universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

DNA methylation analysis for minimal-invasive Breast cancer diagnostic testing

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Matthias Wielscher
Studienrichtung /Studienzweig (lt. Studienblatt):	A 437 Biologie A 441 Mikrobiologie / Genetik
Betreuerin / Betreuer:	Univ. Prof. Dr. Pavel Kovarik

Wien, im Juni 2010

Acknowledgments

I want to thank my family for supporting me in all stages of my efforts to reach an academic degree. Thanks to my parents, grandparents, Franzi, Markus, Maria and Elisabeth. I want to thank my parents for their doubtless belief in me and for pushing me forward. I want to thank Elli for taking care of my son so often, while I was writing this thesis.

Especially I want to thank Karin for her help at this work, her spiritual succour at the writing phase of my diploma theses and her patience at listening to my complains about science. Thank you!

I want to thank all people, who worked with me at the AIT. I want to thank Christa for making things possible, - my lab mates for making my working days easier and more delightful, - my office colleague for nice coffee breaks.
Special thanks are dedicated to Andi, who was a good teacher**,** mentor and boss. Thank you Andi for taking you so much time and always having an open ear for my thoughts, ideas and worries. I want to thank you for many fruitful discussions and a very amicable company

Finally I want to thank my son Jakob for being there.

TABLE OF CONTENTS:

Zusammenfassung

Veränderungen im Methylierungsmuster der DNA geschehen früh in der Entwicklung von Tumoren. Zumeist sind Promoter-Regionen von Tumorsupressor-Genen und Onko-Genen betroffen, die einen Wachstumsvorteil der Krebszellen zur Folge haben. In dieser Studie wurde die Möglichkeit der Detektion dieser epigenetischen Aberrationen in zellfreier DNA aus Serum von Brustkrebspatienten untersucht, da diese potenzielle Biomarker zur minimal-invasiven Diagnostik darstellen.

Hierzu wurden unterschiedliche DNA–Isolationsverfahren und Amplifikationsmethoden zur Anreicherung methylierter DNA getestet. Mit den optimierten Methoden wurde Serum-DNA von Brustkrebspatienten mit A) malignen Neoplasmen (n=40), B) nicht invasiv wachsenden Tumoren (n=18) und C) gesunden Kontrollprobandinnen (n=24) isoliert.

Der DNA Gehalt im Serum von Brustkrebspatienten mit metastasierenden Tumoren [45,64ng/ml ± 32,31ng] war gegenüber gesunden Probanden [10,58ng/ml ± 9,71ng] signifikant erhöht ($p = 0,0013$; Wilcoxon-Rangsummentest).

Die methylierte DNA Fraktion der Proben wurde mit einem Restriktionsenzymen – basierten Verfahren und einer „Rolling Circle Amplifikation“ (RCA) genomweit angereichert. Nach erfolgreicher Prozessierung von 72 der 82 DNA Proben konnte ein Biomarker-Screening zur Unterscheidung der Patientengruppen durchgeführt werden. Dazu wurden 360 ausgewählte DNA Bereiche in Multiplex-PCR Reaktionen amplifiziert und die PCR-Produkte durch Hybridisierung auf einem „targeted-micro-array“ detektiert. Die Micro-Array Analyse erlaubte eine Unterscheidung zwischen den Brustkrebspatienten und den Normal Kontrollen, wobei jedoch nach kritischer Betrachtung ein experimenteller Bias welcher zu diesen Unterschieden beiträgt nicht ausgeschlossen werden kann.

In dieser Arbeit konnte somit ein Protokoll erarbeitet werden, das ein genomweites Biomarker-Screening zur Auffindung von DNA-Methylierungsmarkern für die minimal invasive Diagnostik aus geringsten Mengen zellfreier Serum DNA ermöglicht.

Abstract:

Changes in DNA-methylation patterns are an early event in cancer development. Many promoter regions of tumor suppressor genes or oncogenes change their methylation status due to growth advantages of the cancer cell. During this study we investigated detection of these epigenetic aberrations in cell free DNA derived from serum of breast cancer patients, for elucidation of potential biomarkers for minimal invasive diagnostics.

Therefore several DNA isolation approaches and genome wide amplification methods for enrichment of methylated DNA were tested. Applying these optimized methods serum-DNA was isolated form sera of A) breast cancer patients with malign neoplasm (n=40), B) sera of patients with a non-invasive cancer phenotype (n= 18) and C) healthy normal controls (n=24). The DNA concentrations in serum of patients with a metastasizing tumor (45,64 ng/ml +/- 32,31ng) was significantly increased when compared to serum-concentrations of normal controls (10,58ng/ml +/- 9,71ng), ($p = 0.0013$, Wilcoxon-test).

The methylated fraction of the sample DNA was genome wide enriched, using a restriction enzyme based approach combined with a subsequent “Rolling Circle Amplification” (RCA). After successful sample preparation of 72 out of 82 DNA samples, a biomarker screening to distinguish between the patient study groups was performed. Therefore 360 selected DNA sequences were amplified within a multiplex-PCR approach and amplicons detected upon hybridization onto a target micro-array. Micro-array analyses enabled identification of marker-candidates for classification of study groups. However upon critical examination of our experimental setup, we cannot exclude that some experimental bias contributes to this discrimination.

In this work a protocol was established which enables a genome wide biomarker screening for elucidation of DNA methylation markers for minimal invasive diagnostic testing using spurious amounts of cell free serum DNA.

1. Introduction

1.1 Epigenetics and DNA Methylation

1.1.1 Preface

The term “epigenetics” describes the existence of inheritable factors that contribute to the phenotype of an individual, which are not determined in the DNA sequence. Two mechanisms are available to induce a modification of a cytosine base by addition of a methyl residue. The DNA can be methylated *de novo*, or an existing methylation pattern is maintained during DNA replication by the DNA maintenance methylase. In this case the methylation pattern of the parental DNA stand is transferred to the daughter strands during cell division, as the DNA maintenance methyltransferase modifies the daughter strand according to the methylation pattern of the parental DNA (Lewin et al., 2004).

In most cases, DNA methylation goes along with gene silencing. This can occur via the influence of transcription factors or the formation of heterochromatin. Heterochromatin is the transcriptional inactive form of a DNA molecule and in this context another example for epigenetic regulation in eukaryotes. This chromatin form is packed more tightly than euchromatin and the histones in these regions are modified in a different way than the euchromatin. This tight packed organisation of nucleosoms occurs constitutively in centromer or telomere regions of chromosomes and throughout the inactive X-chromosome (Dimitri *et al.*, 2009).

Different acetylation and methylation patterns of the N-terminus of histones are key-regulators of epigenetic regulation of gene-expression. This is named the „histone code“, which includes additional modifications, like phosphorylation or ubiquitination. Acetylation can lead to a more relaxed chromatin form or to the formation of histone remodelling complexes, which both can facilitate transcription.

The concept of epigenetics includes many different procedures of gene-regulation, which are closely related (Figure 1), not only DNA methylation, investigated in this work, but also RNAi as well as histone modification play an important role in the understanding of the epigenetic influence on gene expression (Figure 1) (Egger *et al.*, 2004).

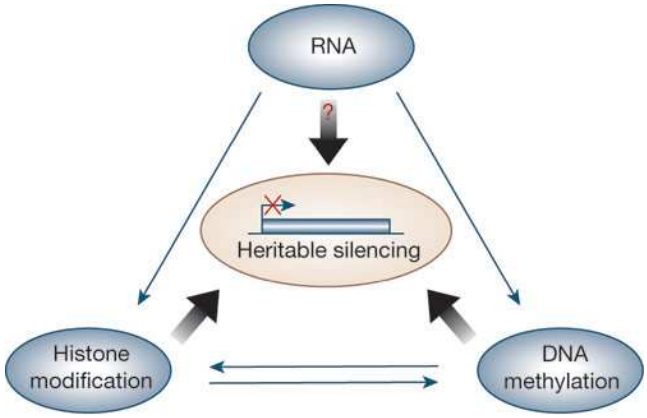


Fig.1: Overview of interaction between epigenetic factors in heritable silencing. Adopted from (Egger *et al.*, 2004)

1.1.2 DNA Methylation

The modification of DNA on specific sites in the genome happens for different reasons. In bacteria DNA methylation is required to distinguish between the own bacterial DNA and foreign DNA. In this case, methylation serves as protection mechanism, as the endonucleases of the bacterium digest alien DNA (Knipers, 2001).

Bacteria require this modification also for additional proof reading during DNA synthesis. For this mechanism it is necessary that the bacterium can recognize the parental DNA strand, which is methylated and correct the errors on the daughter strand, which gets methylated after this reaction.

In eukaryotes, about 20% of the cytosines are methylated and the main known function is not as mentioned above host defence, but control of transcription. Typically the cytosine in CpG dinucleotides is methylated. ^{m}C is always methylated in both DNA strands symmetrically:

$$5' \; {}^{m}CpG \; 3'$$
$$3' \; GpC^{m} \; 5'$$

A vast group of enzymes control DNA methylation in eukaryotic cells. Demethylases like MBD2, MBD4 or TDG (Ooi and Bestor, 2008) remove CH_3 residues from the DNA and methylases like DNMT1, DNMT3a or DNTMT3b add methyl-groups to cytosines (Jair *et al.*, 2006). The methyltransferases can be divided into two groups, the maintenance methylases and the *de novo* methylases. The maintenance methylases recognize hemi-methylated sites, which arise from replication of the DNA. In this case one strand of the DNA is methylated (parental) and one strand is left unmodified. These enzymes work constitutively and convert

every hemi-methylated site into a fully methylated one with virtually 100% efficiency. This process ensures that the methylation pattern is held up after replication.
In contrast to the maintenance methylation system, the *de novo* methylases act only on non-methylated DNA, where they add a methyl-group on one strand. Generally spoken, if a promoter region gets methylated, the gene-expression is truncated via diverse epigenetic mechanisms or binding of repressive acting proteins (Gopalakrishnan *et al.*, 2008).
An important mechanism controlled by DNA methylation is genomic imprinting. This mechanism starts during gametogenesis. First, the existing methylation pattern is erased, followed by recreation of the methylation pattern during maturation of spermatocytes or oocytes. The methylation pattern of maternal and paternal gametes is different, but only a few regions of each genome are unmethylated at this stage of development (Butler, 2009).
The second systematic alteration of the DNA methylation pattern takes place in early embryogenesis. At that time, some sites stay methylated, whereas de-methylation activates many genes. This process results in a specific pattern of methyl groups in germ line cells, which is on imprinted loci dependant on the inheritance of the alleles form either the mother or the father.
Exceptionally we can find the hereditary information in the methylation pattern of the DNA and not in the DNA sequence, in other words one can distinguish between the parental origins of alleles whose sequence is identical by the differential methylation pattern. (Lewin, 2004)

1.1.3 CpG islands

CpG means Cytosine phosphatidyl Guanine. This term was introduced to distinguish easily between GC frequency and CG dinucleotides. The “p” refers to the phosphodiesterbond between every particular base in a DNA molecule. The CpG dinucleotides should statistically appear with a frequency of 4% in the DNA sequence, considering the G-C content of 41% in the human genome. Actually, the commonness of this dinucleotides is 0,8%, in other words only 20% of the expected frequency. The reason for this fact is found in the chemical instability of the base Cytosine, which can get converted into a thymidine by deamination due to diverse stress factors. (Knipers, 2001)
Per definition, a CpG island contains a G-C content of minimum 55% over a length of 200bp in minimum (Egger *et al.*, 2004). Using this standard, about 45000 CpG islands were found, but if CG rich sequences in Alu repeats are subtracted, about 29000 CpG islands remain. The typical CpG island is found around the 5´end of a gene. This G-C rich DNA stretch has a

length of 1kb to 2 kb most of them start in 5´region of the promoter and reach into the coding region of the gene. More precise estimations about the number of genes with a CpG island come to 70% of all genes in the human genome, but it should be mentioned that all housekeeping genes have a CpG island in their 5´region. This means that half of the CpG islands in the human genome are found in housekeeping genes, so they are unmethylated throughout, because these genes are constitutively expressed. The remaining CpG stretches are found in tissue-regulated genes, where their gene-expression is, among other things, also dependent on the methylation status of the CpG island in their promoter-region. In the end, less than 40% of non-constitutive expressed genes bear CpG islands (Lewin, 2004).

DNA methylation can induce several courses in a cell. It may comparably cause the binding of repressors or enhancers dependant on the gene region. Further DNA methylation may silence gene-expression by preventing these factors from binding to the DNA.

However, the current understanding of DNA methylation suggests that if a CpG island is methylated throughout, there is no gene-expression, whereas the unmethylated status of a CpG island is an indication for active gene-expression. In other words a hypo-methylated gene may be active (Figure 2), but the absence of methyl-groups is no evidence for the transcription of a particular gene.

In every case, no universal prediction can be made about linkage between DNA methylation and gene-expression, because transcription also takes place if the CpG island is partially methylated. It seems that a few methyl residues at the core region of a gene are responsible for the silencing of a gene. Anyway, the general rule maintains that DNA methylation prevents genes from expression or at least is a strong indicator for a silenced gene.

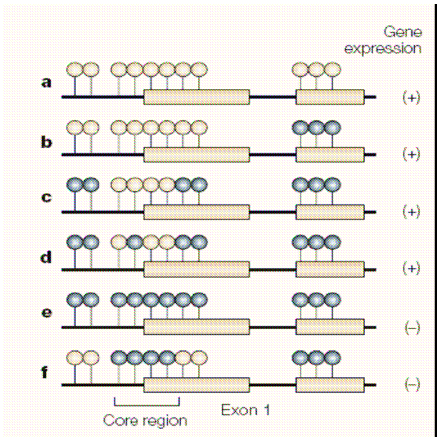


Fig. 2 The effect of the methylation pattern on gene-expression. This schematic drawing shows possible effects of methylation on gene-expression, observed at gene start region of a gene associated with aging chronic inflammation. Normal tissue samples of young individuals are unmethylated at this site. Open circles mark unmethylated CpGs; close circles mark methylated CpG islands; **(a)** promoter region is unmethylated; gene-expression can take place; **(b)** methylation in the 3´region of the promoter does not impair transcription; **(c,d)** core region is not methylated therefore the gene is still expressed. **(d)** Methylation pattern promotes further methylation and results in complete methylation of this site **(e)**; **(f)** Only the core region is methylated, which leads to inactivation of the gene. adopted from: (Ushijima, 2005)

1.2 Breast cancer

1.2.1 Cancer

Hippocrates, who noticed a similarity between a breast abscess and the leg of a crab, introduced the term cancer. Later all superficially detectable abscesses, which infiltrated neighbouring tissues were called cancer.

Individual somatic cells, which lose their ability to grow, divide and undergo apoptosis in a distinct span of time that is predetermined by the normal cell cycle, lead to an illness, called cancer or malignant neoplasm. Cancer cells have quasi lost their growth control and gained the ability to grow in inappropriate locations. In short, these are normal cells that get immortalized due to mutations or epigenetic events.

The appearance of cancer can be understood as a multi-step process that results in an accumulation of mutations and genetic instability of the affected cells. The fundamental model assumes that the initiation of a tumor needs two or more alterations in the genome or epigenetic events of the affected cell, such as changes of methylation pattern or miRNA expression, followed by further changes that strengthen the tumorigenic state of the cancer cells (Knudson, Jr., 1971)

One of the central dogmas in current understanding of cancer, is an increased mutation rate of cancer cells, which leads, on one hand to the accumulation of mutations mentioned above and, on the other hand, to an acceleration of clonal selection of tumor cells, because cells, which divide more aggressively will overgrow the remaining healthy ones. Therefore cancer progression can be seen as selection between and final survival of the fittest cancer cell, which results in an enhancement of cancerous properties of the cells and lastly a development, which invariably goes in the direction of the more malignant cell. (Lewin, 2004)

In principle, tumors can be partitioned into benign neoplasms and malign neoplasms, whereat the benign tumor cells stop to proliferate by themselves (Ueng *et al.*, 2009). Per definition, a benign tumor does not grow unlimited and does not invade neighbouring tissue or show any metastasizing characteristics. In other words, benign tumors sum up all abnormal tissue masses that cannot be classified as cancers. Although these benign neoplasms have the potential to become malignant, it happens very seldom and the benign tumors maintain their typical encapsulated shape, which mostly inhibits the creation of malignant characteristics. In most cases this neoplasm causes no health problems.

Malign neoplasms, also referred to as cancer or invasive tumor, are classified in different grades and stages. A pathologist defines the similarity between the surrounding tissue and the tumor tissue, which leads to a division in different grades of the cancer. This classification starts with G1; where tumor cells resemble normal cells and grow slow, whereby the malign neoplasm shows low aggressive behaviour. G4 is the highest grade, where cells do not look like normal cells anymore and the tumor grows and spreads fast. (http://www.cancertreatmentwatch.org/general/grading.shtml)

The stage of the cancer is defined by the degree of the invasion of the cancer cells in the patients' body. This overall stage grouping ranges from 0 to 4 often in roman numerals; here stage 0 is a carcinoma in situ. This neoplastic cells proliferate in their normal habitat and do not invade into neighbouring tissue, which may lead to a malign neoplasm if long enough untreated. As one can imagine stage IV is the worst development of the illness, where cancers often have metastasized throughout the body.

For diagnoses and cancer treatment of most solid cancer types a staging system called TNM has been accomplished. T refers to the tumor and an added number to the size of the primary tumor and is graded from 1 to 4, whereas the size increases with the number. The parameter N should give an insight in the degree of spread to regional lymph nodes and M reflects the presence of metastases. By means of this assessment, a decision towards the best treatment of the malign neoplasm like surgery, radiation therapy or chemotherapy, is made.

The risk of getting diagnosed with cancer rises rapidly after a certain age (Figure 3). Therefore the need for early and precise diagnoses of any kind of cancer is unquestioned. The data in Figure 3 refer to a survey of SEER the National Cancer Institute's Surveillance, Epidemiology, and End Results, where a hypothetical cohort of 10 million alive birth were analyzed using standard statistical methods. Most of the data were collected in 1987-88 from SEER.

A male person at age of 60 has for example a chance of 6.42% to be diagnosed with prostate cancer till he is 70 years old, but there is a chance of 16% that prostate cancer is diagnosed till the end of his life (http://seer.cancer.gov/). Further there exists a likelihood of about 40% that one gets diagnosed with cancer in a lifetime and a likelihood of 20% that this patient dies of this disease. (http://seer.cancer.gov/)

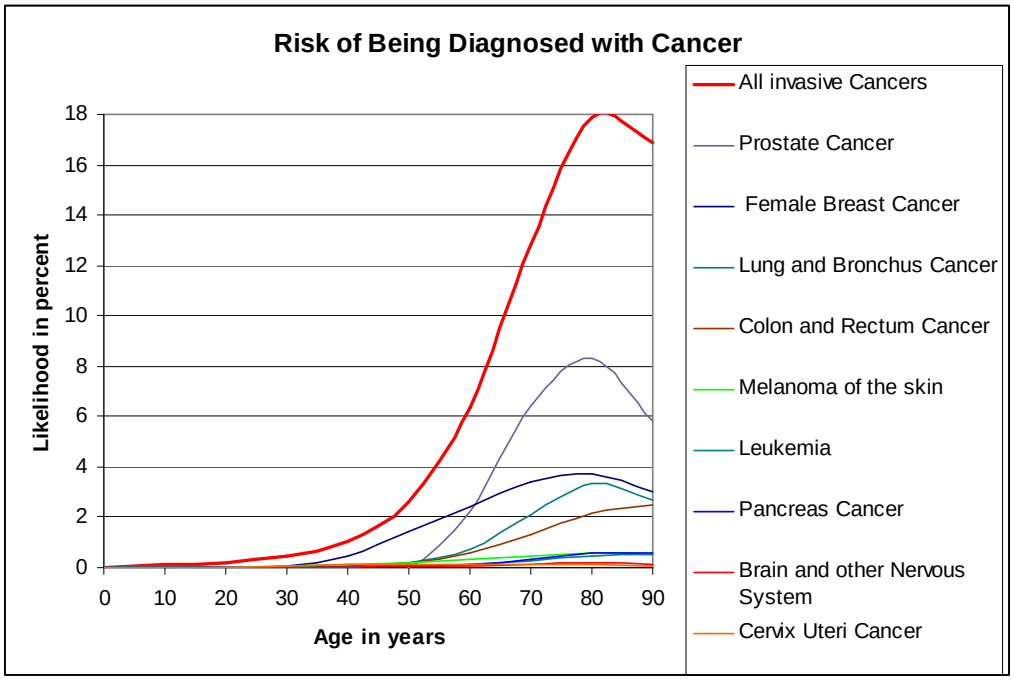


Fig.3: Lifetime risk of being diagnosed with cancer. Data consider both sexes and all ethnicities. Incidence data are from the SEER 17 areas San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California excluding SF/SJM/LA, Kentucky, Louisiana and New Jersey. Source: National Cancer Institute (http://srab.cancer.gov/devcan/)

1.2.2 Breast cancer

Breast cancer is the most common type of cancer in female population. For women there is a likelihood of 12% to get diagnosed with breast cancer and a probability of 2. 8% to die from this illness in a lifetime (http://seer.cancer.gov/). A death rate of 2. 8% implies 411 000 death among female population worldwide per year (Parkin *et al.*, 2005).

Generally, a malign neoplasm in the area of the breast can emerge from milk ducts in 70 to 80% of all cases or from the lobules of the breast in 10% of incidences. Other types like medullar, mucinous, mucoid or colloid breast cancers are rare, but treatment and diagnosis are similar to ductal and lobular breast cancer (Watson, 2001).

If two or more persons in a family get diagnosed with breast cancer there may be a genetic predisposition for this illness in this family due to a germ line mutation in the breast cancer1 (BRCA1) (17q21) or BRCA2 (13q12.3) gene. Current studies revealed that about 20% of all breast cancer incidences arise from a mutation of the BRCA1- or BRCA2-gene. (Risch *et al.*, 2006).

These genes belong to the tumor suppressor genes, where a loss of function leads to a cancer predisposition. In this case, the DNA repair system is impaired, because the zinc finger

motive of the BRCA1 protein binds to branch DNA structures and prevents form diverse nuclease digestions. Further it promotes the DNA repair by NHEJ or HR (Chistiakov *et al.*, 2008).

BRCA2 has a similar function as key protein in double strand break repair. Both proteins, the BRCA1 and BRCA2 gene product, interact with a large number of proteins, which are responsible for accurate DNA repair, and therefore a mutation of these two genes leads to an increased genomic instability (Durant and Nickoloff, 2005), which results in a probability of 90% to develop breast cancer and 24% to develop ovarian cancer for BRCA1 mutation carriers (Risch *et al.*, 2006). More than 1200 different mutations are known, which are distributed on all exons of the BRCA1 gene, showing almost all types of mutation, like single base exchange, insertion, deletion or inversion, leading to frame shift, miss-sense, nonsense and splice-site mutations.

The prevalence of BRCA1 carriers in the overall population is estimated around 1%. Remarkably this value differs between ethnic groups, where Ashkenazi Jewish show a prevalence of 8%, whereas Asian American show a likelihood of 0,5% to carry a BRCA1 mutation (John *et al.*, 2007).

The assessment of the hormone receptor status has developed to a key parameter molecular classification and is part of each tumor diagnoses process. Using immunohistochemistry of the tumor markers estrogenic-receptor (ER) for the steroid hormone “17 – estradiol” as well as Progesterone Receptor (PR) and Her2 are stained and analysed. The expression of the tumor-markers, ER and PR is linearly related to poor prognoses of breast cancer ((Sherman *et al.*, 2007). In other words, the risk of a malign neoplasm increases with a positive hormone receptor status.

An early diagnosis of cancer is essential, because early treatment improves survival. The death rate of the patient rises in the first five years after diagnoses to a value of 12%, after another five years to 20% (http://seer.cancer.gov/). A cheap and minimal invasive diagnostic test that could be accomplished by applying the results of diverse studies performed on blood sera of patients, may contribute to a simplification of the early recognition of cancer (Anker *et al.*, 2003).

1.2.3 Tumor predisposition syndromes associated with breast cancer

Cancer may have hereditary causes, but other effects are said to have a higher impact on the risk of developing a tumor. These carcinogens can be found in every area of life and the vast majority of all cancer incidences are said to be due to such agents. Further exposure to UV-radiation and nuclear radiation as well as some retro viruses were brought in close connection to the arising of cancer (Juzeniene *et al.*, 2009).

