Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Chemistry, Biotechnology and
Food Science

Master Thesis 2014
60 credits

The Impact of Tumor Heterogeneity on DNA Methylation Levels in Breast Cancer

Ingunn Sivertsen

Acknowledgements

The work of this master thesis was carried out at the Department of Genetics, Institute for Cancer Research at the Norwegian Radium Hospital from 2012 to 2014 for the Master's degree in Biotechnology at the Norwegian University of Life Sciences (NMBU).

I would like to thank Professor Anne-Lise Børresen-Dale, the head of the Department of Genetics, for giving me the opportunity to take my master thesis at the department. I am also thankful to Professor Vessela N. Kristensen for having me as a master student in her project group. I would like to express my gratitude to my supervisor Dr. Thomas Fleischer for all the time you have dedicated to me throughout my thesis and for the good scientific discussions, suggestions and follow up. I would also like to thank my co-supervisors Dr. Hege G. Russnes and Dr. Hege Edvardsen for your scientific knowledge, suggestions and enthusiasms. In addition, thanks to my formal supervisor Professor Tor Lea for taking care of all the formalities.

I am grateful to Grethe Irene Grenaker Alnæs for the guidance during the laboratory work and for our discussions. Moreover, thanks to Martina Landschoof Skrede for giving me advice on how to use the cryomicrotome and showing me HE-staining. In addition, thanks to Daniel Nebdal for your technical support. A special thank goes to my colleagues at the department and for including me in the everyday activities.

I am also thankful for all the support I have been given from my friends. Last but not least I want to express my gratitude to my parents and sister for all your support, kindness, patience and encouragement and for always believing in me.

Abstract

Breast cancer is a heterogeneous disease with different clinical outcome. Tumor heterogeneity can be divided into inter-tumor heterogeneity representing variations between tumors from different patients and intra-tumor heterogeneity representing variations within a single tumor. Understanding the role of heterogeneity in tumor evolution and treatment response is important to give the most accurate diagnosis and treatment for each breast cancer patient.

DNA methylation is an epigenetic mechanism important for normal development. Alterations in DNA methylation pattern which might affect gene expression has been observed in breast cancer. Since DNA methylation might influence the gene expression, it might also influence the phenotype. Thus, DNA methylation might contribute to the tumor heterogeneity observed at a phenotypical level. Studies have looked at among other DNA methylation in regard to development, survival and treatment of breast cancer. Few studies have looked at the association between tumor heterogeneity and DNA methylation level in breast cancer. The main aim of this study, designed as a pilot study, was to investigate this association. More specifically the aim was to study the relationship between fraction of tumor cells/tumor percentage and DNA methylation level. Additionally the aim was to study the difference in DNA methylation level between subpopulations with different CD24/CD44 expression patterns within a cell line and between subpopulations with the same CD24/CD44 expression patterns across cell lines with different molecular subtype.

113 bulk tumors (containing tumor cells and surrounding tissue), 3 macrodissected tumors and cell lines sorted into different subpopulations based on phenotype (expression of CD24 and CD44) were used in the present study. Genome-wide DNA methylation data and estimated tumor percentage were avaible for the bulk tumors and DNA methylation analysis of the macrodissected tumors and subpopulations within cell lines was performed by pyrosequencing.

In this study, it was shown that the infiltrating non-neoplastic cells influence the measured DNA methylation level for many CpGs and genes. The genome-wide correlation analysis of 113 bulk tumors showed that the DNA methylation level of CpGs in around 1/3 of the genes in the human genome were associated to tumor percentage. For the macrodissected tumors it was shown that the DNA methylation level of most genes was associated with the amount of tumor cells in at least one of the tumors. The results of the correlation analysis and pyrosequencing corresponded well for the overlapping CpGs. For the macrodissected tumors

the DNA methylation level of more genes was related to the amount of tumor cells in the basal-like tumors than in the luminal tumor. Besides, greater difference in DNA methylation level between the regions was observed in the basal-like tumors. However, of the 113 samples used in the genome-wide correlation analysis, the majority of the samples were luminal tumors. It is therefore likely that infiltrating non-neoplastic cells influence the measurements of DNA methylation level in luminal tumors as well. These findings might pose a challenge for interpretation of studies that use bulk tumors. This might hamper DNA methylation studies like the investigation of DNA methylation as a biomarker, classification of tumors based on the methylation profiles of breast cancer genes and the investigation of drugs for modifying the DNA methylation signal. Further studies should be performed with more samples covering CpGs in larger part of the genome to explore the extent of this tendency. If the same tendency is seen in further studies, it might affect how DNA methylation studies will be performed and interpreted in the future.

Different methylation levels between samples of the same subpopulation within the same cell line (replicates) were revealed for *RASSF1A* and *FOXC1*. Due to too high variation between the replicates, the results weren't trustworthy. The only difference between the replicates is that they had been cultured separately. Studies have shown that epigenetic changes, including DNA methylation, may occur during cell culturing. Further studies should be done to enlighten this area.

Sammendrag

Brystkreft er en heterogen sykdom med ulikt klinisk utfall. Tumor heterogeneitet kan deles inn i intertumorheterogenitet og intratumorheterogenitet. Intertumorheterogenitet vil si variasjoner mellom tumorer fra ulike pasienter og intratumorheterogenitet vil si variasjoner innad i en tumor. Å forstå hvilken rolle hetoerogeneitet kan ha for tumorens utvikling og behandlingsrespons, er viktig for å gi best mulig diagnose og behandling for hver brystkreft pasient.

DNA-metylering er en epigenetisk mekanisme som er viktig for normal utvikling. Endringer i DNA-metyleringsmønsteret som kan påvirke genuttrykk, har blitt observert i brystkreft. Ettersom DNA-metylering kan påvirke genutrykk, kan det også påvirke fenotype. Således kan DNA-metylering bidra til tumor heterogenitet på et fenotypisk nivå. Studier har blant annet sett på DNA-metylering i forhold til utvikling, overlevelse og behandling i brystkreft. Få studier har derimot sett på sammenhengen mellom tumorheterogenitet og DNA-metylering i brystkreft. Hovedmålet i dette studiet, som er et pilotstudiet, var å undersøke denne sammenhengen. Mer spesifikt var målet å studere forholdet mellom mengde tumorceller/ tumorprosent og DNA-metylering. I tillegg var målet å studere forskjellen i DNA-metylering mellom subpopulasjoner med ulikt CD24/C44 mønster uttrykt innad i en cellelinje og mellom subpopulasjoner med samme CD24/CD44 mønster uttrykt på tvers av cellelinjer med ulik molekylær subtype.

Større deler av 113 tumorer (det vil si både tumorceller og omringende vev er inkludert), 3 makrodissekerte tumorer og cellelinjer som ble sortert inn i subpopulasjoner basert på fenotype (uttrykk av CD24 og CD44) ble inkludert i dette studiet. DNA-metylering og estimert tumorprosent gjennom genomet var tilgjengelig for de 113 tumorene og DNA-metyleringsanalyse av de makrodissekerte tumorene og subpopulasjonene innad i cellelinjene ble utført ved pyrosekvensering.

I dette studiet, ble det vist at infiltrerende ikke-neoplastiske celler påvirker den målte DNA-metyleringsverdien for mange CpGer og gener. Korrelasjonsanalysen gjennom genomet for 113 tumorer viste at det var en sammenheng mellom DNA-metylering og tumorprosent for CpGer i omtrent 1/3 av genene i det humane genom. For de tre makrodissekerte tumorene ble det oppdaget en sammenheng mellom DNA-metylering og mengde tumorceller for de fleste av genene i minst en av tumorene. Resultatene fra korelasjonsanslysen og pyrosekvenseringen stemte godt overens for de overlappende CpGene. For de makrodissekerte tumorene ble det

oppdaget en sammenheng mellom DNA-metylering og tumorprosent for flere gener i de basal-lignende tumorene enn for den luminale tumoren. I tillegg ble det observert større forskjeller i DNA-metylering mellom regionene i de basal-lignende tumorene. Hovedmengden av de 113 tumorene inkludert i korrealsjonsanalysen var luminale tumorer. Derfor er det grunn til å tro at infiltrerende ikke-neoplastiske celler også påvirker målingen av DNA-metylering i luminale tumorer også. Disse funnene viser at det å tolke DNA-metyleringsverdier i studier hvor større deler av tumorer er blitt inkludert kan være utfordrende. Dette kan være en hindring for DNA-metyleringsstudier som undersøkelse av DNA-metylering som en mulig kandidat for biomarkør, klassifisering av tumorer basert på DNA-metyleringsmønsteret av gener assosiert til kreft og undersøkelse av medisiner som kan indusere DNA-demetylering. Fremtidige studier med flere prøver hvor flere CpGer i større deler av genomet blir dekket burde bli gjennomført for å undersøke hvor utbredt dette mønsteret er. Dersom det samme mønsteret blir oppdaget, kan det påvirke hvordan DNA-metyleringsstudier blir tolket og utført i fremtiden.

Forskjellig DNA-metyleringsmønster ble oppdaget mellom prøver med samme subpopulasjon innad i samme cellelinje (replikater) for *RASSF1A* og *FOXC1*. På grunn av for store variasjoner mellom replikatene, var ikke resultatene til å stole på. Den eneste forskjellen mellom replikatene er at de ble dyrket separat. I andre studier har det blitt vist at epigenetiske endringer skjer under celledyrking, også for DNA-metylering. Fremtidige studier hvor nettopp dette blir utforsket nærmere burde bli gjennomført.

Contents

1 Introduction

1.1 Cancer

Cancer is a common disease and one of the leading causes of death worldwide (1). In 2012, 14.1 million new cancer cases and 8.2 million deaths related to cancer were registered internationally (2). The cancer diseases that cause most deaths every year overall are lung, stomach, liver, colon and breast cancer (1). In Norway, 30 099 new cancer cases were registered in 2012 (3). At the end of the same year, 224 315 people that had been diagnosed with cancer earlier in life, were still alive (3).

1.1.1 Cancer as a genetic and epigenetic disease

Cancer develops as a result of uncontrolled cell growth and dividing whereby abnormal cells may proliferate and become a tumor (4). For a tumor to develop both epigenetic and genetic alterations are important (4). Mutations in multiple genes have to occur for the cells to become cancerous, thus cancer is considered a polygenic disease (5). During development of cancer, mutations in a cell that are favorable for the tumorigenesis may accumulate as the cell grows and divides. Three groups of genes that are central for normal growth are often highly mutated in cancer (4). These include: tumor suppressor genes, oncogenes and stability genes (4). These mutations can be inherited or occur randomly in a somatic cell (5).

"The modern definition of epigenetics is information heritable during cell division other than the DNA sequence itself" (6). This information refers to changes that modify the DNA sequence, but doesn't change the naked DNA sequence. In tumors, the epigenetic landscape is highly changed and it is well established that epigenetic changes has a role in tumor development and progression (7).

1.1.2 Hallmarks of cancer

Cancer is a group of heterogeneous diseases, and specific properties are thought to be required for a tumor to develop and further progress towards malignancy. These properties are referred to as "the hallmarks of cancer" and was presented by Hanahan and Weinberg in an article from 2000 (8). The hallmarks of cancer are the six acquired capabilities: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis, illustrated in Figure 1, and are common in most if not all cancers (8;9).

In the 2011 update of the article, two new emerging hallmarks were suggested to the list as they are important for tumor development and progression (9). These are: deregulating cellular energetics and avoiding immune destruction (9). In addition, two enabling characteristics: tumor-promoting inflammation and genome instability and mutation were proposed (9). These enabling characteristics are the foundation of the other eight properties. They do not necessarily cause cancer, but rather assist in the progression of tumor. For instance the enabling characteristic genome instability and mutations may lead to mutations in tumor suppressor genes that may result in inactivation of the genes (9). Thereby inappropriate cell growth will be less prevented (9). In this way the cells may acquire the hallmark capability: evading growth suppressors (9). Epigenetic alterations are included in this characteristic as for instance inactivation of a tumor suppressor may be due to epigenetic changes in the gene (9). Acquisition of the hallmarks is possible by the other enabling characteristic as inflammation provides bioactive molecules enhancing the tumor progression, such as survival factors, to avoid cell death (9).

Figure 1: The hallmarks of cancer. The figure represents the six originally acquired capabilities for tumor development: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis. In addition the emerging hallmarks: deregulating cellular energetics and avoiding immune destruction as well as the enabling hallmarks: tumor-promoting inflammation and genome instability and mutation have been added (9).

1.2 Breast Cancer

Breast cancer is the second most common cancer disease across the world (10). In 2012, 1.67 millions new breast cancer incidences were estimated to have occurred that year (11). In Norway, breast cancer is the most common cancer disease for women in the age of 25-69 (12). In 2012, a total of 2984 Norwegians (2956 women and 28 men) were diagnosed with breast cancer and 649 deaths (645 women and 4 men) were registered (13). The five-year survival for breast cancer in Norway for female has increased from 67.9 percent in the period 1973-1977 to 89.1 percent in the period 2008-2012 (13).

1.2.1 Breast anatomy and development

The mammary gland consists of many different types of cells and tissues, which makes the mammary gland a complex structure. The adult mammary gland consists of 15-20 lobes (14). Each of these lobes is composed of smaller lobules which are responsible for milk production (Figure 2). Each lobe contains branched ducts whose function is to transport milk from the lobule to the nipple. Lobules and ducts are composed of a bi-layered structure of luminal epithelial cells and myoepithelial cells surrounded by a basement membrane. The luminal epithelial cells forms the inner layer, while the myoepithelial cells shapes the outer layer and is enclosed by the basement membrane (15). Breast cancer occurs most frequently in the ducts or the lobules (carcinoma), and less frequently in the stroma (sarcoma) (16).

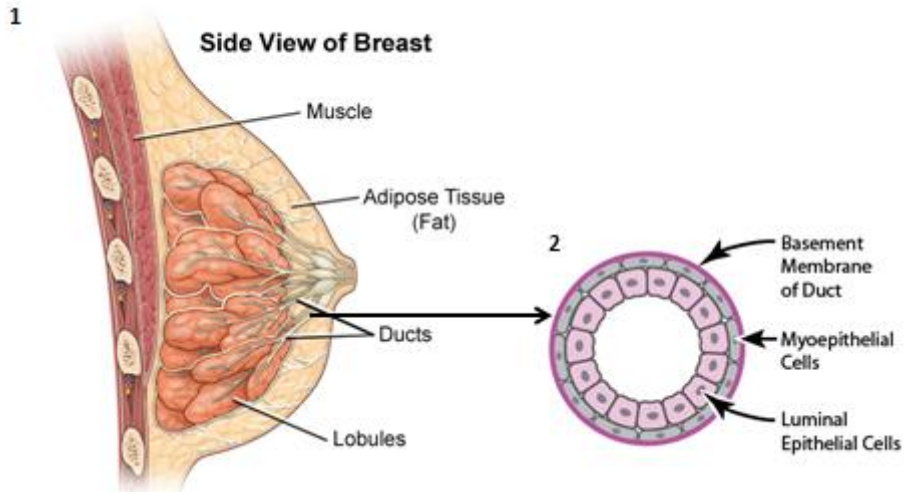


Figure 2: The anatomy of the mammary gland. 1: the mammary gland consists of among other ducts, lobules, adipose tissue (fat) and muscle 2: the duct consists of: luminal epithelial and myoepithelial cells and the basement membrane (17;18).

The development of mammary gland starts during the fetal development (15). The breast will further develop during its lifetime as a result of natural hormone changes (15).

1.2.2 Risk factors

Many risk factors including genetic, epigenetic, hormonal, lifestyle and environmental factors increase the chance of developing breast cancer.

The majority of breast cancers develop as a result of random somatic variants while familial breast cancers, constituting approximately 5-10% of all breast cancers, are caused by germline variants (5). Common variants in genes associated with low risk of developing breast cancer to rare variants in genes associated with high risk of developing breast cancer have been revealed (19) (Figure 3). The latter variants are associated with familial breast cancers (19). Variants in breast-cancer gene 1 (*BRCA1)* and breast-cancer gene 2 (*BRCA2)* are responsible for about 20% of all familial breast cancers (5). The common variants are single nucleotide polymorphisms (SNPs) associated with sporadic breast cancer (19). Changes affecting a single base have to occur in many genes if development of breast cancer should occur. These SNPs have been identified by genome-wide association studies (GWAS) (20;21). In GWAS studies up to several million SNPs are genotyped simultaneously, and so far SNPs in genes such as fibroblast growth factor receptor 2 (*FGFR2*), mitogen-activated kinase 1 (*MAP2K1*) and caspase 8 (*CASP8*) have been identified to be associated with breast cancer (19).

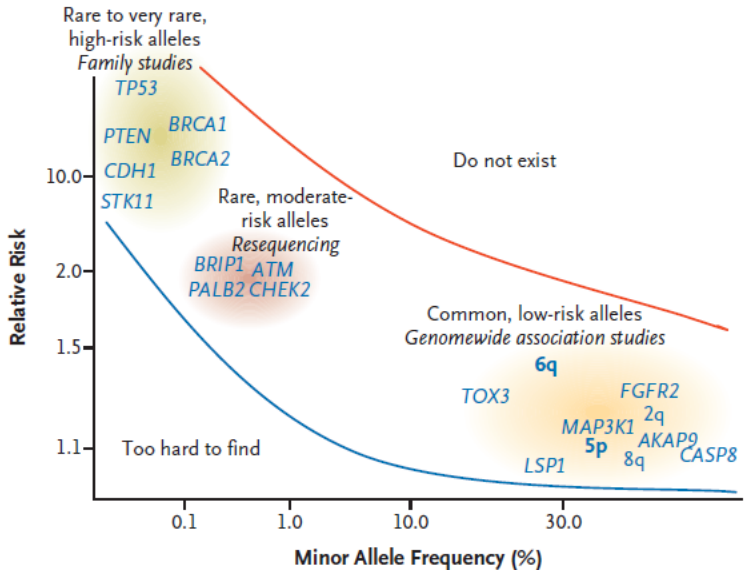


Figure 3: Risk genes for breast cancer. The figure represents known genes, where variants in these genes are associated with an increased risk of developing breast cancer. These genes are shown between the blue and red line; above the red line no genes are thought to exist and below the blue line no genes have been found. The x-axis shows how frequent a variation in a gene is, while the y-axis shows the risk of developing breast cancer, if the mutation is present. The high-risk genes are highlighted in green, the moderate-risk genes in red and the low-risk genes in orange (19).

Breast cancer risk is associated with being female and increasing age. In addition, early menarche, late age at menopause, nulliparity or late age at first birth are all hormonal risk factors that increase the chance of developing breast cancer as these incidences implies prolonged exposure of endogenous estrogen to the breast (22;23). A study of Building (24) showed not only that early menarche and late age at menopause increased the chance of developing cancer but that greater risk is associated with early menarche than late age at menopause. Other hormonal risk factors include exposure of exogenous hormones, for instance oral contraceptive or hormone replacement treatment (23). Lifestyle risk factors such as obesity, smoking and alcohol consumption and environmental factors such as radiation have also been shown to contribute to increased breast cancer risk (22;23). A recent study showed that incidents of breast cancer increases with initiating smoking before first birth (25). Higher mammographic density is a risk factor and it was shown that women with high mammographic breast density diagnosed with benign tumor had a high risk of developing breast cancer (26). This study also showed that women with low mammographic breast density had a low risk of being diagnosed with breast cancer, regardless of being diagnosed with benign tumor (26).

1.2.3 Breast cancer progression

The initiation and progression of breast cancer is hypothetically thought to start with normal epithelial cells that develop and reach the stage atypical hyperplasia which is defined as a stage with increased number of cells (27). From atypical hyperplasia the next stage is carcinoma *in situ*, then invasive carcinoma which finally can have the ability to metastasize (28). Figure 4 represents this progression exemplified by a breast duct. In invasive breast cancers the basement membrane and the myoepithelial cells are lost (28). As a result of this, the cancer cells have to possibility to invade surrounding tissues or spread to distant organs generating secondary tumors (28).

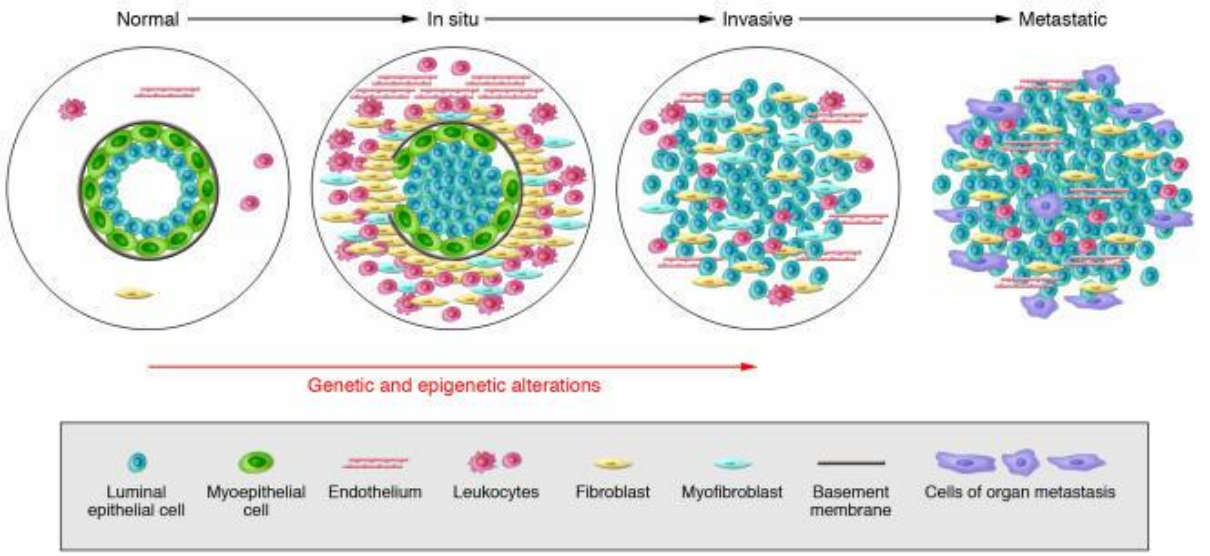


Figure 4: A possibility for how progression of breast cancer occurs in ducts. This illustration shows the progression in ducts from normal epithelial cells to metastatic cells. Normal ducts consist of myoepithelial and epithelial cells that are enclosed by the basement membrane. Stroma composed of fat cells, pericytes, endothelial cells, leukocytes, fibroblasts and myofibroblasts surrounds the basement membrane. Myoepithelial cells and the basement membrane are seen in the earlier stages of development, but are lost in invasive carcinomas (28).

1.2.4 Tumor classification

To assess prognosis and treatment course for a given patient, different parameters have been used to classify the tumors, including histopathological type, grade, stage and receptor status. Of the malignant breast tumors, invasive ductal carcinoma (IDC) is the most common type and constitutes 50-80% of all breast cancers (29). Invasive lobular carcinoma (ILC) constitutes 5-15% of all breast cancers (29). The remaining breast cancer states are rare malignancies such as medullary (constitutes 5%), mucinous (constitutes 2%), tubular (constitutes 1%) and adeno cystic carcinoma (constitutes 1%) (30).

Grade is used to describe the resemblance between cells of the tumor and normal breast epithelial cells and is associated with the aggressiveness of the tumor (31). Grade is based on three features: tubule formation (how much of the tumor consists of normal duct structure), nuclear grade (size and shape of the tumor cells) and mitotic activity (rate of the cell division) (31). Based on these features a score from 1 to 3 is given where tumors with grade 1 have the best prognosis and tumors with grade 3 the worst prognosis (27).

Stage is a parameter used to explain the extent of the disease and is based on the TNM staging system (27). T stands for tumor size, N for lymph nodes and M for metastases. The size of a tumor can be categorized into four stages: T1-T4 (27). For T1-T3 the following applies: the

higher number of T, the larger tumor size (27). T4 on the other hand includes tumors of any size with direct extension to (a) chest wall or (b) skin (27). The lymph nodes involvement are categorized into three stages: N1-N3 and is based on the number and location of cancer cells in the lymph nodes (27). M consists of two categories M0 (no metastases) and M1 (metastases present) (27). Based on T, N and M values the tumors are categorized into stage I to IV, where IV refers to the stage with the greatest extent of the disease, and these patients have the worst prognosis (27).

The receptor status of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) are prognostic markers but also used to decide treatment options. Patients with ER+ and PR+ tumors (meaning the tumors expresses the receptors) have a better prognosis, and may benefit from anti-estrogen treatment (27). HER2 is highly expressed in approximately a quarter of the patients (27). Treatment with medicaments such as the monoclonal antibody trastuzumab has been shown to improve the outcome of these patients (32). However, several HER2+ tumors are non-responsive to this treatment due to trastuzumab resistance (33;34). ER-/PR-/HER2- tumors have the worst prognosis (35). Recently, Ki-67 was included as a prognostic marker. Ki-67 is an antigen associated with the proliferative activity of the breast cancer, and a high Ki67 score indicates a more proliferative tumor (36). Expression of this protein is associated with higher probability of relapse and worse outcome (36).

By using gene expression it has been possible to organize breast cancers into different subtypes. Perou and Sorlie identified five subtypes: luminal A and B, basal-like, HER2 enriched and normal-like (37-39). Luminal tumors are primarily ER+, PR+ and express genes that normally are expressed in normal luminal epithelial cells (39;40). HER2 enriched tumors express HER2 and are mostly ER negative (39). Basal-like subtypes are primarily ER negative, PR negative and HER2 negative and are often called triple-negative (40). Normal-like tumors resemble the normal breast tissue, thus express many of the same genes as expressed in the normal breast tissue, such as genes in the adipose tissue and other non-epithelial cell types (41). Also at other molecular levels breast cancers have been classified into different groups. Both DNA methylation and microRNAs (miRNA) studies have shown that the DNA methylation level and the miRNA expression are associated with the gene expression subtypes (42-45). Moreover, more subtypes than detected by gene expression have been identified by using copy number (46;47) and proteins (48). Molecular classification reflects the heterogeneity of breast cancers. These subtypes have different clinical outcomes

and respond differently to treatment; basal-like tumors have the worst prognosis and luminal A tumors the best prognosis (37).

To characterize tumors' subtype based on gene expression was recommended by the 12th St Gallen International Breast Cancer Conference to be used in the clinic (49). However, the clinical routine of gene expression still remains to be implicated (49). Thus, ER, PR, HER2 and Ki-67 status are used instead for classification of subtype (49).

1.2.5 The tumor microenvironment

The tumor microenvironment in breast cancer surrounds the tumor cells and consists of stroma which includes fat cells, endothelial cells, pericytes, mast cells, macrophages, lymphocytes and fibroblasts as well as extracellular matrix (ECM) (50). The ECM provides structural support for the surrounding cells and consists of mainly 3 types of molecules: 1) fibers: collagen and elastin, 2) specialized proteins like fibrillin and fibronectin and 3) proteoglycans (51). It is now recognized that the tumor microenvironment have a role in tumor initiation and progression (28;50).

Allinen *et al*. (52) demonstrated by studying gene expression profiles on different cell types in normal breast tissue, *in situ* carcinomas and invasive carcinomas that changes in gene expression occurred in all types of cells during cancer progression. They also demonstrated that the chemokine C-X-C motif ligand 14 (CXCL14) is overexpressed in myoepithelial cells while the chemokine C-X-C motif chemokine 12 (CXCL12) is overexpressed in myofibroblasts (52). These chemokines bind to receptors on epithelial cells and enhance proliferation, migration and invasion, suggesting that chemokines play a role in tumorigenesis (52).

In another study by Hu *et al*. (53), DNA methylation was determined for the epithelial cells, the myoepithelial cells and stromal fibroblasts from normal breast tissue and *in situ* and invasive carcinomas. The DNA methylation was altered in all cell types during breast cancer progression, proposing that DNA methylation changes in stromal fibroblasts, is likely to be important for establishing the tumor microenvironment and for tumor progression (53).

Similar, in a study by Fiegl *et al*. (54), it was reported that the DNA methylation of stroma in HER2+ and HER2- tumors was significantly different for the three genes: progesterone receptor (*PGR*), hydroxysteroid (17-beta) dehydrogenase 4 (*HSD17B4*) and H-cadherin (*CDH13*). Since the DNA methylation of stroma in HER2+ and HER2- tumors differed, it has

been suggested that it is possible that the DNA methylation of stroma might be different between different types of tumors (28).

Many studies use bulk tumors (containing tumor cells and surrounding tissue) when performing analysis, for instance DNA methylation analysis. This might be challenging as the non-neoplastic cells might have a different DNA methylation level than the tumor cells, which can make it difficult to interpret the data.

1.3 Epigenetics

The major epigenetic mechanisms can be divided into histone modifications, chromatin remodeling complexes, and DNA methylation (55). Epigenetic mechanisms are known to be important for normal development, but they have also been implicated in different human diseases, such as cancer, Beckwith-Wiedemann syndrome and Rett Syndrome (6). Increasing evidence propose that epigenetic mechanisms are influenced by environmental exposure (6). For instance, DNA methylation and chromatin structure may be altered by environmental toxins like nickel and arsenic (56). These changes may again alter the expression of specific genes.

1.3.1 DNA methylation

DNA methylation is a result of the attachment of a methyl group to the 5' position of cytosine (Figure 5), mainly on CpGs (cytosine and guanine only separated by a phosphate) (55). During this process, the methyl group is transferred to the 5'cytosine from S-adenosylmethionine (SAM) induced by DNA methyltransferases (DNMT) (55). As of today four DNMTs implicated in the process have been identified (DNMT1, DNMT2, DNMT3A and DNMT3B) (55).

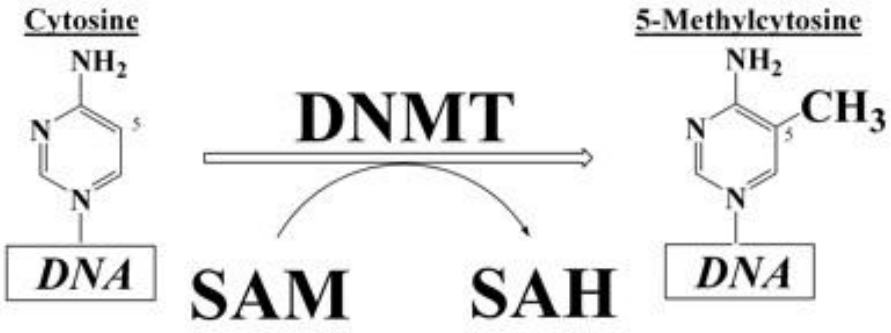


Figure 5: DNA methylation induced by methyltransferase (DNMT). The methyl group is transferred from SAM and the end result is 5-Methylcytosine and SAH (57).

Around 1% of all bases in the mammalian genome consist of 5-methylcytosine (55). Overall, most CpGs in the human genome are methylated (55). Despite this trend, CpGs in CpG islands (that are found in promoter region of many genes) are often unmethylated (55). CpG islands are DNA stretches of at least 200 base pairs (bp) consisting of a C+G content higher than 50% (55). DNA methylation is important for proper normal development (55). It is crucial for genomic imprinting, inactivation of X-chromosome through imprinting and silencing repetitive DNA and endogenous transposons (55). In this way, DNA methylation has a role in maintaining the genomic stability.

1.3.2 DNA methylation and breast cancer

The common view is that cancer is hypomethylated globally and hypermethylated in CpG islands (6). Hypomethylation might cause activation of oncogenes and genomic instability (58) while hypermethylation might cause silencing of tumor suppressor genes (6), all which might lead to the development of cancer. However, the DNA methylation in cancer is more complex than this common view. For instance, methylated promoters in breast cancers have shown to be both positively and negatively correlated to gene expression (43;59). In addition, both hypomethylation and hypermethylation have been shown in the promoter regions of several genes in breast cancers (60). Promoter methylation of among other the genes (*BRCA1*) (61), ras association (RalGDS/AF-6) domain family 1 (*RASSF1A*) (62), *CDH13* (63), cyclin-dependent kinase inhibitor 2A (*CDKN2A*) (64), fragile histidine triad (*FHIT*) (65), glutathione S-transferase pi 1 (*GSTP1*) (66), and phosphate and tensin homolog (*PTEN*) (67) have been reported. Similar, unmethylated promoters in among other the genes flap structure-spesific endonuclease 1 (*FEN1*) (68), cadherin-3 (*CDH3*) (69), interleukin 8 (*IL8*) (60), histone deacetylase 1 (*HDAC1*) (60), B-cell receptor-associated protein 31 (*BCAP31*) (60) and trefoil factor 1 (*TFF1*) (60) have been reported.

As previously mentioned, the DNA methylation profile of cancer related genes has been shown to be associated with the gene expression subtypes (42-44). All of these studies have used unsupervised hierarchical clustering to classify the tumors (42-44), and one example is shown in Figure 6. Two of these studies have also shown that the clusters are associated to clinical parameters. In both Holm *et al*. (42) and Ronneberg *et al*. (43) the clusters were associated with ER status. In addition, in Ronneberg *et al*. (43) the clusters were associated with tumor protein p53 (*TP53*) status, HER2 status and grade (43).

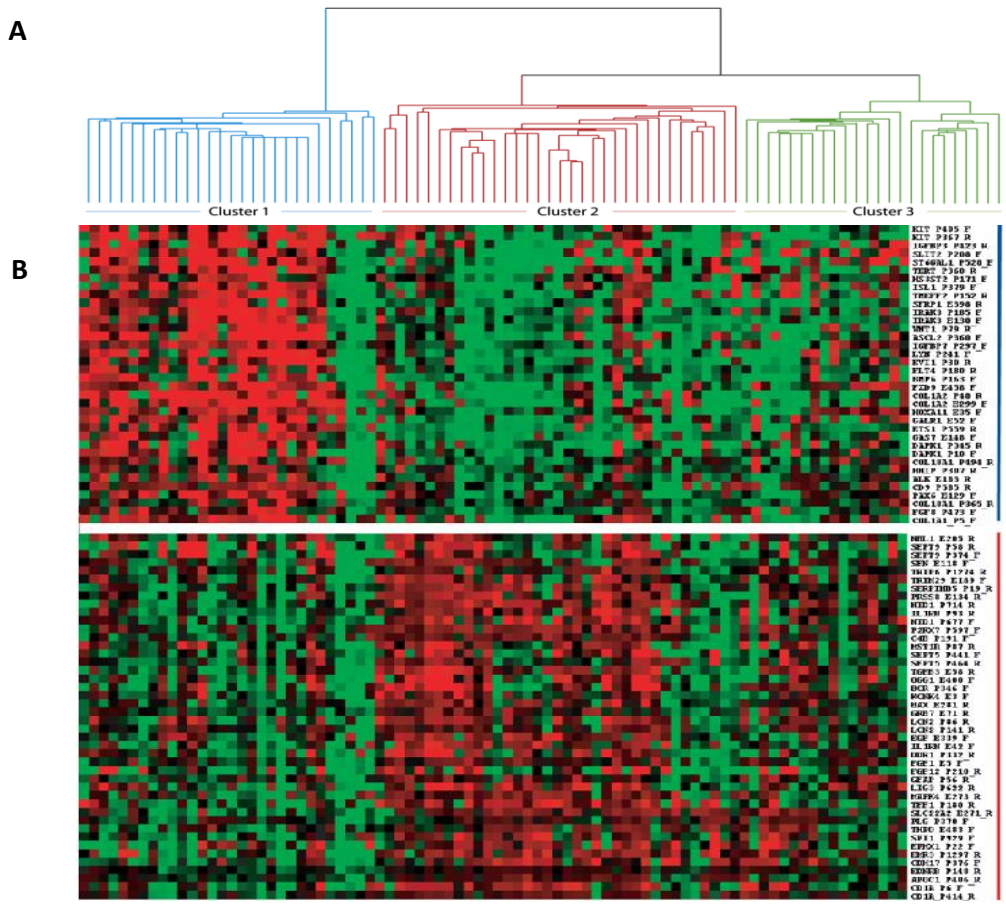


Figure 6: Hierarchical clustering based on the DNA methylation profile of cancer related genes. A) The tumors were divided into 3 clusters based on the DNA methylation in 80 breast tumors and 4 normal breast tumors of 664 genes with 1016 CpG sites. The three clusters consist of 1: mostly luminal tumors, 2: mostly basal-like and HER2 enriched tumors and 3: mostly luminal A tumors (43). B) Heatmap of the CpGs that are most differently methylated between the clusters. The CpGs are marked with green or red, illustrating relative low and high DNA methylation, respectively. The genes in the upper part of the heatmap marked with a blue vertical bar are highly methylated in cluster 1 compared to cluster 2 and 3. Similar, the genes in the lower part of the heatmap marked with a red vertical bar are highly methylated in cluster 2 compared to 1 and 3 (43).

Research on DNA methylation as a biomarker is ongoing. Cancer biomarkers are substances or processes used for cancer assessment. Biomarkers can mainly be used to indicate 5 features: 1) risk assessment, 2) detection and diagnosis 3) determine prognosis, 4) predict response to therapy and 5) monitor recurrence. The benefit of using DNA methylation is that DNA is relatively stable over time (70). Another benefit is that DNA methylation is an early event in breast cancer (71). In addition, many technologies can be used for measuring DNA methylation in many types of samples, like frozen tissue, blood and body fluids (such as plasma and urine) (70).

Recent studies have shown the potential for using DNA methylation as biomarkers. In Ronneberg *et al*. (43) tumors from different patients were clustered into three groups based on

their DNA methylation profile. The three clusters differed in survival, i.e. an example of how to use DNA methylation as a prognostic biomarker. In another study by Evron *et al*. (72) promoter methylation in the genes cyclin D2 (*CCND2*), retinoic acid receptor beta (*RAR-β*) and basic helix-loop-helix transcription factor (*TWIST*) in cells from ductal fluid were compared and cancer-specific DNA methylation in patients with invasive breast cancer and ductal *in situ* carcinoma breast cancer was detected. Besides, cells from healthy women were also used, and in two of the women, abnormal DNA methylation was identified (72). These women was later diagnosed with breast cancer (72). This example shows that DNA methylation can be used to detect cancer in asymptomatic patients. Even though DNA methylation has shown great potential as a biomarker, there are challenges with using DNA as a biomarker. For instance, many of the technologies measuring DNA methylation, like arrays and especially sequencing produce huge amount of data which still makes the interpretation difficult (73).

