Inaugural Dissertation

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Master Sci. Yasser Riazalhosseini

Born in Tehran, Iran

Oral-examination: June 17, 2010

Identification of a DNA methylation signature and distinct microRNA variants in breast cancer

This work was carried out in the Division of Functional Genome Analysis

at the German Cancer Research Center (DKFZ)

Head of division: Dr. Jörg Hoheisel

Referees:

Prof. Dr. Harald Hermmann-Lerdon

Prof. Dr. Jörn Walter

Thesis declaration

I hereby certify that this thesis was written by me. No sources of information other than those listed in the references were used.

Heidelberg, April 27, 2010

……

Yasser Riazalhosseini

Acknowledgement

I am grateful to;

- My parents who were the first teachers in my life and Maria Shahmoradgoli for her constant support throughout our life.
- Jörg Hoheisel for giving me the opportunity to carry out this research towards my doctorate, his invaluable ideas and support, and the unique friendly atmosphere of his group.
- Verena Beier for the introduction to the microarray field and technical guidance.
- Prof. Dr. Harald Hermmann-Lerdon (DKFZ) and Prof. Dr. Jörn Walter (Universität des Saarlandes), members of the advisory committee of my PhD thesis, for their insightful advice.
- Sandeep Kumar Botla and Pedro de Souza Rocha Simonini for their tremendous help in the laboratory work and also in the preparation of the graphical images.
- Amin Moghaddas Gholami (DKFZ) for his broad and valuable bioinformatic analysis.
- Dr. Achim Breiling and Prof. Dr. Frank Lyko (DKFZ) for the ChIP analysis and our fruitful collaboration.
- Michaela Schanne for the translation of the Summary of my thesis from English to German.
- Evgenij Moskalev, Neeme Tõnisson, Michael Böttcher, Armin Pscherer, Özgür Sahin and Doris Mayer for helpful discussions.
- Achim Stephan, Marita Schrenk, Azin Jahangiri for their technical support.
- Christoph Schröder, Jorge Soza Ried and Mahmoud Mostafa Youns for their support and good friendship.
- All my former and current colleagues in the Division of Functional Genome Analysis and my friends at the DKFZ.

Table of Contents

Thesis declaration iii

Acknowledgement iv

Table of Contents v

Abbreviations ix

Summary xi

Zusammenfassung xiii

1 Introduction 1

1.1 Cancer 1

1.1.1 Breast Cancer 2

1.2 Epigenetics 4

1.2.1 DNA methylation 5

1.2.2 Covalent histone modifications 5

1.2.3 microRNAs (miRNAs) 7

1.3 Epi-Genetics of Cancer 7

1.3.1 Abnormal DNA methylation patterns in cancer 9

1.3.2 miRNAs and cancer 14

1.3.3 Epi-Genetics of breast cancer 14

1.3.4 Aim 17

2 Material and Methods 18

2.1 Materials 18

2.1.1 Instruments 18

2.1.2 Chemical reagents, enzymes and general materials 19
2.1.3 Cell culture 21
2.1.4 Kits 21
2.1.5 Media, solutions and buffers 22
2.1.6 Antibodies 24
2.1.7 Plasmids and synthetic RNAs 24
2.1.8 qRT-PCR primers 25
2.1.9 *Genomic* primers for the amplification of CGIs from genomic DNA 25
2.1.10 *SoBi* primers for the amplification of CGIs from deaminated DNA 27
2.1.11 Primers used for ChIP analysis of miR-375 locus 28
2.1.12 Primers used for isolation of miR-375 target sites 29
2.1.13 Cell lines 29
2.2 Methods 29
2.2.1 General molecular biology methods 29
2.2.2 Processing of tissue samples for DNA/RNA isolation 37
2.2.3 Selection of CGIs and designing PCR primers to isolate them from deaminated DNA 38
2.2.4 Preparation of the control samples 40
2.2.5 Sodium bisulfite treatment of DNA samples 43
2.2.6 Isolation of selected CGIs from patients DNA 45
2.2.7 Labeling of DNA by random priming 46
2.2.8 Microarrays for DNA methylation analysis 48
2.2.9 Data analysis 54
2.2.10 Bisulfite-sequencing 56
2.2.11 Cell culture 57

2.2.12 Chromatin Immunoprecipitation (ChIP) assay 59
2.2.13 Expression arrays 59
2.2.14 Functional analysis 62
2.2.15 Statistical analysis 66
3 Results 67
3.1 Microarray based analysis of DNA methylation patterns 67
3.1.1 Developing microarrays for DNA methylation analysis 67
3.1.2 Optimization of probe length for reliable array performance 69
3.1.3 Selection of highly discriminative probes 75
3.2 Screening DNA methylation patterns in tissue samples 75
3.2.1 Isolation of DNA from tissue samples 75
3.2.2 PCR-amplification of CGIs from DNA samples 76
3.2.3 Microarray-based analysis of methylation patterns in tissue samples 77
3.2.4 Bisulfite sequencing data 85
3.2.5 Re-activation of methylated genes following decitabine treatment 89
3.3 Identification of miRNAs involved in ERα activation 90
3.3.1 miR-375 expression is increased in ERα-positive breast cancer cell lines 90
3.3.2 Reciprocal regulation between miR-375 and ERα 92
3.3.3 miR-375 inhibition reduces the proliferation rate of breast cancer cells 93
3.3.4 Epigenetic marks determine the transcriptional state of the miR-375 locus 94
3.3.5 Transcriptional repressors bind to the miR-375 locus 95
3.3.6 RASD1 is a functional target of miR-375 99
3.3.7 Regulatory connection between ERα and RASD1 102
4 Discussion 104

4.1 DNA methylation analysis 104

4.1.1 Development of microarrays 104

4.1.2 DNA methylation patterns of breast tumors 107

4.1.3 Genes affected by hypermethylation in tumors 108

4.1.4 DNA methylator event in breast cancer? 112

4.1.5 DNA methylation signature for diagnosis of breast cancer 114

4.1.6 DNA methylation patterns at signature loci 116

4.2 Identification of miRNAs involved in ERα activation 118

4.2.1 Epigenetic marks at miR-375 locus 119

4.2.2 Proposed model for RASD1 and miR-375 effects on ERα 121

5 Outlook 123

6 References 125

7 Appendix 136

7.1 Supplementary Table S1 136

7.2 Supplementary Table S2 137

7.3 Supplementary Table S3 137

7.4 Supplementary Table S4 138

7.5 Supplementary Table S5 138

7.6 Publications 142

Abbreviations

ATP	Adenosine triphosphate
BSA	Bovine serum albumin
bp	Base pair
cDNA	Complementary DNA
CCD	Charge-Coupled Device
CGI	CpG island
ChIP	Chromatin immunoprecipitation
dNTP	Desoxynucleotide triphosphate
DNA	Desoxyribonucleic acid
DNMT	DNA methyltransferase
dUTP	Desoxyuridine triphosphate
ERα	Estrogen receptor α
ERE	Estrogen responsive element
E. coli	Escherichia coli
FDR	False discovery rate
kb	Kilobase
miRNA	microRNA
mRNA	Messenger RNA
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PVDF	Polyvinylidene fluoride

PBS	Phosphate buffered saline
PcGs	Polycomb group proteins
qRT-PCR	Quantitative real-time polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RT	Room temperature
RPM	Rotations per minute
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
siRNA	Small interfering RNA
TSS	Transcription start site
UTR	Untranslated region
UV	Ultraviolet
v/v	Volume per volume
Vol	Volume
W/v	Weight per volume

Summary

Deregulation of factors involved in the epigenetic regulation of gene expression is a hallmark of cancer. In this dissertation, DNA methylation and microRNA (miRNA) expression, as two components implicated in epigenetic regulation, were studied in breast cancer.

In the first study, the establishment of a DNA methylation signature for breast cancer was aimed at, based on the early occurrence and stability of abnormal DNA methylation patterns in tumors. Specialized oligonucleotide microarrays were developed and optimized to screen DNA methylation patterns of the candidate loci associated with breast cancer. The assay was applied to analyze DNA methylation patterns of breast cancer and control samples. The profiling data was subjected to multiple bioinformatic analyses in order to identify a DNA methylation signature with appropriately high potential for diagnosis. Methylation patterns of CpG sites associated with two genes, *SFRP2* and *GHSR*, were identified as promising markers.

Up-regulation and activation of estrogen receptor α (ERα) signaling is a common feature of almost 70% of breast cancers. However, our understanding of the molecular mechanisms underlying deregulation of this signaling pathway is scarce. In the second study, based on miRNA profiling data of human mammary cell lines, miR-375 was identified as an overexpressed miRNA in ERα-positive cells. Analysis of the miRNA locus revealed that miR-375 overexpression is mainly caused by the loss of epigenetic marks, such as H3K9me2 and local DNA hypomethylation, which, in turn, triggers the dissociation of the transcriptional repressor CCCTC-binding factor (CTCF) from the promoter and enables interactions of ERα with regulatory regions of miR-375. Inhibition of miR-375 in ERα-positive MCF7 cells resulted in reduced ERα activation and cell proliferation. In addition, *RASD1*, an estrogen inducible gene, was identified as a miR-375 target mediating miR-375

effect on ERα. These data indicate the existence of a positive feedback regulation between ERα and miR-375.

The methylation signature identified shows a promising potential to be applicable for diagnosis/risk assessment of breast cancer. Furthermore, our findings from miRNA study provide significant insight into the deregulation of ERα pathway, which may open new therapeutic strategies for breast cancer.

Zusammenfassung

Krebserkrankungen weisen meist eine veränderte epigenetische Regulation der Genexpression auf. In dieser Dissertation werden sowohl DNA-Methylierung als auch die Expression von microRNA (miRNA) als zwei Komponenten der epigenetischen Regulation in Brustkrebs untersucht.

Veränderte DNA-Methylierungsmuster tauchen in Tumoren gewöhnlich früh auf und haben eine hohe Stabilität. Daher hatte die erste Studie das Ziel, eine für Brustkrebs spezifische Signatur der DNA-Methylierung zu identifizieren. Hierzu wurden spezifische Oligonukleotidarrays entwickelt und optimiert, um DNA-Methylierungsmuster ausgewählter Kandidaten-Loci zu untersuchen, die mit Brustkrebs assoziert werden. Um derartige Muster mit einem ausreichend hohen Potential für eine Diagnose auszuwählen, wurden Tumorproben mit gesunden Kontrollen verglichen und einer Vielzahl von bioinformatischen Analysen unterzogen. Hierbei wurden Methylierungsprofile in zwei CpG-Bereichen identifiziert, die mit den Genen *SFRP2* und *GHSR* assoziiert sind und ein Potenzial für eine Diagnosestellung besitzen.

Eine erhöhte Expression und Aktivierung der Estrogenrezeptor α (ERα)-Signalkaskade ist ein gemeinsames Merkmal von fast 70% aller Tumore der Brust. Die zugrunde liegenden molekularen Mechanismen dieser Signalkaskade sind bislang wenig verstanden. Daher wurden in einer zweiten Studie humane Mammakarzinomzelllinien mittels eines miRNA-*Profilings* untersucht. Hierbei wurde eine deutliche Überexpression von miR-375 in Erα-positiven Zellen entdeckt. Eine Analyse des miRNA-Lokus zeigte, dass die Überexpression hauptsächlich durch den Verlust von epigenetischen Markierungen, wie beispielsweise H3K9me2, sowie eine lokale DNA-Hypomethylierung verursacht werden. Die Hypomethylierung löst eine Dissoziation des Transkriptionsrepressors CTCF vom

Promotor aus und ermöglicht Interaktionen von Eα mit den regulatorischen Reg ionen von miR-375. Eine Inhibierung von miR-375 in ERα-positiven MCF7 Zellen resultierte in einer reduzierten ERα-Aktivierung und Zellproliferation. Zusätzlich wurde festgestellt, dass das estrogeninduzierbare Gen *RASD1* als ein Ziel-Gen von miR-375 den Effekt auf Erα vermittelt. Diese Daten dokumentieren eine vorhandene positive Rückkopplung zwischen ERα und miR-375.

Das in dieser Arbeit identifizierte spezifische Methylierungs-Muster zeigt ein vielversprechendes Potenzial für die Anwendung im Rahmen von Brustkrebs-Diagnose und –Prognose. Die Ergebnisse der miRNA-Studie ermöglichen neue Einblicke in die Regulation der ERα-Signalkaskade und könnten neue Möglichkeiten in der Entwicklung theapeutischer Strategien zur Brustkrebsbehandlung eröffnen.

1 Introduction

1.1 Cancer

Cancer is a highly complex and heterogeneous disease. Currently, it is the second leading cause of death after cardiovascular diseases (1). Different types of cancers are named based on the part of the body from which they originate. Although, different cancer types are clinically distinct, they all develop from uncontrolled cell growth resulting in the formation of an unstructured tissue mass, called tumor. Tumors are classified into two categories based on their degree of aggressive growth. Those that grow locally without invading adjacent tissues are classified as benign. Others that invade nearby tissues are malignant tumors. Malignant tumors are able to infiltrate into the blood stream, migrate to the areas distant from their place of origin and seed tumor colonies in other tissues. This behavior is known as metastatic potential and describes the aggressiveness of the corresponding cancer. The metastatic cancers are responsible for 90% of cancer related deaths (2).

The comparatively disorganized tissue architecture of tumors points towards the fact that cancer is a disease of malfunctioning cells. Tumor formation can be originated from transformation of a single cell within the organism. The transformation process can be facilitated by cancerous agents (carcinogens; like toxins, free radicals, UV exposure and some viruses such as human papilloma virus (HPV)) that cause functional mutations in DNA (2). Mutations can also happen spontaneously at an estimated rate of about 10^{-6} mutations per gene per cell division. However, human body has been endowed with various safeguard mechanisms to protect itself from serious mutations and cellular defects. For example, the DNA repair system repairs mutations that might happen during replication. Likewise, the cell division process is highly controlled and can be arrested when necessary. Furthermore, in case of a fundamental damage to DNA -which could not be repaired- or a

pivotal cellular malfunction, the cell undergoes the programmed cell death (apoptosis) process (3).

Many cancers arise when the above mentioned self-protecting mechanisms of the cell do not work properly, mostly due to dysregulation of key genes involved in such processes. Based on their function, these genes are categorized into two classes: oncogenes and tumor suppressor genes. Oncogenes are usually growth-promoting genes that can contribute to the development of cancer by helping cells to escape several growth controlling mechanisms. Unlike oncogenes, tumor suppressor genes are necessary for cells to maintain important control barriers, and balance between cell proliferation and death (2).

1.1.1 Breast Cancer

Breast cancer is the most common malignant tumor in women worldwide. According to estimations by the World Health Organization (WHO), breast cancer leads to about 519,000 deaths per year in the world, and belongs to the top ten mortal diseases (4). Besides the large number of cases, breast carcinoma is a very heterogeneous cancer in terms of histopathological and clinical classification, genetic and genomic background of the tumor mass and clinical outcomes (5).

1.1.1.1 Estrogen Receptor α and breast cancer

Although it is a heterogeneous disease, two-thirds of breast cancers share the common feature of being dependent on the presence and interaction of estrogen with its nuclear receptor, the Estrogen Receptor-α (ERα), for their growth (5, 6). Approximately 70% of invasive breast cancers express ERα in their actively proliferating cells. It has become evident that ERα is up-regulated in luminal mammary epithelial cells during early steps of tumorigenesis process and its overexpression is an important stimulatory factor for proliferation of mammary cells, leading to an un-controlled cell division and eventually to

tumor development (6, 7). ERα functions as a ligand-dependent transcription factor. It acts mainly by regulating the expression of target genes whose promoters contain a specific sequence known as estrogen-responsive element (ERE) (8). The obvious role of ERα signaling in orchestrating the expression of genes involved in growth-related pathways has established ERα as an important therapeutic target breast cancer treatment (6). However, there is a lack of knowledge about the molecular mechanisms underlying the up-regulation of ERα in breast cancer.

1.1.1.2 Breast cancer types

Almost all invasive breast tumors are adenocarcinoma (arisen from secretory epithelia) (96.9%). Among the adenocarcinoma, the largest subgroup is comprised of the invasive or infiltrating ductal carcinoma (IDC) with about 80% of all breast carcinoma, followed by the infiltrating lobular carcinoma (ILC) (Fig. 1). Less common types of breast cancer include mucinous, tubular, papillary, medullary or those not otherwise specified (NOS) that arise only to very low percentages (2.6%, 1.6%, 0.4%, 0.7%, and 1.1%, respectively). Non-invasive breast cancers include the ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). In these types, cancer cells are inside the ducts or lobules and have not invaded the surrounding tissues. These types may develop to invasive forms or predispose the affected person to develop an invasive breast cancer (9).

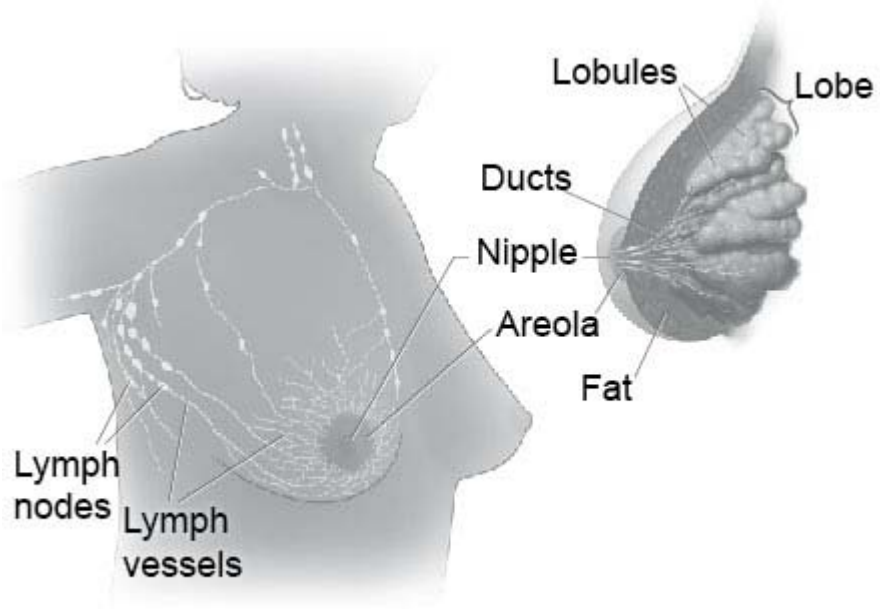


Figure 1. Anatomy of the breast. IDC originates from cells of ducts (milk passage), whereas ILC starts in milk- producing glands in the lobules (Website of National Cancer Institute; www.cancer.gov).

1.2 Epigenetics

Epigenetics was first defined by Conrad Hal Waddington as "the causal interactions between genes and their products, which bring the phenotype into being" (10). This definition initially referred to understanding chromatin structure and its impact on gene function. The current definition of epigenetics is "the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence". Epigenetic mechanisms that govern these expression patterns have been divided into four main categories: DNA methylation, covalent histone modifications, nucleosome remodeling factors and non-coding RNAs including microRNAs (miRNAs) (11).

1.2.1 DNA methylation

DNA methylation is the most-studied epigenetic factor referring to the modification of cytosine residues by methylation at their carbon-5 position. In mammalian genomes, this modification is primarily occurs on cytosine residues that are located 5' to guanine- i.e. CpG dinucleotides (12). DNA methylation is catalyzed by DNA methyltransferases (DNMTs). In human, at least, three DNMTs are involved in establishment and maintenance of DNA methylation patterns; *de novo* DNA methyltransferases DNMT3A and DNMT3B and maintenance DNA methyltransferase DNMT1 (13, 14). CpG dinucleotides are not equally distributed across the human genome but are concentrated in distinct CpG-rich regions called “CpG islands” (CGIs) and in regions of large repetitive sequences. CGIs are mainly found at the 5' regulatory regions of genes. About 60% of human gene promoters are embedded in CGIs. CGIs are also found in first exon and even first intron of several protein coding genes (15).

While most of the non-CGI CpG dinucleotides in the genome are methylated, majority of the CGIs remain unmethylated in normal conditions (16). However, it is known that some CGIs undergo methylation during development resulting in long-term transcriptional silencing of the associate genes. Such a natural silencing mechanism plays a crucial repressing role in X-chromosome inactivation, genomic imprinting, and transcriptional repression of repetitive and centromeric regions. DNA methylation exerts its silencing role in conjunction with histone modifications and nucleosome remodeling (17, 18).

1.2.2 Covalent histone modifications

Histone proteins are building blocks of nucleosomes that can be post-translationally modified by methylation or acetylation, for example. The methylation and acetylation of histones regulate variety of nuclear processes including organization of chromatin structure

and gene expression (19). Generally, it has been observed that regions silenced by DNA methylation show hypoacetylation and hypermethylation of specific histone lysine residues including lysine 9 or 27 in histone H3 (H3K9 or H3K27). Inversely, promoters of actively transcribed genes show hyperacetylation of histones H3 and H4, and methylation of lysine 4 of histone H3 (H3K4). The interactions between DNA methylation and histone modifications mediate gene silencing through recruitment of repressive complexes and methylated DNA binding proteins (20) (Fig. 2).

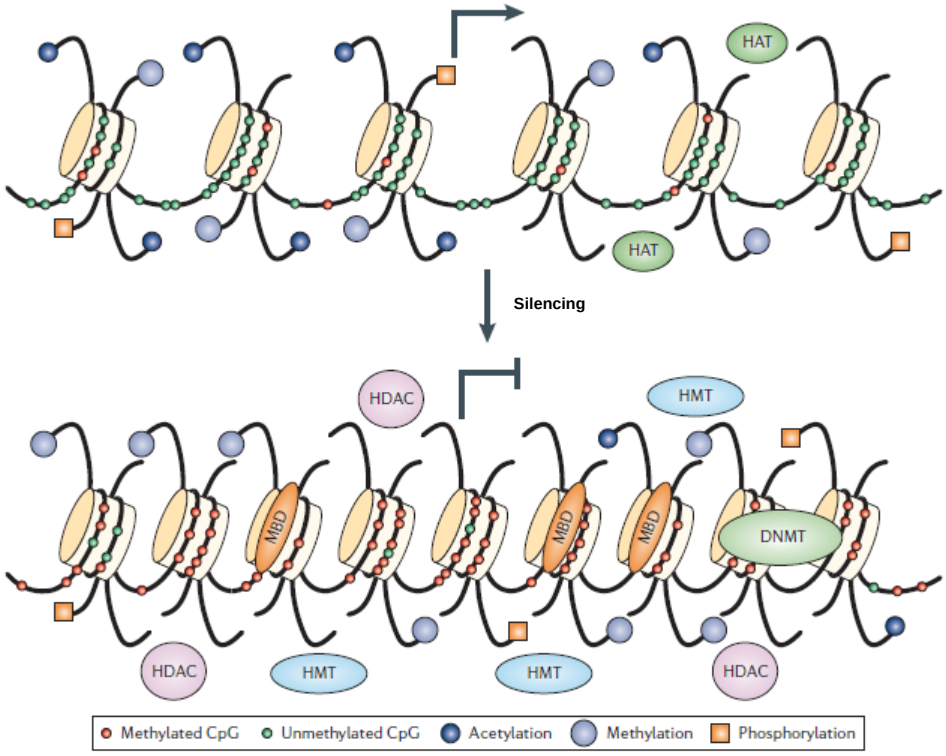


Figure 2. DNA methylation and histone modifications contribute to gene silencing. The promoter region of a transcriptionally active gene is hypomethylated and acetylated in the lysine residues on histone tails. These marks lead to the formation of euchromatin (Upper part). A silenced gene, is characterized by DNA hypermethylation in the promoter region, which attracts methyl-binding proteins such as MBD. Additionally, increased activity of histone deacetylases and histone methylases leads to loss of active markers such as H3K4 and H4K16 acetylation, and establishment of repressive markers such as H3K9 methylation (Lower part). DNMT, DNA methyltransferase; HAT, histone acetyltransferase; HDAC, histone deacetylase; HMT, histone methyltransferase (21).

1.2.3 microRNAs (miRNAs)

MicroRNAs (miRNAs) are endogenous small non-coding RNAs of 20-23 nucleotides in size which are involved in post-transcriptional control of gene expression (22). miRNA transcripts are generated either by RNA polymerase II (23) or III (24) as long primary transcripts (pri-miRNA) carrying the mature miRNA sequence in a stem loop structure. In the nucleus, cleavage of the pri-RNA stem loop by the RNase III endonuclease Drosha releases a 60-70 nucleotide long precursor RNA (25), called pre-RNA, which is subsequently transported into the cytoplasm and is further processed by Argonaute2 and Dicer (26, 27). One strand of the Dicer cleavage product - an RNA duplex with 3' overhangs and 5' phosphate ends – carrying the mature miRNA sequence is further incorporated into the RNA-induced silencing complex (RISC) (28). Due to the complementarities of miRNAs to the 3'UTR of mRNAs, the active RISC recognizes target transcripts and promotes translation inhibition or mRNA destabilization (29-32), both resulting in the reduction of target protein level (Fig. 3).

1.3 Epi-Genetics of Cancer

The inactivation of tumor suppressor genes has been widely observed in almost all cancers. A well-accepted model defining the genetic background of this inactivation is Knudsen's "two-hit" model. Originally, his hypothesis suggested that a germline mutation in one allele of a tumor suppressor gene can lead to an inherited predisposition to retinoblastoma; however, tumorigenesis requires a somatic mutation of the second copy of the respective gene. Non-hereditary cancer of the same type requires the same two hits, but both are

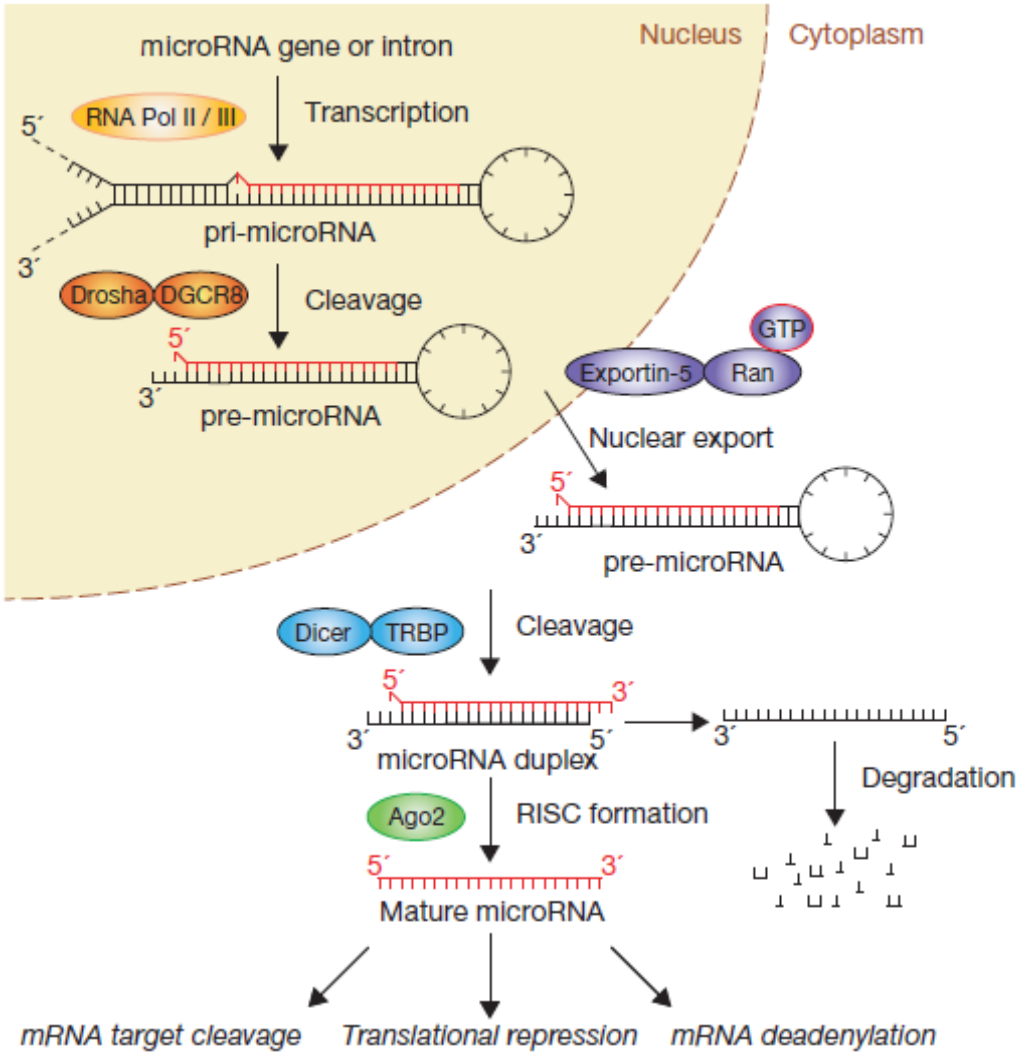


Figure 3. Pathway of miRNA processing. Following different processing and cleavage steps of miRNA transcripts in the nucleus and cytoplasm, the mature form of the miRNA is produced. The functional strand (red) of the mature miRNA is loaded together with Argonaute (Ago2) proteins into the RISC, where it guides RISC to silence target mRNAs through mRNA cleavage, translational repression or deadenylation, whereas the passenger strand (black) is degraded (33).

somatic (34, 35). This very specific finding has been generalized to other cancers and related tumor suppressor genes and has become a key concept in tumor genetics. This model has also been broadly used to find and characterize the genetic defects responsible for hereditary cancers.

In 1993, Bert Vogelstein and Kenneth Kinzler proposed that tumor formation is a multistep process which involves sequential mutations in certain genes associated with the disruption or over-activation of certain pathways that consequently lead to the transformation of

normal cells to cancerous cells (36). This model was originally described for colon carcinoma which shows clear morphological stages during development. Subsequent studies on involvement of genetic defects in cancer, especially in hereditary cancers, provided more evidences on dysregulation of tumor suppressor genes and oncogenes in tumors. Nevertheless, subsequent data revealed that dysregulation of cancer related genes is not always accompanied by causative genetic defects, supporting the possible involvement of "epigenetic" factors in aberrant regulation of genes and pathways involved in tumorigensis (e.g.12, 37).

1.3.1 Abnormal DNA methylation patterns in cancer

Epigenetic patterns of the genome are dramatically different between normal and cancerous cells. Abnormal changes in methylation patterns found in human cancers include global demethylation of DNA with simultaneous hypermethylation of CGIs. Hypomethylation at DNA repetitive elements leads to genomic instability and might also induce overexpression of oncogenes, whereas, hypermethylation of CGIs located at promoter or in the vicinity of transcription start site (TSS) of tumor suppressor genes results in silencing of cognate genes (12, 38) (Fig. 4).

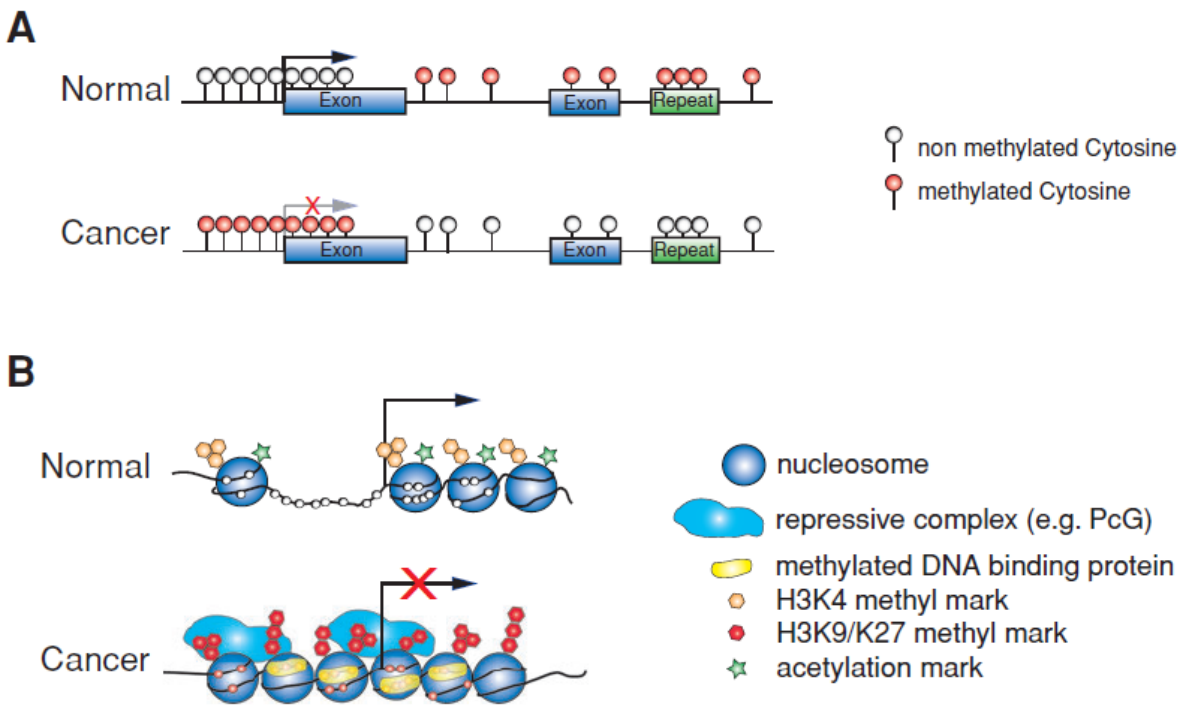


Figure 4. Epigenetic landscapes of normal and cancer cells. (A) DNA methylation. In normal cells, almost all of the CpGs are methylated while CGIs are unmethylated. In cancer cells, many CGIs become hypermethylated leading to silencing of their cognate genes, while global hypomethylation, mostly at repetitive elements, occurs. **(B)** Chromatin and histone modification. Active genes are associated with acetylation of histone tails, methylation of H3K4 and absence of nucleosome at their promoters. The promoters of silenced genes become associated with nucleosomes, lose acetylation and H3K4 methylation marks, and gain repressive methylation marks such as H3K9 or H3K27, which recruit repressive complexes (18).

The first tumor suppressor gene shown to harbor promoter CGI hypermethylation in cancer was *Rb (Retinoblastoma)* gene (39). This discovery was followed by other reports on promoter hypermethylation and silencing of other tumor suppressor genes in different cancers (40-42) confirming the existence of abnormal DNA methylation patterns in cancer. In 1995, Peter Laird showed that induction of hypomethylation in a mouse model for development of intestinal adenomas (early stage of colon cancer) reduces the number of intestinal adenomas (43). This study advanced the field of cancer epigenetics by providing evidence on a direct casual role for DNA methylation in cancer development. This study accompanied by other reports on hypermethylation and silencing of genes with causative roles in different tumors- such as *Rb* in retinoblastoma or *MLH1* in colorectal carcinoma

(44-46) – led to recognition of DNA methylation as an alternative "hit" to DNA sequence-based defects in Knudson's two-hit model (Fig. 5).

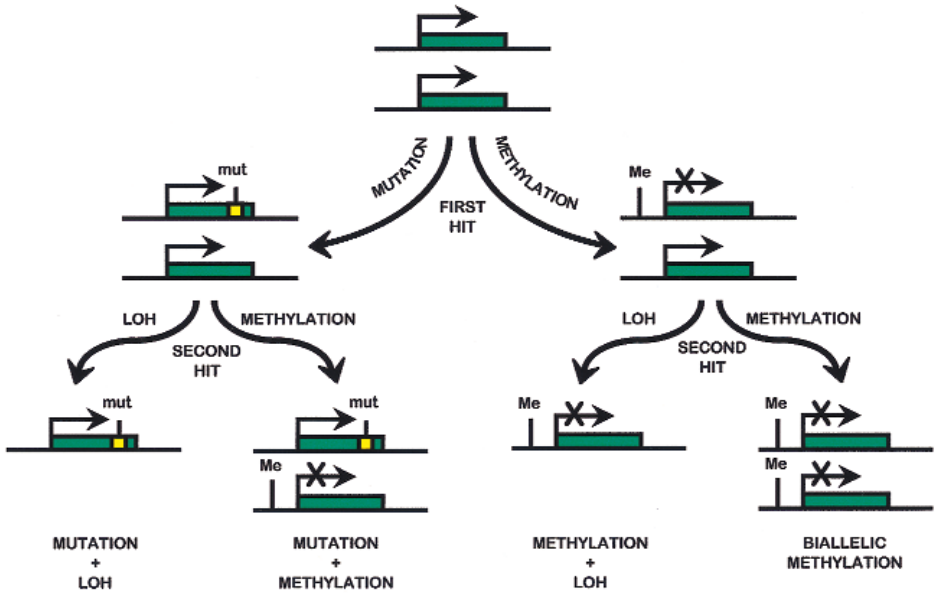


Figure 5. DNA methylation in Knudson's two-hit model. The active alleles of a tumor suppressor gene (green boxes) can become inactivated in the first or second hit or in both by a genetic defect disrupting DNA sequence or DNA methylation (epigenetic effect) resulting in transcriptional inactivation. LOH: loss of heterozygosity (47).

Subsequent studies on individuals affected by hereditary cancer but lacking any causative genetic mutations revealed that, indeed, epimutations (abnormal DNA hypermethylation) affect the responsible genes (48, 49).

1.3.1.1 DNA methylation as a biomarker

Aberrant DNA methylation events are abundant in tumors and also happen in the early stages of tumourigenesis. Also, this modification occurs on DNA molecule (which is more stable than RNA and protein molecules) through a covalent stable modification which can be detected in body fluids e.g. serum. In addition, hypermethylation of DNA occurs mainly at specific CGIs associated with tumor suppressor gene promoters, whereas point mutations often appear at different sites within a given gene in individual tumors. These facts have nominated the DNA methylation as a promising marker for clinical applications in cancer management (50, 51). To date, promoter DNA hypermethylation has been identified for a

large number of tumor suppressor genes (20, 52), among which some are suggested as potential clinical markers for cancer early diagnosis (*GSTP1*; prostate cancer), prognosis (*MGMT*; response to temozolomide treatment in glioma patients) or patient follow-up (*p15*; acute myeloid leukemia) (12).

1.3.1.2 Detection of DNA methylation

DNA methylation patterns are removed by classical molecular biology techniques, such as cloning in bacteria and PCR. Moreover, as the methyl group is located in the major groove of DNA, it is not detected by hybridization methods. Therefore, methylation-dependent pre-treatment of DNA have been developed to uncover the presence of the methyl group at cytosine residues. There are three main approaches used in this pre-treatment step; endonuclease digestion (using methylation sensitive/dependent enzymes), affinity enrichment (using antibodies against 5-methyl cytosine or methyl-binding proteins) and bisulfite conversion. Following the pre-treatment step, variety of molecular biology methods can be applied to analyze DNA methylation patterns, such as sequencing or microarray hybridization (53).

1.3.1.3 Oligonucleotide microarrays for DNA methylation analysis

Microarray-based technologies provide an opportunity for a high-throughput analysis of methylation patterns with a locus-specific resolution (54). Specialized microarrays have been developed to determine methylation patterns, following all the above mentioned pre-treatment steps (55).

Based on the bisulfite conversion of unmethylated cytosine to uracil and, following PCR-amplification, eventually to thymine, methylation patterns represent chemically introduced mutations that can be analyzed using different approaches (56). Therefore, when followed by an appropriate genotyping approach, bisulfite conversion of DNA provides a single-base-pair resolution for DNA methylation analysis (57). This concept has been used, by

different investigators, for the determination of DNA methylation patterns of tumor suppressor genes by oligonucleotide microarrays (58-62). In this microarray platform, methylation status of each CpG site is interrogated by two oligonucleotide probes; one detecting the methylated C and another one detecting unmethylated cytosine (Fig. 6). Despite high resolution in CGI mapping, the previous studies using these assays have been limited to individual genes. Recently, a genome-wide platform using the same principle has been developed by Illumina (63, 64). The current version of this array technology (called Infinium) has the coverage of more than 27,000 CpGs, associated with 14,495 protein coding genes and 110 miRNA gene promoters across the human genome. Nevertheless, the assay resolution is restricted to two CpG dinuclotides per locus, in the optimum condition.

