DETERMINING A PROMOTER HYPERMETHYLATION SIGNATURE FOR THE BLOOD-BASED DETECTION OF BASAL AND BRCA1-ASSOCIATED BREAST CANCER

by

Katrina L. Cristall

A thesis submitted to the Department of Pathology & Molecular Medicine

In conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(September, 2017)

Abstract

Liquid biopsies present an ideal way to non-invasively detect, assess and monitor malignancies. Tumour-specific promoter hypermethylation, a frequent and consistent alteration, in circulating tumour DNA (ctDNA) provides a promising marker for the development of these assays. There have been a number of studies attempting to design such a test but none have achieved the sensitivity required to both reliably distinguish individuals with cancer as well as detect extremely small quantities of ctDNA. This study presents a next generation sequencing (NGS) methylation-based blood test called mDETECT that is able to distinguish between metastatic triple negative breast cancer (TNBC) patients and healthy controls with a sensitivity of 100% and specificity of 95.7%. Furthermore, it has the ability to detect as little as 0.1 genome equivalents of DNA in a background of normal DNA 20,000 times greater. The ability of the hypermethylated regions probed by mDETECT have also been assessed for predictive value through analysis of DNA extracted from patient derived xenograft (PDX) mouse models. These models represent either chemoresistant or chemotherapy naïve disease. Through this it was determined that *TNFRSF10D* ($p = 0.025$) and *FOXA3* ($p = 0.076$) show significantly differential methylation between residual and untreated tumours. Additionally, the response of these models to various chemotherapeutics was previously established. By examining groups of tumours that are sensitive or resistant to various treatments, we discovered that the overall methylation patterns determined by mDETECT could stratify the tumours into two groups that differentially respond to cisplatin ($p=0.029$). Finally, both docetaxel ($p = 0.046$) and cisplatin ($p = 0.013$) resistance could be predicted by hypermethylation of *GIPC2*. Altogether, these results introduce mDETECT as a highly sensitive assay with potential for the early detection and monitoring of

cancer that includes probes with the ability to predict whether or not a TNBC patient is likely to respond to various lines of chemotherapy.

Acknowledgements

I would like to thank my supervisor Dr. Christopher Mueller for giving me the opportunity to work in his lab and for the support and guidance he has given me throughout my degree, especially during the process of generating this thesis. I would also like to thank my supervisory committee members Dr. Scott Davey and Dr. Sonal Varma for their input on my work and for challenging me to think about my project in new ways.

I also gratefully acknowledge the other researchers who contributed to this work. I thank Francois-Clement Bidard, Jean-Yves Pierga, Tatiana Papova and Marc-Henri Stern for the roles they played in the development of mDETECT. I also thank Elisabetta Marangoni for providing me with her patient derived xenograft samples. Lastly, I thank Ashley Huck for the effort she has put into the development of mDETECT and especially for her contributions to the analysis of the patient plasma samples described in this thesis.

Additionally, I want to express my thanks to the ladies of the Mueller lab; Ashley, Catherine and Hilary. Thank you for always being there to commiserate or celebrate, whatever the occasion called for. I don't know what I would have done without your friendship, advice and support.

Finally, I would like to thank my friends and family, both near and far, for all of the texts, phone calls and coffee dates that kept me going throughout the last two years. I truly don't know how I would have done it without you. To Ferne and Rob, I don't think I can ever fully express how much I appreciate you giving me a home away from home. Thank you for always sending me back to Kingston with a full stomach and just a little saner than I arrived. To my parents, Brenda and Neil, thank you for being my constant, stable support system and for filling my head with the crazy idea that I can do anything I set my mind to. Without your never-ending love, patience and understanding I wouldn't be the person I am today much less have made it through this degree.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABCG2	ATP-binding cassette G2
AC	adriamycin + cyclophosphamide
ALL	acute lymphocytic leukemia
ATCC	American Type Culture Collection
AUC	area under the curve
BEAMing	beads, emulsion, amplification and magnetics
BLBC	basal-like breast cancer
bp	base pairs
BRCA1	breast cancer 1
BRCA2	breast cancer 2
CD38	cluster of differentiation 38
CDKL2	cyclin dependent kinase like 2
cfNDA	cell free DNA
CMI	cumulative methylation index
CS1	consensus sequence 1
CS2	consensus sequence 2
CT	cycle threshold
ctDNA	cirulating tumour DNA
DCIS	ductal carcinoma in situ
DMRTA2	DMRT like family A2
dNTP	deoxynucleotide triphosphate
DR4	death receptor 4
DR5	death receptor 5
EGFR	epidermal growth factor receptor
ER	estrogen receptor
ESR1	estrogen receptor 1
EVX1	even-skipped homeobox 1
FEC	5-fluorouracil, epirubicin and cyclophosphamide
FOXA1	forkhead box A1
FOXA2	forkhead box A2
FOXA3	forkhead box A3

GABRA4	gamma-aminobutyric acid type A receptor alpha4 subunit
GIPC1	GIPC PDZ domain containing family member 1
GIPC2	GIPC PDZ domain containing family member 2
GSTP1	glutathione s-transferase pi 1
HER2	human epidermal growth factor receptor 2
HMEC	human mammary epithelial cells
HOXA9	homeobox A9
HOXB13	homeobox B13
HR	homologous recombination
IRF4	interferon regulatory factor 4
IRF8	interferon regulatory factor 8
LAR	luminal androgen receptor
MAST	microtubule associated serine/threonine kinase
mDETECT	methylation detection of circulating tumour DNA
MSP	methylation specific polymerase chain reaction
NF-kB	nuclear factor kappa-light-chain-enhancer of activated B cells
NFIC	nuclear factor 1 c
NGS	next generation sequencing
OTB	Ontario Tumor Bank
PAM50	prediction analysis of microarray 50
PARP	poly (ADP-ribose) polymerase
PAX6	paired box 6
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
pCR	pathological complete response
PDX	patient derived xenograft
PGM	Personal Genome Machine
POU4F1	POU class 4 homeobox 1)
PPFIA3	PTPRF interacting protein alpha 3
PR	progesterone receptor
PRKCB	protein kinase c beta
PTGDR	prostaglandin D2 receptor
QM-MSP	quantitative multiplexed methylation specific polymerase chain reaction
qMSP	quantitative methylation specific polymerase chain reaction

qPCR	quantitative polymerase chain reaction
ROC	receiver operating characteristic
SALL3	spalt like transcription factor 3
SEPT9	septin 9
SOX17	SRY-box 17
SPAG6	sperm associated antigen 6
TAL1	TAL BHLH transcription factor 1, erythroid differentiation factor
TCGA	The Cancer Genome Atlas
TE	tris-EDTA
TIL	tumour infiltrating lymphocytes
TNBC	triple negative breast cancer
TNFRSF10C	TNF receptor superfamily member 10c
TNFRSF10D	TNF receptor superfamily member 10d
TP53	tumour protein P53
TRAIL	tumour necrosis factor-related apoptosis-inducing ligand
TSG	tumour suppressor gene
UCSC	University of California Santa Cruz
VWC2	Von Willebrand factor C domain containing 2

Chapter 1

Introduction

1.1 Breast Cancer

Breast cancer is the most commonly diagnosed malignancy in Canadian women with approximately one in nine individuals receiving a diagnosis within their lifetime[1]. This is estimated to account for 25% of all cancer diagnoses made in Canadian women[1]. Over the last number of decades, increased screening as well as the development of targeted treatments for breast cancer have greatly increased patient survival[1]. However, breast cancer is still the second leading cause of cancer-related death in females and claims the lives of over 5000 Canadian women each year[1]. This demonstrates the need for improved methods to prevent, detect and treat breast cancer.

Breast cancer is a heterogeneous disease that can be divided into a number of different subtypes. This subtyping can be done in two ways[2,3]. The first relies on a tumour's expression of different hormone receptors and results in three groups; tumours that express the Estrogen Receptor (ER), ER-positive, those that overexpress Human Epidermal Growth Factor Receptor 2 (HER2), HER2-positive, and those that do not express either of these receptors or the Progesterone Receptor (PR) and are thus known as Triple Negative Breast Cancer (TNBC)[2]. As the expression of these receptors is determined using immunohistochemistry, these subtypes are known as the histological subtypes of breast cancer. Breast tumours can also by subtyped based on gene expression signatures. This method identifies what are known as the intrinsic subtypes of breast cancer and results in five different subtypes of the disease[3]. These are luminal A, luminal B, HER2 positive, basal-like and normal-like, each of which is defined by a particular

pattern of gene expression[3]. While the two classification systems yield different numbers of subtypes, comparisons can be drawn between the histological and intrinsic subtypes. For example, the majority of basal-like breast cancers are triple negative[4,5]. Determining the subtype of breast cancer an individual has is important as it has implications for both the treatment and prognosis of the disease[6,7].

1.2 Triple Negative & Basal-Like Breast Cancer

Triple negative breast cancers (TNBC) and basal-like breast cancers (BLBC) are two different but highly overlapping subtypes of breast cancer that account for approximately 15 to 20% of breast malignancies[4,5]. Both triple negative and basal-like tumours are aggressive forms of breast cancer that are characterized by a number of similar features such as a high rate of proliferation as well as both Tumour Protein P53 (*TP53)* and germline Breast Cancer Susceptibility 1 (*BRCA1)* gene mutations[8–10]. Additionally, both of these subtypes are associated with a poor prognosis[6,11]. As no targeted treatments are available for TNBC or BLBC[12], with the exception of poly-ADP ribose polymerase (PARP) inhibitors for BRCA1-associated malignancies[13], they are treated with cytotoxic chemotherapeutics[12]. Although targeted therapies could benefit this population, these women generally respond well to chemotherapy in comparison with the other subtypes of breast cancer [14,15]. However, the rate of relapse in women with TNBC or BLBC is high within the five years following diagnosis[14–17].

Despite numerous studies[18–22] there are still few ways to predict which TNBC and BLBC patients will respond to therapy or who is at a heightened risk of a relapse. Various markers have been identified that appear to indicate whether or not a patient will respond to therapy. These include the presence or absence of proliferation markers[21,22],

Epidermal Growth Factor Receptor (EGFR) status combined with metabolic response of the tumor during chemotherapy[19] and homologous recombination (HR) deficiency[23]. Furthermore, the loss of Androgen Receptor (AR) expression[18] as well as a particular microRNA signature[24] correlate with a higher likelihood of recurrence in patients with TNBC. However, none of these factors have been established for use in the clinic. As such, a technique to monitor a patient's disease that can be used to assess response to therapy as well as chance of recurrence would be highly valuable.

In addition to the numerous factors outlined above, TNBC and BLBC are both characterized by particular methylation patterns[25,26]. Overall these forms of breast cancer exhibit high levels of hypomethylation[2]; however, both hyper and hypomethylated regions are observed[26]. The specific DNA methylation patterns observed in these tumours differentiate BLBC from the other subtypes of breast cancer[27,28]. Furthermore, one study found that many of the genes observed to differ in expression between TNBC tumours and normal tissues had altered expression as a result of changes in DNA methylation[29]. Particular patterns of aberrant methylation not only characterize TNBC in general but also have prognostic value[26,29]. Specifically, it has been observed that TNBC tumours can be stratified based on methylation patterns into three groups with differing prognoses[26].

Aside from genome wide methylation patterns, the hypermethylation of specific genes also have particular implications in TN and BL breast tumours[30]. One example of this is the methylation of the *BRCA1* promoter[31–35]. Studies have shown that the *BRCA1* promoter is methylated in approximately 10% of breast cancer cases, primarily those of the TN or BL subtypes[31–35]. Additionally, methylation-induced inactivation of *BRCA1* is linked with the development of tumours displaying similar characteristics to those

malignancies found in carriers of a germline *BRCA1* mutation[32]. This shared phenotype is known as BRCAness[36] and it has been observed that tumours exhibiting this set of characteristics are sensitive to a type of drug known as PARP inhibitors[13]. As a result the identification of *BRCA1* methylation in TN and BL breast tumours may be important in guiding effective treatment of patients[37] as well as assessing the prognosis of the disease[38].

1.3 Germline BRCA1 Mutations and Breast Cancer

The average woman has an approximately 12.5% chance of developing breast cancer throughout their lifetime[1]; however, in women who carry a germline mutation in the *BRCA1* gene, this risk rises to between 60 and 80 percent[39]. BRCA1 is a tumour suppressor gene (TSG) involved in a number of cellular processes including DNA damage repair, transcriptional regulation, and cell cycle control[40,41]. As a result, BRCA1 is essential in maintaining genome integrity and it is through this function that it is thought to exert its tumour suppressing activity[42,43]. When an individual carries an inherited mutation in the *BRCA1* gene, a state of haploinsufficiency results[44]. Although one copy of the gene is still functional, it is not sufficient to entirely make up for the function of the lost allele. This phenomenon appears to cause genomic instability and other cell-type specific changes[45,46], depending on the particular mutation present[47]. This predisposes individuals to the development of breast and ovarian cancer[44]. In these cancers the remaining functional copy of *BRCA1* is often eliminated through loss of heterozygosity (LOH) leading to complete loss of *BRCA1* activity[48].

The breast tumours developed by carriers of a *BRCA1* mutation typically bear a particular set of characteristics[8,49,50]. These malignancies often develop in young women

and nearly all of these tumors are of the basal-like or triple negative subtype[8,49,50]. As such, they also exhibit the characteristics commonly associated with these subtypes such as high rates of proliferation and association with *TP53* mutations[51]. These findings, along with the fact that some studies have developed gene expression profiles capable of distinguishing *BRCA1*-associated tumors from sporadic as well as *BRCA2*-associated tumours[52,53], indicate that a particular tumourigenic process is associated with inactivation of *BRCA1*.

Although identifying a germline *BRCA1* mutation is a powerful tool for predicting an individual's risk of developing breast cancer[54], the options available to these patients are limited. Evidence suggests that the most effective way to prevent breast cancer in women with a *BRCA1* mutation is by performing a preventative bilateral mastectomy[55]. As this is clearly an invasive approach, many women instead opt to undergo regular surveillance in hopes of catching the disease in the early stages when it develops. However, these screening regimens are unfortunately much less effective than would be desired[55,56].

1.4 Breast Cancer Screening

Screening regimens have been widely implemented to monitor particular populations of women for the development of breast cancer[57]. These populations include women over the age of 50[58] and those carrying a germline *BRCA1* mutation[59]. The methods used in these screening programs vary but can include modalities such as clinical breast exam (CBE), mammography, ultrasound and magnetic resonance imaging (MRI)[57]. While some evidence suggests that these screening regimens have improved early detection of breast cancer and reduced the number of advanced stage diagnoses[60],

they are still associated with a number of challenges such as lack of sensitivity and overdiagnosis[61,62].

There are a number of factors to assess when judging how well a diagnostic test works. These characteristics include sensitivity, specificity and positive predictive value (PPV)[63]. Sensitivity refers to the ability of a test to correctly identify individuals who have a disease while specificity reflects the ability of the test to avoid giving a false diagnosis[63]. Furthermore, PPV is the probability that a patient who receives a positive test result actually has the disease[63].

The measures outlined above can be examined for the diagnostic techniques that are typically employed for breast cancer screening. The general population of women over the age of 50 are primarily screened using mammography[57]. Despite being widely used and having sensitivity and specificity of 77.6 and 98.8 respectively in this population, one study found that mammography only has a PPV of 35.8[64]. This means that only 35.8% of women who receive a positive mammogram result will actually have breast cancer. This leads to numerous unnecessary biopsies[61,62], putting undue strain on both falsely identified patients as well as the medical system itself with marginal, if any, impact on breast cancer mortality[61,62].

It has been observed that mammography is less effective in women at a high risk for breast cancer, such as those carrying a *BRCA1* mutation, in comparison with the general population[56]. This is possibly due to high breast density in these individuals[65]. One study found mammography to have a PPV of only 23.7 in women at a high risk of breast cancer[56]. As a result, studies have been conducted to assess the efficacy of other screening modalities in this population and have found MRI to be a more effective option

[56,66,67]. However, even MRI was only found to have a PPV of 50 in high risk women, meaning that only half of positive test results actually reflect the presence of cancer[56]. Furthermore, it has also been observed that the use of MRI for breast cancer screening results in more false positive results than the use of mammography[68].

Although current screening techniques are clearly suboptimal for all populations, regimens for high risk women present a particular set of challenges. This is due to the fact that these malignancies often develop at a young age and present as interval cancers[69], meaning they develop in between annual screening[69]. As a result, these women are typically screened frequently and for long periods of time[57]. This makes the common false negative and positive test results outlined above a particular problem for this population[56,68]. In addition, screening modalities involving radiation, such as mammography, can cause more damage from radiation exposure than the benefit conferred in cases where early detection is made possible[70,71]. Furthermore, although MRI has been identified to more effectively screen high risk populations and is safe to repeat frequently, cost is a major barrier to doing so[72,73]. These observations highlight the need for more effective screening modalities for both general and high risk populations that can be repeated frequently in both a safe and cost effective manner.

1.5 Liquid Biopsies for Cancer Detection & Assessment

Over the last number of years, liquid biopsies have emerged as an interesting alternative to other forms of diagnostic testing in cancer patients[74–76]. These assays monitor the presence of a particular tumour specific factor in the blood. This can include markers such as circulating tumour cells (CTCs)[77], extracellular vesicles[78] or circulating tumour DNA (ctDNA) [79–81]. Furthermore, these markers can be assessed in a wide range

of body fluids such as blood[82], stool[83], urine[84] and saliva[85]. Recently, a number of studies have demonstrated the utility of liquid biopsies in cancer patients with potential for use in early detection[86], assessing prognosis[87,88], guiding treatment decisions[89], monitoring response to treatment[90–93] and assessing risk of relapse[94,95].

Much of the success in liquid biopsy development has been seen in the context of detecting ctDNA in the blood[79–81]. These small fragments of DNA are released into circulation when tumor cells undergo cell death[96]. ctDNA comprises a small portion of the overall cell free DNA (cfDNA), which represents all of the DNA in circulation including that from both normal and tumour cells[97]. When ctDNA is released into the the blood it carries the specific molecular alterations that are found in the tumour with it[98,99], enabling these tumour-specific markers to be identified in the blood. To carry out a liquid biopsy, the cfDNA is interrogated for tumour specific markers such as mutations[89] or aberrant methylation patterns[82,88]. These markers should only be detectable if ctDNA, and therefore presumably a tumour, is present.

Assays focusing on the detection of ctDNA in the blood benefit from a number of advantages relative to other diagnostic tests that allow them to serve in a wide range of applications. First, due to the fact that ctDNA represents the entire genomic landscape of the tumour, these assays are not subject to the same biases as traditional biopsies as a result of tumour heterogeneity[90]. Additionally, these 'biopsies' involve only the extraction of blood, a safe, non-invasive and easy procedure compared with other testing modalities. As a result, these tests can be repeated frequently making them ideal candidates for the screening of high risk populations as well as monitoring a patient's disease over time. Numerous studies have shown that increasing ctDNA levels

correspond with suboptimal response to treatment[91–93]. As a result, these liquid biopsies can be used to assess how well as patient is responding to a given therapy[91–93].

1.6 Mutations in Circulating Tumor DNA

Much of the effort to develop liquid biopsies for the detection, analysis or monitoring of cancer has focused on the identification of tumour-associated mutations in ctDNA[81]. Some studies have focused on the identification of only one particular mutation[100–103]. While these assays may have utility in guiding treatment decisions, such as identifying EGFR mutations in lung cancer[89], they lack sensitivity in terms of early detection or monitoring of disease[101]. Two particular examples focusing on the monitoring of TNBC effectively highlight the difficulties in the use of mutations for the development of liquid biopsies[102,103].

In one study, *TP53* mutations, which are known to occur in 60-90% of TNBC tumors, were used as a way monitor ctDNA levels in 40 metastatic TNBC patients[102]. The tumour of each patient was sequenced using next generation sequencing (NGS) and 84% were found to possess a mutation in the *TP53* gene. For those patients in which a mutation was found, the plasma was then assayed for the presence of the particular mutation observed in the matched tumour. Of those patients found to have a *TP53* mutation in the tumour, 81% were found to have the same mutation in their cfDNA. This study also examined the levels of ctDNA, based on mutant allele frequency of *TP53*, before and after treatment in 12 patients. While 58% of these patients exhibited decreases in mutant allele frequency following treatment, no correlation with response to treatment could be established. Within this study there were a number of patients for which a *TP53* mutation could not be detected in the plasma despite its presence in the tumour. This

makes it difficult to know whether ctDNA level is in fact a poor predictor of response to treatment or whether ctDNA levels were just too low to be picked up using the method these researchers employed, thus giving an inaccurate view of fluctuations in these levels. As NGS has a limit of detection of 1%, meaning that only 1 molecule of ctDNA can be detected in a background of 99 normal cfDNA molecules, this study could possibly benefit from the use of a more sensitive technique[102] allowing for better detection of low levels of ctDNA.

To see if the sensitivity of their technique was the problem with the approach in the above study, another assay was developed to examine *TP53* mutations in the plasma of TNBC patients[103]. The tumours of 51 non-metastatic TNBC patients were sequenced, again using NGS, and for each patient in which a mutation was present, a droplet digital PCR (ddPCR) assay was designed. ddPCR is known to be a particularly sensitive assay and was therefore hoped to be more effective than the approach previously used[102,103]. Again the investigators used mutant allele frequency as a mark of ctDNA in the blood and tracked these levels in patients as they underwent treatment. Samples were collected at baseline, after the first round of chemotherapy, prior to surgery and after surgery. Unfortunately, despite the more sensitive assay, the researchers were essentially unable to detect any ctDNA in the patients following chemotherapy. Although four patients ultimately relapsed, this could not be predicted based on their ctDNA levels[103]. The downfall of this approach seems to be in the analysis of only one gene. ctDNA is generally at low levels in the blood and therefore, in a given blood sample, less than a genome equivalent is often present. This makes the ability to detect a single locus difficult, even with a sensitive approach.

These two studies demonstrate the fact that there are a number of challenges associated with liquid biopsy development. Following the observation of the limitation of a one gene approach, there has been increased success through the use of assays involving the detection of multiple mutations[92,104,105]. However, there are relatively few mutations that are found in a large number of tumours of a given cancer type[2,106]. For example, *TP53* is mutated in around 95% of high grade serous ovarian cancers (HGSOC)[107]. However, the next seven most commonly mutated genes are only mutated in between two and six percent of tumours[107]. Furthermore, even when a particular gene is commonly mutated in malignancies, these mutations can be spread out over large regions of the genome[107]. This necessitates either the personalized development of an assay for each patient based on the molecular alterations identified in their tumour or the use of a genome wide approach. This is not only labor intensive and expensive, but also limits the ability to use this system for the detection of cancer. As a result of these challenges, other markers, such as DNA methylation, have emerged as having the potential for use in the development of liquid biopsies.

1.7 Methylation in Circulating Tumor DNA

Aberrant DNA methylation is known to play an important role in tumourigenesis[108,109]. Generally, there are two pathological methylation patterns observed in cancer cells. Global hypomethylation, primarily in intergenic regions, leads to overall genomic instability[110]. Conversely, hypermethylation of specific CpG islands within promoter regions results in the decreased expression of TSGs[110]. Although genetic alterations have long been thought to drive tumourigenesis, epigenetic events are now known to be of equal importance, possibly representing the earliest events of

transformation[109,110]. Because of this, a number of studies have recently focused on determining genes specifically methylated in tumours as well as overall tumour-associated methylation patterns[26,82,86,111–113]. These signatures, which are often specific to a type[86,114] or subtype[26,113] of cancer, are useful in settings such as tumour stratification[26,111], assessment of prognosis[115–117] and diagnosis[86,114]. Like mutations, these aberrant methylation events can be detected in circulating DNA[118–120].

The detection of tumour specific hypermethylation in ctDNA carries a number of advantages when compared with the detection of mutations. This is most apparent when it comes to prevalence and consistency across tumors[118]. While there are few mutations which occur in a large number of a given cancer type[2,106], methylation events are much more consistent and numerous. There are often many different genes that are hypermethylated across large numbers of a particular form of malignancy[121]. As such, the focusing of ctDNA based tests on aberrant DNA methylation allows for the development of assays that will be able to identify more tumours of a given cancer type.

Many studies have assessed the ability of single hypermethylated genes to detect a particular type of cancer. For example, in breast cancer, promoter hypermethylation of numerous individual genes such as *Sox17*[115], *ESR1*[116] and *GSTP1*[117] have, when identified or tracked in circulating DNA, been found to be indicative of prognosis. However, even when highly prevalent, these markers are not found in all patients and thus typically lack in sensitivity when examined in terms of identifying patients who have the disease[122]. One of the most successful single-gene studies in this field relates to the the methylation of the gene *SEPT9* in colon cancer[122]. Methylation of this gene is detectable in the circulating DNA of patients with relatively high specificity in a

retrospective study[114]. However, in a prospective trial, detection of *SEPT9* methylation only identified the presence of colon cancer with a sensitivity and specificity of 48.2% and 91.5% respectively[122]. This is likely a result of smaller tumours as well as heterogeneity between them in the prospective study compared with the patients selected for the retrospective study. This example highlights the poor performance of single-gene assays when designing an assay for early detection of malignancies.

As in the context of mutation, due to the limited utility of tests examining only one hypermethylated gene, assays employing panels of hypermethylated genes have been developed[82,123,124]. These panels are typically composed of numerous genes that consistently show high levels of methylation in a certain type of tumour but not in normal tissues. These assays have had particular success in the context of prognosis but remain insufficient for the detection or monitoring of disease. One example, an assay termed cMethDNA, was developed for the monitoring and detection of metastatic breast cancer[82]. It incorporated a panel of 10 genes observed to be hypermethylated in breast cancer and was found to have a sensitivity of 91% and a specificity of 96% retrospectively[82]. While this is an improvement on many of the assays that have been developed, it is still unable to identify a number of patients and, if used in a cancer screening setting, would result in a large number of false diagnoses.

While there have been improvements in the development of methylation-based ctDNA assays, there are still many challenges. Many panels continue to use a relatively small number of genes. Because the concentration of ctDNA in the blood is so low, perhaps below a single genome equivalence, incorporating more hypermethylated regions in the assay could help to improve sensitivity. Secondly many approaches design assays

based on methylation patterns in types rather than subtypes of cancer[82]. This issue is especially relevant in the context of breast cancer as evidence suggests that aberrant methylation patterns are subtype specific[26,28,113]. Because of this, focusing attention on methylation patterns in particular subtypes may allow for increased efficacy of these assays. There is already a great deal of promise in cancer monitoring and diagnosis through the detection of tumour-specific methylation in ctDNA and with improvements in design, these tests have the potential to drastically improve cancer patient outcomes.

1.8 Objectives

Because of the fact that TNBC and BLBC breast cancers are highly similar[4,5], aggressive forms of the disease[8–10], and often present as interval cancers, improved screening and monitoring for these populations would have the potential to greatly benefit patient outcomes. As such, we hypothesize that BLBC and BRCA1-mutated tumours are characterized by a particular pattern of promoter hypermethylation and that the detection of this signature in the blood will enable the identification of TNBC patients relative to controls as well as report information regarding patient outcome and response to therapy.

The first objective of this study was to identify regions of tumour-specific promoter hypermethylation in basal-like and BRCA1-associated tumours. A separate analysis was done for these two groups due to the fact that BRCA1-associated tumours represent a small but clinically important subset of BLBC[8,49,50]. The two populations were analyzed separately to ensure that any regions of hypermethylation unique to BRCA1-associated tumours were identified.

Secondly, we aimed to develop and validate a NGS methylation-based assay for the detection of these hypermethylation signatures in circulating DNA. This development involved both the technical optimization of the assay as well as the validation of its efficacy in patient samples. While the hypermethylation signature was developed in basal-like tumours, the assay was targeted at the identification of TNBC. This is due to the fact that, while the basal-like subtype represents a more biologically relevant class of breast cancer, receptor status is typically the only assessment done clinically and therefore the only information available regarding patient samples.

The final aim of this study was to assess whether any of the regions identified to be specifically hypermethylated in BLBC or BRCA1-associated breast cancer are indicative of patient response to treatment and thus capable of acting as predictive biomarkers.

Chapter 2

A Highly Multiplexed Next Generation Sequencing DNA Methylation Based Blood Test for Triple Negative Breast Cancer

2.1 Abstract

Liquid biopsies present an ideal way to non-invasively detect, assess and monitor malignancies. Of particular promise is the detection of tumour-specific promoter hypermethylation, a frequent and consistent marker, in ctDNA. There have been a number of studies attempting to design such a test but none have achieved the sensitivity required to both reliably distinguish individuals with cancer as well as detect extremely small quantities of ctDNA. Here we present a next generation sequencing methylation-based blood test called methylation DETEction of Circulating Tumour DNA (mDETECT) that is able to distinguish women with metastatic triple negative breast cancer from those without, with a sensitivity of 100% and specificity of 95.7%. Furthermore, it has been demonstrated to have the ability to detect as little as 0.1 genome equivalents of DNA in a high background of normal DNA. Based on the tremendous sensitivity displayed by this assay, mDETECT has a great deal of potential for the early detection and monitoring of cancer.

2.2 Introduction

The detection of circulating tumour DNA (ctDNA) has proven to be a useful clinical tool that is able to help guide cancer patient management[74–76]. This ctDNA is released into the blood when cancer cells undergo cell death[96] and carries with it the same molecular markers as the DNA found in tumour cells[98,99]. By interrogating the blood for

the presence of ctDNA as well as the markers it bears, a great deal of information can be learned about an individual's disease. These analyses have utility in early detection[86], monitoring response to treatment[90–93] and prognosis[87,88]. Over time these assays have becoming increasingly complex and to date have primarily been focused on the identification of tumour specific mutations[81]. Significant technical issues have limited the overall sensitivity and specificity of mutation-based approaches. These include elevated background levels due to the presence of high concentrations of normal DNA[125], high error rates in both PCR[126] and sequencing[127] based approaches and limited sensitivity due to the small number of available mutations to detect and the high variability of those mutations[2,106]. A variety of approaches have been used in an attempt to mitigate these issues including BEAMing[128] and sequencing error reduction assays[129], but fundamental issues continue to limit the efficacy of these tests.

DNA methylation is proving to offer significant advantages over mutation-based detection for the analysis of circulating tumour DNA[110]. It has long been known that during the process of tumourigenesis gene specific hypermethylation and genome wide hypomethyation occurs[110]. Hypermethylation is associated with the inactivation of tumour suppressor genes, which provide the tumour with growth advantages[110]. These hypermethylation events may occur relatively early in the transformation process[109,110] and are generally frequent and consistent within tumour types[26,27,86,114]. Regions of hypermethylation have proven to be useful biomarkers for circulating tumour DNA in a variety of cancer types. Individual methylation markers are somewhat limited as they are not uniformly present in all tumours[115–117,122]. The use of multiple hypermethylated

regions has improved the performance of these tests, but PCR based approaches still have significant background issues, limiting their sensitivity[82,124].

TNBC and BLBC are two highly overlapping groups of tumours that represent 15-20% of breast cancers[4,5]. These subtypes are associated with an aggressive phenotype, poor prognosis[6,11,16] and both *TP53*[6] and germline *BRCA1* mutations[9,10]. Current breast cancer screening regimens are insufficient across all populations of women[61,62] but this deficiency is particularly relevant for those women who develop these aggressive subtypes. As such, this population of women would receive a heightened benefit from early diagnosis and effective monitoring of their disease.

In order to overcome the limitations of both mutation and current methylation based blood tests we have developed a highly multiplexed NGS based assay for the detection of tumour specific hypermethylation in blood. Regions of tumour-specific hypermethylation were identified using data from The Cancer Genome Atlas and multiple PCR primers to those regions were designed that preferentially recognize methylated DNA. These probes were then incorporated into a multiplex PCR reaction to amplify DNA from cell free DNA samples. NGS of this multiplex reaction allows for the identification of tumour specific methylation patterns associated with multiple CpG residues within the amplified regions, dramatically increasing the specificity of detection. By using NGS on a multiplex PCR of a large number of probes, compared to PCR alone, increased sensitivity is achieved by both increasing the likelihood that a given tumour carries multiple hypermethylated regions and at the same time allowing for the detection of less than a single genome's worth of DNA in a given sample. We have named this assay "methylation DETEction of Circulating Tumour DNA" (mDETECT). Using Triple

Negative Breast Cancer as a model we demonstrate that this test is superior to a p53 mutation based assay in the metastatic setting.

2.3 Materials and Methods

2.3.1 Test Development

Publicly available Level 3 Illumina 450K methylarray data from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) was obtained for the 771 breast cancer patients and 101 normal breast samples in the Breast invasive carcinoma [BRCA] dataset. The methylation values in this set are from -0.5 to +0.5. PAM50 classification using RNA-Seq data was used to identify tumours of the basal subtype. Tumours containing a germline *BRCA1* mutation were also identified. TCGA methylation data for normal samples from lung, colorectal and prostate were also extracted and the average beta value of all samples of a given tissue type was determined for each methylation probe. All probes were mapped to their chromosomal location and to the tumour and normal tissue data. CpG residues that were differentially methylated between tumours and normal tissues were identified in two different ways. First, where the average methylation value of the tumours was greater than -0.1 and the average methylation for each normal tissue was less than -0.3. Secondly, where at least 50% of tumours had a methylation value greater than -0.1 and where each of the normal tissue averages was less than -0.3. For both of these methods, CpGs were identified only if 2 probes meeting the outlined characteristics were within 300 base pairs of each other. This process was carried out for both basal as well as BRCA1-associated tumours. Identified regions from both populations were then taken forward for further characterization and prioritized based on the possession of multiple CpG residues with the indicated characteristics.

Furthermore, 7 additional regions were identified in the literature in which hypermethylation is found to be a prognostic indicator for TNBC[26,130]. These regions were developed alongside those we identified in our analysis.

Each of the identified regions was characterized by examining the flanking CpGs additional concerted differences in methylation, and a target region for probe development was defined. These regions were characterized using the UCSC Genome Browser[131] MethBase data tracks[132] to ensure that these regions were unmethylated in human mammary epithelial cells (HMEC), as well as peripheral blood mononuclear cells (PBMC) and other specific blood cells. The sequence of these regions were altered to the sequence of bisulphite converted DNA based on all CpGs being methylated using Methyl Primer Express Software v1.0 (ThermoFisher Scientific, Waltham, MA, USA). PCR primers to these regions were developed using either Methyl Primer Express Software v1.0 or Primer Blast (National Center for Biotechnology Information, Bethesda, MD, USA)[133] with the following settings: Minimum PCR product size 75bp, maximum 125bp; melting temperature minimum 64^{o}C, optimal 68^{o}C, maximum 72^{o}C, maximal differential 5^{o}C, primer size minimum 21bp, optimal 28bp, maximum 31bp. Primers with at least 2 CpG residues in the region between the primers were chosen. Concensus sequence 1 (CS1) and concensus sequence 2 (CS2) primer sequences were added to the 5' ends of the primers.

An additional 136 regions had previously been developed using the same parameters for the other subtypes of breast cancer (Luminal A, Luminal B, HER2+ and normal-like) (Mueller, Unpublished). These probes were examined in the basal and BRCA1-associated tumours in the TCGA Illumina 450K methylarray data and any that

exceeded a methylation level of -0.1 in over half of either population were included in the assay.

2.3.2 Preparation of DNA from Breast Cancer Cells Lines

Seven cells lines were used for the purposes of this study. The non-tumourigenic human breast cell line, MCF-10A, the estrogen receptor positive human breast cancer cell line, MCF7, the HER2+ human breast cancer cell line, SKBR3, and the triple negative human breast cancer cell line, MDA-MB-231, were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). The telomerase reverse transcriptase immortalized non-tumourigenic human breast cell line, 184-hTERT, was a gift from Dr. Calvin Roskelley (University of British Columbia, Vancouver, BC, Canada). The triple negative, BRCA1 mutant human breast cancer cell line, HCC1937, was a gift from Dr. Andrew Craig (Queen's University, Kingston, ON, Canada). The triple negative, BRCA1 mutant human breast cancer cell line, MDA-MB-436, was a gift from Dr. Xiaolong Yang (Queen's University, Kingston, ON, Canada). All cell lines were cultured according to the guidelines provided by ATCC. DNA was extracted using the GenElute Mammalian Genomic DNA Mini-Prep Kit (Cat. No. G1N-70, Sigma-Aldrich, Oakville, ON, Canada) according to the manufacturers protocol. The concentrations were determined using spectrophotometry. 1µg of DNA was bisulfite converted as described in section 2.4.2.

2.3.3 Preparation of DNA from Buffy Coat

Buffy coat was obtained from a blood sample taken from a healthy volunteer at the Curie Institute. DNA was extracted from this buffy coat using the QIAmp circulating nucleic acid kit (Cat. No. 1017647, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions. Buffy coat DNA was bisulfite converted using the EpiTect

Fast DNA Bisulfite Kit (Cat. No. 59824, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions and eluted in 10µL of elution buffer.

2.3.4 Primer Validation in Cell Lines

Each primer set was characterized by amplifying DNA from the cell lines 184-hTERT, MCF-10A, MCF7, SKBR3, MDA-MB-231, HCC1937 and MDA-MB-436 as well as buffy coat DNA from a healthy volunteer. PCR was carried out using 0.5U Qiagen Hot Start Taq Plus (Cat. No. 203205, Qiagen, Toronto, ON, Canada) in a 25ul reaction with 1X PCR buffer (Cat. No. 203205, Qiagen, Toronto, ON, Canada), 0.2mM deoxynucleotide triphosphates (dNTP) (Cat. No. N0440S – N0443S, New England Biolabs, Ipswich, MA, USA), 3mM $MgCl_2$ (Cat. No. 203205, Qiagen, Toronto, ON, Canada), 10ng of bisulphite converted DNA and 2pmol of the forward and reverse primer (Sigma-Aldrich, Oakville, ON, Canada). PCR conditions were 95°C for 15 minutes, followed by 35 cycles of 95 °C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds followed by 72°C for 7 minutes and a hold at 4°C. Secondary PCR and barcoding was carried out using 0.5U Qiagen Hot Start Taq Plus (Cat No. 203205, Qiagen, Toronto, ON, Canada) in a 50ul reaction with 1X PCR buffer (Cat No. 203205, Qiagen, Toronto, ON, Canada), 0.2mM dNTPs (Cat No. N0440S – N0443S, New England Biolabs, Ipswich, MA, USA), 3mM $MgCl_2$ (Cat No. 203205, Qiagen, Toronto, ON, Canada), 2ul of primary PCR product and 0.04µM CS1 universal primer with a unique barcode for each sample and CS2 universal primer (Sigma-Aldrich, Oakville, ON, Canada). PCR conditions were 95°C for 15 minutes, followed by 3 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds followed by 72 °C for 7 minutes and a hold at 4°C. Samples were combined after barcoding, purified with Agencourt AmpureXP magnetic

beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturers protocol at a ratio of 1:1.2 (sample to beads) and eluted in 50µL 1X Tris-EDTA (TE) buffer. Following purification samples were sequenced on the Ion Torrent PGM platform.

