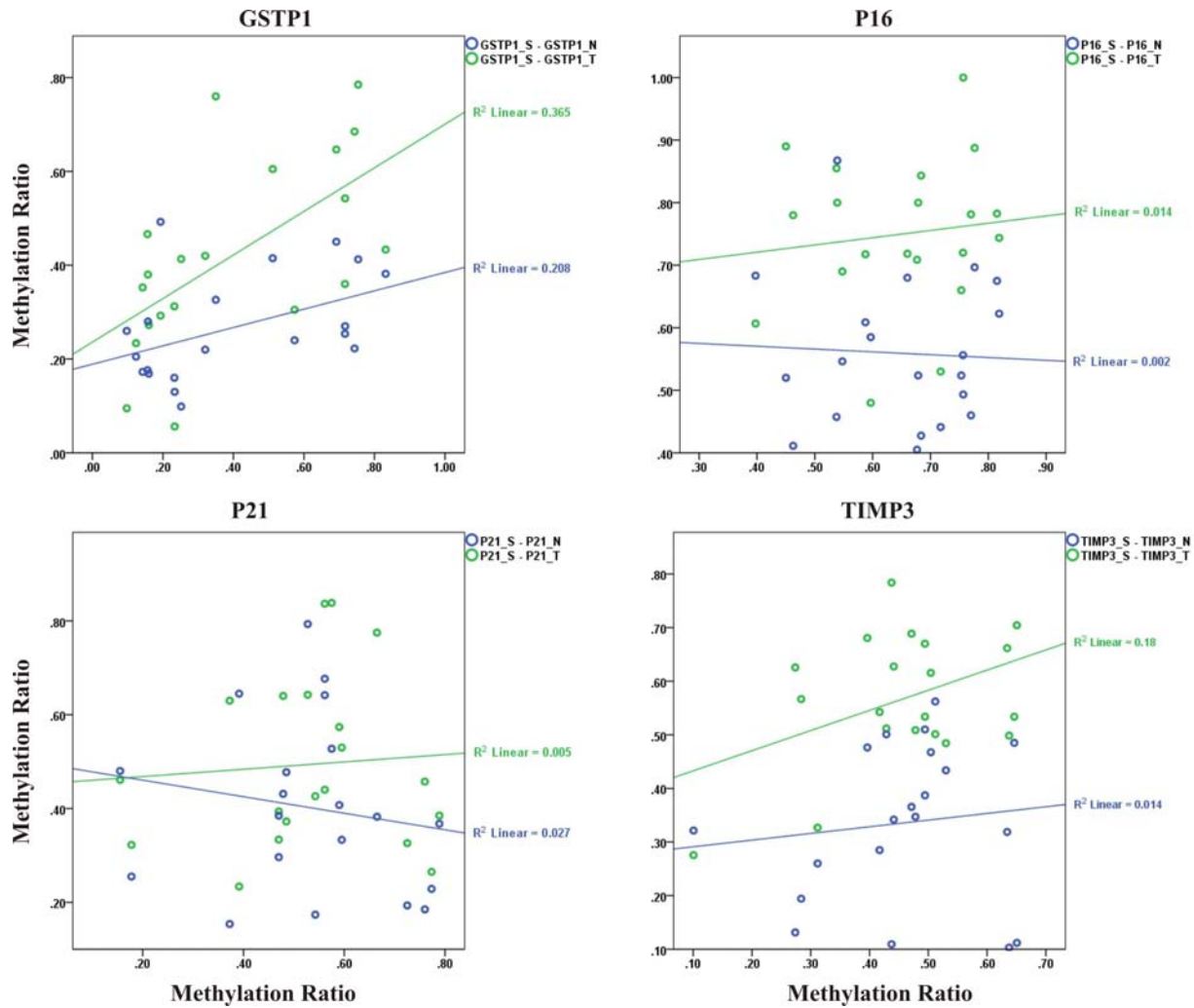
Correlation study of the mean methylation proportion of informative CpG sites for ccfDNA in serum versus tumor and normal samples (S: Serum, N: Normal, T: Tumor).

Dataset S3

Table 1. Correlation study between promoter methylation of 10 studied genes and clinicopathological parameters in 36 plasma samples.

Variables	Group (No. of cases)	Methylation status of 10 studied genes in plasma samples (P-value)									
		APC	BIN1	BMP6	BRCA1	CST6	ESR-b	GSTP1	P16	P21	TIMP3
Age[a]	<50 (8) ≥50 (28)	0.209	0.238	0.577	0.138	0.402	0.834	0.036*	0.098	0.819	0.662
Histological Grade[b]	G1 (11) G2 (18) G3 (7)	0.597	0.239	0.788	0.148	0.077	0.891	0.171	0.337	0.399	0.878
Pathologic Stage[a]	Early (27) Late (9)	0.927	0.674	0.839	0.265	0.297	0.898	0.273	0.017*	0.351	0.927
Lymph node involvement[a]	Positive (19) Negative (17)	0.235	0.295	0.049*	0.401	0.310	0.887	0.466	0.169	0.715	0.173
ER Marker[a]	Positive (28) Negative (8)	0.287	0.170	0.115	0.621	0.313	0.361	0.051	0.780	0.819	0.954
PR Marker[a]	Positive (23) Negative (13)	0.439	0.339	0.265	0.564	0.222	0.729	0.459	0.163	0.882	1.000

a: Mann-Whitney U Test; b: Kruskal-Wallis Test; * Significant correlation.
ER: Estrogen receptor; PR: Progesterone receptor.

Table 2. Correlation study between promoter methylation of 10 studied genes and clinicopathological parameters in 20 serum samples.

Variables	Group (No. of cases)	Methylation status of 10 studied genes in serum samples (P-value)									
		APC	BIN1	BMP6	BRCA1	CST6	ESR-b	GSTP1	P16	P21	TIMP3
Age[a]	<50 (10) ≥50 (10)	0.70	0.705	0.755	0.364	0.545	0.140	0.545	0.151	0.290	0.364
Histological Grade[b]	G1 (0) G2 (5) G3 (15)	0.036*	0.150	0.857	0.965	0.239	0.431	0.206	0.407	0.760	0.570
Pathologic Stage[a]	Early (12) Late (8)	0.671	0.877	0.061	0.064	0.190	0.089	0.123	0.316	0.354	0.011*
Lymph node involvement[a]	Positive (7) Negative (13)	0.874	0.405	0.005*	0.251	0.043*	0.805	0.122	0.322	0.874	0.006*
ER Marker[a]	Positive (4) Negative (16)	0.298	0.508	0.807	0.637	0.147	0.256	0.073	0.925	0.887	0.850
PR Marker[a]	Positive (10) Negative (10)	0.151	0.597	0.198	0.705	0.705	0.198	0.174	0.174	0.734	0.705

a: Mann-Whitney U Test; b: Kruskal-Wallis Test; * Significant correlation.
ER: Estrogen receptor; PR: Progesterone receptor.

8. Published research article:

Correlation of Telomere Length Shortening with Promoter Methylation Profile of p16/Rb and p53/p21 pathways in Breast Cancer

Summary:

In this study, telomere shortening and its potential correlation with down regulation of cell cycle regulatory elements were studied by examination of telomere length and methylation status of the p16INK4a/Rb and/or p53/p21cip1 promoters in a cohort of 104 samples including cancerous and paired adjacent normal breast tissues from 52 patients. The result showed significantly shortened telomere length in the cancerous tissues especially in the grade two and three of breast cancer. After correlation study between the ratio of relative telomere length in the cancerous tissues and the traditional pathological parameters and clinical predictive markers, we could detect significant shortened telomere length in the samples with distant metastasis. Telomere shortening in cancer tissues was correlated with hypermethylation of the *TP53*, *P21* and *P16* promoter regions. The involvement of these cell cycle regulator pathways may allow for continuous cell division and critical telomere shortening. Shortened telomere length and hypermethylation of p53, p21cip1 and p16INK4a promoters may thus serve as biomarkers in breast cancer.

First author s' contribution:

Ramin Radpour was involved in study design, performing the experiment, data analysis and writing the manuscript.
Zeinab Barekati was involved in performing the experiment and writing the manuscript.

Correlation of telomere length shortening with promoter methylation profile of p16/Rb and p53/p21 pathways in breast cancer

Ramin Radpour[1,5], Zeinab Barekati[1,5], Mahdi Montazer Haghighi[2], Corina Kohler[1], Reza Asadollahi[1], Peyman Mohammadi Torbati[3], Wolfgang Holzgreve[4] and Xiao Yan Zhong[1]

[1]Laboratory for Prenatal Medicine and Gynecologic Oncology, Women's Hospital/Department of Biomedicine, University of Basel, Basel, Switzerland; [2]Department of Biology, Faculty of Science, Islamic Azad University, East of Tehran branch, Tehran, Iran; [3]Department of Pathology, Shaheed Beheshti Medical University, Tehran, Iran and [4]Department of Medicine, University Medical Center, Freiburg, Germany

Unregulated cell growth, a major hallmark of cancer, is coupled with telomere shortening. Measurement of telomere length could provide important information on cell replication and proliferation state in cancer tissues. Telomere shortening and its potential correlation with downregulation of cell-cycle regulatory elements were studied by the examination of relative telomere length and methylation status of the *TP53*, *P21* and *P16* promoters in tissues from breast cancer patients. Telomere length was measured in 104 samples (52 tumors and paired adjacent normal breast tissues) by quantitative PCR. Methylation profile of selected genes was analyzed in all samples using a matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI–TOF MS). Our results demonstrated a significant shortening of tumor telomere regions compared with paired adjacent normal tissues ($P < 0.001$). Similarly, telomere lengths were significantly shorter in advanced stage cases and in those with higher histological grades ($P < 0.05$). Telomere shortening in cancer tissues was correlated with a different level of hypermethylation in the *TP53*, *P21* and *P16* promoters ($r = -0.33$, $P = 0.001$; $r = -0.70$, $P < 0.0001$ and $r = -0.71$, $P < 0.0001$, respectively). The results suggested that inactivation of p16/Rb and/or p53/p21 pathways by hypermethylation may be linked to critical telomere shortening, leading to genome instability and ultimately to malignant transformation. Thus, telomere shortening and promoter hypermethylation of related genes both might serve as breast cancer biomarkers.
Modern Pathology (2010) **23**, 763–772; doi:10.1038/modpathol.2009.195; published online 15 January 2010

Keywords: telomere length; breast cancer; biomarker; DNA methylation; MALDI-TOF MS; quantitative real-time PCR

Telomeres are specific repeat sequences $(TTAGGG)_n$ located at chromosome ends. They have a key role in the maintenance of chromosomal stability.[1] The TTAGGG repeats shorten with each cell division because of end replication mispairing, oxidative damage and other end processing events.[2,3] Tumor cells have extremely short telomeres[4,5] in association with increased genomic instability.[6,7] This could suggest that telomere changes are implicated in cancer pathways. Similarly, Meeker *et al*[8] observed that telomere length abnormalities occur early in epithelial carcinogenesis. Therefore, telomere length may serve as a useful biomarker in human cancers.

Regulation of p16/Rb and p53/p21 pathways are important proliferation control mechanisms that are linked to telomere shortening in human cells.[9] Inactivation of p53 enables continued proliferation of cells with dysfunctional telomeres, ultimately promoting chromosomal instability and transformation,[10,11] whereas expression of p16 inhibits phosphorylation of the retinoblastoma protein (pRB), thus preventing cell-cycle progression.[12] Recent

data suggest that the severe genome instability present during telomere crisis, promotes secondary genetic changes that facilitate carcinogenesis.[13,14] Other studies indicated that upregulation of *p16* in senescent cells may also be a consequence of telomere dysfunction.[15]

Optimal telomere length determination is achieved by real-time polymerase chain reaction (PCR).[16] This method relies on the use of telomere primers consisting of a six nucleotide repeated pattern containing four consecutive paired bases, followed by two mismatched bases and a unique 5' sequence. The technique permits calculation of a ratio of telomere repeat copy number to a chosen internal control gene copy number as a relative measure of telomere length unit.

The SEQUENOM's EpiTYPER assay is a methylation quantification method which relies on MALDI–TOF MS.[17] The robustness of this approach for DNA methylation quantification has been previously confirmed by ulterior studies.[18–20] In the current report, this assay was used to quantify the methylation of *TP53*, *P21* and *P16* promoter regions.

The aim of this study was to investigate a potential link between promoter hypermethylation of the *TP53*, *P21* and *P16* genes and telomere length shortening. Paired breast tumor and adjacent normal breast tissues were analyzed by quantitative PCR and MALDI–TOF MS to measure relative telomere length and promoter methylation level of aforementioned genes, respectively. In addition, we compared telomere length in tumor tissues to traditional pathological parameters and clinical predictive markers.

Materials and methods

Samples

The study was approved by the local institutional review board. Cases with demonstrated *BRCA1* or *BRCA2* germline mutations or primary invasive breast carcinoma were excluded. Size of each tumor was evaluated by calculation of surface area (in cm^2) based on the measurement of its two greatest axes. Samples were obtained from 52 paraffin-embedded breast cancer tissues and 52 paired adjacent normal tissues. The corresponding embedded tissue in paraffin blocks of both tumor and adjacent normal tissue were sectioned in $5\,\mu m$ thickness for immunohistochemical staining study. The pathological type, grading and staging were confirmed from the original microscopy slides, which were reviewed separately by two experienced pathologists. Staging and grading was evaluated according to the WHO histological classification.

Remaining part of the paraffin blocks were sectioned (thickness, 10–20 μm) from entirely neoplastic and adjacent normal tissue for DNA extraction. Three to five tissue sections (around 100 mg of tissue) were subjected for DNA extraction using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany).

Immunohistochemical Staining

The standard avidin–biotin–peroxidase-complex and heat-induced antigen retrieval using the microwave method was applied for all immunohistochemical staining. The analysis for the detection of estrogen receptor (ER), progesterone receptor (PR) and HER2/neu (C-ErbB-2) proteins was carried out on $5\,\mu m$-thick sections of paraffin-embedded tissue blocks using primary antibodies for ER (Dako Ltd, Cambridgeshire, UK; clone ID-5, 1:50 dilution), PR (Dako, clone PgR, 1:300 dilution) and HER2/neu antibody (Dako, clone PN2A, 1:200 dilution). Horseradish peroxidase–streptavidin was added and incubated for 30 min followed by dimethylaminoazobenzene (DAB) treatment for 10 min.

Appropriate positive and negative controls were included with each immunohistochemical staining. Normal breast tissues from the studied cases served as an additional internal control. The Quick Score method was used for semi-quantification of ER and PR status as follows: The slides were assessed for average degree of the staining at low power ($\times 10$), and the following scores were allocated: negative (0), weak (1), moderate (2) or strong (3). The percentage of cells with positive nuclei was counted at high power ($\times 40$), and the following scores were allocated: <25% (1), $25 \leq 50\%$ (2), $50 \leq 75\%$ (3) and >75% (4). HER2/neu immunoreactivity was scored according to the conventional scoring system as follows: (0) negative, (1+) negative, (2+) equivocal or weakly positive and (3+) strongly positive.

Quantitative Assessment of the Telomere Length

Telomere repeated sequences and glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*) amplification was carried out for each sample by real-time PCR. *GADPH* was used as internal control (nuclear DNA (nDNA) reference) against which DNA quantification normalization was carried out. Primers were as follows: for *GAPDH*, forward 5'-CCCCACACAC ATGCACTTACC-3' and reverse 5'-CCTAGTCCCA GGGCTTTGATT-3'; for telomere DNA (tDNA), forward 5'-CGGTTTGTTTGGGTTTGGGTTTGGGTTTG GGTTTGGGTT-3' and reverse 5'-GGCTTGCCTTAC CCTTACCCTTACCCTTACCCTTACCCT-3'.[16] Detection and quantification were carried out with the ABI Prism 7000 Sequence Detector and ABI Prism 7000 SDS software (Applied Biosystems, Foster City, CA, USA), respectively.

Reaction was carried out with $5\,\mu l$ of template DNA, $12.5\,\mu l$ SYBR Green PCR Master Mix (Applied Biosystems) and $7.5\,\mu l$ of each $10\,\mu M$ primers, for a final reaction volume of $25\,\mu l$. Thermal cycling was carried out according to the following reaction

Telomere length and promoter methylation
R Radpour *et al*
                                                                                    765

sequence: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 54 °C for 2 min for the telomere amplification, and 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min for *GAPDH*. Following amplification, a dissociation curve was drawn to confirm the specificity of the reaction. Standard and dissociation curves were generated with the ABI Prism 7000 SDS software. R^2 for each standard curve was >0.98.

We examined the amplification efficiency for both *GAPDH* and telomere DNA with a set of serial dilutions, specifying a tDNA and a 97-bp *GAPDH* amplicon concentrations ranging from 3.125×10^4 to 10 pg/μl (including 31 250, 6250, 1250, 250, 50 and 10 pg/μl). There was good correlation of tDNA and *GAPDH* signal on serial dilutions, with comparable efficiencies of amplification. The average threshold cycle number (Ct) values of *GAPDH* and tDNA were obtained from each case in this study. The relative telomere length was calculated using the average Ct of telomere DNA and *GAPDH* (ΔCt = Ct *GAPDH* − Ct tDNA) in the same sample as an exponent of 2 ($2^{\Delta Ct}$). All samples were run in parallel triplicate and the median value was used for calculations. In each run, a standard curve and a negative control (water) were included.

High-Throughput Methylation Analysis using Thymidine-Specific Cleavage Mass Array on MALDI–TOF Silico-Chip

The EpiTYPER assay using MALDI–TOF MS and MassCLEAVE reagent was assessed for high-throughput analysis of DNA methylation patterns of three tumor suppressor gene promoters (*TP53*, *P21* and *P16*) according to the previously published methods.[17,19–21]

Briefly, bisulfite conversion of the target sequence was obtained using the Epitect bisulfite kit (QIAGEN AG, Basel, Switzerland). Primers were designed to cover the promoter regions with the most CpG sites. A T7-promoter tag was added to the reverse primer and a 10mer-tag sequence was added to the forward primer to balance the PCR primer length.[20] Unincorporated dNTPs were dephosphorylated by alkaline phosphatase treatment (SAP; SEQUENOM, San Diego, CA, USA). Typically, 2 μl of the PCR were directly used as a template in a 5 μl of *in vitro* transcription reaction. In all, 20 U of T7 R&DNA polymerase (Epicentre, Madison, WI, USA) was used to incorporate dTTP in the transcripts. Ribonucleotides and dNTPS concentrations were 1 and 2.5 mmol/l, respectively. In the same step, the RNase A (SEQUENOM) was added to cleave the *in vitro* transcript (T-cleavage assay). The mixture was robotically dispensed (nanodispenser) onto silico chips preloaded with matrix (SpectroCHIP; SEQUENOM). Mass spectra were collected using a MassARRAY Compact MALDI–TOF (SEQUENOM) and spectra's methylation ratios were generated by the Epityper software v1.0 (SEQUENOM).

Statistical Analysis

Data analysis was carried out using the SPSS software (Statistical Software Package for Windows, version 17). The Shapiro–Wilk and Kolmogorov–Smirnov tests were used to data distribution analyzis. Both tests similarly demonstrated that our data set was not normally distributed ($P = 0.000/0.000$ by Shapiro–Wilk test and $P = 0.000/0.001$ by Kolmogorov–Smirnov test for tDNA assay and *GAPDH* assay, respectively). The relative telomere length is given as the median. Wilcoxon signed ranks test was used to compare the differences between ranks of each paired samples. The Mann–Whitney *U*-test and Kruskal–Wallis test were used to compare the shortened telomere length in the cancer tissues and paired normal tissues.

Quantitative methylation status of the three previously stated tumor suppressor promoters was compared with relative telomere length. Using the two-way hierarchical cluster analysis, the most variable CpG fragments for each gene were clustered based on pair-wise Euclidean distances and linkage algorithm for all studied samples according to the previously developed method.[19–21] For gene clustering, pair-wise similarity metrics were calculated for each gene separately based on methylation ratio of cancer tissues across the adjacent paired normal tissues. The procedure was carried out using the double dendrogram function of the Gene Expression Statistical System for Microarrays (GESS) version 7.1.19 (NCSS, Kaysville, UT, USA). The non-parametric Spearman's ρ-test was used for correlation study between telomere length and promoter hypermethylation of *TP53*, *P21* and *P16* gene promoters using SPSS software.

Results

Pathological Classification of Samples

According to pathological tumor type and immunohistochemical staining, we separated our patient' samples into two subgroups: ductal carcinoma and lobular carcinoma (Figure 1). Breast cancer characteristics, such as staging, histological grading, hormone receptor status and HER2/neu expression from breast cancer patients are listed in Tables 1 and 2.

Telomere Length Quantification and Methylation Analysis using Paraffin Embedded Tissues

The use of formalin-fixed, paraffin-embedded tissue sections, for which there is a significant degree of degradation, fragmentation and chemical modification of the nucleic acids, is a major limitation for this type of study[19,22] because it can reduce the
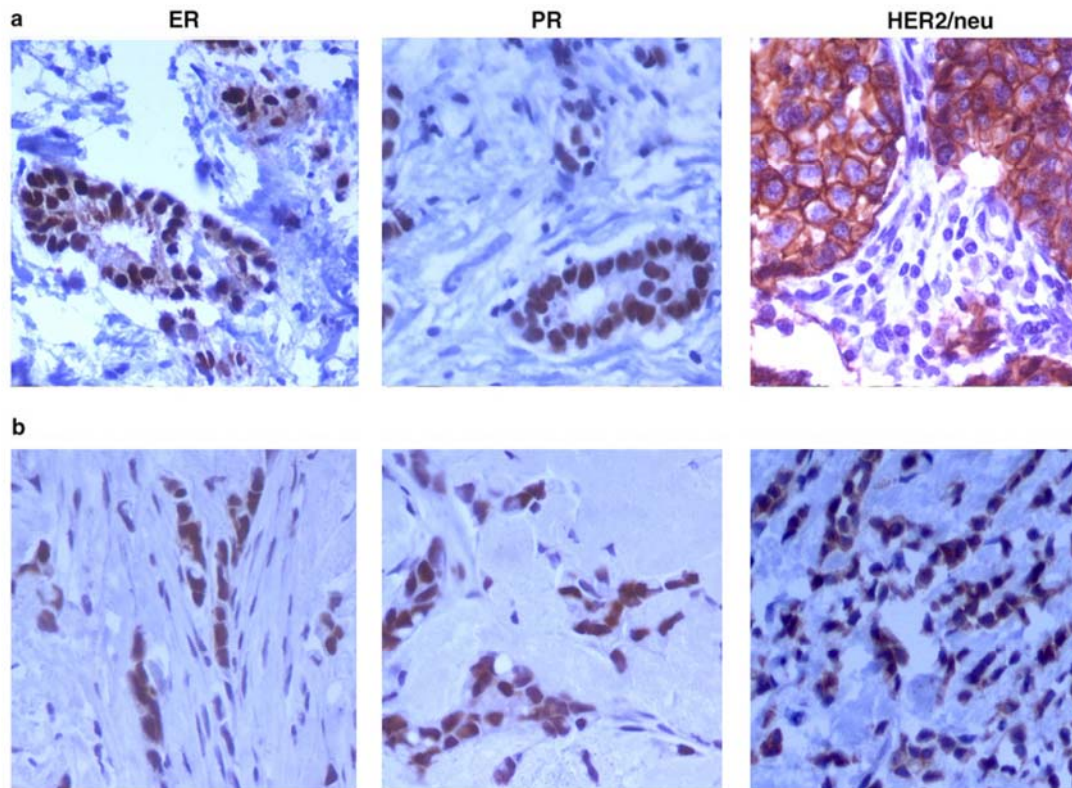
Figure 1 Immunohistochemical staining for estrogen receptor (ER), progesterone receptor (PR) and HER2/neu proteins. (**a**) Ductal carcinoma. (**b**) Lobular carcinoma.

sensitivity of the telomere length measurement.[23,24] In this study, we explored the feasibility of using paraffin-embedded tissues for the quantitative assessment of telomere length together with high-throughput methylation quantification on MALDI–TOF silico-chips.[20,21] Before carrying out telomere length analysis, yield of extracted DNA from formalin-fixed paraffin-embedded tissues was quantified using NanoDrop ND-1000 spectrophotometer (Biolab, Mulgrave, Vic, Australia). Median DNA quantities was determined as 37.92 ng/μl. Samples with <10 ng/μl of DNA were not considered for further experiments and DNA extraction for these samples was repeated.

Relative Telomere Length in Cancer and Paired Normal Breast Tissues

Data analysis demonstrates an 8.58 cycle difference ($\Delta\Delta$Ct, delta Ct) between normal and cancer tissues. The telomere length in cancer tissues was significantly lower than that in normal tissues (Mann–Whitney U-Test: $P < 0.001$) (Figure 2a). Out of 52 paired samples, 41 pairs show telomere length

in normal tissues significantly longer than telomere length in tumors (Wilcoxon signed ranks test: $P < 0.001$).

Correlation Between Shortened Telomere Length and Other Prognostic Factors for Breast Cancer

In this study, associations between the ratio of telomere length in breast cancer tissues and traditional clinical parameters, such as age, tumor type, tumor size, lymph node involvement, extent of metastasis, stage, histological grading, receptor status and pathological biomarkers (HER-2/neu and PS2), were analyzed (Table 2). Shortened telomere length in breast cancer tissues was not associated with age (≥ 50 vs < 50), tumor type (ductal vs lobular), tumor size (T1, T2 and T3), lymph node involvement (N0 and N≥ 1), ER and PR status, HER-2/neu amplification nor presenilin 2 (PS-2) detection (Table 2). However, shorter telomere length was correlated with higher histological grading (grade I vs II and III, $P = 0.007$) (Figure 2c) and distant metastasis ($P = 0.035$) (Figure 3). We could not find any significant telomere length

Telomere length and promoter methylation
R Radpour *et al*
767

Table 1 Clinical characteristics of patient's samples

Breast cancer tumor type	Total no. of patients	Age (years) (mean±s.d. (range))	Side of tumor		Tumor size (cm) (mean±s.d. (range))	No. of patients with lymph node involvement	No. of patients with metastasis	Histological grade		
			R	L				Grade 1	Grade 2	Grade 3
Ductal carcinoma	41	48±11.2 (32–78)	19	22	3.6±3.1 (0.8–12)	33	9	7	15	19
Lobular carcinoma	11	50±10.9 (32–65)	6	5	3.25±1.6 (1.5–6)	6	2	3	3	5

Table 2 Correlation between relative telomere length in cancer tissues and clinical parameters

Variables	Group (cases)	Telomere length (median)	P-value
Age (years)	<50 (33)	0.069	0.294[a]
	≥50 (19)	0.106	
Histological type	Ductal (41)	0.099	0.491[a]
	Lobular (10)	0.035	
Primary tumor	T1 (23)	0.069	0.397[b]
	T2 (16)	0.052	
	T3 (9)	8.439	
Lymph node involvement	Positive (39)	0.069	0.185[a]
	Negative (13)	0.271	
Distant metastasis	No metastasis (41)	4.291	0.035[a]
	Metastasis (11)	0.069	
Stage	I (6)	8.439	0.028[b]
	II (27)	0.541	
	III (8)	1.119	
	IV (11)	0.469	
Histological grading	G1 (10)	2.950	0.07[b]
	G2 (18)	0.036	
	G3 (24)	0.035	
ER	Positive (14)	0.271	0.342[a]
	Negative (18)	0.036	
PR	Positive (18)	0.332	0.138[a]
	Negative (14)	0.041	
HER2/neu	Positive (15)	0.069	0.664[a]
	Negative (17)	0.084	
P53	Positive (4)	2.587	0.185[a]
	Negative (29)	1.069	
PS-2	Positive (12)	0.114	0.836[a]
	Negative (20)	0.524	

ER, estrogen receptor; PR, progesterone receptor.
[a]Mann–Whitney *U*-test.
[b]Kruskal–Wallis test.

changes in the normal samples and their clinical prognostic parameters ($P > 0.05$).

Methylation Status of Three Tumor Suppressor Genes Using MALDI–TOF MS

The quantitative methylation profile of three tumor suppressor promoters (*TP53*, *P21* and *P16*) was compared with relative telomere length in breast cancer and paired normal tissues. The CpG sites were analyzed by T-cleavage assay using MALDI–TOF MS. For these three studied promoters, we analyzed one amplicon per gene containing 83 CpG sites per sample (4316 CpG sites in total).[20] Using the two-way hierarchical cluster analysis, we demonstrated different levels of methylation in the three studied promoters in cancer and adjacent normal tissues (Figure 4). The results showed significant correlation between the telomere length shortening ratio in cancer tissues and the *TP53*, *P21* and *P16* promoters hypermethylation ($r = -0.33$, $P = 0.001$; $r = -0.70$, $P < 0.0001$ and $r = -0.71$, $P < 0.0001$ respectively) (Figure 4).
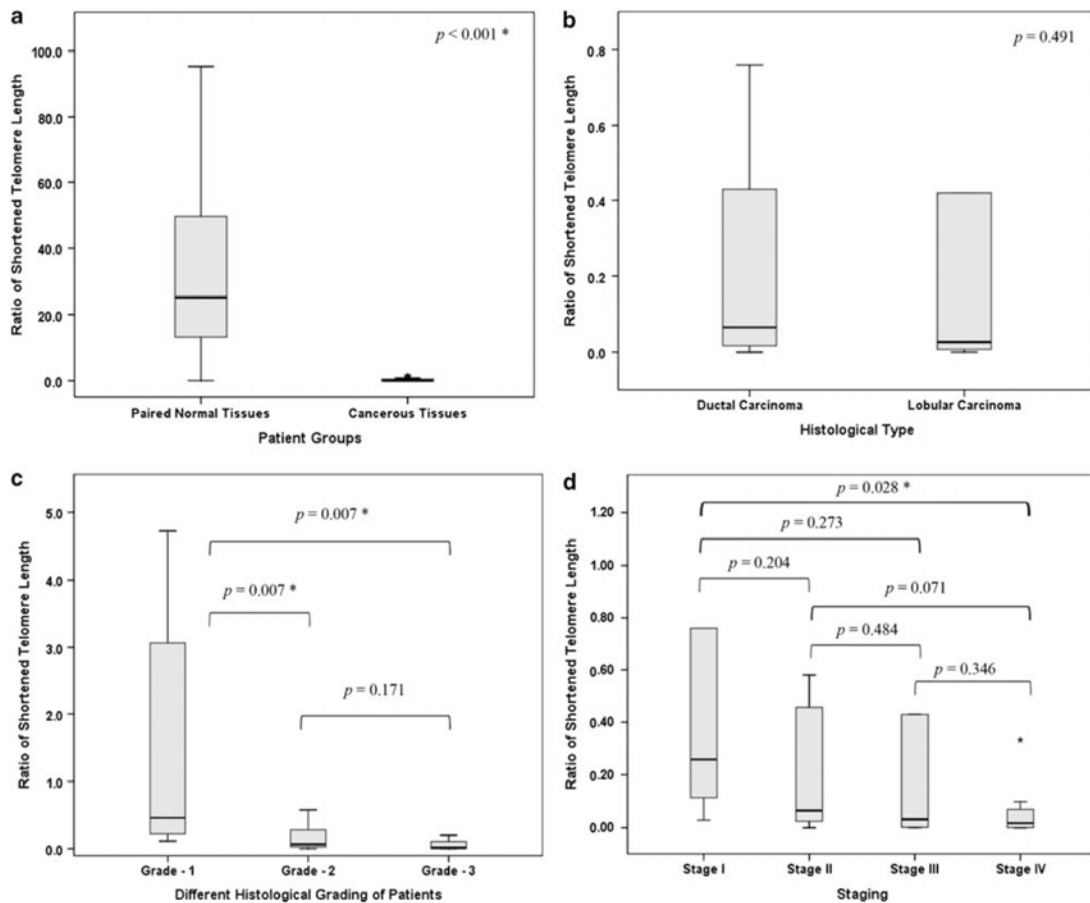
Figure 2 (**a**) Telomere length in the 52 breast cancer tissues and 52 paired normal tissues. (**b**) Telomere length in the two tumor types (ductal carcinoma and lobular carcinoma). (**c**) Telomere length relative to histological grade. (**d**) Telomere length relative to staging. (*significant correlation; Mann–Whitney U-test.)
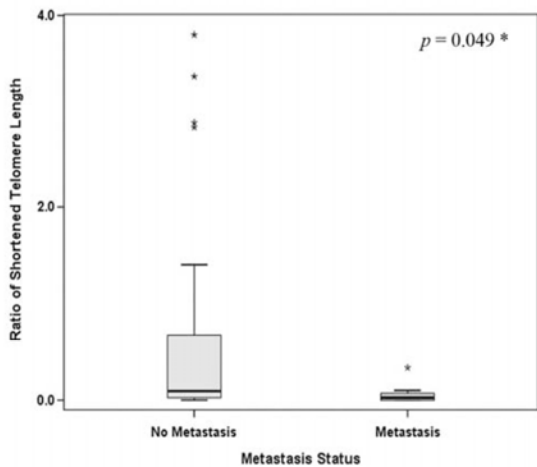
Figure 3 Telomere length in relation to metastasis status. (*Mann–Whitney U-test.)

Discussion

The association between telomere length alterations and cancer has been studied for decades. In the present study, telomere length and promoter methylation status of the p16/Rb and p53/p21 pathways were examined in different grades of breast cancer in 104 paired cancer and normal tissues from 52 patients with breast cancer. Our results show that the average telomere length in breast cancer tissue was significantly shorter than that in the adjacent normal tissue, especially in the histological grades II and III ($P < 0.001$) (Figure 2c). Moreover, hypermethylation of *TP53*, *P21* and *P16* promoters significantly correlated with shortened telomere length in the cancer tissues ($P < 0.001$) (Figure 4).

The most important function of telomeres is the maintenance of genomic integrity and stability.[1,25] Recently, Lin *et al*[24] reviewed the dynamics of telomere length in different types of cancers. Several studies examining telomere length in humans found

Telomere length and promoter methylation
R Radpour *et al*
769
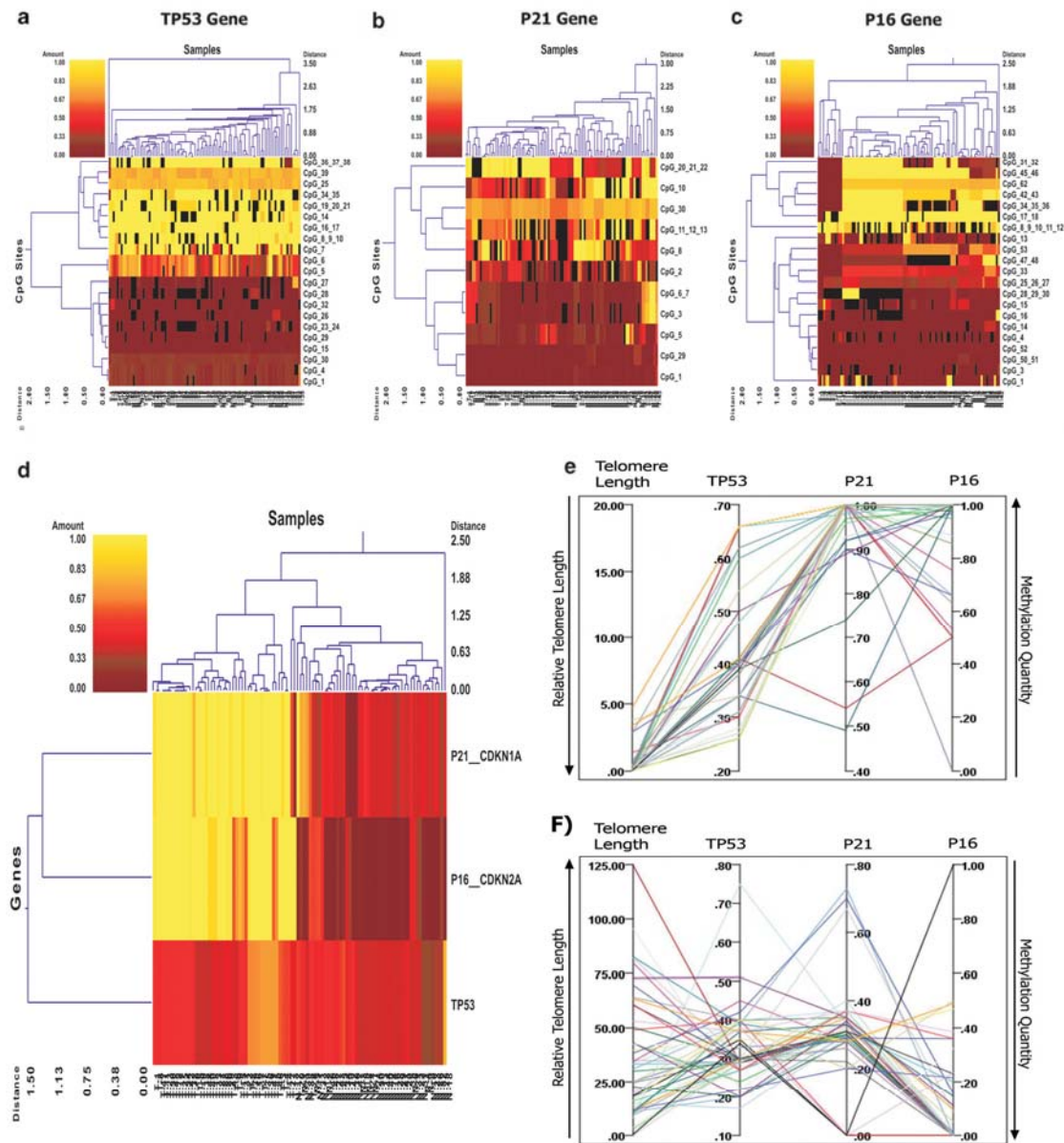
Figure 4 Correlation study between methylation profile of the three tumor suppressor genes (*TP53*, *P21* and *P16*) in cancerous tissues and paired normal tissues with relative telomere length. (**a–c**) High-throughput methylation analysis of three breast cancer-related genes by two-way hierarchical cluster analysis. Double dendrograms of the informative CpG sites for the paired samples of breast cancer patients. (Red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0–100 and black clusters indicate methylation CpG sites not analyzed; T: tumor tissue; N: normal tissue). (**d**) Double dendrogram presenting the methylation profiles of three studied genes (*TP53*, *P21* and *P16*). (**e**) Correlation study between relative telomere length with methylation profile of studied genes in the cancerous tissues. (**f**) Correlation study between relative telomere length with methylation profile of studied genes in the normal tissues. For color figure see online version.

that breast carcinomas had shorter telomeres than normal breast tissue, and high grade (grade III of III) invasive carcinomas had shorter telomeres than low grade (grade I) invasive carcinomas.[26] Our study on paired cancer and normal tissues concurs with those

previous reports, and also demonstrates correlation between shortened telomere length and higher histological grading (Figure 2c and d).