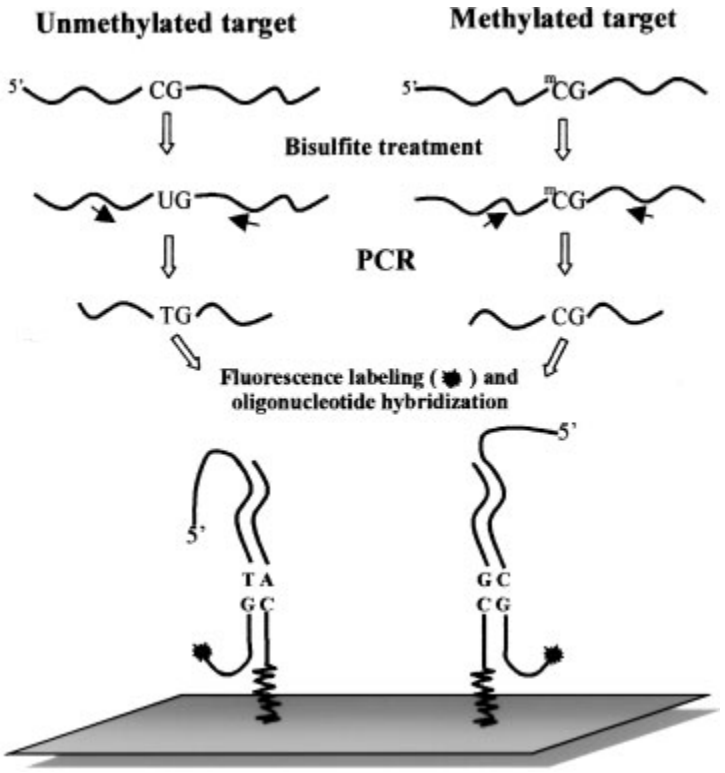


Figure 6. Analysis of DNA methylation by oligonucleotide microarrays following bisulfite conversion. Two oligonucleotide probes are used to interrogate methylation patterns of a CpG site; one detecting unmethylated target and another one detecting the methylated target. The signal intensity ratio of these probes reflect the methylation of the CpG site (62).

1.3.2 miRNAs and cancer

miRNAs are abundantly present in human cells (65). Due to their diversity and target variety - each miRNA can regulate the expression of numerous genes - they are assumed to directly control the expression of a large portion of all human protein-coding transcripts and, thus, are involved in several distinct cellular activities such as metabolism, differentiation, proliferation and apoptosis (reviewed in 66, 67). The observations that all these processes are altered in cancer (reviewed in 68) and that miRNA expression is deregulated in a variety of cancer types (69) suggests that miRNAs expression has profound influence on carcinogenesis (reviewed in 70). Actually, microarray-based expression profiling shows distinct miRNA profiles for different tumor entities (69). Moreover, based on function of their target mRNAs or pathways, several miRNAs have been characterized as tumor suppressors or oncogenes (71). Whereas differential expression patterns of miRNAs between normal and cancer cells are intensively investigated, the cellular function of the majority of miRNAs is still unknown.

The expression of miRNAs can be regulated by various mechanisms including genetic and epigenetic factors. DNA copy number abnormalities of miRNA genes loci, aberrant expression of transcription factors involved in transcription of miRNAs and abnormal epigenetic patterns of miRNA gene loci have been identified in cancer cells. (72-78)

1.3.3 Epi-Genetics of breast cancer

It is known that mutations in two major susceptibility genes, *BRCA1* and *BRCA2*, are directly associated with the development of the disease (79). Screening of these genes has been considered in clinical management of breast cancer patients with a family history (80). In addition, mutations of other genes, involved in different molecular pathways, have also

been reported in familial or sporadic forms of breast cancers (e.g. *TP53*, *PTEN*, *ATM* and *CHEK2*) (81).

1.3.3.1 DNA methylation and breast cancer

The most commonly mutated gene in familial breast cancer, *BRCA1*, is rarely found mutated in sporadic cases of the disease, though it is inactivated in the patients. Esteller and colleagues showed that the *BRCA1* inactivation in sporadic breast cancer was directly due to changes in the methylation level of *BRCA1* promoter (37), indicating the involvement of epigenetic factors on deregulation of key genes in breast cancer. This finding was further supported by identification of aberrant methylation on promoter of different genes associated with breast cancer, such as steroid receptor genes family members (*ER, RARβ2*), mismatch repair genes (*hMLH1*), cell cycle regulators (*p16*), cell adhesion (*CDH1*) and inhibitors of proteases (*TIMP-3*) (82, 83).

Today, perturbations of epigenetic patterns are recognized as early events in breast

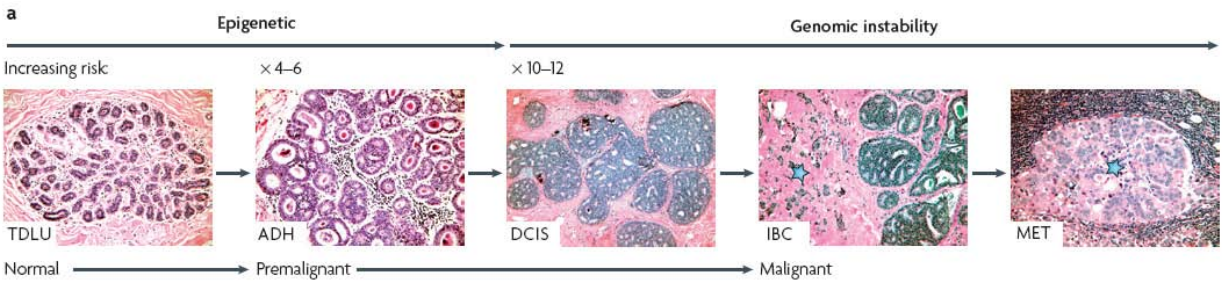


Figure 7 Epi-genetic alterations contribute to development of breast cancer. Deregulation of epigenetic patterns occurs in the early steps of tumorigenesis. The normal breast terminal ductal lobular unit (TDLU) contains lobules and ducts. Atypical ductal hyperplasia (ADH) is a premalignant lesion characterized by abnormal cell layers within the duct or lobule. ADH is thought to be the precursor of ductal carcinoma *in situ* (DCIS). With each stage, the risk of developing malignant or invasive breast cancer (IBC) increases. DCIS may give rise to IBC (indicated by a blue star adjacent to a DCIS lesion). Once cells have invaded, the risk for developing metastasis significantly increases. The lymph nodes are the primary site for breast cancer metastasis (MET; indicated by a blue star) (5).

cancer development (5) (Fig. 7) with a potential value to be used for early diagnosis. In addition, detection of cancerous cells by identification of abnormal DNA methylation has been successfully carried out in fluids obtained by nipple aspiration and ductal lavage (84,

85). Thus, screening for abnormal methylation patterns seems to provide a non-invasive approach for early cancer diagnosis. To this aim, identification of genomic loci whose methylation patterns represent an early diagnostic marker has been a focus of research in recent years.

1.3.3.2 miRNAs and breast cancer

In breast cancer, specific down-regulated and up-regulated miRNA subsets have been identified (86). A key role for de-regualted miRNAs in breast cancer development has recently been established through the identification of target proteins involved in breast cancer establishment and metastasis (87); miR-21 was shown to inhibit the expression of programmed cell death 4 (PDCD4) (88) and tropomyosin 1 (TPM1) (89) promoting cellular proliferation. miR-27a regulates the expression of zinc finger and BTB domain containing 10 (ZBTB10) leading to increased levels of SP proteins (90). miR-125a and miR-125b were shown to regulate the oncogenes ERBB2 and ERB3 in SKBR3 cells (91). miR-206 and miR-222/221 were reported to target estrogen receptor alpha (ERα) (92, 93). Whereas, miR-10b targets homeobox D10 (HOXD10) and initiates metastasis and tumor invasion (94), miR-31 is able to inhibit breast cancer metastasis by regulating a cohort of prometastatic genes (95).

1.3.4 Aim

In this dissertation, two mechanisms involved in epigenetic control of gene expression were studied in breast cancer; DNA methylation and miRNAs.

Microarray based analysis of DNA methylation patterns

The aim of this study was to assess the potential of establishment of a DNA methylation signature, based on available literature data, for diagnosis of female breast cancer. For this purpose, loci with methylation patterns associated with breast cancer had to be identified by literature review, and oligonucleotide microarrays dedicated to these loci had to be developed and optimized to obtain reliable data for each locus. Then, DNA methylation patterns in breast tissue samples were analyzed using the microarrays. After validation in independent sets of samples, the identified methylation signature would be helpful for the diagnosis of breast cancer and/or for the assessment of risk to develop breast cancer.

Identification of miRNAs involved in ERα activation

This study was aimed at evaluation of possible involvement of miRNAs in up-regulation of ERα in breast cancer cells. In the first part of the study, a microarray-based profiling for miRNAs in ERα -positive and -negative mammary cell lines was the method of choice to identify miRNAs with a potential role in ERα signaling. A regulative role on ERα pathway for the candidate miRNA had to be demonstrated functionally. Elucidation of mechanism through which the candidate miRNA exerts its regulative role would help to improve our understanding of mechanisms underlying ERα activation in breast cancer. As ERα has a crucial role in proliferation and survival of vast majority of breast tumors, identified miRNA(s) might serve in possible new strategies for treatment of breast cancer.

2 Material and Methods

2.1 Materials

2.1.1 Instruments

Instrument	Supplier
7900 HT Fast Real Time PCR Systems	Applied Biosystems, Forster city, USA
ABI PRISM 7900 Sequence Detection System	Applied Biosystems, Forster city, USA
Automatic developing machine	Amersham, Freiburg, Germany
Beckman GS-6KR centrifuge	Beckmann, Wiesloch, Germany
Biofuge 13R, Biofuge Fresco, Biofuge pico	Heraeus, Kendro, Hanau, Germany
Centrifuge 5810 R	Eppendorf, Hamburg, Germany
Cell culture incubator	Köttermann, Hänigsen, Germany
Cell culture microscope	Carl Zeiss, Jena, Germany
Centrifuge 580R	Eppendorf, Hamburg, Germany
Cell viability analyzer	Beckmann, Wiesloch, Germany
Dismembrator	B.Braun Biotech, Melsungen, Germany
EL800 Universal Microplate Reader	BIO-TEK, VT, USA
External Hybridization holder	febit biomed GmbH, Heidelberg, Germany
Electrophoresis power supply	E-C Apparatus Corporation, USA
GSA- and SS34- Rotor	DuPont, Boston, USA
Geniom One microarray	febit biomed GmbH, Heidelberg, Germany
Geniom RT analyzer	febit biomed GmbH, Heidelberg, Germany
Heat Block QBT2	Grant Instruments/CLF, Emersacker
LB-940 Mithras Multilabel Reader	Berthold Technologies, Bad Wildbach,
Micro-centrifuge	NeoLab Laborbedarf, Heidelberg, Germany
Mini-Protein 3 gel and electrophoresis system	Bio-Rad Laboratories, Munich, Germany
Multifuge 3SR	Heraeus, Kendro, Hanau, Germany
NanoDrop ND-100 Spectrophotometer	NanoDrop Technologies, San Diego, USA
QuadriPERM plates	Vivascience, Hannover, Germany
Speedvac concentrator apparatus	Eppendorf, Cologne, Germany
Thermomixer compact	Eppendorf, Cologne, Germany
Thermal Cycler	MJ Research, California, USA
Unimax 1010 Shaker	Heidolph Instruments, Schwabach, Germany

Vortex	Scientific Industries Genie-2, New York, USA
Water bath SW22	Julabo Labortechnik, Seelbach, Germany

2.1.2 Chemical reagents, enzymes and general materials

Reagent	Supplier
2-Mercaptoethanol	Roche Diagnostics, Mannheim, Germany
Acrylamide (30%w/v)/Bisacrylamide (29/10.8%)	Bio-Rad Laboratories, Munich, Germany
Agarose	Sigma-Aldrich, Munich, Germany
Ammonium peroxydisulfate (APS)	Sigma-Aldrich, Munich, Germany
Ampicillin	Sigma-Aldrich, Munich, Germany
Bacto-agar	Difco Laboratories, Detroit, USA
Bovine serum albumin (BSA)	Sigma-Aldrich, Munich, Germany
Bacto-trypton	Difco Laboratories, Detroit, USA
Bacto-yeast -extract	Difco Laboratories, Detroit, USA
Biotinylated-dUTP	Roche Diagnostics, Mannheim, Germany
Bromphenol blue	Sigma-Aldrich, Munich, Germany
Blocking-Reagent	Roche Diagnostics, Mannheim, Germany
BstUI	New England Biolabs, Frankfurt, Germany
Chloroform	Merk, Darmstadt, Germany
Chloramphenicol	Sigma-Aldrich, Munich, Germany
DEPC water, molecular biology grade	Roth, Karlsruhe, Germany
Desoxynucleotide–Set	Fermentas, St. Leon-Rot, Germany
Dithiothreitol (DTT)	Invitrogen, Karlsruhe, Germany
Dimethylsulfoxide (DMSO)	Sigma-Aldrich, Munich, Germany
DNA loading buffer	Fermentas, St. Leon-Rot, Germany
DNase I, amplification grade	Invitrogen, Karlsruhe, Germany
ECL Hyperfilm	GE Healthcare Europe, Freiburg, Germany
EDTA	Merk, Darmstadt, Germany
EpiTect PCR control DNA set	Qiagen, Hilden, Germany
Estradiol (E2)	Sigma-Aldrich, Munich, Germany
Ethanol	Merk, Darmstadt, Germany

Ethidiumbromide	Sigma-Aldrich, Munich, Germany
GeneRuller 1kb and 50 bp DNA size markers	Fermentas, St. Leon-Rot, Germany
Glycerin	Roth, Karlsruhe, Germany
Halt™ protease and phosphatase inhibitor single-use cocktail, EDTA-free	Thermo Scientific, Rockford, USA
Hind III	New England Biolabs, Frankfurt, Germany
Human Universal Reference RNA	Agilent Technologies, Waldbronn, Germany
Kanamycin	Roche Diagnostic, Mannheim, Germany
Klenow fragment	New England Biolabs, Frankfurt, Germany
M-PER mammalian protein extraction reagent	Thermo Scientific, Rockford, USA
Magnesiumchloride	Merk, Darmstadt, Germany
Methanol	Merk, Darmstadt, Germany
Milk powder	Sigma-Aldrich, Munich, Germany
Na_2HPO_4	Merk, Darmstadt, Germany
NaH_2PO_4	Merk, Darmstadt, Germany
NaOH	Merk, Darmstadt, Germany
Natriumacetat	Merk, Darmstadt, Germany
Natriumchloride	Merk, Darmstadt, Germany
Natriumcitrate	Merk, Darmstadt, Germany
Methanol	Merk, Darmstadt, Germany
Nuclease free water	Ambion, Austin, USA
Random primer oligonucleotides	Invitrogen, Karlsruhe, Germany
QIAzol	Qiagen, Hilden, Germany
Rainbow Molecular Weight Marker	Amersham, Freiburg, Germany
SAM	New England Biolabs, Frankfurt, Germany
SDS	Sigma-Aldrich, Munich, Germany
SAPE	Invitrogen, Karlsruhe, Germany
SpeI	New England Biolabs, Frankfurt, Germany
SssI methyltransferase	New England Biolabs, Frankfurt, Germany
T4 Gene 32 protein	Roche Diagnostic, Mannheim, Germany
TEMED	Bio-Rad Laboratories, Munich, Germany
Tris-(hydroxymethyl)-aminomethane (Tris)	Sigma-Aldrich, Munich, Germany
Tris-Base	Sigma-Aldrich, Munich, Germany

Triton-X100	Sigma-Aldrich, Munich, Germany
Tween 20 (Polyoxy-ethylen-sorbitan-monolaurate)	Sigma-Aldrich, Munich, Germany

2.1.3 Cell culture

Item	Supplier
Charcoal-stripped FBS	Gibco/ Invitrogen Karlsruhe, Germany
DMEM	Gibco/Invitrogen, Karlsruhe, Germany
Mammary Epithelial Cell Growth medium	Provitro, Berlin, Germany
Foetal bovine serum (FBS)	Biochrom, Berlin, Germany
Phosphate buffered saline (PBS)	Gibco/ Invitrogen Karlsruhe, Germany
Penicillin 1000u/ml-Streptomycin100µg/ml	Gibco/ Invitrogen Karlsruhe, Germany
Trypsin/EDTA solution	Gibco/ Invitrogen Karlsruhe, Germany

2.1.4 Kits

Item	Supplier
Absolute qPCR Green Mixes	ABgene, Epsom, UK
CellTiter Glo Luminescent Cell viability assay	Promega, Mannheim Germany
Caspase-Glo3/7 assay	Promega, Mannheim Germany
DNeasy blood and tissue kit	Qiagen, Hilden, Germany
Effectene Transfection Reagent	Qiagen, Hilden, Germany
ECL + Western Blot Detection Kit	GE Healthcare Europe, Freiburg, Germany
Epitect Bisulfite Kit	Qiagen, Hilden, Germany
miRNAeasy Kit	Qiagen, Hilden, Germany
HotStar Taq DNA polymerase Kit	Qiagen, Hilden, Germany
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
QIAquick PCR purification Kit	Qiagen, Hilden, Germany
Dual-Luciferase Assay Reporter System	Promega, Karlsruhe, Germany
siPORT™ *NeoFX*™ Reverse Transfection Kit	Ambion, Austin, USA
SuperScript III reverse transcription Kit	Invitrogen, Karlsruhe, Germany
Pierce BCA Protein Assay Kit − Reducing	Thermo Scientific, Rockford, USA

Agent Compatible Kit.	
Taq DNA polymerase Kit	Qiagen, Hilden, Germany
TaqMan MicroRNA Reverse Transcription Kit	Applied Biosystems, Darmstadt, Germany
TaqMan Universal PCR Master Mix	Applied Biosystems, Darmstadt, Germany
T4 DNA ligase	Invitrogen, Karlsruhe, Germany
TOPO TA cloning Kit for sequencing	Invitrogen, Karlsruhe, Germany

2.1.5 Media, solutions and buffers

Media, Solutions and Buffers	**Components**
APS	10% APS w/v in ddH_20
LB Medium	1 % (W/V) Bacto-Trypton 0.5 % (W/V) Bacto-Yeast –Extract 1 % (W/V) NaCl In H_2O, pH = 7.5 Autoclave at 121°C and 1000hPa for 20 min
S.O.C. Medium	2% trypton 0.5% yeast extract 10 mM NaCl 2.5 mM KCL 10 mM $MgCL_2$ 10 mM $MgSO_4$ 20 mM glucose
10x PBS	137 mM NaCl 27 mM KCl 100 mM NaH_2PO_4 17 mM KH_2PO_2 dissolved in ddH_2O
1x TBS-T	20 mM Tris-Base 137 mM NaCl 3.8 ml 1M HCl 0.1% (v/v) Tween 20 dissolve in 1l ddH_20

5x Loading buffer	100 mM EDTA 30% (v/v) glycerine 0.25% (w/v) bromophenolblue dissolved in ddH_2O
20% SDS	20% SDS (w/v) dissolve in ddH_20
2x Laemmli buffer	50 mM Tris/HCl pH=8.0 2.4% w/v SDS 8% glycine 0.2% w/v bromphenol blue 2.5% v/v β-mercaptoethanol
5x SDS-PAGE running buffer	25 mM Tris-Base 200 mM glycine 10% (w/v) SDS ddH_20
10x Western blot buffer	25 mM Tris-Base 192 mM Glycine 20% v/v methanol (added when preparing 1× buffer) ddH_20
BSA- blocking buffer	3% BSA in PBS
Resolving gel buffer	1.5M Tris-Base 10% (w/v) SDS ddH_20 PH= 8.8
Stacking gel buffer	0,5M Tris-Base 10% (w/v) SDS ddH_20 PH=6.8
Milk-blocking buffer	5% Milk powder in TBS-Tween (0.1%)
Stripping buffer	100 nM β-Mercaptoethanol 2% SDS 62.5 nM Tris-HCl, pH=6.7 ddH_20

SSPE, 6×	0.9M NaCl 60mM NaH_2PO_4, (pH=7.4) 6 mM Na_2EDTA
TAE buffer, 50×	0.4M Tris base 0.4M acetic acid 20mM EDTA (pH=8)

2.1.6 Antibodies

Antibodies	Supplier	Dilutions
Anti-rabbit IgG, HRP-linked Antibody	Cell Signaling Inc, Danvers,	1:5000
Anti-mouse IgG, HRP-linked Antibody	Cell Signaling Inc, Danvers,	1:5000
Actin	Sigma-Aldrich, Munich,	1:1000
ERα	Cell Signaling Inc, Danvers,	1:1000
GAPDH	Sigma-Aldrich, Munich,	1:25000

2.1.7 Plasmids and synthetic RNAs

Plasmid/synthetic RNAs	Source
Actin-RL renilla luciferase construct	Dorothee Nickles & Michael Boutros, DKFZ
pc-myc-CMV-D12 expression vector	Core Facility of DKFZ
pMIR-REPORT firefly luciferase construct	Ambion, Austin, USA
RASD1 siRNA-2	Qiagen, Hilden, Germany (SI00113197)
RASD1 siRNA-4	Qiagen, Hilden, Germany (SI00113211)
RASD1 siRNA-5	Qiagen, Hilden, Germany (SI03080483)
RASD1 siRNA-6	Qiagen, Hilden, Germany (SI03104115)
ERα siRNA-01	Dharmacon, Rockford, USA (D-003401-01)
ERα siRNA-02	Dharmacon, Rockford, USA (D-003401-02)
ERα siRNA-03	Dharmacon, Rockford, USA (D-003401-03)
ERα siRNA-09	Dharmacon, Rockford, USA (D-003401-09)
Negative control siRNA #1	Ambion, Austin, USA (AM4611)

Pre-miR-375	Ambion, Austin, USA (PM10327)
Pre-miR-Control	Ambion, Austin, USA (AM17110)
Anti-miR-375	Ambion, Austin, USA (AM10327)
Anti-miR-Control	Ambion, Austin, USA (AM 17010)

2.1.8 qRT-PCR primers

Primer Name	Sequence or Cat. number
LAMINB1 forward	5'-CTGGAAATGTTT GCATCGAAGA-3'
LAMINB1 reverse	5'-GCCTCCCATTGGTTGATCC-3'
GHSR reverse	5'-CCGACCTGGACTGGGATG-3'
GHSR reverse	5'-GCGATACCCACCACGAAG-3'
pS2 forward	5'-ATACCATCGACGTCCCTCCA-3'
pS2 reverse	5'-AAGCGTGTCTGAGGTGTCCG-3'
RASD1	QT00201390 (Qiagen, Hilden, Germany)
SFRP1	QT00031927 (Qiagen, Hilden, Germany)
SFRP2	QT00073220 (Qiagen, Hilden, Germany)
SFRP3	QT02238551 (Qiagen, Hilden, Germany)
SFRP4	QT02238537 (Qiagen, Hilden, Germany)
HDPR1	QT01891071 (Qiagen, Hilden, Germany)
miR-375	000564 (Applied Biosystems, Darmstadt, Germany)
RNU6B	001093 (Applied Biosystems, Darmstadt, Germany)
RNU66	001002 (Applied Biosystems, Darmstadt, Germany)

2.1.9 *Genomic* primers for the amplification of CGIs from genomic DNA

CGI	Forward primer (5'→3')	Reverse primer (5'→3')
APCA1	ACTGCCATCAACTTCCTTGC	GCAGCACCTCCATTCTGTCT
BRCA1.A	GGAAGTGAAATATGATCCCTATCCT	GTAGAGACGGGGTTTCACCA
BRCA1.B	TCTCCCTCCCACAAAATCCT	GCAGGCACTTTATGGCAAAC
BRCA1.C	GTGCGCGCTTGTACTTGTC	ACGAAGGGCTCTCTCATCCT

C/EBPα	CAGCAGCTCAAGACCAAGACT	GACTCCATGGGGGAGTTAGAG
CCND2	GTGAAGAAACGCCACCAGAT	GCAGAGTGTGTGCAGAGGAA
CDH1	CAGGTCCCATAACCCACCTA	AATGCGTCCCTCGCAAGT
CDH13	GCCAGAGTGGGAGCCTGTGT	CACCCTGAGGCACTCCCTGT
CST6	GCTGCGGTTGGTAGTTCATT	GGTTGTCCATCGTGGAGTTT
DAPK	CAGGCTGGAGAGAGATTGCTCCCAGTG	TGTTGTCACAGAAGGGCAAG
DKK-3	ACAGGACAGAGGTTGGCTTG	AAGCCTCTCTCAGCCCCTAC
ER	GCACACCCCATTCTATCTGC	CCTCGGGGTAGTTGTACACG
GSTP1.A	AGAATTGCTTGAACCCGAGA	ACTCACTGGTGGCGAAGACT
GSTP1.B	ACCTACTCGGGAGGCTGAAG	GGGAAGCCTTTCCCTCTTT
HDPR1	GGTGCAGTAAACCCGAAAAG	CAGCAAGATGTTCTCCTCCA
HIC1	GTGGGGCATCGGCTAAGAG	GAGAAGGAGCAGGAGGTGAGC
HIN-1.A	TGTGCTTTGCTCTCCATCTG	CGCCTGCCAAGAGGAAGT
HIN-1.B	CGAGCCCACTAAGAAAGCAG	CGAGCCCACTAAGAAAGCAG
HOXA5	TCCCCGAAGGTGATATCTGTAT	TAGCGACCGCAAAATGAGTTTA
HPGD	TGCTCCGCTCTCCTTCTATC	TGGCAATTAGGTTTGGGTTC
MLH1	CAAATAACGCTGGGTCCACT	TCTGTTCACATCCAAAGGCA
OGDHL	GCATTCCCATTTTACAGACGA	AAGGTCAGAGCCAAGGTCAC
P53	ATCCGACGCAGAGCTAAAGA	CCCTAACGTTTTCTCCCAGA
PAK3	ATGTTCCCTCCCCTTCAGTT	GGCTGACGTAGCTTTCAAGG
PGR.A	CCCTAGAGGAGGAGGCGTTG	CGGGAATATAGGGGCAGAGG
PGR.B	CAGCGGACTGCTGGACAGTG	GGAGGACGCAGACGAGACTGA
PTEN	CGAGGCTTAGCTCGTTATCCT	CTCTCATCTCCCTCGCCTGAG
RARB	TGCTGGGAGTTTTTAAGCTCTG	ACAGGCTTTTAGCTGGCTTG
RASSF1A	GTGTGAGGAGGGGACGAA	GATGAAGTCGCCACAGAGGT
SFRP1	CCCTTCTTTTTCTCCCCTTG	GCAAACTTCCAGGGACCTC
SFRP2.A	GCAAAACCAGACCCAGAAAG	GAAGAGCGAGCACAGGAACT
SFRP2.B	GAGATTCGAGGGAAAACGAA	GGCACAGCCAGAGTTTTCTT
SFRP3	AGCTCTTAGCCCACACAGGA	CTGGAAGTCAATGGTGCAGA
SFRP4	TCTAAGGCAGAGGGAGCAAA	ACCAGCTCCTCGTACTGCTC
SFRP5	TGTGGCGGAGAGAGATAAGG	GCCTGCCAGCCATAGTAGTC
SOCS1	GGAGGAGGGAGGGGAGTC	CGAACGGAATGTGCGGAAGT
Stratifin	GAGACACAGAGTCCGGCATT	TGATGAGGGTGCTGTCTTTG
TCF4	TCAAAACAGCTCCTCCGATT	CCCTTAAAGAGCCCTCCATC
TWIST.A	CCCCTCTGGCTCTGCTGCCTTT	CCCGCGAGGTGTCTGGGAGT

TWIST.B	CGGGGGAGGGGGACTGGAA	GCCGCCAGCTTGAGGGTCT
TWIST.C	CAGCGCACCCAGTCGCTGAA	TGGCGCTGACAAAACGGTCCT
WIF1	TCACCATCACCATCACCATC	AGGTACAGGCTCTCCTCCTG

2.1.10 *SoBi* primers for the amplification of CGIs from deaminated DNA

CGI	Forward primer (5'→3')	Reverse primer (5'→3')
APCA1	AGGGTTAGGTAGGTTGTG	ACCAATACAACCACATATC
ARHI	GTTGGGTTAGTTTTTTATAGTTGGTT	AACCAAACAACCTAAAAAACAAATAC
BRCA1.A	GTTTAGGTTGGAGTGTAATGG	ATTTCTCAAATAACTAAACCCCTAC
BRCA1.B	AGTAGAGGGTGAAGGTTTTTTGAG	AAACAAACTAAATAACCAATCCAAAAC
BRCA1.C	TTTGTATTTGTTAATAGTTATGGATTGGA	TTCTTTACAAAATTCCTCCCTTAAAC
C/EBPalp	TAGGTAGGAGGAGGGGGTTT	TCCATAAAAAAATTAAAATTCTCCC
CCND2	GTAAAGATAGTTTTGATTTAAGTATG	TCTAAAAAAATAAAATATAATCCTC
CDH1	ATTTAGTGGAATTAGAATAGTGTAGGTTTT	CTACAACTCCAAAAACCCATAACTAAC
CDH13	ATTTTTTGGAAAAGTGGAATTAGTT	AAAAATACCTAAAATCTCTAAACTC
CST6	TTTTTGTATTGGTATTTGTTGTTGG	CCTATAACCTCCCCCATCTAAAC
DAPK	GGTTGGAGAGAGATTGTTTTTAGTG	ACCCTAAACTACTACCTCTCCTCCA
DKK-3	GTAGGTAGTGAAGGAGATGGTTGTT	TCCACCAAAAAAAACTAAACTCAAC
ER	AGTGTAGTTTTTTTTAGGGTTATTTTATGT	AACCTCCAACTTTAAATACTAATCTCC
GHSR.A	ATTGTTTGTGAAAGGTAAGTATAAA	ACACTAACAAAATAACCCTCTAAAc
GHSR.B	GATTTGGATTGGGATGTTTTTT	CCCTTAATAACCACCACCTTAAC
GSTP1.A	GGAAAGAGGGAAAGGTTTTTT	CCATACTAAAAACTCTAAACCCCATC
GSTP1.B	TGGGGTTTAGAGTTTTTAGTAT	ACTACAACCCCAACCCCTAC
HDPR1	TTGATAGAGGAGGGGAAGTTTG	CCCCTAACCAACAACTCTTAAC
HIC1	GGGTAGGGGAGTTTAGGGTT	ATTACCCCAATTAAAAAAAATAATAC
HIN-1.A	ATTTGTTATATTTTTGTAGGTGATAGA	ATACCCTAACCAACTTCCTACTAC
HIN-1.B	GTAGTAGGAAGTTGGTTAGGGTA	AAACTAAAAAAACCTAAAATCCAC
HOXA5	TATAAAAATAAGAGGGGTTGTAAAA	CCTACCTAATAACCTCTAAAAATAAACTC
HPGD	GGTTTTTAGGGTATTGAAGGAAATT	ATACTCACAACCTCAACTTCAACAA
MLH1	TTTTTTTTAGGAGTGAAGGAGGTTA	TTTAATTAATAAAAATAAAAATATT
OGDHL	GGTTTTAGAGAGGAGTTTTTTTT	CAAAACCAAAATCACAATCTATTC
P53	TTTTGAAAGTATTGTGTTTTTTAGTAT	TCTTATTCATTATTATATTCCTAAATACCT

PAK3	TTTGATTGGTTTAGGTAAGGTGTAGA	AAAAACAAAACATCTACCCAACAAC
PGR.A	TGTTAGAAAGTTGTTTGGTTAGTTTATAGT	CTAATCAACTCCTACCCTTAACCTC
PGR.B	GGATTGTTGGATAGTGTTTTGG	AAACTATCTCCAACCTTACACCC
PTEN	TTATATTGGGTATGTTTAGTAGAGTTTG	AACTTCCATCATAACTACAACTTCC
RARB	TTTGGGAGTTGGTGATGTTAGA	ATTTACCATTTTCCAAACTTACTC
RASSF1A	GGAGGGAAGGAAGGGTAAGG	CAACTCAATAAACTCAAACTCCC
SFRP1	TTTTTAAGGGGTGTTGAGT	CAAACTTCCAAAAACCTCC
SFRP2.A	AAAAGGTTAAGAAAATTTTGGTTGT	CTACTCCAACACCTCCTTCATAATC
SFRP2.B	ATGTTTGGTAATTTAGTAGAAATTT	CAACCAAAATTTTCTTAACCTTTTT
SFRP3	GTGATTTAGGGGAGGAGATATTTTAGA	TTCCAAAACAAAAACTTACACAAAA
SFRP4	AAAGTAGGTTTTTAGTTTTGGGTTG	AAAAAAAACATAACACTACCCTCTC
SFRP5	TAGGGAGTTTTGGGGAGAAA	CCCAAATAAATAACAACCTAC
SOCS1	AGGGGAGTTTAGGGTAGTTAGGAG	CTACCATCCTACAAAAAAAACCAAC
Stratifin	TTTGTGTGTTTTTAGAGTTATGGAGAG	TCTTAACCAAAAAAATAACCTCCTC
TCF4	TTTTATTTGTTTTTTGGGTTTGGTA	AAAAAACCCTCCATCTTACCTCTTA
TWIST.A	GTTGTTTTTTTTATGGATTGGGTT	ACAACAATAACAACAACTTCTACACAATAA
TWIST.B	ATGAGATATTATTTATTGTGTAGAAGTTGT	TTAAAAACAACAATATCATTAACCTAAC
TWIST.C	TTGAGTAAGATTTAGATTTTTAAGTTGG	ATATCTCTATCTCCCCTCCTATCAC
WIF1	GATGTTTTAGGGGTTTTTGAGTGT	CTACAACTCCCTCAACCAAAACTAT
miR-375	***Bisulfite sequencing primers***	
Bs1	GTTTTATAGTGTAGTGGATGAGGGG	AAAAAACAAAAAATCCCAAATAAAC
Bs2	GGGGATTGAATAGGTAGTATAAGAGTATA	TAAAACCTAACAACCCAAAAACTAC

2.1.11 Primers used for ChIP analysis of miR-375 locus

Primer	Forward primer (5'→3')	Reverse primer (5'→3')
CHIP-1	GGATGAGGGGAGTCGAATG	CAAAAAGTCCCAGGTAGGCG
CHIP-2	TCTTCCAGGAGTCTGACCC	AGTGTGGGCGGAGAGGTG
CHIP-3	TGGACAGGTGTGAGTGTGTG	GTCGCCCTCGGTGATCTC
CHIP-4	GCCCTCCGCTCCCGCC	GGTTTCCACCTCCAGAAGG

2.1.12 Primers used for isolation of miR-375 target sites

Primer	Sequence (5'→3')
RASD1_forward	CACACTAGTGGACCTTTTTGTTAAGTC
RASD1_reverse	CACAAGCTTTTAAAGACCTCAGACAC
EBF3_forward	CACACTAGTGCTTTTCCCACCTTTTCC
EBF3_reverse	CACAAGCTTTTTAGGCCTCTGAATTTT

2.1.13 Cell lines

Cell lines	Cellular origin	Tumorigenic	ERα	Source
HEK293T	Embryonic kidney	Yes	N/A	ATCC (CRL-1573)
MCF7	Breast	Yes	+	ATCC (HTB-22)
T47D	Breast	Yes	+	Prof. Peter Altevogt, DKFZ
SK-BR3	Breast	Yes	-	Prof. Peter Altevogt, DKFZ
MDA-MB-435	Breast	Yes	-	Prof. Peter Altevogt, DKFZ
MDA-MB-231	Breast	Yes	-	Prof. Peter Altevogt, DKFZ
HDQ-P1	Breast	Yes	-	Prof. Peter Altevogt, DKFZ
MCF-10A	Breast	No	-	Prof. Annemarie Poustka, DKFZ
MCF-12A	Breast	No	-	Prof. Annemarie Poustka, DKFZ
MELN	Breast	N/A	+	Prof. Doris Mayer, DKFZ

2.2 Methods

2.2.1 General molecular biology methods

2.2.1.1 Transformation in competent bacteria

One Shot® TOP10 chemically competent *E.coli* DH10B (Invitrogen) were used to amplify plasmid DNA. 50µl of electro-competent *E.coli* were thawed on ice, mixed with the complete ligation reaction and incubated on ice for 30 minutes. Thereafter, the bacteria were shifted to 42°C for 30 seconds and briefly equilibrated on ice. 250-300µl of SOC medium was then added and cells were incubated for 1 h at 37°C with shaking (200 rpm). The suspension was plated onto LB-agar plates containing the respective antibiotic (100

mg/ml) and incubated overnight at 37°C. In order to propagate the cultures, either single colonies were picked or frozen glycerol stocks were thawed and inoculated into 10ml LB medium with antibiotic, and shaken overnight at 37°C. For storage of bacteria, a bacterial culture was grown in culture medium to an OD of 0.8 (measured at 600nm). Thereafter, 500µl of the bacterial culture was combined together with 500µl glycerol and subsequently frozen at –80°C.

2.2.1.2 Mini-preparation of plasmid DNA

E.coli containing the plasmid of interest was inoculated in 8ml LB medium, including the appropriate antibiotics (100 mg/ml) and grown overnight at 37°C in a shaking incubator (200 rpm). Plasmid DNA was extracted using the Qiagen Plasmid Mini Kit. First, the bacteria was harvested by centrifugation for 15 minutes at 3500 rpm and re-suspended in 250µl buffer P1 containing 0.1 mg/ml RNase A. Thereafter, 250µl of lysis buffer P2 was added and the tube inverted gently 4–6 times. 350µl neutralization buffer N3 was then added to the tube and mixed again by inversion. The resulting lysate was centrifuged for 10 minutes at maximum speed in a tabletop microcentrifuge (13,000 rpm). The supernatant from this centrifugation was added to a QIAprep spin column placed in a 2ml collection tube and centrifuged for one minute at 13,000 rpm. The flow-through was discarded and the column bound plasmid DNA was washed using 500µl of buffer PB, followed by centrifuging for one minute (13,000 rpm). After discarding the flow-through, the column was washed with 750µl of PE-buffer, followed by centrifuging for 1 minute (13,000 rpm). Thereafter, the column was placed into a clean 1.5ml microcentrifuge tube, 50µl of water was added and incubated at room temperature for one minute. The DNA was then eluted by centrifugation for 1 minute (13,000 rpm). Quantification of the DNA samples were carried out on a Nanodrop spectrophotomer using the DNA measurement setting.

2.2.1.3 DNA isolation and quality assessment

To isolate DNA from powdered tissue samples or cells, the DNeasy Blood & Tissue kit (Qiagen) was used according to the manufacture provided by company; 180µl ATL buffer was added to an eppendorf tube containing 20mg of powdered tissue sample or cell pellet (cells had been harvested by trypsinization after washing with PBS). Then, 20µl of proteinase K was added to the tube and contents were mixed by vortexing. The mixture was incubated at 56°C overnight on a shacking thermomixer. Subsequently, 200µl of buffer AL was added to the sample and mixed by vortexing. Afterwards, 200µl of ethanol (100%) was added and mixed by vortexing. The mixture was transferred into a DNeasy Mini spin column placed in a 2ml collection tube and centrifuged for one minute at 13,000 rpm. The column was placed in a new collection tube and washed first with 500µl AW1 buffer and then with 500µl AW2 buffer. The column was dried by centrifugation at 13000 rpm for 2 minutes. Finally DNA was eluted in 200µl of AE buffer. The isolated patient DNA was then checked for the quality by gel electrophoresis on a 1% agarose gel. The DNA with good quality was then subjected to bisulfite conversion.