Sequencing generated BAM files were converted to FASTA files using GenAp Galaxy[134]. This software was also used to filter out all reads below 100bp to eliminate those reads representing primer dimer. BiqAnalyzer HT[135] was used to analyze the methylation status of each probe. Reference sequences corresponding to the region between PCR probes were used to ensure the methylation status of only the amplified regions were being assessed. Methylation analysis was carried out with a minimum sequence identify of 0.9.

2.3.5 Preparation of DNA from TNBC Tumors

14 TNBC tumours were obtained from the Ontario Tumor Bank (OTB). The DNA was previously extracted from these tumours for a separate study[136] by homogenizing the tissue and extracting DNA using the AllPrep DNA/RNA Mini Kit (Cat No. 80204, Qiagen, Toronto, ON, Canada) and stored at -80^{o}C. Concentrations were determined using spectrophotometry. For the purposes of this project, 100ng of DNA from each tumour or adjacent sample was bisulfite converted as described in section 2.4.2.

2.3.6 Multiplex PCR & Sequencing

Conditions for multiplexing were optimized using bisulfite converted DNA from the TNBC cell line HCC1937. The optimized conditions are as follows. A 25µL multiplex PCR reaction was carried out using 1X Qiagen Multiplex PCR Plus mix (Cat

No. 206152, Qiagen, Toronto, ON, Canada), 5ng of bisulphite converted DNA and 0.015µM of a combined primer mix containing the forward and reverse primers for 86 primer sets. Cycling conditions were 95°C for 15 minutes, 35 cycles of 95°C for 30 seconds, 62°C for 90 seconds, 72°C 90 seconds followed by 68°C for 10 minutes and a hold at 4°C. In order to reduce primer dimers generated in this multiplex reaction being carried on to subsequent amplification, 20µL of the primary PCR reaction was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturers protocol at a ratio of 1:0.8 (sample to beads). Following purification, the sample was eluted in 25µL of water. Secondary PCR and barcoding was carried out essentially according to the protocol described in section 1.4.4 with the completion of 10 PCR cycles. 45ul of this secondary PCR was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturers protocol at a ratio of 1:1.2 (sample to beads) followed by elution in 25ul of water. Products were then sequenced on an Ion Torrent Proton platform.

Sequencing generated BAM files were converted to FASTA files using GenAp Galaxy[134] and analyzed using BiqAnalyzer HT[135] according to the protocol outlined in section 1.4.4. In addition to the previously described BiqAnalyzer analysis, data was also analyzed by restricting analysed reads to a minimum methylation level of 0.8 so as to identify only methylated reads.

2.3.7 Determining the Sensitivity of the Assay

Bisulfite converted DNA from the TNBC cell line HCC1937 was serially diluted into a constant 10ng of bisulfite converted DNA from the buffy coat of a healthy

volunteer. This was done so that six solutions were generated containing the following quantities of HCC1937 DNA: 0.01ng, 0.005ng, 0.0025ng, 0.001ng, 0.0005ng and 0.0001ng in addition to the 10ng of buffy coat DNA. These samples were prepared for Ion Torrent sequencing in the same fashion as was described in section 1.4.6.

Sequencing generated BAM files were converted to FASTA files using GenAp Galaxy[134] and analyzed using BiqAnalyzer HT[135] according to the protocol outlined in section 1.4.6.

Genome equivalents of DNA was determined assuming that 3.3pg of DNA as being one genome equivalent[137].

2.3.8 Preparation of DNA from TNBC Patient Plasma

Patients with metastatic TNBC were drawn from a cohort collected at the Curie Institute[102]. We obtained plasma samples from 14 of the patients in this study. Plasma from 23 normal volunteers was also obtained. cfDNA was extracted as described previously[102]. Briefly, plasma samples of between 1.2 and 2mL volume were processed using the QiaSymphony system (Qiagen, Toronto, ON, Canada), with the QIAmp circulating nucleic acid kit (Cat. No. 1017647, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions. Volumes were made up to 2mL using phosphate buffered saline (PBS). The cfDNA was eluted in 60ul of elution buffer. Of this, 40ul was bisulfite converted using the EpiTect Fast DNA Bisulfite Kit (Cat. No. 59824, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions and eluted in 25µL of water.

2.3.9 Expanded Multiplex PCR & Sequencing

To determine the C_T value for each primer qPCR was carried out in two TNBC cell lines; HCC1937 and MDA-MD-436 using 1X PowerUp™ SYBR® Green Master Mix (Cat. No. A25742, Applied Biosystems, Austin, TX, USA) in a 25µL reaction containing 10ng bisulfite converted DNA and 1µM of forward and reverse primer. Primer sets were ranked according to the lowest C_T value between the two cell lines. These values were used to group primer sets into miniplex groups containing no more than 8 primer sets according to similar C_T values.

To begin the expanded multiplex protocol, a 25µL linear pre-amplification PCR was performed using 1X Qiagen Multiplex PCR Plus mix (Cat No. 206152, Qiagen, Toronto, ON, Canada), 10µL of bisulphite converted cfDNA and 0.02µM of a combined primer mix containing only the reverse primers for 68 primer sets. Cycling conditions were 95°C for 15 minutes, 10 cycles of 95°C for 30 seconds, 62°C for 90 seconds, 72°C 90 seconds followed by 68°C for 10 minutes and a hold at 4°C. The entire 25µL pre-amplification was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturer's instructions at a ratio of 1.2:1 (beads to sample) and eluted in 50µL of water. Primary PCR was carried out for each miniplex reactions for each sample essentially according to the protocol outlined in section 1.4.6 with the addition of 5µL purified linear pre-amplified product and 0.02µM of a combined primer mix. 15µL of each miniplex reaction was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturer's instructions at a ratio of 1.2:1 (beads to sample) and eluted in 20µL of water. Finally, secondary PCR and barcoding was carried out essentially according to the protocol

outlined in section 1.4.4 with the exception that it was conducted in a 25µL reaction. All reagents were kept at the same ratio with the addition of 10ul of purified primary PCR product. Thermocycler conditions were carried out as described in section 1.4.6. Following barcoding, an equal volume of each miniplex reaction for a given sample was pooled. This pooled mix was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturer's instructions at a ratio of 1.2:1 (beads to sample) and eluted in 20µL of water. These products were then sequenced on an Ion Torrent platform.

This protocol was initially optimized using bisulfite converted HCC1937 DNA and using an input of 0.1ng.

Sequencing generated BAM files were converted to FASTA files using GenAp Galaxy[134] and analyzed using BiqAnalyzer HT[135] according to the protocol outlined in section 1.4.6.

2.3.10 Statistical Analyses

Percentage of reads lost to primer dimer was calculated by dividing the number of reads left after filtering in GenAP Galaxy and BiqAnalyzer by the total number of reads for a given sample.

The number of signaling probes for a given sample was determined by calculating the total number of probes for a sample that have greater than one methylated read.

Total normalized methylated reads were determined for each sample by adding together the normalized methylated reads for all probes of a sample. Normalized total methylated reads for each probe was calculate as follows:

$$\frac{average\ total\ reads\ for\ all\ samples}{total\ reads\ for\ sample} \times methylated\ reads\ for\ probe$$

Corrected methylated reads were determined for each probe according to the following formula:

$$\frac{sum\ of\ reads\ exported\ by\ BiqAnalyzer\ for\ a\ sample}{total\ reads\ for\ sample} \times methylated\ reads\ for\ probe$$

Cumulative Methylation Index (CMI) was determined by summing the corrected methylated reads for all probes in a given sample. The CMI for plasma DNA samples was used in receiver-operating characteristic (ROC) analyses to assess the performance of the test by determining the area under the curve (AUC) as well as sensitivity and specificity of the assay. This was done using GraphPad Prism.

The best performing probes were identified based on those with greater than two-fold difference between average corrected methylated reads for the TNBC patient plasma samples relative to the 6 most methylated cfDNA controls.

To assess correlation between ctDNA concentration and CMI, values were plotted and a linear regression was applied. The R^2 value was used to assess the degree of correlation between the variables.

2.4 Results

2.4.1 Identification & Development of Tumor-Specific Hypermethylated Regions

In order to design an appropriate methylation based blood test for the detection of the basal tumour population we established an assay composed of numerous methylation biased primer sets, or 'probes', that amplify small regions of the genome identified to be hypermethylated in tumours but not in normal tissue (Figure 2.1). The test, termed

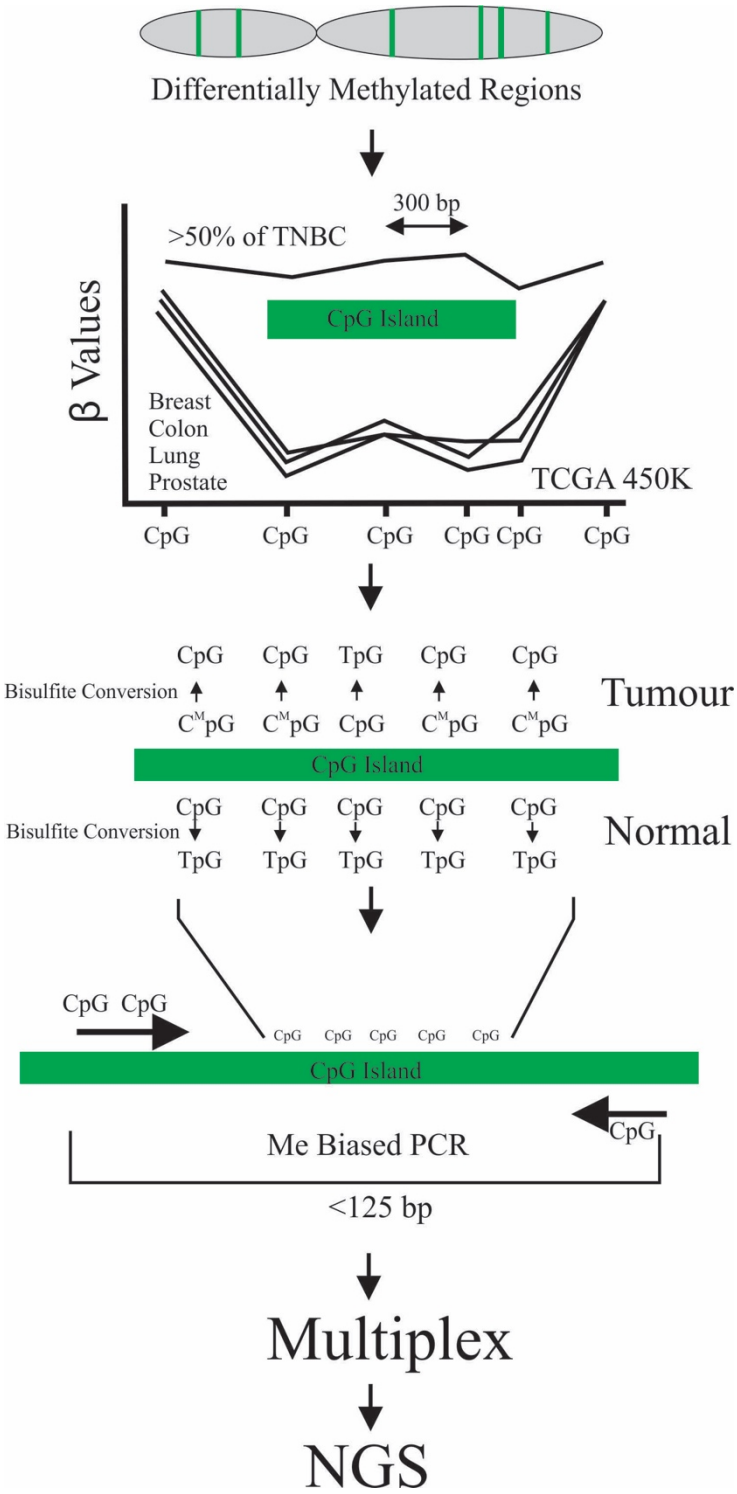

a)
Differentially Methylated Regions
b)
300 bp
>50% of TNBC
CpG Island
β Values
Breast
Colon
Lung
Prostate
TCGA 450K
CpG CpG CpG CpG CpG CpG
c)
CpG CpG TpG CpG CpG
Bisulfite Conversion
C^MpG C^MpG CpG C^MpG C^MpG
Tumour
CpG Island
CpG CpG CpG CpG CpG
Bisulfite Conversion
TpG TpG TpG TpG TpG
Normal
d)
CpG CpG
CpG CpG CpG CpG CpG
CpG Island
CpG
Me Biased PCR
<125 bp
e)
Multiplex
f)
NGS

Figure 2.1 Overview of the mDETECT assay.

mDETECT relies on the blood based detection of differential methylation regions that are specific to tumours. These regions are represented by the green stripes on the grey chromosome (a). To identify these regions, Illumina 450K methylome data was obtained from TGCA for TNBC & BRCA1-associated tumours as well as normal breast, colon, lung and prostate tissue. Those CpG dinucleotides showing high levels of methylation in the tumour populations but low levels in the normal tissues were identified as being differentially hypermethylated. β Values represent methylation levels at a given probe (CpG dinucleotide). These values are plotted on the graph (b) relative to position to show an example of a region that is differentially methylated between tumour and normal tissues. Additionally, only those areas that had two differentially methylated CpGs in close proximity (300 bp) were selected. Bisulfite conversion (c) of the DNA samples of these differentially methylated regions allows the multiple CpG residues in each region to be assessed and subsequently allow for differentiation between tumour and normal samples. Methylation biased primer sets, as indicated by the arrows in (d) designed for the amplification of the tumour-specific hypermethylated regions are used to carry out PCR. These primers preferentially, but not exclusively, amplify methylated DNA, which acts to ensure that small amounts of methylated tumour DNA is amplified even in a high background of normal DNA. Furthermore, these PCR 'probes' were designed to amplify products less than 125bp due to the highly fragmented nature of ctDNA (d). PCR is carried out by multiplexing all probes together (e) ensuring that tumour-specific methylation will not be missed as a result of a fragment of ctDNA being in a different tube than the primer set corresponding to that region. When Next Generation Sequencing (NGS) (f) is carried out, multiple CpG dinucleotides that were methylated in tumours produce a strong redundant signal compared to normal DNA.

mDETECT, involves a highly multiplexed PCR that includes all probes in the assay and is subsequently sequenced using NGS to detect tumour-specific hypermethylation. Here we have designed mDETECT to specifically recognize TNBC-associated hypermethylation; however, this assay could be adapted for the recognition of nearly any malignancy.

To begin, we obtained Illumina 450k methylation data for the TCGA breast cohort. Tumours of the basal subtype were selected based on the PAM50 molecular signature as well as tumours with BRCA1 mutations. This resulted in the identification of 90 basal tumours and 22 BRCA1 mutated tumours. Tumour-specific hypermethylated regions were identified that consisted of at least 2 methylated CpG residues within 300 bp of each other, where there was tumour specific hypermethylation in the majority of samples (>50%) when compared to normal samples from breast, prostate, colon and lung (see Materials and Methods for details). These regions were predominantly associated with CpG islands, typically occurring around the CpG island shore. An example of a differentially methylated region located upstream of Cyclin Dependent Kinase Like 2 (*CDKL2)* is shown in Figure 2.2. In addition to the 64 hypermethylated regions identified here, 6 additional regions were developed that have been observed to be prognostic for TNBC in the literature[26]. The *BRCA1* promoter region was included due to the fact that *BRCA1* hypermethylation has implications for TNBC treatment[37,38] and prognosis[130]. In total, 70 regions were selected for initial development and validation and PCR primers were created to amplify these areas. These primers were designed based on the bisulphite converted sequence where all CpG residues were considered to be methylated. Depending upon the specific primer pair, these either exclusively amplify methylated

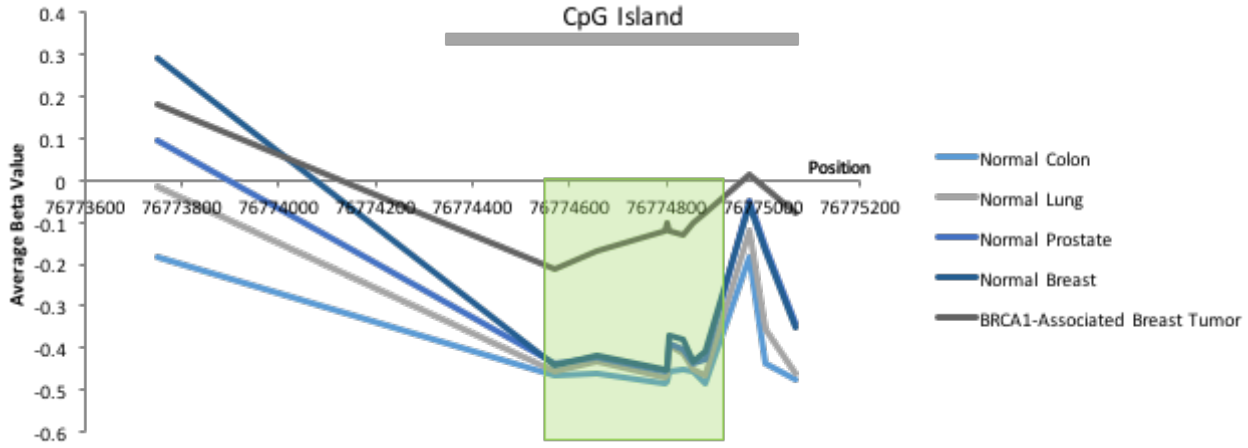


Figure 2.2 Methylation in the promoter region of the gene *CDKL2* in tumor and normal samples.

Illumina 450K methylarray probes in the *CDKL2* promoter were found to show an average beta value (y-axis) greater that -0.1 in the majority of the 22 BRCA1-mutated breast tumours in the TCGA database and less than -0.3 on average in normal colon (Normal Colon), lung (Normal Lung), prostate (Normal Prostate) and breast (Normal Breast) tissues in the TCGA database. The methylation values for the each of these samples were plotted relative to genomic position (x-axis) (human genome 18 (hg18) nucleotide position on Chromosome 4) to determine the extent of differential methylation in surrounding areas. The differentially methylated region is highlighted (green box) and falls directly within a CpG island (gray bar). Outside of this region, the level of methylation begins to converge for the tumour and normal samples.

DNA or amplify a mixture of methylated and non-methylated DNA with a bias towards methylated sequences. These primers were designed to amplify regions less than 125bp due to the fact that ctDNA is highly fragmented[138]. Furthermore, primers were designed to have a higher than average melting temperature as this has been observed to enhance their ability to be multiplexed, allowing greater sensitivity to be achieved. Multiple primers were created per region resulting in a total of 131 PCR 'probes' associated with 70 regions.

2.4.1 Validation of Probes in Cancer Cell Lines

The primers designed for amplifying regions of tumour-specific hypermethylation were assessed for their ability to specifically recognize tumour DNA by individual PCR amplification reactions (singleplex PCR). This was done using two immortalized mammary epithelium cell lines, 184-hTERT and MCF-10A, and ER+ breast cancer cell line, MCF7, a HER2+ breast cancer cell line, SKBR3 and three TNBC cell lines, MDA-MB-231, HCC1937 and MDA-MB-436, as well as DNA from the buffy coat blood fraction of a healthy volunteer.

Singleplex reactions were barcoded by sample type and sequenced on an Ion Torrent platform. Methylation of these regions was assessed using BiqAnalyzer HT[135] by examining the methylation status of CpG residues lying only between PCR primers. All such regions contained between 3 and 17 CpG residues ensuring the detection of genuinely methylated regions. An example of the singleplex analysis for the *NFIC* promoter region in a variety of breast cancer cell lines as well as in buffy coat DNA is

shown in Figure 2.3. There were variable levels of methylation across different breast cancer cell lines highlighting that not all probes will be methylated in all instances of TNBC and that some probes are hypermethylated in other subtypes of breast cancer, not just TNBC. Based on these results it was generally observed that the majority of reads were found to either be entirely methylated or entirely unmethylated across the CpGs in the region. Thirty probes were found to be methylated in buffy coat DNA and, as such, were eliminated from the assay. Several other probes were eliminated due to overlap between primers derived from the same region, making them incompatible for multiplexing. Potential commonalities between primer sequences that could result in primer dimer formation during multiplexing also resulted in probe elimination. These incompatible regions were identified using the Autodimer software[139]. In all, the 70 regions were reduced to 49 and the total number of probes from 131 to 61.

Following validation of the probes designed specifically for the basal and BRCA1-associated tumour populations, another set of probes was assessed that were developed for the other subtypes of breast cancer (luminal A, luminal B, HER2+ and normal-like) (Mueller, Unpublished). Any of the probes developed for these subtypes that were positive in over half of either the basal or BRCA1-associated tumours based on the Illumina 450K methylation data in the TCGA database were added to the assay. This resulted in a total of 86 probes corresponding to 71 regions (Appendix A).

2.4.2 mDETECT Characterization of TNBC Tumors

The probes remaining following validation were then multiplexed together and the multiplex PCR conditions were optimized using the TNBC cell lines MDA-MB-231, HCC1937 and MDA-MB-436 as well as normal buffy coat DNA (Figure 2.4).

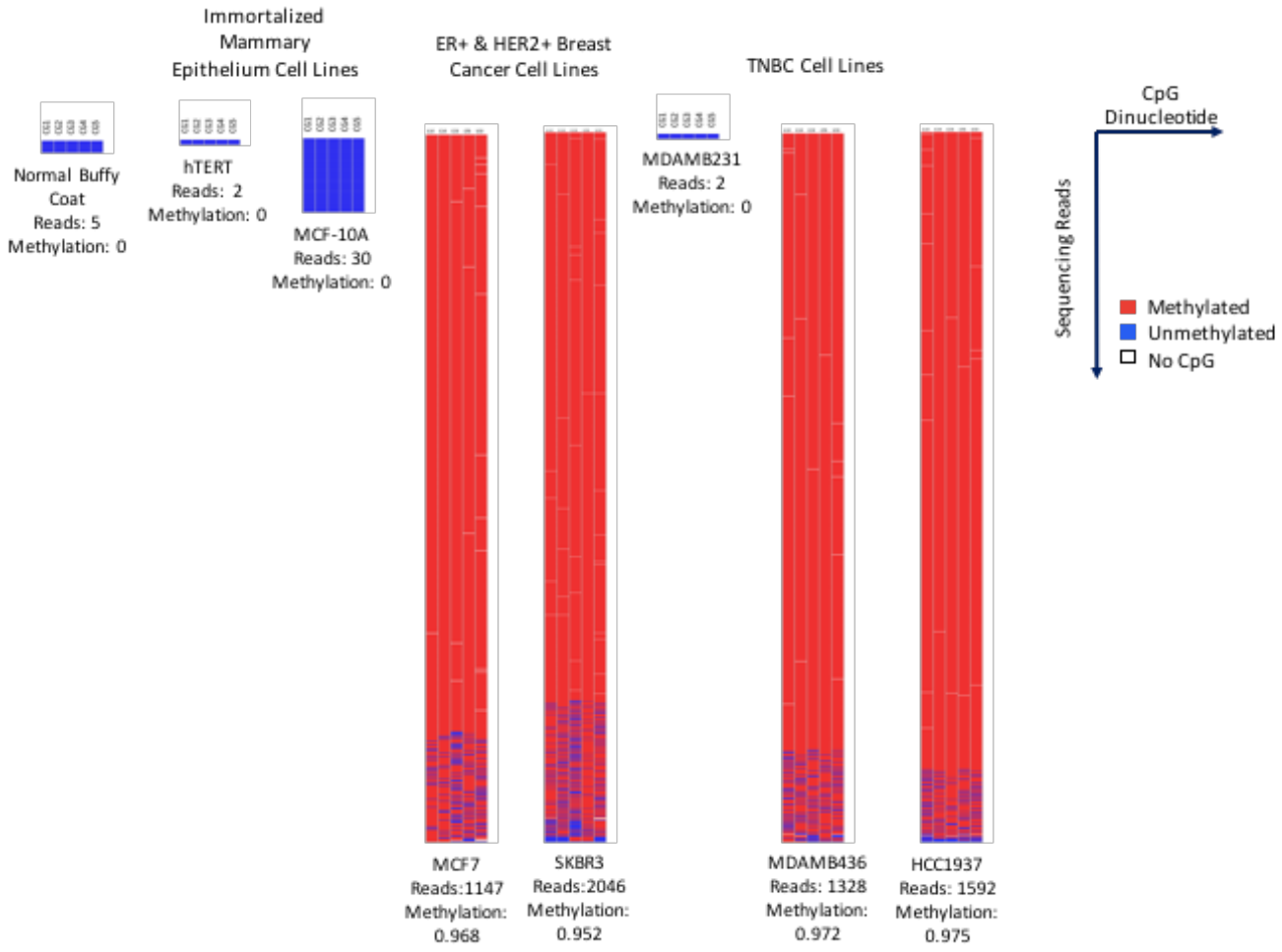


Figure 2.3 Singleplex sequencing results for the probe NFIC-A in cell lines

Shown here are the BiqAnalyzer generated heatmaps of the NGS reads for the probe NFIC-A as analyzed in numerous cell lines. Each column represents a CpG dinucleotide in the amplified region while each row indicates an individual sequencing read. Blue fill indicates that the CpG dinucleotide was not methylated in a particular read while red fill indicates that the CpG dinucleotide was methylated in a particular read. The number of reads and the fraction that are methylated in each probe is indicated below each map. As expected, there is low methylation in the immortalized mammary epithelial cell lines 184-hTERT and MCF-10A as well as the normal buffy coat DNA. Two of three TNBC cell lines (MDA-MB-231, MDA-MB-436 and HCC1937) were methylated demonstrating heterogeneity between individual cancers. Furthermore, MCF7 and SKBR3, an ER+ and HER2+ cell line show methylation, indicating that the probe is not exclusively methylated in TNBC but also other subtypes of cancer. In general, methylation was an all or none phenomena.

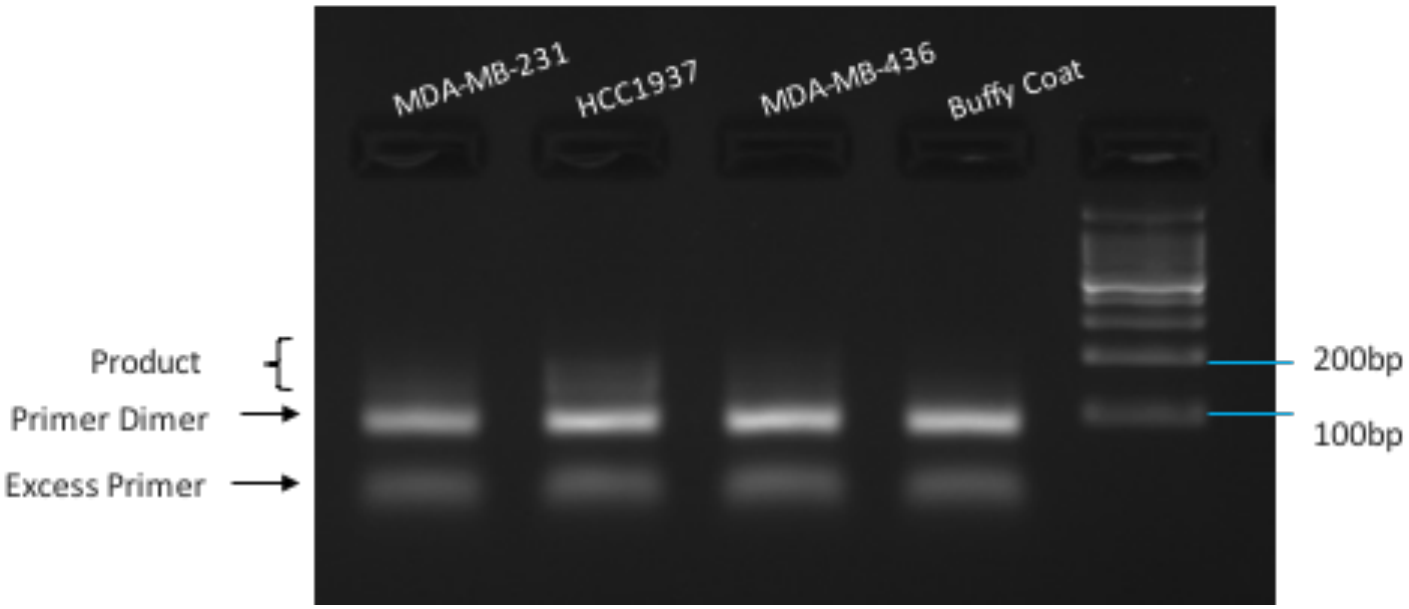


Figure 2.4 Multiplex PCR in TNBC cell lines and normal buffy coat DNA

The optimized primary multiplex PCR in three TNBC cell lines (MDA-MD-231, HCC1937, MDA-MB-436) and buffy coat DNA from a healthy donor (Buffy Coat). Products were run on a 3% agarose gel and visualized using ethidium bromide under UV light. The correct product (*Bracket*, Product), primer dimer (*Arrow*, Primer Dimer) and single stranded primer (*Arrow*, Excess Primer) are shown. DNA size markers in base pairs (bp) are shown on the right. Despite multiple rounds of optimization involving the alteration of conditions such as primer concentration and annealing temperature, there is the appearance of substantially more primer dimer than desired product.

Optimization was conducted by altering primer concentration, melting temperature and number of PCR cycles.

This multiplexed assay was characterized on a collection of 14 primary TNBC tumours from the Ontario Tumor Bank (OTB) (Table 2.1). This was done to examine the distribution and levels of methylation of the various probes in a population of patient samples representative of the TNBC subtype. Between 95.7 and 99.8% of reads were lost to primer dimer with a median of 98.1%. This left samples with between 2,958 and 73,799 total reads across all 86 probes and a median of 30,862 reads. Individual probes had an average of between 0 and 11,593 reads with a median of 28 (Appendix B); however, the reads were distributed quite unevenly between probes (Figure 2.5a). The BiqAnalyzer[135] analysis was filtered so that only reads with 80% or more of CpGs in the region being methylated were exported (Appendix B). This was done to export only those reads considered as methylated. 80% methylation was set as the cut off for identifying methylated reads due to our previous observation that reads are either almost entirely methylated or unmethylated across a given probe. 80% rather than 100% was chosen as the cut off as some regions consistently had one or two umethylated CpGs. A tumour was deemed to give a positive signal for a given probe if it generated more than one methylated read for that probe. These results are output from BiqAnalyzer as a heatmap from which it can be observed that each tumour sample is positive for between 17 and 62 of the 86 probes in the assay (Figure 2.5b). This demonstrates that the mDETECT assay is able to recognize a wide range of tumours with a substantial number of probes reporting for each one.

Table 2.1 Ontario Tumour Bank TNBC Tumour Characteristics

Donor	Breast Cancer Subtype	Age at Sample Collection (years)	Grade
A02354	TNBC	35-39	III
B02275	TNBC	55-59	II
A02282	TNBC	35-39	III
A02338	TNBC	55-59	III
A02397	TNBC	30-34	III
D03248	TNBC	70-74	III
E00700	TNBC	70-74	III
E00931	TNBC	70-74	III
E01003	TNBC	40-44	III
E01672	TNBC	45-49	III
E01684	TNBC	40-44	III
E02243	TNBC	50-54	III
F01103	TNBC	70-74	X
F01172	TNBC	55-59	X

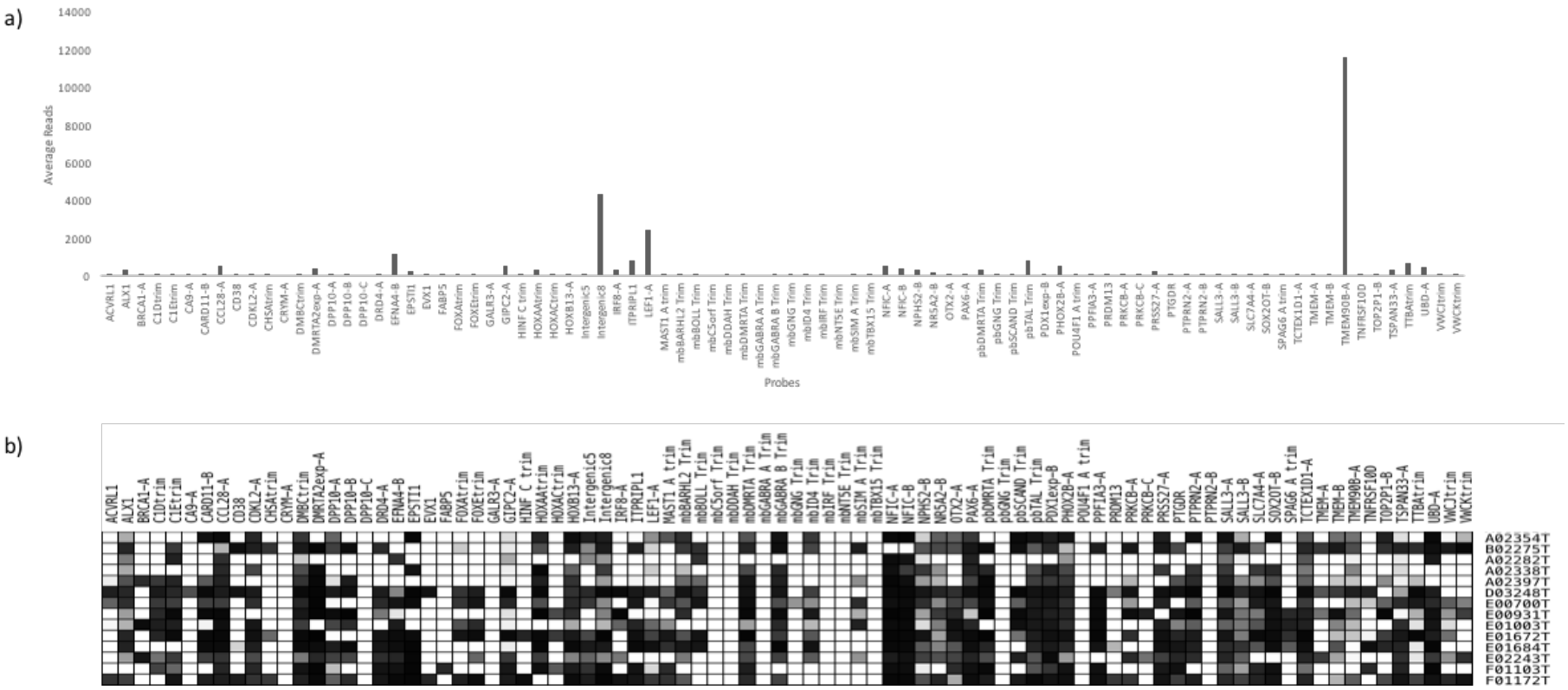


Figure 2.5 mDETECT analysis of 14 TNBC tumours.

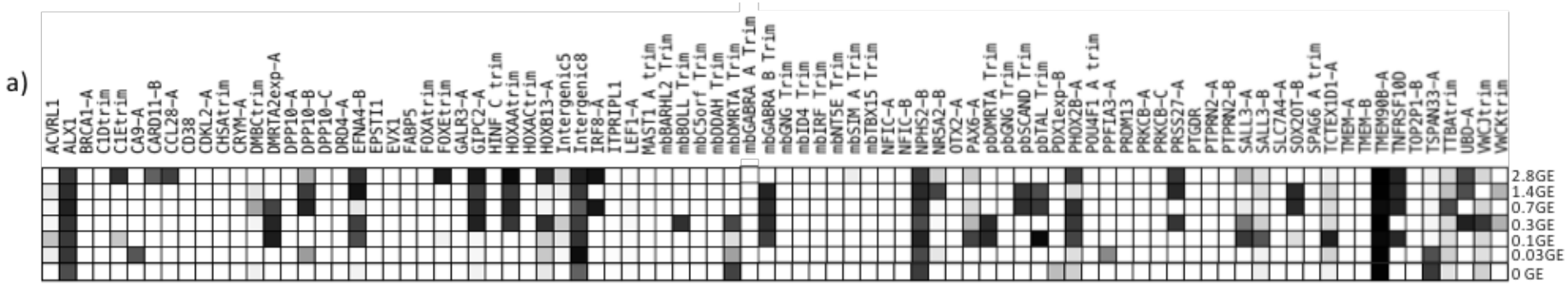
a) The average number of reads (Average Reads) obtained from NGS of all 14 TNBC tumour samples are shown for each of the 86 probes (Probes) used in multiplex PCR. Many probes have very few reads while a select few have dominated the reaction. b) The BiqAnalyzer results displayed in a heatmap representing those reads exhibiting over 80% methylation for each probe in each of the 14 TNBC tumours analyzed using mDETECT. Each column represents an individual probe while each row represents a different sample. The darker grey a box is, the more methylated and numerous the reads for that probe were in a given sample. White boxes indicate that no reads were obtained for a sample however, there is no distinction between samples in which no methylated reads were obtained and those for which no reads at all were obtained. Between 17 and 62 probes were reporting for each sample.