Recently, Svenson *et al*[27] showed that telomere length in peripheral blood cells differs between

breast cancer patients and control subjects, and may serve as a significant prognostic biological marker. In general, malignant tumors show shorter telomeres than corresponding normal tissue, and telomere dysfunction has been indicated as a negative prognostic marker in solid tumors, including breast cancer.[28,29]

Telomere shortening in primary human cells leads to replicative senescence, which is regulated in part by effectors in the p16/Rb and/or p53/p21 pathways.[9,11,30] As well, p53/p21-mediated DNA damage signals are elicited by telomere dysfunction. However, there is still considerable debate on the exact role of telomere attrition on the p16 pathway during senescence.

Inhibition of the p16/Rb and/or p53/p21 pathways enables continuous cell division and critical telomere shortening, a phenomenon known as 'telomere crisis'.[15,31] Using the mass spectrometry on MALDI–TOF silico-chips, we determined quantitative methylation changes of *TP53*, *P21* and *P16* promoters in paired cancer and non-tumor samples. *P21* and *P16* promoters have been reported as being strongly hypermethylated in breast cancer tissues, compared with adjacent normal breast tissues.[20] This study shows significant correlation between telomere length shortening and *TP53*, *P21* and *P16* promoters methylation status (*TP53*:$r = -0.33$, $P = 0.001$; *P21*:$r = -0.70$, $P < 0.0001$ and *P16*:$r = -0.71$, $P < 0.0001$) (Figure 4) in breast cancer tissues. Although our previous study indicated some hypermethylation level of the *TP53* promoter, this hypermethylation level was not significant in breast cancer tissues in comparison with adjacent normal tissues.[20] Our new finding shows a significant correlation between *TP53* hypermethylation and telomere shortening. This suggests that even low level of hypermethylation of *TP53* can influence telomere length shortening. Taken together, these findings could suggest that p16 and p53/p21 may function as a gatekeeper to prevent critical telomere shortening and genome instability.

Changes in telomere length in breast cancer may not be correlated with distant invasion.[32] However, our results showed a significant correlation between telomere length shortening and distant metastasis ($P < 0.05$) (Figure 3). Although we could not find any significant correlation between telomere length and age (≥ 50 *vs* < 50) in cancer tissue (Figure 5a), a significant shortened telomere length was observed in patients over the age of 50 years ($P < 0.05$) (Figure 5b). This finding is in accordance with previous studies on telomere length and aging.[33,34]

It was reported that telomere length is reduced in histologically normal tissues distant at least 1 cm from the adjacent tumor margins.[35] Although the tissue appears normal, it harbors various genetic changes, including telomere dysfunction, that are reflective of the adjacent carcinogenesis process. This seriously complicates the use of histologically normal tissue as control.[35,36] Also examining
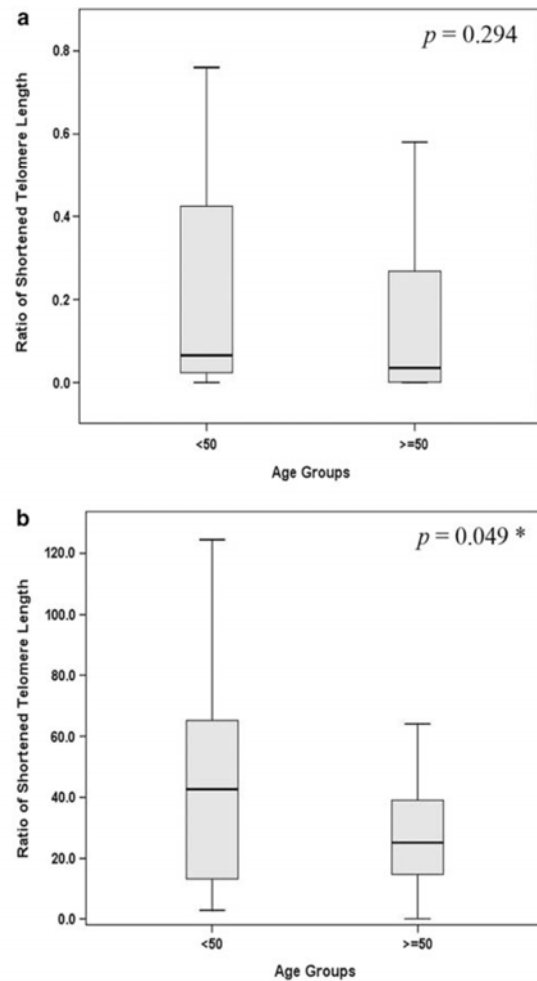
Figure 5 Comparison of telomere shortening in patient's samples separated according to age, in (**a**) cancer tissues (**b**) paired normal tissues. (*significant correlation; Mann–Whitney *U*-test.)

whole tissue sections rather than examining telomere lengths by *in-situ* methodology, which permits distinction of lesional tissue from contaminating normal tissues is one of the limitations of used methodology.[37] To reduce the challenge of contaminating normal tissues, the paired normal samples in our study were selected with distances more than 1.5 cm to the adjacent tumor tissues. In one report, moderate telomere shortening was observed in approximately 50% of histologically benign breast duct lobular units.[37] This kind of telomere shortening might be the result of physiological proliferation, and may delineate the cells at risk for subsequent malignant transformation, but in our work we could not find significant telomere shortening in the far adjacent normal tissues.

In conclusion, our data suggests that shortened telomere length is significantly correlated with carcinogenesis. Moreover, telomere length is also significantly shorter in cancer tissues from patients with distant metastasis compared with those from patients without metastasis. Promoter hypermethylation of the p16/Rb and p53/p21 pathways showed significant correlation with telomere shortening. The involvement of these cell-cycle regulator pathways may allow for continuous cell division and critical telomere shortening. Shortened telomere length and hypermethylation of p53, p21 and p16 promoters might serve as biomarkers in breast cancer.

Acknowledgements

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Blackburn EH. Structure and function of telomeres. Nature 1991;350:569–573.
2 Harley CB. Telomere loss: mitotic clock or genetic time bomb? Mutat Res 1991;256:271–282.
3 Wright WE, Shay JW. Telomere dynamics in cancer progression and prevention: fundamental differences in human and mouse telomere biology. Nat Med 2000;6:849–851.
4 Kim NW, Piatyszek MA, Prowse KR, *et al*. Specific association of human telomerase activity with immortal cells and cancer. Science 1994;266:2011–2015.
5 Shay JW. Telomerase therapeutics: telomeres recognized as a DNA damage signal: commentary re: K. Kraemer *et al.*, antisense-mediated hTERT inhibition specifically reduces the growth of human bladder cancer cells. Clin. Cancer Res., 9: 3794-3800, 2003. Clin Cancer Res 2003;9(10 Pt 1):3521–3525.
6 Hahn WC. Role of telomeres and telomerase in the pathogenesis of human cancer. J Clin Oncol 2003;21:2034–2043.
7 Harley CB. Telomerase and cancer therapeutics. Nat Rev Cancer 2008;8:167–179.
8 Meeker AK, Hicks JL, Iacobuzio-Donahue CA, *et al*. Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. Clin Cancer Res 2004;10:3317–3326.
9 Ohtani N, Yamakoshi K, Takahashi A, *et al*. The p16INK4a-RB pathway: molecular link between cellular senescence and tumor suppression. J Med Invest 2004;51:146–153.
10 Chin L, Artandi SE, Shen Q, *et al*. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. Cell 1999;97:527–538.
11 Artandi SE, Chang S, Lee SL, *et al*. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. Nature 2000;406:641–645.
12 Ben-Porath I, Weinberg RA. The signals and pathways activating cellular senescence. Int J Biochem Cell Biol 2005;37:961–976.
13 Zhang A, Wang J, Zheng B, *et al*. Telomere attrition predominantly occurs in precursor lesions during *in vivo* carcinogenic process of the uterine cervix. Oncogene 2004;23:7441–7447.
14 Gisselsson D, Jonson T, Petersen A, *et al*. Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. Proc Natl Acad Sci USA 2001;98:12683–12688.
15 Jacobs JJ, de Lange T. Significant role for p16INK4a in p53-independent telomere-directed senescence. Curr Biol 2004;14:2302–2308.
16 Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res 2002;30:e47.
17 Ehrich M, Nelson MR, Stanssens P, *et al*. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA 2005;102:15785–15790.
18 Ehrich M, Turner J, Gibbs P, *et al*. Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci USA 2008;105:4844–4849.
19 Radpour R, Haghighi MM, Fan AX, *et al*. High-throughput hacking of the methylation patterns in breast cancer by *in vitro* transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip. Mol Cancer Res 2008;6:1702–1709.
20 Radpour R, Kohler C, Montazer Haghighi M, *et al*. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 2009;28:2969–2978.
21 Radpour R, Sikora M, Grussenmeyer T, *et al*. Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis. J Proteome Res 2009;8:5264–5274.
22 Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 2002;161:1961–1971.
23 Saldanha SN, Andrews LG, Tollefsbol TO. Assessment of telomere length and factors that contribute to its stability. Eur J Biochem 2003;270:389–403.
24 Lin KW, Yan J. The telomere length dynamic and methods of its assessment. J Cell Mol Med 2005;9:977–989.
25 Zakian VA. Telomeres: beginning to understand the end. Science 1995;270:1601–1607.
26 Rha SY, Park KH, Kim TS, *et al*. Changes of telomerase and telomere lengths in paired normal and cancer tissues of breast. Int J Oncol 1999;15:839–845.
27 Svenson U, Nordfjall K, Stegmayr B, *et al*. Breast cancer survival is associated with telomere length in peripheral blood cells. Cancer Res 2008;68:3618–3623.
28 Bisoffi M, Heaphy CM, Griffith JK. Telomeres: prognostic markers for solid tumors. Int J Cancer 2006;119:2255–2260.
29 Heaphy CM, Baumgartner KB, Bisoffi M, *et al*. Telomere DNA content predicts breast cancer-free survival interval. Clin Cancer Res 2007;13:7037–7043.

Telomere length and promoter methylation
772
R Radpour *et al*

30 Kiyono T, Foster SA, Koop JI, *et al.* Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. Nature 1998;396: 84–88.

31 Herbig U, Jobling WA, Chen BP, *et al.* Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). Mol Cell 2004;14:501–513.

32 Mambo E, Chatterjee A, Xing M, *et al.* Tumor-specific changes in mtDNA content in human cancer. Int J Cancer 2005;116:920–924.

33 Aubert G, Lansdorp PM. Telomeres and aging. Physiol Rev 2008;88:557–579.

34 Blasco MA. Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet 2005;6:611–622.

35 Heaphy CM, Bisoffi M, Fordyce CA, *et al.* Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors. Int J Cancer 2006;119:108–116.

36 Dakubo GD, Jakupciak JP, Birch-Machin MA, *et al.* Clinical implications and utility of field cancerization. Cancer Cell Int 2007;7:2.

37 Meeker AK, Hicks JL, Gabrielson E, *et al.* Telomere shortening occurs in subsets of normal breast epithelium as well as *in situ* and invasive carcinoma. Am J Pathol 2004;164:925–935.

9. Submitted research manuscript:

Integrated epigenetics of human breast cancer subtypes: synoptic investigation of targeted genes, microRNAs and proteins upon demethylation treatment

Summary:

Contribution of aberrant DNA methylation in tumorogenesis and silencing of tumor suppressor genes (TSGs) and microRNAs has been thoroughly investigated. Since these epigenetic alterations are reversible, it became of some interest to establish the complete early and late effects of the 5-aza-2'-deoxycytidine (DAC) therapy in breast cancer using a pan-omics approach. In the presented work we have investigated multi-dimensional models to synoptically predict "early and late" systemic stable or transient effects of DAC at the level of the epigenome, the transcriptome, and the proteome for several breast cancer subtypes. In present study, the results give new therapeutic clues based on chemical modification of the pathological methylation patterns in breast cancer as well as investigation of some new molecular DAC targets and involved pathways. The presented approach might become a useful epigenetic treatment model for other human solid tumors in management of cancer patients.

First author's contribution:

Ramin Radpour was involved in study design, making collaborations, performing part of the study, data analysis together with interpretation of biological interactions and writing the manuscript.

Running title: Pan-omics investigation of demethylation treatment

Integrated epigenetics of human breast cancer subtypes: synoptic investigation of targeted genes, microRNAs and proteins upon demethylation treatment

Ramin Radpour[1], Zeinab Barekati[1], Corina Kohler[1], Martin M. Schumacher[2],

Thomas Grussenmeyer[3], Paul Jenoe[4], Nicole Hartmann[2], Suzette Moes[4], Martin Letzkus[2],

Johannes Bitzer[5], Ivan Lefkovits[3*], Frank Staedtler[2*], Xiao Yan Zhong[1*]

[1] Laboratory for Gynecological Oncology, Women´s Hospital/Department of Biomedicine, University of Basel, Switzerland.

[2] Biomarker Development, Novartis Institutes of BioMedical Research, Novartis Pharma AG, Basel, Switzerland.

[3] Department of Biomedicine and Department of cardiac surgery, University hospital Basel, Switzerland.

[4] Biozentrum, University of Basel, Switzerland.

[5] Department of Obstetrics and Gynecology, Women's Hospital, University of Basel, Switzerland.

* Correspondence to:

Prof. Ivan Lefkovits

Head of Proteomics Laboratory, Department of Biomedicine, University hospital Basel, Switzerland.
Tel: +41 61 2673551, E-Mail: ivan.lefkovits@unibas.ch

Dr. Frank Staedtler

Director and head of Genome Technologies, Biomarker Development, Novartis Institutes of
BioMedical Research, Novartis Pharma AG, Basel, Switzerland.
Tel: +41 61 69 61109, E-Mail: frank.staedtler@novartis.com.

Prof. Dr. Xiao Yan Zhong

Head of the Laboratory for Laboratory for Gynecological Oncology, Department of Biomedicine,
University of Basel, Switzerland.
Tel: +41 61 265 9224, Fax: +41 61 265 9399, E-Mail: zhongx@uhbs.ch

Abstract

The contribution of aberrant DNA methylation in silencing of tumor suppressor genes (TSGs) and microRNAs has been investigated. Since these epigenetic alterations are reversible, it became of interest to determine the early and late effects of the 5-aza-2'-deoxycytidine (DAC) demethylation therapy in breast cancer within different molecular levels.

Here we investigate a synoptic model to predict complete DAC treatment effects at the level of epigenome, transcriptome, and proteome for several human breast cancer subtypes. The present study assessed an effective treatment dosage based on the cell viability, cytotoxicity, apoptosis and methylation quantification assays for different breast cancer cell lines. Using the optimal treatment dosage, a highly aggressive and a non-aggressive lines were investigated at various time points using omics approaches such as MALDI-TOF MS, mRNA- and microRNA expression arrays, 2-D gel electrophoresis and LC-MS-MS.

Complete molecular profiles of the studied subtypes including the biological interactions and possible early and late systematic stable or transient effects of the methylation inhibition were determined in depth. Beside the activation of several epigenetically suppressed TSGs, we also showed significant dysregulation of some important oncogenes, oncomirs, oncosuppressors miRNAs as well as drug tolerance genes/miRNAs/proteins in breast cancer related pathways.

In present study, the results give new therapeutic clues based on chemical modification of the pathological methylation patterns in breast cancer as well as investigation of some new molecular DAC targets and involved pathways. The presented approach might become a useful epigenetic treatment model for other human solid tumors in management of cancer patients.

Introduction

Aberrant DNA methylation patterns are associated with various human diseases [1] including cancer development [2]. Hypermethylation of human tumor suppressor genes (TSGs) leads to transcriptional inactivation followed by the gene silencing and carcinogenesis [3].

It was also discovered that microRNAs (miRNAs), endogenous non-coding RNAs with 19-25 nucleotides in size, play important roles in various cellular processes including cellular growth, differentiation and apoptosis [4] that contribute to cancer development and progression [5]. Moreover, emerging studies reported that miRNAs are involved in promoter DNA methylation changes [6]. Also, DNA sequences encoding miRNAs were found to be a target of aberrant DNA methylation as well as protein-coding genes [7].

Genetic changes such as mutation or deletion are resulting in permanent loss of gene expression while epigenetic changes are often reversible and could be a useful target for developing new therapeutic strategies [8,9]. Reversal hypermethylation of silenced TSGs or miRNAs is increasingly being targeted for cancer therapy and prevention [10,11]. Moreover, these approaches are particularly appealing because DNA methylation inhibitors are considerably less toxic in non-cancerous tissues compared to other anti-cancer drugs [12].

The 5-aza-2'-deoxycytidine (decitabine; DAC; Dacogen, Eisai, Inc.), has recently been approved by the Food and Drug Administration (FDA) for the treatment of patients with Myelodysplastic Syndromes (MDS) and leukemia [13,14]. Since DAC is one of the nucleotide analogs that is activated via phosphorylation by cellular deoxcytidine kinase and incorporated into the DNA, the result of this process is thought to lead to the depletion of methyltransferase activity and to demethylation of DNA [15]. Several strategies have been applied to optimize or enhance the activity of DAC in different human cancers including hematologic malignancies and solid tumors as a promising agent for cancer therapy [16].

Pan-omics approaches at multiple molecular levels after DAC treatment for solid tumors are promising in opening new mechanistic insights in this area of cancer biology. These approaches enable to synoptically probe the epigenome, transcriptome and the proteome to understand

complete phenotype of treated cells such as any possible stable or transient changes as well as finding new therapeutic targets or pathways. Therefore we used DAC in this study as a tool compound to study focuses on multi-dimensional models to predict early and late effects of DAC on different subtypes of breast cancer using pan-omics approaches.

Results and Discussion
5-aza-2'-deoxycytidine (DAC) optimal dose-range finding
In the present study, we have first identified the optimal, effective dosage of DAC by investigating the impact of concentrations in a range of 10^1 to 10^4 nM on seven different breast cell lines (six breast cancers and one control cell line) on the cell viability, cytotoxicity, apoptosis and DNA methylation.

Quantification of cell viability, cytotoxicity and apoptosis
The MultiTox-Glo Multiplex Cytotoxicity assay measured the relative number of live and dead cells in cell populations after treatment with increasing doses of DAC and could thus determine cell viability and cytotoxicity (Fig. 1a, b). Increasing the concentration of the DAC treatment (ranging from 10^1 to 10^4 nM) elevates the cytotoxicity and reduced the ratio of the cell viability in all cell lines. The results of these two measurements, cell viability and cytotoxicity, are inversely correlated as expected. The two concentrations, 10^1 and 10^2 nM DAC, showed viability higher than at EC_{50} for all studied cell lines and they revealed lower cytotoxicity compared to the other concentrations.

Two cellular caspases (caspase-3 and caspase-7) play a crucial role in apoptosis and are activated in a sequential cascade of cleavages from their inactive forms. To investigate the extent of cells apoptosis after exposure to different concentration of DAC, the enzymatic activity of caspase-3/7 was measured (Fig. 1c). These cell death markers showed a dose dependent behavior; by increasing the DAC concentration the caspase-3/7 activities in all seven studied cell lines were elevated (Fig. 1c). The mean protease activities in the cells treated with 10^3 and 10^4 nM of drug were significantly higher than the lower dosages ($P=0.001$) while the mean activities of the caspase-3/7 were similar for the both 10^1 and 10^2 nM concentrations of DAC.

Quantitative methylation profiling of six breast cancer candidate genes
Based on our previous study where we assessed the methylation status of more than 42,528 CpG sites in 22 different genes in cancerous breast tissues versus matched normal tissues [17], we selected six hypermethylated TSGs (*BMP6*, *BRCA1*, *CST6*, *CDKN2A*, *CDKN1A* and *TIMP3*) to assess the effective DAC dose range. We analyzed the methylation proportion of the six breast cancer candidate TSGs in all seven cell lines before and after treatment at increasing doses. For all studied genes one amplicon containing CpG rich islands (number of CpG sites higher than 20) was analyzed (Table 1). In total, we assessed six amplicons, containing 171 CpG sites per sample (total of 2,394 sites in all analyzed samples) (Table 1; Fig. 1d). The mean methylation quantity of the informative CpG sites per each gene was used to assess the methylation proportion of the candidate genes (Fig. 1d). DNA methylation proportion of the six studied genes were significantly decreased for all applied concentrations in the studied cell lines ($P<0.01$) (Fig. 1d). There were no significant differences on the extent of demethylation induced by DAC at concentrations of 10^2, 10^3 and 10^4 nM, whereas hypomethylation of DNA achieved with these three mentioned concentrations was significantly more than the lowest treatment dosage (10^1 nM) (Fig. 1e). This data suggests that 10^2, 10^3 and 10^4 nM dosages of DAC have more effective demethylation activity by significantly reducing the methylation proportion in the treated cells.

Based on the results obtained from studies of the cell viability, cytotoxicity, apoptosis and the DNA methylation profile, the concentration 10^2 nM of DAC was identified to be the most favorable because of the less cytotoxicity and good efficiency in demethylation of DNA. This dose of DAC has been chosen for the further study on the early and late effects of the treatment using pan-omics assays for the MDA-MB231 and SKBR3 cell lines which are considered as a highly aggressive and a non-aggressive subtypes of breast cancer, respectively [18], together with the HB2 as a control breast cell line.

MRNA expression profiling

Gene expression analysis was assessed expression microarray before treatment, directly after treatment and at five follow-up passages at "drug holiday" condition. Up- and down-regulated genes in at least in one of the analyzed passages of each cell line were considered as "union genes". The common up- and down-regulated genes in different cell lines were considered as "intersection genes". Annotated lists were created for the 991 union genes for the HB2, 431 union genes for the MDA-MB231 and 2406 union genes for the SKBR3 cell line ($P<0.05$, >2 fold change; Fig. 2-a1; Supplementary data 5). To interpret the significance of the study, first of all we focused on intersection genes within the two breast cancer cell lines and also on the intersection genes within all three studied cell lines. Besides that, expression profiles of tumor suppressor genes and oncogenes within were investigated in each cell line.

The expression profile of 21 candidate genes using qRT-PCR technically confirmed the results of the expression array results and also demonstrated the reliability and reproducibility of the assay using the appropriate gene identifiers [19] with the high stringency criteria to filter out the assay noise [20].

Intersection genes within cancer cell lines. In total 105 intersection genes were differentially expressed in the MDA-MB231and SKBR3 cell lines (Fig. 2-a5). The gene set analysis and GO enrichment scores revealed involvement of these genes in multicellular organismal process and processes that pertinent to the generation and maintenance of rhythms in the physiology of the cells (rhythmic process) (Fig. 2-a6). Pathway analyses showed that 36 out of 105 intersection genes are linked to breast neoplasms or metastasis and most of the significantly altered gene expressions in this category can potentially trigger cell proliferations (Fig. 2-a7). As an example, *CXCL3* (a small cytokine belonging to the CXC chemokine family that is also known as *GRO3* oncogene) was up-regulated after treatment in both cell lines, while *JUP* (a proliferation and oncogenesis marker) was up-regulated in the MDA-MB231 and down-regulated in the SKBR3 line (Table 2).

Intersection genes within all three cell lines. There were 78 intersection genes which are differentially expressed in the two breast cancer cell lines (MDA-MB231and SKBR3) as well as in the control cell line (HB2) (Fig. 2-a2; supplementary data 2). Based on the gene set analysis and GO enrichment scores, these genes are mostly involved in developmental, immune system process and response to stimulus (Fig. 2-a3). Pathway analysis showed that 30 out of 78 intersection genes have roles in breast neoplasms, metastasis and / or cell proliferation (Fig. 2-a4) and below we mention some of these genes.

For example, *CDKN1A* (*P21*) which plays a regulatory role in DNA replication and DNA damage repair and has strong tumor suppressor activity, was significantly up-regulated after treatment in all three cell lines. *GJA1* (*CX43*) is a TSG which is associated with suppression of metastasis [21]. This gene was up-regulated in the HB2 while down-regulated in the MDA-MB231 and SKBR3. One explanation for *GJA1* down-regulation might be strong activation of *IL8* in cancerous cell lines that has been also reported previously as a suppressor of *GJA1* [22]. *VEGFR* was over-expressed in the SKBR3 and HB2. This gene was down-regulated in the MDA-MB231. Since the *VEGFR* gene is augmenting cell proliferation and oncogenesis, it can play an oncogenic role in carcinogenesis [23]. *IL6* was very strongly expressed in all passages of the three studied cell lines as an early and late effect of DAC treatment. IL6 is a multifunctional regulator of immune and inflammatory responses, and an increase of its expression has been detected in multiple epithelial tumors including breast cancer [24]. *IL6* is involved in the regulation of cell proliferation, survival, and metabolism. Additionally, high expression levels of *IL6* are correlated with a poor clinical prognosis which can reflect its oncogenic role [25]. However, the involvement of *IL6* in cancer is still quite controversial, as dichotomous roles for *IL6* in both tumor-promoting and -suppressive activities have been reported [26]. Our data raise the question of whether combination of *IL6*-targeted therapies (*IL6* or *IL6* receptor antagonists) with epigenetic therapy would be effective in treating breast cancer patients.

Differentially expressed TSGs after DAC treatment. It is known that TSGs are hypermethylated in different tumor types as an early phenomenon during carcinogenesis [1,27]. In the present study, expression status of well-known TSGs was analyzed after treatment and follow-

up passages (supplementary data 2). Up-regulation of epigenetically silenced TSGs is considered as a positive effect of DAC treatment. Several TSGs (e.g. *CDKN1A*, *PTEN*, *CST6 BRCA1* and *RASSF1*) were differentially expressed after treatment (Table 2).

The *CDKN1A* (*P21*) gene showed significant hypomethylation and activation in all three studied cell lines. *CDKN1A* was previously reported as a hypermethylated TSG in breast cancerous tissue [17]. *PTEN* is an essential component of the *TP53* gene response upon DNA damage and it regulates p53 function through keeping Akt inactive and making Mdm2 incapable of translocation into the cell nucleus for degradation of p53 [28,29]. Recently, we showed that hypermethylation of the *PTEN* gene in breast cancer patients lacking *TP53* mutations [30]. The epigenetic silencing of lysosomal cysteine protease inhibitor cystatin 6 (*CST6*) is more frequently observed in metastatic lesions than in primary cancers [31]. Our previous study revealed *CST6* to be significantly hypermethylated in breast tumors as compared to the matched normal tissue [17]. *CST6* was significantly reactivated after treatment in MDA-MB231. The two important TSGs, *BRCA1* and *RASSF1*, which were hypermethylated in breast tumors [17] were significantly up-regulated in the SKBR3 after DAC treatment.

Differentially expressed oncogenes after DAC treatment. It has been reported that some oncogenes are epigenetically hypomethylated and activated in human cancers [1]. The expression status of well-known oncogenes was analyzed after treatment and follow-up passages (supplementary data 2). Over-expression of oncogenes might be a negative feature of treatment with DAC, while down-regulation of oncogenes could consider as positive feature of treatment. Oncogenes such as *RAB* family genes, *ETSA*, *CXCL1*, *CXCL2*, *CXCL3*, *ERBB2*, *MAFF*, *MERTK*, *MYC* and *PDGFB* were differentially expressed after treatment (Table 2). For example, *CXCL1* and *-2* are two inflammatory cytokines that have a role in cell growth, proliferation, angiogenesis and diminishing the apoptosis; additionally they have an oncogenic role in carcinogenesis. Furthermore, depletion of *CXCL1* and *-2* expressions can inhibit metastasis [32].

MicroRNA expression profiling

To seek specific miRNA profile after treatment with DAC, the expression analysis was assessed before treatment, directly after treatment and at five follow-up passages at "drug holiday" condition. Annotated lists were created of the 74, 83 and 59 miRNAs which were differentially disregulated (*P*<0.05, >2 fold change) in all of the HB2, MDA-MB231 and SKBR3 cell lines respectively (Fig. 2-b1; Supplementary data 3 & 5). For each cell line, expression profile of intersection and union miRNAs within the passages, and their biological relevance are presented in supplementary data 3. To interpret the miRNA signatures, we focused on the intersection miRNAs within two breast cancer cell lines and the intersection miRNAs within all three studied cell lines.

Intersection miRNAs within cancer cell lines. We detected eight intersection miRNAs (miR-21, -24-2, -494, -193b, -181, let-7f, let-7g and let-7i) which were differentially expressed in the two cancer cell lines (Fig. 2-b4; Table 2). All eight miRNAs have a role in breast neoplasms, metastasis and cell proliferation (Fig. 2-b5).

Intersection miRNAs within all three cell lines. In total 23 intersection miRNAs were differentially expressed in all three studied cell lines (Fig. 2-b2). Pathway analysis showed that 10 out of 23 miRNAs (miR-106b, -126, -130a, -155, 15a, -210, -27a, -27b, -342 and -378) have role in breast neoplasms, metastasis and or cell proliferation (Fig. 2-b3).

Differentially expressed oncosuppressor and oncomirs after DAC treatment. Numerous reports demonstrated that miRNAs contribute to the development and progression of cancer by acting as oncogenes or tumor suppressor genes [4,33]. Several oncosuppressor miRNAs showed significant up- and down-regulation after DAC treatment (Table 2). As an example of up-regulated oncosuppressor miRNAs; miR-155 was over-expressed during the scheduled follow up of the treatment in the both cancer cell lines while down-regulation of this miRNA has been reported in breast cancer and the TSG role was suggested for miR-155 due to decreasing cell proliferation and triggering the apoptosis [34].

As down-regulated oncosuppressor miRNAs after DAC treatment, we can mention the lower-expression of two members of let-7 family, let-7f and let-7g, in both cancerous cell lines after DAC treatment. The expression of several members of the let-7 family including let-7f and let-7g, was previously reported to be down-regulated in breast cancer samples with either lymph node metastasis or distance metastases [34]. In addition, MiR-27b also showed down-regulation after treatment in all three analyzed cell lines. MiR-27b is considered as a regulator of *CYPB1* and acts as tumor suppressor that is suppressed in breast cancer [35].

As oncomirs, different miRNAs were differentially expressed after treatment (Table 2). From this group miR-21 has been reported as a potential oncogene which promotes cell survival and proliferation [36]. In breast cancer, miR-21 is overexpressed [34] and by targeting multiple tumor suppressor genes, such as *PTEN*, *PDCD4*, *TPM1*, and *MASPIN* has been implicated in the acquisition of invasiveness and in promoting tumor metastatic properties in breast cancer [37]. Moreover, miR-21 overexpression has been associated with an advanced clinical stage and lymph node metastasis in human breast cancer [38]. Interestingly, miR-21 was significantly down-regulated in both MDA-MB231 and SKBR3 ($P<0.001$) after treatment that indicates the anti-cancer effect of the treatment. As another important example, miR-27a plays an important role in breast cancer by suppressing the expression of the transcription factors *ZBTB10/RINZF*, and subsequently increasing several angiogenic molecules, such as Survivin, *VEGF* and *VEGFR1* [39]. A previous report indicated over-expression of the miR-27a in breast cancer [40]. Our results demonstrated that DAC treatment can significantly suppress miR-27a expression.

Protein expression profiling

Proteins from the three selected cell lines were analyzed by 2D gel electrophoresis. We could detect a total of 27 intersection significantly up- and down-regulated proteins (4 in HB2, 8 in MDA-MB231 and 15 in SKBR3 cell lines) (Fig. 3A; Supplementary data 4). Principal component analysis (PCA) showed similarity between different treatment passages and in a dendrograms, differentially expressed proteins clustered in different groups based on their expressional profile (Fig. 3A). In the control breast cell line (HB2), the number of differentially expressed proteins was lower than in the other two cancerous cell lines. The dendrogram of HB2 revealed two clusters and the two clusters revealed approximately the same expression level of the untreated cells. In the MDA-MB231 (a highly aggressive cancerous line), proteins whose expression changed significantly, were categorized into two clusters. One cluster showed fluctuating up-regulation along the passages while the other returned to the level of untreated cells. Protein expression profiles in the SKBR3 (non- aggressive cancerous line) were divided into three clusters. Interestingly, all three clusters ended with over-expression and are attributable to late effect of treatment. The protein expression profiles in two cancerous cell lines are related to long term effects of the treatment and could help to better define treatment intervals. The fold change, *P* value, theoretical pI and spot volumes are summarized in supplementary data 4.

In total, 18 candidate spots were excised and submitted to LC-MS-MS analyses for protein identification. The identified proteins, gene ID, theoretical molecular weights (kDa), number of identified peptides and the percentage of amino acid coverage are shown in Table 3. In some excised spots we identified a single protein, in some others multiple isoforms, and yet in others we observed several proteins in a single spot. In all instances where the probability of correct determination was close to 100% (as determined by Scaffold) we have checked the observed molecular mass and pI value (according to spot position on the gel) with the MS identification. We thoroughly checked peptides that might have been due to neighboring contamination (needle transfer). In situations where peptide detection leads to an unambiguous protein identification and spurious a peptide was detected, the presence of the spurious peptide(s) was disregarded. This is highly speculative and doesn't bring you any further: either you identify something properly or not! I'd leave that away.

Notably, in the spot 1543 (SKBR3 cell line) we could identify four different isoforms of the 14-3-3 protein which were confirmed with two different replicates referring to the same spot; their calculated molecular masses or pI values deviated from the observed ones (and also from the main

set of polypeptides). This highly conserved protein family is a large family of 25–30kDa acidic proteins and has seven homologous isoforms in mammalian cells which are involved in a variety of biological interactions, cell-cycle progression, apoptosis, and mitogenic signaling, such as the ATM-*p53* pathway, suggesting that they may play a role in tumorigenesis [41]. 14-3-3 theta is an adapter protein implicated in the regulation of both general and specialized signaling pathways by binding to a large number of partners, usually by recognition of a phosphoserine or phosphothreonine motif.

Interestingly, five differentially expressed genes detected by transcriptomic analysis were also identified by mass spectrometry at the proteome level, which indicated similar expression trends at both, the transcriptome and proteome (Supplementary Data 4). The results implicated that two of these five proteins (Sod2 and Pebp1) were detected in one spot (SKBR3; 1486). These data support the mass spectrometry results for those cases where more than one protein species were detected within one spot (Table 3).

Following the protein identification, the molecular function, biological processes, and subcellular localization was categorized using the PANTHER gene ontology database (http://pantherdb.org) and ResNet® 7 (Mammal) database. The analyzed data showed that 28 out of 41 detected proteins/ isoforms in the three analyzed cell lines are linked to neoplasms, metastasis or carcinogenesis (supplementary data 4).

Integrative, pan-omics

Possible crosstalk between all identified union genes, miRNAs and proteins with a reported role to carcinogenesis, was investigated in this study. In total 11 genes, 11 miRNAs and 10 proteins showed interesting regulatory interactions (Fig. 3B).