2.2.1.4 RNA isolation and quality assessment

Total RNA was extracted from cultured cells or tissue samples using miRNesy kit (Qiagen). To isolate RNA from cultured cells, cells were washed with PBS, then 1ml QIAzol lysis reagent was added to each flask (appropriate amount of lysis solution was used when isolating RNA from wells of 6-well plates) and cells were scraped using a disposable cell scraper. For isolation RNA from tissue samples, 700µl of lysis reagent was added to 25mg powdered sample and mixture was homogenized by pumping using a sterile syringe. The lysate was incubated for 5 minutes at RT and was collected into 1.5ml eppendorf tube. Afterwards, 300µl chloroform was added. The mixture was mixed thoroughly and incubated for another 5 minutes at RT. Thereafter the cell lysate was

centrifuged for 15 minutes in a pre-cooled centrifuge at 4°C at 13,000 rpm. The aqueous phase was placed into a fresh tube and mixed with an equal volume of 70% ethanol, transferred to a RNeasy spin column and centrifuged for one minute at 13,000rpm. The flow-through was discarded and 700µl of wash buffer RW1 was added, and the centrifugation and flow through removal procedure repeated. The same procedure was repeated using 500µl RPE wash buffer. In the end, the RNeasy spin column was placed in a new collection tube and the RNA eluted in 50µl nuclease free water by centrifugation at 13,000 rpm for one minute. Quantification of the RNA sample was carried out using the Nanodrop spectrophotomer using the RNA measurement settings (A260 / A280 ratio).

2.2.1.5 Protein isolation

For protein isolation, the M-PER mammalian lysis buffer (Pierce) was used. During the isolation, the cells/lysates were kept on ice. Medium was aspirated from the cells in one well of a 6-well plate. Cells were washed with 1ml ice-cold PBS containing 1mM Na_4VO_3 and 10mM NaF. Then, 40µl of M-PER solution containing Complete Mini Protease Inhibitor Cocktail, PhosSTOP Phosphatase Inhibitor Cocktail, 1mM Na_4VO_3 and 10mM NaF was added to each well. Cells were detached in the presence of the lysis buffer using a cell scraper. Then, the plate was shacked gently for 10 min on ice. Lysates were then collected and transferred into to an eppendorf tube and were centrifuged for 5 minutes at 13,000 rpm in a tabletop centrifuge at 4°C. The supernatant was collected in a new eppendorf tube and was stored at -80°C.

2.2.1.6 Protein quantification (BCA Protein Quantification)

Total protein concentration was measured using the Pierce BCA Protein Assay. The BCA assay is based on the reduction of Cu+2 to Cu+1 by protein in an alkaline medium followed by colorimetric detection of the cuprous cation using bicinchoninic acid (BCA).

BCA working reagent was prepared by mixing 50 parts of BCA Reagent A to 1 part of BCA Reagent B. 20µl of the BSA standards or samples were plated into single wells of a 96 well microtiter plate well. Thereafter, 100µl of BCA working reagent was added, the plate shaken for 30 seconds and then placed in a 37°C incubator for 30 minutes. The plate was then cooled to room temperature and the absorbance was read at 562nm. A standard curve of absorbance versus micrograms of standard protein was prepared, and the protein concentration of the samples was determined from this graph.

2.2.1.7 Western blot analysis

For 10% SDS-PAGE gel, 10% acrylamide/bisacrylamide in resolving buffer for the resolving part and 5% acrylamide in stacking buffer for the stacking part were prepared. 0.06% (w/v) ammoniumpersulphate and 0.1% (v/v) N,N,N´,N´-tetramethylethylendiamine (TEMED) were added to the gel solution to induce polymerization. 5µg of the protein samples were boiled in Laemmli buffer (1:1) for 5 minutes and subsequently loaded onto the slots of the gel. In addition, a protein size maker was also included. Electrophoresis of the protein gel was applied for 45 minutes at 160V and 70mA. The electrophoretic transfer of polypeptides from the polyacrylamide (PAA) gel to a polyvinylidene floride (PVDF) membrane was performed by using a wet gel transfer apparatus. The PVDF membrane was briefly activated in 100% methanol and soaked for 5 minutes in Western blot buffer for 10 minutes. The PAA gels were carefully removed from the electrophoresis chamber, the stacking gel was cut and the gel was soaked 3 minutes in Western blot buffer. The Western blot ‘sandwich’ was assembled and the transfer was carried out for 2h at 100V and 250mA in ice-cold Western blot buffer. For protein detection with antibodies, the membrane containing the separated polypeptides were placed in a tank and washed with washing buffer (15 minutes) followed by milk-blocking buffer for 1h at RT. After blocking, the membrane was incubated with diluted primary antibody over night at 4°C. After washing

(3×10 minutes), the membrane was incubated with the secondary antibody conjugated with horse radish peroxidase for 1h at RT. After a further set of membrane washings (3×10 minutes), the blots were ready for enhanced chemiluminescence (ECL) protein detection (ECL Western Blot Detection Kit). The ECL substrate was prepared and the blot incubated with the solution for 1 minute in the dark at RT. The solution was then drained off, the membrane placed in a plastic foil to prevent drying, placed in an X-ray cassette and exposed to ECL film (Hyperfilm-ECL™). The exposure time of the film was between 10 seconds and 5 minutes.

2.2.1.8 DNA Sequencing

An aliquot of 30µl containing 30-100ng/µl plasmid DNA was sent to GATC Biotech AG for DNA sequencing.

2.2.1.9 cDNA synthesis

Reverse transcription (RT) is the transcription of RNA into single strand complementary DNA (cDNA). For the RT reaction, in a final volume of 20µl, 1µg of total RNA (in 8µl RNase-free water) was used together with 2µl of 5× 1st strand buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl2) and 1µl (10 units) DNaseI, and incubated for 20 minutes at RT to allow digestion of genomic DNA. Thereafter, 3µl of Master Mix1, which contained equal parts of 25 mM EDTA, dNTP mix (10 mM dNTPs each), and 300ng/µl random hexamer primers, was added to the mix and incubated in a thermocycler, first for 10 minutes at 65°C and then for 10 minutes at 25°C. This step removed secondary structures; heat inactivated the DNaseI and allowed the annealing of primers to the RNA. 5µl of Master Mix 2 containing 5× 1st strand buffer, 0.1 M DTT and nuclease free water at a ratio of 2:2:1 was then added. The mixture was incubated initially at 42°C for 2 minutes, then 1µl of reverse transcriptase and 0.2µl of T4 gene P32 protein were added to the

mixture and further incubated for 50 minutes at 42°C. Thereafter, the reaction was heat inactivated for 10 minutes at 95°C.

2.2.1.10 Quantitative PCR primer design

PCR primers for quantitative real-time (qRT)-PCR experiments were provided by Qiagen company as pre-designed primers or designed with the primer design tool from Roche Applied Science: 'Universal probe library' (https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=UP030000). Intron spanning primers were chosen in order to distinguish between contaminating genomic DNA amplicons, and those resulting from amplification of the correct cDNA template (primer information are provided in section 2.1.8).

2.2.1.11 Real-time PCR analysis

The PCR reaction mixture consisted of 6µl SYBR Green PCR mixture, 100nM forward and reverse primers and 2µl cDNA template in a reaction volume of 12µl. A standard curve consisting of a dilution series of the cDNA produced from Stratagene Universal Human Reference RNA (1:1, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128) was used to determine the efficiency of the PCR reaction. To quantify cDNA templates obtained from RNA samples, real-time PCR was performed in a 7900 HT Fast Real-Time PCR System (Applied Biosystems) with the following settings: 2 minutes at 50°C, 95°C for 15 minutes , then 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. Final steps comprised 15 seconds at 95°C, 15 seconds of 60°C and 15 seconds at 95°C. The heating ramp between the last two steps was increased to 20 minutes to obtain a melting curve of the final qRT-PCR products. This was necessary because SYBR Green fluorescence signals may also be derived from side products such as primer dimers. Calculation of efficiency and relative quantification was performed using the housekeeping genes lamin B1 (*LMNB1*) and Glyceraldehyde 3-

phosphate dehydrogenase (*GAPDH*) as reference. Individual samples were measured in triplicate to ensure reliability.

2.2.1.12 Quantitative real-time PCR for miR-375

Quantification of mature miR-375 expression was performed using a miR-375 TaqMan MicroRNA Assay which provided specific primers for reverse transcription (RT) of miRNA and primers for real-time PCR (Applied Biosystems). 20ng and 50ng of total RNA isolated from cell lines and patient samples, respectively, were reverse transcribed with sequence specific primers by using a TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems). First, a master mix was prepared for each reaction including following materials; 0.15µl dNTPs (100mM), 1µl MultiScribe Reverse Transcriptase (50 U/µl), 1.5µl 10×RT buffer, 0.19µl RNase Inhibitor (20U/µl), 4.16µl Nuclease free water, 3µl RT primer. Then the above mentioned RNA amount in volume of 5µl was added to the prepared mix. Then the reaction was inclubated in a thermalcycler at 16°C for 30 minutes, then 42°C for 30 minutes followed by 5 minutes at 85°C to complete the RT procedure.

Real-time PCR was performed in triplicate in a 384-well plate on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Reaction consisted of 10µl 2×Universal Master Mix (Applied Biosystems), 1µl miR-375 real-time primers, 7µl RNase-free water. 2µl of the cDNA prepared in the RT reaction was added to this master mix in each well. Instrument setting was as follow; 2 minutes at 50°C, 95°C for 15 minutes, then 45 cycles of 15 seconds at 95°C and 1 minute at 60°C. The average expression of RNU6B and RNU66 was used for normalization.

2.2.2 Processing of tissue samples for DNA/RNA isolation

Frozen surgical breast tissue samples were obtained with informed consent of patients after approval of the Institutional Review Board at Tehran University of Medical Sciences, Shahid Beheshti University of Medical Sciences and University of Welfare Sciences and Rehabilitation, Tehran, Iran and Odessa State Medical University, Odessa Ukraine. Samples were stored in –80ºC freezer until processed. They included tumors, farthest available normal tissue from tumors for some of the patients, benign samples and normal breast tissue samples from healthy individuals. Clinical annotations of the samples are presented in Supplementary table S1. Tissue samples were powdered using a tissue dismembrator (Micro Dismembrator from B.Braun Biotech International). Each sample was taken each time in an eppendorf tube (from –80ºC freezer) kept on dry ice. Perior to powdering, eppendorf tube containing the tissue sample was floated in liquid nitrogen (-170ºC). Frozen tissue was immediately transferred into a dry Teflon shaking flask (obtained from Sartorius stedim biotech, and always stored in 4% Sodium Hypochlorite) along with a tungsten ball (obtained from Sartorius stedim biotech, and always stored in 4% Formaldehyde). The Teflon container was then covered with its lid and fixed to a tissue dismembrator which was subsequently vibrated at 3000rpm for 20 seconds. Afterwards, the Teflon container was taken off the dismembrator and dipped in liquid nitrogen. The powdered tissue was immediately transferred from the Teflon container into eppendorf tubes such that each tube accommodated up to 30mg of the powder. The eppendorf tubes were then floated in liquid nitrogen and stored at –80ºC until used for DNA or RNA extraction.

2.2.3 Selection of CGIs and designing PCR primers to isolate them from deaminated DNA

Selection of CGIs was based on available literature. To find CGIs whose methylation patterns might be useful for breast cancer diagnosis, a review of available literature was conducted. Following this review 44 CGIs –associated with 35 protein coding genes- were selected (Table 1). The criteria applied for selection of the genes was as following;

- Genes reported to become silenced by DNA hypermethylation in breast cancer. Most of these genes had been already recognized with tumor suppressor activities.
- Genes found to be hypermethylated in other cancers and involved in growth-related pathways.
- Genes down-regulated in breast cancer and associated with a CGI in vicinity of their transcription start sites.

Table 1 Genes included in DNA methylation analysis. BC: Breast cancer; Dw-reg.: Down-regulated

Gene	Methylation/Down-regulation	Reference	Gene	Methylation/Down-regulation	Reference
14-3-3 sigma	Methylated in BC	(96)	***SFRP2***	Methylated in BC	(97)
APC	Methylated in BC	(98)	***SFRP3***	Methylated in others	(99)
BRCA1	Methylated in BC	(37)	***SFRP4***	Methylated in others -	(100)
C/EBP alpha	Dw-reg. in BC	(101)	***SFRP5***	Methylated in BC	(102)
CCND2	Methylated in BC	(103)	***WIF-1***	Methylated in BC	(104)
CST6	Methylated in BC	(105)	***HDPR1***	Methylated in others -	(106)
CDH1	Methylated in BC	(107)	***TCF4***	Dw-reg. in BC	(108)
CDH13	Methylated in BC	(109)	***PAK3***	Methylated in BC	(110)

ER	Methylated in BC	(111)	***OGDHL***	Methylated in BC	(110)
HIN-1	Methylated in BC	(112)	***hMLH1***	Methylated in BC	(113)
HOXA5	Methylated in BC	(114)	***SOCS1***	Methylated in BC	(115)
HPGD	Methylated in BC	(116)	***PTEN***	Methylated in BC	(117)
GSTP1	Methylated in BC	(118)	***DAPK1***	Methylated in BC	(119)
P53	Methylated in BC	(120)	***HIC***	Methylated in BC	(121)
PGR	Methylated in BC	(122)	***GHSR***	Methylated in BC	(123)
RAR-b2	Methylated in BC	(124)	***ARHI***	Methylated in BC	(125)
RASSF1A	Methylated in BC	(126)	***TWIST***	Methylated in BC	(127)
SFRP1	Methylated in BC	(128)			

We extracted DNA sequence of a region spanning 2kb upstream and 2kb downstream of transcription start site of each gene from human genomic DNA sequences using Ensembl database (www.ensembl.org). The putative CGIs harboured within these sequences were identified by using CGI searcher (http://cpgislands.usc.edu) (129) and Methprimer (http://www.urogene.org/methprimer) (130) softwares and crosschecking with respective previous studies.

Following bisulfite treatment, all cytosines are converted to uracils unless they are located in CpG dinucleotides and had already been methylated. Therefore, the sequence composition of DNA is different before and after bisulfite treatment. To design primer pairs that can be used for PCR-amplification of the selected CGIs from deaminated human DNA (DNA that is subjected to bisulfite treatment) we used Methprimer (130). This software creates a deaminated version of a given DNA sequence and then uses this version as a template for designing PCR primers. The criterion “Pick primers for bisulfite sequencing PCR” was chosen for all loci. This criterion avoids inclusion of CpG dinucleotides in the

primer sequences, thus preventing bias in PCR amplification that might happen due to the methylation status of the binding sites of primers. These primers are hereafter called *SoBi* primers and are listed in section 2.1.10.

2.2.4 Preparation of the control samples

The principle of performance of oligonucleotide microarrays is based on hybridization of the immobilized oligonucleotides (probes) with their complementary sequences available in the target template (target). A common feature of CGIs is their high GC-content which makes significant similarities between their sequence compositions. These similarities can lead to unspecific hybridization of target DNA samples to the probes which, in turn, necessitate the calibration of microarrays by control samples before the main screen (61). To evaluate the microarray performance and calibrate probes, we produced fully methylated and unmethylated versions of all selected CGIs (targets) *in vitro* to serve as control samples in calibration process.

2.2.4.1 Isolation of CGIs from human genomic DNA by PCR

The initial step in producing control samples was isolation of CGIs from genomic DNA by PCR. PCR primers were designed by primer3 software (http://frodo.wi.mit.edu/primer3/) (131) using 'General Primer Picking Conditions' of the software and were ordered from a supplier (Invitorgen). These primers are referred to as *genomic* primers in this manuscript and can be found in section 2.1.9. *Genomic* primers of each CGI were designed such that their PCR product spanned the region amplified by *SoBi* primers of the respective CGI. To find the best annealing temperatures for primer pairs leading to amplification of the respective specific PCR products, a gradient of annealing temperatures between 50°C and 62°C were applied in amplification process for each primer pair. The PCR products were checked on agarose gel and the appropriate annealing temperature was used for

amplification of each amplicon. In case of no amplification or non-specific PCR products in all annealing temperatures, a modified PCR protocol using Q-Solution reagent was used in gradient and final PCR procedures. Addition of Q-Solution to the PCR mixture was based on the fact that this reagent changes the melting behaviour of DNA and improves the PCR amplification of those regions that have a high degree of secondary structure or high GC-content, which is the case for the majority of CGIs (15). The fully methylated and unmethylated versions of *ARHI* and *GHSR* CGIs were amplified from EpiTect control DNA set (Qiagen) using respective *SoBi* primers.

2.2.4.2 Pools of control samples with different methylation levels

At first step, fully methylated and completely unmethylated versions of all CGIs were prepared independently. After establishment of PCR conditions, CGIs were amplified by PCR from genomic DNA using *genomic* primers resulting in an unmethylated amplicon served as unmethylated control after purification. To produce 100% methylated controls, purified PCR products were treated with CpG Methyltransferase (M.SssI; New England Biolabs) in the presence of S-adenosylmethionine (SAM; New England Biolabs). M.SssI methylates all cytosine residues (C^5) within the double-stranded dinucleotide recognition sequence 5'...CG...3'. (132). Methylation reactions were prepared according to the Table 2 and were incubated at 37°C for 3 hours followed by 20 minutes incubation at 65°C (enzyme inactivation).

Table 2. Components of Methylation reaction

Components	Volume (µl)
DNA (1µg)	#
SAM (3.2 mM)	12,5
Buffer 2[1]	5
MSss1[2]	2
ddH_2O	*
Final volume	50

#calculated for 1µg
[1]Buffer 2: 10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl2, 1 mM Dithiothreitol, pH 7.9 @ 25°C
[2]Sss1 methylase (CpG methylase)
*added to final volume of 50µl

The methylation status of the prepared control samples was then examined by incubating them with BstUI restriction endonuclease (New England Biolabs) at 60°C for 2 hours (Table 3) and subsequent analysis of digestion results by gel electrophoresis. BstUI recognizes and cuts "CG/CG" sequence -which is frequently found in CGIs- only if the cytosines are not methylated. Therefore, we expect to observe an undigested DNA band and completely digested DNA fragments for unmethylated and fully methylated control samples, respectively.

Table 3. Components of BstUI digestion reaction

Components	Volume (µl)
DNA (200ng)	*
Buffer 2	2
B*st* UI (10 units/µl)	2
ddH_2O	#
Final volume	20

*calculated for 200ng
#added to total volume of 20µl

After validation of expected methylation levels for *in vitro* prepared control samples (CGIs isolated from human genome using *genomic* primers), they were subjected to bisulfite treatment as will be described in 2.2.5.1-2 sections. Then, each locus was PCR-amplified

from corresponding fully methylated and unmethylated control samples – that had been treated by bisulfite - using *SoBi* primers. Thereafter, these amplicons were column- purified using Qiagen PCR purification kit and were quantified by Nanodrop. Thus, we produced two control bisulfite-treated samples for each CGI representing its 100% and 0% methylation levels. A pool of 0% methylated amplicons was produced by mixing same amount of all unmethylated amplicons together. Same strategy was followed to create a 100% methylation control by pooling fully methylated amplicons. In addition, using individual control samples, additional control pools corresponding to 33% methylation (33ng of each methylated amplicons + 66ng of each unmethylated amplicons) and 50% methylation (50 ng of each methylated amplicons + 50 ng of each unmethylated amplicons) and 66% methylation (66ng of each methylated amplicons + 33ng of each unmethylated amplicons) were prepared. These control pools were subsequently used for calibration and filtration of probes and monitoring the performance of microarrays.

2.2.5 Sodium bisulfite treatment of DNA samples

The control amplicons containing methylated and unmethylated CGIs as well as DNA samples isolated from the patient tissue specimens were subjected to bisulfite conversion. By treating DNA with sodium bisulfite, one can make use of selective conversion of unmethylated cytosines to uracils by this treatment. The bisulfite treated DNA when used as a template for PCR amplification, the uracils in the template will be subsequently replaced by thymine in the amplified product.

2.2.5.1 Deamination of DNA by bisulfite

The treatment of DNA samples was performed using EpiTect Bisulfite Kit (Qiagen) which provides a relatively fast 6-hour procedure for efficient DNA deamination and purification.

The reaction mixtures were prepared according to table 4 and were incubated in the thermocycler with thermal program presented in table 5.

Table 4. Reaction contents in Bisulfite treatment of DNA

Component	Volume per reaction (µl)
DNA solution (1µg)	Variable*
Rnase-free water	Variable*
Bisulfite Mix	85
DNA Protect Buffer	35
Total Volume	140

*The combined volume of DNA and Rnase-free water must total 20µl

Table 5. Incubation prolife of bisulfite treatment reaction

Step	Time	Temperature
Denaturation	5 min	99°C
Incubation	25 min	60°C
Denaturation	5 min	99°C
Incubation	85 min	60°C
Denaturation	5 min	99°C
Incubation	175 min	60°C
Hold	Indefinite*	20°C

2.2.5.2 Purification of deaminated DNA

The EpiTect kit provides necessary materials to clean up the DNA after bisulfite treatment. Briefly, 560µl freshly prepared buffer BL containing 10µg/ml carrier RNA was added to each converted sample and was mixed. The whole mixture was then transferred to an EpiTect spin column and was centrifuged at maximum speed for one minute. The flow-through was removed and column (containing the DNA) was washed with 500µl buffer BW (wash buffer). After removing flow-through, 500µl buffer BD (desulfonation buffer) was added to the column and column was incubated for 15 minutes at room temperature, and then was centrifuged. Flow-through was discarded and column was washed twice with 500µl buffer BW. Column was then placed into a new collection tube and was centrifuged for 2 minutes to remove any residual liquid. Subsequently, DNA was eluted from column in 60µl of elution buffer (10mM Tris-HCL, pH: 8.5) and stored in -20°C. All centrifugation

steps were carried out at 13000 rpm and room temperature for 1 minute unless otherwise mentioned.

2.2.6 Isolation of selected CGIs from patients DNA

DNA isolated from each patient sample was bisulfite treated and was used as a template to PCR-amplify all the 44 CGIs (amplicons) pertaining to the 35 pre-selected genes using *SoBi* primers. Considering the variations between the annealing temperatures of primers related to different CGIs, a PCR program combined of touchdown and gradient PCR procedures was established such that all the 44 amplicons could be amplified at once. Amplification program included activation of Hotstar Taq DNA polymerase at 95ºC for 15 minutes, denaturation at 95ºC for 1 minute, annealing at 65ºC for 2 minutes. This process is continued for 7 cycles going down with the annealing temperature by 1ºC in every cycle. This touchdown amplification was followed by a gradient PCR program consisting of denaturing at 95ºC for 40 seconds, annealing temperature ranging from 48ºC – 62ºC for 30 seconds and 30 seconds with 72ºC elongation temperatures for 38 cycles. The reactions were carried out using a gradient thermocycler (MJ Research, PTC-200, Peltier thermal cycler) in 200µl PCR tubes (Thermo Scientific). All the reactions carried out were 50µl PCR reactions in 96 well plates. 2µl converted DNA was loaded in each well and PCR mastermix comprised of dNTPs (Fermentas Life Sciences), PCR water (Aqua ad iniectabilia, Braun), Hotstar Taq DNA polymerase and PCR buffer (Qiagen) was added to the wells. The primers were then added individually to each reaction. The amounts of material and reagents used in PCR amplification are summarized in table 6. PCR products were then examined on 1% agarose gel.

Table 6. Reaction contents used for PCR amplification of CGIs from patients' samples

Components	Volume (µl)
10x PCR buffer	5
ddH_2O	33.7
dNTPs (2.5 mM each)	4
Forward primer (10 µM)	2.5
Reverse primer (10 µM)	2.5
HotstarTaq Polymerase	0,3
Total	**49**

Following evaluation of amplification of the selected CGIs, the amplicons were purified using 96 well format purification plates (Millipore 96 PCR purification plate) and vacuum manifold. The concentrations of the purified PCR products, was measured using a Nanodrop device (ND-1000 spectrophotometer, Peqlab, Biotechnologie GmbH). Approximately 10ng of each purified amplicon was pooled with other amplicons in a clean 1.5ml eppendorf tube. The pool of all amplicons of each patient was used as a template for labeling.

2.2.7 Labeling of DNA by random priming

To label the DNA for microarray experiments, random priming in the presence of biotinylated uracil was used. The DNA that is to be labeled is denatured by heating and to this denatured DNA, hexamer mix (mixture of short oligonucleotide of 6 nucleotide long) is added. The hexamers bind randomly to the single stranded DNA and the region between two hexamers is then synthesized by the polymerase enzyme used. During this process, the biotinylated nucleotides will be incorporated into the newly synthesized DNA. Following hybridization, this DNA can be detected using fluorescently labeled molecules that have a high affinity for biotin, like Streptavidin. The polymerase used for this process was Klenow Fragment. This enzyme is an N-terminal truncation of DNA polymerase I which retains polymerase activity but has lost the exonuclease activites.

Before labeling, volume of the polled amplicons in the eppendorf tube was adjusted to 26.4µl with DEPC water. The content from the 1.5ml eppendorf tube was transferred to a 200µl PCR tube. The DNA was then incubated at 99ºC for 5 minutes in a thermocycler. In the mean time, a master mix consisting of Random Priming (RP) buffer (febit), dNTPs (Fermentas Life Sciences), biotinylated uracil (Biotin-16- dUTP, Roche Diagnostics GmbH), bovine serum albumin (BSA, Fermentas Life Sciences) and Klenow fragment (New England Biolabs) was prepared according to table 7. Following the denaturation of the pooled DNA, the tubes were immediately transferred onto ice. To this denatured DNA pool, 1.5µl of hexamers mix and 12.1µl of the previously prepared master mix were added. This mixture (40µl in volume) was then incubated at 37ºC for 3 hours in a thermocycler to facilitate the amplification of DNA template (mixture of all the amplicons). Reaction tubes were then incubated for 10 minutes in another thermocycler which was pre-heated to 75ºC to deactivate the polymerase (Klenow fragment).

Table 7. Reaction contents used together with hexamers in random priming procedure

Component	Volume (µl)
RP buffer (5x)	8
dNTPs*	1
BSA solution (10 mg/ml)	1.6
Klenow fragment (5 U/µl)	1.5
Final volume	12.1

*dNTPs mixture consists of dATP, dCTP, dGTP (each 2 mM), dTTP (1.3 mM) and biotinylated dUTP (0.7 mM).

Afterwards, the reaction mixture was transferred to a 1.5ml eppendorf tube. To precipitate the DNA, 4µl of sodium acetate (3M, pH: 5.6) along with 100µl of 100% ethanol (pre-cooled at -20ºC) were added to the mixture and the tube was stored at -20ºC overnight or until the sample was used for hybridization.

2.2.8 Microarrays for DNA methylation analysis

The first step in this study was to establish oligonucleotide microarrays capable of screening DNA methylation patterns of the pre-selected CGIs with the highest possible resolution (as many as possible CpGs of a CGI). To this aim, the Geniom One microarray technology of febit Company was used. This technology enables the user to design and synthesis the customised microarrays on Biochips provided by the company. In addition to the chip synthesis, washing and detection procedures could also be performed within the same device.

2.2.8.1 Designing microarrays

Oligonucleotide probes were designed using MetPat software (febit). Probe sets were designed to analyze methylation status of all CpG dinucleotides within each CGI. Methylation of each CpG was assayed by two probes; one of which complementary to the unmethylated CpG (U; detecting TG) and the other one was matched to the methylated CpG (M; detecting CG). For chip evaluation experiments, probes were designed with different lengths of 17, 21 and 25 nucleotides. Each probe was designed such that the target C (of a CpG nucleotide) could be addressed by the nucleotide located at the central position of the probe. After probe design, list of probes for a given CGI was exported as a text file including the probes name and sequences and was stored for array design. The array design was performed by cumulating the probe sets of all CGIs (analysis probes) and addition of control probes. The control probes acted as positive controls for the occurrence of synthesis and hybridization or represented entirely independent sequences, which indicated the background signal of unspecific binding. Each array could accommodate 6676 probes including 280 probes as positive and 80 probes as negative controls, thus providing capacity for 6316 analysis probes. As analysis probe set was synthesized at least three times

(replicates) on each array, up to 2104 different analysis probes could be synthesized on each array. Array design was loaded into database of the Geniom One system.

2.2.8.2 Generation of microarrays

Oligonucleotide arrays were generated by photo-controlled in situ synthesis within the Geniom device on an activated three-dimensional reaction carrier consisting of a glass-silicon-glass sandwich (biochip) provided by febit. The synthesis procedures and conditions had been already established in the device (133, 134). Each biochip consists of eight individually accessible micro-channels, each of which is referred to as array in this manuscript. The arrays are connected to the micro-fluidic system of the geniom device. Oligonucleotides can be synthesized in parallel in all eight arrays of one biochip.

Prior to synthesis, the glass surface is activated by coating with a spacer. Then, arrays are filled with the four nucleoside phosphoramidite monomers dA, dT, dC and dG, sequentially. Phosphoramidite is a phosphite that has a NR_2 group (R: isopropyl) instead of OH group. The phosphoramidites are protected by the photolabile 2-(2-nitrophenyl)-propoxycarbonyl (NPPOC) group (Fig. 8). Due to the presence of this group, the binding of next phosphoramidite to the 5' end of previously added one is prevented. In each cycle, only distinct areas (spots) to which next phosphoramidite should be added are illuminated. Upon excitation, the photoprotecting groups of the already existing, partial oligonucleotide sequences at these locations, synthesized earlier during the process, are removed and the oligomers are extended by another phosphoramidite after adding the relevant monomer (Fig. 9). There is also another protective group, phenoxyacetyl (Pac), which protects amino groups of the bases with NH_2 groups (Fig. 8). This group is removed at the end of synthesis. All materials and reagents necessary for synthesis were provided by febit in individual bottles and were loaded in special places within the Geniom instrument before starting the synthesis procedure (Fig. 10).

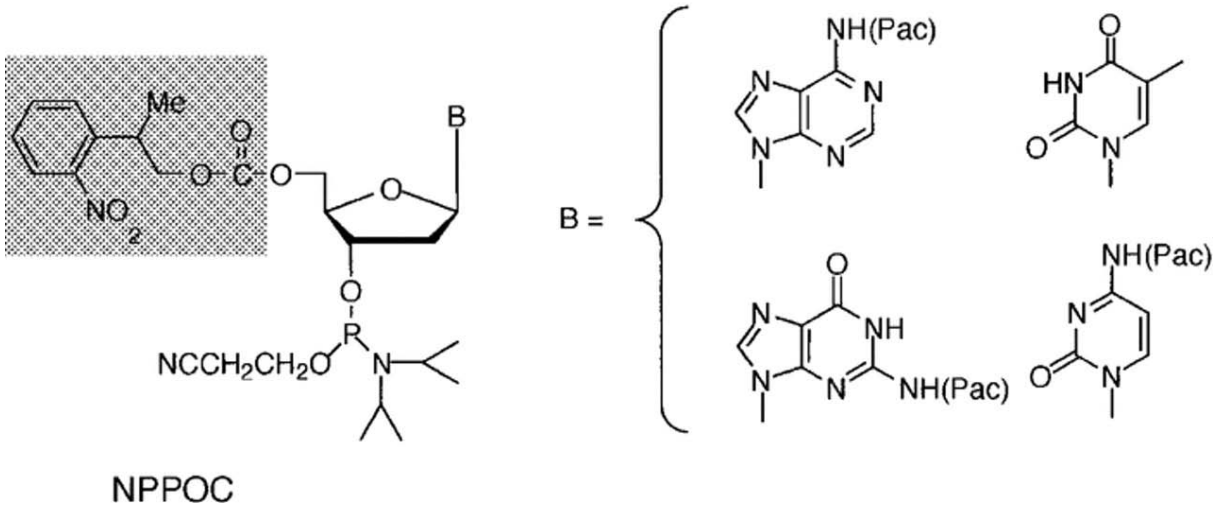


Figure 8 Photolabile protecting group attached to phosporamidite. NPPOC is highlighted with gray background (Adapted from Beier & Hoheisel (133)).

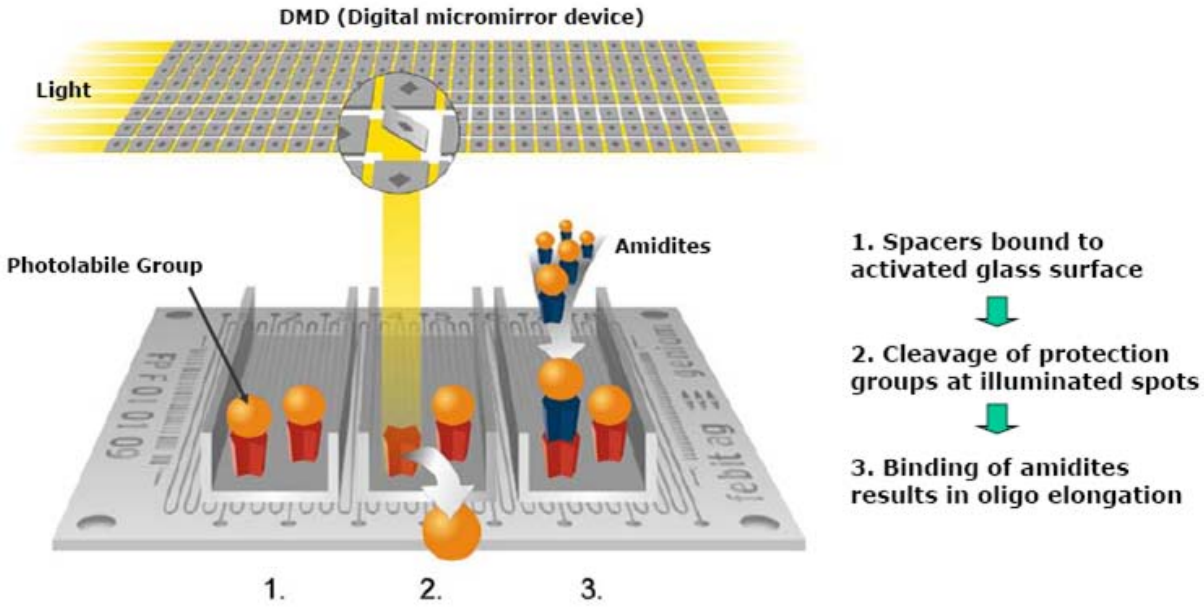


Figure 9- Photo-controlled *in situ* synthesis within the Geniom device. According to the array design, micromirror system of the device directs light to activate the specific spots to which amidites should bind (Provided by febit).

Upon loading the instrument with chemical sets, the synthesis module of the Geniom is activated through Geniom software and array design (already transferred to the Geniom database) is selected as the template for oligo synthesis. Geniom software controls oligo-synthesis steps and dictates the movement of Digital Micromirror Device (DMD) according to the array design. The cyclic process of DNA synthesis steps (photo-activation, flow of

amidites and reagents to the arrays) continues until all oligos are produced. After completion of probe production, the synthesized chip can be stored at 4°C until hybridized.

2.2.8.3 Array hybridization and post-processing

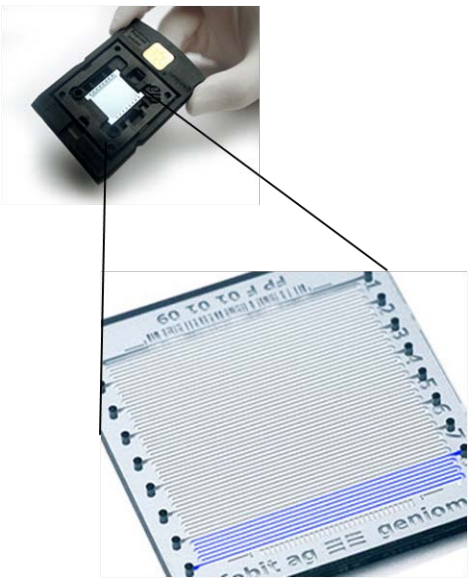
The random primed DNA precipitated at -20ºC was taken out and centrifuged at 15300 rpm and 4ºC for 20 minutes. The supernatant was removed and the DNA pellet was air-dried. The pellet was then dissolved in 3.75µl DEPC-treated water and 11.25µl of hybridization mixture (provided by febit). All the CGIs amplified from one patient DNA were hybridized to one array of a biochip following pooling and labeling.

Before hybridization, synthesized chip was imported into the machine and washed with water at 25°C for 15 minutes. This step is to take care that all the probes are straightened and not collapsed. Hybridization was carried out outside the geniom machine in an external hybridization holder (febit). Initially a 15µl pre-hybridisation buffer (provided by febit) was filled into each array and the chip was incubated for 15 minutes at room temperature. Meanwhile, the labeled DNA dissolved in hybridization mixture was incubated at 99ºC for 5 minutes (for denaturation) and then was kept on ice until loaded into the arrays. Then, pre-hybridization buffer was removed from the arrays and the labeled DNA samples were filled into the arrays and the chip was incubated at 45ºC for 16 hours in a hybridization chamber (febit). Following the incubation, target DNA samples were removed from the chip and the chip was imported into the geniom machine where further washing of the chip and detection were carried out.

To detect the hybridization, the high affinity of streptavidin for biotin (incorporated in the form of biotinylated uracil during random priming procedure) by using the fluorescently labeled streptavidin SAPE, was exploited. SAPE (streptavidin, R-phycoerythrin conjugate,

A

B



C

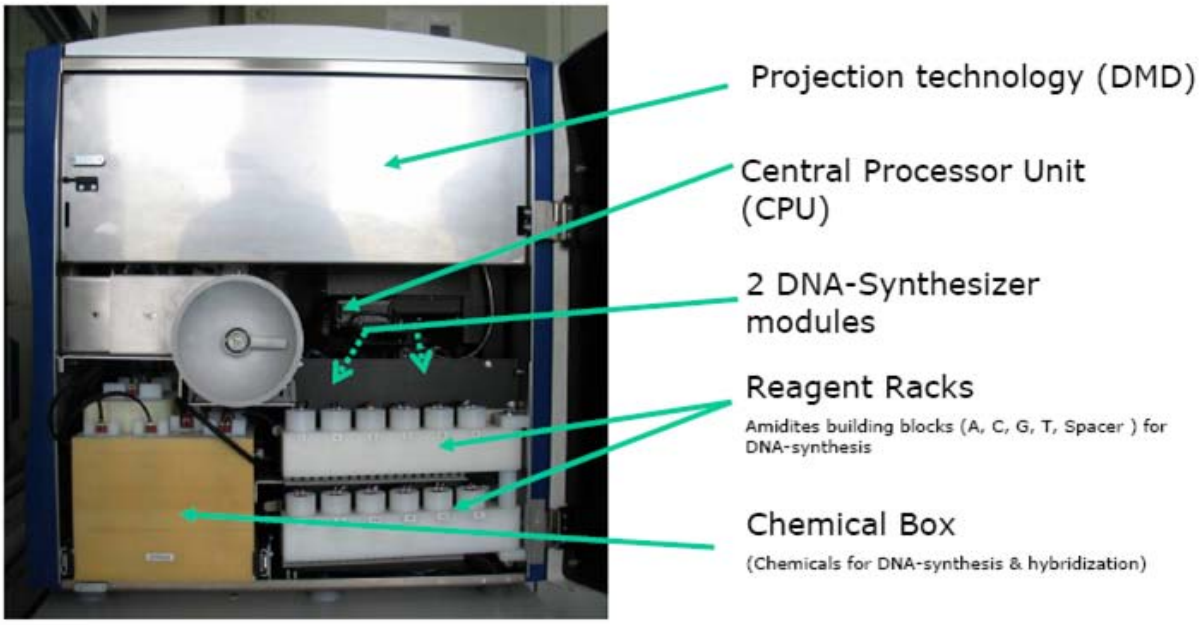


Figure 10. Geniom One instrument. (A) Front view. **(B)** Biochip consisting of 8 arrays. **(C)** Different parts of the Geniom machine.

molecular probes, invitrogen detection technologies) solution was prepared by mixing 44µl of SAPE (1mg/ml) with 9ml of 6× SSPE buffer and was loaded into the geniom instrument. Likewise, the machine was supplied with washing buffers (6× SSPE and 0.5× SSPE). The arrays were first washed with the stringent 0.5× SSPE buffer (prepared from the 20× SSPE buffer, Sigma) (four times; 5 seconds flow, 30 seconds hold) at 45ºC. Then the arrays were filled with SAPE solution and incubated for 15 minutes at 25ºC. Afterwards, washing with non-stringent 6× SSPE buffer was performed at 25ºC (four times; 5 seconds flow, 30 seconds hold). Finally hybridization signals were detected by charge coupled device (CCD) camera system of the geniom using a Cy3 filter. The exposure time for signal detection was calculated automatically by the device based on strength of signals of all spot on the chip providing informative data for vast majority of the spots while avoiding the occurrence of saturation of strong signals (Maximum intensity = 65000). An image of the chip was created and saved by the detection system. This image was first inspected for successful occurrence of probe synthesis and hybridization by verifying the signals from analysis and control probes in different positions of all arrays (Fig. 11). The image was then analyzed by the detection system of the device for calculating signal intensities of all spots. The primary data consisting of intensity values of individual spots were generated by the detection module of the instrument and saved in the database of geniom.