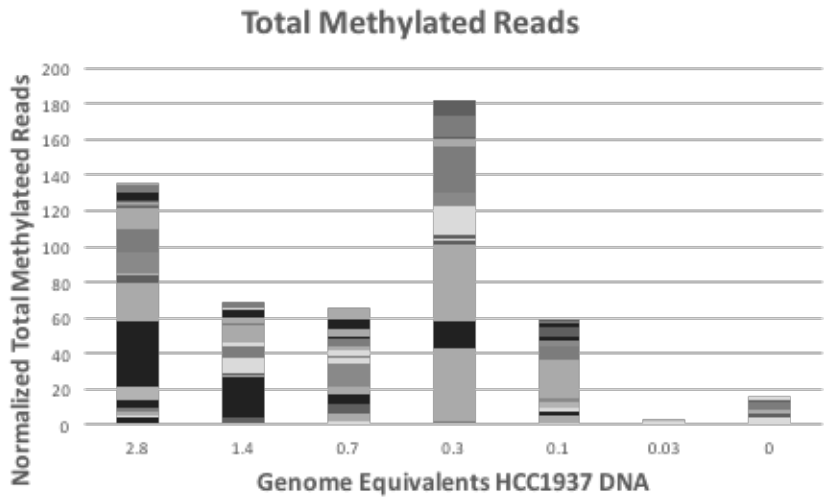
2.4.3 The Sensitivity of the mDETECT Assay

The mDETECT assay was tested for its ability to detect low levels of tumour DNA in a high background of normal cell DNA, by carrying out a dilution experiment. In this experiment, bisulfite converted DNA from the TNBC cell line HCC1937 was diluted in bisulfite converted buffy coat DNA from a healthy donor to reproduce the extreme dilution of ctDNA in cfDNA typical of patient blood samples[125]. As described in the materials and methods, this was done such that quantities from 3 down to as low as 0.03 genome equivalents of TNBC cell line DNA was present in a background of 3300 genome equivalents of normal buffy coat DNA. Additionally, the random distribution of individual molecules of DNA can be seen at the lower concentrations of DNA, representing the fact that less than a genome equivalent of DNA is present. The results of this experiment indicate that much less than a genome equivalent of DNA can be detected against a background of normal DNA that is 20,000 times more abundant (Figure 2.6a). While there were probes reporting methylated reads below this level, they were not distinguishable from the background observed in normal buffy coat DNA. The presence of cell line DNA could be differentiated from normal DNA down to 0.1 genome equivalences based on normalized total methylated reads or on the number of signalling probes (Figure 2.6b & c).

2.4.4 Optimizing the mDETECT Assay

To overcome the unequal amplification of probes that was previously observed as well as to reduce the large quantity of primer dimer present, an expanded multiplex protocol was developed. Prior to this optimization, those of the 86 probes previously assessed that did not perform optimally were eliminated. This involved the removal


a)
ACVRL1
ALX1
BRCA1-A
C1Dtrim
C1Etrim
CA9-A
CARD11-B
CCL28-A
CD38
CDKL2-A
CHSAtrim
CRYM-A
DMBCtrim
DMRTA2exp-A
DPP10-A
DPP10-B
DPP10-C
DRD4-A
EFNA4-B
EPSTI1
EVX1
FABP5
FOXAtrim
FOXEtrim
GALR3-A
GIPC2-A
HINF C trim
HOXAAtrim
HOXACtrim
HOXB13-A
Intergenic5
Intergenic8
IRF8-A
ITPRIPL1
LEF1-A
MAST1 A trim
mbBARHL2 Trim
mbBOLL Trim
mbC5orf Trim
mbDDAH Trim
mbDMRTA Trim
mbGABRA A Trim
mbGABRA B Trim
mbGNG Trim
mbID4 Trim
mbIRF Trim
mbNT5E Trim
mbSIM A Trim
mbTBX15 Trim
NFIC-A
NFIC-B
NPHS2-B
NR5A2-B
OTX2-A
PAX6-A
pbDMRTA Trim
pbGNG Trim
pbSCAND Trim
pbTAL Trim
PDX1exp-B
PHOX2B-A
POU4F1 A trim
PPFIA3-A
PROM13
PRKCB-A
PRKCB-C
PRSS27-A
PTGDR
PTPRN2-A
PTPRN2-B
SALL3-A
SALL3-B
SLC7A4-A
SOX20T-B
SPAG6 A trim
TCTEX1D1-A
TMEM-A
TMEM-B
TMEM90B-A
TNFRSF10D
TOP2P1-B
TSPAN33-A
TTBAtrim
UBD-A
VWCJtrim
VWCKtrim
2.8GE
1.4GE
0.7GE
0.3GE
0.1GE
0.03GE
0 GE


b)
Total Methylated Reads
Normalized Total Methylateed Reads
200
180
160
140
120
100
80
60
40
20
0
2.8
1.4
0.7
0.3
0.1
0.03
0
Genome Equivalents HCC1937 DNA

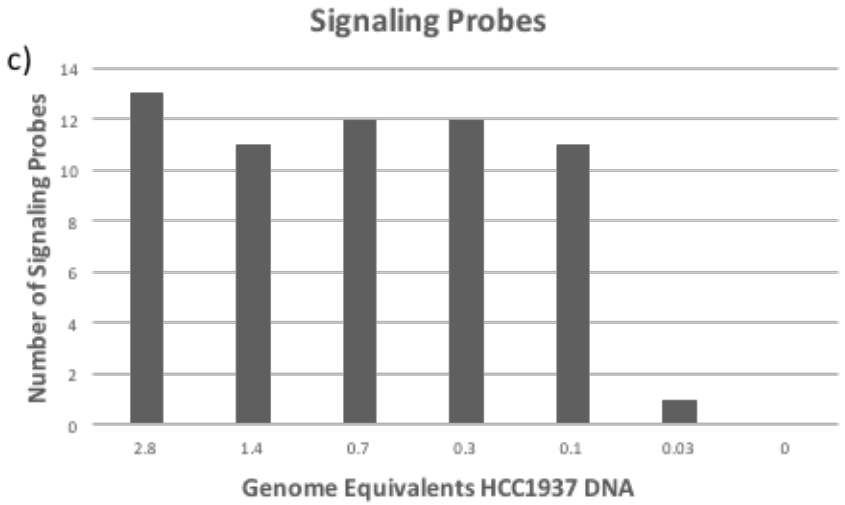

c)
Signaling Probes
Number of Signaling Probes
14
12
10
8
6
4
2
0
2.8
1.4
0.7
0.3
0.1
0.03
0
Genome Equivalents HCC1937 DNA

Figure 2.6 mDETECT Sensitivity

To assess the sensitivity of the mDETECT assay quantities ranging from 2.8 genome equivalents (GE) down to 0.03 GE of bisulfite converted DNA from the TNBC cell line HCC1937 were diluted into a constant background of bisulfite converted normal buffy coat DNA to assess how much ctDNA could be detected in a high background of normal DNA. a) The heatmap shown here was constructed by BiqAnalyzer to represent those reads exhibiting over 80% methylation for the 86 probes used to analyze the sensitivity of mDETECT. Each column represents an individual probe while each row represents a different sample. The darker grey a box is, the more methylated and numerous the reads for that probe in a given sample were. White boxes indicate that no reads were obtained for a sample however, there is no distinction between samples in which no methylated reads were obtained and those for which no reads at all were obtained. The bottom row shows the results for normal buffy coat DNA alone, therefore representing the background signal from the assay (0 GE). Samples containing as low as 0.1 genome equivalents show reporting probes that distinguish them from the background signal, indicating that mDETECT can identify as little as 0.1GE in a background of normal DNA 20,000 times higher in concentration. Additionally, there are particular probes that are not consistent from sample to sample, reflecting the extremely low quantity of DNA that is present and thus the possibility that certain regions of the genome were simply not available for interrogation in certain reactions. b) The total normalized methylated reads for all probes in a given sample are shown (Normalized Total Methylated Reads) for each dilution (2.8 to 0.03 Genome Equivalents HCC1937 DNA) or no tumour DNA (0). Samples with 0.1GE or greater HCC1937 DNA show substantially more normalized methylated reads across all probes than normal buffy coat DNA alone. c) The total number of probes with 2 or more 80% methylated reads reporting for each sample (Number of Signaling Probes) is also substantially higher for all samples with 0.1GE or more of HCC1937 DNA present than the background signal given off by normal buffy coat DNA alone

of those probes that failed to generate NGS reads in all samples or that generated substantially more reads than all of the other probes, thus dominating the sequencing. This resulted in the retention of 68 probes.

To address the unequal distribution of the reads for each probe, they were split into groups based on the efficiency in which they amplify their product. This was done so that probes that amplify with similar efficiency are grouped together to reduce the risk of one primer set outperforming another. The dynamics of amplification of each primer set was determined by qPCR. This was done using 2 different TNBC cell lines, HCC1937 and MDA-MB-436. Following qPCR, these probes were rank ordered based on the lowest C_t in either cell line and grouped into multiplex mixes containing either 4 or 8 probes with the most similar C_t's, for a total of 9 different miniplex mixes (Appendix C). The analysis was carried out in two cell lines since, as a result of their methylation biased nature, some probes fail to amplify unmethylated DNA. By examining two cell lines, both of which were shown to be highly methylated across the majority of probes, the risk of obtaining a falsely high C_T value due to lack of methylation is less likely. In addition to allowing primers of similar efficacy to be amplified together, resulting in more equal amplification, this procedure also decreases the overall levels of primer in each reaction therefore reducing primer-dimer formation.

Dividing the DNA from a blood sample into 9 different reactions would reduce the overall sensitivity of the assay due to dilution of the target DNA. This is particularly

relevant at low concentrations of DNA where individual molecules of ctDNA may not end up in the reaction with the corresponding primer set. To overcome this limitation, a pre-PCR step was added prior to the 9 miniplex reactions where linear amplification of the DNA was carried out using only the reverse primers for each probe as this is the only primer recognizing bisulphite converted DNA. Ten rounds of linear amplification are carried out to avoid the accumulation of primer-dimer, following which excess primer is eliminated using magnetic bead purification. This material was then used in a primary PCR, split between the 9 miniplex reactions. Based on visualization on an agarose gel, primer-dimer formation was greatly reduced and individual PCR reactions clearly exhibited multiple products (Figure 2.7). When samples prepared in this manner were sequenced, this expanded multiplex method demonstrated superior read distribution in comparison with the previously established multiplex protocol (Figure 2.8a,b) for individual probes as well as reduced primer-dimer formation. Although overall read distribution was improved, there is still variation in the read distribution between the probes in each individual miniplex reaction (Figure 2.8b).

2.4.5 mDETECT Characterization of TNBC Patient Plasma Samples

cfDNA was extracted from plasma samples from a previously characterized cohort of metastatic TNBC patients[102] and subjected to the mDETECT test as outlined above. cfDNA from a cohort of normal individuals was also processed in the same manner. Following NGS, both total and methylated read counts were determined for each probe. Reads in which 80% of the CpG sites in the region were methylated were considered as methylated reads. BiqAnalyzer compiled these results as a heatmap

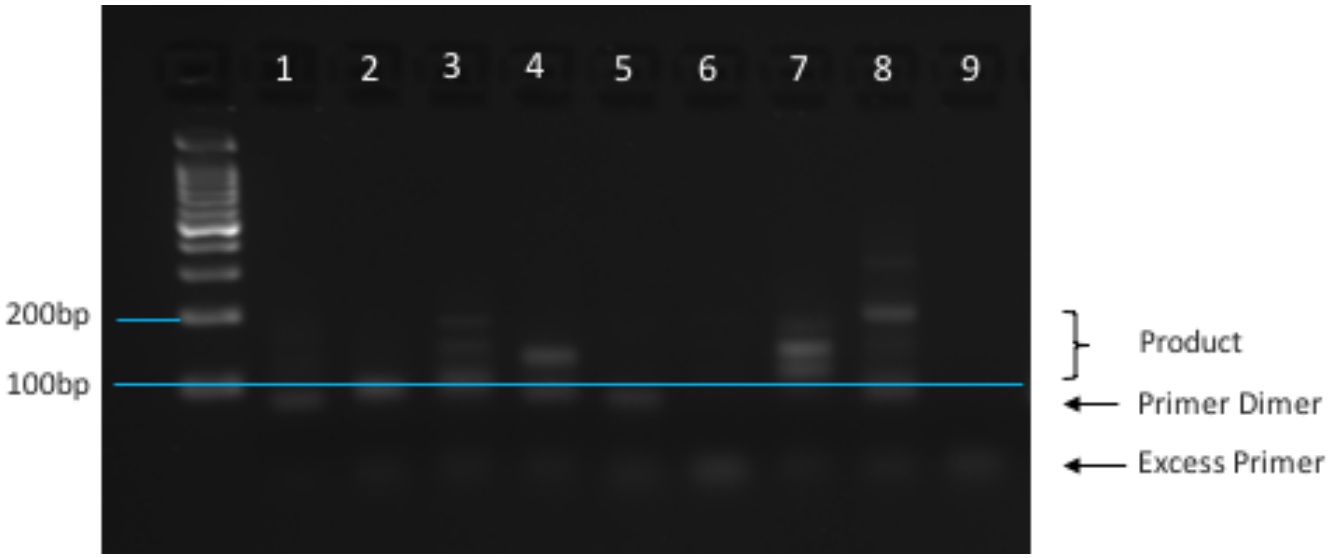


Figure 2.7 Expanded multiplex PCR in HCC1937 DNA

The expanded multiplex PCR protocol outlined in the materials & methods was optimized using bisulfite converted DNA from the TNBC cell line HCC1937. Shown here are the products of primary PCR in 9 miniplex reactions run on a 3% agarose gel and visualized using ethidium bromide under UV light. The correct product (*Bracket*, Product), primer dimer (*Arrow*, Primer Dimer) and single stranded primer (*Arrow*, Excess Primer) are shown. DNA size markers in base pairs (bp) are shown on the left. The primer dimer is reduced compared to previous trials and completely absent in some reactions (6 & 9). Additionally, distinct product bands can be observed for some samples (3, 7, 8). No product can be observed for miniplexes 2, 5 and 9. Although this product cannot be visualized on a gel, subsequent NGS showed that product was in fact present.

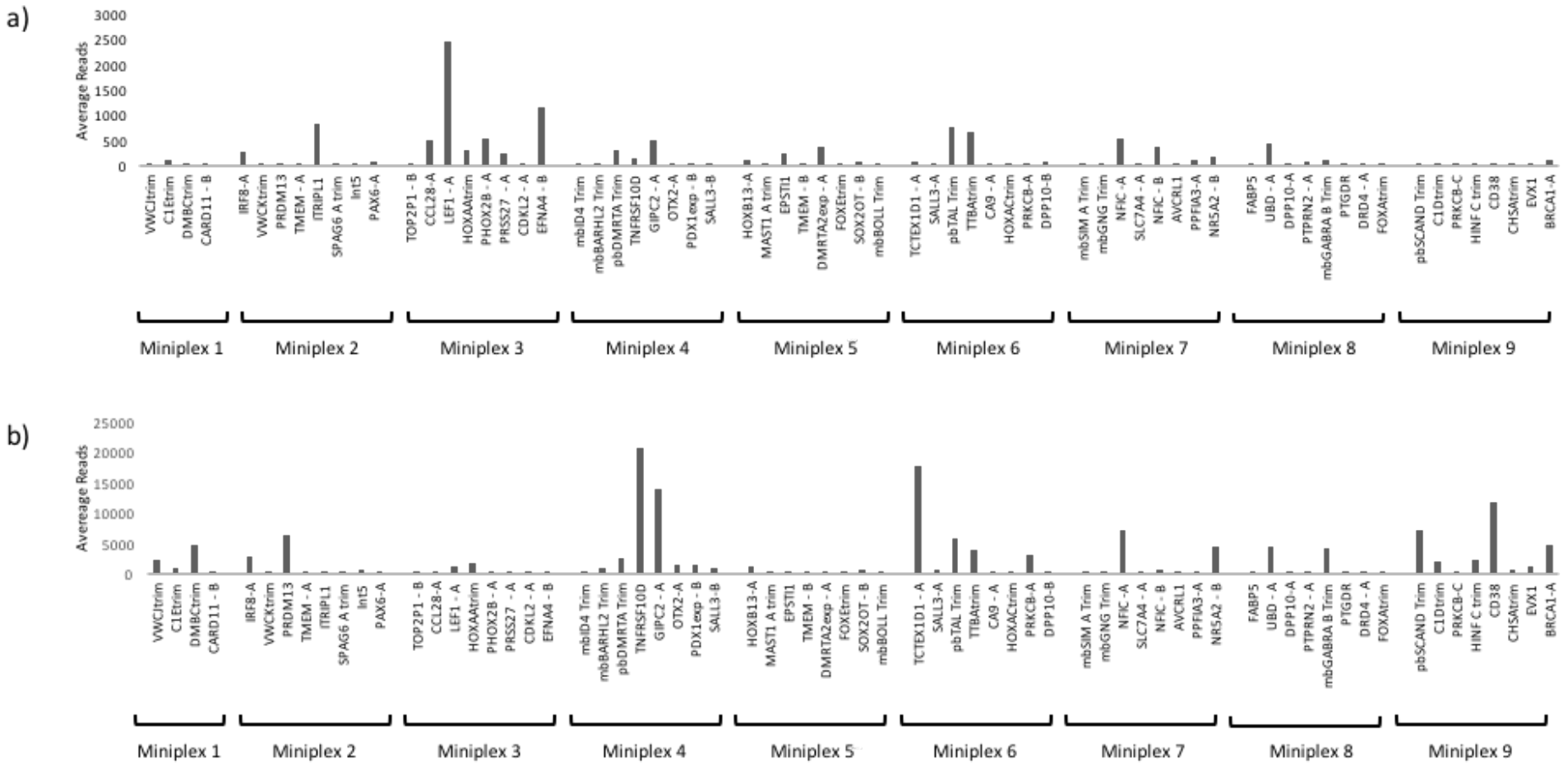


Figure 2.8 Read distribution multiplex versus expanded multiplex

The average reads for the 68 probes included in the expanded multiplex protocol for mDETECT are grouped according to their miniplex grouping for both a) the normal multiplex reaction carried out in TNBC tumors and b) the expanded multiplex reaction in TNBC patient plasma samples. While the read distribution remains uneven overall and within the miniplex groups for the patient plasma samples, there is an observable improvement in the number of probes getting reads, especially within miniplex 1, 4, 6 and 9.

showing that all samples are methylated for between 13 and 62 probes (Figure 2.9). More methylation than was expected was observed in the control samples with one sample exhibiting methylation of 40 probes, including the *BRCA1* promoter.

It was observed that between 66.4 and 94.5% of reads were lost to primer dimer with a median of 85.3%. This left patient samples with between 97,349 and 564,858 total reads across all 68 probes and a median of 231,732 reads. Individual probes for patient samples had an average of between 0 and 47,125 reads with a median of 893 (Appendix D). These numbers were lower when examining cfDNA from healthy controls with between 72.8 and 99.9% of reads were lost to primer dimer and a median of 97.3%. This left control samples with between 350 and 313,164 total reads across all 68 probes and a median of 47,665 reads. Individual probes for control samples had an average of between 0 and 17,409 reads with a median of 43 (Appendix D).

The methylated reads for each probe, those that showed over 80% methylation (Appendix D), were corrected by the fraction of reads left following Galaxy and BiqAnalyzer filtering for a sample compared with the reads left following filtering with Galaxy only. This ratio represents the fraction of the total reads for a sample that are real products as opposed to primer dimer. Individual probe reads were corrected by this number because of our observation that with less amplifiable product being present, there is greater production of primer dimers in the PCR. As such, this correction acts to normalize the number of methylated reads generated by a sample. By summing the corrected methylated reads for each probe of a given sample, a Cumulative Methylation Index (CMI) was determined for each sample. For patient samples, this CMI ranged from approximately 3000 up to just over 115,000 (Table 2.2). For control samples, this metric

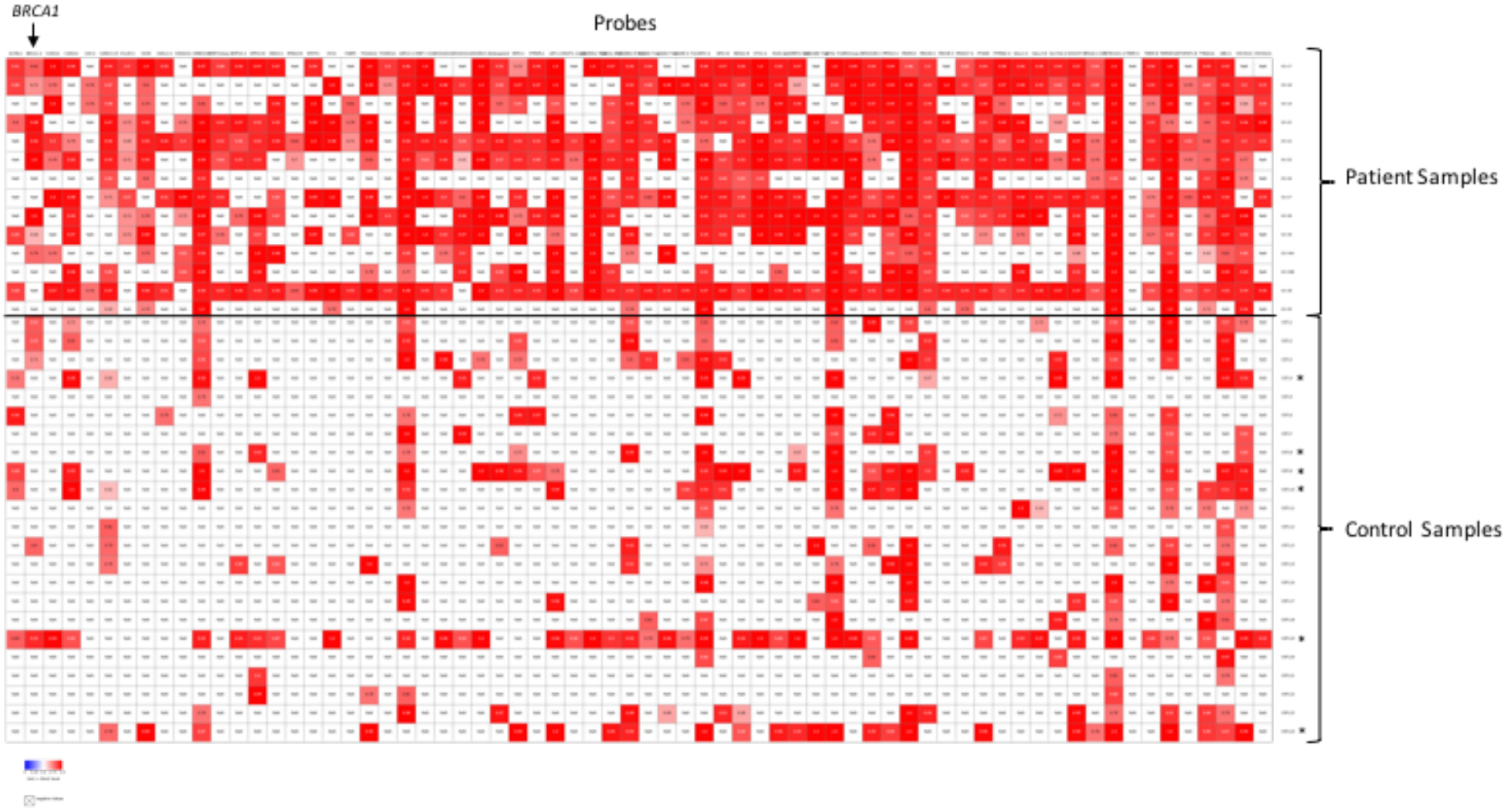

BRCA1
Probes
Patient Samples
Control Samples

Figure 2.9 mDETECT analysis of patient plasma DNA

The DNA from 14 TNBC patient plasma samples (Patient Samples) as well as 23 plasma samples from healthy volunteers (Control Samples) were analyzed using mDETECT. The BiqAnalyzer output representing those reads exhibiting over 80% methylation is displayed here as a heatmap. Each column corresponds to a given probe while each row represents a different sample. As noted by the legend, the darker red a box is, the more methylated that probe is in the corresponding sample. A white box indicates that no methylated reads were present for a given probe. Generally, the overall number of probes that yielded methylated reads is much higher for patient samples when compared with controls however, there are some controls displaying a large number of methylated probes as well (*). The probe for BRCA1 (*Arrow, BRCA1)* was found to be methylated in some controls, indicating the possible presence of early stage cancer or high cancer risk in these individuals.

Table 2.2 Cumulative Methylation Index (CMI) for patient and control plasma samples

	Sample	CMI – 86 Probes	Median – 86 Probes	CMI – 36 Probes	Median – 36 Probes
Patient Samples	03-17	161,103	71,035	128,423	35,126
	03-18	104,796		74,434	
	03-19	31,797		17,011	
	03-22	93,960		50,987	
	03-23	279,544		225,633	
	03-25	217,058		166,589	
	03-26	22,454		21,270	
	03-27	128,704		91,991	
	03-28	48,110		27,850	
	03-30	41,952		25,159	
	03-34A	17,157		8,327	
	03-34B	38,232		15,230	
	03-38	106,450		42,401	
	03-39	12,319		10,963	
Control Samples	Control 1	441	1,692	436	681
	Control 2	9,754		6,347	
	Control 3	2,246		1,760	
	Control 4*	18,557		2,744	
	Control 5	0		0	
	Control 6	6,909		148	
	Control 7	67		64	
	Control 8*	11,099		746	
	Control 9*	36,164		5985	
	Control 10*	15,111		7,317	
	Control 11	44		2	
	Control 12	5		4	
	Control 13	1,288		86	
	Control 14	6,657		6,580	

	Control 16	1,692		1,098	
	Control 17	682		681	
	Control 18	821		476	
	Control 19*	16,764		6,333	
	Control 20	69		67	
	Control 21	4		0	
	Control 22	2		0	
	Control 23	4,302		4,197	
	Control 24*	17,787		9,575	

* Indicates a highly methylated control

ranged from essentially 0 up to around 3600 (Table 2.2). While the CMIs for control samples were clearly substantially lower than patient samples, 6 controls were labelled as having high levels of methylation. These included controls 4, 8, 9, 10, 19 and 24 (Table 2.2, Figure 2.9).

CMI was used to perform ROC analysis to assess the sensitivity and specificity of mDETECT. The assay was found to have a sensitivity of 93%, a specificity of 86% and an area under the ROC curve (AUC) of 0.98 (Figure 2.10a). However, it was noted that not all probes seem to be equally informative and that some appear to show relatively high levels of methylation in the control samples. As such, the 68 probes were assessed to determine which are performing most optimally by identifying those probes that show a greater than two-fold difference in the average corrected methylated reads compared with the six most methylated controls as assessed by CMI (controls 4, 8, 9, 10, 19 and 24). This acted to identify those probes that are, on average, substantially more methylated than controls. This resulted in the identification of 36 optimally performing probes. CMI was determined using only these 36 probes for each sample and these values are shown in table 2.2. The modified CMIs were used to perform ROC analysis again. The results of this analysis were a sensitivity of 100%, a specificity of 95.7% and an AUC of 0.99 (Figure 2.10b).

This cohort was previously analysed using NGS to detect mutant *TP53* in cfDNA[102]. This study also examined the circulating tumour cells (CTCs) for each patient. By assaying for TP53 mutations, 81% of patients with a previously identified mutation in their tumour were found to have a mutation in the cfDNA[102]. Additionally, CTCs were only identified in the plasma of 70%[102]. In contrast, mDETECT was able to identify all

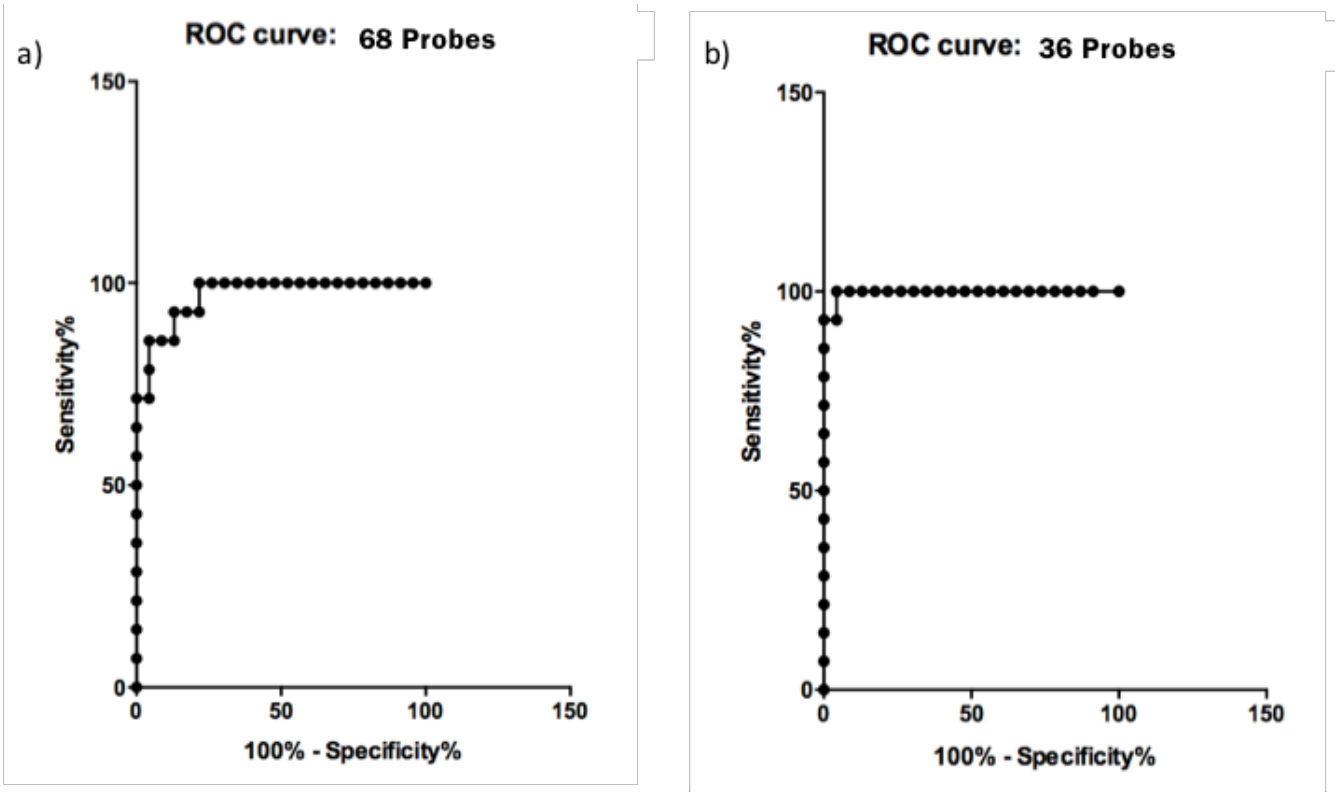


Figure 2.10 ROC Curve Analysis of mDETECT

Receiver operating characteristic (ROC) curve analysis was carried out according to the CMI for each of the 14 patient plasma DNA samples compared with the 23 control samples. This was done with the CMI calculated according a) all 68 probes included in the assay and b) the subset of the 36 best performing probes. When taking all 68 probes into account the AUC is 0.98 (sensitivity = 93%, specificity = 86%) while when only 36 probes are included the AUC is 0.9969 (sensitivity = 100%, specificity = 95.7%). This demonstrates that when the assay is refined to using only those probes that perform optimally, it performs with extremely high accuracy.

TNBC patients, representing a great improvement over single mutation and CTC based tests (Table 2.3). Additionally, it was shown that the ctDNA level determined previously[102] correlated with CMI determined by mDETECT with an R^2 value of 0.775 when all 86 probes were included in the CMI calculation and 0.798 when only the 36 best performing probes were included (Figure 2.11). Two samples were removed from this analysis due to high levels of cfDNA determined previously (Table 2.3). This is because at high levels of cfDNA it is unlikely that mDETECT will perform in a linear fashion, meaning that the CMI determined will not be proportional to the amount of DNA present.

2.5 Discussion

The detection of circulating tumour DNA using mutations is restricted by multiple factors such as the limited and variable spectrum of mutations present in a given class of tumours[2,106] compounded by the difficulty of detecting these changes against a high background of normal DNA[97]. Because of this, promoter hypermethylation has emerged as a promising alternative[118,119]. This is due to the fact that promoter hypermethylation events are both numerous as well as consistent across a particular tumour type[121]. Because of this, we have developed a sensitive NGS based method for the detection of tumour specific promoter hypermethylation in ctDNA. In this study mDETECT has been adapted specifically for the detection of TNBC and has been optimized in order to be able to simultaneously detect cancers arising due to BRCA1 mutations, which we have found to possess a slightly different hypermethylation pattern than the general population of TNBC tumours.

mDETECT incorporates a variety of features to optimize the detection of ctDNA at extremely low levels in the blood. Foremost is the detection of multiple CpG

Table 2.3 *TP53* status, circulating tumour cells (CTC), cfDNA, ctDNA and CMI for TNBC patient plasma samples

	p53 Status	CTCs	CTC Signal	cfDNA1 (ng/ml)	ctDNA1 (ng/ml)	[ctDNA] Signal	CMI – 86 Probes	CMI – 36 Probes	CMI Signal
03-17	+	0	-	22.3	6.467	+	161,103	128,423	+
03-18	+	10	+	14.2	6.39	+	104,796	74,434	+
03-19	+	0	-	11	0.55	+	31,797	17,011	+
03-22	+	95	+	71.7	0	-	93,960	50,987	+
03-23	+	12	+	20	10.8	+	279,544	225,633	+
03-25	+	0	-	34.8	6.96	+	217,058	166,589	+
03-26	+	0	-	12.8	0.256	+	22,454	21,270	+
03-27	+	19	+	153.9	7.0794	+	128,704	91,991	+
03-28	+	0	-	10.7	2.568	+	48,110	27,850	+
03-30*	+	38	+	172.5	10.35	+	41,952	25,159	+
03-38*	+	0	-	233.9	107.594	+	106,450	42,401	+
03-34A	-	1	-	*	*	*	17,157	8,327	+
03-34B	-	-	-	*	*	*	38,232	15,230	+
03-39	-	0	-	*	*	*	12,319	10,963	+

+ indicates the presence of a mutation or signal, - indicates the absence, * indicates a particular metric was not assessed for a particular sample

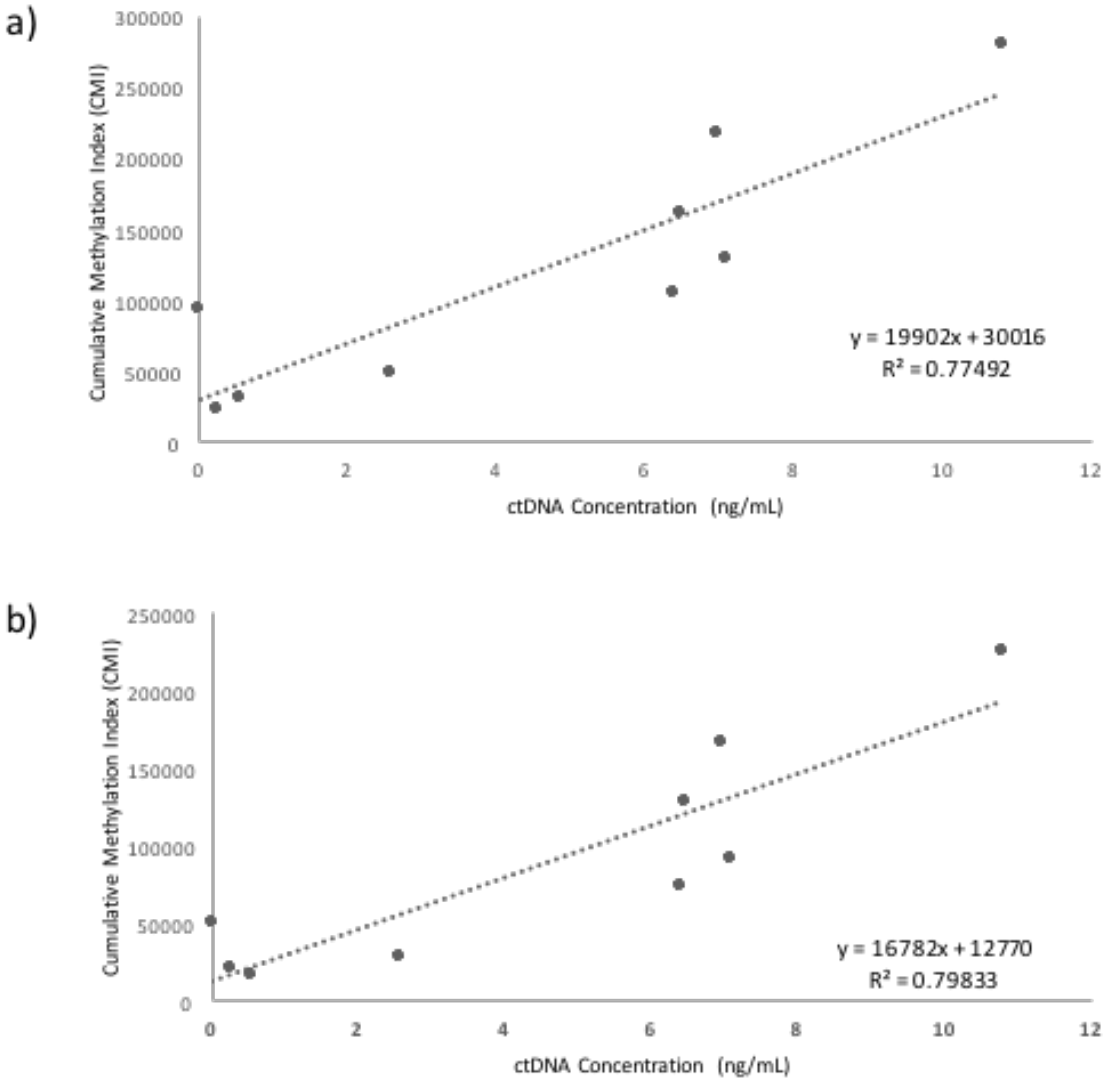


Figure 2.11 Correlation between ctDNA concentration and CMI

The output of the mDETECT assay is CMI which reflects the number of corrected methylated reads for all probes for a given sample. Previously, the ctDNA concentration for each sample was determined using a p53 mutation based assay. Using a linear regression, here we show that with both a) the inclusion of all 68 probes in the assay in the CMI calculation as well as b) the inclusion of only the 36 best performing probes, CMI and ctDNA concentration are correlated. Shown on each graph is the equation of each linear trendline as well as the R^2 value, a measure of how well the regression line fits the data. These results indicate that mDETECT may be semi-quantitative in nature, reflecting the amount of ctDNA in the blood. Thus it may have the ability to detect changes in ctDNA concentration over time, a known indicator for response to treatment.

methylation events in a given read as a way of circumventing the error rates associated with NGS, providing for a robust signal detectable against a high background of normal DNA. The high levels of multiplexing that can be achieved by targeted PCR combined with NGS also allows for the detection of low levels of ctDNA when concentrations are less than a single genome equivalent. A common pre-PCR step ensures that tumour-specific methylation will not be missed as a result of a fragment of ctDNA being in a different tube than the primer set corresponding to that region. PCR targets are less than 125 bp consistent with the small fragment size of ctDNA and through the use of methylation "biased" probes, low levels of methylated DNA still lead to the optimal production of tumour specific products. Overall, this has allowed mDETECT to detect levels of ctDNA at least 20 fold lower than mutation based assays (0.005% vs 0.1%[102]).