Dysregulation of metastasis related genes/miRNAs. In order to screen the direct effect of the DAC treatment on tumor metastasis, we analyzed the expression profile of *PDCD4*, *DHFR*, and *HOXD10*, three of the most promising genes implicated in the metastatic process [40,42]. Among these metastasis related TSGs, only *PDCD4* was altered by the scheduled treatment and was significantly up-regulated in the highly aggressive cell line (MDA-MB231) and significantly down-regulated in the non-aggressive cell line (SKBR3) (Fig. 4A). The two miRNAs (miR-21 and miR-183), have complementary sites for the 3'-UTR of *PDCD4* and are considered to be possible suppressors of this gene [42,43]. Interestingly, the expression analysis revealed down-regulation of miR-21 in the MDA-MB231, which was inversely correlated to *PDCD4* in the same cell line ($P<0.001$) (Fig. 4-a1). This finding is in line with previous reports suggesting miR-21 as a potential down-regulator element for the *PDCD4* gene [42,43]. However, miR-183 did not show any significant correlation with the expression of *PDCD4* in either of the two breast cancer cell lines (Fig. 4A), which is in contradiction with previous reports [44]. This suggests that miR-21 (rather than the miR-183) is responsible for regulation of *PDCD4*.

Furthermore, *CXCR4* as one of 19 known human G protein–coupled chemokine receptors is specifically implicated in cancer metastasis and HIV-1 infection [45]. *CXCR4* has been associated with more than 23 cancer types promoting metastasis, angiogenesis, tumor growth and poor prognosis of patients. Over-expression of *CXCR4* has also been shown in ~10% of the tumors [46]. There is the possibility of using *CXCR4* as a therapeutic target against solid tumors [47]. Interestingly, *CXCR4* was significantly down-regulated in all scheduled follow-ups after demethylation treatment in the MDA-MB231 cell line (Fig. 4-a1). This finding suggests that down-regulation of *CXCR4* can be a secondary response to the demethylation of other upstream regulatory elements in the aggressive form of breast cancer.

Taken together, over-expression of *PDCD4*, inhibitor of cell proliferation and invasion through increasing the apoptosis, and conversely down-regulation of *CXCR4*, activator of cell proliferation and invasion, after treatment with DAC suggest reduction of the invasion and metastasis for the highly aggressive subtype of breast cancer.

Suppression of the ERBB2/HER2 receptor in the non-aggressive breast cancer subtype after DAC treatment. The estrogen receptor (ER), progesterone receptor (PR) and ERBB2/HER2 markers are used as prognosis indicators in breast cancer to stratify patients for appropriately targeted therapies [48]. After DAC treatment there were no significant changes in hormonal receptor expression status (*ESR1*, *ESR2*, *PGR* and *ERBB2*) for the HB2 (breast epithelial cell line) and MDA-MB231 (triple hormonal negative cell line). In the SKBR3 (ERBB2/HER2 positive cell line), *ERBB2* showed a steady down-regulation after treatment (Fig. 4B). The *ERBB2* is a proto-oncogene that encodes a member of the epidermal growth factor (EGF) receptor family of tyrosine kinases. Amplification and/or over-expression of this gene have been reported in numerous cancers, including breast and ovarian tumors [49]. To understand the molecular mechanism of *ERBB2* down-regulation after demethylation treatment with DAC, we analyzed the expression profile of negative upstream regulators (Fig. 4-b1). From 21 candidates, two genes (*BRCA1* and *MYC*) and one miRNA (miR-125b) showed significant up-regulation after the treatment. From these three negative regulators, expression of two (*MYC* and miR-125b) were inversely correlated with *ERBB2* expression (*P*<0.001) (Fig. 4-b2). MiR-125b, which was previously reported to be down-regulated in breast cancer [34], is involved in the regulation of *ERBB2* and *ERBB3*, two important tyrosine kinase receptors, impairing the downstream signaling pathway and the ability of the cells to grow and invade [50]. MiR-125b was significantly over-expressed as early and late effects only in the SKBR3 cells (Table 2). Interestingly, in this regard we could identify a positive upstream regulator protein (Pdia6 protein) by mass spectrometry, which presented a significant correlation with *ERBB2* down-regulation (*P*<0.05) (Fig. 4-b3). Pdia6 has activator role for *ERBB2* can contribute to cell proliferation and breast neoplasm [51].

Deregulation of drug resistance related genes/miRNAs. The relatively rapid acquisition of resistance to cancer drugs remains a key obstacle for a successful cancer therapy [52]. Cancer-initiating cells have been proposed as potential culprits because of their capacity to escape from the drug effects by becoming quiescent [53]. To get insights into the underlying mechanisms of drug resistance against DAC, we checked the expression status of more than 40 candidate genes suggested by the pathway analysis that are involved in a variety of resistance mechanisms. From these candidate genes/proteins, 12 genes detected by differential transcriptomics (*IL6*, *TGFB1*, *VEGFA*, *SERPINB5*, *FGF2*, *SFN*, *ERBB2*, *RAD51*, *CSF2*, *COL18A1*, *MDK* and *TNFSF10*) and four proteins detected proteomically (P4hb, Sod2, Arhgdia and Glo1) were mostly up-regulated after DAC treatment (Table 2; Fig. 5). At the protein expression level, significant alterations over the scheduled follow up were found for the P4hb protein in the HB2 cells, Sod2 and Arhgdia proteins in the MDA-MB231 cells, and for the Glo1 protein in SKBR3 cells (Fig. 5).

The miRNAs expression alteration in response to chemotherapy was previously studied and showed that Let-7i, miR-181a, -221, -27b, -34a, -424, -638 and -768 were up-regulated and miR-17, -21 and -28 were down-regulated after exposure to several drugs [54]. Our analysis of the genome-wide effects of DAC on cellular miRNAs implicated significant down-regulation of miR-28 in all three cell lines, miR-21 in both cancerous cell lines and miR-17 in the non-aggressive cell line (Table 2; supplementary data 5). At least three miRNA families were up-regulated: MiR-638 in all three cell lines, miR-34a in the highly aggressive line, miR-424 in non-aggressive line, in a concordance with previous reports [54]. Other miRNAs were down-regulated upon treatment with DAC in all cell lines (miR-181a and -27b). In some instances Let-7i and miR-221 presented different responses than reported [54]. In conclusion, most of the miRNAs with a role in drug resistance were down-regulated after treatment with DAC (Table 2) which may suggest less resistance controlling behaviors of miRNAs after treatment with this drug.

Prominent regulatory role of miR-24 on the methylated *P16-INK4A* gene in the non-aggressive breast cancer subtype. *P16-INK4A* acts as an inhibitor of CDK4 kinase and in cooperation with *TP53* has regulatory role for cell cycle G1 control. This gene is an important tumor suppressor gene which is frequently silenced (through hypermethylation of specific CpG islands in the promoter, mutation, homozygote deletion or other epigenetic regulators) in many nonendocrine tumors [55], and these alterations may be predictive of recurrence, tumor growth, or aggressiveness [27]. Our previous study showed significant promoter hypermethylation of

CDKN2A (*P16-INK4A*) gene in the breast cancer tissues as well as in the circulating cell free DNA (cff DNA) of patients [17,56]. In the present study, *P16-INK4A* was not transcriptionally active and after treatment with DAC showed no significant change in the expression level in the both cancer cell lines (Fig. 6-a1 & 2). It is reported that *P16-INK4A* has homozygote deletion (c.1_471 del 471) in the MDA-MB231 cell line [57]. Our data from full gene sequencing also confirmed homozygote deletion in this cell line (data not shown).

The SKBR3 cell line showed significant reduction in the methylation proportion of the *P16-INK4A* promoter (*P<0.05*) directly after treatment with DAC (Fig. 6-a3). In spite of *CDKN2A* hypomethylation induced by DAC, there wasn't any change in the expression level of the gene (Fig. 6-a2). In order to screen the possible mutation in the *P16-INK4A*, we performed full gene sequencing on the promoter region, 5'UTR, three exons including exon/intron boundaries (exon 1-3) and 3'UTR. The sequencing result showed two different mutations/polymorphisms, a heterozygote transition mutation in the 5'UTR (c. 81 G>A) and a homozygote transversion mutation in the 3'UTR (c.771 G>C) of the gene (Fig. 6-a4). These two mutations/polymorphisms are located in the UTR part of the gene that may not be the cause of lacking *P16-INK4A* expression in the SKBR3.

Further investigation demonstrated steady up-regulation of miR-24 in the SKBR3 cell line after treatment and at all follow-up passages (Fig. 6-a6). MiR-24 has intriguing complementarities to the 3'-UTRs and controlling region (CR) of the *P16-INK4A* and suppresses the gene (Fig. 6-a5) [58]. The two found mutations/polymorphisms are not located in the complementary site of the miR-24 therefore they cannot change the binding affinity of the miR-24 to the gene (Fig. 6-a7). The expression of miR-24 showed significant inverse correlation to *P16-INK4A* which might provide an explanation for P16-*INK4A* shut down in the non-aggressive cell line. These findings highlight the specific role of the epigenetic regulation of TSGs during carcinogenesis.

Inverse correlation of up-regulated miR-29b as methylation suppressor with expression *of DNMT3A* in the non-aggressive cell line. The miRNA (miR)-29 family (29a, 29b, and 29c) has intriguing complementarities to the 3'-UTRs of DNA methyltransferase (*DNMT*)*3A* and *-3B* (de novo methyltransferases), two key enzymes involved in DNA methylation, that are frequently up-regulated in solid tumors and associated with poor prognosis. In lung squamous cell carcinomas, elevated Dnmt1 expression has been reported as a poor prognosis factor. Moreover, elevated expressions of both Dnmt1 and Dnmt3B have been shown to be correlated with hypermethylation of TSG promoters [59]. It was shown that expression of miR-29s, especially miR-29b, is directly targeting both *DNMT3A* and *-3B* and is inversely correlated to their expression in lung cancer tissues [60,61]. In the present study, after demethylation treatment, the expression of miR-29b was inversely correlated with *DNMT3A* expression in the non-aggressive form of breast cancer (SKBR3) and down-regulated expression levels of *DNMT3A* (Fig. 6B). Additionally, in the SKBR3 cell line with up-regulation of miR-29b and steady down-regulation of *DNMT3A* represented an almost 10 fold increase of number of the differentially expressed genes after demethylation treatment compared to the MDA-MB231 cell line (Fig. 6-b5 vs. Fig. 6-b3) was found. The present data highlights a more prominent role of Dnmt3A in the hypermethylation of promoters in the studied breast cancer subtypes than Dnmt1 and Dnmt3B. This finding also suggests that a combination of chemical demethylation treatment with *DNMT* inhibitors (decitabine or azacitidine) with enforced expression of miR-29b in breast cancer might have a synergistic hypomethylation effects that may turn to give a better disease response in breast cancer along with the more robust gene re-expression especially for the epigenetically silenced TSGs.

To understand the molecular mechanism of miR-29b deregulation, we analyzed the expression profiles of two important upstream suppressors (miR-152 and miR-455) [18]. Both of the miR-29b suppressors, miR-152 and miR-455, were over-expressed in the MDA-MB231 cell line that might have caused miR-29b down-regulation in this cell line. In the SKBR3 cell line along with the over-expression of miR-29b, miR-455 was also up-regulated whereas miR-152 did not show any significant changes. The present data suggest a prominent inhibitory effect of miR-152 rather than miR-455 on the regulation of miR-29b emphasizing on the various controlling behaviors of miRNAs in certain types of cancer.

Conclusions

The aberrant DNA methylations of the genes/miRNAs which are involved in malignant phenotypes provide interesting targets for chemotherapeutic intervention. To the best of our knowledge, the study presented here is the first pan-omics approach to synoptically identify epigenetic, transcriptom and proteome-wide alterations after effective demethylation treatment with DAC in the context of breast cancer subtypes.

The results of our synoptic pan-omics analysis (from a single sample aliquot to reduce heterogeneity between samples), provide a comprehensive view of DAC treatment defined by stable or transient early and late systematic effects in the model of cancer or non-cancer specific changes. We investigated DAC targets in depth (e.g. genes, miRNAs and proteins) and identified new involved pathways as well as biological interactions in between the molecular levels that may correlate with particular steps in breast neoplasm including cell proliferation, cell/tissue invasion, oncogenesis, angiogenesis, apoptosis, neoplasm metastasis and senescence.

Additionally, in the presented study, beside the activation of several epigenetically suppressed TSGs, we also showed significant down-regulation of some miRNAs with oncogenic functions in breast cancer cell lines (e.g. miR-21) as well as over-expression of some miRNAs with tumor suppressor functions (e.g. miR-155) that highlights the potential of a miRNA-based therapy in breast cancer. The presented approach might become a useful model for other human solid tumor malignancies, alone or in combination with other treatments such as enforced targeted therapies for miR-29b and / or IL6 (IL6 or IL6 receptor antagonists).

Achieving a successful therapy with DAC requires the application of repetitive optimal-dose cycles with regular intervals that may allow efficient incorporation of the drug into the newly synthesized DNA undergoing mitosis during each treatment. The obtained results provide a rationale for developing therapeutic strategies based on reactivation of epigenetically silenced genes/miRNAs in breast cancer.

Materials and Methods

An optimal treatment dose screening for DAC was done based on viability, toxicity, apoptosis and methylation alterations for six breast cancer cell lines (MDA-MB-231[62], MCF-7[63], HS578T[64], BT549[65], T47D[66] and SKBR3[67]) and a breast epithelial cell line (HB2)[68] as control.

Three selected cell lines (HB2, MDA-MB231 and SKBR3) were treated with an optimal dose of DAC (10^2 nM, ~ 5 days), then cells were cultured up to 10 passages at "drug holiday" condition. Simultaneous extraction of DNA, RNA, miRNA and proteins was performed according to a previously published protocols [69]. 3-dimentional follow-up including gene expression, microRNA expression and proteomics analysis was assessed before treatment, after treatment and at five point scheduled follow-up (1st, 3rd, 5th, 7th and 10th passages). Microarray expression profiling of mRNAs and miRNAs were conducted using the Human Genome 133 Plus 2.0 GeneChips and Affymetrix GeneChipR miRNA array v1.0 respectively using protocols recommended by the manufacturer (Affymetrix). All analysis was based on expression values generated by using the Robust Multi-array Analysis (RMA) condensing methods. The differentially expressed mRNAs and miRNAs were defined using Partek Genomics Suite software. All mRNA microarray data compiled for this study is made publicly available on GEO (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE28968 and miRNA microarray data under accession number GSE28969. To validate the microarray findings, quantitative real-time (qRT) PCR was performed for 28 candidate genes and 15 candidate miRNAs. The proteomic profile was investigated using two-dimensional gel electrophoresis (2DE) based on the previously published methods [69] and analyzed with Progenesis SameSpot software. The protein spots of interest were excised from the gels and analyzed by capillary liquid chromatography tandem MS (LC-MS-MS). Protein Identification was done using TurboSequest software [70]. Gene networks and canonical pathways including protein-protein interactions as well as genes-miRNAs interactions were identified using the Pathway Studio® software and ResNet® (Mammal) database). The complete materials and method is presented in Supplementary Data 1.

Additional Information

The complete materials and method is available in supplementary data 1. The complete gene expression profiles of three cell lines are presented in supplementary data 2. The complete microRNA expression profiles of three cell lines are summarized in supplementary data 3. The complete proteomic profiles of three cell lines are shown in supplementary data 4.

Conflicts of Interest

The authors declare no conflict of interest in the present study.

Reference

1. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.
2. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.
3. Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253-266.
4. Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 6: 259-269.
5. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.
6. Fabbri M, Ivan M, Cimmino A, Negrini M, Calin GA (2007) Regulatory mechanisms of microRNAs involvement in cancer. Expert Opin Biol Ther 7: 1009-1019.
7. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, et al. (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 67: 1424-1429.
8. Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, et al. (2008) Transient cyclical methylation of promoter DNA. Nature 452: 112-115.
9. Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, et al. (2008) Cyclical DNA methylation of a transcriptionally active promoter. Nature 452: 45-50.
10. Moggs JG, Goodman JI, Trosko JE, Roberts RA (2004) Epigenetics and cancer: implications for drug discovery and safety assessment. Toxicol Appl Pharmacol 196: 422-430.
11. Yoo CB, Jones PA (2006) Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov 5: 37-50.
12. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.
13. Kantarjian H, Issa JP, Rosenfeld CS, Bennett JM, Albitar M, et al. (2006) Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer 106: 1794-1803.
14. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, et al. (2009) Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol 10: 223-232.
15. Issa JP (2003) Decitabine. Curr Opin Oncol 15: 446-451.
16. Momparler RL (2005) Epigenetic therapy of cancer with 5-aza-2'-deoxycytidine (decitabine). Semin Oncol 32: 443-451.
17. Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, et al. (2009) Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28: 2969-2978.
18. Savarese F, Davila A, Nechanitzky R, De La Rosa-Velazquez I, Pereira CF, et al. (2009) Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev 23: 2625-2638.
19. Dai M, Wang P, Boyd AD, Kostov G, Athey B, et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33: e175.
20. Shi L, Campbell G, Jones WD, Campagne F, Wen Z, et al. (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol 28: 827-838.
21. Plante I, Stewart MK, Barr K, Allan AL, Laird DW (2010) Cx43 suppresses mammary tumor metastasis to the lung in a Cx43 mutant mouse model of human disease. Oncogene.

22. Fujita T, Ashikaga A, Shiba H, Uchida Y, Hirono C, et al. (2006) Regulation of IL-8 by Irsogladine maleate is involved in abolishment of Actinobacillus actinomycetemcomitans-induced reduction of gap-junctional intercellular communication. Cytokine 34: 271-277.

23. Li M, Qi SY, Wang Y, Feng SX, Zhang BZ, et al. (2005) Expression and clinical significance of vascular endothelial growth factor, cyclooxygenase-2, and Bcl-2 in borderline ovarian tumors. Arch Gynecol Obstet 272: 48-52.

24. Kishimoto T (2005) Interleukin-6: from basic science to medicine--40 years in immunology. Annu Rev Immunol 23: 1-21.

25. Hong DS, Angelo LS, Kurzrock R (2007) Interleukin-6 and its receptor in cancer: implications for Translational Therapeutics. Cancer 110: 1911-1928.

26. Schafer ZT, Brugge JS (2007) IL-6 involvement in epithelial cancers. J Clin Invest 117: 3660-3663.

27. Radpour R, Barekati Z, Kohler C, Holzgreve W, Zhong XY (2009) New trends in molecular biomarker discovery for breast cancer. Genet Test Mol Biomarkers 13: 565-571.

28. Mayo LD, Dixon JE, Durden DL, Tonks NK, Donner DB (2002) PTEN protects p53 from Mdm2 and sensitizes cancer cells to chemotherapy. J Biol Chem 277: 5484-5489.

29. Zhou BP, Liao Y, Xia W, Zou Y, Spohn B, et al. (2001) HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. Nat Cell Biol 3: 973-982.

30. Barekati Z, Radpour R, Kohler C, Zhang B, Toniolo P, et al. (2010) Methylation profile of TP53 regulatory pathway and mtDNA alterations in breast cancer patients lacking TP53 mutations. Hum Mol Genet 19: 2936-2946.

31. Rivenbark AG, Livasy CA, Boyd CE, Keppler D, Coleman WB (2007) Methylation-dependent silencing of CST6 in primary human breast tumors and metastatic lesions. Exp Mol Pathol 83: 188-197.

32. Bachmeier BE, Mohrenz IV, Mirisola V, Schleicher E, Romeo F, et al. (2008) Curcumin downregulates the inflammatory cytokines CXCL1 and -2 in breast cancer cells via NFkappaB. Carcinogenesis 29: 779-789.

33. Croce CM, Calin GA (2005) miRNAs, cancer, and stem cell division. Cell 122: 6-7.

34. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, et al. (2005) MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65: 7065-7070.

35. Tsuchiya Y, Nakajima M, Takagi S, Taniya T, Yokoi T (2006) MicroRNA regulates the expression of human cytochrome P450 1B1. Cancer Res 66: 9090-9098.

36. Chan JA, Krichevsky AM, Kosik KS (2005) MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res 65: 6029-6033.

37. Zhu S, Wu H, Wu F, Nie D, Sheng S, et al. (2008) MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. Cell Res 18: 350-359.

38. Yan LX, Huang XF, Shao Q, Huang MY, Deng L, et al. (2008) MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA 14: 2348-2360.

39. Mertens-Talcott SU, Chintharlapalli S, Li X, Safe S (2007) The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells. Cancer Res 67: 11001-11011.

40. Ma L, Teruya-Feldstein J, Weinberg RA (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 449: 682-688. Epub 2007 Sep 2026.

41. Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. Nature 408: 307-310.

42. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, et al. (2008) Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. J Biol Chem 283: 1026-1033.

43. Lu Z, Liu M, Stribinskis V, Klinge CM, Ramos KS, et al. (2008) MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. Oncogene 27: 4373-4379.

44. Li J, Fu H, Xu C, Tie Y, Xing R, et al. (2010) miR-183 inhibits TGF-beta1-induced apoptosis by downregulation of PDCD4 expression in human hepatocellular carcinoma cells. BMC Cancer 10: 354.

45. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, et al. (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330: 1066-1071.

46. Muller A, Homey B, Soto H, Ge N, Catron D, et al. (2001) Involvement of chemokine receptors in breast cancer metastasis. Nature 410: 50-56.

47. Teicher BA, Fricker SP (2010) CXCL12 (SDF-1)/CXCR4 pathway in cancer. Clin Cancer Res 16: 2927-2931.

48. Subramaniam DS, Isaacs C (2005) Utilizing prognostic and predictive factors in breast cancer. Curr Treat Options Oncol 6: 147-159.

49. Rubin I, Yarden Y (2001) The basic biology of HER2. Ann Oncol 12 Suppl 1: S3-8.

50. Scott GK, Goga A, Bhaumik D, Berger CE, Sullivan CS, et al. (2007) Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. J Biol Chem 282: 1479-1486.

51. Gumireddy K, Sun F, Klein-Szanto AJ, Gibbins JM, Gimotty PA, et al. (2007) In vivo selection for metastasis promoting genes in the mouse. Proc Natl Acad Sci U S A 104: 6696-6701.
52. Redmond KM, Wilson TR, Johnston PG, Longley DB (2008) Resistance mechanisms to cancer chemotherapy. Front Biosci 13: 5138-5154.
53. Frank NY, Schatton T, Frank MH (2010) The therapeutic promise of the cancer stem cell concept. J Clin Invest 120: 41-50.
54. Ma J, Dong C, Ji C (2010) MicroRNA and drug resistance. Cancer Gene Ther 17: 523-531.
55. Ruas M, Peters G (1998) The p16INK4a/CDKN2A tumor suppressor and its relatives. Biochim Biophys Acta 1378: F115-177.
56. Radpour R, Barekati Z, Kohler C, Lv Q, Bürki N, et al. (2011) Hypermethylation of Tumor Suppressor Genes Involved in Critical Regulatory Pathways for Developing a Blood-Based Test in Breast Cancer. PLoS One 6: e16080.
57. Musgrove EA, Lilischkis R, Cornish AL, Lee CS, Setlur V, et al. (1995) Expression of the cyclin-dependent kinase inhibitors p16INK4, p15INK4B and p21WAF1/CIP1 in human breast cancer. Int J Cancer 63: 584-591.
58. Lal A, Kim HH, Abdelmohsen K, Kuwano Y, Pullmann R, Jr., et al. (2008) p16(INK4a) translation suppressed by miR-24. PLoS One 3: e1864.
59. Lin RK, Hsu HS, Chang JW, Chen CY, Chen JT, et al. (2007) Alteration of DNA methyltransferases contributes to 5'CpG methylation and poor prognosis in lung cancer. Lung Cancer 55: 205-213.
60. Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, et al. (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A 104: 15805-15810.
61. Garzon R, Liu S, Fabbri M, Liu Z, Heaphy CE, et al. (2009) MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. Blood 113: 6411-6418.
62. Chandrasekaran EV, Davidson EA (1979) Glycosaminoglycans of normal and malignant cultured human mammary cells. Cancer research 39: 870-880.
63. Simstein R, Burow M, Parker A, Weldon C, Beckman B (2003) Apoptosis, chemoresistance, and breast cancer: insights from the MCF-7 cell model system. Experimental biology and medicine 228: 995-1003.
64. Smith HS (1979) In vitro properties of epithelial cell lines established from human carcinomas and nonmalignant tissue. Journal of the National Cancer Institute 62: 225-230.
65. Littlewood-Evans AJ, Bilbe G, Bowler WB, Farley D, Wlodarski B, et al. (1997) The osteoclast-associated protease cathepsin K is expressed in human breast carcinoma. Cancer research 57: 5386-5390.
66. Burfeind P, Chernicky CL, Rininsland F, Ilan J (1996) Antisense RNA to the type I insulin-like growth factor receptor suppresses tumor growth and prevents invasion by rat prostate cancer cells in vivo. Proceedings of the National Academy of Sciences of the United States of America 93: 7263-7268.
67. Pollack MS, Heagney SD, Livingston PO, Fogh J (1981) HLA-A, B, C and DR alloantigen expression on forty-six cultured human tumor cell lines. Journal of the National Cancer Institute 66: 1003-1012.
68. Berdichevsky F, Alford D, D'Souza B, Taylor-Papadimitriou J (1994) Branching morphogenesis of human mammary epithelial cells in collagen gels. Journal of cell science 107 (Pt 12): 3557-3568.
69. Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, et al. (2009) Simultaneous Isolation of DNA, RNA, and Proteins for Genetic, Epigenetic, Transcriptomic, and Proteomic Analysis. J Proteome Res 8: 5264-5274.
70. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR, 3rd (2000) Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. Anal Chem 72: 757-763.

Table 1. High-throughput methylation analysis of CpG sites per amplicon for the 10 candidate genes.

Genes	Amplicon size (bp)	Total No. of CpG sites in amplicon	No. of analyzed CpG sites in amplicon	No. of analyzed CpG sites per amplicons	
				Single	Composite
BMP6	397	37	30	11	19
BRCA1	413	30	15	10	5
CST6	445	49	27	15	12
CDKN2A	580	62	36	13	23
CDKN1A	419	30	19	10	9
TIMP3	441	51	44	11	33

The *in silico* digestion was performed for the T-cleavage assay. The percentage of total CpG sites in the amplicon is divided into single sites (single CpG sites) and composite sites (two or more adjacent CpG sites fall within one fragment, or when fragment masses are overlapping).

Table 2. Differentially expressed well-known cancer related and drug tolerance related TSGs, oncogenes, miRNAs and proteins after DAC treatment in breast cancer subtypes.

3D omics analysis	Cancer related genes/miRNAs/proteins	*Up-regulated*			*Down-regulated*		
		HB2	MDA-MB231	SKBR3	HB2	MDA-MB231	SKBR3
Gene expression *	Tumor suppressor genes	CDKN1A GJA1 (CX43)	CDKN1A CST6 PDCD4	BRCA1 CDKN1A FLCN RASSF1	DLGAP5 FAT3 PTEN	GJA1 (CX43)	CADM1 CDKN1B GJA1 (CX43) ING4 STEAP3
	Oncogenes	CXCL1 CXCL2 ETS1 ETV7 LCN2 MAFF MAP3K8 PDGFB RAB33B	CXCL1 CXCL2 CXCL3 JUP LCN2 RAB31	BRAF CXCL2 CXCL3 KLF6 LCN2 MYC	RAB1S	MERTK	ERBB2 JUP VAV1
	Drug tolerance related genes	CSF2 IL6 MDK TNFSF10 VEGFA	CSF2 FGF2 IL6 SERPINB5	CSF2 IL6 MDK RAD51 SERPINB5 SFN	SERPINB5	COL18A1 VEGFA	COL18A1 ERBB2 TGFB1 TNFSF10 VEGFA
MiRNA expression **	Tumor suppressor (oncosuppressor) miRNAs	miR-126	miR-155 miR-193b	miR-155 miR-193b miR-125b	miR-27b miR-155 miR-15a miR-210	miR-27b let-7f let-7g let-7i miR-125a miR-125b miR-126 miR-15a miR-210	miR-27b let-7f let-7g let-7i miR-15a miR-210
	OncomiRs		miR-494 miR-181a miR-378	miR-494 miR-181a miR-378	miR-27a miR-106b miR-130a miR-378	miR-27a miR-21 miR-24-2 miR-106b miR-130a	miR-27a miR-21 miR-24-2 miR-106b
	Drug tolerance related miRNAs	miR-638 miR-768	miR-638 miR-34a	miR-638 miR-424 miR-768	miR-28 miR-181a miR-27b	miR-28 miR-21 miR-181a miR-27b miR-768 let-7i	miR-28 miR-21 miR-17 miR-181a miR-27b let-7i
Protein expression ***	Proteins related to cell proliferation, neoplasms and angiogenesis	Arhgdib Anxa2 Tgf	Gstm2 Lgals7 Pebp1 Sod2	Cbx1			Bat1 Pdia6 Rbbp7 Ywhab
	Proteins related to cell invasion and metastasis	P4hb	Pebp1 Sod2			Arhgdia	Pdia6 Ywhaz
	Drug tolerance related proteins	P4hb	Arhgdia Sod2	Glo1			Pdia6

The complete data include both early and late up- and down-regulations, *P* values and fold changes are presented in supplementary data. * Supplementary data 2 & 5; ** Supplementary data 3 & 5; *** Supplementary data 4.

Table 3. Detected proteins using capillary liquid chromatography tandem MS within three analyzed cell lines.

| Cell line	No. of proteins / spot	Spot no.	Detected proteins	Gene name	Gene ID	Accession No.	MW (kDa)	No. of detected peptides	aa coverage (%)	Comments
HB2	Two	1385	rho GDP-dissociation inhibitor 2	ARHGDIB	397	gi|56676393	23	4	27	
			ran-specific GTPase-activating protein	RANGAP1	5905	gi|542991	23	3	15	
	Three	850	beta-tubulin	TUBB1	81027	gi|2119276	49	5	14	Replica (a)
			protein TRK-fused gene	TFG	10342	gi|21361320	43	4	8	
			protein disulfide-isomerase	P4HB	5034	gi|20070125	57	25	46	Replica (b)
	Four	1044	F-actin-capping protein subunit alpha-1	CAPZA1	829	gi|5453597	33	7	36	Two subunits were detected in replica (a)
			F-actin-capping protein subunit alpha-2	CAPZA2	830	gi|5453599				
			eukaryotic translation initiation factor 3 subunit I	EIF3I	8668	gi|4503513	37	9	28	
			zinc-alpha-2-glycoprotein	AZGP1	563	gi|4502337	34	3	7	Replica (b)
			annexin A2 isoform 2	ANXA2	302	gi|4757756	39	3	3	
MDA-MB231	One	1536	S-phase kinase-associated protein 1	SKP1	6500	gi|25777713	19	2	10	
		1622	peroxiredoxin 5	PRDX5	25824	gi|6912238	22	4	21	
	Two	953	T-complex protein 1 subunit zeta isoform a	CCT7	10574	gi|4502643	58	10	19	Two subunits were detected in replica
			T-complex protein 1 subunit gamma isoform a	CCT3	7203	gi|63162572				
			stress-induced-phosphoprotein 1	STI1	10963	gi|5803181	6	6	12	
		1076	voltage-dependent anion channel 2	VDAC2	7417	gi|48146045	30	7	25	
			Calponin 2	CNN2	1265	gi|4758018	34	5	17	
	Four	1293	tumor protein D54	TPD52L2	7165	gi|40805860	22	8	47	
			rho GDP dissociation inhibitor (GDI)	ARHGDIA	396	gi|36038	23	6		Replica (a)
			Galectin-7	LGALS7	3963	gi|3891470	28	3		
			glutathione S-Transferase M2-3	GSTM2	2946	gi|5822511	26	5	21	Replica (b)
			manganese superoxide dismutase	SOD2	6648	gi|34707	25	5	12	
		1486	phosphatidylethanolamine-binding protein 1	PEBP1	5037	gi|4505621	21	5	30	
			regulator of G-protein signaling 10	RGS10	6001	gi|18266777	20	4	12	
			transgelin-2	TAGLN2	8407	gi|4507357	22	3	10	
SKBR3	One	1121	RCN3 reticulocalbin 3	RCN3	57333	gi|28626510	7	37	21	
		1219	retinoblastoma binding protein 7 (histone-binding protein RBBP7)	RBBP7	5931	gi|4506439	48	4	7	
		1470	ribosomal protein L7	RPL7	6129	gi|35903	29	3	17	
		1543	14-3-3 protein	YWHAQ	10971	gi|5803227	28	6	23	Four different isoforms were detected and confirmed in both replicates
				YWHAB	7529	gi|4507949				
				YWHAZ	7534	gi|4507953				
				YWHAH	7533	gi|4507951				
		1590	chromobox protein homolog 1	CBX1	10951	gi|5803076	21	2	10	
		1633	glyoxalase-I	GLO1	2739	gi|5020074	21	5	22	
		2162	canopy homolog 2 isoform 1 precursor	CNPY2	10330	gi|7657176	21	2	14	
	Two	1727	GTP-binding protein SAR1a	SAR1A	56681	gi|9910542	22	3	17	
			proteasome subunit beta type-2	PSMB2	5690	gi|4506195	23	3	19	
	Four	960	spliceosome RNA helicase BAT1	BAT1	7919	gi|4758112	49	13	32	
			26S proteasome non-ATPase regulatory subunit 5	PSMD5	5711	gi|4826952	56	7	15	
			beta-tubulin	TUBB1	81027	gi|2119276	49	5	14	
			protein disulfide-isomerase A6	PDIA6	10130	gi|5031973	48	5	15	

(aa: amino acids)
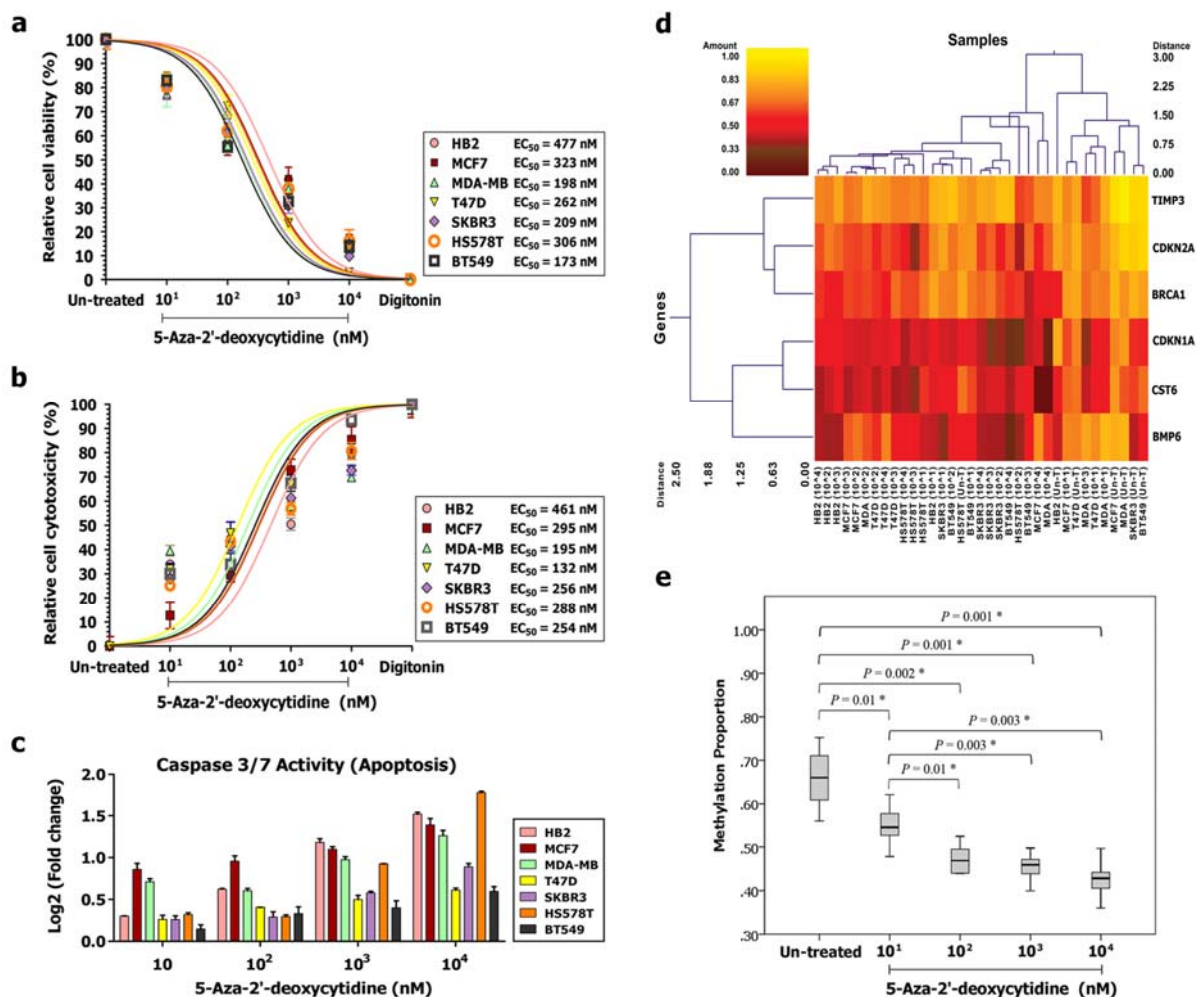
Fig. 1. Dose response screening within six breast cancer cell lines (MCF7, MDA-MB231, T47D, SKBR3, HS578T and BT549) and epithelial breast cell line (HB2) as a control. a-b) Multiplex quantification of the cell viability and cytotoxicity protease activities upon demethylation treatment with DAC (digitonin treated cells were considered as cell death control). c) Quantification of caspase-3 and caspase-7 activities as a measure of apoptosis. d-e) Methylation profiling of six candidate tumor suppressor genes (red clusters indicate 0% methylated, yellow clusters indicate 100% methylated, color gradient between red and yellow indicates methylation ranging from 0-100, and black clusters indicate not analyzed CpG sites).
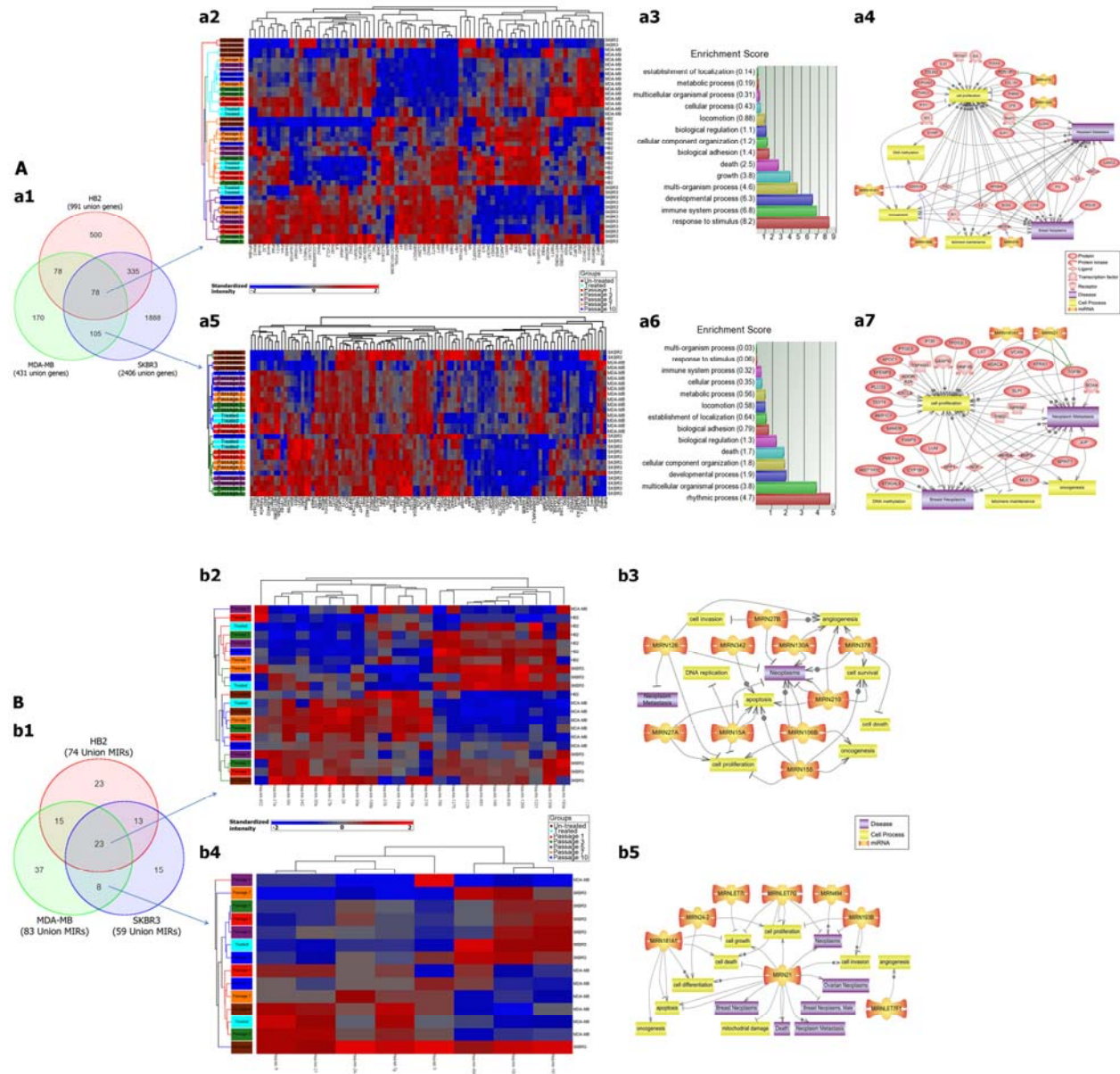
Fig. 2. A) Gene expression matrix of the HB2, MDA-MB231 and SKBR3 cell lines. a1) The number of significant up/down-regulated genes of three analyzed cell lines. a2) Expression profiles of 78 intersect genes among all three cell lines as non-cancer specific changes. a3) Gene set analysis and GO enrichment score of 78 genes (a value of three or higher corresponds to a significant over expression; *P*<0.05). a4) Pathway analysis of 30 out of 78 intersect genes that are linked to breast neoplasms, metastasis and/or cell proliferation. a5) Expression profiles of 105 intersect genes between the MDA-MB231 and SKBR3 as cancer specific changes. a6) Gene set analysis and GO enrichment score of 105 genes. a7) Pathway analysis of 36 out of 105 intersect genes that are linked to breast neoplasms, metastasis and or cell proliferation.