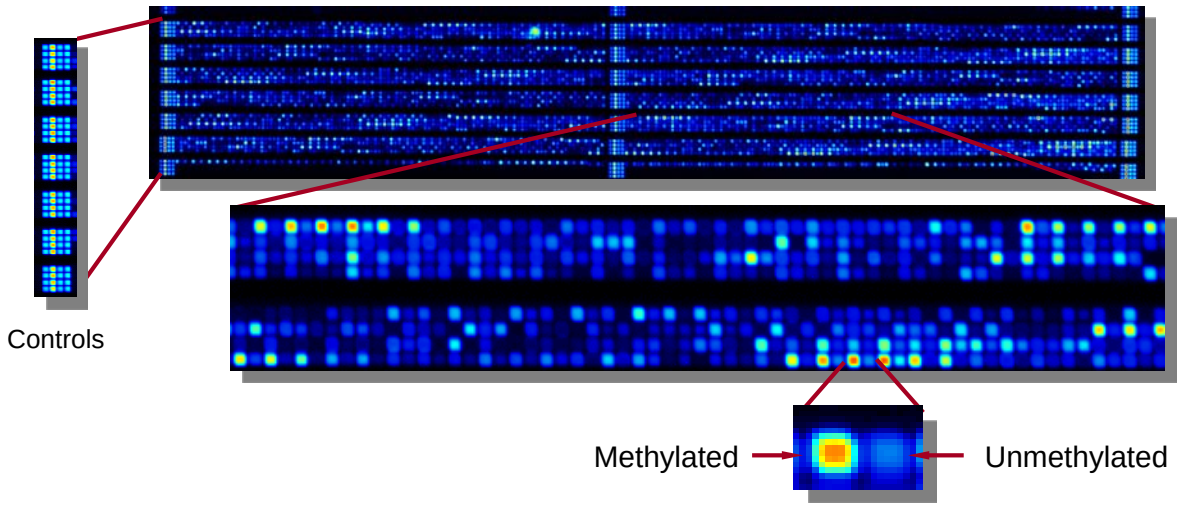


Figure 11. Image taken from a biochip after detection. Controls and a pair of spots related to one CpG site, as an example, are shown.

2.2.9 Data analysis

The signal intensity values measured by the detection module of the geniom machine were used as raw data to analyze the DNA methylation patterns. The data were first exported from geniom database in format of text files consisting of all probe names and median signal intensity of each probe. This data were then process using Microsoft Excel. At the first step, 2% of the lowest signal intensities from each array were taken and their mean was considered as the background signal. This value was subtracted from all the signals of this array to avoid the false positive signals. Then a normalized local methylation index (MI) was defined for each CpG by dividing the signal intensity value of its methylated probe (M) to sum of intensity values of its methylated (M) and adjacent unmethylated (U) probes (MI=M/M+U). Since probe pairs (M and U of each CpG) had been synthesized in triplicate in different positions of each array, three local MIs were defined for each CpG. The MI of each CpG was calculated as the average of the its three local MIs and considered as methylation level of that CpG in the hybridized DNA sample. These normalized data were next used for evaluating array performance or identification of methylation signature.

Probe selection was carried out following calibration of probes based on hybridization of arrays with pools of control samples. Only the probes that could show an appropriate discriminative power between different methylation levels of control samples were considered for further analyses. The methylation indices shown by these probes from each of the patients sample were put together. This combined data set was then used for identification of diagnostic methylation signature. Data analysis towards the identification of the methylation signature was performed in collaboration with Amin Moghaddas Gholami from the Division of Functional Genome Analysis, DKFZ. To identify differentially methylated CpGs between tumors and control samples, datasets were analyzed with the M-CHiPS (Multi-Conditional Hybridization Intensity Processing System) software (135). The data were used to compute p-values via the limma package (136). The limma package uses an approach called linear models to analyze designed microarray experiments. This approach allows very general experiments to be analyzed as a simple replicated experiment. It requires one or two matrices to be specified. The first is the design matrix which indicates which samples have been applied to each array. The second is the contrast matrix which specifies which comparisons of interest between the samples. For very simple experiments, there is no need to specify the contrast matrix. The approach starts by fitting a linear model data which fully models the systematic part of the data. The model is specified by the design matrix. Each row of the design matrix corresponds to an array in the experiment and each column corresponds to a coefficient which is used to describe the samples in the experiment. Consequently there would be as many coefficients as samples in the experiment, thus any set of independent coefficients will describe all treatments. The main purpose of this step is to estimate the variability in the data; hence the systematic part needs to be modeled so it can be distinguished from random variation.

In limma, systematic methylation effect for each gene is described by a linear model using a design matrix and a gene-specific vector of regression coefficients. A complete pairwise-

comparisons contrast matrix was set up to test differential methylation across all samples. In limma, the regression coefficients represent comparison of interest between samples in the experiment. These coefficients are estimated using a least squares linear model fitting procedure and tested for differential methylation with moderated Student's t-statistic via the empirical Bayesian statistics described in the limma package. P-values computed for the *F-statistic* (136) were adjusted for multiple testing to control the False Discovery Rate (FDR) at 5% (137). The adjusted p-values can serve to accept or reject the null-hypothesis based on a significance level.

2.2.10 Bisulfite-sequencing

Bisulfite sequencing was applied to verify the methylation patterns of interesting candidate genes in more details. This procedure included different steps of ligation of the PCR product to the vector, transformation of competent bacteria with the vector, bacteria culture, colony picking and finally the plasmid preparation and sequencing.

2.2.10.1 Ligation of amplicons and transformation in competent bacteria

The ligation reaction and transformation were done according to the protocol of TOPO TA Cloning® Kit for Sequencing (Invitrogen). This kit provides a highly efficient, 5 minutes, one-step cloning strategy for direct insertion of Taq polymerase-amplified PCR products into a vector for sequencing. The plasmid vector is supplied linearized with single 3´deoxythymidine overhangs and a Topoisomerase which is covalently bound to the vector. The Taq polymerase has a non-template-dependent terminal transferase activity which adds a single deoxyadenosine to the 3´end of PCR products. Together with the overhanging 3´deoxythymidine residue of the vector the ligation of the PCR products is performed.

The PCR product was ligated into the pCR®4 TOPO® vector (Invitrogen). For ligation reaction the components described in table 8 were mixed together in a 200µl PCR tube.

Table 8. Contents of TOPO cloning reaction

Component	Volume (µl)
Fresh PCR product	4
Salt solution	1
TOPO vector	1
Final volume	6

The reaction was mixed gently and incubated for 5 minutes at room temperature. Transformation of bacteria and purification of plasmid were performed as described in section 2.2.1.1.

2.2.11 Cell culture

2.2.11.1 Maintenance of cells

Cell lines were maintained at 37°C and 5% CO_2 in the following media: HEK293T, MCF7, T47D, SKBR3, MDA-MB 231, MDA-MB-435 and HDQ-P1 in DMEM with 4.5g/L D-glucose (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (Invitrogen), 100 units/mL penicillin (Invitrogen) and 100µg/mL streptomycin (Invitrogen). MCF-10A and MCF-12A in Mammary Epithelial Cell Growth medium (Provitro) supplemented with 10ng/mL human recombinant Epithelial Growth Factor, 0.5µg/mL hydrocortisone, 5µg/mL insulin, 52µg/mL bovine pituitary extract (all supplements from Provitro), 100units/mL penicillin (Invitrogen) and 100µg/mL streptomycin (Invitrogen). All cells were passaged every 2-3 days cells and diluted 1:4 in fresh medium. Before experimental use, MCF7 and MELN (138) cells were grown for 96 h in phenol red-free DMEM with 4.5g/L D-glucose (Invitrogen) supplemented with 10% dextran-coated charcoal-treated (DCC)-FBS (Invitrogen). Before hormonal treatment with 10nM E2 (Sigma), serum-free medium was added to the cells for 24 hours and cells were kept in this medium until the end of the experiment.

2.2.11.2 Transfection of cells

Chemical transient transfections using Effectene transfection reagent (Qiagen) for transfecting DNA molecules or siPORT™ *NeoFX*™ (Ambion) for transfecting RNA molecules were performed in 24 or 96 well plates. Effectene-based transfections were carried out in 24 well plates using 4×10^4 cells per well and 100ng/well plasmid DNA in a final volume of 400µl culture medium. Ectopic expression was performed for RASD1 using a pc-myc-CMV-D12 expression vector containing RASD1 open reading frame (ORF).

Reverse transfection using siPORT™ *NeoFX* involves the simultaneous transfecting and plating of cells. siPORT™ *NeoFX* transfection agent and RNA molecules are mixed, incubated, distributed to culture wells, and overlaid with cells. Transfection complexes are active and stable in the presence of serum, therefore, in the absence of a cellular stress response, there is no need to remove or replace media after transfection. In the this study, siPORT™ *NeoFX*™-based transfections were carried out in either 96 well plate using 6×10^3 cells per well in a final volume of 100µl culture medium or 24 well plate using 4×10^4 cells per well in a final volume of 500µl culture medium and final concentration of 50nM RNA molecules.

2.2.11.3 Decitabine treatment

Decitabine is a compound that can induce hypomethylation by inhibiting DNMTs (21). One million cells were seeded in 75cm^2 cell culture flask in appropriate medium. After 24 hours, medium was replaced with fresh medium alone or medium containing 0.5µM decitabine. Following 72 hours incubation, DNA and total RNA were isolated from cells as described previously.

2.2.12 Chromatin Immunoprecipitation (ChIP) assay

ChIP-assays were performed by Achim Breiling from Division of Epigenetics as collaboration. All steps were carried out as described previously (139). In brief, cells were fixed in the cell culture dishes by adding fixation solution directly to the medium. Cells were fixed at room temperature for 10 minutes. Fixed cells were washed once with PBS, scratched of the plates and transferred into falcon tubes. Chromatin was then prepared and immunoprecipitated as described using antibodies specific for H3K9me2 (Abcam ab1220) and H3K4me2 (Upstate 07-030) acetylated H3 (Upstate 06-598) and H4 (Upstate 06-599), ZEB1 (Santa Cruz H-102, sc25388X), POLII (8WG16, Abcam ab817), CTCF (Abcam ab70303) and ERa (Santa Cruz HC-20, sc543X). Immunoprecipitates were eluted into 25µl of TE buffer (10 mM Tris-HCL pH 8, 1 mM EDTA) and one µl was analysed by real time PCR using ChIP primers (100nM final concentration, section 2.1.11) in a 10 µl PCR reaction consisting of 6µl Absolute QPCR SYBR Green Mix (Thermo Scientific) in Roche LightCycler 480. PCR conditions were similar to that explained before. Cycle threshold numbers for each amplification were measured with the LightCycler 480 software and enrichments were calculated as percentage of the input.

2.2.13 Expression arrays

2.2.13.1 Illumina mRNA expression array

Three major expression profiling platforms (Affymetrix, Agilent, Illumina) are commercially available. In this study the Illumina platform for expression profiling has been used (service is available at core facility of DKFZ). Illumina's BeadArray technology is based on 3µm silica beads that self assemble in microwells on either of two substrates: fiber optic bundles or planar silica slides. When randomly assembled on one of these two substrates, the beads have a uniform spacing of ~5.7µm. Each bead is covered with

hundreds of thousands of copies of a specific oligonucleotide representing one unique cDNA sequence that act as the capture sequences in the Illumina assay. Characteristics of the BeadArray technology are quality control of every single array and the high feature-redundancy (>20× average) providing high-confidence results. The BeadChip array format used by Illumina utilizes either 6, 8, or 12 different arrays on a glass slide, which are processed simultaneously. In the current study, we applied 6 arrays/chip, each representing > 48,000 probe sequences.

2.2.13.1.1 cRNA synthesis, purification and labeling

In this step, 250ng of total RNA was used for complementary DNA (cDNA) synthesis, followed by an amplification/labelling step (in vitro transcription) to synthesize biotin-labeled cRNA according to the MessageAmp II aRNA Amplification kit (Ambion). Biotin-16-UTP was purchased from Roche Applied Science. The cRNA was column-purified according to TotalPrep RNA Amplification Kit's manual (Ambion) and eluted in 60µl of water. The quality of cRNA was controlled using the RNA Nano Chip Assay on an Agilent 2100 Bioanalyzer and spectrophotometrically quantified using a NanoDrop spectrophotometer.

2.2.13.1.2 Probe labeling and Illumina Sentrix BeadChip array hybridization

Biotin-labelled cRNA samples for hybridization on Illumina Human Sentrix-6 BeadChip arrays (Illumina) were prepared according to the protocol of Illumina's recommended sample labeling procedure (140). Hybridization was performed at 58°C in GEX-HCB buffer (Illumina) at a concentration of 50ng cRNA/µl, unsealed in a wet chamber for 20 hours. Spike-in controls for low, medium and highly abundant RNAs were added, as well as mismatch control and biotinylation control oligonucleotides. The microarrays were washed twice in E1BC buffer (Ambion) at room temperature for 5 minutes. After blocking

for 5 minutes in 4ml of 1% (w/v) Blocker Casein in phosphate buffered saline Hammarsten grade (Pierce Biotechnology), array signals were developed by a 10-minutes incubation in 2ml of 1µg/ml Cy3-streptavidin solution and 1% blocking solution. After a final wash in E1BC buffer, the arrays were dried and scanned.

2.2.13.1.3 Scanning and data analysis

Microarray scanning was done using a Beadstation array scanner, setting adjusted to a scaling factor of 1 and PMT settings at 430. Raw data extraction was performed using the Beadarray R package (svn release 1.7.0) from bioconductor (www.bioconductor.org). Then, outliers were removed when their expression value dropped below a threshold: the median + *MAD (median absolute deviation) expression of all negative control beads. Individual bead types were also flagged as filtered when their bead replicate count dropped below 17. All data were then used for the mean expression value calculations within Beadarray. Finally, a bead type was discarded when the bead type's filter flag was set across all samples. Data analysis was carried out by variance stabilizing and robust spline normalization of the remaining signals using the algorithms from the Lumi R package (release 1.1.0 from bioconductor). Subsequently, values of normal samples were subtracted from values of the tumors. Genes with positive and negative resulting values were considered as up- and down-regulated genes in tumors, respectively.

2.2.13.2 Array-based miRNA profiling

miRNA profiles were generated by using febit's Geniom Realtime Analyzer. The array contained 7 replicates of each human miRNA as annotated in the Sanger miRBase 11.0 (141). Briefly, 3µg of total RNA containing small RNAs were labeled using the FlashTag RNA kit (Genisphere). Following overnight hybridization for 16 hours at 42°C, the biochip was washed automatically as indicated by the supplier and signal intensities were calculated

using the Geniom Wizard Software (febit). All further statistical analyses were provided by febit company using R. Following background correction, the seven replicate intensity values of each miRNA were summarized by their median value. To normalize the data of different arrays, the Variance Stabilizing Normalization (142) was applied by the R 'vsn' package, such that the miRNA profiles were homoscedastic. This normalization transformed the background subtracted raw data, ensuring that the variance was almost constant. Differentially expressed miRNAs between cell line models were identified by using the t test procedure within Significance Analysis of Microarrays (SAM) (143).

2.2.14 Functional analysis

2.2.14.1 Cell proliferation assay

The CellTiter-Glo® Luminescent cell viability assay is a method of determining the number of viable cells in culture based on quantitation of the ATP present, an indicator of metabolically active cells representing the viability and proliferation of cells. The Cell Titer Glo Luminescent Cell Viability Assay was used to verify the viability of the transfected MCF7 cells in opaque-walled multiwell plates (Costar). 48 and 72 hours after transfection, 100µl of CellTiter-Glo® Reagent (Promega) was added to each well of 96 well plates in order to measure the ATP content of the well. After 10 minutes incubation at RT, luminescent signals were measured using a luminometer. Each assay was performed in five replicates. An integration time of 0.25 second per read was applied.

For cell counting, 72 hours post-transfection of cells in 6-well plates (3×10^5 cells per well), the cells were trypsinized and live cells were counted by a cell viability analyzer (Beckman Coulter).

2.2.14.2 Caspase activity measurement using the Caspase-Glo® 3/7assay

The Caspase-Glo® 3/7 Assay (Promega) is a luminescent assay that measures caspase activity. The assay provides a proluminogenic caspase 3/7 substrate in a buffer system optimized for caspase activity, luciferase activity and cell lysis. Luminescence is proportional to the amount of active caspases. Caspase activity assays were performed in 96 well opaque-walled multiwell plates. MCF7 cells were reverse transfected (6×10^3cells/well) with RNA molecules. After incubation for 48 or 72 hours, 100µl of Caspase-Glo reagents for caspases 3/7 were added to each well, incubated for 30 minutes, and the luminescent signal was measured using a luminometer.

2.2.14.3 Luciferase reporter assays

2.2.14.3.1 ERE-Firefly luciferase reporter gene assay in MELN cells

Genetic reporter systems are widely used to study regulation of eukaryotic gene expression. In order to assess a possible role of miR-375 in ERα signaling, ERα protein as well as transcriptional activity levels were measured following ectopic expression of miR-375 in MELN cells. MELN are MCF7 cells which stably express an estrogen responsive element (ERE)-controlled firefly luciferase reporter gene. MELN cells (3×10^5 cells/well in 6-well plates) were reverse-transfected with Pre-miR-control or Pre-miR-375 and incubated for 48 hours before stimulation. Cells were treated or not for 16 hours with 10nM estradiol, washed and lysed with 150µl/well (10 minutes at RT) of Luciferase Cell Culture Lysis Reagent (Promega). After centrifugation (13,000 rpm for 15 minutes), the supernatant was collected and luciferase activity was analyzed using the firefly luciferase assay system from Promega and a luminometer, in triplicates. Luciferase results were normalized to the corresponding protein concentration of cell lysates.

2.2.14.3.2 Dual-luciferase reporter assays in HEK293T and MCF7 cells

The term ‘dual’ refers to the simultaneous expression and measurement of two individual reporter enzymes within a single system. The experimental reporter (*firefly*-luciferase) correlates with specific experimental conditions, while the activity of the co-transfected ‘control’ (*renilla*-luciferase) provides an internal control that serves as a baseline for transfection efficiency. Normalizing the activity of *firefly*-luciferase to the activity of *renilla*- luciferase minimizes experimental variability caused by differences in cell viability or transfection efficiency.

Estrogen receptor activity reporter assay

For analyzing estrogen receptor transcriptional activity in MCF7 cell, cells were reverse transfected with RNAs at a density of 4×10^4 cells/well in 24-well plates. After 24 hours, cells were co-transfected with 50ng/well ERE-Firefly and 10ng/well Actin-Renilla luciferase reporters and 24 hours later reporter activities were assayed by Dual Luciferase Reporter Assay System (Promega).

Reporter assay for miRNA target identification

Segments of the 3' untranslated region (3' UTR) of *RASD1* (600 bp) and *EBF3* (495 bp) containing target sites for miR-375 were amplified from genomic DNA with primers listed inpresented in section 2.1.12. Each amplicon was cloned into the multiple cloning site (MCS) of a pMIR-REPORT Luciferase vector (Ambion) located at the 3'UTR of the *firefly* luciferase gene. To be able to clone the UTRs into the MCS of the vector, recognition sites of SpeI and HindIII endonucleases had been added to the 5' sites of forward and reverse primers, respectively. Following PCR amplification, amplicons and pMIR-REPORT vector were cut with above mentioned restriction enzymes. The digestion reactions were prepared as described in table 9. Reactions were incubated at 37°C for 2 hours and subsequently at 65°C for 20 minutes. Digestion products were analyzed on 0.8% agarose gel and desired

bands were cut from the gel on a UV lamp. DNA was purified from gel using Qiagen gel purification kit. To ligate the cut PCR product to the cut vector, T4 DNA ligase kit (Invitrogen) was used. Ligation reaction was prepared according to the table 10 and was incubated at 22°C for 4 hours.

Table 9. Reaction contents of DNA digestion with SpeI and HindIII.

Components	Volume (µl)
DNA (amplicon/vector)	5µg
ddH_2O	Variable
NEB Buffer 2	5
BSA (2mg/ml)	5
SpeI (10 units/µl)	1
HindIII (20 units/µl)	1
Total	**50**

Table 10. Reaction contents of ligation of 3'UTRs into reporter vector

Components	Volume (µl)
5x ligation buffer	4
T4 ligase	1
Vector	1
Amplicon	10
ddH_2O	4
Final volume	**20**

For miRNA target identification, lusiferase reporter assays were performed in HEK293T and MCF7 cells. MCF7 and HEK293 cells were transfected with Anti-miR-375 and Pre-miR-375, respectively, in quadruplicate in a 24-well plate (4×10^4 cells/well). Reverse transfections with Pre-miR and Anti-miR negative controls were performed as well. One day later, cells were co-transfected with 50ng pMIR-REPORT constructs containing *RASD1* or *EBF3* 3' UTRs and 10ng Actin-*Renilla* Luciferase expression vector. After 24 hours, reporter activities were assayed by Dual Luciferase Reporter Assay System (Promega) as following; transfected cells grown in a 24-well plate were lysed with 100µl of

1× passive lysis buffer (Promega) for 15 minutes at RT on a shaker (310 rpm). 20µl of each lysate were pipetted into a reader plate of the luminator in triplicate. The *firefly*-luciferase activity was measured by adding 100µl of LAR II reagent (Promega) to generate a stabilized luminescent signal. After quantifying the *firefly* luminescence, this reaction was quenched and the *renilla*-luciferase reaction was initiated by adding 100µl of Stop- and Glo reagent to the same well. The Stop- and Glo reagent (Promega) produces a stabilized signal of the *renilla*-luciferase, which was measured in the luminometer (Mithras LB 940 luminometer, Berthold Technologies). After normalizing the *firefly*-luciferase activity to the *renilla*-luciferase activity, a direct comparison of samples regarding firefly activity in different conditions was possible.

2.2.15 Statistical analysis

Data were presented as mean values ± SD from three to five independent experiments. Statistical significance was determined using student's t-test to compare groups of data. *P*-values <0.05 were considered as statistically significant.

3 Results

3.1 Microarray based analysis of DNA methylation patterns

In the present study, the discovery of a DNA methylation signature marker for breast cancer based on screening of DNA from breast tissue samples was investigated. As interrogating the methylation status of every CpG of several candidate loci was sought after, a specialized microarray that combines high throughput and high resolution had to be developed. Then, the microarray was used to analyze the methylation patterns of the candidate loci, which had been PCR-amplified from bisulfite-treated DNA samples of breast tissue specimens. Aiming at identification of a signature discriminating tumors from control samples, a stringent statistical analysis was conducted. Investigation of genomic sequences spanning the signature CpGs, revealed potential involvement of these CpGs in binding behaviour of transcription factors to the regulatory regions of cognate genes, indicating biological relevance of the signature.

3.1.1 Developing microarrays for DNA methylation analysis

3.1.1.1 Control samples representing different levels of DNA methylation

One critical step in production of control samples was the amplification of CGIs from genomic DNA. The high GC content of these regions can strongly influence the denaturing and annealing behaviour of DNA which subsequently affect the PCR performance. Indeed, it was not possible to amplify some of these loci using normal PCR conditions and amplification was only achieved in modified conditions when Q-Solution was used. The best annealing temperature for PCR amplification of each locus was obtained by using a gradient of annealing temperatures in the presence or absence of Q-Solution and analyzing

the PCR products on agarose gel. As an example the agarose gel electrophoresis results for gradient PCR of *GSTP1* CGI is shown in figure 12.

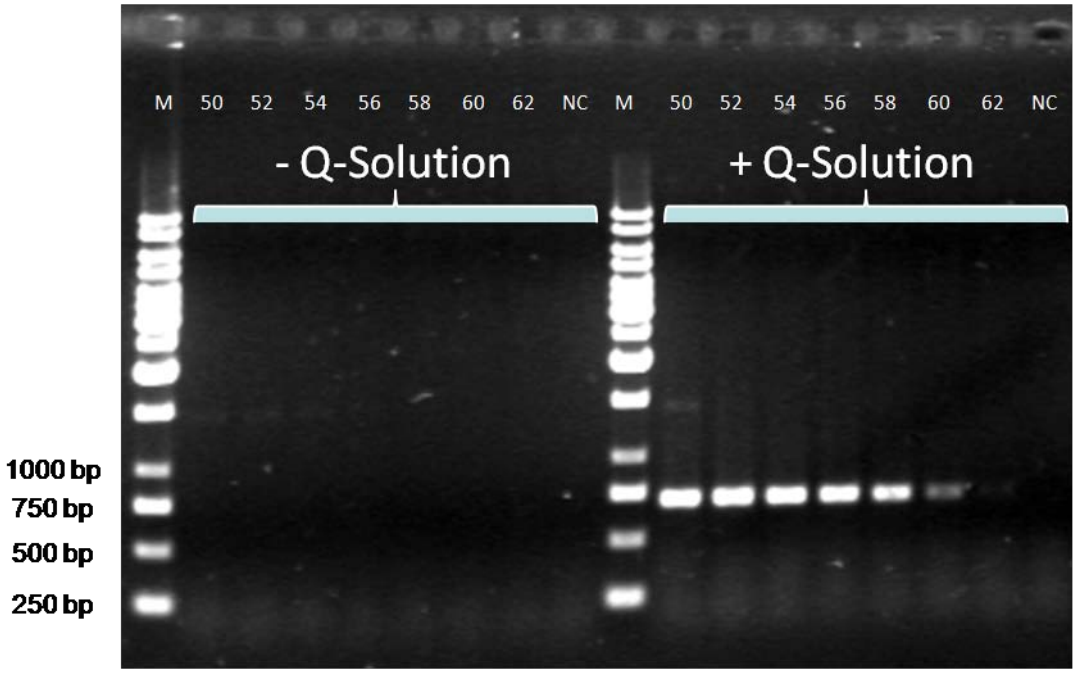


Figure 12. PCR-amplification of *GSTP1* CGI. To isolate the selected GCIs from genomic DNA, a gradient of annealing temperatures (shown in white lettering in upper part of the figure) was applied in the presence or absence of Q-Solution, in PCR amplification process. **M:** DNA Marker; **NC:** Negative Control.

As mentioned in Methods (2.2.4.2), fully-methylated and -unmethylated versions of the CGIs, representing 100% and 0% DNA methylation statuses of all CGIs, were produced. The methylation status of these control samples was confirmed by gel electrophoresis analysis of the product of their digestion with methylation-sensitive restriction endonuclease BstUI. Digestion patterns are shown in figure 13 for *PGR* and *CDH13* genes, as an example.

After validation of expected methylation levels for *in vitro* prepared control samples (CGIs isolated from human genome using *genomic* primers, each locus was PCR-amplified from corresponding fully methylated and unmethylated control samples using *SoBi* primers following treatment of the samples by bisulfite and control pools were prepared as described.

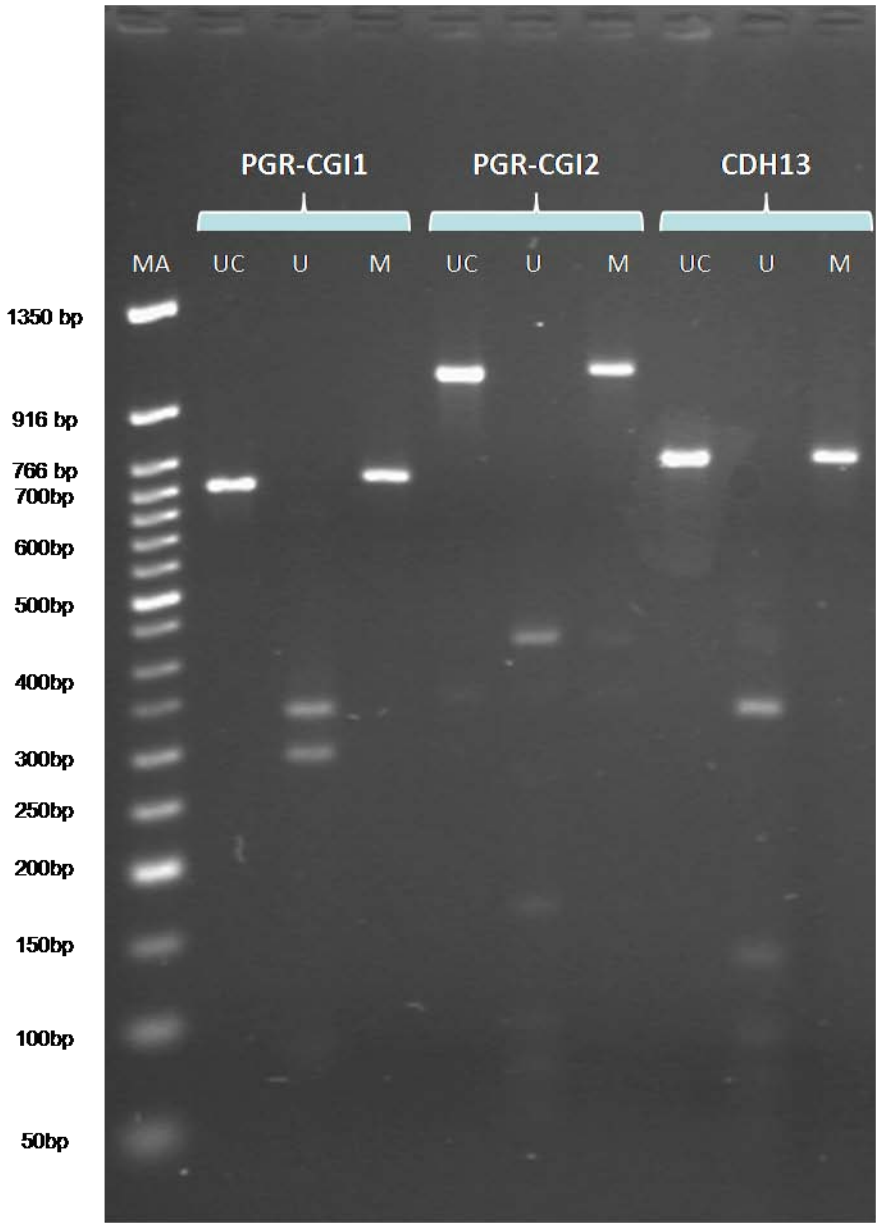


Figure 13. Examining the methylation status of control samples. Following incubation of control samples with BstUI enzyme, the digestion mixtures were analyzed by gel electrophoresis. Digestion patterns are shown for three CGI (indicated in white lettering in upper part of the figure). **MA**: DNA Marker; **UC**: Uncut (CGI without digestion use as control); **U**: Unmethylated version of the CGI; **M**: Methylated version of the CGI.

3.1.2 Optimization of probe length for reliable array performance

The specific and strong hybridization of target to probe is the basis for accurate and reliable performance of oligonucleotide microarrays. The binding between target and probe can be affected by many factors including their sequence composition. For example, high GC

content of probe and target can lead to stronger hybridization. On the other hand, GC-rich targets and probes are prone to form secondary structures that disturb hybridization process. In addition, the specificity of binding of target to the probe can be influenced substantially by probe length. As the length of probe increases the number of mismatches between the probe and non-specific targets increases, leading to a weaker hybridization (in comparison to the specific binding) that can be washed away during washing steps. Nevertheless, long length of probes with high GC content can facilitate the formation of secondary structures. Consideration of above-mentioned facts in array-based DNA methylation analysis is crucial, because vast majority of the targets and probes are GC-rich. This feature can result in production of unreliable data by some probes. Likewise, as probes and targets share similar sequence composition, the probability of occurrence of non-specific hybridization (cross-hybridization) is high. Therefore, it is necessary to find the optimal probe length to achieve the reliable data.

We evaluated the effect of probe length on behaviour and accuracy of array functioning by hybridizing the control targets to their probes, which were synthesized in different lengths. All arrays were synthesized with the same array design consisting of probes with 17, 21 and 25 nucleotides long against control targets. Three control target pools representing 0, 50 and 100% methylation levels of amplicons were used such that each control pool was hybridized to one array. The hybridization and post processing were carried out as described in Methods section. As initial examination, the reproducibility of the data created by microarrays was verified for different probe lengths. For this purpose, pairwise comparisons of the MIs produced by the same oligonucleotide pairs (M and U probes) in two independent hybridizations were performed and R-squared value (R^2) was calculated for each comparison- the higher R^2 value corresponds to more reproducibility. Results of these comparisons are presented in figure 14. The R^2 value obtained from pairwise comparisons when using 21 and 25-mer probes were similar together but higher when

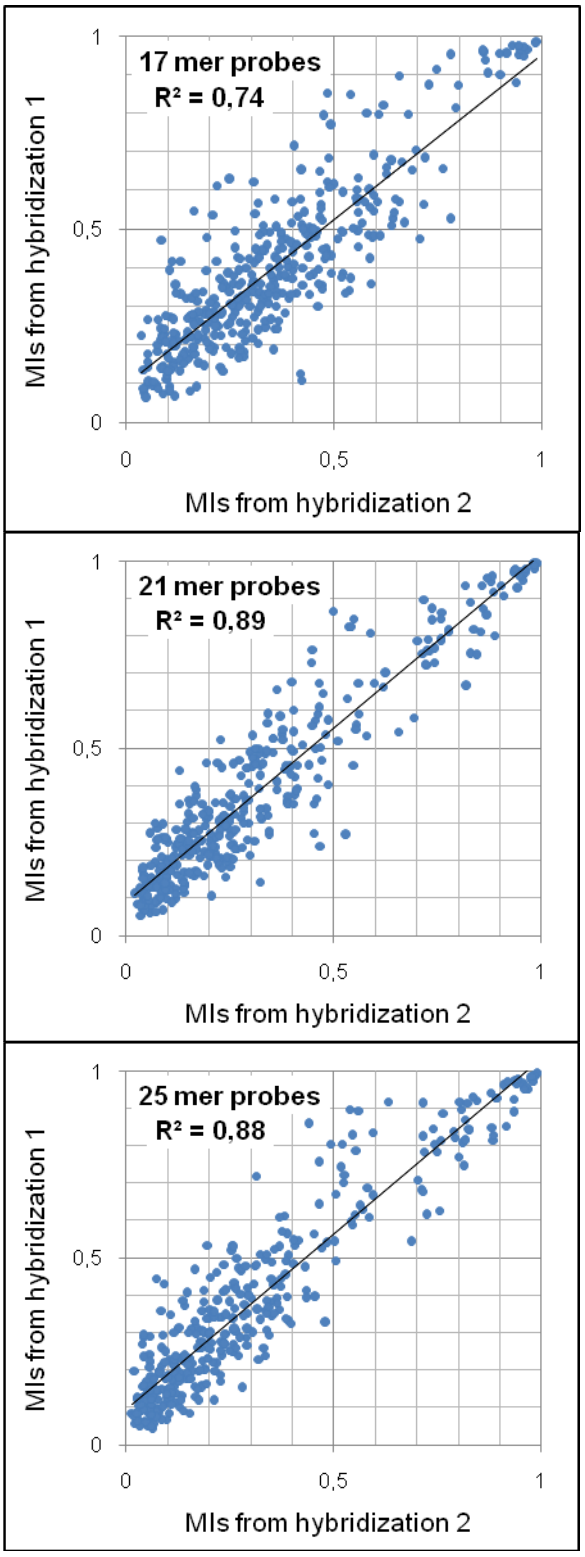


Figure 15. The effect of probe length on reproducibility of array performance. Probes with length of 21 and 25 nucleotides provide higher reproducibility as compared to 17-mer probes.

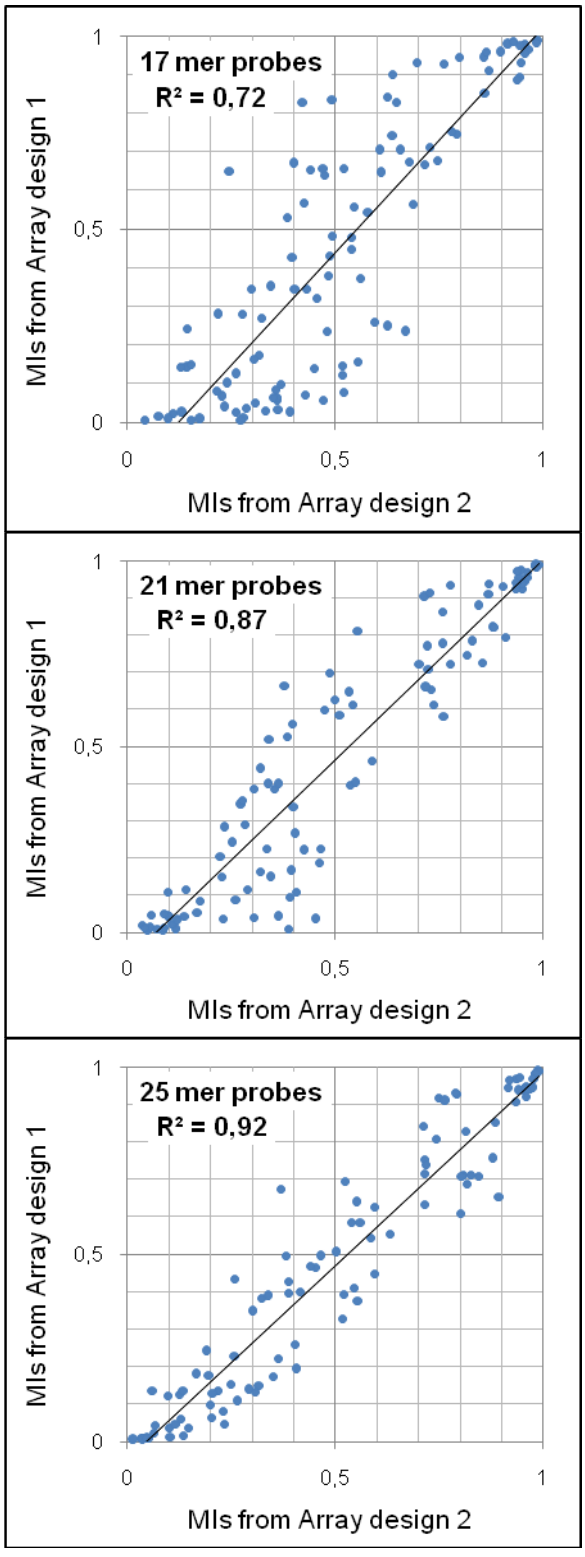


Figure 14. Effect of array composition on performance of probes negatively correlates with probe length. Data show that 25-mer probes provide the highest reproducibility of methylation data in different array contexts.

compared to the R^2 value of comparisons performed with 17-mer probes. These experiments showed that oligonucleotides with lengths of 21 and 25-mer are able to analyze DNA methylation patterns with higher reproducibility than those with length of 17-mer.

To address the effect of probe length on reproducibility of array performance in more detail, another experiment was designed. The aim of this experiment was to assess the possible effect of probe length on the production of reproducible data with same sets of probes in different array designs (array compositions). For this experiment, two different array designs (array designs 1 and 2) were prepared such that array design 1 included probes against amplicons ER-a, ER-b, HIN1-a, HIN1-b and HIN-c, and array design 2 consisted of probes detecting amplicons ER-b, TWIST-a, TWIST-b, TWIST-d. Probes directed against ER-b amplicon were included in both array designs. Three arrays of one biochip were synthesized with each array design to be hybridized with corresponding control sample pools representing 0, 50 and 100% methylation each. After hybridization, detection and data processing, the MIs shown by probe sets against ER-b amplicon (shared target) in two array designs were compared together (Fig. 15). The R^2 values calculated following pairwise comparisons increased with length of probes showing that longer probes are able to produce methylation data with higher reproducibility in different array contexts. The high R^2 value (0.92) observed for 25-mer probes indicated that the functionality of these probes is very slightly influenced by array composition and the most reproducible data can be obtained using these probes.