One of the first approaches to understand the development of cancer is the two hit model of the group around Alfred G. Knudson. They produced the direct evidence that cancer can arise in a few steps as two via statistical analyses of hereditary and sporadic retinoblastoma in the 1970ies (Knudson, Jr., 1971). Several other genetic changes can lead to a cancer predisposition, like a germ line mutation of the gene TP53, which is called Li-Fraumeni syndrome (LFS). This tumor suppressor gene is also known as P53, a DNA binding protein that induces cell cycle arrest, apoptosis and changes in metabolism. Also somatic mutations and aberrant methylation are observed at this locus (discussed in the next section). The proportion of cases in breast cancer is very low. Only about 1% of developed breast cancers bear such a P53 germ line mutation. A germ line mutation of this gene also leads to higher rates of brain cancer, childhood adenocortical carcinoma (ACC), gastric cancer and lung cancer (Hodgson *et al.*, 1995).

Sporadic cancers are caused by somatic mutations, aberrant methylation and various other causes, which were partially mentioned above.

Although, the larger fraction of breast cancer cases can be rated as sporadic cancer, a twin study performed by Lichtenstein P. (Lichtenstein *et al.*, 2000) pointed out that 27% of tested cases were due to a heritable factor. The difficult search for an origin of sporadic cancers produced the cancer stem cell hypothesis (CSC), which describes an alternative scheme to the “stochastic model” of carcinogenesis. This model (CSC) proposes that multiple mutations or other events, like aberrant methylation of a random single cell occurs and is followed by a clonal selection of the fittest cancer cell. Recently a model was proposed (Figure 4) that is based on the observation of the different cancer phenotypes and it combines both approaches in one integrative hypothesis (Melchor and Benitez, 2008). This hypothesis about the origin and the development of breast cancer was created upon the detailed analyses of common cancer phenotypes like the ERBB2 cancer cell line, which over-expresses the ERBB2 locus and the genes located in 17q11. Another appearance of breast cancer cells is the luminal B phenotype. In this case, the cells exhibit many cell surface proteins comparable to basal-like

cancer cells, but the cells are estrogen receptor positive. In the course of this analyses of the certain cancer phenotypes Melchor M. and co-workers (Melchor and Benitez, 2008) found many similarities between the different pathways concerning the arising of cancer, and many different pathways that may result in the same cancer phenotype. These findings finally lead to this model (Figure 4).

This hypothesis may actually be applicable to the development of cancer for the majority of breast cancer and it provides a good insight of the genetic events that lead to cancer and highlights coherences that may facilitate the analyses of breast cancer data. Further, this already simplified illustration (Figure 4) emphasizes the complexity of the illness, which we and many other researchers try to understand.

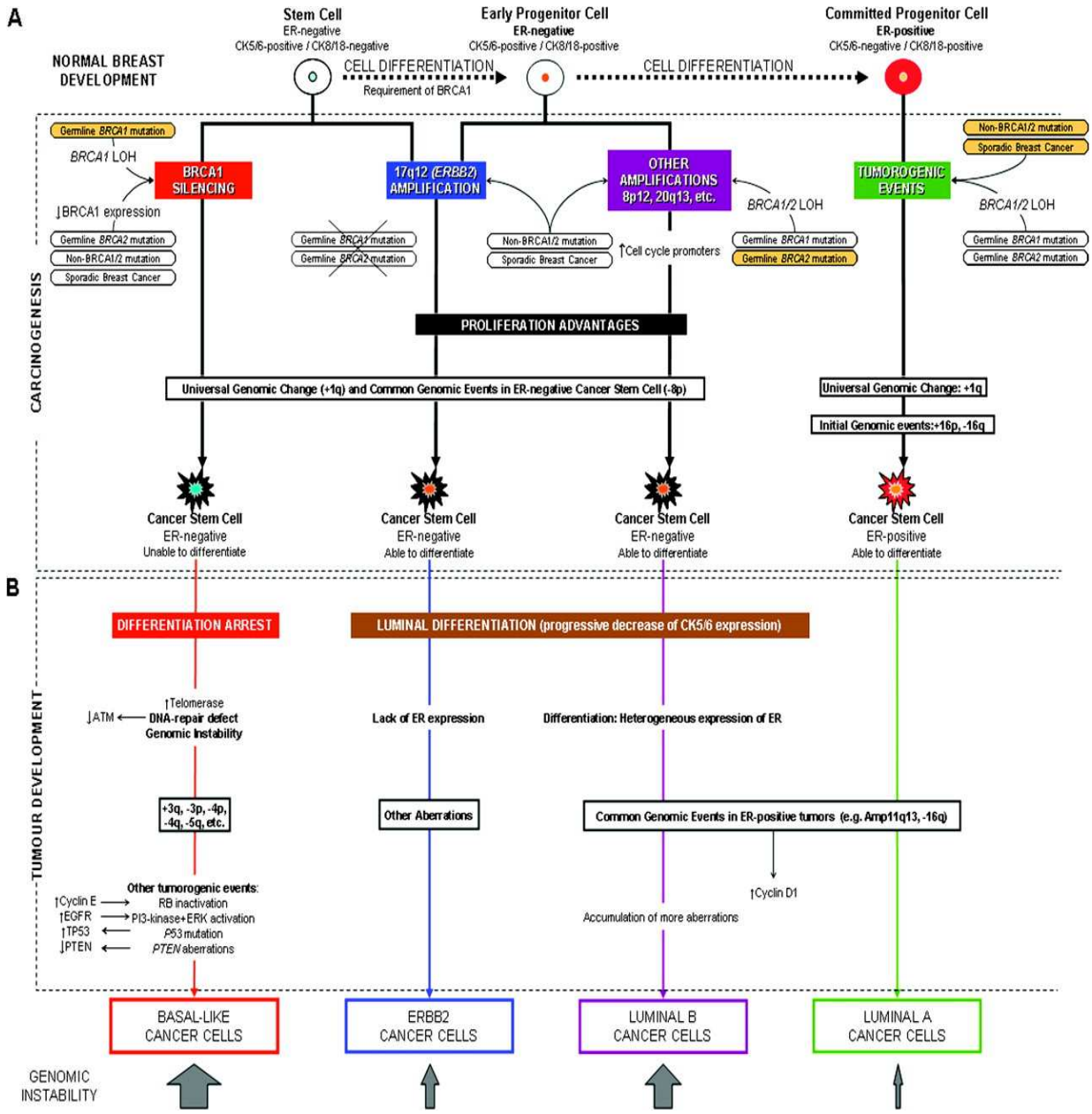


Fig.4: Integrative approach of the CSC hypothesis and the stochastic model. The picture starts with the development of a normal breast cancer cell including their cytokeratin status. Red-labelled boxes represent the basal like cancer cell phenotype including their carcinogenesis (section A). Blue-labelled boxes show ERBB2 phenotype, and so on. Yellow boxes in section A indicate the most frequent pathway of the breast cancer class. Crossed out boxes represent the inability of cells exhibiting a BRCA1/2 mutation to undergo this pathway (ERBB2). The thicknesses of the arrows at the bottom indicate the genomic instability observed in the certain cancer phenotypes. Adopted from (Melchor and Benitez, 2008)

1.2.4 Tumor-suppressor-genes and Oncogenes

The typical tumor-suppressor gene is expressed in normal tissue cells and its inactivation in cancer cells leads to growth advantages of these tumor cells. Frequently this occurs upon methylation of the promoter region of this certain genes, which are the main targets of our research.

One example for such a gene, which is also found methylated in human blood serum of breast cancer patients, is RASSF1a. The promoter hypermethylation of this gene is virtually found in every type of invasive cancers and the silencing goes along with expression-changes from hundreds of genes(Kioulafa *et al.*, 2009). Amongst others the gene product of RASSF1a interacts with XPA, a DNA repair protein, and impair cyclin D1, which induces cell cycle arrest.

Oncogenes have the antithetic effect. They are silenced in normal tissue, and their expression facilitates tumor progression. Prominent examples are genes of the Src family (v-src sarcoma; Schmidt-Ruppin A-2), which are believed to play a critical role in cancer development. The gene products of this family are kinases, which are involved in diverse pathways that enable the rapid growth of cancer cells. Recent reports show that an inhibition of these gene products can slow down tumor cell proliferation significantly (Laird *et al.*, 2003).

1.2.5 Epigenetic events in cancer development

Hereditary cancer can also be due to aberrant methylation of certain genes. In this case a disruption of the imprinting mechanism leads to a cancer predisposition. Insulin-like growth factor 2 (IGF2) is expressed differently in healthy individuals depending whether the mother or the father inherits the allele. An aberration of the methylation pattern in this genomic region (11q15) leads to a cancer predisposition, which is called Beckwith-Wiedemann syndrome that is associated with embryonic tumors like Wilms´tumor or rhabdomyosarcoma. In this case the imprinted cluster of at least 12 genes including IGF2 and H19 is methylated incorrect in early development, which leads to this predisposition (Weksberg *et al.*, 2003).

A loss of this coordinated regulation was recently also observed in a certain fraction of sporadic gastric cancer patients. In this case, the IGF2, an early mitogenic factor and cell proliferation-enhancing factor in healthy individuals, is hypo-methylated in one allele in cancer cells, which leads to biallelic expression of IGF2. Recent studies show that this hypomethylation is not due to environmental influences, but were found in normal tissue and

in malign neoplasm of patients too, which indicates that the methylation status of IGF2 in gastric cancer may be an inherited factor (Lu *et al.*, 2009).

Another current hypothesis proposes the involvement of the polycomb-regulated genes in the development of cancer predisposition due to methylation events. These polycomb group proteins, mainly transcription factors, are essential for human development. The regulation of these genes occurs via the polycomb repressive complex 2 (PRC 2), which silences genes upon histone trimethylation and targets over 1800 genes (Lee *et al.*, 2006). Researchers, investigating the epigenetic changes in cancer progression, observed an enrichment of PRC 2 target genes in several methylation screenings and postulated therefore a model for aberrant promoter hypermethylation in several cancers including breast cancer (Figure 5) (Widschwendter *et al.*, 2007).

The mechanisms of PRC-DNMT cross talk is till today not fully understood, but methylation data for PRC2 targets, which are methylated at a higher frequency at certain phenotypes, suggest a model like figure 5. This theory suggests that the PRC-DNMT crosstalk occurs at low frequency in an early stage of development and does not influence normal differentiation, further this results in a cell with a predisposition for cancer development. A further indication for the accuracy of this model is tumor specific promoter methylation of genes (e.g. NEUROD1), which are not expressed in normal tissue but serve as methylation markers in cancer development without any known function in tumor progression of breast cancer or B-cell lymphoma (Pike *et al.*, 2008).

The most conclusive evidence for this model is a 12-fold more likely cancer specific methylation of PRC2 targets observed by Widschwendter (Widschwendter *et al.*, 2007). Further, McCabe and co workers investigated CpG islands, which are targets of DNMT and showed an aberrant methylation in cancer cell lines in diverse “knock-down” experiments (McCabe *et al.*, 2009). These researchers figured out that about 50% of this so-called methylation–prone CpG islands were targets of the SUZ12, the binding domain of PRC2. This finding also strongly suggests a DNMT-PRC 2 crosstalk.

Applying the stochastic model of cancer development, many researchers screen for methylation events in cancer progression that lead to a growth advantage of the cancer cell, facilitates invasive growth, prevents the cell from apoptosis or accelerates metabolism and so on. This approach is mainly performed upon broad screening methods of cancer samples of multiple cancer types and has lead to the discovery of large cohort of tumor suppressor genes and oncogenes aberrantly expressed upon methylation. Due to the very heterogeneous cancer

subclasses in every cancer type, a clear determination of markers with prognostic significance or fitting as drug target is very rare.

Therefore researchers recently search for methylation signatures. In the field of breast cancer Roll et al. tried to find different methylation signatures in various breast cancer cell lines by over-expression of DNMT 1, 3a, 3b, which is thought to be an early event in carcinogenesis (Roll *et al.*, 2008). Within their 12 tested breast cancer cell lines they could determine two phenotypes: hypermethylator cell lines, where the over-expression of all DNMT gene-products resulted in a significant increase of methylated CpG island and low-frequency methylator cell lines. These two phenotypes base on the methylation and expression of six cancers specific methylated genes and further provided the finding that this phenotype can be associated with primary basal-like breast cancers.

Hence the conclusion can be made that CIMP (CpG island methylator phenotypes), which were initially found exclusively in colon cancers (Dahlin *et al.*, 2010), also can be determined in breast cancer (Paluszczak *et al.*, 2010). This approach provides the possibility of more detailed investigation of cancer progression on the basis of methylation analyses and may enhance the broad search for new methylation markers.

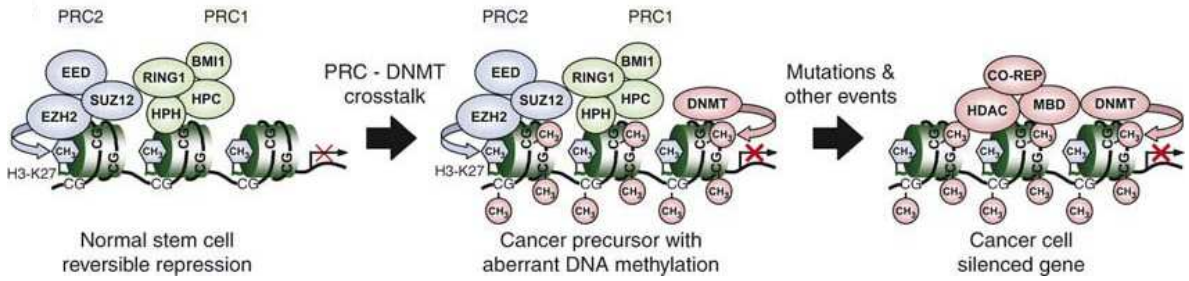


Fig.5: In normal tissue the regulation of polycomb group genes occurs upon binding of PRC2 and PRC1. The hypotheses about creation of cancer precursor cells bases upon the recruitment of DNMT, a *de novo* methlytransferase that methylates the DNA in early precursor cells. The procedure results in an alternative silenced gene, at which a re-induction of expression is facilitated in cancer cells. Adopted from (Widschwendter *et al.*, 2007)

1.2.6 Methylation screening

In principle DNA methylation analysis may be divided into three approaches, which are enzyme digestion, affinity enrichment and sodium bisulphite conversion of the DNA. They can be found in figure 6 and are discussed in brief below. Further one may distinguish between genome scale approaches and loci specific methods, where only a few promoters

were characterized by single base pair resolution. Genome wide approaches like DMH (differential methylation hybridization), RLGS (restriction landmark genome scanning) or MIRA (methylated CpG island recovery assay) have the disadvantage not to reach a single base pair resolution and a creation of a precise epigram of a gene locus was not possible (Laird, 2010). Next generation sequencing combines these two approaches: a genome wide analysis of a methylome and a resolution of a single base pair. Therefore it has been already combined with each of the thee principle approaches (figure 6) to determine the methylation status of DNA (Laird, 2010).

MSP, methylation specific PCR, is till today the most prominent method to assign the methylation status of certain genomic loci. By this method the DNA is treated with sodium bisulfite, which converts all unmethylated cytosines bases to uracil bases. Subsequently primers, specific for methylated or unmethylated targets, are applied to perform an amplification that allows the detection of the differentially methylated targets.

Another option elucidating DNA methylation is using MSRE´s, methyl sensitive restriction enzyme digestion, which are blocked by methylated CpG or digest exclusively methylated CpGs. Upon MSRE-digestion all unmethylated/methylated targets are digested and the un/methylated DNA, which is not affected by the digestion, can be amplified and detected. Dependant on which set of enzymes are used, McrBC for digestion of methylated DNA or HpaII, Hin61 etc. for digestion of unmethylated DNA, the respectively other fraction (i.e. methylated or unmethylated) can be analysed.

MeDIP is a rather new approach to identify differentially methylated DNA sequences. Here the DNA gets sonicated and denatured, then antibodies specific for 5-methylcytosines are used to precipitate the methylated DNA, which subsequently get labelled and hybridized on a micro array or sequenced by new high throughput methods (Mohn *et al.*, 2009). Another possibility to perform an affinity enrichment on methylated DNA (not shown in figure 6) is provided by the utilization of methyl binding proteins for loci specific analyses or at more recent approaches in combination with next generation sequencing (NGS) (Li *et al.*, 2010).

The Next Generation Sequencing approach (developed by 454) to determine methylated DNA regions is a special application of pyrosequencing. The procedure of pyrosequencing is based on “Sequencing by Synthesis” measuring the release of pyrophosphate.

If the template DNA for this reaction has underwent a sodium bisulfite conversion a precise epigram of each template DNA strand can be created.

Pretreatment	Analytical step			
	Locus-specific analysis	Gel-based analysis	Array-based analysis	NGS-based analysis
Enzyme digestion	• *Hpa*II-PCR	• Southern blot • RLGS • MS-AP-PCR • AIMS	• DMH • MCAM • HELP • MethylScope • CHARM • MMASS	• Methyl–seq • MCA–seq • HELP–seq • MSCC
Affinity enrichment	• MeDIP-PCR		• MeDIP • mDIP • mCIP • MIRA	• MeDIP–seq • MIRA–seq
Sodium bisulphite	• MethyLight • EpiTYPER • Pyrosequencing	• Sanger BS • MSP • MS-SNuPE • COBRA	• BiMP • GoldenGate • Infinium	• RRBS • BC–seq • BSPP • WGSBS

Fig.6: Overview of principles of DNA methylation analysis. AIMS (amplification of inter-methylated sites), BC–seq (bisulphite conversion followed by capture and sequencing); BiMP (bisulphitemethylation profiling), BS (bisulphite sequencing), BSPP (bisulphite padlock probes), CHARM (comprehensive high-throughput arrays For relative methylation), COBRA (combined bisulphite restriction analysis), DMH (differential methylation hybridization), HELP (HpaII tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification), MCAM (MCA with micro array hybridization); MeDIP, mDIP and mCIP (methylated DNA immunoprecipitation), MIRA (methylated CpG island recovery assay), MMASS (based-based methylation assessment of single samples), MS-AP-PCR (methylation-sensitive arbitrarily primed PCR), MSCC (methylation-sensitive cut counting), MSP (methylation-specific PCR), MS-SnuPE (methylation-sensitive single nucleotide primer extension), NGS (next-generation sequencing), RLGS (restriction landmark genome scanning), RRBS (reduced representation bisulphite sequencing; –seq, followed by sequencing), WGSBS (whole-genome shotgun bisulphite sequencing); Adopted from (Laird, 2010)

1.2.7 “Methylation” drugs

One of the prime reasons, why the understanding of methylation events in cancer is that important and therefore is addressed by many researchers, is that methylation is a reversible process. This implies that methylation is a preferable drug target for therapy of cancer patients.

Recently some demethylating agents like 5-azacytidine (Vidaza) and 5-aza-2'-deoxycytidine (Decitabine) have been tested on cancer cell lines and researcher suggest these agents as an alternative cancer treatment (Datta *et al.*, 2009). These agents degrade DNMT1 upon a proteasomal pathway, which leads to the reactivation of tumor suppressor genes like P16, MLH and TIMP3 by de-methylation and re-expression and does not show significant toxity to

the used cancer cell lines. The FDA (U.S. Food and Drug Administration) approved these drugs already for the treatment of Myelodysplastic syndrome (MDS) (Szyf, 2009).
A quite different approach to influence methylation in cancer progression is the recruitment of MBD-2, a methyl-binding domain protein. In vitro translation of a recombinant form of MBD-2 induced re-expression of silenced genes in cancer. On the other hand "knock-down" experiments in vitro and in vivo triggered an inhibition of tumor growth, invasiveness and metastases, which let speculation about antisense oligonucleotides, siRNA inhibitors, and MBD2 antagonists as promising cancer drugs. (Szyf, 2009).

1.2.8 DNA in serum

The prognostic potential of cell free nucleic acids in plasma and serum caused a great deal of attention in current efforts to improve diagnoses of diverse malignant cancerous diseases (Lo and Chiu, 2009). Various approaches in this field like the determination of genetic aberrations, the methylation status of the DNA (Fleischhacker and Schmidt, 2007) and the DNA level (Anker *et al.*, 1999) have been reported to have diagnostic potential.
In the sector of breast cancer diagnoses the prevailing situation is the same, although the current diagnostic techniques as physical examinations or diverse imaging methods are convincing the need for a simple blood test would be high. Many methylation markers for breast cancer determination of tissue biopsies are already well established, but the search for aberrant methylated DNA in serum has till now not reached this advance. Although the screening for such markers is a consistent proceeding process and recent reports demonstrated, for example, aberrant methylated promoter region of RASSF1a even at about 20% of tested patients before they got diagnosed with breast cancer, the panel of evaluated breast cancer markers in blood serum is low (Yazici *et al.*, 2009) (Van, I *et al.*, 2009) (Jing *et al.*, 2008) (Shukla *et al.*, 2006)

2. Aim of the thesis

In this work we optimized and evaluated different serum DNA isolation strategies and further assume the DNA binding characteristics of the MBD protein, which was one of our DNA isolation approaches under test.

Moreover we tested several genome wide amplification methods for the enrichment of the methylated DNA in our experimental setup through a comparison of three different approaches.

Finally we aimed at the discovery of methylation markers for clinical diagnoses in serum of breast cancer patients by performing a screening for methylation markers using a panel of 360 genes for multiplexed amplification of the genome wide amplified DNA followed by hybridization.

3. Material and Methods

3.1 Patients and Serum sample preparation

Blood was taken from 12 healthy adults, age and sex matched, for methods setup. After incubation in Vacuette 9ml Z Serum Clot Activator (Greiner Bio One) for 30 minutes the samples were centrifuged at 1,800-x g for 10 minutes at room temperature. Aliquots of 2ml were stored at -80°C.

Serum samples (n=82) obtained form the AKH Wien (General Hospital Vienna) were taken between July 2002 and November 2006, centrifuged at 1000 x g for 15 minutes and stored in 200µl aliquots at -80°C until usage. For this study serum specimen of 40 patients with a malign neoplasm, sera of 12 patients with a benign tumor, 6 patients with a carcinoma in situ and 24 normal controls were processed (table 4).

3.2 DNA isolation from serum

3.2.1 Roche High pure template preparation kit

DNA isolation from serum using common silica membrane based isolation strategy was performed applying the Roche High pure template preparation kit (Roche Applied Science). The protocol was adapted according to Müller H. M. (Muller *et al.*, 2003). Isolation was performed according to manufacturer's instruction except the following protocol steps: the 800µl of serum samples were split into 2 aliquots of each 400µl and each mixed with 400µl of Roche Binding Buffer and 80µl ProteinaseK (Fermentas, 20mg/µl). After 15 minutes of incubation at 55°C, 200µl isopropanol was added to each aliquot. This aliquot was mixed and twice a volume of 1080µl (i.e. two aliquots) was loaded to the column subsequently followed by a centrifugation step of 1 minute at 8000-x g. The flow through after each centrifugation was transferred back onto the same column and centrifuged again. Inhibitor-removal and washing steps were performed according to manufacturers´ protocol, but the DNA was eluted in 55µl of elution buffer.

3.2.2 Qiagen Blood midi kit

Alternatively DNA from 1ml serum was extracted with Quiagen DNA Blood Midi kit for large-scale genomic and viral DNA purification form whole blood, plasma, serum and body fluids (Quiagen, Hilden, Germany) according to the manufactures instructions. Except the elution volume was lowered form recommended 200µl to 100µl.

3.2.3 ZR Serum DNA kit

The ZR Serum DNA kit (Zymo Research, U.S.A.) was applied, as recommended in manufacturers instructions, to isolate DNA form 800µl serum in a silica bead based manner.

3.2.4 MBD based Isolation

3.2.4.a Bead assembly

The Escherichia coli strain BL21, containing the pET6HMBD plasmid, kindly provided by Sally H Cross, was grown in Luria-Bertani medium with 30µg/ml chloramphenicol and 50µg/ml ampicillin. Recombinant protein expression was induced by adding IPTG to an end concentration of 0,4mM, at a measured optical density (600) of 0,6. E.coli cells were broken up by repeated sonication in Lysis Matrix A tubes (MP Biomedicals). The bacterial extract was then centrifuged at 13,000rpm for 10 minutes and supernatant was transferred to a new vial. Centrifugation step was repeated until all cell debris was removed. The pellet was resuspended in buffer A in ratio of 1g pellet plus 2ml buffer A and stored at –20°C until usage.