An essential component of the mDETECT assay is the high level of multiplexing involved, as this is what allows for all probes to be amplified in the presence of all of the available sample. Multiplexing so many primer sets into a single reaction presents two main challenges; primer dimer and unequal amplification of probes leading to an unequal distribution of sequencing reads. Even after numerous rounds of optimization in cell lines, both of these factors remained an issue. Increasing sequencing read depth can improve the issue of unequal read distribution across probes by ensuring that even underrepresented probes are receiving sufficient reads; however, this does less to remedy the challenges presented by primer dimer. This is because NGS favors the sequencing of small products. Because primer dimers are smaller than the probes in the assay, typically around 100bp, these small fragments will be preferentially sequenced and will consume a disproportionate number of reads compared with the longer products of interest.

Using the standard multiplexing protocol that was employed for analysis of the OTB TNBC tumours, between 95.7 and 99.8% of NGS reads were sequences of primer dimer and thus did not represent the products of interest in this assay. Furthermore, there was a massively unequal distribution of reads across the various probes. In an attempt to remedy these problems, the multiplexing protocol was modified to include a low cycle linear pre-amplification followed by 9 miniplex reactions, divided up based on similar C_T values. This enabled an initial amplification in which all probes are amplified in the presence of all available sample, but has the advantage of performing the bulk of the amplification in the presence of lower primer quantities as well as in reactions containing primer sets performing at similar efficiencies.

As predicted, the expanded multiplex protocol was able to decrease primer dimer production and even out read distribution across probes. There has been a substantial improvement in the generation of primer dimer, decreasing from a median of 98.1% initially to a median of 85.3% reads lost to primer dimer when examining ctDNA. Such a decrease was not observed when examining cfDNA from healthy donors. However, a great deal of primer dimer is expected in these samples as the overall lack of methylation results in low product generation from the methylation biased primers and thus more opportunity for the formation of primer dimer.

Despite the steps that have been taken to reduce primer dimer formation, there is still room for improvement. It is likely that the level of primer dimer reduction that can be acheived by modifying PCR protocols has been reached. As such, the next step is to look at the construction of the primers being used. It has been observed that the 3` to 5` exonuclease activity of many DNA polymerases leads to the degradation of PCR

primers[140]. This degradation appears to lead to increased formation of primer dimer[141]. It was observed that the addition of phosphorothioated nucleotides to the 3` end of the primer prevents this exonuclease driven degradation from happening[140]. One study found that by modifying primers in this way, primer dimer production could be decreased from 17% to 6.4%[141]. As such, this provides a promising way in which the primer dimer generated by the mDETECT assay could be further reduced.

mDETECT involves the interrogation of 68 hypermethylated regions of the genome. By assessing these 68 probes, ROC analysis revealed a sensitivity of 93%, specificity of 87% and an AUC of 0.98. However, when the 36 most optimally performing probes were identified and analysis was restricted to this subset of probes, mDETECT was determined to have a specificity of 100%, specificity of 95.7% and an AUC of 0.99. This shows that mDETECT is capable of performing with extremely high accuracy but likely performs optimally with a particular subset of probes. Analysis of a greater number of both patient plasma samples as well as controls will be necessary to determine this optimal group of probes.

There are numerous tests that have been previously been developed to assess tumour specific hypermethylation patterns in ctDNA[82,86,123,142]. However, none involve the number of probes or the robust technique that mDETECT employs, nor have they been able to achieve the sensitivity mDETECT exhibits. Many of the multigene assays developed thus far still only include a few hypermethylated regions[86]. While this does provide improvements on the single gene approach, it is still likely that a large number of tumours will be missed. This is reflected in a study conducted by Hoque *et al.* in which the hypermethylation of four genes were interrogated in the plasma DNA of breast cancer

patients and controls. By defining a positive test results as methylation of at least one of the four genes, the sensitivity of the assay was 62% while the specificity was 87%[86].

Assays that have included additional hypermethylated regions show improvements in the sensitivity and specificity of their assays compared to those with lower numbers. One study, involving an 8 gene panel consisting of candidate tumour suppressor genes, aimed to distinguish between early and late stage breast cancer patients and control samples based on methylation in the blood[123]. This assay was able to achieve a sensitivity of 91.7% and a specificity of 90%[123]. Another study conducted by Fackler *et al.* identified a 10 gene panel, also for the detection of tumour specific promoter hypermethylation in the blood, called cMethDNA[82]. This assay was able to distinguish between metastatic breast cancer patients and normal samples with a sensitivity of 91% and a specificity of 96%[82]. This test, which appears to be the most accurately performing methylated based blood test for the detection of cancer reported in the literature[82], is far less sensitive than mDETECT and has a comparable specificity.

While Fackler *et al.* have been able to develop a relatively sensitive and specific assay based on a 10 gene panel identified for their consistent and high levels of methylation in breast tumours[82], cMethDNA uses a less sensitive approach than mDETECT. This assay uses a quantitative multiplexed methylation-specific PCR (QM-MSP) for the detection of hypermethylation in cfDNA[82]. While this particular assay includes measures to improve sensitivity, such as an initial preamplification step[82,143], qMSP approaches lack resolution in comparison with NGS. qMSP techniques give only a single signal for a given region that contains many CpGs[144]. While this is an improvement when compared with simple MSP assays in which only a yes or no signal

can be obtained[145], it does not rival the resolution obtained by using NGS. NGS enables the assessment of the methylation status of individual CpG dinucleotides. This allows for the assessment of patterns of methylation in a region that can be used to accurately distinguish between tumour and normal tissues and also provide protection against false positive results due to the fact that there are multiple contiguous CpGs in each region, many of which must be methylated to yield a signal.

A further advantage of the mDETECT assay is the subtype-specific approach. mDETECT has the potential to be adapted to any cancer type by identifying hypermethylated regions specific to a certain malignancy. However, by ensuring that regions are identified for a specific type or subtype of cancer, it ensures that the test is as specific as possible. This is particularly important for malignancies like breast cancer that are known to be extremely heterogeneous and include numerous well characterized subtypes likely of having distinct aberrant methylation patterns[2,27]. Many of the methylation based liquid biopsies created previously have examined breast cancer as a single disease as opposed to one composed of subtypes[86,123]. Additionally, while cMethDNA has been designed to ensure coverage of both ER+ and ER- tumours[82], the level of subtyping does not match that of the specific subdivisions employed by mDETECT including basal-like and *BRCA1*-associated tumours.

The study presented here shows that the sensitivity of mDETECT is better than its previously developed counterparts and that the specificity is comparable[82,86,123]. However, this is only true when a subset of the probes in mDETECT are assessed. This is due to the fact that 6 of the controls are methylated for a number of probes. There are numerous explanations for this occurrence such as age-related methylation changes[146,147].

However, it is also possible that these individuals may have had early stage cancer at the time the samples were given or were at a heightened risk of developing it. First, it is possible we are detecting ctDNA in undiagnosed patients. However, it is also possible we are detecting aberrant methylation changes in contaminating DNA derived from blood cells. This is a possibility due to evidence suggesting that particular patterns of aberrant hypermethylation are detectable in the DNA of peripheral blood cells in those individuals at risk of developing cancer[148,149]. An additional rationale for predicting that these controls may have or be at risk of cancer is that two of these controls show methylation in the *BRCA1* promoter. Methylation of this region in healthy cells is unusual and has been shown to be detectable in the peripheral blood cells of individuals at a heightened risk of breast cancer development[150]. Interestingly, *PAX6* was found to be methylated in two of the highly methylated controls observed in this study (controls 19 and 24). This gene is methylated in both advanced breast cancer as well as ductal carcinoma in situ (DCIS), a non-invasive breast abnormality thought to be the earliest stage of breast cancer[151]. This further supports our hypothesis that mDETECT is identifying early stage cancer in these methylated controls. Unfortunately, we have little information regarding the people from which the control samples were taken and have no way of validating our hypothesis. However, as we have shown that the mDETECT assay is able to detect extremely small amounts of ctDNA in a high background of normal DNA, it is likely that we are detecting low quantities of ctDNA in individuals who are in the early stages of cancer development.

In addition to the high sensitivity of mDETECT to identify individuals with TNBC, it also has a high sensitivity in its ability to detect extremely low levels of ctDNA

within a high background of normal cfDNA. This was demonstrated by the ability of the assay to detect as low as little as 0.1 genome equivalents of cancer cell line DNA in a background of normal buffy coat DNA 20,000 times greater. The overall quantity of ctDNA as well as the ratio of ctDNA to total cfDNA in the blood is highly variable between different patients. This makes it difficult to estimate what limit of detection is required for an assay to be able to reliably detect its presence. One study found that, in metastatic breast cancer patients, ctDNA represented between 2 and 70% of overall cfDNA which corresponded to between 48 and 648,030 genome equivalents per millilitre of plasma[102]. Another study that included both patients with primary and metastatic colorectal cancer found that ctDNA comprises between 0.005 and 11.7% of cfDNA[95]. mDETECT has a limit of detection that would enable detection of ctDNA even at the lower end of the quantity estimated by the above studies. This is a promising observation in terms of its ability to detect early stage malignancies as well as to track the progression of disease.

The cfDNA samples used in this study had previously been assessed using a liquid biopsy aimed at identifying *TP53* mutations in plasma DNA[102]. This assay was designed to detect patient specific *TP53* mutations in cfDNA through an NGS approach using the Access Array microfluidic system[102]. This study also examined the ability of circulating tumour cells (CTC) to enable the detection of these patients. *TP53* mutations could only be identified in the plasma of 81% of patients who were found to have a mutation in their tumour[102]. Additionally, CTCs were only identified in the plasma of 70% of patients[102]. Because of this, we anticipated that with the superior sensitivity shown by mDETECT, it would predict the presence of cancer in these patients in a more

accurate fashion than either the presence of a *TP53* mutation or CTC count. In contrast to these tests, mDETECT was able to detect all patients. This demonstrates that mDETECT detects much lower levels of cancer than either a single mutation-based assay or analysis of CTCs. Furthermore, there were two patients in this cohort who were unable to be assessed at all by the *TP53*-based assay as they did not possess a mutation[102]. As a result, these patients that could not even be assayed previously were identified by mDETECT. Lastly, the concentration of ctDNA previously determined shows good correlation with the CMI determined by mDETECT. This highlights the possibly quantitative nature of mDETECT and its potential to monitor changes in the quantity of ctDNA in the blood over time and thus a patient's response to chemotherapy or risk of relapse.

While the majority of the hypermethylated regions in the mDETECT assay have not been previously identified in the literature as being associated with breast cancer, some have been. Additionally, many are associated with other malignancies. One study in particular identified many of the genes in mDETECT as being hypermethylated in HER2+ breast cancer including *IRF4, GABRA4, C1orf114, CDKL2, HOXA9, VWC2, POU4F1, DMRTA2, SALL3, PRKCB, PAX6* and *MAST* [152]. Many of these genes were also identified by other studies as being associated with breast cancer[142,151,153–155]. In addition to those genes that have been found previously to be hypermethylated in breast cancer, numerous genes in mDETECT are hypermethylated in other cancer types. These include *SPAG6*[156], *HOXB13*[157], *PPFIA3*[158], *PTGDR*[159], *CD38*[160], *EVX1*[161], *TAL1*[162] and *CA9*[163].

mDETECT provides a highly sensitive and specific way in which to detect tumour-specific hypermethylation patterns in ctDNA. This study demonstrates that ability

of this assay to detect extremely small amounts of ctDNA in the blood for essentially all members of a cohort of TNBC patients. Furthermore, it appears as though mDETECT has the ability to do this in a semi-quantitative manner. While this study outlines the design and validation of mDETECT in a TNBC population, this assay has the potential to be adapted to any malignancy. Furthermore, this test represents an easy and non-invasive way to assess a patient's tumour. These findings highlight the potential for mDETECT to be used in a number of of settings including the early detection of cancer, assessing patient response to chemotherapy and monitoring for relapse. The results presented here require validation in a separate, larger cohort, but if our findings are substantiated, mDETECT has the potential to drastically improve patient outcomes across numerous malignancies.

Chapter 3

Identification of Potential Markers of Treatment Response in Models of Residual and Treatment Naïve Triple Negative Breast Cancer

3.1 Abstract

The ability to predict which patients with TNBC will respond to chemotherapy remains a major challenge in the treatment of these malignancies. Recently the mDETECT assay was developed for the detection of multiple regions of TNBC specific promoter hypermethylation in the blood and shown to accurately distinguish between patient and control plasma DNA samples. This assay was assessed for the predictive value of these regions using DNA extracted from patient derived xenograft (PDX) mouse models that represent either chemoresistant or chemotherapy naïve disease. It was determined that *TNFRSF10D* ($p = 0.025$) and *FOXA3* ($p = 0.076$) show differential methylation between residual and treatment naïve tumours. The response of these models to various chemotherapeutics was previously established. Using this approach, we discovered that the overall methylation patterns determined by mDETECT could stratify the tumours into two groups that differentially respond to cisplatin ($p=0.029$). Finally, *GIPC2* methylation was associated with resistance to both docetaxel ($p = 0.046$) as well as cisplatin ($p = 0.013$). The results of this study identify potential biomarkers that have the ability to predict whether or not a TNBC patient will respond to different courses of chemotherapy.

3.2 Introduction

TNBC is a particularly aggressive subtype of the disease that is defined by the lack of expression of ER and PR as well as HER2[2]. Due to the absence of these receptors, there are currently no effective targeted therapies for women who develop TNBC. Although these patients respond well to neoadjuvant chemotherapy, they have a high rate of relapse in the years following diagnosis[14–16]. While these TNBC patients often respond well to neoadjuvant treatment compared with the other subtypes of breast cancer[14], there are still a substantial number that do not, with only 20 to 40% achieving pathological complete response (pCR)[15,164–166]. In TNBC, achievement of pCR is indicative of a favorable outcome[167] and it has been observed that those patients who fail to achieve pCR are more likely to relapse[14]. Currently there are limited ways to identify which of these women will not respond to chemotherapy. As such, the development of predictive biomarkers for this population would be highly valuable.

Numerous factors have been explored in an attempt to predict which TNBC patients will respond to chemotherapy and which will not. These include the presence of proliferation markers[22] or the combination of Epidermal Growth Factor Receptor (EGFR) status and the metabolic response of the tumour during chemotherapy[19]. It is also known that the presence of a mutation in a *BRCA* gene, which leads to homologous recombination (HR) deficiency, predicts response to platinum-based chemotherapy in TNBC patients[20165]. Furthermore, it has been observed that in patients who lack a mutation in these genes, deficiency in HR also predicts response to these treatments[20,23,168]. Additionally, the presence of tumour infiltrating lymphocytes (TIL) predicts TNBC response to numerous chemotherapies including anthracycline[164,166,169]

and taxane[164] based agents. Despite the fact that numerous markers have been implicated in predicting the response of TNBC patients to chemotherapy none have proven to be clinically useful[19,23,164,166,168–170]. Because of this, the development of predictive biomarkers for this population remains an important area of research.

Patient derived xenograft (PDX) mouse models provide a way to examine human tumours in an *in vivo* setting[171]. These models involve the transplantation of human tumour cells into mice, enabling the assessment of tumour characteristics such as response to therapy. Previously, PDX models have been created with both treatment naïve tumours[172] and samples from viable TNBC cells remaining in a patient's tumour following neoadjuvant chemotherapy (Marangoni, Unpublished). These models provide an interesting example of resistant disease and enable the identification of predictive biomarkers for TNBC.

We have previously developed an assay called mDETECT for the detection of TNBC-specific methylation patterns in circulating tumour DNA (ctDNA) through interrogation of the ctDNA with multiple PCR probes (Cristall *et al.,* unpublished). This test reliably detects extremely low levels of circulating tumour DNA in the blood of patients with metastatic TNBC (Cristall *et al.*, unpublished). While chosen for their high frequency in TNBC tumours, these hypermethylation patterns are likely linked to specific aspects of the development of these tumours. As there is a pressing need to be able to identify TNBC patients that will fail to respond to chemotherapy, we have assessed the probes in the mDETECT assay for their ability to act as predictive biomarkers in this population. To this end, mDETECT was used to assess DNA from the tumours of the resistant TNBC PDX models mentioned above. This data was analyzed to see if any of

the 86 probes in the mDETECT assay are able to identify those patients who did not respond to neoadjuvant chemotherapy.

3.3 Materials and Methods

3.3.1 Test Development

The identification of genomic regions that are hypermethylated in basal-like and BRCA1-associated breast cancer as well as the development of primers for these regions was previously described (Chapter 2). Briefly, level 3 Illumina 450K methylarray data from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) was used to identify regions that show high levels of methylation in either basal or BRCA1-associated tumours and low levels of methylation in normal breast, colon, lung and prostate tissues. Methylation biased primer sets were designed to amplify these regions using either Methyl Primer Express Software v1.0 (ThermoFisher Scientific, Waltham, MA, USA) or Primer Blast (National Center for Biotechnology Information, Bethesda, MD, USA). The primer sets used for the analysis of samples in this study were validated in cell lines and healthy human buffy coat DNA, also previously described, to ensure functionality (Chapter 2).

3.3.2 Development of PDX TNBC Tumors Models

PDX models of human TNBC tumours were developed at the Curie Institute using a method previously described[172]. In brief, a portion of a patient's tumour was taken at the time of surgery, transplanted into Swiss nude mice and allowed to grow. This was done for 26 patients who did not receive neoadjuvant chemotherapy and 10 patients

that received neo-adjuvant treatment with a combination of 5-fluorouracil, epirubicin and cyclophosphamide (FEC) plus a taxane and had residual disease.

Once established, mouse models were challenged with 4 types of chemotherapy to assess response of both residual as well as naïve tumours. Treatment was carried out using Adriamycin + cyclophosphamide (AC), docetaxel, cisplatin or capecitabine. Response to treatment was assessed by determining percent of tumour growth inhibition following two cycles of chemotherapy. Tumors were termed chemosensitive if their growth was inhibited by 90% or more (Marangoni, unpublished).

3.3.3 Preparation of DNA from PDX tumours

DNA was extracted from the PDX tumours described above using the DNeasy Blood & Tissue Kit (Cat. No. 69504, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions. DNA was quantified using Nanodrop (ThermoFisher Scientific, Waltham, MA, USA). 100ng of DNA from each tumour sample was bisulfite converted using the EpiTect Fast DNA Bisulfite Kit (Cat. No. 59824, Qiagen, Toronto, ON, Canada) according to the manufacturer's instructions and eluted in 20µL of elution buffer.

3.3.4 Multiplex PCR in PDX Tumors

A 25µL multiplex PCR reaction was carried out using 1X Qiagen Multiplex PCR Plus mix (Cat No. 206152, Qiagen, Toronto, ON, Canada), 5ng of bisulphite converted DNA and 0.015µM of a combined primer mix containing the forward and reverse primers for 86 primer sets. Cycling conditions were 95^{o}C for 15 minutes, 35 cycles of 95^{o}C for 30 seconds, 62^{o}C for 90 seconds, 72^{o}C 90 seconds followed by 68^{o}C for 10 minutes and a hold at 4^{o}C. In order to reduce primer dimers generated in this multiplex

reaction being carried on to subsequent amplification, 20µL of the primary PCR reaction was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturers protocol at a ratio of 0.8:1 (beads to sample). Following purification, the sample was eluted in 25µL of water. Secondary PCR and barcoding was carried out using 0.5U Qiagen Hot Start Taq Plus (Cat No. 203205, Qiagen, Toronto, ON, Canada) in a 50ul reaction with 1X PCR buffer (Cat No. 203205, Qiagen, Toronto, ON, Canada), 0.2mM dNTPs (Cat No. N0440S – N0443S, New England Biolabs, Ipswich, MA, USA), 3mM $MgCl_2$ (Cat No. 203205, Qiagen, Toronto, ON, Canada), 10ul of purified primary PCR product and 0.08µM consensus sequence 1 (CS1) universal primer with a unique barcode for each sample and consensus sequence 2 (CS2) universal primer (Sigma-Aldrich, Oakville, ON, Canada). PCR conditions were 95^{o}C for 5 minutes, followed by 10 cycles of 95^{o}C for 30 seconds, 58^{o}C for 30 seconds, 72^{o}C for 30 seconds followed by 72 ^{o}C for 7 minutes and a hold at 4^{o}C. 45ul of this secondary PCR was purified using Agencourt AmpureXP magnetic beads (Cat No. A63881, Beckman-Coulter, Mississauga, ON, Canada) according to the manufacturers protocol at a ratio of 1.2:1 (beads to sample) followed by elution in 25ul of water. Products were then sequenced on an Ion Torrent platform.

Sequencing generated BAM files were converted to FASTA files using GenAp Galaxy[134]. This software was also used to filter out all reads below 100bp to eliminate those reads representing primer dimer. BiqAnalyzer HT[135] was used to analyze the methylation status of each probe. Reference sequences corresponding to the region between PCR probes were used to ensure the methylation status of only the amplified regions were being assessed. Methylation analysis was carried out with a minimum

sequence identify of 0.9, maximum fraction of unrecognized site of 0.0 and a minimum methylation level of either 0, to determine the total reads for a given region, or 0.8 to identify methylated reads.

3.3.5 Statistics

Percentage of reads lost to primer dimer was calculated by dividing the number of reads left after filtering in GenAP Galaxy and BiqAnalyzer by the total number of reads for a given sample.

Normalized total methylated reads was determined for each sample by adding together the total normalized methylated reads for all probes of sample. Normalized total methylated reads for each probe was calculate as follows:

$$\frac{average\ total\ reads\ for\ all\ samples}{total\ reads\ for\ sample} \times methylated\ reads\ for\ probe$$

NGS results for all samples were binarized by setting any probe with greater than one normalized methylated read as '1' and any probes with either one or zero normalized methylated reads as '0'. Methylated reads were set as those with greater than 80% of CpG dinucleotides in the PCR probe being methylated. Once binarized, the number of methylated and unmethylated probes for a given sample were calculated for the group of residual tumours as well as for the group of untreated tumours. These values were used in conducting a two-tailed Fisher's exact test using GraphPad Prism. Significance was set at p-values < 0.05.

Binarized results were also used to cluster samples using the clustergram function in Matlab. Differences in the samples in each clusters response to different drugs was assessed using a two-tailed Fisher's exact test in GraphPad Prism. Significance was set at

p-values < 0.05. TNBC patient plasma sample data previously reported (Cristall *et al.,* unpublished) were clustered based on corrected methylated reads which where determined as previously described (Chapter 2).

3.4 Results

3.4.1 mDETECT Characterization of PDX Tumors

mDETECT is an assay that enables the detection of tumour specific patterns of promoter hypermethylation in circulating tumour DNA (ctDNA) and a version was designed unique to TNBC. The 86 probes (Appendix A) in the assay were combined and previously optimized in a multiplex PCR using the TNBC cell line HCC1937 (Chapter 2). This assay was used to carry out targeted bisulfite sequencing on DNA extracted from the tumours of 36 PDX mouse models. These models consisted of 10 models of chemoresistant disease that were created by transplanting residual tumour cells from women following neoadjuvant chemotherapy into mice (Marangoni, unpublished), as well as 26 models in which the transplanted primary tumours had not been exposed to neoadjuvant treatment[172] (Table 3.1).

Following NGS it was observed that between 93.5 and 99.9% of reads were lost to primer dimer with a median of 98.5%. This left samples with between 312 and 26,222 total reads across all 86 probes and a median of 3,847 reads. Individual probes had an average of between 0 and 1700 reads with a median of 4 (Appendix E).

Despite the poor coverage and low number of reads per probe, it was found that both the residual and untreated tumours could be recognized by mDETECT. To identify methylated reads, the BiqAnalyzer analysis was filtered so that only reads with 80% or more of CpGs in the region being methylated were exported (Appendix E). Generally, the

Table 3.1 PDX Tumour Response to Treatment

	Sample ID	Treatment Status	AC	Docetaxel	Cisplatin	Capecitabine	Cluster
PDX1	BC11	Naïve	S	S	S	S	Group 1
PDX2	BC73	Naïve	R	R	R	R	Group 1
PDX3	BC138	Naïve	R	R	S	S	ND
PDX4	BC143	Naïve	R	R	R	R	Group 2
PDX5	BC146	Naïve	R	R	R	R	Group 2
PDX6	BC147	Naïve	R	R	R	R	Group 2
PDX7	BC152	Naïve	R	R	R	R	Group 2
PDX8	BC162	Naïve	R	R	R	R	Group 1
PDX9	BC173	Residual	R	R	R	R	Group 1
PDX10	BC227	Naïve	S	R	R	R	Group 2
PDX11	BC247	Naïve	R	R	R	R	Group 2
PDX12	BC256	Naïve	S	R	R	S	Group 2
PDX13	BC262	Naïve	R	S	S	R	Group 2
PDX14	BC291	Naïve	S	R	S	R	Group 2
PDX15	BC297	Naïve	R	ND	S	R	Group 1
PDX16	BC326	Naïve	S	R	S	R	Group 2
PDX17	BC381	Naïve	S	R	S	R	Group 2
PDX18	BC408	Residual	S	R	S	S	Group 2
PDX19	BC474MS	Naïve	ND	ND	ND	ND	ND
PDX20	BC509	Naïve	S	S	S	R	Group 1
PDX21	BC664MSC	Naïve	ND	ND	ND	ND	ND
PDX22	BC701NFC	Naïve	ND	ND	ND	ND	ND
PDX23	BC718	Naïve	ND	ND	ND	ND	ND
PDX24	BC762MA	Naïve	ND	ND	ND	ND	ND
PDX25	BC763	Residual	R	R	R	R	Group 1
PDX26	BC801	Residual	R	R	R	S	Group 2
PDX27	BC899	Residual	R	R	R	S	Group 2

PDX28	BC910	Residual	ND	ND	ND	ND	ND
PDX29	BC953	Naïve	ND	ND	ND	ND	ND
PDX30	BC965	Residual	R	R	R	S	Group 2
PDX31	BC976	Residual	R	R	R	R	Group 1
PDX32	BC989	Naïve	ND	ND	ND	ND	ND
PDX33	BC990	Naïve	ND	ND	ND	ND	ND
PDX34	BC999	Naïve	ND	ND	ND	ND	ND
PDX35	BC1006	Residual	R	R	R	R	Group 1
PDX36	BC1014	Residual	ND	ND	ND	ND	ND

S = sensitive, R = resistant, ND = not determined

majority of reads are either entirely methylated or unmethylated across a given probe though an 80% cut off was chosen as some regions had several unmethylated CpGs. A probe was deemed to give a positive signal for a particular sample if it returned greater than one methylated read. These results are presented by BiqAnalyzer as a heatmap from which it can be observed that each of the tumour samples is positive for between 12 and 54 of the 86 probes in the assay (Figure 3.1).

3.4.2 Predictive Biomarkers for TNBC

The 10 PDX models of resistant TNBC were created from residual tumours initially treated in the neoadjuvant setting with 5-fluorouracil, epirubicin and cyclophosphamide (FEC) plus a taxane (Marangoni, unpublished) representing disease resistant to multiple chemotherapeutic challenges. The 26 additional models created by engrafting untreated TNBC tumours into mice (Marangoni, unpublished) represent treatment naïve tumours, some of which may be also be resistant. Analysis of the pattern of mDETECT probes in these models revealed significant association between certain markers and residual tumours. This analysis revealed that, when examining residual versus naïve tumours, methylation of *TNFRSF10D* is significantly associated with residual tumours ($p = 0.025$) while lack of methylation in *FOXA3* is approaching significance in residual tumours ($p = 0.076$).

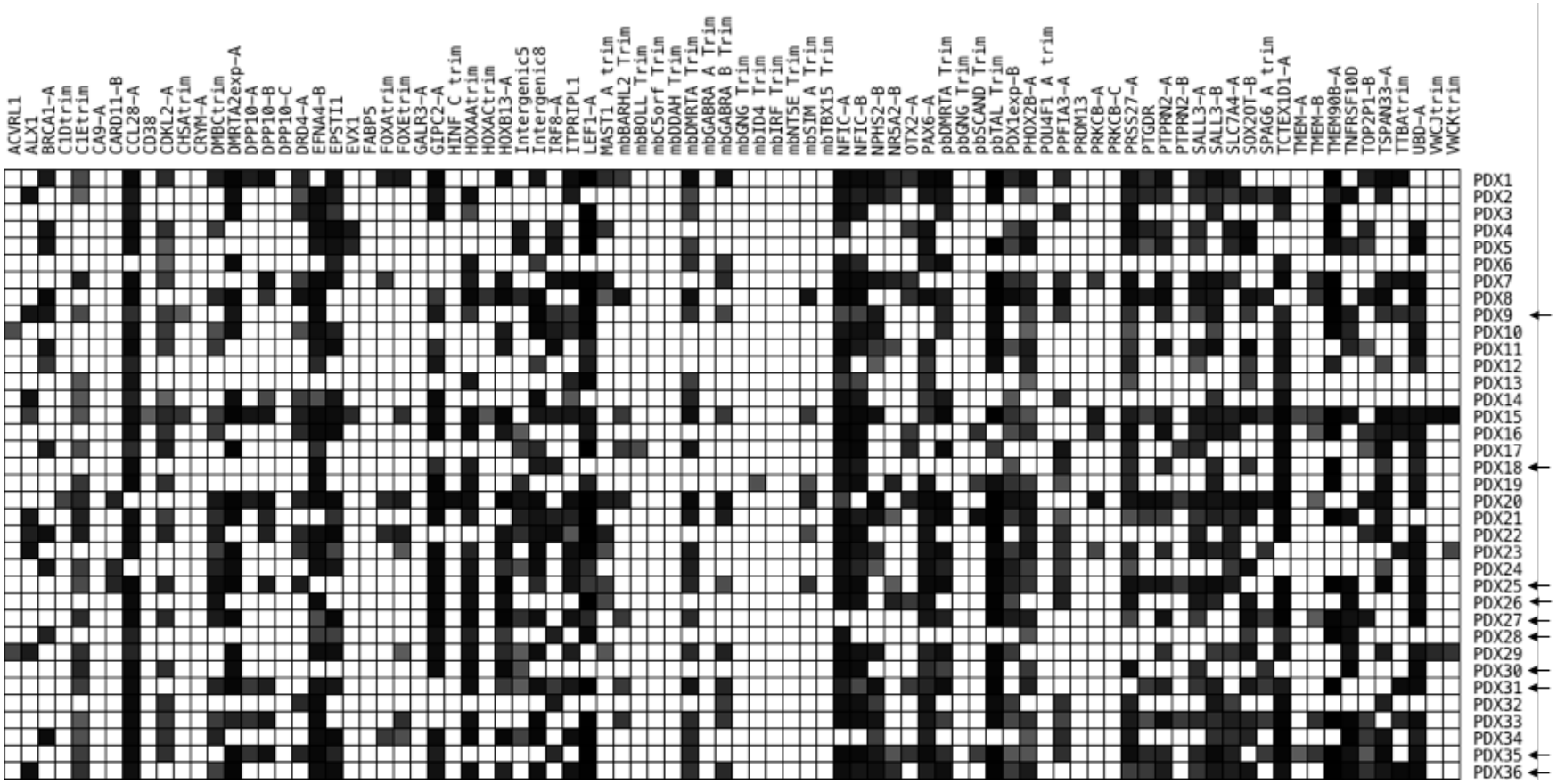


Figure 3.1 mDETECT analysis of TNBC PDX Tumours

The DNA from 36 TNBC PDX tumours were analyzed using mDETECT. This included 10 models of residual disease (indicated by arrows) and 26 models of treatment naïve disease. The BiqAnalyzer output representing those reads exhibiting over 80% methylation is displayed here as a heatmap. Each column corresponds to a given probe while each row represents a different sample. The darker grey a box is, the more methylated the reads for that probe in a given sample were. White boxes indicate that no reads were obtained for a sample; however, there is no distinction between samples in which no methylated reads were obtained and those for which no reads at all were obtained. From this map it can be observed that all tumours are covered by some probes, with between 12 and 54 out of 86 probes reporting for each sample.

Unfortunately, within the group of untreated tumours, we do not know which would or would not have been chemoresistant if the patient had been treated. As such, while differences in methylation can be assessed between the residual and untreated tumours, it is perhaps more informative to examine these differences between various categories of tumours based on how they responded to therapy in the mouse model setting. Following establishment of these naïve models, they were challenged with either Adriamycin + cyclophosphamide (AC), docetaxel, cisplatin or capecitabine and assessed for tumour response based on percent tumour growth inhibition (Marangoni, unpublished) (Table 3.1).

When the binarized NGS results for these PDX tumours (see materials & methods for details on binarization) were clustered, two groups were identified (Figure 3.2). Using a Fisher's exact test it was determined that group 1 is significantly more resistant to cisplatin than group 2 ($p = 0.029$).

Fisher's exact test was also used to examine the ability of each probe in the mDETECT assay to differentiate between those groups found to be resistant or sensitive to Adriamycin + cyclophosphamide (AC), docetaxel, cisplatin or capecitabine. Methylation in *GIPC2* was significantly associated with resistance to both docetaxel ($p = 0.046$) and cisplatin ($p = 0.013$).

Finally, the results obtained from a previous study using mDETECT to analyze plasma DNA from patients with TNBC (Cristall *et al.,* unpublished) was clustered and it was shown that, while the majority of probes cluster together into a larger group, both

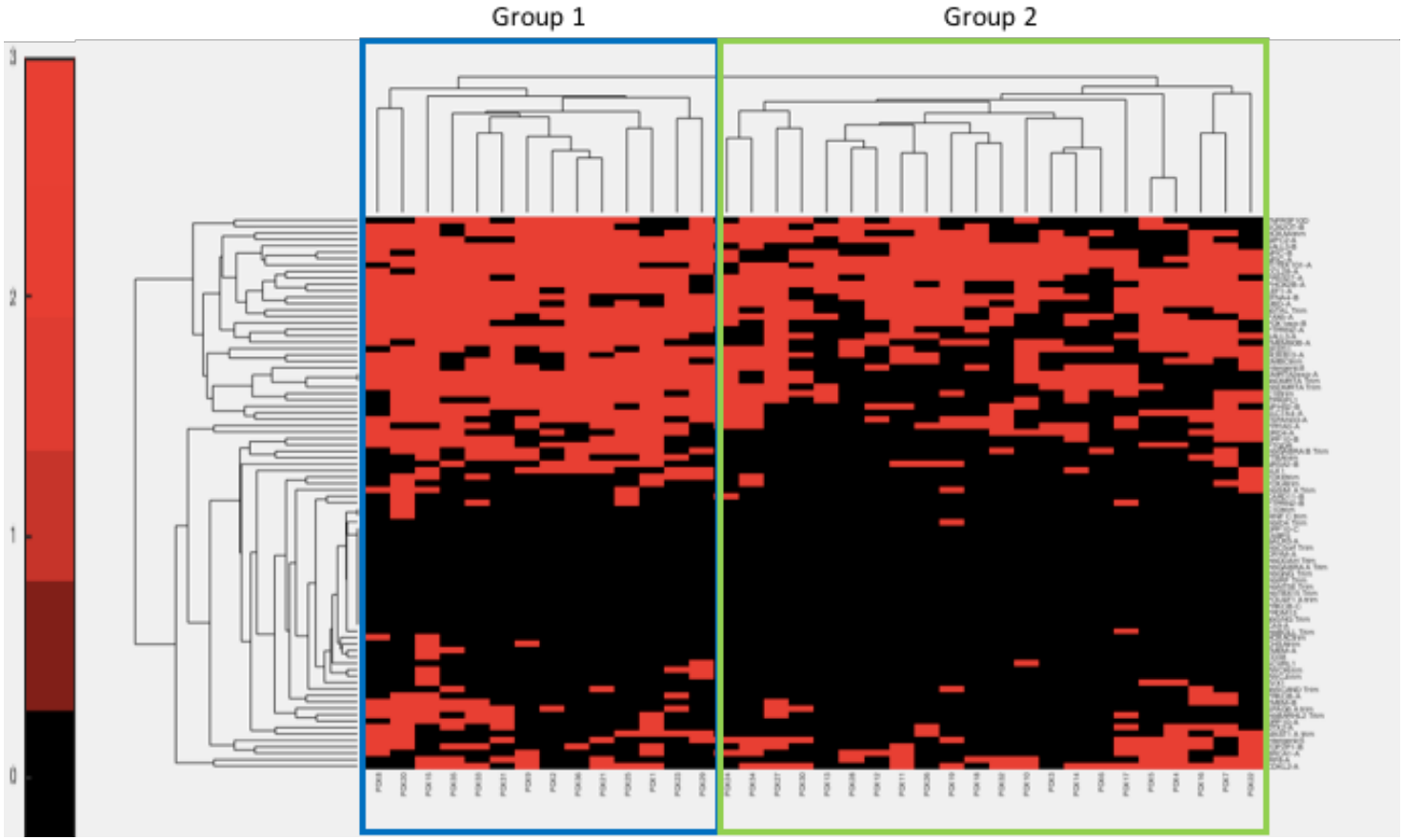


Figure 3.2 Clustering mDETECT results for PDX tumours

The mDETECT results (binarized total methylated reads) for each of the 36 PDX tumour were clustered using the clustergram function in MATLAB. As indicated by the figure legend, black indicates that there were one or no corrected methylated reads for a probe while red indicates that there were more than one. Only these two colors are shown on the clustergram due to the fact that it was constructed using binarized data. This analysis identified that the PDX tumours fall into two groups (Blue box, Green Box). It was then determined that these two groups show a significant difference in their response to cisplatin with group 1 being significantly more resistant ($p = 0.029$).

GIPC2 and TNFRSF10D, two of the resistance associated genes, cluster together in a separate, smaller group (Figure 3.3) along with the CD38 gene.

3.5 Discussion

Through the examination of TNBC PDX tumours, three different potentially predictive biomarkers were identified. *TNFRSF10D* was observed to be significantly hypermethylated in residual tumours when compared with untreated tumours. Conversely, lack of *FOXA3* methylation was also associated with the residual group of tumours. Lastly, hypermethylation of *GIPC2* was specifically correlated with resistance to doxetaxel and cisplatin. Furthermore, it was found that the overall methylation pattern across all the probes in the assay could differentiate between two groups of tumours with differing responses to cisplatin.