In another experiment, we examined the effect of probe length on cross-hybridization of probes to non-specific targets. For this purpose, two conditions were defined; in condition 1, one array was hybridized with pool of all control targets whose probes were present on the array (BRCA1-a, BRCA1-b, BRCA-c, ER-a, ER-b, HIN1-a and HIN1-b targets); in condition 2, another array with same design was hybridized with the same pool of targets

but only one target (ER-b) out of this pool had been labelled and the other ones had undergone a mock labelling. Thus, in condition 2, all signal intensities detected should be specifically referred to ER-b target. The methylation indices produced by the probes against the shared target (ER-b) were obtained from both arrays. The obtained values representing the methylation levels of the CpGs belonging to the shared target (ER-b) were compared between the two conditions for each probe length. This experiment was performed with control samples of 0, 50 and 100% DNA methylation such that both conditions were tested for each control sample. 17 and 25-mer probes were examined. The parwise comparisons of MIs produced by ER-b related probes resulted in R^2 value of 0.53 and 0.85 for 17 and 25-mer probes, respectively. These data demonstrated that the performance of 25-mer probes is less influenced by target mixture composition and they are less subjected to cross-hybridization with non-specific targets (Fig. 16).

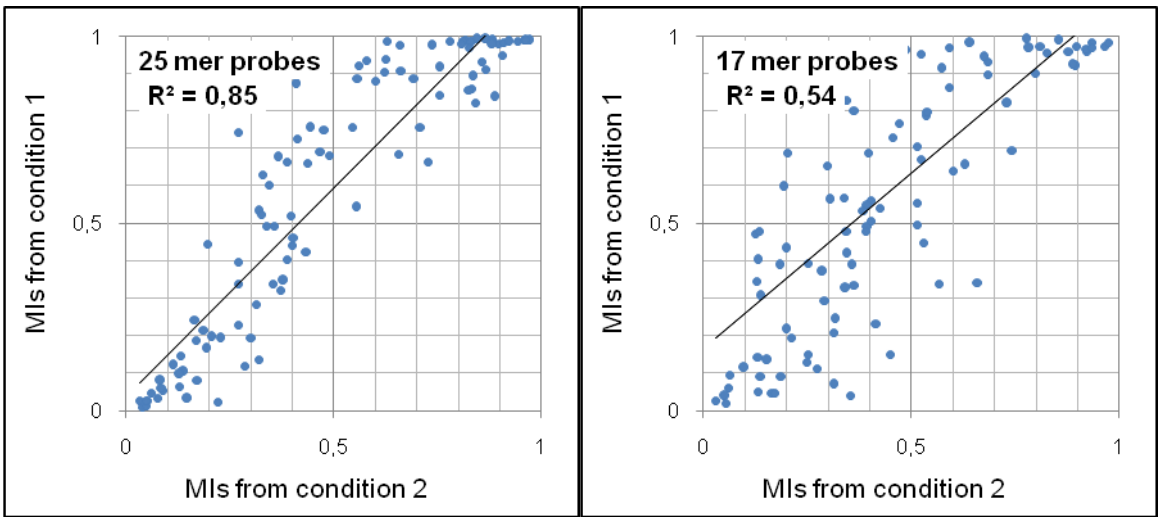


Figure 16. Effect of probe length on cross-hybridization. Two probe lengths of 17 and 25-mer were examined in conditions 1 and 2 as explained in the text. Pairwise comparison clearly shows that methylation data produced by 25-mer probes are similar in both conditions. Thus, cross-hybridization is very low in case of 25-mer probes.

It is known that the performance of oligonucleotide probes is different from one probe to another one (61). They also differ in terms of sensitivity to cross-hybridization. Therefore,

it is possible that not all of the probes work accurately and thus, a pre-selection of functioning probes using control samples before patients screen is indispensable. As the last step to find the optimal probe length, the effect of probe length on frequency of selected discriminative probes was verified. Arrays containing probes with different lengths were hybridized with fully methylated (100% methylation) and completely unmethylated (0% methylation) control samples. Following data analysis, those probes that could produce a MI < 0.2 for 0% methylated control and > 0.8 for 100% methylated control samples were selected. Then, the percentage of the selected probes (number of selected probes / number of total probes used × 100) for each gene was compared between different probe lengths.

This analysis proved that 25 mer probes possess the highest discriminative power for DNA methylation screening. In addition, it showed that using 25 mer probes, it is possible to investigate the methylation state of more CpGs of a given CGI, providing higher resolution for the analysis (Fig. 17). These experiments proved that 25-mer is the optimal length for probes to be used in our study.

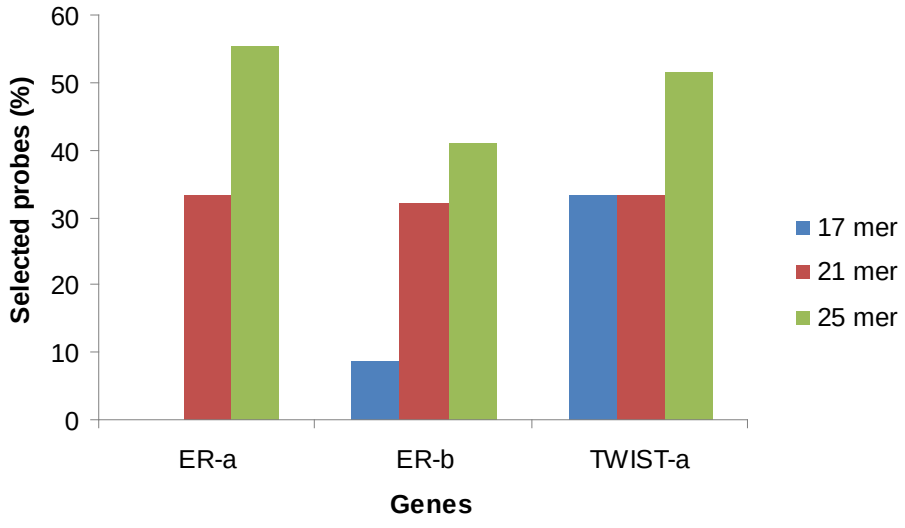


Figure 17. Evaluation of effect of probe length on frequency of selected probes after calibration. Probes targeting CGIs of three genes were examined. For all genes, the highest frequency of selected probes was achieved when using 25-mer long probes.

3.1.3 Selection of highly discriminative probes

The preceding experiments demonstrated that probes with length of 25 nucleotides can create the most reproducible and accurate data in different experimental conditions. It was also observed that the performance of oligonucleotide probes is different from one probe to another one. Although some probes showed MIs compatible with methylation levels of control samples, others could not show the correct methylation indices corresponding to the methylation level of control samples used. This observation necessitated the selection of probes that can discriminate between methylated and unmethylated DNA. To select highly discriminative probes, a calibration was carried out for each probe based on hybridization with control DNA pools designating 0, 33, 66, 100% methylation levels. Only probes that showed a MI of less than 0.2 for 0% methylated control targets, more than 0.8 for 100% methylated targets and an increasing gradient of MI from 0% to 33% to 66% to 100% methylated targets were selected. The final array design included the probes that met the selection criteria and was used for analysing the methylation patterns of patient samples. This array design included 2136 analysis probes interrogating methylation levels of 1068 CpGs in the genome.

3.2 Screening DNA methylation patterns in tissue samples

3.2.1 Isolation of DNA from tissue samples

Following the isolation of DNA from tissue samples, concentration of DNA samples was measured by Nanodrop. Quality of extracted DNA samples was then tested by gel electrophoresis (Fig. 18).

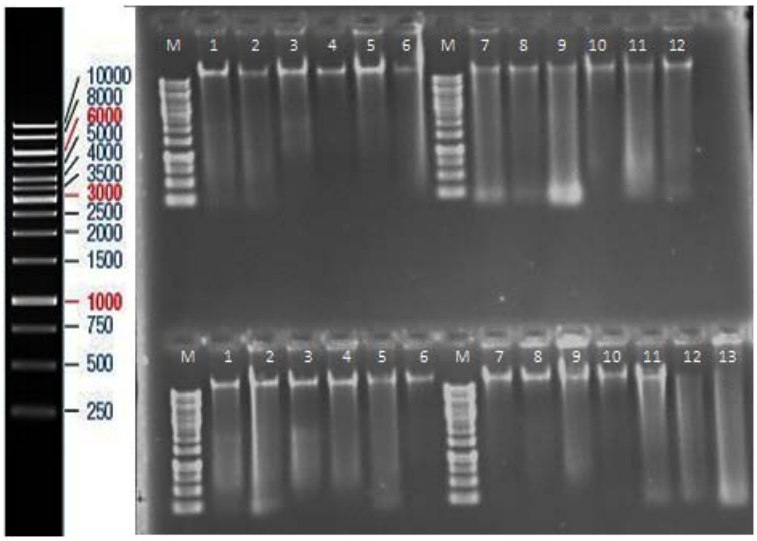


Figure 18. The DNA samples extracted from tissue specimens were loaded on 1% agarose gel to analyze the quality of the extracted DNA.

3.2.2 PCR-amplification of CGIs from DNA samples

All the 44 different CGIs were amplified form the bisulfate-treated DNA template with a single PCR program as described in Methods. Subsequently, PCR products were checked by gel electrophoresis. Gel electrophoresis data for PCR products of one sample is shown in figure 19, as an example.

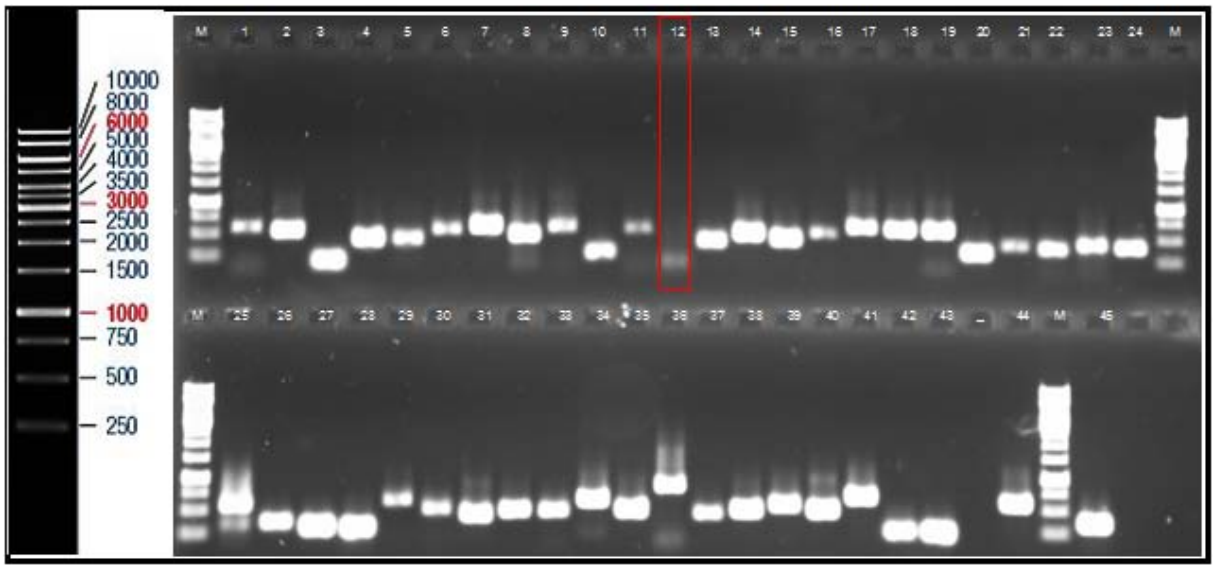


Figure 19. All fourty four amplicons from each patient sample was examined by gel electrophoresis. Electrophoresis data of patient IR-14N are presented as an example. Amplicon number 12 (marked with a red box) which refers to a CGI from the promoter region of the gene

DKK3 could not be amplified from many patient samples. Hence this CGI was excluded from microarray analysis.

3.2.3 Microarray-based analysis of methylation patterns in tissue samples

The last array design prepared following the calibration of probes related to all CGIs was used to produce microarrays for patient screening. This array consisted of 1068 pairs of probes detecting the methylated and unmethylated versions of 1068 CpGs site pertaining to the selected 44 loci. In addition, 280 probes as positive and 80 probes as negative controls were accommodated on each array to ensure the correct occurrence of synthesize and hybridization. The pool of all amplicons of each patient was hybridized to one array subsequent to labelling. The four control DNA samples representing different levels of methylation (0, 33, 66 and 100%) were also included in the analysis. After obtaining signal intensities and calculating MI for each CpG, the hybridization data related to control samples were analyzed carefully. Final filtration - same as filtration explained for probe selection- was performed to select the probes showing reliable data in the very same experiment in which patients samples had been analyzed. By this filtration, 717 probe pairs that passed the selection criteria were selected. The MIs calculated by these probes from each tissue sample were put together. This combined datasets (Supplementary Table S2) were then used for further analyses.

3.2.3.1 DNA methylation signature for diagnosis of breast cancer

To identify differentially methylated CpGs between tumors and control samples, datasets were analyzed with the M-CHiPS and the statistical significance was computed via the limma package as described in the Methods. In order to identify the CpGs whose methylation levels significantly differentiated between all tumor and control samples, different cut-off p-values were tested. As a result, by selection of CpGs that exhibit

adjusted p-values smaller than 0.01 (93 CpGs), all tumors were distinguished from control samples. Heat-map hierarchical clustering of patients' methylation data for these CpGs is shown in figure 20. As shown in this figure, based on the methylation levels of these CpGs, tumor and normal samples are clustered into two separate main clusters indicating a significant divergence of methylation levels between these groups. In addition, this analysis revealed that based on DNA methylation patterns, tumors can be clustered into four sub-groups (shown with A, B, C and D in figure 20). The selected 93 CpGs belonged to eighteen CGIs associated with sixteen genes. These genes and their cellular functions are listed in the table 11.

Although these 93 CpGs were identified as significantly differential methylated sites between tumor and control samples, a heterogeneous pattern of methylation for many of the identified CpGs could be observed among the tumors. Since the aim of this screen was to identify potential DNA methylation markers for diagnosis of breast cancer, we sought to find the most differentiating CpGs sites among these 93 CpGs to reduce this heterogeneity. This also helps to minimize the number of CpG sites necessary for discriminating tumors from normal samples. For this purpose, we applied a filtering on the pre-selected CpGs based on the reducing the p-value cut-off to select the CpGs exhibiting the lowest p-values. To monitor the effect of filtering on sample clustering, correspondence analysis (CA) was used.

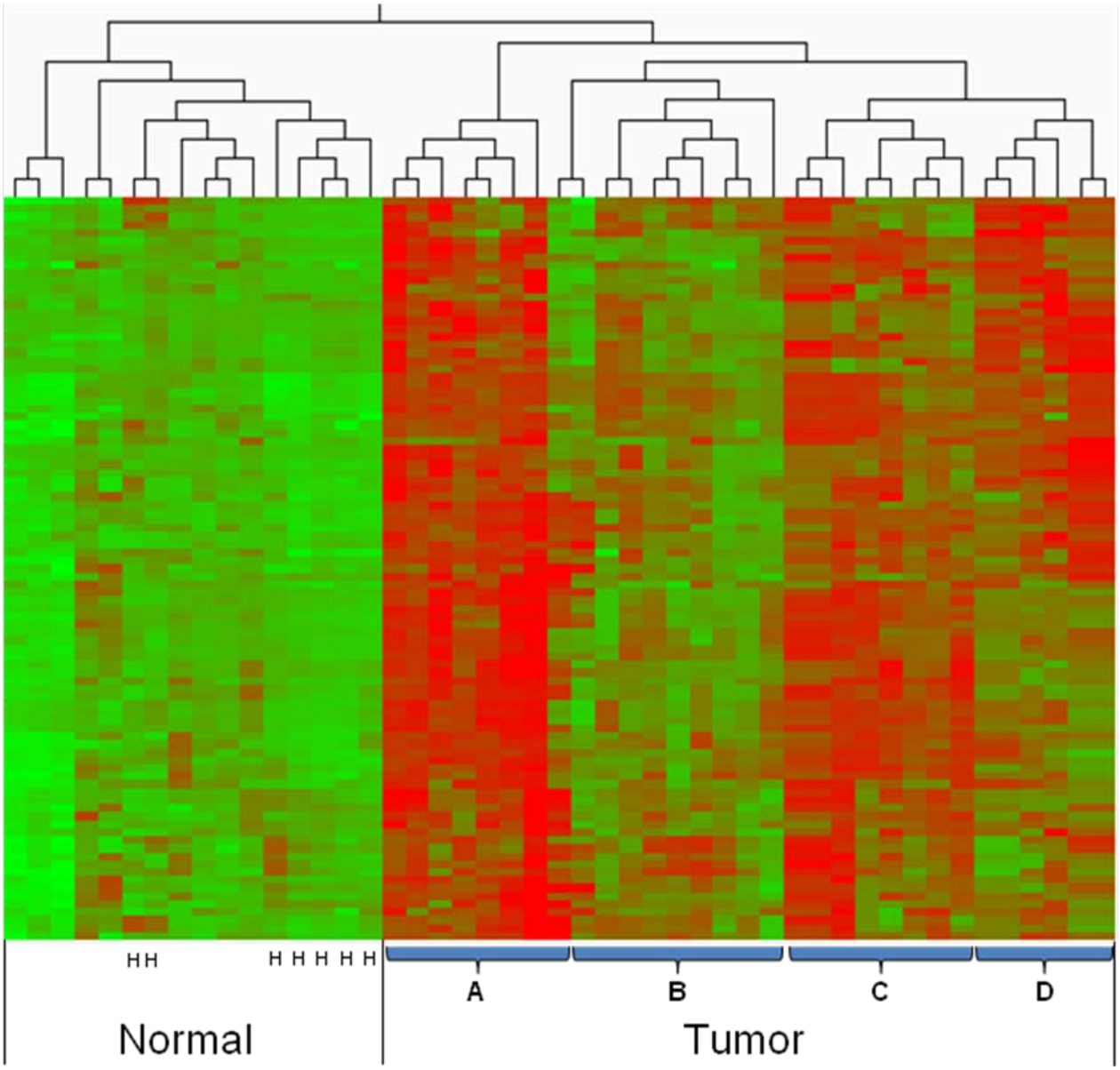


Figure 20. Clustering of samples by DNA methylation levels of the identified 93 CpGs. Green and red colors represent low and high methylation levels, respectively. Tumors and normal samples are clustered into two main groups. Healthy samples are highlighted with "H" among the Normal population. Four sub-groups among tumors identified by DNA methylation patterns are shown with **A**, **B**, **C** and **D**.

Table 11. Genes whose CGIs accommodate the 93 identified CpGs.

Gene	Function (cellular signaling involved)
SFRP1	Modulator of Wnt signaling
SFRP2 (2 CGIs)	Modulator of Wnt signaling
SFRP3	Modulator of Wnt signaling
SFRP4	Modulator of Wnt signaling
SFRP5	Modulator of Wnt signaling
HDPR-1 (DACT1)	Modulator of Wnt signaling
WIF-1	Modulator of Wnt signaling
TCF4	Modulator of Wnt signaling
CST6	Inhibitor of cysteine proteases
C/EBPα	Transcription factor
HIC1	Transcription factor
PAK3	Protein kinase
PTEN	Negative regulator of AKT/PKB signaling
RASSF1A	Cell cycle arrest
GSTP1	Detoxification
GHSR (2 CGIs)	Growth hormone release

Correspondence analysis is a computational method to investigate associations between variables, such as genes and patient samples in a low-dimensional space. Similar to principle component analysis, it displays the data in 2 or 3D projection plot. The data analysis and projection are performed simultaneously for two (or more) variables, thus

revealing associations between them. One major advantage of this process is its ability to represent the associations between samples and genes at the same time in a single biplot (Fig. 21). Not only the genes or CpGs (black dots) but also divergence of the individual samples (colored squares) is presented. Co-localization of genes or CpGs is indicative of a strong association between them; the closer they are the stronger the association is. Likewise, the genes or CpGs that are closer to one cluster, have higher level of expression or methylation in members of that cluster than members of farther clusters. Following this analysis, a DNA methylation signature composed of methylation patterns of fourteen CpG sites was identified, which differentiated between the tumors and the normal breast tissue samples. CA plot depicting the cluster analysis is shown in figure 21. We also added the data of benign samples to this plot for comparison purposes. Clustering of samples based on methylation levels of these CpGs by CA suggested that tumors (green squares) and control samples from healthy people (blue squares) exhibited markedly different methylation profiles and form distinct clusters (Fig. 21). Furthermore, this analysis showed that the principle difference (distance between clusters along the horizental axis) in methylation levels between the two groups exhibits correlation with the malignancy (cancer) phenotype in the analyzed samples. It is also clear that control samples from cancer patients (red squares) are clustered together and are separated from tumors. Interestingly, these samples are located between tumor samples and normal samples from healthy individuals along the horizontal axis on the plot, indicating that the same factor discriminating tumors from healthy samples, distinguishes patient-matched control samples from tumors as well as from healthy samples. It is worth noting that the cluster containing benign samples (pink squares) has a similar position as that of patient-matched control samples along the horizontal axis.

Of significance, all CpGs included in this signature are close to the tumors and far from healthy samples. This suggests that signature CpGs have higher level of methylation in

tumors as compared to control samples. From this plot, it can be predicted that healthy samples have the least amount of DNA methylation among all sample types and DNA methylation level in patient-matched normal and benign samples is higher than those of healthy samples but less than those of tumors.

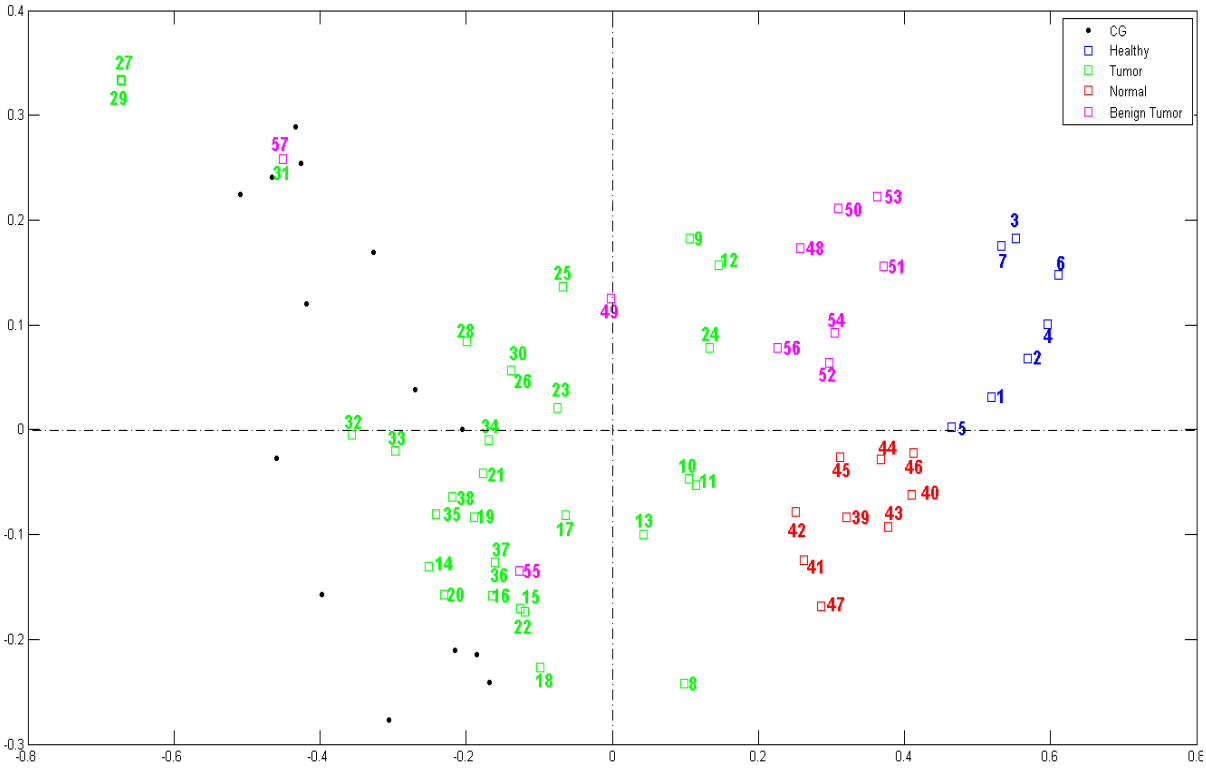


Figure 21. Clustering of samples by DNA methylation signature identified in this study. This figure is the projection of the processed microarray data in Correspondence Analysis plot. The black dots in the figure represent the CpG dinucleotides whose methylation status can differentiate between tumor, paired normal and healthy samples. The green, pink, red and blue squares correspond to tumor, benign, paired normal breast tissue from cancer patients and normal breast tissue from healthy individuals samples, respectively.

3.2.3.2 Signature CpGs are hypermethylated in tumors

The correspondence analysis identified fourteen CpGs significantly differentiating tumors from control sample. These CpGs belong to the CGIs located at either the promoter or the first exon of *SFRP2* and *GHSR* genes, respectively. Data from CA analysis suggested that signature CpGs have higher level of methylation in tumor samples in comparison to other samples. Indeed, by querying the methylation levels of individuals CpGs of this signature, we found that the signature CpGs are hypermethylated in tumors when compared to normal

samples from patients and healthy individuals (Fig. 22). In addition, we observed that majority of these CpG dinucletides showed a relatively higher level of methylation in tumor-matched normal tissue specimens as compared to normal tissue samples from healthy individuals (Fig. 22). The methylation levels of these CpGs sites were also compared between normal, tumor and benign samples. The methylation loads of these sites in benign specimens were lower as compared to those of tumors but higher when compared to their loads in samples from health persons (Fig. 23).

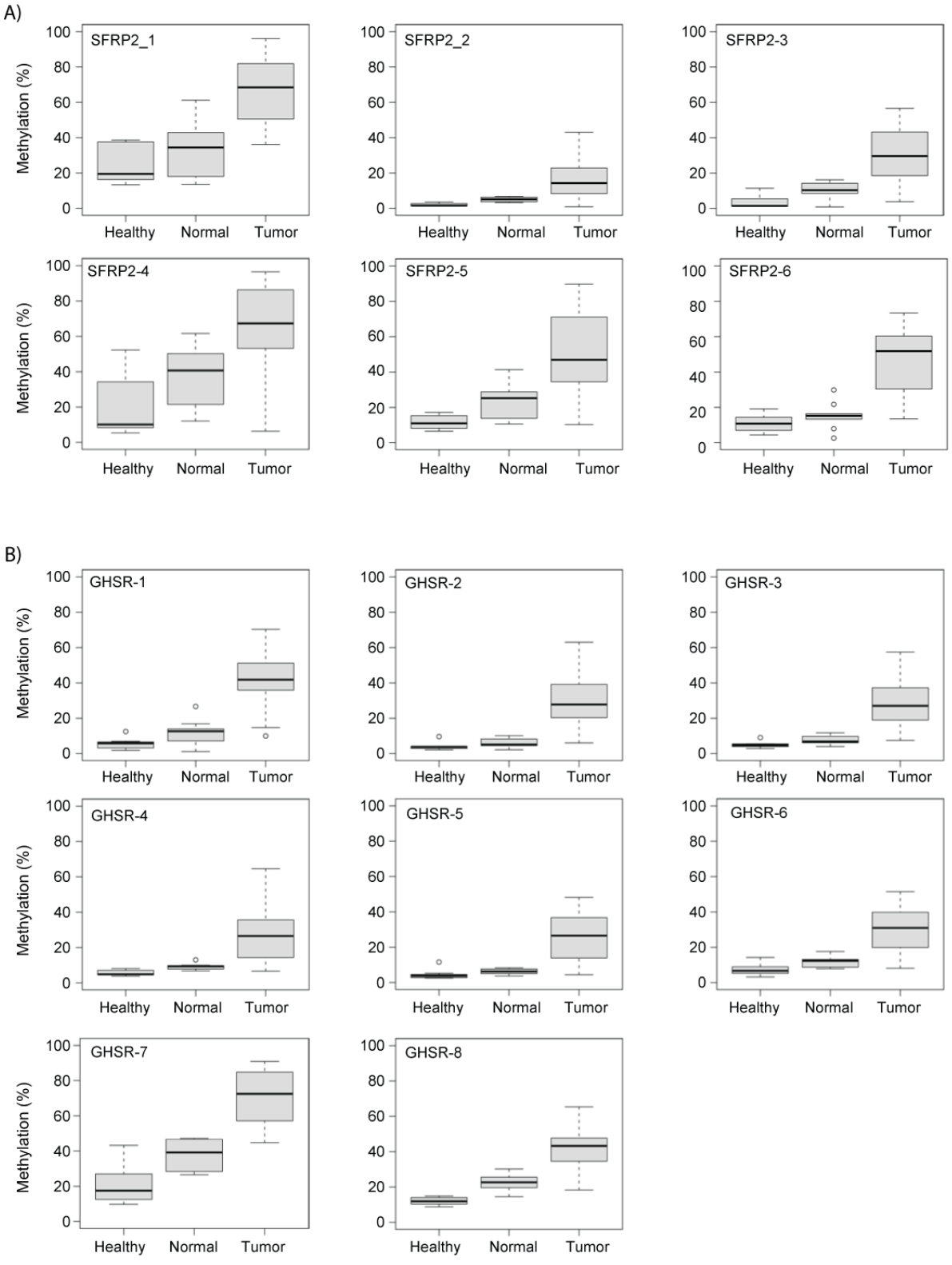


Figure 22. Methylation levels of signature CpGs in different sample types. Levels of methylation on the Y-axis are plotted against the sample types on the X-axis. **(A)** Six signature CpGs are located at the promoter region of *SFRP2* gene. **(B)** Eight signature CpGs belong to a CGI located at the first exon of *GHSR* gene.

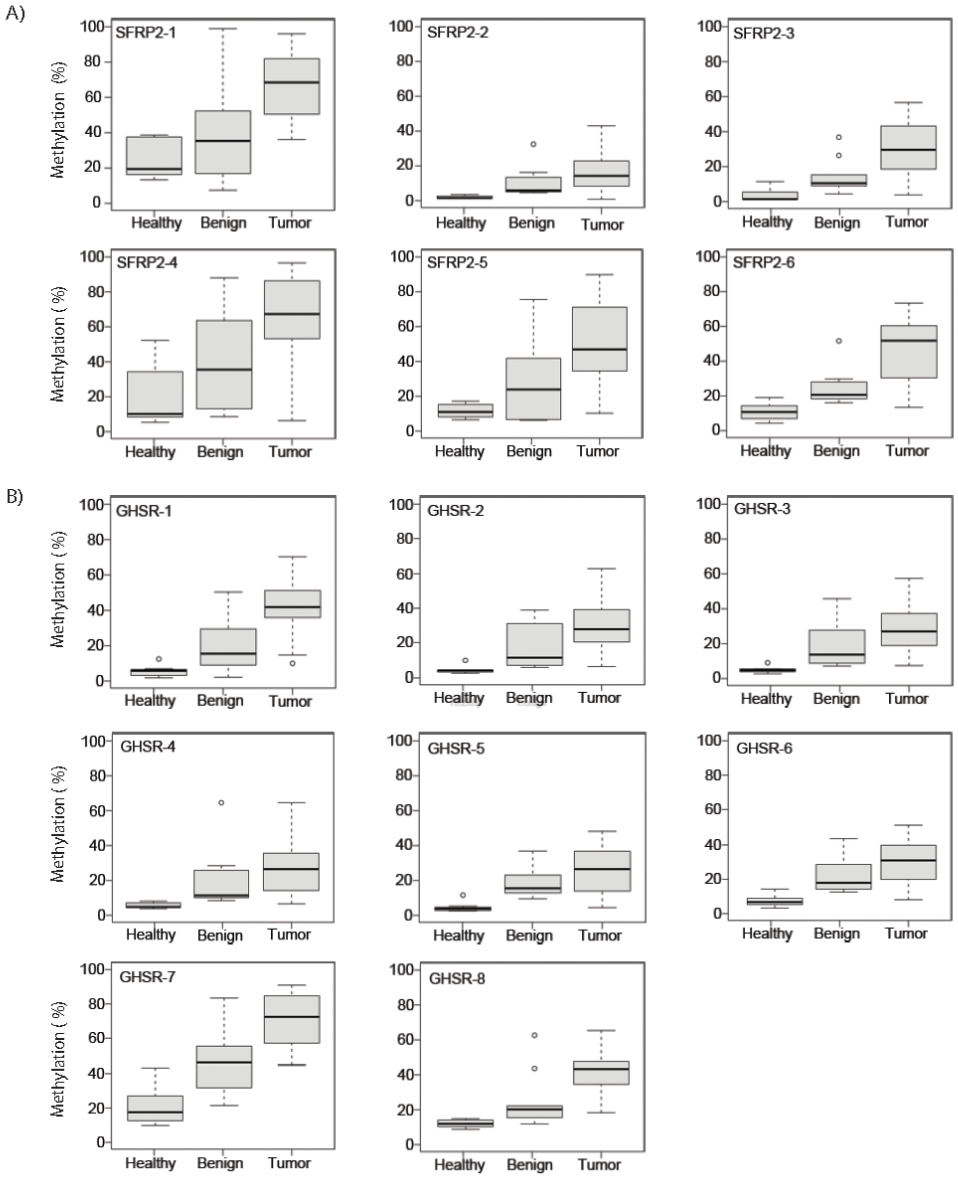


Figure 23. Comparison of DNA methylation of signature CpGs between healthy, benign and tumor samples. DNA methylation levels of benign samples are between those of healthy and tumor samples for the signature CpG dinucleotides.

3.2.4 Bisulfite sequencing data

The signature CpGs were located at CGIs related to *GHSR* and *SFRP2* genes. To obtain a detailed view of DNA methylation patterns of these CGIs, we subjected them to bisulfite sequencing. We compared the DNA methylation maps of these loci between tumors, matched-normal and normal samples from healthy individuals. For this comparison, eight

tumors, their eight matched normal samples and five normal tissue samples from healthy individuals were used. Overall, *GHSR* CGI showed 44.5% methylation in tumors, 11.5% in patient-matched normal and 10% in healthy breast tissue samples (Fig. 24), whereas *SFRP2* CGI methylation levels were 44%, 4.3% and 5% in these samples, respectively (Fig. 25). This observation supports our array results on hypermethylation of these genes in tumors. Additionally, overall methylation levels of these loci were similar in normal samples from patients and cancer-free individuals. Bisulfite sequencing data were then analyzed at single CpG resolution. This analysis revealed several CpGs (in both CGIs) that represent a pattern of increasing gradient in methylation load from healthy people to the normal and then to the tumor samples of cancer patients (represented by red boxes in figures 24 and 25). Interestingly, most of these CpG dinucleotides exist at the regions where signature CpGs were located (signature CpGs are highlighted by blue arrows in figures 24 and 25). These data further support the ability of the signature CpGs to classify the tissue samples into distinct clusters.

The detailed analysis of bisulfate sequencing data showed a selective pattern of hypermethylation for certain CpGs along *GHSR* and *SFRP2* CGIs. To verify whether these CpGs might play an important role in regulation of the cognate genes, genomic sequences of these CGIs were analyzed. *In silico* analysis of the loci using MatInspector software (144) uncovered the consensus binding sites for several transcription factors in these CGIs (Figs. 24 and 25). The binding sites for CCCTC-binding factor (CTCF) protein and PAX family members were found in both CGIs. Strikingly, we found that majority of the CpGs exhibiting the gradient pattern of methylation are located in the annotated binding sites of CTCF. Moreover, a few of them settle in the binding sites of PAX family members (Figs. 24 and 25).

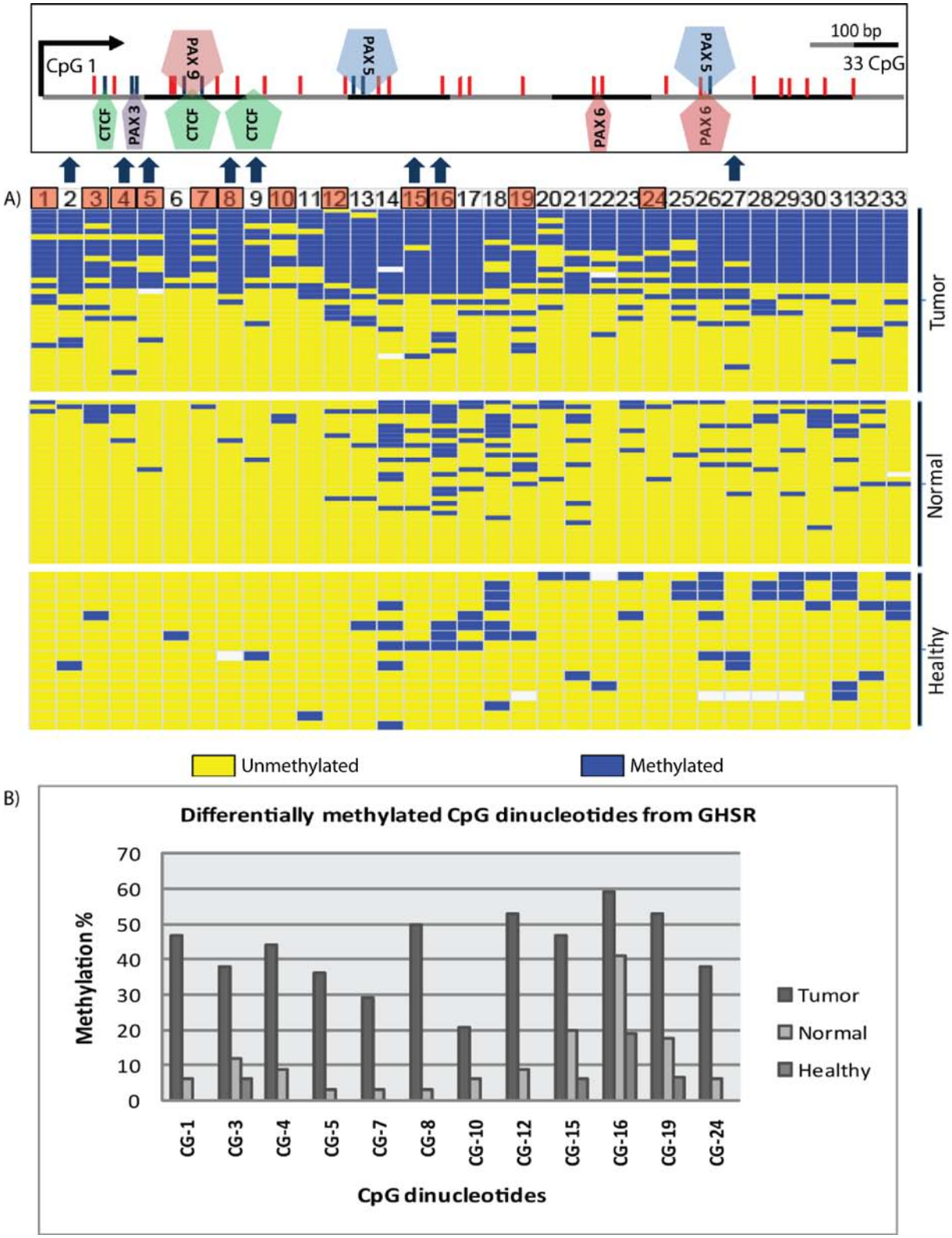


Figure 24. (A) Methylation patterns of the CGI located at the first exon of *GHSR* in tumors and control samples from patients and healthy individuals. Each row is a clone and each column represents one of the 33 CpG dinucleotides in the CGI. The CpGs which are marked in red boxes show differential methylation between the tumors, paired normal and the healthy breast tissue samples, whereas those marked with a blue arrow are constituents of the methylation signature. The panel above is a miniature of the sequence with all the 33 CpGs highlighted with red and blue colored lines. Those belonging to the signature are shown in blue. The transcription factors whose

binding sites are found in the CGI are depicted. **(B)** Methylation levels of differentiating CpGs (in red boxes) from bisulfite sequencing data in different samples.