200 µl of a 50% suspension of Ni-Sepharose beads (Adar Biotech, Rehovot, Israel) were equilibrated through resuspension in 500µl water, followed by centrifugation 1min 1000 x g and removal of the supernatant from the beads. The beads were then equilibrated by adding 500µl buffer A (20mM HEPES, 100mM NaCl, 10% glycerol, 20mM beta-Mercaptoethanol (added daily fresh), 0,5mM PMSF, 0,1mM TritonX, pH=8) and centrifuged at 1000g for 1min. This procedure was repeated twice. Finally the beads were resuspended in 100µl buffer A. His-tagged HMBD-protein was bound to the equilibrated beads by addition of 800µl bacterial extract (derived from approximately 300mg E.coli wet cell weight) and 800µl 2x

buffer A and incubation for two hours on ice at which the beads were held in suspension upon repeated mixing.

Bead-HMBD protein-complexes were washed twice by adding 500µl wash buffer (buffer A, plus 10mM Imidazol) followed by centrifugation at 1000-x g for 1min. The proteins were either eluted from beads with buffer A plus 500mM Imidazol to perform a SDS page gel or residual E.coli DNA was removed with buffer B (1,5M NaCl, 20mM HEPES, 20mM □-Mercaptoethanol, 0,5mM PMSF, 0,1mM TritonX, pH=8 buffer). To apply this fraction for DNA isolation procedure, the beads were resuspended in 100µl Buffer A. Protein concentrations were measured with a BioRad Dc-Protein assay (Biorad, Hercules, California). Protein purification was visualized on a SDS page gel (NuPage Novex Bis-Tris Gel) by applying aliquots of each purification step onto the gel. Proteins were denatured via 10 minutes at 70°C, mixed with loading dye (Biorad, Hercules, California), and loaded onto the gel, where 200V were applied for 35 minutes.

3.2.4.b DNA extraction

For DNA extraction using the MBD immobilized Ni-beads 1 ml of serum was incubated with 1 ml of 2x buffer A and 60µl of prepared 50% bead/Protein-complex suspension in buffer A for 2h on a thermo mixer (Eppendorf) at 450 rpm at room temperature. After centrifugation 1000 x g for 1 min. the reaction was separated into a bead fraction, which was used for further DNA isolation protocol, and a serum fraction, which was tested for protein integrity by a protein micro array, suitable to detect auto-antibody profiles. The DNA-MBD complex was washed twice with buffer C (20mM Hepes, 100mM NaCl, 10% glycerol, pH 8) and resuspended in 145 µl Tris-Cl buffer, pH=8. To elute the DNA from the MBD/bead-complex, a ProteinaseK-digestion was performed in a volume of 150µl by adding 5µl ProteinaseK (Fermentas, 20mg/ml) to the resuspended beads. The reaction was incubated at 55°C for 20 min followed by heat inactivation of the ProteinaseK. DNA was then isolated from the ProteinaseK supernatant using Quiagen MinElute columns. Purification was performed according to manufacturers` instruction with an elution volume of 17µl.

3.3 DNA concentration measurement

DNA concentration was measured with “Quant it” Pico Green (Invitrogen) according to the manufacturers protocol with a reduced reaction volume of 100µl. Samples and standards were excited at 480nm and emission was read at 520nm using a BioRad IQ5 Real time PCR detection system. A five-point lambda DNA standard curve ranged from 15pg/µl to 250pg/µl in a measuring volume of 100µl. Per serum isolate a volume of 5µl was applied to the “Quant it” Pico Green assay.

DNA concentrations within a range of 15ng/µl to 1000ng/µl were measured on a Nanodrop Photometer, measuring the absorption of the DNA samples at 260nm.

3.4 Enrichment of methylated DNA

As illustrated in figure 11 in the section “Serum processing workflow” on page 37, our methylation screening approach is suitable for a genome wide CpG island methylation screen. To meet the requirements for such an approach two restriction enzyme digestions and a ligation is necessary. The first restriction digestion with FspBI fragments the genomic DNA and creates amplify-able DNA fragments sizes as well as a TA-overhang, which is necessary for the subsequent ligation. The second restriction digestion (i.e. after the ligation) cuts all unmethylated fragments that bear a recognition site for the methyl sensitive restriction enzymes. In this way only successful ligated methylated DNA fragments are targets for the genome wide amplification, to enrich the methylated fraction of the genomic DNA.

3.4.1 Creation of sticky ends for ligation dependant WGA

Serum DNA (acc. 3.2.1; conc.: 10-20ng) extracted by Roche High pure template preparation kit was used for restriction digestion with FspBI (10 u/µl, Fermentas). In a total volume of 56,6µl we reached a 1x concentration of 10 x buffer Tango containing BSA (Fermentas) by adding 5,6µl. The creation of TA-overhangs was assured by adding 1µl FspBI (10 u/µl, Fermentas). The reaction was incubated over night at 37°C and heat inactivated by heating the

samples to 65°C for 15 minutes. We assembled a mix including all components (i.e. reaction buffer and enzyme) except the sample at first, and then the combined reagents were aliquoted in this case 6,6µl per vial. Finally the sample (50 µl) and controls were added to the vials to start the reaction. This procedure was performed at all following DNA manipulation and amplification reactions.

3.4.2 Circularization of DNA samples for rolling circle amplification

To create DNA circles upon self-ligation of DNA fragments with TA overhangs the samples were diluted to a concentration smaller than 1 ng/µl. Therefore we used the whole reaction volume of the previous restriction (acc. 3.4.1) digestion of 56,6µl and added 0,283µl of ATP (100mM, Fermentas) to reach the recommended ATP concentration of 0,5mM in the sample volume. Then 24,3µl T4 Ligase buffer (10x Fermentas) was added as well as 215,8µl water (Mayrhofer Phamazeutika, Aqua Bidestillata), to obtain a reaction volume of 300µl, which was necessary to guarantee an appropriate dilution of the DNA. This mix was incubated for 5 minutes at 42°C to dissolve any DNA-concatemers. After the samples were cooled on ice, 3µl of T4 Ligase (5u/µl, Fermentas) were added to every sample. So a reaction end volume of 300µl per sample was maintained. The ligation was performed at 22°C for 8 hours followed by heat inactivation at 75°C for 15 minutes.

3.4.3 Adaptor Ligation

For adaptor ligation mediated amplification an in house developed protocol was applied, where first two complementary ssDNA oligos were mixed to create the ds.adaptor for ligation. Therefore we used oligo A with the sequence: 5´ - TATGAGACTGACTACCAGAT – 3´ and oligo B with the sequence: 5´ - AGTTACATCTGGTAGTCAGTCTCA – 3´. dsDNA were created by mixing two equal volumes (i.e. 10µl) of the two oligos (100µM, Microsynth) followed by heating them to 95°C and a stepwise cooling of the reaction. Thus assembled adaptors were used in the ligation reaction, that consisted of 3,9µl FspBI digested genomic DNA sample (MCF7 cell line DNA), 2µl adaptors, 2µl T4 Ligase buffer (Fermentas), 10,1µl water and 2µl T4 Ligase (5u/µl, Fermentas).

3.4.4 Purification of ligation reaction

Each ligation reaction (acc. 3.4.2) was divided into 3 aliquots á 100µl, which were combined with 300 µl binding buffer of the MinElute reaction clean up kit (Quiagen, Hilden, Germany). The three aliquots per reaction were applied successively to one column. The purification protocol was performed according to the manufacturer's instructions and the DNA was eluted in 10µl of elution buffer.

3.4.5 Methyl sensitive restriction digestion

The entire eluate of the purified ligation reaction (acc. 3.4.4) was included into a methylation sensitive restriction digestion. We combined 10µl DNA sample (acc. 3.4.4) plus 1,5µl buffer tango (10x with BSA, Fermentas), plus 1.6µl of restriction enzyme mix and 1,9µl water to obtain a final volume of 15µl. The enzyme mix consisted of 0.4µl of each of the four methylsensitve restriction enzymes (table 1) namely HpaII (10u/µl, Fermentas); Hin6I (10u/µl, Fermentas); AciI (10u/µl, New England Biolabs) and HpyCH4IV (10u/µl, New England Biolabs). The reactions were incubated at 37°C over night and heat inactivated at 70°C for 20 minutes.

Table 1. Restriction enzymes for methyl sensitive digestion

Name	Recognition Site	blocked by ^{m}CpG
HpaII	5´-C^C G G-3´ 3´-G G C^C-5´	yes
Hin6I	5´-G^C G C-3´ 3´-C G C^G-5´	yes
AciI	5´-C^C G C-3´ 3´-G G C^G-5´	yes
HpyCH4IV	5´-A^C G T-3´ 3´-T G C^A-5´	yes

3.5 Genome wide amplification

3.5.1 Rolling circle amplification

The rolling circle amplification reaction consisted of 13,48µl water; 3µl RepliPHI reaction buffer (10x, RepliPHI phi29 polymerase set, Biozym); 2,7 µl N10 Primer (OD = 125, Microsynth); 2,82µl dNTP (25mM, RepliPHI phi29 polymerase set); 1µl phi29 polymerase (100 u/µl, RepliPHI phi29 polymerase set) and 7µl sample (i.e. restriction reaction), to obtain an end volume of 30µl. Templates in the reaction mix were denatured prior to the amplification by heating the reaction mix at 95°C for 5 minutes followed by cooling on ice water for another 5 minutes. Then 1µl of the phi29 polymerase was added and the reaction was incubated at 30°C overnight. The reaction was stopped by incubation at 65°C for 10 minutes.

3.5.2 APA-PCR

Adapter mediated genome wide amplification was performed with 20µl purified restriction digestion (see section Adaptor Ligation), 10µl dNTP (2mM, Finnzymes), 5µl primer (oligoA, 100pm), 49, 6µl water and 0,625µl Phusion polymerase (Finnzymes). The PCR was 1 min at 98°C followed by 30 cycles of denaturation at 98°C for 20 sec, annealing and extension at 68°C for 1 min, 40 sec and final extension of 5 min at 72°C.

3.5.3 Whole Genome Amplification kit from Sigma-Aldrich

The genome wide amplification using Genome Plex WGA1 kit (Sigma-Aldrich) was performed according to manufactures instruction. 10ng of genomic MCF7 DNA, without previous treatment, served as template for these reactions.

3.6 Control PCR

To ascertain a complete methylation-sensitive restriction digestion we performed a PCR in a multiplex manner, where differentially methylated DNA fragments of the digested DNA samples served as template for PCR amplification (Weinhaeusel *et al.*, 2008). In a total

volume of 20µl a fraction of 13,6µl water plus 1,66µl 2mM dNTPs (Fermentas), plus 1,04µl DMSO (Sigma-Aldrich), 0,5µl Primermix (table2), 2µl 10x Hotstart-Taq-PCR buffer (Quiagen, Hilden, Germany), 0,125µl Hot star Taq (5u/µl, Quiagen, Hilden, Germany) and 1µl sample were mixed and used for amplification. The PCR was 15 min at 95°C followed by 40 cycles of denaturation at 95°C for 30 sec, annealing and extension at 68°C for 1 min, 20 sec and a final extension of 7 min at 72°C. The 2mM dNTP solutions were obtained by mixing using 20µl of each dNTP (dATP, dCTP, dGTP, dTTP 100mM, Fermentas) combined with 920µl of water.

Table 2. Primer list multiplex control PCR

Name	sequences	prod. Length	cut sites
PitXa	fw 5´ - TCCCCGCTGCCCTGGCGCTC - 3´	315bp	
	rev 5´ - GCTGCGCGGCTGGCGATCCA – 3´		14
contrA	fw 5' – TGGGCAGGGGAGGGGAGTGCTTGA –3'	240bp	
	rev 5' – TGACCCCTGGCACATCAGGAAAGGGC –3'		0
XIST	fw 5´-TGCGGCAAGCCCGCCATGATG - 3´	200bp	
	rev 5´- GCATGGTGGTGGACATGTGCGGTC – 3´		4
contrB	fw 5' – TCACAGAGCCAGGCAAGCATGGGTGA –3'	175bp	
	rev 5´– GCGAGCCTGCTGCTCCTCTGGCACC – 3'		0
RB1	fw 5' – AGCGTCCCAGCCCGCGCACC –3'	138bp	
	rev 5' – CCCCGCCCGGCAACTGAGCG – 3'		9
SNRP	fw 5' - CGCTCGCATTGGGGCGCGTC - 3'	124bp	
	rev 5' – TGCCCACTGCGGTTACCCCGCAT –3'		3

All primer pairs were used in one reaction, where we applied appropriate amounts of a 100µM primer stock solution to reach a concentration of 0, 3µM per primer pair.
"Cutsites" refers to the number of recognition sites per fragment including all four methylsensitve restricion enzymes (table 1) used in our experimental setup.

3.7 qPCR tests

All qPCR reactions were performed in 384 well format on a Roche Light Cycler 480. A reaction volume of 10µl contained 0,125µM of each Primer, 0,3U of Hotstart Taq (Qiagen),

5% of DMSO containing Sybr green to reach a dilution of 0,5x in the reaction and 166µM dNTP-mix. 2µl of silica membrane based serum-DNA isolates served as template per qPCR reaction, whereas the qPCR reaction of serum or plasma isolates processed with MBD was performed with 1µl. In both cases about a 0,5ng to 1ng was used per reaction. Ct calls and melt curves were analysed using the Roche Light cycler 480 software, to guarantee that only Ct-calls of PCR product with an appropriate melt curve were used for the comparison of the Ct-calls (section: Results; MBD based isolation; figure 8).

The PCR program was identical to the multiplex PCR program with the exception that for qPCR analyses 50 cycles were performed. Table 3. Primer list Enrichment of methylated DNA fragments

Gene-Symbol	sequence	prod. length	methylated
SALL3	fw 5' – GGCGGGGAAGGCGACCGCAG –3' rev 5' – TCCCCGGGCGGCCATTAGGCA – 3'	212bp	no
ESR1	fw 5' – CCGCCGCCAACGCGCAGGTC – 3' rev 5' – GCCGGCCTCGCGCACCGTGT – 3'	239bp	no
CHFR	fw 5' – GCGGTCCGCGAGTGGGAGCG –3' rev 5' – TGCAGACATTGGCGCGTTCCTCCA – 3'	276bp	no
ZNF502	fw 5'- GGCCCCAGTCCACCTCTGGGAGCG –3' rev 5' – GCCCTACGTCCGGGCAGCACGC – 3'	207bp	yes

Primer sequences used in qPCR with MBD and silica membrane based serum DNA isolates for evaluation of MBD based DNA extraction approach (figure 8). The methylation status of the DNA fragments was ascertained on UCSC genome browser (http://genome.ucsc.edu/) applying the feature „HAIB Methyl-seq" and „HAIB Methyl27" to display the methylation status of diverse human cell lines.

3.8 Agarose gels

DNA electrophoreses was done using a 2% agarose gel (Sigma-Aldrich) in 1 x TBE containing 0,5 µg/ml ethidiumbromide (Merk). We used 6 x DNA loading dye (Fermentas) and Gene Ruler 100bp DNA Ladder Plus. (Fermentas) To achieve a separation of the PCR-product bands and DNA fragments, gels were run at 155 V/ ~20cm was applied for 75 minutes.

3.9 Multiplex PCR of 360 target DNA fragments

Dr. Andreas Weinhäusel at the Austrian Institute of Technology developed the multiplex PCR and the following detection on a target specific micro array. The marker selections based on published work of (Widschwendter *et al.*, 2007), (Shames *et al.*, 2006). A collectivity of 360 target PCR amplifications was divided into 16 reactions for 20-24 amplification products each. Thus each DNA sample was divided into 16 equal aliquots with 25ng input DNA per distinct multiplex PCR reaction, which enabled us to screen 360 target regions for elucidating DNA methylation markers. By the use of biotinylated reverse primers (100µM, Microsynth) the detection with streptavidin cy3 conjugates (Caltag Laboratories) on the target specific micro array was enabled.

One reaction with 16-24 primer pairs, which had a concentration of 0,25µM per primer in a reaction volume of 11,5µl, consisted of 6,29µl water, 1,15µl reaction buffer (10x, PCR buffer, Quiagen), 0,92µl 2mM dNTP (100mM, Fermentas), 0,58µl DMSO (Sigma-Aldrich), 2,5µl primermix, 0,07µl Hot FirePol (Solis Biodyne) and 0,812µl sample. The complexity of this procedure made it necessary to create a mastermix of these reagents without sample and primers. Hence a volume of 167µl mastermix (i.e. 20 times one reaction) was combined with 13µl sample (i.e. 400ng), for which one mastermix was created for all samples and controls and then divided into 167µl aliquots. This mixture of PCR reagents and DNA sample then was aliquoted into 16 vials with 9µl each, followed by addition of 2,5µl primermix (consisting of 16-24 primer pairs), using an eppendorf epMotion 5075-pipette roboter. The PCRs of 24 samples were carried out in 384 well plates. The PCR was 15 min at 95°C followed by 40 cycles of denaturation at 95°C for 40 sec, annealing 40 sec at 65°C, extension at 72°C for 80 sec and final extension of 7 min at 72°C. After amplification all 16 multiplex PCR reactions per sample were pooled.

3.10 Hybridization and Detection of multiplex PCR products

For the detection of the possible 360 PCR products we used a micro array, which was spotted at the AIT with an omnigrid machine (a1 Biotech) At this procedure 0,6nl of about 60nt long oligos with a concentration of 10µM, where each DNA oligo was complementary to one PCR

product, was spotted in triplicates on aldehyde coated glass slides (CEL Associates, VSS 25 Silycated Slides Aldehyde). Each spot had a diameter of 100µm. One micro array consisted of 360 different PCR product specific probes and several controls in triplicates. Three micro arrays were placed on one glass slide.

3.10.1 Sample preparation

The pooled PCR products (approximately 600ng DNA) were combined with hybridization buffer in a ratio of 1:1 to an end volume of 110µl. The hybridisation buffer was pre-warmed to assure a complete dissolving of all components. The hybridization buffer consisted of 7x SSC (20x, Invitrogen), 0,6% SDS (20%, Sigma-Aldrich) and 50% formamide (100%, Merk). To create single stranded DNA for the hybridization, the solution was incubated at 95°C for 10 min and held on 72°C until transfer onto the micro arrays.

3.14.2 Slides and hybridization

Prior to hybridization the micro arrays were blocked for 30 min with a blocking solution containing 3M Urea (Merk), 0,1% SDS (20%, Sigma Aldrich) and water. Followed by a washing step for 5 minutes in wash solution II, which includes 0,1 x SSC (20x, Invitrogen) and 0,2% SDS (20% Sigma-Aldrich) in water. Slides were dried via centrifugation at 900 x g for 1 minute.

Before applying the denatured DNA in 1 x hybridisation buffer onto the micro arrays, the slides and gasket chambers (Agilent) were pre-warmed to 72°C on a heating plate. Hybridization chambers (SureHyb, Agilent) were pre-warmed to 52°C in the DNA micro array hybridization oven (Agilent), which also assures a preheating of the hybridization oven to the required hybridization temperature of 52°C. The handling step for micro array hybridization was as follows: First we put a gasket chamber onto the hybridization chamber. Then one sample was loaded onto each of the three fields framed with a sealing ring. One glass slide containing 3 micro arrays, which were spotted in way that each array fits exactly into one sealing ring framed space, were put onto the gasket slide, which was loaded with 110µl of denatured DNA in 1 x hybridisation buffer. The hybridization chamber was assembled and the thumbscrew of the chambers was tightened. Finally it was checked if there was only one air bubble per micro array, which can move around freely to mix the sample during the hybridization. The hybridization chambers were incubated in a rotary oven (DNA micro array hybridization oven, Agilent) for 16 hours at 52°C. After hybridization chambers

were disassembled and the gasket slides separated form the micro arrays in wash I solution 1 1x SSC (20x, Invitrogen) and 0,2% SDS (20% Sigma-Aldrich). The gasket slides were washed for further usage and the wash I solution discarded. The micro arrays were placed in fresh wash solution I and washed for 5 min followed by 5 minutes in wash II and repeated dipping in wash III. Washings were performed in a colpin char, which was mixed by stirring. The slides were dried by centrifugation 1 min at 1000 x g

3.14.3 Detection

Detection of biotinylated hybridization products was obtained by streptavidin cy3 conjugate (Caltag Laboratories). The streptavidin cy3 conjugate was diluted 1: 400 in 1 x PBST buffer containing 1 x PBS plus 0,1% Tween (Sigma-Aldrich) and 1% milk powder bovine (Sucofin). 80µl of this solution was pipetted onto a cover slip (24 x 60mm Marienfeld). The washed array then was carefully placed narrow above the cover slips in a way that the cover slips were attached to the glass slide via capillary forces of the PBST solution. The arrays were incubated with the detection solution for 30 min at room temperature protected from light. After incubation cover slips were removed from arrays in 1x PBST followed by a 5 min washing step in 1 x PBST and repeated dipping in MilliQ water. Slides were dried by centrifugation 1000 x g for 1 min and scanned on Gene Pix 4000A scanner (Axon Instruments).Cy3 signals were measured the micro arrays via exication at 532nm and emission was read at 560nm with a resolution of 10µm at a photomultiplier tube setting of 700PMT.

4. Results:

4.1 Optimization of cf serum DNA Isolation and genome wide enrichment of methylated DNA

4.1.1 Isolation of cell free DNA from Serum

To figure out the best serum DNA isolation strategy we compared four different approaches to isolate the cell free trace DNA amounts prevailing in human blood serum. For this test, sera of three healthy volunteers were isolated in triplicates in three independent experiments, at which every value of each triplicate was used for creation of the box plot and the calculation of medians and standard deviation.

The ZR Serum DNA kit from Zymo, where the cell free DNA of the serum samples should bind to silica beads, did not work in our hands. Using the Blood midi kit form Qiagen, used widely in the respective literature for serum DNA preparation, we could obtain 7,51 ng DNA+/- 2,82ng (mean +/- s.d.) per ml of serum. Although the kit was easier to handle than the other silica membrane based isolation kit (Roche), the difference between the isolated DNA amounts was too big to legitimate the use of this less laborious kit.

The MBD based isolation strategy, established during practical work before starting this master thesis resulted in a total DNA amount of 2,06 ng DNA +/-1,17 (mean +/-s.d.) per one ml serum.

The optimized Roche isolation protocol resulted in 10. 31ng/ml +/- 4.13ng (mean +/- s.d.), which was the highest DNA level isolated from 1 ml serum in this experimental setup. To reach a reproducible high amount of extracted serum DNA the volumes of the aliquots per sample as well as the incubation times and temperatures were varied until the highest reproducible DNA yield was gained.

These facts finally lead to the decision to use the Roche high pure template prepation kit to isolate all clinical serum samples from breast cancer patients and accordant normal controls to perform a methylation screen.

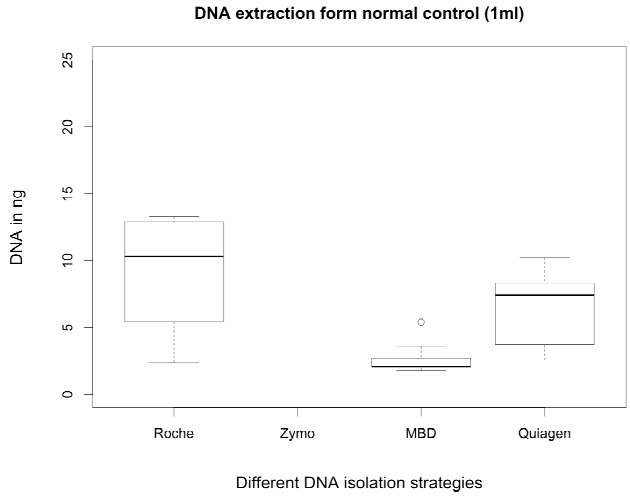


Fig.7: Boxplot of DNA amounts isolated from 1ml serum using different DNA purification kits or MBD based DNA isolation strategy. The sera of three healthy controls were isolated in triplicates. The DNA levels were measured with Pico Green.

4.1.2 Evaluation of (silica and MBD based) DNA isolations for methylation testing

At the analyses of the data obtained from different isolation strategies (figure 7) we addressed the question, why DNA levels prevailing in 1ml of serum were significantly lower using the MBD isolation approach than applying the membrane-based protocols (i.e. Quiagen or Roche). Therefore we hypothesized that this effect might be due to the higher affinity of the MBD protein for methylated DNA.