TNFRSF10D, also known as decoy receptor 2 (DCR2), is a receptor for the tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) [173]. The pro-apoptotic message carried by TRAIL is transmitted to the cell through the cell surface death receptors DR4 and DR5[173]. However, TRAIL can also bind to the decoy receptors TNFRSF10C and TNFRSF10D[173]. TNFRSF10C and TNFRSF10D compete with DR4 and 5 to bind to TRAIL and when this occurs, the pro-apoptotic signal typically carried by TRAIL is not passed on to the cell. This inhibition of TRAIL signaling occurs in the majority of normal cells[173]. As such, TNFRSF10C and TNFRSF10D have been considered to have an anti-apoptotic function[173].

Interestingly, it has been found that decoy receptor expression is frequently lost as a result of promoter hypermethylation in a large number of malignancies including breast, lung, prostate, ovarian and blood cancers[174]. Although these genes are frequently

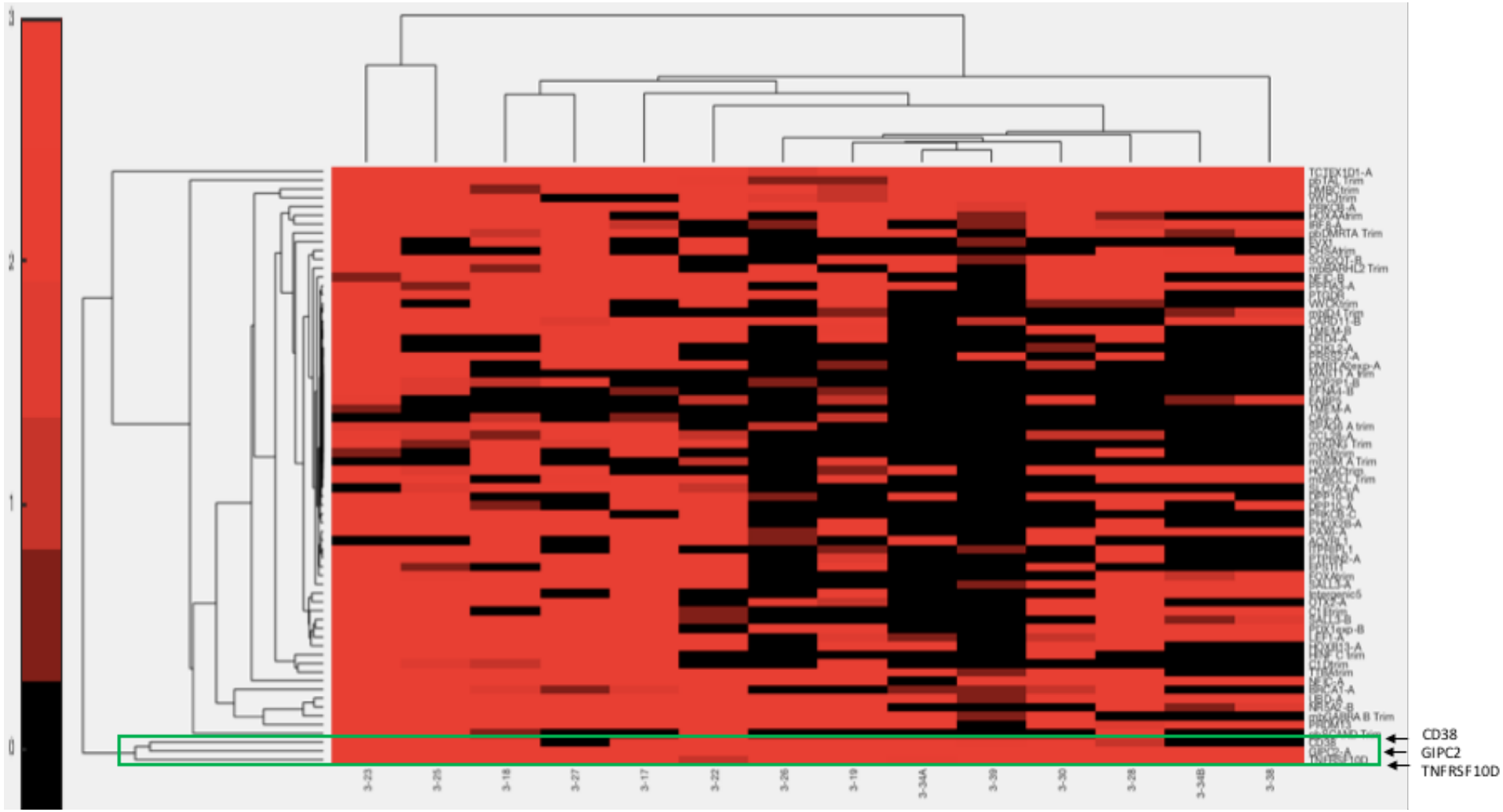

3
2
1
0
CD38
GIPC2
TNFRSF10D

Figure 3.3 Clustering mDETECT results for TNBC patient plasma samples

The mDETECT results (corrected methylated reads) from the previous analysis of 14 TNBC patient plasma samples was clustered using the clustergram function in MATLAB. As indicated by the figure legend, black indicates that no corrected methylated reads were present for a probe. Furthermore, the gradient of red indicates increasing numbers of corrected methylated reads for a given probe in a certain sample. This revealed that, while the majority of probes cluster together, 3 segregate into a separate group (boxed in green). These 3 probes are GIPC2, TNFRSF10D and CD38 (Green box, Arrows). This study identified both *GIPC2* and *TNFRSF10D* hypermethylation to be indicative of chemotherapy resistance. This result lends further support to this finding, indicating that a particular pattern of hypermethylation is associated with those probes that predict resistance.

lost in tumour cells, they are often expressed in normal tissues[174]. Hypermethylation of these genes has not been extensively explored but has been examined more closely in melanoma[175,176], prostate cancer[177] and colorectal cancer[178]. *TNFRSF10D* is also hypermethylated in as many as 74% of breast cancer cell lines and 63% of tumours[174]. Furthermore, in a previous study we identified that 36% (5/14) primary TNBC tumours as well as 79% (11/14) of ctDNA samples from metastatic TNBC patients showed hypermethylation of *TNFRSF10D,* but not in buffy coat DNA from a healthy donor (Chapter 2).

The implications of *TNFRSF10C* and *D* loss of expression are not entirely evident and continue to be an area of interest due to the fact that the loss of anti-apoptotic signaling in tumour cells seems counterintuitive. Nonetheless, the loss of these molecules seems to be an important event in numerous malignancies and may represent an alteration that occurs in the transition from pre-malignancy to malignancy of the lung[179]. One possible explanation revolves around the idea that, under certain conditions, TRAIL signaling can actually initiate NF-kB signaling[180]. As such, decoy receptor expression may not only inhibit TRAIL-mediated apoptosis but also NF-kB related pathways. This would deprive the cell of survival signals and thus would need to be circumvented by the tumour, resulting in the silencing of decoy receptors[174]. Additionally, it is possible that decoy receptors have additional roles separate from their function of silencing TRAIL signaling. It is possible that in certain contexts they assist in transmitting the TRAIL-induced pro-apoptotic signaling through interactions with other receptors and thus require silencing for tumour cell survival.

Decoy receptor methylation is not just characteristic of different tumour types but also appears to have implications for outcome and response to treatment[176,178,181]. For example, *TNFRSF10D* methylation in melanoma is indicative of poor overall as well as relapse-free survival[176]. Furthermore, *TNFRSF10C* methylation may predict a lack of response to irinotecan-based chemotherapy in metastatic colorectal cancer[178]. These results support our finding that *TNFRSF10D* methylation is associated with chemoresistant tumours. However, another study observed conflicting results in which overexpression of *TNFRSF10D* predicted lack of response of non-small cell lung cancer and ER+ breast cancer cell lines to chemotherapy[181].

There are many ways in which resistance to chemotherapy can arise. These can be extremely specific to a particular drug or can be a more general resistance mechanism that lead to resistance to many treatments[182]. The tumours used in this study have been treated with combination chemotherapy of 5-fluorouracil, epirubicin and cyclophosphamide (FEC) plus a taxane. As a result, these tumours are theoretically resistant to all four of these drugs. Because of this, biomarkers identified in the residual tumour population are likely to confer resistance to numerous types of chemotherapy through mechanisms such as resistance to apoptosis or increased survival signaling[182]. Due to the fact that TNFRSF10D is involved in both of these pathways[173,180], it is of particular interest as a predictive biomarker.

In contrast to *TNFRSF10D*, *FOXA3* is unmethylated in residual tumours when compared with treatment naïve tumours. While this finding was only approaching significance, this is possibly a result of the small sample size of this study, the lack of NGS reads or the unclear distinction between sensitive and resistant tumours.

Furthermore, it was observed that *FOXA3* was unmethylated in all residual tumours. As a result, we thought it worthwhile to explore this finding.

The forkhead box A (*FOXA*) family of genes are known as pioneer factors for their ability to modify chromatin structure, allowing other transcription factors to bind[183]. This family consists of three highly related members, FOXA1, A2 and A3, all of which are important for organ development[184]. FOXA3 is particularly important for liver development[185] and regeneration[186] as well as glucose homeostasis[187]. This family of proteins has recently been increasingly associated with various types of cancers[188,189]; however, there is little information regarding FOXA3 in this context.

FOXA1 is the most well characterized member of the FOXA family in relation to cancer[189]. Expression of this gene is linked with a number of cancers including prostate[190,191] and breast cancer[192–194]. The role played by this molecule appears to be context dependent and has been linked with both positive[195] as well as poor prognosis[191] depending on the cancer type. In one example, *FOXA1* is predominantly expressed in ER+ breast cancers[192–194]. Although *FOXA1* expression is predominantly expressed in ER+ breast cancers, one study that attempted to identify subtypes of TNBC found that *FOXA1* expression is also specifically associated with a particular from of TNBC known as the luminal androgen receptor (LAR) subtype[196]. Interestingly, the LAR subtype has one of the lowest responses to chemotherapy out of the seven subtypes identified based on pathological complete response (pCR)[197]. Because of the high level of homology between FOXA1 and FOXA3[184], this leads us to hypothesize that perhaps expression of *FOXA3*, as marked by hypomethylation, is also indicative of an aggressive subtype of TNBC and predicts a lack of response to chemotherapy.

As opposed to *TNFRSF10D* and *FOXA3*, which distinguish between residual and untreated tumours, hypermethylation of *GIPC2* is instead indicative of those tumours that are specifically resistant to docetaxel and cisplatin. The members of the GIPC family are involved in a number of cellular processes such as transmembrane protein trafficking as well as the regulation of cell proliferation and migration[198].

Unfortunately, the roles of GIPC2 are not well characterized but the involvement of this molecule in malignancy has been observed. *GIPC2* has been documented to be upregulated in some gastric cancers[199] but downregulated in kidney cancer[200]. Furthermore, the *GIPC2* promoter has been found to be hypermethylated in all of the 23 leukemia cell lines included in one study as well as many of the acute lymphocytic leukemia (ALL) patient samples that were examined[201]. It has also been observed that depletion of GIPC1, a homolog of GIPC2[200], leads to the secretion of the drug resistance molecule ATP-binding cassette G2 (ABCG2) in exosomes in prostate cancer cells[202]. This removes it from the cell, sensitizing cells to gemcitabine[202]. GIPC1 depletion in ER+ and ER- breast cancer cell lines leads to apoptosis in these cells and GIPC1 expression also promotes breast tumour progression[203]. These findings appear to be in direct contrast with the ones presented here in which GIPC2 promoter hypermethylation leads to drug resistance. However, it is possible that GIPC2 serves a slightly different function that GIPC1 and thus, when depleted leads to resistance. The involvement of GIPC family members with the surface expression of ABC transporters is of interest as ABC transporter expression is known to mediate resistance to numerous chemotherapies[204]. Although there is little information regarding GIPC2 specifically, it is clear that it plays a role in malignancy, likely in a context dependent manner based on the differing patterns

of expression observed in various cancer types[199,200]. As such, it presents as an interesting potential biomarker for predicting docetaxel and cisplatin resistance in TNBC.

Interestingly, when the mDETECT results of the analysis of TNBC patient plasma DNA samples previously analyzed (Cristall *et al.,* unpublished) were clustered, that majority of probes clustered together while 3 formed a separate group. Two of the probes in the category were *GIPC2* and *TNFRSF10D*. The fact that these two genes cluster together, indicating that there is a particular hypermethylation pattern associated with resistance determining genes, further support their potential as predictive biomarkers. The third gene clustering in this group is *CD38*, which was not identified as a marker of resistance in our study. However, the *CD38* probe generated few reads across all PDX tumours analyzed in this study which could have resulted in it being missed in our analysis. The expression of this predicts lack of response to chemotherapy in chronic lymphocytic leukemia[205] and, as such, is a potential biomarker of resistance.

To date, a number of biomarkers have been identified that are though to predict which TNBC patients will or will not respond to chemotherapy. These markers include factors such as the identification of a proliferation signature[22], tumour glucose metabolism[19], homologous recombination deficiencies, whether through loss of BRCA1[20,38,165] or other means[23], and the presence of tumour infiltrating lymphocytes[164,166,169]. Unfortunately, none of these markers have been validated for use in the clinic. Furthermore, none of these markers appear to have utility beyond the prediction of treatment response, making them potentially costly additions to the tests that already have to be performed. Here we provide evidence that an assay already demonstrated to have potential utility in terms of monitoring TNBC progression, and

possibly early detection, may also have the ability to predict which patients will or will not respond to treatment. Furthermore, none of the previously identified biomarkers are liquid biopsies. All require examination of the tumour itself, meaning they can only assess disease at baseline. Because mDETECT is a blood test, it can easily and safely be repeated over time to assess the evolution of the tumour and the development of resistance. The multi-use nature of this liquid biopsy provides a practical, economical assay with a great deal of potential to improve the management of TNBC.

Little research has been conducted regarding the predictive potential of methylation based biomarkers in TNBC. The one example in which this has been previously identified is in the context of *BRCA1* methylation, which has been observed to predict response to platinum[38] and anthracycline[170] based therapy. Furthermore, stratification of TNBC based on aberrant DNA methylation patterns has led to the identification of groups of TNBC with differing prognoses however, these categories were not associated with response to treatment[26]. As such, this study represents one of the first incidences of promoter hypermethylation being identified as a biomarker of TNBC chemotherapy resistance.

One limitation to this study was the number of reads obtained through NGS and subsequently lost to primer dimer. mDETECT involves the use of many probes multiplexed into one reaction. As a result of this high primer concentration, substantial primer dimer is generated that are sequenced preferentially over the product of interest. Because as many as 99.9% of sequencing reads were lost as a result of this primer dimer, many probes failed to generate reads. This likely resulting in the missed identification of potential biomarkers. Since this study was conducted, improvements have been made to

the technique used to carry out the mDETECT assay that reduce primer dimer and evens out read distribution (Cristall *et al.,* unpublished). As such, repeating this experiment using the new mDETECT protocol could result in the identification of additional TNBC predictive biomarkers.

A further limitation to this study is the fact that we were unable to complete a prospective analysis of tumours that were later found to be sensitive or resistant to treatment. In the future our plan is to carry out a prospective analysis to confirm the findings of this study. We have collected samples from a previously characterized cohort of non-metastatic TNBC patients that were recruited at the Curie Institute[103]. Plasma samples were taken from these patients at four time points; at the time of diagnosis, following one round of chemotherapy, following the completion of chemotherapy and after surgery. This cohort is associated with information regarding patients that did or did not respond to treatment. Because of this, we will be able to carry out mDETECT on these samples and confirm whether or not the markers identified in this study can actually identify at baseline those patients who did or did not go on to respond to treatment.

In summary, we have used the previously developed mDETECT test, an 86 probe assay designed to amplify regions of TNBC specific hypermethylation, to characterize a set of PDX tumours. These samples represent residual and treatment naïve triple negative tumours. Through this analysis we have identified three potential biomarkers capable of predicting response of TNBC patients to chemotherapy. Both *TNFRSF10D* hypermethylation and a lack of methylation in *FOXA3* are associated with residual tumours when compared with untreated tumours. Furthermore, promoter hypermethylation of *GIPC2* is prominent in those tumours that did not respond to

docetaxel or cisplatin. We have previously demonstrated that mDETECT is capable of identifying patients who have metastatic TNBC based on patterns of promoter hypermethylation in ctDNA. Furthermore, based on the extremely sensitive nature of the assay's ability to detect small amounts of ctDNA in the blood, it is possible that this test may be able to be used for monitoring and possibly even early detection of these malignancies. In this study we further identify potential utility for mDETECT and establish that particular probes within the assay have the possible capability of predicting which TNBC patients will or will not respond to therapy. When these results are validated in a prospective setting, these findings have the ability to improve the care received by TNBC patients by enabling the alternative treatment of those women unlikely to respond to the standard of care.

Chapter 4

Conclusion

The primary goal of mDETECT is the identification of the presence of cancer. As such, the regions probed in the assay were not chosen for their biological relevance but simply for their reliably hypermethylated state in triple negative tumors in comparison with normal tissues. Because of this, there is no indication of whether these alterations are driver or passenger events. However, due to the important role aberrant DNA methylation plays in the development of the hallmarks of cancer[110,206,207], it is likely that the hypermethylated regions in mDETECT include at least some that are important in the promotion of the malignant phenotype and, as such, are able to act as biomarkers indicative of such factors as prognosis and treatment response[208]. Based on the ability of some probes to predict response to treatment, it is clear that some must be linked to basic characteristics of the tumour. As a result, mDETECT represents a test that not only detects the presence of cancer in the blood in an extremely sensitive manner, but also has biological relevance, potentially enabling the assay to report clinically important information.

The hallmarks of cancer have long been accepted as representing the fundamental characteristics of malignancies and represent traits that cells must acquire for cancer development to occur[209]. These include sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis and resisting cell death[209,210]. When these hallmarks were recognized[209], as well as when they were revisited ten years after their establishment[210],

genetic aberrations were highlighted as the predominant driving force behind these features. However, there is a constantly growing body of evidence that suggest epigenetic mechanisms, in addition to the genetic mechanisms already identified, are capable of driving each of the six hallmarks of cancer[110,206,207].

DNA methylation is one of the principle epigenetic systems that can alter gene expression over the long term[211]. Aberrant methylation has long been recognized as a characteristic of cancer cells[212]. This includes an overall global hypomethylation that destabilizes the genome and activates oncogenes, as well as the specific hypermethylation of certain promoters leading to loss of tumor suppressor gene expression[110]. It has been shown that these alterations in DNA methylation have the ability to drive the development of many, if not all, of the hallmarks of cancer[110,206,207]. In fact, promoter hypermethylation alone has been shown to silence numerous genes, subsequently resulting in sustained proliferative signaling[213], evasion of growth suppressors[214], induction of angiogenesis[215] and resistance to cell death[216]. For example, hypermethylation of interferon regulatory factor 8 (*IRF8*), a gene identified in this study, is capable of mediating resistance to apoptosis in metastatic tumors[216]. Furthermore, homeobox B13 (*HOXB13*), also a gene included in mDETECT, appears to halt the growth of colon cancer cells[157]. The hypermethylation of this gene represents a way in which tumor cells evade growth suppressors[157]. Based on these findings, it is not unreasonable to hypothesize that DNA hypermethylation alone can establish each of the hallmarks of cancer.

Epigenetic changes are constantly occurring in healthy cells as a result of changes in the environment[217], some of which may be indicative of cancer risk or pre-

malignancy[150,151]. Aberrant DNA hypermethylation can arise from exposure to environmental factors, such as smoking tobacco, as well as from the normal aging process[218]. It is difficult to determine whether these epigenetic alterations are indicative of cancer, cancer risk or just a result of aging. However, perhaps such a distinction is unnecessary due to the fact that aging is such a substantial risk factor for cancer[219]. mDETECT has identified hypermethylation in controls that may have the potential to indicate cancer risk or the early presence of cancer. For example, methylation of *BRCA1* in blood cells has been closely tied with risk of breast cancer[150] while *PAX6* methylation has been shown in DCIS, a non-invasive breast abnormality potentially representing the the earliest stage of breast cancer[151]. The fact that mDETECT is able to detect these aberrations in controls indicates an ability to identify cancer risk, pre-malignancy, as well as the tumour itself, supporting its potential use as a screening modality.

The study presented here demonstrates that mDETECT holds great potential for use as a screening, monitoring and predictive assay for TNBC. However, it's potential utility extends far beyond this population. Based on analysis of the current assay in cell lines, some probes clearly show hypermethylation not only in triple negative tumours but in the other subtypes of breast cancer as well. As such, with the careful selection of a particular subset of probes, this assay could be adapted to provide coverage of all breast cancer subtypes. This would be particularly important in terms of implementing mDETECT as a screening modality as it would enable the use of a single assay for the detection of all breast malignancies, much like the current image based methods. In addition to mDETECT's potential to be modified for coverage of all breast cancer subtypes, the premise of mDETECT has the ability to be adapted for the identification of

nearly any malignancy. As a result, screening and monitoring assays for many if not all cancer types could be created. Through this capability, it is possible that with the identification of an optimal set of probes as well as further improvements in multiplex PCR technology, the assays for all of these malignancies could be combined to create a universal blood test for cancer.

References

1. Candian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2017. Toronto, ON: Canadian Cancer Society 2017.

2. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

3. Sørlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci.* **100,** 8418–8423 (2003).

4. Rakha, E. A. *et al.* Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clin. Cancer Res.* **15,** 2302–2310 (2009).

5. Bertucci, F. *et al.* How basal are triple-negative breast cancers? *Int. J. cancer* **123,** 236–240 (2008).

6. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10869–10874 (2001).

7. Rouzier, R. Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy. *Clin. Cancer Res.* **11,** 5678–5685 (2005).

8. Turner, N. C. & Reis-filho, J. S. Basal-like breast cancer and the BRCA1 phenotype. 5846–5853 (2006). doi:10.1038/sj.onc.1209876

9. Seong, M.-W. *et al.* A multi-institutional study on the association between BRCA1/BRCA2 mutational status and triple-negative breast cancer in familial breast cancer patients. *Breast Cancer Res. Treat.* **146,** 63–9 (2014).

10. Prat, A., Cruz, C., Hoadley, K. A. & Dı, O. Molecular features of the basal-like breast cancer subtype based on BRCA1 mutation status. 185–191 (2014). doi:10.1007/s10549-014-3056-x

11. Sotiriou, C. *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci.* **100,** 10393–10398 (2003).

12. Perou, C. M. Molecular Stratification of Triple-Negative Breast Cancers. *Oncologist* **16,** 61–70 (2011).

13. Fong, P. C. *et al.* Inhibition of poly(ADP-Ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* **361,** 123–134 (2009).

14. Carey, L. A. *et al.* Triple Negative Paradox: Primary Tumor Chemosensitivity of Breast Cancer Subtypes. *Clin. Cancer Res.* **13,** 2329–2335 (2007).

15. Liedtke, C. *et al.* Response to Neoadjuvant Therapy and Long-Term Survival in Patients With Triple-Negative Breast Cancer. *J. Clin. Oncol.* **26,** 1275–1281 (2008).

16. Dent, R. *et al.* Triple-Negative Breast Cancer: Clinical Features and Patterns of

Recurrence. *Clin. Cancer Res.* **13,** 4429–4435 (2007).

17. Banerjee, S. *et al.* Basal-like breast carcinomas: clinical outcome and response to chemotherapy. *J. Clin. Pathol.* **59,** 729–735 (2006).

18. Thike, A. A. *et al.* Loss of androgen receptor expression predicts early recurrence in triple-negative and basal-like breast cancer. *Mod. Pathol.* **27,** 352–360 (2014).

19. Humbert, O. *et al.* Identification of Biomarkers Including FDG-PET/CT for Early Prediction of Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clin. Cancer Res.* **21,** 5460–5469 (2015).

20. Isakoff, S. J. *et al.* A Multicenter Phase II Clinical Trial of Platinum Monotherapy With Biomarker Assessment in Metastatic Triple-Negative Breast Cancer. *J. Clin. Oncol.* **33,** 1902–1909 (2015).

21. Keam, B. *et al.* Ki-67 can be used for further classification of triple negative breast cancer into two subtypes with different response and prognosis. *Breast Cancer Res.* **13,** (2011).

22. Prat, A. *et al.* Predicting response and survival in chemotherapy-treated triple-negative breast cancer. *Br. J. Cancer* **111,** 1532–1541 (2014).

23. Telli, M. L. *et al.* Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* **22,** 3764–3774 (2016).

24. Sahlberg, K. K. *et al.* A Serum MicroRNA Signature Predicts Tumor Relapse and Survival in Triple-Negative Breast Cancer Patients. *Clin. Cancer Res.* **21,** 1207–1215 (2015).

25. Lee, J. S. *et al.* Basal-like breast cancer displays distinct patterns of promoter methylation Basal-like breast cancer displays distinct patterns of promoter methylation. *Cancer Biol. Ther.* **9,** 1017–1024 (2010).

26. Stirzaker, C. *et al.* Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat. Commun.* **6,** 5899 (2015).

27. Holm, K. *et al.* Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res.* **12,** R36 (2010).

28. Stefansson, O. A. *et al.* A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol. Oncol.* **9,** 555–568 (2015).

29. Mathe, A. *et al.* DNA methylation profile of triple negative breast cancer-specific genes comparing lymph node positive patients to lymph node negative patients. *Sci. Rep.* **6,** 1–15 (2016).

30. Parrella, P. *et al.* Nonrandom Distribution of Aberrant Promoter Methylation of Cancer-Related Genes in Sporadic Breast Tumors Nonrandom Distribution of Aberrant Promoter Methylation of Cancer-Related Genes in Sporadic Breast Tumors. **10,** 5349–5354 (2004).

31. Wei, M. *et al.* BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. *Cancer Res.* **65,** 10692–10699 (2005).

32. Birgisdottir, V. *et al.* Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. *Breast Cancer Res.* **8,** R38 (2006).

33. Turner, N. C. *et al.* BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* **26,** 2126–2132 (2007).

34. Matros, E. *et al.* BRCA1 promoter methylation in sporadic breast tumors: relationship to gene expression profiles. *Breast Cancer Res Treat* **91,** 179–186 (2005).

35. Lips, E. H. *et al.* Triple-negative breast cancer: BRCAness and concordance of clinical features with BRCA1-mutation carriers. *Br. J. Cancer* **108,** 2172–2177 (2013).

36. Turner, N., Tutt, A. & Ashworth, A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* **4,** 1–6 (2004).

37. Veeck, J. *et al.* BRCA1 CpG Island Hypermethylation Predicts Sensitivity to Poly (Adenosine Diphosphate) - Ribose Polymerase Inhibitors. *J. Clin. Oncol.* **28,** 563–564 (2010).

38. Xu, Y. *et al.* Promoter methylation of BRCA1 in triple-negative breast cancer predicts sensitivity to adjuvant chemotherapy. *Ann. Oncol.* **24,** 1498–1505 (2013).

39. Gabai-Kapara, E. *et al.* Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 14205–10 (2014).

40. Mullan, P. B., Quinn, J. E. & Harkin, D. P. The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene* **25,** 5854–5863 (2006).

41. Jiang, Q. & Greenberg, R. A. Deciphering the BRCA1 tumor suppressor network. *J. Biol. Chem.* **290,** 17724–17732 (2015).

42. Zhang, J. & Powell, S. N. The role of the BRCA1 tumor suppressor in DNA double-strand break repair. *Mol. Cancer Res.* **3,** 531–9 (2005).

43. Silver, D. P. & Livingston, D. M. Mechanisms of BRCA1 tumor suppression. *Cancer Discov.* **2,** 679–684 (2012).

44. Konishi, H. *et al.* Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17773–8 (2011).

45. Sedic, M. *et al.* Haploinsufficiency for BRCA1 leads to cell-type-specific genomic instability and premature senescence. *Nat. Commun.* **6,** 7505 (2015).

46. Pathania, S. *et al.* BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **5,** 5496 (2014).

47. Vaclova, T. *et al.* DNA repair capacity is impaired in healthy BRCA1 heterozygous mutation carriers. *Breast Cancer Res. Treat.* (2015).

48. Dworkin, A. M., Spearman, A. D., Tseng, S. Y., Sweet, K. & Toland, A. E. Methylation not a frequent 'second hit' in tumors with germline BRCA mutations. *Fam. Cancer* **8,** 339–346 (2009).

49. Miyoshi, Y., Murase, K. & Oh, K. Basal-like subtype and BRCA1 dysfunction in breast cancers. *Int. J. Clin. Oncol.* **13,** 395–400 (2008).

50. Tung, N. *et al.* Prevalence and predictors of loss of wild type BRCA1 in estrogen receptor positive and negative BRCA1-associated breast cancers. *Breast Cancer Res.* **12,** R95 (2010).

51. Foulkes, W. D. *et al.* Germline BRCA1 Mutations and a Basal Epithelial Phenotype in Breast Cancer. *J. Natl. Cancer Inst.* **95,** 1482–1485 (2003).

52. Hedenfalk, I. *et al.* Gene-Expression Profiles in Hereditary Breast Cancer. *N. Engl. J. Med.* **344,** 539–548 (2001).

53. Jönsson, G. *et al.* Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. *Cancer Res.* **65,** 7612–7621 (2005).

54. Metcalfe, K. A. *et al.* A comparison of the detection of BRCA mutation carriers through the provision of Jewish population-based genetic testing compared with clinic-based genetic testing. *Br. J. Cancer* **109,** 777–9 (2013).

55. Heemskerk-Gerritsen, B. A. M. *et al.* Substantial breast cancer risk reduction and potential survival benefit after bilateral mastectomy when compared with surveillance in healthy BRCA1 and BRCA2 mutation carriers: A prospective analysis. *Ann. Oncol.* **24,** 2029–2035 (2013).

56. Kuhl, C. K. *et al.* Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J. Clin. Oncol.* **23,** 8469–8476 (2005).

57. Lee, C. H. *et al.* Breast Cancer Screening With Imaging: Recommendations From the Society of Breast Imaging and the ACR on the Use of Mammography , Breast MRI , Breast Ultrasound , and Other Technologies for the Detection of Clinically Occult Breast Cancer. *JACR* **7,** 18–27 (2010).

58. Oeffinger, K. C. *et al.* Breast Cancer Screening for Women at Average Risk 2015 Guideline Update From the American Cancer Society. *JAMA* **314,** 1599–1614 (2015).

59. Armstrong, A. C. & Evans, D. G. Management of women at high risk of breast cancer Management of women at high risk of breast cancer. *BMJ* **348,** 28–32 (2014).

60. Nelson, H. D. *et al.* Effectiveness of Breast Cancer Screening: Systematic Review and Force Recommendation. *Ann. Intern. Med.* **164,** 244–255 (2016).

61. Bleyer, A. & Welch, H. G. Effect of three decades of screening mammography on breast-

cancer incidence. *N. Engl. J. Med.* **367,** 1998–2005 (2012).

62. Miller, A. B. *et al.* Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* **348,** 1–10 (2014).

63. Wong, H. B. & Lim, G. H. Measures of Diagnostic Accuracy: Sensitivity , Specificity , PPV and NPV. *Proc. Singapore Healthc.* **20,** 316–318 (2011).

64. Kolb, T. M., Lichy, J. & Newhouse, J. H. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. *Radiology* **225,** 165–175 (2002).

65. Tilanus-Linthorst, M. *et al.* A BRCA1/2 mutation, high breast density and prominent pushing margins of a tumor independently contribute to a frequent false-negative mammography. *Int. J. cancer* **102,** 91–95 (2002).

66. Riedl, C. C. *et al.* Triple-Modality Screening Trial for Familial Breast Cancer Underlines the Importance of Magnetic Resonance Imaging and Questions the Role of Mammography and Ultrasound Regardless of Patient Mutation Status , Age , and Breast Density. *J. Clin. Oncol.* **33,** 1128–1135 (2015).

67. Passaperuma, K. *et al.* Long-term results of screening with magnetic resonance imaging in women with BRCA mutations. *Br. J. Cancer* **107,** 24–30 (2012).

68. Berg, W. A. *et al.* Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single Screening MRI. *JAMA* **307,** 1394–1404 (2012).

69. Komenaka, I. K. *et al.* The Development of Interval Breast Malignancies in Patients with BRCA Mutations. *Cancer* **100,** 2079–2083 (2004).

70. Berrington De Gonzalez, A., Berg, C. D., Visvanathan, K. & Robson, M. Estimated risk of radiation-induced breast cancer from mammographic screening for young BRCA mutation carriers. *J. Natl. Cancer Inst.* **101,** 205–209 (2009).

71. Pijpe, A. *et al.* Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (GENE-RAD-RISK). *BMJ* **345,** 1–15 (2012).

72. Moore, S. G., Shenoy, P. J., Fanucchi, L., Tumeh, J. W. & Flowers, C. R. Cost-effectiveness of MRI compared to mammography for breast cancer screening in a high risk population. *BMC Health Serv. Res.* **9,** 1–8 (2009).

73. Ahern, C. H., Shih, Y. T., Dong, W., Parmigiani, G. & Shen, Y. Cost-effectiveness of alternative strategies for integrating MRI into breast cancer screening for women at high risk. *Br. J. Cancer* **111,** 1542–1551 (2014).

74. Pantel, K. & Alix-panabieres, C. Liquid biopsy: Potential and challenges. *Mol. Oncol.* **10,** 371–373 (2016).

75. Alix-panabières, C. & Pantel, K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov.* **6,** 479–491 (2016).

76. Yong, E. Written in Blood. *Nature* **511,** 524–526 (2014).

77. Alix-Panabieres, C. & Pantel, K. Circulating Tumor Cells: Liquid Biopsy of Cancer. *Clin. Chem.* **59,** 110–118 (2013).

78. Yoshioka, Y. *et al.* Ultra-sensitive liquid biopsy of circulating extracellular vesicles using ExoScreen. *Nat. Commun.* **5,** 1–8 (2014).

79. Cheng, F., Su, L. & Qian, C. Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer The Biology of ct DNA. *Oncotarget* **7,** 48832–48841 (2016).

80. Ulz, P., Heitzer, E., Geigl, J. & Speicher, M. R. Patient monitoring through liquid biopsies using circulating tumor DNA. *Int. J. cancer* **141,** 887–896 (2017).

81. Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature* **17,** 223–238 (2017).

82. Fackler, M. J. *et al.* Novel methylated biomarkers and a robust assay to detect circulating tumor dna in metastatic breast cancer. *Cancer Res.* **74,** 2160–2170 (2014).

83. Imperiale, T. F. *et al.* Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *N Engl J Med* **370,** 1287–1297 (2014).

84. Christensen, E. *et al.* Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur. Urol.* **71,** 961–969 (2017).

85. Wei, F. *et al.* Noninvasive Saliva-based EGFR Gene Mutation Detection in Patients with Lung Cancer. *Am. J. Respir. Crit. Care Med.* **190,** 1117–1126 (2014).

86. Hoque, M. O. *et al.* Detection of Aberrant Methylation of Four Genes in Plasma DNA for the Detection of Breast Cancer. *J. Clin. Oncol.* **24,** 4262–4269 (2006).

87. Bidard, F. *et al.* Detection rate and prognostic value of circulating tumor cells and circulating tumor DNA in metastatic uveal melanoma. *Int. J. Cancer* **134,** 1207–1213 (2013).

88. Visvanathan, K. *et al.* Monitoring of Serum DNA Methylation as an Early Independent Marker of Response and Survival in Metastatic Breast Cancer : TBCRC 005 Prospective Biomarker Study. *J. Clin. Oncol.* **35,** 751–758 (2017).

89. Remon, J. *et al.* Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. *Ann. Oncol.* **28,** 784–790 (2017).

90. De Mattos-Arruda, L. *et al.* Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann. Oncol.* **25,** 1729–1735 (2014).

91. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14,** 985–990

(2008).

92. Dawson, S. *et al.* Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. *N Engl J Med* **368,** 1199–1209 (2013).

93. Gray, E. S. *et al.* Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. *Oncotarget* **6,** 42008–42018 (2014).

94. Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl. Med.* **7,** (2015).

95. Tie, J. *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8,** (2016).

96. Schwarzenbach, H., Hoon, D. S. B. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nature* **11,** 426–437 (2011).

97. Thierry, A. R., Messaoudi, S. El, Gahan, P. B., Anker, P. & Stroun, M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35,** 347–376 (2016).

98. Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J. & Magenheim, J. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci.* **113,** E1826–E1834 (2016).

99. Snyder, M. W. *et al.* Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164,** 57–68 (2016).

100. Kimura, H. *et al.* Detection of Epidermal Growth Factor Receptor Mutations in Serum as a Predictor of the Response to Gefitinib in Patients with Non-Small-Cell Lung Cancer. *Clin. Cancer Res.* **12,** 3915–3922 (2006).

101. Swisher, E. M. *et al.* Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. *Am. J. Obstet. Gynecol.* **193,** 662–667 (2005).

102. Madic, J. *et al.* Circulating tumor DNA and circulating tumor cells in metastatic triple negative breast cancer patients. *Int. J. cancer* **136,** 2158–2165 (2015).

103. Riva, F. *et al.* Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clin. Chem.* **63,** (2017).

104. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20,** 548–554 (2014).

105. Forshew, T. *et al.* Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Sci. Transl. Med.* **4,** (2012).

106. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science (80-.).* **339,** 1546–1558 (2013).

107. The Cancer Genome Atlas Network. Integrated genomic analyses of ovarian carcinoma.

Nature **474,** 609–615 (2011).

108. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* **7,** 21–33 (2006).

109. De Carvalho, D. D. *et al.* DNA Methylation Screening Identifies Driver Epigenetic Events of Cancer Cell Survival. *Cancer Cell* **21,** 655–667 (2012).

110. Shen, H. & Laird, P. W. Interplay between the Cancer Genome and Epigenome. *Cell* **153,** 38–55 (2013).

111. Castelo-Branco, P. *et al.* Methylation of the TERT promoter and risk stratification of childhood brain tumours: an integrative genomic and molecular study. *Lancet Oncol.* **14,** 534–542 (2013).

112. Beukers, W. *et al.* Stratification based on methylation of TBX2 and TBX3 into three molecular grades predicts progression in patients with pTa-bladder cancer. *Mod. Pathol.* 1–8 (2014).

113. Holm, K. *et al.* Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res.* 1–16 (2010).

114. DeVos, T. *et al.* Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin. Chem.* **55,** 1337–1346 (2009).

115. Fu, D., Ren, C., Tan, H., Wei, J. & Zhu, Y. Sox17 Promoter Methylation in Plasma DNA Is Associated With Poor Survival and Can Be Used as a Prognostic Factor in Breast Cancer. *Medicine.* **94,** 1–8 (2015).