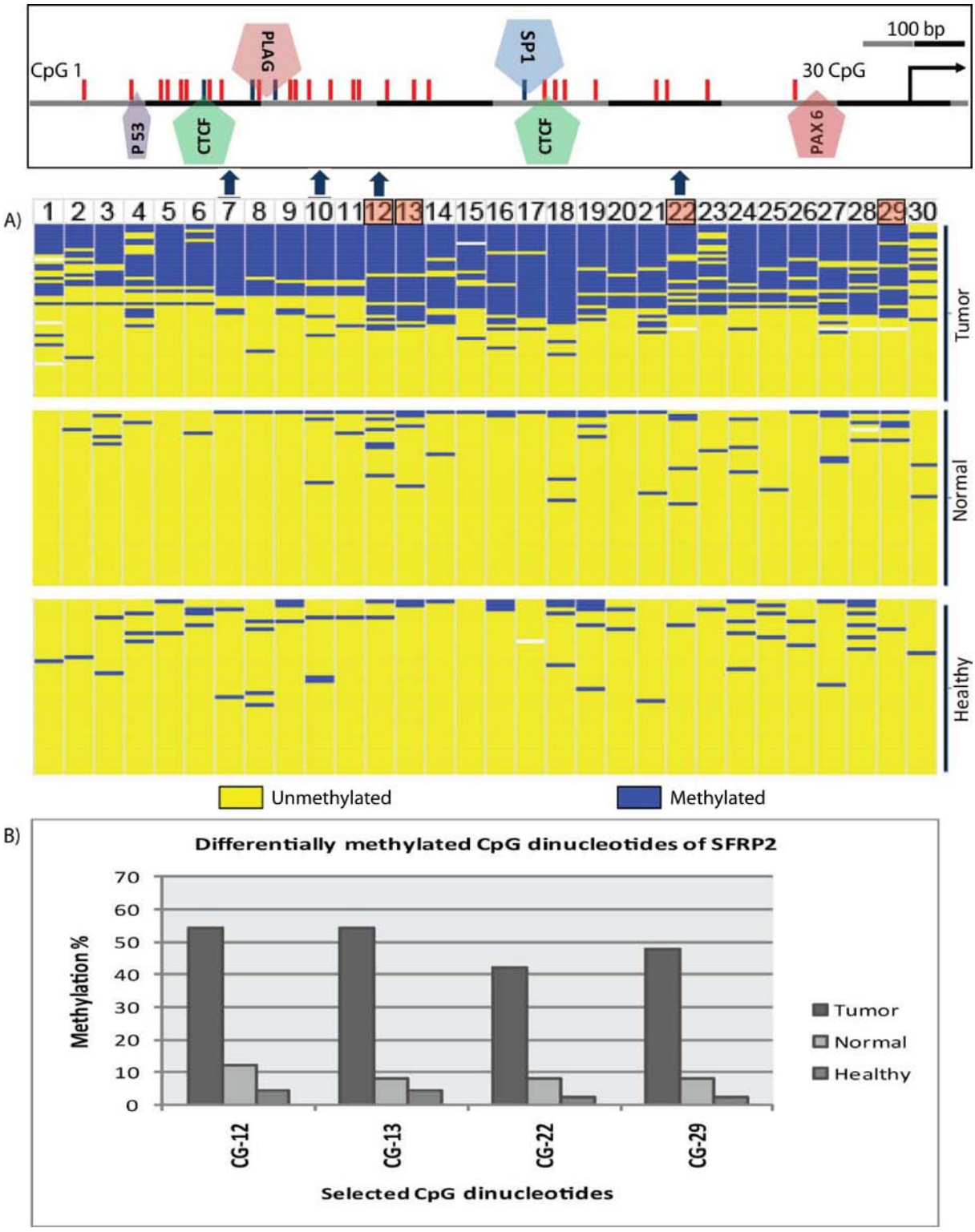


Figure 25. (A) Methylation patterns of the CGI located upstream of the TSS of *SFRP2* in tumors and control samples from patients and healthy individuals. Each row is a clone and each column represents one of the 30 CpG dinucleotides in the CGI. The CpGs which are marked in red boxes show differential methylation between the tumors, paired normal and the healthy breast tissue samples. Those CpG sites marked with a blue arrow are constituents of the signature CpGs. The

panel above is a miniature of the sequence with all the 30 CpGs highlighted with red and blue colored lines. Those belonging to the signature are highlighted in blue. The transcription factors whose binding sites are found in the CGI are depicted. **(B)** Methylation levels of differentiating CpGs (in red boxes) from bisulfite sequencing data in different samples.

3.2.5 Re-activation of methylated genes following decitabine treatment

To verify whether inhibitaion of DNA methylation can affect the expression of signature genes in breast cancer cells, cells were treated or not with decitabine and the mRNA levels of the candidate genes were compared between treated and untreated cells. This analysis was performed for other Wnt pathway-related gene listed in table 11. Figure 26 shows the mRNA levels of these genes in decitabine treated cells relative to their levels in untreated cells.

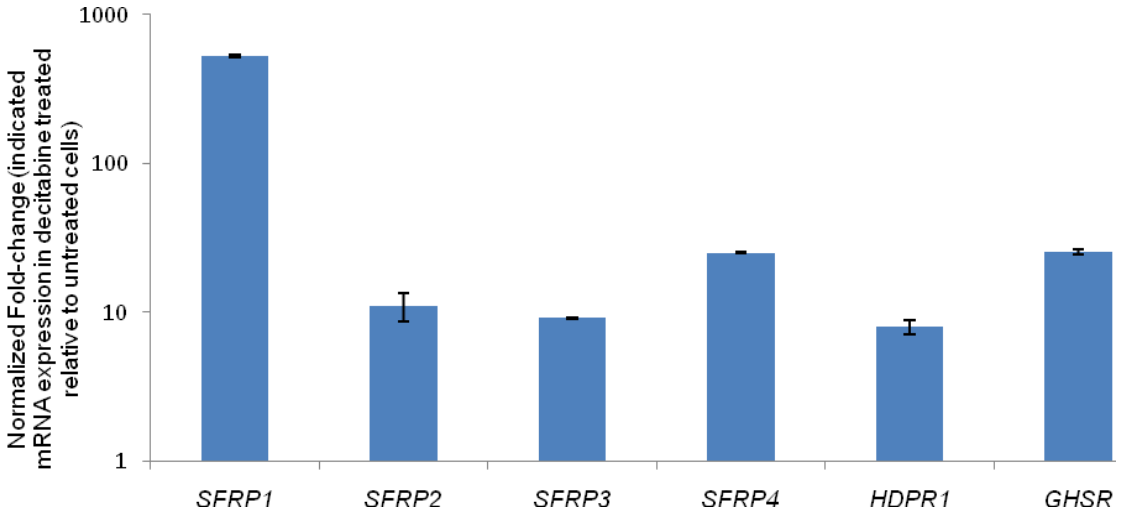


Figure 26. Re-activation of methylated genes in MDA-MB-231 breast cancer cells following decitabine treatment. qRT-PCR data for the indicated genes. mRNA levels of genes were normalized to the level of *GAPDH* mRNA in treated and untreated cells. The values are represented as ratio of the normalized expression of each mRNA in treated cells versus untreated cells.

3.3 Identification of miRNAs involved in ERα activation

3.3.1 miR-375 expression is increased in ERα-positive breast cancer cell lines

In an initial attempt to identify miRNAs that might be involved in the regulation of ERα pathway and in the development of breast cancer, we performed miRNA profiling of 8 human mammary cell lines, which included non-tumorigenic immortalized cells (MCF-10A and MCF-12A), ERα-positive (MCF7 and T47D) and ERα-negative (SK-BR3, MDA-MB-231 and MDA-MB-435) cancer cell lines using microarray technology. Microarray raw data was processed, controlled for quality and subjected to normalization as described previously in methods. Normalized data are available in supplementary table S3. miRNA expression profiles of ERα-positive cells were compared to ERα-negative cancer cell lines as well as to non-cancer cells and differentially expressed miRNAs were shortlisted (Table 12). Aiming at identification of miRNAs which might positively regulate ERα activity, we sought for up-regulated miRNAs in ERα-positive cell lines. Interestingly, we identified miR-375 as the second most significantly up-regulated miRNA in ERα-positive cells when compared to both ERα-negative and non-tumorigenic cell lines. Quantitative real-time PCR (qRT-PCR) confirmed that the expression level of mature miR-375 was higher in MCF7 and T47D (ERα-positive) cancer cell lines when compared to ERα-negative cancer and the non-tumorigenic breast cell lines (Fig. 27). These results confirmed our miRNA profiling data and indicated that miR-375 is overexpressed specifically in ERα- positive breast cancer cell lines.

Table 12. Differentially expressed miRNAs in cell lines

ERα+ compared to ERα- Cancer cells		ERα+ compared to non-cancer cells	
miRNA	Fold-change	miRNA	Fold-change
5% most up-regulated		5% most up-regulated	
hsa-miR-203	7,4	hsa-miR-200a	7,7
hsa-miR-375	5,4	hsa-miR-375	6,9
hsa-miR-205	4,6	hsa-miR-200b	5,4
hsa-miR-148a	4,3	hsa-miR-203	4,9
hsa-miR-615-3p	4,0	hsa-miR-200b*	4,3
hsa-miR-196a	3,9	hsa-miR-196a	4,2
hsa-miR-200c	2,9	hsa-miR-615-3p	3,5
hsa-miR-421	2,8	hsa-miR-429	3,5
5% most down-regulated		5% most down-regulated	
hsa-miR-146b-5p	-3,0	hsa-miR-34c-5p	-3,7
hsa-miR-29a	-3,3	hsa-miR-29a	-3,9
hsa-miR-31*	-4,0	hsa-miR-146b-5p	-4,5
hsa-miR-146a	-4,9	hsa-miR-224	-5,3
hsa-miR-155	-6,4	hsa-miR-31*	-6,5
hsa-miR-31	-6,7	hsa-miR-221	-8,6
hsa-miR-221	-8,2	hsa-miR-31	-9,3
hsa-miR-222	-9,7	hsa-miR-222	-10,3

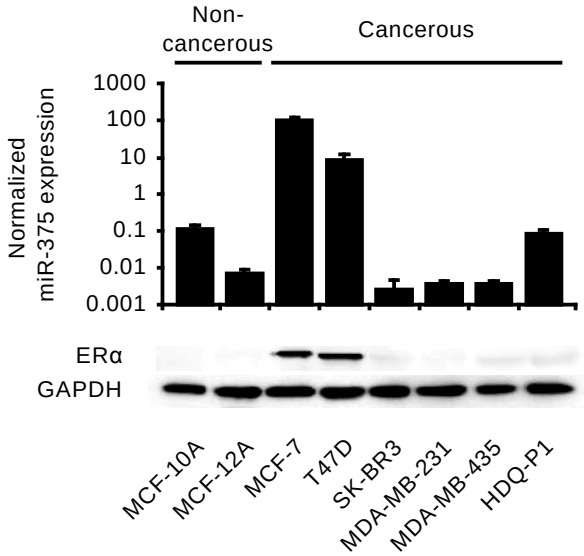


Figure 27. Correlation between miR-375 expression and ERα in mammary cell lines. miR-375 levels were measured by qRT-PCR and ERα presence was verified by western blot.

3.3.2 Reciprocal regulation between miR-375 and ERα

Given the positive correlation between miR-375 and ERα expression in cell lines, we sought to examine the possible regulatory connection between these molecules. In order to assess a possible role of miR-375 in ERα signaling, ERα protein as well as transcriptional activity levels were measured following ectopic expression of miR-375 in MELN cells. MELN are MCF7 cells which stably express an estrogen responsive element (ERE)-controlled firefly luciferase reporter gene (138). Overexpression of miR-375 in MELN cells resulted in a 2-fold induction of luciferase activity as compared to the control, following stimulation of cells with estradiol (E2) (Fig. 28 left). In agreement with this data, an almost 2-fold increase of ERα protein levels was observed following transfection of cells with pre-

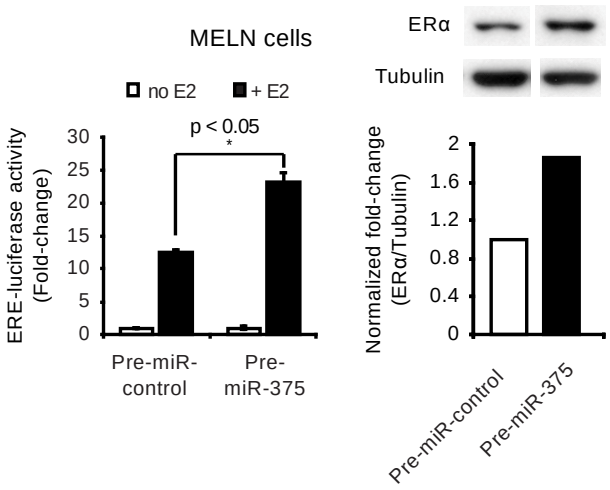


Figure 28. miR-375 positively regulates ERα activity and expression in MELN cells. The effect of miR-375 expression on ERα activity and expression was measured after modulation of miR-375 expression in MELN cells. Before measuring ERE-driven luciferase activity cells were treated or not with 10 nM E2 for 16h (left). ERα expression assayed by immunoblotting in cells with no E2 treatment (right).

miR-375 (Fig. 28 right). This effect was reproducible in MCF7 cells transiently transfected with an ERE-controlled *Firefly* luciferase vector (Fig. 29A). In addition, blockade of miR-375 resulted in decreased ERα levels in MCF7 cells (Fig. 29B). These findings provide evidence that miR-375 positively regulates ERα both at the protein and transcriptional activity levels.

According to the correlation observed between miR-375 and ERα levels in different cell lines, it is conceivable that ERα presence is a determinant for miR-375 overexpression. To

investigate a potential role for ERα in the regulation of miR-375, miR-375 levels were quantified in MCF7 cells 48 hours after transfection with siRNAs against ERα. Quantitative real-time PCR analysis showed that miR-375 levels decreased by 50% (as compared to negative controls) after transfection with ERα specific siRNAs (Fig. 29C). These data indicate that miR-375 expression in MCF7 cells is partially dependent on ERα.

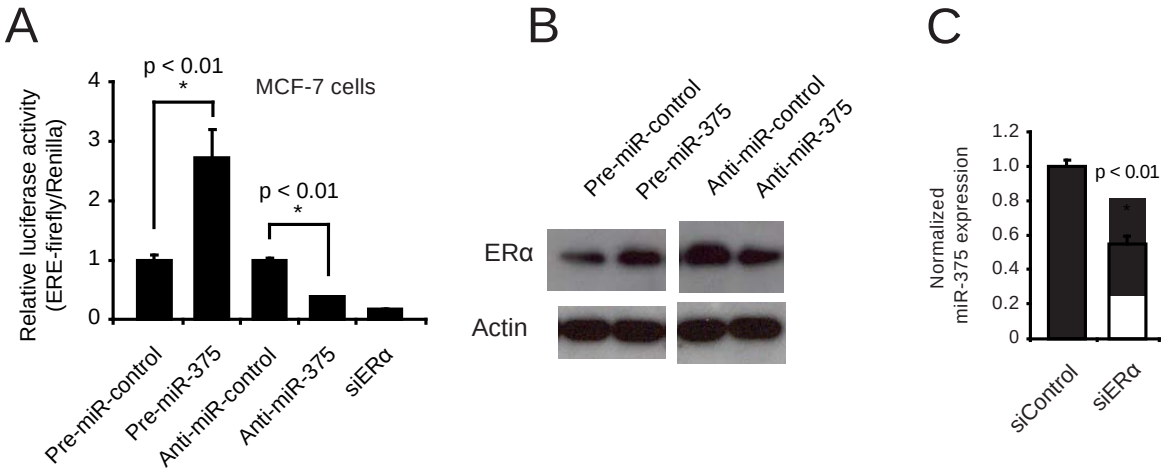


Figure 29. Reciprocal regulation between miR-375 and ERα in MCF7 cells. (A) MCF7 cells were co-transfected with an ERE-controlled firefly luciferase reporter vector and with pre-, anti-miR-375 or the control RNAs. A co-transfected renilla luciferase vector was used for transfection normalization. siERα was used as positive control. **(B)** The effect of transfection of MCF7 cells with pre- and anti-miR-375 on ERα expression was measured by western blot. **(C)** miR-375 expression in MCF7 cells was measured by qRT-PCR 48h after transfection with siERα or control siRNAs.

3.3.3 miR-375 inhibition reduces the proliferation rate of breast cancer cells

Since ERα signaling is an important factor in the growth and proliferation of ERα-positive cancer cells, the potential contribution of miR-375 to the proliferation of these cells was evaluated through blocking miR-375 activity with anti-miR-375 in MCF7 cells. Compared to the control experiments (transfection with anti-miR-control), cell proliferation decreased in miR-375 inhibited cells to almost 50%, 72 hours after transfection (Fig. 30A-B). However, inhibition of miR-375 did not result in an induction of caspase activation (Fig. 30C), suggesting that the antiproliferative activity of miR-375 inhibition is not due to apoptosis induction. These observations, together with the positive regulatory effect of

miR-375 on ERα activity, indicated that the positive effect of miR-375 on cell proliferation could be mediated by the positive modulation of ERα activity in cancer cells.

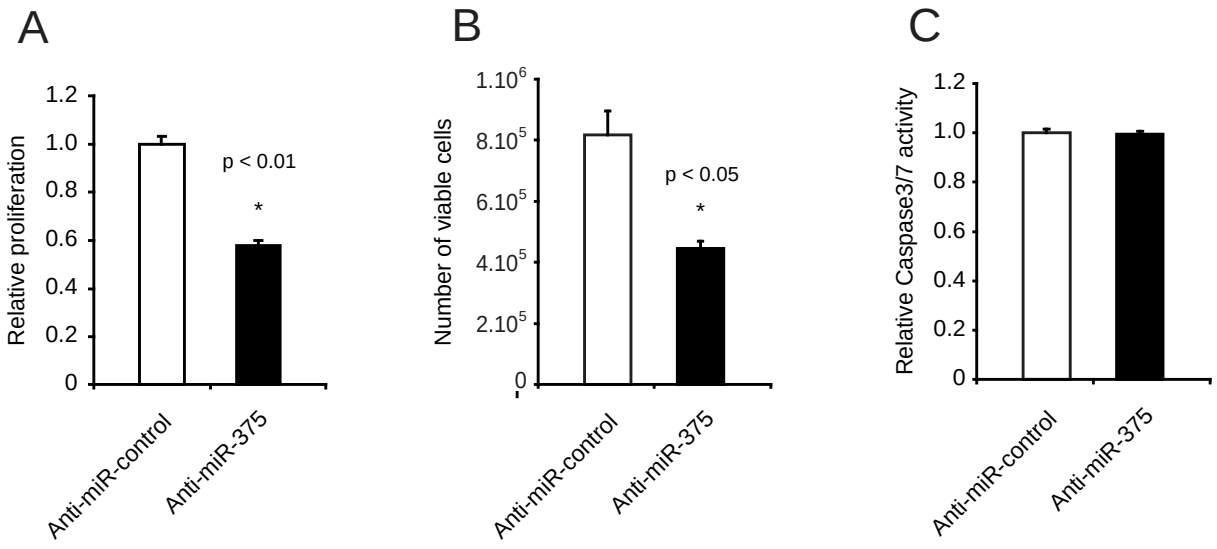


Figure 30. The effect of anti-miR-375 transfection on proliferation of MCF7 cells. Inhibition of miR-375 decreases proliferation of MCF7 cells. Proliferation of MCF7 cells transfected with anti-miR-375 or anti-miR-control was measured 72h after transfection using **(A)** the Celltiter Glo assay or **(B)** by cell counting. **(C)** miR-375 inhibition does not induce apoptosis. Induction of apoptosis was assayed by measuring Caspase 3/7 activity in MCF7 cells after miR-375 inhibition using Caspase 3/7 Glo assay.

3.3.4 Epigenetic marks determine the transcriptional state of the miR-375 locus

To further explore the regulatory mechanisms governing miR-375 expression, the genomic region upstream and spanning the miR-375 gene was analyzed *in silico*. The coding region of the miR-375 gene is located in a large CGI (CGI 2) of around 850 bp which contains at its most distal part a region homologous to the miR-375 promoter identified in mouse (145) (Fig. 31A). Another CGI (CGI 1) resides 550 bp upstream this region. As several miRNAs have been reported to be regulated by DNA methylation and histone modifications (78), we analyzed the epigenetic modification pattern of the miR-375 locus in MCF7, T47D (cells with high miR-375 expression) and MCF-10A, MCF-12A and MDA-MB-231 (cells with low miR-375 expression) cell lines by bisulfate sequencing and chromatin immunoprecipitation (ChIP). Bisulfite sequencing results showed that CGI 1 is methylated

in the cell lines with high expression of miR-375, whereas MCF-10A, MCF-12A and MDA-MB-231 cells showed specific hypomethylation in the distal part of this region (Fig. 31B). In contrast, CGI 2 was mostly unmethylated in MCF7, T47D and MDA-MB-231 cells, while MCF-10A and MCF-12A showed strong DNA methylation in the proximal part of the region (Fig. 31A).

To characterize the chromatin state of the miR-375 locus, we employed four different antibodies recognizing distinct covalent histone modifications in a ChIP experiment. ChIP analysis revealed a peak of histone H3 dimethylated at lysine 4 (H3K4me2), a marker of active transcription, in the CGI 1 in all cell lines analyzed (Fig. 31C). Repressive histone H3 lysine 9 dimethylation (H3K9me2) was found throughout the locus in the three cell lines with low miR-375 expression, whereas H3K9me2 levels were found to be low in both CGIs in MCF7 and T47D. H3 and H4 acetylation, a marker of active transcription, was generally low and was only slightly enriched in CGI 1 in T47D cells (Fig. 31A).

Together, these findings led us to conclude that an active miR-375 epiallele is characterized by a fully methylated CGI 1, an unmethylated CGI 2 spanning the gene body, H3K4me2 enrichment in the CGI 1 and low overall H3K9me2 levels. The repressed epiallele is characterized by local hypomethylation around CpG 18 of the CGI 1 (see box in Fig. 31B), a methylated gene body (with the exception of the MDA-MB-231 cell line), H3K4me2 enrichment in the CGI 1 and overall high levels of H3K9me2. These data suggest that H3K9 methylation is a major repressive mark of the miR-375 locus.

3.3.5 Transcriptional repressors bind to the miR-375 locus

The bisulfite sequencing data indicated that one feature of the repressed epiallele of miR-375 is local hypomethylation around CpG 18 of CGI 1, which in the case of MCF-12A and MDA-MB-231 cells also became detectable throughout the CGI (Fig. 31B). Analysis of the

miR-375 locus with MatInspector software (144) revealed the presence of consensus binding sites for the CCCTC-binding factor (CTCF) protein in this locus and especially in the hypomethylated region. CTCF is a highly conserved multifunctional zinc finger protein involved in transcriptional repression and activation, insulation, imprinting and X-inactivation that binds preferentially to unmethylated DNA (146, 147). CTCF is a very widely expressed factor that is expressed in many breast cancer cell lines including MDA-MB-231 and MCF7, but also in non-tumorigenic breast cell lines like MCF-12A (148). Moreover, we identified several Z- and E-boxes that are potential binding sites for the ZEB1, a transcriptional repressor that has been found to be involved in the regulation of several cancer-associated genes (149, 150). ZEB1 has been described to be expressed in MCF-10A and MDA-MB-231 cells; however, almost no expression was reported in MCF7 and T47D cell lines (149). We therefore performed ChIP with antibodies against CTCF and ZEB1 in MCF7, MCF-12A and MDA-MB-231 cells. ChIP was also performed with antibodies against ERα and the largest subunit of RNA polymerase II to obtain information about ERα-binding and active or paused transcription events. We confirmed the binding of CTCF not only to the predicted binding site around CpG 18 (ChIP-1) in MCF12A cells but also to sites in the proximal region of CGI 1 (ChIP-2) and in CGI 2 (ChIP-3 and 4) (Fig. 32). In MDA-MB-231 cells, a similar pattern was found, though peak binding was observed in all but the region amplified by the ChIP-1 primers. In contrast, MCF7 cells showed a weak enrichment for CTCF-immunoprecipitated DNA only in the proximal region of CGI-1. These data suggest that CTCF regulates the miR-375 locus by interacting with several hypomethylated binding sites and creating a higher order chromatin structure that prevents active transcription.

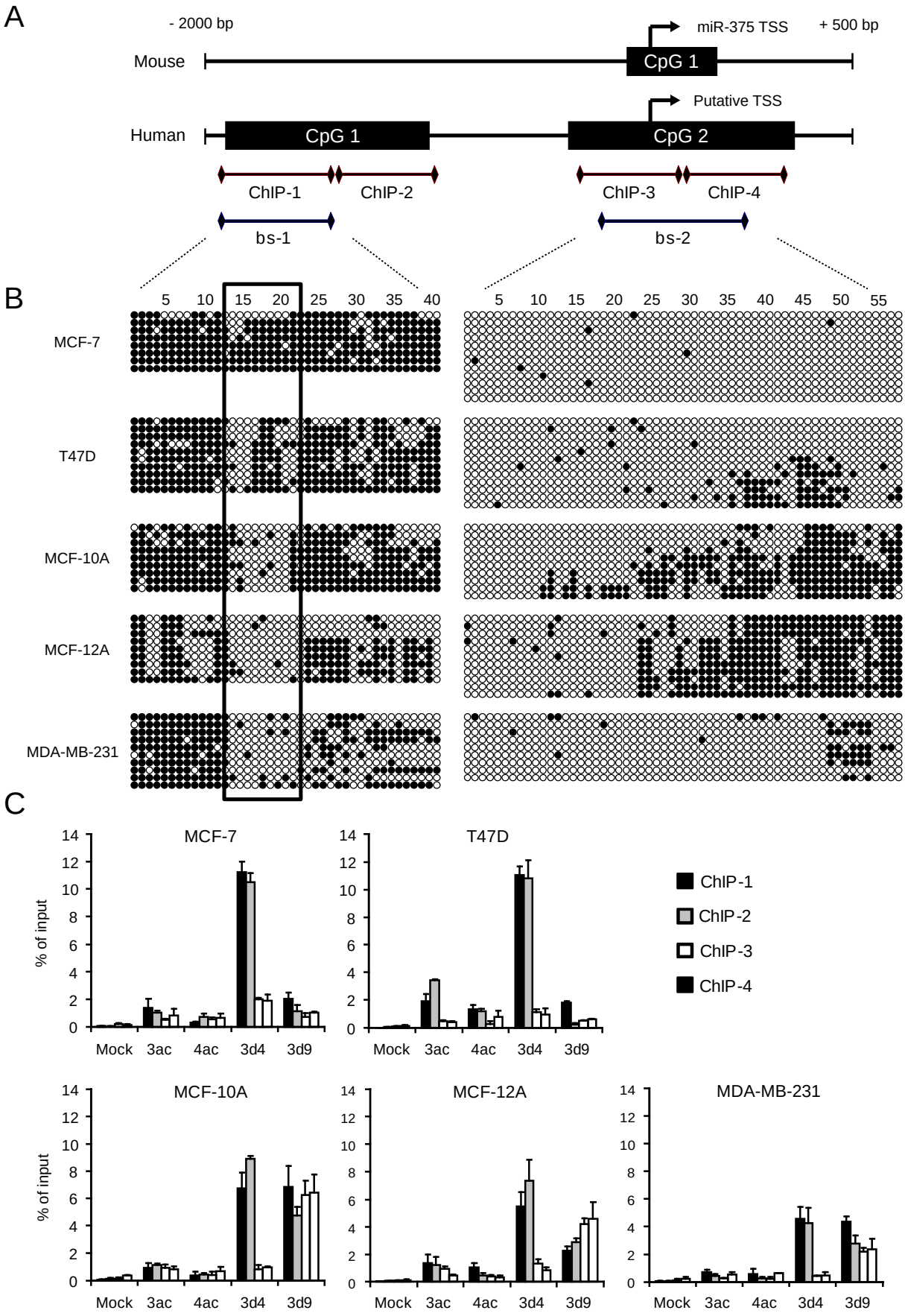


Figure 31. Epigenetic marks determine the transcriptional state of the miR-375 locus. (A) Comparison of the miR-375 locus in human and mouse. CGIs (CGI 1 and 2) are shown. The

transcription start site (TSS) of miR-375 is indicated by arrows. The locations of two primer pairs used for bisulfite sequencing (bs-1 and bs-2) and the four primer pairs employed for the ChIP analysis (ChIP-1, ChIP-2, ChIP-3, ChIP-4) are depicted. **(B)** Bisulfite sequencing of the CGIs in five breast cell lines. Black and open circles represent methylated and unmethylated CpGs, respectively. The region with specific hypermethylation in MCF7 and T47D cells is shown in quadrangle. **(C)** ChIP analysis of the miR-375 locus in cell lines. Crosslinked chromatin of each cell line was immunoprecipitated with antibodies specific for acetylated histone H3 (3ac), acetylated histone H4 (4ac), dimethylated lysine 4 of histone H3 (3d4) and dimethylated lysine 9 of histone H3 (3d9). Purified DNA was amplified with the four ChIP primer pairs (see A)

RNA POLII was detected in most regions with CTCF enrichment, presumably representing paused polymerase molecules interacting with CTCF (Fig. 32). Consistently, low levels of RNA POLII were detected in the miR-375 coding region (ChIP-4) in MCF-12A and MDA-MB-231 cells. ZEB1 binding was restricted to CGI 2 in the cell lines with low miR-375 levels, which correlates well with the presence of E-boxes in the ChIP-3 region (Fig. 32). We found no ZEB1 binding in MCF7 cells, but very prominent peaks of ERα and RNA POLII in the miR-375 coding region (ChIP-4), adjacent to the putative miRNA promoter (Fig. 32). Collectively, these findings support a role of CTCF and ZEB1 in the repression and ERα in the activation of miR-375 expression. The binding of ERα to the putative miRNA promoter further supports our preceding findings on a key role of ERα in miR-375 overexpression in MCF7 cells and indicates the existence of a positive feedback regulation between these molecules.

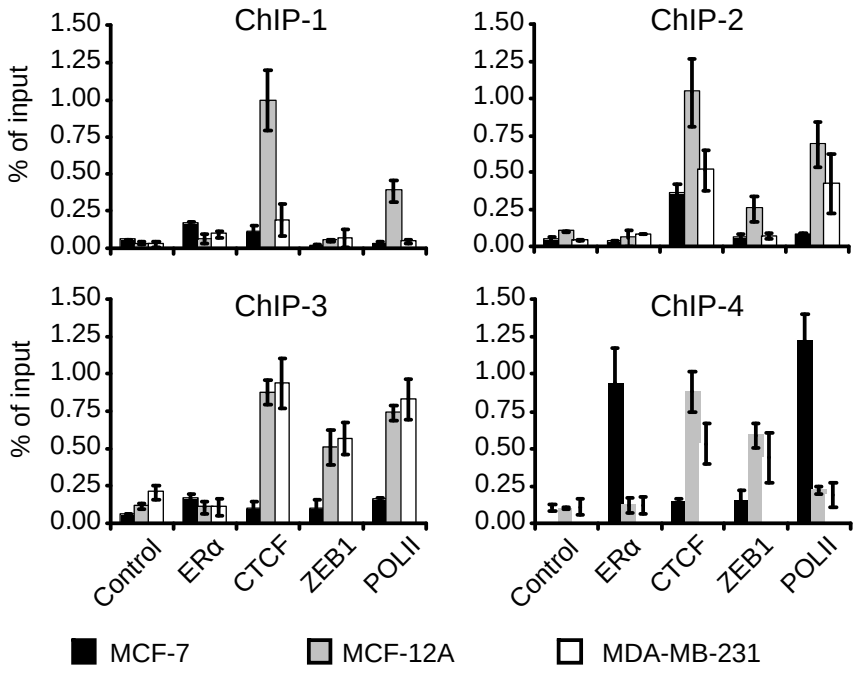


Figure 32. Transcriptional repressors bind to the miR-375 locus. Crosslinked chromatin derived from MCF7, MCF-12A and MDA-MB-231 cells was immunoprecipitated with antibodies specific for ERα, CTCF, ZEB1 and DNA polymerase II (POLII). Purified DNA was amplified with the four ChIP primer pairs (see figure 31A). Results are shown as percentage of the input (normalized against input).

3.3.6 RASD1 is a functional target of miR-375

We expanded our functional analyses by measuring the expression of miR-375 in nine pairs of primary breast carcinomas and adjacent normal tissues from breast cancer patients using quantitative real-time PCR. Although not specific to ERα-positive tumors (Supplementary Table S4), miR-375 was up-regulated in seven of nine analyzed tumors (Fig. 33A). To identify miR-375 targets, mRNA expression profiles of tumor and normal breast tissue specimens of 4 patients (patients 1, 2, 6, 7) showing differential miR-375 expression were analyzed by Illumina microarrays. We identified 125 genes commonly down-regulated in tumors over-expressing miR-375 (Supplementary Table S5). In parallel, 144 potential miR-375 targets were predicted using the TargetScan algorithm (151). Combining microarray profiling and target prediction data, we identified two genes, Ras dexamethasone-induced 1

(*RASD1*) and early B-cell factor 3 (*EBF3*), as potential miR-375 targets (Fig. 33B). We cloned segments of the 3'UTRs of both genes into luciferase reporter vectors and performed luciferase assays upon overexpressing miR-375 in HEK293T cells (that do not express endogenous miR-375) or blocking its activity in MCF7 cells. The EBF3 luciferase construct showed no sensitivity to miR-375. In contrast, reporter assays for RASD1 showed significant changes in luciferase activity (Fig. 33C). The observation that modulation of miR-375 caused consistent expression changes in *RASD1*-luciferase construct in both cell lines strongly suggests that RASD1 is a functional target of miR-375. Because of the lack of appropriate antibodies, the effect of miR-375 on RASD1 protein levels could not be evaluated. However, inhibition of miR-375 in MCF7 cells in more than 2.5 fold-induction in *RASD1* mRNA levels, as measured by quantitative RT-PCR (Fig. 33D).

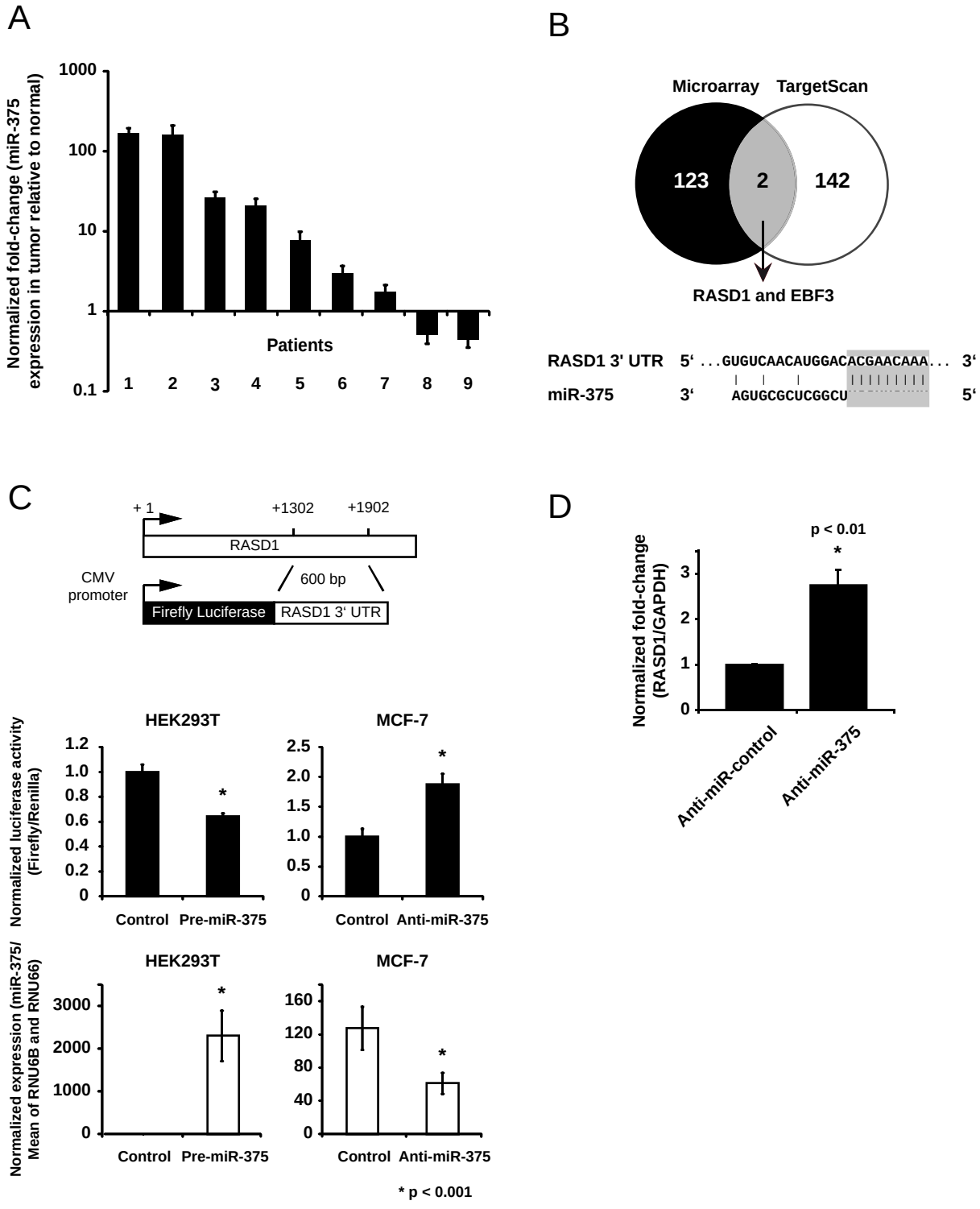


Figure 33. RASD1 is a target of miR-375. (A) miR-375 expression in breast cancer patients measured by qRT-PCR. Values are represented as the ratio of miR-375 expression in tumors *versus* matched normal tissues. **(B)** Comparison between 125 genes commonly down-regulated in patients showing overexpression of miR-375 and 144 predicted targets identified EBF3 and RASD1 as potential miR-375 targets. **(C)** Luciferase assays in HEK293T and MCF7 cells transfected with the *RASD1* 3' UTR reporter construct (middle). The respective miR-375 expression is indicated (lower part). **(D)** *RASD1* mRNA expression in MCF7 cells after transfection with anti-miR-375 measured by qRT-PCR.

3.3.7 Regulatory connection between ERα and RASD1

It has been reported that RASD1 is able to suppress cell growth of breast cancer cells and to inhibit clonogenic growth of MCF7 cells (152). This report, combined with our findings, led us to hypothesize that RASD1 could function as a negative regulator of ERα. The effect of RASD1 on ERα was tested by loss- and gain-of-function experiments in MCF7 cells. While silencing of RASD1 by specific siRNAs gave rise to an increase in ERα levels, overexpression of RASD1 had the opposite effect and led to downregulation of ERα (Fig. 34A). These findings support a negative regulatory effect of RASD1 on ERα expression.

It has been shown that steroid hormones can positively modulate RASD1 expression in AtT-20 corticotrope cells (153). Therefore, we asked whether RASD1 could be regulated by estrogen in MCF7 cells. To address this question, MCF7 cells were treated with estradiol (E2) and *RASD1* mRNA levels were monitored over time. Interestingly, *RASD1* was induced more than 10-fold 3 hours after treatment (Fig. 34B), whereas mRNA of *pS2* (*TFF1*), a well-known classical estrogen responsive gene, became elevated after 16 hours (Fig. 34B). The reduction of *RASD1* transcript levels after 3 hours may be explained by the finding that ligand activation of ERα (e.g. activation by estrogens) is followed by its destabilization through degradation by the proteasome pathway (154). To investigate an association between E2-dependent *RASD1* induction and ERα downregulation, we analyzed ERα protein levels in different time points after treatment with E2. We observed that substantial reduction of ERα protein levels occurred simultaneous to the E2-dependent up-regulation of *RASD1* mRNA (Fig. 34C). Together, our data show that; i) E2 induces *RASD1* mRNA levels in a rapid and effective manner, ii) the level of ERα protein, which is the main mediator of E2, is reduced simultaneously with RASD1 up-regulation and iii) ERα is negatively regulated by RASD1 (as shown by the silencing and overexpression

experiments). These observations suggest a negative feedback regulation of ERα signaling which is mediated by RASD1.