B) MicroRNA expression matrix of the HB2, MDA-MB231 and SKBR3 cell lines. b1) The number of significant up/down-regulated miRs of three analyzed cell lines. b2) Expression profiles of 23 intersection miRs among all three analyzed cell lines as non-cancer specific changes. b3) Pathway analysis of 10 out of 23 intersection miRs that are linked to breast neoplasms, metastasis and/or cell proliferation. b4) Expression profiles of 8 intersection miRs between the MDA-MB231 and SKBR3 as cancer specific changes. b5) Pathway analysis of 8 intersection miRs as cancer specific changes that are linked to breast neoplasms, metastasis and or cell proliferation.
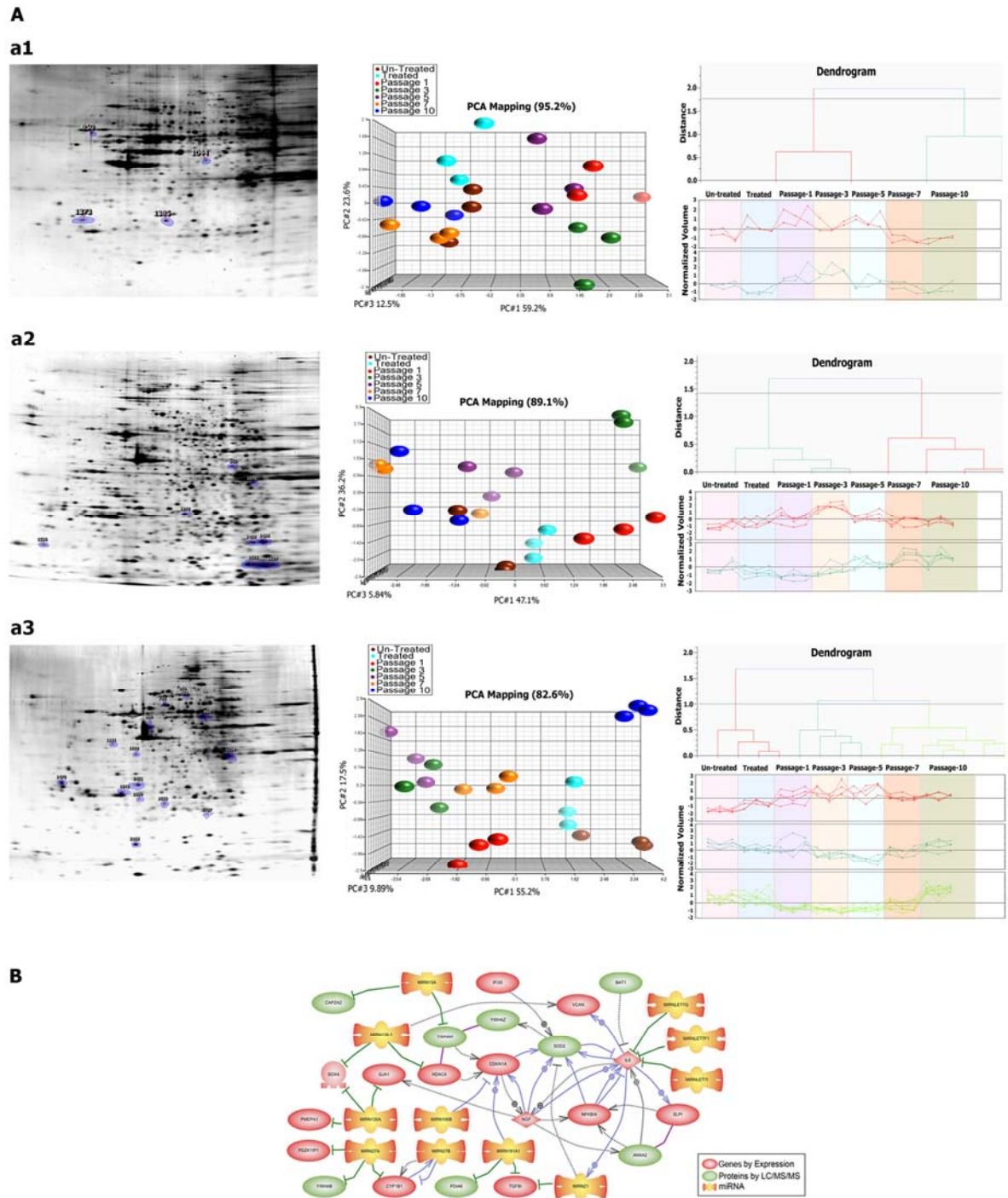
Fig. 3. A) Proteomics analysis of the three studied cell lines. (a1 – a3 left panel) Proteomic images of representative gels of the three cell lines HB2, MDA-MB231 and SKBR3 with indicated spot positions excised for mass spectrometry analysis; spot picking performed from up/down-regulated intersection proteins. (a1 – a3 middle panel) PCA analysis of intersection proteins from passages 1, 3, 5, 7 and 10 from the three cell lines. (a1 – a3 right panel) Clustering and classification of significantly up/down-regulated proteins based on their profiles throughout the passages.
B) Crosstalk between 3-dimensional Omics. Predicted interactions between intersection genes, miRNAs and proteins.
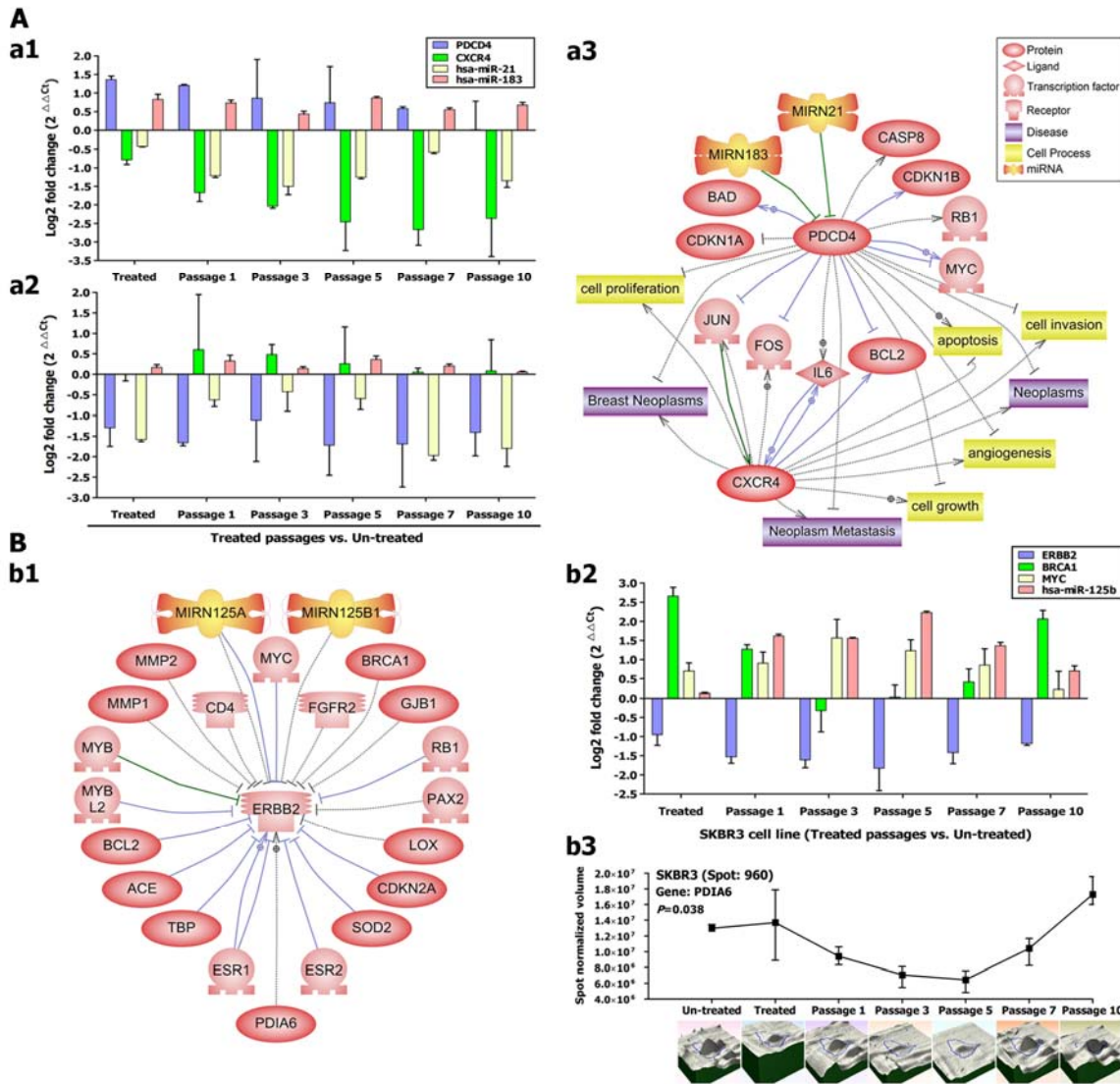
Fig. 4. A) Regulation of *PDCD4* and *CXCR4* genes as important indicators of metastasis and prognosis before and after treatment in two studied cancerous cell lines. a1-2) Expression profiles of *PDCD4* and *CXCR4* genes and regulatory microRNAs in the MDA-MB231 and SKBR3 cell lines, respectively. a3) Pathway analysis predicts the role of *PDCD4* and *CXCR4* in cancerogenesis and shows the upstream regulatory miRs. B) Mechanism of *ERBB2* receptor down-regulation in the SKBR3 (non-aggressive breast cancer cell line). b1) Pathway analysis shows negative upstream regulators of *ERBB2* gene and *PDIA6* as an example of positive regulators. b2) Expression profiles of *ERBB2* and some negative regulators (*BRCA1*, *MYC* and miR-125b) that demonstrated significant changes during treatment. b3) Expression profile of Pdia6 protein as a positive regulator of ERBB2 that is detected by proteomics analysis.
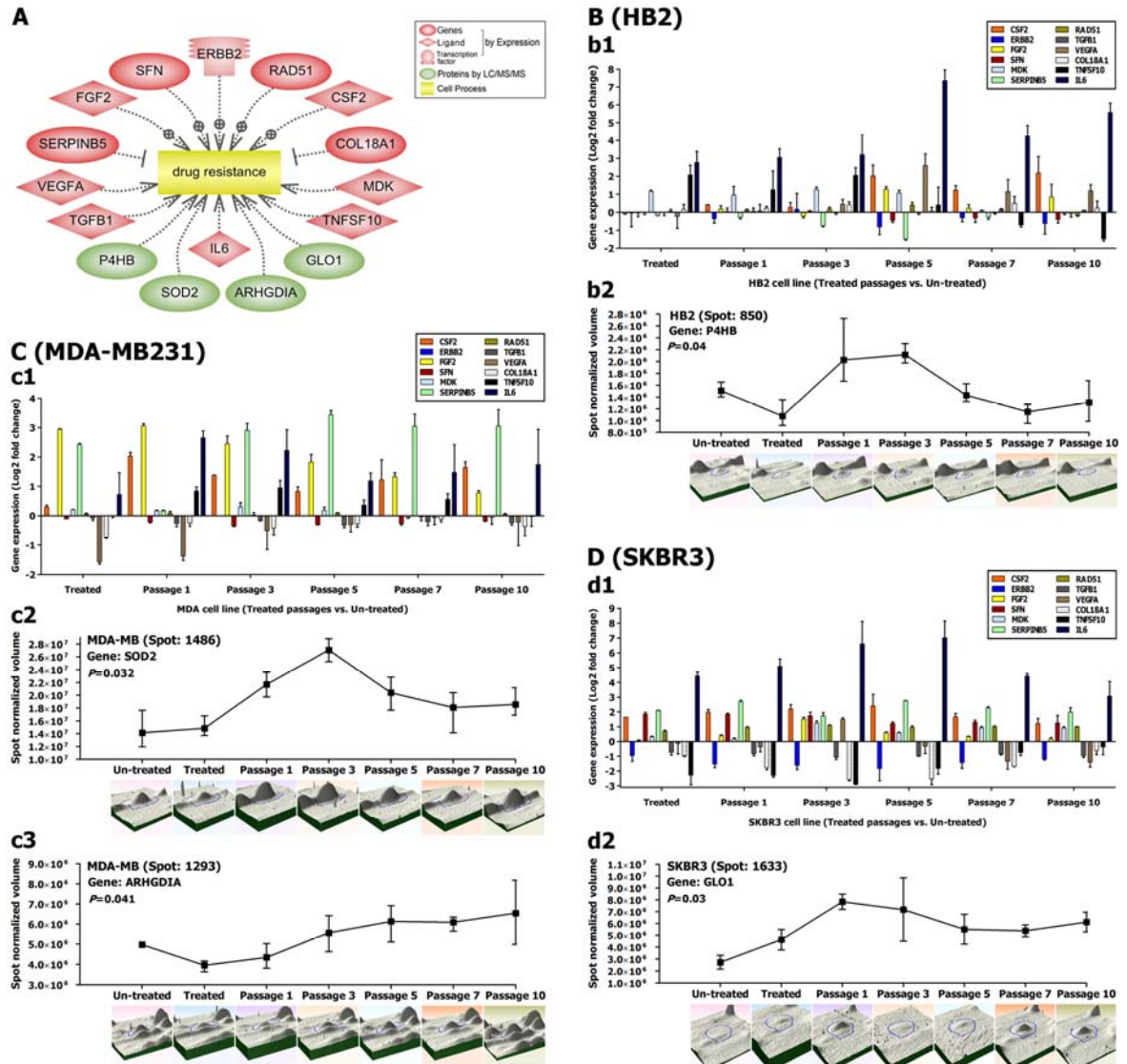
Fig. 5. Drug resistance after treatment with DAC. Candidate genes and proteins that are involved in drug resistance with significant expression changes after treatment (A). Cell lines: HB2 (B), MDA-MB231 (C) and SKBR3 (D). Expression profiles of 12 candidate genes (*TGFB1*, *VEGFA*, *SERPINB5*, *FGF2*, *SFN*, *ERBB2*, *RAD51*, *CSF2*, *COL18A1*, *MDK*, *TNFSF10* and *IL6*) (b1, c1 and d1). Expression profiles of candidate proteins: P4hb (b2), Sod2 (c2), Arhgdia (c3) and Glo1 (d2).
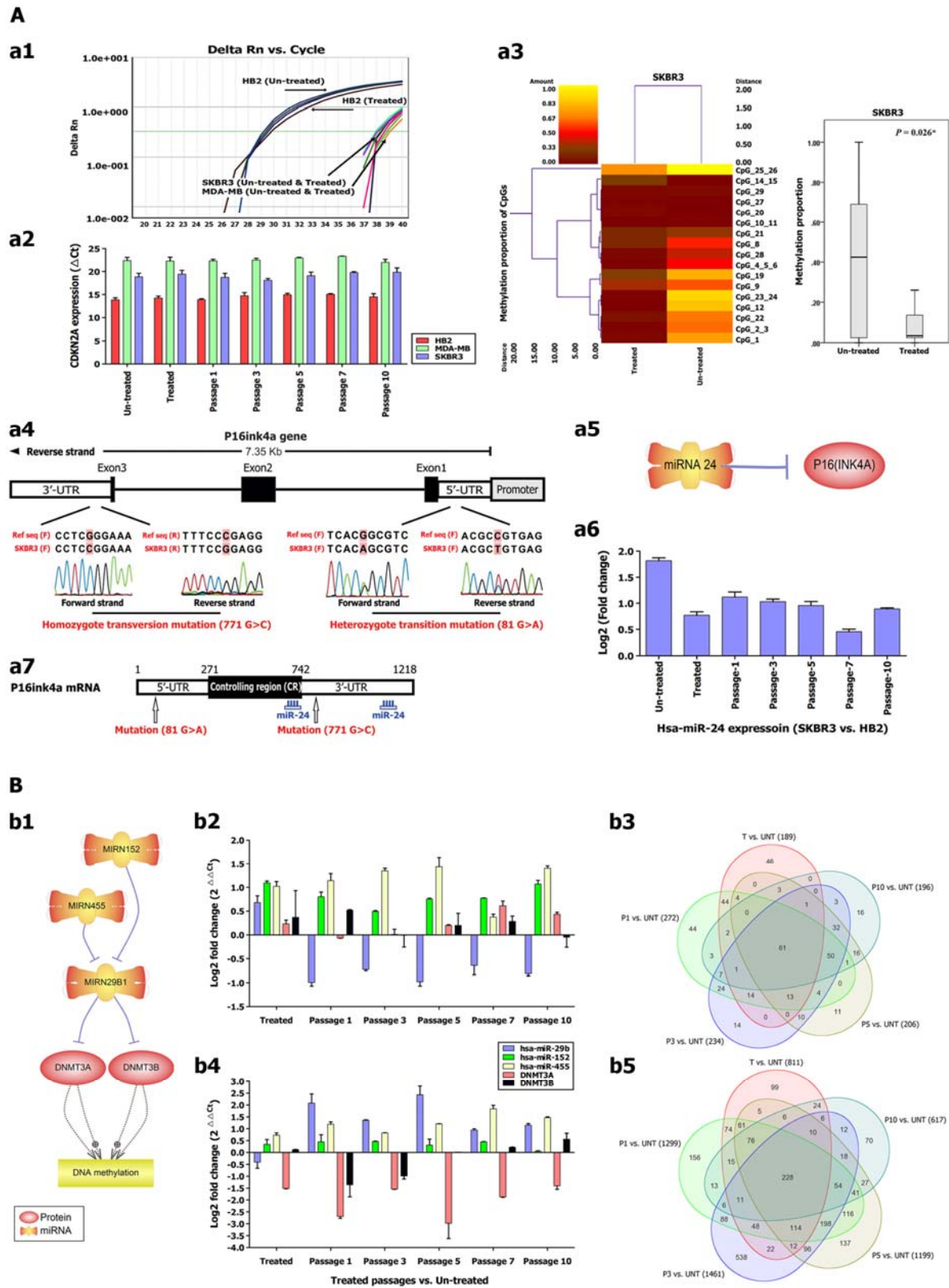
Fig. 6. A) Epigenetic regulation of *P16-INK4A* tumor suppressor gene. a1) Expression profile of *P16-INK4A* in treated and untreated cell lines; in the cancerous cell lines MDA-MB231 and SKBR3 no expression was detected. a2) Expression profile of *P16-INK4A* within all analyzed passages according to normalized Ct values (ΔCt). a3) Methylation proportion of informative CpG sites of *P16-INK4A* in the SKBR3 cell line. a4) Full gene sequencing of *P16-INK4A* revealed a heterozygote transition mutation in 5'UTR and a homozygote transversion mutation in 3'UTR of the gene. a5) Pathway analysis predicts the role of has-miR-24 (MIRN24) as down-regulator of the *P16-INK4A*. a6)

Expression profile of has-miR-24 in the non-aggressive breast cancerous cell line (SKBR3) versus control breast cell line (HB2). a7) Comparison of miR-24 recognition site with the location of found mutations in the mature *P16-INK4A* mRNA.

B) Role of has-miR-29b (MIRN29b) as methylation suppressor during chemical methylation treatment. b1) Pathway analysis predicts the inhibitory effects of MIRN29b for *DNMT3A* and *DNMT3B* genes and shows negative upstream regulators of MIRN29b. b2) Expression profiles of MIRN29b, *DNMT3A*, DNMT3B and regulatory microRNAs in the MDA-MB231 cell line. b3) The number of significant up/down-regulated genes within different passages of the MDA-MB231. b4) Expression profiles of MIRN29b, *DNMT3A*, DNMT3B and regulatory microRNAs in the SKBR3 cell line. b5) The number of significant up/down-regulated genes within different passages of the SKBR3.

Supplementary Data 1
Complete Materials and Methods

Eligibility criteria and study design

An optimal DAC concentration for the effective treatment was screened on the cell viability, toxicity, apoptosis and DNA methylation for the seven breast cell lines (six breast cancers and a breast epithelial cell line as control). Pan-omics analysis at multiple molecular levels including epigenetic, transcriptomic (mRNA and microRNA expression) and proteomic analysis was performed before treatment, directly after treatment and at five follow-up passages at "drug holiday" condition for the two selected breast cancer cell lines and the control cell line of the step (a) for up to 10 passages. The study design workflow is illustrated in figure 1.
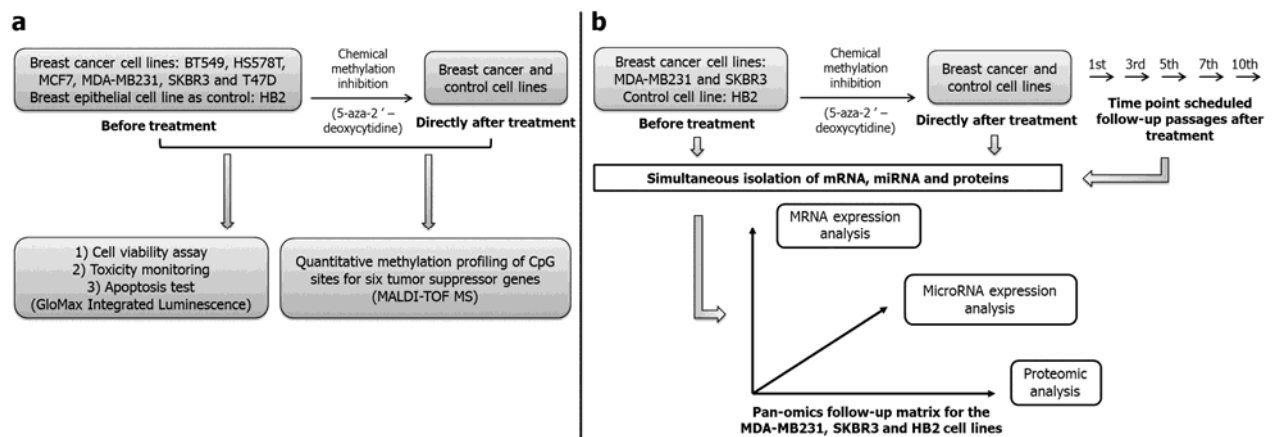
Fig. 1. The study design workflow. a) Screening for the optimal DAC treatment concentration on the cell viability, toxicity, apoptosis and DNA methylation for the six cancerous and a normal breast cell lines. b) Pan-omics follow-up (mRNA expression, microRNA expression and proteomics analysis) after treatment up to passage 10 for the two cancerous and a normal breast cell lines.

Cell lines and culture conditions

A total of six breast cancer cell lines (MDA-MB-231, MCF-7, HS578T, BT549, T47D and SKBR3) and a breast epithelial cell line (HB2) were subjected for this study. The characteristics and media conditions are summarized in the table 1.

Table 1. Source, clinical and pathological features, and culture conditions of used cell lines.

Cell line	Gene cluster	Receptor status			TP53	Source	Tumor Type	Culture media	Culture conditions
		ER	PR	ERBB2/HER2					
HB2	Lu	+	+	-	$++^{WT}$	P.Br	-	DMEM, 10% FBS, I, H	$37^{o}C, 5\% CO_2$
BT549	BaB	-	-	-	$++^{M}$	P.Br	IDC, pap	RPMI, 10% FBS	$37^{o}C, 5\% CO_2$
HS578T	BaB	-	-	-	$+^{M}$	P.Br	IDC	DMEM, 10% FBS	$37^{o}C, 5\% CO_2$
MCF7	Lu	+	+	-	$+/-^{WT}$	PE	IDC	DMEM, 10% FBS	$37^{o}C, 5\% CO_2$
MDA-MB231	BaB	-	-	-	$++^{M}$	PE	AC	DMEM, 10% FBS	$37^{o}C, 5\% CO_2$
SKBR3	Lu	-	-	+	+	PE	AC	McCoys 5A, 10% FBS	$37^{o}C, 5\% CO_2$
T47D	Lu	+	+	-	$++^{M}$	PE	IDC	RPMI, 10% FBS	$37^{o}C, 5\% CO_2$

AC, adenocarcinoma; BaB, Basal B; IDC, invasive ductal carcinoma; Lu, luminal; Pap, papillary; P.Br, primary breast; PE, pleural effusion.

ER/PR/HER2/TP53 status: ER/PR positivity, HER2 overexpression, and TP53 protein levels and mutational status (obtained from the Sanger web site; M, mutant protein; WT, wild-type protein) are indicated.

Media conditions: FBS, fetal bovine serum; I, Insulin (0.01 mg/ml); H, hydrocortisone (500 ng/ml); DMEM, Dulbecco's modified Eagle's medium, GIBCO #11965-092; RPMI, RPMI medium 1640, GIBCO #27016-021.

DAC treatments

Cells were seeded (2 X 10^5 cells / T75 flasks) for 24h then media were removed and cells were freshly exposed to the DAC (Sigma Chemical Co., Switzerland) at concentrations of 10^1, 10^2, 10^3 or 10^4 nM in suspension culture every 24h until cultured cells reached 80-85 % confluence (~5 days). The cells were split for 10 follow-up passages at "drug holiday" condition.

Multiplex quantification of cell viability and cytotoxicity protease activities

Multiplex quantification of the cell viability and cytotoxicity protease activities were assessed using the MultiTox-Glo Multiplex Cytotoxicity® Assay kit (Promega AG, Dübendorf, Switzerland) according to the manufacturer´s instructions. In this assay the two protease activities were measured (one as a marker of cell viability, and the other as a marker of cytotoxicity). The live-cell protease activity which is restricted to intact viable cells was measured using a fluorogenic, cell-permeant, peptide substrate (glycyl-phenylalanylaminofluorocoumarin, GF-AFC). This substrate enters intact cells, where it is cleaved by the live-cell protease activity to release AFC and generate a fluorescent signal that is proportional to the number of living cells. For this assay, 50 µL of the GF-AFC reagent was added to each well of a white-walled plate including ~5000 cells followed by mixing and incubation for one hour at 37^oC. The live-cell fluorescence was measured at ~$400nm_{Ex}$/~$505nm_{Em}$ on a GloMax®-Multi Microplate Multimode Reader (Promega AG, Dübendorf, Switzerland).

A second, luminogenic cell-impermeant peptide substrate (alanyl-alanyl-phenylalanyl-aminoluciferin; AAF-Glo™ Substrate) was used to measure dead-cell protease activity, which is released from cells that have lost membrane integrity. The AAF-Glo™ Substrate is not cell-permeant; therefore, essentially no signal is generated from this substrate by intact, viable cells. Briefly, 50 µL of AAF-Glo™ reagent was added to each well from viability assay followed by orbital shaking, and incubate for 15 minutes at room temperature. The dead-cell fluorescence was measured at ~400nmEx/~505nmEm. Each sample was matured in triplex format and data were collected as relative fluorescence units (RFU).

Quantification of caspase-3 and caspase-7 activities

The caspases-3 and -7 activities were measured using the Caspase-Glo®3/7 Assay kit (Promega AG, Dübendorf, Switzerland) according to the manufacturer´s instructions. This kit is based on the cleavage of the amino acid sequence DEVD of a luminogenic substrate by the caspases-3 and -7 which results in a luminescent signal. For this assay 50 µL of freshly prepared Caspase-Glo Reagent was added to each well of a white-walled plate including ~5000 cells followed by two hours incubation at room temperature. Luminescence was measured at ~$485nm_{Ex}$/~$527nm_{Em}$ on a GloMax®-Multi Microplate Multimode Reader (Promega AG, Dübendorf, Switzerland). Each sample was matured in triplex format and data were collected as relative fluorescence units (RFU).

Simultaneous isolation of DNA, RNA miRNA and proteins

The selected cell lines (MDA-MB231, SKBR3 and HB2) were subject for simultaneous DNA, RNA, miRNA and proteins isolation ($5×10^6$ cells per sample) before treatment, after treatment and at five point follow-ups (1st, 3rd, 5th, 7th and 10th passages) using AllPrep® DNA/RNA/Protein Mini Kit (QIAGEN AG, Basel, Switzerland) according to the published protocol [1]. The RNeasy MinElute Cleanup Kit (QIAGEN AG, Basel, Switzerland) was used for isolation and purification of enriched miRNAs. The quantity of extracted species was assessed using a NanoDrop ND-1000 spectrophotometer (Biolab, Mulgrave, VIC, Australia). The extracted RNA samples were analyzed for the size fractionation using RNA 6000 Nano LabChip® as well as using RNA 6000 Pico LabChip® for the enriched miRNAs (Agilent Technologies, GmbH, Waldbronn, Germany).

Methylation quantification of candidate tumor suppressor genes (TSGs) using thymidine-specific cleavage mass array on MALDI-TOF silico-chip

The EpiTYPER™ assay using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and MassCLEAVE[TM] reagent was assessed for the quantification of DNA methylation patterns of four TSGs (*BMP6*, *BRCA1*, *CST6*, *CDKN1A*, *CDKN2A*, *P16-INK4A* and *TIMP3*) according to the previously published methods [2,3,4,5].

Bisulfite treatment. To perform bisulfite conversion of the target sequence, the Epitect[®] Bisulfite Kit (QIAGEN AG, Basel, Switzerland) was used.

Primer designing and PCR-tagging for EpiTYPER[TM] assay. CpG density and CpG sites of target sequences in six TSGs were analyzed for the PCR primer design. We used previously designed and tagged primers (reverse primer with T7-promoter tag and forward primer with 10mer tag sequence as balance) for the candidate genes. The primer sequences are summarized in the table 2. Selected amplicons were mostly located in the promoter regions, or started from the promoter and partially covered the first exon [3]. For the PCR on bisulfite-treated genomic DNA (gDNA), the following PCR conditions were used: 1x: 95°C for 10 min; 48x: 95°C for 30s, Ta for 40s, 72°C for 1 min; 1x 72°C for 5 min. The PCR cocktail was: 2μL DNA (2.00μL of at least 10 ng/μL DNA for a final concentration of 2ng/μL per reaction) in a 10μL total volume using 10 pmol of each primer, 200μM dNTP, 0.2 unit Hot Start Taq DNA polymerase, 1.5mM MgCl$_2$ and the buffer supplied with the enzyme.

In vitro transcription, T-cleavage assay and Mass spectrometry. *In vitro* transcription and T-cleavage were assessed according to the previously published methods [3,5,6]. Twenty-two nanoliters of cleavage reaction were robotically dispensed (nanodispenser) onto silicon chips preloaded with matrix (SpectroCHIP; SEQUENOM, San Diego). Mass spectra were collected using a MassARRAY Compact MALDI-TOF (SEQUENOM) and spectra's methylation ratios were generated by the Epityper software v1.0 (SEQUENOM, San Diego).

Table 2. The sequence of PCR tagged primers for *in vitro* transcription.