In this tests it turned out, that applying the MBD for serum DNA isolation, on average 26,1% (P=0.0002 by student t-test) of the DNA levels gained with Roche High pure template preparation kit could be obtained from 1ml serum. The reason for this effect was investigated by qPCR tests on serum isolates using primers for the gene loci SALL3, HIC1, ESR1 and ZNF502. These tests were performed in an independent test series using sera of 12 healthy volunteers (age and sex matched).

The differences between CT-values highlight the reduced amount of DNA, isolated with MBD except for ZNF502, which is methylated in healthy adults. Hence, there is no significant difference between the CT-values (i.e. DNA amounts) using the primer for ZNF502 for both extraction methods (Figure 8). We determined an average difference for CT-calls at unmethylated DNA fragments (i.e. SALL, ESR1, CHFR) of 2,81 by subtracting the CT calls form MBD isolated DNA from Roche isolated DNA. Doing the same calculation for methylated DNA fragments (i.e. ZNF502) a value of 0,09 was computed. These values were converted into percent values, which lead to the suggestion that in comparison to the overall

membrane based isolated DNA, 95% of methylated DNA fragments and 10% of unmethylated DNA fragments were isolated using the MBD procedure.

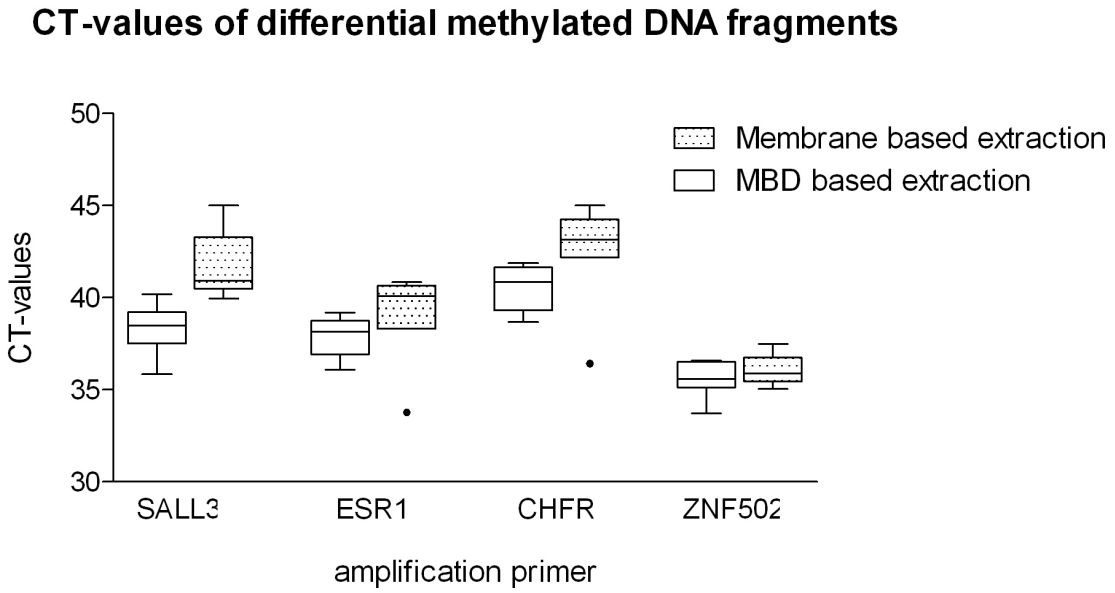


Fig.8: Box plot of normal control DNA (n=12) isolated with MBD workflow or silica membrane based using primer for differential methylated DNA fragments The methylation status of the DNA fragments was ascertained on UCSC genome browser (http://genome.ucsc.edu/), whereas the amplified gene loci of SALL3, ESR1 and CHFR are unmethylated and the gene region amplified with ZNF502 is supposed to be heavily methylated in normal tissue.

This significant enrichment of methylated DNA fitted perfect into our research field, where differences between methylation patterns of healthy persons versus breast cancer patients are shown, but the reduced amount of isolated DNA upon MBD-isolation (26% compared to Roche) made DNA manipulation and genome wide amplification very difficult.

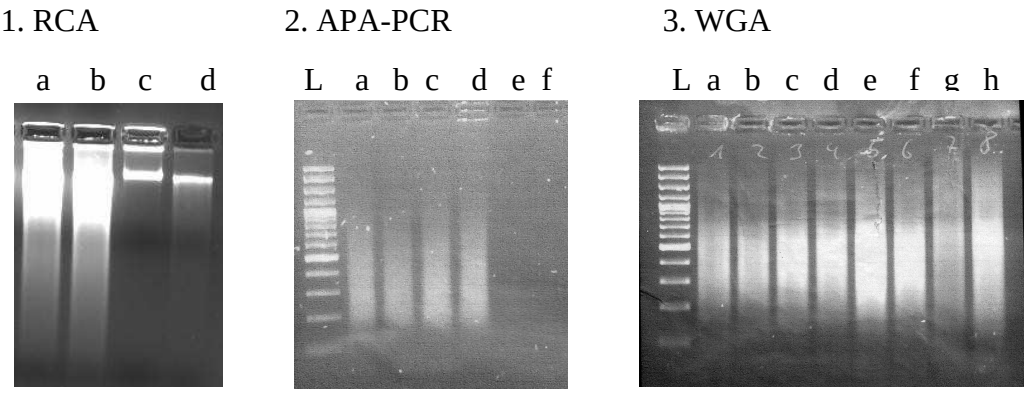
Based on several control PCRs results (data not shown) the MBD approach could not be used for accurate methylation studies without a methyl sensitive restriction digestion, which was due to the fact that a significant amount (10%) of the unmethylated DNA fraction was isolated from serum. In several experimental series for enrichment of the methylated DNA fraction, results were not consistent and enrichment was not only dependant on the methylation content when several DMRs were analysed, but rather being influenced by the neighbouring surrounding sequence context.

Therefore this challenging method, where we started with the expression of the MBD protein in E.coli, purified the protein upon binding to nickel beads and then used these beads for serum DNA isolation was not implemented into our standard sample processing procedure.

4.1.3 Comparison of genome wide amplification systems

When starting with trace DNA amounts prevailing in serum (figure 7, figure 13) the genome-wide amplification of DNA samples became a key process of this work, providing sufficient amounts of DNA for methylation screenings. Therefore three different amplification systems were compared according their DNA yield and recovery of the methylation pattern after amplification, which was achieved by control PCRs.

A



B

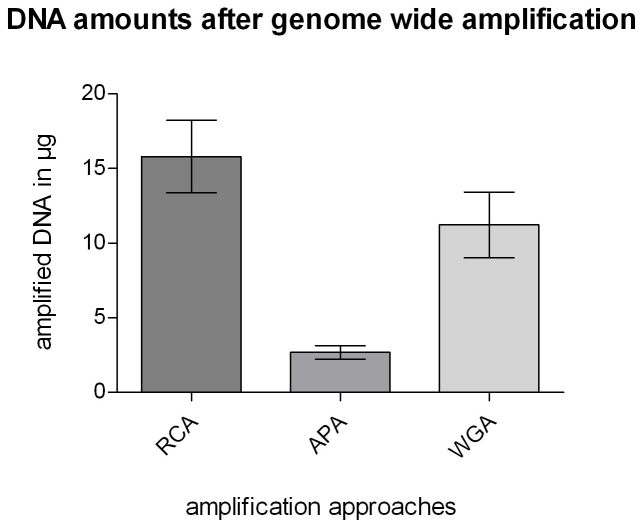


Fig.9: (A) representative Gel images of whole genome amplification products using different approaches. In case of RCA (1) 2µl of the reaction volume were loaded onto the gel; for APA-PCR (2) and WGA (Sigma-Aldrich) (3) 10µl of amplification reaction volume were loaded. (B) Means of DNA amounts yielded after various amplification reactions; error bars indicate the standard derivation. As input for each reaction served 20ng of genomic DNA.

As template for these reaction served 50ng of genomic MCF7 DNA, which was treated according to the requirements of each method. These gel images gave first hints on the functioning and the performance of the diverse amplification methods.

In contrast to PCR based adapter-mediated amplification (APA-PCR and the WGA from Sigma Aldrich), the RCA products are of high molecular weight. This is due to the amplification process, where circles in a range from 100bp to 1,5kb produced huge DNA molecules with consecutive repeats of the template circle. This isothermal reaction process resulted in high molecular weight DNA, which was observed in the region of 10kb and above.

The PCR based adaptor mediated amplification resulted in amplified DNA fragments in a region of 200 to 800bp, a pattern that could be determined as typical for common PCR reactions, which highlights the basically differing process of the RCA to the PCR-based methods.

The purified amplification reactions were measured for DNA content with Nanodrop. We observed big differences between these methods. Highest DNA amounts were measured after rolling circle amplification (RCA), where 15,7µg +/- 2,42µg (mean +/- s.d.) were detected. In case of the RCA the water controls (Figure 9, image A1, lane *c* and *d*) of the reaction showed products on the gel and we could detect 2,9µg DNA.

The APA-PCR produced 2,68µg +/- 0,45µg (mean +/- s.d.) DNA and this amplification method showed a clear gel image without any contamination or products of side reactions, (Figure 9, image A2, lane *e* and *f*). On contrary to the WGA from Sigma, which showed throughout contaminated water controls (Figure 9, image A3, lane *g* and *h*), amounts of 4,5µg and more DNA were detected in these controls. These high amounts of unspecific amplified DNA or unknown side reactions lead to the exclusion of the WGA kit from further testing, although the DNA yield of 11,83µg +/- 2,2µg (mean +/- s.d.) was sufficient (Figure 9).

To test whether the genomic DNA was amplified entirely or some parts of the genomic information got lost during the amplification process, we performed control PCRs on amplification products. Therefore 1µl of the purified amplification reaction served as template for the PCR reactions.

The gel images (Figure 10) clarified, that information losses with both amplification methods, were existent. The positive controls, where unamplified MCF7 DNA served as template, showed 6 fragments (PitXa 315bp; contrA 240bp; Xist 200bp; contrB 175bp; RB1 138bp; SNRP 124bp). At the second positive control one fragment (RB1) got lost upon methyl

sensitive restriction digestion, but the remaining fragments should show up as product of each PCR reaction.

This was not the case for neither of the two approaches for genome wide enrichment of methylated DNA, but apparently the RCA products contained more amplifiable fragments for succeeding PCR reactions.

We concluded from these observations that the rolling circle amplification had the disadvantage of contaminated negative controls, but had two big advantages over the APA-PCR. On the one hand about six times higher DNA yields were observed, and on the other hand a more reliable representation of the methylation-pattern after the amplification step was achieved.

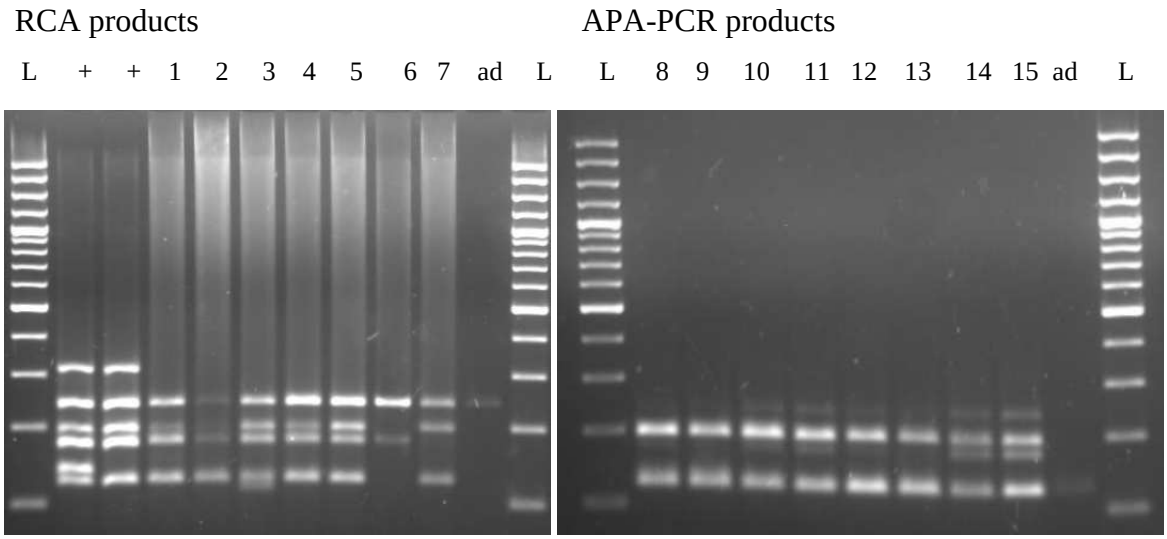


Fig.10: Control PCR with genome wide amplified DNA to assume the DNA recovery. L = ladder; + = positive control: first lane undigested 2nd lane methyl sensitive digested MCF7 DNA both without previous amplification. Lanes 1 – 15 methyl sensitive digested MCF7 DNA, which served as template for each genome wide amplification strategy, ad = water control of amplification reaction

4.2 Sample preparation for methylation screening in serum of breast cancer patients

4.2.1 Serum processing workflow

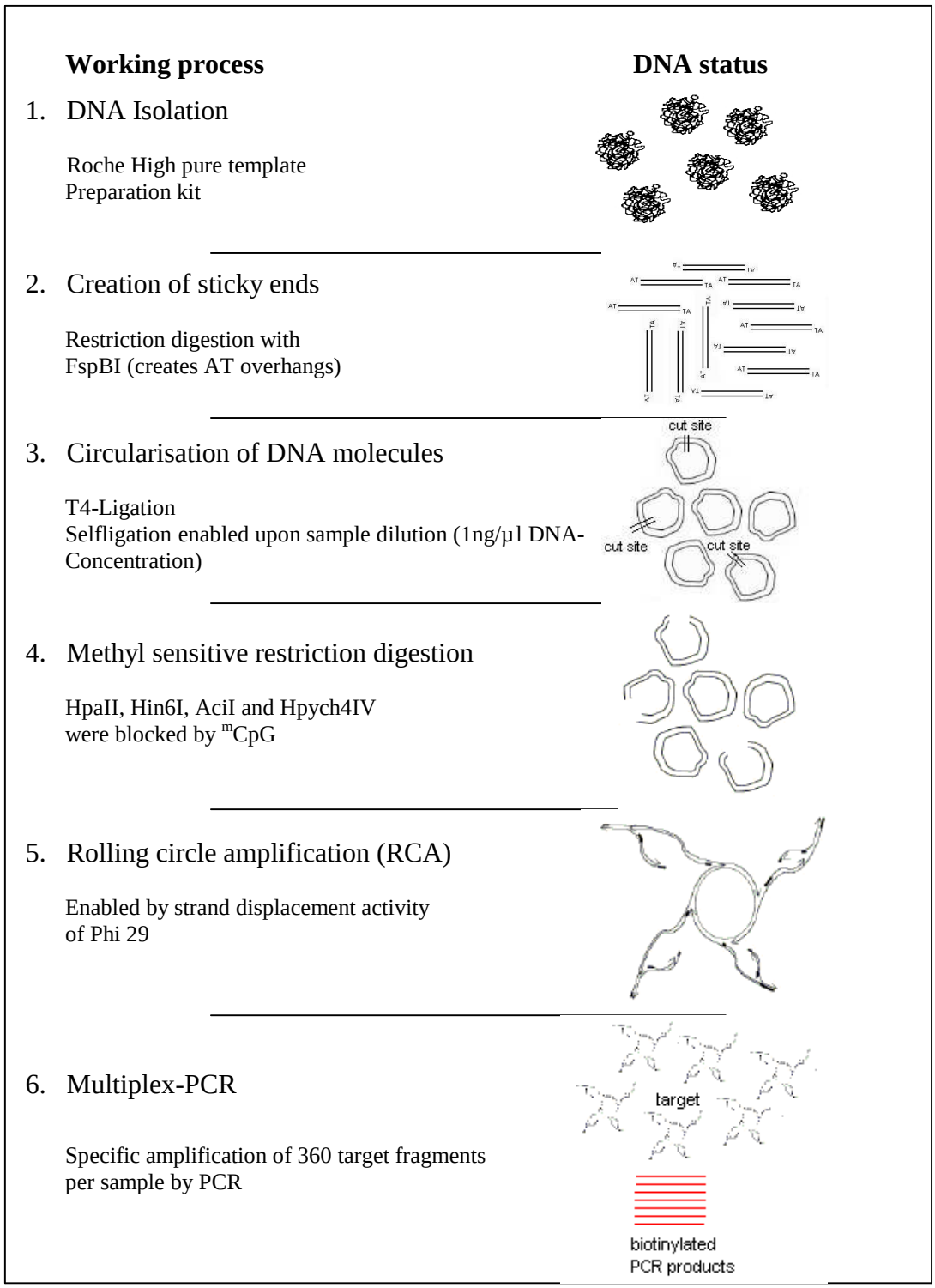


Fig.11: Serum processing workflow. Images for this figure were adapted form the publications of (Dean *et al.*, 2002) and (Wang *et al.*, 2004) and modified as required.

The work with clinical samples and normal controls followed a strict processing scheme. The samples were randomized after the DNA isolation process and this random order kept until hybridization on the micro arrays. All samples and controls described in the following part were isolated and further processed in one experimental run to minimize experimental variances.

The collective of 96 samples (i.e. clinical samples, normal controls and assay controls) were treated in the way outlined in the serum processing workflow (Figure 11). Control PCRs were performed before and after rolling circle amplification, as well as diverse argarose-gels after control PCRs, RCA and Multiplex-PCR (not shown in the workflow graph).

4.2.2 Patients information and study groups

Overview basic patient information

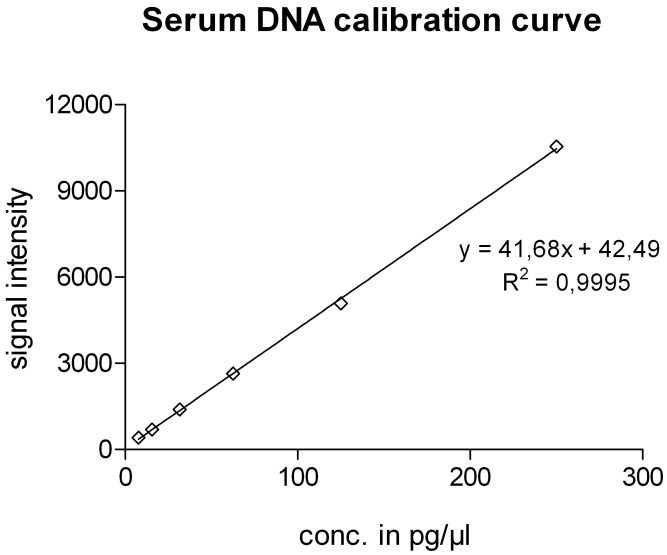
Study group	histological finding				estrogenic Rec. pos	pN >0
	G1	G2	G3	other		
M1		1	3	2	3	4
Relapse	2	3			3	2
G3			6		6	2
G2		6			5	3
G1	6				6	2
C.I.S	1	4	2		6	0
Benign			12		0	0
BRCA1/2					0	0

Tab.4: The histological grading (G1-G3) refers to the affection of tumor surrounding tissue of the invasive lobular or ductal tumors. Estrogenic receptor tests are included in standard breast cancer diagnoses process and serve as additional indication. pN staging ≥ 0 shows a local affection of lymph nodes.

The investigated samples were selected from a collective of clinical samples obtained at initial diagnoses and provided by Prof. T. Wagne from AKH Vienna. On the basis of supplied clinical data form the hospital the patient samples were divided into test groups, to allow addressing of distinct questions, which were expected to be answered upon this test series.
The 8 study groups were then named according to their clinical representation. M1 samples were derived form patient with a metastasizing tumor. The relapse group resembled patient sera, where the malign neoplasm raised again in a time period of 5 years after the first treatment. The group identifier G1-G3 was according to the histological grading.
C.I.S. indicates a carcinoma in situ, where an epithelial tumor was diagnosed, without growing invasively. The group of benign tumor samples was collected from patients with fibro adenomas, which did not affect-surrounding tissue. The term BRCA1 or 2 refers to patients with hereditary breast cancer bearing a germ-line-mutation in one of these two genes.

4.2.3 Measurement of trace DNA amounts

For measurement of small DNA amounts like prevailing in blood serum we developed an adjusted protocol of the “Quant it” Pico Green kit from Invitrogen. Therefore we reduced the assay volume from 1ml to 100µl, which enabled us to measure the DNA concentration of 5µl sample diluted in 95µl working solution. Then the values were calculated for the entire sample volume of 55µl.
The lambda DNA dilutions, serving as standard, were prepared in a volume of 50µl and ranged from 7,8pg/µl to 250pg/µl (Figure 12). This principle enables exact DNA concentration measurements of the low amounts of DNA extracted from the serum samples



Conc. in pg/µl	Value
250	10543
125	5080
62,5	2642
31,25	1395
15,6	699
7,8	419

Fig.10: DNA standard curve for Pico Green-DNA measurements: All reagents including the lambda-DNA dilution of the assay were freshly prepared for every measurement.

4.2.4 Cell free DNA derived form patients and controls

DNA from serum samples (800µl) was isolated using the optimized Roche protocol. Figure 13 shows an increased level of serum DNA in breast cancer patients with a metastasizing etiopathology, which illustrated one of the key findings of this work. We detected 45,64 ng/ml +/- 32,31ng (mean +/- s.d.) in the serum of patients with a metastasizing disease in contrast to 8 ng/ml +/- 5,48ng (mean +/- s.d.) for patient with a solid tumor of grading G3.

9,86 ng/ml +/- 11,4ng (mean +/- s.d.) were detected for benign cancers and 10,58 ng/ml +/- 9,71ng (mean +/- s.d.) for normal controls. These DNA concentrations are shown in the box plots, which enhanced the hypothesis that there is only an increase of serum DNA from patients with a metastasizing disease.

As reported in current literature about 60% of patients with metastasizing cancer show a significant increase of cell free serum DNA. This was the case in our study too, but the height of the 60% of serum isolates with increased DNA levels lead to a significant increase of the entire group of M1 patients. Hence the conclusion out of these observations was a significant increase of serum DNA levels only for the M1 study group.

Further there was no increased DNA level in study group G1-G3 or in the serum of patients with benign cancers as claimed in some recent reports.

A slight higher median of 23,46 ng/µl +/- 13,48ng (median +/- s.d.) was observed in the group of patients with a new arising of their cancer (i.e. Relapse). The reason for this observation was not cleared entirely, because of absent patient information a few years after blood donation. We do not know if some of these patients developed neoplasm on other sites of their body. Anyway we could not distinguish between these two groups (i.e. M1 and Relapse) because of a P-value of P = 0,45 (by Wilcox rang sum test).

Hence we excluded this sample cohort for the calculation of P-values. Between study group M1 and the other groups the DNA amounts differed significantly by a P value of 0,0013 (by Wilcox rang sum test). Further the P value was computed for discrimination between the M1 group and each study group, whereupon P < 0,005 (Wilcox rang sum test) were observed.