Unlike genetic changes, DNA methylation and other epigenetics mechanisms are reversible. Thus, research on drugs that inhibit the DNA methyltransferase activity which again might lead to reactivation of silenced genes are on-going. Many drugs shown to be great inhibitors of the activity of DNA methyltransferase, like hydralazine, procaine, procainamide and the azanucleosides: azacytidine (5-azacytidine) and decitabine (2'-deoxy-5-azacytidine) have been developed (74). Treatment with azacytidine and decitabine has increased the overall survival of patients with myelodysplastic syndrome (MDS) (75;76). In addition, low-dose treatment with decitabine has shown good response for acute myeloid leukemia patients (AML) (77). Also, procainamide and hydralazine have shown to inhibit DNA methylation in human cancer cell lines (78). Treatment of MCF-7 breast cancer cell lines with procaine have been reported to induce DNA demethylation and the expression of silenced suppressor tumor genes (79). A problem with these agents is that they not specifically inhibit DNA methylation of only the target sequence but likely other parts of the genome as well (7). As a result other genes may become expressed as well.

1.4 Heterogeneity

Breast cancer is a heterogeneous disease associated with different clinical outcome. The heterogeneity is divided into intra-tumor heterogeneity and inter-tumor heterogeneity. Intra-tumor heterogeneity is variation within a tumor, while inter-tumor heterogeneity is variation between patients (80) (Figure 7). Both inter-tumor and intra-tumor heterogeneity might among other complicate the diagnosis and treatment. Therefore, understanding the

heterogeneity's role in tumor evolution and treatment response is important for giving the most accurate diagnosis and treatment for each breast cancer patient. One of the challenges is that a biopsy from a tumor is only a small part of the tumor, and might not necessarily give the adequate reflection of the biological features of the whole tumor (81). Further, what treatment a patient will be given is based on among other stage, grade and receptor status determined from a biopsy (81). In this way, the score from grade and stage as well as the receptor status might be misleading if minor subpopulations in other parts of the tumors with other biological features are not accounted for (81).

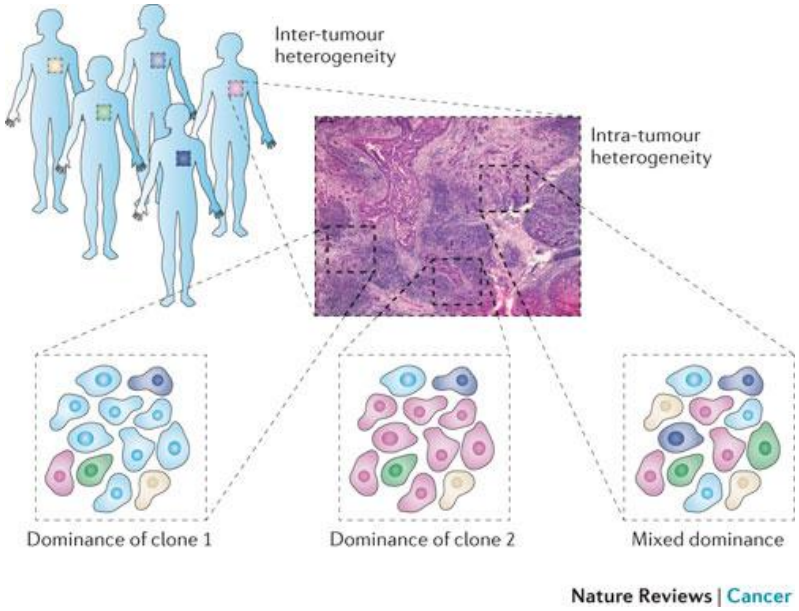


Figure 7: Illustration of inter-tumor heterogeneity and intra-tumor heterogeneity in breast cancer. This figure illustrates inter-tumor heterogeneity by showing several people with breast cancer that differs from each other, shown by marking the breast cancer with different colors. In addition, intra-tumor heterogeneity is illustrated by a biopsy from one patient which consists of three regions with different mix of cells (81).

1.4.1 Inter-tumor and intra-tumor heterogeneity

Inter-tumor heterogeneity can be observed at many levels, for instance at a phenotypical level. This might among other be caused by variations in DNA methylation between tumors, which as explained previously has been detected (42-44). Variations in DNA methylation between tumors might influence the gene expression, which again might influence the phenotype. Also at a genetic level, inter-tumor heterogeneity might be observed. For instance variations in copy number between tumors have been observed (46;47). The inter-tumor heterogeneity is also illustrated when deciding the diagnosis and treatment option of a tumor, as the tumors are classified into stage, grade etcetera.

The intra-tumor heterogeneity is thought to arise during tumor development. Different theories have been proposed, resulting in different types of intra-tumor heterogeneity (Figure 8). One of these is the clonal evolution theory, resulting in a tumor consisting of either one clone (monoclonal evolution) or several clones (polyclonal evolution) (80). In this theory, a tumor develops from one or several abnormal cells where repeated rounds of genetic, epigenetic and environmental changes that benefit the cells may lead to clonal expansion and formation of a tumor (4;28). All cells have the ability to contribute to tumor progression and drug resistance as they have the possibility to self-renewing division (28). Another theory is the cancer stem cell theory. In this theory, the tumor is proposed to arise from one or several stem cells or progenitors (4;80). Due to epigenetic, environmental or/and genetic changes these cells further evolve to cancer stem cells which give rise to differentiated cancer cells (4). At the same time, the cancer stem cells have the ability to self-renewing division (28). In this way, only the cancer stem cells contribute to tumor progression and drug resistance (28). In addition, a third model, the mutator phenotype, have been suggested (80). In this model, the tumors are proposed to progress from gradual and random accumulations of mutations, resulting in highly diverse tumors (80). All of these models explain how intra-tumor heterogeneity may arise, as all of these models might result in different subpopulations within a tumor (80).

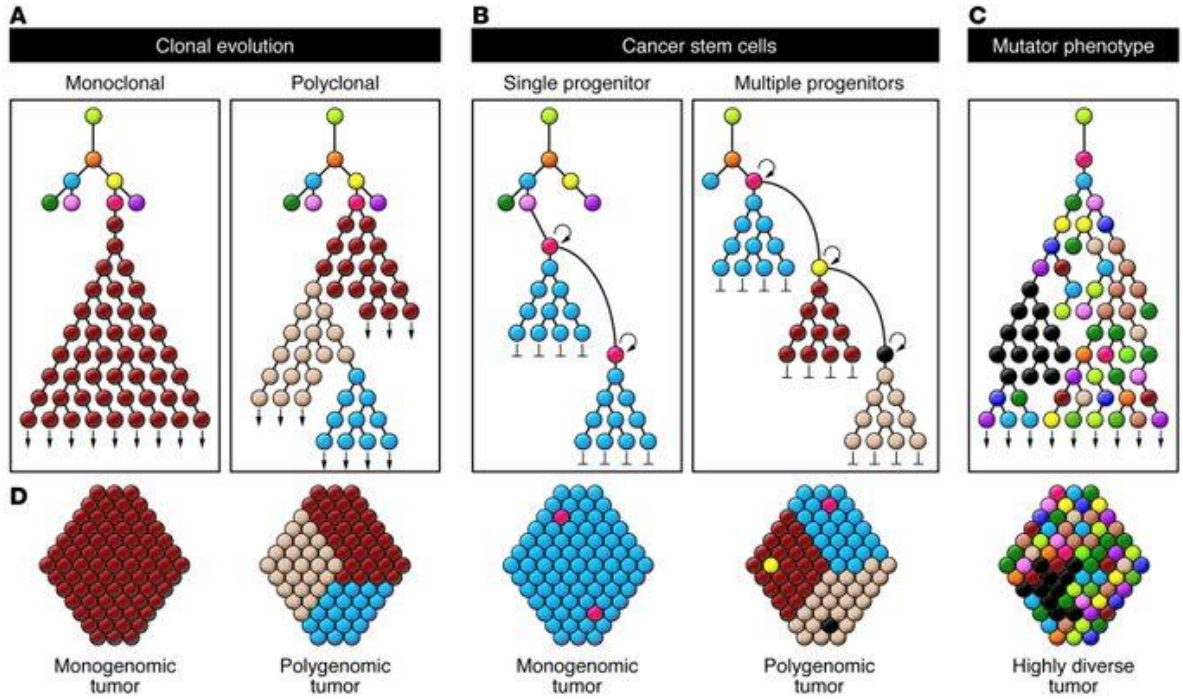


Figure 8: Hypothetical models of intra-tumor heterogeneity. The different models: clonal evolution (A), cancer stem cells (B) and mutator phenotype (C) can explain different types of intra-tumor heterogeneity (D) (80).

Like inter-tumor heterogeneity, intra-tumor heterogeneity can be observed at several levels. For instance, the heterogeneity within a tumor is well known by pathologists who have observed different morphology between different regions within a tumor (82;83). Intra-tumor heterogeneity has also been studied at a genomic level. For instance in Nik-Zainal *et al*. (84) the genomic history of 21 breast cancers was reconstructed. This was achieved by using newly developed bioinformatic algorithm (84). By exploring all genomic changes occurring within a tumor, they could define a pattern of which genomic changes that had most likely occurred at a given level during the development of a tumor (84). In another study by Shah *et al*. (85) next generation sequencing was also used which revealed that the fraction of cells with specific mutations varied between primary tumor and the metastases.

Other studies have looked closer at the intra-tumor heterogeneity at a phenotypical level. In these studies, cell membrane molecules often have been used, such as CD (cluster of differentiation) 24 and CD44. These cell membrane molecules are often associated with a characteristically feature within the cell (86). CD44+ and CD24+ have been suggested to be associated with stem-cell like properties and a more differentiated phenotype, respectively (87;88). Generally, higher content of CD24 cells than CD44 cells have been seen in luminal cells and vice versa for basal-like cells in breast cancer (89). It has been shown that CD44+ and CD24+ cells have differentially methylated and expressed genes (87;89). Genes implicated in motility, angiogenesis and chemotaxis are highly expressed in CD44+ cells, while genes implicated in the carbohydrate metabolism and RNA splicing were highly expressed in CD24+ cells (87). In addition, Shipitsin *et al*. (87) showed that the gene expression profile of CD44+ cells from normal breast tissue and breast cancer tissue are more similar to each other than the gene expression profile of CD24+ and CD44+ cells from the same tissue (meaning normal breast tissue or breast cancer tissue). The overall survival of patients with breast cancer associated with CD44+ has been shown to be worse than of patients with breast cancer associated with CD24+ (87).

1.5 Genes used in the present study

In the present study 22 genes, including 13 protein-coding genes and 9 miRNAs were assessed, see Table 1.

Table 1: Overview of the 22 genes including 13 protein-coding genes and 9 miRNAs used in the study.

Gene	Description
BCAN	Brevican
BNIPL	BCL2/adenovirus E1B 19KD interacting protein like
CTSA	Cathepsin A
FOXC1	Forkhead box C1
FGFBP2	Fibroblast growth factor binding protein 2
GYPE	Glycophorin E
IL1A	Interleukin 1, alpha
IL1R2	Interleukin 1 receptor, type II
IL5RA	Interleukin 5 receptor, alpha
LTC4S	Leukotriene C4 synthase
PCK1	Phosphoenolpyruvate carboxykinase 1 (soluble)
RASSF1A	Ras association (RalGDS/AF-6) domain family member 1
CDKN2A	Cyclin-dependent kinase inhibitor 2A
MIR199A	MicroRNA 199a
MIR135B	MicroRNA 135b
MIR16-2	MicroRNA 16-2
MIR887	MicroRNA 887
MIR148A	MicroRNA 148a
MIR200C/141	MicroRNA 200c and microRNA 141
MIR17-92	miR-17-92 cluster host gene
MIR142	MicroRNA 142
MIR150	MicroRNA 150

There are several publications on these genes in cancers including breast cancer. The genes *BCAN*, *BNIPL*, *CTSA*, *FGFBP2*, *GYPE*, *IL1A*, *IL1R2*, *IL5RA*, *LTC4S* and *PCK1* have been shown to be differently methylated between breast cancer samples between before and after radiation treatment (90). Of these, *BCAN*, *BNIPL*, *CTSA*, *FGFBP2*, *GYPE*, *IL5RA*, *LTC4S* and *PCK1* were part of the top ten genes most significantly different methylated between samples before and after radiation (90). The DNA methylation level of *LTC4S*, *CTSA*, *IL5RA*, *BCAN*, *GYPE* and *PCK1* have been associated with good response to treatment in the samples before radiation, after radiation or both (90).

The expression of the miRNAs has been found to be associated with DNA copy number alteration and/or DNA methylation in breast cancer (91). The expression of MIR200C/141, and MIR17, MIR19B, and MIR20a, which are part of the MIR 17-92 cluster, have been reported as upregulated in malignant breast tissue compared to normal tissue (92). Moreover, the expression of MIR199A has been reported as downregulated in malignant breast tissue compared to normal tissue (93). Interestingly the promoter of MIR148A was reported as hypermethylated in lymph node metastatic cancer cells and the expression of the miRNA was

inactivated (94). In normal tissue the same miRNA was hypomethylated (94). By treating the cancer cells with DNA demethylation agents, the expression of the same miRNA was reactivated (94). Followed by this reactivation, tumor growth was reduced and metastases in xenograft models was inhibited (94).

The genes *RASSF1A* and *CDKN2A* are tumor suppressors (95), and promoter methylation of these genes have been reported in breast cancer (62;64). *FOXC1* has been reported as differently methylated in the promoter region between normal samples versus DCIS (96) and as hypomethylated in the promoter region and expressed in CD44+ breast cancer cells (89). In addition, different DNA methylation levels in the promoter region has been shown between early and late stage breast cancers for the genes *FOXC1*, *RASSF1A* and *CDKN2A* (97). CpGs in *RASSF1A* and *FOXC1* has also been shown to be differently methylated between the subtypes or different clusters with different amount of the different subtypes (43;44).

2 Aim

The present study was designed as a pilot study where the main aim was to investigate the impact of tumor heterogeneity on DNA methylation levels in breast cancer. Few studies have investigated this association, and such associations may have important implications for both existing and future studies. The specific aims were to:

1. Study the relationship between DNA methylation level and tumor percentage genome-wide for 113 bulk tumors
2. Study the difference in DNA methylation level between regions with different fraction of tumor cells in each of 3 tumors for a selected panel of genes
3. Study the difference in DNA methylation level between subpopulations with different CD24/CD44 expression patterns within a cell line and between subpopulations with the same CD24/CD44 expression patterns across cell lines with different molecular subtype in a panel of breast cancer cell lines for the genes *RASSF1A* and *FOXC1*

3 Materials

In this part, the materials and a brief description of the methods previously performed on the materials are described. The materials can be divided into two parts: breast cancer cell lines and breast tissue samples. The present study includes both the continuation on a prior project regarding the breast cancer cell lines, an original wet lab and a bioinformatical analysis of already available data, Figure 9.

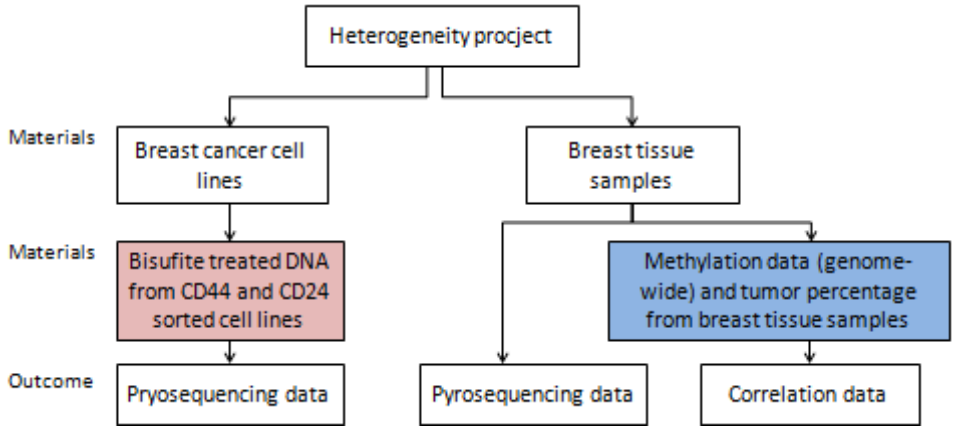


Figure 9: Flowchart of the project. Pink: materials from a previous study and blue: already available data.

3.1 Breast cancer cell lines

From a study at University of Barcelona, bisulfite-treated DNA from CD24 and CD44 sorted cells from basal-like and luminal cell lines were available (98). The breast cancer cell lines available were: MCF-7, ZR-75-1, MDA-MB-453, MDA-MB-468, MDA-MB-231 and T47D, and their characteristics are given in Table 2. The outcome of the cell sorting is shown in Table 3.

Table 2: Characteristics of the breast cancer cell lines included in this thesis (99-101). Subtype, ER, PR, HER2, source and tumor type are listed. The subtype of a cell line is approximated by the expression of estrogen, progesterone and human epidermal growth factor (49).

Cell line	Subtype	ER	PR	HER2	Source	Tumor Type
MCF-7	Luminal	+	+	-	Pleural effusion	Invasive ductal carcinoma
ZR-75-1	Luminal	+	-	+	Ascites fluid	Invasive ductal carcinoma
MDA-MB-453	Luminal	-	-	+	Pericardial effusion	Adenocarcinoma
T47D	Luminal	+	+	-	Pleural effusion	Invasive ductal carcinoma
MDA-MB-468	Basal-like	-	-	-	Pleural effusion	Adenocarcinoma
MDA-MB-231	Basal-like	-	-	-	Pleural effusion	Adenocarcinoma

Table 3: The CD24/CD44 sorted breast cancer cell lines. An overview over the cell lines and their CD44/CD24 expression patterns, average of DNA concentration and cell number are shown as well as if they were sorted and/or concentrated using a SpeedVac®.

Cell line name	CD44/CD24 patterns	Sample ID	Average of DNA concentration (ng/µl)	Average of number of cells	Sorted	Samples concentrated using SpeedVac®
MDA-MB-453	CD44-/CD24+	1-9	144.0	4566666.7	No	No
ZR-75-1	CD44-/CD24+	10-12	253.7	4583333.3	No	No
MDA-MB-468	CD44+/CD24+	13-18	285.4	5000000.0	No	No
MDA-MB-468	CD44-/CD24+	19,22	16.4	15823.5	Yes	Yes
MDA-MB-468	CD44-/CD24+	20,21	88.6	1012682.5	Yes	No
MDA-MB-231	CD44+/CD24-	23, 25-26	104.8	263508.7	Yes	No
MDA-MB-231	CD44+/CD24+	24	3.8	506.0	Yes	No
MDA-MB-231	CD44+/CD24+	27-28	13.7	NA*	Yes	Yes
MCF-7	CD44-/CD24+	32,35	53.5	903121.5	Yes	No
MCF-7	CD44-/CD24-	33,39,41	30.8	335546.7	Yes	No
MCF-7	CD44+/CD24-	34,40	49.5	663465.0	Yes	No
MCF-7	CD44+/CD24-	36	15.0	1169.0	Yes	Yes
MCF-7	CD44+/CD24+	37-38	19.9	684.5	Yes	Yes
T47D	CD44-/CD24+	42,44-45	80.8	617191.0	Yes	No
T47D	CD44-/CD24-	43,47,49	24.62	141428.7	Yes	Yes
T47D	CD44+/CD24+	46,48	36.5	40672.5	Yes	Yes

* NA: not available

In this previous study, 3-9 replicas of the six cell lines described in Table 2 were cultured and consequently submitted to flow cytometry analysis to identify subpopulations of the cells based on CD44/CD24 expression pattern. For this purpose, the cells were incubated with the primary antibodies anti-CD44-FITC (BD Bioscience) and anti-CD24-R-PE (BD Bioscience) to target the antigens respectively and sorted in a BD Biosciences FACSAria III cell sorter (BD Bioscience). Cell sorting was not necessary for MDA-MB-453 and ZR-75-1, since they both presented a high percentage (>96%) of the major subpopulation CD44-/CD24+. Only some of the four subpopulations (CD44-/CD24+, CD44+/CD24+, CD44-/CD24-, CD44+/CD24-) were present in the cell lines expect for MCF-7. On these sorted cells, DNA was isolated using QIAamp DNA mini kit (Qiagen) (102), and the DNA was quantified by using the Nanodrop® ND-1000 Spectrophotometer (Saveen Werner A/S) (103). The low concentration of DNA in some of the fractions was increased by using a DNA120 SpeedVac® (Thermo Savant). The isolated DNA was bisulfite-treated by using the Epitect bisulfite kit (Qiagen) (104). As a result methylated and unmethylated cytosines were distinguished which is required when studying DNA methylation level. Bisulfite treatment was performed three

times to enable enough DNA for downstream methods. These samples will be further used in the lab experiments.

3.2 Breast tissue samples

3.2.1 Macrodissected breast tumor

Macrodissection and DNA methylation analysis by pyrosequencing was performed on breast tumor tissue from three breast cancer samples, see Table 4. In addition, tissue from three healthy women that went through breast tissue reduction was available and used. These samples are called RP2, RP10 and RP11. These samples will be further used in the lab experiments.

Table 4: The breast cancer samples. The different receptors expressed and the TNM classification are listed, as well as the number of lymph nodes checked and how many of these that was detected with cancer.

Breast cancer sample	**ER**	**PR**	**T[1] (mm)**	**N[2]**	**M[3]**	**Lymph nodes analyzed**	**Lymph nodes with detected tumor cells**
137	-	-	T3(60)	N2	M0	10	5
142	+	+	T3 (51)	N2	M0	8	4
155	-	-	T4 (51)	N2	M0	10	4

TNM classification:

[1] T (primary tumor): T3: tumor>5cm, " T4: tumor of any size with direct extension to a) chest wall or b) skin" (105).

[2] N (lymph nodes): "N2: Metastasis to ipsilateral axillary lymph node(s) fixed to each other or to other structures" (105).

[3] M (metastasis):" M0: No distant metastasis" (105).

3.2.2 Bulk tumors from breast cancer samples

Whole genome methylation data and estimated tumor percentage were available for bulk tumors from breast tumor tissue from 200 breast cancer samples. Of these samples, 113 (explained under **chapter 4.6.1**) where used and showed in Table 5.

Table 5: The 113 breast cancer samples. An overview over the patient characteristics and the patients' ratio and tumor % range are listed.

Characteristic	**Patients (ratio)**	**Tumor % range**
PHENOTYPE		
DCIS	7/113	21-85
Mixed	8/113	32-85
Invasive	98/113	25-86
Expression subtype		
Luminal A	26/113	25-78
Luminal B	20/113	25-82

Characteristic	Patients (ratio)	Tumor % range
HER2 enriched	8/113	32-85
Basal-like	31/113	32-86
Normal-like	13/113	21-73
Unknown	15/113	37-76
ER		
+	76/113	21-86
-	28/113	25-85
Unknown	9/113	32-76
PR		
+	57/113	21-86
-	46/113	25-85
Unknown	10/113	32-76
Her2		
+	11/113	26-58
-	67/113	21-86
Unknown	35/113	21-58
Grade		
1	9/113	42-66
2	60/113	26-86
3	33/113	25-85
Unknown	11/113	21-76
T stage		
T1	33/113	26-86
T2	37/113	25-82
T3	5/113	25-75
T4	2/113	44-55
Unknown	36/113	21-85
N stage		
+	41/113	26-81
-	34/113	25-86
Unknown	38/113	21-85

Previous methods performed on the material are whole genome methylation and SNP analysis. The Infinium® HD Assay Super Protocol Guide (Illumina) (106) and the Infinium® Assay Methylation Protocol Guide (Illumina) (107) were used for the SNP analysis and the DNA methylation analysis, respectively.

The Infinium® HumanMethylation450 BeadChip covers 485764 methylation sites distributed in CpG islands, CpG shores, CpG shelves and isolated CpGs (108). The BeadChip covers CpGs in the promoter region (200 bases or 1500 bases upstream of the transcription start site (TSS200, TSS1500), 5'UTR, first exon) the gene body ,3'UTR and intergenic sequences (108).

The Infinium® HD Human660w-quadBeadChip covers 657366 genomic markers for SNP and comparative genomic hybridization (CGH) (109). The whole genome SNP data was analyzed using the bioinformatic approach, allele-specific copy number analysis of tumors (ASCAT) (described in detail in (110)) to determinate tumor percentage of each of the 113 tumor tissue samples.

The DNA methylation- and tumor percentages will be further used in the bioinformatical analysis.

4 Methods

In this part, the laboratory- and statistical/bioinformatical methods used in the present study are described. The suppliers with internet address, purity and catalog number/article number for reagents and equipments as well as suppliers for the instruments are presented in **Appendix A**

From **chapter 4.1** to **chapter 4.5** the description regards only the breast tumor tissue from 3 breast cancer samples.

4.1 Macrodissection, cryomicrotome, haematoxylin and eosin staining and microscopy

Fresh frozen breast tumor tissue was cut into small sections using a cryomicrotome. The sections were microscopic evaluated, which was executed in collaboration with a pathologist. Prior to microscopy the sections were stained with haematoxylin and eosin (HE-stained). HE-staining involves haematein (oxidant of haematoxylin) combined with aluminums ions which stains the nuclei of the cells purple (111). Eosin stains positive charged proteins in cytoplasm and connective tissue pink/red (111). When looking at a HE-stained section of a tumor sample in a microscope, the tumor cells are recognized by color, shape and size. Their nuclei are colored purple and are often big in size and have an irregular shape. Around the nuclei, cytoplasm can be seen in pink/red. Based on the microscopic evaluation, the tumors were macrodissected into different regions and each region was sectioned for DNA isolation.

Procedure

Figure 10 presents the workflow on how this part was performed. The cryomicrotome Leica CM1950 was used to make 5 µm sections of each of the three tumor samples: 137, 142 and 155. In this procedure, Tissue-Tek was used to mount the tissue on a plate that follows the machine to allow steady sectioning. The sections were put on microscope slides and HE-staining was performed (see **Appendix B** for recipe).The HE-stained sections were evaluated by microscopy, and based on varying amount of tumor cells in different regions of each of the tumor samples, the tumors were macrodissected into different regions, Figures 11-13. When defining which cells the different regions consist of, lymphocytes have been distinguished from tumor and stroma. The reason is that lymphocytes are transient cells. They circulate in the blood stream and can migrate out of the blood stream upon stimulating signals. In this way lymphocytes can be found in many places, both stroma and tumor.

For tumor 137 three regions were selected, as shown in Figure 11. Region B contained mainly stromal cells. In addition, a few normal epithelial cells as well as scattered tumor cells were present (not shown in Figure 11). Region C contained mostly tumor cells but also some lymphocytes and a few stromal cells were present. Region D had a lot of lymphocytes as well as some stromal cells and tumor cells. For tumor 155 two regions were selected, as shown in Figure 12. Region A had a high number of stromal cells, as well as some tumor cells and a few lymphocytes. Region C had a high tumor percentage, but a few stromal cells were also seen. For tumor 142, three regions were selected, shown in Figure 13. A high number of tumor cells and few stromal cells were found in region A, while there were more stromal cells and less tumor cells in region B. Region C contained stromal cells as well as normal epithelial cells. In addition, some tumor cells were present (not shown in the Figure 13).

For each region, a 5 µm section was made for HE-staining to make sure that the cutting of the tissue was correct. Then, for each region, 30-45 sections of 20 µm were made and allocated into three separate tubes for DNA isolation. Further, new 5 µm sections were made for HE-staining to see if there were any changes in the amount and type of tissue. Small differences in amount and type of cells and tissue were observed in 137 and 155 when comparing HE-staining from both ends of the included tumor piece for DNA methylation. The regions which consisted of mostly stroma at one end of the included tumor piece, contained a few more tumor cells further down in the tumor. For sample 142 on the other hand, the amount and type of tissue changed a lot further down in the tumor. Region A consisted of more stromal cells and region C consisted of more tumor cells further down in the tumor. In this way, the 3 regions consisted of more or less the same amount of tumor cells further down in the tumor. The regions were defined to belong to one of 5 groups based on the amount of tumor cells from both HE-sections, Table 6. The groups from least to most tumor were: 1 = least tumor, 2 = less tumor, 3 = some tumor, 4 = more tumor and 5 = mostly tumor.

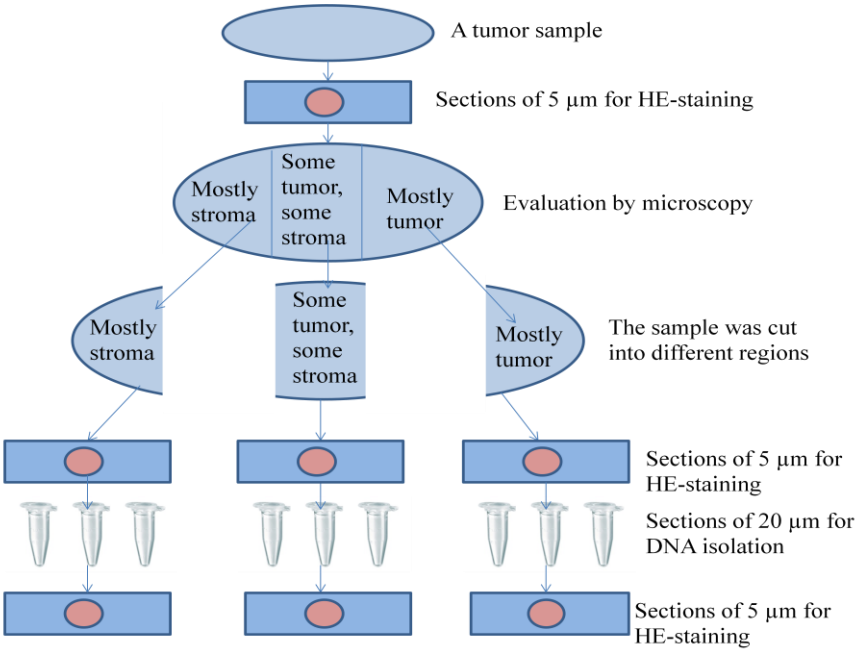


Figure 10: Workflow showing the steps for how to isolate tumor tissue from different regions within a tumor sample. First the tumor sample was sectioned, HE-stained and evaluated by microscopy. Then, the tumor sample was cut into different regions containing varying number of tumor cells and stroma. Each region was sectioned for DNA isolation. Sections for HE-staining from both ends of the included tumor piece were made.

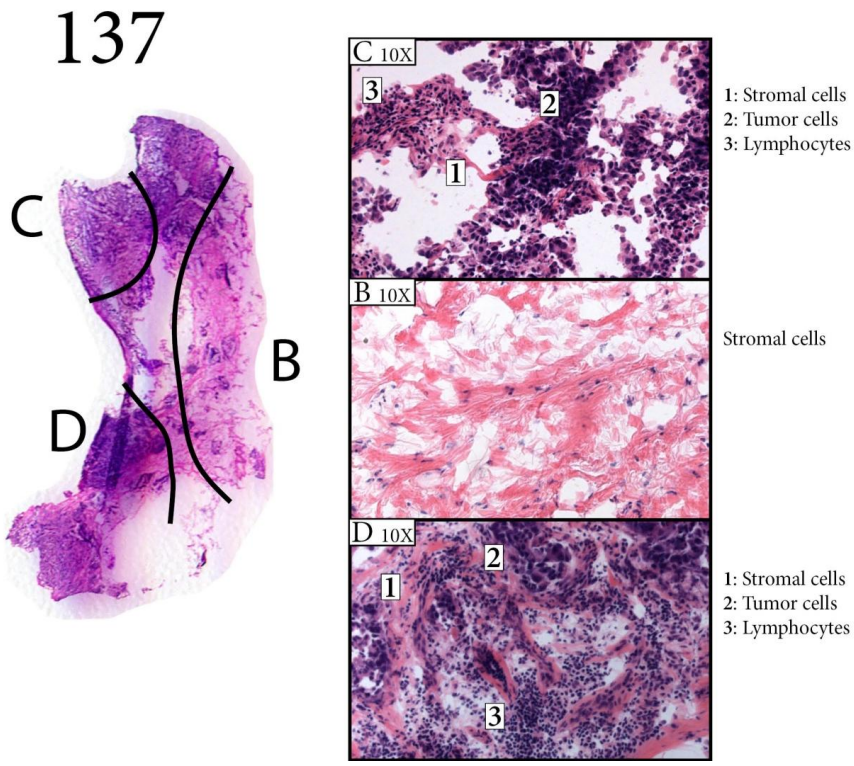


Figure 11: Image by light microscopy (microscope: Olympus BX45) of the tumor sample 137. To the left side of the figure, the HE-stained tumor samples are shown with the regions that were selected to be studied for DNA methylation levels individually (denoted B, C or D). The selected regions are shown in enlargements to the right. The numbers 1, 2 and 3 refers to enriched regions with specific cell types.

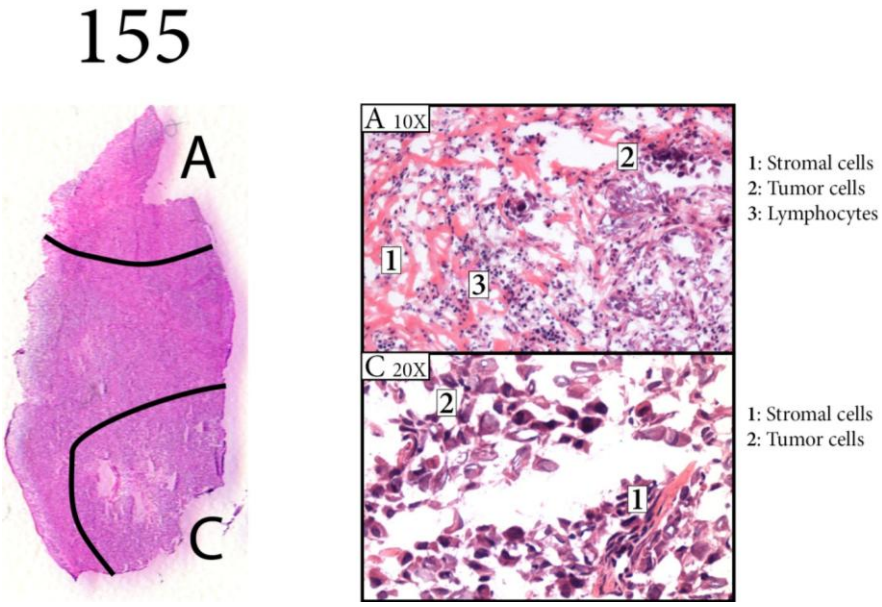


Figure 12: Image by light microscopy (microscope: Olympus BX45) of the tumor sample 155. To the left side of the figure, the HE-stained tumor samples are shown with the regions that were selected to be studied for DNA methylation levels individually (denoted A or C). The selected regions are shown in enlargements to the right. The numbers 1, 2 and 3 refers to enriched regions with specific cell types.

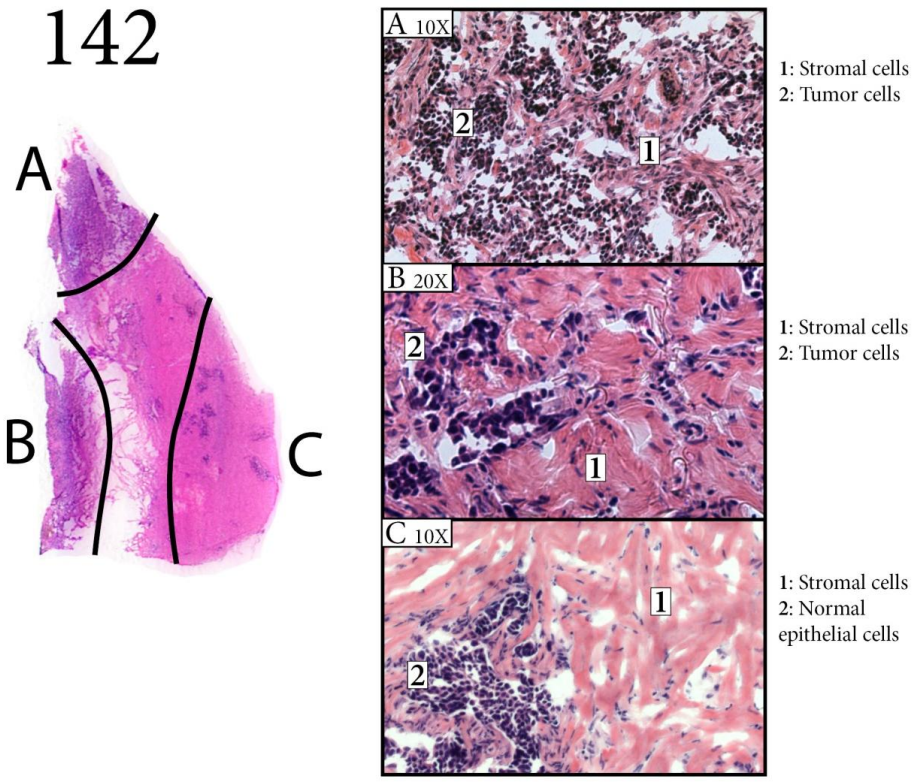


Figure 13: Image by light microscopy (microscope: Olympus BX45) of the tumor sample 142. To the left side of the figure, the HE-stained tumor samples are shown with the regions that were selected to be studied for DNA methylation levels individually (denoted A, B or C). The selected regions are shown in enlargements to the right. The numbers 1 and 2 refers to enriched regions with specific cell types.