Gene	Primer	Sequence (5'→3')	Length	T$_a$	Prodoct Size (bp)
BMP6	tag-EN1-FW	AGGAAGAGAGGGGGTAAATTTTATGGTGGTTT	22+10	57	397
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**CCTTCCTAACCCTCAATCCTTA	22+31		
BRCA1	tag-EN1-FW	AGGAAGAGAGAATTGGAGATTTTTATTAGG	20+10	56	413
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**AAATCTCAACRAACTCAC	18+31		
CST6	tag-EN1-FW	AGGAAGAGAGGTTGGTAGTTTATTTTGGATAGTTT	25+10	59	445
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**CAAATCCCRAAATTCTCC	18+31		
CDKN1A	tag-EN1-FW	AGGAAGAGAGGGTAAATTTTTGTTTGTTAGAGTGG	25+10	60	419
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**TAACTTCRACAACTACTCACACCT	24+31		
CDKN2A	tag-EN1-FW	AGGAAGAGAGGGTTGTTTTTGGTAGGG	17+10	58	580
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**ATATAAACCACRAAAACCC	19+31		
P16-INK4A	tag-EN1-FW	AGGAAGAGAGTGGGGTTTTTATAATTAGGAAAGAATA	27+10	57	533
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**CCACCCTCTAATAACCAACCAA	22+31		
TIMP3	tag-EN1-FW	AGGAAGAGAGTTTTGTTATTGGTTTGAGGG	20+10	59	441
	T7-EN1-RV	CAGT**AATACGACTCACTATAGGG**AGA**AGGCT**CCAAACTCCAACTACCCA	18+31		

Microarray analysis and qRT-PCR validations

MRNA expression analysis. Microarray analyses of the mRNA samples at various follow-up passages including two independent biological replicates were conducted using the Affymetrix Genome 133 Plus 2.0 GeneChips (Affymetrix Inc., Santa Clara, U.S.). The synthesis of cDNA was carried out with a starting amount of approximately 50 ng total RNA using the NuGEN Ovation RNA Amplification System V2 (NuGEN Technologies Inc.; San Carlos, U.S.). The arrays were hybridized with the biotinlabeled fragments and rotated in the hybridization ovens for 18 hours at 45°C and 60 rpm (miRNA arrays rotated 16 hours at 48°C. The arrays were washed and stained with a streptavidin phycoerythrin conjugate on GeneChip Fluidics 450 Workstations, and scanned on a GeneChip Scanner 3000 7G (Affymetrix Inc.) The expression data were acquired using the Affymetrix GeneChip Operating Software (GCOS). The systems were used to generate the numerical values of the probe intensity (Signal). The HGU133Plus2 annotations (v2.1.0, ENTREZG) supplied by University of Michigan was used as probe identifier (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/12.1.0/entrezg.asp).

MicroRNA expression analysis. Microarray analyses of the enriched miRNAs samples at various follow-up passages were carried out using Affymetrix GeneChipR miRNA array (v1.0). The labeling of the microRNAs was carried out with a starting amount of approximately 100 ng small RNA using the FlashTag Biotin HSR RNA Labeling Kit (Genisphere). The labeling process was performed in two stages (i) Poly (A) Tailing (ii) FlashTag Biotin HSR Ligation. The expression data were acquired using the Affymetrix GeneChip Command Console Software (AGCC). The miRNA-1_0 annotations (20081203) supplied by Affymetrix was used as probe identifier.

Data analysis. After log transformation and Robust Multi-array Analysis (RMA) normalization, differentially expressed mRNAs/miRNAs were defined by applying three filtering criteria (mean intensity greater than six, fold change greater than two and ANOVA set to $P<0.05$) using Partek Genomics Suite software v6.5 (Partek Incorporated, Missouri, USA). MiRNAs that were differentially expressed between untreated and treated or follow-up passages were identified using a paired t–test at $P<0.05$ significance. Present study reported differentially expressed mature miRNAs. Unsupervised hierarchical clustering of significant up- and down-regulated genes/miRNAs were applied using Partek software with standard Pearson's correlation as similarity measurement, and Ward'd method for clustering the data.

qRT-PCR validation. To confirm the microarray findings, quantitative real-time (qRT) PCR was used for several candidate mRNAs and miRNAs (Table 3 and 4). For each cell line, RNA used for microarray analysis was also used to synthesize cDNA with High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA). Real-time primers were then designed for each gene using Primerquest Software (Integrated DNA Technologies). For qRT-PCR analysis, synthesized cDNAs amplified with specific gene primers using SYBR® Green 2X PCR Master Mix (Applied Biosystems). In order to analyze miRNAs with qRT-PCR the miScript PCR system kit was used with the commercially available primers for mature miRNAs (QIAGEN AG, Basel, Switzerland). Raw values were normalized using geometric mean of seven reference genes (*18S*, *GAPDH*, *GRB2*, *Hmox2*, *TMEM184B*, *USP7* and *ZNF398*) for mRNA expression and two reference miRNAs (miR-151 and miR-193a) for miRNA expression as internal controls. These reference genes/miRNAs were selected according to their low variability overall microarray data in all the analyzed samples [7,8,9,10]. Real-time PCR reactions were performed in two replicates and including no-template controls using ABI Prism 7000 Sequence Detection System (Applied Biosystems). The fold difference for each sample was calculated using the comparative Ct method [11].

Table 3. The primer sequence of the selected mRNAs for qRT-PCR.

Gene Name	Ensembl Gene ID	Primer	Sequence (5'→3')	Length (bp)	Tₐ	Prodoct Size (bp)
18S*	ENST00000445125	FW	CGCCGCTAGAGGTGAAATTCT	21	60.1	66
		RV	CATTCTTGGCAAATGCTTTCG	21		
BRCA1	ENST00000357654	FW	AGTTGGTCTGAGTGACAAGG	20	60.7	109
		RV	CTGCTTCACCTAAGTTTGAATCC	23		
CDKN1A (P21)	ENST00000244741	FW	TGCGTTCACAGGTGTTTCTG	20	60.3	217
		RV	GTCACCCTCCAGTGGTGTCT	20		
P16-INK4A	ENST00000304494	FW	GAGGAAGAAAGAGGAGGG	17	59.5	265
		RV	CATCATGACCTGGATCGG	18		
CST6	ENST00000312134	FW	CTACTTCCGAGACACGCACA	20	59.8	201
		RV	GGAACCACAAGGACCTCAAA	20		
CXCR4	ENST00000241393	FW	CCAGAAACTTCAGTTTGTTGG	21	60.79	242
		RV	ATGATGGAGTAGATGGTGGG	20		
DNMT1	ENST00000340748	FW	CTGGCTTTGAGAGTTATGAGG	21	60.4	182
		RV	CATTAACACCACCTTCAAGAG	21		
DNMT3A	ENST00000264709	FW	GATGATTGATGCCAAAGAAGTG	22	60.2	272
		RV	CCAAATACCCTTTCCATTTCAG	22		
DNMT3B	ENST00000328111	FW	CAAGGAAATACGAGAACAAGAC	22	59.3	195
		RV	GACAAACAGCCATCTTCCAG	20		
ERBB2	ENST00000269571	FW	GCCTGTCCCTACAACTACC	19	61	178
		RV	GTAACTGCCCTCACCTCTC	19		
ESR1	ENST00000206249	FW	CAGACACTTTGATCCACCTG	20	60.63	115
		RV	GCCTTTGTTACTCATGTGCC	20		
ESR2	ENST00000341099	FW	GGATGGAGGTGTTAATGATGG	21	60.53	233
		RV	GAGGGTACATACTGGAATTGAG	22		
ETS1	ENST00000319397	FW	TCACTAAAGAACAGCAACGAC	21	60.2	250
		RV	TGGTTTCACATCCTCTTTCTG	21		
GAPDH*	ENST00000229239	FW	GAAGGTGAAGGTCGGAGT	18	60	226
		RV	GAAGATGGTGATGGGATTTC	20		
GJA1 (CX43)	ENST00000282561	FW	AGGAAAGTACCAAACAGCAG	20	60.2	209
		RV	CAGTTGAGTAGGCTTGAACC	20		
GRB2*	ENST00000316804	FW	AGAAATGAAACCACATCCGT	20	60.3	155
		RV	ACATCGTTTCCAAACTTGACAG	22		
Hmox2*	ENST00000219700	FW	AACCAAATGAGAATGGCTGAC	21	61.2	164
		RV	GGCTGAGTATGTGAAGTAAAGTG	23		
IL6	ENST00000404625	FW	CACTCACCTCTTCAGAACGA	20	61.08	196
		RV	GCAAGTCTCCTCATTGAATCC	21		
MERTK	ENST00000295408	FW	TCAACATCAAAGCAATTCCCTC	22	61.05	205
		RV	GATTTGGTACAGATGTGGTAAGG	23		
MYC	ENST00000377970	FW	GATTCTCTGCTCTCCTCGAC	20	61.17	110
		RV	TTCTTGTTCCTCCTCAGAGTC	21		
PDCD4	ENST00000280154	FW	GAAGTTGCGGAAATGTTAAGAG	22	60.17	259
		RV	ACAGCTCTAGCAATAAACTGG	21		
PGR	ENST00000263463	FW	CAATGGAAGGGCAGCAC	17	60.58	231
		RV	CCACTGACGTGTTTGTAGG	19		
PTEN	ENST00000371953	FW	CGGAACTTGCAATCCTCAG	19	60.8	233
		RV	GTTTCCTCTGGTCCTGGT	19		
RASSF1	ENST00000359365	FW	AAGTTCACCTGCCCACTACC	19	60.4	217
		RV	CCGTCCTTGTTCAAGCTC	18		
TMEM184B*	ENST00000361906	FW	CACCAGATCTACATGCACCT	20	61.3	`277
		RV	TACATACAGCTGGACTCAATGG	22		
USP7*	ENST00000344836	FW	ATTCCTAACATTGCCACCAG	20	60.1	256
		RV	ATTTACACCATTTGCCATCC	20		
VEGFA	ENST00000372067	FW	GAGTACCCTGATGAGATCGAG	21	60.05	213
		RV	CTTTCTTTGGTCTGCATTCAC	21		
ZNF398*	ENST00000420008	FW	CAGGTATTAAGGGAGATATCCCA	23	60.2	161
		RV	TAGCATAATCCATGGAGATGAG	22		

* selected as reference gene

Table 4. The primer sequence and ID of validated miRNAs by qRT-PCR.

MIR ID	Sanger ID	Sanger Accession	Sequence 5' → 3'
hsa-mir-125b-1	hsa-miR-125b	MIMAT0000423	UCCCUGAGACCCUAACUUGUGA
hsa-mir-130a	hsa-miR-130a	MIMAT0000425	CAGUGCAAUGUUAAAAGGGCAU
hsa-mir-151**	hsa-miR-151-3p	MIMAT0000757	CUAGACUGAAGCUCCUUGAGG
hsa-mir-152	hsa-mir-152	MIMAT0000438	UCAGUGCAUGACAGAACUUGG
hsa-mir-181a	hsa-miR-181a	MIMAT0000256	AACAUUCAACGCUGUCGGUGAGU
hsa-mir-183	hsa-mir-183	MIMAT0000261	UAUGGCACUGGUAGAAUUCACU
hsa-mir-193a**	hsa-miR-193a-5p	MIMAT0004614	UGGGUCUUUGCGGGCGAGAUGA
hsa-miR-21_2	hsa-miR-21	MIMAT0000076	UAGCUUAUCAGACUGAUGUUGA
hsa-miR-24_1	hsa-miR-24	MIMAT0000080	UGGCUCAGUUCAGCAGGAACAG
hsa-mir-27a	hsa-miR-27a	MIMAT0000084	UUCACAGUGGCUAAGUUCCGC
hsa-mir-27b	hsa-miR-27b	MIMAT0000419	UUCACAGUGGCUAAGUUCUGC
hsa-miR-29b_1	hsa-miR-29b	MIMAT0000100	UAGCACCAUUUGAAAUCAGUGUU
hsa-mir-378	hsa-miR-378	MIMAT0000732	ACUGGACUUGGAGUCAGAAGG
hsa-mir-455	hsa-miR-455-3p	MIMAT0004784	GCAGUCCAUGGGCAUAUACAC
hsa-miR-99a_1	hsa-miR-99a	MIMAT0000097	AACCCGUAGAUCCGAUCUUGUG

** selected as reference miRNAs

Mutation screening of P16-INK4A gene

Full gene sequencing of *P16-INK4A* has been performed on the promoter region, 5'UTR, three exons including exon/intron boundaries (exon 1-3) and 3'UTR. Information of primer sequences are listed
in Table 5. The PCR for each region carried out in a 50 μL total volume containing 100ng DNA, 200μM of each dNTP, 20 pmoL of each primer, 2.5 U Hot Start Taq DNA polymerase, 1X PCR buffer and 1.5 mM MgCl2**.** The PCR was performed in 40 cycles under following condition: denaturation at 94°C for 15 sec, annealing step for 15 sec at appropriate T_a that is mentioned in Table 5, and 60 sec primer extension at 72°C. Direct DNA sequencing was performed using a Big Dye terminator v3.1 cycle sequencing kit and automated sequencer was performed (ABI 3130, Applied Biosystem). The results of DNA sequence analysis were compared with the reference sequences of the gene bank (www.ncbi.nlm.mih.gov/nucleotide) using DNASTAR sequence alignment software (DNASTAR Lasergene 8, Inc., Madison, USA). All sequencing reactions were performed in both directions and confirmed for concordance.

Table 5. The primer sequence for full *P16-INK4A* gene sequencing.

Gene Name	Primer	Sequence (5'→3')	Length (bp)	T_a	Prodoct Size (bp)
P16-INK4A (Promoter)	FW	GCTCCTGAAAATCAAGGGTTG	21	58	648
	RV	CCTGCTCTCCCCCTCTCC	18		
P16-INK4A (5'UTR + Exon1*)*	FW	AGTCCTCCTTCCTTGCCAAC	20	57	620
	RV	CTTCTGAAAACTCCCCAGGA	20		
P16-INK4A (Exon2)	FW	AATTAGACACCTGGGGCTTG	20	57	582
	RV	AGGGCGATAGGGAGACTCAG	20		
P16-INK4A (Exon3 + 3'UTR)	FW	GTAGGGACGGCAAGAGAGG	19	57	660
	RV	TGAAACAACAGTGTCAGAAACG	22		

Proteomic profiling

Protein pallets from AllPrep® extraction (5 x 10^6 cells per sample) were directly solubilized in a buffer consisting 7M urea, 2M thiourea, 4% CHAPS, 20mM DTT and 2% ampholines, pH 3-10 (Invitrogen) for a total volume of 0.1mL.

Two-dimensional gel electrophoresis (2DE) and spot visualization the first dimension was run using the ISODALT system (ampholines pH 3-10 (Invitrogen)) [12,13]; and for the second dimension 11-19% linear acrylamide gradient was used. Protein spots were visualized by silver staining and wet silver-stained gels were scanned by Pharmacia Image Scanner with 300 dpi, 16 bit based on published method [1].

Image analysis and spot quantification. Each protein sample was run by 2DE in triplicate in order to evaluate gel reproducibility and improve the reliability of the qualitative and qualitative changes in protein expression measurement by means of electrophoresis. Progenesis SameSpot software (v 4.0, NonLinear Dynamics, UK) was used for gel alignment, spot detection, spot quantification, and normalization for the total spot volume in each gel, and the data were statistically analyzed using the incorporated statistical package. Gel images were automatically aligned after manually assigning 20 landmark vectors. The samples variables were expressed as mean of each replicate (SE or percentages, and were statistically analyzed by ANOVA). The cutoff level for a differentially expressed protein was defined based on an ANOVA at significance of $P<0.05$ using the SameSpots software considering a minimum of 1.5-fold change (normalized volume). In addition, we applied separate non-parametric statistical analyses between each pair of samples including untreated vs. treated and other follow-up passages.

Liquid Chromatography - Mass Spectrometry and Liquid Chromatography - Tandem Mass Spectrometry (LC-MS-MS)

The protein spots of interest were manually excised from the gels and underwent in-gel digestion with trypsin. The trypsin digested proteins were analyzed by capillary liquid chromatography tandem MS (LC-MS-MS) using a setup of a ProteoCol trap C-18 column (0.15 x 10mm, 3 μm particle size, 300Å) (SGE Analytical Science, Victoria, AU) and a separating column (0.1mm x10cm) that had been packed with Magic 300Å C18 reverse-phase material (5 mm particle size, Swiss Bioanalytics, Birsfelden, Switzerland). The columns were connected on line to an Orbitrap FT hybrid instrument (Thermo Finnigan, San Jose, CA, USA). The solvents used for peptide separation consisted 0.1% acetic acid in water (solvent A) and 0.1% acetic acid and 80% acetonitrile in water (solvent B). Peptides were injected via a 2 μL loop onto the trap column with the capillary pump of an Agilent 1200 system set to 5 μL/min. After 15 min, the trap column was switched into the flow path of the separating column. A linear gradient from 2 to 35% solvent B in solvent A in 60 min was delivered with an Agilent 1200 nano pump at a flow rate of 500 nL/min. After 60 min the percentage of solvent B was increased to 60% in ten minutes and further increased to 80% in 2 min. The eluting peptides were ionized at 1.7 kV. The mass spectrometer was operated in a data-dependent fashion. The precursor scan was done in the Orbitrap set to 60,000 resolutions, while the fragment ions were mass analyzed in the LTQ instrument. A top five method was run so that the five most intense precursors were selected for fragmentation.

Peptide identification. The MS-MS spectra were then searched against the human data bank (NCBI non-redundant, version October 1st, 2010) using TurboSequest software [14]. The data bank was searched with 10 ppm precursor ion tolerance, while the fragment ions were set to 0.5 Da tolerances. Cleavage rules were set to fully enzymatic – cleaves at both ends, allowing 2 missed cleavages. Post filtering was set to the following parameters: ΔCN, 0.1; Xcorr versus charge state was 1.50 (1+), 2.00 (2+), 2.50 (3+); peptide probability, 0.01; protein probability 0.01. Valid identification required two or more peptides independently matching the same protein sequence, a significant peptide score ($P<0.05$), and the manual confirmation of agreement between the spectra and peptide sequence. At least two peptides were required for protein identification. Protein quantification and data validation was performed using Scaffold (version Scaffold 3.00.06; Proteome Software, Portland, OR, USA), which models the score distributions of the entire dataset of spectra. Database search files generated by TurboSequest were imported into Scaffold and analyzed using the tandem searches against the same protein sequence database the same search parameters as the associated TurboSequest search.

Gene ontology Enrichment

The gene ontology (GO) enrichment was assessed using Partek Genomics Suite software v6.5 (Partek Incorporated, Missouri, USA), which could take a list of significantly expressed genes and see how they group in the functional hierarchy. The enrichment score were calculated using a chi-square test comparing the proportion of the gene list in a group to the proportion of the background

in the group. If a functional group had an enrichment score over 1, the functional category was over expressed. A value of 3 or higher corresponds to significant over expression ($P<0.05$).

In silico prediction of miRNA targets

Conserved and non-conserved targets of detected miRNAs were identified using TargetScan 5.1 (http://www.targetscan.org/). List of predicted targets was obtained from TargetScan data download and further investigated using pathway analysis. For important miRNAs, the mRNA targets were validated to find exact miRNA complementary sites using RNA22 program (http://cbcsrv.watson.ibm.com/rna22.html). The default stringency settings were used: maximum number of allowed UN-base paired bases =0 in seed/nucleus of 7 nucleotides, and minimum number of paired-up bases in heteroduplex = 14; maximum folding energy for heteroduplex (Kcal/mol =225).

Cell signaling and pathway analysis

Gene networks and canonical pathways representing key genes, miRNAs and proteins were identified using the Pathway Studio® software version 7 and ResNet® 7 (Mammal) database (Ariadne Genomics, Inc., Rockville, USA). The data set containing gene/miRNA identifiers and corresponding fold changes were uploaded into the Pathway Studio and each gene identifier was mapped to its corresponding gene object. The functional analysis identified the biological functions and/or diseases that were most relevant to the data sets and facilitated the understanding beyond their functional link to breast neoplasm. This analysis provided an approach to compare different kinds of cellular interactions, including protein-protein interactions as well as genes-miRNAs interactions.

Reference
1. Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, et al. (2009) Simultaneous Isolation of DNA, RNA, and Proteins for Genetic, Epigenetic, Transcriptomic, and Proteomic Analysis. J Proteome Res 8: 5264-5274.
2. Radpour R, Barekati Z, Haghighi MM, Kohler C, Asadollahi R, et al. (2010) Correlation of telomere length shortening with promoter methylation profile of p16/Rb and p53/p21 pathways in breast cancer. Mod Pathol 23: 763-772.
3. Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, et al. (2009) Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28: 2969-2978.
4. Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M, et al. (2008) Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci U S A 105: 4844-4849.
5. Radpour R, Haghighi MM, Fan AX, Torbati PM, Hahn S, et al. (2008) High-Throughput Hacking of the Methylation Patterns in Breast Cancer by In vitro Transcription and Thymidine-Specific Cleavage Mass Array on MALDI-TOF Silico-Chip. Mol Cancer Res 6: 1702-1709.
6. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, et al. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 102: 15785-15790.
7. Gur-Dedeoglu B, Konu O, Bozkurt B, Ergul G, Seckin S, et al. (2009) Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. Oncol Res 17: 353-365.
8. Kwon MJ, Oh E, Lee S, Roh MR, Kim SE, et al. (2009) Identification of novel reference genes using multiplatform expression data and their validation for quantitative gene expression analysis. PLoS One 4: e6162.
9. Peltier HJ, Latham GJ (2008) Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. RNA 14: 844-852.
10. Latham GJ (2010) Normalization of microRNA quantitative RT-PCR data in reduced scale experimental designs. Methods Mol Biol 667: 19-31.
11. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.
12. Grussenmeyer T, Meili-Butz S, Dieterle T, Traunecker E, Carrel TP, et al. (2008) Quantitative Proteome Analysis in Cardiovascular Physiology and Pathology. I. Data Processing. J Proteome Res.
13. Anderson NL, Anderson NG (1978) Analytical techniques for cell fractions. XXII. Two-dimensional analysis of serum and tissue proteins: multiple gradient-slab gel electrophoresis. Anal Biochem 85: 341-354.
14. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR, 3rd (2000) Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. Anal Chem 72: 757-763.

Supplementary Data 2

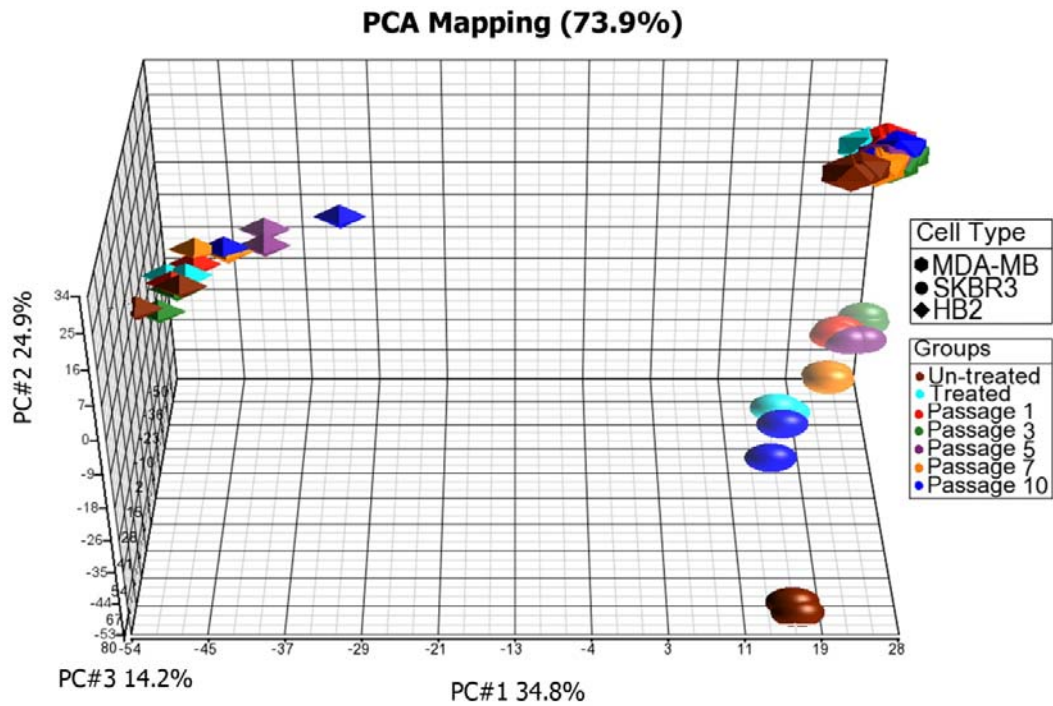**MRNA Expression Profiles**

Fig. 1. PCA analysis of three different studied cell lines (HB2, MDA-MB231 and SKBR3).
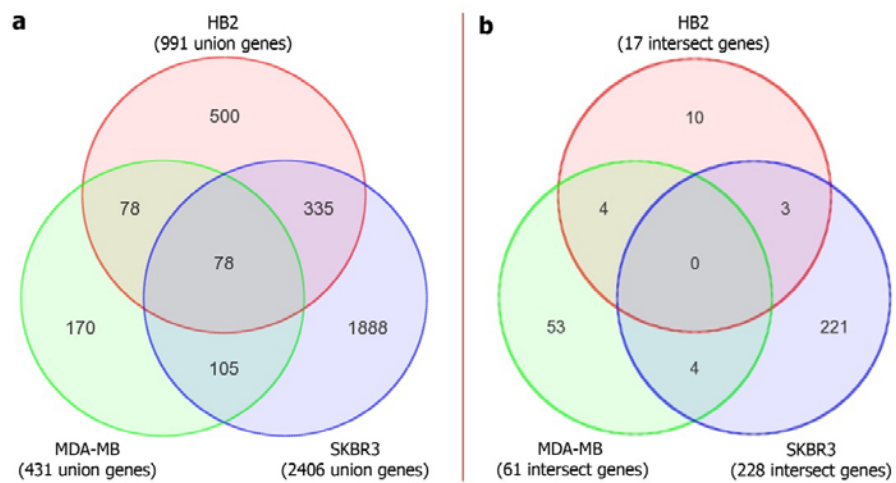

Fig. 2. a) Comparison of significant up- and down-regulated union genes within three analyzed cell lines. b) Comparison of significant up- and down-regulated intersect genes within three analyzed cell lines.
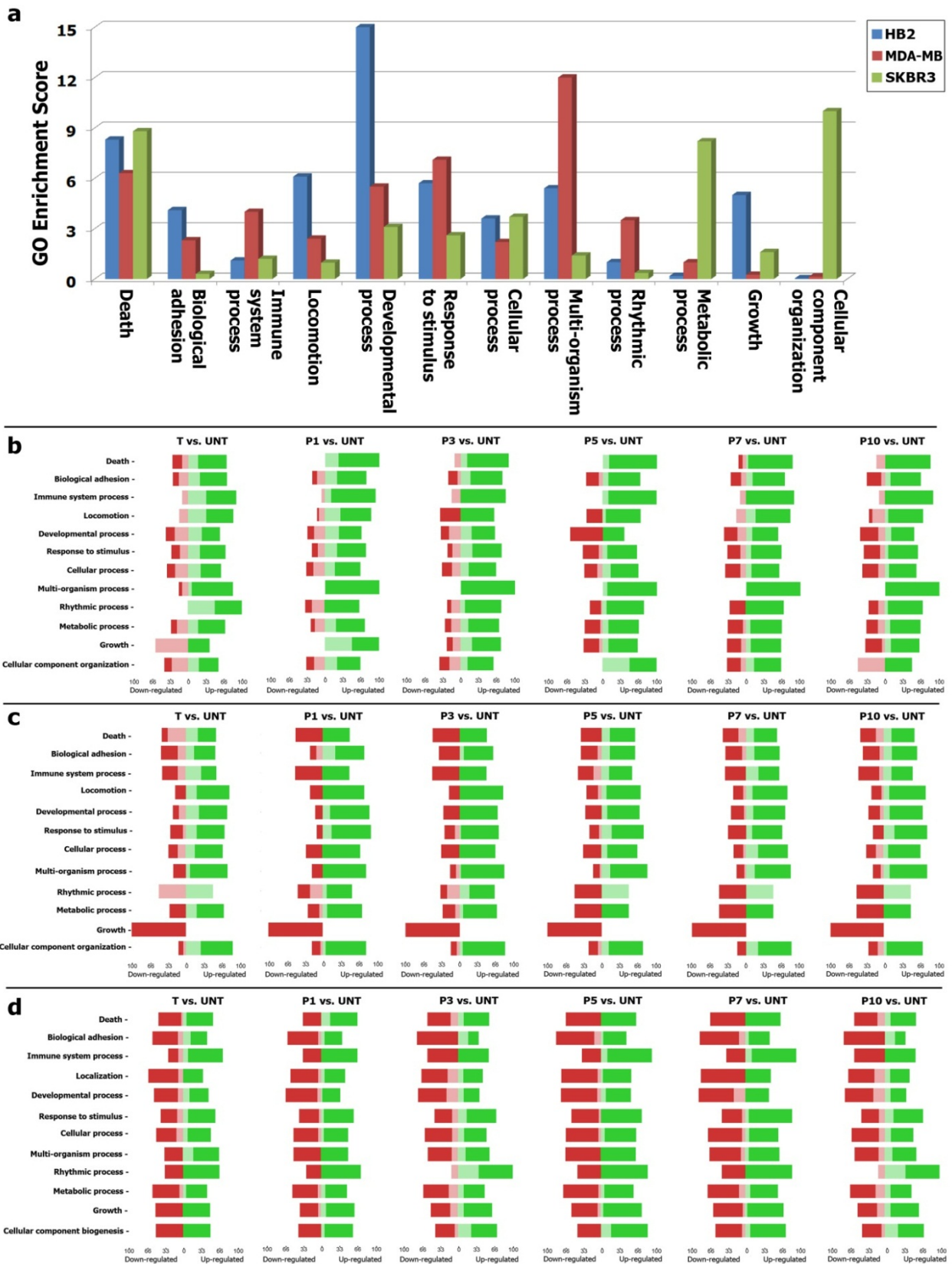
Fig. 3. a) Gene set analysis and GO enrichment score for union genes of three analyzed cell lines. A value of 3 or higher corresponds to significant over expression (chi-square test; *P*<0.05). b) Forest plot analysis shows up- and down-regulated of HB2 union genes based on their ontology enrichment. c) Forest plot analysis of MDA-MB231 union genes. d) Forest plot analysis of SKBR3 union genes.