116. Martínez-Galán, J. *et al.* ESR1 gene promoter region methylation in free circulating DNA and its correlation with estrogen receptor protein expression in tumor tissue in breast cancer patients. *BMC Cancer* **14,** 1–8 (2014).

117. Sharma, G., Mirza, S., Parshad, R., Srivastava, A. & Datta, S. Clinical significance of promoter hypermethylation of DNA repair genes in tumor and serum DNA in invasive ductal breast carcinoma patients. *Life Sci.* **87,** 83–91 (2010).

118. Warton, K. & Samimi, G. Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front. Mol. Biosci.* **2,** 1–10 (2015).

119. Warton, K., Mahon, K. L. & Samimi, G. Methylated circulating tumor DNA in blood: power in cancer prognosis and response. *Endocr. Relat. Cancer* **23,** R157–R171 (2016).

120. Wai Yin Cheuk, I., Cheuk, Y., Shin, V. Y. & Kwong, A. Breast Cancer Detection of Methylated Circulating DNA as Noninvasive Biomarkers for Breast Cancer Diagnosis. *J. Breast Cancer* **20,** 12–19 (2017).

121. Montavon, C. *et al.* Prognostic and diagnostic significance of DNA methylation patterns in high grade serous ovarian cancer. *Gynecol. Oncol.* **124,** 582–588 (2012).

122. Church, T. R. *et al.* Prospective evaluation of methylated SEPT9 in plasma for detection

of asymptomatic colorectal cancer. *Gut* **63,** 317–325 (2014).

123. Radpour, R. *et al.* Hypermethylation of Tumor Suppressor Genes Involved in Critical Regulatory Pathways for Developing a Blood- Based Test in Breast Cancer. *PLoS One* **6,** (2011).

124. Hulbert, A. *et al.* Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum. *Clin. Cancer Res.* **23,** 1998–2006 (2017).

125. Mouliere, F. & Thierry, A. R. The importance of examining the proportion of circulating DNA originating from tumor, microenvironment and normal cells in colorectal cancer patients The importance of examining the proportion of circulating DNA originating from tumor, microenvironment . *Expert Opin. Biol. Ther.* **12,** (2012).

126. Lundberg, K. S. *et al.* High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108,** 1–6 (1991).

127. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13,** (2012).

128. Li, M. *et al.* Sensitive digital quantification of DNA methylation in clinical samples. *Nat. Biotechnol.* **27,** 858–863 (2009).

129. Narayan, A. *et al.* Ultrasensitive Measurement of Hotspot Mutations in Tumor DNA in Blood Using Error-Suppressed Multiplexed Deep Sequencing. *Cancer Res.* **72,** 3492–3499 (2012).

130. Sharma, P. *et al.* The prognostic value of BRCA1 promoter methylation in early stage triple negative breast cancer. *J. cancer Ther. Res.* **3,** 1–11 (2014).

131. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

132. Raney, B. J. *et al.* Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30,** 1003–1005 (2014).

133. Ye, J. *et al.* Primer-BLAST : A tool to design target-specific primers for polymerase chain reaction. *Bioinformatics* **13,** (2012).

134. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44,** 3–10 (2016).

135. Bock, C., Reither, S., Mikeska, T., Paulsen, M. & Lengauer, T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21,** 4067–4068 (2005).

136. Nesset, K. A. *et al.* Frequent promoter hypermethylation and expression reduction of the glucocorticoid receptor gene in breast tumors Frequent promoter hypermethylation and expression reduction of the glucocorticoid receptor gene in breast tumors. *Epigenetics* **9,** 851–859 (2014).

137. Devonshire, A. S. *et al.* Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal. Bioanal. Chem.* **406,** 6499–6512 (2014).

138. Mouliere, F. *et al.* High Fragmentation Characterizes Tumour-Derived Circulating DNA. *PLoS One* **6,** (2011).

139. Vallone, P. M. & Butler, J. M. Autodimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37,** 226–231 (2004).

140. Skerra, A. Phosphorothioate primers improve the amplification of DNA sequences by DNA polymerases with proofreading activity. *Nucleic Acids Res.* **20,** 3551–3554 (1992).

141. Schöfl, G. *et al.* 2 .7 million samples genotyped for HLA by next generation sequencing: lessons learned. *BMC Genomics* **18,** 1–16 (2017).

142. Legendre, C. *et al.* Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. *Clin. Epigenetics* **7,** 1–10 (2015).

143. Fackler, M. J. *et al.* Quantitative Multiplex Methylation-Specific PCR Assay for the Detection of Promoter Hypermethylation in Multiple Genes in Breast Cancer. *Cancer Res.* **64,** 4442–4452 (2004).

144. Lehmann, U. *et al.* Quantitative Assessment of Promoter Hypermethylation during Breast Cancer Development. *Am. J. Pathol.* **160,** 605–612 (2002).

145. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci.* **93,** 9821–9826 (1996).

146. Maegawa, S. *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.* **20,** 332–340 (2010).

147. Bell, J. T. *et al.* Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. *PLoS One* **8,** (2012).

148. Widschwendter, M. *et al.* Epigenotyping in peripheral blood cell DNA and breast cancer risk: A proof of principle study. *PLoS One* **3,** 1–9 (2008).

149. Koestler, D. C. *et al.* Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol. Biomarkers Prev.* **21,** 1293–1302 (2012).

150. Iwamoto, T., Yamamoto, N., Taguchi, T., Tamaki, Y. & Noguchi, S. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. *Breast Cancer Res. Treat.* **129,** 69–77 (2011).

151. Moelans, C. B., Verschuur-maes, A. H. J. & Diest, P. J. Van. Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. *J. Pathol.* **225,** 222–231 (2011).

152. Lindqvist, B. M. *et al.* Whole genome DNA methylation signature of HER2-positive breast cancer. *Epigenetics* **9,** 1149–1162 (2014).

153. Hill, V. K. *et al.* Genome-Wide DNA Methylation Profiling of CpG Islands in Breast Cancer Identifies Novel Genes Associated with Tumorigenicity. **71,** (2011).

154. Noetzel, E. *et al.* Promoter methylation-associated loss of ID4 expression is a marker of tumour recurrence in human breast cancer. *BMC Cancer* **8,** 1–11 (2008).

155. Miyamoto, K., Fukutomi, T., Akashi-tanaka, S., Hasegawa, T. & Asahara, T. Identification of 20 genes aberrantly methylated in human breast cancers. *Int. J. cancer* **116,** 407–414 (2005).

156. Kitchen, M. O. *et al.* Methylation of HOXA9 and ISL1 Predicts Patient Outcome in High-Grade Non-Invasive Bladder Cancer. *PLoS One* **10,** 1–12 (2015).

157. Ghoshal, K. *et al.* HOXB13, a Target of DNMT3B , Is Methylated at an Upstream CpG Island, and Functions as a Tumor Suppressor in Primary Colorectal Tumors. *PLoS One* **5,** (2010).

158. Li, W. *et al.* Detection of OSR2, VAV3, and PPFIA3 Methylation in the Serum of Patients with Gastric Cancer Wen-han. *Dis. Markers* **2016,** (2016).

159. Kalmár, A. *et al.* DNA hypermethylation and decreased mRNA expression of MAL, PRIMA1, PTGDR and SFRP1 in colorectal adenoma and cancer. *BMC Cancer* **15,** 1–14 (2015).

160. Irving, L. *et al.* Methylation markers identify high risk patients in IGHV mutated chronic lymphocytic leukemia. *Epigenetics* **6,** 300–306 (2011).

161. Truong, M. *et al.* Even-skipped homeobox 1 is frequently hypermethylated in prostate cancer and predicts PSA recurrence. *Br. J. Cancer* **107,** 100–107 (2012).

162. Borssen, M., Palmqvist, L., Karrman, K., Abrahamsson, J. & Behrendtz, M. Promoter DNA Methylation Pattern Identifies Prognostic Subgroups in Childhood T-Cell Acute Lymphoblastic Leukemia. *PLoS One* **8,** (2013).

163. Nakamura, J., Kitajima, Y. & Kai, K. Expression of Hypoxic Marker CA IX Is Regulated by Site-Specific DNA Methylation and Is Associated with the Histology of Gastric Cancer. *AJPA* **178,** 515–524 (2011).

164. Ono, M. *et al.* Tumor-infiltrating lymphocytes are correlated with response to neoadjuvant chemotherapy in triple-negative breast cancer. *Breast Cancer Res Treat* **132,** 793–805 (2012).

165. Silver, D. P. *et al.* Efficacy of Neoadjuvant Cisplatin in Triple-Negative Breast Cancer. *J. Clin. Oncol.* **28,** 1145–1153 (2010).

166. West, N. R. *et al.* Tumor-infiltrating lymphocytes predict response to anthracycline-based chemotherapy in estrogen receptor-negative breast cancer. *Breast Cancer Res.* **12,** (2011).

167. Minckwitz, G. Von *et al.* Definition and Impact of Pathologic Complete Response on Prognosis After Neoadjuvant Chemotherapy in Various Intrinsic Breast Cancer Subtypes. *J. Clin. Oncol.* **30,** 1796–1804 (2011).

168. Telli, M. L. *et al.* Phase II Study of Gemcitabine, Carboplatin, and Iniparib As Neoadjuvant Therapy for Triple-Negative and BRCA1 / 2 Mutation – Associated Breast Cancer With Assessment of a Tumor-Based Measure of Genomic Instability: PrECOG 0105. *J. Clin. Oncol.* **33,** 1895–1901 (2015).

169. Denkert, C. *et al.* Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy With or Without Carboplatin in Human Epidermal Growth Factor Receptor 2 – Positive and Triple-Negative Primary Breast Cancers. *J. Clin. Oncol.* **33,** 983–991 (2015).

170. Ignatov, T. *et al.* BRCA1 promoter methylation is a marker of better response to anthracycline-based therapy in sporadic TNBC. *Breast Cancer Res Treat* **141,** 205–212 (2013).

171. Siolas, D. & Hannon, G. J. Patient-Derived Tumor Xenografts: Transforming Clinical Samples into Mouse Models. *Cancer Res.* **73,** 5315–5320 (2013).

172. Marangoni, E. *et al.* A New Model of Patient Tumor-Derived Breast Cancer Xenografts for Preclinical Assays. *Clin. Cancer Res.* **13,** 3989–3999 (2007).

173. Ashkenazi, A. & Dixit, V. M. Death Receptors: Signaling and Modulation. *Science (80-.).* **281,** (1998).

174. Shivapurkar, N. *et al.* Aberrant methylation of TRAIL decoy receptor genes is frequent in multiple tumor types. *Int. J. cancer* **109,** 786–792 (2004).

175. Bonazzi, V. F. *et al.* Cross-Platform Array Screening Identifies COL1A2 , THBS1 , TNFRSF10D and UCHL1 as Genes Frequently Silenced by Methylation in Melanoma. **6,** (2011).

176. Ratzinger, G., Mitteregger, S., Wolf, B. & Berger, R. Association of TNFRSF10D DNA-Methylation with the Survival of Melanoma Patients. *Int. J. Mol. Sci.* **15,** 11984–11995 (2014).

177. Murphy, T. M., Perry, A. S. & Lawler, M. The emergence of DNA methylation as a key modulator of aberrant cell death in prostate cancer. *Endocr. Relat. Cancer* **15,** 11–25 (2008).

178. Bosch, L. J. W. *et al.* Decoy receptor 1 (DCR1) promoter hypermethylation and response to irinotecan in metastatic colorectal cancer. *Oncotarget* (2017).

179. Collado, M. *et al.* Senescence in premalignant tumours. *Nature* **436,** 642 (2005).

180. Malhi, H. & Gores, G. J. TRAIL resistance results in cancer progression: a TRAIL to perdition? *Oncogene* **2,** 7333–7335 (2006).

181. Liu, X., Yue, P., Khuri, F. R. & Sun, S. Decoy Receptor 2 (DcR2) Is a p53 Target Gene

and Regulates Chemosensitivity. *Cancer Res.* **2,** 9169–9176 (2005).

182. Holohan, C., van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **13,** 714–726 (2013).

183. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25,** 2227–2241 (2011).

184. Kaestner, K. H. The FoxA Factors in Organogenesis and Differentiation. *Curr. Opin. Genet. Dev.* **20,** 527–532 (2010).

185. le Lay, J. & Kaestner, K. H. The Fox Genes in the Liver: From Organogenesis to Functional Integration. *Physiol. Rev.* **90,** 1–22 (2010).

186. Wangensteen, K. J., Zhang, S., Greenbaum, L. E. & Kaestner, K. H. A genetic screen reveals Foxa3 and TNFR1 as key regulators of liver repopulation. *Genes Dev.* **29,** 904–909 (2015).

187. Shen, W., Scearce, L. M., Brestelli, J. E., Sund, N. J. & Kaestner, K. H. Foxa3 (Hepatocyte Nuclear Factor 3) Is Required for the Regulation of Hepatic GLUT2 Expression and the Maintenance of Glucose Homeostasis during a Prolonged Fast. *J. Biol. Chem.* **276,** 42812–42817 (2001).

188. Myatt, S. S. & Lam, E. W. F. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat. Rev. Cancer* **7,** 847–859 (2007).

189. Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H. & Katoh, M. Cancer genetics and genomics of human FOX family genes. *Cancer Lett.* **328,** 198–206 (2013).

190. Shah, S., Prasad, S. & Knudsen, K. E. Targeting Pioneering Factor and Hormone Receptor Cooperative Pathways to Suppress Tumor Progression. *Cancer Res.* **72,** 1248–1260 (2012).

191. Mills, I. G. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. *Nat. re* **14,** 187–198 (2014).

192. Zheng, L. *et al.* FOXA1 positively regulates gene expression by changing gene methylation status in human breast cancer MCF-7 cells. *Int. J. Exp. Pathol.* **8,** 96–106 (2015).

193. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43,** 27–33 (2011).

194. Carroll, J. S. *et al.* Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. *Cell* **122,** 33–43 (2005).

195. Hisamatsu, Y., Tokunaga, E., Yamashita, N. & Akiyoshi, S. Impact of FOXA1 Expression on the Prognosis of Patients with Hormone Receptor-Positive Breast Cancer. *Ann. Surg. Oncol.* **19,** 1145–1152 (2012).

196. Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. **121,** (2011).

197. Masuda, H. *et al.* Differential Response to Neoadjuvant Chemotherapy Among 7 Triple-Negative Breast Cancer Molecular Subtypes. *Clin. Cancer Res.* **19,** 5533–5540 (2013).

198. Katoh, M. Functional proteomics, human genetics and cancer biology of GIPC family members. *Exp. Mol. Med.* **45,** e26–9 (2013).

199. Kirikoshi, H. & Katoh, M. Up-regulation of GIPC2 in human gastric cancer. *Int. J. Oncol.* **20,** 1183–1187 (2002).

200. Kirikoshi, H. & Katoh, M. Molecular cloning and characterization of human GIPC2, a novel gene homologous to human GIPC1 and Xenopus Kermit. *Int. J. Oncol.* **20,** 571–576 (2002).

201. Kuang, S.-Q. *et al.* Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. *Leukemia* **22,** 1529–1538 (2008).

202. Bhattacharya, S., Pal, K., Sharma, A. K. & Dutta, S. K. GAIP Interacting Protein C-Terminus Regulates Autophagy and Exosome Biogenesis of Pancreatic Cancer through Metabolic Pathways. *PLoS One* **9,** 1–20 (2014).

203. Wang, L., Lau, J. S., Patra, C. R., Cao, Y. & Bhattacharya, S. RGS-GAIP – Interacting Protein Controls Breast Cancer Progression. *Mol. Cancer Res.* **8,** 1591–1601 (2010).

204. Dean, M. ABC Transporters, Drug Resistance, and Cancer Stem Cells. *J Mammary Gland Biol Neoplasia* **14,** 3–9 (2009).

205. Durig, J. *et al.* Expression of ribosomal and translation-associated genes is correlated with a favorable clinical course in chronic lymphocytic leukemia. *Blood* **101,** 2748–2755 (2003).

206. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science (80-.).* **357,** (2017).

207. Baylin, S. B. A decade of exploring the cancer epigenome — biological and translational implications. *Nat. Rev. Cancer* **11,** 726–734 (2012).

208. Kalari, S. & Pfeifer, G. P. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.* **70,** 227–308 (2011).

209. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70 (2000).

210. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

211. Razin, A. & Cedar, H. DNA Methylation and Gene Expression. *Microbiol. Rev.* **55,** 451–458 (1991).

212. Baylin, S. B. *et al.* Aberrant patterns of DNA methylation , chromatin formation and gene

expression in cancer. **10,** 687–692 (2001).

213. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529,** 110–114 (2016).

214. Yoshikawa, H. *et al.* SOCS-1, a negative regulator of the JAK / STAT pathway, is silenced by methylation in human hepatocellular carcinoma and shows growth-suppression activity. *Nat. Genet.* **28,** 29–35 (2001).

215. Cowey, C. L. & Rathmell, W. K. VHL Gene Mutations in Renal Cell Carcinoma: Role as a Biomarker of Disease Outcome and Drug Efficacy C. *Curr. Oncol. Rep.* **11,** 94–101 (2010).

216. Yang, D. *et al.* Repression of IFN Regulatory Factor 8 by DNA Methylation Is a Molecular Determinant of Apoptotic Resistance and Metastatic Phenotype in Metastatic Tumor Cells. *Cancer Res.* **67,** 3301–3309 (2007).

217. Aguilera, O., Fernández, A. F., Muñoz, A. & Fraga, M. F. Epigenetics and environment: a complex relationship. *J. Appl. Physiol.* **109,** 243–251 (2010).

218. Christensen, B. C. *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *PLoS One* **5,** (2009).

219. Kamińska, M., Ciszewski, T., Łopacka-szatan, K., Miotła, P. & Starosławska, E. Breast cancer risk factors. *Menopause Rev.* **14,** 196–202 (2015).

Appendix A

The 86 primer sets in the mDETECT multiplex

Table A.4.1 Locations, sequences and names of the primer sets in the mDETECT multiplex

Gene	Location	Probe Name	Primer Sequence
-**	chr10:43138416-43138530	Int5f	TAGTGATTGGTTATTTGGGCGCGGGGC
		Int5r	AAACGACATCCATCATCTCCCTCGACCC
-**	chr2:236737778-236737883	Int8f	CGTCGTAGTATTTGGCGGCGCGTTTC
		Int8r	AACGTACCTAATCCCCAAACCCACTCCT
ACVRL1	chr12:50587308-50587443	ACVRL1f	GGATGTGGGAGGTTCGGTTCGGGTG
		ACVRL1r	CCGCTCGCCCCTCGCTAAAACTACA
ALX1	chr12:84198385-84198501	ALX1f	AGGATAGTAGCGGTGAGTCGTTAGCGTT
		ALX1r	CGCTCCCACTTTTCTCCTTTCTCCCTCC
BARHL2	chr1: 90967815 - 90967900	mbBARHL f	TGGTTTTTTTCGGTTTTTGTTCGA
		mbBARHL r	ACTTTCTCCCAATTCCAATATCCA
BOLL	chr2: 198359331 - 198359461	mbBOLL f	GAGGGTTCGGTTTTGGGATTTTT
		mbBOLL r	ACCAATCCCAAAAACCGACTCT
BRCA1*	chr17:38530874-38530968	BRCA1 Af	GGTAACGGAAAAGCGCGGGAATTATAGA
		BRCA1 Ar	CCCACAACCTATCCCCCGTCCAAAAA
C1orf114	chr1: 167663228 - 167663361	C1Df	TTGAGGTAAAGGAGATTTCGGT
		C1Dr	ACATACGCCTACGCAAATTTTTA
	chr1: 167663398 - 167663508	C1Ef	TTCGGTGTTTGCGAAGGGTTA
		C1Er	TCACAACCAACACAACGACACTT
C5orf39**	chr5: 43076267 - 43076390	mbC5orf f	AATTGTTTGTTAGGGTTTTGTTTTAA
		mbC5orf r	CGACGCCAAAACGAATATTTATTTA
CA9	chr9:35666101-35666239	CA9 Af	TTTCGGGCGGGAGTATCGGGTTTTGTAG
		CA9 Ar	GCTCCTTTACCCCTTCTCGACCAACTCC
CARD11	chr7:3049955-3050087	CARD11 Bf	TTTAGGGGTTGTAGGGTCGGG
		CARD11Br	ATTTTACATTTCCCTCCCCCGC
CCL28	chr5:43433001-	CCL28 Af	AGGGTGGAGTTTTAGGTAGCGC

	43433128	CCL28 Ar	AACAACCCGCGATAAACTAAACC
CD38**	chr4:15389377-15389501	CD38f	GCGATTAAGGCGTATCGGTGGGTATTGC
		CD38r	AACACCACCCGACGAACTCTCGACTAAC
CDKL2	chr4:76774785-76774935	CDKL2 Af	ATTGGTCGAGTCGAGTCGTTAC
		CDKL2 Ar	ACAAAAAAACGCCTCCTAACGAA
CHST11**	chr12: 103376278 - 103376397	CHSAf	GGGCGGCGTGGGAATGAATTTT
		CHSAr	CTTTCCCTCGCACCCCTAAA
CRYM	chr16:21202786-21202934	CRYM Af	GCAGTGAGTGCTCGGGAGCCCC
		CRYM Ar	GGTTTTCATTTGTTAGAGGCGCGCG
DDAH2**	chr6: 31806097 - 31806189	mbDDAH f	TTTTTTGTTAGAAATCGGTATCGTTTA
		mbDDAH r	AAAATCTACGAAACGTTTACAACC
DMBX1	chr1: 46723889 - 46724003	DMBCf	TGTTTAGGAGATGGTTCGTGGT
		DMBCr	GCATCTACGCCGTCCGCAAC
DMRTA2	chr1: 50659325 - 50659469	pbDMRTA f	CGAAGATTTCGTAGGCGGGT
		pbDMRTA r	ACGACGCAAATAACGCTACGCA
	chr1: 50659366 - 50659489	DMRTAexp Af	GCGGCGGTTAGCGTTAGTTTTTCGGTAG
		DMRTAexp Ar	CGAAACGCCAACGTATCATAACGACGCA
DPP10	chr2:115635638-115635739	DPP10 Af	CGGATTGCGGGAAGAAGGTAC
		DPP10 Ar	GACGAAACGAAACCAAACGACAA
	chr2:115635947-115636088	DPP10 Bf	CGGTTTTCGAGTTTGAAGCGTTC
		DPP10 Br	TACCGACTCTCACCTAATCCGC
	chr2:115635821-115635943	DPP10 Cf	TTACGACGGGGAGTTCGTTC
		DPP10 Cr	CTTAACAACGTTCGCAAATCACGA
DRD4	chr11:627035-627175	DRD4 Af	CGTCGGGCGATGTTGGTTTGTTCGTG
		DRD4 Ar	GCGACGCTCCACCGTAAACCCAATATTTA
EFNA4	chr1:153310139-153310288	ERNA Bf	TGTGCGAAAGAGACGCGGGGTTTAGTTA
		ERNA Br	CCCGTAATCGCTAAAACATCCGCCCTTA
EPSTI1	chr13:42464253-42464369	EPSTI1 Af	TAGGGGAGGCGTCGAGTTC
		EPSTI1 Ar	ACTCGCTAAACGTCCCAACC
EVX1**	chr7:27248900-27249019	EVX1f	AGGAGGATGATAGTTTAGAAAGAAGAGGGT
		EVX1r	CGCGACCGCGACGATAACGATAAAAACT

FABP5**	chr8:82355078-82355157	FABP5f	GAAACGTGTAGGCGTCGGCGTTTATGAG
		FABP5r	CGACCTCTCGAACGCCTCCTACAAACAA
FOXA3**	chr19: 51071936 - 51072038	FOXAf	CGAGGTAGGAAGTTTTGCGG
		FOXAr	CGACTCCTCCCGCGAAATAA
	chr19: 51071734 - 51071824	FOXEf	CGTTTTGTCGGTTGCGTGTTA
		FOXEr	ATTCCCCGACCTACCCAAAAC
GABRA4**	chr4: 46690411 - 46690545	mbGAB Af	TTTAGCGTTTAATGTGTATGTAGA
		mbGAB Ar	CGAAATTACAATCGAAACAAACTTAC
	chr4: 46690248 - 46690398	mbGAB Bf	GATGTTTTGAGTAGGGTGCGAG
		mbGAB Br	AAACGAAAAAAACAAATTCCGCCT
GALR3	chr22:36550907-36551049	GALR3 Af	GGTTGTGGTGAGTTTGGTTTACGGGCG
		GALR3 Ar	CGTAAAACGCGACCACCGCCAACATA
GIPC2	chr1:78284199-78284289	GIPC Af	GTTTAGGGGTGGAGGTCGGGGTTTTGA
		GIPC Ar	CCGAACCCCGCGCAAATAAAAACAACCT
GNG4	chr1: 233880823 - 233880971	pbGNG f	ACGTAGTGTTGGTAAGATTTGTAGA
		pbGNG r	ACAAAAACCGCTTATAAACGACGA
	chr1: 233880677 - 233880817	mbGNG f	GTAGGTTTTTGCGTTGGAGATT
		mbGNG r	ATTTTCGTTACTTCTCTATTCCCAAA
HINF1B**	chr17: 33177654 - 33177773	HINF Cf	GTTAAGGTTTAGTTTCGAAATCGC
		HINF Cr	CTAAAAAACCGAACGATTCCCTAA
HOXA9	chr7: 27171666 - 27171765	HOXAAf	GTAATAATTTGGTGGTATCGGGGG
		HOXAAr	TCTACTAAACGAACACGTAACGC
	chr7: 27171878 - 27172029	HOXACf	TGGGGTTTGTTTTAATTGTGGTT
		HOXACr	GCGAAACCCGCGCCTTCTTAAT
HOXB13*	chr17:44157793-44157885	HOXB13 Af	TCGCGGGTTATAAATATTTGGTTGCGGC
		HOXB13 Ar	GACCGCCACTACCTCGAAAACATTTCCC
ID4**	chr6: 19945031 - 19945148	mbID f	GGAATTTTACGGATGGAGTGATG
		mbID r	CTTATCCCGACTAAACTACTAAAAAA
IRF4**	chr6: 336298 - 336405	mbIRF f	GTTATACGTTTTGTAAAGCGAAGTTT
		mbIRF r	AAACCAATCAATCACTAAACTACA
IRF8		IRF8 Af	TTACGATTTTTTTTAAGGTCGCGC

		IRF8 Ar	ACTATACCTACCTACCGCCGTC
ITPRIPL1**	chr2:96354715-96354857	ITPRIPL1f	TTTTGTACGTTGGGTTACGGGGGTTTGG
		ITPRIPL1r	TAAACGCGATAAACCCCTACGACCCCCA
LEF1	chr4:109307444-109307547	LEF1 Af	AGAGTTGGGGGCGGTATAGTTAGGGTGT
		LEF1 Ar	TTCAATCCCTACGACCCCAACGCCTAAA
MAST1	chr19: 12839386 - 12839533	MAST1 Af	GGAGAAGGTGGTCGGTAAGC
		MAST1 Ar	AAAACGTAATTATAAAAAACACGCC
NFIC	chr19:3386117-3386219	NFIC Af	CGTGGATACGAGTTTTGGCGGCGATTAT
		NFIC Ar	GCCACCAACCCTACCTCCTTCCATATCC
	chr19:3386234-3386379	NFIC Bf	TTTTTCGGTTTGAGTTATCGTGGCGGGA
		NFIC Br	CGAACCGTACTTCCAACCAAACGCAACT
NPHS2	chr1:177811817-177811891	NPHS2 Bf	TGGTGGAGTTGTGGATTGCG
		NPHS2 Br	TCCCACCCAAACCTCTCTCT
NR5A2	chr1:198278432-198278527	NR5A2 Bf	GGTAGGGTTTCGGTTGCGTA
		NR5A2 Br	TATTTCCCGAAAACTCCACATCCA
NT5E	chr6: 86215769 - 86215912	mbNT5E f	GTTGTCGGATTTTATTTTAATCGTG
		mbNT5E r	CTTAAACAAAAAAATCTCAAAAACTAAAA
OTX2*	chr14:56331673-56331781	OTX2 Af	AGGGATTGTATTTCGAGGTGGTCGAGGT
		OTX2 Ar	CCGACAAATCGAAACCTTCGCCCGAAAC
PAX6	chr11:31783206-31783322	PAX6 Af	GTTAATTTGGATGTTTCGCGTTTC
		PAX6 Ar	GTTTATCGCTACGACCCGACTAA
PDX1*	chr13:27396588-27396685	PDX1exp Bf	GGAAAAAGGAGGAGGATAAGAAGCGCGG
		PDX1exp Br	CTCGCCGAAAATCACGACGCAATCCTAC
PHOX2B	chr4:41447111-41447255	PHOX Af	GCGGACGTAGTAATGGATTAAACGGGGA
		PHOX Ar	AAATCCGACTCCCTACACTCCCGACTTT
POU4F1**	chr13: 78075512 - 78075652	POU4F1 Af	AAGTGCGTACGTCGGTTTATTC
		POU4F1 Ar	GCGACGCTCTACGCGATCAAA
PPFIA3	chr19:54337953-54338094	PPFIA3 Af	AGATACGGAGATTTAGCGCGAGATCGGT
		PPFIA3 Ar	AAATTAACCGCCGAACACTCACAATACG

PRDM13**	chr6:100168753-100168844	PRDM13f	AAGTTTCGTCGAGTTGGGGTCGTTGGTT
		PRDM13r	GACCCTTCCCGACAACCATCTCGAACA
PRKCB	chr16:23754928-23755052	PRKCB Af	TCGGTTCGTAGTTCGCGGTTTC
		PRKCB Ar	GCACGATACTCTCCTCGCCCT
	chr16:23755076-23755216	PRKCB Cf	CGGTAGAAGAACGTGTATGAGGT
		PRKCB Cr	GCTACCCTCGAAAACCCGAA
PRSS27	chr16:2705610-2705748	PRSS Af	GGGAGGTTATTCGTAGGATTTGGCGCGG
		PRSS Ar	ATCCTAACGACTACGCACTACTTCCGCA
PTGDR**	chr14:51804089-51804198	TGDRf	AAGAGGGGTGTGATTCGCGAGTTTAGAT
		TGDRr	CCGCGCGCGACTCGAACGAAAAA
PTPRN2	chr7:157176188-157176342	PTPRN2 Af	TCGGAGGAGATAAAGGTGTCGC
		PTPRN2 Ar	CCAACGTACCTAACCCGAAAAC
	chr7:157176379-157176463	PTPRN2 Bf	CGGACGGTTTCGGTAGGGTC
		PTPRN2 Br	GCGCCGAACCGAATATAAAACGA
SALL3	chr18:74841456-74841551	SALL3 Af	TGTTTTTTGCGGGTAAGCGTTC
		SALL3 Ar	GCCCACAACTCTCTCGACGAC
	chr18:74840051-74840163	SALL3 Bf	ATTTCGGGAAAGGGTGGGTC
		SALL3 Br	ACCCTAATCCCCCTTCACCA
SCAND3**	chr6: 28618249 - 28618359	pbSCAND f	TTAATTCGTTTCGCGACGTGAG
		pbSCAND r	ACACGCCTTAAAACCTACTCAT
SIM1**	chr6: 101019624 - 101019757	mbSIM Af	AGGCGTTTATTGGTTAATAGGG
		mbSIM Ar	CGACCCGACACCCTAAACTCAT
SLC7A4	chr22:19716858-19717005	SLC Af	GAGTTCGTTTAGTTCGTCGGCGTC
		SLC Ar	AACCCCGATAAACTCCGATAACGACCT
SOX2OT	chr3:182919884-182919969	SOX2OT Bf	GGGTTACGGTTTCGGGTTGT
		SOX2OT Br	CGCGTCGACTCTCAACACTA
SPAG6	chr10: 22674453 - 22674529	SPAG6 Af	GTCGAGTCGTCGTTACGATC
		SPAG6 Ar	CTACCCTCCTCGAACTCTACG
TAL1	chr1: 47470631 - 47470738	pbTAL f	GTATTGTCGCGGGTTCGTTC
		pbTAL r	CTCAACCAATCCCCACTCCC
TBX15**	chr1: 119331740 -	mbTBX f	GGTGAGGTTTTAGGTCGTGTGT

	119331881	mbTBX r	AAAACCTTAATCGACTCAAATAAAA
TCTEX1D1	chr1:66990668-66990782	TCTEX Af	CGGGGAGGGTCGAGGGTTTTGTTTGAG
		TCTEX Ar	GCGTCCCAAACTTCATTCAACCGACGAC
TMEM132C	chr12:127317663-127317786	TMEM Af	CGGGAGAAAAGTTGTTTCGGTC
		TMEM Ar	CCGCTACGTCTCTACTATCCGA
	chr12:127318043-127318179	TMEM Bf	GAGTTCGGGGTGAGGGTAGTC
		TMEM Br	GAATCCCGACGCCCAACTAAAAA
TMEM90B	chr20:24398648-24398747	TMEM90 Af	TAGGAAGGGGTCGATGTTGGTTTGGGTT
		TMEM90 Ar	TCTCACCAACTCCCATCGAATTCGCACA
TNFRSF10D**	chr8:23077397-23077475	TNFRSf	AAGGGAGGAGGGTGGATCGAAAGCGTTA
		TNFRSr	CGAAAACCTTTACACGCGCACAAACTACG
TOP2P1	chr6:28283447-28283568	TOP Bf	TGATTTGGGTGGATGTAGAGGTTGTGGT
		TOP Br	TTTCGAATAACGCTACTCCGAACCGCGA
TSPAN33	chr7:128596487-128596593	TSPAN Af	GGGGGTTGTGTTAGTTGTTTGTTTAGCGA
		TSPAN Ar	CGAAACTATTTCCCGCCAAACCGAACCC
TTBK1	chr6: 43319189 - 43319287	TTBAf	CGCGGTGTATTGTGGGTAGT
		TTBAr	CCTTCCGACCCGAATCATCC
UBD	chr6:29629437-29629564	UBD Af	TCGGTTGCGTAAATCGCGTTTTTGGTTG
		UBD Ar	TTCTCGATAATATCTCCGTCGCCTCCGC
VWC2	chr7: 49783871 - 49784003	VWCJf	TATTTCGGTTGTCGGGTTTGGA
		VWCJr	CCCTCAATCGCTCATCCTCC
	chr7: 49784151 - 49784279	VWCKf	TCGTCGGTCGGTTTAGGATG
		VWCKr	AAAACCGACGCCAAACCTACAT

- denotes probes in intergenic regions not associated with a particular gene, * indicates those probes developed for prognostic relevance in basal breast cancer, ** indicates those probes taken from the set of probes developed for the other subtypes of breast cancer

Appendix B

NGS reads for OTB TNBC tumours using the mDETECT multiplex protocol

Table B.1 All BiqAnalyzer exported reads for each probe in the mDETECT analysis of OTB TNBC tumours

Probes	A02354T	B02275T	A02282T	A02338T	A02397T	D03248T	E00700T	E00931T	E01003T	E01672T	E01684T	E02243T	F01103T	F01172T
ACVRL1	1	0	0	0	2	106	69	2	7	4	0	1	1	121
ALX1	51	295	18	5	173	571	670	109	66	1014	825	66	1	295
BRCA1-A	0	0	0	0	19	0	0	0	312	0	0	226	1	0
C1Dtrim	0	0	0	0	2	8	2	4	13	7	0	3	3	0
C1Etrim	0	22	16	1	61	115	87	215	215	221	1	185	49	35
CA9-A	0	0	0	1	1	3	0	0	0	1	0	1	1	0
CARD11-B	20	1	0	0	2	81	21	55	2	124	53	18	0	33
CCL28-A	123	13	55	21	54	284	972	358	1438	1832	225	1114	396	12
CD38	0	35	0	1	0	73	0	0	0	0	2	0	0	29
CDKL2-A	9	10	0	8	0	27	118	0	232	135	141	11	18	18
CHSAtrim	0	172	0	0	0	0	0	0	3	64	0	0	0	13
CRYM-A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DMBCtrim	13	20	3	2	5	136	43	56	61	69	115	2	127	26
DMRTA2exp-A	0	280	0	32	171	585	616	354	0	1257	681	217	513	237
DPP10-A	2	14	0	0	2	54	3	52	36	4	42	4	2	14

DPP10-B	0	100	0	1	8	182	1	271	83	0	325	2	0	87
DPP10-C	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DRD4-A	0	9	0	0	0	19	16	2	0	27	34	54	24	19
EFNA4-B	0	0	93	0	1	24	2122	15	4346	6437	1147	56	1373	609
EPSTI1	104	0	0	12	0	392	377	0	0	445	401	999	302	291
EVX1	0	0	0	1	0	27	0	1	0	1	1	0	0	12
FABP5	0	1	0	0	0	0	0	0	0	0	0	0	4	0
FOXAtrim	1	3	0	0	0	275	77	3	52	130	0	0	0	116
FOXEtrim	0	0	0	0	1	88	54	0	56	137	0	4	0	39
GALR3-A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIPC2-A	52	59	32	17	46	857	622	128	210	3261	161	506	700	295
HINF C trim	0	0	0	0	0	0	0	0	0	5	0	0	0	21
HOXAAtrim	124	5	0	128	42	817	7	537	0	1157	631	30	515	0
HOXACtrim	0	0	0	0	0	0	0	3	1	2	0	0	2	0
HOXB13-A	41	85	0	2	39	220	166	134	47	196	231	7	228	141
Intergenic5	13	11	3	1	0	21	25	0	2	28	41	92	13	34
Intergenic8	1136	4241	73	755	327	8898	783	750	455	12232	14298	292	7702	8159
IRF8-A	0	0	0	0	0	443	0	629	1202	0	0	894	0	571
ITPRIPL1	11	338	7	3	37	875	1696	23	1967	2304	1461	28	1657	1005
LEF1-A	1371	844	251	258	1053	2205	2731	2054	7204	6185	3650	2103	1733	2532
MAST1 A trim	2	1	0	0	0	7	22	0	3	35	46	0	33	14
mbBARHL2 Trim	3	5	0	2	6	21	11	7	1	9	14	14	1	5

mbBOLL Trim	1	3	3	0	6	7	17	0	0	19	17	0	0	0
mbC5orf Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbDDAH Trim	0	0	0	0	0	0	0	0	0	1	0	0	0	1
mbDMRTA Trim	2	41	0	10	28	105	80	71	5	247	17	40	93	31
mbGABRA A Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbGABRA B Trim	61	55	0	7	47	191	28	90	250	18	204	247	1	8
mbGNG Trim	0	1	0	0	0	0	2	0	0	0	0	0	0	0
mbID4 Trim	2	3	0	0	1	5	3	8	5	9	12	2	0	0
mbIRF Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	1
mbNT5E Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbSIM A Trim	0	9	0	0	4	0	9	4	14	3	0	2	0	1
mbTBX15 Trim	0	0	0	0	0	0	0	0	0	0	0	1	0	1
NFIC-A	182	0	163	48	108	523	761	988	1015	1172	666	689	680	234
NFIC-B	137	0	55	14	67	481	583	716	837	528	361	655	382	338
NPHS2-B	188	282	27	84	219	716	543	299	382	590	252	378	65	248
NR5A2-B	123	49	2	11	23	785	44	385	214	52	441	158	3	158
OTX2-A	2	3	1	1	1	3	4	4	13	10	6	8	7	4
PAX6-A	2	64	20	9	30	88	51	3	202	133	43	234	92	59
pbDMRTA Trim	0	177	0	12	95	510	542	358	0	1138	438	252	384	213
pbGNG Trim	0	0	0	0	0	0	0	0	0	0	1	0	0	0
pbSCAND Trim	56	25	0	0	1	133	19	124	23	14	0	6	0	26
pbTAL Trim	193	624	1	231	127	1674	1162	1100	1097	529	1336	1311	413	805