Table 6: The definition of the regions within the tumors based on the amount of tumor cells. The regions were defined into one of 5 groups based on the amount of tumor cells from both HE-staining sections. The 5 groups from least to most tumor were: 1 = least tumor, 2 = less tumor, 3 = some tumor, 4 = more tumor and 5 = mostly tumor

Tumor	Region	Definition
137	B	1= Least tumor
	D	3= Some tumor
	C	5= Mostly tumor
155	A	2= Less tumor
	C	5= Mostly tumor
142	C	2= Less tumor
	B	3= Some tumor
	A	4= More tumor

4.2 DNA isolation by MAXWELL® 16 Instrument

In this method, the Maxwell® 16 Instrument was used to isolate DNA from the breast cancer tissue. DNA is isolated by using MagneSilR paramagnetic particles which optimizes capture, elution and washing of the sample (112). The sample is transferred between different wells with different solutions with help from a plunger (112). The first well consists of lysis buffer, the second well of MagneSilR paramagnetic particles and in the rest of the wells there are washing buffers (112). In well two, MagneSilR paramagnetic particles will be picked up by the plungers (112). In well one, the plungers will help destructing the tissue and cells, and DNA that binds to the paramagnetic particles will be picked up. In the rest of the wells, different washing steps will be performed. The DNA isolation procedure was performed according to Technical Manual Maxwell® 16 DNA Purification Kits (112).

4.3 Nanodrop

To quantify the DNA concentration and determine the pureness of the DNA samples the NanoDrop® ND-1000 spectrophotometer was used. This instrument measures the concentration of DNA by determining the absorbance of light at 260 nm. The concentration of the sample and the ratios 260/280 and 260/230, which are used to determine the purity of DNA (and also RNA), were calculated by the ND-1000 software version 3.8.1 (103). A 260/280 ratio under 1.8 for DNA suggests that there are presence of other components like proteins and phenols in the sample (103). A 260/230 ratio between 1.8-2.2 suggests that the sample is pure, and a lower number may indicate that the sample is contaminated (103) for instance with carbohydrates or guanidine (113). The procedure was performed according to NanoDrop 1000 Spectrophotometer V3.7 User's Manual (103).

4.4 Bisulfite treatment

Distinguishing between methylated and unmethylated cytosines is crucial for determination of DNA methylation status and is achieved in the present study by bisulfite treatment (114). By using this treatment the unmethylated cytosines will be converted to uracil, leaving the methylated cytosines unaltered (114). The conversion of unmethylated cytosine to uracil consists of several steps including denaturation, deamination, sulfonation and desulfonation (114). In a thermal cycler the DNA is denatured followed by being sulfonated and deaminated (114). Desulfonation is achieved in the clean-up process of the bisulfite-treated DNA (114). This clean-up process is also important for purifying the bisulfite-treated DNA by removing bisulfite salts and chemicals used in the conversion proves as well as desulfonation agents (114).

The first time this procedure was performed, a dilution series (1 ng,10 ng, 20ng, 50 ng, 100 ng, 200 ng, 300 ng, 400 ng, 500 ng) in a volume of 20 µl of DNA sample number 13 (**Appendix D**) was made to find the lowest concentration of DNA giving a good pyrosequencing result. Optimal results were achieved with 500 ng in 20 µl (data not shown) and this was used for the rest of the thesis.

Bisulfite treatment was performed 4 times to enable enough DNA for the downstream analysis according to the EpiTect® Fast 96 Bisulfite Conversion Handbook (114).

Procedure

Conversion process

- For each sample, DNA (500 ng), RNAse-free water, 85 µl Bisulfite solution, 35 µl DNA protect buffer were mixed in a ThermoFast 96 well PCR-plate. The total volume of DNA and RNAse-free water must be 20 µl and was adjusted dependent on the concentration of the DNA samples.
- The ThermoFast 96 well PCR-plate was capped with ThermoFast cap stripes, and the solution was mixed thoroughly.
- The ThermoFast 96 well PCR-plate was put on the GeneAMP® PCR System 2700 to execute the thermal profile of the bisulfite conversion which includes the steps: 5 minutes denaturation at 95 ºC, 20 minutes incubation at 60 ºC, 5 minutes denaturation at 95 ºC, 20 minutes incubation at 60 ºC and indefinite hold for 20 ºC.

Clean-up process of bisulfite-treated DNA

- Reagents preparations: 30 ml ethanol was added to buffer washing buffer (BW), and 27 ml ethanol was added to buffer desulfonation buffer (BD).
- The ThermoFast 96 well PCR-plate with samples from the GeneAMP® PCR System 2700 was centrifuged briefly and the complete reactions volume from each well was transferred to 1.5 ml Eppendorf tubes.
- 310 µl Buffer BL (contains a guanidine salt) was added to each sample. The solutions were mixed by vortexing before briefly centrifuged.
- 250 µl ethanol was added to each sample. The solutions were mixed by vortexing before briefly centrifuged.
- The mixture from each tube was transferred to Mini Elute DNA spin columns with collection tubes (hereafter denoted: spin columns).
- The spin columns were centrifuged at maximum speed (=13400 revolutions per minute (rpm)) for 1 minute. The flow-through was discarded and the spin columns were placed back into the collection tubes.
- 500 µl buffer BW was added to each spin column, and centrifuged for 1 minute at maximum speed. The flow-through was discarded and the spin columns were placed back into the collection tubes.
- 500 µl buffer BD was added to each spin column. After 15 minutes incubation the spin columns were centrifuged at maximum speed for 1 minute. The flow-through was discarded and the spin columns were placed back to the collection tubes.
- 500 µl buffer BW was added to each spin column, and the spin columns centrifuged at maximum speed for 1 minute. The flow through was discarded and the spin columns were placed back to the collection tubes. This step was repeated once.
- The spin columns were transferred to new collection tubes and centrifuged again at maximum speed for 1 minute.
- The spin columns were placed into new 1.5 ml Eppendorf tubes and 30 µl elution buffer was added to the column filter. After incubating for 1 minute, the tubes with the spin columns were centrifuged for 1 minute at 12000 rpm to elute the DNA.
- The NanoDrop® ND-1000 Spectrophotometer was used to determine the concentration of the samples and the samples with high concentration were diluted to approximately 15 ng/µl.

4.5 Analyzing level of DNA methylation

In this thesis the percentage of methylated cytosines in CpGs was determined by pyrosequencing® technology. This method is carried out on specific PCR amplified genomic regions.

The 22 genes including 13 protein-coding genes and 9 miRNAs (MIR 200C/141 and MIR17-92 miRNA are clusters) included in this study are shown in Table 7. The PCR- and sequencing primers used were designed by collaborators at Centre National de Génotypage, CEA-Institut de Génomique; Evry, France. Eurofins MWG Operon produced the primers. Two PCR products cover MIR142. To distinguish these two from each other they are named 142(1) and 142(2), as shown in Table 7. Replicates of the same subpopulation in a cell line were avaible and each CD24/CD44 sorted cell line (each replicate) was pyrosequenced in triplicates for *RASSF1A* and *FOXC1*. The tumor regions were also pyrosequenced in triplicates for the protein-coding genes. The miRNAs were available from another study (91) and a colleague at the Radium Hospital pyrosequenced the tumor regions for these miRNAs. The 5% of the total pyrosequencing runs were run in duplicates for the tumor regions for the miRNAs. These miRNAs were designed to match probes at Infinium® HumanMethylation450 BeadChip (Illumina). More detailed information on the primers and the CpG positions are shown in **Appendix C.**

Table 7: The genes included in this study. An overview of the genes including their chromosome (denoted: Chr), PCR amplified region and size of PCR product (denoted Size) are listed. For the protein-coding genes the related gene region are shown and for the miRNAs the corresponding probe at the Infinium® HumanMethylation450 BeadChip (denoted 450k array) with related miRNA region are shown. TSS1500 and TSS200: 1500 and 200 bases, respectively, upstream of the transcription start site of the miRNA mentioned. Body: the body of the gene. UTR: untranslated region.

Gene	Chromosome and PCR product	Size	Corre-sponding probe at the 450k array	Protein-coding gene region/miRNA region related to the probe
BCAN	1: 156612048-156612262	215		Intron 1
BNIPL	1: 151009541-151009755	214		Intron 1
CTSA	20: 44518768-44519064	297		5UTR
FOXC1	6: 1611034-1611248	215		Exon 1
FGFBP2	4: 15962780-15963077	297		5UTR/Exon 1
GYPE	4: 144793451-144793598	148		Exon 1
IL1A	2: 113533276-113533452	177		Exon1/Intron1
IL1R2	2: 102608130-102608305	176		5UTR
IL5RA	3: 3108999-3109234	236		5UTR/Exon 1
LTC4S	5: 179220910-179221185	276		5UTR/Exon 1

Gene	Chromosome and PCR product	Size	Corre-sponding probe at the 450k array	Protein-coding gene region/miRNA region related to the probe
PCK1	20: 56135850-56136126	277		5UTR/Exon 1
RASSF1A	3: 50378398-50378208	191		Exon 1
CDKN2A	9: 21974920-21974659	276		Exon 1/Intron 1
MIR199A	1: 172113895-172114186	292	cg24002149	TSS200
MIR135B	1: 205418231-205418469	239	cg13061767	TSS1500
MIR16-2	3: 160122450-160122618	169	cg09100593	TSS200
MIR887	5: 15500029-15500139	111	cg19632594	NA*
MIR148A	7: 25989666-25989855	190	cg03853208	TSS200
MIR200C/141	12: 7072539-7072761	223	cg27534624	TSS1500-MIR141 // TSS200-MIR200C
MIR17-92	13: 92002221-92002484	264	cg23665802	TSS1500
MIR142 (1)•	17: 56408439-56408731	293	cg00057966	TSS200
MIR142 (2)•	17:56408940-56409074	135	cg10530767	TSS1500
MIR150	19: 50004089-50004332	244	cg27388703	TSS200

*NA: not available
•MIR 142 (1) and MIR142 (2): the numbers in the brackets are added to distinguish between different amplified region of the same miRNA.

4.5.1 Polymerase chain reaction (PCR)

PCR is a method to amplify selected regions of DNA. Here the PyroMark® PCR kit was used (115).

Procedure

The PCR primer sequences are presented in **Appendix C**. Each PCR reaction was optimized for each region of the tumors and cell line sample. To achieve PCR products of the subgrouped cell lines that had been concentrated (overview of all samples in **Appendix D**) more volume into the PCR was needed to get a strong band on the agarose gel. Optimizing is important because proper amount of PCR-product are required in the pyrosequencing run. The master mix alternatives are presented in Table 8. The PCR was run in the DNA Engine Tetrad 2 Peltier Thermal Cycler with the alternative PCR programs shown in Table 9. The optimized conditions for all the PCR reactions are presented in Table 10. Negative and positive controls, Epitect methylated human control DNA (100% methylated) and Epitect unmethylated human control DNA (0% methylated) were used.

Table 8: The master mix alternatives. 6 master mix alternatives were used for the genes. Alt: alternative

PyroMark PCR Kit	Alt. 1 Volume (µl/well)	Alt. 2 Volume (µl/well)	Alt. 3 Volume (µl/well)	Alt. 4 Volume (µl/well)	Alt. 5 Volume (µl/well)	Alt. 6 Volume (µl/well)
PyroMark PCR Master Mix, 2x	12.5	12.5	12.5	12.5	12.5	12.5
CoraLoad	2.5	2.5	2.5	2.5	2.5	2.5
25 mM $MgCl_2$			0.5	1	2	2
Forward primer	0.17	0.17	0.17	0.17	0.17	0.17
Reverse primer	0.17	0.17	0.17	0.17	0.17	0.17
RNAse-free water	9.16	8.66	8.16	7.66	6.66	5.66
DNA	0.5	1	1	1	1	2
Total	25	25	25	25	25	25

Table 9: The PCR programs. 3 different PCR programs were used.

PCR program 1			PCR program 2			PCR program 3		
Temperature profile (ºC)	Time profile	No of cycles	Temperature profile (ºC)	Time profile	No of cycles	Temperature profile (ºC)	Time profile	No of cycles
96	15 min.		95	15min.		95	15min.	
96	45 sec.	40	94	30 sec.	46	94	30 sec.	35
X	45 sec.		X	30 sec.		X	30 sec.	
72	45 sec.		72	30 sec.		72	30 sec.	
72	7 min.		72	7min.		72	7min.	
8	∞		8	∞		8	∞	

Table 10: Optimized PCR conditions. P: program, An: annealing, Temp: temperature, alt: alternative

PCR P	An Temp	Master mix alt. 1	Master mix alt. 2	Master mix alt. 3	Master mix alt. 4	Master mix alt. 5	Master mix alt. 6
2	48			*IL5RA,GYPE, IL1A*			
	50		*IL1R2, LTC4S*	*PCK1, BNIPL, CTSA, FGFBP2,*			
	52		*BCAN*				
1	62	*FOXC1**	*FOXC1*, CDKN2A*				
	58	*RASSF1A**	*RASSF1A**				
3	50				MIR17-92 cluster		MIR16-2, MIR148A
	52			MIR199A MIR135B MIR887 MIR200C/141 cluster MIR142 (2) • MIR150		MIR 142(1)•	

•MIR 142 (1) and MIR142 (2): the numbers in the brackets are added to distinguish between different amplified region of the same miRNA.
**FOXC1* and *RASSF1A*: 1 µl DNA was used for the cell lines that were concentrated and the tumor regions, for the rest of the cell lines 0.5 µl were used.

The presence of PCR-product was analyzed by agarose gel electrophoresis (see **Appendix B** for recipe).

4.5.2 Pyrosequencing

Pyrosequencing is a quantitative method where the principle is based on sequencing by synthesis (116). In this method, when studying DNA methylation levels, bisulfite-treated DNA is used. First, the PCR-product is generated with one of the primers biotinylated, resulting in one of the strands being biotinylated (117). In the subsequent clean-up process using the PyroMark Q96 Vacuum Workstation, the biotinylated PCR-strand will bind to streptavidin beads which again will bind to the filter on the vacuum prep tool. During this clean-up procedure, the biotinylated strand of the PCR-product is separated from the un-biotinylated strand when submerged in PyroMark Denaturation solution. In the end, the probes on the vacuum prep tool are lowered into the pyrosequencing plate (containing the sequencing primers) and shacked. As a result, the streptavidin beads together with the template are released from the vacuum prep tool. In the PyroMark® Q96 MD instrument, the sequencing primer anneals to the template and deoxyribonucleotide triphosphates (dNTPs) are added one by one (117). A cascade reaction follows: for each dNTP that is added, a pyrophosphate (PPi) is released (117). PPi and adenosine- 5'-phosphosulfate (APS) will in presence of Adenosine- 5'-phosphate (ATP) sulfuryase convert to ATP (Figure 14A) (117). Luciferase will in turn utilize the ATP to convert luciferin to oxyluciferin which generates visible light in amounts that is proportional with the amounts of ATP, and the number of PPi groups released and through this the number of dNTPs incorporated (117). This light is detected by a charge couple device (CCD) camera, and can be seen as a peak in the instrument's software (pyrogram) (116;117). The amount incorporated dNTPs are equal the height of a peak seen in the pyrogram (117) (Figure 14B). The dNTPs and ATPs that are not incorporated are degraded by apyrase (117).

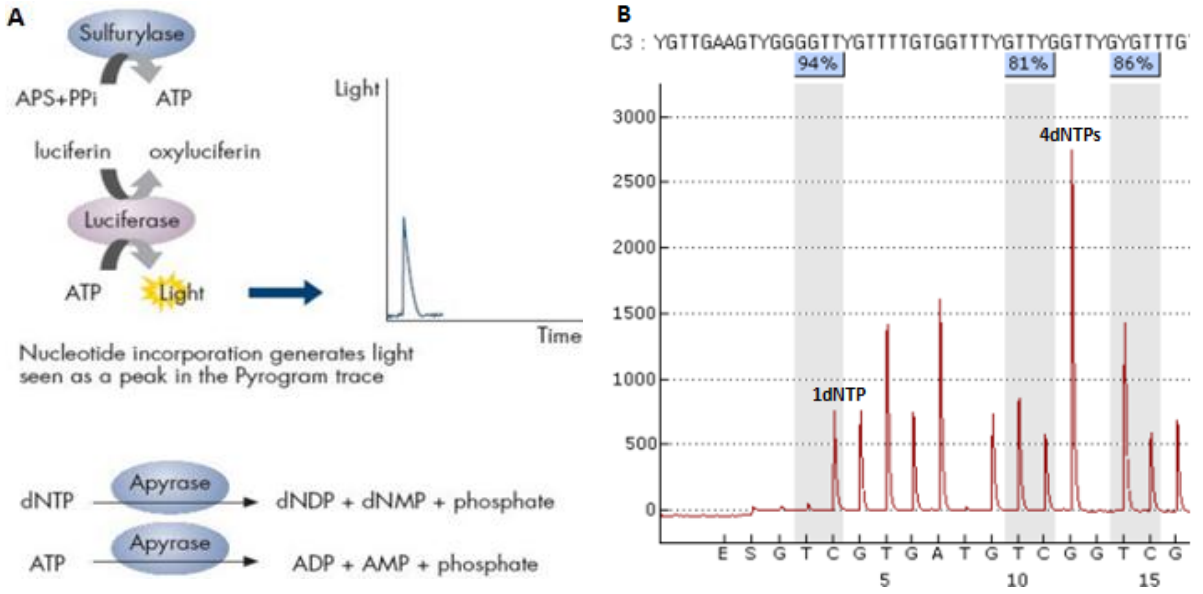


Figure 14: An overview over the pyrosequencing method. A) The cascade reaction from adding deoxyribonucleotide triphosphate (dNTP) to visible light. ADP: Adenosine diphosphate AMP: adenosine monophosphate, dNDP: deoxyribonucleoside diphosphate, dNMP: deoxyribonucleoside monophosphate, APS: adenosine- 5'-phosphosulfate, PPi: phosphate group and ATP: Adenosine triphosphate (117). B) Peak heights are proportional to the amount of dNTPs incorporated (Figure B is from one of the pyrosequencing runs in this thesis). The x-axis shows the order of the bases that are incorporated while the y-axis shows the relative light unit (RLU).

4.5.2.1 Verifying PyroMark systems by using PyroMark control oligo

PyroMark control oligo was used to check the quality of the PyroMark Q96 Vacuum Workstation and PyroMark Gold Q96 MD System. PyroMark control oligo is a biotinylated oligonucleotide with the ability of forming an internal steem loop (118). In this way, the oligonucleotide performs self priming (118). Thus, no need of sequencing primer is required before performing sequencing by DNA polymerase (118). The procedure of this PyroMark control kit was performed according to PyroMark® Control Oligo Handbook (118). This procedure is very similar the procedure described below.

4.5.2.2 Pyrosequencing procedure

The pyrosequencing procedure was performed according to the Pyromark® Q96 MD User Manual (119).

Immobilization of PCR product to streptavidin beads

- The reagents in Table 11 were mixed. Depending on the strength of the PCR-product, analyzed by agarose gel electrophoresis, 10 µl, 15 µl, 20 µl or 25 µl of the PCR product was used. The Streptavidin Sepharose High Performance beads 5 ml (hereafter

denoted: streptavidin beads) was carefully shaken from side to side to get a homogeneous solution.

Table 11: Reagents in the mix used for immobilization of biotinylated PCR product to streptavidin beads.

	If 10 µl of the PCR is used and 70 µl of the following mix	**If 15 µl of the PCR is used and 65 µl of the following mix**	**If 20 µl of the PCR is used and 65 µl of the following mix**	**If 25 µl of the PCR is used and 65 µl of the following mix**
Reagent	**µl/well**	**µl/well**	**µl/well**	**µl/well**
Streptavidin beads	1	1	1	1
1x PyroMark binding buffer	40	40	40	40
Milli-Q H_2O	29	24	19	14
Total volume	70	65	55	50

- 50/55/65/70 µl of the mix was added to each well on a ThermoFast 96 well PCR-plate. To avoid precipitation of streptavidin beads, the mix was shaken from side to side before every pipetting.
- 15/10/20/25 µl of the PCR product was added to each well (samples with too low pyrosequencing signals were redone using more PCR-product). Each tumor region and CD24/CD44 sorted cell line sample was run in triplicates.
- The plate was sealed with Abgene® Thermo Scientific seals and thoroughly mixed on a Thermomixer comfort at 13000 rpm for 10 minutes, to make sure that the PCR product binds to the beads.

Clean-up of streptavidin-bound PCR-product, denaturation and annealing of sequencing primer

- 11.6 µl 1x PyroMark Annealing buffer and 0.4 µl primer were mixed per well used in a run. A list of sequencing primers is given in the **Appendix C**.
- 12 µl of the mix was added to each well of a PSQ HS 96 Plate (hereafter denoted: pyrosequencing plate).
- The PyroMark Q96 Vacuum Workstation (hereafter denoted: workstation) was used to clean-up the streptavidin-bound biotin labeled PCR-product. The workstation was prepared by adding 90 ml of Denaturation solution and 110 ml of Milli-Q H_2O, 70% ethanol, PyroMark Wash buffer in trays and placing the pyrosequencing plate in the designed location.
- The vacuum pump and the vacuum on the workstation were switched on.

- PyroMark Vacuum Prep Filter Probe (100) (hereafter denoted: vacuum prep tool) was transferred to the Milli-Q H_2O parking tray and placed there until it was emptied to wash the filter probes. This step was redone once.
- The vacuum prep tool was lowered into the ThermoFast 96 well PCR-plate for 10 seconds or until the wells were emptied.
- Then, the vacuum prep tool was transferred to the 70% ethanol tray for 8 seconds, then to the PyroMark Denaturation Solution (500) for 5 seconds and then to the PyroMark Wash buffer for 10 seconds.
- The vacuum prep tool was held 90º vertical, to empty the hose.
- The vacuum prep tool was held just above the pyrosequencing plate. The vacuum on the workstation was switched off, and the vacuum prep tool was lowered into the pyrosequencing plate and shaken for 30 seconds. The pyrosequencing plate was checked to transfer the streptavidin beads to the wells. They can be seen as white precipitates.
- The pyrosequencing plate was heated at 82ºC for 2 minutes on the heat block QBT2 to ensure that eventual secondary structures of the template are broken.
- The probes on the vacuum prep tool was cleaned by the following steps:
    - ❖ First, the probes were transferred to the Milli-Q H_2O for 10 seconds while shaken.
    - ❖ The vacuum on the workstation was switched on.
    - ❖ The probes were transferred to the Milli-Q H_2O parking tray and placed there until it was emptied to wash the filter probes. This step was redone once.
    - ❖ To empty the hose the vacuum prep tool was held in the air. The vacuum on the workstation and vacuum pump was switched off.
    - ❖ To make sure that the hose was emptied, the vacuum was switched on again.

PyroMark Gold Q96 MD System

- The DNA methylation analysis was set up and run using the PyroMark CpG SW 1.0 software. The instrument parameters corresponding to the dispensing tips where chosen (0004NDT, reagent pressure = 400 mbar, nucleotide pressure = 650 mbar, reagent pulse time = 13.4 milliseconds and nucleotide pulse time = 27.5 milliseconds) and the sequences to be analyzed and the samples name were directed to each well.

- 620 µl Milli-Q H_2O were added to the lyophilized enzyme and substrate mix from the PyroMark Gold Q96 Reagents and shaken until homogenous solution was obtained.
- To prevent clogging of the PyroMark Q96 HS Nucleotide Dispensing Tip's, the dNTPs from the PyroMark Gold Q96 Reagents were centrifuged for three minutes.
- The enzyme and substrate mix were added to the PyroMark Q96 HS Reagent Dispensing Tips, and the dNTPs were added to the PyroMark Q96 HS Nucleotide Dispensing Tips. The amounts of reagents needed for the pyrosequencing run were calculated by the software. For the dNTPs some excess volume were applied.
- Bubbles that can cause clogging of the dispensing tips were removed by careful pipetting.
- The dispensing tips were automatically tested in the PyroMark instrument before starting the run.
- The pyrosequencing plate containing the PCR-product and the sequencing primer was inserted to the PyroMark instrument and the run was started.
- When the run was finished, the dispensing tips were washed with Milli-Q H_2O and set for drying in a box with lint-free towel.

4.5.2.3 Analysis of pyrograms

After the runs were finished, the pyrograms with the sequences were carefully evaluated and analyzed. In the pyrogram each sample is shown with their analyzed sequence and DNA methylation percentage over each CpG. Colors behind the DNA methylation percentage were evaluated. These colors indicate how certain the percentage is; blue indicates high certainty, yellow that the percentage is uncertain and red that the percentage is not trustworthy. Messages about the analyzed sequence were also evaluated. Uncertainty due to low signals is one example of a message that was seen in some occasions. Some of these samples benefitted of being redone with more input volume into the pyrosequencing. Some CpGs were not accepted after the pyrosequencing run due to too low signal regardless of being redone with more input volume into the pyrosequencing. Others were not accepted due to uncertainties in the reference sequences pattern. Uncertain reference peaks were also removed to increase the certainty of the DNA methylation percentages.

4.5.2.4 Processing the pyrosequencing results

Replicates of the same subpopulation in a cell line were available in the present study and each replicate was pyrosequenced in triplicates for the protein-coding genes. Likewise, each tumor region was pyrosequenced in triplicates for the protein-coding genes. For each region within a tumor the mean of methylation levels and standard deviation were calculated for each CpG for the protein-coding genes. The threshold for standard deviation was set to 5, since the detection limit for pyrosequencing is between 5-10%. With a standard deviation above this, the variation between the triplicates was deemed too high to be trustworthy. Likewise, for all replicates of the same subpopulation in a cell line (each replicate has been pyrosequenced in triplicates) the mean of methylation levels and standard deviation were calculated for each CpG within a protein-coding gene. Due to that the mean is based on more values for subpopulations in a cell line than for the regions in a tumor, the threshold for standard deviation was set to 4. With a standard deviation above this, the variation was deemed too high to be trustworthy.

4.5.2.5 Optimizing and methodological observations

Optimized pyrosequencing results were achieved for the protein-coding genes in the tumor regions when increasing the DNA input to the PCR to approximately 15 ng. In almost all cases the standard deviation was lower than 5. An example is shown for gene *IL5RA* in sample 142C, Figure 15. In this example, an input of 7-10 ng DNA into the PCR led to the standard deviation 6.5 and 8.2 for CpG 3 and 4, respectively. When the input of DNA into the PCR was increased to 15-20 ng, the standard deviation was to 1.9 and 1.7 for CpG 3 and CpG 5, respectively. This optimization was taken into consideration when performing pyrosequencing on the tumor regions for the miRNAs which was performed by a colleague and on the cell line samples for the protein-coding genes. Thus, approximately 15 ng DNA was used as input into the PCR for all the genes. The 5% of the total pyrosequencing runs were run in duplicates for the miRNAs in the tumor regions. These duplicates showed low variation in DNA methylation level. The standard deviation of the DNA methylation values of each CD24/CD44 sorted sample/replicate (run in triplicates) was low.

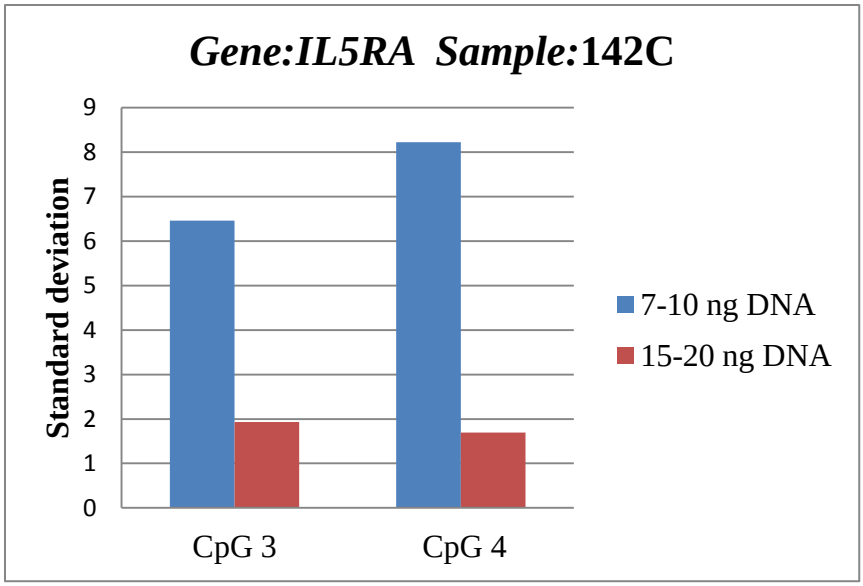


Figure 15: Column plot showing the optimized pyrosequencing result for gene *IL5RA* in sample 142C. The y-axis represents the standard deviation of the triplicates from the pyrosequencing run for CpG 3 and 4 in gene *IL5RA* for sample 142C. Decreasing standard deviation with increasing DNA concentration was observed.

The positive controls were in principle supposed to be 0% and 100% methylated. Unexpectedly, this was not the case. For the 0% and 100% methylated control, relatively low DNA methylation values (from approximately 0% up to 20%) and high DNA methylation values (from approximately 60% up to 100%) were obtained, respectively. Moreover, the methylation level differed from CpG to CpG within a gene.

On the other hand, the PyroMark control oligo samples were supposed to have a DNA methylation value around 50%. All of the samples had a DNA methylation value of exactly 50%. The background signal from the blank dispensations was supposed to be not more than 3%, which was the case.

For the subgrouped cell lines (overview of all samples in **Appendix D**), no PCR product was achieved for the samples 24, 34, 36-38, 49. All of these samples, except 24 and 34 were concentrated by using a SpeedVac®. Consequently, these were excluded in the analysis.

For the tumor regions pyrosequencing results were achieved for more of the selected CpGs within the miRNAs than for the selected CpGs in the protein-coding genes. In addition, for the protein-coding genes pyrosequencing results were achieved for fewer CpGs in the normal samples called RP2, RP10 and RP11 than in the tumor samples.

4.6 Bioinformatics and statistics

4.6.1 Correlation

Correlation analysis was performed on the DNA methylation level and tumor percentage data from **chapter 3.2.2** to estimate for each individual probe/CpG the relationship between DNA methylation level and tumor percentage, which are continuous variables.

Of the originally 200 available samples after performing DNA methylation analysis and estimating tumor percentage, only the samples with tumor percentage between 10% and 90% were included since the accuracy of the ASCAT cannot distinguish between 0% tumor and 100% tumor. The reason for this is that when the tumor percentage is near 100% and the copy number changes are few; there might be 100% normal cells and a copy number of 2 just as well as 100% tumor and a copy number of 2. When taking this into consideration only 113 samples were used for the correlation analysis.

The strength and direction of the relationship between two variables is determined by performing a correlation analysis (120). The correlation coefficient provides the strength and direction between the two variables and the coefficient ranges from -1 to +1, where -1 and+1 is a perfect negative and positive correlation, respectively (120). There are different techniques used for studying correlation, which are both parametric and non-parametric. A standard parametric method is Pearson (120). This method requires normal distributed data (121). A standard non-parametric method is Spearman which do not requires normal distributed data (120;121).

The correlation analysis was performed in the statistical software **R** (http://cran.r-project.org/). Spearman's rank correlation was chosen since the data are not normal distributed because of the many outliers that are presented in these datasets. This analysis is a non-parametric test and replaces the scores with rank values (121). The lowest score gets the lowest rank value and the highest score the highest rank value. The spearman's correlation coefficient or the Spearman's rho uses the rank values to determine the strength and direction of the relationship between two datasets (120).

Overall, p-values<0.05 were considered statistical significant. A p-value for one test, for instance a p-value of 0.05, means that it is a 5% chance of getting a false positive result for that test. This is acceptable for one test. However, if the dataset consists of many tests, the p-value of 0.05 might result in a lot of false positives (type I error). This is what is called the

multiple testing problem. To correct for multiple testing, false discovery rate (FDR) was used calculated by Benjamini-Hochberg method and FDR<0.05 were considered statistical significant. FDR adjusts the p-values by correcting for erroneously accepting false positives by defining the expected false positive rate among the significant results (122).

4.6.2 Ingenuity Pathway analysis

Ingenuity Pathway Analysis (IPA) (Ingenuity® systems, www.ingenuity.com) was used to analyze the list of genes with high correlation identified through correlation analysis. IPA is a software that can be used for understanding the input data better by identifying among other networks, molecular and cellular functions, diseases and disorders and canonical pathways enriched within a list of genes. These results are often given a p-value, FDR value or range (Benjamini-Hockberg method) and a ratio. For functions and pathways the p-value is the likelihood for the input gene set to be associated with a pathway or function due to random chance (123). To calculate the p-value this test takes into consideration: # genes used as input in IPA, # genes from the uploaded gene list in IPA involved in the pathway/function, # genes in the pathways or functions and # genes (123). A FDR value of 0.05 was chosen and is calculated by using the Fisher's Exact Test (123). FDR-range is provided for the functions, due to many sub functions tested for each function. Each of these sub-functions has a FDR value. A pathway, on the other hand is a single entity. Therefore, a single FDR value is provided. The ratio for the canonical pathway is the number of genes in the input gene set that is associated with a specific pathway, divided by the number of genes in the reference set in IPA that is associated with the same pathway (123).

4.6.3 Heatmap

In the present study heatmaps were used to study the differences in DNA methylation between the regions within a tumor for the chosen genes. The heat maps were made in **R** and generated from the mean of methylation levels of all CpGs within a gene cluster for each region. The DNA methylation values were centered on the mean of methylation levels of the regions of each tumor and gene. The heatmaps were scaled so that the difference in colors represents the same amount of change in DNA methylation.

4.6.4 Correlation analysis of the overlapping CpGs from the Infinium® HumanMethylation450 BeadChip array and the pyrosequencing runs

Correlation analysis between DNA methylation level and tumor percentage from the 113 breast tumor samples were assessed like above in **chapter 4.6.1** but only on the CpGs from

the Infinium® HumanMethylation450 BeadChip array that overlapped with the ones used in the pyrosequencing. The overlap included 32 genes. A correlation analysis of these 32 was examined, and a FDR value with a threshold of 0.05 was chosen.

5 Results

The results can be divided into four main parts: 1) Genome-wide correlation between DNA methylation level and tumor percentage, 2) The difference in DNA methylation level between the regions with different amount of tumor cells in each of the 3 tumor samples for a panel of genes, 3) Comparison of the result of the genome-wide correlation analysis with the result of the pyrosequencing of the macrodissected tumors for the overlapping CpGs and 4) The difference in DNA methylation level between subpopulations with different CD24/CD44 expression patterns within a cell line and between subpopulations with the same CD24/CD44 expression patterns across cell lines with different molecular subtype in a panel of breast cancer cell lines for *RASSF1A* and *FOXC1*.

5.1 Genome-wide correlation between DNA methylation level and tumor percentage

The possible influence of heterogeneity on DNA methylation levels was investigated by studying the association between tumor percentage and DNA methylation level genome-wide.

5.1.1 Correlation between DNA methylation and tumor percentage for 113 patients

The results showed that the DNA methylation level of 21569 individual CpGs representing 7723 genes were associated to tumor percentage ($p<0.05$, FDR<0.05), Table 12. When performing the analysis using different FDR cutoffs, 1494 individual CpGs (representing 804 genes) and 47 individual CpGs (representing 25 genes) were identified when using FDR<0.005 and FDR<0.001, respectively. Both positive and negative correlation was revealed within the promoter region as well as within regions outside of the promoter region. The Spearman correlation coefficient (Rho) ranged from -0.52 to 0.60 irrespective of the FDR cutoff.

The CpGs which DNA methylation level was most strongly associated to tumor percentage (FDR<0.001) are shown in Table 13. The CpGs were found on most chromosomes. About half (22/47) of the CpGs were found in regions outside known genes. Both positive (n=39) and negative (n=8) correlations were observed. The association between tumor percentage and DNA methylation is illustrated for nine genes (Figure 16), including the six top genes and three genes known to be important in breast cancer. Positive correlation means that both tumor percentage and DNA methylation level increase or decrease. On the other hand,

negative correlation means that when the tumor percentage increases the DNA methylation level decreases and vice versa.

Table 12: The significant correlation between DNA methylation level and tumor percentage. Number of significant associations for three different FDR cutoffs is shown.

FDR	# significant individual CpGs	# significant probes with negative rho	# significant probes with positive rho	Rho (range)	# significant genes
0.05	21569	8175	13394	(-0.52) – 0.60	7723
0.005	1494	379	1115	(-0.52) – 0.60	804
0.001	47	8	39	(-0.52) – 0.60	25

Table 13: Top CpGs which DNA methylation level was most strongly associated to tumor percentage (FDR<0.001). Probe (one probe represents one CpG), represented gene, gene region, chromosome and genomic position are shown, as well as the Spearman correlation coefficient (Rho) and the associated p-value.