Gene Expression Profiles of HB2
(Breast Epithelial Cell Line)
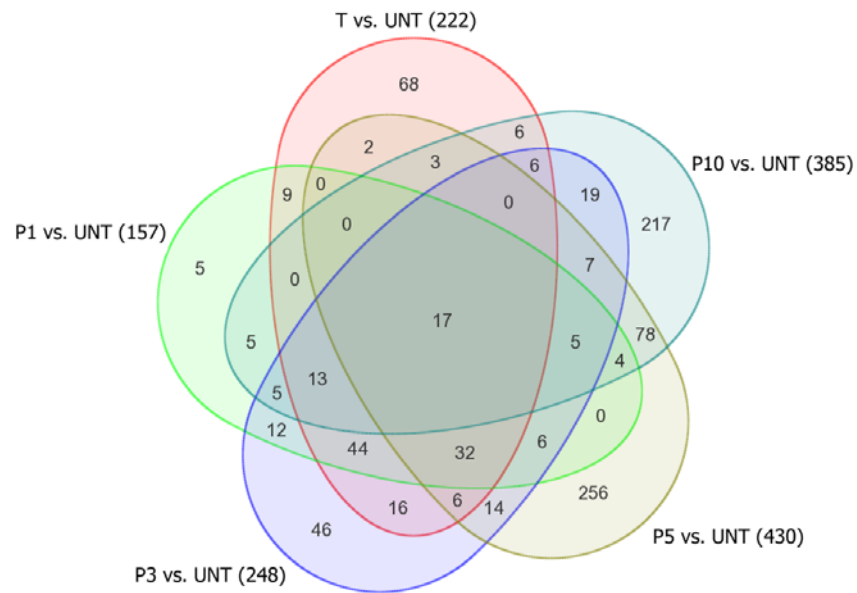
Fig. 4. The number of significant up- and down-regulated genes within different passages of HB2.
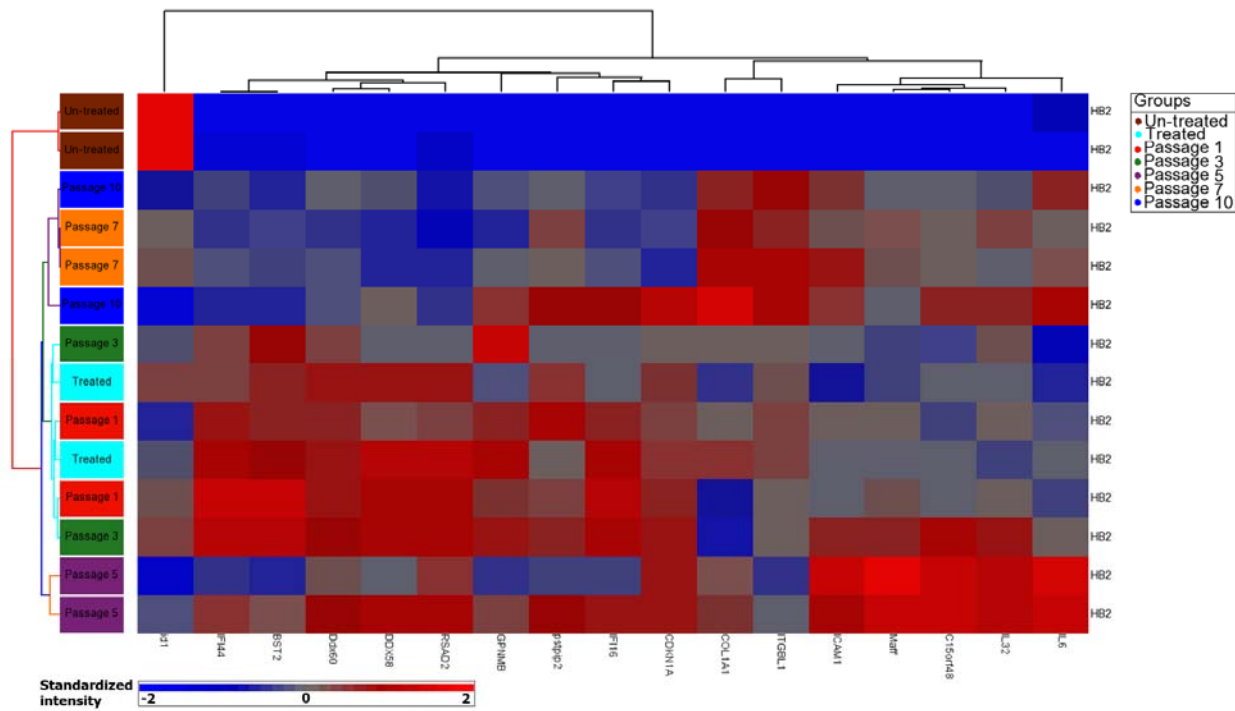
Fig. 5. Heatmap shows significant up- and down-regulated 17 intersect genes of HB2.
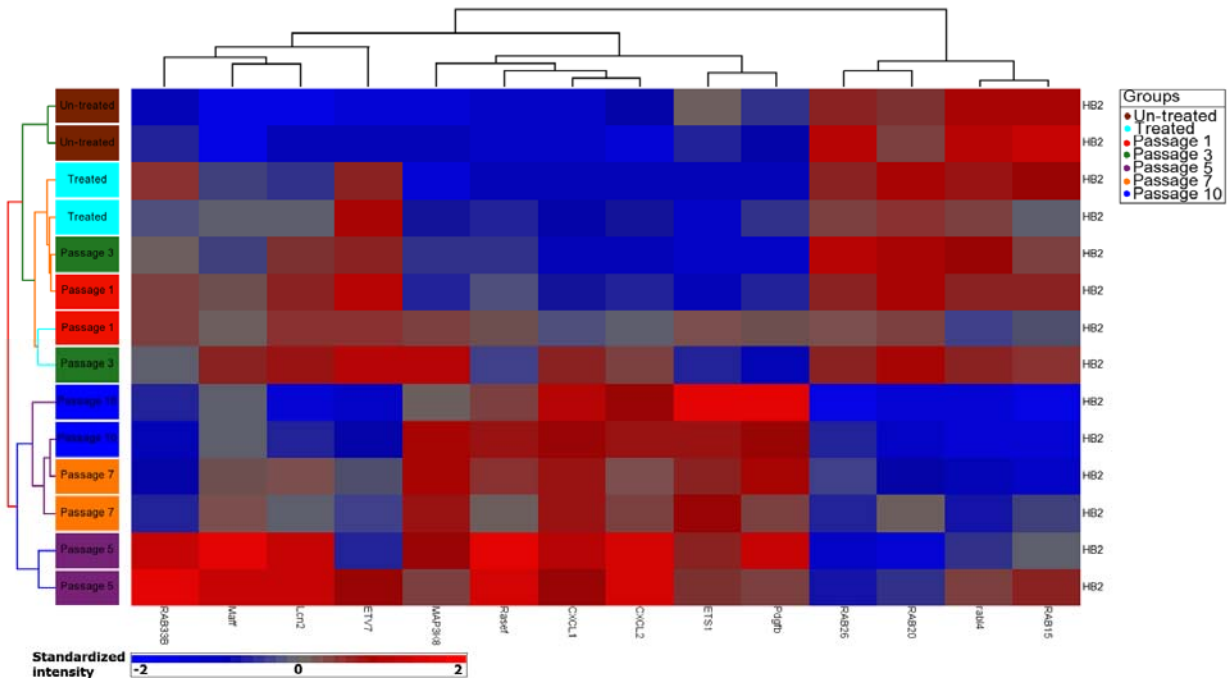
Fig. 6. Heatmap shows significant up- and down-regulated 14 oncogenes out of 991 union genes in HB2.
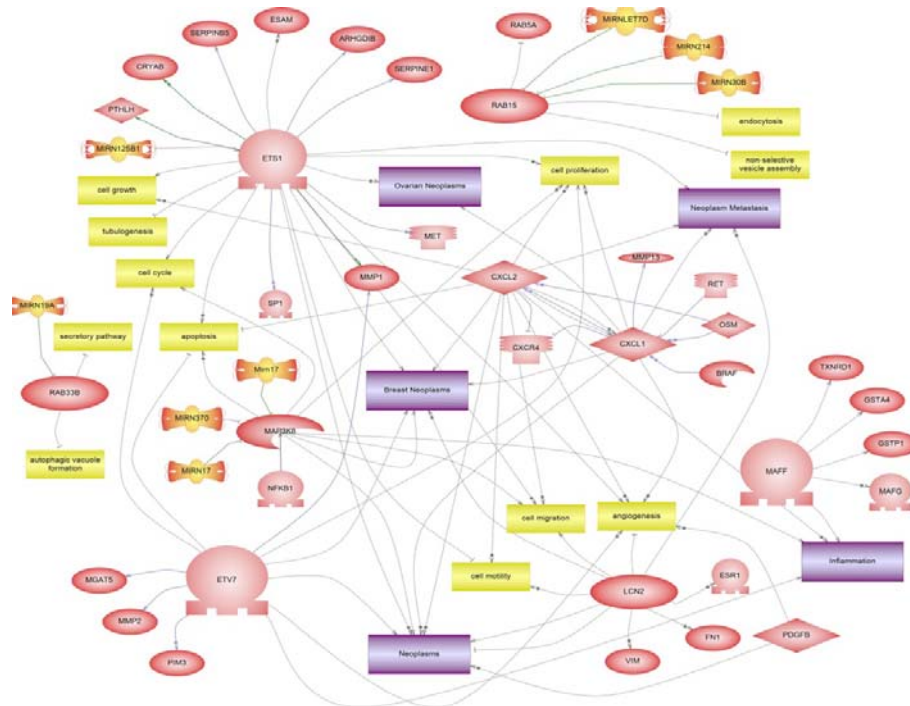
Fig. 7. Pathway analysis of 10 out of 14 oncogenes (*ETS1*, *RAB1S*, *CXCL1*, *CXCL2*, *MAFF*, *LCN2*, *PDGFB*, *ETV7*, *RAB33B* and *MAP3K8*) that are linked to breast neoplasms and metastasis. These genes are shown as significant up- and down-regulated during the different passages of HB2.
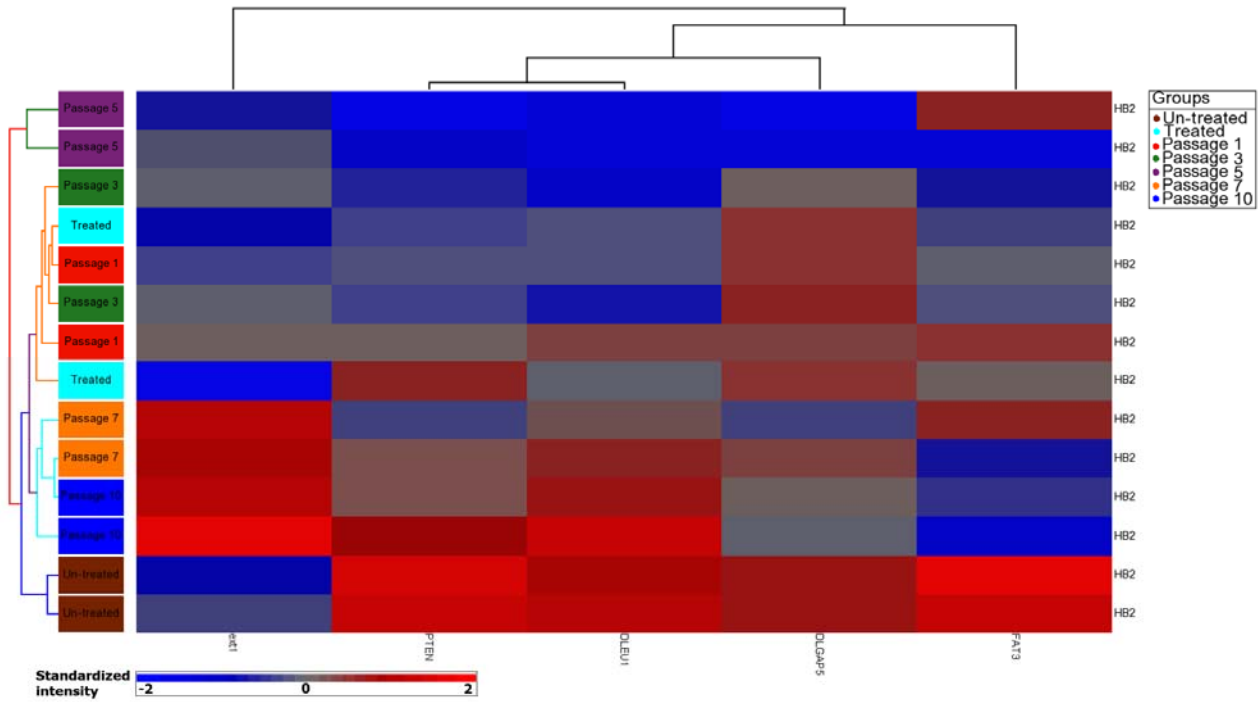
Fig. 8. Heatmap shows significant up- and down-regulated five tumor suppressor genes out of 991 union genes in HB2.
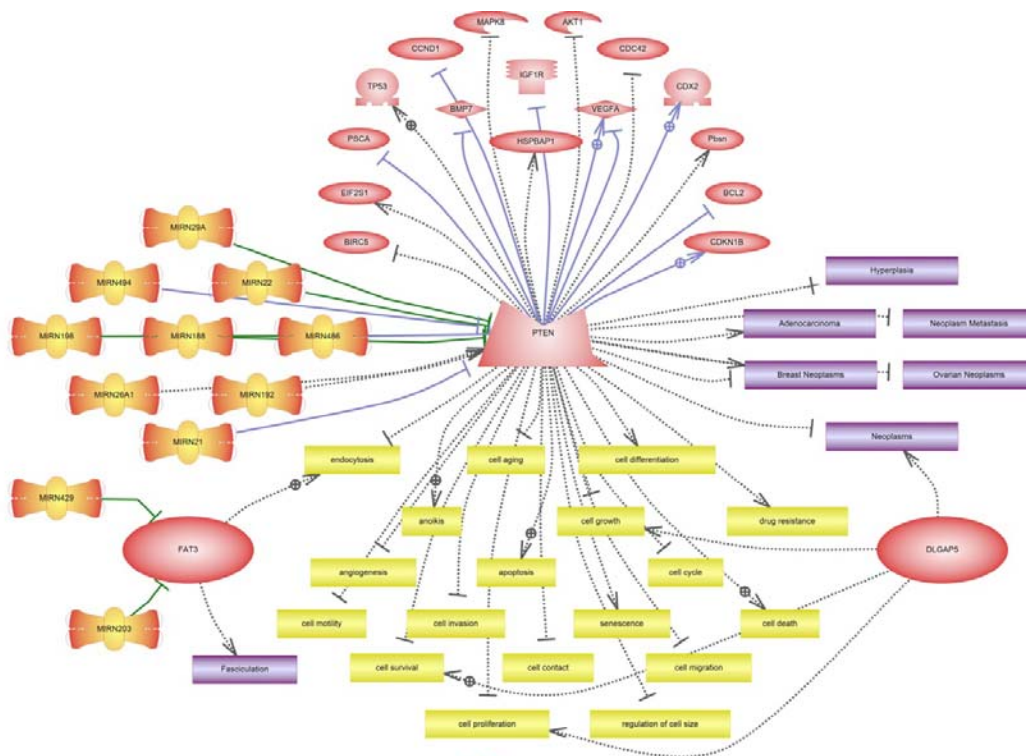
Fig. 9. Pathway analysis of three out of five tumor suppressor genes (*FAT3*, *PTEN* and *DLGAP5*) that are linked to breast neoplasms and metastasis. These are significant up- and down-regulated genes during the different passages of HB2.
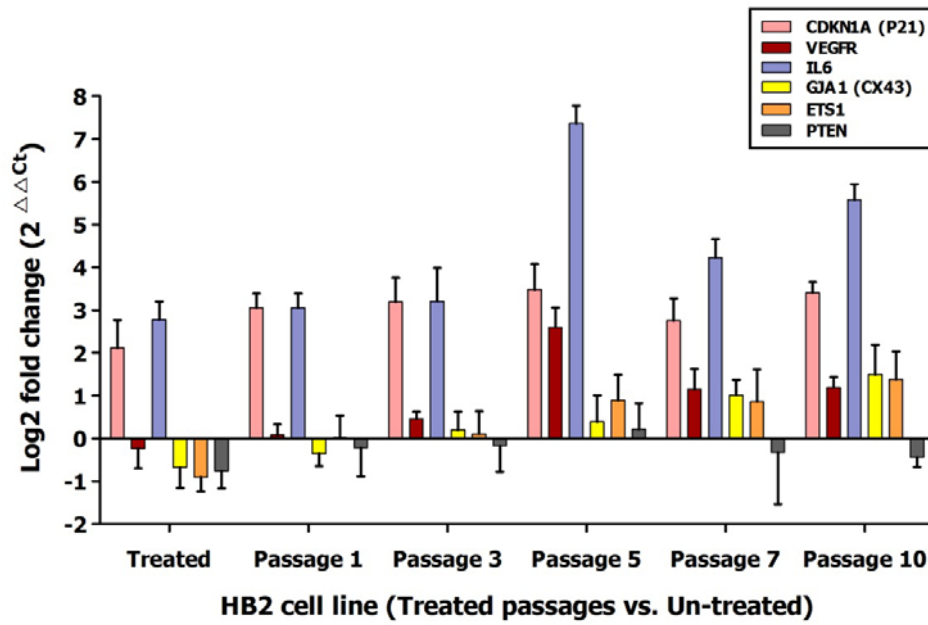
Fig. 10. Expression level of six cancer related genes after treatment using qPCR. Four genes (*CDKN1A*, *VEGFR*, *IL6* and *GJA1*) as intersection genes in the three analyzed cell lines, *ETS1* and *PTEN* as examples of significantly changed oncogenes and tumor suppressor genes respectively.

Gene Expression Profiles of MDA-MB231
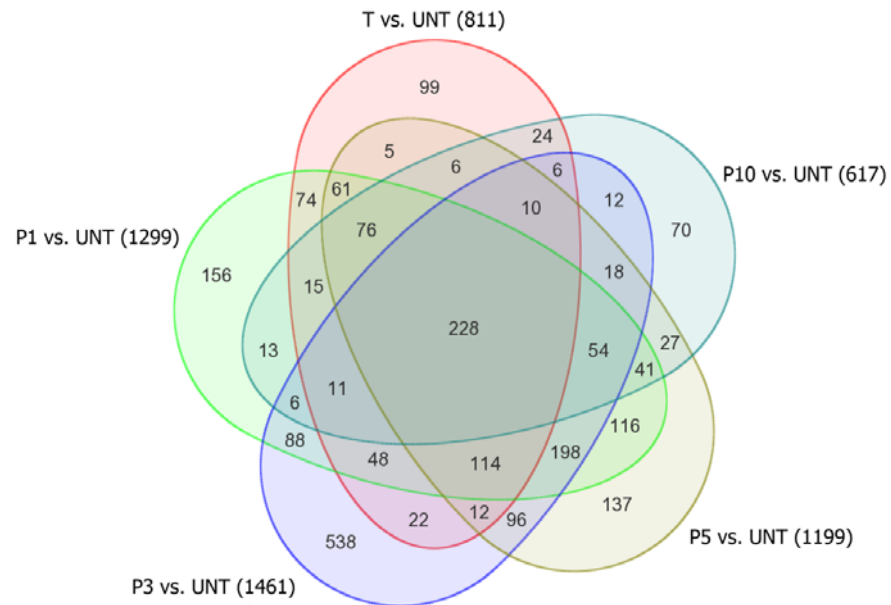**(Highly Aggressive Breast Cancer Cell Line)**
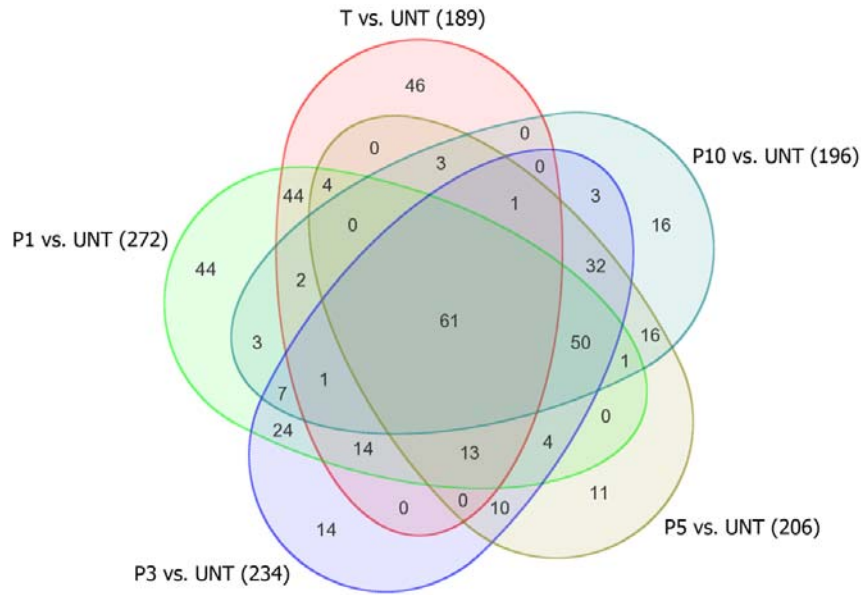
Fig. 11. The number of significant up- and down-regulated genes within different passages of MDA-MB231.
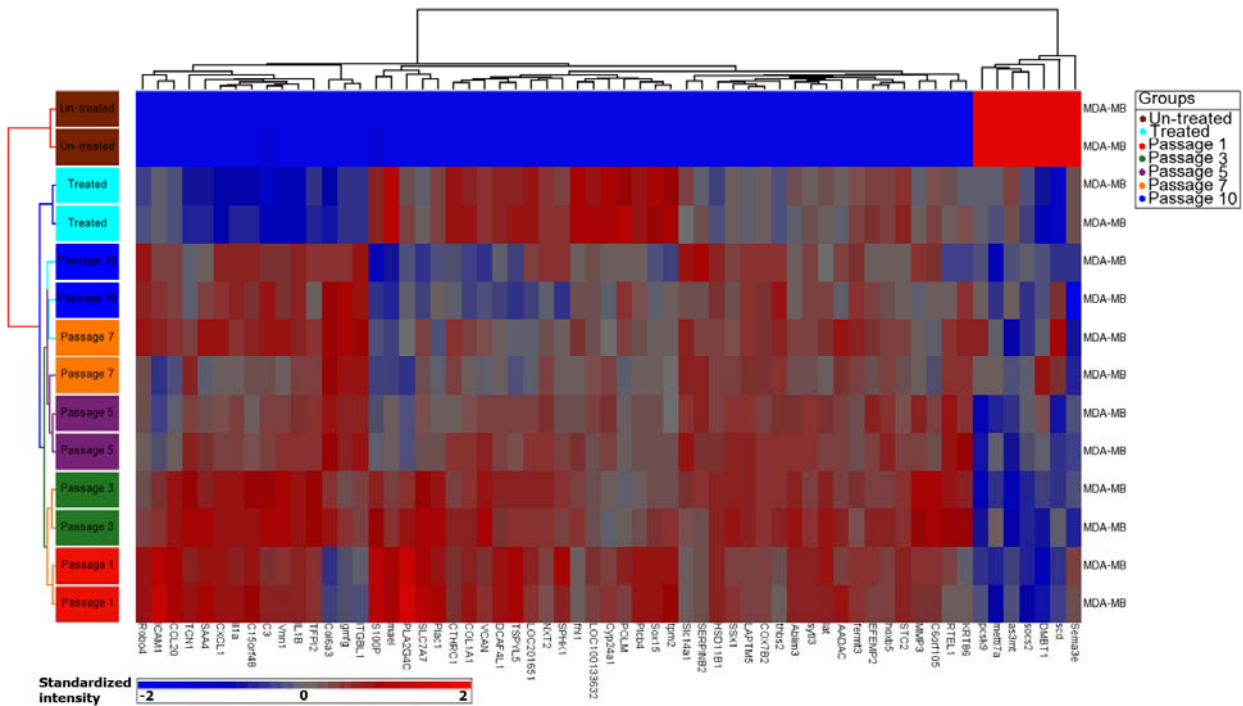
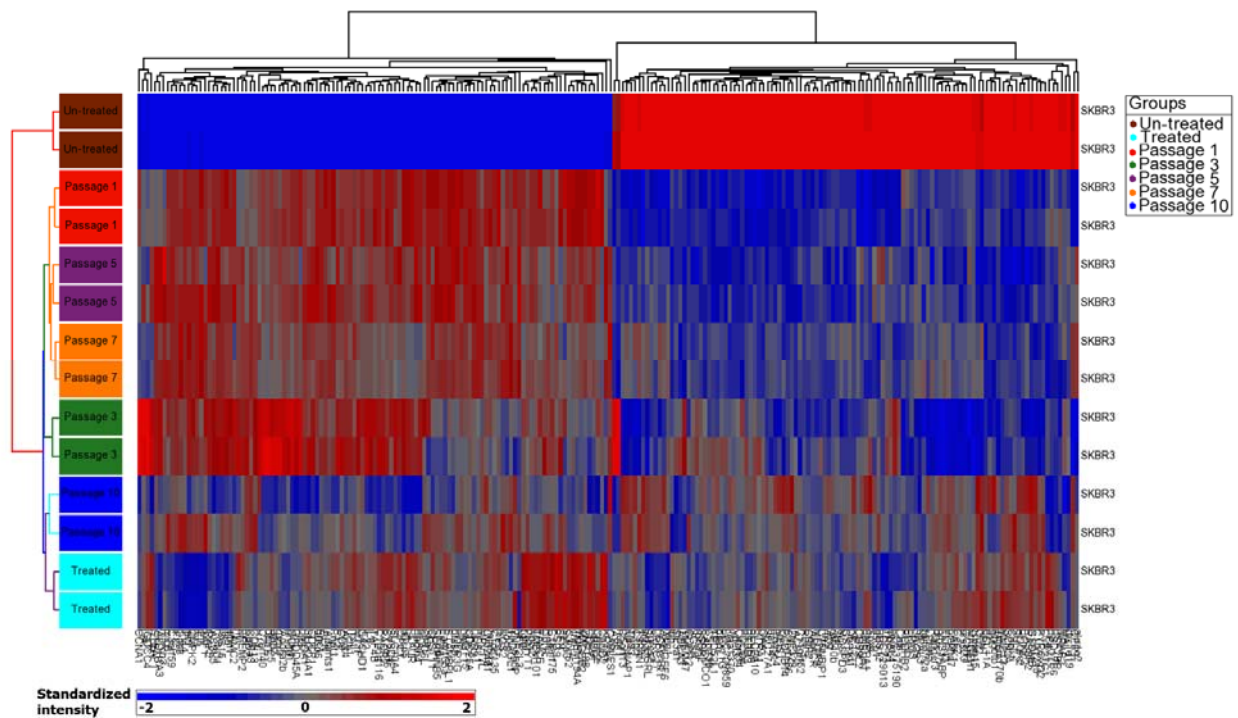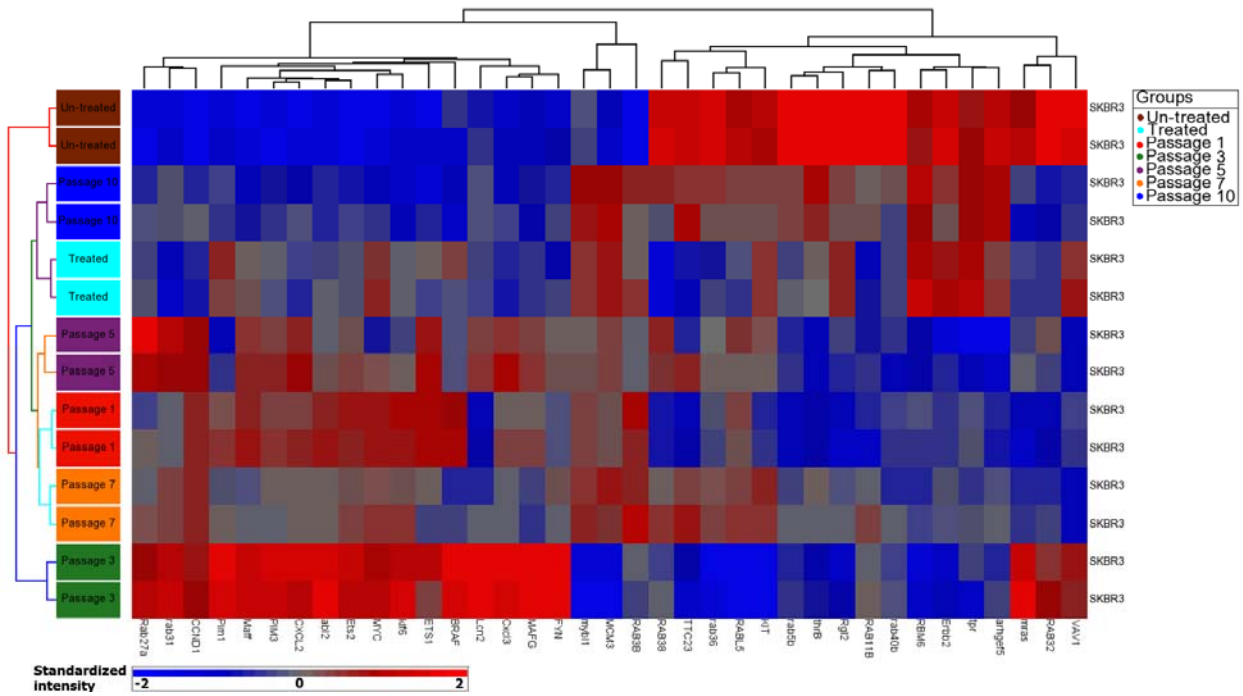Fig. 12. Heatmap shows significant up- and down-regulated 61 intersect genes of MDA-MB231.
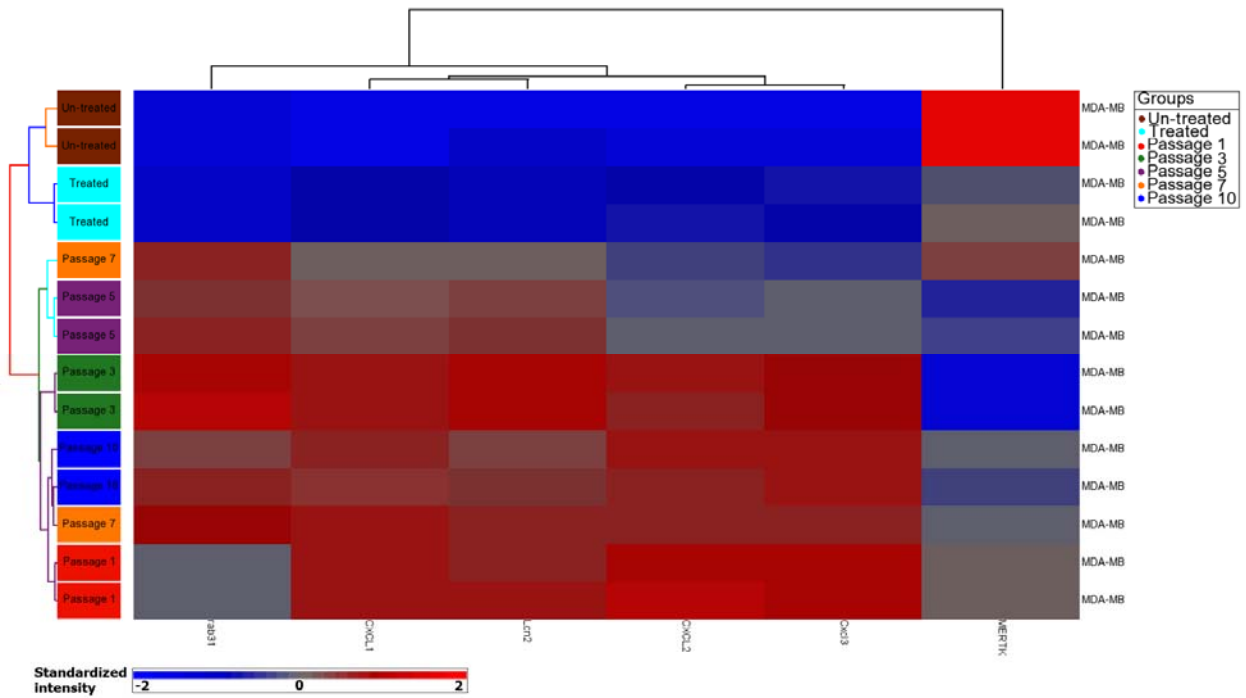
Fig. 13. Heatmap shows significant up- and down-regulated six oncogenes out of 431 union genes in MDA-MB231.
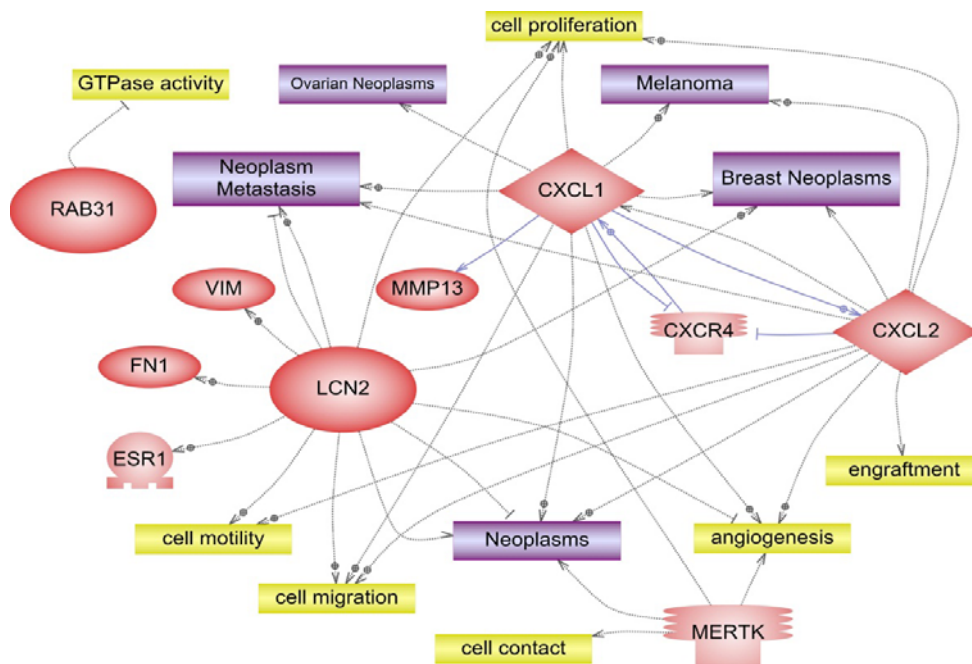
Fig. 14. Pathway analysis of five out of six oncogenes (*MERTK*, *RAB31*, *CXCL1*, *CXCL2* and *LCN2*) that are linked to breast neoplasms and metastasis. These genes are shown as significant up- and down-regulated during the different passages of MDA-MB231.
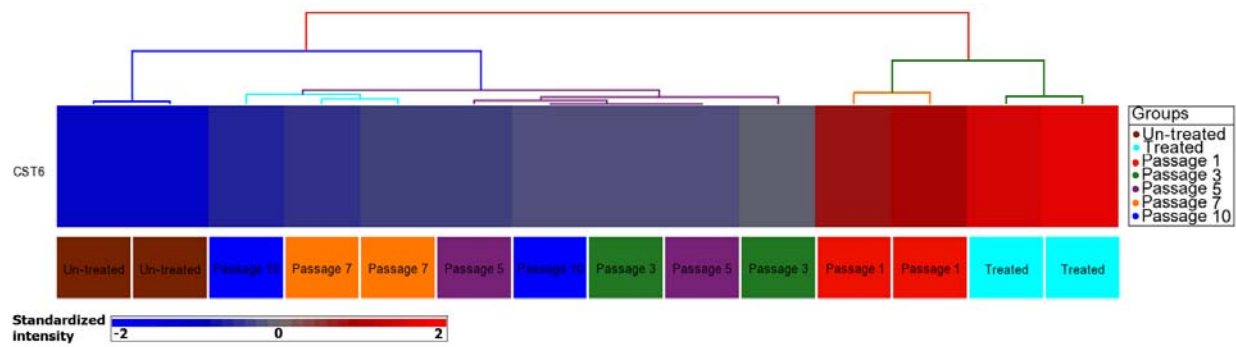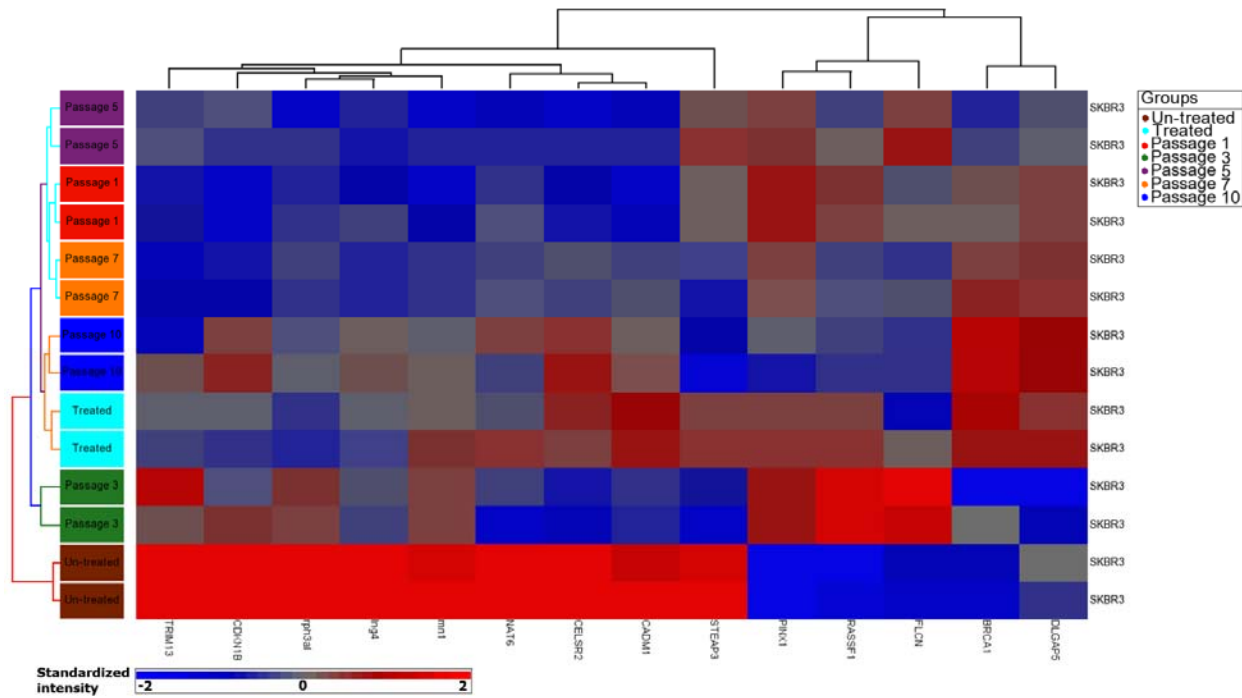
Fig. 15. Heatmap shows significant up- and down-regulated tumor suppressor gene out of 431 union genes
in MDA-MB231.
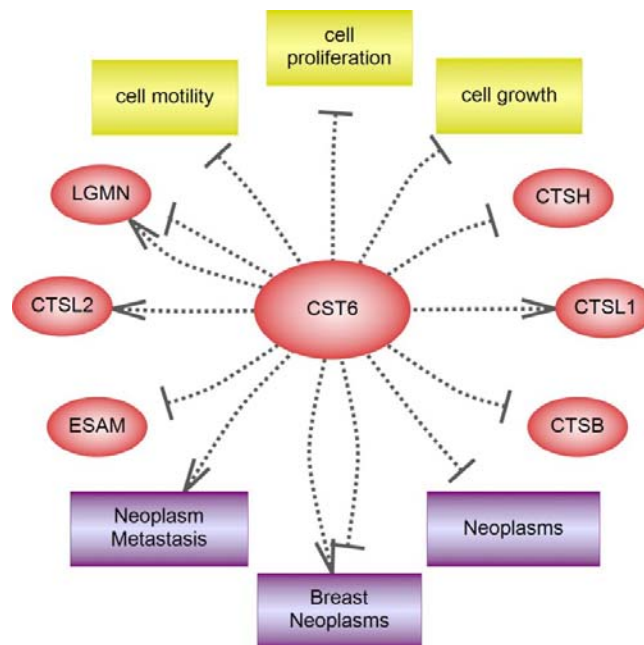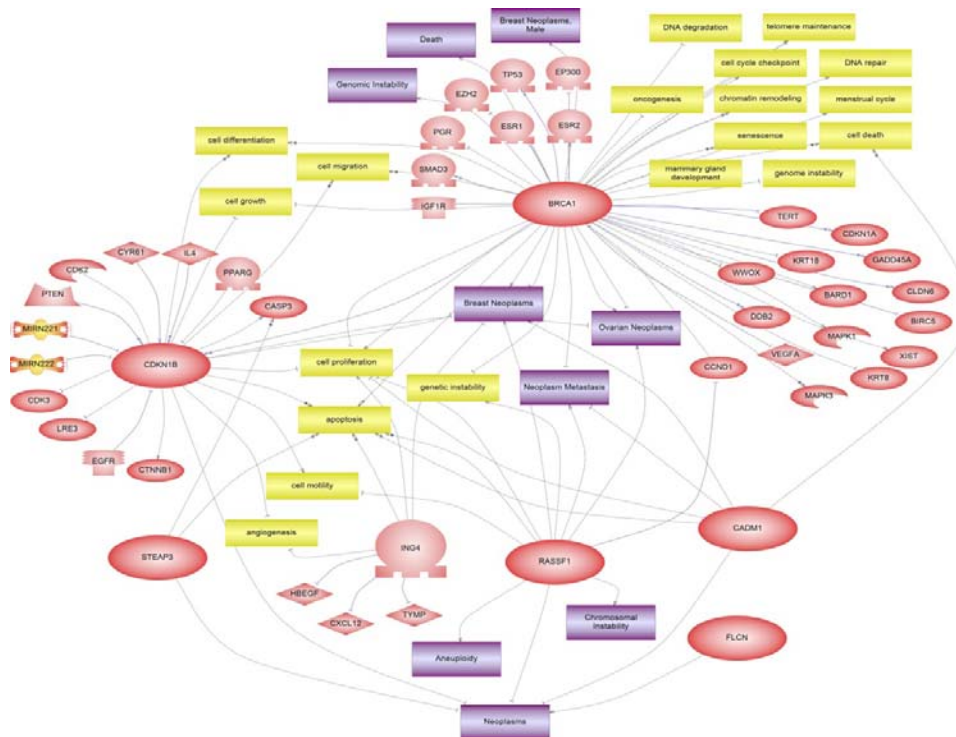

Fig. 16. Pathway analysis of *CST6* tumor suppressor gene that are linked to breast neoplasms and metastasis. This gene is marked as significant up- and down-regulated during the different passages of MDA-MB231.
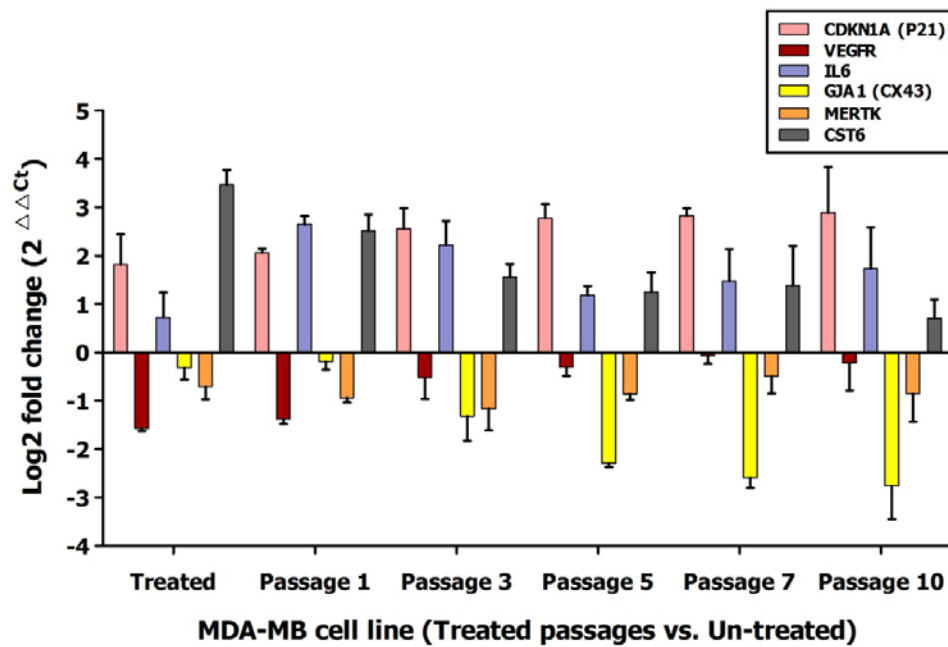
Fig. 17. Expression level of six cancer related genes after treatment using qPCR. Four genes (*CDKN1A*, *VEGFR*, *IL6* and *GJA1*) as intersection genes in the three analyzed cell lines, *MERTK* and *CST6* as examples of significantly changed oncogenes and tumor suppressor genes respectively.