PDX1exp-B	6	9	0	10	2	31	14	23	49	34	15	55	15	26
PHOX2B-A	209	13	16	32	157	476	768	804	1206	1386	303	153	1320	371
POU4F1 A trim	0	1	0	0	0	0	1	0	0	0	0	0	0	0
PPFIA3-A	0	47	0	0	0	155	272	140	390	220	59	0	155	32
PRDM13	0	0	0	1	1	5	1	1	0	0	3	0	0	2
PRKCB-A	0	18	0	0	3	31	8	53	0	24	0	108	0	10
PRKCB-C	0	0	0	0	0	0	0	6	0	0	0	3	0	0
PRSS27-A	121	164	0	2	0	279	199	0	555	727	698	118	160	199
PTGDR	0	12	0	0	8	58	0	23	2	80	47	88	48	25
PTPRN2-A	31	0	0	11	22	204	6	483	3	131	81	43	0	26
PTPRN2-B	1	0	0	0	0	0	0	0	1	0	0	0	0	1
SALL3-A	13	25	22	4	17	110	51	32	32	28	7	38	6	10
SALL3-B	9	13	10	1	24	59	34	34	177	22	8	44	8	22
SLC7A4-A	0	4	0	2	3	48	97	7	56	76	5	0	71	27
SOX2OT-B	21	1	0	10	6	85	240	34	105	58	129	1	237	40
SPAG6 A trim	0	1	0	0	0	1	1	4	6	3	3	0	0	0
TCTEX1D1-A	36	47	16	5	14	122	52	52	173	91	147	101	16	48
TMEM-A	0	2	1	0	0	2	0	2	0	0	0	7	1	0
TMEM-B	18	11	0	1	14	25	0	29	8	3	13	2	0	9
TMEM90B-A	9304	10018	2001	3361	9803	13761	17483	20760	14468	23498	6137	10504	8161	13049
TNFRSF10D	1	0	1	8	11	262	42	4	8	4	592	5	708	3
TOP2P1-B	23	5	0	1	6	1	12	32	0	50	12	0	0	6

TSPAN33-A	39	132	12	9	31	90	391	126	621	926	550	611	380	253
TTBAtrim	0	355	0	1	8	3003	1008	22	5	4038	9	13	8	717
UBD-A	311	216	56	0	1	639	994	442	970	479	674	2	720	248
VWCJtrim	8	67	0	1	0	0	3	11	9	135	0	2	0	86
VWCKtrim	3	60	0	0	0	0	4	3	1	0	0	3	0	71
Total	14180	19121	2958	5138	12940	42082	36840	33039	40916	73799	37812	22813	29533	32192

Table B.2 80% methylated reads for each probe in the mDETECT analysis of OTB TNBC tumours

Probe	A02354T	B02275T	A02282T	A02338T	A02397T	D03248T	E00700T	E00931T	E01003T	E01672T	E01684T	E02243T	F01103T	F01172T
ACVRL1	0	0	0	0	0	104	22	0	0	0	0	0	0	92
ALX1	13	240	6	2	102	468	520	11	8	826	711	16	1	194
BRCA1-A	0	0	0	0	18	0	0	0	312	0	0	209	1	0
C1Dtrim	0	0	0	0	2	8	2	4	13	7	0	2	3	0
C1Etrim	0	20	16	1	55	100	58	213	191	217	0	122	18	34
CA9-A	0	0	0	1	0	3	0	0	0	1	0	1	1	0
CARD11-B	20	1	0	0	2	81	19	55	0	123	52	13	0	30
CCL28-A	123	0	35	16	38	253	445	358	1402	1810	222	549	281	12
CD38	0	35	0	1	0	69	0	0	0	0	0	0	0	27
CDKL2-A	9	10	0	7	0	25	91	0	213	133	127	9	18	15
CHSAtrim	0	113	0	0	0	0	0	0	0	14	0	0	0	11
CRYM-A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DMBCtrim	11	19	0	2	5	136	38	48	30	69	111	1	118	23
DMRTA2exp-A	0	280	0	32	164	575	612	353	0	1246	666	145	511	228
DPP10-A	0	12	0	0	0	46	0	52	17	0	36	4	0	5
DPP10-B	0	99	0	1	8	178	0	271	75	0	325	2	0	68
DPP10-C	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DRD4-A	0	8	0	0	0	19	14	1	0	27	32	36	24	19
EFNA4-B	0	0	41	0	1	9	1794	0	3299	6006	1045	32	1244	590
EPSTI1	104	0	0	12	0	392	354	0	0	430	399	982	302	286

EVX1	0	0	0	0	0	27	0	0	0	1	1	0	0	9
FABP5	0	1	0	0	0	0	0	0	0	0	0	0	4	0
FOXAtrim	0	0	0	0	0	112	29	0	8	92	0	0	0	116
FOXEtrim	0	0	0	0	0	85	52	0	50	133	0	3	0	39
GALR3-A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIPC2-A	7	5	11	2	1	845	516	5	8	3208	5	400	646	265
HINF C trim	0	0	0	0	0	0	0	0	0	5	0	0	0	21
HOXAAtrim	95	5	0	125	40	811	6	537	0	1144	621	29	510	0
HOXACtrim	0	0	0	0	0	0	0	3	0	2	0	0	2	0
HOXB13-A	37	65	0	1	18	105	93	99	0	141	205	1	172	124
Intergenic5	13	4	2	1	0	18	18	0	0	22	39	92	12	33
Intergenic8	971	3933	7	563	77	8742	165	158	32	12205	13190	104	7447	8073
IRF8-A	0	0	0	0	0	443	0	629	142	0	0	91	0	525
ITPRIPL1	0	266	0	0	0	748	1518	0	1422	2198	1420	0	1496	910
LEF1-A	529	41	0	47	11	1840	0	0	3569	3705	3248	0	132	1536
MAST1 A trim	2	0	0	0	0	3	17	0	1	35	38	0	24	13
mbBARHL2 Trim	3	4	0	0	4	19	8	6	0	6	13	3	1	3
mbBOLL Trim	1	3	3	0	5	7	17	0	0	19	17	0	0	0
mbC5orf Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbDDAH Trim	0	0	0	0	0	0	0	0	0	1	0	0	0	1
mbDMRTA	2	31	0	4	28	97	77	69	0	246	7	13	82	23

Trim														
mbGABRA A Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbGABRA B Trim	59	54	0	7	36	177	5	90	151	14	204	77	1	7
mbGNG Trim	0	1	0	0	0	0	0	0	0	0	0	0	0	0
mbID4 Trim	2	3	0	0	1	4	1	8	2	4	12	2	0	0
mbIRF Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	1
mbNT5E Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbSIM A Trim	0	9	0	0	3	0	8	1	14	0	0	0	0	1
mbTBX15 Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NFIC-A	182	0	161	48	108	518	740	980	1005	1170	661	603	662	233
NFIC-B	137	0	55	11	63	478	575	716	828	527	357	601	381	332
NPHS2-B	4	178	0	62	59	662	447	74	14	548	143	252	2	138
NR5A2-B	54	24	0	2	0	635	12	238	44	0	441	79	0	128
OTX2-A	1	2	0	1	0	3	3	4	13	10	5	3	6	0
PAX6-A	2	63	20	9	20	88	28	0	148	131	43	177	91	59
pbDMRTA Trim	0	177	0	11	95	507	526	358	0	1136	355	103	382	172
pbGNG Trim	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pbSCAND Trim	51	24	0	0	1	122	4	108	22	0	0	6	0	22
pbTAL Trim	153	87	1	81	56	1218	797	786	761	350	707	372	112	700

PDX1exp-B	6	9	0	10	2	31	14	21	45	34	15	55	14	26
PHOX2B-A	193	2	4	23	120	423	610	710	977	1332	267	38	1238	361
POU4F1 A trim	0	1	0	0	0	0	0	0	0	0	0	0	0	0
PPFIA3-A	0	38	0	0	0	147	270	136	380	217	54	0	148	32
PRDM13	0	0	0	1	0	3	1	0	0	0	0	0	0	2
PRKCB-A	0	18	0	0	0	31	0	53	0	16	0	64	0	8
PRKCB-C	0	0	0	0	0	0	0	6	0	0	0	1	0	0
PRSS27-A	107	148	0	2	0	261	149	0	448	701	667	57	145	189
PTGDR	0	7	0	0	7	55	0	23	1	67	44	77	46	17
PTPRN2-A	31	0	0	9	18	204	4	483	2	125	77	38	0	26
PTPRN2-B	1	0	0	0	0	0	0	0	1	0	0	0	0	1
SALL3-A	13	20	4	4	16	110	14	31	11	26	7	34	0	10
SALL3-B	4	13	0	1	11	46	7	19	24	13	8	18	3	22
SLC7A4-A	0	1	0	2	0	47	82	0	13	71	3	0	56	27
SOX2OT-B	21	0	0	10	4	85	205	34	94	56	126	0	219	40
SPAG6 A trim	0	1	0	0	0	1	1	4	6	3	2	0	0	0
TCTEX1D1-A	21	35	11	2	7	106	33	42	130	86	121	75	4	48
TMEM-A	0	2	0	0	0	2	0	1	0	0	0	6	1	0
TMEM-B	3	11	0	0	0	16	0	29	0	0	5	0	0	9
TMEM90B-A	4209	6360	13	1154	1239	11465	1922	10393	3236	1776	0	99	0	13049
TNFRSF10D	0	0	0	1	0	228	10	0	1	1	562	0	597	3
TOP2P1-B	20	5	0	1	1	1	7	32	0	28	12	0	0	6

TSPAN33-A	2	120	2	2	1	32	323	0	516	853	485	509	332	253
TTBAtrim	0	292	0	1	0	2999	921	6	1	4023	0	0	0	717
UBD-A	303	211	54	0	0	538	708	398	904	401	665	2	709	248
VWCJtrim	0	62	0	0	0	0	1	0	0	8	0	0	0	86
VWCKtrim	0	60	0	0	0	0	1	3	0	0	0	3	0	71
Total Methylated Reads	7519	13233	446	2273	2447	37611	14914	18659	20584	47799	28578	6210	18199	30370

Appendix C

Probes in the mDETECT Expanded Multiplex

Table C.1 mDETECT probes included in the expanded multiplex protocol according to miniplex group

	CT	△CT	Probe Name
Group 1	23.925	5.39	VWCJtrim
	28.722		C1Etrim
	29.168		DMBCtrim
	29.315		CARD11 - B
Group 2	31.489	1.53	IRF8-A
	31.617		VWCKtrim
	31.941		PRDM13
	32.325		TMEM - A
	32.331		ITRIPL1
	32.465		SPAG6 A trim
	32.604		Int5
	33.016		PAX6-A
Group 3	33.151	0.71	TOP2P1 - B
	33.275		CCL28-A

	33.311		LEF1 - A
	33.336		HOXAAtrim
	33.581		PHOX2B - A
	33.584		PRSS27 - A
	33.761		CDKL2 - A
	33.860		EFNA4 - B
Group 4	33.894	0.57	mbID4 Trim
	33.964		mbBARHL2 Trim
	33.964		pbDMRTA Trim
	33.972		TNFRSF10D
	34.149		GIPC2 - A
	34.305		OTX2-A
	34.356		PDX1exp - B
	34.461		SALL3-B
Group 5	34.537	0.34	HOXB13-A
	34.654		MAST1 A trim
	34.710		EPSTI1
	34.732		TMEM - B
	34.795		DMRTA2exp - A
	34.829		FOXEtrim

	34.856		SOX2OT - B
	34.878		mbBOLL Trim
Group 6	34.908	0.31	TCTEX1D1 - A
	34.917		SALL3-A
	34.972		pbTAL Trim
	34.999		TTBAtrim
	35.105		CA9 - A
	35.110		HOXACtrim
	35.121		PRKCB-A
	35.222		DPP10-B
Group 7	35.230	0.28	mbSIM A Trim
	35.318		mbGNG Trim
	35.332		NFIC -A
	35.415		SLC7A4 - A
	35.426		NFIC - B
	35.468		AVCRL1
	35.503		PPFIA3-A
	35.506		NR5A2 - B
Group 8	35.514	2.89	FABP5
	35.542		UBD - A

	35.60		DPP10-A
	35.93		PTPRN2 - A
	36.17		mbGABRA B Trim
	37.34		PTGDR
	38.32		DRD4 - A
	38.403		FOXAtrim
Group 9	38.768	-	pbSCAND Trim
	-		C1Dtrim
	-		PRKCB-C
	-		HINF C trim
	-		CD38
	-		CHSAtrim
	-		EVX1
	-		BRCA1-A

Appendix D

NGS reads for patient and control plasma samples using the mDETECT expanded multiplex protocol

Table D.1 All BiqAnalyzer exported reads for each probe in the expanded multiplex mDETECT analysis of patient and control plasma samples

Probes	ACVRL1	BRCA1-A	C1Dtrim
CNTL24	7	0	0
CNTL23	42	0	0
CNTL22	0	0	1
CNTL21	0	0	0
CNTL20	0	0	0
CNTL19	302	12125	136
CNTL18	0	0	0
CNTL17	0	0	0
CNTL16	0	0	0
CNTL14	0	0	0
CNTL13	0	8314	0
CNTL12	0	1	0
CNTL11	0	0	0
CNTL10	111	0	0
CNTL9	120	0	0
CNTL8	9	1	0
CNTL7	0	0	0
CNTL6	225	1	0
CNTL5	0	0	0
CNTL4	360	0	0
CNTL3	0	3	0
CNTL2	0	8	0
CNTL1	0	6	0
03-39	196	1	0
03-38	2366	1	2719
03-34B	84	1293	3
03-34A	37	6	2
03-30	166	2	0
03-28	120	8798	0
03-27	84	1	1771
03-26	95	0	0
03-25	59	355	2
03-23	56	45179	229
03-22	337	12787	0
03-19	0	0	2226
03-18	516	3	2
03-17	1493	3	22985

C1Etrim	CA9-A	CARD11-B	CCL28-A	CD38	CDKL2-A	CHSAtrim
1	0	5	0	373	0	0
0	0	9	0	24074	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
127	1	0	1	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	1	0	0
0	0	2	0	0	0	0
0	0	5	0	1	0	0
1	0	3	0	1	0	0
1	0	0	0	0	0	0
1117	0	2	0	1	0	0
688	0	1	0	0	0	1
0	0	6	0	0	0	0
0	0	3	0	1	1	1
0	0	0	0	0	47	0
1	0	1	0	0	0	0
987	0	4	0	1	0	0
0	0	0	0	0	0	0
3	0	1	0	1	0	0
3	0	0	0	0	0	0
0	0	10	0	843	0	0
3221	2	254	0	38963	76	0
1614	0	15	3	310	0	13
0	0	1	0	8	0	545
733	2	6	2	4970	1	0
1027	1	3	3	2	0	1972
1726	0	3	9	0	104	11999
9	0	9	0	324	0	0
165	0	197	2	64556	0	1084
136	0	431	6	58792	80	16234
152	1	192	2	38325	0	15427
16	2	6	0	17190	0	1
0	2	12	1	7119	0	0
6380	1	38	16	14755	46	330

DMBCtrim	DMRTA2exp-A	DPP10-A	DPP10-B	DRD4-A	EFNA4-B	EPSTI1
1041	0	0	0	0	0	0
2	0	0	0	0	0	0
1	0	0	141	0	0	1
0	0	0	902	1	0	0
0	0	0	0	0	0	0
1465	2	214	366	79	0	0
0	0	0	0	0	0	1
1	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	541	0	17	0	1
0	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
6351	0	0	0	0	0	0
458	0	0	0	11	0	0
3	0	0	1009	0	0	0
0	0	0	0	0	0	0
2	0	0	0	0	0	3951
2	0	0	0	0	0	0
4735	0	0	252	10	0	0
6	0	0	0	0	0	0
4	0	0	0	0	0	0
2	0	0	1	0	0	0
9401	0	0	1	0	0	0
13220	53	732	433	872	4	8347
10655	0	0	381	0	0	0
5372	0	0	336	51	0	0
4182	2	0	294	0	0	105
5073	0	7	137	34	0	231
4704	16	0	0	7	0	167
49	0	0	1	0	0	0
842	4	102	41	0	4	1
280	23	170	56	199	6	375
1402	34	131	239	218	0	192
24	1	0	0	43	1	370
2661	0	1	0	0	0	0
16900	10	555	629	195	2	3225

EVX1	FABP5	FOXAtrim	FOXEtrim	GIPC2-A	HINF C trim	HOXAAtrim
1	0	216	0	38135	0	0
0	0	0	0	114305	0	0
0	0	2	0	102	0	0
0	0	1	0	2	0	1
0	0	0	0	15071	0	0
4552	0	0	0	58269	0	186
0	0	0	0	57	0	0
0	0	1	0	6383	0	0
0	0	0	0	9402	0	0
1	0	11796	0	5817	0	0
0	0	1	0	10453	0	0
0	0	1	0	33830	0	0
0	0	1	0	7183	0	0
1	0	1	0	14916	0	0
2486	0	226	0	21742	0	0
0	0	0	0	27539	0	0
0	0	0	0	3117	0	0
1	0	1	0	21	1	0
0	0	1	0	3	0	0
0	0	0	0	751	0	1
2	0	0	0	14856	0	60
2	0	1	0	18454	0	0
0	0	0	0	21	0	0
2	0	0	0	61451	0	1
2752	55	12374	46	25379	38807	945
0	1	657	0	39100	0	132
1	0	2	0	56575	0	2
0	5	0	4	46421	2738	10
0	0	781	8	60846	0	1
15849	0	291	0	75717	5617	4012
0	0	1	0	51693	0	0
201	1	3	0	42588	7	855
58	3	253	8	60984	6841	15859
2328	6	2661	0	51827	0	950
0	3	392	0	34079	0	375
2226	0	988	8	19848	1714	4852
0	0	1967	14	45331	13249	0

HOXACtrim	HOXB13-A	Intergenic5	IRF8-A	ITPRIPL1	LEF1-A	MAST1 A trim
0	126	33	123	54	3651	0
0	1	49	1	0	230	0
0	0	0	1	0	100	0
0	0	0	0	2	3	0
0	0	0	0	7	102	0
8	2319	2057	0	328	17687	4
0	0	1	0	0	3	0
0	0	0	0	1	45	0
0	10	0	0	0	72	0
0	1	0	1	0	4	0
0	0	4	0	0	231	0
0	1	1	0	0	58	0
5	0	0	0	0	123	0
0	1	70	1	1	365	0
0	517	54	26692	105	142	0
0	1	0	3	0	405	0
131	1	0	0	0	251	0
0	6	2	1514	35	378	0
0	0	0	0	0	245	0
85	0	0	1	63	3	0
0	4	2	2	0	45	0
0	5	18	8	0	257	0
0	13	0	0	2	308	0
0	124	905	1	261	10307	0
1	20945	8305	23464	3020	18097	6
22	446	147	233	506	9993	0
8	1	1	0	0	652	0
58	603	648	6922	326	16422	0
26	892	369	4	517	32113	0
31	857	46	11146	150	25827	0
0	51	60	1	5	6381	0
2	837	2435	1665	174	20791	16
30	1270	666	9243	387	4953	38
30	2330	2062	0	223	34965	0
1	967	636	7	14	1652	0
33	2482	2915	8491	912	6859	0
0	13745	5253	3	1841	15526	0

mbBARHL2	mbBOLL Trim	mbGABRA B	mbGNG Trim	mbID4 Trim	mbSIM A	NFIC-A
5	25	265	0	0	0	1916
0	0	17927	1	274	0	1
0	0	1	0	0	0	1
0	0	30	0	0	0	1
0	0	0	0	0	0	297
1648	18	938	3	917	86	101
0	0	0	3	0	0	982
0	0	2	0	0	0	1
0	0	4	0	0	0	174
0	0	875	0	1	0	3
0	0	368	0	0	0	1
0	0	8	0	0	0	2
994	0	0	0	0	0	4
11	0	0	1	0	34	2548
0	0	127	0	0	0	1160
0	0	89	0	0	0	1228
2	0	1	0	0	0	0
0	0	1	0	1	0	462
0	0	1	0	0	0	0
0	1	2	0	0	0	302
0	0	14	9	0	26	3257
0	0	1220	0	0	0	211
0	0	86	0	0	0	3
0	0	3	0	1	1	7318
1414	215	1310	29	340	173	33160
2002	10	1	3	1	0	1659
2010	0	71	0	1557	0	0
4083	4	106	4	860	61	14471
629	88	232	2	56	1	23498
2944	10	928	9	175	0	3705
564	3	418	0	0	3	1741
213	37	26474	1	156	0	307
939	31	29530	18	520	10	107
0	10	10628	14	0	131	6752
0	6	333	0	1	5	5710
1	11	970	6	455	93	6841
1851	157	4119	26	0	12	18944

NFIC-B	NR5A2-B	OTX2-A	PAX6-A	pbDMRTA	pbSCAND	pbTAL Trim
0	1036	751	37	2432	10866	5753
12	10993	2	0	0	0	1
1	0	57	0	0	1	0
0	0	0	1	0	0	2
0	1	108	0	0	0	0
0	118	5230	31	1029	4892	746
0	0	0	0	1	1	5469
1	0	1	0	1	3	3241
0	222	0	0	1	0	2627
0	0	0	0	0	0	4
0	0	0	0	0	687	1
0	0	1	0	0	0	2
0	2	0	0	0	0	2
188	222	2	0	0	0	5366
223	1165	0	0	9403	0	8322
0	1	88	0	2	0	3950
0	0	0	0	0	0	5
0	0	1	0	0	0	32443
0	1	0	0	0	0	1
1	415	0	0	0	0	15433
258	5	0	1	0	1	1
1	1	1	0	0	2	4
0	1	0	0	0	1	4
0	0	110	0	0	0	1905
5378	3279	11179	2942	4685	7	12981
0	282	1225	178	1	0	19842
0	4	0	0	1	1	17028
2477	24	3656	55	3489	12327	4407
2179	403	2586	90	1769	4642	11106
416	842	3784	626	6550	0	21949
1051	184	1926	1	0	0	2
207	29959	1527	164	431	37458	4431
96	44969	796	480	1291	3853	5751
148	15331	0	171	0	33911	7
44	862	1915	124	14058	0	5
386	1696	5491	126	2	3	1608
1253	5272	7355	1540	3320	0	13283

PDX1exp-B	PHOX2B-A	PPFIA3-A	PRDM13	PRKCB-A	PRKCB-C	PRSS27-A
0	269	81	802	0	0	0
0	1	0	39804	176	0	0
0	0	1	0	0	0	0
0	0	0	5	0	0	0
3	3	0	1	0	0	0
902	380	0	1659	0	0	0
0	1	0	0	0	0	0
0	0	0	32	0	0	0
0	0	0	970	0	0	0
0	0	61	810	0	0	0
38	8	0	208	0	0	0
0	1	0	1	0	0	0
0	0	0	28	1	0	0
0	47	14	340	1	0	0
0	602	28	1050	395	0	32
0	0	1	642	11088	0	0
0	20	43	1	0	0	0
0	1	42	148	0	0	0
0	2	0	0	0	0	0
1	2	0	2	2	0	0
0	1	0	934	2971	0	0
0	1	0	4	1390	0	0
1	95	0	968	1	0	0
235	131	0	903	5	0	7
3807	3120	1007	2908	2707	762	432
186	1	94	1063	1153	0	7
0	3	82	341	8785	0	0
2231	24	202	128	2780	0	3
859	63	151	842	814	0	11
1604	1896	49	1071	940	886	158
1135	65	0	2732	2514	0	3
692	225	1	44843	9780	602	18
330	147	874	40396	12766	141	38
0	452	529	16011	5911	242	13
1048	63	73	2090	2768	0	0
1851	613	142	157	2820	939	10
8402	888	930	2946	144	0	18

PTGDR	PTPRN2-A	SALL3-A	SALL3-B	SLC7A4-A	SOX2OT-B	SPAG6 A
49	0	0	2	0	94	2
25	0	0	0	0	125	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1475	6	0	0
124	0	557	6087	29	424	3
0	0	0	1	16	0	0
0	0	0	0	0	166	0
0	0	0	0	0	0	0
43	198	0	0	0	0	0
0	60	0	0	0	0	0
0	0	0	0	0	0	0
0	0	660	2	0	0	0
0	0	0	15	0	0	0
0	0	0	0	211	157	0
0	0	0	2	1	0	0
0	0	0	72	0	0	0
0	0	0	1	97	0	0
1	0	0	85	0	1	0
0	0	0	0	124	0	0
0	0	0	1	513	0	0
0	0	1	1	0	1	0
0	0	0	2	0	0	0
0	0	1	1708	0	1	0
2202	634	375	796	497	9756	6
0	0	643	6715	0	150	0
0	0	0	69	0	2	0
35	0	36	4822	0	94	0
40	12	214	2342	0	395	0
45	34	1083	2578	244	323	11
26	0	0	967	0	0	2
313	48	72	1237	2	25	2
677	91	707	883	48	5344	6
575	130	2753	1	80	17	0
94	4	0	280	0	115	0
63	1163	776	1759	35	970	6
566	1894	1375	6110	287	1742	4

TCTEX1D1-A	TMEM-A	TMEM-B	TNFRSF10D	TOP2P1-B	TTBAtrim	UBD-A
14401	0	0	20558	0	4764	156
16	0	0	8604	0	20527	1218
3237	0	0	0	0	3	0
7	0	0	1	0	0	2
7738	0	0	30	0	1	36
37936	0	148	653	0	2461	552
5	0	0	0	0	6559	5
9	0	1	8958	0	1	2
3155	0	0	3	0	2755	93
2	0	0	27138	0	4	4520
4	0	0	6	0	0	3
1	0	0	1	0	9565	9
12	0	0	3	0	4	1
21233	0	0	4	0	4681	165
29904	0	0	10757	1	4	2771
49430	0	0	7	0	3	1
6	0	0	5	0	0	0
35	0	0	8	0	19199	1
2	0	1	2	0	0	0
37411	0	0	1	0	4	108
14	0	0	9	0	6308	80
22998	0	0	31388	0	2	35
18	0	0	10009	0	578	233
10571	0	0	14212	0	58	1
47966	0	622	38054	20	15651	16576
43774	0	0	11824	0	2604	259
38983	0	0	120	0	6	145
37293	0	38	139	0	9703	2403
52819	0	45	9126	1	4341	1765
48707	0	109	45390	7	8294	1488
28542	0	0	23737	1	20102	281
34384	0	85	58371	2	33807	19701
23710	1	407	101784	11	32280	30611
79622	0	211	2	0	19311	14744
34817	0	17	20210	0	2779	1102
53643	0	90	72133	2	10704	3139
45546	0	372	20273	0	21353	5872

VWCJtrim	VWCKtrim	Total
164	0	108194
2	0	238422
2	0	3654
1	0	962
0	0	24879
542	195	168638
1	0	13106
0	0	18850
1	0	19491
1	0	51841
1	0	20394
1	0	43490
2	0	9029
3584	0	61394
633	0	120188
1402	0	96911
456	0	4117
1	0	58626
0	0	350
2658	0	63720
0	0	29383
0	0	76023
3	0	12359
7562	0	128237
15406	3924	469779
9428	0	168713
1118	0	133926
2728	2	193307
372	1	234459
1	19	317009
190	0	144869
206	0	442770
1586	784	564858
802	252	375582
591	28	147148
1049	334	231732
3	3	340352

Table D.2 80% methylated reads for each probe in the expanded multiplex mDETECT analysis of patient and control plasma samples

Probes	ACVRL1	BRCA1-A	C1Dtrim	C1Etrim	CA9-A
CTRL 24	0	0	1	0	5
CTRL 23	0	0	0	0	0
CTRL 22	0	1	0	0	0
CTRL 21	0	0	0	0	0
CTRL 20	0	0	0	0	0
CTRL 19	7	11210	136	34	0
CTRL 18	0	0	0	0	0
CTRL 17	0	0	0	0	0
CTRL 16	0	0	0	0	1
CTRL 14	0	0	0	0	2
CTRL 13	0	8258	0	0	5
CTRL 12	0	1	1	3	0
CTRL 11	0	1	0	0	0
CTRL 10	107	1117	0	2	0
CTRL 9	84	519	0	1	0
CTRL 8	0	1	0	0	0
CTRL 7	0	0	0	1	0
CTRL 6	186	1	0	0	0
CTRL 5	0	0	1	0	0
CTRL 4	7	987	0	3	0
CTRL 3	0	3	0	0	0
CTRL 2	0	8	0	3	0
CTRL 1	0	6	0	3	0
03 39	0	1	0	0	0
03 38	0	0	0	1584	0
03 34 B	0	0	0	1584	0
03 34 A	82	2	0	603	0
03 30	82	2	0	603	0
03 28	0	8798	0	713	0
03 27	0	1	1670	1715	0
03 26	1	0	0	0	0
03 25	0	355	2	62	0
03 23	0	43866	151	13	0
03 22	179	12777	0	1	0
03 19	0	0	2226	0	2
03 18	175	3	2	0	2
03 17	1433	3	22981	6191	1

CARD11-B	CCL28-A	CD38	CDKL2-A	CHSAtrim	DMBCtrim	DMRTA2exp-A
0	372	0	0	703	0	0
0	0	0	0	2	0	0
0	0	0	1	0	0	141
0	0	0	0	0	0	902
0	0	0	0	0	0	0
0	1	1	0	0	574	0
0	0	0	0	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	458	0
0	1	0	0	0	0	0
1	0	0	1	0	0	0
0	0	0	0	0	0	0
0	0	5391	0	0	0	0
0	1	458	0	0	11	0
0	0	0	0	0	3	0
1	1	0	0	0	0	0
0	0	45	0	0	0	0
0	0	0	0	2	0	0
0	0	0	4719	0	0	252
0	0	0	0	5	0	0
1	0	0	0	0	4	0
0	0	0	2	0	1	0
3	0	6	0	0	9400	0
11	0	0	0	4	10549	0
11	0	0	0	4	10549	0
0	2	4943	1	0	4167	2
0	2	4943	1	0	4167	2
0	2	2	0	5	4736	0
3	9	0	99	3807	4680	16
7	0	315	0	0	18	0
192	2	58669	0	0	385	4
415	3	50415	73	3877	140	23
171	2	24702	0	9	617	31
6	0	85	0	0	3	1
9	1	7004	0	0	1	0
32	16	14755	46	0	15495	10

DPP10-A	DPP10-B	DRD4-A	EFNA4-B	EPSTI1	EVX1	FABP5
0	0	0	0	0	111	0
0	0	0	0	0	0	0
0	0	0	0	0	2	0
1	0	0	0	0	0	0
0	0	0	0	0	0	0
110	252	24	0	368	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	95
0	0	0	0	0	6727	0
17	0	0	1	0	7276	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	5	0	0
0	0	0	0	6757	0	0
0	0	0	0	0	6754	0
1009	0	0	0	0	0	2
0	0	0	0	0	3107	0
0	0	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	1	0
0	0	0	0	0	4	0
0	0	0	0	0	2	0
381	0	0	0	0	0	1
0	381	0	0	0	0	1
0	294	0	0	101	0	5
0	294	0	0	101	0	5
7	137	31	0	0	0	0
0	0	7	0	99	11836	0
0	1	0	0	0	0	0
77	41	0	2	1	0	0
24	42	179	6	375	58	3
131	239	197	0	192	2308	2
0	0	43	1	370	0	3
1	0	0	0	0	2212	0
554	624	169	1	2985	0	0

FOXAtrim	FOXEtrim	GIPC2-A	HINF C trim	HOXAAtrim	HOXACtrim	HOXB13-A
0	0	0	0	0	0	123
104	0	0	0	1	29	1
3	0	0	0	0	0	0
1	0	0	0	0	0	0
0	0	0	0	0	0	0
0	3840	0	129	4	427	1
0	0	0	0	0	0	0
0	0	0	0	0	0	43
0	0	0	0	0	1	0
1	0	0	0	0	1	0
0	0	0	3	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
0	1	0	1	0	73	0
0	342	54	25591	34	2	0
0	0	0	0	0	3	0
131	0	0	0	0	0	0
2	1	0	0	1	1	1514
0	0	0	0	0	0	0
0	1	85	0	0	1	63
9358	60	0	2	0	2	0
10828	0	0	0	1	0	7
0	1	0	0	0	0	0
0	0	4389	0	1	0	0
2	0	11	0	0	22	0
2	0	11	0	0	22	0
0	0	2273	2738	10	48	381
0	0	2273	2738	10	48	381
779	8	4766	0	1	26	757
45	0	56163	5617	3305	25	81
0	0	19567	0	0	0	0
3	0	24588	7	618	2	113
48	1	46017	5455	14874	26	11
1654	0	3556	0	872	0	1838
0	0	11028	0	375	1	763
166	7	7034	1714	4850	33	1052
1208	12	13714	13239	0	0	11544

Intergenic5	IRF8-A	ITPRIPL1	LEF1-A	MAST1 A trim	mbBARHL2	mbBOLL Trim
0	986	0	0	24	112	0
0	0	0	0	0	237	0
0	0	0	0	0	0	0
0	0	0	0	0	0	1
0	0	0	0	0	0	296
0	0	55	4	202	15	300
0	0	0	3	0	977	0
0	0	0	1	0	0	1
0	0	0	0	0	173	0
0	0	0	0	53	0	0
0	0	365	0	0	0	1
0	0	0	0	0	0	0
0	0	0	0	0	4	0
0	0	0	0	34	2536	188
0	0	0	0	0	0	1159
0	0	0	88	0	0	0
0	0	1	0	0	0	0
35	0	0	0	0	1	0
0	1	0	0	0	0	1
0	0	0	1	1	0	0
0	0	0	0	14	8	0
0	0	0	0	0	883	0
0	84	0	0	0	3	0
0	1	1	0	0	0	0
5	233	0	76	0	413	9
5	233	0	76	0	413	9
0	6922	0	2	0	2475	0
0	6922	0	2	0	2475	0
80	3	19	39	0	485	67
0	6464	0	4324	0	2663	4
0	1	0	0	0	39	0
1560	1662	36	16	16	147	33
118	7251	277	84	11	866	22
0	0	0	196	0	0	8
204	6	1	4	0	0	5
2551	8484	788	1578	0	1	0
4386	2	1762	9847	0	1664	157

mbGABRA B	mbGNG Trim	mbID4 Trim	mbSIM A	NFIC-A	NFIC-B	NR5A2-B
0	0	1916	0	763	0	37
3	0	1	12	3	1	0
0	1	0	0	0	0	0
0	0	0	0	0	0	1
0	0	0	0	0	0	0
2	526	7	101	0	46	1131
0	0	0	1	1	5453	0
1	0	1	0	1	3	4
0	0	0	1	2617	0	0
0	3	0	0	0	0	0
0	0	0	0	0	686	1
2	0	0	1	0	0	1
1	0	0	0	2	0	0
0	0	0	0	0	5352	0
223	555	0	0	9399	8291	0
1228	0	0	0	0	2	3900
0	0	0	0	0	5	0
0	0	462	0	0	0	0
0	0	0	1	0	1	0
0	302	1	415	0	0	0
26	2512	258	0	0	1	0
0	0	209	1	0	0	0
1	0	0	0	1	3	1
2	0	0	0	7318	0	0
0	0	1	0	1652	0	1
0	0	1	0	1652	0	1
32	0	0	0	13754	1749	0
32	0	0	0	13754	1749	0
0	0	0	0	19425	1336	398
920	3	161	0	3700	386	598
418	0	0	0	1352	881	181
22676	1	132	0	307	167	17803
24174	7	519	0	106	1	22865
8563	10	0	2	6689	147	11784
275	0	1	5	5710	44	860
968	6	299	93	6835	386	945
1243	14	0	0	16504	1211	4300

OTX2-A	PAX6-A	pbDMRTA	pbSCAND	pbTAL Trim	PDX1exp-B	PHOX2B-A
2428	10821	3953	0	255	81	465
0	0	0	0	1	0	1759
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	3	0	0	0	0	0
5	1006	1	583	380	53	0
0	0	0	0	0	0	0
0	0	31	0	0	0	0
0	960	0	0	0	0	0
0	2	0	0	61	809	0
0	3	0	204	0	0	0
1	0	1	0	0	0	0
0	1	1	0	660	2	0
47	14	338	1	0	0	0
437	28	462	393	32	0	0
0	0	1	0	11034	0	0
20	43	1	0	0	0	0
0	32344	0	0	42	1	0
0	0	0	1	0	0	0
0	15372	1	1	0	0	2
1	1	0	0	0	448	2964
0	1	3	0	0	0	0
95	0	6	1	0	0	0
0	0	0	0	1898	0	0
0	177	1	0	19378	153	0
0	177	1	0	19378	153	0
1943	54	3482	0	3129	1637	0
1943	54	3482	0	3129	1637	0
1757	90	1769	4618	11053	461	22
2965	626	6492	0	14334	1587	1197
5	1	0	0	1	858	0
1009	164	431	37296	1711	525	35
607	318	750	3474	357	236	4
0	171	0	33793	4	0	56
2	124	14026	0	1	915	55
5255	126	2	1	11	1846	537
5560	1537	3310	0	5577	8392	806