Probe/CpG	Gene	Gene region	Chr*	Genome position	Rho	p-value
cg02770546	*BUD31*	Body	7	99013242	0.6031	1.56E-12
cg06281648	*C16orf57*	Body	16	58039281	0.5389	7.41E-10
cg04339860	*WWP2*	TSS200	16	69796075	0.5313	1.41E-09
cg13487422	*C8orf34*	Body	8	69655786	0.5245	2.47E-09
cg00980625	*GRP*	Body	18	56889548	-0.523	2.83E-09
cg02063165	NA‣	NA‣	1	149294388	0.5224	2.95E-09
cg02483126	NA‣	NA‣	19	22466877	0.5052	1.15E-08
cg06325346	*VN1R2*	TSS1500	19	53761180	0.5038	1.29E-08
cg02603518	NA‣	NA‣	20	61717939	-0.504	1.31E-08
cg02706910	*ZNF254*	5UTR	19	24269975	0.5011	1.58E-08
cg00982799	*PTPRE*	Body	10	129868087	0.4988	1.88E-08
cg00733315	NA‣	NA‣	6	11791912	0.4976	2.05E-08
cg10884614	*EIF5A*	3UTR	17	7215252	0.4972	2.11E-08
cg25420945	NA‣	NA‣	5	95046754	0.4971	2.13E-08
cg14061609	NA‣	NA‣	12	114645716	-0.496	2.29E-08
cg14941172	*ZSCAN5A*	Body	19	56733045	0.495	2.50E-08
cg20434562	NA‣	NA‣	18	76320786	-0.494	2.70E-08
cg06927217	NA‣	NA‣	1	149193722	0.4922	3.07E-08
cg17372841	*BUD31*	Body	7	99013357	0.4912	3.32E-08
cg24762029	*VN1R2*	TSS1500	19	53761256	0.491	3.35E-08
cg14137698	*POMT2*	3UTR	14	77741448	0.4899	3.66E-08
cg00983520	*CHKB-CPT1B, CPT1B*	Body, 5UTR	22	51017067	0.4891	3.86E-08
cg24192505	NA‣	NA‣	6	163776782	0.489	3.90E-08
cg06832449	*PARD3*	Body	10	35069982	0.4887	3.97E-08
cg26682900	*HIP1R*	Body	12	123344689	0.4885	4.05E-08
cg02596994	*HPD*	TSS200	12	122296950	0.4882	4.12E-08
cg19714308	NA‣	NA‣	18	75996301	-0.488	4.29E-08
cg10272977	NA‣	NA‣	6	101674899	0.4864	4.71E-08

Probe/CpG	Gene	Gene region	Chr*	Genome position	Rho	p-value
cg00356694	*SPNS3*	TSS200	17	4337202	0.4863	4.75E-08
cg05528999	*INTS4*	Body	11	77597626	0.4853	5.09E-08
cg17270100	NA‣	NA‣	10	130326232	-0.485	5.29E-08
cg27561567	NA‣	NA‣	18	37380258	0.4847	5.34E-08
cg24251218	NA‣	NA‣	11	10347429	0.4819	6.50E-08
cg23989156	*CRYBA1*	TSS1500	17	27572583	0.4813	6.82E-08
cg13220123	NA‣	NA‣	18	76160277	-0.481	6.96E-08
cg23612095	NA‣	NA‣	2	124445492	0.4808	7.04E-08
cg07577824	SNORD17, *SNX5*	Body	20	17943403	0.48	7.46E-08
cg24279744	*ATP2B2*	Body	3	10408402	0.4796	7.66E-08
cg25605290	NA‣	NA‣	19	12096091	0.4793	7.87E-08
cg05164968	NA‣	NA‣	1	149679932	0.4786	8.24E-08
cg10936879	NA‣	NA‣	10	74397277	0.4786	8.25E-08
cg19483007	*WWTR1*	Body	3	149327651	0.4785	8.29E-08
cg02055607	NA‣	NA‣	5	141152021	0.4783	8.40E-08
cg21055978	*VPS41*	TSS200	7	38948895	0.4782	8.47E-08
cg09221159	NA‣	NA‣	10	90031426	0.4776	8.85E-08
cg27015499	NA‣	NA‣	3	37274798	-0.477	9.16E-08
cg08105529	*HFM1*	Body	1	91852884	0.477	9.20E-08

Chr*: Chromosome
NA‣: Intergenic region

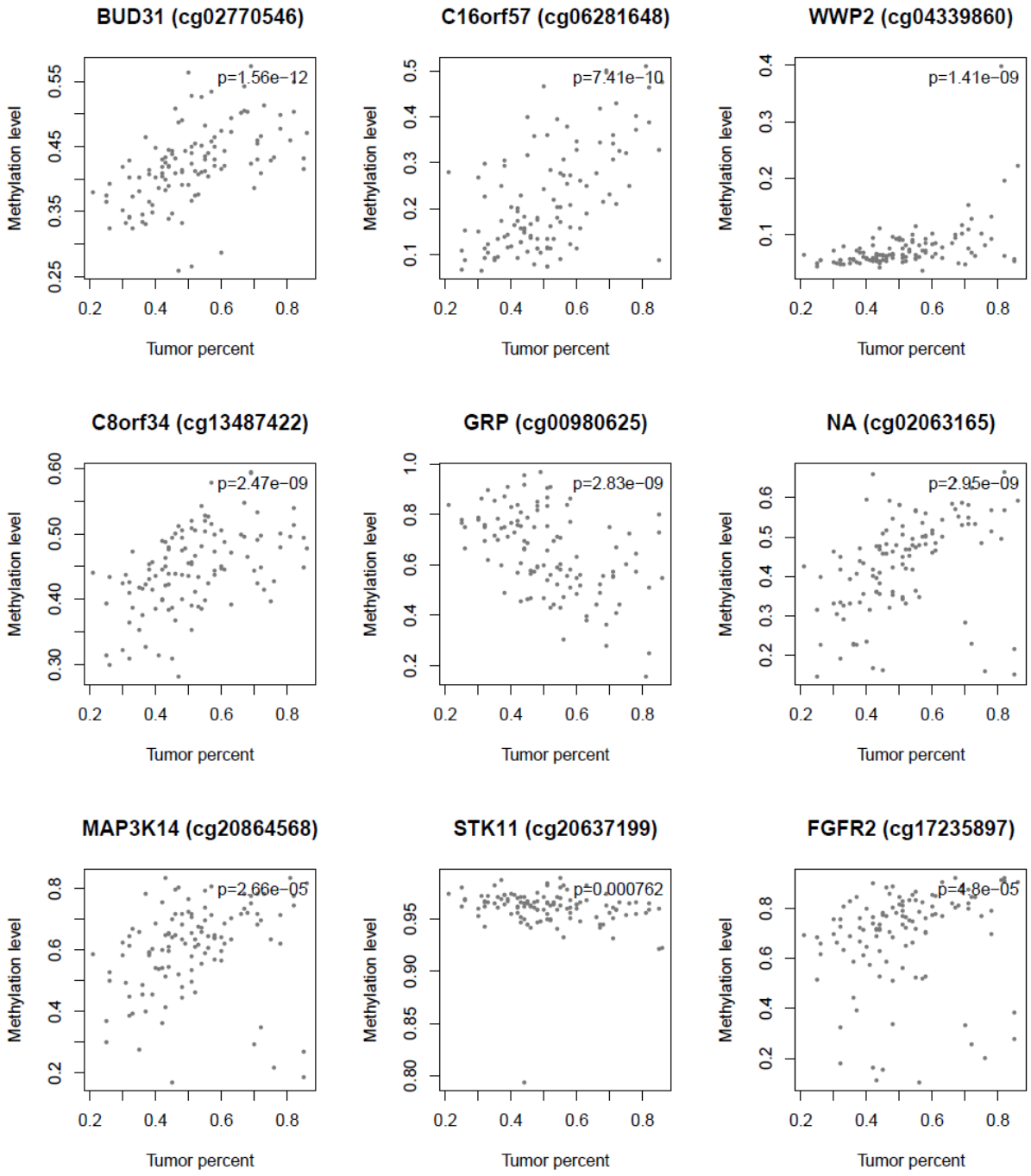


Figure 16: **Plots illustrating the correlation between tumor percentage and DNA methylation level.** The x-axis represents the tumor percent and the y-axis the methylation level. Each grey circle represents a patient. For each plot the current gene or intergenic region (illustrated as NA) as well as the probe in the brackets representing an individual CpG within the gene/intergenic region is shown. The first six plots represent the top six most significantly correlated CpGs while the three others represent significantly correlated CpGs representing genes known as risk genes of breast cancer.

5.1.2 Ingenuity pathways (IPA)

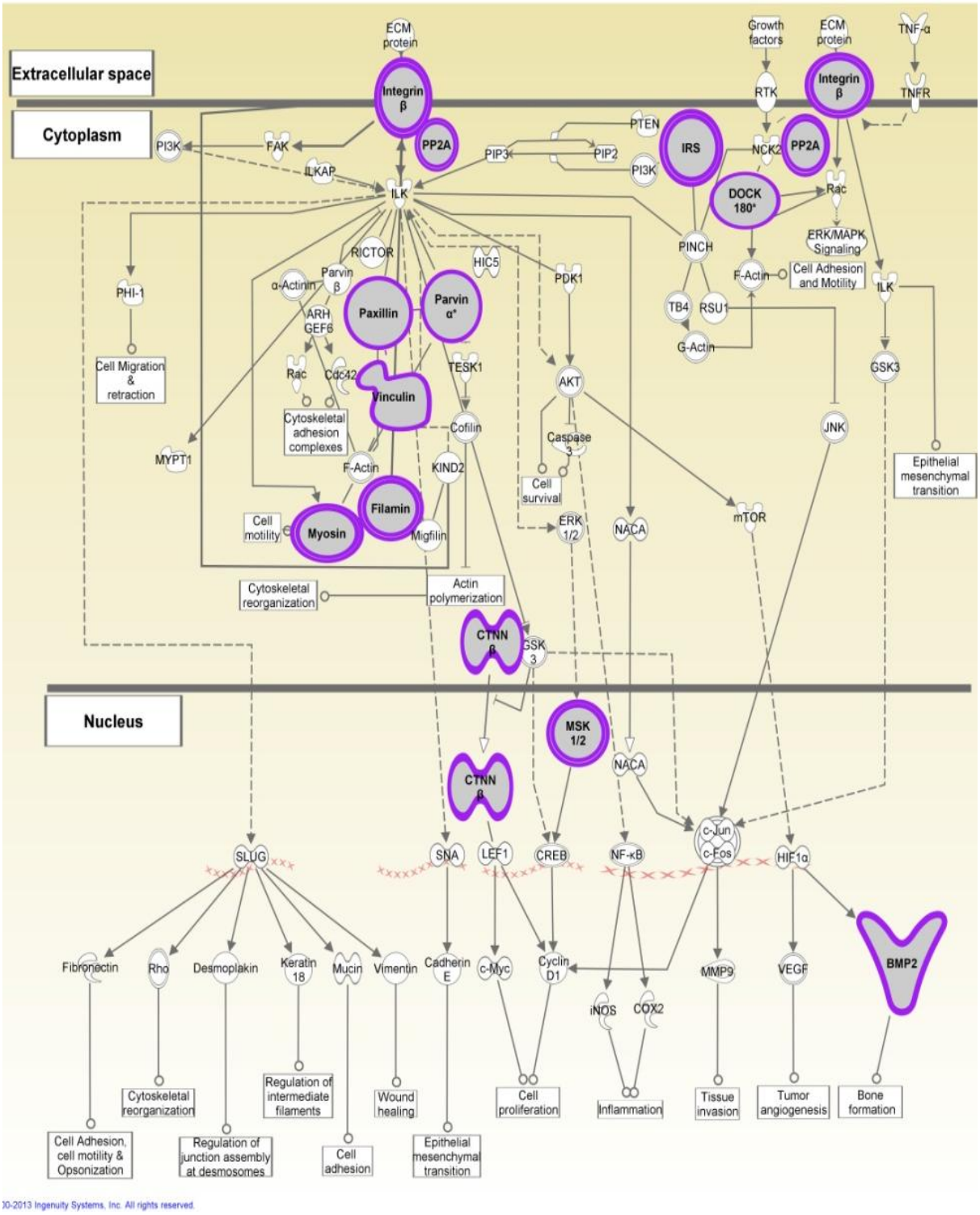
The genes significantly correlated (FDR<0.005, chosen since there is a limit of amount genes that can be used as an input) was uploaded to IPA to identify functions and canonical pathways overrepresented in the gene list (the genes can be seen in **Appendix E)**. Of these 804 genes, 780 were mapped and 24 unmapped. The top 5 molecular and cellular functions identified include cell-cell signaling and interaction, cellular assembly and organization, cell death and survival, cellular movement and cellular function and maintenance (Table 14). For these functions associated molecules from the uploaded gene list ranging from 128-253 molecules were reported. The top and lowest FDR-range for these functions was 3.58E-17-1.78E-01 and 3.64E-04-1.75E-01, respectively. 3 canonical pathways were revealed (FDR<0.05), Table 15. The 3 canonical pathways: ILK-Signaling, Epithelial Adherens Junction Signaling and Tight Junction Signaling have a role in enabling cell-cell signaling or cell-extracellular matrix adhesion. ILK- signaling pathway was identified as the most significant pathway, and genes in the extracellular matrix, the cytoplasm and nucleus were identified, Figure 17.

Table 14: The top 5 molecular and cellular functions enriched in the uploaded gene list. For each molecular and cellular function, FDR- range and the # associated molecules from the uploaded gene list are provided.

Molecular and cellular functions	FDR-range	# molecules
Cell-To-Cell Signaling and Interaction	3.58E-17-1.78E-01	132
Cellular Assembly and Organization	3.58E-17-1.75E-01	149
Cell Death and Survival	4.29E-06-1.78E-01	253
Cellular Movement	9.19E-05-1.78E-01	148
Cellular Function and Maintenance	3.64E-04-1.75E-01	128

Table 15: The 3 significant canonical pathways enriched in the uploaded gene list (FDR<0.05). The p-value, FDR, genes from the uploaded gene list associated with each pathway and the ratio between the genes from the uploaded gene list and the total amount of genes in the pathway are displayed.

Top canonical pathways	p-value	FDR	Ratio	Genes from the uploaded gen list involved in the pathway
ILK-Signaling	8.34E-05	2.47E-02	19/190	*BMP2, CTNNB1, DOCK1, FLNC, IRS1, ITGB2, ITGB4, ITGB5, ITGB7, MYH2, MYH3, MYH8, MYH10, MYH13, PARVA, PPAP2B, PPP2R5C, PXN, RPS6KA4*
Epithelial Adherens Junction Signaling	1.31E-04	2.47E-02	16/146	*BMPR2, CDH2, CTNNB1, FER, FGF1, MYH2, MYH3, MYH8, MYH10, MYH13, PARD3, PVRL1, TGFBR1, TGFBR2, TGFBR3, ZYX*
Tight Junction Signaling	2.27E-04	2.84E-02	16/158	*CLDN4, CLDN15, CTNNB1, INADL, MYH2, MYH3, MYH8, MYH10, MYH13, NGFR, PPP2R5C, PRKCZ, PVRL1, SAFB, TGFBR1, TGFBR2*



)0-2013 Ingenuity Systems, Inc. All rights reserved.

Figure 17: The ILK pathway. The genes in the ILK pathway and where they are located are illustrated. The genes from the uploaded gene list are shown with purple ring around the genes.

5.2 DNA methylation levels of regions within a tumor for a panel of genes

The possible influence of heterogeneity on DNA methylation levels was investigated by assessing DNA methylation level in regions with varying amount of tumor cells from the same tumor (Figures 11-13). DNA methylation level of candidate genes was measured by pyrosequencing for a set of selected CpGs.

5.2.1 The difference in DNA methylation level between the regions within a tumor

The DNA methylation values of the CpGs in the genes for the regions within the tumor samples are shown in **Appendix F**.

The methylation percentage of the CpGs in each gene for each region within the tumors as well for the normal sample (comprises the average of the three normal samples: RP2, RP10 and RP11) was plotted. Overall, similar shape of the curves (representing regions and normal sample) within a gene was observed. An example of this is shown in Figure 18 were the shape of the curves in the tumors for gene *LTC4S* was found to be similar. No clear differences between the two basal-like tumors (137 and 155) and the luminal tumor (142) were revealed.

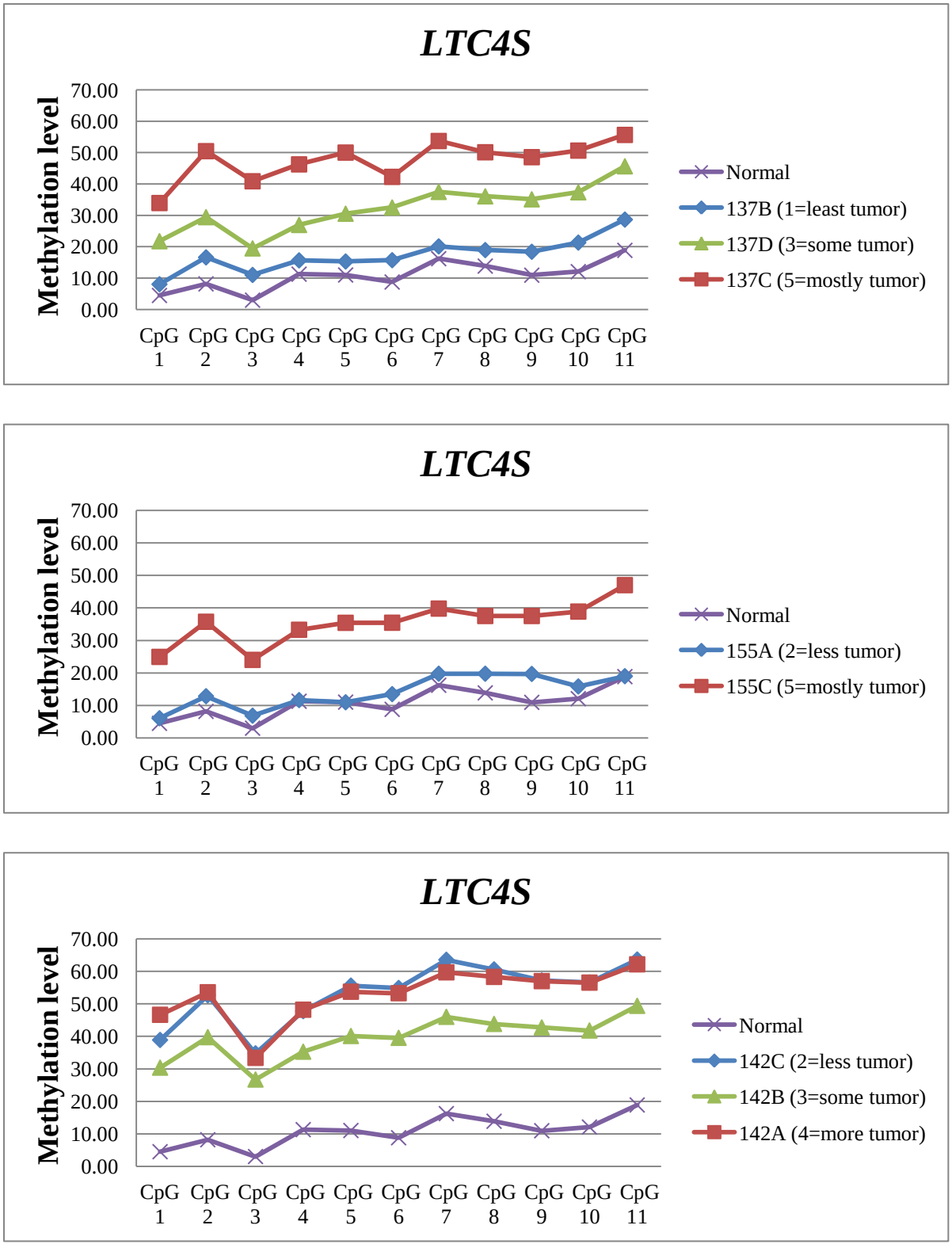


Figure 18: The DNA methylation level of 11 CpGs in the gene *LTC4S* for the different regions in the basal-like tumors (137,155) and the luminal tumor (142) as well as for the normal sample. The x-axis illustrates the CpGs and the y-axis the mean methylation of the CpGs within *LTC4S*. Each data point is an average of the pyrosequencing run in triplicates. The data points for the curve called Normal (marked in purple) are the average of the three normal samples: RP2, RP10 and RP11 run in triplicates.

Since in general similar shape of the curves within a tumor for each gene was detected, the mean methylation of all CpGs in each gene for each region can be used to determine if there are methylation changes between the regions within a tumor. Regions were considered differentially methylated if the mean of methylation levels of all CpGs in a gene differed by more than 10%. The reason for choosing 10% is that the detection limit for pyrosequencing is around 5%-10%. In the cases where the methylation level between the regions were less than 10%, the plots of the methylation level of the CpGs in the genes were further investigated to see if there were any cases where using the mean methylation would disguise differences. This was the case for the genes: *BNIPL* and MIR148A in sample 142 (Figure 19). Differences in methylation level between the regions were observed, but this was not captured using the mean of methylation levels (Table16).

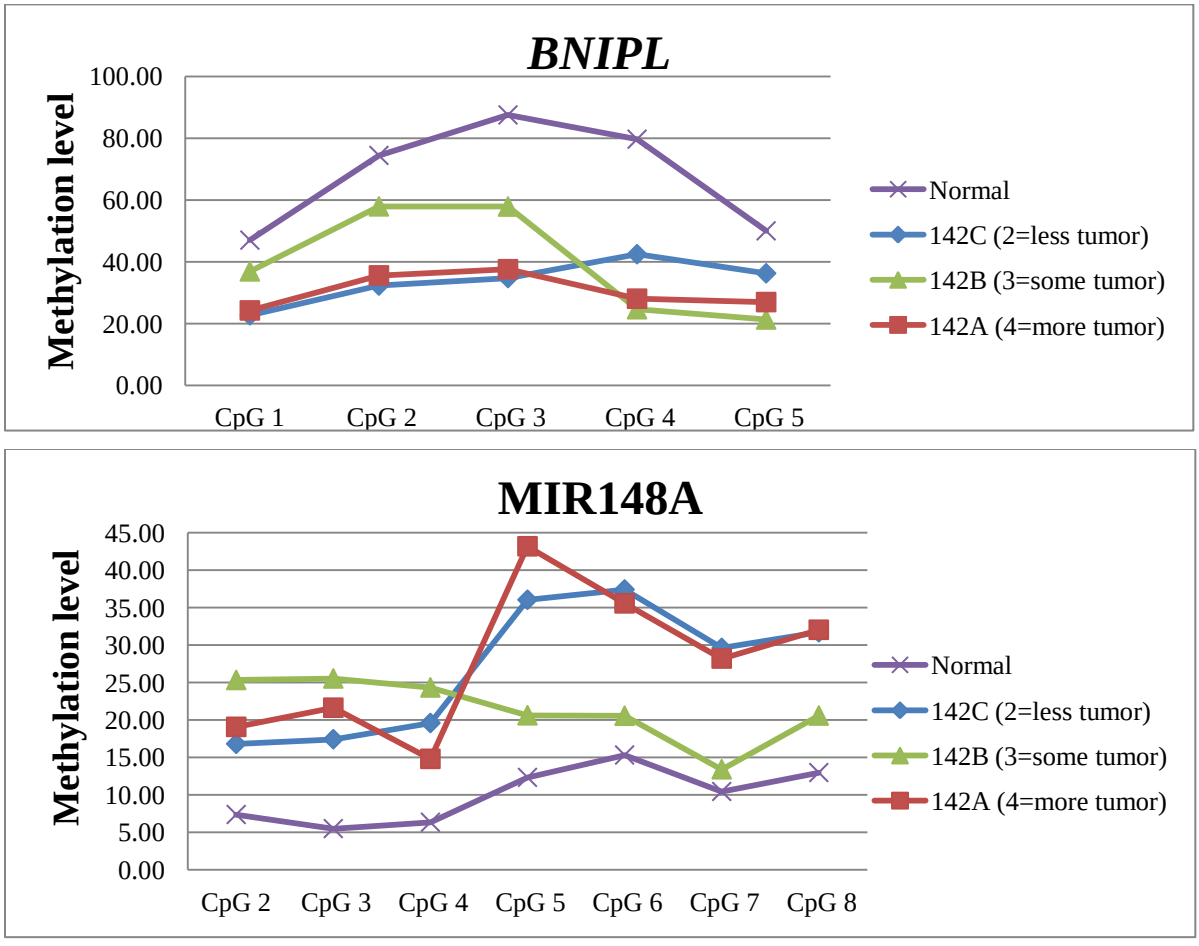


Figure 19: The DNA methylation level of studied CpGs in *BNIPL* and MIR148 for the different regions in tumor 142 (luminal). The x-axis illustrates the CpGs and the y-axis the methylation of the CpGs within *BNIPL* and MIR148A. Each data point is an average of the pyrosequencing run in triplicates for *BNIPL*. The data points for the curve called Normal (marked in purple) are the average of the three normal samples: RP2, RP10 and RP11 run in triplicates. The curves intersect reflecting that comparing mean level of methylation with the amount of tumor would not reflect the true variation. However, by looking at the graphs, differences in methylation level between the regions are clearly observed.

In general, the results showed that for almost all genes a relationship between the DNA methylation level and the amount of tumor cells was detected in one or more tumors, Table 16. For 6 genes the DNA methylation level was related to amount of tumor cells in all tumors. The DNA methylation level of only 3 genes was not related to the amount of tumor cells for any of the 3 tumors.

Table 16: Illustration of the relationship between amount of tumor and DNA methylation level for the genes in the basal-like tumors (155, 137) and the luminal tumor (142). The mean methylation of the CpGs in each gene for all regions in the 3 tumors and for the normal samples is shown. The mean of the 3 normal samples (RP2, RP10 and RP11) is shown. Blue represents genes where differences in methylation level between the regions in all of the tumors were observed, while white represents genes were no relationship between methylation and amount of tumor in some or all tumors were revealed. Pink and yellow represents genes where differences in methylation level between the regions were observed in a basal-like tumor and a luminal tumor, respectively.

Gene	137 (basal-like)			155 (basal-like)		142 (luminal)			Normal
	B (1= least tumor)	D (3= some tumor)	C (5= mostly tumor)	A (2= less tumor)	C (5= mostly tumor)	C (2= less tumor)	B (3= some tumor)	A (4= more tumor)	RP2 RP10 RP11
BCAN	3.82	3.71	1.93	9.08	23.53	42.50	47.44	47.85	5.13
BNIPL	80.70	62.81	54.88	75.48	48.74	33.67	39.72	30.46	69.72
CTSA	7.65	4.68	3.28	6.84	4.31	34.46	24.89	37.31	3.23
FOXC1	3.36	2.81	1.99	3.85	2.54	6.30	8.23	6.17	3.62
FGFBP2	31.79	24.56	17.24	36.20	19.49	20.48	17.59	15.47	NA*
GYPE	70.06	71.91	80.28	56.43	51.02	57.95	59.62	55.81	60.83
IL1A	32.39	28.39	15.02	30.55	17.71	40.32	51.15	45.71	26.3
IL1R2	33.60	25.23	18.88	39.15	49.25	25.09	26.34	28.55	55.53
IL5RA	55.05	69.19	78.18	33.95	17.45	18.75	15.20	12.46	31.12
LTC4S	17.27	32.05	47.53	14.17	35.41	53.27	39.58	52.94	9.71
PCK1	41.96	36.70	37.19	39.18	26.99	15.16	10.53	10.88	25.83
RASSF1A	1.13	0.91	1.16	1.35	1.31	23.80	32.94	35.54	2.26
CDKN2A	1.88	2.08	1.47	2.26	1.86	1.77	2.12	2.56	2.53
MIR199A	64.07	75.42	77.49	66.57	66.20	69.86	67.02	70.63	29.24
MIR135B	56.84	28.66	17.25	89.16	88.10	82.08	81.90	80.70	84.90
MIR16-2	52.95	49.04	39.17	52.83	50.21	39.31	36.81	24.74	60.80
MIR887	8.95	7.11	3.23	10.94	5.12	2.69	1.30	2.55	2.85
MIR148A	14.49	14.54	7.76	21.51	9.34	26.94	21.48	27.78	10.02
MIR200C/141	39.46	27.92	16.06	56.94	38.13	23.05	19.05	16.81	51.01
MIR17-92	39.76	36.71	26.27	48.72	35.26	41.25	43.26	44.32	60.31
MIR142(1)•	24.91	32.74	51.15	19.49	33.02	67.73	50.48	64.78	62.30
MIR142(2)•	42.49	46.48	62.49	29.73	50.66	64.11	81.68	75.58	74.39
MIR150	13.90	26.16	35.60	8.98	28.49	40.21	47.33	50.37	21.60

•MIR 142(1) and MIR142(2): two different amplified regions of the same miRNA.
NA*: not available due too to low signal in the pyrosequencing

5.2.2 Pattern of DNA methylation for the 3 tumors

The difference in mean of DNA methylation level of all regions in a tumor for each gene between the 3 tumors was investigated. The basal-like tumors showed different DNA methylation pattern (10% or more) from the luminal tumor in around half of the genes.

Differences between the 3 tumors for the genes which an association between the tumor amount and DNA methylation level was revealed were studied. The DNA methylation of more genes was related to the amount of non-neoplastic cells in the basal-like tumors (137 and 155) than in the luminal tumor (142), as shown in Table 16 and Figure 20. In addition, greater differences in DNA methylation level between the regions were observed in the basal-like tumors than in the luminal tumor. This is illustrated well in the heat map, Figure 20, where the variation between the colors for sample 142 were less than for 137 and 155, indicating smaller differences in DNA methylation level.

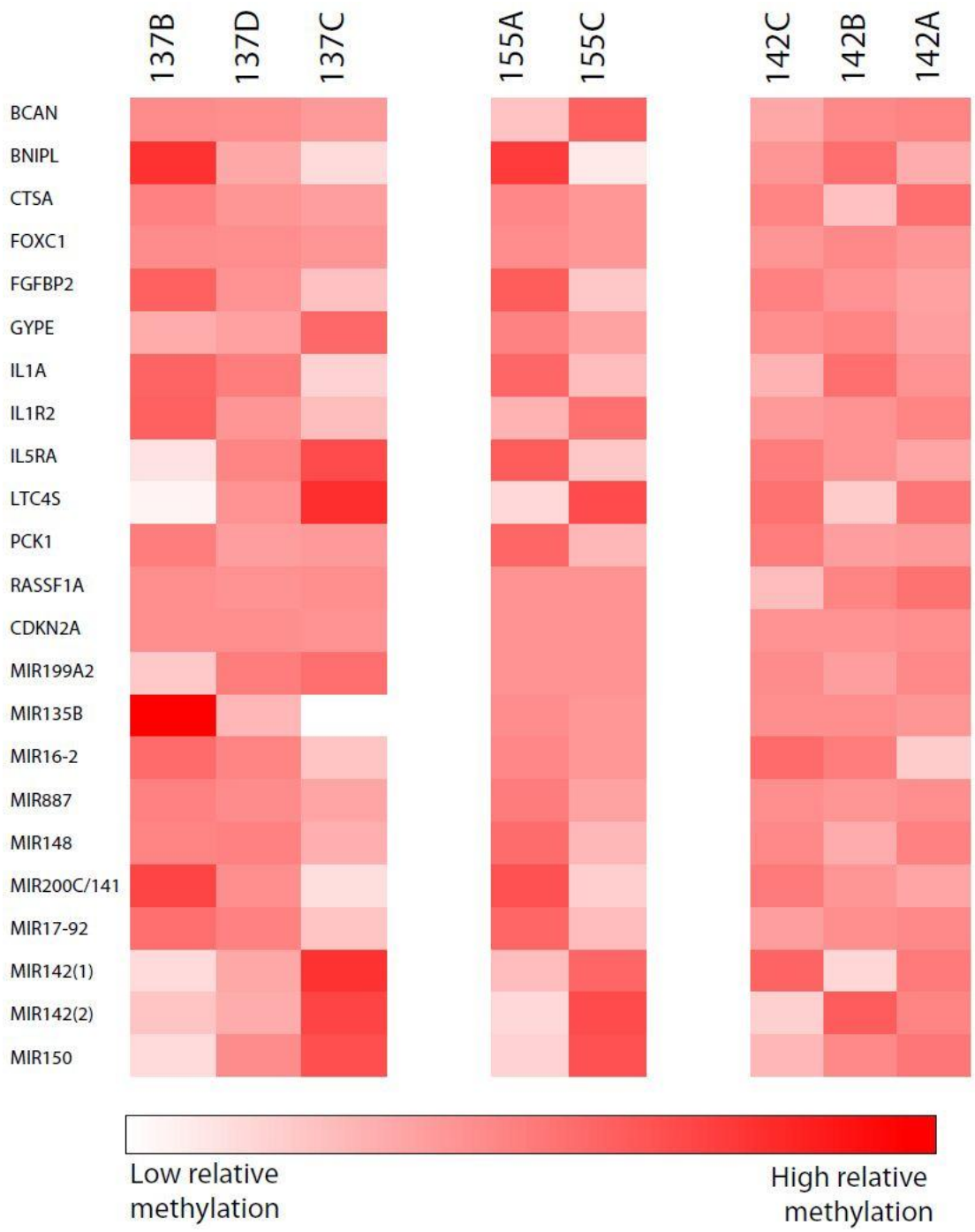


Figure 20: Heatmaps reflecting the differences in DNA methylation level between the regions in each of the basal-like tumors (137 and 155) and the luminal tumor (142) for each gene. The heatmaps are graphical illustrations generated from the mean methylation of the CpGs within a gene for each region within the tumors. For each gene, the methylation levels in the regions (put in order with the region with the least tumor cells to the region with the most tumor cells from left to right) were represented by a color scale from white to red. The more variation in color between the regions, the higher is the difference in methylation level. The heatmaps were generated by centering the DNA methylation values on the mean of methylation levels of the regions of each tumor and gene. The heatmaps were scaled so that the difference in colors represents the same amount of change in DNA methylation. MIR142(1) and MIR142(2): the number in brackets is added to distinguish between different amplified region of the same miRNA.

If a monotone trend was present or not was examined. A monotone trend is present if a function of two variables (in this study DNA methylation level and tumor percentage) constantly increases or decreases. A criterion for this examination was that three or more regions within a tumor were available. Therefore, tumor 155 was not included in this regard. However, if the direction of the association between the tumor- and DNA methylation level was positive or negative was examined for this tumor.

For sample 137 a monotone trend was observed for all genes, Table 17. For 9 of 11 genes, the direction of the monotone trend/association between DNA methylation level and tumor percentage was the same for both basal-like tumors meaning: when the tumor % increased the methylation level in both basal-like tumors either decreased or increased. However, for *IL1R2* and *IL5RA* opposite direction was observed between 137 and 155.

In contrast, for the luminal tumor, a monotone trend was rarely discovered, Table 17. For the luminal tumor the DNA methylation level of 10 genes was shown to be different between the regions. For two of these, MIR148A and *BNIPL*, the pattern of how the amount of tumor was associated with the methylation level was difficult to investigate as the curves of the regions representing the methylation level of the CpGs intersect. Of the other 8, a monotone trend was discovered for only 3 of the genes. For the other genes region B which consisted of some tumor had either the highest or lowest mean of methylation level. For three of these genes: *CTSA*, *LTC4S* and MIR142(1) the region with more tumor and the less tumor were most alike with regard to methylation level. The mean methylation of 6 genes (MIR148A and *BNIPL* not considered) was associated with the amount of tumor in the luminal tumor and one or both of the basal-like tumors. Of these, a monotone trend between the methylation level and the amount of tumor was observed for only 2 genes, and the direction of the monotone trend was equal the direction of the monotone trend/association between tumor percentage and methylation level of the basal-like tumors.

Which region the normal sample was closest to in DNA methylation level was also examined and differed among the genes in each tumor, Table 17. However, for sample 137 the normal sample was closest to the region with least tumor for the majority of the genes.

Table 17: Trends observed for the basal-like tumors (137 and 155) and the luminal tumor (142) for the genes which a relationship between the amount of tumor and the DNA methylation profile was detected. Blue represents genes where the methylation level of the CpGs between the regions intersects. Green represents genes were a monotone trend was present or the direction of the association between methylation level and tumor percentage (regards tumor 155) was stated. Purple represents genes were a monotone trend was not observed. D: if difference in methylation level between the regions was observed M: if a monotone trend was observed, Dir: the direction of the monotone trend/association; negative=neg, positive=pos, mostly tumor=MT, more tumor=MOT, some tumor=ST, less tumor=LET, least tumor=LT. M 155=NA: a monotone trend was not considered for sample 155 since it only consisted of 2 regions.

	137 (basal-like)			155 (basal-like)			142 (luminal)			137	155	142
Gene	**D**	**M**	**Dir**	**D**	**M**	**Dir**	**D**	**M**	**Dir**	**The region in the tumor that had DNA methylation level closets to the normal sample**		
BCAN	no			yes	NA*	pos	no				LT	
BNIPL	yes	yes	neg	yes	NA*	neg	yes			ST	LT	
CTSA	no			no			yes	no				ST
FGFBP2	yes	yes	neg	yes	NA*	neg	no			ND*	ND*	ND*
GYPE	yes	yes	pos	no			no			LT		
IL1A	yes	yes	neg	yes	NA*	neg	yes	no		ST	MT	LET
IL1R2	yes	yes	neg	yes	NA*	pos	no			LT	MT	
IL5RA	yes	yes	pos	yes	NA*	neg	no			LT	LT	
LTC4S	yes	yes	pos	yes	NA*	pos	yes	no		LT	LT	ST
PCK1	no			yes	NA*	neg	no				MT	
RASSF1A	no			no			yes	yes	pos			LET
MIR199A	yes	yes	pos	no			no			LT		
MIR135B	yes	yes	neg	no			no			LT		
MIR16-2	yes	yes	neg	no			yes	yes	neg	LT		LET
MIR148A	no			yes	NA*	neg	yes				MT	
MIR200C/141	yes	yes	neg	yes	NA*	neg	no			LT	LT	
MIR17-92	yes	yes	neg	yes	NA*	neg	no			LT	LT	
MIR142 (1)•	yes	yes	pos	yes	NA*	pos	yes	no		MT	MT	MOT
MIR142 (2)•	yes	yes	pos	yes	NA*	pos	yes	no		MT	MT	MOT
MIR150	yes	yes	pos	yes	NA*	pos	yes	yes	pos	ST	MT	LET

•MIR 142 (1) and MIR142 (2): the numbers in the brackets are added to distinguish between different amplified region of the same miRNA.
*NA: not decided since methylation level of the gene was not available of the normal sample due to too low signal in the pyrosequencing.

5.3 Comparison of the result of the genome-wide correlation analysis with the result of the pyrosequencing of the macrodissected tumors for the overlapping CpGs

Of the 205 CpGs used in the pyrosequencing, 32 were found on the Infinium® HumanMethylation450 BeadChip array. Only a few of these were mapped to protein-coding genes. Of these 32 CpGs, the DNA methylation level of 15 probes/CpGs was significantly correlated to tumor percentage. These 15 CpGs were mapped to 6 of 22 genes used for the pyrosequencing, seen in Table 18. Only one of these represents a protein-coding gene.