Gene Expression Profiles of SKBR3
(Non-aggressive Breast Cancer Cell Line)

Fig. 18. The number of significant up- and down-regulated genes within different passages of SKBR3.

Fig. 19. Heatmap shows significant up- and down-regulated 228 intersect genes of SKBR3.

Fig. 20. Heatmap shows significant up- and down-regulated 37 oncogenes out of 2406 union genes in SKBR3.
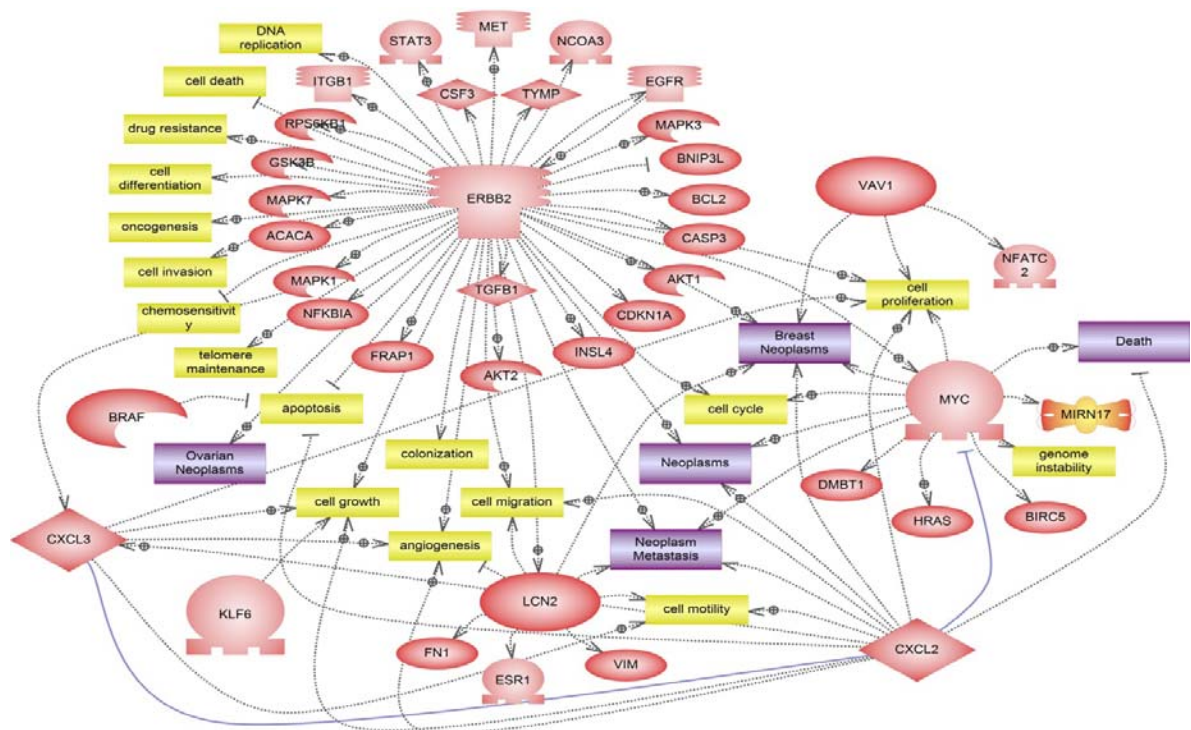
Fig. 21. Pathway analysis of six out of 37 oncogenes (*CXCL2*, *CXCL3*, *ERBB2*, *VAV1*, *MYC*, *BRAF*, *KLF6* and *LCN2*) that are linked to breast neoplasms and metastasis. These genes are shown as significant up- and down-regulated during the different passages of SKBR3.

Fig. 22. Heatmap shows significant up- and down-regulated 14 tumor suppressor genes out of 2406 union genes in SKBR3.

Fig. 23. Pathway analysis of seven out of 14 tumor suppressor genes (*BRCA1*, *CDKN1B*, *STEAP3*, *ING4*, *RASSF1*, *CADM1* and *FLCN*) that are linked to breast neoplasms and metastasis. These genes are shown as significant up- and down-regulated during the different passages of SKBR3.
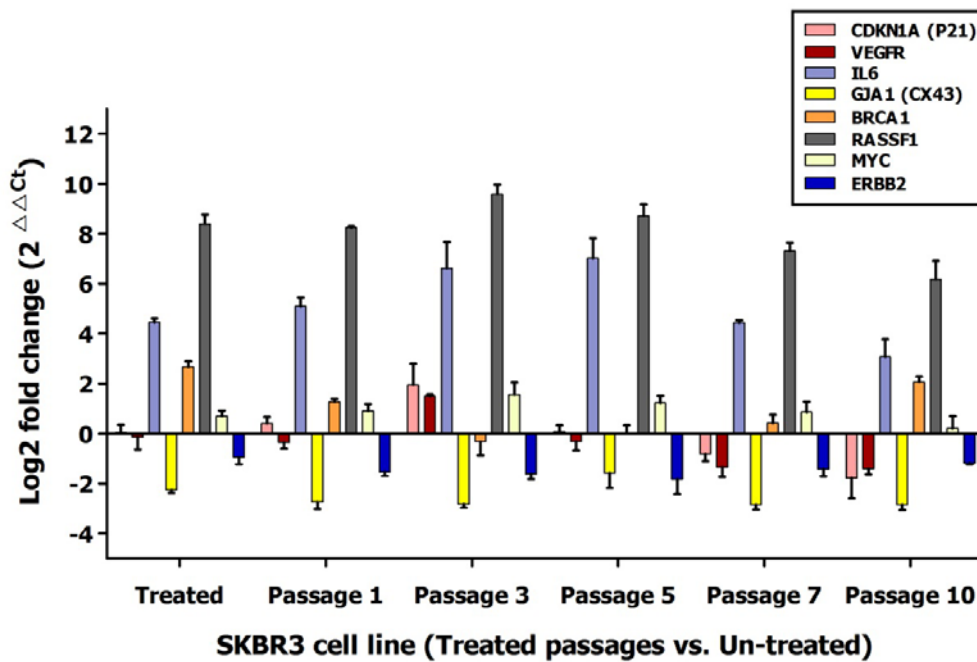
Fig. 24. Expression level of eight cancer related genes after treatment using qPCR. Four genes (*CDKN1A*, *VEGFR*, *IL6* and *GJA1*) as intersection genes in the three analyzed cell lines, *MYC* and *ERBB2* as examples of significantly changed oncogenes, *BRCA1* and *RASSF1* as an example of tumor suppressor genes.

Supplementary Data 3
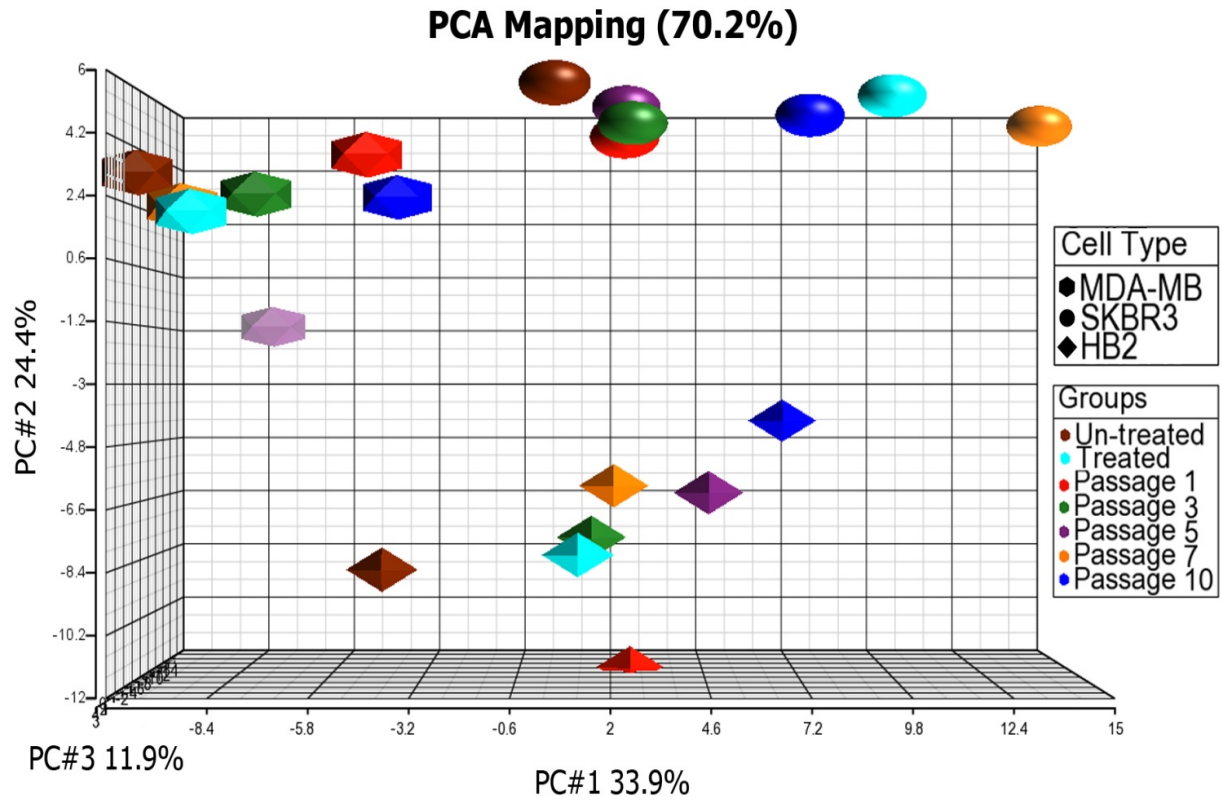
MicroRNA Expression Profiles

Fig. 1. PCA analysis of three different studied cell lines (HB2, MDA-MB231 and SKBR3).

MicroRNA Expression Profiles of HB2
(Breast Epithelial Cell Line)
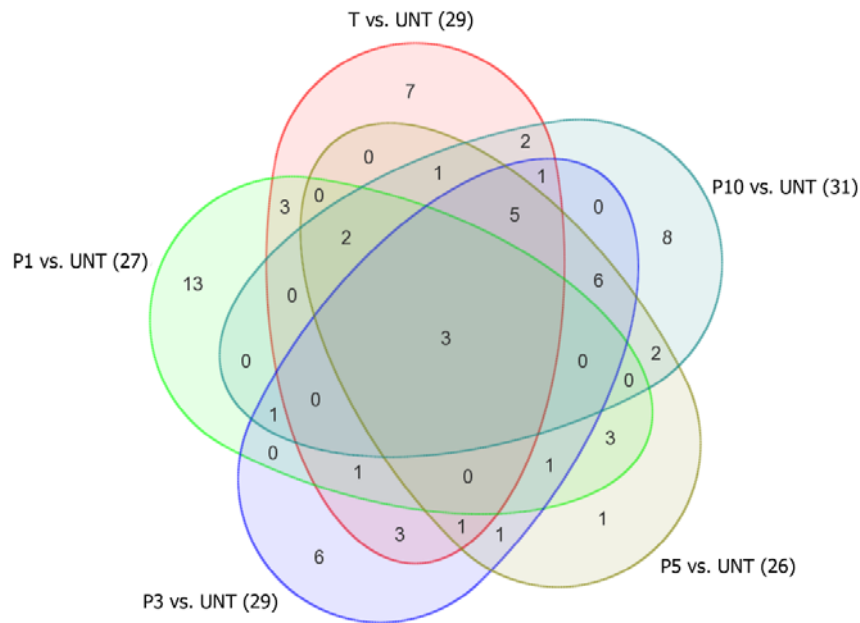
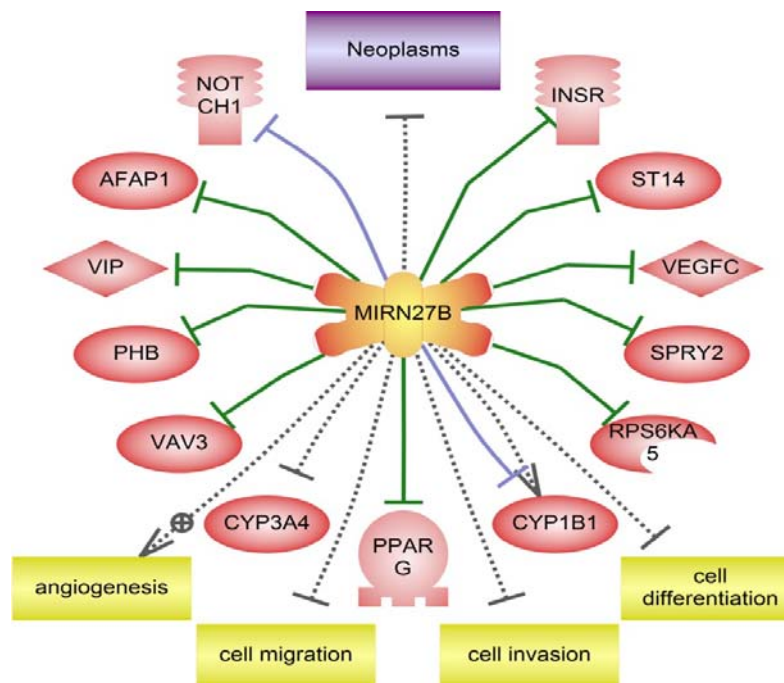Fig. 2. The number of significant up- and down-regulated miRNAs within different passages of HB2.

Fig. 3. Pathway analysis of one miRNAs out of three intersect miRNAs in HB2 that is linked to neoplasms, metastasis or carcinogenesis.
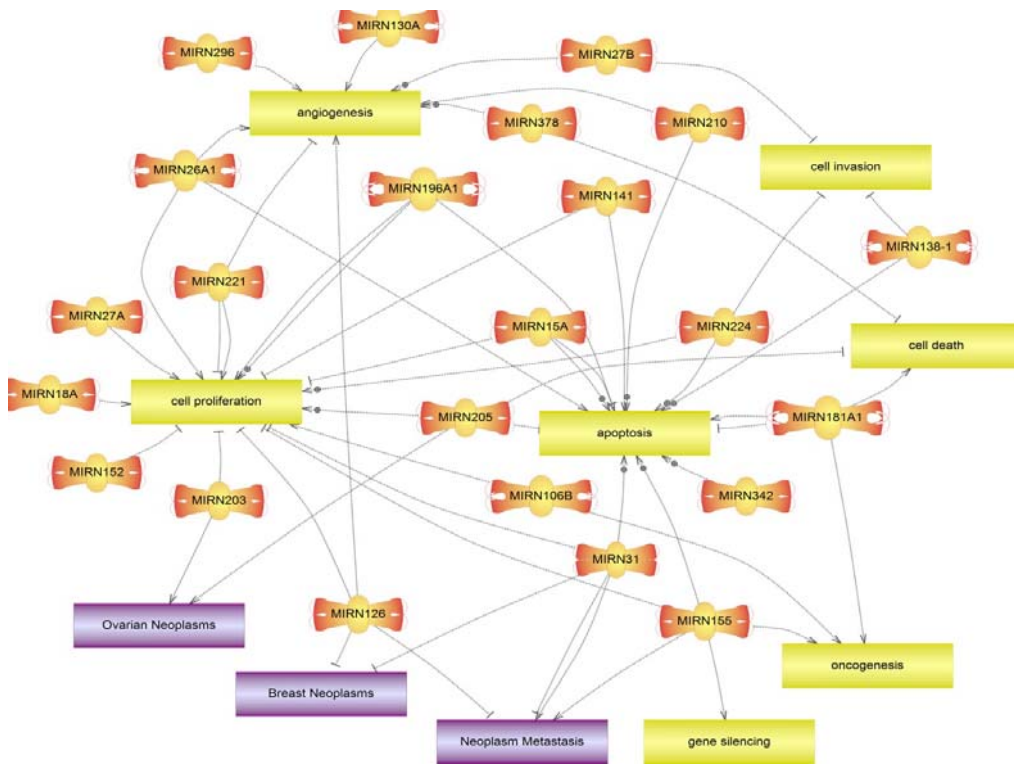
Fig. 4. Pathway analysis of 23 union miRNAs in HB2 that are linked to neoplasms, metastasis or carcinogenesis.
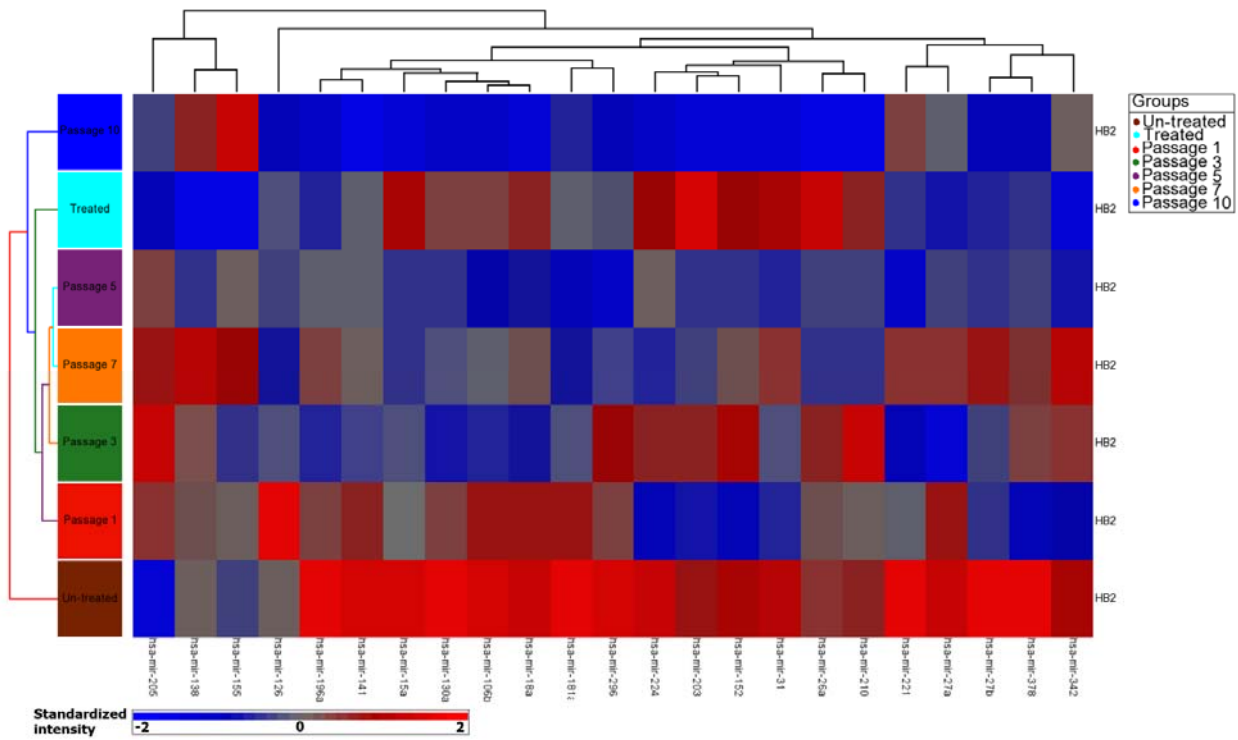

Fig. 5. Heatmap shows significant up- and down-regulated 23 selected union miRNAs in HB2.

MicroRNA Expression Profiles of MDA-MB231 (Highly Aggressive Breast Cancer Cell Line)
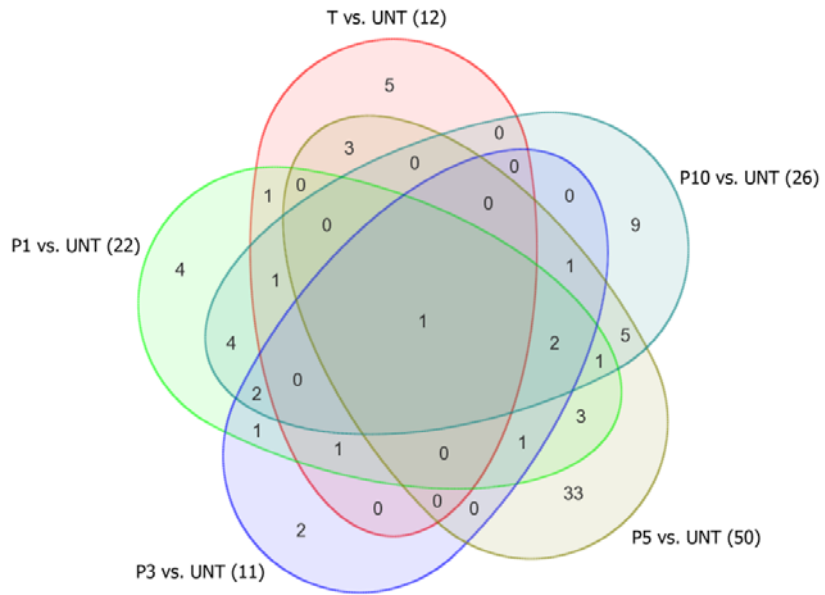

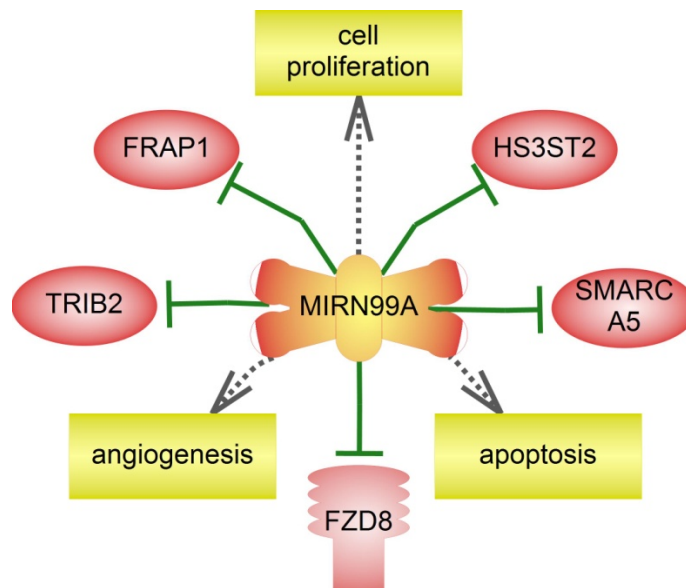Fig. 6. The number of significant up- and down-regulated miRNAs within different passages of MDA-MB231.

Fig. 7. Pathway analysis of one intersect miRNAs in MDA-MB231 that is linked to neoplasms, metastasis or carcinogenesis.
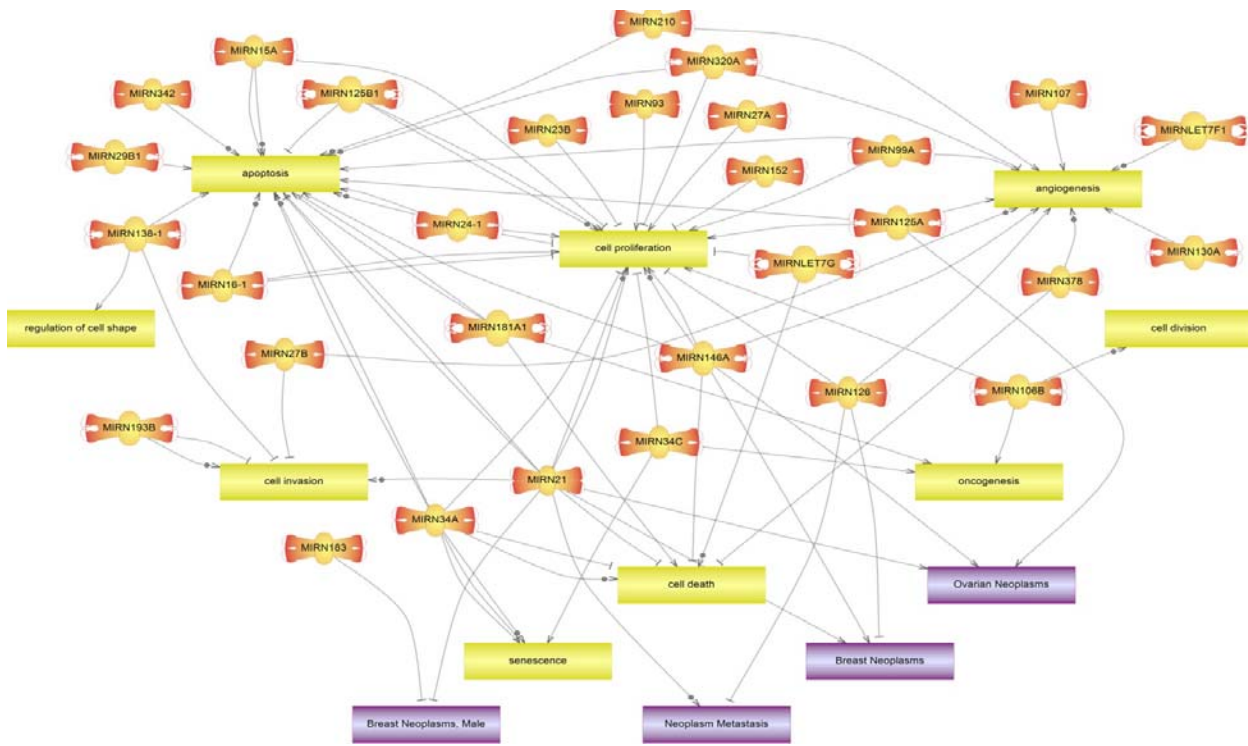
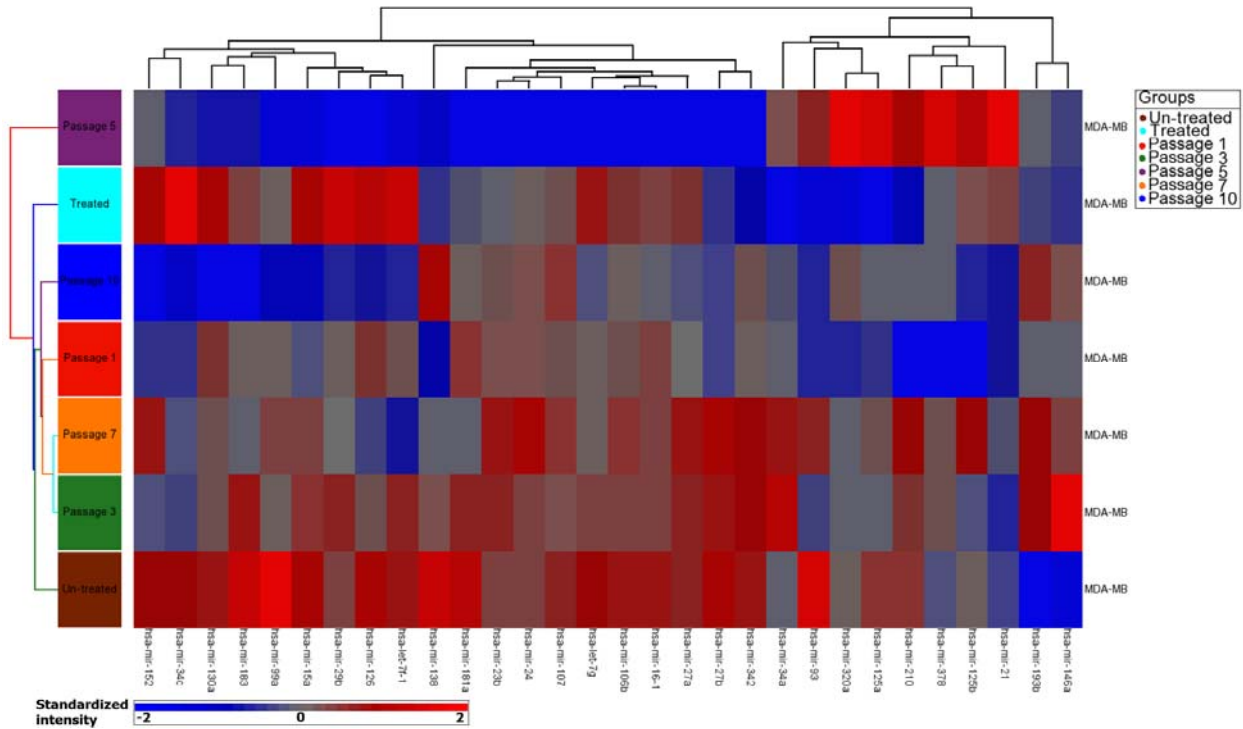Fig. 8. Pathway analysis of 30 union miRNAs in MDA-MB231 that are linked to neoplasms, metastasis or carcinogenesis.

Fig. 9. Heatmap shows significant up- and down-regulated 30 selected union miRNAs in MDA-MB231.

MicroRNA Expression Profiles of SKBR3
(Non-aggressive Breast Cancer Cell Line)
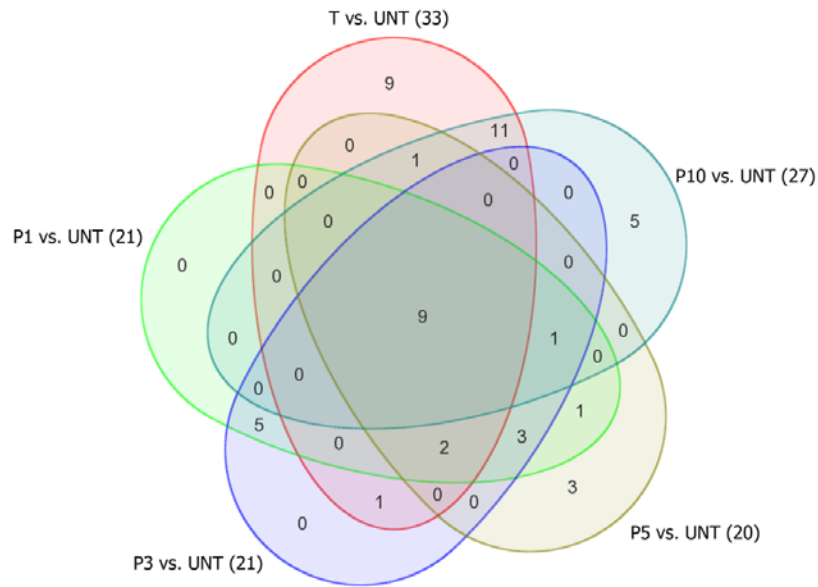
Fig. 10. The number of significant up- and down-regulated miRNAs within different passages of SKBR3.