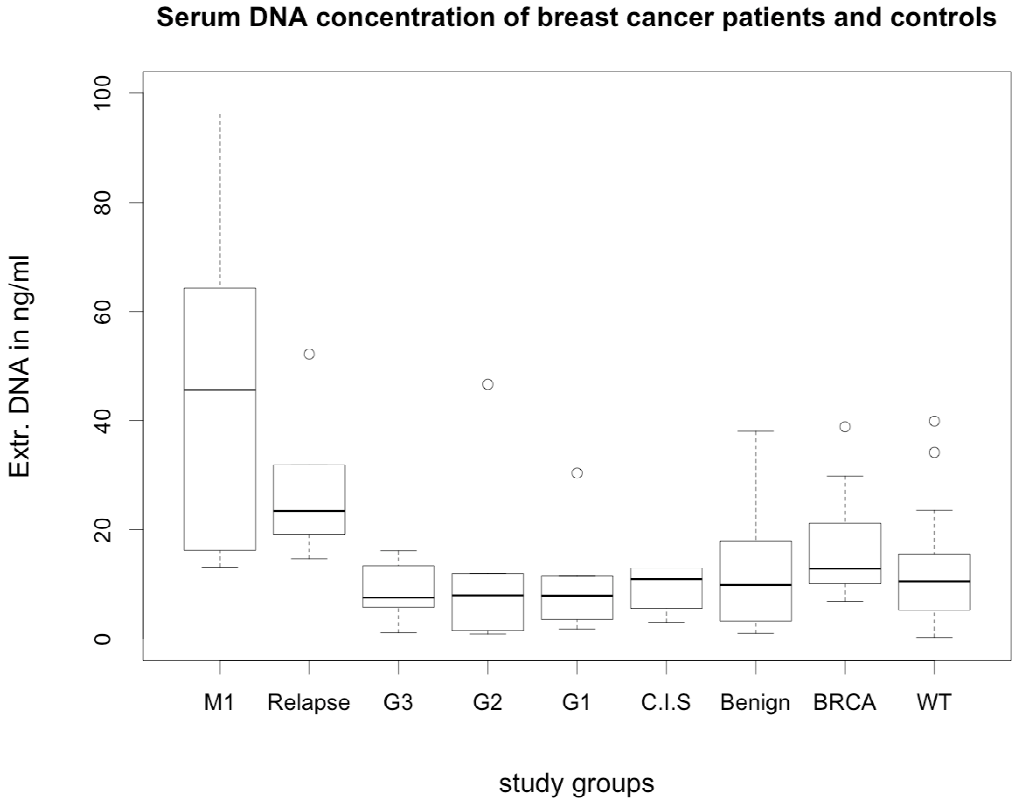


Fig.11: Box plots of extracted DNA per 1ml human blood serum. Each study group consisted of at least 6 serum samples or more. The abbreviations (i.e. M1, G1-3, etc) refer to the study groups as outlined in the section patient information and study groups (table 4). WT applies to the group of normal controls (n=24).

4.2.5 Evaluation of sample quality and methylation sensitive restriction digestion

Ancillary to concentration measurements and separation by gel electrophoreses to control amplification, the control PCRs were the main measures to monitor the DNA samples during manipulation and amplification in our experimental procedure. The first control PCR in our assay was performed after methyl sensitive restriction digestion(figure 11, step 4).

On basis of the number of amplified PCR fragments we could evaluate the sample quality. So the more fragments per sample were detected, the better was the sample quality and the more complete was the genomic DNA of each sample.

The second information provided by this PCR, gave insight into the restriction digestion. The two positive controls in Figure 14 represent a complete methylation sensitive restriction digestion. First lane displays an undigested PCR-reaction of normal DNA, the second lane a digested one, at which the fragments with a length of 315bp (i.e. PitXa) and a length of 138bp (i.e. RB1) are not present, because they are known to be unmethylated in healthy blood DNA.

Therefore complete digested samples had to show four fragments as the second positive control. In this case it was assured that the restriction digestion was complete and the sample quality was high. Figure 14 shows that all samples shown here had underwent a complete methyl sensitive restriction digestion and all remaining possible fragments were amplified.

In Gel 1 (lane *a-p*) amplicons of clinical serum samples of any kind were loaded (e.g. *a* refers to sample P300 of the Relapse group or *p* refers to sample WT181 of the normal control group), at which 1µl of extracted DNA served as template for the PCR reaction. From all samples the expected number of fragments were amplified. At lane *h* and *q* water controls were loaded and in lane *m* an operator mistake occurred and the sample was loaded again onto another gel.

Gel 2 (right section of figure 14) shows two water controls, in this case the two water controls (i.e. *d* and *p*) passed through all steps like the clinical DNA samples. The last lane was again reserved for the water control of the PCR reaction.

A close view on gel 2 (right section of figure 14) reveals an additional fragment of the samples *j-o*. These PCR products originated from assay positive controls, where 50ng of MCF7 DNA was processed. These controls were treated parallel to the clinical samples. MCF7 DNA is derived from a breast cancer cell line, at which the PitX2 locus is methylated. Hence there was no digestion of the PitX-DNA fragment by the methyl sensitive restriction digestion.

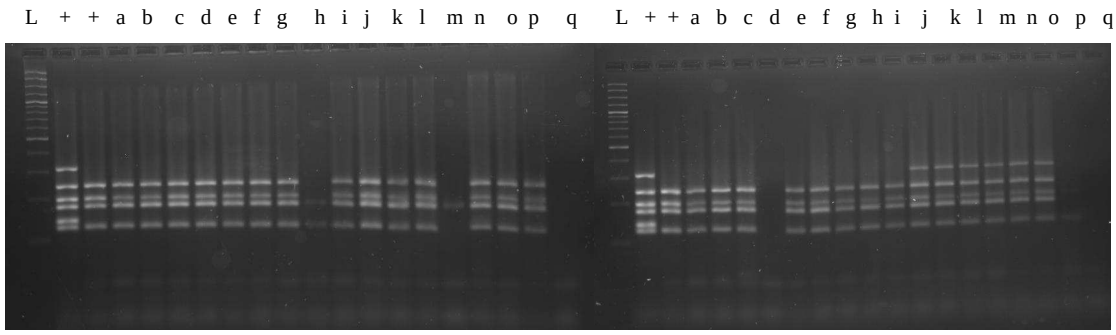


Fig.14: Chosen control PCRs of serum isolates and controls after ligation reaction and before genome wide amplification. L refers to the 100bp-ladder. The lanes marked with + are positive controls, at which the second is methyl sensitive digested. The abbreviations a-q of each gel refers to samples; assay controls (right section: j-o) and water controls (left section: h, q; right section d, p, q). PCR products: PitX (314bp); contrA (240bp); Xist (196bp); contrB (174bp); RB (138bp); SNRP (124bp).

4.2.6 Rolling circle amplification

An accurate executed RCA reaction was ascertained by two independent controls. Therefore directly after the reaction an aliquot of 1µl (approx. 500ng) was loaded onto a gel. As outlined in the pre-test section of this work and shown in figure 15, the reaction created DNA molecules of high molecular weight. The image also indicates the DNA amounts yielded by this reaction, which were about 15µg for each 30µlreaction.

Figure 15 shows representative RCA-amplicons, for example, *c* refers to sample P23 of the study group M1 and *m* refers to WT212 of the normal control group. The RCA product of a water control was loaded onto lane *i*. Here one could assume large DNA amounts produced by the reaction, which was confirmed upon Nanodrop measurement, where we determined 7.9µg in the whole reaction, which was previously purified as outlined in the methods section. Despite this high DNA concentration a significant contamination of this water control was not detected in the second control PCR (Figure 16: left section, sample *o*).

In lane *o* (figure 15) a normal control sample was loaded. In this case the expected small DNA amount was not verified on Nanodrop, where 18,5 µg DNA was determined in the purified reaction plus a positive result of the second control PCR.

The box plots section in Figure 15 gives an overview of the DNA yield after genome wide amplification. We expected same DNA amounts for all study groups and samples, which was not the case in our experimental setting.

The plot represents a cohort of samples where the RCA worked well, which was observed in study group M1, Relapse, BRCA and WT.

So we assumed throughout complete RCA reactions for these samples. The fact that some researchers reported that they could amplify only every second serum DNA isolate upon RCA, makes a definite analyses of the diverse DNA amounts and their associated serum samples difficult. Although 10% of the samples showed an incomplete RCA reaction, analysed upon concentration measurements, gel images and multiplex-PCR yield, all samples were further processed.

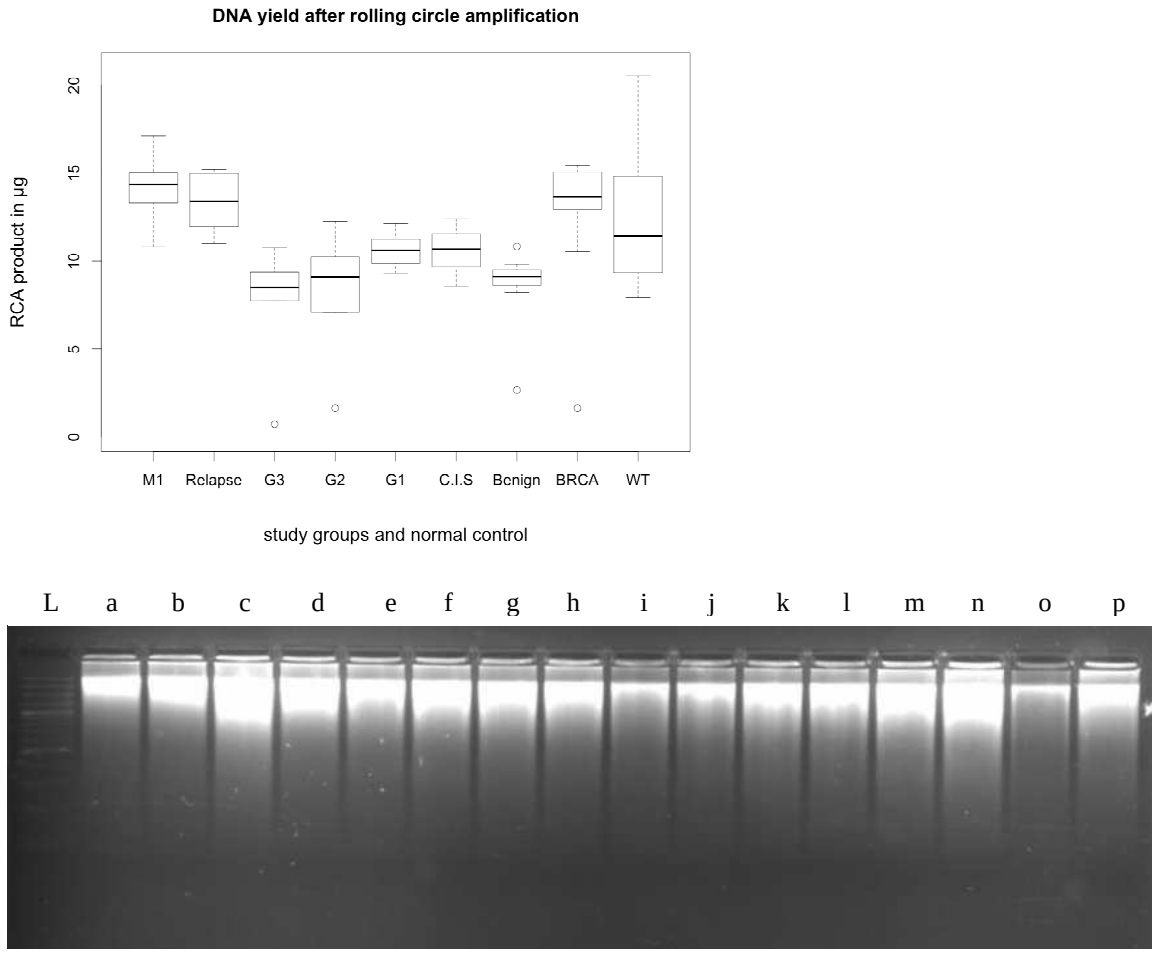


Fig.15: Rolling circle amplification. Box plot of obtained DNA amounts after RCA of the different breast cancer study groups. Beneath a representative gel image of 1µl rolling circle reaction loaded onto a 2% agarose-gel. In the lane marked with L 100bp ladder was loaded. The labels *a-p* refers to diverse samples and controls.

4.2.7 Evaluation of enrichment of methylated DNA upon RCA

In order to get an impression whether the enrichment of methylated DNA was successful or not we performed control-PCRs on the RCA amplicons using the same 6 primer pairs as for the first control-PCR (figure 14).

These representative gel images follow the same pattern as the gels loaded with amplicons of the first control PCR (figure14). The ladder is followed by two positive controls, at which the first shows untreated genomic normal control DNA and second PCR was performed with a methyl sensitive digested DNA template, both controls did not undergo the RCA.

In the left section of figure 16 clinical samples are shown except for sample *a*, which is genomic unamplified MCF7 DNA as well as sample *c* in the right section of figure 16, where the PCR reaction did not work. Both samples served as positive control of the following multiplex PCR.

The control PCR water controls were loaded onto lane *p* in the left section and the right section. Another water control, which was included into the assay at the initial isolation process, was loaded onto lane *o* in left section of figure 16.

Provided that the control PCR should result in 4 PCR products and bearing in mind that complex genomic/chromosomal aberrations are common to cancer cells of which the cell free serum DNA might have derived, we could assume that the whole genome amplification worked sufficiently well (e.g. right section, lanes *i*-*k*).

We observed that the RCA did not produce such sufficient results for all samples. Many samples like *f* and *g* in the right section of figure 16 were amplified incomplete, which could be deduced from incomplete band patterns or too weak gel bands. This suggested that the templates for this PCR reaction were not equally amplified by the genome wide amplification.

The combination of these three gel images (Fig. 14, 15, 16), which were created for each processed sample, enabled us to evaluate the experimental course of every sample. For example one may track the sample WT213, which was loaded on left section in figure 14 in lane *n*, further in figure 15 in lane n and finally it was shown in Figure 16 in lane *l* on the right section.

For the example of WT213 the experimental course could be rated as satisfying, because the first control PCR suggested a complete methyl sensitive restriction digestion and a good sample quality. The control gel of the RCA suggested a huge amount of DNA, which was confirmed by a Nanodrop measurement of 18,64µg DNA. Finally control PCR 2 indicated an

almost complete amplification of the DNA isolated form serum, because three of four strong bands are visible on gel 2 (lane *l*).

This process could be followed for each sample in our assay and was used in combination with concentration measurements to confirm the accuracy of sample manipulation.

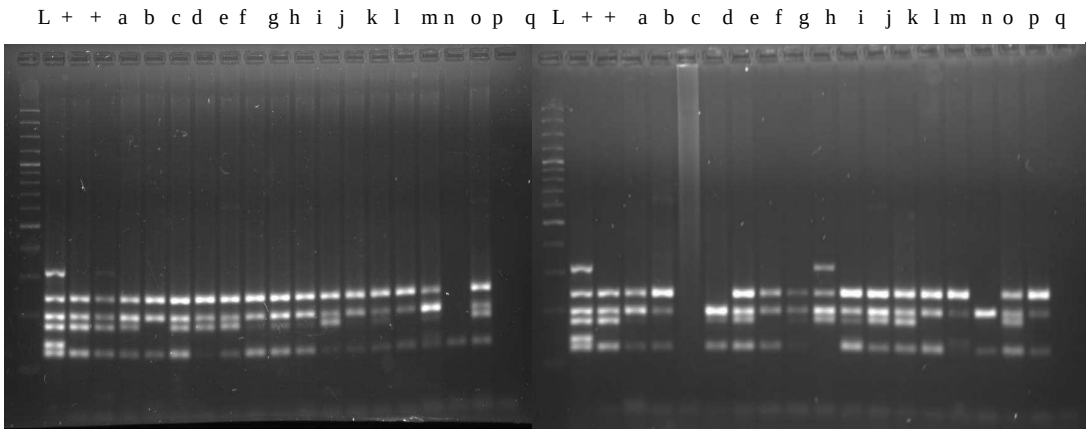


Fig. 16: Representative gel image of the second control PCR. The reaction was performed with genome wide amplified DNA samples. The abbreviations a-q of each gel refers to samples assay controls (left section: c; right section: h) and water controls (left section: o, q; right section c, q). PCR products: PitX (314bp); contrA (240bp); Xist (196bp); contrB (174bp); RB (138bp); SNRP (124bp).

4.2.8 Multiplex-PCR control

As outlined in the methods section we performed a multiplex-PCR consisting of 12 reactions with 16-24 primer pairs each to create a suitable amplicon for the target micro array. To assure an accurate activity of the PCR-reaction 10µl of pooled amplicon were loaded onto a 2% agarose gel.

Noticeable in Figure 17 are the strong blazing gel bands in a range smaller than 100bp, which originate from primer dimers of the multiplex PCR reaction. Primers are available in the PCR reaction in excess and therefore give such a huge gel band.

Thus 50µl of this unpurified reaction were used for micro array hybridization, as outlined in the materials and methods section of this work.

The weak bands in the range of 200bp indicate specific products derived from CpG360-PCR, which should amplify 360 targets. This would indicate that the RCA bare enrichment of methylated DNA.

The gel image analyses were also used for a last validation of the samples before micro array hybridization. In this connection we could control if the sample was amplified accurately and

had a closer view on samples that provided small DNA yields like on this gel sample P114 from sample group M1 (figure 17, lane *p*).

In this case, the course of the sample from first PCR (figure 14, sample *f*, left section) upon RCA (figure 15, sample *f*) and second PCR, where we detected three strong bands (data not shown), was promising, but the multiplex control gel seems to show very little products. Exclusion or any other final conclusion by means of the gel images in this section would not have been meaningful.

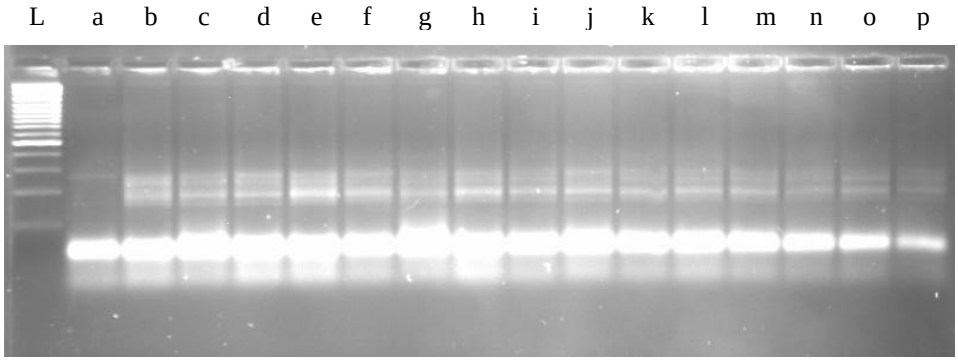


Fig.17: Gel image of the pooled CpG 360-multiplex PCR control gel. Pooled amplicons (10µl) of each sample were loaded. L refers to the 100bp ladder. In lane *a*n assay water control was loaded (Same control is shown in lane *o* in Fig.9 and Fig.10 gel 1), the rest of the lanes were loaded with amplicons of clinical samples.

4.2.9 Hybridization process

After micro array hybridisation and detection, an accurate function of these processes was assured via control of the scanned images (figure 18). On basis of the signal intensity of each spot a data file was created by the GenPix software. The values were further log2 transformed, a median was created out of the probe replicates, and then the data were used for analyses.

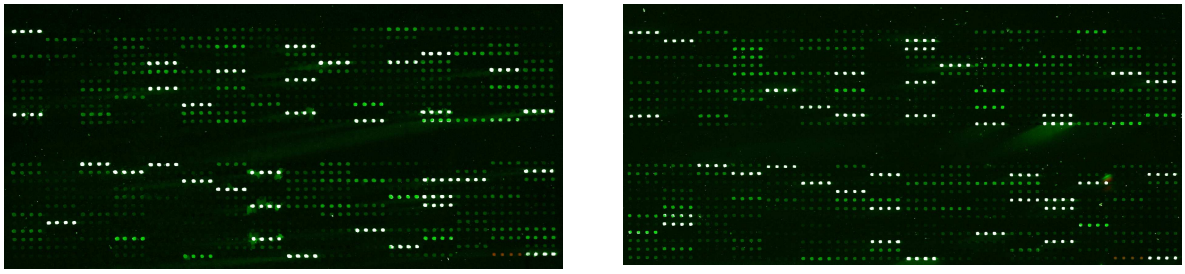


Fig.18: Representative images of scanned micro arrays. The left micro array is derived from a normal control sample. The right one originates from a sample of the study group M1.

4.3 Analyses of the CpG360-micro array data

4.3.1 Normalization

To identify new methylation markers in the serum of breast cancer patients, we first had to assure that all array data could be compared to each other without being influenced by any experimental effect or methodological failure.
So, if an array did not show at least two times the median signal intensity of the water control it was excluded form further analyses. Concretely that means (Figure 19), if an array showed an overall signal intensity below 11 it was excluded form the analyses. Therefore the data had to be normalized

Overview of samples used for data analyses

Study group	DNA isolation	Analysed	excluded in %
M1	6	5	17
Relapse	6	6	0
G3	6	5	17
G2	6	5	17
G1	6	6	0
C.I.S	6	6	0
Benign	12	12	0
BRCA1/2	10	9	10
WT	24	19	21
Total	82	72	12

Tab.5: DNA isolation reflects the number of samples, which were processed in this assay according to table 4. Number of analysed samples represents samples that were successful manipulated and amplified. Excluded in % indicates the fraction of samples per study group, which were excluded from methylation screening.

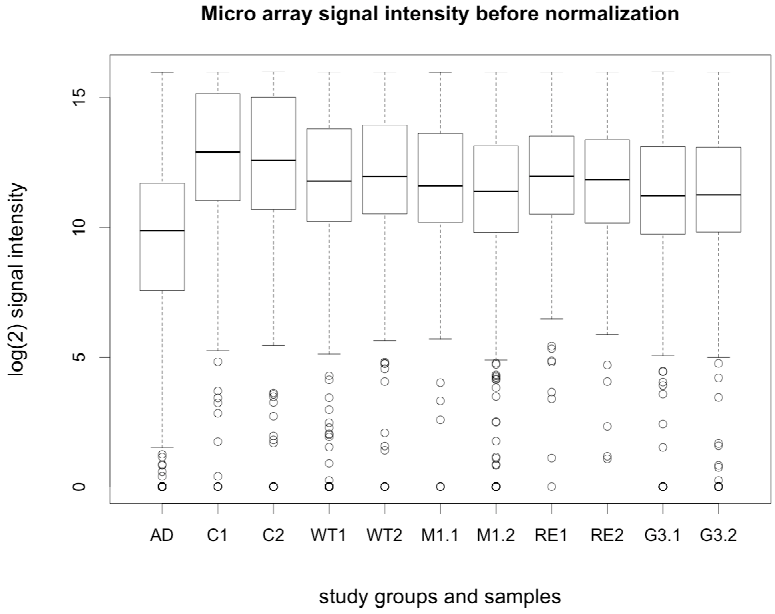
All water and positive controls were excluded before normalization and the clinical relevant samples were normalized using gene-by-gene median normalization from the software package BRB array tools for Microsoft Excel (Simon *et al.*, 2007).

Box plots of different representative samples before and after normalisation are shown in figure 19. The box plot shows comparable signal intensities derived from different patient groups. After normalization, not all samples did reveal the same median, which is due to the gene-by-gene normalization strategy.

The median normalization of BRB array tools first selected a reference array according to certain selection parameters, at which the median log intensity should equal the overall median of all arrays. Then, a gene-by-gene difference between each array and the reference array was computed. This median difference of each array and the reference array was then subtracted from the normalized arrays, in a way that the gene-by-gene difference between the normalized array and the reference array became zero.

Further a gene was filtered out, if more than 50% of the sample values from one probe were absent.

A



B

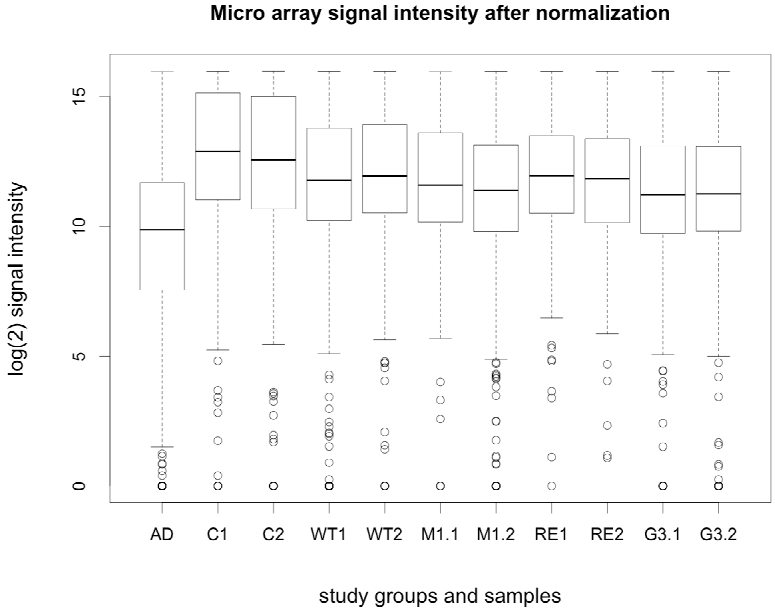


Fig.19: Box plots of log (2) values from signal intensities after DNA hybridization from chosen samples before (A) and after (B) gene-by-gene median normalization, which were grouped by their study group. Representative selected C1 and C2 refer to assay controls; WT data were taken from representative normal controls; M1, RE and G3 refer to their study group label.