PPFIA3-A	PRDM13	PRKCB-A	PRKCB-C	PRSS27-A	PTGDR	PTPRN2-A
0	0	0	49	0	0	1
175	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	3	0	0
0	6	0	1	0	0	1
83	0	0	88	0	500	1258
0	0	0	16	0	2	0
166	4	0	0	8953	0	0
0	0	3139	0	0	2	0
0	21	198	0	0	0	0
0	60	0	0	0	0	3
0	1	0	0	1	0	0
0	7	0	2	0	4	1
0	21124	0	4	0	4680	163
0	211	157	0	26664	0	0
0	0	0	0	20686	0	0
0	0	2	0	0	4	0
0	0	0	0	0	3	0
0	1	0	0	1	1	0
0	0	0	0	0	0	123
0	0	0	0	0	510	0
1382	0	0	0	0	0	1
2	0	0	0	5	0	10002
0	0	5	0	7	0	0
94	501	973	0	0	0	0
94	501	973	0	0	0	0
202	102	2765	0	0	8	0
202	102	2765	0	0	8	0
151	3	811	0	7	40	12
41	401	552	886	125	45	34
0	2380	2500	0	0	26	0
1	36582	9260	294	16	313	48
851	33677	12075	126	18	663	37
527	11178	5265	242	0	566	130
73	1531	2764	0	0	93	4
142	138	2307	939	8	63	1163
929	5	118	0	16	491	1894

SALL3-A	SALL3-B	SLC7A4-A	SOX2OT-B	SPAG6 A trim	TCTEX1D1-A	TMEM-A
94	2	3143	0	0	13467	0
125	2	0	0	9	0	5027
6	0	0	0	0	1	0
1	0	0	2	1	0	0
0	1	36	0	0	0	0
0	263	0	4091	0	77	2
0	0	0	6558	3	1	0
2	0	0	0	0	0	0
2681	93	1	0	0	0	0
1	0	0	27105	0	0	4092
0	0	6	0	0	2	1
8	1	0	0	0	0	0
2	0	0	0	0	0	0
3566	0	0	0	0	0	0
14	1	0	2595	632	0	0
5	0	1	1	1312	0	0
0	0	448	0	0	0	0
3	0	0	8	0	0	1
0	0	0	0	0	0	0
0	37243	0	0	1	0	0
6	0	0	8	0	1	80
14222	0	0	17138	0	1	35
0	0	232	2	0	0	0
1	0	0	1	0	2263	0
612	1	0	90	0	20315	0
612	1	0	90	0	20315	0
10	0	0	94	0	12189	0
10	0	0	94	0	12189	0
214	808	0	392	0	6624	0
833	2382	165	317	11	33265	0
0	0	0	0	2	4	0
72	433	2	25	2	18896	0
707	869	0	5183	2	4099	1
2706	1	2	0	0	19361	0
0	0	0	114	0	18378	0
776	802	35	964	6	39156	0
1352	6049	287	1736	4	39671	0

TMEM-B	TNFRSF10D	TOP2P1-B	TTBAtrim	UBD-A	VWCJtrim	VWCKtrim
4720	155	163	0	0	0	0
2	1	0	0	0	0	0
1	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	1712	1	135	173	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	0	0	0	0	0	0
0	0	0	0	0	0	0
108	2594	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	5152	0	2	1	7498	0
0	3684	0	0	130	8640	0
0	3684	0	0	130	8640	0
16	6	0	2890	1010	1993	1
16	6	0	2890	1010	1993	1
33	7750	0	275	1397	305	1
9	40011	6	7566	1364	0	18
0	15732	1	3122	275	4	0
28	47607	2	1904	15877	5	0
354	73926	8	2397	23134	929	659
196	2	0	2086	10004	581	250
5	12460	0	2779	511	2	27
88	67510	2	6775	3022	906	309
312	11896	0	20552	4303	0	0

Total
44950
7495
158
914
344
29918
13015
9307
16396
40101
9599
25
694
51496
85104
39276
3765
34653
11
62283
16268
44730
10455
37952
69704
69704
72117
72117
81213
223633
47692
302217
387857
163998
75886
180089
260900

Appendix E

NGS reads for PDX Tumours using mDETECT

Table E.1 All BiqAnalyzer exported reads for each probe in the multiplex mDETECT analysis of PDX tumours

Probes	ACVRL1	ALX1	BRCA1-A	C1Dtrim
PDX36	0	56	0	0
PDX35	0	0	0	0
PDX34	2	1	55	0
PDX33	0	0	0	0
PDX32	0	0	0	0
PDX31	0	0	0	0
PDX30	0	0	0	0
PDX29	10	53	0	0
PDX28	0	0	9	0
PDX27	1	0	0	1
PDX26	0	0	0	0
PDX25	1	0	0	0
PDX24	0	0	38	1
PDX23	0	18	0	0
PDX22	0	19	55	1
PDX21	0	61	0	0
PDX20	0	0	1	3
PDX19	0	0	1	0
PDX18	0	0	0	0
PDX17	2	3	323	0
PDX16	0	1	0	0
PDX15	3	23	0	1
PDX14	0	59	0	0
PDX13	0	0	0	0
PDX12	1	1	6	0
PDX11	2	0	25	0
PDX10	2	2	0	0
PDX9	0	21	16	0
PDX8	0	1	149	0
PDX7	0	0	0	1
PDX6	0	0	0	0
PDX5	0	0	31	0
PDX4	1	0	45	0
PDX3	0	1	0	0
PDX2	0	85	0	0
PDX1	1	0	30	1

C1Etrim	CA9-A	CARD11-B	CCL28-A	CD38	CDKL2-A	CHSAtrim
3	0	0	80	0	0	0
6	0	0	215	0	5	0
10	1	1	519	0	12	0
6	0	0	337	0	2	8
0	0	0	243	0	3	0
5	0	1	90	0	0	0
10	0	0	409	0	6	0
4	0	0	29	0	1	0
0	0	0	9	0	0	0
11	0	2	750	0	4	0
0	0	0	305	0	0	0
4	0	5	715	0	5	0
8	0	15	0	0	0	0
0	0	1	426	0	0	0
15	0	0	570	0	11	0
6	0	0	109	0	4	0
24	1	17	0	1	0	0
6	0	0	452	0	8	0
0	0	0	105	0	0	0
19	0	0	825	0	26	0
0	0	0	172	0	11	0
7	0	0	244	2	7	61
4	0	0	40	0	2	0
3	0	0	89	0	0	0
2	0	0	40	0	0	0
0	0	1	236	1	3	0
0	0	0	48	0	2	0
4	0	0	7	0	5	4
0	0	1	1016	0	0	0
14	0	1	275	0	5	1
0	0	0	0	0	2	0
0	0	0	168	0	6	0
0	0	0	415	0	11	0
0	0	0	49	0	0	0
4	0	0	29	0	0	0
4	0	1	413	0	9	0

CRYM-A	DMBCtrim	DMRTA2exp-A	DPP10-A	DPP10-B	DPP10-C	DRD4-A
0	4	100	0	0	0	0
0	1	182	9	107	1	10
0	29	0	0	0	0	0
0	14	32	14	102	0	0
0	0	0	0	0	0	4
0	7	139	4	45	0	1
0	0	63	0	0	0	0
0	1	84	0	0	0	0
0	0	1	0	0	0	0
0	43	145	0	1	0	0
0	23	0	0	0	0	1
0	34	425	14	171	0	7
0	13	45	0	0	0	1
0	5	369	0	0	0	2
0	11	0	12	78	0	9
0	9	197	1	25	0	0
0	38	1425	11	121	1	8
0	0	0	0	1	0	0
0	0	0	1	0	0	0
0	0	2405	0	1	0	1
0	6	0	0	0	0	6
0	6	603	14	147	0	12
0	0	45	1	5	0	5
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	2	0	0	1	0	0
0	2	30	1	2	0	2
0	1	15	0	0	0	0
0	12	325	24	64	1	55
0	2	1	4	60	0	6
0	0	70	0	0	0	1
0	0	0	0	0	0	0
0	5	0	0	0	0	0
0	0	70	0	0	0	6
0	1	89	0	0	0	3
0	13	376	9	151	1	9

EFNA4-B	EPSTI1	EVX1	FABP5	FOXAtrim	FOXEtrim	GALR3-A
167	44	0	0	0	0	0
566	5	0	1	0	0	2
474	78	0	0	5	7	0
575	0	2	0	6	6	3
323	0	0	0	0	0	0
42	20	0	0	0	0	0
0	0	0	0	0	0	0
153	34	0	0	0	2	0
5	3	0	0	0	0	0
2	208	0	0	0	0	0
241	0	0	0	0	0	0
1	0	0	0	0	0	1
461	125	0	0	0	0	0
361	8	0	0	4	3	0
619	56	0	0	9	12	0
1	9	0	0	0	0	0
1874	175	0	0	11	15	0
156	4	0	0	0	0	0
134	0	0	0	0	0	0
4	0	0	0	0	0	0
632	85	0	0	0	0	0
531	63	4	1	4	3	0
12	11	0	0	0	0	0
1	0	0	0	0	0	0
10	0	0	0	0	0	0
40	58	0	0	0	0	0
41	13	0	0	3	1	0
39	0	0	0	0	0	0
645	0	0	0	0	0	0
512	57	0	0	2	0	0
1	12	1	0	0	0	0
10	37	7	0	0	0	0
143	101	9	0	0	0	2
155	7	0	0	0	0	0
127	14	0	0	0	1	0
0	56	0	0	12	16	0

GIPC2-A	HINF C trim	HOXAAtrim	HOXACtrim	HOXB13-A	Intergenic5	Intergenic8
50	0	60	0	9	0	1078
129	0	0	0	21	20	0
177	0	686	1	38	31	2621
6	0	90	0	0	23	2671
4	0	0	0	0	0	1
3	0	120	0	6	4	727
179	0	178	0	31	13	0
77	0	57	0	10	5	1519
13	0	16	0	5	0	0
559	0	366	0	118	42	5224
103	1	67	0	20	0	0
147	0	627	0	69	0	3259
165	0	37	0	13	32	1069
125	0	5	0	27	0	1958
8	0	1	1	22	12	1242
83	0	96	1	3	18	1426
567	28	389	0	83	49	9689
60	0	197	0	10	22	0
13	0	50	2	0	0	230
65	0	0	0	1	97	7
191	0	111	1	24	7	0
131	0	487	3	62	0	2586
49	0	0	0	4	0	251
1	0	39	0	1	0	0
16	0	0	0	0	0	2
25	0	0	0	11	0	0
0	0	0	0	9	0	352
13	0	5	0	0	4	152
32	0	606	3	37	18	5413
5	0	79	0	16	5	10
4	0	24	0	0	0	2
3	0	0	0	0	17	0
6	0	0	0	0	44	0
40	0	3	0	0	2	1
16	0	37	0	0	10	0
271	0	12	0	32	0	3529

IRF8-A	ITPRIPL1	LEF1-A	MAST1 A trim	mbBARHL2	mbBOLL Trim	mbC5orf Trim
1	112	146	1	0	0	0
1778	3	659	1	0	0	0
0	832	770	0	0	0	0
0	123	604	0	3	0	0
10	3	67	0	1	0	0
125	118	425	1	2	0	0
0	1	87	0	0	0	0
0	54	71	2	2	0	0
65	1	29	0	0	0	0
0	417	306	0	2	1	0
0	0	624	3	0	0	0
1	474	137	8	0	1	0
0	315	91	0	1	0	0
0	442	189	5	0	0	0
501	5	1042	7	0	0	0
218	38	364	1	0	0	0
0	619	4101	41	6	1	0
191	1	147	0	2	0	0
136	2	59	0	0	0	0
0	62	2780	0	3	3	1
0	2	176	0	0	0	0
890	444	416	0	3	0	0
0	5	20	0	0	0	0
0	6	89	0	0	0	0
19	0	30	0	0	0	0
411	0	82	0	1	0	0
56	9	155	0	0	0	0
3	13	102	0	0	0	0
0	3	1877	5	14	0	0
772	98	309	8	0	0	0
0	0	49	0	0	0	0
177	0	473	4	0	0	0
189	0	533	3	0	0	0
0	0	248	0	0	1	0
0	95	23	1	0	0	0
0	353	510	14	2	0	0

mbDDAH	mbDMRTA	mbGABRA A	mbGABRA B	mbGNG Trim	mbID4 Trim	mbIRF Trim
0	10	0	6	0	0	0
0	15	0	5	0	0	0
0	57	0	0	0	1	0
0	10	0	15	0	0	0
0	0	0	3	0	0	0
0	23	0	11	0	0	0
0	3	0	0	0	0	0
0	14	0	12	0	0	0
0	0	0	0	0	0	0
0	36	0	0	0	0	0
0	0	0	1	0	0	0
0	43	0	4	0	2	0
0	21	0	0	0	0	0
0	17	0	1	0	0	0
0	0	0	0	0	0	0
0	50	0	16	0	0	0
0	123	0	113	0	0	0
0	0	0	0	0	3	0
0	0	0	1	0	0	0
0	380	0	4	0	0	0
0	0	0	0	0	0	0
0	57	0	2	0	0	0
0	6	0	1	0	0	0
0	8	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	2	0	0	0	0	0
0	5	0	2	0	0	0
0	168	0	7	0	0	0
0	0	0	5	0	0	0
0	7	0	5	0	0	0
0	5	0	0	0	0	0
0	0	0	0	0	0	0
0	14	0	0	0	0	0
0	4	0	5	0	0	0
0	53	0	70	0	0	0

mbNT5E Trim	mbSIM A	mbTBX15	NFIC-A	NFIC-B	NPHS2-B	NR5A2-B
0	0	0	52	19	25	17
0	1	0	144	214	141	11
0	0	0	1	0	67	0
0	0	0	180	140	139	11
0	0	0	88	34	92	14
0	0	0	14	6	172	10
0	0	0	184	141	21	6
0	1	0	118	41	39	1
0	0	0	26	1	1	0
0	0	0	0	0	59	0
0	0	0	80	45	5	7
0	3	0	300	257	25	11
0	1	0	149	54	44	0
0	0	0	129	33	17	0
0	2	0	172	154	7	0
0	0	0	50	22	35	11
0	8	0	8	1	476	6
0	2	0	146	61	2	35
0	0	0	47	12	3	5
0	0	0	924	341	27	0
0	0	0	279	126	6	0
0	2	0	268	297	114	3
0	1	0	57	20	2	0
0	0	0	9	3	1	0
0	0	0	14	1	5	0
0	0	0	92	33	11	3
0	0	0	27	16	20	1
0	0	0	4	4	16	0
0	16	0	811	72	2	28
0	1	0	131	85	37	17
0	1	0	36	11	2	0
0	0	0	0	0	8	0
0	0	0	3	1	12	0
0	0	0	63	24	4	1
0	1	0	40	19	32	7
0	0	0	213	106	61	20

OTX2-A	PAX6-A	pbDMRTA	pbGNG Trim	pbSCAND	pbTAL Trim	PDX1exp-B
0	10	44	0	0	126	4
3	40	82	0	3	355	2
0	64	0	0	0	19	0
0	17	13	0	0	66	3
0	25	0	0	1	43	4
3	11	48	0	0	100	1
0	27	36	0	0	374	0
0	11	40	0	0	125	2
0	0	0	0	0	4	0
0	41	64	0	0	363	13
2	16	0	0	0	291	3
0	59	204	0	0	164	8
1	104	13	0	0	244	10
0	49	191	0	0	86	4
1	49	0	0	0	499	0
0	25	104	0	17	269	2
5	135	783	1	0	1164	56
1	13	0	0	4	293	5
0	0	0	0	0	0	3
6	1	1174	0	0	2	1
2	0	0	0	5	0	4
0	25	344	0	2	475	6
0	6	20	0	0	0	2
0	3	0	0	0	0	1
0	2	0	0	0	44	0
0	27	0	0	0	136	4
1	1	14	0	0	36	0
0	1	5	0	0	40	0
4	278	104	0	0	293	61
3	1	1	0	0	208	3
1	10	43	0	0	0	0
1	40	0	0	0	397	7
2	33	0	0	0	0	5
1	0	24	0	0	0	0
0	22	44	0	0	140	1
3	101	181	0	0	443	7

PHOX2B-A	POU4F1 A	PPFIA3-A	PRDM13	PRKCB-A	PRKCB-C	PRSS27-A
33	0	1	0	0	0	6
39	0	7	0	3	0	87
30	0	6	0	0	0	175
26	0	3	0	0	0	68
0	0	3	0	0	0	48
60	1	1	0	0	0	0
66	0	0	0	0	0	63
0	0	0	0	4	0	0
4	0	0	0	0	0	0
377	0	0	0	0	0	55
1	0	6	0	0	0	72
68	0	10	0	0	0	207
6	0	6	0	0	0	12
81	0	5	0	3	0	33
192	0	10	0	0	0	6
43	1	3	0	1	0	7
514	0	0	0	49	1	442
18	0	2	0	0	0	18
3	0	5	0	0	0	13
6	0	99	1	0	0	219
152	0	0	0	6	0	72
15	0	1	0	31	1	111
1	0	4	0	0	0	0
2	0	1	0	0	0	3
11	0	0	0	0	0	11
99	0	0	0	0	0	8
4	0	1	0	0	0	2
11	0	4	0	3	0	5
684	0	205	0	0	0	201
57	0	6	0	6	1	69
0	0	0	0	0	0	0
164	0	0	0	0	0	55
130	0	0	1	0	0	235
1	0	6	0	0	0	45
21	0	0	0	3	0	18
99	0	4	0	0	0	102

PTGDR	PTPRN2-A	PTPRN2-B	SALL3-A	SALL3-B	SLC7A4-A	SOX2OT-B
2	21	0	5	12	7	19
7	19	0	38	66	13	9
0	0	0	0	32	15	25
7	48	2	7	59	15	12
0	0	0	1	65	5	0
2	13	0	7	26	0	19
0	0	0	27	0	0	0
0	5	0	0	11	6	25
0	1	1	0	0	1	2
0	255	1	12	96	1	81
0	0	0	22	155	0	18
19	49	2	19	29	20	0
1	0	0	0	0	6	31
1	4	0	20	33	18	0
4	0	0	9	3	11	0
9	7	0	7	35	7	19
47	147	2	52	209	31	156
0	0	0	16	168	0	0
0	11	0	0	38	0	3
0	0	2	84	9	0	0
0	68	0	17	109	13	0
13	23	1	19	65	20	56
0	0	0	1	6	0	3
0	0	0	0	0	0	5
0	0	0	2	0	0	2
2	15	1	10	34	0	71
1	1	0	1	5	0	0
0	7	0	2	17	1	2
35	60	0	35	36	0	102
1	15	1	26	46	4	0
0	0	0	0	25	0	0
7	19	0	8	23	8	51
13	11	0	16	0	22	66
0	0	0	0	23	0	0
4	7	0	6	6	3	13
3	44	0	8	56	18	0

SPAG6 A trim	TCTEX1D1-A	TMEM-A	TMEM-B	TMEM90B-A	TNFRSF10D	TOP2P1-B
0	16	0	1	747	27	5
2	67	2	4	3587	19	2
0	112	0	0	1201	98	4
2	76	0	7	3620	67	3
0	29	0	0	391	0	0
2	9	0	2	1100	1	5
3	2	0	0	924	74	0
0	23	0	0	292	13	2
0	3	0	0	286	21	0
3	217	0	5	3421	49	22
0	58	0	0	536	30	2
0	93	0	3	2583	36	0
0	0	0	0	1185	0	0
3	2	0	0	728	3	4
1	63	0	0	1170	0	4
0	32	0	1	690	16	0
10	282	0	6	1964	2	17
0	38	0	0	1010	0	1
0	24	0	1	912	0	0
5	179	0	0	11599	0	3
1	24	0	4	3672	0	7
3	84	2	8	3647	24	0
0	14	0	0	164	0	2
0	4	0	0	42	0	0
1	13	1	1	132	0	0
0	14	1	1	284	7	2
0	10	0	0	308	5	0
0	6	0	1	194	10	1
9	0	0	4	3943	0	7
0	34	0	4	1726	0	1
0	8	0	0	117	0	0
0	1	0	0	1043	7	4
1	0	0	0	3344	0	9
0	8	0	0	1244	0	0
2	21	1	0	194	33	1
0	0	0	0	3192	0	6

TSPAN33-A	TTBAtrim	UBD-A	VWCJtrim	VWCKtrim	Total
20	337	46	0	0	3529
42	13	242	0	0	9005
19	0	216	0	0	8492
7	494	464	0	0	10208
8	1	0	0	0	1514
1	594	102	0	0	4229
6	0	193	0	0	3127
10	3	58	7	4	3035
1	0	0	0	0	508
31	2	941	0	0	14348
3	0	164	0	0	2905
89	0	211	0	0	10624
11	1	146	0	0	4476
2	91	140	0	2	5648
88	1	344	2	1	7111
17	0	0	0	0	4166
51	9	53	0	0	26222
17	0	91	0	0	3202
2	0	33	0	0	1848
421	0	1380	0	0	23495
24	13	99	0	0	6128
73	348	155	129	121	13303
0	0	7	0	0	830
1	0	0	0	0	312
4	0	13	0	0	384
3	0	77	0	0	1835
1	1	34	0	0	1222
3	11	4	0	0	763
137	0	106	0	0	17709
13	13	111	0	0	4869
0	0	1	0	0	433
2	0	83	0	0	2866
7	0	85	0	0	5503
3	0	0	0	0	2044
9	0	0	0	0	1192
60	621	1	0	0	11297

Table E.4.2 BiqAnalyzer methylated reads for each probe in the multiplex mDETECT analysis of PDX tumours

Probes	VWCKtrim	VWCJtrim	UBD-A	TTBAtrim	TSPAN33-A	TOP2P1-B
PDX36	0	0	46	337	20	5
PDX35	0	0	223	0	42	2
PDX34	0	0	169	0	17	4
PDX33	0	0	449	75	0	2
PDX32	0	0	0	0	6	0
PDX31	0	0	101	568	0	1
PDX30	0	0	189	0	0	0
PDX29	4	7	52	0	10	0
PDX28	0	0	0	0	0	0
PDX27	0	0	918	0	0	21
PDX26	0	0	118	0	0	0
PDX25	0	0	205	0	89	0
PDX24	0	0	129	1	4	0
PDX23	2	0	132	81	0	0
PDX22	0	0	325	0	86	4
PDX21	0	0	0	0	17	0
PDX20	0	0	48	0	36	17
PDX19	0	0	72	0	15	0
PDX18	0	0	27	0	2	0
PDX17	0	0	943	0	399	2
PDX16	0	0	89	13	20	5
PDX15	121	115	102	328	72	0
PDX14	0	0	0	0	0	1
PDX13	0	0	0	0	0	0
PDX12	0	0	13	0	3	0
PDX11	0	0	75	0	0	2
PDX10	0	0	32	0	1	0
PDX9	0	0	4	11	3	1
PDX8	0	0	106	0	137	7
PDX7	0	0	106	12	10	1
PDX6	0	0	1	0	0	0
PDX5	0	0	81	0	0	4
PDX4	0	0	79	0	1	9
PDX3	0	0	0	0	0	0
PDX2	0	0	0	0	9	1
PDX1	0	0	1	607	60	6

TNFRSF10D	TMEM90B-A	TMEM-B	TMEM-A	TCTEX1D1-A	SPAG6 A trim	SOX2OT-B
27	655	1	0	16	0	18
14	416	4	2	56	2	0
84	1	0	0	88	0	16
50	2499	7	0	67	2	11
0	0	0	0	24	0	0
0	465	0	0	8	2	17
66	0	0	0	0	3	0
12	0	0	0	20	0	25
16	121	0	0	3	0	2
13	46	2	0	177	2	79
26	0	0	0	52	0	18
32	908	0	0	78	0	0
0	0	0	0	0	0	31
0	1	0	0	0	3	0
0	0	0	0	49	1	0
14	250	1	0	27	0	19
1	0	2	0	244	9	153
0	89	0	0	34	0	0
0	281	0	0	23	0	3
0	1	0	0	34	1	0
0	2300	3	0	22	0	0
14	3114	4	2	67	3	55
0	0	0	0	13	0	0
0	0	0	0	4	0	4
0	6	0	1	12	0	2
5	0	0	1	13	0	69
5	125	0	0	8	0	0
8	149	1	0	5	0	2
0	895	4	0	0	7	92
0	1514	4	0	29	0	0
0	0	0	0	8	0	0
5	14	0	0	0	0	48
0	223	0	0	0	1	65
0	668	0	0	5	0	0
22	5	0	1	14	2	12
0	2761	0	0	0	0	0

SLC7A4-A	SALL3-B	SALL3-A	PTPRN2-B	PTPRN2-A	PTGDR	PRSS27-A
7	12	5	0	21	2	4
11	66	33	0	19	7	84
11	27	0	0	0	0	131
12	34	7	2	48	6	58
5	58	1	0	0	0	30
0	24	6	0	11	2	0
0	0	22	0	0	0	62
4	11	0	0	5	0	0
1	0	0	1	1	0	0
0	1	7	1	254	0	4
0	142	22	0	0	0	68
17	29	19	2	49	19	199
5	0	0	0	0	1	3
15	26	19	0	4	0	19
10	0	9	0	0	0	0
7	31	6	0	6	9	3
30	200	52	2	146	47	393
0	149	1	0	0	0	15
0	26	0	0	11	0	9
0	0	49	2	0	0	29
12	99	14	0	68	0	72
16	65	19	1	23	9	79
0	5	1	0	0	0	0
0	0	0	0	0	0	3
0	0	2	0	0	0	11
0	34	10	1	15	0	7
0	4	1	0	1	1	2
1	17	2	0	7	0	5
0	35	34	0	60	35	197
4	46	26	1	15	1	67
0	0	0	0	0	0	0
8	1	6	0	19	4	47
20	0	13	0	11	12	233
0	23	0	0	0	0	44
2	5	6	0	7	4	17
18	56	8	0	44	3	96

PRKCB-C	PRKCB-A	PRDM13	PPFIA3-A	POU4F1 A	PHOX2B-A	PDX1exp-B
0	0	0	1	0	33	4
0	0	0	7	0	4	2
0	0	0	6	0	3	0
0	0	0	3	0	3	3
0	0	0	3	0	0	4
0	0	0	1	1	47	1
0	0	0	0	0	33	0
0	0	0	0	0	0	2
0	0	0	0	0	4	0
0	0	0	0	0	332	13
0	0	0	6	0	1	3
0	0	0	10	0	31	8
0	0	0	6	0	6	10
0	2	0	5	0	79	4
0	0	0	10	0	144	0
0	0	0	2	1	39	2
1	49	0	0	0	469	56
0	0	0	2	0	7	5
0	0	0	5	0	0	3
0	0	0	86	0	0	1
0	5	0	0	0	87	4
1	28	0	1	0	4	6
0	0	0	3	0	0	2
0	0	0	1	0	2	1
0	0	0	0	0	10	0
0	0	0	0	0	86	4
0	0	0	1	0	2	0
0	1	0	4	0	9	0
0	0	0	205	0	654	61
1	5	0	6	0	19	3
0	0	0	0	0	0	0
0	0	0	0	0	160	7
0	0	0	0	0	113	5
0	0	0	6	0	0	0
0	1	0	0	0	3	0
0	0	0	4	0	97	7

pbTAL Trim	pbSCAND	pbGNG	pbDMRTA	PAX6-A	OTX2-A	NR5A2-B
101	0	0	44	10	0	14
245	2	0	60	40	3	3
0	0	0	0	43	0	0
37	0	0	5	17	0	0
22	0	0	0	25	0	0
87	0	0	48	8	2	0
58	0	0	8	19	0	0
71	0	0	40	11	0	1
0	0	0	0	0	0	0
95	0	0	55	41	0	0
156	0	0	0	16	2	6
71	0	0	204	59	0	2
161	0	0	12	102	0	0
27	0	0	191	49	0	0
324	0	0	0	46	1	0
199	17	0	103	25	0	2
957	0	1	782	134	5	0
16	2	0	0	13	1	29
0	0	0	0	0	0	0
2	0	0	1166	1	4	0
0	3	0	0	0	2	0
264	1	0	342	14	0	1
0	0	0	20	6	0	0
0	0	0	0	3	0	0
15	0	0	0	2	0	0
60	0	0	0	27	0	3
23	0	0	13	1	0	1
27	0	0	5	1	0	0
178	0	0	101	265	3	1
132	0	0	1	1	3	17
0	0	0	43	10	0	0
153	0	0	0	32	1	0
0	0	0	0	33	2	0
0	0	0	23	0	0	0
43	0	0	43	22	0	6
385	0	0	180	100	3	10

NPHS2-B	NFIC-B	NFIC-A	mbTBX15	mbSIM A	mbNT5E	mbIRF Trim
25	19	52	0	0	0	0
53	211	141	0	1	0	0
6	0	1	0	0	0	0
82	136	178	0	0	0	0
12	34	88	0	0	0	0
158	4	14	0	0	0	0
0	141	179	0	0	0	0
29	41	118	0	1	0	0
0	1	24	0	0	0	0
1	0	0	0	0	0	0
0	43	80	0	0	0	0
0	253	297	0	3	0	0
2	29	137	0	1	0	0
5	33	127	0	0	0	0
1	139	169	0	2	0	0
33	21	42	0	0	0	0
470	1	8	0	8	0	0
0	61	145	0	2	0	0
0	12	47	0	0	0	0
0	341	912	0	0	0	0
0	125	270	0	0	0	0
108	296	260	0	2	0	0
0	20	55	0	1	0	0
0	3	9	0	0	0	0
2	1	14	0	0	0	0
9	31	92	0	0	0	0
17	16	27	0	0	0	0
16	4	4	0	0	0	0
0	72	787	0	16	0	0
25	85	130	0	1	0	0
0	11	34	0	1	0	0
1	0	0	0	0	0	0
0	1	3	0	0	0	0
0	24	63	0	0	0	0
28	18	38	0	1	0	0
49	105	201	0	0	0	0

mbID4 Trim	mbGNG Trim	mbGABRA B	mbGABRA A	mbDMRTA	mbDDAH	mbC5orf Trim
0	0	6	0	9	0	0
0	0	1	0	8	0	0
1	0	0	0	40	0	0
0	0	10	0	9	0	0
0	0	2	0	0	0	0
0	0	10	0	23	0	0
0	0	0	0	0	0	0
0	0	10	0	12	0	0
0	0	0	0	0	0	0
0	0	0	0	32	0	0
0	0	0	0	0	0	0
1	0	3	0	39	0	0
0	0	0	0	17	0	0
0	0	1	0	17	0	0
0	0	0	0	0	0	0
0	0	13	0	49	0	0
0	0	111	0	120	0	0
3	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	313	0	1
0	0	0	0	0	0	0
0	0	2	0	56	0	0
0	0	0	0	5	0	0
0	0	0	0	4	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	2	0	5	0	0
0	0	1	0	167	0	0
0	0	5	0	0	0	0
0	0	4	0	7	0	0
0	0	0	0	5	0	0
0	0	0	0	0	0	0
0	0	0	0	8	0	0
0	0	1	0	3	0	0
0	0	55	0	51	0	0

mbBOLL	mbBARHL2	MAST1 A	LEF1-A	ITPRIPL1	IRF8-A	Intergenic8
0	0	1	130	112	1	1013
0	0	0	568	0	867	0
0	0	0	657	795	0	2562
0	3	0	427	0	0	2628
0	0	0	1	0	10	1
0	2	1	386	83	13	635
0	0	0	1	0	0	0
0	1	2	0	41	0	1444
0	0	0	24	1	26	0
1	2	0	4	349	0	4987
0	0	3	537	0	0	0
1	0	7	2	421	1	2724
0	1	0	3	254	0	1053
0	0	4	74	388	0	1941
0	0	7	953	2	489	1080
0	0	1	331	34	127	1188
1	6	36	3733	592	0	9522
0	1	0	6	0	0	0
0	0	0	0	0	80	227
3	2	0	738	0	0	7
0	0	0	9	0	0	0
0	2	0	11	307	532	1803
0	0	0	0	2	0	251
0	0	0	82	5	0	0
0	0	0	8	0	0	2
0	1	0	4	0	410	0
0	0	0	142	9	37	338
0	0	0	92	8	3	152
0	13	2	1696	0	0	5312
0	0	8	224	93	747	0
0	0	0	0	0	0	2
0	0	1	444	0	69	0
0	0	3	414	0	116	0
1	0	0	225	0	0	1
0	0	1	0	66	0	0
0	2	13	460	345	0	3244

Intergenic5	HOXB13-A	HOXACtrim	HOXAAtrim	HINF C trim	GIPC2-A	GALR3-A
0	7	0	58	0	50	0
0	0	0	0	0	118	0
8	35	1	686	0	164	0
1	0	0	85	0	0	0
0	0	0	0	0	0	0
3	2	0	116	0	0	0
5	28	0	177	0	176	0
2	4	0	57	0	76	0
0	2	0	16	0	13	0
24	72	0	361	0	554	0
0	16	0	61	0	98	0
0	54	0	616	0	143	0
6	3	0	37	0	165	0
0	19	0	5	0	123	0
12	12	0	0	0	0	0
18	1	1	93	0	83	0
48	65	0	374	27	563	0
16	0	0	186	0	57	0
0	0	1	45	0	10	0
12	0	0	0	0	0	0
3	9	1	82	0	189	0
0	45	3	481	0	126	0
0	1	0	0	0	49	0
0	1	0	39	0	0	0
0	0	0	0	0	15	0
0	10	0	0	0	25	0
0	8	0	0	0	0	0
0	0	0	5	0	13	0
16	36	3	606	0	2	0
1	11	0	74	0	0	0
0	0	0	23	0	0	0
17	0	0	0	0	0	0
39	0	0	0	0	0	0
0	0	0	3	0	39	0
0	0	0	35	0	15	0
0	20	0	0	0	268	0

FOXEtrim	FOXAtrim	FABP5	EVX1	EPSTI1	EFNA4-B	DRD4-A
0	0	0	0	43	158	0
0	0	1	0	5	537	9
5	2	0	0	76	407	0
6	1	0	1	0	485	0
0	0	0	0	0	294	4
0	0	0	0	18	36	1
0	0	0	0	0	0	0
2	0	0	0	33	140	0
0	0	0	0	3	3	0
0	0	0	0	204	2	0
0	0	0	0	0	203	1
0	0	0	0	0	1	7
0	0	0	0	110	336	1
3	1	0	0	4	311	2
11	4	0	0	55	469	8
0	0	0	0	9	1	0
15	4	0	0	169	1712	8
0	0	0	0	1	137	0
0	0	0	0	0	123	0
0	0	0	0	0	3	1
0	0	0	0	85	599	6
3	1	1	4	62	502	11
0	0	0	0	10	2	5
0	0	0	0	0	1	0
0	0	0	0	0	10	0
0	0	0	0	56	39	0
1	1	0	0	13	38	2
0	0	0	0	0	39	0
0	0	0	0	0	608	53
0	2	0	0	52	447	4
0	0	0	0	11	1	1
0	0	0	5	37	10	0
0	0	0	9	95	113	0
0	0	0	0	7	141	6
1	0	0	0	14	117	2
13	7	0	0	55	0	9

DPP10-C	DPP10-B	DPP10-A	DMRTA2exp-A	DMBCtrim	CRYM-A	CHSAtrim
0	0	0	99	3	0	0
0	103	7	180	0	0	0
0	0	0	0	14	0	0
0	101	7	22	4	0	0
0	0	0	0	0	0	0
0	38	2	134	5	0	0
0	0	0	62	0	0	0
0	0	0	83	1	0	0
0	0	0	1	0	0	0
0	0	0	141	32	0	0
0	0	0	0	22	0	0
0	160	0	423	31	0	0
0	0	0	45	11	0	0
0	0	0	353	5	0	0
0	63	0	0	11	0	0
0	12	0	194	6	0	0
1	121	10	1411	34	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	1	0	2350	0	0	0
0	0	0	0	4	0	0
0	126	14	597	4	0	12
0	5	1	45	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	1	0	0	2	0	0
0	0	1	30	2	0	0
0	0	0	15	1	0	2
0	28	0	316	12	0	0
0	58	1	1	1	0	0
0	0	0	70	0	0	0
0	0	0	0	0	0	0
0	0	0	0	4	0	0
0	0	0	70	0	0	0
0	0	0	83	0	0	0
1	151	9	375	12	0	0

CDKL2-A	CD38	CCL28-A	CARD11-B	CA9-A	C1Etrim	C1Dtrim
0	0	80	0	0	3	0
3	0	205	0	0	5	0
10	0	518	0	1	5	0
2	0	335	0	0	3	0
3	0	242	0	0	0	0
0	0	89	1	0	5	0
6	0	395	0	0	8	0
1	0	20	0	0	3	0
0	0	9	0	0	0	0
4	0	748	0	0	8	0
0	0	297	0	0	0	0
5	0	714	5	0	3	0
0	0	0	13	0	5	1
0	0	379	1	0	0	0
10	0	569	0	0	15	1
4	0	108	0	0	6	0
0	1	0	17	1	16	3
8	0	452	0	0	6	0
0	0	105	0	0	0	0
20	0	144	0	0	6	0
10	0	163	0	0	0	0
7	2	218	0	0	4	1
2	0	38	0	0	3	0
0	0	89	0	0	3	0
0	0	28	0	0	1	0
2	1	236	0	0	0	0
2	0	48	0	0	0	0
5	0	7	0	0	3	0
0	0	1015	0	0	0	0
4	0	270	1	0	14	1
2	0	0	0	0	0	0
3	0	142	0	0	0	0
10	0	410	0	0	0	0
0	0	46	0	0	0	0
0	0	29	0	0	2	0
9	0	410	1	0	3	1

BRCA1-A	ALX1	ACVRL1	Total
0	56	0	3308
0	0	0	4370
55	0	0	6649
0	0	0	7923
0	0	0	869
0	0	0	3179
0	0	0	1638
0	34	8	2438
9	0	0	281
0	0	0	9587
0	0	0	1997
0	0	0	7945
38	0	0	2739
0	15	0	4470
55	19	0	5165
0	59	0	3214
1	0	0	23043
1	0	0	1533
0	0	0	1041
250	0	0	7825
0	0	0	4373
0	9	1	10484
0	52	0	598
0	0	0	254
6	1	0	165
25	0	0	1356
0	0	2	955
16	21	0	676
149	1	0	13988
0	0	0	4284
0	0	0	229
31	0	0	1355
44	0	0	2081
0	0	0	1403
0	71	0	750
29	0	0	10344