For the same 15 CpGs that were pyrosequenced, the amount of tumor cells in a tumor was determined to be associated with the DNA methylation level if the difference in methylation level between the regions was 10% or more different. Overall, the pyrosequencing result and the result from the genome-wide correlation corresponded well, Table 18. For tumor 137, the methylation level of all except 1 CpG was associated with the amount of tumor. For these CpGs a monotone trend was observed and the direction of the monotone trend was equal the direction of Spearman's rank correlation (shown by rho being positive or negative). For sample 155, the methylation level of the majority of the CpGs was associated with the amount of tumor. The direction of the association between methylation level and tumor percentage was the same as the direction of the Spearman's rank correlation. For sample 142 the methylation of the majority of the CpGs was associated with amount normal infiltration. However, only 1 of these showed a monotone trend and the direction of this trend was equal the direction of Spearman's rank correlation. Consequently, the pyrosequencing result of the basal-like tumors corresponded better with the genome-wide correlation analysis than the luminal tumor did.

Table 18: Comparison of the result from the genome-wide correlation analysis with the pyrosequencing result of the macrodissected tumors for the overlapping CpGs. This table illustrates the 15 CpGs where the DNA methylation level was found significantly correlated to tumor percentage of the 32 CpGs used in the pyrosequencing found on the Infinium® HumanMethylation450. These 15 CpGs were compared to the outcome of the same CpGs pyrosequenced. D: if differences in the methylation level between the regions were observed or not for a CpG, M: if a monotone trend was observed and Di: direction of monotone trend/association between methylation level and tumor percentage. M 155=NA: a monotone trend was not considered for sample 155 since it only consisted of 2 regions. pos: positive and neg: negative

Genome-wide correlation		Pyrosequencing										
Probe	Rho	CpG	Gene	Tumor								
				137 (basal-like)			155 (basal-like)			142 (luminal)		
				D	M	Di	D	M	Di	D	M	Di
cg19759135	0.33	8	*LTC4S*	yes	yes	pos	yes	NA•	pos	yes	no	
cg11394785	0.33	11	*LTC4S*	yes	yes	pos	yes	NA•	pos	yes	no	
cg23665802	-0.27	2	MIR17-92	yes	yes	neg	yes	NA•	neg	no		
cg00366413	-0.31	12	MIR200C/141	yes	yes	neg	yes	NA•	neg	yes	no	
cg27534624	-0.24	14	MIR200C/141	yes	yes	neg	yes	NA•	neg	yes	no	
cg15035143	0.34	11	MIR142(1)*	yes	yes	pos	yes	NA•	pos	yes	no	
cg21232937	0.36	12	MIR142(1)*	yes	yes	pos	yes	NA•	pos	yes	no	
cg00057966	0.34	13	MIR142(1)*	yes	yes	pos	no			yes	no	
cg26112797	0.24	2	MIR142(2)*	yes	yes	pos	yes	NA•	pos	yes	no	
cg10530767	024	3	MIR142(2)*	yes	yes	pos	yes	NA•	pos	yes	no	
cg27388703	0.26	9	MIR150	yes	yes	pos	yes	NA•	pos	yes	yes	pos
cg06105296	0.27	10	MIR150	yes	yes	pos	yes	NA•	pos	no		
cg24002149	0.28	1	MIR199A	no			no			no		
cg21450888	0.29	3	MIR199A	yes	yes	pos	no			no		
cg17178220	0.29	5	MIR199A	yes	yes	pos	no			no		

*MIR 142(1) and MIR142(2): the numbers in the brackets are added to distinguish between different amplified region in the pyrosequencing of the same miRNA.
•NA: not available

5.4 The DNA methylation levels of subpopulations in breast cancer cell lines

The DNA methylation levels, obtained by pyrosequencing, of *RASSF1A* and *FOXC1* in CD24/CD44 sorted breast cancer cell lines were investigated to assess the difference in DNA methylation level between subpopulations with different CD24/44 expression patterns within a cell line and between subpopulations with the same CD24/CD44 expression patterns across cell lines with different molecular subtype.

5.4.1 Different DNA methylation levels between samples of the same subpopulation in a cell line (replicates)

Different DNA methylation levels between the samples of the same subpopulation in a cell line (replicates) were revealed for *RASSF1A* and *FOXC1*. Since the variation in DNA methylation level was high between the replicates the results weren't trustworthy, indicating by a standard deviation above 4. However, the variation in DNA methylation level of each

CD24/CD44 sorted cell line/replicate pyrosequenced in triplicates was low, indicating by a low standard deviation.

An example is shown for *FOXC1* in the samples 10-12, replicates of the CD44-/CD24+ subpopulation in the ZR-75-1 cell line. Great difference in mean methylation was observed between the replicates. For sample 10 considerately lower mean methylation level was observed than for sample 11 and 12 (Figure 21). This was further reflected by the high standard deviation of the methylation levels of all the samples 10-12 (run in triplicates), as seen in Figure 22. Despite this observation, the standard deviation of the methylation levels of each replicate pyrosequenced in triplicates was low. The same pattern was observed for *RASSF1A*. An example is shown for the sample 42, 44 and 45, replicates of the CD44-/CD24+ subpopulation in the T47D cell line, where great differences in mean methylation between the replicates were discovered. The mean methylation for sample 44 differed greatly from 42 and 45 (Figure 23). This was further reflected by the high standard deviation of the methylation levels taking of all the samples 42, 44 and 45 (run in triplicates), as seen in Figure 24**.** Despite this observation, the standard deviation of the methylation levels of each replicate pyrosequenced in triplicates was low.

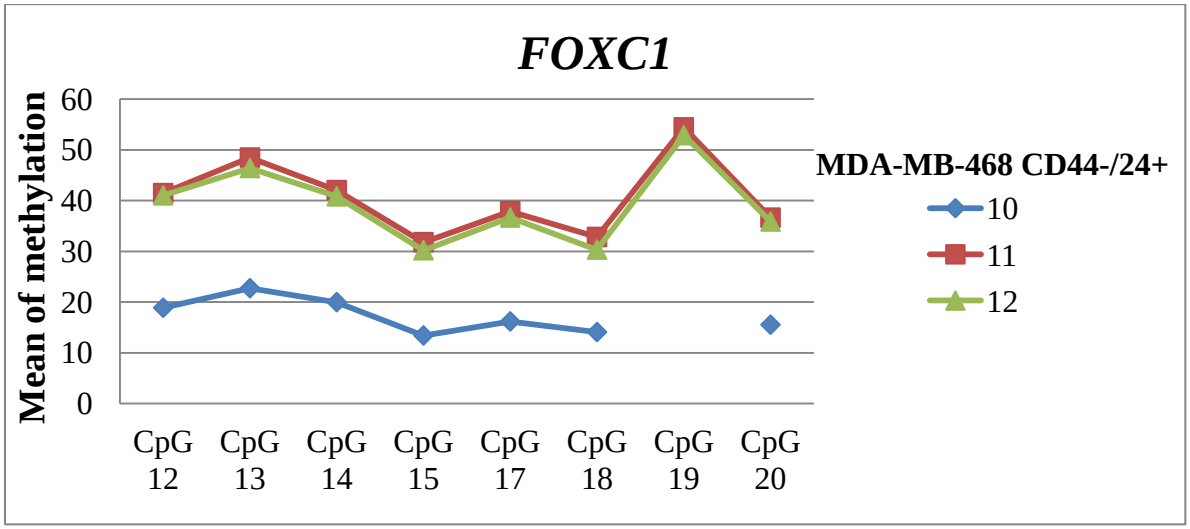


Figure 21: The mean methylation of CpG 12-15 and 17-20 in *FOXC1* for sample 10, 11 and 12 with DNA from CD44-/CD24+ sorted ZR-75-1 cell line. The x-axis represents the CpGs and the y-axis the mean methylation of the CpGs in *FOXC1*. The mean methylation of each CpG is generated from triplicates of each sample. The mean methylation of CpG 19 for sample 10 was not calculated due to none acceptable results from the pyrosequencing.

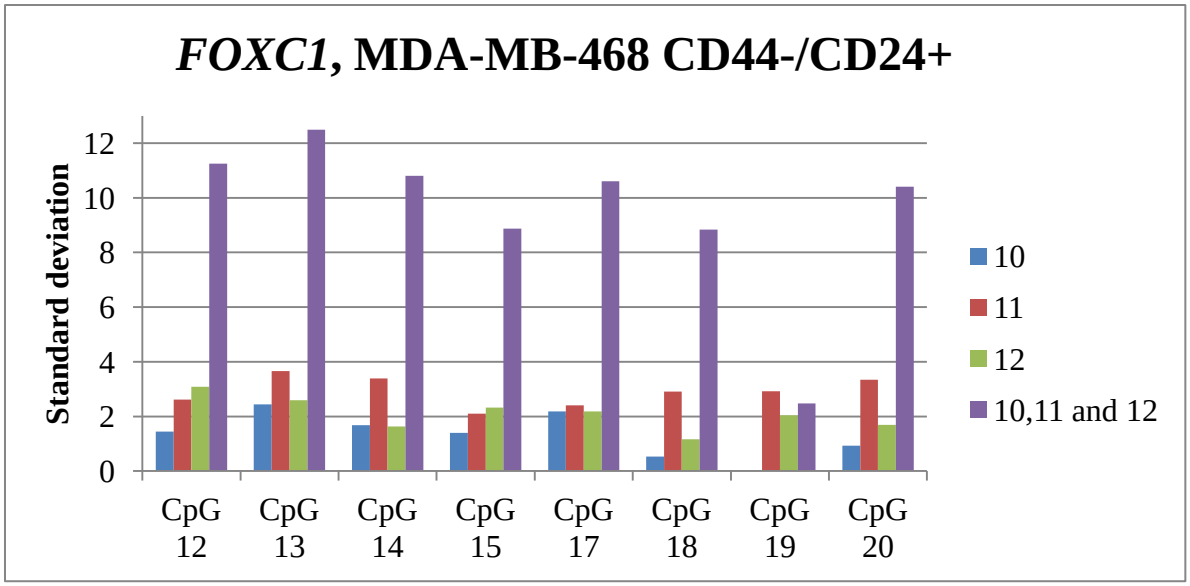


Figure 22: The standard deviation of the DNA methylation values of CpG 12-15 and 17-20 in *FOXC1* for each of the samples and all samples with DNA from the CD44-/CD24+ sorted ZR-75-1 cell line. The x-axis represents the CpGs. The y-axis represents the standard deviation of the methylation levels of each sample run in triplicates and all samples 10-12 with DNA from the same cell line with the same CD24/44 pattern. The standard deviation of CpG 19 for sample 10 was not be calculated due to none acceptable results from the pyrosequencing.

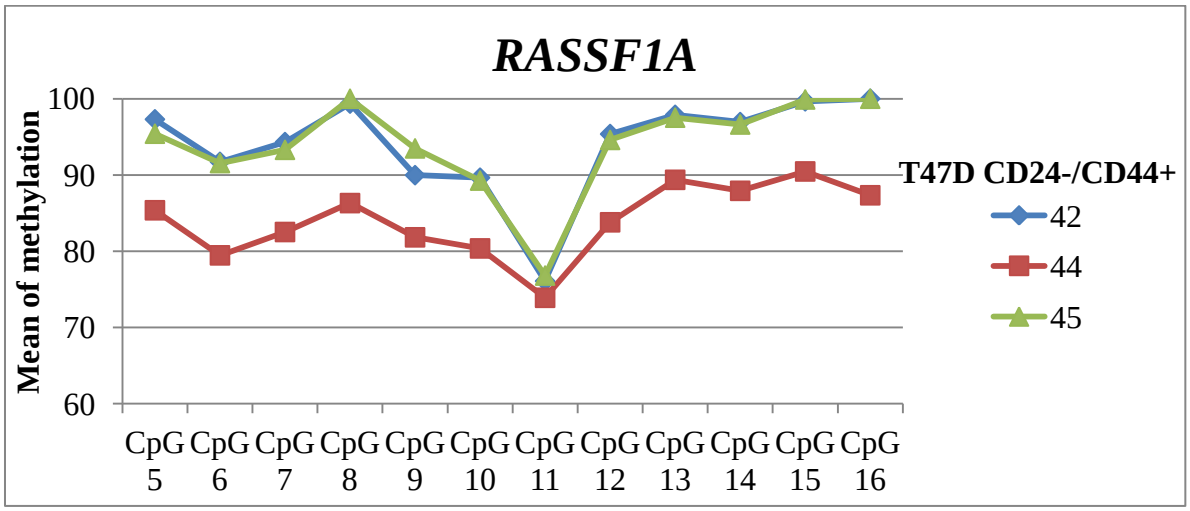


Figure 23: The mean methylation of CpG 5-16 in *RASSF1A* for sample 42, 44 and 45 with DNA from the CD24+/CD44- sorted T47D cell line. The x-axis represents the CpGs and the y-axis the mean methylation of the CpGs in *RASSF1A*. The mean methylation of each CpG is generated from triplicates of each sample.

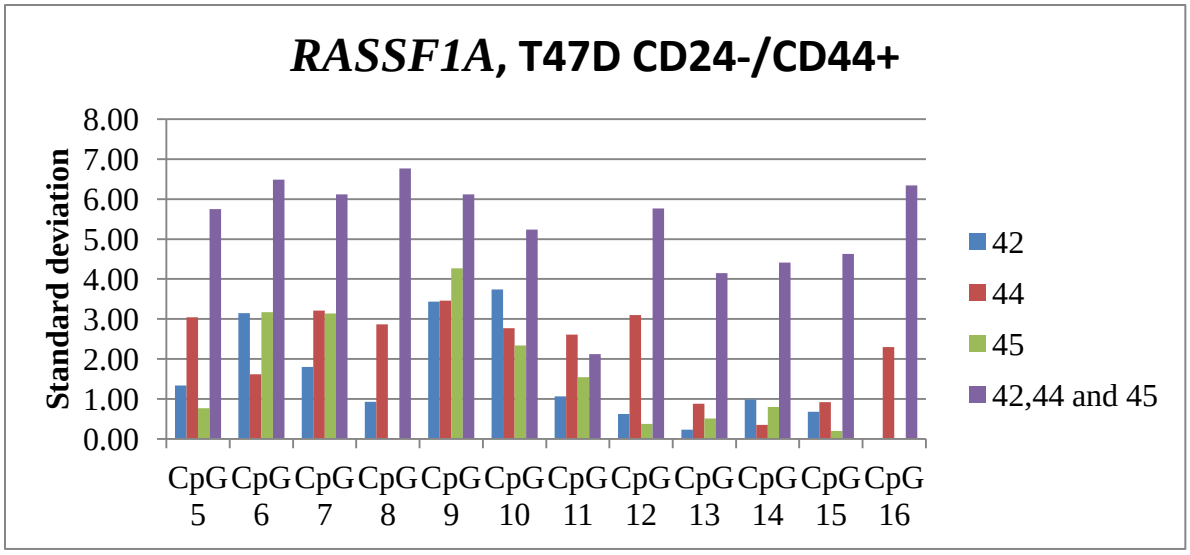


Figure 24: The standard deviation of the DNA methylation values of CpG 12-15 and 17-20 in *FOXC1* for each of the samples and all samples with DNA from the CD24+/CD44- sorted T47D cell line. The x-axis represents the CpGs. The y-axis represents the standard deviation of the methylation levels of each sample run in triplicates and all samples 42, 44 and 45 with DNA from the same cell line with the same CD24/44 pattern.

6 Discussion

DNA methylation studies have look at among other DNA methylation in regard to development, treatment and survival of breast cancer. Few studies have investigated the impact of tumor heterogeneity on DNA methylation levels in breast cancer. This was investigated in this study. This study was designed as a pilot study to explore if this topic is worth studying further. More specifically the pilot study was initiated to assess if infiltrating non-neoplastic cells intermingled with the tumor cells might modify the DNA methylation level. This is of importance as many studies use bulk tumors (containing tumor cells and surrounding tissue). The finding is important for how DNA methylation studies will be interpreted and performed in the future. Additionally, this pilot study was initiated to investigate the difference in DNA methylation levels between subpopulations with different CD24/CD44 expression patterns within a cell line and between subpopulations with the same CD24/CD44 expression patterns across cell lines with different molecular subtype. This is of importance for investigating the robustness of these markers beyond cell lines with different molecular subtype. Different DNA methylation levels between luminal and basal-like tumors have been shown (42-44). It has also been suggested that the different cell types (CD24/CD44) are associated with different feature and shown to have different DNA methylation profile (87-89). However, few studies have looked at DNA methylation of subpopulations with different CD24/44 expression patterns in cell lines with different molecular subtype. Especially, this study is relevant since tumors associated with CD44 have shown worse outcome (87).

6.1 Biological considerations

6.1.1 The relationship between methylation level and tumor percentage/amount of tumor

The results from the genome-wide correlation analysis showed that the DNA methylation level of individual CpGs in around 1/3 of the genes in the human genome was associated to tumor percentage. Several of the CpGs which DNA methylation was most strongly associated to tumor percentage are related to genes associated with cancer. Three of these genes are WW domain containing E3 ubiquitin protein ligase (*WWP2*), gastrin-releasing peptide (*GRP*) and eukaryotic translation initiation factor 5A (*EIF5A*). *WWP2* has been shown to deregulate the activity of *PTEN*, an important tumor suppressor playing a role in cell growth and apoptosis (124;125). Overexpression of *GRP* has been shown in malignant breast tumors, but not in

begins lesions, and the gene has been suggested to play a role in resistance to apoptosis (126). *EIF5A* has been shown to regulate the activity of TP53, a tumor suppressor known to be important for cell growth and dividing, DNA repair and genome stability (127;128). As previously described, variants in this gene is associated with high risk of breast cancer (19). Further, the genome-wide correlation revealed both positive and negative correlation within and outside the promoter region. This illustrates that with increased tumor amount, both hypermethylation and hypomethylation were revealed within and outside the promoter region. This reflects that the DNA methylation pattern in cancer is more complex than the common view which is that cancer is hypomethylated globally and hypermethylated in CpG islands (often found in the promoter region) (6).

For the three macrodissected tumors studied by pyrosequencing it was also shown that the DNA methylation level of most genes was associated with the amount of non-neoplastic cells in at least one of the tumors. For most genes, it varied from tumor to tumor if an association between DNA methylation and amount of non-neoplastic cells was present. A possibility for why the relationship is different might be due to inter-tumor heterogeneity. The association was observed for a few of the same genes in all tumors. Furthermore, the association was detected for some of the same genes in the tumors with the same subtype. For other genes, the association was observed for only one of the tumors.

A relationship between DNA methylation level and tumor percentage/amount of tumor was observed both genome-wide for 113 bulk tumors and for selected CpGs in a panel of genes in 3 macrodissected tumors. From previous studies, it has been established different DNA methylation levels between healthy women and women with breast cancers. For instance Suijkerbuijk *et al*. (129) used quantitative multiplex methylation-specific PCR (QM-MSP) on breast cancer samples and normal breast samples and revealed different DNA methylation levels between the two groups. Other studies like Ronneberg *et al*. (43) and Kamalakran *et al*. (44) have shown the same trend. In these studies unsupervised hierarchical clustering was performed based on the DNA methylation level of cancer related genes in breast tissue samples as well as normal samples. The normal samples were clustered together in their own branch. In addition, the hierarchical clusters illustrated that the variation of the DNA methylation values were less between the normal samples than the breast cancer samples.

These studies above have studied normal breast tissue and breast cancers from different patients. Unlike these studies, other studies have looked at the DNA methylation level of

normal sample and breast tissue in the same patient. For instance, a recent study by Hansen *et al*. (130) looked closer at the DNA methylation level of tumor tissue and matched normal samples from breast cancer as well as from colon, lung, thyroid and kidney cancer. The results showed that the DNA methylation level of a normal sample was closer to other normal samples than their own matched tumor tissue. Likewise, another study showed the same pattern in breast cancer by Bediaga *et al*. (131). Ergo, the DNA methylation level of normal tissue and tumor tissue from the same patient has been shown to be different. In the present study, different DNA methylation levels between the tumor cells and infiltrating non-neoplastic cells intermingled with the tumor cells were observed. These infiltrating non-neoplastic cells are neither tumor cells nor "complete normal cells" as signals between the tumor and the surrounding non-neoplastic cells are thought to occur (9) like in Cheng *et al*. (132). This result is not unexpected when knowing that the DNA methylation level of normal tissue and tumor tissue from the same patient has been shown to be different.

This study is also in line with another study. Moelans *et al.* (133) used microdissection to obtain 2 to 5 sub regions of 21 breast cancer patients. Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) was used on these sub regions for promoter regions of 24 tumor suppressor genes. They revealed different DNA methylation levels between the sub regions for 95% of the tumors for at least one locus and the most heterogeneous loci were found in the represented genes: *RASSF1A*, cyclin-dependent kinase inhibitor 2B (*CDKN2B*), death-associated protein kinase 1 (*DAPK1*), cyclin-dependent kinase inhibitor 1B (*CDKN1B*), adenomatous polyposis coli (*APC*), *GSTP1* and *CDH13*. Of these genes, CpGs in the promoter region of *RASSF1A*, *APC* and *GSTP1* were covered by the Infinium® HumanMethylation450 BeadChip in this study. The DNA methylation level of one or more CpGs in these genes was shown to be associated with the tumor percentage. Also, for the 3 tumors assessed in this study, differences in DNA methylation level between the regions were revealed for the luminal tumor. In addition, the article showed that similar DNA methylation level was found between the regions in the promoter region of *CDKN2A* which is in agreement with the 3 tumor macrodissected of this study.

As a relationship between DNA methylation level and tumor amount/percentage has been revealed in this study, it has been shown that infiltrating non-neoplastic cells influence the measurements of DNA methylation level for many CpGs and genes. This might be a limitation when studying DNA methylation of bulk tumors (meaning tumor + surrounding tissue) which are often used, as the DNA methylation measurements represents the DNA

methylation of the tumor part and cells in the surrounding tissue. For instance if the DNA methylation level of infiltrated non-neoplastic cells is less methylated than the tumor cells, a tumor with high infiltration of non-neoplastic cells would look like having a lower DNA methylation level than what is real. The opposite would be the case if the infiltrated non-neoplastic was more methylated than the tumor. In this way, tumor samples with high contamination of non-neoplastic cells will on the wrong basis perhaps look more similar to each other with regard to DNA methylation than they in reality are.

Several examples of DNA methylation studies that might be affected by using bulk tumors exist. One example is studies that based on DNA methylation of cancer genes classify tumors. In previous studies an association between the gene expression subtypes and the DNA methylation profile has been detected (42-44). However, genes might be wrongly reported to have a DNA methylation pattern that is correlated to a subtype since perhaps the amount of infiltrated non-neoplastic cells is different between the samples with different subtypes. The genes found in the genome-wide correlation analysis in the present study where the DNA methylation level was correlated to tumor percentage were compared to two previous studies (42;43), reporting that the DNA methylation level of cancer related genes were associated with the subtypes. Interestingly, around 1/3 of the genes in these articles were found in this study (FDR<0.05) to have a DNA methylation level correlated to tumor percentage. This illustrates the importance of considering tumor percentage when reporting these results, to avoid reporting findings on improper basis. Since the CpGs used in these articles are from the promoter region, they were compared to the genes with CpGs in the promoter region found in this study. However, the CpGs might not have the same genomic position, thus this result should be interpreted with caution. In the future it is important to compare CpGs with the same genomic.

Another example where DNA methylation studies might be affected by using bulk tumors regards studying DNA methylation as a potential biomarker. DNA methylation has shown great potential as a biomarker, but challenges by using DNA methylation as a biomarker has also been reported. The finding in the present study presents another challenge. The fact that the intra-tumor heterogeneity was shown to be associated with DNA methylation as revealed in this study, shows that it does matter where in the sample DNA methylation is measured. As a consequence, this might prevent accurate treatment options, prognostication, prediction of therapy response, diagnosis and risk of recurrence assessment. In addition, this might prevent validation of DNA methylation from previous findings.

Another possible challenge in the future where DNA methylation analysis might be affected by using bulk tumors regards drugs used to induce DNA demethylation and re-expression of target genes. Currently, drugs used to induce DNA demethylation is not used as a treatment option in Norway for breast cancer, but has been used in clinical trials in other countries like USA (134). In these clinical trials the selection of which patients that are treated with drugs inducing DNA demethylation has not been decided based on DNA methylation as a marker. However, if this would be the case in the future, this could be a challenge due to the intra-tumor heterogeneity. Since the results shown that the amount of tumor influence the DNA methylation, the measured DNA methylation of a bulk tumor will not necessary give the right indication. For instance if a patient has a high DNA methylation level in the region with the most tumor and a low DNA methylation level in the region with mostly stroma, the DNA methylation measurement would be lower than what the tumor part in realty is. This patient would perhaps not be treated with DNA demethylation agents due to too low DNA methylation measurement, even though the DNA methylation level of the tumor cells are higher.

6.1.1.2 Ingenuity pathways (IPA)

The gene list uploaded into IPA represents genes showing the most difference in DNA methylation level between non-neoplastic cells and tumor cells. Many of the functions that were overrepresented are not surprising since they are important in the development and progression of cancer. For instance, the top function that was overrepresented in the list is cell-to-cell interaction and signaling, which is important for acquiring many of the hallmarks. For instance, cell-to-cell interaction and signaling is important for sustaining proliferative signaling (9). This capability can be maintained in several ways. One alternative is that a tumor sends signals to the non-neoplastic cells. The non-neoplastic cells "reply" by sending growth factors back to the tumor, and in this way proliferative signaling is maintained. In Cheng *et al*. (132) it was shown that "TGF-β signaling deficient fibroblasts enhance Hepatocyte Growth Factor signaling in mammary carcinoma cells". Another function that was overrepresented is cell death and survival which is of great importance for the cancer cells. The functions overrepresented in the list might in varying degree be active in the non-neoplastic cells compared to the tumor cells due to different DNA methylation levels of the genes related to these functions. This might again lead to different level of gene expression, which might lead to different activity of the present function.

Three pathways were overrepresented in the list. These pathways enable cell-to-cell or cell-to-ECM interactions. Previous studies have shown that essential molecules of these pathways might lead to cancer. Aberrant expression of E-cadherin important for Epithelial Adherens Junctions Signaling has shown to play a role in invasion and metastases of breast cancer (135;136). In the pathway Tight Junction Signaling claudin 7 has shown to be downregulated in esophageal cancer (137). An increased level of ILK which has a role in ILK Signaling has been reported in various cancers and in general tumors with increased tumor grade have increased ILK expression, such as prostate, colon and breast cancer (138-140). Like for the functions above, the activity of these pathways may be different between non-neoplastic cells and tumor cells possibly due to different DNA methylation profiles.

6.1.1.3 Pattern of DNA methylation of the 3 tumors

For the three tumors studied by pyrosequencing it was shown that for around half of the genes the basal-like tumors had different DNA methylation level than the luminal tumor when taking the mean methylation of all regions in each tumor. This is in agreement with what other studies have found (42;43).

It was also observed that the DNA methylation level of more genes was related to the amount of non-neoplastic cells in the basal-like tumors than the luminal tumor. Besides, greater difference in DNA methylation level between the regions was observed in the basal-like tumors. This is however not surprising since the difference in amount of tumor between the regions was less in the luminal tumor versus the basal-like tumors when considering both the HE-staining sections.

Even if this result can be explained by the fact that the luminal tumor had less difference in DNA methylation levels between the regions, other theories may also partly be applicable. For instance, this result may partly be due to different types of tumor cells and/or stromal cells in the basal-like tumors versus the luminal tumor. DNA from stroma comes from different types of cells such as fat cells, endothelial cells, pericytes, mast cells, macrophages, lymphocytes and fibroblasts. Also, the tumors might be separated further into different types, such as CD24+ - and CD44+ tumor cells. These different cells might have different DNA methylation profiles. If so, and the basal-like tumors consist of different amount and composition of these cells than the luminal tumor, this might partly play a role for the difference observed between the luminal tumor and the basal-like tumors. Also, the basal-like tumors generally consist of much more lymphocytes than the luminal ones (141). This was

shown in this study as well, and might also partly play a role for the difference observed between the basal-like tumors and the luminal tumor.

A monotone trend is expected to be observed in the tumors where 1) observable differences in amount of tumor between the regions were observed and 2) an association between tumor amount and DNA methylation level was detected. The basal-like tumors showed differences in amount of tumor between the regions. As expected a monotone trend was detected for the basal-like tumor 137 for those genes where an association between the DNA methylation level and tumor amount was observed. For the basal-like tumor 155, a monotone trend was not considered as explained earlier. Remarkably, in the basal-like tumors, the direction of the monotone trend/association between DNA methylation and tumor percentage was remarkably the same for all genes expect two. On the other hand, a monotone trend was only seen for 3 genes in the luminal tumor. For the other genes, region B which consisted of some tumor had either the highest or lowest mean of methylation. For three of these genes the region with more tumor and less tumor were most similar with regard to DNA methylation level. This is perhaps not surprising since the HE-stained section of the end further down in the included tumor piece barely showed any differences between the regions. Moreover, the difference in DNA methylation that was shown may be due to other reasons than as a consequence of apparent differences in amount of tumor. Like above, this difference might be a result of different types of tumor cells and/or stromal cells between the regions that have different DNA methylation profiles. Another possibility might be that there are different clones of tumor cells in the different regions. These clones may have obtained different DNA methylation level of selected genes.

Interestingly, which region the normal sample was closets to with regard to DNA methylation level differed between tumors and between genes. The region with the least tumor (meaning a lot of stroma) could be assumed to be the closest to normal sample since stroma is part of the normal tissue. Since the normal sample is not a match to the unhealthy patients (tumor 137,142 and 155) but rather contains the mean of three healthy patients this was first believed as a possible explanation for why the region closest to the normal varies. However, as previously described, the DNA methylation level of normal samples have shown to be similar (43;44). Therefore, this argument is less likely.

It is important when discussing the trends to consider that there only were 3 tumors used in this study, and only one luminal tumor. This is a weakness which makes interpretation difficult. Despite that only 3 tumors were available, trends were observed.

6.1.2 Comparison of the result of the genome-wide correlation analysis with the result of pyrosequencing of the macrodissected tumors for the overlapping CpGs

Of the 205 CpGs used in the pyrosequencing, 32 were observed on the Infinium® HumanMethylation450 BeadChip array. 15 probes/CpGs of these were significantly correlated and related to 6 of 22 genes used for the pyrosequencing.

These two studies were used for investigating the same issue. However, some differences between them are worth to be aware of. The genome-wide correlation is a statistical method and the significance was based on all samples (n=113). The DNA methylation results were obtained genome wide by using the Infinium® HumanMethylation450 BeadChip and the tumor percentage was obtained by using ASCAT. On the other hand, the other DNA methylation study was based on only 3 tumors. Due to the low number of tumors in this study, a statistical analysis was not performed. The DNA methylation values were obtained by pyrosequencing selected CpGs in a panel of genes, and the reporting was based on each sample. In addition, the tumor percentage was not detected, however, if a region consisted of many or few tumor cells was stated.

The DNA methylation study performed by pyrosequencing only consisted of 3 tumors. However, the result of the overlapping CpGs from the correlation analysis and pyrosequencing corresponded remarkably well, especially for the basal-like tumors. The reason for why the luminal didn't correspond that well, is probably mostly due to the fact that the luminal tumor barely showed any difference in amount of tumor between the regions further down in the included tumor piece. Of the 113 samples used for the genome-wide correlation analysis, the majority of the samples were luminal tumors. Therefore, it is likely that also the level of DNA methylation of the luminal tumors is associated to the amount of tumor.

Of the 32 CpGs that were overlapping, only a few of these were related to protein-coding genes. This is because the PCR-primers of the miRNAs were designed to match CpGs at the Infinium® HumanMethylation450 BeadChip (Illumina). Therefore, it is not unexpected that of the significant correlated CpGs matched to the CpGs used in pyrosequencing representing 6 genes; only 1 was a protein-coding gene.

6.1.3 The DNA methylation levels of subpopulations in breast cancer cell lines

Different DNA methylation levels between samples of the same subpopulation in a cell line (replicates) were revealed for *RASSF1A* and *FOXC1*. Since the variation in DNA methylation level was high between the replicates, the results weren't trustworthy, indicating by a standard deviation above 4. The only difference between the samples is that they have been cultured separately.

The literature reflects the complexity of using breast cancer cell lines to study DNA methylation. On one hand, many other studies have used DNA methylation data from breast cancer cell line to explore biological theories (142-144). On the other hand, studies have shown that genetic and epigenetic changes, including DNA methylation, may occur in the cell lines during cell culturing (145-147) . Thus, the results in the present study may not be unexpected. Furthermore, a recent study has shown that replicas of the same cell line, in all cell lines studied were differently methylated for a specific group of CpGs (148). Further studies should be done to enlighten this area.

6.2 Methodological considerations

6.2.1 Dissection of tumor

In this study macrodissection was performed to distinguish between regions with infiltrated non-neoplastic cells and tumor cells. However, it was difficult to achieve since tumor cells and non-neoplastic cells are often intermingled. Thus, the tumors were crudely separated into regions with greater or smaller fractions of tumor cells compared to stromal cells.

6.2.2 Modification of DNA

The methods used for methylation profiling: the Infinium® HumanMethylation450 BeadChip and pyrosequencing depend on the unmethylated cytosines to be converted to uracils. This is achieved by bisulfite treatment. This method is rough, and might result in fragmentation of up to 90% of the template DNA (149). Also, the chemicals and the salts in the bisulfite treatment may inhibit downstream methods such as PCR. Following, it is important to use enough DNA to make sure that enough DNA templates are available for downstream methods.

6.2.3 Nanodrop

Nanodrop was used as an estimate to quantify the DNA concentration. A limitation with this method is that it uses the absorbance of light at 260 nm to determine the DNA concentration. In this way, too fragmented DNA that cannot be used as template for further analysis or

components that absorb at the same wavelength are measured. Because of this, the method can estimate a higher amount of DNA in a sample, than what is the real or can be used as a template. Therefore, more than 10 ng which is the minimum amount to input to the PCR were used (**see chapter 6.2.5**).

6.2.4 PCR product of the subpopulations of the breast cancer cell lines

6 samples were taken out of analysis due to no PCR product. 4 of these were concentrated by a SpeedVac®. All of the concentrated samples needed more concentration than the rest to get a strong enough band on the gel. Due to the limits with bisulfite treatment, written about above, enough DNA must be available. Besides, concentration of DNA is also a rough method. In this method the DNA may be more fragmented. Thus, expectedly no PCR product was obtained for some of the samples that had been concentrated. The two others, that had not been concentrated, might for instance have been fragmented to a great extent during the bisulfite treatment. As a consequence, a PCR product was not obtained.

6.2.5 Methylation profiling

The DNA methylation profile was assessed genome-wide by using the Infinium® HumanMethylation450 BeadChip and for selected CpGs within a panel of genes by using pyrosequencing. The first method is a high throughput analysis, enabling as many as 485764 CpGs analyzed at the same time (108). Besides, this array covers many CpG sites in the promoter, body and 3'UTR region of protein-coding genes and miRNAs as well as intergenic regions (108). The detection limit is 10%. A limitation with this method is that the CpGs are preselected by the manufacturer. However, this is not of importance in this study since this is a pilot study, and it first of all is important to test the hypothesis for any CpG. Another limitation is that relatively limited CpGs are analyzed. By using a sequencing approach more CpGs could be analyzed simultaneously. However, sequencing is still more expensive and involves more challenging bioinformatical pipelines than microarray based methods.

Pyrosequencing on the other hand, is a lower throughput analysis, but in return CpG sites in close proximity can be analyzed (116). In contrast, the CpGs are usually some distance away from each other on the Infinium® HumanMethylation450 BeadChip. Unlike the Infinium® HumanMethylation450 BeadChip, the CpGs to be studied by pyrosequencing are not pre-selected. To perform pyrosequencing, primers for sequencing and PCR have to be designed and made upfront. It is important that these do not contain SNPs or CpGs and that the sequence to be read contains at least one cytosine without guanine after, as a control for

successful bisulfite treatment (116). Besides, no template control is important to use during PCR and pyrosequencing as a control of contamination. Likewise, another control of contamination is the extra bases added in the sequence to be read in the pyrogram. No signal should be observed for these bases. The detection limit of pyrosequencing is around 5-10%, and the minimum amount DNA into the PCR should be at least 10 ng to ensure good reproducibility (150). With a reduced amount, random amplification of one strand can occur more frequently during the first cycles in PCR (150). A limitation with this method is the short reading length. The PCR product should not be more than 350 base pair to avoid any secondary structures (116).

6.2.6 Observations regarding pyrosequencing

Optimized pyrosequencing results were achieved in the tumor regions for the protein-coding genes when increasing the DNA input to the PCR to approximately 15 ng. This optimization was taken into account when performing pyrosequencing for the miRNAs in the tumor regions which was performed by a colleague and for the protein-coding genes in the CD24/44 sorted cell line samples. Therefore, approximately 15 ng DNA was used as input into the PCR for all genes. The 5% of the total pyrosequencing runs were run in duplicates for the tumor regions for the miRNAs. These duplicates showed low variation in DNA methylation level. This suggests that also these results seem to be trustworthy. The variation in regard to DNA methylation level was also low for each CD24/CD44 sorted sample/replicate, indicating that the pyrosequencing method itself was functioning correctly.

The positive controls were not 0% and 100% methylated as they were supposed to be. However, for the 0% and 100% methylated control, relatively low and high DNA methylation values were obtained, respectively. Moreover, the DNA methylation level differed from CpG to CpG within a gene. Therefore, these controls cannot be used as a control of the DNA methylation percentage, but rather as a quality control to help understand what can have gone wrong if pyrosequencing results aren't detected. The manufacturer of the controls has later stated that these controls are not 0% or 100% methylated, as the protocol indicates. Instead their DNA methylation level was low or high and differed from CpG to CpG within a gene. Thus, the results of the controls were as expected.