4.3.2 Reproducibility of targeted methylation screen

An accurate multiplex PCR and hybridization process was ensured by insertion of twice the same sample at the experimental stage of the CpG360 multiplex PCR. In other words the RCA product of the samples P114 and P43 were randomly allocated into the multiplex amplification and hybridized on different arrays.

By means of this high correlation of r = 0,924 and r = 0,912 for the samples P114 and P43 respectively we could exclude that the following discrimination between diverse cancer samples and normal controls was due to PCR bias or variations in the hybridization process. These correlations also confirms the reproducibility of this multiplexing strategy.

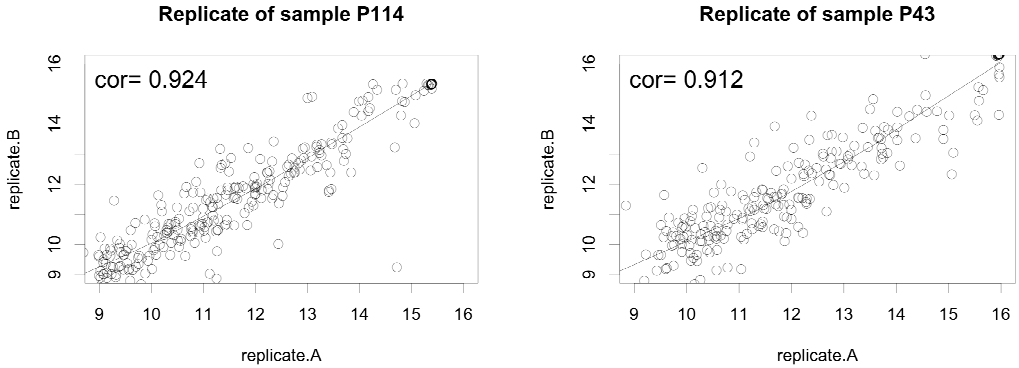


Fig.20: Correlation of normalized values of replicate samples. Sample P114 originates from the study group M1, whereas sample P43 is part of the study group benign breast cancer samples.

4.3.3 Assay controls

To ascertain the accuracy of the sample data we implemented triplicate samples of MCF7 and normal control DNA as well as negative controls in the experimental setup, which will be discussed in the following section.

The dendrogram in Figure 21 displays the assay controls, which illustrate on the one hand, the discrimination of MCF7 tumor cell line versus normal DNA of the multiplex PCR followed by the detection of the PCR fragments on the micro array, and on the other hand again assure a correct function of the CpG 360 assay.

Figure 21 provides an overview of positive controls in our assay (i.e. normal DNA isolated from whole blood of healthy adults and MCF7 cell line DNA), which enabled the discrimination between MCF7 breast cancer DNA and wild type DNA. This should show the optimal experimental setup for the multiplex PCR procedure, at which the DNA sample was not genome wide amplified and therefore the DNA has an optimum DNA quality for analyses.

The dendrogram was created in BRB array tools and compares the pair wise similarities of amplified PCR products from each sample. In this way, clusters of similar genes are created. Then the clusters were compared to each other, calculating a Pearson correlation, at which finally, one minus the Pearson correlation serves as metric distance between the samples as shown in Figure 21. (e.g. distance in the dendrogram between control MCF7 A and MCF B). In other words, we observed correlations between MCF7 replicates of r = 0,89 and between wild type DNA replicates a correlation of 0,88 was determined, being a measure for the entire procedure enrichment of methylated DNA.

Via the class comparison tool from the BRB array tools package, a calculation algorithm described in detail in the section “Discrimination between study group M1 and normal controls”, we could determine 24 significant genes with a P-value of $P < 0{,}005$, which allowed the discrimination between breast cancer DNA and wild type DNA.

These introductory examples confirmed the accuracy and reproducibility of this method as well as a prove of the suitability of this detection tool. After checking the performance of our test by analyses of several controls, we started data analyses for elucidation of methylation markers in serum of breast cancer patients.

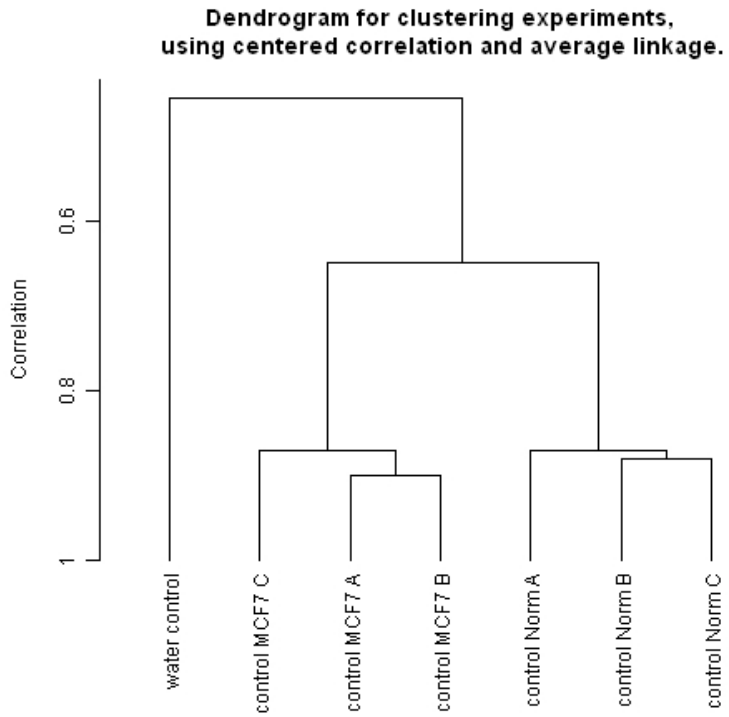


Fig.21: Performance testing of the CpG369 assay. Controls were gene-by-gene median normalized. Triplicates of the methyl sensitive digested DNA from MCF7 cell line or genomic wild type DNA were amplified and read out upon micro array hybridization without previous genome wide amplification.

4.3.4 Discrimination between study group M1 and normal controls

The CpG360-data analysis of the clinical samples started with the attempt to distinguish between patients with metastasizing cancer and normal controls. Due to the elevated serum DNA levels this strategy seemed most plausible, because the most significant difference were expected between these two groups.

Therefore, we performed cluster analyses of the two classes. This analyses may not always be biological meaningful, because all genes, which pass the filtering criteria, were included into analyses. As shown in Figure 22, two main clusters emerged in the dendrogram, which could be allocated to the two different study groups. One outlier was observed, namely sample WT181. This artefact could be explained, by having a look on the experimental course of the sample. Although we extracted a rather high amount of 35ng of DNA from the serum of sample 181 there were only about 9µg of DNA detected after genome wide amplification, which indicates low performance of the RCA reaction. In addition, the median signal intensity on the micro array was 9,8 (log (2) of the absolute value), which is distinct from other samples, and should be excluded from further analyses.

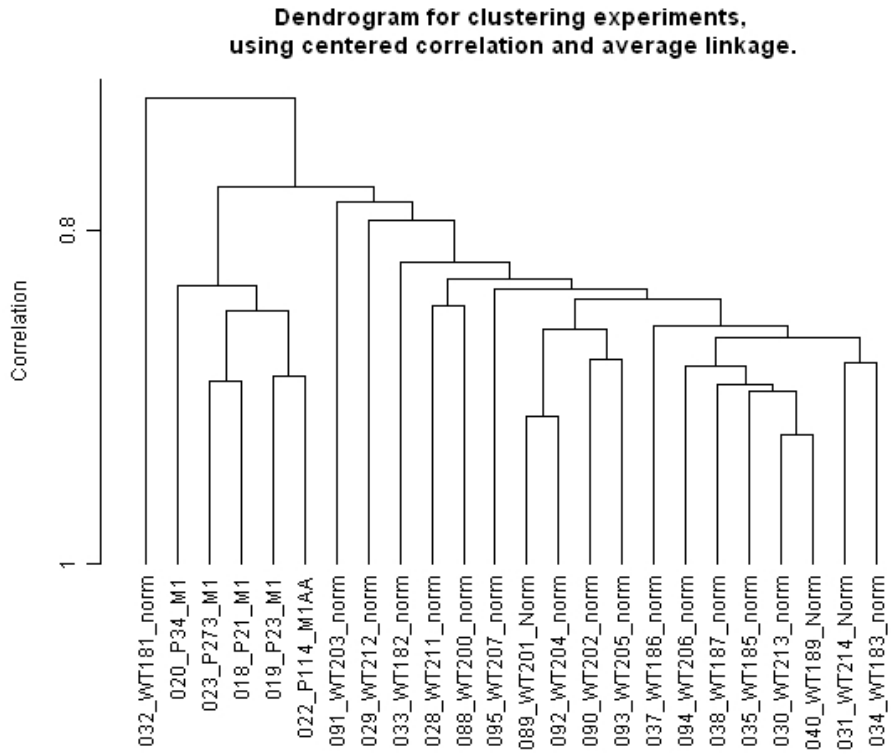


Fig.22: Cluster analyses of normal controls (indicated as "WT_" and "norm") and study group M1. Samples form the distinct groups are clustered. Two M1 samples (i.e. P23 and P114) show a correlation of r = 0.88, just like two WT samples (i.e. WT213 and WT 189) that show a correlation of 0.92. In contrast to that, two samples of different groups (i.e. P21 and WT213) show a correlation of 0,74.

As next step of our array data analysis we performed class comparisons between the two study groups (M1 versus WT), using BRB array tools. This calculation applies a two-sample t-test for each gene on the normalized log intensities of the DNA samples by comparing the values among the groups. Further, a univariante permutations test estimating the false discovery rate of each gene was computed. The method of Benjamini and Hochberg was applied for this calculation (Benjamini *et al.*, 2001).

The negative log (10) value of the P-value is plotted in Figure 23. For the creation of the plot (Figure 23) all significant genes at $P < 0.1$ are shown. Because of this minimal stringent application all false positive were displayed. This plot should just reflect the large number of genes, which are statistically significant but did not support discrimination between the study groups.

The genes, which lead to the separation of the array data into two groups, are highlighted red. These genes were chosen on the basis of the P-value, median log intensity, fold changes between the study groups and their false discovery rate. During analyses no genes with a false discovery above 5% were accepted and we did not mandatory select the genes with the highest P-values. The median signal intensities of the suggested significant genes were controlled first. This was done to exclude genes, which were differentially positive or negative. If a gene had a median log intensity of 13 in the WT group and a median intensity of 15 in group M1, the gene was maintained as significant although both values are clearly positive and indicating a successful amplification of the fragment in both study groups.

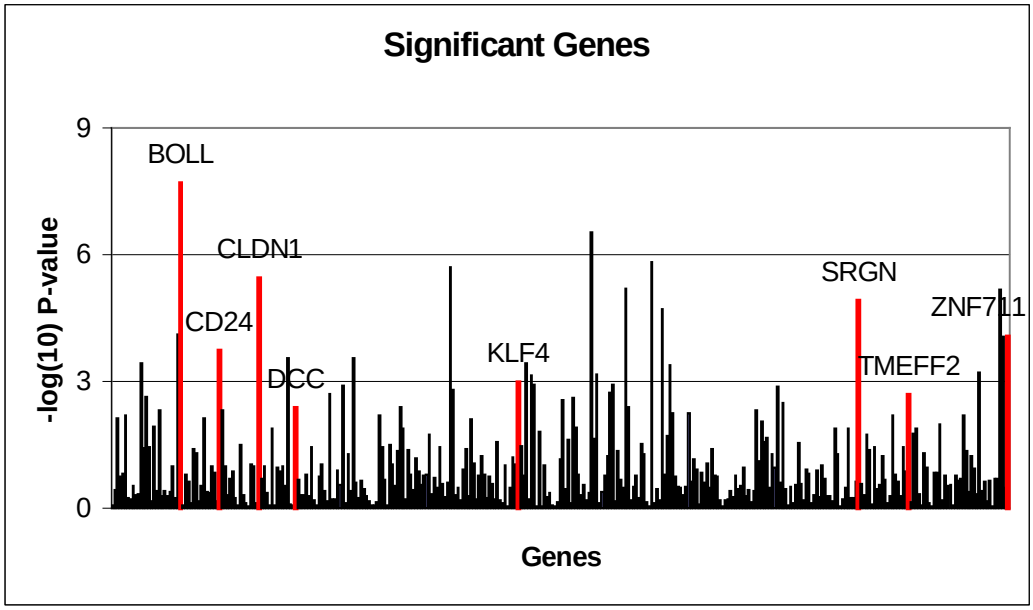


Fig.23: Negative log (10) value of P-values derived form two sample t-tests for all genes on the micro array in alphabetical order. Thus the higher the value in the plot the smaller is the concordant P-value.

Result of class comparison analysis from the 8 selected genes

Gene name	P-Value	FDR	Signal intensity of M1 samples	Signal intensity of WT samples
BOLL	<0,0005	<0,0005	15,8	10,1
SRGN	<0,0005	0,0006	15,7	5,3
CLDN1	<0,0005	<0,0005	15,7	11,1
ZNF711	<0,0005	0,003	14,7	5,8
CD24	<0,0005	0,004	13,4	6,5
DCC	0,002	0,04	11,9	9,5
KLF4	0.0005	0,01	10,5	11,5
TMEFF	0,004	0,04	12,3	9,8

Tab.6: Genes plus values, which lead to the selection of this marker set. FDR refers to the false discovery rate. The signal intensity of the two study groups reflects the median log (2) value of each gene for all samples of one study group.

Before we started our attempts to classify the other study groups (i.e. Relapse, G3, G2, G1, C.I.S., Benign and BRCA 1/2) by means of the 8 significant genes that enabled a separation of study group M1 and normal controls, we searched for genes that differ in their methylation profile between each study group and normal controls. We also checked if we could find genes that were differentially methylated in the specific study groups. These analyses were performed in the same manner as the search for significant genes in study group M1 via a two-sample t-test for each gene. This analysis approach did not elucidate marker for discrimination of the study groups. Therefore our further analysis was focused on the 8 selected genes (table 6), which enabled discrimination of M1 vs. WT samples. Before further statistical analyses, we had a closer view on the percentage of positive signals of each study group based on the introduction of a cut off value of 11 (table 7) as well as the percentage of positive signals matched to patient information (table 9).

Overview of positive signals in percent of each study group for 8 genes according table 6.

Study group	Gene names								
	BOLL	CD24	CLDN1	DCC	KLF4	SRGN	TMEFF2	ZNF711	Σ
WT (n=19)	10,53%	21,05%	15,79%	5,26%	84,21%	15,79%	0,0%	15,79%	12,5%
M1 (n=5)	100%	100%	100%	80%	0,0%	100%	80%	100%	95%
Relapse (n=6)	66,6%	66,6%	83,3%	50%	0.0%	66,6%	66,6%	66,6%	70,8%
G3 (n=5)	20%	60%	20%	0,0%	60%	60%	20%	40%	32,5%
G2 (n=5)	40%	20%	20%	40%	60%	60%	20%	40%	37,5%
G1 (n=6)	50%	33,3%	66,7%	33,3%	66,6%	33,3%	33,3%	33,3%	43,4%
C.I.S (n=6)	16,6%	16,6%	16,6%	66,6%	83,3%	33,3%	16,6%	16,6%	24,8%
Benign (n=11)	18,2%	18,2%	9,1%	18,1%	63,3%	9,1%	9,1%	18,2%	17,1%
BRCA1/2 (n=9)	44,4%	33,3%	33,3%	33,3%	66,6%	33,3%	22,2%	22,2%	

Tab.7: Percentage of positive signals per study group per gene.

Median log (2) signal intensity of selected genes according to study groups

Study group	Gene names							
	BOLL	CD24	CLDN1	DCC	KLF4	SRGN	TMEFF2	ZNF711
WT (n=19)	9,9	4,3	10,7	10,1	11,6	3,7	10.1	3,7
M1 (n=5)	15,9	13,3	15,7	12,1	10,5	15,9	12,7	14,7
Relapse (n=6)	15,4	12,6	15,3	11,5	10,6	15,4	12,4	14,3
G3 (n=5)	9,8	11,4	10,6	10,6	11,6	15,3	10,3	10,4
G2 (n=5)	10,6	3,7	10,4	10,6	11,4	14,32	10,2	3,7
G1 (n=6)	12,2	10,9	15	10,5	11,3	4,6	10,3	6,7
C.I.S (n=6)	10	6,6	10,5	11,2	11,4	5,2	10,4	3,4
Benign (n=11)	9,9	4,1	10,6	10,5	11,3	3,7	10,1	3,7
BRCA1/2 (n=9)	10,4	9,6	10,3	10,8	11,1	4,1	10,4	3,8

Tab.8: The median signal from all arrays of one study group was calculated for each gene.

Overview of positive signals in percent of 8 genes matched with different clinical information

Clinical information	Gene names								
	BOLL	CD24	CLDN1	DCC	KLF4	SRGN	TMEFF2	ZNF711	Σ
Hysterectomy (n=10)	60%	40%	50%	60%	50%	50%	50%	50%	51,2%
Lymph nodes (n=13)	61,5%	46,2%	53,9%	38,5%	38,5%	46,2%	38,5%	46,2%	46,1%
E.R.+ (n=16)	50%	56,2%	62,5%	56,6%	43,8%	62,5%q	50%	50%	53,8%
E.R.+/- (n=7)	28,6%	28,6%	14,35	28,6%	57,1%	42,9%	14,3%	28,6%	30,3%
P.R. + (n=5)	20%	40%	20%	40%	80%	40%	20%	20%	35%
P.R.+/- (n=10)	30%	30%	60%	30%	50%	30%	30%	30%	36,3%

Tab.9: The assessment of positive signals according to clinical markers did not reveal a significant elevation of positive signals of any gene for a specific clinical marker. The term Lymph nodes refers to confirmation of affected Lymph nodes of these patients. In other words these patients had local lymph node metastases. Estrogenic rec.+ implies that these patients had a strong positive detection of estrogenic receptors. Whereas E.R. +/- implies a moderate detection. This notation is valid for progesterone receptors as well

.

4.3.5 Discrimination between all study groups

Table 7 and table 8 give an overview of positives and signal intensities of 8 selected genes according to sample groups and clinical presentation. Keeping in mind that more than one false positive value increased the log value of median signal intensity, one can estimate the difference between the median values of study groups with differential methylated genes.

To make the calculated values in table 7 and 9 easier to interpret, we will briefly discuss one example. We compared the signals of the CLDN1 probe from WT to the M1 study group. In the M1 study group 100% of the samples were positive for CLDN1 obtaining a log value of the median signal intensity of 15.7. The log value of the median signal of CLDN1 from the WT group is 9.9, which does not imply that the difference between a positive CLDN1 signal and a negative one was 5.8, because in the WT group 15,6% of all arrays gave a positive signal (i.e. three arrays), when we recalculate the median signal intensity of the WT samples for CLDN1 without these three arrays a median log intensity of 9,5 is observed for this gene in this study group. This example also shows that the cut off value of 11 was rather used to rate gene signals positive than negative.

Particularly table 7 shows the result of this assay. One can see that the M1 group, which showed 100% of the expected signals for 6 out of 8 markers, can be clearly differentiated from the WT group, where 6 out of 8 marker showed fewer than 20% false positives, as well as from all other study groups. In the M1 study group all genes, except KL4, provided a clear positive signal for the selected genes. In contrast to that, the signals of these genes were overall negative in the WT samples. The methylation status of the gene KL4 seemed to be the other way round. Here we observed negative signals in the study group “M1” and “Relapse” and positive signals in the normal controls. Analysing the values of each candidate marker in table 7 and 8 concerning the difference of the median log intensities between the study groups and the number of false positive signals per candidate marker one could determine a ranking of the gene markers with most biological meaning and best discrimination between groups, which starts with SRGN followed by ZNF711, BOLL and CD24. This ranking of significant genes was confirmed partially by statistic test in the section “4.3.6. Statistical analyses”. These tables further highlight the prognostic potential of the gene SRGN, since high median log intensities were observed also in study group G3 and G2. Therefore the SRGN marker was considered as the most promising candidate marker of our assay. On the one hand, big differences in the median log intensities were observed, which indicates a biologic meaningful observation, and on the other hand, a promising course of the methylation status of this gene

can be assumed, because increasing positive signals could be observed with rising severity of the illness.

Despite this positive outcome of our experiments, results have shown that a distinct discrimination between the normal control group and “Benign”, “C.I.S”, “BRCA1/2” and “G1” was not possible and the differences between normal controls and study group G2 and G3 were rather small. It should be mentioned further that due to incomplete patient information the BRCA1/2 group might consist of healthy individuals, which had no tumor at time of blood donation, and patients with tumors of undefined grade. Because of this inaccuracy, we had to split the group into healthy patient and patients with breast cancer of undefined grade, which made the study group to small for a meaningful analysis.

The 66% of positive signal of the candidate marker gene DCC in study group C.I.S is still under investigation, because on basis of these experiments we could not predict whether this observation was due to false positive results or a biological reason.

The evaluation of the array data was continued by matching the positive or negative signals of the selected genes to the clinical information of the diverse patients (table 9). To establish a connection between the clinical features and our results, the array data were grouped according to this information. The organisations of the samples in study groups were broke up for this analyses approach. We observed a superior number of positive median signal intensities from patients, who underwent a hysterectomy, for all genes analysed. Higher numbers of positive signals of the selected genes were confirmed at a strong positive estrogenic receptor throughout the study cohort. Further analyses of the patient information like lymph node metastases, strong or moderate positive progesterone or estrogene receptors did not provide a significant increased or decreased levels of positive signals for the 8 analysed genes.

Figure 25 should give an insight into our decision to create a test set consisting of study group M1 and normal controls and a following evaluation process with all study groups. This plot of negative log (10) P-values (Figure 25) was created in the exact same way as the plot in Figure 23 and it shows clearly that the selected significant genes vanish in the surrounding P-values of genes with no biological meaning.

Finally we want to point out that a determination of significant genes, upon analyses of all samples at once, is more complicated and more error-prone than a division into two evaluation steps for elucidating markers, which are suitable to distinguish between study groups.

In figure 25 a throughout decrease of all P-values can be observed. This fact reflects also the restricted possibility to distinguish between groups as discussed above and shown in table 7 and table 8. Although there is a significant difference between normal controls and study group M1 and relapse, and the assumption that on basis of SRGN methylation also tumors of Grade 2 and 3 can be verified, this would have been difficult to figure out if we started our analyses with all samples at once. Figure 25 should also indicate that for the analyses of serum samples a two step analyses approach, where first exclusively sera of patients with metastasizing breast cancer and normal controls were analysed, is more suitable than an analyses of all available samples at once.

On basis of table 7 and 8 we considered if the study groups C.I.S and benign tumors should rather be rated as a wild type serum sample than a cancer sample. In other words we kept in mind that possibly only invasive malign neoplasms can be detected by methylation-markers in patient serum.

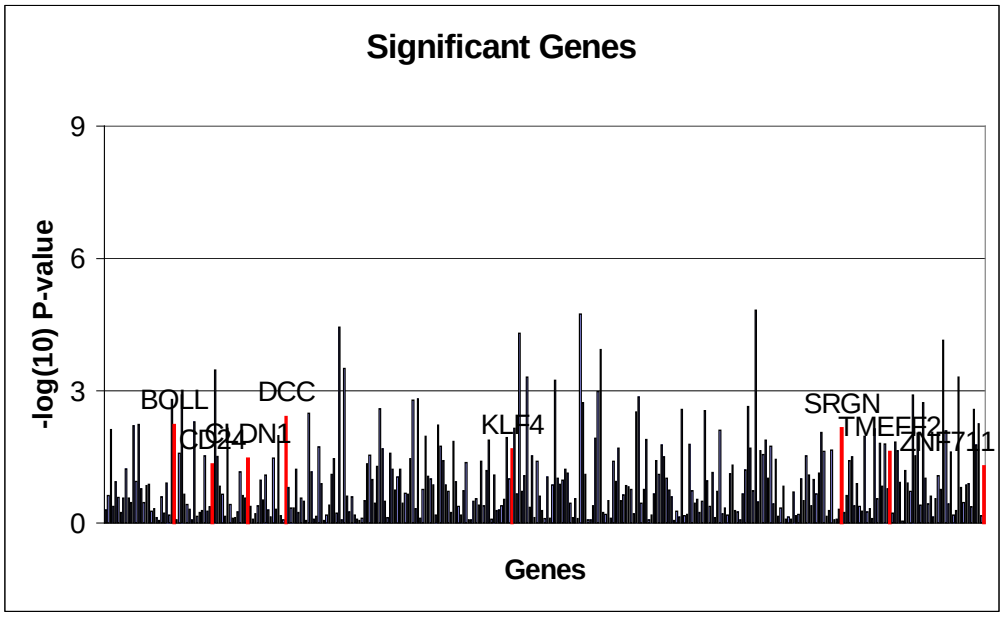


Fig.25: Negative log (10) value of P-values derived form two sample t-test for all genes on the micro array in alphabetical order. Values were computed, including all samples of all study groups into the calculation

4.3.6 Class prediction

The nearest neighbour prediction is another common method to classify array data. This parameter free method bases on a trait vector, which is assembled upon the log intensities of the selected genes. The classification can be lead back to the k nearest neighbours of already classified samples form each sample. The nearest neighbour of the new classified sample is chosen via the Euclidean distance between the samples.