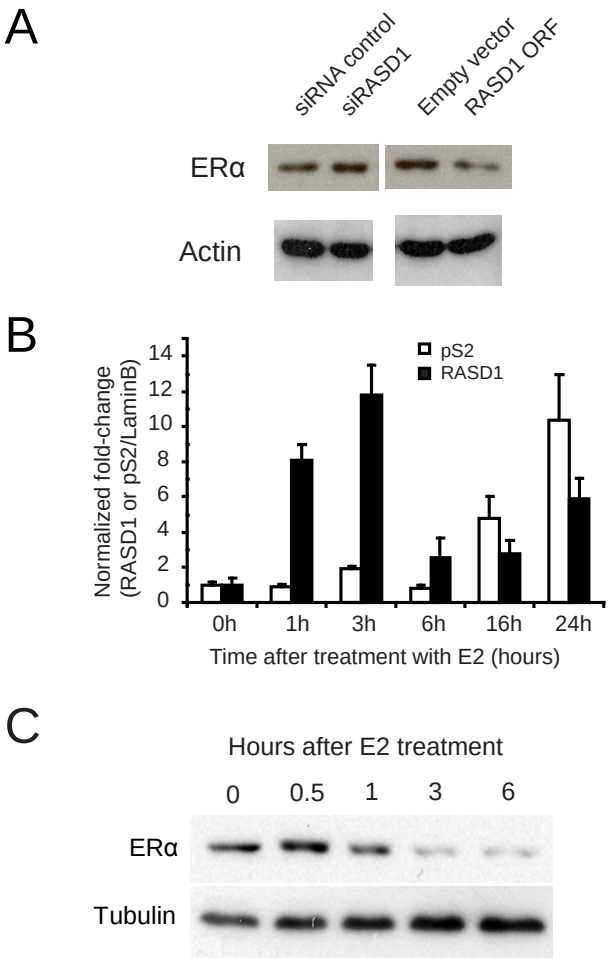


Figure 34. Regulatory connection between ERα and RASD1. (A) RASD1 negatively regulates ERα. Immunoblots for ERα in MCF7 cells 48h after silencing or overexpression of RASD1. **(B)** qRT-PCR data for *RASD1* and *pS2* transcripts in E2-treated MCF7 cells. **(C)** Immunoblots for ERα in MCF7 cells following E2 treatment.

4 Discussion

4.1 DNA methylation analysis

4.1.1 Development of microarrays

Although the first technique used for broad DNA methylation profiling was based on two-dimensional gel electrophoresis (155), microarray technology soon found its way into the field. The adaptation of microarrays to the output of all major DNA pre-treatment procedures (mentioned in the Introduction) has resulted in diverse array-based techniques for determining DNA methylation profiles (53). Microarray formats that are used for methylation analysis are mainly based on either “CpG island arrays” (CGI arrays) or “oligonucleotide arrays”. CGI arrays have enabled the examination of the methylation patterns of multiple CGIs- following treatment of DNA with an endonuclease or affinity enrichment- in order to find the CGIs which show hypermethylation in tumors. However, these arrays provide low resolution about the methylation patters within a CGI. On the other hand, in addition to endonuclease treated- or affinity enriched-DNA, oligonucleotide arrays have also been adapted to bisulfite-treated DNA, providing an opportunity for high resolution analyses (54).

The first step of this study was aimed at the development of a dedicated oligonucleotide microarray to screen DNA methylation patterns of multiple candidate loci on bisulfite-treated DNA. The ultimate goal of the project was the establishment of a DNA methylation signature potentially useful for diagnosis of breast cancer. As it has been reported that CpG sites within a CGI can differ in their methylation statuses, a high resolution analysis of each locus was required. Oligonucleotides can be designed against any DNA sequence of interest, thus oligonucleotide arrays would be theoretically capable of dissecting the methylome at single-CpG-resolution. However, indeed, array hybridization methods are not

easily adapted to bisulfite-converted DNA. This is due to the fact that following bisulfite-conversion, DNA sequence consists of essentially three different bases instead of four, with most cytosines being converted into thymines after PCR, resulting in reduced complexity of the genomic sequence. This feature of treated-DNA results in much sequence redundancy and, thus, decreased hybridization specificity. This problem has been reflected in the previous studies that applied bisulfite-treated DNA to oligonucleotide arrays. In these reports, either despite the high resolution of analyses, the coverage of assays was limited to a single gene (59, 156), or when several genes were analyzed, the assay resolution was low (58, 62).

Considering the above mentioned facts, the specificity of array performance was assessed through different approaches in the initial step of microarray development process. Owing to its flexibility and fast performance- which is a valuable advantage in the process of technical optimization of customized arrays- the Geniom One technology (febit) was selected for this analysis. Several parameters related to the labeling of samples, duration and temperature applied in array hybridization and washing procedures had already been optimized (61). However, the effect of probe length- which directly influences the specificity of probe hybridization- had not been addressed systematically. Therefore the evaluation process was focused on the length of probes. According to the suggestions made in the previous reports (61, 62), the lengths of 17, 21 and 25 nucleotides were chosen for this experiment.

The first analysis, examining the effect of probe length on array reproducibility, showed that probes with length of 17 nucleotides are weaker than those with 21 and 25 nucleotide in terms of reproducing the same data. Twenty one and 25 nucleotide probes showed a similar power in overall array reproducibility. However, in the second experiment, which was more dedicated toward reproducibility of probes functioning in different array layouts (designs), 25 nucleotide probes clearly showed an advantage to other probes. These

experiments provided evidence that probes in length of 25 nucleotides are more stable -in terms of performance- and less influenced by potential experimental factors (like sample handling) and environmental features such as array composition.

The third experiment was planed such that cross-hybridization issue could be directly analyzed. While some of the detected signals in “condition 1” could originate from non-specific target binding, all signals obtained in “condition 2” were from specific binding of the studied target to the related probes. Hence, the difference observed between data produced in these conditions could be referred to the binding of non-specific targets to the probes analyzed. This experiment showed that while cross-hybridization can significantly influence the performance of 17 nucleotide probes, it can not disturb the functioning of 25 nucleotide oligos, effectively. Taken together, these analyses suggested 25 nucleotide long as the optimum length of the oligonucleotide probes.

During the array development procedure, it was realized that individual probes with same length (e.g. 25 nucleotides) exhibit different abilities in showing the correct methylatin levels. Based on this observation, probe calibration and selection were performed. As expected, using 25 nucleotide as the probe length, more probes were selected (according to the selection criteria), referring to the fact that probes with this length are more able to produce reliable data. It is worth noting that vast majority of the probes were able to show the correct methylation values for completely unmethylated and fully methylated samples. The difference between high and low discriminative probes was revealed when we subjected them to the control samples representing partial DNA methylation.

The complexity and difficulty of adaptation of oligonucleotide arrays to bisfulfite converted DNA may reduce when targeting non-CGI CpGs. In contrast to CGIs, CpGs are not densely packed together in non-CGI regions. This feature decreases the sequence redundancy between these loci, providing a chance to use longer probes in order to increase the specificity of binding. Therefore, it seems that oligonucleotide arrays can be implemented

as powerful tools for genome-scale profiling of methylation patterns of non-CGI CpGs. Furthermore, the usefulness of this platform to screen CGIs methylation patterns is indubitable; however, a careful selection of highly discriminative probes is necessary prior to the analysis of actual clinical samples.

4.1.2 DNA methylation patterns of breast tumors

Analysis of the DNA methylation profile of samples was directed such that methylation markers that can distinguish tumors from control samples could be identified. At first step of the analysis, 93 CpGs were identified whose methylation levels significantly separated tumors from control samples as depicted in clustering projection in figure 20. These CpG sites pertained to CGIs associated with sixteen genes. Additionally, this methylation profile divided tumor samples into four subgroups. Due to the lack of complete clinical data (e.g. hormone receptor or menopausal status) a detailed analysis on patients of different groups was not possible. However, using available data, we did not find a difference in tumor staging or lymph nodes status between the four groups. The relative lower DNA methylation levels of group B-patients might partially be explained by their ages as they are, in average, younger than other patients. The mean of ages of the members was 50.8 ± 11, 42.8 ± 7.2, 52 ± 11.4 and 56 ± 14 years in groups A, B, C and D, respectively. Nevertheless, this does not affect the cancer-association of these methylation markers, because they are able to discriminate between tumors from their patient-matched control (non-tumor) samples (Fig. 35).

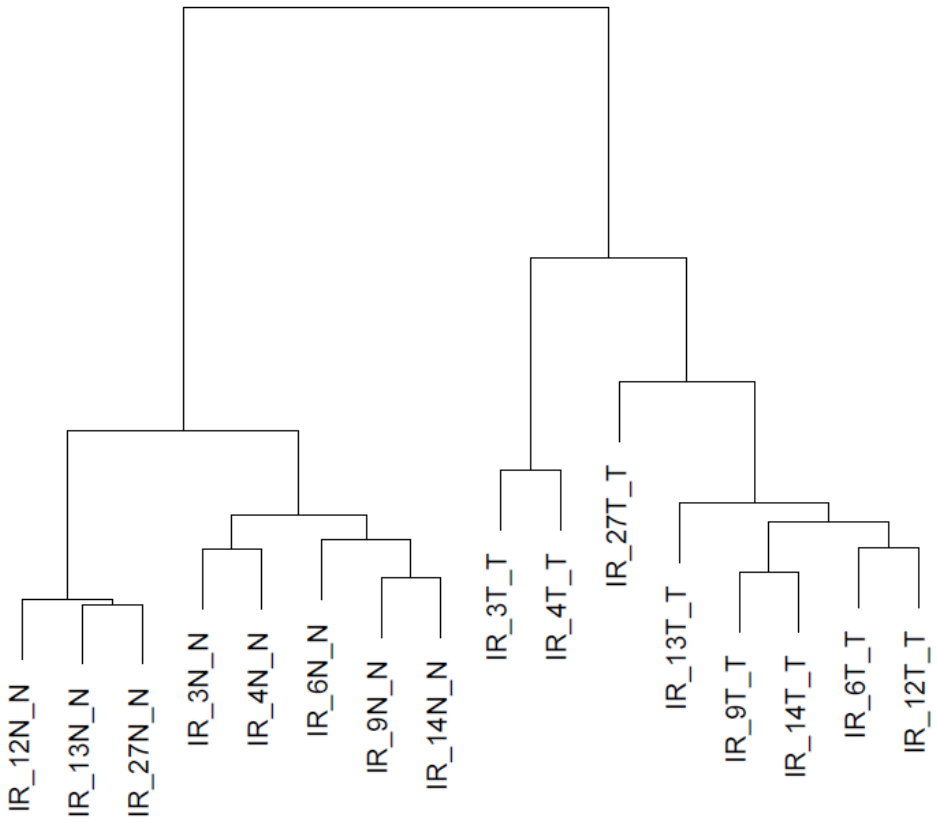


Figure 35. Dendrogram clustering of tumors and their corresponding patient-matched control samples by the 93 CpGs methylation markers. T: Tumor**, N:** Normal from same patient.

In this regard, recent studies have reported on an age-related DNA methylation of CGIs that are target of polycomb group proteins (PcGs) (157). This age-dependent DNA methylation has been shown to be a hallmark of cancer (158). Intriguingly, among the hypermethylated genes identified in our study, *SFRP1*, *SFRP2*, *SFRP4*, *SFRP5*, *HDPR1*, *HIC1*, *RASSF1A* and *GHSR* are *bona fide* PcG targets (159-161).

4.1.3 Genes affected by hypermethylation in tumors

Wnt signaling inhibitors or associated genes

Wnt signaling is a pathway that contributes to cell proliferation by transducing mitogenic signals from cell membrane to the nucleus. In addition, this pathway helps cells to remain in a relatively undifferentiated state. The pathway is activated by binding of Wnt factors to the Frizzled receptors on cell surface (2). It has been demonstrated that hyperactivity of

Wnt signaling has oncogenic effects on mammary epithelium (162). Interestingly, almost half of the hypermethylated genes which were identified in this study produce proteins that have an association with Wnt signaling; *SFRP1, SFRP2, SFRP3, SFRP4, SFRP5, HDPR1, WIF-1, TCF4*.

The first five genes encode members of secreted frizzled-related (SFRP) protein family that function as soluble modulator of Wnt signaling. SFRPs contain an N-terminal domain homologous to the cysteine-rich domain of Frizzled receptors. This provides them with a capacity to inhibit Wnt signaling by competing with Frizzled receptors for Wnt binding or by binding directly to Frizzled (163). *WIF1* encodes Wnt inhibitory factor-1 (WIF1), a secreted protein that binds to Wnt proteins and inhibits their activities (164). *HDPR1* (*DACT1*) is coding Dapper (Dpr) protein. This protein is a negative regulator of Dishevelled (Dvl) protein which acts as a signal transducer in Wnt signaling (165). Transcription factor 4 is coded by *TCF4* gene. A role for TCF4 as a transcriptional repressor of Wnt signaling has been found in breast cells (108). Therefore, products of all these genes are implicated in the control of Wnt pathway activation.

CST6

Cystatin M (Cystatin E/M), encoded by *CST6*, is lysosomal cysteine proteinase inhibitor that is down-regulated in metastatic breast tumor cells as compared to primary tumor cells (166). It regulates the activity of two cysteine proteases, the cathepsins B and L. Deregulation of expression and activity of lysosomal cysteine proteases has been implicated in cancer progression. In line with this, imbalances between these proteases and their inhibitory cystatins can lead to tumor development, invasion and metastasis. This gene is silenced in breast cancer due to the DNA hypermethylation (167).

C/EBPα

CCAAT/enhancer binding protein α (C/EBPα) is a member of C/EBP family of transcription factors. These proteins are implicated in the regulation of cell proliferation and growth arrest in different tissues including mammary gland (168). Down-regulation of C/EBPα in both mRNA and protein levels has been reported in primary breast cancer samples. As ectopic expression of this protein in breast cancer cell lines results in inhibition of cell growth, a tumor suppressor activity has been suggested for this protein (101).

HIC1

Hypermethylated in cancer 1 (HIC1) gene was identified as a candidate tumor suppressor gene in 1995. The product of this gene is a transcriptional repressor which is ubiquitously expressed in normal tissues but down-regulated in tumor cells (169). A role for DNA hypermethylation in down-regulation of this gene in breast cancer was reported in 1998 (121). Recently, based on its ability to inhibit the growth of breast cancer cells, a tumor suppressor activity for HIC1 has been suggested (170).

PAK3

PAK3 is a member of the p21-activated kinases (PAKs) family. PAK proteins are critical effectors that link Rho GTPases to cytoskeleton reorganization and nuclear signaling. In mammalian cells, PAKs are one of the targets of Cdc42 and Rac, regulators of formation and functions of motile structures in mammalian cells (171). Although involvement of PAK1 in cell motility and invasiveness of breast cancer cells has been documented (172), no similar effect for PAK3 has been reported so far. *PAK3* hypermethylation has been reported in esophagus, lung, cervix, head and neck, bladder and breast cancers (110).

PTEN

PTEN is a well-known tumor suppressor that controls the activation of Akt/PKB pathway. *PTEN* is inactivated in variety of human cancers by mutations or DNA methylation. This inactivation leads to hyperactivation of Akt/PKB pathway which, in turn, has positive effects on cell survival, growth and proliferation (173). Recently aberrant DNA methylation of PTEN promoter has been reported in early invasive breast cancers (174).

RASSF1A

Ras-associated factor 1 is a member of RASSF family of proteins that mediate some of the growth inhibitory functions of Ras proteins. *RASSF1A* is frequently inactivated in a huge variety of cancers by promoter hypermethylation, deletion and rarely by mutation. The protein acts as tumor suppressor through various independent pathways including apoptosis, the regulation of microtubule dynamics and mitotic arrest via multiple effectors. Thus, RASSF1A is involved in the control of both cell growth and migration (175). Although hypermethylation of this gene in breast cancer has been known for years, an association between this hypermethylation and estrogen receptor-status of the tumor has recently been identified (174).

GSTP1

Glutathione *S*-transferase P1 catalyzes intracellular detoxification reactions including inactivation of electrophilic carcinogens by conjugating reactive electrophiles to glutathione. *GSTP1* has been found to be silenced by DNA methylation in different cancers and frequent hypermethylation of this gene has been reported in different stages of breast carcinomas (176). Recent data show that DNA hypermethylation at *GSTP1* locus is more frequent in estrogen receptor positive breast tumors (174).

GHSR

GHSR gene encodes a member of the G-protein coupled receptor family called growth hormone secretagogue receptor. This protein acts as the receptor for the circulating peptide hormone ghrelin which has a stimulatory effect on growth hormone release (177). Hypermethylation of *GHSR* accompanied by its down-regulation has been reported in breast cancer (123). The involvement of GHSR in breast cancer formation or inhibition is not clear, as no functional study has been done on this protein in breast cancer. However, it has been proposed that the deregulation of the gherlin/GHSR axis may play an autocrine/paracrine role in hormone-dependent cancers (178).

To our knowledge, among the genes reviewed above, *SFRP3*, *SFRP4*, *HDPR1*, *TCF4* and *C/EBPα* are found with differential DNA methylation between tumor and normal breast tissues for the first time in this study. We could examine the effect of DNA methylation inhibition on the expression of the three former genes. As re-expression of these genes was observed following treatment of breast cancer cells with decitabine, we suggest that DNA hyperemethylation is a potential factor involved in down-regulation of these genes in breast cancer. Further examinations of possible reactivation of *TCF4* and *C/EBPα* in breast cancer cells following treatment with DNA methylation inhibitors will shed light on involvement of DNA methylation in their regulation.

4.1.4 DNA methylator event in breast cancer?

Further examinations of methylation profiles indicated that a heterogeneous pattern of DNA methylation exists within the tumors. For instance, among the 93 identified CpGs, hypermethylation were more prevalent for some CpG sites. In addition, a group of patients (Group A in Fig. 20) were found affected by a high degree of methylation in almost all of these loci, suggesting the existence of a DNA methylator event in these samples. Such a concordant methylation of multiple loci has been observed in different tumor entities

including colorectal cancer, ovarian cancer, acute lymphocytice leukemia, liver cancer, esophageal cancer, glioblastomas, just to name a few (179). This concept has been studied widely in colorectal cancer where it was referred to the "CpG Island Methylator Phenotype (CIMP)" for the first time (180). Toyota and colleagues showed that subset of patients suffering from colorectal cancer can be distinguished from others by simultaneous hypermethylation of several genes in their genome. Majority of patients positive for CIMP were also diagnosed with a mutation in *BRAF* gene and microsatellite instability (MSI) (181). Following the first report on CIMP in colorectal cancer, by Toyota et al., different studies have been conducted to investigate the existence and to characterize the CIMP in other cancers. It is now accepted that various genes can be the target of methylator event in different cancers (179).

Concomitant DNA methylation between multiple genes has also been reported in breast cancer. For instance, a correlation between hypermethylation of *RASSF1A* and *GSTP1* has been reported in different studies (e.g. (174, 182)). In line with the previous reports, these genes are among the hypermethylated genes identified in group A-patients in the current study. Likewise, we observed a concurrent hypermethylation for genes coding proteins known as Wnt antagonist family members. Majority of these genes have been reported as being hypermethylated in different cancers including acute lymphoblastic leukemia (ALL) (106), acute myeloid leukemia (AML) (183) and medulloblastoma (184). Among Wnt antagonist family genes, methylation dependent down-regulation has already been confirmed for *SFRP1*, *SFRP2*, *SFRP5* and *WIF1* in breast tumors (102, 104). Our data show that *SFRP3*, *SFRP4*, *HDPR1* and *TCF4* are also hypermethylated in breast cancer specimens. Strikingly, all these genes –which produce proteins that are functionally related- are simultaneously targeted by DNA methylation in group-A patients. The fact that these genes are located in different chromosomes but are silenced concomitantly by DNA methylation suggests that a connected epigenetic program might underlie their

hypermethylation. Such a selective process targeting essential tumor suppressor genes has been proposed as a reason for nonrandom distribution of the abnormal promoter CGI methylation in breast tumors (185). It is not clear to which extend the proposed methylator phenotype in breast cancer is similar to the CIMP phenomenon reported in colorectal cancer. Comparisons between mutation spectrums of the "would-be" CIMP-positive and CIMP-negative breast tumors will help to verify whether or not this phenomenon shares the same genetic background in both breast and colorectal cancers. Moreover, a detailed clinical characterization and follow-up data of patients affected by this DNA methylator event would help to verify if this event is associated with specific clinical manifestations.

4.1.5 DNA methylation signature for diagnosis of breast cancer

The study presented here aimed at identification of a DNA methylation signature which allows for diagnosis of breast cancer. Primarily, three types of samples were included in the analysis: breast tumors (IDC type); patient-matched normal breast tissue specimens from some of the patients and normal breast tissue samples from healthy (cancer-free) individuals. Among these, the patient-matched normal and healthy breast samples were used as controls each from different perspective. The normal breast tissues from patients served the purpose of establishing a tumor associated marker by excluding the methylation patterns induced mainly by factors such as environmental influences and ageing (186). By including samples from cancer-free individuals, we aimed to avoid missing the methylation markers that might act as predisposing factors to breast cancer.

The statistical analysis applied to the microarray data led to discovery of 14 CpG sites (pertaining to *SFRP2* and *GHSR* gene CGIs) that possess the most discriminative power between tumor and non-tumor samples (exhibiting adjusted P-values ranging between 10^{-6} and 10^{-12}). On careful examination of the correspondance analysis plot (Fig. 21) one can

observe that three benign samples (samples 49, 55 and 57 coded in pink) have been seperated from benign cluster and are among the tumor samples (represented in green). Following a request to the clinical partners on re-analysis of their pathology tests, sample number 55 was diagnosed with IDC which fits very well with the pathology reports of its neighbours in the CA plot. The other two patients (numbers 49 and 57) - which are seperated from the core tumor cluster on the plot- were reported to be affected by intraductal neoplasia. Therefore, the clinical re-evaluation data supported the ability of the identified signature to diagnose invasive ductal carcinoma.

One important finding of the present study is the observation that the signature CpGs have higher levels of DNA methylation in normal samples collected from cancer patients as compared to those obtained from healthy donors. Consistent with our data, Yan *et al*. have identified loci hypermethylated in breast tumors that showed no or low methylation in normal breast cells of cancer-free individuals. Notably, these loci were frequently hypermethylated in normal tissues adjacent to the tumors with hypermethylation in same loci (187). Considering these data, it becomes clearer that hypermethylation of DNA in normal-appearing tissue of patients affected by cancer is more prevalent than previously recognised. This suggests that early cancer-associated methylation patterns might have occurred also in these tissues. Being aware of this point is particularly important in marker identification when the goal is DNA methylation markers that may be useful for early cancer diagnosis and risk assessment. In such a case, using patient-matched normal tissues from regions adjacent to the actual tumor as the sole control is probably insufficient. Interestingly, the intensities of DNA methylation at signature CpG sites in benign samples were higher than those found in healthy controls but lower than observed in tumor samples. In line with this observation, Huang and colleagues have shown that genes frequently methylated in breast tumors are also methylated in benign breast lesions, albeit at a lower level (188). Hypermethylation of tumour suppressor genes has also been reported in women

who are not affected by cancer but are at risk to develop breast cancer. This abnormal change occurs more frequently in benign breast epithelium of women at high risk for breast cancer than in people at low risk (189, 190). Strikingly, abnormally methylated DNA has even been identified in mammary epithelial cells with normal morphology in high-risk women (191). Combining these data with the possible seminal role of epigenetic abnormalities in the earliest steps of cancer initiation (192, 193) suggests a potential cancer predisposing role for DNA methylation and support the emerging evidence on an involvement of aberrant DNA methylation patterns in the proposed "field defect" (or "field cancerization") concept (194-196). Among the two genes associated with our signature CpGs, *SFRP2* methylation in normal breast tissues adjacent to tumors has also been noted in another recent report (102). As discussed earlier, SFRP2 is involved in modulation of Wnt signaling pathway. Wnt signaling is conserved among various adult stem cells and its deregulation is associated with oncogenesis (197). In mammary gland, hyperactivity of Wnt signaling can result in deregulated self-renewal and can provoke dedifferentiation to a stem-like state; a feature of "cancer stem" (or "tumor initiating") cells (162). Taken together, it is conceivable that hypermethylation of genes implicated in regulation of Wnt pathway (e.g. *SFRP2*) is among the early events in breast cancer tumorigenesis which exert first predisposing steps toward malignancy. To which extend GHSR might be implicated in such a cancer-predisposing event is not clear. Nevertheless, it is plausible that *GHSR* hypermethylation is mediated by the same mechanism which is responsible for hypermethylation of *SFRP2* as both genes are targeted by PcG proteins (159).

4.1.6 DNA methylation patterns at signature loci

Another aspect that was addressed in this study was the pattern of DNA methylation in the interesting loci. Substantial evidence has been amassed showing that CGI methylation is a step-by-step process which starts initially at certain CpG dinucleotides in a given CGI and

then spreads all over the CGI over time (by ageing) (179). It is demonstrated *in vitro* that CGI methylation is acquired gradually over many cell cycles in the benign breast cells escaping senescence (198). Moreover, recent studies have indicated that substantial differences exist in DNA methylation patterns of different parts of a given CGI (199). These findings emphasize on the importance of the position in a given CGI where DNA methylation is studied to draw valid conclusion in a studied biological context. By analyzing the methylation patterns of the signature loci, we realized that some certain CpG sites exhibit an increasing gradient in the load of methylation from healthy samples to the tumors, as described in the results. *In silico* analysis revealed that most of these CpGs lie in the annotated binding sites of CTCF protein. CTCF is an important component of the epigenetic system regulating the complex interplay between DNA methylation, higher order chromatin structure and gene expression. Existing evidence suggests a role of CTCF in maintaining the methylation patterns in a cell and protecting the CpG islands from *de novo* methylation (146). CTCF binding to the DNA can be disturbed by DNA methylation, even if only one CpG dinucleotide in the binding site is modified by methylation (200, 201). In combination, one can speculate that methylation of CpGs located in the binding sites of CTCF, and thereby inhibition of CTCF binding, is one of the initial events in the process of establishment of abnormal DNA methylation patterns on CGIs associated with CFCT binding, such as *SFRP2* CGI.

We observed that some of the signature CpGs located at *GHSR* CGI are located in the annotated binding sites of PAX family transcription factors, particularly PAX5. PAX proteins are reported to be very important in the early developmental stages. These transcription factors regulate various cellular processes like tissue development and cellular differentiation in embryos. They are involved in cellular proliferation, cell-lineage specification, migration and survival. Although expression of PAX proteins is attenuated in normal conditions after the completion of development (202), a constitutive expression of

these proteins has been reported in cancer (203). Interestingly, nuclear accumulation of PAX5 has recently been associated with the autocrine growth hormone-induced cell proliferation of MCF7 breast cancer cells (204). As an autocrine/paracrine role for GHSR in hormone-dependent cancers has been suggested (178), it seems likely that GHSR may be involved in PAX5 related cell proliferation.

Whether CTCF and PAX proteins bind to the *SFRP2* and *GHSR* CGIs could not be elucidated in this study. Empirical evidence including data on the effect of DNA methylation patterns on binding of these proteins are needed to draw valid conclusions about epigenetic regulation of these genes.

4.2 Identification of miRNAs involved in ERα activation

Emerging evidences emphasize a fundamental role for miRNAs in different steps of tumor formation and progression. In the current study, we demonstrate that miR-375 functions as an activator for ER-α signaling in breast cancer cells and its depletion gives rise to an attenuated activity of ER-α pathway and reduced cell proliferation.

We used microarray technology to simultaneously monitor the expression patterns of hundreds of miRNAs in human mammary cell line models. Due to its high expression level in ERα-positive cancer cells when compared to other cell lines, we identified miR-375 as a potential new factor in breast cancer development. Interestingly, miR-375 has not been identified as a breast cancer-associated miRNA in previous studies analyzing miRNA expression patterns. In some studies, this was due to the fact that miR-375 was not present on the microarrays (86, 205). In other reports, the experimental settings and the parameters applied in the data analysis did not identify a differential expression of miR-375 between breast cancer cell lines (206, 207). Notably, our screening revealed that miR-375 is overexpressed specifically in ERα-positive breast cancer cells. To date, only a limited number of miRNAs with a regulative connection to the ERα pathway have been discovered.

Of these, miR-18a, miR-22, miR-206 and miR-222/221 directly target ERα mRNA and reduce ERα expression (92, 93, 208, 209), whereas miR-17-5p attenuates ERα activity - and subsequently cell proliferation - by inhibiting *AIB1* (an ERα coactivator) translation (210). Interestingly, all these miRNAs act as inhibitors of ERα signaling pathways. In contrast, miR-375 is the first identified miRNA that has the capacity to enhance ERα signaling in breast cells and, thus, to promote cell proliferation. The proliferative activity of miR-375 seems to strongly dependent on the cellular context, as its overexpression in gastric carcinoma cells led to decreased number of viable cells by induction of apoptosis (211). In agreement with the proliferative activity of miR-375, Poy and colleagues showed that miR-375 knockout mice have impaired proliferation in pancreatic β-cells in which a high endogenous expression of miR-375 is found (212). In another recent study, miRNA profiling of patients with Barrett's-associated adenocarcinoma identified miR-375 as a potential marker for cell proliferation (213). Of significance, we also observed overexpression of miR-375 in ERα-negative primary tumors. These observations suggest that miR-375 might modulate cell proliferation through other mechanisms than ERα signaling as well. More work is required to dissect the molecular networks which are influenced by miR-375.

4.2.1 Epigenetic marks at miR-375 locus

miRNA genes have been previously described to be epigenetically regulated by DNA methylation (78). For miR-375, increased DNA methylation in CGI 2 (containing the pre-miR-375 hairpin sequence) was detected in non-cancerous cell lines, but not in MDA-MB-231 cells, although low levels of miR-375 are present in all those cells. However, we found that DNA methylation patterns in a distal part of CGI 1, which contains potential binding sites for the insulator protein CTCF, are crucial for the repression of miR-375. DNA demethylation in this region correlated with CTCF binding and low expression of miR-375.

As CTCF binds preferentially to unmethylated DNA, we suggest that hypomethylation of this region is necessary for CTCF recruitment and subsequent silencing of the miR-375 locus. CTCF binding also correlated with the repressive mark H3K9me2 which was found throughout the locus in cells with a low miR-375 expression. In summary, our data indicate the presence of a repressive, CTCF-dependent, chromatin structure at the miRNA locus in cells with low miR-375 expression. Such a repressive structure would either directly prevent transcription, or inhibit the interaction of the miR-375 promoter region with activating factors and enhancer elements (insulation). Additionally, the binding of the transcriptional repressor ZEB1 was detected at a region encompassing the putative miR-375 promoter. This region contains several evolutionary conserved E-boxes, ZEB1-binding sites, suggesting that ZEB1 contributes to the repression of miR-375. However, the miR-375 locus maintains the potential to be reactivated, as indicated by a peak of H3K4me2 on the distal part of CGI 1 in cell lines with a silenced locus. In MCF7 cells, we found significantly reduced levels of H3K9me2, weak CTCF and ZEB1 interactions with their target sites and high abundance of ERα and POLII near the putative promoter of miR-375 and in its transcribed locus. These findings are in agreement with recent data showing that CTCF confines estrogen receptor action on a genome-wide scale (214). Furthermore, the CTCF antagonist BORIS, which seems to interact in a methylation-independent way with CTCF-binding sites (215), has been found to be aberrantly expressed in breast tumors, which correlated in many cases with over-expression of ERα (216). A role of CTCF as tumor suppressor has previously been suggested, as CTCF has been found to inhibit cell growth and induce cell cycle arrest. CTCF has a repressive role in the regulation of several prominent oncogenes and also seems to be important in preventing epigenetic silencing of growth suppressor genes (147).

4.2.2 Proposed model for RASD1 and miR-375 effects on ERα

Our data show that the modulation of ERα by miR-375 is achieved through the repression of negative regulators of the ERα signaling, such as RASD1. RASD1 is known as a guanine exchange factor for $G\alpha_{i/o}$ proteins (217) which is able to inhibit adenylyl cyclase, reducing Gsα -induced cAMP production in HEK293T cells (218). Vaidyanathan et al. demonstrated that the anti-proliferative effect of RASD1 in MCF7 cells is independent of the activation of heterotrimeric G-proteins (152). In this context, Nguyen and Watts have proposed a mechanism by which RASD1 could indirectly regulate $G\alpha_{i/o}$ signaling by negatively regulating protein kinase C delta (PKC δ) activity (219). PKC δ is the predominant PKC isoform in MCF7 cells (220) and its activation supports proliferation in breast cancer cells (221). Interestingly, PKC δ was also reported to be involved in ERα signaling. PKC δ overexpression confers estrogen-dependent breast cancer cells a proliferative advantage (222) and resistance to tamoxifen (223). Taken together, it seems likely that RASD1 exerts its regulatory influence on ERα, at least in part, via inhibition of PKC δ activity in breast cells. Nevertheless, considering the rapid induction of *RASD1* mRNA by E2, other mechanisms of action of RASD1 leading to the destabilization and degradation of ERα seem conceivable. More research is warranted to explain the specific mechanisms by which RASD1 down-regulate ERα signaling.

Previous studies have described a negative feedback loop for ERα through several miRNAs which are induced and then down-regulate ERα upon estrogenic stimulation (224, 225). The present work suggests that, in abnormal conditions such as in cancer, ERα can induce aberrant up-regulation of miRNAs that disrupt its natural negative feedback regulation (Fig. 36). We showed that the up-regulation of miR-375 in MCF7 cells is dependent on the presence of ERα and that ERα binds near the putative promoter of miR-375. Because the

establishment of abnormal epigenetic patterns is an early event in tumorigenesis, miR-375 deregulation may prove to be involved in the up-regulation of ERα expression in mammary cells in early steps of tumor formation.

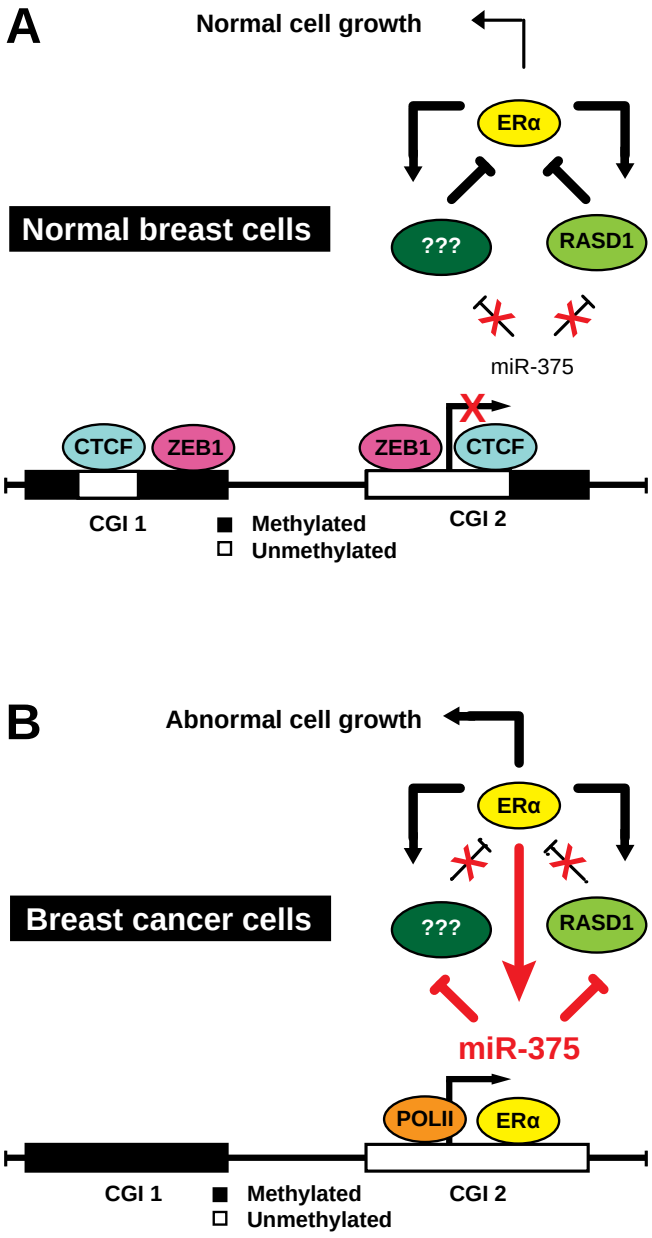


Figure 36. Proposed model describing miR-375 role in breast cancer development. (A) Under normal conditions, ERα activity is controlled through a negative feedback loop by estrogen inducible genes, such as RASD1. A low endogenous level of miR-375 is assured by a repressive combination of DNA methylation and the binding of the transcriptional repressors CTCF and ZEB1. Under these conditions cells grow normally. **(B)** In cancer, changes in DNA methylation and the consequential removal of repressive factors from miR-375 locus give rise to an open chromatin structure which attracts transcription factors to the promoter region and induces the transcription of miR-375. High amounts of miR-375 destabilize the balance of the negative feedback regulation of ERα by targeting RASD1 and increasing ERα activity. As a consequence, ERα can bind to miR-375 promoter and activates its expression which, in turn, creates a positive feedback loop leading to aberrant cell growth.

5 Outlook

In the first part of the thesis presented here, an oligonucleotide microarray-based analysis was applied for the identification of a DNA methylation signature capable of diagnosing breast cancer. The signature identified showed a promising potential to be applicable for this purpose, as it distinguished actual tumors from both control (normal) and benign samples. However, as all tumors included in this study were diagnosed as IDC type, this signature may represent an IDC-specific DNA methylation signature. In a new study pursuing the one presented here, the identified DNA methylation signature is tested in another set of tumor (IDC and non-IDC) and control tissue samples using a quantitative approach. The results of this analysis will allow for an evaluation of the robustness of this signature. In addition, quantitative data of the new study will be helpful for the analysis of the abnormal DNA methylation patterns in patient-matched normal-appearing tissue samples to address the possible association of these patterns to the field defect.

The DNA methylation mapping of the signature loci indicated a potential involvement of certain transcription factors in the establishment/maintenance of the aberrant epigenetic control of expression of the associated genes. Investigation of the relationship between these factors and between them and DNA methylation patterns may be helpful for understanding of the mechanisms underlying the establishment of abnormal DNA methylation patterns in breast cancer.

In the second part of this thesis, we reported on miR-375 as the first identified miRNA with a capacity to enhance ERα signaling in breast cancer cells. RASD1 was identified as a miR-375 target that negatively regulates ERα. However, the involvement of other potential molecules mediating miR-375 effect on ERα was not excluded. A few known targets of miR-375, which have been reported in other tissues, are implicated in signaling pathways that have crosstalk

with ERα signaling, such as PDK1 or Cav1. Investigating the potential regulative connection between these molecules and the ERα pathway as well as miR-375 in breast cancer cells may be helpful to elaborate on the molecular network involved in the positive feedback loop between miR-375 and ERα.

6 References

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006;367(9524):1747-57.
2. Weinberg RA. The Biology of Cancer. New York and London: Garland Science; 2007.
3. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. New York and London; 2007.
4. WHO. Fact sheet N°297, Cancer. WHO Fact sheets; 2009.
5. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nat Rev Cancer 2007;7(9):659-72.
6. McCafferty MP, McNeill RE, Miller N, Kerin MJ. Interactions between the estrogen receptor, its cofactors and microRNAs in breast cancer. Breast Cancer Res Treat 2009;116(3):425-32.
7. Shoker BS, Jarvis C, Clarke RB, *et al.* Estrogen receptor-positive proliferating cells in the normal and precancerous breast. Am J Pathol 1999;155(6):1811-5.
8. Platet N, Cathiard AM, Gleizes M, Garcia M. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol 2004;51(1):55-67.
9. Altekruse S, Kosary C, Krapcho M, *et al.* SEER Cancer Statistics Review, 1975-2007, based on November 2009 SEER data submission. Bethesda: National Cancer Institute; 2010.
10. Waddington CH. The epigenotype. Endeavour 1942;1:18-20.
11. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis;31(1):27-36.
12. Esteller M. Epigenetics in cancer. N Engl J Med 2008;358(11):1148-59.
13. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99(3):247-57.
14. Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J 2002;21(15):4183-95.
15. Jones PA. The DNA methylation paradox. Trends Genet 1999;15(1):34-7.
16. Ting AH, McGarvey KM, Baylin SB. The cancer epigenome--components and functional correlates. Genes Dev 2006;20(23):3215-31.
17. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16(1):6-21.
18. Gal-Yam EN, Saito Y, Egger G, Jones PA. Cancer epigenetics: modifications, screening, and therapy. Annu Rev Med 2008;59:267-80.
19. Kouzarides T. Chromatin modifications and their function. Cell 2007;128(4):693-705.
20. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 2007;8(4):286-98.
21. Yoo CB, Jones PA. Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov 2006;5(1):37-50.
22. Ambros V. The functions of animal microRNAs. Nature 2004;431(7006):350-5.
23. Lee Y, Kim M, Han J, *et al.* MicroRNA genes are transcribed by RNA polymerase II. EMBO J 2004;23(20):4051-60.

24. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol 2006;13(12):1097-101.
25. Lee Y, Ahn C, Han J, *et al.* The nuclear RNase III Drosha initiates microRNA processing. Nature 2003;425(6956):415-9.
26. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 2001;293(5531):834-8.
27. Diederichs S, Haber DA. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell 2007;131(6):1097-108.
28. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 2002;297(5589):2056-60.
29. Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. Genes Dev 2003;17(4):438-42.
30. Liu J, Carmell MA, Rivas FV, *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. Science 2004;305(5689):1437-41.
31. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 2004;15(2):185-97.
32. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA. Short RNAs repress translation after initiation in mammalian cells. Mol Cell 2006;21(4):533-42.
33. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009;11(3):228-34.
34. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68(4):820-3.
35. Knudson AG. Hereditary cancer: two hits revisited. J Cancer Res Clin Oncol 1996;122(3):135-40.
36. Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends Genet 1993;9(4):138-41.
37. Esteller M, Silva JM, Dominguez G, *et al.* Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst 2000;92(7):564-9.
38. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128(4):683-92.
39. Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet 1989;83(2):155-8.
40. Herman JG, Latif F, Weng Y, *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A 1994;91(21):9700-4.
41. Gonzalez-Zulueta M, Bender CM, Yang AS, *et al.* Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res 1995;55(20):4531-5.
42. Kane MF, Loda M, Gaida GM, *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res 1997;57(5):808-11.
43. Laird PW, Jackson-Grusby L, Fazeli A, *et al.* Suppression of intestinal neoplasia by DNA hypomethylation. Cell 1995;81(2):197-205.
44. Herman JG, Umar A, Polyak K, *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A 1998;95(12):6870-5.
45. Lengauer C, Kinzler KW, Vogelstein B. DNA methylation and genetic instability in colorectal cancer cells. Proc Natl Acad Sci U S A 1997;94(6):2545-50.

46. Ohtani-Fujita N, Fujita T, Aoike A, Osifchin NE, Robbins PD, Sakai T. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. Oncogene 1993;8(4):1063-7.
47. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999;21(2):163-7.
48. Hitchins MP, Wong JJ, Suthers G, *et al.* Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 2007;356(7):697-705.
49. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 2004;36(5):497-501.
50. Widschwendter M, Jones PA. The potential prognostic, predictive, and therapeutic values of DNA methylation in cancer. Commentary re: J. Kwong et al., Promoter hypermethylation of multiple genes in nasopharyngeal carcinoma. Clin. Cancer Res., 8: 131-137, 2002, and H-Z. Zou et al., Detection of aberrant p16 methylation in the serum of colorectal cancer patients. Clin. Cancer Res., 8: 188-191, 2002. Clin Cancer Res 2002;8(1):17-21.
51. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003;3(4):253-66.
52. Ballestar E, Esteller M. SnapShot: the human DNA methylome in health and disease. Cell 2008;135(6):1144- e1.
53. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet 2010;11(3):191-203.
54. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 2006;7(3):200-10.
55. Zuo T, Tycko B, Liu T, Lin HJ, Huang THM. Methods in DNA methylation profiling. Epigenomics 2009;1(2):331-45.
56. Frommer M, McDonald LE, Millar DS, *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 1992;89(5):1827-31.
57. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22(15):2990-7.
58. Adorjan P, Distler J, Lipscher E, *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res 2002;30(5):e21.
59. Gitan RS, Shi H, Chen CM, Yan PS, Huang TH. Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. Genome Res 2002;12(1):158-64.
60. Kimura N, Nagasaka T, Murakami J, *et al.* Methylation profiles of genes utilizing newly developed CpG island methylation microarray on colorectal cancer patients. Nucleic Acids Res 2005;33(5):e46.
61. Mund C, Beier V, Bewerunge P, Dahms M, Lyko F, Hoheisel JD. Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16INK4A promoter in colon carcinoma cell lines. Nucleic Acids Res 2005;33(8):e73.
62. Shi H, Maier S, Nimmrich I, *et al.* Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. J Cell Biochem 2003;88(1):138-43.
63. Bibikova M, Lin Z, Zhou L, *et al.* High-throughput DNA methylation profiling using universal bead arrays. Genome Res 2006;16(3):383-93.
64. Brunner AL, Johnson DS, Kim SW, *et al.* Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res 2009;19(6):1044-56.
65. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001;294(5543):853-8.

66. Kloosterman WP, Plasterk RH. The diverse functions of microRNAs in animal development and disease. Dev Cell 2006;11(4):441-50.
67. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136(2):215-33.
68. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100(1):57-70.
69. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006;6(11):857-66.
70. Garzon R, Fabbri M, Cimmino A, Calin GA, Croce CM. MicroRNA expression and function in cancer. Trends Mol Med 2006;12(12):580-7.
71. Ventura A, Jacks T. MicroRNAs and cancer: short RNAs go a long way. Cell 2009;136(4):586-91.
72. Zhang L, Huang J, Yang N, *et al.* microRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci U S A 2006;103(24):9136-41.
73. Brueckner B, Stresemann C, Kuner R, *et al.* The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Res 2007;67(4):1419-23.
74. Deng S, Calin GA, Croce CM, Coukos G, Zhang L. Mechanisms of microRNA deregulation in human cancer. Cell Cycle 2008;7(17):2643-6.
75. Saito Y, Liang G, Egger G, *et al.* Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell 2006;9(6):435-43.
76. Bueno MJ, Perez de Castro I, Gomez de Cedron M, *et al.* Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. Cancer Cell 2008;13(6):496-506.
77. Weber B, Stresemann C, Brueckner B, Lyko F. Methylation of human microRNA genes in normal and neoplastic cells. Cell Cycle 2007;6(9):1001-5.
78. Lujambio A, Calin GA, Villanueva A, *et al.* A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A 2008;105(36):13556-61.
79. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer 2000;83(10):1301-8.
80. Domchek SM, Weber BL. Clinical management of BRCA1 and BRCA2 mutation carriers. Oncogene 2006;25(43):5825-31.
81. Sjoblom T, Jones S, Wood LD, *et al.* The consensus coding sequences of human breast and colorectal cancers. Science 2006;314(5797):268-74.
82. Szyf M, Pakneshan P, Rabbani SA. DNA methylation and breast cancer. Biochem Pharmacol 2004;68(6):1187-97.
83. Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. Oncogene 2002;21(35):5462-82.
84. Evron E, Dooley WC, Umbricht CB, *et al.* Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. Lancet 2001;357(9265):1335-6.
85. Krassenstein R, Sauter E, Dulaimi E, *et al.* Detection of breast cancer in nipple aspirate fluid by CpG island hypermethylation. Clin Cancer Res 2004;10(1 Pt 1):28-32.
86. Iorio MV, Ferracin M, Liu CG, *et al.* MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65(16):7065-70.
87. Spizzo R, Nicoloso MS, Croce CM, Calin GA. SnapShot: MicroRNAs in Cancer. Cell 2009;137(3):586- e1.
88. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, Lund AH. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. J Biol Chem 2008;283(2):1026-33.

89. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J Biol Chem 2007;282(19):14328-36.
90. Mertens-Talcott SU, Chintharlapalli S, Li X, Safe S. 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 2007;67(22):11001-11.
91. Scott GK, Goga A, Bhaumik D, Berger CE, Sullivan CS, Benz CC. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. J Biol Chem 2007;282(2):1479-86.
92. Adams BD, Furneaux H, White BA. The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. Mol Endocrinol 2007;21(5):1132-47.
93. Zhao JJ, Lin J, Yang H, *et al.* MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. J Biol Chem 2008;283(45):31079-86.
94. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 2007;449(7163):682-8.
95. Valastyan S, Reinhardt F, Benaich N, *et al.* A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. Cell 2009;137(6):1032-46.
96. Ferguson AT, Evron E, Umbricht CB, *et al.* High frequency of hypermethylation at the 14-3-3 sigma locus leads to gene silencing in breast cancer. Proc Natl Acad Sci U S A 2000;97(11):6049-54.
97. Veeck J, Noetzel E, Bektas N, *et al.* Promoter hypermethylation of the SFRP2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. Mol Cancer 2008;7:83.
98. Jin Z, Tamura G, Tsuchiya T, *et al.* Adenomatous polyposis coli (APC) gene promoter hypermethylation in primary breast cancers. Br J Cancer 2001;85(1):69-73.
99. Lee AY, He B, You L, *et al.* Expression of the secreted frizzled-related protein gene family is downregulated in human mesothelioma. Oncogene 2004;23(39):6672-6.
100. Urakami S, Shiina H, Enokida H, *et al.* Combination analysis of hypermethylated Wnt-antagonist family genes as a novel epigenetic biomarker panel for bladder cancer detection. Clin Cancer Res 2006;12(7 Pt 1):2109-16.
101. Gery S, Tanosaki S, Bose S, Bose N, Vadgama J, Koeffler HP. Down-regulation and growth inhibitory role of C/EBPalpha in breast cancer. Clin Cancer Res 2005;11(9):3184-90.
102. Suzuki H, Toyota M, Carraway H, *et al.* Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer. Br J Cancer 2008;98(6):1147-56.
103. Evron E, Umbricht CB, Korz D, *et al.* Loss of cyclin D2 expression in the majority of breast cancers is associated with promoter hypermethylation. Cancer Res 2001;61(6):2782-7.
104. Ai L, Tao Q, Zhong S, *et al.* Inactivation of Wnt inhibitory factor-1 (WIF1) expression by epigenetic silencing is a common event in breast cancer. Carcinogenesis 2006;27(7):1341-8.
105. Schagdarsurengin U, Pfeifer GP, Dammann R. Frequent epigenetic inactivation of cystatin M in breast carcinoma. Oncogene 2007;26(21):3089-94.
106. Roman-Gomez J, Cordeu L, Agirre X, *et al.* Epigenetic regulation of Wnt-signaling pathway in acute lymphoblastic leukemia. Blood 2007;109(8):3462-9.
107. Graff JR, Gabrielson E, Fujii H, Baylin SB, Herman JG. Methylation patterns of the E-cadherin 5' CpG island are unstable and reflect the dynamic, heterogeneous loss of E-cadherin expression during metastatic progression. J Biol Chem 2000;275(4):2727-32.

108. Shulewitz M, Soloviev I, Wu T, Koeppen H, Polakis P, Sakanaka C. Repressor roles for TCF-4 and Sfrp1 in Wnt signaling in breast cancer. Oncogene 2006;25(31):4361-9.
109. Toyooka KO, Toyooka S, Virmani AK, *et al.* Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. Cancer Res 2001;61(11):4556-60.
110. Hoque MO, Kim MS, Ostrow KL, *et al.* Genome-wide promoter analysis uncovers portions of the cancer methylome. Cancer Res 2008;68(8):2661-70.
111. Keen JC, Garrett-Mayer E, Pettit C, *et al.* Epigenetic regulation of protein phosphatase 2A (PP2A), lymphotactin (XCL1) and estrogen receptor alpha (ER) expression in human breast cancer cells. Cancer Biol Ther 2004;3(12):1304-12.
112. Shigematsu H, Suzuki M, Takahashi T, *et al.* Aberrant methylation of HIN-1 (high in normal-1) is a frequent event in many human malignancies. Int J Cancer 2005;113(4):600-4.
113. Murata H, Khattar NH, Kang Y, Gu L, Li GM. Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. Oncogene 2002;21(37):5696-703.
114. Raman V, Martensen SA, Reisman D, *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. Nature 2000;405(6789):974-8.
115. Sutherland KD, Lindeman GJ, Choong DY, *et al.* Differential hypermethylation of SOCS genes in ovarian and breast carcinomas. Oncogene 2004;23(46):7726-33.
116. Wolf I, O'Kelly J, Rubinek T, *et al.* 15-hydroxyprostaglandin dehydrogenase is a tumor suppressor of human breast cancer. Cancer Res 2006;66(15):7818-23.
117. Garcia JM, Silva J, Pena C, *et al.* Promoter methylation of the PTEN gene is a common molecular change in breast cancer. Genes Chromosomes Cancer 2004;41(2):117-24.
118. Jhaveri MS, Morrow CS. Methylation-mediated regulation of the glutathione S-transferase P1 gene in human breast cancer cells. Gene 1998;210(1):1-7.
119. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. Cancer Res 2001;61(8):3225-9.
120. Kang JH, Kim SJ, Noh DY, *et al.* Methylation in the p53 promoter is a supplementary route to breast carcinogenesis: correlation between CpG methylation in the p53 promoter and the mutation of the p53 gene in the progression from ductal carcinoma in situ to invasive ductal carcinoma. Lab Invest 2001;81(4):573-9.
121. Fujii H, Biel MA, Zhou W, Weitzman SA, Baylin SB, Gabrielson E. Methylation of the HIC-1 candidate tumor suppressor gene in human breast cancer. Oncogene 1998;16(16):2159-64.
122. Leu YW, Yan PS, Fan M, *et al.* Loss of estrogen receptor signaling triggers epigenetic silencing of downstream targets in breast cancer. Cancer Res 2004;64(22):8184-92.
123. Ordway JM, Budiman MA, Korshunova Y, *et al.* Identification of novel high-frequency DNA methylation changes in breast cancer. PLoS One 2007;2(12):e1314.
124. Widschwendter M, Berger J, Hermann M, *et al.* Methylation and silencing of the retinoic acid receptor-beta2 gene in breast cancer. J Natl Cancer Inst 2000;92(10):826-32.
125. Yuan J, Luo RZ, Fujii S, *et al.* Aberrant methylation and silencing of ARHI, an imprinted tumor suppressor gene in which the function is lost in breast cancers. Cancer Res 2003;63(14):4174-80.
126. Dammann R, Yang G, Pfeifer GP. Hypermethylation of the cpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers. Cancer Res 2001;61(7):3105-9.

127. Shinozaki M, Hoon DS, Giuliano AE, *et al.* Distinct hypermethylation profile of primary breast cancer is associated with sentinel lymph node metastasis. Clin Cancer Res 2005;11(6):2156-62.
128. Veeck J, Niederacher D, An H, *et al.* Aberrant methylation of the Wnt antagonist SFRP1 in breast cancer is associated with unfavourable prognosis. Oncogene 2006;25(24):3479-88.
129. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 2002;99(6):3740-5.
130. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics 2002;18(11):1427-31.
131. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132:365-86.
132. Nur I, Szyf M, Razin A, Glaser G, Rottem S, Razin S. Procaryotic and eucaryotic traits of DNA methylation in spiroplasmas (mycoplasmas). J Bacteriol 1985;164(1):19-24.
133. Beier M, Hoheisel JD. Production by quantitative photolithographic synthesis of individually quality checked DNA microarrays. Nucleic Acids Res 2000;28(4):E11.
134. Baum M, Bielau S, Rittner N, *et al.* Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. Nucleic Acids Res 2003;31(23):e151.
135. Fellenberg K, Hauser NC, Brors B, Hoheisel JD, Vingron M. Microarray data warehouse allowing for inclusion of experiment annotations in statistical analysis. Bioinformatics 2002;18(3):423-33.
136. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.
137. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 1995;57:289-300.
138. Balaguer P, Boussioux AM, Demirpence E, Nicolas JC. Reporter cell lines are useful tools for monitoring biological activity of nuclear receptor ligands. Luminescence 2001;16(2):153-8.
139. Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA 2007;13(2):223-39.
140. Eberwine J, Yeh H, Miyashiro K, *et al.* Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A 1992;89(7):3010-4.
141. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008;36(Database issue):D154-8.
142. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002;18 Suppl 1:S96-104.
143. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98(9):5116-21.
144. Werner T. Computer-assisted analysis of transcription control regions. Matinspector and other programs. Methods Mol Biol 2000;132:337-49.
145. Avnit-Sagi T, Kantorovich L, Kredo-Russo S, Hornstein E, Walker MD. The promoter of the pri-miR-375 gene directs expression selectively to the endocrine pancreas. PLoS One 2009;4(4):e5033.
146. Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell 2009;137(7):1194-211.
147. Filippova GN. Genetics and epigenetics of the multifunctional protein CTCF. Curr Top Dev Biol 2008;80:337-60.

148. Docquier F, Farrar D, D'Arcy V, *et al.* Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. Cancer Res 2005;65(12):5112-22.
149. Eger A, Aigner K, Sonderegger S, *et al.* DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene 2005;24(14):2375-85.
150. Spaderna S, Schmalhofer O, Wahlbuhl M, *et al.* The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer Res 2008;68(2):537-44.
151. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007;27(1):91-105.
152. Vaidyanathan G, Cismowski MJ, Wang G, Vincent TS, Brown KD, Lanier SM. The Ras-related protein AGS1/RASD1 suppresses cell growth. Oncogene 2004;23(34):5858-63.
153. Brogan MD, Behrend EN, Kemppainen RJ. Regulation of Dexras1 expression by endogenous steroids. Neuroendocrinology 2001;74(4):244-50.
154. Reid G, Hubner MR, Metivier R, *et al.* Cyclic, proteasome-mediated turnover of unliganded and liganded ERalpha on responsive promoters is an integral feature of estrogen signaling. Mol Cell 2003;11(3):695-707.
155. Plass C, Shibata H, Kalcheva I, *et al.* Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat Genet 1996;14(1):106-9.
156. Yan PS, Shi H, Rahmatpanah F, *et al.* Differential distribution of DNA methylation within the RASSF1A CpG island in breast cancer. Cancer Res 2003;63(19):6178-86.
157. Maegawa S, Hinkal G, Kim HS, *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 2010;20(3):332-40.
158. Teschendorff AE, Menon U, Gentry-Maharaj A, *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 2010;20(4):440-6.
159. Lee TI, Jenner RG, Boyer LA, *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 2006;125(2):301-13.
160. Ohm JE, McGarvey KM, Yu X, *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 2007;39(2):237-42.
161. Widschwendter M, Fiegl H, Egle D, *et al.* Epigenetic stem cell signature in cancer. Nat Genet 2007;39(2):157-8.
162. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev 2009;23(22):2563-77.
163. Jones SE, Jomary C. Secreted Frizzled-related proteins: searching for relationships and patterns. Bioessays 2002;24(9):811-20.
164. Hsieh JC, Kodjabachian L, Rebbert ML, *et al.* A new secreted protein that binds to Wnt proteins and inhibits their activities. Nature 1999;398(6726):431-6.
165. Cheyette BN, Waxman JS, Miller JR, *et al.* Dapper, a Dishevelled-associated antagonist of beta-catenin and JNK signaling, is required for notochord formation. Dev Cell 2002;2(4):449-61.
166. Sotiropoulou G, Anisowicz A, Sager R. Identification, cloning, and characterization of cystatin M, a novel cysteine proteinase inhibitor, down-regulated in breast cancer. J Biol Chem 1997;272(2):903-10.
167. Kioulafa M, Balkouranidou I, Sotiropoulou G, *et al.* Methylation of cystatin M promoter is associated with unfavorable prognosis in operable breast cancer. Int J Cancer 2009;125(12):2887-92.
168. McKnight SL. McBindall--a better name for CCAAT/enhancer binding proteins? Cell 2001;107(3):259-61.

169. Wales MM, Biel MA, el Deiry W, *et al.* p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3. Nat Med 1995;1(6):570-7.
170. Zhang W, Zeng X, Briggs KJ, *et al.* A potential tumor suppressor role for Hic1 in breast cancer through transcriptional repression of ephrin-A1. Oncogene.
171. Manser E, Huang HY, Loo TH, *et al.* Expression of constitutively active alpha-PAK reveals effects of the kinase on actin and focal complexes. Mol Cell Biol 1997;17(3):1129-43.
172. Kumar R, Gururaj AE, Barnes CJ. p21-activated kinases in cancer. Nat Rev Cancer 2006;6(6):459-71.
173. Jiang BH, Liu LZ. PI3K/PTEN signaling in angiogenesis and tumorigenesis. Adv Cancer Res 2009;102:19-65.
174. Muggerud AA, Ronneberg JA, Warnberg F, *et al.* Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Res 2010;12(1):R3.
175. Hesson LB, Cooper WN, Latif F. The role of RASSF1A methylation in cancer. Dis Markers 2007;23(1-2):73-87.
176. Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res 1998;58(20):4515-8.
177. McKee KK, Tan CP, Palyha OC, *et al.* Cloning and characterization of two human G protein-coupled receptor genes (GPR38 and GPR39) related to the growth hormone secretagogue and neurotensin receptors. Genomics 1997;46(3):426-34.
178. Nikolopoulos D, Theocharis S, Kouraklis G. Ghrelin's role on gastrointestinal tract cancer. Surg Oncol 2010;19(1):e2-e10.
179. Issa JP. CpG island methylator phenotype in cancer. Nat Rev Cancer 2004;4(12):988-93.
180. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 1999;96(15):8681-6.
181. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 2006;38(7):787-93.
182. Sunami E, Shinozaki M, Sim MS, *et al.* Estrogen receptor and HER2/neu status affect epigenetic differences of tumor-related genes in primary breast tumors. Breast Cancer Res 2008;10(3):R46.
183. Valencia A, Roman-Gomez J, Cervera J, *et al.* Wnt signaling pathway is epigenetically regulated by methylation of Wnt antagonists in acute myeloid leukemia. Leukemia 2009;23(9):1658-66.
184. Kongkham PN, Northcott PA, Croul SE, Smith CA, Taylor MD, Rutka JT. The SFRP family of WNT inhibitors function as novel tumor suppressor genes epigenetically silenced in medulloblastoma. Oncogene 2010.
185. Parrella P, Poeta ML, Gallo AP, *et al.* Nonrandom distribution of aberrant promoter methylation of cancer-related genes in sporadic breast tumors. Clin Cancer Res 2004;10(16):5349-54.
186. Riazalhosseini Y, Hoheisel J. Do we use the appropriate controls for the identification of informative methylation markers for early cancer detection? Genome Biology 2008;9(11):405-.
187. Yan PS, Venkataramu C, Ibrahim A, *et al.* Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue. Clin Cancer Res 2006;12(22):6626-36.
188. Huang TH, Perry MR, Laux DE. Methylation profiling of CpG islands in human breast cancer cells. Hum Mol Genet 1999;8(3):459-70.

189. Euhus DM, Bu D, Milchgrub S, *et al.* DNA methylation in benign breast epithelium in relation to age and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2008;17(5):1051-9.
190. Lewis CM, Cler LR, Bu DW, *et al.* Promoter hypermethylation in benign breast epithelium in relation to predicted breast cancer risk. Clin Cancer Res 2005;11(1):166-72.
191. Bean GR, Bryson AD, Pilie PG, *et al.* Morphologically normal-appearing mammary epithelial cells obtained from high-risk women exhibit methylation silencing of INK4a/ARF. Clin Cancer Res 2007;13(22 Pt 1):6834-41.
192. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 2006;6(2):107-16.
193. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nat Rev Genet 2006;7(1):21-33.
194. Giovannucci E, Ogino S. DNA methylation, field effects, and colorectal cancer. J Natl Cancer Inst 2005;97(18):1317-9.
195. Shen L, Kondo Y, Rosner GL, *et al.* MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst 2005;97(18):1330-8.
196. Liang G, Wolff EM, Chihara Y, *et al.* DNA methylation profile of 1501 loci in bladder cancer shows evidence of a field defect. The 99th Annual meeting of the American Association for Cancer Research; 2008 12-16 April; San Diego; 2008.
197. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature 2005;434(7035):843-50.
198. Wong DJ, Foster SA, Galloway DA, Reid BJ. Progressive region-specific de novo methylation of the p16 CpG island in primary human mammary epithelial cell strains during escape from M(0) growth arrest. Mol Cell Biol 1999;19(8):5642-51.
199. Zhang Y, Rohde C, Tierling S, *et al.* DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet 2009;5(3):e1000438.
200. Renaud S, Loukinov D, Abdullaev Z, *et al.* Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene. Nucleic Acids Res 2007;35(4):1245-56.
201. Renda M, Baglivo I, Burgess-Beusse B, *et al.* Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 2007;282(46):33336-45.
202. Dahl E, Koseki H, Balling R. Pax genes and organogenesis. Bioessays 1997;19(9):755-65.
203. Muratovska A, Zhou C, He S, Goodyer P, Eccles MR. Paired-Box genes are frequently expressed in cancer and often required for cancer cell survival. Oncogene 2003;22(39):7989-97.
204. Vouyovitch CM, Vidal L, Borges S, *et al.* Proteomic analysis of autocrine/paracrine effects of human growth hormone in human mammary carcinoma cells. Adv Exp Med Biol 2008;617:493-500.
205. Volinia S, Calin GA, Liu CG, *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103(7):2257-61.
206. Blenkiron C, Goldstein LD, Thorne NP, *et al.* MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol 2007;8(10):R214.
207. Sempere LF, Christensen M, Silahtaroglu A, *et al.* Altered MicroRNA expression confined to specific epithelial cell subpopulations in breast cancer. Cancer Res 2007;67(24):11612-20.

208. Pandey DP, Picard D. miR-22 inhibits estrogen signaling by directly targeting the estrogen receptor alpha mRNA. Mol Cell Biol 2009;29(13):3783-90.
209. Liu WH, Yeh SH, Lu CC, *et al.* MicroRNA-18a prevents estrogen receptor-alpha expression, promoting proliferation of hepatocellular carcinoma cells. Gastroenterology 2009;136(2):683-93.
210. Hossain A, Kuo MT, Saunders GF. Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. Mol Cell Biol 2006;26(21):8191-201.
211. Tsukamoto Y, Nakada C, Noguchi T, *et al.* MicroRNA-375 is downregulated in gastric carcinomas and regulates cell survival by targeting PDK1 and 14-3-3zeta. Cancer Res;70(6):2339-49.
212. Poy MN, Hausser J, Trajkovski M, *et al.* miR-375 maintains normal pancreatic alpha- and beta-cell mass. Proc Natl Acad Sci U S A 2009;106(14):5813-8.
213. Mathe EA, Nguyen GH, Bowman ED, *et al.* MicroRNA expression in squamous cell carcinoma and adenocarcinoma of the esophagus: associations with survival. Clin Cancer Res 2009;15(19):6192-200.
214. Chan CS, Song JS. CCCTC-binding factor confines the distal action of estrogen receptor. Cancer Res 2008;68(21):9041-9.
215. Nguyen P, Cui H, Bisht KS, *et al.* CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. Cancer Res 2008;68(14):5546-51.
216. D'Arcy V, Pore N, Docquier F, *et al.* BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. Br J Cancer 2008;98(3):571-9.
217. Cismowski MJ, Ma C, Ribas C, *et al.* Activation of heterotrimeric G-protein signaling by a ras-related protein. Implications for signal integration. J Biol Chem 2000;275(31):23421-4.
218. Graham TE, Qiao Z, Dorin RI. Dexras1 inhibits adenylyl cyclase. Biochem Biophys Res Commun 2004;316(2):307-12.
219. Nguyen CH, Watts VJ. Dexamethasone-induced Ras protein 1 negatively regulates protein kinase C delta: implications for adenylyl cyclase 2 signaling. Mol Pharmacol 2006;69(5):1763-71.
220. Shanmugam M, Krett NL, Maizels ET, *et al.* Regulation of protein kinase C delta by estrogen in the MCF-7 human breast cancer cell line. Mol Cell Endocrinol 1999;148(1-2):109-18.
221. McCracken MA, Miraglia LJ, McKay RA, Strobl JS. Protein kinase C delta is a prosurvival factor in human breast tumor cell lines. Mol Cancer Ther 2003;2(3):273-81.
222. De Servi B, Hermani A, Medunjanin S, Mayer D. Impact of PKCdelta on estrogen receptor localization and activity in breast cancer cells. Oncogene 2005;24(31):4946-55.
223. Nabha SM, Glaros S, Hong M, *et al.* Upregulation of PKC-delta contributes to antiestrogen resistance in mammary tumor cells. Oncogene 2005;24(19):3166-76.
224. Bhat-Nakshatri P, Wang G, Collins NR, *et al.* Estradiol-regulated microRNAs control estradiol response in breast cancer cells. Nucleic Acids Res 2009;37(14):4850-61.
225. Castellano L, Giamas G, Jacob J, *et al.* The estrogen receptor-alpha-induced microRNA signature regulates itself and its transcriptional response. Proc Natl Acad Sci U S A 2009;106(37):15732-7.

7 Appendix

7.1 Supplementary Table S1

Supplementary Table S1. Clinical information of the patients and control samples used in this study.

Patient	Age	Diagnosis	Grade	No. of Lymph nodes affected	Normal sample
IR3	45	IDC	III	8	+
IR4	39	IDC	III	N/A	+
IR6	74	IDC	III	N/A	+
IR7	45	IDC	III	11	+
IR9	42	IDC	III	9	+
IR12	49	IDC	III	3	+
IR13	37	ILC	III	3	+
IR14	53	IDC	III	14	+
IR-27	51	IDC	III	4	+
IR5	45	IDC	III	3	-
IR17	38	IDC	II	10	-
IR18	50	IDC	III	1	-
IR19	67	IDC	N/A	0	-
IR20	60	IDC	I	1	-
IR23	40	IDC	III	5	-
IR24	59	ILC	I	1	-
IR-27	51	IDC	III	4	-
IR-28	35	IDC	I	N/A	-
IR-30	61	IDC	I	18	-
IR-31	53	IDC	I	1	-
IR-32	45	IDC	I	3	-
IR-34	64	IDC	III	0	-
IR-35	45	IDC	II	4	-
IR-36	48	IDC	III	1	-
IR-37	60	IDC	II	1	-
IR-38	47	IDC	II	0	-

IR-43	33	IDC	N/A	N/A	-
IR-45	35	IDC	N/A	N/A	-
IR-46	56	IDC	II	2	-
IR-47	43	IDC	III	10	-
IR-48	56	IDC	I	0	-
IR-49	35	IDC	I	0	-
IR-50	75	IDC	II	0	-
UKR48	36	Adenofibroma	-	-	-
UKR49	45	Adenofibroma	-	-	-
UKR 50	35	Adenofibroma	-	-	-
UKR 51	42	Adenofibroma	-	-	-
UKR 52	38	Adenofibroma	-	-	-
UKR 53	48	Adenofibroma	-	-	-
UKR 54	43	Adenofibroma	-	-	-
UKR 55	52	Adenofibroma	-	-	-
UKR 58	58	Adenofibroma	-	-	-
UKR 59	28	Adenofibroma	-	-	-
IR-105	29	Healthy	-	-	-
IR-106	31	Healthy	-	-	-
UKR5	31	Healthy	-	-	-
UKR9	22	Healthy	-	-	-
UKR10	27	Healthy	-	-	-
UKR13	37	Healthy	-	-	-
UKR14	40	Healthy	-	-	-

7.2 Supplementary Table S2

The table including the methylation profiling data of all samples can be accessed online via: http://www.dkfz.de/funct_genome/Riazalhosseini-thesis-data.html

7.3 Supplementary Table S3

The table including miRNA profiling data of the mammary cell lines can be accessed online via: http://www.dkfz.de/funct_genome/Riazalhosseini-thesis-data.html

7.4 Supplementary Table S4

Supplementary Table S4. Clinical information of breast cancer specimens used for miR-375 quantification and target identification. IDC: Invasive ductal carcinoma; AMC: Atypical medullary carcinoma; FD: Fibrocystic disease; P: positive; N: Negative; N/A: Not applicable

#	Age	Histological subtype	ERα	Grade	Lymph node metastasis
1	45	IDC	P	III	Yes
2	49	IDC	N	III	Yes
3	51	IDC	P	II	Yes
4	46	AMC	N	III	Yes
5	74	IDC	P	III	Yes
6	45	IDC	N	III	Yes
7	63	FD	P	N/A	N/A
8	42	IDC	P	III	Yes
9	53	IDC	P	III	Yes

7.5 Supplementary Table S5

Supplementary Table S5. Transcripts commonly dowregulated in the tumors with overexpression of miR-375.

Symbol	Probe ID	NCBI Reference Sequence	Mean Fold-change in tumor relative to normal
HBB	5340674	NM_000518	-33.87009985
HBA1	360554	NM_000558	-33.82576067
ADH1A	6020441	NM_000667	-25.93859944
PI16	7560070	NM_153370	-22.17755738
CLEC3B	4880187	NM_003278	-19.08968666
GPD1	6290452	NM_005276	-15.06624523
CFD	5720243	NM_001928	-14.8521602
DCN	6650196	NM_003467	-11.20582049

GPX3	840050	NM_002084	-11.10483516
FHL1	2320475	NM_001449	-10.57238041
FABP4	3440477	NM_001442	-10.17169134
GYG2	5310521	NM_001079855	-9.615858196
IGFBP6	2230242	NM_002178	-9.613800418
CD36	4260433	NM_001001548	-9.567190595
MAOA	6650278	NM_000240	-8.784490748
DCN	3400682	NM_133503	-8.589691606
GPC3	3130739	NM_004484	-8.545181193
CD209	1740300	NM_021155	-8.425745046
ABCA8	580689	NM_007168	-8.234910062
CA4	3990296	NM_000717	-8.128796321
CALB2	6130364	NM_001740	-8.108158855
CD36	3990167	NM_000072	-7.539840991
CXCL12	2350482	NM_001033886	-7.397941133
CIDEC	5490736	NM_022094	-7.252977302
DPT	1440474	NM_001937	-7.156513683
PDK4	380400	NM_002612	-7.143417879
ADM	6650020	NM_001124	-7.047800702
HSPB6	5310719	NM_144617	-6.855199543
RDH5	2810609	NM_002905	-6.732952428
PLAC9	5420184	NM_001012973	-6.636228744
CCL8	7510360	NM_005623	-6.486231866
MAMDC2	5260762	NM_153267	-6.267896618
ABCA6	3520370	NM_080284	-6.101352302
PLTP	3130154	NM_182676	-6.051027217
CD163	5570414	NM_203416	-5.94335589
GPAM	6590301	NM_020918	-5.925120424
CDO1	730202	NM_001801	-5.846683556
ALDH1A1	50100	NM_000689	-5.794897295
CXCL12	4670202	NM_199168	-5.744843023
SEPP1	6760343	NM_005410	-5.674047839
ABCA9	5090370	NM_080283	-5.643664088
F13A1	1300095	NM_000129	-5.549224755
RNASE1	540670	NM_198232	-5.53853568

CXCR7	7050682	NM_020311	-5.480711243
ACACB	3180139	NM_001093	-5.440757069
MGST1	3370072	NM_020300	-5.358188051
LOC221091	5720446	XM_001130636	-5.327153756
PTGIS	2900722	NM_000961	-5.325847786
PMP22	3890113	NM_153321	-5.311701041
RASD1	6180692	NM_016084	-5.153772132
NMB	770437	NM_205858	-5.153296035
CYP26B1	2510619	NM_019885	-5.071875192
AGPAT2	6860039	NM_006412	-5.048047974
SRPX	3400703	NM_006307	-5.044800086
VSIG4	4200008	NM_007268	-5.032407928
CTSG	830619	NM_001911	-4.900024243
GSN	1510379	NM_198252	-4.89190932
CAV2	2030195	NM_001233	-4.691028693
C10orf116	6020017	NM_006829	-4.578504376
CDKN2C	5690167	NM_078626	-4.401851127
AOX1	4250291	NM_001159	-4.347345237
C1R	4220364	NM_001733	-4.23608185
PPARG	7160711	NM_138711	-4.181030631
MMD	360671	NM_012329	-4.114168911
MFAP4	2710167	NM_002404	-4.087520182
GFPT2	5420053	NM_005110	-3.975438749
EFEMP1	5820689	NM_001039349	-3.971058827
ANGPTL4	2630609	NM_139314	-3.964492939
PMP22	5340630	NM_000304	-3.947155391
SORBS1	4150343	NM_001034956	-3.830232343
C1S	1260255	NM_201442	-3.749741705
MOSC1	1980082	NM_022746	-3.744157261
CD14	990398	NM_000591	-3.738037941
FLJ21986	4730064	NM_024913	-3.677525416
ATOH8	130068	NM_032827	-3.648054215
EMILIN2	10070	NM_032048	-3.586175202
EPDR1	5560255	NM_017549	-3.529878588
GYPC	1580538	NM_002101	-3.528260833

RNF150	2340341	NM_020724	-3.476769089
PNPLA2	5550168	NM_020376	-3.451442902
PFKFB3	1710273	NM_004566	-3.372097251
MYOM1	2370477	NM_003803	-3.309374056
LAMA2	5390373	NM_001079823	-3.293220836
MT1M	5870398	NM_176870	-3.288630437
FXYD1	3060767	NM_005031	-3.270451783
C13orf33	7650468	NM_032849	-3.267193935
CLDND2	10551	NM_152353	-3.256551554
CD248	2350292	NM_020404	-3.240887585
PPP2R1B	3930196	NM_181699	-3.223894069
FBLN2	2120243	NM_001004019	-3.161848086
GAS7	6860193	NM_201433	-3.141857829
MGST1	3940575	NM_145791	-3.12341643
MSX1	650546	NM_002448	-3.122182544
CCDC3	4810692	NM_031455	-3.106450045
MAPK10	240451	NM_138981	-3.09583022
SMPDL3A	2070411	NM_006714	-3.066785896
ZBTB16	6040296	NM_006006	-2.979691267
NID1	670739	NM_002508	-2.928778345
FCGRT	4540041	NM_004107	-2.89735884
TYRO3	2450403	NM_006293	-2.889187399
CGI-38	1400722	NM_015964	-2.874736608
EBF3	4280482	NM_001005463	-2.852552376
GSTM5	5570221	NM_000851	-2.840638135
HSPB2	1580239	NM_001541	-2.804975762
CDKN1C	3610601	NM_000076	-2.804949084
TUBB6	1940156	NM_032525	-2.797839665
HSPB8	5290215	NM_014365	-2.692440056
MAF	6400369	NM_005360	-2.668694992
CYP4F12	4390273	NM_023944	-2.603727871
AVPI1	6840722	NM_021732	-2.568960106
ANTXR2	240468	NM_058172	-2.538804922
GAS1	2370762	NM_002048	-2.522079771
FMOD	3940113	NM_002023	-2.507249207

PDE2A	5260102	NM_002599	-2.493511866
PEMT	2450102	NM_148173	-2.444707749
BIN1	5890333	NM_139348	-2.441973248
GLDN	1570615	NM_181789	-2.430553479
HRH1	5310286	NM_000861	-2.377187947
TUBB6	1410059	NM_032525	-2.298365414
CYP4B1	3890767	NM_000779	-2.280069175
DHRS3	3840170	NM_004753	-2.278422379
S100A13	4010168	NM_001024212	-2.214367175
AIFM2	5690070	NM_032797	-2.148792422
RBPMS2	4230164	NM_194272	-2.012409927
OLFML2A	2480471	NM_182487	-1.944303204

7.6 Publications

Articles based on this thesis:

S. Kumar Botla, M. Malekpour, A. Moghaddas Gholami, E. Moskalev, V. Bubnov, H. Najmabadi, JD. Hoheisel, **Y. Riazalhosseini** (2010) A breast cancer DNA methylation signature associated with field defect (*In preparation*).

P. de Souza Rocha Simonini, A. Breiling, N. Gupta, M. Malekpour, M. Youns, R. Omranipour, F. Malekpour S. Volinia, C. Croce, H. Najmabadi, S. Diederichs, Ö. Sahin, D. Mayer, F. Lyko, JD. Hoheisel, **Y. Riazalhosseini**. (2010) Epigenetically de-regulated microRNA-375 is involved in a positive feedback loop with Estrogen Receptor α in breast cancer cells. (*Submitted*).

Y. Riazalhosseini and JD. Hoheisel (2008). Do we use the appropriate controls for the identification of informative methylation markers for early cancer detection? *Genome Biol; 9 (11): 405.*