Fig. 11. Pathway analysis of five miRNAs out of nine intersect miRNAs in SKBR3 that are linked to neoplasms,
metastasis or carcinogenesis.
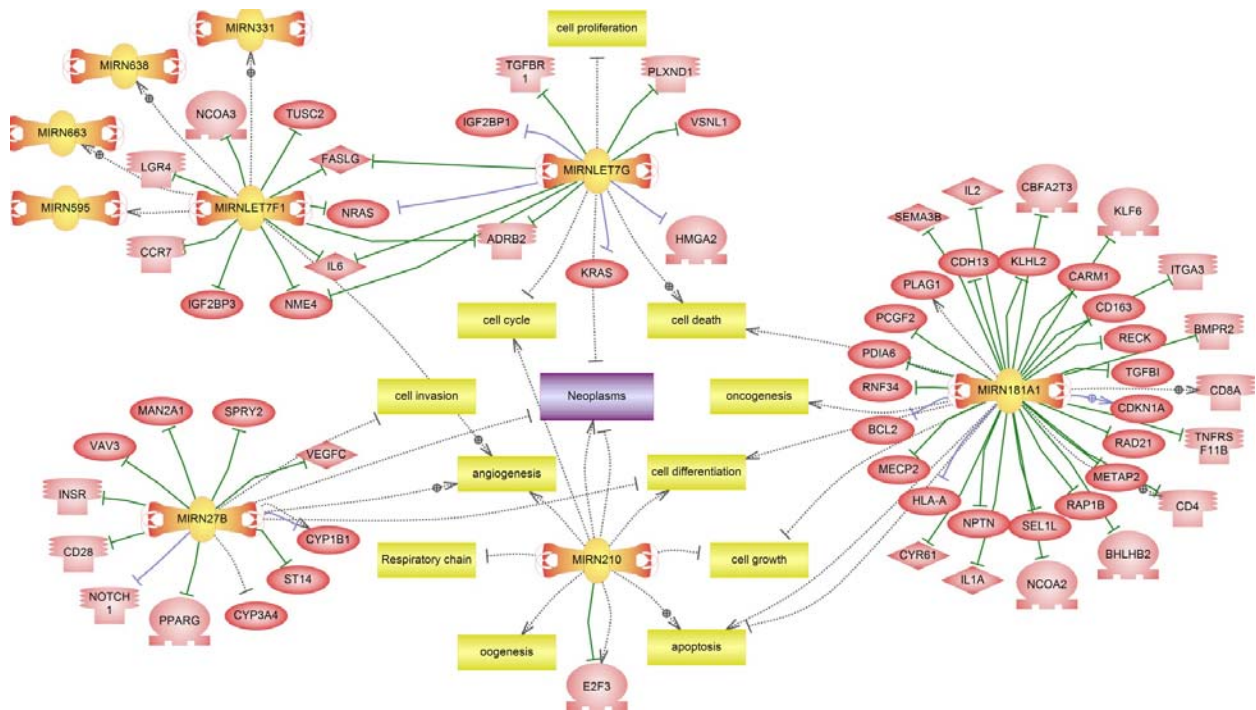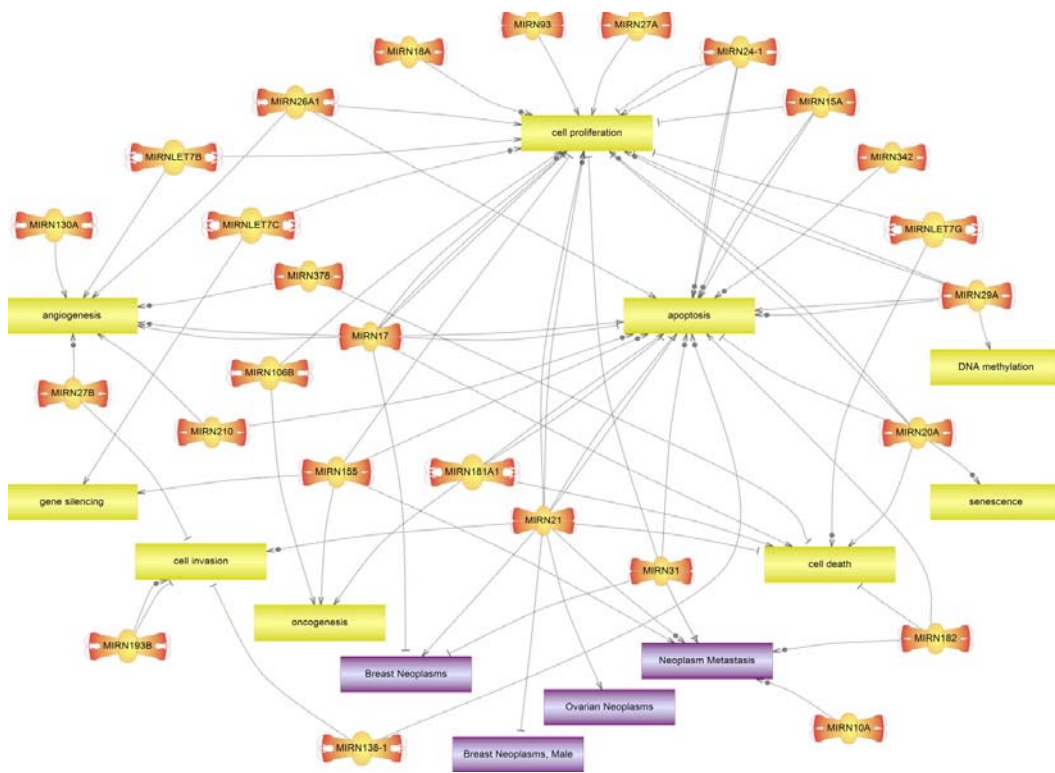
Fig. 12. Pathway analysis of 27 union miRNAs in SKBR3 that are linked to neoplasms, metastasis or carcinogenesis.
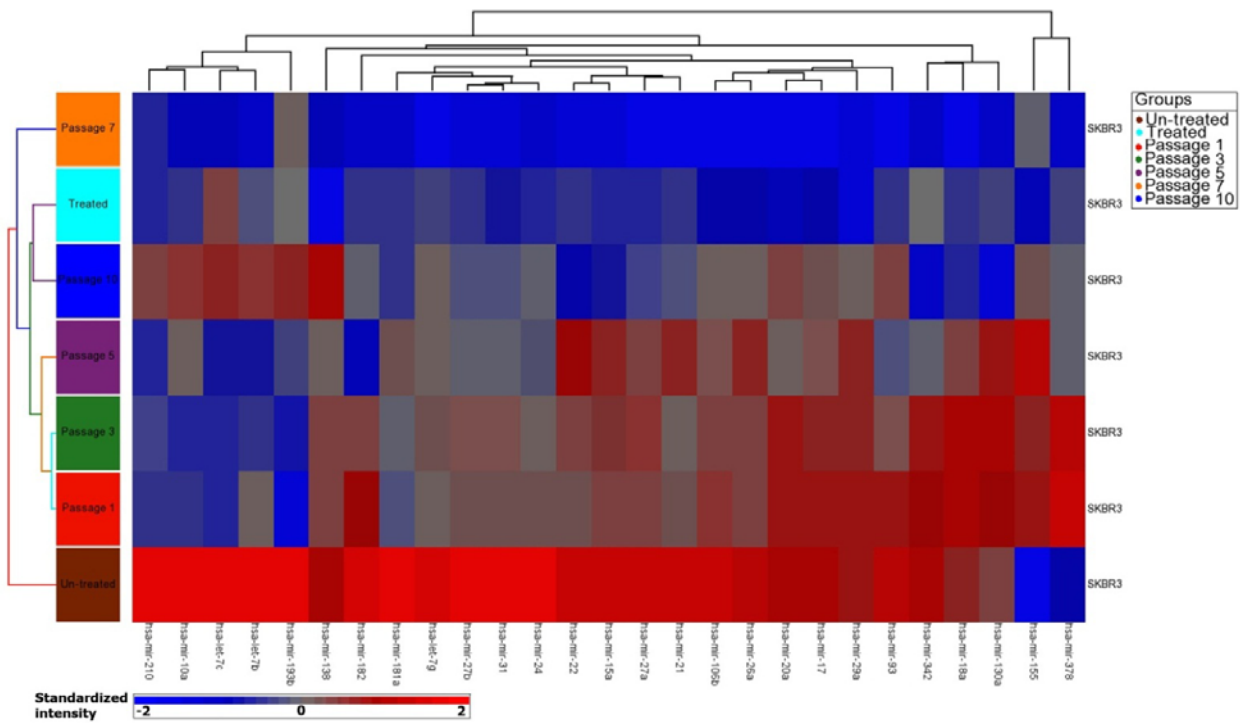
Fig. 13. Heatmap shows significant up- and down-regulated 27 selected union miRNAs in SKBR3.
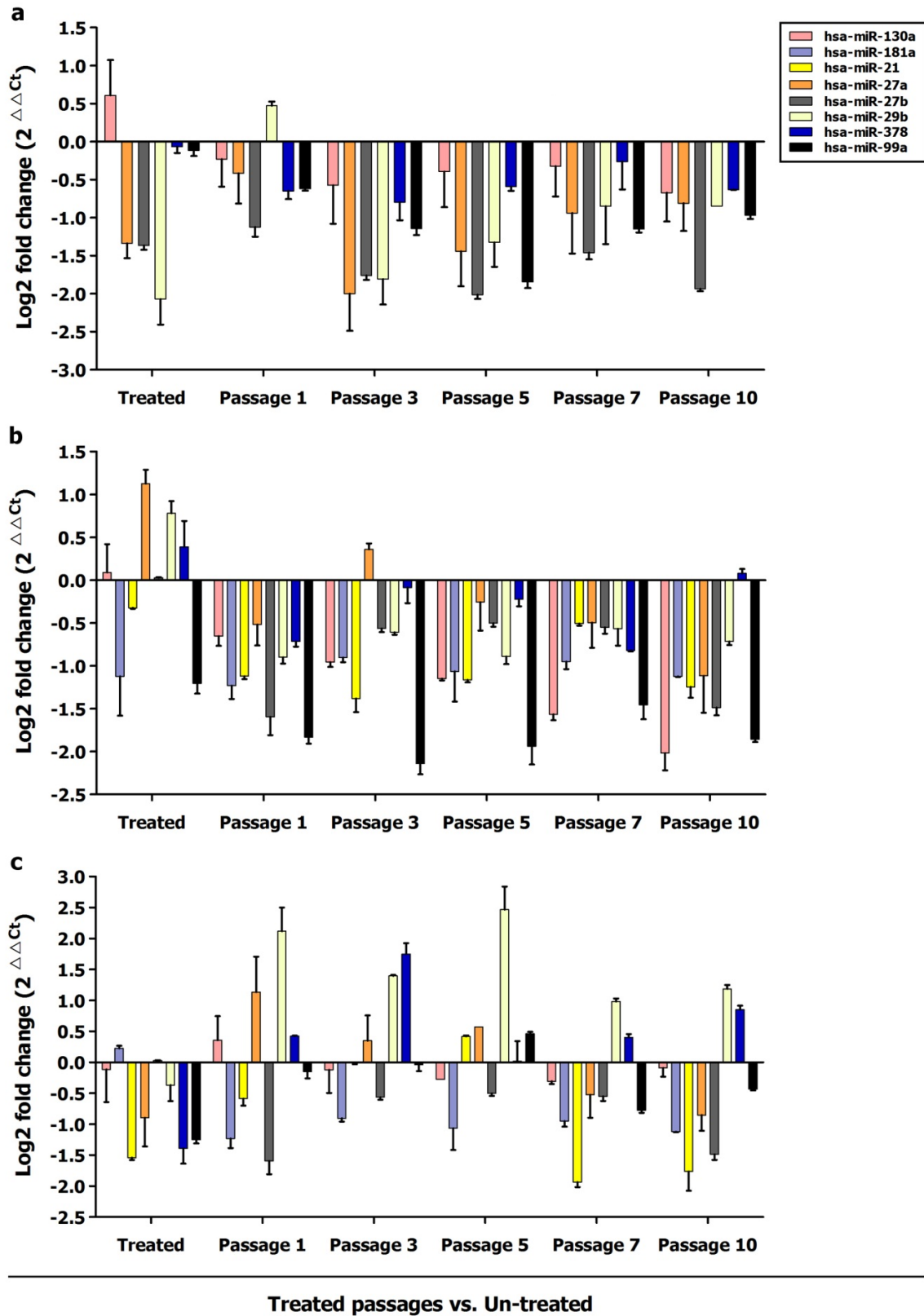
Fig. 14. Expression level of important cancer related miRNAs using qPCR.
a) HB2 cell line. b) MDA-MB231 cell line. c) SKBR3 cell line.

Supplementary Data 4

Proteomics Profiles

HB2 (Breast Epithelial Cell Line)
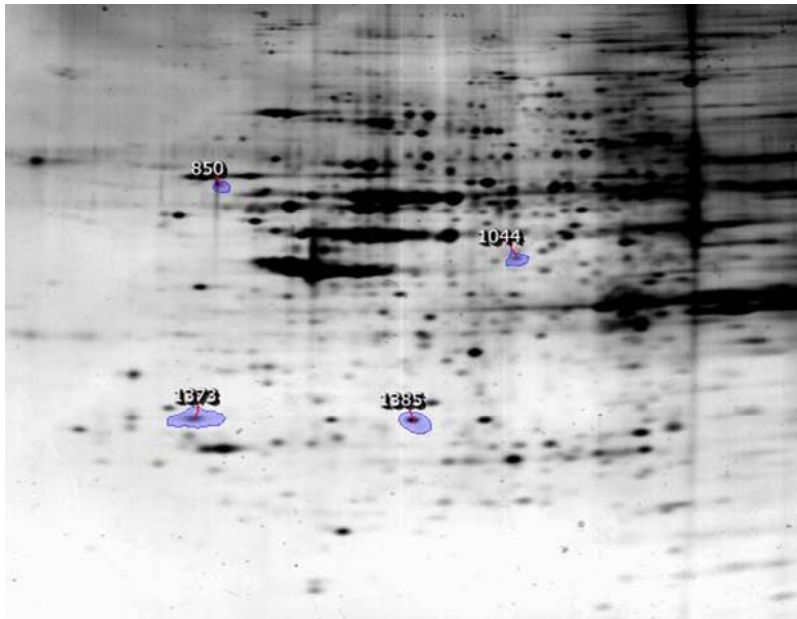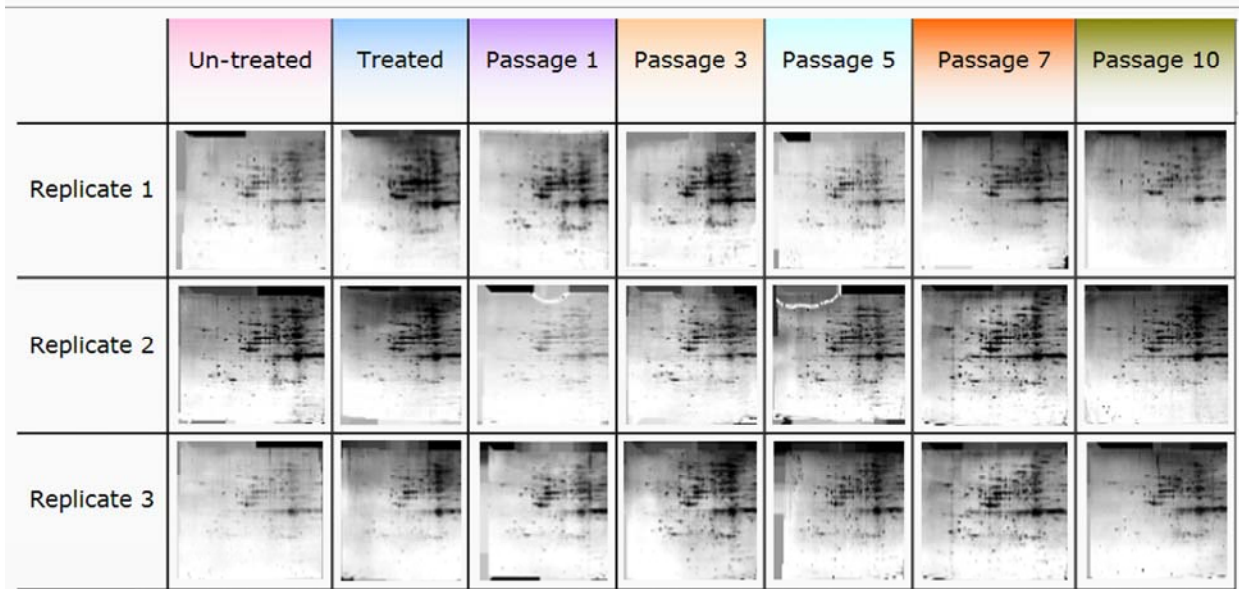
Spot No.	Anova (p)	Fold change	pI
1373	0.022	4.9	5.07
850	0.040	2.0	5.3
1044	0.044	1.6	7.75
1385	0.050	2.4	6.94

Spot picking gel for significant up- and down-regulated intersect proteins.

MDA-MB231 (Highly Aggressive Breast Cancer Cell Line)

	Un-treated	Treated	Passage 1	Passage 3	Passage 5	Passage 7	Passage 10
Replicate 1							
Replicate 2							
Replicate 3							
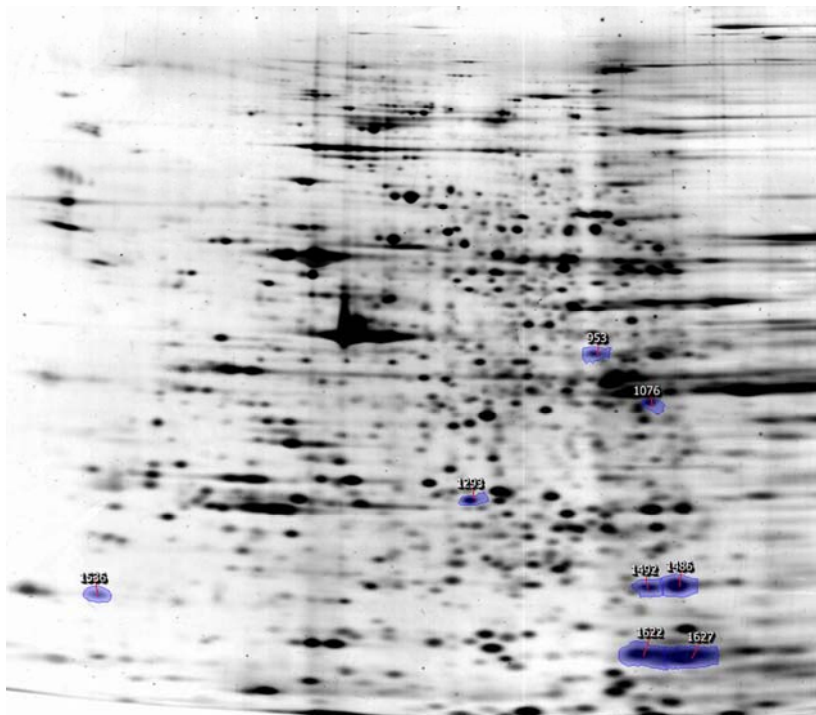
Spot No.	Anova (p)	Fold change	pI
1622	0.007	2.8	8.36
953	0.030	1.6	7.98
1486	0.032	1.9	8.7
1492	0.036	2.0	8.46
1627	0.038	2.6	8.83
1536	0.038	1.6	3.72
1293	0.041	1.7	6.94
1076	0.047	1.9	8.47

Spot picking gel for significant up- and down-regulated intersect proteins.

SKBR3 (Non- aggressive Breast Cancer Cell Line)

	Un-treated	Treated	Passage 1	Passage 3	Passage 5	Passage 7	Passage 10
Replicate 1							
Replicate 2							
Replicate 3							
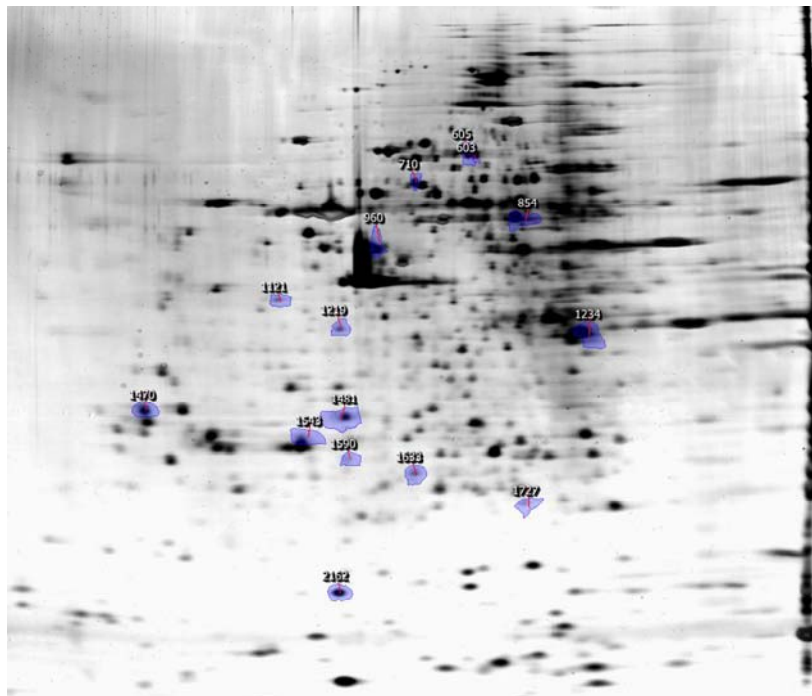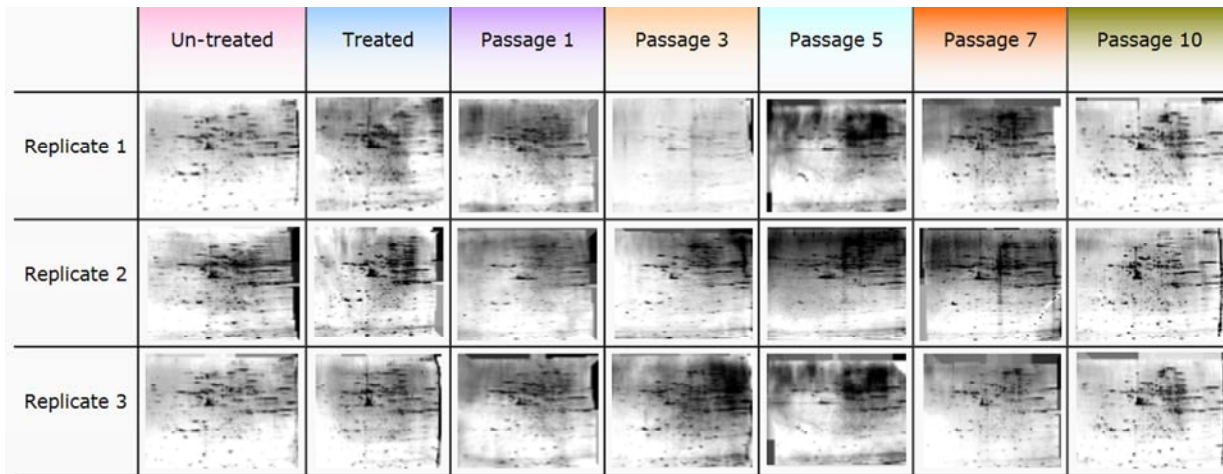
Spots No.	Anova (p)	Fold change	pI
710	0.005	3.1	7.2
1543	0.013	2.2	6.11
2162	0.016	2.6	6.42
1727	0.016	1.9	8.24
603	0.017	7.3	7.76
1470	0.021	6.6	4.67
1121	0.025	3.2	5.87
854	0.030	1.9	8.16
1633	0.030	2.9	7.18
960	0.038	2.7	6.8
1219	0.041	2.2	6.47
1234	0.045	2.0	8.78
605	0.047	3.3	7.7
1590	0.047	2.5	6.53
1481	0.048	2.7	6.53

Spot picking gel for significant up- and down-regulated intersect proteins.
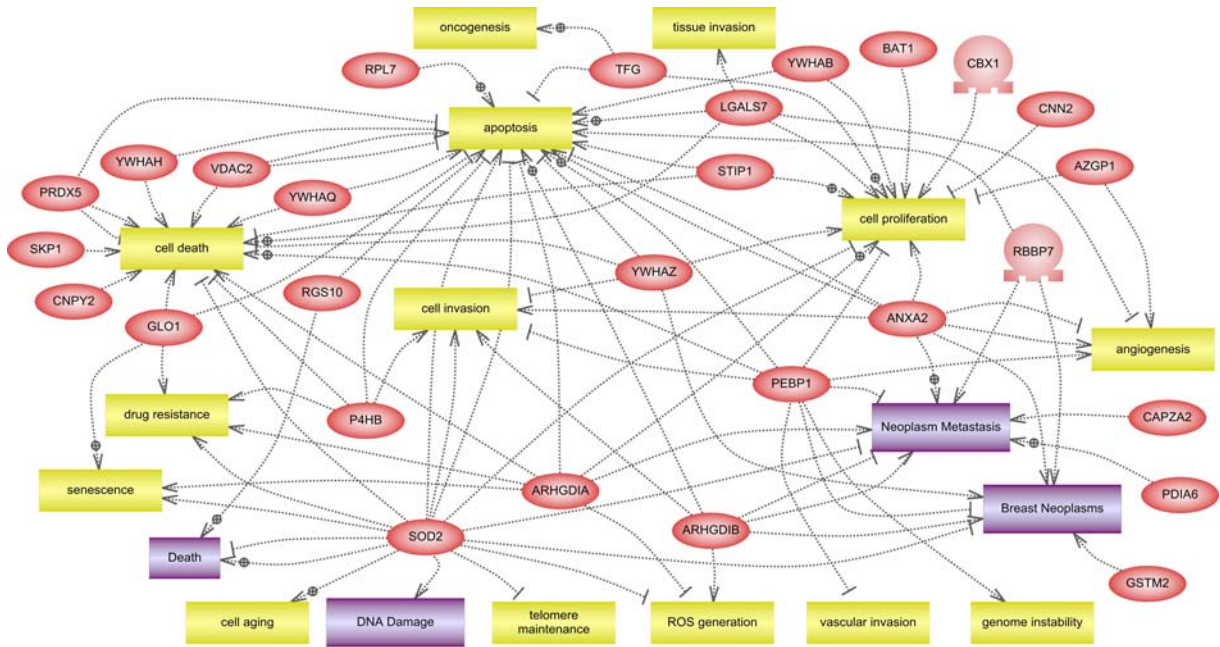
Pathway analysis of 28 out of 41 detected proteins/ isoforms in the three analyzed cell lines (HB2, MDA-MB231, SKBR3) that are linked to neoplasms, metastasis or carcinogenesis.
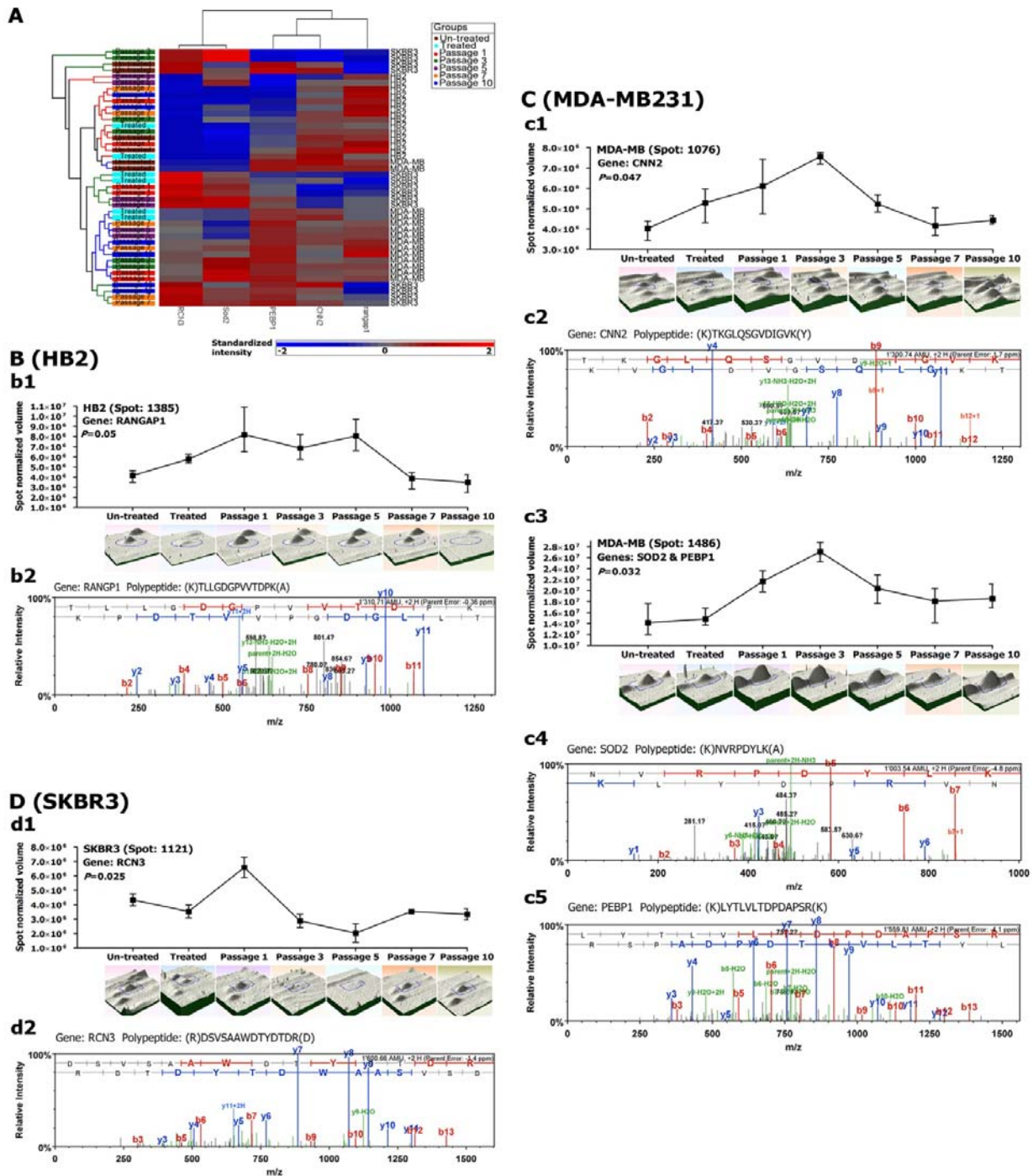
Expression profiles of five genes/proteins that were detected similarly in both transcriptomic and proteomic analysis. A) Expression profiles of five differentially expressed genes in transcriptomic analysis. B-D) Expression profiles of five proteins as significant changing markers in proteomic analysis for the HB2, MDA-MB231 and SKBR3 cell lines. b1-2) Proteomic profile of Rangp1 protein including expression pattern and LC/MS/MS detection. c1-2) Proteomic profile of Cnn2 protein including expression pattern and LC/MS/MS detection. c3-5) Proteomic profiles of Sod2 and Pebp1 proteins including expression pattern and LC/MS/MS detection. d1-2) Proteomic profile of Rnc3 protein including expression pattern and LC/MS/MS detection.

10. Summary and final conclusion

The research for biomarker discovery for the diagnosis and prognosis of cancer patients has been going on for decades, yielding only a few major breakthroughs. Our increasing understanding of cancer biology, including genetic, molecular and cellular mechanisms, and epigenetic background is now providing objectives for the early detection of some malignancies including breast cancer. Such a progress has a direct impact on current activities dedicated to the search for sensitive and specific biomarkers for the early detection and diagnosis of cancers.

In the present study, to find an efficient and high-throughput method for analyzing the methylation profile in breast cancer, we developed a method that allows for the simultaneous detection of multiple targets CpG residue by employing thymidine-specific cleavage on MALDI-TOF MS. In this study, the large-scale analysis was the first high-throughput implementation of the method for quantification of methylation in breast cancer. We were able to conclude that T-specific assay in combination with MALDI-TOF MS is a sensitive and accurate technique for high-throughput methylation analysis.

Using the optimized method on MALDI-TOF silico-chips, we determined quantitative methylation changes of 22 candidate genes in breast cancer tissues. Firstly, we analyzed the methylation status of a total of 42,528 CpG dinucleotides on 22 genes in 96 different paraffin-embedded tissues (48 breast cancerous tissues and 48 paired normal tissues). A two-way hierarchical cluster analysis was used to classify methylation profiles. In this study, 10 hypermethylated genes (*APC, BIN1, BMP6, BRCA1, CST6, ESRb, GSTP1, P16, P21* and *TIMP3*) were identified to distinguish between cancerous and normal tissues according to the extent of methylation. Individual assessment of the methylation status for each CpG dinucleotide indicated that cytosine hypermethylation in the cancerous tissue samples was mostly located near the consensus sequences of the transcription factor binding sites. These hypermethylated genes may serve as biomarkers for clinical molecular diagnosis and targeted treatments of patients with breast cancer.

To achieve a gene panel for developing a breast cancer blood-based test according to the pathologic methylation changes, we quantitatively assessed the DNA methylation proportion of 248 CpG sites per sample (total of 31,248 sites in all analyzed samples) on 10 candidate genes (*APC, BIN1, BMP6, BRCA1, CST6, ESR-b, GSTP1, P16, P21* and *TIMP3*). The number of 126 samples consisting of two different cohorts was used (first cohort: plasma samples from breast cancer patients and normal controls; second cohort: triple matched samples including cancerous tissue, matched normal tissue and serum samples). Circulating cell free methylated DNA of the 8 tumor suppressor genes (TSGs) was significantly higher in patients with breast cancer compared to normal controls. Using eight genes as a panel to develop a blood-based test for breast cancer, a sensitivity and specificity of more than 90% could be achieved in distinguishing between tumor and normal samples. Presented data is promising to design a gene panel and develop a blood-based screening method for breast cancer which relies on pathologic methylation changes. Tissue specific and blood-based methylation markers might provide valuable information as diagnostic and predictive markers for breast cancer, as well as for developing novel targeted therapeutic strategies.

Analyzing the relative telomere length in paired breast cancer and matched normal tissue demonstrated a significant shortening of tumor telomere regions compared to paired adjacent normal tissues. Similarly, telomere lengths were significantly shorter in advanced stage cases, and in those with higher histological grades. In conclusion, our data suggests that shortened telomere length is significantly correlated with breast carcinogenesis. Moreover, promoter hypermethylation of the p16/Rb and p53/p21 pathways showed significant correlation with telomere shortening. The results suggested that inactivation of p16/Rb and/or p53/p21 pathways by hypermethylation may be linked to critical telomere shortening, leading to genome instability and ultimately to malignant transformation. Thus, shortened telomere length and hypermethylation of p53, p21 and p16 promoters might serve as biomarkers in breast cancer.

Epigenetic changes can be pharmacologically reversible and could be a useful target to develop new therapeutic strategies for cancer therapy. Reversal hypermethylation of silenced genes/miRNAs inhibiting malignant phenotypes is increasingly being targeted for cancer therapy

and prevention strategies. In our study, the therapeutic value of detected hypermethylated tumor suppressor genes in different subtypes of breast cancer was assessed after treatment with a demethylating agent (5-aza-2'-deoxycytidine; DAC). Additionally, we investigated multidimensional models to predict effects of DAC treatment at the level of the genome, epigenome and proteome-wide alterations in highly aggressive and non-aggressive subtypes of breast cancer. The results provided early and late effects of this treatment in 3-dimentional follow-up omics model as cancer and non-cancer specific changes that may correlate with particular steps in breast neoplasm including cell proliferations, cell/tissue invasion, oncogenesis, angiogenesis, apoptosis, neoplasm metastasis and senescence. Additionally, beside the activation of epigenetically suppressed TSGs, we also showed significant down-regulation of some miRNAs with oncogenic functions in breast cancer cell lines as well as up-regulation of some other miRNAs with tumor suppressor functions that highlights the usefulness potential of a miRNA-based therapy in breast cancer. Present approach is suggested to be extended for other human solid tumor malignancies, alone or in combination with other treatments such as enforced targeted therapies for miR-29b.

Achieving successful therapy with DAC required the application of repetitive optimal-dose cycles with the regular intervals that may allow efficient incorporation of the drug into the newly synthesized DNA undergoing mitosis during each treatment. Moreover, the obtained results provided a rationale for developing therapeutic strategies based on reactivation of epigenetically silenced genes/miRNAs in breast cancer.

11. Appendix

Short Curriculum Vitae

Last name: Radpour

First name: Ramin

Birthday: 09.12.1977

Education

2011	**Dr. phil. (PhD):** (Grade: Summa Cum Laude (Outstanding)) Department of Biomedicine / Women's Hospital, University of Basel, Switzerland. **Specialty:** *Genetics* • Thesis title: Epigenetic and genomic biomarker discovery in breast cancer.
2002	**Master of Science (MSc):** (Grade: "excellent") Department of Human Genetics, University of Social Welfare and Rehabilitation Sciences (USER), Tehran, Iran. **Specialty:** *Human Genetics* • Thesis title: Molecular study of *PKD1* & *PKD2* genes in the autosomal dominant polycystic kidney disease (ADPKD) by linkage analysis.
2000	**Bachelor of Science (BSc):** (Grade: "very good") AZAD University, Tehran, Iran. **Specialty:** *Molecular Cell Biology*

Professional experience

2007-2011	Department of Biomedicine / Women's Hospital, University of Basel, Switzerland. • Epigenetic and genomic biomarker discovery in breast cancer (PhD thesis project) • Noninvasive prenatal diagnosis (as research partner) • Genetic analysis of colorectal cancer biomarkers (as research partner)
2005-2007	Department of Reproductive Genetics, Royan Institute, Tehran, Iran. • Molecular analysis and mutation screening of *CFTR* gene in Iranian CBAVD patients (as researcher and project director) • Association analysis of CAG and GGN repeats of androgene receptor gene (AR) with idiopathic male infertility and cryptorchidism (as research partner)
2003-2005	Instructor at the Faculty of Health and Therapeutics, Medical Army University, Tehran, Iran. (as military service)
2000-2003	Department of Human Genetics, University of Social Welfare and Rehabilitation Sciences (USER), Tehran, Iran. • Molecular study of *PKD1* & *PKD2* genes in several Iranian families with autosomal dominant polycystic kidney disease by linkage analysis (MSc thesis project)

Awards and honors

2010	**The European Human Genetics (ESHG) Conference 2010** (Gothenburg, Sweden). Candidate for young investigator award.

2010 **The European Human Genetics (ESHG) Conference 2010** (Gothenburg, Sweden). Candidate for young investigator award.

2007 **8th Annual Royan International Research Award** (reproductive biomedicine and stem cell research) (Tehran, Iran). Selected as Best National Researcher.

2006 **12th Annual Razi Research Festival on Medical Sciences** (Scientific award of ministry of health, treatment and education of medical sciences) (Tehran, Iran). The winner of young scientists as FIRST GRADE in the field of basic sciences.

2006 **8th Khwarizmi International Youth Award** (Scientific award of ministry of education) (Tehran, Iran). The winner of best scientists as THIRD GRADE in the field of basic science research.

2003 **10th University Student's Year Book Award** (Tehran, Dec. 2003). Selected as outstanding youth author in the field medicine and basic science. The book title: Genetic Terminology (ISBN: 964-460-391-5).

Publications related to the PhD work

- *Radpour R*, Barekati Z, Kohler C, Zhang B, Lv Q, Bürki N, Diesch C, Bitzer J, Zheng H, Schmid S, Zhong XY. **Hypermethylation of tumor suppressor genes involved in critical regulatory pathways for developing a blood-based test in breast cancer.** PLoS One. 2011 Jan 24;6(1):e16080.

- *Radpour R*, Barekati Z, Haghighi MM, Kohler C, Asadollahi R, Torbati PM, Holzgreve W, Zhong XY. **Correlation of telomere length shortening with promoter methylation profile of p16/Rb and p53/p21 pathways in breast cancer.** Mod Pathol. 2010 May;23(5):763-72.

- Barekati Z, *Radpour R*, Kohler C, Zhang B, Toniolo P, Lenner P, Lv Q, Zheng H, Zhong XY. **Methylation profile of TP53 regulatory pathway and mtDNA alterations in breast cancer patients lacking TP53 mutations.** Hum Mol Genet. 2010 Aug 1;19(15):2936-46.

- *Radpour R*, Barekati Z, Kohler C, Holzgreve W, Zhong XY. **New trends in molecular biomarker discovery for breast cancer.** Genet Test Mol Biomarkers. 2009 Oct;13(5):565-71. *Review*.

- *Radpour R*, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, Holzgreve W, Lefkovits I, Zhong XY. **Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis.** J Proteome Res. 2009 Nov;8(11):5264-74.

- *Radpour R*, Kohler C, Haghighi MM, Fan AX, Holzgreve W, Zhong XY. **Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array.** Oncogene. 2009 Aug 20;28(33):2969-78.

- *Radpour R*, Haghighi MM, Fan AX, Torbati PM, Hahn S, Holzgreve W, Zhong XY. **High-throughput hacking of the methylation patterns in breast cancer by in vitro transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip.** Mol Cancer Res. 2008 Nov;6(11):1702-9.

Presentations at congresses related to the PhD work

Oral presentations

1. European Society of Human Genetics (ESHG) conference, June 12-15, 2010, Gothenburg, Sweden. **Title: Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array**.

2. 16th Annual EuCC Symposium for Clinical and Experimental Oncology, May 15, 2009, Freiburg, Germany. **Title: Quantitative assessment of methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array**.

Poster presentations

1. European Society of Human Genetics (ESHG) conference, May 28-31, 2011, Amsterdam, Netherlands. **Title: Complete transcriptomic, epigenomic and proteomic signature upon demethylation treatment of human breast cancer subtypes approached by multi-platform 'omics analyses.**

2. 17th Annual EuCC Symposium for Clinical and Experimental Oncology, May 7, 2010, Strasbourg, France. **Title: A model prediction panel based on methylation profile of *APC*, *BIN1*, *BMP6*, *BRCA1*, *CST6*, *ESR-b*, *GSTP1*, *P16*, *P21*, and *TIMP3* genes in circulation, paired tumor and adjacent tissues from breast cancer patients**.
Complete transcriptomic, epigenomic and proteomic signature upon demethylation treatment of human breast cancer subtypes approached by multi-platform 'omics analyses