The Euclidean distance is the stereoscopic distance between the k trait vectors of the k samples involved in the calculation. Based on this distance to the k neighbours of a new classified sample a majority decision is made whether the sample belongs to a group (e.g. normal controls) or not. In the application of this algorithm for our data we used a nearest centroid prediction, where only two classes of samples are considered. The result of this calculation is yes or no (Nigsch *et al.*, 2006). In our case if a sample belongs to a certain group or not. The nearest neighbour prediction algorithm, which was applied to create the data for figure 27, provided an insight into the statistical relevance of the 8 significant rated genes.

The study group Relapse, which consists of patients with recurrent malign neoplasm, can also be distinguished form wild type samples with a reliability of 60%. Including the other study groups into this prediction process the prognostic power of our data decreases. We observed a likelihood of less than 60 % for patients with a tumor of grade 3 and less than 50% for patients with a G1 or G2 tumor to be predicted accurate.

The study groups Benign and C.I.S. were rated as wild type samples and showed probabilities of more than 70% to belong to cancer samples, which is in the range of normal control samples. The Benign and C.I.S. samples were rated as cancer samples in another run of this prediction method (data not shown) were they also reached a likelihood of between 30% and 40% to belong to cancer samples.

For this prediction method it was true that there was highly significant differentiation between normal controls and serum of patients with metastasizing tumor. Due to the results of the nearest neighbour prediction the performance of the study group “Relapse” has to be discussed critically.

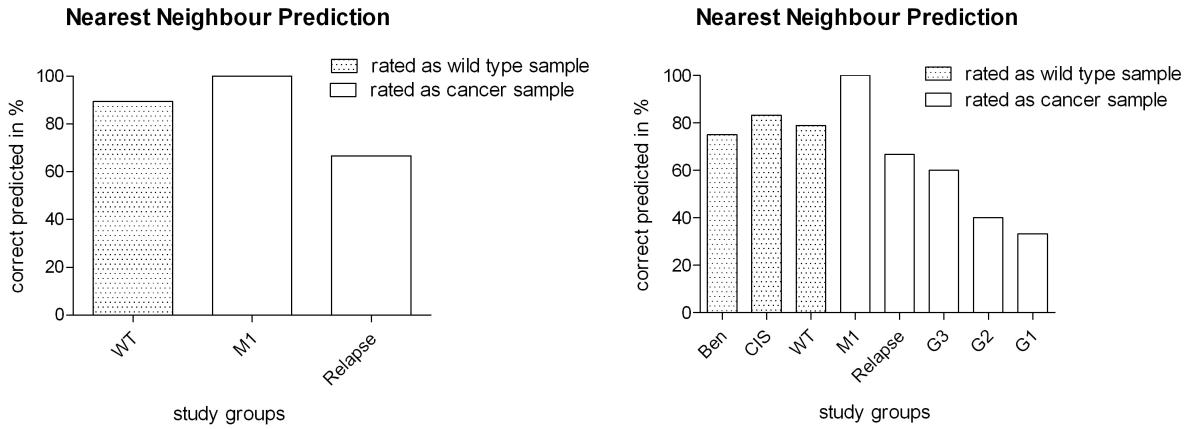


Fig.27: Nearest neighbour Prediction of all clinical samples applying the 8 significant rated genes (table 6). Study groups are indicated below. The bars show the portion of correct predicted samples per study group upon nearest centred prediction. The same numbers of samples were used as listed in table 5.

4.3.7. Binary tree Prediction

The binary tree prediction is a method that applies common prediction methods like in this case bayesian compound covariant predictor that uses this algorithm not to predict the analysed samples into classes, but it attempts to define two or more subclasses. For this analyses the samples are divided into several classes for example tumor, normal and benign. The algorithm then tries to split the samples into three groups by creation of subgroups that show the fewest cross-validated misclassification errors.

The Bayesian compound covariant predictor method calculates a linear predictor score (LPS) for each sample by introducing a scaling factor, which is based on a t-test for the difference between the two study groups. In this context the LPS is a linear combination of the median signal intensity values and distribution of these within the study groups, which should be normal distributed inside the study groups. In this way a mean variance is calculated for these normal distributions. This LPS distribution of each subgroup is then used to estimate a likelihood of a new sample, to belong to one of the study groups by an application Bayes´ rule. This rule allows, simplified expressed, a conversion of conclusions. This method was also used as an application of BRB array tools and was first described by (Wright *et al.*, 2003).

In our binary tree prediction analysis we created three subgroups, followed by the creation of a prediction set and a training set. The prediction set creates a “predictor” by means of the

samples in the prediction set and an evaluation is performed with the training set. Therefore we divided the samples into a tumor group, a group of normal controls and a group for Benign tumors. Then we splitted the samples according to their multiplex-PCR run in the experiment. In this way we had a prediction set consisting of study groups of the first two multiplex-PCR runs and a training set with the remaining two PCR runs. This analysis was also performed the other way around. Using the first two PCRs as “prediction set” we had over 40% of misclassified Tumor samples. As we performed the analysis the other way around no creation of subgroups was possible. These facts indicated towards a PCR-bias or hybridisation artefacts, because the first two PCR-runs were hybridized on one day and the hybridisation of the second two PCRs followed on the next day. Therefore we decided to evaluate our analysis strategy and then reanalyse the data with a different normalization strategy.

4.3.8. Evaluation of micro array analyses

The receiver operator characteristics curve (ROC-curve) is a graphical plot of sensitivity, which represents the false negative signals versus specivity that represents the false positives signals per sample. If a sample bears neither a false positive signal or a false negative signal it is found in the right top corner of the plot, which means 100% sensitivity and specivity. We used this analyses method to evaluate the quality and accuracy of our data analyses, because if the area under the ROC-curve is 0,5 the data are random distributed, which indicates that the analyses was not successful. This is the case if the curve is diagonal through the plot. We performed this analysis with various sample collections. The initial approach was to analyse only study group M1 versus normal controls, in this case the area under the ROC curve was almost one. Subsequently we added all study groups to the analyses. In figure 26 a prediction and creation of the ROC-curve was performed analysing the data of the significant rated genes of all clinical samples. We observed the most significant decrease of the area under the ROC-curve adding the study group “Benign” to the collective of samples although we rated them like C.I.S. samples as normal control samples. Therefore we decided to reanalyse our data aiming at the detection of markers, which discriminate between study group Benign, and cancer samples.

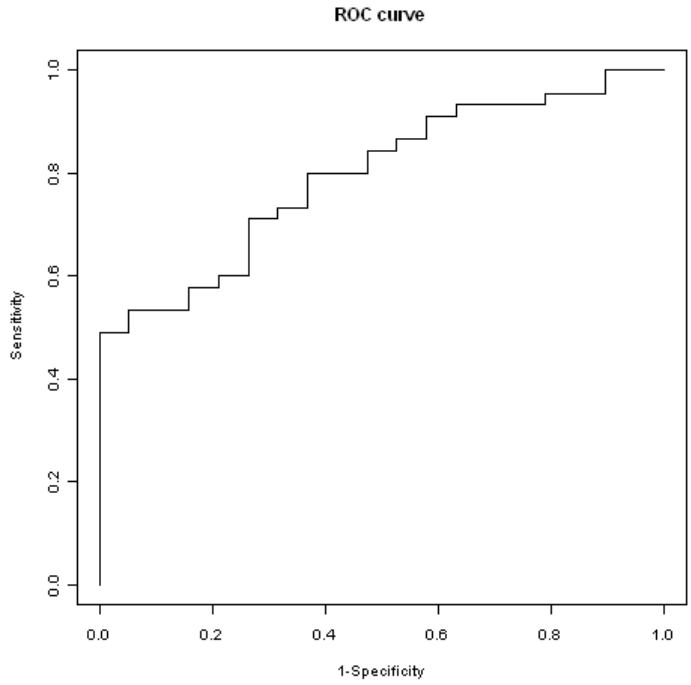


Fig. 26: ROC-curve of clinical samples using the 8 genes rated significant. Specifity refers to the false positives of each sample; sensitivity to the false negative signals of each sample.

4.3.9. Quantile Normalization

The most powerful normalization approach to counter variances between arrays of experimental origin is the quantile normalization, but this method has a very strong influence on the values of every data point within the array. This normalization algorithm ranks the data points of every array. Simplified expressed, the highest value gets “position one” (i.e. quantil 1) the second “position two” and so on. Then the median value for each data row, which must not consist of equal probes (i.e. gene loci), is calculated and inserted in every array at this position. This method results in the same distribution of signal intensities on different arrays. In this way experimental variances that lead to array specific signal intensity differences have no more influence on the data (figure 28).

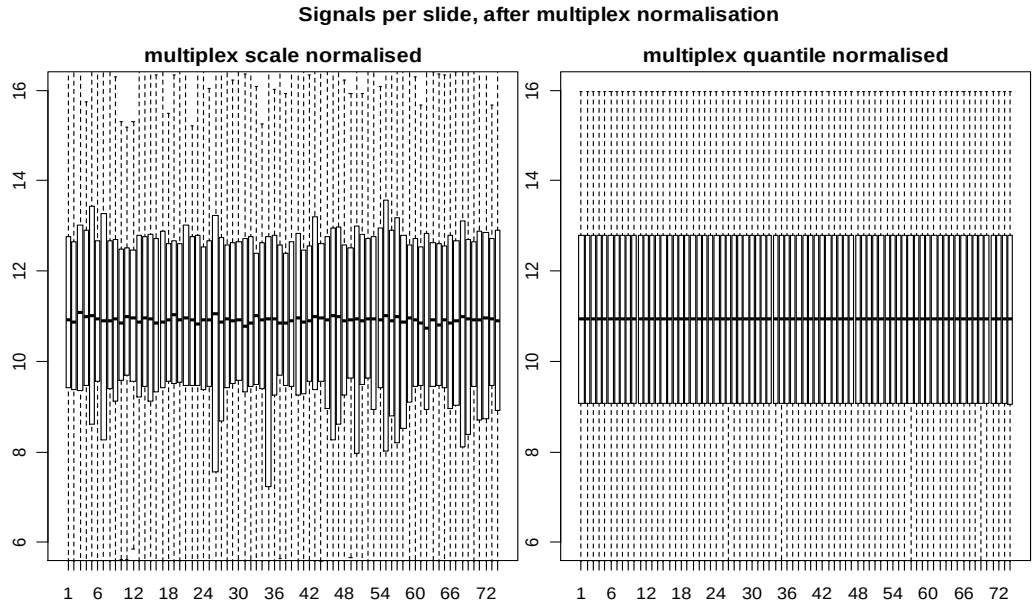


Fig.28: Box plots of normalized signal intensities per slide. Left section after scale normalization (comparable to figure 19); right section quantile normalized data. Numbers given below indicate the number of normalized arrays.

4.3.10. Binary tree prediction

We performed a binary tree prediction on the quantil normalized data dividing the samples again into three groups namely normal controls, cancer samples and samples of patients with a benign neoplasm. A "predictor" was calculated using all in all 21 genes (table 10) among them 6 of the 8 significant rated genes discussed in section 4.3.4 and 4.3.5. At this analysis approach about 25% of the cancer samples were classified wrong, which was either due to the inclusion of the study group Benign as own subgroup into the calculation or the already discussed PCR-bias.

Further we repeated the binary tree prediction as outlined in the 4.3.7 applying the same prediction and training set composition, which provided the same result consisting of over 40% misclassified samples or it was not possible to apply this algorithm if training and test set were interchanged.

Result of binary tree prediction		
Gene name	P-Value	% CV support
MYOD1	<0,0005	100%
CTNNB1	<0,0005	100%
ONECUT2	0,0003	83%
SRGN	0,0006	100%
NRCAM	0,0007	99%
CDH1	0,001	73%
TTC3	0,002	93%
DCC	0,002	100%
BOLL	0,003	100%
CDH1	0,004	100%
TP53	0,004	100%

Result of binary tree prediction		
Gene name	P-Value	%CV support
NANOS1	0,004	100%
NCL	0,008	100%
CXADR	0,01	100%
GDNF	0,01	100%
CD24	0,01	100%
CLDN1	0,01	100%
ZNF502	0,01	93%
TFF1	0,02	100%
ZNF711	0,03	100%
PITX2	0,04	59%

Tab.10: Output of binary tree prediction. Many gene loci show a significant difference between the study groups, but 25% of the cancer samples were misclassified. % CV support indicates the number of samples in % at which the application of a cross validation algorithm was possible and therefore reflects the prediction error for a test set.

4.3.11. Class prediction

As last step of our analyses we performed Class prediction with multiple methods as well as binary tree prediction for the PCR runs separately. Applying this strategy we aimed also at the discovery of novel marker for a differentiation between Benign neoplasms and cancer patients. If a PCR run or prediction strategy is not listed in table 11, the analyses did not provide any significant results. Compared to table 6 the P-values are significantly lower, which is due to the reduced numbers of samples in the analyses. The division of the samples according to their PCR runs was necessary, to guarantee a determination of markers, at which a PCR-bias or hybridization artefacts are minimized. Notably, we found two markers that may have the potential to discriminate between the serum of normal controls and the study group “Benign”. Anyway, both sets of markers (table 6 and table 11) have to be validated via qPCR tests before clear conclusions can be made about their diagnostic relevance.

Result of multiple prediction analysis

Gene name	P-Value	method	PCR run	discrimination	% CV support
TFF	0,02	ClassPred.	C+D	TU vs Norm	100%
TJP2	0,02	ClassPred.	C	TU vs Norm	100%
TIMP1	0,019	ClassPred.	C	TU vs Norm	84%
SMAD3	0,04	ClassPred.	C	TU vs Norm	47%
TFF	0,006	ClassPred.	D	TU vs Norm	100%
LZTS1	0,04	ClassPred.	D	TU vs Norm	23%
IL1B	0.01	ClassPred.	D	TU vs Norm	100%
TFF	0,006	Bin.TreePred.	D	TU vs Norm+Ben	100%
TP53	0,046	Bin.TreePred.	D	TU vs Norm+Ben	100%
CDKN2A	0,009	Bin.TreePred.	D	Ben vs Norm	33%
C5ORF4	0,027	Bin.TeePred.	D	Ben vs Norm	61%

Tab.11: Output of diverse Prediction methods to reveal serum markers. Method reveres to the applied Prediction algorithm. PCR run indicates the used data in the analysis. Discrimination shows division of the samples in subgroups within the analysis. % CV support indicates the number of samples in % at which the application of a cross validation algorithm was possible and therefore reflects the prediction error for a test set.

5. Discussion

An increased level of cell free (cf) DNA in human blood serum was first detected in association with systemic lupus in the 1960ies (Tan *et al.*, 1966). These finding fell into oblivion for a few decades, but now they are causing a great deal of attention since higher amounts of cell free DNA were detected in sera of cancer patients (Shapiro *et al.*, 1983). Current research estimates a high diagnostic potential of these free circulating nucleic acids in the blood serum.

In serum, the content of cell free (cf) DNA of healthy adults (Zhong *et al.*, 2007) as well as breast cancer patients (Seefeld *et al.*, 2008) is low, but the diagnostic potential of these ng amounts is promising (Gahan and Swaminathan, 2008).

Notably we could affirm the widely discussed increase of cf serum DNA levels of breast cancer patient (Kohler *et al.*, 2009) (Van, I *et al.*, 2009). However this elevated DNA amounts were observed only in patients with a metastatic disease. Therefore our data rather agree with observations of metastatic cancerous diseases (Tokuhisa *et al.*, 2007) than with an increase of cf serum DNA of overall malignant cancers (Wu *et al.*, 2002).

Since the technology of qPCR emerged and stood in as gold standard for appointment of small DNA amounts, endpoint measurements like Pico Green had become rare in current research. We took the decision to adjust the Pico green method and base our work on a simple measurement to determine the DNA amounts of the various isolation methods. By this approach we measured serum DNA concentrations within the range of 0,72ng – 96,21ng per ml. This range is smaller and there are only few outliers compared to recent reported DNA concentration measurements by qPCR (Deligezer *et al.*, 2008;Boni *et al.*, 2007;van der Drift *et al.*, 2010). Comparing qPCR and Pico green concentration appointments in our study, there was a higher correlation of replicate measurements for Pico Green (r = 0,933) than for qPCR, where no correlation above 0,84 was achieved (data not shown). A similar satisfactory performance and correlation of both methods, like reported by Szpechcinski A. (Szpechcinski *et al.*, 2008), could not be accomplished in our experimental setting.

The decision for an optimized protocol of the Roche High pure template preparation kit was evaluated by testing three other DNA purification approaches too. Two bead based methods like MBD isolation and ZR Serum DNA kit form Zymo did not provide sufficient amount for further DNA processing, although in case of the MBD based approach the methylated fraction

of the DNA was already enriched (Wielscher et al., 2010). In case of the membrane based isolation strategies the yielded DNA amounts using the optimized Roche protocol was significantly higher than using the Quiagen blood midi kit, which is widespread approach in current literature (Xue *et al.*, 2009).

To control the sample quality and the entirety of the extracted DNA, we performed a multiplex PCR on serum isolates, which led to the detection of 4 amplification products that showed product lengths of 124bp up to 315 bp for 4 different gene loci in nearly all analyzed samples. These experiments also gave an insight into the DNA integrity by the comparison of serum DNA of healthy patients and cancer patients, which did not indicate a decrease of DNA integrity.
The term DNA integrity index was defined in recent reports as the unequal distribution of long DNA fragments in relation to shorter ones whereupon the cell free serum or plasma DNA of normal controls should show a relative abundance of longer DNA fragments (i.e. PCR-products) starting at a size of 300bp. (Umetani *et al.*, 2006;Jiang *et al.*, 2006). By means of our observation, performed on 30 serum samples with a malign neoplasm and 24 normal controls, we could not support this assumption. Therefore additional qPCR tests on plasma isolated DNA were performed where no significant decrease of DNA fragments longer than 150bp in normal control DNA were detected too (data not shown). So this work emphasizes first critically considerations that do not consider an increased DNA integrity index as prognostic tool in cancer diagnoses (Schmidt *et al.*, 2008).

Three approaches to achieve a genome wide amplification were tested in this work. On the one hand adaptor mediated genome wide amplification systems like APA-PCR and the Whole Genome Amplification kit from Sigma-Aldrich, and on the other hand a strand displacement dependant approach like the RCA were evaluated. We decided to perform the genome wide amplification with the RCA system, because it performed best with the trace input DNA amounts isolated from serum.
The genome wide amplification was performed upon rolling circle amplification (RCA), a reliable method with many different applications, ranging from the amplification of vectors for cloning to template dependant amplification system in an isothermal reaction (Lizardi *et al.*, 1998). In this study, however, we aimed, using the “phi29” Polymerase for amplification of methylated DNA fragments, at a establishing of the multiplex-PCR mediated detection on our target arrays. Methylation sensitive restriction digestion guaranteed the enrichment of the methylated DNA fraction, because only methylated DNA fragments were left behind, which

not got digested and therefore were amplified by the “phi29” Polymerase. In this way DNA amounts in range of at least 10µg were produced out of a minimal sample input (10ng). We found that the genome wide amplification reaction is extremely dependant on the input DNA amount and the sample quality (Figure 15). The analyses of the gel images after RCA and the second control PCR (Figure 17) lead to the conclusion that the DNA recovery in this assay averaged out at 70%. We found a slightly higher number of complete RCA reactions as observed by Wang G. and colleagues (Wang *et al.*, 2004), who also performed RCA reaction on serum isolates. This might be due to different reaction conditions. Other voices, on the contrary, claim to obtain a DNA profile after rolling circle amplification with a similarity of 80% (complete RCA reactions) and more to the unamplified serum or plasma sample via SNP or micro satellite DNA analyses. (Nakamoto *et al.*, 2008;Croft, Jr. *et al.*, 2008). Further Croft D.T.jr. and colleagues achieved to isolate on average 366ng +/-27ng out of 200µl human blood plasma, a DNA amount that was never observed again in corresponding literature before or after this publication, might be due to remaining cells, which were avoided in our setting.

The detection of differential methylated DNA fragments was guaranteed via multiplex PCR followed by hybridization on a target array. After confirmation of an accurate performance of the detection system (section 4.3.2 and 4.3.3) we started the micro array analysis.

In our micro array analyses, we used class comparison methods to distinguish between two sample groups, which lead to a first determination of significant genes and creation of overview result tables. These tests provided first results that lead to a discrimination between the normal control group and patients with a metastasizing tumor with high level of significance. In further tests we detected that significant genes might be also due to an influence of the experimental setting. Therefore the accuracy of this selection of genes has to be questioned critically, but a qPCR-based evaluation of these possible markers will solve this question.

Additional analyses with reduced sample numbers were performed, in way that an influence of experimental variances is excluded, which lead to the discovery of another 10 possible methylation markers to distinguish between patients with malign neoplasms and normal controls. In this case the level of significance was lower ($P< 0,05$).

To give an overview of the biological meaning and possible clinical potential of the revealed methylation markers, we selected a few of them and discussed their biological function and the influence of their methylation on the affected cell in detail.

One example for a methylation marker we found in our study is the aberrant methylation of the promoter region of TP53 better known as p53, which is hyper-methylated in cancer. The gene product of p53 is a DNA binding protein that acts as transcription factor, which inhibits expression of growth and invasion factors and therefore is a typical tumor suppressor gene. In healthy cells p53 is expressed at a low level and leads via induction of several genes to cell cycle arrest or apoptosis as response to cellular stress (Amaral *et al.*, 2010). Besides germ line mutation (Li-Fraumeni syndrome, section 1.2.3) also aberrant methylation of the promoter region was observed in many cancer types (Patra, 2008). Our study confirms the finding of recent research reports that p53 is a possible methylation marker prevailing in serum of cancer patients (Fleischhacker and Schmidt, 2007).

The methylation status of the TMEFF2 gene in breast cancer was already described by Suzuki M. and co-workers (Suzuki *et al.*, 2005). They compared 4 normal cell lines where the TMEFF2 gene product was expressed in all 4 of them and methylated in none, to 10-breast cancer cell lines were the gene was expressed in 50% of the cell lines and methylated in the other 50% (Suzuki *et al.*, 2005). Further they observed a promoter hypermethylation of the gene in 35% of clinical samples, dependent on TNM staging of the tumor, where upon the frequency of the methylated tumors increased in accordance to higher TNM stages (Suzuki *et al.*, 2005). Our data complete these observations, because we detected TMEFF2 methylation in 80% of the tested metastasizing serum tumor samples. TMEFF2 methylation frequency in samples of less invasive neoplasm was 60% in the study group relapse, and 20% at G2 and G3. The gene product of TMEFF2 acts as epidermal growth factor and its aberrant promoter methylation was verified in colorectal tumors, gastric adenocarcinomas and other cancer types (Shibata *et al.*, 2002), where this locus was methylated on average in 80% of the tested sample material.

The situation at the analyses of KLF4 was the other way round. The KLF4 promoter region was methylated in normal controls and an absence of the KLF4 methylation in the study groups M1 and Relapse. KLF4 denotes the Kruppel-like factor 4, which was identified as tumor suppressor gene in colorectal cancer (Zhao *et al.*, 2004). KLF4 is a zinc-finger protein that acts as repressor or activator of transcription and its over-expression leads to an arrest of the cell cycle at the stage of G1/S1. This model was confirmed by several in vitro experiments and gene knockouts, (Ghaleb *et al.*, 2005) but the authors also reported increased levels of KLF4 mRNA detected in breast cancer and squamous cell carcinoma. These findings were due to a change of the biological behaviour of KLF4 in conjunction with different tumor

models. Our study may contribute to the elucidation of the reason for this, till now, not well-understood process, because we showed that the KLF4 methylation is decreased in breast cancer stages M1 and relapse, which would explain an increase of expression found by other work (Ghaleb *et al.*, 2005).