The accuracy of the DNA methylation values in each gene was not controlled since the controls weren't working as supposed. Anyhow, PyroMark control oligo samples were used to control the accuracy of the DNA methylation values obtained by the pyrogram. All of these

samples showed exactly a DNA methylation level of 50%, which they were supposed to. The PyroMark Q96 MD system can therefore be concluded to measure correctly. This strengthens the reliability of the DNA methylation levels obtained in this study. Nonetheless, the PyroMark control oligo samples do not require any sequencing primer during the procedure. In this way, the genes itself cannot be controlled. Therefore, there is a chance that a SNP that isn't known can be part of the sequencing primer. This SNP can be linked to the DNA methylation value. Ergo, the template DNA that will not be pyrosequenced due to a SNP, may have another methylation profile, which could have changed the DNA methylation level. A positive control added in the pyrosequenced run can only notice this if the SNP is included in the control. If so, the sequencing primer could be moved to another preferential place in the DNA sequence. For this possibility to occur, many coincidences must be present, which makes this possibility less likely.

The camera in the PyroMark Q96 MD system did not detect the light proportional with the amount of ATP that well. In this way, it was often necessary to have stronger bands on the gel and a higher volume of the PCR product to be used as a sequencing template than normal. However, the camera varied in how well it detected the light. Therefore, there was a narrow range of how much amount of the PCR product that could be used as input to avoid supersaturating when the camera worked well.

Pyrosequencing results for the tumor regions were achieved for more CpGs within the miRNAs than for the CpGs in the protein-coding genes. This might be due to that longer sequence templates often were used for the protein-coding genes. The substrate and enzyme mix are added in the beginning of a pyrosequencing run. The longer the sequence template is, the weaker the signal on the pyrogram will be towards the end of the signal due to lower amount of substrate and enzyme. Additionally, for the protein-coding genes pyrosequencing results were achieved for fewer CpGs in the normal samples called RP2, RP10 and RP11 than in the tumor samples. This was unexpected since both the tumor samples and the RP samples had similar strength on their bands when performing agarose gel electrophoresis. Besides, both bands were very strong. The samples were prepared for pyrosequencing by the following step: clean-up of streptavidin-bound PCR-product, denaturation and annealing of sequencing primer. In this procedure, it is important that streptavidin beads with the template DNA are transferred to the wells with the sequencing primer, and this was the case for the RP samples. However, a possibility is that the sequencing primer doesn't attach that well for these genes, due to a SNP in the RP samples. If a SNP was present in the sequence was once more checked

(already checked once when designing the primers), and no SNP was found. However, there might be a SNP that isn't known yet that is present in the sequence. Anyway, it is not very likely that this would be the case for all genes, and is still not known.

6.2.7 Statistics and bioinformatics

ASCAT is an algorithm estimating the copy number of tumors, the tumor ploidy and the tumor percentage (110). The limitation with ASCAT is that it can't distinguish between 0% and 100% tumor, as explained previously under **chapter 4.6.1**.

IPA is a database that is manually curated by a set of people to ensure accurate scientific data at all times (151). The wealth of information provided by IPA is also a challenge, which might make it difficult to interpret data.

7 Conclusion

In this pilot study it has been shown that infiltrating non-neoplastic cells influence the measurements of the DNA methylation level for many CpGs and genes. This was performed genome-wide for 113 bulk tumors and for a panel of genes in 3 macrodissected tumors. This is a challenge since many studies use bulk tumors, thus the DNA methylation level comprises the DNA methylation levels of tumor cells as well as the DNA methylation levels of stromal cells. This might hamper DNA methylation studies like investigation of DNA methylation as a biomarker, classification of tumors based on the DNA methylation profiles of cancer genes and investigation of drugs used to induce DNA demethylation. Since the non-neoplastic cells have been shown to influence the measurements of the DNA methylation profile for many CpGs and genes, it should be investigated further in larger studies with more samples covering CpGs in larger part of the genome. If the same pattern is seen in future studies, it might affect how DNA methylation studies will be performed and interpreted in the future.

The results from the correlation analysis and the pyrosequencing corresponded well for the overlapping CpGs. For the macrodissected tumors it was also observed that the DNA methylation level of more genes was related to the amount of tumor cells in the basal-like tumors than in the luminal tumor. Besides, greater difference in DNA methylation level between the regions was observed in the basal-like tumors. This might be mainly due to that the difference in amount of tumor between the regions was less in the luminal tumor versus the basal-like tumors. Of the 113 samples used for the genome-wide correlation analysis, the majority of the samples were luminal tumors. It is therefore likely that infiltrating non-neoplastic cells influence the measurements of DNA methylation level for the luminal tumors as well.

Different DNA methylation levels of *RASSF1A* and *FOXC1* were revealed between the samples of the same subpopulation of a cell line (replicates). As a consequence the results weren't trustworthy. The only difference between the replicates is that they have been cultured separately. Studies have shown that epigenetic changes, including DNA methylation, may occur during cell culturing. Further studies should be done to enlighten this area.

8 Future perspectives

In this thesis it has been shown in a pilot study that for many CpGs and genes the infiltrating non-neoplastic cells influence the measurements of DNA methylation level. Further studies should be done to enlighten this area in more samples. Next generation sequencing approaches, such as reduced representation bisulfite sequencing (RRBS), methylated DNA immunoprecipitation (MeDip-seq), and methylated DNA capture by affinity purification (methylCAP-seq) could be used to investigate more CpGs in the genome per sample. If further studies show the same trend, it might be important to take this into account when working with DNA methylation and circumvent the obstacle. One way is to check the genes/CpGs before being used against lists of genes/CpGs shown to have a DNA methylation level significantly associated to tumor percentage. It may also be possible to circumvent the obstacle, by getting rid of the slope that is created when plotting the DNA methylation level and tumor percentage of each CpG in a graph. The slope is an effect of the tumor percentage. The slope can be removed by using a linear regression model on each CpG. The absolute DNA methylation level of each CpG will be replaced with the residual of the regression model. However, this approach has never been reported before, and the validity of this approach must be checked thoroughly through further studies.

In further studies it could also be interesting to investigate the genes with CpGs where the DNA methylation level has been shown to be most strongly associated with tumor percentage in the correlation analysis.

In the present study, the regions of a tumor were not completely distinguished between tumor and stroma by using macrodissection. In further studies, regions with either tumor or stroma could be obtained by using microdissection. This approach is however time consuming and a lot of slices would be needed to obtain enough DNA. As previously written, stroma and tumor consist of different types of cells, and the DNA methylation levels of these types of cells might differ. Therefore, it could be also be beneficial to separate different cells types. Microdissection can also be used to distinguish between single cells. However, it might not be so easily to do due to cell-cell adhesion. In addition, relatively large amount of DNA needs to be obtained for the bisulfite treatment and pyrosequencing, which makes this method less suitable. Other and perhaps better approaches could be to use immunoprecipitation methods where antibodies attach to proteins prior to cell sorting. To be able to perform these methods, viable cells or fresh tumor tissue has to be available. To separate different cells fluorescence activated cell sorting or cell sorting by Dynabeads could for instance be used.

In this thesis the purpose of the breast cancer cell lines were not studied due to different DNA methylation levels between the samples of the same subpopulation of a cell line (replicates). Further studies with more cell lines covering CpGs in larger part of the genome should be investigated to explore the extent of this pattern for instance by next generation sequencing approaches. In this way, it is possible to explore if this pattern is applicable for larger or smaller regions in the genome, and if it is applicable for the same region for every cell line or if it's varying. If further studies show that that this tendency is applicable for a smaller part of the genome, further studies could be done to study the purpose in this study by using the breast cancer cell lines studying the regions within the genome that doesn't seem to have their epigenetic landscape altered upon cell culturing. Then, it would be beneficial to use cell lines with high content of subpopulations to avoid concentration of the cells by SpeedVac® since this might lead to fragmented DNA which might be difficult to make use of when studying DNA methylation.

In this study, nanodrop was used to measure the DNA concentration. Nanodrop is as explained previously not ideal for this method as it measures all DNA no matter how fragmented it is and other components that absorb at the same wavelength. This is a challenge in this thesis as both concentration of DNA by using SpeedVac® and bisulfite-treatment might lead to fragmented DNA. As a consequence, the DNA concentration measured at the nanodrop might consist of DNA that cannot be used in the following PCR since it is too fragmented. Since the input of DNA into the PCR has been shown in this thesis to be crucial for proper analysis, it is important that the concentration of the DNA reflects as much as possible DNA that can be used in the PCR. To overcome this challenge in this thesis, more DNA than required was used in the PCR. A better method could be to use quantitative PCR (qPCR) before running PCR. By using qPCR, the concentration of DNA will be measured based on the PCR products. To generate a PCR product, a certain length of the DNA must be available, and in this way, it is more believable that the concentration of DNA reflects in higher degree DNA template that can be used in the PCR.

Reference List

(1) World Health Organization, Cancer
[Online] Available at:
http://www.who.int/mediacentre/factsheets/fs297/en/index.html
Accessed date: 8.6.2013

(2) World Health Organization, GLOBOCAN 2012, Esitmated Cancer Incidence, Mortality and Prevalence Worldwide in 2012
[Online] Available at:
http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx
Accessed date: 29.9.2014

(3) CANCER Registry of Norway, 30 099 nordmenn fikk kreft i 2012
[Online] Available at:
http://www.kreftregisteret.no/no/Generelt/Nyheter/-30-099-tilfeller-av-kreft-i-2012/
Accessed date: 29.9.2014

(4) Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of THE CELL. 5 ed. Garland Science; 2008.

(5) Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 2003 March;33 Suppl:238-44.:238-44.

(6) Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature 2007 May 24;447(7143):433-40.

(7) Esteller M. Epigenetic changes in cancer. F1000 Biol Rep 2011;3:9. doi: 10.3410/B3-9. Epub;%2011 May 3.:9.

(8) Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000 January 7;100(1):57-70.

(9) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011 March 4;144(5):646-74.

(10) Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010 December 15;127(12):2893-917.

(11) Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2014 September 13;10.

(12) KREFTregisteret, 28 271 tilfeller av kreft i 2010
[Online] Available at:
http://www.kreftregisteret.no/no/Generelt/Nyheter/Cancer-in-Norway-2010/
Accessed date: 8.6.2013

(13) KREFTregisteret, Brystkreft

[Online] Available at:
http://www.kreftregisteret.no/no/Generelt/Fakta-om-kreft-test/Brystkreft-Alt2/
Accessed date: 14.12.2014

(14) Ali S, Coombes RC. Endocrine-responsive breast cancer and strategies for combating resistance. Nat Rev Cancer 2002 February;2(2):101-12.

(15) Lanigan F, O'Connor D, Martin F, Gallagher WM. Molecular links between mammary gland development and breast cancer. Cell Mol Life Sci 2007 December;64(24):3159-84.

(16) BREAST CANCER.ORG, What is breast cancer?
[Online] Available at:
http://www.breastcancer.org/symptoms/understand_bc/what_is_bc
Accessed date: 8.6.2013

(17) Stephan Pam, Epithelial Cells Definition
[Online] Available at:
http://breastcancer.about.com/od/cgterms/g/Mammary-Epithelial-Cells-Definition.htm
Accessed date: 3.11.2014

(18) Johns Hopkins Medicine, Breast Biopsy
[Online] Available at:
http://www.hopkinsmedicine.org/healthlibrary/test_procedures/gynecology/breast_biopsy_92,P07763/
Accessed date: 3.11.2014

(19) Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med 2008 November 13;359(20):2143-53.

(20) Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013 April;45(4):353-2.

(21) Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013 April;45(4):392.

(22) Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Annu Rev Public Health 1996;17:47-67.:47-67.

(23) McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ 2000 September 9;321(7261):624-8.

(24) Richard Doll Building. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol 2012 November;13(11):1141-51.

(25) Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. J Natl Cancer Inst 2013 April 17;105(8):515-25.

(26) Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. J Natl Cancer Inst 2013 July 17;105(14):1043-9.

(27) Lakhani S.R DSAFCJDA. BASIC PATHOLOGY AN INTRODUCTION TO THE MECHANISMS OF DISEASE. Hodder Arnold; 2003.

(28) Polyak K. Breast cancer: origins and evolution. J Clin Invest 2007 November;117(11):3155-63.

(29) Weigelt B, Horlings HM, Kreike B, Hayes MM, Hauptmann M, Wessels LF et al. Refinement of breast cancer classification by molecular characterization of histological special types. J Pathol 2008 October;216(2):141-50.

(30) CANCER RESEARCH UK, Rare types of breast cancer
[Online] Available at:
http://www.cancerresearchuk.org/cancer-help/type/breast-cancer/about/types/rare-types-of-breast-cancer.
Accessed date: 10.6.2013

(31) Ignatiadis M, Sotiriou C. Understanding the molecular basis of histologic grade. Pathobiology 2008;75(2):104-11.

(32) Murphy CG, Fornier M. HER2-positive breast cancer: beyond trastuzumab. Oncology (Williston Park) 2010 April 30;24(5):410-5.

(33) Wilken JA, Maihle NJ. Primary trastuzumab resistance: new tricks for an old drug. Ann N Y Acad Sci 2010 October;1210:53-65. doi: 10.1111/j.1749-6632.2010.05782.x.:53-65.

(34) Harris LN, You F, Schnitt SJ, Witkiewicz A, Lu X, Sgroi D et al. Predictors of resistance to preoperative trastuzumab and vinorelbine for HER2-positive early breast cancer. Clin Cancer Res 2007 February 15;13(4):1198-207.

(35) Nishimura R, Arima N. Is triple negative a prognostic factor in breast cancer? Breast Cancer 2008;15(4):303-8.

(36) de AE, Cardoso F, de CG, Jr., Colozza M, Mano MS, Durbecq V et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer 2007 May 21;96(10):1504-13.

(37) Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003 July 8;100(14):8418-23.

(38) Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001 September 11;98(19):10869-74.

(39) Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. Nature 2000 August 17;406(6797):747-52.

(40) Perou CM, Borresen-Dale AL. Systems biology and genomics of breast cancer. Cold Spring Harb Perspect Biol 2011 February 1;3(2):a003293.

(41) Sorlie T. Molecular classification of breast tumors: toward improved diagnostics and treatments. Methods Mol Biol 2007;360:91-114.:91-114.

(42) Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast Cancer Res 2010;12(3):R36.

(43) Ronneberg JA, Fleischer T, Solvang HK, Nordgard SH, Edvardsen H, Potapenko I et al. Methylation profiling with a panel of cancer related genes: association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. Mol Oncol 2011 February;5(1):61-76.

(44) Kamalakaran S, Varadan V, Giercksky Russnes HE, Levy D, Kendall J, Janevski A et al. DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. Mol Oncol 2011 February;5(1):77-92.

(45) Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG et al. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. PLoS One 2011 February 22;6(2):e16915.

(46) Russnes HG, Vollan HK, Lingjaerde OC, Krasnitz A, Lundin P, Naume B et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. Sci Transl Med 2010 June 30;2(38):38ra47.

(47) Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012 April 18;486(7403):346-52.

(48) Gonzalez-Angulo AM, Hennessy BT, Meric-Bernstam F, Sahin A, Liu W, Ju Z et al. Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. Clin Proteomics 2011 July 8;8(1):11-8.

(49) Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 2011 August;22(8):1736-47.

(50) Place AE, Jin HS, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. Breast Cancer Res 2011;13(6):227.

(51) Kim JB, Stein R, O'Hare MJ. Tumour-stromal interactions in breast cancer: the role of stroma in tumourigenesis. Tumour Biol 2005 July;26(4):173-85.

(52) Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 2004 July;6(1):17-32.

(53) Hu M, Yao J, Cai L, Bachman KE, van den Brule F, Velculescu V et al. Distinct epigenetic changes in the stromal cells of breast cancers. Nat Genet 2005 August;37(8):899-905.

(54) Fiegl H, Millinger S, Goebel G, Muller-Holzner E, Marth C, Laird PW et al. Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. Cancer Res 2006 January 1;66(1):29-33.

(55) Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. Mol Biotechnol 2010 January;44(1):71-81.

(56) Sutherland JE, Costa M. Epigenetics and the environment. Ann N Y Acad Sci 2003 March;983:151-60.:151-60.

(57) Lee WJ, Zhu BT. Inhibition of DNA methylation by caffeic acid and chlorogenic acid, two common catechol-containing coffee polyphenols. Carcinogenesis 2006 February;27(2):269-77.

(58) Hamilton JP. Epigenetics: principles and practice. Dig Dis 2011;29(2):130-5.

(59) Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT et al. Dynamic changes in the human methylome during differentiation. Genome Res 2010 March;20(3):320-31.

(60) Fleischer T, Edvardsen H, Solvang HK, Daviaud C, Naume B, Borresen-Dale AL et al. Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. Int J Cancer 2014 June 1;134(11):2615-25.

(61) Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst 2000 April 5;92(7):564-9.

(62) Dammann R, Yang G, Pfeifer GP. Hypermethylation of the cpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers. Cancer Res 2001 April 1;61(7):3105-9.

(63) Toyooka KO, Toyooka S, Virmani AK, Sathyanarayana UG, Euhus DM, Gilcrease M et al. Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. Cancer Res 2001 June 1;61(11):4556-60.

(64) Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 1995 October 15;55(20):4525-30.

(65) Zochbauer-Muller S, Fong KM, Maitra A, Lam S, Geradts J, Ashfaq R et al. 5' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer. Cancer Res 2001 May 1;61(9):3581-5.

(66) Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res 1998 October 15;58(20):4515-8.

(67) Khan S, Kumagai T, Vora J, Bose N, Sehgal I, Koeffler PH et al. PTEN promoter is methylated in a proportion of invasive breast cancers. Int J Cancer 2004 November 10;112(3):407-10.

(68) Singh P, Yang M, Dai H, Yu D, Huang Q, Tan W et al. Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. Mol Cancer Res 2008 November;6(11):1710-7.

(69) Paredes J, Albergaria A, Oliveira JT, Jeronimo C, Milanezi F, Schmitt FC. P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. Clin Cancer Res 2005 August 15;11(16):5869-77.

(70) How KA, Nielsen HM, Tost J. DNA methylation based biomarkers: practical considerations and applications. Biochimie 2012 November;94(11):2314-37.

(71) Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. Genome Biol 2014;15(8):435-2333349012841587.

(72) Evron E, Dooley WC, Umbricht CB, Rosenthal D, Sacchi N, Gabrielson E et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. Lancet 2001 April 28;357(9265):1335-6.

(73) Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet 2012 October;13(10):679-92.

(74) Brueckner B, Kuck D, Lyko F. DNA methyltransferase inhibitors for cancer therapy. Cancer J 2007 January;13(1):17-22.

(75) Silverman LR, Demakos EP, Peterson BL, Kornblith AB, Holland JC, Odchimar-Reissig R et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. J Clin Oncol 2002 May 15;20(10):2429-40.

(76) Kantarjian H, Issa JP, Rosenfeld CS, Bennett JM, Albitar M, DiPersio J et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer 2006 April 15;106(8):1794-803.

(77) Issa JP, Garcia-Manero G, Giles FJ, Mannari R, Thomas D, Faderl S et al. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies. Blood 2004 March 1;103(5):1635-40.

(78) Segura-Pacheco B, Trejo-Becerril C, Perez-Cardenas E, Taja-Chayeb L, Mariscal I, Chavez A et al. Reactivation of tumor suppressor genes by the cardiovascular drugs hydralazine and procainamide and their potential use in cancer therapy. Clin Cancer Res 2003 May;9(5):1596-603.

(79) Villar-Garea A, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. Cancer Res 2003 August 15;63(16):4984-9.

(80) Russnes HG, Navin N, Hicks J, Borresen-Dale AL. Insight into the heterogeneity of breast cancer through next-generation sequencing. J Clin Invest 2011 October;121(10):3810-8.

(81) Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer 2012 April;%19;12(5):323-34.

(82) van der Poel HG, Oosterhof GO, Schaafsma HE, Debruyne FM, Schalken JA. Intratumoral nuclear morphologic heterogeneity in prostate cancer. Urology 1997 April;49(4):652-7.

(83) Fitzgerald PJ. Homogeneity and heterogeneity in pancreas cancer: presence of predominant and minor morphological types and implications. Int J Pancreatol 1986 July;1(2):91-4.

(84) Nik-Zainal S, Van LP, Wedge DC, Alexandrov LB, Greenman CD, Lau KW et al. The life history of 21 breast cancers. Cell 2012 May 25;149(5):994-1007.

(85) Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009 October 8;461(7265):809-13.

(86) Lea T. Immunologi og immunologiske teknikker. 3 ed. Fagbokforlaget; 2006.

(87) Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J et al. Molecular definition of breast tumor heterogeneity. Cancer Cell 2007 March;11(3):259-73.

(88) Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003 April 1;100(7):3983-8.

(89) Bloushtain-Qimron N, Yao J, Snyder EL, Shipitsin M, Campbell LL, Mani SA et al. Cell type-specific DNA methylation patterns in the human breast. Proc Natl Acad Sci U S A 2008 September 16;105(37):14076-81.

(90) Halvorsen AR, Helland A, Fleischer T, Haug KM, Grenaker Alnaes GI, Nebdal D et al. Differential DNA methylation analysis of breast cancer reveals the impact of immune signaling in radiation therapy. Int J Cancer 2014 November 1;135(9):2085-95.

(91) Aure MR, Leivonen SK, Fleischer T, Zhu Q, Overgaard J, Alsner J et al. Individual and combined effects of DNA methylation and copy number alterations on miRNA expression in breast tumors. Genome Biol 2013 November;%20;14(11):R126-14.

(92) Chen L, Li Y, Fu Y, Peng J, Mo MH, Stamatakos M et al. Role of deregulated microRNAs in breast cancer progression using FFPE tissue. PLoS One 2013;8(1):e54213.

(93) Tahiri A, Leivonen SK, Luders T, Steinfeld I, Ragle AM, Geisler J et al. Deregulation of cancer-related miRNAs is a common event in both benign and malignant human breast tumors. Carcinogenesis 2014 January;35(1):76-85.

(94) Lujambio A, Calin GA, Villanueva A, Ropero S, Sanchez-Cespedes M, Blanco D et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A 2008 September 9;105(36):13556-61.

(95) Weinberg RA. The biology of cancer. Garland Science; 2007.

(96) Muggerud AA, Ronneberg JA, Warnberg F, Botling J, Busato F, Jovanovic J et al. Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Res 2010;12(1):R3.

(97) Klajic J, Fleischer T, Dejeux E, Edvardsen H, Warnberg F, Bukholm I et al. Quantitative DNA methylation analyses reveal stage dependent DNA methylation and association to clinico-pathological factors in breast tumors. BMC Cancer 2013 October 5;13:456. doi: 10.1186/1471-2407-13-456.:456-13.

(98) Escale MM, Fuster G, Rusness H, Almendro V. FOXC1 Methylation status in CD44/CD24 cell subpopulations from luminal and basal breast cancer cell lines 2012.

(99) Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 2006 December;10(6):515-27.

(100) Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol Oncol 2007 June;1(1):84-96.

(101) ATCC, Product Description
[Online] Available at:
http://www.lgcstandards-atcc.org/
Accessed date: 9.2.2013

(102) Qiagen, QIAamp® DNA Mini and Blood Mini Handbook 2010
[Online] Available at:
http://www.qiagen.com
Accessed date: 29.11.2012

(103) Nanodrop, NanoDrop 1000 Spectrophotometer V 3.7 User's Manual 2008
[Online] Available at:
http://nanodrop.com/Library/nd-1000-v3.7-users-manual-8.5x11.pdf
Accessed date: 1.12.2012

(104) Qiagen, EpiTect® 96 Bisulfite Handbook 2009
[Online] Available at:
http://www.qiagen.com
Accessed date: 17.10.2013

(105) The American Joint Committee on Cancer (AJCC) staging system, TNM classification

[Online] Available at:
http://ccm.ucdavis.edu/bcancercd/311/tnm_staging.html
Accessed date: 3.12.2012

(106) Illumina. Infinium HD Assay Super Protocol Guide. Part # 11322427, REV.C. 2010.

(107) Illumina. Infinium® Assay Methylation Protocol Guide. Part # 11286471,Rev. A. 2008.

(108) Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 2011 June;6(6):692-702.

(109) Illumina, Interpreting Infinium® Assay Data for Whole-Genome Structural Variation
[Online] Available at:
http://res.illumina.com/documents/products/technotes/technote_cytoanalysis.pdf
Accessed date: 21.10.2013

(110) Van LP, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A 2010 September 28;107(39):16910-5.

(111) COOK D.J. CELLULAR PATHOLOGY. 2th ed. Scion; 2006.

(112) Promega, Techical Manual Maxwell® 16 DNA Purification Kits 2012
[Online] Available at:
http://www.promega.com/~/media/files/resources/protocols/technical%20manuals/0/maxwell%2016%20dna%20purification%20kits%20protocol.pdf?la=en
Accessed date: 30.11.2012

(113) Thermo Scientific, Assessment of Nucleic Acid Purity
[Online] Available at:
http://www.nanodrop.com/Library/T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf
Accessed date: 18.6.2013

(114) Qiagen, Epitec® Fast Bisulfite Conversion Handbook 2012
[Online] Available at:
www.qiagen.com
Accessed date: 29.11.2012

(115) Qiagen, PyroMark® PCR Kit 2011
[Online] Available at:
http://www.qiagen.com
Accessed date: 29.11.2012

(116) Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nat Protoc 2007;2(9):2265-75.

(117) Qiagen, Pyrosequencing® - the syngergy of sequencing and quantification 2010
[Online] Available at:
http://www.qiagen.com

Accessed date: 29.11.2012

(118) Qiagen, PyroMark® Control Oligo Handbook 2013
[Online] Available at:
www.qiagen.no
Accessed date: 30.5.2014

(119) Qiagen. PyroMark Q96 User Manual. 2010.

(120) Pallant J. SPSS Survival Manual. 3 ed. The McGraw - Hill Companies; 2007.

(121) Altman D.G. PRACTICAL STATISTICS FOR MEDICAL RESEARCH. 1 ed. CHAPMAN & HALL/CRC; 1999.

(122) Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics 2005 July 1;21(13):3017-24.

(123) Ingenuity® systems, Statistics used in IPA
[Online] Available at:
http://www.ingenuity.com/
Accessed date: 14.10.2013

(124) Maddika S, Kavela S, Rani N, Palicharla VR, Pokorny JL, Sarkaria JN et al. WWP2 is an E3 ubiquitin ligase for PTEN. Nat Cell Biol 2011 June;13(6):728-33.

(125) Carnero A, Blanco-Aparicio C, Renner O, Link W, Leal JF. The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. Curr Cancer Drug Targets 2008 May;8(3):187-98.

(126) Fernandez PM, Tabbara SO, Jacobs LK, Manning FC, Tsangaris TN, Schwartz AM et al. Overexpression of the glucose-regulated stress gene GRP78 in malignant but not benign human breast lesions. Breast Cancer Res Treat 2000 January;59(1):15-26.

(127) Li AL, Li HY, Jin BF, Ye QN, Zhou T, Yu XD et al. A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. J Biol Chem 2004 November;%19;279(47):49251-8.

(128) Ziyaie D, Hupp TR, Thompson AM. P53 and breast cancer. Breast 2000 October;9(5):239-46.

(129) Suijkerbuijk KP, Fackler MJ, Sukumar S, van Gils CH, van LT, van der Wall E et al. Methylation is less abundant in BRCA1-associated compared with sporadic breast cancer. Ann Oncol 2008 November;19(11):1870-4.

(130) Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet 2011 June 26;43(8):768-75.

(131) Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, Ruiz D, I et al. DNA methylation epigenotypes in breast cancer molecular subtypes. Breast Cancer Res 2010;12(5):R77.

(132) Cheng N, Chytil A, Shyr Y, Joly A, Moses HL. Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion. Mol Cancer Res 2008 October;6(10):1521-33.

(133) Moelans CB, de Groot JS, Pan X, van der Wall E, van Diest PJ. Clonal intratumor heterogeneity of promoter hypermethylation in breast cancer by MS-MLPA. Mod Pathol 2014 June;27(6):869-74.

(134) The U.S.National Institutes of Health, Department of Health and Human Services, ClinicalTrials.gov
[Online] Available at:
http://clinicaltrials.gov
Accessed date: 4.8.2014

(135) Oka H, Shiozaki H, Kobayashi K, Inoue M, Tahara H, Kobayashi T et al. Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis. Cancer Res 1993 April 1;53(7):1696-701.

(136) Kowalski PJ, Rubin MA, Kleer CG. E-cadherin expression in primary carcinomas of the breast and its distant metastases. Breast Cancer Res 2003;5(6):R217-R222.

(137) Lioni M, Brafford P, Andl C, Rustgi A, El-Deiry W, Herlyn M et al. Dysregulation of claudin-7 leads to loss of E-cadherin expression and the increased invasion of esophageal squamous cell carcinoma cells. Am J Pathol 2007 February;170(2):709-21.

(138) Marotta A, Tan C, Gray V, Malik S, Gallinger S, Sanghera J et al. Dysregulation of integrin-linked kinase (ILK) signaling in colonic polyposis. Oncogene 2001 September 27;20(43):6250-7.

(139) Graff JR, Deddens JA, Konicek BW, Colligan BM, Hurst BM, Carter HW et al. Integrin-linked kinase expression increases with prostate tumor grade. Clin Cancer Res 2001 July;7(7):1987-91.

(140) White DE, Cardiff RD, Dedhar S, Muller WJ. Mammary epithelial-specific expression of the integrin-linked kinase (ILK) results in the induction of mammary gland hyperplasias and tumors in transgenic mice. Oncogene 2001 October 25;20(48):7064-72.

(141) Tsang JY, Ni YB, Chan SK, Yamaguchi R, Tanaka M, Tan PH et al. Increased lymphocytic infiltration in breast cancer correlated with molecular subtypes and HER2 gene amplification. Histopathology 2013 May;62(6):963-5.

(142) Liu G, Liu YJ, Lian WJ, Zhao ZW, Yi T, Zhou HY. Reduced BMP6 expression by DNA methylation contributes to EMT and drug resistance in breast cancer cells. Oncol Rep 2014 August;32(2):581-8.

(143) Pathiraja TN, Nayak SR, Xi Y, Jiang S, Garee JP, Edwards DP et al. Epigenetic reprogramming of HOXC10 in endocrine-resistant breast cancer. Sci Transl Med 2014 March 26;6(229):229ra41.

(144) Jiao F, Bai SY, Ma Y, Yan ZH, Yue Z, Yu Y et al. DNA methylation of heparanase promoter influences its expression and associated with the progression of human breast cancer. PLoS One 2014 March 14;9(3):e92190.

(145) Schmidt M, Khan A, Schmidt AM, Heinze B, Hack E, Waltenberger J et al. A novel breast cancer cell line initially established from pleural effusion: evolution towards a more aggressive phenotype. Int J Oncol 2007 March;30(3):565-72.

(146) Maitra A, Arking DE, Shivapurkar N, Ikeda M, Stastny V, Kassauei K et al. Genomic alterations in cultured human embryonic stem cells. Nat Genet 2005 October;37(10):1099-103.

(147) Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD et al. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. Aging Cell 2010 February;9(1):54-63.

(148) Cocozza S, Scala G, Miele G, Castaldo I, Monticelli A. A distinct group of CpG islands shows differential DNA methylation between replicas of the same cell line in vitro. BMC Genomics 2013 October 10;14:692. doi: 10.1186/1471-2164-14-692.:692-14.

(149) Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 2001 July 1;29(13):E65.

(150) Dupont JM, Tost J, Jammes H, Gut IG. De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal Biochem 2004 October 1;333(1):119-27.

(151) Ingenuity® Knowledge Base, Knowledge Base
[Online] Available at:
http://www.ingenuity.com/science/knowledge-base
Accessed date: 22.9.2014

(152) Thermo Scientific, O'GeneRuler DNA Ladder Mix, Ready - to - Use 100-10000 bp
[Online] Available at:
http://www.thermoscientificbio.com/EktronTemplates/ProductLayout.aspx?id=17179919945&terms=GeneRuler+DNA+Ladder+Mix
Accessed date: 29.11.2012

Appendix A: reagents, equipment and instruments

Table A1: Reagents and equipments for cryomicrotome sectioning, HE-staining, microscopy DNA isolation by Maxwell® 16 instrument and NanoDrop 1000 Spectrophotometer.

Reagent /equipments	Supplier with address	Catalog number/ article number
Tissue-Tek	Chemi-Teknik as http://chemi-teknik.no/	4583
Microscope slides	Thermo Scientific http://www.thermoscientific.com	10143562CE
Glycerol 99.5% analar normapur pa 1 * 1 l	Prolabo https://no.vwr.com/app/Home	24388.295
acetic acid 100% normapur pa 1 * 1 l	Prolabo https://no.vwr.com/app/Home	20104298
Haematoxylin cryst.	Chroma-Gesellschaft http://www.merckmillipore.se	104302
Periodic acid normapur pa 1 * 50 g	Prolabo https://no.vwr.com/app/Home	20593151
Aluminium sulph. about 14 hydr. 1 * 1 kg	Prolabo https://no.vwr.com/app/Home	21070297
Absolute alcohol (ethanol)	Kemetyl http://www.kemetyl.no	
Eosin Y	Chroma-Gesellschaft www.chemdat.merck.de	115935
Xylen	Chemi-teknik http://chemi-teknik.no/	22500
Ammonia solution 25%	Merck www.chemdat.merck.de	105432
Microscope cover slips	Menzel-Glaser http://www.menzel.de	BB024050A1
Formalin	Chemi-teknik http://chemi-teknik.no/	1220-00A
Maxwell® 16 Tissue DNA Purification Kit	Nerliens Meszansky http://www.nmas.no	AS1030
Magentic elution rack	Nerliens Meszansky/ follow the machine	
Cartridge rack	Nerliens Meszansky/follow the machine	
Screw cap micro tube 2ml	Sarstedt http://www.sarstedt.com/php/main.php?newlanguage=en	72.693.005
Standard elution volume hardware kit	Nerliens meszansky http://www.nmas.no	AS1200

Table A2: Instruments used for cryomicrotome sectioning, HE-staining, microscopy DNA isolation by Maxwell® 16 instrument and NanoDrop 1000 Spectrophotometer

Instruments	Supplier
Maxwell® 16 Instrument	Promega
NanoDrop® ND-1000 Spectrophotometer	Saveen Werner A/S
Leica CM1950	Leica biosystems
Microm CTM6	Thermo Scientific
Olympus BX45	Olympus

Table A3: Reagents and equipments used for bisulfite treatment

Reagent/equipment	Supplier with address	Catalog number/ article number
ThermoFast 96 well PCR-plate	VWR https://no.vwr.com/app/Home	6.9724-22
EpiTect Fast DNA Bisulfite Kit	Qiagen http://www.qiagen.com/default.aspx	59824
ThermoFast cap strips	VWR https://no.vwr.com/app/Home	6.9724-26
Absolute alcohol (ethanol)	Kemetyl http://www.kemetyl.no	
1,5 ml Eppendorf tubes	VWR https://no.vwr.com/app/Home	213100-328

Table A4: Instruments used for bisulfite treatment

Instruments	Supplier
GeneAMP® PCR System 2700	Applied biosystems

Table A5: Reagents and equipments used for PCR, agarose gel electrophoresis and pyrosequencing

Reagent/equipment	Supplier with address	Catalog number / article number	Purity
Pyromark PCR kit (200)	Qiagen http://www.qiagen.com/default.aspx	978703	
Epitect unmethylated human control DNA	Qiagen http://www.qiagen.com/default.aspx	59665	
Epitect methylated human control DNA	Qiagen http://www.qiagen.com/default.aspx	59655	
GelRed Gel Stain 10.000x in water 1 *10 ml	VWR https://no.vwr.com/app/Home	730-2960	
GeneRulerTM DNA Ladder Mix	Thermo scientific http://www.thermoscientific.com	SM0333	
Certified Molecular Biology Agarose	Bio-Rad http://www.bio-rad.com	161-3102	
Trizma® base	Sigma-Aldrich http://www.sigmaaldrich.com/norway.html	93362	99.9%
EDTA	VWR https://no.vwr.com/app/Home	1.08418.1000	
Glacial acetic acid	VWR https://no.vwr.com/app/Home	1.00063.1000	
PyroMark Denaturation Solution (500 ml)	Qiagen http://www.qiagen.com/default.aspx	979007	
Absolut alcolhol (ethanol)	Kemetyl http://www.kemetyl.no		
Pyromark Gold Q96 Reagent kit	Qiagen http://www.qiagen.com/default.aspx	97804	
ThermoFast 96 well PCR-plate	VWR https://no.vwr.com/app/Home	6.9724-22	
ThermoFast cap stripes	VWR https://no.vwr.com/app/Home	6.9724-26	
PSQ HS 96 Plate	Qiagen http://www.qiagen.com/default.aspx	979101	
PyroMark Vacuum Prep Filter Probe(100)	Qiagen http://www.qiagen.com/default.aspx	979010	
PyromMark Q96 HS Reagent Dispensing Tip	Qiagen http://www.qiagen.com/default.aspx	979102	

Reagent/equipment	Supplier with address	Catalog number / article number	Purity
PyroMark Q96 HS Nucleotide Dispensing Tip	Qiagen http://www.qiagen.com/default.aspx	979103	
Abgene ® Thermo Scientific, 100 seals	VWR https://no.vwr.com/app/Home	732-0065	
PyroMark Control Oligo	Qiagen http://www.qiagen.com/default.aspx	979203	
PyroMark Binding Buffer	Qiagen http://www.qiagen.com/default.aspx	979006	
Streptavidin Sepharose High Performance 5 ml	Qiagen http://www.qiagen.com/default.aspx	17-5113-01	
PyroMark Wash Buffer	Qiagen http://www.qiagen.com/default.aspx	979008	
PyroMark Annealing Buffer	Qiagen http://www.qiagen.com/default.aspx	979009	

Table A6: Instruments used for PCR, agarose gel electrophoresis and pyrosequencing

Type of instruments	Supplier
Cooktronic M712	Philips
Sub-cell® Model 192	BIO-RAD
Genegenius	Syngene
Thermofixer comfort	Eppendorf
Heat block QBT2	Qiagen
PyroMark Q96 Vacuum work station	Qiagen
PyroMark Q96 MD System	Qiagen

Appendix B: recipes

HE-staining

Procedure

Solutions prepared before HE-staining:

Haematoxylin:

- 200 ml glycerol 99.5% analar normapur pa 1 * 1 l
- 100 ml acetic acid 100% normapur pa 1 * 1 l
- 650 ml distilled water
- 5 g haematoxylin cryst.
- 50 ml 1% periodic acid normapur pa 1 * 50 g
- 41.3 g aluminium sulph. about 14 hydr. 1 * 1 kg

Eosin:

- 800 ml absolute alcohol (ethanol)
- 200 ml distilled water
- 2.5 g eosin Y
- 5 ml acetic acid 100% normapur pa 1 * 1 l

Ammonia:

- Distilled water and 10-15 drops of ammonia solution 25%

HE-staining was performed by transporting the slides through the following steps:

- 4 minutes in 4% formalin
- Washed in water
- 90 seconds in haematoxylin
- Washed in water
- 15 seconds in ammonia
- Washed in water
- 60 seconds in eosin
- Washed in water

- 5 seconds in each of the steps: 70% ethanol - 96% ethanol - absolute alcohol (ethanol) - absolute alcohol (ethanol) - xylen-xylen
- Microscope cover slips were put on by the Microm CTM6.

Agarose gel electrophoresis

PCR-products checked on 1.5% Agarose gel electrophoresis stained with GelRed gel stain 10.000X in water 1* 10 ml was used to check the success of the PCR reaction. The principal with this method is that molecules of different size can be separated. In addition, the size of the PCR product achieved can be checked by comparing it to a gene ruler with different DNA fragments of known size that is applied to one lane of the gel (Figure B1).

Procedure

- 5.25 g certified Molecular Biology Agarose and 350 ml 1xTAEbuffer (Trizma® base, Glacial acetic acid, and edetic acid (EDTA)) were mixed and heated in the Cooktronic M712 microwave until the solution was resolved.
- The solution was cooled down before 35 µl GelRed gel stain 10.000X in water 1* 10 ml was added. The gel was added to a gel chamber and left until it was stiffened, around 30 minutes.
- The agarose gel was placed in the Sub-cell® Model 192 electrophoresis chamber.
- 2 µl of GeneRuler™ DNA Ladder Mix was applied in one lane of the gel.
- 5 µl PCR-product was applied to the other lanes of the gel.
- The Power Pac 300 was turned on to turn on the electric field and was running on 200 V and 400 mA for 30 minutes.
- The gel was visualized by using Genegenius.

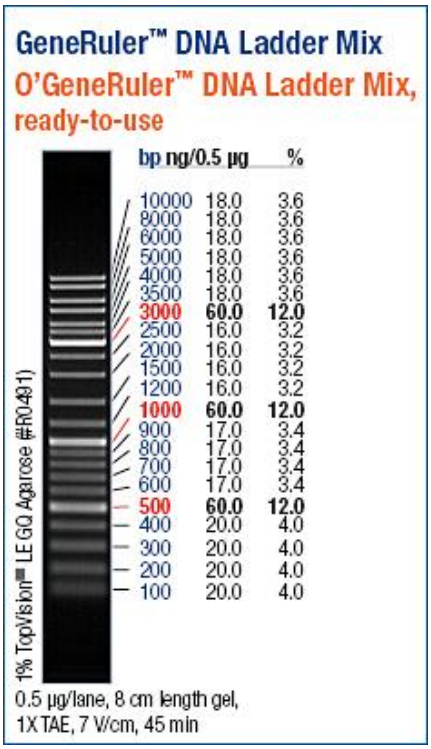


Figure B1: Gene Ruler™ DNA Ladder Mix. The ladder was used as size marker in agarose gel electrophoresis (152).

Appendix C: primers used in the pyrosequencing analysis

Table C1: An overview over the PCR- and sequencing primers used in the pyrosequencing analysis of the DNA methylation status. Primer name, primer type, primer sequence and pyrosequencing product are listed. PT: primer type, P: PCR and PS: pyrosequencing

Gene	Primer name	PT	Primer sequence	Pyrosequencing product	
				CpG	Position
BCAN	BCAN_1309f26	P	TTAATTTTAGTTTTGGGTAGGTGAGT		
	BCAN_1523_9r26		B-CAAAAATAACAACAACAACTACAAAC		
	BCAN_AMP1_pyro1f18	PS	TGGGTAGGTGAGTGTTTT	1	156612079
				2	156612087
				3	156612090
				4	156612094
				5	156612097
				6	156612105
	BCAN_AMP1_pyro2f18		GAGTGTTTGTTTGGTTTG	7	156612140
				8	156612147
				9	156612166
				10	156612188
				11	156612191
				12	156612204
				13	156612209
				14	156612232
				15	156612236
BNIPL	BNIPL_1513f27	P	TGGTTATTAGGAAAGTTGAATTAGAAT		
	BNIPL_1726_9r25		B-ATAACTAAAACCCAACTAAACCCAC		
	BNIPL_AMP1_pyro1f20	PS	TTATTAGGAAAGTTGAATTA	1	151009568
				2	151009588
				3	151009668
	BNIPL_AMP1_pyro2f17		GGTGGAGGAAAGGTAGG	4	151009699
				5	151009707
CTSA	CTSA_178f26	P	AGATTTTTTTATTTTTAAATTAGTGT		
	CTSA_474_9r21		B-CCTAAATATCCTCTTTTACCC		
	CTSA_AMP1_pyro3f17	PS	TTTTTGTGGAAGTAAAT	8	44518937
				9	44518947
				10	44518949
				11	44518962
				12	44518978
				13	44518982
				14	44518986
				15	44519000
				16	44519003
	CTSA_AMP1_pyro4f20		GTTGATGATGATGTGTATGT	17	44519024
				18	44519028

				19	44519036
				20	44519041
				21	44519043
FOXC1	FOXC1_3354f25	P	TAAGTAGGGTTGGTAGAATAGTATT		
	FOXC1_3568_9r25		B-ACAACCTATCCTTCTCCTCCTTATC		
	FOXC1_MSP3_PYRO4F20	PS	GGTAGAATAGTATTYGTTAT	2	1611072
				3	1611079
	FOXC1_MSP3_PYRO4F20/ FOXC1_MSP3_PYRO3F17		GGTAGAATAGTATTYGTTAT /TYGTTTAAYGAGTGTTT	4	1611088
				5	1611099
				6	1611101
				7	1611103
				8	1611147
				9	1611117
				10	1611138
				11	1611147
	FOXC1_MSP3_PYRO2F20		TTCGGATTTTTATAATATGT	12	1611166
				13	1611172
				14	1611185
				15	1611188
				16	1611191
				17	1611194
				18	1611197
				19	1611204
				20	1611207
FGFBP2	KSP37_1215r25	P	CTAAAAATAAAATTCCTCCCCAATA		
	KSP37_919_9f25		B-TGGTTGTTTTGTTTTTGTTAAGTGT		
	KSP37_AMP1_pyro3r18	PS	CCCACAAACCTTAAACAC	1	15962808
	KSP37_AMP1_pyro2r15/ KSP37_AMP1_pyro1r19		AAAATCCCCAAACAA/CCCAATAYTTCCTTACTTT	4	15962966
				5	15962978
	KSP37_AMP1_pyro1r19		CCCAATAYTTCCTTACTTT	6	15963034
GYPE	GYPE_1433f30	P	TTAAATTAAAGAGTTATTAATAAAGGTGTT		
	GYPE_1580_9r24		B-TTAATAATAAACCCACAAAAAAAC		
	GYPE_AMP1_pyro1f20	PS	AGTTAAAAGTATGAATATTG	1	144793560
				2	144793574
IL1A	IL1A_1785_9f25	P	B-AAAAATATAAAATTGAATTTTTTTT		
	IL1A_1961r25		AAACAATACCTTTACTAACTCAAAC		
	IL1A_AMP1_pyro2r20	PS	AACCTAAATCAATCTTCTTC	1	113533312
				2	113533320
				3	113533323
				4	113533337
				5	113533345
	IL1A_AMP1_pyro2r20/ IL1A_AMP1_pyro1r20		AACCTAAATCAATCTTCTTC/ CAATACCTTTACTAACTCAA	6	113533371
	IL1A_AMP1_pyro1r20		CAATACCTTTACTAACTCAA	7	113533427

IL1R2	IL1R2_1000_9r25	P	B-AAACACCCCCAAACTACAATACTAA		
	IL1R2_825f25		TTAATGGGGTTATAGTTGGGTAAAA		
	IL1R2_AMP1_pyro1f18	PS	TGGGGTTATAGTTGGGTA	1	102608155
				2	102608180
				3	102608192
				4	102608204
	IL1R2_AMP1_pyro2f20		GAAAAATAGGGAAATTTATG	5	102608233
				6	102608236
IL5RA	IL5RA_1227_9r26	P	B-AAACAACTATACAAAAACTAAACTAA		
	IL5RA_992f28		AAGGTATATAAATAAAATAGAAATGTAA		
	IL5RA_AMP1_pyro1f20	PS	TAAATAAAATAGAAATGTAA	1	3109027
				2	3109043
	IL5RA_AMP1_pyro2f20		TTAGATAATGGTTATTTGGA	3	3109166
				4	3109179
LTC4S	LTC4S_1200_9r25	P	B-ATAAAAATCTAAAACCCACCCTCTT		
	LTC4S_925f26		TTTGGTTTTGTGTGGTATGGTTATAT		
	LTC4S_AMP1_pyro1f20	PS	GTGTGGTATGGTTATATTTT	1	179220939
				2	179220965
				3	179220977
				4	179221002
	LTC4S_AMP1_pyro2f17		GTTTTTTGAGTAGTAGA	5	179221020
				6	179221030
				7	179221042
				8	179221059
				9	179221072
				10	179221075
				11	179221090
PCK1	PCK1_714f25	P	ATTTGTTTGTTAAGAAAAGGGTGTT		
	PCK1_990_9r25		B-AAAAATAACCCTACCCCTCAACTAC		
	PCK1_AMP1_pyro1f18	PS	TTATTATGGTTGGTTGAG	1	56135957
				2	56136012
	PCK1_AMP1_pyro2f18		AGTATAATTGATTTTGGT	3	56136053
				4	56136078
				5	56136087
RASSF1A	Rassf1A_2970f25	P	AGTTTTTGTATTTAGGTTTTTATTG		
	Rassf1A_3160_9r23		B-AACTCAATAAACTCAAACTCCCC		
	Rassf1A_MSP2_Pyro1f15	PS	GGATTTTGGGGGAGG	5	50378322
				6	50378313
				7	50378306
				8	50378293
				9	50378289
				10	50378284
				11	50378382

				12	50378272
	Rassf1A_MSP2_Pyro2f18		TGTTAGYGTTTAAAGTTA	13	50378259
				14	50378252
				15	50378242
				16	50378232
CDKN2A	P16_MSP1f22	P	GAGGGGTTGGTTGGTTATTAGA		
	P16_MSP1_9r24		B-TACAAACCCTCTACCCACCTAAAT		
	P16_MSP1_pyro1f16	PS	TGGTTATTAGAGGGTG	1	21974890
				2	21974885
				3	21974883
				4	21974879
				5	21974875
				6	21974872
				7	21974866
	P16_MSP1_pyro2f18		GAGGGGGAGAGTAGGTAG	8	21974844
				9	21974840
				10	21974837
				11	21974819
				12	21974816
				13	21974813
	P16_MSP1_pyro3f20		GGGGAGTAGTATGGAGTTTT	14	21974792
				15	21974774
				16	21974770
				17	21974768
				18	21974763
				19	21974757
	P16_MSP1_pyro3f20/ P16_MSP1_pyro4f15		GGGGAGTAGTATGGAGTTTT /GGGTYGGGTAGAGGA	20	21974742
				21	21974738
	P16_MSP1_pyro4f15		GGGTYGGGTAGAGGA	22	21974726
				23	21974720
				24	21974710
				25	21974705
				26	21974695
				27	21974691
				28	21974684
MIR199A	cg24002149F	P	B-TAAGGGTTGTGATTTTTAGTTTTGA		
	cg24002149R		AAACCCACTTCCTACCCAATTAA		
	cg24002149py3	PS	AATCTACAACAAATATACCA	1	172113920
	cg24002149py2		ATACCATTTTATACACAAAC	2	172114039
				3	172114047
				4	172114059
				5	172114068
	cg24002149py1		CTAAAAATAAAATAACTACT	6	172114124
MIR135B	cg13061767F	P	TGGAGTTATTTTTGATAAGGTTGTG		
	cg13061767R		B-AAAACCCTAAAAACCCACTTCTCTA		
	cg13061767py1	PS	GGTTGTGGGAGGGG	1	205418265

				2	205418278
	cg13061767py2		TGGATTGAAGAGTTGAGTTA	3	205418348
				4	205418410
				5	205418421
MIR16-2	cg09100593F	P	TTTTAGAAATTTAAGGAAATTTATTTAAAA		
	cg09100593R		B-TCCCTATCACACTAAAACAACACAA		
	cg09100593py1	PS	TGTTTTTATTATTAGATGTT	1	160122503
				2	160122550
				3	160122562
MIR887	cg19632594F	P	TGGGGGTTAAGTTTGTAAAAGAATA		
	cg19632594R		B-CCAAACCCCAACCTAAACAC		
	cg19632594py1	PS	AGTAAAATAGAGTTAATAGG	1	15500077
				2	15500085
				3	15500090
				4	15500099
				5	15500101
				6	15500103
				7	15500108
				8	15500110
				9	15500118
MIR148	cg03853208F	P	TTTTTAGTTAAGATAAAGAAGAATAATT		
	cg03853208R		B-TACAACCCCTTTACTATAACATTAC		
	cg03853208py1	PS	AAAAGTAATTTAGTTTAGGA	2	25989720
				3	25989724
				4	25989735
	cg03853208py2		GGTTATTAATTATGGAAAGT	5	25989763
				6	25989811
				7	25989816
				8	25989830
MIR200C/ 141 cluster	cg27534624F	P	GGTTGAGTTTGGGATTGTAGAG		
	cg27534624R		B-CCCTAAAAACACTTCCTAATAAACC		
	cg27534624py1	PS	GATGAGGGTGGGTAAAT	1	7072582
				2	7072592
				3	7072594
				4	7072599
				5	7072612
	cg27534624py2		GGAAGGTGGTTTAG	6	7072639
				7	7072642
				8	7072647
				9	7072650
				10	7072654
				11	7072664
	cg27534624py3		GTTTGTTTGGATTGTAATTT	12	7072696
				13	7072705
				14	7072727

MIR17-92 cluster	cg23665802F	P	TTTGTATATTTTTGGAAATTGGTTTATG		
	cg23665802R		B-ATACCTTTTAAAAAAAATCTTCACATCC		
	cg23665802py1	PS	TGGAAATTGGTTTATGTAGT	1	92002256
	cg23665802py2		TGATAGATTTAAATAGGATA	2	92002338
	cg23665802py3		GAGGAGGAAAATGTTTTG	3	92002454
MIR142(1)	cg00057966F	P	AATTAGGAAGGGTAGGAAAGTTATG		
	cg00057966R		B-CAAAAATCAAAAAACCTAAACAACC		
	cg00057966py1	PS	AGGTATTTGGGTAGGGT	1	56408501
				2	56408523
				3	56408532
				4	56408536
				5	56408539
	cg00057966py2		GAGGAAGATGGTGG	6	56408554
				7	56408561
				8	56408565
				9	56408577
				10	56408586
	cg00057966py3		GGGTGATTGTATTGTTTGT	11	56408688
				12	56408691
				13	56408694
MIR142(2)	cg10530767F	P	TAGGTAGAGATTAGGGTTTTTGGAG		
	cg10530767R		B-AAAAACCAAAACCTTAACAAAAAAAC		
	cg10530767py1	PS	ATTTTTAGTTTTGTGAATTT	1	56408999
				2	56409011
				3	56409028
MIR150	cg27388703F	P	TATTGGTATAAGGGTTGGGAGATAG		
	cg27388703R		B-ACCCCCAACATAAAATAAAATAAATATAC		
	cg27388703py1	PS	GGTTGGGAGATAGGGTTA	1	50004132
				2	50004135
				3	50004141
				4	50004152
				5	50004158
				6	50004169
	cg27388703py2		TTAGGTTTAGGGAGG	7	50004210
				8	50004238
	cg27388703py3		GGGGTAGGGATAGGTT	9	50004281
				10	50004294

Appendix D: All samples obtained from the CD24/CD44 sorting of breast cancer cell lines

Table D1: The CD24/44 sorted breast cancer cell lines. An overview over the cell lines and their CD44/CD24 expression patterns, DNA concentration and cell number are shown as well as if they were sorted and/or concentrated using a SpeedVac®.

Cell line name	CD24/CD44 Patterns	Sample ID	DNA concentration (ng/µl)	Number of cells	Sorted	Samples concentrated using SpeedVac®
MDA-MB-453	CD44-/CD24+	1	101.6	4.600.000 cells	No	No
MDA-MB-453	CD44-/CD24+	2	152.6	4.500.000 cells	No	No
MDA-MB-453	CD44-/CD24+	3	101.1	4.600.000 cells	No	No
MDA-MB-453	CD44-/CD24+	4	136.2	4.600.000 cells	No	No
MDA-MB-453	CD44-/CD24+	5	171.8	4.600.000 cells	No	No
MDA-MB-453	CD44-/CD24+	6	148.3	4.500.000 cells	No	No
MDA-MB-453	CD44-/CD24+	7	163.7	4.500.000 cells	No	No
MDA-MB-453	CD44-/CD24+	8	102.5	4.600.000 cells	No	No
MDA-MB-453	CD44-/CD24+	9	218.6	4.600.000 cells	No	No
ZR-75-1 1	CD44-/CD24+	10	385.4	4.500.000 cells	No	No
ZR-75-1 2	CD44-/CD24+	11	143.9	4.125.000 cells	No	No
ZR-75-1 3	CD44-/CD24+	12	231.8	5.125.000 cells	No	No
MDA-MB-468	CD44+/CD24+	13	381	5.000.000 cells	No	No
MDA-MB-468	CD44+/CD24+	14	316.7	5.000.000 cells	No	No
MDA-MB-468	CD44+/CD24+	15	208	5,000,000 cells	No	No
MDA-MB-468	CD44+/CD24+	16	296.7	5.000.000 cells	No	No
MDA-MB-468	CD44+/CD24+	17	217.8	5.000.000 cells	No	No
MDA-MB-468	CD44+/CD24+	18	292.1	5.000.000 cells	No	No
MDA-MB-468	CD44-/CD24+	19	19.3	14.879 cells	Yes	Yes
MDA-MB-468	CD44+/CD24+	20	93.32	1.015.489 cells	Yes	No
MDA-MB-468	CD44+/CD24+	21	83.87	1.009.876 cells	Yes	No
MDA-MB-468	CD44-/CD24+	22	13.4	16.768 cells	Yes	Yes
MDA-MB-231	CD44+/CD24-	23	91.54	789.211 cells	Yes	No
MDA-MB-231	CD44+/CD24+	24	3.75	506 cells	Yes	No
MDA-MB-231	CD44+/CD24-	25	117.7	635.000 cells	Yes	No
MDA-MB-231	CD44+/CD24-	26	105.02	680.000 cells	Yes	No
MDA-MB-231	CD44+/CD24+	27	14.98	NDA	Yes	Yes
MDA-MB-231	CD44+/CD24+	28	12.5	NDA	Yes	Yes
MDA-MB-231	CD44-/CD24-	29	37.47	NDA	Yes	Yes
MDA-MB-231	CD44-/CD24-	30	13.69	NDA	Yes	Yes
MDA-MB-231	CD44-/CD24-	31	68.97	NDA	Yes	Yes
MCF-7	CD44-/CD24+	32	44.32	1.185.000 cells	Yes	No
MCF-7	CD44-/CD24-	33	34.06	459.740 cells	Yes	No

Cell line name	CD24/CD44 Patterns	Sample ID	DNA concentration (ng/µl)	Number of cells	Sorted	Samples concentrated using SpeedVac®
MCF-7	CD44+/CD24-	34	17.43	759 cells	Yes	No
MCF-7	CD44-/CD24+	35	62.64	621.243 cells	Yes	No
MCF-7	CD44+/CD24-	36	15	1.169 cells	Yes	Yes
MCF-7	CD44+/CD24+	37	19.94	1.264 cells	Yes	Yes
MCF-7	CD44+/CD24+	38	19.76	105 cells	Yes	Yes
MCF-7	CD44-/CD24-	39	32.02	453.499 cells	Yes	No
MCF-7	CD44+/CD24-	40	81.59	1.326.171 cells	Yes	No
MCF-7	CD44-/CD24-	41	26.23	507.167 cells	Yes	No
T47D	CD44-/CD24+	42	119.7	NDA	Yes	No
T47D	CD44-/CD24-	43	32.08	189.625 cells	Yes	Yes
T47D	CD44-/CD24+	44	107.65	1.006.046 cells	Yes	No
T47D	CD44-/CD24+	45	14.95	422.767 cells	Yes	No
T47D	CD44+/CD24+	46	24.41	21.561 cells	Yes	Yes
T47D	CD44-/CD24-	47	17.994	126.116 cells	Yes	Yes
T47D	CD44+/CD24+	48	48.56	59.784 cells	Yes	Yes
T47D	CD44-/CD24-	49	23.79	108.545 cells	Yes	Yes

Appendix E: genes used as input in IPA

Table E1: The genes significantly correlated (FDR<0.005) that were imported into IPA.

Genes						
AACSL	*CA5A*	*FEZ1*	*KIF4B*	*NCRNA00171*	*RAPGEF5*	*SYTL3*
ABAT	*CABLES1*	*FEZF1*	*KIF9*	*NCRNA00203*	*RASAL3*	*TAC3*
ABCA2	*CACNA1B*	*FGF1*	*KIFC1*	*NEDD4L*	*RASEF*	*TACC2*
ABCC8	*CACNG3*	*FGGY*	*KIRREL3*	*NEDD9*	*RASSF5*	*TAF4B*
ABCD2	*CAPN2*	*FHL3*	*KLC2*	*NEK6*	*RBFOX1*	*TBC1D16*
ABHD12	*CAPN3*	*FIGN*	*KLF10*	*NEURL4*	*RBM47*	*TBX15*
ABI3	*CBL*	*FILIP1*	*KLHL23*	*NGFR*	*RBMS1*	*TCOF1*
ABLIM1	*CBX4*	*FKBP14*	*KLHL6*	*NHSL1*	*RBMX2*	*TEAD1*
ABTB2	*CC2D2A*	*FKBPL*	*KRT9*	*NKAPL*	*REEP6*	*TEAD3*
ACADSB	*CCDC19*	*FLJ22536*	*KRTAP1-5*	*NLRP1*	*REG1A*	*TEAD4*
ACER2	*CCDC62*	*FLJ40504*	*KRTAP2-4*	*NOXA1*	*REN*	*TECR*
ACOT11	*CCDC69*	*FLJ42289*	*KRTAP4-3*	*NRF1*	*RERE*	*TET1*
ACOT7	*CCL22*	*FLNC*	*KRTAP5-6*	*NRXN1*	*RFPL2*	*TFAP2A*
ACOT8	*CCR4*	*FLT4*	*LAPTM5*	*NSUN7*	*RGS12*	*TFAP2B*
ACOX3	*CCT6A*	*FMNL1*	*LETM1*	*NTRK1*	*RGS7BP*	*TGFBR1*
ACSL4	*CD209*	*FMNL3*	*LIF*	*NUP54*	*RHBDF2*	*TGFBR2*
ADAM19	*CD226*	*FOLR1*	*LILRB2*	*NXN*	*RIMBP2*	*TGFBR3*
ADAMTS13	*CD44*	*FOSL1*	*LIMK1*	*OBSCN*	*RNASE1*	*THOC5*
ADAMTS2	*CD82*	*FOXN4*	*LMO2*	*ODF4*	*RNASE6*	*THSD7B*
ADAP2	*CDC42EP3*	*FOXRED1*	*LMX1B*	*OPCML*	*RNASE7*	*TIMP2*
ADARB2	*CDH2*	*FRAS1*	*LOC100129066*	*OR2S2*	*RNASE8*	*TKT*
ADCY9	*CDH4*	*FRMD1*	*LOC100188947*	*OR4C16*	*RNF216*	*TMC7*
ADD3	*CDR2*	*FRMD4B*	*LOC100192378*	*OR7D4*	*RNF220*	*TMEM132D*
AEN	*CDX4*	*FXYD1*	*LOC100294362*	*OR7E91P*	*RPH3AL*	*TMEM181*
AFAP1L2	*CEACAM6*	*GABRA4*	*LOC154449*	*P2RX1*	*RPL23AP82*	*TMEM211*
AGAP1	*CECR7*	*GAD2*	*LOC157627*	*P2RY14*	*RPL23P8*	*TMEM217*
AHDC1	*CELA2A*	*GAK*	*LOC158381*	*PACRG*	*RPS6KA1*	*TMEM8B*
AIF1	*CGB1*	*GALNTL1*	*LOC284100*	*PACSIN1*	*RPS6KA4*	*TMPRSS4*
AIF1L	*CHD9*	*GARNL3*	*LOC285419*	*PADI4*	*RPS6KL1*	*TNFAIP8L1*
AKAP12	*CHID1*	*GATS*	*LOC285780*	*PADI6*	*RPTOR*	*TNFRSF10A*
AKAP13	*CHKB-CPT1B*	*GFOD2*	*LOC286359*	*PAFAH1B3*	*RRP12*	*TNFRSF8*
AKAP7	*CHMP1A*	*GGN*	*LOC339240*	*PAGE2*	*RYR2*	*TNS1*
ALDOB	*CHRNB1*	*GGT1*	*LOC388796*	*PANX1*	*SAFB*	*TNXB*
ALPL	*CHST3*	*GH1*	*LOC399753*	*PAQR8*	*SAMD3*	*TPTE2*
ALS2CR11	*CHST9*	*GJC3*	*LOC440040*	*PARD3*	*SBF2*	*TRAM2*
AMICA1	*CIB2*	*GLRA4*	*LOC440335*	*PARP15*	*SC5DL*	*TRIM61*
AMT	*CILP*	*GMFG*	*LOC441294*	*PARP16*	*SCHIP1*	*TRPM1*
ANGEL2	*CLDN15*	*GNAQ*	*LOC645166*	*PARVA*	*SCML1*	*TRPV2*
ANKMY1	*CLDN4*	*GNG12*	*LOC645323*	*PAX7*	*SCML4*	*TSHZ2*
ANKRD11	*CLNK*	*GNGT2*	*LOC646214*	*PBX2*	*SCO2*	*TSPAN14*
ANKRD26P1	*CMTM1*	*GPR123*	*LOC728264*	*PCDHGA1*	*SCUBE1*	*TSSC1*
ANKRD27	*CMTM5*	*GPR132*	*LOC728927*	*PCDHGA10*	*SDPR*	*TTC12*
ANKRD9	*CNTN4*	*GPRIN1*	*LOC729176*	*PCDHGA11*	*SEC1*	*TTLL10*
ANKS1B	*COL1A1*	*GPX6*	*LOC730755*	*PCDHGA12*	*SEC31B*	*TUBGCP6*
AP2A2	*COL7A1*	*GRAP2*	*LOC80154*	*PCDHGA2*	*SEMA5B*	*TULP1*
ARGLU1	*CPEB1*	*GRB10*	*LOC96610*	*PCDHGA3*	*SEMA6A*	*TYMP*
ARHGAP10	*CPEB3*	*GRHL2*	*LONP2*	*PCDHGA4*	*SEMA7A*	*UBAC2*

Genes						
ARHGAP25	*CPLX4*	*GRHL3*	*LOXL2*	*PCDHGA5*	*SERINC5*	*UBASH3A*
ARHGAP31	*CPNE5*	*GRIK2*	*LPAL2*	*PCDHGA6*	*SERPINB5*	*UBE2D2*
ARHGAP36	*CPT1B*	*GRIN1*	*LPAR1*	*PCDHGA7*	*SFRS8*	*UBTD1*
ARHGAP39	*CREG2*	*GRIN2B*	*LRRC15*	*PCDHGA8*	*SGCD*	*UBXN10*
ARHGAP8	*CRYBA1*	*GRK5*	*LRRC27*	*PCDHGA9*	*SH2B3*	*UCN2*
ARHGEF15	*CRYBB2*	*GRK6*	*LRRC33*	*PCDHGB1*	*SH3RF3*	*UCN3*
ARHGEF4	*CSF1R*	*GRM1*	*LRRC37A3*	*PCDHGB2*	*SH3TC2*	*UNC119B*
ARHGEF40	*CSNK2A2*	*GRM4*	*LRRC4B*	*PCDHGB3*	*SHANK2*	*UNC5B*
ARID1A	*CTAGE6*	*GRN*	*LRRFIP1*	*PCDHGB4*	*SHC4*	*UPP2*
ARVCF	*CTBP2*	*GRP*	*LSM2*	*PCDHGB5*	*SHISA5*	*USP7*
ASAP2	*CTNNB1*	*GSN*	*LST1*	*PCDHGB6*	*SIGLEC15*	*VAMP5*
ASB1	*CTTN*	*GSPT1*	*LY6G6C*	*PCDHGB7*	*SIM1*	*VCAM1*
ATF6B	*CUX2*	*GYPC*	*LY86*	*PCDHGC3*	*SIPA1*	*VN1R2*
ATP11A	*CXCR2*	*H19*	*LYPLA2*	*PCDHGC4*	*SKI*	*VPS41*
ATP2B2	*CXorf26*	*HCG22*	*MACROD1*	*PCDHGC5*	*SLAMF8*	*WDFY4*
ATXN1	*CYB561D2*	*HCRTR2*	*MAD1L1*	*PCDP1*	*SLC13A4*	*WDR51B*
ATXN1L	*CYP2D7P1*	*HCST*	*MAML2*	*PCNT*	*SLC16A12*	*WDR93*
B3GNT7	*CYP4A22*	*HDAC9*	*MAP1LC3B*	*PCSK6*	*SLC16A3*	*WNK4*
B4GALNT1	*DAB2IP*	*HECA*	*MAP3K5*	*PCTP*	*SLC1A7*	*WNT3*
BAI1	*DAP3*	*HECTD2*	*MAP3K8*	*PDE8B*	*SLC22A11*	*WWP2*
BAT2	*DCHS1*	*HECW2*	*MBP*	*PDZRN3*	*SLC22A3*	*WWTR1*
BAT5	*DDAH1*	*HFM1*	*MBTPS1*	*PEMT*	*SLC29A2*	*XAB2*
BBS5	*DDR1*	*HIP1R*	*MCOLN3*	*PEX11A*	*SLC2A5*	*XAF1*
BCAR3	*DDX58*	*HIVEP3*	*MCPH1*	*PEX11G*	*SLC38A10*	*ZBTB12*
BDNF	*DENND1C*	*HLA-DMA*	*MDFI*	*PHF1*	*SLC38A5*	*ZC3H12D*
BMP2	*DGKA*	*HOXC4*	*MED12L*	*PHF15*	*SLC44A3*	*ZC3H6*
BMPR2	*DHFR*	*HOXC5*	*MEF2D*	*PHOSPHO2*	*SLC6A1*	*ZCCHC24*
BPESC1	*DHODH*	*HOXC6*	*MEIS1*	*PIKFYVE*	*SLC6A4*	*ZEB2*
BPIL2	*DHRS3*	*HPD*	*MEIS3P1*	*PIP4K2A*	*SLC7A1*	*ZMIZ1*
BRD7	*DIRC3*	*HPSE2*	*MGC14436*	*PIP5K1C*	*SLC9A3R2*	*ZNF254*
BRF1	*DLG2*	*HRK*	*MGC45800*	*PIRT*	*SLCO1B1*	*ZNF257*
BRSK2	*DLG4*	*IFFO1*	MIR130A	*PKHD1*	*SLCO3A1*	*ZNF28*
BSDC1	*DNA2*	*IGDCC4*	MIR141	*PKMYT1*	*SLCO5A1*	*ZNF430*
BTBD19	*DNAH1*	*IKZF4*	MIR144	*PKP3*	*SMAD6*	*ZNF445*
BTBD7	*DNAH17*	*IL17D*	MIR145	*PLA2G4B*	*SMCR7L*	*ZNF506*
BUD31	*DNAH5*	*IL17RD*	MIR195	*PLAG1*	*SMCR8*	*ZNF559*
BZRAP1	*DNAJB6*	*IL27*	MIR1973	*PLCG2*	*SMO*	*ZNF619*
C10orf122	*DNAJC8*	*IL29*	MIR200C	*PLEKHB1*	*SNAR-G1*	*ZNF69*
C10orf79	*DOCK1*	*INADL*	MIR219-2	*PLEKHH1*	*SNORA15*	*ZNF697*
C11orf41	*DOCK7*	*INHBC*	MIR2276	*PLIN4*	*SNORA38*	*ZNF710*
C12orf34	*DPEP1*	*INSRR*	MIR451	*PLXND1*	*SNORA71B*	*ZNF791*
C12orf56	*DPYSL2*	*INTS1*	MIR497	*POLR1A*	*SNORD17*	*ZNF826*
C12orf76	*DUSP7*	*INTS4*	MIR609	*POLR2A*	*SNORD1A*	*ZNF90*
C14orf43	*EBF3*	*IPO5*	MIR675	*POMT2*	*SNORD1B*	*ZSCAN1*
C15orf32	*EDA*	*IRAK2*	*MLXIP*	*POR*	*SNX21*	*ZSCAN12*
C16orf57	*EHD1*	*IRS1*	*MOB2/YY1AP1*	*PPAP2B*	*SNX29*	*ZSCAN5A*
C17orf68	*EHMT1*	*ITFG3*	*MORN1*	*PPARGC1B*	*SNX5*	*ZYX*
C19orf50	*EI24*	*ITGB2*	*MPL*	*PPFIA4*	*SOLH*	
C1orf172	*EIF3D*	*ITGB4*	*MRGPRF*	*PPFIBP2*	*SORL1*	
C1orf54	*EIF5A*	*ITGB5*	*MTUS1*	*PPP2R5C*	*SOX2OT*	
C1S	*EPHA2*	*ITGB7*	*MUC2*	*PPP5C*	*SPATA13*	

Genes						
C20orf117	*ESR1*	*ITIH5*	*MUC21*	*PRAM1*	*SPG7*	
C20orf202	*ESRRG*	*ITPK1*	*MUSTN1*	*PRB4*	*SPINT2*	
C20orf72	*EXD3*	*ITPR3*	*MYBPC1*	*PRDM11*	*SPNS3*	
C22orf15	*EXOSC10*	*IZUMO1*	*MYH10*	*PRDM16*	*SPOCK3*	
C22orf34	*EYA4*	*JMJD7*	*MYH13*	*PRKAG3*	*SPRY4*	
C2orf54	*F10*	*JPH4*	*MYH2*	*PRKCZ*	*SPSB4*	
C2orf58	*F7*	*KAZN*	*MYH3*	*PROKR2*	*SRPR*	
C2orf72	*FABP6*	*KCNAB2*	*MYH8*	*PRR19*	*SSH1*	
C2orf77	*FAM171A2*	*KCNC1*	*MYO10*	*PRR5*	*ST3GAL1*	
C3AR1	*FAM175B*	*KCNG2*	*MYO15A*	*PRRX2*	*ST5*	
C3orf67	*FAM180B*	*KCNIP4*	*MYO3B*	*PSMD7*	*STAT3*	
C5orf49	*FAM71F1*	*KCNMA1*	*MYOCD*	*PTGER4*	*STIM1*	
C6orf103	*FAM78A*	*KCNT1*	*NAALADL1*	*PTPRE*	*STK24*	
C6orf106	*FAM83H*	*KDM4B*	*NBPF4*	*PTPRN2*	*STK4*	
C6orf145	*FAM86B1*	*KIAA0922*	*NBPF7*	*PUSL1*	*STRA8*	
C6orf25	*FAS*	*KIAA0947*	*NCF4*	*PVRIG*	*STX18*	
C6orf27	*FBXL21*	*KIAA1161*	*NCKAP5L*	*PVRL1*	*SUFU*	
C7orf50	*FBXO40*	*KIAA1949*	*NCR3*	*PXN*	*SVIL*	
C8orf34	*FBXW11*	*KIF12*	*NCRNA00099*	*PYGB*	*SYNE2*	
CA12	*FER*	*KIF25*	*NCRNA00114*	*RADIL*	*SYNPO*	

Appendix F: the DNA methylation values of the CpGs in the genes for the regions within the tumors

The DNA methylation levels are shown for the CpGs within miRNAs as the regions of the tumors were pyrosequenced once. While for the protein-coding genes the mean of methylation level and standard deviation are shown for the CpGs within the protein-coding genes as the regions of the tumors were pyrosequenced in triplicates.

Table F1: DNA Methylation level of CpG 1-6 in MIR199A for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M%: methylation percentage.

MIR199A											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	71.93	79.43	79.93	71.68	71.53	76.87	73.05	75.11	24.52	30.66	38.59
2	63.00	72.16	77.45	63.43	61.82	65.38	61.99	66.87	27.82	29.78	34.85
3	67.64	82.89	83.53	73.86	75.10	75.92	69.90	74.43	26.98	28.48	33.90
4	65.72	77.42	79.36	68.93	62.66	72.22	66.79	70.40	26.43	28.56	34.05
5	66.10	76.05	79.27	67.26	70.41	69.69	65.33	70.26	26.72	27.49	32.68
6	50.02	64.59	65.41	54.26	55.66	59.06	65.05	66.70	24.20	17.51	33.04

Table F2: DNA Methylation level of CpG 1-5 in MIR135B for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage. NA: samples that were not accepted after the pyrosequencing run.