The DCC gene product is well described as a transmembrane protein called netrin 1 receptor that induces apoptosis of the cell in presence of its ligand nectin 1. Therefore it was classified as tumor suppressor gene, since it was found to be deleted in colorectal cancer (DCC) for the first time in 1990. The methylation of this locus and the successional loss of gene expression were shown for a handful of cancer types like gastric, head and neck, oesophageal squamous cell carcinoma and breast cancer (Park *et al.*, 2008). In our study we observed 40.7% methylated promoter regions of DCC including all the study groups with a malign neoplasm M1, Relapse and G1-G3, which agreed with previous findings of the methylation status of DCC in breast cancer (Miyamoto *et al.*, 2005).

Claudins have been found a potential diagnostic marker for gynaecological, lung, pancreas and other cancers. There is only little published about CLDN1 methylation and for CLDN3 and CLDN5. Some reports are suggesting that CLND3 and CLDN5 are aberrantly methylated in a few cancer types (Sato *et al.*, 2003). For all three genes, precise expression data of each gene in diverse cancers, including breast cancer, are available. In the case of the gene claudin1 (CLDN1), a trans membrane protein that forms tight junctions in epithelial cells, the down regulation was associated with carcinogenesis in breast cancer (Morohashi *et al.*, 2007). A coherence of down-regulated CLDN1 gene expression and the recurrence of breast cancer as well as an increased loss of expression in metastasizing tumors could be demonstrated. Mutation analysis of this gene did not provide any significant results in connection to its behaviour in breast cancer (Kramer *et al.*, 2000). Our data on the one hand clarify the mutation analysis study, because our data strongly indicate that the silencing of CLDN1 might be promoter-methylation mediated, and on the other hand, support the a decrease in expression data of CLDN1 related to breast cancer. We detected a methylation frequency of 100% from patients with metastasizing disease and a frequency of 83% for patients with recurrent breast cancer (study group Relapse). Furthermore the prevalence of the methylated promoter region of CLDN1 decreased rapidly to 28% in samples with less pronounced cancer phenotypes (i.e. study group G1-G3). A follow-up study to access whether the cancer of these 28% of patients developed a metastasizing stage or was recurrent would be interesting.

Another methylation marker with high-predicted statistical significance was the promoter region of BOLL, a member of the DAZ-gene family, which plays an important role in germ cell development. The loss of a functioning gene product of BOLL was brought into connection with infertility of both sexes, because this translational regulator promotes the formation of haploid gametes (Kee *et al.*, 2009). Apart from embryonic stem cell researchers, this gene attracted the attention of Shames DS (Shames *et al.*, 2006), who observed a four time higher expression of the BOLL gene, upon the induction of expression of methylated genes by 5-aza-2´-deoxycytidine treatment, compared to the results of cell lines derived from normal tissue that were treated the same way. The aberrant methylation of BOLL in breast cancer is indeed difficult to interpret, because this gene is predominantly expressed in germ cells (Yen, 2004). In this case we might have discovered a possible new marker for breast cancer, but to determine a functional role for this aberrant methylation, further tests would be necessary. A few assumptions in current literature could be applied for this finding, like a silencing process of genes that are not crucial for differentiation into tumor cells (Widschwendter *et al.*, 2007). Here a repression of the polycomb group proteins (Lee *et al.*, 2006), factors that regulate developmental genes upon trimethylation of histone H3 in embryonic stem cells is suggested, which would lead to an aberrant methylation pattern of this hypothesized cancer precursor cells.

Each gene discussed here would have a great potential to serve diagnostic marker or contribute to the understanding of malign neoplasms, but they have to undergo a precise evaluation before a further investigations would be meaningful. However, we developed a reproducible working protocol suitable for genome wide methylation screening of cell free serum DNA.

6. Conclusion

1. We established a serum DNA isolation protocol that enabled us to purify high DNA amounts in reproduce able manner from 1ml of patient serum. The application of this protocol resulted in a detection of significant increased DNA amounts in serum of breast cancer patients with a metastasizing tumor.
2. We successfully applied a methylated DNA enrichment procedure for the generation of satisfying DNA amounts out of limited DNA samples.
3. The utilization of the targeted “CpG-360” test elucidated 8 candidate marker genes, which might have the potential to discriminate between patients with a metastasising breast cancer and normal controls. Although our analysis also turned out that these markers could have been influenced by an experimental bias.
4. Further we could define 10 possible methylation markers, at which, due to our analysis approach, an influence of experimental variances is excluded.
5. Possibly this work is another step towards the development of a minimal invasive breast cancer test. Further qPCR tests could elucidate if these biomarkers, revealed here, have the potential to severe as serum methylation markers for breast cancer.

7. References

Research Papers

Amaral JD, Xavier JM, Steer CJ, Rodrigues CM (2010) The role of p53 in apoptosis. *Discov Med* 9 (45): 145-152

Anker P, Mulcahy H, Chen XQ, Stroun M (1999) Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev* 18 (1): 65-73

Anker P, Mulcahy H, Stroun M (2003) Circulating nucleic acids in plasma and serum as a noninvasive investigation for cancer: time for large-scale clinical studies? *Int J Cancer* 103 (2): 149-152

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125 (1-2): 279-284

Boni L, Cassinotti E, Canziani M, Dionigi G, Rovera F, Dionigi R (2007) Free circulating DNA as possible tumour marker in colorectal cancer. *Surg Oncol* 16 Suppl 1 S29-S31

Butler MG (2009) Genomic imprinting disorders in humans: a mini-review. *J Assist Reprod Genet* 26 (9-10): 477-486

Chistiakov DA, Voronova NV, Chistiakov PA (2008) Genetic variations in DNA repair genes, radiosensitivity to cancer and susceptibility to acute tissue reactions in radiotherapy-treated cancer patients. *Acta Oncol* 47 (5): 809-824

Croft DT, Jr., Jordan RM, Patney HL, Shriver CD, Vernalis MN, Orchard TJ, Ellsworth DL (2008) Performance of whole-genome amplified DNA isolated from serum and plasma on high-density single nucleotide polymorphism arrays. *J Mol Diagn* 10 (3): 249-257

Dahlin AM, Palmqvist R, Henriksson ML, Jacobsson M, Eklof V, Rutegard J, Oberg A, Van Guelpen BR (2010) The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clin Cancer Res* 16 (6): 1845-1855

Datta J, Ghoshal K, Denny WA, Gamage SA, Brooke DG, Phiasivongsa P, Redkar S, Jacob ST (2009) A new class of quinoline-based DNA hypomethylating agents reactivates tumor suppressor genes by blocking DNA methyltransferase 1 activity and inducing its degradation. *Cancer Res* 69 (10): 4277-4285

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* 99 (8): 5261-5266

Deligezer U, Eralp Y, Akisik EE, Akisik EZ, Saip P, Topuz E, Dalay N (2008) Size distribution of circulating cell-free DNA in sera of breast cancer patients in the course of adjuvant chemotherapy. *Clin Chem Lab Med* 46 (3): 311-317

Dimitri P, Caizzi R, Giordano E, Carmela AM, Lattanzi G, Biamonti G (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. *Chromosoma* 118 (4): 419-435

Durant ST, Nickoloff JA (2005) Good timing in the cell cycle for precise DNA repair by BRCA1. *Cell Cycle* 4 (9): 1216-1222

Egger G, Liang G, Aparicio A, Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429 (6990): 457-463

Fleischhacker M, Schmidt B (2007) Circulating nucleic acids (CNAs) and cancer--a survey. *Biochim Biophys Acta* 1775 (1): 181-232

Gahan PB, Swaminathan R (2008) Circulating nucleic acids in plasma and serum. Recent developments. *Ann N Y Acad Sci* 1137 1-6

Ghaleb AM, Nandan MO, Chanchevalap S, Dalton WB, Hisamuddin IM, Yang VW (2005) Kruppel-like factors 4 and 5: the yin and yang regulators of cellular proliferation. *Cell Res* 15 (2): 92-96

Gopalakrishnan S, Van Emburgh BO, Robertson KD (2008) DNA methylation in development and human disease. *Mutat Res* 647 (1-2): 30-38

Hodgson SV, Bishop DT, Dunlop MG, Evans DG, Northover JM (1995) Suggested screening guidelines for familial colorectal cancer. *J Med Screen* 2 (1): 45-51

Jair KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW, Baylin SB, Schuebel KE (2006) De novo CpG island methylation in human cancer cells. *Cancer Res* 66 (2): 682-692

Jiang WW, Zahurak M, Goldenberg D, Milman Y, Park HL, Westra WH, Koch W, Sidransky D, Califano J (2006) Increased plasma DNA integrity index in head and neck cancer patients. *Int J Cancer* 119 (11): 2673-2676

Jing F, Jun L, Yong Z, Wang Y, Fei X, Zhang J, Hu L (2008) Multigene methylation in serum of sporadic Chinese female breast cancer patients as a prognostic biomarker. *Oncology* 75 (1-2): 60-66

John EM, Miron A, Gong G, Phipps AI, Felberg A, Li FP, West DW, Whittemore AS (2007) Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA* 298 (24): 2869-2876

Juzeniene A, Setlow R, Porojnicu A, Steindal AH, Moan J (2009) Development of different human skin colors: a review highlighting photobiological and photobiophysical aspects. *J Photochem Photobiol B* 96 (2): 93-100

Kee K, Angeles VT, Flores M, Nguyen HN, Reijo Pera RA (2009) Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* 462 (7270): 222-225

Kioulafa M, Kaklamanis L, Mavroudis D, Georgoulias V, Lianidou ES (2009) Prognostic significance of RASSF1A promoter methylation in operable breast cancer. *Clin Biochem* 42 (10-11): 970-975

Knudson AG, Jr. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68 (4): 820-823

Kohler C, Radpour R, Barekati Z, Asadollahi R, Bitzer J, Wight E, Burki N, Diesch C, Holzgreve W, Zhong XY (2009) Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. *Mol Cancer* 8 105

Kramer F, White K, Kubbies M, Swisshelm K, Weber BH (2000) Genomic organization of claudin-1 and its assessment in hereditary and sporadic breast cancer. *Hum Genet* 107 (3): 249-256

Laird AD, Li G, Moss KG, Blake RA, Broome MA, Cherrington JM, Mendel DB (2003) Src family kinase activity is required for signal tranducer and activator of transcription 3 and focal adhesion kinase phosphorylation and vascular endothelial growth factor signaling in vivo and for anchorage-dependent and -independent growth of human tumor cells. *Mol Cancer Ther* 2 (5): 461-469

Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 11 (3): 191-203

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125 (2): 301-313

Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Zhang X, Wang J (2010) Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods*

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343 (2): 78-85

Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* 19 (3): 225-232

Lo YM, Chiu RW (2009) Next-generation sequencing of plasma/serum DNA: an emerging research and molecular diagnostic tool. *Clin Chem* 55 (4): 607-608

Lu Y, Lu P, Zhu Z, Xu H, Zhu X (2009) Loss of imprinting of insulin-like growth factor 2 is associated with increased risk of lymph node metastasis and gastric corpus cancer. *J Exp Clin Cancer Res* 28 125

McCabe MT, Lee EK, Vertino PM (2009) A multifactorial signature of DNA sequence and polycomb binding predicts aberrant CpG island methylation. *Cancer Res* 69 (1): 282-291

Melchor L, Benitez J (2008) An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes. *Carcinogenesis* 29 (8): 1475-1482

Miyamoto K, Fukutomi T, Akashi-Tanaka S, Hasegawa T, Asahara T, Sugimura T, Ushijima T (2005) Identification of 20 genes aberrantly methylated in human breast cancers. *Int J Cancer* 116 (3): 407-414

Mohn F, Weber M, Schubeler D, Roloff TC (2009) Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol* 507 55-64

Morohashi S, Kusumi T, Sato F, Odagiri H, Chiba H, Yoshihara S, Hakamada K, Sasaki M, Kijima H (2007) Decreased expression of claudin-1 correlates with recurrence status in breast cancer. *Int J Mol Med* 20 (2): 139-143

Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. *Cancer Res* 63 (22): 7641-7645

Nakamoto D, Yamamoto N, Takagi R, Katakura A, Mizoe JE, Shibahara T (2008) Detection of microsatellite alterations in plasma DNA of malignant mucosal melanoma using whole genome amplification. *Bull Tokyo Dent Coll* 49 (2): 77-87

Nigsch F, Bender A, van BB, Tissen J, Nigsch E, Mitchell JB (2006) Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization. *J Chem Inf Model* 46 (6): 2412-2422

Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. *Cell* 133 (7): 1145-1148

Paluszczak J, Krajka-Kuzniak V, Baer-Dubowska W (2010) The effect of dietary polyphenols on the epigenetic regulation of gene expression in MCF7 breast cancer cells. *Toxicol Lett* 192 (2): 119-125

Park HL, Kim MS, Yamashita K, Westra W, Carvalho AL, Lee J, Jiang WW, Baek JH, Liu J, Osada M, Moon CS, Califano JA, Mori M, Sidransky D (2008) DCC promoter hypermethylation in esophageal squamous cell carcinoma. *Int J Cancer* 122 (11): 2498-2502

Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. *CA Cancer J Clin* 55 (2): 74-108

Patra SK (2008) Ras regulation of DNA-methylation and cancer. *Exp Cell Res* 314 (6): 1193-1201

Pike BL, Greiner TC, Wang X, Weisenburger DD, Hsu YH, Renaud G, Wolfsberg TG, Kim M, Weisenberger DJ, Siegmund KD, Ye W, Groshen S, Mehrian-Shai R, Delabie J, Chan WC, Laird PW, Hacia JG (2008) DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status. *Leukemia* 22 (5): 1035-1043

Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Fan I, Tang J, Li S, Zhang S, Shaw PA, Narod SA (2006) Population BRCA1 and BRCA2 mutation frequencies and cancer penetrances: a kin-cohort study in Ontario, Canada. *J Natl Cancer Inst* 98 (23): 1694-1706

Roll JD, Rivenbark AG, Jones WD, Coleman WB (2008) DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. *Mol Cancer* 7 15

Sato N, Fukushima N, Maitra A, Matsubayashi H, Yeo CJ, Cameron JL, Hruban RH, Goggins M (2003) Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res* 63 (13): 3735-3742

Schmidt B, Weickmann S, Witt C, Fleischhacker M (2008) Integrity of cell-free plasma DNA in patients with lung cancer and nonmalignant lung disease. *Ann N Y Acad Sci* 1137 207-213

Seefeld M, El TS, Fan AX, Hahn S, Holzgreve W, Zhong XY (2008) Parallel assessment of circulatory cell-free DNA by PCR and nucleosomes by ELISA in breast tumors. *Int J Biol Markers* 23 (2): 69-73

Shames DS, Girard L, Gao B, Sato M, Lewis CM, Shivapurkar N, Jiang A, Perou CM, Kim YH, Pollack JR, Fong KM, Lam CL, Wong M, Shyr Y, Nanda R, Olopade OI, Gerald W, Euhus DM, Shay JW, Gazdar AF, Minna JD (2006) A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med* 3 (12): e486

Shapiro B, Chakrabarty M, Cohn EM, Leon SA (1983) Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease. *Cancer* 51 (11): 2116-2120

Sherman ME, Rimm DL, Yang XR, Chatterjee N, Brinton LA, Lissowska J, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, Kalaylioglu Z, Harigopal M, Charrette L, Falk RT, Richesson D, Anderson WF, Hewitt SM, Garcia-Closas M (2007) Variation in breast cancer hormone receptor and HER2 levels by etiologic factors: a population-based analysis. *Int J Cancer* 121 (5): 1079-1085

Shibata DM, Sato F, Mori Y, Perry K, Yin J, Wang S, Xu Y, Olaru A, Selaru F, Spring K, Young J, Abraham JM, Meltzer SJ (2002) Hypermethylation of HPP1 is associated with hMLH1 hypermethylation in gastric adenocarcinomas. *Cancer Res* 62 (20): 5637-5640

Shukla S, Mirza S, Sharma G, Parshad R, Gupta SD, Ralhan R (2006) Detection of RASSF1A and RARbeta hypermethylation in serum DNA from breast cancer patients. *Epigenetics* 1 (2): 88-93

Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y (2007) Analysis of Gene Expression Data Using BRB-Array Tools. *Cancer Inform* 3 11-17

Suzuki M, Shigematsu H, Shames DS, Sunaga N, Takahashi T, Shivapurkar N, Iizasa T, Frenkel EP, Minna JD, Fujisawa T, Gazdar AF (2005) DNA methylation-associated inactivation of TGFbeta-related genes DRM/Gremlin, RUNX3, and HPP1 in human cancers. *Br J Cancer* 93 (9): 1029-1037

Szpechcinski A, Struniawska R, Zaleska J, Chabowski M, Orlowski T, Roszkowski K, Chorostowska-Wynimko J (2008) Evaluation of fluorescence-based methods for total vs. amplifiable DNA quantification in plasma of lung cancer patients. *J Physiol Pharmacol* 59 Suppl 6 675-681

Szyf M (2009) Epigenetics, DNA methylation, and chromatin modifying drugs. *Annu Rev Pharmacol Toxicol* 49 243-263

Tan EM, Schur PH, Carr RI, Kunkel HG (1966) Deoxybonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. *J Clin Invest* 45 (11): 1732-1740

Tokuhisa Y, Iizuka N, Sakaida I, Moribe T, Fujita N, Miura T, Tamatsukuri S, Ishitsuka H, Uchida K, Terai S, Sakamoto K, Tamesa T, Oka M (2007) Circulating cell-free DNA as a predictive marker for distant metastasis of hepatitis C virus-related hepatocellular carcinoma. *Br J Cancer* 97 (10): 1399-1403

Ueng SH, Mezzetti T, Tavassoli FA (2009) Papillary neoplasms of the breast: a review. *Arch Pathol Lab Med* 133 (6): 893-907

Umetani N, Giuliano AE, Hiramatsu SH, Amersi F, Nakagawa T, Martino S, Hoon DS (2006) Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 24 (26): 4270-4276

Ushijima T (2005) Detection and interpretation of altered methylation patterns in cancer cells. *Nat Rev Cancer* 5 (3): 223-231

van der Drift MA, Hol BE, Klaassen CH, Prinsen CF, van Aarssen YA, Donders R, van der Stappen JW, Dekhuijzen PN, van der Heijden HF, Thunnissen FB (2010) Circulating DNA is a non-invasive prognostic factor for survival in non-small cell lung cancer. *Lung Cancer* 68 (2): 283-287

Van dA, I, Elst HJ, Van Laere SJ, Maes H, Huget P, van DP, Van Marck EA, Vermeulen PB, Dirix LY (2009) The presence of circulating total DNA and methylated genes is associated with circulating tumour cells in blood from breast cancer patients. *Br J Cancer* 100 (8): 1277-1286

Wang G, Maher E, Brennan C, Chin L, Leo C, Kaur M, Zhu P, Rook M, Wolfe JL, Makrigiorgos GM (2004) DNA amplification method tolerant to sample degradation. *Genome Res* 14 (11): 2357-2366

Watson L (2001) Breast cancer: diagnosis, treatment and prognosis. *Radiol Technol* 73 (1): 45-61

Weinhaeusel A, Thiele S, Hofner M, Hiort O, Noehammer C (2008) PCR-based analysis of differentially methylated regions of GNAS enables convenient diagnostic testing of pseudohypoparathyroidism type Ib. *Clin Chem* 54 (9): 1537-1545

Weksberg R, Smith AC, Squire J, Sadowski P (2003) Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development. *Hum Mol Genet* 12 Spec No 1 R61-R68

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW (2007) Epigenetic stem cell signature in cancer. *Nat Genet* 39 (2): 157-158

Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM (2003) A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A* 100 (17): 9991-9996

Wu TL, Zhang D, Chia JH, Tsao KH, Sun CF, Wu JT (2002) Cell-free DNA: measurement in various carcinomas and establishment of normal reference range. *Clin Chim Acta* 321 (1-2): 77-87

Xue X, Teare MD, Holen I, Zhu YM, Woll PJ (2009) Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. *Clin Chim Acta* 404 (2): 100-104

Yazici H, Terry MB, Cho YH, Senie RT, Liao Y, Andrulis I, Santella RM (2009) Aberrant methylation of RASSF1A in plasma DNA before breast cancer diagnosis in the Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev* 18 (10): 2723-2725

Yen PH (2004) Putative biological functions of the DAZ family. *Int J Androl* 27 (3): 125-129

Zhao W, Hisamuddin IM, Nandan MO, Babbin BA, Lamb NE, Yang VW (2004) Identification of Kruppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer. *Oncogene* 23 (2): 395-402

Zhong XY, Hahn S, Kiefer V, Holzgreve W (2007) Is the quantity of circulatory cell-free DNA in human plasma and serum samples associated with gender, age and frequency of blood donations? *Ann Hematol* 86 (2): 139-143

Books:

Hodgson S.V, Foulkes W.D, Eng C., Maher E.R. (2007) A Practical Guide to Human Cancer Genetics. ISBN 13-978-0-521-68563-4

Knipers Rolf (2001) Molekulare Genetik ISBN 3-13-477008-3

Lewin Benjamin (2004) Genes VIII ISBN 0-13-123924-4

Internet pages:

http://seer.cancer.gov/

http://srab.cancer.gov/devcan/

http://www.cancertreatmentwatch.org/general/grading.shtml

http://www.food.gov.uk/

http://www.wikipedia.org/

Curriculum Vitae

Name:	Wielscher Matthias
Address:	Castellezgasse 2/13 1020 Wien Tel.: 0650-8641043
Date of Birth:	29th of October 1979
Place of Birth:	Villach (Carinthia)
Parents:	Josef Wielscher Angela Wielscher

Academical Background

Nov. 07 – Dec. 08	Diploma thesis at the department Health and Environment – Molecular Medicine, Austrian Institute of Technology Title: "DNA methylation analyses for minimal-invasive breast cancer diagnostic testing"
2004	A 441 "Genetics/Microbiology" at the University of Vienna
10th of August, 2004	1st diploma
2002 – 2004	A 437 "Biology" at the University of Vienna
2002	University entrance diploma at Advanced Training Program for employees, Graz

Other activities

Oct. 09 – May 10	Care taking of my son (paternal leave)
Jan. 09 – Sept. 09	Member of research staff at the Austrian Institute of Technology

Soft skills:

- languages: English fluently
- Advanced knowledge: BRB array tools for micro array analyses, Windows, Office
- Basic knowledge: „R", Graphpad PRISM, Reference Manager12, Leopard 10.5 (system software Apple), diverse online tools (Primer3, NEBcutter, Blast, etc.)

Posters:

Vierlinger, K.; Mansfeld, M.; Schönthaler, S.; **Wielscher, M.**; Weinhaeusel, A.; and Noehammer, C.
Molecular classification for benign and malignant thyroid nodules.
ECE: 12th European Congress of Endocrinology 2010, Prague 24.04-28.04. 2010

Pichler, R.; Hofner, M.; Fürhauser, C.; **Wielscher, M.**; Schönthaler, S.; Singer, C.; Kandioler, D.; Panzer, R.; Liloglou, T.; Noehammer, C.; Weinhaeusel, A.
MSRE based DNA-Methylation Microarrays are Powerful Tools for Tumor-Marker Definition.
DECHEMA: 12th Status Seminar Chip Technologies, Sequencing and Functional Genomics 2010, Frankfurt / Main, 04.02-05.02. 2010

Weinhaeusel, A.; Hofner, M.; Fürhauser, C.; **Wielscher, M.**; Schönthaler, S.; Pichler, R.; Singer, C.; Kandioler, D.; Panzer, R.; Noehammer, C.
Multiplexed DNA methylation testing for efficient tumor marker definition and microarray validation.
Genau: European Human Genetics Conference 2009, Vienna, 23.05.2009 – 26.05.2009

Wielscher, M.; Hofner, M.; Nöhammer, C.; Weinhäusel, A.
Methyl-binding domain (MBD) protein based isolation of methylated DNA from human serum
Genau: European Human Genetics Conference 2009, Vienna, 23.05.2009 – 26.05.2009

Paper in Preparation:

Wielscher, M.; Hofner, M.; Peham, J.; Pulverer, W.; Nöhammer, C.; and Weinhäusel, A. MBD-based DNA isolation from human blood serum. (paper in preparation)