MIR135B											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%
1	57.35	29.64	15.60	90.44	93.51	NA	75.11	68.33	81.11	91.81	82.90
2	60.37	28.78	16.93	100.00	94.44	NA	90.81	91.27	82.12	99.43	82.54
3	57.79	29.53	18.49	88.11	86.79	90.02	88.82	88.87	83.56	95.54	81.34
4	59.30	30.68	20.53	95.01	95.83	90.80	87.84	87.03	97.23	99.83	86.40
5	49.39	24.68	14.71	72.25	69.93	65.42	66.93	68.01	68.10	75.17	66.37

Table F3: DNA Methylation level of CpG 1-3 in MIR16-2 for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage.

MIR16-2											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	62.15	54.51	40.32	61.34	49.35	44.92	36.11	24.32	82.49	66.51	78.65
2	52.98	51.28	40.83	55.07	55.50	43.28	46.48	28.13	65.24	65.53	65.71
3	43.71	41.34	36.37	42.07	45.79	29.73	27.85	21.77	43.32	40.85	38.87

Table F4: DNA Methylation level of CpG 1-9 in MIR887 for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M%: methylation percentage.

MIR887											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	3.24	4.07	1.55	4.64	2.10	1.13	0.00	2.08	0.49	1.28	1.11
2	10.55	8.38	3.61	14.06	5.22	3.74	1.68	2.98	2.73	3.45	4.82
3	14.87	12.75	5.33	20.04	10.14	4.72	2.49	4.18	3.07	5.72	5.91
4	12.24	9.55	4.40	14.35	7.21	3.39	2.02	2.68	2.67	4.17	4.77
5	6.38	5.21	2.18	7.95	3.04	1.77	1.16	1.40	1.28	1.50	3.33
6	8.35	6.60	2.66	10.40	5.43	2.17	1.54	1.95	2.09	1.89	3.80
7	6.43	5.62	2.25	7.47	3.85	1.91	1.33	1.64	1.16	2.22	2.67
8	7.40	4.76	2.43	8.43	3.84	2.14	1.48	2.57	1.16	1.84	2.98
9	11.12	7.03	4.68	11.11	5.25	3.24	0.00	3.50	2.33	2.76	5.67

Table F5: DNA Methylation level of CpG 2-8 in MIR148A for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage.

MIR148A											
CpG	137			155		142			RP		
	137B	137D	137C	155A	155C	142C	142B	142A	RP2	RP10	RP11
	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%
2	20.02	15.72	8.56	25.31	8.81	16.80	25.33	19.06	4.24	9.68	8.15
3	9.48	15.39	6.07	18.47	7.18	17.41	25.53	21.62	4.22	7.52	4.69
4	14.60	13.30	4.53	19.61	8.93	19.58	24.33	14.79	4.19	8.88	5.90
5	17.24	15.29	9.54	27.97	11.28	36.04	20.61	43.18	23.97	6.65	6.33
6	15.87	15.88	11.10	20.62	11.73	37.41	20.55	35.56	28.86	8.53	8.52
7	8.61	9.89	5.18	14.19	5.81	29.61	13.40	28.19	25.24	2.71	3.36
8	15.59	16.28	9.37	24.41	11.65	31.70	20.58	32.03	23.17	7.44	8.26

Table F6: DNA Methylation level of CpG 1-14 in MIR200C/141 cluster for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage.

MIR200C/141 cluster											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	41.70	27.91	14.30	65.33	39.14	26.93	19.60	16.36	56.94	65.45	47.60
2	49.76	35.56	20.71	65.13	45.85	33.49	20.98	20.85	70.39	77.10	54.26
3	41.22	28.64	17.05	60.15	38.91	31.62	19.26	18.37	55.79	69.68	45.82
4	47.29	31.59	18.54	65.62	46.78	32.95	21.35	21.26	65.48	77.35	51.73
5	48.61	31.70	18.97	62.73	44.72	33.91	20.61	23.19	65.50	80.99	50.68
6	40.20	26.07	16.91	58.90	39.38	17.01	23.89	14.98	45.57	59.76	41.81
7	41.48	30.71	22.69	55.53	38.15	17.38	23.85	16.83	48.91	63.56	47.31
8	36.57	29.62	15.34	54.27	34.79	13.22	18.57	12.74	36.27	49.67	36.83
9	35.40	27.26	14.99	53.38	34.17	12.82	17.60	13.72	34.15	54.84	32.37
10	38.75	28.14	15.18	54.44	36.10	14.48	20.79	13.69	39.47	49.74	35.66
11	39.17	28.00	12.55	55.58	37.84	17.09	20.49	15.93	46.08	62.30	41.90
12	31.78	20.83	12.93	44.61	29.83	19.14	10.65	15.25	53.46	47.10	31.76
13	31.77	24.66	13.86	54.69	36.28	27.97	15.95	17.28	40.01	61.23	34.56
14	28.75	20.15	10.77	46.75	31.82	24.72	13.10	14.93	34.63	44.66	34.12

Table F7: DNA Methylation level of CpG 1-3 in MIR17-92 cluster for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage. NA: samples that were not accepted after the pyrosequencing run.

MIR17-92 cluster											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	55.64	49.08	34.95	65.52	48.22	51.77	52.59	52.26	75.13	75.64	75.84
2	37.97	NA	28.22	46.21	37.55	38.35	41.95	NA	66.76	53.80	70.00
3	25.68	24.33	15.65	34.44	20.01	33.62	35.25	36.37	41.93	36.84	46.89

Table F8: DNA Methylation level of CpG 1-13 in MIR 142 (1) for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M %: methylation percentage. NA: samples that were not accepted after the pyrosequencing run.

MIR142(1)											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%	M%
1	31.59	37.12	58.24	25.38	38.24	78.62	55.67	72.18	72.96	62.64	66.87
2	19.99	25.69	46.36	19.38	31.41	48.43	33.51	48.87	52.82	46.87	50.47
3	25.34	32.55	52.22	23.86	37.79	66.20	45.35	63.90	67.52	53.82	77.35
4	23.97	29.03	46.60	17.49	32.18	62.94	42.83	60.65	64.54	52.72	55.22
5	22.74	30.61	51.67	17.22	34.18	62.05	43.27	61.81	64.00	53.39	57.16
6	21.97	32.04	47.29	19.89	32.65	67.07	55.85	64.08	67.10	56.86	66.27
7	25.30	36.36	53.82	21.67	37.72	72.77	59.30	70.12	69.99	60.44	71.74
8	28.47	37.20	52.61	19.33	38.71	73.02	58.12	71.02	73.35	58.02	67.53
9	24.04	35.81	51.25	18.17	37.10	71.18	55.45	68.65	68.50	56.99	67.27
10	30.39	36.11	57.59	22.30	41.78	71.11	59.22	69.70	83.12	67.39	64.26
11	24.88	34.29	52.46	18.21	27.07	72.38	50.51	65.43	61.17	NA	62.51
12	22.59	30.11	49.93	15.17	22.27	71.25	53.76	64.02	59.85	NA	52.64
13	22.53	28.75	44.91	15.32	18.21	63.45	43.41	61.67	55.94	NA	53.60

Table F9: DNA Methylation level of CpG 1-3 in MIR142 (2) for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M%: methylation percentage.

MIR142(2)											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	43.34	48.02	62.31	31.60	50.25	63.97	83.88	76.95	81.85	66.00	81.47
2	45.78	49.86	67.51	31.48	53.04	66.13	82.73	77.83	80.64	69.00	82.37
3	38.35	41.57	57.64	26.10	48.68	62.22	78.42	71.96	71.58	63.51	73.11

Table F10: DNA Methylation level of CpG 1-10 in MIR150 for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. M%: methylation percentage.

MIR150											
CpG	137			155		142			RP		
	B	D	C	A	C	C	B	A	2	10	11
	M %	M%	M %	M%	M%	M%	M%	M%	M%	M%	M%
1	15.95	32.08	36.56	11.90	37.54	42.48	49.23	45.54	27.28	19.43	36.76
2	8.75	27.71	31.18	7.19	28.25	42.57	56.91	53.88	12.76	8.68	14.69
3	14.36	27.64	44.04	8.35	30.05	38.95	44.51	46.24	20.66	15.11	20.73
4	6.51	13.34	17.72	5.35	16.26	24.17	30.29	40.83	16.89	8.01	14.97
5	11.19	23.83	27.27	7.56	28.47	36.73	41.15	52.47	20.39	15.59	19.96
6	9.06	9.84	21.41	5.94	11.87	18.98	22.19	21.75	17.05	6.49	15.51
7	16.96	34.37	42.52	9.75	31.62	42.40	53.20	57.69	28.60	19.48	31.01
8	13.22	22.27	31.37	6.98	29.74	44.20	48.99	57.67	20.04	7.80	19.65
9	27.47	41.93	61.92	18.81	42.89	62.12	71.18	73.84	46.45	49.45	53.63
10	15.53	28.54	42.01	8.01	28.22	49.47	55.68	53.78	22.03	20.10	18.77

Table F11: Methylation level of CpG 1-15 in *BCAN* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *BCAN*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	4.85	1.70	3.31	1.84	1.39	1.96	8.49	1.96	27.89	1.29	38.98	1.53	46.56	1.04	51.01	2.41	NA	NA	3.10	0.23	NA	NA
2	2.22	2.33	1.36	1.17	0.82	1.15	7.55	0.66	24.48	1.87	45.45	1.81	42.67	0.78	43.58	2.14	NA	NA	1.96	0.47	NA	NA
3	4.39	1.50	2.38	1.03	2.53	1.24	10.35	1.83	26.81	0.85	41.04	1.41	51.04	2.94	51.07	2.05	NA	NA	1.40	1.98	NA	NA
4	4.75	1.97	3.33	1.33	3.96	1.44	10.58	2.57	20.91	2.88	44.15	2.32	44.97	0.31	44.57	2.45	NA	NA	2.53	1.44	NA	NA
5	3.76	1.17	2.42	1.69	1.14	1.61	5.70	2.70	15.78	0.75	43.64	0.69	49.27	0.42	50.16	4.27	NA	NA	2.80	2.15	NA	NA
6	0.00	0.00	0.00	0.00	0.00	0.00	2.50	2.19	4.79	1.92	40.85	1.70	36.94	1.61	40.64	0.60	NA	NA	0.85	1.20	NA	NA
7	12.55	3.35	8.16	4.95	5.14	3.24	13.52	1.00	37.25	1.61	49.18	0.89	56.39	0.64	62.02	0.19	11.18	0.47	8.20	0.66	NA	NA
8	8.97	0.54	6.50	1.68	1.75	2.47	12.09	1.13	28.19	0.87	47.84	0.74	60.99	0.39	57.30	3.00	12.48	0.16	9.96	1.53	NA	NA
9	2.17	3.07	6.02	1.68	2.35	3.32	8.76	0.40	19.65	0.25	53.58	4.83	53.68	0.72	59.92	3.03	6.47	2.99	4.37	1.27	NA	NA
10	0.00	0.00	1.70	2.40	1.54	2.18	8.42	0.71	25.07	4.01	33.44	0.08	35.10	0.76	30.87	0.45	5.70	2.01	3.66	1.94	NA	NA
11	2.22	3.14	1.26	1.77	1.14	1.61	7.82	1.62	18.59	1.68	34.26	4.77	43.16	3.37	47.67	0.07	4.34	3.41	4.43	2.58	NA	NA
12	0.00	0.00	8.15	4.04	1.45	2.04	13.15	2.23	33.02	4.34	43.41	4.09	51.23	0.25	54.34	0.13	6.26	4.48	6.60	2.31	NA	NA
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	40.54	4.06	45.37	0.70	36.60	1.10	2.53	3.58	4.99	1.76	NA	NA
14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	39.20	2.30	48.26	0.21	47.33	1.11	3.00	4.24	6.47	3.15	NA	NA
15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	41.91	3.44	46.00	1.92	40.74	0.11	2.60	3.68	7.24	3.94	NA	NA

Table F12: Methylation level of CpG 1-5 in *BNIPL* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *BNIPL*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	55.85	2.96	38.42	2.05	33.39	2.08	57.97	3.63	35.58	2.53	22.65	1.25	36.88	1.14	24.23	1.15	NA	NA	NA	NA	47.00	2.10
2	92.51	3.45	76.31	3.85	71.99	1.50	91.65	4.25	57.69	2.35	32.29	3.17	57.93	2.40	35.57	1.40	NA	NA	NA	NA	74.43	1.30
3	98.55	1.96	79.97	3.41	77.09	1.27	94.85	3.06	59.19	3.99	34.72	2.22	57.88	0.74	37.57	1.18	NA	NA	NA	NA	87.54	1.32
4	84.13	2.70	71.03	1.75	57.74	0.21	71.15	3.53	49.54	0.08	42.44	0.98	24.63	0.79	28.02	0.43	NA	NA	85.04	4.43	74.32	2.32
5	72.46	1.32	48.31	0.65	34.20	0.93	61.79	4.57	41.70	3.12	36.25	1.47	21.30	2.46	26.93	4.16	NA	NA	NA	6.44	49.99	0.82

Table F13: Methylation level of CpG 8-21 in *CTSA* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *CTSA*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
8	5.75	1.58	5.39	1.05	2.69	0.64	6.89	1.65	4.48	2.34	35.98	3.25	27.55	1.86	37.29	1.70	NA	NA	3.32	0.45	3.50	1.90
9	5.82	2.14	5.66	0.69	3.55	0.32	8.82	0.32	4.32	0.98	35.72	2.59	27.75	0.93	37.45	0.32	NA	NA	3.12	0.57	3.33	0.51
10	7.51	4.46	6.04	1.14	3.09	0.34	9.30	2.05	4.91	1.56	30.66	2.75	24.79	1.39	32.63	3.18	NA	NA	1.05	1.48	3.35	0.64
11	8.81	1.71	5.71	1.77	5.17	0.31	10.75	1.30	6.55	0.01	36.24	1.63	27.17	0.82	36.33	3.05	NA	NA	1.94	2.74	4.22	0.26
12	6.77	1.70	3.75	0.20	4.42	1.03	8.41	1.71	5.77	1.38	35.23	2.05	23.10	1.93	35.22	3.37	NA	NA	5.41	1.32	4.60	0.41
13	8.46	1.06	7.17	0.59	3.47	0.93	8.31	0.49	5.53	0.66	29.35	3.02	21.55	0.60	32.61	2.40	NA	NA	4.06	1.39	4.99	0.49
14	8.79	1.63	3.16	0.68	3.46	0.66	3.62	3.15	3.90	1.01	33.54	2.19	23.93	0.82	39.81	2.10	NA	NA	0.00	0.00	4.04	1.84
15	9.95	1.19	7.99	1.40	4.64	1.88	8.45	1.07	5.36	0.11	32.97	3.27	23.90	1.54	37.08	2.84	NA	NA	2.98	4.21	1.95	2.76
16	8.88	0.55	4.77	0.18	2.00	2.82	6.76	0.27	5.26	2.06	32.26	1.93	24.91	0.15	39.89	3.05	NA	NA	5.89	2.28	NA	NA
17	7.25	3.39	3.42	0.38	4.50	2.60	5.58	0.59	0.00	0.00	NA	6.58	24.69	1.09	38.82	3.57	0.90	1.27	1.13	1.60	3.16	0.71
18	6.83	4.52	2.92	0.96	0.98	1.38	4.95	0.56	0.00	0.00	34.85	3.69	24.37	0.57	41.03	2.67	4.03	3.83	0.95	1.34	3.23	0.71
19	7.34	4.50	4.20	0.65	5.68	3.45	5.89	0.68	11.55	2.26	41.72	2.87	28.45	0.58	41.48	3.37	5.10	3.22	4.11	0.45	4.14	1.55
20	8.63	4.51	2.90	0.38	1.22	1.73	4.66	0.32	0.00	0.00	35.03	4.94	24.26	1.12	38.30	1.73	4.88	3.21	2.51	0.55	3.63	0.80
21	6.28	3.44	2.43	0.91	1.04	1.46	3.42	0.91	2.69	3.80	NA	5.40	22.02	0.69	34.35	2.58	3.48	2.72	3.29	0.77	0.97	1.69

Table F14: Methylation level of CpG 1 and 4-6 in *FGFBP2* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *FGFBP2*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	42.43	1.28	31.96	0.73	21.02	0.90	45.42	2.69	28.32	2.76	12.31	0.66	20.72	1.44	17.01	1.29	NA	NA	NA	NA	NA	NA
4	37.57	1.33	27.04	0.43	20.30	0.16	44.50	0.86	26.45	1.47	30.74	1.16	18.80	0.74	19.08	0.31	NA	NA	NA	NA	NA	NA
5	40.82	0.60	32.76	0.24	22.41	1.14	47.27	1.12	17.05	1.62	31.51	3.00	21.65	0.54	19.41	0.49	NA	NA	NA	NA	NA	NA
6	6.34	1.40	6.48	0.72	5.22	0.86	7.61	0.83	6.15	0.59	7.35	1.61	9.18	0.72	6.40	1.63	NA	NA	NA	NA	NA	NA

Table F15: Methylation level of CpG 2-20 in *FOXC1* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *FOXC1*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
2	3.55	2.55	2.90	0.05	2.42	1.51	3.97	0.25	2.17	0.25	7.54	0.89	7.77	2.00	16.92	0.74	NA	NA	3.07	1.63	4.44	0.74
3	0.76	1.07	1.49	0.08	1.73	0.85	2.40	0.18	1.38	0.12	3.33	0.62	5.20	1.63	4.85	0.10	NA	NA	2.06	0.86	1.78	0.37
4	3.68	1.30	3.37	0.71	2.95	0.92	3.79	0.50	2.01	0.74	7.62	1.33	9.10	1.19	8.04	1.63	NA	NA	4.05	1.53	4.58	1.72
5	3.72	0.38	2.91	0.33	2.71	0.68	3.39	0.36	2.89	0.74	6.55	0.84	7.19	0.52	6.35	0.56	NA	NA	4.35	1.21	2.91	0.70
6	1.13	0.20	1.12	0.31	0.73	1.05	1.49	0.12	1.09	0.26	1.74	0.35	2.95	0.61	1.53	0.56	NA	NA	1.16	1.37	1.23	0.99
7	2.05	0.75	1.50	0.41	1.09	0.85	2.70	0.34	1.53	0.29	2.37	0.30	6.05	1.19	2.43	0.24	NA	NA	2.62	1.07	1.50	1.08
8	4.28	0.98	2.95	0.53	0.83	1.09	5.03	0.76	2.61	0.90	8.51	0.56	10.29	1.97	5.56	1.26	NA	NA	2.79	1.92	3.94	1.24
9	5.63	1.43	4.67	0.54	3.18	1.53	6.11	1.26	3.67	1.02	9.03	0.94	11.63	2.56	7.98	1.02	NA	NA	3.88	2.64	4.03	0.62
10	5.73	1.70	4.70	0.92	4.55	1.82	6.29	0.81	3.84	0.84	9.28	1.46	12.98	1.60	8.21	2.38	NA	NA	6.74	1.87	6.15	1.34
11	NA	NA	4.24	1.74	1.60	1.39	4.18	0.77	3.29	1.05	7.63	0.93	12.42	1.45	5.63	1.61	NA	NA	6.02	1.23	5.08	1.58
12	3.17	0.22	2.85	0.89	2.21	0.76	3.33	0.86	2.24	0.10	8.73	1.82	7.32	1.66	4.35	0.67	4.54	1.41	5.80	1.36	5.27	2.84
13	3.11	0.13	2.38	0.09	2.21	0.78	3.78	0.76	2.70	0.88	6.73	1.22	7.81	1.85	4.71	1.40	4.32	0.37	5.47	1.95	4.15	1.74
14	4.54	0.60	4.59	0.51	2.84	0.28	6.20	0.91	4.57	1.51	11.16	0.61	13.89	2.24	7.24	1.36	5.10	2.58	5.69	2.06	3.38	1.20
15	2.84	1.19	2.24	0.80	0.86	0.81	3.05	0.32	1.96	0.62	5.48	1.01	8.97	1.14	6.73	0.74	4.29	0.43	5.16	1.82	3.32	1.42
16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17	4.63	3.29	1.97	0.39	1.24	1.08	3.62	1.13	2.61	0.81	4.75	1.41	6.07	0.66	7.76	0.61	3.88	0.01	0.94	1.63	2.17	2.02
18	2.09	0.78	1.98	0.76	0.93	0.36	2.19	0.69	1.91	0.60	2.56	0.21	4.41	0.45	2.99	0.41	3.37	2.09	0.00	0.00	2.40	1.06
19	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
20	2.91	0.45	1.98	0.71	1.76	0.75	3.85	0.87	2.65	0.78	4.03	0.38	5.82	1.11	3.60	0.49	1.14	1.61	2.84	2.66	2.75	1.25

Table F16: Methylation level of CpG 1-2 in *GYPE* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *GYPE*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	69.30	0.35	72.43	3.34	73.97	1.52	52.90	1.41	43.39	0.43	53.36	0.38	55.09	1.36	50.14	1.46	NA	NA	NA	NA	60.14	2.12
2	70.83	0.21	71.39	0.72	86.59	0.39	59.96	3.79	58.65	0.45	62.55	1.85	64.15	0.80	61.47	0.28	NA	NA	NA	NA	61.53	1.14

Table F17: Methylation level of CpG 1-7 in *IL1A* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *IL1A*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	NA	NA	29.41	2.01	14.29	1.74	33.66	1.40	15.01	1.05	44.95	3.69	56.14	0.37	38.52	0.20	25.3	1.7	NA	NA	23.8	2.6
2	NA	NA	25.90	1.55	12.79	1.71	26.39	1.72	10.13	0.62	40.82	0.66	54.01	0.24	48.65	1.14	24.0	4.1	NA	NA	17.2	2.0
3	37.73	2.28	32.53	0.93	18.59	1.99	36.19	1.81	27.13	1.49	44.93	0.23	57.06	0.37	49.33	1.25	31.0	2.0	NA	NA	29.1	2.8
4	33.58	2.88	23.65	1.35	12.35	1.70	29.46	2.73	18.02	1.00	46.55	0.49	55.58	0.16	47.19	2.31	28.5	2.5	NA	NA	21.6	3.7
5	32.44	3.90	34.37	2.16	19.68	2.67	32.39	1.55	20.05	0.78	44.90	0.71	59.20	0.08	58.68	1.07	29.2	1.4	30.5	2.6	25.2	2.1
6	38.71	2.08	33.62	3.38	18.80	4.93	37.54	1.50	20.02	1.88	48.55	2.97	64.92	0.86	59.89	2.34	26.4	1.3	31.9	2.2	23.8	0.8
7	19.50	2.95	19.26	1.70	8.64	1.88	18.24	1.77	13.62	1.32	11.58	0.23	11.17	0.67	17.68	1.07	NA	NA	24.3	3.3	16.2	0.5

Table F18: Methylation level of CpG 1-6 in *IL1R2* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *IL1R2*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	43.29	0.62	38.02	0.27	29.56	0.37	44.79	1.82	63.84	1.92	22.42	1.40	25.10	0.82	26.12	1.99	NA	NA	63.67	3.71	66.92	2.96
2	38.15	0.58	29.74	0.18	19.39	1.10	42.42	2.11	51.38	2.21	22.56	1.93	22.47	1.23	26.10	1.93	NA	NA	61.55	3.05	67.17	3.06
3	27.76	0.03	21.05	1.99	12.78	2.66	33.19	0.97	41.93	3.86	17.75	2.36	21.12	1.10	27.76	1.09	NA	NA	45.13	1.84	45.23	4.51
4	22.53	1.72	16.07	1.30	10.13	1.17	30.10	0.38	43.69	1.37	25.02	2.90	26.83	0.35	35.28	1.03	NA	NA	46.78	0.81	47.81	1.53
5	34.70	4.35	24.83	0.08	23.67	3.21	47.44	1.71	55.54	0.64	37.19	1.50	38.04	1.69	33.32	1.57	NA	NA	NA	NA	NA	NA
6	35.21	0.76	21.66	2.64	17.75	0.35	36.96	2.02	39.12	0.28	25.62	3.43	24.48	0.88	22.72	1.87	NA	NA	NA	NA	NA	NA

Table F19: Methylation level of CpG 1-4 in *IL5RA* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *IL5RA*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	35.83	3.05	53.61	1.74	64.57	3.32	14.24	1.66	11.94	0.76	14.66	2.07	6.89	1.97	7.53	1.09	21.06	0.10	19.58	2.93	20.15	1.96
2	NA	NA	NA	NA	NA	NA	22.58	3.42	13.32	1.85	18.57	1.18	12.06	1.64	11.27	2.11	29.11	0.56	25.13	4.14	28.75	4.00
3	61.19	2.13	74.44	1.38	83.20	0.08	45.05	1.36	19.46	0.50	19.09	1.94	18.31	1.81	12.16	0.84	41.29	5.38	43.90	4.43	NA	NA
4	68.15	3.03	79.53	1.92	86.77	1.00	53.93	1.06	25.07	1.40	22.68	1.69	23.55	2.53	18.86	2.84	51.12	2.348	NA	NA	NA	NA

Table F20: Methylation level of CpG 1-11 in *LTC4S* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *LTC4S*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	8.06	0.81	21.78	1.23	33.96	2.07	6.05	2.75	24.93	2.51	38.87	0.88	30.36	4.43	46.62	3.14	NA	NA	7.05	3.26	1.94	2.74
2	16.63	3.10	29.45	0.40	50.50	1.09	12.83	1.27	35.74	2.33	52.64	1.44	39.74	2.18	53.56	0.05	NA	NA	8.49	0.87	7.86	0.41
3	11.03	1.88	19.50	2.26	40.90	2.25	6.84	1.44	24.01	1.93	34.76	0.61	26.70	2.50	33.36	0.62	NA	NA	0.00	0.00	5.99	2.28
4	15.68	0.62	26.97	1.80	46.28	2.53	11.65	1.34	33.28	2.45	47.75	1.86	35.27	2.28	48.18	2.98	NA	NA	11.75	2.51	10.82	1.03
5	15.32	0.65	30.55	4.00	50.04	0.34	10.97	1.80	35.42	2.50	55.61	1.65	40.10	0.77	53.72	1.06	NA	NA	11.01	1.99	NA	NA
6	15.73	1.05	32.54	0.13	42.30	3.70	13.48	1.96	35.44	2.50	54.88	2.17	39.53	1.42	53.26	0.57	NA	NA	8.78	0.95	NA	NA
7	20.13	0.69	37.50	0.54	53.74	0.34	19.76	2.28	39.80	2.09	63.54	3.40	46.00	2.25	59.69	3.07	NA	NA	16.21	0.25	NA	NA
8	18.96	2.18	36.09	2.57	50.15	4.02	19.76	1.55	37.52	1.17	60.59	2.94	43.82	2.67	58.29	0.82	NA	NA	13.86	1.42	NA	NA
9	18.42	2.11	35.14	0.15	48.58	1.56	19.69	2.93	37.52	1.62	57.19	0.92	42.73	0.58	56.97	2.66	NA	NA	10.95	0.63	NA	NA
10	21.35	0.23	37.38	0.09	50.68	0.55	15.84	1.47	38.88	0.89	56.56	0.65	41.79	1.74	56.52	0.69	NA	NA	12.08	2.79	NA	NA
11	28.68	0.23	45.67	0.72	55.68	0.81	19.00	1.31	47.01	1.86	63.63	1.89	49.38	2.76	62.15	0.43	NA	NA	18.87	1.29	NA	NA

Table F21: Methylation level of CpG 1-5 in *PCK1* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

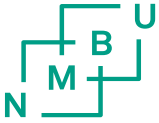
Gene *PCK1*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	56.47	3.68	45.98	4.31	44.01	2.76	57.29	3.39	38.32	1.01	20.06	1.63	16.51	0.08	12.54	0.98	45.24	4.28	NA	7.94	39.99	3.44
2	53.93	0.25	47.73	1.20	51.70	0.35	50.73	3.10	36.47	0.25	23.09	0.31	14.93	0.13	14.99	2.00	34.51	1.53	34.56	1.22	33.56	1.26
3	40.42	1.64	37.88	1.53	33.98	3.38	40.59	0.66	24.45	2.04	10.85	1.94	6.71	1.82	10.49	2.13	23.44	0.42	26.26	2.71	21.80	0.36
4	28.61	1.70	25.61	1.88	24.53	0.96	26.84	0.38	18.43	0.84	12.42	0.79	7.91	1.37	9.53	2.52	15.71	3.69	17.43	1.75	13.40	0.45
5	30.37	2.27	26.29	2.04	31.73	3.13	20.48	1.16	17.26	1.27	9.39	0.37	6.62	1.19	6.86	1.36	20.17	2.25	20.41	0.72	15.22	1.25

Table F22: Methylation level of CpG 5-16 in *RASSF1A* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *RASSF1A*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
5	1.17	0.54	1.41	0.02	0.66	0.66	1.57	0.86	1.06	0.30	23.13	4.51	32.36	2.54	33.94	1.39	5.24	4.72	4.45	2.73	2.58	0.95
6	1.83	0.06	2.31	0.31	1.63	1.63	2.81	0.50	2.03	0.08	25.82	4.35	33.71	1.75	37.03	0.35	1.15	1.62	4.19	0.97	3.06	0.32
7	0.40	0.56	0.56	0.79	0.51	0.51	0.74	0.65	0.99	0.19	23.33	3.77	33.88	2.19	36.78	1.94	0.69	0.97	4.38	1.24	2.13	0.90
8	2.97	0.25	2.25	0.52	3.71	3.71	3.74	0.98	2.98	0.01	24.60	4.57	35.72	1.72	38.35	2.05	6.29	3.56	5.56	2.50	3.36	0.59
9	0.57	0.80	1.18	1.67	0.70	0.70	0.82	0.78	1.16	0.21	22.06	4.52	32.75	2.24	34.35	1.73	4.89	4.34	4.58	2.12	2.22	1.24
10	0.83	1.17	0.00	0.00	1.43	1.43	1.32	1.17	1.39	0.30	21.82	3.40	30.30	1.91	33.89	1.32	0.82	1.16	1.92	1.67	1.18	1.04
11	0.32	0.45	0.00	0.00	0.00	0.00	0.15	0.27	0.34	0.30	18.05	3.81	28.17	2.03	29.73	1.27	0.57	0.80	0.49	0.84	0.31	0.54
12	0.99	1.40	0.99	1.39	2.82	2.82	2.43	0.80	1.88	0.09	20.97	3.41	29.96	0.97	32.74	3.39	NA	NA	3.74	2.52	2.34	0.11
13	NA	NA	0.00	0.00	0.90	0.90	0.00	0.00	0.00	0.00	26.57	0.82	32.63	2.41	34.42	0.62	0.00	0.00	2.39	2.08	0.38	0.66
14	NA	NA	0.61	0.86	1.52	1.52	0.80	0.33	1.26	1.19	27.26	1.68	37.76	1.16	41.78	0.65	0.76	1.31	2.57	2.50	0.31	0.54
15	NA	NA	0.62	0.87	0.00	0.00	0.73	0.64	1.55	1.65	25.58	1.95	32.64	1.93	35.91	3.98	2.29	2.90	0.75	1.30	0.37	0.65
16	NA	NA	0.97	1.37	0.00	0.00	1.12	0.15	1.06	0.95	26.43	1.48	35.43	1.35	37.54	3.03	0.46	0.79	2.50	2.61	0.29	0.50

Table F23: Methylation level of CpG 1-28 in *CDKN2A* for the regions within 137, 142 and 155 as well as for the normal samples (RP2, RP10 and RP11). For each CpG the mean of methylation levels (%) of triplicates and standard deviation (Stdev) are shown. The regions within a tumor are put in order with the region with the least tumor cells to the region with the most tumor cells from left to right. The mean (%) of a CpG with higher Stdev than 5 was not accepted and is marked NA. Both the mean (%) and Stdev are marked NA in the cases where the pyrosequencing result of a CpG was not accepted.

Gene *CDKN2A*																						
	137						155				142						RP					
	B		D		C		A		C		C		B		A		2		10		11	
CpG	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev	Mean (%)	Stdev
1	2.23	2.06	0.52	0.73	0.00	0.00	1.40	0.14	1.37	2.37	0.61	0.07	0.58	0.50	0.90	0.13	0.00	0.00	NA	NA	NA	NA
2	3.04	2.93	3.96	0.16	3.92	0.37	3.44	0.37	3.53	1.08	3.83	0.26	3.22	0.25	4.12	0.45	4.65	0.02	NA	NA	NA	NA
3	1.49	1.38	2.14	0.91	0.69	0.98	1.32	0.23	0.52	0.90	1.51	0.14	1.16	0.14	1.59	0.04	0.87	1.23	NA	NA	NA	NA
4	3.13	0.96	2.29	0.95	0.70	0.98	1.74	0.25	2.45	0.39	1.33	0.03	1.48	0.09	1.72	0.21	3.43	1.21	NA	NA	NA	NA
5	1.70	2.94	0.62	0.88	0.86	1.22	1.03	0.89	1.60	1.51	0.92	0.13	1.02	0.30	0.92	0.07	0.00	0.00	NA	NA	NA	NA
6	0.00	0.00	0.55	0.78	0.62	0.88	0.98	0.13	0.39	0.68	0.73	0.01	0.87	0.77	1.19	0.20	0.82	1.16	NA	NA	NA	NA
7	2.10	1.82	3.21	0.69	0.95	1.34	2.33	0.24	1.64	1.74	1.82	0.45	1.77	0.16	2.10	0.06	2.41	0.60	NA	NA	NA	NA
8	1.40	0.08	1.46	0.13	1.30	0.11	3.48	1.50	2.15	0.82	3.06	0.88	1.76	0.50	3.14	1.07	NA	NA	NA	NA	NA	NA
9	1.49	0.23	0.97	0.45	0.73	1.03	0.45	0.78	2.51	1.21	0.58	1.00	0.28	0.49	1.17	1.65	NA	NA	NA	NA	NA	NA
10	1.26	0.27	1.13	0.04	0.00	0.00	0.43	0.74	2.34	0.73	0.00	0.00	0.77	0.72	0.00	0.00	NA	NA	NA	NA	NA	NA
11	2.47	0.01	3.21	0.68	2.50	0.15	2.24	2.42	2.81	0.81	4.04	1.45	3.25	1.19	4.52	1.37	NA	NA	NA	NA	NA	NA
12	1.54	0.40	1.50	0.28	1.63	0.42	1.36	1.46	1.39	1.32	0.85	1.48	1.15	1.10	1.19	1.68	NA	NA	NA	NA	NA	NA
13	0.79	1.11	1.61	0.72	0.94	1.32	0.62	1.07	1.43	1.36	2.53	2.31	0.39	0.68	0.00	0.00	NA	NA	NA	NA	NA	NA
14	6.08	0.83	6.28	2.29	6.50	2.05	7.62	3.06	NA	NA	4.10	0.96	6.01	0.41	10.04	4.92	7.17	3.34	1.16	0.39	NA	NA
15	4.04	0.49	3.37	0.77	4.54	1.55	2.37	3.35	NA	NA	3.38	0.82	6.78	2.96	6.24	2.46	5.44	0.88	2.47	1.12	NA	NA
16	2.78	0.82	1.64	2.31	4.06	2.08	0.00	0.00	NA	NA	1.13	1.59	1.67	2.35	1.39	1.97	0.00	0.00	1.78	0.79	NA	NA
17	0.68	0.96	1.77	2.50	0.00	0.00	0.00	0.00	NA	NA	1.68	2.38	0.00	0.00	1.21	1.70	0.00	0.00	0.00	0.00	NA	NA
18	0.66	0.93	0.00	0.00	1.05	1.48	0.00	0.00	NA	NA	0.00	0.00	2.48	3.51	1.52	2.15	0.00	0.00	1.75	1.54	NA	NA
19	0.77	1.09	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA	0.00	0.00	2.98	4.21	1.69	2.38	0.00	0.00	1.74	1.53	NA	NA
20	1.11	1.57	3.72	0.59	0.00	0.00	1.04	2.08	NA	NA	1.21	1.76	4.68	1.07	4.46	4.87	1.80	2.58	3.37	0.25	1.67	1.50
21	0.82	1.16	3.78	1.28	0.00	0.00	0.00	0.00	NA	NA	1.17	1.64	2.21	3.13	4.78	4.88	1.99	2.66	1.92	2.71	2.50	2.88
22	NA	NA	NA	NA	NA	NA	4.43	0.19	NA	NA	2.80	2.58	NA	NA	NA	NA	4.86	2.45	NA	NA	4.36	1.48
23	NA	NA	NA	NA	NA	NA	0.00	0.00	NA	NA	0.00	0.00	NA	NA	NA	NA	1.36	1.92	NA	NA	0.99	1.72
24	NA	NA	NA	NA	NA	NA	6.88	0.84	NA	NA	3.61	3.28	NA	NA	NA	NA	4.98	0.57	NA	NA	3.02	2.65
25	NA	NA	NA	NA	NA	NA	5.66	0.72	NA	NA	0.00	0.00	NA	NA	NA	NA	1.87	2.64	NA	NA	2.20	3.11
26	NA	NA	NA	NA	NA	NA	5.48	0.12	NA	NA	1.66	2.87	NA	NA	NA	NA	5.27	3.05	NA	NA	4.59	1.20
27	NA	NA	NA	NA	NA	NA	2.57	3.63	NA	NA	1.37	2.37	NA	NA	NA	NA	4.39	2.26	NA	NA	1.05	1.81
28	NA	NA	NA	NA	NA	NA	6.42	1.27	NA	NA	5.75	1.05	NA	NA	NA	NA	6.80	1.85	NA	NA	6.03	3.58



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no